UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS

School of Geography

Developing GIS Analysis Techniques for The Measurement of Safe Drinking Water Access

by

Shawky Abou Elghit Ali Mansour

A thesis for the degree of Doctor of Philosophy

Supervisors

Prof. David Martin

Dr Jim Wright

July 2011

Abstract

Geographic Information Systems (GIS) have provided effective and useful methods that are widely used to measure spatial access to services and to inform the planning of public facilities and infrastructures. The development of innovative GIS tools has informed approaches for researching demographic and socioeconomic problems. Nevertheless, GIS methods have not yet been developed for construction of water indices that can provide measures of household accessibility to safe drinking water at the finest spatial scales such as district and sub-district. There has, to date, not been any integrated measurement of access to safe drinking water that can be used as a practical tool to quantitatively identify small areas suffering from poor water supplies or households who use unimproved water sources. A comprehensive and structured review of the literature indicates that although there is international emphasis on drinking water problems in developing countries, there is a wide gap in relation to the current international measurements and programmes set up to measure and monitor access to safe drinking water. This includes efforts made by the United Nations (UN) and the Millennium Development Goals (MDGs) to measure and monitor access to safe water. These measurements are not only limited in terms of the socioeconomic factors impacting safe water access but also have not addressed measurement at the sub-national resolution level of residential communities and spatial variation within individual countries. Therefore, the available measurements neither provide a convincing picture of water accessibility nor a spatially detailed measurement. To bridge this gap, this study introduces new GIS analysis techniques for the measurement of access to safe drinking water as a global problem in the developing world. The main objective is to develop a multivariate index which measures current access to safe drinking water using datasets commonly available in lower and lower middle income countries. It draws on a wider range of data than current United Nations monitoring efforts.

GPS coordinates are increasingly collected as a part of household surveys, particularly in the developing world. This offers great opportunities to enhance national census data about drinking water by spatial linkage with other survey sources. Egypt was chosen as a case study and spatial linkage was undertaken between the 2005 Egyptian Demographic and Health Survey (DHS) and the 2006 Egyptian population census in order to develop indicators that reflect household access to safe drinking water. Understanding the spatial uncertainty associated with linking points (DHS GPS clusters) to polygons (sub-governorate census districts) was essential. Consequently, positional error relevant to DHS GPS clusters was detected, validated, measured and modelled. Selection criteria were developed for choosing the index components and a method for scoring these components was implemented. This was followed by standardisation and weighting of the components before combining them all into a single index. A map demonstrating index values for all Egyptian districts has been created. The index was evaluated against health factors to examine the influences of water accessibility on public health. The same types of datasets (DHS and local census) about another country (Jordan) were explored to examine potential index transferability. This was based on assessing the positional accuracy of the GPS clusters of the Jordanian DHS and evaluating the index construction. Potential limitations of this measurement were discussed and recommendations for further research suggested. Future policies options with a wider incorporation and implementation of GIS and spatial analysis methods were also considered.

Table of content

Abstract i

Table of content iii

List of figures ix

List of tables xiii

Acknowledgements xvii

List of Acronyms xix

Chapter 1: Introduction 1

1.1 Overview 1

1.2 Thesis theme 1

1.3 Aims and objectives 3

1.4 Thesis structure 4

Chapter 2: Literature review 9

2.1 Overview 9

2.2 Drinking water and health 9

2.2.1 The Seventh Millennium Development Goal 9

2.2.2 Measuring access to drinking water 12

2.2.3 Water supply interruption and quality issues 16

2.2.4 Sustainability of water supply and spatial coverage: rural and urban 17

2.2.5 Access to safe water and home treatment 17

2.2.6 Problems of measuring access to safe water 18

2. 3 Measurements issue and data sources: Demographic and Health Survey 19

2.3.1 What is the DHS? 19

2.3.2 The DHS questions about drinking water 19

2.3.3 DHS: the topics of interest 22

2.3.4 DHS applications using GIS and spatial analysis 24

2.4 Sampling strategy and GPS clusters in the DHS 29

2.4.1 Sampling strategy and design 29

2.4.2 Selected sampling methods 31

2.4.3 Sampling errors 31

2.4.4 GPS clusters of the DHS 32

2.5 Spatial measurements: methodology of GIS data linkage 33

2.5.1 GIS data linkage and uncertainty 33

2.5.2 Linkage of DHS clusters and administrative boundaries 34

2.6 Drinking water and health in Egypt 35
2.6.1 Water resources 36
2.6.2 Population growth and water demand 37
2.6.3 Informal settlement growth 38
2.6.4 Egyptian households living in the cemetery yards in Cairo 40
2.6.5 New city policies 43
2.7 Egyptian Demographic and Health Survey (EGYPTIAN DHS) 44
2.7.1 Sampling strategy in the 2005 Egyptian DHS 48
2.8 Summary 50

Chapter 3: Research Methodology [an evaluation of geospatial data linkage for measuring safe water access in Egypt] 51
3.1 Overview 51
3.2 Administrative Geographies 51
3.2.1 Census geography 51
3.2.2 Spatial characteristics of census boundaries 61
3.2.3 DHS Geography 63
3.2.4 Primary Data Collection and Field Observations 64
3.2.5 Linkage of DHS and census datasets 74
3.3 Spatial error and modelling 75
3.3.1 Spatial data validation 75
3.3.2 Spatial error of GPS clusters 77
3.3.3 Measuring spatial error 81
3.3.4 Analyzing spatial error 84
3.4 Analysing spatial errors when linking GPS clusters with shyakha geography 102
3.4.1 Creating a list of shyakhas sampled in the DHS 102
3.4.2 Geocoding of sampled shyakhas names 102
3.4.3 Overlaying the layer of shyakhas sampled for the DHS with the GPS cluster layer 103
3.5 Discussion and Conclusions 109
3.6 Summary 110

Chapter 4: Composite indicators for Safe Drinking Water Index (SDWI) – a review of the conceptual framework and background 113
4.1 Overview 113
4.2 Concept of area classification and indices 113
4.2.1 Overview of area classification 114

4.2.2 Overview of indices and their construction 115
4.3 Application of area classification 116
4.3.1 Applications of area classification in developed countries 116
4.3.2 Applications of area classification in developing countries......................... 118
4.4 Deprivation indices 119
4.4.1 Deprivation indices in developed countries 119
4.4.2 Deprivation indices in developing countries 121
4.5 Water index review 123
4.6 Classification or index?.. 128
4.7 Criteria for component selection.. 130
4.8 Key differences between census and survey data 131
4.8.1 Aims and targets 133
4.8.2 Time differences.. 133
4.8.3 Geographic levels and spatial linkage... 133
4.9 Summary 134

Chapter 5: Developing SDWI for Egypt at sub-national scale............................. 135
5.1 Overview... 135
5.2 Scope and domains of the SDWI 135
5.2.1 Water domain.. 139
5.2.2 Environmental domain... 144
5.2.3 Socioeconomic domain 145
5.2.4 Education domain 146
5.2.5 Demographic domain 146
5.2.6 Components excluded from the final index 146
5.3 Method for scoring the index components... 148
5.3.1 Water supply type indicator 148
5.3.2 Home water treatment indicator 150
5.3.3 Household water fetching indicator 154
5.3.4 Household water interruption indicator.. 154
5.3.5 Sanitation type indicator 155
5.4 Assessing the distribution of scores 156
5.4.1 Normal distribution 156
5.4.2 Bimodal distribution.. 158
5.4.3 Left-skewed distributions.. 158
5.4.5 Right-skewed distributions 160
5.5 Correlation matrix 162

5.6 Standardisation 163
5.6.1 Standardization methods 163
5.6.2 Results of standardisation process 165
5.7 Spatial distribution of the SDWI in sampled districts 166
5.8 Summary 169

Chapter 6: Estimating SDWI values for all Egyptian districts 171
6.1 Overview 171
6.2 Review of methods for estimating safe water access in unsampled districts 171
6.2.1 Small area estimation review 171
6.2.2 Review of areal interpolation methods 174
6.3 Estimating safe water access for all Egyptian districts 177
6.3.1 Accounting for DHS survey design in calculating kism level estimates 177
6.3.2 Tobler and Kennedy interpolation method 180
6.3.3 Resultant map of the SDWI for all Egyptian districts 183
6.4 Summary 196

Chapter 7: Evaluation and transferability of the Safe Drinking Water Index (SDWI) 197
7. 1 Overview 197
7.2 Evaluation of the Safe Drinking Water Index (SDWI) 197
7.2.1 Evaluation against Health outcomes 197
7.2.2 Index Evaluation through expert reviews 201
7.2.3 Summary of evaluation 202
7.3 The international context 203
7.3.1 DHS and census data availability in other lower middle income countries 204
7.3.2 International progress towards MDG Target 7C 207
7.3.3 International variation in factors affecting water safety 207
7.4 Transferability case study – Jordan 209
7.4.1 Justification for choice of Jordan 209
7.4.2 Jordanian data sets: census geography and DHS 210
7.4.3. Data linkage issues in Jordan 215
7.4.4. Index construction issues in Jordan 220
7.4.5 Options for evaluating a Jordanian index of safe drinking water 223
7.5 Summary 224

Conclusion **225**
8.1 Chapter overview 225
8.2 Main findings 225
8.2.1 The assessment of spatial uncertainty in the DHS clusters 225
8.2.3 The SDWI as a mean for measuring water access 230
8.3 Recommendations for further research 235
8.4 Policy implications 236
8.5 Limitations and research gaps 238
8.6 The future for monitoring safe drinking water access 239
Reference list **243**

List of figures

Figure 2.1 The DHS topics and spatial analysis application.. 24
Figure 2.2 GIS and spatial analysis application of the DHS.. 29
Figure 2.3 Population living in slum areas in Egyptian Governorates in 2005 40
Figure 2.4 Housing crisis and the problem of cemetery yard households in Cairo 42
Figure 2.5 The number of housing units that have been constructed by Cairo Governorate to accommodate Cairenes in the historical places during the 2002-2008 42
Figure 2.6 The hierarchy of Egyptian geographies based on the DHS, adminstrative governorates and regions.. 46
Figure2.7Method of selecting the DHS samples in the 2005 Egyptian DHS................ 49

Figure 3.1 The new Egyptian administrative boundaries since August 2008............... 52
Figure 3.2 The previous Egyptian administrative boundaries before August 2008 54
Figure 3.3 Egyptian sub-governorate boundaries showing kism/markaz boundaries .. 55
Figure 3.4 6th of October Governorate, showing kism/markaz boundaries 56
Figure 3.5 Helwan Governorate, showing kism/markaz boundaries 57
Figure 3.6 The new boundaries of Cairo Governorate in August 2008, showing kism/markaz boundaries 58
Figure 3.7 The new boundaries of Giza Governorate in August 2008, showing kism/markaz boundaries 59
Figure 3.8 The old boundaries of Cairo Governorate.. 60
Figure 3. 9 Distribution of total number of populations across Egyptian Governorates in 2006 61
Figure 3.10 Distribution of DHS GPS clusters by main Egyptian regions 64
Figure 3.11 locations of field study sites in Upper and Lower Egypt............................ 67
Figure 3.12 Picture illustrates a roof tank for storing drinking water used as a solution against water interruption 69
Figure 3.13 Picture shows a form of public drinking water provision, the electrical cooler (Coldair) which is used instead of the more traditional Zir vessel 71
Figure 3.14 GIS linkage process of DHS and census datasets................................... 75
Figure 3.15 Distribution of DHS GPS clusters within 2005 governorate boundaries.... 77
Figure 3.16 Flowchart illustrating how positional error in misallocated GPS points was detected.. 79
Figure 3.17 Positional error of GPS points in Alexandria Governorate 79
Figure 3.18 Distribution of positional error of GPS cluster points in all Egyptian Governorates.. 80

Figure 3.19 Flowchart representing the GIS process of point to polygon distance calculation in the Idrisi software 82
Figure 3.20 Histogram representing the distribution of distances for all misallocated points and frequency of positional error in metres 86
Figure 3.21 The probability of GPS clusters lying in the correct kism/markaz polygon according to the second logistic regression model 92
Figure 3.22 The frequency of cluster probabilities of being in the correct ksim/markaz according the second model 93
Figure 3.23 Map showing the probability of being in the incorrect kism/markaz 95
Figure 3.24 The probability of being in correct kism/markaz predicted by the fourth regression model 99
Figure 3.25 The frequency of cluster probabilities of being in the correct kism/markaz according to the fourth regression model 101
Figure 3.26 Geographical distribution of GPS clusters across Egyptian shyakha geographies 104
Figure 3.27 Geographical distribution of GPS clusters across Egyptian shyakha geographies using the ESRI ArcGIS online services "world Imagery function" 105
Figure 3.28 Geographical distribution of GPS clusters across Alexandria shyakh boundaries using the ESRI ArcGIS online services "world imagery function" 106
Figure 3.29 showing the distribution of distances of misallocated GPS clusters from shyakhas sampled for the DHS 107
Figure 3.30 Comparison of positional accuracy between census polygons and GPS clusters using the ERSI ARCGIS online services "world imagery function" 108

Figure 4.1 Developing a health domain in the Safe Drinking Water Index (SDWI) 116

Figure 5.1 Flowchart demonstrating the process of Safe Drinking Water Index construction 140
Figure 5.2 The relationship between percentage of Egyptian households living in flats and household wealth factors 147
Figure 5.3 Histogram of educational indicator score 157
Figure 5.4 Histogram of water interruption score 157
Figure 5.5 Histogram of household wealth component 158
Figure 5.6 Population density score 159
Figure 5.7 Histogram of home water treatment (theoretical range) 159
Figure 5.8 Stretched histogram of water quantity score 160

Figure 5.9 Histogram of water source type (census) score 161
Figure 5.10 Histogram of water source type (DHS) score 161
Figure 5.11 Histogram of sanitation score 162
Figure 5.12 Scores for the safe drinking water variables, standardised using the theoretical range for each variable 165
Figure 5.13 Distribution of SDWI values across Egyptian districts that contain DHS GPS 167

Figure 6.1 Sequence of steps to construct the SDWI taking into account the DHS design 179
Figure 6.2 Upper and lower 95% confidence limits for the mean values of the water interruption component index in kisms whithin Cairo governorate 180
Figure 6.3 Combination of Access and Excel steps to implement Tobler and Kenndy's areal interpolation method 182
Figure 6.4 Comparison of the interpolated versus measured values of the DHS derived component of the SDWI for 259 kisms and markazes 183
Figure 6.5 Map of the SDWI of Egypt at district level (quintiles) 184
Figure 6.6 Histogram showing the normal distribution of the SDWI values for Lower districts compared with Upper Egypt districts 186
Figure 6.7 The SDWI in Greater Cairo kisms 188
Figure 6.8 The SDWI in districts of Lower Egypt 190
Figure 6.9 The SDWI values in Alexandria kisms 191
Figure 6.10 The SDWI values in the Suez Canal region highlighting, with the urban governorate of Port Said as an inset map 192
Figure 6.11 The SDWI values in rural and urban districts of Upper Egypt governorates 193
Figure 6. 12 The average values of the DHS index components for different Egyptian regions 194
Figure 6.13 The average values of all index components (DHS and census) across Egyptian regions 195

Figure 7.1 Lowess smoothing line showing the correlation between diarrhoeal disease and combined DHS safe drinking water index components for 259 districts.............. 199
Figure 7.2 Lowess smoothing line showing the correlation between child diarrhoeal disease and the SDWI for 259 districts 200
Figure 7.3 Lower and lower middle income countries undertaking the DHS survey .. 206
Figure 7.4 Boundaries of Jordanian Governorates.. 211

Figure 7.5 Jordanian sub-governorate boundaries 212
Figure 7.6 Distribution of DHS GPS clusters by main Jordanian regions 214
Figure 7.7 Distribution of positional error of GPS cluster points in all Jordanian governorates 216
Figure 7.8 Distribution of misallocated GPS cluster points compared with Jordanian governorate boundaries using ESRI ARCGIS online world imagery 219
Figure 7.9 Distribution of misallocated GPS cluster points in the 2005 Ethiopian DHS compared with Ethiopian Governorate boundaries..220

List of tables

Table 2.1 Improved and unimproved water sources 11

Table 2.2 Representing the DHS questions about drinking water in the 2005 survey and before 22

Table 2.3 GIS and spatial analysis studies that have been done internationally so far using data from the DHS 25

Table 2.4 The characteristics of studies that published in academic journals using data from the Egyptian DHS... 47

Table 3.1 X & Y coordinate system in the initial command of rasterising process 83

Table 3.2 The number of the DHS GPS points falling outside governorate boundaries 85

Table 3.3 The output of Cuzick and Edwards' cluster test for the GPS points with spatial error 88

Table 3.4 The output of the first regression model 90

Table 3.5 The outputs of the second regression model with squared roots of distances 91

Table 3.6 The percentage of GPS clusters by the survey methodology codes GPS or AVE 96

Table 3.7 GPS cluster error in urban or rural areas 97

Table 3.8 The coefficients of the third regression model... 98

Table 3.9 The coefficients of the fourth regression model.. 99

Table 4. 1 Water indices, published internationally 126

Table 4. 2 Shows criteria selection for the SDWI components.................................. 132

Table 5.1 Safe water access index components, with associated data sources, weights and selection criteria 136

Table 5.2 The average of weighted score for each index component calculated from stakeholder weighting scheme 138

Table 5.3 Methodology for scoring water source types 151

Table 5.4 Methodology for scoring household water treatment................................ 153

Table 5.5 Methodology for scoring water quantity.. 153

Table 5.6 Methodology for scoring water interruption 153

Table 5.7 Methodology for scoring sanitation type.. 156

Table 5.8 Correlation matrix for the Safe Drinking Water Index component scores... 163

Table 6.1 Deterministic interpolation techniques definitions 175

Table 7.1 Matrix showing the correlation between the DHS index components and the child diarrhoeal disease variable 201
Table 7.2 Confusion matrix illustrating the number of GPS clusters in each governorate and those that have governorate labels inconsistent with their locations 217
Table 7.3 variable and their values in Egyptian and Jordanian censuses that could be used to characterise the SDWI domains 222
Table 7.4 Overall compliance of Jordanian utility piped water supplies with WHO guideline values (GVs) 223

DECLARATION OF AUTHORSHIP

I, Shawky Mansour declare that the thesis entitled “Developing GIS Analysis Techniques for the Measurement of Safe Drinking Water Access” and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- ~~none of this work has been published before submission~~, or [delete as appropriate] parts of this work have been published as: [please list references]

Mansour, S.A, Martin, D.J and Wright, J.A. (2010) Positional accuracy and spatial linkage for a new health dataset: GPS clusters in the Egyptian Demographic and health survey. In, Proceedings of the GIS Research UK (GISRUK) ,the 18Annual Conference 2010, University College London, London, UK on 14-16 April 2010. GIS Research UK (GISRUK), 6pp, 517-522.

Mansour, S.A, Martin, D.J. and Wright, J. (2011) Problems of spatial linkage of a geo-referenced demographic and health survey (DHS) dataset to a population census: A case study of Egypt. *Computers Environment and Urban Systems* (doi:10.1016/j.compenvurbsys.2011.04.001) (In Press)

Signed: …………………………………………………………………………..

Date:……………………………………………………………………………….

Acknowledgements

From the initial idea to print, this thesis brought me great opportunities to learn from and work with astonishing, talented, and generous people in the School of Geography, University of Southampton. Foremost I would like to express my sincere gratitude and thank my supervisors, Professor David Martin and Dr Jim Wright for their encouragement, support, guidance, and suggestions provided throughout the entire course of this thesis. I am forever indebted to your advice and lessons I learnt to overcome challenges and this always will be cause of admiration.

I am particularly grateful to Professor Fathy Abou Ayana, and Professor Fathy Abou Raddy from Alexandria University, Egypt and Dr Steve Pedley University of Surrey, for their help in assessing the weighting scheme of the SDWI. I would like also to thank the MEASURE DHS in the United States and the Central Agency for Public Mobilization and Statistics (CAPMAS) in Egypt for provision of datasets.

I am greatly in debt to all friends and colleagues of research groups in the School of Geography. During the long days and nights of research work, the postgraduates and staff provided company and entertainment. Particular thanks to Booker, Naill, Niladry, Dave, Rong, Mark, Jagan, Gary, Sam, Niamh, Eloise, Nicola, Chris, Simon, Trev and Hassan. I would like to thank all of you for the warm atmosphere and sense of community over the whole period of study. The values of your friendship will not be forgotten.

Special thanks go to my parents, sister, uncle and all members of my extended family for their emotional support and pursuing. I really have felt your love and initial faith in me. Last but not least, I would like to express my heart-felt gratitude to my beloved wife and my stars Ahmed and Sherif. You have really been a constant source of immeasurable love, concern, support and strength over the research period.

List of Acronyms

AIS	AIDS Indicator Surveys
ANC	Antenatal Care
BCM	Billion Cubic Metres
CAPMAS	Agency for Public Mobilisation and Statistics (Egypt)
CPS	Contraceptive Prevalence Survey
DHS	Demographic and Health Survey
EA	Enumeration Area
EGYPTIAN DHS	Egyptian Demographic and Health Survey
EPSEM	Equal Probability of Selection Method
ESIS	Egypt State Information Services
GIS	Geographic Information System
GEMS	Global Environment Monitoring System
GPS	Global Positioning System
GWQI	Global Water Quality Index
HWTS	Home Water Treatment and Storage
HIV	Human Immunodeficiency Virus
IDW	Inverse Distance Weighted
JMP	Joint Monitoring Programme
MARA	Malaria Risk in Africa
MAUP	Modifiable Areal Unit Problem
MDGs	Millennium Development Goals
MICS	Multiple Indicator Cluster Surveys
MWRI	Ministry of Water Resources and Irrigation
ORC	Opinion Research Corporation
PSU	Primary Sampling Unit
SDWI	Safe Drinking Water Index
UNICEF	United Nations Children's Fund
UN	United Nations
USAID	Unite States Agency for International Development
UTM	Universal Transverse Mercator
WHO	World Health Organization
WFS	World Fertility Survey
WPM	Water Point Mapping

Chapter 1: Introduction

1.1 Overview

This chapter provides an introduction to this research. The first section presents the research theme by introducing the central concept of developing GIS analysis techniques for the measurement of access to safe drinking water and the spatial linkage of two datasets (an international survey and a local census) to construct a new means of measurement. The effects of positional error on data set linkage are considered, focusing on the geography of the household survey. The following section of this chapter explains the thesis aims and objectives and finally the chapter outlines the structure of the thesis.

1.2 Thesis theme

The rapid development of Geographic Information Systems (GIS) and spatial analysis techniques during the last decade offers researchers great opportunities to analyse, model and measure the interrelationships between social and environmental phenomena within their places and spaces. In addition, the potential progress that has been made in collecting georeferenced data either by remote sensors or Global Positing Systems (GPS) particularly through international surveys provides a concrete platform to foster socioeconomic applications. Such resources help research into various global problems such as climate change, sustainable development, poverty, water scarcity and others that may impact on population health and human life. GIS has been presented as a technology which promotes problem-solving using spatial and attribute datasets concerning geographic phenomena (Longley et al. 2005). The growth of GIS methods has enriched spatial analysis research particularly with more focus on a quantitative approach, comprising sophisticated operations and processes (O'Kelly, 1994). GIS has unique capabilities in dataset integration which creates a clearer picture of a given research problem and provides powerful approaches for identifying relationships between human and environmental factors (Shamsi, 2005).

One of the most important developing data themes used within a GIS framework is socioeconomic and population-related data. This type of data relates to (groups of) household members or individuals distributed across space. Examples of this kind of data include censuses and surveys undertaken by local agencies and health authorities. GIS methods are used widely to analyse these data and investigate their interrelationships with other physical and environmental objects that exist within the same geographic space (Martin, 1996). Nevertheless, few GIS and spatial analysis approaches have been developed to model and research household access to safe drinking water. Furthermore, the monitoring approaches developed for the Millennium Development Goals (MDGs) by the United Nations (UN) and World Health Organization (WHO) to measure access to safe water in the developing world are limited to geographic representation at small scales. They do not incorporate socioeconomic effects on this phenomenon such as water quantity, water interruption and other household characteristics.

Although comprehensive research has been carried out on constructing drinking water indices (e.g. Sullivan, 2002: Sullivan et al. 2006; Lawrence et al. 2002; Mlote et al. 2002), no single water index exists which adequately measure water access at a sub-national scale (provincial or district level) covering the whole country. Consequently, the representation of spatial variations across one country is missing from such measurements. Similarly, the current measurements lack any integrated tool which quantifies and captures small areas where households have access only to inadequate and poor water sources.

In this study, data that are collected internationally (e.g. the Demographic and Health Survey DHS or other surveys) and locally such censuses are spatially linked to construct an integrated index. This index is internationally applicable where particular forms of survey that collect data about drinking supplies is undertaken. Spatial linkage has been recognized as one of the more important GIS applications (Longley et al. 1999). The integration of different datasets through spatial co-location provides powerful and reliable approaches to address problems associated with spatial and socioeconomic relationships.

GPS coordinates are increasingly available as spatial references for population surveys in the developing world, where high-resolution address and street mapping are absent. This potentially offers opportunities to enhance national census data by spatial linkage with survey sources. This thesis intends to develop a spatial linkage which focuses on linking census and sample survey data in a developing world setting in order to construct a Safe Drinking Water Index (SDWI). Egypt is the case study where the 2005 Egyptian DHS is linked to the 2006 Egyptian population census. The importance of assessment of spatial uncertainty in such a context and its impacts on data linkage are considered. More specifically, this is a form of spatial data manipulation named point-in-polygon where the point is the DHS GPS cluster while the census boundary is the polygon. The new approach developed is to identify attribute labels that were inconsistent with the locational data for the points and polygons. Information about feature attributes that are inconsistent with location is used to model uncertainty in the point-in-polygon operation undertaken. Accordingly, the spatial linkage process will also offer a novel approach for assessing and modelling the positional error of the DHS GPS clusters based on an examination of inconsistencies between location and attribute labels.

The theme of this research then is the development of GIS analysis techniques for the measurement of access to safe drinking water as a global problem. Practical GIS techniques are used to construct a quantitative, practical and spatial means for measuring household access to safe drinking water in the developing world that will offer a window into a more sophisticated and accurate measurement. These techniques then provide new instruments that enable a variety of datasets about drinking water and relevant socioeconomic and demographic factors to be incorporated into the measurement of safe water access.

1.3 Aims and objectives

In the new millennium, access to potable drinking water has become a central issue for the population in developing countries. Lack of access to safe water is a leading cause of disease and is thus a public health problem. According to WHO, 1.1 billion people globally lacked access to safe drinking water (WHO/UNICEF 2004). Partly as a result, diarrhoeal disease causes four million deaths every year among children under the age of five. Accessibility to safe drinking water can be associated with various geographic characteristics including spatial patterns of housing, rurality, distance to water source, supply networks, and urban growth. In addition, socioeconomic characteristics of households such as income, deprivation, and education are often correlated with potable drinking water access. Thus, GIS integration of different datasets representing these factors could have potential for measuring drinking water accessibility. The powerful tools and methods of GIS and spatial analysis techniques have been widely used in exploring the interrelationships between spatial and socioeconomic factors related to public health issues (Cromely & Mclafferty, 2002). Nonetheless, as will be shown in Chapter 2, there has been a scarcity of GIS linkage methods that could be used to integrate such data and thereby measure safe drinking water access at spatial micro scales. Therefore, the aims of this thesis are the following:

- To develop a methodology for GIS-based measurement of safe drinking water access at sub-national scale using both national census and international DHS survey sources.
- To use the resultant metric in an initial analysis of the relationship between safe drinking water access and health outcomes and assess to what extent the index could be transferred to other settings

The objectives

The following objectives are formulated to achieve the two aims above:

I. Review the literature on GIS and spatial analysis, monitoring of drinking water access, and index development to find relevant techniques and make appropriate assumptions to underpin the subsequent analysis.
II. Use GIS techniques to link Egyptian DHS and Census geographies as a case study.
III. Measure and model the spatial error that might occur related to GIS linkage of GPS survey clusters and census administrative boundaries.
IV. Identify the spatial patterns associated with drinking water supplies and piped water interruptions in Egypt to identify areas with the most and least safe water access.
V. Produce a SDWI for Egypt and a map of the index values at subnational level (kism/markaz).
VI. Examine whether there is a relationship between access to unsafe drinking water and infectious water-borne diseases, in particular diarrhoeal disease and evaluate the index through expert consultation.
VII. Examine the transferability of the constructed SDWI internationally through a case study of Jordan.

1.4 Thesis structure

The thesis will start by examining the evidence for the effects of drinking water on public health, covering the issue of access to safe drinking water in the light of the MDGs. More specifically, the weaknesses and limitations of the measurement of household water accessibility in developing countries will be investigated critically. Notably, the existing measurement of drinking water accessibility has not taken into account the geographic and socioeconomic factors associated with this problem. Therefore, a potential gap arises in the research literature about the problem of access to safe drinking water, particularly the absence of methods and techniques to construct an integrated measurement for this accessibility targeted at small communities.

Then the thesis presents the GIS methods used in this study to undertake the spatial linkage of the DHS to a national population census in order to construct the SDWI as a practical tool for measuring drinking water access. An approach for evaluating the spatial uncertainty with regards to the DHS geo-referenced data will be developed. Moreover, positional accuracy of the DHS GPS cluster locations at three census levels (governorate, district, and sub-district) will be investigated, modelled, and analysed. After that, the focus of the study will shift to the conceptual framework of the indicators for the composite SDWI, especially to set out selection criteria for the index components. This will be followed by the development of a methodology to score the index components and then standardise, weight and combine them into a single index.

Each methodological section begins with a review of the literature to explore relevant techniques which were used previously to develop water indices, other kind of indices or in wider GIS applications. Thus, there are two chapters involving comprehensive reviews followed by the methods developed within this particular research. As a more detailed outline, this thesis will be composed of eight chapters including this introduction and they are as follows.

Chapter one mainly discusses the overall research theme, introduces research aims and objectives and presents thesis structure.

Chapter two will present the results of a literature review, summarising the theoretical context and conceptualisation of drinking water and health. In addition it will highlight the crucial role of GIS and spatial analysis methods in this context. It will start by addressing the nature of access to safe water in general and according to the international policy of the MDGs specifically. It will be concerned with drinking water monitoring associated with Target 7C of the MDGs especially challenges of measuring drinking water internationally. Some of the international measurements for access to safe drinking water that are conducted by international programs such as Joint Monitoring Program (JMP) will be reviewed and the limitations of these metrics will be highlighted. Then the chapter will concentrate on data sources that could be used in the measurements, particularly the DHS. The nature of the survey, its questions about drinking water, it topics of interest and related applications using GIS methods will be discussed. In addition to this, issues of collecting georeferenced data in the DHS will be described, particularly the sampling strategy and spatial representation as GPS clusters. Methods of GIS data linkage will be reviewed to provide a clear understanding of spatial uncertainty in survey data. Finally, the chapter will close with the problem of drinking water in Egypt as a case study and thus issues relevant to the 2005 Egyptian DHS, such as characteristics of published studies that are based on the Egyptian DHS and the sampling strategies behind such a survey.

Chapter three will concentrate on methodology, particularly the evaluation of geospatial data linkage issues. It will start with a brief overview of the Egyptian DHS survey and census geographies. It will then describe GIS techniques and tools used to undertake spatial linkage. This will involve spatial data validation for the DHS GPS clusters and an approach to identify positional error of these clusters. A GIS process for measuring the spatial error will be discussed before analysing and modelling the error by fitting a series of logistic regression models at governorate level. Finally, it will describe the spatial linkage of the DHS GPS clusters to the lowest Egyptian census level (shyakha) to examine the magnitude of uncertainty in matching sub-district census boundaries to the GPS clusters.

The following chapters will focus on the construction of the SDWI. Chapter four will explain the conceptual background behind the index construction. In particular, a comparison will be made between index construction and area classification, so as to choose the appropriate method. The chapter will argue that drinking water indices constructed to date suffer from various drawbacks and weaknesses, particularly failure to fully incorporate spatial and socioeconomic factors in measuring access to safe drinking water. Similarly, in general the resolution of existing indices is generally regional or national and thus spatial variations among small communities within one nation cannot be detected. The chapter will end with developing selection criteria for the index components and exclude those that will not contribute to the measurement of safe drinking water access.

Chapter five will primarily focus on developing the SDWI at sub-national scale. It will determine the scope and domains of the SDWI. The ultimate goal will be to explain the methodologies behind the index construction particularly scoring, standardising and weighting the index components. Moreover, it will depict the spatial distribution of the index values for all Egyptian districts sampled as part of the DHS, paying attention to the variation between rural (markazes) and urban (kisms) districts.

Chapter six explains the estimation of SDWI values for all Egyptian districts, not solely those sampled as part of the DHS. It will describe the areal interpolation method which will be used to estimate the unknown DHS values based on the surrounding known values. It will discuss the importance of taking the survey design and weights into account before applying the interpolation method. Finally, it will generate a map of the SDWI values for the whole of Egypt and examine the spatial patterns and geographic variation of the index values across the whole country.

Chapter seven will evaluate the output of the SDWI against health outcomes using rates of child diarrhoeal disease. It will briefly present an evaluation of the index undertaken by two Egyptian experts. It will also investigate international issues concerning monitoring of access to safe drinking water and other global factors that might impact on future monitoring of water accessibility and availability in the developing world. Finally it will discuss issues of SDWI transferability through a case study of a selected country (Jordan). In this context, it will present a brief analysis of the spatial uncertainty in the Jordanian DHS GPS clusters. In addition, the characteristics of the Jordanian census geographies and the available census attribute data available as potential index components will be discussed.

The final eighth chapter will summarise the main finding of this research. It will draw conclusions and recommendations, particularly how the research may be developed further and how the ideas of this thesis could be achieved within a wider research context. The limitations of this research and the future for monitoring drinking water will be discussed, especially through the development of further GIS and spatial analysis techniques.

Chapter 2: Literature review

2.1 Overview

This chapter reviews the literature on the research problem and data sources. Three aspects are discussed. Firstly, access to safe drinking water is a key issue in population health and is the focus of Target 7C in the MDGs (Hunter et al. 2009; Valdivia & Para El Desarrollo, 2006). Approximately 1.1 billion people in developing countries lack access to potable drinking water. Therefore, unsafe drinking water is a major factor in the prevalence of water borne diseases such as diarrhoea which leads to higher rates of mortality. It is estimated that 1.5 million children under the age of five die every year. Water borne diseases are thus a major global health challenge (Parashar et al. 2003 & WHO/UNICEF, 2004). Secondly, the literature that reports research on this issue using spatial and non-spatial analyses was reviewed. Specifically I considered the importance of GIS tools and spatial analysis techniques to address the geographic and geodemographic factors that might be related to the problem. Thirdly, the factors associated with drinking water problem in Egypt are reviewed, particularly water resources; population distribution and density; slums and deprived areas in the capital; unplanned housing growth on the agricultural land and spatial factors related to drinking water quality supplies. Finally, the DHS as a data source generally and the 2005 Egyptian DHS in particular are discussed in detail. Applications concerning the usefulness of the DHS data in different contexts are reviewed.

2.2 Drinking water and health

2.2.1 The Seventh Millennium Development Goal

The MDGs were set in September 2000 by the UN Summit in New York as aims to be achieved by 2015. However, the expectation of achieving these goals varies from one country to another based on socioeconomic problems and economic development progress (Keyzer & Van Wesenbeek, 2006). The MDGs aim to decrease poverty rates and solve sustainable development problems that affect human development in the developing countries. There has been an increasing amount of literature on the MDGs, particularly for poverty and standard of living problems. The DHS provides an effective dataset for studying and analyzing the MDGs particularly for addressing the nature of health problems and health inequalities between geographic regions (spatial differences within one country) and population groups (Attaran, 2005).

The seventh goal of the MDGs focuses on “ensuring environment sustainability” while Target 7C as a part of this goal covers improving drinking water and basic sanitation “Halve, by 2015, the proportion of people without sustainable access to safe drinking water and basic sanitation” (UN, 2000). Sahn & Stifel (2003) analyzed data from the DHS to investigate six of the MDGs and identify progress in achieving those goals in the African countries. However, the study would have been more persuasive if the authors had included Target 7C of the MDG because of the importance of improving water supply and sanitation systems to reducing poverty rates and increasing quality of life. DHS data have been identified as a major source for researching and investigating progress towards the MDGs. Several studies have addressed health outcomes of population to build health information systems. Examples include: infant and child health including morbidity, mortality rates and prevalence of infectious diseases related to socioeconomic conditions (AbouZahr & Boerma, 2005 ; Fay et al. 2005). Access to potable and safe drinking water is an important component in human health, particularly in developing countries where it seems to be associated with geodemographic and socioeconomic circumstances of the communities. Therefore, it is expected that fresh water and piped supplies are related to developed and urban areas while contaminated and unsafe water are related to slum and rural areas. Likewise, using filters and treating drinking water are correlated with educated and rich households while poor access to clean water and using unimproved sources are related to deprived households and poor communities. As a result, households who do not have access to safe drinking water are more likely to be infected by water borne diseases such as cholera, typhoid and diarrhoea.

Although access to safe drinking water is a concern in almost all African countries, Sahn & Stifel (2003) have researched the progress that has been met in Africa regarding all the MDGs except the seventh goal that focuses on environmental resources and sustainable development. Access to unsafe drinking water has been seen by Fay et al. (2003) as a part of many other socioeconomic problems such as deprivation, literacy, slum areas, inadequate sanitation and poor population health. Therefore, Mathew (2004) pointed out that access to clean and safe drinking water as a target in the MDGs is related to many others targets and goals in the MDGs. The intersection between access to safe drinking water and other goals has been addressed in more detail elsewhere (Mathew, 2004).

According to the international policy of the MDGs access to improved and unimproved water is defined as the following quotation and table 2.1 (WHO/UNICEF, 2009; UN, 2003):

"An improved drinking-water source is defined as one that, by nature of its construction or through active intervention, is protected from outside contamination, in particular from contamination with faecal matter"

Improved	**Unimproved**
Piped water into dwelling, plot or yard	Unprotected dug well
Public tap/standpipe	Unprotected spring
Tube well / borehole	Small cart with tank/drum
Protected dug well	Tanker truck
Protected spring	Bottled water *
Rainwater	Surface water (river, dam, lake, pond,

* Bottled water is considered to be improved only when the household uses water from another improved source for cooking and personal hygiene; where this information is not available, bottled water is classified on a case-by-case basis

Source: WHO and UNICEF.2009, the Joint Monitoring Programme for Water supply and Sanitation

Table 2.1 Improved and unimproved water sources

Regarding water quantity and access distance, the JMP defines access to drinking water as "*the availability of at least 20 litres per person per day from an "improved source" within 1km of the user's dwelling". (*Shordt et al, 2004). However, one of the limitations with this definition is that it does not consider either households connected to piped water with regular daily, weekly or monthly interrupted water problems, or the water treatment processes of private or public sources. In the treatment process water quality might be affected. Another weakness is that walking time and distance for fetching water defined as 1km from household location is still a long distance. Therefore, fetching and storing water could be major issue for households and thus, water is more likely to be contaminated between the spatial location of collection and the point of use (Wright et al. 2004). Furthermore, there is an ambiguity related to rainwater collection as an improved source of drinking water whereby the catchment area of rainwater varies from one country to another according to environmental factors. In addition, variations can be found in techniques and approaches to collecting rainwater.

Many challenges are faced in achieving Target 7C of the MDG focusing on access to improved sources of drinking water, particularly in some regions such as Sub-Saharan

Africa. Thus, more effort is needed to meet the goals. The rapid population growth in general and on the margin of towns and cities in particular, increases the population who do not have access to safe drinking water and adequate sanitation. For instance, the number of people living in urban areas with no access to safe drinking water doubled in Sub-Saharan Africa between 1990 and 2004 while the number in rural areas in 2004 was five times higher than in urban areas (UNICEF &WHO, 2006). Achieving this goal is necessary since a lack of clean drinking water can lead to several diseases affecting children and younger people of poor households. Therefore, it is expected that the MDGs on health issues such as child mortality, maternal health, malaria, diarrhoea and other diseases are actually correlated with drinking water quality. As a result, any progress achieved in access to improved source of drinking water will definitely impact on other MDGs which relate to population health outcomes.

2.2.2 Measuring access to drinking water

Despite Target 7C in the MDG aiming to improve drinking water sources and increase the proportion of households who have access to safe drinking water and sanitation facilities (WHO& UNICEF 2006), it does not indicate any methods, techniques or approaches for measurement of drinking water access indicators. Furthermore, there has been vagueness and fuzziness in relation to the term “safe water” which needs to be more clearly specified. Including spatial factors that impact water accessibility requires accurate measurement of access to safe water sources. Previous studies have introduced methods to measure indicators of drinking water access such as Water Point Mapping (WPM) (Jimenez & Foguet 2009; Jimenez & Foguet 2008; WaterAid ODI, 2009); water access based on quality and quantity definitions (Devi & Bostoen, 2009); and measuring water system capability in maintaining and improving water supply infrastructures of urban areas, with respect to the percentage of the population with access to portable drinking water (Milman & Short, 2008).

Difficulties arise for many reasons, however, when an attempt is made to investigate measurement of drinking water access internationally. Firstly, developing countries vary in terms of database availability, capacity and government policy to achieve the MDG targets. Secondly, there is a lack of standard methods or techniques to measure water access and progress towards the drinking water target. Thirdly, variations have been found in socioeconomic and spatial factors impacting safe water access between developing countries. For instance, containers, tanks and vessels used for water storage vary from one country to another. Some of these simple technologies are appropriate and have significant impacts on microbiological quality of water (Sobsey, 2002). Therefore, researching and investigating drinking water is often undertaken based on country-specific data and information. Thus, many studies have conducted a country based survey or a case study to address improved water sources and supplies (O'Hara et al. 2008; Jimenez & Foguet 2007).

The JMP is undertaken by the WHO/UNICEF and aims to track the efforts which are made by government policies of developing countries to measure improvement in water supplies and sanitation systems. According to JMP the term safe drinking water refers to all connecting technologies of improved water sources such as public standpipes, protected boreholes with hand pumps, protected wells, protected springs and rainwater collection (WHO/UNICEF, 2009a). To be safe, all these water sources must give reasonable and ready access. Therefore, tanker trucks and bottled water are considered unsafe water sources. JMP implement the monitoring process by gathering information at global and national levels about drinking water using official sources particularly censuses and surveys such as DHS and Multiple Indicator Cluster Survey (MICS). The programme plays a key role in building an international database to monitor and track the progress of drinking water access in each country. This is achieved by harmonising the datasets and preparing the questions and indicators which are used in the national surveys. The information collected by JMP is used widely by international organisations to produce several official documents and reports, for instance the World Water Development Report and World Health Report, as well as information and reports on the general status of drinking water and sanitation sectors (WHO/UNICEF, 2009b).These reports are useful for decision makers as they provide comprehensive and appropriate guidance on drinking water access.

WaterAid is an international organisation that works on a non-profit basis to enable deprived people in developing countries to increase access to improved and safe water sources. The programme provides households with simple technologies to help them obtain safe water. The programme was launched in 1981 in the UK as a result of the Third World Conference that was held in the UK by the National Water Council (WaterAid ODI, 2009). Many water industry organizations contributed funding for projects in developing countries particularly African countries such as Zambia, Ethiopia, Tanzania and Uganda. WaterAid has set a basic strategy (2005-2010) that aims to support the MDG targets in terms of access to safe drinking water and clean sanitation facilities (WaterAid ODI, 2009). WaterAid works in the poorest communities in rural and urban areas to identify areas where households do not have access to safe water because of socioeconomic factors particularly disability, ethnicity, gender, age and other social difficulties and problems. In urban areas the problem of access to safe drinking water is often found in slums and marginalized settlements. Thus, the WaterAid organization is undertaking its projects in these communities. Water Point Mapping (WPM) is an effective method used by WaterAid to analyse drinking water access in target communities.

WPM may be broadly defined as collecting spatial demographic and infrastructure information about improved water points in a particular geographic area using GPS and questionnaire methods. This is followed by GIS analysis of the collected information to establish, analyse and display factors which impact on improved water points. This method has been applied as a successful approach in some developing countries particularly in Africa (Jimenez & Foguet, 2008). There are three significant aspects strongly related to Target 7C in the MDGs (WaterAid, ODI, 2009). Firstly, there is physical access which is associated with distance to water point and the size of the population who that is served by it. Secondly, there is the definition of safe water which varies nationally due to differing national standards. It is assumed that improved water sources provide safe water. Thirdly, there is the sustainability concept since continuity of water supply infrastructures and service management matter. Thus, providing water supply for the long term is considered a sustainable source. WPM has been applied by WaterAid in different developing countries aiming to achieve two targets:

- Exploring the distribution of water points in a geographic region to support local government policy in terms of water access.
- Focusing on the definition of water access indicators to be displayed at the sub-national level based on dataset availability.

The relationship between access to improved water sources and the service provided in reality has been investigated by Jimenez and Foguet (2008). They addressed the WPM in three geographic regions in Tanzania: Dodoma, Tabora and Singida, where datasets analysed came from a WPM campaign. Access in the Tanzania case study was defined by a maximum distance and a number of people for each water point (WP) (250 people within 400m radius). The findings of this study showed that the number of households with access to safe water was 55% lower when it is estimated using WPM, compared with official census figures based on improved supply coverage (excepting one area, Singida). These differences appear when other indicators such as quality and seasonality are included. This suggests that the JMP should be reviewed and these factors need to be included in the international access standard.

Beside the absence of a clear definition of the access term, one major criticism of Target 7C of the MDGs is that it considers water sources protected by technologies to be safe water sources while the water itself might be contaminated. For instance, water in public stand pipes could be contaminated through the treatment process in the public plants. Similarly, protected wells and pumps are considered to be improved sources, and consequently safe, while the ground water itself may contain harmful components that affect human health. Furthermore, Wright et al. (2004) pointed out that a water source could be improved and safe while it is located further from household dwellings so that the water is more likely to be contaminated through fetching and storage between source and point of use.

A national case of monitoring access to safe drinking water has been presented by O'Hara et al. (2008). In this study, a survey about access to water and sanitation was undertaken in 2005 to monitor the progress of Target 7C in Kazakhstan. A questionnaire survey was designed to cover 0.05% of the population in rural and urban areas. Additionally, 250 interviews were conducted with individuals living in urban and rural settlements and others with government and private sector organisations that were working on water supply related and health domains. The survey results showed that almost 55% of people have access to public piped water networks. However, the percentage varies greatly across the country, being higher in urban areas than rural and lower in the north than other parts. Furthermore, a high percentage of surveyed households reported incidences of interrupted water supply. This interruption was irregular, particularly in small rural clustered households. Households who do not have access to public piped networks tend to fetch drinking water from common and general standpipes located further away from their dwellings. Other drinking water sources such as wells, hand pumps and boreholes supply almost 50% of households who are unconnected to the public water supply. On the other hand, 20% of households have access to unsafe water sources such as reservoirs, streams, rivers and water delivery-tankers.

Investigating the spatial variables affecting access to safe water sources, the findings indicated that lower percentage of households fetch drinking water from sources which are located more than 1km away while the majority (70%) of people takes water from outside but within 100m. Nonetheless, great variations among urban and rural settlements have been reported. The findings also highlighted that most of the surveyed households with access to piped water reported problems in terms of water contamination and interruption and therefore access to piped water supply does not mean access to safe and improved drinking water, either in quality or quantity.

2.2.3 Water supply interruption and quality issues

Water access and quality are influenced by the supply process by which water is taken from natural sources until it is distributed and delivered to households in their dwellings. Water supply interruption has been recognized as a problem that affects improved drinking water and may demonstrate health effects, particularly diarrhoeal disease in developing countries (Blum & Feachem 1983). A recent study by Hunter et al. (2009) has used data on the water supply system of Uganda to investigate the daily infection risk due to raw water consumption when a water supply system is not able to deliver treated water. The findings of this study have shown that an interrupted water supply has a significant health impact. Moreover, the probability of water-related infections increases as the number of days of interruption increases. Geographically, the problem of water supply interruption is more likely to arise in rural areas than urban due to inadequate infrastructures. In many cases, water supply tools are found but they do not work effectively as a result of damage, old age on construction defects (Davis, 2000). Water supply interruptions have many effects on public health such as:

- Interrupted water supplies make households store the water they need and thus water is more likely to be contaminated due to container materials, transmitted bacteria and other hygiene issues.
- Water quantity is more likely to be low and thus people drink and consume less which impacts their normal life and hygiene practices.
- Changes in water pressure occur during the time of interruption and inflow of contaminated water from outside the supply system may occur, particularly in damaged and old infrastructures.
- During times of interruption, particularly long periods, households are more likely to fetch water from unimproved sources or unsafe water from sources that are located further way from their dwellings.

De Zuane (1997) has suggested that a cooperative should be established between health and drinking water authorities to monitor and tackle the water distribution system. Thus, water quality could be evaluated by assessing the water delivery system. As a result, wastewater, leakage and other supply flow problems could be determined, measured and eliminated. Fewtrell et al. (2005) reviewed the literature on drinking water quality and health impacts and have pointed out that improvements in water quality could be estimated based on water supply since water supply interventions are considered an effective method in reducing the burden of water related diseases such as diarrhoea.

2.2.4 Sustainability of water supply and spatial coverage: rural and urban

There has been an increasing amount of literature on drinking water supply problems based on the differentiation between rural and urban communities (Briscoe et al. 1990, Showers 2002; Harvey & Reed 2004). However, these studies make no attempt to highlight water access measurement or investigate the impact of spatial factors that vary between these environments. Water services and infrastructures in rural areas often neither provide a reliable and adequate water supply nor acceptable level of quality, which impact on the sustainability of drinking water. Thus high morbidity and mortality related to unsafe drinking water are found in rural areas (Gine & Perez-Foguet, 2008). Demberere et al. (n.d) have found that water quality has a great impact on sustainability since households usually reject drinking water sources such as boreholes, wells and pumps that have a changeable and unsatisfactory taste. Milman & Short (2008) have investigated water provision resilience as an indicator of the sustainability of drinking water supply in urban areas. They argue that despite access to safe drinking water being a cornerstone in public health reducing infectious water diseases and consequently promoting human development, there are many barriers that challenge the existing level of drinking water access in urban communities. These include unpredictable population growth, migration from rural to urban areas and slum growth on the margins of cities and towns.

2.2.5 Access to safe water and home treatment

Drinking water treatment is a process which aims to provide people with safe drinking water and to reduce the risks associated with microbial and infectious diseases and therefore it primarily addresses the quality rather than the quantity issue. A lack of access to safe water and poor quality lead to a higher number of deaths from water borne diseases in developing countries, in particular children under five (WHO/UNICEF 2000). Thus, access to safe drinking water and sanitation facilities have been seen as effectual ways to reduce morbidity and mortality rates (Wright & Gundry 2009). A substantial amount of literature has been published on drinking water treatment. These studies have addressed water treatment issues from different perspectives such as monitoring and assessing water treatment at plants (Jofre et al. 1995; Lasheen et al.

2008, Rodriguze-Mozaz et al. 2004), home drinking water treatment and storage facilities (Wright & Gundry 2009; Mintez et al.1995; Nicolaisen 2003; Ali et al. 2003, Lehtola et al.2002; Crump et al. 2005; Sobsey 2002; Clasen 2006) and drinking water treatment and health effects particularly water-borne diseases and contaminated underground water (Clasen et al.2006; Jousilahti et al.2005; Laurent 2005; Reller et al. 2003). Despite this extensive literature, there has to date been little consideration of integrating approaches to the measurement of access to safe drinking water internationally and nationally. Furthermore, home water treatment as a process is undertaken by households who lack access to safe water and obtain drinking water from unimproved sources. It is expected that some households treat their drinking water using filtration methods although this water is considered improved and safe such as home tap water.

2.2.6 Problems of measuring access to safe water

Although the majority of water supply problems seem to appear in rural areas of developing countries, where people lack access to public utilities and water networks, it is expected that water accessibility and sustainable safe water sources will be the greatest challenge in urban areas in the next decades (WHO/UNICEF&JMP, 2000). The interrelationships between rural and urban environments lead to socioeconomic phenomena such as population growth, migration and slum communities that affect water supplies (Gronwall, 2008). Therefore, the performance of water services is more likely to be low and the percentage of dwellings with no access to safe water is increasing dramatically, as slum and unplanned settlements are increasing.

The problem of access to safe drinking water and its effects are found in almost all communities of developing countries in Asia, Africa, and Latin America. Research on drinking water is usually related Target 7C in the MDGs and its definition of improved and unimproved water sources. Nonetheless, this definition and its criteria for distinguishing between safe and unsafe sources do not cover all spatial and demographic factors that affect access to potable water. Furthermore, no integrating quantitative technique has yet been introduced to measure access to safe water, either internationally or nationally. Consequently, difficulties arise when any attempt is made to include socioeconomic indicators in measuring accessibility to safe water sources in urban and rural areas. Standardisation of the term “access” to safe drinking water remains a great challenge.

2. 3 Measurements issue and data sources: Demographic and Health Survey

2.3.1 What is the DHS?

This section focuses on the nature and the structure of the DHS as a source of attribute and spatial data covering 84 developing countries in Africa, Asia and Latin America. The DHS topics and applications relevant to GIS and spatial analysis research are reviewed. Finally, the 2005 Egyptian DHS is considered in terms of questions of drinking water and geographic regions that were covered by the survey. The DHS programme was launched in 1984 and is funded by the United States Agency for International Development (USAID). The survey is carried out by Opinion Research Corporation (ORC Macro). Two world surveys had been carried out before DHS: The World Fertility Survey (WFS) and Contraceptive Prevalence Survey (CPS). These concentrated on population health issues in developing countries. Nevertheless, DHS has different applications and has been implemented for many aspects of health and population. The core set of questions of the DHS concentrates on household characteristics in terms of socioeconomic domains (MEASURE DHS, 2008a).

The DHS programme is considered an effective and useful data source for demographic and health issues, especially exploring and measuring the relationship between environmental influences and demographic and health variables for local communities. The survey provides researchers with geographic information that enables them to interpret and highlight the effect of spatial differences on health outcomes of populations. Despite there being some other international surveys which are water related such as the MICS, the DHS is more powerful and widely used. Moreover, the DHS provides a useful and large section on drinking water with data that could be analysed spatially, as a result of collecting latitude and longitude coordinates using GPS technology. The section on drinking water included internationally in the DHS survey had focused just on water supply types and interruptions before it was developed by adding questions about water treatments in 2005.

2.3.2 The DHS questions about drinking water

Table 2.2 illustrates nine questions related to drinking water which were introduced in the 2003 and 2005 Egyptian DHS surveys and seven of them were in both surveys:

- A question about drinking water sources for households that focuses on water sources whereby connection to public networks and piped supplies are defined as safe water sources. In contrast, other unprotected sources are unimproved and thus unsafe. For instance, bottled water, fetching water by tanker truck, and unprotected wells, pumps, springs, boreholes, surface water are all unsafe sources.

- A question about water source location addresses whether the water source is inside or outside the household dwelling, followed by another question addressing time spent to fetch water from outside household dwellings.
- A question about people who are responsible for fetching water whether they are male, female, children or adults. Two questions were added in the 2005 Egyptian DHS.
- Two questions investigate interrupted water supplies and frequency.

In 2005 two important new questions were added into the DHS. The first one considers source of water used by households for cooking and whether the source is the same as for drinking water or not. The second question focuses on household water treatment and which method is usually used by household to make drinking water safer.

	Question	**Answer**	**Survey year**
1	What is the main source of drinking water for the members of your households?	**Piped water** Piped in to dwelling Piped to yard/plot Public tap/standpipe **Tube well** **Dug well** Protected well Unprotected well **Water from spring** Protected spring Unprotected spring **Tanker truck** **Cart with small tank** **Surface water** (River/Dam/Lake/Pond/Stream/Canal/ Irrigation Channel) **Bottled water** **Other** …………………….. (Specify)	**2005/ before**

2	What is the main source of water used by your household for other purposes such as cooking and hand washing?	**Piped water** Piped in to dwelling Piped to yard/plot Public tap/standpipe **Tube well** **Dug well** Protected well Unprotected well **Water from spring** Protected spring Unprotected spring **Tanker truck** **Cart with small tank** **Surface water** (River/Dam/Lake/Pond/Stream/Canal/ Irrigation Channel) **Other** …………………….. (Specify)	**Since 2005**
3	Where is (SOURCE IN 1 OR 2) located?	**In own dwelling** **In own yard** **Else where…….**	**2005/ before**
4	How long does it take to go there, get water, and come back?	**Minutes** **Do not know**	**2005/ before**
5	Who usually goes to this source to fetch the water for your household?	**Adult woman 15+** **Adult man 15+** **Female Child** Under 15 year old **Male Child** Under 15 year old Other……….. (Specify)	**2005/ before**
6	During the last two weeks, was there any time when water was not available from (SOURCE IN 1 OR 2)?	**Yes** **No** **Do not know**	**2005/ before**

7	Did this happen on a daily or almost daily basis, only a few times per week, or less frequently?	**Daily/almost daily** **Few times per week** **Less frequently** **Do not know**	**2005/ before**
8	Do you treat your water in any way to make it safer to drink?	**Yes** **No** **Do not know**	**Since 2005**
9	What do you usually do to the water to make it safer to drink?	**Boil** **Add bleach/Chlorine** **Strain through a cloth/cotton** **Use water filter (Ceramic, Sand, composite etc)** **Solar disinfection** **Let it stand to settle** **Other……………. (specify)** **Do not know**	**Since 2005**

Table 2.2 Representing the DHS questions about drinking water in the 2005 survey and before

Source: 2003 and 2005 Egyptian DHS

2.3.3 DHS: the topics of interest

Although the drinking water issue has been addressed in the DHS as a part of household characteristics, the survey incorporates many socioeconomic, demographic, medical, epidemiologic and geographical themes. Figure 2.1 presents the intersection of DHS research topics and spatial analysis applications. Concerning the data collected about households, a spatial perspective is a major relevant area for research and analysis of environmental effects on the demographic and health status of communities.

The HIV (Human Immunodeficiency Virus) is an important indicator that can be analysed from a spatial perspective. For the purpose of understanding how the prevalence of HIV varies geographically, the relationships between spatial factors and socioeconomic circumstances are important. However, estimating HIV prevalence before 2001 was based on monitoring HIV rates in pregnant women attending antenatal care (MEASURE DHS, 2008b).) Montana et al. (2008) used GIS based analysis to identify the AIDS clusters related to antenatal care (ANC) sites. In this study, data from DHS for three African countries, Ethiopia, Kenya and Malawi had been used as well as two AIDS Indicators Surveys (AIS) from Tanzania and Uganda. The study made a comparison between DHS/AIS estimates and ANC surveillance estimates. The results proved that HIV prevalence was higher among women who lived in urban areas near ANC sites compared with men of the same areas in all countries.

Malaria is an infectious disease found widely in developing countries especially Sub-Saharan countries. In addition to approximately 1 million deaths annually, this disease causes about 500 million infections and is considered to be one of the top 5 diseases causes of death among children globally (Preman et al. 2004). Kazembe et al. (2007) argued that malaria affects child mortality particularly in under 5-year olds. They used 2000 Malawi DHS data to research biodemographic and socioeconomic conditions related to malaria. The study suggests that the spatial variations in malaria endemicity and its impacts on child mortality are related to many factors such as characteristics of places of residence, maternal age and education. Furthermore, malaria was associated with early childhood mortality in Malawi.

DHS collects valuable data to analyse the associations between family health and women’s empowerment and status. Especially in the developing countries, gender presents different social problems related to cultural backgrounds. Data on the effects of domestic violence on women and children’s health have been collected in 20 developing countries (MEASURE DHS, 2008c).

The biomarkers test focuses on health care services and health conditions related to diseases such as diabetes, vitamin A, herpes and syphilis (MEASURE DHS, 2008d). These survey data gives the opportunity to address standards of life of the local community. Furthermore, these dataset of health conditions could be linked with the socioeconomic structure of households to identify geodemographic and spatial indicators that impact on disease prevalence.

The DHS highlights many youth issues, with data collected on surveys conducted in 30 countries in Asia, Africa, Latin America and Eastern Europe since 2000. Data relating to the 15-24 age group focus on socioeconomic and cultural information particularly education, unemployment, media exposure, nutrition, general health issues such as HIV prevalence, and fertility rates (MEASURE DHS, 2008e).

Geographic information is a crucial part of the DHS datasets. It aims to improve understanding of the interrelationships between population health outcomes and spatial components. Mapping and modelling techniques and other powerful GIS tools have the potential to help decision making related to public health policies. There have been 12 studies found of DHS geographic information published (MEASURE DHS, 2011), mostly focusing on mapping health indicators and analysis of the spatial variation of infectious diseases in Sub-Saharan countries.

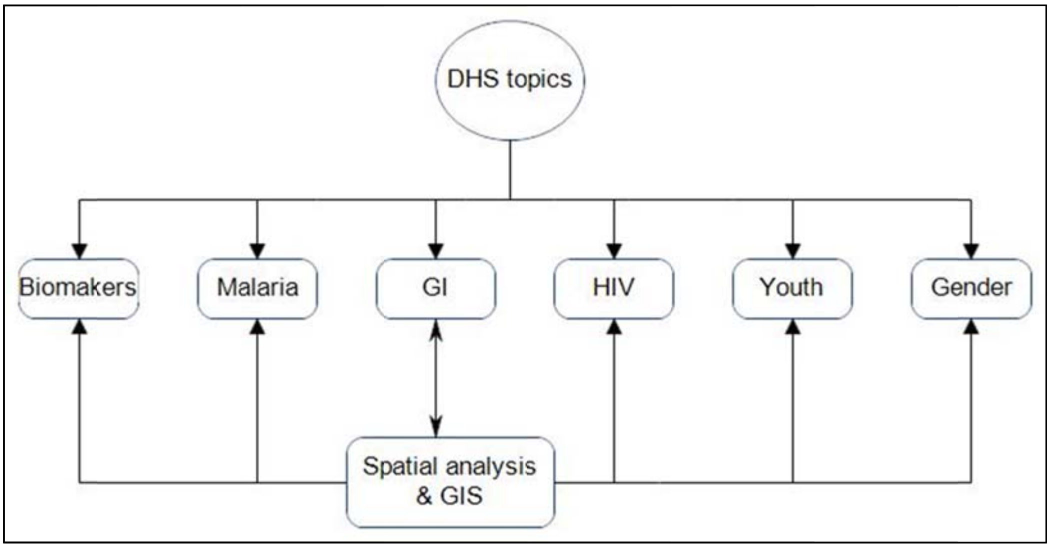


Figure 2.1 The DHS topics and spatial analysis application

2.3.4 DHS applications using GIS and spatial analysis

Analysis of the impact of places on people's health and access to health care requires GIS software for processing, manipulating, analyzing and representing spatial data. Decision making needs GIS analysis to help in tackling health problems and providing policy makers with effective interventions. DHS provides different datasets which can be analyzed with GIS tools. Therefore, many GIS projects have been developed to build health information systems and units in developing countries such as Jordan, Cambodia, West Bank and Gaza and other countries in sub-Saharan and West Africa (MEASURE DHS, 2008f).

Table (2.3) demonstrates the number of spatial analysis and GIS studies internationally that have used data from the DHS. Most of the published work has focused on family health and planning, malaria and HIV, fertility and women's empowerment. Few of these studies considered the spatial dimensions in any country and in Egypt particularly. Concerning drinking water and health outcomes of Egyptian households, although there is considerable evidence of different geographic problems and barriers of access to improved drinking water, little or nothing has been written, except the Wright and Gundry (2009) study, about the correlations between social indicators and drinking water supply, facilities, patterns and health situations.

	Family planning	Malaria, HIV and other disease	Child health	Infant mortality	Education and female empowering	Household characteristics (drinking water)	Other household characteristics	Maternal health	Wealth and socioeconomic
Mapping	3	6	3	2	1			1	
Modelling	3	1	5			1			
Geospatial techniques									
Spatial visualisation			1						
GIS and decision making									
Socioeconomic application	1								
Geostatistics									

Table 2.3 GIS and spatial analysis studies that have been done internationally so far using data from the DHS

The Global Positioning System (GPS) has been used to obtain geographic data with local interviewers collecting the longitude and latitude coordinates of household locations. GPS units are used to collect georeferenced data from the field, and these data are released as a separate DHS data file that displays community clusters at national and sub-national scales (MEASURE DHS, 2008g). The GPS data provide a great opportunity for measuring, analysing and modeling the spatial relationships between phenomena covered by DHS. Montana and Spencer (2004) have discussed how GPS data are collected and used in the DHS surveys. They emphasize the importance of the association between geographic locations, population and health outcomes where spatial applications can be successfully carried out using GPS data from DHS. Many world organizations use the GPS data collection standards which are being prepared by MEASURE DHS such as the World Bank and UNICEF. Geographic data are being used to improve and assess spatial aspects of policy planning for population and health issues in developing countries with complex health problems. In the DHS, groups of households are georeferenced to a central, averaged location, known as a GPS cluster. The mean GPS cluster size is 17 households (range: 4 to 81) within the 2005 Egyptian DHS and there are 1,359 GPS clusters within the survey. Two different methodologies are used to generate GPS cluster locations within the DHS. Using one methodology (labelled as ‘AVE’ within the DHS data), the location of each household within a cluster is surveyed individually using GPS and the coordinates of all households within the cluster are then averaged. Under the second methodology (labeled ‘GPS’), in situations where household clusters are dispersed, an approximate centre point for the cluster is assigned using a single GPS reading (Burgert, 2009).

A number of studies have utilized analysis of the spatial relationships between health inequalities and demographic variables in the DHS such as age, gender, area of residence, occupation and education level (Wirth et al, 2006). Wang (2002) studied the health outcomes of 60 low income countries using DHS data between 1990 and 1999 based on national and local levels (rural and urban). He studied the child mortality gap that was observed between rural and urban regions. Significant associations were found between health outcomes and socioeconomic determinants among poor people in these countries. Moreover, many variables can be obtained from DHS to assess and measure the relationship between deprived areas and health outcomes, such as the association between child mortality rates and access to safe drinking water, electricity, income and unemployment of households. Tanser & Sueur (2002) used GPS data from DHS to address the impact of climate on children’s nutritional status and the associations between economic diversity and spatial variations in health indicators. They confirmed the importance of the geo-coded DHS data that allows different disciplines to contribute to the study of socioeconomic and geographic patterns of demographic and health phenomena. The availability of these data presents more GIS and spatially based analysis opportunities. On health care much more has been done, particularly interpreting women’s household positions as cultural variables and impacts on maternal health data (Salway, 2006). Mathews & Bina (2004) used data from the 2001 DHS of Nepal to explore the direct impact of ecological and environmental conditions on the use of ANC behaviours by married women and the spatial variations of gender development across Nepal’s districts. They linked DHS clusters with contextual data at district level. The results suggested there were many variables affecting use of ANC services, particularly level of education of married women, distance to nearest hospital, and husband’s level of education.

Root (1997) used data from the Zimbabwean DHS to study spatial variations in child mortality in Zimbabwe. The main findings have confirmed regional variations between two provinces (Shona and Ndebele) according to population density which affected the transmission of infectious diseases. High population density and poverty were associated with high rates of child mortality, especially diarrhoea, malaria and measles. Kandala et al. (2006) studied the spatial patterns of child mortality and morbidity due to diarrhoea and fever prevalence in Malawi using 2000 DHS data. They examined the spatial distribution of the two diseases based on a geo-additive probit model and mapping approach. The results suggested that there were some differences between locations of high and low rates of diarrhoea and febrile disease. Urban areas were more likely to show lower fever percentages and higher rates of diarrhoea diseases especially in the capital. These areas might be associated with a particular pattern of socioeconomic factors such as parental education, especially maternal educational attainment, child's place of residence, household size and household economic status.

Uthman (2008) has analysed data from the 2003 Nigerian DHS examining the spatial relations and contextual impacts of initiation of sexual intercourse among women in Nigeria. However, weakness is apparent in this study where no description of data linkage has been presented. Despite the study showing extensive and clear explanations of statistical analysis, DHS sampling techniques and Nigerian hierarchical boundary structure, no information is given regarding the GIS linkage methods and processes used to complete this work. Gemperli et al. (2004) used spatial analysis methods to focus on geographic differences in infant mortality in Mali using data from DHS 1995-1996. In this study, a high rate of infant mortality has been reported in the inland Delta of the Niger River because of malaria transmission, which results in relationships between infant mortality and place of residence. The result of a geostatistical logistic model showed a strong impact of socioeconomic factors, particularly mother's age at infant's birth, gender, birth order, and mother's education and residence area.

Gemperli (2003) have studied the Malaria Risk in Africa (MARA) datasets to examine the relationship between malaria and infant mortality risk in Mali using two point-referenced datasets from the DHS. These datasets were linked using a geostatistical model. Geostatistics is defined as a statistical description of natural variables distributed in space or in time and space. Examples of these variables are thickness and depth of geological layers, soil categories in a region, pollution level in a contaminated site (Chiles & Delfiner 1999; Webster, 2007). A logistic model was applied to assess malaria prevalence among children aged less than one year to 10 years old in 101 different sites. Other spatial variables obtained from remotely sensed imagery were included in the statistical model, such as rainfall, temperature and vegetation index. Kandala et al. (2007) have studied the effects of risk factors for child mortality in Nigeria using data from 1999 and 2003 Nigerian DHS. Multivariate analysis was used to investigate the spatial patterns of childhood diseases and related socioeconomic variables. Although they mapped the spatial distribution of child morbidity especially diarrhoea, fever, and cough at Nigerian district level, no explanation is given of the linkage of DHS data and boundaries.

The spatial multilevel approach has been used for modelling interrelationships between epidemiological and health phenomena and geographical and socioeconomic factors. Kaggwa et al. (2008) has demonstrated differential impacts of individual and community normative factors of contraceptive methods among women in Mali. The study used data from the 2001 DHS of Mali and the analysis was undertaken using 403 DHS clusters. The dependent variable was "current use of modern contraceptive" and different independent variables were categorized into two groups: individual and community variables. This study, however, would have been much more persuasive if the work had considered the linkage between the DHS dataset and administrative boundaries behind the multilevel analysis. Munch et al. (2003) studied tuberculosis prevalence in the Cape Town suburbs of Ravensmead and Uitsig. They argued that according to the South African DHS there was a high rate of this disease related to social factors such as overcrowded housing, deprivation and poor lifestyle. Amin et al. (2002) used data from the 1993–94 Bangladesh DHS to analyze the spatial patterns of contraceptive use in Bangladeshi districts. They found similarities in contraceptive use, and consequently decline in fertility, across the Bangladeshi-Indian border. These similarities might be explained by cultural and social factors such as the Bengali language which is a common on both sides of the Bangladesh–West Bengal border.

A spatial modeling approach is applied widely in demographic research, particularly global issues such as population growth, climate change and water (Weeks, 2004). Figure (2.2) shows the possibilities of using DHS data in different applications and research topics. GPS data permit GIS and spatial analysis applications and modeling techniques and allow better understanding of spatial relationships between socioeconomic and geodemographic variables. GPS clusters are available as a useful part of the DHS survey which supports spatial research and provides opportunities to link it with other spatial data. For example, using a dataset from the DHS for exploring public health might require integration with other demographic variables associated with population health outcomes.

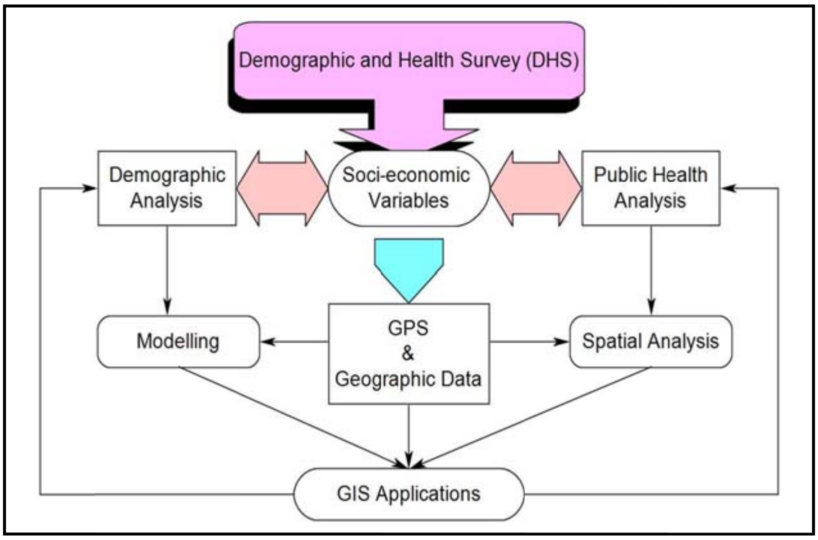


Figure 2.2 GIS and spatial analysis application of the DHS

2.4 Sampling strategy and GPS clusters in the DHS

The following sections provide a review of the strategies followed in sampling design and GPS use to collect georeferenced data in the DHS. The DHS is designed around a sampling technique that gives every eligible geographic unit the chance to be included in the survey. In other words, the sample is designed to be nationally representative. Nevertheless, in some countries some areas may be excluded based on access difficulties or scattered settlements. An explanation of the DHS sampling strategy, technique and errors has been published by Macro International Inc (MEASURE DHS, 2009a) and is summarised below:-

2.4.1 Sampling strategy and design

- The DHS are always undertaken using probability sampling. In other words, each areal unit, household, and individual has a known, non-zero probability of selection. In some cases, a self-weighting sample is used, in which each sampled unit has an equal, non-zero probability of selection. The self-weighting process is applied according to the objectives of the survey and is computed by field staff if statistical offices are not present in the country being surveyed. Spatial coverage of all national regions must be considered because the exclusion of many dispersed areas from the survey may make any interpretation or analysis invalid. The sample size is typically 5,000 to 6,000 eligible women of reproductive age (15-49).
- The sample size is mainly affected by the typical proportion of married women as there are many questions relating to family planning, fertility and contraceptive use. In some countries larger samples are required when the

tabulation of key variables is based on six or more geographical regions. In this case each geographical region should have 1,000 women in the sample. Thus, sample size depends on survey implementation, time and funding.

- An area sampling frame is needed at an early stage of planning the DHS survey. The sampling frame is often obtained from the census organization where a recent census has been conducted. The sampling frame should present a list of enumeration areas (EAs) of the survey country, with paper maps. Each geographical unit must have a unique name and a measurement size such as population size or number of households.
- Census enumeration areas in most survey countries do not vary greatly in population size. Nonetheless, in some countries they are large (1,000 to 2,000 population), and come without household or individual lists. Census maps, if they exist, may not be accurate enough. In this situation, more work is required to sub-divide large enumeration areas in the planning period of the survey and to obtain household lists for the sub-divided areal units. These smallest areal units are known as segments.
- For these smallest areal units, over time and based on experience, the DHS has developed a segment size of 500 people for the areal sampling frame. Each larger geographic unit can then be assigned a measure of size equal to the number of segments it contains by dividing its population by 500 and rounding the results to the nearest whole number.
- Stratification is applied in the DHS survey whereby the survey population is divided into subgroups. The sample is then selected independently for each stratum. This process is different from the survey domain where the country is divided into urban and rural areas for the purposes of reporting and analysis rather than sample selection. The main objective of stratification is to reduce the sampling error related to the variations that exist within the strata not between them. The most important reason for applying stratification is the variations that might exist among different populations such as urban and rural. Therefore, a stratum allows separate selection of the sampling allocation for each subgroup.
- The size of each sample in each cluster depends on the variable of interest. For instance, contraceptive use needs more clusters for comparison purposes between geographic units. Therefore, the sample size is suggested to be 15 to 20 women per cluster. By contrast, some other variables such as fertility are less clustered. Although a large cluster “take” costs less, the DHS tends to avoid this because there are many varied objectives to be achieved by the

survey. Thus, 30-40 women is the typical cluster size in rural areas and 20-25 women in urban sectors.

- Segmentation should be done in the field and each areal unit is visited, followed by selecting the sample segment for each areal unit. Mapping all segments of areal units is required, as well as listing dwellings with households residing and those absent at the time of the survey. After completing this stage, the outputs from the mapping and listing work are sent to the central office of the survey in order for household selection to be undertaken.

2.4.2 Selected sampling methods

The most important thing in the selection stage are the final elementary units, for instance eligible women, who often have an equal chance to be selected and included in the survey. Where the survey structure and sampling design means that the chance of being selected is equal for all women, this process is known as the equal probability of selection method (EPSEM). By applying this process, weighting or self-weighting will be done directly since the results are treated as a representative of the population concerned without the need for weighting in any analysis. However, to obtain the best design for the survey domains, a higher sampling probability is sometimes recommended to be applied, in particular when the population variance is larger and the unit cost is lower. The sampling error increases as the sample size decreases. Therefore, oversampling is required for every small domain to reduce sampling error. Where the probability of selection of eligible women does vary, then a technique known as Probability Proportional to Size (PPS) weighting is used to estimate sample weights that reflect the sizes of the areal units within which women and households are located.

2.4.3 Sampling errors

Two error types have been encountered in the DHS survey:

- Nonsampling error which refers to mistakes made during the survey such as incorrect allocation of households, interviewing a household which was not selected, misunderstanding survey questions and all the various mistakes in the data entry processes.

- Secondly, sampling errors such as coverage and non-response errors. Coverage error is a lack of correspondence between the sampling, as designed prior to the survey, and its implementation on the ground by interviewers. Non-response error occurs when an interview is attempted but the household does not respond to the survey because of unavailability, absence or otherwise refuses to take part in the survey. Much more detail about the DHS in terms of selection techniques and sampling can be found in MEASURE DHS (2009b).

2.4.4 GPS clusters of the DHS

Locational information is a particularly important factor in health and demographic analysis, particularly if we consider the fact that many diseases and health problems have strong geographic patterns. For these reasons, and to broaden the research utility of the data, MEASURE DHS has adopted a GPS method to collect geographic data and include them as part of the survey. The initial samples of the DHS cover different socioeconomic content. Topics of interest vary from sub-Saharan Africa to Asian and Latin America countries. In general the DHS produces useful attribute data that could be linked to cluster points with, or without, other spatial data for research. Point data can also be aggregated to the boundaries of administrative units which creates great opportunities to undertake multivariate analysis at different geographic levels (Montana & Spencer, 2004; Vaessen et al. 2004).

- **Coordinate system and cluster error**

GPS technology has made obtaining the latitude and longitude of any location much easier. Latitude and longitude are angles (degree or grid) from a point on the earth's surface to the earth's centre (Kennedy & Kopp, 2001). Latitude defines spatial position based on distance from the equatorial line with positive coordinate values to positions located north of the equator and negative ones to the south. Longitude defines positions east and west based on Prime Meridian Line (Greenwich), with positive values to the east and negative values to the west (Montana and Spencer, 2004). In his study of spatial sampling design for DHS, Kumar (2007) has argued that GPS and GIS technologies have an important impact on spatial sampling design for collecting demographic and health data where household location is required. Using GPS to identify the residential areas is considered a helpful measurement for constructing a sampling frame of households. Montana & Spencer (2004) have demonstrated how GPS is used in DHS for collecting sample clusters. Longitude and latitude are taken with attention to GPS accuracy, so that any receiver should collect data from at least four satellite signals and calculate the position of the cluster point according to the best strength of signal. However, topographic phenomena in the field such as tall buildings, mountains, hills, plateaus and tree canopies still affect signals and produce errors. In addition, there are data entry mistakes produced by field workers such as incorrect identification of site information. These mistakes are critical errors, as they cannot be corrected if they are identified after leaving field sites.

2.5 Spatial measurements: methodology of GIS data linkage

2.5.1 GIS data linkage and uncertainty

A clear interpretation of uncertainty in GIS is presented by Longley et al. (2005) who define uncertainty as the observed differences between any spatial phenomenon in the real world and the way it is being represented in a GIS context. Thus, uncertainty may affect the structure and the quality of represented datasets. They conclude that, although GIS and spatial analysis processes are nowadays implemented by powerful computational software, positional errors and uncertainty in representation of digital data are increasing. Therefore greater efforts are required to assess data quality by looking at different scales of study area or different datasets for a single phenomenon. Uncertainty also relates to concepts such as vagueness, ambiguity and fuzziness. Vagueness arises when determining the exact location of polygons, lines or points. Ambiguity can be distinguished when there is a doubt as to which class or geographic level spatial data should be assigned. Fuzziness appears in GIS fields in which classification and categories are major steps such as land use, soil studies, land cover and vegetation (Longley et al. 2005). Foody & Atkinson (2003) state that much attention has been given to the issue of uncertainty in GIS compared with remote sensing. Nevertheless, in all spatial sciences uncertainty can be investigated from the perspective of ambiguity and vagueness. They argue that ambiguity reflects the uncertainty related to a crisp term while vagueness expresses uncertainty in terms of fuzziness and roughness. The term uncertainty has been used in relation to bias, accuracy and precision.

In 1972 Gale introduced the term "fuzzy" and discussed its geographical applications. Robinson and Strahler (1984) provided in-depth interpretation of the term in a GIS context (Fisher, 2000). Robinson (2003) mentioned that the role of fuzzy sets in solving various problems in spatial analysis is increasing due to increased complexity and sophistication of GIS tools. In their study of the impact of uncertainty on decision making, Abbspour et al. (2003) make a clear distinction between uncertainty and error. Uncertainty means lack of knowledge and perfect information about spatial phenomenon which causes doubt, sureness and suspicion of the results. Meanwhile positional error reflects variations between the measurement of the observation and the phenomenon in the real world. Crosetto et al. (2000) pointed out that positional error of vector data often leads to a low degree of accuracy and high level of uncertainty. However, it is important to consider data type. Some spatial phenomena such as roads, houses and land parcel are well defined. On the other hand, soil classes and vegetation boundaries are less accurate being more reliant on the data collecter and sources. The spatial error that occurs in GIS data input leads to further error propagation problems; consequently uncertainty is associated with the findings (Heuvelink 1998).

A relationship exists between accuracy and precision concepts and positional errors in GIS. Dutton (1989) differentiated between accuracy and precision, defining accuracy as the degree of congruence and agreement with external realities, while precision refers to the level of detail recorded of the spatial phenomenon. Maceachren (1985) considered the input of database characteristics to illustrate great impact on the accuracy of spatial units. He identified four factors which may produce spatial error in thematic maps: error in the mapping process, error in collecting and recording data, error in data classification and errors in data representation. Kennedy (1989) considers the accuracy issue in demographic research to be a matter especially associated with the problem of small numbers and the outputs of percentages and ratios. Aggregation of spatial units, as a solution of this problem, is subject to the modifiable areal unit problem (MAUP) (Green & Flowerdew, 1996) and aggregation over time conceals temporal trends. A considerable and growing amount of literature has investigated point-in-polygon problems in GIS and spatial analysis research. Zalik & Kolingerous (2001) have described an algorithm mainly consisting of a polygon laid on a grid of equally sized cells. Points are tested to whether each one falls inside or outside the cell, and the local position of any point can be determined when the cell contains a polygon border. Li et al. (2007) have mentioned point-in-polygon query as an important measurement process in computational geometry applications generally and in GIS particularly. Testing a query point inclusion in a polygon is a popular problem in computer science, computational geometry, and spatial analysis and GIS (Schirra, 2008). This issue is of particular relevance when considering the aggregation of DHS GPS clusters to administrative boundaries.

2.5.2 Linkage of DHS clusters and administrative boundaries

Linkage of two datasets such as census and household surveys is one of the most significant and effective methods in GIS and spatial analysis research. In this context, spatial analysis linkage means the combination of attribute data from two different map layers into a single integrated file. Sester et al. (1998) discussed in more detail aspects of the linkage and matching together of datasets of similar or different scales using GIS and spatial analysis tools. They argued that the tremendous variety of data sets, which come from many sources, allow researchers to build effective analysis processes. A study by Chen et al. (2003) used a conflation technique to automatically integrate satellite imagery data with vector data which were in different geographic projections. Nonetheless, implementing linkage processes in general requires an accuracy assessment of the matching operation and its outputs. The DHS data structure offers opportunities for building useful comparisons for different socioeconomic and geodemographic variables. In his explanation of the role of spatial analysis in demography research, Weeks (2004) has pointed out that the DHS is an important source of georeferenced data, since it uses GPS to record the geographic coordinates for every household included in the survey. These coordinates are then averaged for each cluster of households. Moreover, researchers in the spatial sciences can use these datasets and link them to census boundaries, locating each cluster within each administrative zone. There has not been much coverage in the literature of linkage between DHS household clusters and administrative boundary data. Only one such study was identified (Pande et al. 2008) who linked DHS cluster points to other spatial datasets and investigated diarrhoea prevalence in Benin. A logit regression method was applied to assess geographical and socioeconomic variables that affect the prevalence of diarrhoeal disease. They linked DHS household clusters to various spatial datasets, including diarrhoeal prevalence estimates linked to county boundaries to identify factors that might be responsible for diarrhoea prevalence, particularly drinking water sources and water quality. One of the limitations with this study is that neither explains in detail how they linked the DHS and other data sets to administrative boundaries, nor whether there was a problem in making the spatial link. One question that needs to be asked, is whether there was actually a problem with their point-in-polygon linkage which has either not been identified or has not been completely correctly performed. These questions cannot be answered as there is no explanation of the methodology used. It seems that data attached to administrative areas have seldom been used with the DHS, and where they have been used, the spatial linkage processes have not been investigated.

2.6 Drinking water and health in Egypt

Egypt lies in the northeast part of Africa and covers an area of about one million km^2 and it is over populated country with high fertility, rapid population growth and high density of residential areas. The current population of the country is around 75,821,700 (CAPMAS, 2008). Population growth in Egypt represents challenges for sustainable development and is seen as a barrier, particularly in socioeconomic planning, because of the nature of Egypt as a dry and desert country. Population is concentrated in a narrow area around the Nile River and its Delta. The Egyptian population lives in just

7% of the total land area while the remaining areas (93%) are desert and mountain regions. It is the most populous country in the Arab world and basically a desert state. More than 90% of the population lives within the Nile Valley and Delta which make up less than 4 percent of Egypt´s total area (1,001,449 km²).The demographic situation in Egypt is causing many socioeconomic problems especially in rural areas and overcrowded urban districts in Greater Cairo and other big cities which are suffering from low income, deprivation, low standard of living and the absence of public services.

2.6.1 Water resources

The climate is generally hot and dry with insufficient rains on the northern coasts and Delta governorates during the winter season. Therefore, the climatic situation does not generate sufficient water resources. The Nile River is the main source of water from which Egypt receives 55.5 billion cubic meters (BCM) per annum based on the agreement signed between Egypt and the Sudan in 1959 (ESIS, 2008). Other sources that contribute to the total water resources are underground, rain, and drainage water. Concerning the underground water aquifer, Egypt is divided into 6 regions: the Western Desert, the Eastern Desert, the Nile valley and its Delta, the Sinai Peninsula, the northern coastal zone, and Wadi EI-Natrun (Idris & Nour, 1990). Except the Nile and Delta region aquifer, all underground water aquifers are important sources of fresh and potable drinking water as they are located far away from domestic, agricultural, and industrial pollutants. In terms of developing Egyptian water resources, the Egyptian government is carrying out a plan to finish by 2017 which aims to provide 5.9 billion cubic meters (BCM) of underground water (ESIS, 2008). Rainfall in Egypt is rare and mainly on the northwest coasts between 20mm to 150mm annually. Therefore, rains remain a limited and unreliable source of water while their valuable benefits are related to pasture cultivation in desert regions on the north coasts. The drainage water system in Egypt can be divided geographically in to sections; Upper Egypt (between Aswan and Cairo) and Lower Egypt (Delta region). Drainage water in Upper Egypt is considered a pollution source for the Nile whereby water goes from the canals to the river and increases water salinity which in turn affects water quality. The length of drainage system in the Delta is about 1,600 km and serves a wide area of agricultural land of about 1.9 million hectare (Willardson et al. 1997).

A key problem with Egyptian water sources is the pollution that has a significant impact on drinking water quality. Firstly, the River Nile is being polluted in different ways particularly by industrial and agricultural activities which lead to deep changes in physical and chemical water components. Secondly, the underground water aquifer in the Nile valley and Delta has been much polluted by domestic activities and private sanitation. The quality of the Western desert aquifer does not impact on the drinking water system because of its location outside the settled area and it is mainly being used by the private sector as a source of bottled drinking water. Thirdly, despite the reuse of drainage water in irrigation where drainage water is mixed with fresh water from the Nile, the drainage system impacts underground water negatively through raising the salinity level, especially in the northern governorates. These environmental circumstances, in addition to rapid population growth, confront Egypt with great challenges of drinking water supply.

2.6.2 Population growth and water demand

The availability of safe drinking water is a critical requirement for densely populated countries, particularly for those located in dry climate with fresh water limitations. The total population of Egypt increased from 22 million in 1950 to 75,821,700 in January 2009 (CAPMAS, 2009a). Population is likely to increase to above 120 million by 2050. This means that population size increased fivefold since 1950 while the potable water supply has been fixed. Khalifa et al. (2000) demonstrated that there have been two reasons for the continued population growth: firstly, the fertility rate is still higher in many deprived rural areas with increased health problems, child mortality and morbidity and scarce clean water per capita. Secondly, population momentum means that after the country reaches replacement-level fertility (two children per woman) a large proportion of women will still be of childbearing age. Egypt is facing a challenge in water resources where it is has been below the international standard of water scarcity since 1997 and this decline will probably continue in the coming decades (Goujon et al. 2007). A high consumption rate of potable drinking water has been continuing with rapidly population growth in Egypt. This leads to an intensive withdrawal of water resources and reduces the per capita supply of freshwater which was 880 m^3 in 2000, lower than the international level of potable water per capita by about 38% (Yang & Zehnder, 2002). It is expected that the proportion of rural population will decrease from 57% in 1997 to 52% in 2017 as a consequence of the establishment of new cities in some desert regions and the growth of some villages into towns (MWRI, 2005).The continued increase of urbanization may cause problems in terms of reallocation of water services and public networks.

It is becoming increasingly difficult to ignore the population growth problem related to drinking water supplies in Egypt. Population distribution and density in the past decade has been seen as the main challenge to sustainable development. Furthermore, residential areas are growing out of control horizontally on agricultural land because of the limitation of desert flat land in each governorate, particularly in Upper Egypt. It is assumed that the public water network should serve all Egyptian communities, villages and household clusters. Moreover, people tend to build houses and dwellings outside the geographical coverage of water services which makes it difficult to connect isolated clusters with public networks and infrastructures. As a result, households who have no access to the public network use pumps and wells as a main source of drinking water.

The problem of population growth significantly affects the water supply system in both urban and rural areas. Increasing water demand with fixed water resources and supply reduces water quality and increases the proportion of households who have no access to the public network and potable drinking water. Population growth increases water demand in all industrial, agricultural and domestic sectors. For instance, agricultural sector consumes around 60% of the Egyptian water allocation from the Nile River (Hamza & Mason, 2004). Furthermore, water services suffer from pressure of consumers and interruption particularly in districts with high density and slum areas in the Greater Cairo region. Thus, the water supply and drainage system might not work sufficiently. In addition, a problem of pollution of underground water is apparent at both old and new industrial communities in Cairo such as Shobra, Mostrod and Helwan (El Arabi, 1999).

2.6.3 Informal settlement growth

Establishing new urban communities has been targeted through formal strategies of the Egyptian government during the last two decades. The importance of new city policies comes from the demographic and housing crisis in Egypt. There have been many obvious reasons for establishing new cities within desert areas and outside the narrow populated communities. First, moving out the Nile valley and Delta where the community is suffering from many demographic, spatial and environmental problems such as high density, pollution, deprived districts and random housing. Second, there is a high proportion of population with low income looking for affordable accommodation, so creating new suitable housing units within the uninhabited desert areas is an effective solution to unwanted housing growth on valuable agricultural land. Third, new cities are being established with full utilities and infrastructure which guarantees connection to all the necessary public network services such as transport, electricity and drinking water (ESIS, 2007).

