UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES

Mathematical Sciences

Non-Gaussian Spatial Modeling in Index Flood Estimation

by

Marinah Muhammad

Thesis submitted for the degree of Doctor of Philosophy

October 2018

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES
Mathematical Sciences

Doctor of Philosophy

NON-GAUSSIAN SPATIAL MODELING IN INDEX FLOOD ESTIMATION

by Marinah Muhammad

Index flood estimation is important in regionalization procedure to solve an issue of ungauged catchment that has received great attention among hydrologists. The UK index flood estimation model known as the FEH-QMED model is a well established one with nonlinear effect of explanatory variables identified. However the shortcomings of current research in literature such as not taking into account a spatial dependency and non-Gausianity that exist in the flooding data motivated us to investigate further. This thesis aims to improve the existing methodology in index flood estimation model, where FEH-QMED model is chosen as the benchmark. Three objectives that have been developed in this thesis are: (i) to explore the shortcomings of the current research with a possibly improved model in estimating the UK index flood, (ii) to develop a more efficient statistical model in estimating the index flood that better fits the UK flooding data, and (iii) to discover more relevant predictive catchment characteristics that may improve the index flood estimation model. To answer the objectives, statistical methods have been proposed and applied into the UK flooding data analysis to establish new index flood regression models detailedly discussed in chapters of the thesis respectively. In Chapter 2 we apply the spatial additive and spatial error analyses into the UK flooding data to explore possibly improved models for the UK index flood estimation. Chapter 3 proposes a new spatial error model with skewed normal distribution for residuals and develops a maximum likelihood computational algorithm to apply into the UK flooding data for the purpose of establishing a new index flood estimation model. Chapter 4 is focused on model selection of index flood estimation model by proposing a panelized likelihood estimation method that utilizes adaptive Lasso as a regularization tool in variable selection of all available catchment characteristics in the UK flooding data source. We also present the simulations to investigate the finite sample performance of the proposed statistical methods. In comparison study, AIC scores have been used as model selection criteria, while to measure the performance of different models, the percentage improvement in mean square prediction error relative to the updated FEH-QMED model in Kjeldsen and Jones (2010) is applied. The obtained results demonstrate that the skewed spatial error flood model that is established by using statistical method suggested in Chapter 4 outperforms the others and can significantly improve the FEH-QMED model in estimating the UK index flood.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Marinah Muhammad , declare that the thesis entitled *Non-Gaussian Spatial Modeling in Index Flood Estimation* 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;

Signed:...

Date:...

Acknowledgements

First and foremost, I would like to express my deep gratitude and respect to my main supervisor, Professor Zudi Lu. He taught me the delight of studying statistics and motivated me to investigate the area of spatial additive and autoregressive model and apply it in the field of flood risk study. He inspired me to be an independent researcher and encouraged me to do an interdisciplinary research which is more practical for my career and research path in the future. He was always there to give me sound advice and generous support whenever I had a hard time either research and financial. I am appreciate all his contributions of ideas, guidance, encouragement and financial support to make my Ph.D experience productive and stimulating.

My sincere thanks must also go to other members of my committee: Professor Sujit Sahu (Second Supervisor), Professor Wei Liu and Dr. Maria Kyriacou (Independent assessors for yearly study reviews). They generously gave their time to offer me valuable suggestion towards improving my work.

I gratefully acknowledge the National River Flow Archive team at the Centre for Ecology and Hydrology that helped on data retrieval to make it relevant for my research project. Not forget to Dr Thomas Kjeldsen from Bath University that previously was a senior hydrologist at Centre for Ecology and Hydrology for his constructive comments on my first manuscript of this research work.

I want to extend my appreciation to the School of Mathematical Sciences, where I have been provided with qualified education, excellent teaching and research environment, and generous administrative assistantship from post graduate office during the four years at University of Southampton.

I would also like to acknowledge the Ministry of Higher Education of Malaysia and Universiti Malaysia Kelantan for the scholarship awarded and financial support to pursue my studies and fulfill my dream.

Last but not least, I thank with love to my beloved family. I am very grateful to my parents and parents in law for their faith on me and allowing me to follow my ambition. Very special thanks to my husband, Mohd Borhan Mat Ail for being my best friend, my lover and my soulmate. I could never have accomplished this dissertation without his constant understanding, patience and support. Finally to my lovely children Dhiya, Azim and my new baby that expected in October 2018. They make my life so much happier and more meaningful than I could imagine. All praise to Allah for giving me this opportunity to experience this very humble journey.

Chapter 1

Introduction

Flooding is one of the hydrological disasters that may occur in all parts of the world and give devastating impacts on human life and properties. In order to provide a good flood defense scheme, well-organized spatial planning and quality hydraulic infrastructures are essential in all countries around the world. To make that be provided, flood estimation needs to be made beforehand. However, due to the complexity of the hydrological system, flooding process is not fully understood and therefore reliable statistical models that can represent flood estimation need to be continuously developed. According to Grimaldi et al. (2011), many statistical methods have been used in the literature and practice of hydrological applications with different aims; among them the most important ones are simulation, forecasting, uncertainty analysis, spatial interpolation and risk analysis. This research attempts to explore and improve the statistical methods in flood estimation at ungauged catchment area, which can be grouped under the aim of risk analysis. Flood estimation is an important task for hydrologists in risk analysis because it provides the knowledge of the frequency and magnitude of flood peak discharges at a particular area. Human property, health and lives will be at risk if floods are not properly estimated in designing infrastructures and planning land use.

Based on the literature and practical reports of the hydrological applications, the methods for flood estimation in risk analysis can be grouped into two main types, which are data-based methods and physically-based methods (c.f, Sene, 2008; Solomatine and Wagener, 2011; Samsudin et al., 2011). The complexity of the processes in flooding has made many hydrologists and researchers focusing on flood estimation using the data-based methods. This type of methods only considers the past recorded and related flood data and has received a rapid development with minimum information requirement (Adamowski and Sun, 2010; Kişi, 2004; Wang et al., 2009). Data-based methods focus on extracting and re-using the information that is implicitly contained in the hydrological data, without having to directly take into consideration the physical law of the flooding processes. Although the data-based methods may lack the ability to provide physical interpretation and insight into other inter-related factors in flood process, it is able to provide relatively accurate flood prediction from the particular characteristics of flood data under study. The main method used for flood estimation in risk analysis among the data-based methods is Flood frequency analysis (FFA), while, for the physically-based methods, it is rainfall-runoff modeling (RRM).

Flood frequency analysis (FFA) is a statistical approach that is used to predict the possible flood magnitude over a certain period of time, known as return period. This analysis can also estimate the return period with the emergence of flood magnitude that may occur. The FFA has been used to estimate the flood at sites of interest located at gauged catchment area. Gauged catchment area is referred to the sites that have enough data for the purpose of flood estimation. This analysis will be explained in detail in Section 1.1. Nevertheless, many parts of the world are ungauged or poorly gauged because monitoring stations are only located at specific, strategic and important locations (Mamun et al., 2011). Therefore, many sites of interest for flood estimation have no or short-recorded data. The issue of ungauged catchments is still an open problem among hydrologists and it has received great attention. In response to this issue, statistical approach is particularly important for developing hydrological analyses without observations of flood information at the ungauged area (Grimaldi et al., 2011). When development projects are located at an ungauged catchment area, hydrologists need to estimate the design flood using available information from the nearby gauged sites. This approach is known as regionalization procedure. Regionalization procedure is the first principle proposed by NRC-US (1988) in hydro-meteorological modeling with the basic understanding of "substitute time for space". It is very important and mainly used in solving the issue of ungauged catchments where hydrological information collected from different locations is used to compensate for the limited or absent information at the site of interest. There are many types of regionalization procedure available (Cunnane, 1988, 1989), among which the most widely used is index flood method. Index flood method was introduced by Dalrymple (1960), it is based on the assumption that flood magnitudes at all sites in a region follow the same frequency distribution except for a scaling factor, "index flood" (c.f, Hamed and Rao, 1999; Grimaldi et al., 2011).

Flood frequency analysis (FFA) with regionalization procedure is known as regional flood frequency analysis (RFFA) which will be introduced in detail in Section 1.2. RFFA is used by hydrologists in estimating flood at ungauged catchments area. Regional flood frequency analysis (RFFA) with index flood method follows the same steps as FFA except that it is more complicated as it requires the estimation of index flood and analysis of construction of the regional growth curves. The flood at any locations of ungauged catchment area could be estimated by multiplying estimated value of the index flood and estimated value of the growth factor from constructed growth curve. The accuracy of flood estimation at ungauged catchment area is dependent on the accurate estimated value of the index flood and growth factor. Therefore, the improvement of existing methodology of both criteria are needed for better accuracy of flood estimation at any location of ungauged catchment area in the future.

This research project aims to improve the existing method of estimating the index flood because it contributes a large amount of uncertainty to the regionalization procedure for flood estimation at ungauged catchment area (Kjeldsen and Jones, 2007). The literature contains numerous studies to improve the existing method in estimating the growth factor compared to the index flood. However, countless practical applications (see e.g., IH, 1999; Pilgrim and Institution of Engineers, 2001; Wilson et al., 2011) highlighted the importance of having an accurate estimation of the index flood in regionalization procedure. It is common practice that multiple regression models are used to indirectly estimate the index flood at ungauged basin by using catchments characteristics (c.f., Grimaldi et al., 2011; Castellarin et al., 2007; Brath et al., 2001). This is because, it is generally more accurate than conceptual indirect models for estimating index flood at ungauged catchment area (Brath et al., 2001). Many index flood estimation models have been developed by using multiple regression models (NERC, 1975; Canuti and Moisello, 1982; Acreman, 1985; Mimikou and ordios, 1989; Garde and Kothyari, 1990; Reimers, 1990). Index flood estimation model from Flood Estimation Handbook (FEH) is one of them (IH, 1999). It is well-established in the UK and therefore it has been chosen as the benchmark model for this study. The aforementioned issues have motivated us to further investigate and improve the existing methodologies in the index flood estimation models from a statistical point of view.

1.1 Flood Frequency Analysis

This study attempts to respond to the issue of uncertainty in index flood estimation models, therefore a general understanding of a methodology in flood frequency analysis (FFA) is needed beforehand. This study will first define the meaning of a return period and how it is connected to the probability of occurrence of flood events. The return period is denoted as T and also known as the recurrence interval that can be defined as average interval between occurrence of floods that exceed a particular magnitude. It is measure of rarity of the flood because naturally large floods have large return periods and vice versa. It is common to see terms like “100-year flood” or the “500-year flood” being used in the news media and within professional organizations. A “100-year flood” is defined as a flood that can occur on average of every 100 years, therefore the return period is defined as $T = 100$. However the time interval between flood varies. So “100-year flood” do not occur at nicely spaced interval of 100 years. A “100-year flood” can also be defined as a flood that has a 0.01 chance of occurring in any given year. This is called as an exceedance probability or probability of occurrence which can represented as $\frac{1}{T}$.

A relationship between the probability of occurrence of a flood and its return period can be justified through the fundamental concept in statistical frequency analysis. Flood observations are made at regular interval at some sites of interest. Let a random variable Q, defined as the magnitude of the flood event that occurs at the given time and at the given site, where $0 < Q < \infty$. While the observed data space,q provide a sample of realization of Q and it is assumed that each q is statistically independent of each other. From frequency analysis, the question of how frequently the possible value of Q occur could be specified. From the definition of return period, a given flood q with a return period T may be exceeded once in T years. Hence, the exceedance probability is $P(Q_T > q) = \frac{1}{T}$. The cumulative probability of non-exceedance $F(Q_T)$ is given by

$$F(Q_T) = P(Q_T \leq q) = 1 - P(Q_T \geq q) = 1 - \frac{1}{T}. \tag{1.1}$$

Equation 1.1 is the basis principle of FFA for estimating the magnitude of a flood at any return period, Q_T. Substituting $F(Q_T) = 1 - \frac{1}{T}$ in a known statistical distribution function one can solve for the magnitude of Q_T.

The primary objective of FFA is to relate the magnitude of extreme flood events to their frequency of occurrence through the selected statistical distributions (Chow et al., 1988) with the assumption that all events in the observed flood series represent a process that can be described by one single statistical distribution. Due to that, it involves fitting a theoretical statistical distribution to the observed flood data using selected parameter estimation method. Many different statistical distributions are available and commonly applied for FFA but no underlying justification for the most suitable one (Coles et al., 2001). Nevertheless, the literature recommended to used L-Moment ratio diagram to identify the best fitting distribution and to test the acceptability (e.g., Hosking and Wallis, 1997; Vogel and Fennessey, 1993; Peel et al., 2001). Hence, the most suitable fitted distribution can be expressed in mathematical equation form in terms of the return period, T to represent as a flood frequency curve for the site of the used flood data. It is always helpful to plot flood frequency curves in the flood frequency diagram with flood magnitudes on the y-axis and return period on the x-axis. Observed flood data can be added to the flood frequency diagram to visually investigate the extreme flood value. The reasons for using the flood frequency diagram are to provide a simple way in relating flood magnitude and return period, and to compare possible frequency curves with observed flood behavior.

FFA concentrates on single-site frequency analysis that has sufficiently long flood data observed over an extended period of time in a river system. The data are assumed to be independent and identically distributed, stationary and may even be assumed to be space and time independent. Further, it is assumed that the floods have not been affected by natural or environmental changes such as climate change, land use changes, urbanization and extensive tree felling. How long of this observed flood data to be sufficient for FFA differs in each practical application. For example, the Norwegian guidelines for flood estimation recommended practicing FFA for stations that have more than 30 years recorded flood data by using 2 parameters distribution if the records are up to 50 years long and using 3 parameters distribution for stations that have more than 50 years recorded data (c.f., Wilson et al., 2011). In the UK through Flood Estimation Handbook (IH, 1999), they are practicing FFA for the stations that have recorded flood data 2 times of target return period. For example if they want to estimate 50-year flood at a particular site,100 years of recorded flood data is needed.

1.2 Flood regionalization approach

Regionalization procedure is required if the flood records are insufficient. Short and no recorded flood data at the site of interest has always been a challenge to hydrologists in estimating the floods before any project could be designed or developed. Applying FFA without regionalization procedure at a site with insufficient length of flood data records introduces significant uncertainty into the flood estimates. With regionalization procedure, hydrologists could use available data and information from nearby gauged stations to reduce the uncertainty in flood estimation at the ungauged sites of interest. One could say space replace time to increase the sample size and reduce sample uncertainty. However, research has shown that more accurate values of flood estimates are obtained from FFA with regionalization procedure than without it (Lettenmaier et al., 1986). Many types of regionalization procedures are available (Cunnane, 1988, 1989). A traditional and simple approach which has been used for a long time is index-flood method introduced by Dalrymple (1960). However, this method is still currently relevant in flood risk studies since it outperforms others regionalization approaches in some recent literature (see e.g., Malekinezhad et al. 2011)

The flood regionalization by using index flood procedure is known as classical regionalization approach that contain two main steps which are:- i) estimation of the index flood, μ_Q, ii) estimation of regional growth factor, q_T. Some issues in the second main step of the index flood method that involved several important tasks including identification of homogeneous region of the sites of interest are considered well studied (Grimaldi et al., 2011). Therefore, this research is focused on improving the existing methodologies in the first main step of the index flood procedure that relatively few found in the literature. A method used in the first main step of index flood procedure which is estimating the index flood, μ_Q, at ungauged catchment area will be introduced in Section 1.3. However, the understanding about index flood procedure is needed beforehand. Therefore, we first introduce the detail definition and methodology on index flood procedure in the following subsection.

1.2.1 Index flood procedure

The key assumption in index-flood procedure is that the distributions of flood at the different sites in a region is the same except for a scaling factor, the “index flood” (Hamed and Rao, 1999; Grimaldi et al., 2011). The index flood can be thought as typical flood for a particular catchment. It tends to reflect with local hydrological conditions such as catchment size, annual rainfall, the catchment’s soil types and many more. The index value of annual maximum series of flood peak varies according to different methodologies. As an example, Dalrymple (1960) and Hosking and Wallis (1993) used the observed sample mean as estimator of the index flood, Robson and Reed (1999) suggested the observed median which is the middle ranking value in the series of annual maximum flood, while Sveinsson et al. (2001) proposed the location parameter determined on a site to site basis as index flood. Standard procedure for flood frequency estimation in the

UK is provided by the Flood Estimation Handbook and according to Robson and Reed (1999), the median is a more robust measure, less affected by the size of an exceptionally large flood event, whereas the mean can change markedly. In this study the index flood is defined to be median annual maximum flood, (QMED). Since QMED is a median value, the index flood of this study is the two-year return period flood on annual maximum scale. This is because on average half of all observed annual flood data at the site of interest are greater then QMED. This mean exceedance probability, $\frac{1}{T}$ is half at QMED, which implies the return period, $T = 2$.

FFA represents the simplest flood estimation case of single gauged site through flood frequency curve diagram. Flood frequency curve diagram shows the relationship between flood magnitude and flood frequency together with its return period. Flood magnitude with return period of T years, Q_T can be estimated directly from flood frequency curve diagram. However, through index flood procedure, Q_T can be estimated by

$$Q_T = \mu_Q q_T, \tag{1.2}$$

where μ_Q is the index flood at the given site and q_T is a dimensionless growth factor for the given return period. The dimensionless growth factor, q_T, is the value that could be extracted from the regional growth curve diagram at the particular return period. The regional growth curve diagram which also describes the relationship between flood magnitude and return period can be thought of as a scaled version of the flood frequency curves. It can be similarly constructed as the flood frequency curves diagram, but it has to be scaled to have the value of 1 at the index flood. This is applied in this case of study at two years return period for the purpose of comparison between other different catchments. Therefore, by using the procedure in index flood method, FFA can also be applied in the case of ungauged sites and known as regional flood frequency analysis, (RFFA). In the UK, the index flood procedure has been practically used to estimate the flood at ungauged catchment area since 1965. It then, has been well established through Flood Estimation Handbook, (FEH) published by IH (1999) with latest improved done by Kjeldsen et al. (2008). A detail methodology with application of the index flood procedure can be referred in these two scientific reports. The index flood procedure has also currently been used in research studies all around the world and shown the good performance of this method compared to other regionalization approaches. One can be found in Malekinezhad et al. (2011) that comparing index flood and multiple regression methods based on L-moment which are applied to several sites in the Namak-Lake basin in central Iran.

The method of constructing the regional growth curve, in RFFA is similar to construct the flood frequency curve, in FFA. Nevertheless, constructing the regional growth curve, is more complicated than constructing the flood frequency curve because the homogeneous region must be first identified before other similar steps could be applied. The literature recommended using the L-moment method introduced by Hosking and Wallis (1997) for identifying homogeneous region and estimating the regional frequency distribution (see e.g., Malekinezhad et al. 2011; Kjeldsen et al. 2008; Peel et al. 2001). The identification of the homogeneous regions is essential in RFFA because the flow data information from the identified regions will be used in the data transformation method since the site of interest has no data to construct the flood frequency curve. Each identified regions will use the data from several gauged sites to represent the flow information at the interest ungauged site. As a result, each region that represent the interest ungauged site has different flood frequency curves with respective index flood values (i.e., mean or median of flood data). Therefore, those flood frequency curves have to be changed to the flood growth curves by dividing them with the index flood values respectively. Working with flood growth curves allows data from different sites with differing index flood values to be combined hence the best regional growth curves can be developed. By using the best regional growth curves, the dimensionless growth factor, q_T at return period T for the interest ungauged site can be determined. Nevertheless, this research not interested in improving the existing methodology of the second step of flood index procedure which is estimation the dimensionless growth factor, q_T but we are interested in improving the existing methodology of the first main step of index flood procedure which is estimation the index flood value,μ_Q. The good estimate of index flood, $\hat{\mu_Q}$ is essential because it used to multiply with the estimated growth factor, $\hat{q_T}$ to estimate the magnitude of the flood at the return period T for the interest ungauged site, $\hat{Q_T}$.

The method of estimating the index flood, μ_Q also differs in FFA and RFFA. Let consider the case of this study in which index flood, μ_Q is assumed to be the median of the flood distribution. The estimation of the index flood in FFA is easy because the site of interest is gauged and the length of the recorded flood data is sufficiently long. In this case μ_Q can be obtained directly by calculating the arithmetic median of the available observations. Indirect methods have to be used in ungauged sites, instead. Thus, for RFFA index flood is estimated using statistical equation that underlying the relationship between the index flood with catchment characteristics. Most commonly used indirect method is multiple regression model. Therefore many index flood regression models have been developed by active researchers in this field to be used in estimating the index flood at ungauged site. Most of the developed models are standard multiple regression model which are not take into account any unknown and unexpected features in flooding data used (see e.g., Malekinezhad et al. 2011; Wan Jaafar et al. 2011). The aforementioned shortcomings of current research in literature have motivated us to investigate further the existing index flood regression models to ensure a good statistical method could be proposed that probably improve the estimating the index flood, μ_Q at ungauged catchment area by using RFFA.

RFFA based on the index flood method stated in Equation (1.2) clearly shows that flood at ungauged catchment area could be estimated precisely if we had an accurate reliable method for estimating index flood and constructing the regional growth curves that can give good estimates of growth factor. The literature contains numerous studies and research efforts toward improving the existing methodologies in constructing regional growth curves and relatively few on estimation of index flood (Brath et al., 2001; Bocchiola et al., 2003). Nevertheless, the countless practical applications (see e.g., IH, 1999; Pilgrim and Institution of Engineers, 2001; Wilson et al., 2011) have highlighted the importance of index flood estimation in RFFA. These applications discussed the difficulty in obtaining reliable models to estimate the index flood which requires the merging of concepts in statistical and physical hydrology to reduce the presence of uncertainties. The high variability of index flood in different location of river basins and throughout the river network of a given catchment reflects hydrological diversity encompassing river geomorphology, basin land use, geology, lithology and micro-climate. This can lead to non-trivial difficulties in the index flood estimation exercise and therefore large uncertainties may be introduced in the estimated T-year flood at interest site (Bocchiola et al., 2003). Wan Jaafar et al. (2011) listed several potential difficulties in index flood estimation and modeling which the difficulties are in getting the required data, selecting the covariates effectively and efficiently, finding a suitable model structure and estimating the model parameters.

Index flood estimation contributes to the large amount of uncertainties in RFFA and it has become an issue among hydrologists (see e.g., Kjeldsen and Jones, 2007; Jaafar et al., 2012). Due to this, it still requires great deal of effort among researchers to improve the existing methodologies in estimating the index flood. Some studies done to improve the existing methods of estimating index flood were conducted by Shu and Burn (2004) and Kjeldsen and Jones (2007) and the latest is by Wan Jaafar et al. (2011). The definition of guidelines for the identification of the most suitable method in estimating index flood can be found in Bocchiola et al. (2003). A comprehensive review about issues and problems in RFFA has been documented by Grimaldi et al. (2011) that stated uncertainty in index flood estimation is one of them .

Based on the aforementioned literature review, it clearly shows the urgent need for research to improve the accuracy in index flood estimation since it contributes to a large amount of uncertainty in RFFA for flood estimation at ungauged catchments area. Furthermore some of the issues in constructing regional growth curves in RFFA have been considered to be well studied so the margin of improvement in the accuracy of regional estimate associated with the issues are probably rather limited. Therefore this research is determined to focus on improving the existing methodologies in estimating the index flood, μ_Q, by using median as the key statistics of response variable, catchment characteristics as the input variables and FEH index flood model as the benchmark.

1.3 Estimating the index flood

It is common practice for regression models to be used to forge a link between a certain hydrological parameters and a set of catchment characteristics. It is based on the assumption that the flood characteristics can be explained by catchment characteristics (Mazvimavi et al., 2005). The estimation of index flood, μ_Q, is straightforward when the site of interest is gauged and the record length of available data is sufficiently long. In this case, μ_Q can be obtained directly by calculating the arithmetic mean or median of the available observations. In other circumstances, indirect methods need to be used at ungauged sites. The most used indirect methods are multi-regression models, which link to an appropriate set of morphological and climatic descriptors of the basins through statistical relations using power form equation with multiplicative or additive of error term (Brath et al., 2001; Grimaldi et al., 2011; Castellarin et al., 2007):

$$\mu_Q = a x_1^{\beta_1} x_2^{\beta_2} \ldots x_p^{\beta_p}, \tag{1.3}$$

where x_i is the i^{th} catchment descriptor, for $i = 1, \cdots, p$, with p the number of the catchment descriptors taken as the explanatory variables for the model, and a and β_i are the model parameters. Brath et al. (2001) and Bocchiola et al. (2003) compared the regression approach to conceptual indirect methods such as rational model and geomorphoclimatic model. Based on an assessment of the prediction errors, Brath et al. (2001) indicated that statistical indirect models such as regression models are generally more accurate than conceptual indirect models for predicting index flood at ungauged catchments.

By this approach, many index flood estimation models have been developed, (c.f., NERC, 1975; Canuti and Moisello, 1982; Acreman, 1985; Mimikou and ordios, 1989; Garde and Kothyari, 1990; Reimers, 1990), including efforts for improved models (Shu and Burn, 2004; Kjeldsen and Jones, 2007; Wan Jaafar et al., 2011). See also guidelines for choice of suitable methods in (Bocchiola et al., 2003) and potential difficulties in (Wan Jaafar et al., 2011) for the index flood estimation and modeling. Among them, the index flood estimation model known as QMED model in Flood Estimation Handbook (FEH) is a well-established of this kind in the UK. The Flood Estimation Handbook (FEH) was first published in 1999 (IH, 1999) and later improved in 2008 by Centre for Ecology and Hydrology (CEH) (Kjeldsen et al., 2008). In the UK, FEH is used for examples in flood defense planning, flood risk analysis, new development planning and rarity assessment of notable rainfalls or floods. Please refer to IH (1999) for further details.

Practically in the UK, the FEH defines the index flood as the median of annual maximum series denoted as QMED (IH, 1999), which is slightly different from traditional choice of mean as an index flood in hydrological application analysis such as RFFA. At any interest locations of ungauged catchments, the index flood is estimated using a combination of the QMED model and data transfer procedure from geographically close and hydrologically similar gauged catchments. This is to ensure for more accurate estimates rather than using the index flood regression model alone. The FEH index flood regression model and data transfer procedure have each been updated recently (Kjeldsen and Jones, 2007, 2009; Kjeldsen et al., 2008). This research attempted to improve the FEH-QMED model by incorporating geographical proximity directly into the index flood regression model. Hence, the improvements were determined by using proposed prediction method based on estimation model selection criteria.

1.4 Uncertainty in index flood estimation

The estimation of index flood at ungauged catchments is an important aspect in the regional flood frequency analysis (RRFA). A review on the recent advance in index flood estimation was presented by Bocchiola et al. (2003) ranging from simple direct method through complex indirect conceptual method. Brath et al. (2001) conducted a comparison study on different methods for estimating the index flood at ungauged catchments in Northern Italy. They found that a regression model linking index flood to a set of catchment characteristics provides the most efficient estimates of the index flood at ungauged catchments. Such regression models are widely used with numerous examples in the hydrological literature (Canuti and Moisello, 1982; Acreman, 1985; Mimikou and ordios, 1989; Tasker and Stedinger, 1989; Garde and Kothyari, 1990; Reimers, 1990). From the statistical point of view, there quite some questions that need to be investigated: (1) potentially non-linear impact of covariates, (2) autocorrelation in the observations, (3) non-Gaussian distribution of observations and (4) potentially unconsidered covariates.

As mentioned in Section 1.3, there are two types of power form equation available to forge a link between flood statistics with catchments characteristics. Power form equation with multiplicative error terms can be linearized by a logarithmic transformation and the parameters of the linearized model can be estimated by a linear regression technique. Due to the fact that linear regression technique is a well established framework and is easy to interpret and understand, the power form equation with multiplicative error term has become the most widely used method in practice. Nevertheless, in many real data situations, the relationship between some covariate(s), X and response y is not considered so straightforward. In dynamic processes such as flooding, the relationship between response variable and its covariate(s) always seem to have non-linear pattern. One of the common approaches that is widely used is to add a transformation of the predictor X as a covariate in a linear regression. This approach could overcome some of the non-linear features in the model, but it may lose some information of the original data and it is unclear if this transformation is reliable. In addition, it has been proven in the literature, that even though the estimation of the linearized model parameters may be unbiased in the logarithmic domain, but in the real data flow domain it is still biased (Shu and Burn, 2004; Pandey and Nguyen, 1999; McCuen et al., 1990). One of the solutions to overcome this issue is by using non-linear regression technique. Comparative study between linear and non-linear regression in application of flood quantile and index flood estimation has been done by Pandey and Nguyen (1999) and they have found that non-linear regression produced better estimates than linearized the regression. See also, Wan Jaafar et al. (2011) on non-linear regression technique in estimating the parameters of power form equation with additive error term in index flood estimation. Another solution that can be used to model non-linear impact of covariates through additive model (c.f., Lu et al., 2007; Gao et al., 2006; Friedman and Stuetzle, 1981). The additive spatial regression model is more flexible than the linearized regression model where it replaces linear predictor with some smoothing functions of covariates, to investigate other potentially non-linear impacts of catchment characteristics identified by FEH-QMED model.

In practice, the quality of flood data has to be ensured before it can be used in any hydrological application analysis such as the RFFA. However, some unknown and unexpected features in flood data cannot be controlled due to its nature. For example, water level or river flow that represent flood peak discharge at a particular point in a river has well-defined spatial effect due to the flow of the river and the lay of the land while the direction in which water will flow within catchment is generally known. Therefore, local factors influencing index flood may not be adequately represented in the regression models, and it may be beneficial to adjust the estimates using local data from neighboring gauged catchments (Kjeldsen and Jones, 2010). The Flood Estimation Handbook (FEH) (IH, 1999) recommended the use of the data transfer procedure together with the FEH-QMED model when estimating the index flood at ungauged site, where one or more suitable gauged sites can be found among geographically close and hydrological similar gauged catchments. There is no specific guidance as to the limit of the geographical area or necessary degree of hydrological similarity. Nevertheless, studies have shown that geographical proximity gives superior impact in estimating the index flood at ungauged sites in comparison to hydrological similarity (Merz and Blöschl, 2005; Eng et al., 2007; Kjeldsen and Jones, 2007). Consecutive studies by Kjeldsen and Jones (2007, 2009, 2010) concluded that geographical proximity representing the spatial impact in the data transfer procedure is very important and useful to improve the regression results. Therefore, in this study we will incorporate spatial effects into the FEH-QMED model using spatial error analysis and fit the UK flood data by using spatial econometrics models. The performance of these models will then be determined by using prediction method based on estimation model selection criteria. This can be seen in Chapter 2.

Another aspect that contributes to the uncertainty in index flood estimation at ungauged catchments is the assumption of Gaussian distribution in the index flood regression model. The Gaussian distribution is widely used in describing data in many applications but it is no longer suitable in environmental, hydrological and ecological studies since the observed spatial variables in those fields are known to have skewed distributions (Zhang and El-Shaarawi, 2010). In the hydrological application analysis such as the RFFA, the step to choose appropriate statistical distribution called as parent distribution has been well-studied and numerous examples are available in hydrological application studies (see, e.g., Stedinger, 1993; Peel et al., 2001; Vogel and Fennessey, 1993; Hosking and Wallis, 1997; Robson and Reed, 1999). This is because selection of parent distributions is compulsory for constructing regional growth curves to estimate the growth factor that needs to be multiplied with index flood before floods at any interest sites can be estimated. In practice, IH (1999), Wilson et al. (2011) and Pilgrim and Institution of Engineers (2001) it have shown that the distribution with three parameters such as location, scale and skewness outperform the two parameters distributions in certain criteria for the purpose of constructing the growth curves. In estimating the index flood by using regression models, Gaussian assumption was mostly used. This has motivated us to develop a new index flood regression model taking into account the spatial dependency and non-Gaussianity. Therefore, we will propose a spatial error model with skew-normal distribution for regression residuals with maximum likelihood estimation developed in Chapter 3. The skew-normal distribution is chosen because it has a desirable property which is skewness parameter that can represent the real data situations. However, drawing the inference for the parameters in skew-normal distribution is complex and challenging. The challenges will be explored in Chapter 3, which may give improve results in index flood regression model, hence contributing to new knowledge in the related field.

Flooding is a complex process that is caused not only by hydrological factors, but also by a lot of other physical factors. Numerous examples of index flood regression models that are available in hydrological literature were constructed based on limited number of explanatory catchment characteristic variables (Kjeldsen and Jones, 2010). Such models could be biased in the variable selection if there are more catchment characteristics variables available in data sets but they are not considered in the models. It can contribute to the uncertainty in index flood regression model. Nevertheless with many potential candidate predictors and some of them are detected to be highly correlated could give regression problem that require an efficient statistical model selection to find an optimal model, one that is as simple as possible while still providing good predictive performance. In general, a model should not be too simple, which may not capture the process behavior at reasonable degree, but if the model is too complex, it would involve too many parameters to estimate (Nelles, 2001; Hong and Mitchell, 2007). As a trade off, there should be an optimal model which contains the most appropriate number of parameters to be generalized well when tested using further data sets. Traditional variable selection procedure such as stepwise selection methods, Akaike's information criteria and Bayesian information criteria, suffer from high variability (Breiman, 1996) due to the estimation and variable selection are resolved separately. To overcome the aforementioned drawback in the traditional variable selection methods and enhance the prediction accuracy, a variety of penalized methods such as bridge regression, least absolute shrinkage and selection operator (LASSO), and elastic net have been developed in linear regression setting and gained popularity (see, e.g., Frank and Friedman, 1993; Tibshirani, 1996; Zou and Hastie, 2005). Unlike traditional methods, penalized regression methods perform variable selection and coefficient estimation simultaneously that can produce more stable results for correlated data or data where the number of predictors is much larger than sample size. Therefore, this study will utilized a novel application of adaptive lasso introduced by Zou (2006) as a tool in variable selection for our established spatial error model with skewed normal residuals. Adaptive lasso has been chosen because it proved not only retains the good features of Lasso but also enjoys the Oracle property where Fan and Li (2001) speculated that Lasso does not have it. The simulations of the performance for this method will then be investigated before it can be applied into the UK flooding data to discover relevant predictive catchment characteristics variables that may improve the index flood estimation. This can be seen in Chapter 4.

1.5 Research framework

This thesis aims to improve the existing methodology in index flood estimation model where FEH-QMED model is chosen as the benchmark. Three objectives have been developed based on the shortcomings of current research in literature that detailedly discussed in Section 1.4. The developed objectives are (i) to explore the shortcomings of the current research with a possibly improved model for estimation the index flood in the UK, (ii) to develop a more efficient statistical model in estimating the index flood that better fits the UK flooding data, (iii) to discover more relevant predictive catchment characteristics variables that may improve the index flood estimation model. We have constructed our methods based on these objectives and with that our research framework can be divided into three stages which are model exploration, model development and model selection that have been shown in Figure 1.1.

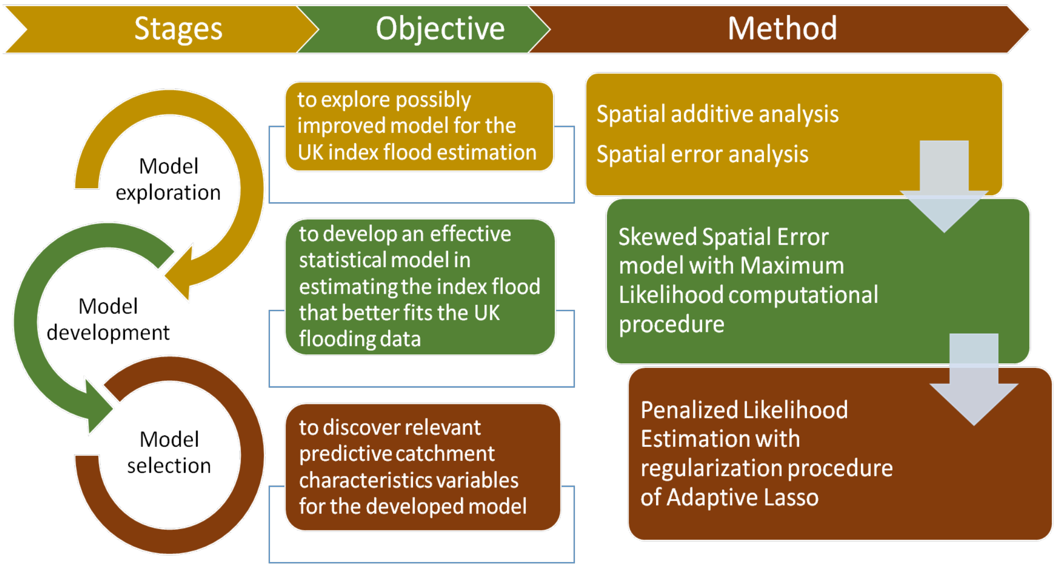


Figure 1.1: Research framework

In model exploration stage which is covers the first objective, we are exploring two fundamental questions on the FEH-QMED model based on the flooding dataset consisting of 586 gauged stations around the UK: (i) Is the FEH-QMED model reliable in characterizing the nonlinear features of the catchment characteristics for the UK flood estimation? (ii) Can the FEH-QMED model be better improved in estimating the index flood with spatial information well identified and utilized? To answer these questions, the FEH-QMED model have been chosen as a benchmark due to its worldwide recognition and for being the basis of most studies on flood risk estimation and modeling. In the UK, updated flood information of gauged stations and high quality data of flood peak are available through the National River Flow Archive. The Centre for Ecology and Hydrology (CEH) has as well published the FEH CD-ROM that provides a large number of catchment characteristics, which is now been replace by web service https://fehweb.ceh.ac.uk/. These sources of data make it possible for us to explore our analyses. In order to investigate the reliability of the identified nonlinear effects of the catchment characteristics in the FEH-QMED model (see Question (i) above), this study starts by applying a spatial additive regression analysis (c.f., Lu et al., 2007) to examine the nonlinear effects of the catchment characteristic covariates in the flood index estimation. This is a kind of semi-parametric model (Gao et al., 2006), by which some unknown or unexpected complex nonlinear features, if existed, would be extracted from the data. Further, in order to check potential spatial dependence, a diagnostic test called Moran's I popular in exploring spatial dependence (c.f., Moran, 1950 and Paradis, 2016) is used to examine the regression residuals. Note that spatial correlation or neighboring impact does have been noticed in the updated FEH procedure by using the QMED model together with the data transfer procedure in estimating the index flood at an ungauged catchment (c.f., Kjeldsen and Jones, 2007, 2009, 2010). The FEH-QMED model is a general linear model with covariates taken as the identified nonlinear functions of the catchment characteristics, where it is as usual assumed that the residuals are independent from each other. After identification of the spatial dependence in the residuals of the FEH-QMED model, this study will further consider spatial econometric models in fitting of the UK flood data, in order to answer Question (ii) above in which to explore the possible improvement over the FEH-QMED model. It is hoped that the results from these analyses will answer Questions (i) and (ii) and therefore, the first objective of this research could be achieved. Interestingly, our additive analysis confirms the nonlinear effects of the catchment descriptors identified in the FEH-QMED model, while our Moran test and spatial error analysis do find that the spatial autocorrelation significantly exists in the flood data, which supports the earlier finding in Kjeldsen and Jones (2007). This fact has led us in developing spatial autoregressive index flood estimation models which we will demonstrate outperform the updated FEH model (Kjeldsen and Jones, 2010) in prediction.

The second objective of this study is motivated by the results found from the empirical works of model exploration stage in which spatial error model has been found to be an improved model in estimating the index flood in the UK. Furthermore, exploratory analysis has shown that the UK flooding data and the residuals of spatial error model could have a kind of skewed distribution. This has led us to the second stage of the research in developing a new statistical model for flood index estimation by taking into account spatial dependency and non-Gaussianity. The fact that asymmetry in data makes the Gaussian assumption not valid, has called for spatial error model for skewed UK flooding data. Thus, a skew-normal distribution of Azzalini (1985) is suggested for the distribution of the UK flooding data under study which takes normal distribution as a special case. While the normal distribution with its symmetry has only location and scale parameters, the skew normal distribution has an additional asymmetry parameter describing the skewness. This is very desirable property that perhaps can represent the UK flooding data situations. To develop maximum likelihood procedure for this new model, we firstly need to replace the distribution of the residuals in spatial error model from normal to skew normal. By using some common work with location and scale transformation we specify a skewed normal distribution of the residuals for spatial error model with appropriate location and scale parameter in term of asymmetry parameter such that the mean is zero and variance is one. Hence, the probability density function for the skew normal distribution of the residuals in spatial error model can be determined by using transformation rule. By having this function the likelihood for the joint vector of response variable observation y of spatial error model with skew-normal distribution can be obtained. From this maximum likelihood principle, we get a complicated likelihood function. Thus, it gives nonlinear equations of the score functions. Because of its nonlinearity, we cannot find the maximum likelihood estimators by solving the equations in a closed form. Therefore, we suggest the computational procedure to solve the maximum likelihood estimation for the proposed statistical method. We also present the simulations to investigate the finite sample performance of the proposed statistical methods. It has been proved that the estimators of the proposed model are unbiased and consistent. This has led us to apply the method into the UK flooding data which is demonstrated outperform the model established in the first stage of this research and the updated FEH model (Kjeldsen and Jones, 2010) from prediction perspective.

The objective in the final stage of this research is concerned about variable selection in index flood estimation model because there are more catchment characteristics variables are available in the UK flooding dataset that never been considered in model development. Practically, maximum likelihood estimation with traditional selection procedure such as such as stepwise selection methods, Akaike's information criteria and Bayesian information criteria are used for model estimation and model selection in flood regression models in the literature. This traditional method has a number of drawbacks including high variability due to the estimation and variable selection have been executed in separate steps especially when the number of predictor, p larger than the sample size n or high correlated in predictor variables. This drawback has forced them to reduce the number of considered catchment characteristics especially the one that has been detected to be highly correlated to ensure for efficiently approximating the solution in those flood regression models with underlying truth. This procedure could bias in variable selection since not all available catchment characteristics are equally considered. Furthermore, the variable selection in those flood regression models was made under independent error structure. This could also bias the variable selection for the model. Aforementioned problems have motivated us to suggest a penalized estimation procedure by utilizing adaptive lasso to estimate the important catchment characteristics predictors for the UK index flood estimation under spatial dependent error models by considering all available catchment characteristics in the dataset. A computational procedure is suggested to determine the solution for panelized likelihood estimators. With this suggested computational procedure the simulations of the performance for finite sample of the proposed panelized likelihood estimation method is investigated with regards to 'sensitivity and specificity test' and 'unbiasedness and consistency of the estimators'. It has been proved that the proposed penalized likelihood method that utilizing adaptive Lasso as a tool in regularization procedure has an ability in correctly identifying the sparse solutions for the regression coefficients and can produce the unbiased and consistent estimators of skewed spatial error model. This has led us to apply the suggested computational procedure of panelized likelihood estimation with adaptive Lasso regularization tool into the UK flooding data which is demonstrated outperform the best model established in the first and second stage of his research based on AIC score which was shown by the largest percentage improvement in mean square prediction error, MSPE relative to the updated FEH-QMED model (Kjeldsen and Jones, 2010) .

1.6 Structure of the thesis

The main contributions of this thesis are summarized as follows:

(i). We have found that the spatial error model with contiguity based weight matrix, by incorporating spatial dependence directly into the FEH-QMED model, can better fit the UK flooding data, which has been illustrated to outperform the latest revised FEH-QMED model in Kjeldsen and Jones (2010) from the prediction perceptive.

(ii). This research has proposed and developed a new statistical model structure for spatial error model with the skewed normal distribution in responding to the issue of non-Gaussianity in the real flooding data, which improves the index flood estimation model.

(iii). This study has suggested a new computational procedure detailed in Algorithm 1 to solve the proposed maximum likelihood estimation method for the skewed spatial error model.

(iv). We have proposed a penalize maximum likelihood estimation for variable selection with our established spatial error model with skewed normal residuals by utilizing adaptive lasso as a tool, and a new computational procedure detailed in Algorithm 2 is also suggested to solve the proposed panelized likelihood estimation method.

The structure of the following chapters in this thesis is as follows. Chapter 2 gives a detailed discussion on empirical work by using spatial additive and spatial error analyses in exploring possibly improve models for index flood estimation in the UK where FEH-QMED model is chosen as the benchmark. Information on the sources of the UK flood dataset with its catchment characteristics variables, an in-depth explanation of the FEH-QMED model as well as questions and methodology in investigating the FEH-QMED is also provided in Chapter 2. The results of the application by using the spatial additive and spatial autoregressive models into the UK flooding data have been reported together in Chapter 2 with its discussion and conclusion. Work on Non-Gaussianity in the spatial index flood estimation model is discussed in Chapter 3. This chapter contains maximum likelihood estimation procedure for spatial error model with a skewed normal distribution of the error terms, including both methodology and computational parts. The results of the simulation in investigating the finite sample performance and the application of the proposed method have been reported together with discussions and conclusions. Chapter 4 concerns on variable selection in Index flood Estimation model. This chapter contains the penalized maximum likelihood estimation with some background on adaptive lasso and computational algorithm procedure. As in the previous chapter, the results of the simulation in investigating the performance of finite sample and the application of the proposed panelized likelihood estimation method with discussions and conclusions are also reported together. Chapter 5 concluding remarks as chapters within this thesis have their own discussion and conclusion section.

Chapter 2

Exploring the UK Index Flood Estimation Model

2.1 Introduction

Empirical work is very important in applied statistical studies because it deals with real data to solve real problems. This chapter has empirically attempted to investigate the reliability and explore the possibly improved model for the UK index flood estimation. Better understanding of real dataset is essential to ensure appropriate framework of statistical method is applied in developing mathematical models. Due to the emergence of modern data acquisition techniques together with powerful and intelligent computational tools, voluminous hydrological data have been and continue to be collected. As a result, many hydrological process models have been developed by active researchers in this area. There are urgent needs for efficient statistical method to extract unknown and unexpected complex features in datasets as well as detailed investigation on checking the reliability of those models requiring new advance methods in statistics. Catchment areas in the UK have been chosen as the study area because it is equipped with updated flood information and it has many gauged stations. Besides, Flood Estimation Handbook (FEH) published in 1999 by Institute of Hydrology (IH) which is now known as the Centre for Ecology and Hydrology (CEH) is a well known handbook and has become the basis for most studies on flood frequency estimation. Therefore, index flood estimation model known as QMED model introduced by FEH is chosen as the benchmark model to assess the alternative index flood models developed in this research.

There are about one thousand and five hundred gauged stations around the UK, but not all stations are suitable for flood risk study. The data sets are based on 602 sample stations around the UK that were selected by Kjeldsen et al. (2008) in the science report of "Improving the FEH statistical procedures for flood frequency estimation" published by Environment Agency. Dataset provided by the HiFlows-UK project released on 1st August 2005 superseded the FEH dataset provided in Volume 3 of the Flood Estimation Handbook (IH, 1999). Since that, updates and changes have been made to UK flood data through the HiFlows-UK Notes to ensure the quality of the dataset for the purpose of improvement in flood risk study. To date, eight HiFlows-UK Notes have been released and the latest version 3.3.4 was released in August 2014. The important changes made to improve the quality of the dataset of HiFlows-UK project ware in regard to the list of stations that have been changed, removed and closed. Out of 602 stations, 6 stations have two different station numbers with the same station names due to incorrect information, 8 stations were removed and 2 stations were closed due to lack of data quality. Therefore, only 586 rural stations have been considered in this study and the locations are shown in Figure 2.1.

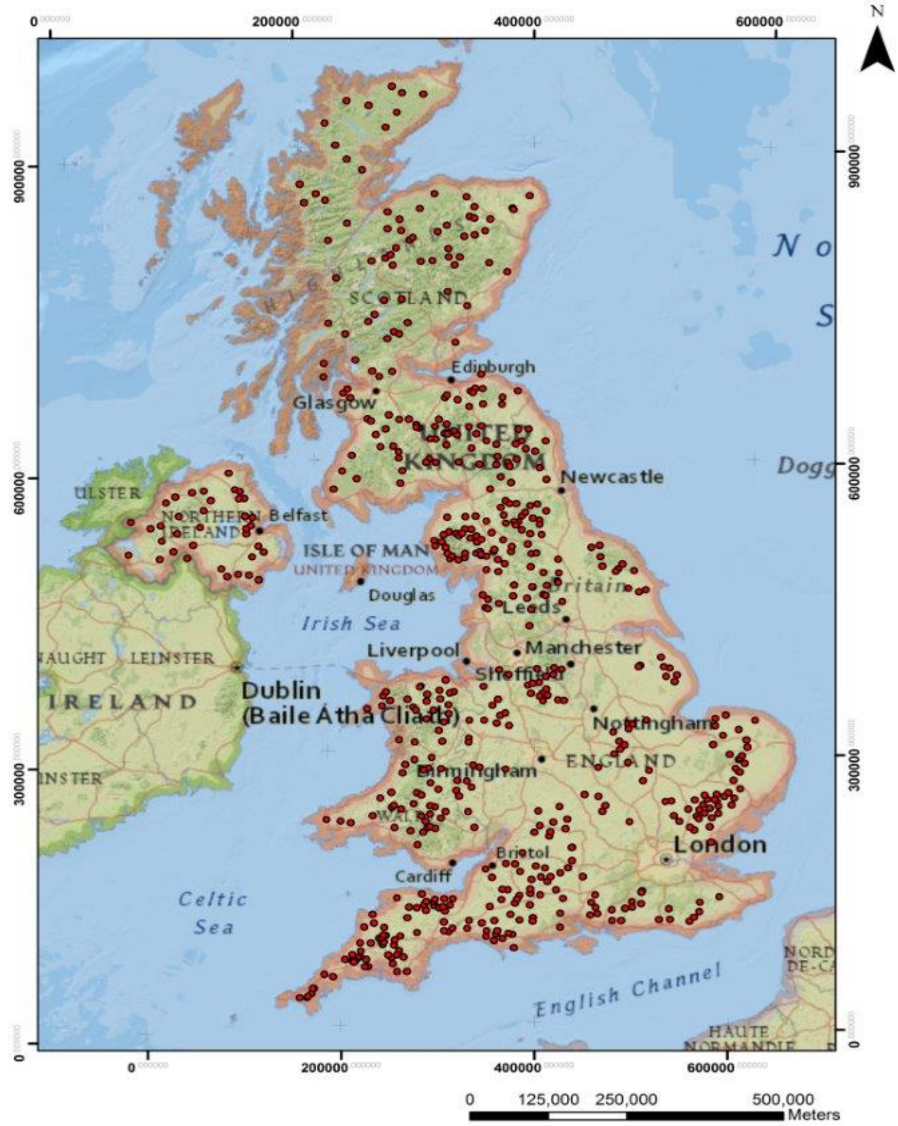


Figure 2.1: Location of 586 gauging/monitoring stations on rural catchments providing annual maximum flood peak data

This chapter will;

(i). apply the spatial additive and spatial error analyses to explore the FEH-QMED model that has been chosen as a benchmark for the purpose of improving the UK index flood estimation model.

(ii). fit the spatial autoregressive models into the UK flooding data to develop a new index flood estimation model that taking into account the spatial autocorrelation in the observations.

(iii). develop a method to compute predictive random values for spatial autoregressive models based on leave one out cross validation (LOOCV) technique for the purpose of evaluating the established models.

(iv). determine the estimated distribution of the residuals in the established models for further research direction of this thesis.

Perhaps, by completing the aforementioned tasks in this chapter, the first objective of this research study which is to explore the shortcomings of the current research with a possibly improved model in estimating the UK index flood can be achieved.

2.2 Data sources

Two types of data used in this thesis are observed flood peaks data and physical catchment characteristics data. The observed flood peaks data have been taken from HiFlows-UK project that are available at http://nrfa.ceh.ac.uk/data/search, while the physical catchment characteristics data are from FEH CD-ROM 3.0 which is now been replace by web service https://fehweb.ceh.ac.uk/. These websites belong to the National River Flow Archive (NRFA) which it is the UK's focal point for river flow data under the Centre of Ecology and Hydrology (CEH). The NRFA collates, quality controls, and archives hydrometric data from gauging station networks across the UK including the extensive networks operated by the Environment Agency (England), Natural Resources Wales, the Scottish Environment Protection Agency and the Rivers Agency (Northern Ireland). NRFA data also includes catchment rainfall totals derived from Met Office data and various spatial data sets such as digital elevation data, land cover, geology and hydrogeology that developed by CEH and the British Geological Survey (BGS).

2.2.1 Flood peak data

Throughout the year, there are many flood peaks associated with individual storm events and have been recorded as continuous time series stream flow discharges. Two types of flood peaks data that can be extracted from such a record are annual maximum (AM) series and partial duration (PD) series or peak over a threshold (POT) series. These two types of data have traditionally been used in statistical flood risk and frequency analysis. The AM series are formed by extracting the maximum discharge in each year. This yields the series where $\{w_1 \ldots w_n\}$, with w_i for the maximum discharge in the i^{th} year of the n-year record. The data in the AM series can be used to estimate the probability of maximum flood discharge in a year that exceeds a particular magnitude which is called as annual exceedance probability. The PD or POT series are extracted from the record of every statistically independent peak discharges (that exceeds the threshold discharge). This yields the series where $\{v_1 \ldots v_m\}$, with v_i is the peak discharge associated with the i^{th} statistically independent flood event in the n-year record. The threshold must be the smallest AM series value within a selected period of study so that m is about 2 to 3 times greater than n.

