UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS

School of Ocean and Earth Sciences

Surface Forcing of the North Atlantic: Accuracy and Variability

by

David Inglis Berry

Thesis for the degree of Doctor of Philosophy

November 2009

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS

SCHOOL OF OCEAN AND EARTH SCIENCES

Doctor of Philosophy

SURFACE FORCING OF THE NORTH ATLANTIC: ACCURACY AND VARIABILITY

by David Inglis Berry

A new methodology to estimate the turbulent air – sea heat and moisture fluxes and their uncertainty is developed and assessed using Voluntary Observing Ship (VOS) observations. Whilst important drivers of the global oceanic and atmospheric circulation these fluxes remain poorly quantified, both in terms of mean value and uncertainty. The new methodology addresses both of these issues and is extensible to other data sources.

The individual observations are first bias and height adjusted to remove systematic errors and the impact of changing observing heights. They are then characterised in terms of random errors using a semi-variogram analysis and a range of variogram models. The data quality and sampling are then taken into account using optimal interpolation (OI) to grid the observations, producing daily mean fields and uncertainty estimates. These are then used to estimate the fluxes and flux uncertainty on both daily and monthly time scales.

Comparisons of the mean fields and fluxes to the original input data and to independent buoy observations show the fields not to be significantly biased. The adjustments applied before gridding and flux calculation are also shown to improve the agreement with the buoy observations. The uncertainty estimates are assessed using a series of cross validation experiments and 3-way error analyses to make alternative estimates of the uncertainty. These alternative estimates are shown to be of the same order of magnitude as the OI uncertainty estimates and generally to be within 10 – 20% of the OI estimate. Whilst all three estimates are similar there are some systematic differences. The OI uncertainty estimates tend to be lower (higher) than the alternative estimates in high (low) variability regions.

The representation of the variability in the new dataset is examined and shown to be improved compared to previous VOS based datasets. The adjustments are shown to have little impact on the temporal trends in temperature and humidity whilst reducing the wind speed and sensible and latent heat flux trends. These reduced trends are thought to be more realistic. The wind speed trend after adjustment is more similar to the trends reported in previous studies using reanalysis model output. However, there are still some differences in the trends, with the VOS based estimates larger, leading to uncertainty in trend estimates. The trends in the adjusted latent and sensible heat flux estimates are similar to those seen in other flux datasets but when compared to changes in the upper ocean heat content may still be too large. This may be due to the overestimate of the wind speed trend. Overall the uncertainty in the wind speed trend gives the largest uncertainty in the flux trends.

Finally, the advances made in developing the new methodology are summarised and the potential uses of the new dataset identified. Future work and improvements are then suggested.

List of Contents

ABSTRACT ..... *ii*

List of Contents ..... *iii*

List of Figures ..... *viii*

List of Tables ..... *viii*

DECLARATION OF AUTHORSHIP ..... *xvii*

Acknowledgements ..... *xviii*

1 Chapter 1 - Introduction *1*

1.1 Air – sea interaction and the surface fluxes **1**

1.2 Observing and Estimating the Surface Fluxes **1**

1.3 Research Aims and Layout of Thesis **3**

2 Surface Fluxes and the Bulk Formulae *5*

2.1 Introduction **5**

2.2 Derivation of the Bulk Formulae **5**

2.3 Parameterisations **9**

2.3.1 Introduction 9

2.3.2 Transfer Coefficients 9

2.3.3 Stability Correction 11

2.3.4 Non turbulent fluxes 14

2.4 Summary and Discussion **16**

3 Chapter 3 – Data Sources and Surface Flux Data Sets *18*

3.1 Introduction **18**

3.2 Observational Data Sources **18**

3.2.1 Air Temperature 18

3.2.2 Humidity 22

3.2.3 Sea Surface Temperature 26

3.2.4 Wind speed 29

3.2.5 Sea level pressure 32

3.2.6 Cloud Cover 32

3.3 Model Data Sources 32

3.4 Previous Flux Datasets Construction Methods 34

3.4.1 Arithmetic Mean 34

3.4.2 Sampling and Classical Fluxes 34

3.4.3 Successive Correction 35

3.4.4 Objective Analysis 37

3.4.5 Optimal Interpolation 37

3.5 Flux Datasets 40

3.5.1 In situ based estimates 40

3.5.2 Reanalysis based estimates 41

3.5.3 Satellite and combined satellite / reanalysis based estimates 42

3.6 Random Errors and Flux Uncertainties 43

3.6.1 Random and Systematic Uncertainty 43

3.6.2 Semi-variogram analysis 44

3.6.3 Sampling Error 46

3.6.4 Flux Uncertainty 48

3.7 Discussion and Summary 50

4 Development of Dataset Construction Methodology 52

4.1 Introduction 52

4.2 Outline of method and Rationale 52

4.3 Data Sources Used 54

4.3.1 VOS Observations and Metadata 54

4.3.2 Quality Control 55

4.4 Bias Corrections 56

4.4.1 Air Temperature 56

4.4.2 Humidity 58

4.4.3 Visual Wind Speeds 59

4.4.4 Height Corrections 59

4.5 Random and Bias Uncertainty 60

4.5.1 Random Errors 60
4.5.2 Bias Uncertainties 61
4.6 Optimal Interpolation 61
4.6.1 Application to VOS observations 61
4.7 Flux Calculation and Uncertainty Estimates 64
4.8 Calculating the Monthly Mean Uncertainty 65
4.9 Discussion 68
4.9.1 Quality Control 68
4.9.2 Bias Corrections 68
4.9.3 Random Error Estimation 69
4.9.4 Optimal Interpolation and Error Correlations 69
4.10 Summary 70
5 Results: Initial Processing and Flux Estimates 71
5.1 Introduction 71
5.2 Quality control and track checking 71
5.3 Air Temperature Bias Estimates and Fitting the Bias Adjustment 72
5.3.1 Heating Error Bias Estimates 72
5.3.2 Bias Adjustment Coefficients 73
5.4 Random Errors 75
5.4.1 Example semi-variograms and choice of variogram model 75
5.4.2 Impact of Bias and Height Corrections 78
5.4.3 Global Error Estimates 82
5.5 Optimal Interpolation and Flux Calculation 83
5.5.1 Mean values 83
5.5.2 Random and Sampling Uncertainty 85
5.5.3 Bias Uncertainty 87
5.5.4 Flux Total Uncertainty 89
5.5.5 Impact of Bias Corrections on the Mean Fields 90
5.6 Summary 92
6 Cross Validation and Comparisons 93

6.1 Introduction 93

6.2 Comparison to VOS Observations 93

6.3 Comparison to WHOI UOP Mooring Data Archive 98

6.3.1 Introduction 98

6.3.2 Case Study – Comparison to the North East Subduction Buoy 99

6.3.3 Comparison of Mean Differences 105

6.4 3 way Error Analysis 116

6.5 Cross Validation 121

6.6 Discussion 125

6.6.1 Comparisons of mean differences 125

6.6.2 Comparisons of error estimates 127

6.7 Summary 128

7 Monthly means, trends and variability 130

7.1 Introduction 130

7.2 Variability 130

7.2.1 Comparison with the gridded dataset 130

7.2.2 Comparison with the SOC Climatology 133

7.2.3 Geophysical Consistency 137

7.3 Trends and Impact of Adjustments 138

7.3.1 Introduction 138

7.3.2 Trend Estimation 139

7.3.3 Impact of Adjustments 142

7.4 Discussion 146

7.4.1 Air temperature, humidity and SST trends 146

7.4.2 Wind Speed Trends 148

7.4.3 Heat Flux Trends 148

7.5 Summary 150

8 Summary, Further Work and Conclusions 151

8.1 Summary of aims and results 151

8.2 Key Improvements 151

8.3 Requirements for surface flux datasets and uses of the new dataset 153

8.4 Future Work 154

8.4.1 Bias adjustments and bias uncertainties 154

8.4.2 Gridding 155

8.4.3 Flux Calculation and Uncertainty 156

8.4.4 Other Improvements 156

8.5 Summary 156

Appendix A – Acronyms, Abbreviations and Definitions 158

Appendix B – Nomenclature 160

Appendix C – Coefficients for air temperature bias correction 165

List of References 166

List of Figures

Figure 2-1: Momentum stability corrections (ψ_m) from the Smith (1980, 1988) parameterisation (black) and COARE 3.0 (Fairall *et al*. 2003) algorithm (red) as a function of stability (ζ).....14

Figure 3-1: Monthly mean number of VOS observations of air temperature from ICOADS (e.g. Worley *et al*. 2005) during 1975 (top left), 1985 (top right), 1995 (bottom left) and 2005 (bottom right). 19

Figure 3-2: Stacked area plot showing number of VOS air temperature observations per month between 40S – 70N and 100W – 30E for each of the main air temperature exposure observing methods. 19

Figure 3-3: Mean air temperature observation height over the region 40S – 70N and 100W – 30E from ICOADS and Pub. 47 (see Section 4.3). A 12 month running mean filter has been applied.....21

Figure 3-4: Grid cells with air temperature observations at least once every 6 hours from moored buoys during: 1975 (top left); 1985 (top right); 1995 (bottom left); and 2005 (bottom right). 22

Figure 3-5: As Figure 3-2 but for humidity sensor exposure.....23

Figure 3-6: Impact of bias corrections of Kent *et al*. (1993b) (solid) and Josey *et al*. (1999) (dashed) on specific humidity as a function of specific humidity.....25

Figure 3-7: As Figure 3-2 but for the SST measurement method.....27

Figure 3-8: As Figure 3-1 but for SST observations flagged as coming from drifting buoys during 1975 (top left), 1985 (top right), 1995 (bottom left) and 2005 (bottom right).....28

Figure 3-9: As Figure 3-2 but for the wind speed measurement method. 30

Figure 3-10: As Figure 3-3 but for anemometer height. 31

Figure 3-11: Examples of the different variogram models: exponential (red); Gaussian (black); linear (green). $C_0 = 2$, $C_1 = 8$ and $a_{sv} = 500$ for the Gaussian and exponential models. $C_0 = 2$ and $h_{sv} = 1/150$ for the linear model.....46

Figure 4-1: Differences between screen and psychrometer specific humidity observations averaged globally (left) and zonally (right). Differences before

correction (black), after correction following Kent *et al*. (1993b) (green) and after correction following Josey *et al*. (1999) (red) are shown. The differences after a reduction of 3.3 % in the screen specific humidity observations are also shown (blue). The inset shows the regions with at least 3 observations from both methods and at least 5 years of data. 58

Figure 5-1: Number of observations for each variable per year before (black) and after quality control (red). Also shown are the number of observations per year after QC and track checking (green). 72

Figure 5-2: Diurnal cycle of estimated bias in air temperature observations (black squares) and residual bias (red squares) after applying the new bias correction. The shaded region indicates the bias uncertainty. 74

Figure 5-3: Diurnal cycle of the bias correction using the parameters of the best fit from 1995 (solid line) and fixed environmental conditions of: relative wind speed = 7 ms^{-1}, cloud cover = 3 oktas; latitude = 40 °N. Also shown is the average bias correction from the top 10 runs (diamonds) calculated using the same environmental conditions. The error bars give the standard deviation correction from the top 10 fits. 74

Figure 5-4: a) example semi-variograms (left) and b) difference (residual) between the modelled and observed variograms (right) over the North Atlantic during July 1993 for air temperature. The models and residuals shown are the linear (red), exponential (blue) and Gaussian (green) models. The sample semi-variogram (black) is also shown in the left hand plot. 76

Figure 5-5: As Figure 5-4 but for wind speed observations and February 1997. 77

Figure 5-6: Sample semi-variograms for unadjusted (black), height corrected (red) and bias and height adjusted (green) air temperature observations during July 1993 and between 30N – 60N and 60W – 30W. 79

Figure 5-7: Time series of random error estimates for unadjusted (black), height adjusted (red) and bias and height adjusted (green) air temperature observations made in the region 40S – 70N and 100W – 30E. A 12 month running mean filter has been applied. 80

Figure 5-8: As Figure 5-7 but for specific humidity. 81

Figure 5-9: As Figure 5-7 but for wind speed observations. 81

Figure 5-10: Time series of random error estimates averaged globally for the basic variables with 12 month running mean filter applied. Estimates for unadjusted (black); height adjusted (red) and bias and height adjusted (green) observations are shown. The horizontal lines show the mean values. 83

Figure 5-11a: Mean daily values (top) averaged over the period December 1992 – February 1993 for air temperature, sea surface temperature, wind speed and specific humidity. Also shown is the mean daily uncertainty for the different fields (bottom). 84

Figure 5-12a: As Figure 5-11a but for June – August 1993. 86

Figure 5-13: Mean daily bias uncertainty averaged over December 1992 – February 1993 (top) and June – August 1993 (bottom) for Sea Surface Temperature, latent heat flux and sensible heat flux. 88

Figure 5-14: Total uncertainty estimates in the daily fluxes averaged over December 1992 – February 1993 (top) and June – August 1993 (bottom). 89

Figure 5-15: Impact of the bias corrections on daily estimates of: air temperature (left); specific humidity (left middle); latent heat flux (right middle); and sensible heat flux (right) averaged over December 1992 – February 1993 (top) and June – August 1993 (bottom). 91

Figure 6-1a: Histograms of the residuals (VOS Observations – optimally interpolated values) for each of the main variables during January 1993. Also shown are the mean residuals as a function of the OI estimates (inset plots). 94

Figure 6-2: Comparison of residuals from simulated wind speed dataset as a function of wind speed. 96

Figure 6-3: Time series of monthly mean residuals (black) from the comparison of VOS observations with the output of the OI over the Atlantic (40S – 70N, 100W – 30 E). Also shown are the RMS errors (red) between the VOS observations and the output of the OI. Smoothed values, using a 12 month running mean filter, are also shown. 97

Figure 6-4: Location of the buoys listed in Table 6-2. 99

Figure 6-5: Comparison of daily values from the new dataset (red) and values from the North East Subduction buoy (black) during June – August 1992. The grey shading indicates the uncertainty in the output from the OI due to random and sampling errors. The bias uncertainty is not included. 102

Figure 6-6: As Figure 6-5 but for December 1992 – February 1993. 104

Figure 6-7: Mean daily air temperature (left) and mean daily uncertainty (right) during January 1992 from the new dataset. The locations of the different Subduction buoys are shown as white squares....106

Figure 6-8: Mean differences (product – buoy) between the North West Subduction buoy observations and a number of different flux products for each of the main variables and fluxes. An x on the abscissa indicates that the data is unavailable for that dataset....107

Figure 6-9: As Figure 6-8 but for the South West Subduction buoy. 108

Figure 6-10: As Figure 6-8 but for the MLML91 mooring. 109

Figure 6-11: As Figure 6-8 but for the CMO mooring. 112

Figure 6-12: As Figure 6-8 but for the SESMOOR mooring. 113

Figure 6-13: Ratio of the corrected standard deviation of the daily air temperature values across the ensemble of OI runs to the mean daily uncertainty estimate averaged for the two month periods shown. Values greater than 1 indicate an underestimate of the uncertainty by the OI whilst values less than 1 indicate an overestimate. The '1' contour is in bold. 122

Figure 6-14: As Figure 6-13 but for sea surface temperature 123

Figure 6-15: As Figure 6-13 but for specific humidity 124

Figure 6-16: As Figure 6-13 but for sea level pressure....124

Figure 6-17: As Figure 6-13 but for wind speed....125

Figure 7-1: Ratio of intermonth variability from the new dataset to the variability from the gridded data averaged over the calendar months for 1970 – 2006....132

Figure 7-2: Correlation between air temperature and SST anomalies from the gridded data, smoothed on a number of different scales, and NOCS 2.0. 133

Figure 7-3: Ratio of intermonth variability from the new dataset to the variability from the SOC climatology averaged over the calendar months. The contour line indicates a ratio of 1. 134

Figure 7-4: Comparison of intermonth variability of the basic variables between the NOCS 2.0 and the SOC Climatology. Results are for the Atlantic (40S – 70N, 100W – 30E) and the period 1980 – 2005. Well-sampled regions are shown in green, poorly sampled regions in red and intermediate regions in black. The blue line indicates a one to one relationship (equality) 135

Figure 7-5: Time series of air temperature estimates at 27.5 N 54.5 W from the SOC climatology (black) and NOCS 2.0 (red). 136

Figure 7-6: Correlation between the surface turbulent heat flux (latent + sensible) and the SST tendency from the: gridded data (top left); SOC climatology (top right); and new dataset (NOCS 2.0, bottom left). The correlations have been calculated over the period 1980 to 2002. 137

Figure 7-7: Results of fit of linear model to air temperature anomalies using an OLS regression (black) and the CO procedure (red). Shown are: time series of the air temperature anomalies (green) and linear models estimated using OLS (black) and the CO procedure (red) (top left); time series of the error terms (ε_t and u_t for the OLS and CO respectively) from the two regressions (middle left); auto-correlation function for the error terms (bottom left); standardized relative histogram of the errors (top right); and quantile – quantile plot for the standardized errors (sample quantiles) vs quantiles from the normal distribution (theoretical quantiles) (bottom right). w is the Shapiro-Wilk test statistic. 141

Figure 7-8: Time series of monthly anomalies (relative to the period 1970 – 2006) for each of the main variables and fluxes for the Atlantic. Adjusted (black) and unadjusted (red) time series are shown together with a model of the linear trend (bold lines). 145

Figure 7-9: Air temperature anomalies calculated from the SOC climatology (red) and the new dataset (black) for the period 1980 – 2002 averaged over the Atlantic. Also shown are linear trend models fitted to both time series. 146

Figure 7-10: Difference between anomalies of air temperature, humidity and SST averaged over 1983 to 1986 (NAO +ve) and 1978 to 1981 (NAO –ve) (1983 to 1986 minus 1978 to 1981) from the adjusted dataset. 147

List of Tables

Table 2-1: Recent global flux datasets and bulk formulae used 9

Table 2-2: Drag coefficients (C_{D10n}) or momentum roughness length (z_0) used in the different bulk formulae. u_{10n} is the 10 m neutral wind speed; ν the kinematic viscosity of air; a_c the Charnock parameter; g the acceleration due to gravity; L_p the wavelength associated with the dominant wave period; h_s the significant wave height and C_p the phase speed of the dominant wave. 10

Table 2-3: As Table 2-2 but for heat and moisture transfer coefficients or roughness lengths. R_r is the roughness Reynolds number; a_1, a_2, b_1 and b_2 are constants given by Liu et al (1979).....11

Table 3-1: Specific humidity accuracy as a function of temperature for a relative humidity of 80 % and an accuracy of 2 % (i.e. RH = 80 ± 2 %). A pressure of 101.3 kPa has been used. 26

Table 3-2: Random and systematic uncertainty estimates from Blanc (1986) and Gleckler and Weare (1997). V_s = wind speed (m s^{-1}).....43

Table 3-3: Globally averaged (weighted by ocean area) random error estimates from Kent and Berry (2005). 45

Table 3-4: Commonly used variogram models. $\gamma(h_{sv})$ is the modelled variogram at separation h_{sv}; $C_0 \geq 0$ and gives the value at zero separation (i.e. the intercept), $C_0 + C_1$ gives the sill value (i.e. the value at which the variogram levels off). The range is given by $3a_{sv}$ for the exponential and Gaussian models. b_{sv} gives the gradient for the linear model. 45

Table 4-1 ICOADS trimming flags. σ_l is the difference between the 1st and 3rd sextiles and σ_u the difference between the 3rd and 5th sextile. δ is the difference between the observation and the climatological median value. 55

Table 4-2: Trimming flags checked for each variable used 55

Table 4-3: Lower and upper limits for initial x values in fitting routine. 57

Table 5-1: Percentage of observations removed by the quality control (QC) and track checking for each variable averaged over 1970 – 2006. 71

Table 5-2: Summary of fitting statistics for the different variogram models. The mean random error estimates, number of months and grid cells with valid fits and mean r^2 are given. The uncertainty on the mean random errors is the standard deviation of the error estimates over all valid grid cells. The rows in bold indicate the model chosen for each field. 78

Table 5-3: Globally averaged random error estimates for the each variable. The random errors have additionally been averaged over 1970 – 2006 and the uncertainties given are one standard deviation. The values listed for air temperature, humidity and wind speed from Kent and Berry (2005) are the height adjusted values with the 3.5σ trimming limits applied.....82

Table 6-1: Monthly mean residuals over 1970 – 2006 for each variable. Also listed for each variable are: the maximum absolute monthly mean residual; the mean RMS error averaged over 1970 – 2006; and the random error estimates for unadjusted observations and using the chosen variogram models (Table 5-2). 95

Table 6-2: Buoys observations used from the WHOI UOP Data Archive. Also listed are the buoy locations and the periods with valid data. 98

Table 6-3: Typical long-term accuracy of the basic observations made by the WHOI buoys (e.g. Crescenti and Weller 1992; Moyer and Weller 1997). Also listed are the bias uncertainties for the VOS observations (see Chapter 4). The values listed for latent and sensible heat fluxes are based on the Subduction buoys (Moyer and Weller 1997). 99

Table 6-4: Mean (μ) and standard deviations (σ) of the variables shown in Figure 6-5 and Figure 6-6 for the new dataset (NOCS OI) and the NE Subduction buoy. The mean difference and root mean squared error between the new dataset and the buoy observations are also listed together with the mean random and sampling uncertainty for the daily OI values (S). 100

Table 6-5: Mean values (± standard deviation) of the buoy observations for the different Subduction buoy deployments and number of days with valid data. Also shown are mean differences (Δ) (product – buoy) from the different datasets and the standard deviation (σ) of the differences. Also listed for the OI are the total error estimates and t-statistic for each comparison. t-statistics marked with a star (*) are significant at the 95 % confidence level. 110

Table 6-6: As Table 6-5 but for the MLML91 mooring 114

Table 6-7: As Table 6-5 but for the SESMOOR and CMO moorings.....115

Table 6-8: 3-way error analysis performed at each buoy location.....117

Table 6-9: Uncertainty estimates from the OI compared to the error estimates from the 3-way error analysis of the new dataset, the buoy observations and the dataset listed. The error range gives the error estimate from the 3-way analysis assuming no correlation between datasets and assuming zero error in the buoy observations. 118

Table 7-1: Estimated trends for the different variables, air – sea gradients and fluxes before and after adjustment. The uncertainties shown are the 95 % confidence interval for the trend estimates (±1.96 σ). Negative trends in the air – sea gradients and fluxes indicate an increase in the magnitude of the trends. For comparison, the trends from the first SOC climatology (1980 – 2002) are listed together with the trends from the new adjusted dataset over the same period. The values in italics are adversely affected by large outliers in the SOC climatology.....143

DECLARATION OF AUTHORSHIP

I, DAVID INGLIS BERRY

declare that the thesis entitled

SURFACE FORCING OF THE NORTH ATLANTIC: ACCURACY AND VARIABILITY

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;
- parts of this work have been published as:

Berry, D. I., and E. C. Kent, 2009: A New Air - Sea Interaction Gridded Data Set from ICOADS with Uncertainty Estimates. *Bulletin Of The American Meteorological Society*, **90,** 645 - 656, DOI: 10.1175/2008BAMS2639.1.

Berry, D. I., and E. C. Kent, 2010: Air-Sea Fluxes from ICOADS: The Construction of a New Gridded Dataset with Uncertainty Estimates. *International Journal of Climatology*, **In press**, DOI: 10.1002/joc.2059.

Berry, D. I., E. C. Kent, and P. K. Taylor, 2004: An analytical model of heating errors in marine air temperatures from ships. *Journal of Atmospheric and Oceanic Technology*, **21,** 1198-1215, DOI: 10.1175/1520-0426(2004)021<1198:AAMOHE>2.0.CO;2.

Signed: ……..

Date:……

Acknowledgements

I would like to thank my supervisors – Elizabeth Kent, Peter Taylor and Harry Bryden – for providing guidance and agreeing to supervise the project. I am also grateful to Peter Killworth for acting as panel chair. Special thanks also go to Elizabeth Kent for suggesting the project and providing continual encouragement throughout the project. Richard W. Reynolds and Tom Smith from the National Climatic Data Center and Diane Stokes from the National Centers for Environmental Prediction provided the original optimal interpolation code. The original flux calculation subroutines were provided by the James Rennell Division's meteorological team. Additional guidance, comments on the work and encouragement were provided by Margaret Yelland and Simon Josey amongst others. Finally, I would like to thank David Parker and Neil Wells for providing useful comments and suggestions during the viva that have improved this thesis.

1 Chapter 1 - Introduction

1.1 Air – sea interaction and the surface fluxes

The circulation of the oceans and atmosphere is influenced and driven by the interaction and fluxes across the air – sea interface. For example, at high latitudes the overlying atmosphere cools the ocean surface, leading to the sinking and formation of deep waters. This forms one of the driving mechanisms of the thermohaline circulation. Another example is the tropical atmospheric convection. At low latitudes the strong warming and evaporation from the ocean surface reduces the density of the overlying atmosphere, leading to the strong atmospheric convection typical of the tropics and the formation of the upward limb of the tropical Hadley Cells. Understanding the interactions at the air – sea interface is therefore key to understanding the atmospheric and oceanic circulations and the Earth's climate system.

The main physical interactions at the interface are the turbulent heat, moisture (evaporation and precipitation) and momentum fluxes between the ocean and atmosphere, the radiative warming of the ocean surface by solar (shortwave) radiation and the net cooling of the ocean surface by the thermal (longwave) radiation. The oceans also act as a source or sink for other climatically important quantities such as CO_2 and other gases and aerosols (e.g. Huebert *et al*. 2004). Whilst the radiative and biogeochemical fluxes are important, this study is restricted to the conductive and convective heat transfer between the ocean and atmosphere and the evaporation from the ocean surface, i.e. the turbulent latent and sensible heat fluxes, over the Atlantic Basin (defined as 40S – 70N, 100W – 30E in this thesis) and their uncertainty.

1.2 Observing and Estimating the Surface Fluxes

Direct measurements of the heat, moisture and momentum fluxes are possible (e.g. Edson *et al*. 1991). However, accurate, high frequency measurements corrected for the platform motions and flow distortion are required. This has so far resulted in direct measurements only being made on air – sea interaction research cruises and dedicated moored buoys and platforms. In turn, this has limited the availability of direct measurements in space and time. As a result, global marine flux datasets have typically been constructed using bulk estimates of the mean meteorological parameters and parameterisations known as "bulk formulae". These parameterisations have been

developed using co-incident measurements of the direct fluxes and mean meteorological variables (e.g. Fairall *et al*. 1996, 2003).

The bulk formulae require estimates of the air temperature and humidity, sea surface temperature, wind speed and sea level pressure as input. Various sources exist for these parameters, including *in situ* measurements made by Voluntary Observing Ships (VOS), moored and drifting buoys, satellite retrievals and model output. Each source has its own advantages and disadvantages. The satellite estimates give improved spatial coverage compared to *in situ* measurements but cannot accurately recover all the variables required for flux calculation. Satellite estimates are also limited to the last three decades. The moored buoys can give more accurate measurements but are limited spatially. The drifting buoys improve spatial coverage in some regions, but only report a subset of the required parameters and only provide data globally for the last two decades. The atmospheric reanalysis model estimates give global coverage with a relatively high time resolution. However, the low spatial resolutions typical of the models may lead to problems in coastal and high variability regions. Errors in the model physics can also lead to biased flux estimates from the reanalyses. The VOS observations are known to contain significant biases, to be of variable quality and to have uneven sampling with the observations clustered over the major shipping lanes. However, the VOS observations have been well characterised in terms of random errors (e.g. Kent and Berry 2005), bias (e.g. Cardone *et al*. 1990; Kent *et al*. 1993b; Folland and Parker 1995; Kent and Kaplan 2006) and metadata on observing practices (e.g. Kent *et al*. 2007). Additionally, the VOS provide all the parameters required to estimate the fluxes and give a multi-decadal series of observations extending back over 50 years.

There is a long history of estimating regional and global air – sea fluxes from the VOS data. Early examples include the atlases of Bunker (1976), Hsuing (1986) and Oberhuber (1988). Bunker (1976) estimated the fluxes for individual VOS observations over the North Atlantic and averaged these onto a monthly semi-regular grid dependant on the data density and natural variability. Hsiung (1986) and Oberhuber (1988) calculated monthly mean values for the bulk meteorological variables and used these to calculate the fluxes. More recently, da Silva *et al*. (1994) and Josey *et al*. (1999) estimated the fluxes for individual VOS reports and averaged the individual estimates onto a regular grid to give monthly mean and climatological values of the fluxes. Both

datasets used successive correction (e.g. da Silva *et al*. 1994) to smooth and fill gaps in the fields produced.

In addition to the *in situ* based estimates, flux datasets have been generated using satellite and atmospheric reanalysis model data. Estimates are directly available from reanalysis models (Kalnay *et al*. 1996; Uppala *et al*. 2005; Onogi *et al*. 2007) on sub-daily timescales. Daily and monthly estimates have been produced using satellite measurements (Chou *et al*. 1995; Chou *et al*. 1997; Bakan *et al*. 2000; Kubota *et al*. 2002; Bentamy *et al*. 2003; Chou *et al*. 2003). The majority of these datasets used data from reanalysis models to supplement the satellite data for variables, such as air temperature, that cannot be reliably retrieved by satellite measurements. The Woods Hole Objectively Analyzed air – sea Fluxes (OAFlux) (Yu *et al*. 2004a, 2004b; Yu and Weller 2007; Yu *et al*. 2008) dataset used a blend of data from multiple satellites and reanalysis models to estimate the bulk meteorological variables on a daily 1° grid. Fluxes were then calculated from these bulk estimates.

Each of these datasets differed in the construction methods, bias and error estimation and bulk formulae used. As a result, there are differences between flux products, even in the case where the same input data are used. For example, da Silva *et al*. (1994) and Josey *et al*. (1999) both used VOS observations from the (International) Comprehensive Ocean Atmosphere Data Set (e.g. Worley *et al*. 2005) but applied different bias corrections, quality control and slightly different gridding procedures. In addition to the uncertainty that this introduces, the variable, and in the case of the VOS, often sparse sampling, observational errors and un-parameterised effects lead to an artificial imbalance in the net heat fluxes with the oceans gaining an unrealistic amount of heat. In order to reconcile the differences between the different datasets, to understand the closure problem and to understand the natural variability, realistic estimates of the uncertainties in the flux estimates are needed. None of the flux products or datasets currently available includes realistic uncertainty estimates.

1.3 Research Aims and Layout of Thesis

There are two aims to this thesis. The first aim is to develop a methodology for making a comprehensive error analysis and calculating the uncertainty in the fluxes routinely as an integral part of the dataset production. Estimates of the uncertainty due to the random, sampling, and bias uncertainties will be presented alongside the gridded

fields of the fluxes and meteorological variables for the first time. The second aim is to perform an assessment of the method developed, determining whether the uncertainty estimates produced are realistic and examining the impact of the improvements on the mean fluxes, their trends and variability. Initially the method development and analysis will be performed for the Atlantic Basin (40S – 70N, 100W – 30E). A longer-term goal is to then use the method developed to produce a new version of the SOC (now renamed NOC) climatology, including the wind stress and radiative fluxes.

Chapters 2 and 3 review previous work on the surface fluxes. The bulk formulae, their derivation, and a number of commonly used parameterisations, are summarised in Chapter 2. Chapter 3 discusses different sources of the mean meteorological parameters and past flux datasets. Reviews of the gridding methods used in those datasets and of previous flux uncertainty estimates are also given.

Chapter 4 describes the methodology developed in this research, including the rationale behind the choices made. The individual components and methods used are described together with a description of the data and quality control applied. This is followed by a description of the bias corrections developed and applied to the individual observations and the estimation of the uncertainty in those observations. The gridding procedure, which takes into account the uncertainty in the individual observations, is then described. This is followed by the estimation of the fluxes, their uncertainty and monthly mean values.

The results of the quality control, bias correction and uncertainty estimation are described in Chapter 5. The basic output from the gridding procedure, estimates of the basic variables, the fluxes and their uncertainty are examined. Chapter 6 compares the results of the gridding and flux estimation to other data sources and evaluates the uncertainty estimates based on the comparisons. The improvements made and the trends and variability in the new dataset are examined and evaluated in Chapter 7. A discussion and summary of the results and method developed is given in Chapter 8.

2 Surface Fluxes and the Bulk Formulae

2.1 Introduction

Direct measurements of the turbulent fluxes of heat and moisture over the oceans are both difficult and expensive to make. These measurements require continuous high frequency observations (e.g. Large and Pond 1981) together with detailed information on the motion of the observing platform (e.g. Edson *et al*. 1998) and the flow distortion over the platform (e.g. Yelland *et al*. 2002). Hence, direct flux measurements are only made on research vessels and towers or other stable platforms and, as a result, are limited in both time and space. In order to get large-scale estimates of the fluxes over the oceans we need to turn to other sources of data and indirect flux estimation methods.

Based on direct measurements of the fluxes made during dedicated research cruises, together with observations of the bulk meteorological variables, many different parameterisations have been developed for the turbulent fluxes (e.g. Liu *et al*. 1979; Smith 1980, 1988; Clayson *et al*. 1996; Fairall *et al*. 2003). These parameterisations have been used to make global estimates of the heat fluxes over the oceans (see Chapter 3). In this chapter the derivation of the bulk formulae is given (Section 2.2) followed by a review of commonly used parameterisations (Section 2.3). A discussion of the different bulk formulae and their application to the VOS data is given (Section 2.4).

2.2 Derivation of the Bulk Formulae

The boundary layer is defined as the bottom layer of the troposphere that is directly influenced by the surface forcing on time scales of an hour or less (e.g. Stull 1988). This layer can vary from several hundred to several thousand metres in depth depending on the weather and atmospheric conditions. The bottom 10 % of this layer is usually referred to as the surface layer, within which the turbulent fluxes and stresses typically vary by less than 10 % of their magnitude. Studies of air – sea interaction are concerned with fluxes between this surface layer and the underlying ocean.

The bulk formulae and flux parameterisations have been developed making the assumption that the horizontal fluxes in the surface layer are homogeneous, invariant with height and that the mean vertical wind speed is zero (e.g. Busch 1972). However, small-scale features in the surface layer such as vertical plumes and convergence lines

will influence the fluxes and result in a non-zero local vertical wind speed (e.g. Stull 1988). On the scale of the boundary layer, atmospheric organization, such as horizontal roll vortices, will also act to moderate the fluxes and result in local non-zero mean vertical wind speeds. These small scale and boundary layer scale features have not been explicitly included in the parameterisations. They will, however, be included implicitly in parameterizations developed under a wide range of conditions. This implicit inclusion will act to increase the uncertainty in the parameterizations.

Making the assumption of horizontally homogeneous fluxes, zero mean vertical wind speed and using Reynolds averaging allows the vertical fluxes of momentum (τ), sensible (H) and latent (E) heat to be written as (e.g. Busch 1972)

$$\tau = -\rho_0 \overline{u'w'} \qquad \text{Eq. 2-1 a}$$

$$H = c_p \rho_0 \overline{\theta'w'} \qquad \text{Eq. 2-1 b}$$

$$E = L_v \rho_0 \overline{q'w'} \qquad \text{Eq. 2-1 c}$$

where ρ_0 is the density of air (kg m^{-3}), c_p the specific heat capacity of air at constant pressure (J kg K^{-1}), L_v the latent heat of vaporization (J kg^{-1}) and u', w', θ' and q' the turbulent components of the horizontal wind speed (m s^{-1}), vertical wind speed (m s^{-1}), potential temperature (K) and specific humidity (g kg^{-1}) respectively. The overbar indicates a time average. These fluxes are assumed constant with height in the surface layer (e.g. Busch 1972) and can be used to define vertically invariant scaling parameters: the friction velocity u_*, the characteristic temperature T_* and the characteristic humidity q_*.

$$u_*^2 = \overline{u'w'} \qquad \text{Eq. 2-2 a}$$

$$u_* T_* = \overline{\theta'w'} \qquad \text{Eq. 2-2 b}$$

$$u_* q_* = \overline{q'w'} \qquad \text{Eq. 2-2 c}$$

These scaling parameters can then be related to dimensionless profiles of the wind speed, humidity and temperature (Stull 1988)

$$\frac{kz}{u_*}\frac{\partial \bar{u}}{\partial z} = \phi_m \qquad \text{Eq. 2-3 a}$$

$$\frac{kz}{T_*}\frac{\partial \bar{T}}{\partial z} = \phi_t \qquad \text{Eq. 2-3 b}$$

$$\frac{kz}{q_*}\frac{\partial \bar{q}}{\partial z} = \phi_q \qquad \text{Eq. 2-3 c}$$

where k is the von Karman constant, z the observation height, ϕ_m the dimensionless wind shear, ϕ_t and ϕ_q the dimensionless gradients for temperature and humidity respectively. From empirical estimates of the dimensionless profiles and measurements of the wind speed, air temperature and humidity at two different heights the fluxes can be estimated by integrating Eq. 2-3 between those two heights. Over the oceans the integration is usually done between a height just above the surface, z_0 (known as the momentum roughness length), and the observation height, z. Analogous roughness lengths are also defined for heat (z_{ot}) and humidity (z_{oq}). Estimates of the mean variables at the sea surface are then used to represent the values just above the surface. The integrals of Eq. 2-3 are given by (e.g. WGASF 2000)

$$u_z = u_0 + \frac{u_*}{k}\left(\ln\left(\frac{z}{z_0}\right) - \psi_m\right) \qquad \text{Eq. 2-4 a}$$

$$T_z = T_0 + \frac{T_*}{k}\left(\ln\left(\frac{z_t}{z_{0t}}\right) - \psi_t\right) \qquad \text{Eq. 2-4 b}$$

$$q_z = q_0 + \frac{q_*}{k}\left(\ln\left(\frac{z_q}{z_{0q}}\right) - \psi_q\right) \qquad \text{Eq. 2-4 c}$$

where u_0, T_0 and q_0 are the wind speed, temperature and humidity at the air – sea interface. Similarly, u_z, T_z and q_z are the wind speed, air temperature and humidity at heights z, z_t and z_q respectively. The ψ terms are commonly referred to as stability corrections and given by

$$\psi_x = \int_{z_0}^{z} \frac{1-\phi_x}{z}\,dz \qquad \text{Eq. 2-5}$$

where the subscript x refers to the momentum, humidity or temperature profile as appropriate. Rearranging Eq. 2-4 so the scaling parameters are on the LHS we have

$$u_* = k\left(\ln\left(\frac{z}{z_0}\right) - \psi_m\right)^{-1}\left(u_z - u_0\right) \qquad \text{Eq. 2-6 a}$$

$$T_* = k\left(\ln\left(\frac{z_t}{z_{0t}}\right) - \psi_t\right)^{-1}\left(T_z - T_0\right) \qquad \text{Eq. 2-6 b}$$

$$q_* = k\left(\ln\left(\frac{z_q}{z_{0q}}\right) - \psi_q\right)^{-1}\left(q_z - q_0\right) \qquad \text{Eq. 2-6 c}$$

Substituting Eq. 2-6 into Eq. 2-1 gives the bulk formulae in terms of the mean meteorological variables and transfer coefficients

$$\tau = -\rho_0 C_D\left(u_z - u_0\right)^2 \qquad \text{Eq. 2-7 a}$$

$$H = c_p \rho_0 C_H\left(u_z - u_0\right)\left(T_z - T_0\right) \qquad \text{Eq. 2-7 b}$$

$$E = L_v \rho_0 C_E\left(u_z - u_0\right)\left(q_z - q_0\right) \qquad \text{Eq. 2-7 c}$$

where C_D, C_H and C_E are the drag, heat and moisture transfer coefficients respectively and are given by

$$C_D = k^2\left(\ln\left(\frac{z_u}{z_0}\right) - \psi_m\right)^{-2} \qquad \text{Eq. 2-8 a}$$

$$C_H = k^2\left(\ln\left(\frac{z_u}{z_0}\right) - \psi_m\right)^{-1}\left(\ln\left(\frac{z_t}{z_{0t}}\right) - \psi_t\right)^{-1} \qquad \text{Eq. 2-8 b}$$

$$C_E = k^2\left(\ln\left(\frac{z_u}{z_0}\right) - \psi_m\right)^{-1}\left(\ln\left(\frac{z_q}{z_{0q}}\right) - \psi_q\right)^{-1} \qquad \text{Eq. 2-8 c}$$

Under neutral conditions the stability corrections, ψ_x, are equal to zero and the neutral transfer coefficients at a reference height, usually chosen to be 10 metres, given by

$$C_{D10n} = k^2\left(\ln\left(\frac{10}{z_0}\right)\right)^{-2} \qquad \text{Eq. 2-9 a}$$

$$C_{H10n} = k^2\left(\ln\left(\frac{10}{z_0}\right)\right)^{-1}\left(\ln\left(\frac{10}{z_{0t}}\right)\right)^{-1} \qquad \text{Eq. 2-9 b}$$

$$C_{E10n} = k^2\left(\ln\left(\frac{10}{z_0}\right)\right)^{-1}\left(\ln\left(\frac{10}{z_{0q}}\right)\right)^{-1} \qquad \text{Eq. 2-9 c}$$

Based on estimates of the neutral transfer coefficients equations Eq. 2-7 to Eq. 2-9 can be solved iteratively to give the fluxes and estimates of the mean wind speed, air temperature and humidity at a 10 m reference level. The estimates of the wind speed (u_{10s}), air temperature (T_{10s}) and humidity (q_{10s}) height corrected to 10 m are given by

$$u_{10s} = u_z - \frac{u_*}{k}\left(\ln\left(\frac{z_u}{10}\right) - \psi_m\left(\frac{z_u}{L}\right) + \psi_m\left(\frac{10}{L}\right)\right) \quad \text{Eq. 2-10 a}$$

$$T_{10s} = T_z - \frac{t_*}{k}\left(\ln\left(\frac{z_t}{10}\right) - \psi_t\left(\frac{z_t}{L}\right) + \psi_t\left(\frac{10}{L}\right)\right) \quad \text{Eq. 2-10 b}$$

$$q_{10s} = q_z - \frac{q_*}{k}\left(\ln\left(\frac{z_q}{10}\right) - \psi_q\left(\frac{z_q}{L}\right) + \psi_q\left(\frac{10}{L}\right)\right) \quad \text{Eq. 2-10 c}$$

2.3 Parameterisations

2.3.1 Introduction

Different parameterisations exist for the bulk formulae, varying in: the parameterisations used for dimensionless profiles (ϕ_m, ϕ_t and ϕ_q); how the values at the sea surface interface are estimated; and how the values of the roughness lengths or neutral transfer coefficients are estimated. Sources of the bulk formulae used in recent flux global products are given in Table 2-1.

Table 2-1: Recent global flux datasets and bulk formulae used

Flux dataset	Observations	Grid	Bulk Formula
UWM/COADS (da Silva *et al*. 1994)	*In situ* (VOS)	1° Monthly	Large and Pond (1981, 1982)
SOC (Josey *et al*. 1999)	*In situ* (VOS)	1° Monthly	Smith (1980, 1988)
OAFlux (Yu and Weller 2007)	Satellite / Reanalysis	1° Daily	COARE 3.0 (Fairall *et al*. 2003)
HOAPS (Bakan *et al*. 2000)	Satellite / Reanalysis	1° Daily	Smith (1988)
J-OFURU (Kubota *et al*. 2002)	Satellite / Reanalysis	1° Daily	Large and Pond (1982) and Kondo (1975)
GSSTF 2.0 (Chou *et al*. 2003)	Satellite / Reanalysis	1 ° Daily	Liu et al (1979) and Chou (1997; 2003)

2.3.2 Transfer Coefficients

The parameterisations listed in Table 2-1 use one of two different approaches, either estimating the neutral drag coefficient directly and using constant values for the neutral heat and moisture transfer coefficients or estimating the different roughness

lengths and using those to estimate the neutral transfer coefficients (Table 2-2 and Table 2-3). The neutral drag coefficients of Large and Pond (1981, 1982) and Smith (1980) are a linear function of wind speed derived from data obtained in moderate to high wind speed conditions. These have been extrapolated for low wind speeds in the datasets listed in Table 2-1. The Smith (1988), COARE 3.0 (Fairall *et al*. 2003) and Chou (2003) parameterisations take into account the viscous shear under low wind speeds (i.e. smooth flow over the sea surface) and the roughness of the sea surface due to gravity waves under moderate to high wind speeds. Both Smith (1988) and Chou (2003) use a simple relationship between the roughness length and the wind stress (u_*^2), gravity and a proportionality parameter, a_c, to estimate the roughness length (Charnock 1955). This relationship is expanded in the COARE 3.0 algorithm to allow for a linear increase in the roughness, via a variable Charnock parameter, with increasing wind speed. The COARE 3.0 algorithm also provides parameterisations for the roughness based on either a wave age (Oost *et al*. 2002) or a wave steepness (Taylor and Yelland 2001) model. In all three parameterisations (Smith 1988; Chou *et al*. 2003; Fairall *et al*. 2003) the roughness length under smooth flow is proportional to the ratio of the viscosity of air to the friction velocity.

Table 2-2: Drag coefficients (C_{D10n}) or momentum roughness length (z_0) used in the different bulk formulae. u_{10n} is the 10 m neutral wind speed; ν the kinematic viscosity of air; a_c the Charnock parameter; g the acceleration due to gravity; L_p the wavelength associated with the dominant wave period; h_s the significant wave height and C_p the phase speed of the dominant wave.

Formula	Drag coefficient or momentum roughness length
Smith (1980)	$10^3 C_{D10n} = (0.61 + 0.063 u_{10n})$
Large and Pond (1981)	$10^3 C_{D10n} = \begin{cases} 1.2 & 4 \le u_{10n} < 11 \quad ms^{-1} \\ 0.49 + 0.065 u_{10n} & 11 \le u_{10n} < 25 \quad ms^{-1} \end{cases}$
Smith (1988), Chou (2003)	$z_0 = \frac{a_c u_*^2}{g} + \frac{0.11\nu}{u_*}, \quad a_c = 0.011$ (Smith, 1988); $a_c = 0.0144$ (Chou et al., 2003)
COARE 3.0 (Fairall *et al*. 2003)	$z_0 = \frac{a_c u_*^2}{g} + \frac{0.11\nu}{u_*}, \quad \begin{cases} a_c = 0.011 & u_{10n} < 10 \\ a_c = 0.011 + (0.007/8)(u_{10n} - 10) & 10 \le u_{10n} < 18 \\ a_c = 0.018 & 18 \le u_{10n} \end{cases}$

Formula	Drag coefficient or momentum roughness length
COARE 3.0 (using Taylor and Yelland (2001) wave height model)	$z_{0TY} = 1200 h_s \left(\frac{h_s}{L_p} \right)^{4.5} + \frac{0.11\nu}{u_*}$
COARE 3.0 (using Oost (2002) wage age model)	$z_{0Oo} = \frac{50}{2\pi} L_p \left(\frac{u_*}{C_p} \right)^{4.5} + \frac{0.11\nu}{u_*}$

Constant values are used for the neutral heat and moisture transfer coefficients in the parameterisations of Large and Pond (1981, 1982) and Smith (1980, 1988). The COARE 3.0 algorithm (Fairall *et al*. 2003) and parameterisation used by Chou (2003) use surface renewal theory (e.g. Liu *et al*. 1979) to estimate the roughness lengths for heat and moisture.

Table 2-3: As Table 2-2 but for heat and moisture transfer coefficients or roughness lengths. R_r is the roughness Reynolds number; a_1, a_2, b_1 and b_2 are constants given by Liu et al (1979).

Formula	Transfer coefficient or roughness length
Smith (1988)	$10^3 C_{H10n} = 1.0$ $10^3 C_{E10n} = 1.2$
Large and Pond (1982)	$10^3 C_{H10n} = \begin{cases} 1.13 & \zeta < 0, \text{ unstable}, 4 < \mathrm{u_{10n}} < 25 \text{ m s}^{-1} \\ 0.66 & \zeta > 0, \text{ stable}, 6 < \mathrm{u_{10n}} < 20 \text{ m s}^{-1} \end{cases}$ $10^3 C_{E10n} = 1.15 \quad \zeta < 0, \text{ unstable}, 4 < \mathrm{u_{10n}} < 14 \text{ m s}^{-1}$
COARE 3.0 (Fairall *et al*. 2003)	$z_{0q} = z_{0t} = \min\left(1.1 \times 10^{-4}, 5.5 \times 10^{-5} R_r^{-0.6}\right)$ $R_r = \frac{u_* z_0}{\nu}$
Chou (2003)	$z_{0t} = \frac{\nu}{u_*}\left(a_1 R_r^{b_1}\right)$ $z_{0q} = \frac{\nu}{u_*}\left(a_2 R_r^{b_2}\right)$

2.3.3 Stability Correction

The dimensionless profiles, required for the stability correction, are normally parameterised in terms of a dimensionless stability parameter, ζ, defined by

$$\zeta = \frac{z}{L} \qquad \text{Eq. 2-11}$$

$$L = -\frac{u_*^3 T_v}{gkw'T_v'} \qquad \text{Eq. 2-12}$$

where L is the Monin-Obukhov length and T_v the virtual temperature of air. ζ is equal to the buoyancy production or consumption term from the turbulent kinetic energy (TKE) budget normalized by $-u_*^3/kz$ (e.g. Stull 1988). When ζ is negative there is a production of TKE due to buoyancy forcing and the atmosphere is said to be unstable. When ζ is positive there is a consumption of TKE, surface exchange is inhibited, and the surface layer said to be stable.

The relationships between the dimensionless stability parameter and the dimensionless profiles have been determined empirically based on measurements of the flux profiles under a range of stability conditions. With the exception of the COARE 3.0 algorithm, the different bulk formulae use versions of the dimensionless profiles recommended by Dyer (1974) and summarised by WGASF (2000) as

$$\phi_m = (1 - \alpha_s \zeta)^{-\beta_s} \qquad \zeta < 0 \qquad \text{Eq. 2-13 a}$$

$$\phi_m = (1 + \gamma_s \zeta) \qquad \zeta > 0 \qquad \text{Eq. 2-13 b}$$

for the wind shear. α_s, β_s and γ_s are empirically determined coefficients. These are set to $\alpha_s = 16$, $\beta_s = 1/4$, $\gamma_s = 5$ in the Smith (1980, 1988) and Large and Pond (1981) parameterisations. Following comparison with direct flux measurements made on Sable Island, Large and Pond (1982) update the parameterisation for stable conditions to $\gamma_s = 7$. Chou *et al*. (2003) use the values given by Large and Pond (1982).

The dimensionless temperature and humidity profiles have been similarly summarised by WGASF (2000) as

$$\phi_t = \phi_q = \phi_m^2 \qquad \zeta < 0 \qquad \text{Eq. 2-14 a}$$

$$\phi_t = \phi_q = \phi_m \qquad \zeta > 0 \qquad \text{Eq. 2-14 b}$$

When the profiles recommended by Dyer (1974) are used the stability corrections (ψ_x, Eq. 2-5) for unstable conditions are given by (e.g. Paulson 1970)

$$\psi_m = 2\ln\left(\frac{1+\phi_m^{-1}}{2}\right) + \ln\left(\frac{1+\phi_m^{-2}}{2}\right) - 2\tan^{-1}\phi_m^{-1} + \frac{\pi}{2} \qquad \text{Eq. 2-15 a}$$

$$\psi_t = \psi_q = 2\ln\left(\frac{1+\phi_t^{-1}}{2}\right) \qquad \text{Eq. 2-15 b}$$

Under stable conditions the stability corrections are given by

$$\psi_m = \psi_t = \psi_q = 1 - \phi_m \qquad \text{Eq. 2-15 c}$$

Under unstable conditions the COARE 3.0 algorithm uses the same dimensionless profiles recommend by Dyer (1974) but blended with a form that obeys the theoretical scaling limits in convective conditions with $\beta_s = 1/3$ (Fairall *et al*. 1996). Under stable conditions the profiles recommended by Dyer (1974) have been replaced by those of Beljaars and Holtslag (1991) to better represent the flux profiles at extreme stability (Fairall *et al*. 2003). The stability correction used in the COARE 3.0 algorithm under unstable conditions is given by (Fairall *et al*. 1996)

$$\psi_m = \frac{1}{1+\zeta^2}\psi_{mK} + \frac{\zeta^2}{1+\zeta^2}\psi_{mC} \qquad \text{Eq. 2-16a}$$

where ψ_{mK} is the stability correction given by Eq. 2-15a and ψ_{mC} the stability correction under convective conditions given by (Fairall *et al*. 1996)

$$\psi_{mC} = 1.5\ln\left(\frac{\phi_{mC}^{-2} + \phi_{mC}^{-1} + 1}{3}\right) - \sqrt{3}\tan^{-1}\left(\frac{2\phi_{mC}^{-1}+1}{\sqrt{3}}\right) + \frac{\pi}{\sqrt{3}} \qquad \text{Eq. 2-16b}$$

$$\phi_{mC} = \left(1 - \alpha\zeta\right)^{-1/3}$$

Under stable conditions, COARE 3.0 uses the stability corrections of Beljaars and Holtslag (1991)

$$\psi_m = -\left(a_{bh}\zeta + b_{bh}\left(\zeta - \frac{c_{bh}}{d_{bh}}\right)\exp\left(-d_{bh}\zeta\right) + \frac{b_{bh}c_{bh}}{d_{bh}}\right) \qquad \text{Eq. 2-16c}$$

Eq. 2-16d

where $a_{bh} = 1$, $b_{bh} = 0.667$, c_{bh}=5 and d_{bh}= 0.35.

Whilst the COARE 3.0 algorithm uses different parameterisations for the dimensionless profiles and stability corrections the resulting differences in the stability corrections are small over the range of stabilities typical of the oceans ($-1 < \zeta < 0.5$). Figure 2-1 shows a comparison of the stability corrections plotted against ζ for the range of stabilities typically found over the oceans. The close agreement at near neutral stabilities can be seen together with the divergence of the stability corrections as the magnitude of ζ increases. The extreme stabilities, where the different stability

corrections diverge, rarely occur over the oceans. The choice of stability correction is therefore expected to have little impact on the daily and monthly flux fields. Whilst the choice of stability correction may make little difference to the flux estimates there may still be large differences between flux estimates from different algorithms due to the choice of neutral transfer coefficients (Table 2-3).

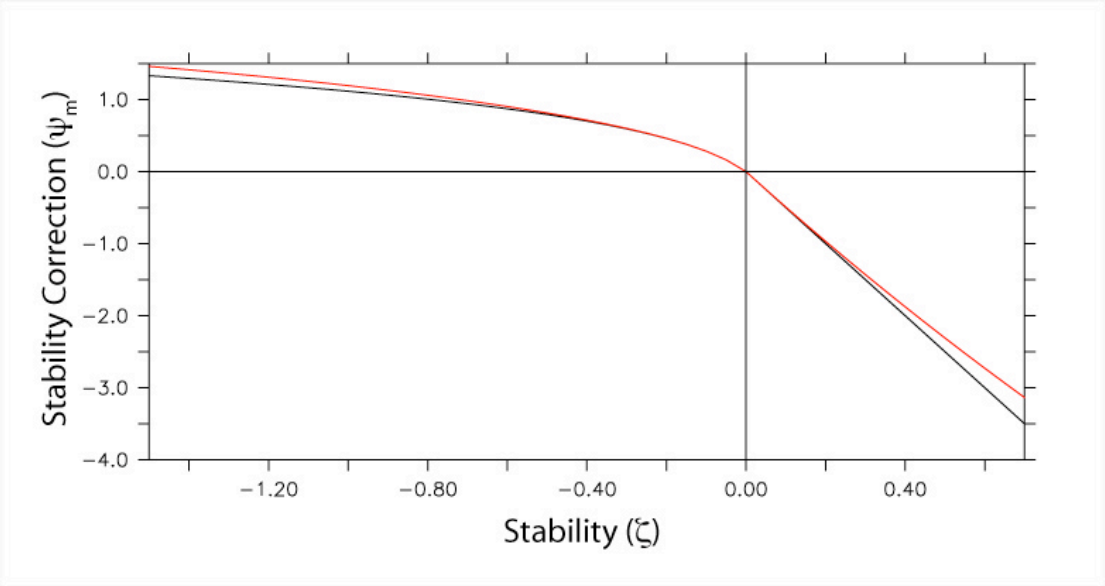

Figure 2-1: Momentum stability corrections (ψ_m) from the Smith (1980, 1988) parameterisation (black) and COARE 3.0 (Fairall *et al.* 2003) algorithm (red) as a function of stability (ζ).

2.3.4 Non turbulent fluxes

In addition to differences between the bulk formulae described above the COARE 3.0 algorithm also estimates non-turbulent components of the fluxes not included in the other parameterisations. The effects of precipitation and free convection under highly unstable conditions ($\zeta << -1.0$) are included. The free convection has been accounted for through the inclusion of a term that depends on $\zeta^{1/3}$ in the stability correction (i.e. $\beta = 1/3$ in Eq. 2-13). The underestimation of the scalar wind speed calculated from vector means is also allowed for in terms of a gustiness parameter. This is given by (Fairall *et al*. 1996)

$$w_g = \beta_{conv}\left(\frac{g}{T}\left[\frac{H}{\rho_a c_p} + 0.61T\frac{E}{\rho_a L_v}\right]z_i\right)^{1/3}$$

Where ρ_a is the density of air; c_p the specific heat capacity of air at constant pressure; L_v the latent heat of vaporisation; g the acceleration due to gravity; T the air temperature, z_i the height of the boundary layer and β_{conv} an empirical coefficient. Under stable conditions w_g is set to zero. The wind speed used in the COARE 3.0 algorithm is then given by $S_{coare} = \sqrt{u^2 + w_g^2}$. This term should not be required for scalar wind speed measurements.

Precipitation acts to cool and freshen the sea surface, transferring heat, freshwater and momentum between the atmosphere and underlying ocean. Whilst the sensible heat flux due to precipitation will be small on long time scales the flux can be significant on hourly and sub-hourly time scales. For example, during the COARE experiment mean precipitation heat fluxes of 2.5 W m^{-2} were observed but with hourly fluxes of up to 75 W m^{-2} (Fairall *et al*. 1996). Under heavy tropical rainfall conditions, instantaneous precipitation heat fluxes up to 400 W m^{-2} have been reported with sustained values of over 100 W m^{-2} lasting an hour or longer (Flament and Sawyer 1995). The COARE 3.0 algorithm includes the sensible heat flux due to precipitation using the parameterisation of Gosnell (1995), given by

$$H_{sr} = -Rc_{pw}\alpha_w\left(1 - B_0^{-1}\right)\Delta T$$

where R is the rain rate; c_{pw} the specific heat capacity of liquid water; α_w the Clausius-Clapeyron wet bulb factor; B_0 the bulk Bowen ratio; and ΔT the air – sea temperature difference.

Sea spray also moderates and contributes to the fluxes between the oceans and atmosphere under high wind speeds (e.g. Andreas and De Cosmo 2002). This contribution is implicitly included in the COARE 3.0 parameterisation and other algorithms whose development included higher wind speeds. An earlier version of the COARE algorithm has also been modified by Andreas *et al*. (2002; 2008) to explicitly model the spray contribution, partitioning the fluxes into interfacial components and spray components. A comparison of the fluxes calculated using the modified COARE 2.5b algorithm and direct fluxes measured from a fixed platform in the North Sea (Katsaros *et al*. 1994) showed better agreement than using the COARE 2.5b algorithm alone (Andreas and De Cosmo 2002). In the comparison the spray fluxes were shown to contribute up to 10 % of the total fluxes under high wind speed ($u_{10} > 15$ ms^{-1}) and high flux conditions. However, it should be noted that the same data are used by Andreas *et*

al. (2002) to both tune the modified version of the COARE2.5b algorithm and to verify their modifications. As a result, the verification presented in Andreas *et al*. (2002) is not independent.