Over the past three decades there has been a dramatic increase in the number of housing units required for the increasing population in Egypt, with a huge concentration in urban areas with high density and poor environmental conditions, particularly in the Greater Cairo region. Demographic estimation of future need for housing units confirms that 7,532,919 new units will be required by 2022 across all Egyptian governorates. 1,799,446 units are required for the Greater Cairo region only (CAPMAS, 2009b). In 1996, Stewart classified the new urban communities in to four categories based on Egyptian government terminology. Firstly satellite cities which are big planned community constructed within the desert but close to Greater Cairo such as Al Abour and 6th of October Cities, each city targets between 250,000 to 500,000 people. Secondly, new towns constructed mainly based on industrial and tourist activities such as Al Amraya within Alexandria region and Sadat city located in the mid-way between Cairo and Alexandria and very close to the Delta governorates. Thirdly, new settlements located outside and far away from Greater Cairo region such as New Beni Soueif, New El Menia in the north of Upper Egypt. Finally, special natural communities which are new communities developed and constructed close to the existing urban centres such as Al Salhya and Al Nobarya where scientific agricultural activities for industrial purposes are found. The expansion of unplanned urban clusters is a critical problem in Egypt and causes a reduction in cultivated land in all rural regions and leads to slum housing in urban areas (El-Raey et al. 2000). As a consequence of unplanned housing, it appears that a high number of slum areas is found in Egypt.

The official figures indicate that there have been 909 slum areas, expanding over 326,335 km^2. The population living in these areas was 3,086,078 (4.3%) in 2005 (CAPMAS, 2009c). Dakahlya Governorate has the largest number of slum areas (121) while Giza has the greatest population living in slums (769,055) (Figure 2.3).

Constructing new cities is an effective policy for the eradication of urban slums and solving the problem of housing growth over agricultural land. However, finding suitable places for establishing housing units is a problem because Egyptian people always prefer to live close to the Nile Valley and its Delta known as the Nile settlements where there is no place for communities to expanding. Therefore, reclamation areas which are located in the marginal area between the desert and the Nile Valley and its Delta are successful places for constructing new communities.

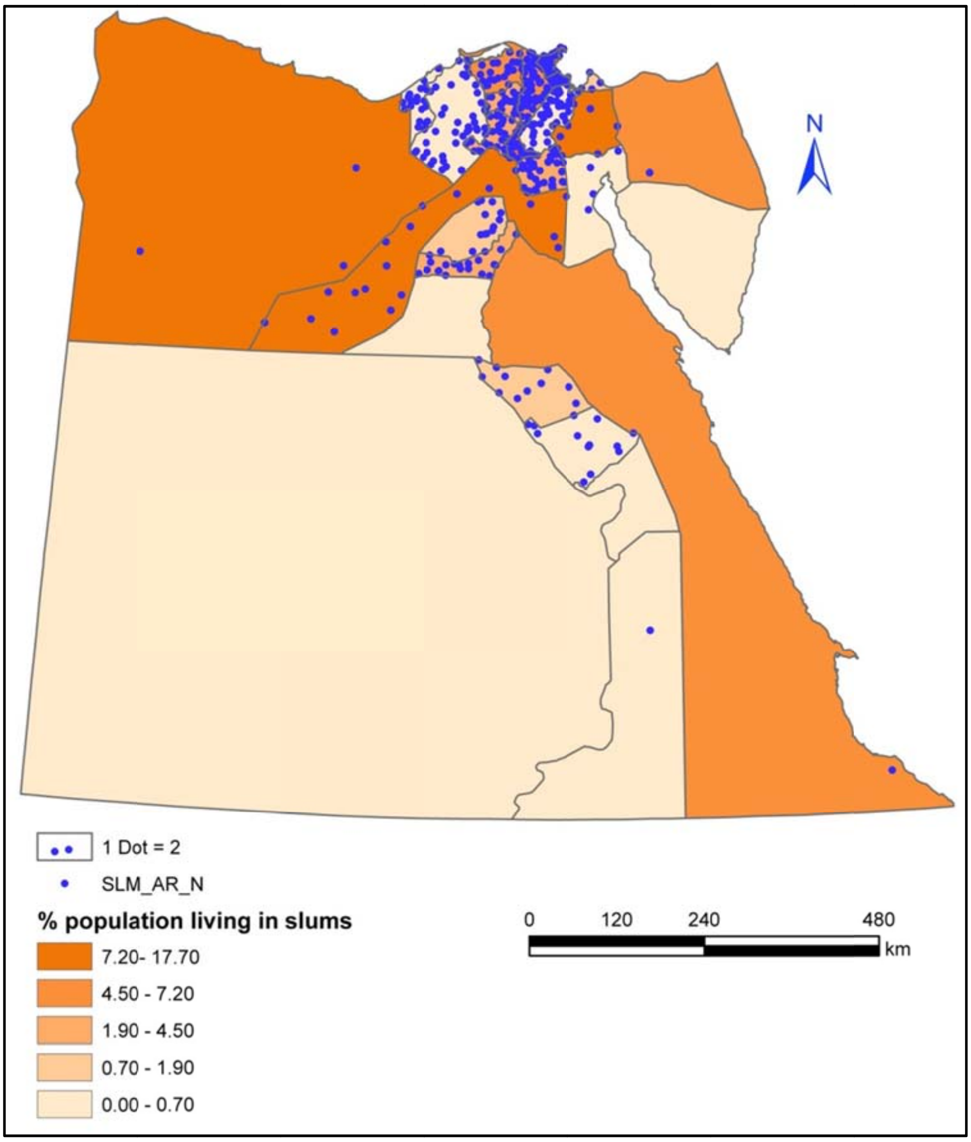


Figure 2.3 Population living in slum areas in Egyptian Governorates in 2005
Data source: CAPMAS, 2009

2.6.4 Egyptian households living in the cemetery yards in Cairo

The issue of informal and illegal housing in Cairo, as one of the largest cities in the Middle East, has been a controversial and much disputed subject within the field of urbanization. Due to the housing crisis in the capital of Egypt, the spatial structure of cemetery yard living has become a distinguishing and interesting phenomenon of ancient Cairo. In previous ages the big cemetery was outside the settlement area of the city while it is now considered a part of the housing community of Cairenes. Many socioeconomic factors are responsible for this kind of illegal housing such as rural migration, high price of accommodation, poverty and the location of this area close to the centre of the capital. Almost 2,250 households are living in the cemetery yards in Cairo (CAPMAS, 2006b). Egyptian households in these areas are suffering from poverty, particularly low income, unemployment and high crime rates. As Cairo is considered an attractive area for migrants from all Egyptian regions, rural migration increases the number of people living in the cemetery. Several attempts have been made to explore the problem of migration flows to Greater Cairo; Zohry (2002) has showed that socioeconomic conditions such as agricultural landlessness, low pay rates and low standards of living are the major contributing factors for the high proportion of migrants from rural areas, mainly Upper Egypt, to Cairo. Migrants come to Cairo to live and find work. However, they face challenges to find job opportunities or even any work in the informal economic sector. Nedoroscik (1997) has researched the anthropological and social dimensions of living in the cemetery. He found that the ancient necropolises in Cairo appear to be constructed in a particular way where every tomb includes a wide area and one or two rooms. The size and the number of rooms depend on the wealth of each family. The old tombs were built from stones while the newer are designed from bricks covered with plaster. It can be seen that the construction of these cemeteries leads to illegal housing where deprived households live in these empty rooms as temporary solution. They usually cannot find affordable accommodation in Cairo and over time they become a part of this cemetery community and live permanently.

The cemetery community is the most deprived area in the capital and a considerable proportion of the population lives without public infrastructure and service networks. As these households live in illegal dwellings no public services can be connected to their properties, except some houses that were constructed by households themselves on the marginal areas of the cemeteries and close to the modern districts. Over the past few decades, a strong relationship has been reported in the literature between lack of public services and informal housing in Egypt (Soliman, 1992) . The main problem is that the public service of drinking water is not found in the cemetery yards. As a consequence, households living in the centre of the ancient cemetery not only do not have access to the public drinking water network but also they cannot construct any private sources of drinking water such as pumps or wells. So they are more likely to fetch their drinking water from the neighboring areas, particularly from public establishments such as mosques and schools which are open all day. This water may be stored for days (Figure 2.4). Fetching drinking water and storing it without any disinfection probably exposes water to contamination and causes many infectious diseases. Wright et al. (2004) investigated the differential impact of fetching water from outside the dwelling on the quality of drinking water in developing countries. They found that many studies confirmed strong evidence regarding microbiological contamination of drinking water between source and point of use. As a future scenario, finding a means to shift all Cairene households living in the cemetery yards and accommodating them in new public settlements, probably in one of the new towns in El Salam Kism, could be an effective solution for the cemetery yard housing. Moreover, the scenario could be a part of the "Historical Cairo" housing program which aims to preserve the unique architectural and cultural of the historical places in Cairo (Figure 2.5). This program has been conducted by Cairo governorate and has accommodated 794 households who had been living the archaeological areas in the time 1997-2008 (Cairo Governorate, 2009).

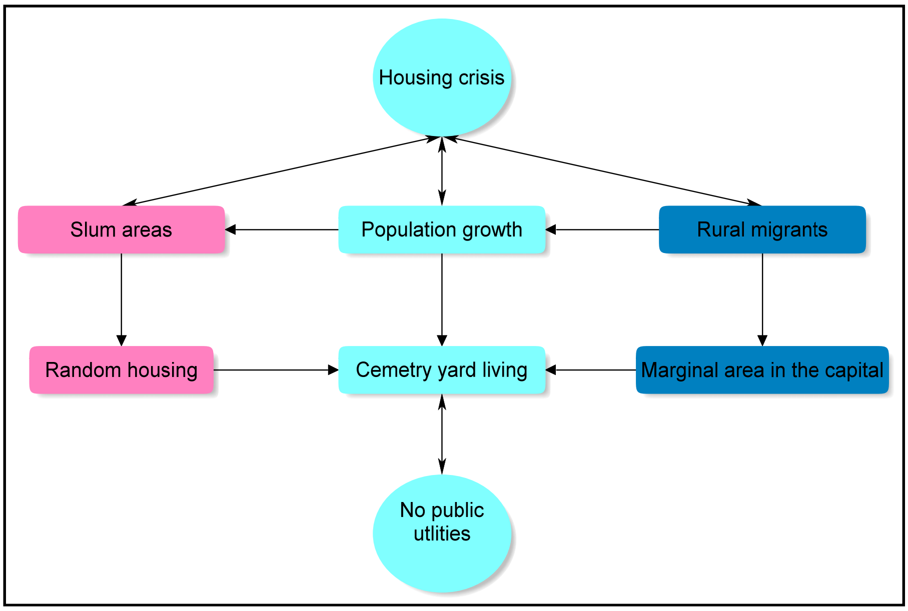


Figure 2.4 Housing crisis and the problem of cemetery yard households in Cairo

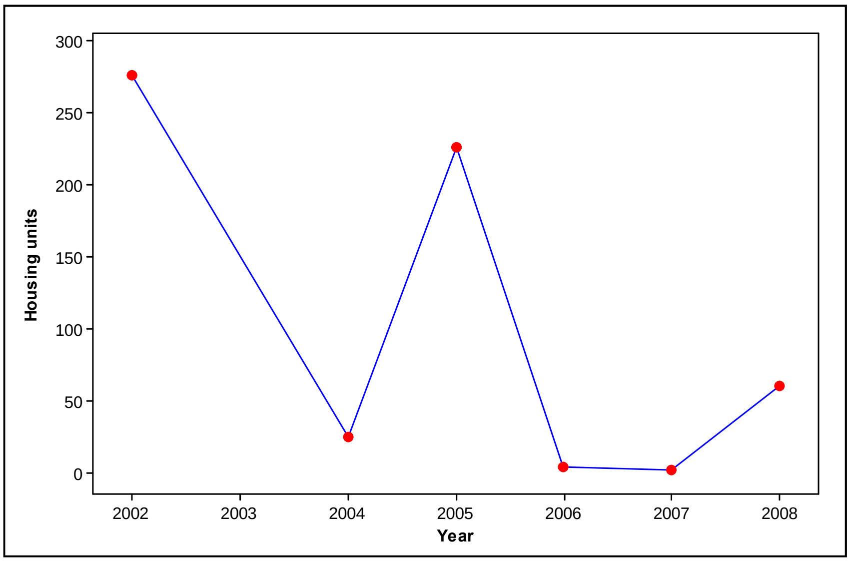


Figure 2.5 The number of housing units that have been constructed by Cairo Governorate to accommodate Cairenes in the historical places during the 2002-2008

2.6.5 New city policies

Rapid informal urbanization growth impacts environmental resources badly, imposing pressures and creating deprived communities lacking public services and infrastructure. These circumstances confront sustainable development and increase socioeconomic problems such as unemployment, pollution, illness and low quality of life (Zetter & Hassan, 2002). El Hefnawi (2005) argued that solving the problem of conflict between informal settlements and losing agricultural land requires a focus on desert urbanization and establishing new industrial, service economic communities. These communities would target all population groups who are not working on farming activities. Moreover, he discussed how the Egyptian Government would accommodate the increased population until 2020 where just 20 % of Egyptian villages have desert land as expansion areas of housing growth. Hence, the remaining (80%) of rural villages are still suffering from informal housing growth over agricultural land. New communities in those desert areas behind the villages would absorb approximately 3 million people until 2020 while there will be an additional 10 million people needing to obtain accommodation outside the Nile valley and its Delta otherwise they will construct their dwelling over arable land.

The Egyptian government had started implementing a new plan for establishing new cities in the desert in 1976 when they constructed the10th of Ramadan city, located about 48 km east of Cairo, followed by Sadat city in 1986. During the past two decades, formal housing plans have constructed 22 new cities in all Egyptian regions. These cities are classified in to three generations (Scott, 2009; Ministry of Housing, 2009) :-

- The first generation (1974-1975)

Three big cities had been constructed near to Greater Cairo, 6th of October, 10th of Ramadan and 15th of May and these are targeting people looking for accommodation in the capital and the region around it. Sadat and New Borg Elarab are located on the Western Delta aiming to accommodate population live in Alexandria and governorates of Western Delta. New Salhyia and New Damietta are located in the Eastern Delta for housing units required for Damietta and El Sharkia Governorates.

- The second generation (the mid-1980s)

This generation includes Badr, Obour and Shiekh Zaied cities within Greater Cairo region and Al Nobariya in the middle way between Alexandria and Cairo. New Bani Souif, New El Menya are located in the north of Upper Egypt.

- The third generation (1997-2000)

In this generation, many cities have been constructed in Southern Egypt alongside the Nile valley such as New Assuit, New Sohag, New Qena, New Akhmeim and New Asswan and New Fayoum city within Fayoum Governorate.
Ali (2003) argued that randomly selected locations of new cities in desert land areas cause a lack of independence of these new communities. Moreover, they suffer from inappropriate environmental characteristics such as site topography, soil, water sources and economic variation. For instance, cities are located close to flood spillways which affect new communities and expose them to environmental hazards. Therefore, establishing new urban communities in the Egyptian desert land areas requires consideration of the desert environment and its spatial relationships with the socioeconomic mechanisms of the existing communities. Although it is always thought that constructing new cities will be an effective solution for reducing the growing rate of slum areas and random housing, the location of any new community in important. The attractiveness of a new city or town increases when the distance from the existing community is small and vice versa. This spatial relationship occurs because Egyptian households prefer to live in the Nile valley and its Delta and connect with their relatives and probably their work in the original community. Despite these considerations, the new cities policy is sufficiently successful at producing the requirements of housing units located in alternative places outside the existing overcrowded areas. Furthermore, they are constructed with all services and utilities that are needed for the community especially drinking water networks.

2.7 Egyptian Demographic and Health Survey (EGYPTIAN DHS)

Egyptian policy makers have been making efforts to redistribute population, establishing new communities with complete infrastructure and solving socioeconomic problems in existing residential areas. In April 2008 a decision was made to change the Egyptian administrative boundaries where two places (Helwan and 6th October City with some other wards and districts) had been cut from the Greater Cairo region to be independent governorates. Therefore, Egypt administratively has become 28 governorates in addition to Luxor City. Figure (2.6) shows the hierarchy of the Egyptian geographies based on DHS classification and Egyptian regions and administrative boundaries.

The Egyptian DHS is a comprehensive survey covering various socioeconomic and demographic issues. The most important aspects of this survey are household characteristics, women's health, family planning, maternal healthcare, child health, HIV/AIDS and infectious diseases (El Zanaty & way, 2006). Concerning the application of Egyptian DHS in spatial analysis and GIS fields, table (2.4) illustrates the studies that have been published in academic journals. It can be seen that there are fewer applications of spatial analysis than other approaches such as family planning and women's health. Two studies made use of GIS and spatial analysis: Bachieri (2007) explored the relationships between socioeconomic modernization and desired fertility in the Egyptian community using GIS and mapping techniques. GPS data from EGYPTIAN DHS were used to assess the effect of many spatial factors, particularly land use, transport networks, land cover, population density and urbanization on modernizations and the fertility behaviour. Wright & Gundry (2009) interpreted the association between Egyptian household characteristics and home water treatments according to the responses to the new question related on drinking water treatment. They found that 5.9% of households used home water treatment in different ways. Letting water stand to settle is a common method in rural and deprived areas. On the other hand, using filters constructed in the water tap or constructed separately and filled with water is less common, and found in urban areas where rich and highly educated households usually do not drink tap water.

This review of the number of studies using data from Egyptian DHS suggests that many subjects related to household characteristics and other socioeconomic and demographic variables are yet to be researched from a spatial perspective. This research could clarify the spatial dimensions the relationships between health outcomes and geographic variations of drinking water in Egypt. The Egyptian DHS includes various housing characteristics concerning access to improved drinking water which makes possible research and analysis of the spatial patterns and geodemographic factors of drinking water and household health.

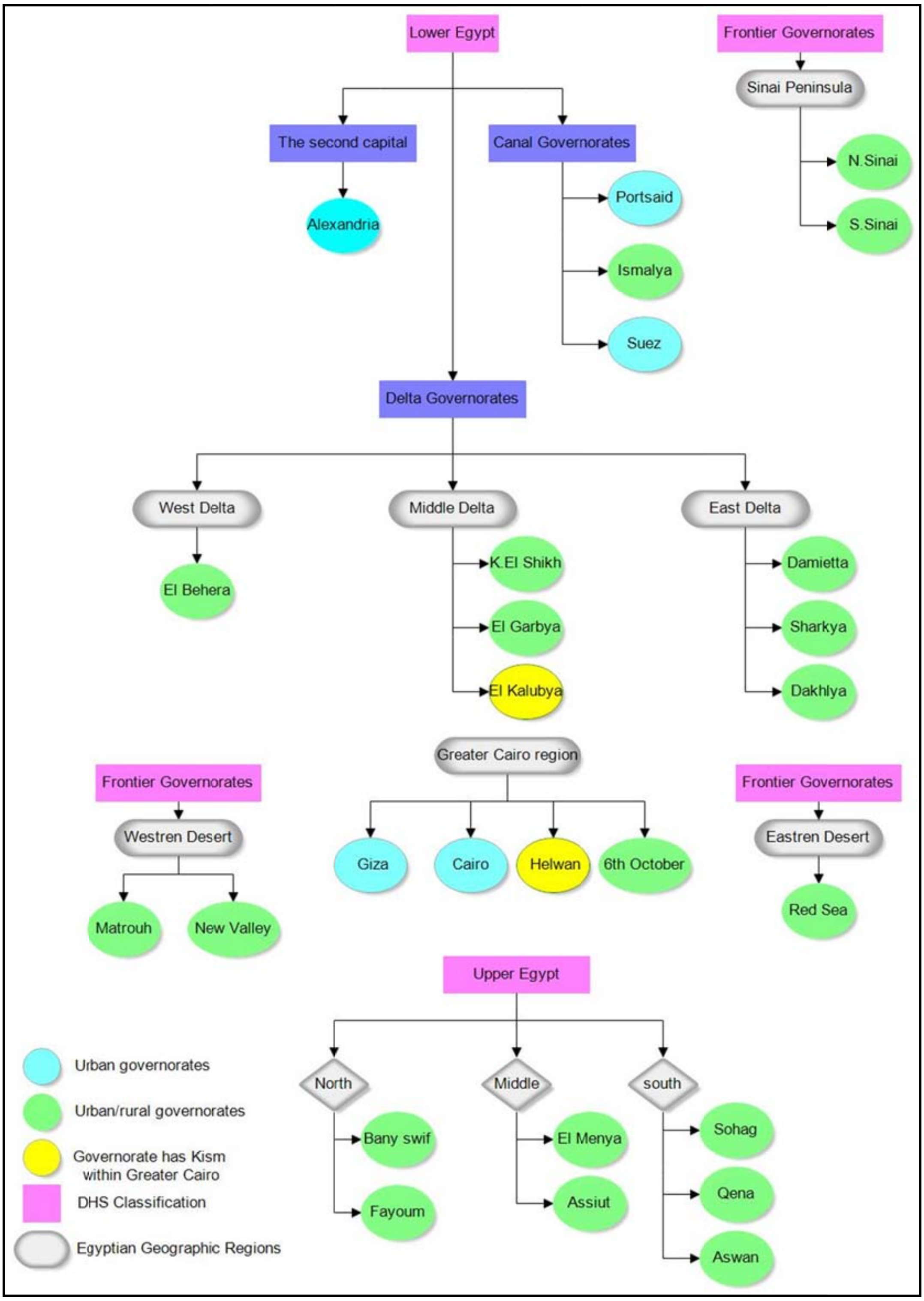


Figure 2.6 The hierarchy of Egyptian geographies based on the DHS, adminstrative governorates and regions

Study Title	GIS Applications	Spatial Analysis	Drinking Water	Youth	Gender and Women	Family Planning	Malaria and HIV
Weinreb (2008) Characteristics of Women in consanguineous marriages in Egypt, 1988-2000					✓	✓	
Tiziana et al. (2008) Community factors affecting rising caesarean section rates in developing countries: An analysis of six countries					✓	✓	
Khawaja et al. (2004) Determinants of caesarean section in Egypt: Evidence from the Demographic and Health Survey					✓	✓	
Guisti and Vignoli (2006) Determinants of contraceptive use in Egypt: A multilevel approach					✓	✓	
Marston and Cleland (2003) Do unintended pregnancies carried to term lead to adverse outcomes for mother and child? An assessment in five developing countries					✓	✓	
Baschieri (2007) Effects of modernisation on desired fertility in Egypt	✓	✓			✓	✓	
Hong et al.(2006) Family planning services quality as a determinant of use of IUD in Egypt					✓	✓	
Vignoli (2006) Fertility Change in Egypt: From Second to Third Birth					✓	✓	
Zaky (2004) Fertility transition and female rational choices in Egypt					✓	✓	
Khella et al. (2004) Lactational amennorhea as a method of family planning in Egypt					✓	✓	
Afi (2007) Lactational amenorrhoea method for family planning and women empowerment in Egypt					✓	✓	
Khawaja et al. (2004) Rising trends in cesarean section rates in Egypt						✓	
Langsten and Salem (2008) Two approaches to measuring women's work in developing countries: A comparison of survey data from Egypt						✓	
Casterline et al. (2003) Unmet need and unintended fertility: Longitudinal evidence from Upper Egypt							
Wright and Gundry (2009) Household characteristics associated with home water treatment: an analysis of the Egyptian Demographic and Health Survey		✓	✓				

Table 2.4 The characteristics of studies that published in academic journals using data from the Egyptian DHS

2.7.1 Sampling strategy in the 2005 Egyptian DHS

The sampling strategy of the 2005 Egyptian DHS aims to meet the survey goals in terms of obtaining accurate estimation of health outcomes and demographic factors of the entire Egyptian population. The spatial division of the country was mainly based on six major subdivisions, each of which is classified as both urban and rural except urban governorates where there is no rural area (see section 2.8 and figure 2.6). With the exception of the1995 and 2000 Egyptian DHS all other surveys before 2005 had not included the Frontier Governorates as a part of the geographic domain. Nonetheless, the samples that were taken from the Frontier Governorates did not include a sufficient number of household compared with other governorates. This is because of the demographic characteristic of these governorates whose population sizes are small. A clear explanation of the sample design and implementation has been provided in Appendix B of the 2005 Egyptian DHS report (El-Zanaty & Way, 2006,):

Calculating sampling error estimates for any survey is an essential procedure if we are to make any statistical comparison between population groups within the survey. Additionally, it is important for assessing the reliability and accuracy of the survey results (Turner et al. 2001). the major issue at the sampling design stage is to plan a strategy that offers an adequate number of cases with acceptable precision and low level of sampling error. However, two types of sampling error may be encountered at the stage of selecting Primary Sampling Units (PSUs). The first error comes from the variation between PSUs that are selected, while the second error appears as a result to the variation within each PSU. The greatest impact comes from the first error. Therefore, the total number of PSUs is an important factor in controlling the size of sampling error and a small number of PSUs may lead to great variation and consequently large errors. CAPMAS was the main source of lists of villages and shyakhas* that provided the sampling frames for both urban and rural areas. The list of names of spatial units reflected the geographic situation at the time of survey which had been updated after their use for the 1996 Egyptian Census. Geographic stratification was applied to the list of shyakhas, towns and villages in each governorate, listing their locations in meandering order from north to south (Figure 2.7). In the first stage, 682 PSUs were selected from three spatial units shyakhas, villages and towns. In the second stage, many steps were undertaken.

(*) Shyakha comes originally from Shiakh who used to be the governor manager of a small community in both rural and urban areas. It is the smallest unit in the Egyptian Census and it is also used as a geographic unit for grouping people who have the right to vote in an election.

Firstly, each PSU was divided into number of equally size parts(5000 people for each size of 20,000 and more) while smaller towns, village and shyakhas one part was chosen. Secondly, each part was divided into number of equal size. Thirdly, two segments were selected from each part in small towns, village and shyakhas while in the large one where there were two segment one part from each segment was selected.

A sample of 22,807 households was chosen for the 2005 Egyptian DHS. All married women aged 15-45 were eligible for the survey. One-third of all households were selected, as a subsample, for anemia-testing. The results of sampling implementation showed that 10,565 eligible women were successfully interviewed with a high response rate (99.5%) and response rate of 98.9% for all households involved in the survey which means an overall response rate of 98.5% was achieved.

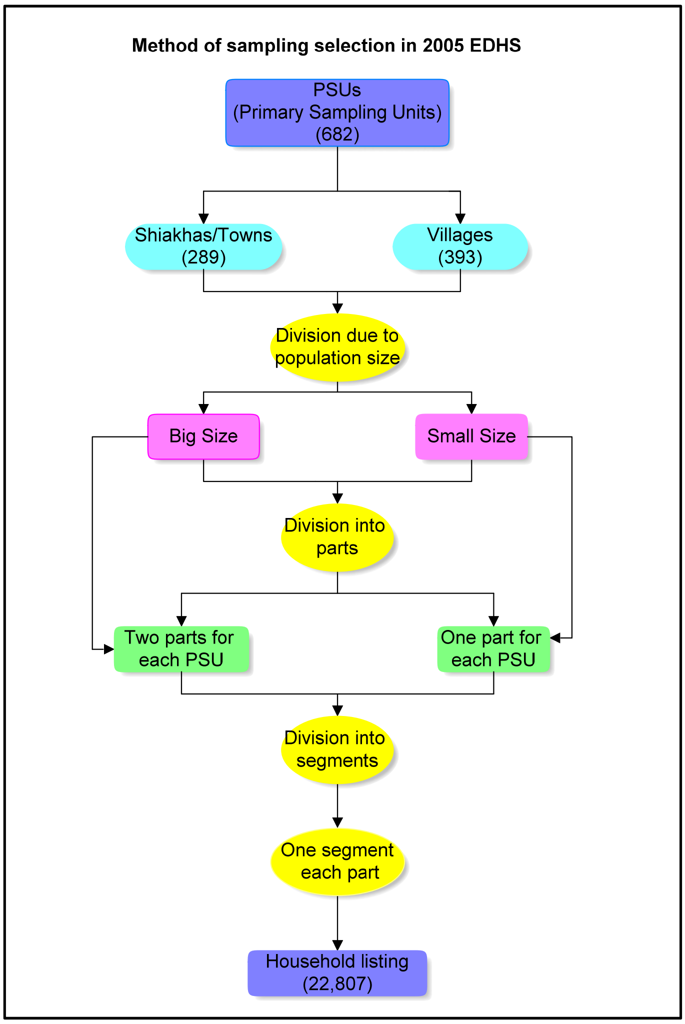


Figure2.7Method of selecting the DHS samples in the 2005 Egyptian DHS

The 1,298 clusters are distributed across 259 kisms and markazes, meaning that there is on average approximately 5 GPS clusters per kism or markaz. Thus, whilst the survey is not designed to provide district-level estimates, the risk of a highly biased estimate that might result from a single cluster falling within an unrepresentative neighbourhood within a district is reduced. For smaller geographic units such as shyakhas, which would likely contain only a single GPS cluster, the potential for bias would be much greater.

2.8 Summary

This chapter started by introducing drinking water and health issues by acknowledging international efforts to address access to safe drinking water through the MDG seventh goal this concerns the problem of household accessibility to drinking water in developing countries and defines what improved and unimproved sources. Furthermore, it focuses on how access to safe drinking water is measured internationally by highlighting programs such as “WaterAid” and sustainability of water supplies in rural and urban areas. Data sources and measurement issues for drinking water accessibility, particularly the DHS as a source of attribute and spatial data covering many developing countries worldwide, were also discussed. Specifically questions about drinking water in this survey were addressed. As many published studies have used DHS data, this chapter examined specifically GIS and spatial analyses that have used DHS as a source to examine spatial relationships. The importance of georeferenced data in the DHS was reviewed by discussing the sampling strategy and GPS clusters in the DHS survey. Linkage of DHS GPS clusters and census administrative boundaries was examined and it was clear that there has not been much coverage in the literature about this process, particularly regarding spatial uncertainty. Lastly, access to drinking water in Egypt and related issues were highlighted, especially water resources, population growth and water demand, slums and informal settlement growth, and Egyptian new city policies. The 2005 Egyptian DHS was introduced as one of the main data sources of this particular research. Sampling design, selection and implementation of the Egyptian DHS were described and case studies using Egyptian DHS in general were identified and reviewed.

Chapter 3: Research Methodology [an evaluation of geospatial data linkage for measuring safe water access in Egypt]

3.1 Overview

The aim of this chapter is to investigate the issues involved in linking the Egyptian DHS with census data using spatial overlay. It begins with a brief overview of census and survey geography, highlighting spatial characteristics of Egyptian administrative boundaries. Moreover, it discusses the field work particularly its objectives to obtains deep understanding about topology of drinking water supplies in different Egyptian regions and the local factors may influences drinking water accessibility. In section two, GIS analysis tools were used to measure and analyse the spatial error of Egyptian DHS GPS clusters, detected by examining the province codes for each GPS cluster, as coded in the Egyptian DHS and according to cluster location. Finally, in section three, the spatial error was modelled using logistic regression and spatial statistical techniques to obtain the probability of each GPS cluster falling within its incorrect sub-national boundary unit, kism/markaz. Thus, the chapter examines whether there is a significant pattern to this spatial error before undertaking any linkage of the DHS clusters to sub-national boundaries. The method described here and subsequent results form the basis of a paper by the author accepted for publication in Computers, Environment and Urban Systems, which is reproduced as appendix.1 to this thesis.

3.2 Administrative Geographies

3.2.1 Census geography

Egypt is divided into four main regions based on demographic and spatial characteristics: Upper or Valley Governorates, Lower or Delta Governorates, Urban Governorates, and Frontier or Desert Governorates. Each region has distinctive spatial features. For instance, Lower Governorates are characterised by high population density and plain lands where there is no major relief such as plateaus, mountains, or hilly areas. Frontier Governorates are located on desert land with very low densities of population and settlement. Most of those governorates' areas are inappropriate land for construction of new settlements and communities because of their geographical features that are completely different from the old communities that are located around the Nile River and its delta. Therefore, those areas do not attract population, not even the new formal cities and towns that have been established by government near to the

Valley and Delta. This situation affects population distribution and leads to rapid growth of slums and informal housing that lack public services and utilities, within the existing communities. The shapes of census and administrative boundaries are influenced by environmental phenomena, spatial and demographic characteristics. Therefore, changes to the boundaries of governorates are mostly due to increased urbanization, population growth and the establishment of new communities over reclamation land in the marginal areas of the Valley and Delta regions. However, Frontier Governorates are always large as they contain huge desert areas with almost no population. A map of the sub provincial level, kism and markaz, especially in Upper Egypt, consists of a narrow populated band of agricultural land while plateau areas east and west of the Nile have been removed as there are no communities present.

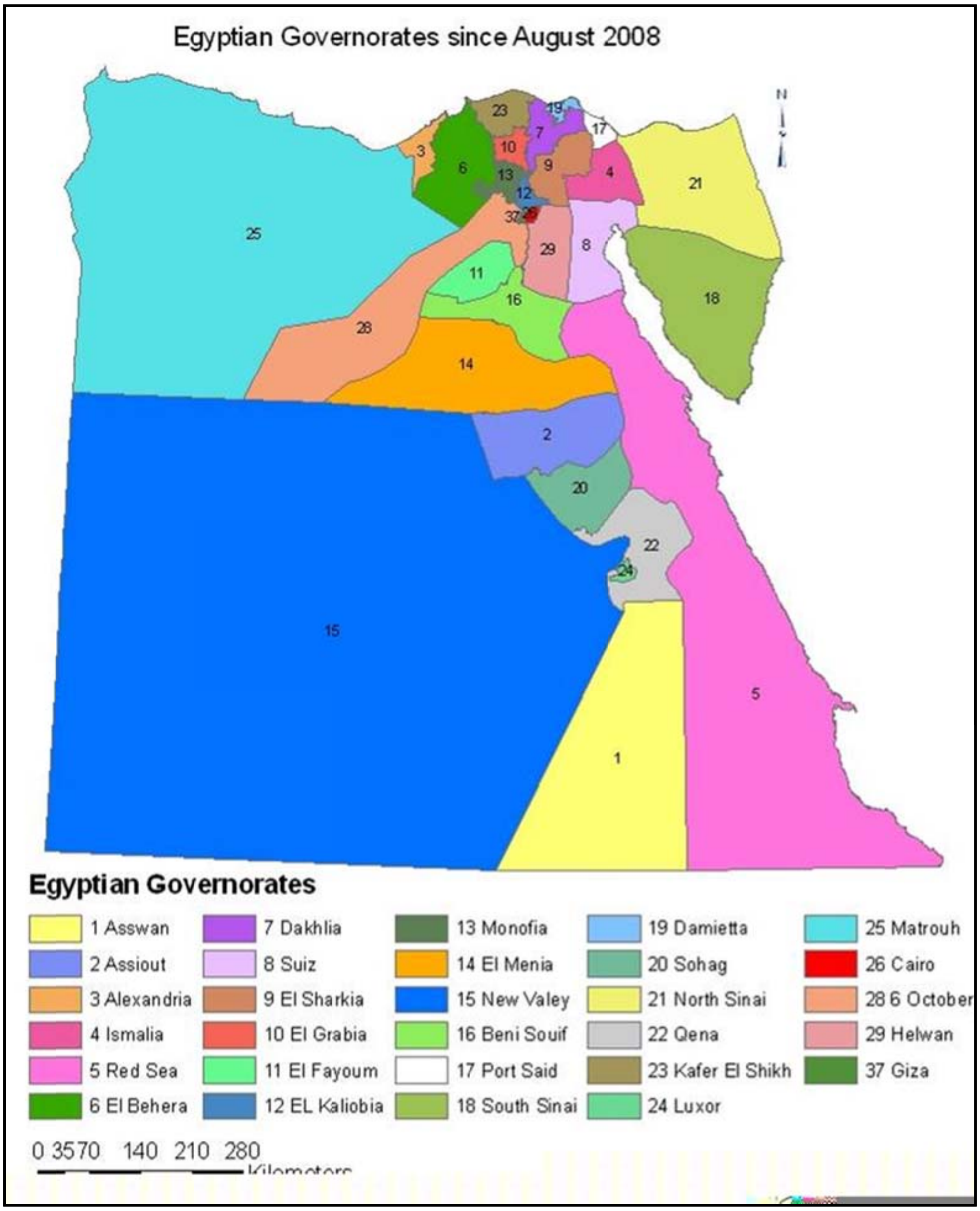


Figure 3.1 The new Egyptian administrative boundaries since August 2008

CAPMAS is the official organization responsible for censuses, surveys, public statistics and formal data and information. This formal source provides researchers, universities, organizations, and institutions with statistical data that can help in terms of studying and researching Egyptian society. Furthermore, it is the main source of digital administrative boundaries. Before August 2008, the number of Egyptian Governorates was 26 in addition to Luxor Supreme Council (Figures 3.1 and 3.2). However, changes took place on the 1st August 2009 at governorate level by establishing new two governorates; Helwan and 6th of October (Ministry of Local Development, 2009). This change was made to modify administrative boundaries of the Greater Cairo region by separating some new cities from Cairo such as New Cairo First, Second, and Third, Badr, 15th May, and El Shorok and adding them to Helwan and separating rural areas from Giza such as El Ayat, El Badrasheen, El Hawamdia, El Wahat, and Imbaba Markaz and adding them to the new 6th of October Governorate (Figures 3.4 and 3.5; 3.6 and 3.7 and 3.8). The main target of creating those new governorates was to reallocate public services and population distribution and reduce the pressures concentrated on the capital region. The total number of Level 2 (district) census units is 360 kisms/markas. Five Urban Governorates, Cairo, Alexandria, Giza, Port Said, and Suez, are comprised of 86 Kisms (an urban division) while all other governorates and Luxor Supreme Council contain markaz (both an urban and rural division) within their census boundaries (Figure 3.3).

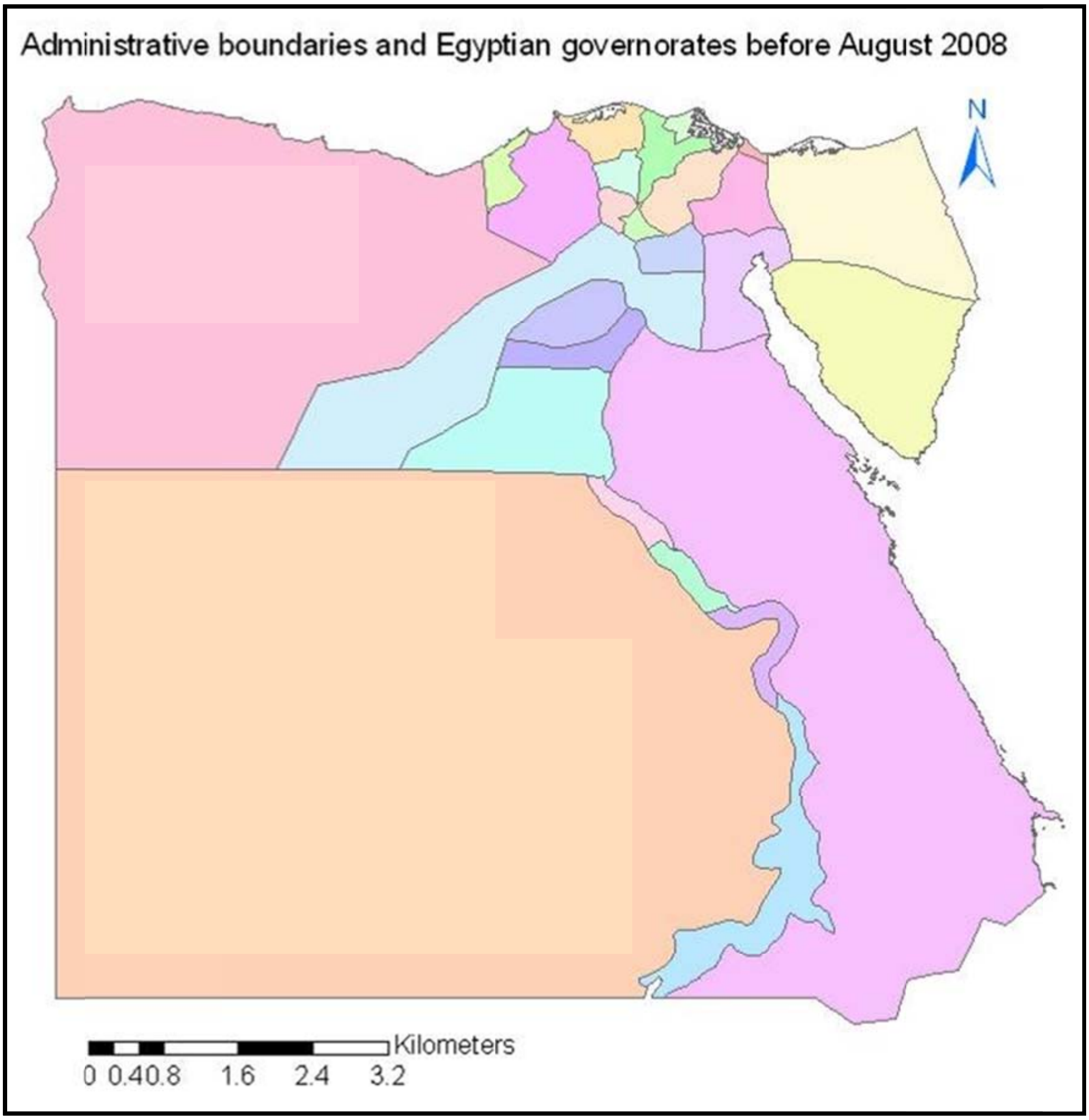


Figure 3.2 The previous Egyptian administrative boundaries before August 2008

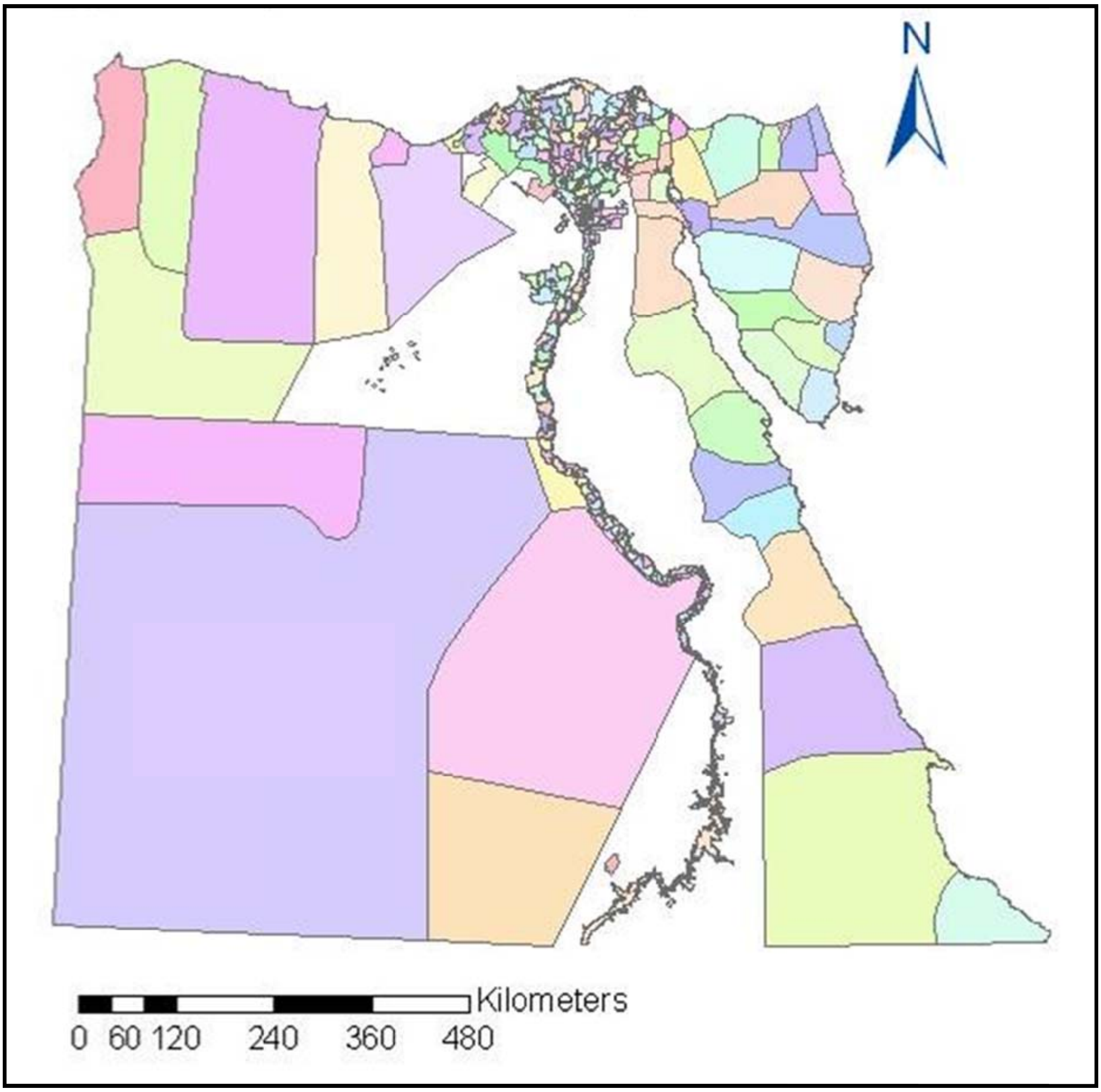


Figure 3.3 Egyptian sub-governorate boundaries showing kism/markaz boundaries

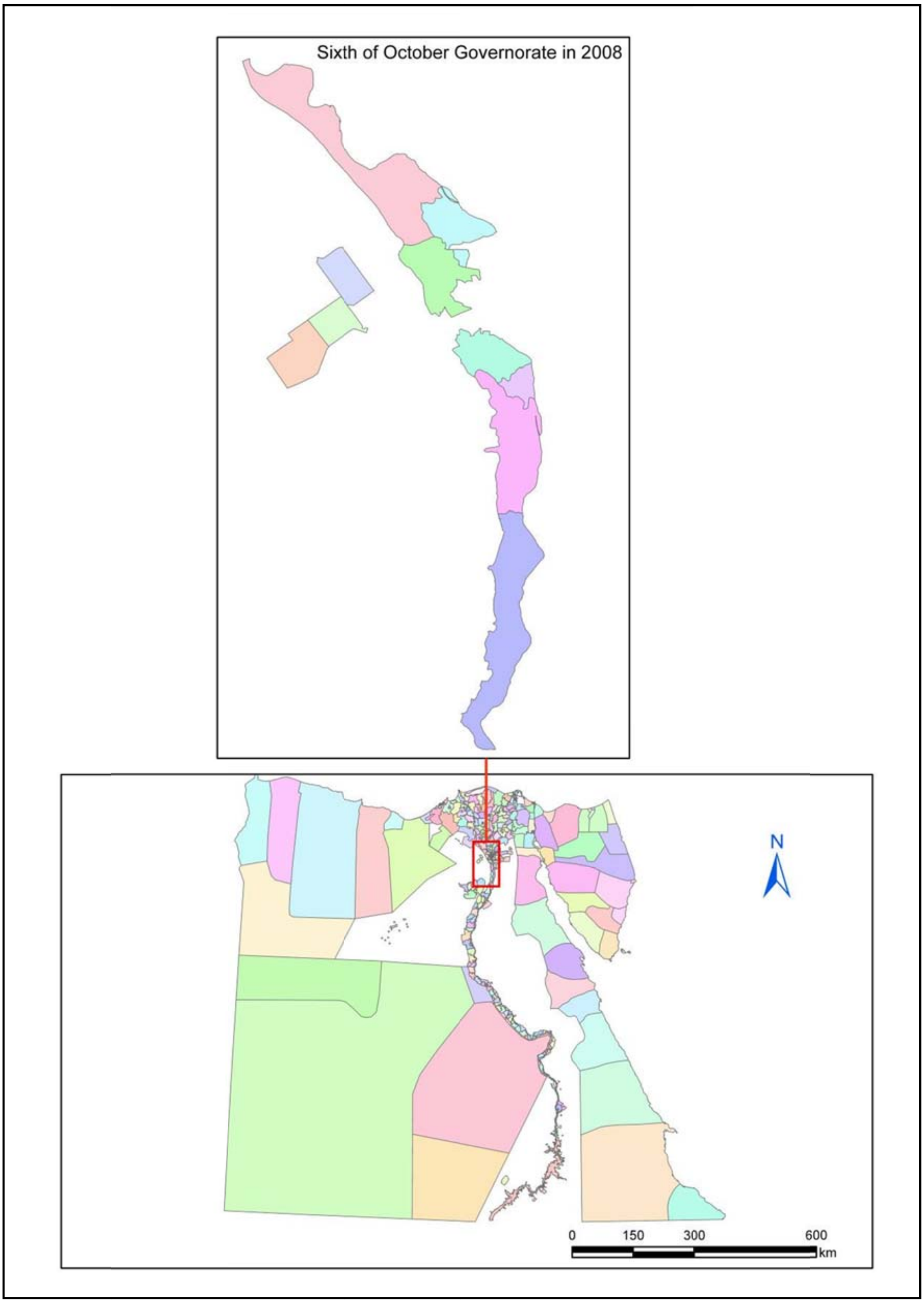


Figure 3.4 6th of October Governorate, showing kism/markaz boundaries

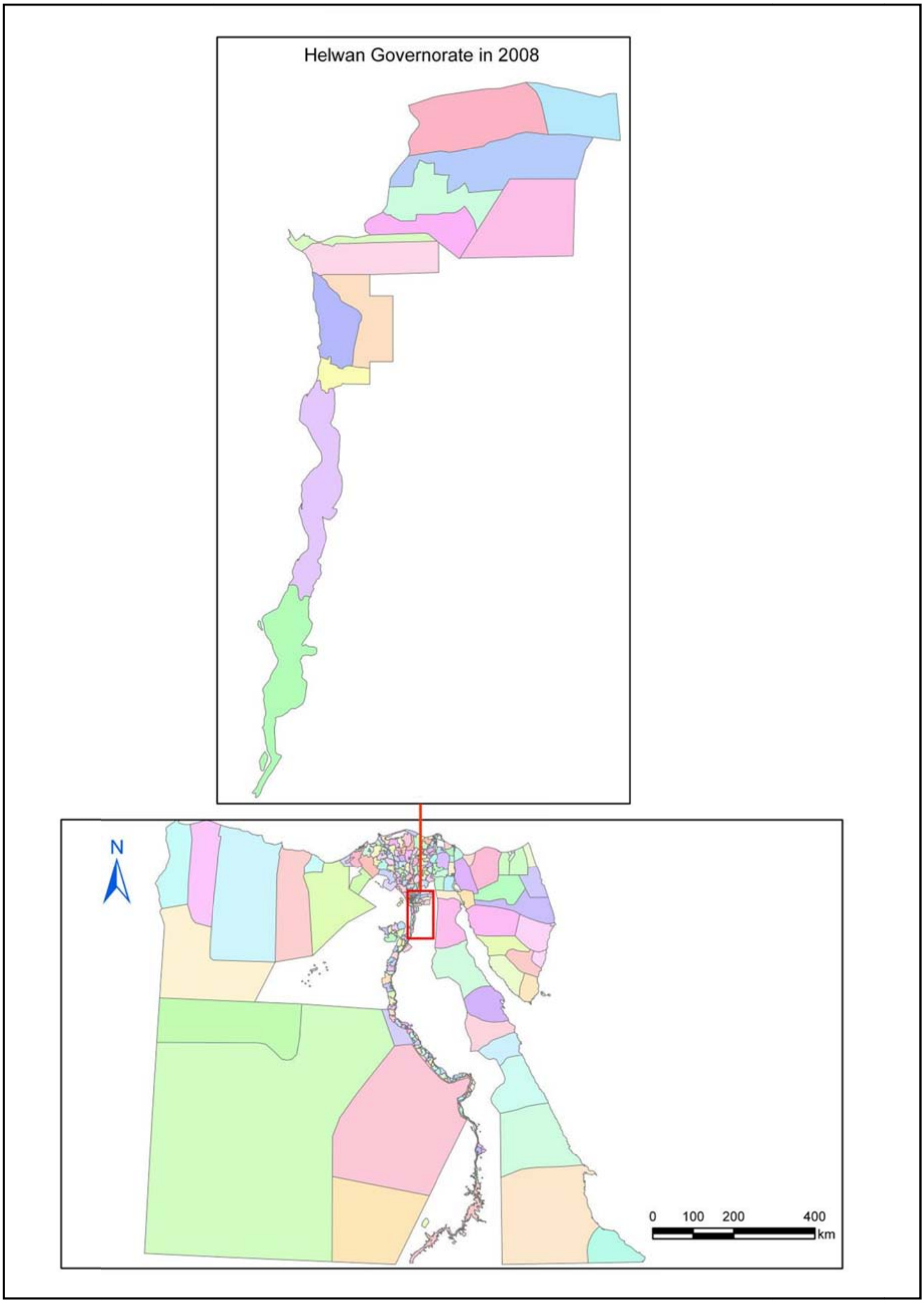


Figure 3.5 Helwan Governorate, showing kism/markaz boundaries

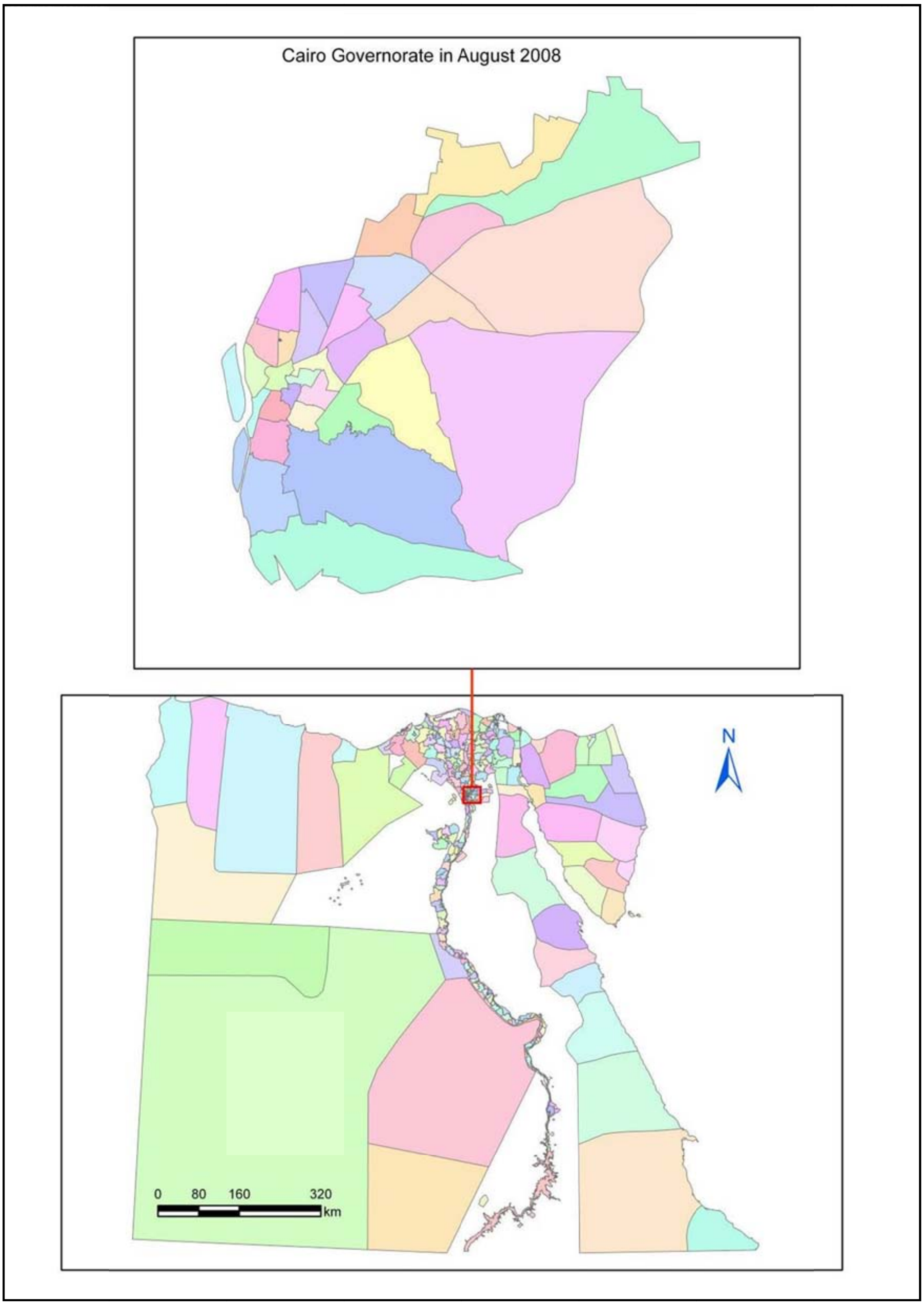


Figure 3.6 The new boundaries of Cairo Governorate in August 2008, showing kism/markaz boundaries

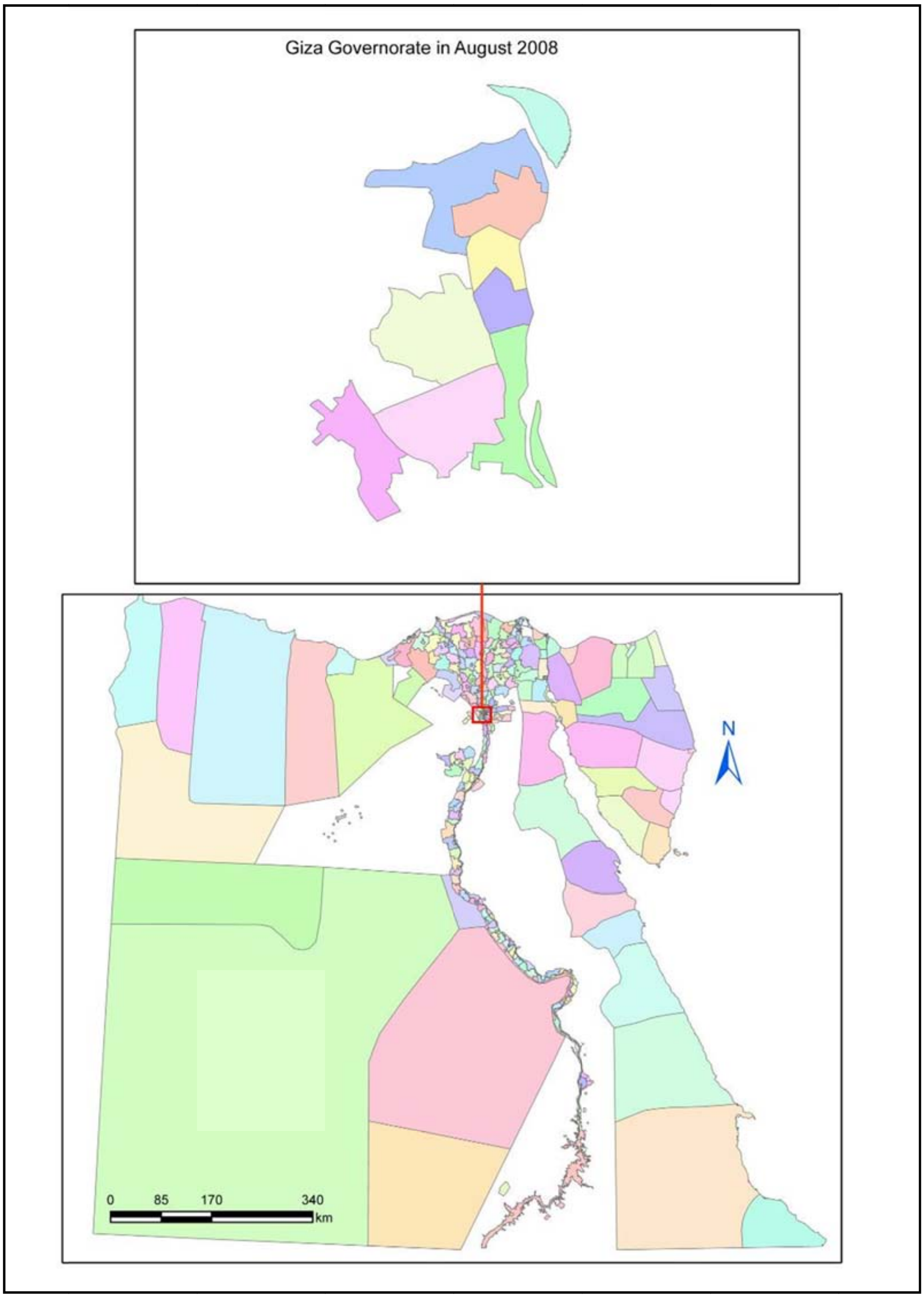


Figure 3.7 The new boundaries of Giza Governorate in August 2008, showing kism/markaz boundaries

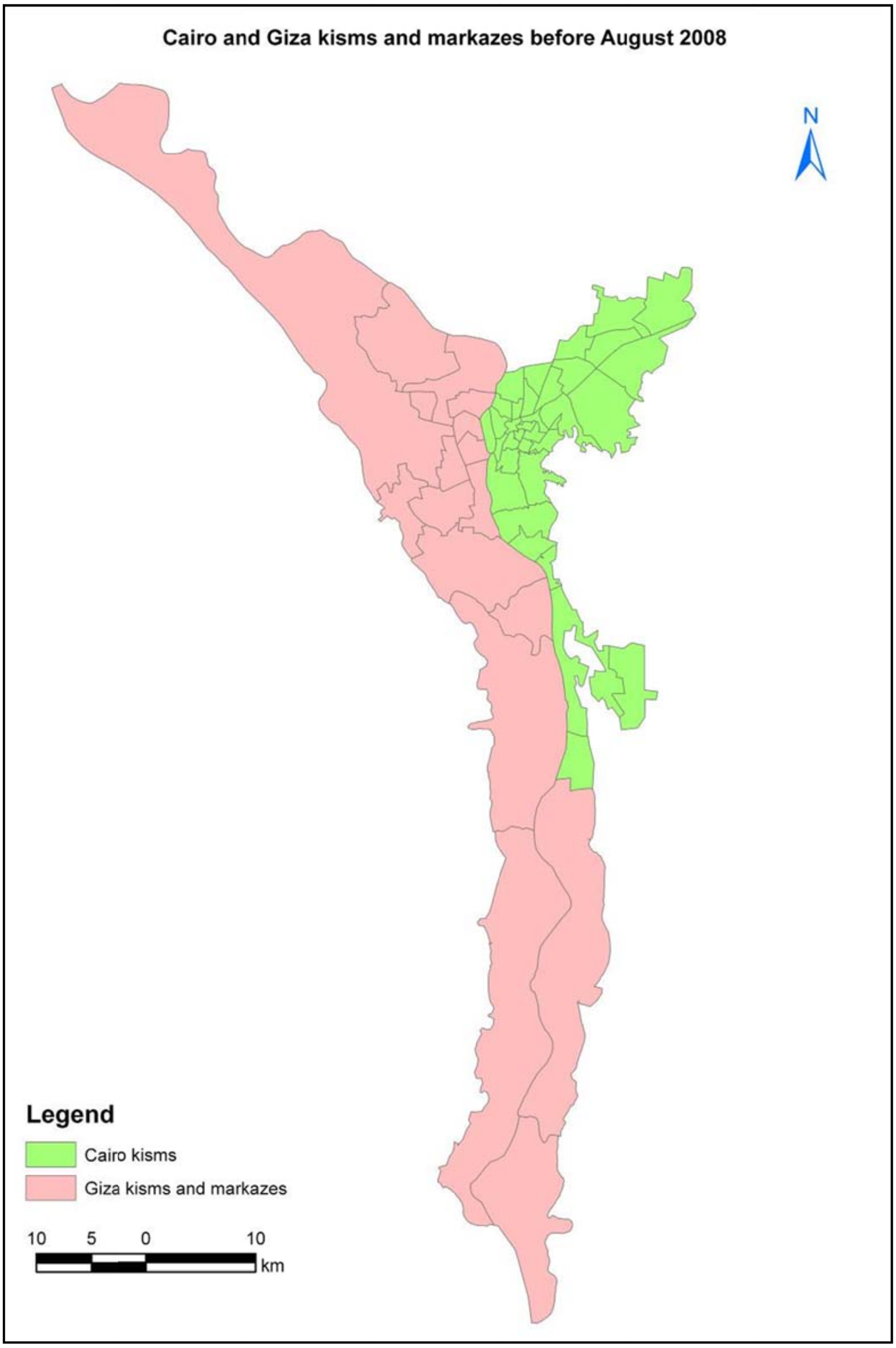


Figure 3.8 The old boundaries of Cairo Governorate

3.2.2 Spatial characteristics of census boundaries

Looking at the map representing the total number of population in each governorate in 2006 (Figure 3.9), it can be seen that the direction of settlement growth has followed the Nile River and its branches. The concentration of population in the Nile River corridor and the Delta can be described as the "flood plain settlements" and this term will be used in the remainder of the discussion. The Egyptian boundaries have been affected by those dimensions of population growth, and each governorate's shape reflects its spatial location (Figure 3.1). Moreover, the total land area in the west and east banks of the river in Valley Governorates are very narrow by comparison with the locations of east and west plateaus in particular Aswan, Qena, Sohag, and Assiout. Therefore, Valley Governorate rural and urban settlements run parallel to the Nile River direction from the south to the north.

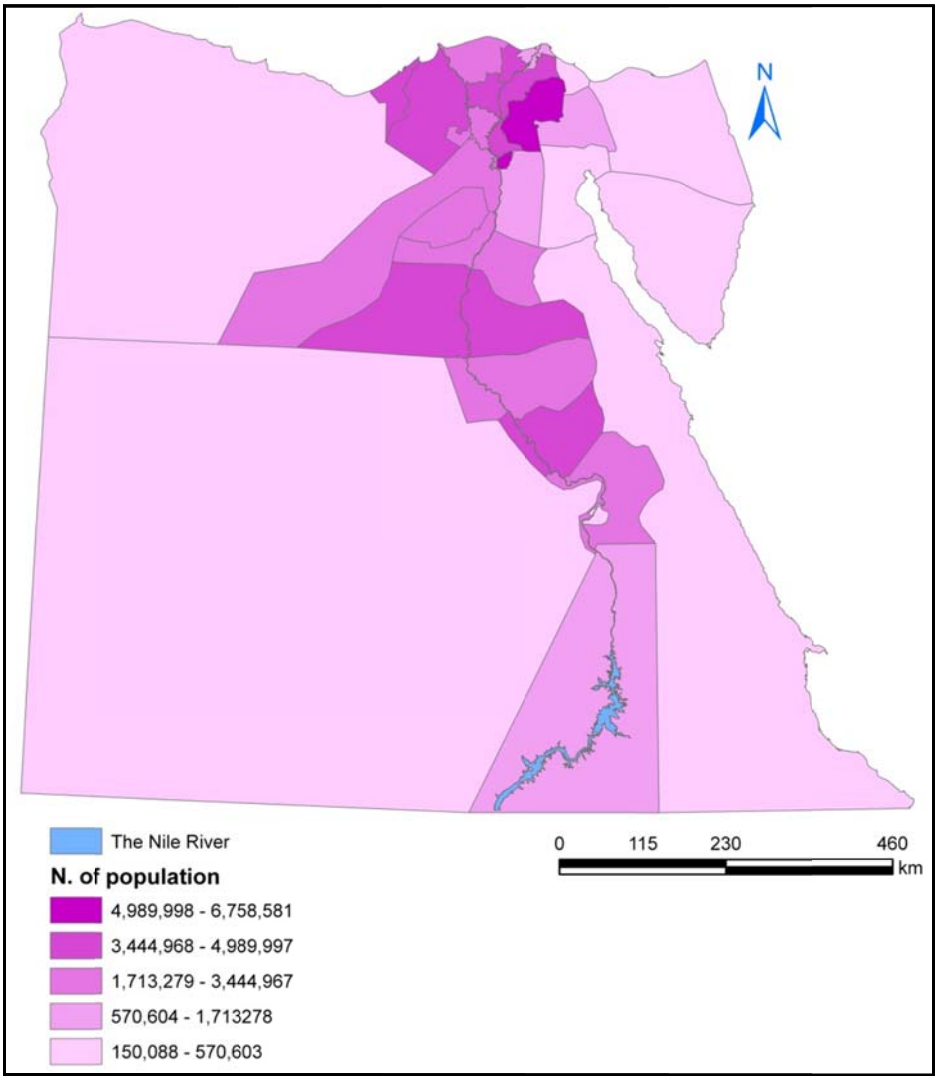


Figure 3. 9 Distribution of total number of populations across Egyptian Governorates in 2006

The Egyptian Government is targeting modification of administrative boundaries by adding some reclaimed desert land to each governorate to be alternative areas for new communities and stopping loss of the productive agricultural land to urbanisation in the Nile Valley and Delta governorates. However, Egyptian settlements and population are still growing and expanding, following the existing pattern of the flood plain settlements which are concentrated between and around the Nile Branches and canals. Southern Egypt represents the lowest part of the communities of the Nile flood plain which is a line from south to north following the Nile River direction, with population concentrated around it. However, the delta settlements spread out between and around the Damietta branch of the Nile in the east and the Rosetta branch in the west. Therefore, the Delta Governorates' shapes reflect the wide spaces around and between the two Nile branches compared with the linear Valley Governorates. This spatial characteristic of administrative boundaries reflects the Upper and Lower Governorates' respective locations in the Nile valley and Delta.

The Urban Governorates are Cairo, the capital of Egypt, Alexandria, Port Said and Suez. With the new modification of Egyptian boundaries, Giza Governorate has become an urban governorate because all rural areas that were parts of it had been added to the new governorate 6th of October. As a result, five Egyptian Governorates are urban and they do not contain any markaz census units. The two urban Nile flood plain settlements, Cairo and Giza are located at the base of the Nile River Delta. While Cairo is located on the east bank of the Nile River, Giza is located on the west bank. Both governorates are the major destinations of Egyptian migrants from the north and south of Egypt. Alexandria, the second largest city, is located along the coast of the Mediterranean Sea in the west of the Rosetta branch of the Nile River Delta, and is considered the second capital of Egypt. Furthermore, it is considered the demographic and urban centre of the Western Delta region and the biggest port of the country. On the east of upper Damietta branch of the delta, Port Said is located along the coast of the Mediterranean Sea at the northern gateway of the Suez Canal. It is the smallest governorate with an area of about 72 km². On the other hand, Suez Governorate is the largest urban governorate (17,840 km²) and it is located on the Northern Suez Gulf. It is situated in the Eastern Delta and considered as the southern gateway of the Suez Canal (CAPMAS, 2009d).

Frontier Governorates are located in desert regions. New Valley is located in the south west of Egypt and it is the largest governorate in Egypt and indeed in the whole of Africa (376,505 km²) with the lowest population (187,263 million in 2006) compared with other governorates. It consists of four Markaz, each of which is an oasis. Despite the New Valley project started in 1997 by the Egyptian government to redevelop Southern Egypt and partly address the rapidly growing population, it has the lowest population density. Matrouh Governorate is located in the north west of Egypt along 450 km of the Mediterranean Sea coast and along the Libyan border and south 400 km to the New Valley Governorate border, with a total area of 166,563 km² and a population of 334,246 people in 2006. It is the second largest Governorate in terms of area (CAPMAS, 2009e). Red Sea Governorate lies in the Egyptian Eastern Desert and has the longest borders. It is located in the south east of Egypt between the Nile River on the west and Red Sea on the east. The total land area is 203,685 km² and it has a small population of 288,233. Hurghada city is the capital and it is a famous place of recreational, underwater and aquatic sports tourist activities. The Sinai Peninsula is divided into two frontier governorates. North Sinai covers an area of 28,992 km² and its population is 339,800. The governorate is described by two environments; coastal and marine areas along the Mediterranean Sea and El Bardaweel Lake in the north and desert, mountain, and valleys in the south. South Sinai Governorate is located between the Suez Gulf in the west and Aqaba Gulf in the east and covers an area of 31,272 km² with a population of 149,300 in 2006. It contains the most famous places in Egypt in terms of international tourism in particular Sham El Sheikh, Dahab, and Nuweiba. In addition it produces 30% of Egyptian petroleum and it does not have any slum areas (CAPMAS, 2009f).

3.2.3 DHS Geography

The geographical referencing of the EGYPTIAN DHS takes two forms, the inclusion of a governorate identifier and a DHS GPS point. The GPS points represent clusters of DHS household survey locations and are therefore an important basis for geographical analysis of the DHS data. They represent a far higher spatial resolution than the governorate level. The 2005 EGYPTIAN DHS contains 1,298 GPS cluster points that are distributed across 26 Egyptian Governorates and Luxor City Council. According to the sampling design of Egyptian DHS (Figure3.10), the survey strata are based on an urban/rural classification of the four major Egyptian regions; the Urban Governorates, urban Lower Egypt, rural Lower Egypt, urban Upper Egypt, rural Upper Egypt, and the Frontier Governorates (El-Zanaty & Way, 2006)

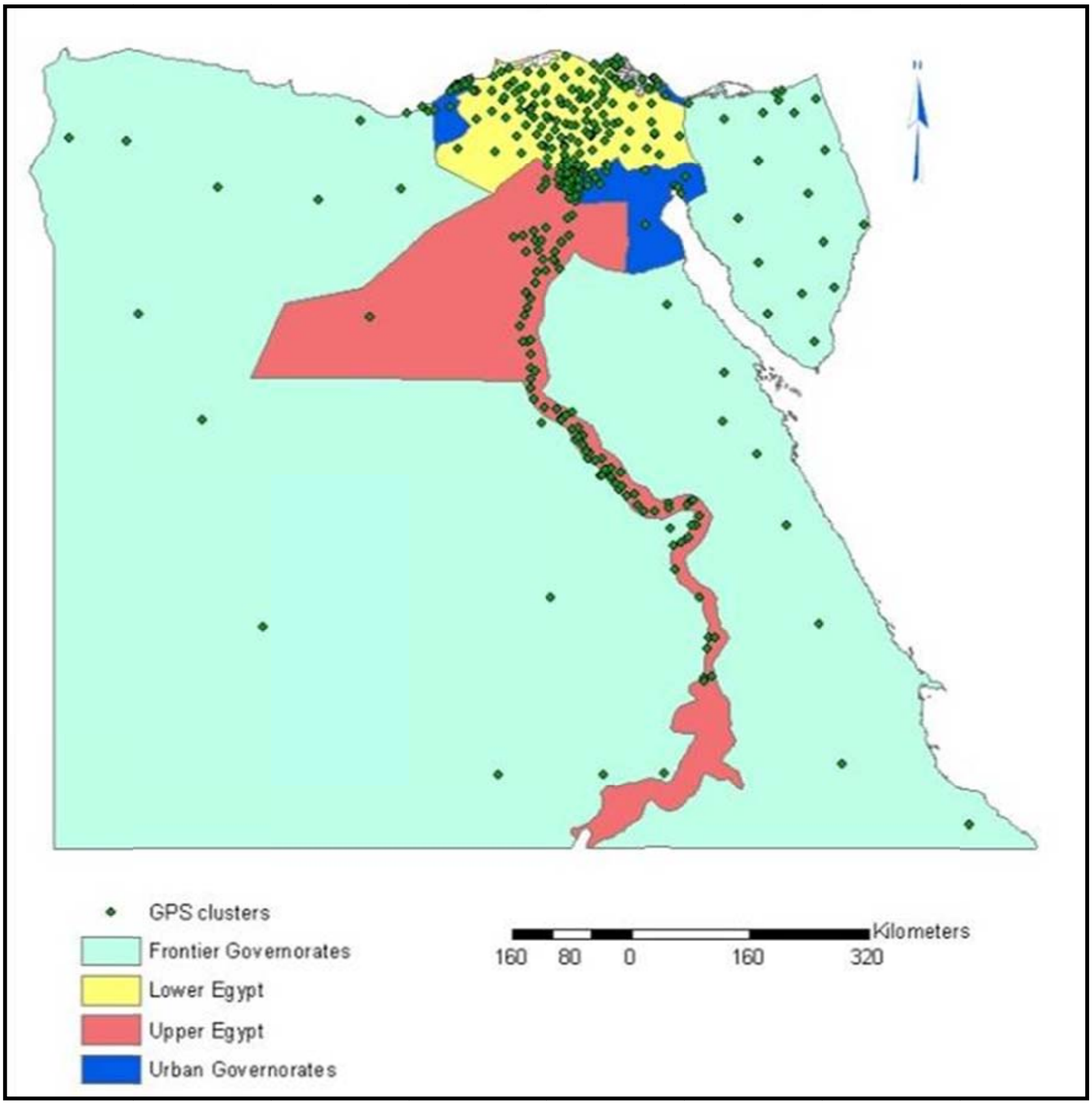


Figure 3.10 Distribution of DHS GPS clusters by main Egyptian regions

3.2.4 Primary Data Collection and Field Observations

Introduction

The major aim of this field study is to investigate and explore the potential spatial variations of drinking water supply patterns across Egypt to determine the challenges that affect some Egyptian communities. Moreover, it also sought to:

- Gain better understanding of spatial and environmental factors that affect access to safe drinking water.
- Understand the context of the secondary data sets analysed elsewhere in the thesis.

In order to achieve these aims various fieldwork activities were carried out in November 2009. Informal meetings were undertaken with some Egyptian households to enquire about problems of drinking water supply in rural areas. Field observations were made about drinking water issues particularly interruption, storage, treatment and private supply patterns. Observations were recorded using photographs and field notes were made after each visit recording drinking water patterns and whether any distinctive water infra-structure and facilities had been constructing to support the accessibility of safe drinking water. The theoretical framework of this field study is based on Rapid Rural Appraisal (RRA) as a convenient approach to be followed. The Rapid Rural Appraisal (RRA) is a range of investigation procedures using more informal methods, in particular applied to rural environments and communities. The method includes informal meetings, household interviews, and collecting information via direct observations at field study sites. It has been used as an effective method in socioeconomic research to avoid disadvantages of traditional research methods such as challenges of formal interviews, non-sampling errors, protracted fieldwork campaigns and high costs (Crawford 1997).

Visited sites were chosen so as to provide contrasting case study areas from Upper and Lower Egypt, with site selection further restricted to safe and accessible sites. Therefore, a plan to visit a cemetery site in ancient Cairo was cancelled because of safety issues. The hierarchy of Egyptian geographies and settlements outlined in section (4.4) was used as a basis for selecting contrasting case study areas. In Lower Egypt sites were chosen based on their locations near the Mediterranean Sea and Northern Lakes where Egyptian households suffer from water supply interruptions as a significant problem. On the other hand, isolated household clusters suffered from problems related to hand pump usage in the sites visited in Upper Egypt.
In this field work the observations were made in different areas of Egypt because of the great variations in water network supplies between regions and rural and urban communities. For instance, the Desert Governorates present specific water supply problems as they are located far from Nile River, the main source of drinking water in Egypt. Therefore, in addition to salty underground water, hand-pumps and wells are the main source of drinking water and present a problem for households living in those areas. Simultaneously, in the northern areas of the Delta Governorates where many villages located close to the Mediterranean Sea suffer from drinking water problems, particularly lack of capacity in public networks and interrupted supplies and contaminated water following the long journey of water from the Nile and other canals and drains. In these areas, agricultural pollutants and solid wastewater disposal are potential hazards impacting drinking water and human health.

In Upper Egypt as a zone of excessive population pressure, most cities, towns, villages and communities are much closer to the Nile, the main surface water source, compared with some Delta communities. This is because of the geographical fact that the alluvial plain of the Nile valley is very narrow and is surrounded by two plateaus from the east and the west. However, drinking water problems still arise where new settlements are on productive land with rapid conversion of critical agricultural land resources. These new isolated household clusters are usually not covered by any public utilities such as electricity and water networks. As a result, households traditionally dig wells and pumps to have access to water. Initially they use groundwater to produce mud bricks with clay soils and after finishing constructing the house the pump or well continues as a source of drinking water. Moreover, groundwater in Upper Egypt can potentially be contaminated with sewage disposal since there is no public sanitation network in many communities. Private sewage is a common pattern in the southern governorates and it represents a potential hazard for groundwater contamination.

Field trips to Upper Egypt

Field observation in Upper Egypt took place in Assuit Governorate which is located in the middle of Valley Governorates. El Ghanaem Markaz, where the field work observations were taken, is located on the west bank of the Nile River and close to the western plateau (Figure 3.11). Specifically, fieldwork concentrated on hand pumps as a private pattern of access to drinking water especially in the new isolated homestead clusters which have been growing rapidly on agricultural land.

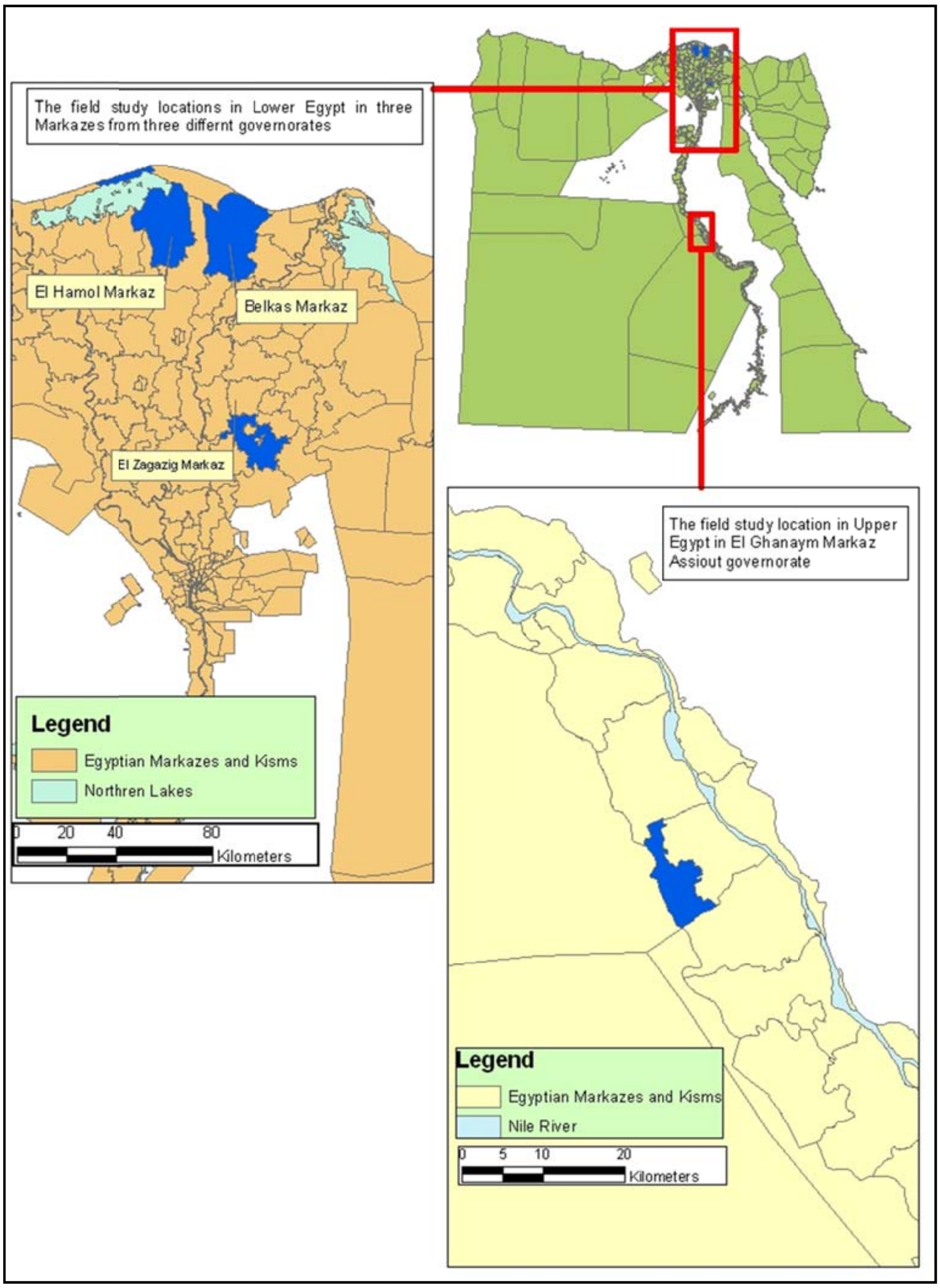


Figure 3.11 locations of field study sites in Upper and Lower Egypt

An interview was conducted with a 67 year old household member who has constructed a hand pump. He spoke in more detail about the benefits of constructing a hand pump when any household wants to build a new house as it provide people with water for many purposes:

- Irrigation - if the household continues to grow crops or fruit trees inside the new house for a while before constructing the house for living.

- As a water supply used in making bricks which were used in building the house. Water supply is the most important thing in brick-making where other materials are available free or at minimal cost such as appropriate types of soil. This process saves much money compared with buying bricks from the market.
- As a water supply for household members when they live in the new house, which continues to be used until the village council connects the area to the public water network and other infrastructures.

Households who use a hand pump to obtain drinking water are not aware of the risks posed by nitrate, arsenic, fluoride, and other toxic pollutants. Tests are not routinely performed before drinking from hand pumps. The interviewed household member confirmed that sometimes charity associations collect samples from pumps and send them to environmental and physical labs in the main capital of the governorate. The results are often positive but households who have pumps do not take any action regarding metals removals or disinfection treatment processes. It is believed that pumped water quality is better than tap water, “water comes clean, fresh and cooler than tap water. However, dozens of pumps have been abandoned after a change in water color and taste” the interviewed man added. Moreover, a great advantage of pumped water supply is that there is no interruption in water supply, so it plays an effective role when the public supply is interrupted. Members of households, usually young girls and women, fetch drinking water from the nearest pump. Although they walk a long distance and spend much time collecting water as no pumps are found on the village margins, they collect pumped water for free as a kind of community act for the village.

Historically, water was abstracted using bucket wells instead of constructing pumps but that was a long time ago when tools used for constructing hand pumps were not available. There were many bucket wells that had been constructed far away from the case study area but they are abandoned now either because the area is fully covered by the public water network or hand pumps have been constructed. Motorized pumps are rare in rural areas in Upper Egypt except for irrigation purposes in desert areas and new reclamation land. In fact, in most rural areas households do not need to construct motorized wells or pumps as tall houses with many floors also are rare. However, this kind of pump is sometimes constructed in some urban areas particularly slum settlements around city margins. Water quantity is not restricted in any of the sites visited, since private water supplies are widely used when the public network is interrupted. Regarding drinking water contamination, some households believed that water was potable, pure and fresh whilst others argued that scientific tests of most pumped water samples by charity organizations indicated contamination.

The observations suggest that water storage is not common. If any households store drinking water, this is not as a treatment process when fetching water from unsafe sources, but for making water cool or when a piped supply is interrupted. Storage containers include traditional water storage vessels such as the traditional Egyptian Zir in many shapes and other plastic containers and metallic tanks that are constructed on house roofs (Figure 3.12). However, Zir vessels are used widely because of the hot weather in summer compared with other Egyptian regions and when households do not have refrigerator.

Figure 3.12 Picture illustrates a roof tank for storing drinking water used as a solution against water interruption

Field trips to Lower Egypt

Lack of access to safe drinking water in Lower Egypt and Delta governorates is a significant problem and a major factor that affects daily life in many Egyptian communities. The sites which were visited are located in three governorates Kafr El Shekh, El Dakahlaya and El Sharkya. The first two sites are villages in El Hamol Markaz which forms one of the vast majority of markazes suffering from water shortages and lack of access to potable drinking water. Another site is located in Bilkas

Markaz where households struggle with interrupted water supplies and use the same storage methods to solve the problem, rather like El Homol markaz. Another site is Banyoos village in Zagazig Markas El Sharkya governorate where motorised pumps were observed as a private source of drinking water. In certain parts of Delta governorates, especially across northern parts near Brols Lake and Mediterranean Sea, most villages and rural isolated households suffer from daily interrupted tap water. According to some households living in these sites, water is not available for approximately 15 to 20 hours every day. They explained that throughout the last three months tap water was completely unavailable for more than two days every week. Sometimes, the Markaz council provided drinking water using truck tanks which came every day and stopped in every street to provide drinking water for households.

In certain northern parts of Kafr El Shekh and El Dakahlaya governorates, water storage water tanks can be easily recognised, constructed on house roofs. Primarily, households purchase water tanks when they build a new house, since they know that the water interruption problem will be ongoing and they cannot access drinking water without having these tanks. There are two main kinds of tanks either plastic or metal and they come in different size and shapes. A big plastic tank is preferred and used by rich households while a metal one is considered as a public tank and widely used. Some households prefer tanks with a black colour as they absorb solar radiation and make the water warm or even hot particularly in the summer. There are various benefits of this kind of water storage facility. Firstly, water tanks provide a household with their required daily water for both drinking and cooking. Secondly, these tanks guarantee tap water supply and make it available at all times during the day. Thirdly, piped water is stored in these tanks daily and reduces the need to fetch drinking water from outside the home or construct private sources and consequently use of untreated water. The problem of water interruption even affects public facilities which are constructed in the street to introduce drinking water for people such as the *Coldair,* a form of electric water cooler. It is usual that a household is in charge of a *Coldair* and the cooled water is available for public consumption. Typically, the responsible household constructs a tank over it to store water and to guarantee continuous supply (Figure 3.13).

Figure 3.13 Picture shows a form of public drinking water provision, the electrical cooler (Coldair) which is used instead of the more traditional Zir vessel

A meeting with some members of a household living in El Hamol Markaz revealed that although water storage and water tank facilities are public features of these villages of northern Egypt, digging pumps took place at an early stage alongside village construction, after the government allocated land in Kafr El Shekh and El Dakahlaya. The real problem in these villages lies with the public network supplies which do not provide these rural communities with sufficient drinking water. Therefore, households use their own facilities to solve drinking water access challenges either by piped water storage using water tanks or, depending on groundwater resources, by constructing private supplies such as pumps or motorized wells.

Access to drinking water and public water supplies in rural areas of the northern Delta governorates seems to be affected by various spatial and environmental factors compared with Upper Egypt:

- Most household communities in Delta governorates are located far way from either Rosetta or Damietta Nile branches or other drainage canals.
- River Nile water in the main Nile branches and canals is contaminated because of industrial wastewater from many factories, domestic pollutants and runoff or subsurface flow from irrigated agriculture.
- In many villages and towns, the sanitation system and sewage networks rely on open drainage and there are no safe underground sewage infrastructures. This contaminates underground water, the main source for many water plants and private supplies.
- Increasingly, rice cultivation requires large amounts of water and is completely concentrated in the Delta regions, which leads to water shortages and impacts on water delivery.
- Groundwater quality is low in all towns and villages located close to the Mediterranean Sea and the shores of Egyptian Northern Lakes.
- Saline intrusion produces a variable water taste and most households use piped water just for washing. Moreover, water interruptions happen on a daily basis because of the lack of capacity in drinking water plants.
- The public supply network in these rural areas has not always met global and national standards for providing households with access to safe drinking water supply. In addition, home drinking water treatment such as filtration is rare and this may affect human health and increase the rate of morbidity and mortality.

Conclusion

This field study has drawn clear evidence that although access to safe drinking water is a significant challenge in many Egyptian communities, particularly in rural areas, there are great variations among Egyptian regions in terms of water supply patterns and potable drinking water delivery. In most of the rural areas of Upper Egypt private supplies such as pumps and wells are the most common pattern, related to housing growth on agricultural land and brick industries. In contrast, field observation suggests that northern rural areas of Delta governorates are suffering from water shortage and interruptions, which has created a specific water storage strategy that can be recognised easily as a part of any community structure. Urban households in all Egyptian regions have better access to piped water than those who live in rural areas.

However, motorized pumps were identified as the preferred solution in some marginal settlements and slums areas to the unreliability of piped water supplies. Groundbreaking work on drinking water supplies in developing countries entitled “Drawers of water” was undertaken by White et al. (1972) based on field work and observations from 34 rural and urban communities in three countries of East Africa: Kenya, Tanzania and Uganda. The study discussed the use of drinking water in the domestic life of households. In addition to this, the impact of supply patterns on household health was highlighted. After almost 30 years Thompson et al. (2001) published ‘’Drawers of Water II’’ which involved field observations on both health and water in the same study sites as the original White et al. study. Such long-term studies have drawn clear results about problems of access to drinking water in developing countries; they are limited for communities of East Africa and it would have been more useful if these studied had considered the spatial variations and focused on different countries inside Africa or even outside. Drawers of Water raised many questions about households who were spending much time fetching drinking water and how they coped with difficulties in accessing drinking water sources. Moreover, it showed how households in East Africa were regularly using traditional vessels to store water. After three decades and after new water infra-structure had been installed, improvements in water supplies were expected to happen. However, Drawers of Water II showed that households are still suffering from poor water delivery and spend much time in collecting water as a result of breakdowns and deficiencies in water services. Consequently, people still collect and store water but in plastic containers rather than traditional ones. This field study suggests that in spite of the majority of Egyptian households having access to public piped water, they have developed unique strategies such as water tanks and motorised pumps to cope with interrupted supplies and poor water delivery.