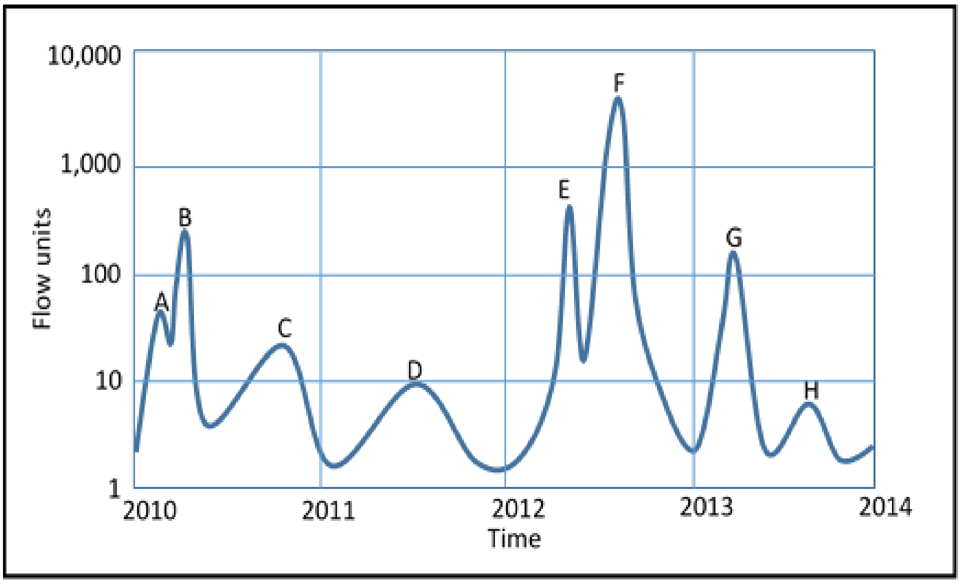


Figure 2.2: Illustration of time series stream flow discharge

The meaning of peak flow can vary in practice. Ideally, an instantaneous maximum peak flow will be used. The instantaneous peak flow is the peak flow occurring at any moment in a day. Unfortunately, such values are not always available. Instead, the annual maximum daily flow is often used. This is based on the average daily flow values where the average daily flow is the average flow over a 24hour period. In a year we will have 365 (or 366) daily flow values. Each of these values represent the average flow over the calendar year. The annual maximum daily flow is the greatest of those values. So in the case of AM series that use average daily flow in the UK it is defined as the greatest 24hour average flow within each water year. The water year is defined as a year that started from October of the year until September of the next year in the calender year. There are situations when it is advantageous to represent more than one flood peak per water year. For example, floods that frequently occur has a reasonable chance to happen more than once per water year. It is also possible that the second or third largest peak in a particular water year may be larger than the maximum flood of some other water year. In AM series, such additional events were ignored since the largest annual event is only allowed. The

POT series includes all events above some arbitrary threshold value. The threshold value must be the smallest AM series value within a selected period of study. Figure 2.2 clearly shown that, from water year 2010 until 2014, the AM series contained the value of peaks B, D, F and G, while peaks A, B, C, D, E, F and G are included in POT series by considering 10 flow units as the threshold.

This research project used AM series because first, as mentioned by Kjeldsen et al. (2008), POT series from HiFlows-UK database was found to be inadequate and in addition to that there is no significant difference to the number of affected stations if either POT or AM series are used. The second reason is because, this study wants to avoid ambiguity in selecting flood peaks. It would be difficult to select independent peaks for POT series in each water year, because the flood peaks data used for valid frequency analysis should constitute a random sample of independent values that ideally come from a homogeneous population. The independence requirement means floods occur individually and do not influenced each other, while the homogeneous requirement means floods occur under the same type of conditions. Perhaps by using AM series, this ambiguity that is illustrated by Figure 2.2 could be avoided. Refer to Figure 2.2 if POT series are used therefore all peaks values can be accepted except for H value. This will allow some water years to have two or more peaks values. On the aforementioned note, three ambiguities will arise when selecting independent peaks for POT series for each water year from 2010 until 2014. The first is the independence between peak A and B, the second is the independence between peak C from peaks A and B and the third is the independence between peak E and F. Perhaps, by using AM series issue of independence and homogeneity in flood peaks data can be avoided

The AM series data used in this research project are from 586 stations with the length of years varying from 7 to 127 years. These AM series are used to determine the important hydrological parameters in the index flood estimation model known as median annual maximum flood (QMED), that is used as the index flood. The AM series used in this study is the current peaks flow dataset available in NRFA website. This dataset is based on HiFlows-UK note released in April 2014 that contains data for stations in England and Wales until 30 September 2012 while data for stations in Scotland are until 30 September 2006. There are a few stations that give data ended before both mentioned water years but are still considered in this study due to the history of the catchment areas that are suitable for flood risk study.

2.2.2 Catchment characteristics

The Central for Ecology and Hydrology (CEH) formerly known as the Institute of Hydrology (IH) in the UK has provided a large number of catchment characteristics, or also known as catchment descriptors, in Flood Estimation Handbook (FEH). Catchment descriptors in the FEH have been derived from the CEH proprietary map on soil, land use, terrain, etc. (IH, 1999). The FEH uses the Institute of Hydrology Digital Terrain Model (IHDTM) to define a range of catchment characteristics. They can be extracted from the FEH CD-ROM based on the raster data set at 50 m resolution. The FEH CD-ROM has been updated to version 3.0, which is now been replace by web service https://fehweb.ceh.ac.uk/ for the purpose of flood risk assessment studies (CEH, 2009). The number of potential catchment characteristics available are large, but only the variables that have previously found to be useful in flood risk study will be included in our study; see Table 2.1 below for details. The selected catchments characteristic variables are based on Kjeldsen et al. (2008). However, two of the catchment characteristics which are the steepness of design rainfall growth curves (PRAT) and the annual evaporation (EVAP) are no longer stored in the database of the UK National River Flow Archive at the CEH because as argued by them these descriptors are deemed to be not useful in the flood model development. Therefore they are replaced by other catchment descriptors of the standard percentage runoff (SPRHOST) and the longest drainage path (LDP), which claimed to be useful by Robson and Reed (1999) and Badyalina and Shabri (2013). The observations of all the variables in Table 2.1 are also based on the same 586 stations around the UK.

Table 2.1: Summary of catchment characteristics

Descriptors name	Unit	Range	Note
AREA	km^2	$[0; \infty]$	Catchment drainage area.
SAAR	mm	$[0; \infty]$	Average annual rainfall.
FARL		$[0; 1]$	Flood attenuation by reservoirs and lakes.
BFIHOST		$[0; 1]$	Base flow index derive by Hydrology of Soil Type data.
PROPWET		$[0; 1]$	Proportion of time when soil moisture deficit $\leq$ 6mm.
DPSBAR	m/km^{-1}	$[0; \infty]$	Mean catchment slope.
FPEXT		$[0; 1]$	Floodplain extent.
RMED(1day)	mm	$[0; \infty]$	Median annual maximum 1-day rainfall.
SPRHOST		$[0; 1]$	Standard percentage runoff.
LDP	km	$[0; \infty]$	Longest drainage path.

2.3 Details of FEH-QMED model

The used regression models in previous studies linking the index flood to the catchment characteristics or descriptors in the UK were established as early as in 1965 by Cole (1965) where the mean annual maximum flood (QBAR) was used as an index flood linked to the catchment area (AREA) as an explanatory variable in the model. The use of QBAR as the index flood for flood frequency estimation in the UK continued until the Natural Environmental Research Council published a report on flood studies called Flood Studies Report (FSR), but the number of the catchment characteristics used has increased from one to five variables (NERC, 1975).

In 1999 the Institute of Hydrology (IH) published a book called Flood Estimation Handbook (FEH), which has been used as a standard procedure for flood risk assessment in the UK till the present. Since the publication of FEH, the index flood has evolved from using the QBAR to the median annual maximum flood (QMED) (IH, 1999). The FEH introduced the regression model of QMED with six explanatory catchment variables for general use in the UK, as the

QMED was considered to be more robust to outliers in short series. Improvement over the FEH statistical method had been made and a new report (Kjeldsen et al., 2008) was published by the Environmental Agency. This report is an extension of the work undertaken in the FEH by considering the recommendations made by Morris (2003) and the identification of the link between the model error structure of the QMED regression model and the benefit obtained from the use of the data transfer through a series of publications by Kjeldsen and Jones (2006, 2007, 2009). As a result, Kjeldsen et al. (2008) have developed the new regression model to link the QMED to the catchment descriptors and the revised data transfer procedure emphasized by the FEH as an important procedure in enhancing the initial regression estimates.

The FEH has been the most widely used procedure in flood risk assessment in the UK and it has provided the excellent catchment descriptors (CDs) as well as providing high-quality QMED model for the purpose of estimating flood index at ungauged catchment. Both the original and the updated ones are a kind of multiple regression model that links the catchment characteristics to the flood flow. They are based on Equation (1.3), the most commonly used power form equation (Brath et al., 2001; Grimaldi et al., 2011), and use generalized least square procedure or maximum likelihood method for parameter estimation (Draper and Smith, 1981; Stedinger and Tasker, 1986; Tasker and Stedinger, 1989) from the linearized equation using logarithm transformation,

$$\ln(\mu_Q) = \ln(a) + \beta_1 \ln(x_1) + \beta_2 \ln(x_2) + \ldots + \beta_n \ln(x_p). \tag{2.1}$$

Variable selection was carried out by using an exhaustive search to find a relatively small set of variables, with stopping criteria such as the coefficient of determination R^2, adjusted R^2 or the predicted error sum of squares used, which provided a good statistical fit to the QMED. The optimal set of the variables suggested by the good fit to the QMED data from 728 gauged catchments located throughout the UK is the six variables (AREA, lnAREA, SAAR, FARL, SPRHOST and BFIHOST) as the best model, while by using the data from 602 gauged catchments, Kjeldsen et al. (2008) improved the model performance through a new QMED model with only four variables which are the AREA, SAAR, FARL and BFIHOST (in Section 2.1 detail was described about the reduced number of the gauged stations considered in this study). Equation (2.2) to Equation (2.5) are the original and the updated QMED models from the FEH in the linearized power forms, respectively as follow:

<u>The original QMED model (IH, 1999)</u>

$$\begin{aligned} \ln(QMED) \quad = \quad & 0.159 + \ln(AREA) - 0.015\ln(AREA)\ln\left(\frac{AREA}{0.5}\right) \\ & + 1.560\ln\left(\frac{SAAR}{1000}\right) + 2.642\ln(FARL) + 1.211\ln\left(\frac{SPRHOST}{100}\right) \\ & - 3.923\Big[BFIHOST + 1.30\Big(\frac{SPRHOST}{100}\Big)\Big] - 0.987, \end{aligned} \tag{2.2}$$

$$\begin{aligned} QMED &= 1.172 \times AREA^{1-0.015\ln\left(\frac{AREA}{0.5}\right)} \times \left(\frac{SAAR}{1000}\right)^{1.560} \times FARL^{2.642} \\ &\times \left(\frac{SPRHOST}{100}\right)^{1.211} \times 0.0198^{BFIHOST+1.30\left(\frac{SPRHOST}{100}\right)-0.987}. \quad (2.3) \end{aligned}$$

The new QMED model (Kjeldsen et al., 2008)

$$\begin{aligned} \ln(QMED) &= 2.1170 + 0.8510\ln(AREA) - 1.8734\left(\frac{1000}{SAAR}\right) + 3.4451\ln(FARL) \\ &- 3.080BFIHOST^2, \quad (2.4) \\ QMED &= 8.3062 \times AREA^{0.8510} \times 0.1536^{\left(\frac{1000}{SAAR}\right)} \times FARL^{3.4451} \\ &\times 0.0460^{BFIHOST^2}. \quad (2.5) \end{aligned}$$

In practice, the quality of flood data has to be ensured before it can be used in any hydrological application analysis. However, some unknown and unexpected features in flood data cannot be controlled due to its nature. For example, water level or river flow that represent flood peak discharge at a particular point in a river has well-defined spatial effect due to the flow of the river and the lay of the land while the direction in which water will flow within catchment is generally known. Therefore, local factors influencing index flood may not be adequately represented in the regression models, and it may be beneficial to adjust the estimates using local data from neighboring gauged catchments (Kjeldsen and Jones, 2010). The Flood Estimation Handbook (FEH) (IH, 1999) recommended the use of the data transfer procedure together with the QMED regression model when estimating the index flood at ungauged site, where one or more suitable gauged sites can be found among geographically close and hydrological similar gauged catchments. There is no specific guidance as to the limit of the geographical area or necessary degree of hydrological similarity. Nevertheless, studies have shown that geographical proximity gives superior impact in estimating the index flood at ungauged sites in comparison to hydrological similarity (Merz and Blöschl, 2005; Eng et al., 2007; Kjeldsen and Jones, 2007). Consecutive studies by Kjeldsen and Jones (2007, 2009, 2010) concluded that geographical proximity representing the spatial impact in the data transfer procedure is very important and useful to improve the regression results. As Kjeldsen and Jones (2010) has showed that the new QMED model (2.4) or (2.5) of the FEH together with revised data-transfer method performs better in estimating index flood at ungauged catchment, we will use it as a benchmark model as a comparison in the following sections of Chapter 2.

2.4 Questions to investigate FEH-QMED model

It is interesting to note that the FEH-QMED model (2.4) is a linear multiple regression model with nonlinear effects of the catchment descriptors identified. Is this a reliable model in estimating index flood in the UK? As indicated in Chapter 1 in achieving the first objective of this study, two fundamental questions will be investigated on the basis of the flooding dataset consisting of 586 gauged stations around the UK:

(i) Is the FEH model reliable in characterizing the nonlinear effects of the catchment characteristics for the UK flooding?

(ii) Could we improve the FEH model with a better accuracy in estimation of index flood?

For these purposes, as the FEH-QMED model (2.4) was chosen as the benchmark, the first objective of this research could be achieved by exploring possibly improved model for estimation of the index flood. Here the median annual maximum flood (QMED) is considered with the response variable, $Y = \ln(QMED)$, which is a function of four predictor variables which are the catchment drainage area (AREA), the average annual rainfall (SAAR), the flood attenuation by reservoirs and lakes (FARL) and the base flow index by hydrology of soil type data, (BFIHOST). Two assumptions underlying Equation (2.4) are: (i) the linearity between $Y = \ln(QMED)$ and the explanatory covariates

$$(X_1, X_2, X_3, X_4) = (\ln(AREA), \frac{1000}{SAAR}, \ln(FARL), BFIHOST^2), \tag{2.6}$$

and (ii) the independence in residuals between the observations for multiple regression model (2.4). These will be investigated by using spatial additive regression analysis for model structure of explanatory variables and by spatial error analysis for diagnostic test to achieve an improved model for index flood estimation. The background information on these methods are provided in Subsections 2.5.1 and 2.5.2.

2.5 Methods to explore FEH-QMED model

2.5.1 Spatial additive analysis

Given the response variable, Y, and the $p = 4$ explanatory variables, $X_1 \dots X_p$, as in Section 2.4, with their spatial data, $(Y(s_i), X_1(s_i), \dots, X_p(s_i)),\ i = 1, 2, \cdots, n$, observed from the $n = 586$ gauged stations around the UK (see Section 2.1), where $s_i = (u_i, v_i)$ standing for the location (latitude,longitude) of the ith gauged station. In general, the mean of Y is a function of $X_1, \dots, X_p$, say $\mu = E(Y|X_1, \dots, X_p) = f(X_1 \dots X_p)$. If the FEH-QMED model (2.4) is true, then it is a linear regression

$$\mu = \beta_0 + \sum_{j=1}^{p} \beta_j X_j. \tag{2.7}$$

Equation (2.7) shows that the predictors are multiplied by coefficients, $\beta_1, \dots, \beta_p$.

To explore if Equation (2.4) or (2.7) is true, we are considering a more general model, an additive spatial regression model (c.f., Lu et al., 2007; Gao et al., 2006; Friedman and Stuetzle, 1981) that is more flexible than (2.7) by replacing linear predictor with a sum of smoothing functions of covariates, $\sum_{j=1}^{p} f_j(X_j)$ in the form

$$\mu = \beta_0 + \sum_{j=1}^{p} f_j(X_j), \tag{2.8}$$

where the smooth function, f_j, is a non-linear predictor relating to the expected value of Y under the $p = 4$ covariates (X_1, X_2, X_3, X_4) defined in (2.6), and for identifiability it is often assumed that $Ef_j(X_j) = 0$. By the data $(Y(s_i), X_1(s_i), \ldots, X_p(s_i))$, $i = 1, 2, \cdots, n$, we can obtain the estimates of the unknown (nonparametric) functions, $\hat{f}_j(X_j)$ for $j = 1, \cdots, p$, and $\hat{\beta}_0$ using the so-called backfitting or other semiparametric methods (c.f., Lu et al., 2007; Gao et al., 2006; Friedman and Stuetzle, 1981). Clearly, if the resultant estimate $\hat{f}_j(X_j)$ is a linear function of X_j, then it just demonstrates that Equation (2.4) or (2.7) is true, i.e., the nonlinear effects of the catchment descriptors (AREA, SAAR, FARL, BFIHOST) identified in (2.6) are well presented in the QMED model (2.4) developed by Kjeldsen et al. (2008).

2.5.2 Spatial error analysis

After a regression model such as (2.7) or (2.8) is established, we need to check if the obtained model is appropriate by examining the residuals of the regression. We can define residuals $e_i \equiv e(s_i) = Y(s_i) - \mu(s_i)$, for $i = 1, 2, \cdots, n$, where $\mu(s_i) = \beta_0 + \sum_{j=1}^{p} \beta_j X_j(s_i)$ for (2.7) or $\mu(s_i) = \beta_0 + \sum_{j=1}^{p} f_j(X_j(s_i))$ for (2.8). In applications, β_j's and f_j's are based on their estimated values in calculating the residuals. Two statistical tests to detect the presence of spatial autocorrelation in the residuals are introduced in this study. Here a spatial weight matrix (c.f., Anselin (1988)) need to be constructed based on the locations of the $n = 586$ monitoring stations used in this study. Two types of spatial weight matrices are considered in this study, which are contiguity based and distance based, respectively. Both types need to be row-standardized, which means the weights need to sum up to one on each row.

The spatial weight matrix is defined as a $n \times n$ matrix, W, for a set of n locations, with elements w_{ij} indicating whether the locations s_i and s_j of the ith and the jth observations are spatially close, where $i \neq j$, and $w_{ii} = 0$ as usual. For a contiguity based matrix, the (i, j)th element is define as

$$w_{ij} = \begin{cases} 1, & \text{if } i \text{ is contiguous to } j, \\ 0, & \text{otherwise.} \end{cases} \tag{2.9}$$

Here, if stations i and j are neighbors in space, the spatial weight w_{ij} is one, and otherwise zero, by the definition of neighbor based on sharing triangle edges using the Delaunay triangulation concept (c.f., Bivand and Portnov, 2004; Estivill-Castro and Houle, 2001; Chen et al., 2014; Maus and Drange, 2010). For a spatial weight matrix based on distance, the element is defined as

$$w_{ij} = \begin{cases} 1, & 0 < d_{ij} \leq D, \\ 0, & d_{ij} > D, \end{cases} \tag{2.10}$$

where d_{ij} is a distance between the locations i and j, with D a defined distance band. Equation (2.10) indicates that stations within a defined distance band from each other are categorized as neighbors and have spatial weight of one, and otherwise zero. Typically, zero is set as the lower bound here, and the upper bound D is taken as the greatest value of the Euclidean distance between the list of the first order nearest neighbor (c.f., Anselin, 2003; Bivand, 2017).

The first test for spatial dependence in the disturbances of a regression model is Moran's *I*-statistic (Moran (1950)). It is the most widely used method in detecting spatial autocorrelation of a geo-referenced variable (see e.g., Dray et al., 2008; Kosfeld and Dreger, 2006; Kazembe et al., 2006). By having a spatial weight matrix W, with entries w_{ij} as defined in (2.9) or (2.10), Moran's *I* formula for an unstandardized weight matrix W^* with entries w^*_{ij} reads as:

$$I = \frac{ne'W^*e}{S_0 e'e}, \tag{2.11}$$

where $S_0 = \sum_{i=1}^n \sum_{j=1}^n w^*_{ij}$ and $e = (e_1, e_2, \cdots, e_n)'$ is an $n \times 1$ vector of the regression residuals, or, when using a row-wise standardized weight matrix W with $S_0 = n$,

$$I = \frac{e'We}{e'e}. \tag{2.12}$$

The statistic I asymptotically follows a standard normal distribution under the null hypothesis of no spatial dependence (c.f., Cliff and Ord, 1972, 1973, 1981). In R, this test is implemented using function `lm.morantest`.

The Moran's I test statistics has high power against a range of spatial alternatives, but does not provide much help in terms of which alternative spatial regression model would be most appropriate. An alternative test for spatial dependence is called Lagrange Multiplier Test, which allows a distinction between spatial error models and spatial lag models. Unlike the Moran test, the Lagrange multiplier test relies on a well-structured hypothesis for test of spatial error dependence or spatial lag dependence.

The Lagrange multiplier test for spatial error dependence, or the so-called LM error test, is based on the least-squares residuals of spatial error model (2.15) below with calculations involving a spatial weight matrix W, conditional on having a λ parameter not equal to zero in the model. The LM error test statistic takes the form (Anselin, 1988):

$$LM_{err} = \frac{(e'We/s^2)^2}{T} \sim \chi^2(1) \tag{2.13}$$

with $s^2 = e'e/n$, e denote a vector of the least-squares residuals and $T = tr[(W + W')W]$. The Lagrange multiplier test for spatial lag dependence, or the so-called LM lag test, is based on the residuals from the spatial lag model (2.16). It can be used to examine whether inclusion of the spatial lag term eliminates spatial dependence in the residuals of the model or not. This test differs from the two tests outlined above because it allows for the presence of the spatial lag variable in the model. It is conditional on having a ρ parameter not equal to zero in the model, rather than relying on the least-squares residuals. The LM lag test takes the form (Anselin, 1988):

$$LM_{lag} = \frac{(e'Wy/s^2)^2}{nJ} \sim \chi^2(1), \tag{2.14}$$

where $nJ = T + (WX\beta)'M(WX\beta)/s^2$, $T = tr[(w + w')w]$, $M = I - X(X'X)^{-1}X'$. Both Lagrange multiplier tests as well as their robust forms are implemented in the R function `lm.LMtest`.

As mentioned above, two cases of spatial dependence could arise in (2.7) or (2.8). The spatial error model is appropriate when the focus of interest is in correcting spatial autocorrelation due to use of spatial data, irrespective of whether the model of interest is spatial or not. The spatial error model can be expressed as

$$y = X\beta + e, \tag{2.15}$$

$$e = \lambda We + \varepsilon, \;\; \varepsilon \sim N(0, \sigma^2 I_n).$$

Here the $n \times 1$ vector y contains the dependent variables and X represents the usual $n \times p$ data matrix containing explanatory variables, W is a spatial weight matrix and the parameter λ is the coefficient on the spatially correlated errors. The parameter β reflects the influence of the explanatory variables on variation in the dependent variable y. Maximum likelihood estimation of the spatial error model (Anselin, 1988) can be implemented by the R function `errorsarlm`.

The alternative model is a spatial lag model, which is appropriate when the focus is on the spatial interactions of the dependent variables. Here the existence and strength of spatial interactions are the focus of interest to be assessed. The spatial lag model takes the form:

$$y = \rho Wy + X\beta + \varepsilon, \tag{2.16}$$

$$\varepsilon \sim N(0, \sigma^2 I_n),$$

where y, X and W are similarly as defined in (2.15). The parameter ρ is the coefficient of the spatially lagged dependent variable, Wy, and the parameter β reflect the influence of the explanatory variables on variation in the dependent variable y. Maximum likelihood method for estimating the parameters of this model Anselin (1988) can be carried out with the R function `lagsarlm`.

2.5.3 Model selection criteria

Comparative evaluation between the fitted models in this study are based on Akaike information criterion (AIC) and generalized cross validation (GCV) score. AIC is one of the most popular approaches in model selection that compares multiple models by taking into account both descriptive accuracy and parsimony (Akaike, 1974; Burnham and Anderson, 2003). The basic idea of the AIC is that, if the distribution of the data is known and there are choices of the model in fitting the data set, a preferred model could be figured out by noting which has a lower KullbackLeibler divergence that minimized the information loss from the distribution by using selected models. However, in the real cases the distribution of the data are always unknown, but the distribution of each models could be estimated from the dataset, hence how much (or less) information lost could be estimated via AIC through Equation (2.17), where K is the number of estimated parameters in the model and L is the maximum value of likelihood function for the model.

$$AIC = 2K - 2\ln(L). \tag{2.17}$$

The GCV score can be taken as an estimate of the mean square prediction error (MSPE) on leave-one-out-cross validation (LOOCV) estimation process. In LOOCV, the model is estimated using all observations except the observation i , and the squared residual predicting observations i from the model is noted. This process is repeated for all observations. However, the GCV score is an efficient measure of this concept that does not actually require fitting all those models and overcomes other issues. It is the score that is minimized when determining the specific nature of the smooth. On its own it does not tell us much, but its usage is similar to AIC as a comparative measure to choose among different models, with lower being better. The GCV score can be defined as Equation 2.18 and for the initial model the GCV score can be found as Equation 2.19, where n is the number of data observations, D is the deviance, s is the scale parameter estimator, p is the number of parametric terms, and edf is the effective degrees of freedom of the model.

$$GCV = \frac{nD}{(n - edf)^2}, \tag{2.18}$$

$$GCV = \frac{ns}{(n - edf - p)^2}. \tag{2.19}$$

2.6 Performance measure

The key point in evaluating the performance of a developed model is the ability to accurately measure its prediction error, $\hat{\varepsilon}_i$. Predicted random value from spatial models is not as easy to compute as that from linear regression model because it involves spatial neighboring effect in the regression. This study will make the evaluation of a method of computing predicted random values by using leave one out cross validation (LOOCV) technique. In this method, for each left observation j of the $n = 586$ observations of the data set, $(n - 1)$ of them are used for model training and the one left out is for testing of the prediction of the jth observation. The model training is performed on the training data set of $(n - 1)$ observations except observation j to estimate the model parameters, $\hat{\beta}$ and $\hat{\lambda}$ or $\hat{\rho}$, which are used to get the predicted random value

$\hat{y}_j$ that is shown in Table 2.2, where w_{ik}, $k \neq j$, is the (i, k)-th element of standardized spatial weight matrix, W, i.e., $\sum_{k \neq j} w_{ik} = 1$, for each j. This process is repeated for j from 1 until n.

Table 2.2: Computation of predicted random values

Model	**Estimation** (**using** $i = 1, 2, \ldots, j-1, j+1, \ldots, n$)	**Prediction** ($j = 1, 2, \ldots, n$)
Linear	$y_i = x_i\beta + \varepsilon_i$	$\hat{y}_j^{LS} = x_j\hat{\beta}$
Spatial error	$y_i = x_i\beta + \varepsilon_i;\ \varepsilon_i = \lambda \sum_{k \neq j} w_{ik}\varepsilon_k + e_i$	$\hat{y}_j^{SE} = x_j\hat{\beta} + \hat{\lambda} \sum_{k \neq j} w_{jk}(y_k - x_k\hat{\beta})$
Spatial lag	$y_i = x_i\beta + \rho \sum_{k \neq j} w_{ik}y_k + \varepsilon_i$	$\hat{y}_j^{SL} = x_j\hat{\beta} + \hat{\rho} \sum_{k \neq j} w_{jk}y_k$

Hence, the prediction error for each model can be measured using Equation (2.20):

$$\hat{\varepsilon}_j = y_j - \hat{y}_j, \ \ j = 1, 2, \cdots, n. \tag{2.20}$$

The performance of established spatial flood models with their prediction by using proposed methods in Chapter 2, 3 and 4 can be compared using two performance indexes which are mean square prediction error ($MSPE$) and mean absolute prediction error ($MAPE$). These indexes are defined as follows:

$$MSPE = \frac{1}{n} \sum_{i=m+1}^{n} \hat{e}_i^2, \tag{2.21}$$

$$MAPE = \frac{1}{n} \sum_{i=m+1}^{n} |\hat{e}_i|. \tag{2.22}$$

2.7 Analysis of exploring the FEH-QMED model

2.7.1 Spatial additive analysis

The analysis started with fitting of the dataset introduced in section 2.2 for the variables used in the FEH-QMED model by using general linear model (2.7) in section 2.5. After that, spatial additive model (2.8) is fitted to the dataset to look for potentially nonlinear effect of each covariate. Hence the estimated smoothing functions from the fitted additive model are plotted with their respective covariates to visualize clearly the nonlinear effect of the four considered covariates (2.6) in this study. This analysis is extended to other available catchment characteristics listed in Table2.1 for the purpose of extracting in detail the potentially nonlinear effect of the covariates in the linear regression model. The performance of the general linear and the spatial additive models could be determined by using model performance criteria such as the generalized Akaike information criterion (AIC) and generalized cross validation (GCV) score.

Table 2.3: Comparison between general linear and additive models

Coefficient	QMED Model	Linear model	Additive model		
Intercept	2.1170*	2.1826*	3.6263*		
			edf	F-value	Prob
log(AREA)	0.8510*	0.8527*	2.561	1218.13	0.000
1000/(SAAR)	-1.8734*	-1.8410*	5.098	208.99	0.000
log(FARL)	3.4451*	3.0335*	4.655	24.13	0.000
$(BFIHOST)^2$	-3.0800*	-3.3479*	1.000	1116.56	0.000
R-Square	0.935	0.935	0.938		
AIC	-	575.898	553.398		
GCV	-	0.154	0.150		

[1] edf = effective degree of freedom

[*] the estimated parameter is significant at 0.000

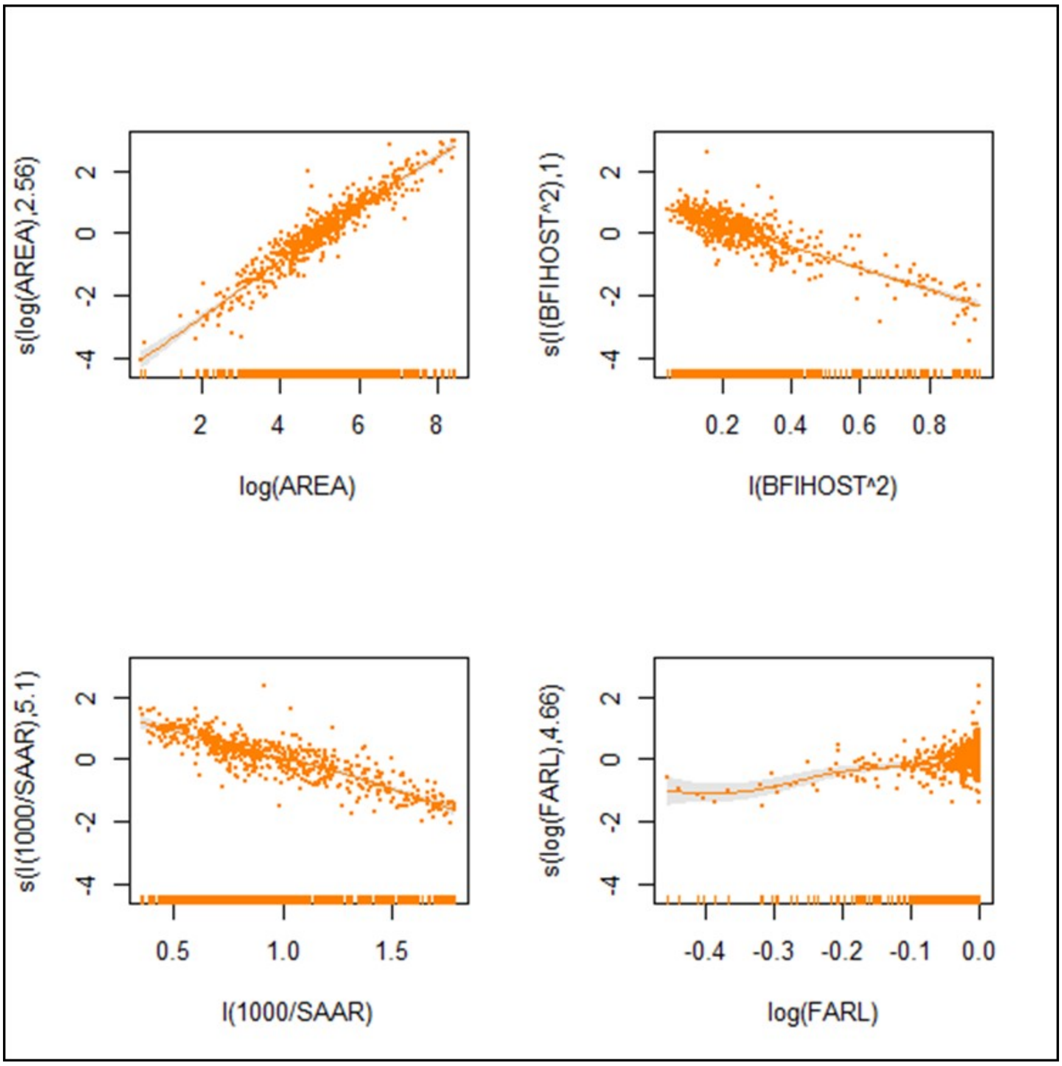


Figure 2.3: Graphical display on effects of four catchment characteristics in additive model

The results from the fitted spatial additive (2.8) and general linear (2.7) for the UK flooding data of 586 gauged stations with $Y = \ln(QMED)$ and four catchment characteristics given in (2.6) are shown in Table 2.3, where the coefficients of the Kjeldsen et al. (2008) model are also provided for a comparison. As mentioned in subsection 2.5.1 the FEH-QMED model (2.4) is represented by linear regression model (2.7). The estimated values from the linear model and the FEH-QMED model shown in Table 2.3 are slightly different because both models were fitted by using different data sets (c.f. Section 2.1 for detail about the reduced number of the gauged stations considered in this study). The comparison between the linear model and the additive model is made. Statistically, it shows the same conclusion in regards to the significance of the individual effects of the covariates in both models. The effective degree of freedom (edf) for variable (BFIHOST)^2 is 1, suggesting that it has essentially a simple linear effect. In terms of model comparison by GCV and AIC it seems safe to conclude that spatial additive model is a better fit than the linear model. Nevertheless, the improvement in performance is just 0.3 percent in terms of the adjusted R-square. Further, from the estimated smoothing functions of the fitted additive model (2.8) with $p = 4$, plotted in Figure 2.3, all four transformed explanatory covariates given in (2.6) seem to be rather linear. This implies that the identified nonlinear effect of the catchment descriptors reported in the FEH-QMED model given in (2.4) seems satisfactorily reliable for the UK flooding data.

Table 2.4: Additive spatial model of 10 catchment covariates

Parametric coefficients:				
	Estimate	**Std. error**	**t value**	**Prob**
(Intercept)	3.62625	0.01492	243	0.00000
Approximate significance of smooth terms:				
	edf	Ref.df	F	p-value
s(log(AREA))	1.000	1.000	346.517	0.00000
s(BFIHOST^2)	3.841	4.870	16.814	0.00000
s(log(DPSBAR))	1.000	1.000	1.16	0.28196
s(log(FARL))	5.925	7.049	22.346	0.00000
s(log(FPEXT))	2.85	3.636	5.132	0.00083
s(log(LDP))	2.033	2.607	1.483	0.21224
s(log(PROPWET))	2.392	3.104	5.303	0.00118
s(log(RMED1D))	1.000	1.000	19.951	0.00001
s(1000/SAAR)	3.623	4.610	10.721	0.00000
s(log(SPRHOST))	7.825	8.610	2.003	0.12804

R-sq.(adj) = 0.938 GCV = 0.15074

[1] edf = effective degree of freedom

[2] Ref.df = Original degree of freedom before deductions for smoothing fits

For detailed investigation in extracting other potentially nonlinear effects of the covariates, more catchment characteristics listed in Table 2.1 are considered as covariates in fitting the spatial additive analysis (2.8). Analysis started with 10 catchment characteristics that are believed to be important covariates in flood risk analysis. The result of this analysis is summarized in Table 2.4. The result clearly shows that model performance is slightly improved by GCV score compare to the model with only four covariates. The result could not also support on the nonlinear effect of covariates due to its small value of edf in all considered catchment characteristics.

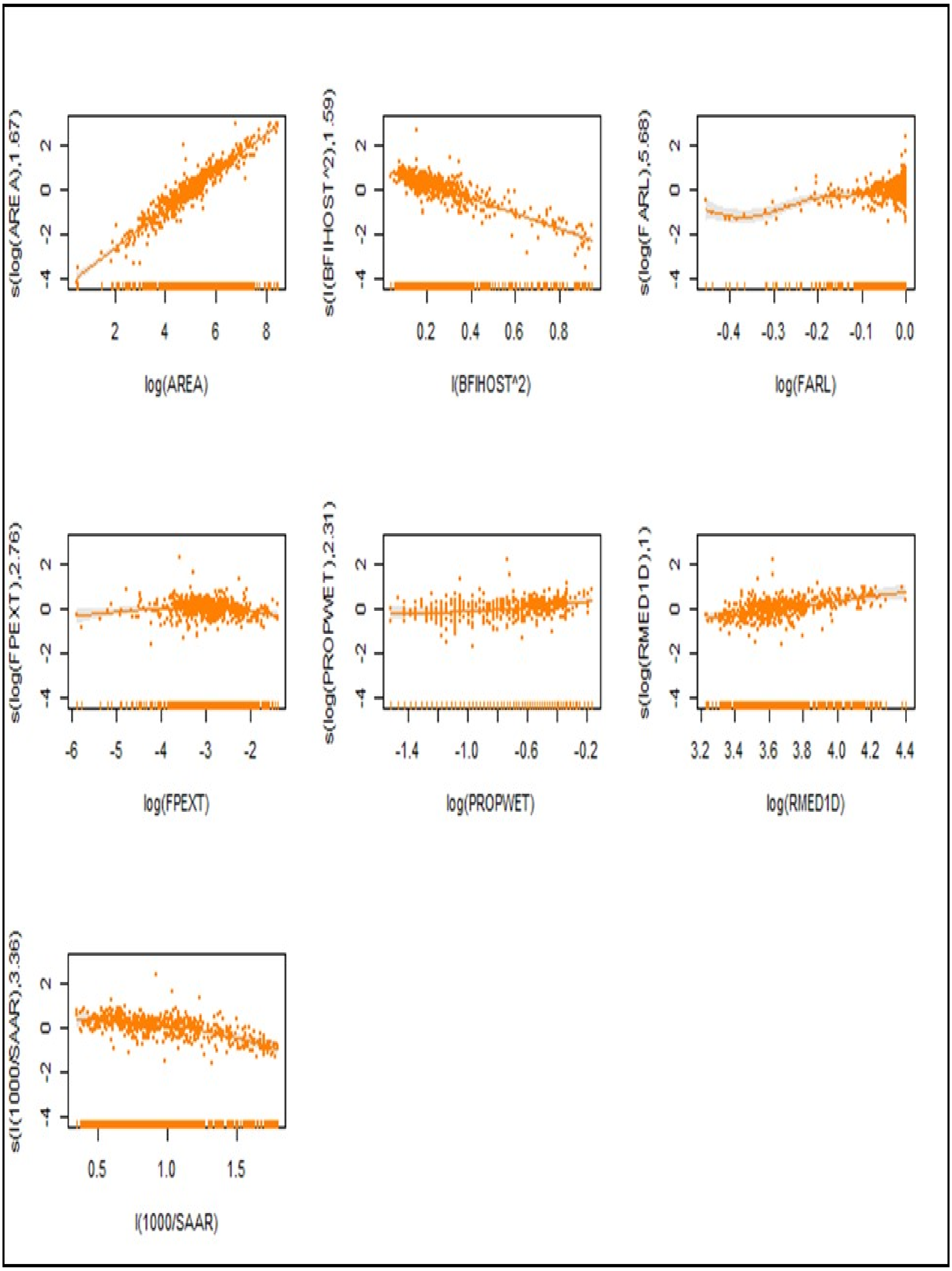
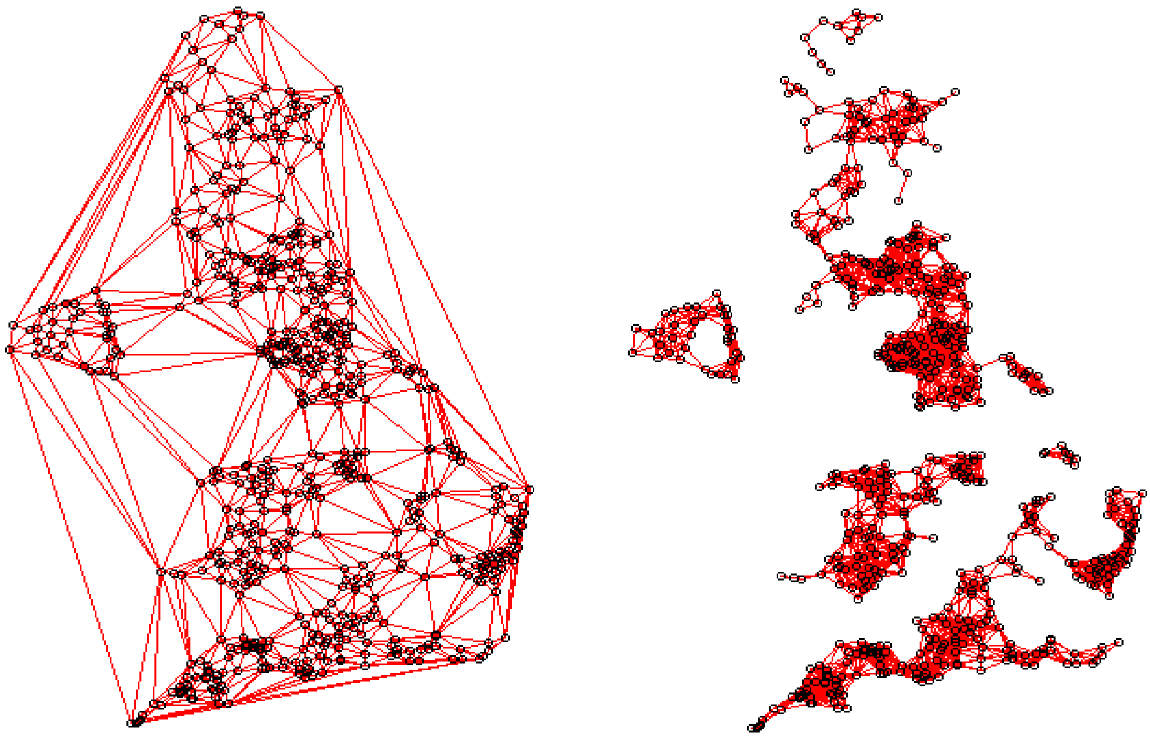


Figure 2.4: Graphical display on effects of seven catchment characteristics in additive model

Further, it is also clearly shown in Table 2.4, 3 covariates have been essentially reduced to a simple linear effect by referring to the value of edf are AREA, DPSBAR and RMED1D. Out of 10 covariates, 3 covariates are not significant to be included in the additive model: DPSBAR, LDP and SPRHOST. Therefore, the spatial additive analysis has been fitted again by using seven significant covariates and estimated smoothing functions of the fitted additive model (2.8) with $p = 7$ plotted in Figure 2.4. From the figure, all seven transformed explanatory covariates, seem to be rather linear. Clearly, it can be concluded that the nonlinear impact of the catchment characteristics identified by the FEH-QMED model appears reliable for the UK flooding data since there are only linear feature detected in the plots of estimated smoothing functions and its covariate in spatial additive analysis.

2.7.2 Test of spatial autocorrelation



(a) Contiguity based matrix (b) Distance based matrix

Figure 2.5: Illustration of spatial relationships in the UK flood peak data

Further investigation has been made in checking the spatial autocorrelation in the residuals of the spatial additive or the FEH-QMED model. This is done by using Moran's I and Lagrange Multiplier tests. This study has considered two types of spatial weight matrices widely used in constructing the spatial relationship among the observations, which are the contiguity based matrix and the distance based matrix. Figure 2.5 gives graphical display of the spatial neighborhoods under (a) contiguity based weights and (b) distance based weights. The background on these spatial weight matrices is provided in Subsection 2.5.2.

Table 2.5 gives the summary of diagnostics tests of spatial dependence in the residuals. Six tests are performed to assess the spatial dependence. First, Moran's I score of 0.213 is highly significant, indicating strong spatial autocorrelation of the residuals. This conclusion is further confirmed by other five tests including the simple Lagrange multiplier test for a missing spatially lag dependent variable, the simple Lagrange multiplier test for error dependence, the robust Lagrange multiplier test for a missing spatially lag dependent variable in the presence of error dependence, the robust Lagrange multiplier test for error dependence in the presence of a missing spatially lag dependent variable, and a portmanteau test. From Table 2.5, it clearly shows that both simple tests of the lag and error are significant, indicating presence of spatial autocorrelation in the model for both types of weight matrices. The robust tests give better understanding in terms of which alternative spatial dependence would be most appropriate for model fitting, showing that the spatial error model appears more acceptable than the lag model.

Table 2.5: Diagnostics for spatial dependence in Linear model

Test	**Contiguity matrix**			**Distance matrix**		
	DF	**Value**	**Prob**	**DF**	**Value**	**Prob**
Moran I (error)	-	0.213	0.0000	-	0.162	0.00000
Lagrange Multiplier (lag)	1	20.961	0.0000	1	19.563	0.00001
Lagrange Multiplier (error)	1	77.196	0.0000	1	76.611	0.00000
Robust Lagrange Multiplier (lag)	1	3.306	0.0690	1	4.672	0.03065
Robust Lagrange Multiplier (error)	1	59.541	0.0000	1	64.721	0.00000
Lagrange Multiplier (SARMA)	2	80.502	0.0000	2	84.284	0.00000

2.7.3 Spatial autoregressive flooding analysis

Based on the identified features in Subsections 2.7.1 and 2.7.2, further analysis is made by fitting the spatial error and spatial lag models introduced in Subsection 2.5.2 to the UK flooding data by using both types of spatial weight matrices considered in this study. The results are summarized in Table 2.6, which gives the estimated values for each of the parameters in all models considered in this analysis, where the value in the round bracket represents the standard error of the estimate. All estimated parameters of the regressive coefficients have also been found to be significant for each model. In addition, the spatial coefficients λ and ρ are significant in both the spatial error and the spatial lag models, given in (2.15) and (2.16), respectively, for both types of spatial weight matrices, demonstrating that the spatial relationship in the UK flood peak data seems to be well explained by both contiguity based matrix and distance based matrix. As before, the linear FEH-QMED model is compared with both spatial error and spatial lag models that have been fitted by using two types of spatial weight matrices. By using the AIC value, it is clearly shown that using the spatial dependence can improve the performance of the FEH-QMED model. It is also found that the spatial error model performs better than the spatial lag model, with use of the contiguity weight matrix the best with the lowest AIC value.

Table 2.6: Comparison between general linear and Spatial autoregressive models

Coefficients	QMED	Linear	Contiguity matrix		Distance matrix	
			SpErr	SpLag	SpErr	SpLag
Intercept	2.1170	2.1826	2.1494	1.6800	2.1430	1.6576
		(0.0848)	(0.0911)	(0.1385)	(0.1024)	(0.1454)
log(AREA)	0.8510	0.8527	0.8497	0.8320	0.8472	0.8327
		(0.0131)	(0.0129)	(0.0137)	(0.0130)	(0.0137)
1000/(SAAR)	-1.8734	-1.8410	-1.8181	-1.6554	-1.7923	-1.6499
		(0.0523)	(0.0748)	(0.0660)	(0.0796)	(0.0674)
log(FARL)	3.4451	3.0335	2.9526	2.8210	3.1500	2.9076
		(0.2629)	(0.2725)	(0.2617)	(0.2627)	(0.2596)
$(BFIHOST)^2$	-3.0800	-3.3479	-3.2655	-3.1228	-3.2959	-3.1251
		(0.0969)	(0.1144)	(0.1077)	(0.1169)	(0.1087)
Lambda			0.4351		0.5001	
			(0.0546)		(0.0599)	
Rho				0.0975		0.1007
				(0.0220)		(0.0232)
AIC		575.90	519.34	558.03	525.78	559.49

The values in the () represent as standard error for each estimated parameters
[1] SpErr = Spatial error model
[2] SpLag = Spatial lag model

Table 2.7: Diagnostics for spatial dependence in residuals of spatial autoregresive models

SPATIAL ECONOMETRIC MODEL	MORAN-I TEST			
	Contiguity matrix		Distance matrix	
	Value	Prob	Value	Prob
Spatial lag model	0.1600	0.0000	0.1230	0.0000
Spatial error model	-0.0114	0.6889	-0.0029	0.9459

Moran's I test is also examined for the residuals in both the fitted spatial error and spatial lag models to investigate whether these spatial flooding models are sufficiently acceptable. The results in Table 2.7 clearly show that spatial autocorrelation in the spatial lag model is still existent but not in the spatial error model. This indicate that by allowing the FEH-QMED model error terms modeled by a spatial error model it does not only improve the model fit but also makes the spatial effect disappear. This further demonstrates that the spatial error model is more suitable for the UK flooding data compared to the spatial lag model and other models in Table 2.6.

2.7.4 Model evaluation

Model evaluation is important in assessing the quality and accuracy of the developed models for the purpose of practical usage in prediction. In particular, we are comparing the established spatial flooding models above with the earlier updated FEH-QMED model of Kjeldsen and Jones (2010) based on a revised data-transfer method that partially takes account of spatial correlation.

To evaluate the different models in this study, the leave one out cross validation method given in Section 2.6 is applied. Two types of spatial neighboring relationship, i.e., contiguity based and distance based matrices, are used in evaluating model performance in prediction. The prediction for the updated FEH-QMED model is performed just using the given models in Kjeldsen and Jones (2010). The results of the mean squared prediction error (MSPE) for different models are summaries in Table 2.8. Here the percentage of the improvement of the established spatial autoregressive flooding models relative to the updated FEH-QMED model of Kjeldsen and Jones (2010) is also reported. It is clearly shown that both proposed spatial flooding models outperform the benchmark model from the prediction perspective. The results also give a further hand to support the finding in the above that spatial error model is the best estimation model for the UK index flood based on the AIC score, and the spatial weight matrix based on contiguity is preferred than the distance based, with 13.8 percent of improvement over the updated FEH-QMED model in reducing the MSPE.