2.4 Summary and Discussion

In summary, there are a number of different bulk parameterizations that can be used to estimate the surface turbulent fluxes over the oceans. These range from relatively simple parameterizations, such as Smith (1980), that use either fixed or wind speed dependent transfer coefficients to more complicated algorithms based on surface renewal theory (Liu *et al*. 1979) and that explicitly model the underlying physics and the effects of precipitation and free convection (e.g. Fairall *et al*. 2003).

Due to the explicit modeling of the underlying physics, the more complicated parameterizations may be expected to give better results. However, the implementation of the more complicated algorithms requires estimates of parameters such as the skin sea surface temperature or the boundary layer height. These are not routinely measured over the oceans, instead model estimates, such as those from the reanalysis models (see Chapter 3), or estimates based on the available parameters need to be used to estimate the fluxes globally. The impact of estimating the missing parameters on the fluxes and their uncertainty is uncertain and, as a result, the benefits gained from using the more complicated algorithms unclear.

In contrast to the more complicated algorithms, the simpler algorithms use bulk meteorological parameters that are widely available over the oceans from numerous sources. These parameterizations also have the advantage of having been developed using bulk observations similar to those widely available and will, on average, include the effects of gustiness, rain and the cool skin implicitly and result in unbiased flux estimates. This implicit inclusion of these effects will however increase the uncertainty in the fluxes.

Ideally, a comprehensive comparison of the different parameterizations, comparing the results to independent direct flux measurements, would be used to identify which parameterization to use. However, no recent independent comparisons have been performed. The most recent comparison (Brunke *et al*. 2003) found the COARE 3.0 algorithm to be the least problematic for estimating the surface fluxes over the oceans. However, a number of the datasets used in the comparison were also used in

the development of the COARE 3.0 algorithm, removing the independence of the comparison. Their conclusions therefore need to be treated with caution.

Due to the uncertainty over the benefits gained by using one of the more complicated algorithms two of the simpler parameterizations will be used in this study. The parameterization of Smith (1980) is used for the drag coefficient and Smith (1988) used for the heat and moisture transfer coefficients. These choices have, in part, been guided by the use of the VOS observations in this study (see Chapter 4 for choice of data) and the history of the use of these parameterizations with VOS data (e.g. Josey *et al*. 1999). It should be noted that Smith (1988) also provides a parameterization for the wind stress. However, Drennan *et al*. (2005) concluded that this parameterisation may underestimate the roughness length under strongly forced conditions ($u_{10} > 12$ ms^{-1}) and that the parameterisation of Smith (1980) is to be preferred for the drag coefficient. The implementation of the Smith (1980, 1988) will be discussed in Chapter 4.

3 Chapter 3 – Data Sources and Surface Flux Data Sets

3.1 Introduction

Estimates of the air temperature, humidity, wind speed, sea level pressure and sea surface temperature are required over the oceans to calculate the fluxes using the bulk formulae. Estimates of the cloud cover are also used in the bias correction of air temperature observations (Chapter 4). Various sources exist for these parameters, including: *in situ* measurements such as buoy observations and those made by merchant ships as part of the World Meteorological Organisation (WMO) Voluntary Observing Ships (VOS) Scheme; retrievals from satellites; and atmospheric model output. An overview of the observational sources is given in Section 3.2. A description of the model sources is given in Section 3.3. Gridding methods used in previous surface flux and marine meteorological datasets are reviewed in Section 3.4 and the datasets themselves reviewed in Section 3.5. Uncertainty estimates in the observations and fluxes are reviewed in Section 3.6. A discussion and summary follow in Section 3.7.

3.2 Observational Data Sources

3.2.1 Air Temperature

Air temperature observations are made on board merchant ships as part of the VOS Scheme (e.g. Worley *et al*. 2005). These observations are typically made every 3 or 6 hours and tend to be clustered over the major shipping lanes. As an example, Figure 3-1 shows the average number of VOS air temperature observations per month for sample years over the past 4 decades over the Atlantic on a 1° grid (see Section 4.3 for the data used). The clustering of the observations over the shipping lanes can clearly be seen with over 100 observations per month in the coastal regions of Europe, North America and West Africa. The mid-ocean shipping lanes typically have more than 10 observations per month whilst the regions outside the shipping lanes typically have less than 5 observations per month. A decline in the number of VOS and VOS observations during the 1990's can be seen in the bottom two panels.

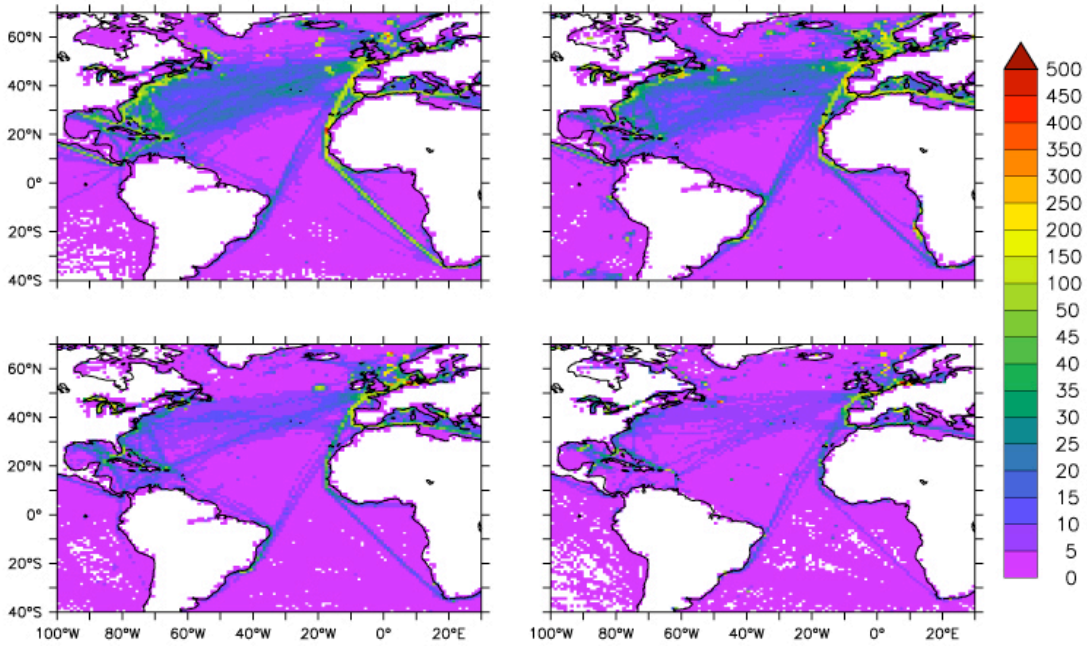

Figure 3-1: Monthly mean number of VOS observations of air temperature from ICOADS (e.g. Worley *et al.* 2005) during 1975 (top left), 1985 (top right), 1995 (bottom left) and 2005 (bottom right).

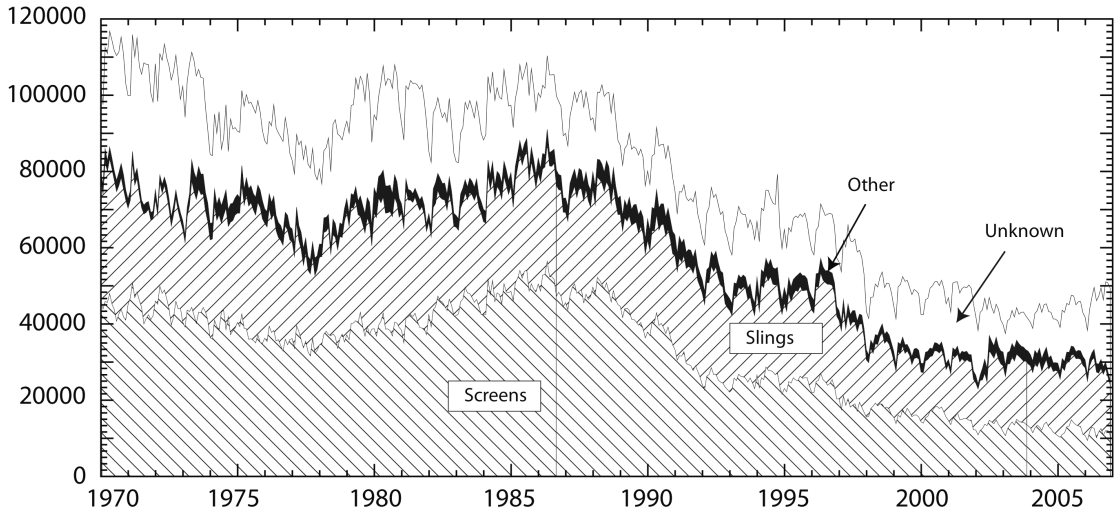

Figure 3-2: Stacked area plot showing number of VOS air temperature observations per month between 40S – 70N and 100W – 30E for each of the main air temperature exposure observing methods.

The air temperature observations from the VOS are typically made using mercury thermometers exposed in either marine screens or sling psychrometers (e.g. Kent *et al*. 2007). A small number of observations have also been made using unscreened thermometers or thermometers housed in aspirated Assman type

psychrometers. Figure 3-2 shows the number of air temperature observations made by the VOS in the Atlantic and the method of thermometer exposure. Slightly more observations are made using marine screens at the start of the period shown, however, by the end roughly equal numbers are made using sling psychrometers (slings) and marine screens (screens).

The VOS observations contain random errors of around 1 °C (Kent *et al*. 1999; Kent and Berry 2005) and can contain biases of up to 2 °C due to the solar heating of the instruments and ships superstructure (e.g. Glahn 1933; Folland 1971; Goerrs and Duchon 1980; Blanc 1986; Kent *et al*. 1993a). A bias correction was proposed to remove the warm bias due to the solar heating (Kent *et al*. 1993a) and used in the SOC climatology (Josey *et al*. 1999). The bias correction was given by

$$\Delta T_{cor} = 2.3 \times 10^{-3} R_{SW} - 1.646 \times 10^{-5} R_{SW} V_{rel} \qquad \text{Eq. 3-1}$$

where ΔT_{cor} is the bias correction (°C), R_{SW} the incident solar radiation (W m^{-2}) and V_{rel} the relative wind speed (m s^{-1}). Whilst this bias correction removes the mean effect of the solar heating, the bias is overestimated in the mornings and underestimated in the afternoons due to the neglect of the heat storage by the ships superstructure. A new correction has therefore been developed as part of this research taking into account the heat storage terms and is described in Chapter 4. Other authors have used night-time only observations to avoid this heating bias (e.g. Rayner *et al*. 2003). However, the number of observations is approximately halved, increasing the uncertainty due to random and sampling errors.

In addition to biases due to solar heating, spurious trends exist in the surface marine air temperature record due to the changing size of the VOS and the height above the sea surface at which the observations are made. The observing height has increased from around 6 m in the late 1800's (Parker *et al*. 1995) to 16 m by 1973 and to over 24 m by 2006 (Kent *et al*. 2007). Figure 3-3 shows how the average measurement height over the Atlantic has changed since 1970. If the changing height is not accounted for a downward change of 0.25°C would be introduced over the last 100 years (Rayner *et al*. 2003). Over the past 30 years the change would be ~ 0.1 °C (or a trend of ~0.03 °C decade^{-1}). The spurious component of the trend due to the changing platform heights can be removed using the bulk formulae (see Chapter 2) to reduce the observations to a standard reference height (usually 10 m). This has been done in both the first SOC

climatology (Josey *et al*. 1999) and in the HadNMAT (e.g. Rayner *et al*. 2003) series of datasets.

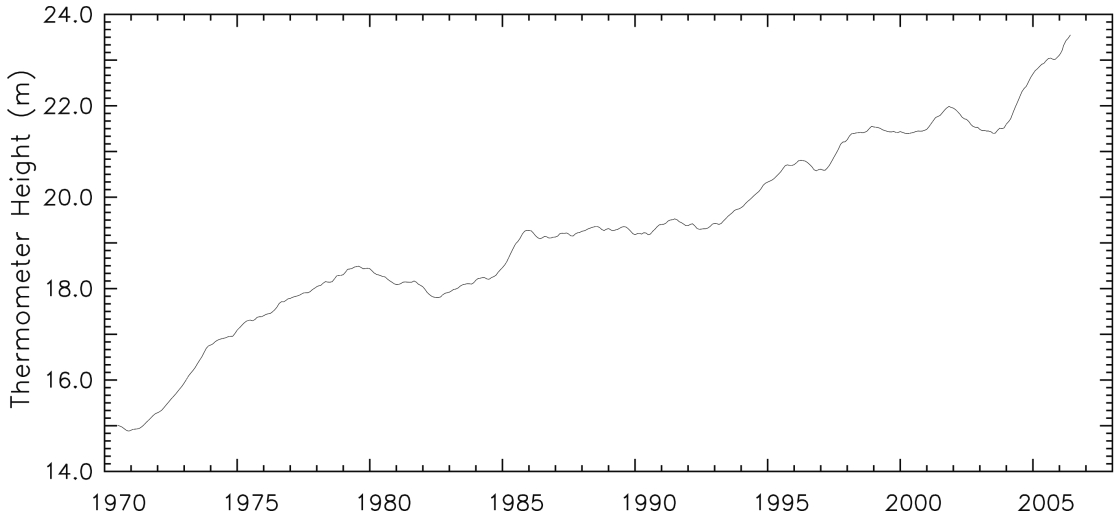

Figure 3-3: Mean air temperature observation height over the region 40S – 70N and 100W – 30E from ICOADS and Pub. 47 (see Section 4.3). A 12 month running mean filter has been applied.

In addition to the VOS observations, measurements of the air temperature are made on moored buoys and research vessels. The moored buoy network comprises of both operational and research moorings and has expanded from a few buoys in the Gulf of Mexico during the 1970's to include the majority of the North American and European coastlines by the mid 1990's (Figure 3-4). Research moorings and smaller arrays have typically been deployed for periods ranging from several months to several years (e.g. Moyer and Weller 1997). The larger arrays, such as the Pilot Research Moored Array in the Tropical Atlantic (PIRATA) (Servain *et al*. 1998) and Tropical Atmosphere Ocean (TAO) (e.g. McPhaden *et al*. 1998) arrays are part of the Global Ocean Observing System and are planned as sustained deployments. The earliest observations from these arrays are from 1984. Contributions from the TAO and PIRATA arrays can be seen in the tropical Pacific and Atlantic respectively in the bottom two panels of Figure 3-4. The observations from the moored buoys have similar accuracy to the VOS observations and un-aspirated instruments similarly contain biases due to heating of the instruments by solar radiation (Anderson and Baumgartner 1998). Research vessel air temperature observations tend to be of higher quality with fewer biases and random errors and a higher reporting frequency (e.g. Smith *et al*. 2001).

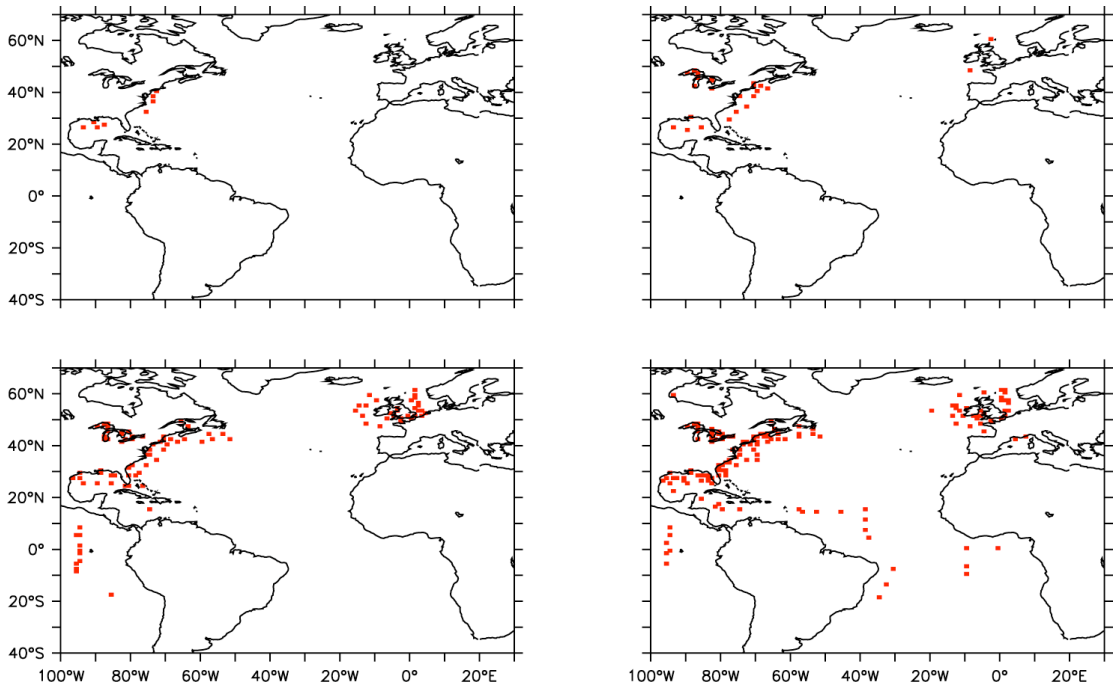

Figure 3-4: Grid cells with air temperature observations at least once every 6 hours from moored buoys during: 1975 (top left); 1985 (top right); 1995 (bottom left); and 2005 (bottom right).

3.2.2 Humidity

Humidity observations have been made routinely by the VOS since the early 1950's and are usually made using wet and dry bulb thermometers housed in marine screens or sling psychrometers (Kent *et al*. 2007). Figure 3-5 shows the proportion of VOS observations of humidity made using each method over the North Atlantic for the period 1970 – 2006. Observations with a known method are divided approximately equally between screens and sling psychrometers. A small number of observations are made using either unscreened thermometers or Assman type aspirated psychrometers. In addition to the paired wet and dry bulb thermometers, a small number of VOS use hair hygrometers and electric sensors (Kent *et al*. 2007).

For the paired wet and dry bulb thermometer readings the vapour pressure is calculated using the psychrometric formula (e.g. List 1951)

$$e = \left(e_s' - a_p(1 + 0.00115T_{wet})p_p\left(T_{dry} - T_{wet}\right)\right) \qquad \text{Eq. 3-2}$$

where e_s' is the saturation vapour pressure at the wet bulb temperature (mb); a_p the psychrometric coefficient; p_p the pressure (mb); T_{wet} the wet bulb temperature (°C) and

T_{dry} the dry bulb temperature (°C). The value of the psychrometric coefficient depends on the ventilation of the wet bulb thermometers with 0.66 x 10^{-3} or 0.791 x 10^{-3} typically used for whirling psychrometers and marine screens respectively (e.g. UK Meteorological Office 1964b, 1964a). The saturation vapour pressure (hPa) at a given temperature is given by (Henderson-Sellers 1984)

$$e_s = 2.1718 \times 10^8 \exp\left(\frac{-4157}{T - 33.91 - 0.16}\right) \qquad \text{Eq. 3-3}$$

where T is the temperature (K) at which the saturation vapour pressure is required. The factor 0.16 has been included in the denominator as Henderson-Sellers (1984) assume absolute zero to be 273.0 K (Taylor 1995) and by including this factor better agreement is found with the values given in the Smithsonian Tables (List 1951). If only the dew point temperature is reported by the VOS the vapour pressure can be calculated using Eq. 3-3 and the dew point temperature. The specific humidity can be calculated from the vapour pressure using (List 1951)

$$q = 1000\left(\frac{0.622e}{p_p - 0.378e}\right) \qquad \text{Eq. 3-4}$$

where q is the specific humidity (g Kg^{-1}); p_p the pressure (mb); and e the vapour pressure (mb).

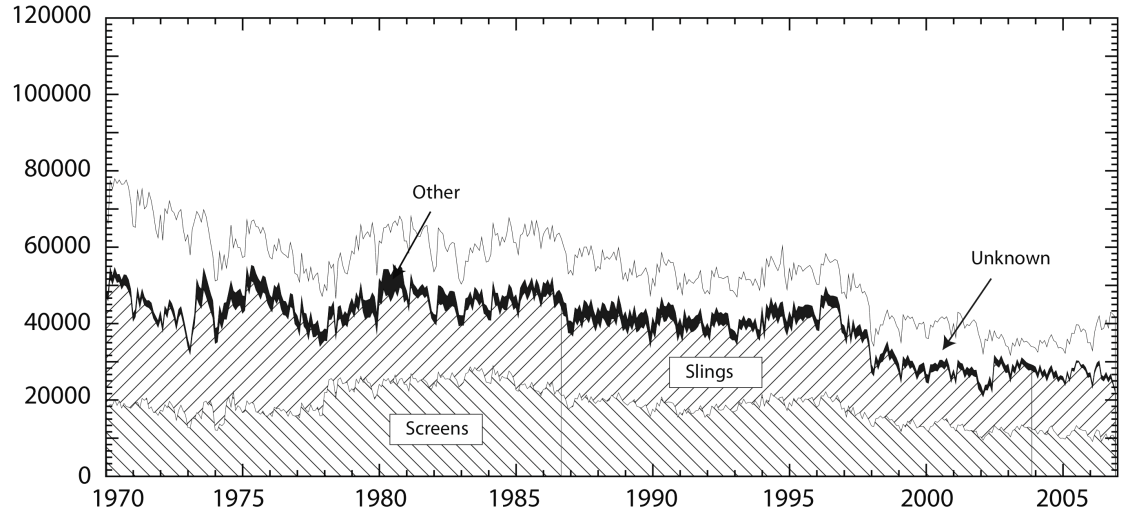

Figure 3-5: As Figure 3-2 but for humidity sensor exposure.

As with the VOS air temperature measurements, measurements of the humidity are noisy and can contain systematic biases. Random errors have been estimated for both the dew point temperature measurements and specific humidity estimates, with

errors of around 1 °C and 1 g Kg^{-1} reported respectively (Gleckler and Weare 1997; Kent *et al*. 1999; Kent and Berry 2005). The observation height has also increased over the past 50 years and the observations need to be height adjusted to avoid spurious trends being introduced. This can be done using the bulk formulae (see Chapter 2, e.g. Josey *et al*. 1999). Despite the required data being available this has not been done in recent studies (e.g. Dai 2006; Willett *et al*. 2008).

Biases have been reported in the humidity observations made using marine screens due to the inadequate ventilation of the wet bulb thermometers (Kent *et al*. 1993b). This inadequate ventilation leads to an overestimate of the vapour pressure and causes the dew point temperature, relative and specific humidity estimates from marine screens to be biased high. This bias has been examined by previous authors (Kent *et al*. 1993b) and a bias correction for the dew point temperatures proposed. The correction of Kent *et al*. (1993b) is given by

$$T_{dew}^{cor} = 1.032T_{dew} - 0.839 \qquad \text{Eq. 3-5}$$

where T_{dew}^{cor} and T_{dew} are the corrected and uncorrected dew point temperatures respectively in °C. A modified version of this correction has been used in the SOC climatology (Josey *et al*. 1999)

$$T_{dew}^{cor} = 1.029T_{dew} - 1.080 \qquad \text{Eq. 3-6}$$

Again, with T_{dew}^{cor} and T_{dew}, the corrected and uncorrected dew point temperatures (°C). The reason for the change in Josey *et al*. (1999) is not given.

Both corrections to the dew point temperature give corrections to the specific humidity that peak at low to moderate values before decreasing at higher specific humidity values (see Figure 3-6). These bias corrections are examined in further detail in Section 4 and a new bias correction proposed.

In addition to biases due to the poor ventilation of the wet bulb thermometers in marine screens, a moist bias has been suggested for observations made before 1982 (Willett *et al*. 2008). These authors found a shift in relative and specific humidity estimates over the oceans around 1982 without similar shifts in air temperature or sea surface temperature measurements. This shift was then tentatively attributed to a change in reporting practices around this time period with a significantly larger proportion of dew point temperature measurements made in whole degrees before 1982 compared to the period after. In this thesis similar changes are presented in both air temperature and

sea surface temperature estimates, suggesting the changes may be climatic rather than due to reporting practices. This is discussed in Chapter 7.

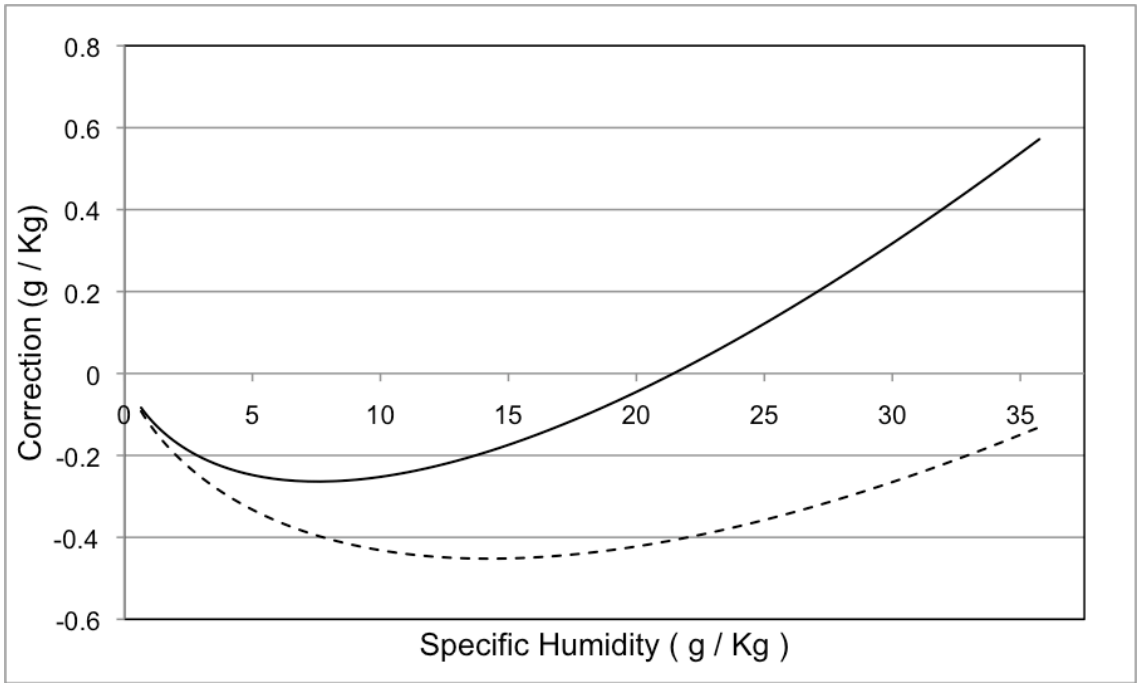

Figure 3-6: Impact of bias corrections of Kent *et al.* (1993b) (solid) and Josey *et al.* (1999) (dashed) on specific humidity as a function of specific humidity.

Humidity estimates are also available from research moorings, vessels and the tropical buoy arrays. The buoys usually report relative humidity whilst research vessels report either the relative humidity or paired wet and dry bulb temperatures. The accuracy of the relative humidity observations are in the range 2 – 4 % (e.g. Hosom *et al*. 1995; Moyer and Weller 1997; Servain *et al*. 1998). For a 2 % relative humidity accuracy, the specific humidity accuracy ranges from 0.05 g Kg^{-1} at an air temperature of -5 °C to 0.7 g Kg^{-1} at 35 °C (Table 3-1). Whilst the research vessel observations are more accurate they are limited in space and time. This reduces their usefulness for constructing global datasets.

Estimates of the surface specific humidity are available from satellites measurements such as those made by the Special Sensor Microwave / Imager (SSM/I) sensors on board the Defense Meteorological Satellite Program (DMSP) series of satellites (e.g. Wentz 1997). RMS errors of order 1 – 2 g Kg^{-1} are usually quoted for the parameterisations used to estimate the surface (10 m) specific humidity from the SSM/I sensors (e.g. Chou *et al*. 1995; Schulz *et al*. 1997). However, this level of accuracy is only achievable under ideal conditions, and seasonally and regionally varying biases

have been reported when the satellite measurements are compared to *in situ* observations (Chou *et al*. 1995). Additionally, the estimates are only available in rain free conditions (e.g. Wentz 1997), leading to a potential fair weather bias.

Table 3-1: Specific humidity accuracy as a function of temperature for a relative humidity of 80 % and an accuracy of 2 % (i.e. RH = 80 ± 2 %). A pressure of 1013 hPa has been used.

Air Temperature (°C)	Accuracy (g Kg^{-1})
-5.00	0.05
0.00	0.08
5.00	0.11
10.00	0.15
15.00	0.21
20.00	0.29
25.00	0.40
30.00	0.53
35.00	0.71

3.2.3 Sea Surface Temperature

Sea surface temperature measurements made on board the VOS over the past 40 years have typically been made using insulated buckets to collect water samples or by measuring the temperature of the water in the engine room intake (ERI). A small but increasing number of measurements are also made using hull contact sensors (e.g. Kent *et al*. 2007). Figure 3-7 shows the proportion of observations made each month using the different methods over the Atlantic. Initially, a higher proportion of observations are made using bucket measurements. However, this proportion declines as more ships switch over to using the engine room intake or hull contact sensors.

The VOS SST observations typically contain random errors of order 1°C (e.g. Gleckler and Weare 1997; Kent and Challenor 2006) and biases have been reported in both the bucket and the ERI measurements. When the ocean is warmer than the overlying atmosphere the buckets can be biased cold due to the heat loss from the water sample between collection and measurement of its temperature (e.g. Kent and Kaplan 2006; Kent and Taylor 2006). Engine room intake measurements can be biased warm if the intake temperature is measured close to the engine (e.g. Kent and Kaplan 2006).

Corrections have been proposed for both the insulated bucket measurements (Kent and Kaplan 2006) and engine intake measurements (Josey *et al*. 1999; Kent and

Kaplan 2006). Corrections have also been proposed for measurements prior to World War 2 made using un-insulated buckets (Folland and Parker 1995). The ERI measurements used in the SOC climatology were reduced by 0.35 °C to correct for the warm bias where information on the observing practices existed. Where this information was missing a default correction of 0.2 °C was applied by Josey *et al* (1999) based on the proportion of ships with a known method and making ERI measurements (60%). Kent and Kaplan (2006) made new estimates of the bias for both measurement methods for night time observations. Bucket measurements were found to be biased cold by just over 10 % of the air – sea temperature difference. The ERI measurements were found, on average, to be biased warm by up to 0.2 °C in the 1970's and 1980's, decreasing to a cold bias of 0.1°C relative to the buckets in the 1990's. This change in the relative bias in ERI measurements is likely to be due to an improvement in the location of the thermometers used and improved pipe insulation. Whilst different corrections have been suggested there are still large uncertainties in the magnitude of these biases and even the sign (Elizabeth Kent, pers. comm., also Berry and Kent 2010).

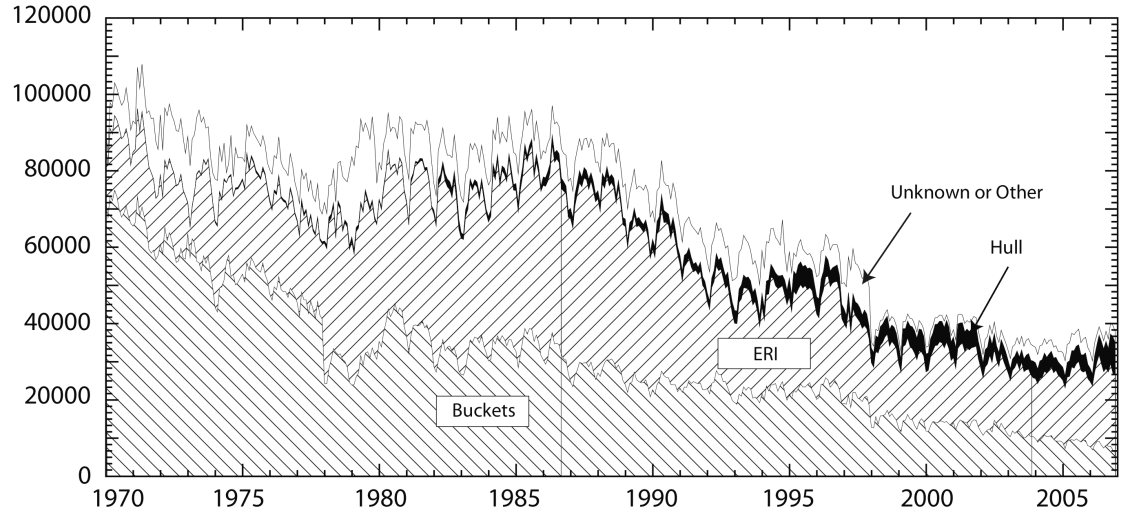

Figure 3-7: As Figure 3-2 but for the SST measurement method

Sea surface temperature measurements are also available from moored and drifting buoys, research vessels and satellites. As with the other variables, observations made as part of research cruises are typically of higher accuracy and precision than the operational sources and provide data on hourly or sub-hourly timescales (e.g. Smith *et al*. 2001). Similarly, the observations made as part of research moorings are more accurate than the operational sources and report on sub-hourly timescales. However the measurements from both the research and operational moorings, together with research

vessel data, are limited in spatial distribution. The research data are also of limited duration.

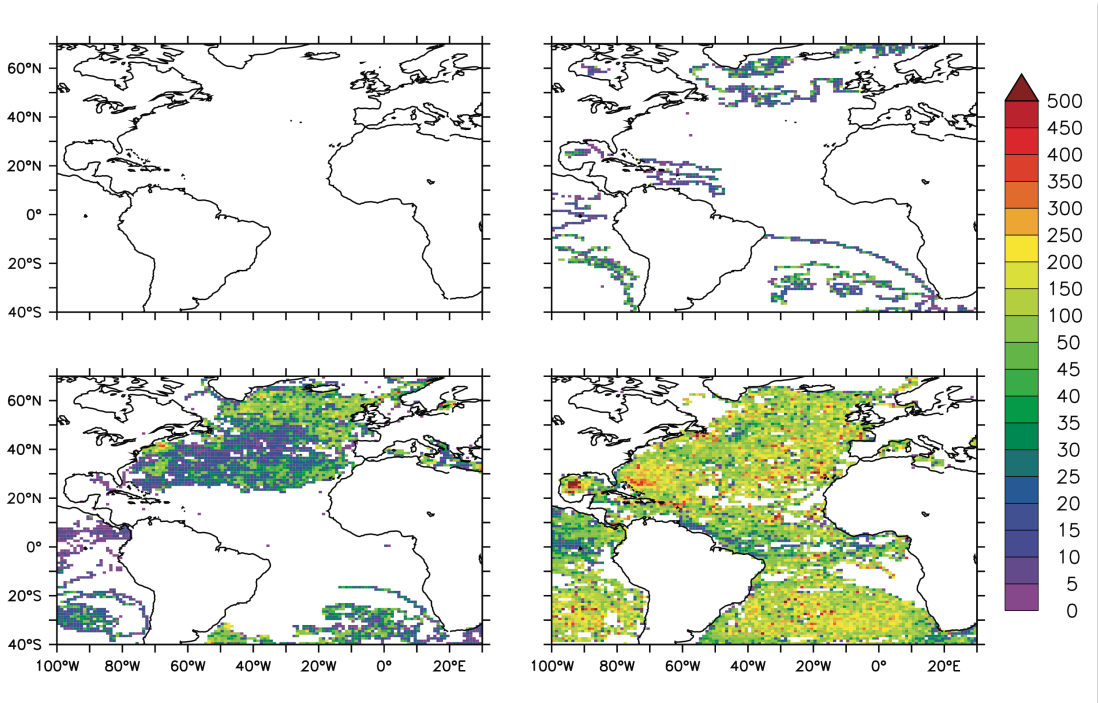

Figure 3-8: As Figure 3-1 but for SST observations flagged as coming from drifting buoys during 1975 (top left), 1985 (top right), 1995 (bottom left) and 2005 (bottom right).

The drifting buoy observations first appear in the surface record in the late 1970's and were originally limited in their spatial distribution. However, over the last 10 – 15 years there has been a large increase in the number of drifting buoys deployed. As an example, Figure 3-8 shows the average number of drifting buoy SST observations per month from ICOADS (see Section 4.3) during the last 4 decades. The expanding drifting buoy network and increasing number of observations can be clearly seen. A comparison of drifting buoy observations made within an hour of each other and no more than 50 km apart (Emery *et al*. 2001) has shown the uncertainty in the drifting buoy measurements to be ~ 0.15°C when the spatial variability is taken into account. Similar uncertainty estimates have been made for the drifting buoy observations through a three way error analysis (O'Carroll *et al*. 2008). In this comparison the uncertainty in the buoy observations was estimated to be 0.2°C. Drifting buoy observations have also been compared to nearby VOS observations (Emery *et al*. 2001) and a mean difference (drifting buoy – VOS) of -0.2 °C reported. This has been attributed to the VOS ERI measurements being biased high.

The satellite estimates of the sea surface temperature are based on infrared and microwave measurements using instruments such as the Advanced Very High Resolution Radiometer (AVHRR) (e.g. Scambos *et al*. 2000 (updated 2002)) and Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E) (e.g. O'Carroll *et al*. 2008) respectively. The recent satellite measurements can provide spatial resolutions as high as 1.25 km with twice daily measurements. The infrared radiometers give the higher resolutions and tend to be more accurate than the microwave measurements. However, the estimates can be prevented or contaminated by the presence of clouds and can be adversely affected by aerosols (e.g. Reynolds 1993; Merchant *et al*. 2008). In contrast, the microwave sensors can provide SST estimates under cloudy, rain free conditions. In a recent comparison, RMS errors of 0.16 K, 0.42 K and 0.23 K have been reported for the Advanced Along Track Scanning Radiometer (AATSR), AMSR-E and drifting buoys respectively in a 3-way comparison (O'Carroll *et al*. 2008). Infrared and microwave radiometers have routinely been flown on satellites since the late 1970's and 1980's (e.g. WGASF 2000) with the infrared sensors providing the longer time series. Whilst multi-decadal time series are now available from the satellite measurements, the different instruments can have different biases and need to be homogenized to create a climate quality dataset (e.g. Merchant *et al*. 2008).

3.2.4 Wind speed

Observations of the wind speed and direction over the oceans by the VOS are made by either visually estimating the wind speed and direction or by taking anemometer measurements (e.g. Kent *et al*. 2007), Figure 3-9. The visual wind speed estimates are based on the sea state and reported using the WMO1100 Beaufort Equivalent Scale (e.g. Thomas *et al*. 2008). The anemometers are usually either cup and vane or combined propeller vane anemometers. As with other VOS observations the wind speed estimates tend to be noisy with random errors of around 2 m s^{-1} (e.g. Kent and Berry 2005). The observations in the ambiguous category currently have an unknown method due to choices made when the data were originally digitised and may be recoverable from the original logbooks.

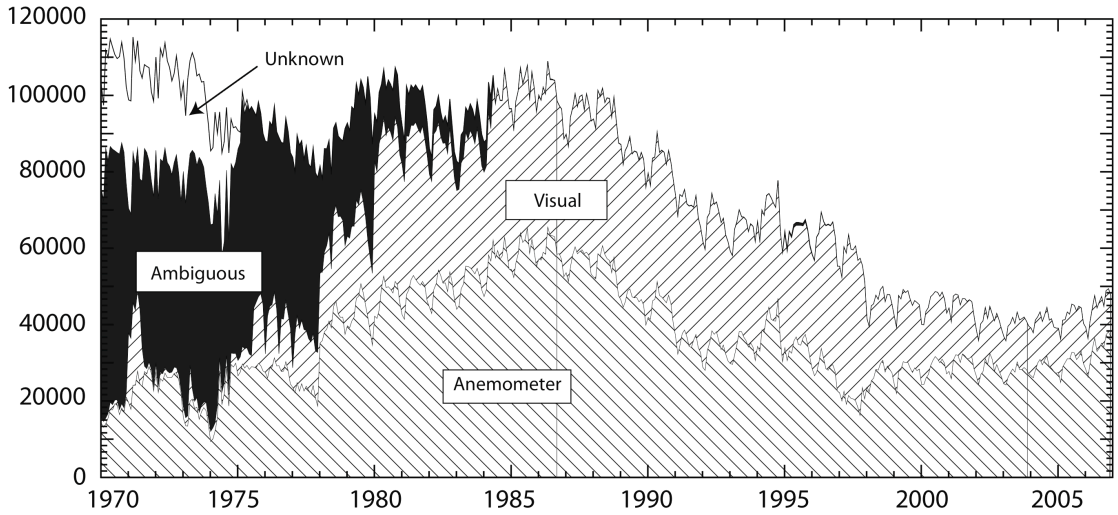

Figure 3-9: As Figure 3-2 but for the wind speed measurement method.

The WMO1100 scale contains non-linear biases due to the statistics and small sample size used to develop the scale with an under-estimation of low wind speeds and an overestimation of high wind speeds (Isemer and Hasse 1991). These biases result in an underestimate of the climatological values and various corrections and alternative scales have been suggested (e.g. da Silva *et al*. 1994; Lindau 1995; Thomas *et al*. 2005). Based on comparisons of visual and anemometer wind speed estimates, Kent and Taylor (1997) found the scale of Lindau (1995) to be the most consistent with height adjusted monthly mean anemometer winds. A polynomial has been fitted through the mid-points of Lindau's scale by Thomas *et al*. (2005) with the corrected wind speed given by

$$U_{EL} = 0.0161 + 1.1888U_E - 0.0221U_E^2 + 0.0004U_E^2 \qquad \text{Eq. 3-7}$$

where U_{EL} is the Lindau adjusted wind speed and U_E the visually estimated wind speed.

The VOS anemometer measurements contain biases due to flow distortion by the ships superstructure (e.g. Moat *et al*. 2006a, 2006b) and a spurious upwards trend due to the changing measurement height (e.g. Thomas *et al*. 2008). The biases due to the flow distortion can be of either sign depending on the location of the anemometers and the relative wind direction (Moat *et al*. 2006b). Currently there is not enough information on the location of the anemometers on board the VOS to estimate the effects of flow distortion routinely. The spurious trend due to the increasing vessel size and measurement height (Figure 3-10) can be accounted for by adjusting the wind speed observations to a standard reference height. This can be done either using the bulk formulae (Chapter 2) with the atmospheric stability taken into account or using a logarithmic wind speed profile (e.g. Thomas *et al*. 2005).

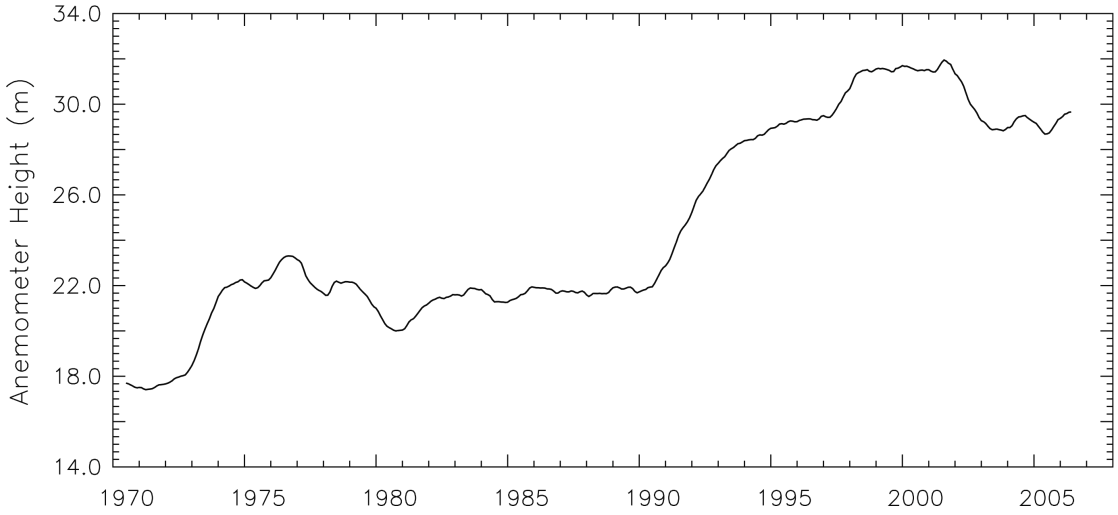

Figure 3-10: As Figure 3-3 but for anemometer height.

Wind speed measurements are also made on board research vessels and moored buoys. The measurements from research vessels contain smaller random errors than the VOS. However, they need to be height corrected and can also contain biases due to flow distortion. Similar to the VOS observations, the height corrections can be made using either the bulk formulae or a logarithmic wind speed profile. Additionally, the impact of flow distortion has been estimated for a number of research vessels allowing the effects of flow distortion to be taken into account (Yelland *et al*. 2002). The buoy observations have smaller random errors than the VOS observations, but also need to be height corrected and suffer from different biases. These include those due to wave induced motions and possible sheltering effects (e.g. Thomas *et al*. 2005) and instrument degradation on long deployments.

Wind speed estimates are also available from passive (radiometer) and active (scatterometer) microwave sensors on board satellites. The earliest sensors (both passive and active) were flown on board Seasat in 1978 (e.g. Wentz *et al*. 1986). Routine measurements have been available from the late 1980's and early 1990s for radiometers and scatterometers respectively. Current scatterometer measurements have accuracies of around 1 ms^{-1}, nominal spatial resolutions of 25 km and, at best, daily coverage of ~ 92% of the global ice-free oceans (e.g. Ebuchi *et al*. 2002). The microwave radiometers have a similar accuracy on a 50 km resolution (Wentz 1997). The earlier measurements will have lower accuracies and resolutions.

3.2.5 Sea level pressure

Measurements of the sea level pressure are only available from *in situ* sources such as moored and drifting buoys, VOS and research vessels. The VOS usually use aneroid barometers to measure the pressure but with an increasing number of digital aneroid barometer measurements (e.g. Kent *et al*. 2007). A small number in the early 1970's also use mercury barometers (Kent *et al*. 2007). The random errors in the VOS observations have been estimated at around 2.2 hPa (Kent and Berry 2005). Whilst little information is available in the published literature on the accuracy and precision of the buoy and research vessel observations these are expected to be at least as accurate as the VOS observations.

3.2.6 Cloud Cover

Whilst cloud cover is not required to estimate the turbulent fluxes (see Chapter 2) cloud cover is used in estimating the incident solar radiation for the air temperature correction (see Chapter 4). Estimates of the cloud cover are available from the VOS observations with the total fractional cloud cover reported together with the amount and types of low, middle and high clouds. Random errors of 2/8 have previously been reported in the VOS cloud cover observations (Gleckler and Weare 1997) based on the monthly variability of the cloud cover.

3.3 Model Data Sources

Estimates of the basic variables also exist based on the output of numerical weather prediction (NWP) models. Operational NWP models are typically updated every few months to take advantage of improved model physics and parameterisations and new data sources to improve the forecasts. As a result, the data from operational models contain inhomogeneities and are of limited use for climate studies (e.g. Kalnay *et al*. 1996). This has resulted in a number of projects using historical observations and atmospheric models with fixed model physics based on state of the art NWP models, to reanalyse the state of the global atmosphere over multiple decades. Examples include: the National Centers for Environmental Prediction (NCEP) / National Center for Atmospheric Research (NCAR) 40 year reanalysis project (NCEP1) (Kalnay *et al*. 1996); the NCEP – Department of Energy (DOE) Atmospheric Model Intercomparison

Project (AMIP-II) reanalysis 2 (R2, hereafter NCEP2) (Kanamitsu *et al*. 2002); the European Centre for Medium-Range Weather Forecasts (ECMWF) 40 year reanalysis (ERA40) (Uppala *et al*. 2005); and the Japanese 25 year Reanalysis Project (JRA-25) (Onogi *et al*. 2007). Using frozen models avoids the inhomogeneities due to changing models seen in the operational model output. The output usually includes all the basic variables (temperature, wind speed and humidity) at different pressure levels, including at a nominal 10 m above the surface for wind speed and 2 m for temperature and humidity. Estimates of the surface fluxes and surface temperature (SST over the oceans, soil temperature over the land) are also given in the output of the models.

Whilst using frozen NWP models avoids the inhomogeneities due to changing model physics and parameterisations, they are constrained by the available observations. As a result, and due to the changing observing system, inhomogeneities are introduced (Stendel *et al*. 2000; Kistler *et al*. 2001; Simmons *et al*. 2004; Sterl 2004). Good examples of such inhomogeneities are given in both Sterl (2004) for the oceans and Simmons *et al*. (2004) for the land. In the first study, a comparison of pressure fields from NCEP1 and ERA40 is given together with an examination of the correlation between the SST tendency and latent heat flux in ERA40. In both cases, the data coverage was found to be too low to constrain the models in the Southern Hemisphere before 1980, with large disagreements between the pressure fields. The correlation between the latent heat flux and SST tendency in the Southern Hemisphere was also reported to be too low. After 1980, an improved agreement between the pressure fields and a more realistic correlation between the latent heat flux and SST tendency were found due to the increased data coverage from satellite and drifting buoy observations constraining the model estimates.

In the second study, Simmons *et al*. (2004) examine anomalies and trends in the surface temperature over land, finding disagreements between observations and the estimates from ERA40 before the 1970s. After an improvement in the observing network and an increase in data volumes and quality in the 1970s, they found better agreement between the observations and model output. The improvement found is likely to be due to the model being increasingly constrained by the data. Similar results are expected over the oceans.

Due to these problems and a lack of quality indicators, such as uncertainty estimates, the mean meteorological parameters from reanalyses need to be used with

caution in climate studies and in estimating the fluxes over long timescales. If the inhomogeneities due to the changing observing system are not taken into account, the changes observed in the models may be either incorrectly attributed to, or too much or too little weight given to, changing climate and weather patterns.

3.4 Previous Flux Datasets Construction Methods

3.4.1 Arithmetic Mean

The arithmetic mean has been used in the majority of flux datasets based on the *in situ* VOS observations and on the monthly satellite products. The arithmetic mean has the advantage of easily understood statistics. However, this method doesn't take into account the spatial (or temporal) distribution of the observations or the quality of the data. As a result the uncertainties in the mean values will be underestimated and mean values may be biased in regions with a strong gradient or when large grid cells are used. This is due to the inhomogeneous sampling of the VOS and the clustering of the observations along the major shipping routes. When smaller grid cells are used with the VOS observations the fields are incomplete due to the sparseness of the data in some regions. A number of authors (da Silva *et al*. 1994; Josey *et al*. 1999) have generated globally complete fields for the ice-free oceans using the successive correction method (see section 3.4.3) to fill and smooth the fields based on the arithmetic mean.

3.4.2 Sampling and Classical Fluxes

Two different averaging strategies have been used to estimate the fluxes from the VOS observations. The first, known as the classical method, calculates monthly mean values for the basic variables and then uses these to estimate the fluxes. The second, known as the sampling method, calculates the fluxes for the individual VOS observations and then averages these fluxes to give monthly mean values. Both methods have their advantages. The classical method is computationally efficient, requiring the iterative flux calculation to be performed only once for each month and grid cell, an important consideration for the developers of early flux datasets. However, biases are introduced in the flux fields due to the loss of the synoptic scale correlation between the basic variables (e.g. Josey *et al*. 1995). In contrast, the sampling method maintains the

correlations between the variables but at a cost of computational expense and loss of data where observations are not complete.

3.4.3 Successive Correction

Fields based on the VOS observations tend to be noisy and contain gaps due to the inhomogeneous sampling by the VOS. As a results, the fields in the UWM/COADS (da Silva *et al*. 1994) and SOC (Josey *et al*. 1999) flux climatologies have been filled and smoothed using successive correction. This method is described in full in da Silva *et al*. (1994) and a summary is given in this section.

In successive correction the observations are first gridded onto a regular grid (raw data) and an initial first guess (first guess) generated (see below). Deviations between the raw data and the first guess are then calculated, i.e.

$$D_{ij}^{k} = R_{ij} - F_{ij}^{k}$$

where D is the data deviation, R the raw data and F the first guess. The subscripts i and j refer to the grid box and the superscript k refers to the iteration. For each grid box (i, j) there are n grid boxes (i_s, j_s), and associated data deviations, within a set radius of influence and these data deviations are averaged to give an analysis increment using a weighted sum, i.e.

$$C_{ij}^{k} = \sum_{s=1}^{n} W_s D_{i_s, j_s}^{k} \Bigg/ \sum_{s=1}^{n} W_s$$

where C is the analysis increment and W_s a weighting function. Both da Silva *et al*. (1994) and Josey *et al*. (1999) use a Gaussian weighting function given by

$$W_s = \exp\left(\frac{-4 r_{sc}^{\ 2}}{R^k}\right)$$

where r_{sc} is the distance between the analysis grid box (i, j) and the s^{th} grid box (i_s, j_s) and R the radius of influence. The first guess is then updated by adding on the analysis increment, i.e.

$$F_{ij}^{k+1} = F_{ij}^{k} + C_{ij}^{k}$$

This new first guess is then smoothed to remove noise and the process repeated using a smaller radius of influence.

Similar implementations of successive correction are used in the UWM/COADS and SOC climatologies, with the same smoothing functions (described in da Silva *et al*. 1994) used and 4 iterations of the successive correction performed. In the UWM/COADS climatology influence radii of 1541, 1221, 881 and 771 km are used, the successive correction performed on monthly mean anomalies and the initial first guess set to zero. The SOC climatology used slightly different radii of 1541, 1221, 771 and 331 km, with the lower limits changed to better define frontal features such as those found in the Gulf Stream (Josey *et al*. 1998). Additionally, the successive correction in the SOC climatology was performed on monthly mean fields (rather than anomalies) and the initial first guess set to the zonal average.

Whilst the successive correction method produces globally complete fields problems have been reported with the method (Kent *et al*. 2000; Sterl 2001). In the first of these two studies, Kent *et al*. (2000) found successive correction to adversely affect the variability of the gridded fields in the SOC climatology. In low variability but well sampled regions, successive correction was found to artificially increase the variability due to the inclusion of noise from adjacent, poorly sampled regions. In high variability regions, the variability was found to be smoothed and artificially reduced. The true resolution of the dataset, i.e. the resolution at which grid cells are effectively independent of each other, was also examined in this study and the true resolution found to be typically around 3° rather than the 1° grid used to grid the data. In data sparse regions, the resolution was found to be significantly lower.

In the second study, Sterl (2001) compared the impact of two different gap filling algorithms, successive correction and reduced space optimal interpolation (RSOI, e.g. Kaplan *et al*. 1998), on the variability of SST and SLP anomalies through comparison with output from the NCEP1 reanalysis model. The RSOI based estimates, Kaplan *et al*. (1998) and Kaplan *et al*. (2000) for SST and SLP respectively, were found to agree better with the model output in poorly sampled regions compared to successive correction based estimates from da Silva (1994). Additionally, the geophysical consistency between the latent heat flux and SST tendency was found to be poor in the Southern Hemisphere in the successive correction based dataset. It should be noted that no similar analysis was made for RSOI fluxes due to only SST and SLP estimates being available. Based on these results, Sterl (2001) concluded successive correction was inadequate to fill gaps between the VOS observations.

3.4.4 Objective Analysis

Objective analysis combines measurements or estimates of a variable from different sources to give a field that is optimal in terms of a minimised cost function. The cost function gives a measure of the lack of fit of the analysed value to the data according to a set of prescribed conditions that can be dynamically or statistically motivated (e.g. Legler *et al*. 1989). For example, Yu *et al*. (2004a) used objective analysis to estimate gridded fields of the basic variables required to calculate the fluxes from the output of several reanalysis and operational models together with satellite data. The cost function used by Yu *et al*. (2004) minimised the weighted sum of the squared differences between the analysed value and the different estimates. Similarly, Legler *et al*. (1989) used objective analysis to combine monthly wind stress estimates over the Indian ocean from VOS observations with climatological estimates. Legler *et al*. (1989) used a cost function that minimised the squared differences between the analysed values and the two estimates (observed and climatological) as well as two kinematic constraints. The kinematic constraints compared the divergence and curl of the analysed wind stress fields with the climatological values.

3.4.5 Optimal Interpolation

Optimal interpolation (OI) has previously been used to generate gridded fields of the sea surface temperature (e.g. Reynolds 1988; Reynolds and Smith 1994; Kaplan *et al*. 1998; Smith *et al*. 1998; Rayner *et al*. 2003; Reynolds *et al*. 2007), sea level pressure (Kaplan *et al*. 2000) and to assimilate observations into numerical models (e.g. Lorenc 1981). The method estimates unknown or unobserved values of a field based on a weighted sum of observations at nearby locations. i.e.

$$A_k = \sum_{i=1}^{n} w_{ik} B_i \qquad \text{Eq. 3-8}$$

where A_k is the unknown value at location k; B_i is the ith observation and w_{ik} the weight applied to observation i in the interpolation for point k. The weights (w_{ik}) are determined in order to minimize the expected error variance in the estimated value (A_k). This is given by a function of the distribution of the observations, the error characteristics of the observations and the spatial correlation structure of the field. The minimized expected error variance then gives an estimate the uncertainty in the interpolated values.

Whilst various OI schemes exist, such as the RSOI used by Kaplan *et al*. (1997), only the scheme developed by Lorenc (1981) is reviewed in this chapter and used in this thesis. A good description of the OI method and its various implementations can be found in Cressie (1993). The deviation of the interpolated, observed and first guess values from the "true" value can be expressed as

$$\begin{aligned} a &= A - T \\ b &= B - T \\ p &= P - T \end{aligned} \qquad \text{Eq. 3-9}$$

where T is the true value of the field; A, B and P the interpolated, observed and predicted (first guess) values respectively. a, b and p are then the deviations of the interpolated, observed and predicted values from the true value. As noted by Lorenc (1981), the value T is not necessarily the actual true value since we do not wish to analyze features below a certain scale. The deviations will be due to a combination of true small-scale variability and measurement errors (random and systematic) in the observations. Over a large number of realisations we have an associated set of error terms (E) and normalised error terms (ε)

$$\begin{aligned} E^a &= \left\langle a^2 \right\rangle^{1/2} \\ E^b &= \left\langle b^2 \right\rangle^{1/2} \\ E^p &= \left\langle p^2 \right\rangle^{1/2} \\ \varepsilon^a &= \frac{E^a}{E^p} \\ \varepsilon^b &= \frac{E^b}{E^p} \end{aligned} \qquad \text{Eq. 3-10}$$

The angular brackets denote an ensemble mean value.

Expressing the interpolated values and observations as a deviation from the first guess and normalising by the error terms in Eq. 3-10 the interpolation, Eq. 3-8, can be expressed as

$$\frac{A_k - P_k}{E_k^p} = \sum_{i=1}^{N} w_{ik} \frac{\left(B_i - P_i\right)}{E_i^p} \qquad \text{Eq. 3-11}$$

The subscripts i (and j used later) refer to individual observations whilst the subscript k refers to location. A_k gives the OI analysis value at point k; P_k is a first guess value for point k; w_{ik} the optimal weight for observation i used in the analysis at point k; B_i the i^{th}

observation; P_i the first guess value at the location of the i^{th} observation; and E^p the uncertainty or error in the predicted values. It should be noted that an observation can contribute to the analysis value at several different locations.