In conclusion, water supply problems in Egyptian governorates vary geographically and may impact on human health as a result of lack of access to safe water. In most cases, consuming untreated water from private supplies or storage of water for a long time as a solution to interruptions and the former in particular may cause serious infectious diseases and illness. One of the most important findings to emerge from this study is also that rapid appraisal techniques such as field observations and household interviews can support secondary data analysis by providing the researcher with additional context and understanding.

3.2.5 Linkage of DHS and census datasets

Diagram 3.14 shows the GIS linkage process of the Egyptian DHS and census datasets. The process involves linking two kinds of attribute data via the associated spatial features. Firstly, the DHS attribute data has been processed using the SPSS package to obtain the percentage for each variable. This process starts with the recoding function to manage, rename, and select the variables needed for the analysis. An aggregation command has then been conducted for each variable where cluster number was the break variable. These two steps are followed by computing the percentage of each variable, representing either households or persons. The result is an SPSS file that includes the percentage of each variable in each cluster and this can simply be linked with the GPS coordinates for mapping and other GIS analysis.

The second stage involved maintaining and reorganizing the demographic and socioeconomic variables of the 2006 Egyptian census to be aggregated with sub provincial administrative boundaries (kism/markaz level). This was followed by computing the percentage of each variable. Thirdly, the most important process is spatial joining to link the GPS coordinates, as a target feature, with Egyptian sub provincial boundaries, as a join feature, through an overlay function in ArcGIS. This process aims to create one database file containing each cluster number within each administrative unit (kism or markaz). Therefore, SPSS was used to link this file which contains cluster numbers and areas names with DHS variables through the merge and add variables functions and selecting administrative names to be added. In addition, the aggregation function was used, choosing admin names (kism/markaz) as the break variable. The resultant file contains household clusters in each administrative area for comparison purposes and GIS analysis.

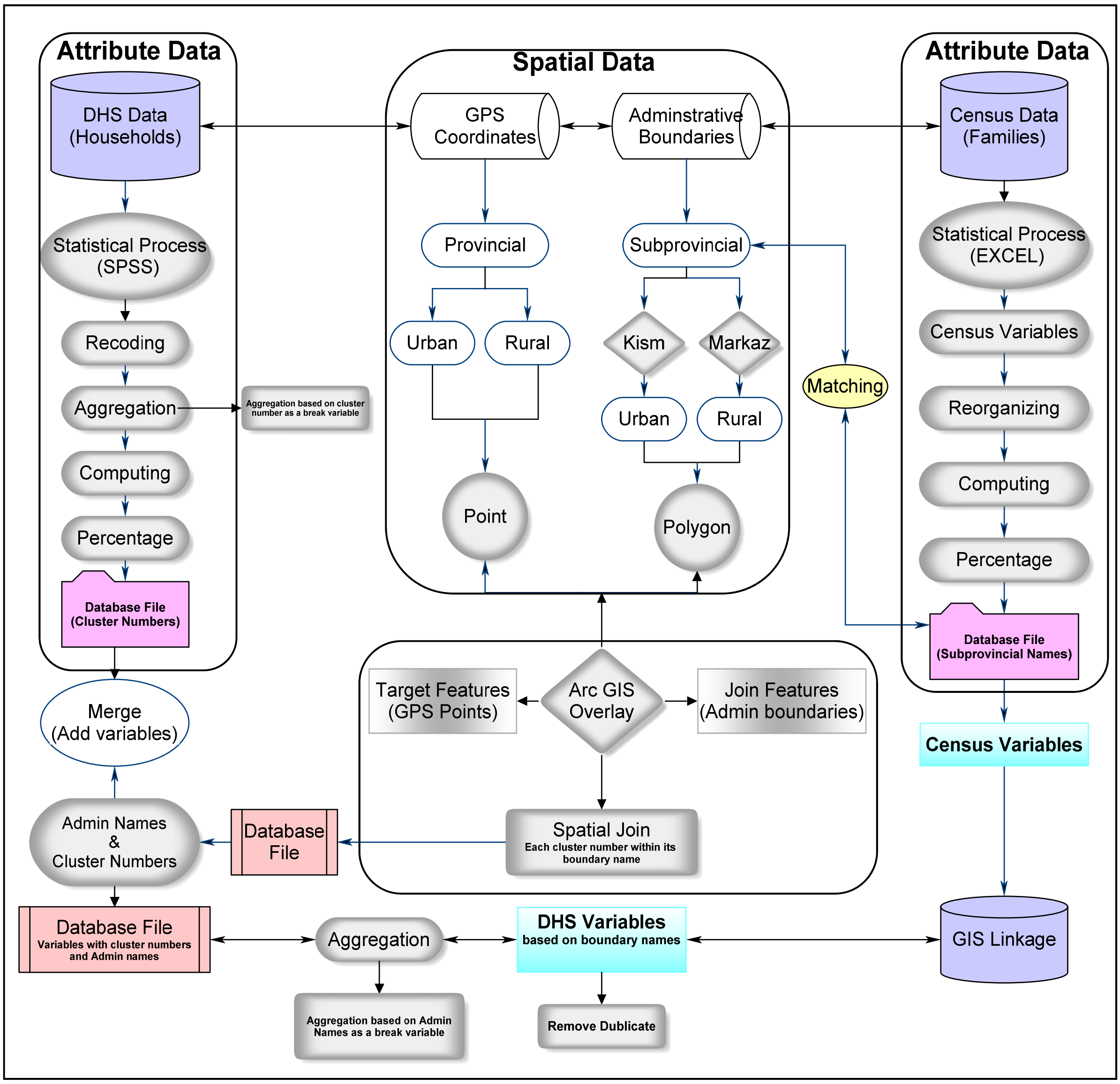


Figure 3.14 GIS linkage process of DHS and census datasets

3.3 Spatial error and modelling

3.3.1 Spatial data validation

The utility of the DHS can be considerably enhanced by linkage to available census data, which provides coverage of the entire population, albeit with less detailed information on some key health and lifestyle topics. However, these datasets cannot be directly linked using standard published lookup tables, as might be possible in a developed world setting (Simpson & Yu, 2003) and it is therefore necessary to develop appropriate GIS approaches for spatial linkage between the two sources. However, an assessment of the positional accuracy of the DHS and its implications for GIS-based linkage to census data is essential (figure 3.15). The focus here is on the assessment and accommodation of spatial uncertainty in the DHS to census linkage, so as to facilitate an important enhancement to the data available for analysis.

There is an extensive literature on spatial uncertainty, its impacts on subsequent analysis and data linkage operations, and its communication to spatial data users (Heuvelink & Burrough, 2002). Within the over-arching concept of uncertainty, Fisher (1999) distinguishes between error affecting well-defined spatial objects, and vagueness and ambiguity which affect poorly-defined spatial objects. Since the spatial entities in this study (census units and GPS clusters) are well defined, the primarily concern is with error, namely the difference between locations or attributes held digitally and their true values. Fisher (1999) also distinguishes a further sub-category of uncertainty in spatial databases that he terms controlled uncertainty, in which uncertainty is deliberately introduced into a socio-economic data set. Given the deliberate perturbation of locational data in the DHS, this concept is also relevant to this case study.

More specifically, in this case study the concern will be with a specific form of spatial data manipulation. The process entailed a point-in-polygon operation, with the points being DHS GPS clusters and the polygons being census boundaries. Several authors have developed procedures for handling uncertainty in point-in-polygon operations, for example using probability density functions or fuzzy set theory (Leung & Yan, 1997; Cheung et al. 2004) or by reducing the problem to a point-in-triangle problem (Leung, et al. 2004). As an alternative approach, Hunter (1999) systematically perturbed the co-ordinates of points and polygon boundaries according to known probability density functions. He then used Monte Carlo simulation to assess the influence of positional uncertainty on point-in-polygon results. However, these methods require that the characteristics of any positional uncertainty are known from the outset. Ratcliffe (2001) characterised uncertainty in point-in-polygon operations involving address locations and census tracts without any prior knowledge of the uncertainty characteristics of the two data sets used. His approach involved a comparison of two different methods (geocoding and cadastral) of generating address locations. However, in this study, no information was available concerning the positional accuracy of point and polygon data. This is not uncommon in developing world contexts. Furthermore, generating locations for point or polygon data via different methods was not possible. Therefore, rather than following the methodology of these previous studies, the approach was to identify attribute labels that were inconsistent with the locational data for points and polygons.

Information about feature attributes that were inconsistent with location was used to model uncertainty in the point-in-polygon operation undertaking.

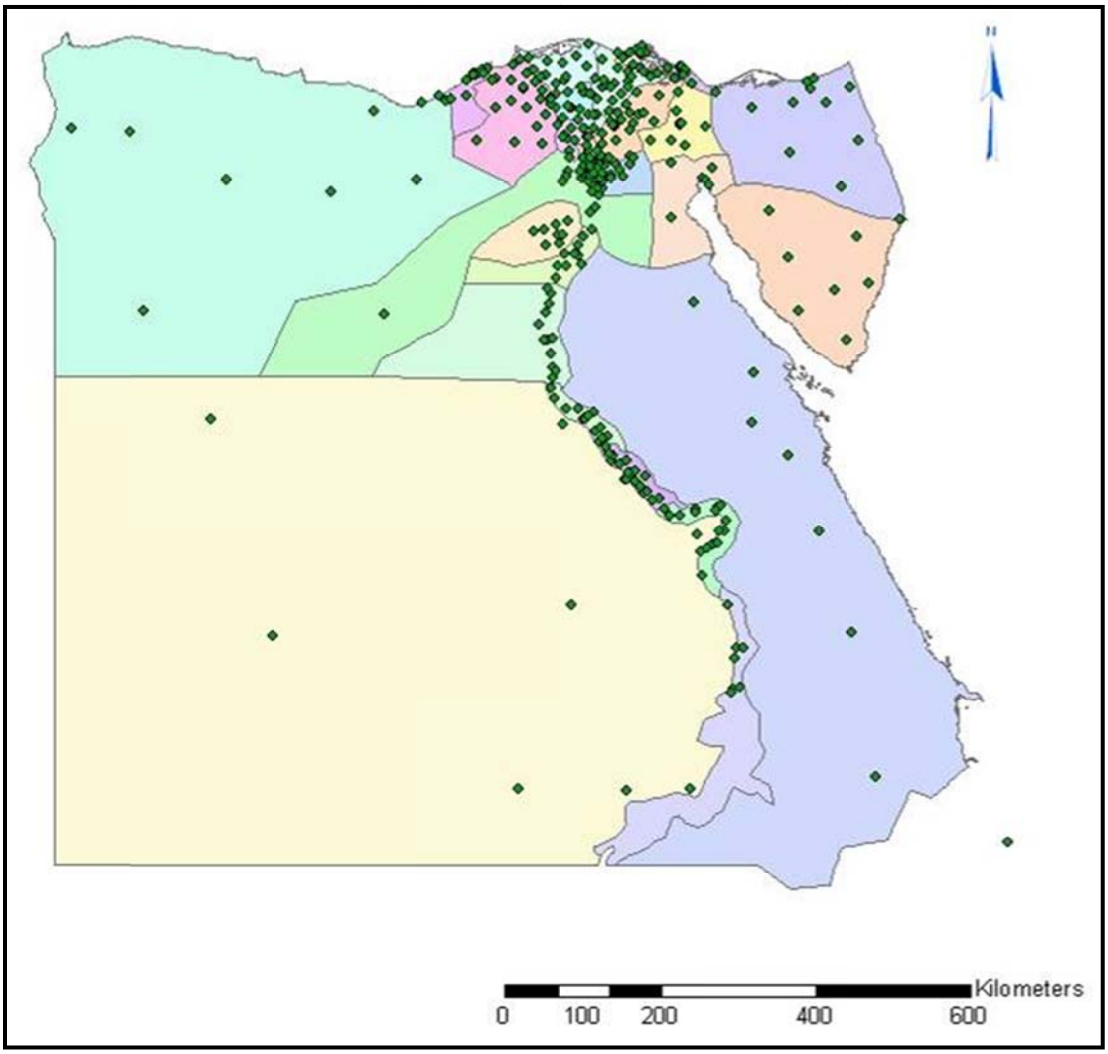


Figure 3.15 Distribution of DHS GPS clusters within 2005 governorate boundaries

3.3.2 Spatial error of GPS clusters

The accuracy of GIS datasets is a matter of great importance when researching spatial phenomena. This requires measuring, validating and focusing on spatial error that might be present in datasets. In general, this is particularly relevant to some GIS data sources, such as surveys and sampling design, where the probability of inaccuracy and uncertainty is likely to be higher than for other sources (Zhang & Goodchild, 2002). Goodchild (1989) reported that although different kind of GIS processes have been applied to spatial data to produce a variety of graphical and tabular outputs, little regard has been given to the level of error that might occur. According to Gong & Mu (2000) the spatial error in GIS is the difference between the measured value and true value and relates to uncertainties of positional accuracy. There are many kinds of spatial error which can occur in the coordinate values of polygons, lines and points.

One objective of this study is to make a comparison between specific census and DHS variables that might help in a clearer understanding of the spatial context of drinking water and health in Egypt. For this purpose, overlaying the two datasets, GPS point and census polygon, in ArcGIS was initially undertaken to examine the extent to which each variable in the DHS data agreed with the same variable in the census data. Such overlay analysis can play an important role in GIS applications in terms of explaining, assessing and identifying spatial error in data sources (Leung et al. 2004).

When overlaying polygons of Egyptian census boundaries with GPS cluster points, it is necessary to look at the distribution of DHS cluster points and it is to be expected that each cluster point will fall within the correct governorate. However, spatial error has been detected where there are 10 cluster points which fall entirely outside Egyptian boundaries. Figure 3.16 illustrates how some of these misallocated points were detected. Five points coded to be within Alexandria Governorate are located in the Mediterranean Sea. This implies that there may also be some GPS points not in their true positions inside Egyptian boundaries. Undertaking a spatial join operation between the two data layers in ArcGIS could capture this error. The outputs illustrate that there are 63 points whose spatial locations and governorate codes do not agree (Figure 3.17).Obviously, those points are distributed closer to governorate boundaries, in particular Greater Cairo region, Valley, Delta, and Costal Governorates (Figure 3.18). However, no spatial error in cluster points was found in the Frontier Governorates such as New Valley, North and South Sinai Governorates where no communities are found close to borders. In spite of Matrouh being considered as a Frontier Governorate with a large area, it has one cluster point with spatial error falling in the Mediterranean Sea. Furthermore, two cluster points in Red Sea Governorate fall within the sea. It is generally observed that except for North and south Sinai Governorates all other coastal governorates have spatial error of at least some GPS points. However, just four governorates have GPS points which are located in the sea; Alexandria, Matrouh, Port Said, and Red Sea, while other governorates such as El Behera, Kafr El Shekh, Dakahlyia and Damietta show spatial error inside their boundaries.

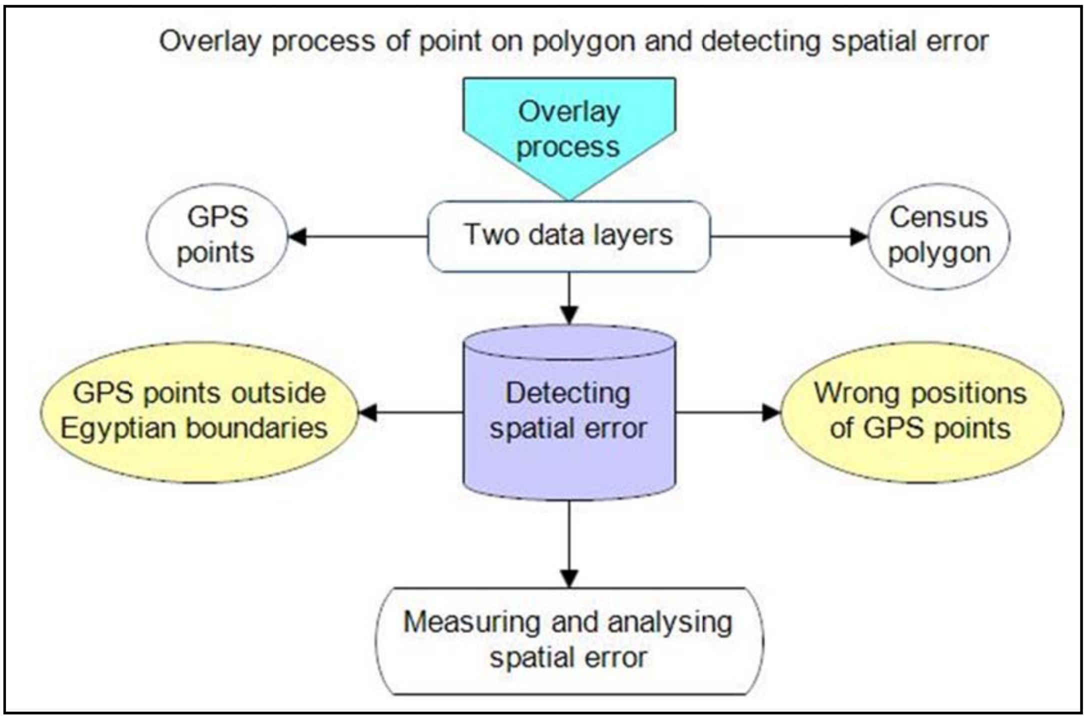


Figure 3.16 Flowchart illustrating how positional error in misallocated GPS points was detected

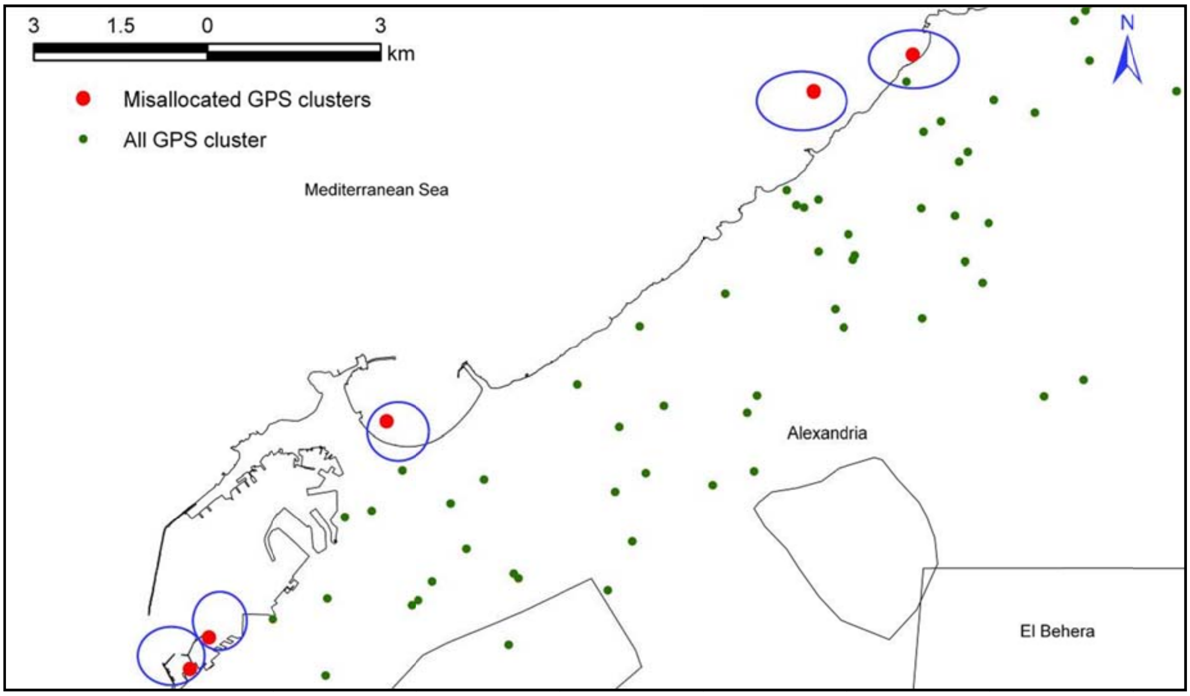


Figure 3.17 Positional error of GPS points in Alexandria Governorate

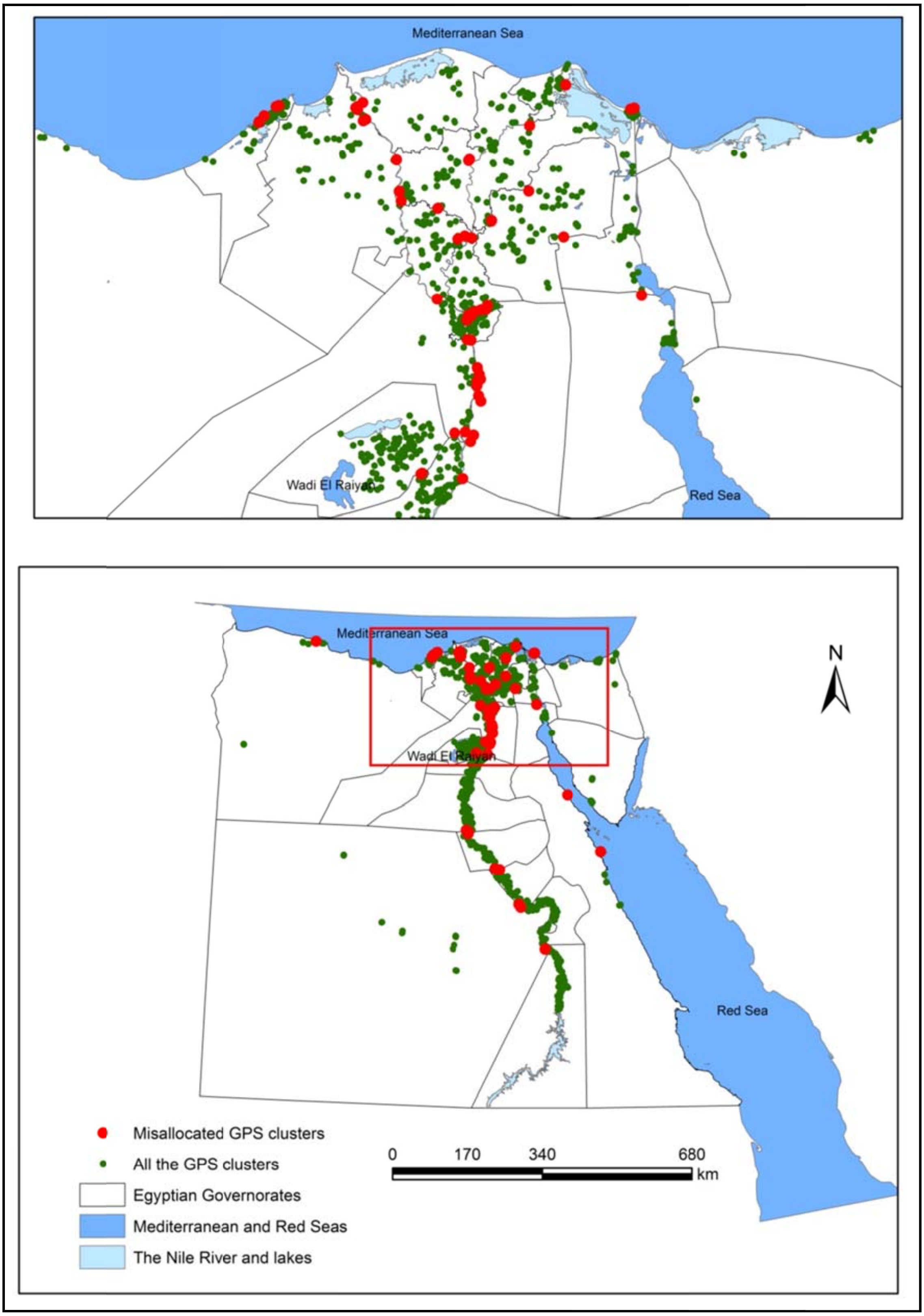


Figure 3.18 Distribution of positional error of GPS cluster points in all Egyptian Governorates

3.3.3 Measuring spatial error

Previous studies have reported that spatial error has been given attention and identified as a major problem that might arise from different sources (Shi et al. 2003). These sources could be related to the spatial data themselves or arising from GIS and spatial analysis processes. In addition, measuring positional error appears to be increasingly as a necessary operation in spatial analysis and GIS research, in addition to efforts to improve locational accuracy of any georeferenced data that are being used. This is an important process when spatial error may affect the findings of statistical or spatial analysis (Griffith et al. 2007).

Three possible sources might cause this apparent positional error: incorrect coordinates, incorrect projection or incorrect labelling. Incorrect coordinates would refer to individual GPS clusters being recorded in the wrong locations: this seems possible and requires further investigation. Incorrect projection would refer to one or more entire data layers being wrongly georeferenced. Examination of both data layers here shows that they have the same projection and errors are widely distributed across the study area, therefore it does not seem to be the problem. Incorrect labelling can expected in survey data collection using GPS methods. However, this kind of error does not seem to be present because the majority of misallocated points are concentrated around governorate boundaries. In addition, 10 cluster points fall within the Mediterranean and Red seas. Therefore, it seems that the apparent spatial error must be mainly due to GPS coordinate error.

In this study, exploration of the positional error of the GPS clusters when overlaid on census boundaries is required, calculating distances between each polygon, governorate, and each GPS point. The main target of this process is to know the distance between each point with a positional error and its correct governorate. ArcGIS software did not offer a clear or direct way to apply the necessary point-in-polygon distance operation and capture the resulting distance distribution. After several experiments using ArcGIS and Idrisi software, the most successful and efficient method for calculating these distances was determined as follows, which is based on raster processing in Idrisi (Figure 3.19):

Step 1 Importing shape files

The two Shapefiles, governorate polygons and GPS cluster points were imported into the Idrisi software, based on the latitude and longitude referencing system and degrees reference unit.

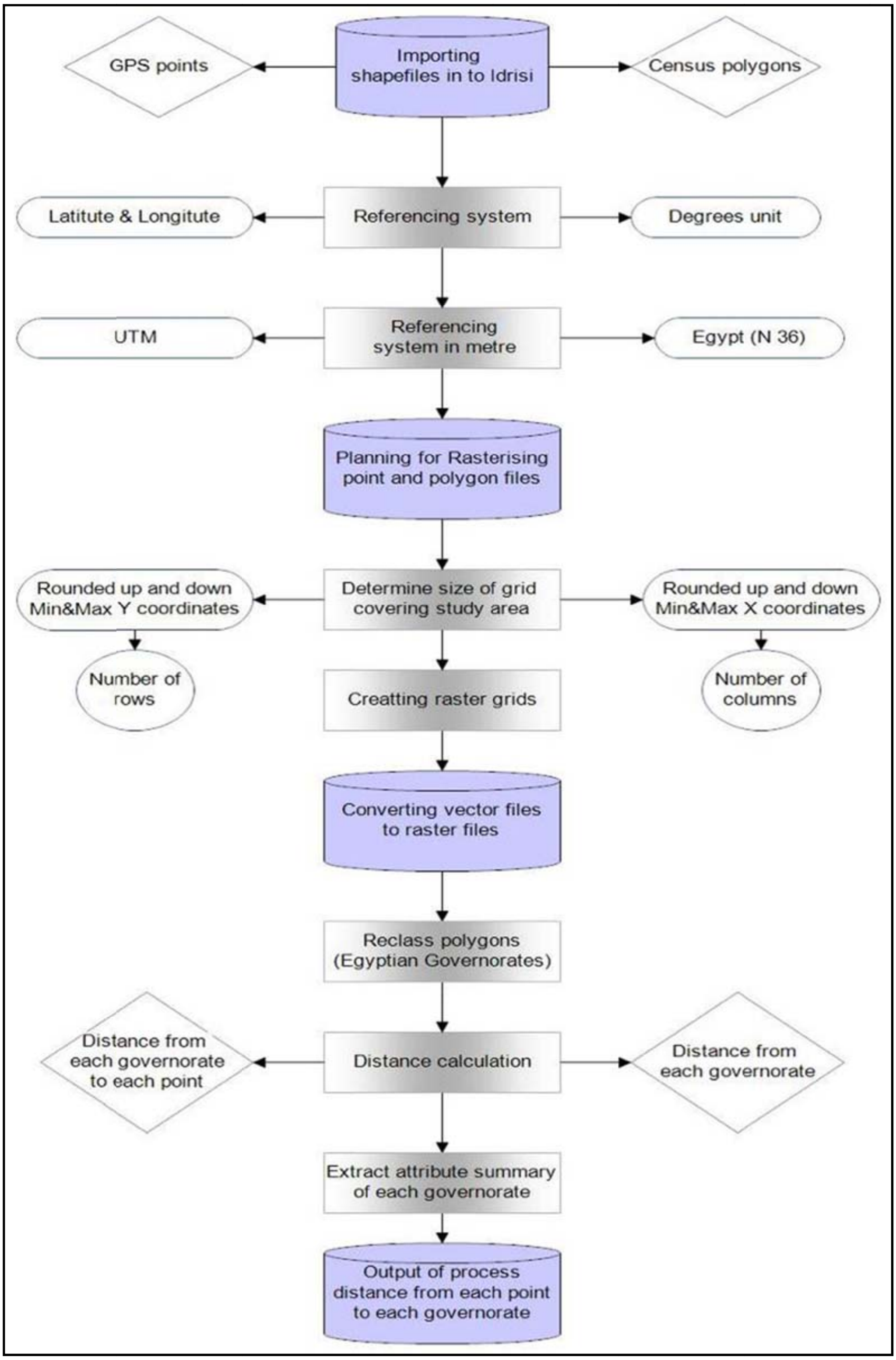


Figure 3.19 Flowchart representing the GIS process of point to polygon distance calculation in the Idrisi software

Step 2 Converting shape files

In this step, Shapefiles were converted to a reference system in metres which is, UTM N36, the Universal Transverse Mercator (UTM) and zone 36 in the northern hemisphere where Egypt is located (N36).

Step 3 planning the rasterization process

This step is necessary in order to determine all values that are required in the raster initialization process in Idrisi. Firstly, the VDC data files from the previous step were opened and minimum X&Y and maximum X&Y values were read and recorded as in table 3.1. Secondly, 100m x 100m was chosen as a grid cell size that would cover the entire study area while allowing relatively small errors to be detected. Finally, the minimum and maximum coordinates were rounded to exact multiples of the grid cell size. The number of columns and rows of the grid based can be determined by subtracting minimum from maximum coordinates and dividing the result by the cell size (100m):

Number of columns = 899000 + 326300 / 100 = 12253

Number of rows = 3532300 – 2455400 / 100 = 10769

Coordinate Systems	Original	Rounded up
Minimum X	-326334	-326300
Maximum X	898922	899000
Minimum Y	2455449	2455400
Maximum Y	3532299	3532300

Table 3.1 X & Y coordinate system in the initial command of rasterising process

Step 4 Creating raster grids

From the last step, all spatial parameters needed to generate blank raster grids have been specified. Blank grids can therefore be initialised for census polygons and GPS points.

Step 5 Rasterising vector files

The governorate polygons and GPS points are rasterised using Idrisi's "polygon to raster" and an empty target raster from Step 4 is selected as the target. With the "point to raster" command "Change cells to record the identifiers of points" is selected as the operation type.

Step 6 Reclass operation

The purpose of this step is to reclassify the image file of polygon boundaries to pick up each governorate separately. So, 29 polygons (28 governorates plus Luxor City) have been classified with their names.

Step 7 Calculation distance from each polygon

Using the distance operator within the "GIS Analysis" commands, the Euclidean distance is measured between each cell and the nearest polygon feature.

Step 8 Calculation distance from each polygon to each point

The last Idrisi step is to extract governorate identifiers and distance values for each GPS cluster point. The results for each governorate are sorted in a file named file.avl where "file" is the name of the governorate. Finally, file.avl can be opened by Microsoft Excel

Looking at the result, each file.avl contains two columns, the first column contains the serial numbers of DHS Clusters and the second distance in metres from each point to the relevant governorate. Zero values indicate that the point falls inside the governorate border.

3.3.4 Analyzing spatial error

Spatial error of GPS points was measured in terms of distances from cluster points to governorate boundaries. Distance from each point to each governorate polygon was obtained and all misallocated points were identified by comparing their governorate code from the DHS with that in which the point is located. Moreover, recognising cluster points falling outside Egyptian boundaries, mainly within seas, was represented by applying MIN formula in the spreadsheet for each record. Therefore, any record which does not show zero as a minimum value indicates that the cluster point falls outside all Egyptian Governorates and thus outside the boundary of Egypt. From the finding, it is apparent that 10 points are misallocated outside all polygons, eight of them in the Mediterranean Sea. One point belongs to Matrouh, five belong to Alexandria, and two points belong to Port Said Governorate. On the other hand, two points fall within the Red Sea belong to Red Sea Governorate (table 3.2).

DHS Cluster ID	Egyptian Governorate	Distance in Metre	Location Outside boundaries
510301	Alexandria	806.23	Mediterranean Sea
590307	Alexandria	100.00	Mediterranean Sea
740106	Alexandria	360.56	Mediterranean Sea
800403	Alexandria	100.00	Mediterranean Sea
810502	Alexandria	360.56	Mediterranean Sea
6680101	Matrouh	707.11	Mediterranean Sea
910302	Port Said	100.00	Mediterranean Sea
970102	Port Said	1000.00	Mediterranean Sea
6490506	Red Sea	223.61	Red Sea
6540501	Red Sea	100.00	Red Sea

Table 3.2 The number of the DHS GPS points falling outside governorate boundaries

Overall, the total number of misallocated points is 63 of which the mean positional error is 1,410m. The minimum distance is 100m (one cell width) while the maximum distance is 6,122m. Thus, all points located further away from boundary lines than this distance are considered to be within their correct polygons. Figure 3.20 illustrates the distribution of positional error for misallocated points. It can be seen that the majority of misallocated points have a positional error of distance between 400 to 2,000m. In addition, roughly 20% of points show positional error less than 200m while 10% of the points show positional error between 200m to 400m. A low percentage (6%) of the cluster points have positional errors of more than 4,000m and less than 5,200m. This means that the frequency of mislabelled points at shorter distances is greater than that at longer distances.

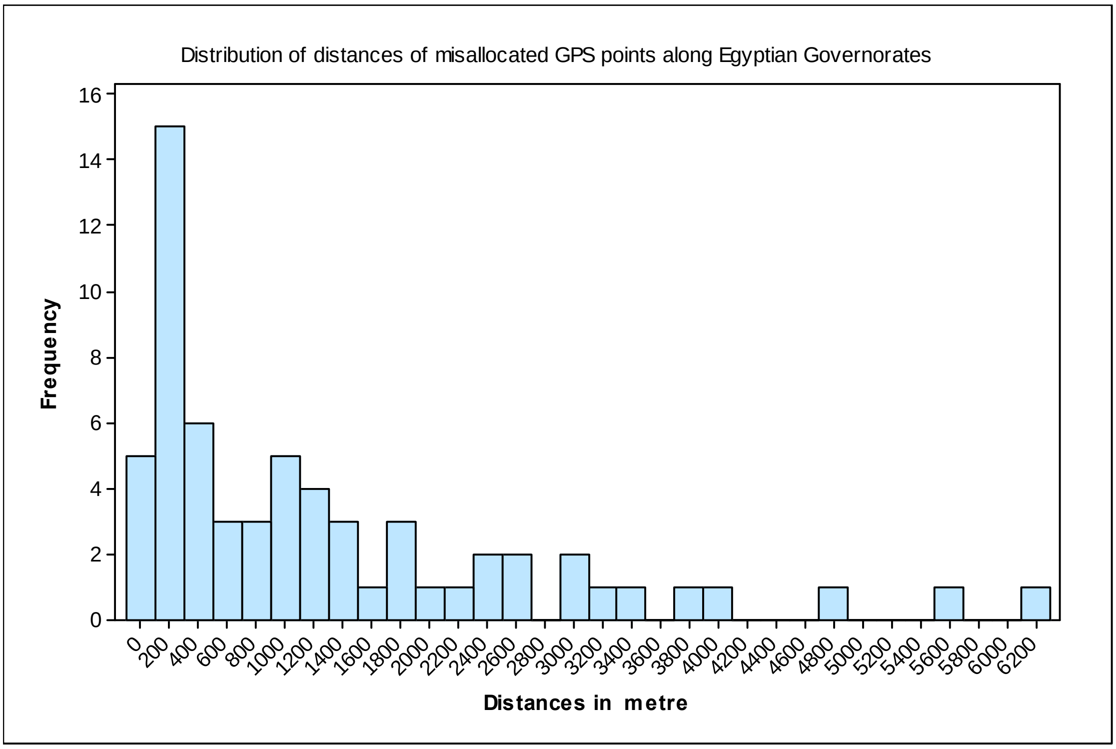


Figure 3.20 Histogram representing the distribution of distances for all misallocated points and frequency of positional error in metres

3.3.5 Modelling the spatial error in DHS clusters

In statistical and spatial analysis when the outcome variable consists of two possible values, logistic regression is an appropriate approach for modelling the relationship between two variables. Therefore, it has become a standard method for describing the relationship between a dependent variable, the outcome or response, and one or more independent variable(s), the predictor(s).The difference between linear and logistic regression is that the dependent variable in the logistic regression is always binary or dichotomous having values of zero or one (Hosmer and Lemeshow, 2004). Several studies have investigated the popularity of logistic regression model (Kleinbaum & Klein,1992; Hosmer et al. 1991; Allison 2001). They observe that dichotomous variables are common in the social sciences, so the researchers in these disciplines apply logistic functions to estimate the relationship when the outcome variable is dichotomous. Furthermore, logistic regression models are widely available using standard software.

Is there spatial clustering?

Detecting unusually high aggregations of data units from those expected is a common process in spatial data analysis. A cluster is a high density of points in space where the distance between any two points in the cluster is shorter than the distance between any point in the cluster and any outside it. Where clustering is present, there are usually geographic factors behind clustering phenomena (Laswon & Dension 2002; Jain and Dubes 1988). Cluster analysis is a technique which aims to detect the grouping patterns found in a dataset and it is widely used. Nonetheless, every research discipline may have different techniques and algorithms for cluster analysis and exploring spatial data structure (Kogan, 2007).

A test for global clustering is used for investigating whether clustering can be detected in the entire study area or not. A global clustering test is common in the epidemiology and public health domains where researchers are looking for clustering patterns that may occur because of infectious diseases (Kulldorff et al. 2006). Haining (1998) has shown three types of clustering related to epidemiology and public health research: higher risk around a prior identified point, global clustering and local clustering. The first kind requires a focused test where the hypothesis tests whether the pre-specified local cluster displays unusually elevated values. The second type is known as a global clustering which examines whether there is any clustering pattern present throughout the study area. This pattern is found when there is a disease case near another case and high global clustering means all cases are close to each other. The third type determines locally whether specific cluster locations are significant or not.

In order to test whether the 63 DHS GPS clusters are clustered or not, the ClusterSeer Software was chosen because it offers a wide range of methods for testing the pattern of spatial clustering. A global spatial test was applied to answer the question whether there are any unusual spatial patterns. The null hypothesis is simply that there is no clustered pattern (ClusterSeer, 2009). Columns containing the ‘cases’, the GPS clusters with spatial errors, and the ‘controls’, the GPS clusters in their correct places, were joined to the whole GPS cluster shapefile to be processed in ClusterSeer. Cuzick and Edwards’ test is an appropriate test which detects global clustering in individual level case-control data when the cases are coded as (1) and the controls are coded (0) (Cuzick and Edward 1990). Therefore, it was the appropriate test to be performed on the data. The output of Cuzick-Edwards k-nearest neighbour test (table 3.3) shows evidence of clustering patterns. In the output of the test k indicates the number of neighbours while T/k refers to how many cases neighbour other cases. The upper tail P-value (0.000) and P-value via Monte Carlo simulation (0.001) are both significant. This means that the null hypothesis is strongly rejected. The significance of the cluster statistic can be evaluated by looking at upper-tail P-values and P-values via Monte Carlo simulation that are both provided by ClusterSeer. The first value represents the probability of the null hypothesis and a small value indicates that the result is significant (and vice versa).The second value is generated by ClusterSeer for each k by scrambling all the data, cases and controls, randomly for each of the geographic locations. Monte Carlo randomization is obtained in three steps: randomizing the observations in the original dataset, recalculating the statistics for the randomized data and repeating the previous two steps many times. The output from each simulation run is used to produce the P-value for the observed statistic by comparing the observed statistic with the reference distribution. Combined P-value for Monte Carlo randomization provides two values: Bonferroni P-value and Simes P-value. The Cuzick and Edward method does a multiple test which calculates a statistic at each value when k>1. In this calculation a combined P-value will be created for all tests that were performed at one initial alpha level (0.05). The two values can be compared with the alpha level to check whether the tests provide a significant result or not.

k	T[k]	E[T]	Var[T]	z	Upper-tail-P value	Monte Carlo P-value
1	24	3.011	4.627	9.758	0.000	0.001
2	42	6.023	9.361	11.758	0.000	0.001
3	56	9.034	14.29	12.42	0.000	0.001
4	62	12.04	19.21	11.39	0.000	0.111
5	74	15.05	24.27	11.96	0.000	0.001
6	83	19.06	29.23	12.00	0.000	0.009
7	90	21.08	34.29	11.76	0.000	0.035
8	97	24.09	39.40	11.61	0.000	0.035
9	104	27.10	44.58	11.51	0.000	0.028
10	112	30.11	44.71	11.61	0.000	0.022

Combined P-value for Normal Approximation

Bonferroni P-value: 0.000; Simes P-value: 0.000

Monte Carlo randomization summary

Number of Monte Carlo simulations = 999

Combined P-value for Monte Carlo Randomization:

Bonferroni P-value: 0.010

Statistical Distance Test statistic = 18.44

Simes P-value: 0.006

P-value from Monte Carlo simulation = 0.001

Table 3.3 The output of Cuzick and Edwards' cluster test for the GPS points with spatial error

The first model

This section aims to model the spatial error of the GIS linkage process between GPS clusters and subnational census boundaries, kism and markaz level, and to measure the probability of each cluster point being in its incorrect spatial location. This may be quite useful to inform the confidence with which we can make any comparison between the two datasets in terms of the socioeconomic variables being addressed. Exploring the geographic positions of the GPS clusters that were identified with a spatial error and that fall outside their correct governorates (see figure 3.18), it can be seen that most of the points with spatial errors are located very close to administrative boundaries and therefore distance from these boundaries is likely to affect the spatial error pattern. Thus, measured distance to governorate boundaries was chosen to be an independent variable in the regression model.

Using GIS analysis in Idrisi software was an appropriate method to convert kism and markaz polygons into lines and calculate distances from them. Logistic regression was applied to fit an equation for these distances and obtain the probability of each DHS cluster being assigned to an incorrect polygon. Data analysis was performed using the Stata software to fit the logistic regression relationship. The distance from each DHS cluster to the nearest governorate boundary was the independent variable in the regression equation while the dependent variable was dichotomous (0, 1) where each cluster point in the correct governorate polygon has been coded zero as there is no spatial error. On the other hand, a value of one has been assigned to the 63 clusters points that had been identified as being in their incorrect polygons. Logistic regression was then applied to Kism and Markaz distances using the two coefficients of the previous equation. The descriptive statistics and outputs of first regression model are shown on table 3.4. The value of the Prob>chi² reveals that the model is statistically significant and although the coefficient value of the independent variable, governorate distance, is very small (-0.0007), it is statistically significant since P>Z value = (0.000) which is less than alpha. The P value is very close to zero and the Z value (6.15) is less than -2. The independent variable in this model consists of very large distances and for this reason the coefficient is so small. To modify the large distances and minimize their impact during model fitting, a square-root transformation was applied to governorate distances. This step was considered in the second model to improve the performance of the model in predicating the relationship between the spatial error (the dependent) and distances from governorate boundaries (independent variable).

Variables	*Coef.*	*Std.Err*	*Z value*	*95% conf.*	**P> Z**
Governorate distance	**-0.0007**	**.0001**	**-6.15**	**-0.0009**	**0.000**
Constant	**-0.93**	**.20**	**-4.57**	**-1.3**	**0.000**

Log likelihood = 183.17576 Prob > chi² = 0.000

Pseudo R^2 = 0.2733 LR chi² (1) = 137.75

Linktest **hat** (P>|z|) 0.000 Linktest **hatsq** (P>|z|) 0.002

Table 3.4 The output of the first regression model

The second model

Assessing the logit of the model outcome involves checking whether all relevant predictors have been included or not. In other words, it is important to consider the linear combination of the predictors and whether the relationship between the logit outcome and the predictors is linear or not. In either case there may be a specification error in the model and refitting it is strongly recommended. Stata offers a useful command named Linktest which can be applied to test the model output and consequently detect this kind of error. Preliminary work on examining the adequacy of the linear model and its link function through linktest was undertaken by Pregibon (1980). The main idea of Linktest is that after constructing the model properly, finding any additional predictor that is statistically significant and needs to be added, should not be found except by chance Pregibon ,1980. The output of the Linktest command consists of two components: linear predicted value (**hat**) and the same linear value squared (**hatsq**). The value of **hat** should be significant as it is the predicted value in the model otherwise the whole model is misspecified. By contrast, **hatsq** value should not be powerfully predictive except by chance. Therefore, the interpretation of the linktest command outputs depends on the hatsq value and when this is significant refitting the model is necessary, because either the link function is not accurately specified or other variable(s) need to be included (Chen et al. 2009).

By applying the Linktest command to the first model (table 3.4) it can be seen that the P-value of **hatsq** is (0.002) which is significant despite the **hat** value (0.000) confirming that the independent variable is a meaningful predictor. Therefore, the model needs to be refitted either by including a new variable or changing the included predictor variable. Thus, a square root formula was applied to the governorate distance and the output was used as a new predictor in the second model. After experiment with many

functions such as opposite, reciprocal and slope, the square root of distance was chosen which is a widely used as a function of distance.
The outputs of the second regression model (table 3.5) show an improved fit to the data. There were many advantages to using the squared root of distance:

- The calculation output intuitively makes sense for clusters that lie exactly on a boundary. When any distance equals zero the probability of being in their correct place is approximately 0.5.
- Squared root is a common transformation method in spatial analysis research. In particular Inverse Distance Weighted (IDW), when used in interpolation, gives larger weights to the cluster points nearer to the predicted position than those further away (Naoum and Tsanis, 2004).
- Similarly, the square root model fits the available data slightly better as the Pseudo R^2 value is slightly greater (0.29) than in the first model (0.27). Therefore, the model presents a better fit and it is slightly more powerful at predicting the probability of the error in the spatial location of DHS clusters.
- The regression model based on the square root of distance passed the Linktest where the linktest was not significant and hatsq was larger than 0.05.

Variables	***Coef.***	***Std.Err***	***Z value***	***95% conf.***	**P> Z**
√(Governorate distance)	-.0587	.0073	-8.08	-.073	**0.000**
Constant	**.048**	**.28**	**0.17**	**-.499**	**0.863**

Log likelihood = -179.67576 Prob > chi^2 = 0.000
Pseudo R^2 = 0.2871 LR chi^2 (1) = 144.75
Linktest hat (P>|z|) 0.002 Linktest hatsq (P>|z|) 0.981

Table 3.5 The outputs of the second regression model with squared roots of distances

The probability of each cluster point being in the incorrect kism/markaz was calculated by applying the logistic regression coefficients of the Stata model for all kism/markaz distances using an Excel sheet. The logistic regression equation is the following:

P =1/ (1+2.718^ (-1(-0.0586306*(SQRT (K/Mdist+0.0481082)))))*

Where P is the probability of being in an incorrect Kism/markaz and K/Mdist is the distance for each cluster point from the nearest kism or markaz sub-administrative boundary. The next step is calculating the probability of being in the correct position as *q= 1-P*. The results obtained from applying the previous equation are presented in figure 3.20. The majority of cluster points which are located close to boundaries by less than 1000 metres show a high probability of being in incorrect places. Nonetheless, the values vary between 0.2, 0.3 and 0.4. Moreover, a value of 0.5 indicates that a cluster point is lying exactly on a boundary line and demonstrates an equal chance of being within the correct or incorrect kism/markaz. On the other hand, as the distances from the boundaries increase, the probability of each cluster point falling within the correct kism/markaz increases. It is apparent from the data in figures 3.21 and 3.22 that the majority of DHS cluster points show higher probabilities of being in their correct spatial positions. These values ranges between 0.7 and 0.9 and a few clusters present a value of one which means they are certainly within their correct kism/markaz polygons. The most frequent values were between 0.60 and 0.80 while almost 50 clusters had probability values less than 0.50 indicating they are either lying on boundary lines or located very near to them.

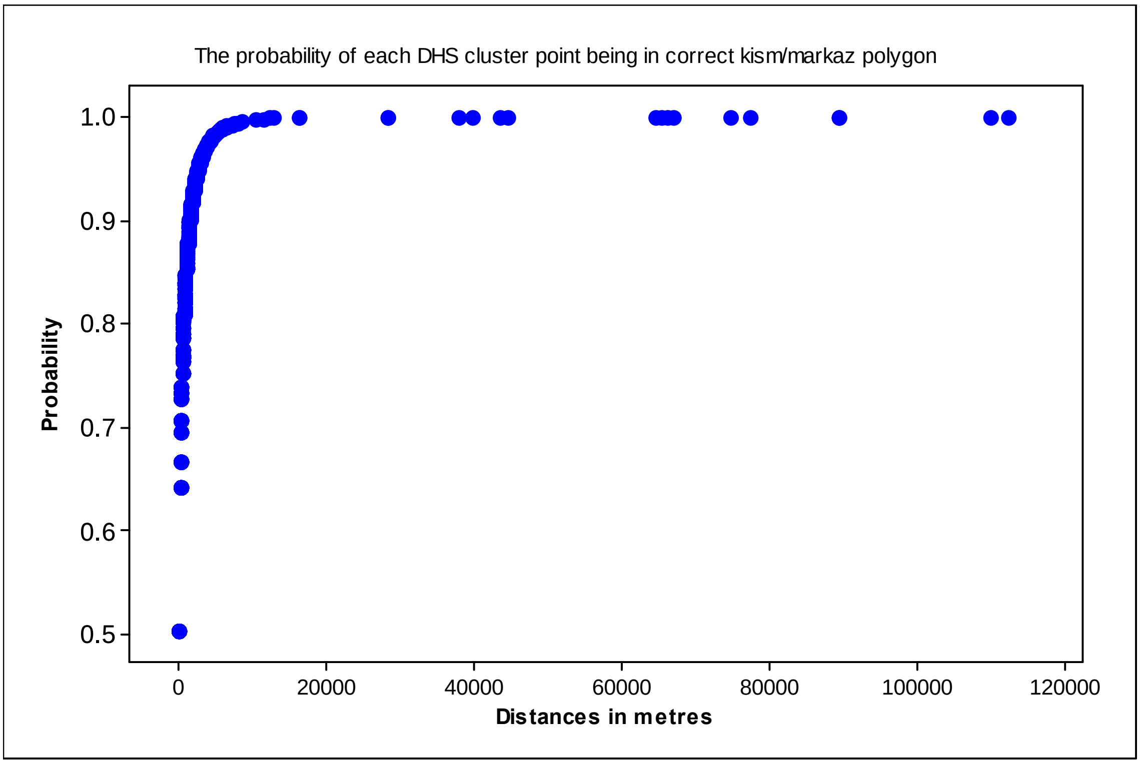


Figure 3.21 The probability of GPS clusters lying in the correct kism/markaz polygon according to the second logistic regression model

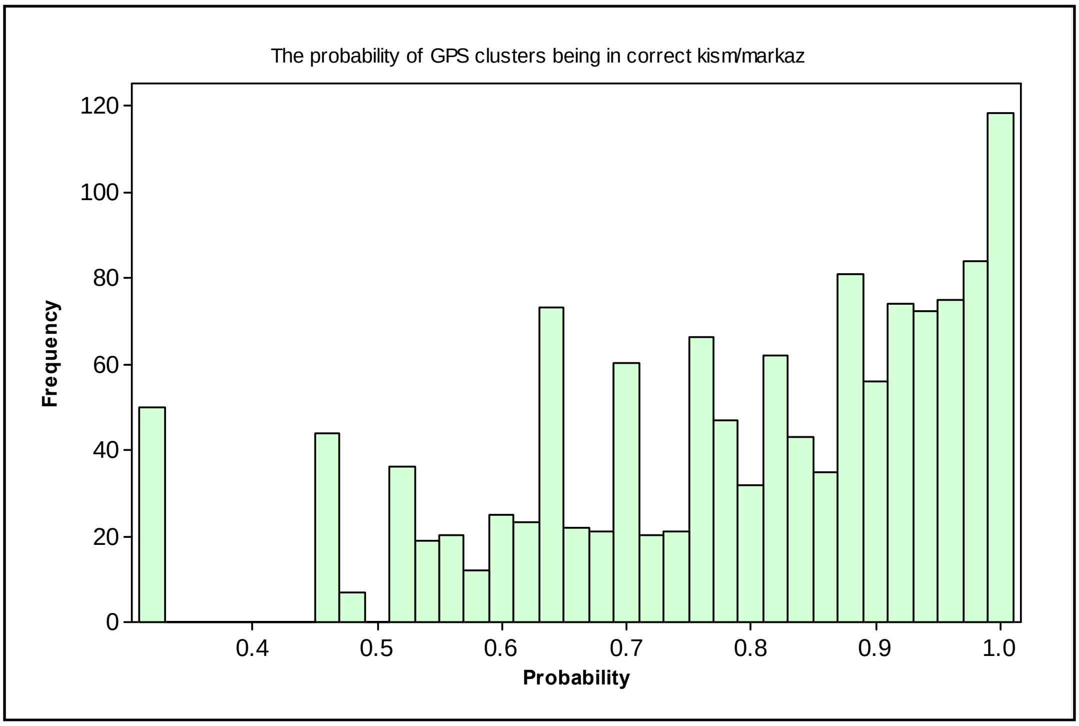


Figure 3.22 The frequency of cluster probabilities of being in the correct ksim/markaz according the second model

Mapping probability values

Visualising the calculated probabilities from the logistic regression equation is an important process to present the spatial differences that impact on chances of cluster points falling within incorrect positions due to their locations. The Idrisi software was used to create a map of the probability values based on the previous logistic regression equation using the following steps:

- **Multiply distance map by the coefficient**

Having been prepared for obtaining distances from sub-governorate boundaries, the distance map was used as an input layer in all steps of this process. The first step involved using the transform operation to obtain the square root of distances to match the regression equation. This was followed by the Scalar operation to multiply squared distance to kisms/markaz by the coefficient for distance from the regression (-**0.0586306***SQRD distance).

- **Adding the constant term from the regression analysis**

The constant value from the regression model (0.0481082) was added to the previous output using the Scalar operation (-0.0586306*SQRD distance**+0.0481082**).

- **Multiply the output by -1**

The output from the previous step was multiplied by -1, so this part of regression equation was applied (**-1***(a*SQRD distance +b).

- **Finding the natural antilog**

In this step, the transform operation was used to find the natural antilog for the output from the last step. The natural antilogarithm of any distance value is found by raising e (2.718) to the exponential of that distance (**e^** (-1*(a*SQRD distance +b).

- **Adding one to the resultant value**

The Scalar operation was used to add one value to the resultant value. So the equation parts that have been applied so far are (**1+**e^ (-1*(a*SQRD distance +b).

- **Obtaining the final probability**

In this final step the reciprocal option was selected from the transform options with the previous result as the input map layer, thereby completing the regression equation and producing the probability map **1/** (1+e^ (-1*(a*SQRD distance +b).

Figure 3.23 shows a map of the distribution of probability values in Egyptian sub-governorate boundaries. It can be seen that higher values of being in incorrect places are mainly concentrated in the Delta and Nile valley. Here the areas of kisms or markazs in all governorates of agricultural land are smaller than in the desert regions. However, the areas of kisms in Urban Governorates and Greater Cairo are smaller than both Upper Egypt and Lower Egypt. Therefore, DHS clusters located in these regions are more likely to be very close to boundary lines and, thus, are associated with high probabilities of spatial error. In contrast, DHS clusters located within Frontier Governorates tend to show the highest probabilities of being in their correct kisms or markazes. Consequently, all probability values of one are located in kisms or markazs in these governorates with large areas and low population density.

The third model

According to the allocation of the clusters and GPS coordinates in the DHS survey, it is expected that the geographic distribution of the GPS clusters will follow the spatial distribution of Egyptian population and settlements. As a result, the majority of GPS clusters are concentrated alongside the Nile River and its Delta while a few clusters represent desert communities in the Frontier Governorates in Western, Eastern, and Sinai Peninsula Deserts. Despite the fact that the Greater Cairo region shows the highest population density in Egypt, governorates which are located in the north of Upper Egypt in particular Bani Sueif, El Fayoum and El Menia show the highest concentration of clusters. Similarly, governorates in the Southern Delta such Kaliobya and Monfyaa show a higher concentration of GPS clusters compared with the governorates which are located in the north. This can be attributed to the location of these governorates around the Greater Cairo region and because they include the most populous rural areas. Moreover, these governorates are smaller in size than the others, but they have an equal or greater number of Primary Sampling Units (PSUs) in the survey.

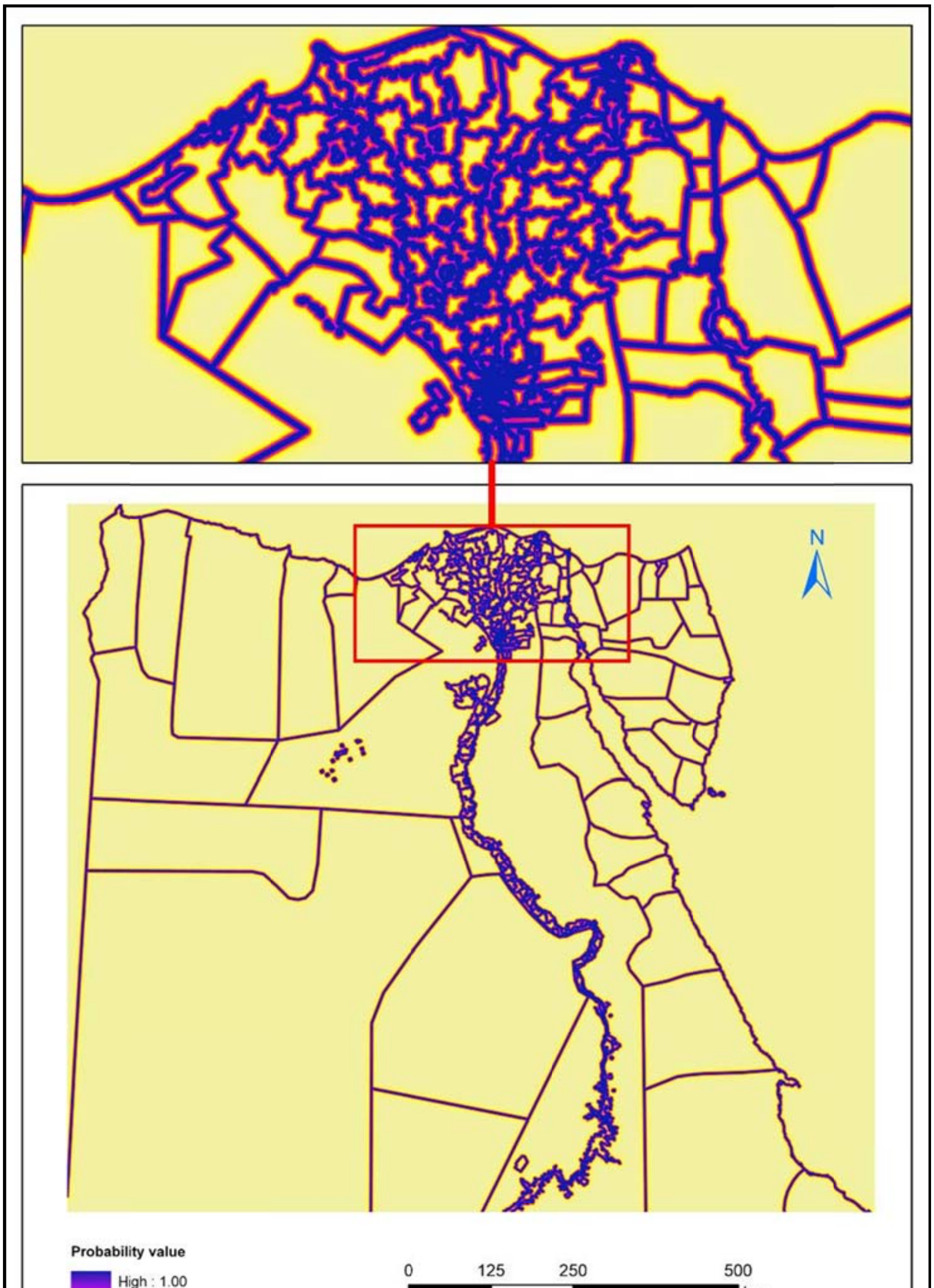


Figure 3.23 Map showing the probability of being in the incorrect kism/markaz

Although applying Linktest to the second model is not significant and a slight improvement in the result was found, fitting another model would be worthwhile since the Pseudo R^2 value is still low (0.287). Therefore, including a new independent variable is strongly recommended. In this case, increasing the capability of the predictive model requires adding new predictors. Thus, another variable “GPS-AVE” was created according to whether each cluster has been assigned as GPS or AVE in the main shapefile. I contacted MEASURE DHS to understand the differences between GPS and AVE as a survey method code for the GPS clusters. In the main shape file where there is a column titled “source” GPS means that just one GPS point was taken for that cluster while AVE means many GPS points were taken for one cluster. For instance, a cluster might be fairly large and consequently many points were surveyed and then averaged to obtain the approximate centre of that cluster (Burgert, 2009). Table 3.6 illustrates the number of GPS clusters in correct and incorrect locations broken down by the GPS or AVE codes in all Egyptian Governorates.

	GPS		AVE		
	Count	Percentage	Count	Percentage	Total
Correct	1097	95	138	95.8	1235
Incorrect	57	5	6	4.2	63
Total	1154	100	144	100	1298

Table 3.6 The percentage of GPS clusters by the survey methodology codes GPS or AVE

In order to protect the confidentiality of the participant households in the DHS survey, scrambling is a method undertaken by MEASURE DHS to anonymise the GPS cluster codes. Although scrambling is undertaken initially for the HIV test survey, it is applied to the DHS survey as a precaution for individual records of households. Therefore, a geographic error in any direction is randomly added to the cluster locations of up to 2 km for urban areas and 5 km for rural areas. As a result it is impossible to associate an individual record with a particular geographic location. Based on the assumption that the location of GPS clusters being in rural or urban areas might impact on the distribution of the identified spatial errors, all the GPS cluster were classified into two groups urban and rural according to the column included in the GPS shapefile entitled “ARUBAN-RURAL”. Therefore, a new variable was constructed to be included in the new model. This was followed by calculating the percentage of GIS clusters lying in correct and incorrect positions in both groups in all Egyptian governorates. Table 3.7

illustrates the distribution of the identified spatial errors of GPS clusters when they were linked to administrative boundaries. It can be seen that 716 out of 764 GPS clusters (93.7%) that are located in rural areas in the correct positions. By contrast 15 GPS clusters out of 63 with incorrect positions appeared in urban areas.

	Urban		Rural		
	Count	Percentage	Count	Percentage	Total
Correct	519	97.2	716	93.7	1235
Incorrect	15	2.8	48	6.3	63
Total	534	100	764	100	1298

Table 3.7 GPS cluster error in urban or rural areas

The third regression model was implemented by creating a new dichotomous variable, urban or rural location, and adding it to the previous model as an independent variable. The GPS clusters were categorised into two groups, urban and rural, and each cluster located in an urban area was coded (1) while those in rural area were coded (0). The model is made up of four variables: a binary response variable, two dichotomous variables, urban-rural and GPS-AVE, and the distance variable to nearest governorate boundary. The main purpose of this model is to increase the strength of prediction. The GPS_AVE variable was not statistically significant (P value = 0.271) and was dropped.

The findings of the third model clearly show a small improvement in the explanation of the observed data through this regression model compared with the previous model; this is in response to the new variable that was added. Table 3.8 shows the outputs of this model. The interpretation of the model as a whole is statistically significant since the value of P> Chi^2 is 0.000 which means the null hypothesis must be rejected and the independent variables together have impacted the dependent variable. The values of P>Z suggest that all variables included in this model were statistically significant since they have P-values less than alpha (0.05).Therefore, the coefficient values of the independent variables are valid to estimate the amount of increase in the predicted log odds of spatial error (the dependent variable). The value of the Pseudo R^2 (0.36) indicates an improvement in the strength of the relationship between the variables and it expresses a better fit of the probability of spatial error of GPS clusters compared with the previous model (0.29).

Variables	*Coef.*	*Std.Err*	*Z value*	*95% conf.*	P>|z|
√(Governorate distance)	-0.064	.008	-8.61	-.078	0.000
Urban	**-1.825**	**.331**	**-5.51**	**-2.48**	**0.000**
Constant	**1.026**	**.330**	**3.10**	**.377**	**0.002**

Log likelihood = -161.60 Prob > Chi^2 = 0.000

Pseudo R^2 = 0.36 LR chi2 (1) = 180.89

Urban: yes = 1 No = 0 Linktest **hat** (P>|z|) 0.595

Linktest **hatsq** (P>|z|) 0.005

Table 3.8 The coefficients of the third regression model

The fourth model

In spite of the new predictor added in the previous model being meaningful, the finding from applying Linktest illustrates that **hatsq** is significant (0.005). This means that there may be a specification error in the third model and refitting a new one is required. An initial step has already been taken towards improving the logistic regression model by identifying the variables that best explain the observed pattern of errors. Hence, all the relevant independent variables, predictors, appear to be included in the model. Interaction between the two independent variables previously introduced can be a potential source of error in the model. Therefore, a new interaction variable (urban-dist) has been created and used instead of the dummy urban variable in the previous model. This was created by multiplying the value of the urban variable (1, 0) by the squared distances variable (the distance for each GPS cluster to the nearest boundary). It seems intuitive that this model is the best one and its summary statistics suggest a stronger relationship between the response variable and the explanatory variables. Table 3.9 shows the findings of applying the new logistic regression model. The value of P> chi^2 is 0.000 and comparison of the four models shows that this model has the best Pseudo R^2 value (0.39). The model shows better predictive abilities than the others. Therefore, the coefficients of the independent variables are more useful in estimating the probability of the spatial error of GPS clusters and, thus, they were used in the regression equation to predict the probability of spatial error based on distance of GPS clusters from kism/markaz administrative boundaries.

Variables	Coef.	Std.Err	Z value	95% conf.	P> Z
√(Governorate distance)	-0.056	-8.37	-8.37	-0.069	0.000
Urban*squared distances	-0.078	0.015	-5.08	-0.108	0.000
Constant	0.740	0.299	2.48	0.154	0.013

Log likelihood = -154.438 Prob > chi2 = 0.000

Pseudo R2 = 0.39 LR chi2 (1) = 195.22

Linktest **hat** (P>|z|) 0.157 linktest **hatsq** (P>|z|) 0.050

Table 3.9 The coefficients of the fourth regression model

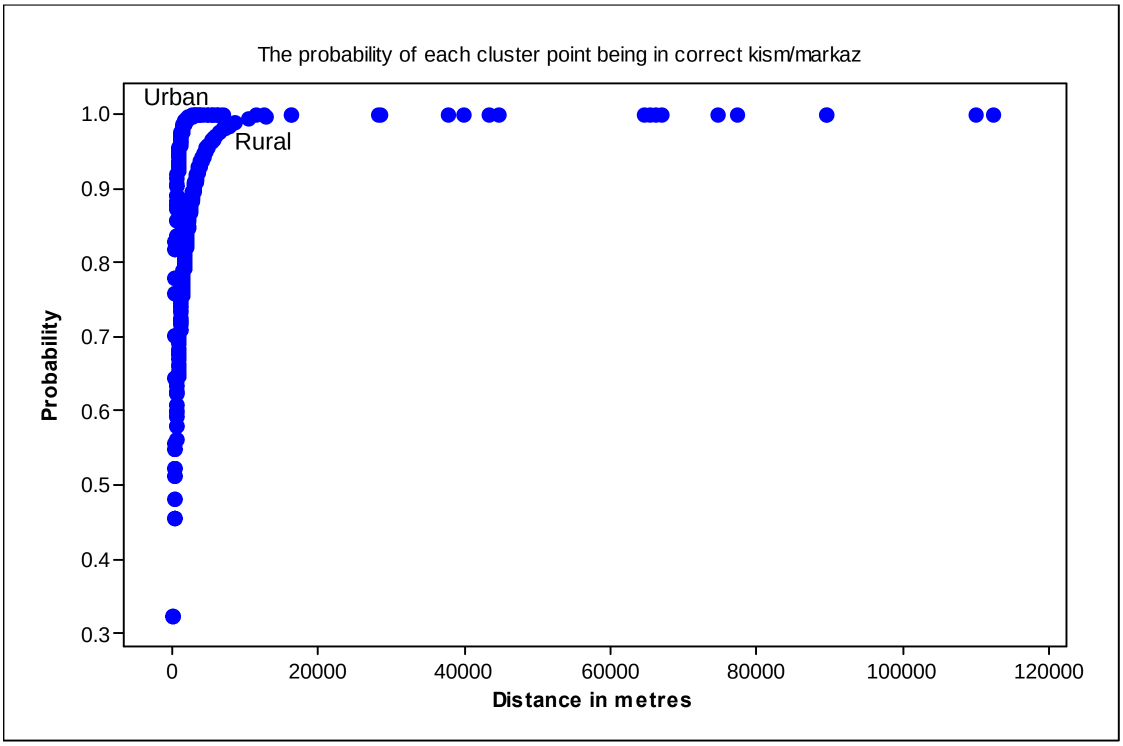


Figure 3.24 The probability of being in correct kism/markaz predicted by the fourth regression model

The following equation was applied to kism/markaz distances within an Excel spreadsheet to produce the probabilities of spatial error of GPS clusters:

P =1/ (1+2.718^ (-1(-0.0558526*SQRT (K/M distances)-0.0778706*SQRT (Urban*distances) +0.7402697)))*

Likewise *q= 1-P* was applied to produce the probabilities of GPS clusters being in their correct positions. The result of the previous logistic regression equation applied to kism/markaz distances is presented in figure 3.24 which fits the relationship between distances and the probability of GPS clusters being in correct positions. It can be seen that the pattern of predictions of this model is different from the previous ones as there are two curve lines plotted instead of one. This originally comes from the interaction variable that was included in the model, produced by multiplying the distance squared by the dummy variable urban (1)-rural (0).Therefore, the line to the right side represents clusters located in rural areas that have distances of zero while clusters present in the left side close to the Y-axis are located in urban areas with the actual value of squared distance (as it was multiplied by one, and, thus, the curve was pushed to the left side). The graph follows the hypothesis that scrambling the GPS clusters in rural areas is greater than urban (MEASURE DHS, 2008f) and therefore the two curves indicate the difference in distance, about three kilometres, that has been added as a spatial error between clusters located in rural and urban areas.

It is apparent that a few clusters show low probabilities of being within the correct boundary, 0.32 to 0.45, and these either lie on boundary lines and have a distance of zero or they are very near boundary lines with distances less than 200m. On the other hand, the majority of clusters are concentrated around higher values, greater than 0.5, of the probability of being within the correct administrative boundary. Thus, there is a low likelihood that the spatial error of the GPS cluster occurs even if they are located within a small distance (500m) of the kism/markaz administrative boundaries). Furthermore, the curve of the model shows that the clusters which are located near kism/markaz boundaries provide high values, 0.7 to 0.9, of probability of being in the correct administrative unit. The model predicts that a GPS cluster located 2060 metres and above away from an administrative boundary will definitely be in the correct kism/markaz polygon. Figure 3.25 illustrates the fourth model's histogram. It is clearly seen that the most frequent probability value were between 0.99 and 1.00 (118 clusters) and probability values between 0.87 and 0.99 were quite frequent (165 clusters). However clusters with low probability values of being in the correct administrative unit, less than 0.5, were less frequent (101). It is concluded that the new variable, urban squared distance, has improved the predictive ability of the model and it is more useful than the three previous models.

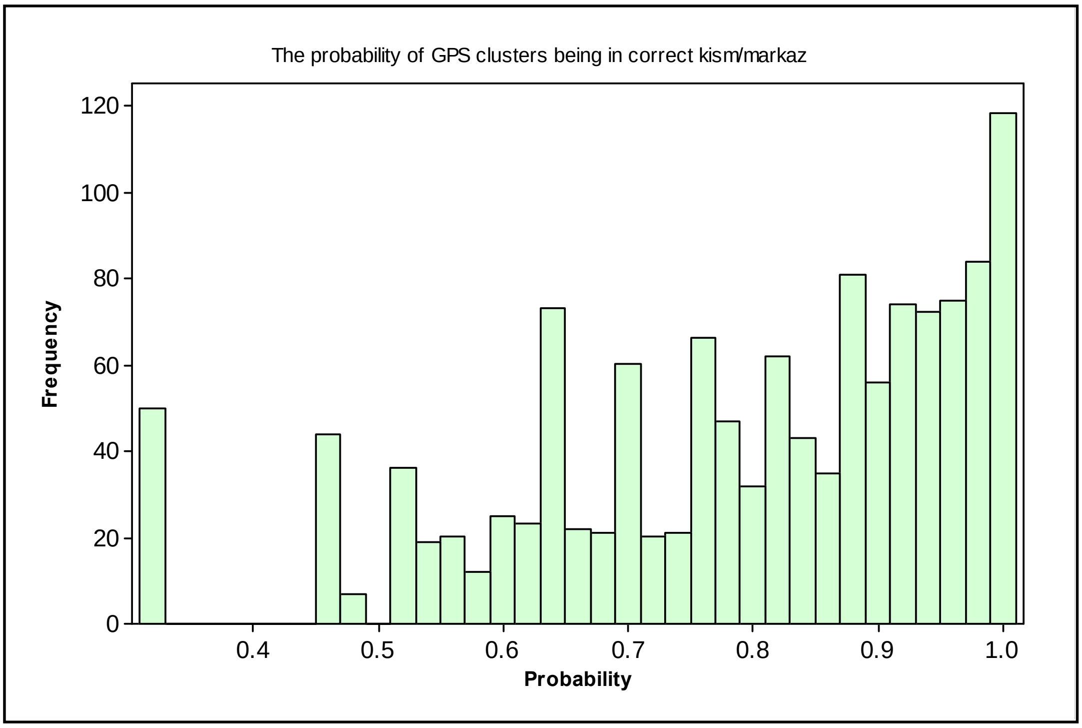


Figure 3.25 The frequency of cluster probabilities of being in the correct kism/markaz according to the fourth regression model

Residuals and Moran's I test

Residuals in regression model are known as the portion of errors for an individual measurement or the differences between the observed values and the predicted values of the dependent variable (Hosmer et al. 1991). Computing the residuals in the regression model is an important step in evaluating the assumptions of the fitted model (Cook and Weisberg 1982). The residuals of the model were computed using the Stata commands: Postestimation and standardised Pearson residuals. Moran's I is a common way of measuring and assessing the absence or the presence of spatial autocorrelation in areal datasets (Rogerson 2006). The output of the Moran's I test includes the Moran's I index value which measures two kinds of correlation, either positive or negative. A positive value close to one means clustering of high or low values while close to minus one means no clustering is found and the spatial pattern is dispersed. In addition, the Z score indicates the significance of the index value (Haining, 2006; Zhand and Lin 2006). The global spatial autocorrelation metric, the Moran's I method was applied to the residuals of the regression model. The test statistic shows that there is spatial autocorrelation between the residual values and clustering does exist, but with a low index value (Moran I index = 0.01, Z score = 2.41). However, it was greater than zero and therefore, the null hypothesis should be rejected and although it is a weak pattern, it could not have occurred by chance.

3.4 Analysing spatial errors when linking GPS clusters with shyakha geography

Egyptian census geography consists of three levels; firstly, the governorate level which is used by MEASURE DHS administratively as the names of the DHS samples and cluster codes. According to the overlay process which was applied earlier using this census geography level and the DHS clusters, the spatial error was identified, measured, analysed and modeled. Secondly, the kism/markaz level is comprised of sub-administrative boundaries which are lower than governorate and higher than shyakha. The parameters produced from the modeling process were applied to this geography level to predict the probability of spatial error for each GPS cluster being in its correct kism/markaz. Finally, the shyakha level is the smallest census geography and it is important to examine the magnitude of spatial error expected to be found when the GPS clusters layer of the DHS is linked to this level. To prepare spatial linkage at the shyakha level, the following steps were undertaken:-

3.4.1 Creating a list of shyakhas sampled in the DHS

To create a new list of sampled shyakhas for the DHS, the shyakhas/towns that were used by MEASURE DHS in the 2005 survey (El-Zanaty & Way, 2006) were manipulated and a database table was created to match the Egyptian Census shyakha layer to be ready for geocoding.

3.4.2 Geocoding of sampled shyakhas names

Before undertaking the geocoding process it was essential to modify the attribute table of the Egyptian Census shyakha as it contains 28 governorates. It includes the two new governorate codes, Helwan and October, which do not match the DHS shykha table that contains only 26 governorates. Therefore, a new column called "Old governorates" was created and added to the census shyakha attribute table.

1) The standard geocoding function in ArcGIS 9.3 was used to create an address locator by selecting "world cities with country" as an address locator style and locating the modified census shyakha layer as a "reference data" used by the address locator. shyakha name was determined to be "City" in the input field name while Old governorate was the "country".
2) Starting geocoding addresses command by selecting the created address locator and choosing the DHS sampled shyakha table as an address table. In the input field "shyakha" was chosen as "City" and "Governorate" as "Country". Geocoding

options including spelling sensitivity, minimum candidate score and minimum match score were left at default values in the first step.

3) This process produced a high proportion of unmatched shyakha names (457) while 188 shyakhas were matched and one shyakha was “Tied “. Thus, a rematching process was needed to improve and increase the number of matched shyakhas.
4) The geocoding addresses command was rerun, but using low spelling sensitivity, minimum candidate and minimum match parameters since shyakha names had been transcribed from Arabic into English. The output of this process produced a high percentage of matched shyakhas (58%) while the percentage of unmatched shyakhas decreased to 42%.
5) Using the “interactive rematch” facility the number of matched shyakhas was increased to be more than 500 out of 647 shyakhas.
6) Because of local name variations and issues in transcribing Arabic shyakha names in English letters, the remaining unmatched shyakhas had to be manipulated and rematched manually. By rematching all unmatched candidate shyakhas, a new layer was created for the DHS sampled shyakhas/towns containing 636 polygons. However, a few shyakhas listed as sampled in the DHS documentation were missed because either their names do not exist in the census layer or their names probably have been changed.

3.4.3 Overlaying the layer of shyakhas sampled for the DHS with the GPS cluster layer

Looking at the distribution of the GPS cluster points overlaid with the DHS sampled shyakha, it can be seen that although the household clusters are concentrated around the sampled shyakhas, a spatial error is widely apparent across the whole of the census boundaries (figures 3.26, 3.27 and 3.28). Thus, a proximity analysis using the near function was undertaken to determine distance from each GPS cluster to the nearest shyakha boundary.

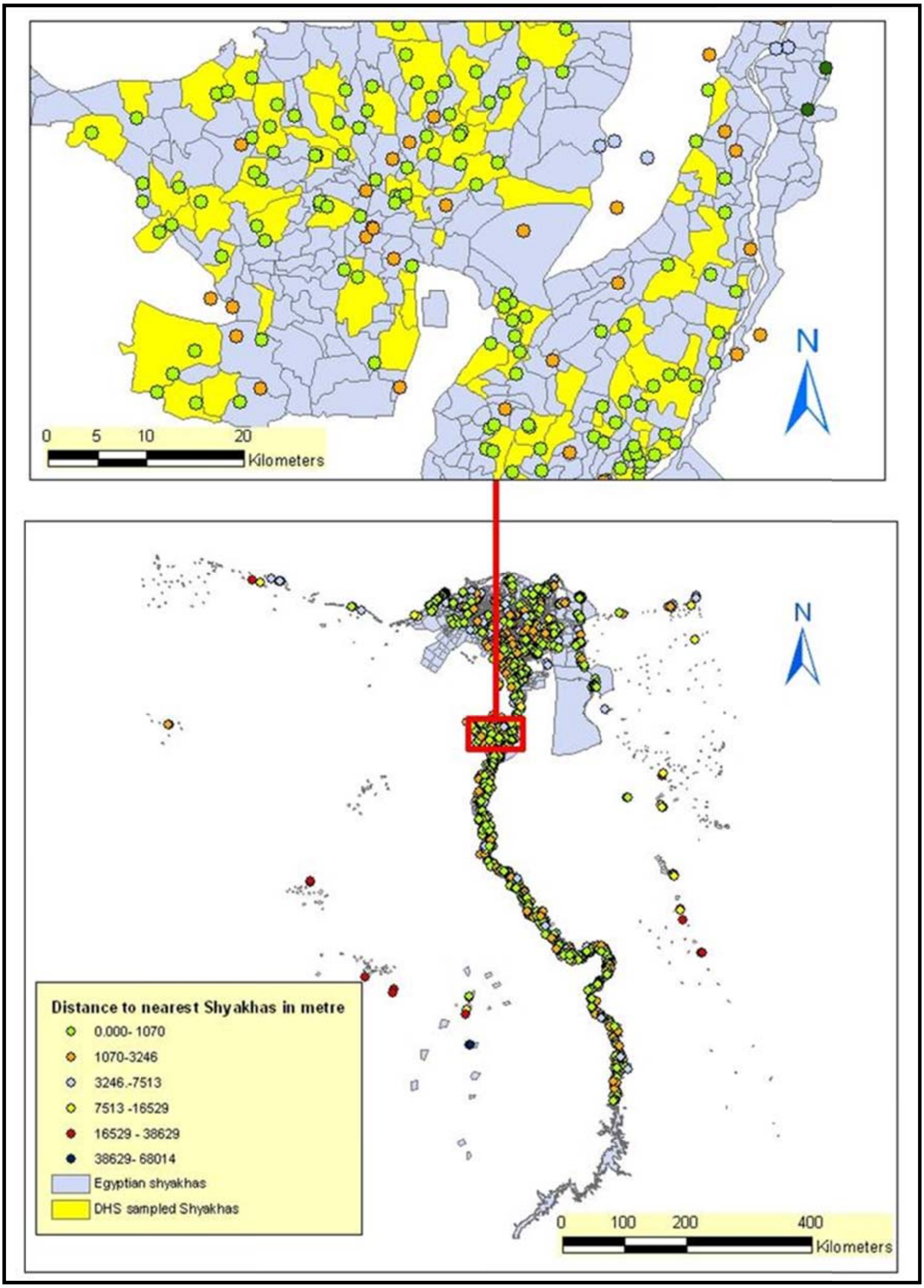


Figure 3.26 Geographical distribution of GPS clusters across Egyptian shyakha geographies

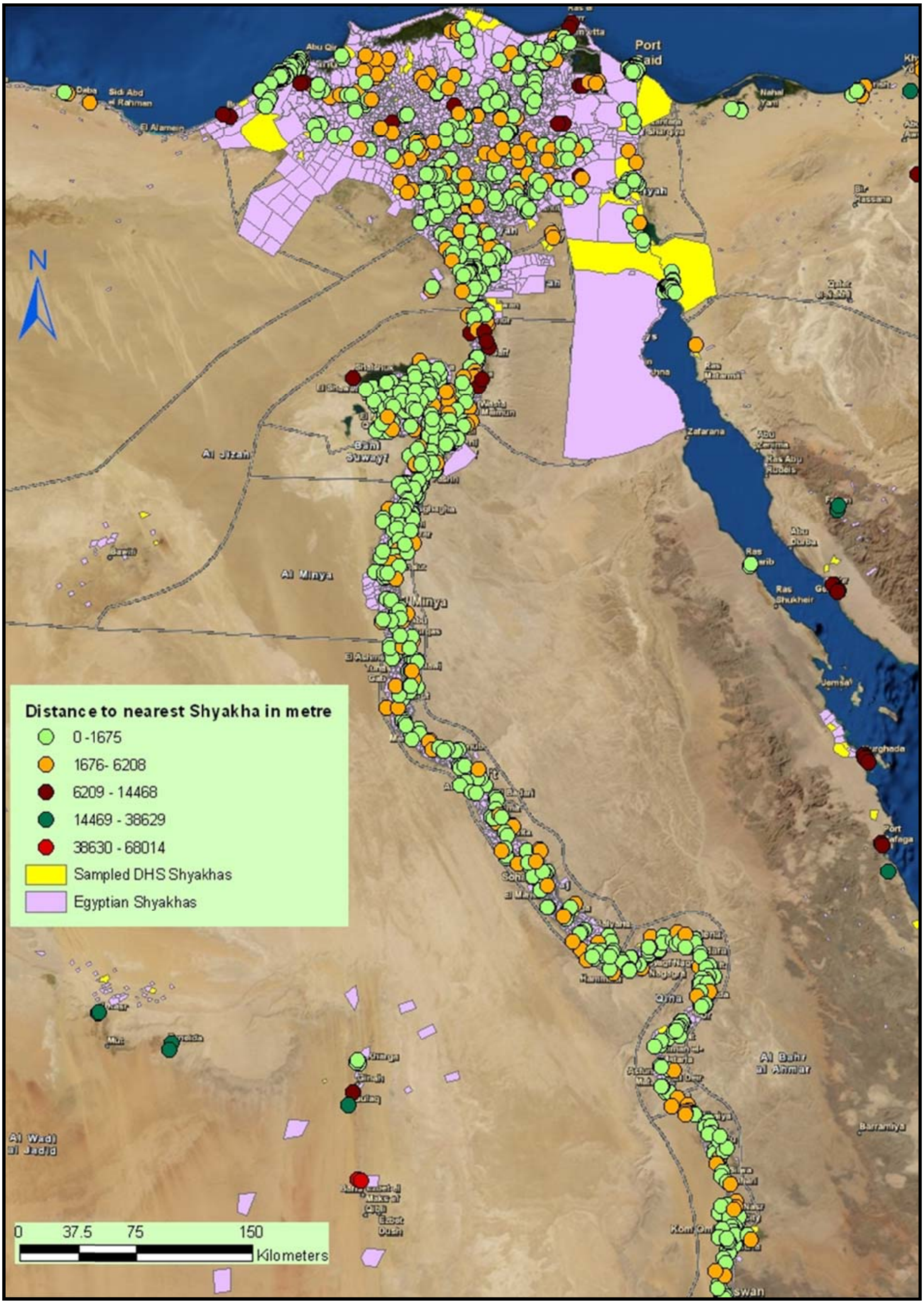


Figure 3.27 Geographical distribution of GPS clusters across Egyptian shyakha geographies using the ESRI ArcGIS online services "world Imagery function"

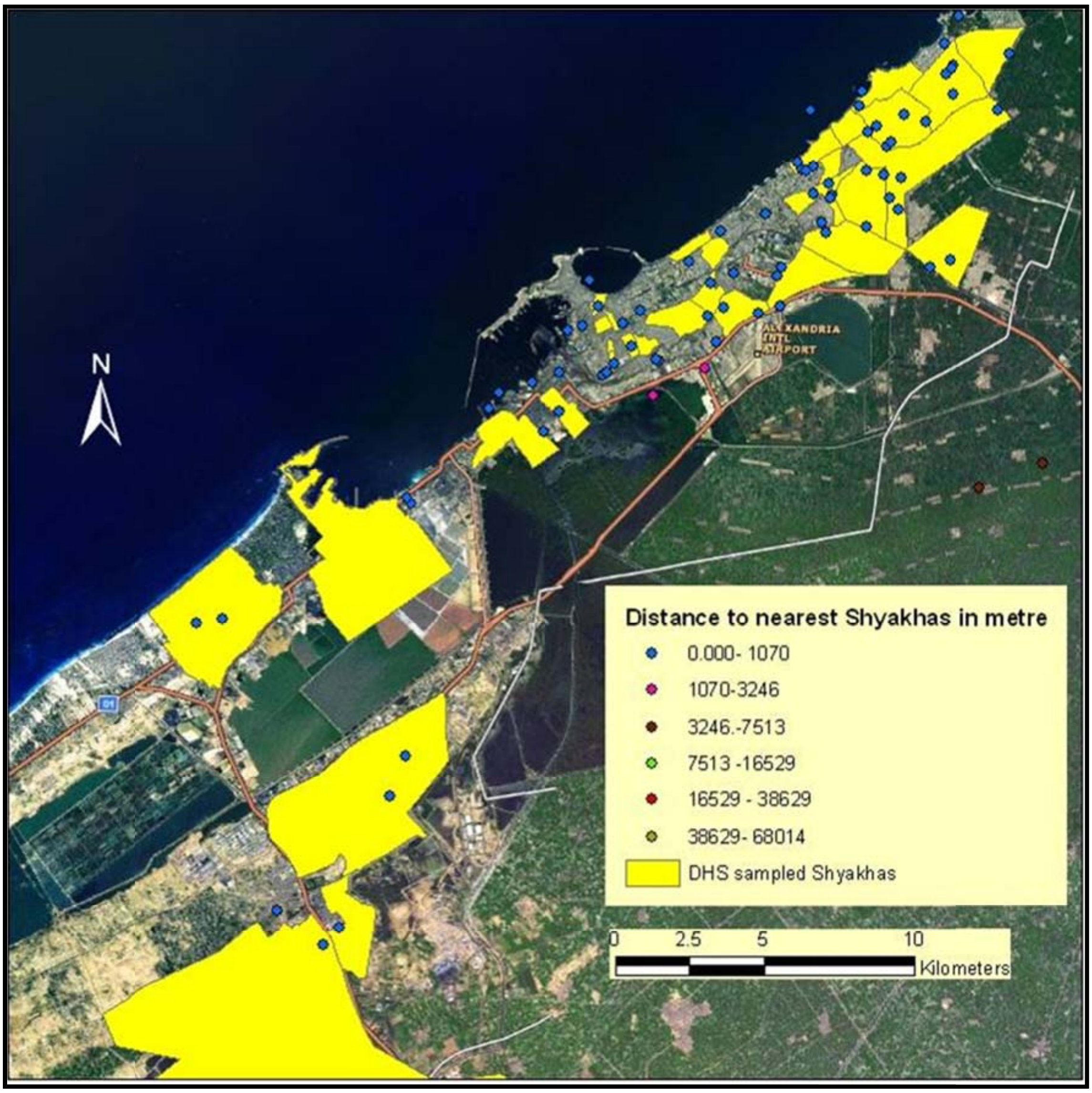


Figure 3.28 Geographical distribution of GPS clusters across Alexandria shyakh boundaries using the ESRI ArcGIS online services "world imagery function"

Figure 3.29 illustrates the output of the proximity analysis and the near function. The histogram shows that there is a large number of GPS clusters which fall outside their sampled shyakha boundaries. The majority of clusters are located far away from nearest boundaries by one to one and one and half km (almost 160 clusters). Similarly, roughly the same number of clusters fall within a distance equal to or less than one km. Most of these clusters are distributed along the Lower and Upper Egypt Governorates. The third category of misallocated clusters is located more than one and half and less than 2 km away from the nearest sampled shyakha boundaries. The remaining groups of GPS clusters are distributed at distance between 1 and 5 km. On the other hand, a number of clusters fall a long way outside the sampled shyakha boundaries, more than six km, in particular in the desert governorates. Moreover, there is a GPS cluster group reflected in some outlier values, with distance more than seven kilometers, and this is located in the desert governorates especially New Valley. Using the ESRI ArcGIS online services "world imagery" resource, figure 3.30 illustrates how the El Kasr shyakha census polygon is located further away from the settlement pattern on the imagery, while the GPS clusters seem to be in their correct location. This suggests a positional error in the shyakha polygons where some of these, particularly in desert governorates, do not match their true locations.