Table 2.8: Comparison of model performance using LOOCV method

Model	**Spatial weight**	**MSPE**	**Improvement**
FEH-QMED	-	0.15859	-
Spatial error	Contiguity	0.13671	13.8 percent
	Distance	0.13879	12.5 percent
Spatial lag	Contiguity	0.15094	4.8 percent
	Distance	0.15139	4.5 percent

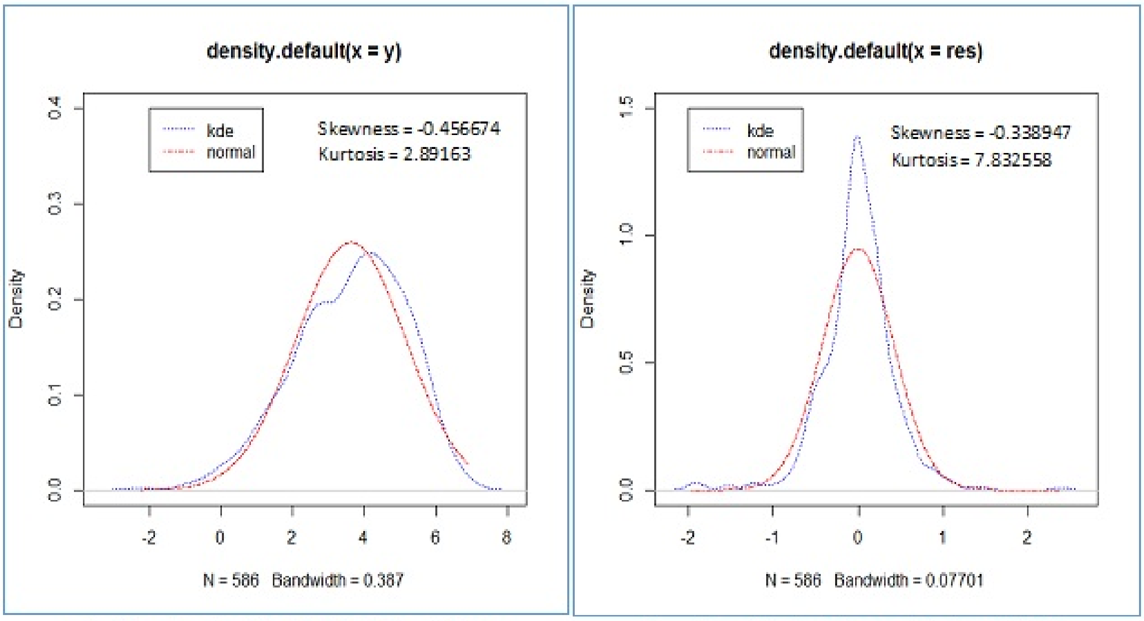
2.8 Discussion and Conclusion

2.8.1 Discussion

Results in subsection 2.7.3 and 2.7.4 allow us to conclude that, an improved model for estimation of index flood in the UK is spatial error model. This is based on the AIC criteria which was shown with the largest percentage of improvement in mean squared prediction error in comparison with FEH-QMED model that is represented by linear regression model. Furthermore, diagnostics Moran's I test has shown that spatial autocorrelation has been disappeared by allowing the error terms to be spatially correlated in the FEH-QMED model. The improved index flood estimation model for the UK flooding data named as spatial index flood estimation (SIFE) model can be expressed as

$$\begin{aligned} \ln(QMED_i) &= 2.1494 + 0.8497\ln(AREA_i) - 1.8181(1000SAAR_i^{-1}) + \\ &\quad 2.9526\ln(FARL_i) - 3.2655(BFIHOST_i)^2 + e_i, \\ e &= 0.4351We, \end{aligned} \tag{2.23}$$

where $e = (e_1, e_2, \cdots, e_n)'$ and W is a $n \times n$ standardized contiguity based spatial weight matrix, with model error terms, $\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)'$ where ε_i beung i.i.d. of skewed normal distribution .



a) Density plot for the UK flood data b) Density plot for residuals of SEM

Figure 2.6: Estimated distribution for the UK flood data

The SIFE model (2.23) is the spatial error model based on Gaussian assumption for the error terms. Further investigating was done to the residuals of the Equation (2.23) to check the validity of the Gaussian assumption in the SIFE model. Kernel density estimation was used to estimate the distribution of the residuals and also the UK flooding data. For more clear visualization, normal distribution with same mean and variance of residuals and the UK flooding data are plotted in the same diagram. Figure 2.6 (a) and (b) give the kernel density plots for the UK flooding data and residuals of spatial error model with their skewness and kurtosis value reported together respectively. The results indicated that we should to consider a kind of skewed distribution in the UK flooding data and residuals of spatial error model. Therefore, this findings lead us to the current research in developing new statistical index flood estimation model that need to take into account spatial dependency and non-Gaussianity.

2.8.2 Conclusion

Flood estimation for ungauged catchments is a challenging task for hydrologists. With modern geographical information system and advanced computational tools, many hydrological models were developed by active researchers in hydrological area. However, there are urgent needs for efficient statistical method to extract unknown and unexpected complex features from data sets, as well as check the reliable of the developed models. For this purposes a well-established index flood model is needed as a study case with the related available dataset. As such, the index flood regression model known as QMED model from Flood Estimation Handbook (FEH) is chosen as the benchmark model for this study due to its worldwide recognition and for being the basis of most studies on flood risk estimation and modeling. Apart from that the UK is well equipped with updated flood information and dataset in conjunction with many gauged stations and this is very important for the empirical work of this project. This study is not intended to develop a new QMED model to replace the existing FEH-QMED model. Instead, this study attempted to explore unknown and unexpected features in the UK flood peak dataset that might be overlooked and appropriated for the FEH-QMED model . This is very important for the accuracy improvement in the flood estimation model. FEH study provides us a high-quality QMED model, good-quality flood peak data and excellent catchment descriptors CD-ROM. However, the detailed investigating and exploring on the FEH-QMED model using advanced statistical method is also needed because the findings will be very useful to fill in the knowledge gap in hydrology field through statistical perspective.

Several findings that have been found through empirical works using the UK flood data in exploring FEH-QMED model are;

1. The nonlinear impacts of explanatory catchment characteristics variables identified by FEH-QMED model appears reliable the UK flooding data through spatial additive analysis.

2. To further improved the FEH-QMED model, spatial neighboring effect was taken into account through spatial error analysis, since there are spatial autocorrelation is detected in the FEH-QMED model.

3. Spatial error model is an improved FEH-QMED model based on the AIC criteria which was shown with the largest percentage of improvement in mean squared prediction error in comparison with the linear regression model.

4. The UK flooding data are more suitable with spatial error model rather than spatial lag model because by allowing the error terms to be spatially correlated could make spatial effect to not exist in the index flood estimation model.

5. The UK flooding data and the residuals of spatial error model were found to be skewed which was identified by using kernel density estimation analysis.

These empirical findings have motivated us to consider further research in developing more effective statistical model for index flood estimation by taking into account spatial dependence and non-Gaussianity. Actually, if developing the model for such data sets relies on Gaussian assumption, then biased estimate would be produced. Further research will therefore be pursued in Chapter 3 which could overcome the potential problem.

Chapter 3

Non-Gaussian Spatial Index Flood Estimation Model

3.1 Introduction

The Gaussian distribution is widely used in describing data in many application. In the hydrological application analysis such as the Regional Flood Frequency Analysis (RFFA), Gaussian assumption is mostly used in estimating the index flood by using regression models (NERC, 1975; Canuti and Moisello, 1982; Acreman, 1985; Mimikou and ordios, 1989; Garde and Kothyari, 1990; Reimers, 1990; Shu and Burn, 2004; Kjeldsen and Jones, 2007; Wan Jaafar et al., 2011). Nevertheless, the celebrated Gaussian distribution is no longer suitable in environmental, hydrological and ecological studies since the observed spatial variables in those fields are known to have skewed distributions (Zhang and El-Shaarawi, 2010). In such situation, a Gaussian assumption is too ideal to be used in developing a model and could provide a biased estimate. Furthermore, large amount of the uncertainty contributed by the index flood estimation in the RFFA is still an open problem among hydrologists. Index flood estimation is one of the main steps in the flood regionalization procedure of RFFA by using index flood method. Accurate estimates of the index flood from index flood regression models are essential for precise prediction of flood design at any locations of ungauged catchments area. Therefore, this chapter attempts to respond to the both aforementioned issues by proposing a more reliable index flood estimation model that takes into account the spatial dependence and non-Gaussianity. This new model, perhaps can improve existing methodology in the index flood regression model from statistical point of view that may not be noticeable to hydrologists.

Empirical work in Chapter 2 has found that spatial index flood estimation (SIFE) model which has taken into account the spatial dependency is an improved model in estimating the UK index flood. SIFE models have been established by fitting the considered spatial autoregressive models which are spatial lag and spatial error models into the UK flooding data with two types of spatial relationships i.e. contiguity and distance based weight matrices. The best SIFE model established in Chapter 2 that shown in Equation 2.23 has been developed by using spatial error model with contiguity weight matrix. This best SIFE model was selected based on the AIC criteria which was shown with the largest percentage of improvement in mean squared prediction error (MSPE) in comparison with the updated FEH-QMED in Kjeldsen and Jones (2010). However, this best SIFE model is based on a Gaussian assumption of the error terms. The results from kernel density estimation have shown that the UK flooding data and the residuals of the best SIFE model could have a kind of skewed distribution. This has motivated us to the further research through this chapter to improve the SIFE models by proposing a new statistical model structure for spatial error model with non-Gaussian of error terms. This research is relevant to Zhang and El-Shaarawi (2010) who has claimed that the celebrated Gaussian distribution is no longer suitable in environmental, hydrological and ecological studies since the observed spatial variables in those fields are well-known to have skewed distributions. Furthermore, the assumption in the distribution for the data or random variable in drawing statistical inference under a reasonable parametric framework is important for parameter estimation.

A relatively complete treatment for the spatial autoregressive models from the perspective of maximum likelihood with Gaussian assumption can be found in Anselin (1988), Chapter 6, see also LeSage (2008) and Arbia (2014). The fact that asymmetry in data makes the Gaussian assumption not valid, has called for spatial error model for skewed UK flooding data. Thus, a skew-normal distribution of Azzalini (1985) is suggested for the distribution of the UK flooding data under study which takes normal distribution as a special case. This distribution is one of many near normal distributions that have been proposed by researchers including Mudholkar and Hutson (2000), Turner (1960) and Prentice (1975). Skew-normal distribution is mathematically tractable, which can deals with the deviations from symmetry in the normal distribution and accommodating reasonable values of skewness and kurtosis.

The skew-normal distribution has desirable properties which has asymmetry parameter representing the skewness of real data situations. However, this leads to complexity and challenges in drawing inference on the parameters. This issue has been discussed by Sartori (2006), Dalla Valle (2004), Pewsey (2000), Monti (2003) among others. On the other hand, Azzalini and Capitanio (1999), Liseo and Loperfido (2003), Monti (2003) and Sartori (2006) have proposed some different methods in overcoming the estimation issues. Latest, Dey (2010) has proposed simulation method for maximum likelihood estimation in drawing the inference of the parameters of the skew-normal distribution based on a simple linear regression. Therefore, a suggested computational procedure will be developed for the maximum likelihood estimation under spatial error model with the skew-normal distribution of error terms, where the simple linear regression of Dey (2010) can be seen as a special case of our proposed models.

This chapter will;

(i). propose a new statistical structure for spatial error model with a skewed normal distribution for residuals to improve the efficiency of the established Gaussian SIFE model in Chapter 2.

(ii). suggest a maximum likelihood computational algorithm procedure to solve the proposed model in (i).

(iii). present the simulation to investigate the finite sample performance of the proposed statistical method in (i) by applying the computational algorithm procedure that has been developed in (ii).

(iv). evaluate the performance of the non-Gaussian SIFE models that have been established from the proposed statistical method in (i) by using leave one out cross validation technique together with computational algorithm procedure suggested in (ii).

Perhaps, by completing the aforementioned tasks in this chapter, the second objective of this research study which is to develop a more efficient statistical model in estimating the index flood that better fit the UK flooding data can be accomplished.

3.2 A spatial error model from normal to skew normal for error terms

A spatial error model that was selected is introduced in Chapter 2 and can be expressed as,

$$y = X\beta + u, \tag{3.1}$$

$$u = \lambda W u + \varepsilon,$$

$$\varepsilon \sim N(0, \sigma^2 I_n).$$

The $n \times 1$ vector y contains the dependent variables and X represents the usual $n \times k$ data matrix containing explanatory variables. W is a spatial weight matrix and the parameter λ is a coefficient on the spatially correlated errors, analogous to the serial correlation problem in time series models. The parameter β reflects the influence of the explanatory variables on variation in the dependent variable y. Since the UK flooding data used for this study has been found to be skewed, the normality of error term in equation (3.1) is not an adequate modeling assumption anymore. Due to that, skewed normal distribution has been proposed as a candidate distribution for the UK flooding data.

From equation (3.1), let $\varepsilon = \sigma e$, then $e_i = (e_1, \cdots, e_n)^T \sim N(0, I_n)$. Now, we want to replace $e_i \sim N(0,1)$ by $e_i \sim$ a skewed normal, $SN(\xi_\alpha, \omega_\alpha, \alpha)$ with mean zero, $E(e) = 0$ and variance one, $Var(e) = 1$. How can we specify a skewed normal distribution with appropriate ξ_α and ω_α such that $E(e) = 0$ and $Var(e) = 1$? We first introduce skewed normal distribution in section 3.2.1.

3.2.1 Skewed normal distribution

According to Azzalini (1985), a random variable Z is said to have a skew-normal distribution with notation $Z \sim SN(0,1,\alpha)$ if its probability density function is given by

$$Z \sim f_z(z) = \phi(z;0,1,\alpha) = 2\phi(z)\Phi(\alpha z); \qquad -\infty < x < \infty, \tag{3.2}$$

where α, a real number, is the skewness parameter, meanwhile, ϕ and Φ are respectively the density and cumulative distribution function of the standard normal distribution. As is well known, for the special case which is $\alpha = 0$ the density function of ordinary normal distribution follows. Therefore, the k-th moment of a random variable Z can be computed directly from the integral

$$E[Z^k] = \int_{-\infty}^{\infty} z^k 2\phi(z)\Phi(\alpha z)dz.$$

From the results of the first and second moment, the $E(Z)$ and Var(Z) could be determine as follows:

$$\mu_z = E[Z] = \frac{\sqrt{2}\alpha}{\sqrt{\pi}\sqrt{1+\alpha^2}}, \tag{3.3}$$

$$\sigma_z^2 = Var[Z] = 1 - \frac{2\alpha^2}{\pi(1+\alpha^2)}. \tag{3.4}$$

Further, with $E(Z)$ and Var(Z) the skewness (γ_1) and kurtosis (γ_2) for $Z \sim SN(0,1,\alpha)$ could be defined as follows:

$$\gamma_1 = \frac{4-\pi}{2}\frac{\left(\frac{\sqrt{2}\alpha}{\sqrt{\pi}\sqrt{1+\alpha^2}}\right)^3}{\left(1-\frac{2\alpha^2}{\pi(1+\alpha^2)}\right)^{\frac{3}{2}}}, \tag{3.5}$$

$$\gamma_2 = 2(\pi-3)\frac{\left(\frac{\sqrt{2}\alpha}{\sqrt{\pi}\sqrt{1+\alpha^2}}\right)^4}{\left(1-\frac{2\alpha^2}{\pi(1+\alpha^2)}\right)^{2}}. \tag{3.6}$$

3.2.2 Spatial error model with skewed normal distribution

In this section, we propose a spatial error model with skewed normal distribution. From equation (3.1) spatial error model with skewed normal distribution can be expressed as

$$y = X\beta + u, \tag{3.7}$$

$$u = \lambda W u + \varepsilon,$$

$$\varepsilon = \sigma e,$$

where $e_i = (e_1, \cdots, e_n)^T \sim SN(\xi_\alpha, \omega_\alpha, \alpha)$ such that $E(e_i) = 0$ and $Var(e_i) = 1$. By using some common work with location and scale transformation in $Z_i \sim SN(0, 1, \alpha)$, e_i can be determined by

$$e_i = \xi_\alpha + \omega_\alpha Z_i, \tag{3.8}$$

where the expectation and variance of e_i are given by

$$E(e_i) = \xi_\alpha + \omega_\alpha E(Z) = \xi_\alpha + \omega_\alpha \mu_z,$$

$$Var(e_i) = \omega_\alpha^2 Var(Z) = \omega_\alpha^2 \sigma_z^2.$$

In order that $Var(e_i) = 1$, we let $\omega_\alpha = \dfrac{1}{\sigma_z}$; meanwhile set $\xi_\alpha = -\dfrac{\mu_z}{\sigma_z}$ such that $E(e_i) = 0$. Therefore, the distribution of e_i can be denoted as

$$e_i \sim SN\Big(\xi_\alpha, \omega_\alpha, \alpha\Big),$$

where

$$\xi_\alpha = -\frac{\mu_z}{\sigma_z} = -\frac{\sqrt{2}\alpha}{\sqrt{\pi + (\pi - 2)\alpha^2}},$$

$$\omega_\alpha = \frac{1}{\sigma_z} = \sqrt{\frac{\pi + (\pi - 2)\alpha^2}{\pi(1 + \alpha^2)}}.$$

Hence, by using the transformation rule, the density for $e_i \sim SN(\xi_\alpha, \omega_\alpha, \alpha)$ can be determined following from (3.8) by

$$\begin{aligned} f_e(e) &= f_z\Big(\frac{e - \xi_\alpha}{\omega_\alpha}\frac{1}{\omega_\alpha}\Big) \\ &= f_z\left(\frac{e + \dfrac{\mu_z}{\sigma_z}}{\dfrac{1}{\sigma_z}}\right)\frac{1}{\dfrac{1}{\sigma_z}} \\ &= f_z(\mu_z + \sigma_z e)\sigma_z, \end{aligned}$$

where $f_z(\cdot)$ is defined in (3.2). Therefore

$$e_i \sim f_e(e) = 2\phi(\mu_z + \sigma_z e)\Phi(\alpha(\mu_z + \sigma_z e))\sigma_z. \tag{3.9}$$

3.3 Maximum likelihood estimation

3.3.1 The likelihood

From equation (3.1) spatial error model could be expressed in a reduce form as shown in the equation (3.10)

$$\begin{aligned} y &= X\beta + \lambda W u + \varepsilon \\ y - X\beta &= \lambda W(y - X\beta) + \varepsilon \\ y - \lambda W y &= X\beta - \lambda W X\beta + \varepsilon \\ (I - \lambda W)y &= (I - \lambda W)X\beta + \varepsilon, \end{aligned} \tag{3.10}$$

while for data generation process form as in equation (3.11)

$$\begin{aligned} y &= (I - \lambda W)^{-1}(I - \lambda W)X\beta + (I - \lambda W)^{-1}\varepsilon \\ y &= X\beta + (I - \lambda W)^{-1}\varepsilon. \end{aligned} \tag{3.11}$$

Suppose $\varepsilon = \sigma e$, then the equation (3.11) could be denoted as

$$y - X\beta = (I - \lambda W)^{-1}\sigma e. \tag{3.12}$$

From equation (3.12), we could express the error term, e in the term of regression residual, u and determine the determinant of the transformation's Jacobian.

$$u = (I - \lambda W)^{-1}\sigma e,$$

$$e = \frac{1}{\sigma}(I - \lambda W)u,$$

$$|\frac{\delta e}{\delta u}| = |\frac{1}{\sigma}(I - \lambda W)| = \frac{1}{\sigma^n}|I - \lambda W|.$$

Then, by the independence of $e_i \sim SN(\xi_\alpha, \omega_\alpha, \alpha)$ the joint distribution of $e_i = (e_1, \cdots, e_n)^T$ is defined as the product of the probability density functions

$$f_e(e_1, \cdots, e_n) = \prod_{i=1}^{n} f_{e_i}(e_i). \tag{3.13}$$

By using the rule of transformation for probability density function the density of regression residuals u, can be determined by

$$\begin{aligned} f_u(u) &= f_e(e)|\frac{\delta e}{\delta u}| \\ &= f_e\Big[\frac{1}{\sigma}(I-\lambda W)u\Big]|\frac{\delta e}{\delta u}| \\ &= f_e\Big[\frac{1}{\sigma}(I-\lambda W)u\Big]\frac{1}{\sigma^n}|I-\lambda W|. \end{aligned} \tag{3.14}$$

Similarly, the density for $y=(y_1,\cdots,y_n)^T$ can also be computed as follow

$$\begin{aligned} f_y(y) &= f_u(y-X\beta) \\ &= f_e\Big[\frac{1}{\sigma}(I-\lambda W)(y-X\beta)\Big]\frac{1}{\sigma^n}|I-\lambda W|. \end{aligned} \tag{3.15}$$

Hence, the likelihood function for the joint vector of observation y could be obtained by using equation (3.15) with joint skew-normal distribution of error term e

$$\begin{aligned} f_y(y_i,\cdots,y_n) &= \prod_{i=1}^n f_{e_i}\Big[\frac{1}{\sigma}(I-\lambda W)_i(y-X\beta)\Big]\frac{1}{\sigma^n}|I-\lambda W| \\ &= \Bigg[\prod_{i=1}^n 2\phi\Big[\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big]\Phi\Big[\alpha\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\Big]\sigma_z\Bigg] \\ &\quad \frac{1}{\sigma^n}|I-\lambda W|, \end{aligned} \tag{3.16}$$

where $(I-\lambda W)_i$ stands for the i-th row of $(I-\lambda W)$. Then, the log-likelihood is as follows

$$\begin{aligned} \ell \;=\;& n\log 2+n\log\sigma_z+\log|I-\lambda W|-n\log\sigma \\ &+\sum_{i=1}^n \log\phi\Big[\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big] \\ &+\sum_{i=1}^n \log\Phi\Big[\alpha\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\Big]. \end{aligned} \tag{3.17}$$

This likelihood ℓ is a complex function of the parameters $\theta=(\beta,\lambda,\sigma,\alpha)^T$. We shall suggest a computational algorithm for the maximum likelihood estimation.

3.3.2 A suggested computational procedure

Maximum likelihood principle in section 3.3.1 gives complex likelihood function. With the complex likelihood function we cannot find the maximum likelihood estimators by solving the equations in a closed form. We have to rely on computer software to compute maximum likelihood estimators numerically using efficient algorithm method. Therefore, in this subsection, we are developing a computational algorithm procedure to determine the values of unknown parameters $\hat{\theta} = (\hat{\underset{\sim}{\beta}}, \hat{\lambda}, \hat{\sigma}, \hat{\alpha})$ i.e., a maximum likelihood estimators that maximize

$$\begin{aligned}\ell = n\log 2 + n\log\sigma_z + \log|I-\lambda W| - n\log\sigma + \\ \sum_{i=1}^{n}\log\phi\Big[\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big] + \\ \sum_{i=1}^{n}\log\Phi\Big[\alpha\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\Big],\end{aligned}$$

$$\text{where } \mu_z = \frac{\sqrt{2}\alpha}{\sqrt{\pi}\sqrt{1+\alpha^2}} \text{ and } \sigma_z = \sqrt{1-\frac{2\alpha^2}{\pi(1+\alpha^2)}}.$$

Algorithm 1, gives the suggested maximum likelihood computational procedure under skewed spatial error model.

Algorithm 1

(i). Fit the spatial error model into the UK flooding data using `errorsarlm` function from `spdep` package in `R` (Bivand et al., 2015) to determine the estimated values of $\hat{\underset{\sim}{\beta}}, \hat{\lambda}$ and $\hat{\sigma}$.

(ii). Develop a profile log-likelihood function for α by substituting the estimated values of $\hat{\underset{\sim}{\beta}}, \hat{\lambda}$ and $\hat{\sigma}$ from (i) into log-likelihood function 3.17.

(iii). Determine the estimated value of alpha, $\hat{\alpha}$. This can be solve by using `optimize` function in `R` (Team, 2014) to numerically maximize the profile log-likelihood that has been developed in (ii).

(iv). Determine the solution for MLE by using `optim` function in `R` (Team, 2014). We could well use this function by giving the initial guess for each unknown parameters that need to be determined; different initial guesses may return different solutions. For this purpose we use $\hat{\underset{\sim}{\beta}}, \hat{\lambda}, \hat{\sigma}$ and $\hat{\alpha}$ determined in (i) and (iii) as a good initial values to numerically maximizes the log-likelihood function (3.17) in order to get the maximum likelihood estimators $\hat{\theta} = (\hat{\underset{\sim}{\beta}}, \hat{\lambda}, \hat{\sigma}, \hat{\alpha})$.

3.4 Properties of estimators

A suggested maximum likelihood computational procedure under skewed spatial error model detailed in Algorithm 1 is developed with the purpose to be applied into the UK flooding data in establishing the efficient index flood estimation model. However, the performance of the proposed estimation procedure need to be investigated beforehand. Two important properties of maximum likelihood estimators that need to be examined are on unbiasedness and consistency of the estimators.

The unbiasedness and consistency of the maximum likelihood estimators under the proposed statistical method suggested in this chapter can be examined through the fourth step of suggested computational procedure (Algorithm 1) in Subsection 3.3.2 using simulations. A finite number of simulated sample data is drawn based on data generation process (DGP) defined by equation (3.11) that characterizes the population from skewed normal distribution where the model residuals are independently and identically distributed by skewed normal distribution, $e_i = (e_1, \cdots, e_n)^T \sim SN(\xi_\alpha, \omega_\alpha, \alpha)$, can refer 3.2 for details. By fixing the values of parameters β, λ, σ and α as true values and the skewed normal of error terms are generated from `rskewnorm` function in `R` from `VGAM` package (Yee and Yee, 2018), we generate the finite number of samples that then can be examined one by one through repeating simulation. Let the number of simulations is denoted by $S = (1, 2, \cdots, s)$ and θ be a parameter of interest for which the inference need to be made. With the first simulated sample of observation $Y^{(1)} = \{y_1^{(1)}, y_2^{(1)}, \cdots, y_n^{(1)}\}$ together with our propose technique, an estimate of the parameter, $\hat{\theta}(Y^{(1)})$ can be obtained. Similarly, with another simulated sample $Y^{(s)}$ another estimate, $\hat{\theta}(Y^{(s)})$ can be obtained. This make $\hat{\theta}(Y)$ a realization of a random variable which can be denoted a s $\hat{\theta}$. This random variable is called estimator. The standard error of the estimated parameter, $\hat{\theta}(Y)$ is equal to standard deviation for the random variable $\hat{\theta}$. The estimators are asymptotically unbiased if they tend to be centered very near to the true values that specified in the DGP. And this unbiased estimator are consistent which means, they convergence to the true parameter values when their standard error decreases to zero as the sample size, n increases.

3.5 Application of skewed spatial error model in the UK index flood estimation

3.5.1 Evaluation on the performance of the estimators

Table 3.1: The five summary of simulated distribution for $\hat{\theta}$ and its bias

$\hat{\theta}$	**True**	**Min.**	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max.**	**Bias**
$\hat{\beta}_0$	2.000	1.811	1.958	1.999	2.002	2.046	2.280	0.002
$\hat{\beta}_1$	1.000	0.974	0.994	1.000	1.000	1.006	1.038	0.000
$\hat{\beta}_2$	-2.000	-2.167	-2.035	-2.001	-2.000	-1.964	-1.862	0.000
$\hat{\beta}_3$	3.000	2.332	2.881	3.006	3.002	3.143	3.514	0.002
$\hat{\beta}_4$	-3.000	-3.311	-3.058	-3.000	-3.001	-2.940	-2.743	-0.001
$\hat{\lambda}$	0.400	0.210	0.348	0.389	0.388	0.429	0.579	-0.012
$\hat{\sigma}$	0.300	0.270	0.294	0.299	0.299	0.305	0.327	-0.001
$\hat{\alpha}$	0.800	-1.220	0.646	0.811	0.795	0.963	1.642	-0.005

The simulations of the performance for the proposed skewed spatial error model need to be investigated before it can be applied into the UK flooding data to develop Non-Gaussian SIFE models. Two basic properties of statistical estimator which are unbiasedness and consistency for this proposed statistical method have been examined by using simulation procedure mentioned in subsection 3.4.

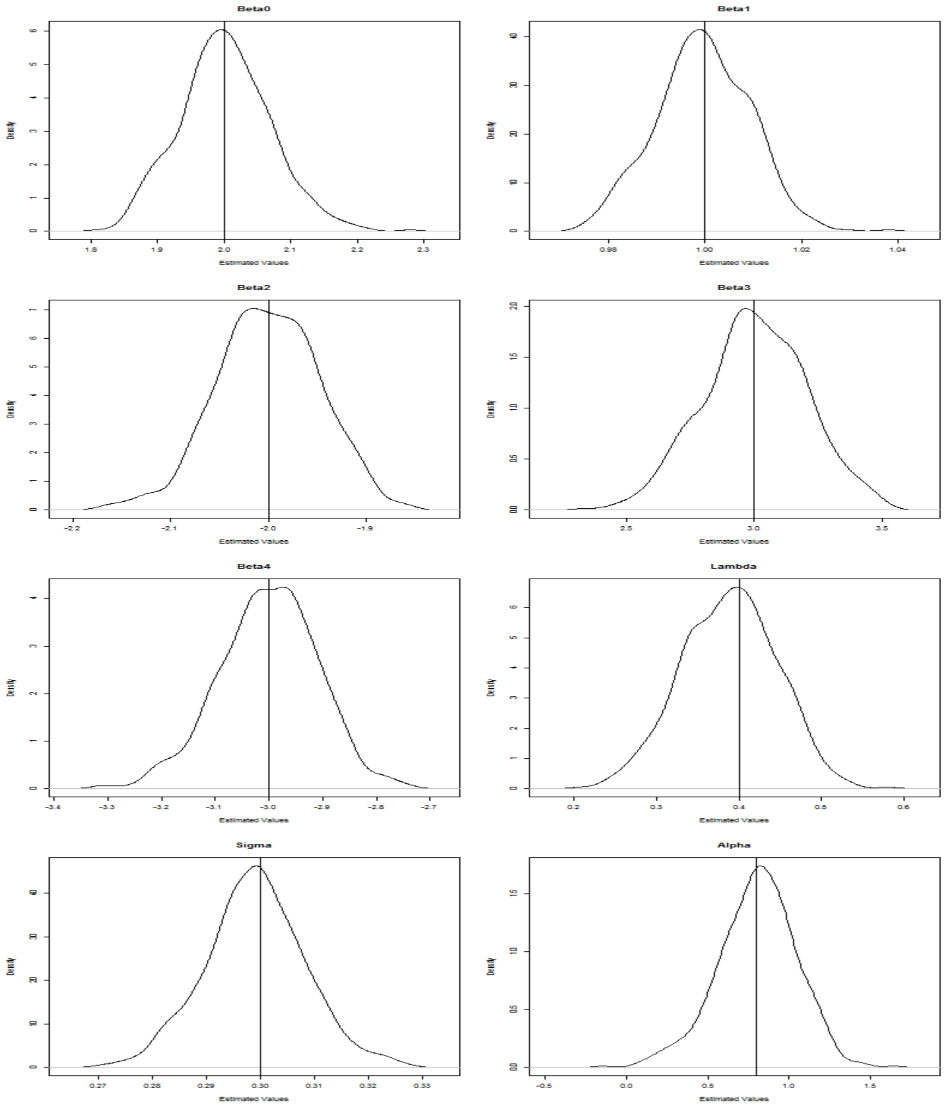
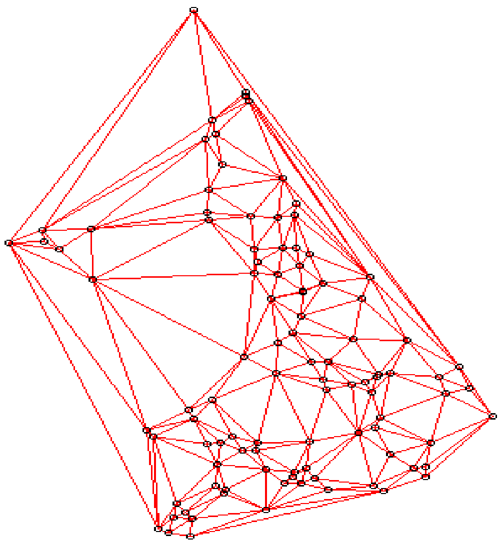


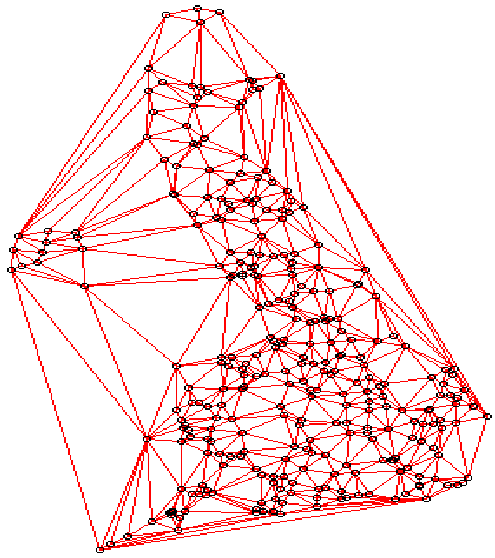
Figure 3.1: Simulated distribution for estimated values of unknown parameters, $\hat{\theta}$

In evaluating the unbiasedness of our estimators, we used catchment characteristics data of 586 observations in the original sample (see Section 2.1) as explanatory variables to generate random variable describe in data generation process (DGP) (3.11), by fixing a true value for each unknown parameters and the error terms have been randomly generated using `rskewnorm` function from `VGAM` package in `R` (Yee and Yee, 2018). Then, this simulated sample data of 586 observations are analyzed by using step (iv) described in Algorithm 1 to obtain an estimate of each parameters. Both processes are repeated 500 times through simulation, yielding a realization of the estimators. This realization then can be analyzed to examine the unbiasedness of the estimators. Several different combinations of true value for the parameters $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \lambda, \sigma$ and α have been used in data generation proses using Equation (3.11), but we present here the outcomes for just one set of parameter values which is when

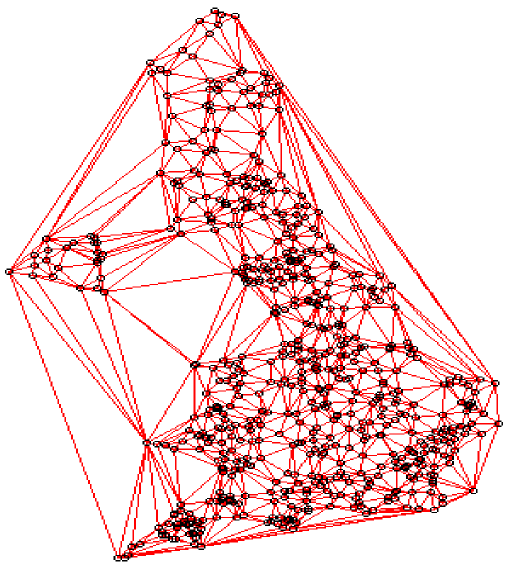
$\beta_0 = 2.0, \beta_1 = 1.0, \beta_2 = -2.0, \beta_3 = 3.0, \beta_4 = -3.0, \lambda = 0.4, \sigma = 0.3$ and $\alpha = 0.8$. The five summary of the estimators are shown in Table 3.1 with the true value and bias are reported together while, simulated distributions of each estimators have been visualized in Figure 3.1. Both results show that, the estimated values, $\hat{\theta}$ tend to be centered very near to the true values, θ specify in the DGP with bell-shaped distribution. This means our suggested estimation procedure, 3.3 can produce unbiased estimators, and there is no persistent tendency to under and overestimate the parameters, θ.



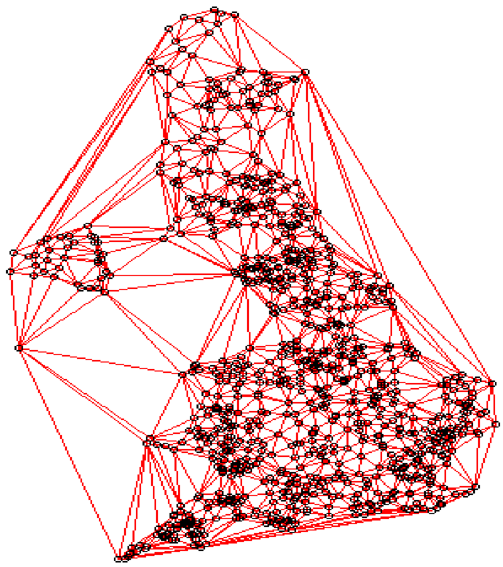
(a) spatial data from 100 stations



(b) spatial data from 300 stations



(c) spatial data from 600 stations



(d) spatial data from 900 stations

Figure 3.2: Illustration of spatial relationship for different sample sizes

Next, we also want to investigate the consistency of our unbiased estimators. Although some bias may be acceptable in an estimator, but the requirements that related to the idea of consistency need the bias tend to 0 as the sample size, n tends to ∞ and the variance tend to 0 as the sample size, n tends to ∞. The idea of consistency can be illustrated by using simulation of different sample sizes. Therefore, four different sample sizes i.e. $n = 100, 300, 600, 900$ are considered to examine the consistency of our unbiased estimators. As mentioned in Subsection 2.5.1, number of sample size, n is related to the number of gauged stations around the UK used to observed the spatial data which is standing for the location (latitude, longitude) of each gauged station. Repeated spatial data cannot be used in the analysis that involved a spatial relationship, therefore we need the spatial data that observe from 900 gauge stations due to the largest number of sample size considered in investigating the consistency of our unbiased estimator is 900. In the UK there are 1500 monitoring/gauged stations are available but only 928 of them are active in recording an updated annual peak flow data. The spatial dataset of size $n = 100, 300, 600, 900$ are sampled from these 928 gauged stations around the UK. The spatial relationships using contiguity based matrix that have been analyzed using these four different spatial samples data are illustrated in Figure 3.2.

Table 3.2: The scale of simulated distribution for $\hat{\theta}$ with its bias and variance when sample size increase

$\hat{\theta}$	True	n=100					n=300				
	value	**Min**	**Mean**	**Max**	**Bias**	**Variance**	**Min**	**Mean**	**Max**	**Bias**	**Variance**
$\hat{\beta}_0$	2.000	1.417	1.995	2.539	-0.005	0.02527	1.700	2.005	2.312	0.005	0.00732
$\hat{\beta}_1$	1.000	0.906	0.999	1.097	-0.001	0.00061	0.961	0.999	1.043	-0.001	0.00017
$\hat{\beta}_2$	-2.000	-2.336	-1.995	-1.696	0.005	0.00831	-2.208	-2.003	-1.840	-0.003	0.00279
$\hat{\beta}_3$	3.000	1.316	2.982	5.040	-0.018	0.27870	2.234	3.013	3.815	0.013	0.06889
$\hat{\beta}_4$	-3.000	-3.473	-2.996	-2.446	0.004	0.02655	-3.368	-2.998	-2.613	0.002	0.00963
$\hat{\lambda}$	0.400	-0.112	0.359	0.779	-0.041	0.02069	0.107	0.387	0.625	-0.013	0.00627
$\hat{\sigma}$	0.300	0.234	0.291	0.372	-0.009	0.00049	0.251	0.297	0.337	-0.003	0.00017
$\hat{\alpha}$	0.800	-2.032	0.891	3.691	0.091	0.31206	-0.855	0.839	2.023	0.039	0.11799
$\hat{\theta}$	**True**	**n=600**					**n=900**				
	value	**Min**	**Mean**	**Max**	**Bias**	**Variance**	**Min**	**Mean**	**Max**	**Bias**	**Variance**
$\hat{\beta}_0$	2.000	1.793	2.002	2.207	0.002	0.00391	1.831	2.001	2.164	0.001	0.00222
$\hat{\beta}_1$	1.000	0.969	1.000	1.027	0.000	0.00008	0.977	1.000	1.020	0.000	0.00005
$\hat{\beta}_2$	-2.000	-2.132	-2.002	-1.883	-0.002	0.00148	-2.117	-2.001	-1.909	-0.001	0.00099
$\hat{\beta}_3$	3.000	2.393	3.011	3.702	0.011	0.03500	2.513	3.007	3.485	0.007	0.01742
$\hat{\beta}_4$	-3.000	-3.199	-2.999	-2.795	0.001	0.00433	-3.174	-3.001	-2.792	-0.001	0.00318
$\hat{\lambda}$	0.400	0.194	0.393	0.541	-0.007	0.00320	0.256	0.399	0.529	-0.001	0.00200
$\hat{\sigma}$	0.300	0.272	0.298	0.324	-0.002	0.00008	0.275	0.299	0.320	-0.001	0.00005
$\hat{\alpha}$	0.800	-0.194	0.810	1.636	0.010	0.05892	-0.692	0.804	1.657	0.004	0.04402

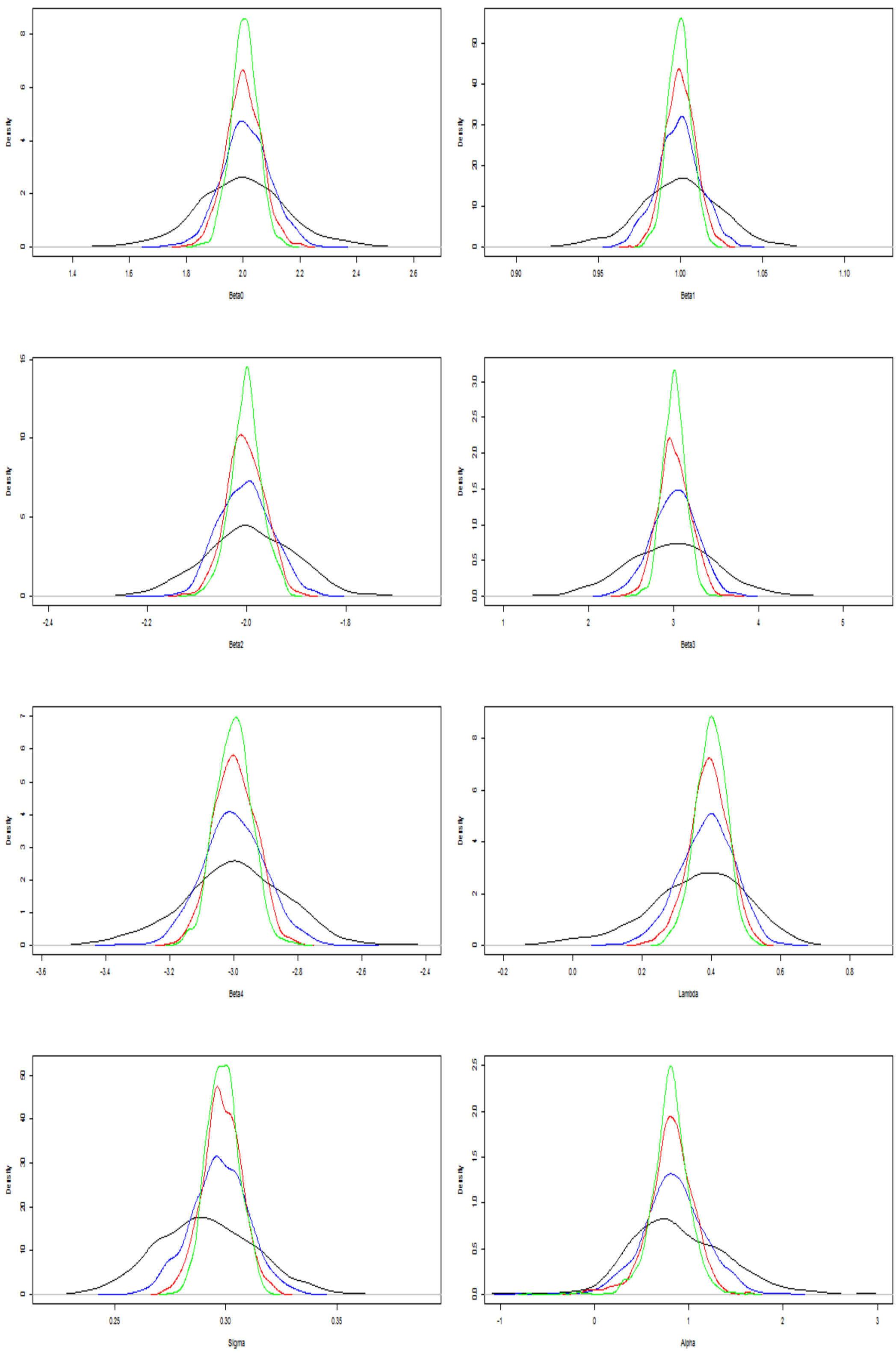


Figure 3.3: Simulated distribution for estimated values of unknown parameters, $\hat{\theta}$ when sample size, n increases (i.e.; black line for $n = 100$, blue line for $n = 300$, red line for $n = 600$, green line for $n = 900$).

Besides the spatial data, catchment characteristics data is also needed as the explanatory variables in data generation proses using Equation (3.11) . The datasets of catchment characteristics for each sample sizes have been randomly selected with replacement from 586 rural stations that introduced in Section 2.1. By fixing $\beta_0 = 2.0, \beta_1 = 1.0, \beta_2 = -2.0, \beta_3 = 3.0, \beta_4 = -3.0, \lambda = 0.4, \sigma = 0.3$ and $\alpha = 0.8$ with the error terms are randomly generated using `rskewnorm` function from `VGAM` package, we generated 1000 samples that then, their consistency can be examined one by one through the similar repeating simulation process and analysis that have been done in investigating the unbiasedness of the estimators. The results are summarized in Table 3.2 and the simulated distribution for each parameters with different sample sizes are visualized in the same diagram, Figure 3.3 for comparison. From Table 3.2 and Figure 3.3, we can say our unbiased estimators are consistent because they converge to the true parameter values specify in DGP when their standard error decrease to zero as the sample size increase.

3.5.2 Model estimation and selection

The satisfying result from the simulations of the performance for the proposed statistical method shown in subsection 3.5.1 made us no doubt to apply the developed maximum likelihood computational algorithm under skewed spatial error model (Algorithm 1) into the UK flooding data for non-Gaussian SIFE model estimation and selection. In this application, both spatial neighboring weight matrices which are contiguity based and distance based have been considered for a comparison purposes. Initially, four catchment characteristics data have been considered as explanatory variables, $(X_1, X_2, X_3, X_4) = (\ln(AREA), \frac{1000}{SAAR}, \ln(FARL), BFIHOST^2)$ and median annual maximum flood data (QMED) as the response variable, $y = \ln(QMED)$ in developing Non-Gaussian SIFE models shown in Table 3.3. A general linear regression model that is represented by FEH-QMED model and Gaussian SIFE model that has been established using the proposed statistical method in Chapter 2 are reported together in the table as a comparison. There is nothing interesting that can be expressed through Table 3.3 due to both models namely the Gaussian and Non-Gaussian SIFE models seems like giving a slightly similar result for both cases of spatial weight matrices. Based on AIC values, there is no improvement shows in the SIFE models if we change the distribution of error terms from normal to skewed-normal in the case of contiguity based weight matrix and very small improvement shows for distance based weight matrix. However, based on the standard error of some parameters in Non-Gaussian SIFE models leads us to consider more catchment characteristics as explanatory variables for further data analysis in Non-Gaussian SIFE model estimation and selection. These explanatory variables are based on ten catchment characteristics that usually used in flood risk analysis listed in Table 2.1 and defined as follows;

$$\begin{aligned} &(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}) \\ &= (\ln(AREA), \frac{1000}{SAAR}, \ln(FARL), BFIHOST^2, PROPWET, DPSBAR,) \\ &\quad FPEXT, RMED, SPRHOST, LDP) \end{aligned} \tag{3.18}$$

Table 3.3: Comparison between Linear, Gaussian SIFE, and non-Gaussian SIFE models with four initial catchment covariates.

Coefficients	**Linear model**			**G-SIFE model**			**nG-SIFE model**		
	Est.	**S.E**	**P-val**	**Est.**	**S.E**	**P-val**	**Est.**	**S.E**	**P-val**
Intercept	2.1826	0.0848	0.0000	2.1494	0.0911	0.0000	2.1457	0.0986	0.0000
ln(AREA)	0.8527	0.0131	0.0000	0.8497	0.0129	0.0000	0.8519	0.0130	0.0000
1000/SAAR	-1.8410	0.0523	0.0000	-1.8181	0.0748	0.0000	-1.8181	0.0744	0.0000
ln(FARL)	3.0335	0.2629	0.0000	2.9526	0.2725	0.0000	2.9599	0.2734	0.0000
(BFIHOST)^2	-3.3480	0.0969	0.0000	-3.2655	0.1144	0.0000	-3.2880	0.1163	0.0000
Lambda				0.4351	0.0546	0.0000	0.4308	0.0528	0.0000
Sigma				0.3657			0.3656	0.0109	0.0000
Alpha							0.7294	0.2570	0.0046
AIC	**575.90**			**519.34**			**519.90**		

(a) Spatial error analysis with contiguity weight matrix

Coefficients	**Linear model**			**G-SIFE model**			**nG-SIFE model**		
	Est.	**S.E**	**P-val**	**Est.**	**S.E**	**P-val**	**Est.**	**S.E**	**P-val**
Intercept	2.1826	0.0848	0.0000	2.1430	0.1024	0.0000	2.1405	0.1025	0.0000
ln(AREA)	0.8527	0.0131	0.0000	0.8472	0.0130	0.0000	0.8501	0.0130	0.0000
1000/SAAR	-1.8410	0.0523	0.0000	-1.7923	0.0796	0.0000	-1.7933	0.0800	0.0000
ln(FARL)	3.0335	0.2629	0.0000	3.1500	0.2627	0.0000	3.1602	0.2653	0.0000
(BFIHOST)^2	-3.3480	0.0969	0.0000	-3.2959	0.1169	0.0000	-3.3278	0.1184	0.0000
Lambda				0.5001	0.0599	0.0000	0.4997	0.0620	0.0000
Sigma				0.3689			0.3684	0.0110	0.0000
Alpha							0.8494	0.2286	0.0001
AIC	**575.90**			**525.78**			**524.74**		

(b) Spatial error analysis with distance weight matrix

Notice that the first four explanatory variables, X_1, X_2, X_3 and X_4 in Equation (3.18) are identified to be nonlinear by FEH-QMED model in Kjeldsen et al. (2008). The nonlinear effect in this four initial covariates has been defined reliable by using spatial additive model (2.8). Therefore, we remained to use the same transformation for these four initial covariates in non-Gaussian SIFE model estimation and selection by using ten selected covariates (3.18). For more comprehensive comparison, both cases which are without and with the logarithmic transformation for the additional covariates, X_5, X_6, X_7, X_8, X_9, and X_{10} have been used in this further data analysis. Here, AIC criteria and backward selection method have been used as a tool in model selection. Two spatial neighboring weight matrices which are contiguity weight matrix and distance weight matrix are also used in this application. Therefore, for non-Gaussian SIFE model estimation with contiguity based spatial weight matrix, twelve different models have been established, while another twelve non-Gaussian SIFE models developed using distance based spatial weight matrix. Those models are summarized in Table 3.4 and 3.5 for both cases: without and with logarithmic transformation of additional covariates respectively.