Rearranging and making a number of substitutions we have

$$\alpha_k \varepsilon_k^a = \pi_k + \sum_{i=1}^{N} w_{ik}\left(\beta_i \varepsilon_i^b - \pi_i\right) \qquad \text{Eq. 3-12}$$

where

$$\alpha = \frac{a}{E^a}$$
$$\beta = \frac{b}{E^b} \qquad \text{Eq. 3-13}$$
$$\pi = \frac{p}{E^p}$$

Squaring Eq. 3-12 and taking ensemble means gives

$$\left(\varepsilon_k^a\right)^2 = 1 + 2\sum_{i=1}^{N} w_{ik}\left(\varepsilon_i^b \langle \pi_k \beta_i \rangle - \langle \pi_k \pi_i \rangle\right) + \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ik} w_{jk} \begin{pmatrix} \langle \pi_i \pi_j \rangle + \varepsilon_i^b \varepsilon_j^b \langle \beta_i \beta_j \rangle \\ -\varepsilon_i^b \langle \beta_i \pi_j \rangle - \varepsilon_j^b \langle \pi_i \beta_j \rangle \end{pmatrix} \qquad \text{Eq. 3-14}$$

This then gives the expected error variance in the interpolated field. In matrix notation Eq. 3-14 becomes

$$\left(\varepsilon_k^a\right)^2 = 1 - 2\mathbf{w_k^T h_k} + \mathbf{w_k^T M w_k} \qquad \text{Eq. 3-15}$$

where

$$\mathbf{w_k} = \left[w_{ik}\right]$$
$$\mathbf{h_k} = \left[\langle \pi_k \pi_i \rangle - \varepsilon_i^b \langle \pi_k \beta_i \rangle\right] \qquad \text{Eq. 3-16}$$
$$\mathbf{M} = \left[\langle \pi_i \pi_j \rangle + \varepsilon_i^b \varepsilon_j^b \langle \beta_i \beta_j \rangle - \varepsilon_i^b \langle \beta_i \pi_j \rangle - \varepsilon_j^b \langle \pi_i \beta_j \rangle\right]$$

The $<\pi_i\pi_j>$ terms are the correlation between the first guess errors and $<\beta_i\beta_j>$ correlations between the observation errors. The $<\beta_i\pi_j>$ terms give the correlation between the errors in our first guess field and the errors in our observations. Minimizing Eq. 3-15 by differentiating with respect to the weights for each observation and equating to zero we have a set of linear equations

$$\mathbf{w_k} = \mathbf{M^{-1} h_k} \qquad \text{Eq. 3-17}$$

The analysis and analysis errors are then given by

$$A_k = \mathbf{h_k^T M^{-1}}\left(\mathbf{B} - \mathbf{P}\right) + P_k \qquad \text{Eq. 3-18}$$

$$\left(E_k^a\right)^2 = \left(E_k^p\right)^2\left(1 - \mathbf{h_k^T M^{-1} h_k}\right) \qquad \text{Eq. 3-19}$$

The observation errors are usually assumed to be uncorrelated (e.g. Reynolds and Smith 1994) for the VOS observations and the error correlations, $<\beta_i\beta_j>$, set to the Kronecker delta δ_{ij}. Likewise, the first guess errors and observation errors are usually assumed to be uncorrelated and their correlation terms, $<\beta_i\pi_j>$, set to zero. Lorenc (1981) and Reynolds and Smith (e.g. Reynolds and Smith 1994) both use a Gaussian function to estimate the correlation between the first guess errors. Other authors use empirical orthogonal functions to estimate the correlation between the first guess errors (e.g. Kaplan *et al*. 1997; Rayner *et al*. 2003).

3.5 Flux Datasets

3.5.1 In situ based estimates

Early surface flux datasets are either regional (e.g. Bunker 1976) or based on fluxes calculated using the monthly mean meteorology (e.g. Hsiung 1986; Oberhuber 1988). Bunker (1976) estimated the fluxes for individual weather observations over the North Atlantic using the bulk formulae and averaged the fluxes onto a semi-regular grid based on the data density. Hsiung (1986) and Oberhuber (1988) estimated the fluxes globally on a 5° and 2° monthly grid respectively from monthly mean estimates of the basic variables.

More recently, fluxes have been estimated globally from individual VOS reports and averaged to give monthly mean and climatological values in the UWM/COADS (da Silva *et al*. 1994) and SOC (Josey *et al*. 1999) climatologies. Both datasets discard VOS reports which do not contain all the information required for flux calculation, apply bias corrections and calculate the fluxes before averaging to give monthly mean fluxes. The monthly fluxes are then smoothed and filled using the successive correction method. Bias corrections are applied to the visual wind speed estimates in the UWM/COADS climatology and the anemometer winds height adjusted to a reference height of 20 m. In the SOC climatology the air temperature is corrected to account for the solar heating errors using the correction of Kent *et al*. (1993a), the humidity corrected to account for the inadequate ventilation of the wet bulb thermometers (Kent *et al*. 1993b; Josey *et al*.

1999) and the SST corrected to account for warm bias in ERI measurements (Kent *et al*. 1993b; Josey *et al*. 1999). Height adjustments are also applied to the air temperature, humidity and wind speed observations to adjust to a reference height of 10 m.

One goal of the global flux atlases is to produce balanced estimates of the net heat flux between the ocean and atmosphere, with the incoming solar radiation balanced by the heat loss from the ocean surface due to the turbulent heat fluxes and radiative cooling. However, with only bias and height adjustments applied, the fluxes in observational datasets tend to be imbalanced, with the oceans gaining of the order 30 W m^{-2} more heat than they lose (e.g. Josey *et al*. 1999). This has led to an inverse analysis of the heat fluxes in the UWM/COADS (da Silva *et al*. 1994) and SOC climatologies (Grist and Josey 2003). Inverse analysis uses basin scale net heat fluxes estimated from hydrographic transects to adjust the different components of the in situ estimates and to bring the net heat flux into agreement with the hydrographic estimates (e.g. da Silva *et al*. 1994; Grist and Josey 2003). By applying inverse analysis to the SOC climatology Grist and Josey (2003) reduce the net imbalance from a gain of 30 W m^{-2} by the oceans to a small heat loss of -5 W m^{-2} after adjustment. Whilst inverse adjustment brings the net heat fluxes into balance, biases are introduced compared to research buoy observations (Grist and Josey 2003).

3.5.2 Reanalysis based estimates

In addition to providing estimates of the meteorological parameters required to estimate the fluxes, the reanalysis models described in Section 3.3 (ERA40, JMA, NCEP1 and NCEP2) also provide estimates of the turbulent fluxes. These have the same spatial and temporal resolutions as the basic variables, typically 6 hourly and on a 2.5° grid, and provide balanced flux estimates. However, whilst the flux estimates are balanced, the estimates suffer from the same problems as the basic variables. Inhomogeneities exist in the flux estimates due to the changing observing system, as described in Section 3.3, and the fluxes are poorly represented in coastal and high variability regions due to the relatively coarse spatial resolutions. In addition to these problems, the bulk formulae used in the models to calculate the fluxes have been shown to give biased estimates of the fluxes compared to direct measurements (e.g. Brunke *et al*. 2003).

3.5.3 Satellite and combined satellite / reanalysis based estimates

Several datasets have been generated combining meteorological parameters from one or more satellites and reanalysis model output. Examples include the Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data (HOAPS) (Bakan *et al*. 2000), the Goddard Satellite based Surface Turbulent Fluxes version 2 (GSSTF2) dataset (Chou *et al*. 2003), and the Woods Hole Oceanographic Institution's Objectively Analyzed air – sea Fluxes (OAFlux) data set (Yu *et al*. 2004a; Yu and Weller 2007; Yu *et al*. 2008).

The HOAPS dataset estimates the surface turbulent fluxes, net longwave heat flux and balance on a variety of space (0.5°, 1° and 2.5°) and time (1 day, pentad and monthly) scales for the period 1987 - 1997. Estimates of the wind speed and atmospheric specific humidity are estimated from SSM/I data and the sea surface temperature from the AVHRR Pathfinder dataset. The specific humidity at the sea surface is estimated from the SST data assuming saturation and the air temperature estimated assuming a constant 80 % relative humidity. The turbulent fluxes are estimated using the Smith (1988) parameterisation for the transfer coefficients. The assumption of the 80% relative humidity will lead to potentially large biases in both the air temperature estimates and sensible heat fluxes. The impact on the latent heat flux, through the stability correction, will be smaller than for sensible heat flux and largest at low wind speeds where the effects of stability are greatest.

The Goddard Satellite based Surface Turbulent Fluxes version 2 (GSSTF2) dataset (Chou *et al*. 2003) estimates the surface turbulent fluxes on a 1° daily grid for the period 1987 – 2000. Monthly mean fluxes are also estimated based on the average of the daily values. The fluxes are calculated using the roughness length parameterisations of Liu *et al*. (1979) and stability profiles of Large and Pond (1982) for stable conditions. The profiles of Businger et al (1971) are used for unstable conditions. Estimates of humidity and wind speed from SSM/I data are used together with sea surface temperature, air temperature and sea level pressure estimates from the NCEP1 reanalysis project.

Estimates of the turbulent fluxes are made on a 1° daily grid in the OAFlux data set (Yu *et al*. 2004a; Yu and Weller 2007; Yu *et al*. 2008). The basic variables are first estimated on a 1° daily grid using objective analysis and a number of input datasets.

Fluxes are then calculated from the daily fields using the COARE 3.0 algorithm (Fairall *et al*. 2003). Estimates of the wind speed from SSM/I and AMSR-E passive microwave radiometers are used together with scatterometer estimates from the QuickSCAT sensor on the SeaWinds satellite. Estimates of the wind speed from the NCEP1, NCEP2 and ERA40 reanalysis models are also used. The humidity is estimated based on the output of the NCEP1, NCEP2 and ERA40 reanalysis models together with estimates from the SSM/I data. SST estimates from the NOAA Optimum Interpolation daily 0.25° SST dataset (Reynolds *et al*. 2007) are used together with the output from the NCEP1, NCEP2 and ERA40 models. Air temperature is estimated based on the output of the ERA40, NCEP1 and NCEP2 models. OAFlux also contains estimates of the fluxes (and basic variables) for the pre-satellite era. These are a combination of the reanalysis models used in OAFlux and will suffer from the same inhomogeneity problems.

3.6 Random Errors and Flux Uncertainties

3.6.1 Random and Systematic Uncertainty

Table 3-2: Random and systematic uncertainty estimates from Blanc (1986) and Gleckler and Weare (1997). V_s = wind speed (m s^{-1}).

Variable	Random uncertainty		Systematic uncertainty	
	Blanc	Gleckler and Weare	Blanc	Gleckler and Weare
Air temperature	0.3 °C	1 °C	0.5 °C (day) 0.0 °C (night)	0.5 °C
Wet bulb temperature	0.3 °C	N/A	0.7 °C (day) 0.2 °C (night)	N/A
Dew point temperature	N/A	1°C	N/A	0.25 °C
Sea surface temperature	0.5 °C	1 °C	0.3 °C	0.5 °C
Wind speed	0.5 m s^{-1}	0.5 V_s	0.1 V_s	1.5 m s^{-1}
Sea level pressure	1 hPa	N/A	N/A	N/A
Cloud cover	N/A	2/8	N/A	N/A
Specific humidity	0.2 g Kg^{-1}	N/A	0.1 g Kg^{-1}	N/A

Random errors and bias uncertainties for the VOS observations and flux estimates have been reviewed and summarised by Blanc (1986) and Gleckler and Weare (1997) (Table 3-2). The values given by Blanc (1986) are typical of measurements from

an ocean weather ship. The values listed by Gleckler and Weare (1997) are for observations typical of the VOS and have been based on a review of previous studies. More recent estimates have been made using semi-variogram analyses.

3.6.2 Semi-variogram analysis

The semi-variogram analysis method has been used previously to estimate the random errors in VOS observations (Lindau 1995; Kent *et al*. 1999; Lindau 2003; Kent and Berry 2005; Kent and Challenor 2006). In this method the total error variance (i.e. squared differences) between pairs of observations is plotted against separation distance. This is usually referred to as the variogram. These differences are a function of the true variability of the field and observational errors (e.g. Lindau 1995). By fitting a regression model through the observed error variances, or sample variogram, the variogram can be extrapolated to zero separation distance and the contribution from natural variability removed. The value at zero separation distance will then be due to observational errors and small scale variations. By assuming the small scale variations are small compared to the observational errors, the observational errors can then be estimated as the square root of half the intercept value (e.g. Kent *et al*. 1999). This assumes that the contribution to the differences between pairs of observations from each observation is equal. The square root of half the squared differences plotted as a function of separation distance is usually called the semi-variogram.

In previous studies using semi-variograms to estimate the random errors linear models have been used to extrapolate the semi-variogram to zero separation distance (e.g. Lindau 1995; Kent *et al*. 1999; Kent and Berry 2005; Kent and Challenor 2006). The regression has been performed for both individual squared differences (e.g. Kent *et al*. 1999) and averages of the squared differences binned on separation distance (e.g. Kent and Berry 2005). The regression against the individual squared differences can be adversely affected by outliers due to the use of squared differences. As a result, Kent and Challenor (Kent and Challenor 2006) attempted to minimise the influence of the outliers, examining the use of the individual squared differences, average squared differences and a generalized linear model with a gamma error function in the regression. Little difference was found in the results of the different regressions with the use of the binned differences chosen as being more statistically sound. The random

error estimates from the most recent study (Kent and Berry 2005) can be found in Table 3-3.

Table 3-3: Globally averaged (weighted by ocean area) random error estimates from Kent and Berry (2005).

Variable	No Height Correction	Neutral Stability Height Correction
Pressure	2.1 ± 0.2 hPa	N/A
Wind speed	2.2 ± 0.1 m s^{-1}	2.0 ± 0.1 m s^{-1}
Air Temperature	1.2 ± 0.1 °C	1.0 ± 0.1 °C
Sea Surface Temperature	1.2 ± 0.1 °C	N/A
Specific Humidity	1.2 ± 0.1 g Kg^{-1}	1.1 ± 0.1 g Kg^{-1}

These estimates are similar to those given in Gleckler and Weare (1997) with the exception of wind speed. Gleckler and Weare (1997) estimated the random error in the wind speed based on the average monthly standard deviation assuming any variability is due to random errors and acknowledging that they are likely to be overestimating the errors.

Table 3-4: Commonly used variogram models. $\gamma(h_{sv})$ is the modelled variogram at separation h_{sv}; $C_0 \geq 0$ and gives the value at zero separation (i.e. the intercept), $C_0 + C_1$ gives the sill value (i.e. the value at which the variogram levels off). The range is given by $3a_{sv}$ for the exponential and Gaussian models. b_{sv} gives the gradient for the linear model.

Model	Formula	Equation
Exponential	$\gamma(h_{sv}) = C_0 + C_1\left(1 - \exp\left(-\frac{h_{sv}}{a_{sv}}\right)\right)$	Eq. 3-20
Gaussian	$\gamma(h_{sv}) = C_0 + C_1\left(1 - \exp\left(-\frac{h_{sv}^2}{a_{sv}^2}\right)\right)$	Eq. 3-21
Linear	$\gamma(h_{sv}) = C_0 + b_{sv}h_{sv}$	Eq. 3-22

Whilst using a linear model, Kent *et al*. (1999) noted that a non-linear model may give a better fit to the observed variograms. However the use of a non-linear model was ruled out on the basis that a non-linear model could give a serious underestimate of the intercept value, especially if data near zero separation distance are sparse. It will be shown in Chapter 5 that this is not necessarily the case and that other models can give better results. Table 3-4 and Figure 3-11 give examples of some of the commonly used models in geostatistics (e.g. Isaaks and Srivastava 1990). The linear model is the simplest of the three, increasing linearly from a zero or positive (nugget) value at the

intercept. The exponential model increases almost linearly from the intercept before levelling off asymptotically towards a sill at higher separation distances. The Gaussian model shows parabolic behaviour at low separation distances before increasing almost linearly and then asymptotically approaching a sill.

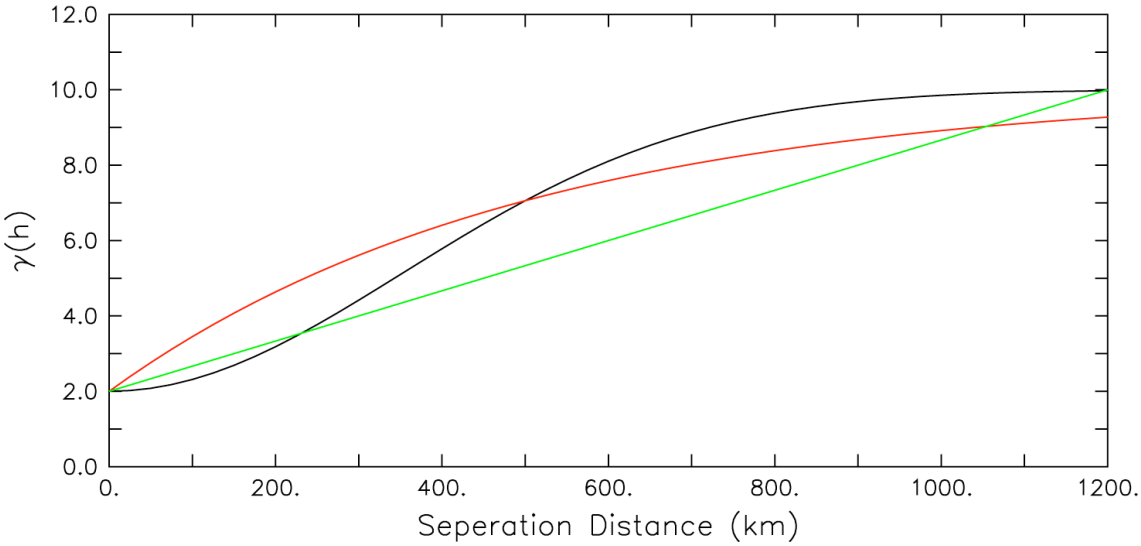

Figure 3-11: Examples of the different variogram models: exponential (red); Gaussian (black); linear (green). C_0 = 2, C_1 = 8 and a_{sv} = 500 for the Gaussian and exponential models. C_0 = 2 and h_{sv} = 1/150 for the linear model.

3.6.3 Sampling Error

Sampling errors are errors in the mean value due to incomplete sampling and have typically been estimated using one of two methods. The first empirically estimates the errors by sub-sampling either the actual data or idealized fields, such as model output, and examining the impact of the sub-sampling on the gridded fields. The second uses a combination of propagation of errors and analysis of variance to estimate the expected error variance due to under sampling.

As an example of sub-sampling, Gulev *et al*. (2007b, 2007a) sub-sampled the output from the NCEP1 and ERA40 reanalysis models using either random sampling of the number of VOS observations or sampling based on the actual times and locations of VOS observations (VOS sampling). The errors in the monthly mean estimates from both sampling methods were estimated by comparison with the actual model values and the differences attributed to sampling errors. For the random sampling, errors of 2.5 – 3 °C were reported for air temperature, 2 – 2.5 g Kg^{-1} for specific humidity and 3 m s^{-1} for

wind speed in the poorly sampled regions outside of the shipping lanes. In well-sampled regions, such as the shipping lanes and coastal regions, the errors decreased to a minimum of 0.2 °C for air temperature and 0.1 – 0.4 m s^{-1} for wind speed. No minimum values are given for the specific humidity. When the actual VOS sampling was used the sampling uncertainties were 20 – 30 % larger. The larger errors found in the VOS-like sampling were due to a smaller effective number of independent measurements due to the clustering of the VOS observations in time and space (Gulev *et al*. 2007a).

Examples of using propagation of errors and analysis of variance to estimate the sampling errors can be found in Jones *et al*. (1997) and Rayner *et al*. (2006). For example, Rayner *et al*. (2006) estimated the grid box and large-scale (i.e. regional and global) uncertainty in the HadSST2 dataset due to random, sampling and bias uncertainties. The contribution by random and sampling errors was estimated as the increase in the variance of grid box anomalies due to under sampling and measurement errors. This increase was calculated by modelling the observed variance of the anomalies in a grid box as a function of the 'true' climatological variability, the measurement errors and the number of observations. For a given grid box, the variance of the monthly anomalies based on n observations is modelled as (Rayner *et al*. 2006)

$$S_n^2 = S_{ne}^2 + S_{nt}^2 \qquad \text{Eq. 3-23}$$

where S_n^2 is the observed variance of the monthly mean anomalies based on n_{se} observations, S_{ne}^2 is the contribution to the variance by measurement errors and S_{nt}^2 is the variance of the grid box average of n_{se} correlated observations. The contribution from measurement errors is given by

$$S_{ne}^2 = \frac{m^2}{n_{se}} \qquad \text{Eq. 3-24}$$

where m is the measurement error for an individual observation. The variance of the grid box average is then modelled using propagation of errors following Yevjevich (1972), i.e.

$$S_{nt}^2 = \frac{c^2\left(1 + \left(n_{se} - 1\right)\bar{r}\right)}{n_{se}} \qquad \text{Eq. 3-25}$$

where c^2 is the true temporal variance within the grid box and $\bar{r}$ the mean correlation of the true point values in the grid box with every other point value in the grid box. Eq. 3-23 becomes

$$S_n^2 = \frac{c^2 + c^2(n_{se} - 1)\bar{r} + m^2}{n_{se}} \qquad \text{Eq. 3-26}$$

For each grid box Rayner *et al*. (2006) fitted the true climatological variability, the mean correlation between the points within the box and the individual measurement errors based on the observed variance of the anomalies. The increase in the variance of the anomalies due to random and sampling errors was then estimated as the difference between Eq. 3-26 with the actual number of observations and when *n* tends to infinity. i.e.

$$S_{err}^2 = S_n^2 - S_\infty^2 = \frac{c^2(1 - \bar{r}) + m^2}{n_{se}} \qquad \text{Eq. 3-27}$$

where $S_{err}{}^2$ is the contribution from random and sampling errors.

Based on the results of the variance partitioning, Rayner *et al*. (2006) generally found the individual grid box uncertainties to be less than 0.1 °C during September 2003. During this period the oceans were generally well sampled in terms of SST measurements from the VOS and drifting buoys.

3.6.4 Flux Uncertainty

The uncertainty in a function of several variables due to the uncertainty in each variable can be estimated using propagation of errors (e.g. Taylor 1997, Section 9.2). The uncertainty can be expressed as

$$\sigma_V^2 = \sum_{i=1}^{N} \left(\frac{\partial V}{\partial v_i} \right)^2 \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{\partial V}{\partial v_i} \frac{\partial V}{\partial v_j} \rho_{ij} \sigma_i \sigma_j \qquad \text{Eq. 3-28}$$

where V denotes a function of N variables v, i.e. $V = f(v_1, v_2 \ldots v_N)$; σ_i is the uncertainty in the ith variable, σ_j the uncertainty in the jth variable and ρ_{ij} the correlation between the errors in the ith and jth variables. This method has been used in several studies to estimate the uncertainty in the fluxes with different assumptions made about the correlation terms.

Blanc (1986) estimated the uncertainty in fluxes calculated for individual observations made by an Ocean Weather Ship (OWS) using the error estimates in Table 3-2 and propagation of errors. The errors in the observations were assumed to be uncorrelated across the variables and at low flux values large uncertainties (as a percentage of the fluxes) reported. At higher flux values the uncertainties were found to

decrease and at around 20 – 30 W m^{-2} and 40 W m^{-2} for sensible and latent heat fluxes respectively the uncertainties were found to have decreased to 30 % of the flux value. At higher magnitudes the uncertainty continued decreasing before levelling at around 10 % for each mean flux estimate. The greatest contributions to the uncertainty in the fluxes were found to be from, in order of greatest contribution, the wind speed, air temperature and humidity and sea surface temperature measurements. Out of the different bulk formulae examined Blanc (1986) found the parameterisations of Liu *et al*. (1979) to be the most adversely affected by random errors.

Gleckler and Weare (1997) estimated the uncertainties in the fluxes from the Oberhuber flux atlas (Oberhuber 1988) due to random and systematic errors (see Table 3-2). The correlation terms between the different variables were estimated from the monthly mean values for each variable and the random errors reduced by $N^{-1/2}$ (where N is the number of observations in a given month). Over all regions, Gleckler and Weare (1997) found the random uncertainties to be small and the total uncertainty to be dominated by the systematic component. The main contributions to the systematic uncertainties, in order of size, were those from the wind speed measurements, transfer coefficient, air – sea humidity gradient and correlation terms. Globally, the uncertainties in the annual mean latent heat fluxes were found to be greater than 10 W m^{-2} everywhere, often exceeding 30 W m^{-2} in the tropical western oceans and almost 50 W m^{-2} over the western boundary currents. Similar patterns were found in the sensible heat flux estimates but with smaller magnitudes. Over the western boundary currents the total uncertainty in the sensible heat flux was found to be larger than 15 W m^{-2} and between 5 – 10 W m^{-2} elsewhere.

Uncertainty estimates are also provided for the OAFlux dataset. The different input data sources are assumed independent of each other in the estimation of the flux uncertainty estimates (Yu *et al*. 2008). However, due to the choice of similar data sources (such as for SST), this is unlikely to be the case and the estimates of the basic variables from the different products, and their errors, likely to be correlated. If they are correlated this assumption would result in an underestimation of the uncertainties. Little further information is given on the calculation of the uncertainty estimates in the available documentation (Yu *et al*. 2004a, 2004b; Yu and Weller 2007; Yu *et al*. 2008) making it difficult to assess whether or not this is the case and whether the uncertainty estimates are realistic.

In addition to the uncertainty in globally gridded flux datasets, the uncertainty due to choice of bulk formula has been examined by Brunke *et al*. (2002) through comparison of the fluxes calculated using a number of different parameterisations and 4 different observational datasets. Slight differences were found in hourly flux estimates from the different parameterisations, with the differences largest at both low and high wind speeds. Large differences were also found in the latent heat flux and wind stress at moderate tropical sea surface temperatures (~27 °C) and also for latent heat flux under very unstable conditions. Similar results were found for monthly mean fluxes, with maximum differences in the monthly latent heat flux of ~ 23 W m^{-2} (~ 16 % of the flux averaged across the different algorithms) at moderate tropical SSTs. The uncertainty in the fluxes due to the choice of parameterisation is difficult to estimate from the results of Brunke *et al*. (2002) since the standard deviation of the flux estimates across the parameterisations was not given.

3.7 Discussion and Summary

There are a number of previous flux datasets based on a variety of sources, ranging from *in situ* observations to satellite measurements and model output. A number of estimates of the uncertainty in the different flux products have also been made. As a result, the question as to why a new flux dataset and new uncertainty estimates are needed needs to be answered.

Each of the previous surface flux datasets contains problems as discussed in this Chapter, and recent research means improved estimates can be made using the different sources available. The VOS based estimates have either lacked the metadata required to properly height and bias adjust the observations prior to flux calculation, resulting in biases in the resultant flux fields, or problems have been highlighted with the averaging and flux calculation strategies.

The current reanalysis based estimates have a relatively poor spatial resolution, resulting in poor flux estimates in high variability and coastal regions. Additionally, the flux parameterisations used in the models have been shown to give biased estimates. It should be noted that a number of these deficiencies are being addressed by a new generation of reanalysis models in development (e.g. ERA Interim), but the issue of a changing observing system and the inhomogeneities introduced is unlikely to be resolved.

The satellite and combined satellite / reanalysis datasets only provide relatively short time series compared to the VOS and reanalysis based estimates. Additionally, the satellite estimates can suffer from inhomogeneities due to sensor drift, changing satellites and changing times of the satellite overpasses. The combined satellite and reanalysis datasets will also suffer from the inhomogeneities and problems seen in the reanalyses datasets.

The objective analysis approach used in the OAFlux dataset has the potential to improve the flux estimates. However, each source needs to be carefully treated and homogenized. Additionally, the error co-variances between the different data sources need to be fully understood in order to accurately estimate the uncertainty in the resulting flux fields and to gain the maximum benefit from each dataset used. This homogenization and error characterisation is currently lacking in the OAFlux dataset. As a result, the fluxes are likely to contain severe inhomogeneities due to the choice of source datasets used and the uncertainty estimates are likely to be an underestimate due to the assumption of independence between the different datasets.

Overall, the recent advances in understanding the different sources of data, and the sources of error in those data, coupled with improved gridding techniques and increased computer power mean that improved flux and flux uncertainty estimates can be made. In this thesis, a number of the different methods discussed in this chapter will be bought together to form a coherent method for estimating the fluxes and their uncertainty in a single study. This will be the first time that the gridding and flux calculation method has been focused on estimating the uncertainty in the fluxes and on explicitly minimizing the error variances.

4 Development of Dataset Construction Methodology

4.1 Introduction

The method developed in this chapter brings together a number of disparate techniques described in the previous chapter to give flux estimates with known (or estimated) minimized error variances. An overview of the method developed is first given together with the rationale behind the different choices made. The data used are then described followed by the adjustments made to the observations to account for the different sources of bias. A number of new bias adjustments are developed where previous ones have been found to be inadequate, This is then followed by a description of the implementation of the different techniques used and their application to the chosen data sources.

4.2 Outline of method and Rationale

In order to accurately estimate the turbulent air – sea fluxes and their uncertainty a number of factors need to be taken into account and appropriate choices made. The main choices are: the data source, or sources, to use; the quality control procedures applied; the choice of bias corrections to apply (if needed); and the flux calculation and averaging strategy. The main factors that need to be taken into account are the error characteristics of the data used, i.e. random and systematic errors, and how the observations are distributed in space and time. The impact that these have on the averaging method chosen and flux calculation strategy also need to be taken into consideration.

The first choice, that of the data to use, is relatively straightforward. One of the longer term goals of the work presented in this thesis is to use the method developed to update and produce a new version of the NOC (formerly SOC, Josey *et al*. 1999) climatology. Hence only the VOS observations have been used. This also has the advantage of maintaining the independence of the dataset from recent satellite based flux estimates and from buoy observations.

Following the choice of data to use, the Quality Control (QC), bias corrections and method of estimating the uncertainty in the source data need to be decided. As seen in the previous chapter, the VOS observations are relatively well understood in terms of systematic and random errors. The International Comprehensive Ocean – Atmosphere

Data Set (ICOADS), used as the source of VOS observations in this thesis, contains quality indicator flags that will be used to QC the data. Whilst bias adjustments already exist for some variables, problems have been highlighted, notably for air temperature and humidity (see Chapter 3). As a result two new bias corrections (air temperature and humidity) have been developed and applied to the VOS observations as part of this thesis. A third new bias correction is applied to visual wind speed estimates based on other authors work. The random errors for the individual observations have also been re-estimated using improved error models.

The next set of choices include how to grid the observations and whether to calculate the fluxes before or after gridding. The order of flux calculation is determined by the characteristics of the observations chosen. The potentially large random errors seen in the individual VOS observations will impact the flux estimates in two ways. Firstly, the random errors will contribute directly to the uncertainty in the fluxes. Secondly, due to the non-linearity of the bulk formulae, the random errors in the basic variables will not necessarily cancel out in the flux estimates even when averaged over a large number of observations and may lead to biased estimates. As a result, the random errors in the input variables need to be minimized before flux calculation. It should be noted that this effect has been ignored in all previous calculations. However, whilst we want to minimize the random errors we need to be careful to maintain the synoptic scale correlations between the basic variables (i.e. Sampling vs Classical fluxes, Section 3.4.2). These conflicting constraints have led to the choice of gridding the data on a 1 ° daily grid before calculating the fluxes. In regions where we have multiple observations the random errors will be reduced and the synoptic scale correlation between the variables maintained. In regions with only a single observation or relatively few observations the results will be comparable to fluxes calculated using the sampling method.

The gridding method is also guided by the choice of observations and the requirement to estimate the uncertainty in the gridded fields and fluxes routinely. The inhomogeneous sampling by the VOS observations may lead to large sampling errors and needs to be taken into account in order to make accurate uncertainty estimates. Additionally, the choice of gridding the observations on a 1° daily grid will lead to gaps in the field if a simple method, such as using the arithmetic mean, is used. OI meets the requirements of taking the spatial sampling of the observations into account when

gridding the observations into complete fields and producing uncertainty estimates routinely as part of the output. All observations within a defined radius are used, rather than just the observations within the grid cell, and the weights are determined optimally to minimize the expected error variance (i.e. uncertainty) in the interpolated fields. As a result, grid cells without observations are filled based on the observations in nearby grid cells. OI has therefore been chosen to grid the observations and the fluxes calculated using the output from the OI. The uncertainty in the fluxes will then be estimated using propagation of errors with the uncertainty estimates produced by the OI.

The final choice to make is how to calculate the monthly mean values from the daily flux estimates. Several different methods could be used to estimate the monthly mean values for the different variables from the daily OI fields and flux estimates. The individual days could be given equal weighting or weighted by their uncertainty with higher uncertainties given lower weights. Alternatively, only days with an analysis value (as opposed to the incremented previous days analysis) could be used. One of the goals of this thesis is to provide a dataset with realistic uncertainty estimates. Applying a non-uniform weighting to the individual days will lead to days with lower uncertainty being given higher weight. Similarly, the lower uncertainty estimates will be given a higher weight in the estimate of the uncertainty in the monthly mean. This will lead to the calculated uncertainty in the monthly mean being biased towards the well sampled days and underestimated as a result. This leads to the choice of giving each day's analysis equal weight in the estimation of the monthly mean values and accounting for the correlation between days in estimating the uncertainty.

4.3 Data Sources Used

4.3.1 VOS Observations and Metadata

Observations from the VOS contained in ICOADS v2.4 (Worley *et al*. 2005) and made between 1970 and 2006 have been used to estimate the fluxes together with the metadata information from the WMO Publication Number 47 series (Pub. 47) (e.g. Kent *et al*. 2007). The datasets have been merged following Josey *et al*. (1999), matching the call-signs in ICOADS with those in Pub. 47 and assigning measurement methods and height to the observations. The period 1970 – 2006 has been chosen due to the availability of metadata on the observing practices and heights required for the

various bias and height corrections (Section 4.4). Over this period all the variables required to calculate the fluxes (air temperature and humidity, wind speed, sea level pressure, sea surface temperature and cloud cover) are available from the VOS observations.

4.3.2 Quality Control

Monthly summary files containing statistics, such as the mean and median values, have been calculated and released as part of ICOADS (Wolter 1997). These have been used in ICOADS to generate "trimming flags" for the different variables for use in quality controlling of the observations. Table 4-1 gives the information on the different limits available. Only observations within 4.5 σ (trimming flag <= 5) have been used in this study. These limits have been chosen to discard the outliers whilst retaining climatologically extreme but valid events (Wolter 1997). If a quantity is derived from a combination of variables, such as the specific humidity, each component must pass the QC (see Table 4-2 for a list of variables and flags checked).

Table 4-1 ICOADS trimming flags. σ_l is the difference between the 1st and 3rd sextiles and σ_u the difference between the 3rd and 5th sextile. δ is the difference between the observation and the climatological median value.

Trimming Flag	Limits	Trimming Flag	Limits
1	$-2.8\sigma_l <= \delta <= 2.8\sigma_u$	6	$\delta < -4.5\sigma_l$
2	$-3.5\sigma_l <= \delta < -2.8\sigma_l$	7	$\delta > 4.5\sigma_u$
3	$2.8\sigma_u < \delta <= 3.5\sigma_u$	14	data unusable
4	$-4.5\sigma_l <= \delta < -3.5\sigma_l$	15	data missing or not computable
5	$3.5\sigma_u < \delta <= 4.5\sigma_u$		

Table 4-2: Trimming flags checked for each variable used

Variable	Trimming Flags Checked
Air Temperature	Air Temperature
Sea Surface Temperature	Sea Surface Temperature
Wind speed	Zonal and meridional wind speed
Sea Level Pressure	Sea Level Pressure
Specific Humidity	Air temperature and relative humidity
Cloud cover	No flags checked.

In addition to the ICOADS trimming flags, a track checking algorithm has been applied to the observations following Kent and Challenor (2006). To perform the track check, the observations for a given call-sign or unique identifier were first split into

blocks of successive observations with gaps no longer than a week between observations. Based on the time of the observations and the reported latitude and longitude the speed of the VOS making the observation was estimated and any observation requiring a ship speed > 100 km h^{-1} discarded to remove any observations with a gross error in its location. For observations where a track check was not possible, either through a missing call-sign or a non-unique identifier, the observations have been retained. The impact of the quality control and track checking on the amount of data available will be shown in Chapter 5.

4.4 Bias Corrections

Bias corrections are only applied to the air temperature and humidity measurements and the visual wind speed estimates. Whilst corrections have been proposed for SST observations from both methods commonly used by the VOS, ERI and insulated buckets, there is a large uncertainty both in the sign and magnitude of the bias (Elizabeth Kent, pers. comm., also Berry and Kent 2009). As a result, unadjusted SST observations have been used. Similarly, the bias correction for anemometer winds suggested by Moat *et al*. (2006a) has not been applied due to a lack of knowledge of the location of the anemometers on board the VOS. The sea level pressure and cloud cover observations are not thought to be biased.

4.4.1 Air Temperature

Air temperature observations are corrected using a bias correction developed as part of this thesis building on the work of previous authors (e.g. Kent *et al*. 1993a; Anderson and Baumgartner 1998). The correction models the heat budget of the ships structure and sensor between sunrise and the time of measurement, balancing the heat storage, convective and conductive cooling with the amount of solar insolation received. The heat budget is given by

$$\frac{d\left(\Delta T_{cor}\right)}{dt} + \overbrace{x_2\left(x_3 V_{rel}^{x_4} + x_5\right)}^{\text{convective and conductive cooling}} = \overbrace{x_1 R_{SW}}^{\text{solar radiative heating}} \qquad \text{Eq. 4-1}$$

where ΔT_{cor} (°C) is the radiative heating error, t (s) is time, V_{rel} (m s^{-1}) the relative wind speed and R_{sw} (W m^{-2}) the incident solar radiation. x_1 through x_5 are empirically determined coefficients and relate to the typical thermal properties of the ships and the sensor exposure. The form of the term accounting for the convective and conductive

cooling are based on the heat loss from a simple block shaped structure, further details can be found in Berry *et al*. (2004). A term to account for the longwave radiative heating and cooling of the sensors and ships' superstructure was originally included in Eq. 4-1. This term was found to be negligible compared to the other terms and discarded.

In order to apply the correction Eq. 4-1 needs to be solved for ΔT, integrating between sunrise and the time of observation, and the empirical coefficients determined. The Okta model of Dobson and Smith (1988) is used to estimate the solar radiation and the full derivation and integration of Eq. 4-1 is given in Berry *et al*. (2004). In order to determine the empirical coefficients, estimates of the bias are needed. These have been estimated as the difference between the observations and a night-time only version of the optimally interpolated (see section 4.6) air temperatures. A subset of the observations have then been selected for each year by randomly selecting 1000 observations each month to ensure the annual cycle is fully sampled. For each year a non-linear regression (using the NAG subroutine NAG E04USF (NAG 2008)) is performed, minimizing the sum of squared differences between the error estimates and the solution of Eq. 4-1. The regression is performed on annual subsets of data to account for changing ship characteristics and a large sample size used to minimize the noise in the VOS observations and error estimates.

Due to the non-linearity of the solution, the results of the fit will depend strongly on the initial coefficients chosen, if the choice is poor the fit returned may not be optimal. As a result a Monte Carlo approach has been used with the fitting routine run 500 times. For each run the initial coefficients have been randomly selected between the limits given in Table 4-3. Results of the bias estimation, fitting of the coefficients and bias correction are given in Chapter 5. The impact of the bias correction on the random error estimates will also be shown.

Table 4-3: Lower and upper limits for initial x values in fitting routine.

Coefficient	Lower limit	Upper limit
x_1	1×10^{-4}	1×10^{-2}
x_2	1×10^{-3}	1×10^{-3}
x_3	1	1×10^{3}
x_4	1	1
x_5	10	1×10^{3}

4.4.2 Humidity

Measurements of the humidity made using screens are biased high due to the inadequate ventilation of the wet bulb thermometers. Kent *et al*. (1993b) and Josey *et al*. (1999) both propose corrections to the dew point temperature from screen measurements. However, as noted in Chapter 3 (Section 3.2.2) the corrections peak at moderate or moderate to high specific humidity values before decreasing. Thus, it appears likely that both of these corrections under correct high specific humidity estimates and possibly over correct at lower humidity values.

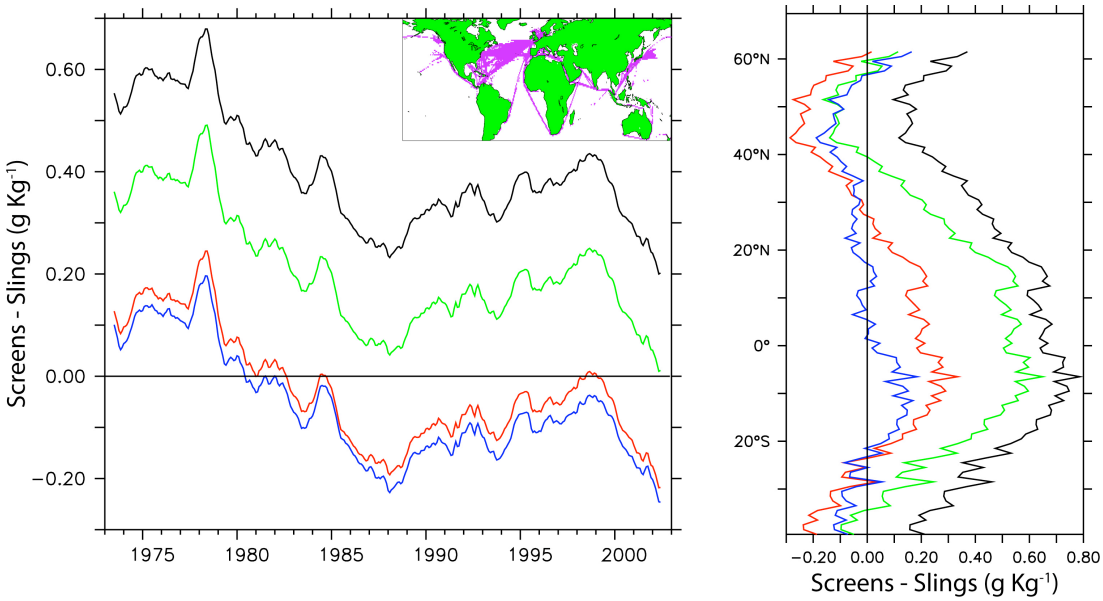

Figure 4-1: Differences between screen and psychrometer specific humidity observations averaged globally (left) and zonally (right). Differences before correction (black), after correction following Kent *et al.* (1993b) (green) and after correction following Josey *et al.* (1999) (red) are shown. The differences after a reduction of 3.3 % in the screen specific humidity observations are also shown (blue). The inset shows the regions with at least 3 observations from both methods and at least 5 years of data.

Figure 4-1 shows a comparison of the humidity estimates from both methods before and after correction. The screen humidity estimates have also been reduced by the average ratio of the differences between screen and sling measurements to the sling measurements (3.3%). The relative wet bias in the uncorrected screen observations can be clearly seen with the screens biased high by 0.3 – 0.6 g Kg^{-1} when averaged globally. The variation with latitude can also be seen with higher biases in the tropics compared

to the higher latitudes. The under correction of the correction proposed by Kent *et al*. (1993b) can be seen both in time and zonally. The correction used by Josey *et al*. (1999) and the reduction of 3.3 % have a similar impact when averaged globally. However, the correction used by Josey *et al*. (1999) under corrects at lower latitudes and over corrects at higher latitudes (as expected from Figure 3-6).

Reducing the screen humidity observations by 3.3 % also over and under corrects at higher and lower latitudes respectively, but differences are much smaller compared to the other two corrections. Whilst this represents a slight over (under) correction at low (high) humidity values it has been adopted in this thesis. It should be possible to improve this correction using an approach similar to Kent and Kaplan (2006) to compare co-located observations from each method.

4.4.3 Visual Wind Speeds

The visual winds are adjusted following Lindau (1995) and using the polynomial (Eq. 3-7) suggested by Thomas *et al*. (2005) to account for the biases in the WMO1100 scale. Thomas *et al*. (2008) examined the trends in adjusted visual and anemometer winds, finding a residual spurious increasing trend in the visual wind speeds. The cause of this is speculated as being due to the observers making the visual wind reports being increasingly influenced by anemometers (without height corrections applied) on board the vessels. i.e. after making the visual report they are checking the ships anemometer and adjusted their visual estimates accordingly. An adjustment has been made to account for this influence. Applying a factor of 1 prior to the end of 1985 and then a factor that linearly decreases to 0.95 by 2000 brings the trends seen in the visual winds into agreement with those seen in height corrected anemometer winds (Elizabeth Kent, pers. comm., also Berry and Kent 2010). This time varying factor has been applied to the adjusted visual winds in this thesis.

4.4.4 Height Corrections

The observations used in this study have been height corrected to a reference height of 10 m using the bulk formula (Chapter 2) before gridding but after the other bias adjustments have been applied. The parameterisations of Smith (1980, 1988) are used.

4.5 Random and Bias Uncertainty

4.5.1 Random Errors

Random errors in the VOS observations have been estimated using the semi-variogram method (see Section 3.6.2). The sample semi-variograms are estimated for each variable and 30° ocean region and month, using all pairs of observations made on the same day and within 500 km of each other. The squared differences are averaged into 30 km separation distance bins and the process repeated for the observations before and after bias and height correction. The time separation requirements have been relaxed from simultaneous observations, as used in Kent and Berry (2005), to observations made the same day. This was done because the OI scheme requires the expected error variance between the observations and the daily mean value (see Section 4.6).

Unlike previous authors, three different models are used to model the observed sample semi-variograms and to extrapolate the variograms to zero separation distance. These are the linear, exponential and Gaussian models (see Table 3-4). For each variable, the parameters (range, sill and intercept or intercept and gradient) in the different models are fitted to the observed sample semi-variograms at every 30° grid cell and month using a (non-linear) least squares regression. The models are then assessed based on the number of grid cells and months with a valid fit and the mean R^2 (i.e. Pearson correlation coefficient) value for each model. The validity of the fits are assessed by testing for significant auto-correlation using a Durbin-Watson test and any fit showing significant auto-correlation at the 95% level discarded. The model giving the largest number of regressions with a valid fit and the highest mean R^2 value is then used to estimate the uncertainty for that variable. If different models are selected based on these criteria the simpler of the two models is used (i.e. linear model in preference to exponential model and exponential model in preference to Gaussian). Results of the estimation of the sample semi-variograms, regression of the model semi-variograms and comparisons are given in Chapter 5. The effects of the height and bias corrections on the random errors in wind speed, air temperature and humidity observation will also be shown.

4.5.2 Bias Uncertainties

In order to correctly estimate the total uncertainty in the mean fields of the basic variables and fluxes, estimates of the bias uncertainty are required. In the case where we can apply bias corrections the residual bias uncertainty will be small due to imperfections in any correction. Where we know biases to exist but not the magnitude or sign the uncertainty will be larger.

The bias uncertainty in the corrected air temperature observations has been estimated at 0.2 °C (Berry *et al*. 2004). Similarly, for specific humidity the bias uncertainty has been estimated at 0.2 g Kg $^{-1}$ based on the range of residual biases shown in Figure 4-1. The cloud cover and sea level pressure are assumed to be unbiased and to have zero bias uncertainty. Biases are known to exist in the SST observations but no correction is applied. Instead, the bias uncertainty, based on Kent and Kaplan (2006), has been estimated as the greater of 0.15°C and $0.1 \cdot \left| T_{sea} - T_{air} \right|$ °C (Elizabeth Kent, pers. comm., see also Berry and Kent 2009). The true SST value is thought to lie between the two different measurement methods and this estimate of the bias uncertainty is based on the range of the possible biases in the ERI and bucket measurements respectively. The bias uncertainty in the corrected wind speeds has been estimated at 0.2 m s^{-1} in view of the results of Thomas *et al*. (2008).

It should be noted that the estimation of the bias uncertainty is, by definition, very difficult where we don't have enough information to adjust for the bias. In these cases, we are trying to estimate the size and variability of a quantity that we do not fully understand.

4.6 Optimal Interpolation

4.6.1 Application to VOS observations

The OI scheme developed by Lorenc (1981) and implemented by Reynolds and Smith (e.g. Reynolds and Smith 1994) is used in this study to grid the VOS observations. A description of the scheme was given in section 3.4.5.

In order to apply the OI estimates of the first guess field, uncertainty in the first guess and the uncertainty in the observations are required. Estimates of the correlation terms in Eq. 3-16 are also required. The globally and temporally averaged uncertainty

estimates from the semi-variogram analysis (Section 4.5 and Chapter 5) are used for the observation uncertainty estimates (E_b). The first guess and its uncertainty can either be set to the climatological mean value, the previous analysis value, or another estimate such as model output. In this study we are interested in capturing the synoptic scale variability between the variables required for the flux calculation and also in estimating the trends and variability in the fields. We also wish the analysis to be independent from other sources. Hence, following Reynolds and Smith (1994) the previous day's analysis and analysis error have been used as the basis for our first guess and uncertainty estimates. For the initial day, climatological values are used as the first guess and first guess error.

Whilst we follow Reynolds and Smith (1994) we have far fewer observations as they include satellite data in their analysis. As a result, in data sparse regions there can be an extended period of time without data and we need to ensure we maintain the annual cycle in the fields and also account for any period without data in our uncertainty estimates. The annual cycle is maintained by incrementing the previous days analysis to allow for the annual cycle, with the daily increments given by

$$\Delta C_t = \frac{C_m - C_{m-1}}{D_m - D_{m-1}} \qquad \text{Eq. 4-2}$$

where ΔC_t is the daily increment for day t; C_m the climatological monthly mean for month m; D_m the day of year for the mid-point of month m and $D_{m-1} < t < D_m$. Hence the first guess is given by

$$P_{k,t} = A_{k,t-1} + \Delta C_t \qquad \text{Eq. 4-3}$$

where $P_{k,t}$ is the first guess for point k at time t, $A_{k,t-1}$ the analysis from the previous day and ΔC_t the increment to allow for the annual cycle. The climatological monthly means have been estimated by averaging the VOS observations onto a 1° monthly grid between 1973 and 2002, applying the same QA, bias and height adjustments as applied to the input to the OI. The monthly files have then been averaged to give a monthly climatology and any gaps filled by linearly interpolating between adjacent grid cells. Any gaps still remaining have then been filled using the zonal mean value and a 1-2-1 smoother applied to remove the effects of random noise on the scale of the gridded climatology.

The uncertainty in the first guess has been set to the previous day's analysis uncertainty plus an exponential decay model that allows the uncertainty in the first

guess to rise to a preset value after a period without any data. This preset value is set to the uncertainty that would be introduced by using the climatological mean value as the estimate of the field rather than the analysis value. This has been estimated as the sum of the mean standard deviation of the observations in each grid box and the standard deviation of the grid boxes in time added in quadrature. For grid boxes filled using the zonal mean the uncertainty has been set to the sum of mean standard deviation of the observations across the grid boxes and the standard deviation of the grid boxes used to calculate the zonal mean added in quadrature. A 1-2-1 smoother was then applied. The uncertainty in the first guess is given by

$$\left(E_t^p\right)^2 = \left(E_{t-1}^a\right)^2 + \left(\sigma_{c\,\mathrm{lim}}^2 - \left(E_{t0}^a\right)^2\right)\left(f(n)^2 - f(n-1)^2\right) \qquad \text{Eq. 4-4}$$

where E_t^p is the first guess error for the current time step; E_{t-1}^a the analysis error from the previous time step; E_{t0}^a the analysis error from the last time step with data; $\sigma_{c\,\mathrm{lim}}^2$ the uncertainty in the climatological monthly mean for the current month; *n* the number of days since the last data for the current location and *f(n)* a function allowing the decay of the uncertainty back to the climatological standard deviation. The function *f(n)* is given by

$$f(n) = 1 - \exp\left(\frac{-n}{\lambda_t}\right) \qquad \text{Eq. 4-5}$$

where λ_t is an e-folding timescale and *n* the number of days since the last observation. E_{t0}^a can be calculated from the previous day's analysis error as

$$\left(E_{t0}^a\right)^2 = \frac{\left(E_{t-1}^a\right)^2 - \sigma_{c\,\mathrm{lim}}^2 f(n-1)^2}{1 - f(n-1)^2} \qquad \text{Eq. 4-6}$$

In order to calculate the weights for the OI the correlation terms in Eq. 3-16 are required. As a first approximation, each observation in the analysis is assumed to come from a separate source and the observation to be uncorrelated. Hence the correlations between the observation errors, $<\beta_i\beta_j>$, have been set to the Kronecker delta, δ_{ij}. The observations are independent of the first guess fields, hence the correlations between the observations and first guess errors, $<\beta_i\pi_j>$, are set to zero.

The remaining correlation terms in Eq. 3-16, the correlation between the first guess errors at different locations ($<\pi_i\pi_j>$) are estimated using a relatively simple isotropic Gaussian function, i.e.

$$\left\langle \pi_i \pi_j \right\rangle = \exp\left[\frac{-\left(x_i - x_j\right)^2}{\lambda^2}\right] \qquad \text{Eq. 4-7}$$

where $(x_i - x_j)$ is the scalar distance between observations i and j and λ the e-folding space scale for the field. The e-folding timescales for both the spatial (Eq. 4-7) and temporal (Eq. 4-5) correlations have been set to 300 km and 3 days respectively. These have been chosen for pragmatic reasons, balancing the capture of the synoptic scale variability within the constraints of a limited number of observations. A discussion of the choice of length scales and correlation model is given in Section 4.9.

Once all the terms required in Eq. 3-17 are known the system of linear equations can be solved using a Cholesky decomposition and the analysis and analysis errors calculated using Eq. 3-18 and Eq. 3-19. As noted by Lorenc (1981) and Reynolds and Smith (1994) the inversion of **M** is independent of the grid point being analysed and only needs to be calculated once. Hence, for computational efficiency we follow Reynolds and Smith (1994), performing the OI analysis on 4° x 4° boxes, using all observations and centred within a larger 12° x 12° box. The bounding box has been increased from the 8° x 8° box used in Reynolds and Smith (1994) to handle the smaller size of the 1° grid boxes at higher latitudes. When the observations are too close together the matrix **M** can become singular and it is necessary to reduce the number of observations close together. This is done following Reynolds and Smith (1994), averaging all pairs of observations within a set distance of each other when **M** starts to become singular. If **M** is still singular the minimum separation distance is increased and the thinning repeated.

4.7 Flux Calculation and Uncertainty Estimates

As discussed in Chapter 2, the turbulent sensible and latent heat flux are calculated using the bulk formulae. The parameterisation of Smith (1980) has been used for the drag coefficient and Smith (1988) for the heat and moisture transfer coefficients. Following previous authors, the humidity at the sea surface has been calculated at the SST, assuming saturation, and then multiplied by a factor of 0.98 to account for the reduction in vapour pressure over a saline surface (Kraus and Businger 1994). At low wind speeds (below 0.5 ms^{-1}) the iterative flux calculation becomes numerically unstable. A lower limit of 0.5 ms^{-1} has been used to avoid this. Propagation of errors (see Section 3.6.4) has been used to estimate the uncertainty in the daily flux estimates.

As a first approximation, the terms in Eq. 3-28 due to uncertainties in the density and specific heat capacity of air and latent heat of vapourisation have been assumed to be negligible. The correlation terms have also been assumed to be negligible as a first approximation. This should be a reasonable assumption for the random and bias errors in the temperature, humidity and wind speed terms since these are not expected to be correlated across variables. However, there will be some correlation in the sampling errors for these variables due to the VOS making co-located measurements of the observations. However, it is possible for the correlation to either increase or decrease the uncertainty estimates, and further work is required to quantify these terms. The correlation between the sampling errors in the different variables have therefore been set to zero. Uncertainties in the transfer coefficient scheme have not been included in the calculation. The uncertainty in the fluxes then becomes

$$\sigma_H^2 = \left(\frac{\partial H}{\partial u_z}\right)^2 \sigma_{u_z}^2 + \left(\frac{\partial H}{\partial T_z}\right)^2 \sigma_{T_z}^2 + \left(\frac{\partial H}{\partial T_0}\right)^2 \sigma_{T_0}^2 \qquad \text{Eq. 4-8a}$$

$$\sigma_E^2 = \left(\frac{\partial E}{\partial u_z}\right)^2 \sigma_{u_z}^2 + \left(\frac{\partial E}{\partial q_z}\right)^2 \sigma_{q_z}^2 + \left(\frac{\partial E}{\partial q_0}\right)^2 \sigma_{q_0}^2 \qquad \text{Eq. 4-8b}$$

where the sigma terms are the uncertainty in the respective basic meteorological variables given by the subscripts.

In this study the uncertainty due to bias uncertainty has been evaluated separately from that due to random and sampling errors. The total uncertainty in each flux is then given by evaluating Eq. 4-8 separately for the uncertainty in the output from the OI and for the bias uncertainty terms from Section 4.5.2. The uncertainty due to random and sampling errors and due to bias uncertainty are then added in quadrature (Gleckler and Weare 1997) to give the total uncertainty in the daily flux estimates.

4.8 Calculating the Monthly Mean Uncertainty

Estimates of the monthly mean fields and fluxes have been calculated by averaging the daily values with each day given equal weight. However, the individual daily fields from the OI will be correlated with each other due to the use of the previous days analysis in the first guess. The degree of correlation will depend on the amount of data available for each day's analysis and needs to be taken into account when calculating the uncertainty in monthly mean estimates based on the daily fields. By treating each analysis value as a separate variable, propagation of errors (Eq. 3-28) can

be used to estimate the uncertainty in the average of multiple analysis values. The uncertainty terms, σ, are now the uncertainty in the individual daily analyses. If each day is given equal weight the derivatives ($\partial V/\partial v$) are simply 1 divided by the number of days being averaged. The correlation between variables (days) can be estimated from the output of the OI. It should be noted that the uncertainty due to the bias errors will not reduce on averaging.

From the OI, Eq. 3-11, the daily analysis at single point is given by

$$\frac{A_k}{E_k^p} = \sum_{i=1}^{N} w_{ik} \frac{\left(B_i - P_i\right)}{E_i^p} + \frac{P_k}{E_k^p}$$

where A_k is our analysis field at the point of interest k; P_k the first guess value at point k; w_{ik} the weight for observation i; B_i the ith observation; P_i the value of the first guess at the location of observation i and E^p the uncertainty in the first guess at the respective locations. As a first approximation, assuming the differences between the first guess values at the different locations and between the first guess errors are small we can multiply by the uncertainty values and rewrite Eq. 3-11 for point k as

$$A_k = P + \sum_{i=1}^{N} w_i\left(B_i - P\right) = \left(1 - \sum_{i=1}^{N} w_i\right)P + \sum_{i=1}^{N} w_i B_i \qquad \text{Eq. 4-9}$$

where $P \approx P_k \approx P_i$. The first term on the RHS gives the contribution from the first guess to the current analysis and the second term is the contribution from the observations for the current day. The first guess terms, P, are given by the analysis from the previous days incremented to allow for the annual cycle as the first guess. Hence, for day t we can rewrite Eq. 4-9 as

$$A_t = \left(1 - \sum_{i=1}^{n_t} w_{t,i}\right)\left(A_{t-1} + \Delta C_t\right) + \sum_{i=1}^{n_t} w_{t,i} B_{t,i} \qquad \text{Eq. 4-10}$$

where ΔC_t is the daily increment for day t (Eq. 4-2). The increments to allow for the annual cycle will be small for any given day. Assuming the increments are negligible we can rewrite Eq. 4-10 as

$$A_t = \rho_{t,t-1} A_{t-1} + \beta_t$$

where

$$\rho_{t,t-1} = \left(1 - \sum_{i=1}^{n_t} w_{t,i}\right) \qquad \text{Eq. 4-11a}$$

$$\beta_t = \sum_{i=1}^{n_t} w_{t,i} B_{t,i} \qquad \text{Eq. 4-11b}$$

i.e. β_t is now the contribution from the observations for the current day and $\rho_{t,t-1}$ gives the correlation between analyses one day apart due to the analysis process and excluding any natural autocorrelation. The correlation between analyses several days apart will then be given by the product of the correlations of the analyses one day apart. For example, the correlation between the fifth and third day would be given by

$$\rho_{5,3} = \rho_{5,4}\rho_{4,3} = \prod_{t=4}^{5} \rho_{t,t-1} \qquad \text{Eq. 4-12}$$

This correlation, imposed by the method, will also give the correlation between the errors in the analysis. Representing each day as a separate variable and using Eq. 3-28 the uncertainty in the monthly mean due to random uncertainty in the daily values is then given by

$$\sigma^2_{random} = \sum_{i=1}^{N} \left(\frac{1}{N}\right)^2 \sigma^2_{r,i} + 2\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left[\left(\frac{1}{N}\right)^2 \sigma_{r,i}\sigma_{r,j} \left(\prod_{k=i+1}^{j} \rho_{k,k-1} \right) \right] \qquad \text{Eq. 4-13}$$

where σ_{random} is the uncertainty in the mean of the analyses due to random and sampling errors, $\sigma_{r,i}$ the uncertainty in the analysis for day *i*, and $\rho_{k,k-1}$ the correlation between analyses for days *k* and *k-1* calculated using Eq. 4-11 and Eq. 4-12. The uncertainty in the monthly mean fluxes has been estimated in a similar manner. In order to estimate the correlation between the daily flux estimates, Eq. 4-11 and Eq. 4-12 have been used, with the minimum sum of weights from the basic variables used to calculate the fluxes each day. This will act as an upper limit for the uncertainty in the fluxes.