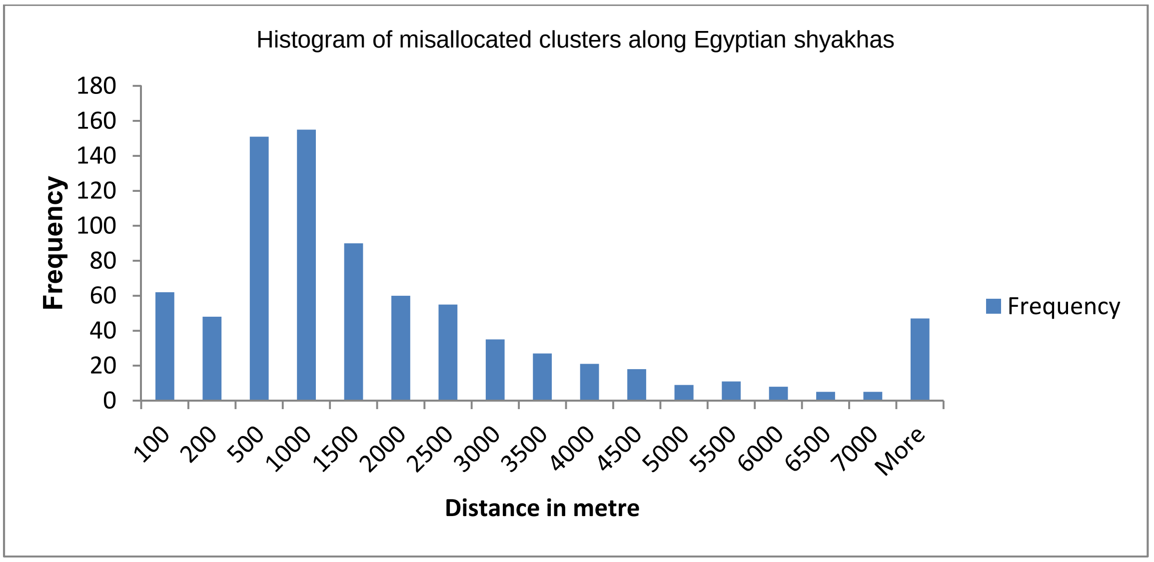


Figure 3.29 showing the distribution of distances of misallocated GPS clusters from shyakhas sampled for the DHS

In conclusion, the shyakha linkage process did not produce reliable matching between the sampled DHS shyakhas and the GPS clusters and there was a considerably higher percentage of the GPS clusters falling outside their correct shyakhas, compared with the earlier governorate level analysis. As a result, the misallocated clusters are either located outside all of the available administrative boundaries or located in the wrong Shyakha polygons. Thus, this suggests there is much uncertainty in assigning these clusters to shyakhas. Since governorates are too large to show much geographic variation, the kism/markaz census geography thus appears to be the best geographic level for undertaking spatial linkage and establishing a standard comparable method to be applied for linkage between an international dataset, the DHS, and a national census dataset.

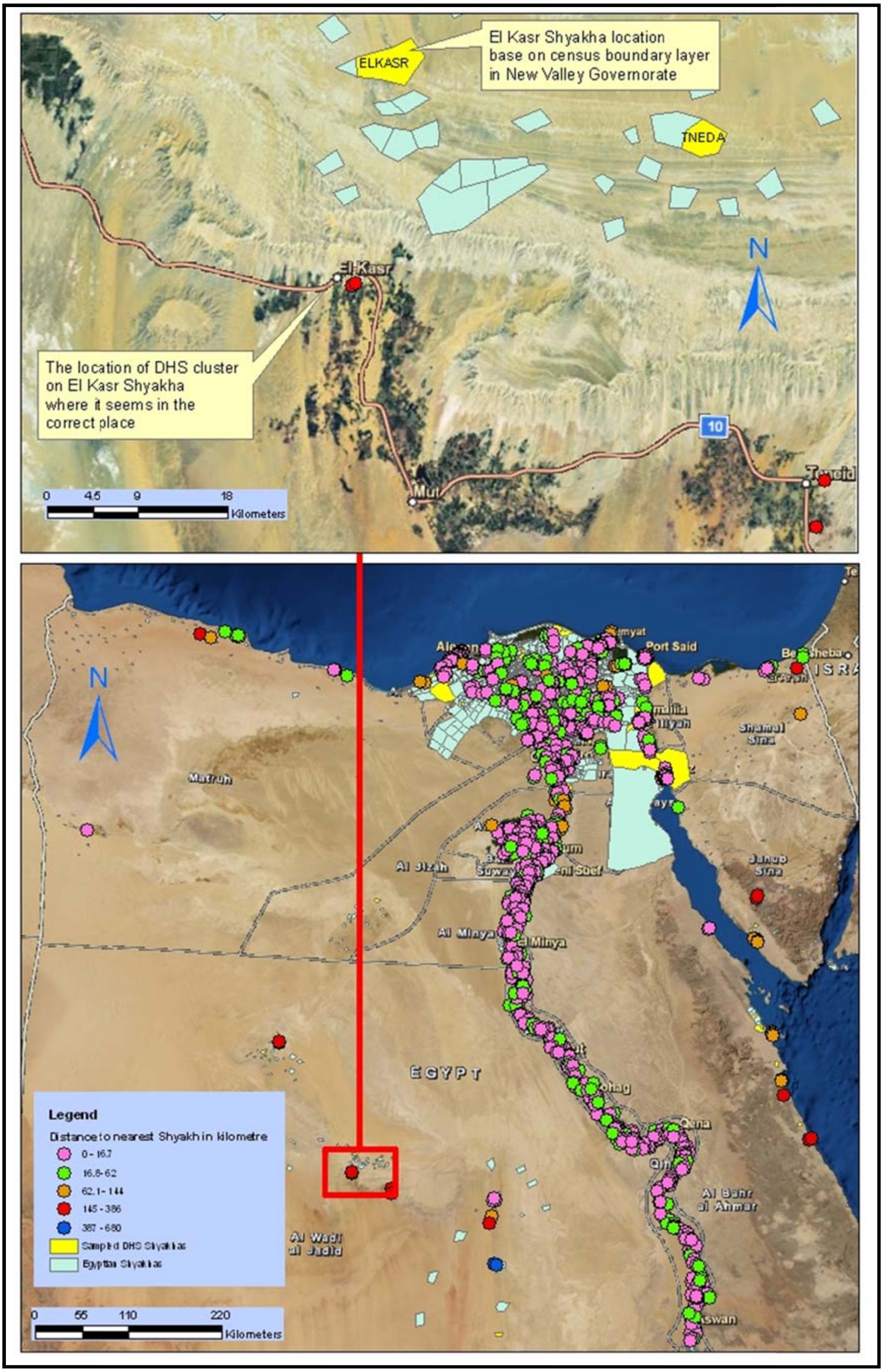


Figure 3.30 Comparison of positional accuracy between census polygons and GPS clusters using the ERSI ARCGIS online services "world imagery function"

3.5 Discussion and Conclusions

These results suggest it may be problematic to link data on DHS GPS clusters to census small areas (such as shyakhas). Similarly, there is little to be gained from undertaking spatial linkage at the governorate level, since governorate labels are already provided for each GPS cluster for this purpose. Consequently, given the spatial uncertainties apparent in the data set, the evidence here suggests that the district (kism or markaz) level is the most appropriate for conducting spatial linkage. Examination of both the DHS GPS point and census map layers shows that they have the same projection and errors are widely distributed across the study area. Therefore, there is no evidence in this analysis to suggest a systematic error in either of the map layers analysed, such as a discrepancy in map datums used or some region-specific mislabeling of governorates. Since the majority of misallocated points are concentrated around governorate boundaries, it seems likely that the observed inconsistencies arise because of a positional error rather than mislabeling of GPS cluster points.

In addition to the spatial error that was found at governorate level, there are some additional sources of error that will have influenced the shyakha level analysis:

- It was not possible to geocode some shyakhas sampled through the Egyptian DHS, whilst other shyakhas may have been incorrectly geocoded because of local place name variations and issues in translating Arabic shyakha names into English. This may explain why some GPS cluster points are apparently located far from sampled shyakhas, as shown in the histogram in Figure 3.29
- At the shyakha level, a GPS cluster was only identified as erroneous and a misallocated point if it lay in an unsampled shyakha. However, this process does not identify all GPS clusters that lie outside their equivalent shyakha boundary, since some GPS clusters may be displaced to a neighboring sampled shyakha. This will result in an under-estimate of the extent of spatial linkage problems at the shyakha level.

Thus, care should be taken in interpreting the error rate in linking GPS cluster points to shyakhas relative to those for kism/markaz and governorate level.

The addition of GPS clusters to the DHS offers increased analytical potential in GIS and global health research. However, this analysis suggests administrative labels and spatial locations of GPS clusters cannot be used uncritically and error may affect spatial analysis of DHS datasets. This is most likely due to deliberate modification of cluster locations noted earlier, intended to protect the confidentiality of participant households in the DHS. A strong recommendation to the MEASURE DHS organization would be to adapt the data modification algorithm to avoid gross misallocation between administrative areas. In combining DHS data from GPS clusters with census data for kizms/markazes or other administrative units, any linkage process will need to take account of the likely impact of positional uncertainty. At the shyakha scale there would be unacceptable levels of uncertainty in matching shyakha boundaries to GPS clusters. Thus, the kism/markaz census geography appears to be the best geographic level for undertaking spatial linkage.

In summary, there is a growing literature that enriches DHS data by using GPS cluster locations to link to other spatial data sets, including census data. However, none of these studies has examined the effect of locational uncertainty on these linkage operations. I propose that the issues raised here are of broad significance when attempting GIS use for analytical data linkage in developing country settings, particularly in the increasingly common situation where hand-held GPS are used to georeference social survey data which must then be linked with existing national sources. Furthermore, in using attribute labels inconsistent with location to model spatial error, it adds to the existing literature on uncertainty and the point-in-polygon problem.

3.6 Summary

Throughout this investigation, the aim was to assess, model and understand the spatial error that was identified by linking the DHS clusters to Egyptian boundaries. The finding of applying cluster tests was significant, with the interpretation that GPS clusters with spatial errors were clustered together. Exploring the geographic pattern of the distribution of the GPS clusters with spatial errors, most of them are located near to the governorate boundaries. This indicates a relationship between the distance from the location of the GPS cluster and the probability of being in their incorrect governorates. Logistic regression was an appropriate technique to model this error where the dependent variable was dichotomous and the distance to boundaries was an independent variable. Despite the first regression model being significant, the Linktest method indicated a specific error in the linear combination and thus a square root transformation was applied to the distances to create a new independent variable to be included in the second model. However, model performance was further enhanced by adding new variables. Considering the DHS procedures and scrambling method that are applied to the GPS clusters, GPS-AVE and urban-rural variables were added as independent variables in the model. The result of the third model has shown that GPS-AVE was not significant and thus it was removed.

Applying the Linktest function to the model to examine the linear combination of the predictors, the results suggested refitting a new model would be valuable since a specification error could be found in the linear relationship and a relevant variable might be included to the model. In the fourth model, an interaction variable, the squared root of distance multiplied by urban-rural, was added instead of just urban-rural and therefore, the linktest was not significant and Pseudo R^2 value was greater than in the previous models. This means that it is the best model in terms of its structure and as a predictor of the probability of spatial error. The most obvious finding to be drawn is that a clustering pattern of spatial error of GPS clusters appeared to be related to the nature of the spatial data. Nonetheless, this clustering pattern is not very strong. Moving forward and applying the coefficients of the best regression model to the lower level boundaries, kism/markaz, the results clearly show that there is a reasonable chance of obtaining about one third of the GPS clusters falling in their incorrect geographic locations. Furthermore, the distance of the DHS clusters from administrative boundaries and rural or urban locations both contribute to understanding of the spatial error when linking DHS clusters to census boundaries. Using the list of shyakhas published by MEASURE DHS, positional uncertainty was analysed and evaluated. However, undertaking geocoding of sampled shyakha names was problematic because of issues in translating Arabic shyakha names into English and thus matching sampled shyakhas to DHS GPS clusters would produce unacceptable level of uncertainties.

Chapter 4: Composite indicators for Safe Drinking Water Index (SDWI) – a review of the conceptual framework and background

4.1 Overview

The principal aim of this chapter is to review the literature on area classification systems and index construction. This is in order to explore the concepts behind two methods that have been widely used in investigating the socio-spatial structure of small geographic areas in developed and developing countries. In the broadest sense, reviewing the most relevant and common water and non-water (especially deprivation) indices is necessary to obtain deeper insights regarding the problem of household water accessibility and its impacts on public health in developing countries. This section begins by discussing concepts of area classification and index construction and their applications in developed and developing countries. In addition, it considers lessons learnt from previous water applications of these methods, particularly highlighting previous methods used in index construction. It identifies some limitations in existing research on measuring drinking water access and identifies a gap which needs to be investigated. Furthermore, the review is used to identify the most appropriate method for developing a spatial measurement of safe drinking water access in Egypt. Selection criteria for the inputs to a new safe drinking water index will be explained. Finally, a comparison between census and survey is presented to highlight the differences between the two major data sources used in this research.

4.2 Concept of area classification and indices

The main reason for reviewing area classification and index methods here is to assess their suitability for analysing the interrelationships between drinking water and population health in Egyptian communities. For this thesis, a measure of access to safe drinking water in developing countries is needed. Developing an area classification system or index of safe drinking water access is crucial for many reasons. Firstly, there is a diversity of georeferenced datasets about drinking water which are collected nationally and internationally (the DHS and local censuses) which could be used to describe spatial patterns of access to safe drinking water across the entire country at the subprovincial scale. Secondly, there are likely to be many benefits from generating such a spatial measurement, particularly integrating different datasets about drinking water into a single summary measure. Thirdly, this tool would allow direct comparison of progress in improving safe drinking water access, both subnationally and internationally among many developing countries. Finally, spatial measurement construction has been initiated because of the desire to create an integrated and comprehensive tool to help policy planners and decision makers to track drinking water problems and their health impacts on households. Considering this interest, the question is how to develop an integrated spatial measurement of safe drinking water accessibility in the study area at a detailed scale. Either area classification or an index could be an effective tool to achieve the desired objectives relating to problems of drinking water supplies in developing countries. Geodemographics and multivariate indices play an important role as targeting tools for small communities, capturing the spatial homogeneity or heterogeneity of geographic units. The application of the two methods depends on combining appropriate socioeconomic variables and components for different areas, derived either from census or noncensus databases.

4.2.1 Overview of area classification

During the last three decades, area classification and geodemographics in particular have been applied in a variety of successful applications (Singleton & Longley 2009). Parker et al. (2007) argued that the development of place classifications is strongly related to the progress achieved in geodemographic technologies, and therefore, this method is fundamentally related to class analysis undertaken by new geographic software. Following a review of previous work on geodemographic classification for demographic phenomena generally and water supply issues particularly, a decision could be made whether a classification system would be valuable in developing a new measurement for safe drinking water and health in Egypt. The main purpose in undertaking an area classification is to assign every geographic area into one of a number of classes according to its socioeconomic and demographic characteristics (Beaumont, 1991; Mitchell, 1997). There are a variety of techniques that are used to build area classifications with applications in private companies and the public sector. However, there are specific processes that are commonly undertaken in every classification system. Harris et al. (2005) pointed to the most frequently used approaches in building area classifications. Firstly, data are gathered from different sources and data reduction is applied to these data. Secondly, the most appropriate variables are selected, which should be reliable, robust and not repeated. Thirdly, input variables are processed, including evaluation and standardisation. Fourthly, the appropriate weighting method is determined and applied. Finally, cluster analysis techniques are applied to produce classes.

4.2.2 Overview of indices and their construction

The second candidate method is index construction, which consists of comprehensive domains, indicators, components and subcomponents representing demographic and socioeconomic characteristics derived from both census and noncensus datasets. The main objective here is to score areas according to access to safe drinking water and consequently capture the spatial variations of water supply patterns for Egyptian households at a sub-provincial spatial scale.

- An index is a statistical tool comprising different indicators and components. It can be used to measure quantity and changes in the represented phenomenon and it is target as scores in which index factors can be ranked. Each index may consist of domains, indicators, components and, in some cases, subcomponents (see figure 4.1).
- The index domain can be defined as a main area or field such as health or education that provides components from which the index can be built.
- In many indices, domains effectively comprise only a single component and so the terms indicator, domain and component are synonymous. However, an index component is considered as part of a larger domain or indicator such as tap or pumped water as a part of water supply while water supply is an indicator for the water quality domain.
- A subcomponent of an index is a lower level of data as part of a component, for instance boiling water is a subcomponent of home water treatment (component) which is a part of larger indicator (water quality). Figure 4.1 illustrates an index domain and its relationships with other indicators, components and subcomponents. Matern et al. (2009) used the term deprivation index to mean a list of indicators, components and subcomponents that have to be seen as necessary utilities or services for households within a community to have a given standard of living.

Creating an index involves many steps:-

- Exploring the available datasets which are assumed to cover the phenomenon being measured. In this step the work initially focuses on identifying the key indicators and components to be selected as the best parameters relevant to the scope of the index.
- Determining an appropriate calculation method. Although there are many methods used to construct indices in different scientific disciplines, the most frequently used is composite index construction where normalisation or standardisation of the index components is applied. This method minimises

differences between the eligible components and results in a common measurement unit for all index components.

- Undertaking a correlation analysis for the selected components.
- Weighting the index components according to some measure of their relative importance.
- Combination of the selected components from each domain into the newly developed index.
- Application and use in relation to a specific research problem.

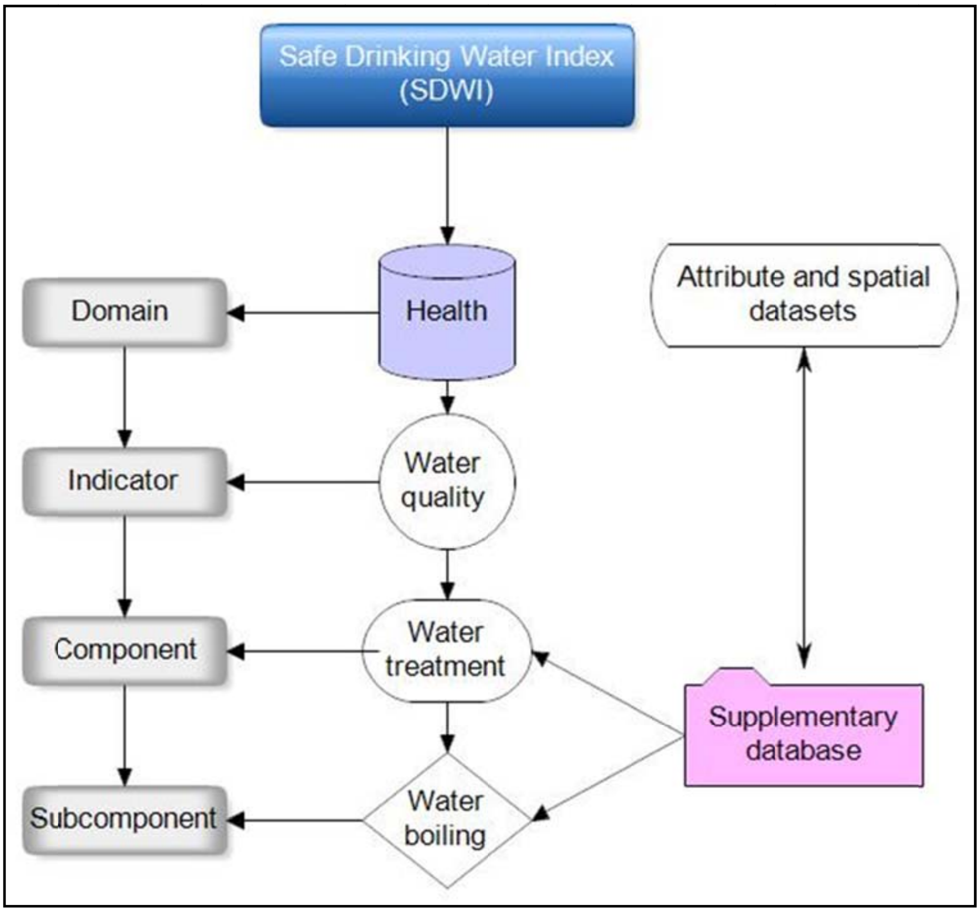


Figure 4.1 Developing a health domain in the Safe Drinking Water Index (SDWI)

4.3 Application of area classification

4.3.1 Applications of area classification in developed countries

Geodemographics is the classification of small spatial units according to their socioeconomic characteristics. Classification approaches have been developed based on the idea that people who live close to one another are more likely to share the same socioeconomic and demographic characteristics and consequently consumption behaviours (Rothman, 1989; Debenham et al. 2001). The preliminary concepts and ideas of geodemographic classification were developed gradually, and at the same time, in the USA and the UK. A geodemographic classification system based on computer software was created originally by Jonathan Robbin in the USA in the 1960s (Goss, 1995; Harries et al.2005; Uprichard et al. 2009). In the 1970s he created the Potential Rating Index for ZIP Markets (PRIZM) which was considered to be the most effective system in the commercial sector and urban policy planning at that time. This system was based on a cluster analysis algorithm using consumer survey data and USA census datasets. For all 36000 USA ZIP codes the system produced 40 exclusive lifestyle clusters. In the UK, Webber studied the spatial pattern of urban deprivation in the city of Liverpool and developed software to identify spatial clusters of neighbourhoods. Webber's work in this field led to the first UK-based geodemographic classification known as 'A Classification of Residential Neighbourhoods' (ACORN) (Burrows & Gane 2006; Harries et al.2005; Weiss 2000). In 1985 Webber developed another classification system called MOSAIC, which included non-census data such as credit card activity and county court judgements. Webber developed these systems for neighbourhood classification and they have been extensively used in business and marketing, in both the public and private domains (Webber, 2004).

Mitchell & McGoldrick (2004) discussed the importance of geodemographic classification. They consider the strength and benefits of the ACORN classification system to be a part of the wide range of methods and techniques introduced to marketing and decision makers, particularly customer profiling, segmentation, demand forecasting, branch location analysis and media selection. These methods often involve the linkage of different spatial datasets. Because of this, geodemographic classification has become an important resource for public services management in developed nations. It has potential as a policing tool to assess public safety and crime occurrence at different spatial scales such as households, postcode or police beat (Ashby & Longley 2005). Batey & Brown (2007) used a geodemographic classification system to generate a socioeconomic profile for area targeting initiatives in relatively poor communities in eight provincial English cities. The main target was to identify population groups who share the same deprivation circumstances and demographic features. They concluded that geodemographic classification offers an appropriate method to address and evaluate the success of spatially targeted urban policy initiatives.

Longley et al. (2006) developed a geodemographic classification in which they investigated the spatial variations of access to Information Communication Technology (ICT) across Great Britain. Many details were produced at household level to express quality of life and socioeconomic characteristics of population according to the level of use of ICT. This classification system could be of value for planning policies that target improving the quality of life in developed nations. Walford (2007) studied migrant flows between rural and urban areas. Using geodeomgraphic classification he addressed the interrelationships between the origins and destinations of migrants who moved into Mid Wales in the last decades of the 20^{th} Century. The main objective was to investigate whether there were area types more likely to supply the most migrants. Voas & Willamson (2001) analysed data from the 1991 census for England and Wales at district, ward and census enumeration district levels. They argued that geodemographic classification is one of the most popular methods of spatial discrimination and description of geographic characteristic of neighbourhoods. However, they criticised small area classification systems as in most cases, a specific classification is required a variety of datasets to produce comprehensive characterisation.

It seems that geodemographic classification is a popular method used in variety of spatial applications concerning services, business and socioeconomic policies in the developed world. This is due to the wide range of available datasets from different sources either census (national census) or non-census such as register statistics and local surveys (Harries et al. 2005). Abbas et al. (2009) discussed the potential application of geodemographics in the health domain. They pointed out that classification mechanisms offer better opportunities to explore health inequalities among population groups. In addition, developing spatial analysis for small geographical units is useful in terms of measuring the socioeconomic variations associated with public health inequalities. In spite of the widespread use of geodemographic classification in many public and private domains in the developed world, few attempts have been made yet to use this method of analysis in studying and investigating water supply issues (Khatri & Vairavamoorthy, 2009). An implication of this is possibly that households in most developed countries, where geodemographic methods were developed, have access to safe drinking water and water supply infrastructures are improved and adequate.

4.3.2 Applications of area classification in developing countries

It seems that geodemographic classification methods have rarely been used or applied in developing countries. To date, only two papers have been identified (Ingwa et al 2008, Allo 2010) and both about Nigeria. Ingwa et al (2008) conducted geodemographic analysis to address the ICT provided by the Nigerian Federal

Government for secondary education organisations under deprived circumstances. Using data from the 2006 Nigerian census and another dataset from 1991 Population Commission reports, they analysed the geographic distribution of internet services for schools and identified spatial inequalities across the country, particularly noting that web services provided to schools were concentrated in the capital region (Abuja). Allo (2010) introduced a framework of geodemographic method to disaggregate the Nigerian postcode to create postcode structures and reducing the spatial resolution from local provincial area to a unit level.

More broadly, geodemographics is a powerful tool for investigating socioeconomic patterns and specifically for describing sociodemographic characteristics of small spatial units. However, the availability of socioeconomic datasets plays the main role in building and developing any classification. Thus, small area statistics are not widely used in spatial analysis applications in the developing countries either in public or private sectors.

4.4 Deprivation indices

4.4.1 Deprivation indices in developed countries

The term deprivation generally refers to lack of basic human needs such as potable water, education, health and other public utilities. There are two types of deprivation: absolute and relative. Gough el al. (2009) defined absolute deprivation as insufficient resources, particularly in healthcare, living environment, food and nutrition. This kind of deprivation is associated with developing countries where deprived circumstances often influence population health where consequently the mortality rate increases and life expectancy decreases. On the other hand, relative deprivation means some individuals or groups of people are less able to meet their basic needs compared with the national average and this concept of deprivation is applicable to both developed and developing countries. In broad spatial terms, deprivation tends to be measured at neighbourhood level comparing households which suffer from some socioeconomic exclusion in urban or rural environments. Deprivation indices have been developed to identify patterns of deprived areas, and often for small geographic areas. Nevertheless, not all households who live in the identified deprived areas experience deprivation, for instance in England on average less than a third of the population who live in deprived areas experience a low income (Syrett & North, 2008).

Deprivation indices are one major tool for identifying poor communities in small spatial units. Senior (2002) reviewed the most popular deprivation indices in the UK. He pointed out that constructing deprivation indices used to depend on census data as the only rich datasets covering the entire country. However, in developed countries non-census data are increasingly being used for constructing small area classifications and indices. The majority of deprivation indices were commissioned for purposes such as analysing spatial variations or basic research description as in Townsend et al. (1988) which led to further work on health inequalities, particularly morbidity and mortality. In the policy planning domain, census data were used effectively to assess local resources in the Index of Local Conditions and the Index of Multiple Deprivation in England, Scotland and Wales (Niggebrugge et al.2005; Noble et al. 2000). In England the 2000 Index of Multiple Deprivation was developed at the ward level to identify populations experiencing relative deprivation and with poor accessibility to local services (Niggebrugge et al.2005; Harris & Longley 2002). The 2004 Indices of Deprivation used multiple domains of deprivation to measure variations in poverty across the UK based on the lower super output area scale to identify deprivation. In this context, a deprived area is considered to contain a large percentage of deprived people and its population suffers from a lack of access to services and other environmental factors (Noble et al. 2006). The Index of Deprivation (ID) (2004) was intended to update the IMD 2000. Seven domains were used to create the index; Income, health, employment, education skills and training, living environment, barriers to housing services and crime domain (Noble et al. 2004). Using the same previous domains, the Index of Deprivation 2007 (2007) was a first update to the ID 2004 using the same spatial units (Noble et al. 2008). The Index of Multiple Deprivation 2010 tended to focus on deprivation based on national and sub-national scale. Most of the statistical indicators used to construct this index were from 2008 (IMD, 2010).

A case study of 175 wards in the city of Bristol in England was developed by Harris & Longley (2002) to measure deprivation. The study had two aims: firstly, 1991 census data were used to predict deprivation by analysing socioeconomic circumstances at different spatial levels. Secondly, the Breadline Britain index* used lifestyle survey data, which are a source of income data at small scale, to identify deprived areas. The findings demonstrated heterogeneity in census enumeration districts (EDs).

*The Breadline Britain index is a poverty score was developed based on a survey undertaken in London TV programme in 1990 to investigate the relative poverty of British people using six variables; unemployment, lack of owner occupied accommodation and lack of car ownership, limiting long term illness, lone parent households, and low social class (Frayman et al. 1991)

Point patterns and aggregation analysis were used to identify the central locations of household clusters classified as having low income levels. They argued that developing effective measures of deprivation in the UK should consider the impact of census boundaries that may lead to artificial segregation of deprived clusters. However, the study was useful in terms of measuring the correlation between the Breadline Britain index and a measure of univariate income taken from lifestyle survey data.

4.4.2 Deprivation indices in developing countries

Similar to the infrequent use of geodemographic methods in developing countries, deprivation indices are also rare. To date, the only major deprivation index at subnational scale (province) in developing countries was conducted by Noble et al. (2006a) for South Africa. The Provincial Index of Multiple Deprivation (PIMD) was constructed using socioeconomic datasets derived from the 2001 South African Census. A variety of components and domains were developed to calculate the index. The five domains and 13 components are summarised as follows:-

i. Income domain

Three components were developed and used to create this domain:

- Number of individuals living in a household with income below 40% of the average income of all households.
- Number of individuals living without a refrigerator.
- Number of individuals living without TV or Radio.

All these numbers were used as numerators while the denominator was the total population

ii. Employment domain

Two components were developed and used to create this domain:

- Number of individuals aged 15-65 who are unemployed.
- Number of individuals aged 15-65 who are unemployed due to illness or disabilities.

These numbers were used as numerators while the denominators were calculated based on adding the total numbers who are economically active to the total number of unemployed individuals due to illness or disabilities.

iii. Health domain

One component was developed (premature deaths) using directly standardised age and sex to measure the health domain.

iv. Education domain

One component was used to in this domain; the number of individuals aged 18-65 who did not attend school at secondary level or above. This number was used as the numerator while the denominator was the total population aged 18-65.

v. Living circumstance domain

This domain includes six components mainly developed from census data that express population lifestyle and household accessibility to public utilities and services;

- Number of individuals living in households that have no access to a landline or telephone.
- Number of individuals living in households that have no access to piped drinking water inside their dwelling, yard or plot.
- Number of individuals living in households that have no access to public electricity network.
- Number of individuals living in households where the dwelling is a shack.
- Number of individuals living in households that have no access to adequate sanitation system particularly a pit latrine or flush toilet.
- Number of individuals living in households that have two or more people living in one room.

The total population was used as denominator while the above numbers were used as numerators to calculate each component as percentage.

Noble et al. (2010) discussed the Provincial Index of Multiple Deprivation (PIMD) and the census datasets used to construct the deprivation index from different discrete dimensions. The Eastern Cape province was used as a case study area. Three main methods (standardisation, transformation and weighting) were used to combine each domain into the index which provided useful clues about the spatial distribution of socioeconomic circumstances of population at small area level, and subsequently identifying deprived areas within each province. Details about the chosen methodologies are found elsewhere (Noble et al. 2006b). The PIMD is a useful tool for national and international companies, organizations and charities interested in deprivation. The authors concluded that many fields could benefit from the PIMD including policy planning for resource allocation and academic research.

Using datasets from the same 2001 South African census, Barnes et al. (2007) constructed the South African Index of Multiple Deprivation for Children (SAIMDC). The ‘adequate care deprivation’ domain was developed and added as a new domain in addition to four similar domains to those used in the PIMD. The index reflects the effects of poverty on child development in terms of health, education and basic living needs rights. This index is the first attempt to analyse and visualise the spatial distribution of child deprivation in South Africa.

4.5 Water index review

This study aims to develop an integrated index which measures access to safe drinking water at sub-national scale taking into account that the index should ideally be applicable internationally, particularly in the developing world. However, compiling an index to measure and assess water related problems is a difficult task. Because of the variety in water programmes, applications and techniques used to address water issues, there have been a number of water indices created from different disciplines and perspectives. However, indices which focus on water quality, scarcity, resources and accessibility are the most widely used to analyse water shortage and measure household access to potable drinking water. Developing an integrated water index at sub-national scale which could be used globally is still a major challenge. One part of the problem is related to datasets but there is also variation in the methodologies and approached applied for index components. Summaries of the most popular water indices are given in table (3.1). The following is a full review for each water index:-

In her major study, Sullivan (2002) introduced the most popular water index (Water Poverty Index) as a new approach for analysis and assessment of water poverty. The index was developed mainly to examine how linking environmental and socioeconomic variables would be an effective approach to water related problems particularly water scarcity, availability, resources and supply accessibility. The index addressed the relationship between the water resource crisis and lack of access to adequate drinking water supplies and sanitation. In the spatial sciences determining the scale at which any index should be generated is an important challenge (Gibson et al. 2000; Cash et al. 2006), but in the Water Poverty Index various methods were suggested for development of any Water Poverty Index, summarised as follows:-

- A conventional composite index where each component in the index is given a weighted value using a national or local scale.
- The gap method in which the target is to assess by how much water components (such as supply, use, resources, access) deviate from a predetermined standard. The standard could be a quantitative or qualitative assessment of groups within society focusing on ecosystems, health system and economic welfare.

- A matrix approach where the index components could be represented in a two dimensional matrix and the individual components identified in each spatial unit (country or region).
- A simple time analysis method in which WPI is determined based on time being spent to access drinking water. A simple equation is applied where WPI = $T/1000\ m^3$ and T refers to the time required for any household to obtain access to a particular quantity of drinking water.

Sullivan pointed out that although there are many approaches that could be followed to produce WPI, geo-referencing of datasets that are collected internationally is needed to make any linkage between macro (hydrological and other environmental factors) and micro (household) spatial data levels.

Building on the earlier work by Sullivan (2002), Lawrence et al. (2002) created an index (The Water Poverty Index) in which five water components, water resources, access, capacity, use and environment were used to measure access to reliable and sufficient water supplies. Consequently, the index highlights communities that are in water stress and suffering water poverty. These indices were linked to different socioeconomic factors such as income level and water affordability, the impact of educational level on managing water supply, consumption of fresh water and health issues. Moreover, the index focused on water resources for each country where water availability for irrigation and agricultural use was examined. Normalisation, standardisation and weighting the components were applied and an analysis of correlation between components was undertaken. A range of aggregated data covering world resources was used to assess water poverty globally.

Application of this index showed that countries gaining higher scores are developed or rich developing countries while a few developing countries scored high scores in one indicator and low in others. A large body of literature has investigated water resources and environment based on this WPI using environmental and socioeconomic variables. This integrated approach has been used to assess water scarcity globally and poor households locally (Sullivan & Might 2003; Sullivan et al. 2003; Sullivan & Might 2007, Molle & Mollinga 2003; Heidecke 2006; Komnenic et al. 2009).

An attempt was made by Sullivan et al. (2006) to develop an application of the WPI THATwould be applicable at different spatial scales. They showed some examples of how the WPI could be applied locally to assess water related problems such as water resources, quality, access, and water environmental issues. Some sites in Tanzania, Sri Lanka and South Africa were demonstrated as case studies at community level. However, the datasets used in the index were collected specifically to cover just 1500 households. However, the index would have been more useful if data from international or national surveys were used, especially as these countries have conducted some of them, such as the DHS. Mlote et al. (2002) found that hydrological and environmental modelling could provide valuable estimates of water quantity and availability. Moreover, improved water supplies and capable infrastructures are important factors in terms of gaining accessibility to clean drinking water. Thus, most communities suffering from water-related problems are likely to be classified as deprived areas.

Sullivan et al. (2009) developed The Rural Water Livelihood Index (RWLI) based on rural factors that affect drinking water accessibility particularly appropriate supplies, adequate water facilities, water allowance for growing crops, environmental pollutant risks which impact water sources and the corruption rank of each country. The index included components, such as clean water environment, and sub components such as pressure on water resources and water pollutants. Weighted averaging was applied to the four components to produce standardized score values ranging from 0 to 100. However, water supply and accessibility issues are always associated with the local circumstances of communities and addressing them according to large geographic scales may omit site-specific variation in drinking water problems (Schouten & Moriarty, 2003). Thus, a WPI to measure water accessibility in rural places should be generated at a sub-national scale. In addition to this, measuring water poverty in rural areas at national scale may fail accurately to represent the variations of water access challenges between different communities. The physical and socioeconomic characteristics of any rural community vary from region to region. Furthermore, the corruption component is not clearly defined, nor is its relation to water poverty in rural communities. Consequently, it does not produce a meaningful contribution to the index, which therefore suffers some limitations in scale, data sources and component section.

Index Name	Author	Index scope	Components	Methodology	Data source	Spatial coverage and resolution
Water poverty index (WPI)	Sullivan (2002)	Water quality, resources and accessibility	Resources, access, use, capacity and environment	composite index , the gap method, matrix approach, time analysis	Statistical datasets suggested to be used from World Bank/JMP/UNICEF	**Coverage**: the whole world **Resolution**: national scale
	Lawrence et al.2002	Water resources, quality, accessibility, use	Resources, access, use, capacity and environment	Normalization , standardization and combination of scores	Aggregate national data from World Bank/ World Resources Institute	**Coverage**: the whole world **Resolution**: national scale
	Mlote et al. (2002)	Water resources availability, quality,	Resources, access, use, capacity and environment	composite index , the gap method, matrix approach, time analysis	Household survey, national datasets from different government departments	**Coverage**: four pilot sites in Tanzania, Sri Lanka and South Africa **Resolution**: locally
	Sullivan et al. (2006)	Water scarcity, water resources and accessibility	Resources, access, use, capacity and environment	Statistical computation to obtain standardized score values range 0 to 100 for each indicator.	national data from World Resources Institute and World Economic Forum Columbia University	**Coverage**: regional (Tanzania, Sri Lanka and South Africa) **Resolution**: community level
The Rural Water Livelihoods Index (RWLI)	Sullivan et al. (2009)	Accessibility to water in rural environment at national level	Accessibility, livestock water security, potable water environment, secure and equitable water entitlement	Normalization , standardization and combination of scores	JMP(WHO/UNICEF) FAO Transparency International	**Coverage**: the whole world **Resolution**: national scale
Index of Drinking Water Adequacy (IDWA-I)	Kallidaikurichi and Roa (2009)	Water accessibility and availability in 23 countries of the Asian Development Bank	1.Water capacity 2.Available water resources 3.Water quantity 4.Water quality 5. Water accessibility	Statistical computation to obtain standardized score values range 0 to 100 for each indicator.	World Development Indicators (WDI) WHO/UNICEF	**Coverage** :regional (23 member countries of the Asian Development Bank) **Resolution**: national scale
Global Drinking Water Quality Index (GDWQI)	Rickwood and Carr (2007)	Water quality, accessibility and environmental issues.	Assesses water quality against WHO guidelines	Assessment of water quality parameters by comparing observation against the accepted WHO guidelines	The United Nations GEMS/Water Program (GEMS/Water)	**Coverage**: the whole world **Resolution**: national scale

Table 4. 1 Water indices, published internationally

Kallidaikurichi & Roa (2009) developed The Index of Drinking Water Adequacy (IDWA) in which they investigate drinking water adequacy applied to 23 Asian countries which are members of the Asian Development Bank. The index aims to investigate water quality, availability and accessibility for each country based on the available water resources and consequently water provision coverage and reliable quantity. The index components were the same as the WPI (resources, access, capacity, use, and environment) but data from JMP were used to examine households connected to public drinking water facilities according to an urban or rural classification. However, the index is limited in terms of data quality and spatial scale since the household water sector needs data at sub-national level to strengthen measurement of spatial distribution and spatial relationships. This may help in obtaining significant socioeconomic and environmental parameters for estimation of household accessibility to potable drinking water.

The *WHO Guidelines for Drinking Water Quality* support developing effective strategies to improve water supply systems and safe drinking water accessibility. Thus, the guidelines cover managing, measuring and analysing water quality for protecting public health from water related disease risks (WHO, 2008). Rickwood and Carr (2007) developed the Global Water Quality Index (GWQI), supervised by the Global Environment Monitoring System (GEMS)/Water Programme. Data from this programme were used in the index as well as selected parameters from the *WHO Guidelines for Drinking Water Quality* to represent different water quality aspects. Classification of the parameters was undertaken according to the WHO guidelines, and three indices were developed to address water quality, its human effects and associated accessibility issues: a Drinking Water Quality Index (DWQI) focusing on chemical parameters, a Health Water Quality Index (HWQI) and an Accessibility Water Quality Index (AWQI). However, due to lack of available monitoring data the indices do not provide any indication of water source safety and the applications were essentially concerned with analysing physical water components in the sources and environments at the monitoring stations that measure water quality parameters, such as pH, temperature, dissolved oxygen, on an annual basis.

There have been some international programmes such as UN-HABITAT that monitor the progress of water and sanitation coverage according to the MDGs (MDG Target 7C). This project focuses upon using local spatial data collected from household surveys (Urban Inquiry Survey) and high resolution remote sensing images to create databases about water infrastructure and public utilisation in many towns and urban centres at the East African countries (UN-HABITAT, 2010). These spatial databases at sub-national scales are useful to investigate geographic variations and problems of water supply and sanitation systems. Despite the substantial benefits which have been found from applications such as these, most existing water indices focus on the interrelationships between water quality and scarcity of water resources. Furthermore, water domains which were used in the actual measurements lack the spatial detail that enable sub-national patterns to be investigated. Clearly, spatial measurements of water access and supply issues have not been taken into account and therefore spatial analysis of geographic patterns of household water accessibility at sub-national scale has still not been addressed.

This review of the literature on water indices has demonstrated that water resource and quality issues have been addressed widely in physical and environmental contexts and, in many cases, water accessibility was highlighted as a part of the index components. This review has also shown a variety of methods used to calculate the water indices. These include comparison and assessment of water parameters, composite index and matrix approaches. Standardisation and combination of scores was used widely. Although there is a considerable and growing literature on developing water indices, spatial measurements and analysis of access to adequate water supplies and public health at the sub-national level have not been clearly addressed. This is either because of index generation at national or regional scale used or the absence of datasets that cover a whole country for small geographic units. Thus, a few applications investigated drinking water challenges at local community levels but this was based on collection of datasets from small household surveys, national departments or international projects that cover only a few sites and locations. As a result, no water index has yet been developed to show an integrated view that covers one nation and provides detailed knowledge and better understanding of spatial variations in drinking water problems.

4.6 Classification or index?

Regardless of whether a classification or index should be constructed to reflect drinking water accessibility, the fact remains that both methods have not yet been used to develop any spatial measure of access to safe drinking water sub-nationally and covering the whole country. Nevertheless, there are many potential benefits that both methods share as follows:-

- Developing either of the two methods would provide a tool for measuring spatial patterns of drinking water in urban and rural communities.

- Using international data in both methods to develop a spatial measure of safe drinking water access, which is transferable to many developing countries, could contribute to the field of drinking water and health.
- It is expected that the newly developed classification or index would be capable of filling an existing gap by measuring the drinking water access patterns across one country in both rural and urban communities.
- Either method could draw on the DHS international surveys that include a rich dataset about drinking water and other relevant socioeconomic and demographic factors. These data could be linked to other local data such the census and then used to construct a geodemographic classification or a safe drinking water index which explores drinking water issues nationally and subnationally.
- Constructing an integrated and comprehensive methodology at sub-national spatial scale across a whole country is an effective tool that could be used internationally to explore water problems and their socioeconomic impacts in developing countries where the DHS survey is undertaken.

Although the two methods can be implemented at a sub-provincial spatial scale and are appropriate for analysing drinking water problems, there are many advantages to constructing a Safe Drinking Water Index (SDWI) rather than a geodemographic classification:-

- An index has not yet been developed to cover lack of safe drinking water and its impacts on human health in line with the MDGs for small spatial units. Thus, it could be used initially to track progress towards the MDG relating to drinking water in many developing countries.
- An area classification system is a descriptive and categorical method and it tends to be used usually in the market, business, retail planning and private sectors while an index is an appropriate statistical tool for ranking areas relative to one another. Therefore, constructing an index is better than building an area classification system in producing spatial measurement which could be used to assess access to safe drinking water.
- An index develops numeric values which facilitates area ranking and thus could be used to identify trends over time using the same datasets. In contrast, an area classification often uses different data over time and may also produce different output classes for different periods. Thus, monitoring based on the classes is not possible.

- Creating an integrated index and possibly updating it over time would be valuable for decision makers and policy planners in developing countries to gain better understanding of spatial variation in trends in drinking water problems. As a result, they could solve drinking water problems in a way sensitive to local variation in safe water access patterns.
- Ranking spatial areas over space and time is very important and considered one of the advantages of developing an index.

4.7 Criteria for component selection

Exploring the available datasets from the international DHS survey and local censuses, five domains (water, environmental, socioeconomic, demographic and education) were identified as appropriate comprehensive indicators and components. They were selected to include all possible valuable parameters related to safe water accessibility. Each domain contains different components that are capable of providing useful information about the relationships between access to safe drinking water and public health. Furthermore, each index component is expected to contribute a significant value to safe drinking water measurement and should not correlate with other components. Criteria of indicator selection include relevance, validity, availability and longitudinal were used to develop indices (Noble et al. 2001; Noble et al. 2004,). Based on the literature, table 4.2 shows an original summary of the selection criteria to be used for SDWI components on the basis of the attributes and spatial data sources for drinking water which could be used to construct the index. They are explained as the following:

- The first criterion is “water relevance” where each indicator should provide useful information about drinking water and thus the component is excluded if it does not have relevance to the SDWI aim and objectives.
- The second criterion is that the component has to be “valid and specific”. This criterion means the component should be valid for measuring safe drinking water. Also it should describe and measure a specific part of the entire domain and thus be not confounded or overlapping with other components.
- Each component should be “available and achievable” which means that datasets for generating the component are available. This is a crucial criterion as a component may be can effective in achieving the index objectives but the data are not accessible - for instance data related to water quality or water borne-disease are not available.

- Each component should be “plausible and complementary”. This means that each component is logically linked to other components in the same domain.
- The ability to produce the same component and construct an equivalent measurement at different times using the same datasets is a very important consideration, so the component should be “longitudinal and repeatable” to ensure that the measurement could be repeated.
- As the SDWI is to be a quantitative metric, each component should be “measurable and reliable”. The component should be quantifiable and in numeric values such as percentages to precisely address safe drinking water.
- The SDWI will be implemented within a GIS framework and thus each component should be mappable at a sub-national geographic scale (kisms/markazes).

4.8 Key differences between census and survey data

The spatial linkage of the DHS and census was undertaken to investigate relationships between drinking water and health of the Egyptian population. The DHS and census seem to be similar in terms of question definition and demographic structure while they do not always present the same values. One of the main definitions of a census given by Rees et al. (2002) is ‘a device for counting populations and recording their characteristics’’. A census is a statistical enumeration conducted essentially to collect demographic and socioeconomic datasets that cover the entire population and geographies of a country.

“Census” is defined by the Oxford Dictionary as “(*denoting a poll tax): from Latin, applied to the registration of citizens and property in ancient Rome, usually for taxation, from censere 'assess'* (Oxford dictionary, 2011). In ancient civilizations governments had taken censuses to collect data for management and planning purposes (ONS, 2010). Historically, the first census is believed to have been taken 6000 years ago by the Babylonians. Similarly, the Egyptian census in 2500 BC was related to general statistics of Pyramids builders and Roman people took the first complete census in 500 B.C for military services and taxation purposes. Recently, censuses have been used in many fields such as agriculture, industry, trade labour, health (Australian Bureau statistics, 2010). Within Egypt a census has been conducted every 10 years and the last census took place in 2006, conducted by CAPMAS. It is considered the most important source of social, economic and demographic statistics.

Component criteria	Criterion principle and definition	Drinking water context
Water-related and relevant	The component should be relevant to the index objectives.	A component about sanitation type is relevant to environmental circumstances and much linked to drinking water as access to improved drinking water source is associated with adequate sanitation system. Likewise poor sanitation can cause water contamination.
Valid and specific	A particular component which describes a specific part of the entire domain to which it belongs and does not affect multiple domains.	Data about water sources could be used to construct a valid component which measures the spatial variations in access to public water supply.
Available and achievable	The required data for developing the component should be available and accessible.	DHS datasets are free and easily accessible and include useful questions about drinking water issues. Most local censuses collect data about drinking water patterns and coverage.
Plausible and complementary	There should be a credible link between the component and the index domain to which it belongs; the component should not overlap substantially on components selected elsewhere.	For instance in the SDWI, a component about home water treatment is intuitively related to water quality and therefore safe water access.
Longitudinal and repeatable	The component should be capable of producing the same result about the same groups at different points in time using the same data. This is useful in terms of analysing trends.	Drinking water components in the SDWI were selected from the DHS which is undertaken annually in most developing countries and census which is decennial (every 10 years).This gives great opportunities to develop the same component over time.
Measurable and reliable	The component should be based on reliable measurements and quantifiable in values such as percentages or ratios.	For example in the SDWI water quantity and availability could be measured by the percentage households experiencing water supply interruptions.
Demonstrable and mappable	For spatial analysis purposes, the component must be capable of being implemented in a GIS and mappable at sub-provincial level.	The datasets used for developing the SDWI components are georeferenced and thus each component is capable of being implemented within a GIS and is mappable.

Table 4. 2 Shows criteria selection for the SDWI components

The survey is another valuable method that has been designed for researching issues regarding populations and households. Surveys are taken in many subject domains, particularly in the social sciences. Surveys differ from traditional censuses in that they are not compulsory and no person has to take a part if they do not want to answer the questions (ONS, 2010). Surveys have many advantages such as determining social change outside census years. Furthermore, researchers can obtain different kinds of information related to many subjects that are not included in the census, for instance data and information on smoking and alcohol drinking and other socioeconomic variables.

4.8.1 Aims and targets

Census and survey aims and targets are different. The Egyptian census aims to collect data from the entire Egyptian population in all governorates according to geographic census divisions (provinces, kisms/markazes, and shyakhas). On the other hand, the DHS survey is a sample survey aiming to collect statistical data and the main target is Egyptian households with women and children. Despite the DHS samples being taken from all Egyptian governorates, they covered only a few clusters of households in each governorate.

4.8.2 Time differences

While censuses are usually held every decade, sample surveys are undertaken more frequently, depending on organizational decisions and the need for specific socioeconomic estimates and measurements. The Egyptian census from which data are used here was held in 2006 by government's statistics agency (CAPMAS). By contrast, the Egyptian DHS was conducted by the Ministry of Health and Population (MOHP) and National Population Council (NPC) and implemented by El-Zanaty & Associates with technical support provided by ORC Macro in 2005.

4.8.3 Geographic levels and spatial linkage

A major difference between the EGYPTIAN DHS and the Egyptian census is the geographic level of collected data which impacts on any spatial or statistical analysis. In the Egyptian census the attribute data could be aggregated to the smallest geographic unit (shyakha), the middle unit (kism/markaz) or the biggest unit (governorate). Although GPS units have been used in the majority of the DHS surveys, and therefore households clusters are georeferenced, there is little spatial linkage for the GPS DHS to administrative boundaries. This is because spatial error is deliberately applied to the GPS cluster locations. Moreover, the DHS samples are always drawn from higher spatial units: province (governorate) and regional (rural and urban).

4.9 Summary

In this chapter, the construction of area classifications and indices, both water and non-water, have been reviewed. Differences between geodemographic area classification and index methods have been considered. For instance, the main purpose of using area classification is to assign each small geographic unit to one class, while indices are statistical tools resulting in numeric measures built from various domains and indicators. Although there is a rich range of applications of geodemographic classifications and indices in the developed world, in the developing nations these techniques have seldom been used and when found, it is at the governorate or regional scale. Most water indices focus on water quality and environmental factors and when an accessibility variable is included, the sub-national scale is absent. Thus, highlighting problems of access to safe drinking water at the local level requires construction of an index at an appropriate spatial scale.

Based on the literature, criteria were presented for constructing an index to measure Egyptian household accessibility to safe drinking water at the sub-national scale. Such an index could be updated over time using the same data sources or adding others and this potentially gives a great opportunity to capture and highlight spatial variations and track identified problems. Some major differences between the Egyptian census and surveys as sources of data were highlighted, especially in terms of their aims and objectives, time difference and spatial scale and coverage. These considerations are important because these are the key data sources for index construction.

Chapter 5: Developing SDWI for Egypt at sub-national scale

5.1 Overview

This chapter discusses the key points concerning the process of constructing a SDWI for Egypt at sub-national scale. It starts by explaining the index scope and domains in more detail and underlines the criteria developed for evaluation of components in order to construct a multivariate index. It also considers indicators and components from the literature, particularly the importance of each domain in terms of investigating the spatial patterns of drinking water. The explanation highlights both the index components excluded as well as those included. The second half of this chapter outlines the methodological approach used to score the index components, specifically those concerning water, such as: water quantity and fetching, water interruption, water treatment and supply types. Similar consideration is given to a sanitation system indicator drawn from census data that cover the whole country. An assessment of the distribution of scores has been undertaken and all the indicator scores graphed to identify their distributional characteristics. To ensure that each index component does not overlap with other, correlation coefficients between the indicators are examined. All indicators have been converted to standardized units. Finally, the spatial distribution of the SDWI in sampled kisms/markazes was investigated and analysed.

5.2 Scope and domains of the SDWI

The components of the SDWI were selected and included according to the selection criteria (see section 4.7 and table 4.2). Selecting domains and indicators of the SDWI initially requires addressing and evaluating all factors associated with household drinking water issues in Egypt. All domains were developed based on the available data sources at sub-national geographic scales. Each index domain was constructed separately and then combined into a single measure of safe drinking water access. A set of candidate components (12 components) were generated from the available census and DHS datasets. These components were evaluated and assessed against the criteria. Scores were assigned against each criteria for each index component and then components with lower scores were excluded from the final index. In the final multivariate index, eight variables were identified as potential components, based on a set of seven criteria such as relevance, repeatability, and reliability of measurements (see table 5.1). Two components were excluded from the index as they did not score highly enough (characteristics of person who fetches drinking water and water used for cooking) while two components (Dwelling types and child diarrhoeal disease) were excluded as they demonstrated high correlation with other components. The index components were all range-standardised prior to application of a weighting scheme and combination. The selection criteria which were used to choose valid and valuable components for constructing the SDWI were represented in section 4.7.

Table 5.1 Safe water access index components, with associated data sources, weights and selection criteria

Criteria for selection (1 to 7)	Water supply	Water quantity	Water interruption	Water treatment	Sanitation system	Population density	Household education	Wealth factor
1. Valid and specific*	5	4	4	5	4	3	2	3
2. Available and achievable*	4	4	4	3	4	3	2	2
3. Plausible & complementary *	4	4	4	3	3	2	2	2
4. Water related and relevant *	5	5	5	4	4	1	1	3
5. Longitudinal and repeatable*	4	3	3	2	5	4	3	2
6. Measurable and reliable*	4	3	3	4	3	2	2	1
7. Demonstrable and mappable*	5	3	4	5	4	5	1	1
Data source	DHS/Census	DHS	DHS	DHS	Census	Census	Census	DHS
Mean of weight scores **	8.33	5	5.33	4	7	6.67	8	7
Range**	2	2	4	3	0	3	2	0

**Scale for weights 1=least important; 5= of intermediate importance; 9= most important

*Scale of evaluation for criteria: 5=Very strong; 4=Strong; 3=Medium; 2=Weak;

Applying weights to the SDWI components

In developing a multivariate index, eight variables were identified as potential index components, based on a set of seven criteria such as relevance, repeatability, and reliability of measurements (see table 5.1). The index components were all range-standardised prior to combination. Numeric weights for each index component were developed through consultation with three independent experts (2 Egyptian, 1 International) who rated the importance of each component in an Egyptian context.

The two Egyptian experts are geographers and professors at Alexandria University. One is a human geographer, with particular interest in urban and rural Egyptian geographies, and has broad experience in terms of demographic and socioeconomic phenomena in Egyptian communities. His interests also include deprivation and water poverty. The second Egyptian expert is a physical geographer with experience in environment and spatial factors that impact Egyptian household living especially environmental hazards. They both offered spatial perspectives on drinking water problems in the Egyptian context. The third expert is an academic at the University of Surrey. He is a specialist in the development and application of microbiological indicators of water quality in piped water systems. He has wide experience on underground water and sanitation system problems in developing countries and has conducted many projects on drinking water quality in Sub-Saharan and Asian countries.

Weighting of index components is intended to reflect the importance of the underlying concept being assessed, rather than the indicator's suitability for measuring a given concept (Mlote et al.2002). The group of experts participated in evaluating the importance of each index component and its influence and relevance to measuring safe drinking water accessibility. Each expert was initially approached to obtain their consent for taking part in this exercise. They then assigned a score between 0 and10 where 0 means that the component is least important while 10 means it is most important. A letter was sent to each external expert inviting them to assign scores to all the selected index components (see appendix 2). The letter starts with an introduction which explains the general rationale behind constructing the SDWI. A brief explanation of each selected index component was presented in the letter. The letter asks the stakeholder to give a score as a weighting value for each component according to the importance of each component for the measurement of drinking water accessibility in an Egyptian context. Finally, the letter ends with three questions: (i) whether there are other components that should have been included in the index; (ii), whether there is

one or more indicators that should be excluded from the index and (iii) any general comments the respondent wants to add about the index creation.
The main objective behind the weighting scheme process was to generate a weighting value for each of the eight components which will determine their contribution to the total combined index value. The group of experts replied with all index components assigned to weighted scores (table 5.1). A mean weight for each component was then calculated and this mean will be multiplied by the standardised value to produce the final index value for each component before integrating component values into a single index. The components were then combined through weighted linear combination. Index values were calculated at the district (kism or markaz) level.

The responses of these experts were studied and compared to each other. They all agreed that the SDWI would be a useful contribution to measuring safe drinking water accessibility internationally. Furthermore, water and environmental components were given the highest rankings in terms of their importance and relevance for the index construction while disagreement between the respondents was mainly about the water treatment component. Egyptian experts indicated that water treatment is not widely recognized among Egyptian households while bottled water is more important, particularly in urban areas. The international expert suggested for future consideration that the final assessment be incorporated into a piece of software which could then be used in a mobile phone questionnaire to generate large-scale spatially linked data.

	Water supply	Water quantity	Water interruption	water treatment	Sanitation system	Population density	Household education	Wealth factor
Stakeholder 1	7	6	4	3	7	8	9	7
Stakeholder 2	9	5	4	3	7	7	8	7
Stakeholder 3	9	4	8	6	7	5	7	7
Average	8.33	5.00	5.33	4.00	7.00	6.67	8.00	7.00

Table 5.2 The average of weighted score for each index component calculated from stakeholder weighting scheme

The index consists of five domains: water, demography, socioeconomic, education and environment. Each domain contains one or more indicators (figure 5.1) with eight indicators overall. For instance, the water domain contains four indicators (supply types, time taken to fetch water, water interruption and water treatment) while the socioeconomic domain comprises just one indicator which is "wealth factor":

- **Water domain**

This is the most important domain in terms of explaining spatial patterns for household access to safe drinking water across the entire country. It covers supply type, water interruption, water quantity and water treatment indicators.

- **Demographic domain**

This domain consists of a population density indicator i.e. number of people in each geographic unit (kism/markaz) from the census. It covers the interrelationship between population density and household access to safe water. It is expected that overcrowded and unplanned settlements suffer from poor and unreliable drinking water supplies.

- **Socioeconomic domain**

This domain is intended to represent the relationship between deprived households and poor water supplies. It consists of a wealth indicator which measures cumulative living standard of households in each geographic unit (kism/markaz) based on data from the DHS survey. Wealth is calculated from data collected about ownership of cars, televisions, housing, furniture and other household materials. The higher score is the richest while the low is the poorest.

- **Educational status domain**

This domain consists of one indicator which expresses the percentage of Egyptian people who are able to read and write in each kism/markaz based on the 2006 Egyptian census. It is assumed that highly educated households are more likely to be concerned with potable drinking water delivery and supply.

- **Environmental domain**

The domain consists mainly of a sanitation system indicator. It expresses the percentage of Egyptian households in each kism/markaz based on their access to each type of sanitation using data from the 2006 Egyptian census.

5.2.1 Water domain

Populations living in the East Africa region suffer from big challenges in terms of water quantity and access to safe drinking water (Thompson, 2000). This was confirmed by Davi & Bostoen (2009) for all Eastern African countries included in their study (Uganda, Kenya, Eritrea, Ethiopia, and Madagascar). However, no attempt was made to consider other factors that might affect water quantity. Despite time taken to collect water being an important indicator in measuring water quantity and consequently identifying its impact on household health, there are still many indicators, particularly water interruption and public network capabilities that can affect confidence in such results produced using just one variable. For instance, time taken could be zero while households suffer from water interruption and shortage. In this case, the quantity will be affected and therefore other water practices such as water storage and using tankers are usually followed by households to guarantee water availability. These home practices certainly have impacts on health and challenge the idea that a water tap inside a dwelling equals to water supply coverage. Finally, time taken to collect water suggests that the point of source is different from the point of use and water could be contaminated during storage and as a result of using different types of containers (Gundry et al. 2006).

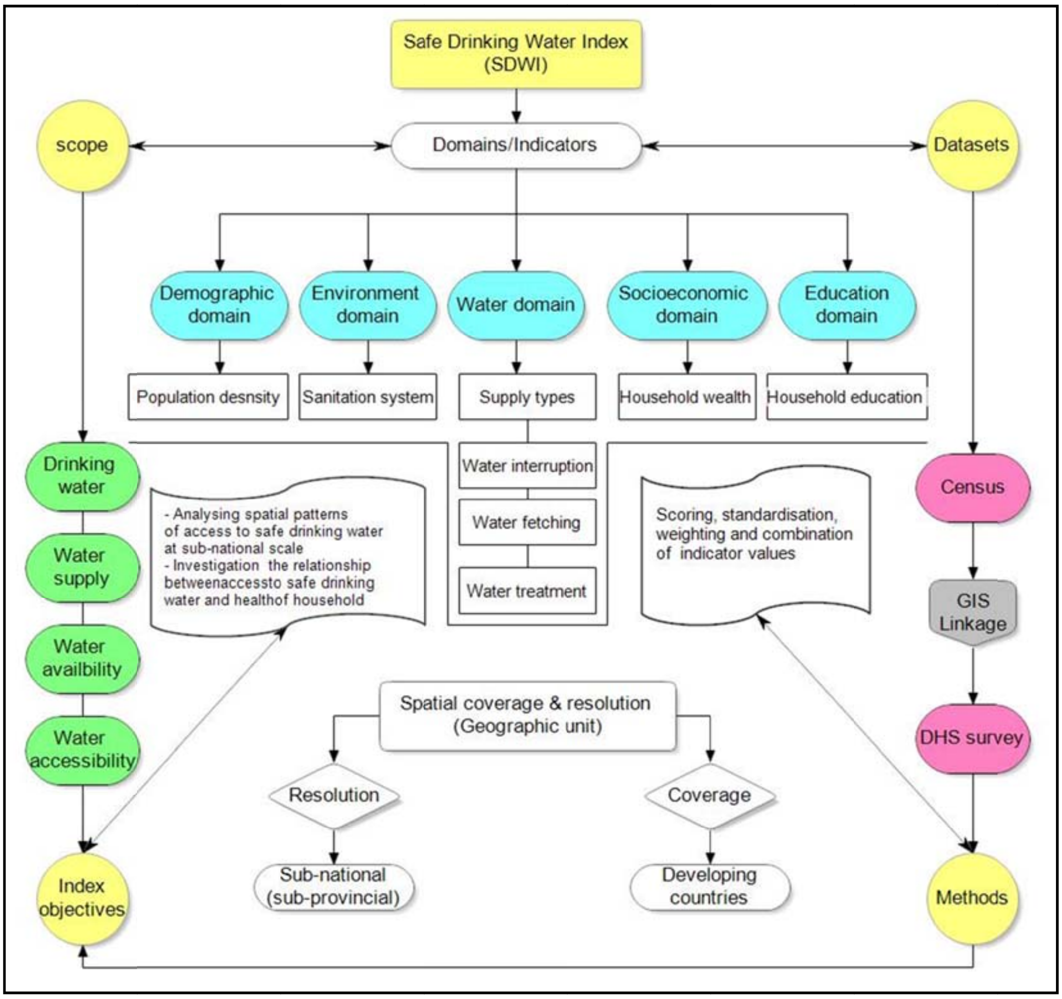


Figure 5.1 Flowchart demonstrating the process of Safe Drinking Water Index construction

Time taken to collect water component

Devi & Bostoen (2009) addressed the quantity of available water as an essential factor which influences access to safe drinking water and improved water sources. Different definitions of access to drinking water were addressed particularly WHO/UNICEF and JMP. The East African countries of Uganda, Kenya, Madagascar, Eritrea and Sudan were selected as a study area because of the low percentage of households who are connected to public water supply and the spread of infectious water diseases among populations. Datasets mainly from the DHS, MICS covering 1997-2003 were examined, including water quantity as well as quality in determining accessibility to potable water. Two models were created. The first JMP model does not consider water quantity while the authors' model does. A comparison was made between the results produced from the two models and it was apparent from the findings that the percentage of households who had access to safe water decreases when the second model, considering time taken to collect water, is applied. There was an inverse relationship between water quantity and time taken to collect water and distance to water source. Much time spent in fetching drinking water from outside the dwelling equals a long distance to water source.

This reflects challenges in the amount of potable water available to households. Thus, time taken to collect water is an effective proxy in measuring water quantity. Furthermore, it is considered an indicator for microbiological water quality and health problems that are likely to occur during the time between water collection and storage such as faecal contamination due to diarrhoeal symptoms (Clasen & Bastable, 2003; Maraj et al. 2009).

Water interruption component

Hunter et al. (2009) have shown strong evidence that there are associations between the reliability of the potable drinking water supply and the risk of water borne diseases. They have argued that failure of water provision systems and poor water delivery could lead to consumption of untreated water and as a consequence, increasing water-related illness, particularly among children aged less than one year. The findings draw attention to the impact of water interruption on household health. Furthermore, a few days of interrupted and poor water supplies are likely to increase household exposure to consumption of raw and untreated water and consequently increased risk of infection. Hunter et al. (2005) found that diarrhoreal disease is strongly associated with water interruption and low water pressure. Surprisingly, the findings were found in a developed country setting using a case control study of sporadic cryptosporidiosis conducted in Wales and the northwest of England. Failure of the water infrastructure to deliver water affects both water quality and quantity if the interruption of piped connections occurs regularly. Thus, water interruption has significant impacts on increasing diarrhoreal infectious diseases, particularly in the rural areas of many developing countries where households do not treat drinking water, and usually consume raw water which is more likely to be contaminated. Hence, it could be hypothesised that improving public drinking water systems and maintaining water utilities would have the effect of reducing diarrhoeal diseases rate in developing countries.

Drinking water source component

Drinking water sources include groundwater via boreholes, wells, rivers, and water tankers. This component can reflect the safety level of water sources and whether they are improved or unimproved. The definitions and explanations of these terms have been discussed in detail in chapter two (section 2.2). Drinking water source is a significant factor in investigating household drinking water and water quality indicators. In the drinking water literature, many studies have been published on water source and its impacts on household accessibility to safe water and other water quality issues, particularly in developing countries (Xue et al. 2009; Ferguson et al. 2009). Zvidzai et al. (2007) considered traditional drinking water sources in rural communities as they might be microbially contaminated and consequently associated with water-borne diseases.

As a part of the SDWI, a water source component can reflect water quality and highlight spatial variations of household water sources in rural and urban areas. The DHS and census provide data about water source types which are capable of development into a reliable component to identifying the spatial distribution of water source types at subprovincial scale across the country.

Water supply type component

Water supply types are related to water provision and utilities especially in deprived areas in rural and semi urban communities. Investigating water supply coverage gives a clear indication of water accessibility and quality. Thus, it is expected that drinking water from a tap is different from pumped underground water as the first type is assumed to be public and protected while the second is private and possibly unprotected. MacGillivary et al. (2006) discussed the water distribution system as it consists of various parts such as water source, pipes, pumps, tanks, plants and catchments. These variables work together to successfully deliver drinking water to households in local communities. They argued that water network analysis which models water supply distribution could be used to assess household demands and measure water supply availability. Farooqui et al. (2007) claimed that poor villages and communities suffer from drinking water problems, particularly unreliable water supplies and inadequate sanitation facilities, and thus are vulnerable to illness and infectious diseases such as typhoid.

This component could contribute to the SDWI as a parameter measuring the spatial distribution of household accessibility to public water networks across all subprovincial areas. Likewise, it would identify areas with poor access to piped supplies, where private water supplies such as pumps, unprotected springs, and unprotected wells are common.

Water storage component

Several studies have revealed that a significant relationship exists between home drinking water storage and health risk exposure, associated with collecting drinking water from surfaces and unimproved sources (Fosso-Kankeu et al. 2008; Jagals et al. 2003; Chemuliti et al. 2002). A case study of drinking water storage in Northern Coastal Ecuador was published by Levy et al. (2008). They aimed to explore water quality and contamination due to home water storage and whether drinking water changes its nature over time. The findings of this study showed that uncovered containers of drinking water are likely to be responsible for increased microbial contamination factors. Consuming raw and surface water was related to significant health problems compared with tap water. However, reduction of contamination risk is likely to be associated with home water treatment especially using boiling and chlorination disinfection.

Regarding water storage as a component of the SDWI, it indicates water quality and hygiene practices. The data used to develop this component are available from the DHS for the households who store their drinking water. This component is expected to be a reliable measure of household water storage over rural and urban communities. It might be correlated with the water interruption component since water storage containers and tankers are used in rural deprived villages with poor water supplies (see section 3.2.4). In addition, the impacts of contaminated water on household health, might be related to storage containers, are are considered another element which would be reflected in the water storage component.

Water treatment component

In the literature, there has been a large and growing body of research highlighting the impact of drinking water sources and household treatment on population health in the developing countries (Clasen et al. 2006; Rosa & Clasen, 2010; Mausezahl et al. 2009, Hunter, 2009; Tiwaryi et al. 2009). However, no attempt has been made yet to include the effects of household water treatment on public health as an element of any water index.

In this component, the main objective is to identify the most common and specific home water treatment approaches used by households to make drinking water safer and potable. This component consists of many subcomponents such as water boiling, adding chlorine, water straining using sand or ceramic filters, solar disinfection and letting water stand to settle.

5.2.2 Environmental domain

From the available datasets, two components could potentially represent the environmental domain: access to an adequate sanitation system. The component reflects the effects of household environment on drinking water.

Sanitation system component

Tsuzuki (2009) pointed out that vulnerability of the local water environment could be estimated based on pollutant discharge parameters. Improving water quality requires decreasing pollutant discharge, particularly from sanitation systems, in developing countries. In the MDGs, halving the percentage of population without access to improved sanitation systems is targeted as well as halving the percentage of households without access to safe drinking water (UNICEF/WHO, 2006). Lack of access to adequate sanitation systems certainly affects public health and the burden of microbiological diseases. Usually improved drinking water supplies are associated with improved sanitation systems and vice versa. Fewtrell et al. (2005) published a paper in which they implemented a systematic review to identify studies of reducing mortality, morbidity and illness rates, particularly diarrhoeal disease, through improving accessibility to safe drinking water and sanitation facilities. They suggested that giving more attention to increasing the proportion of households with access to potable drinking water and adequate sanitation are the most effective factors in reducing diarrhoeal disease in low income countries. Pruss et al. (2002) divided the world's countries into 14 regions to analyse the relationships between waterborne diseases, particularly diarrhoeal disease and sanitation. They found that 4% of death and 5% of morbidity cases are related to water-sanitation problems.

Therefore, public health is affected by the improvement of water supply and sanitation systems. Checkley et al. (2004) argued that unimproved water sources and sanitation facilities influence children health and increase mortality and morbidity rates especially deprived rural and semi urban areas in the developing countries. Bartram et al. (2005) pointed out that water supply and sanitation coverage should be substantially expanded to all unserved poor areas. Policy planners and development programmes should also identify these communities and connect them to potable water supplies and adequate sewage systems. As result, improving environmental and socioeconomic circumstances in these communities leads to advances towards meeting the water and sanitation MDG goals (Hutton & Haller, 2004).

5.2.3 Socioeconomic domain

In this domain, components were developed to represent household wealth factor. Bezerra et al. (2007) used drinking water supply to assess the effects of socioeconomic conditions on population health in Ceara, Brazil. Low income, illiteracy, children working, healthcare, sanitation and water tap variables were used as indicators to estimate infant mortality rates. They concluded that urban and rural areas with better access to public utilities, particularly fresh drinking water and appropriate sanitation systems were associated with decreased infant mortality.

Household wealth component

In the water poverty index, Sullivan (2002) pointed out that household income is a strong factor in determining water affordability and availability. Dungumaro (2007) reported that in the developing countries that suffer from water shortage, drinking water availability and affordability are related to the socioeconomic characteristics of households. He used logistic regression methods to model the probability of household accessibility to safe water sources in South Africa. The finding was a significant relationship between socioeconomic conditions and obtaining water from an improved source. Moreover, households who were characterised by low income and deprived conditions lived in traditional dwellings that were not provided with clean water supplies. Consequently, these households suffered from waterborne diseases. In the DHS, data on household wealth could be used to develop a component that might be capable of reflecting socioeconomic status, as related to drinking water supply patterns. In this sense, the household wealth component in the SDWI would work as a socioeconomic indicator of drinking water affordability, where safe water is a cost option.

5.2.4 Education domain

Household educational level is an important aspect of socioeconomic characteristics and can be used as a determinant of home drinking water quality. Wright & Gundry (2006) analysed Egyptian household characteristics based on home water treatment. One of the major findings was that filtering home drinking water was correlated with high educational levels. Therefore, educational level might be a useful parameter in the SDWI to estimate household attention to home water purifications. Household educational levels could be calculated from the DHS and census datasets as a specific socioeconomic indicator.

5.2.5 Demographic domain

In the demographic domain, the population density component was developed from census data to represent the spatial variations of density across urban and rural areas. The inhabited areas in the Nile Valley and Delta Governorates suffer from population pressure as there are many people and not enough space for horizontal settlement growth and thus public infrastructures, particularly water supply networks, may be influenced by population density.

Population density

Population density measures how closely populations live together. Because access to utilities and public infrastructure varies between urban, semi urban, desert and rural regions, particularly in Egypt, the percentage of households connected to adequate public water supplies is greater in urban than in rural and desert communities. Consequently, poor water supplies and unsafe drinking water may be expected to be associated with deprived rural areas and slum settlements often located at city margins with high population density and high room occupation (Angueletou-Marteau, 2008). In the SDWI, a population density component may be expressed as the number of people per squared kilometre in each kism/markaz.

5.2.6 Components excluded from the final index

Dwelling type component

The main objective of developing this component was to investigate dwelling types. It is expected that water supply varies greatly from one area to another, based on dwelling types. For example, villas, flats, or urban houses are often provided with tap water and public supplies while tents, yards, rural houses, shacks and shanty dwellings could depend on private sources such as pumps, wells, unprotected springs or indeed lack access to any water supply. In addition, many households may live in one dwelling consisting of many floors but with one supply type. Therefore, developing this component in the SDWI could be used to measure whether there is a significant correlation between household accommodation type and water supply type. This component is relevant to measuring and assessment spatial relationships between community and household living environment and access to safe drinking water. The data required to develop this component are included in the DHS and census. The measure could be the percentage of each dwelling type in each subprovincial geographic unit. However, this indicator was excluded from the final socioeconomic domain as it overlaps and correlates (figure 5.2) with the household wealth component (Pearson correlation coefficient 0.86).

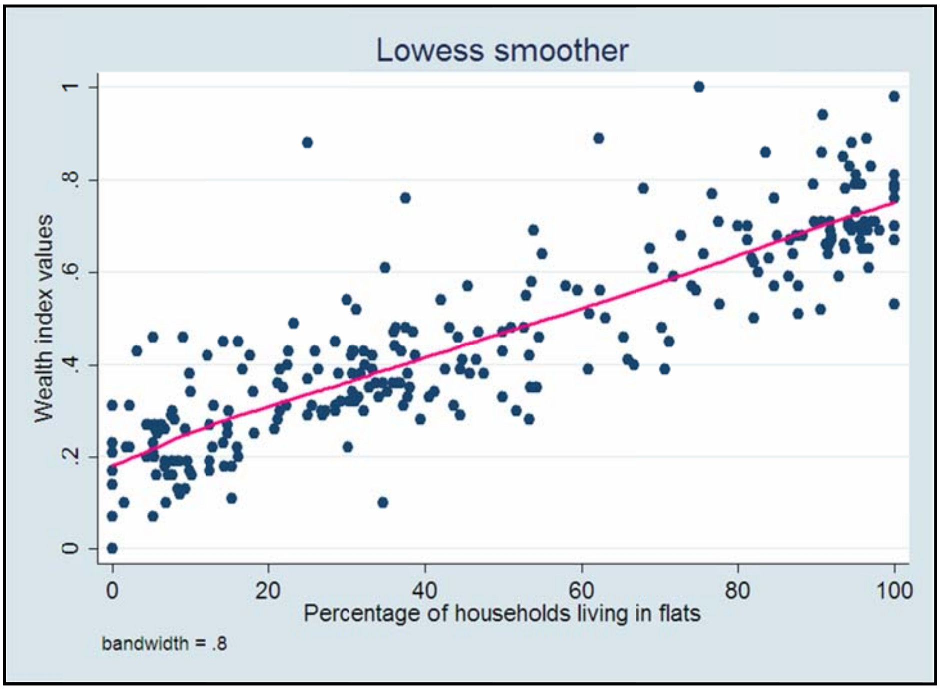


Figure 5.2 The relationship between percentage of Egyptian households living in flats and household wealth factors

Diarrhoeal disease

Despite inadequate sanitation and food sources being clearly related to pathogenic infectious diseases, unimproved drinking water remains as an important source of gastrointestinal disease particularly diarrhoea, which is responsible for high child death rates in developing countries (Ashbolt, 2004). Jensen el al. (2003) assessed the impacts on child health of adding chlorine to drinking water in rural Pakistan. They monitored the relationship between diarrhoea symptoms among children aged less than five years and consumption of drinking water from private and public supplies. The findings indicated that children who lived in dwellings provided with an underground water supply were less likely to be infected with diarrhoea. On the other hand, children who consumed water from public water supplies (where the chlorination scheme was applied) had a higher risk of diarrhoeal infection. Obi et al. (2009) evaluated the effects of poor water quality and supplies on diarrhoeal disease in a small community in Limpopo Province in South Africa. They identified diarrhoeal disease cases that were clustered and associated with inadequate microbial water quality. Data about diarrhoeal disease were only available in the DHS and could be used to construct this component to reflect illness based on lack of access to safe water.

Despite there being several water poverty indices published, they are all limited to investigating and addressing the significant relationships between population health outcomes and access to improved water supplies. Although diarrhoeal disease was excluded from the final index components, a section in this research will examine the interrelationships between the health of Egyptian households and safe drinking water access. This section will focus on examining whether there is a correlation between the percentage of child diarrhoeal disease in each kism/markaz and the final index values.

5.3 Method for scoring the index components

The core purpose of developing the SDWI is to measure spatial variations in access to safe drinking water and their impact on people’s health in small Egyptian communities. Hence, water components from both the DHS and census are the basic inputs to the overall index. The main objective of this section is to develop a methodology by which all the potential components in each indicator can be assigned scores. Scores are necessary in order that the observed values or categories of indicators related to the safety of drinking water can be assigned numerical values and combined in a controlled way into a single overall index. The water and sanitation indicators are particularly important parts of the index. Except for the sanitation type indicator, for which scores were based on the JMP and MDG, a new scoring method has been devised for each of the other indicators. Each of the indicators is considered here in turn.

5.3.1 Water supply type indicator

The water supply type indicator was created using both DHS and census datasets. It was composed of the most common and essential water sources, tap water and underground water. The presence of tap water suggests that a household is connected to a public water supply, while underground water sources are water pumps, protected, unprotected and tube wells. In terms of household safe water accessibility, underground sources usually indicate that households consume untreated water and suffer from a lack of access to potable public supplies. Therefore, these water sources are more likely to be unimproved and unsafe. This is in accordance with the assessment presented in the MDGs where water sources are classified into just two classes, “improved” and “unimproved” (see section 2.2.1). However, reviewing the literature on household water quality and safety it was apparent that significant microbiological and chemical variations exist between sources within the same category (improved or unimproved). Thus, it is appropriate to develop a more detailed scoring system which takes into consideration variation in drinking water quality by source type.

To do this, a literature review was conducted to identify Egyptian studies of drinking water quality for different source types as defined by the census:

- Tap water: Analyses of water sample contamination by *E.coli* bacteria and faecal streptococci were conducted as microbiological evaluations to assess tap water safety and quality in Egyptian drinking networks (Lasheen et al. 2008; Diab 2002; El-Taweel & Shaban 2001; El Abagy et al. 1988).
- Groundwater: Underground water samples from different Egyptian sites were investigated as well as tap water to examine and count microbial contaminants and their influences on human health (Abo-Amer et al. 2008; El Abagy et al. 1988). Chemical contaminates in drinking water samples from Egypt were presented in some studies and it was concluded that there were variations in chemical concentration in household drinking water supplies (Lasheen et al. 2008; Soltan, 1999; Saleh et al. 2001).
- Other source types: Except for microbial evaluation of bottled water in Egypt (El Abagy et al. 1988), no attempt has been made to evaluate other water supply types such as tanker trucks or carts with tanks. Therefore, a study from a developing country, Indonesia (Gupta et al. 2007) was used instead to evaluate tanker trucks as a water source component. This study was the only one to examine and measure quantitatively the water quality of tanker trucks as a source of drinking water in developing countries.

In each study an overall percentage safety score was obtained and assigned to the relevant water supply type. A score of 100 indicates that the source type was consistently safe in terms of its water quality, whilst a score of 0 indicates that the source type was consistently unsafe in terms of its measured water quality. Intermediate values reflect the proportion of samples considered unsafe according the *WHO Guidelines for Drinking Water Quality (WHO, 2006).* A median was then calculated from these in order to produce a score for each supply type (table 5.3). To obtain overall scores for each geographic unit (kism/markaz), this median score was multiplied by the percentage of households using each water supply type.

5.3.2 Home water treatment indicator

The home water treatment indicator was based on data available only in the DHS since no data are available in the Egyptian census relevant to water treatment. There are many lessons that can be learnt from studying household water treatment, particularly capturing the interrelationships between household treatment practices and public health. In addition, water treatment could be an effective indicator in measuring water quality and deliveries in small communities. There were two possible approaches that could be followed to develop a scoring method for each household water treatment type:-

- Considering the large volume of published studies describing the influence of home water treatment on diarrhoeal disease (see section 2.2.5), one way would be to address the impact of each household water treatment type (e.g. filtration, boiling) on reducing diarrhoeal disease.
- A second way would be to score household water treatment types according to their influence on water quality. This was considered to be the preferable approach because it is comparable with the scoring method applied to the water source indicator in relation to evaluation of supply types.

In light of this approach, it was necessary to review the literature relating to evaluation of household water treatment types according to microbiological parameters. However, no single research study has been identified which presents a suitable assessment of water quality in Egypt by addressing household water treatment. Instead, global studies were used to obtain relative risks for each treatment type. Relative risk is defined as "multiple of risk of the outcome in one group compared with another group and is expressed as the risk ratio in cohort studies and clinical trials" (Zhang &Yu, 1998, p 690). It is a common statistical method used in epidemiological research, mainly in disease incidence estimation (Kelsall & Wakefeld, 2002). Four studies were identified which calculate a relative risk of stored water being uncontaminated with indicator bacteria (versus contaminated) for each household treatment type. Thus, a relative risk of 1 indicates a treatment that does not improve stored drinking water quality, values greater than 1 indicate increasing effectiveness of the treatment types, and values less than 1 an adverse effect of the treatment on water quality

Table 5.3 Methodology for scoring water source types

Study	% safety	Evaluation parameters	Components	Indicator	Median Score	Aggregation
El-Taweel and Shaban (2001)	95	Microbiological E.Coli.	Tap water	Network coverage	**96%**	**Sum**
Diab (2002)	80	Microbiological E.Coli.	Tap water			
El Atar et al. (1982)	96	Microbiological E.Coli.	Tap water			
El Abagy et al. (1988)	100	Microbiological E.Coli.	Tap water			
Lasheen et al. (2007)	100	Chemical Analysis	Tap water			
Abo-Amer et al. (2007)	85	Microbiological E.Coli.	Pump	Underground water	**90%**	**Sum**
Soltan (1999)	90	Chemical Analysis	well			
El Abagy et al. (1988)	90	Microbiological E.Coli.	Unprotected well			
El Abagy et al. (1988)	95	Microbiological E.Coli.	Protected well			
El Abagy et al. (1988)	95	Microbiological E.Coli.	Tube well			
Saleh et al. (2001)	97	Chemical Analysis	Bottled water	Bottled water	**97.5**	**Sum**
El Abagy et al. (1988)	98	Microbiological E.Coli.	Bottled water			
Gupta et al. (2007)	86	Microbiological E.Coli.	Water tanker*	Tanker trucks	**86%**	**Sum**

(*) Searching the literature, no study has been found of water tankers in Egypt. Therefore, a study from Indonesia was used instead

This score was multiplied by the percentage of households using each treatment type to obtain the overall score for each geographic unit. Table 5.4 shows relative risk scores for the main five water treatment methods:

- Arnold & Colford (2007) undertook a meta-analysis of 10 studies for point-of-use drinking water treatment by chlorination. E. coli contamination of stored drinking water and raw water was used to calculate the overall relative risk which was 0.71. The reciprocal of this number (1/0.71) gave the final relative risk score for the chlorination treatment method (1/0.71= 1.41).
- A relative risk score for the use of ceramic filters was calculated from Du Preez et al. (2008). They calculated the number of stored water samples contaminated with E.Coli for control and intervention households in South Africa and Zimbabwe. The relative risk was 0.61 and then adjusted to be (1/0.61= 1.64) for the risk of safe, rather than contaminated water.
- Boiling water showed the highest relative risk (2.00). Calculation of this score was based on Clasen et al. (2008). They presented percentage faecal contamination levels in water from sources versus stored boiled water (18% boiled stored water and 59% water from sources). The relative risk of uncontaminated stored water is thus (100-18%)/ (100-59%) = 82%/41% = 2.00.
- In the absence of quantitative evidence, a relative risk for letting water stand to settle (1.20) was approximated from a report by Skinner & Shaw (2003) who describe the effectiveness of different household water treatment methods versus consuming raw water in relative terms. 1.2 was chosen as the value for this treatment because Skinner & Shaw (2003) describe letting water stand to settle as being more effective than no treatment at all, but less effective than boiling, filtration or chlorination.
- A relative risk value similar to use of a ceramic filter was given to solar disinfection treatment in the Skinner & Shaw (2003) study where both treatment types showed the same value in terms of effectiveness of treating pathogens.

Study	Relative risk	Evaluation parameters	Components	Indicator	Aggregation
Skinner and Shaw 2003	1.2	Microbiological E.Coli.	Let water stand to settle	Home water treatment	
Arlond and Colford 2007	1.41	Microbiological E.Coli.	Chlorination	Home water treatment	
Clasen et al.2008	2.00	Microbiological E.Coli.	Boiling	Home water treatment	Sum
Du preez et al.2008	1.64	Microbiological E.Coli.	Ceramic filter	Home water treatment	
Skinner and Shaw 2003	1.64	Microbiological E.Coli.	Solar disinfection	Home water treatment	

Table 5.4 Methodology for scoring household water treatment

Study	Water quantity (LPCD*)	Evaluation parameters	Components	Indicator	Aggregation
	30	1-5 minutes			
Howard and Bartram 2003	12	6-30 minutes	Time to get to water source	Water fetching	Sum
	5	31-< minutes			
	50	zero minutes			

Table 5.5 Methodology for scoring water quantity

Study	Score	Evaluation parameters (within the last fortnight)	Components	Indicator	Aggregation
Due to variable label and comparison between categories	1	Water was unavailable daily			
	2	Unavailable a few times	Frequency of water unavailability	Water interruption	sum
	7	Unavailable less frequently			
	25	Water available			

Table 5.6 Methodology for scoring water interruption

5.3.3 Household water fetching indicator

This indicator was based on the only available data from the DHS. Fetching water and the time taken to the water source are considered an important indication of safe water accessibility. If much time is needed to reach the water source, this will influence water quantity. Howard & Bartram (2003) studied the relationship between household water availability, accessibility and quantity per capita per day. They reviewed many studies of water consumption and trip distance, measured in terms of the time taken to fetch water, and graphed trip travel time (in minutes) versus water quantity (litres per capita per day). This relationship was used to score the water fetching indicator for availability in cubic litres per capita per day (CLPCPD) based on time spent fetching water. Time taken to fetch water was recoded from the DHS into four categories and each category was assigned a water quantity score. For instance, for households spending from one up to five minutes fetching water it was estimated that 30 litres per capita per day would be used (see table 5.5)

5.3.4 Household water interruption indicator

Data about water supply interruptions are only available in the DHS. As noted previously, water supply interruptions cause several problems such as consuming less water quantity, fetching unsafe water from unimproved sources, decreasing piped supply pressure and storing water (see section 2.2.3). These impacts significantly affect public health. The scoring of the water interruption indicator attempted to capture all these effects by translating the four DHS answer categories relating to frequency of interruption during the last two weeks (see table 5.6) into a measure of the frequency of interruption (measured as an approximate time interval in days between successive supply interruptions):-

- Water unavailable daily. This category was scored 1 (i.e. 1 supply interruption every day).
- Water unavailable a few times. This category was scored 2 which indicates that a water interruption event approximately once every 2 days.
- Water unavailable less frequently. This category was scored 7 indicating an interruption once every week.
- Water was available. It is hard to assign a numerical value to this category as the lack of interruption in the last fortnight does not mean that the supply was never interrupted in the longer term. A score of 25 has been assigned in this case.

5.3.5 Sanitation type indicator

Data about sanitation, like water, are available in both the DHS and census. Sanitation types from the census were chosen for inclusion in the SDWI because of the advantage that the census covers the entire country and provides representative estimates at the kism/markaz level. However, the questions in the DHS and census data sources are different and thus sanitation types are described differently. In addition, the JMP uses a third different classification system. The classification criteria in each type are different from one another:-

- The DHS classifies sanitation according to both toilet facilities and drainage. The DHS toilet facility categories are modern flush toilet, traditional (tank flush), traditional (bucket flush), pit latrine/bucket toilet, other/missing, no facility. The DHS drainage categories are public sewer, vault (*Bayara*), septic system, pipe to canal, pipe to groundwater, emptied (no connection), other and no toilet facility.
- The JMP classification system is a mixture of both drainage and toilet facilities (WHO/UNICEF, 2010). The JMP improved sanitation categories are flush toilet, piped sewer system, septic tank, flush/pour flush to pit latrine, ventilated improved pit latrine, pit latrine with slab, composting toilet and special case while the JMP unimproved sanitation categories are flush/pour flush to elsewhere, pit latrine without slab, bucket, hanging toilet or hanging latrine and no facilities or bush or field. The JMP classifies sanitation facilities into improved and unimproved based largely on the risk of faecal contamination from flies. However, this sort of classification takes little account of groundwater contamination.
- The Egyptian census classifies sanitation categories based on whether household sanitation is connected to a public network or not. Difficulties arise, however, when an attempt is made to match census sanitation types to the DHS and JMP because the sanitation categories used only partially match. In particular, not all census sanitation types are documented and the census question form is not available, even for academic purposes (CAPMAS, 2010). Reviewing the literature, it was not possible to find any scoring system that matches this type of Egyptian sanitation. Thus, the best way for scoring these sanitation types is to compare the percentage of each type in the DHS and census. Following the 2005 EGYPTIAN DHS final report, It was concluded that 91.3% of Egyptian households are connected to improved sources (sewer, Bayara or vault and septic system) while 8.7% are connected to unimproved

sanitation (El-Zanaty & Way, 2006). On the other hand, 3.7 (641,344) % of households are offline (Not connected) according to the 2006 Egyptian census. Since it proved impossible to match the census *'tranch'* and 'local network' sanitation types to equivalent JMP or DHS sanitation categories, all census sanitation types except for 'offline' were considered improved (public network, tranch and local network). This somewhat overestimates the percentage of households with improved sanitation, but proved the most pragmatic approach given the definitional mismatches between data sources.