Table 3.4: Twelve established non-Gaussian SIFE models using contiguity based spatial weight matrix for model selection

Coefficients	Model 1			Model 2			Model 3			Model 4			Model 5			Model 6		
	Est.	S.E	P-val	Est.	S.E	P-val	Est.	S.E	P-val	Est.	S.E	P-val	Est.	S.E	P-val	Est.	S.E	P-val
Intercept	1.2032	0.4594	0.0088	0.9333	0.4098	0.0226	1.4550	0.2875	0.0000	1.4420	0.2870	0.0000	1.8851	0.1147	0.0000	2.0737	0.0988	0.0000
ln(AREA)	0.9308	0.0242	0.0000	0.9327	0.0243	0.0000	0.9370	0.0242	0.0000	0.9331	0.0241	0.0000	0.9277	0.0238	0.0000	0.8651	0.0133	0.0000
1000/SAAR	-1.3462	0.1678	0.0000	-1.3688	0.1643	0.0000	-1.5650	0.1237	0.0000	-1.5325	0.1214	0.0000	-1.6858	0.0811	0.0000	-1.6774	0.0815	0.0000
ln(FARL)	3.0544	0.2712	0.0000	3.0638	0.2714	0.0000	3.0297	0.2717	0.0000	3.0557	0.2718	0.0000	2.9702	0.2676	0.0000	2.9422	0.2694	0.0000
$(BFIHOST)^2$	-3.6459	0.2824	0.0000	-3.2855	0.1163	0.0000	-3.3398	0.1137	0.0000	-3.3393	0.1135	0.0000	-3.3449	0.1143	0.0000	-3.3481	0.1150	0.0000
PROPWET	0.5941	0.3133	0.0574	0.5357	0.3062	0.0802												
DPSBAR	-0.0009	0.0005	0.0734	-0.0010	0.0005	0.0644	-0.0007	0.0005	0.1442									
FPEXT	-2.3747	0.5364	0.0000	-2.3469	0.5349	0.0000	-2.2196	0.5336	0.0000	-1.9712	0.5059	0.0000	-1.9177	0.5074	0.0000	-1.9306	0.5109	0.0000
RMED	0.0124	0.0045	0.0062	0.0113	0.0044	0.0102	0.0090	0.0043	0.0348	0.0067	0.0040	0.0910						
SPRHOST	-0.0069	0.0049	0.1616															
LDP	-0.0027	0.0010	0.0060	-0.0028	0.0010	0.0054	-0.0031	0.0010	0.0016	-0.0031	0.0010	0.0018	-0.0031	0.0010	0.0016			
Lambda	0.4040	0.0584	0.0000	0.3859	0.0575	0.0000	0.4120	0.0552	0.0000	0.4070	0.0554	0.0000	0.4181	0.0544	0.0000	0.4143	0.0545	0.0000
Sigma	0.3560	0.0106	0.0000	0.3572	0.0106	0.0000	0.3572	0.0106	0.0000	0.3580	0.0106	0.0000	0.3585	0.0107	0.0000	0.3617	0.0107	0.0000
Alpha	0.7780	0.2362	0.0010	0.7861	0.2359	0.0008	0.7527	0.2448	0.0020	0.7431	0.2488	0.0028	0.7221	0.2575	0.0052	0.6945	0.2686	0.0096
AIC		**497.91**			**496.90**			**498.87**			**499.02**			**499.81**			**507.79**	

(a) Additional covariates without transformation

Coefficients	Model 7			Model 8			Model 9			Model 10			Model 11			Model 12		
	Est.	S.E	P-val	Est.	S.E	P-val	Est.	S.E	P-val	Est.	S.E	P-val	Est.	S.E	P-val	Est.	S.E	P-val
Intercept	-0.2063	1.0709	0.9282	-0.4943	0.9503	0.6030	-0.5922	0.9329	0.5286	-0.5453	0.9337	0.5620	-0.2612	0.9284	0.7794	0.4103	0.8997	0.6456
ln(AREA)	0.8928	0.0464	0.0000	0.8895	0.0460	0.0000	0.8649	0.0163	0.0000	0.8701	0.0145	0.0000	0.8587	0.0133	0.0000	0.8583	0.0134	0.0000
1000/SAAR	-1.1852	0.2024	0.0000	-1.1949	0.2013	0.0000	-1.1840	0.2000	0.0000	-1.2112	0.1970	0.0000	-1.2736	0.1962	0.0000	-1.5976	0.1352	0.0000
ln(FARL)	3.0578	0.2737	0.0000	3.0602	0.2737	0.0000	3.0677	0.2733	0.0000	3.0730	0.2733	0.0000	3.0787	0.2750	0.0000	3.0462	0.2761	0.0000
$(BFIHOST)^2$	-3.4328	0.3528	0.0000	-3.2376	0.1183	0.0000	-3.2400	0.1182	0.0000	-3.2347	0.1182	0.0000	-3.2099	0.1180	0.0000	-3.2821	0.1154	0.0000
ln(PROPWET)	0.3109	0.1551	0.0456	0.3134	0.1547	0.0424	0.3216	0.1538	0.0366	0.3401	0.1520	0.0250	0.3405	0.1532	0.0264			
ln(DPSBAR)	0.0550	0.0668	0.4122	0.0494	0.0660	0.4532	0.0465	0.0658	0.4778									
ln(FPEXT)	-0.0417	0.0428	0.3320	-0.0438	0.0427	0.3030	-0.0456	0.0425	0.2846	-0.0643	0.0333	0.0536						
ln(RMED)	0.5200	0.2229	0.0198	0.5104	0.2219	0.0214	0.5286	0.2192	0.0160	0.5609	0.2150	0.0090	0.5665	0.2162	0.0088	0.4054	0.2088	0.0518
ln(SPRHOST)	-0.0846	0.1438	0.5552															
ln(LDP)	-0.0518	0.0822	0.5352	-0.0468	0.0817	0.5686												
Lambda	0.3811	0.0574	0.0000	0.3787	0.0574	0.0000	0.3763	0.0572	0.0000	0.3793	0.0570	0.0000	0.3858	0.0562	0.0000	0.4160	0.0541	0.0000
Sigma	0.3629	0.0108	0.0000	0.3631	0.0108	0.0000	0.3633	0.0108	0.0000	0.3634	0.0108	0.0000	0.3644	0.0108	0.0000	0.3649	0.0109	0.0000
Alpha	0.7350	0.2513	0.0036	0.7288	0.2528	0.0040	0.7345	0.2519	0.0036	0.7414	0.2489	0.0028	0.7816	0.2383	0.0010	0.7455	0.2486	0.0026
AIC		**518.53**			**516.88**			**515.20**			**513.70**			**515.42**			**518.20**	

(b) Additional covariates with logarithmic transformation

Table 3.5: Twelve established non-Gaussian SIFE models using distance based spatial weight matrix for model selection

Coefficients	Model 13			Model 14			Model 15			Model 16			Model 17			Model 18		
	Est.	S.E	P-val	Est.	S.E	P-val	Est.	S.E	P-val	Est.	S.E	P-val	Est.	S.E	P-val	Est.	S.E	P-val
Intercept	1.1008	0.4616	0.0174	0.9233	0.4115	0.0250	1.5079	0.2904	0.0000	1.4957	0.2895	0.0000	1.8876	0.1180	0.0000	2.0817	0.1016	0.0000
ln(AREA)	0.9281	0.0245	0.0000	0.9294	0.0245	0.0000	0.9354	0.0243	0.0000	0.9306	0.0242	0.0000	0.9255	0.0240	0.0000	0.8621	0.0133	0.0000
1000/SAAR	-1.3412	0.1705	0.0000	-1.3505	0.1687	0.0000	-1.5719	0.1298	0.0000	-1.5348	0.1271	0.0000	-1.6737	0.0863	0.0000	-1.6631	0.0867	0.0000
ln(FARL)	3.2384	0.2644	0.0000	3.2368	0.2645	0.0000	3.2294	0.2657	0.0000	3.2554	0.2655	0.0000	3.1810	0.2612	0.0000	3.1484	0.2634	0.0000
(BFIHOST)^2	-3.5538	0.2804	0.0000	-3.3294	0.1186	0.0000	-3.3857	0.1169	0.0000	-3.3863	0.1166	0.0000	-3.3907	0.1174	0.0000	-3.4018	0.1181	0.0000
PROPWET	0.6302	0.3115	0.0434	0.6024	0.3073	0.0500												
DPSBAR	-0.0010	0.0005	0.0444	-0.0010	0.0005	0.0414	-0.0008	0.0005	0.0910									
FPEXT	-2.1936	0.5405	0.0000	-2.1768	0.5396	0.0000	-2.0437	0.5391	0.0000	-1.7740	0.5142	0.0000	-1.7001	0.5144	0.0010	-1.7606	0.5169	0.0000
RMED	0.0116	0.0044	0.0086	0.0110	0.0044	0.0118	0.0085	0.0042	0.0456	0.0058	0.0039	0.1388						
SPRHOST	-0.0043	0.0048	0.3788															
LDP	-0.0028	0.0010	0.0058	-0.0028	0.0010	0.0056	-0.0032	0.0010	0.0014	-0.0031	0.0010	0.0018	-0.0031	0.0010	0.0016			
Lambda	0.4523	0.0699	0.0000	0.4435	0.0697	0.0000	0.4779	0.0667	0.0000	0.4678	0.0674	0.0000	0.4797	0.0652	0.0000	0.4743	0.0654	0.0000
Sigma	0.3600	0.0107	0.0000	0.3604	0.0107	0.0000	0.3607	0.0107	0.0000	0.3618	0.0108	0.0000	0.3621	0.0108	0.0000	0.3654	0.0109	0.0000
Alpha	0.8949	0.2140	0.0000	0.8974	0.2146	0.0000	0.8743	0.2184	0.0000	0.8635	0.2223	0.0000	0.8477	0.2266	0.0000	0.8324	0.2310	0.0001
AIC		**505.59**			**504.37**			**506.03**			**506.96**			**507.13**			**515.27**	

(a) Additional covariates without transformation

Coefficients	Model 19			Model 20			Model 21			Model 22			Model 23			Model 24		
	Est.	S.E	P-val	Est.	S.E	P-val	Est.	S.E	P-val	Est.	S.E	P-val	Est.	S.E	P-val	Est.	S.E	P-val
Intercept	-0.0066	1.0947	0.9920	-0.2024	0.9660	0.8336	-0.2847	0.9479	0.7642	-0.2394	0.9458	0.8026	0.0026	0.9379	0.9999	2.1633	0.1015	0.0000
ln(AREA)	0.8830	0.0464	0.0000	0.8813	0.0462	0.0000	0.8612	0.0165	0.0000	0.8644	0.0148	0.0000	0.8552	0.0134	0.0000	0.8479	0.0130	0.0000
1000/SAAR	-1.2002	0.2059	0.0000	-1.2052	0.2052	0.0000	-1.1963	0.2039	0.0000	-1.2172	0.1991	0.0000	-1.2672	0.1981	0.0000	-1.6262	0.1273	0.0000
ln(FARL)	3.2283	0.2663	0.0000	3.2285	0.2663	0.0000	3.2329	0.2662	0.0000	3.2361	0.2662	0.0000	3.2330	0.2668	0.0000	3.1347	0.2648	0.0000
(BFIHOST)^2	-3.4024	0.3593	0.0000	-3.2731	0.1208	0.0000	-3.2759	0.1206	0.0000	-3.2734	0.1207	0.0000	-3.2477	0.1198	0.0000	-3.2819	0.1206	0.0000
ln(PROPWET)	0.3491	0.1540	0.0232	0.3510	0.1537	0.0226	0.3567	0.1531	0.0198	0.3648	0.1524	0.0168	0.3674	0.1535	0.0168	0.2478	0.1477	0.0930
ln(DPSBAR)	0.0348	0.0674	0.6030	0.0310	0.0666	0.6384	0.0292	0.0665	0.6600									
ln(FPEXT)	-0.0348	0.0420	0.4066	-0.0361	0.0419	0.3898	-0.0374	0.0418	0.3682	-0.0486	0.0332	0.1442						
ln(RMED)	0.4813	0.2229	0.0308	0.4761	0.2223	0.0324	0.4911	0.2196	0.0250	0.5079	0.2169	0.0192	0.5059	0.2182	0.0204			
ln(SPRHOST)	-0.0561	0.1466	0.7040															
ln(LDP)	-0.0405	0.0816	0.6170	-0.0379	0.0813	0.6384												
Lambda	0.4338	0.0706	0.0000	0.4328	0.0707	0.0000	0.4304	0.0705	0.0000	0.4349	0.0696	0.0000	0.4463	0.0676	0.0000	0.4836	0.0627	0.0000
Sigma	0.3666	0.0109	0.0000	0.3667	0.0109	0.0000	0.3668	0.0109	0.0000	0.3668	0.0109	0.0000	0.3673	0.0109	0.0000	0.3680	0.0109	0.0000
Alpha	0.8555	0.2246	0.0000	0.8521	0.2251	0.0000	0.8552	0.2251	0.0000	0.8596	0.2235	0.0000	0.8793	0.2203	0.0000	0.8604	0.2271	0.0000
AIC		**526.10**			**524.25**			**522.46**			**520.66**			**520.78**			**523.96**	

(b) Additional covariates with logarithmic transformation

Table 3.6: Comparison between the best models selected from linear, Gaussian SIFE and non-Gaussian that have been fitted into the UK flooding data with the usually used ten catchment characteristics in flood risk study.

Coefficients	Linear model			G-SIFE model			nG-SIFE model		
	Est.	S.E	P-val	Est.	S.E	P-val	Est.	S.E	P-val
Intercept	0.6138	0.3352	0.0676	0.9968	0.4062	0.0141	0.9333	0.4098	0.0226
ln(AREA)	0.9258	0.0257	0.0000	0.9290	0.0241	0.0000	0.9327	0.0243	0.0000
1000/SAAR	-1.2429	0.1303	0.0000	-1.3848	0.1648	0.0000	-1.3688	0.1643	0.0000
ln(FARL)	3.1879	0.2657	0.0000	3.0440	0.2696	0.0000	3.0638	0.2714	0.0000
(BFIHOST)^2	-3.2340	0.1020	0.0000	-3.2611	0.1156	0.0000	-3.2855	0.1163	0.0000
PROPWET	0.8257	0.2555	0.0013	0.4966	0.3011	0.0991	0.5357	0.3062	0.0802
DPSBAR	-0.0009	0.0005	0.0522	-0.0009	0.0005	0.0774	-0.0010	0.0005	0.0644
FPEXT	-2.7062	0.4967	0.0000	-2.3718	0.5347	0.0000	-2.3469	0.5349	0.0000
RMED	0.0133	0.0037	0.0003	0.0108	0.0044	0.0147	0.0113	0.0044	0.0102
LDP	-0.0024	0.0011	0.0248	-0.0027	0.0010	0.0057	-0.0028	0.0010	0.0054
Lambda				0.3937	0.0566	0.0000	0.3859	0.0575	0.0000
Sigma				0.3573			0.3572	0.0106	0.0000
Alpha							0.7861	0.2359	0.0008
AIC	**537.71**			**498.09**			**496.90**		

(a) Spatial error analysis with contiguity weight matrix

Coefficients	Linear model			G-SIFE model			nG-SIFE model		
	Est.	S.E	P-val	Est.	S.E	P-val	Est.	S.E	P-val
Intercept	0.6138	0.3352	0.0676	0.9953	0.4087	0.0149	0.9233	0.4115	0.0250
ln(AREA)	0.9258	0.0257	0.0000	0.9249	0.0244	0.0000	0.9294	0.0245	0.0000
1000/SAAR	-1.2429	0.1303	0.0000	-1.3683	0.1700	0.0000	-1.3505	0.1687	0.0000
ln(FARL)	3.1879	0.2657	0.0000	3.2115	0.2618	0.0000	3.2368	0.2645	0.0000
(BFIHOST)^2	-3.2340	0.1020	0.0000	-3.2949	0.1174	0.0000	-3.3294	0.1186	0.0000
PROPWET	0.8257	0.2555	0.0013	0.5632	0.3031	0.0631	0.6024	0.3073	0.0500
DPSBAR	-0.0009	0.0005	0.0522	-0.0010	0.0005	0.0529	-0.0010	0.0005	0.0414
FPEXT	-2.7062	0.4967	0.0000	-2.2003	0.5380	0.0000	-2.1768	0.5396	0.0000
RMED	0.0133	0.0037	0.0003	0.0103	0.0044	0.0185	0.0110	0.0044	0.0118
LDP	-0.0024	0.0011	0.0248	-0.0028	0.0010	0.0059	-0.0028	0.0010	0.0056
Lambda				0.4462	0.0645	0.0000	0.4435	0.0697	0.0000
Sigma				0.3611			0.3604	0.0107	0.0000
Alpha							0.8974	0.2146	0.0000
AIC	**537.71**			**506.50**			**504.37**		

(b) Spatial error analysis with distance weight matrix

Table 3.4 and 3.5 clearly show that Model 2 and model 14 are the best models among the twelve established non-Gaussian SIFE models in each type of spatial weight matrix, respectively. This selection is based on the smallest AIC value among them. The similar method of model estimation and selection (i.e. AIC criteria with backward selection) have also been used to determine the best model for the linear and Gaussian SIFE models with the same UK flooding dataset of the usually used ten considered catchment characteristics. The best models for each different statistical methods are summarized separately by type of spatial weight matrix in Table 3.6 for comparison. From the table it is clearly shown that there is an improvement in index flood estimation model when we change the distribution for the residuals of the spatial error model from normal to skewed normal distribution because the non-Gaussian SIFE model gives the lowest AIC value among others in both cases of spatial weight matrices. Also note, these three best models having the same nine significant catchment characteristics variables in both cases of spatial weight matrices. From Table 3.4 and 3.5, AIC values for all models with additional covariates without transformation are smaller than the AIC values for all models with log-transformed of additional covariates in both spatial neighboring impact. Therefore, explanatory variables defined in Equation (3.18) will be used for further analysis in evaluating established non-Gaussian SIFE models by using the leave one out cross validation (LOOCV) method for prediction in next subsection.

3.5.3 Model evaluation

Table 3.7: Gaussian SIFE and non-Gaussian SIFE models performance relative to the best Gaussian SIFE model in Chapter 2 under the LOOCV method with contiguity spatial weight matrix

Model	Covariates	Performance		Improvement (%)	
		MSPE	MAPE	MSPE	MAPE
G-SIFE	x1,x2,x3,x4,x5,x6,x7,x8,x9,x10	0.13225	0.26046	3.3	3.6
	x1,x2,x3,x4,x5,x6,x7,x8,x10	0.13269	0.26087	2.9	3.4
	x1,x2,x3,x4,x6,x7,x8,x10	0.13221	0.26095	3.3	3.4
	x1,x2,x3,x4,x7,x8,x10	0.13242	0.26170	3.1	3.1
	x1,x2,x3,x4,x7,x10	0.13233	0.26271	3.2	2.7
	x1,x2,x3,x4,x7	0.13430	0.26531	1.8	1.8
	x1,x2,x3,x4	0.13671	0.27013		
nG-SIFE	x1,x2,x3,x4,x5,x6,x7,x8,x9,x10	0.13093	0.25932	4.2	4.1
	x1,x2,x3,x4,x5,x6,x7,x8,x10	0.13155	0.25995	3.8	3.8
	x1,x2,x3,x4,x6,x7,x8,x10	0.13161	0.26031	3.7	3.6
	x1,x2,x3,x4,x7,x8,x10	0.13175	0.26121	3.6	3.3
	x1,x2,x3,x4,x7,x10	0.13183	0.26244	3.6	2.8
	x1,x2,x3,x4,x7	0.13401	0.26522	2.0	1.8
	x1,x2,x3,x4	0.13666	0.27013	0.0	0.0

The leave one out cross validation (LOOCV) method given in section 2.6 together with maximum likelihood computational procedure detailed in Algorithm 1 is applied to evaluate all non-Gaussian SIFE models that have been developed and summarized in Table 3.4 and 3.5. For more comprehensive comparison, the predictions for all established Gaussian SIFE models are also performed by using statistical models proposed in Chapter 2. Consistently, two types of spatial neighboring relationship, i.e., contiguity based and distance based matrices are used in evaluating model performance in prediction. The results of performance measurements, i.e., mean squared prediction error, (MSPE) and mean absolute prediction error, (MAPE) for different models are summarized in Table 3.7 and 3.8 for each considered spatial weight matrices respectively. Here, the percentage improvement for both performance measures of the established spatial index flood estimation models in both cases, i.e. Normal and Skewed Normal distribution of error terms, relative to the best Gaussian SIFE model of Chapter 2 (highlighted in blue) is also reported. The results from Table 3.7 and 3.8 show, the percentage improvement for all non-Gaussian SIFE models are slightly bigger than Gaussian SIFE models under both performance measures. Based on the AIC criteria Which was shown with the largest percentage improvement in both performance measures, the best non-Gaussian SIFE model for each considered spatial weight matrices have been selected and highlighted in pink (see Table 3.7 and 3.8). Therefore, these best models have been considered in a further investigation to justify the efficiency of the proposed statistical method in this chapter that can improve the performance of the UK index flood estimation model.

Table 3.8: Gaussian SIFE and non-Gaussian SIFE models performance relative to the best Gaussian SIFE model in Chapter 2 under the LOOCV method with distance spatial weight matrix

Model	Covariates	Performance		Improvement (%)	
		MSPE	MAPE	MSPE	MAPE
G-SIFE	x1,x2,x3,x4,x5,x6,x7,x8,x9,x10	0.13518	0.26594	2.6	2.4
	x1,x2,x3,x4,x5,x6,x7,x8,x10	0.13501	0.26557	2.7	2.4
	x1,x2,x3,x4,x6,x7,x8,x10	0.13471	0.26524	2.9	2.0
	x1,x2,x3,x4,x7,x8,x10	0.13514	0.26676	2.6	2.1
	x1,x2,x3,x4,x7,x10	0.13492	0.26689	2.8	2.0
	x1,x2,x3,x4,x7	0.13698	0.26908	1.3	1.2
	x1,x2,x3,x4	0.13879	0.27242		
nG-SIFE	x1,x2,x3,x4,x5,x6,x7,x8,x9,x10	0.13370	0.26427	3.7	3.0
	x1,x2,x3,x4,x5,x6,x7,x8,x10	0.13393	0.26443	3.5	2.9
	x1,x2,x3,x4,x6,x7,x8,x10	0.13390	0.26452	3.5	2.9
	x1,x2,x3,x4,x7,x8,x10	0.13434	0.26602	3.1	2.3
	x1,x2,x3,x4,x7,x10	0.13447	0.26663	3.1	2.1
	x1,x2,x3,x4,x7	0.13674	0.26891	1.5	1.3
	x1,x2,x3,x4	0.13866	0.27231	0.1	0.0

In particular for further investigation, we are comparing the established best Gaussian and non-Gaussian models that have been established in Chapter 2 and Chapter 3 with the updated FEH-QMED model of Kjeldsen and Jones (2010) based on the revised data-transfer method that partially takes account of spatial correlation. Consistently, the leave one out cross validation (LOOCV) method given in Section 2.6 together with two types of spatial neighboring relationship, i.e., contiguity based and distance based matrices, are used in evaluating model performance in prediction. The prediction for updated FEH-QMED model is also performed by using LOOCV method that has been applied into the given models in Kjeldsen and Jones (2010). The results of the mean squared prediction error (MSPE) for different models are summaries in Table 3.9. Here the percentage improvement of the established spatial error flooding models in both cases: normal and skewed normal distribution of residuals relative to the updated FEH-QMED model in Kjeldsen and Jones (2010) is also reported. Those selected models have been chosen for a comparison based on their outperform performance in two different methods developed in Chapter 2 and Chapter 3 respectively. Clearly, the best Gaussian and non-Gaussian SIFE models in Chapter 2 and 3 are performed well that shown by the decreasing of the MSPE values, but the non-Gaussian SIFE model of Chapter 3 with contiguity spatial weight matrix outperform the other models with 17.1 percent improvement over the updated FEH-QMED model in reducing the MSPE. Referring to the results in Table 3.9, we can say that, by changing the distribution of residuals in spatial error model from normal to skewed normal distribution, can develop the effective statistical method for index flood estimation that better fit the UK flooding data. With that, we can justify the new statistical model structure proposed in this chapter can improve the UK index flood estimation model.

Table 3.9: Comparison of model performance using leave one out cross validation method

Model	Covariates	Spatial weigt	MSPE	Improvement
FEH-QMED	X_1, X_2, X_3, X_4	-	0.15859	-
G-SIFE	X_1, X_2, X_3, X_4	Contiguity	0.13671	13.8 percent
	X_1, X_2, X_3, X_4	Distance	0.13879	12.5 percent
nG-SIFE	$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_{10}$	Contiguity	0.13155	17.1 percent
	$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_{10}$	Distance	0.13393	15.6 percent

3.6 Discussion and conclusion

3.6.1 Discussion

The main objective of this chapter is to develop a more efficient statistical model in estimating the index flood that better fit the UK flooding data. Spatial error model with the skewed normal distribution of error terms have been developed with a maximum likelihood computational procedure is suggested together for the application into the UK flooding data. several findings that have found through computational tasks of the suggested statistical procedure. The findings are listed below:-

i. The estimators determined by the proposed maximum likelihood estimation method under skewed spatial error model have been proved to be unbiased and consistent through the simulation in investigating the finite sample performance. Hence, the suggested computational procedure to solve the maximum likelihood estimation of the proposed statistical model has been successfully applied to the UK flooding data in estimating the index flood.

ii. Spatial error model with skewed normal distribution error by using contiguity weight matrix is an improved model based on AIC criteria which was shown by the percentage improvement in mean square prediction error, (MSPE) and mean absolute prediction error, (MAPE) relative to the best Gaussian SIFE model is among the largest.

iii. The reliability and accuracy of the improved statistical model suggested in this chapter have been justified through the largest percentage improvement of MSPE relative to the updated FEH-QMED model based on a revised data-transfer method by Kjeldsen and Jones (2010).

From those findings, we can say that the new statistical model structure of spatial error model that has been developed by changing the distribution of its residuals from normal distribution to skewed normal distribution is justified as an efficient statistical method in developing the UK index flood estimation model. The suggested maximum likelihood estimation computational procedure for this new statistical model structure have been successfully applied into the UK flooding data of the usually used ten catchment characteristics in flood risk study as the covariates to develop the skewed spatial error flood model that outperform the other models based on AIC score which was shown by the largest percentage improvement relative to updated FEH-QMED model in Kjeldsen and Jones (2010).

Actually, there are some more catchment characteristics available in the UK flooding dataset. It could be better if we can use all the catchment variables in the data fitting with the new statistical method proposed in this chapter. However, from hydrological literature, some of the catchment characteristics have been detected to be highly correlated. Maximum likelihood estimation methods have many remarkable properties, but in the case of highly correlated predictor variables it will cause unsatisfactory in regression problems including large variability due to the estimation and variable selection are executed in separate steps. Due to this reason, many flood risk study in the literature have to reduce the number of considered catchment characteristics especially the one that has been detected highly correlated in their analysis of developing the flood regression model. This problem leads us to suggest panelized likelihood estimation method for further research to discover relevant predictive catchment characteristic variables for the developed model in this chapter. This is because penalized regression methods have been developed in linear regression setting to overcome the drawbacks in the traditional variable selection method that has been used in this chapter. For this purpose, we will use the novel adaptive Lasso as a regularization tool and with no doubt, all available catchment characteristics available in the UK flooding dataset can be used and have an equal change in variable selection in developing the UK index flood estimation model.

3.6.2 Conclusion

The Gaussian distribution is widely used in describing data in many application. In the hydrological application analysis such as the Regional Flood Frequency Analysis (RFFA), a Gaussian assumption was mostly used in estimating the index flood by using regression models. Nevertheless, from the literature, the celebrated Gaussian distribution is no longer suitable in environmental, hydrological and ecological studies since the observed spatial variables in those fields are known to have skewed distributions. In such a situation, the Gaussian assumption is too ideal to be used in developing a model and could provide a biased estimate. Furthermore, a large amount of the uncertainty contributed by the index flood estimation in the RFFA is still an open problem among hydrologists. Index flood estimation is one of the main steps in the flood regionalization procedure of RFFA by using index flood method. Accurate estimates of the index flood from index flood regression models are essential for precise prediction of flood design at any locations of ungauged catchments area. Therefore, this study attempts to respond to both aforementioned issues by proposing a more reliable index flood estimation model that takes into account the spatial dependence and non-Gaussianity.

Results in subsection 3.5.2 and 3.5.3 allow us to conclude that by changing the distribution of residuals from normal to skewed normal in spatial error model the estimation of index flood in the UK can be improved. This is based on AIC criteria which was shown by the percentage of improvement in mean square prediction error, (MSPE) and mean absolute prediction error, (MAPE) in comparison with the best model in Chapter 2 are among the largest. Based on those values: AIC, MSPE, and MAPE justify the proposed method in this chapter can improve the method that has been applied in Chapter 2 in estimating the UK index flood at ungauged catchments area. This justification hence can answer the main objectives of this study which is to develop a more efficient statistical model for index flood estimation by taking into account spatial correlation and non-Gaussianity. The improved index flood estimation model for the UK flooding data named as non-Gaussian SIFE model can be expressed as

$$\begin{aligned}\ln(QMED_i) &= 0.9333 + 0.9327\ln(AREA_i) - 1.3688(1000SAAR_i^{-1}) + 3.0638 \\ &\quad \ln(FARL_i) - 3.2855(BFIHOST_i)^2 + 0.5357(PROPWET_i) - \\ &\quad 0.0010(DPSBAR_i) - 2.3469(FPEXT_i) + 0.0113(RMED_i) - \\ &\quad 0.0028(LDP) + u_i, \qquad (3.19) \\ u &= 0.3859Wu + 0.3572e.\end{aligned}$$

where $u = (u_1, u_2, \cdots, u_n)'$ and W is a $n \times n$ standardized contiguity based spatial weight matrix, with model error terms, $e = (e_1, e_2, \cdots, e_n)'$ where e_i being i.i.d. of skewed normal distribution.

Besides in ensuring high prediction accuracy, discovering relevant predictive variables is also important in statistical method. Apparently, the proposed maximum likelihood estimation method of skewed spatial error model has been applied to the UK flooding data of ten considered catchment characteristics as covariates whereas there are some more catchment characteristics available in the dataset. Further, the estimation and variable selection in this proposed statistical method has been executed in separate steps by using Akaike's information criterion (AIC) as a model selection criteria. This kind of traditional model selection methods achieve simplicity but have been shown to yield models that have low prediction accuracy due to their number of limitations and drawbacks. Both aforementioned issues have motivated us to pursue a further investigation in discovering any other relevant predictive variables for the skewed spatial index flood estimation model by considering nineteen catchment characteristics that are available in the FEH-CD ROM which are suggested to be particularly suitable to describe a catchment's hydrological behavior in the UK. Nevertheless, simultaneous estimation and variable selection will be suggested for further research in our proposed skewed spatial error model through the penalized likelihood estimation method by utilizing the novel technique of Adaptive Lasso. This is to overcome the limitations and drawbacks of the traditional variable selection method and enhance the prediction accuracy in estimating the index flood. For the details will be demonstrated in Chapter 4.

Chapter 4

Variable Selection in Index Flood Estimation Model by Adaptive Lasso

4.1 Introduction

Flooding is a complex process that is caused not only by the hydrological factor but also by a lot of physical factors. With modern data acquisition techniques, the observed variables from those mentioned flooding factors have been and continue to be collected. There are one thousand and five hundred gauged stations throughout the UK. Advance in technology makes it possible for them to provide an updated flood information and flood peaks data with high quality through the National River Flow Archive. The Center for Ecology and Hydrology, (CEH) has as well published FEH-CD ROM that provides a large number of catchment characteristics, which has now been replaced by a web-serviced: https://fehweb.ceh.ac.uk/. These characteristics are suggested to be particularly suitable to describe a catchment's hydrological behavior in the UK. There are twenty five catchment characteristics available which have been grouped into landform descriptors, floodplain descriptors, climate and soil descriptors, and urban suburban land cover descriptors. Some of these catchment characteristics are highly correlated which will be detail explained in Section 4.2.

The use of multiple regression method is a common practice in operational hydrological analysis to established a statistical model that can perform the underlying process between a certain hydrological parameter and a set of catchment characteristics. The index flood estimation model in Flood Estimation Handbook (FEH) known as QMED model is a well established model in the UK that has been developed by using this approach (IH, 1999). With many potential candidate predictors and some of them being detected to be highly correlated, it could give rise to a regression problem that requires an efficient statistical model selection to find an optimal model. This optimal model should be as simple as possible while still providing good predictive performance. Traditional variable selection procedures, such as stepwise selection methods, Akaike's information criteria or Bayesian information criteria, suffer from high variability

(Breiman, 1996) because the estimation and variable selection steps are resolved separately. In estimation step, these traditional methods are often trapped into a local optimal solution rather than a global optimal one (Zou, 2006). Moreover, in the variable selection step, the stochastic error or uncertainty is not taken into account (Fan and Li, 2001). These could yield models with low prediction accuracy, especially when there are correlated predictors or when the number of predictors is large.

Penalized regression methods, such as bridge regression, least absolute shrinkage and selection operator (Lasso) and elastic net have been developed in linear regression setting to overcome the aforementioned drawbacks in the traditional variable selection methods. The high prediction accuracy and computational efficiency of penalized regression methods have brought them increasing attention over the last decade. In addition, those methods can also produce more stable results for correlated data and data where the number of predictors is much larger than the sample size. Unlike traditional methods, panelized regression method do not explicitly select the variables, instead they minimize the sum of squares of residuals by using a penalty on the size of regression coefficients. This penalty results in the regression coefficients to be shrunk toward zero. If the shrinkage is large enough, some regression coefficients are set to zero exactly. Thus, panelized regression methods perform variable selection and coefficient estimation simultaneously. This suggested statistical method will utilize a novel application of adaptive lasso introduced by Zou (2006) as a tool in variable selection for spatial error model with skewed normal residuals that have been established in Chapter 3. Adaptive lasso has been chosen because it not only retains the good features of Lasso but also enjoys the Oracle property which Lasso does not possess (see Fan and Li, 2001).

In the literature, catchment characteristics used in the analysis of developing flood regression models are only based on those listed in Table 2.1 (Kjeldsen et al., 2008; Robson and Reed, 1999; Kjeldsen and Jones, 2010). Since they used traditional methods in model estimation and selection, it is necessary for them to reduce the number of characteristics especially for the one that has been detected to be highly correlated in Wagener et al. (2004). This is to ensure the approximation of the solution to the underlying truth is efficient. Therefore, not all available catchment characteristics are considered in their model estimation and selection for the purposed of developing flood regression models. Furthermore, the variable selection in those flood regression models is made under independent error structure, which may bias the variable selection for the models. These problems have motivated us to propose a penalized estimation procedure that utilizes adaptive lasso to estimate the important catchment characteristics predictors for the UK index flood estimation under spatial dependent error models by considering all available catchment characteristics in the dataset. Notice that, the number of available catchment characteristics (being supposed to be greater than a sample size) is not too large. Nevertheless, the use of penalized regression method for the further investigation in this research attempts to response the instability in traditional variable selection due to the correlated predictors.

This chapter will;

(i). suggest a penalized likelihood estimation method by utilizing adaptive lasso as a tool for the non-Gaussian spatial index flood model selection.

(ii). propose a new computational algorithm procedure to solve the penalize likelihood estimation method suggested in (i).

(iii). present the simulation of the finite samples performance for the proposed algorithm procedure of non-Gaussian spatial index flood estimation model.

(iv). apply the proposed algorithm procedure of the non-Gaussian spatial index flood estimation model to the UK flooding data and evaluate its performance.

Hopefully, the last objective of this research project that is to identify significant explanatory variables of catchment characteristics and enhance prediction performance of the fitted non-Gaussian spatial index flood estimation model can be achieved by completing the aforementioned tasks.

4.2 The UK catchment characteristics

Flood Estimation Handbook (FEH) published by Institute of Hydrology (IH) which is now known as Centre for Ecology and Hydrology (CEH), has accomplished together with CD ROM. It contains data where catchment characteristics can be extracted. The FEH-CD ROM has been updated from time to time to ensure the quality of the updated flood data in the UK. Nowadays it has been replaced by a website available at https://fehweb.ceh.ac.uk/. All catchment characteristics listed in this study are taken from the latest FEH-CD ROM 3.0. These characteristics are suggested to be particularly suitable for describing the UK catchment's hydrological behavior. Table 4.1, explains the abbreviation for the FEH catchment characteristics. The first ten catchment characteristics listed in Table 4.1 have been introduced in Chapter 2 before. These ten catchment characteristics are the commonly used variables in the flood risk studies and have also been used as explanatory covariates for application of our proposed statistical methods in Chapters 2 and Chapter 3. To apply the penalized likelihood estimation method by utilizing adaptive lasso that suggested in this chapter, 19 catchment characteristics will be taken into consideration as explanatory covariates, namely (4.1).

$$\begin{aligned}&(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}, X_{18}, X_{19})\\&= (\ln(AREA), \frac{1000}{SAAR}, \ln(FARL), BFIHOST^2, PROPWET, DPSBAR,)\\&\quad FPEXT, RMED, SPRHOST, LDP, ALTBAR, ASPBAR, ASPVAR,\\&\quad DPLBAR, RMED-1H, RMED-2D, SAAR_{4170}, FPLOC, FPDBAR) \qquad (4.1)\end{aligned}$$

Notice that, we do not consider the rest six catchment characteristics listed in Table 4.1. This is because there are a lot of missing values in these six variables. In general, the FEH catchment characteristics belong to the following four groups (Wagener et al., 2004)

i. Landform descriptors (AREA, LDP, DPLBAR, DPSBAR, ALTBAR, ASPBAR, ASPVBAR)

ii. Floodplain descriptors (FARL, FPEXT, FPLOC, FPDBAR)

iii. Climate and soil descriptors (SAAR, $SAAR_{4170}$, RMED-1H, RMED-1D, RMED-2D, SPRHOST, BFIHOST, PROPWET)

iv. Urban and suburban descriptors ($URBCONC_{1990}$, $URBCONC_{2000}$, $URBLOC_{1990}$, $URBLOC_{2000}$, $URBEXT_{1990}$, $URBEXT_{2000}$)

Table 4.1: The UK catchment characteristics

Name	**Description of catchment characteristics**
AREA	Catchment drainage area
SAAR	Standard annual average rainfall (1961-1990)
FARL	Lake, reservoir or loch flood attenuation index
BFIHOST	Base ow index derive by Hydrology of Soil
PROPWET	Proportion of time when soil moisture decit
DPSBAR	Mean drainage path slope
FPEXT	Floodplain extent
RMED-1D	Median annual maximum 1-day rainfall
SPRHOST	Standard percentage runoff
LDP	Longest drainage path length
ALTBAR	Mean elevation
ASPBAR	Mean of dominat aspect of the catchment slope
ASPVAR	Mean of invariability aspect of the catchment slope
DPLBAR	Mean drainage path length
RMED-1H	Median annual maximum 1-hour rainfall
RMED-2D	Median annual maximum 2-day rainfall
$SAAR_{4170}$	Standard annual average rainfall (1941-1970)
FPLOC	Floodplain location
FPDBAR	Mean flood depth upstream of the catchment outlet
$URBEXT_{1990}$	Extend of urban and suburban land cover (1990)
$URBCONC_{1990}$	Concentration of urban and suburban land cover (1990)
$URBLOC_{1990}$	Location of urban and suburban land cover (1990)
$URBEXT_{2000}$	Extend of urban and suburban land cover (2000)
$URBCONC_{2000}$	Concentration of urban and suburban land cover (2000)
$URBLOC_{2000}$	Location of urban and suburban land cover (2000)

Some of these characteristics are highly correlated especially the one in the same group of descriptors (Wagener et al., 2004). For example SAAR and $SAAR_{4170}$ are highly correlated because both of them measured the standard annual average rainfall but in different time series (1961 to 1990 for SAAR and 1941 to 1970 for $SAAR_{4170}$). Another example of strong correlation occurs between DPLBAR with AREA and LDP because DPLBAR is an index that combines catchment size (AREA) and drainage path (LDP) configuration. Last example happens between baseflow index (BFIHOST) and standard percentage runoff (SPRHOST) as both of them are measures of catchment's runoff response to rainfall input. It is therefore necessary for hydrologists to analysis this correlation and reduce the number of considered catchment characteristics especially for the one that has been detected highly correlated in index flood regression model development, which is regards as the purpose of reducing the unsatisfactory in regression problem.

4.3 Penalized maximum likelihood

In previous chapter we were dealing with maximum likelihood estimation to determine values of unknown parameters $\hat{\theta} = (\hat{\underset{\sim}{\beta}}, \hat{\lambda}, \hat{\sigma}, \hat{\alpha})$ i.e., a maximum likelihood estimator that maximizes $\ell(\theta|y)$, a log likelihood function of spatial error model with skewed normal residuals (3.17) seen in Subsection 3.3.1. Maximum likelihood estimation has many remarkable properties, but is often unsatisfactory in regression problems for two reasons: (i) large variability due to the number of predictor p being larger than the sample size n or highly correlated feature in predictor variables, (ii) lack of interpretability if the number of predictor p is large. Therefore, an efficient variable selection method is important in dealing with this problem. In previous chapter, backward selection method with Akaike information criterion (AIC) has been used in model selection. This traditional method has several drawbacks including instability (Breiman, 1996), as the estimation and variable selection are executed in separate step.

Penalized maximum likelihood is one of a different ways in statistical method. It is able to overcome the limitations of the traditional variable selection methods. Instead of maximizing $\ell(\theta|y)$, it maximizes the function

$$M(\theta) = \ell(\theta|y) - \eta P(\theta), \tag{4.2}$$

where

$p(\theta)$ = penalty function that penalizes the less realistic values of the unknown parameters,

η = tuning parameter that control the trade of between log likelihood and penalty function, and

$M(\theta)$ = objective function.

In regression, minimizing a loss function is more realized than maximizing likelihood. Therefore, an equivalent formulation for Equation (4.2) in estimating θ is to minimize

$$M(\theta) = L(\theta|y) + \eta P(\theta), \tag{4.3}$$

where $L(\theta|y)$ is a loss function, a quantity that is usually proportional to a negative log likelihood such as the residual sum of squares. Base on Equation 4.3, the penalized maximum likelihood estimation for spatial error model with skewed normal residual can be expressed. For this purpose we utilize adaptive lasso penalty function to impose on the regression coefficients since this method has been suggested as a tool in variable selection of our established spatial flood model.

4.3.1 Adaptive LASSO

In model selection, the Oracle property (Zou, 2006) of statistical estimation method is desirable because it assures two important asymptotic properties: (i) consistency on model selection, i.e., as the sample size approaches infinity, the selected set of predictor variables approaches the true set of predictor variables with probability 1, (ii) consistency in estimation where the estimators are asymptotically normal with the same mean and covariance that they would have by maximum likelihood estimation, when the zero coefficients are know in advance. Fan and Li (2001) speculated that Lasso estimator does not have the Oracle property. Therefore, adaptive Lasso developed by Zou (2006) modifies the Lasso penalty function by applying different weights to different coefficients, and demonstrated that adaptive Lasso enjoys Oracle properties and eases application. These weights control the zero coefficients more than they control shrinking the nonzero coefficients. The adaptive Lasso penalty function is expressed as,

$$\sum_{j=1}^{p} \hat{\tau}_j |\theta_j|, \tag{4.4}$$

where $\hat{\tau}_j = 1/|\hat{\theta}|^{\gamma}$, by picking $\gamma > 0$ and $\hat{\theta}$ is initial estimator. We will detail discuss it in next subsection.

By substituting adaptive lasso penalty function (4.4) into Equation (4.3), we then can determine the adaptive Lasso estimate, $\hat{\theta}^*$ for established non-Gaussian spatial flood model by

$$\hat{\theta}^* = \arg\min_{\theta}(-\ell(\theta|y)) + \eta \sum_{j=1}^{p} \hat{\tau}_j |\theta_j|, \tag{4.5}$$

where

$$\ell(\theta|y) = n\log 2 + \sum_{i=1}^{n} \log\phi\Big[\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y - X\beta)\Big] + n\log\sigma_z - n\log\sigma +$$
$$\sum_{i=1}^{n} \log\Phi\Big[\alpha\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y - X\beta)\Big)\Big] + \log|I - \lambda W|,$$

$$\text{with } \mu_z = \frac{\sqrt{2}\alpha}{\sqrt{\pi}\sqrt{1+\alpha^2}} \text{ and } \sigma_z = \sqrt{1 - \frac{2\alpha^2}{\pi(1+\alpha^2)}}.$$

4.3.2 Initial estimator

As mentioned above, the adaptive weights can be calculated through $\hat{\tau}_j = 1/|\hat{\theta}|^{\gamma}$. Apparently, the weight depends on the initial estimator $\hat{\theta}$ regardless $\gamma = 1$ has been chosen. By default, algorithms in any statistical computational software use the OLS estimate $\hat{\theta}$(ols) of the regression coefficients as the initial estimator to form the adaptive Lasso weights, $\hat{\tau}_j$. However, the algorithms also provide options so that any initial estimators can be used in calculating the desire weights. For example if collinearity is a concern, $\hat{\theta}$(ridge) from the best ridge regression fit is suggested to be used as the initial estimator, because it is more stable then $\hat{\theta}$(ols). Here, estimate values, $\hat{\theta}$(ssem) from the best spatial error model with skew normal distribution established in Chapter 3, will be used as the initial estimator in calculating the adaptive Lasso weights.

4.3.3 Selection of tuning parameter

It is very critical to choose a proper tuning parameter η since it determines the sparsity of the selected model. This can be illustrated by using Equation (4.5). Note that, η is a nonnegative value. When $\eta = 0$, the optimum $\hat{\theta}^*$ is equal to spatial error model with skew normal residuals (SSEM) solution. As η increases, more shrinkage is imposed on the model coefficients, hence it shrinks from SSEM solution toward zero. Therefore, an optimal tuning parameter can result in parsimonious model which is simple with great explanatory predictive power. Wang, Li and Tsai (2007) and Wang, Li and Leng (2009) showed that Bayesian Information criterion (BIC) is consistence in model selection. Here, BIC-type criterion is employed to choose the tuning parameter. For a given η, we can obtain an estimate, $\hat{\theta}^*_\eta$, with which a model M_η can be established. Therefore, the tuning parameter η that can give the smallest BIC value of the M_η will be chosen. Let d_η be the number of non zero components of $\hat{\theta}^*_\eta$ and ℓ_η be the log likelihood value for M_η. The BIC-type criterion in selecting the tuning parameter is defined by

$$BIC(\eta) = -2(\ell_\eta) + (\log(n))d_\eta \tag{4.6}$$

4.4 Computational aspects

Computational aspects are important for the optimization of the penalized likelihood estimation of skewed spatial error model by using Adaptive Lasso which is define in Equation (4.5). The second term in (4.5), is the so-called "ℓ_1 penalty" which is the weighted L_1 norm of the regression coefficients. L_1 norm is non-differentiable penalty which makes optimization of the objective function in (4.5) inapplicable by using `optim` function in `R` core package `stats` that provides a variety of general purpose optimization for differentiable objectives. It is also inapplicable to solve L_1 penalization by using other alternative packages of optimization such as `penalized`, `glmnet` and `mlegp`. Those packages are developed for solving the particular problem functions with difficulties to make an extension or modification, which means they do not permit to solve general optimization of L_1 regularized functions. The `lbfgs` package addresses this issue, which allows to deal with an optimization of an objective with an L_1 penalty through both the Limited-memory Broyden-Fletcher-Goldfarb-Shanno(L-BFGS) and the Orthant-Wise Limited-memory Quasi-Newton (OWL-QN) (Coppola et al., 2014).

All optimization routines in `lbfgs` package are handled by the `lbfgs` function. The objective and gradient functions are the basic input and should be supplied when using `lbfgs` function (Coppola et al., 2014). From Equation (4.5) the first term of the objective function is negative value of the log-likelihood function (3.17) for the skewed spatial error model. While, the gradient function can be derived by taking the partial derivatives of objective function with respect to the each parameters. So here, the gradient function is similar to a negative score functions. The score function then, can be derived by taking partial derivatives of the log likelihood (3.17) with respect to the parameters. Since gradient function is necessary in `lbfgs` function, the next step is to derive the score function to be used as gradient functions in a suggested computational algorithm procedure. We will explain this in Subsection 4.4.2.

4.4.1 The score functions

The score function can be derived by taking partial derivatives of the log likelihood in Equation (3.17) with respect to the parameters β, λ, α, and σ. The solutions for each of the partial derivatives are shown below while their detailed derivations are provided in Appendix A, B, C and D at the end of this thesis.

<u>Partial derivative of log-likelihood with respect to β</u>

There are two expressions that give a nonzero derivatives when differentiating the log-likelihood function with respect to β:

The first expression is

$$\frac{\partial}{\partial\beta}\left(\sum_{i=1}^{n}\log\phi\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\right)$$
$$=\frac{\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big).$$

Detailed derivation is provided in Appendix A, page 107

The second expression is

$$\frac{\partial}{\partial\beta}\left(\sum_{i=1}^{n}\log\Phi\Big(\alpha\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\Big)\right)$$
$$=-\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}\left(X^T(I-\lambda W)_i^T\left\{\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}\right).$$

Detailed derivation is provided in Appendix A, page 108

Hence

$$\begin{aligned}\frac{\partial\ell}{\partial\beta} \;&=\; \frac{\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)-\\ &\quad\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}\left(X^T(I-\lambda W)_i^T\left\{\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}\right).\end{aligned}$$

Partial derivative of log-likelihood with respect to λ

There are three expressions that give a nonzero derivatives when differentiating the log-likelihood function with respect to λ:

The first expression is

$$\frac{\partial}{\partial\lambda}\left(\sum_{i=1}^{n}\log\phi\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\right)$$
$$=\frac{\sigma_z}{\sigma}\Big(W\big(y-X\beta\big)\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big).$$

Detailed derivation is provided in Appendix B, page 109

The second expression is

$$\frac{\partial}{\partial\lambda}\left(\sum_{i=1}^{n}\log\Phi\Big(\alpha\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\Big)\right)$$
$$=-\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}\left(W_i^T\big(y-X\beta\big)\left\{\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}\right).$$

Detailed derivation is provided in Appendix B, page 110

The third expression is

$$\frac{\partial}{\partial\lambda}\Big(\log|I-\lambda W|\Big) = -tr\Big((I-\lambda W)^{-1}W\Big).$$

Detailed derivation is provided in Appendix B, page 110

Hence

$$\begin{aligned}\frac{\partial \ell}{\partial \lambda} &= \frac{\sigma_z}{\sigma}\Big(W\big(y-X\beta\big)\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big) - tr\Big((I-\lambda W)^{-1}W\Big) - \\ &\quad \frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}\left(W_i^T\big(y-X\beta\big)\left\{\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}\right).\end{aligned}$$

Partial derivative of log-likelihood with respect to α

From Equation (3.3) and (3.4) μ_z and σ_z are known as the function of α. Three expressions that give a nonzero solution when differentiating the log-likelihood function with respect to α:

The first expression is

$$\begin{aligned}&\frac{\partial}{\partial\alpha}\left(\sum_{i=1}^{n}\log\phi\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\right)\\ &= \frac{\mu_z^2}{\alpha(1+\alpha^2)\sigma\sigma_z}\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\Big((I-\lambda W)(y-X\beta)\Big) - \\ &\quad \frac{\sqrt{2}}{1+\alpha^2}\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\cdot\mathbb{1}.\end{aligned}$$

Detailed derivation is provided in Appendix D, pages 113-114

The second expression is

$$\begin{aligned}&\frac{\partial}{\partial\alpha}\left(\sum_{i=1}^{n}\log\Phi\Big(\alpha\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\Big)\right)\\ &= \sum_{i=1}^{n}\left(\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)+\frac{\sqrt{2}\alpha}{1+\alpha^2}-\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)\right.\\ &\quad \left.\left\{\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}\right).\end{aligned}$$

Detailed derivation is provided in Appendix D, pages 114-115

The third expression is

$$\frac{\partial}{\partial\alpha}\Big(n\log\sigma_z\Big) = -\frac{n\mu_z^2}{\alpha(1+\alpha^2)\sigma_z^2}.$$

Detailed derivation is provided in Appendix D, page 115

Hence

$$\begin{aligned}\frac{\partial \ell}{\partial \alpha} \;=\;& \frac{\mu_z^2}{\alpha(1+\alpha^2)\sigma\sigma_z}\Big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\Big((I-\lambda W)(y-X\beta)\Big) - \\ & \frac{\sqrt{2}}{1+\alpha^2}\Big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T \cdot \mathbb{1} - \frac{n\mu_z^2}{\alpha(1+\alpha^2)\sigma_z^2} + \\ & \sum_{i=1}^{n}\Bigg(\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta) + \frac{\sqrt{2}\alpha}{1+\alpha^2} - \frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big) \\ & \left\{\frac{\phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}\Bigg).\end{aligned}$$

To ensure the computational task is defined well, we substitute $\frac{\mu_z^2}{\alpha(1+\alpha^2)}$ with $\frac{2\alpha}{\pi(1+\alpha^2)^2}$ in $\frac{\delta \ell}{\delta \alpha}$, that derived from Equation (3.3) as follows;

$$\mu_z = \frac{\sqrt{2}\alpha}{\sqrt{\pi(1+\alpha^2)}}$$

$$\mu_z^2 = \frac{2\alpha^2}{\pi(1+\alpha^2)}$$

$$\frac{\mu_z^2}{\alpha(1+\alpha^2)} = \frac{\frac{2\alpha^2}{\pi(1+\alpha^2)}}{\alpha(1+\alpha^2)} = \frac{2\alpha}{\pi(1+\alpha^2)^2}$$

Then $\frac{\delta \ell}{\delta \alpha}$ can be rewritten as

$$\begin{aligned}\frac{\partial \ell}{\partial \alpha} \;=\;& \frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\Big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\Big((I-\lambda W)(y-X\beta)\Big) - \\ & \frac{\sqrt{2}}{1+\alpha^2}\Big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T \cdot \mathbb{1} - \frac{n2\alpha}{\pi(1+\alpha^2)^2\sigma_z^2} + \\ & \sum_{i=1}^{n}\Bigg(\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta) + \frac{\sqrt{2}\alpha}{1+\alpha^2} - \frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big) \\ & \left\{\frac{\phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}\Bigg).\end{aligned}$$

Partial derivative of log-likelihood with respect to σ

There are three expressions that give a nonzero solution when differentiating the log-likelihood function with respect to σ:

The first expression is

$$\frac{\partial}{\partial\sigma}\Big(-n\log\sigma\Big) = -n\Big(\frac{1}{\sigma}\Big) = -\frac{n}{\sigma}.$$

The second expression is

$$\frac{\partial}{\partial\sigma}\left(\sum_{i=1}^{n}\log\phi\left(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\right)\right)$$
$$= \frac{\sigma_z}{\sigma^2}\Big((\mu_z\cdot\mathbb{1})^T(I-\lambda W)(y-X\beta)\Big) + \frac{\sigma_z^2}{\sigma^3}\Big(\big((I-\lambda W)(y-X\beta)\big)^T\big((I-\lambda W)(y-X\beta)\big)\Big).$$

Detailed derivation is provided in Appendix C, page 111

The third expression is

$$\frac{\partial}{\partial\sigma}\left(\sum_{i=1}^{n}\log\Phi\left(\alpha\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\right)\right)$$
$$= -\frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}\left(\Big((I-\lambda W)_i(y-X\beta)\Big)\left\{\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}\right).$$

Detailed derivation is provided in Appendix C, page 112

Hence

$$\begin{aligned}\frac{\partial\ell}{\partial\sigma} &= \frac{\sigma_z}{\sigma^2}\Big((\mu_z\cdot\mathbb{1})^T(I-\lambda W)(y-X\beta)\Big) + \frac{\sigma_z^2}{\sigma^3}\Big(\big((I-\lambda W)(y-X\beta)\big)^T\big((I-\lambda W)(y-X\beta)\big)\Big) \\ &\quad -\frac{n}{\sigma} - \frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}\left(\Big((I-\lambda W)_i(y-X\beta)\Big)\left\{\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}\right).\end{aligned}$$

For convenience of reference below, we summarize the key score function before concluding this subsection. To simplify the notation, let

$$A_i = \left\{\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\} \tag{4.7}$$

Then the score function of the log-likelihood can be expressed as follows:

$$\frac{\partial\ell}{\partial\beta} = \frac{\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big) - \frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}\left(X^T(I-\lambda W)_i^T A_i\right), \tag{4.8}$$

$$\begin{aligned}\frac{\partial \ell}{\partial \lambda} &= \frac{\sigma_z}{\sigma}\Big(W\big(y-X\beta\big)\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)-\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}\Bigg(W_i^T\big(y-X\beta\big)A_i\Bigg)\\ &\quad -tr\Big((I-\lambda W)^{-1}W\Big),\end{aligned} \tag{4.9}$$

$$\begin{aligned}\frac{\partial \ell}{\partial \alpha} &= \frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\Big((I-\lambda W)(y-X\beta)\Big)-\frac{n2\alpha}{\pi(1+\alpha^2)^2\sigma_z^2}\\ &\quad -\frac{\sqrt{2}}{1+\alpha^2}\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\cdot\mathbb{1}+\sum_{i=1}^{n}\Bigg(\Bigg(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\\ &\quad +\frac{\sqrt{2}\alpha}{1+\alpha^2}-\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Bigg)A_i\Bigg),\end{aligned} \tag{4.10}$$

$$\begin{aligned}\frac{\partial \ell}{\partial \sigma} &= \frac{\sigma_z^2}{\sigma^3}\Big(\big((I-\lambda W)(y-X\beta)\big)^T\big((I-\lambda W)(y-X\beta)\big)\Big)+\frac{\sigma_z}{\sigma^2}\Big((\mu_z\cdot\mathbb{1})^T(I-\lambda W)(y-X\beta)\Big)\\ &\quad -\frac{n}{\sigma}-\frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}\Bigg(\Bigg((I-\lambda W)_i(y-X\beta)\Bigg)A_i\Bigg).\end{aligned} \tag{4.11}$$

4.4.2 A suggested computational procedure

In this subsection we will propose a new computational procedure for model selection, but some computational issues need to be discussed beforehand.