In addition to sampling and random errors in the OI fields we also have uncertainty due to systematic biases, either due to uncertainties in the bias corrections applied, or where we know potential biases exist in the observations but are not able to apply a bias correction. These uncertainties are assumed to be constant in time and correlated across the daily analysis fields. As a result these will not be reduced by averaging across days. The total uncertainty in our monthly fields is then given by

$$\sigma^2_{monthly} = \sigma^2_{bias} + \sigma^2_{random} \qquad \text{Eq. 4-14}$$

Where $\sigma_{monthly}$ is the uncertainty in the monthly mean field, σ_{bias} the estimate of the bias uncertainty in the monthly mean fields and σ_{random} the uncertainty due to random and sampling errors in the monthly mean fields.

4.9 Discussion

4.9.1 Quality Control

The VOS observations have been quality controlled by discarding observations more than 4.5σ from the climatological median value using the trimming flags included in ICOADS. Whilst the ICOADS trimming flags are adequate for quality controlling the data there is scope for improvement such as the use of a nearest neighbour or buddy check for each observation. Alternatively, the individual observations could be compared to the dataset produced as part of this thesis, taking into account the uncertainties in both the observations and gridded fields, and outliers flagged. The use of alternative methods of QC will be explored in future versions of the dataset.

4.9.2 Bias Corrections

4.9.2.1 Air Temperature

The bias correction applied to the air temperature observations depends on the fit of the parameters in the bias correction. In this thesis these parameters have been determined each year by fitting the correction to bias estimates for a subset of the observations used. The biases have been estimated as the difference between the observed value and a night-time only estimate. Whilst this risks removing the true diurnal cycle for the observed air temperatures this is preferable to the other options. Leaving the observations uncorrected would result in air temperature estimates with significant warm biases and biased flux estimates as a result. Using night-time only observations would result in significantly reduced data quantities and result in higher random errors. In Chapter 6 it will be shown that the bias correction does not over-correct and remove the diurnal cycle.

4.9.2.2 Specific humidity

Whilst the bias correction applied to the humidity observations made by the VOS using marine screens is not perfect it is an improvement on the previous versions (as seen in Figure 4-1). The effects of the correction will be shown to improve the humidity estimates in the new dataset compared to uncorrected values (Chapters 5 and 6). Future improvements to the bias correction should be possible and are discussed in Chapter 8.

4.9.3 Random Error Estimation

Improvements have been made to the method used to estimate the random errors for the individual observations and the results will be shown in the next Chapter. Further improvements are still possible, but it is unclear what gains will be made in terms of estimating the random errors in the VOS observations. Previous studies using variograms and other estimates have shown similar values to each other and the results shown in Chapter 5 are also similar to these previous studies.

The uncertainty estimates for the individual VOS observations vary spatially and temporally, but this variation is relatively small (see Chapter 5). Due to this small variability, and in order to simplify understanding of the uncertainty estimates from the output of the OI, a constant (temporal and spatial) global value has been used for each variable.

4.9.4 Optimal Interpolation and Error Correlations

The OI applied to the VOS observations requires knowledge or estimates of the correlation between the errors in the first guess field and the correlation between the errors in the observations. The errors in the observations will contain contributions from temporal sampling errors, random errors and biases. The biases will be composed of biases in the individual instruments used to make the measurements, i.e. calibration errors, and biases due to environmental conditions such as the radiative heating errors in the air temperature observations. The calibration errors should be independent between ships and normally distributed. As a result, in this thesis these errors have been included as part of the random error estimates and assumed to be uncorrelated. However, the VOS make several observations over the course of a day and, as a result, the OI for a given location is likely to contain multiple observations from the same ship. Hence, the random error will contain some correlation between observations. This effect is likely to be small compared to the contributions from the true random and sampling errors.

The errors in the first guess field, and their correlation, will be a function of the method used to make the first guess. By using the previous days analysis as the first guess the errors will be strongly dependent on the available observations used in the previous days analysis and the autocorrelation of the field itself. This will include the spatial structure of the autocorrelation, i.e. how similar the fields are from one day to

the next and how spatially coherent this similarity is. As a result, the true correlation length scales will vary with both region and sampling, with shorter length scales in high variability regions, such as the Gulf Stream, and longer length scales in the lower variability regions. Whilst the length scales are expected to vary regionally a relatively simple first approximation is used in this thesis, with a Gaussian error correlation model used and with fixed length scales of 300 km. The impact of this choice on the uncertainty estimates and the mean fields is examined and discussed in Chapter 6 and possible improvements discussed in Chapter 8.

4.10 Summary

In this Chapter the outline of the methodology developed to routinely estimate both the fluxes and their uncertainties from the VOS observations has been described. The reasons behind the different choices made are also given. The methodology developed includes basic pre-processing of the observations used such as the quality control and bias correction of the observations. After the pre-processing a semi-variogram analysis is performed to estimate the uncertainty in the individual observations due to random errors and temporal sampling errors on a daily time scale. The observations are then gridded using OI to take into account the spatial distribution of the observations and the uncertainty in those observations. The fluxes and their uncertainty are calculated from the output of the OI on a daily timescale. These daily estimates are then combined to give monthly mean flux fields and the uncertainty in the monthly mean values.

5 Results: Initial Processing and Flux Estimates

5.1 Introduction

This Chapter presents the results of the quality control, bias adjustments and random error estimation together with a qualitative examination of the output from the optimal interpolation. The impact of the quality control and track checking applied to the observations on the amount of data available is described in Section 5.2. Section 5.3 presents the results of the bias adjustment for air temperature heating errors. The results of the random error estimation is described in Section 5.4 including the impact of the height and bias adjustments on the random error estimates. A qualitative examination of the output from the OI and the fluxes is given in Section 5.5 whilst Section 5.6 gives a summary of the results presented.

5.2 Quality control and track checking

Figure 5-1 shows the number of observations per month made over the Atlantic before (black) and after (red) the quality control is applied. Also shown are the number of observations per year for each variable after both the track checking and QC have been applied (green). For each variable, excluding humidity, there are between 1 and 1.2 million observations per year over the period 1970 – 1990. After 1990 the number of observations decreases steadily, reducing to between 400k and 600k observations per year by 2006. For humidity, there are initially around 900k observations per year during the early 1970s. However, as with the other variables, the number decreases steadily over time to around 500k observations per year in 2006. The effects of the quality control and track checking on the number of observations available is small (Table 5-1).

Table 5-1: Percentage of observations removed by the quality control (QC) and track checking for each variable averaged over 1970 – 2006.

Variable	QC	QC + Track Check
Air Temperature	1.3	3.3
Sea Surface Temperature	2.5	4.4
Specific Humidity	2.2	4.0
Wind Speed	3.7	5.7
Sea Level Pressure	0.8	2.9

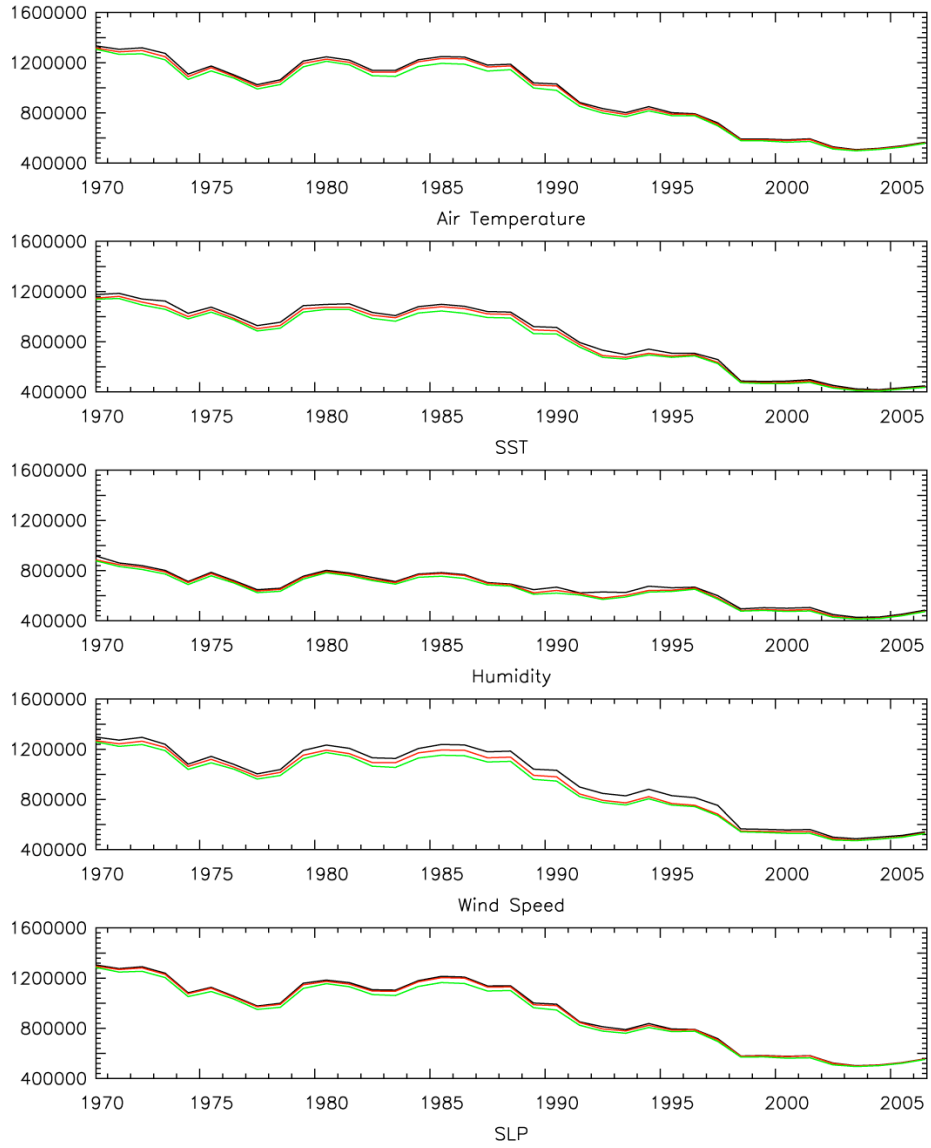

Figure 5-1: Number of observations for each variable per year before (black) and after quality control (red). Also shown are the number of observations per year after QC and track checking (green).

5.3 Air Temperature Bias Estimates and Fitting the Bias Adjustment

5.3.1 Heating Error Bias Estimates

In order to highlight the diurnal cycle in the bias estimates the estimates have been binned against local solar time. Examples of the bias estimates (black squares) during January and July for two sample years (1995 and 2005) are shown in Figure 5-2. All observations in the Northern Hemisphere that pass the QC and track checks have been used. A strong diurnal warming is visible, peaking around 1400h local time before

decaying into the evening. The warming is asymmetric about midday, i.e. the errors are larger in the afternoon compared to the same time before midday, and the warming greater in July than in January. A peak warming of almost 2 °C occurs in July and between 0.5 °C and 1°C in January. Similar results are seen for the other years (not shown). As noted in Chapter 4 this procedure removes the true diurnal cycle and will therefore overestimate the heating errors and daytime observations may be overcorrected as a result. However, the resulting bias will be small compared to the heating errors. In the next Chapter (Section 6.3.3.1) a comparison of the OI fields to buoy observations shows that this overcorrection is small.

5.3.2 Bias Adjustment Coefficients

The results of fitting the coefficients used in the bias adjustment are generally consistent across the top few runs from the Monte Carlo approach. As an example, Figure 5-3 shows the diurnal cycle of the bias correction calculated using the best fit (solid line) and fixed environmental conditions. Also shown are the mean values (diamonds) and standard deviation (error bars) of the bias correction calculated using each of the sets of parameters for the top 10 runs. There is little difference between the different fits and the standard deviations are generally small (< 0.1°C). Similar results are seen for the other years and coefficients for each year are listed in Appendix C.

As an example of the impact of the bias correction, Figure 5-2 shows the residual biases (red squares) after the bias correction has been applied for the labelled months and years. The grey shaded bar indicates the estimated uncertainty in the bias correction (± 0.2 °C). From Figure 5-2, it can be seen that the correction removes the strong diurnal signal seen in the estimated biases with residual biases close to 0 °C over the full 24 hour period and generally within the bias uncertainty estimate. The correction also models the larger warming in the afternoon due to the heat storage by the ships, a key component missing from the previous correction of Kent *et al*. (1993a).

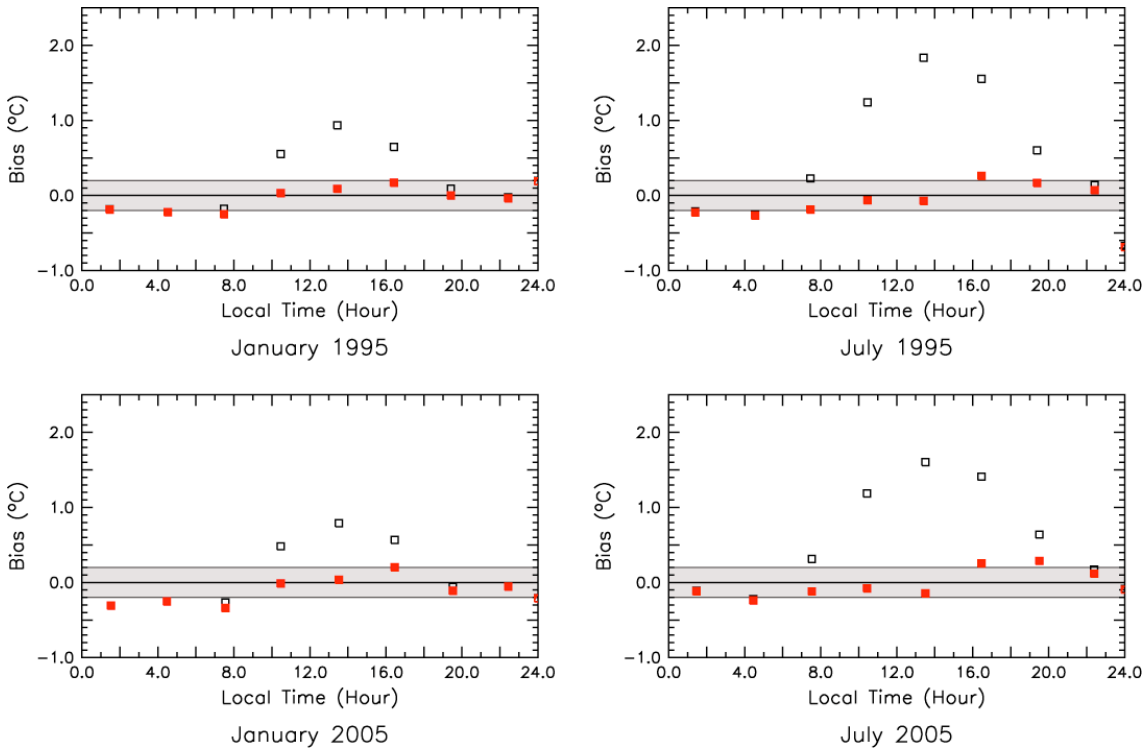

Figure 5-2: Diurnal cycle of estimated bias in air temperature observations (black squares) and residual bias (red squares) after applying the new bias correction. The shaded region indicates the bias uncertainty.

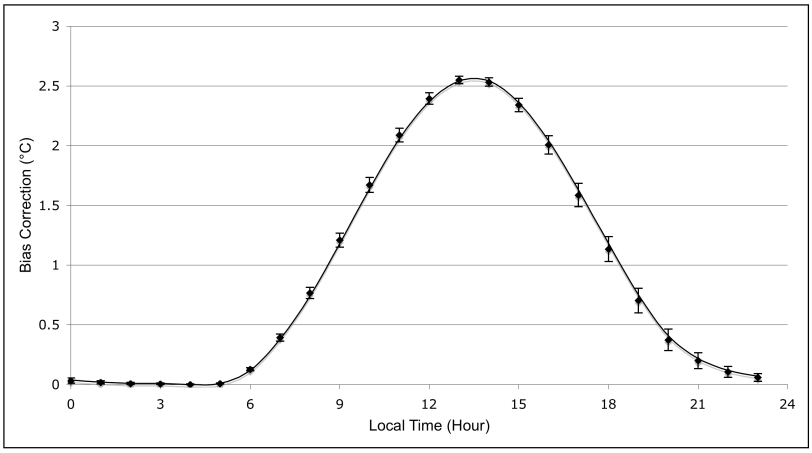

Figure 5-3: Diurnal cycle of the bias correction using the parameters of the best fit from 1995 (solid line) and fixed environmental conditions of: relative wind speed = 7 ms^{-1}, cloud cover = 3 oktas; latitude = 40 °N. Also shown is the average bias correction from the top 10 runs (diamonds) calculated using the same environmental conditions. The error bars give the standard deviation correction from the top 10 fits.

5.4 Random Errors

5.4.1 Example semi-variograms and choice of variogram model

As an example of the variogram estimation, Figure 5-4a (left panel) shows the sample semi-variograms for air temperature observations during July 1993 over the North Atlantic (black squares and line) and for different 30° regions. Also shown are the results of fitting three different variogram models. The transformed models shown (i.e. $\sqrt{0.5 \cdot \gamma(h)}$ where $\gamma(h)$ is the variogram model) are: linear model (red); Gaussian model (green); and exponential model (blue). Figure 5-4b (right panel) shows the residuals for the different models as a function of separation distance. Over the Southern and Eastern Atlantic there is little to distinguish between the different models, with each accurately representing the sample semi-variograms. This can be seen in the residuals, with values close to zero and no significant auto-correlation in the residuals. Performing an F-test on the residuals shows no improvement through using a Gaussian model compared to the linear model ($0.873 \leq F \leq 1.046$, df = 15,14) and a Durbin-Watson test confirms the lack of auto-correlation at lag 1 (i.e. between 30 km separation bins) in the residuals for all three models.

In the Central and North Western Atlantic and over the Caribbean Sea (i.e. 4 panels in the top left of Figure 5-4) there are significant differences between the models. Both the linear and exponential models underestimate the sample semi-variogram at short separation distances and over estimate at mid separations. This is clearly seen in the residuals with significant autocorrelation at the 95% confidence level ($d < 1.106$ for the linear model and $d < 0.982$ for the exponential model). In contrast, the Gaussian model accurately models the sample semi-variogram over all separation distances with residuals close to zero and no significant auto-correlation. An F-test confirms the improvement through the use of the Gaussian model compared to the linear and exponential models at the 95% confidence level ($F > 3.8$; df = 15,14).

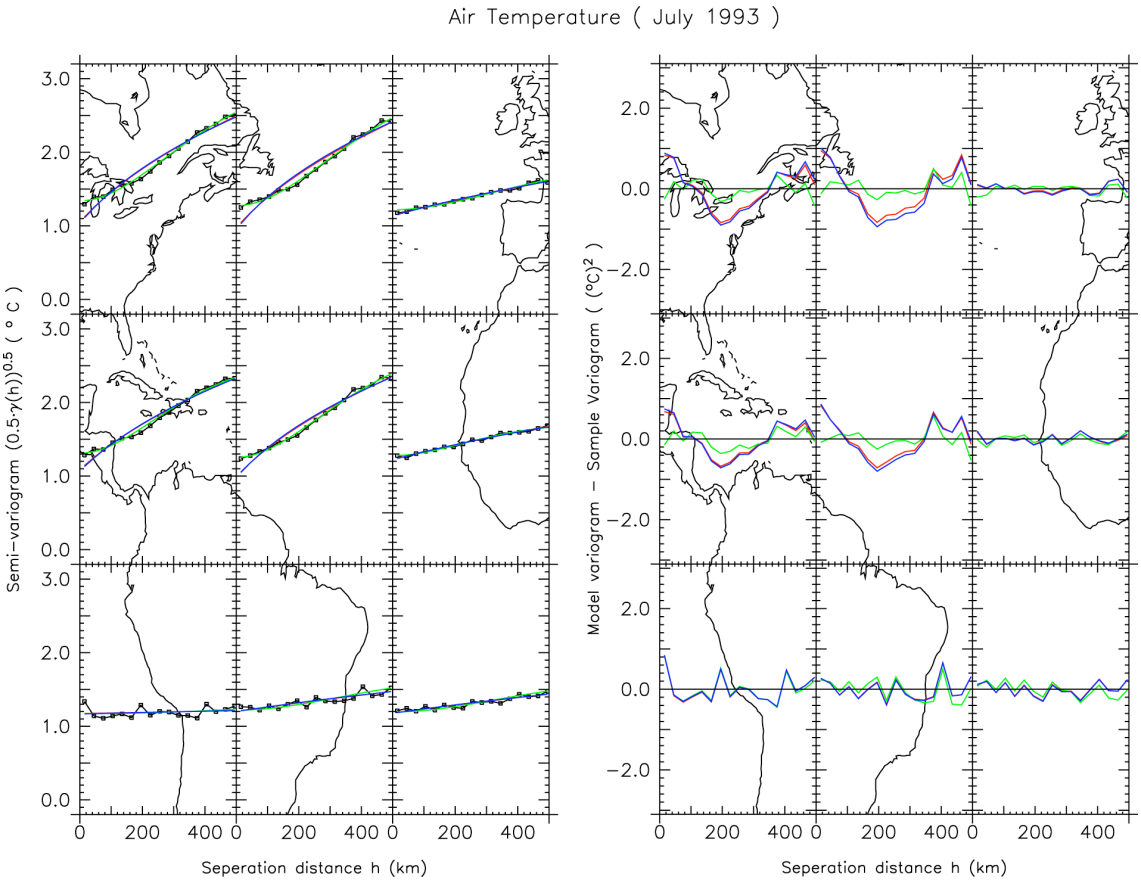

Figure 5-4: a) example semi-variograms (left) and b) difference (residual) between the modelled and observed variograms (right) over the North Atlantic during July 1993 for air temperature. The models and residuals shown are the linear (red), exponential (blue) and Gaussian (green) models. The sample semi-variogram (black) is also shown in the left hand plot.

The Gaussian model performs well for all variables except wind speed where the exponential and linear models perform better. This is shown in Figure 5-5. Over the majority of the region shown the three models perform similarly, however, over the northern and north-eastern regions (top centre and top right panels) the Gaussian model performs poorly and the linear model gives the best results.

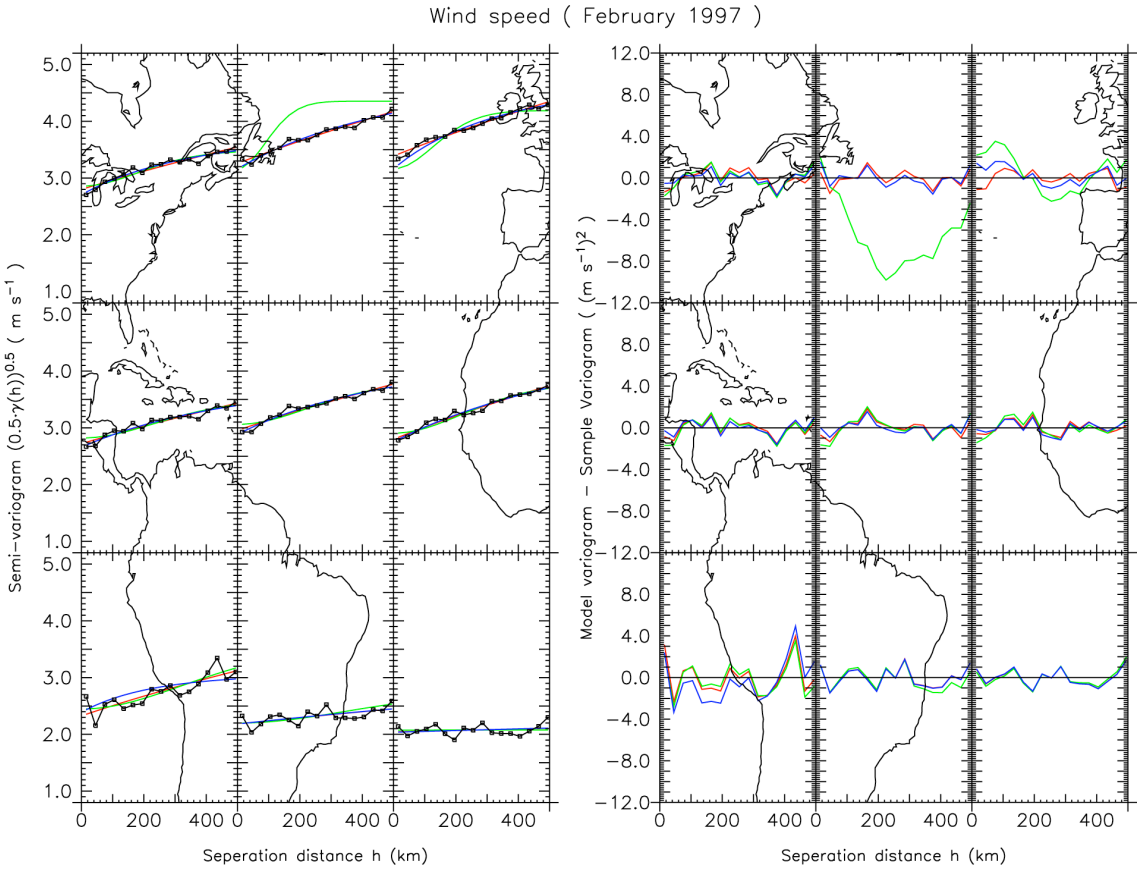

Figure 5-5: As Figure 5-4 but for wind speed observations and February 1997.

From the examples shown in Figure 5-4 and Figure 5-5 it can be seen that the model giving the more accurate uncertainty estimate varies both with region and with variable. Initially, and for consistency, we want to apply a single model to each variable and need a systematic way to determine which model gives the best estimate of the random errors. This has been done by fitting the different variogram models to the sample variograms for each variable, 30 ° grid cell and month between 1970 and 2006. Any month or grid cell failing a Durbin-Watson test for autocorrelation is discarded and the number of grid cells and months with a valid fit to the sample variogram counted for each model. The mean Pearson correlation coefficient (R^2) from the regressions is also calculated for each model. The model giving the largest number of fits and the largest mean R^2 value for a given variable is then chosen to estimate the random errors for that variable. For variables where two different models are chosen based on number of valid fits and based on the mean R^2 value the simpler of the models is used.

A summary of the model selection for the different variables is given in Table 5-2, with the mean random error estimate (i.e. the square root of half the intercept), the number of grid cells with valid fits and the mean R^2 coefficient. From Table 5-2 it can

be seen that there is little to differentiate between the variogram models on a global scale for the different fields in terms of the R^2 values, with R^2 values usually above 0.97 (significant at the 1 % level). However, there are significant differences in the number of grid cells with valid fits. For example, the Gaussian model has almost 5500, or 40%, more grid cells with valid fits than the linear model for air temperature. The model selected to estimate the random errors for each field is highlighted in bold.

Table 5-2: Summary of fitting statistics for the different variogram models. The mean random error estimates, number of months and grid cells with valid fits and mean r^2 are given. The uncertainty on the mean random errors is the standard deviation of the error estimates over all valid grid cells. The rows in bold indicate the model chosen for each field.

Field	Model	Mean Random Error	Number of good fits	Mean R^2
Air temperature (°C)	Linear	1.13 ± 0.21	13649	0.977
	Exponential	1.09 ± 0.22	14161	0.977
	Gaussian	**1.25 ± 0.21**	**19112**	**0.983**
Specific humidity (g Kg^{-1})	Linear	1.23 ± 0.27	15712	0.975
	Exponential	1.20 ± 0.27	16213	0.975
	Gaussian	**1.29 ± 0.24**	**19165**	**0.979**
Sea level pressure (hPa)	Linear	2.03 ± 0.65	14129	0.966
	Exponential	1.92 ± 0.57	13722	0.964
	Gaussian	**2.10 ± 0.53**	**16641**	**0.969**
SST (°C)	Linear	1.12 ± 0.27	13055	0.977
	Exponential	1.08 ± 0.27	13923	0.978
	Gaussian	**1.24 ± 0.27**	**17576**	**0.982**
Scalar wind speed (m/s)	Linear	2.38 ± 0.42	17049	0.982
	Exponential	**2.30 ± 0.43**	**18857**	**0.983**
	Gaussian	2.41 ± 0.39	18123	0.982

5.4.2 Impact of Bias and Height Corrections

As noted in Chapter 4, if a bias adjustment improves the consistency of the observations the magnitude of the random error estimates will reduce. In this section the impact of the bias and height adjustments described in Chapter 4 are shown. Height and corrections are applied to the air temperature, wind speed and humidity observations. Corrections for systematic biases are also applied to the air temperature, humidity and wind speed observations. None of the other fields are bias or height corrected.

5.4.2.1 Air Temperature

Figure 5-6 shows the impact of the bias and height corrections on the sample variogram for the 30° grid box centred on 45°N 45°W for July 1993. As expected the height corrected observations (red) have smaller squared differences compared to the unadjusted observations (black) over all separation distances. Similarly, bias correcting and then height correcting the observations (green) reduces the squared differences further. It should be noted that the same observations have gone into each variogram estimate.

Figure 5-7 show a time series of the random errors estimated for the air temperature over the Atlantic region (40 S – 70 N and 100W – 30E). The uncorrected data is shown in black, the height corrected in red and bias and then height corrected in green. The impact of the bias and height corrections is clearly visible, with the estimated random errors reduced with each adjustment or correction applied. The average value for the unadjusted observations is 1.27 °C, 1.20 °C for the height corrected data and 1.10 °C for the bias and then height corrected data. This reduction in the random error estimates gives confidence in the adjustments and that the data are improved by applying the corrections.

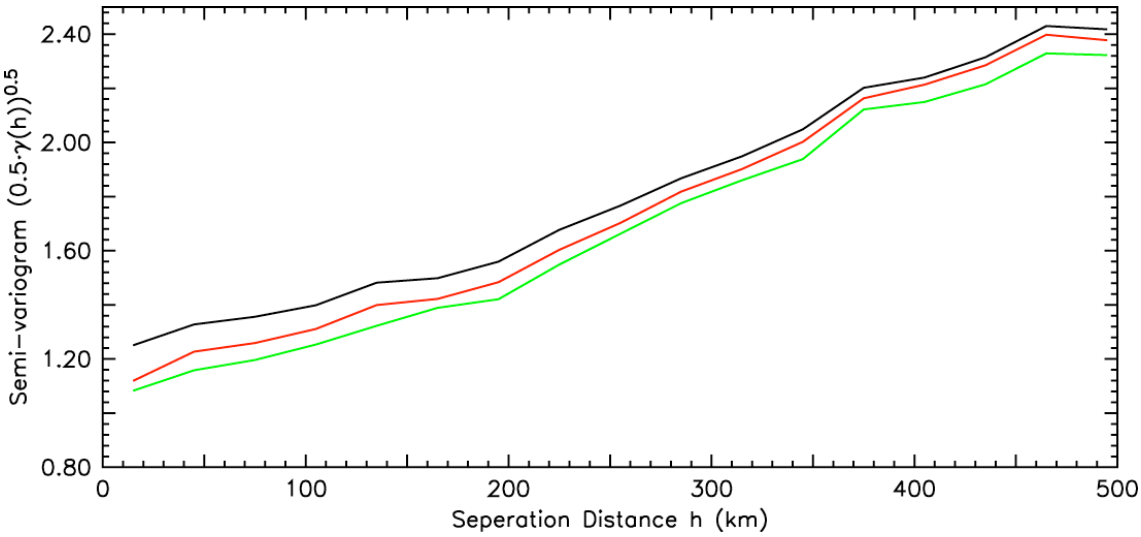

Figure 5-6: Sample semi-variograms for unadjusted (black), height corrected (red) and bias and height adjusted (green) air temperature observations during July 1993 and between 30N – 60N and 60W – 30W.

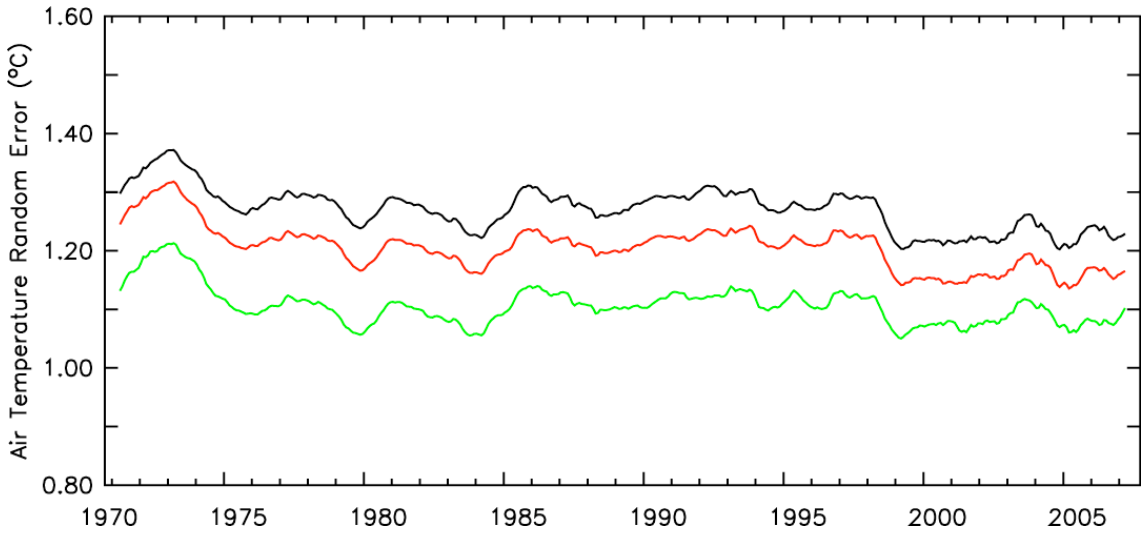

Figure 5-7: Time series of random error estimates for unadjusted (black), height adjusted (red) and bias and height adjusted (green) air temperature observations made in the region 40S – 70N and 100W – 30E. A 12 month running mean filter has been applied.

5.4.2.2 Specific Humidity

Figure 5-8 shows the time series of random error estimates for the specific humidity unadjusted (black), height corrected (red) and bias and then height corrected (green) over the Atlantic region. As with the air temperature, both adjustments act to reduce the random errors. The height adjustment produces the largest decrease in the random errors, the fully adjusted having only slightly lower random errors than the height adjusted data over the majority of the period. The average random errors over the 30 year period are 1.27 g Kg^{-1}, 1.25 g Kg^{-1} and 1.24 g Kg^{-1} for the unadjusted, height adjusted and fully adjusted data respectively. The improvement in the consistency of the humidity observations through the application of the bias adjustment is less obvious for humidity compared to the air temperature observations.

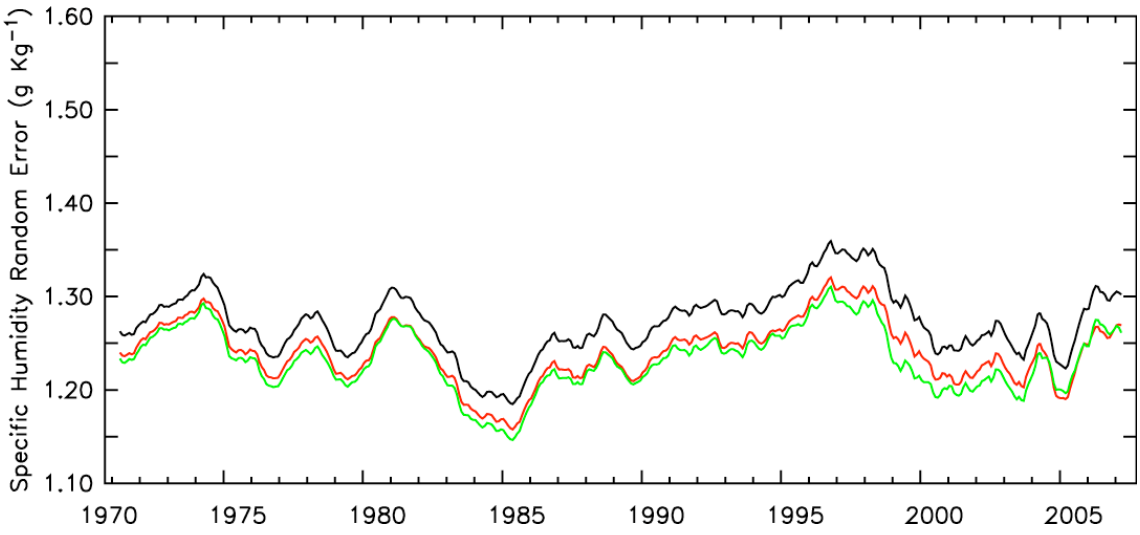

Figure 5-8: As Figure 5-7 but for specific humidity.

5.4.2.3 Wind speed

Figure 5-9 shows a time series of the random error estimates for the unadjusted (black) and height adjusted (red) wind speeds over the Atlantic region. The errors have been estimated using the exponential variogram model. Again, the height correction reduces the random error estimates giving confidence in the adjustments applied. The average values of the random errors for the unadjusted and height adjusted wind speeds are 2.31 ms^{-1} and 2.17 ms^{-1} respectively. The slight upward trend in random errors for unadjusted data is reduced by the adjustment, suggesting that the wind speed observing heights are becoming more variable over time.

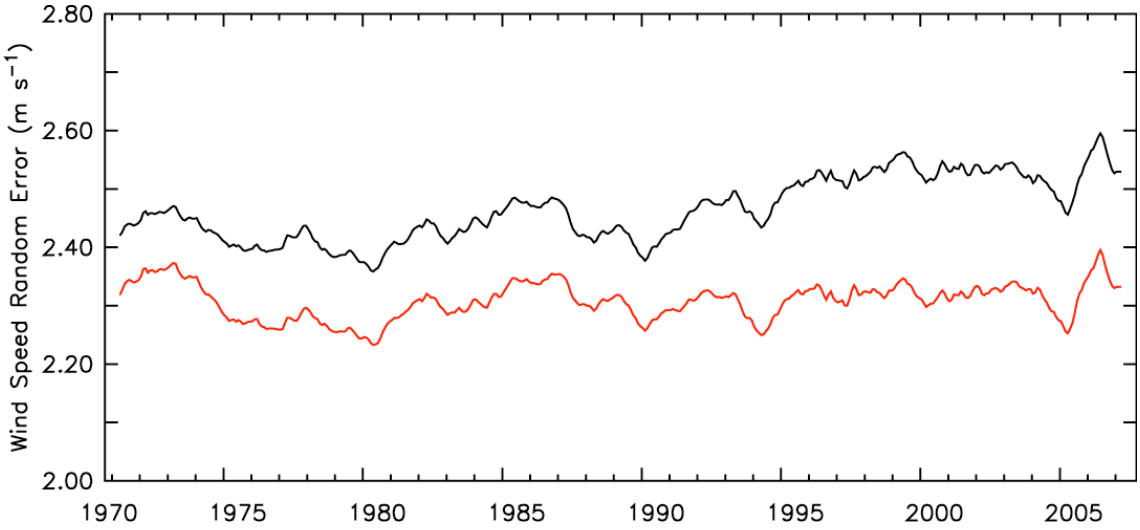

Figure 5-9: As Figure 5-7 but for wind speed observations.

5.4.3 Global Error Estimates

As noted in Chapter 4, different random error estimates can be assigned to each observation in the OI with the errors varying spatially and temporally. However, in order to simplify the understanding of the output from the OI a single global value is used for each variable. These have been estimated by averaging the different random error estimates globally and in time with each grid cell weighted by ocean area. Any grid cell for which the fitted model does not show a significant correlation with the sample variogram or which exhibits significant autocorrelation is discarded. These globally averaged estimates are shown in Figure 5-10 after a 12 month running mean filter has been applied. Table 5-3 lists the globally and temporally averaged values before and after the different height and bias adjustments. The values used in the OI are given in bold. Whilst a random error estimate is listed for the height and bias adjusted wind speed this was calculated after the OI had been run as the bias adjustments applied to the visual wind speed estimates were finalised at a late stage. Calculation of the random error estimates is computationally intensive and revising the estimate would have significantly delayed dataset production. In view of the modest reduction to the error estimate as a result of the bias adjustment (Table 5.3), this was not a major compromise. Similarly, use of the global averages, in view of the modest standard deviations, is not a major compromise.

Table 5-3: Globally averaged random error estimates (± 1 standard deviation) for the each variable. The random errors have been averaged over 1970 – 2006. The values listed for air temperature, humidity and wind speed from Kent and Berry (2005) are the height adjusted values with the 3.5σ trimming limits applied.

Field	Model	Unadjusted	Height Corrected	Height and Bias Corrected	Kent and Berry (2005)
Air Temperature (°C)	Gaussian	1.25 ± 0.21	1.17 ± 0.20	**1.07 ± 0.21**	1.0 ± 0.1
Specific Humidity (g Kg^{-1})	Gaussian	1.29 ± 0.24	1.26 ± 0.23	**1.25 ± 0.23**	1.1 ± 0.1
SST (°C)	Gaussian	**1.24 ± 0.27**	-	-	1.2 ± 0.1
SLP (hPa)	Gaussian	**2.10 ± 0.53**	-	-	2.1 ± 0.2
Wind speed (ms^{-1})	Exponential	2.30 ± 0.43	**2.15 ± 0.40**	2.09 ± 0.39	2.0 ± 0.1

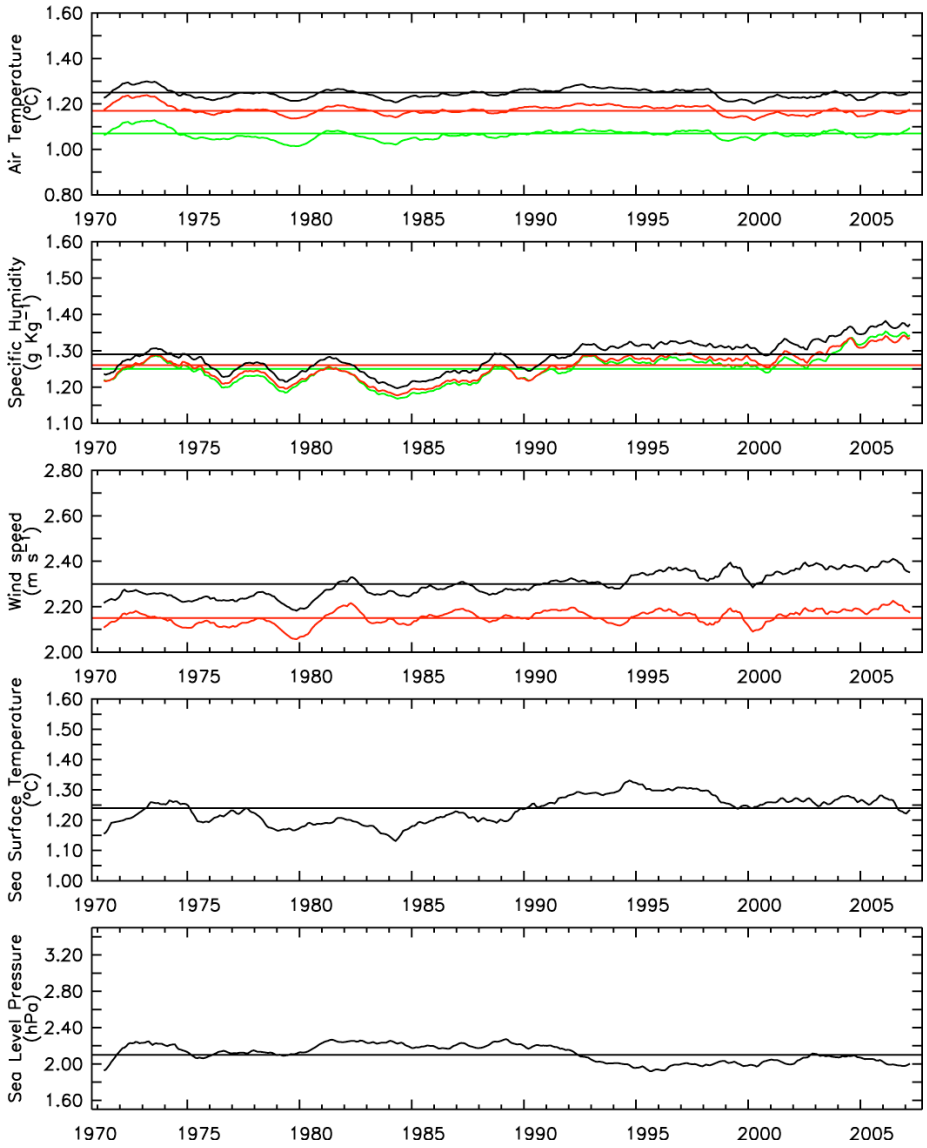

Figure 5-10: Time series of random error estimates averaged globally for the basic variables with 12 month running mean filter applied. Estimates for unadjusted (black); height adjusted (red) and bias and height adjusted (green) observations are shown. The horizontal lines show the mean values.

5.5 Optimal Interpolation and Flux Calculation

5.5.1 Mean values

Examples of the estimates of the basic variables and fluxes from the OI are shown in Figure 5-11 and Figure 5-12. Figure 5-11 shows the daily fields averaged over the 1992 / 1993 winter season whilst Figure 5-12 shows the fields averaged over the

summer months during 1993. Also shown are the mean daily uncertainty estimates from the OI.

The sea surface temperature, air temperature and humidity all show similar spatial patterns, with highest values in the tropics and lowest values in high latitudes as expected. There are also East – West gradients, increasing (decreasing) from East to West in the tropics (high latitudes). The winds are highest in the North-Western Atlantic and trade wind belts and lowest in the centres of the subtropical gyres. The sea level pressure shows the Azores high – Iceland low dipole whilst the heat fluxes out of the oceans are greatest over the Gulf Stream in the western Atlantic in the northern winter and around 10 °S in the northern summer.

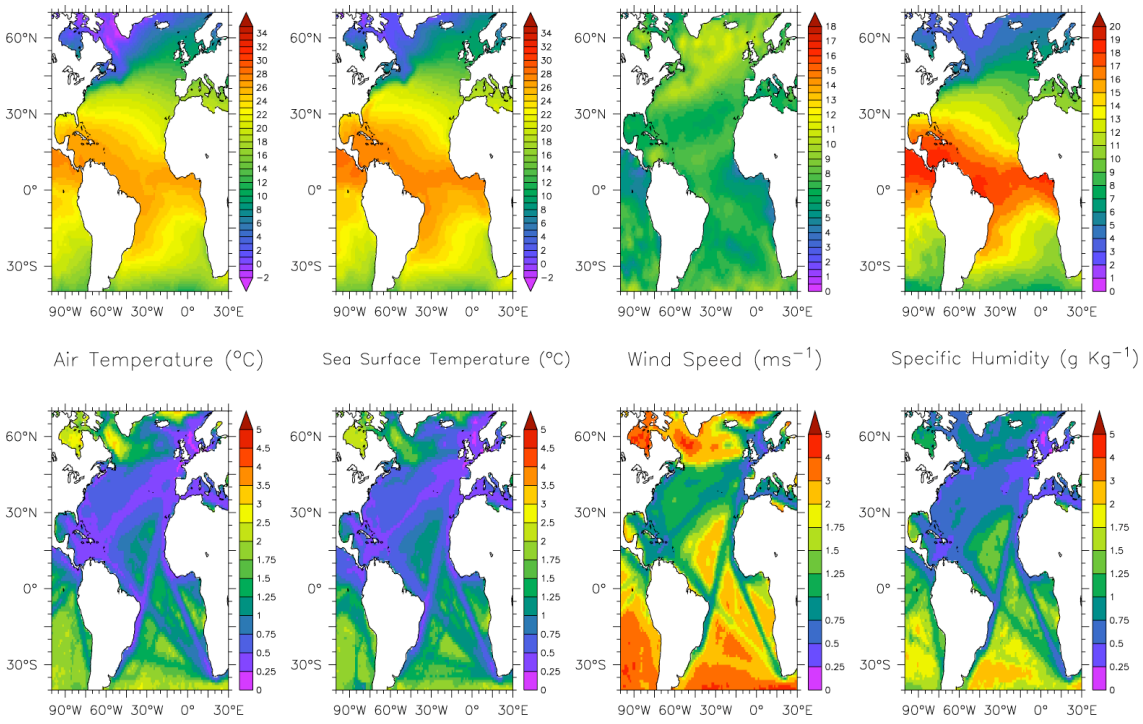

Figure 5-11a: Mean daily values (top) averaged over the period December 1992 – February 1993 for air temperature, sea surface temperature, wind speed and specific humidity. Also shown is the mean daily uncertainty for the different fields (bottom).

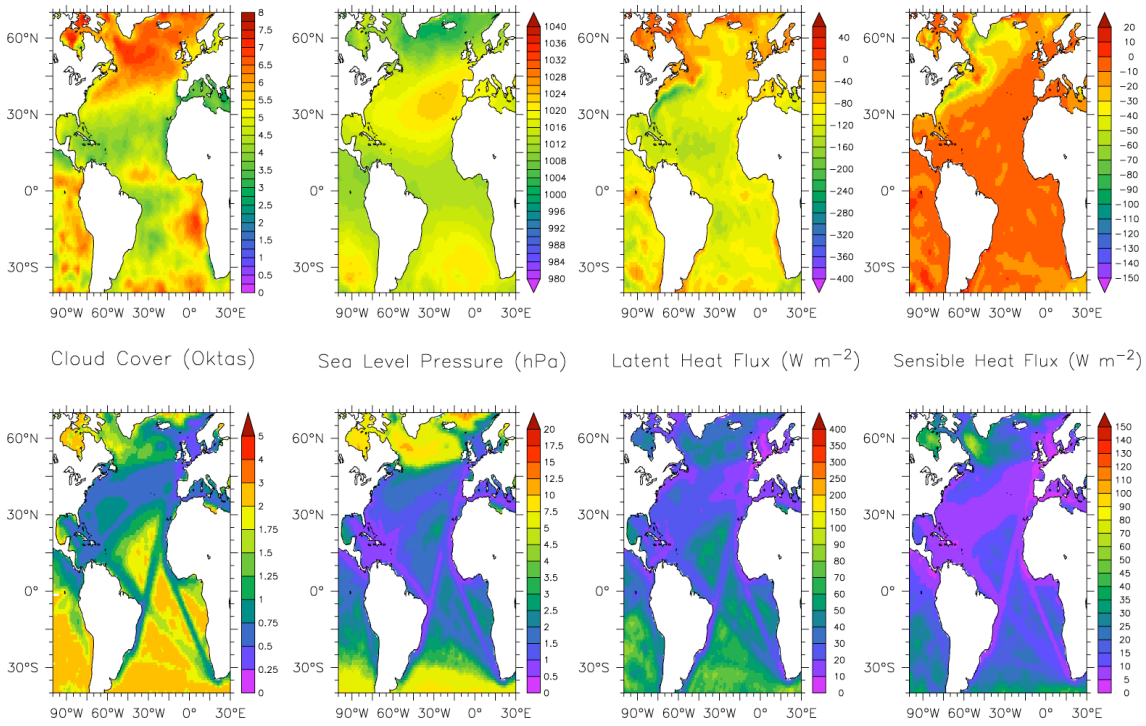

Figure 5-11b: as Figure 5-11a but for cloud cover, sea level pressure, latent and sensible heat fluxes. Positive (negative) fluxes indicate a heat gain (loss) by the ocean.

5.5.2 Random and Sampling Uncertainty

The uncertainties in the daily values due to random and sampling errors are smallest over the shipping lanes and largest in the poorly sampled regions. The smallest values can be found in the North Sea between the United Kingdom and Northern Europe with uncertainties in the air temperature and sea surface temperature below 0.25 °C. Similarly, the uncertainty in the specific humidity is less than 0.25 g Kg^{-1} in the North Sea. Over the shipping lanes between Europe, North and South America and Africa the uncertainty is greater but is generally less than 1°C and 1 g Kg^{-1} for the temperatures (air and sea) and humidity respectively. Outside the shipping lanes the uncertainty in temperature is between 1 °C and 2 °C or higher. The highest uncertainties in the specific humidity can be found in the poorly sampled regions of the South Atlantic with uncertainties of between 2 and 3 g Kg^{-1}. Uncertainties in the wind speed are also lowest over the main shipping lanes, with minimum values in the region 0.75 – 1.25 ms^{-1}, and highest in the poorly sampled and high variability regions with uncertainties of between 3 – 5 ms^{-1} in the Labrador Sea and the North Atlantic between

the UK and Newfoundland. The uncertainties in the cloud cover and sea level pressure show similar patterns to those seen in the other variables.

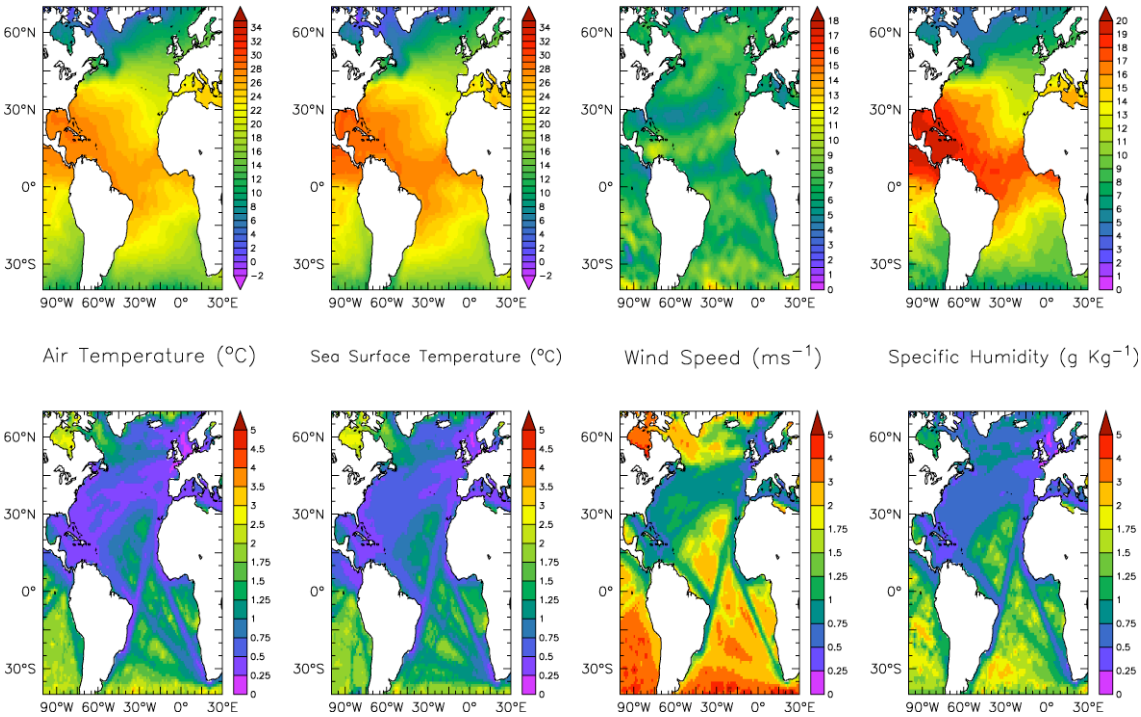

Figure 5-12a: As Figure 5-11a but for June – August 1993.

The random and sampling uncertainty in the fluxes also show the lowest values over the shipping lanes and highest values in poorly sampled regions. During the northern winter months the uncertainty in the daily latent heat flux is generally below 30 W m^{-2} over the shipping lanes, decreasing to below 20 W m^{-2} in the coastal regions of Europe. Outside of the shipping lanes in the Tropics the uncertainties in the daily values increase to 50 – 70 W m^{-2}. The highest uncertainty in the latent heat flux can be found in the poorly sampled South Atlantic with uncertainty estimates of over 100 W m^{-2}. For the sensible heat flux, the minimum uncertainty due to random and sampling errors can be found over the shipping lanes with uncertainties of < 10 W m^{-2}. Outside of the shipping lanes the uncertainty increases to 20 – 30 W m^{-2}. The largest values are found over the Labrador Sea with uncertainty estimates greater than 70 W m^{-2}. During the northern summer months, when the heat fluxes are generally lower, the uncertainties in the latent and sensible heat fluxes are smaller but show similar spatial patterns to the winter estimates.

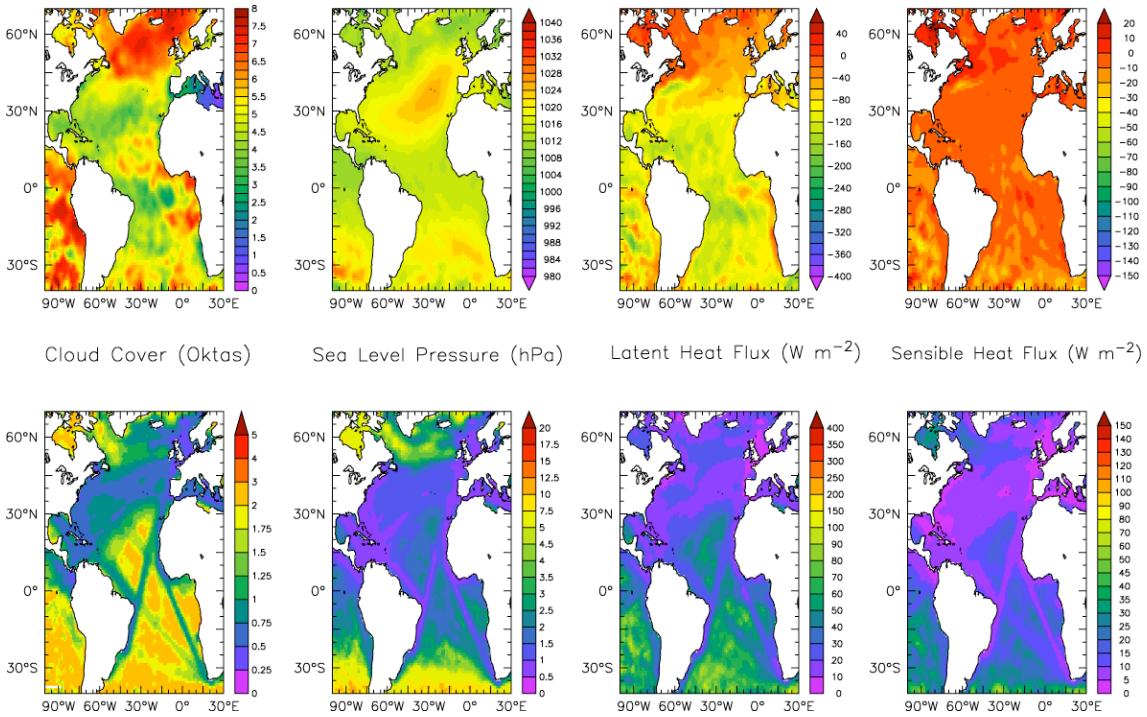

Figure 5-12b: As Figure 5-11b but for June – August 1993.