Sanitation Types (Census)	**Parameters**	**Sanitation Score**	**Indicator**	**Aggregation**
Public sanitation	Improved	1.00	Sanitation system	Sum
Tranch sanitation	improved	1.00		
Local sanitation	Improved	1.00		
Offline	Unimproved	0.00		

Table 5.7 Methodology for scoring sanitation type

The MDG criteria (improved and unimproved sources) were thus followed approximately to score each census sanitation type (see table 5.7). A value of one was given for each improved type and a value of zero for each unimproved type. Four types of sanitation from the census were classified and scored as the following:

- Households connected to public network (1).
- Households connected to *tranch* (1).
- Households connected to local network (1).
- Households offline (not connected) (0).

5.4 Assessing the distribution of scores

To investigate variations in the spatial distribution of the indicators, all scores were graphed, aggregated by district (kism or markaz). Producing a histogram for each indicator score can be used to assess the characteristics of the frequency distributions. Looking at all histograms of indicator scores, their distributions can be interpreted as follows:

5.4.1 Normal distribution

Figure 5.3 shows a symmetrical histogram of the educational indicator (population group who read and write). It can be seen that the histogram represents a normal distribution where most of the score values occur around the mean (highest bar) and there is a gradual decrease in the frequency towards the left and right. A histogram of the water interruption indicator score (figure 5.4) demonstrates more or less a normal distribution where the frequency is high in the middle of the scores and decreases approximately towards the left and right.

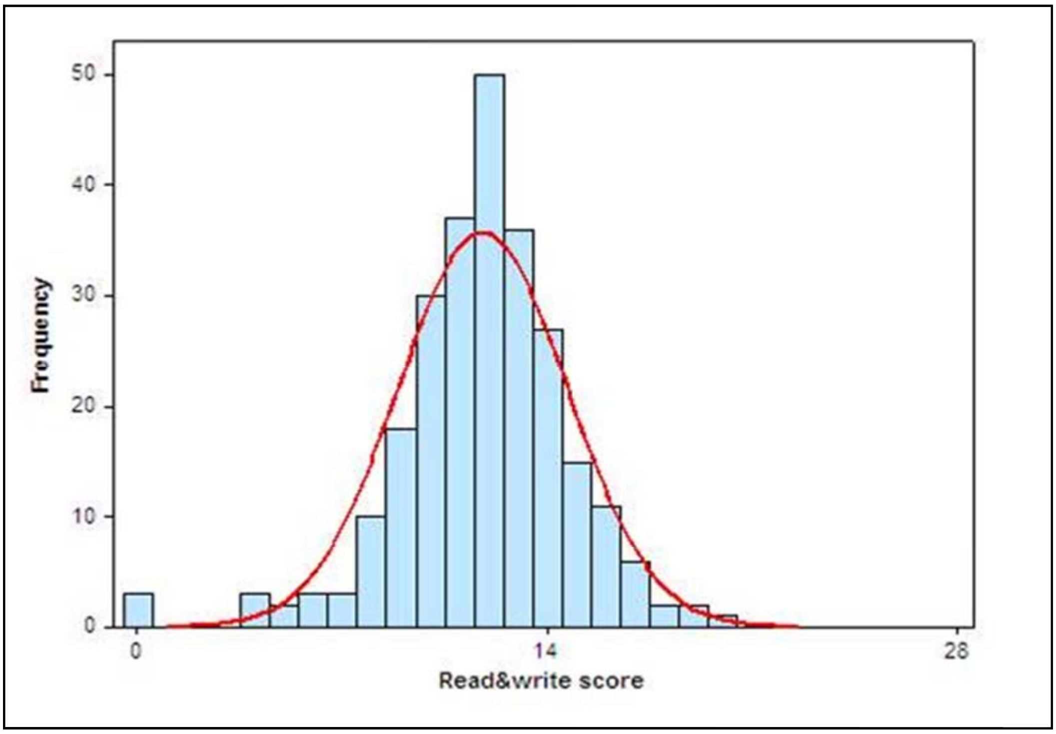


Figure 5.3 Histogram of educational indicator score

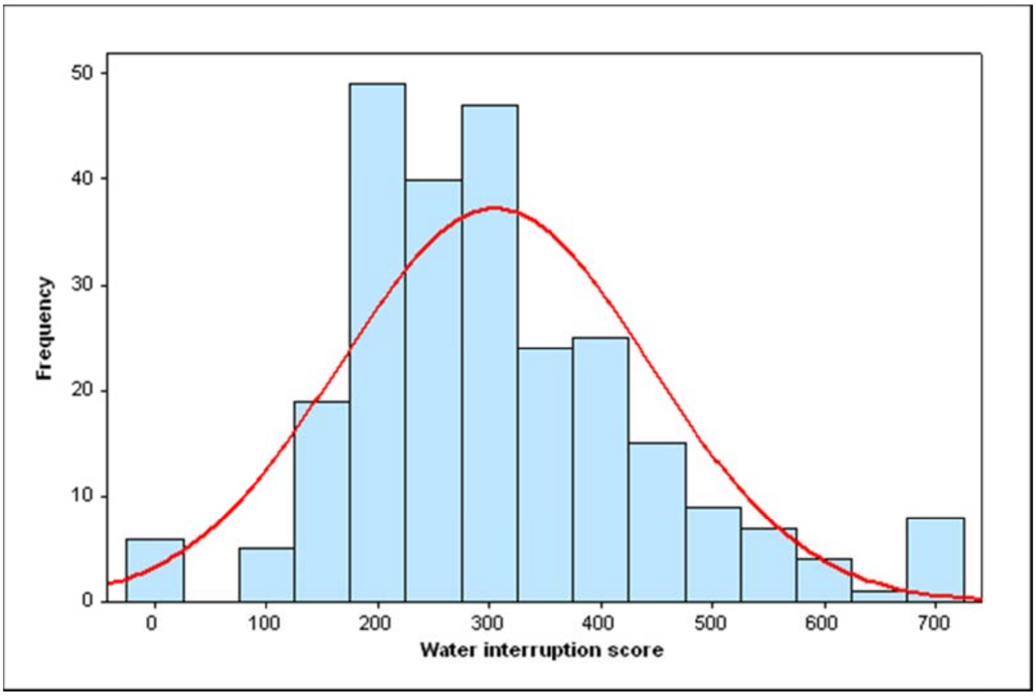


Figure 5.4 Histogram of water interruption score

5.4.2 Bimodal distribution

The wealth indicator displays a bimodal distribution (figure 5.5). This type of histogram is distinguished by two peaks (modes) of the distribution. It can be seen that the two peaks are related to the central tendency for Egyptian household wealth classes.

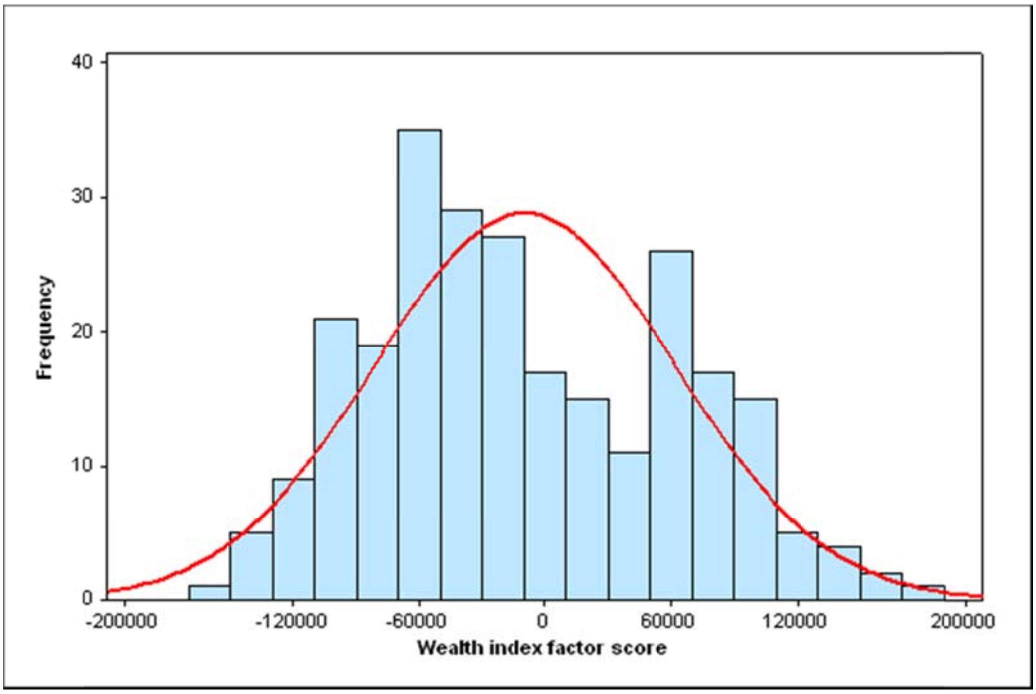


Figure 5.5 Histogram of household wealth component

5.4.3 Left-skewed distributions

Figures 5.6 and 5.7 show histograms of the population density and home water treatment indicators. The two indicators exhibit a left-skewed distribution where most of the score values are close to zero. The histogram of water treatment scores was displayed using the theoretically possible range of scores (0.00 to 200, the latter being the relative risk of 2.0 multiplied by 100% uptake. The observed scores only ranged 0 - 40). On the other hand, the histogram of population density represents the actual spatial distribution (0.00 to 80,000). In the two histograms, the peak is at the far left which indicates low scores. With regard to population density, the majority of subprovincial spatial units (kism/markazes) tend to have low population density (between 0 and 4,000 persons per km^2), particularly in Frontier and Desert governorates. Lower score values of home water treatment can be interpreted as lower proportions of Egyptian households who treat their drinking water.

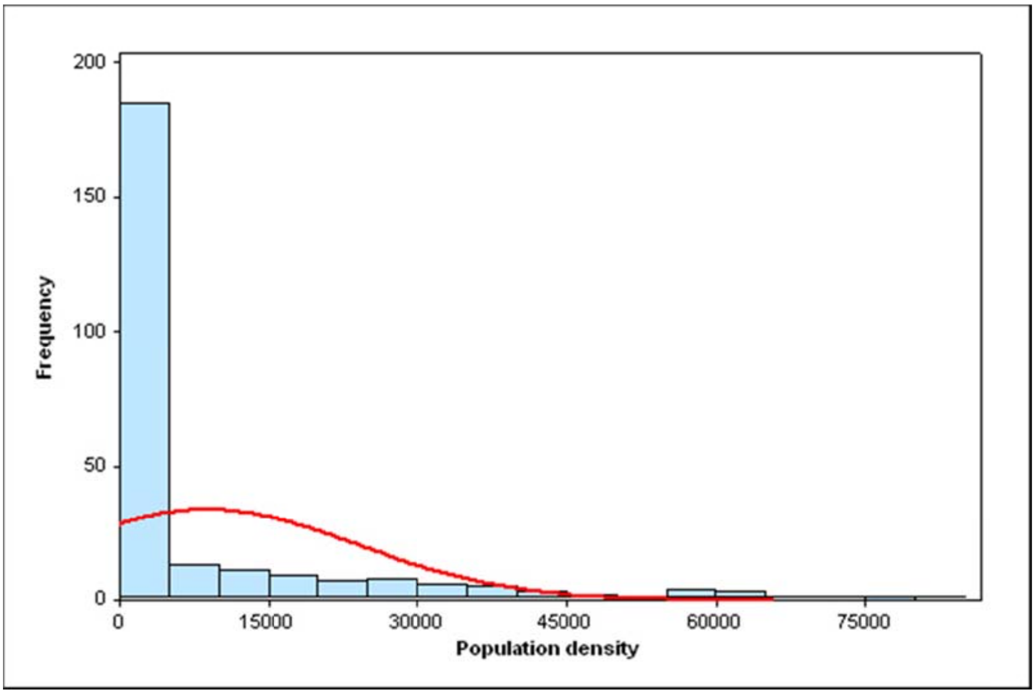


Figure 5.6 Population density score

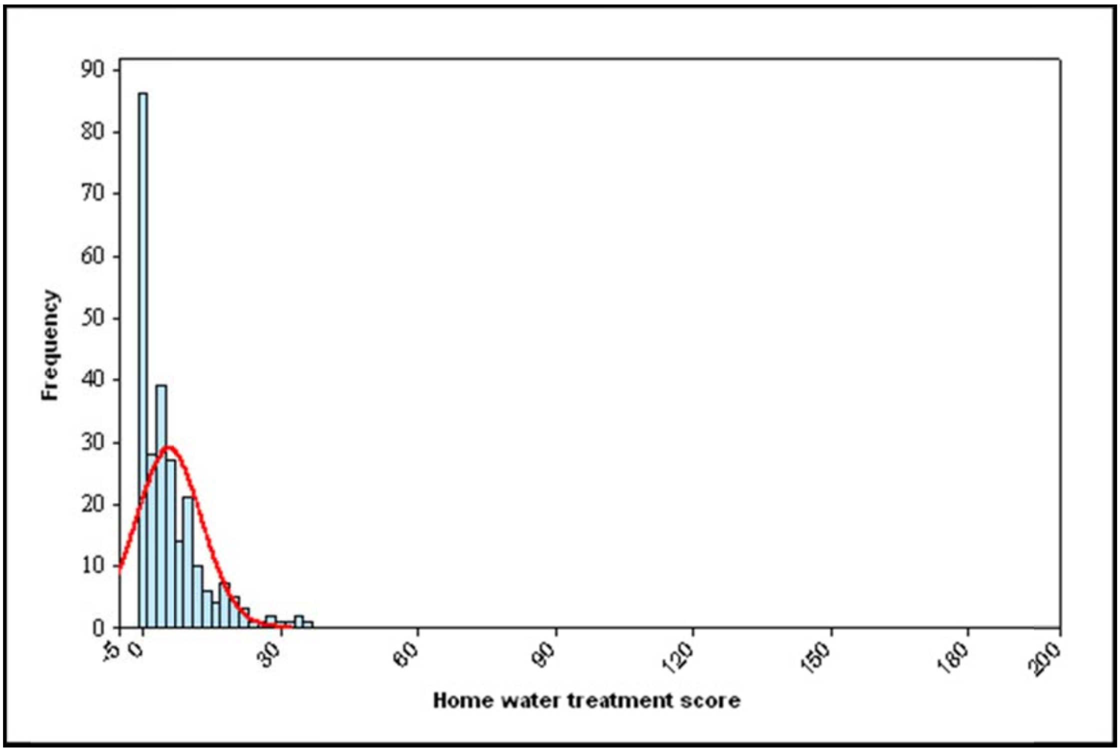


Figure 5.7 Histogram of home water treatment (theoretical range)

5.4.5 Right-skewed distributions

There are four indicator scores (water quantity, water sources from the DHS, water sources from the census and sanitation system) with a right skewed distribution since the right side of each histogram contains most of the observations (figures 5.8 to 5.11). The water quantity indicator was displayed using the theoretically possible range (5 to 50 which are the minimum and maximum water quantities in litres per capita per day calculated according to time taken to reach water source). The theoretically possible range for the sanitation type score is the same as the observed actual range of the data.

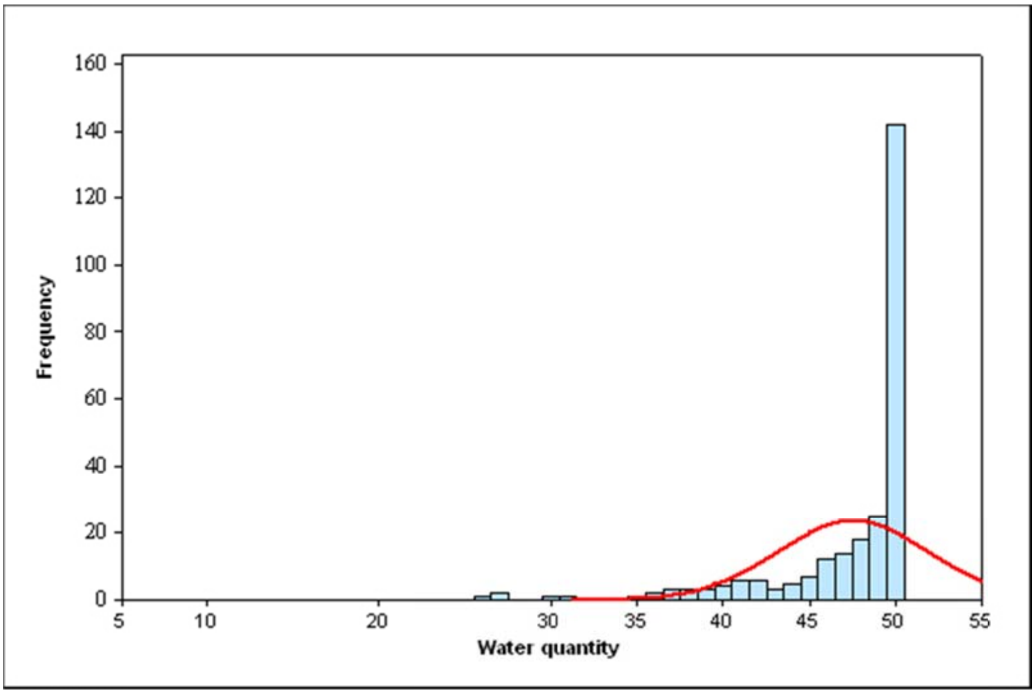


Figure 5.8 Stretched histogram of water quantity score

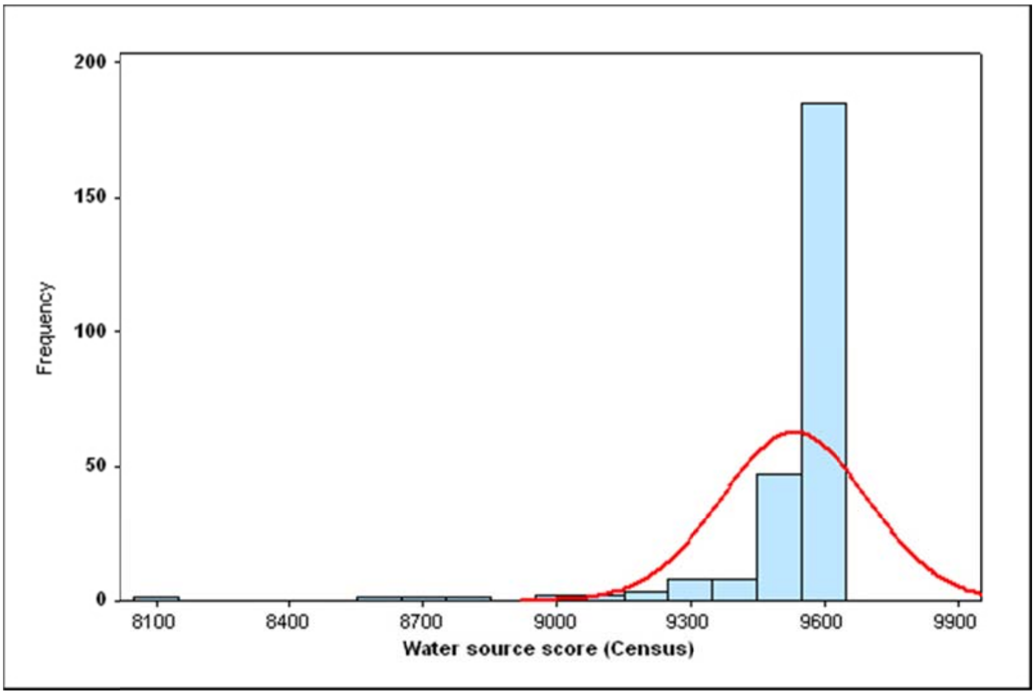


Figure 5.9 Histogram of water source type (census) score

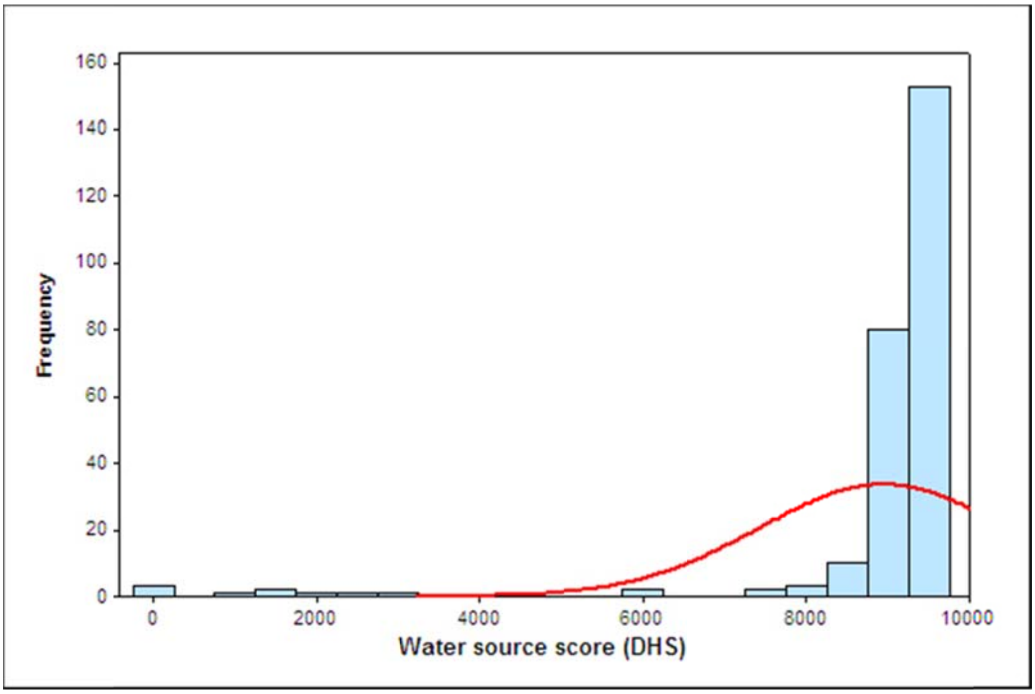


Figure 5.10 Histogram of water source type (DHS) score

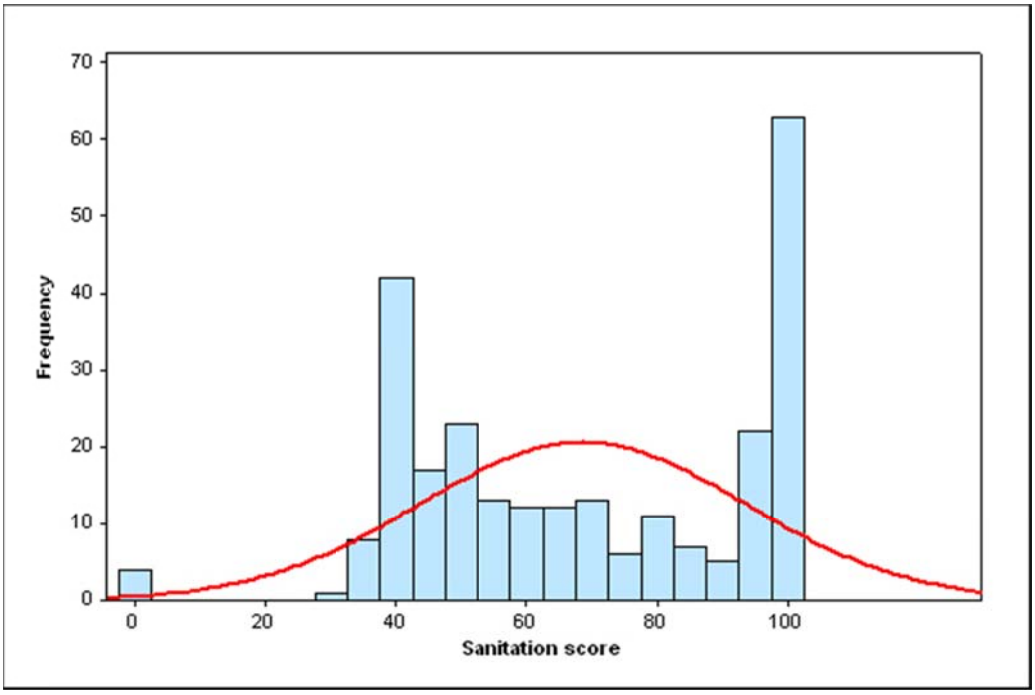


Figure 5.11 Histogram of sanitation score

5.5 Correlation matrix

The Pearson correlation coefficient has been used to measure the degree of association between each pair of variables. Thus, the objective is to investigate how variables co-vary where the covariance concept is the average value of the product of two variables (Rogerson, 2006). It was important to examine the validity and complementary of each indicator and ensure that it does not overlap extensively with other indicators. Hence, a correlation matrix was created by calculating the Pearson correlation coefficient between all the scored indicators. The correlation matrix (table 5.7) indicates no strong positive or negative relationship between the indicators. In general, indicators such as population density, educational status (ability to read and write), water supply types (in both the DHS and census) and water interruption show very weak or no association with other indicators ($r < \pm 0.4$). However, weak relationship were found between water source types (DHS) and water quantity ($r = 0.504$), home water treatment and wealth ($r = 0.541$), water quantity and wealth ($r = 0.422$). The correlation of water treatment and wealth is similar to the reported correlation found by Gundry & Wright (2009).

Variables		Population density	Read and write	Wealth index	Water quantity Score	Water types (DHS)	Water types (census)	Water interruption	Water treatment	Sanitation
Population density	Pearson	1.00								
	Sig. (2-									
	N									
Read and write	Pearson	-0.05	1.00							
	Sig. (2-	0.42								
	N	259.00	259.00							
Wealth index	Pearson	-0.02	0.01	1.00						
	Sig. (2-	0.81	0.83							
	N	259.00	259.00	259.00						
Water quantity score	Pearson	-0.07	0.00	.422**	1.00					
	Sig. (2-	0.29	1.00	0.00						
	N	259.00	259.00	259.00	259.00					
Water types (DHS)	Pearson	-.167**	0.05	.273**	.504**	1.00				
	Sig. (2-	0.01	0.40	0.00	0.00					
	N	259.00	259.00	259.00	259.00	259.00				
Water types (census)	Pearson	0.11	0.07	0.03	-0.05	-0.10	1.00			
	Sig. (2-	0.08	0.27	0.60	0.45	0.13				
	N	259.00	259.00	259.00	259.00	259.00	259.00			
Water interruption	Pearson	0.03	0.10	.256**	.158*	0.00	-0.02	1.00		
	Sig. (2-	0.70	0.13	0.00	0.01	0.97	0.70			
	N	259.00	259.00	259.00	259.00	259.00	259.00	259.00		
Water treatment	Pearson	-0.08	0.04	.541**	.210**	0.11	0.04	0.05	1.00	
	Sig. (2-	0.22	0.56	0.00	0.00	0.09	0.48	0.40		
	N	259.00	259.00	259.00	259.00	259.00	259.00	259.00	259.00	
Sanitation	Pearson	.201**	-.292**	-0.03	-0.08	-0.08	.318**	-0.02	-0.06	1.00
	Sig. (2-	0.00	0.00	0.66	0.19	0.18	0.00	0.72	0.37	
	N	259.00	259.00	259.00	259.00	259.00	259.00	259.00	259.00	259.00
**. Correlation is significant at the 0.01 level (2-tailed).										
*. Correlation is significant at the 0.05 level (2-tailed).										

Table 5.8 Correlation matrix for the Safe Drinking Water Index component scores

5.6 Standardisation

An index contains many indicators and components that are relevant to the index topic and objectives. However, the indicators are usually measured in different units. Thus, standardisation is a common method used to convert a group of variables or indicators into common measurement units so they can be combined (Senior, 2006).

5.6.1 Standardization methods

There are many possible standardisation methods for indicator scores. However, the most common methods are Z score, ranking, chi squared and range standardisation:-

Z score method

This is a widely used method in data standardisation. A Z score compares each case X¡ to the mean (*μ)* and then divides by the standard deviation SD. However, it works best when the data are approximately normally distributed.

Ranking method

This method was used in the Index of Multiple Deprivation 2000 (Joradn et al. 2004; Noble et al. 2000). It is the simplest standardisation method to apply. The method is based on giving a rank to each unit on each variable or indicator in the study. The main advantage of this technique is its resilience to outliers. The loss of information on the absolute levels is considered to be one of its disadvantages (Nardo et al. 2005).

Signed chi-squared

This method was used to standardise the 1991 Index of Local Conditions. This method solves the problem when the standardisation is based on a small denominator which often occurs for small spatial units. Signed chi-squared incorporates absolute values and references rates by local denominators (Senior, 2006).

Range standardisation

This method compares each value X_i to the X_{min} value of the indicator and divides by the difference between the maximum X_{max} and the minimum X_{min} of the indicator. Range standardisation has its disadvantage: when the data contain outliers it is less effective. A decision was made to use range standardisation for the following reasons:-

- Looking at the histograms of all components, it can be seen that, except for the educational status and water interruption components, all other components are not normally distributed. Therefore, the Z score method would not be valid to apply without initial transformation of the raw data.
- The data distribution for some indicators (e.g. sanitation; water source) showed gaps and thus using the ranking method would lead to loss of information about absolute levels of water safety for such scores.
- Range standardisation is quite useful in terms of maintaining the actual score distribution. Moreover, the technique is easy to interpret and implement.
- Since the kism and markaz are relatively large units in population terms, the signed chi squared method also appeared inappropriate because it is designed for small areas.
- In some cases, the distribution of indicator scores covered the entire theoretically possible range of values, such as the sanitation indicator (0.00 to 100). However, for some other indicators, only a small part of the theoretically possible range of scores was observed. In particular, home water treatment is a rare behaviour in Egypt and a maximum score of 40 was observed compared to the theoretically possible maximum of 200. There is thus a danger of inflating

the importance of rare behaviours during standardisation of scores. Using a theoretically possible maximum and minimum value for range standardisation for some indicators, rather than the observed maximum and minimum, is useful to avoid this effect.

5.6.2 Results of standardisation process

All scored variables were range standardized. Figure 5.12 shows the variables in the SDWI standardised using the range of theoretically possible values. It can be seen that the distribution of the standardised variables varies from normal to skewed distributions, as did the unstandardised scores. Three variables (water interruption, wealth index and sanitation) have approximately symmetric distributions while the other variables have either left or right skewed distributions. The observed range of values of the water treatment, educational status (ability to read and write) and population density variables tend to be narrower relative to the theoretically possible limits, compared with other variables. Water treatment values are 0 to 0.15; educational status values are 0 to 0.20 while population density values are 0 to 0.80 with a long tail towards the right side. Therefore, these variables exhibit a left skewed distribution. On the other hand, three variables, water sources (DHS), water sources (census) and water quantity display more high values and thus they are right skewed distributed. Water source (DHS) shows the highest observed range of standardised scores 0.09 to 1.00 while the range of standardised scores for water sources (census) is 0.20 to 1.00. Values of water quantity range from 0.55 to 1.00.

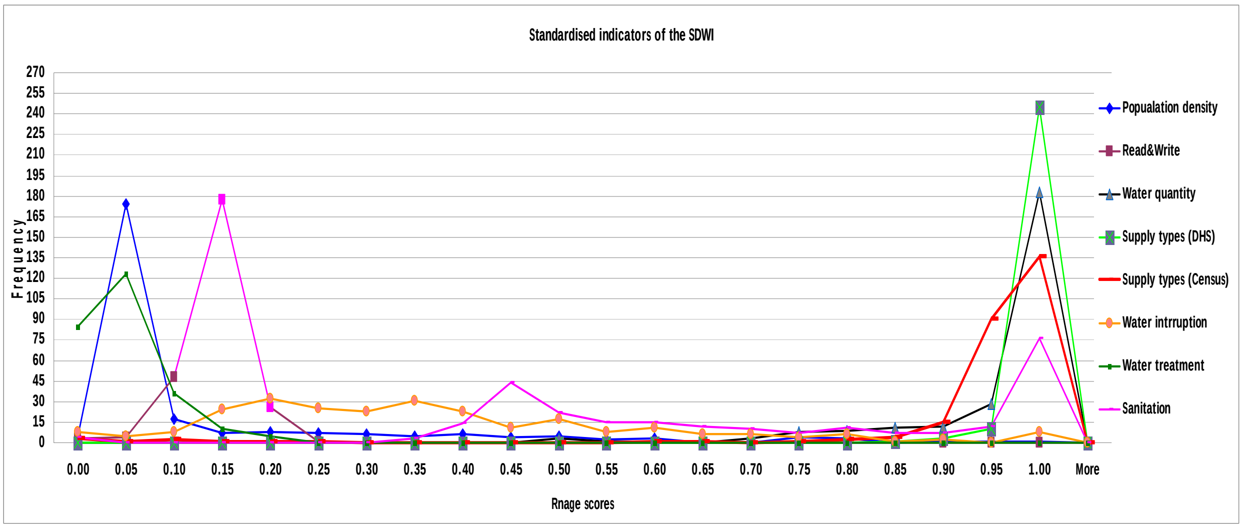


Figure 5.12 Scores for the safe drinking water variables, standardised using the theoretical range for each variable

5.7 Spatial distribution of the SDWI in sampled districts

Figure 5.13 shows a map of the SDWI values that are standardised, weighted, combined and then represented in those kisms/markazes sampled by the DHS. In general, it can be seen that the index of the SDWI for the sampled Egyptian kisms/markazes shows the following geographic patterns:

Desert districts (both urban and rural) illustrate the lower and middle values of safe drinking water access. However, El Wahat El Kharga and El Wahat El Edakhla in New Valley governorate and Bear El Abd in North Sinai show higher values compared with other desert kisms/markazes. This is probably because the majority of households live in urban settlements and higher percentages of households have tap water inside houses or flats which are connected to public networks. Middle Egypt (as a part of Upper Egypt region), particularly the markazes of Helwan, El Fayoum, Bani Swef, El Menia, and Assuit governorates have lower index values. Although the drinking water uptakes in these governorates are closer to the Nile River (the main source of drinking water in Egypt), higher percentages of households construct pumps and use wells. In addition, there is a higher percentage of water interruption and households fetch water from long distances particularly in rural villages. Most of the public networks and purification plants are old and need replacement and renovation. In the south of Upper Egypt kisms and markazes of three governorates (Sohag, Qena and Asswan) especially kisms and markazes located on the west bank of the Nile River show lower values of SDWI compared to other Valley governorates.The areas located on the east bank of the Nile River appear in the middle classes of accessibility.

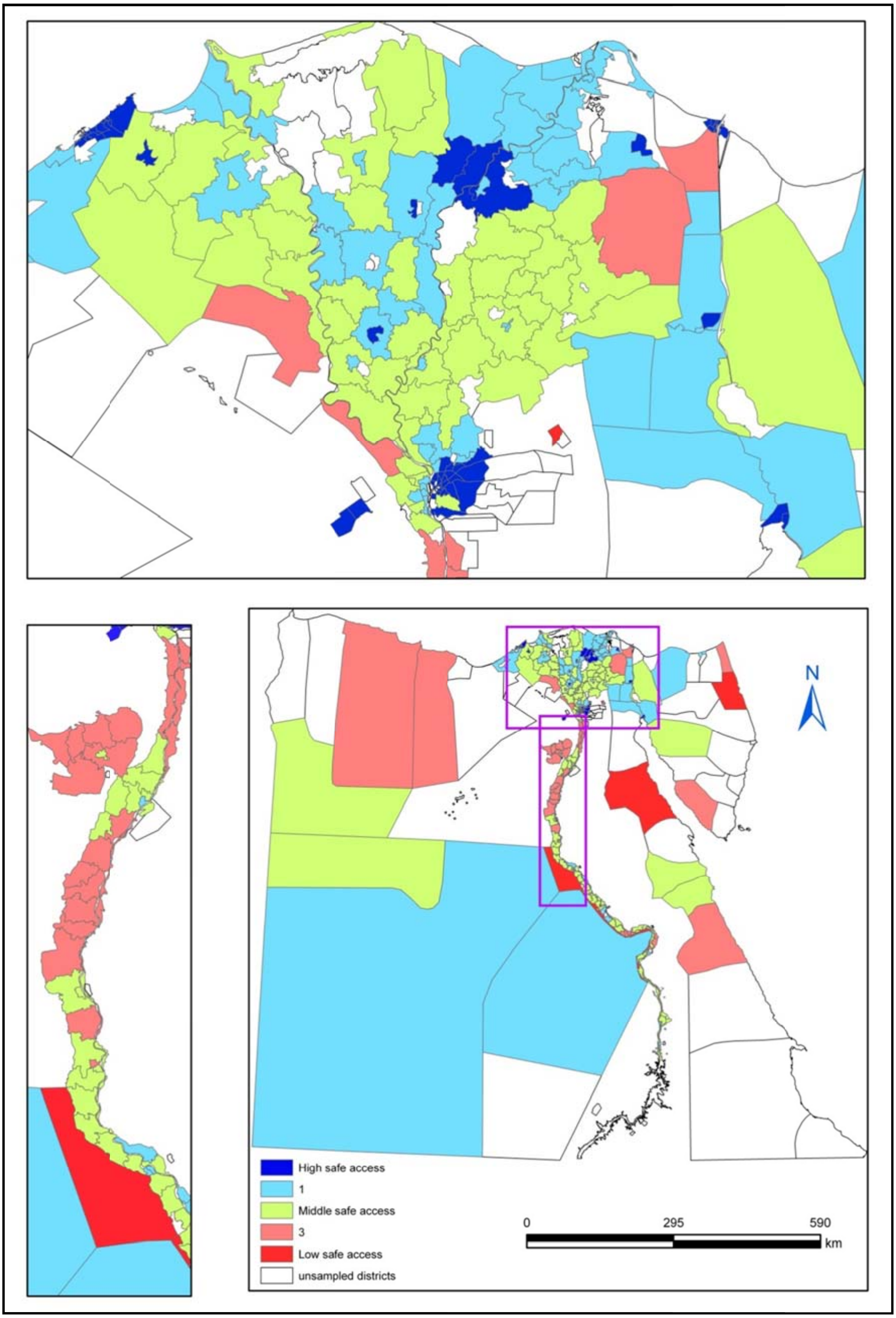


Figure 5.13 Distribution of SDWI values across Egyptian districts that contain DHS GPS clusters. Class boundaries are based on the quintiles for SDWI values

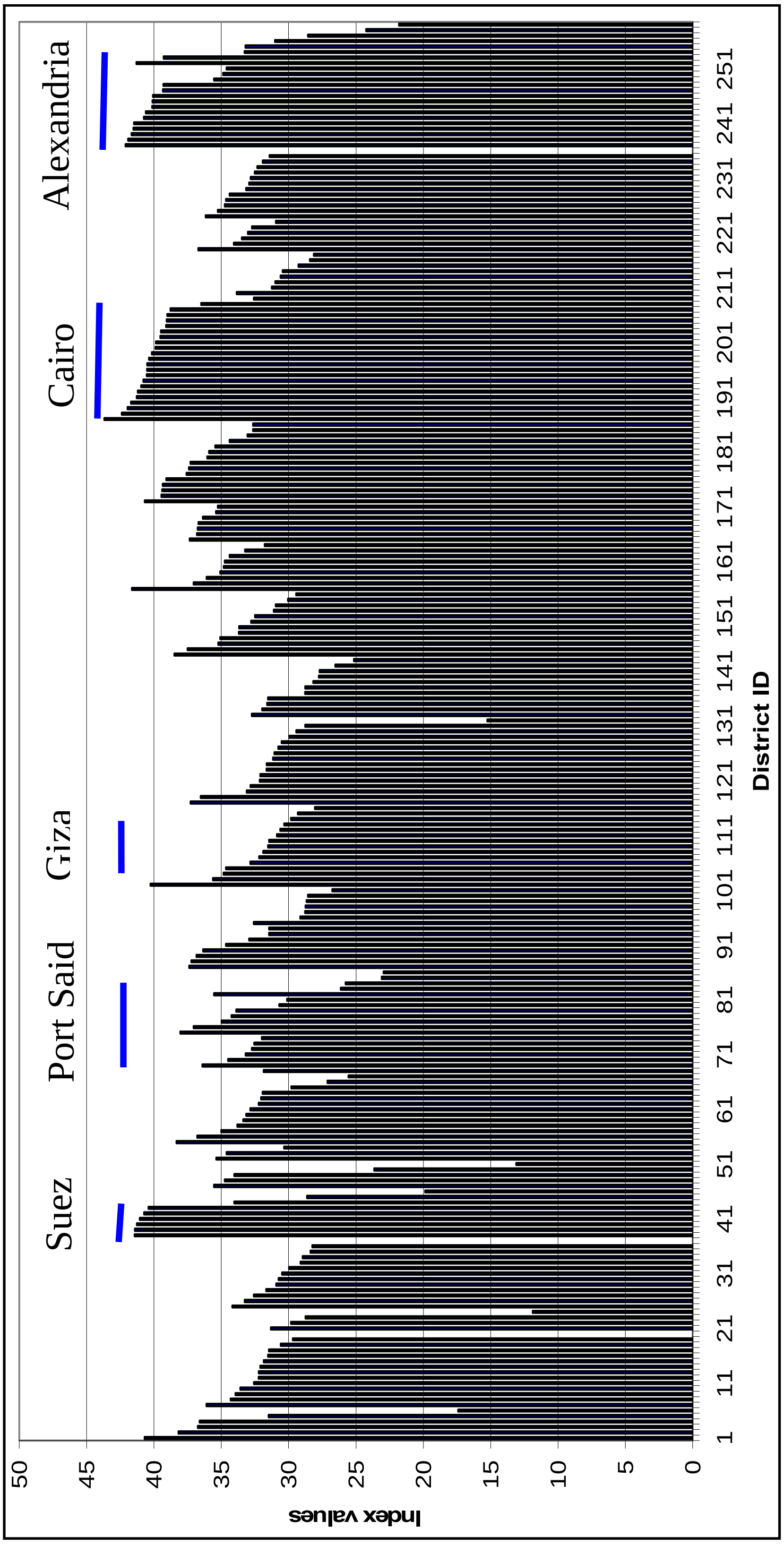


Figure 5.14 Safe drinking water index values for Egyptian districts (kism/markaz) with blocks of values highlighted for urban governorates

Figure 5.14 illustrates the index results at district level (kism or markaz), sorted firstly by governorate and secondly by index value. It can be seen that the urban centres of Cairo and Alexandria have the greatest values for the SDWI of any Egyptian governorate. The three other wholly urban governorates of Giza, Port Said and Suez have somewhat lower SDWI values more comparable with rural governorates. The lower values in Suez are likely to result from water supply difficulties in land reclaimed from desert. It is also apparent that there are quite large variations in safe water access within each of these major urban centres. It can be seen that districts that are located in the eastern desert and Sinai have low values compared with other desert districts in the western desert (Matrouh and New valley) such as El Ksimaa (North Sinai governorate), Rass Gareb (Red Sea governorate), El Toor (South Sinai governorate) and Rafh (North Sinai governorate). Although districts in both regions (western and eastern deserts) are not accessible to Nile water, higher values in the western desert district can be investigated as better drinking water accessibility associated with richness in underground water aquifers, particularly Oasis wells in New Valley and Matrouh governorates. These aquifers are geologically parts of the Nubian Sandstone Aquifer System (NSAS) which is saturated with water originating from equatorial rainfall zone (Hefny et al. 1992).

5.8 Summary

This chapter has described the scope and domains of the SDWI. The final index consists of five domains: water, environmental, socioeconomic, demographic and household educational domains. Each domain contains one or more indicators that were reviewed for potential inclusion in the index. Seven specific criteria for selection were introduced (i.e. water related, valid and specific, available and achievable, plausible and complementary, longitudinal and repeatable, measurable and reliable, and demonstrable and mappable) and each indicator was given a score. For instance, in terms of valid and specific criteria, the water supply indicator scored higher while educational status of household scored lower. The chapter also has provided a detailed description of how each indicator was scored through developing a scoring system by which indicator contributions can be identified and distinguished from the others. Since water supply types are the most effective contributors to the overall index, focus was given to all Egyptian publications that analysed water sample contamination by *E.coli* bacteria and faecal streptococci. From each study an overall safety score was obtained. A score of 100 indicated that the water source was consistently safe while 0 reflected that the source was inconsistently unsafe. Other indicators such as water treatment, water quantity and water interruption had different scoring systems. While each water treatment method (i.e. boiling and using filter ceramic) was scored based on its influence on water quality, the water quantity indicator was scored according to time taken to fetch drinking water from the source. Another section in this chapter has covered assessment of the scoring methods and approaches involving consideration of spatial distribution. Correlation matrix was used to examine overlap between indicators and showed no strong positive or negative relationships between the selected indicators.The standardisation method has been clearly explained and the justification for using range standardisation described. The last section of this chapter has presented the spatial distribution of the index values across sampled kisms and markazes and described geographic patterns and characteristics of this distribution.

Chapter 6: Estimating SDWI values for all Egyptian districts

6.1 Overview

We now have SDWI values for kism/matkazes sampled by the DHS and for which all DHS and census indicator are available.This chapter initially discusses the conceptual methodology behind estimating the unknown values of the unsampled DHS districts to construct the SDWI for the whole of Egypt. The first section of this chapter reviews the literature on small area estimation and commonly used methods to generate such estimates. It then reviews areal interpolation methods specifically and ends by describing the interpolation problem of this particular research. The second section focuses on the method used to interpolate the unknown values and develop the SDWI. The section also describes the sampling design of the DHS and the importance of taking it into account when the unknown values are estimated. This is followed by an explanation of Tobler and Kennedy's (1985) interpolation method and how it was applied within an overall GIS framework. Finally, the last section presents the results of the interpolation process and the map of the SDWI for the whole of Egypt at sub-provincial level. It describes and explores the factors that contribute towards the geographic patterns of safe drinking water accessibility among Egyptian districts and regions in terms of the index values.

6.2 Review of methods for estimating safe water access in unsampled districts

6.2.1 Small area estimation review

Much literature has been published on small area estimation describing, the role of this method in analysing spatial variations across socioeconomic domains. Ghosh and Rao (1994) defined a small area as a small local geographic unit or census division such as a county, district or municipality. A small area could also be a small domain, for instance a subpopulation group (e.g. the 15 to 25 year age group of population) within a geographic area. Sample surveys are considered an effective and useful means to derive information on different socioeconomic and demographic domains. Most sample surveys are used as estimators and provide reliable estimates for either the total population of interest or subgroups of the entire population community (Rao, 2003). Pfefferman (2002) investigated the important role of small area estimation as a result of the growing demand for small area statistics which are usually obtained from the sample surveys conducted for these areas or domains.

Applications

Saei & Chambers (2003) reviewed the literature on methods of small area estimation and described it as a technique used to produce direct and accurate estimate for small areas or sub-national domains of population. Small area estimation has been used in a wide variety of applications. Singh et al. (2002) used small area estimation to estimate crop yield in India based on remote sensing and crop yield surveys datasets. To target and analyse the distribution of poor people by administrative area and produce a poverty map, Simler & Nhate (2005) studied the spatial distribution of poverty and inequalities in Mozambique using small area estimates. Two datasets, one from a 1996 national household survey and the other from a 1997 population census were combined. Household survey data were then used to estimate the relationship between a dependent variable (welfare status) and a group of independent variables in a regression equation. This was followed by applying the estimated regression coefficients to census data to estimate consumption values per capita for each household in the census. Simpson & Tranmer (2005) conducted small area estimation by combining sample and census datasets using the Iterative Proportion Fitting (IPF) technique. They estimated a two dimensional table of car ownership by tenure using data from the 1991 UK census assuming an independent relationship between the two variables.

Methods of small area estimation

Although there has been an increasing demand for small area estimation in planning and decision making, the main purpose of producing small area estimates is when there are no available data at the fine level that will serve the required objective. The principal methods of small area estimation were summarised by the Australian Bureau of Statistics as the following (ABS, 2006):-

- **Simple small area methods**

This method includes three simple methods: the direct estimator, broad area estimator and calibration estimator. They do not require a high level of statistical knowledge to be implemented and they are used widely by government agencies to produce rough small area estimates:-

1. **Direct estimator**

This method produces direct estimators that are obtained after applying survey weights to the sample units in each small area. Most surveys are designed to provide estimates at national level or higher geographic unit such as province or state. Therefore, small area estimation with higher accuracy is produced using other kinds of techniques.

2. **Broad area ratio estimator**

This kind of estimation is calculated by pro-rating a broad area direct estimate based on a small area to broad area population ratio. The method works by applying the reliable broad area estimate proportionately to all small areas that are included in the broad region. However, there are some requirements that need to be considered to produce successful estimates. For instance it is important for the broad area to be large enough to afford a direct estimate but small enough that all small areas are homogenous.

3. **Calibration estimator**

As described by Chambers (2005) in this method the original survey weights are adapted into new values and calibrated on some available auxiliary variables. The small area estimate for this method should agree with the known population total. It is a simple and straightforward method but the auxiliary variable has to be chosen carefully and to be related to the variable being estimated.

- **Regression methods**

Regression methods are an effective means applied when a higher level of accuracy is needed for small area estimation. They differ from the previous methods since they require a higher level of statistical knowledge. In general these methods are divided into two models:

1. **Synthetic regression models**

The main idea behind this kind of regression is to use available auxiliary data to develop a particular relationship between the auxiliary variable and variable of interest (response) which needs to be predicted. Synthetic regression models are based on the assumption that the systematic spatial variability in the response variable is clearly explained by the variability of auxiliary variables. The difference between the predicted value for the response variable and the observed value of the data is called a residual or random error. This kind of model is useful and valuable when the auxiliary variables are accurate and available.

2. **Random effects regression models**

In synthetic regression models, residuals usually show considerable variation between areas which indicates that there is another systematic variation in the response variable and among different areas. This means that the synthetic model is missing some auxiliary variables. To solve this kind of problem, a random effect is added into the model. Consequently, the constant in the model is treated as fixed, plus the random effect component. The advantage of including this component is the production of more precise and robust estimates. However, random effect models are more complex than synthetic models because of their technical nature.

6.2.2 Review of areal interpolation methods

Spatial interpolation is a term which refers to an estimation process to predict unknown values of some locations based on known values of surrounding locations in a study area, following a mathematical procedure. There are some specific terms related to interpolation methods such as pycnophylactic, kriging, spatial, point and areal interpolation. In addition, there are different kind of deterministic techniques for interpolation (see table During the past three decades these terms were defined in the literature (e.g. Tobler, 1979; Goodchild & Lam 1980; Lam, 1983, Flowerdew et al. 1991). Pycnophylactic interpolation refers to an interpolation method which was introduced by Tobler (1979) mainly for volume preservation and isopleth mapping. Volume preservation means preserving the total value of a variable within each source zone and it is highly desirable for areal interpolation as the estimated values of the target zones are less subject to error as a consequence (Lam, 1983). Kriging is the earliest spatial interpolation method introduced by G. Krige in the late 1970s. It refers to the prediction of unknown values according to statistical models assuming that the spatial surface has stationarity and can be described statistically. Point interpolation is a method used to estimate unknown values of sample point locations based on known points within the study area. Areal interpolation is a mathematical procedure which interpolates values of one or more variables for a set of polygons (target zones) based on known values (source zones) for the same variable in the study area.

Lam (2009) addressed spatial interpolation methods emphasizing that there is a continuous and increasing use of interpolated surfaces in many spatial and socioeconomic applications, particularly cartography, climatology, geology, population density and public health. Additionally, the development and growth of advanced GIS tools has produced many interpolation and estimation methods and made them more popular (Mitas & Mitasova, 1999). Interpolation methods were classified by Lam and Goodchild into point and areal interpolation according to their spatial nature (Goodchild & Lam, 1980; Lam, 1983):-

- Point interpolation is used when the dataset is comprised of points, for instance socioeconomic survey points, and is divided methodologically into exact and approximate methods. In exact interpolation, the values of all known points are presented. Exact interpolation of point data is preferred when the original sample points are more accurate, while approximate methods are highly recommended when the collected sample points show some noise. Another classification for point interpolation is based on the nature of the spatial operation (global or local methods). In a global method, all sample points are used to predict the unknown values while in a local one only the surrounding (neighbouring) points are used.

Deterministic techniques for interpolation	Definition	How it works	Suitable for datasets	limitations
Inverse distance weighting (IDW)	An exact interpolator used to predict an unknown value for a location based on its neighbors' measured values.	It works based on the assumption that closer phenomena are more similar than those that are further away. A weight is given to each surrounding measured value according to how far is it to the prediction location. As the distance increases the weight decreases.	When the spatial surface of the dataset shows local variations. Also when the measured points are evenly distributed throughout the area	It does not work well when the measured values are clustered.
Global polynomial interpolation	An inexact interpolator used to fit a smooth surface to sample points. It identifies the long range variation in the data and thus it is known as trend surface analysis	It works by fitting a smooth surface defined by a mathematical function to the input sample points. This method is applied when there a gentle and gradual change in the spatial surface	No assumption or data requirements are needed to apply the method	The more complex the polynomial, the more difficult to interpret its physical meaning. The surface produced is highly affected by outliers
Local polynomial interpolation	An inexact interpolator used to fit a polynomial for specified overlapping neighborhoods. There is an option to define the search neighborhood and maximize or minimize the number of points used.	It fits a specified order (e.g., zero, first, second, or third) polynomial using all points only within the defined neighborhood. The value at the centre of the neighborhood of the fitted polynomial is used to predict the unknown value.	It is better used when the data show short range variation.	The method is sensitive to the neighborhood distances
Radial basis function	An exact interpolator which predicts values above the maximum and below the minimum measured values.	It produces a surface which must go through all measured values.	Large number of data points	The method is when there are large changes in the surface within short distances.

Table 6.1 Deterministic interpolation techniques definitions

Source: reproduced from ArcGIS manual (2001). Geostatistical analysis, ESRI, USA

- Areal interpolation is used when the dataset is aggregated into areas or polygons. The polygons with known values are called source zones while polygons where values are needed to be predicted are called target zones. Two approaches are usually used in this kind of interpolation: a non-preserving approach which does not preserve the original source zone values and a more recent one which preserves the original values and is considered more accurate.

There are many different situations when areal interpolation method is used. Mrozinski & Cromley (1999) identified four areal interpolation categories; missing data, alternative geography, overlay of a polygon layer and an area class layer, and overlay of two polygon data layers:-

1. In the case of missing data, areal interpolation is used to estimate unknown values for one or a set of zones within one polygon layer. This problem does not require any intersection or data disaggregation as just one map layer is used.

2. Alternative geography is a type of areal interpolation used to transfer data values from one specific geography of areal units to another.

3. A third problem is the desire to estimate attribute values for a new polygon layer which is created as a result of overlaying a polygon layer and an area class layer.

4. The last situation where areal interpolation can be used is to estimate unknown values of a new polygon layer which is found as a result of overlaying two choropelth (polygon) layers.

The interpolation problem encountered in this research is the first type where there are many unknown attribute values of spatial zones (kism/ markazes) within the polygon layer (the whole Egyptian kism/ markazes) which lack survey data (DHS survey). Bennet et al. (1984) reviewed missing data on spatial surfaces and classified the problem into two types: missing information and missing data. Missing information is the situation in which data were collected and recorded but are not available. On the other hand, missing data refers to a correctly collected dataset for an area or region but where some points or zones are missing either because of the nature of data collection (sample surveys) or data availability. Therefore, solving this problem requires selecting suitable interpolation techniques to estimate the missing items. Tobler & Kennedy (1985) used spatial proximity to implement an interpolation method to solve this type of problem. The missing value of each targeted areal location is basically estimated by taking an average from its neighbours. They pointed out that assigning an average value to a location or locations in which data are missing is considered an effective spatial interpolation method. Gregory (2002) examined the accuracy of areal interpolation methods for the purpose of comparing different population censuses and historical long term data. He concluded that areal interpolation enables researchers to analyse geographic changes at a satisfactory spatial resolution. Furthermore, areal interpolation is considered a useful method for maximising the benefits of the available data by interpolating values from multiple datasets and representing them as a single standardised spatial coverage. However, areal interpolation error has to be seen as a fact of life and thus examining the assumptions of the interpolation method used with regard to the variable of interest is important.

6.3 Estimating safe water access for all Egyptian districts

6.3.1 Accounting for DHS survey design in calculating kism level estimates

Sample weighting is defined as a procedure which adjusts for differences in probability of selection among cases within a survey. These differences might happen either because of the sampling design frame or happenstance. Sample weighting is a key procedure in the DHS to correct the imperfections and errors that might lead to incorrect inferences between samples and population (Measure DHS, 2010). The sample weight adjustment could be to account for selection of survey cases with unequal probability, non-coverage of population due to geographic barriers or isolation, or to compensate for non-response among selected households. The adjustment of sampling weights is explained further below following Yansaneh, (2003) & Mohadjer, (n.d):-

- **The adjustment of sample weights for unequal probabilities of selection**

In multiphase household surveys, the main purpose of applying the weighting process is to adjust for differences in probability of selection. When cases in the sample are selected with unequal probabilities, unbiased estimates can be produced by inflating the sample outputs by the reciprocal of the probability of selection.

- **The adjustment of sample weights for non-response**

Where some households do not answer the survey questionnaire in full or in part (non-response), not all of the information targeted by the survey design is actually obtained. Thus, reducing the non-response rate is required to avoid the bias. Two factors are considered as a function which reflects the size of the non-response bias for a sample mean:-

1. The proportion of non-response households.
2. The size of the difference in population means between respondent and non-respondent households.

There are many techniques to solve the non-response problem such as drawing a larger sample than is needed or using a substitution technique where each non-responding household is replaced by another which was in the sample. However, adjustment of sample weights is the most effective technique for handling this problem. The base weights of all eligible non-responding sampled households are transferred by the adjustment techniques to the responding households.

- **The adjustment of sample weights for non-coverage**

Non-coverage in the survey occurs when the sampling frame does not cover all of the target cases. For example, there may be some geographic areas that are inaccessible because of civil war, natural disaster or other geographic barriers and therefore, some sampled cases have no probability of being selected in survey samples. Most developing country household surveys are multiphase surveys (such as the DHS). In this kind of survey the design involves multi-stage area probability. The first stage units are called primary sampling units (PSUs), which are geographic areas (for more details about sampling selection in the 2005 Egyptian DHS see section 2.5.2). The second stage involves creating a list of households or dwellings while the third stage is to create a list of household members from which the sample of people is drawn. Handling the non-coverage problem is usually addressed in two ways; the first is to enhance field procedures including producing accurate household lists, whilst the second is to construct a statistical adjustment for the weights. This is by adjusting the weighted sampling distribution for some variables to correspond to a known population distribution.

Aside from affecting the probability of selection for households, the sampling design of the DHS survey influences estimation results in other ways. Although cluster sampling compared with other techniques is useful in different ways, particularly in reducing budget costs and drawing larger samples, there are many disadvantages such as producing larger errors than other techniques. Large samples are needed to avoid bias errors and cluster sampling is considered less precise than other forms of sampling. The clustered nature of the DHS sampling strategy introduces a ‘design effect’, which results in wider confidence limits than would be encountered with simple random sampling (National Audit Office, 1999; Stat Trek, 2011).

It is clear that survey weights have to be included as an essential step in any household survey analysis. Thus, in this study, aggregate estimates at kism level were calculated for all the DHS variables related to household drinking water, taking into account the sample design and weighting scheme. Figure 6.1 depicts how this process was implemented using the STATA software and the *svy* set of commands (see appendix 3. The steps of this process are explained as follows:-

- Firstly, a linking file which contains kism/markaz names (prepared previously using GIS linkage between the census polygon layer and the DHS GPS layer) was joined to the original 2005 DHS household file.
- Secondly, the design characteristics of the survey were declared using the *svyset* command as follows 'svyset hv021 [pweight=hv005], strata(hv023) vce(linearized) single unit(missing)'.
- Thirdly, the five relevant DHS variables were recoded and scored. Kism-level aggregate index values were generated for each variable using this code 'svy linearized : mean index_interrupt, over(kism)'
- Finally, the data for each index component was exported into an Excel spreadsheet and the confidence limits and estimated mean index values displayed graphically.

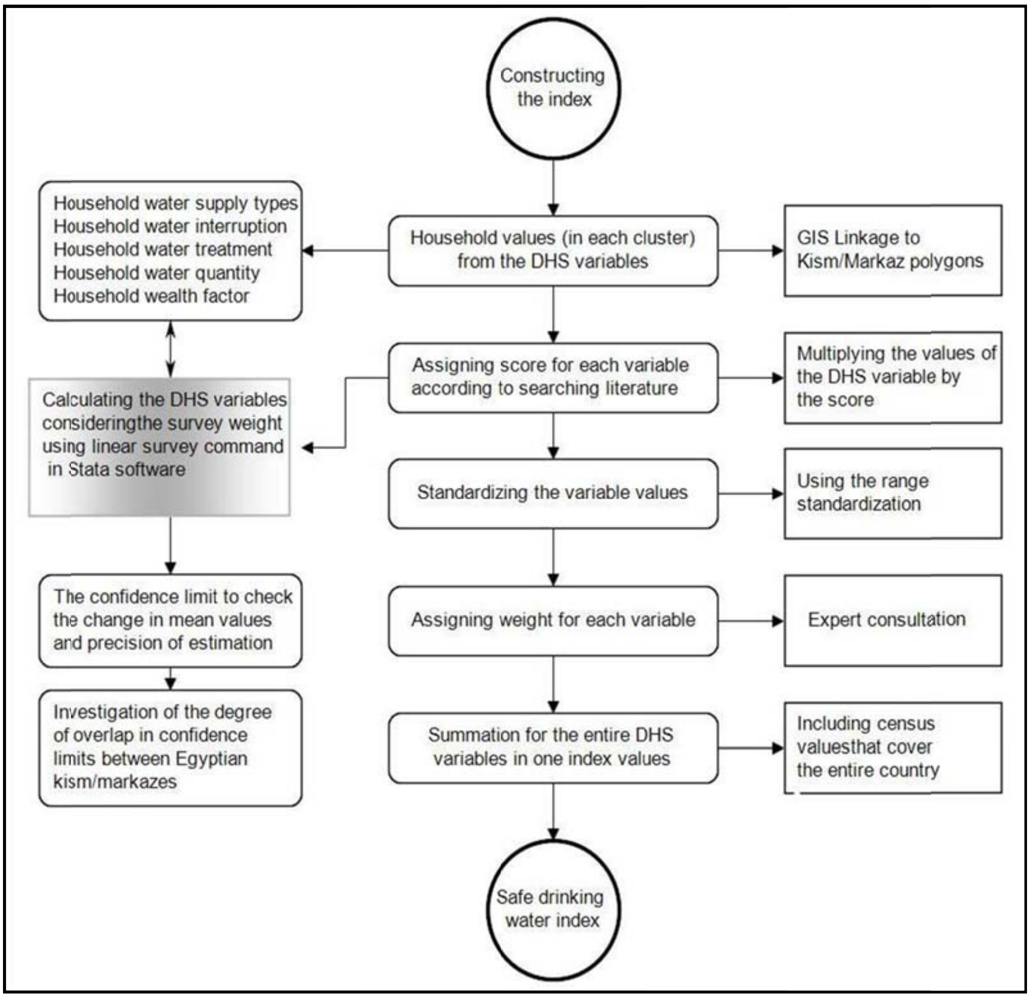


Figure 6.1 Sequence of steps to construct the SDWI taking into account the DHS design

Snedecor & Cochran (1989) defined the confidence limits for the mean as an interval estimate for the mean. It is expected that this estimate varies from one sample to another and thus constructing interval estimates is recommended.

Upper and lower confidence limits for the mean can be generated which give a clear indication of how the estimated mean may differ from the true one. The wider the interval, the less precise is the estimate. Figure 6.2 illustrates the upper and lower 95% confidence limits for the mean values of the water interruption index component for Cairo Governorate kisms. Note that it is not possible to calculate confidence limits where all households within a kism share the same value for the supply interruption component. It can be seen that for some kisms, the confidence limits do not overlap. For example, El Galmalia's confidence limits do not overlap with those of the El Marag kism. This suggests that even allowing for uncertainty from sampling, it is possible to distinguish differences between at least some kisms using the DHS data.

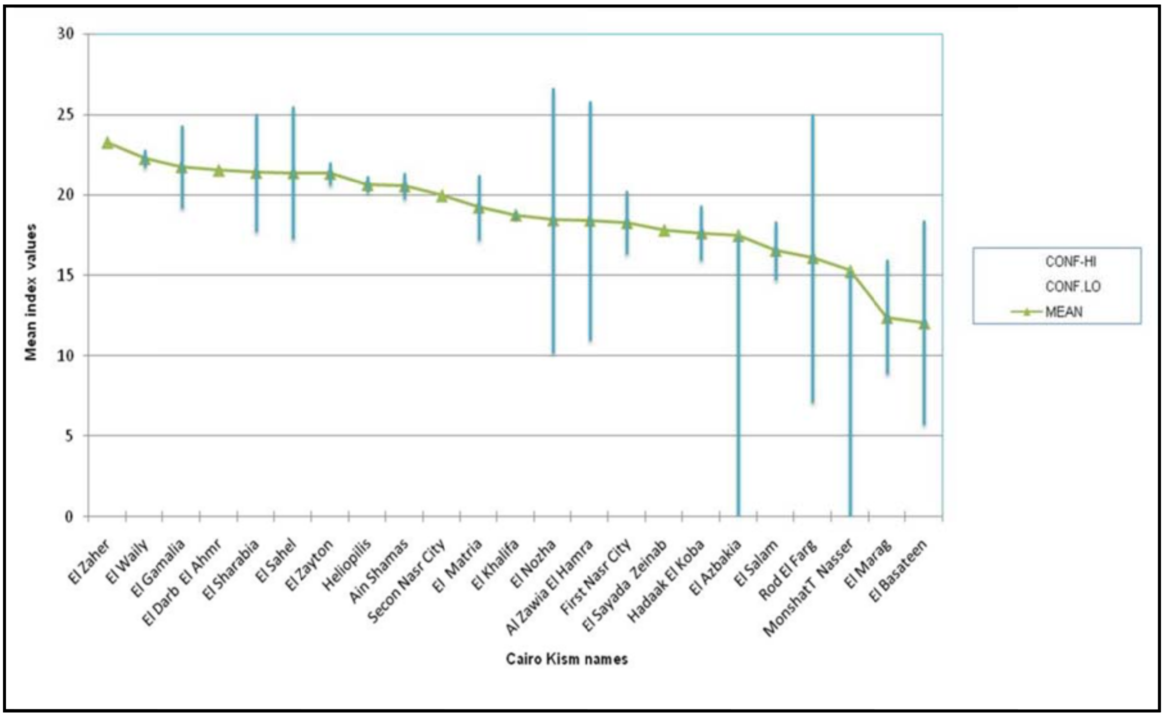


Figure 6.2 Upper and lower 95% confidence limits for the mean values of the water interruption component index in kisms whithin Cairo governorate

6.3.2 Tobler and Kennedy interpolation method

To estimate the unknown DHS values for the Egyptian kisms/markaz Tobler &Kennedy's (1985) areal interpolation method was implemented:-

In this method, the initial objective is to weight the values of neighbouring polygons (sources) according to the length of their shared common boundaries with the polygons whose values are unknown (targets). ArcGIS does not offer a direct technique to obtain the length of these boundaries and thus a contiguity tool for ArcMap developed using Visual Basic for Applications (VBA) and downloaded from http://forums.esri.com/Thread.asp?c=93&f=993&t=217068 and used to achieve this process and provide the length of shared borders for every adjacent combination of 360 polygons. However, some completely isolated polygons that are not contiguous with any other polygons (new cities in the desert or islands in the Nile River) are necessarily omitted from this calculation. Similarly, no values can be computed for polygons that are entirely bordered by other polygons which themselves have no DHS values. Consequently this process generated missing data as outputs for nine isolated polygons and three other polygons with neighbours lacking DHS values. It is important to note that the method as implemented here only takes account of the values of first-order neighbours.

To join the lengths of the shared boundaries (the outputs of the contiguity tool) to the known DHS values (a single column comprised of a composite of DHS variables) a multiple step sequence of data transfer back and forth between Microsoft Excel and Access was implemented (see figure 6.3) until a set of estimates had been produced for each polygon, based on the weighted combination of its first order neighbours in proportion to the lengths of common boundary segments. Completing the interpolation method, except for the 12 missing polygons mentioned earlier, values for both polygons lacking DHS data and the majority of polygons for which data were available were estimated. Comparing the interpolated values against the known, original values, Figure 6.4 illustrates that there is a significant (P value = 0.000) correlation between the interpolated and observed values, where both were available (Pearson's R = 0.634).

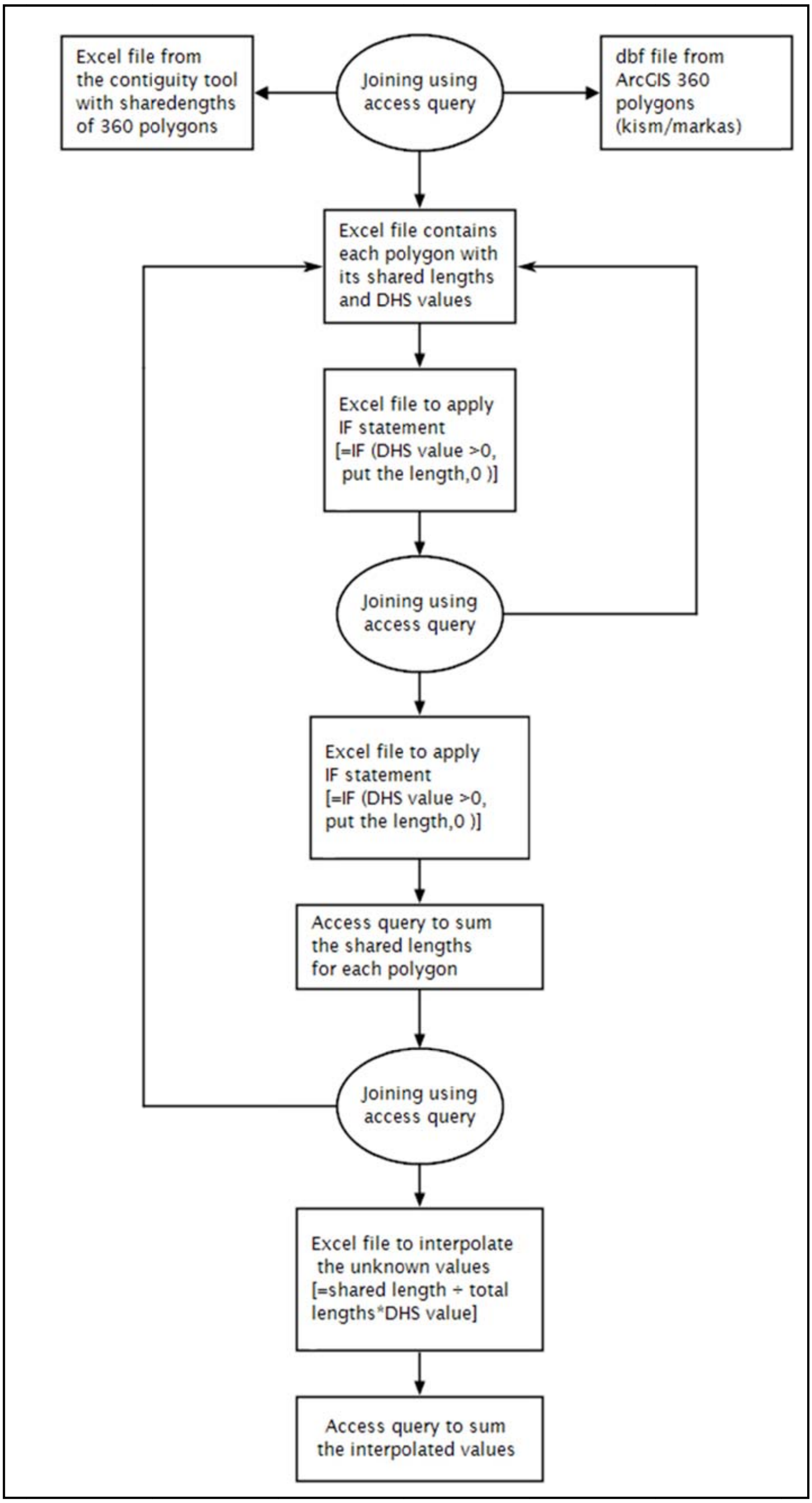


Figure 6.3 Combination of Access and Excel steps to implement Tobler and Kenndy's areal interpolation method

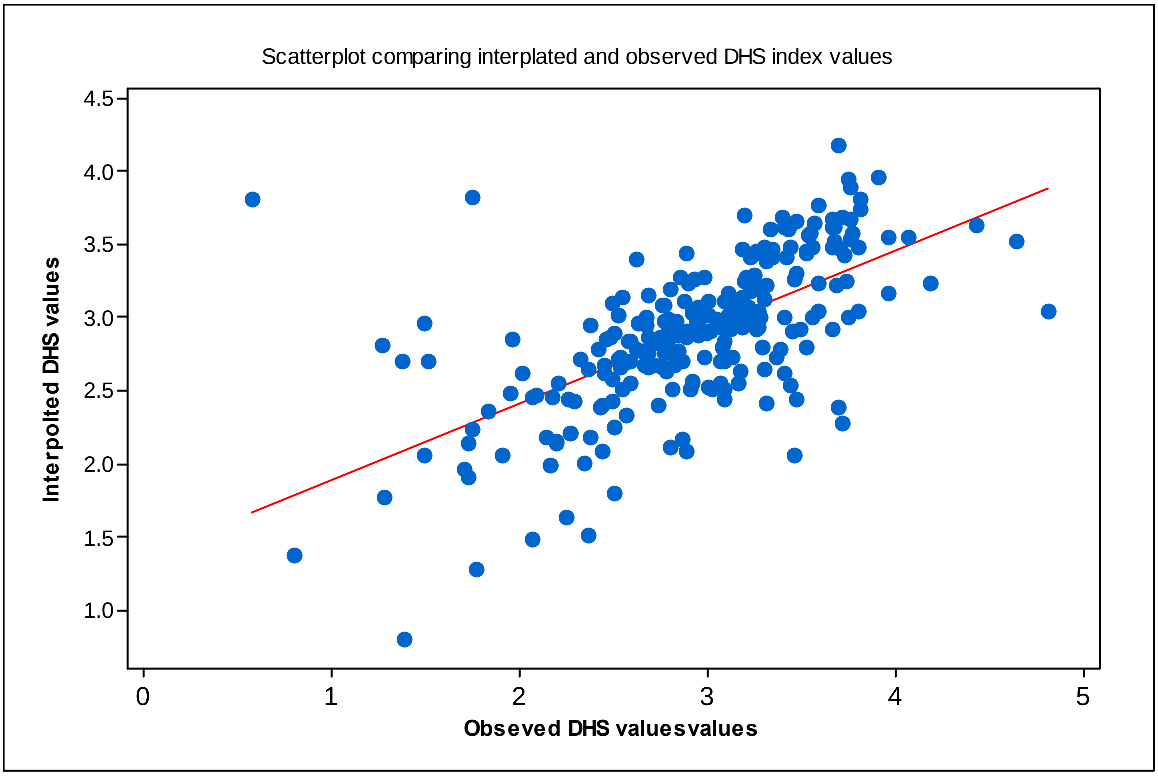


Figure 6.4 Comparison of the interpolated versus measured values of the DHS derived component of the SDWI for 259 kisms and markazes

6.3.3 Resultant map of the SDWI for all Egyptian districts

The efforts that have been made so far to measure household access to safe water in developing communities either by the MDG, JMP or other international programmes are limited in scope, resolution and coverage. In these measurements, there has been no attempt to integrate household socioeconomic characteristics, with a focus just on drinking water sources and classifying these sources into two categories "improved and unimproved" (see section 2.2.1). Furthermore, measuring access to safe drinking water has to be implemented by a quantitative and flexible tool such as an index. As a result these measurements lack sufficient geographic detail to address drinking water problems nationally based on small geographic areas in each country. The SDWI has been designed to highlight drinking water availability, accessibility and identify districts which suffer from poor drinking water supplies and delivery. The GIS techniques developed, particularly spatial linkage for survey and census variables, have focused on integrating two datasets which are available in many developing countries. This is in order to provide a means which is longitudinal and repeatable and which can be constructed in other countries. Combining different indicators into a single index gives a valuable opportunity to take account of all relevant variables that may influence safe water accessibility. So far there has been no research conducted on the spatial distribution of household access to safe drinking water in Egypt, either at provincial or sub-provincial geographic scale.

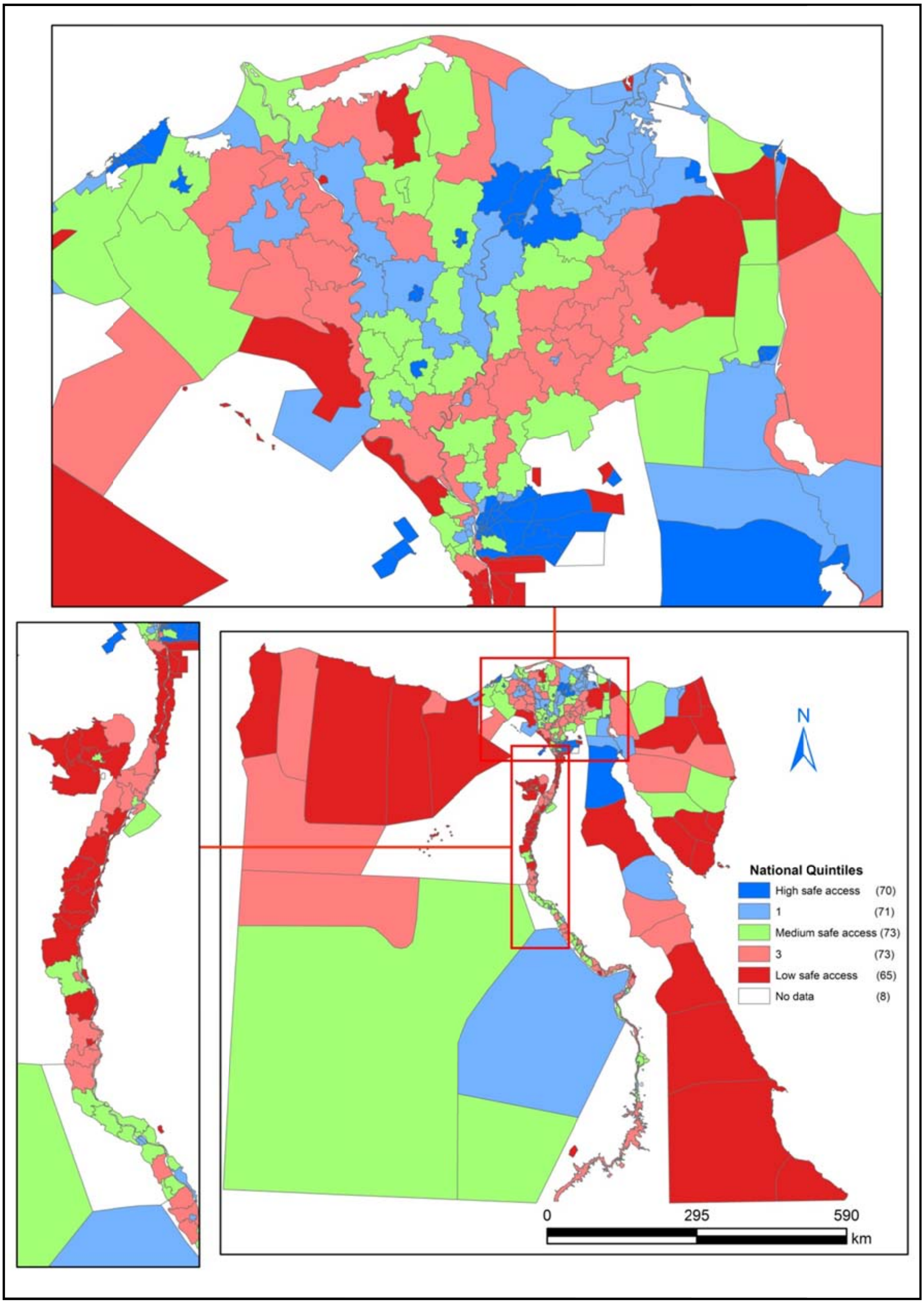


Figure 6.5 Map of the SDWI of Egypt at district level (quintiles)

Districts of Frontier governorates

Districts of Mattrouh governorate show the least safe access to water sources among all desert governorates while districts of New Valley governorate tend to show higher values compared with other governorates. Among the districts of Red Sea governorate Second Hergada kism scores higher while the other districts score among the lowest values of the index. In South Sinai, although Sharm El Sheik is an international tourist city, it scores relatively low as does Saint Catherina city. One reason could be that water supplies in these cities particularly Sharm El Sheik depend on desalinisation and on bottled drinking water. In North Sinai, except for the El Areesh kisms (First, Second, Third, and Fourth) and Beer Al Abd that score medium values, all other districts score lower values and are among the lowest for access to safe drinking water. In a broader sense, there is no great variation among districts of Egyptian Frontier governorates and most of them have lower index values. This is because of the nature of desert communities that depend on groundwater as a source for everything. Most households in these governorates live remotely and move from place to place with their livestock.

Like many developing countries, drinking water accessibility in Egypt is associated with various demographic, socioeconomic, environmental and human factors. Hence, the SDWI has identified and combined most of these variables to present the real picture of access to safe drinking water in a comprehensive and clear way at sub-national level. The map of the SDWI (Figure 6.5) illustrates clearly various spatial patterns of drinking water accessibility across all Egyptian kism/markaz:

Rural/urban pattern

Broadly, the first spatial pattern that can be drawn from the map is the difference between kism (urban districts) and markaz (rural districts), where urban areas show higher index values and therefore higher access to safe drinking water. However, although drinking water sources in Middle and Upper Egypt are closer to the Nile river than those in the Delta governorates, most rural districts in Delta governorates show higher values compared to Middle and Upper Egypt governorates (figure 6.6). This could be due to weaknesses and failure in drinking water infrastructure and plants which leads to more frequent water supply interruptions, as well as a high percentage of wells and pumps as sources of drinking water. Similarly, sanitation systems in rural districts of the Delta governorates reported a higher percentage of household access to public sanitation than middle and south of Egypt. Another fact is that rural Lower Egypt in general is closer to the capital Cairo and thus Egyptian decision makers may give more attention to these areas in terms of development, constructing utilities and planning.

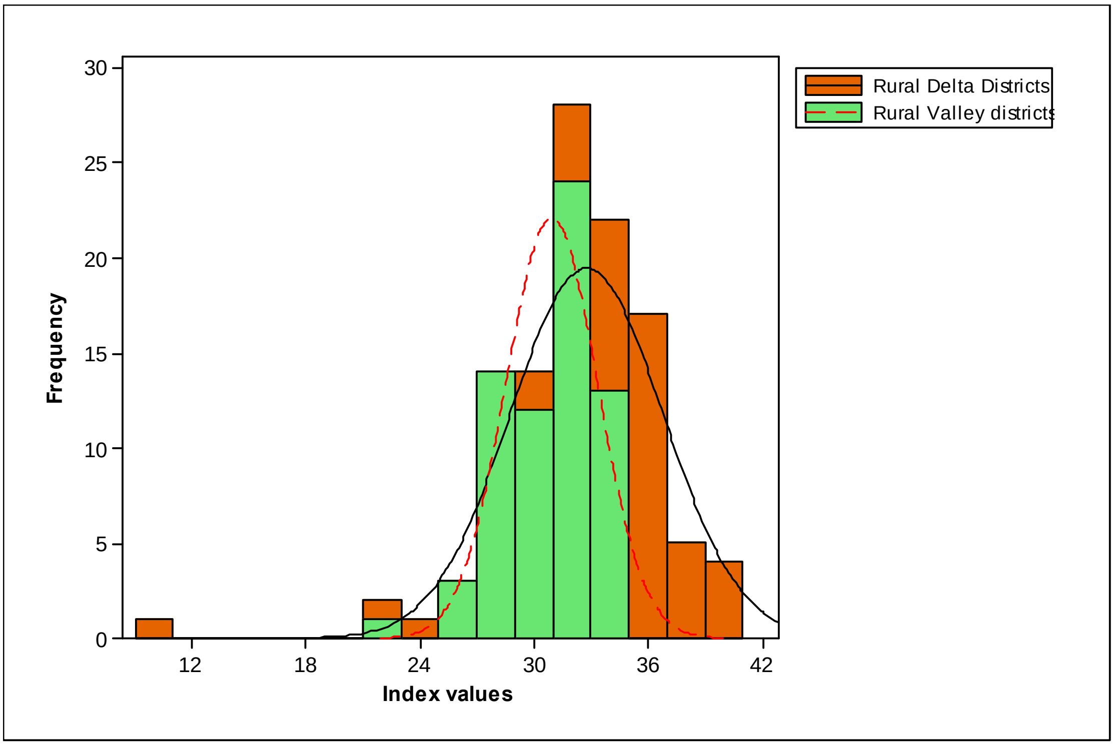


Figure 6.6 Histogram showing the normal distribution of the SDWI values for Lower districts compared with Upper Egypt districts

Districts of Greater Cairo

Districts of greater Cairo show the safest access compared with other urban kisms across Egyptian governorates (figure 6.7). However, kisms located on the east bank of the Nile River such as First and Second Nasr City kisms, El Zayton, El Nozha, and New Cairo kisms show higher values than others. Most of wards within these districts most of their wards are well planned where rich households live in flats, villas or houses. Other kisms such as Al Zawia El Hamra, Shobra, and Rod El Farg show high values of access to safe drinking water although they have large populations, deprived households and include some crowded wards. This can be explained by the higher percentage of households connected to the public network and water taps and lower frequency of water supply interruption. In the east Ancient Cairo shows lower index values while in the west bank of the Nile River El Warak kism and Island have the least safe access. Households living in this Island have access to neither drinking water nor sanitation public networks. In the south all districts located in Helwan governorate such as El Maadi, Tora and el Tebeen show the least safe access.

Districts of Lower Egypt

Three major patterns of the SDWI are clustered in the rural districts of Delta region (Figure 6.8). The first pattern is located between the Damietta and Rosetta Nile branches and consists of rural districts with higher index values. Such districts include Monf, Berket El Sabh, Shbeen El Kom (El Monfya Governorate), Tanta, Samnod, Basyon and Kafr El zayat (El Gharbya governorate) and Kafr El Shyakh (Kafr El Shikh governorate) while Mansora, Talkah and Nabroh (El Dakhlya governorate) score the highest values in the region. Kotor (El Gharbya governorate) and Belaa (Kafr El Shikh governorate) scored the lowest values in the middle Delta. The second pattern is centred mainly around the east of the Damietta Nile branch. This pattern includes Ashmon, El Bagor (El Monfya Governorate), Banha, Kafr Shoker (El Kaliobya governorate), El Zagazik, Menia El Kamah, Hehya, Abou Hammad, El Abrahemia, Fakous, Abou Kpeer, Awlad Saker and El Hosainia (El Sharkya governorate). This reflects the fact that El Sharkya governorate is the poorest in access to safe drinking water. One reason could be that it does not share the Damietta Nile branch which means that there is no direct drinking water uptake from the Nile River. In addition, it seems that this governorate has a higher percentage of households who use pumps as a source of drinking water in the rural areas of this governorate. For example 6% of households in Dearb Negm use pumps as a source of drinking water while 8% of households do so in Menia El Kamah. Rural districts of this governorate have a lower percentage of households who have access to public sanitation, for instance Abou Hamad (19%), Abou Kpeer (28%), and El Hosaina (2%).

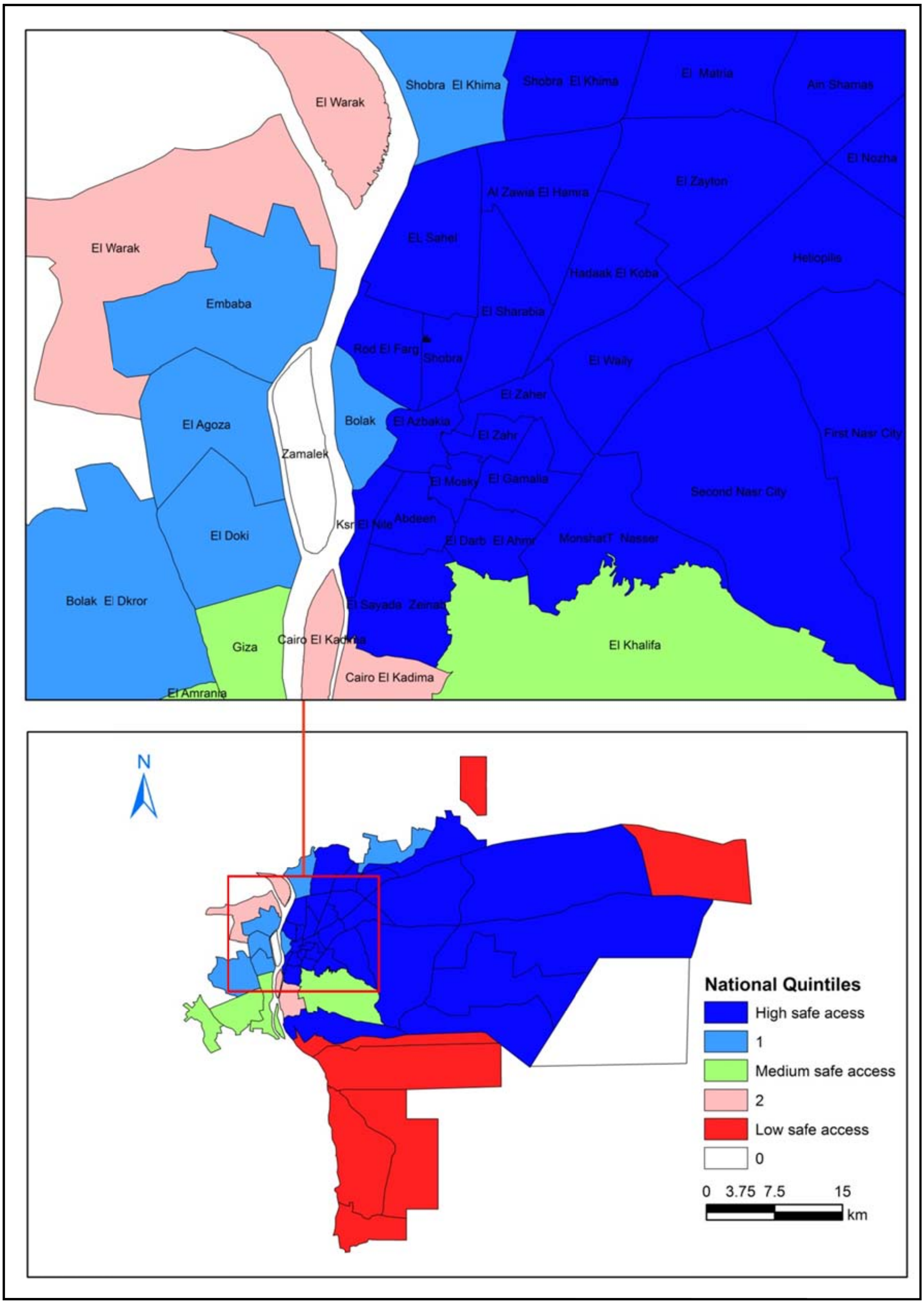


Figure 6.7 The SDWI in Greater Cairo kisms

The third pattern also relates to the lowest values of safe water accessibility and is located to the west of the Rosetta Nile branch and within El Behera governorate in kisms such as Badr, Kom Hamada, El Dalangat, Etay El Barod, Shabrakheet, and El Mahmoudia. Although these share a border with the Rosetta Nile branch, they include reclamation land where households rely on groundwater and use wells and pumps as drinking water supplies. For instance, 8% of households in Kom Hamada markaz have access to pumps and wells as a source of drinking water while just 12 % have access to public sanitation. In El Dalangat markaz 21.5% of households use wells and pumps as drinking water supply types while just 15% have access to public sanitation and 66.6% of households are connected to the public drinking water network. Areas located in the south and east of the Suez Canal scored medium index values, particularly East El Kantra while El Kantra markaz in the west of the Suez Canal scores the lowest value in the region.

Regarding the urban districts of Lower Egypt, most of the kisms scored medium or high values of the SDWI particularly in the middle and east of the Delta region such as El Mahala El Kobra kism and Tanta kism (El Gharbya governorate) and El Mansora First and Second kisms, and Meet Gamer kism (El Dakhlya governorate). Despite the fact that rural areas of El Sharkya governorate scored the lowest values in the east Delta, urban kisms of this governorate scored medium to high values such El Zagazig First and second kism (the capital of the governorate) and Fakous kism.