First, as mentioned in Section 4.3.1 this chapter want to estimate the values of unknown parameters $\hat{\theta} = (\hat{\underset{\sim}{\beta}}, \hat{\lambda}, \hat{\sigma}, \hat{\alpha})$ through penalized maximum likelihood by utilizing adaptive lasso for a purpose of variable selection. The variable selection just involves with the regression coefficient, β which is needed to be sparse. Let $\theta = (\beta^T, \varphi^T)^T$ denoted as $p \times 1$ vector of model parameters, where $\varphi = (\lambda, \sigma, \alpha)$ is $q \times 1$ vector, therefore the adaptive lasso estimation in Equation (4.5) can be rewritten as,

$$\hat{\beta}^* = \arg\min_{\beta}(-\ell(\beta|y, X, W, \varphi)) + \eta\sum_{j=1}^{p-q}\hat{\tau}_j|\beta_j|, \tag{4.12}$$

where

$$\begin{aligned}\ell(\beta|y,X,W,\varphi) &= n\log 2+\sum_{i=1}^{n}\log\phi\Big[\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big]+n\log\sigma_z-n\log\sigma+\\ &\quad \sum_{i=1}^{n}\log\Phi\Big[\alpha\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\Big]+\log|I-\lambda W|,\end{aligned}$$

$$\text{with } \mu_z = \frac{\sqrt{2}\alpha}{\sqrt{\pi}\sqrt{1+\alpha^2}} \text{ and } \sigma_z = \sqrt{1-\frac{2\alpha^2}{\pi(1+\alpha^2)}}.$$

Second, the adaptive lasso estimate in (4.12) can be solved by the LARS algorithm which is computationally efficient to solve lasso problem by path-wise coordinate optimization (Efron et al., 2004). Given the weight, $\hat{\tau}_j = \frac{1}{|(\hat{\beta_s})_j|}$, where $\hat{\beta}_s$ is estimated value of β from SSEM solution, we define $X^{**}_{ij} = \frac{X_{ij}}{\hat{\tau}_j} = X_{ij}|(\hat{\beta_s})_j|$ where $j = 1, 2, ..., p$. Then, by letting $\hat{\tau}_j \beta_j = \tilde{\beta}_j$ where $j = 1, 2, ..., p$, we can change the adaptive Lasso estimate (4.12) into Lasso estimate, that is

$$\hat{\beta}^{**} = \arg\min_{\tilde{\beta}}(-\ell(\tilde{\beta}|y, X^{**}, W, \varphi)) + \eta \sum_{j=1}^{p-q} |\tilde{\beta}_j|, \tag{4.13}$$

where

$$\ell(\tilde{\beta}|y, X^{**}, W, \varphi) = n \log 2 + \sum_{i=1}^{n} \log \phi\Big[\mu_z + \frac{\sigma_z}{\sigma}(I - \lambda W)_i(y - X^{**}\tilde{\beta})\Big] + n \log \sigma_z - n \log \sigma +$$
$$\sum_{i=1}^{n} \log \Phi\Big[\alpha\Big(\mu_z + \frac{\sigma_z}{\sigma}(I - \lambda W)_i(y - X^{**}\tilde{\beta})\Big)\Big] + \log |I - \lambda W|,$$

$$\text{with } \mu_z = \frac{\sqrt{2}\alpha}{\sqrt{\pi}\sqrt{1+\alpha^2}} \text{ and } \sigma_z = \sqrt{1 - \frac{2\alpha^2}{\pi(1+\alpha^2)}}.$$

Path-wise coordinate optimization can be adopted to solve the Lasso problem (4.13). Hence, the estimate of adaptive Lasso is defined by $\hat{\beta}^* = \frac{\hat{\beta}^{**}}{\hat{\tau}} = \hat{\beta}^{**}|\hat{\beta}_s|$.

Third issue is to estimate unknown parameters $\varphi = (\lambda, \sigma, \alpha)$. This can be solved through optimization of profile log-likelihood food each parameters, φ. For this purpose, adaptive lasso estimate, $\hat{\beta}^*$ is needed. The details of the suggested computational procedure are illustrated in Algorithm 2.

Algorithm 2

(i). Determine the skewed spatial error model estimates, $\hat{\theta_s} = (\hat{\beta}_s, \hat{\lambda_s}, \hat{\sigma_s}, \hat{\alpha_s})$ by applying steps (i) to (iv) in Algorithm 1. But here, we used `multiroot` function from `rootSolve` package in `R` (Soetaert, 2014) to solve the MLE of the skewed spatial error model by numerically maximizing the score function.

(ii). Compute the weight of adaptive Lasso, $\hat{\tau}_j = \frac{1}{|\hat{\beta}_s|}$.

(iii). Define $X^{**}_{ij} = \frac{X_{ij}}{\hat{\tau}_j} = X_{ij}|(\hat{\beta_s})_j|$ for $j = 1, 2, \cdots, p$.

(iv). Solve the Lasso problem (4.13) for all tuning parameter, η,

$$\hat{\beta}^{**} = \arg\min_{\tilde{\beta}}(-\ell(\tilde{\beta}|y, X^{**}, W, \varphi)) + \eta \sum_{j=1}^{p-q} |\tilde{\beta}_j|.$$

We use `lbfgs` function from `lbfgs` package in `R` (Coppola et al., 2014) to solve aforementioned equation. To use this R-package, the objective function, $-\ell(\tilde{\beta}|y, X^{**}, W, \varphi)$

and gradient function which is the negative value of the score functions derived in subsection 4.4.1 are two basic inputs that need to be specified for the purpose of optimization with L_1 penalization.

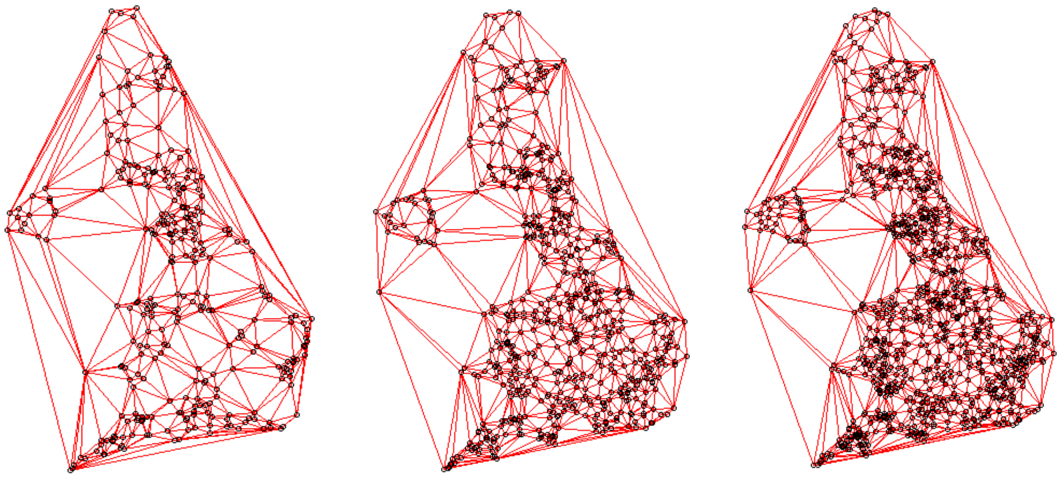
(v). Compute the adaptive Lasso estimate, $\hat{\beta}^* = \frac{\hat{\beta}^{**}}{\hat{\tau}_j} = \hat{\beta}^{**}|\hat{\beta}_s|$ for $j = 1, 2, \cdots, p$.

(vi). Define $\hat{\beta}^*_\eta$ as the best $\hat{\beta}^*$ at the given optimal tuning parameter, η based on the smallest value of BIC.

(vii). Determine the adaptive Lasso estimate for other unknown parameters $\hat{\lambda}^*_\eta, \hat{\sigma}^*_\eta$ and $\hat{\alpha}^*_\eta$ by solving the profile log-likelihood for each parameters φ, given the values of $\hat{\beta}^*_\eta, \lambda_s, \sigma_s$ and α_s using `optimize` function in `R` (Team, 2014).

(viii). Output, $\hat{\theta}^*_\eta = (\hat{\beta}^*_\eta, \hat{\lambda}^*_\eta, \hat{\sigma}^*_\eta, \hat{\alpha}^*_\eta)$.

4.5 Simulation of the performance



(a) spatial data from 300 stations (b) spatial data from 600 stations (c) spatial data from 900 stations

Figure 4.1: Illustration of spatial relationship for different sample sizes

Simulation study is used to demonstrate the performance of the proposed penalized likelihood estimator of skewed spatial error model before it is applied into the UK flooding data for variable selection in index flood estimation model. Similarly with the maximum likelihood estimation in Chapter 3, the properties of statistical estimator need to be examined. However, for penalized likelihood estimation, the sensitivity and specificity test are also essential to be investigated. Here, sensitivity is defined as an ability to determine the zero estimators while specificity is defined as an ability to determine the non-zero estimators. We will give a detailed discuss in Subsection 4.5.1. In this simulation study, the simulation datasets with three different sample sizes of $n = 300, 600, 900$ are considered to investigate the performance of both aforementioned expects. A spatial relationship using contiguity based matrix for these three spatial datasets is illustrated in Figure 4.1. A procedure to simulate the data is similar to what we have done in Chapter 3 (see Subsection 3.5.1) but with the different set of true values which are $\beta_0 = 1, \beta_1 = 1, \beta_2 = -1.5, \beta_3 = 3, \beta_4 = -3, \beta_5 = 0, \beta_6 = 0, \beta_7 = -1.5, \beta_8 = 0, \beta_9 = 0, \beta_{10} = -0.01, \beta_{11} = 0, \beta_{12} = 0, \beta_{13} = 0.4, \beta_{14} = 0, \beta_{15} = 0, \beta_{16} = 0.01, \beta_{17} = 0, \beta_{18} = 0, \beta_{19} = 0, \lambda = 0.4, \sigma = 0.3, \alpha = 0.8$. Then, Algorithm 2 in suggested computational procedure 4.4.2, excluding step (i) has been applied to 100 simulation datasets, with yielding 100 simulation outputs of $\hat{\theta}^*_\eta$ for each considered sample sizes. The optimal tuning parameter for each simulation output is selected based on smallest BIC value from entire regularization path solutions of tuning parameter η from 0 to 3 by 0.1. Finally, the simulation outputs from three different sample sizes are analyzed to examine both aforementioned performance aspects (i.e., sensitivity and specificity test, and properties of estimator) for the proposed penalized likelihood estimation method of skewed spatial error model.

4.5.1 Sensitivity and specificity test

As defined, sensitivity is the ability to determine the zero estimators while specificity is the ability to determine the non-zero estimators. For the purpose of investigating the sensitivity and specificity in variable selection of 19 considered covariates by using our proposed regularization method thought adaptive Lasso procedure, 9 out of 20 regression coefficients $\underset{\sim}{\beta}$ including β_0 are set to be non zero true values while the rest are set to be zeros in the simulation study. This is because, for the variable selection purpose, only regression coefficient $\underset{\sim}{\beta}$ is needed to be sparse. This causes the total number of true non zero estimator, $TNZ = 9$ and the total number of zero estimator, $TZ = 11$ for 20 regression coefficients, $\underset{\sim}{\beta}$. High value of sensitivity and specificity for example greater than 0.5 demonstrate a good performance of the ability to determine zero and non zero coefficient of the proposed penalized likelihood estimator. To calculate the values, firstly the observations of each simulation outputs $\hat{\theta}^*_\eta$ need to be summarized into two by two contingency table where $CEZ, ICEZ, CENZ$ and $ICENZ$ are defined as follows;

True value	Number of observations	
	Correcly estimated	**Incorrectly estimated**
Zero	CEZ	$ICEZ$
Non Zero	$CENZ$	$ICENZ$

CEZ = number of correctly estimated on true zero

$ICEZ$ = number of incorrectly estimated on true zero

$CENZ$ = number of correctly estimated on true non zero

$ICEZ$ = number of incorrectly estimated on true non zero

By having those values, sensitivity and specificity test can be done and specified as;

$$Sensitivity = \frac{CEZ}{CEZ + ICEZ},$$

$$Specificity = \frac{CENZ}{TNZ}.$$

4.5.2 Unbiasedness and consistency of penalized likelihood estimator

Besides the sensitivity and specificity test that have investigated the ability of the penalized likelihood estimation to correctly estimate sparse and non sparse solutions for regression coefficients, we also need to ensure that, the non zero estimators determined by penalized likelihood are complied with the properties of estimators requirement. The details about properties of estimators considered in this research for investigating the performance of proposed statistical methods have been written in Section 3.4. For this purpose, two basic properties of estimators which are unbiasedness and consistency will be examined before the proposed penalized likelihood estimation for skewed spatial error model utilizing the adaptive Lasso regularization technique can be applied into the UK flooding data.

4.6 Adaptive Lasso regularization in the UK skewed spatial Index flood estimation model

4.6.1 Performance of the estimators

Table 4.2: Sensitivity and specificity test results by sample size.

Sample size	**Sensitivity**	**Specificity**
n=300	0.940	0.918
n=600	0.988	0.984
n=900	0.997	0.996

As previous chapter, the simulation of the performance for proposed statistical method in this chapter is also need to be investigated before it can be applied into the UK flooding data of nineteen catchment characteristics for the purpose of index flood estimation. Two aspects as mentioned in Section 4.5 which are sensitivity and specificity test, and unbiasedness and consistency of estimators are examined through simulation study. The simulation method has also been explained in Section 4.5 and the results are reported here. Table 4.2 gives the values of sensitivity and specificity for each different sizes of simulated datasets. The value of sensitivity measures how effective the proposed panelized estimation method for skewed spatial error model in identifying the zero estimators while the value of specificity measures the effectiveness of identifying the non zero estimators. The higher the value of both measurements the better but cannot exceed 1. It is clearly shown that from Table 4.2, as the number of sample size increases the sensitivity and specificity approached to the almost perfect value 1. With the results we can declare our proposed penalized likelihood method has the ability to correctly identify the sparse solutions for the regression coefficients of skewed spatial error model.

Table 4.3: The scale of simulated distribution for non zero parameters $\hat{\theta}$ with its bias and variance when sample size increase

Nonzero	**True**	**Sample size, n=300**				
Parameter	**value**	**Min**	**Mean**	**Max**	**Bias**	**Variance**
$\hat{\beta}_0$	1.00	0.233	1.103	2.203	0.103	0.13096
$\hat{\beta}_1$	1.00	0.909	0.989	1.082	-0.011	0.00148
$\hat{\beta}_2$	-1.50	-1.978	-1.532	-1.182	-0.032	0.02095
$\hat{\beta}_3$	3.00	1.918	2.912	3.769	-0.088	0.10428
$\hat{\beta}_4$	-3.00	-3.382	-2.983	-2.718	0.017	0.01492
$\hat{\beta}_7$	-1.50	-3.366	-1.422	0.000	0.078	0.45807
$\hat{\beta}_{10}$	-0.01	-0.012	-0.010	-0.007	0.000	0.00000
$\hat{\beta}_{13}$	0.40	0.000	0.344	1.099	-0.056	0.11987
$\hat{\beta}_{16}$	0.01	0.001	0.009	0.018	-0.001	0.00001
$\hat{\lambda}$	0.40	0.154	0.378	0.564	-0.022	0.00830
$\hat{\sigma}$	0.30	0.265	0.297	0.328	-0.003	0.00016
$\hat{\alpha}$	0.80	-0.273	0.843	2.033	0.043	0.11361
Nonzero	**True**	**Sample size, n=600**				
Parameter	**value**	**Min**	**Mean**	**Max**	**Bias**	**Variance**
$\hat{\beta}_0$	1.00	0.314	1.009	1.680	0.009	0.06004
$\hat{\beta}_1$	1.00	0.917	0.999	1.055	-0.001	0.00069
$\hat{\beta}_2$	-1.50	-1.714	-1.504	-1.280	-0.004	0.00679
$\hat{\beta}_3$	3.00	2.504	2.971	3.442	-0.029	0.03680
$\hat{\beta}_4$	-3.00	-3.159	-2.999	-2.839	0.001	0.00411
$\hat{\beta}_7$	-1.50	-2.551	-1.486	0.000	0.014	0.19030
$\hat{\beta}_{10}$	-0.01	-0.012	-0.010	-0.007	0.000	0.00000
$\hat{\beta}_{13}$	0.40	0.000	0.381	0.744	-0.019	0.03865
$\hat{\beta}_{16}$	0.01	0.005	0.010	0.016	0.000	0.00000
$\hat{\lambda}$	0.40	0.214	0.384	0.504	-0.016	0.00269
$\hat{\sigma}$	0.30	0.282	0.299	0.324	-0.001	0.00006
$\hat{\alpha}$	0.80	0.088	0.797	1.377	-0.003	0.05316
Nonzero	**True**	**Sample size, n=900**				
Parameter	**value**	**Min**	**Mean**	**Max**	**Bias**	**Variance**
$\hat{\beta}_0$	1.00	0.553	0.999	1.469	-0.001	0.03419
$\hat{\beta}_1$	1.00	0.955	1.000	1.041	0.000	0.00032
$\hat{\beta}_2$	-1.50	-1.670	-1.503	-1.350	-0.003	0.00309
$\hat{\beta}_3$	3.00	2.581	3.014	3.380	0.014	0.02379
$\hat{\beta}_4$	-3.00	-3.113	-2.999	-2.798	0.001	0.00307
$\hat{\beta}_7$	-1.50	-2.204	-1.504	-0.856	-0.004	0.08342
$\hat{\beta}_{10}$	-0.01	-0.011	-0.010	-0.008	0.000	0.00000
$\hat{\beta}_{13}$	0.40	0.000	0.404	0.705	0.004	0.01712
$\hat{\beta}_{16}$	0.01	0.006	0.010	0.014	0.000	0.00000
$\hat{\lambda}$	0.40	0.312	0.398	0.498	-0.002	0.00177
$\hat{\sigma}$	0.30	0.282	0.299	0.312	-0.001	0.00005
$\hat{\alpha}$	0.80	0.335	0.803	1.473	0.003	0.03581

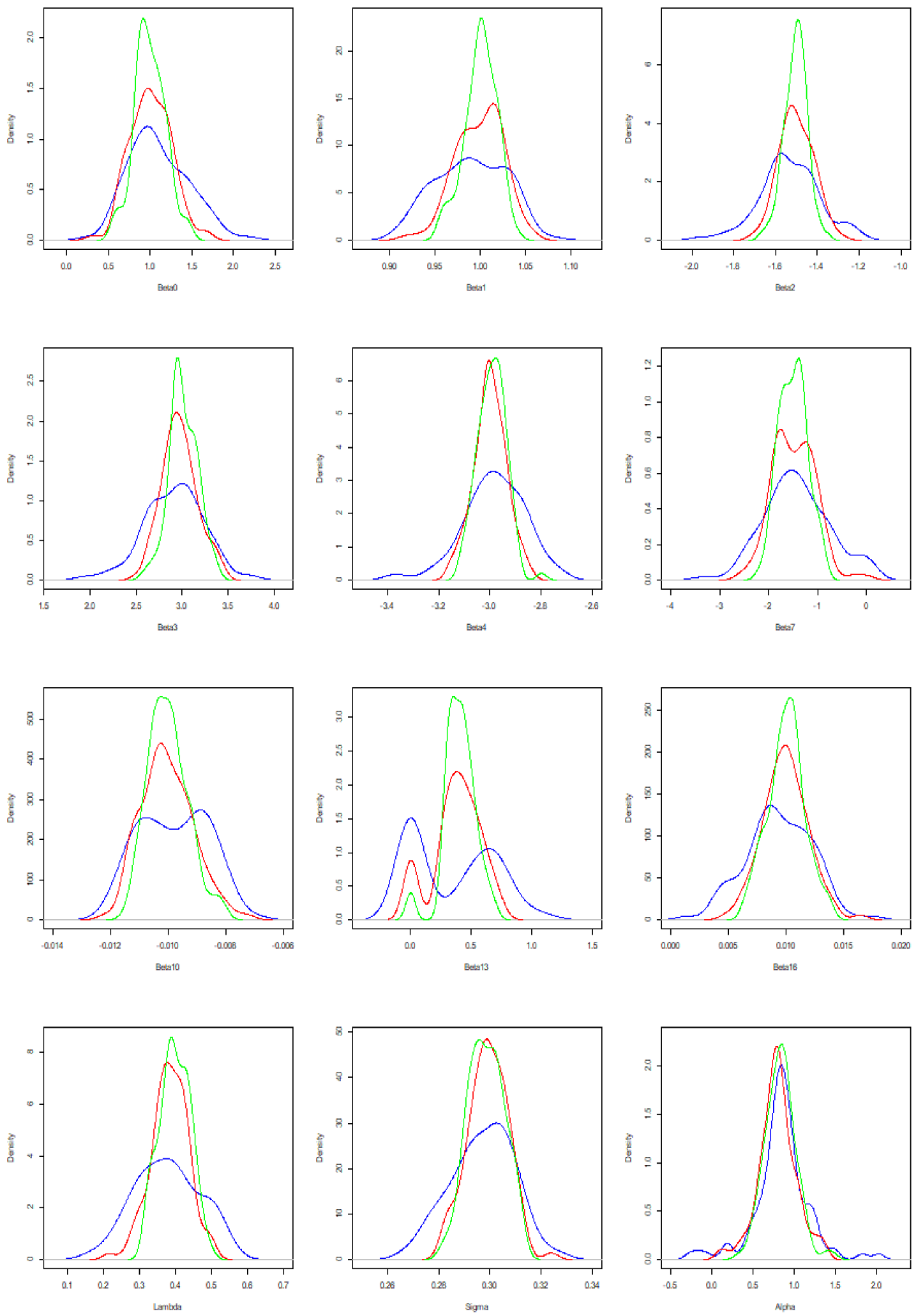


Figure 4.2: Simulated distribution for estimated values of non zero unknown parameters, $\hat{\theta}$ when sample size, n increases (i.e.; blue line for $n = 300$, red line for $n = 600$, green line for $n = 900$).

Besides the sparse solutions, we also want to investigate the performance of the non zero estimators by examining their unbiasedness and consistency. The analysis of both properties for each different size of simulated datasets have been summarized in Table 4.3 while the simulated distribution of each non zero parameters are visualized in Figure 4.2. Table 4.3 shows that, the estimated values of each parameters tend to be centered and be very near to the true values specified in the data generation process (DGP) with small variance that reported together in the table especially for the datasets with sample size of 600 and 900. In addition, each parameter is shown to have a bell shaped distribution visualized in Figure 4.2, with becoming clearer and smoother when the sample size become larger.

By referring to the aforementioned results, it is possible to say the estimators determined by the penalized likelihood estimation method for skewed spatial error model proposed in this chapter are unbiased estimators since its estimated value tends to be centered to the true value and has a small variance with bell shape distribution. However, some reasonable explanations should be given for parameter β_{13} due to its visualization shown in Figure 4.2 has two peaks where the first peak is centered near to the true value specified in the DGP and second peak is centered at zero. The second peak in $\hat{\beta_{13}}$ is due to sparsity of coefficients, $\hat{\theta}^*_\eta$ of selected model in each simulated samples determined by optimal tunning parameter, η based on the lowest BIC value (see Subsection 4.3.3). However, this sparse solution has been examined through specificity and sensitivity test that make the distribution result for parameter $\hat{\beta_{13}}$ shown in Figure 4.2 is acceptable. Lastly, in order to investigate the consistency of our unbiased estimators, results in Table 4.3 and Figure 4.2 are also referred. It is clearly shown that, the unbiased estimators are consistent because they converge to the true parameter values specified in DGP when their variances decrease to zero as sample size increases. With the findings of both expects of simulation performance (i.e.; sensitivity and specificity test, and unbiasedness and consistency of estimators) we now do not hesitate to applied the penalized likelihood estimation procedure proposed by Algorithm 2 into the UK flooding data for the purpose of index flood model estimation and selection where the results will be reported in next Subsection.

4.6.2 Model estimation and selection

In this section, we report all results on application of the proposed panelized likelihood estimation utilizing adaptive Lasso as a tool for regularization procedure of the skewed spatial error model into the UK flooding data. A computational procedure has been detailed described in Algorithm 2 (see Subsection 4.4.2). In this application, both spatial neighboring weight matrices i.e., contiguity based and distance based have been also considered. Table 4.4 gives the adaptive Lasso regularization path solutions of skewed spatial error index flood model with contiguity spatial weight relationship for the scale of tuning parameter, η from 0 to 4 by 0.1 difference. We named these established models as adaptive Lasso non-Gaussian spatial index flood estimation (SIFE) model. From Table 4.4 we have 41 adaptive Lasso non-Gaussian SIFE models to be selected with their BIC values reported together. Apparently, when $\eta = 2.5$ (the row highlighted by pink colour) the selected model has the smallest BIC value 518.37. Shortly afterwards, the BIC values performed an increasing pattern for the rest of the regularization path. Therefore, we decide to run Algorithm 2 again with the scale of tuning parameter, η from 2.4 to 2.6 (highlighted in light blue) by 0.01 difference. The regularization path solutions for this second optimization procedure are summarized in Table 4.5. By referring to this table, the best adaptive Lasso non-Gaussian SIFE model with contiguity spatial weight matrix has the BIC value 518.31 when $\eta = 2.44$.

The similar steps of computational task by applying Algorithm 2 into the UK flooding data for skewed spatial index flood model estimation and selection with distance spatial weight matrix have also been done. The first regularization path solutions for tuning parameter, η from 0 to 6 by 0.1 difference have been summarized in Table 4.6. Due to margin space the table just gave the path solutions up to 4.3. However, starting from selected model, the pink row with lowest BIC value 530.34 when $\eta = 3.7$, an increasing pattern of BIC values has been detected for the rest of path solutions up to $\eta = 6$. As previous Algorithm 2 is run again for second time but here, the scale of tuning parameter, η is from 3.6 to 3.8 by 0.01 difference. The entire regularization path solutions from the second optimization task have been summarized in Table 4.7. Apparently from the table when $\eta = 3.61$, the BIC value is 530.21 which is the best selected model for adaptive Lasso non-Gaussian SIFE model with distance spatial weight matrix.

Table 4.4: Adaptive Lasso non-Gaussian SIFE models with contiguity based spatial weight matrix and tuning parameter η from 0 to 4 by 0.1

η	$\hat{\theta}$											
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$	$\hat{\beta}_9$	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$
0.0	1.180	0.971	-1.209	3.143	-3.674	0.365	-0.00112	-2.570	-0.0206	-0.0072	-0.010	1E-05
0.1	1.174	0.971	-1.248	3.136	-3.642	0.359	-0.00104	-2.422	-0.0194	-0.0064	-0.010	0
0.2	1.162	0.971	-1.273	3.131	-3.613	0.346	-0.00097	-2.372	-0.0182	-0.0058	-0.010	0
0.3	1.144	0.971	-1.285	3.130	-3.590	0.326	-0.00092	-2.346	-0.0170	-0.0054	-0.009	0
0.4	1.126	0.971	-1.296	3.128	-3.567	0.305	-0.00087	-2.319	-0.0159	-0.0049	-0.009	0
0.5	1.109	0.971	-1.308	3.127	-3.543	0.285	-0.00082	-2.293	-0.0147	-0.0044	-0.008	0
0.6	1.091	0.972	-1.319	3.125	-3.520	0.264	-0.00077	-2.266	-0.0136	-0.0040	-0.008	0
0.7	1.074	0.972	-1.331	3.124	-3.497	0.244	-0.00072	-2.240	-0.0124	-0.0035	-0.007	0
0.8	1.056	0.972	-1.343	3.122	-3.474	0.223	-0.00067	-2.213	-0.0113	-0.0030	-0.007	0
0.9	1.038	0.972	-1.354	3.121	-3.451	0.203	-0.00062	-2.187	-0.0101	-0.0026	-0.007	0
1.0	1.021	0.972	-1.366	3.119	-3.428	0.182	-0.00057	-2.161	-0.0090	-0.0021	-0.006	0
1.1	1.003	0.972	-1.377	3.118	-3.405	0.162	-0.00052	-2.134	-0.0078	-0.0016	-0.006	0
1.2	0.985	0.972	-1.389	3.116	-3.382	0.142	-0.00048	-2.108	-0.0067	-0.0012	-0.005	0
1.3	0.968	0.972	-1.401	3.115	-3.359	0.121	-0.00043	-2.081	-0.0056	-0.0007	-0.005	0
1.4	0.953	0.972	-1.411	3.112	-3.336	0.102	-0.00038	-2.056	-0.0044	-0.0002	-0.004	0
1.5	0.951	0.972	-1.420	3.107	-3.325	0.086	-0.00034	-2.033	-0.0033	0	-0.004	0
1.6	0.959	0.972	-1.427	3.102	-3.326	0.073	-0.00030	-2.013	-0.0022	0	-0.003	0
1.7	0.967	0.971	-1.434	3.098	-3.328	0.060	-0.00027	-1.991	-0.0012	0	-0.003	0
1.8	0.976	0.971	-1.441	3.093	-3.329	0.048	-0.00023	-1.969	-0.0002	0	-0.003	0
1.9	0.987	0.970	-1.448	3.089	-3.331	0.033	-0.00019	-1.946	0	0	-0.003	0
2.0	0.998	0.970	-1.456	3.085	-3.332	0.017	-0.00015	-1.923	0	0	-0.003	0
2.1	1.009	0.970	-1.465	3.081	-3.334	0.001	-0.00011	-1.899	0	0	-0.003	0
2.2	1.003	0.969	-1.468	3.077	-3.334	0	-0.00008	-1.879	0	0	-0.003	0
2.3	0.996	0.969	-1.472	3.074	-3.334	0	-0.00004	-1.858	0	0	-0.003	0
2.4	0.994	0.968	-1.473	3.071	-3.334	0	-0.00001	-1.838	0	0	-0.003	0
2.5	1.000	0.967	-1.474	3.068	-3.334	0	0	-1.826	0	0	-0.003	0
2.6	1.005	0.967	-1.474	3.065	-3.333	0	0	-1.818	0	0	-0.003	0
2.7	1.011	0.966	-1.475	3.062	-3.333	0	0	-1.809	0	0	-0.003	0
2.8	1.017	0.965	-1.476	3.059	-3.333	0	0	-1.800	0	0	-0.003	0
2.9	1.022	0.964	-1.476	3.056	-3.333	0	0	-1.792	0	0	-0.003	0
3.0	1.028	0.963	-1.477	3.052	-3.332	0	0	-1.783	0	0	-0.003	0
3.1	1.034	0.963	-1.478	3.049	-3.332	0	0	-1.774	0	0	-0.003	0
3.2	1.039	0.962	-1.478	3.046	-3.332	0	0	-1.766	0	0	-0.003	0
3.3	1.045	0.961	-1.479	3.043	-3.332	0	0	-1.757	0	0	-0.003	0
3.4	1.050	0.960	-1.480	3.040	-3.331	0	0	-1.748	0	0	-0.003	0
3.5	1.056	0.960	-1.480	3.037	-3.331	0	0	-1.740	0	0	-0.003	0
3.6	1.062	0.959	-1.481	3.034	-3.331	0	0	-1.731	0	0	-0.003	0
3.7	1.067	0.958	-1.482	3.031	-3.331	0	0	-1.722	0	0	-0.003	0
3.8	1.073	0.957	-1.482	3.028	-3.331	0	0	-1.714	0	0	-0.003	0
3.9	1.079	0.956	-1.483	3.024	-3.330	0	0	-1.705	0	0	-0.003	0
4.0	1.084	0.956	-1.484	3.021	-3.330	0	0	-1.697	0	0	-0.003	0

η	$\hat{\theta}$											
	$\hat{\beta}_{12}$	$\hat{\beta}_{13}$	$\hat{\beta}_{14}$	$\hat{\beta}_{15}$	$\hat{\beta}_{16}$	$\hat{\beta}_{17}$	$\hat{\beta}_{18}$	$\hat{\beta}_{19}$	$\hat{\lambda}$	$\hat{\sigma}$	$\hat{\alpha}$	**BIC**
0.0	-2E-05	0.689	0.0133	-0.031	0.027	7E-05	-0.170	0.01	0.381	0.349	0.841	**572.41**
0.1	0	0.688	0.0126	-0.029	0.027	3E-05	-0.158	0	0.378	0.349	0.840	**553.45**
0.2	0	0.689	0.0118	-0.027	0.026	0	-0.144	0	0.376	0.349	0.838	**547.32**
0.3	0	0.692	0.0110	-0.026	0.025	0	-0.132	0	0.376	0.350	0.836	**547.55**
0.4	0	0.695	0.0101	-0.025	0.024	0	-0.120	0	0.376	0.350	0.834	**547.89**
0.5	0	0.698	0.0093	-0.024	0.022	0	-0.107	0	0.376	0.350	0.832	**548.31**
0.6	0	0.701	0.0085	-0.023	0.021	0	-0.095	0	0.376	0.350	0.830	**548.83**
0.7	0	0.705	0.0077	-0.022	0.020	0	-0.083	0	0.376	0.350	0.828	**549.45**
0.8	0	0.708	0.0068	-0.021	0.019	0	-0.070	0	0.376	0.350	0.826	**550.16**
0.9	0	0.711	0.0060	-0.020	0.018	0	-0.058	0	0.376	0.351	0.824	**550.97**
1.0	0	0.714	0.0052	-0.019	0.017	0	-0.046	0	0.376	0.351	0.822	**551.87**
1.1	0	0.717	0.0044	-0.017	0.016	0	-0.033	0	0.377	0.351	0.820	**552.86**
1.2	0	0.721	0.0035	-0.016	0.015	0	-0.021	0	0.377	0.351	0.818	**553.94**
1.3	0	0.724	0.0027	-0.015	0.013	0	-0.009	0	0.378	0.352	0.815	**555.12**
1.4	0	0.725	0.0019	-0.014	0.012	0	0	0	0.379	0.352	0.812	**549.94**
1.5	0	0.721	0.0011	-0.013	0.011	0	0	0	0.380	0.352	0.809	**544.58**
1.6	0	0.716	0.0003	-0.012	0.010	0	0	0	0.381	0.353	0.805	**545.50**
1.7	0	0.710	0	-0.011	0.009	0	0	0	0.383	0.353	0.802	**539.88**
1.8	0	0.704	0	-0.010	0.008	0	0	0	0.384	0.353	0.798	**540.56**
1.9	0	0.698	0	-0.008	0.008	0	0	0	0.386	0.353	0.796	**534.76**
2.0	0	0.693	0	-0.007	0.008	0	0	0	0.388	0.354	0.793	**535.32**
2.1	0	0.687	0	-0.005	0.008	0	0	0	0.390	0.354	0.791	**535.91**
2.2	0	0.682	0	-0.003	0.007	0	0	0	0.391	0.354	0.789	**530.05**
2.3	0	0.677	0	-0.001	0.007	0	0	0	0.393	0.354	0.787	**530.57**
2.4	0	0.672	0	0	0.007	0	0	0	0.394	0.354	0.786	**524.61**
2.5	0	0.665	0	0	0.007	0	0	0	0.394	0.354	0.785	**518.37**
2.6	0	0.657	0	0	0.007	0	0	0	0.395	0.354	0.785	**518.45**
2.7	0	0.649	0	0	0.007	0	0	0	0.395	0.354	0.784	**518.54**
2.8	0	0.642	0	0	0.007	0	0	0	0.395	0.354	0.784	**518.64**
2.9	0	0.634	0	0	0.007	0	0	0	0.395	0.354	0.783	**518.74**
3.0	0	0.626	0	0	0.007	0	0	0	0.396	0.354	0.783	**518.84**
3.1	0	0.618	0	0	0.007	0	0	0	0.396	0.354	0.782	**518.94**
3.2	0	0.611	0	0	0.007	0	0	0	0.396	0.354	0.782	**519.05**
3.3	0	0.603	0	0	0.007	0	0	0	0.397	0.354	0.781	**519.16**
3.4	0	0.595	0	0	0.007	0	0	0	0.397	0.354	0.781	**519.28**
3.5	0	0.588	0	0	0.007	0	0	0	0.397	0.355	0.780	**519.40**
3.6	0	0.580	0	0	0.007	0	0	0	0.398	0.355	0.780	**519.52**
3.7	0	0.572	0	0	0.007	0	0	0	0.398	0.355	0.780	**519.64**
3.8	0	0.565	0	0	0.007	0	0	0	0.398	0.355	0.779	**519.77**
3.9	0	0.557	0	0	0.006	0	0	0	0.399	0.355	0.779	**519.90**
4.0	0	0.549	0	0	0.006	0	0	0	0.399	0.355	0.778	**520.04**

Table 4.5: Adaptive Lasso non-Gaussian SIFE models with contiguity based spatial weight matrix and tuning parameter η from 2.4 to 2.6 by 0.01

η	$\hat{\theta}$											
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$	$\hat{\beta}_9$	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$
2.40	0.994	0.968	-1.473	3.071	-3.334	0	-1E-05	-1.838	0	0	-0.003	0
2.41	0.995	0.968	-1.473	3.071	-3.334	0	-7E-06	-1.837	0	0	-0.003	0
2.42	0.995	0.968	-1.473	3.070	-3.334	0	-5E-06	-1.835	0	0	-0.003	0
2.43	0.996	0.968	-1.473	3.070	-3.334	0	-2E-06	-1.833	0	0	-0.003	0
2.44	0.996	0.968	-1.473	3.070	-3.334	0	0	-1.832	0	0	-0.003	0
2.45	0.997	0.968	-1.474	3.070	-3.334	0	0	-1.831	0	0	-0.003	0
2.46	0.997	0.968	-1.474	3.069	-3.334	0	0	-1.830	0	0	-0.003	0
2.47	0.998	0.968	-1.474	3.069	-3.334	0	0	-1.829	0	0	-0.003	0
2.48	0.999	0.968	-1.474	3.069	-3.334	0	0	-1.828	0	0	-0.003	0
2.49	0.999	0.967	-1.474	3.068	-3.334	0	0	-1.827	0	0	-0.003	0
2.50	1.000	0.967	-1.474	3.068	-3.334	0	0	-1.826	0	0	-0.003	0
2.51	1.000	0.967	-1.474	3.068	-3.334	0	0	-1.826	0	0	-0.003	0
2.52	1.001	0.967	-1.474	3.067	-3.334	0	0	-1.825	0	0	-0.003	0
2.53	1.001	0.967	-1.474	3.067	-3.334	0	0	-1.824	0	0	-0.003	0
2.54	1.002	0.967	-1.474	3.067	-3.333	0	0	-1.823	0	0	-0.003	0
2.55	1.003	0.967	-1.474	3.066	-3.333	0	0	-1.822	0	0	-0.003	0
2.56	1.003	0.967	-1.474	3.066	-3.333	0	0	-1.821	0	0	-0.003	0
2.57	1.004	0.967	-1.474	3.066	-3.333	0	0	-1.820	0	0	-0.003	0
2.58	1.004	0.967	-1.474	3.066	-3.333	0	0	-1.819	0	0	-0.003	0
2.59	1.005	0.967	-1.474	3.065	-3.333	0	0	-1.819	0	0	-0.003	0
2.60	1.005	0.967	-1.474	3.065	-3.333	0	0	-1.818	0	0	-0.003	0

$\hat{\theta}$											
$\hat{\beta}_{12}$	$\hat{\beta}_{13}$	$\hat{\beta}_{14}$	$\hat{\beta}_{15}$	$\hat{\beta}_{16}$	$\hat{\beta}_{17}$	$\hat{\beta}_{18}$	$\hat{\beta}_{19}$	$\hat{\lambda}$	$\hat{\sigma}$	$\hat{\alpha}$	BIC
0	0.672	0	0	0.007	0	0	0	0.394	0.354	0.786	524.61
0	0.671	0	0	0.007	0	0	0	0.394	0.354	0.786	524.63
0	0.670	0	0	0.007	0	0	0	0.394	0.354	0.786	524.65
0	0.670	0	0	0.007	0	0	0	0.394	0.354	0.786	524.67
0	0.669	0	0	0.007	0	0	0	0.394	0.354	0.785	518.31
0	0.668	0	0	0.007	0	0	0	0.394	0.354	0.785	518.32
0	0.668	0	0	0.007	0	0	0	0.394	0.354	0.785	518.33
0	0.667	0	0	0.007	0	0	0	0.394	0.354	0.785	518.34
0	0.666	0	0	0.007	0	0	0	0.394	0.354	0.785	518.35
0	0.665	0	0	0.007	0	0	0	0.394	0.354	0.785	518.36
0	0.665	0	0	0.007	0	0	0	0.394	0.354	0.785	518.37
0	0.664	0	0	0.007	0	0	0	0.394	0.354	0.785	518.37
0	0.663	0	0	0.007	0	0	0	0.394	0.354	0.785	518.38
0	0.662	0	0	0.007	0	0	0	0.394	0.354	0.785	518.39
0	0.662	0	0	0.007	0	0	0	0.394	0.354	0.785	518.40
0	0.661	0	0	0.007	0	0	0	0.394	0.354	0.785	518.41
0	0.660	0	0	0.007	0	0	0	0.394	0.354	0.785	518.42
0	0.659	0	0	0.007	0	0	0	0.394	0.354	0.785	518.43
0	0.658	0	0	0.007	0	0	0	0.394	0.354	0.785	518.43
0	0.658	0	0	0.007	0	0	0	0.394	0.354	0.785	518.44
0	0.657	0	0	0.007	0	0	0	0.395	0.354	0.785	518.45

Table 4.6: Adaptive Lasso non-Gaussian SIFE models with distance based spatial weight matrix and tuning parameter η from 0 to 4.3 by 0.1

η	$\hat{\theta}$																							
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$	$\hat{\beta}_9$	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{13}$	$\hat{\beta}_{14}$	$\hat{\beta}_{15}$	$\hat{\beta}_{16}$	$\hat{\beta}_{17}$	$\hat{\beta}_{18}$	$\hat{\beta}_{19}$	$\hat{\lambda}$	$\hat{\sigma}$	$\hat{\alpha}$	**BIC**
0.0	1.19	0.97	-1.17	3.30	-3.64	0.414	-1E-03	-2.29	-0.017	-0.006	-0.011	1E-04	-4E-06	0.617	0.015	-0.042	0.024	8E-05	-0.182	-0.001	0.43	0.35	0.94	**579.76**
0.1	1.17	0.96	-1.21	3.30	-3.58	0.411	-1E-03	-2.27	-0.016	-0.005	-0.011	8E-05	0	0.621	0.014	-0.041	0.024	4E-05	-0.166	0	0.42	0.35	0.94	**567.16**
0.2	1.16	0.96	-1.25	3.30	-3.53	0.406	-1E-03	-2.24	-0.015	-0.003	-0.011	2E-05	0	0.625	0.014	-0.039	0.024	0	-0.151	0	0.42	0.35	0.94	**561.20**
0.3	1.14	0.96	-1.26	3.30	-3.50	0.390	-1E-03	-2.22	-0.013	-0.003	-0.010	0	0	0.628	0.013	-0.039	0.022	0	-0.139	0	0.42	0.35	0.94	**555.12**
0.4	1.12	0.96	-1.27	3.30	-3.47	0.373	-1E-03	-2.20	-0.012	-0.002	-0.010	0	0	0.629	0.012	-0.038	0.021	0	-0.127	0	0.42	0.35	0.94	**555.44**
0.5	1.10	0.96	-1.28	3.30	-3.44	0.356	-9E-04	-2.17	-0.011	-0.002	-0.009	0	0	0.630	0.011	-0.038	0.020	0	-0.115	0	0.42	0.35	0.94	**555.86**
0.6	1.08	0.96	-1.29	3.30	-3.42	0.339	-9E-04	-2.15	-0.009	-0.001	-0.009	0	0	0.631	0.011	-0.037	0.019	0	-0.103	0	0.42	0.35	0.93	**556.37**
0.7	1.06	0.96	-1.30	3.30	-3.39	0.322	-9E-04	-2.12	-0.008	-0.001	-0.009	0	0	0.632	0.010	-0.036	0.018	0	-0.092	0	0.42	0.35	0.93	**556.98**
0.8	1.03	0.96	-1.31	3.30	-3.36	0.306	-8E-04	-2.10	-0.006	-2E-04	-0.008	0	0	0.633	0.009	-0.036	0.016	0	-0.080	0	0.42	0.35	0.93	**557.67**
0.9	1.03	0.96	-1.32	3.30	-3.35	0.291	-8E-04	-2.08	-0.005	0	-0.008	0	0	0.632	0.008	-0.035	0.015	0	-0.069	0	0.42	0.35	0.93	**551.97**
1.0	1.03	0.96	-1.33	3.29	-3.35	0.278	-8E-04	-2.06	-0.004	0	-0.007	0	0	0.630	0.008	-0.034	0.014	0	-0.058	0	0.42	0.35	0.93	**552.63**
1.1	1.03	0.96	-1.33	3.29	-3.35	0.265	-7E-04	-2.04	-0.002	0	-0.007	0	0	0.627	0.007	-0.034	0.013	0	-0.048	0	0.42	0.36	0.92	**553.36**
1.2	1.03	0.96	-1.34	3.29	-3.36	0.253	-7E-04	-2.02	-0.001	0	-0.007	0	0	0.625	0.006	-0.033	0.012	0	-0.037	0	0.42	0.36	0.92	**554.15**
1.3	1.03	0.96	-1.35	3.29	-3.36	0.238	-6E-04	-1.99	0	0	-0.006	0	0	0.623	0.006	-0.032	0.011	0	-0.026	0	0.42	0.36	0.92	**548.57**
1.4	1.04	0.96	-1.36	3.29	-3.36	0.221	-6E-04	-1.97	0	0	-0.006	0	0	0.622	0.005	-0.031	0.011	0	-0.016	0	0.42	0.36	0.92	**549.24**
1.5	1.04	0.96	-1.37	3.29	-3.36	0.203	-6E-04	-1.95	0	0	-0.006	0	0	0.621	0.004	-0.030	0.010	0	-0.006	0	0.42	0.36	0.92	**549.96**
1.6	1.05	0.96	-1.38	3.28	-3.36	0.186	-5E-04	-1.93	0	0	-0.005	0	0	0.618	0.004	-0.028	0.010	0	0	0	0.43	0.36	0.92	**544.27**
1.7	1.06	0.96	-1.38	3.28	-3.36	0.171	-5E-04	-1.91	0	0	-0.005	0	0	0.612	0.003	-0.027	0.010	0	0	0	0.43	0.36	0.91	**544.89**
1.8	1.07	0.96	-1.39	3.28	-3.36	0.155	-5E-04	-1.89	0	0	-0.005	0	0	0.606	0.002	-0.026	0.010	0	0	0	0.43	0.36	0.91	**545.55**
1.9	1.08	0.96	-1.40	3.27	-3.36	0.140	-4E-04	-1.87	0	0	-0.004	0	0	0.601	0.002	-0.024	0.009	0	0	0	0.43	0.36	0.91	**546.25**
2.0	1.09	0.96	-1.41	3.27	-3.36	0.124	-4E-04	-1.85	0	0	-0.004	0	0	0.595	0.001	-0.023	0.009	0	0	0	0.43	0.36	0.91	**546.98**
2.1	1.11	0.96	-1.42	3.26	-3.37	0.109	-4E-04	-1.83	0	0	-0.003	0	0	0.589	4E-04	-0.022	0.009	0	0	0	0.43	0.36	0.90	**547.74**
2.2	1.12	0.96	-1.42	3.26	-3.37	0.094	-3E-04	-1.81	0	0	-0.003	0	0	0.583	0	-0.020	0.009	0	0	0	0.43	0.36	0.90	**542.05**
2.3	1.13	0.96	-1.43	3.26	-3.37	0.078	-3E-04	-1.79	0	0	-0.003	0	0	0.576	0	-0.019	0.008	0	0	0	0.43	0.36	0.90	**542.55**
2.4	1.14	0.96	-1.44	3.25	-3.37	0.063	-3E-04	-1.77	0	0	-0.003	0	0	0.569	0	-0.018	0.008	0	0	0	0.44	0.36	0.90	**543.08**
2.5	1.15	0.96	-1.45	3.25	-3.37	0.048	-3E-04	-1.74	0	0	-0.003	0	0	0.562	0	-0.017	0.008	0	0	0	0.44	0.36	0.90	**543.62**
2.6	1.17	0.96	-1.45	3.25	-3.37	0.032	-2E-04	-1.72	0	0	-0.003	0	0	0.555	0	-0.015	0.008	0	0	0	0.44	0.36	0.89	**544.18**
2.7	1.18	0.96	-1.46	3.24	-3.37	0.017	-2E-04	-1.70	0	0	-0.003	0	0	0.547	0	-0.014	0.008	0	0	0	0.44	0.36	0.89	**544.77**
2.8	1.19	0.96	-1.47	3.24	-3.37	0.001	-2E-04	-1.68	0	0	-0.003	0	0	0.540	0	-0.013	0.007	0	0	0	0.44	0.36	0.89	**545.37**
2.9	1.19	0.96	-1.47	3.24	-3.37	0	-1E-04	-1.66	0	0	-0.003	0	0	0.534	0	-0.011	0.007	0	0	0	0.44	0.36	0.89	**539.50**
3.0	1.19	0.96	-1.47	3.24	-3.37	0	-1E-04	-1.64	0	0	-0.003	0	0	0.527	0	-1E-02	0.007	0	0	0	0.45	0.36	0.89	**540.00**
3.1	1.18	0.95	-1.48	3.23	-3.37	0	-8E-05	-1.62	0	0	-0.003	0	0	0.520	0	-8E-03	0.007	0	0	0	0.45	0.36	0.88	**540.53**
3.2	1.18	0.95	-1.48	3.23	-3.37	0	-5E-05	-1.60	0	0	-0.003	0	0	0.514	0	-6E-03	0.007	0	0	0	0.45	0.36	0.88	**541.06**
3.3	1.17	0.95	-1.48	3.23	-3.37	0	-2E-05	-1.58	0	0	-0.003	0	0	0.507	0	-5E-03	0.006	0	0	0	0.45	0.36	0.88	**541.62**
3.4	1.17	0.95	-1.48	3.22	-3.37	0	0	-1.56	0	0	-0.003	0	0	0.500	0	-3E-03	0.006	0	0	0	0.45	0.36	0.88	**535.79**
3.5	1.17	0.95	-1.49	3.22	-3.37	0	0	-1.55	0	0	-0.003	0	0	0.492	0	-2E-03	0.006	0	0	0	0.45	0.36	0.88	**536.17**
3.6	1.16	0.95	-1.49	3.21	-3.37	0	0	-1.54	0	0	-0.003	0	0	0.483	0	-7E-05	0.006	0	0	0	0.46	0.36	0.88	**536.56**
3.7	1.17	0.95	-1.49	3.21	-3.37	0	0	-1.53	0	0	-0.003	0	0	0.475	0	0	0.006	0	0	0	0.46	0.36	0.88	**530.34**
3.8	1.18	0.95	-1.49	3.21	-3.37	0	0	-1.52	0	0	-0.003	0	0	0.47	0	0	0.006	0	0	0	0.46	0.36	0.88	**530.50**
3.9	1.18	0.95	-1.49	3.20	-3.37	0	0	-1.51	0	0	-0.003	0	0	0.46	0	0	0.006	0	0	0	0.46	0.36	0.88	**530.65**
4.0	1.19	0.95	-1.49	3.20	-3.37	0	0	-1.50	0	0	-0.003	0	0	0.449	0	0	0.006	0	0	0	0.46	0.36	0.88	**530.81**
4.1	1.20	0.95	-1.49	3.20	-3.37	0	0	-1.49	0	0	-0.003	0	0	0.441	0	0	0.006	0	0	0	0.46	0.36	0.88	**530.98**
4.2	1.20	0.95	-1.50	3.20	-3.37	0	0	-1.48	0	0	-0.003	0	0	0.432	0	0	0.006	0	0	0	0.46	0.36	0.87	**531.14**
4.3	1.21	0.95	-1.50	3.19	-3.37	0	0	-1.47	0	0	-0.003	0	0	0.424	0	0	0.006	0	0	0	0.46	0.36	0.87	**531.32**

Table 4.7: Adaptive Lasso non-Gaussian SIFE models with distance based spatial weight matrix and tuning parameter η from 3.6 to 3.8 by 0.01

η	$\hat{\theta}$											
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$	$\hat{\beta}_9$	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$
3.60	1.164	0.952	-1.491	3.214	-3.373	0	0	-1.542	0	0	-0.003	0
3.61	1.164	0.952	-1.491	3.214	-3.373	0	0	-1.541	0	0	-0.003	0
3.62	1.165	0.952	-1.491	3.214	-3.373	0	0	-1.540	0	0	-0.003	0
3.63	1.166	0.952	-1.491	3.213	-3.373	0	0	-1.539	0	0	-0.003	0
3.64	1.166	0.951	-1.491	3.213	-3.372	0	0	-1.537	0	0	-0.003	0
3.65	1.167	0.951	-1.491	3.213	-3.372	0	0	-1.536	0	0	-0.003	0
3.66	1.168	0.951	-1.491	3.212	-3.372	0	0	-1.535	0	0	-0.003	0
3.67	1.168	0.951	-1.491	3.212	-3.372	0	0	-1.534	0	0	-0.003	0
3.68	1.169	0.951	-1.491	3.212	-3.372	0	0	-1.533	0	0	-0.003	0
3.69	1.169	0.951	-1.491	3.211	-3.372	0	0	-1.532	0	0	-0.003	0
3.70	1.170	0.951	-1.491	3.211	-3.372	0	0	-1.531	0	0	-0.003	0
3.71	1.171	0.951	-1.492	3.211	-3.372	0	0	-1.530	0	0	-0.003	0
3.72	1.171	0.951	-1.492	3.210	-3.372	0	0	-1.529	0	0	-0.003	0
3.73	1.172	0.951	-1.492	3.210	-3.372	0	0	-1.528	0	0	-0.003	0
3.74	1.173	0.951	-1.492	3.210	-3.372	0	0	-1.527	0	0	-0.003	0
3.75	1.173	0.951	-1.492	3.210	-3.372	0	0	-1.526	0	0	-0.003	0
3.76	1.174	0.951	-1.492	3.209	-3.372	0	0	-1.525	0	0	-0.003	0
3.77	1.174	0.950	-1.492	3.209	-3.372	0	0	-1.524	0	0	-0.003	0
3.78	1.175	0.950	-1.492	3.209	-3.372	0	0	-1.523	0	0	-0.003	0
3.79	1.176	0.950	-1.492	3.208	-3.372	0	0	-1.522	0	0	-0.003	0
3.80	1.176	0.950	-1.492	3.208	-3.372	0	0	-1.521	0	0	-0.003	0

$\hat{\theta}$											
$\hat{\beta}_{12}$	$\hat{\beta}_{13}$	$\hat{\beta}_{14}$	$\hat{\beta}_{15}$	$\hat{\beta}_{16}$	$\hat{\beta}_{17}$	$\hat{\beta}_{18}$	$\hat{\beta}_{19}$	$\hat{\lambda}$	$\hat{\sigma}$	$\hat{\alpha}$	BIC
0	0.483	0	-7E-05	0.006	0	0	0	0.455	0.359	0.877	536.56
0	0.482	0	0	0.006	0	0	0	0.455	0.359	0.877	530.21
0	0.481	0	0	0.006	0	0	0	0.455	0.359	0.876	530.23
0	0.481	0	0	0.006	0	0	0	0.455	0.359	0.876	530.24
0	0.480	0	0	0.006	0	0	0	0.455	0.359	0.876	530.26
0	0.479	0	0	0.006	0	0	0	0.455	0.359	0.876	530.27
0	0.478	0	0	0.006	0	0	0	0.455	0.359	0.876	530.29
0	0.477	0	0	0.006	0	0	0	0.455	0.359	0.876	530.30
0	0.476	0	0	0.006	0	0	0	0.456	0.359	0.876	530.31
0	0.476	0	0	0.006	0	0	0	0.456	0.359	0.876	530.33
0	0.475	0	0	0.006	0	0	0	0.456	0.359	0.876	530.34
0	0.474	0	0	0.006	0	0	0	0.456	0.359	0.876	530.36
0	0.473	0	0	0.006	0	0	0	0.456	0.359	0.876	530.37
0	0.472	0	0	0.006	0	0	0	0.456	0.359	0.876	530.39
0	0.471	0	0	0.006	0	0	0	0.456	0.359	0.876	530.41
0	0.470	0	0	0.006	0	0	0	0.456	0.359	0.876	530.42
0	0.470	0	0	0.006	0	0	0	0.456	0.359	0.876	530.44
0	0.469	0	0	0.006	0	0	0	0.456	0.359	0.876	530.45
0	0.468	0	0	0.006	0	0	0	0.456	0.359	0.876	530.47
0	0.467	0	0	0.006	0	0	0	0.456	0.359	0.876	530.48
0	0.466	0	0	0.006	0	0	0	0.456	0.359	0.876	530.50

Table 4.8: Comparison among the best SIFE models with contiguity based weight matrix that established by using proposed statistical methods in Chapter 3 and Chapter 4.