5.5.3 Bias Uncertainty

Figure 5-13 shows an example of the estimated bias uncertainty in the daily sea surface temperature for December 1992 – February 1993 (top) and June – August 1993 (bottom). This has been estimated as the maximum of 0.15 °C and $0.1 \cdot \left| T_{sea} - T_{air} \right|$ °C (Chapter 4.5.2). The bias uncertainty in the other variables required for the fluxes (not shown) have been estimated at (Chapter 4.5.2): air temperature - 0.2 °C; specific humidity - 0.2 g Kg^{-1}; and wind speed - 0.2 ms^{-1}. The resulting bias uncertainties in the latent and sensible heat fluxes are shown in the middle and right hand panels of Figure 5-13.

The bias uncertainty in the sea surface temperature peaks over the Gulf Stream and the Labrador Sea during the winter months, with values as high as 0.6 – 0.8 °C. Away from the regions of the maximum air – sea temperature differences, i.e. away from the western boundary current and Labrador Sea, the bias uncertainty decreases to below 0.2 °C. During the summer months the bias uncertainty in the SST is almost everywhere below 0.2 °C, with peak values less than 0.4 °C

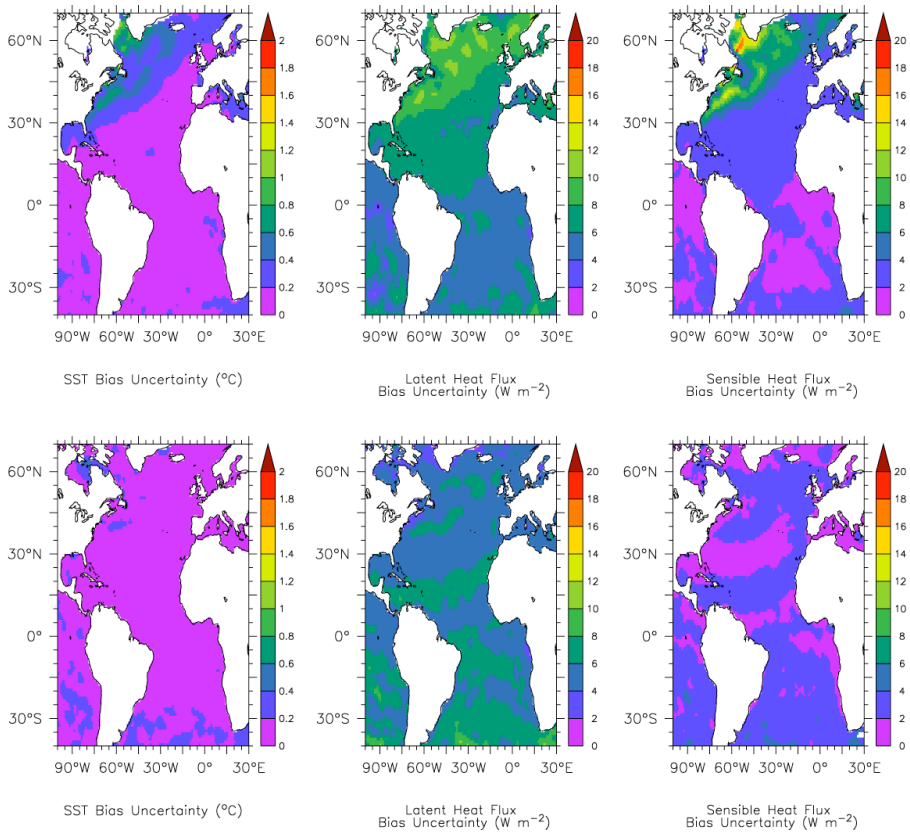

Figure 5-13: Mean daily bias uncertainty averaged over December 1992 – February 1993 (top) and June – August 1993 (bottom) for Sea Surface Temperature, latent heat flux and sensible heat flux.

The impact of the bias uncertainty on the flux estimates is greatest in regions of high flux and large air – sea temperature differences, peaking in the winter months and over the western boundary current. For latent heat flux, the bias uncertainty is between 10 – 12 W m^{-2} over the western boundary current region and Labrador seas during the winter. Outside of this region, the uncertainty decreases to below 8 W m^{-2} in the North Atlantic and below 6 W m^{-2} in the South Atlantic. During the summer the uncertainty over the Gulf Stream has decreased to 6 – 8 W m^{-2} and below 6 W m^{-2} for the rest of the North Atlantic outside of the tropics. In the northern tropics there is little change in the bias uncertainty estimates between northern winter and summer. In the South Atlantic there is an increase in the bias uncertainty to 6 – 8 W m^{-2} in northern summer. Similar patterns, but generally smaller in magnitude, are seen in the sensible heat flux bias uncertainty estimates over the majority of the Atlantic. However, in the Labrador Sea and Gulf Stream regions in the northern winter the impact of the SST bias uncertainty can be seen on the estimates for the sensible heat flux. Over these regions

the bias uncertainty estimates in the northern winter are generally larger than for latent heat flux and peak at over 14 W m^{-2}.

5.5.4 Flux Total Uncertainty

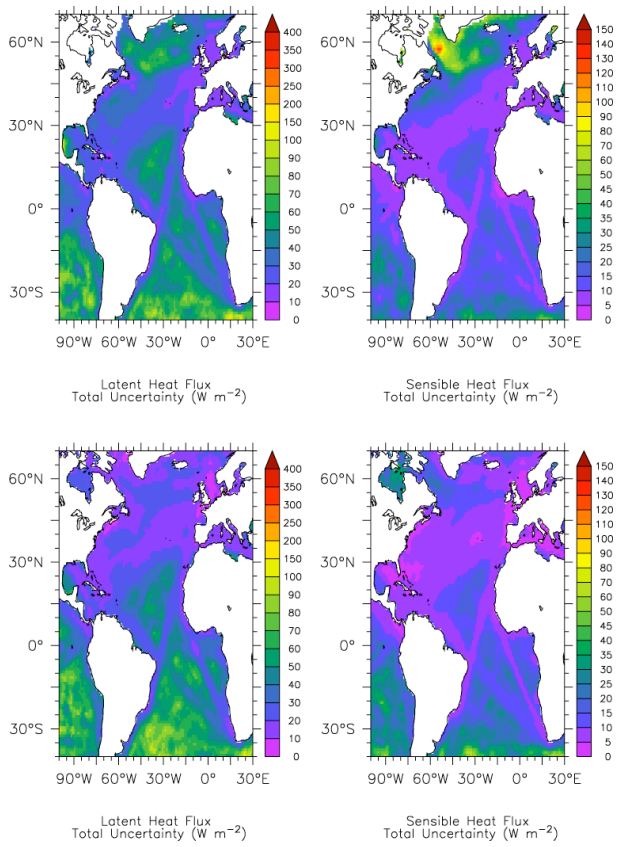

Figure 5-14: Total uncertainty estimates in the daily fluxes averaged over December 1992 – February 1993 (top) and June – August 1993 (bottom).

Figure 5-14 shows an example of the total uncertainty (i.e. bias uncertainty and random and sampling uncertainty added in quadrature) in the daily latent (left) and sensible (right) heat flux estimates averaged over December 1992 – February 1993 (top) and June – August 1993 (bottom). The total daily uncertainty for both variables is dominated by the random and sampling uncertainties. For monthly values, the random and sampling uncertainties will be reduced in well sampled regions and the bias uncertainties will dominate. The increase due to the bias uncertainty can be seen over the Gulf Stream region where the bias uncertainties are relatively large and in the North Sea where the random uncertainties are small. In both cases the mean total uncertainty is increased by several W m^{-2} by the inclusion of the bias uncertainty. This increase is typically 3 – 4 W m^{-2} over the Gulf Stream for sensible heat flux and 3 – 4 W m^{-2} over

the North Sea for latent heat flux during the winter. Smaller increases are seen during the summer months.

5.5.5 Impact of Bias Corrections on the Mean Fields

Figure 5-15 shows an example of the impact of the bias adjustments applied to the air temperature and humidity observations on the interpolated fields of air temperature, humidity and the surface fluxes averaged over the 1992/1993 winter and 1993 summer. Differences between OI runs using combined height and bias adjusted and height adjusted observations (height and bias adjusted – height adjusted) are shown and include the impact of the bias corrections on the height adjustments applied. In addition to the impact of the bias corrections on the height adjustments the datasets used in the two sets of OI runs are slightly different. This difference is due to a number of observations being discarded in the bias and height adjusted version where parameters required to calculate the bias corrections are missing or where the height adjustment does not converge after the bias corrections have been applied. These differences are only noticeable in very poorly sampled regions and are within the uncertainty estimates.

The impact of the air temperature bias correction peaks in the tropics with maximum differences of almost 1 °C between the datasets with and without bias adjustments applied. In the northern winter the maximum negative values are found in the southern equatorial regions of the Atlantic. During the northern summer months the region of maximum correction encompasses the northern equatorial Atlantic. The warm region seen in the plot for December 1992 – February 1993 to the north east of Iceland is due to the exclusion of observations from the bias corrected dataset as mentioned above. The humidity bias correction also has largest impact in the tropics, with peak values between 0.4 and 0.6 g Kg^{-1} in the tropics for both seasons. During the winter, the correction reduces the humidity in the mid latitude North Atlantic by less than 0.2 g Kg^{-1}. The humidity is reduced by between 0.2 and 0.4 g Kg^{-1} there during the summer months. As with air temperature, the regions where the input datasets differ can be seen in the humidity plots with the sign of the bias adjustment being opposite to that expected.

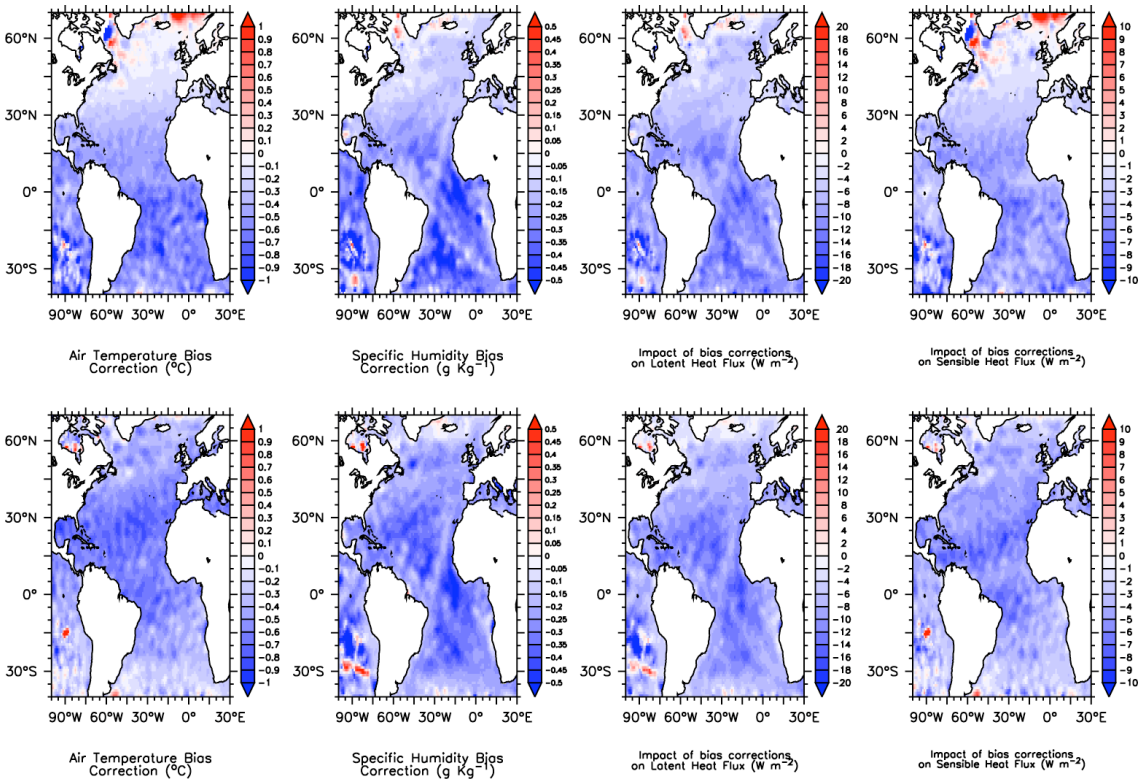

Figure 5-15: Impact of the bias corrections on daily estimates of: air temperature (left); specific humidity (left middle); latent heat flux (right middle); and sensible heat flux (right) averaged over December 1992 – February 1993 (top) and June – August 1993 (bottom).

The impact of the bias corrections on the latent and sensible heat fluxes can be clearly seen. The spatial patterns are similar to those seen in air temperature and humidity for the sensible and latent heat fluxes respectively. The bias adjustments, through increasing the air – sea humidity difference, increase the magnitude of the latent heat flux (i.e. more negative) during the winter by an average of 8.2 W m^{-2} with a standard deviation of 6.0 W m^{-2}. The differences are greatest in the tropics with the magnitude of the adjusted latent heat fluxes 10 to 15 W m^{-2} larger than unadjusted values. In the mid latitude North Atlantic, the differences are less than 10 W m^{-2}. During northern summer, mean differences are 8.4 ± 5.9 W m^{-2} and the focus of major differences is further north, matching the humidity correction.

The magnitude of the sensible heat fluxes are increased by an average of 3.8 ± 4.2 W m^{-2} during the winter and 4.0 ± 3.1 W m^{-2} during the summer. As with latent heat flux, this increase is due to the bias adjustment increasing the air – sea temperature difference. Regions where differences have the opposite sign to the bias corrections are

due to small differences in the input data and these poorly sampled regions have been excluded from the calculated averages.

5.6 Summary

In this Chapter, the impact of the QC, track checking, bias and height adjustment has been shown together with the results of the random error estimation and the OI. Combined, the QC and track checking remove a small amount of data for each variable, typically less than 5 % of the available observations. The use of different semi-variogram models has been shown to give improved estimates of random errors compared to the model used by previous authors, and the height and bias adjustments have been shown to reduce the estimated random errors. This reduction gives confidence in those adjustments, and the improvement made to the mean fields by the adjustments will be shown in Chapter 6.

The mean fields of the basic variables match the known climate, with the highest temperatures and humidities occurring in the tropics and with a general East – West gradient towards higher values over the tropical and subtropical western boundary current. The northern sub-tropical gyre can be seen in the wind speed estimates with the lowest wind speed values occurring in the centre of the gyre and highest values occurring in the trade wind belt and storm track regions. The fluxes peak over the Gulf Stream region and are highest in the winter months. The fluxes are increased slightly (< 20 W m^{-2} in total) by the application of the air temperature and humidity bias adjustments because the adjustments increase the air – sea temperature and humidity gradients.

The uncertainty estimates due to random and sampling errors for each variable are lowest over the shipping lanes and highest in poorly sampled regions. On a daily time scale, the bias uncertainties are small compared to the random and sampling uncertainties. On monthly time scales the bias uncertainties are expected to dominate in well sampled regions. In poorly sampled regions the random errors will still dominate the total uncertainty estimates.

6 Cross Validation and Comparisons

6.1 Introduction

This Chapter presents a quantitative assessment of the OI daily fields and their uncertainties. The original observations are compared to the optimally interpolated values in Section 6.2 to identify any bias compared to the input data. Independent buoy observations from the Woods Hole Oceanographic Institutions (WHOI) Upper Ocean Processes (UOP) group data archive are compared to the fields from the new dataset in Section 6.3. Differences between the buoy observations and a number of other surface flux and meteorology datasets are also shown for comparison.

Sections 6.4 and 6.5 examine the uncertainty estimates in the new dataset. The uncertainty estimates from the OI are compared to alternative estimates made using a 3-way error analysis in Section 6.4. In Section 6.5 the uncertainty estimates are cross validated, performing an ensemble of runs of the OI with 50 % of the observations excluded from each run. The standard deviation of the daily fields from the ensemble is then compared to the uncertainty estimates to identify regions where we may be over or underestimating the uncertainty. The results presented in the Chapter are then discussed and summarised in Sections 6.6 and 6.7 respectively.

6.2 Comparison to VOS Observations

In order to check that the OI is working and that the output is unbiased compared to the input data the OI has been run using unadjusted VOS observations (i.e. without height or bias adjustments applied) and then compared to those same observations. Figure 6-1 shows histograms of differences between the unadjusted observations and the OI fields for observations of the basic variables made during January and July 1993 over the Atlantic Basin. Also shown, as inset plots are the mean differences averaged against the estimates from the interpolation.

For each variable, the histograms are approximately symmetrical and centred on zero, as expected if the OI were working correctly. Similar results are seen in all months (not shown). When the residuals are binned and plotted against the daily estimates from the OI no significant systematic variations are visible for air or sea surface temperatures or for specific humidity. The slight deviations at either extreme are a function of the random errors in the both the interpolated values and original data. This impact of the

random errors is most extreme for the wind speed and sea level pressure estimates. Similar results have been shown before in comparisons of individual VOS wind speed observations and scatterometer derived wind speeds (e.g. Kent *et al*. 1998).

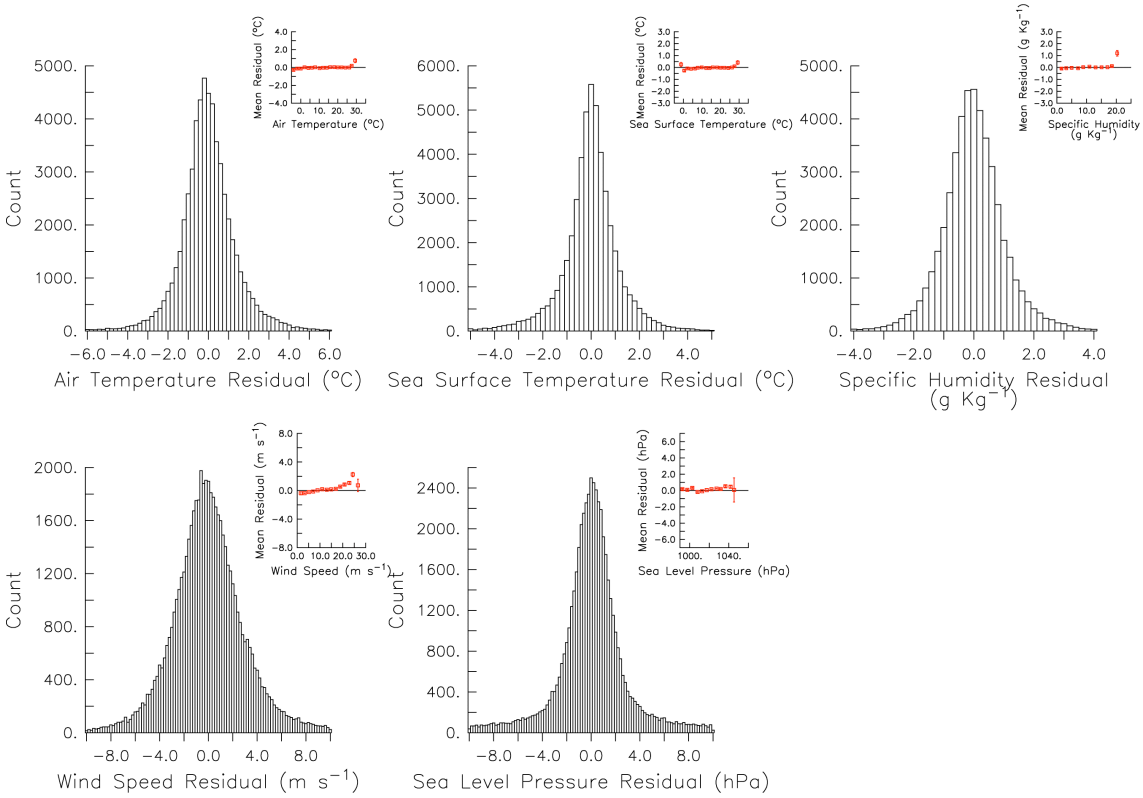

Figure 6-1a: Histograms of the residuals (VOS Observations – optimally interpolated values) for each of the main variables during January 1993. Also shown are the mean residuals as a function of the OI estimates (inset plots).

To demonstrate that the deviations are explicable by random errors we follow Kent *et al*. (1998) by attempting to reproduce these results using simulated data. Two datasets have been generated using VOS wind speed observations from ICOADS made during July 1993. For the first dataset ('VOS like'), the wind speed observations have been perturbed by adding normally distributed random errors typical of the VOS with a standard deviation of 2 ms^{-1}. For the second dataset ('OI like'), the same procedure has been followed but with random errors with a standard deviation of 1 ms^{-1} added. This lower value is more typical of the uncertainty estimates for the daily wind speed estimates from the OI. Differences between the two datasets have then been calculated ('VOS like' – 'OI like') and plotted as a function of the 'VOS like' wind speeds (Figure 6-2). A similar trend to the comparison of the actual VOS observations and the OI output is seen, with low wind speed values over estimated and high wind speeds

underestimated in the 'OI like' data compared to the 'VOS like' data. In addition to the impact of random errors, the smoothing implicit in the OI can be expected to lead to similar results with minima increased and maxima decreased.

Table 6-1: Monthly mean residuals over 1970 – 2006 for each variable. Also listed for each variable are: the maximum absolute monthly mean residual; the mean RMS error averaged over 1970 – 2006; and the random error estimates for unadjusted observations and using the chosen variogram models (Table 5-2).

Variable	Mean Residual	Max |Residual|	Mean RMS error	Random Error
Air Temperature (°C)	-0.01	0.01	1.47	1.25
Specific Humidity (g Kg^{-1})	0.00	0.03	1.12	1.29
Sea Surface Temperature (°C)	0.00	0.06	1.26	1.24
Sea Level Pressure (hPa)	0.00	0.18	2.48	2.10
Wind Speed (ms^{-1})	0.02	0.08	2.62	2.30

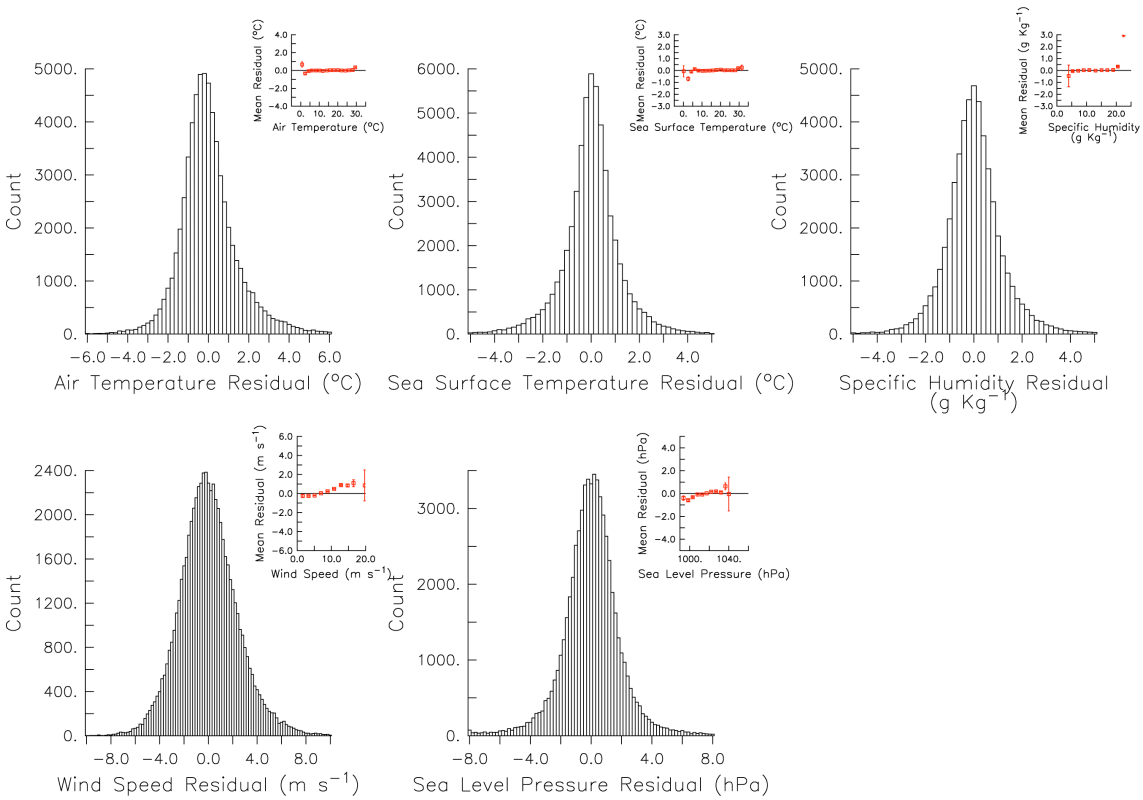

Figure 6-1b: As Figure 6-1a but for July 1993

Figure 6-3 shows the mean residuals (VOS – OI) plotted against time for each of the variables over the Atlantic and is summarised in Table 6-1. The mean differences for each variable are close to zero. The RMS errors are similar in magnitude to the

random error estimates from the semi-variograms, giving us confidence in our random error estimates. The annual cycle seen in the RMS errors is due to the increased variability in the winter months. Overall, the results demonstrate that the fields are unbiased compared to the input data.

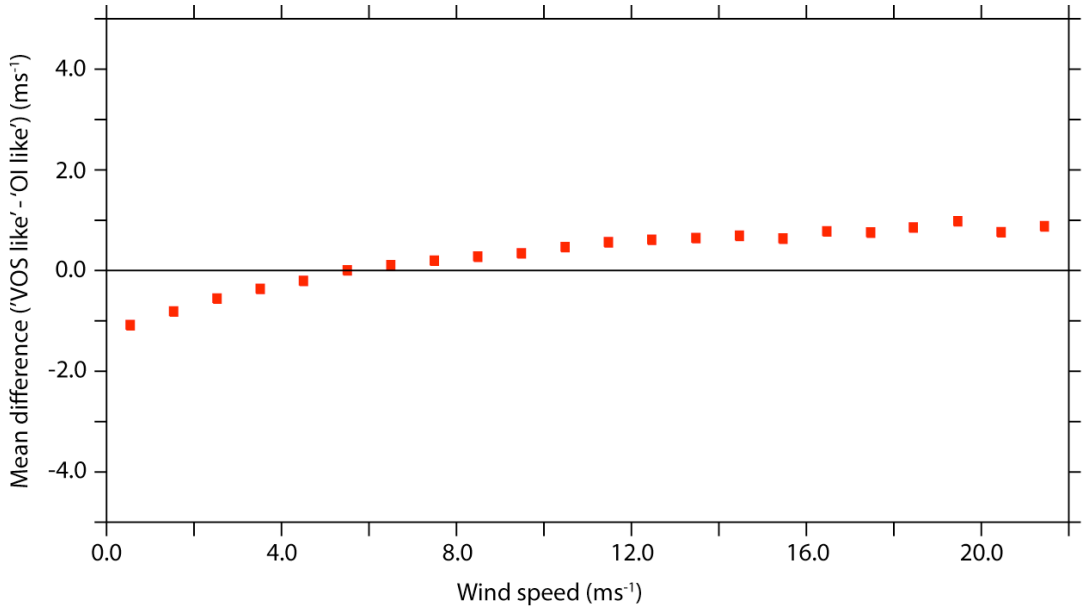

Figure 6-2: Comparison of residuals from simulated wind speed dataset as a function of wind speed.

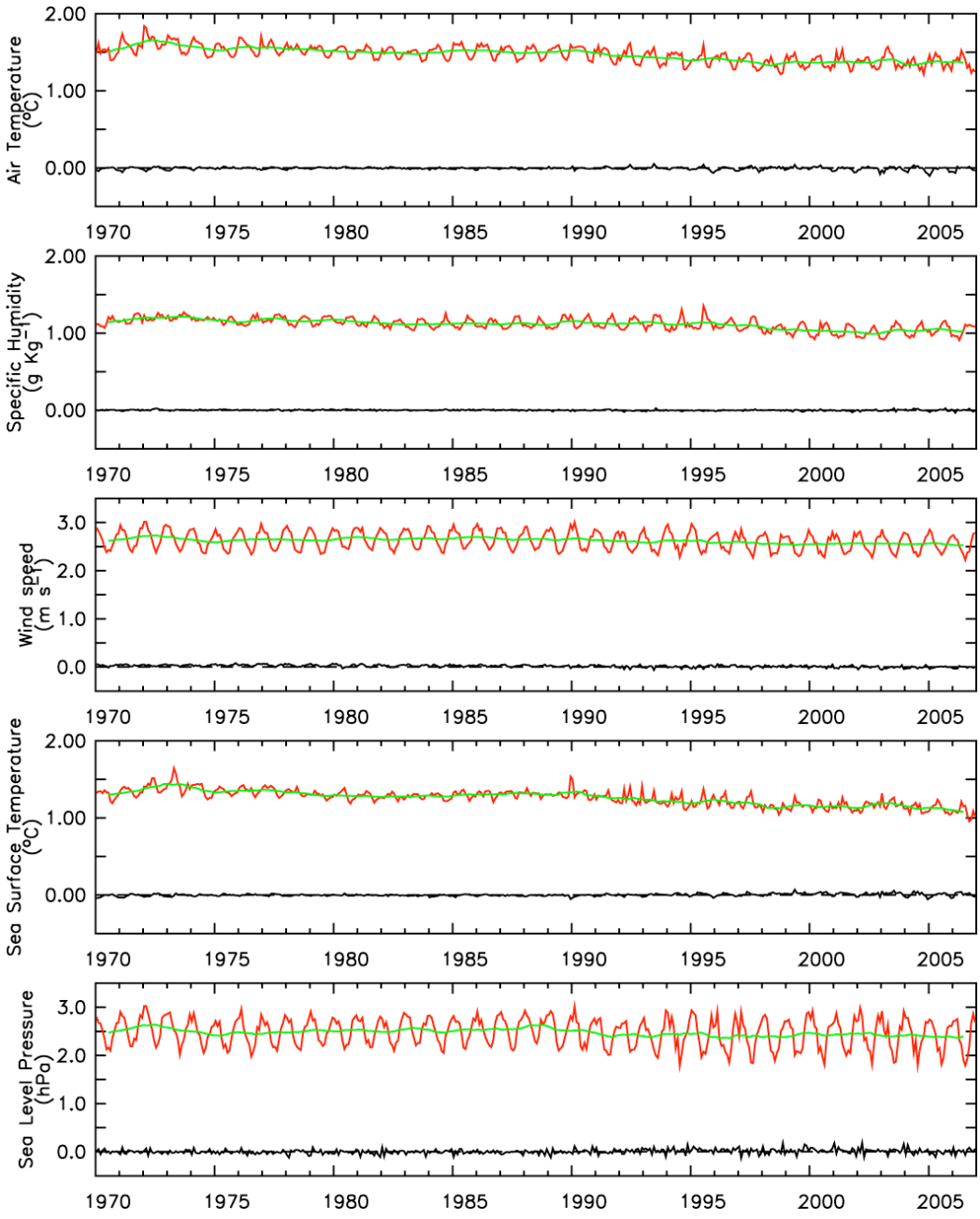

Figure 6-3: Time series of monthly mean residuals (black) from the comparison of VOS observations with the output of the OI over the Atlantic (40S – 70N, 100W – 30 E). Also shown are the RMS errors (red) between the VOS observations and the output of the OI. Smoothed values, using a 12 month running mean filter, are also shown.

6.3 Comparison to WHOI UOP Mooring Data Archive

6.3.1 Introduction

The observations from the WHOI UOP mooring data archive are withheld from the ICOADS and form an independent validation dataset. These observations are used to validate the output from the OI. Table 6-2 lists locations and time periods covered by the moorings used, the locations are also shown in Figure 6-4. The buoy data include all the parameters required to calculate the fluxes and have been averaged to give daily mean values. The typical long term accuracy of the buoy observations are listed in Table 6-3 (e.g. Crescenti and Weller 1992; Moyer and Weller 1997). In the following comparisons the buoy observations have been height adjusted to 10 m and the fluxes calculated using the bulk formulae (see Chapter 3). The parameterisations used in the new dataset (Smith 1980, 1988) have also been used with the buoy observations.

Table 6-2: Buoys observations used from the WHOI UOP Data Archive. Also listed are the buoy locations and the periods with valid data.

Buoy	**Location**	**Period**
South West Subduction buoy (SW)	18 °N, 34 °W	October 1992 – June 1993
South East Subduction buoy (SE)	18 °N, 22 °W	March 1992 – June 1993
Central Subduction buoy (CENT)	25.5 °N, 29 °W	June 1991 – June 1993
North West Subduction buoy (NW)	33 °N, 34 °W	March 1992 – March 1993
North East Subduction buoy (NE)	33 °N, 22 °W	June 1991 – June 1993
Severe Environment Surface Mooring (SESMOOR)	42° 33'N, 61° 14'W	October 1988 – March 1989
Coastal Mixing and Optics Experiment (CMO)	40° 30' N, 70 30' W	August 1996 – June 1997
Marine Light – Mixed Layer Experiment (MLML91)	50 °N 29', 20° 50' W	April 1991 – September 1991

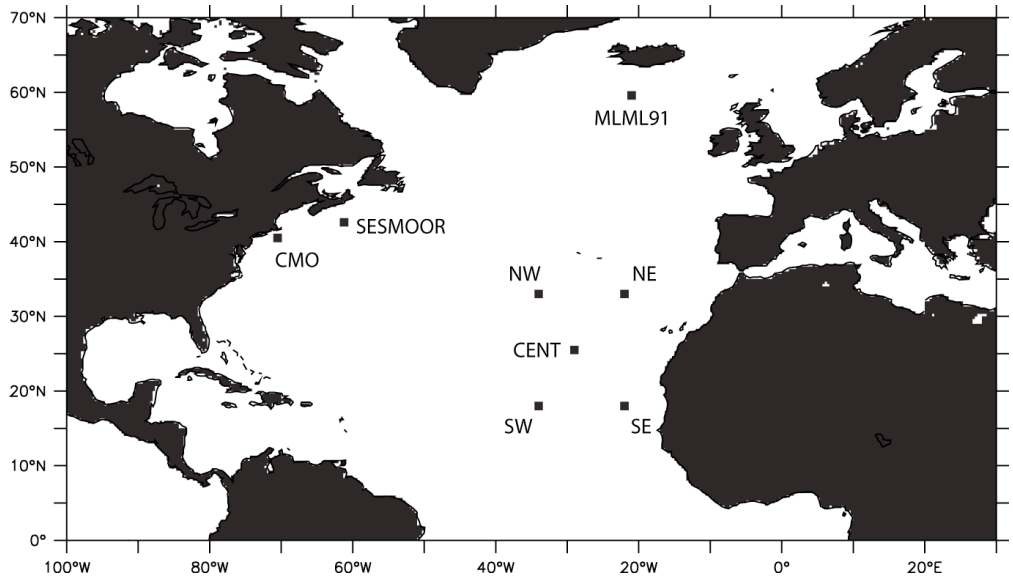

Figure 6-4: Location of the buoys listed in Table 6-2.

Table 6-3: Typical long-term accuracy of the basic observations made by the WHOI buoys (e.g. Crescenti and Weller 1992; Moyer and Weller 1997). Also listed are the bias uncertainties for the VOS observations (see Chapter 4). The values listed for latent and sensible heat fluxes are based on the Subduction buoys (Moyer and Weller 1997).

Variable	Long term accuracy	VOS Bias Uncertainty Estimate
Sea surface temperature	0.1 °C	$\max(0.15, 0.1 \cdot \lvert T_{sea} - T_{air} \rvert)$
Air Temperature	0.2 °C	0.2°C
Relative Humidity	3 %	
Specific Humidity	0.2 g Kg^{-1}	0.2 g Kg^{-1}
Wind speed	0.3 – 0.5 ms^{-1}	0.2 ms^{-1}
Sea Level Pressure	0.2 – 0.4 hPa	0 hPa
Latent Heat Flux	13 – 18.4 Wm^{-2}	
Sensible Heat Flux	2 – 2.5 Wm^{-2}	

6.3.2 Case Study – Comparison to the North East Subduction Buoy

A comparison of daily meteorological parameters and flux estimates from the new dataset and the North East Subduction Buoy (33N 22W) during the summer (JJA) of 1992 is shown in Figure 6-5 and during the 1992 / 1993 winter (DJF) in Figure 6-6. The buoy measurements are shown in black, the values from the new dataset in red and the uncertainty in the new dataset indicated by the grey shaded region. The mean values

and standard deviation of the basic variables and fluxes are listed in Table 6-4 for both the new dataset and the buoy measurements. Also listed are: the mean differences and the root mean squared error of the differences (RMSE); the mean daily random and sampling error uncertainty (*S*) from the OI (Section 5.5.2); and an estimate of the total error for the mean differences.

The total error has been estimated as sum of the mean random and sampling uncertainty from the OI divided by the square root of the number of days, the buoy accuracy (Table 6-3) and the bias uncertainty for the OI dataset (Table 6-3 and Section 5.5.3) added in quadrature. Where a range is listed for the buoy long term accuracies listed in Table 6-3 the mid point has been used. The total error estimate gives an estimate of the uncertainty in the mean value and can be used in place of the standard error of the mean in a Students t-test to test whether the mean differences are significantly different to zero.

Table 6-4: Mean (μ) and standard deviations (σ) of the variables shown in Figure 6-5 and Figure 6-6 for the new dataset (NOCS OI) and the NE Subduction buoy. The mean difference and root mean squared error between the new dataset and the buoy observations are also listed together with the mean random and sampling uncertainty for the daily OI values (*S*).

	Variable	NOCS OI ($\mu \pm \sigma$)	Buoy ($\mu \pm \sigma$)	Mean Difference	RMSE	*S*	Total Error
Jun - Aug 1992	T_{10S} (°C)	21.3 ± 1.8	21.4 ± 1.7	-0.1	0.5	0.7	0.3
	SST (°C)	22.4 ± 1.6	22.0 ± 1.5	0.4	0.5	0.8	0.2
	q_{10S} (g Kg^{-1})	12.3 ± 1.7	12.4 ± 1.9	-0.1	0.8	0.8	0.3
	U_{10S} (m s^{-1})	6.3 ± 1.6	5.8 ± 1.7	0.5	1.4	1.3	0.5
	SLP (hPa)	1024 ± 2	1023 ± 3	0.9	1.6	1.4	0.3
	LHF (W m^{-2})	-101 ± 32	-81 ± 35	-19.0	30.4	27.2	17.4
	SHF (W m^{-2})	-8.6 ± 5.0	-5.0 ± 3.8	-3.6	5.5	9.3	3.4
Dec 1992 – Feb 1993	T_{10S} (°C)	18 ± 1.0	17.9 ± 1.1	0.1	0.5	0.6	0.3
	SST (°C)	19.1 ± 1.0	19.4 ± 0.9	-0.4	0.6	0.6	0.2
	q_{10S} (g Kg^{-1})	9.2 ± 0.9	8.6 ± 1.1	0.6	0.9	0.7	0.3
	U_{10S} (m s^{-1})	6.9 ± 1.9	6.8 ± 2.6	0.1	1.5	1.3	0.5
	SLP (hPa)	1024 ± 4	1023 ± 4	0.8	1.6	1.8	0.4
	LHF (W m^{-2})	-112 ± 50	-140 ± 60	27.5	45.7	25.9	17.5
	SHF (W m^{-2})	-10.2 ± 10.1	-15.7 ± 11.7	5.5	10.6	8.1	3.5

Overall, there is good agreement between the buoy observations and the estimates from the new dataset. Typically, the daily averages from the buoy lie within the OI uncertainty range from the new dataset (Figure 6-5 and Figure 6-6). The daily differences between the two datasets for the different variables have values of either

sign over the periods shown and mean differences sometimes smaller than the estimate of the total error. For each variable, excluding the SLP during the summer, the t-statistic (mean difference / total error) is smaller than the critical value (t-critical = 2.28, degrees of freedom = 90, $\alpha = 0.05$) suggesting the mean biases are within the accuracy of the two datasets at the 95 % confidence level. The t-statistic for the mean sea level pressure difference during the summer is 3.0 suggesting a significant difference between the two datasets. However, this will have only a minor impact on the fluxes.

Whilst the daily values from the buoy are generally within the uncertainty range from the OI and the mean differences within the accuracy of the two datasets the mean differences are still relatively large for the SST during both seasons and for the humidity during the winter. For example, the SST during the summer tends to be slightly higher than the buoy measurements with a mean difference of 0.4 °C and colder in the winter with a mean difference of -0.4°C. These mean differences result in relatively large mean differences between the flux estimates from the two datasets. However, these are still within the estimated accuracy of the two datasets.

When the variability between the two datasets is compared we generally see agreement on timescales of several days or longer with the major features seen in both. For example, a period of relatively high humidity occurs at the end of June 1992 (Figure 6-5). This can be seen in the buoy observations and is also captured in the daily estimates from the new dataset. Whilst the variability is generally captured on timescales of several days or longer the daily variability is underestimated. For example, the wind speed estimates during the winter have similar mean values (mean difference of 0.1 ms^{-1}), but the buoy observations have much greater variability and more extreme values. This can be seen in Figure 6-6 and by comparing the standard deviations of the wind speed from the two datasets, 2.6 ms^{-1} for the buoy observations compared to 1.9 ms^{-1} for the new dataset. Whilst the OI is not capturing the full daily variability we are comparing gridded mean values with point measurements in space and, as a result, would expect to lower variability in the gridded data.

These results suggest the mean fields from the new dataset are not significantly biased at this buoy location compared to the total uncertainty estimates and that the majority of the variability on timescales of several days or longer is being captured. The next section examines the mean differences over a wider number of buoys and under different conditions of natural variability and sampling.

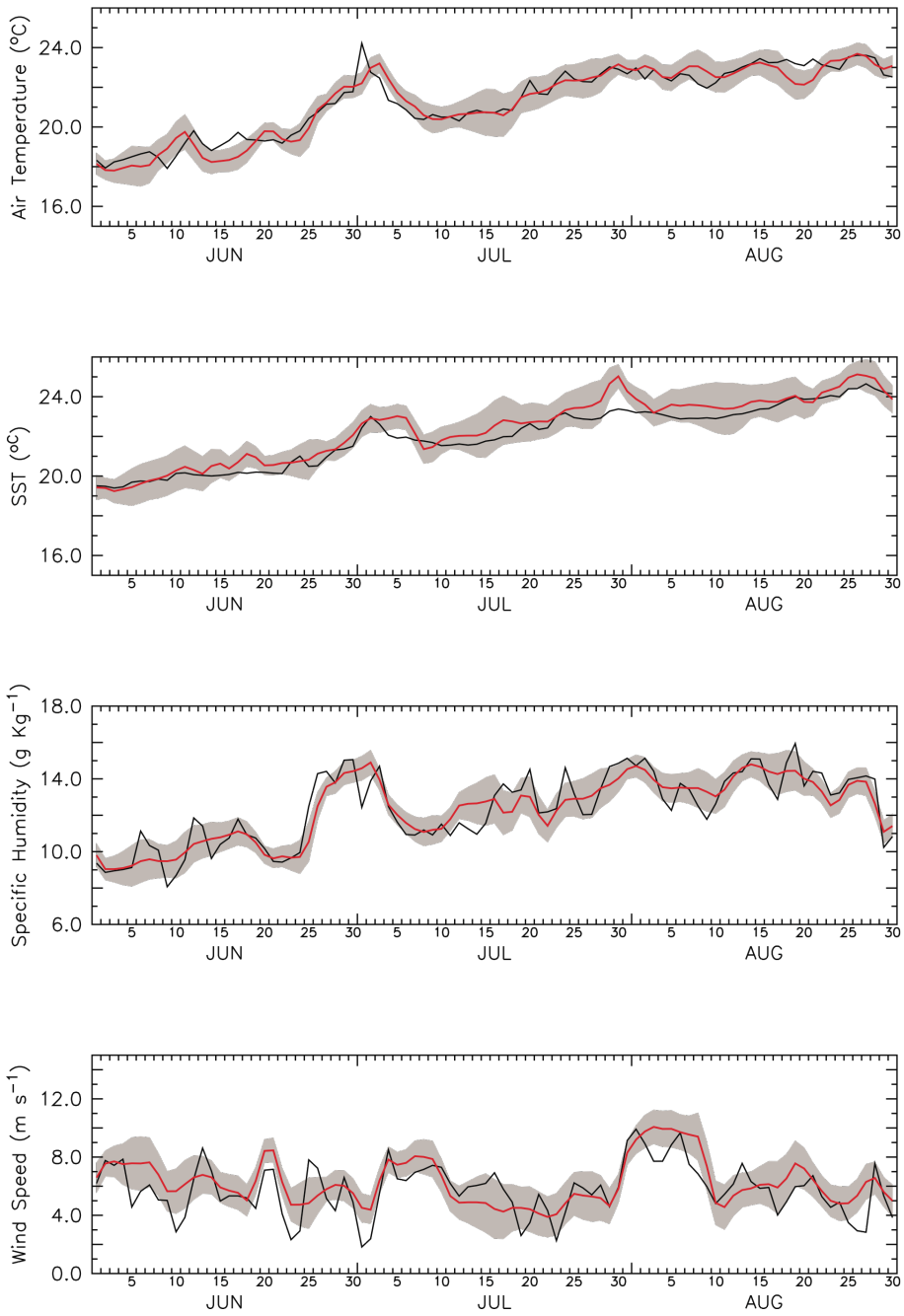

Figure 6-5: Comparison of daily values from the new dataset (red) and values from the North East Subduction buoy (black) during June – August 1992. The grey shading indicates the uncertainty in the output from the OI due to random and sampling errors. The bias uncertainty is not included.

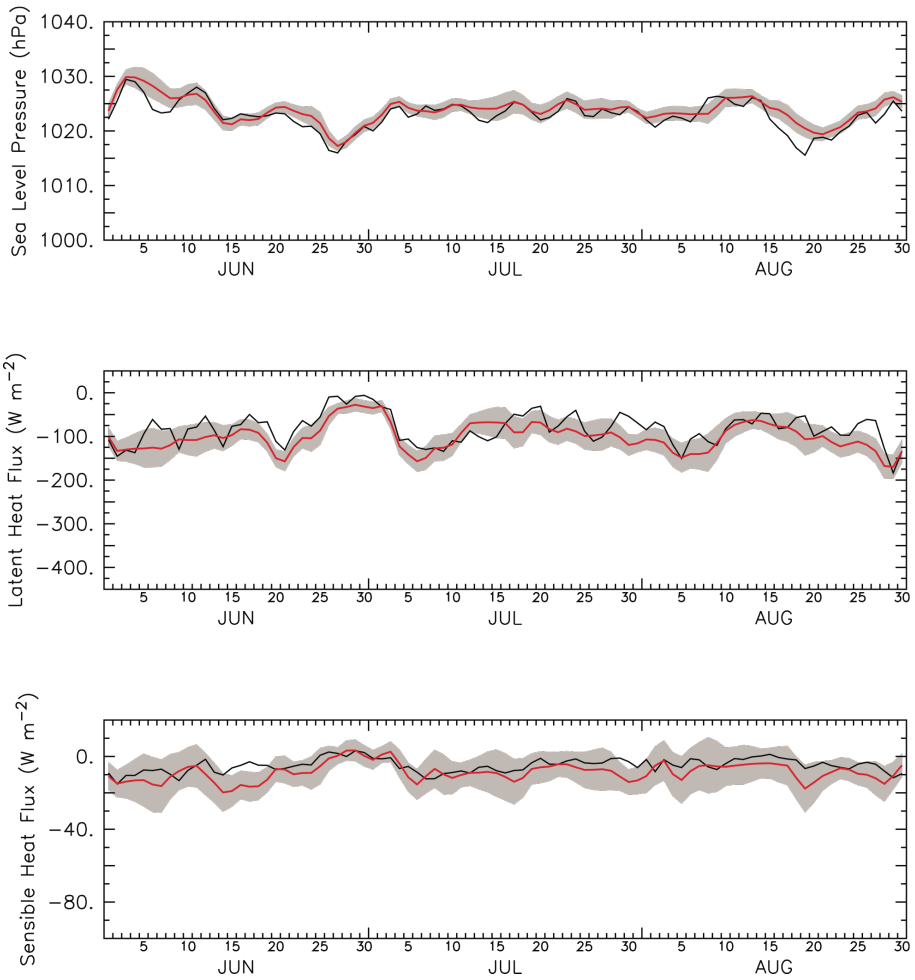

Figure 6-5 continued.

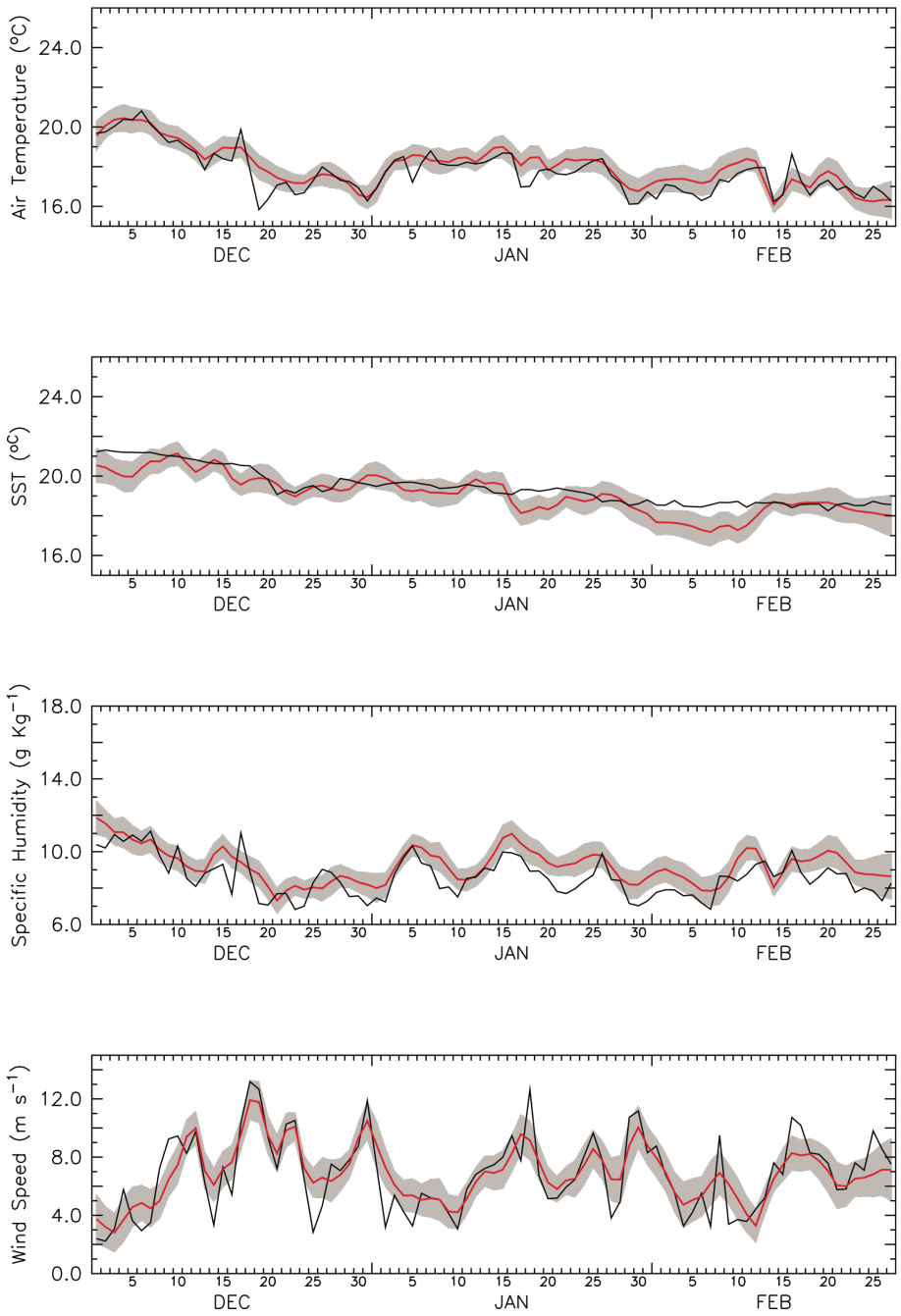

Figure 6-6: As Figure 6-5 but for December 1992 – February 1993.

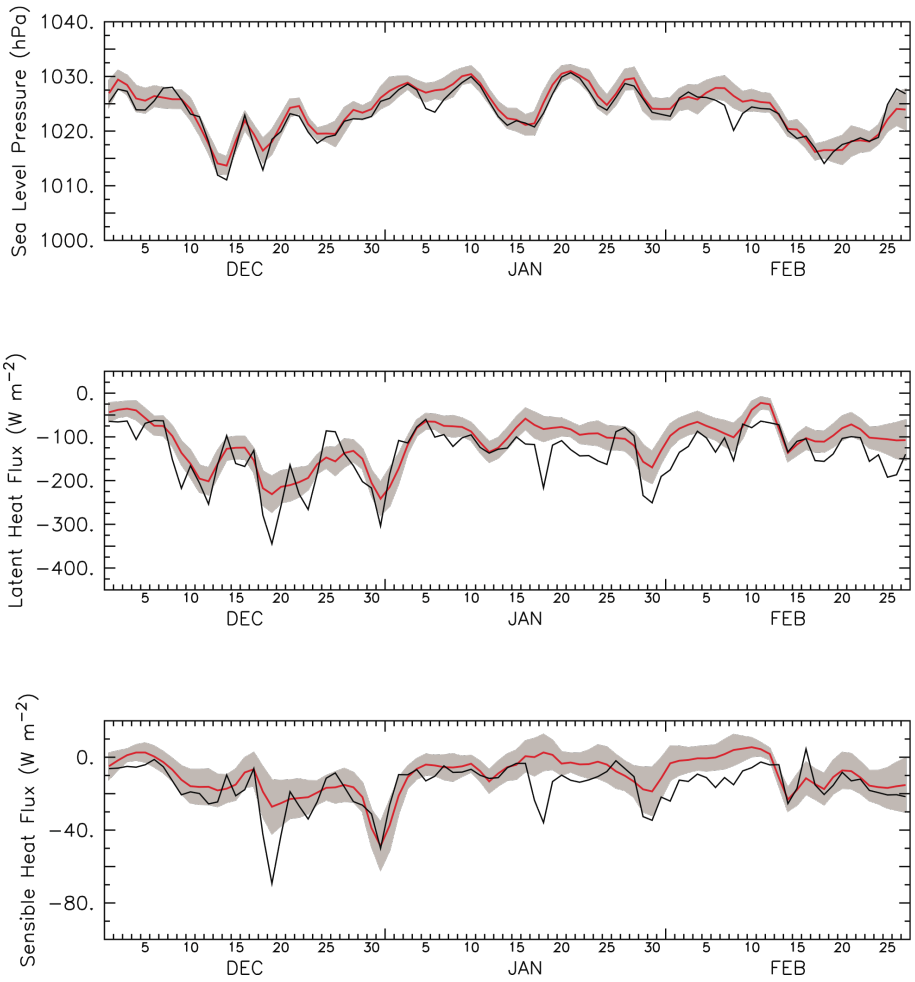

Figure 6-6 continued.

6.3.3 Comparison of Mean Differences

6.3.3.1 Subduction Buoys

The Subduction buoys are located in a region of large-scale horizontal gradients in the surface forcing (e.g. Moyer and Weller 1997). Three of the buoys (NW, NE and SE) are within or on the edge of shipping lanes and fairly well sampled. The remaining two buoys (SW and Central) are in poorly sampled areas between the shipping lanes. As an example, Figure 6-7 shows the mean daily air temperature (left) and mean daily uncertainty (right) for January 1992. Also shown are the locations of the different Subduction buoys. The low spatial gradient in the air temperature can be seen at all locations together with the low uncertainties at the three well sampled buoys. Similar results are seen for the other variables. Due to this low variability and relatively good sampling mean differences between the OI fields and the buoy observations are expected to be small.

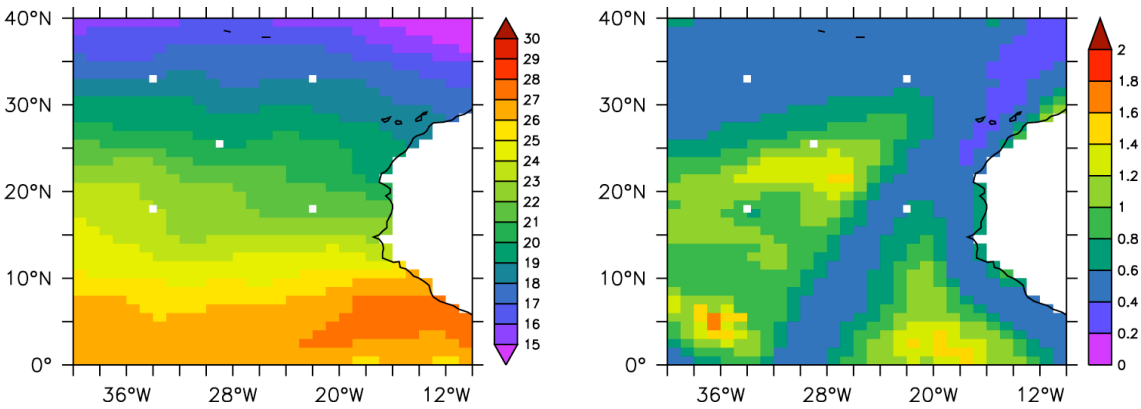

Figure 6-7: Mean daily air temperature (left) and mean daily uncertainty (right) during January 1992 from the new dataset. The locations of the different Subduction buoys are shown as white squares.

Figure 6-8 and Figure 6-9 show examples of the mean differences (dataset – buoy) between the new dataset and the North West buoy (well sampled) and the South West buoy (relatively poorly sampled) respectively. The mean differences for each of the basic variables and the fluxes are shown. Also shown are the differences between the buoy observations and a non-bias adjusted version of the new dataset (but with height adjustments applied) and also 4 gridded datasets (OAFlux, ERA40, NCEP1 and NCEP2). The comparison for these two buoys and the other three Subduction buoys are summarised in Table 6-5.

The mean differences for the basic variables (excluding humidity) from the new dataset are generally small and of either sign. For example, the mean differences for air temperature range between -0.1 °C and 0.3 °C across the different buoys and with a mean value of 0.1 °C when averaged across all buoys. For the humidity, the mean differences are also small at the different buoy locations. However, they are all positive, varying between 0.2 and 0.6 g Kg^{-1} and with a mean value of 0.4 g Kg^{-1} when averaged across all buoys. Whilst all positive, possibly indicating a positive bias in the new dataset or a negative bias in the buoy observations, the differences are still within the estimated accuracy of the two datasets. Overall, these results suggest that the air temperature, SST, wind speed and sea level pressure from the new dataset are unbiased compared to the buoy observations.

The impact of the bias corrections on the different variables can clearly be seen in Figures 6-8 and 6-9, with the air temperature and humidity brought into better agreement with the buoy observations after adjustment. In contrast, after adjustment, the

wind speed estimates from the new dataset are in worse agreement with the buoys compared to the non-bias adjusted data, with the mean difference increased by 0.1 – 0.2 ms^{-1}. This is due to the adjustment of the visual wind speed estimates following Lindau (1995) and Thomas *et al*. (2005) increasing the wind speed at the low to moderate values typical of the Subduction array region and increasing the mean differences as a result. This is discussed further in Section 6.6.1.