Northern urban governorates (Alexandria and Port Said)

Three urban governorates are located in the north of Egypt (Alexandria, Port Said, and Suez) while the other two (Cairo and Giza) are considered part of the Greater Cairo region. Regarding the Alexandria kisms, except for two kisms (Porg el Arab and El Amria) all kisms in the city scored the safest access, particularly in the east of the city such as El montzh, Second El Rmal and Bab Sharki (Figure 6.9). Borg El Arab is a new industrial city while El Amria kism is a main destination of migrants particularly from the Delta regions. Consequently, the lowest percentage of households who have access to public sanitation (11%) was found in Borg El Arab and El Amria (31%). Similarly, in borg El Arab kism there is a very high percentage of households who use pumps and wells (0.75%) compared with the majority of other kisms (0.0%) and almost 7% of households in El Amria lack a tapwater supply inside their dwellings. Port Said governorate is located on the North east of the Nile Delta and considered the Northern gateway to the Suez Canal. Kisms of Port Said governorate scored the highest values and safest accessibility compared with other kisms in the Suez Canal area (Figure 6.10). However, El Ganob kism shows the lowest value.

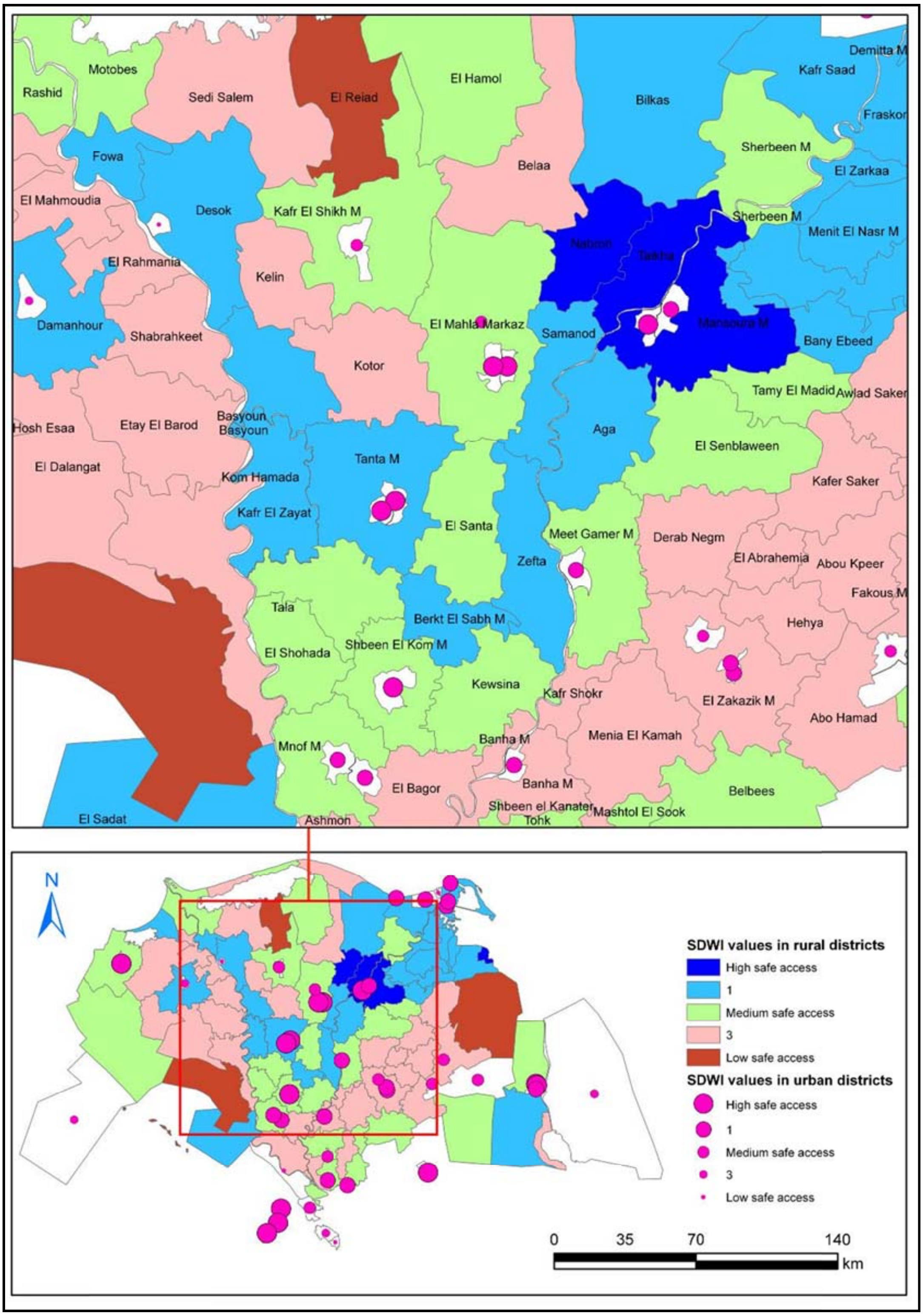


Figure 6.8 The SDWI in districts of Lower Egypt

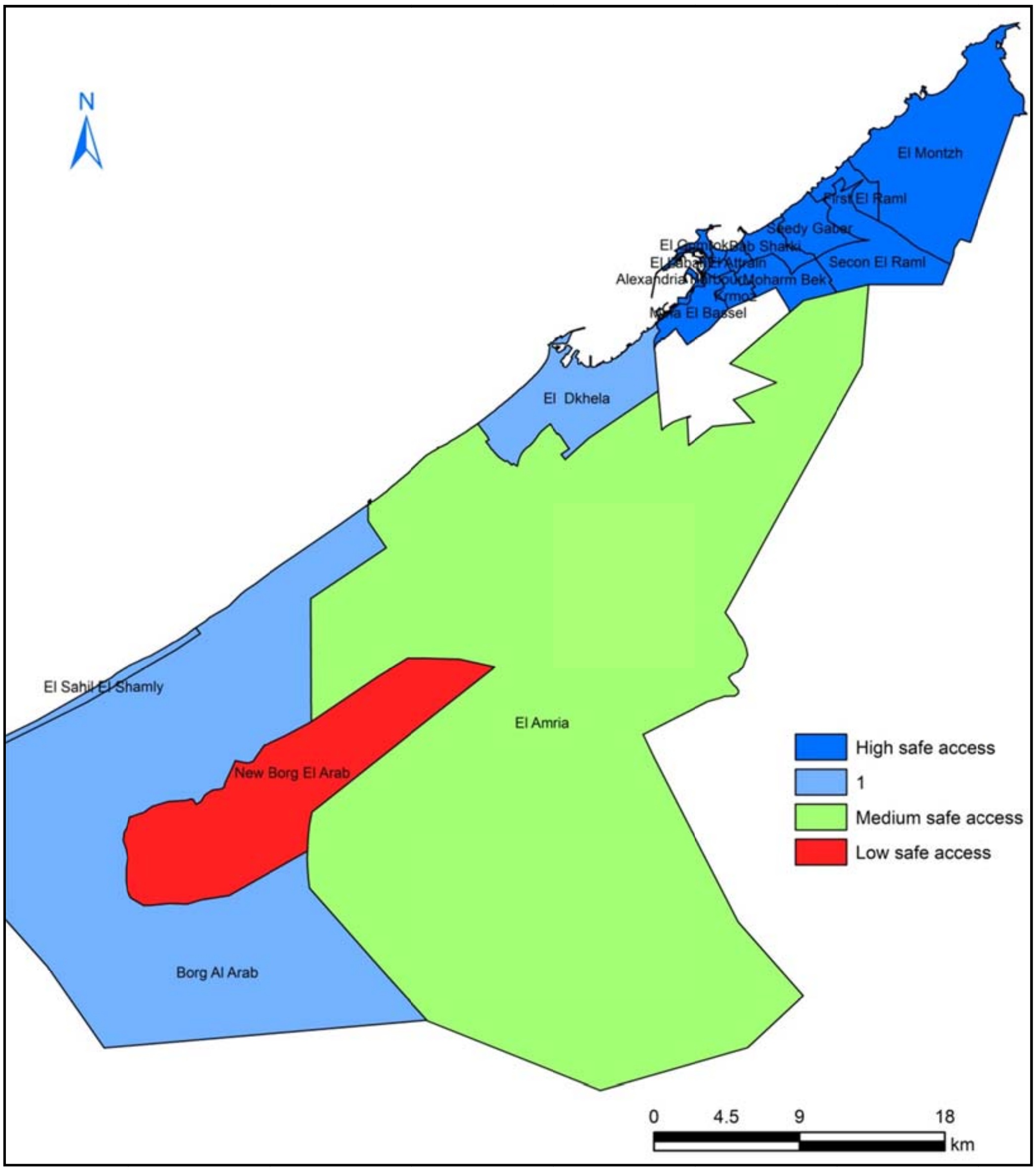


Figure 6.9 The SDWI values in Alexandria kisms

Districts of Upper Egypt

The map of the index provides a clear indication that districts in Upper Egypt governorates show lower index values compared to lower Egypt governorates, with districts in the Delta region scoring better than elsewhere (figure 6.11). Districts of El Fayoum and Bani Souif governorates generally score lower values, as illustrated by kisms such as Beba and El Fashin (Bani Souif), Atas, Sanors, and abshowy (El Fayoum). The SDWI scores well in the south of El Menia districts such as El Menia kism and markas, Malwy, and der Mows while it scores lower in northern districts such as Maghagh, Bani Mazar, and Samlot. Going further down to the south, districts of Assiout governorate overall show the safest access compared to other Upper Egypt governorates. El Badari and Assiout first show high index values whilst the rest of the districts have medium index values except for Abou Teeg markas which has a lower value. For the districts of the three governorates located in southern Egypt (Sohag, Qena, and Aswan) the index values are generally in the low to medium range except for the cities of Luxor and Aswan. These cities are the famous tourist cities in Egypt and thus have well-constructed infrastructure including public networks for drinking water, electricity, and sanitation.

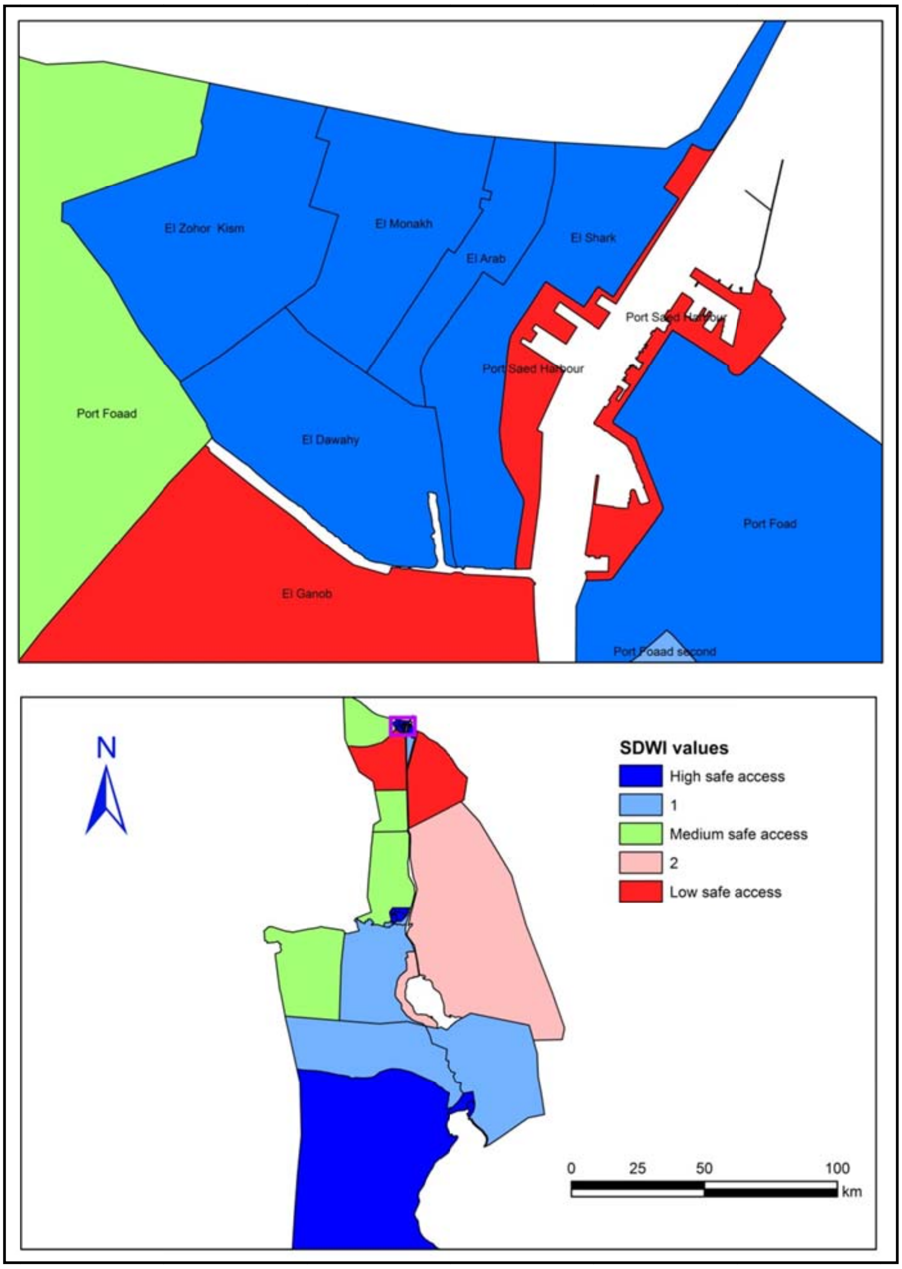


Figure 6.10 The SDWI values in the Suez Canal region highlighting, with the urban governorate of Port Said as an inset map

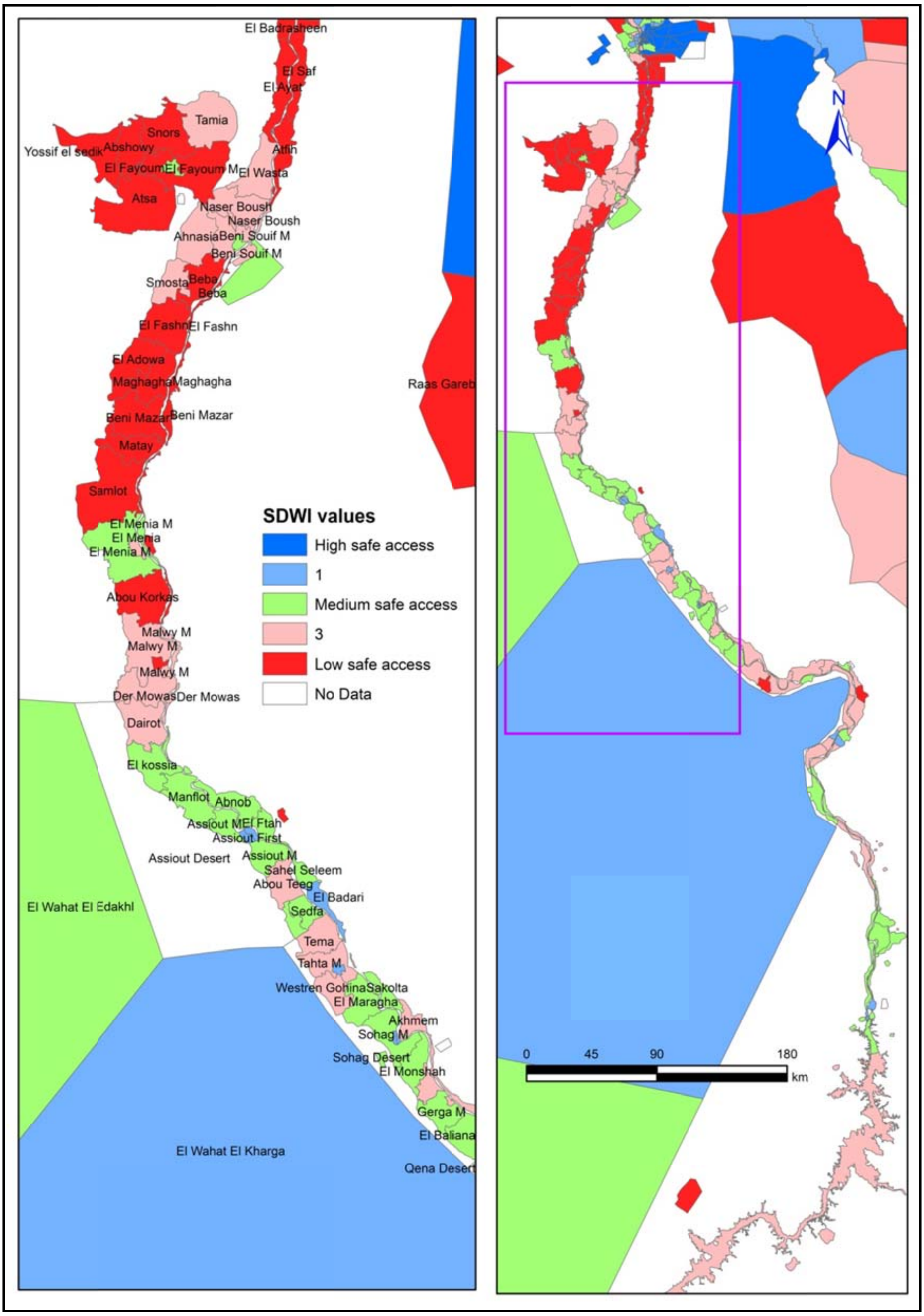


Figure 6.11 The SDWI values in rural and urban districts of Upper Egypt governorates

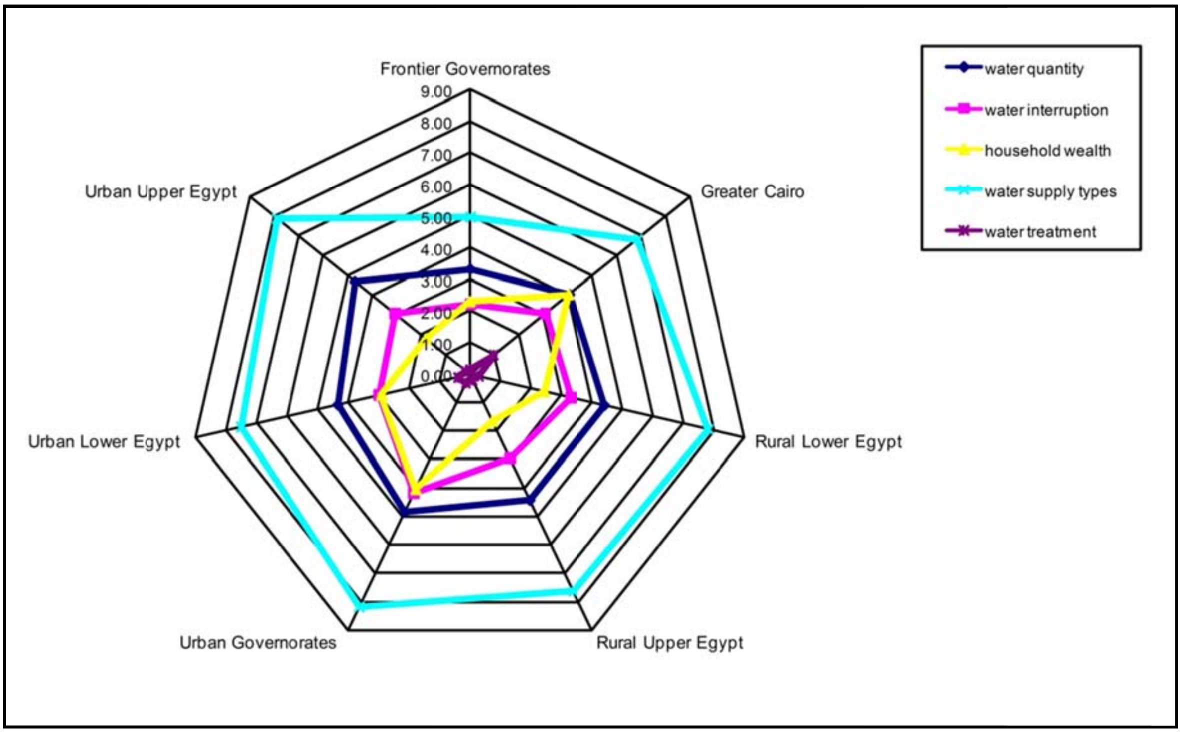


Figure 6. 12 The average values of the DHS index components for different Egyptian regions

Based on the five interpolated DHS index components, figure (6.12) shows the variation between Egyptian regions. It can clearly be seen that although the average values of the index components, particularly water supply types, are close for each region, Frontier governorates compared with other regions scored lower values in most of the index components especially water supply types and water quantity. By contrast, urban governorates scored higher in almost all components. In terms of water quantity, Greater Cairo region and urban governorates scored the highest values followed by urban Upper Egyptian and urban Lower Egypt. The average score for water interruption in rural Lower Egypt was higher than in rural Upper Egypt while the worst situation was in the Frontier governorates which scored the lowest average for water interruption. Urban governorates scored the highest averages for index values. Regarding the household wealth component, urban governorates and Greater Cairo regions scored the highest averages. The lowest average household wealth is found in Upper Egypt in both rural and urban districts. This reflects the fact that deprivation in Upper Egypt is higher than in other Egyptian regions including Frontier governorates. Although all Egyptian regions scored low averages in terms of drinking water treatment (less than 1.00), Greater Cairo scored the highest average value and this reflects the high percentage of Egyptian households living in Cairo and Giza governorates who treat their drinking water using filters.

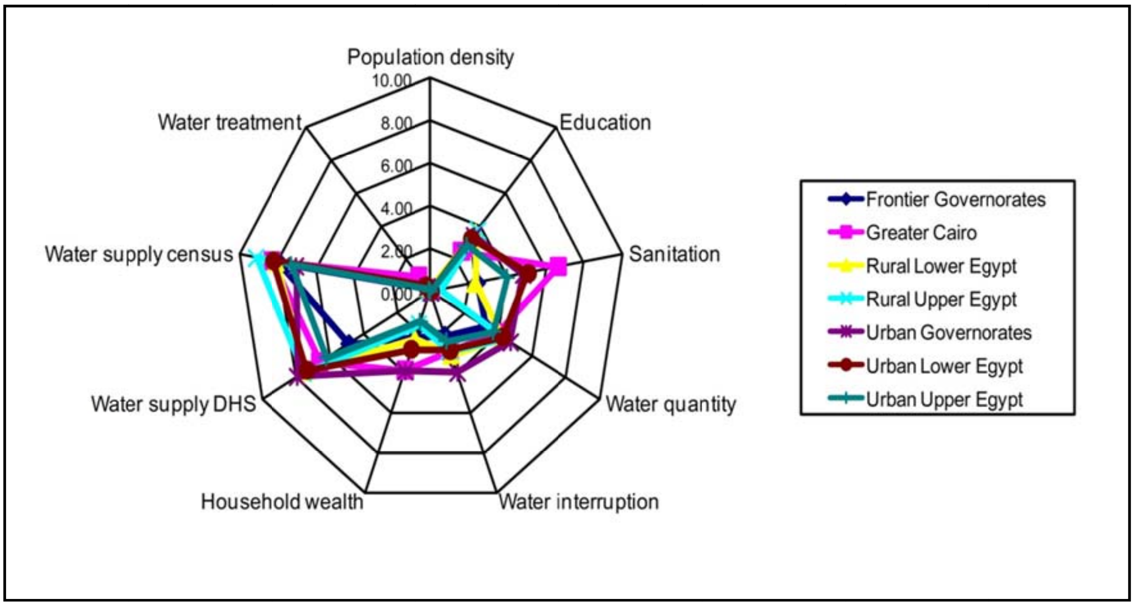


Figure 6.13 The average values of all index components (DHS and census) across Egyptian regions

Figure (6.13) shows all eight SDWI components drawn from both census and DHS datasets. The diagram shows that the values of water supply types represented using census data were higher than supply types from the DHS dataset. The average value of the public sanitation component was higher in Greater Cairo region while the poorest sanitation situation was found in Upper Egypt, particularly among rural districts. The education component was lower in Greater Cairo and urban Upper Egypt but higher in rural Upper Egypt and Lower Egypt. The lowest values for population density were in the Frontier governorates and rural Upper Egypt while the highest were in Lower Egypt and Greater Cairo.The population density component is also low for other Egyptian districts that are large in size with small populations such as districts located in the western and eastern deserts and the Sinai peninsula. Similarly all rural districts neighbouring reclamation land have lower population density values.

The SDWI suggests that there are significant spatial variations of safe drinking water accessibility across Egyptian districts and from one region to another. Geographic patterns can be recognized that differentiate rural, urban and desert districts. However, rural and urban patterns varied between Southern Egypt or Valley governorates to Northern Egypt or Delta governorates. Urban districts of Greater Cairo had better access to safe water than others while urban districts of Alexandria governorate were better than other urban districts that are located outside Greater Cairo (Port Said and Suez governorates). The districts that are located in the Frontier governorates scored the lowest values for access to safe drinking water and this can be explained mainly through a lack of access to public networks for both water and sanitation with households using wells and pumps as sources of drinking water as a consequence.

6.4 Summary

This chapter has covered the small area estimation problem, together with related methods and applications. The most obvious finding to emerge from the literature was that small area estimation has been widely used in different geographic and demographic domains, particularly to research the spatial distribution of socioeconomic phenomena and estimate unknown values of survey data for administrative units. Nevertheless, interpolation methods have seldom been used to estimate index component values for small areas, and not been used for drinking water index construction. The chapter has described the types of small area estimation, particularly the simple small area methods and regression methods. The areal interpolation review revealed that interpolation is a widely used method when survey datasets are aggregated into a group of polygons or zones. Sample weighting as an important procedure of any survey design was discussed, especially the way that sampling weights are related to the probability of household selection in the survey. This was followed by an explanation of how this study took account of the DHS sample design and weight scheme. Tobler and Kenneddy's interpolation method was successfully applied using ArcGIS, Excel, and Access software. The correlation between the interpolated and known DHS values was significant. Consequently, the DHS and census values of all index components were combined and presented in map format for almost all Egyptian districts. A conclusion was drawn from the map of the index that spatial patterns could be discerned in the SDWI values across all Egyptian regions. Generally, the safest access is found in urban areas rather than rural and in Greater Cairo districts in particular, especially districts located in the east. The least safe access is found in desert districts and in Upper Egypt, rather than the Lower Egypt districts.

Chapter 7: Evaluation and transferability of the Safe Drinking Water Index (SDWI)

7. 1 Overview

The aim of this chapter is to evaluate the constructed SDWI for Egypt and discuss the transferability issues, particularly the international situation for household drinking water in the light of the MDG. The first section evaluates the SDWI values against health outcomes by correlating these values against child diarrhoeal disease. It also presents an evaluation from Egyptian stakeholder perspectives of the SDWI outputs. The following section describes the international context of measuring access to safe drinking water and the datasets that might be available from national censuses and elsewhere in lower middle income countries. It summarizes drinking water accessibility as a target of the MDG and discusses the progress of Egypt towards this target relative to other similar countries where the index methodology could potentially be used. The second section outlines the most widely encountered global factors that impact drinking water availability in the developing world, and how these might differ in Egypt compared with other low or lower middle income countries. Both issues affect the transferability of the index. The final section focuses on a transferability case study of Jordan. It discusses the justification for choosing Jordan as a case study and explains Jordanian DHS and census geographies. The section then turns to address spatial linkage of Jordanian DHS GPS clusters to census polygons. This is followed by a further analysis of misallocated GPS clusters relative to provinces within the 2005 Ethiopian DHS. The chapter ends with a review of index construction issues in Jordan, particularly the available census data and how to evaluate these as a step towards completing the SDWI for Jordan.

7.2 Evaluation of the Safe Drinking Water Index (SDWI)

7.2.1 Evaluation against Health outcomes

Investigating the interrelationships between drinking water and population health outcomes is one objective of this current research. However, health data availability limits the study's ability to draw a wide picture about this relationship. In Egypt as with other developing countries, it is difficult to find data about population health available particularly for sub-national geographic units (kism/markas) the scale at which the SDWI has been constructed. The 2006 Egyptian census did not involve any questions associated with health conditions of households nor did it collect information about illness and disease and the only available attribute data were about different types of disability. For various reasons therefore, a decision was made to use child diarrhoeal disease data from the DHS survey. These are explained as follows:

- In most developing countries where there is a high percentage of mortality among children aged less than 5 years, health registration systems for reporting causes of death are limited and where they exist often do not work well (Mathers et al. 2005; Black et al.2003). Furthermore, when there are death registration data that are supplied to the WHO they are at national or regional geographic scales (WHO, 2006).
- Relatively few people in most developing countries go to public clinics to treat diarrhoeal disease and usually they either follow traditional treatments such as providing children with natural herbs or other remedies, particularly where fever and vomiting accompany diarrhoea as symptoms or according to other cultural beliefs (Herman & Bentley, 1993).
- Available data sources such as censuses, surveys and official data for Egypt lack any sort of data relevant to individual health status. For example, in CAPMAS, data about health focus on health services such as total number of GPs and medical staff, nurses and public and private clinic in each governorate.
- The level of aggregation for health data in the Egyptian Ministry of Health is the governorate and national rural and urban breakdowns, so health data are not available for academics to undertake spatial analysis of health and epidemiological issues.

Consequently, since geographically disaggregated health data about Egypt are generally unavailable, the decision was made to use child diarrhoeal disease from the DHS survey to examine whether there is a correlation between access to safe drinking water and health. The diarrhea question in the 2005 Egyptian DHS was posed to mothers. The question was whether any of their children aged less than five years had had diarrhea in the last two weeks preceding the survey (El-Zanaty & Way, 2006). Data about child diarrhoeal disease were aggregated to the 259 sampled DHS kisms/markazes and thus a variable about child diarrhoeal disease was generated. Using Stata 10 software, a Lowess smoother was applied to the generated variable concerning child diarrhea set as a dependent variable against the SDWI values for the 259 sampled DHS districts. A lowess smoother was implemented using the lowess command in Stata. Lowess smoothing is a locally weighted regression of a Y variable on an X variable, often used to fit a smoothed trend line to a scatterplot (NIST/SEMATECH, 2011).

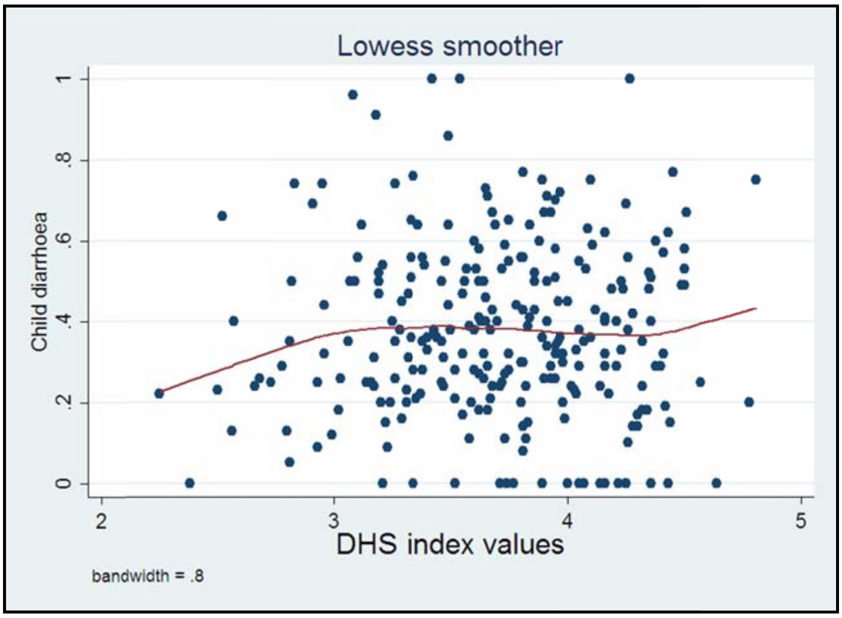


Figure 7.1 Lowess smoothing line showing the correlation between diarrhoeal disease and combined DHS safe drinking water index components for 259 districts

Figure 7.1 illustrates the relationship between the DHS index values and diarrhoeal disease in the 259 sampled Egyptian districts. It can be seen that correlation between the two variables statistically is not significant. The Pearson's correlation coefficient was 0.02. A correlation matrix was calculated to measure the linear relationship between the values of DHS index components and diarrhoeal disease values (table 7.1). None were correlated with child diarrhea rates. Also no correlation was found between the SDWI values including both (DHS and census) and child diarrhoeal disease (Figure 7.2). These findings may reflect the limitations of using child diarrhoeal disease as a health indicator. However, this finding was not unexpected and there are many potential reasons why child diarrhoea and other index components were uncorrelated:

- The generated variable of percentage child diarrhoeal disease in the 259 districts was estimated based on the DHS sample households, which are not designed to provide direct level estimation.
- Despite the fact that unsafe drinking water, microbially contaminated water and supply interruptions are regarded as major sources of diarrhoeal disease (Hunter et al.2009; Fewtrell et al. 2005; Ashbolt, 2004; Esrey et al. 1991; Blum & Feachem 1983), such infection comes also from other pathways such as food, poor nutritional status, hygiene practices and inadequate sanitation practices (Usfar et al. 2010; Halvorson, 2004; Haddad et al.1998).

- In the literature many studies have examined the correlation between drinking water indicators and diarrhoea, particularly in the developing world (Moy et al.1991; Gundry et al.2009). For instance, Gundry et al. (2009) reviewed the literature and assessed risk factors for child dysentery and watery diarrhoea. They concluded that since diarrhoeal disease symptoms and conditions may relate to many transmission pathways it is difficult to quantify the risk associated with water-related factors.
- The DHS is a seasonal survey and it is possible that children show peak levels of diarrhoeal disease during periods of time other than that when the survey takes place. The incidence of diarrhoeal disease may be higher in one season (e.g. in the winter) than another (e.g. in the summer).
- The DHS relies on the mother or child's carer to recall diarrhea experienced by the child. It is therefore difficult to obtain a direct and clear answer about the occurrence of child diarrhoeal disease by asking the adult household whether the child had diarrhoea during the last two weeks.

This examination of the relationship between drinking water and health outcomes of Egyptian households using child diarrhoeal disease is limited because of restricted data availability for water-borne diseases in Egypt. Nevertheless, such an examination might be a valuable measurement of the relationship between safe water access and health if alternative, accurate health datasets relevant to diarrhoea or other water borne disease were to be available at district level.

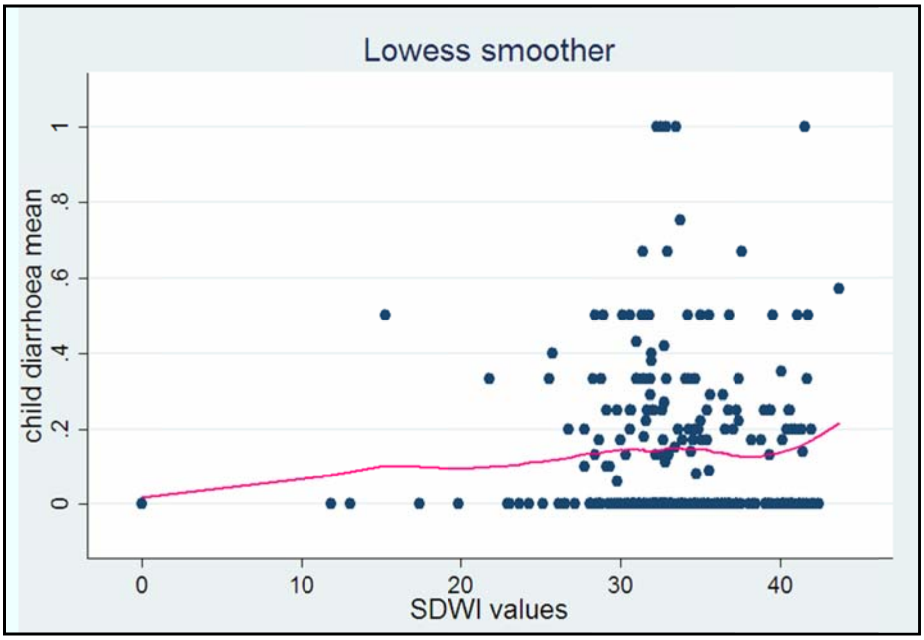


Figure 7.2 Lowess smoothing line showing the correlation between child diarrhoeal disease and the SDWI for 259 districts

		Water quantity	water interruption	Household wealth	Water supply types	Water treatment	Child diarrhoea
Water quantity	Pearson Correlation	1					
	Sig. (2-tailed)						
	N	259					
water interrpution	Pearson Correlation	.113	1				
	Sig. (2-tailed)	.069					
	N	259	259				
Household wealth	Pearson Correlation	.458**	.259**	1			
	Sig. (2-tailed)	.000	.000				
	N	259	259	259			
Water supply types	Pearson Correlation	.801**	.008	.364**	1		
	Sig. (2-tailed)	.000	.898	.000			
	N	259	259	259	259		
Water treatment	Pearson Correlation	.162**	.100	.580**	.145*	1	
	Sig. (2-tailed)	.009	.107	.000	.020		
	N	259	259	259	259	259	
Child diarrhoea	Pearson Correlation	.049	-.058	.053	.084	.005	1
	Sig. (2-tailed)	.432	.358	.392	.180	.935	
	N	258	258	258	258	258	258

Table 7.1 Matrix showing the correlation between the DHS index components and the child diarrhoeal disease variable

7.2.2 Index Evaluation through expert reviews

The use of expert judgment can provide useful information on index validity, given the absence of publicly available data that addresses the spatial distribution of drinking water accessibility or quality in Egypt. Expert opinions and judgments are therefore necessary to evaluate the constructed SDWI. Consequently, the SDWI output was assessed by the two Egyptian experts who participated previously in the derivation of weightings for the index components. Using this approach, a set of questions was developed to allow these experts to comment on the spatial patterns of the SDWI map across all Egyptian districts (kism/markaz). A letter including a map depicting the SDWI values for all Egyptian districts was sent to the Egyptian experts (appendix 4). They were asked to comment on and evaluate the output index values, particularly the geographical patterns observable across Egyptian regions and governorates. Their responses and comments were overall that the spatial distribution of index values reflected the general socioeconomic situation in Egypt. For instance, urban district scored higher values than rural districts especially rural areas in Upper Egypt, where overall human development is less than in urban and rural areas of Lower Egypt. The distribution of index values for Greater Cairo was seen by both experts as comprising two patterns: east with greatest access and west with least access. However, in general, the Greater Cairo region having the highest index values was seen as consistent with the benefits of the central administrative management associated with the capital. Both experts agreed that the index values for Alexandria Governorates reflected the real geographic situation. The districts located in the west alongside Maryut Lake and the desert scored lower values while the more modern urban areas in the east scored higher. Regarding the interrelationship between the index values and diarrhoeal disease, they commented that they expected a significant correlation. However, as noted earlier, it is known that diarrhea is caused by other transmission pathways, particularly food-borne, and furthermore, data reliability (e.g. recall errors concerning diarrhea episodes) could impact on the correlation observed.

7.2.3 Summary of evaluation

Because of the absence of a single agreed measure of safe drinking water access, the evaluation of the SDWI was a difficult task. The diarrhoeal disease question from the DHS was used as the only available data but this indicated no correlation. However, the finding regarding this health outcome-based evaluation (a non-significant correlation) does not invalidate the index for many reasons. For instance, the rationale for constructing the SDWI is not only investigation of the relationship between access to safe drinking water and population health. In a broad sense, there are many uses of the SDWI and the most important benefit is developing a new means for measuring access to safe drinking water using datasets commonly available in developing countries. The methodology outlined in this study for creating a measurement of safe drinking water access is based on the development of GIS techniques to construct an index which quantifies variations in safe drinking water access at sub-provincial scale (district level).

Moreover, the index draws on a wider range of data than currently used by the United Nations for drinking water monitoring in the developed world. The linkage of census and survey datasets gives great opportunities to generate a variety of indicators and components and then combine them into a single multivariate index. As a result, the impacts of several socioeconomic, demographic, environmental and geospatial factors are identified, taken into account and included to provide a reasonable measurement.

In the Egyptian context, the SDWI highlighted the spatial distribution of drinking water accessibility using several indicators. It provided a reliable picture about the current situation of water supplies and sources. Investigating the output index values, it was clear that their geographic patterns were consistent with real water issues across Egyptian regions. Inadequate water supplies and access to unsafe sources were identified in most of rural, slums and desert districts. Spatial variations among urban districts were identified also across the entire country. Districts within urban governorates scored higher than other urban districts. Urban districts located in the east of Greater Cairo region scored higher than others located in the west and south. Rural Upper Egypt scored lower than rural of Delta Governorates. Interestingly, a geographic pattern with low scores presented alongside the Rosetta Nile branch and within rural of El Behera Governorate in the west of Delta and another pattern with similar characteristics in the east of the Damietta Nile branch and within rural areas of El Sharkya Governorate.

The SDWI has raised some serious issues relating to measuring safe and unsafe drinking water accessibility. Yet measuring safe drinking water accessibility in developing countries lacks a robust method which quantifies the drinking water situation for small areas and covers the entire country. As a result, constructing the SDWI and producing a map of household accessibility to drinking water at sub-national level can fill this gap, particularly through developing GIS based analysis techniques and using datasets that are available in most low or lower middle income developing countries. By constructing the SDWI for Egypt, there may be an opportunity to monitor change in the drinking water situation in the future using the same kind of datasets or perhaps adding others to them as they become available. This will be useful work to explore the spatial pattern of household drinking water accessibility, sources, and availability.

7.3 The international context

This section examines the international issues associated with household access to safe drinking water in lower middle income countries. Particular attention will be given to the DHS and census data availability in these countries. It will discuss the seventh goal of the MDG which highlights access to potable and improved sources of drinking water and how to halve the number of households who do not have access to adequate drinking water sources in developing countries by 2015. It investigates the global factors that affect access to safe drinking and how far these differ between Egypt and other countries where the index could potentially be deployed.

7.3.1 DHS and census data availability in other lower middle income countries

The main objective of this particular research is to develop an index which measures access to safe drinking water using existing datasets from both DHS and population census. The index should be applicable internationally particularly in the developing world and low income countries. Developing an integrated water index at sub-national scale (below province) which could be used globally is still a major challenge. However, many difficulties arise when an attempt is made to investigate measurement of drinking water access internationally because of socioeconomic variations impacting safe water access, capacity and government policy to achieve the MDGs targets and database availability across low income countries. The index will be applicable for many developing countries in Asia, Africa, South America and even Eastern Europe. The DHS is conducted over 85 countries worldwide particularly in lower income developing countries (MEASURE DHS, 2008a). As result, DHS datasets for these countries are available and can be used with national census data for constructing the SDWI and monitoring drinking water situation by updating the index (Figure 7.3).

Taking Europe as an example, achieving the Millennium Development Goal concerning access to safe drinking water remains an important issue in urban and rural communities of Eastern Europe. According to the United Nations, for instance whilst urban Egypt achieves almost 100% access to improved water sources, urban Russia 98% coverage, urban Moldova 96% coverage and urban Azerbaijan 88% coverage. The Albanian DHS suggests just 73% of the population have safe water access (institute of Statistics, 2010). Factors that potentially compromise urban water quality in Egypt also occur in Eastern Europe. Whilst 3% of urban Egyptians lack access to improved sanitation (a potential contamination hazard for drinking water), 12% of urban Romanians lack improved sanitation. Similarly, the water supply interruptions recorded for some urban Egyptians have also been reported by 5% of urban Moldovans (institute of Statistics, 2010).

The DHS is undertaken in many lower income Asian countries that are located in Central, South, west and Southeast Asia. As an example of these countries is Kazakhstan and Uzbekistan in Central Asia, Afghanistan and Bangladesh in South Asia and Indonesia and Sri Lanka in southeast. In west Asia region, the DHS is conducted in Azerbaijan, Jordan and Yemen. According to the UN official site for the MDG indicators, in Kazakhstan the total proportion of household who access safe drinking water decreased from 96% in 2005 to 95% in 2008 while in Uzbekistan the proportion of households use improved source of drinking water decreased from 88% in 2005 into 87% in 2008. Although the worst situation was in Afghanistan, the proportion of households who have access to safe drinking water increased from 41% in 2005 to 48% in 2008. In Bangladesh the UN reported that 80% of households accessed to improved source of drinking water in 2008 while the official figures of Bangladeshi census refer to 89% of households using tube-wells as a source of drinking water in 2005 (BBS, 2011).

The DHS is conducted in almost all Sub-Saharan Africa countries and it is also conducted in Egypt, Morocco and Tunisia in North Africa. For instance, in Kenya 56% of household had access to improved sources of drinking water in 2005 while this proportion increased to 59% in 2008 (UN,2011).

According to the Central Statistical Agency of Ethiopia (CSA) 28% of household had drinking water taps at home (CAS, 2011) while the UN reported that 35% of household accessed to improved sources of drinking water in 2005 and this proportion increased to 38% in 2008 (UN,2011). It seems that most African countries suffer from drinking water problems particularly in rural areas and poor urban settlements. In the countries of Latin America and Caribbean the DHS is undertaken in 15 countries such as Brazil, Ecuador, Haiti, Mexico and Peru. A higher proportion of households (95%) had access to improved source of drinking water in Brazil in 2008 and it was slightly lower in Ecuador (94%). A similar proportion was reported by the UN in Mexico (94%) while it was 63% in Haiti in the same year (UN, 2011). The data sources manipulated to construct the SDWI have become available for many developing countries. Egyptian communities share a common challenge with other communities of lower income in ensuring safe water access in the face of water scarcity and population growth. Monitoring safe water access is one component in meeting that challenge.

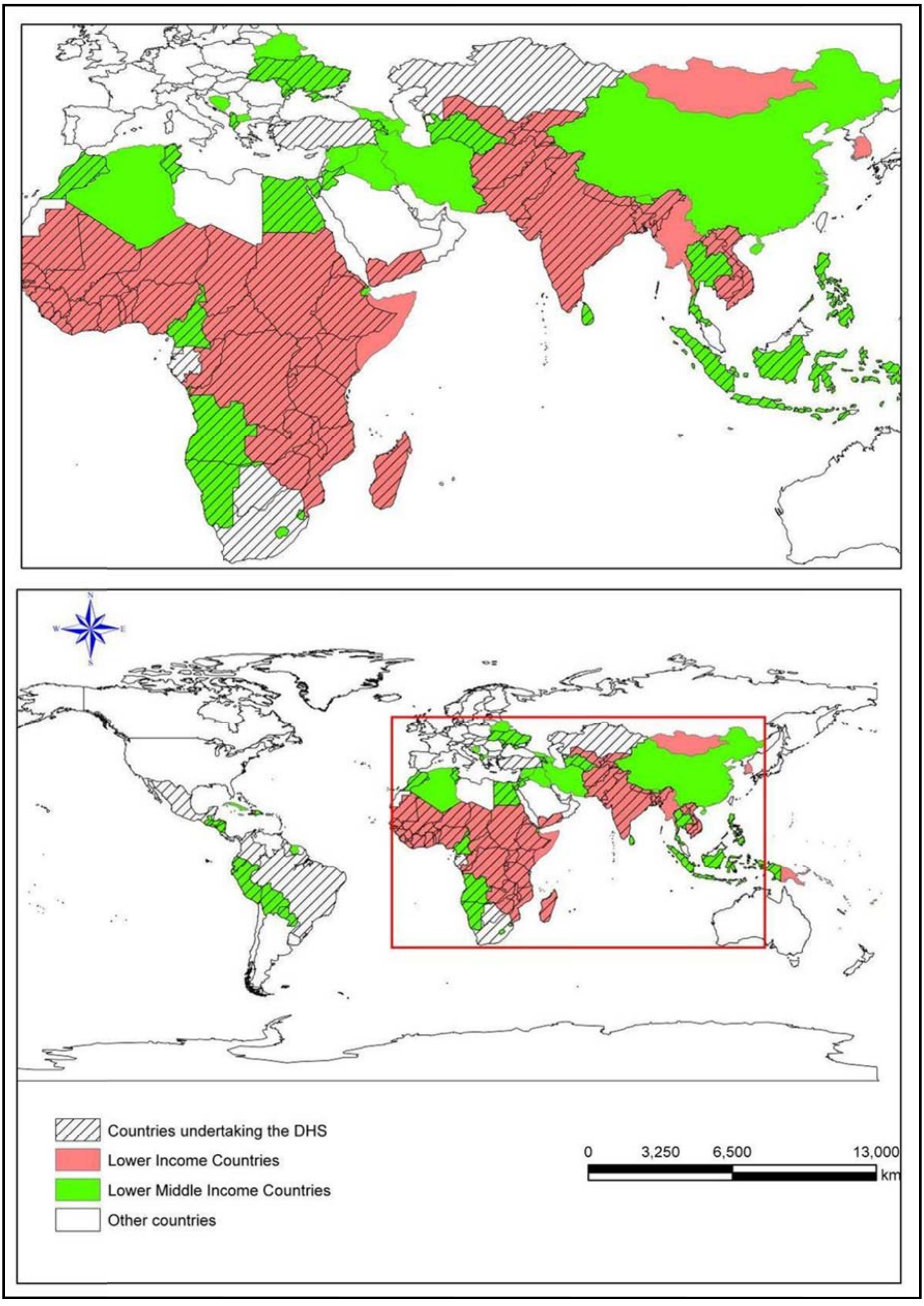


Figure 7.3 Lower and lower middle income countries undertaking the DHS survey

7.3.2 International progress towards MDG Target 7C

Safe drinking water accessibility is a key issue in population health and is the focus of target C of the seventh Millennium Development Goal (MDG) (Hunter et al. 2009). Halving the number of households with access to unsafe drinking water sources by 2015 is one aim of the seventh goal and it is also considered a challenge.

Looking at the situation of access to safe drinking water in the light of Target 7C, Egypt has been in line with the goal. According to the United Nations Joint Monitoring Programme in 1990, the baseline for the MDGs, 90 % of Egyptian population had access to improved water sources (urban 96%, and rural 86%). In 2008 99% of the Egyptian population had access to safe drinking water (urban 100 % while rural was 98%). This indicates that Egypt is on track for meeting Target 7C. However, other low and lower income countries where the DHS is undertaken and the SDWI could potentially be deployed may have not achieved the same progress. For instance, in Ethiopia only 38% of households had access to improved sources of drinking water in 2008 while only 48% of households in Niger had access to improved sources. In Asia, while in Azerbaijan total 80% of households had access to an improved source of drinking water in 2008, the percentage of total households in Cambodia was 61% (UN, 2011). Therefore, whilst in Egypt it is apparent that the SDWI will produce a very different sub-national picture of safe water access to the JMP methodology, in other countries with lower improved source use, the difference between a JMP-based and SDWI safe water access map may not be so apparent.

7.3.3 International variation in factors affecting water safety

Household accessibility to safe drinking water is threatened by various factors, which can be classified into two main categories, national and international factors. National factors include supply interruptions, water contamination and the nature of the public distribution system. International factors include climate change influencing hydrological systems, transnational water transfers and related geopolitical conflicts. In the Egyptian case, some of these factors currently have direct impacts on sustainable accessibility to safe drinking water such as water interruption and distribution systems, contamination of underground water by sewage and poor sanitation system while other factors particularly climate change and geopolitical conflict associated with the Nile water allocation will certainly affect the water situation in the future. Since the SDWI is intended to be a measure of present-day water access, such longer-term threats to water safety are excluded from it. Furthermore, the national level factors that influence water safety and might be associated with water accessibility can vary from country to country and this could affect the transferability of the index from Egypt to other low and middle income countries. For example, in general, at national level, water contamination risks are dominated by microbiological pathogens, whose presence is measured using indicator bacteria (e.g. *E. coli*) and this appears true of Egypt. However, there are other countries that could adopt the SDWI where chemical contamination by fluoride, arsenic, and nitrate is also important, affecting the quality of water sources particularly groundwater. For instance, fluoride and arsenic contamination is quite widespread in Bangladesh and Pakistan (Alam et al. 2003; Farooqi et al. 2007) and nitrate contamination in the Philippines and Thailand (Bouman et al. 2002;Asadi et al. 2002). Whilst microbiological contamination rates were used to score water supply types with different values in Egypt, chemical contamination rates wouldalso therefore need to be considered in some other low and lower middle income countries. In other words, the local factors associated with each individual country need be incorporated in the measurement of safe water access.

The pattern of available data may also vary, as well as the factors affecting water safety and accessibility. In some other countries, such as Bangladesh and Cambodia, contamination from poorly regulated chemical industries is a particular potential threat to water safety while underground water is a major source for supplies. In such countries, one might consider using a map layer depicting the concentration of such industries, if one were available.

In conclusion, the factors that are important for water safety in Egypt will not necessarily be the same in other countries, e.g. in terms of chemical versus microbiological contamination. The pattern of available data may be different in other countries too, e.g. there may be additional ancillary geospatial data that could be used to construct an index in other countries. Therefore, the SDWI would likely need some adaptation to the local context in at least some of the low and middle income countries described above.

The previous global factors emphasis on the importance of water management strategies particularly water supply, quality, quantity and accessibility in Egypt. Thus, the generated index provides a firm method to measure and evaluate drinking water accessibility across the entire countries at sub-national geographic scale. The constructed SDWI has a powerful role can play in terms of drinking water policies such as the availability of safe drinking water sources, problems of household access to these sources, the relationship between the current accessibility situation and health and wellbeing issues, and identifying the poorest districts where greater efforts are needed to provide adequate supplies.

7.4 Transferability case study – Jordan

This section addresses transferability of the SDWI to other countries that have the same geographic and socioeconomic characteristics. Transferability as a concept is defined as the degree to which the data, method and findings of a study are transferable and applicable to other different settings or cases (Antonanzas et al. 2009). Jordan was chosen to be the transferability case study of the SDWI and the reasons for this choice will be explained. The Jordanian datasets including census and DHS will be described and spatial linkage of these datasets will also be discussed. The method of construction of an SDWI for Jordan in a similar way to that for Egypt will be explained and possible index components evaluated. Although the transferability of the SDWI focuses mainly on Jordan, an additional example from Ethiopia concerning DHS GPS clusters and their spatial uncertainty will also be presented.

7.4.1 Justification for choice of Jordan

Although there are many African countries in the Sub-Saharan and North African regions where the DHS is conducted, Jordan was chosen as a transferability case study for several reasons:

- *A neighbouring country with a similar climate and DHS data*: DHS surveys have been conducted in two countries that share borders with Egypt. Of these, Sudan where in theory there is sufficient water since it receives regular rainfall (Gleick, 1993). Thus it does not suffer from water resource scarcity compared to Egypt and its allocation from the Nile River covers water demand for irrigation and drinking. The other country, Jordan, has an arid climate and the annual rainfall is exceedingly variable. Furthermore, Jordan is considered one of the ten most water-scarce nations on the world (Potter & Darmame, 2010).
- *Similar water source type*: Drinking water comes mainly from underground sources (80%) while the remaining percentage comes from surface waters. Drinking water demand is increasing and many Jordanian households buy drinking water from tanker trucks (WHO&UNICEF, 2010b).
- *Similar population growth and water demand*: Various factors impact on water scarcity in Jordan, particularly aridity, shortage of rainfall and population growth which increases the demand of drinking water and affects access to safe drinking water (Abdulla & Al-Shareef, 2009). As a result, there is a need for an

SDWI there to represent the spatial distribution of household access to potable drinking water and highlight areas that suffer from problems of drinking water supplies. The Jordanian population was 6 million in February 2010 and it is estimated that it will be 7,500,000 in 2020 (DOS, 2011). With such population growth, water management and identifying improved sources of drinking water are real challenges for Jordanian households (Jaber et al.1997). Jordan is water scarce and there has been increasing demand for drinking water in the domestic sector which impacts other sectors particularly agriculture and industry (Potter & Darmame, 2010).

- *Similar census data*: Similar to the Egyptian census, the Jordanian census contains data about households' main source of drinking water. In addition, data about household access to sewage and sanitation systems are also collected.
- *Similar population distribution*: There are some similarities between Egypt and Jordan regarding the spatial distribution of population since a high proportion of the total population live in the capital. For instance in Jordan 37% of Jordanians live in Amman and while the Egyptian population almost all live in about 8% of the country's area, the Jordanian population almost all live in 10% of the total area of Jordan. Most of Jordan's population is concentrated in four governorates (Amman, Balqa, Zarqa and Madaba).
- *Similar adaptations to water shortages*: Jordan has a drinking water shortage problem where the consumed water quantity is less that 100 liter per household per day. Furthermore, supply systems suffer from interruptions and consequently water quality is influenced by water storage, particularly tanks that are made of steel, galvanized fiber glass or concrete (Sunna, 2003). This storage system is quite similar to what is found in the Northern Delta region in Egypt. In conclusion, Jordan as it is similar to Egypt it seems to be the most appropriate country to examine the transferability of the Egyptian SDWI.

7.4.2 Jordanian data sets: census geography and DHS

Jordan occupies an area of approximately 96,188 km² including the Dead Sea. Most of its population live in the western part of the country which is a Mediterranean climate (hot and dry in the summer and wet in the winter). Most of the country (75%) has a desert climate with less than 200 mm of annual rainfall. It is divided into twelve governorates (figure 7.5). The capital is Amman with an area of 8,231 km² and its population was 1,824,177 in 2004. The smallest governorate is Jarash (401 km²) while the largest governorate is Maan (33,163 km²) and its population was 96,179 (DOS, 2004). The Jordanian governorates are divided into 51 sub-national administrative units (districts) and each unit is named Lewa in Arabic and it is equivalent to the kism/markaz in Egyptian census geography. These districts (figure 7.6) are used for statistical purposes in the Jordanian censuses and surveys. The lower-level administrative boundary of Jordan consists of sub-districts named Kada in Arabic which are equivalent to shyakhas in Egyptian administrative geography. Each Lewa (district) is divided into one or more Kada and the total number of sub-district units is 89 across the country.

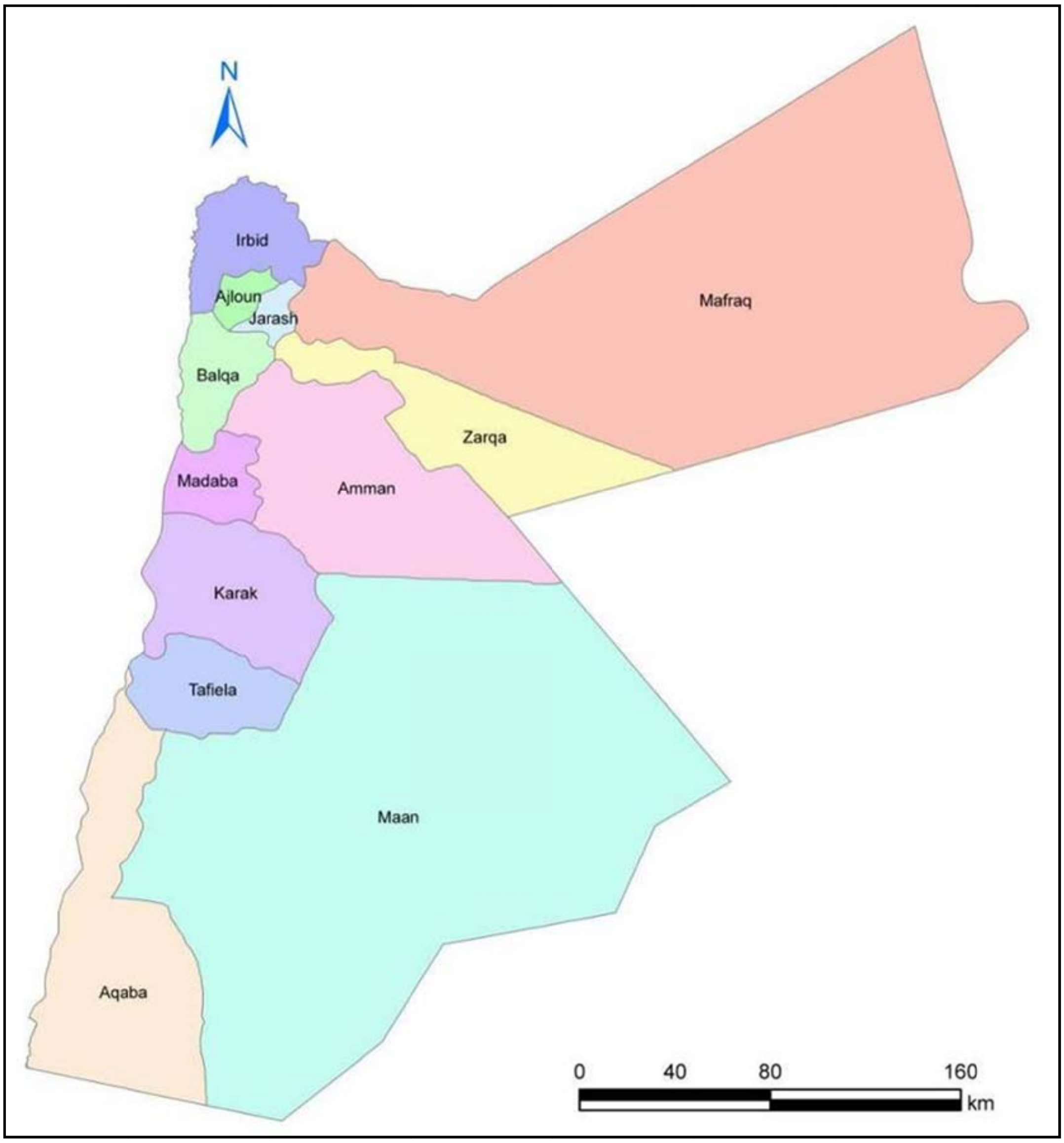


Figure 7.4 Boundaries of Jordanian Governorates

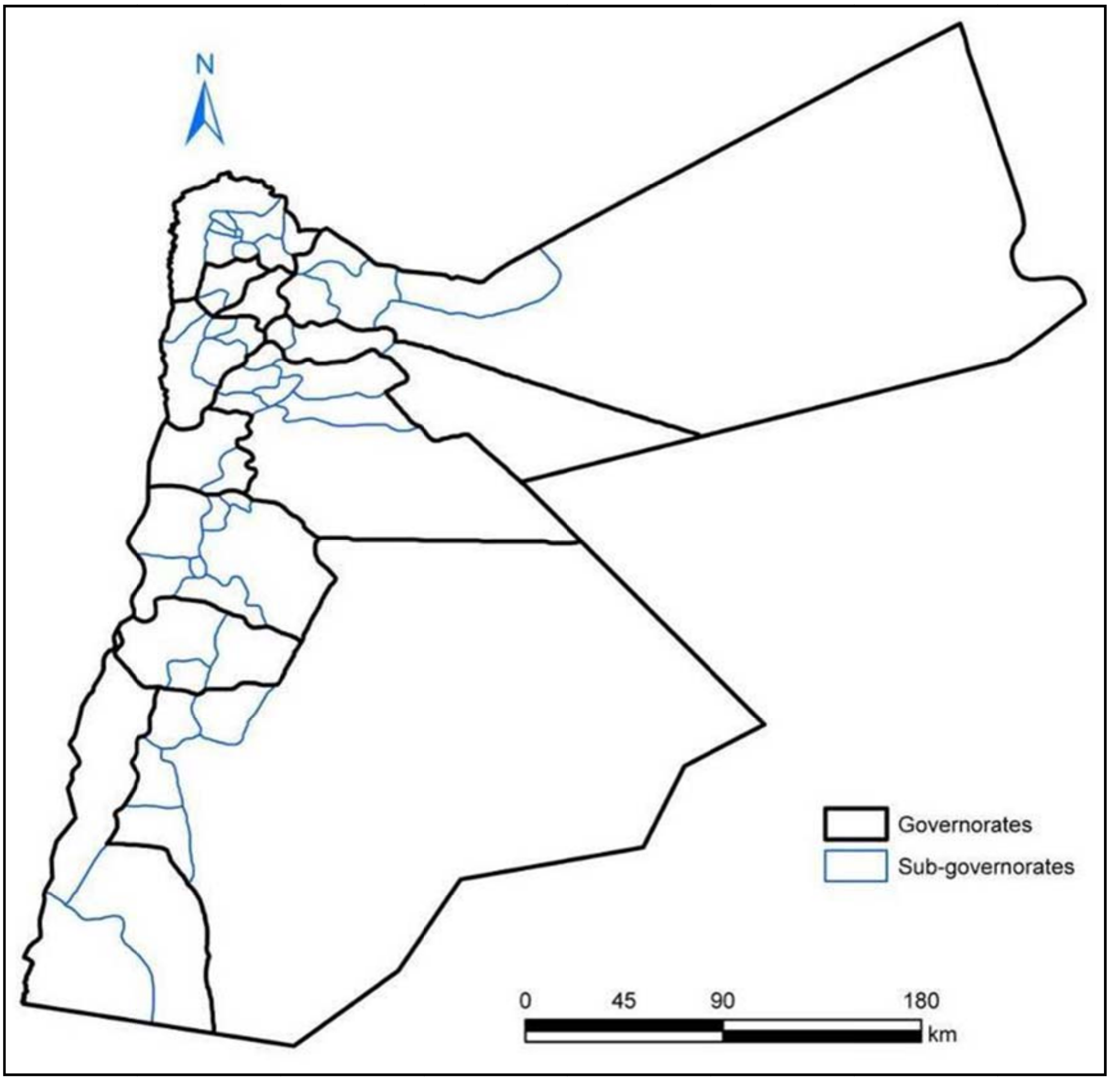


Figure 7.5 Jordanian sub-governorate boundaries

Historically, five censuses were undertaken in Jordan, with the first housing census conducted in 1952 while the second one was carried out in 1961. The third census of population and housing was conducted in 1979 while the fourth one was carried out in December 1994. On the second of October 2004, the Department of Statistics (DOS) undertook the fifth Jordanian population census (DOS, 2004). The Jordanian census aimed to collect a comprehensive statistical dataset covering Jordanian population characteristics and households, particularly their socioeconomic and demographic characteristics based on the lower-level geographic units across the country. The final 2004 Jordanian census reported that Jordanian population was 5.1 million compared to 4,130,950 in 1994 (DOS, 2004). The DOS has provided a clear definition of four major places and housing unit types in the Jordanian census and they are as follows:-

- **Locality** which refers to a large group of permanent or temporary housing unit for households which differ in terms of population size and consist of one block or more.
- **Block** which is defined as a group of houses or buildings.. Its boundaries are clearly defined and it is connected to all kinds of public services and infrastructure such as drinking water and electricity networks, highways and railways etc.
- **Building** which is a permanent or temporary construction on land or water and made of different materials. A building could consist of one or more floors and ceiling used for residency or work. There are many types of buildings for instance, an Amarah (building which is used for residential purposes only) or a Dar which is a traditional house that could be old or modern in fashion.
- **Rural and urban**, the 2004 Jordanian census distinguished between rural and urban areas based on the total number of population living in a locality. Each locality with 5000 population or more is considered an urban unit while all localities that have a population less than this are considered rural.

Although the Jordanian census is clearly organized and its framework and procedures are well defined compared to the Egyptian census, it shares some limitations and drawbacks with the Egyptian census in that it does not collect health data. Both Egyptian and Jordanian censuses provide data about population disabilities and health services and insurance, but these types of datasets are less valuable and not helpful for addressing population health, morbidity or mortality phenomena in a water-related context.

The 2007 Jordanian Demographic and Health survey (JDHS) (also known as the Population and Family Health Survey (JPFHS)) was conducted by DOS in collaboration with Macro International Inc with technical assistance provided by MEASURE DHS. The survey was funded by the Jordanian government, USAID, UNFPA and UNICEF (DOS and Macro International Inc, 2008). As part of the DHS program three standard surveys (JDHS 2002, 1997, 1990) were conducted before the 2007 JDHS and another one was conducted in 2009. Similar to other DHS, the JDHS is nationally representative. The sample drawn was 14,564 households and 10,876 married women aged 15-49 were interviewed in the survey. The total number of clusters is 930 although four of these lack coordinates, and so the 2007 JDHS contains 926 valid GPS cluster points distributed across three regions and twelve governorates. The average size of a cluster is 74 households in urban areas, which is where 83% of Jordanian population live (urban localities having 5000 residents or more) while the average is 62 household in rural areas. The overall average size of a cluster is 72 households (DOS and Macro International Inc, 2008). Figure 7.7 illustrates the Jordanian DHS geography according to the 2007 JDHS. It can be seen that the density of DHS GPS clusters is higher in the north and central regions than in the south. This reflects the population distribution across the country since 38% of the population is concentrated in the capital (central region) and 34% in Irbid (north region) and Zarka (central region) governorates while 28% of the population is distributed over the reminder of the country.

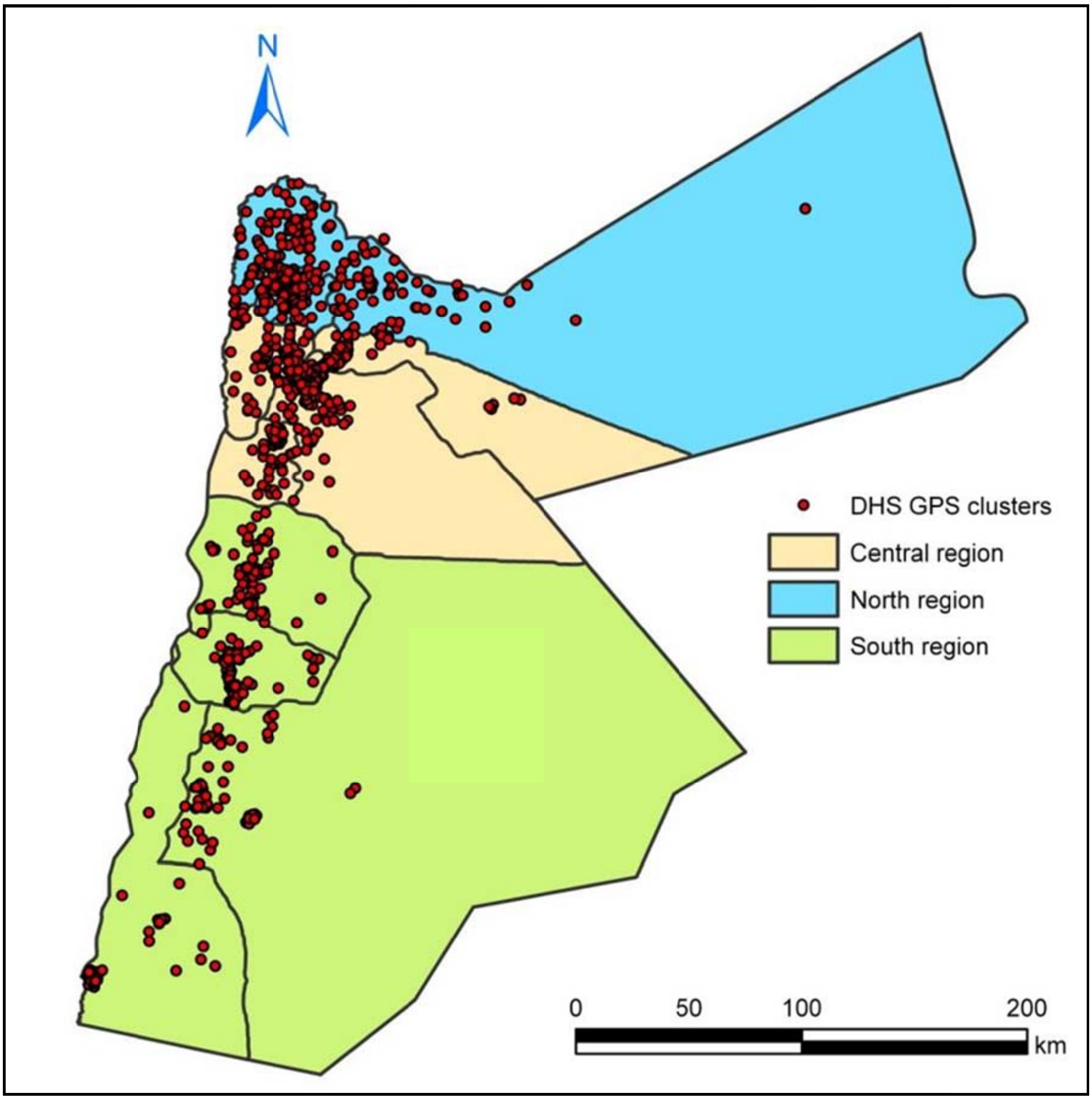


Figure 7.6 Distribution of DHS GPS clusters by main Jordanian regions

7.4.3. Data linkage issues in Jordan

Assessing spatial uncertainty in DHS GPS clusters is an important step before undertaking any spatial linkage between the GPS dataset and administrative boundaries. I shall here introduce another example (similar to the Egyptian case) of spatial error and uncertainty associated with the DHS GPS clusters locations. Furthermore, this error which might be present in the Jordanian GPS clusters needs to be measured, modeled and evaluated to assess the influences of uncertainties on the spatial linkage process. To investigate this issue in the 2007 Jordanian DHS geo-referenced dataset, the GPS cluster points and governorate census polygons were overlaid in ArcGIS to examine to what extent there is agreement between governorate administrative labels and the spatial location of GPS clusters. A spatial join on the two layers was undertaken in ArcGIS to identify such spatial error. Figure 7.7 demonstrates the detected spatial error and misallocated GPS clusters across Jordanian governorates. The findings indicated that there was a discrepancy between some of the DHS governorate labels for GPS clusters and census governorate polygons. Overall there were 103 GPS clusters out of 926 that lay outside the governorate boundaries that corresponded to the DHS provincial attribute label for the cluster.

Figure 7.8 and table 7.2 presents the number of GPS clusters with spatial error in each governorate and their incorrect locations by governorate. According to this method, the percentage overall accuracy for GPS clusters in Jordan was higher (11%) than the Egyptian case (4%). No such errors were identified in the GPS cluster points belonging to Aqaba, Karak, Maan and Tafilah governorates where all clusters were located in their correct governorate. Thus GPS clusters within four out of 12 governorates are error free at this administrative level. On the other hand, the 103 misallocated GPS clusters belong to seven governorates (Ajlun, Amman, Balqa, Jarash, Madaba, Mafrak, and Zarka) lying in incorrect positions. The greatest percentage of these cluster points (41%) was in Jarash (40 clusters points), whilst 37 cluster points lie within Ajlun, two points within Balqa and one within Irbid. The lowest number was in Irbid (one cluster point) while 17 clusters points belonging to Ajlun were identified as entirely lying within Irbid. 15 cluster points belonging to Amman lie incorrectly within Balqa (13 cluster points) and Madaba (two cluster points). Three cluster points belonging to Balqa were identified lying in Ajlun (one cluster point), Irbid (one cluster point) and a third cluster point lies outside the governorate census polygons. Two cluster points belonging to Madaba were identified in Amman and nine cluster points belonging to Mafraq were identified as misallocated, of which eight were within Irbid (five cluster ponts), Jarash (two points), Zarqa (one cluster point) and one cluster point lay outside any census boundary. 16 misallocated cluster points belong to Zarka and were identified as lying within Amman (10 cluster points), Mafraq (four cluster points) and Jarash (two cluster point). One cluster point belonging to Irbid lies outside any census administrative boundary. Overall, three cluster points lie entirely outside the census boundaries.

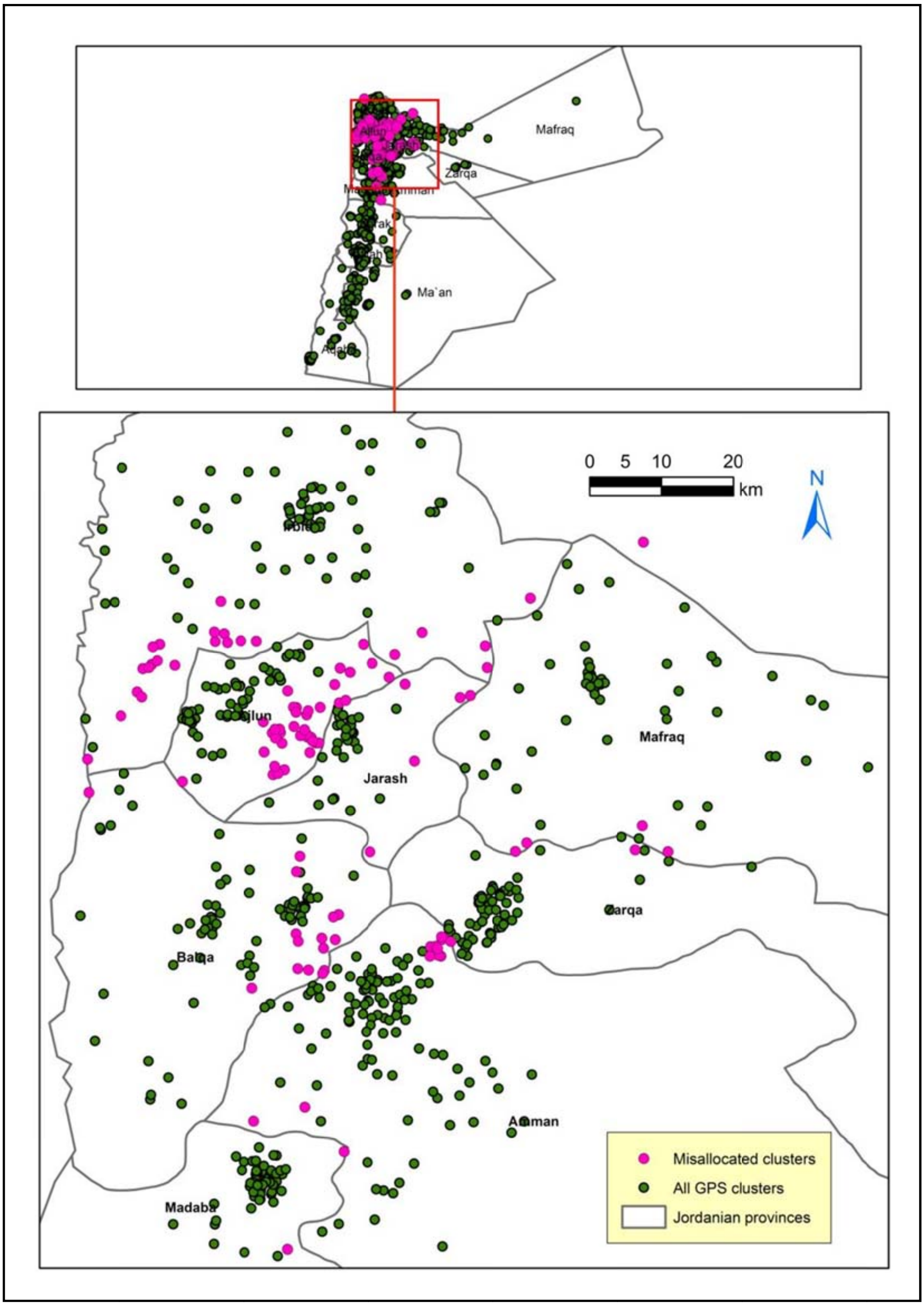


Figure 7.7 Distribution of positional error of GPS cluster points in all Jordanian governorates

Comparing the Jordanian and Egyptian case studies, it seems that there are some differences, particularly in the nature of the spatial distribution of misallocated clusters. It is not entirely clear why the misallocated GPS clusters were not distributed across all Jordanian governorates, but only exist in eight governorates out of twelve. Nevertheless, it can be speculated on the causes behind these differences as follows:

- The DHS GPS clusters that were free of positional errors belong to four governorates (Aqaba, Karak, Maan and Tafilah) located entirely in the south region. They are large in area but have lower population density compared to the central and north regions (see figure 7.6) and this may explain this pattern.
- The misallocated GPS clusters appeared to gather around the governorate census boundaries. Furthermore, misallocated GPS clusters belonging to governorates that are located in the east (e.g. Amman) fall incorrectly within governorate located in the west (e.g Madaba and Balqa) and vice versa.

	Ajlun	Amman	Aqaba	Balqa	Irbid	Jarash	Karak	Maan	Madaba	Mafraq	Tafilah	Zarqa
Ajlun	72	0	0	1	0	37	0	0	0	0	0	0
Amman	0	119	0	0	0	0	0	0	2	0	0	10
Aqaba	0	0	72	0	0	0	0	0	0	0	0	0
Balqa	0	12	0	72	0	2	0	0	0	0	0	0
Irbid	17	0	0	1	75	1	0	0	0	5	0	0
Jarash	0	0	0	0	0	72	0	0	0	2	0	2
Karak	0	0	0	0	0	0	72	0	0	0	0	0
Maan	0	0	0	0	0	0	0	72	0	0	0	0
Madaba	0	3	0	0	0	0	0	0	72	0	0	0
Mafraq	0	0	0	0	0	0	0	0	0	72	0	4
Tafilah	0	0	0	0	0	0	0	0	0	0	72	0
Zarqa	0	0	0	0	0	0	0	0	0	1	0	84

Correct Misallocated

(DHS governorates are columns and spatial join results are rows)

Table 7.2 Confusion matrix illustrating the number of GPS clusters in each governorate and those that have governorate labels inconsistent with their locations

- Three cluster points are entirely outside any census boundary and they belong to three governorates: Balqa in the west of Jordan (central region), Mafraq in the north east of Jordan (northern region), and Irbid in the north west of the country. This means that positional error of misallocated GPS cluster is likely to be unsystematic and randomly distributed.

- As there were some GPS clusters blocked further away from the nearest administrative boundaries (Figure 7.7), it was necessary to check the revision history of Jordanian provincial boundaries. However, the last change happened in 1996 where the Ajlun and Jarash governorates split from Irbid. Aqaba split from Maan, whilst Madaba split from Amman. A strip of land in the northeastern part of Amman was transferred to Zarqa (Statoids, 2011).

Using the ESRI world imagery service (Figure 7.9), the GPS cluster layer was overlaid with Jordanian census polygons. Checking whether there is any systematic displacement of sub-district Jordanian census, no any displacement was found particularly in the isolated and uninhabited areas. Looking at the distribution of misallocated GPS clusters in another developing country such as Ethiopia (Figure 7.10), it can be seen that the positional errors of GPS clusters are widely distributed across the entire country and most of the misallocated GPS clusters fall near census administrate boundaries and thus this distribution pattern is quite similar to the Egyptian example.

Although there were some of misallocated GPS clusters in the Jordanian case that do not follow the distribution pattern of the Egyptian case nor the Ethiopian one, the overall findings indicate that the identified inconsistencies between governorate labels and cluster GPS locations in the Jordan and Ethiopian cases arise because of a positional error similar to the Egyptian case. Despite the difference in the spatial distribution of positional errors between Jordanian and Ethiopian DHS GPS clusters, the scrambling process undertaken by Measure DHS and distance to nearest census boundaries may be effective variables in measuring, modeling and evaluating spatial uncertainty. As a result, it may be problematic to undertake any spatial linkage of Jordanian DHS GPS clusters to census administrative boundary without addressing and measuring the GPS uncertainty and modeling spatial errors associated with the GPS dataset.

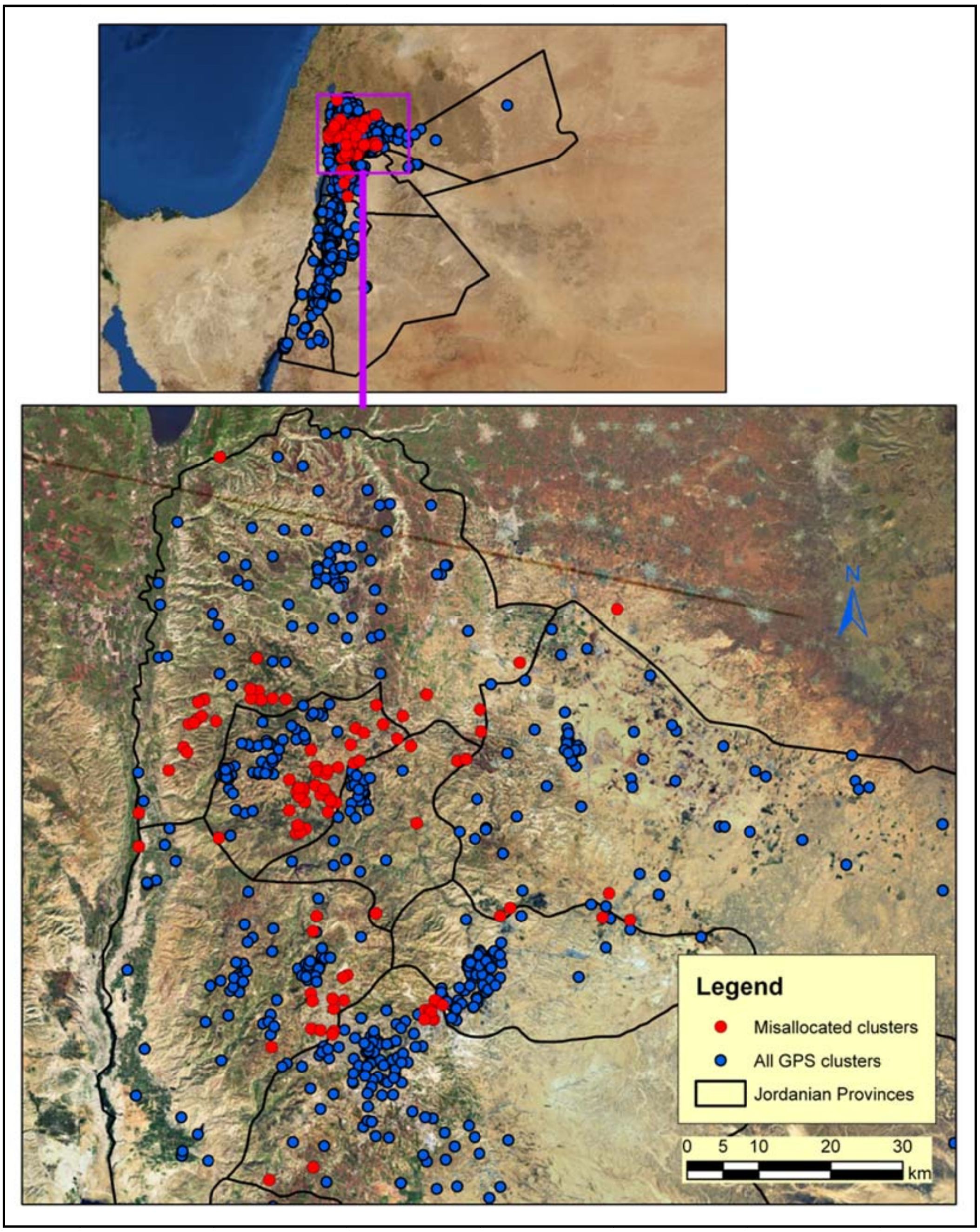


Figure 7.8 Distribution of misallocated GPS cluster points compared with Jordanian governorate boundaries using ESRI ARCGIS online world imagery

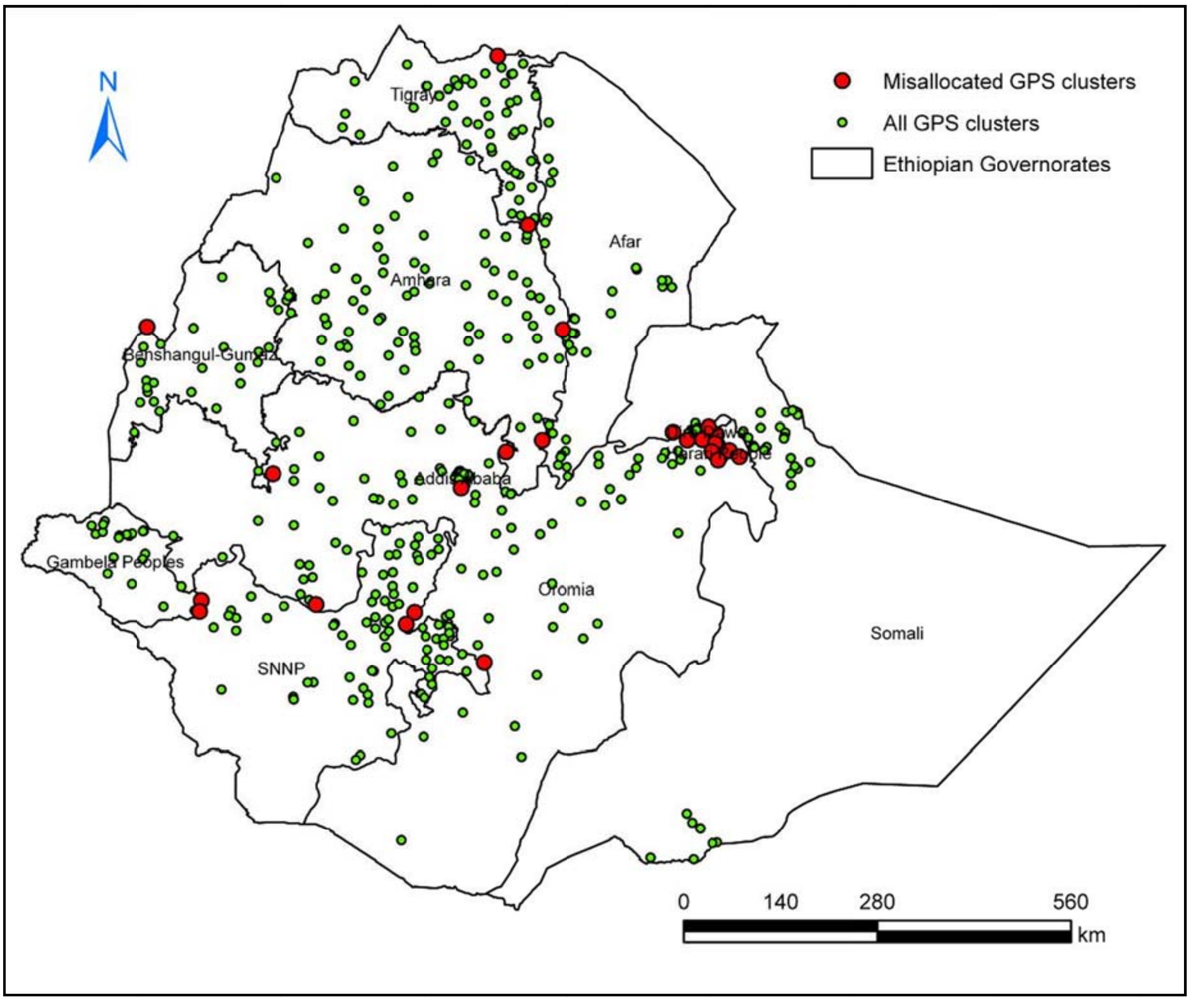


Figure 7.9 Distribution of misallocated GPS cluster points in the 2005 Ethiopian DHS compared with Ethiopian Governorate boundaries

7.4.4. Index construction issues in Jordan

Exploring the available datasets from both Jordanian DHS and census, five domains can be generated similar to the SDWI for Egypt. These domains are environmental, spatial, socioeconomic, water and demographic domains that include the most relevant parameters related to household safe water accessibility. However, the sampling uncertainty in district-level parameter estimates and consequent effects on the interpolation process have not been investigated here. Greater sampling uncertainty or weaker spatial autocorrelation would make the index difficult to implement in Jordan.

All index indicators from the Jordanian DHS are virtually identical to the Egyptian DHS while the difference was expected to be in the census dataset. However, the Jordanian census provides data about the spatial distribution of Jordanian households according to the source of drinking water and also based on access to different types of sanitation.

Table 7.3 shows a comparison between the SDWI components for Egypt and possible index components for Jordan. From this table it can be seen that datasets about socioeconomic, demographic and education domains provided by Egyptian and Jordanian are virtually identical while some differences are found regarding sanitation system and water supply types. There are four sanitation types available in the Egyptian census (public network, local, *tranch*, and offline) while just three types are recorded in the Jordanian census (public network, Cess Pool, and none). There are six supply types in the Jordanian census (public network, tanker, well or rain water, Artesian well, bottled water, and spring) while in the Egyptian census just three types exist (public network, pumps, and well). Thus, the distribution of households according to source of drinking water covers a wider variety of sources in the Jordanian census, particularly bottled water, tanker, and spring types that are not collected in the Egyptian case. These three types are also available in the DHS and therefore, the supply type component from Jordanian census data would need some modification to be relevant to access to safe drinking water in Jordan. Similarly, the sanitation system component would need some modification from the Egyptian version to distinguish between households connected to public sewage and those who use cess pool systems which might be a potential risk factor for groundwater safety.