Covariates	Coefficients	Model 1			Model 2			Model 3
		Est.	S.E	P-val	Est.	S.E	P-val	Estimate
	β_0	0.9333	0.4098	0.0226	0.8637	0.3022	0.0044	0.99635
ln(AREA)	β_1	0.9327	0.0243	0.0000	0.9861	0.0272	0.0000	0.96782
1000/SAAR	β_2	-1.3688	0.1643	0.0000	-1.4576	0.1230	0.0000	-1.47344
ln(FARL)	β_3	3.0638	0.2714	0.0000	3.1431	0.2686	0.0000	3.06991
(BFIHOST)^2	β_4	-3.2855	0.1163	0.0000	-3.335	0.1104	0.0000	-3.33371
PROPWET	β_5	0.5357	0.3062	0.0802				0.00000
DPSBAR	β_6	-0.0010	0.0005	0.0644				0.00000
FPEXT	β_7	-2.3469	0.5349	0.0000	-2.0495	0.4966	0.0001	-1.83157
RMED-1D	β_8	0.0113	0.0044	0.0102				0.00000
SPRHOST	β_9							0.00000
LDP	β_{10}	-0.0028	0.0010	0.0054	-0.0037	0.0010	0.0002	-0.00327
ALTBAR	β_{11}							0.00000
ASPBAR	β_{12}							0.00000
ASPVAR	β_{13}				0.8536	0.2203	0.0003	0.66926
DPLBAR	β_{14}							0.00000
RMED-1H	β_{15}							0.00000
RMED-2D	β_{16}				0.0073	0.0028	0.0098	0.00679
$SAAR_{4170}$	β_{17}							0.00000
FPLOC	β_{18}							0.00000
FPDBAR	β_{19}							0.00000
	λ	0.3859	0.0575	0.0000	0.3848	0.0565	0.0000	0.39403
	σ	0.3572	0.0106	0.0000	0.3536	0.0105	0.0000	0.35416
	α	0.7861	0.2359	0.0008	0.7982	0.2347	0.0006	0.78540
AIC		**496.90**			**485.94**			**484.96**

Note: **Pink empty cells** represent the unused covariates while **blue empty cells** for non significant covariates in model fitting; **nG-SIFE** = non-Gaussian SIFE model; **ALnG-SIFE** = Adaptive Lasso non-Gaussian SIFE model; **CSWM** = Contiguity based Spatial weight matrix.
Model 1 = The best nG-SIFE model with CSWM developed by applying proposed statistical method in Chapter 3 to the UK flooding data of 10 covariates.
Model 2 = The best nG-SIFE model with CSWM developed by applying proposed statistical method in Chapter 3 to the UK flooding data of 19 covariates.
Model 3 = The best ALnG-SIFE model with CSWM developed by applying proposed statistical method in Chapter 4 to the UK flooding data of 19 covariates.

Both selected models in Table 4.5 and 4.7 that have been highlighted by pink colour are declared as a best model of Chapter 4 for each considered spatial weight matrix. These best adaptive Lasso non-Gaussian SIFE models represented as Model 3 and Model 6 in Table 4.8 and 4.9 are then compared with the best non-Gaussian SIFE models from Chapter 3 represented as Model 1 and Model 4 in the same tables respectively. Note that, Model 1 and Model 4 are developed by using the maximum likelihood computational procedure detailed in Algorithm 1 into the UK flooding data with 10 selected catchment characteristics being the covariates. Meanwhile, Model 3 and Model 6 are developed by applying penalized maximum likelihood that utilizes adaptive Lasso as regularization tool into the UK flooding data with 19 catchment characteristics where its computational optimization procedure is detailed in Algorithm 2. Further, for a more comprehensive comparison study, we also develop Model 2 and Model 5 by using the same computational optimization procedure as in Model 1 and Model 4, but it is applied into the UK flooding data with 19 available catchment characteristics being the covariates. The similar model estimation and selection described in Subsection 3.5.2 have been used in establishing Model 2 and Model 5, where the non significant covariates from 19 fitted covariates are removed one by one in the model fitting process by using Algorithm 1, until the best model with lowest AIC value and all significant covariates being obtained. Apparently, the coefficients in Model 2 and Model 5 can be used as initial values for Algorithm 2 to deal with the UK flooding data to get the best adaptive Lasso non-Gaussian SIFE model for each considered spatial weight matrix (i.e., Model 3 and Model 6).

Table 4.9: Comparison among the best SIFE models with distance based weight matrix that established by using proposed statistical methods in Chapter 3 and Chapter 4.

Covariates	**Coefficients**	**Model 4**			**Model 5**			**Model 6**
		Est.	**S.E**	**P-val**	**Est.**	**S.E**	**P-val**	**Estimate**
	β_0	0.9233	0.4115	0.0250	0.9428	0.3042	0.0002	1.16442
ln(AREA)	β_1	0.9294	0.0245	0.0000	0.9807	0.0275	0.0000	0.95174
1000/SAAR	β_2	-1.3505	0.1687	0.0000	-1.4594	0.1274	0.0000	-1.49070
ln(FARL)	β_3	3.2368	0.2645	0.0000	3.3240	0.2635	0.0000	3.21381
(BFIHOST)^2	β_4	-3.3294	0.1186	0.0000	-3.3790	0.1132	0.0000	-3.37257
PROPWET	β_5	0.6024	0.3073	0.0500				0.00000
DPSBAR	β_6	-0.0010	0.0005	0.0414				0.00000
FPEXT	β_7	-2.1768	0.5396	0.0000	-1.9176	0.5044	0.0001	-1.54059
RMED-1D	β_8	0.0110	0.0044	0.0118				0.00000
SPRHOST	β_9							0.00000
LDP	β_{10}	-0.0028	0.0010	0.0056	-0.0037	0.0010	0.0001	-0.00312
ALTBAR	β_{11}							0.00000
ASPBAR	β_{12}							0.00000
ASPVAR	β_{13}				0.7844	0.2197	0.0002	0.48231
DPLBAR	β_{14}							0.00000
RMED-1H	β_{15}							0.00000
RMED-2D	β_{16}				0.0067	0.0028	0.0164	0.00584
$SAAR_{4170}$	β_{17}							0.00000
FPLOC	β_{18}							0.00000
FPDBAR	β_{19}							0.00000
	λ	0.4435	0.0697	0.0000	0.4336	0.0713	0.0000	0.45516
	σ	0.3604	0.0107	0.0000	0.3581	0.0107	0.0000	0.35924
	α	0.8974	0.2146	0.0000	0.8949	0.2183	0.0000	0.87653
AIC		**504.37**			**497.22**			**496.85**

Note: **Pink empty cells** represent the unused covariates while **blue empty cells** for non significant covariates in model fitting; **nG-SIFE** = non-Gaussian SIFE model; **ALnG-SIFE** = Adaptive Lasso non-Gaussian SIFE model; **DSWM** = Distance based Spatial weight matrix.
Model 4 = The best nG-SIFE model with DSWM developed by applying proposed statistical method in Chapter 3 to the UK flooding data of 10 covariates.
Model 5 = The best nG-SIFE model with DSWM developed by applying proposed statistical method in Chapter 3 to the UK flooding data of 19 covariates.
Model 6 = The best ALnG-SIFE model with DSWM developed by applying proposed statistical method in Chapter 4 to the UK flooding data of 19 covariates.

Table 4.8 and 4.9 give the summary of the comparison studies of aforementioned models for contiguity and distance spatial weight matrices respectively. Here, AIC is used as a model selection criteria. From Table 4.8 , Model 3 is outperform Model 1 and 2, while from Table 4.9 Model 6 outperform than Model 4 and 5 based on the lowest value of AIC. The obtained results demonstrate that, by using adaptive Lasso regularization method in panelized likelihood estimation procedure it can significantly improve non-Gaussian SIFE models. However, the performance in prediction also need to be investigated before this result can be concluded. Therefore, in this further investigation only models 2,3,5 and 6 will be taken into consideration. The prediction of those models will be performed by using the leave one out cross validation (LOOCV) method and the results are demonstrated in next Subsection.

4.6.3 Model evaluation

Table 4.10: Maximum likelihood vs panelized likelihood estimation (PLE) method in non-Gaussian SIFE model performance based on the value of mean square prediction error (MSPE)

Model	Method	Spatial weight	MSPE
Model 2	Maximum likelihood estimation	Contiguity	0.12893
Model 3	PLE with Adaptive Lasso		0.12596
Model 5	Maximum likelihood estimation	Distance	0.13235
Model 6	PLE with Adaptive Lasso		0.12982

In model estimation perspective based on AIC model selection criteria, the non-Gaussian SIFE models established by using panelized likelihood estimation method outperform the non-Gaussian SIFE models developed by using maximum likelihood estimation method for both considered spatial weight matrices. This is clearly shown in Tables 4.8 and 4.9 when we compare model 2 with model 3 and model 5 with model 6 respectively. The ability to accurately measure their prediction error is another perspective that has to evaluate in those established models. For this purpose, the prediction for Model 2 and Model 5 are performed by using the LOOCV method together with Algorithm 1 in Chapter 3 meanwhile, the prediction for Model 3 and Model 6 are performed by using the combination of the LOOCV method with Algorithm 2 in Chapter 4. There is no issue of time efficiency in computational task when we combine LOOCV method with Algorithm 1 but this issue exists if the combination is with Algorithm 2. Therefore, to make the computational task of LOOCV with algorithm 2 more efficient, we just consider tuning parameter η from 2.4 to 3.0 by 0.1 difference to solve optimization problem with contiguity spatial weight matrix and η from 3.6 to 4.1 by 0.1 difference to solve optimization problem with distance spatial weight matrix. Both considered scale tuning parameters used are based on the pattern of BIC values shown in Table 4.4 and 4.6. The results of the aforementioned computational tasks are summarized in Table 4.10. Notice that, we use the same name for each established models for continuity from Table 4.8 and 4.9. It is clearly shown from Table 4.10, panelized likelihood estimation with adaptive Lasso regularization techniques has an ability to accurately measure the prediction error which is better than maximum likelihood estimation method based on the smaller value of mean square prediction error. This allows us to suggest Model 3 and Model 6 as best adaptive Lasso non-Gaussian SIFE model in each considered spatial weight matrix for this chapter.

Table 4.11: The percentage improvement for three types of the SIFE model relative to the FEH-QMED model

Model	Significant covariates	Spatial weigt	MSPE	Improvement
FEH-QMED	x1,x2,x3,x4		0.15859	
G-SIFE	x1,x2,x3,x4	Contiguity	0.13671	13.8 percent
	x1,x2,x3,x4	Distance	0.13879	12.5 percent
nG-SIFE	x1,x2,x3,x4,x5,x6,x7,x8,x10	Contiguity	0.13155	17.1 percent
	x1,x2,x3,x4,x5,x6,x7,x8,x10	Distance	0.13393	15.6 percent
ALnG-SIFE	x1,x2,x3,x4,x7,x10,x13,x16	Contiguity	0.12596	20.6 percent
	x1,x2,x3,x4,x7,x10,x13,x16	Distance	0.12982	18.1 percent

To justify the efficiency of the statistical method proposed in this chapter, we are also comparing the prediction of the best SIFE models developed in Chapter 2, 3 and 4 (i.e., Gaussian, non-Gaussian, and adaptive Lasso non-Gaussian, respectively) with the updated FEH-QMED model by Kjeldsen and Jones (2010). The prediction for updated FEH-QMED model is also performed by using LOOCV method that has been applied into the given models in Kjeldsen and Jones (2010). The results of mean square prediction error together with its percentage of improvement relative to the updated FEH-QMED model of Kjeldsen and Jones (2010) are summaries in Table 4.11. All the best models show improvement relative to the benchmark model. Nevertheless, the best adaptive Lasso non-Gaussian SIFE model with contiguity spatial weight matrix from Chapter 4 outperforms the other models with 20.6 percent improvement over the updated FEH-QMED model in reducing the MSPE. From this finding, we can suggest that the panelized likelihood estimation method with adaptive Lasso under skewed spatial error model suggested in Chapter 4 is the most efficient statistical method compared to other methods developed by the chapters within this thesis in estimating the UK index flood with relevant predictive catchment characteristics.

4.7 Discussion and conclusion

4.7.1 Discussion

The main objective of this chapter is to discover relevant predictive catchment characteristics for the skewed spatial error model developed in Chapter 3. Panelized likelihood estimation method utilizing adaptive Lasso as a regularization tool is suggested for variable selection of 19 catchment characteristics in the UK flooding data for the developed models. Several findings can be observed through computational tasks of the suggested statistical procedure. They are:-

i. Panelized likelihood estimation with adaptive Lasso regularization procedure has an ability to correctly estimate the sparse solutions and can determine the unbiased and consistent estimators for skewed spatial error model that has been justified through simulation of the finite sample performance.

ii. Among the best SIFE models that developed under different statistical methods with both considered spatial weight matrices suggested in the chapters within this thesis, adaptive Lasso non-Gaussian SIFE model outperforms others with the lowest value of AIC and largest percentage of improvement in mean square prediction error, (MSPE) relative to the updated FEH-QMED model by Kjeldsen and Jones (2010) in the literature.

iii. The panelized likelihood estimation method utilizing adaptive Lasso as a regularization tool for skewed spatial error model suggested in this chapter is the most efficient statistical method compared to other statistical methods suggested in the previous two chapters which can improve the UK index flood estimation model with relevant predictive catchment characteristics.

All different statistical procedure suggested in this thesis seems to give reliable improvement comparing with the updated FEH-QMED model in Kjeldsen and Jones (2010). Among them, panelized likelihood estimation method using adaptive Lasso as a regularization tool for skewed spatial error model suggested in Chapter 4 is the most efficient method to establish the UK index flood estimation model. From this finding we can conclude that, simultaneous estimation and variable selection through this efficient statistical estimation procedure can result a parsimonious model which is simple with great explanatory predictive power. It can be justified through the reduction of the number of significant variables when we compare the best model of Chapter 4 with the best model in Chapter 3. However, this is not strictly a fair comparison because only ten covariates considered in fitting process by using maximum likelihood estimation method for skewed spatial error model suggested in Chapter 3. Therefore, we applied again the statistical method proposed in Chapter 3 into the UK flooding data of 19 catchment characteristics as covariates. Surprisingly, the developed models from these two different statistical methods are slightly similar in terms of the value of the coefficients and the significant covariates (see Table 4.8 and 4.9). This will raise the question on the compatibility of using panelized likelihood estimation instead of maximum likelihood estimation method.

This surprising result is literally expected to happen, since our limitation on the total number of available covariates is not too large which complies for the usage of penalized likelihood estimation method. Nevertheless the suggestion of using this method with adaptive Lasso as the regularization tool in this further investigation of our study attempts to response the instability in traditional variable selection method due to the correlated predictors where this problem has been stated in hydrological literature (see Section 4.1). This problem has been addressed through the findings obtained which allow us to conclude that an improve model for estimation of index flood in the UK is the adaptive Lasso non-Gaussian SIFE model based on AIC criteria which was shown with the largest percentage of improvement in mean square prediction error (MSPE) in comparison with the updated FEH-QMED model by Kjeldsen and Jones (2010).

In order to response on the aforementioned limitation, a new statistical model structure needs to be developed for future research to make it is more interesting to investigate. For example, in our statistical model structure proposed in Chapter 3, it just has one unknown parameter representing spatial dependency in model estimation which is λ, since the spatial relationship is treated as known one through considering spatial weight matrices. By changing the spatial relationship between the observed spatial data of n monitoring stations from being known to being unknown, it will increase the number of unknown parameters related to spatial dependency to be estimated from single value 1×1 to a vector $n \times 1$, where n represents the number of sample size. With these newly introduced unknown parameters coupled with the unknown parameters from regression coefficients and error terms $(\underset{\sim}{\beta}, \sigma, \alpha)$, make the number of unknown parameter p that need to be estimated in this new statistical model structure will be greater than sample size n. Apparently, solving the case of $p > n$ panelized likelihood estimation method is a most relevant alternative statistical method.

4.7.2 Conclusion

The use of multiple regression method is a common practice in operational hydrological analysis to establish a statistical model that can perform the underlying process between a certain hydrological parameter and a set of catchment characteristics. The FEH index flood estimation model (IH, 1999) known as the QMED model is well established one in the UK by using this approach. Practically, maximum likelihood estimation is used for model estimation with traditional selection procedure such as stepwise selection methods, Akaike's information criteria and Bayesian information criteria for model selection. These methods have many remarkble properties, but are often unsatisfactory in regression problems for two reasons: (i) large variability due to the number of predictor, p larger than the sample size n or high correlated in predictor variables, (ii) lack of interpretability if the number of predictor, p is large. Therefore, many flood risk study in the literature used just a selected catchment characteristics as the covariates (see Subsection (2.2.2) in the analysis of developing flood regression models. The traditional selection methods suffer from high variability due to the estimation and variable selection steps being resolved separately, which forces them to reduce the number of considered catchment characteristics especially the one that has been detected to be highly correlated (see Subsection 4.2). This is to ensure efficiently approximating the solution in those flood regression models with underlying truth, but it could bias in variable selection since not all available catchment characteristics are equally considered. Furthermore, the variable selection in those flood regression models was made under independent error structure, which could also bias the variable selection for the model.

Aforementioned problems have motivated us to suggest a penalized estimation procedure via utilizing adaptive lasso to estimate the important catchment characteristics predictors for the UK index flood estimation under spatial dependent error models with considering all available catchment characteristics in the dataset. Many penalized regression methods have been developed in linear regression setting to overcome the drawbacks in the traditional variable selection methods. The high prediction accuracy and computational efficiency of penalized regression methods have brought them increasing attention over the last decade. In addition, these methods can also produce more stable results for correlated data and data with the number of predictors being much larger than the sample size. Notice that, the number of available catchment characteristics is not too large, in which supposes to be greater than a sample size, however further investigation in this research attempt to response the instability in traditional variable selection due to the correlated predictors.

The findings listed in Subsection 4.7.1 allow us to conclude that by applying simultaneous estimation and variable selection through panelized likelihood estimation method with adaptive Lasso as a regularization tool for skewed spatial error model can result in a parsimonious model which is simple with great explanatory predictive power. This is based on AIC criteria which was shown with the largest percentage of improvement in mean square prediction error, (MSPE) n comparison with the updated FEH-QMED model by Kjeldsen and Jones (2010) in the literature. Furthermore, its simplicity can be justified through the reduction of the number of significant variables containing in the best model of this chapter when compared to the best model in Chapter 3. This justification hence can answer the main objective of this chapter, that is discovering relevant predictive catchment characteristics for the proposed skewed spatial error models in Chapter 3. The improved index flood estimation model for the UK flooding data established in this chapter named as adaptive Lasso non-Gaussian SIFE model can be expressed as

$$\begin{aligned} \ln(QMED_i) &= 0.9964 + 0.9678\ln(AREA_i) - 1.4734(1000SAAR_i^{-1}) + 3.0699 \\ &\quad \ln(FARL_i) - 3.3371(BFIHOST_i)^2 - 1.8316(FPEXT_i) - 0.0033 \\ &\quad (LDP_i) + 0.6693(ASPVAR_i) + 0.00679(RMED2D_i) + u_i, \qquad (4.14) \\ u &= 0.3940Wu + 0.3542e. \end{aligned}$$

where $u = (u_1, u_2, \cdots, u_n)'$ and W is a $n \times n$ standardized contiguity based spatial weight matrix with model error terms, $e = (e_1, e_2, \cdots, e_n)'$ where e_i beung i.i.d. of skewed normal distribution .

Model 4.14 has been developed by applying the suggested statistical method in this chapter into the UK flooding data of 19 catchment characteristics as the covariates. For the strictly fair comparison, this model need to be compared with the one that has been developed by using the statistical method suggested in Chapter 3 with the same catchment characteristics as the covariates in model fitting. In this fair comparison, we found that the significant covariates from these two different models are similar with just being slightly different in term of its coefficients. This result is probably due to the limitation regarding to the total number of available covariates being not too large. Nevertheless, the penalized likelihood estimation method suggested in this chapter attempts to respond to the problem of instability in the traditional variable selection method caused by the highly correlation in several catchment characteristics variables of the UK flooding data. This problem has been successful addressed when we found that adaptive Lasso non-Gaussian SIFE model outperforms other best models developed by using suggested statistical methods in previous two chapters. This finding justifies us to conclude that, the penalized likelihood estimation method that utilizes the adaptive lasso as regularization tool can improve the UK index flood estimation model with more relevant predictive catchment characteristics.

For future research, one could develop a more general statistical model structure with the ability to deal with much more unknown parameters, especially the case where the number of unknown parameters is greater than the sample size. This can be done for example by changing the known spatial relationship that represented by considered spatial weight matrices into unknown parameters that need to be estimated. The usage of panelized likelihood estimation method by utilizing adaptive lasso as a tool in variable selection hopefully will be more beneficial to this new statistical model structure. With that, more interesting study in skewed spatial error model can be investigated for the purpose of improving the index flood estimation model.

Chapter 5

Concluding Remarks

This chapter summaries and concludes the entire thesis entitled "Non Gaussian Spatial Modeling in Index Flood Estimation". Since this research project is interested in novel statistical methods for flood risk estimation, we need to fully understand the issues in flood risk study. From intensive reading and research, we found that the issue of how to estimate index flood at ungauged catchments is still an open, difficult problem among hydrologists though it has received great attention. The ungauged catchment area is referred to as the sites where there is no data for the purpose of flood estimation. Many parts of the world are ungauged or poorly gauged because monitoring stations are only located at specific, strategic and important locations (Mamun et al., 2011). Therefore, at many sites of interest for flood estimation we have no, or only have short-recorded, data. In response to this issue, statistical approach is particularly important for developing hydrological analysis with no observations of flood information at the ungauged area (Grimaldi et al., 2011). One of this kind of hydrological analysis is to estimate the design flood using available information from the nearby gauged sites when development projects are located at an ungauged catchment area. This approach is known as regionalization procedure.

Many types of regionalization procedures are available (Cunnane, 1988, 1989). A traditional and simple approach which has been used for a long time is index-flood procedures introduced by Dalrymple (1960). It consists of two main steps: 1) estimating the index flood, and 2) estimating the flood quantile. The literature contains numerous studies to improve the existing method in estimating the flood quantile compared to the index flood. Countless practical applications (see e.g., IH, 1999; Pilgrim and Institution of Engineers, 2001; Wilson et al., 2011) have highlighted the need and importance of having an accurate estimation of the index flood in regionalization procedure. The first step in the index flood procedure contributes a large amount of uncertainty to the regionalization procedure for flood estimation at ungauged catchment area (Kjeldsen and Jones, 2007). It is a common practice that multiple regression models are used to indirectly estimate the index flood at ungauged area by using catchments characteristics (c.f., Grimaldi et al., 2011; Castellarin et al., 2007; Brath et al., 2001). Many index flood estimation models were developed by using multiple regression models (NERC, 1975; Canuti and Moisello, 1982; Acreman, 1985; Mimikou and ordios, 1989; Garde and Kothyari, 1990; Reimers, 1990). Index flood estimation model from Flood Estimation Handbook (FEH) known as FEH-QMED model is one of them (IH, 1999), which was well-established in the UK. Therefore it has been chosen as the benchmark model for this research. The aforementioned issues have motivated us to further investigate and improve the existing statistical methods in the index flood estimation model development.

Three objectives have been developed based on the main purpose of this study to improve the existing index flood estimation model, with the FEH-QMED model taken as the benchmark. These objectives are (i) to explore the potential to possibly improve the model for estimation of the index flood in the UK, (ii) to develop a more effective statistical model in estimating the index flood that can better fit the UK flooding data, and (iii) to discover the more relevant predictive catchment characteristics variables for the developed model. We have developed the methods of this research based on the aforementioned objectives, and with that, our research can be divided into three stages which are model exploration, model development and model selection that are represented by the chapters within this thesis, Chapter 2, Chapter 3 and Chapter 4, respectively.

In operational hydrological analysis, using multiple regression models has been a common practice to establish functional relationship between the flow statistics and the associated physical catchment properties. The benchmark UK index flood estimation model known as FEH-QMED model is a well established model by using this approach. It is interesting to note that the FEH-QMED model is of a linear multiple regression model structure with nonlinear effects of explanatory variables identified. By applying a spatial additive regression analysis on the flooding data sets (see Chapter 2), we have found that the nonlinear impact of the catchment characteristics identified by the FEH-QMED model appear to be reliable. However, spatial Moran's I and Lagrange Multiplier tests both show that there is a strong spatial autocorrelation in the regression residuals of the FEH-QMED model. Note that hydrologists used not just the FEH-QMED model alone in estimating index flood, but used it together with data transferring procedure to take account of the spatial autocorrelation in the index flood regression residual separately. We therefore proposed incorporating the potential spatial neighboring impacts (i.e., spatial autocorrelation, see chapter 2) into the FEH-QMED model by using spatial error model analysis, which simplifies and improves the method of using the FEH-QMED model. Results have shown the Gaussian SIFE model that has been established by using spatial error model with contiguity based weight matrix can better fit the UK flooding data, which has been illustrated to outperform the latest revised FEH-QMED model in Kjeldsen and Jones (2010) from the prediction perceptive. This concludes that, by incorporating spatial neighboring impact directly into the FEH-QMED model, we can significantly improve the estimation of the UK index flood and simplify the previous methodology.

It has been stated in the literature, the celebrated Gaussian distribution is no longer suitable for hydrological data since the observed spatial variables in this field is known to have skewed distribution (Zhang and El-Shaarawi, 2010). Notice that, the established spatial index flood estimation (SIFE) model in Chapter 2 is based on a Gaussian assumption of the residuals. A further investigation has therefore been made to determine the distribution of the regression residuals and the UK flooding data by using Kernel density estimation plot. The results showed that a kind of skewed distribution need to be taken into account in both random variables. The fact that asymmetry in data makes the Gaussian assumption not valid, has called for a spatial error model for skewed UK flooding data. Thus, we proposed a maximum likelihood with computational procedure detailed in Algorithm 1 for spatial error model with skewed normal distribution of residuals (see Chapter 3). The simulation of the performance for the proposed estimation method has been investigated before it could be applied into the UK flooding data to establish a new index flood estimation model. This new model called non-Gaussian SIFE model has been illustrated to outperform the Gaussian SIFE model in Chapter 2 and the updated FEH-QMED model in Kjeldsen and Jones (2010) from the prediction perspective. This concludes that, by changing the distribution of the residuals in the spatial error model from normal to a skewed normal distribution can improve the estimation of the UK index flood model.

Flooding is a complex process that is caused not only by hydrological factors, but also by a lot of other physical factors. The objective in the final stage of this research is concerned about variable selection in index flood estimation model because there are some more catchment characteristics variables available in the UK flooding dataset that have not been considered in model development. In Chapters 2 and 3 within this thesis, the traditional selection criteria with AIC score have been used in model selection. This traditional method has a number of drawbacks including high variability due to the estimation and variable selection have been executed in separate steps (Breiman, 1996). Furthermore, the significant variables of the FEH-QMED model in Kjeldsen et al. (2008) have been selected from ten considered catchment characteristics under independent error structure which could also bias the variable selection for the model. These two issues motivated us to propose a penalized likelihood estimation method by utilizing adaptive Lasso as a tool in variable selection for the skewed spatial error model that has been developed in Chapter 3. A new computational procedure detailed in Algorithm 2 is suggested to solve the optimization problem in the proposed panelized likelihood estimation method with adaptive Lasso as a regularization tool (see Chapter 4). The simulations of the performance for the proposed estimation procedure in terms of the ability in correctly identifying the sparse solutions for the regression coefficients and determining unbiased and consistent non-zero estimators have been investigated before it could be applied into the UK flooding data to establish a new index flood estimation model. The established spatial index flood estimation models from this new statistical method have been illustrated to outperform the best Gaussian SIFE model from Chapter 2, the best non-Gaussian SIFE model from Chapter 3 and the updated FEH-QMED model in Kjeldsen and Jones (2010) from the prediction perspective. This concludes that, by applying simultaneous estimation and variable selection in the spatial error model through regularization procedure of adaptive Lasso can improve the estimation of the UK index flood.

All different statistical methods suggested in this research seem to give reliable improvements compared with the benchmark model in the literature,that is the updated FEH-QMED model of Kjeldsen and Jones (2010). Chapter 2 suggested the statistical method that took into account the spatial dependency in the residuals of the flood regression model by fitting spatial autoregressive models into the UK flooding data. The conclusion through the findings from Chapter 2 leads us to develop the new statistical model structure in Chapter 3 for the spatial error model with the skewed normal distribution of the error terms. This is to take into account the non-Gausianity that presented in the UK flooding data and the residuals of spatial autoregressive models. A Maximum likelihood estimation computational procedure has been suggested in Algorithm 1 to solve the proposed model. Finally, the panelized likelihood estimation with adaptive Lasso as a regularization tool of variable selection has been suggested in Chapter 4 to overcome the limitations in maximum likelihood estimation method declared in the conclusion of Chapter 3. A suggested computational procedure to solve this new statistical method can be referred to in Algorithm 2. It has been shown that, the improvement made in statistical methods for each chapter enhanced the performance of the developed models based on the AIC score and percentage of improvement in the MSPE relative to that of the benchmark model. Accordingly, we can conclude that all the proposed statistical methods have their own ability in enhancing the performance to determine the reliable estimation models that can produce a good predictive outcome for the index flood estimation. With this ability, each chapter within this thesis can be developed independently and evolve for more interesting future research that could be explored.

It is my hope and intention that the work within this thesis has an impact in the application research of the real problems in the environmental, hydrological and ecological studies where the observed spatial variables collected through modern data acquisition techniques in those fields are well-known to have skewed distribution. There are an endless number of possible empirical applications of the methods proposed within this thesis. But the most important, this research can be a base for more interesting research that could be explored to improve the spatial autoregressive models through new statistical model structure that can be developed. Some examples will be briefly explained in Section 5.2 that gives the recommendation for the future works. Before that, we would like to emphasize the contributions of this study that have been made to the existing literature in the following Section 5.1.

5.1 Contribution of this research

It is my intention to do the research which can be beneficial in practical usage to the response in the open issues of the real spatial problems whether it is in the environmental, hydrological or ecological area of studies. With modern data acquisition techniques the observed spatial variables in those aforementioned fields have been and continue to be collected. These sources of data are important to make it possible for us to do an application study. For this purpose we have decided to develop novel statistical methods that can be applied into the flood risk modeling. Therefore, a lot of intensive reading and research have been done to make sure we have sufficient recent knowledge about the theories and issues in the flood risk study. With that effort, we have finally found the gaps from statistical perspective in flood risk modeling. The gaps are clearer when we do the empirical work by using the available flooding data in exploring the UK index flood estimation model. The findings of the empirical work have led us to the further research to improve existing methodology in the index flood regression model from statistical point of view that may not be noticeable to hydrologists. The empirical works have also made us fully understand about the data used in this study and motivated us to investigate further on it and contribute new knowledge in the related field. The contributions from this research study that have been made to meet the drawbacks and challenges in the existing literature are:-

i. We have found that the Gaussian spatial index flood estimation (SIFE) model with contiguity based weight matrix, by incorporating spatial dependence directly into the FEH-QMED model, can better fit the UK flooding data, which has been illustrated to outperform the latest revised FEH-QMED model in Kjeldsen and Jones (2010) from the prediction perceptive.

   The FEH index flood model known as the QMED model is a well established index flood regression model. As a national model it is reliable to the UK flood data overall, but spatial autocorrelation in the model residuals is still existent and need to be further investigated. In practice, hydrologists did not use the FEH-QMED model alone to estimate the index flood but adjusted it by using data transferring procedure which takes into account geographical proximity and hydrological similarity, where the spatial autocorrelation in regression residual could be encountered (Kjeldsen and Jones, 2010). It could be great if the spatial dependency would be incorporated directly into the FEH-QMED model. Therefore, this research responsed to this problem by incorporating spatial dependence directly into the FEH-QMED model through spatial autoregressive models. Interestingly, we found that the Gaussian SIFE model which has been developed by using spatial error model with contiguity based weight matrix can better fit the UK flooding data which has been illustrated outperform the latest revised FEH-QMED model in Kjeldsen and Jones (2010) from the prediction perceptive.

ii. This research has proposed and developed a new statistical model structure for spatial error model with the skewed normal distribution in responding to the issue of non-Gaussinity in the real flooding data, which improves the index flood estimation model.

   Exploratory analyses into the UK flooding data and regression residuals of the spatial error model have shown that a kind of skewed distribution need to be considered in the model development. Furthermore, it is well-known that most real data, especially the environmental, hydrological and ecological data, have skewed distribution (Zhang and El-Shaarawi, 2010). This makes a Gaussian assumption is too ideal to be used in developing a model and could provide a bias estimate. Therefore, this research has developed a new statistical model structure for spatial error model by changing the normal distribution of the error terms into the skewed normal distribution in responding to the issue of non-Gaussinity in the real flooding data and hence improves the index flood estimation model.

iii. This study has suggested a new computational procedure detailed in Algorithm 1 to solve the proposed maximum likelihood estimation method for the skewed spatial error model.

Skewed normal distribution used to represent the non-Gausianity detected in the UK flooding data has desirable properties, with an asymmetry parameter representing the skewness of the real data situations. However, this leads to complexities in drawing inference on the parameters of the proposed models. This issue of complexity has been discussed by Sartori (2006), Dalla Valle (2004), Pewsey (2000), Monti (2003) among others. To overcome the complexities problem in its optimization, this study suggested a new computational procedure detailed in Algorithm 1 to solve the proposed maximum likelihood estimation method for the skewed spatial error model. The simulations on the performance for finite samples have been done to ensure the suggested estimators are compliant with the two basic properties of an estimator, that is unbiasedness and consistency, before it can be applied into the UK flooding data for index flood estimation. By applying this statistical method into the UK flooding data we found that, the non-Gaussian SIFE model which has been established by using skewed spatial error model with contiguity based weight matrix can better fit the UK flooding data. This best non-Gaussian SIFE model has been illustrated to outperform the best Gaussian SIFE model in Chapter 2 and the latest revised FEH-QMED model in Kjeldsen and Jones (2010) from the prediction perceptive.

iv. We have proposed a penalize maximum likelihood estimation for variable selection with our established spatial error model with skewed normal residuals by utilizing adaptive lasso as a tool, and a new computational procedure detailed in Algorithm 2 is also suggested to solve the proposed panelized likelihood estimation method.

In addition to ensuring high predictive accuracy, discovering relevant predictive variables is another fundamental goal in developing statistical model. A lot of catchment characteristics availables in the FEH CD-ROM that can be considered as the predictors in developing flood index regression model. With many potential candidate predictors, some of them have been detected to be highly correlated in Wagener et al. (2004). This could give a regression problem that requires an efficient statistical model selection to find an optimal model which is simple but still provides good predictive performance. This regression problems can be even more problematic due to the instability of stepwise selection method as the estimation and variable selection are executed separately (Breiman, 1996). This challenge has been countered through the proposed penalize maximum likelihood estimation for our established spatial error model with skewed normal residuals by utilizing adaptive lasso as a tool in variable selection. A new computational procedure detailed in Algorithm 2 is also suggested to solve the proposed panelized likelihood estimation method. The simulations on the performance for finite samples have been done to ensure the ability in correctly identifying the sparse solution for the regression coefficient and determining unbiased and consistent estimators before it can be applied into the UK flooding data for index flood estimation. The established models from this proposed method are also found to outperform the best Gaussian SIFE model in Chapter 2, the best non-Gaussian SIFE model in Chapter 3 and the latest revised FEH-QMED model in Kjeldsen and Jones (2010) from the prediction perceptive.

5.2 Recommendations for further work

As mentioned earlier in this chapter, the proposed statistical methods in this thesis have their own ability in enhancing the performance to determine the reliable estimation models that can produce a good predictive outcome for the index flood estimation. This has been demonstrated through the findings of the study, where all the models that have been developed through the proposed statistical methods outperform the benchmark model of this study set as the latest revised FEH-QMED model in Kjeldsen and Jones (2010). Therefore, it is believed that each chapter within this thesis can be developed independently and evolve for more interesting future research that could be explored.

Chapter 2 in this thesis has been presented by emphasizing more on the empirical work purposes, which is to explore the potential to possibly improve a model for estimation of the index flood in the UK. A suggested spatial autoregressive model has been incorporated into a linear model due to our benchmark model being a linear multiple regression model with non linear effect of the explanatory variables that have been identified reliable by using spatial additive analysis. This empirical work can be represented by introducing a new statistical model structure that combines additive and spatial autoregressive models. With this statistical model structure, we can potentially design an iterative solution for the model by using a backfitting computational procedure for the additive structure together with the spatial autoregressive model as we did in Chapter 2. This established model can also be further developed based on a nonlinear model structure for future research and by introducing more covariates in its application.

Through Chapter 3, one of the further work that could be considered is to extend the developed statistical methods by using other candidate distribution for the error terms. Besides using skewed normal distribution we can have a try to take into account other distribution that probably could represent well the non Gausianity in the UK flooding data for example a generalized error distribution.

This study also can be continued with future work that has been suggested in Chapter 4. The suggested future work through Chapter 4 is to focus on estimating the spatial neighboring impact as in our current research where we treat this spatial relationship as a known parameter in the spatial weight matrix that was represented by spatial contiguity and distance weight matrices. It will also be interesting if we can combine these spatial weights and propose a new structure of the non-Gaussian spatial error model and use a penalized likelihood estimation method that utilizes adaptive Lasso as a tool in variable selection in establishing a new UK index flood estimation model.

Besides that, derivation of the Fisher-information matrix for theoretical contribution in proving the asymptotic normality of our proposed method can also be considered for the future work. We have derived the second partial derivatives of the log-likelihood function 3.17 where the details are shown in Appendix E,F,G and H, pages 117-143. The next step for future work is to use them to derive the elements in the fisher information matrix by determining the expectation of the second partial derivatives and construct the Fisher Information matrix. With the Fisher information matrix, the asymptotic normality of the proposed statistical methods in this thesis and the statistical methods that will be proposed in the future work can be well investigated.

It is my personal hope and intention that the works within this thesis have an impact in the application research of the real problems generally in the environmental, hydrological and ecological studies, where the observed spatial variables collected through modern data acquisition techniques are known to have non-Gaussian distribution. There are an endless number of possible empirical applications of the statistical methods proposed within this thesis. The works that have been done in this thesis are a good start for my future research path that is really close to the one of the niche research area of Faculty of Earth Sciences at Univesiti Malaysia Kelantan where I am servicing. In this faculty, I belong to the research cluster group of "Geo-Computing, Spatial Modeling and Simulation" which really need my contribution since they have rich spatial data that can be explored in statistical point of view and I can contribute a new knowledge in the interdisciplinary field.

Appendix A

First order partial derivative of the log-likelihood with respect to β

There are two expressions in log-likelihood function give a nonzero solution when differentiating it with respect to β;

First expression,

$$
\begin{aligned}
&\frac{\partial}{\partial\beta}\Bigg(\sum_{i=1}^{n}\log\phi\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\Bigg)\\
&=\frac{\partial}{\partial\beta}\Bigg(\sum_{i=1}^{n}\log\Bigg(\frac{1}{\sqrt{2\pi}}\exp\Big\{\frac{-\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)^2}{2}\Big\}\Bigg)\Bigg)\\
&=\frac{\partial}{\partial\beta}\Bigg(-\frac{n}{2}\log(2\pi)-\frac{1}{2}\sum_{i=1}^{n}\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)^2\Bigg)\\
&=-\frac{1}{2}\sum_{i=1}^{n}\Bigg(2\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\cdot\frac{\partial}{\partial\beta}\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\Bigg)\\
&=-\frac{1}{2}\sum_{i=1}^{n}\Bigg(2\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\cdot\frac{\sigma_z}{\sigma}\frac{\partial}{\partial\beta}\Big(-(I-\lambda W)_iX\beta\Big)\Bigg)\\
&=-\frac{1}{2}\sum_{i=1}^{n}\Bigg(2\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\cdot\frac{\sigma_z}{\sigma}\Big(-X^T(I-\lambda W)_i^T\Big)\Bigg)\\
&=\frac{\sigma_z}{\sigma}\sum_{i=1}^{n}X^T(I-\lambda W)_i^T\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\\
&=\frac{\sigma_z}{\sigma}\sum_{i=1}^{n}\Big((I-\lambda W)_iX\Big)^T\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\\
&=\frac{\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)
\end{aligned}
$$

This solution is stated in Chapter 4 page 73.

Second expression,

$$
\begin{aligned}
&\frac{\partial}{\partial\beta}\left(\sum_{i=1}^{n}\log\Phi\Big(\alpha\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\Big)\right)\\
&=\sum_{i=1}^{n}\left(\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\cdot\frac{\partial}{\partial\beta}\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)\right)\\
&=\sum_{i=1}^{n}\left(\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\cdot\frac{\alpha\sigma_z}{\sigma}\left(\frac{\partial}{\partial\beta}\Big((I-\lambda W)_i(y-X\beta)\Big)\right)\right)\\
&=\sum_{i=1}^{n}\left(\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\cdot\frac{\alpha\sigma_z}{\sigma}\left(\frac{\partial}{\partial\beta}\Big(-(I-\lambda W)_iX\beta\Big)\right)\right)\\
&=\sum_{i=1}^{n}\left(\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\cdot\frac{\alpha\sigma_z}{\sigma}\Big(-X^T(I-\lambda W)_i^T\Big)\right)\\
&=-\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}\left(X^T(I-\lambda W)_i^T\left\{\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}\right)
\end{aligned}
$$

This solution is stated in Chapter 4 page 73.

Hence,

$$
\begin{aligned}
\frac{\partial\ell}{\partial\beta} \;=\;& \frac{\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)-\\
&\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}\left(X^T(I-\lambda W)_i^T\left\{\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}\right)
\end{aligned}
$$

Appendix B

First order partial derivative of the log-likelihood with respect to λ

First expression,

$$
\begin{aligned}
&\frac{\partial}{\partial\lambda}\left(\sum_{i=1}^{n}\log\phi\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\right)\\
&=\frac{\partial}{\partial\lambda}\left(\sum_{i=1}^{n}\log\left(\frac{1}{\sqrt{2\pi}}\exp\left\{\frac{-\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)^2}{2}\right\}\right)\right)\\
&=\frac{\partial}{\partial\lambda}\left(-\frac{n}{2}\log(2\pi)-\frac{1}{2}\sum_{i=1}^{n}\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)^2\right)\\
&=-\frac{1}{2}\sum_{i=1}^{n}\left(2\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\cdot\frac{\partial}{\partial\lambda}\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\right)\\
&=-\frac{1}{2}\sum_{i=1}^{n}\left(2\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\cdot\frac{\sigma_z}{\sigma}\Big\{\frac{\partial}{\partial\lambda}((I-\lambda W)_i y)-\frac{\partial}{\partial\lambda}((I-\lambda W)_i X\beta)\Big\}\right)\\
&=-\frac{1}{2}\sum_{i=1}^{n}\left(2\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\cdot\frac{\sigma_z}{\sigma}\Big(-W_i^T y+W_i^T X\beta\Big)\right)\\
&=\frac{\sigma_z}{\sigma}\sum_{i=1}^{n}\left(\Big(W_i^T\big(y-XB\big)\Big)\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\right)\\
&=\frac{\sigma_z}{\sigma}\Big(W\big(y-XB\big)\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)
\end{aligned}
$$

This solution is stated in Chapter 4 page 73.

Second expression,

$$\begin{aligned}
&\frac{\partial}{\partial\lambda}\Bigg(\sum_{i=1}^{n}\log\Phi\Big(\alpha\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\Big)\Bigg)\\
&=\sum_{i=1}^{n}\left(\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\cdot\frac{\partial}{\partial\lambda}\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)\right)\\
&=\sum_{i=1}^{n}\left(\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\cdot\frac{\alpha\sigma_z}{\sigma}\left(\frac{\partial}{\partial\lambda}\Big((I-\lambda W)_i(y-X\beta)\Big)\right)\right)\\
&=\sum_{i=1}^{n}\left(\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\cdot\frac{\alpha\sigma_z}{\sigma}\left(\frac{\partial}{\partial\lambda}\Big((I-\lambda W)_iy-(I-\lambda W)_iX\beta\Big)\right)\right)\\
&=\sum_{i=1}^{n}\left(\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\cdot\frac{\alpha\sigma_z}{\sigma}\left(\frac{\partial}{\partial\lambda}\big(-\lambda W_iy+\lambda W_iX\beta\big)\right)\right)\\
&=\sum_{i=1}^{n}\left(\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\cdot\left(-\frac{\alpha\sigma_z}{\sigma}\Big(W_i^Ty-W_i^TX\beta\Big)\right)\right)\\
&=\sum_{i=1}^{n}\left(\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\cdot\left(-\frac{\alpha\sigma_z}{\sigma}\Big(W_i^T(y-X\beta)\Big)\right)\right)\\
&=-\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}\left(W_i^T\big(y-X\beta\big)\left\{\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}\right)
\end{aligned}$$

This solution is stated in Chapter 4 page 73.

Third expression,

$$\begin{aligned}
\frac{\partial}{\partial\lambda}\Big(\log|I-\lambda W|\Big) &= \frac{1}{|I-\lambda W|}\cdot\frac{\partial}{\partial\lambda}|I-\lambda W|\\
&= \frac{1}{|I-\lambda W|}\left(|I-\lambda W|tr\Big((I-\lambda W)^{-1}\frac{\partial}{\partial\lambda}(I-\lambda W)\Big)\right)\\
&= \frac{1}{|I-\lambda W|}|I-\lambda W|tr\Big((I-\lambda W)^{-1}(-W)\Big)\\
&= -tr\Big((I-\lambda W)^{-1}W\Big)
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{\partial\ell}{\partial\lambda} \;=\; & \frac{\sigma_z}{\sigma}\Big(W\big(y-XB\big)\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)-tr\Big((I-\lambda W)^{-1}W\Big)-\\
& \frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}\left(W_i^T\big(y-X\beta\big)\left\{\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}\right)
\end{aligned}$$

Appendix C

First order partial derivative of the log-likelihood with respect to σ

There are three expressions in log-likelihood function give a nonzero solution when differentiating it with respect to σ;

First expression,

$$\frac{\partial}{\partial\sigma}\Big(-n\log\sigma\Big) = -n\Big(\frac{1}{\sigma}\Big) = -\frac{n}{\sigma}$$

Second expression,

$$\frac{\partial}{\partial\sigma}\left(\sum_{i=1}^{n}\log\phi\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\right)$$
$$= \frac{\partial}{\partial\sigma}\left(\sum_{i=1}^{n}\log\left(\frac{1}{\sqrt{2\pi}}\exp\left\{\frac{-\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)^2}{2}\right\}\right)\right)$$
$$= \frac{\partial}{\partial\sigma}\left(-\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)^2\right)$$
$$= -\frac{1}{2}\sum_{i=1}^{n}\left(2\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\cdot\frac{\partial}{\partial\sigma}\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\right)$$
$$= -\frac{1}{2}\sum_{i=1}^{n}\left(2\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\Big(-\frac{\sigma_z}{\sigma^2}(I-\lambda W)_i(y-X\beta)\Big)\right)$$
$$= \frac{\sigma_z}{\sigma^2}\Big((\mu_z\cdot\mathbb{1})^T(I-\lambda W)(y-X\beta)\Big) + \frac{\sigma_z^2}{\sigma^3}\Big(\big((I-\lambda W)(y-X\beta)\big)^T\big((I-\lambda W)(y-X\beta)\big)\Big)$$

This solution is stated in Chapter 4 page 76.