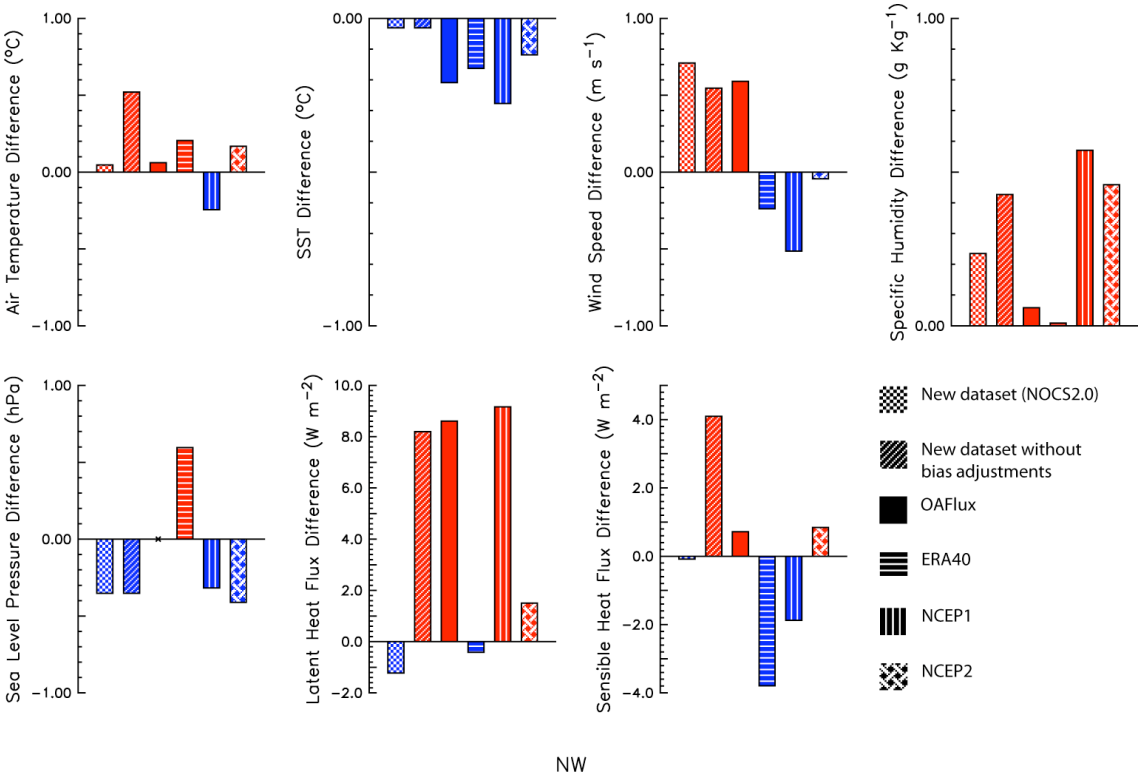

Figure 6-8: Mean differences (product – buoy) between the North West Subduction buoy observations and a number of different flux products for each of the main variables and fluxes. An x on the abscissa indicates that the data is unavailable for that dataset.

When the fluxes from the new dataset are compared to the fluxes calculated using the buoy observations similar results are seen. The mean differences are small, both for individual buoys and averaged across the buoys, and generally less than 10% of the mean flux. Averaged across the buoys, the mean differences are 5.3 W m^{-2} and -0.3 W m^{-2} for latent and sensible heat flux respectively compared to mean fluxes of -116 W m^{-2} and 8.3 W m^{-2}. The improvements to the estimates in the new dataset made by the bias adjustments can again be seen compared to the un-adjusted version with smaller mean differences in the adjusted version. The relatively large differences (as a

percentage of the flux) in the sensible heat flux at the location of the South West buoy, the most poorly sampled of the buoys, are due to the overestimation of the wind speed and sea surface temperature at this location in the new dataset.

The comparisons of the basic variables and fluxes from the other datasets to the buoy observations show similar results to the comparisons for the new dataset. Generally, the mean differences are small at the different buoy locations for each variable and close to zero when averaged across the different buoys. Interestingly, the over estimate of the humidity seen in the new dataset is also seen in the other datasets with similar sized mean differences suggesting the humidity is also overestimated in these datasets. There are also noticeable differences for the wind speed estimates from ERA40 and NCEP1 are underestimated compared to the buoy observations.

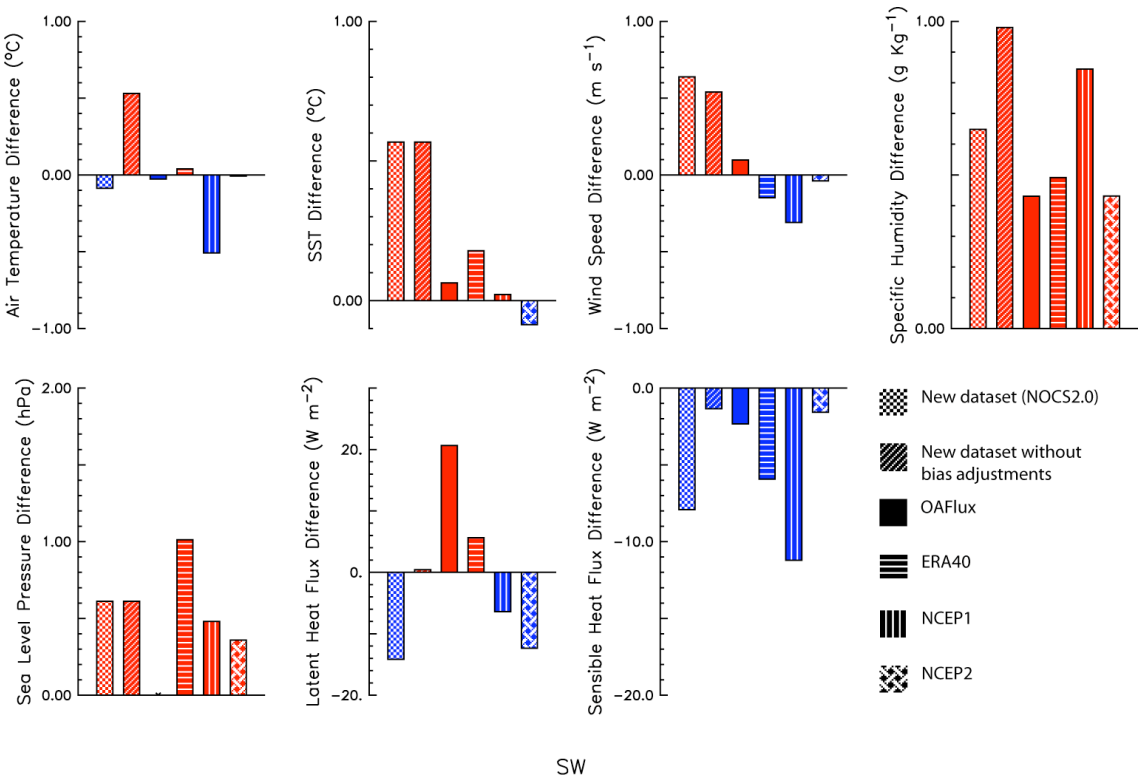

Figure 6-9: As Figure 6-8 but for the South West Subduction buoy.

Overall, the comparisons between the different variables from the new dataset and the observations from the Subduction buoys suggest the fields in the new dataset are not significantly biased in this area. A small positive bias is present in the humidity estimates from the OI, but is not significant compared to the total error estimates. In addition to having unbiased fields, in this relatively well sampled region the new dataset

is comparable to other gridded datasets in terms of mean differences from the buoy observations.

6.3.3.2 Marine Light-Mixed Layer Experiment (MLML91)

The Marine Light-Mixed Layer Experiment mooring was also located in a region with relatively large spatial scales of variability and with moderate sampling. Figure 6-10 and Table 6-6 show a comparison of the different datasets and buoys observations. As expected, from the fairly good sampling, small mean differences are seen for the different variables from the new dataset and the other datasets examined. The improvements made by the air temperature and humidity bias adjustments can be seen. As with the Subduction array, the bias adjustments made to the visual wind speed estimates increase the mean differences in the new dataset compared to non-bias adjusted data.

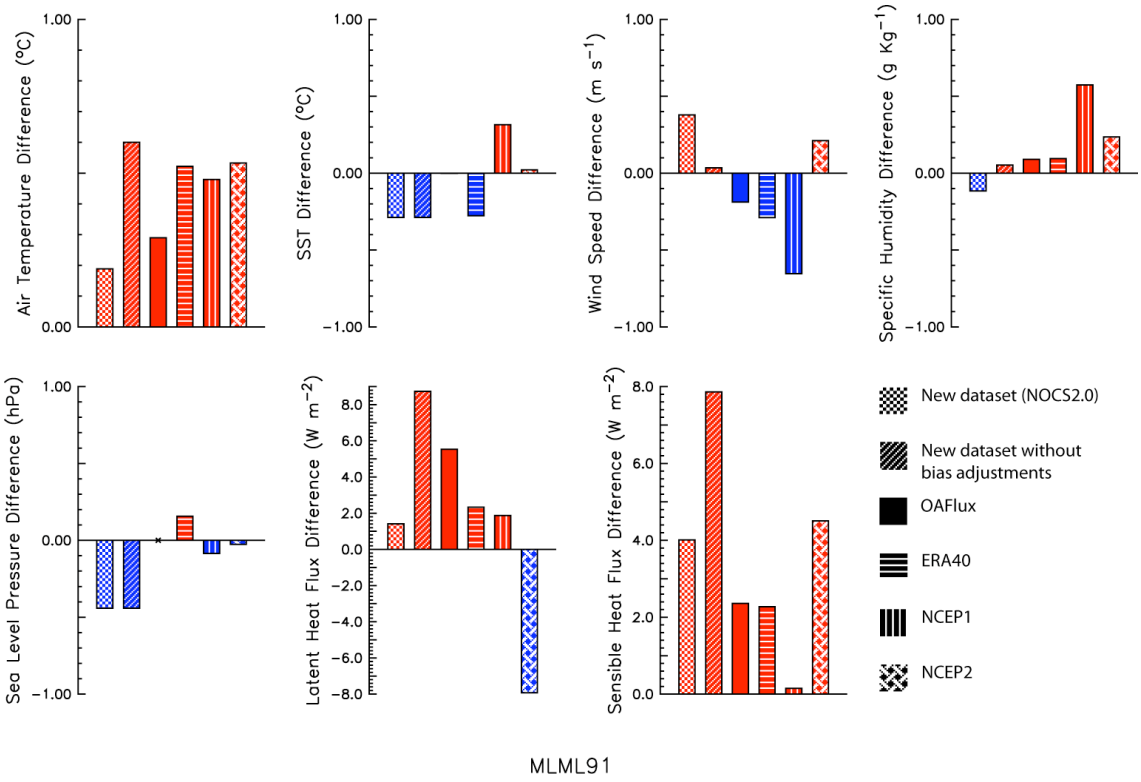

Figure 6-10: As Figure 6-8 but for the MLML91 mooring.

Table 6-5: Mean values (± standard deviation) of the buoy observations for the different Subduction buoy deployments and number of days with valid data. Also shown are mean differences (Δ) (product – buoy) from the different datasets and the standard deviation (σ) of the differences. Also listed for the OI are the total error estimates and t-statistic for each comparison. t-statistics marked with a star (*) are significant at the 95 % confidence level.

			#	OI			OAFlux	ERA40	NCEP1	NCEP2
	Buoy	Mean ± σ	Days	Δ ± σ	Total error	t-stat	Δ ± σ	Δ ± σ	Δ ± σ	Δ ± σ
Air Temperature	SW	23.5 ± 1.3	264	-0.1 ± 0.4	0.29	0.303	0.0 ± 0.4	0.0 ± 0.5	-0.5 ± 0.6	0.0 ± 0.4
	SE	21.8 ± 1.6	476	0.3 ± 0.6	0.28	1.145	0.4 ± 0.3	1.1 ± 0.5	-0.1 ± 0.5	0.4 ± 0.4
	CENT	22.1 ± 1.9	725	0.1 ± 0.7	0.29	0.252	0.0 ± 0.3	0.1 ± 0.5	-0.3 ± 0.5	0.1 ± 0.5
	NW	20.7 ± 2.8	395	0.0 ± 0.6	0.28	0.165	0.1 ± 0.4	0.2 ± 0.5	-0.2 ± 0.5	0.2 ± 0.5
	NE	19.8 ± 2.5	728	0.1 ± 0.6	0.28	0.489	0.1 ± 0.4	0.4 ± 0.6	-0.2 ± 0.4	0.1 ± 0.4
	Mean			**0.1 ± 0.6**			**0.1 ± 0.3**	**0.4 ± 0.5**	**-0.3 ± 0.5**	**0.2 ± 0.5**
SST	SW	24.2 ± 1.2	264	0.6 ± 0.7	0.23	2.512*	0.1 ± 0.3	0.2 ± 0.3	0.0 ± 0.3	-0.1 ± 0.3
	SE	22.5 ± 1.5	476	0.4 ± 0.7	0.19	1.907	-0.1 ± 0.4	0.2 ± 0.4	-0.2 ± 0.4	-0.1 ± 0.4
	CENT	23.0 ± 1.7	725	0.1 ± 0.7	0.20	0.595	0.0 ± 0.3	0.0 ± 0.3	-0.1 ± 0.2	-0.1 ± 0.3
	NW	21.7 ± 2.7	395	0.0 ± 0.6	0.19	0.162	-0.2 ± 0.4	-0.2 ± 0.5	-0.3 ± 0.4	-0.1 ± 0.5
	NE	20.9 ± 2.4	728	-0.2 ± 0.7	0.20	1.016	-0.2 ± 0.3	-0.3 ± 0.3	-0.2 ± 0.3	-0.2 ± 0.4
	Mean			**0.1 ± 0.7**			**-0.1 ± 0.3**	**0.0 ± 0.4**	**-0.1 ± 0.3**	**-0.1 ± 0.4**
Wind Speed	SW	7.3 ± 1.9	264	0.6 ± 1.8	0.47	1.372	0.1 ± 0.7	-0.1 ± 0.6	-0.3 ± 1.0	0.0 ± 1.2
	SE	8.1 ± 1.9	476	-0.1 ± 1.4	0.45	0.167	-0.2 ± 0.7	-0.8 ± 0.8	-1.0 ± 0.8	-0.6 ± 0.9
	CENT	6.4 ± 2.2	725	-0.1 ± 2.0	0.45	0.272	0.1 ± 0.7	-0.3 ± 0.8	-0.6 ± 0.9	-0.4 ± 1.1
	NW	6.1 ± 2.9	395	0.7 ± 1.6	0.45	1.575	0.6 ± 1.2	-0.2 ± 0.9	-0.5 ± 1.2	0.0 ± 1.3
	NE	6.3 ± 2.4	728	0.4 ± 1.6	0.45	0.849	0.4 ± 0.9	-0.5 ± 0.8	-0.3 ± 1.1	0.1 ± 1.2
	Mean			**0.2 ± 1.7**			**0.2 ± 0.8**	**-0.4 ± 0.8**	**-0.5 ± 1.0**	**-0.2 ± 1.1**

	Buoy	Mean ± σ	# Days	OI Δ ± σ	OI Total error	OI t-stat	OAFlux Δ ± σ	ERA40 Δ ± σ	NCEP1 Δ ± σ	NCEP2 Δ ± σ
Specific Humidity	**SW**	13.0 ± 1.6	264	0.6 ± 0.9	0.29	2.206*	0.4 ± 0.5	0.5 ± 0.6	0.8 ± 0.9	0.4 ± 0.7
	SE	12.8 ± 1.8	476	0.4 ± 0.9	0.29	1.336	0.0 ± 0.5	0.2 ± 0.6	0.2 ± 0.7	-0.2 ± 0.6
	CENT	12.2 ± 2.0	725	0.5 ± 1.2	0.29	1.629	0.1 ± 0.7	0.1 ± 0.8	0.6 ± 0.8	0.4 ± 0.8
	NW	11.7 ± 2.6	395	0.2 ± 0.8	0.28	0.825	0.1 ± 0.5	0.0 ± 0.6	0.6 ± 0.7	0.5 ± 0.7
	NE	10.7 ± 2.4	728	0.5 ± 1.0	0.28	1.635	0.2 ± 0.6	0.2 ± 0.7	0.8 ± 0.7	0.6 ± 0.7
	Mean			**0.4 ± 1.0**			**0.1 ± 0.6**	**0.2 ± 0.7**	**0.6 ± 0.7**	**0.4 ± 0.7**
Sea Level Pressure	**SW**	1017.0 ± 1.8	264	0.6 ± 1.4	0.32	1.901	0.6 ± 1.4	1.0 ± 0.5	0.5 ± 0.4	0.4 ± 0.5
	SE	1015.0 ± 1.6	476	0.4 ± 1.0	0.30	1.195	0.4 ± 1.0	1.0 ± 0.4	0.3 ± 0.4	0.3 ± 0.4
	CENT	1020.0 ± 3.0	725	0.0 ± 2.3	0.31	0.017	0.0 ± 2.3	1.0 ± 0.5	0.3 ± 0.5	0.2 ± 0.6
	NW	1024.0 ± 4.8	395	-0.4 ± 2.0	0.31	1.141	-0.4 ± 2.0	0.6 ± 0.6	-0.3 ± 0.7	-0.4 ± 0.7
	NE	1022.0 ± 5.3	728	0.4 ± 2.3	0.31	1.423	0.4 ± 2.3	1.2 ± 0.7	0.0 ± 0.8	-0.1 ± 0.9
	Mean			**0.2 ± 2.0**			**0.2 ± 2.0**	**1.0 ± 0.5**	**0.2 ± 0.6**	**0.0 ± 0.7**
Latent Heat Flux	**SW**	-151.1 ± 41.8	264	-14.2 ± 47.1	17.61	0.803	20.6 ± 20.1	5.6 ± 21.1	-6.4 ± 29.1	-12.4 ± 34.4
	SE	-114.4 ± 47.4	476	1.3 ± 43.1	17.12	0.074	11.1 ± 20.4	7.9 ± 20.7	-6.4 ± 25.4	-14.6 ± 26.5
	CENT	-120.2 ± 44.3	725	10.2 ± 45.2	16.89	0.606	12.9 ± 22.7	1.2 ± 20.8	7.2 ± 28.2	2.3 ± 31.5
	NW	-99.3 ± 59.3	395	-1.2 ± 34.3	16.94	0.072	8.6 ± 22.9	-0.4 ± 22.0	9.2 ± 25.8	1.5 ± 30.0
	NE	-110.6 ± 54.9	728	13.7 ± 38.6	16.87	0.811	13.1 ± 22.3	7.5 ± 19.0	8.7 ± 24.4	2.9 ± 28.5
	Mean			**5.3 ± 41.8**			**12.8 ± 22.0**	**4.4 ± 20.5**	**4.0 ± 26.4**	**-2.3 ± 29.9**
Sensible Heat Flux	**SW**	-7.4 ± 3.6	264	-7.9 ± 8.9	3.95	2.006*	-2.3 ± 4.8	-5.9 ± 5.5	-11.2 ± 10.4	-1.6 ± 7.2
	SE	-7.9 ± 5.7	476	-0.2 ± 8.0	3.55	0.059	4.7 ± 5.5	-0.3 ± 4.9	-1.7 ± 7.8	5.5 ± 5.7
	CENT	-7.4 ± 5.4	725	-0.5 ± 8.1	3.28	0.165	-1.3 ± 4.3	-4.5 ± 5.0	-4.4 ± 7.5	0.5 ± 6.2
	NW	-7.9 ± 9.9	395	-0.1 ± 7.2	3.41	0.024	0.7 ± 6.8	-3.8 ± 7.6	-1.9 ± 9.9	0.8 ± 8.8
	NE	-10.1 ± 9.1	728	2.7 ± 7.6	3.37	0.798	0.2 ± 5.1	-2.5 ± 5.6	-4.1 ± 8.4	1.0 ± 6.6
	Mean			**-0.3 ± 7.9**			**0.4 ± 5.2**	**-3.2 ± 5.7**	**-4.1 ± 8.5**	**1.4 ± 6.8**

6.3.3.3 Coastal Mixing and Optics Experiment (CMO) and Severe Environment Surface Mooring (SESMOOR) moorings

The Severe Environment Surface Mooing (SESMOOR) and Coastal Mixing and Optics Experiment (CMO) buoys were located in the Gulf Stream region of the North Atlantic. Whilst relatively well sampled, this region has strong spatial gradients and short time scales in the surface forcing fields. As a result, in this region we would expect the new dataset to have larger sampling errors and perform less well than over the Subduction and MLML91 regions shown in the previous two sections. Figure 6-11 and Figure 6-12 show the mean differences between the buoy observations and the different datasets at location of the CMO and SESMOOR buoys respectively. Table 6-7 summarises these results.

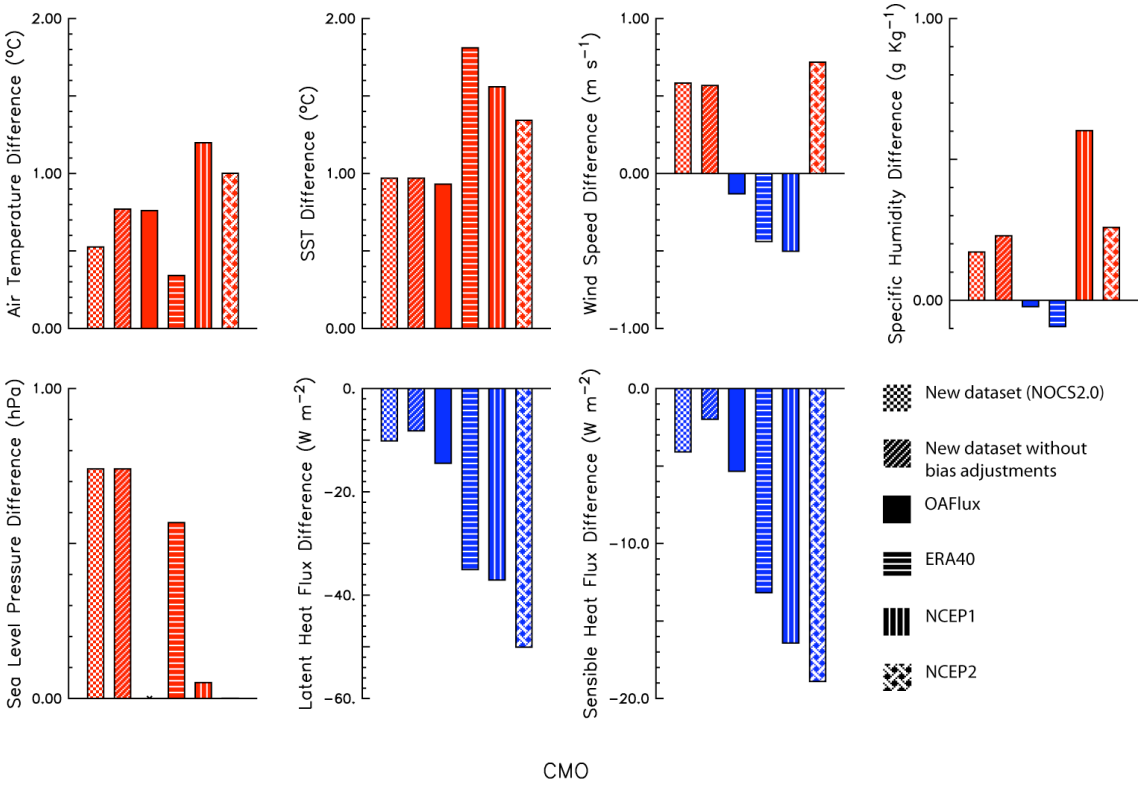

Figure 6-11: As Figure 6-8 but for the CMO mooring.

At both buoy locations, the air temperature and sea surface temperature are overestimated for all of the datasets shown. Large differences can be expected between the buoy observations and the gridded datasets which have spatial resolutions that are too coarse to capture the large horizontal gradients in temperature (both air and sea) seen in this region (e.g. Baumgartner and Anderson 1999). The mean differences for

SST in each dataset are larger than the respective air temperature differences, suggesting the air-sea temperature difference is also overestimated. Additionally, a warm bias in the SST will also lead to the humidity at the air – sea interface being overestimated and result in an overestimate of the humidity gradient in the lower atmosphere. The impact of this overestimate of the temperature and humidity gradients in the lower atmosphere can be seen in the heat flux estimates with the fluxes overestimated (i.e. more negative) for each of the products shown compared to the buoy estimates. This is true even in the cases where the wind speed is underestimated in the different datasets.

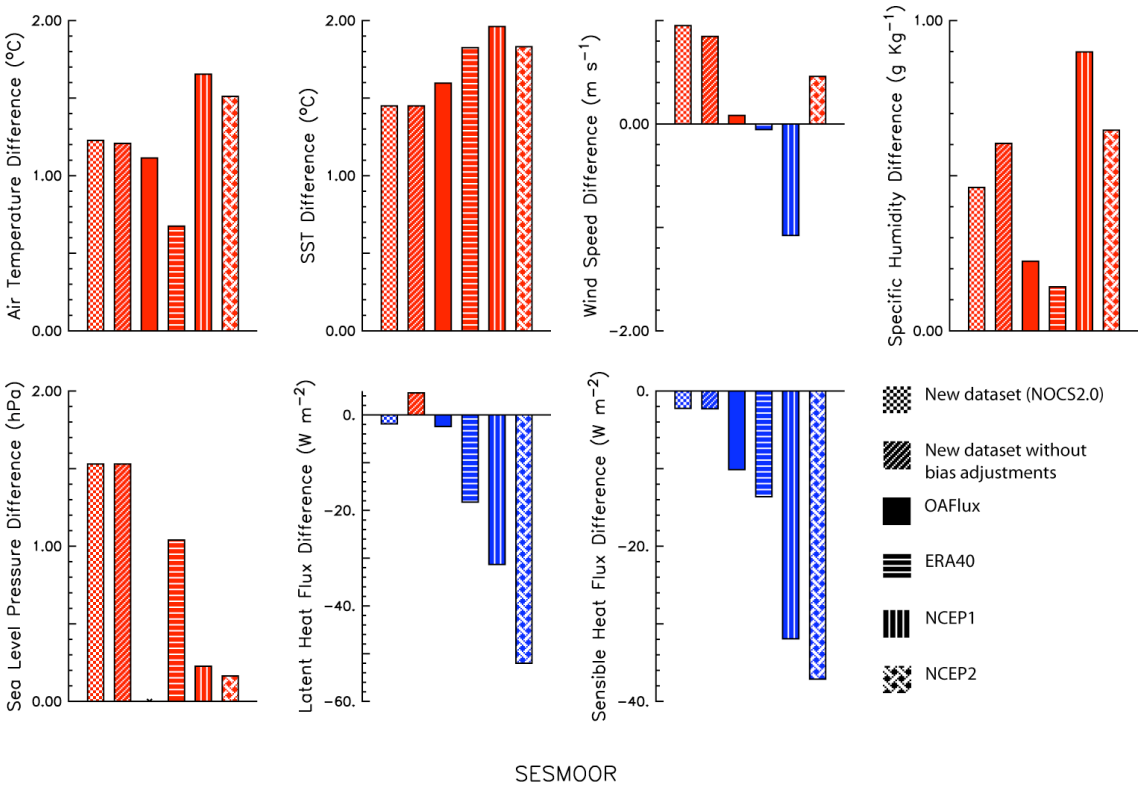

Figure 6-12: As Figure 6-8 but for the SESMOOR mooring.

Table 6-6: As Table 6-5 but for the MLML91 mooring

	Mean ± σ	# Days	OI			OAFlux	ERA40	NCEP1	NCEP2
			Δ ± σ	Total error	t-stat	Δ ± σ	Δ ± σ	Δ ± σ	Δ ± σ
Air Temperature	10.8 ± 2.1	122	0.2 ± 0.5	0.29	0.654	0.3 ± 0.4	0.5 ± 0.6	0.5 ± 0.4	0.5 ± 0.5
SST	11.4 ± 2.0	122	-0.3 ± 0.6	0.19	1.499	0.0 ± 0.4	-0.3 ± 0.4	0.3 ± 0.4	0.0 ± 0.5
Wind Speed	7.6 ± 2.6	122	0.4 ± 1.9	0.47	0.813	-0.2 ± 1.1	-0.3 ± 1.3	-0.7 ± 1.2	0.2 ± 1.4
Specific Humidity	6.8 ± 1.3	122	-0.1 ± 0.6	0.29	0.404	0.1 ± 0.3	0.1 ± 0.4	0.6 ± 0.4	0.2 ± 0.3
Sea Level Pressure	1012.0 ± 10.7	122	-0.4 ± 5.1	0.40	1.092	-0.4 ± 5.1	0.2 ± 1.2	-0.1 ± 1.2	0.0 ± 1.3
Latent Heat Flux	-43.4 ± 28.7	122	1.4 ± 22.5	17.05	0.083	5.5 ± 13.2	2.3 ± 11.8	1.9 ± 13.8	-7.9 ± 19.7
Sensible Heat Flux	-7.0 ± 8.0	122	4.0 ± 9.7	3.62	1.107	2.4 ± 5.6	2.3 ± 7.3	0.1 ± 8.4	4.5 ± 7.9

Table 6-7: As Table 6-5 but for the SESMOOR and CMO moorings

	Buoy	Mean ± σ	# Days	OI Δ ± σ	OI Total error	OI t-stat	OAFlux Δ ± σ	ERA40 Δ ± σ	NCEP1 Δ ± σ	NCEP2 Δ ± σ
AT	SESMOOR	4.8 ± 5.3	139	1.2 ± 3.3	0.29	4.239*	1.1 ± 3.3	0.7 ± 4.0	1.7 ± 3.4	1.5 ± 3.5
	CMO	9.2 ± 5.5	317	0.5 ± 1.8	0.29	1.843	0.8 ± 0.8	0.3 ± 1.2	1.2 ± 1.1	1.0 ± 1.2
	Mean			**0.7 ± 2.4**			**0.9 ± 1.9**	**0.4 ± 2.4**	**1.3 ± 2.1**	**1.2 ± 2.2**
SST	SESMOOR	7.4 ± 5.1	139	1.4 ± 2.6	0.43	3.335*	1.6 ± 2.2	1.8 ± 2.5	2.0 ± 2.1	1.8 ± 2.1
	CMO	9.9 ± 4.3	317	1.0 ± 1.3	0.31	3.145*	0.9 ± 0.8	1.8 ± 0.9	1.6 ± 1.2	1.3 ± 1.1
	Mean			**1.1 ± 1.8**			**1.1 ± 1.4**	**1.8 ± 1.6**	**1.7 ± 1.5**	**1.5 ± 1.5**
WSPD	SESMOOR	9.2 ± 3.3	139	0.9 ± 2.9	0.46	2.079*	0.1 ± 4.0	-0.1 ± 4.5	-1.1 ± 4.1	0.5 ± 4.7
	CMO	7.5 ± 3.0	317	0.6 ± 2.0	0.45	1.294	-0.1 ± 1.3	-0.4 ± 1.6	-0.5 ± 1.7	0.7 ± 2.1
	Mean			**0.7 ± 2.3**			**-0.1 ± 2.5**	**-0.3 ± 2.8**	**-0.7 ± 2.7**	**0.6 ± 3.1**
QAIR	SESMOOR	4.5 ± 1.9	139	0.5 ± 1.4	0.29	1.608	0.2 ± 1.4	0.1 ± 1.8	0.9 ± 1.6	0.6 ± 1.6
	CMO	6.6 ± 3.2	317	0.2 ± 0.9	0.28	0.601	0.0 ± 0.5	-0.1 ± 0.7	0.6 ± 0.6	0.3 ± 0.6
	Mean			**0.3 ± 1.1**			**0.1 ± 0.9**	**0.0 ± 1.1**	**0.7 ± 1.0**	**0.4 ± 1.1**
SLP	SESMOOR	1015.0 ± 9.4	139	1.5 ± 8.1	0.36	4.298*	1.5 ± 8.1	1.0 ± 10.5	0.2 ± 8.2	0.2 ± 8.2
	CMO	1016.0 ± 8.5	317	0.7 ± 6.6	0.32	2.337*	0.7 ± 6.6	0.6 ± 1.3	0.1 ± 1.1	0.0 ± 1.2
	Mean			**1.0 ± 7.1**			**1.0 ± 7.1**	**0.7 ± 5.9**	**0.1 ± 4.6**	**0.1 ± 4.6**
LHF	SESMOOR	-85.7 ± 95.0	139	-1.9 ± 73.1	18.01	0.104	-2.4 ± 76.8	-18.3 ± 90.1	-31.3 ± 88.8	-52.0 ± 99.0
	CMO	-38.5 ± 58.9	317	-10.2 ± 36.2	17.10	0.594	-14.5 ± 19.6	-35.0 ± 33.0	-37.1 ± 43.9	-50.1 ± 49.5
	Mean			**-7.6 ± 50.6**			**-10.8 ± 45.6**	**-29.9 ± 57.1**	**-35.3 ± 61.4**	**-50.6 ± 68.8**
SHF	SESMOOR	-40.3 ± 62.5	139	-2.2 ± 52.3	8.06	0.277	-10.1 ± 56.6	-13.6 ± 73.2	-31.9 ± 78.6	-37.1 ± 86.0
	CMO	-11.5 ± 39.2	317	-4.1 ± 26.3	5.33	0.768	-5.3 ± 15.7	-13.2 ± 26.5	-16.4 ± 38.5	-18.9 ± 47.0
	Mean			**-3.5 ± 36.4**			**-6.8 ± 34.0**	**-13.3 ± 46.2**	**-21.1 ± 54.1**	**-24.5 ± 61.7**

6.4 3-Way Error Analysis

So far, only the values of meteorological parameters and fluxes have been compared to other sources. In this section, the uncertainty estimates due to random and sampling errors from the OI are compared to estimates of these errors made using a series of 3-way error analyses.

In a 3-way error analysis, the error (or uncertainty) in each dataset can be estimated by calculating the variances of the differences between the different datasets and by partitioning the variances into contributions from each dataset (e.g. O'Carroll *et al*. 2008). For 3 different datasets, the error variances can be estimated by

$$\begin{aligned}
\sigma_1^2 &= \frac{1}{2}\left(V_{12} + V_{31} - V_{23}\right) + \left(r_{12}\sigma_1\sigma_2 + r_{31}\sigma_3\sigma_1 - r_{23}\sigma_2\sigma_3\right) \\
\sigma_2^2 &= \frac{1}{2}\left(V_{23} + V_{12} - V_{31}\right) + \left(r_{23}\sigma_2\sigma_3 + r_{12}\sigma_1\sigma_2 - r_{31}\sigma_3\sigma_1\right) \\
\sigma_3^2 &= \frac{1}{2}\left(V_{31} + V_{23} - V_{12}\right) + \left(r_{31}\sigma_3\sigma_1 + r_{23}\sigma_2\sigma_3 - r_{12}\sigma_1\sigma_2\right)
\end{aligned} \qquad \text{Eq. 6-1}$$

where σ_1^2, σ_2^2 and σ_3^2 are the error variances for the three data sets being compared; V_{ij} the variance of the differences between the i^{th} and j^{th} datasets; and r_{ij} the correlation between the errors in the i^{th} and j^{th} datasets.

In this study, four sets of 3-way error analyses have been performed at each buoy location (Table 6-8) with the values from the new dataset and the buoy observations used in each analysis. The buoy observations should be independent from the other datasets used, allowing the correlation terms in Eq. 6-1 that include the buoy data to be dropped as a first approximation. The gridded datasets are likely to contain correlations between the errors, either through using the same data or through sampling errors since we are comparing point estimates (i.e. the buoy observations) with gridded mean values. Discarding the terms that include the buoy correlations (dataset 2) but allowing for the correlation between the other two datasets (datasets 1 and 3) Eq. 6-1 then becomes

$$\begin{aligned}
\sigma_1^2 &= \frac{1}{2}\left(V_{12} + V_{31} - V_{23}\right) + r_{31}\sigma_3\sigma_1 \\
\sigma_2^2 &= \frac{1}{2}\left(V_{23} + V_{12} - V_{31}\right) - r_{31}\sigma_3\sigma_1 \\
\sigma_3^2 &= \frac{1}{2}\left(V_{31} + V_{23} - V_{12}\right) + r_{31}\sigma_3\sigma_1
\end{aligned} \qquad \text{Eq. 6-2}$$

where σ_1^2 is now the error variance for the new dataset; σ_2^2 the error variance for the buoy observations; σ_3^2 the error variance in the third dataset being compared; and r_{31} the correlation between the errors in the new dataset and the third dataset being compared.

The correlation between the errors in the new dataset and the other datasets (i.e. r_{31}) is not known. However, Eq. 6-2 can still be used to estimate the range of error values for the new dataset. Setting the correlation terms in Eq. 6-2 to zero and finding the 3-way error analysis that gives the lowest estimate of σ_2^2 will give the upper limit for the buoy errors. This will also give the lower limit for the errors in the new dataset. Assuming the buoy errors are zero for the same combination of 3 datasets will then give the upper limit for the error analysis.

Table 6-8: 3-way error analysis performed at each buoy location

Comparison	Dataset 1	Dataset 2	Dataset 3
1	NOCS	Buoy observations	OAFlux
2	NOCS	Buoy observations	ERA40
3	NOCS	Buoy observations	NCEP1
4	NOCS	Buoy observations	NCEP2

Table 6-9 summarises the results of the different 3-way error analyses for each buoy location and variable, comparing the error range from the different 3-way error analyses with the uncertainty estimates from the new dataset. For a number of cases negative error variances have been returned by Eq. 6-2 when the correlation terms have been set to zero. This indicates that significant correlations are present between at least two of the datasets in the 3-way analysis and that the results may be unreliable. The error estimates for these cases have been discarded.

The uncertainty estimates for the air temperature and SST from the OI dataset are larger than or close to the upper limits of the error estimates from the three-way error analyses at the buoy locations with low spatial variability (i.e. the Subduction buoys and MLML91 mooring), suggesting the uncertainties are being overestimated by the OI. For the buoy locations in the Gulf Stream region (SESMOOR and CMO) the uncertainty estimates are significantly lower than the estimates from the 3-way analysis suggesting that the uncertainties are being underestimated. These results will be discussed in further detail in Section 6.6. When the error estimates are compared for humidity similar results are seen. The uncertainty estimates for humidity at the majority of the buoy locations, including the SESMOOR mooring but excluding the CMO site, are equal to or larger than the lower limits from the 3-way error analyses and in several

cases larger than the upper limit. This suggests the uncertainties from the OI are either the right magnitude or slightly high compared to the 3-way error estimates. Only at the location of the CMO mooring do the results suggest an under estimate of the uncertainty in the humidity by the OI.

Table 6-9: Uncertainty estimates from the OI compared to the error estimates from the 3-way error analysis of the new dataset, the buoy observations and the dataset listed. The error range gives the error estimate from the 3-way analysis assuming no correlation between datasets and assuming zero error in the buoy observations.

Variable	Buoy	OI Uncertainty	3-way Error Range (min – max)
Air Temperature	SW	0.93	0.35 - 0.40
	SE	0.62	0.50 - 0.57
	CENT	1.12	0.63 - 0.67
	NW	0.51	0.43 - 0.56
	NE	0.62	0.49 - 0.56
	SESMOOR	0.7	1.53 - 3.33
	CMO	0.59	1.68 - 1.84
	MLML91	0.59	0.48 - 0.54
SST	SW	1.08	0.73 - 0.74
	SE	0.7	0.63 - 0.68
	CENT	1.06	0.59 - 0.66
	NW	0.57	0.47 - 0.62
	NE	0.68	0.63 - 0.69
	SESMOOR	0.84	1.21 - 2.59
	CMO	0.64	1.12 - 1.27
	MLML91	0.61	0.38 - 0.56
Specific Humidity	SW	1.26	0.79 - 0.95
	SE	0.76	0.65 - 0.87
	CENT	1.37	0.96 - 1.21
	NW	0.65	0.64 - 0.83
	NE	0.76	0.75 - 0.97
	SESMOOR	0.57	0.58 - 1.37
	CMO	0.57	0.82 - 0.93
	MLML91	0.59	0.48 - 0.62
Wind Speed	SW	2.02	1.79 - 1.82
	SE	1.21	1.32 - 1.42
	CENT	2.11	1.99 - 2.02
	NW	1.08	1.50 - 1.60
	NE	1.27	1.63 - 1.65
	SESMOOR	1.07	0.50 - 2.86
	CMO	0.93	1.96 - 1.99
	MLML91	1.36	1.90 - 1.90

Variable	Buoy	OI Uncertainty	3-way Error Range (min – max)
Sea Level Pressure	**SW**	1.84	1.39 - 1.43
	SE	1.08	0.98 - 0.98
	CENT	2.38	2.28 - 2.32
	NW	1.41	1.97 - 1.99
	NE	1.68	2.31 - 2.31
	SESMOOR	2.09	3.67 - 8.10
	CMO	1.59	N/A
	MLML91	2.69	N/A
Latent Heat Flux	**SW**	52.48	43.98 - 47.14
	SE	27.66	39.34 - 43.04
	CENT	46.23	40.13 - 45.22
	NW	22.48	28.95 - 34.25
	NE	25.75	33.10 - 38.64
	SESMOOR	24.95	28.90 - 73.07
	CMO	18.43	33.82 - 36.18
	MLML91	19.28	22.34 - 22.47
Sensible Heat Flux	**SW**	15.6	8.34 - 8.94
	SE	9.73	7.88 - 7.95
	CENT	13.51	N/A
	NW	7.12	7.12 - 7.23
	NE	8.35	7.31 - 7.58
	SESMOOR	16.53	26.69 - 52.32
	CMO	9.79	N/A
	MLML91	8.36	8.02 - 9.69

In contrast to the other basic variables, the uncertainty estimates for the sea level pressure and wind speed from the OI are significantly smaller than the estimates from the 3-way error analysis at the majority of the locations. The uncertainties are of a similar size or larger than the estimates from the 3-way error analyses only at the two Southern Subduction buoys and at the Central Subduction buoys. This suggests the uncertainty in these fields is underestimated in the OI. Additionally, at the location of the CMO and MLML91 moorings negative error variances are returned for the sea level pressure when the correlation terms are discarded, suggesting there may be significant error correlations between two or more of the datasets and that the results of the 3-way error analyses may be unreliable at these locations.

It should be noted that whilst the results for sea level pressure and wind speed have been included there may be serious problems with this analysis. Significant error correlations between the gridded datasets are likely to exist due to the assimilation of VOS wind speed and pressure observations into the reanalyses, which are then used in the OAFlux product. As a result we'd expect the correlation terms in Eq. 6-2 to be

significant. However, these are set to zero and this will lead to an underestimate of the error variances in the gridded datasets (σ_1^2 and σ_3^2) and an overestimate of the error variances for the buoy observations (σ_2^2). This can be seen by negative variances being returned for sea level pressure error estimates at the CMO and MLML91 mooring sites when the correlation terms are discarded. Even in the cases where positive error variances are returned there may be a serious underestimation of the errors in the gridded datasets and an overestimate for the buoy. Whilst the results for pressure and wind speed may be unreliable they have been included for completeness.

When the two different uncertainty estimates for latent heat flux are compared similar results to the wind speed comparisons are seen. For example, the uncertainties from the new dataset are larger at the location of the South West and Central Subduction buoys and lower at the other locations compared to the estimates from the 3-way error analyses. The uncertainty estimates in the latent heat flux from the new dataset are based on the uncertainty estimates in the basic variables and calculated using propagation of errors. The results shown above for wind speed and humidity suggest that the uncertainty estimates for humidity are either the right magnitude or slightly high whilst the wind speed uncertainties are underestimated. Hence, these results suggest that any under or overestimation is primarily controlled by the over and underestimation of the errors in the wind speed. Improving the wind speed estimates should in turn improve the latent heat flux uncertainty estimates. In contrast to the latent heat flux, the uncertainty estimates for the sensible heat flux are generally similar or larger in the output of the OI compared to the 3-way error analyses except at the location of the SESMOOR mooring. These results suggest the uncertainty estimates are dominated by the uncertainties in the air and sea surface temperatures, which appear to have been overestimated.

Overall, these results suggest the uncertainty estimates for the air and sea surface temperatures from the OI are overestimating the uncertainty due to random and sampling errors over the majority of the ocean and underestimating the errors over the Gulf Stream region. Similar results were shown for the humidity but with a smaller overestimation of the errors. The comparisons for wind speed and pressure are less conclusive due to problems with the comparisons. However, the results suggest there may be an underestimation of the uncertainties in wind speed and pressure.

6.5 Cross Validation

In this section an alternative approach to validating the uncertainty estimates is used. The observations used as input to the OI are sub-sampled and the OI re-run to produce different estimates of the error variances and daily fields. The OI has been run 10 times for the first five years (1970 – 1974) with 50 % of the observations randomly discarded each run. On average, we may expect 50 % of the observations used in any two runs to be the same and the expected correlation between data in the runs is 0.5. This will reduce the standard deviation across the ensemble by $\sqrt{0.5}$ (i.e. $\sqrt{(1-r)}$) in well sampled regions. Correcting for this, by multiplying the standard deviation by $\sqrt{2}$, the standard deviation across the ensemble of runs can be used as an alternate measure of the uncertainty in the daily fields in data rich regions.

In poorly sampled regions, the daily analysis fields from the OI will be strongly influenced by the first guess field. As a result, in these regions, the daily analysis values will be strongly correlated with each other across the ensemble of runs and the standard deviation will give a poor estimate of the uncertainty in the daily values. Hence, grid cells that are poorly sampled, indicated by a sum of weights in the OI of less than 0.5, have been discarded from this analysis. As a sensitivity test, ensembles for each variable have also been run excluding 70 % of data (r = 0.3), with the standard deviations corrected to account for the correlation between data (multiplying by $0.7^{-1/2}$) and the results compared to the runs excluding 50 % of the data. These results will be discussed in this section.

Figure 6-13 shows the ratio of the corrected ensemble standard deviation of the daily air temperature values to the mean daily uncertainty estimate averaged bimonthly for 1974. Regions where the ratio is greater than one indicate an underestimate of the uncertainty by the OI. In regions where the ratio is significantly less than one the output from the OI is likely to be overestimating the uncertainty. In regions with ratios less than but close to one the conclusions are less clear. This is due to the correlation between the different runs.

Excluding the western boundary current and North Atlantic storm track regions, the ratios shown in Figure 6-13 are significantly less than 1 suggesting an overestimate of the uncertainty by the OI. Over the western boundary current and North Atlantic storm track the ratio typically varies between 1 and 2 and peaks at over 2 where the

Labrador and North Atlantic Currents meet. This suggests that the OI underestimates the uncertainty in these regions. During the late northern spring and summer months there is a slight reduction in the region of underestimation, reaching a minimum area in June and July before increasing again in the autumn and winter months. Similar results are seen for the SST (Figure 6-14) and the other 4 years (not shown). When 70 % of the data are excluded similar results are also seen.

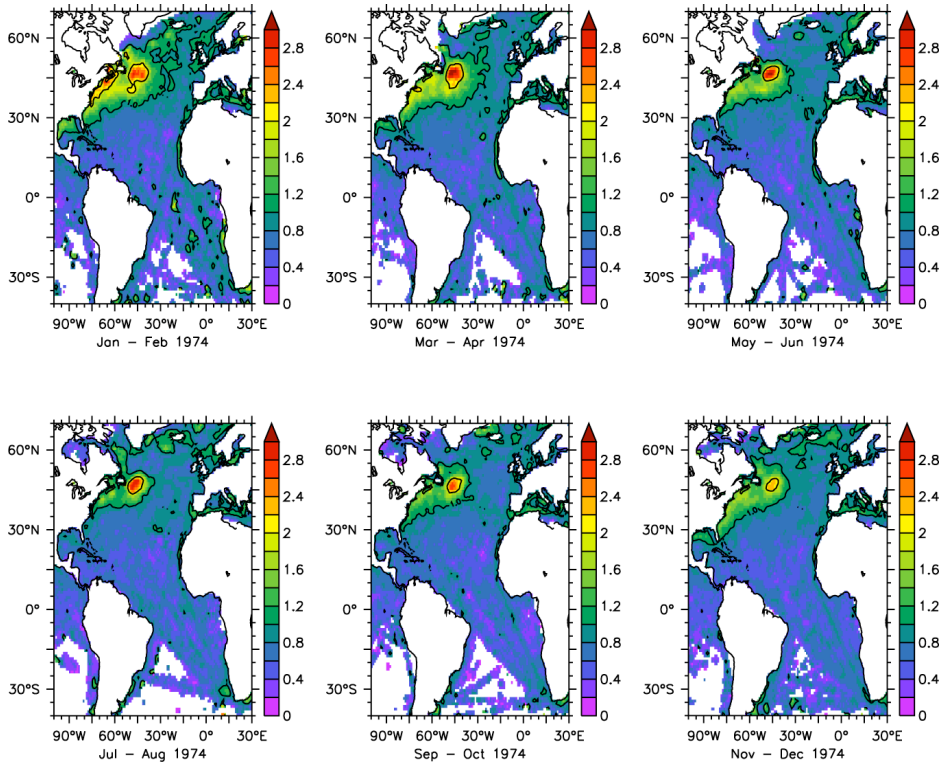

Figure 6-13: Ratio of the corrected standard deviation of the daily air temperature values across the ensemble of OI runs to the mean daily uncertainty estimate averaged for the two month periods shown. Values greater than 1 indicate an underestimate of the uncertainty by the OI whilst values less than 1 indicate an overestimate. The '1' contour is in bold.

When the ratio of the two different error estimates is calculated for humidity (Figure 6-15) similar results are seen but with a smaller area where the OI underestimates the uncertainty. For humidity, the region of underestimation is restricted to the coastal regions of the Gulf Stream and to where the North Atlantic and Labrador Currents meet. Similar results are seen for the other years and excluding 70 % of the data only has a relatively small impact. This suggests these results are relatively stable and that the uncertainty in the humidity is being overestimated by the OI in the majority of regions.

Figure 6-16 shows the ratio of the corrected ensemble standard deviation of the daily values to the mean uncertainty estimates for sea level pressure. Compared to SST and air temperature, the region of underestimation of the uncertainty by the OI is larger. During the more variable northern winter months the region of underestimation covers the majority of the North Atlantic north of approximately 35°N. This region of underestimation decreases throughout the year, reaching a minimum during the summer before increasing again towards the autumn and winter months. South of 35°N, the ratios are less than 1 throughout the year suggesting an overestimation of the uncertainty by the OI. Similar results are also seen for wind speed (Figure 6-17) but with a slightly smaller area of underestimation. During the northern summer this region of underestimation almost disappears. As with the other variables, when 70 % of the data are excluded for wind speed and SLP similar results are seen, suggesting the results presented in this section are relatively stable.

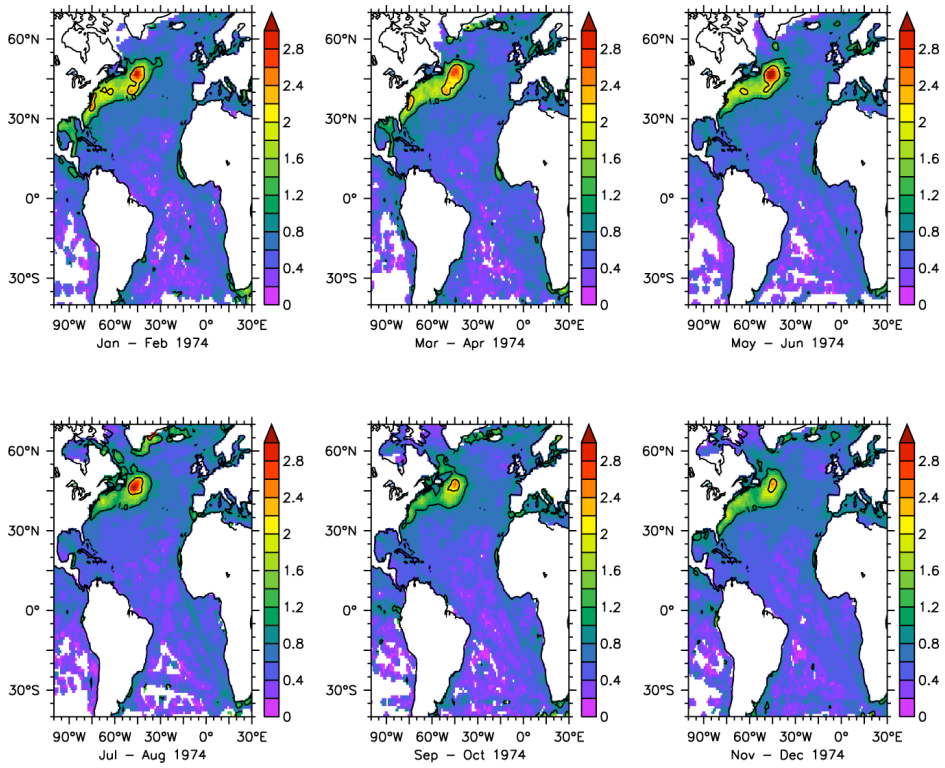

Figure 6-14: As Figure 6-13 but for sea surface temperature

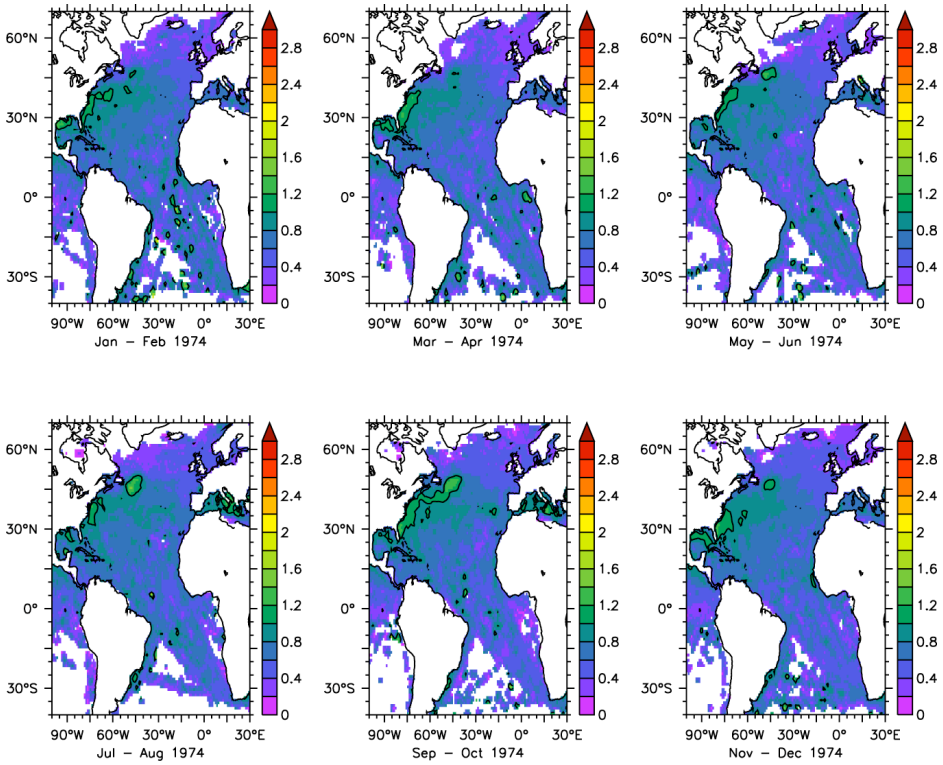

Figure 6-15: As Figure 6-13 but for specific humidity

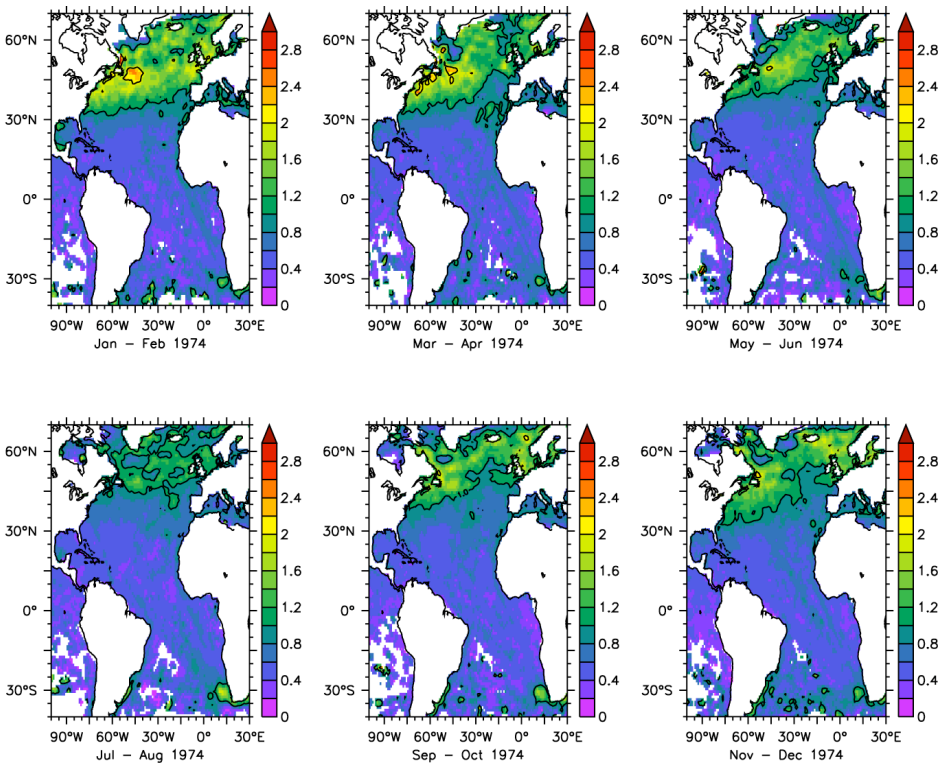

Figure 6-16: As Figure 6-13 but for sea level pressure.

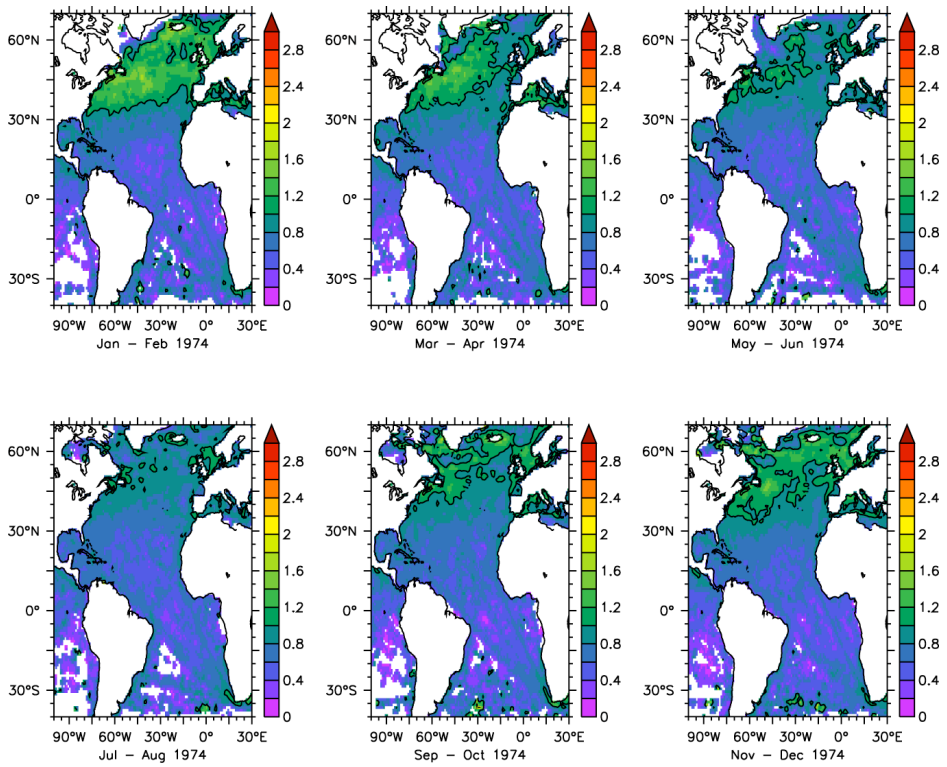

Figure 6-17: As Figure 6-13 but for wind speed

6.6 Discussion

6.6.1 Comparisons of mean differences

A comparison of the daily OI values to the original VOS observations and to independent buoy observations has been made. The comparison to the original data showed the OI to be unbiased compared to the input data and also highlighted the impact of random errors. The OI truncates the error distributions at either extreme and its implicit smoothing also reduces the extremes; this leads to offsets that need to be taken into account when interpreting the results. It should be noted that this comparison is not independent as the same observations are used in both the comparison and the OI.