According to the scoring methodology developed in the SDWI, each indicator can be assigned a score (numerical value). For instance, a score is assigned to each water source type where this value reflects overall compliance with microbiological and chemical parameters. In the Egyptian SDWI, drinking water sources were scored nationally while in the Jordanian case study, utility piped supply types can be scored at provincial level. Reviewing the literature, Jordan is one of six countries (China, Ethiopia, Jordan, Nicaragua, Nigeria, and Tajikistan) selected by WHO/UNICEF to implement a pilot project on Rapid Assessment of Drinking Water Quality (RADWQ). The methodology in this project was based on the MICS survey design and it is intended to generate nationally representative water quality statistics. It follows a clustered sampling approach which is based on selecting individual water sources across all Jordanian governorates (WHO/UNICEF,RADWQ, 2010). Table 7.4 shows the overall compliance of Jordanian utility piped water supplies for all governorates according to WHO and national standards. The compliance was calculated for the coliforms, arsenic, fluoride and nitrate. Therefore, it is possible to incorporate a wider range assessment of water quality parameters than in the Egyptian case. Generally Jordanian governorates scored highly as 97.8% met the WHO guideline values. Four out of twelve governorates (Amman, Zarqa, Mafraq, and Ajlun) scored less than (100%). The remaining components of the index could be scored similar to the Egyptian case study. For example there were two possible ways to score water treatment types, either based on the influences on diarrhoeal disease or on water quality. In the Egyptian case, water quality was the preferred approach. Water quantity was scored according to the time taken to reach water source while water interruption was scored due to frequency of water unavailability. On the other hand the sanitation type component was scored according to whether the type is improved or not (see section 5.3 in chapter five).

Jordanian census indicators					Egyptian census indicators				
Demographic domain	**Education domain**	**Socioeconomic domain**	**Environmental domain**	**Water domain**	**Demographic domain**	**Education domain**	**Socioeconomic domain**	**Environmental Domain**	**Water domain**
Population density	Enrolment Status	Building characteristics	**Sewage type:** public network Cess Pool None	**Supply types:** Public network Tanker Well/Rain water Artesian well Bottled water Spring	Population density	Enrolment Status	Building characteristics	**Sewage types:** public network Local sanitation Tranch offline	**Supply types:** Water tap inside house (public network) Water tap inside building (public network) Water tap outside building (public network) Pumps Wells
	Educational Institutions by Grade,	Economic Activity Status,				Educational Institutions by Grade,	Economic Activity Status,		
	Enrolled in Educational Institutions	Employment Status and Governorates				Enrolled in Educational Institutions	Employment Status and Governorate		
	Education level	Current Occupation and Governorates				Education level	Current Occupation and		
		Distribution of Employed Persons					Distribution of Employed Persons		
		Sector of Work and Governorates					Sector of Work and Governorate		

Table 7.3 variable and their values in Egyptian and Jordanian censuses that could be used to characterise the SDWI domains

Jordanian Governorates	Overall compliance	
	No. of assessed samples	Compliance with WHO GVs (%)
Ajlun	611	94.4
Amman	146	100.0
Aqaba	244	99.6
Balqa	25	100.0
Irbid	270	100.0
Jarash	73	98.6
Karak	73	93.9
Maan	49	100.0
Madaba	73	100.0
Mafraq	25	100.0
Tafilah	25	100.0
Zarqa	25	100.0
National	1639	97.8

Table 7.4 Overall compliance of Jordanian utility piped water supplies with WHO guideline values (GVs)

Source: RADWQ, 2010, p21

7.4.5 Options for evaluating a Jordanian index of safe drinking water

As with Egypt, it would be essential to evaluate the proposed SDWI index for Jordan. For this purpose, fieldwork could be undertaken for a short period to explore patterns of drinking water access in a sample of districts with differing index values. Such fieldwork could be similar to that undertaken at the start of this study (section 3.2.4 in chapter three). Since water interruption and storage as a resultant coping strategy were reported in Jordan (Sunna, 2003) it would be worth assessing supply and tanker storage systems and their impacts on water contamination, quality and public health. Similarly, contacting Jordanian stakeholders who work for drinking water or public health organizations might be helpful in evaluating factors relevant to drinking water issues and assessing or judging the SDWI results. Moreover, experts and stakeholders may provide valuable advice on health or other ancillary data sources that might be available, especially those associated with water borne diseases or dataset about water supply types, interruption and, water home treatment. These three dimensions (field work, stakeholders and experts, health data from local institutions) could potentially play a significant role in evaluating the SDWI in Jordan.

7.5 Summary

In this chapter, the SDWI was evaluated through the interrelationship between the SDWI values and diarrhoeal disease, which is influenced by unsafe water accessibility. However, DHS-derived child diarrhoea data are only the available source of health data, an evaluation of the index values against such data was limited because of various weaknesses, notably that there are many other diarrhoeal disease transmission pathways. In addition to this, the question about diarrhoea in the DHS is subject to recall error. Consequently, there was not a significant correlation between the index values and the percentage of children who had been infected by diarrhoea across the sampled Egyptian districts (kism/markaz). The index outputs were also evaluated using Egyptian stakeholder judgment. In broad terms, the variation across Egyptian districts depicted by the index appeared to show geographically realistic patterns according to these stakeholders. For instance, poor access and low safe water index values were generated in desert and rural districts and within those governorates that do not border on the Nile River as a major drinking water source (Sharkya). On the other hand, the safest access was found in urban governorates but districts within Greater Cairo and Alexandria scored higher than other urban districts.

The chapter then explained the international context behind the SDWI, particularly the drinking water target in the MDGs. It was clear that access to safe drinking water for the MDGs is difficult to measure at sub-national scale. International progress towards Target 7C was discussed, in particular how access to safe drinking water varies between Egypt and other low and lower middle income countries where the DHS is undertaken. Furthermore, the international and national factors that impact on safe water accessibility vary from one country to another according to differing spatial, environmental, and socioeconomic circumstances. There were specific similarities between Jordanian and Egyptian censuses, particularly in collecting data about housing units, household conditions, and water supply types. Furthermore, similarities in terms of water resources were apparent in that Jordan suffers from drinking water scarcity and increasing demand from the domestic sector. The spatial linkage of DHS GPS clusters over Jordanian census polygons indicated that there were 103 clusters with governorate labels inconsistent with their GPS coordinates, out of 926 GPS clusters in total. They were distributed across eight governorates while GPS cluster within four governorates were free of any error. Finally, field work, stakeholder consultation and health data from Jordan were suggested as approaches for evaluation of the proposed Jordanian index

Chapter 8: Conclusion

8.1 Chapter overview

In this chapter, the first section reviews the main findings of this research, particularly the assessment, modelling and analysis of spatial uncertainty in the DHS GPS dataset. It discusses how to undertake a spatial linkage to this dataset with local census data in order to construct the SDWI as a metric and as a practical tool for measuring safe drinking water access. In the second section, recommendations are made for further research which could be conducted to develop measurement of drinking water using a variety of GIS and spatial analysis techniques, in particular new data sources that may be obtained, wider stakeholder consultation, or using the same datasets for different time periods. Policy implications of this research are highlighted, especially the MDGs and monitoring of the drinking water situation in the Egyptian case. In the third section, the strengths and limitations of this research are discussed such as lack of data and metadata, and establishment of the link between health and drinking water. Finally, I briefly outline how GIS methods can play an important role in addressing the future strategic challenge of measuring safe water access in the light of water scarcity and geoconflicts between developing nations over water allocations.

8.2 Main findings

The major challenges for this thesis were to develop GIS based analysis techniques that can be used to create a measurement of safe drinking water access, particularly in developing nations. The measure had to give a general and a clear picture of drinking water access at a sub-national scale and according to a combination of geographic, demographic and socioeconomic factors that may affect household access to potable water supplies. To undertake this measurement method a spatial linkage of two datasets (the DHS and local census) was conducted. Consequently, a validation was undertaken of DHS GPS clusters linked to administrative census boundaries. Positional error was identified and thus this error had to be assessed, measured and modeled before implementing the spatial linkage of the two datasets.

8.2.1 The assessment of spatial uncertainty in the DHS clusters

The literature on DHS applications using GIS and spatial analysis was reviewed to investigate the nature of the DHS data sources and explore the topics covered by the survey, especially those associated with population health and socioeconomic issues. The review demonstrated that there is a growing literature on DHS applications using GIS and spatial analysis techniques, particularly in child mortality, morbidity and disease, family planning, fertility and women's empowerment, and socioeconomic characteristics of households. Collecting household coordinates provides geo-referenced datasets which enrich the utility of the DHS. Using the DHS GPS clusters is considered a significant addition to the survey as it provides great opportunities to link the DHS to other local datasets. Thus, the utility of the DHS can be considerably enhanced by linkage to available census data, which provides coverage of the entire population, albeit with less detailed information on some key health and lifestyle topics. Nevertheless, the DHS GPS clusters should not be directly linked to any other data sources without uncertainty assessment. As there is a deliberate scrambling process for the GPS clusters implemented by the organization which undertakes the DHS, it is therefore necessary to develop appropriate GIS approaches for spatial linkage between the two sources. An assessment of the positional accuracy of the DHS and its implications for GIS-based linkage to census data is essential. The focus should be on the assessment and accommodation of spatial uncertainty in the DHS to census linkage, so as to facilitate an important enhancement to the data available for analysis.

The spatial uncertainty relevant to the geo-referenced DHS data and its impacts on subsequent analysis and data linkage was reviewed. Within the concept of spatial uncertainty there is commonly differentiation between errors affecting well defined spatial objects and vagueness and ambiguity which affect poorly-defined spatial objects (chapter three). Although the data sources (DHS points and census polygons) were well defined, spatial error was identified. This error was described as the difference between the recorded locations or attributes and their true values. The assessment process was based on a point-in-polygon operation, with the points being DHS GPS clusters and the polygons being census boundaries. In the literature, a further sub-category of uncertainty in spatial databases was distinguished by Fisher (1999) that he terms controlled uncertainty, in which uncertainty is deliberately introduced into a socio-economic data set. Given the deliberate perturbation of locational data in the DHS to protect the confidentiality of survey participants, this concept was also relevant to the positional error identified in this study.

Several authors have developed procedures for handling uncertainty in point-in-polygon operations, using probability density functions or fuzzy set theory (Leung and Yan, 1997; Cheung et al. 2004), or by reducing the problem to a point-in-triangle problem (Leung et al. 2004), using Monte Carlo simulation (Hunter, 1999), and comparison of two different methods (geocoding and cadastral) of generating address locations Ratcliffe (2001). However, the previous procedures were not applicable to the spatial uncertainty in this study as no information was available concerning the positional accuracy of DHS GPS point and polygon data. This is not uncommon in developing world contexts. Furthermore, it was not possible to generate locations for point or polygon data via different methods. Therefore, rather than following the methodology of these previous studies, the approach was to identify attribute labels that were inconsistent with the locational data for GPS points and polygons. Information about feature attributes that were inconsistent with location were used to model uncertainty in the undertaken point-in-polygon operation.

When the DHS GPS clusters were overlaid with Egyptian census polygons to examine the agreement between the two dataset, spatial error was detected. A spatial join process was implemented to check the consistency between the DHS GPS identifier and census polygon names. The output indicated that 53 GPS clusters that were coded as belonging to specific governorates were misallocated within other governorates and another 10 GPS clusters lie entirely outside the Egyptian administrative boundaries. Overall, there were 63 GPS clusters whose spatial locations and governorate codes did not agree. Looking at the distribution of these misallocated clusters it was found that their incorrect position were close toadministrative boundaries. It was necessary to ensure that incorrect projection and labeling are not responsible for the detected spatial error and thus the map projection of the two layers were checked and found to be the same. The spatial distribution of misallocated GPS clusters appeared to be related to proximity to census boundaries, which indicated that incorrect GPS labels are not the cause of this positional error.

Measurement of spatial error was done by applying a point-in-polygon distance operation using a combination steps in ArcGIS and Idrisi software to calculate the distance between each GPS cluster and census polygon. As a result, distance between each misallocated GPS cluster and its correct polygon was determined. Analyzing the spatial distribution of the positional error across the entire census polygons, the overall mean distance of misallocated GPS cluster from the nearest census boundaries was 1410m. The minimum distance was 100m while the maximum was 6,122m which means that each GPS cluster located further from boundaries than the maximum distance was within its correct governorate.

The final section of the geospatial data linkage evaluation was to model the spatial error detected in the DHS GPS clusters. Because the outcome variable was binary or dichotomous (the spatial error as a dependent variable) and consists of two possible values (0 as GPS cluster with no error and 1 as GPS cluster with no error), logistic regression was appropriate to model the relationship between the positional error and the independent variables. The first step in the modeling process was testing for global clustering to examine whether the 63 misallocated GPS clusters were clustered or not. To implement this test, Clusterseer software was used as it offers a wide range of methods for testing the pattern of spatial clustering. Cuzick and Edwards' test was applied as a test for spatial clustering in mislabeled GPS clusters. The output of Cuzick and Edwards' k-nearest neighbor statistic suggested that there was spatial clustering in GPS clusters with governorate labels inconsistent with their locations (T=24; k=1; p=0.001). The spatial clustering was generally significant at the 99% level when up to 6 neighbors were considered.

A series of logistic regression models was developed in Stata software using different predictors (GPS survey methodology used, rural versus urban clusters, and distance to governorate boundary) and transformations to improve the strength of the model. Alogistic regression model of distance to nearest governorate boundary was fitted first, but the *linktest* diagnostic suggested that this model was misspecified. Therefore, a square root transformation was applied to this distance variable in the second model to correct for this mis-specification. In the third model, two additional dichotomous covariates were introduced, namely the rural / urban classification of clusters and the survey methodology used to locate the GPS clusters (i.e. based on averaging the coordinates for each household in the cluster or based on a single GPS reading at an approximate cluster centre). Goodness of fit statistics suggested that rurality improved the fit of the model and reduced spatial autocorrelation in residuals, but the survey methodology variable was not significant and was therefore dropped. To arrive at the final model, an interaction term between distance to governorate boundary and rurality was introduced to correct for a further misspecification caused by the introduction of this new variable.

The model explained the presence of misallocated cluster points using two independent variables; square root of distance to the nearest governorate boundary and an interaction variable created by multiplying this distance variable by a dichotomous rural-urban variable (where each cluster point is coded one if it is urban and zero if not). The output of this model indicated that both these distance-related variables are significant and together explain around 40% of the variation in inconsistent governorate labels (pseudo $R^2 = 0.39$). The model also confirmed that the distance a GPS cluster point lies from a governorate boundary affects the probability of it having the correct governorate label and this effect varies between rural and urban areas. The Moran's I test statistic showed that there is very weak (albeit significant) spatial autocorrelation in the residuals of this model (Moran's I = 0.01, Z score = 2.41).

Although individual clusters are not identified by administrative identifiers lower than governorate names (kism/markaz or shyakha), assessment of spatial uncertainty in the GPS clusters at lower levels was undertaken:-

- The logistic regression model derived at governorate level was applied to district level (kism/markaz) data by calculating the distance from each GPS cluster to the nearest census boundary. Consequently, the probability of a point lying in the incorrect kism/markaz was calculated. The findings of this process indicated that kism/markaz differs between urban and rural cluster points.
- 7.9% of the GPS clusters showed a low probability value (less than 0.5) of being in their correct position. Monte Carlo simulation using these probabilities suggested that on average 285 GPS clusters (confidence limits 257 -313) would fall within an incorrect kism or markaz

- To analyse the spatial uncertainty at the lowest census boundary (shyakha level), a new list of sampled shyakhas for the Egyptian DHS was created by matching the shyakha/towns and villages list with the census shyakha layer via geocoding within ArcGIS. Experimentation with geocoding parameters suggested that it was necessary to use low spelling sensitivity, minimum candidate and minimum match parameters since shyakha names had been transcribed from Arabic into English.
- However, the findings of this process revealed that 58% of 647 sampled shyakhas were successfully matched automatically. The remaining DHS shyakha names were matched manually to equivalent census names while 13 out of 647 shyakhas remained unmatched, having no equivalently named units in the census map layer. Furthermore, some shyakhas may have been incorrectly geocoded because of local place name variations and issues in translating Arabic shyakha names into English.

At the shyakha level, 807 GPS cluster points (62% of those with co-ordinates) lay outside the boundaries of the shyakhas listed as sampled for the DHS and a systematic displacement of shyakha census boundaries in some isolated areas was noted, particularly southern New Valley governorate, but it was not possible to identify any overall pattern to this error. Thus, care should be taken in interpreting the error rate in linking GPS cluster points to shyakhas relative to those for kism/markaz and governorate level.

From these experiments with spatial uncertainty analysis at three geographic levels, it can be concluded that positional errors may affect spatial analysis of the DHS GPS clusters. Therefore, spatial linkage to other map layers, particularly small area census data, cannot be used uncritically. The district level is the most appropriate spatial scale at which to undertake spatial linkage between local census boundary and the DHS GPS clusters. As noted earlier (chapter two), there is a growing body of research that enriches DHS data by using GPS cluster locations to link to other spatial data sets, including census data. Nevertheless, none of these studies has examined the effect of locational uncertainty on these linkage operations. It is proposed that the issues raised here are of broad significance when attempting GIS use for analytical data linkage in developing country settings, particularly in the increasingly common situation where hand-held GPS are used to georeference social survey data which must then be linked with existing national sources.

Particularly where HIV data are collected as part of a DHS survey, positional uncertainty may mean that in multi-level modeling studies (e.g. Feldacker et al. 2010), community-level variables are derived from the incorrect higher level spatial units. The findings of this thesis suggest that this effect is likely to be more pronounced where small areas (below district level) are used to derive community-level covariates. This may reduce the ability of such studies to detect contextual effects as opposed to individual-level effects. In small area estimation studies, this same effect is likely to occur and may increase the uncertainty over small area estimates of malnutrition prevalence, for example. Although this would be more difficult to model, positional uncertainty in GPS cluster points is also likely to affect spatial linkage to other types of data, such as land cover or transportation infra-structure, particularly where these vary greatly within the space of several km. Furthermore, in using attribute labels inconsistent with location to model spatial error, this study has contributed to the existing literature on uncertainty and the point-in-polygon problem.

8.2.3 The SDWI as a mean for measuring water access

After the linkage process had been successfully undertaken, work has been undertaken on constructing a tool for the measurement of safe drinking water access. Two concepts, area classification and index construction, were explored as they have been widely used to address the socio-spatial structure of data for small geographic. Making a comparison between the two candidate methods was necessary to choose the appropriate one. Consequently, a critical review of existing research on area classification and index construction was given (chapter four). Despite the fact that geodemographic classification has been applied in a variety of socioeconomic research, few applications in developing countries could be identified and none of these was relevant to drinking water access. The review of existing literature on index construction showed that deprivation indices are popular methods for identifying poor communities in small geographic units and the UK deprivation indices are considered the most popular. In developing countries, the experience of index construction in general was similar to that of geodemographic classification methods, with both being infrequently applied. The Provincial Index of Multiple Deprivation (PIMD) was constructed by Noble et al. (2006) at sub-national scale in South Africa. In addition, the South African Index of Multiple Deprivation (SAIMD) at small area level (datazone) (Noble et al. 2009). As the SDWI aims to measure water access at sub-national geographic scale, a critical review of water indices was undertaken. It became clear that there is an absence of any water index which measures safe drinking water accessibility for small geographic units and which covers an entire country. By contrast, the water indices which could be found focus on water quality, scarcity and availability issues but within physical and environmental contexts. Furthermore, although an accessibility indicator is included in some indices (e.g. water poverty index and index for drinking water adequacy), their coverage is global and the resolutions are national or regional. Hence, this demonstrated a gap in the area of safe water access measurement, particularly associated with the measurement introduced in the MDGs and the way water sources are classified. After comprehensive review and comparison of the area classification and index methods, a decision was made that constructing a SDWI would be a new, valuable and practical measurement for safe drinking water access.

The starting point for SDWI construction was exploration of the available datasets from the DHS and local census to determine the domains relevant to all possible drinking water parameters and development of the index indicators and components. Selection criteria were developed, against which 12 candidate index components were evaluated and assessed. Using these criteria, four potential components were excluded while eight were retained as the potential index components. Except for the sanitation component, it was necessary to develop appropriate scoring methods for each of the index components. Each component was assigned to a numeric value based on relevant literature and Egyptian studies that evaluated drinking water supplies according to microbiological and chemical contamination. Assessment for the normal distribution of scored components was applied and a Pearson correlation matrix was created to measure the degree of association between pairs of components. This showed that there was no significant positive or negative correlation between the index components.

A standardisation method was required to convert all the component values into a single integrated index. Exploring many standardisation methods (e.g. Z score, ranking, Chi square), range standardisation was chosen. As an example of the reasons for this, the distribution of some scored components showed gaps which led to exclusion of the ranking method. The output of the linkage process between census polygons (kism/markaz) showed that 101kism/matkaz were unsampled by the DHS and thus their characteristics needed to be estimated. Before undertaking this estimation process, the conceptual methodology behind estimating the unknown values of the unsampled DHS districts was reviewed and discussed. Spatial interpolation was chosen as the most appropriate method to be used. It primarily predicts unknown values according to surrounding known values, which was the case in the DHS unsampled districts. To quantify the influence of sampling design of the DHS on the estimated district-level values, the Stata software was used to estimate district-level values for all relevant DHS components taking into account the survey sample design and weighting scheme.

Tobler and Kennedy’s (1985) areal interpolation method was implemented to predict the unknown DHS values at the level of the Egyptian kism/markaz. The method was appropriate since it gives a weight for each neighbouring polygon (with known DHS value) based on its shared boundary length to polygons whose DHS value is unknown. However, ArcGIS did not offer any command to calculate boundaries length and a contiguity tool for ArcMap using VBA was downloaded and used to produce these boundary lengths. Estimation was completed using a combination of Microsoft Excel and Access software. Except for 12 kisms/markazes that are islands in the Nile River or new cities in the desert and which therefore do not share boundaries with any other polygons, all unknown DHS values were successfully estimated. The correlation between the interpolated and observed values was measured and was significant (Pearson’s R = 0.634).

A weighting scheme was developed to reflect the importance of components in terms of their contributions to measuring safe drinking water access. An international consultation consisting of two Egyptian experts and one from the UK rated each component. A numeric weight was then calculated based on the mean of these three scores. The weights for all index components from the DHS and census were applied and the components were integrated into one index.

A map of the SDWI for all Egyptian districts was created to present the SDWI values for all Egyptian districts. This showed that there was spatial variation across the entire country but geographic patterns could be seen in almost all Egyptian regions. These patterns reflected the nature of spatial variations of household water access. Overall, the safest access was found in urban districts particularly in Greater Cairo and more specifically in the east of the region. On the other hand, the poorest access was found in desert and rural districts but this varies between Lower and Upper Egypt. A group of clustered rural districts demonstrated the least safe access in the region of East Delta, particularly distributed across El Sharkiya governorate. Similarly, rural districts of El Behera governorate in the west of the Rosetta Nile branch showed the least safe access. Spatial variation also was recognized across Upper Egypt where districts located in the south demonstrated higher values than those in the north.

Evaluating the SDWI values against the percentage of child diarrhoeal disease at district level, no significant correlation was found. This could be the result of many reasons. Particularly, the literature indicated that diarrhoeal symptoms are associated with many sources of which drinking water is just one and it was difficult to find a direct correlation between diarrhoea and drinking water. The way the question about diarrhoea is posed in the DHS might be another limitation. In addition, the DHS is undertaken seasonally while children can be infected in other periods during the year according to a variety of environmental and local factors. Furthermore, health data and data about water quality are not available to researchers in most of the developing countries, either at sub-national or even provincial scale. Regarding the transferability of the SDWI, the drinking water situation in developing countries similar to Egypt was addressed, particularly lower middle income countries. Problems of access to safe drinking water and adequate sanitation were highlighted by the UN in many regions in many countries within Asia, Africa, South America and Eastern Europe. As the DHS is conducted in 85 developing countries and these datasets are available, they can be linked to national population censuses in order to construct an SDWI. The problems associated with household access to potable drinking water and the availability of the

DHS datasets with local population census globally suggest that the SDWI will be applicable and that the newly developed method for the measurement of drinking water access could be transferred to other developing countries.

The selected case study for the index transferability (Jordan) was addressed and investigated. Jordan was chosen especially as it is one of the ten countries which suffers the greatest water scarcity globally (Potter & Darmame, 2010). Moreover, there are similarities between the Egyptian and Jordanian censuses in terms of collecting and providing data about conditions of household living, particularly the spatial distribution of households according to source of drinking water. Jordanian census geographies are similar to Egypt, consisting of three levels (governorate, district, and sub-district). Spatial linkage of the Jordanian DHS GPS clusters to census boundaries was conducted. Positional error was identified, whereby 103 GPS clusters were not allocated to their correct governorates and three lie entirely outside the census boundaries. Nevertheless, GPS clusters that belong to four governorates (Tafilaf, Maan,Irbid, and Aqaba) were free of spatial displacement. Although the distribution of misallocated GPS clusters in the Jordanian case study was different in detail from the Egyptian one, errors in both cases arise because of positional error. The scrambling process which is applied by the MEASURE DHS to protect participant confidentiality was a major cause of this error. ESRI world imagery service was used to overlay the GPS cluster layer over census boundaries to ensure that the distribution of misallocated clusters is the same as it was in the previous overlay process. No change was found relating to the distribution of misallocated Jordanian GPS clusters. For further validation of this issue, spatial linkage was undertaken for another developing country. The GPS clusters of the 2005 Ethiopian DHS were joined to census boundaries and the findings indicated similar positional error whereby 25 GPS cluster out of 529 were not allocated to their correct governorates. The distribution of Ethiopian misallocated clusters was again similar to the Egyptian case since, misallocated GPS clusters were found very close to census boundaries while those located far away from boundary lines were in their correct governorates.

Exploring the available census dataset for Jordan to address the possible SDWI indicators, it was clear that the number of water supply sources provided in the Jordanian census (public network, tanker, well/rain water, Artesian well, bottled water, and spring) is higher than those in the Egyptian census (public taps, well, pump). Thus, a water domain could be generated similar to the water domain in the Egyptian case but with richer census data. The scoring methodology developed in the Egyptian SDWI can be used effectively in the proposed Jordanian SDWI. In addition to this, Jordan is one of the countries selected by WHO/UNICEF to implement a pilot study on Rapid Assessment of Drinking Water Quality (RADWQ). In the published report of this project, the compliance of Jordanian water supplies with various water quality parameters at governorate basis can be used as scoring method for public water supplies, while other water components can be scored similar to the Egyptian SDWI.

8.3 Recommendations for further research

- Regarding data sources, a strong recommendation to the MEASURE DHS organization would be to adapt the data modification algorithm to avoid gross misallocation between governorate –level polygons. For example, the algorithm could be modified to prevent GPS cluster locations being moved across governorate boundaries as a result of the perturbation of their coordinates which would potentially enhance the utility of the data.
- The SDWI could be successfully constructed using other survey data sources such as MICS, the AIDS Indicator Survey and Malaria Indicator Surveys (both are run by the Measure DHS consortium), Living Standards Measurement, and other possible spatial data that are collected locally. These sources can be linked to population censuses. Similar methods to that used in this research could be followed to assess spatial uncertainty which might affect the linkage and after that the SDWI components could be generated.
- A variety of small area estimation methods could be used to estimate values for unsampled DHS districts, either simple methods that do not require advanced statistical knowledge (e.g. director estimator and calibration methods) or more sophisticated small area estimation methods such as regression methods.
- Building a framework for a wider consultation of experts is needed as it provides significant in-depth insight for developing and generating the indicators. This would be most beneficial if the selected experts have a variety of experiences relevant to the index domains (water, health, socioeconomic, demographic and environment etc...). For instance, it would be useful if there is expert input relevant to each index domain as this would assist in evaluating and weighting the components of the SDWI.
- Problems of access to safe drinking water emphasize the need for increasing the engagement of a variety of specialists, academics and planners. Thus, involving local and international planners with interests in sustainable resource

development in the refinement of an SDWI would contribute towards the development of a stronger measurement of safe water accessibility.

- A variety of problems are associated with drinking water in developing countries such as water scarcity, rapid population growth, increasing water demand and climate change. Accordingly, there is a need to develop a system which focuses on monitoring access to safe water geographically. This should particularly consider socioeconomic factors affecting the drinking water sector. The DHS is conducted approximately every three years, which provides a great opportunity to establish an effective monitoring system for household water accessibility at small community level and across entire countries. Through constructing the SDWI regularly, drinking water accessibility and health outcome of households can be monitored and assessed.
- As an alternative to the methodology used in this study, nonlinear combination of the index components can be applied. One of the benefits of applying nonlinear combination is that it can capture the interaction of two or more factors that impact on water access. For example, a high percentage of households using pit latrines for sanitation may only seriously affect water access in combination with high population densities (where the 'carrying capacity' of a groundwater body is exceeded) (Cronin et al. 2007). High pit latrine use alone or high population density alone would compromise water safety far less. A non-linear combination of these two factors would capture such influences better than linear combination. Despite such advantages of non-linear combination, neither the deprivation indices nor water indices constructed to date have applied this technique yet.

8.4 Policy implications

Within the context of increasingly complex demographic and environmental factors affecting access to safe drinking water in the developing world, the findings of this study suggest a series of policy implications for the measurement of safe drinking water, particularly relevant to the MDGs which run till 2015. Most important are meeting the MDG associated with drinking water and halving the number of households with access to unsafe water sources. However, this research has identified several weaknesses not only related to the definition and classification of safe and unsafe (improved and unimproved) water sources (chapter two section 2.2.1) but also in the methods and approaches that are followed to monitor the progress that has been made by each country. Although there are many socioeconomic and environmental factors that affect drinking water accessibility such as population growth, sewage type, household educational status and household wealth – and most of them are shared by developing nations – measuring and monitoring drinking water access to assess progress towards the MDGs has never included these factors. The national policies for each country are different and thus an integrated tool to measure access to safe water is needed. The MDG monitoring programs lack any measurement based on small geographic areas and covering entire countries. This means that the MDG monitoring has never considered many significant spatial variations, not only between rural and urban households but also among small communities within each region or governorate within one country. The SDWI provides an integrated approach, capable of capturing spatial variations in safe drinking water accessibility among small communities and internationally.

Egypt is confronted by a challenging situation both nationally and internationally, as scarce water resources limit the available amount of water in the face of rapid population growth which increases water demand. In addition to this, various socioeconomic and geographic factors influence the current and probably the future situation of drinking water supplies. The findings of the SDWI constructed in this research indicated that households living in many rural and desert kisms/markazes suffer from poor accessibility to safe drinking water. Meanwhile, several urban districts within Greater Cairo and other urban governorates also showed lower values of access to safe drinking water. Although this probably results from different factors, establishing public policy and strategies for monitoring the progress of improving household accessibility to adequate water supply is needed. Except for the “2017 water resources plan” which was set up by the Ministry of Water Resources and Irrigation in March 2005 (ESIS, Water Resource, 2008) and focuses mainly on water resource management, canal and drainage systems, no other future plan has been found relating to household access to safe water supplies in Egypt, nor to the spatial distribution of water supply types. To date, no spatial research has been published to address geographic and socioeconomic factors impacting on drinking water and health of the population in Egypt. Also, drinking water issues like water interruptions and quantity still have not been investigated from an academic perspective. Consequently, the Egyptian government should make efforts to establish an independent official organization for the drinking water sector. This organization could consist of academics including geographers, planners, stakeholders and experts in water issues. The major aims of this organization could be addressing, measuring and monitoring the improvement in drinking water availability and accessibility at small geographic scale and across all Egyptian districts and villages.

8.5 Limitations and research gaps

The lack of data about drinking water and population health in Egypt and most developing countries is not an uncommon challenge. This limits many aspects of spatial research, particularly constructing accurate measurement of drinking water and public health. The principal limitation of this research was mainly associated with lacking geospatial and attribute data about water quality, availability and public health of the population at any geographic scale. Moreover, the spatial census data lack metadata that describe their characteristics and sources. For example, in Egypt no attempt has yet been made to create a link between the GIS unit and census department in CAPMAS. Consequently, the two datasets (census attribute and spatial datasets) are created separately and provided by different units. As a result there were difficulties in undertaking successful geocoding for shyakah level analysis as a part of the uncertainty analysis of the DHS GPS clusters. This was caused mainly by transcription of polygons names from Arabic into English. Similarly, a lack of metadata in the vector census layer may have an impact on applying GIS and spatial analysis processes particularly when information about data source, quality, accuracy, representation, and maintenance are absent.

The definitional discrepancies related to socioeconomic variables were another limitation. These variations are related to the attribute DHS and census datasets, especially water sources and sanitation types. Neither Egyptian nor Jordanian censuses provided clear definitions of these types and categories.

An important shortfall was found in understanding the link between drinking water contamination and health. Thus, a further limitation was to understand the influences of sanitation risks on underground water. No clear definition exists for either the private or *tranch* sewage types which are recorded as sanitation systems in the census data. Although they are in general kinds of pit or traditional latrines (the way they are constructed poses a potential hazard to groundwater which is the main source for most pumps, wells and even water plants), it was hard to assess or evaluate risks to groundwater. Source type-specific water contamination rates were used as a method to score water supply types. However, as a result of the lack of any published research relevant to some of these sources, difficulties arise when scoring specific supply types such as tankers and carriers. Overall, addressing the impacts of drinking water on public health is limited by data availability. For instance, water-borne diseases could be an important indicator of the link between water quality and population health but such data may be not reliably gathered and when they are recorded, are not available for researchers. This thesis has argued that collecting geo-referenced data about water quality can help in establishing a clearer link between drinking water and public health. Consequently, research addressing this relationship by the use of more robust GIS techniques should be developed.

8.6 The future for monitoring safe drinking water access

Future efforts should be made to use advanced GIS techniques in order to develop practical, accurate, quantitative measurements of safe water access. The capabilities of powerful GIS tools enable effective measurement and analysis of drinking water accessibility for small geographic units and, potentially, can enable further research for identifying its influence on public health. The use of innovative GIS tools will be a significant addition in developing strategic planning for water policies. The water indices constructed to date lack geospatial and socioeconomic factors that impact on access to safe water. Monitoring safe water access in the face of increasing water scarcity increases the need to strategic planning for the future, particularly with regard to the following challenges:

Population growth: Future strategic planning of water policy should give more attention to insufficient and poor quality drinking water supplies in developing nations. Unsafe water supplies are driven by population growth, deprived urban and rural housing conditions, and slum expansion. Meanwhile, the increase of population means more settlements are needed to accommodate new households. As a result, expanding coverage of public water supply networks is crucial to deliver safe water to households in both urban and rural areas. This thesis has also argued that population growth has not been yet included in the measurement of safe drinking water access. Monitoring household access to safe drinking water should include reference to demographic problems as a key challenge of provision of potable water supplies.

Climate change: Iglesias et al. (2007) argued that there have been recent episodes of drought in the Mediterranean Sea Region that affect the demand for drinking water and make it higher than the available water sources, particularly in the populous countries like Egypt. As a result, the Mediterranean countries especially in Africa and Asia confront global environmental challenges, particularly the water scarcity problem and how to reduce the proportion of households who have access to unsafe drinking water.

The impacts of climate change on potable water resources are expected to affect drinking water resources in the future (Brookes et al. 2010; Vorosmarty et al. 2000). The effects of global and environmental hazards on water resources vary among developing countries and thus water shortages vary also from one country to another. For instance, rising sea level in coastal areas might affect groundwater resources as safe water supplies. Moreover, the expected change in rainfall and floods could impact water quantity by reducing aquifers recharge. Similarly, climate change might impact on drinking water infrastructure and utilities for example as flooding may damage pipes or as soils fracture and dry. Consequently drinking water quality, quantity, availability and accessibility will be influenced by these changes. Therefore, taking greater account of environmental factors in the measurement of access to safe drinking water will be necessary when the SDWI is constructed.

Water demand from agriculture: Use of water resources is related to domestic, industrial and agricultural demand. However, agricultural demand is by far the largest consumer of water resources, particularly when irrigation systems are traditional (Playan & Mateos, 2006; Araus, 2004). Hence, measuring and monitoring access to safe drinking water should consider the influences of water demand in agriculture and crop production. For instance, available data about the amount of water for irrigation and its source in each rural district could be used to capture the competing agricultural demand for drinking water resources. Moreover, using these data can generate useful indicators about water demand for agriculture as a part of the SDWI. Taking this variable into account will be crucial in understanding how various environmental and spatial factors affect the future of access to safe drinking water.

Geopolitical conflict: Allocation of water from Nile River is a clear example for geopolitical conflict. Ten countries share the basin of the Nile River: Burundi, Egypt, Eritrea, Ethiopia, Kenya, Rwanda, Sudan, Tanzania, Uganda, and the Democratic Republic of the Congo. About 300 million people live within the ten countries and it is expected that this number will double within the next 25 years (Kameri-Mbote, 2007; El-Fadel et al.2003). In the last three years, a conflict has developed between the countries of the Nile River Basin about allocation of river water. While downstream countries (Sudan and Egypt) have the right to use 100% of the water that enters their territory, Egypt must consent to other countries' use of the Nile's water according to two agreements. The first one was signed in 1929 between Egypt and Britain (as a colonial power in Sudan, Kenya, Tanzania and Uganda), while the second was between Egypt and Sudan in 1959. Upstream countries, particularly Ethiopia, demanded the revocation of these two agreements and a new system of basin water management (Kameri-Mbote, 2007). It is clear that any decreasing of Egypt's allocation from the Nile River will influence Egyptian irrigation systems and decrease available water quantity per capita per day, particularly given rapidly population growth in Egypt and the Nile as the major source of drinking water.

Generating an indicator or index component about geopolitical conflict is not possible as the SDWI coverage is at sub-national geographic unit. However, sustainable access to safe drinking water may be influenced by water conflict and tensions over water resources and allocations. Therefore, countries that rely on a trans-border river or lake as a major water source and do not receive sufficient rainfall, such as Egypt, should establish water management policies which longitudinally monitor access to safe water, taking into account the international factors impacting water availability. Planning for long term policy to measure and monitor access to safe drinking water should incorporate GIS analysis. An initial development for the SDWI could be expanding the measurement components by incorporating more factors and applying new spatial methods and thus signifying new directions for future research on drinking water utilisation and public health

Reference list

ABBAS, J., OJO, A. & ORANGE, S. 2009. Geodemographics--a tool for health intelligence? *Public Health,* 123**,** e35-9.

ABBASPOUR, R. A., DELAVAR, M. R. & BATOULI, R. 2003. The Issue of Uncertainty Propagation in Spatial Decision Making. *In:* Paper presented at Scandinavian Research Conference, , 2003 Finland, http://www.scangis.org/scangis2003/papers/13.pdf, accessed on 20th April 2009.

ABDEL-SHAFY, H. & ALY, R. 2002. Water issue in Egypt: resources, pollution and protection endeavors. *Central European Journal of Occupational and Environmental Medicine Hungary, avialable at* http://www.omfi.hu/cejoem/Volume8/Vol8No1/CE02_1-01.html, accessed on 12th April 2011.

ABDULLA, F. A. & AL-SHAREEF, A. 2009. Roof rainwater harvesting systems for household water supply in Jordan. *Desalination,* 243**,** 195-207.

ABO-AMER, A., SOLTAN, E. & ABU-GHARBIA, M. 2008. Molecular approach and bacterial quality of drinking water of urban and rural communities in Egypt. *Acta Microbiologica et Immunologica Hungarica,* 55**,** 311-326.

ABOUZAHR, C. & BOERMA, T. 2005. Health information systems: the foundations of public health. *Bulletin of the World Health Organization,* 83**,** 578-583.

ABS. 2006. *Australian Bureau of Statistics, National Statistical Service, Small Area Estimates Manual Version 1.0,* available at http://www.nss.gov.au/nss/home.nsf/pages/small+area+estimates+sections+1-4?OpenDocument, accessed on 10thOctober 2010

ALAM, M., SNOW, E. & TANAKA, A. 2003. Arsenic and heavy metal contamination of vegetables grown in Samta village, Bangladesh. *The Science of the total environment,* 308**,** 83-96.

ALEMAYEHU, T., LEGESSE, D., AYENEW, T., MOHAMMED, N., WALTENIGUS, S., UNEP, N., UNESCO, P. & XU, Y. 2006. *Degree of groundwater vulnerability to pollution in Addis Ababa, Ethiopia*, Taylor & Francis/Balkema, Leiden (The Netherlands).

ALI, E. E. D. 2003. Evalution of The Egyptian Experment In establishing The New Towns In The Desert Areas. *Journal of Engineering Sciences, Assiut University,* 31**,** 1-14.

ALLISON, P. D. 2001. *Logistic regression using the SAS system: theory and application,* SAS Institute, North Carolina, USA.

ALLO, N. 2010. Disaggregating the Nigerian postcode: A step to creating an environment for geomarketing in developing countries, In the proceedings of the GIS Research UK (GISRUK) , the 18Annual Conference 2010, University College London, London, UK on 14-16 April 2010. GIS Research UK

(GISRUK), 183-191

AMIN, S., BASU, A. M. & STEPHENSON, R. 2002. Spatial variation in contraceptive use in Bangladesh: Looking beyond the borders. *Demography,* 39**,** 251-267.

ANGUELETOU-MARTEAU, A. 2008. Informal water suppliers meeting water needs in the peri-urban territories of Mumbai, an Indian perspective, Global Changes and Water Resources: Confronting the Expanding and Diversifying Pressures :XIIIth World Water Congress, MONTPELLIER : France (2008).

ANTONANZAS, F., RODRÍGUEZ IBEAS, R., JUÁREZ, C., HUTTER, F., LORENTE, R. & PINILLOS, M. 2009. Transferability indices for health economic evaluations: methods and applications. *Health Economics,* 18**,** 629-643.

ARAUS, J. 2004. The problems of sustainable water use in the Mediterranean and research requirements for agriculture. *Annals of Applied Biology,* 144**,** 259-272.

ARNOLD, B. & COLFORD JR, J. 2007. Treating water with chlorine at point-of-use to improve water quality and reduce child diarrhoea in developing countries: a systematic review and meta-analysis. *The American Journal of Tropical Medicine and Hygiene,* 76**,** 354.

ASADI, M. E., CLEMENTE, R. S., GUPTA, A. D., LOOF, R. & HANSEN, G. K. 2002. Impacts of fertigation via sprinkler irrigation on nitrate leaching and corn yield in an acid-sulphate soil in Thailand. *Agricultural water management,* 52**,** 197-213.

ASHBOLT, N. 2004. Microbial contamination of drinking water and disease outcomes in developing regions. *Toxicology,* 198**,** 229-238.

ASHBY, D. & LONGLEY, P. 2005. Geocomputation, geodemographics and resource allocation for local policing. *Transactions in GIS,* 9**,** 53-72.

ATTARAN, A. 2005. An immeasurable crisis? A criticism of the millennium development goals and why they cannot be measured. *PLoS Med,* 2**,** e318.

Baschieri, A. (2007). Effects of modernisation on desired fertility in Egypt. Population, Space and Place 13(5): 353-376.

BAFFREY, R. M. N. 2005. Development of program implementation, evaluation, and selection tools for household water treatment and safe storage systems in developing countries, Massachusetts Institute of Technology

BARNES, H., WRIGHT, G., NOBLE, M. & DAWES, A. 2007. The South African Index of Multiple Deprivation for Children Census 2001, Cape Twon, South Africa.

BARTRAM, J., LEWIS, K., LENTON, R. & WRIGHT, A. 2005. Focusing on improved water and sanitation for health. *The Lancet,* 365**,** 810-812.

BASCHIERI, A. 2007. Effects of modernisation on desired fertility in Egypt. *Population, Space and Place,* 13**,** 353–376.

BATEY, P. & BROWN, P. 2007. The spatial targeting of urban policy initiatives: a geodemographic assessment tool. *Environment and Planning A,* 39**,** 2774-

2793.

BBS & BANGLADESH BUREAU OF STATISTICS 2011. Percentage distributions of households by sources of drinking water and residence, available at http://www.bbs.gov.bd/RptHIES_3_3.aspx, accessed on 10th April 2011.

BEAUMONT, J. & INGLIS, K. 1989. Geodemographics in practice: developments in Britain and Europe. *Environment and Planning A,* 21**,** 587-604.

BENNETT, R., HAINING, R. & GRIFFITH, D. 1984. Review Article: The Problem of Missing Data on Spatial Surfaces. *Annals of the Association of American Geographers,* 74**,** 138-156.

BESSONG, P. O., ODIYO, J. O., MUSEKENE, J. N. & TESSEMA, A. 2009. Spatial Distribution of Diarrhoea and Microbial Quality of Domestic Water during an Outbreak of Diarrhoea in the Tshikuwi Community in Venda, South Africa. *Journal of Health Population and Nutrition,* 27**,** 652-659.

BEZERRA, J. G., PONTES, L., MINA, D. D. & BARRETO, M. L. 2007. Infant mortality and sociodemographic conditions in Ceara, Brazil, 1991 and 2000. *Revista De Saude Publica,* 41**,** 1022-1029.

BILLIG, P., BENDAHMANE, D. & SWINDALE, A. 1999. Water and sanitation indicators measurement guide. Washington, DC; USAID.

BLACK, R. E., MORRIS, S. S. & BRYCE, J. 2003. Where and why are 10 million children dying every year? *The Lancet,* 361**,** 2226-2234.

BLUM, D. & FEACHEM, R. 1983. Measuring the impact of water supply and sanitation investments on diarrhoeal diseases: problems of methodology. *International Journal of Epidemiology,* 12**,** 357.

BOUMAN, B., CASTA EDA, A. & BHUIYAN, S. 2002. Nitrate and pesticide contamination of groundwater under rice-based cropping systems: past and current evidence from the Philippines. *Agriculture, Ecosystems & Environment,* 92**,** 185-199.

BREMAN, J., ALILIO, M. & MILLS, A. 2004. Conquering the intolerable burden of malaria: what's new, what's needed: a summary. *The American Journal of Tropical Medicine and Hygiene,* 71**,** 1.

BRISCOE, J., DE CASTRO, P., GRIFFIN, C., NORTH, J. & OLSEN, O. 1990. Toward equitable and sustainable rural water supplies: a contingent valuation study in Brazil. *The World Bank Economic Review,* 4**,** 115.

BROOKES, J. D., AINGER, C. M., HOWE, C., NORTON, J. & SCHLADOW, G. 2010. Water and climate change: challenges for the 21st century. *Journal of Water and Climate Change,* 1**,** 1-1.

BURGERT, C. 2009. *RE: The different "source" options in the dataset (Personal comunication)*

BURROUGH, P. A., MCDONNELL, R. A., BURROUGH, P. A. & MCDONNELL, R. 1998. *Principles of geographical information systems*, Oxford University Press Oxford.

BURROWS, R. & GANE, N. 2006. Geodemographics, software and class. *Sociology,* 40**,** 793.

CAIRO GOVERNORATE. 2009. *Housing in Cairo Governorate.Cairo, Egypt, In arabic Language,* http://www.cairo.gov.egC15/C8/EHome/default.aspx, accessed on 25th April 2009.

CAPMAS. 2008. Egypt today. Cairo, Centeral Agency for Public Mobilization and Statistics. http://www.msrintranet.capmas.gov.eg/pls/fdl/tst12l?action=1&lname=, accessed on 10th of December 2008.

CAPMAS. 2009a. Population clock. http://www.msrintranet.capmas.gov.eg/pls/fdl/tst12e?action=&lname=, accessed on 23rd January 2009.

CAPMAS. 2009b. *2006 Egyptian Census* [Online]. Cairo: Centeral Agency for Public Mobilization and Statistics, accessed on 25th Februray 2009.

CAPMAS. 2009c. Sulm Areas in Egypt in 2005, Arabic report. Centeral Agency for Public Mobilisation and Statistics, data available for registered people,http://www.msrintranet.capmas.gov.eg/pls/fdl/tst12l?action=1&lname=. accessed on 10th Februray 2009

CAPMAS. 2009d. Egyptian Governorates in numbers, Suez governorate. Central Agency for Public Mobilisation and Statistics. accessed on 10th Februray 2009

CAPMAS. 2009e. Egyptian Governorate in Number, Matrouh and New Valley governorates Central Agency for Public Mobilisation and Statistics. accessed on 12th Februray 2009

CAPMAS. 2009f. Egyptian Governorates in number, Red Sea. North and South Sinai Governorates Central Agency for Public Mobilisation and Statistics. accessed on 12th Februray 2009

CAPMAS. 2010. Sanitation types, Personal contact, 15th May 2010.

CASH, D., ADGER, W., BERKES, F., GARDEN, P., LEBEL, L., OLSSON, P., PRITCHARD, L. & YOUNG, O. 2006. Scale and Cross-Scale Dynamics: Governance and Information in a Multilevel World. *Ecology and Society,* 11**,** 8.

CENTRAL STATISTICAL AGENCY (CAS). 2011. Basic welfare Indicator, available at http://www.csa.gov.et/pdf/Basic_welfare_Indicator/Households%20by%20sources%20of%20drinking%20water.pdf, accessed on 12th April 2011.

CHAMBERS, R. & TZAVIDIS, N. 2006. M-quantile models for small area estimation. *Biometrika,* 93**,** 255-268.

CHECKLEY, W., GILMAN, R. H., BLACK, R. E., EPSTEIN, L. D., CABRERA, L., STERLING, C. R. & MOULTON, L. H. 2004. Effect of water and sanitation on childhood health in a poor Peruvian peri-urban community. *The Lancet,* 363**,** 112-18.

CHEMULITI, J., GATHURA, P., KYULE, M. & NJERUH, F. 2002. Bacteriological qualities of indoor and out-door drinking water in Kibera sub-location of Nairobi, Kenya. *East African Medical Journal,* 79**,** 271-273.

CHEN, C., ENDER, P., MITCHELL, M., WELLS, C. . 2009. *logistic regression diagnostic lesson 3, Stata Web Books,* [Online]. Academic Technology Services,http://www.ats.ucla.edu/stat/stata/webbooks/logistic/, accessed on 16th of August 2009.

CHEN, C. C., THAKKAR, S., KNOBLOCK, C. & SHAHABI, C. 2003. Automatically annotating and integrating spatial datasets. *Lecture Notes in Computer Science,* 2750**,** 469-488.

CHEUNG, C. K., SHI, W. & ZHOU, X. 2004. A probability-based uncertainty model for point-in-polygon analysis in GIS. *GeoInformatica,* 8**,** 71-98.

CHILÈS, J. & DELFINER, P. 1999. *Geostatistics: modeling spatial uncertainty,* New York, USA, Wiley-Interscience.

CLASEN 2009. Scaling Up Household Water Treatment Among Low-Income Populations. *World Health Organization 2009*, available at http://whqlibdoc.who.int/hq/2009/WHO_HSE_WSH_09.02_eng.pdf, accessed on 22nd September 2009.

CLASEN, T. & BASTABLE, A. 2003. Faecal contamination of drinking water during collection and household storage: the need to extend protection to the point of use. *Journal of Water and Health,* 1**,** 109-116.

CLASEN, T., MCLAUGHLIN, C., NAYAAR, N., BOISSON, S., GUPTA, R., DESAI, D. & SHAH, N. 2008. Microbiological effectiveness and cost of disinfecting water by boiling in semi-urban India. *The American Journal of Tropical Medicine and Hygiene,* 79**,** 407.

CLASEN, T., ROBERT, I., RABIE, T., SCHMIDT, W. & CAIRNCROSS, S. 2006. Interventions to improve water quality for preventing diarrhoea (review). *Cochrane Database of Systematic Reviews,* 3**,** 104.

CLASEN, T., SCHMIDT, W., RABIE, T., ROBERTS, I. & CAIRNCROSS, S. 2007. Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis. *British Medical Journal,* 334**,** 782.

CLASEN, T. F. & BASTABLE, A. 2003. Faecal contamination of drinking water during collection and household storage: the need to extend protection to the point of use. *Journal of Water and Health,* 1**,** 109-116.

CLELAND, J. & VERMA, V. 1989. The World Fertility Survey: An Appraisal of Methodology. *Journal of the American Statistical Association***,** 756-767.

COGAN, J. 2007. *Introduction to Clustering Large and High-Dimensional Data,* London, Cameridge University Press.

COOK, R. D. & WEISBERG, S. 1982. *Residuals and influence in regression,* New York, Chapman and Hall New York.

CRAWFORD, I. 1997. *Marketing research and information systems,* Food & Agriculture

Org, Rome, Itly,

CROMLEY, E. K. & MCLAFFERTY, S. 2002. *GIS and public health*, The Guilford Press.

CRONIN, A. A., HOADLEY, A. W., GIBSON, J., BRESLIN, N., KOMOU, F. K., HALDIN, L. & PEDLEY, S. 2007. Urbanisation effects on groundwater chemical quality: findings focusing on the nitrate problem from 2 African cities reliant on on-site sanitation. *Journal of Water and Health,* 5**,** 441-454.

CROSETTO, M., TARANTOLA, S. & SALTELLI, A. 2000. Sensitivity and uncertainty analysis in spatial modelling based on GIS. *Agriculture, Ecosystems and Environment,* 81**,** 71-79.

CRUMP, J., OTIENO, P., SLUTSKER, L., KESWICK, B., ROSEN, D., HOEKSTRA, R., VULULE, J. & LUBY, S. 2005. Household based treatment of drinking water with flocculant-disinfectant for preventing diarrhoea in areas with turbid source water in rural western Kenya: cluster randomised controlled trial. *British Medical Journal,* 331**,** 478-481.

CUZICK, J. & EDWARDS, R. 1990. Spatial clustering for inhomogeneous populations. *Journal of the Royal Statistical Society. Series B (Methodological)***,** 73-104.

DAVIS, D. 2000. Agent-based decision-support framework for water supply infrastructure rehabilitation and development. *Computers, Environment and Urban Systems,* 24**,** 173-190.

DE ZUANE, J. 1997. *Handbook of drinking water quality,* London, Wiley.

DEBENHAM, J., CLARKE, G. & STILLWELL, J. 2001. Deriving supply-side variables to extend geodemographic classification. *Working paper-School of Geography University of Leeds.*

DEICHMANN, U. 1999. Geographic aspects of inequality and poverty. *World Bank web site on inequality, poverty, and socioeconomic performance, http://www.worldbank.org/poverty/inequal/index.htm,* accessed on 15th March 2009.

DEICHMANN, U., GOODCHILD, M. & ANSELIN, L. 1990. A general framework for the spatial interpolation of socioeconomic data. *Advanced Computing in the Social Sciences, Williamsburg, VA*.

DEMBERERE, T., SIWADI, J. & HOKO, Z. nd. An evaluation of the sustainability of a water supply and sanitation project in Mt Darwin District: Zimbabwe- Water for people. Available at http://www.nmun.org/ny11_downloads/ny11_committee_guides/igos/NY11_AFDB.pdf, accessed on 9th May 2009.

DEPARTMENT OF STATISTICS (DOS) [JORDAN] AND MACRO INTERNATIONAL INC. 2008. Jordan Population and Family Health Survey 2007, Calverton, Maryland, USA: Department of Statistics and Macro International Inc.

DEVI, A. & BOSTOEN, K. 2009. Extending the critical aspects of the water access indicator using East Africa as an example. *International Journal of Environmental Health Research,* 19**,** 329-341.

DIAB, A. 2002. Bacteriological Studies on the Potability, Efficacy and EIA of Desalination Operations at Sharm El-Sheikh Region, Egypt. *OnLine Journal of Biological Sciences,* 2**,** 411-414.

DONIA, N. 2007. SURVEY OF POTABLE WATER QUALITY PROBLEMS IN EGYPT. *In:* Proceeding of the Eleventh International Water Technology Conference, 2007 Sharm El-Sheikh, Egypt 1049-1058.

DEPARTMENT OF STATISTICS (DOS) THE HASHEMITE KINGDOM OFJORDAN. 2011. *Population and Family Health 2009, Available at* http://www.dos.gov.jo/sdb_pop/sdb_pop_e/index.htm, accessed on 14th April 2011.

DEPARTMENT OF STATISTICS (DOS) THE HASHEMITE KINGDOM OFJORDAN 2004. 2004 census, Census subjects, General characteristics of persons, available at http://www.dos.gov.jo/sdb_pop/sdb_pop_e/index.htm, accessed on 14th April 2011

DU PREEZ, M., CONROY, R., WRIGHT, J., MOYO, S., POTGIETER, N. & GUNDRY, S. 2008. Use of ceramic water filtration in the prevention of diarrheal disease: a randomized controlled trial in rural South Africa and Zimbabwe. *The American Journal of Tropical Medicine and Hygiene,* 79**,** 696.

DUNGUMARO, E. 2007. Socioeconomic differentials and availability of domestic water in South Africa. *Physics and Chemistry of the Earth,* 32**,** 1141-1147.

DUTTON, G. 1989. Modeling locational uncertainty via hierarchical tessellation. *In:* GOODCHILD, M., AND GOPAL, S. (ed.) *Accuracy of Spatial Databases.* London: Taylor and Francis, 125-140.

EIBNER, C. & STURM, R. 2006. US-based indices of area-level deprivation: Results from HealthCare for Communities. *Social Science & Medicine,* 62**,** 348-359.

EL ARABI, N. 1999. Problems of groundwater quality related to the urban environment in Greater Cairo. *IAHS PUBLICATION***,** 29-38.

EL-ABAGY, M., DUTKA, B., KAMEL, M. & EL ZANFALY, H. 1988. Incidence of coliphage in potable water supplies. *Applied and Environmental Microbiology,* 54**,** 1632.

EL-FADEL, M., EL-SAYEGH, Y., EL-FADL, K. & KHORBOTLY, D. 2003. The Nile River Basin: a case study in surface water conflict resolution. *Journal of Natural Resources and Life Sciences Education,* 32**,** 107-117.

EL-HEFNAWI, A. 2005. Protecting'agricultural land from urbanization or 'Managing'the conflict between informal urban growth while meeting the demands of the communities (Lessons learnt from the Egyptian policy reforms). *Paper presented at the Third Urban Research Symposium on "Land Development Urban Policy and Poverty Reduction"*. 4-6 April 2005, Brasilia, DF, Brazil

EL-RAEY, M., FOUDA, Y. & GAL, P. 2000. GIS for environmental assessment of the impacts of urban encroachment on Rosetta Region, Egypt. *Environmental Monitoring and Assessment,* 60**,** 217-233.

EL-TAWEEL, G. & SHABAN, A. 2001. Microbiological quality of drinking water at eight water treatment plants. *International Journal of Environmental Health Research,* 11**,** 285-290.

EL-ZANATY, F. & WAY, A. 2006. *Egypt demographic and health survey, 2005,* Cairo,Ministry of Health and Population, National Population Council, El- Zanaty and Associates, and ORC Macro, Ministry of Health and Population.

EL-ZANATY, F. W., A. 2006. Egypt demographic and health survey, 2005,Cairo, Egypt, Ministry of Health and Population.

ESIS 2007. The land and the people Egyptian Yearbook, 2007. Cairo, Egypt: Egypt State Information System.

ESIS 2008. Egypt State Information Services, water Resources. http://www.sis.gov.eg/En/Pub/waterinegypt/110900000000000001.htm, accessed on 10th of December 2008.

ESREY, S. A., POTASH, J. B., ROBERTS, L. & SHIFF, C. 1991. Effects of improved water supply and sanitation on ascariasis, diarrhoea, dracunculiasis, hookworm infection, schistosomiasis, and trachoma. *Bulletin of the World Health Organization,* 69**,** 609.

FAROOQI, A., MASUDA, H. & FIRDOUS, N. 2007. Toxic fluoride and arsenic contaminated groundwater in the Lahore and Kasur districts, Punjab, Pakistan and possible contaminant sources. *Environmental Pollution,* 145**,** 839-849.

FARR, M. & WEBBER, R. 2001. MOSAIC: From an area classification system to individual classification. *Journal of Targeting, Measurement and Analysis for Marketing,* 10**,** 55-65.

FAY, M., LEIPZIGER, D., WODON, Q. & YEPES, T. 2003. Achieving the Millennium Development Goals: The Role of Infrastructure. *Policy Research Working Paper Series,* 3163, The World Bank.

FAY, M., LEIPZIGER, D., WODON, Q. & YEPES, T. 2005. Achieving child-health-related millennium development goals: the role of infrastructure. *World Development,* 33**,** 1267-1284.

FERGUSON, C. M., CHARLES, K. & DEERE, D. A. 2009. Quantification of Microbial Sources in Drinking-Water Catchments. *Critical Reviews in Environmental Science and Technology,* 39**,** 1-40.

FEWTRELL, L., FUGE, R. & KAY, D. 2005. An estimation of the global burden of disease due to skin lesions caused by arsenic in drinking water. *Journal of Water and Health,* 3**,** 101-107.

FEWTRELL, L., KAUFMANN, R. B., KAY, D., ENANORIA, W., HALLER, L. & COLFORD, J. M. 2005. Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. *The Lancet Infectious Diseases,* 5**,** 42-52.

FIELDHOUSE, E. & TYE, R. 1996. Deprived people or deprived places? Exploring the ecological fallacy in studies of deprivation with the Samples of Anonymised Records. *Environment and Planning A,* 28**,** 237-259.

FISHER 2000. Fuzzy modelling. *In:* OPENSHAW, S. & ABRAHART, R. J. (eds.) *Geocomputing, P161-186* London: Taylor and Francis, London.

FISHER, P. F. 1999. Models of uncertainty in spatial data. *Geographical Information Systems,* 1**,** 191-205.

FLOWERDEW, R., GREEN, M. & KEHRIS, E. 1991. Using areal interpolation methods in geographic information systems. *Papers in regional science,* 70**,** 303-315.

FOODY, G. M. & ATKINSON, P. M. 2003. *Uncertainty in remote sensing and GIS,* Chichester, England, John Wiley and Sons.

FOSSO-KANKEU, E., JAGALS, P. & DU PREEZ, H. 2008. Exposure of rural households to toxic cyanobacteria in container-stored water. *Water SA (Online),* 34**,** 631-636.

FRAYMAN, H., MACK, J., LANSLEY, S., GORDON, D. & HILLS, J. 1991. Breadline Britain In The 1990s: The Findings Of The Television Series. *London, Domino Films and London Weekend Television*.

GEMPERLI, A. 2003. *Development of spatial statistical methods for modelling point-referenced spatial data in malaria epidemiology.* PhD thesis, University of Basel.

GEMPERLI, A., VOUNATSOU, P., KLEINSCHMIDT, I., BAGAYOKO, M., LENGELER, C. & SMITH, T. 2004. Spatial patterns of infant mortality in Mali: the effect of malaria endemicity. *American Journal of Epidemiology,* 159**,** 64-72.

GHOSH, M. & RAO, J. 1994. Small area estimation: an appraisal. *Statistical Science,* 9**,** 55-76.

GIBSON, C., OSTROM, E. & AHN, T. 2000. The concept of scale and the human dimensions of global change: a survey. *Ecological Economics,* 32**,** 217-239.

GINE, R. & PEREZ-FOGUET, A. 2008. Sustainability assessment of national rural water supply program in Tanzania, *Natural resources forum*, November 2008, 32, 327-342.

GLEICK, P. H. 1993. Water and conflict: Fresh water resources and international security. *International Security,* 18**,** 79-112.

GOMBOS I, M. & ŽALIK, B. 2005. Point-in-polygon tests for geometric buffers. *Computers and Geosciences,* 31**,** 1201-1212.

GONG, P. & MU, L. 2000. Error detection through consistency checking. Annals of GIS 6(2): 188-193.

GOODCHILD, M., & LAM, N. 1980. *Areal interpolation: a variant of the traditional spatial problem*, Dept. of Geography, University of Western Ontario.

GOODCHILD, M. F. & GOPAL, S. 1989. *The accuracy of spatial databases,* London, CRC.

GOSS, J. 1995. We know who you are and we know where you live- the instrumental rationality of geodemographic systems. *Economic Geography,* 71**,** 171-188.

GOUGH, J., EISENSCHITZ, A. & MCCULLOCH, A. 2009. Spaces of Social Exclusion. *Economic Geography,* 84**,** 237-238.

GOUJON, A., ALKITKAT, H., LUTZ, W., PROMMER, I. & HORDIJK, L. 2007. Population and Human Capital Growth in Egypt: Projections for Governorates to 2051. Laxenburg, Austria: International Institute for Applied Systems Analysis.

GREEN, M. & FLOWERDEW, R. 1996. New evidence on the modifiable areal unit problem. *Spatial analysis: Modelling in a GIS environment***,** 41-54.

GREGORY, I. N. 2002. The accuracy of areal interpolation techniques: standardising 19th and 20th century census data to allow long-term comparisons. *Computers, Environment and Urban Systems,* 26**,** 293-314.

GRIFFITH, D. 1989. Statistical analysis of spatial data in the presence of missing observations: a methodological guide and an application to urban census data. *Environment and Planning A,* 21**,** 1511-1523.

GRIFFITH, D. A., MILLONES, M., VINCENT, M., JOHNSON, D. L. & HUNT, A. 2007. Impacts of Positional Error on Spatial Regression Analysis: A Case Study of Address Locations in Syracuse, New York. *Transactions in GIS,* 11**,** 655-679.

GRONWALL, J. T. 2008. Access to water: Rights, obligations and the Bangalore situation, Department of Water and Environmental Studies, Linköping University,SE-581 83 Linköping, Sweden, available fromLinköping University Electronic Press at http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-11686, accessed on 21st September 2009

GUCHI, E. 2007. Evaluating the efficiency of slow sand filtration in clay pots in removing coliform bacteria and turbidity from drinking water at household level and assessment of the contamination level at the point-of-use at home in the central highlands of Ethiopia (Yubdo Legebatu PA). Addis Ababa University, Ethiopia.(MSc thesis Applied Microbiology).

GUNDRY, S., WRIGHT, J., CONROY, R., DU PREEZ, M., GENTHE, B., MOYO, S., MUTISI, C., NDAMBA, J. & POTGIETER, N. 2006. Contamination of drinking water between source and point-of-use in rural households of South Africa and Zimbabwe: implications for monitoring the Millennium Development Goal for water. *Water Practice and Technology,* 1.

GUNDRY, S. W., WRIGHT, J. A., CONROY, R. M., PREEZ, M., GENTHE, B., MOYO, S., MUTISI, C. & POTGIETER, N. 2009. Child dysentery in the Limpopo Valley: a cohort study of water, sanitation and hygiene risk factors. *Journal of Water and Health,* 7**,** 259-266.

GUPTA, S., SUANTIO, A., GRAY, A., WIDYASTUTI, E., JAIN, N., ROLOS, R., HOEKSTRA, R. & QUICK, R. 2007. Factors associated with E. coli contamination of household drinking water among tsunami and earthquake survivors, Indonesia. *The American Journal of Tropical Medicine and Hygiene,* 76**,** 1158.

HADADIN, N., QAQISH, M., AKAWWI, E. & BDOUR, A. 2010. Water shortage in Jordan -- Sustainable solutions. *Desalination,* 250**,** 197-202.

HADDAD, L., BHATTARAI, S., IMMINK, M. & KUMAR, S. 1998. Estimating the interactions between household food security and preschool diarrhea. *Food Policy,* 23**,** 241-261.

HAINING, R. 1998. *Spatial statistics and the analysis of health data,* London, Taylor and Francis.

HALLER, L., HUTTON, G. & BARTRAM, J. 2007. Estimating the costs and health benefits of water and sanitation improvements at global level. *Journal of Water and Health,* 5**,** 467-480.

HALVORSON, S. J. 2004. Women's management of the household health environment: responding to childhood diarrhea in the Northern Areas, Pakistan. *Health & Place,* 10**,** 43-58.

HAMZA, W. & MASON, S. 2004. Water availability and food security challenges in Egypt. *In:* International Forum on Food Security Under Water Scarcity in the Middle East: Problems and Solutions, 2004 Como, Italy, 24-27 Nov.

HARRIS, R. & LONGLEY, P. 2002. Creating small area measures of urban deprivation. *Environment and Planning A,* 34**,** 1073-1094.

HARRIS, R., SLEIGHT, P. & WEBBER, R. 2005. *Geodemographics, GIS and Neighborhoods Targeting,* Chichester, England, Wiley & Sons Inc.

HARVEY, P. & REED, B. 2004. *Rural water supply in Africa: Building blocks for handpump sustainability*, WEDC, Loughborough University.

HAVARD, S., DEGUEN, S., BODIN, J., LOUIS, K., LAURENT, O. & BARD, D. 2008. A small-area index of socioeconomic deprivation to capture health inequalities in France. *Social Science and Medicine,* 67**,** 2007-2016.

HEFNY, K., FARID, M. S. & HUSSEIN, M. 1992. Groundwater assessment in Egypt. *International Journal of Water Resources Development,* 8**,** 126-134.

HEIDECKE, C. 2006. *Development and evaluation of a regional water poverty index for Benin,* Washington DC, USA, Intl Food Policy Res Inst.

HERMAN, E. & BENTLEY, M. 1993. *Rapid assessment procedures, RAP: to improve the household management of diarrhea*, International Nutrition Foundation for Developing Countries. United Nations University, Tokyo, Japan.

HEUVELINK, G. B. M. 1998. *Error Propagation in Environmental Modelling with GIS,* London, Taylor & Francis.

HEUVELINK, G. B. M. & BURROUGH, P. A. 2002. Developments in statistical approaches to spatial uncertainty and its propagation. *International Journal of GIS,* 16**,** 111-113.

HOSMER, D. W. & LEMESHOW, S. 2004. *Applied Logistic Regression,* USA, Wiley-Interscience.

HOSMER, D. W., TABER, S. & LEMESHOW, S. 1991. The importance of assessing the fit of logistic regression models: a case study. *American Journal of Public Health,* 81**,** 1630-1635.

HOWARD, G. & BARTRAM, J. 2003. *Domestic Water Quantity, Service Level, and Health,* Geneva, Switzerland, World Health Organization.

HOWARD, G. & BARTRAM, J. 2005. Effective water supply surveillance in urban areas of developing countries. *Journal of Water and Health,* 3**,** 31-43.

HUNTER, G. J. 1999. New Tools For Handling Spatial Data Quality: Moving from Academic Concepts to Practical Reality. *URISA Journal,* 11, 25 - 34

HUNTER, P. R. 2009. Household Water Treatment in Developing Countries: Comparing Different Intervention Types Using Meta-Regression. *Environmental Science and Technology,* 43**,** 8991-8997.

HUNTER, P. R., CHALMERS, R. M., HUGHES, S. & SYED, Q. 2005. Self-reported diarrhoea in a control group: a strong association with reporting of low-pressure events in tap water. *Clinical Infectious Diseases,* 40**,** 32-34.

HUNTER, P. R., ZMIROU-NAVIER, D. & HARTEMANN, P. 2009. Estimating the impact on health of poor reliability of drinking water interventions in developing countries. *Science of the Total Environment,* 407**,** 2621-2624.

IDRIS, H. & NOUR, S. 1990. Present groundwater status in Egypt and the environmental impacts. *Environmental Geology,* 16**,** 171-177.

IGLESIAS, A., GARROTE, L., FLORES, F. & MONEO, M. 2007. Challenges to manage the risk of water scarcity and climate change in the Mediterranean. *Water Resources Management,* 21**,** 775-788.

INGWE, R., OTU, J., AGI, S., EJA, E. & UKWAYI, J. 2008. Federal government assistance for web-based secondary education under poor socio-economic development conditions in Nigeria: Geodemographic and qualitative analyses. *Journal of Geography and Regional Planning,* 1**,** 72-84.

INSTITUE OF STATISTICS.2010. Albania Demographic and Health Survey 2008-09. Institute of Statistics / Institute of Public Health, Tirana, Albania & ICF Macro, Calverton, Maryland.

JABER, J. O., PROBERT, S. D. & BADR, O. 1997. Energy and environmental issues for Jordan. *Applied Energy,* 57**,** 45-101.

JAGALS, P., JAGALS, C. & BOKAKO, T. 2003. Effect of container-biofilm on the microbiological quality of water used from plastic household containers. *Journal of Water and Health,* 1**,** 101-8.

JENSEN, P., ENSINK, J., JAYASINGHE, G., VAN DER HOEK, W., CAIRNCROSS, S. & DALSGAARD, A. 2003. Effect of chlorination of drinking-water on water quality and childhood diarrhoea in a village in Pakistan. *Journal of Health Population and Nutrition,* 21**,** 26-31.

JIMENEZ, A., AND FOGUET, A. 2008. Quality and sustainability aspects in water access indicators: an example from Same District, Tanzania. *Access to*

Sanitation and Save Water:Global Partnerships and Local Actions. Paper presented at 33rd WEDC International Conference, Accra, Ghana,2008.

JIMENEZ, A., AND FOGUET, A. 2009. Access to safe and year round functional water:an estimation of coverage for three central region in Tanzania. *Water,Sanitation and Hygiene: Sustainable Development and Multisectoral Approaches.* Paper presented at 34th WEDC International Confernce, Addis Ababa, Ethiopia.

JOFRE, J., OLLE, E., RIBAS, F., VIDAL, A. & LUCENA, F. 1995. Potential usefulness of bacteriophages that infect Bacteroides fragilis as model organisms for monitoring virus removal in drinking water treatment plants. *Applied and Environmental Microbiology,* 61**,** 3227.

JORDAN, H., RODERICK, P. & MARTIN, D. 2004. The Index of Multiple Deprivation 2000 and accessibility effects on health. *Journal of Epidemiology and Community Health,* 58**,** 250.

JOUSILAHTI, P., MADKOUR, S., LAMBRECHTS, T. & SHERWIN, E. 1997. Diarrhoeal disease morbidity and home treatment practices in Egypt. *Public Health,* 111**,** 5-10.

KAGGWA, E. B., DIOP, N. & STOREY, J. D. 2008. The Role of Individual and Community Normative Factors: A Multilevel Analysis of Contraceptive Use Among Women in Mali. *International Family Planning Perspectives,* 34**,** 79-88.

KHALIFA, M., DAVANZO, J. & ADAMSON, D. 2000. Population growth in Egypt: A continuing policy challenge. Centre for Middle East public policy, *population matters project*. Issue paper, IP-183.

KALLIDAIKURICHI, S. & RAO, B. 2009. Index of Drinking Water Adequacy for the Asian Economies. *working paper series, Serial No IWP/WP, No.2/2009,* http://www.spp.nus.edu.sg/iwp/WorkingPaper_Series/IDWA_I.pdf,accessed 10th March 2010.

KAMERI-MBOTE, P. 2007. Water, Conflict, and Cooperation: lessons from the nile river Basin, available at http://www.wilsoncenter.org/topics/pubs/navigatingpeaceissuepkm.pdf. accessed on 13th April 2011.

KANDALA, N. B., JI, C., STALLARD, N., STRANGES, S. & CAPPUCCIO, F. P. 2007. Spatial analysis of risk factors for childhood morbidity in Nigeria. *The American Journal of Tropical Medicine and Hygiene,* 77**,** 770.

KANDALA, N. B., MAGADI, M. A. & MADISE, N. J. 2006. An investigation of district spatial variations of childhood diarrhoea and fever morbidity in Malawi. *Social Science and Medicine,* 62**,** 1138-1152.

KAZEMBE, L. N., APPLETON, C. C. & KLEINSCHMIDT, I. 2007. Spatial analysis of the relationship between early childhood mortality and malaria endemicity in Malawi. *Geospatial Health,* 2**,** 41-50.

KELSALL, J. & WAKEFIELD, J. 2002. Modeling spatial variation in disease risk. *Journal of the American Statistical Association,* 97**,** 692-701.

KENNEDY, M. & KOPP, S. 2001. *Understanding Map Projections,* New York, USA, Environmental Systems Research Institute.

KENNEDY, S. 1989. The small number problem and the accuracy of spatial databases. *In:* GOODCHILD, M., AND GOPAL S. (ed.) *The accuracy of spatial databases.* London: Taylor and Francis, p 187–196.

KEYZER, M. & VAN WESENBEECK, L. 2006. The Millennium Development Goals, how realistic are they? *De Economist,* 154**,** 443-466.

KHAIRY, A., EL SEBAIE, O., GAWAD, A. & EL ATTAR, L. 1982. The sanitary condition of rural drinking water in a Nile Delta village. I. Parasitological assessment of 'zir' stored and direct tap water. *The Journal of Hygiene,* 88**,** 57.

KHATRI, K. & VAIRAVAMOORTHY, K. 2009. Water Demand Forecasting for the City of the Future against the Uncertainties and the Global Change Pressures: Case of Birmingham. *Paper presented at the World Environmental & Water Resources Congress EWRI.* Kansas, USA

KLEINBAUM, D. & KLEIN, M. 2002. *Logistic regression: a self-learning text (Statistics for Biology and Health)*, Springer, New York, USA.

KOGAN, J. 2007. *Introduction to clustering large and high-dimensional data*, Cambridge Univ Pr. New York, USA

KOMNENIC, V., AHLERS, R. & ZAAG, P. 2009. Assessing the usefulness of the water poverty index by applying it to a special case: Can one be water poor with high levels of access? *Physics and Chemistry of the Earth,* 34**,** 219-224.

KULLDORFF, M., SONG, C., GREGORIO, D., SAMOCIUK, H. & DECHELLO, L. 2006. Cancer map patterns are they random or not? *American Journal of Preventive Medicine,* 30**,** 37-49.

KUMAR, N. 2007. Spatial sampling design for a demographic and health survey. *Population Research and Policy Review,* 26**,** 581-599.

LAM, N. 1980. *Methods and problems of areal interpolation, PhD thesis.* The University of Western Ontario.

LAM, N. 1983. Spatial interpolation methods: a review. *Cartography and Geographic Information Science,* 10**,** 129-150.

LAM, N. S. 2009. Spatial Interpolation. *In:* Kitchen, R and Thrift, N,. (eds.) *International Encyclopedia of Human Geography.* Oxford: Elsevier.

LASHEEN, M., EL-KHOLY, G., SHARABY, C., ELSHERIF, I. & EL-WAKEEL, S. 2008. Assessment of selected heavy metals in some water treatment plants and household tap water in Greater Cairo, Egypt. *Management of Environmental Quality: An International Journal,* 19**,** 367-376.

LASWON, A., DENSION,D. 2002. Spatial Cluster Modelling:An overview. *In:* LASWON, A., DENSION,D. (ed.) *Spatial Cluster Modelling.* London: A CRC press.

LAURENT, P. 2005. Household drinking water systems and their impact on people with weakened immunity. *Geneva: World Health Organization.* http://www. who. int/household_water/research/HWTS_impacts_on_weakened_immunity. Pdf, accessed on 25th June 2009.

LAWRENCE, P., MEIGH, J. & SULLIVAN, C. 2002. The water poverty index: an international comparison. *Keele Economics Research Papers,Department of Economics, Keele, Staffordshire,* www.keele.ac.uk/depts/ec/web/wpapers/kerp0219.pdf ,accessed on 9th March 2010.

LEE, E. & SCHWAB, K. 2005. Deficiencies in drinking water distribution systems in developing countries. *Journal of Water and Health,* 3**,** 109-127.

LEHTOLA, M., MIETTINEN, I., VARTIAINEN, T. & MARTIKAINEN, P. 2002. Changes in content of microbially available phosphorus, assimilable organic carbon and microbial growth potential during drinking water treatment processes. *Water Research,* 36**,** 3681-3690.

LEUNG, Y., MA, J. H. & GOODCHILD, M. F. 2004. A general framework for error analysis in measurement-based GIS Part 1: The basic measurement-error model and related concepts. *Journal of Geographical Systems,* 6**,** 325-354.

LEUNG, Y. & YAN, J. 1997. Point-in-polygon analysis under certainty and uncertainty. *GeoInformatica,* 1**,** 93-114.

LEVY, K., NELSON, K. L., HUBBARD, A. & EISENBERG, J. N. S. 2008. Following the Water: A Controlled Study of Drinking Water Storage in Northern Coastal Ecuador. *Environmental Health Perspectives,* 116**,** 1533-1540.

LI, J., WANG, W. & WU, E. 2007. Technical Section: Point-in-polygon tests by convex decomposition. *Computers and Graphics,* 31**,** 636-648.

LONGLEY, P. A.,GOODCHILD, M. F.,MAGUIRE, D. J., & RHIND, D. W. (1999). Data transformation and linkage. In P. A. Longley, M. F. Goodchild, D. J. Maguire, & D. W. Rhind (Eds.). Geographical Information Systems: Principles and Technical Issues(vol. 1, pp. 477–479). Chichester: Wiley.

LONGLEY, P., ASHBY, D., WEBBER, R. & LI, C. 2006. Geodemographic classifications, the digital divide and understanding customer take-up of new technologies. *Technology Journal,* 24**,** 67-74.

LONGLEY, P. & BATTY, M. 2003. *Advanced spatial analysis: the CASA book of GIS,* New York, USA, Esri Press.

LONGLEY, P., GOODCHILD, M. F., MAGUIRE, D. J. & RHIND, D. W. 2005. *Geographical Information Systems and Science,* Chichester, England, Wiley & Sons Inc.

MACEACHREN, A. M. 1985. Accuracy of thematic maps/Implications of choropleth symbolization. *Cartographica: The International Journal for Geographic Information and Geovisualization,* 22**,** 38-58.

MACGILLIVRAY, B., HAMILTON, P., STRUTT, J. & POLLARD, S. 2006. Risk analysis strategies in the water utility sector: An inventory of applications for better and more credible decision making. *Critical Reviews in Environmental Science and Technology,* 36**,** 85-139.

MANGYO, E. 2008. The effect of water accessibility on child health in China. *Journal of Health Economics,* 27**,** 1343-1356.

MARAJ, S., RODDA, N., JACKSON, S., BUCKLEY, C. & MACLEOD, N. 2009. Microbial deterioration of stored water for users supplied by stand-pipes and ground-tanks in a peri-urban community. *Water SA,* 32.

MARTIN, D. 1996. *Geographic Information Systems: Socioeconomic Applications,* London, Routledge

MARTIN, D., SENIOR, M. & WILLIAMS, H. 1994. On measures of deprivation and the spatial allocation of resources for primary health care. *Environment and Planning A,* 26**,** 1911-1911.

MATERN, R., MENDELSON, M., OLIPHANT, M. & BANK, D. 2009. Developing a Deprivation Index: The Research Process, Daily Bread Food Bank and the Caledon Institute of Social Policy, Available at

http://www.dailybread.ca/learningcentre/documents/DevelopingaDeprivationIndexFINAL.pdf, accessed on 14th February 2010

MATHERS, C. D., MA FAT, D., INOUE, M., RAO, C. & LOPEZ, A. D. 2005. Counting the dead and what they died from: an assessment of the global status of cause of death data. *Bulletin of the World Health Organization,* 83**,** 171-177c.

MATHEW, B. 2004. *Ensuring sustained beneficial outcomes for water and sanitation (WATSAN) programmes in the developing world.* PhD thesis, Granfield University

MATHEWS, A., and BINA, G. 2004. Contextual influences on the use of antenatal care in Nepal. DHS Geographic Studies 2. Calverton, Maryland USA: ORC Macro.

MÄUSEZAHL, D., CHRISTEN, A., PACHECO, G., TELLEZ, F., IRIARTE, M., ZAPATA, M., CEVALLOS, M., HATTENDORF, J., CATTANEO, M. & ARNOLD, B. 2009. Solar drinking water disinfection (SODIS) to reduce childhood diarrhoea in rural Bolivia: a cluster-randomized, controlled trial. *Plos Medicine,* 6**,** e1000125.

MEASURE DHS.2004.Contextual influences on the use of antenatal care in Nepal. *DHS Geographic Studies, Calverton, Maryland USA: ORC Macro*.

MEASURE DHS 2008a. About DHS. http://www.measurEgyptian DHS.com/aboutdhs/whoweare.cfm, accessed on 25th of December 2008.

MEASURE DHS 2008b. HIV Prevalence Estimates. http://www.measurEgyptian DHS.com/pubs/pub_details.cfm?ID=792, accessed on 10th January 2008.

MEASURE DHS 2008c. DHS topics, Gender. http://www.measurEgyptian DHS.com/topics/gender/start.cfm, accessed on 10th of December 2008.

MEASURE DHS 2008d. DHS topic, Biomarker. http://www.measurEgyptian

DHS.com/topics/Biomaker/start.cfm, accessed on 28th of November 2008.

MEASURE DHS 2008e. DHS topic, Youth. http://www.measurEgyptian DHS.com/topics/Youth/start.cfm, accessed on 28th of December 2008.

MEASURE DHS 2008f. DHS methodology. http://www.measurEgyptian DHS.com/aboutsurveys/dhs/methodology.cfm, accessed on 28th December 2008.

MEASURE DHS 2008g. Surveys,Geographic Information System. http://www.measurEgyptian DHS.com/aboutsurveys/gis/start.cfm, accessed on 10th of January 2009.

MEASURE DHS 2009a. DHS sampling manual,Macro International Inc,Calverton, Maryland, available at http://www.measurEgyptian DHS.com/pubs/pdf/DHSM4/DHS_III_Sampling_Manual.pdf, accessed on 29th July 2009.

MEASURE DHS 2009b. Egyptian and Demographic Health Survey, Final report, Appendix C, http://www.measurEgyptian DHS.com/pubs/pub_details.cfm?ID=586, *accessed on* 30th July 2009.

MEASUREDHS 2011. Publication Search Results, http://www.measuredhs.com/pubs/Search/search_results.cfm?newSrch=1&srchTp=advanced&topic=20&tnm=Geographic%20Information, accessed on 10th June, 2011.

MILMAN, A. & SHORT, A. 2008. Incorporating resilience into sustainability indicators: An example for the urban water sector. *Global Environmental Change,* 18**,** 758-767.

MINISTRY OF HOUSING UTILITIES AND URBAN DEVELOPMENT. 2009. The new cities, New Urban Communities Authority, http://www.housing-utility.gov.eg/nuca.asp, accessed on 10th April 2009.

MINTZ, E., REIFF, F. & TAUXE, R. 1995. Safe water treatment and storage in the home: a practical new strategy to prevent waterborne disease. Journal of American Medical Association *JAMA,* 273**,** 948.

MISHRA, V., VAESSEN, M., BOERMA, J. T., ARNOLD, F., WAY, A., BARRERE, B., CROSS, A., HONG, R. & SANGHA, J. 2006. HIV testing in national population-based surveys: experience from the Demographic and Health Surveys. *Bulletin of the World Health Organization,* 84**,** 537-545.

MITAS, L. & MITASOVA, H. 1999. Spatial interpolation In: *Geographical Information Systems: Principles, Techniques, Management and Applications, Wiley,* 481, New York, USA.

MITCHELL, R. 1997. *An integration of aggregate and disaggregate census data* Unpublished PhD thesis, University of Southampton.

MITCHELL, V. & MCGOLDRICK, P. 1994. The role of geodemographics in segmenting and targeting consumer markets: a Delphi study. *European Journal of Marketing,* 28**,** 54-72.

MLD 2009. Administrative Boundaries of Egyptian Governorates, Ministry of Local Development, http://www.mold.gov.eg/Arabic/Left/DataBaseSearch/Egypt Accessed on 12th of June 2009.

MLOTE, S., SULLIVAN2A, C. & MEIGH2B, J. 2002. Water Poverty Index: a Tool for Integrated Water Management. Paper presented at *the 3rd WaterNet/Warfsa Symposium 'Water Demand Management for Sustainable Development.* Dar es Salaam.

MOE, C. & RHEINGANS, R. 2006. Global challenges in water, sanitation and health. *Journal of Water and Health,* 4**,** 41.

MOHADJER, L. 1999. Sample weights for households with multiphase data collection approaches. *Proceedings actes of the 52 nd ISI Session, available at* http://www.stat.fi/isi99/proceedings/arkisto/varasto/moha0259.pdf, accessed on 19th January 2011.

MOHSEN, M. S. 2007. Water strategies and potential of desalination in Jordan. *Desalination,* 203**,** 27-46.

MOLLE, F. & MOLLINGA, P. 2003. Water poverty indicators: conceptual problems and policy issues. *Water Policy,* 5**,** 529-544.

MONTANA, L. & SPENCER, J. 2004. Incorporating Geographic Information Into MEASURE Surveys: A Field Guide to GPS Data,Spencer. Macro International Publication, http://www.measurEgyptian DHS.com/basicdoc/gps/DHS_GPS_Manual.pdf accessed on 24th of July 2009

MONTANA, L. S., MISHRA, V. & HONG, R. 2008. Comparison of HIV prevalence estimates from antenatal care surveillance and population-based surveys in sub-Saharan Africa. *British Medical Journal,* 84**,** 178.

MOY, R., BOOTH, I., CHOTO, R. G. A. B. & MCNEISH, A. 1991. Risk factors for high diarrhoea frequency: a study in rural Zimbabwe. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 85**,** 814-818.

MROZINSKI, R. & CROMLEY, R. 1999. Singly and doubly constrained methods of areal interpolation for vector based GIS. *Transactions in GIS,* 3**,** 285-301.

MUNCH, Z., VAN LILL, S. W. P., BOOYSEN, C. N., ZIETSMAN, H. L., ENARSON, D. A. & BEYERS, N. 2003. Tuberculosis transmission patterns in a high-incidence area: a spatial analysis. *International Journal of Tuberculosis and Lung Disease,* 7**,** 271-277.

MWRI 2005. Integrated water resources plan, Egypt, Water for the future. Ministry of Water Resources and Irrigation, http://www.wldelft.nl/cons/area/rbm/wrp/index.html, accessed on 25th August 2009

NATIONAL AUDIT OFFICE 1999. A practical guide to sampling, available at http://www.nao.org.uk/publications/Samplingguide.pdf,accessed on 22nd January 2011.

NEDOROSCIK, J. A. 1997. *The city of the dead: a history of Cairo's cemetery communities,* USA, JF Bergin & Garvey.

NICOLAISEN, B. 2003. Developments in membrane technology for water treatment. *Desalination,* 153**,** 355-360.

NIGGEBRUGGE, A., HAYNES, R., JONES, A., LOVETT, A. & HARVEY, I. 2005. The index of multiple deprivation 2000 access domain: a useful indicator for public health? *Social Science and Medicine,* 60**,** 2743-2753.

NIST/SEMATECH 2011. E-handbook of statistical methods, http://www.itl.nist.gov/div898/handbook/index.htm,accessed on 11th June 2011.

NOBLE, M., BABITA, M., BARNES, H., DIBBEN, C., MAGASELA, W., NOBLE, S., NTSHONGWANA, P., PHILLIPS, H., RAMA, S. & ROBERTS, B. 2006a. The Provincial Indices of Multiple Deprivation for South Africa 2001: technical report. Oxford: Centre for the Analysis of South African Social Policy, University of Oxford.

NOBLE, M., BARNES, H., WRIGHT, G. & ROBERTS, B. 2010. Small Area Indices of Multiple Deprivation in South Africa. *Social Indicators Research,* 95**,** 281-297.

NOBLE, M., MCLENNAN, D., WILKINSON, K., WHITWORTH, A., BARNES, H. & DIBBEN, C. 2008. The English Indices of Deprivation 2007. *London: Communities and Local Government*.

NOBLE, M., PENHALE, B., SMITH, G., WRIGHT, G., DIBBEN, C., OWEN, T., LLOYD, M., FIRTH, D. & PAYNE, C. 2000. Indices of Deprivation 2000. *Regeneration research summary,* 31.

NOBLE, M., WRIGHT, G., SMITH, G. & DIBBEN, C. 2006b. Measuring multiple deprivation at the small-area level. *Environment and Planning A,* 38**,** 169-185.

NOBLE, M., SMITH, G., WRIGHT, G., DIBBEN, C., LLOYD, M. & PENHALE, B. 2000. *Welsh Index of Multiple Deprivation*. National Assembly for Wales, Cardiff Bay, Cardiff CF99 1NA.

NOBLE, M., WRIGHT, G., DIBBEN, C., SMITH, G., MCLENNAN, D., ANTTILA, C., BARNES, H., MOKHTAR, C., NOBLE, S. & GARDNER, J. 2004. *The English Indices of Deprivation 2004*. *London: Neighbourhood Renewal Unit, Office of the Deputy Prime Minister*.

OBI, C., MOMBA, M., SAMIE, A., IGUMBOR, J., GREEN, E. & MUSIE, E. 2009. Microbiological, physico-chemical and management parameters impinging on the efficiency of small water treatment plants in the Limpopo and Mpumalanga Provinces of South Africa. *Water Sa,* 33.