Third expression,

$$\frac{\partial}{\partial\sigma}\left(\sum_{i=1}^{n}\log\Phi\Big(\alpha\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\Big)\right)$$

$$=\sum_{i=1}^{n}\left(\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\cdot\frac{\partial}{\partial\sigma}\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)\right)$$

$$=-\frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}\left(\Big((I-\lambda W)_i(y-X\beta)\Big)\left\{\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}\right)$$

This solution is stated in Chapter 4 page 76.

Hence,

$$\begin{aligned}\frac{\partial\ell}{\partial\sigma} &= \frac{\sigma_z}{\sigma^2}\Big((\mu_z\cdot\mathbb{1})^T(I-\lambda W)(y-X\beta)\Big)+\frac{\sigma_z^2}{\sigma^3}\Big(\big((I-\lambda W)(y-X\beta)\big)^T\big((I-\lambda W)(y-X\beta)\big)\Big)\\ &\quad -\frac{n}{\sigma}-\frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}\left(\Big((I-\lambda W)_i(y-X\beta)\Big)\left\{\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}\right)\end{aligned}$$

Appendix D

First order partial derivative of the log-likelihood with respect to α

μ_z and σ_z are known as a function of α from Equation (3.3) and (3.4), where $\dfrac{\partial \mu_z}{\partial \alpha} = \dfrac{\mu_z}{\alpha(1+\alpha^2)}$

First expression,

$$
\begin{aligned}
&\frac{\partial}{\partial \alpha}\left(\sum_{i=1}^{n} \log \phi\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\right)\\
&= \frac{\partial}{\partial \alpha}\left(\sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi}}\exp\left\{\frac{-\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)^2}{2}\right\}\right)\right)\\
&= \frac{\partial}{\partial \alpha}\left(-\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)^2\right)\\
&= -\frac{1}{2}\sum_{i=1}^{n}\left(2\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\cdot\frac{\partial}{\partial \alpha}\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\right)\\
&= -\frac{1}{2}\sum_{i=1}^{n}\left(2\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\cdot\frac{\partial}{\partial \alpha}\left\{\frac{\sqrt{2}\alpha}{\sqrt{\pi(1+\alpha^2)}} + \right.\right.\\
&\qquad \left.\left.\frac{1}{\sigma}\sqrt{1-\frac{2\alpha^2}{\pi(1+\alpha^2)}}(I-\lambda W)_i(y-X\beta)\right\}^*\right)\\
&= -\frac{1}{2}\sum_{i=1}^{n}\left(2\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\left(\frac{\sqrt{2}}{1+\alpha^2} - \frac{\mu_z^2}{\alpha(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\right)\right)
\end{aligned}
$$

$$= \frac{\mu_z^2}{\alpha(1+\alpha^2)\sigma\sigma_z}\sum_{i=1}^{n}\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\Big((I-\lambda W)_i(y-X\beta)\Big)-$$
$$\frac{\sqrt{2}}{1+\alpha^2}\sum_{i=1}^{n}\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)$$
$$= \frac{\mu_z^2}{\alpha(1+\alpha^2)\sigma\sigma_z}\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\Big((I-\lambda W)(y-X\beta)\Big)-$$
$$\frac{\sqrt{2}}{1+\alpha^2}\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\cdot\mathbb{1}$$

This solution is stated in Chapter 4 page 74.

Second expression,

$$\frac{\partial}{\partial\alpha}\left(\sum_{i=1}^{n}\log\Phi\Big(\alpha\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)\Big)\right)$$
$$= \sum_{i=1}^{n}\left(\frac{\phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\cdot\frac{\partial}{\partial\alpha}\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)\right)$$
$$= \sum_{i=1}^{n}\left(\frac{\phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\cdot\Bigg(\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)+\right.$$
$$\left.\alpha\Big(\frac{\partial}{\partial\alpha}\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)\Bigg)\right)$$
$$= \sum_{i=1}^{n}\left(\frac{\phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\cdot\Bigg(\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)+\right.$$
$$\left.\alpha\Big(\frac{\partial}{\partial\alpha}\Big\{\frac{\sqrt{2}\alpha}{\sqrt{\pi(1+\alpha^2)}} + \frac{1}{\sigma}\sqrt{1-\frac{2\alpha^2}{\pi(1+\alpha^2)}}(I-\lambda W)_i(y-X\beta)\Big\}^*\Big)\Bigg)\right)$$
$$= \sum_{i=1}^{n}\left(\frac{\phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\Bigg(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)+\right.$$
$$\left.\alpha\Big(\frac{\sqrt{2}}{1+\alpha^2} - \frac{\mu_z^2}{\alpha(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)\Bigg)\right)$$
$$= \sum_{i=1}^{n}\left(\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta) + \frac{\sqrt{2}\alpha}{1+\alpha^2} - \frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)\right.$$
$$\left.\left\{\frac{\phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}\right)$$

This solution is stated in Chapter 4 page 74.

Where, the solution for * is shown below,

$$\begin{aligned}\frac{\partial}{\partial\alpha}\left(\frac{\sqrt{2}\alpha}{\sqrt{\pi(1+\alpha^2)}}\right) &= \frac{\sqrt{2}\big(\pi(1+\alpha^2)\big)^{\frac{1}{2}} - \frac{1}{2}\big(\pi(1+\alpha^2)\big)^{-\frac{1}{2}}2\pi\alpha\sqrt{2}\alpha}{\pi(1+\alpha^2)} \\ &= \frac{\sqrt{2}\pi(1+\alpha^2) - \sqrt{2}\pi\alpha^2}{\pi(1+\alpha^2)} \\ &= \frac{\sqrt{2}\pi + \sqrt{2}\pi\alpha^2 - \sqrt{2}\pi\alpha^2}{\pi(1+\alpha^2)} \\ &= \frac{\sqrt{2}}{1+\alpha^2}\end{aligned}$$

$$\begin{aligned}&\frac{\partial}{\partial\alpha}\left(\frac{1}{\sigma}\sqrt{1-\frac{2\alpha^2}{\pi(1+\alpha^2)}}(I-\lambda W)_i(y-X\beta)\right) \\ &= \frac{(I-\lambda W)_i(y-X\beta)}{\sigma}\left(\frac{\partial}{\partial\alpha}\sqrt{1-\frac{2\alpha^2}{\pi(1+\alpha^2)}}\right) \\ &= \frac{(I-\lambda W)_i(y-X\beta)}{\sigma}\left(\frac{1}{2}\Big(1-\frac{2\alpha^2}{\pi(1+\alpha^2)}\Big)^{-\frac{1}{2}}\Big(\frac{-\big(4\alpha\pi(1+\alpha^2)\big) - 2\pi\alpha 2\alpha^2}{\big(\pi(1+\alpha^2\big)^2}\Big)\right) \\ &= \frac{1}{2\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\left(\frac{4\pi\alpha^3 - 4\pi\alpha(1+\alpha^2)}{\big(\pi(1+\alpha^2)\big)^2}\right) \\ &= \frac{1}{2\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\left(\frac{4\pi\alpha^4 - 4\pi\alpha^2(1+\alpha^2)}{\alpha\big(\pi(1+\alpha^2)\big)^2}\right) \\ &= \frac{1}{2\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big(\frac{\sqrt{2}\alpha}{\sqrt{\pi(1+\alpha^2)}}\Big)^2\Big(\frac{2\pi\alpha^2 - 2\pi(1+\alpha^2)}{\alpha\pi(1+\alpha^2)}\Big) \\ &= \frac{\mu_z^2}{2\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big(\frac{2\pi\alpha^2 - 2\pi - 2\pi\alpha^2}{\alpha\pi(1+\alpha^2)}\Big) \\ &= \frac{-\mu_z^2}{\alpha(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\end{aligned}$$

Third expression,

$$\begin{aligned}\frac{\partial}{\partial\alpha}\Big(n\log\sigma_z\Big) &= \frac{\partial}{\partial\alpha}\left(n\log\sqrt{1-\frac{2\alpha^2}{\pi(1+\alpha^2}}\right) \\ &= n\left(\frac{\frac{1}{2}\Big(1-\frac{2\alpha^2}{\pi(1+\alpha^2}\Big)^{-\frac{1}{2}}}{\Big(1-\frac{2\alpha^2}{\pi(1+\alpha^2}\Big)^{\frac{1}{2}}}\Big(\frac{-\Big(4\alpha\big(\pi(1+\alpha^2)\big)\Big) - 2\pi\alpha 2\alpha^2}{\big(\pi(1+\alpha^2)\big)}\Big)\right) \\ &= \frac{n}{2\sigma_z^2}\left(\frac{\sqrt{2}\alpha}{\sqrt{\pi(1+\alpha^2)}}\right)^2\left(\frac{2\pi\alpha^2 - 2\pi - 2\pi\alpha^2}{\alpha\pi(1+\alpha^2)}\right) \\ &= -\frac{n\mu_z^2}{\alpha(1+\alpha^2)\sigma_z^2}\end{aligned}$$

Hence,

$$\begin{aligned}\frac{\partial \ell}{\partial \alpha} &= \frac{\mu_z^2}{\alpha(1+\alpha^2)\sigma\sigma_z}\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\Big((I-\lambda W)(y-X\beta)\Big)-\\ &\quad \frac{\sqrt{2}}{1+\alpha^2}\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\cdot\mathbb{1}-\frac{n\mu_z^2}{\alpha(1+\alpha^2)\sigma_z^2}+\\ &\quad \sum_{i=1}^{n}\Bigg(\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)+\frac{\sqrt{2}\alpha}{1+\alpha^2}-\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)\\ &\quad \left\{\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}\Bigg)\end{aligned}$$

To ensure the computational task is defined well, we substitute $\frac{\mu_z^2}{\alpha(1+\alpha^2)}$ with $\frac{2\alpha}{\pi(1+\alpha^2)^2}$ in $\frac{\delta\ell}{\delta\alpha}$, that derived from Equation (3.3) as follows;

$$\mu_z = \frac{\sqrt{2}\alpha}{\sqrt{\pi(1+\alpha^2)}}$$

$$\mu_z^2 = \frac{2\alpha^2}{\pi(1+\alpha^2)}$$

$$\frac{\mu_z^2}{\alpha(1+\alpha^2)} = \frac{\frac{2\alpha^2}{\pi(1+\alpha^2)}}{\alpha(1+\alpha^2)} = \frac{2\alpha}{\pi(1+\alpha^2)^2}$$

Then $\frac{\delta\ell}{\delta\alpha}$ can be rewritten as

$$\begin{aligned}\frac{\partial \ell}{\partial \alpha} &= \frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\Big((I-\lambda W)(y-X\beta)\Big)-\\ &\quad \frac{\sqrt{2}}{1+\alpha^2}\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\cdot\mathbb{1}-\frac{n2\alpha}{\pi(1+\alpha^2)^2\sigma_z^2}+\\ &\quad \sum_{i=1}^{n}\Bigg(\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)+\frac{\sqrt{2}\alpha}{1+\alpha^2}-\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)\\ &\quad \left\{\frac{\phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}\Bigg).\end{aligned}$$

Appendix E

Second order partial derivative of the log-likelihood with respect to β

Lets derive the second order partial derivatives of the log-likelihood with respect to β. First we have,

$$\frac{\partial^2 \ell}{\partial \beta^2} = \frac{\partial}{\partial \beta}\Big(\frac{\partial \ell}{\partial \beta}\Big)$$
$$= \frac{\partial}{\partial \beta}\Bigg(\frac{\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T\Big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big) - \frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}(I-\lambda W)_i X A_i\Bigg)$$

where, $A_i = \left\{\dfrac{\phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}$. Let $u = \mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)$, such that $\frac{\partial u}{\partial \beta} = -\frac{\sigma_z}{\sigma}(I-\lambda W)X$, and $\frac{\partial(\alpha u)}{\partial \beta} = -\frac{\alpha\sigma_z}{\sigma}(I-\lambda W)X$, then,

$$\frac{\partial(A_i)}{\partial \beta} = \frac{\partial}{\partial \beta}\left(\frac{\frac{1}{\sqrt{2\pi}}\exp\{\frac{-(\alpha u)^2}{2}\}}{\Phi(\alpha u)}\right)$$
$$= \frac{\phi(\alpha u)\Big(\frac{-2(\alpha u)}{2}\Big)\Big(\frac{\partial(\alpha u)}{\partial \beta}\Big)\Phi(\alpha u) - \phi(\alpha u)\Big(\frac{\partial(\alpha u)}{\partial \beta}\Big)\phi(\alpha u)}{\Big(\Phi(\alpha u)\Big)^2}$$
$$= \frac{\phi(\alpha u)\Big(\frac{-2(\alpha u)}{2}\Big)\Big(-\frac{\alpha\sigma_z}{\sigma}(I-\lambda W)_i X\Big)\Phi(\alpha u) - \phi(\alpha u)\Big(-\frac{\alpha\sigma_z}{\sigma}(I-\lambda W)_i X\Big)\phi(\alpha u)}{\Big(\Phi(\alpha u)\Big)^2}$$

$$
\begin{aligned}
&= \frac{\frac{\alpha\sigma_z}{\sigma}(I-\lambda W)_i X\phi(\alpha u)\big((\alpha u)\Phi(\alpha u)+\phi(\alpha u)\big)}{\Phi(\alpha u)\cdot\Phi(\alpha u)} \\
&= \frac{\alpha\sigma_z}{\sigma}(I-\lambda W)_i X A_i\big(\alpha u + A_i\big) \\
&= \frac{\alpha\sigma_z}{\sigma}(I-\lambda W)_i X A_i\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big) + A_i\Big)
\end{aligned}
$$

Now, we start differentiating first expression in $\frac{\partial \ell}{\partial \beta}$ with respect to β.

$$
\begin{aligned}
&\frac{\partial}{\partial\beta}\bigg(\frac{\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)\bigg) \\
&= \frac{\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T\Big(-\frac{\sigma_z}{\sigma}\Big((I-\lambda W)X\Big) \\
&= -\frac{\sigma_z^2}{\sigma^2}\Big((I-\lambda W)X\Big)^T\Big((I-\lambda W)X\Big).
\end{aligned}
$$

Then, we differentiate second expression in $\frac{\partial \ell}{\partial \beta}$ with respect to β.

$$
\begin{aligned}
&\frac{\partial}{\partial\beta}\Big(-\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}(I-\lambda W)_i X A_i\Big) \\
&= -\frac{\alpha\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T\sum_{i=1}^{n}\frac{\partial(A_i)}{\partial\beta} \\
&= -\frac{\alpha\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T\sum_{i=1}^{n}\frac{\alpha\sigma_z}{\sigma}(I-\lambda W)_i X A_i\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)+A_i\Big) \\
&= -\frac{\alpha^2\sigma_z^2}{\sigma^2}\Big((I-\lambda W)X\Big)^T\Big(\alpha\big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big)+A_i\Big)A^T(I-\lambda W)X
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\frac{\partial^2\ell}{\partial\beta^2} = &-\frac{\sigma_z^2}{\sigma^2}\Big((I-\lambda W)X\Big)^T\Big((I-\lambda W)X\Big) - \frac{\alpha^2\sigma_z^2}{\sigma^2}\Big((I-\lambda W)X\Big)^T \\
&\Big(\alpha\big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big)+A_i\Big)A^T(I-\lambda W)X
\end{aligned}
$$

Second we have,

$$
\begin{aligned}
\frac{\partial^2\ell}{\partial\lambda\partial\beta} &= \frac{\partial}{\partial\beta}\Big(\frac{\partial\ell}{\partial\lambda}\Big) \\
&= \frac{\partial}{\partial\beta}\bigg(\frac{\sigma_z}{\sigma}\Big(W\big(y-X\beta\big)\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big) - \\
&\quad \frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}W_i\big(y-X\beta\big)A_i - tr\Big((I-\lambda W)^{-1}W\Big)\bigg)
\end{aligned}
$$

Let start differentiating the first expression in $\frac{\partial \ell}{\partial \lambda}$ with respect to β,

$$\frac{\partial}{\partial \beta}\left(\frac{\sigma_z}{\sigma}\Big(W\big(y-X\beta\big)\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)\right)$$
$$=-\frac{\sigma_z}{\sigma}(WX)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)-\frac{\sigma_z^2}{\sigma^2}\Big((I-\lambda W)X\Big)^T\Big(W\big(y-X\beta\big)\Big)$$

Then, we differentiate second expression in $\frac{\partial \ell}{\partial \lambda}$ with respect to β.

$$\frac{\partial}{\partial \beta}\left(-\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}W_i(y-X\beta)A_i\right)$$
$$=\frac{\alpha\sigma_z}{\sigma}(WX)^TA-\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}W_i(y-X\beta)\left(\frac{\partial(A_i)}{\partial\beta}\right)$$
$$=\frac{\alpha\sigma_z}{\sigma}(WX)^TA-\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}W_i(y-X\beta)\bigg(\frac{\alpha\sigma_z}{\sigma}(I-\lambda W)_iXA_i\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i$$
$$(y-X\beta)\big)+A_i\Big)\bigg)$$
$$=\frac{\alpha\sigma_z}{\sigma}(WX)^TA-\frac{\alpha^2\sigma_z^2}{\sigma^2}\Big((I-\lambda W)X\Big)^TA\Big(W(y-X\beta)\Big)^T\Big(\alpha\big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)$$
$$(y-X\beta)\big)+A\Big)$$

The third expression in $\frac{\partial \ell}{\partial \lambda}$ give a zero solution of partial derivative with respect to β. Hence,

$$\frac{\partial^2\ell}{\partial\lambda\partial\beta}=\frac{\alpha\sigma_z}{\sigma}(WX)^TA-\frac{\alpha^2\sigma_z^2}{\sigma^2}\Big((I-\lambda W)X\Big)^TA\Big(W(y-X\beta)\Big)^T\Big(\alpha\big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}$$
$$(I-\lambda W)(y-X\beta)\big)+A\Big)-\frac{\sigma_z}{\sigma}(WX)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)-$$
$$\frac{\sigma_z^2}{\sigma^2}\Big((I-\lambda W)X\Big)^T\Big(W\big(y-X\beta\big)\Big)$$

Third we have,

$$\frac{\partial^2\ell}{\partial\sigma\partial\beta}=\frac{\partial}{\partial\beta}\left(\frac{\partial\ell}{\partial\sigma}\right)$$
$$=\frac{\partial}{\partial\beta}\bigg(\frac{\sigma_z}{\sigma^2}(\mu_z\cdot\mathbb{1})^T(I-\lambda W)(y-X\beta)+\frac{\sigma_z^2}{\sigma^3}\big((I-\lambda W)(y-X\beta)\big)^T\big((I-\lambda W)$$
$$(y-X\beta)\big)-\frac{n}{\sigma}-\frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}(I-\lambda W)_i(y-X\beta)A_i\bigg)$$

Let start differentiating the first expression in $\frac{\partial \ell}{\partial \sigma}$ with respect to β.

$$\frac{\partial}{\partial\beta}\left(\frac{\sigma_z}{\sigma^2}(\mu_z\cdot\mathbb{1})^T(I-\lambda W)(y-X\beta)\right)=-\frac{\sigma_z}{\sigma^2}\Big((I-\lambda W)X\Big)^T(\mu_z\cdot\mathbb{1})$$

Then, we differentiate the second expression in $\dfrac{\partial \ell}{\partial \sigma}$ with respect to β.

$$\begin{aligned}
&\frac{\partial}{\partial \beta}\left(\frac{\sigma_z^2}{\sigma^3}\big((I-\lambda W)(y-X\beta)\big)^T\big((I-\lambda W)(y-X\beta)\big)\right)\\
&=-\frac{\sigma_z^2}{\sigma^3}\big((I-\lambda W)X\big)^T\big((I-\lambda W)(y-X\beta)\big)-\frac{\sigma_z^2}{\sigma^3}\Big(\big((I-\lambda W)(y-X\beta)\big)^T(I-\lambda W)X\Big)^T\\
&=-\frac{2\sigma_z^2}{\sigma^3}\big((I-\lambda W)X\big)^T\big((I-\lambda W)(y-X\beta)\big)
\end{aligned}$$

Lastly, we differentiate fourth expression in $\dfrac{\partial \ell}{\partial \sigma}$ with respect to β.

$$\begin{aligned}
&\frac{\partial}{\partial \beta}\left(-\frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}(I-\lambda W)_i(y-X\beta)A_i\right)\\
&=-\frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}-(I-\lambda W)_iXA_i-\frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}(I-\lambda W)_i(y-X\beta)\left(\frac{\partial(A_i)}{\partial \beta}\right)\\
&=\frac{\alpha\sigma_z}{\sigma^2}\big((I-\lambda W)X\big)^TA-\frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}(I-\lambda W)_i(y-X\beta)\Bigg(\frac{\alpha\sigma_z}{\sigma}(I-\lambda W)_iXA_i\\
&\quad\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)+A_i\Big)\Bigg)\\
&=\frac{\alpha\sigma_z}{\sigma^2}\big((I-\lambda W)X\big)^TA-\frac{\alpha^2\sigma_z^2}{\sigma^3}\big((I-\lambda W)X\big)^T(I-\lambda W)(y-X\beta)A^T\\
&\quad\Big(\alpha\big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big)+A\Big)
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \sigma \partial \beta}=&-\frac{\sigma_z}{\sigma^2}\Big((I-\lambda W)X\Big)^T(\mu_z\cdot\mathbb{1})-\frac{2\sigma_z^2}{\sigma^3}\big((I-\lambda W)X\big)^T\big((I-\lambda W)(y-X\beta)\big)+\\
&\frac{\alpha\sigma_z}{\sigma^2}\big((I-\lambda W)X\big)^TA-\frac{\alpha^2\sigma_z^2}{\sigma^3}\big((I-\lambda W)X\big)^T(I-\lambda W)(y-X\beta)A^T\\
&\Big(\alpha\big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big)+A\Big)
\end{aligned}$$

Fourth we have,

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \alpha \partial \beta}&=\frac{\partial}{\partial \beta}\left(\frac{\partial \ell}{\partial \alpha}\right)\\
&=\frac{\partial}{\partial \beta}\Bigg(\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\Big((I-\lambda W)(y-X\beta)\Big)-\\
&\quad\frac{\sqrt{2}}{1+\alpha^2}\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\cdot\mathbb{1}-\frac{n2\alpha}{\pi(1+\alpha^2)^2\sigma_z^2}+\\
&\quad\sum_{i=1}^{n}\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)+\frac{\sqrt{2}\alpha}{1+\alpha^2}-\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)A_i\Bigg)
\end{aligned}$$

We start differentiating the first expression in $\dfrac{\partial \ell}{\partial \alpha}$ with respect to β.

$$\frac{\partial}{\partial \beta}\left(\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\left(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\right)^T\left((I-\lambda W)(y-X\beta)\right)\right)$$
$$=-\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma^2}\big((I-\lambda W)X\big)^T\big((I-\lambda W)(y-X\beta)\big)-$$
$$\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\left(\big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big)^T(I-\lambda W)X\right)^T$$

Then, we differentiate second expression in $\dfrac{\partial \ell}{\partial \alpha}$ with respect to β.

$$\frac{\partial}{\partial \beta}\left(-\frac{\sqrt{2}}{1+\alpha^2}\left(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\right)^T\cdot\mathbb{1}\right)=\frac{\sqrt{2}\sigma_z}{(1+\alpha^2)\sigma}\big((I-\lambda W)X\big)^T\cdot\mathbb{1}$$

Lastly, we differentiate the fourth expression in $\dfrac{\partial \ell}{\partial \alpha}$ with respect to β.

$$\frac{\partial}{\partial \beta}\left(\sum_{i=1}^{n}\left(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)+\frac{\sqrt{2}\alpha}{1+\alpha^2}-\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\right)A_i\right)$$
$$=\left(\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)X-\frac{\sigma_z}{\sigma}(I-\lambda W)X\right)^TA+\sum_{i=1}^{n}\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)+$$
$$\frac{\sqrt{2}\alpha}{1+\alpha^2}-\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)\Big(\frac{\alpha\sigma_z}{\sigma}(I-\lambda W)_iXA_i\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i$$
$$(y-X\beta)\big)+A_i\Big)\Big)$$
$$=\left(\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)X-\frac{\sigma_z}{\sigma}(I-\lambda W)X\right)^TA+\frac{\alpha\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}$$
$$(I-\lambda W)(y-X\beta)+\frac{\sqrt{2}\alpha}{1+\alpha^2}\cdot\mathbb{1}-\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)(y-X\beta)\Big)A^T$$
$$\Big(\alpha\big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big)+A\Big)$$

Hence,

$$\frac{\partial^2\ell}{\partial\alpha\partial\beta}=\frac{\sqrt{2}\sigma_z}{\alpha(1+\alpha^2)\sigma}\big((I-\lambda W)X\big)^T\cdot\mathbb{1}-\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma^2}\big((I-\lambda W)X\big)^T\big((I-\lambda W)$$
$$(y-X\beta)\big)-\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\Big((I-\lambda W)X\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)+$$
$$\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}\Big((I-\lambda W)X\Big)^TA-\frac{\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^TA+\frac{\alpha\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T$$
$$\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)+\frac{\sqrt{2}\alpha}{1+\alpha^2}\cdot\mathbb{1}-\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)(y-X\beta)\Big)$$
$$A^T\Big(\alpha\big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big)+A\Big)$$

Appendix F

Second order partial derivative of the log-likelihood with respect to λ

Lets derive the second order partial derivatives of the log-likelihood with respect to λ. First we have,

$$\frac{\partial^2 \ell}{\partial\beta\partial\lambda} = \frac{\partial}{\partial\lambda}\Big(\frac{\partial\ell}{\partial\beta}\Big)$$
$$= \frac{\partial}{\partial\lambda}\left(\frac{\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big) - \frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}(I-\lambda W)_i X A_i\right)$$

where, $A_i = \left\{\dfrac{\phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}$. Let $u = \mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)$, such that $\frac{\partial u}{\partial\lambda} = -\frac{\sigma_z}{\sigma}W(y-X\beta)$, and $\frac{\partial(\alpha u)}{\partial\lambda} = -\frac{\alpha\sigma_z}{\sigma}W(y-X\beta)$, then,

$$\frac{\partial(A_i)}{\partial\lambda} = \frac{\partial}{\partial\lambda}\left(\frac{\frac{1}{\sqrt{2\pi}}\exp\{\frac{-(\alpha u)^2}{2}\}}{\Phi(\alpha u)}\right)$$
$$= \frac{\phi(\alpha u)\Big(\frac{-2(\alpha u)}{2}\Big)\Big(\frac{\partial(\alpha u)}{\partial\lambda}\Big)\Phi(\alpha u) - \phi(\alpha u)\Big(\frac{\partial(\alpha u)}{\partial\lambda}\Big)\phi(\alpha u)}{\Big(\Phi(\alpha u)\Big)^2}$$
$$= \frac{\phi(\alpha u)\Big(\frac{-2(\alpha u)}{2}\Big)\Big(-\frac{\alpha\sigma_z}{\sigma}W_i(y-X\beta)\Big)\Phi(\alpha u) - \phi(\alpha u)\Big(-\frac{\alpha\sigma_z}{\sigma}W_i(y-X\beta)\Big)\phi(\alpha u)}{\Big(\Phi(\alpha u)\Big)^2}$$

$$
\begin{aligned}
&= \frac{\frac{\alpha\sigma_z}{\sigma} W_i(y - X\beta)\phi(\alpha u)\big((\alpha u)\Phi(\alpha u) + \phi(\alpha u)\big)}{\Phi(\alpha u)\cdot\Phi(\alpha u)} \\
&= \frac{\alpha\sigma_z}{\sigma} W_i(y - X\beta)A_i\big(\alpha u + A_i\big) \\
&= \frac{\alpha\sigma_z}{\sigma} W_i(y - X\beta)A_i\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I - \lambda W)_i(y - X\beta)\big) + A_i\Big)
\end{aligned}
$$

Now, we start differentiating first expression in $\dfrac{\partial \ell}{\partial \beta}$ with respect to λ.

$$
\begin{aligned}
&\frac{\partial}{\partial\lambda}\Bigg(\frac{\sigma_z}{\sigma}\Big((I - \lambda W)X\Big)^T\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I - \lambda W)(y - X\beta)\Big)\Bigg) \\
&= -\frac{\sigma_z}{\sigma}\Big(WX\Big)^T\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I - \lambda W)(y - X\beta)\Big) + \frac{\sigma_z}{\sigma}\Big((I - \lambda W)X\Big)^T \\
&\quad \Big(-\frac{\sigma_z}{\sigma}W(y - X\beta)\Big) \\
&= -\frac{\sigma_z}{\sigma}\Big(WX\Big)^T\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I - \lambda W)(y - X\beta)\Big) - \frac{\sigma_z^2}{\sigma^2}\Big((I - \lambda W)X\Big)^T W(y - X\beta)
\end{aligned}
$$

Then, we differentiate second expression in $\dfrac{\partial \ell}{\partial \beta}$ with respect to λ.

$$
\begin{aligned}
&\frac{\partial}{\partial\lambda}\Big(-\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}(I - \lambda W)_i X A_i\Big) \\
&= -\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n} -\Big(W_i X\Big)A_i - \frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}(I - \lambda W)_i X\frac{\partial(A_i)}{\partial\lambda} \\
&= \frac{\alpha\sigma_z}{\sigma}\Big(WX\Big)^T A - \frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}(I - \lambda W)_i X\Big(\frac{\alpha\sigma_z}{\sigma}W_i(y - X\beta)A_i\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I - \lambda W)_i \\
&\quad (y - X\beta)\big) + A_i\Big) \\
&= \frac{\alpha\sigma_z}{\sigma}\Big(WX\Big)^T A - \frac{\alpha^2\sigma_z^2}{\sigma^2}\Big((I - \lambda W)X\Big)^T W(y - X\beta)A^T\Big(\alpha\big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I - \lambda W) \\
&\quad (y - X\beta)\big) + A\Big)
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial\beta\partial\lambda} = {}& \frac{\alpha\sigma_z}{\sigma}\Big(WX\Big)^T A - \frac{\sigma_z}{\sigma}\Big(WX\Big)^T\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I - \lambda W)(y - X\beta)\Big) - \\
& \frac{\sigma_z^2}{\sigma^2}\Big((I - \lambda W)X\Big)^T W(y - X\beta) - \frac{\alpha^2\sigma_z^2}{\sigma^2}\Big((I - \lambda W)X\Big)^T W(y - X\beta)A^T \\
& \Big(\alpha\big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I - \lambda W)(y - X\beta)\big) + A\Big)
\end{aligned}
$$

Second we have,

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \lambda^2} &= \frac{\partial}{\partial \lambda}\left(\frac{\partial \ell}{\partial \lambda}\right)\\ &= \frac{\partial}{\partial \lambda}\left(\frac{\sigma_z}{\sigma}\Big(W\big(y - X\beta\big)\Big)^T\Big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y - X\beta)\Big) - \right.\\ &\quad \left.\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n} W_i\big(y - X\beta\big)A_i - tr\Big((I-\lambda W)^{-1}W\Big)\right)\end{aligned}$$

Let start differentiating the first expression in $\frac{\partial \ell}{\partial \lambda}$ with respect to λ,

$$\begin{aligned}&\frac{\partial}{\partial \lambda}\left(\frac{\sigma_z}{\sigma}\Big(W\big(y - X\beta\big)\Big)^T\Big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y - X\beta)\Big)\right)\\ &= -\frac{\sigma_z^2}{\sigma^2}\Big(W\big(y - X\beta\big)\Big)^T\Big(W\big(y - X\beta\big)\Big)\end{aligned}$$

Then, we differentiate second expression in $\frac{\partial \ell}{\partial \lambda}$ with respect to λ.

$$\begin{aligned}&\frac{\partial}{\partial \lambda}\left(-\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n} W_i(y - X\beta)A_i\right)\\ &= -\frac{\alpha\sigma_z}{\sigma}\Big(W(y - X\beta)\Big)^T\left(\frac{\partial(A_i)}{\partial \lambda}\right)\\ &= -\frac{\alpha\sigma_z}{\sigma}\Big(W(y - X\beta)\Big)^T\left(\sum_{i=1}^{n}\frac{\alpha\sigma_z}{\sigma}W_i(y - X\beta)A_i\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y - X\beta)\big) + A_i\Big)\right)\\ &= -\frac{\alpha^2\sigma_z^2}{\sigma^2}\Big(W(y - X\beta)\Big)^T W(y - X\beta)A^T\Big(\alpha\big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y - X\beta)\big) + A\Big)\end{aligned}$$

Lastly we differentiate the third expression in $\frac{\partial \ell}{\partial \lambda}$ with respect to λ.

$$\frac{\partial}{\partial \lambda}\Big(-tr\Big((I-\lambda W)^{-1}W\Big)\Big) = -tr\Big(W(I-\lambda W)^{-2}W\Big)$$

Hence,

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \lambda^2} &= -\frac{\sigma_z^2}{\sigma^2}\Big(W\big(y - X\beta\big)\Big)^T\Big(W\big(y - X\beta\big)\Big) - tr\Big(W(I-\lambda W)^{-2}W\Big) - \\ &\quad \frac{\alpha^2\sigma_z^2}{\sigma^2}\Big(W(y - X\beta)\Big)^T W(y - X\beta)A^T\Big(\alpha\big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y - X\beta)\big) + A\Big)\end{aligned}$$

Third we have,

$$\frac{\partial^2 \ell}{\partial\sigma\partial\lambda} = \frac{\partial}{\partial\lambda}\left(\frac{\partial\ell}{\partial\sigma}\right)$$
$$= \frac{\partial}{\partial\lambda}\Bigg(\frac{\sigma_z}{\sigma^2}(\mu_z \cdot \mathbb{1})^T(I-\lambda W)(y-X\beta) + \frac{\sigma_z^2}{\sigma^3}\big((I-\lambda W)(y-X\beta)\big)^T\big((I-\lambda W)$$
$$(y-X\beta)\big) - \frac{n}{\sigma} - \frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}(I-\lambda W)_i(y-X\beta)A_i\Bigg)$$

Let start differentiating the first expression in $\frac{\partial\ell}{\partial\sigma}$ with respect to λ.

$$\frac{\partial}{\partial\lambda}\left(\frac{\sigma_z}{\sigma^2}(\mu_z \cdot \mathbb{1})^T(I-\lambda W)(y-X\beta)\right) = -\frac{\sigma_z}{\sigma^2}(\mu_z \cdot \mathbb{1})^T\Big(W(y-X\beta\Big)$$

Then, we differentiate the second expression in $\frac{\partial\ell}{\partial\sigma}$ with respect to λ.

$$\frac{\partial}{\partial\lambda}\left(\frac{\sigma_z^2}{\sigma^3}\big((I-\lambda W)(y-X\beta)\big)^T\big((I-\lambda W)(y-X\beta)\big)\right)$$
$$= -\frac{\sigma_z^2}{\sigma^3}\big(W(y-X\beta)\big)^T\big((I-\lambda W)(y-X\beta)\big) - \frac{\sigma_z^2}{\sigma^3}\Big(\big((I-\lambda W)(y-X\beta)\big)^T W(y-X\beta)\Big)$$

Lastly, we differentiate fourth expression in $\frac{\partial\ell}{\partial\sigma}$ with respect to λ.

$$\frac{\partial}{\partial\lambda}\left(-\frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}(I-\lambda W)_i(y-X\beta)A_i\right)$$
$$= -\frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n} -W_i(y-X\beta)A_i - \frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}(I-\lambda W)_i(y-X\beta)\left(\frac{\partial(A_i)}{\partial\lambda}\right)$$
$$= \frac{\alpha\sigma_z}{\sigma^2}\big(W(y-X\beta)\big)^T A - \frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}(I-\lambda W)_i(y-X\beta)\Bigg(\frac{\alpha\sigma_z}{\sigma}W_i(y-X\beta)A_i$$
$$\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big) + A_i\Big)\Bigg)$$
$$= \frac{\alpha\sigma_z}{\sigma^2}\big(W(y-X\beta)\big)^T A - \frac{\alpha^2\sigma_z^2}{\sigma^3}\big(W(y-X\beta\big)^T(I-\lambda W)(y-X\beta)A^T$$
$$\Big(\alpha\big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big) + A\Big)$$

Hence,

$$\frac{\partial^2\ell}{\partial\sigma\partial\lambda} = \frac{\alpha\sigma_z}{\sigma^2}\big(W(y-X\beta)\big)^T A - \frac{\alpha^2\sigma_z^2}{\sigma^3}\big(W(y-X\beta\big)^T(I-\lambda W)(y-X\beta)A^T$$
$$\Big(\alpha\big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big) + A\Big) - \frac{\sigma_z}{\sigma^2}(\mu_z \cdot \mathbb{1})^T\Big(W(y-X\beta\Big) -$$
$$\frac{2\sigma_z^2}{\sigma^3}\big(W(y-X\beta)\big)^T\big((I-\lambda W)(y-X\beta)\big)$$

Fourth we have,

$$\frac{\partial^2 \ell}{\partial\alpha\partial\lambda} = \frac{\partial}{\partial\lambda}\Big(\frac{\partial\ell}{\partial\alpha}\Big)$$
$$= \frac{\partial}{\partial\lambda}\Bigg(\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\Big((I-\lambda W)(y-X\beta)\Big)-$$
$$\frac{\sqrt{2}}{1+\alpha^2}\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\cdot\mathbb{1}-\frac{n2\alpha}{\pi(1+\alpha^2)^2\sigma_z^2}+$$
$$\sum_{i=1}^{n}\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)+\frac{\sqrt{2}\alpha}{1+\alpha^2}-\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)A_i\Bigg)$$

We start differentiating the first expression in $\frac{\partial\ell}{\partial\alpha}$ with respect to λ.

$$\frac{\partial}{\partial\lambda}\Bigg(\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\Big((I-\lambda W)(y-X\beta)\Big)\Bigg)$$
$$= -\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma^2}\big(W(y-X\beta)^T\big((I-\lambda W)(y-X\beta)\big)-$$
$$\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big)^T W(y-X\beta)$$

Then, we differentiate second expression in $\frac{\partial\ell}{\partial\alpha}$ with respect to λ.

$$\frac{\partial}{\partial\lambda}\Bigg(-\frac{\sqrt{2}}{1+\alpha^2}\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\cdot\mathbb{1}\Bigg) = \frac{\sqrt{2}\sigma_z}{(1+\alpha^2)\sigma}\big(W(y-X\beta\big)^T\cdot\mathbb{1}$$

Lastly, we differentiate the fourth expression in $\frac{\partial\ell}{\partial\alpha}$ with respect to λ.

$$\frac{\partial}{\partial\lambda}\Bigg(\sum_{i=1}^{n}\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)+\frac{\sqrt{2}\alpha}{1+\alpha^2}-\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)A_i\Bigg)$$
$$= \Big(\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}W(y-X\beta)-\frac{\sigma_z}{\sigma}W(y-X\beta\Big)^T A+\sum_{i=1}^{n}\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)+$$
$$\frac{\sqrt{2}\alpha}{1+\alpha^2}-\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)\Big(\frac{\alpha\sigma_z}{\sigma}W_i(y-X\beta)A_i\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i$$
$$(y-X\beta)\big)+A_i\Big)\Big)$$
$$= \Big(\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}W(y-X\beta)-\frac{\sigma_z}{\sigma}W(y-X\beta\Big)^T A+\frac{\alpha\sigma_z}{\sigma}\Big(W(y-X\beta\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}$$
$$(I-\lambda W)(y-X\beta)+\frac{\sqrt{2}\alpha}{1+\alpha^2}\cdot\mathbb{1}-\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)(y-X\beta)\Big)A^T$$
$$\Big(\alpha\big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big)+A\Big)$$

Hence,

$$\frac{\partial^2 \ell}{\partial\alpha\partial\lambda} = \frac{\sqrt{2}\sigma_z}{(1+\alpha^2)\sigma}\big(W(y-X\beta)\big)^T \cdot \mathbb{1} - \frac{2\alpha}{\pi(1+\alpha^2)^2\sigma^2}\big(W(y-X\beta)\big)^T\big((I-\lambda W)(y-X\beta)\big)$$
$$- \frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\Big(W(y-X\beta)\Big)^T\Big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big) +$$
$$\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}\Big(W(y-X\beta)\Big)^T A - \frac{\sigma_z}{\sigma}\Big(W(y-X\beta\Big)^T A + \frac{\alpha\sigma_z}{\sigma}\Big(W(y-X\beta\Big)^T$$
$$\Big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta) + \frac{\sqrt{2}\alpha}{1+\alpha^2}\cdot \mathbb{1} - \frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)(y-X\beta)\Big)$$
$$A^T\Big(\alpha\big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big) + A\Big)$$

Appendix G

Second order partial derivative of the log-likelihood with respect to σ

Lets derive the second order partial derivatives of the log-likelihood with respect to σ. First we have,

$$\frac{\partial^2 \ell}{\partial\beta\partial\sigma} = \frac{\partial}{\partial\sigma}\left(\frac{\partial \ell}{\partial \beta}\right)$$
$$= \frac{\partial}{\partial\sigma}\left(\frac{\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T\Big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big) - \frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}(I-\lambda W)_i X A_i\right)$$

where, $A_i = \left\{\dfrac{\phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}$. Let $u = \mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)$, such that $\frac{\partial u}{\partial \sigma} = -\frac{\sigma_z}{\sigma^2}(I-\lambda W)(y-X\beta)$, and $\frac{\partial(\alpha u)}{\partial\sigma} = -\frac{\alpha\sigma_z}{\sigma^2}(I-\lambda W)(y-X\beta)$, then,

$$\frac{\partial(A_i)}{\partial\sigma} = \frac{\partial}{\partial\sigma}\left(\frac{\frac{1}{\sqrt{2\pi}}\exp\{\frac{-(\alpha u)^2}{2}\}}{\Phi(\alpha u)}\right)$$
$$= \frac{\phi(\alpha u)\Big(\frac{-2(\alpha u)}{2}\Big)\Big(\frac{\partial(\alpha u)}{\partial\sigma}\Big)\Phi(\alpha u) - \phi(\alpha u)\Big(\frac{\partial(\alpha u)}{\partial\sigma}\Big)\phi(\alpha u)}{\Big(\Phi(\alpha u)\Big)^2}$$
$$= \frac{\frac{\alpha\sigma_z}{\sigma^2}(I-\lambda W)_i(y-X\beta)\phi(\alpha u)\big((\alpha u)\Phi(\alpha u) + \phi(\alpha u)\big)}{\Phi(\alpha u)\cdot\Phi(\alpha u)}$$

$$= \frac{\alpha\sigma_z}{\sigma^2}(I-\lambda W)_i(y-X\beta)A_i\big(\alpha u + A_i\big)$$
$$= \frac{\alpha\sigma_z}{\sigma^2}(I-\lambda W)_i(y-X\beta)A_i\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big) + A_i\Big)$$

Now, we start differentiating first expression in $\dfrac{\partial \ell}{\partial \beta}$ with respect to σ.

$$
\begin{aligned}
&\frac{\partial}{\partial \sigma}\left(\frac{\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)\right)\\
&=-\frac{\sigma_z}{\sigma^2}\Big((I-\lambda W)X\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)+\frac{\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T\\
&\quad\Big(-\frac{\sigma_z}{\sigma^2}(I-\lambda W)(y-X\beta)\Big)\\
&=-\frac{\sigma_z}{\sigma^2}\Big((I-\lambda W)X\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)-\frac{\sigma_z^2}{\sigma^3}\Big((I-\lambda W)X\Big)^T\\
&\quad(I-\lambda W)(y-X\beta)\\
&=-\frac{\sigma_z}{\sigma^2}\Big((I-\lambda W)X\Big)^T\mu_z\cdot\mathbb{1}-\frac{2\sigma_z^2}{\sigma^3}\Big((I-\lambda W)X\Big)^T(I-\lambda W)(y-X\beta)
\end{aligned}
$$

Then, we differentiate second expression in $\dfrac{\partial \ell}{\partial \beta}$ with respect to σ.

$$
\begin{aligned}
&\frac{\partial}{\partial \sigma}\Big(-\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}(I-\lambda W)_i XA_i\Big)\\
&=\frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}(I-\lambda W)_i XA_i-\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}(I-\lambda W)_i X\Big(\frac{\partial(A_i)}{\partial\sigma}\Big)\\
&=\frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}(I-\lambda W)_i XA_i-\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}(I-\lambda W)_i X\Bigg(\frac{\alpha\sigma_z}{\sigma^2}(I-\lambda W)_i(y-X\beta)A_i\\
&\quad\Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)+A_i\Big)\Bigg)\\
&=\frac{\alpha\sigma_z}{\sigma^2}(\Big(I-\lambda W)X\Big)^T A-\frac{\alpha^2\sigma_z^2}{\sigma^3}\Big((I-\lambda W)X\Big)^T(I-\lambda W)(y-X\beta)A^T\\
&\quad\Big(\alpha\big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big)+A\Big)\Big)
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\frac{\partial^2\ell}{\partial\beta\partial\sigma}=&-\frac{\sigma_z}{\sigma^2}\Big((I-\lambda W)X\Big)^T\mu_z\cdot\mathbb{1}-\frac{2\sigma_z^2}{\sigma^3}\Big((I-\lambda W)X\Big)^T(I-\lambda W)(y-X\beta)+\\
&\frac{\alpha\sigma_z}{\sigma^2}(\Big(I-\lambda W)X\Big)^T A-\frac{\alpha^2\sigma_z^2}{\sigma^3}\Big((I-\lambda W)X\Big)^T(I-\lambda W)(y-X\beta)A^T\\
&\Big(\alpha\big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big)+A\Big)\Big)
\end{aligned}
$$

Second we have,

$$\frac{\partial^2 \ell}{\partial\lambda\partial\sigma} = \frac{\partial}{\partial\sigma}\left(\frac{\partial\ell}{\partial\lambda}\right)$$
$$= \frac{\partial}{\partial\sigma}\left(\frac{\sigma_z}{\sigma}\Big(W\big(y-X\beta\big)\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)- \frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}W_i\big(y-X\beta\big)A_i - tr\Big((I-\lambda W)^{-1}W\Big)\right)$$

Let start differentiating the first expression in $\frac{\partial\ell}{\partial\lambda}$ with respect to σ,

$$\frac{\partial}{\partial\sigma}\left(\frac{\sigma_z}{\sigma}\Big(W\big(y-X\beta\big)\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)\right)$$
$$= -\frac{\sigma_z}{\sigma^2}\Big(W\big(y-X\beta\big)\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)-\frac{\sigma_z^2}{\sigma^3}\Big(W\big(y-X\beta\big)\Big)^T (I-\lambda W)(y-X\beta)\Big)$$
$$= -\frac{\sigma_z}{\sigma^2}\Big(W\big(y-X\beta\big)\Big)^T\mu_z\cdot\mathbb{1}-\frac{2\sigma_z^2}{\sigma^3}\Big(W\big(y-X\beta\big)\Big)^T(I-\lambda W)(y-X\beta)$$

Then, we differentiate second expression in $\frac{\partial\ell}{\partial\lambda}$ with respect to σ.

$$\frac{\partial}{\partial\sigma}\left(-\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}W_i(y-X\beta)A_i\right)$$
$$= \frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}W_i(y-X\beta)A_i-\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}W_i(y-X\beta)\left(\frac{\partial(A_i)}{\partial\sigma}\right)$$
$$= \frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}W_i(y-X\beta)A_i-\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}W_i(y-X\beta)\left(\frac{\alpha\sigma_z}{\sigma^2}(I-\lambda W)_i(y-X\beta)A_i \Big(\alpha\big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)+A_i\Big)\right)$$
$$= \frac{\alpha\sigma_z}{\sigma^2}\Big(W(y-X\beta)\Big)^T A-\frac{\alpha^2\sigma_z^2}{\sigma^3}\Big(W(y-X\beta)\Big)^T(I-\lambda W)(y-X\beta)A^T \Big(\alpha\big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big)+A\Big)$$

The third expression in $\frac{\partial\ell}{\partial\lambda}$ give a zero solution of partial derivative with respect to λ. Hence,

Hence,

$$\frac{\partial^2\ell}{\partial\lambda\partial\sigma} = \frac{\alpha\sigma_z}{\sigma^2}\Big(W(y-X\beta)\Big)^T A-\frac{\alpha^2\sigma_z^2}{\sigma^3}\Big(W(y-X\beta)\Big)^T(I-\lambda W)(y-X\beta)A^T \Big(\alpha\big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big)+A\Big)-\frac{\sigma_z}{\sigma^2}\Big(W\big(y-X\beta\big)\Big)^T\mu_z\cdot\mathbb{1}- \frac{2\sigma_z^2}{\sigma^3}\Big(W\big(y-X\beta\big)\Big)^T(I-\lambda W)(y-X\beta)\Big)$$

Third we have,

$$\frac{\partial^2 \ell}{\partial \sigma^2} = \frac{\partial}{\partial \sigma}\left(\frac{\partial \ell}{\partial \sigma}\right)$$
$$= \frac{\partial}{\partial \sigma}\Bigg(\frac{\sigma_z}{\sigma^2}(\mu_z \cdot \mathbb{1})^T(I-\lambda W)(y-X\beta) + \frac{\sigma_z^2}{\sigma^3}\big((I-\lambda W)(y-X\beta)\big)^T(I-\lambda W)$$
$$(y-X\beta) - \frac{n}{\sigma} - \frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}(I-\lambda W)_i(y-X\beta)A_i\Bigg)$$

Let start differentiating the first expression in $\frac{\partial \ell}{\partial \sigma}$ with respect to σ.

$$\frac{\partial}{\partial \sigma}\left(\frac{\sigma_z}{\sigma^2}(\mu_z \cdot \mathbb{1})^T(I-\lambda W)(y-X\beta)\right) = -\frac{2\sigma_z}{\sigma^3}(\mu_z \cdot \mathbb{1})^T(I-\lambda W)(y-X\beta)$$

Then, we differentiate the second expression in $\frac{\partial \ell}{\partial \sigma}$ with respect to σ.

$$\frac{\partial}{\partial \sigma}\left(\frac{\sigma_z^2}{\sigma^3}\big((I-\lambda W)(y-X\beta)\big)^T\big((I-\lambda W)(y-X\beta)\big)\right)$$
$$= -\frac{3\sigma_z^2}{\sigma^4}\big((I-\lambda W)(y-X\beta)\big)^T\big((I-\lambda W)(y-X\beta)\big)$$

Then, we differentiate the third expression in $\frac{\partial \ell}{\partial \sigma}$ with respect to σ.

$$\frac{\partial}{\partial \sigma}\left(-\frac{n}{\sigma}\right) = \frac{n}{\sigma^2}$$

Lastly, we differentiate fourth expression in $\frac{\partial \ell}{\partial \sigma}$ with respect to σ.

$$\frac{\partial}{\partial \sigma}\left(-\frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}(I-\lambda W)_i(y-X\beta)A_i\right)$$
$$= \frac{2\alpha\sigma_z}{\sigma^3}\sum_{i=1}^{n}(I-\lambda W)_i(y-X\beta)A_i - \frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}(I-\lambda W)_i(y-X\beta)\left(\frac{\partial(A_i)}{\partial \sigma}\right)$$
$$= \frac{2\alpha\sigma_z}{\sigma^3}\big((I-\lambda W)_i(y-X\beta)\big)^T A - \frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}(I-\lambda W)_i(y-X\beta)\Bigg(\frac{\alpha\sigma_z}{\sigma^2}(I-\lambda W)_i$$
$$(y-X\beta)A_i\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big) + A_i\Big)\Bigg)$$
$$= \frac{2\alpha\sigma_z}{\sigma^3}\big((I-\lambda W)_i(y-X\beta)\big)^T A - \frac{\alpha^2\sigma_z^2}{\sigma^4}\big((y-\lambda W)(y-X\beta\big)^T(I-\lambda W)(y-X\beta)A^T$$
$$\Big(\alpha\big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big) + A\Big)$$

Hence,

$$\frac{\partial^2 \ell}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{2\sigma_z}{\sigma^3}(\mu_z \cdot \mathbb{1})^T (I-\lambda W)(y-X\beta) - \frac{3\sigma_z^2}{\sigma^4}\big((I-\lambda W)(y-X\beta)\big)^T \big((I-\lambda W)$$
$$(y-X\beta)\big) + \frac{2\alpha\sigma_z}{\sigma^3}\big((I-\lambda W)_i(y-X\beta)\big)^T A - \frac{\alpha^2\sigma_z^2}{\sigma^4}\big((y-\lambda W)(y-X\beta)^T$$
$$(I-\lambda W)(y-X\beta)A^T\Big(\alpha\big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big) + A\Big)$$

Fourth we have,

$$\frac{\partial^2 \ell}{\partial \alpha \partial \sigma} = \frac{\partial}{\partial \sigma}\left(\frac{\partial \ell}{\partial \alpha}\right)$$
$$= \frac{\partial}{\partial \sigma}\left(\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\Big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T \Big((I-\lambda W)(y-X\beta)\Big) - \right.$$
$$\frac{\sqrt{2}}{1+\alpha^2}\Big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T \cdot \mathbb{1} - \frac{n2\alpha}{\pi(1+\alpha^2)^2\sigma_z^2} +$$
$$\left.\sum_{i=1}^{n}\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta) + \frac{\sqrt{2}\alpha}{1+\alpha^2} - \frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)A_i\right)$$

We start differentiating the first expression in $\dfrac{\partial \ell}{\partial \alpha}$ with respect to σ.

$$\frac{\partial}{\partial \sigma}\left(\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\Big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T \Big((I-\lambda W)(y-X\beta)\Big)\right)$$
$$= -\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma^2\sigma_z}\Big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T \big((I-\lambda W)(y-X\beta)\big) +$$
$$\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\Big(-\frac{\sigma_z}{\sigma^2}(I-\lambda W)(y-X\beta)\big)^T (I-\lambda W)(y-X\beta)\Big)$$
$$= -\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma^2\sigma_z}\Big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T \big((I-\lambda W)(y-X\beta)\big) -$$
$$\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma^3}\big((I-\lambda W)(y-X\beta)\big)^T (I-\lambda W)(y-X\beta)$$
$$= -\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma^2\sigma_z}\Big(\mu_z \cdot \mathbb{1}\Big)^T (I-\lambda W)(y-X\beta) - \frac{4\alpha}{\pi(1+\alpha^2)^2\sigma^3}\big((I-\lambda W)(y-X\beta)\big)^T$$
$$(I-\lambda W)(y-X\beta)$$

Then, we differentiate second expression in $\dfrac{\partial \ell}{\partial \alpha}$ with respect to σ.

$$\frac{\partial}{\partial \sigma}\left(-\frac{\sqrt{2}}{1+\alpha^2}\Big(\mu_z \cdot \mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T \cdot \mathbb{1}\right)$$
$$= -\frac{\sqrt{2}}{(1+\alpha^2)\sigma}\big(-\frac{\sigma_z}{\sigma^2}(I-\lambda W)(y-X\beta)\big)^T \cdot \mathbb{1}$$
$$= \frac{\sqrt{2}}{(1+\alpha^2)\sigma^2}\big((I-\lambda W)(y-X\beta)\big)^T \cdot \mathbb{1}$$

Lastly, we differentiate the fourth expression in $\frac{\partial \ell}{\partial \alpha}$ with respect to σ.

$$\frac{\partial}{\partial \sigma}\left(\sum_{i=1}^{n}\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta) + \frac{\sqrt{2}\alpha}{1+\alpha^2} - \frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)A_i\right)$$
$$= \sum_{i=1}^{n}\Big(-\frac{\sigma_z}{\sigma^2}(I-\lambda W)_i(y-X\beta) + \frac{\mu_z^2}{(1+\alpha^2)\sigma^2\sigma_z}(I-\lambda W_i)(y-X\beta)\Big)A_i + \sum_{i=1}^{n}\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta) + \frac{\sqrt{2}\alpha}{1+\alpha^2} - \frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)\left(\frac{\partial(A_i)}{\partial \sigma}\right)$$
$$= \Big(\frac{\mu_z^2}{(1+\alpha^2)\sigma^2\sigma_z}(I-\lambda W)(y-X\beta) - \frac{\sigma_z}{\sigma^2}(I-\lambda W)(y-X\beta\Big)^T A + \sum_{i=1}^{n}\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta) + \frac{\sqrt{2}\alpha}{1+\alpha^2} - \frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)\Big(\frac{\alpha\sigma_z}{\sigma^2}(I-\lambda W)_i (y-X\beta)A_i\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big) + A_i\Big)\Big)$$
$$= \frac{\mu_z^2}{(1+\alpha^2)\sigma^2\sigma_z}\Big((I-\lambda W)(y-X\beta)\Big)^T A - \frac{\sigma_z}{\sigma^2}\Big((I-\lambda W)(y-X\beta\Big)^T A + \frac{\alpha\sigma_z}{\sigma^2}A^T\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta) + \frac{\sqrt{2}\alpha}{1+\alpha^2}\cdot\mathbb{1} - \frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)(y-X\beta)\Big)\Big((I-\lambda W) (y-X\beta)\Big)^T\Big(\alpha\big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big) + A\Big)$$

Hence,

$$\frac{\partial^2 \ell}{\partial\alpha\partial\sigma} = \frac{\sqrt{2}\sigma_z}{1+\alpha^2\sigma^2}\big((I-\lambda W)(y-X\beta\big)^T\cdot\mathbb{1} - \frac{2\alpha}{\pi(1+\alpha^2)^2\sigma^2\sigma_z}\Big(\mu_z\cdot\mathbb{1}\Big)^T(I-\lambda W)(y-X\beta) - \frac{4\alpha}{\pi(1+\alpha^2)^2\sigma^3}\big((I-\lambda W)(y-X\beta)\big)^T(I-\lambda W)(y-X\beta) + \frac{\mu_z^2}{(1+\alpha^2)\sigma^2\sigma_z}\Big((I-\lambda W) (y-X\beta)\Big)^T A - \frac{\sigma_z}{\sigma^2}\Big((I-\lambda W)(y-X\beta\Big)^T A + \frac{\alpha\sigma_z}{\sigma^2}A^T\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta) + \frac{\sqrt{2}\alpha}{1+\alpha^2}\cdot\mathbb{1} - \frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)(y-X\beta)\Big)\Big((I-\lambda W)(y-X\beta)\Big)^T\Big(\alpha\big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big) + A\Big)$$

Appendix H

Second order partial derivative of the log-likelihood with respect to α

μ_z and σ_z are known as a function of α from Equation (3.3) and (3.4), such that $\dfrac{\partial(\mu_z)}{\partial\alpha} = \dfrac{\sqrt{2}}{1+\alpha^2}$ and $\dfrac{\partial(\sigma_z)}{\partial\alpha} = \dfrac{-\mu_z^2}{\alpha(1+\alpha^2)\sigma\sigma_z}$. By letting $u = \mu_z \cdot \mathbb{1} + \dfrac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)$, we have $\dfrac{\partial u}{\partial\alpha} = \dfrac{\sqrt{2}}{1+\alpha^2}\cdot\mathbb{1} - \dfrac{\mu_z^2}{\alpha(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)(y-X\beta)$ and $\dfrac{\partial(\alpha u)}{\partial\alpha} = \mu_z\cdot\mathbb{1} + \dfrac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta) + \dfrac{\sqrt{2}\alpha}{1+\alpha^2}\cdot\mathbb{1} - \dfrac{\mu_z^2}{(1+\alpha^2)\sigma^2\sigma_z}(I-\lambda W)(y-X\beta)$. From Equation (3.3) we can also express $\dfrac{\mu_z^2}{\alpha(1+\alpha^2)} = \dfrac{2\alpha}{\pi(1+\alpha^2)^2}$ to ensure the computational task is defined well.