To validate the OI fields, the buoy observations from the WHOI UOP data archive have been used as an independent source for comparison. Generally, the estimates from the new dataset were shown to be unbiased with respect to the buoy observations, with the mean differences within the total accuracy estimates of the two datasets. The air temperature and humidity bias corrections were found to improve the comparisons, giving further confidence in the corrections. As noted in Chapter 5, the air

temperature correction could potentially over correct the air temperature observations during the daytime. There is no evidence of this in the results shown.

In contrast to the air temperature and humidity bias adjustments, the bias adjustments made to the visual wind speed estimates typically increase the mean differences between the OI and the buoy data. This increase is due to the adjustment made to account for systematic biases in the WMO1100 scale following Lindau (1995). This adjustment has been designed to bring the visual estimates into agreement with height adjusted anemometer measurements by increasing low to moderate visual wind speed estimates (< 12 ms^{-1}) and decreasing higher wind speed estimates. Indeed at 6 of the 8 buoy locations examined in Section 6.3 comparisons of VOS visual and anemometer wind speeds show improved agreement between visual and anemometer mean wind speeds following application of the Lindau (1995) adjustment. At the buoy locations average winds speeds are all less than 10 ms^{-1}, and the adjustment of Lindau (1995) therefore increases the mean. As the height adjusted anemometer winds speeds are typically greater than those measured by the buoys, the change in the visual winds acts to worsen agreement of the OI wind speed with the buoy measurements. This result is consistent with Thomas *et al*. (2005) who reported adjusted VOS wind observations to be typically order 6% higher than observations from an operational moored buoy. This further highlights the requirement for further research on marine winds.

Whilst the fields were generally shown to be unbiased, there were some notable differences between the OI and the buoy data. The humidity estimates from the OI were found to be generally higher than the buoy measurements, confirming the results of previous authors (e.g. Kent *et al*. 1993b). However, these differences were still within the combined uncertainty range of the buoy observations and the new datasets. For example, the accuracy of the new dataset, ignoring random and sampling errors, and the accuracy of the buoy observations have both been estimated at 0.2 g Kg^{-1}, giving a total error estimate of 0.28 g Kg^{-1} when combined. Given the large number of daily values the critical t value for a two tailed Students t-test approaches 2.25 for the 95% confidence interval. As a result, only absolute mean differences greater than 0.63 g Kg^{-1} are significantly different from zero, suggesting the results are within the combined accuracy of the two datasets.

In addition to the small moist bias in the new dataset, large mean differences were found over the Gulf Stream region for the temperature variables due to large

interpolation errors and the inadequate resolution of the new dataset to capture the high spatial gradients in this region. It may be possible to increase the resolution of the OI of the VOS observations in this region since it is relatively well sampled. However, further work is required to characterise the spatial and temporal variability of the fields and error estimates in this region (see next Section).

A number of additional gridded datasets (OAFlux, ERA40, NCEP1 and NCEP2) were also compared to the buoy observations. Comparable results were generally seen across the different datasets, with generally small mean differences at the different buoy locations. These datasets also performed poorly over the Gulf Stream region, with similar interpolation errors in the temperature estimates at the location of the CMO and SESMOOR buoys. Additionally, the estimates of the wind speed from the NCEP1 and ERA40 reanalysis models were generally lower than the buoy estimates suggesting an underestimate by these models. This generally supports the conclusions of previous authors in comparisons of the reanalysis wind speeds with VOS observations (Thomas *et al*. 2008) and research vessel observations (Smith *et al*. 2001).

6.6.2 Comparisons of error estimates

Two different methods have been used to examine the random and sampling uncertainty estimates in the new dataset. The first makes use of a series of 3-way error analyses, estimating the errors in the new dataset, buoy observations and the third comparison dataset used (either OAFlux, ERA40, NCEP1 and NCEP2). The results generally show the error estimates in the new dataset to be high compared to the error estimates from the 3-way analyses with the exception of observations over the Gulf Stream region. Over the Gulf Stream region the error estimates tended to be underestimated by the OI. The second method sub-sampled the VOS observations and re-ran the OI a number of times. The standard deviation of the daily values from the ensemble, corrected to account for the correlation between runs, was then used as an alternative estimate of the uncertainty and compared to the mean uncertainty estimates from the OI. The results were generally consistent with the 3-way error analyses, with the uncertainty typically overestimated by the OI over the majority of the oceans and underestimated over the Gulf Stream region (and North Atlantic Storm track for wind speed and pressure).

The general agreement between the two comparisons (i.e. the uncertainties are under- or overestimated by the OI in similar regions) gives confidence in the results and suggests that the OI is generally over estimating the uncertainty outside regions of high variability and underestimating in the more variable regions. This over- and underestimation is due to the choice of time and space scales used in the OI. For example, the choice of 3 days (Section 4.6) for the time scales over the Gulf Stream region is likely to be too long and will result in the uncertainty in the first guess due to sampling errors being underestimated. In low variability regions the uncertainty due to sampling errors is likely will be overestimated. Improving the space and time scales used should therefore improve the uncertainty estimates and is discussed further in Chapter 8.4.

6.7 Summary

In this Chapter a quantitative assessment of the daily fields and uncertainty estimates from the OI has been made together with an assessment of the fluxes. The daily fields from the interpolation have been shown to be unbiased with respect to the input data, with negligible mean differences and RMS errors similar in magnitude to the random errors in individual VOS observations. A comparison of the daily fields and flux estimates from the new dataset to independent buoy observations has shown the basic variables and fluxes not to be significantly biased compared to the buoy observations. Small mean differences were found between the buoy observations and the estimates from the new dataset. These were generally within the combined accuracy of the two datasets. The impact of the bias adjustments has been shown, for both the air temperature and humidity variables and also the fluxes, with the adjustments bringing the new dataset and buoy observations into better agreement. For comparison, mean differences between the buoy observations and a number of similar gridded datasets (OAFlux, ERA40, NCEP1 and NCEP2) were also calculated and similar sized differences found.

The assessment of the uncertainty estimates from the OI used two different methods, comparing the results to a 3-way error analysis and performing a series of cross validation experiments, with similar results. The estimates of the random and sampling uncertainty in the meteorological parameters from the OI tend to be too low in regions of high variability, such as the Gulf Stream, and too high in regions of lower

variability. This over- and underestimation of the uncertainty in the basic variables also leads to an over- and underestimation of the uncertainty in the fluxes in the low and high variability regions respectively. Refining the temporal and spatial scales used in the OI will improve the error estimates.

Overall, the results presented in this Chapter give confidence in the new dataset. The OI does not give biased results compared to the input data and does not increase the random noise in the gridded fields. When compared to independent buoy measurements, with a few exceptions, the fields are unbiased. In the locations where biases are reported these are generally due to interpolation errors and similar results are seen in comparator datasets. The adjustments applied to the observations have been shown to increase the agreement between the new dataset and the buoy observations for air temperature and humidity, giving confidence that the adjustments have improved the dataset, both in terms of bias and random errors. When the uncertainty estimates from the OI are compared to estimates made using a series of 3-way error analyses and a series of cross validation experiments they are found to be generally of the right magnitude.

7 Monthly means, variability and trends

7.1 Introduction

The OI used to grid the observations and the adjustments applied will have an impact on both the variability and the trends observed in the new dataset (e.g. Sterl 2001). The aim of this section is to demonstrate that the methods used have improved the quality of the monthly mean estimates, both of the basic meteorological variables and of the fluxes, and hence the trends and variability within the dataset. It is expected that the careful consideration of data uncertainty and bias used in the construction of the new dataset will lead to improved dataset characteristics. Spurious trends should be reduced and noise suppressed whilst the real variability is maintained. This will be tested, assessing the improvements relative to a simple dataset based on the arithmetic mean of the observations used in the OI (calculated on a 1 degree grid) and to an extended version (1980 – 2002) of the first SOC climatology (Josey *et al*. 1999). In this Chapter the dataset based on the monthly arithmetic mean will be referred to as the gridded data.

7.2 Variability

7.2.1 Comparison with the gridded dataset

The first comparison in this Section examines whether the variability in the new dataset is more realistic than that in the gridded data by comparing the intermonth variability, defined as the standard deviation across calendar months. Figure 7-1 shows the average ratio of the variability from the new dataset to that from the gridded data using only grid cells with monthly mean values in both datasets. Similar results are also seen for the individual calendar months (not shown). In the poorly sampled regions the contribution by random errors (i.e. noise) to the variability in the gridded data is largest due to the low number of observations and the variability is therefore overestimated. In well-sampled regions the contribution to the variability by random errors will be small. As a result, the OI has the largest impact on the variability in the poorly sampled regions, with a larger reduction in random errors, and hence variability, in these regions. This can be clearly seen in Figure 7-1 with the smallest ratios (i.e. largest reduction in variability) in poorly sampled regions and ratios approaching unity over the relatively

well-sampled shipping lanes for air temperature, humidity and SST. The ratios for fluxes are smaller, i.e. the variability is less in the new dataset, due to the combined impact of reducing the random errors in the other variables.

Whilst the variability is reduced in the new dataset compared to the gridded data this could be due to a loss of part of the true variability by excessive smoothing in the interpolation process. To test this, the correlation between sea surface temperature and air temperature has been examined. If the decrease in variability seen in the new dataset is due to smoothing we would expect to see an increase in the correlation everywhere relative to the 1° gridded dataset, including in those regions where the variability is high and the correlation known to be smaller. If only the noise is being reduced we would also expect to see an increase in the correlation. However, we would expect this increase to be much less in high variability regions.

Figure 7-2 shows the correlation between the SST and air temperature anomalies for the new dataset. The correlation is also shown for the 1 ° gridded data, unsmoothed and smoothed over 2° and 5° using a two dimensional running mean filter. The impact of the smoothing on the correlation in the gridded data can clearly be seen, with the correlation increasing as the scales are increased over all regions. This includes the Gulf Stream region where we know the variability to be higher and the correlation lower. In the new dataset the correlation is also increased compared to the 1° gridded data and similar in magnitude to that seen the data smoothed on a 2° scale over the majority of the Atlantic. However, in the Gulf Stream region the increase is much less, suggesting the decrease in variability seen in the new dataset is not due to an excessive smoothing by the OI.

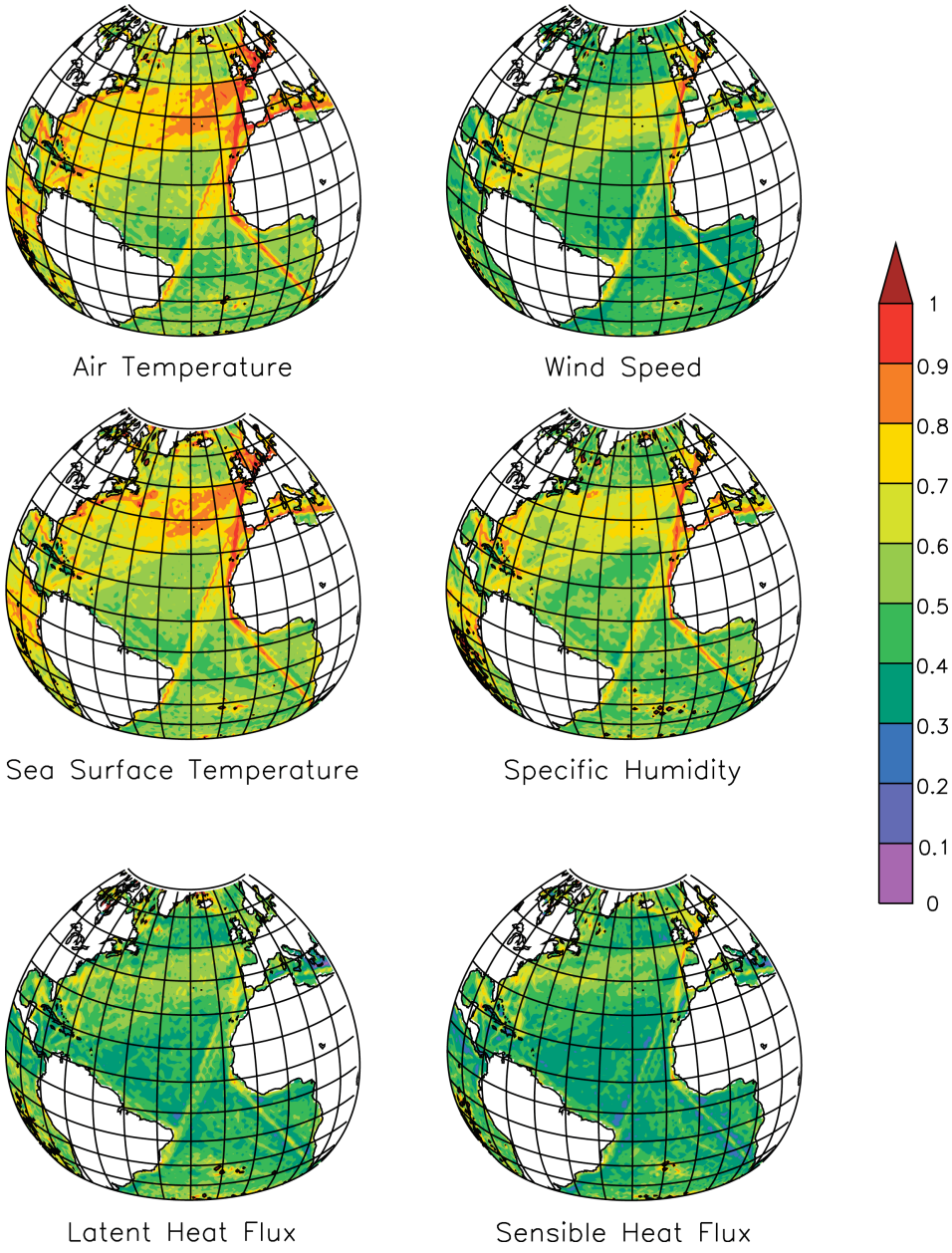

Figure 7-1: Ratio of intermonth variability from the new dataset to the variability from the gridded data averaged over the calendar months for 1970 – 2006.

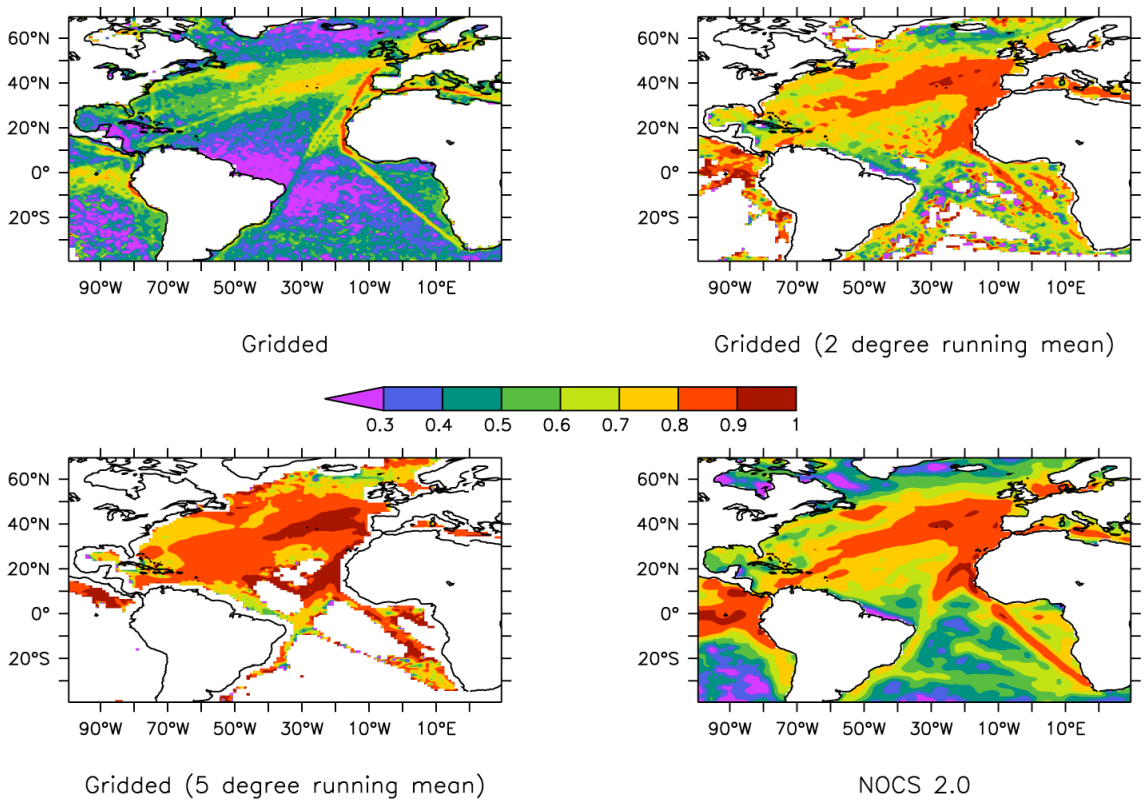

Figure 7-2: Correlation between air temperature and SST anomalies from the gridded data, smoothed on a number of different scales, and NOCS 2.0.

7.2.2 Comparison with the SOC Climatology

The second comparison in this section compares the variability in the new dataset to that in the SOC climatology. Figure 7-3 shows the average ratio of the intermonth variability from the new dataset to that from the SOC climatology. The ratios for wind speed, air temperature and humidity are typically smaller than one over the Atlantic, indicating the variability is smaller in the new dataset for these variables. For wind speed and humidity the impact is greatest in poorly sampled regions and smallest over the shipping lanes, similar to the comparisons of the new dataset to gridded data. For air temperature the difference between regions is less clear. Instead the variability in the tropics is much lower in the new dataset compared to the SOC climatology. This lower variability is due to erroneous air temperature values in the SOC climatology, causing the variability estimates from the SOC climatology to be overestimated. This will be demonstrated later.

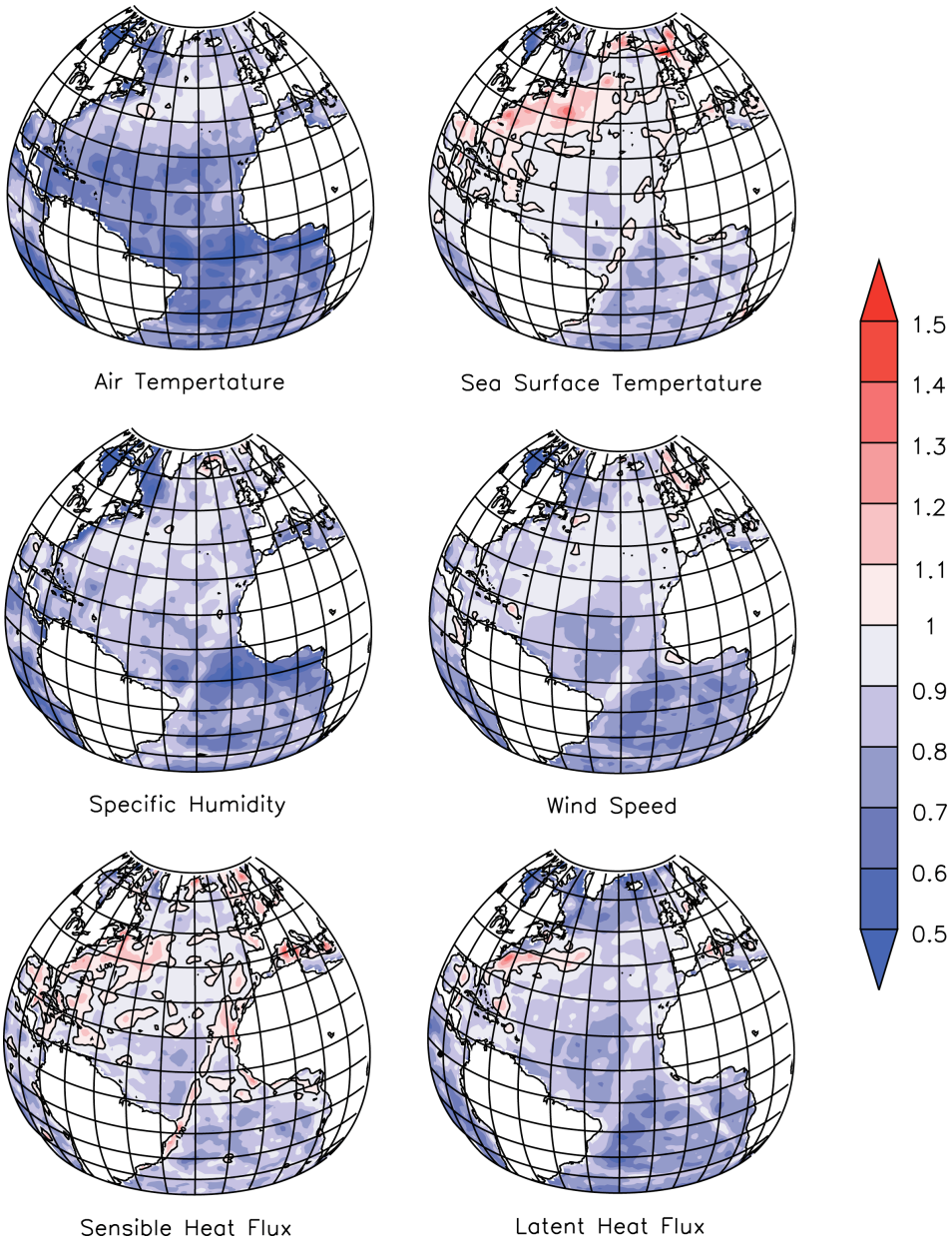

Figure 7-3: Ratio of intermonth variability from the new dataset to the variability from the SOC climatology averaged over the calendar months. The contour line indicates a ratio of 1.

The variability in the new dataset is also lower for the SST and fluxes compared to that in the SOC climatology in the poorly sampled regions. However, in contrast to the other variables, the variability is higher in a number of regions in the new dataset. Over the Gulf Stream and North Sea regions the intermonth variability for SST is higher in the new dataset. Additionally, the variability is slightly higher in the shipping lanes between Europe and Africa and South America. The impact of this higher variability in SST is seen in the variability of the sensible heat flux, and to a lesser degree, the latent

heat flux. This lower variability in the SOC climatology is due to the successive correction used by Josey *et al*. (1999) and that has been previously shown to artificially reduce the observed variability in high variability but well sampled regions (e.g. Kent *et al*. 2000).

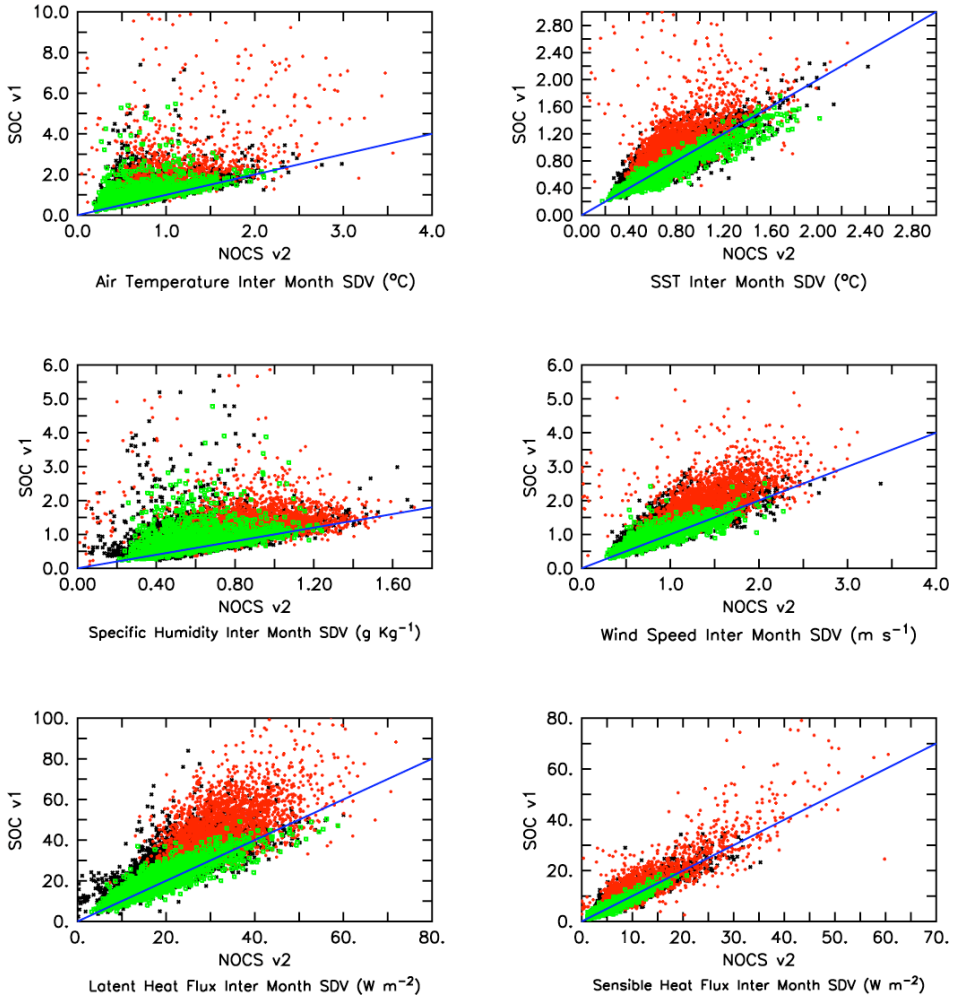

Figure 7-4: Comparison of intermonth variability of the basic variables between the NOCS 2.0 and the SOC Climatology. Results are for the Atlantic (40S – 70N, 100W – 30E) and the period 1980 – 2005. Well-sampled regions are shown in green, poorly sampled regions in red and intermediate regions in black. The blue line indicates a one to one relationship (equality).

In order to show the impact of the sampling on the variability comparisons Figure 7-4 shows a scatter plot of the intermonth variability estimates from the new dataset plotted against those from the SOC climatology for each grid cell and calendar month. Well sampled regions have been selected based on the grid cells with mean

uncertainty values in the lower quartile of all grid cells. Similarly, poorly sampled regions have been selected using grid cells with mean uncertainty estimates in the upper quartile. The impact of the sampling is clear in these plots, with the poorly sampled regions typically having higher variability in the SOC climatology. In contrast, the variability estimates in well sampled regions are more similar between the two datasets and scattered along the line of equality.

The higher variability in the SST from the new dataset over the Gulf Stream seen in Figure 7-3 can also be seen in this figure, with higher variability estimates from the new dataset in the well sampled regions. For both air temperature and humidity there are a significant number of points in well sampled regions where the variability in the SOC climatology is much higher than that seen in the new dataset. This is caused by outliers in the SOC climatology skewing the variability estimates, an example of which is shown in Figure 7-5. These outliers are thought to be caused by problems with the height corrected data under low wind speed conditions in the SOC climatology before gridding (Elizabeth Kent, pers. comm.). Whilst the same process is used in the SOC climatology to calculate the height corrected wind speed and fluxes no evidence for similar outliers in these variables could be found.

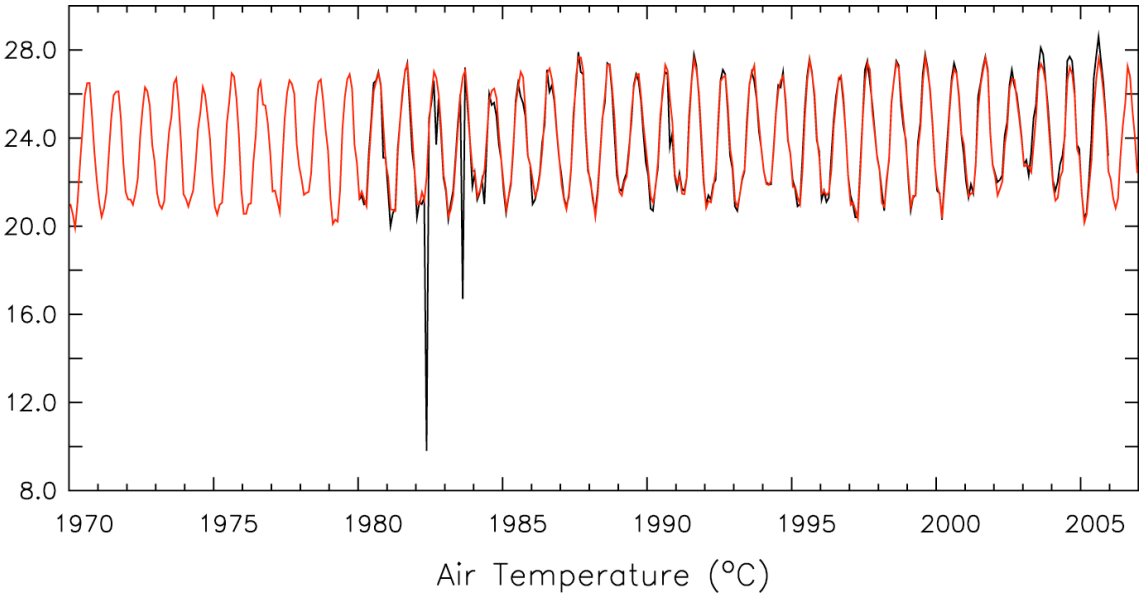

Figure 7-5: Time series of air temperature estimates at 27.5 N 54.5 W from the SOC climatology (black) and NOCS 2.0 (red).

7.2.3 Geophysical Consistency

The final test examining the variability in the new dataset examines the geophysical consistency, in this study defined as the correlation between the total turbulent heat flux (latent + sensible heat flux) and the SST tendency (i.e. $dSST/dT$) (e.g. Cayan 1992). Other authors have used the correlation between the latent heat flux and SST tendency (e.g. Sterl, 2001) as a measure of the geophysical consistency. A strong correlation between the total turbulent heat flux and the SST tendency has been previously shown by Cayan (1992) for the North Atlantic and, as noted by Sterl (2001), it is reasonable to expect the same relationship to hold for the South Atlantic, with stronger fluxes leading to a cooling of the SST.

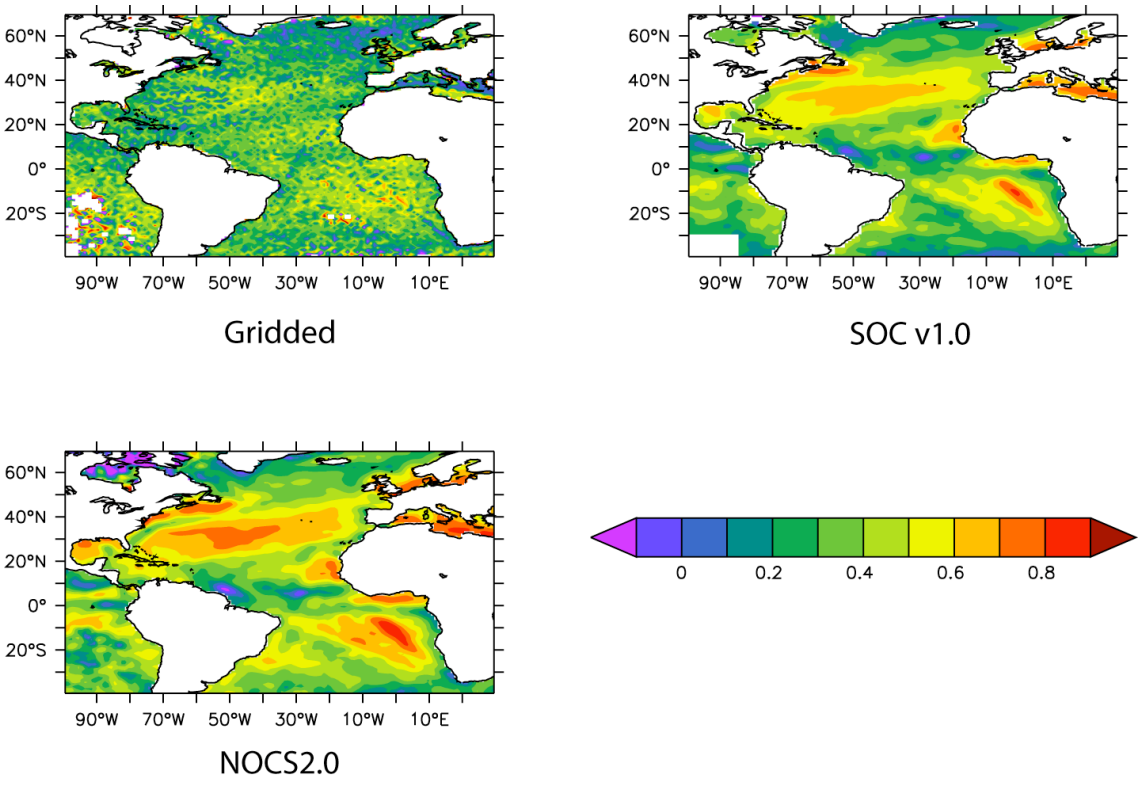

Figure 7-6: Correlation between the surface turbulent heat flux (latent + sensible) and the SST tendency from the: gridded data (top left); SOC climatology (top right); and new dataset (NOCS 2.0, bottom left). The correlations have been calculated over the period 1980 to 2002.

Figure 7-6 shows the correlation between the heat fluxes and the SST tendency estimated from: the gridded data, the SOC climatology, and the new dataset. The improvements in the geophysical consistency in the new dataset are clear, with an increase in correlation over most regions compared to both the SOC climatology and the

gridded data. Regions where we would expect the correlation to be lower, such as upwelling zones and the Gulf Stream, are clearly visible. This gives confidence that the increased correlation is not due to a smoothing of the data. However, whilst the correlation is increased in the new dataset, the correlation in the South Atlantic is still thought to be too low, especially south of 20 S and close to the South American coast. Improving the estimates of the correlation between first guess errors (see Chapter 4.9) should help to improve the consistency in this region. The spatial variability in this region is relatively low, and as a result, the correlation between first guess errors is likely to be underestimated through the use of the 300 km length scales in the OI. Increasing the length scales would increase the error correlations and the amount of data used from nearby locations.

7.3 Trends and Impact of Adjustments

7.3.1 Introduction

Chapter 4 described the adjustments applied to the individual VOS observations to take into account biases, changing observing practices and measurement height. A review of the adjustments in the context of their impact on the trends is given in this section.

Air Temperature

Both a bias adjustment and a height adjustment have been applied to the VOS observations to account for radiative heating errors during the daytime and the changing observing height respectively. The coefficients used in the adjustment for daytime heating errors have been fitted for each year (Appendix C), allowing them to vary systematically with time. As a result the adjustment may also vary systematically with time through changes in instrument exposure (e.g. Berry and Kent 2005) or changing environmental conditions, such as increasing or decreasing cloud cover. In addition to a bias adjustment that may vary in time, the height adjustment is known to vary systematically due to the increase in observing height. This increase in observing height, about 7 m over 1970 – 2006, would lead to an increase in the required adjustment of approximately 0.02 °C decade^{-1} based on the adiabatic lapse rate. The actual increase will depend on the stability dependent height correction being applied.

Specific Humidity

Two sets of adjustments have been applied to the humidity observations to account for biases in screen observations and the changing observing height. The impact of the bias adjustment depends on the proportion of observations made using screens compared to other methods. There are only small changes to this proportion over the period of the new dataset and as a result the impact of the bias adjustment is not expected to vary systematically in time and should not affect the trend estimate. The height adjustment, as with the air temperature height adjustment, will vary systematically with time and is expected to increase the observed trend in specific humidity over time.

Wind Speed

Several different adjustments have been applied to the wind speed observations. Two different adjustments have been applied to the visual winds, the first to take into account biases in the Beaufort Equivalent Scale (BES) used to report the visual estimates and the second to correct for the increasing influence of anemometers on the visual measurements. Both of these systematically change with time. In a previous study, Thomas *et al*. (2008) showed the adjustment made to account for the biases in the BES to reduce the trend by 0.02 ms^{-1} $decade^{-1}$ over the period 1982 – 2002. The second adjustment to the visual winds is expected to have a larger impact, reducing the trends by order 0.1 ms^{-1} $decade^{-1}$ over 1970 - 2006. This is based on a 5 % reduction being applied to the wind speeds incrementally over a 15 year period, ranging from 0 % in 1985 to 5 % in 2000, and a mean wind speed of around 8 ms^{-1}. The height adjustment to the anemometer winds is also expected to have a significant impact, with Thomas *et al*. (2008) reporting a reduction in trend of 0.12 ms^{-1} $decade^{-1}$over the period 1982 – 2002 for height corrected anemometer wind speed measurements.

7.3.2 Trend Estimation

Time series of the meteorological variables and fluxes over the oceans contain low and high frequency variability together with long term trends. The low frequency variability, such as that due to the North Atlantic Oscillation (NAO), will lead to autocorrelated errors when estimating linear trends in the time series. If ordinary least squares (OLS) regression is used to estimate the trends, the autocorrelation will cause

the degrees of freedom to be overestimated and the uncertainty in the fitted coefficients to be underestimated. In order to account for any autocorrelation in the time series, and to produce unbiased estimates of the confidence intervals, linear trends have been fitted to the time series' using the Cochrane-Orcutt (CO) procedure (e.g. Thejll and Schmith 2005). It should be noted that other methods, such as the Restricted Maximum Likelihood (REML) method (e.g. Rayner *et al*. 2003), are available for taking the autocorrelation into account. The CO procedure has been chosen due to the ease with which it can be implemented.

In the CO procedure the errors terms in the regression are assumed to be well represented by a first order autoregressive process (AR1), i.e.

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \qquad \text{Eq. 7-1}$$

$$\varepsilon_t = \rho_{ac1}\varepsilon_{t-1} + u_t \qquad \text{Eq. 7-2}$$

where Y_t and X_t are the dependent (i.e. anomaly time series) and independent (time) variables respectively at time t; β_0 and β_1 the intercept and slope regression coefficients; ε_t the error term (or residual) from the regression; ρ_{ac1} the lag 1 autocorrelation in the error terms; and u_t an independent normally distributed error. The autocorrelation in the error terms is removed by transforming the dependent and independent variables using

$$X_t' = X_t - \rho_{ac1}X_{t-1} \qquad \text{Eq. 7-3}$$

$$Y_t' = Y_t - \rho_{ac1}Y_{t-1} \qquad \text{Eq. 7-4}$$

where X_t' and Y_t' are the transformed variables. An OLS regression is then performed using the transformed variables, i.e.

$$Y_t' = \beta_0' + \beta_1' X_t' + u_t \qquad \text{Eq. 7-5}$$

The untransformed regression coefficients and their uncertainties (i.e. standard deviations) are then given by

$$\beta_0 = \frac{\beta_0'}{1-\rho_{ac1}} \qquad \text{Eq. 7-6}$$

$$\beta_1 = \beta_1' \qquad \text{Eq. 7-7}$$

$$s\{\beta_0\} = \frac{s\{\beta_0'\}}{1-\rho_{ac1}} \qquad \text{Eq. 7-8}$$

$$s\{\beta_1\} = s\{\beta_1'\} \qquad \text{Eq. 7-9}$$

where $s\{\beta_0'\}$ and $s\{\beta_1'\}$ are the standard deviations of the OLS coefficients from Eq. 7-5. It should be noted that the uncertainties in the regression coefficients are only due

to regression model errors and do not take into account any uncertainty due to changing observing practices or sampling.

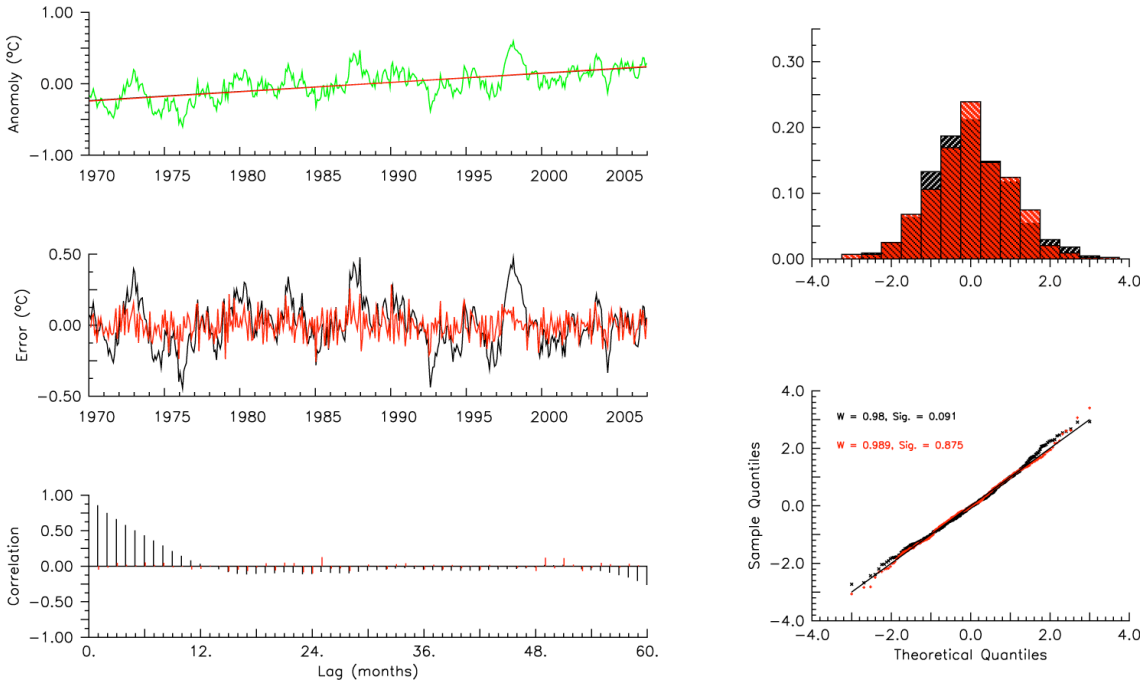

Figure 7-7: Results of fit of linear model to air temperature anomalies using an OLS regression (black) and the CO procedure (red). Shown are: time series of the air temperature anomalies (green) and linear models estimated using OLS (black) and the CO procedure (red) (top left); time series of the error terms (ε_t and u_t for the OLS and CO respectively) from the two regressions (middle left); auto-correlation function for the error terms (bottom left); standardized relative histogram of the errors (top right); and quantile – quantile plot for the standardized errors (sample quantiles) vs quantiles from the normal distribution (theoretical quantiles) (bottom right). w is the Shapiro-Wilk test statistic.

As an example of the process Figure 7-7 shows the results of fitting a linear trend model using OLS (black) and the CO procedure (red) to air temperature anomalies averaged over the Atlantic. Both methods of fitting the linear trend model produce similar results in terms of trend estimates, with little or no difference between the models (top left). However, whilst the models are similar there are significant differences in the error terms with the terms from the OLS containing significant auto-correlation (middle and bottom left plots), leading to an underestimate of the uncertainty in the fitted parameters. In contrast, the errors terms from the CO procedure contain no significant auto-correlation at the 95% significance level based on the Durbin-Watson

test. For comparison, the estimated decadal trends (± 1 standard deviation) from the two methods are 0.162 (± 0.008) °C decade^{-1} and 0.165 (± 0.021) °C decade^{-1} for the OLS and CO procedures respectively. For completeness, both sets of error terms have been tested for normality (right hand panels) using a Shapiro-Wilk's test. The null hypothesis (i.e. normally distributed errors) is accepted for both sets of error terms at the 95% significance level.

7.3.3 Impact of Height and Bias Adjustments

Trends have been estimated from monthly mean anomalies for each of the variables averaged over the Atlantic (40S – 70N, 100W – 30E) with (adjusted) and without (unadjusted) the height and bias adjustments applied. Only regions that are well sampled have been used in the calculation of the anomalies to restrict the impact of random errors on the comparisons, selecting only regions where the mean uncertainty is smaller than the mean daily standard deviation. A linear trend model has then been fitted through the averaged anomalies using the CO procedure.

Anomalies and trend models are shown in Figure 7-8 for both the adjusted (black) and unadjusted datasets (red) and are summarised in Table 7-1. For comparison the trend in unadjusted SST is also listed. The trends in the air temperature and humidity are increased slightly by the adjustments as expected, however these are within the 95% confidence intervals, suggesting the differences are not significant and that the impact of the adjustments cannot be detected reliably. In order to reduce the uncertainty and improve the detection of the adjustments the analysis has been repeated on differences in the air temperature and humidity between the adjusted and unadjusted dataset. Similar results are seen, with trends in the differences between the adjusted and unadjusted datasets of 0.011 ± 0.017 g Kg^{-1} and 0.012 ± 0.012 °C decade^{-1} for humidity and air temperature respectively. Similar results are also seen for the air – sea temperature and humidity differences, with only small differences in the trend estimates between the adjusted and unadjusted datasets. Overall, these results suggest that it is difficult to detect the impact of the adjustments on the trend estimates for air temperature and humidity and also on the air – sea gradients above the background of natural variability. Conversely, these results also suggest that it would be difficult to detect the impact of the increasing observing heights on unadjusted data without longer time series and lower uncertainties in the trend estimates.

Table 7-1: Estimated trends for the different variables, air – sea gradients and fluxes before and after adjustment. The uncertainties shown are the 95 % confidence interval for the trend estimates (±1.96 σ). Negative trends in the air – sea gradients and fluxes indicate an increase in the magnitude of the trends. For comparison, the trends from the first SOC climatology (1980 – 2002) are listed together with the trends from the new adjusted dataset over the same period. The values in italics are adversely affected by large outliers in the SOC climatology.

Variable	NOCS 2.0			SOC
	Unadjusted	Adjusted	Adjusted	
	1970 - 2006	1970 – 2006	1980 - 2002	1980 – 2002
Air temperature (°C decade^{-1})	0.15 ± 0.04	0.16 ± 0.04	0.13 ± 0.09	*0.24 ± 0.09*
Sea surface temperature (°C decade^{-1})	0.18 ± 0.05	N/A	0.19 ± 0.10	0.21 ± 0.09
Specific humidity (g Kg^{-1} decade^{-1})	0.12 ± 0.04	0.13 ± 0.03	0.16 ± 0.08	*0.26 ± 0.07*
Wind speed (m s^{-1} decade^{-1})	0.31 ± 0.02	0.15 ± 0.02	0.16 ± 0.03	0.28 ± 0.03
Air – sea temperature difference (°C decade^{-1})	-0.02 ± 0.01	-0.01 ± 0.02	-0.05 ± 0.03	*0.05 ± 0.02*
Air – sea humidity difference (g Kg^{-1} decade^{-1})	-0.04 ± 0.02	-0.03 ± 0.02	-0.00 ± 0.03	*0.08 ± 0.04*
Latent heat flux (W m^{-2} decade^{-1})	-4.82 ± 0.52	-2.72 ± 0.42	-2.37 ± 0.89	-6.88 ± 0.88
Sensible heat flux (W m^{-2} decade^{-1})	-0.32 ± 0.11	-0.12 ± 0.12	-0.32 ± 0.24	-0.61 ± 0.20

In contrast to the temperature and humidity adjustments, the adjustments made to the wind speed observations significantly alter the trends observed. By applying the adjustments to the wind speed observations the trend is reduced by ~ 50% compared to the unadjusted dataset. This large decrease is expected since the VOS wind speed observations are known to contain a spurious component to the trend (Thomas *et al*. 2008) and the height and bias adjustments have been implemented and applied to remove this spurious component. The reduction in the wind speed trend also reduces the observed trends in the latent and sensible heat flux estimates (negative fluxes indicate a heat loss by the ocean), reducing the combined trend magnitude from 5.1 W m^{-2}

decade^{-1} to 2.8 W m^{-2} decade^{-1}. The magnitude of this trend will be discussed in the next Section.

In addition to showing the improvement over the unadjusted dataset the trends in the adjusted dataset have been compared to those in the extended version of SOC climatology. The period 1980 – 2002 has been used rather than the full available period of 1980 to 2005 due to an unexplained increase in air temperature estimates from the SOC climatology from 2003 onwards. In addition to this jump, the outliers in air temperature and humidity shown in the previous section have an adverse impact on the trends estimates, leading to inaccurate trend estimates. As an example Figure 7-9 shows the time series of air temperature anomalies and estimated trends from the SOC climatology and adjusted version of the new dataset averaged over the Atlantic. Generally, the two time series are similar, but the impact of the outliers in the SOC climatology can be seen in the early 1980s and also in the trend estimates (Table 7-1). The impact of these outliers on the humidity and air temperature anomalies and trends estimates makes it is difficult to assess how the impact of the adjustments to air temperature and humidity varies between datasets.

Whilst there are problems with the air temperature and humidity estimates from the SOC climatology the wind speed, SST and fluxes do not seem to be affected by outliers. Similar sized trends are seen in the SST anomalies with slightly larger estimates in the SOC climatology. The source of this difference is unclear since both datasets use the same observations, the trend estimates are for well sampled regions and the bias correction applied to the observations used in the SOC climatology is expected to reduce the trend as the proportion of VOS using engine room intakes increases. The uncertainties in the trend estimates are large and the difference between the different trend estimates could be due to errors in the regression rather than real differences. In contrast to the SST, the trends in the wind speed and fluxes in the SOC climatology are significantly larger compared to the new (adjusted) dataset due to the adjustment applied to the visual wind speed estimates in the new dataset (see Chapter 4). This reduction in wind speed trend leads to a smaller trend in the fluxes.

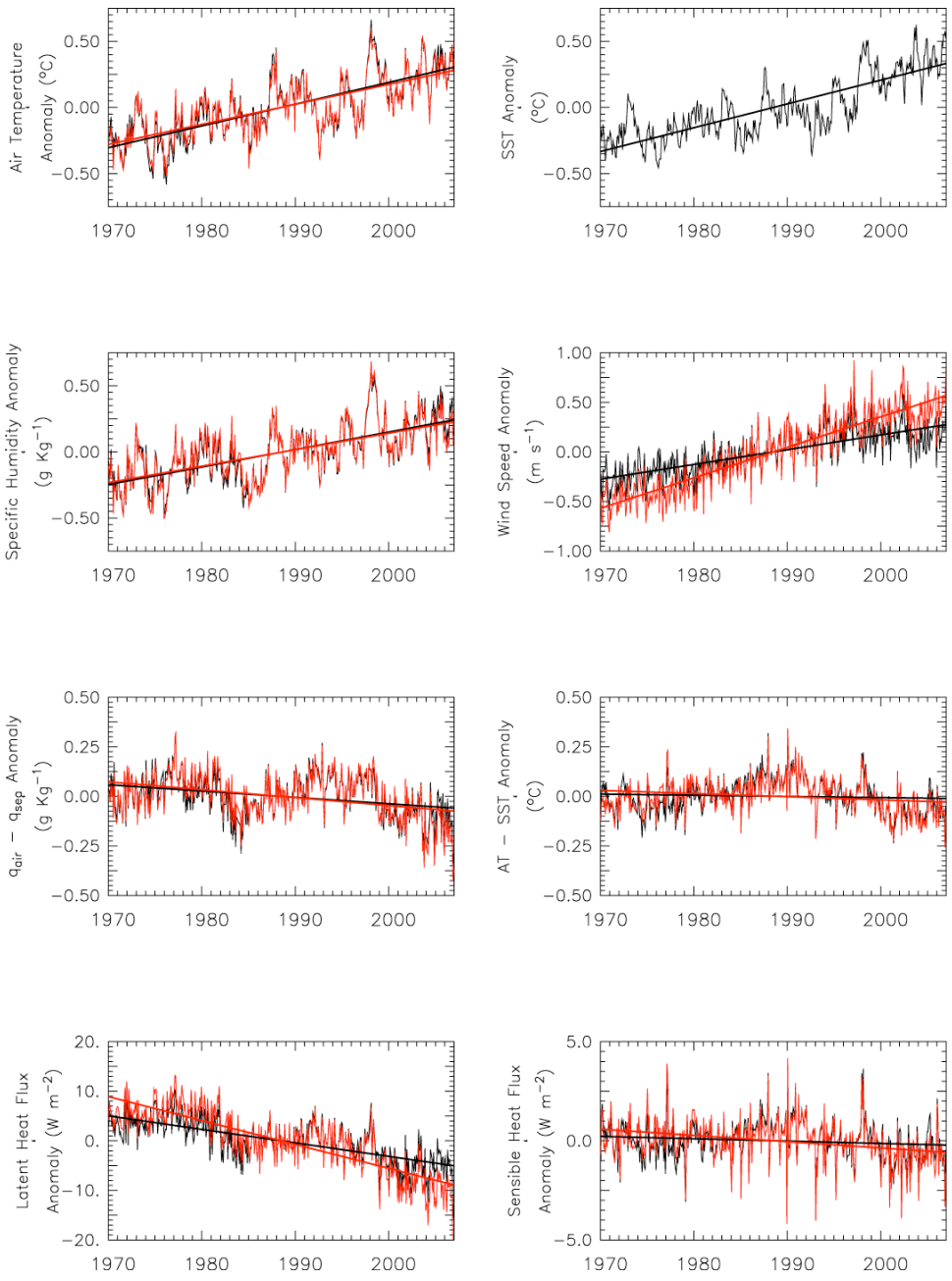

Figure 7-8: Time series of monthly anomalies (relative to the period 1970 – 2006) for each of the main variables and fluxes for the Atlantic. Adjusted (black) and unadjusted (red) time series are shown together with a model of the linear trend (bold lines).

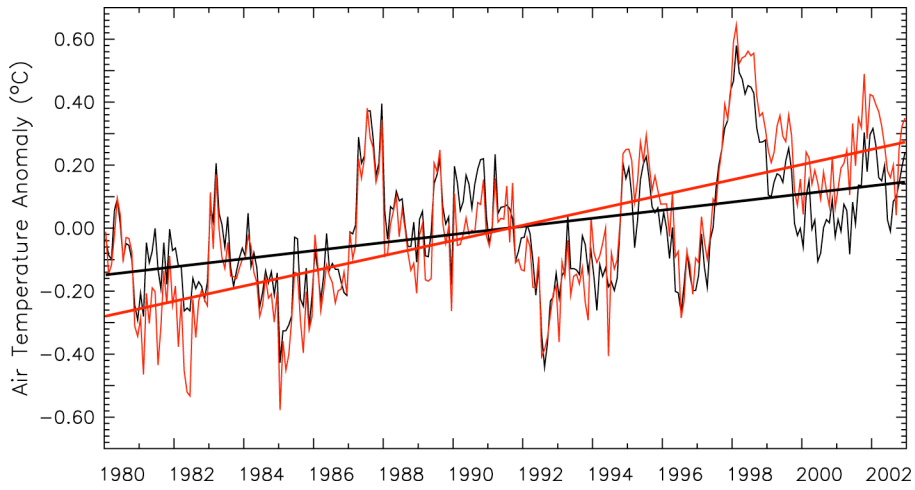

Figure 7-9: Air temperature anomalies calculated from the SOC climatology (red) and the new dataset (black) for the period 1980 – 2002 averaged over the Atlantic. Also shown are linear trend models fitted to both time series.

7.4 Discussion

7.4.1 Air temperature, humidity and SST trends

The recent changes in surface temperature over the oceans are generally undisputed with recent authors reporting similar sized trends to those seen in the new dataset. For example, Rayner *et al*. (2003) reported trends in annually averaged global (excluding the poorly sampled Southern Ocean) temperature anomalies for a number of different periods. For the recent period, 1982 – 1999, a trend of 0.15 ± 0.08 °C decade^{-1} has been reported for both night marine air temperature and sea surface temperature. These are also similar in size to those reported by Yu and Weller (2007) in the OAFlux dataset over the period 1981 - 2005, with trends in the Atlantic of around 0.2 °C decade^{-1} and with strong regional variations.

Based on the increase in the surface temperature and the Clausius-Clapeyron relationship an increase in the surface specific humidity is also expected. For example, when the specific humidity is estimated from the SST and an assumed constant relative humidity of 80% a trend of 0.14 ± 0.02 g Kg^{-1} decade^{-1} is seen for the period 1970 - 2006. This is similar in magnitude to that seen in the actual specific humidity estimates from the new dataset (0.13 ± 0.03 g Kg^{-1} decade^{-1}). Increases in the specific humidity have also been reported in previous studies with Willett *et al*. (2008), Dai (2006) and Yu and Weller (2007) all reporting positive trends. The estimates from both Dai (2006)

and Willett *et al*. (2008) are smaller than those seen in the new dataset with Willett *et al*. (2008) reporting a trend of 0.07 g Kg^{-1} $decade^{-1}$ in the marine global humidity estimates for 1973 – 2003 and Dai (2006) reporting 0.06 g Kg^{-1} $decade^{-1}$ in global land and marine humidity estimates for 1976 - 2004. Whilst these estimates are lower, they have been averaged globally rather than for the Atlantic and so some difference is expected due to the strong regional dependence of the trends. This can be seen in both the trend maps shown in Yu and Weller (2007) and in the regional estimates from Willett *et al*. (2008) and Dai (2006). Some difference due to the slightly shorter periods analysed is also to be expected.

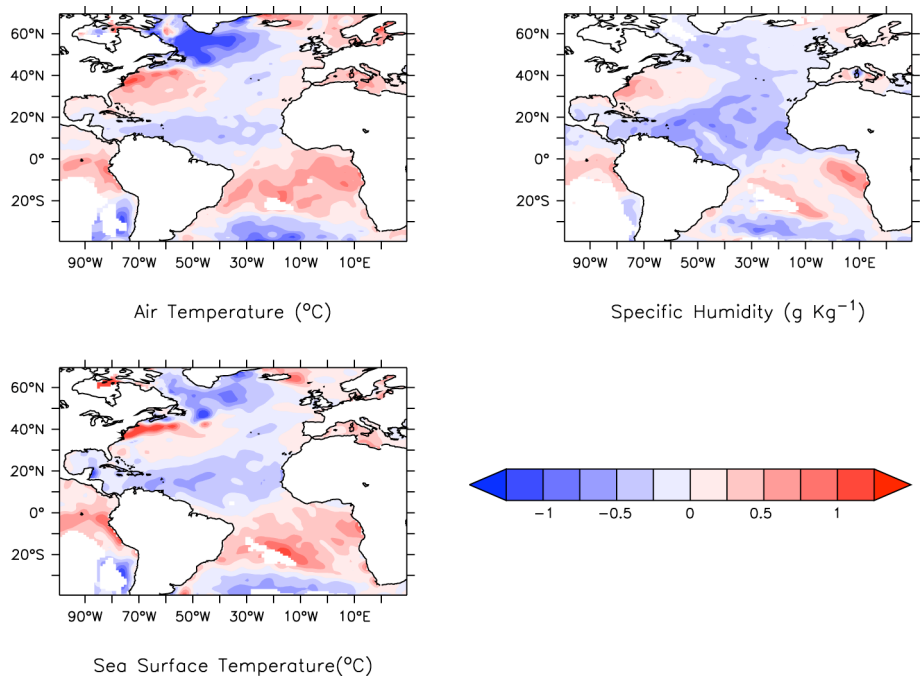

Figure 7-10: Difference between anomalies of air temperature, humidity and SST averaged over 1983 to 1986 (NAO +ve) and 1978 to 1981 (NAO –ve) (1983 to 1986 minus 1978 to 1981) from the adjusted dataset.

It should be noted that whilst the shift in humidity around 1982, reported by Willett *et al*. (2008) and attributed to changing reporting practices, can be seen in the new dataset it is not thought to be due to changing reporting practices. The spatial variability of this shift can also be seen in air temperature and SST estimates (Figure 7-10), suggesting the reported changes in 1982 may be due to climatic changes. Similar results are seen for the SST estimates from the HadISST dataset (e.g. Rayner *et al*. 2006) (not shown), with the tri-pole patterns strongly resembling those expected under

positive NAO conditions (e.g. Marshall *et al*. 2001a). Similar shifts in the humidity are not seen for other shifts in the NAO however.