O'HARA, S., HANNAN, T. & GENINA, M. 2008. Assessing access to safe water and monitoring progress on MDG7 target 10 (access to safe water and basic sanitation): Lessons from Kazakhstan. *Water Policy, IWA,* 10**,** 1-24.

O'KELLY, M. E. 1994. Spatial analysis and GIS. *Spatial Analysis in GIS***,** 65-79.

ONLINEOXFORDDICTIONARY 2011. Census, available at http://oxforddictionaries.com/, accessed on 20th May 2011.

ONS:OFFICE FOR NATIONAL STATISTICS 2010 The census in England and Wales, London, Titchfield, and Newport, available at: http://www.ons.gov.uk/census/index.html, accessed: on 11th May 2010.

OXFORD ENGLISH DICTIONARY ONLINE. 2011. Census, available at http://oxforddictionaries.com/definition/census, accessed on 10th May 2011.

PAMPALON, R., HAMEL, D., GAMACHE, P. & RAYMOND, G. 2009. A deprivation index for health planning in Canada. *Chronic Diseases in Canada,* 29.

PANDE, S., KEYZER, M. A., AROUNA, A. & SONNEVELD, B. 2008. Addressing diarrhea prevalence in the West African Middle Belt: social and geographic dimensions in a case study for Benin. *International Journal of Health Geographics,* 7**,** 17.

PARASHAR, U., BRESEE, J. & GLASS, R. 2003. The global burden of diarrhoeal disease in children. *Bulletin of the World Health Organization,* 81**,** 236-236.

PARKER, S., UPRICHARD, E. & BURROWS, R. 2007. Class places and place classes geodemographics and the spatialization of class. *Information Communication and Society,* 10**,** 902-921.

PFEFFERMANN, D. 2002. Small area estimation new developments and directions. *International Statistical Review,* 70**,** 125-143.

PLAYÁN, E. & MATEOS, L. 2006. Modernization and optimization of irrigation systems to increase water productivity. *Agricultural Water Management,* 80**,** 100-116.

POTTER, R. B. & DARMAME, K. 2010. Contemporary social variations in household water use, management strategies and awareness under conditions of water stress: The case of Greater Amman, Jordan. *Habitat International,* 34**,** 115-124.

PREGIBON, D. 1980. Goodness of link tests for generalized linear models. Journal of the *Royal Statistical Society,* 29**,** 15-24

PRÜSS, A., KAY, D., FEWTRELL, L. & BARTRAM, J. 2002. Estimating the burden of disease from water, sanitation, and hygiene at a global level. *Environmental Health Perspectives,* 110**,** 537-542.

QUALITY, D. 2006. *Guidelines for Drinking-water Quality, World Health Organization,* Geneva, Switzerland.

RAO, J. 2003. *Small Area Estimation,* New Jersey, USA, Wiley-Interscience.

RATCLIFFE, J. H. 2001. On the accuracy of TIGER-type geocoded address data in relation to cadastral and census areal units. *International Journal of Geographical Information Science,* 15**,** 473-485.

REES P, Martin. D., and WILLIAMSON P (EDS.) 2002 *The Census Data System*, John

Wiley & Sons, Chichester, England.

RELLER, M., MENDOZA, C., LOPEZ, M., ALVAREZ, M., HOEKSTRA, R., OLSON, C., BAIER, K., KESWICK, B. & LUBY, S. 2003. A randomized controlled trial of household-based flocculant-disinfectant drinking water treatment for diarrhea prevention in rural Guatemala. *The American Journal of Tropical Medicine and Hygiene,* 69**,** 411.

RICKWOOD, C. & CARR, G. 2009. Development and sensitivity analysis of a global drinking water quality index. *Environmental Monitoring and Assessment,* 156**,** 73-90.

ROBINSON, V. B. 2003. A perspective on the fundamentals of fuzzy sets and their use in geographic information systems. *Transactions in GIS,* 7**,** 3-30.

RODRIGUEZ-MOZAZ, S., DE ALDA, M. & BARCELÓ, D. 2004. Monitoring of estrogens, pesticides and bisphenol A in natural waters and drinking water treatment plants by solid-phase extraction–liquid chromatography–mass spectrometry. *Journal of Chromatography A,* 1045**,** 85-92.

ROGERSON, P. A. 2006. Statistical methods for the detection of spatial clustering in case–control data. *Statistics in Medicine,* 25**,** 811-823.

ROOT, G. 1997. Population density and spatial differentials in child mortality in Zimbabwe. *Social Science and Medicine,* 44**,** 413-421.

ROSA, G. & CLASEN, T. 2010. Estimating the scope of household water treatment in low- and medium-Income countries. *American Journal of Tropical Medicine and Hygiene,* 82**,** 289-300.

ROTHMAN, J. 1989. *Editorial, Journal of the Market Research Society*, 31,1: 1-7

SAEI, A. & CHAMBERS, R. 2003. Small Area Estimation: A review of methods based on the application of mixed models, Enhancing small area estimation techniques to meet European Needs" project IST 2000-26290, a report in the European Union's Fifth Research And Technological Development Framework Programme.

SAHN, D. E. & STIFEL, D. C. 2003. Progress toward the millennium development goals in Africa. *World Development,* 31**,** 23-52.

SALEH, M., EWANE, E., JONES, J. & WILSON, B. 2001. Chemical evaluation of commercial bottled drinking water from Egypt. *Journal of Food Composition and Analysis,* 14**,** 127-152.

SALVATI, N., TZAVIDIS, N., PRATESI, M. & CHAMBERS, R. 2007. Small area estimation via M-quantile geographically weighted regression. Unpublished paper from authors

SALWAY, S. 2006. Women's position within the household as a determinant of maternal health care use in Nepal. *International family planning perspectives,* 32**,** 17-27.

SÄRNDAL, C., SWENSSON, B. & WRETMAN, J. 2003. *Model Assisted Survey Sampling,* New York, USA, Springer Verlag.

SAUNDERS, J. 1998. Weighted Census-based deprivation indices: their use in small areas. *Journal of Public Health Medicine,* 20**,** 253-260.

SCHIRRA, S. 2008. How reliable are practical point-in-polygon strategies? *In:* Proceedings of the 16th annual European symposium on Algorithms 2008. Springer, 744-755.

SCHOUTEN, T. & MORIARTY, J. 2003. *Community water, community management: from system to service in rural areas*, Intermediate Technology. London: ITDG Publishing

SCOTT, A. J. 2009. World Development Report 2009: reshaping economic geography. *Journal of Economic Geography,* 9**,** 583.

SCOTT, C. A., EL-NASER, H., HAGAN, R. E. & HIJAZI, A. 2003. Facing Water Scarcity in Jordan. *Water international,* 28**,** 209-216.

SENIOR, M. 2002. Deprivation indicators. *In:* Rees P, Martin D and Williamson P (eds)*. The Census Data System, pp123-139*. Chichester, England

SESTER, M., ANDERS, K. H. & WALTER, V. 1998. Linking objects of different spatial data sets by integration and aggregation. *GeoInformatica,* 2**,** 335-358.

SHAMSI, U. M. 2005. *GIS Applications for Water, Wastewater, and Stormwater Systems*, CRC Press. Pennsylvania, USA

SHI, W., CHEUNG, C. K. & ZHU, C. 2003. Modelling error propagation in vector-based buffer analysis. *International Journal of Geographical Information Science,* 17**,** 251-271.

SHORDT, K., WIJK, C., BRIKKÉ, F. & HESSELBARTH, S. 2004. *Monitoring Millennium Development Goals for Water and Sanitation: A review of experiences and challenges.* Delft, the Netherlands: IRC, International Water and Sanitation Centre and KFW.

SHOWERS, K. 2002. Water Scarcity and urban Africa: an overview of urban-rural water linkages. *World Development,* 30**,** 621-648.

SIMLER, K. & NHATE, V. 2005. Poverty, inequality, and geographic targeting: Evidence from small-area estimates in Mozambique. Food consumption and nutrition division discussion, Paper No. 192. International Food Policy Research Institute, Washington, DC, pp. 43

SIMPSON, L. & TRANMER, M. 2005. Combining sample and census data in small area estimates: iterative proportional fitting with standard software. *The Professional Geographer,* 57**,** 222-234.

SIMPSON, L. & YU, A. 2003. Public access to conversion of data between geographies, with multiple look up tables derived from a postal directory. *Computers, Environment and Urban Systems,* 27**,** 283-307.

SINGH, R., SEMWAL, D., RAI, A. & CHHIKARA, R. 2002. Small area estimation of crop yield using remote sensing satellite data. *International Journal of Remote Sensing,* 23**,** 49-56.

SINGLETON, A. D. & LONGLEY, P. A. 2009. Geodemographics, visualisation, and social networks in applied geography. *Applied Geography,* 29**,** 289-298.

SIRI, J. G., LINDBLADE, K. A., ROSEN, D. H., ONYANGO, B., VULULE, J. M., SLUTSKER, L. & WILSON, M. L. 2008. A census-weighted, spatially-stratified household sampling strategy for urban malaria epidemiology. *Malaria Journal,* 7**,** 39.

SKINNER, B. & SHAW, R. 2003. Household water treatment. *Journal of Appropriate Technologies for Water Supply and Sanitation* 15-18.

SMITH, A., LINGAS, E. & RAHMAN, M. 2000. Contamination of drinking water by arsenic in Bangladesh: a public health emergency. *Bulletin of the World Health Organization,* 78**,** 1093-1103.

SNEDECOR, W. & COCHRAN, G. (1989). *Statistical Methods,* Eighth Edition , Iowa State University Press, Iowa, USA.

SOBSEY, M. 2002. Managing water in the home: accelerated health gains from improved water supply. *, Geneva, World Health Organization*

SOLIMAN, A. M. 1992. Housing consolidation and the urban poor: the case of Hagar El Nawateyah, Alexandria. *Environment and Urbanization,* 4**,** 184-195.

SOLTAN, M. 1999. Evaluation of ground water quality In Dakhla Oasis (Egyptian Western Desert). *Environmental Monitoring and Assessment,* 57**,** 157-168.

SOUTER, P., CRUICKSHANK, G., TANKERVILLE, M., KESWICK, B., ELLIS, B., LANGWORTHY, D., METZ, K., APPLEBY, M., HAMILTON, N. & JONES, A. 2003. Evaluation of a new water treatment for point-of-use household applications to remove microorganisms and arsenic from drinking water. *Journal of Water and Health,* 1**,** 73-84.

STAT TREK, Statistical Tutorial 2011. Cluster Sampling, available on http://stattrek.com/Lesson6/CLS.aspx, accessed on 22nd January 20111.

SULLIVAN, A. C., CHOHEN, A., FAURES, J., & SANTINI, J 2009. The Rural Water Livelihoods Index. *In:* IN WATER, F. E. (ed.) *Available at* http://www.fao.org/nr/water/docs/FAOW_RWLI_paper.pdf, accessed on 18th February 2010.

SULLIVAN, C. 2002. Calculating a Water Poverty Index. *World Development,* 30**,** 1195-1210.

SULLIVAN, C. & MEIGH, J. 2003. Considering the Water Poverty Index in the context of poverty alleviation. *Water policy,* 5**,** 513-528.

SULLIVAN, C., MEIGH, J., GIACOMELLO, A., FEDIW, T., LAWRENCE, P., SAMAD, M., MLOTE, S., HUTTON, C., ALLAN, J. & SCHULZE, R. Year. The Water Poverty Index: Development and application at the community scale. Natural Resources Forum, 27, pp. 189 – 199.

SULLIVAN, C., MEIGH, J. & LAWRENCE, P. 2006. Application of the water poverty index at different scales: A cautionary tale. *Water international,* 31**,** 412-426.

SUNNA, N. 2003. Water quality and intermittent water supplies, consultation on minimum household water security requirement and health, Amman, available at http://www.emro.who.int/ceha/media/powerpoint/quality.pps, accessed on 13th April 2011

SYRETT, S. & NORTH, D. 2008. *Renewing neighbourhoods: work, enterprise and governance*, Policy Press, University of Bristol, England.

TANSER, F. C. & LE SUEUR, D. 2002. The application of geographical information systems to important public health problems in Africa. *International Journal of Health Geographics,* 1**,** 1-9.

THOMPSON, J. 2001. *Drawers of water II: 30 years of change in domestic water use & environmental health in east Africa. Summary*, International Institute for Environment and Development, London, England

THOMPSON, J., PORRAS, I. T., WOOD, E., TUMWINE, J. K., MUJWAHUZI, M. R., KATUI-KATUA, M. & JOHNSTONE, N. 2000. Waiting at the tap: Changes in urban water use in East Africa over three decades. *Environment and Urbanization,* 12**,** 37.

TIWARI, S. S. K., SCHMIDT, W. P., DARBY, J., KARIUKI, Z. G. & JENKINS, M. W. 2009. Intermittent slow sand filtration for preventing diarrhoea among children in Kenyan households using unimproved water sources: randomized controlled trial. *Tropical Medicine & International Health,* 14**,** 1374-1382.

TOBLER, W. 1979. Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association,* 74**,** 519-530.

TOBLER, W. & KENNEDY, S. 1985. Smooth Multidimensional Interpolation. *Geographical Analysis,* 17**,** 251.

TOWNSEND, P., PHILLIMORE, P. & BEATTIE, A. 1988. *Health and Deprivation: Inequalities and the North*. London: Croom Helm.

TSUZUKI, Y. 2009. Comparison of pollutant discharge per capita (PDC) and its relationships with economic development: An indicator for ambient water quality improvement as well as the Millennium Development Goals (MDGs) sanitation indicator. *Ecological Indicators,* 9**,** 971-981.

TURNER, A. 2001. *Sampling Manual for Facility Surveys: For Population, Maternal Health, Child Health and STD Programs in Developing Countries*, MEASURE Evaluation Project, Carolina Population Center, University of North Carolina, Chapel Hill.

UN 2000. United Nations Millennium Declaration, available at http://www.un.org/millennium/declaration/ares552e.htm, accessed on 12th September 2009

UN. 2011. *Millennium Development Goals Indicators*, available: http://unstats.un.org/unsd/mdg/Default.aspx, accessed 20th January 2011.

UN-HABITAT 2010. The Lake Victoria Region Water and Sanitation Initiative. *In:* UNITED NATIONS OFFICE IN NAIROBI (UNON) JAMES OHAYO, T. O. (ed.) http://www.unhabitat.org/pmss/listItemDetails.aspx?publicationID=2578, accessed data 26th February 2010.

UNICEF & WHO 2006 Meeting the MDG drinking water and sanitation target : the urban and rural challenge of the decade,WHO Library Cataloguing-in-Publication Data, available at http://www.who.int/water_sanitation_health/monitoring/jmp2006/en/index.html, accessed on 13th July 2009.

UN, 2003. Indicators for Monitoring the Millennium Development Goals. New York, NY: United Nations. available at: http://devdata.worldbank.org/gmis/mdg/UNDG%20document_final.pdf, accessed on 8th September 2009.

UPRICHARD, E., BURROWS, R. & PARKER, S. 2009. Geodemographic code and the production of space. *Environment and Planning A,* 41**,** 2823-2835.

USFAR, A. A., ISWARAWANTI, D. N., DAVELYNA, D. & DILLON, D. 2010. Food and personal hygiene perceptions and practices among caregivers whose children have diarrhea: a qualitative study of urban mothers in Tangerang, Indonesia. *Journal of Nutrition Education and Behavior,* 42**,** 33-40.

UTHMAN, O. A. & KONGNYUY, E. J. 2008. A multilevel analysis of effect of neighbourhood and individual wealth status on sexual behaviour among women: evidence from Nigeria 2003 Demographic and Health Survey. *BMC International Health and Human Rights,* 8**,** 9.

VAESSEN, M., THIAM, M. & THANH, L. 2004. *The Demographic and Health Surveys. Household sample surveys in developing and transition countries: design, implementation and analysis*. New York: United Nations, Department of Economic and Social Affairs.

VALDIVIA, M. & PARA EL DESARROLLO, G. A. 2006. Institutional Change in the Health Sector and the Millennium Development Goals, Paper presented at the 7th Annual Global Development Network Conference, Saint Petersburg, Russia; January 19-21, 2006,.

VERMA, V. & LÊ, T. 1996. An Analysis of Sampling Errors for the Demographic and Health Surveys. *International Statistical Review,* 64**,** 265-294.

VOAS, D. & WILLIAMSON, P. 2001. The diversity of diversity: a critique of geodemographic classification, 33 (1): 63-76.

VÖRÖSMARTY, C. J., GREEN, P., SALISBURY, J. & LAMMERS, R. B. 2000. Global water resources: vulnerability from climate change and population growth. *Science*, 289, 284.

WALFORD, N. 2007. Geographical and geodemographic connections between different types of small area as the origins and destinations of migrants to Mid-Wales. *Journal of Rural Studies,* 23**,** 318-331.

WANG, L. 2002. *Health outcomes in low-income countries and policy implications: Empirical findings from demographic and health surveys,* World Bank, Environment Department, available at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.17.518&rep=rep1&type=pdf,accessed on 12th October 2009

WATERAID, ODI. 2009. Learning for advocacy and good practice-WaterAid water point mappingOverseas Development Insitute, available at http://www.odi.org.uk/resources/download/2952.pdf, accessed on 5th September 2009.

WEBBER, R. 2004. Designing geodemographic classifications to meet contemporary business needs. *Interactive Marketing,* 5**,** 219-237.

WEBBER, R. & LONGLEY, P. 2003. Geodemographic analysis of similarity and proximity: their roles in the understanding of the geography of need. ESRI Press, Redlands, USA

WEBSTER, R. & OLIVER, M. A. 2007. *Geostatistics for Environmental Scientists,* Chichester, England, Wiley.

WEEKS, J. R. 2004. *The Role of Spatial Analysis in Demographic Desearch*Oxford University press, New York, USA,.

WEISS, M. 2000. *The clustered world: How we live, what we buy, and what it all means about who we are*, Little, Brown and Company, New York, USA.

WHITE, G., BRADLEY, D. & WHITE, A. 2002. Drawers of water: domestic water use in East Africa. *Bulletin of the World Health Organization,* 80**,** 63-73.

WHO 2004. Guide to promotion of drinking water disinfection in emergencies, World Health Organization Regional Office for the Eastern Mediterranean,Amman, Jordan, available at http://www.emro.who.int/ceha/pdf/DrinkingWater_Disinfection_En.pdf, accessed on 5th October 2009.

WHO 2006.*World Health Statistics2006*, WHO press, World Health Organization Geneva. Switzerland.

WHO, Unicef & Joint Monitoring Programme for Water Supply and Sanitation. 2000. Global water supply and sanitation assessment 2000 report,available at http://www.who.int/water_sanitation_health/monitoring/globalassess/en/ accessed on 10th of November 2010

WHO, W. 2008. *Guidelines for Drinking-Water Quality*. WHO, Geneva, Volume 1**,** 392–394.

WHO/UNICEF 2000. *Global Water Supply and Sanitation Assessment 2000 Report,* Geneva, World Health Organization.

WHO/UNICEF 2009a. *The Joint Monitoring Programme for Water supply and Sanitation, The Joint Monitoring Programme: Definitions,* http://www.wssinfo.org/en/122_definitions.html, accessed on 6th September 2009.

WHO/UNICEF 2009b. *The Joint Monitoring Programme for Water supply and*

Sanitation, The Joint Monitoring Programme: Data collection and Analysis, http://www.wssinfo.org/en/122_definitions.html, accessed on 7th September 2009.

WHO/UNICEF 2010*. Rapid assessment of drinking water quality in the Hashemite Kingdom of Jordan, implementation in 2004-2005*, Geneva , Switzerland.

WHO/UNICEF 2010 *The Joint Monitoring Programme for Water supply and Sanitation, The Joint Monitoring Programme: Data collection and Analysis*, http://www.wssinfo.org/definitions/infrastructure.html ,accessed on 25th ptember 2010.

WHO/UNICEF 2010b. *Rapid assessment of drinking water quality in the Hashemite Kingdom of Jordan, country report of the pilot project*, available at http://www.wssinfo.org/fileadmin/user_upload/resources/RADWQ_Jordan.pdf, accessed on 20th march 2011.

WILLARDSON, L. S., BOELS, D. & SMEDEMA, L. K. 1997. Reuse of drainage water from irrigated areas. *Irrigation and Drainage Systems,* 11**,** 215-239.

WILLIAMS, B. G., GOUWS, E., BOSCHI-PINTO, C., BRYCE, J. & DYE, C. 2002. Estimates of world-wide distribution of child deaths from acute respiratory infections. *The Lancet Infectious Diseases,* 2**,** 25-32.

WIRTH, M. E., BALK, D., DELAMONICA, E., STOREYGARD, A., SACKS, E. & MINUJIN, A. 2006. Setting the stage for equity-sensitive monitoring of the maternal and child health Millennium Development Goals. *Bulletin of the World Health Organization,* 84**,** 519-527.

WORLD HEALTH ORGANIZATION, G. 2004. *Guidelines for drinking-water quality*, World Health Organization, Distribution and Sales Geneva 27 CH-1211 Switzerland.

WRIGHT, J. & GUNDRY, S. 2009. Household characteristics associated with home water treatment: an analysis of the Egyptian Demographic and Health Survey. *Journal of Water and Health,* 7**,** 21.

WRIGHT, J., GUNDRY, S. & CONROY, R. 2004. Household drinking water in developing countries: a systematic review of microbiological contamination between source and point-of-use. *Tropical Medicine and International Health,* 9**,** 106.

WRIGHT, J. & GUNDRY, S. W. 2009. Household characteristics associated with home water treatment: an analysis of the Egyptian Demographic and Health Survey. *Journal of Water and Health,* 7**,** 21-30.

XUE, L., WU, C., GENG, F. & WU, Y. 2009. Water quality health and hazard risk assessment on drinking water sources. *In:* 2009 3rd International Conference on Bioinformatics and Biomedical Engineering (iCBBE 2009),
11-16 June 2009 Beijing, China.

YANG, H. & ZEHNDER, A. J. B. 2002. Water scarcity and food import: A case study for southern Mediterranean countries. *World Development,* 30**,** 1413-1430.

YANSANEH, I. S. 2005. Construction and use of sample weights. *Designing*

Household Surveys Samples: Practical Guidelines, New York, United Nations.

ŽALIK, B. & KOLINGEROVA, I. 2001. A cell-based point-in-polygon algorithm suitable for large sets of points. *Computers and Geosciences,* 27**,** 1135-1145.

ZETTER, R. & HASSAN, A. M. 2002. Urban economy or environmental policy? The case of Egypt. *Journal of Environmental Policy & Planning,* 4**,** 169-184.

ZHANG, J. & GOODCHILD, M. F. 2002. *Uncertainty in geographical information,* London, Taylor and Francis.

ZHANG, J. & YU, K. 1998. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *Journal of the American Medical Association,* 280**,** 1690-1691.

ZOHRY, A. G. 2002. *Rural-to-urban labor migration: a study of Upper Egyptian laborers in Cairo.* Doctor of Philosophy (D.Phil.), University of Sussex.

ZVIDZAI, C., MUKUTIRWA, T., MUNDEMBE, R. & SITHOLE-NIANG, I. 2007. Microbial community analysis of drinking water sources from rural areas of Zimbabwe. *African Journal of Microbiology Research,* 1**,** 100-104.

Computers, Environment and Urban Systems xxx (2011) xxx–xxx

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/compenvurbsys

Problems of spatial linkage of a geo-referenced Demographic and Health Survey (DHS) dataset to a population census: A case study of Egypt

Shawky Mansour [a,b,*], David Martin [a], Jim Wright [a]

[a] School of Geography, University of Southampton, Southampton, SO17 1BJ, UK
[b] Assistant Lecturer at the Department of Geography and GIS, Alexandria University, Alexandria, Egypt

ARTICLE INFO

Article history:
Received 10 November 2010
Received in revised form 3 April 2011
Accepted 5 April 2011
Available online xxxx

Keywords:
GIS
Spatial linkage
DHS
GPS
Uncertainty
Egypt

ABSTRACT

GPS coordinates are increasingly available as spatial references on population surveys in the developing world, where high-resolution address and street mapping are absent. This potentially offers opportunities to enhance national census data by spatial linkage with survey sources. The paper explores the use of GPS-referenced Demographic and Health Survey (DHS) data in combination with census data in Egypt and identifies errors in coordinate referencing. The study develops a practical approach to the measurement of spatial uncertainty in this situation and assessment of its impact on data linkage. The analysis specifically addresses the analytical implications at three different spatial scales and is internationally relevant to the handling of GPS-referenced DHS data in GIS.

1. Introduction

Data linkage has long been cited as one of the most important practical applications of GIS (Longley, Goodchild, Maguire, & Rhind, 1999). The integration of different datasets through spatial co-location potentially provides enormous power for addressing questions about geographic and socioeconomic relationships and can lead to the development of rich spatial analytic research in a variety of application fields. This paper addresses one such application field, namely the linkage of census and sample survey data in a developing world setting. Our study demonstrates the importance of assessment of spatial error in such a context and its impacts on data linkage. Specifically, we are concerned with the linkage of a geo-referenced Demographic and Health Survey (DHS) dataset to a national population census. The DHS is a global survey, launched in 1984 by the United States Agency for International Development (USAID), currently covering 84 developing countries (MEASURE DHS 2010a). The core set of questions covered by the DHS concentrates on household characteristics, particularly focusing on fertility, family planning, infant and child mortality, reproductive health, demography, child health, nutrition, malaria, and socioeconomic characteristics. The survey thus provides researchers with important socioeconomic and health information in contexts, where this will not usually be available from other sources. It is also a spatial dataset, providing geographic information that enables spatial patterns in population characteristics to be explored.

Since the early 1990s Global Positioning System (GPS) technology has been increasingly used to georeference DHS data (Weeks, 2004). Montana and Spencer (2004) describe GPS data collection and use in the DHS: surveys are conducted with multiple households within local sampling units and these results are collectively referenced by a single 'GPS cluster' location. Previously, georeferencing was limited to high level (regional, provincial) area codes, so the use of GPS makes recent DHS data far more amenable to spatial analysis and linkage to other data sources using GIS.

There is a growing literature on DHS applications using GIS and spatial analysis techniques, particularly in child mortality and disease (Uthman & Kongnyuy, 2008; Wirth et al., 2006; Gemperli et al., 2004). Aside from spatial analyses of DHS variables, there are several studies that have enriched DHS data by using GPS clusters to link to non-census datasets. For example, Stephenson, Baschieri, Clements, Hennink, and Madise (2007) linked data for DHS GPS clusters in six African countries to data on rainfall, habitat and transportation infra-structure in their study of contextual influences on contraceptive use. Similarly, in a multi-level modeling study of influences on fertility, Baschieri (2007) used GPS cluster locations from the 2000 Egyptian DHS to link to land use and transportation infra-structure data. Internationally, Balk et al. (2005) linked GPS clusters for 19 African countries to elevation, soils, and land cover data in their analysis of child malnutrition.

* Corresponding author at: School of Geography, University of Southampton, Southampton, SO17 1BJ, UK. Tel.: +44 (0) 238059 2215; fax: +44 (0) 23 8059 3295.
E-mail addresses: Saem1c08@soton.ac.uk (S. Mansour), D.J.Martin@soton.ac.uk (D. Martin), J.A.Wright@soton.ac.uk (J. Wright).

0198-9715/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compenvurbsys.2011.04.001

There have also been several studies which linked DHS cluster points to census boundaries and thus to census variables. Since DHS data do not provide statistically representative estimates of household characteristics at the sub-provincial level, such studies to date have used census data in combination with DHS data in one of two ways: to derive co-variates that might explain individual or household characteristics within the DHS, or to undertake small area estimation. For example, Johnson, Chandra, Brown, and Padmadas (2010) combined data from the 2003 Ghanaian DHS into the 2000 Population and Housing Census to study maternal health and estimate institutional births in Ghana. They used small area estimation to model the proportion of institutional births at district level. Similarly, to study sub-provincial child malnutrition prevalence in Cambodia, Fujii (2005) carried out a small area estimation (commune-level) procedure by linking the 2000 DHS with individual level data from the 1998 Cambodian census. McGuire (2000) used a different approach to aggregate DHS data to district level in parts of four Sahel countries. They used a spatial filtering approach to draw on data for GPS clusters neighboring a given district, arguing that the resultant increased sample size improved the robustness of the district-level estimates. Pande, Keyzer, Arouna, and Sonneveld (2008) studied diarrhea prevalence across Benin by linking the DHS household clusters to various spatial datasets, including diarrheal prevalence estimates linked to county boundaries, to identify factors that might be responsible for diarrhea prevalence, particularly drinking water sources and water quality. Feldacker, Emch, and Ennett (2010) used GPS cluster locations to link DHS individual level data on HIV status to socio-economic variables attached to groups of enumeration areas, so as to investigate community-level risk factors for HIV in Malawi. However, none of these studies considers the impact of spatial uncertainty on the point-in-polygon linkage process. Therefore, there has been little coverage in the literature of linkage between DHS clusters and administrative boundary data.

Thus, the utility of the DHS can be considerably enhanced by linkage to available census data, which provides coverage of the entire population, albeit with less detailed information on some key health and lifestyle topics. However, these datasets cannot be directly linked using standard published lookup tables, as might be possible in a developed world setting (Simpson & Yu, 2003) and it is therefore necessary to develop appropriate GIS approaches for spatial linkage between the two sources. However, an assessment of the positional accuracy of the DHS and its implications for GIS-based linkage to census data is essential. We here concentrate on the assessment and accommodation of spatial uncertainty in the DHS to census linkage, so as to facilitate an important enhancement to the data available for analysis.

There is an extensive literature on spatial uncertainty, its impacts on subsequent analysis and data linkage operations, and its communication to spatial data users (Heuvelink & Burrough, 2002). Within the over-arching concept of uncertainty, Fisher (1999) distinguishes between error affecting well-defined spatial objects, and vagueness and ambiguity which affect poorly-defined spatial objects. Since the spatial entities in our study (census units and GPS clusters) are well defined, we are primarily concerned with error, namely the difference between locations or attributes held digitally and their true values. Fisher (1999) also distinguishes a further sub-category of uncertainty in spatial databases that he terms controlled uncertainty, in which uncertainty is deliberately introduced into a socio-economic data set. Given the deliberate perturbation of locational data in the DHS, this concept is also relevant to our case study.

More specifically, in our case study we were concerned with a specific form of spatial data manipulation. Our study entailed a point-in-polygon operation, with the points being DHS GPS clusters and the polygons being census boundaries. Several authors have developed procedures for handling uncertainty in point-in-polygon operations, for example using probability density functions or fuzzy set theory (Cheung, Shi, & Zhou, 2004; Leung & Yan, 1997), or by reducing the problem to a point-in-triangle problem (Leung, Ma, & Goodchild, 2004). As an alternative approach, Hunter (1999) systematically perturbed the coordinates of points and polygon boundaries according to known probability density functions. He then used Monte Carlo simulation to assess the influence of positional uncertainty on point-in-polygon results. However, these methods require that the characteristics of any positional uncertainty are known from the outset. Ratcliffe (2001) characterized uncertainty in point-in-polygon operations involving address locations and census tracts without any prior knowledge of the uncertainty characteristics of the two data sets used. His approach involved a comparison of two different methods (geocoding and cadastral) of generating address locations. However, in our study, no information was available concerning the positional accuracy of our point and polygon data. This is not uncommon in developing world contexts. Furthermore, we were unable to generate locations for our point or polygon data via different methods. Therefore, rather than following the methodology of these previous studies, our approach was to identify attribute labels that were inconsistent with the locational data for our points and polygons. We then used information about feature attributes that were inconsistent with location to model uncertainty in the point-in-polygon operation that we were undertaking.

The remainder of this paper is structured as follows: the next section describes the available datasets in more detail and vdescribes our methodology. In Section 3, we present our findings, before evaluating their importance in Section 4 and drawing conclusions in Section 5.

2. Materials and methods

2.1. Data sources

The paper uses data from both the 2005 Egyptian DHS and 2006 national census. The Egyptian Central Agency for Public Mobilisation and Statistics (CAPMAS) is responsible for censuses, as well as surveys, and other public statistics. There are three levels in census administrative geography; governorate, kism/markaz and shyakha:

- Governorates consist of several districts and an administrative capital. Administratively, Egypt is divided into 28 governorates plus Luxor Supreme City Council. In the 2006 Egyptian census, on average Egyptian governorates comprised 3,089,029 people and 735,961 households.
- Kisms (urban) and markazes (both urban and rural) are the next level below the governorate and they combine a number of neighborhoods and villages (an Egyptian village is a rural community consists of a group of settlements smaller than a town and larger than a shyakha with a largely agricultural workforce). In 2006, there were 360 kisms/markazes with an average population of 202,000 each.
- Shyakhas are the smallest units in the Egyptian Census and are also used as electoral units. In 2006, there were 5705 shyakhas, with an average population of 12,750 each.

Cairo is the most populous governorate (population 6,758,581) while Suez is the least (population 512,135). There are also five frontier (desert) governorates (Matroh, New Valley, North Sinai, South Sinai and Red Sea) with very low population densities (on average 291,726 people and 60,364 households) A map of Egyptian governorates, showing their names, is provided in Fig. 1. Although

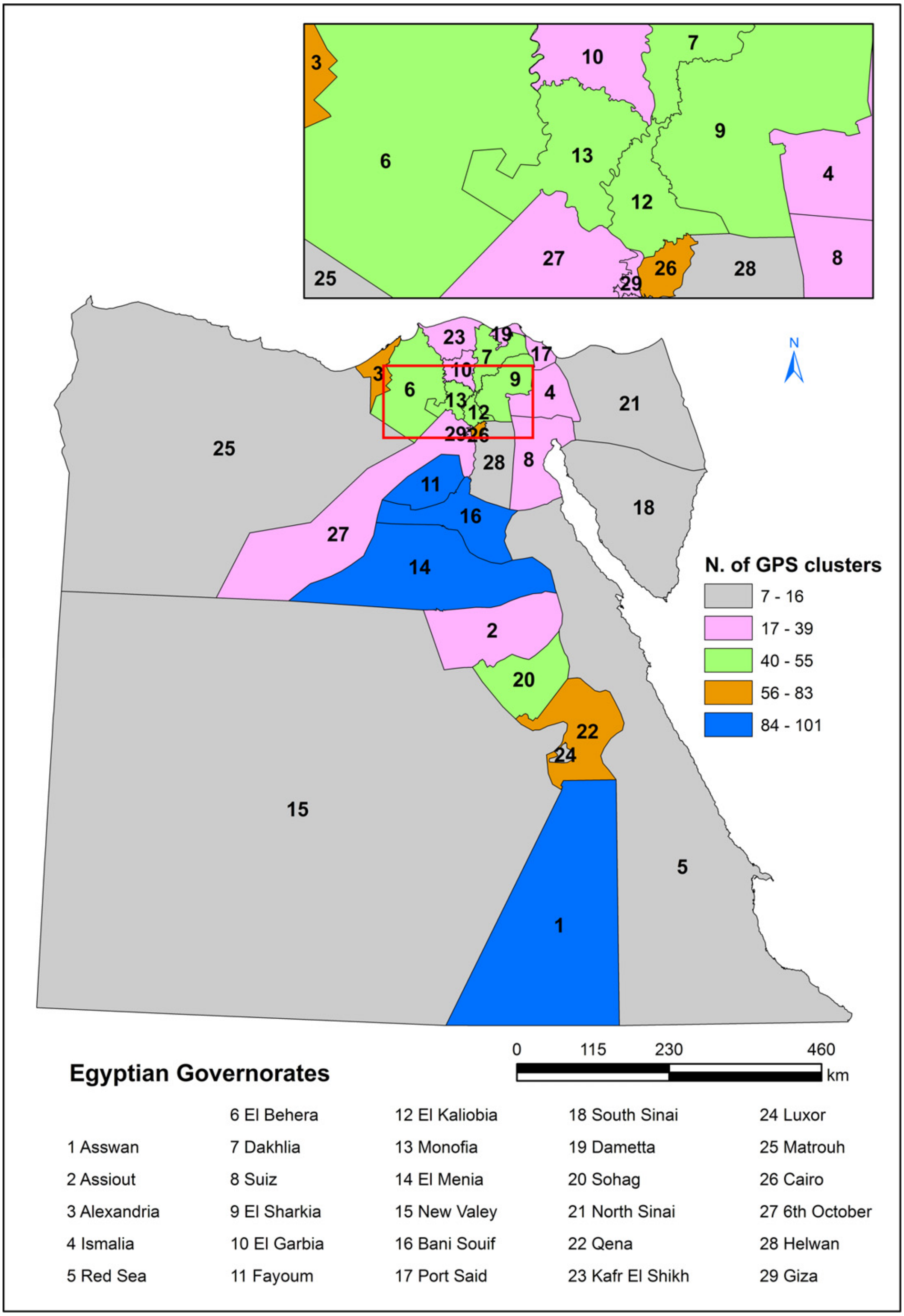


Fig. 1. Distribution of GPS clusters from the 2005 Egyptian DHS across Egyptian governorates.

covered by governorate boundaries, some uninhabited desert areas do not form part of any kism, markaz or shyakha. The GIS Center at CAPMAS in Egypt maintains the three digital census geographies (governorate, kism/marqaz, shyakha). We used CAPMAS digital map layers in the analysis that follows: whilst boundaries for each administrative layer were created in the same way, the scale of the source data is not documented.

The MEASURE DHS international system of surveys is funded by USAID and implemented by an ICF International Company (ICF Macro) in partnership with other organizations (MEASURE DHS, 2010b). The 2005 Egyptian DHS was conducted by the Ministry of Health and Population (MOHP) and National Population Council (NPC) and implemented by El-Zanaty and Associates with technical support provided by ORC Macro. In the DHS, groups of households are geo-referenced to a central, averaged location, known as a GPS cluster. The mean GPS cluster size is 17 households (range: 4–81) within the 2005 Egyptian DHS and there are 1359 GPS clusters within the survey. Two different methodologies are used to generate GPS cluster locations within the DHS. Using one methodology (labeled as 'AVE' within the DHS data), the location of each household within a cluster is surveyed individually using GPS and the coordinates of all households within the cluster are then averaged. Under the second methodology (labeled 'GPS'), in situations, where household clusters are dispersed, an approximate center point for the cluster is assigned using a single GPS reading (Burgert, 2009). Each cluster is also labeled with a governorate name. Therefore, geographical referencing of the Egyptian DHS takes two forms, the inclusion of a governorate identifier and a DHS GPS point, the latter having a far higher spatial resolution than the governorate level. There are 61 clusters that lack coordinates, and so the 2005 Egyptian DHS contains 1298 GPS cluster points distributed across 28 Egyptian governorates and Luxor City Council. The shading in Fig. 1 indicates the distribution of these locations, reflecting the massive concentration of population along the River Nile and its delta. Our correspondence with the survey team indicates that the GPS clusters locations are scrambled by MEASURE DHS to preserve the confidentiality of survey respondents. Specifically, latitudes and longitudes of cluster are projected to Universal Transverse Mercator coordinates, and displaced a random distance in a random direction. The displacement distance is greater in rural areas (up to 5 km) than urban areas (up to 2 km) and only takes place in DHS surveys with an HIV/AIDS component. Every hundredth rural cluster is displaced by up to 10 km. The Egyptian DHS is based on a three-stage probability sample design (El-Zanaty & Way, 2006). The first stage draws on an urban/rural classification for the six major Egyptian regions: the Urban Governorates, Urban Lower Egypt, Rural Lower Egypt, Urban Upper Egypt, Rural Upper Egypt and the Frontier Governorates. CAPMAS provided lists of villages, towns and shyakhas as the sample frames for both urban and rural areas. The list of names of spatial units reflected the geographic situation at the time of survey. 682 primary sampling units (PSUs) were selected from this list and their names are recorded in the documentation. In the second stage, each PSU was divided into a number of equally sized parts of approximately 5000 people. For large PSUs (with populations of 20,000 or more), two of these parts were selected, while from smaller towns, villages and shyakhas one part was chosen. Each part was divided into a number of equally sized segments and two segments were selected from each PSU. In the third stage, a list of households was drawn up for each segment and a systematic sample chosen from those households that included married women aged 15–49. This gave a total sample of 22,807 households selected for the survey (El-Zanaty & Way, 2006). The DHS survey provides representative estimates of household characteristics (e.g. of the percentage of households using groundwater sources) at national and governorate level and for rural versus urban areas, but does not generate representative estimates at sub-governorate level.

2.2. Evaluation of positional uncertainty

The GPS clusters are not identified by individual kism/markaz or shyakha names so we are reliant on spatial linkage to make any type of association between the two data sources. We might expect this to be a straightforward process with each cluster point simply being allocated to the polygon within which it falls. However, a simple overlay of GPS cluster locations on governorate boundaries identified several cluster points that lay completely outside governorate boundaries. Therefore, to investigate potential uncertainty involved in this linkage operation further, we conducted separate analyses at each of the three administrative levels:

2.2.1. Governorate level analysis

The GPS cluster points and census polygons were overlaid using ArcGIS 9.3 software to identify discrepancies between DHS governorate labels for GPS clusters and census governorate polygons. Distance from each DHS cluster point to each governorate polygon was obtained for mislabeled points, i.e. those, where the DHS governorate label and point locations were inconsistent. We also tested for spatial clustering in mislabeled points using Cuzick and Edwards (1990) test. Since mislabeled points appeared to be more common near governorate boundaries, we used logistic regression to model the effect of distance to nearest governorate boundary on governorate labeling errors and obtain the probability of each cluster point being assigned to an incorrect polygon. We also investigated whether the error pattern differed between rural and urban GPS clusters and according to the survey methodology used to generate GPS clusters. A series of logistic regression models was developed in Stata using different predictors (GPS survey methodology used, rural versus urban clusters, and distance to governorate boundary) and transformations to improve the strength of the model. We evaluated the performance of the model at each stage by assessing overall goodness of fit through the log likelihood and pseudo R^2 statistics, significance of individual coefficients, and testing for possible misspecification using the *linktest* command (Hosmer & Lemeshow, 2000). The global Moran's I statistic was also used to test for spatial autocorrelation in the standardized Pearson residuals of each logistic regression model.

2.2.2. Kism/markaz level analysis

Other than spatial location, no ancillary information is provided at the kism level to match DHS GPS clusters to individual kism/markaz. Therefore, the logistic regression coefficients from the governorate level modeling process were applied to distances calculated using the kism/markaz level boundaries. We used these to predict the probability of each GPS cluster being in its correct kism/markaz. Based on these probabilities, we then estimated the mean number of times that a GPS cluster would lie outside the correct kism through a Monte Carlo simulation. Specifically, for each GPS cluster we generated 100 random numbers from a uniform distribution between 0 and 1 and compared these to the probabilities generated by the logistic regression equation.

2.2.3. Shyakha level analysis

Since MEASURE DHS publishes a list of the shyakhas that were sampled as part of the 2005 Egyptian DHS, it is also possible to evaluate positional uncertainty using information about these lower level administrative units, even though individual clusters are not identified by shyakha names. To create a new list of sampled shyakhas for the Egyptian DHS, each of the shyakhas/towns used by MEASURE DHS in the 2005 survey (El-Zanaty & Way, 2006) were matched to the Egyptian Census shyakha layer via a geocoding process within ArcGIS. This took account of shyakha locations within governorates in matching DHS shyakha names to those from the census. Experimentation with geocoding parameters suggested that it was necessary to use low spelling sensitivity, minimum candidate and minimum match parameters since shyakha names had been transcribed from Arabic into English. Interactive, manual matching was used to geocode sampled shyakha names, where no automatic match was possible. Having thus generated a map layer depicting the shyakhas sampled for the EGYPTIAN DHS within each governorate, we then superimposed GPS cluster locations. We identified those GPS clusters that lay within sampled shyakha boundaries and for the remaining GPS clusters, calculated distances to the nearest sampled shyakha boundary.

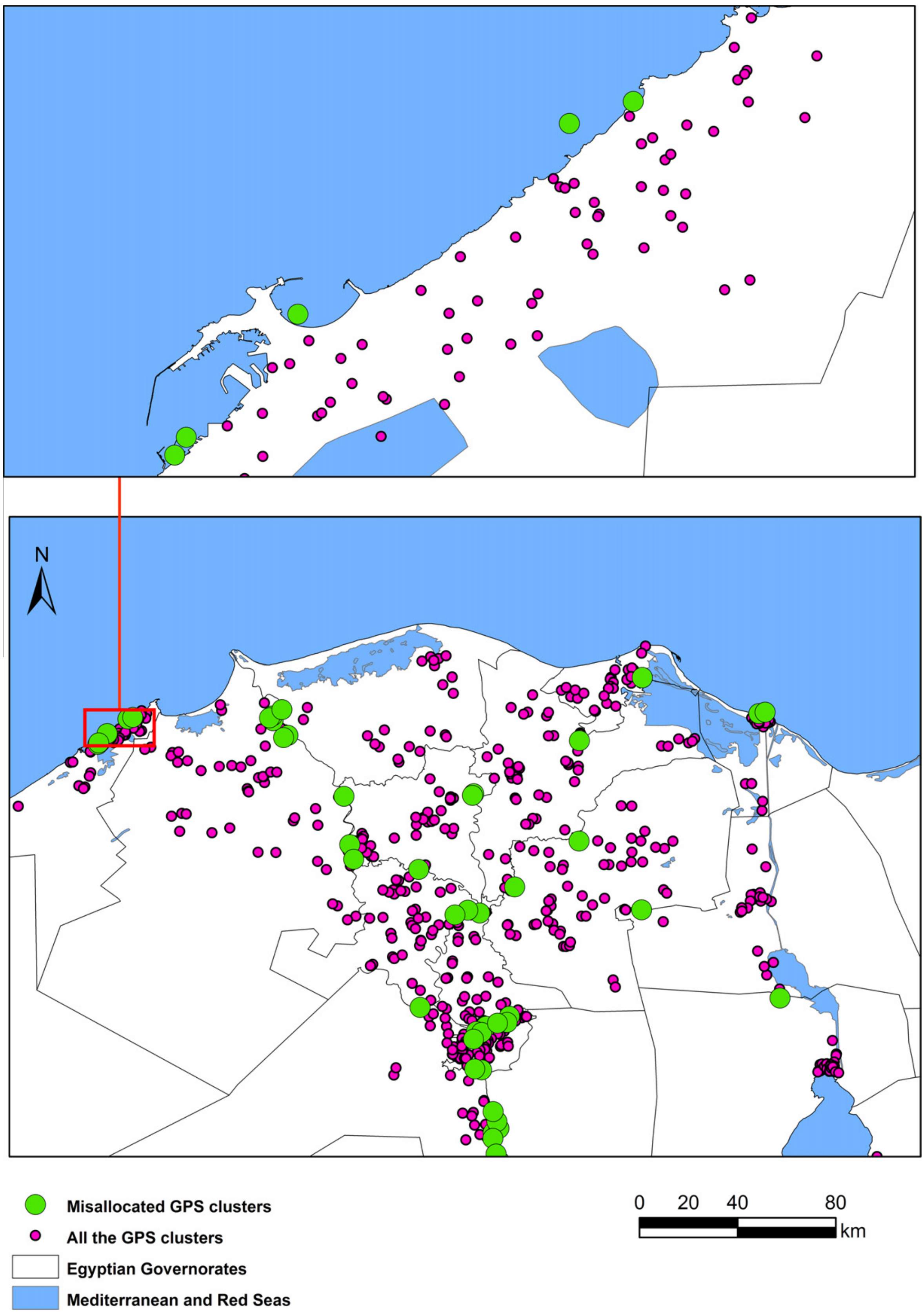


Fig. 2. Spatial distribution of DHS GPS clusters in Delta Governorates and Greater Cairo highlighting those misallocated outside correct boundaries.

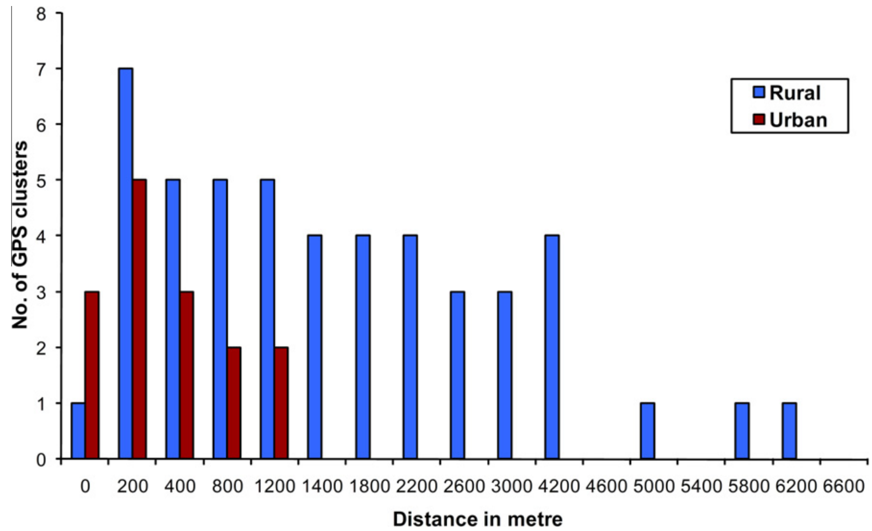


Fig. 3. Distribution of misallocated rural and urban GPS clusters by distance to nearest Egyptian governorate boundary.

Table 1
Coefficients for logistic regression model, predicting DHS GPS clusters with governorate labels inconsistent with their locations (based on 1298 GPS clusters).

Variable	Coefficient (95% confidence limits)	Z value
All clusters: $\sqrt{}$(distance in meters to nearest governorate boundary)	−0.056 (−0.069 to −0.043)	−8.37([a])
Urban clusters only: $\sqrt{}$(distance in meters to nearest governorate boundary)	−0.078 (−0.108 to −0.048)	−5.08([a])
Constant	0.74 (0.15 to 1.33)	2.48([b])

[a] Indicates significance at the 99% level.
[b] Indicates significance at the 95% level; model pseudo $R^2 = 0.39$; log likehood = −154.4.

3. Results

3.1. Governorate level analysis

When governorate labels for DHS GPS clusters are compared with their locations, there are at least 63 cluster points either mislabeled or lying in incorrect positions (Fig. 2). As is apparent from this illustration, such points lie closer to governorates boundaries, particularly in the Greater Cairo Region, Valley, Delta, and Coastal Governorates (Fig. 2 shows an example of misallocated points in the Delta with an inset showing Alexandria, where points in the sea are evident). No such errors were identified in the cluster points for the Frontier Governorates such as New Valley, North and South Sinai Governorates, where no communities are found close to governorate borders. Overall, 10 points lie outside all governorate polygons, eight of them in the Mediterranean Sea. One point belongs to Matrouh, five belong to Alexandria, two points belong to Port Said Governorate and two to the Red Sea Governorate. Except for the North and south Sinai Governorates, all other coastal governorates have GPS cluster points that lie outside the corresponding governorate boundary.

Fig. 3 shows the distribution of distances from misallocated cluster points to the governorates to which they belonged. It can be seen that most misallocated points were displaced by distances between 400 and 2000 m. Roughly 20% of points showed displacement distances of less than 200 m while 10% of the points showed displacement distances of between 200 m and 400 m. Six percent of points were displaced by distances of more than 4000 m and less than 5200 m.

The output of Cuzick and Edwards' k-nearest neighbor statistic suggested that there was spatial clustering in GPS clusters with governorate labels inconsistent with their locations ($T = 24$; $k = 1$; $p = 0.001$). The spatial clustering was generally significant at the 99% level when up to six neighbors were considered.

We began by fitting a logistic regression model of distance to nearest governorate boundary, but the *linktest* diagnostic suggested that this model was misspecified. We therefore applied a square root transformation to this distance variable in our second model to correct for this misspecification. In our third model, we introduced two additional dichotomous covariates, namely the rural/urban classification of clusters and the survey methodology used to locate the GPS clusters (i.e. based on averaging the coordinates for each household in the cluster or based on a single GPS reading at an approximate cluster center). Goodness of fit statistics suggested that rurality improved the fit of the model and reduced spatial autocorrelation in residuals, but the survey methodology variable was not significant and was therefore dropped. To arrive at our final model, we introduced an interaction term between distance to governorate boundary and rurality to correct for a further misspecification caused by the introduction of this new variable.

Table 1 shows the resultant final model that best explained the observed errors. This model explained the presence of misallocated cluster points using two independent variables; square root of distance to the nearest governorate boundary and an interaction variable created by multiplying this distance variable by a dichotomous rural–urban variable (where each cluster point is coded one if it is urban and zero if not). As shown by the Z values, both these distance-related variables are significant and together explain around 40% of the variation in inconsistent governorate labels (pseudo $R^2 = 0.39$). The model confirms that the distance a GPS cluster point lies from a governorate boundary affects the probability of it having the correct governorate label and this effect varies between rural and urban areas. The Moran's I test statistic shows that there is very weak (albeit significant) spatial autocorrelation in the residuals of this model (Moran's $I = 0.01$, Z score = 2.41).

3.2. Kism/markaz level analysis

Since Measure DHS publish a list of sampled shyakhas (see Section 3.3 below), it would in theory have been possible for us to generate a list of sampled kisms and markazes by identifying those kisms/markazes that contain one or more sampled shyakhas. We would then be able to identify those GPS clusters lying outside

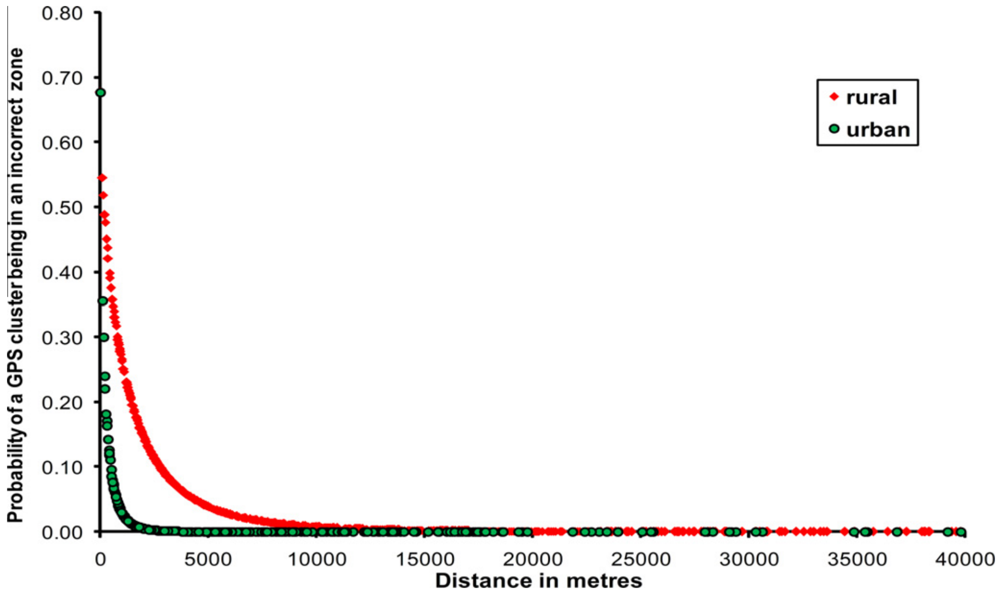


Fig. 4. The probability of a DHS GPS cluster being in an incorrect kism/markaz according to the regression model presented in Table 1.

the boundaries of the sampled shyakhas and so characterize the pattern of spatial error at this level. In practice, however, this is not straightforward: as described in Section 3.3 below, we had some difficulties in geocoding the names of sampled shyakhas. Furthermore, even allowing for these geocoding problems, we found that the majority of kisms and markazes had been sampled, with 259 out of 360 containing at least one sampled shyakha. This made the use of this methodology to characterize spatial error at the kism/markaz level problematic.

We therefore applied the logistic regression model derived at governorate level to kism/markaz level data by calculating the distance from each GPS cluster to the nearest kism/markaz boundary. Fig. 4 shows how the probability of a point lying in the incorrect kism/markaz differs between urban and rural cluster points. 7.9% of the GPS clusters showed a low probability value (less than 0.5) of being in their correct position. Monte Carlo simulation using these probabilities suggested that on average 285 GPS clusters (confidence limits 257–313) would fall within an incorrect kism or markaz.

3.3. Shyakha level analysis

Using the geocoding parameters described earlier, 58% of 647 sampled shyakhas were successfully matched automatically. The remaining DHS shyakha names were matched manually to equivalent census names. However, 13 out of 647 shyakhas remained unmatched, having no equivalently named units in the census map layer.

Fig. 5 illustrates the distances that GPS clusters lay from the nearest shyakha that was sampled as part of the Egyptian DHS. Relative to the governorate level analysis, the histogram shows that there is a larger number of GPS clusters which fall outside sampled shyakha boundaries (807 out of 1298). Many of these clusters are located 1–1.5 km away from the nearest sampled shyakha (almost 160 clusters). Similarly, roughly the same number of clusters fall within a distance equal to or less than 1 km. Most of these clusters are distributed along the Lower and Upper Egypt Governorates. The third category of misallocated clusters is located more than 1.5 and less than 2 km further way from the nearest shyakha boundaries. The remaining groups of GPS clusters are distributed over distances from more than 2 to less than 5 km. Unlike the governorate level analysis in Fig. 3, some GPS clusters lie more than 6 km from the nearest sampled shyakha, particularly in the desert governorates. GPS clusters more than 7 km from the nearest sampled shyakha, notably those in the New Valley desert governorate, were examined relative to online digital satellite imagery (ESRI ARCGIS online services "world imagery" resource). This reveals that some shyakha polygons in

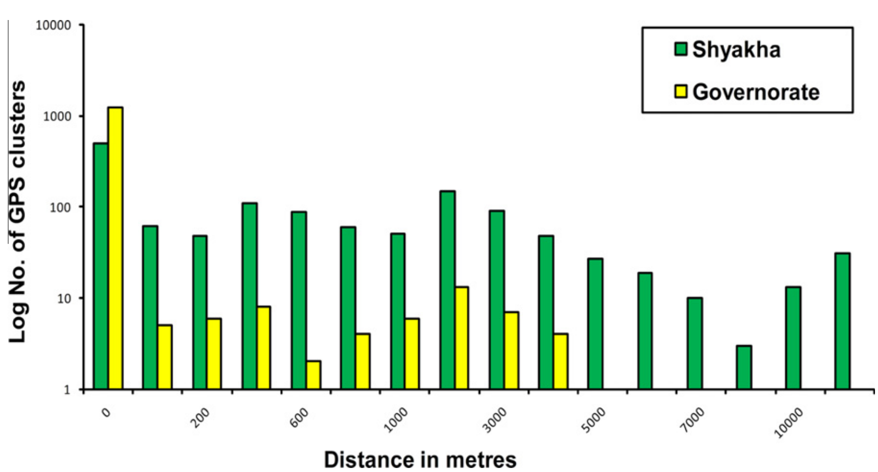


Fig. 5. Distances from GPS clusters to nearest Shyakha and governorate administrative boundaries.

Please cite this article in press as: Mansour, S., et al. Problems of spatial linkage of a geo-referenced Demographic and Health Survey (DHS) dataset to a population census: A case study of Egypt. *Computers, Environment and Urban Systems* (2011), doi:10.1016/j.compenvurbsys.2011.04.001

the census map layer are incorrectly located relative to observable human settlement in the image, while the GPS clusters seem to be in their correct locations.

Overall, there were 63 GPS cluster points (5% of those with coordinates) whose locations were inconsistent with governorate labels. Such GPS cluster points lay closer to governorate boundaries, an effect moderated by rurality. At the lowest census tier, the shyakha level, 807 GPS cluster points (62% of those with coordinates) lay outside the boundaries of shyakhas listed as sampled for the DHS. It is not possible to calculate an equivalent figure for the intermediate tier, the kism or markaz level. However, a simulation suggests that 285 GPS cluster points (22%, 95% confidence limits 20–24%) would lie in the incorrect kism or markaz. This simulation assumes that the spatial error at kism/markaz level follows that at governorate level.

4. Discussion

4.1. Understanding spatial uncertainty

These results suggest it may be problematic to link data on DHS GPS clusters to census small areas (such as shyakhas). Similarly, there is little to be gained from undertaking spatial linkage at the governorate level, since governorate labels are already provided for each GPS cluster for this purpose. Consequently, given the spatial uncertainties apparent in the data set, the evidence here suggests that the district (kism or markaz) level is the most appropriate for conducting spatial linkage. Examination of both the DHS GPS point and census map layers shows that they have the same projection and errors are widely distributed across the study area. Therefore, there is no evidence in our analysis to suggest a systematic error in either of the map layers analysed, such as a discrepancy in map datums used or some region-specific mislabeling of governorates. Since the majority of misallocated points are concentrated around governorate boundaries, it seems likely that the observed inconsistencies arise because of a positional error rather than mislabeling of GPS cluster points. In general, there seem to be two sources of positional error in the map layers analysed:

- There appears to be a randomly distributed spatial error in the GPS cluster points, which is greater in rural than urban areas (Table 1; Fig. 3). The difference in the distance decay curves for rural versus urban areas (Fig. 4) is likely to be caused by the deliberate scrambling of GPS cluster locations, described earlier for the 2005 Egyptian DHS as a form of controlled uncertainty (Fisher, 1999).
- We noted a systematic displacement of shyakha census boundaries in some isolated areas. This occurred only in some parts of the desert governorates, particularly southern New Valley governorate, but we were unable to identify any overall pattern to this error.

Several authors have noted how a change in the scale of geographic units can influence the strength of relationships in socioeconomic data and patterns of uncertainty (Cockings & Martin, 2005). This may raise concerns about our application of a logistic regression model derived using governorate-level data to data at the kism/markaz level. However, the main pattern in this governorate-level model is the rural–urban differences in the distance decay curves (Fig. 5), which as noted above appears to result from the deliberate scrambling of GPS cluster locations. This process will influence the observed pattern at all three levels (governorate, kism/markaz, and shyakha) and thus we would argue that the decision to apply the model at the kism/markaz level is justified.

In addition to the spatial error that was found at governorate level, there are some additional sources of error that will have influenced our shyakha level analysis:

- We were unable to geocode some shyakhas sampled through the Egyptian DHS, whilst other shyakhas may have been incorrectly geocoded because of local place name variations and issues in translating Arabic shyakha names into English. This may explain why some GPS cluster points are apparently located far from sampled shyakhas, as shown in the histogram in Fig. 4.
- At the shyakha level, a GPS cluster was only identified as erroneous and a misallocated point if it lay in an unsampled shyakha. However, this process does not identify all GPS clusters that lie outside their equivalent shyakha boundary, since some GPS clusters may be displaced to a neighboring *sampled* shyakha. This will result in an under-estimate of the extent of spatial linkage problems at the shyakha level.

Thus, care should be taken in interpreting the error rate in linking GPS cluster points to shyakhas relative to those for kism/markaz and governorate level.

As noted in the introduction, there is a growing literature that enriches DHS data by using GPS cluster locations to link to other spatial data sets, including census data. However, none of these studies has examined the effect of locational uncertainty on these linkage operations. We propose that the issues raised here are of broad significance when attempting GIS use for analytical data linkage in developing country settings, particularly in the increasingly common situation, where hand-held GPS are used to georeference social survey data which must then be linked with existing national sources. Furthermore, in using attribute labels inconsistent with location to model spatial error, we add to the existing literature on uncertainty and the point-in-polygon problem.

Particularly, where HIV data are collected as part of a DHS survey, positional uncertainty may mean that in multi-level modeling studies (e.g. Feldacker et al., 2010), community-level variables are derived from the incorrect higher level spatial units. Our findings suggest that this effect is likely to be more pronounced, where small areas (below district level) are used to derive community-level covariates. This may reduce the ability of such studies to detect contextual effects as opposed to individual-level effects. In small area estimation studies, this same effect is likely to occur and may increase the uncertainty over small area estimates of malnutrition prevalence, for example. Although this would be more difficult to model, positional uncertainty in GPS cluster points is also likely to affect spatial linkage to other types of data, such as land cover or transportation infra-structure, particularly, where these vary greatly within the space of several km.

Our study is subject to several weaknesses, notably that we have only a limited knowledge of the characteristics of the census boundary data that we used, owing to a lack of metadata. Similarly, we had limited success in geocoding shyakha names and this will have impacted on our analysis at this level. However, these problems are likely to occur in many other developing countries, which have limited resources for spatial data infra-structures and placenames that have been transcribed from non-European languages into English.

4.2. Future research implications

It would be possible to repeat the analysis of uncertainty in spatial linkage presented here using other, similar data sets. For example, GPS data were collected for other DHS surveys in Egypt, namely the 1992, 1995, 2000, and 2008 standard DHS surveys, as well as the 2003 interim DHS survey. These surveys could be analysed in the same way as we have done here. Alternatively, there are many other developing countries for which comparable standard DHS and census data are available, for which this methodology could be repeated. There are also other international surveys that record the locations of participating households using GPS. Such surveys include the AIDS Indicator Survey and Malaria Indicator Surveys, both run by the Measure DHS consortium, as well as some Living Standards Measurement Surveys and Multiple Indicator Cluster Surveys (Bocci & Petrucci, n.d.). Using similar methods, there may be potential to assess how far positional uncertainty influences spatial linkage for these other household surveys.

It may also be possible to modify the algorithm currently used to prevent inadvertent disclosure of the identities of households participating in the DHS, so that the algorithm does not generate inconsistencies between the location and governorate labels attached to each cluster. For example, the algorithm could be modified to prevent GPS cluster locations being moved across governorate boundaries as a result of the perturbation of their coordinates which would potentially enhance the utility of the data.

5. Conclusion

To our knowledge, ours is the first study to assess the impact of uncertainty on spatial linkage operations involving DHS GPS cluster points. This is despite there being an extensive literature that links DHS data to other spatial data sets such as land cover, rainfall, transportation infra-structure, and census data, as described earlier. The methodology described here provides a means of identifying GPS cluster points likely to be located within an incorrect census polygon, enabling the impact of linkage problems to be explored in subsequent analysis. In terms of the literature on uncertainty and its impact on point-in-polygon operations, the approach adopted here is practical and represents a new, alternative means of quantifying uncertainty in the specific situation, where there are inconsistencies between location and attribute label.

The addition of GPS clusters to the DHS offers increased analytical potential in GIS and global health research. However, our analysis suggests administrative labels and spatial locations of GPS clusters cannot be used uncritically and error may affect spatial analysis of DHS datasets. This is most likely due to deliberate modification of cluster locations noted earlier, intended to protect the confidentiality of participant households in the DHS. A strong recommendation to the MEASURE DHS organization would be to adapt the data modification algorithm to avoid gross misallocation between administrative areas. In combining DHS data from GPS clusters with census data for kizms/markazes or other administrative units, any linkage process will need to take account of the likely impact of positional uncertainty. At the shyakha scale there would be unacceptable levels of uncertainty in matching shyakha boundaries to GPS clusters. Thus, the kism/markaz census geography appears to be the best geographic level for undertaking spatial linkage. All researchers working with previously undocumented GPS-based datasets should consider investigation and modeling of spatial error before undertaking GIS linkage with administrative boundaries.

Acknowledgements

This research was supported financially by Department of Geography and GIS at Alexandria University, Egypt. We would also like to thank MEASURE DHS in the United States and the Central Agency for Mobilisation and Statistics (CAPMAS) in Egypt for provision of datasets. The authors alone are responsible for the design, conduct and presentation of this research.

References

Balk, D., Storeygard, A., Levy, M., Gaskell, J., Sharma, M., & Flor, R. (2005). Child hunger in the developing world: An analysis of environmental and social correlates. *Food Policy, 30*, 584–611.

Baschieri, A. (2007). Effects of modernisation on desired fertility in Egypt. *Population Space & Place, 13*, 353–376.

Bocci, C., & Petrucci, A. (n.d). Statistical methods for the analysis of spatial patterns: A geoadditive approach. <http://www.homes.stat.unipd.it/mgri/SIS2010/Program/3-SSIII_Giusti/901-1565-1-RV.pdf> Accessed 19.09.10.

Burgert, C. (2009). *The different "source" options in the DHS GPS dataset*, personal communication.

Cheung, C., Shi, W., & Zhou, X. (2004). A probability-based uncertainty model for point-in-polygon analysis in GIS. *GeoInformatica, 8*, 71–98.

Cockings, S., & Martin, D. (2005). Zone design for environment and health studies using pre-aggregated data. *Social Science & Medicine, 60*(12), 2729–2742.

Cuzick, J., & Edwards, R. (1990). Spatial clustering for inhomogeneous populations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 52*(1), 73–104.

El-Zanaty, F., & Way, A. (2006). *Egypt demographic and health survey, 2005*. Cairo: Ministry of Health and Population.

Fisher, P. (Ed.). (1999). *Models of uncertainty in spatial data* (2nd ed.. Chichester: John Wiley and Sons.

Feldacker, C., Emch, M., & Ennett, S. (2010). The who and where of HIV in rural Malawi: Exploring the effects of person and place on individual HIV status. *Health & Place, 16*, 996–1006.

Fujii, T. (2005). *Micro-level estimation of child malnutrition indicators and its application in Cambodia*. Washington DC: World Bank, Development Research Group, Poverty Team.

Gemperli, A., Vounatsou, P., Kleinschmidt, I., Bagayoko, M., Lengeler, C., & Smith, T. (2004). Spatial patterns of infant mortality in Mali: The effect of malaria endemicity. *American Journal of Epidemiology, 159*, 64–72.

Heuvelink, G., & Burrough, P. (2002). Developments in statistical approaches to spatial uncertainty and its propagation. *International Journal of Geographical Information Science, 16*, 111–113.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). Chichester: John Wiley and Sons.

Hunter, G. (1999). New tools for handling spatial data quality: Moving from academic concepts to practical reality. *URISA Journal, 11*, 2.

Johnson, F. A., Chandra, H., Brown, J. J., & Padmadas, S. S. (2010). District-level estimates of institutional births in ghana: Application of small area estimation technique using census and DHS data. *Journal of Official Statistics, 26*, 341–359.

Leung, Y., Ma, J., & Goodchild, M. (2004). A general framework for error analysis in measurementbased GIS Part 2: The algebra-based probability model for point-in-polygon analysis. *Journal of Geographical Systems, 6*, 355–379.

Leung, Y., & Yan, J. (1997). Point-in-polygon analysis under certainty and uncertainty. *GeoInformatica, 1*, 93–114.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (1999). Data transformation and linkage. In P. A. Longley, M. F. Goodchild, D. J. Maguire, & D. W. Rhind (Eds.). *Geographical Information Systems: Principles and Technical Issues* (vol. 1, pp. 477–479). Chichester: Wiley.

McGuire, M. (2000). Evaluation of food security in the Sahel: An analysis using the Demographic and Health Survey (DHS) data with a geographical information system. In D. Bigman & H. Fofack (Eds.), *Geographic targeting for poverty alleviation: Methodology and application*. Washington DC: World Bank.

MEASURE DHS (2010a). About DHS. <http://www.measuredhs.com/aboutdhs/whoweare.cfm> Accessed 19.09.10.

MEASURE DHS (2010b). About DHS, overview. <http://www.measuredhs.com/aboutdhs/start.cfm> Accessed 19.09.10.

Montana, L., & Spencer, J. (2004). *Incorporating geographic information into MEASURE surveys: A field guide to GPS data collection*. Chapel Hill, NC: Carolina Population Center [CPC], MEASURE.

Pande, S., Keyzer, M., Arouna, A., & Sonneveld, B. (2008). Addressing diarrhea prevalence in the West African Middle Belt: Social and geographic dimensions in a case study for Benin. *International Journal of Health Geographics, 7*, 17.

Ratcliffe, J. (2001). On the accuracy of TIGER-type geocoded address data in relation to cadastral and census areal units. *International Journal of Geographical Information Science, 15*, 473–485.

Simpson, L., & Yu, A. (2003). Public access to conversion of data between geographies, with multiple look up tables derived from a postal directory. *Computers, Environment and Urban Systems, 27*, 283–307.

Stephenson, R., Baschieri, A., Clements, S., Hennink, M., & Madise, N. (2007). Contextual influences on the use of health facilities for childbirth in Africa. *American Journal of Public Health, 96*, 84.

Uthman, O., & Kongnyuy, E. (2008). A multilevel analysis of effect of neighbourhood and individual wealth status on sexual behaviour among women: Evidence from Nigeria 2003 Demographic and Health Survey. *BMC International Health & Human Rights, 8*, 9.

Weeks, J. (2004). The role of spatial analysis in demographic research. In M. F. Goodchild & D. G. Janelle (Eds.), *Spatially integrated social science* (pp. 381–399). Oxford: Oxford University Press.

Wirth, M., Balk, D., Delamonica, E., Storeygard, A., Sacks, E., & Minujin, A. (2006). Setting the stage for equity-sensitive monitoring of the maternal and child health Millennium Development Goals. *Bulletin of the World Health Organization, 84*, 519–527.

Please cite this article in press as: Mansour, S., et al. Problems of spatial linkage of a geo-referenced Demographic and Health Survey (DHS) dataset to a population census: A case study of Egypt. *Computers, Environment and Urban Systems* (2011), doi:10.1016/j.compenvurbsys.2011.04.001

Dear Professor / Dr

As a key part of my PhD project, I have been developing a GIS-based measurement for assessment of current drinking water supplies and health in Egypt. Consequently, you are cordially invited to participate with other experts from Egypt to rate the components of my proposed Safe Drinking Water Index SDWI in terms of their importance for safe drinking water access in an Egyptian context. Your comments will not be attributed to you within the dissertation manuscript, but I would like to include an acknowledgement of your assistance unless you wish to remain completely anonymous in the final manuscript. I would be most grateful for your response by 1st September 2010.

The index is a new spatial methodology of measuring household access to improved drinking water sources at sub-provincial geographic level (Kism/Marqaz). The index should be applicable internationally particularly in the developing world and could be used globally as well.

According to data availability and an undertaken spatial linkage for two Egyptian datasets (Census and survey), several spatial, demographic and socioeconomic indicators were selected to compose the SDWI. These components are: population density, wealth factor, educational status, water supply types, water quantity, water interruption, home water treatment and sanitation type. It is proposed that the SDWI will be a weighted combination of these indicators.

The following is a brief explanation for each indicator:-

Population density

The number of people per km^2 in each Kism or Marqaz based on 2006 census data. It is expected that households live in overcrowded, density and unplanned settlements suffer from poor and unreliable drinking water supplies.

Wealth factor

Wealth indicator measures cumulative living standard of households in each Kism or Marqaz based on data from the survey. Wealth factor is calculated according to collected data about household ownerships such car, television, housing furniture and other household materials. The higher score is the richest while the lowest is poorer.

Educational status

This indicator consists of percentage of Egyptian population who are able to read and write in each Kism/Marqaz based on 2005 Egyptian census. It is assumed that high educated households are more likely to use filtration methods and consume potable water.

Supply types

This indicator consists of public and private supplies and therefore improved and unimproved drinking water sources using data from the survey and census. For instance, public water tap, pump, well, standpipe, tap inside or outside the building.

Water quantity

The available drinking water quantity in litres per capita per day for each household based on supply location using survey data. It is assumed that drinking water supply inside household dwelling is associated with high water quantity compared with long trip to reach water source.

Water interruption

This indicator was composed using survey data and consists of the number of water interruption incidences and water frequency in the last two weeks.

Home water treatment

Home water treatment indicator was composed of percentage of households who treated their drinking water in each Kism/Marqaz using survey data. Home water treatment is a process which aims to reduce microbial contaminations and other risks associated with water borne diseases.

Sanitation type

Sanitation type is percentage of Egyptian households in each Kism or Marqaz according to their access to each type of sanitation systems using census data.

Indicators of the SDWI	Weighting score (0 -10)
Population density	
Wealth factor	
Educational status	
Water supply types	
Water quantity	
Water interruption	
Home water treatment	
Sanitation type	

- 0 value means “least important” for safe drinking water access
- 10 value means “most important” for safe drinking water access

Please give a score for each of the following indicators as a weighting value (score 0 to10) and according to the importance, influence and relevant to safe drinking water accessibility:-

I believe that your view is crucially important so if you would like to comment on any of these indicators or the overall index components. In particular I would like your thought on the following:-

Do you think there are other indicators should be included in the index?

Do you think any of these indicators should not be included in the index?

Do you have any potential or general comment on the creation of this index?.

Please send your respond back to me by email (preferred) or by post:

Email (preferred method): saem1c08@soton.ac.uk

Postal address:

Shawky Mansour

School of geography

University of Southampton

Southampton, Highfield campus

SO 17 1JB England, United Kingdom

Acknowledgement

I understand that my respond will not be individually identifiable.

☐ I would be grateful to be acknowledged in the final thesis.

☐ I would prefer my name not to appear at all.

Thank you very much in advance for your support and look forward to hearing from you soon.

Sincerely

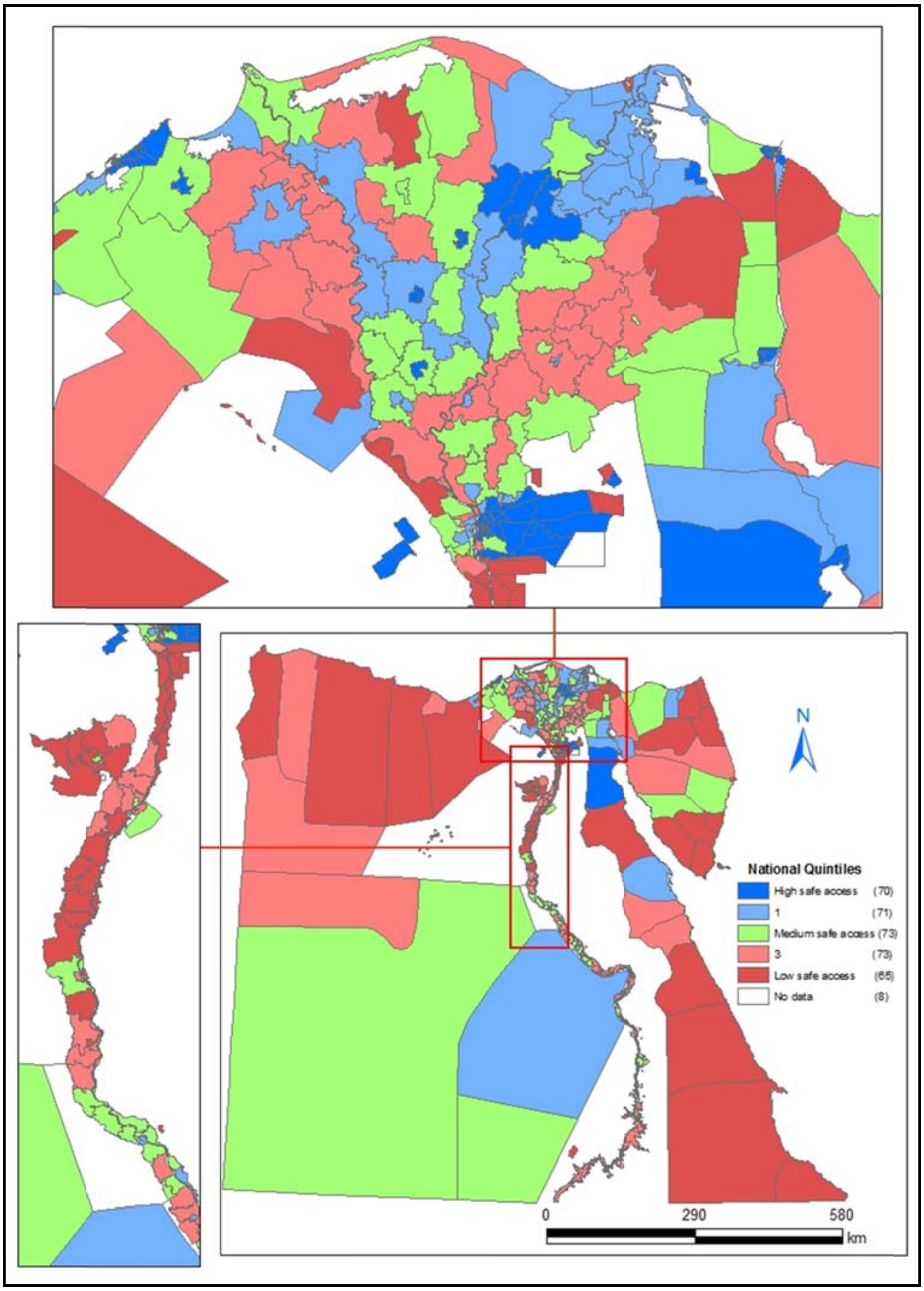
Shawky Mansour

```
use " X:\thesis_work\interpolation\Stata_StSpCo\JoinedFile",
*this command merge two data files based on one-to-many
uniqusing

* these lines recode and calculate the water interruption scores
recode  sh109 (1=1) (2=2) (3=7) (8=.) , gen(index_interrupt)
replace index_interrupt = 25 if sh108 == 0

* these lines declare the survey design and send output to a file
encode sec_ename, generate(kism)
svyset hv021 [pweight=hv005], strata(hv023) vce(linearized) singleunit(missing)
log using "c:\temp\shawky_interrup.txt", text

* this line calculates mean interruption scores by kism/markaz
svy linearized : mean index_interrupt, over(kism)
log close

* these lines recode and calculate mean water quantity scores by kism/markaz
recode  hv204  (1/5=30) (6/30=12) (31/121=5) (996=50) (.=50) (998=.) ,
gen(index_watquant)
gen index_watquantln = ln(index_watquant)
log using "c:\temp\shawky_quantity.txt", text
svy linearized : mean index_watquantln, over(kism)
log close

* this line calculates mean for household wealth index factor by kism/markaz
log using "c:\temp\shawky_wealth.txt", text
svy linearized : mean hv271, over(kism)
log close

* these lines recode and calculate mean for water supply types scores
recode hv201 (11/13=960) (20/33=900) (40/41=0) (61/62=860) (71=975) (96=.),
gen(sourcescore)
log using "c:\temp\shawky_sourcescores.txt", text
svy linearized : mean sourcescore, over(kism)
log close
* these lines recode and calculate mean for water treatment scores
```

```
recode sh111a (1=200) (0=100), gen(treatment_score)
replace treatment_score=141 if sh111b==1
replace treatment_score=164 if sh111d==1
replace treatment_score=164 if sh111e==1
replace treatment_score=120 if sh111f==1
log using "c:\temp\shawky_treatscores.txt", text
svy linearized : mean treatment_score, over(kism)
log close
```

Dear Prof/Dr

It is my pleasure to contact you again. To complete the expert evaluation task of my project, you are cordially invited to comment on the geographic patterns of the calculated index values. The following map illustrates the spatial distribution of the SDWI index values across Egyptian districts (kism/markaz)

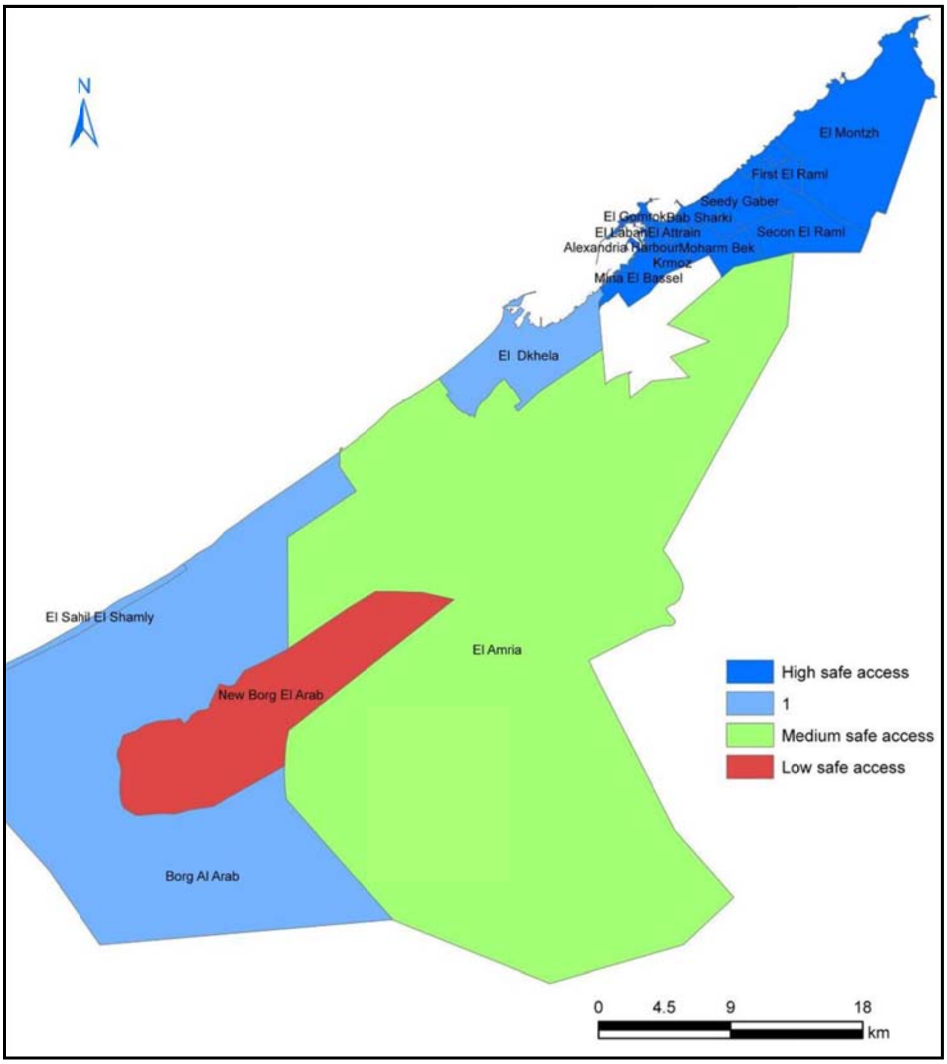


According to the map, please give a comment on the geographic patterns of the index values for each Egyptian region:

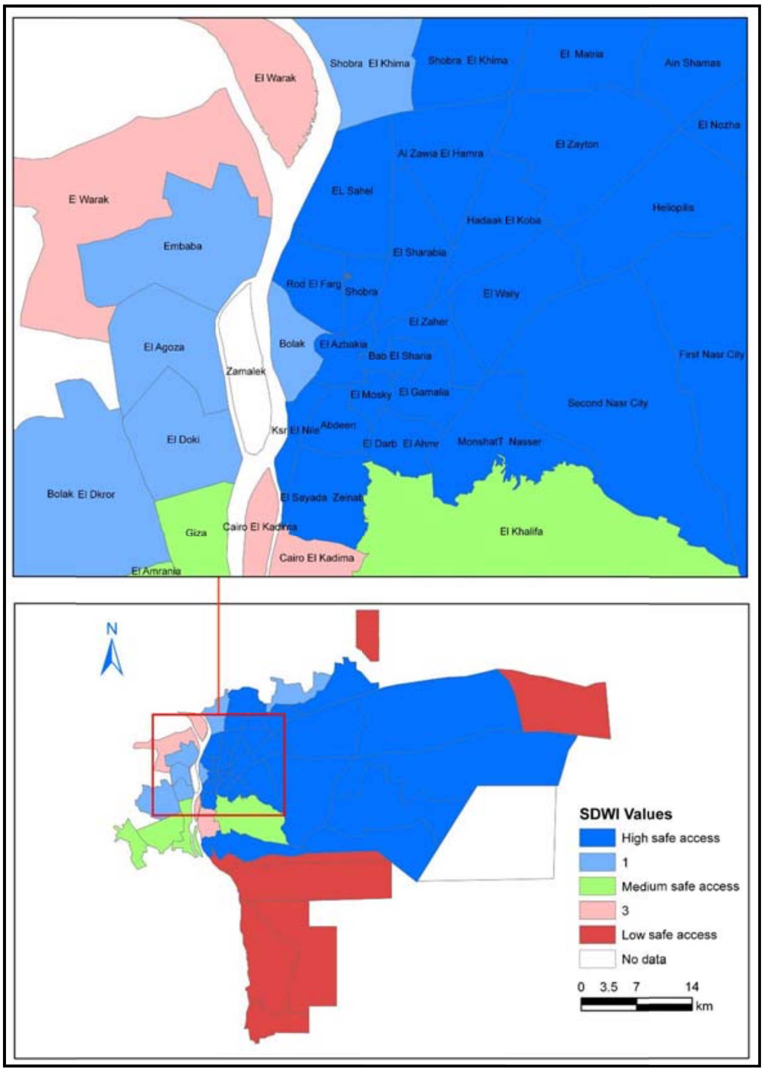
The spatial pattern of the index values in Delta governorates:

The spatial pattern of the index values in Valley governorates:

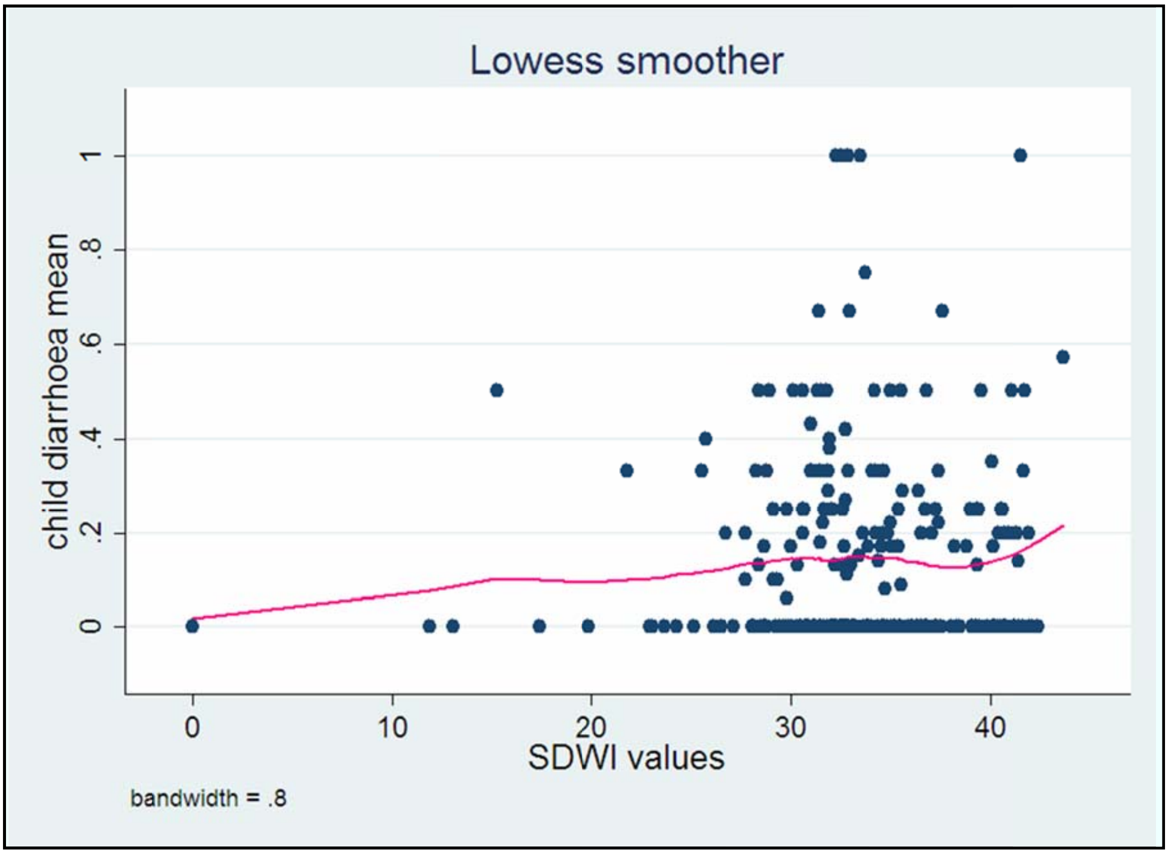
The spatial pattern of the index values in Frontier governorates:



The spatial pattern of the index values for Alexandria districts:



The spatial pattern of the index values for Greater Cairo districts:



This figure demonstrates the correlation between child diarrhoeal disease and the SDWI values. It seems that no significant correlation was found between the two variables.

Please give any comment on this figure:

Do you have any potential or general comment on the output of this index?

Please send your respond back to me by email (preferred) or by post:

Email (preferred method): saem1c08@soton.ac.uk

Postal address:

Shawky Mansour

School of geography

University of Southampton

Southampton, Highfield campus

SO 17 1JB England, United Kingdom

Thank you very much in advance for your support and look forward to hearing from you soon.

Sincerely

Shawky Mansour