Now we derive the second order partial derivatives of the log-likelihood with respect to α.

First we have,

$$\begin{aligned}\frac{\partial^2\ell}{\partial\beta\partial\alpha} &= \frac{\partial}{\partial\alpha}\left(\frac{\partial\ell}{\partial\beta}\right)\\ &= \frac{\partial}{\partial\alpha}\left(\frac{\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big) - \frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}(I-\lambda W)_i X A_i\right)\end{aligned}$$

where, $A_i = \left\{\dfrac{\phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}{\Phi\Big(\alpha\big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\big)\Big)}\right\}$,such that,

$$
\begin{aligned}
\frac{\partial(A_i)}{\partial\alpha} &= \frac{\partial}{\partial\alpha}\left(\frac{\frac{1}{\sqrt{2\pi}}\exp\{\frac{-(\alpha u)^2}{2}\}}{\Phi(\alpha u)}\right)\\
&= \frac{\phi(\alpha u)\Big(\frac{-2(\alpha u)}{2}\Big)\Big(\frac{\partial(\alpha u)}{\partial\alpha}\Big)\Phi(\alpha u) - \phi(\alpha u)\Big(\frac{\partial(\alpha u)}{\partial\alpha}\Big)\phi(\alpha u)}{\Big(\Phi(\alpha u)\Big)^2}\\
&= \frac{-\phi(\alpha u)\Big(\frac{\partial(\alpha u)}{\partial\alpha}\Big)\Big((\alpha u)\Phi(\alpha u) + \phi(\alpha u)\Big)}{\Phi(\alpha u)\Phi(\alpha u)}\\
&= -A_i\Big(\frac{\partial(\alpha u)}{\partial\alpha}\Big)\Big((\alpha u) + A_i\Big)\\
&= -A_i\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta) + \frac{\sqrt{2}\alpha}{1+\alpha^2} - \frac{\mu_z^2}{(1+\alpha^2)\sigma^2\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)\\
&\quad \Big(\alpha\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big) + A_i\Big)
\end{aligned}
$$

Let start differentiating first expression in $\dfrac{\partial\ell}{\partial\beta}$ with respect to α,

$$
\begin{aligned}
&\frac{\partial}{\partial\alpha}\left(\frac{\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)\right)\\
&= \frac{1}{\sigma}\Big(\frac{\partial(\sigma_z)}{\partial\alpha}\Big)\Big((I-\lambda W)X\Big)^T\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big) + \frac{\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T\Big(\frac{\partial u}{\partial\alpha}\Big)\\
&= \frac{-\mu_z^2}{\alpha(1+\alpha^2)\sigma_z\sigma}\Big((I-\lambda W)X\Big)^T\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big) + \frac{\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T\\
&\quad\Big(\frac{\sqrt{2}}{1+\alpha^2}\cdot\mathbb{1} - \frac{\mu_z^2}{\alpha(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)(y-X\beta)\Big)\\
&= \frac{\sqrt{2}\sigma_z}{1+\alpha^2\sigma}\Big((I-\lambda W)X\Big)^T\cdot\mathbb{1} - \frac{2\alpha}{\pi(1+\alpha^2)^2\sigma^3}\Big((I-\lambda W)X\Big)^T(I-\lambda W)(y-X\beta) -\\
&\quad \frac{2\alpha}{\pi(1+\alpha^2)^2\sigma_z\sigma}\Big((I-\lambda W)X\Big)^T\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)
\end{aligned}
$$

Next we differentiate second expression in $\dfrac{\partial\ell}{\partial\beta}$ with respect to α

$$
\begin{aligned}
&\frac{\partial}{\partial\alpha}\left(-\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}(I-\lambda W)_i X A_i\right)\\
&= -\frac{1}{\sigma}\Big(\sigma_z + \alpha\Big(\frac{\partial(\sigma_z)}{\partial\alpha}\Big)\Big)\sum_{i=1}^{n}(I-\lambda W)_i X A_i - \frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}(I-\lambda W)_i X\Big(\frac{\partial(A_i)}{\partial\alpha}\Big)
\end{aligned}
$$

$$= -\frac{1}{\sigma}\Big(\sigma_z - \frac{\alpha\mu_z^2}{\alpha(1+\alpha^2)\sigma\sigma_z}\Big)\sum_{i=1}^{n}(I-\lambda W)_i X A_i + \frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}(I-\lambda W)_i X\Big(\alpha\Big(\mu_z + \frac{\sigma_z}{\sigma}$$
$$(I-\lambda W)_i(y-X\beta)\Big) + A_i\Big)A_i\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta) + \frac{\sqrt{2}\alpha}{1+\alpha^2} -$$
$$\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)$$
$$= \frac{\mu_z^2}{(1+\alpha^2)\sigma_z\sigma}\Big((I-\lambda W)X\Big)^T A - \frac{\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T A + \frac{\alpha\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T$$
$$\Big(\alpha\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big) + A\Big)A^T\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta) +$$
$$\frac{\sqrt{2}\alpha}{1+\alpha^2}\cdot\mathbb{1} - \frac{\mu_z^2}{(1+\alpha^2)\sigma^2\sigma_z}(I-\lambda W)(y-X\beta)\Big)$$

Hence,

$$\frac{\partial^2\ell}{\partial\beta\partial\alpha} = \frac{\sqrt{2}\sigma_z}{(1+\alpha^2)\sigma}\big((I-\lambda W)X\big)^T\cdot\mathbb{1} - \frac{2\alpha}{\pi(1+\alpha^2)^2\sigma^3}\big((I-\lambda W)X\big)^T\big((I-\lambda W)$$
$$(y-X\beta)\big) - \frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\Big((I-\lambda W)X\Big)^T\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big) +$$
$$\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}\Big((I-\lambda W)X\Big)^T A - \frac{\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T A + \frac{\alpha\sigma_z}{\sigma}\Big((I-\lambda W)X\Big)^T$$
$$\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta) + \frac{\sqrt{2}\alpha}{1+\alpha^2}\cdot\mathbb{1} - \frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)(y-X\beta)\Big)$$
$$A^T\Big(\alpha\big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big) + A\Big)$$

Second we have,

$$\frac{\partial^2\ell}{\partial\lambda\partial\alpha} = \frac{\partial}{\partial\alpha}\Big(\frac{\partial\ell}{\partial\lambda}\Big)$$
$$= \frac{\partial}{\partial\alpha}\Big(\frac{\sigma_z}{\sigma}\Big(W\big(y-X\beta\big)\Big)^T\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big) -$$
$$\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}W_i\big(y-X\beta\big)A_i - tr\Big((I-\lambda W)^{-1}W\Big)\Big)$$

Let start differentiating first expression in $\dfrac{\partial\ell}{\partial\lambda}$ with respect to α,

$$\frac{\partial}{\partial\alpha}\Big(\frac{\sigma_z}{\sigma}\Big(W\big(y-X\beta\big)\Big)^T\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)\Big)$$
$$= \frac{1}{\sigma}\Big(\frac{\partial(\sigma_z)}{\partial\alpha}\Big)\Big(W(y-X\beta)\Big)^T\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big) +$$
$$\frac{\sigma_z}{\sigma}\Big(W(y-X\beta)\Big)^T\Big(\frac{\partial u}{\partial\alpha}\Big)$$

$$= \frac{-\mu_z^2}{\alpha(1+\alpha^2)\sigma_z\sigma}\Big(W(y-X\beta)\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)+\frac{\sigma_z}{\sigma}\Big(W(y-X\beta)\Big)^T$$
$$\Big(\frac{\sqrt{2}}{1+\alpha^2}\cdot\mathbb{1}-\frac{\mu_z^2}{\alpha(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)(y-X\beta)\Big)$$
$$= \frac{\sqrt{2}\sigma_z}{1+\alpha^2\sigma}\Big(W(y-X\beta)\Big)^T\cdot\mathbb{1}-\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma^2}\Big(W(y-X\beta)\Big)^T(I-\lambda W)(y-X\beta)$$
$$-\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma_z\sigma^2}\Big(W(y-X\beta)\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)$$

Next we differentiate second expression in $\frac{\partial\ell}{\partial\lambda}$ with respect to α.

$$\frac{\partial}{\partial\alpha}\Big(-\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}W_i(y-X\beta)A_i\Big)$$
$$= -\frac{1}{\sigma}\Big(\sigma_z+\alpha\Big(\frac{\partial(\sigma_z)}{\partial\alpha}\Big)\Big)\sum_{i=1}^{n}W_i(y-X\beta)A_i-\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}W(y-X\beta)\Big(\frac{\partial(A_i)}{\partial\alpha}\Big)$$
$$= -\frac{1}{\sigma}\Big(\sigma_z-\frac{\alpha\mu_z^2}{\alpha(1+\alpha^2)\sigma\sigma_z}\Big)\sum_{i=1}^{n}W(y-X\beta)A_i+\frac{\alpha\sigma_z}{\sigma}\sum_{i=1}^{n}W(y-X\beta)\Big(\alpha\Big(\mu_z+\frac{\sigma_z}{\sigma}$$
$$(I-\lambda W)_i(y-X\beta)\Big)+A_i\Big)A_i\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)+\frac{\sqrt{2}\alpha}{1+\alpha^2}-$$
$$\frac{\mu_z^2}{(1+\alpha^2)\sigma^2\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)$$
$$= \quad \frac{\mu_z^2}{(1+\alpha^2)\sigma^2\sigma_z}\Big(W(y-X\beta)\Big)^TA-\frac{\sigma_z}{\sigma}\Big(W(y-X\beta\Big)^TA+\frac{\alpha\sigma_z}{\sigma}\Big(W(y-X\beta)X\Big)^T$$
$$\Big(\alpha\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)+A\Big)A^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)+$$
$$\frac{\sqrt{2}\alpha}{1+\alpha^2}\cdot\mathbb{1}-\frac{\mu_z^2}{(1+\alpha^2)\sigma^2\sigma_z}(I-\lambda W)(y-X\beta)\Big)$$

The third expression in $\frac{\partial\ell}{\partial\lambda}$ give a zero solution of partial derivative with respect to α. Hence,

$$\frac{\partial^2\ell}{\partial\lambda\partial\alpha} = \frac{\sqrt{2}\sigma_z}{(1+\alpha^2)\sigma}\big(W(y-X\beta\big)^T\cdot\mathbb{1}-\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma^2}\big(W(y-X\beta\big)^T\big((I-\lambda W)(y-X\beta)\big)$$
$$-\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma^2\sigma_z}\Big(W(y-X\beta)\Big)^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)+$$
$$\frac{\mu_z^2}{(1+\alpha^2)\sigma^2\sigma_z}\Big(W(y-X\beta)\Big)^TA-\frac{\sigma_z}{\sigma}\Big(W(y-X\beta\Big)^TA+\frac{\alpha\sigma_z}{\sigma}\Big(W(y-X\beta\Big)^T$$
$$\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)+\frac{\sqrt{2}\alpha}{1+\alpha^2}\cdot\mathbb{1}-\frac{\mu_z^2}{(1+\alpha^2)\sigma^2\sigma_z}(I-\lambda W)(y-X\beta)\Big)$$
$$A^T\Big(\alpha\big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big)+A\Big)$$

Third we have,

$$\frac{\partial^2 \ell}{\partial\sigma\partial\alpha} = \frac{\partial}{\partial\alpha}\Big(\frac{\partial\ell}{\partial\sigma}\Big)$$
$$= \frac{\partial}{\partial\alpha}\Big(\frac{\sigma_z}{\sigma^2}(\mu_z \cdot \mathbb{1})^T(I-\lambda W)(y-X\beta) + \frac{\sigma_z^2}{\sigma^3}\Big((I-\lambda W)(y-X\beta)\Big)^T$$
$$\Big((I-\lambda W)(y-X\beta)\Big) - \frac{n}{\sigma} - \frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}(I-\lambda W)_i(y-X\beta)A_i\Big)$$

Let start differentiating the first expression in $\frac{\partial\ell}{\partial\sigma}$ with respect to α.

$$\frac{\partial}{\partial\alpha}\Big(\frac{\sigma_z}{\sigma^2}(\mu_z \cdot \mathbb{1})^T(I-\lambda W)(y-X\beta)\Big)$$
$$= \frac{1}{\sigma^2}(I-\lambda W)(y-X\beta)\Big(\Big(\frac{\partial(\sigma_z)}{\partial\alpha}\Big)(\mu_z \cdot \mathbb{1})^T + \sigma_z(\Big(\frac{\partial(\mu_z)}{\partial\alpha}\Big) \cdot \mathbb{1}\Big)$$
$$= \frac{1}{\sigma^2}(I-\lambda W)(y-X\beta)\Big(\frac{-\mu_z^2}{\alpha(1+\alpha^2)\sigma\sigma_z}(\mu_z \cdot \mathbb{1})^T + \sigma_z\Big(\frac{\sqrt{2}}{1+\alpha^2}\Big) \cdot \mathbb{1}\Big)$$
$$= \frac{\sqrt{2}\sigma_z}{1+\alpha^2}\Big((I-\lambda W)(y-X\beta)\Big)^T \cdot \mathbb{1} - \frac{2\alpha}{\pi(1+\alpha^2)^2\sigma_z\sigma^3}(\mu_z \cdot \mathbb{1})^T(I-\lambda W)(y-X\beta)$$

Next, we differentiate the second expression in $\frac{\partial\ell}{\partial\sigma}$ with respect to α.

$$\frac{\partial}{\partial\alpha}\Big(\frac{\sigma_z^2}{\sigma^3}\big((I-\lambda W)(y-X\beta)\big)^T\big((I-\lambda W)(y-X\beta)\big)\Big)$$
$$= -\frac{4\alpha}{\pi(1+\alpha^2)^2\sigma^4}\big((I-\lambda W)(y-X\beta)\big)^T\big((I-\lambda W)(y-X\beta)\big)$$

Then, we differentiate fourth expression in $\frac{\partial\ell}{\partial\sigma}$ with respect to α.

$$\frac{\partial}{\partial\alpha}\Big(-\frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}(I-\lambda W)_i(y-X\beta)A_i\Big)$$
$$= -\frac{1}{\sigma^2}\Big(\sigma_z + \alpha\Big(\frac{\partial(\sigma_z)}{\partial\alpha}\Big)\Big)\sum_{i=1}^{n}(I-\lambda W)_i(y-X\beta)A_i - \frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}(I-\lambda W)_i(y-X\beta)\Big(\frac{\partial(A_i)}{\partial\alpha}\Big)$$
$$= -\frac{1}{\sigma^2}\Big(\sigma_z - \frac{\alpha\mu_z^2}{\alpha(1+\alpha^2)\sigma\sigma_z}\Big)\sum_{i=1}^{n}(I-\lambda W)_i(y-X\beta)A_i + \frac{\alpha\sigma_z}{\sigma^2}\sum_{i=1}^{n}(I-\lambda W)_i(y-X\beta)$$
$$\Big(\alpha\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big) + A_i\Big)A_i\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta) + \frac{\sqrt{2}\alpha}{1+\alpha^2} -$$
$$\frac{\mu_z^2}{(1+\alpha^2)\sigma^2\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)$$

$$= \frac{\mu_z^2}{(1+\alpha^2)\sigma_z\sigma^3}\Big((I-\lambda W)(y-X\beta)\Big)^T A - \frac{\sigma_z}{\sigma^2}\Big((I-\lambda W)(y-X\beta)\Big)^T A + \frac{\alpha\sigma_z}{\sigma^2}$$
$$\Big((I-\lambda W)(y-X\beta)\Big)^T\Big(\alpha\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)+A\Big)A^T\Big(\mu_z\cdot\mathbb{1}+$$
$$\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)+\frac{\sqrt{2}\alpha}{1+\alpha^2}\cdot\mathbb{1}-\frac{\mu_z^2}{(1+\alpha^2)\sigma^2\sigma_z}(I-\lambda W)(y-X\beta)\Big)$$

The third expression in $\frac{\partial \ell}{\partial \sigma}$ give a zero solution of partial derivative with respect to α. Hence,

$$\frac{\partial^2\ell}{\partial\sigma\partial\alpha} = \frac{\sqrt{2}\sigma_z}{(1+\alpha^2)\sigma^2}\big((I-\lambda W)(y-X\beta\big)^T\cdot\mathbb{1} - \frac{2\alpha}{\pi(1+\alpha^2)^2\sigma^3\sigma_z}\Big(\mu_z\cdot\mathbb{1}\Big)^T(I-\lambda W)$$
$$(y-X\beta) - \frac{4\alpha}{\pi(1+\alpha^2)^2\sigma^4}\big((I-\lambda W)(y-X\beta)\big)^T(I-\lambda W)(y-X\beta)+$$
$$\frac{\mu_z^2}{(1+\alpha^2)\sigma^2\sigma_z}\Big((I-\lambda W)(y-X\beta)\Big)^T A - \frac{\sigma_z}{\sigma^2}\Big((I-\lambda W)(y-X\beta\Big)^T A+$$
$$\frac{\alpha\sigma_z}{\sigma^2}A^T\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)+\frac{\sqrt{2}\alpha}{1+\alpha^2}\cdot\mathbb{1}-\frac{\mu_z^2}{(1+\alpha^2)\sigma^2\sigma_z}(I-\lambda W)$$
$$(y-X\beta)\Big)\Big((I-\lambda W)(y-X\beta)\Big)^T\Big(\alpha\big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\big)+A\Big)$$

Fourth we have,

$$\frac{\partial^2\ell}{\partial\alpha^2} = \frac{\partial}{\partial\alpha}\Big(\frac{\partial\ell}{\partial\alpha}\Big)$$
$$= \frac{\partial}{\partial\alpha}\Bigg(\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\Big((I-\lambda W)(y-X\beta)\Big)-$$
$$\frac{\sqrt{2}}{1+\alpha^2}\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\cdot\mathbb{1}-\frac{n2\alpha}{\pi(1+\alpha^2)^2\sigma_z^2}+$$
$$\sum_{i=1}^{n}\Big(\mu_z+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)+\frac{\sqrt{2}\alpha}{1+\alpha^2}-\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)A_i\Bigg)$$

Let start differentiating the first expression in $\frac{\partial\ell}{\partial\alpha}$ with respect to α.

$$\frac{\partial}{\partial\alpha}\Big(\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\Big((I-\lambda W)(y-X\beta)\Big)\Big)$$
$$= \frac{\partial}{\partial\alpha}\Big(\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\Big)\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\Big((I-\lambda W)(y-X\beta)\Big)+$$
$$\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\Big(\frac{\partial u}{\partial\alpha}\Big)^T\Big((I-\lambda W)(y-X\beta)\Big)$$
$$= \frac{2\sigma(1+\alpha^2)-8\alpha^2\sigma-2\mu_z^2}{\pi(1+\alpha^2)^3\sigma^2\sigma_z}\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\Big((I-\lambda W)(y-X\beta)\Big)+$$
$$\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\Big(\frac{\sqrt{2}}{1+\alpha^2}\cdot\mathbb{1}-\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}(I-\lambda W)(y-X\beta)\Big)^T\Big((I-\lambda W)(y-X\beta)\Big)$$

Where,

$$
\begin{aligned}
&\frac{\partial}{\partial\alpha}\left(\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\right)\\
&=\frac{2}{\pi\sigma}\frac{\partial}{\partial\alpha}\left(\alpha(1+\alpha^2)^{-2}\sigma_z^{-1}\right)\\
&=\frac{2}{\pi\sigma}\left(\left((1+\alpha^2)^{-2}-4\alpha^2(1+\alpha^2)^{-3}\right)\sigma_z^{-1}+\alpha(1+\alpha^2)^{-2}\left(\frac{-\mu_z^2}{\alpha(1+\alpha^2)\sigma\sigma_z}\right)\right)\\
&=\frac{2}{\pi\sigma}\left(\frac{1}{(1+\alpha^2)^2\sigma_z}-\frac{4\alpha^2}{(1+\alpha^2)^3\sigma_z}-\frac{\alpha\mu_z^2}{\alpha(1+\alpha^2)^3\sigma\sigma_z}\right)\\
&=\frac{2}{\pi(1+\alpha^2)^2\sigma\sigma_z}-\frac{8\alpha^2}{\pi(1+\alpha^2)^3\sigma_z}-\frac{2\mu_z^2}{\pi(1+\alpha^2)^3\sigma^2\sigma_z}\\
&=\frac{2\sigma(1+\alpha^2)-8\alpha^2\sigma-2\mu_z^2}{\pi(1+\alpha^2)^3\sigma^2\sigma_z}
\end{aligned}
$$

Next, we differentiate second expression in $\dfrac{\partial\ell}{\partial\alpha}$ with respect to α.

$$
\begin{aligned}
&\frac{\partial}{\partial\alpha}\left(-\frac{\sqrt{2}}{1+\alpha^2}\left(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\right)^T\cdot\mathbb{1}\right)\\
&=\frac{\partial}{\partial\alpha}\left(-\frac{\sqrt{2}}{1+\alpha^2}\right)\left(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\right)^T\cdot\mathbb{1}-\frac{\sqrt{2}}{1+\alpha^2}\left(\frac{\partial u}{\partial\alpha}\right)^T\cdot\mathbb{1}\\
&=\frac{\sqrt{8}\alpha}{(1+\alpha^2)^2}\left(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\right)^T\cdot\mathbb{1}-\frac{\sqrt{2}}{1+\alpha^2}\\
&\quad\left(\frac{\sqrt{2}}{1+\alpha^2}\cdot\mathbb{1}-\frac{\mu_z^2}{\alpha(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)(y-X\beta)\right)^T\cdot\mathbb{1}\\
&=\frac{\sqrt{8}\alpha}{(1+\alpha^2)^2}\left(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\right)^T\cdot\mathbb{1}-\frac{\sqrt{2}}{1+\alpha^2}\\
&\quad\left(\frac{\sqrt{2}}{1+\alpha^2}\cdot\mathbb{1}-\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}(I-\lambda W)(y-X\beta)\right)^T\cdot\mathbb{1}
\end{aligned}
$$

Where,

$$
\begin{aligned}
\frac{\partial}{\partial\alpha}\left(-\frac{\sqrt{2}}{1+\alpha^2}\right)&=\sqrt{2}(1+\alpha^2)^{-2}(2\alpha)\\
&=\frac{\sqrt{8}\alpha}{(1+\alpha^2)^2}
\end{aligned}
$$

Next we differentiate third expression in $\dfrac{\partial \ell}{\partial \alpha}$ with respect to α.

$$
\begin{aligned}
&\frac{\partial}{\partial \alpha}\Big(- \frac{n2\alpha}{\pi(1+\alpha^2)^2\sigma_z^2}\Big) \\
&= \frac{-2n\pi(1+\alpha^2)^2\sigma_z^2 - \Big(2\pi(1+\alpha^2)2\alpha\sigma_z^2 + 2\sigma_z \frac{\partial(\sigma_z)}{\partial\alpha}\pi(1+\alpha^2)^2\Big)\Big(-n2\alpha\Big)}{\pi^2(1+\alpha^2)^4\sigma_z^4} \\
&= \frac{-2n\pi(1+\alpha^2)^2\sigma_z^2 + 8n\pi\alpha^2(1+\alpha^2)\sigma_z^2 + 4n\pi\alpha(1+\alpha^2)^2\sigma_z^2\Big(\frac{-\mu_z}{\alpha(1+\alpha^2)\sigma\sigma_z^2}\Big)}{\pi^2(1+\alpha^2)^4\sigma_z^4} \\
&= \frac{8n\pi\alpha^2(1+\alpha^2)\sigma\sigma_z^2 - 2n\pi(1+\alpha^2)\sigma\sigma_z^2 - 4n\pi\mu_z(1+\alpha^2)\sigma_z^2}{\pi^2(1+\alpha^2)^4\sigma_z^4} \\
&= \frac{8n\alpha^2\sigma\sigma_z^2 - 2n(1+\alpha^2)\sigma\sigma_z - 4n\mu_z}{\pi(1+\alpha^2)^3\sigma_z^3}
\end{aligned}
$$

Finally, we differentiate the fourth expression in $\dfrac{\partial \ell}{\partial \alpha}$ with respect to α.

$$
\begin{aligned}
&\frac{\partial}{\partial\alpha}\Bigg(\sum_{i=1}^{n}\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta) + \frac{\sqrt{2}\alpha}{1+\alpha^2} - \frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)A_i\Bigg) \\
&= \sum_{i=1}^{n}\Bigg(\frac{\partial}{\partial\alpha}\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta) + \frac{\sqrt{2}\alpha}{1+\alpha^2} - \frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)A_i\Bigg) + \\
&\quad \sum_{i=1}^{n}\Bigg(\Big(\mu_z + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta) + \frac{\sqrt{2}\alpha}{1+\alpha^2} - \frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Big)\Big(\frac{\partial(A_i)}{\partial\alpha}\Big)\Bigg) \\
&= \Bigg(\frac{\sqrt{2}}{1+\alpha^2}\cdot\mathbb{1} - \frac{\mu_z^2}{\alpha(1+\alpha^2)\sigma^2\sigma_z}(I-\lambda W)(y-X\beta) + \frac{\sqrt{2}(1-\alpha^2)}{(1+\alpha^2)^2}\cdot\mathbb{1} + \\
&\quad (\frac{2\pi\mu_z^2\alpha(1+\alpha^2)\sigma\sigma_z^2 - \sqrt{8}\pi\mu_z(1+\alpha^2)\sigma\sigma_z^2 - 2\alpha\mu_z^2)}{\pi(1+\alpha^2)^3\sigma^2\sigma_z^3})(I-\lambda W)_i(y-X\beta)\Bigg)^T A - \\
&\quad \Bigg(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta) + \frac{\sqrt{2}\alpha}{1+\alpha^2}\cdot\mathbb{1} - \frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Bigg)^T A \\
&\quad \Bigg(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta) + \frac{\sqrt{2}\alpha}{1+\alpha^2}\cdot\mathbb{1} - \frac{\mu_z^2}{(1+\alpha^2)\sigma^2\sigma_z}(I-\lambda W)_i(y-X\beta)\Bigg)^T \\
&\quad \Bigg(\alpha\Big(\mu_z\cdot\mathbb{1} + \frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big) + A_i\Bigg)
\end{aligned}
$$

Where,

$$
\begin{aligned}
\frac{\partial}{\partial\alpha}\Big(\frac{\sqrt{2}\alpha}{1+\alpha^2}\Big) &= \frac{\sqrt{2}(1+\alpha^2) - 2\alpha\sqrt{2}\alpha}{(1+\alpha^2)^2} \\
&= \frac{\sqrt{2} + \sqrt{2}\alpha^2 - 2\sqrt{2}\alpha^2}{(1+\alpha^2)^2} \\
&= \frac{\sqrt{2}(1-\alpha^2)}{(1+\alpha^2)^2}
\end{aligned}
$$

and,

$$\begin{aligned}
&\frac{\partial}{\partial\alpha}\Big(-\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}\Big)\\
&=\frac{-2\mu_z\dfrac{\partial(\mu_z)}{\partial\alpha}(1+\alpha^2)\sigma\sigma_z-\Big((2\alpha)\sigma\sigma_z+(1+\alpha^2)\sigma\dfrac{\partial(\sigma_z)}{\partial\alpha}\Big)\Big(-\mu_z^2\Big)}{\Big((1+\alpha^2)\sigma\sigma_z\Big)^2}\\
&=\frac{-2\mu_z\Big(\dfrac{\sqrt{2}}{1+\alpha^2}\Big)(1+\alpha^2)\sigma\sigma_z+2\alpha\sigma\sigma_z\mu_z^2+(1+\alpha^2)\sigma\mu_z^2\Big(\dfrac{-\mu_z^2}{\alpha(1+\alpha^2)\sigma\sigma_z}\Big)}{\Big((1+\alpha^2)\sigma\sigma_z\Big)^2}\\
&=\frac{-\sqrt{8}\mu_z\sigma\sigma_z+2\alpha\mu_z^2\sigma\sigma_z-(1+\alpha^2)\sigma\mu_z^2\Big(\dfrac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\Big)}{(1+\alpha^2)^2\sigma^2\sigma_z^2}\\
&=\frac{2\pi\mu_z^2\alpha(1+\alpha^2)\sigma\sigma_z^2-\sqrt{8}\pi\mu_z(1+\alpha^2)\sigma\sigma_z^2-2\alpha\mu_z^2}{\pi(1+\alpha^2)^3\sigma^2\sigma_z^3}
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{\partial^2\ell}{\partial\alpha^2}=&\frac{2\sigma(1+\alpha^2)-8\alpha^2\sigma-2\mu_z^2}{\pi(1+\alpha^2)^3\sigma^2\sigma_z}\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\Big((I-\lambda W)(y-X\beta)\Big)+\\
&\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}\Big(\frac{\sqrt{2}}{1+\alpha^2}\cdot\mathbb{1}-\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}(I-\lambda W)(y-X\beta)\Big)^T\Big((I-\lambda W)(y-X\beta)\Big)+\\
&\frac{\sqrt{8}\alpha}{(1+\alpha^2)^2}\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)(y-X\beta)\Big)^T\cdot\mathbb{1}-\frac{\sqrt{2}}{1+\alpha^2}\\
&\Big(\frac{\sqrt{2}}{1+\alpha^2}\cdot\mathbb{1}-\frac{2\alpha}{\pi(1+\alpha^2)^2\sigma\sigma_z}(I-\lambda W)(y-X\beta)\Big)^T\cdot\mathbb{1}+\\
&\Bigg(\frac{8n\alpha^2\sigma\sigma_z^2-2n(1+\alpha^2)\sigma\sigma_z-4n\mu_z}{\pi(1+\alpha^2)^3\sigma_z^3}\Bigg)\cdot\mathbb{1}+\\
&\Bigg(\frac{\sqrt{2}}{1+\alpha^2}\cdot\mathbb{1}-\frac{\mu_z^2}{\alpha(1+\alpha^2)\sigma^2\sigma_z}(I-\lambda W)(y-X\beta)+\frac{\sqrt{2}(1-\alpha^2)}{(1+\alpha^2)^2}\cdot\mathbb{1}+\\
&\quad(\frac{2\pi\mu_z^2\alpha(1+\alpha^2)\sigma\sigma_z^2-\sqrt{8}\pi\mu_z(1+\alpha^2)\sigma\sigma_z^2-2\alpha\mu_z^2)}{\pi(1+\alpha^2)^3\sigma^2\sigma_z^3})(I-\lambda W)_i(y-X\beta)\Bigg)^T A-\\
&\Bigg(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)+\frac{\sqrt{2}\alpha}{1+\alpha^2}\cdot\mathbb{1}-\frac{\mu_z^2}{(1+\alpha^2)\sigma\sigma_z}(I-\lambda W)_i(y-X\beta)\Bigg)^T A\\
&\Bigg(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)+\frac{\sqrt{2}\alpha}{1+\alpha^2}\cdot\mathbb{1}-\frac{\mu_z^2}{(1+\alpha^2)\sigma^2\sigma_z}(I-\lambda W)_i(y-X\beta)\Bigg)^T\\
&\Bigg(\alpha\Big(\mu_z\cdot\mathbb{1}+\frac{\sigma_z}{\sigma}(I-\lambda W)_i(y-X\beta)\Big)+A_i\Bigg)
\end{aligned}$$

References

Acreman, M. C. (1985). Predicting the mean annual flood from basin characteristics in scotland. *Hydrological sciences journal*, 30(1):37–49.

Adamowski, J. and Sun, K. (2010). Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology*, 390(1):85–91.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*, volume 4. Springer Science & Business Media.

Anselin, L. (2003). Data and spatial weights in spdep, notes and illustrations. *Spatial Analysis Laboratory, University of Illinois at Urbana-Champaign. Retrieved online https:// geodacenter. asu. edu/ system/ files/ dataweights. pdf.*

Arbia, G. (2014). *A primer for spatial econometrics: with applications in R*. Springer.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, pages 171–178.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):579–602.

Badyalina, B. and Shabri, A. (2013). Streamflow forecasting at ungauged sites using multiple linear regression. *MATEMATIKA*, 29:67–75.

Bivand, R. (2017). Creating neighbours. 2015. *Available online: cran. r-project. org/web/packages/spdep/vignettes/nb. pdf (accessed on 17 March 2015).*

Bivand, R., Altman, M., Anselin, L., Assunção, R., Berke, O., Bernat, A., and Blanchet, G. (2015). Package spdep. *See ftp://garr. tucows. com/mirrors/CRAN/web/packages/spdep/spdep. pdf (accessed 9 December 2015).*

Bivand, R. S. and Portnov, B. A. (2004). Exploring spatial data analysis techniques using r: The case of observations with no neighbors. In *Advances in Spatial Econometrics*, pages 121–142. Springer.

Bocchiola, D., De Michele, C., and Rosso, R. (2003). Review of recent advances in index flood estimation. *Hydrology and Earth System Sciences Discussions*, 7(3):283–296.

Brath, A., Castellarin, A., Franchini, M., and Galeati, G. (2001). Estimating the index flood using indirect methods. *Hydrological sciences journal*, 46(3):399–418.

Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *The annals of statistics*, 24(6):2350–2383.

Burnham, K. P. and Anderson, D. R. (2003). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media.

Canuti, P. and Moisello, U. (1982). Relationship between the yearly maxima of peak and daily discharge for some basins in tuscany. *Hydrological Sciences Journal*, 27(2):111–128.

Castellarin, A., Camorani, G., and Brath, A. (2007). Predicting annual and long-term flow-duration curves in ungauged basins. *Advances in Water Resources*, 30(4):937–953.

CEH (2009). FEH CD-ROM 3.0. Centre for Ecology and Hydrology, Wallingford, U.K.

Chen, K., Anthony, S. M., and Granick, S. (2014). Extending particle tracking capability with delaunay triangulation. *Langmuir*, 30(16):4760–4766.

Chow, V. T., Maidment, D. R., and Mays, L. W. (1988). *Applied hydrology*.

Cliff, A. and Ord, K. (1972). Testing for spatial autocorrelation among regression residuals. *Geographical analysis*, 4(3):267–284.

Cliff, A. D. and Ord, J. K. (1973). *Spatial autocorrelation*, volume 5. Pion London.

Cliff, A. D. and Ord, J. K. (1981). *Spatial processes: models & applications*, volume 44. Pion London.

Cole, G. (1965). An application of the regional analysis of flood flows. In *Proceedings of the Symposium on River Flood Hydrology*. Institution of Civil Engineers.

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*, volume 208. Springer.

Coppola, A., Stewart, B., and Okazaki, N. (2014). *lbfgs: Limited-memory BFGS Optimization*. R package version 1.2.1.

Cunnane, C. (1988). Methods and merits of regional flood frequency analysis. *Journal of Hydrology*, 100(1):269–290.

Cunnane, C. (1989). *Statistical distributions for flood frequency analysis*. WMO (Series). Secretariat of the World Meteorological Organization.

Dalla Valle, A. (2004). The skew-normal distribution. *Skew-elliptical distributions and their applications: A journey beyond normality*, pages 3–24.

Dalrymple, T. (1960). Flood-frequency analyses, manual of hydrology: Part 3. Technical report, USGPO,.

Dey, D. (2010). *Estimation of the parameters of skew normal distribution by approximating the ratio of the normal density and distribution functions*. PhD thesis, UNIVERSITY OF CALIFORNIA, RIVERSIDE.

Draper, N. and Smith, H. (1981). *Applied Regression Analysis*. Number pt. 766 in Applied Regression Analysis. Wiley.

Dray, S., Saïd, S., and Débias, F. (2008). Spatial ordination of vegetation data using a generalization of wartenberg's multivariate spatial correlation. *Journal of Vegetation Science*, 19(1):45–56.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., et al. (2004). Least angle regression. *The Annals of statistics*, 32(2):407–499.

Eng, K., Milly, P., and Tasker, G. D. (2007). Flood regionalization: a hybrid geographic and predictor-variable region-of-influence regression method. *Journal of Hydrologic Engineering*, 12(6):585–591.

Estivill-Castro, V. and Houle, M. E. (2001). Robust distance-based clustering with applications to spatial data mining. *Algorithmica*, 30(2):216–242.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360.

Frank, L. E. and Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, 35(2):109–135.

Friedman, J. H. and Stuetzle, W. (1981). Projection pursuit regression. *Journal of the American statistical Association*, 76(376):817–823.

Gao, J., Lu, Z., and Tjøstheim, D. (2006). Estimation in semiparametric spatial regression. *The Annals of Statistics*, 34(3):1395–1435.

Garde, R. and Kothyari, U. (1990). Flood estimation in indian catchments. *Journal of Hydrology*, 113(1):135–146.

Grimaldi, S., Kao, S.-C., Castellarin, A., Papalexiou, S.-M., Viglione, A., Laio, F., Aksoy, H., and Gedikli, A. (2011). 2.18 - statistical hydrology. In Wilderer, P., editor, *Treatise on Water Science*, pages 479 – 517. Elsevier, Oxford.

Hamed, K. and Rao, A. R. (1999). *Flood frequency analysis*. CRC press.

Hong, X. and Mitchell, R. J. (2007). Backward elimination model construction for regression and classification using leave-one-out criteria. *International journal of systems science*, 38(2):101–113.

Hosking, J. and Wallis, J. (1993). Some statistics useful in regional frequency analysis. *Water resources research*, 29(2):271–281.

Hosking, J. and Wallis, J. R. (1997). Regional frequency analysis. *Regional Frequency Analysis, by JRM Hosking and James R. Wallis, pp. 240. ISBN 0521430453. Cambridge, UK: Cambridge University Press, April 1997.*, page 240.

IH (1999). *Flood Estimation Handbook, 5 Volumes*. Institute of Hydrology.

Jaafar, W., Zurina, W., and Han, D. (2012). Uncertainty in index flood modelling due to calibration data sizes. *Hydrological Processes*, 26(2):189–201.

Kazembe, L. N., Kleinschmidt, I., Holtz, T. H., and Sharp, B. L. (2006). Spatial analysis and mapping of malaria risk in malawi using point-referenced prevalence of infection data. *International Journal of Health Geographics*, 5(1):1.

Kişi, Ö. (2004). River flow modeling using artificial neural networks. *Journal of Hydrologic Engineering*, 9(1):60–63.

Kjeldsen, T. R. and Jones, D. (2007). Estimation of an index flood using data transfer in the uk. *Hydrological sciences journal*, 52(1):86–98.

Kjeldsen, T. R. and Jones, D. A. (2006). Prediction uncertainty in a median-based index flood method using l moments. *Water resources research*, 42(7).

Kjeldsen, T. R. and Jones, D. A. (2009). An exploratory analysis of error components in hydrological regression modeling. *Water resources research*, 45(2).

Kjeldsen, T. R. and Jones, D. A. (2010). Predicting the index flood in ungauged uk catchments: on the link between data-transfer and spatial model error structure. *Journal of Hydrology*, 387(1):1–9.

Kjeldsen, T. R., Jones, D. A., and Bayliss, A. C. (2008). *Improving the FEH statistical procedures for flood frequency estimation*. Environment Agency.

Kosfeld, R. and Dreger, C. (2006). Thresholds for employment and unemployment: A spatial analysis of german regional labour markets, 1992–2000. *Papers in Regional Science*, 85(4):523–542.

LeSage, J. P. (2008). An introduction to spatial econometrics. *Revue d'économie industrielle*, (3):19–44.

Lettenmaier, D. P., Wallis, J. R., and Wood, E. F. (1986). *Effect of Regional Heterogeniety on Flood Frequency Estimation*. IBM Thomas J. Watson Research Division.

Liseo, B. and Loperfido, N. (2003). A bayesian interpretation of the multivariate skew-normal distribution. *Statistics & probability letters*, 61(4):395–401.

Lu, Z., Lundervold, A., Tjøstheim, D., and Yao, Q. (2007). Exploring spatial nonlinearity using additive approximation. *Bernoulli*, pages 447–472.

Malekinezhad, H., Nachtnebel, H., and Klik, A. (2011). Comparing the index-flood and multiple-regression methods using l-moments. *Physics and Chemistry of the Earth, Parts A/B/C*, 36(1-4):54–60.

Mamun, A. A., Hashim, A., and Amir, Z. (2011). Regional statistical models for the estimation of flood peak values at ungauged catchments: Peninsular malaysia. *Journal of Hydrologic Engineering*, 17(4):547–553.

Maus, A. and Drange, J. M. (2010). All closest neighbors are proper delaunay edges generalized, and its application to parallel algorithms. *Proceedings of Norwegian informatikkonferanse*, pages 1–12.

Mazvimavi, D., Meijerink, A., Savenije, H., and Stein, A. (2005). Prediction of flow characteristics using multiple regression and neural networks: a case study in zimbabwe. *Physics and Chemistry of the Earth, Parts A/B/C*, 30(11):639–647.

McCuen, R. H., Leahy, R. B., and Johnson, P. A. (1990). Problems with logarithmic transformations in regression. *Journal of Hydraulic Engineering*, 116(3):414–428.

Merz, R. and Blöschl, G. (2005). Flood frequency regionalisationspatial proximity vs. catchment attributes. *Journal of Hydrology*, 302(1):283–306.

Mimikou, M. and ordios, J. (1989). Predicting the mean annual flood and flood quantiles for ungauged catchments in greece. *Hydrological sciences journal*, 34(2):169–184.

Monti, A. C. (2003). A note on the estimation of the skew normal and the skew exponential power distributions. *Metron*, 61(2):205–219.

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2):17–23.

Morris, D. (2003). Automation and appraisal of the feh statistical procedures for flood frequency estimation. *Science report FD1603 to Defra, Centre for Ecology and Hydrology, Wallingford, UK*.

Mudholkar, G. S. and Hutson, A. D. (2000). The epsilon–skew–normal distribution for analyzing near-normal data. *Journal of Statistical Planning and Inference*, 83(2):291–309.

Nelles, O. (2001). Nonlinear system identification: from classical approach to neuro-fuzzy identification.

NERC (1975). *Flood Studies Report, 5 Volumes*. Natural Environment Research Council.

NRC-US (1988). *Estimating Probabilities of Extreme Floods: Methods and Recommended Research*. National Academy Press.

Pandey, G. and Nguyen, V.-T.-V. (1999). A comparative study of regression based methods in regional flood frequency analysis. *Journal of Hydrology*, 225(1):92–101.

Paradis, E. (2016). Morans autocorrelation coefficient in comparative methods. [Retrived Online;https://cran.r-project.org/web/packages/ape/vignettes/ MoranI.pdf].

Peel, M. C., Wang, Q., Vogel, R. M., and McMAHON, T. A. (2001). The utility of l-moment ratio diagrams for selecting a regional probability distribution. *Hydrological Sciences Journal*, 46(1):147–155.

Pewsey, A. (2000). Problems of inference for azzalini's skewnormal distribution. *Journal of Applied Statistics*, 27(7):859–870.

Pilgrim, D. and Institution of Engineers, A. (2001). *Australian Rainfall and Runoff: A Guide to Flood Estimation*. Australian Rainfall and Runoff: A Guide to Flood Estimation. Vol. 1. Institution of Engineers, Australia.

Prentice, R. L. (1975). Discrimination among some parametric models. *Biometrika*, 62(3):607–614.

Reimers, W. (1990). Estimating hydrological parameters from basin characteristics for large semiarid catchments. *IAHS Publication n*, 191:187–194.

Robson, A. J. and Reed, D. W. (1999). *Flood Estimation Handbook, Vol. 3: Statistical Procedures for Flood Frequency Estimation*. Institute of Hydrology.

Samsudin, R., Saad, P., and Shabri, A. (2011). River flow time series using least squares support vector machines. *Hydrology and Earth System Sciences*, 15(6):1835.

Sartori, N. (2006). Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. *Journal of Statistical Planning and Inference*, 136(12):4259–4275.

Sene, K. (2008). *Flood warning, forecasting and emergency response*. Springer Science & Business Media.

Shu, C. and Burn, D. H. (2004). Artificial neural network ensembles and their application in pooled flood frequency analysis. *Water Resources Research*, 40(9).

Soetaert, K. (2014). Package rootsolve: roots, gradients and steady-states in r.

Solomatine, D. and Wagener, T. (2011). Hydrological modeling. *Treatise on Water Science*, 2:435–457.

Stedinger, J. R. (1993). Frequency analysis of extreme events. *Handbook of hydrology*, 18.

Stedinger, J. R. and Tasker, G. D. (1986). Regional hydrologic analysis, 2, model-error estimators, estimation of sigma and log-pearson type 3 distributions. *Water Resources Research*, 22(10):1487–1499.

Sveinsson, O. G., Boes, D. C., and Salas, J. D. (2001). Population index flood method for regional frequency analysis. *Water Resources Research*, 37(11):2733–2748.

Tasker, G. D. and Stedinger, J. R. (1989). An operational gls model for hydrologic regression. *Journal of Hydrology*, 111(1):361–375.

Team, R. C. (2014). R: A language and environment for statistical computing.. r foundation for statistical computing, vienna, austria. url http://www. r-project. org/(date of access 01/12/2014). Technical report, Accessed 01/12.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

Turner, M. E. (1960). 150. note: On heuristic estimation methods. *Biometrics*, 16(2):299–301.

Vogel, R. M. and Fennessey, N. M. (1993). L moment diagrams should replace product moment diagrams. *Water Resources Research*, 29(6):1745–1752.

Wagener, T., Wheater, H. S., and Gupta, H. V. (2004). *Rainfall-runoff modelling in gauged and ungauged catchments*. World Scientific.

Wan Jaafar, W., Liu, J., and Han, D. (2011). Input variable selection for median flood regionalization. *Water Resources Research*, 47(7).

Wang, W.-C., Chau, K.-W., Cheng, C.-T., and Qiu, L. (2009). A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of hydrology*, 374(3):294–306.

Wilson, D., Fleig, A. K., Lawrence, D., Hisdal, H., Pettersson, L.-E., and Holmqvist, E. (2011). A review of nves flood frequency estimation procedures. *NVE Report*, pages 9–2011.

Yee, T. W. and Yee, M. T. (2018). Package vgam. R package version 1.0-5.

Zhang, H. and El-Shaarawi, A. (2010). On spatial skew-gaussian processes and applications. *Environmetrics*, 21(1):33–47.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.