7.4.2 Wind Speed Trends

The increase seen in the wind speed estimates in the new dataset is generally consistent with those seen in the increases in the NAO index since 1980 (e.g. Marshall *et al*. 2001b) and in significant wave height in the North Atlantic (e.g. Gulev and Grigorieva 2006). However, the magnitude of the trend may be too large due to a residual spurious trend component in the adjusted wind speed observations (Thomas *et al*. 2008). Thomas *et al*. (2008) report trends of 0.20 ms^{-1} $decade^{-1}$ for adjusted anemometer wind speed observations from the well sampled regions of the globe for the period 1982 – 2002 compared to trends of 0.05 and 0.06 ms^{-1} $decade^{-1}$ from the ERA40 and NCEP reanalysis models. Over this period, the reanalysis models are fairly well constrained by the available observations and the trends thought to be reasonable. As a result, this difference in trends is more likely to be due to a spurious component remaining in the adjusted observations rather than in the model values.

As a test to examine whether the magnitude of the remaining trend seen in the adjusted dataset is real or spurious seasonal trends have been estimated. The natural modes of variability have strong seasonal tendencies (e.g. Marshall *et al*. 2001b) and, as a result, we would also expect the trends to show strong seasonal tendencies. This is seen for the temperature and humidity variables, with the trends varying by 43% across the seasons. In contrast, seasonally estimated trends in wind speed vary by 14 % between minimum and maximum, ranging from 0.15 m s^{-1} $decade^{-1}$ during the spring to 0.17 m s^{-1} $decade^{-1}$ during the winter. This small variation suggests there may be a spurious component remaining in the adjusted wind speed observations and that the linear trend estimate of 0.15 ms^{-1} $decade^{-1}$ is an overestimate. Any overestimate of the trend in the wind speed observations will also impact on the trend estimates for the fluxes.

7.4.3 Heat Flux Trends

An increase in the evaporation from the sea surface is consistent with a warming ocean coupled with a warming and moistening atmosphere (i.e. increasing specific humidity) and also the increasing salinity seen in the Atlantic between 15 S and 42 N

(Bindoff *et al*. 2007). However, the increase seen in the new dataset, 3.9 W m^{-2} $decade^{-1}$ when estimated globally (2.7 W m^{-2} $decade^{-1}$ for the Atlantic), is thought to be an overestimate due to the overestimate of the wind speed trend. Similar sized trends are seen in a comparable dataset (OAFlux). In the preliminary version of OAFlux, restricted to the North Atlantic and for the period 1988 – 1999, Yu *et al*. (2004a) report similar sized fluxes and trends to the SOC climatology. In the more recent, global version of covering the period 1981 – 2005 Yu and Weller (2007) report a trend of 4 W m^{-2} $decade^{-1}$ in global latent heat flux anomalies . As a result, due to the similarity in the magnitude of the latent heat flux trends and since OAFlux does not directly use the VOS wind speed observations it is unclear from this comparison whether or not the trends are overestimated in the new dataset.

To try and resolve this and put the trends seen in the latent heat flux estimates into context it is useful to compare them to estimates of the global net heat flux balance at the ocean surface and to the contribution to radiative forcing by anthropogenic emissions. Based on changes to the net ocean heat content and rises in sea level over the period 1993 – 2003 Willis *et al*. (2004) estimate the net heat flux balance at the ocean surface to be + 0.86 W m^{-2} (i.e. heat gain by the oceans). This small net balance suggests that the increase in the latent heat flux must be balanced by either an increase in the radiative forcing or a decrease in sensible heat flux if the latent heat flux trends are realistic. However, the trend in sensible heat flux from the new dataset (and from OAFlux) is not significantly different to zero and an increase of sufficient magnitude in the radiative forcing of order 3 W m^{-2} $decade^{-1}$ is unrealistic. For example, the combined contribution to radiative forcing by anthropogenic sources has been estimated at only 1.6 W m^{-2} since 1750 (Forster *et al*. 2007). Assuming this increase is linear, the change equates to ~0.06 W m^{-2} $decade^{-1}$, an order of magnitude smaller than that required to balance the increase in the latent heat flux. Even if we assume that this increase occurred over the last couple of decades the increase is still insufficient to compensate for the increase in latent heat flux. As a result, the trends seen in the latent heat flux in the new dataset (and in OAFlux) are likely to be unrealistic and a significant overestimate of the trend in evaporation.

7.5 Summary

Two different sets of results have been presented in this Chapter. First, the variability of the monthly mean values in the new dataset have been compared to estimates from a dataset based on the arithmetic mean of the available observations and also to estimates from the first SOC climatology. Additionally, the geophysical consistency is examined in the different datasets by calculating the correlation between the fluxes and SST tendency. In the comparisons of the variability, the variability estimates from the new dataset have been shown to be an improvement compared to gridded data and to the SOC climatology. In both cases, there is a reduction in the noise whilst maintaining realistic correlation between variables where supported by the data. Additionally, the problem previously reported for the SOC climatology, with an artificial smoothing and reduction of variability in high variability regions, is not seen in the new dataset. The examination of the geophysical consistency showed improvements over the previous SOC climatology and gridded data. However, there are still problems in the South Atlantic due to poor sampling, with the correlation between the heat fluxes and SST tendency too low.

The second set of results examines the impact of the bias and height adjustment on the trends. The adjustments made to the air temperature and humidity observations increase the decadal trends by a small amount. However, the increases are within the respective 95% confidence intervals suggesting that the changes due to the increasing observing height are not easily detectable. Additionally, the trends in the air temperature and humidity and sea surface temperature are similar to those reported by previous authors giving confidence in the trend estimates for those variables.

In contrast to the air temperature and humidity, the wind speed adjustments make a significant impact on the trends observed in the new dataset with a reduction of almost 50%. Whilst the trend is significantly reduced, a spurious increase over time probably still exists in the wind speed observations and the trend estimated from the new dataset is likely to be an overestimate. The spurious component is also thought to adversely affect the trend estimates for latent heat flux with the trends unrealistically large compared to known changes in the ocean heat content and the radiative forcing due to anthropogenic emissions. It should be noted that the unrealistically large trend in the latent heat flux is not restricted to the new dataset but is also present in the OAFlux dataset and is expected to be present in other observational flux datasets.

8 Summary, Further Work and Conclusions

8.1 Summary of aims and results

The aims of the research presented in this thesis were to develop, and validate, a methodology for calculating the surface turbulent heat fluxes and their uncertainties over the Atlantic Ocean. These have been achieved and this research also contributes to the wider goal of generating a new, globally complete, surface flux dataset for the ice-free oceans to replace the first version of the SOC climatology (Josey *et al*. 1999).

The methodology developed characterizes bias adjusted VOS observations in terms of random errors using a semi-variogram analysis. The adjusted observations, and uncertainty estimates, are then used in an OI scheme to estimate daily mean fields, and associated uncertainty estimates, from the VOS observations. The surface fluxes and uncertainties are then calculated from the daily estimates and averaged to give monthly mean values. These mean fluxes were validated and shown to be realistic in well-sampled regions (Chapter 6). The uncertainty estimates were also found to be realistic (Chapter 6). The variability and trends in the new dataset were shown to be improved compared to the first SOC climatology and also to a simple gridded dataset (Chapter 7).

The method developed has subsequently been used to estimate the surface fluxes globally and to generate a new surface flux dataset. This also includes radiative heat fluxes not discussed in this thesis and is known as the NOCS Flux Dataset v2.0. The new dataset is available for research and documented in Berry and Kent (2009, 2010).

8.2 Key Improvements

The research presented in this thesis makes a number of important improvements to the surface flux estimates compared to previous studies. Central to the improvements is the calculation of realistic uncertainty estimates, bringing together a number of methods from different independent studies and forming a coherent method for estimating the surface fluxes and their uncertainty from the VOS observations.

The first improvement made is in the flux calculation strategy, estimating mean daily fields from the VOS observations and using these to calculate daily flux estimates. Previous authors have estimated monthly mean fluxes using either the classical or sampling methods (see Chapter 3). However, there are problems with both methods,

with either the synoptic scale correlations between variables lost through averaging (as in the classical method) or random errors adversely affecting the flux estimates due to the non-linearity of the bulk formulae (as in the sampling method). The use of daily mean values forms a compromise between minimizing the random errors and maintaining the correlation between the different variables on a synoptic time scale. In well-sampled regions the errors will be reduced and the synoptic scale correlations maintained. In poorly sampled regions, the results from the method developed in this thesis and from the sampling and classical methods will be similar.

The next major advance uses OI to estimate the daily mean fields from the VOS observations. This builds on the work of previous authors (e.g. Lorenc 1981; Reynolds and Smith 1994) who have used OI to develop gridded SST datasets based on combined satellite and *in situ* data sources and also for data assimilation. However, a number of adaptations have had to be made to use OI with the VOS observations on a daily time scale to account for the sparseness of the data. These adaptations are described in detail in Chapter 4 and in summary are: incrementing the previous days analysis to allow for the annual cycle and using this as the first guess for the OI; allowing the uncertainty to decay back to the climatological uncertainty to account for any period without observations and sampling errors; and using a relatively simple error correlation model for the first guess errors. The use of OI allows the errors to be minimized in the daily mean fields and provides realistic estimates of the uncertainty in those fields by taking into account the observational errors (i.e. uncertainties) and the spatial sampling of the VOS observations. Without this error minimization and associated uncertainty estimates it would not be possible to estimate the fluxes and flux uncertainty on a daily timescale.

The final major advance made in this research is the calculation of the monthly mean flux fields and their uncertainty from the daily values, taking the methodological contributions to the autocorrelation between daily fields into account. This allows the sampling by the VOS and the natural variability to be taken into account and for realistic uncertainty estimates to be made for the monthly mean values. If the autocorrelation were not taken into account the uncertainty in the monthly mean fields would have been underestimated.

A number of other improvements have also been made compared to previous studies using the VOS observations. Briefly, these are: improved bias adjustments for air temperature, humidity and wind speed observations; improved estimates of the

random errors by improving the semi-variogram method used; and estimates of the bias uncertainties made.

8.3 Requirements for surface flux datasets and uses of the new dataset

The development of the new dataset has been guided by the requirements of the potential users of the dataset. A number of users, those requiring high resolutions such as for forcing models, have been excluded from the outset in the development of the dataset by the use of VOS observations. For these studies, the requirements can only be met through other model based estimates or through the use of satellite data.

Out of the remaining potential users and applications targeted (in italics), the most stringent accuracy requirement is for *climate change* applications. For detection and attribution studies, stabilities of order 1 W m^{-2} are required for the net heat fluxes. The minimum resolutions required for these studies are typically 2° monthly over the majority of the oceans but with higher resolutions required in some key regions such as the Labrador Sea (WGASF 2000). For *climate variability* studies high accuracy is also desirable, but accuracies of 10 – 20 W m^{-2} in the mid latitudes and 7 – 10 W m^{-2} in the tropics are usable.

For *validation of NWP and reanalysis model output* and the output from the model data assimilation schemes a direct comparison would require the same time and space resolutions as the model. Similar resolutions are also required for *validating satellite flux estimates*. Fluxes on lower resolutions can be useful for these studies providing the true resolution of the both the flux dataset and the model output are known. For *validation applications* it is important that the validation dataset is independent from the source being validated. Realistic uncertainty estimates are also required. Due to the frequency of updates applied to NWP models and their data assimilation schemes the fluxes also need to be available within a few months of the forecast for the validation studies to be useful. For the *validation of the fluxes in climate models* the resolution requirements are relaxed, but the accuracy of the long term mean and variability are important. Accuracies less than 10 W m^{-2} are typically required.

Generally, with the exception of the climate change studies, the new dataset meets the requirements of the different applications. For the climate change studies the stability requirement is not met due to the potential spurious trend in the wind speed and latent heat flux estimates. It should be noted that this potential spurious trend may also

be present in other observational datasets such as OAFlux. For the climate variability studies, the lower accuracy requirements are generally met over the North Atlantic and also in the well-sampled regions of the South Atlantic. The regions where the accuracy requirements cannot be met can be readily identified in the new dataset by using the uncertainty estimates. These are not present in any of the other flux datasets.

For the validation studies, the higher resolution requirements cannot be met globally with the *in situ* sources. However, the lower resolution requirements are met with the new dataset. Additionally, the new dataset has the realistic uncertainty estimates required for validation applications and also has a greater degree of independence compared to other flux datasets. It is this greater independence and the realistic uncertainty estimates that differentiate the new dataset from other flux datasets. No other flux dataset has realistic uncertainty estimates and the different satellite flux products and model outputs cannot be used to validate each other as they are not independent from each other.

8.4 Future Work

8.4.1 Bias adjustments and bias uncertainties

Several improvements can be applied to the different bias adjustments. Characterising and removing biases in the wind speed should be a priority as the potential spurious trend in the wind speed estimates leads to the biggest uncertainty in the fluxes and flux trends. This may then allow the new dataset to be used for studying the impact of climate change on the fluxes.

The changing bias thought to exist in SST will also have an impact on the trend in SST and also in the fluxes. However, this changing bias is thought to be much less important for the fluxes than the changing biases in wind speed. However, once the issue of the potential spurious trend in wind speed is resolved improving the bias adjustments for SST should be a priority.

Whilst the adjustment applied to the screen humidity observations has been shown to improve the humidity estimates in the new dataset and agreement with buoy observations, the humidity estimates are still biased high compared to the buoys. Improving the bias adjustments, including investigating whether there is also a bias

present in the measurements from sling psychrometers, will improve the humidity and flux estimates.

The bias uncertainty estimates used in this research are based on either the differences between the different observing methods or on the RMS errors for the parameterisation of the coefficients used in the air temperature bias adjustment. As a result, these estimates are relatively crude and may be underestimating the accuracy of the different methods and observations. However, it is currently unclear how to improve these estimates.

8.4.2 Gridding

Whilst the uncertainty and flux fields have generally been shown to be realistic, improvements can be made to the gridding method used to generate these fields. The uncertainty estimates have been shown to be of the correct magnitude, however, they have also been shown to be overestimating the uncertainty in the low variability regions and underestimating in the high variability regions. Problems have also been shown with the geophysical consistency in the South Atlantic away from the shipping lanes.

Improving the estimates of the length scales used and the correlation between first guess errors should improve both of these problems. Examining the autocorrelation in the different fields in well-sampled regions should allow improved estimates of the temporal length scales to be made. Estimates from poorly sampled regions and model sources cannot be used due to forced auto-correlation by the models and OI process. An alternative method for examining the temporal length scales would be to perform a semi-variogram analysis on the time axis for the individual observations.

The correlation in the first guess errors could be improved by either performing an EOF analysis on the daily fields or by using a semi-variogram analysis on the data deviations (i.e. difference between the observations and first guess). Both methods are likely to give similar results in well sampled regions, especially if anisotropic variograms are used. However, in poorly sampled regions the EOF analysis may perform poorly due to potentially large errors leading to poor estimates of the correlation between grid cells. In contrast, the semi-variogram analysis is less sensitive to random errors and incomplete fields and should give better results in the poorly sampled regions.

8.4.3 Flux Calculation and Uncertainty

A number of different bulk formulae are available for estimating the surface turbulent heat fluxes from estimates of the bulk meteorological parameters. Some of these require additional approximations to be made, such as the boundary layer height and skin SST. These additional approximations will increase the uncertainty in the flux estimates without necessarily improving the accuracy. As a result, for completeness, a comparison of fluxes and flux uncertainties calculated using the different bulk formulae and the meteorological parameters from the new dataset is needed. Such a comparison will also help in assigning a value to the uncertainty estimates for the transfer coefficients, further improving the uncertainty estimates.

In addition to estimating the uncertainty in the transfer coefficients, estimating the correlation between the errors in the different variables will improve the uncertainty estimates. It is currently unclear what impact this will have on the uncertainty estimates since correlated error terms in the air and sea temperature and humidity estimates will tend to cancel in the bulk formulae.

8.4.4 Other Improvements

In addition to improving the method and adjustments applied to the VOS observations, it is intended in future work to include other data sources such as satellite and buoy observations. Whilst this will remove the independence of the dataset it is intended to have two parallel versions. The first of these will use all available observational sources and aim to provide the best estimate of the surface fluxes at the highest resolution possible. The second dataset, based on the present, will be used to improve the estimates of the autocorrelation in the daily fields and to improve the estimates of the spatial correlation length scales. However, the independence of the dataset will be maintained through using only VOS observations. It should be noted that any additional data must be fully characterised in terms of random and systematic errors.

8.5 Summary

Overall, the aims of the research presented in this thesis have been achieved, with the development of a new, integrated method for estimating the fluxes and their

uncertainty. This method has been assessed and future improvements identified. In well-sampled regions the fluxes and their uncertainty have been shown to be realistic. Additionally, the method has been shown to improve the variability and trend estimates in the fluxes compared to simple gridding of the VOS observations and to the estimates from the first SOC climatology. Finally, the method has been used to produce a new global flux dataset (Berry and Kent 2009, 2010) that will be an important resource for the independent validation and verification of flux estimates from model and satellite sources. We also expect to use the new dataset to examine the climate variability of the fluxes and related variables.

Appendix A – Acronyms, Abbreviations and Definitions

AATSR	Advanced Along Track Scanning Radiometer
AMIP-II	2nd Atmospheric Model Inter-comparison Project
AMSR-E	Advanced Microwave Scanning Radiometer – Earth Observing System
AR1	Autoregressive 1
AVHRR	Advanced Very High Resolution Radiometer
BES	Beaufort Equivalent Scale
CO	Cochrane-Orcutt, as in Cochrane-Orcutt procedure
COADS	Comprehensive Ocean-Atmosphere Data Set (now renamed ICOADS)
COARE	Coupled Ocean-Atmosphere Response Experiment
COARE 3.0	Version 3 of the COARE flux algorithm
DJF	December, January and February
DOE	Department of Energy
ECMWF	European Centre for Medium-range Weather Forecasting
ERA40	ECMWF Reanalysis Project
ERI	Engine room intake
ICOADS	International Comprehensive Ocean-Atmosphere Data Set
JJA	June, July and August
JMA	Japanese Meteorological Agency
JRA-25	25 Year Japanese Reanalysis Project
NAO	North Atlantic Oscillation
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NCEP1	NCEP / NCAR Reanalysis project
NCEP2	NCEP-DOE Reanalysis

NOC	National Oceanography Centre, Southampton
NWP	Numerical Weather Prediction
OI	Optimal Interpolation
OLS	Ordinary Least Squares
QC	Quality control
REML	Restricted Maximum Likelihood
RMS	Root mean squared, as in RMS error
RSOI	Reduced Space Optimal Interpolation
SLP	Sea level pressure
SOC	Southampton Oceanography Centre (now renamed National Oceanography Centre, Southampton)
SST	Sea surface temperature
TKE	Turbulent Kinetic Energy
VOS	Voluntary Observing Ship
WGASF	Working Group on Air – Sea Fluxes
WHOI	Woods Hole Oceanographic Institution
WHOI UOP	WHOI Upper Ocean Processes Group
WMO	World Meteorological Organization

Appendix B – Nomenclature

Symbol	Description	Units / value
A	Analysis value in OI	
a	Deviation of interpolated value from true value in OI	
a_1	Coefficient from surface renewal theory (Lui *et al*. 1979)	
a_2	Coefficient from surface renewal theory (Lui *et al*. 1979)	
a_{bh}	Parameter in stable stability correction of Beljaars and Holtslag (1991)	1
a_c	Charnock parameter	
A_k	Analysis value at location k in optimal interpolation	
a_p	Psychrometric coefficient	
a_{sv}	Range parameter for exponential and Gaussian variogram models	[km]
α	Deviation of analysis value (a) normalised by analysis error	
α_s	Coefficient in dimensionless profiles	
α_w	Clausius – Clapeyron wet bulb factor	
b	Deviation of observed value from true value in OI	
B	Observed value in OI	
B_0	Bulk Bowen ratio	
b_1	Coefficient from surface renewal theory (Liu *et al*. 1979)	
b_2	Coefficient from surface renewal theory (Liu *et al*. 1979)	
b_{bh}	Parameter in stable stability correction of Beljaars and Holtstag	0.667
B_i	Observation *i* in OI	
b_{sv}	Gradient in linear variogram model	
β	Deviation of observed value normalised by observational error	
β_{conv}	Empirical coefficient for gustiness parameter	
β_s	Coefficient in dimensionless profiles	
β_t	Contribution to analysis value by observations	
β_0	Intercept in linear AR1 time series model	
β_0'	Transformed intercept from linear AR1 time series model used in Cochrane-Orcutt procedure	
β_1	Gradient in linear AR1 time series model	
β_1'	Transformed gradient from linear AR1 time series model used in Cochrane-Orcutt procedure	
C_0	Intercept of model variogram	
C_1	Sill parameter for Gaussian and exponential variogram models	
c^2	True temporal variance within grid box	
c_{bh}	Parameter in stable stability correction of Beljaars and Holtslag (1991)	5
C_D	Momentum transfer coefficient	
C_{D10n}	Neutral momentum transfer coefficient at 10m	
C_E	Moisture transfer coefficient	
C_{E10n}	Neutral moisture transfer coefficient at 10 m	
C_H	Heat transfer coefficient	
C_{H10n}	Neutral heat transfer coefficient at 10 m	
C_{ij}^{k}	Successive correction analysis increment for grid box (i,j)	

Symbol	Description	Units / value
	and iteration k	
C_m	Climatological mean value for month m	
C_p	Phase speed of dominant wave	
c_p	Specific heat capacity of air at constant pressure	[J kg K^{-1}]
c_{pw}	Specific heat capacity of water	[J kg^{-1} k^{-1}]
d_{bh}	Parameter in stable stability correction of Beljaars and Holtslag (1991)	0.35
D_{ij}^{k}	Data deviation in successive correction for grid box (i,j) and iteration k.	
D_m	Mid point (day) of month m	
ΔC_t	Daily increment to OI analysis to give first guess for day t	
ΔT	Air – sea temperature difference	[°C]
ΔT_{cor}	Air temperature bias correction	[°C]
E_{t0}^{a}	Analysis error for last time step with data	
E^a	Expected analysis error (standard deviation) in OI	
E^p	Expected first guess error	
E^b	Expected observation error	
E_k^p	First guess error at location k	
E_t^p	First guess error at time step t	
ε^a	Normalised analysis error	
ε^b	Normalised observation error	
ε_t	Error term in linear AR1 time series model at time step t	
E	Moisture flux	[kg m^{-2} s^{-1}]
e	Vapour pressure	[mb]
e_s	Saturation vapour pressure at a given temperature	[mb]
e_s'	Saturation vapour pressure at wet bulb temperature	[mb]
F_{ij}^{k}	First guess for successive correction for grid box (i,j) and iteration k	
ϕm	Dimensionless wind shear	
ϕ_q	Dimensionless humidity gradient	
ϕ_t	Dimensionless temperature gradient	
g	Acceleration due to gravity	[m s^{-2}]
$\gamma(h)$	Modelled variogram at separation h	
γ_s	Coefficient in dimensionless profiles	
H	Sensible heat flux	[W m^{-2}]
h_s	Significant wave height	[m]
H_{sr}	Sensible heat flux due to precipitation in COARE 3.0 algorithm	[W m^{-2}]
h_{sv}	Separation distance between observations in variogram analysis	[km]
k	von Karmen constant	0.4
L	Monin-Obukhov length	[m]
L_p	Wave length associated with dominant wave period	[m]
L_v	Latent heat of vaporization	[J kg^{-1}]
λ	Spatial e-folding scale used for setting error correlations in OI	[km]
λ_t	Time e-folding scale used in setting of first guess error	[days]
m	Measurement error for individual observation	

Symbol	Description	Units / value
n	Number of days since last observations	
n_{se}	Number of observations in grid box in estimation of sampling errors	
ν	Kinematic viscosity of air	[m^2 s^{-1}]
N	Number of variables in function V	
p	Deviation of predicted value from true value	
P	Predicted (first guess) value in OI	
P_i	Predicted value at location i	
p_p	Pressure	[mb]
π	Deviation of predicted value (a) normalised by first guess error	
q_*	Characteristic humidity	[g Kg^{-1}]
$\overline{q}$	Mean specific humidity	[g kg^{-1}]
q'	Turbulent fluctuation of specific humidity	[g kg^{-1}]
q_0	Specific humidity at air – sea interface	[g kg^{-1}]
q_{10s}	Stability dependent height adjusted 10 m specific humidity	[g kg^{-1}]
q_z	Specific humidity at height z_q	[g kg^{-1}]
θ'	Turbulent fluctuation of potential temperature	[K]
$\overline{r}$	Mean correlation of every true point value in grid box with every other point value in grid box	
r_{sc}	Distance between grid boxes in successive correction	[km]
R	Rain rate	[mm hr^{-1}]
R^2	Pearsons correlation coefficient	
R_{ij}	Raw data in successive correction at grid box (i,j)	
R^k	Radius of influence in successive correction at iteration k	
R_r	Roughness Reynolds number	
R_{SW}	Incident solar radiation	[W m^{-2}]
ρ_0	Density of air	[kg m^{-3}]
ρ_{ij}	Correlation between errors in i^{th} and j^{th} variables in propagation of errors	
$\rho_{t,t-1}$	Correlation between errors in analyses on days t and t-1	
ρ_{ac1}	Lag 1 autocorrelation in linear AR1 time series model	
S_{ne}^2	Contribution to observed variance by measurement errors	
S_n^2	Observed variance of monthly mean anomalies based on n observations	
σ_{clim}	Uncertainty in climatological monthly mean fields	
S	Mean daily uncertainty in OI fields	
S_{nt}^2	Variance of grid box average of n correlated observations in estimation of sampling error	
S_{coare}	Scalar wind speed used in COARE 3.0 algorithm	[m s^{-1}]
σ_{bias}	Bias uncertainty in monthly mean value	
σ_i	Error in i^{th} variable	
$\sigma_{monthly}$	Total uncertainty in monthly mean value	
σ_{random}	Random and sampling uncertainty in monthly mean value	
σ_V	Error (uncertainty) in function V	
T_*	Characteristic temperature	[°C]
T_{dew}^{cor}	Corrected dew point temperature	[°C]
$\overline{T}$	Mean air temperature	[°C]

Symbol	Description	Units / value
T_v'	Turbulent fluctuation of virtual air temperature	[K]
T	True value of field in OI	
T_0	Temperature at air – sea interface	[°C]
T_{10s}	Stability dependent height adjusted 10 m air temperature (°C)	[°C]
T_{dew}	Dew point temperature	[°C]
T_{dry}	Dry bulb temperature	[°C]
T_v	Virtual temperature of air	[K]
T_{wet}	Wet bulb temperature	[°C]
T_z	Temperature at height z	[°C]
τ	Wind stress	[N m^{-2}]
u_*	Friction velocity	[m s^{-1}]
$\overline{u}$	Mean horizontal wind speed	[m s^{-1}]
u'	Turbulent component of horizontal wind speed	[m s^{-1}]
u_0	Wind speed at air – sea interface	[m s^{-1}]
u_{10n}	Height adjusted wind speed assuming neutral stability	[m s^{-1}]
u_{10s}	Stability dependent height adjusted 10 m wind speed	[m s^{-1}]
U_E	Visual wind speed estimates before adjustment	[m s^{-1}]
U_{EL}	Visual wind speed estimates adjusted following Lindau (1995)	[m s^{-1}]
u_t	Independent, normally distributed error term in linear AR1 time series model	
u_z	Wind speed at height z	[m s^{-1}]
V	Function of several variables	
v_i	i^{th} variable in function V	
V_{rel}	Relative wind speed	[m s^{-1}]
V_s	Wind speed	[m s^{-1}]
w'	Turbulent component of vertical wind speed	[m s^{-1}]
w_g	Gustiness wind parameter for COARE 3.0 algorithm	[m s^{-1}]
w_{ik}	Weight applied to observation i in optimal interpolation for point k	
W_s	Weighting function used in successive correction	
$x_1 - x_5$	Empirical coefficients for air temperature radiative heating error adjustment	
X_t	Independent variable (time) in linear AR1 time series model	
X_t'	Independent variable in linear AR1 time series model transformed using Cochrane – Orcutt procedure	
Y_t	Dependent variable in linear AR1 time series model	
Y_t'	Dependent variable in linear AR1 time series model transformed using Cochrane – Orcutt procedure	
ψ_m	Momentum stability correction	
ψ_{mC}	Convective stability correction component of COARE 3.0 algorithm	
ψ_{mK}	Kansas stability correction component of COARE 3.0 algorithm	
ψ_q	Specific humidity stability correction	
ψ_t	Temperature stability correction	
z	Observation height (subscripts refer to variables)	[m]

Symbol	Description	Units / value
z_0	Momentum roughness length	[m]
z_{0Oo}	Momentum roughness length according to Oost (2003) wave age model	[m]
z_{0q}	Humidity roughness length	[m]
z_{0t}	Temperature roughness length	[m]
z_{0TY}	Momentum roughness length according to Taylor and Yelland (2001) wave height model (m)	[m]
z_i	Height of boundary layer	[m]
ζ	Dimensionless stability parameter	

Appendix C – Coefficients for air temperature bias correction

Year	x1	x2	x3	x4	x5	x6
1970	0.00196	0.001	30.1	1	697.0	0
1971	0.00182	0.001	21.8	1	585.1	0
1972	0.00144	0.001	25.9	1	456.3	0
1973	0.00201	0.001	42.4	1	538.3	0
1974	0.00162	0.001	15.9	1	494.7	0
1975	0.00146	0.001	29.1	1	311.2	0
1976	0.00189	0.001	17.7	1	574.5	0
1977	0.00176	0.001	14.8	1	532.4	0
1978	0.00153	0.001	14.7	1	487.1	0
1979	0.00152	0.001	11.9	1	517.7	0
1980	0.00187	0.001	23.9	1	532.8	0
1981	0.00183	0.001	10.6	1	699.3	0
1982	0.00180	0.001	33.3	1	504.8	0
1983	0.00173	0.001	17.0	1	626.9	0
1984	0.00172	0.001	30.6	1	499.3	0
1985	0.00206	0.001	46.0	1	547.2	0
1986	0.00214	0.001	33.0	1	668.7	0
1987	0.00261	0.001	46.6	1	819.5	0
1988	0.00169	0.001	33.7	1	437.8	0
1989	0.00177	0.001	15.3	1	645.0	0
1990	0.00219	0.001	55.6	1	486.0	0
1991	0.00173	0.001	22.1	1	540.5	0
1992	0.00209	0.001	37.1	1	569.6	0
1993	0.00184	0.001	32.3	1	508.0	0
1994	0.00203	0.001	31.5	1	564.7	0
1995	0.00160	0.001	20.3	1	461.5	0
1996	0.00174	0.001	25.5	1	472.8	0
1997	0.00177	0.001	22.1	1	564.7	0
1998	0.00147	0.001	23.5	1	445.8	0
1999	0.00160	0.001	33.9	1	369.5	0
2000	0.00201	0.001	34.0	1	530.3	0
2001	0.00217	0.001	14.2	1	759.4	0
2002	0.00160	0.001	28.9	1	412.8	0
2003	0.00173	0.001	24.6	1	480.7	0
2004	0.00204	0.001	29.4	1	563.3	0
2005	0.00181	0.001	40.9	1	431.2	0
2006	0.00161	0.001	43.0	1	254.8	0

List of References

Anderson, S. P., and M. F. Baumgartner, 1998: Radiative heating errors in naturally ventilated air temperature measurements made from buoys. *Journal of Atmospheric and Oceanic Technology*, 15, 157 - 173, DOI: 10.1175/1520-0426(1998)015<0157:RHEINV>2.0.CO;2.

Andreas, E. L., and J. De Cosmo, 2002: The Signature of Sea Spray in the Hexos Turbulent Heat Flux Data. *Boundary Layer Meteorology*, 103, 303 - 333, DOI: 10.1023/A:1014564513650.

Andreas, E. L., P. O. G. Persson, and J. E. Hare, 2008: A Bulk Turbulent Air-Sea Flux Algorithm for High-Wind, Spray Conditions. *Journal of Physical Oceanography*, 38, 1581 - 1596, DOI: 10.1175/2007JPO3813.1.

Bakan, S., V. Jost, and K. Fennig, 2000: Satellite derived water balance climatology for the North Atlantic: First results. *Physics and Chemistry of the Earth*, 25, 121 - 128, DOI: 10.1016/S1464-1909(99)00137-9.

Baumgartner, M. F., and S. P. Anderson, 1999: Evaluation of regional numerical weather prediction model surface fields over the Middle Atlantic Bight. *Journal of Geophysical Research*, 104, 18141 - 18158.

Beljaars, A. C. M., and A. A. M. Holtslag, 1991: Flux parameterisation over land surfaces for atmospheric models. *Journal of Applied Meteorology*, 30, 327 - 341, DOI: 10.1175/1520-0450(1991)030<0327:FPOLSF>2.0.CO;2.

Bentamy, A., K. B. Katsaros, A. M. Mestas-Nunez, W. M. Drennan, E. B. Forde, and H. Roquet, 2003: Satellite Estimates of Wind Speed and Latent Heat Flux over the Global Oceans. *Journal of Climate*, 16, 637 - 656, DOI: 10.1175/1520-0442(2003)016<0637:SEOWSA>2.0.CO;2.

Berry, D. I., and E. C. Kent, 2005: The effect of instrument exposure on marine air temperatures: An assessment using VOSClim data. *International Journal of Climatology*, 25, 1007 - 1022, DOI: 10.1002/joc.1178.

Berry, D. I., and E. C. Kent, 2009: A New Air - Sea Interaction Gridded Data Set from ICOADS with Uncertainty Estimates. *Bulletin Of The American Meteorological Society*, 90, 645 - 656, DOI: 10.1175/2008BAMS2639.1.

Berry, D. I., and E. C. Kent, 2010: Air-Sea Fluxes from ICOADS: The Construction of a New Gridded Dataset with Uncertainty Estimates. *International Journal of Climatology*, In press, DOI: 10.1002/joc.2059.

Berry, D. I., E. C. Kent, and P. K. Taylor, 2004: An analytical model of heating errors in marine air temperatures from ships. *Journal of Atmospheric and Oceanic Technology*, 21, 1198-1215, DOI: 10.1175/1520-0426(2004)021<1198:AAMOHE>2.0.CO;2.

Bindoff, N. L., and Coauthors, 2007: Observations: Oceanic Climate Change and Sea Level. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon, and Coauthors, Eds., Cambridge University Press.

Blanc, T. V., 1986: The effect of inaccuracies in weather-ship data on bulk-derived estimates of flux, stability and sea surface roughness. *Journal of Atmospheric and Oceanic Technology*, 3, 12 - 26, DOI: 10.1175/1520-0426(1986)003<0012:TEOIIW>2.0.CO;2.

Brunke, M. A., X. Zeng, and S. Anderson, 2002: Uncertainties in sea surface turbulent algorithms and data sets. *Journal of Geophysical Research*, 107, DOI: 10.1029/2001JC000992.

Brunke, M. A., C. W. Fairall, X. Zeng, L. Eymard, and J. A. Curry, 2003: Which Bulk Aerodynamic Algorithms are Least Problematic in Computing Ocean Surface Turbulent Fluxes? *Journal of Climate*, 16, 619 - 635, DOI: 10.1175/1520-0442(2003)016<0619:WBAAAL>2.0.CO;2.

Bunker, A. F., 1976: Computations of Surface Energy Flux and Annual Air-Sea Interaction Cycles of the North Atlantic Ocean. *Monthly Weather Review*, 104, 1122 - 1140, DOI: 10.1175/1520-0493(1976)104<1122:COSEFA>2.0.CO;2.

Busch, N. E., 1972: On the mechanics of atmospheric turbulence. *Workshop on Micrometeorology*, D. A. Haugen, Ed., American Meteorological Society, 1 - 61.

Businger, J. A., J. C. Wyngaard, Y. Izumi, and E. F. Bradley, 1971: Flux profile relationships in the atmospheric surface layer. *Journal of the Atmospheric Sciences*, 28, 181 - 189, DOI: 10.1175/1520-0469(1971)028<0181:FPRITA>2.0.CO;2.

Cardone, V. J., J. G. Greenwood, and M. A. Cane, 1990: On Trends in Historical Marine Wind Data. *Journal of Climate*, 3, 113 - 127, DOI: 10.1175/1520-0442(1990)003<0113:OTIHMW>2.0.CO;2.

Cayan, D. R., 1992: Latent and Sensible Heat Flux Anomalies over the Northern Oceans: Driving the Sea Surface Temperature. *Journal of Physical Oceanography*, 22, 859 - 881, DOI: 10.1175/1520-0485(1996)026<1165:DAISVI>2.0.CO;2.

Charnock, H., 1955: Wind stress on a water surface. *Quarterly Journal of the Royal Meteorological Society*, 81, 639 - 640, DOI: 10.1002/qj.49708135027.

Chou, S. H., R. M. Atlas, C. L. Shie, and J. Ardizzone, 1995: Estimates Of Surface Humidity And Latent-Heat Fluxes Over Oceans From SSM/I Data. *Monthly Weather Review*, 123, 2405-2425, DOI: 10.1175/1520-0493(1995)123<2405:EOSHAL>2.0.CO;2.

Chou, S. H., C. L. Shie, R. M. Atlas, and J. Ardizzone, 1997: Air-sea fluxes retrieved from special sensor microwave imager data. *Journal Of Geophysical Research-Oceans*, 102, 12705-12726.

Chou, S. H., E. Nelkin, J. Ardizzone, R. M. Atlas, and C. L. Shie, 2003: Surface Turbulent Heat and Momentum Fluxes over Global Oceans Based on the Goddard Satellite Retrievals, Version 2 (GSSTF2). *Journal of Climate*, 16, 3256 - 3273, DOI: 10.1175/1520-0442(2003)016<3256:STHAMF>2.0.CO;2.

Clayson, C. A., C. W. Fairall, and J. A. Curry, 1996: Evaluation of turbulent fluxes at the ocean surface using surface renewal theory. *Journal of Geophysical Research*, 101, 28503 - 28513.

Crescenti, G. H., and R. A. Weller, 1992: Analysis of Surface Fluxes in the Marine Atmospheric Boundary Layer in the Vicinity of Rapidly Intensifying Cyclones. *Journal of Applied Meteorology*, 31, 831 - 848, DOI: 10.1175/1520-0450(1992)031<0831:AOSFIT>2.0.CO;2.

Cressie, N. A. C., 1993: *Statistics for Spatial Data*. Wiley and Sons, 900 pp.

da Silva, A. M., C. C. Young, and S. Levitus, 1994: Atlas of Surface Marine Data 1994 Volume 1: Algorithms and Procedures. NOAA Atlas NESDIS 6, 74 pp pp.

Dai, A., 2006: Recent Climatology, Variability, and Trends in Global Surface Humidity. *Journal of Climate*, 19, 3589 - 3606, DOI: 10.1175/JCLI3816.1.

Dobson, F. W., and S. D. Smith, 1988: Bulk models of solar radiation at sea. *Quarterly Journal of the Royal Meteorological Society*, 114, 165 - 182, DOI: 10.1002/qj.49711447909.

Drennan, W. M., P. K. Taylor, and M. J. Yelland, 2005: Parameterizing the Sea Surface Roughness. *Journal of Physical Oceanography*, 35, 835 - 848, DOI: 10.1175/JPO2704.1.

Dyer, A. J., 1974: A review of flux-profile relationships. *Boundary Layer Meteorology*, 7, 363 - 372, DOI: 10.1007/BF00240838.

Ebuchi, N., H. C. Graber, and M. J. Caruso, 2002: Evaluation of Wind Vectors Observed by QuikSCAT/SeaWinds Using Ocean Buoy Data. *Journal of Atmospheric and Oceanic Technology*, 19, 2049 - 2062, DOI: 10.1175/1520-0426(2002)019<2049:EOWVOB>2.0.CO;2.

Edson, J. B., C. W. Fairall, P. G. Metsayer, and S. E. Larsen, 1991: A Study of the Inertial-Dissipation Method for Computing Air-Sea Fluxes. *Journal of Geophysical Research*, 96, 10689 - 10711.

Edson, J. B., A. A. Hinton, K. E. Prada, J. E. Hare, and C. W. Fairall, 1998: Direct Covariance Flux Estimates from Mobile Platforms at Sea. *Journal of Atmospheric and Oceanic Technology*, 15, 547 - 562, DOI: 10.1175/1520-0426(1998)015<0547:DCFEFM>2.0.CO;2.

Emery, W. J., D. J. Baldwin, P. Schlussel, and R. W. Reynolds, 2001: Accuracy of in situ sea surface temperatures used to calibrate infrared satellite measurements. *Journal of Geophysical Research*, 106, 2387 - 2405.

Fairall, C. W., E. F. Bradley, D. P. Rogers, J. B. Edson, and G. S. Young, 1996: Bulk parameterization of air-sea fluxes for Tropical Ocean-Global Atmosphere Coupled-Ocean Atmosphere Response Experiment. *Journal of Geophysical Research*, 101, 3747 - 3764.

Fairall, C. W., E. F. Bradley, J. E. Hare, A. A. Grachev, and J. B. Edson, 2003: Bulk Parameterization of Air–Sea Fluxes: Updates and Verification for the COARE Algorithm. *Journal of Climate*, 16, 571 - 591, DOI: 10.1175/1520-0442(2003)016<0571:BPOASF>2.0.CO;2.

Flament, P., and M. Sawyer, 1995: Observations Of The Effect Of Rain Temperature On The Surface Heat-Flux In The Intertropical Convergence Zone. *Journal of Physical Oceanography*, 25, 413-419, DOI: 10.1175/1520-0485(1995)025<0413:OOTEOR>2.0.CO;2.

Folland, C. K., 1971: Daytime temperature measurements on weather ship 'Weather Reporter'. *Meteor. Mag.*, 100, 6 - 14.

Folland, C. K., and D. E. Parker, 1995: Correction of instrumental biases in historical sea surface temperature data. *Quarterly Journal of the Royal Meteorological Society*, 121, 319 - 367, DOI: 10.1002/qj.49712152206.

Forster, P., and Coauthors, 2007: Changes in Atmospheric Constituents and in Radiative Forcing. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon, and Coauthors, Eds., Cambridge University Press.

Glahn, W., 1933: False measurements of air temperatures on ships. *Der Seewart*, 2, 250 - 256.

Gleckler, P. J., and B. C. Weare, 1997: Uncertainties in Global Ocean Surface Heat Flux Climatologies Derived from Ship Observations. *Journal of Climate*, 10, 2764 - 2781, DOI: 10.1175/1520-0442(1997)010<2764:UIGOSH>2.0.CO;2.

Goerrs, J. S., and C. E. Duchon, 1980: Effect of ship heating on dry bulb temperature measurements in GATE. *Journal of Physical Oceanography*, 10, 478 - 479, DOI: 10.1175/1520-0485(1980)010<0478:EOSHOD>2.0.CO;2.

Gosnell, R., C. W. Fairall, and P. J. Webster, 1995: The sensible heat of rainfall in the tropical ocean. *Journal of Geophysical Research*, 100, 18437 - 18442.

Grist, J. P., and S. A. Josey, 2003: Inverse Analysis Adjustment of the SOC Air–Sea Flux Climatology Using Ocean Heat Transport Constraints. *Journal of Climate*, 16, 3274 - 3295, DOI: 10.1175/1520-0442(2003)016<3274:IAAOTS>2.0.CO;2.

Gulev, S., T. Jung, and E. Ruprecht, 2007a: Estimation of the Impact of Sampling Errors in the VOS Observations on Air–Sea Fluxes. Part I: Uncertainties in Climate Means. *Journal of Climate*, 20, 279 - 301, DOI: 10.1175/JCLI4010.1.

Gulev, S., T. Jung, and E. Ruprecht, 2007b: Estimation of the Impact of Sampling Errors in the VOS Observations on Air–Sea Fluxes. Part II: Impact on Trends and Interannual Variability. *Journal of Climate*, 20, 302 - 315, DOI: 10.1175/JCLI4008.1.

Gulev, S. K., and V. Grigorieva, 2006: Variability of the Winter Wind Waves and Swell in the North Atlantic and North Pacific as Revealed by the Voluntary Observing Ship Data. *Journal of Climate*, 19, 5667 - 5685, DOI: 10.1175/JCLI3936.1.

Henderson-Sellers, B., 1984: A new formula for latent heat of vaporization of water as a function of temperature. *Quarterly Journal of the Royal Meteorological Society*, 110, 1186-1190, DOI: 10.1002/qj.49711046626.

Hosom, D. S., R. A. Weller, R. E. Payne, and K. E. Prada, 1995: The IMET (Improved Meteorology) Ship and Buoy Systems. *Journal of Atmospheric and Oceanic Technology*, 12, 527 - 540, DOI: 10.1175/1520-0426(1995)012<0527:TIMSAB>2.0.CO;2.

Hsiung, J., 1986: Mean surface energy fluxes over the global ocean. *Journal of Geophysical Research*, 91, 10585 - 10606.

Huebert, B. J., B. W. Blomquist, J. E. Hare, C. W. Fairall, J. E. Johnson, and T. S. Bates, 2004: Measurement of the sea-air DMS flux and transfer velocity using eddy correlation. *Geophysical Research Letters*, 31, DOI: 10.1029/2004GL021567.

Isaaks, E. H., and R. M. Srivastava, 1990: *An Introduction to Applied Geostatistics*. Oxford University Press Inc., USA, 592 pp.

Isemer, H. J., and L. Hasse, 1991: The Scientific Beaufort Equivalent Scale: Effects on Wind Statistics and Climatological Air-Sea Flux Estimates in the North Atlantic Ocean. *Journal of Climate*, 4, 819 - 836, DOI: 10.1175/1520-0442(1991)004<0819:TSBESE>2.0.CO;2.

Jones, P. D., T. J. Osborn, and K. R. Briffa, 1997: Estimating Sampling Errors in Large-Scale Temperature Averages. *Journal of Climate*, 10, 2548 - 2568, DOI: 10.1175/1520-0442(1997)010<2548:ESEILS>2.0.CO;2.

Josey, S. A., E. C. Kent, and P. K. Taylor, 1995: Seasonal-Variations Between Sampling And Classical Mean Turbulent Heat-Flux Estimates In The Eastern North-Atlantic. *Annales Geophysicae-Atmospheres Hydrospheres And Space Sciences*, 13, 1054-1064.

Josey, S. A., E. C. Kent, and P. K. Taylor, 1998: The Southampton Oceanography Centre (SOC) Ocean - Atmosphere Heat, Momentum and Freshwater Flux Atlas. James Rennell Division, Southampton Oceanography Centre, UK, 59.

Josey, S. A., E. C. Kent, and P. K. Taylor, 1999: New Insights into the Ocean Heat Budget Closure Problem from Analysis of the SOC Air–Sea Flux Climatology. *Journal of Climate*, 12, 2856 - 2880, DOI: 10.1175/1520-0442(1999)012<2856:NIITOH>2.0.CO;2.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bulletin Of The American Meteorological Society*, 77, 437-471, DOI: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP-DOE AMIP-II Reanalysis (R-2). *Bulletin Of The American Meteorological Society*, 83, 1631 - 1643, DOI: 10.1175/BAMS-83-11-1631.

Kaplan, A., Y. Kushnir, and M. A. Cane, 2000: Reduced Space Optimal Interpolation of Historical Marine Sea Level Pressure: 1854–1992. *Journal of Climate*, 13, 2987 - 3002, DOI: 10.1175/1520-0442(2000)013<2987:RSOIOH>2.0.CO;2.

Kaplan, A., Y. Kushnir, M. A. Cane, and M. B. Blumenthal, 1997: Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures. *Journal Of Geophysical Research-Oceans*, 102, 27835-27860.

Kaplan, A., M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature 1856-1991. *Journal Of Geophysical Research-Oceans*, 103, 18567-18589.

Katsaros, K. B., and Coauthors, 1994: Measurements of Humidity and Temperature in the Marine Environment during the HEXOS Main Experiment. *Journal of Atmospheric and Oceanic Technology*, 11, 964 - 981, DOI: 10.1175/1520-0426(1994)011<0964:MOHATI>2.0.CO;2.

Kent, E. C., and P. K. Taylor, 1997: Choice of a Beaufort Equivalent Scale. *Journal of Atmospheric and Oceanic Technology*, 14, 228 - 242, DOI: 10.1175/1520-0426(1997)014<0228:COABES>2.0.CO;2.

Kent, E. C., and D. I. Berry, 2005: Quantifying random measurement errors in Voluntary Observing Ship meteorological observations. *International Journal of Climatology*, 25, 843 - 852, DOI: 10.1002/joc.1167.

Kent, E. C., and P. K. Taylor, 2006: Toward Estimating Climatic Trends in SST. Part I: Methods of Measurement. *Journal of Atmospheric and Oceanic Technology*, 23, 464 - 475, DOI: 10.1175/JTECH1843.1.

Kent, E. C., and P. G. Challenor, 2006: Toward Estimating Climatic Trends in SST. Part II: Random Errors. *Journal of Atmospheric and Oceanic Technology*, 23, 476 - 486, DOI: 10.1175/JTECH1844.1.

Kent, E. C., and A. Kaplan, 2006: Toward Estimating Climatic Trends in SST. Part III: Systematic Biases. *Journal of Atmospheric and Oceanic Technology*, 23, 487 - 500, DOI: 10.1175/JTECH1845.1.

Kent, E. C., R. J. Tiddy, and P. K. Taylor, 1993a: Correction of Marine Air Temperature Observations for Solar Radiation Effects. *Journal of Atmospheric and Oceanic Technology*, 10, 900 - 906, DOI: 10.1175/1520-0426(1993)010<0900:COMATO>2.0.CO;2.

Kent, E. C., P. K. Taylor, and P. G. Challenor, 1998: A comparison of ship- and scatterometer-derived wind speed data in open ocean and coastal areas *International Journal of Remote Sensing*, 19, 3361 - 3381, DOI: 10.1080/014311698214046.

Kent, E. C., P. G. Challenor, and P. K. Taylor, 1999: A Statistical Determination of the Random Observational Errors Present in Voluntary Observing Ships Meteorological Reports. *Journal of Atmospheric and Oceanic Technology*, 16, 905 - 914, DOI: 10.1175/1520-0426(1999)016<0905:ASDOTR>2.0.CO;2.

Kent, E. C., P. K. Taylor, and P. G. Challenor, 2000: The Effect of Successive Correction on Variability Estimates for Climatological Datasets. *Journal of Climate*, 13, 1845 - 1857, DOI: 10.1175/1520-0442(2000)013<1845:TEOSCO>2.0.CO;2.

Kent, E. C., S. D. Woodruff, and D. I. Berry, 2007: Metadata from WMO Publication No. 47 and an Assessment of Voluntary Observing Ship Observation Heights in ICOADS. *Journal of Atmospheric and Oceanic Technology*, 24, 214 - 234, DOI: 10.1175/JTECH1949.1.

Kent, E. C., P. K. Taylor, B. S. Truscott, and J. S. Hopkins, 1993b: The Accuracy of Voluntary Observing Ships' Meteorological Observations-Results of the VSOP-NA. *Journal of Atmospheric and Oceanic Technology*, 10, 591 - 608, DOI: 10.1175/1520-0426(1993)010<0591:TAOVOS>2.0.CO;2.

Kistler, R., and Coauthors, 2001: The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bulletin Of The American Meteorological Society*, 82, 247-267, DOI: 10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2.

Kondo, J., 1975: Air-sea bulk transfer coefficients in diabatic conditions. *Boundary Layer Meteorology*, 9, 91 - 112, DOI: 10.1007/BF00232256.

Kraus, E. B., and J. A. Businger, 1994: *Atmosphere - ocean interaction*. 2 ed., 362 pp.

Kubota, M., N. Iwasaka, S. Kizu, M. Konda, and K. Kutsuwada, 2002: Japanese ocean flux data sets with use of remote sensing observations (J-OFURO). *J. Oceanogr.*, 58, 213 - 225, DOI: 10.1023/A:1015845321836.

Large, W. G., and S. Pond, 1981: Open ocean momentum flux measurements in moderate to strong winds. *Journal of Physical Oceanography*, 11, 324 - 336, DOI: 10.1175/1520-0485(1981)011<0324:OOMFMI>2.0.CO;2.

Large, W. G., and S. Pond, 1982: Sensible and latent heat fluxmeasurements over the ocean. *Journal of Physical Oceanography*, 12, 464 - 482, DOI: 10.1175/1520-0485(1982)012<0464:SALHFM>2.0.CO;2.

Legler, D. M., I. M. Navon, and J. J. Obrien, 1989: Objective Analysis Of Pseudostress Over The Indian-Ocean Using A Direct-Minimization Approach. *Monthly Weather Review*, 117, 709-720, DOI: 10.1175/1520-0493(1989)117<0709:OAOPOT>2.0.CO;2.

Lindau, R., 1995: A new Beaufort Equivalent Scale. *International COADS Winds Workshop*, Kiel, Germany, Institut fur Meereskunde, Kiel, and National Oceanic and Atmospheric Administration, 232 - 252.

Lindau, R., 2003: Errors of Atlantic Air–Sea Fluxes Derived from Ship Observations. *Journal of Climate*, 16, 783 - 788, DOI: 10.1175/1520-0442(2003)016<0783:EOAASF>2.0.CO;2.

List, R. J., 1951: *Smithsonian Meteorological Tables*. Sixth ed. Vol. 114, Smithsonian Institution, 527 pp.

Liu, W. T., K. B. Katsaros, and J. A. Businger, 1979: Bulk parameterization of air - sea exchanges of heat and water vapour including the molecular constraints at the interface. *Journal of the Atmospheric Sciences*, 36, 1722 - 1735, DOI: 10.1175/1520-0469(1979)036<1722:BPOASE>2.0.CO;2.

Lorenc, A. C., 1981: A Global Three - Dimensional Multivariate Statistical Interpolation Scheme. *Monthly Weather Review*, 109, 701-721, DOI: 10.1175/1520-0493(1981)109<0701:AGTDMS>2.0.CO;2.

Marshall, J., H. Johnson, and J. Goodman, 2001a: A Study of the Interaction of the North Atlantic Oscillation with Ocean Circulation. *Journal of Climate*, 14, 1399 - 1421, DOI: 10.1175/1520-0442(2001)014<1399:ASOTIO>2.0.CO;2.

Marshall, J., and Coauthors, 2001b: North Atlantic Climate Variability: Phenomena, impacts and mechanisms. *International Journal of Climatology*, 21, 1863 - 1898, DOI: 10.1002/joc.693.

McPhaden, M. J., and Coauthors, 1998: The tropical ocean global atmosphere observing system: A decade of progress. *Journal Of Geophysical Research-Oceans*, 103, 14169-14240.

Merchant, C. J., and Coauthors, 2008: Deriving a sea surface temperature record suitable for climate change research from the along track scanning radioemeters. *Advances in Space Research*, 41, 1 - 11, DOI: 10.1016/j.asr.2007.07.041.

Moat, B. I., M. J. Yelland, and A. F. Molland, 2006a: Quantifying the Airflow Distortion over Merchant Ships. Part II: Application of the Model Results. *Journal of Atmospheric and Oceanic Technology*, 23, 351 - 360, DOI: 10.1175/JTECH1859.1.

Moat, B. I., M. J. Yelland, R. W. Pascal, and A. F. Molland, 2006b: Quantifying the Airflow Distortion over Merchant Ships. Part I: Validation of a CFD Model. *Journal of Atmospheric and Oceanic Technology*, 23, 341 - 350, DOI: 10.1175/JTECH1858.1.

Moyer, K. A., and R. A. Weller, 1997: Observations of Surface Forcing from the Subduction Experiment: A Comparison with Global Model Products and Climatological Datasets. *Journal of Climate*, 10, 2725 - 2742, DOI: 10.1175/1520-0442(1997)010<2725:OOSFFT>2.0.CO;2.

NAG, 2008: *The NAG Fortran Library Manual, Mark 21*. The Numerical Algorithms Group Ltd.

O'Carroll, A. G., J. R. Eyre, and R. W. Saunders, 2008: Three-Way Error Analysis between AATSR, AMSR-E, and In Situ Sea Surface Temperature Observations. *Journal of Atmospheric and Oceanic Technology*, 25, 1197 - 1207, DOI: 10.1175/2007JTECHO542.1.

Oberhuber, J. M., 1988: An atlas based on the COADS data set: The budgets of heat, buoyancy and turbulent kinetic energy at the surface of the global ocean. MPI Rep. 15, 20 pp. plus figures pp.

Onogi, K., and Coauthors, 2007: The JRA-25 Reanalysis. *Journal of the Meteorological Society of Japan*, 85, 369 - 432, DOI: 10.2151/jmsj.85.369.

Oost, W. A., G. J. Komen, C. M. J. Jacobs, and C. van Oort, 2002: New evidence for a relation between wind stress and wave age from measurements during ASGAMAGE. *Boundary Layer Meteorology*, 103, 409 - 438, DOI: 10.1023/A:1014913624535.

Parker, D. E., C. K. Folland, and M. Jackson, 1995: Marine surface temperature: Observed variations and data requirements. *Climatic Change*, 31, 559-600, DOI: 10.1007/BF01095162.

Paulson, C. A., 1970: The Mathematical Representation of Wind Speed and Temperature Profiles in the Unstable Atmospheric Surface Layer. *Journal of Applied Meteorology*, 9, 857 - 861, DOI: 10.1175/1520-0450(1970)009<0857:TMROWS>2.0.CO;2.

Rayner, N. A., and Coauthors, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, 108, DOI: 10.1029/2002JD002670.

Rayner, N. A., and Coauthors, 2006: Improved Analyses of Changes and Uncertainties in Sea Surface Temperature Measured In Situ since the Mid-Nineteenth Century: The HadSST2 Dataset. *Journal of Climate*, 19, 446 - 469, DOI: 10.1175/JCLI3637.1.

Reynolds, R. W., 1988: A Real-Time Global Sea Surface Temperature Analysis. *Journal of Climate*, 1, 75 - 87, DOI: 10.1175/1520-0442(1988)001<0075:ARTGSS>2.0.CO;2.

Reynolds, R. W., 1993: Impact of Mount Pinatubo Aerosols on Satellite-derived Sea Surface Temperatures. *Journal of Climate*, 6, 768 - 774, DOI: 10.1175/1520-0442(1993)006<0768:IOMPAO>2.0.CO;2.

Reynolds, R. W., and T. M. Smith, 1994: Improved Global Sea Surface Temperature Analyses Using Optimum Interpolation. *Journal of Climate*, 7, 929 - 948, DOI: 10.1175/1520-0442(1994)007<0929:IGSSTA>2.0.CO;2.

Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax, 2007: Daily High-Resolution-Blended Analyses for Sea Surface Temperature. *Journal of Climate*, 20, 5473 - 5496, DOI: 10.1175/2007JCLI1824.1.

Scambos, T., T. Haran, C. Fowler, J. Maslanik, J. Key, and W. Emmery, 2000 (updated 2002): AVHRR Polar Pathfinder twice daily 1.25 km EASE Grid composites. National Snow and Ice Data Center.

Schulz, J., J. Meywerk, S. Ewald, and P. Schlussel, 1997: Evaluation of Satellite-Derived Latent Heat Fluxes. *Journal of Climate*, 10, 2782 - 2795, DOI: 10.1175/1520-0442(1997)010<2782:EOSDLH>2.0.CO;2.

Servain, J., A. J. Busalacchi, M. J. McPhaden, A. D. Moura, G. Reverdin, M. Vianna, and S. E. Zebiak, 1998: A Pilot Research Moored Array in the Tropical Atlantic (PIRATA). *Bulletin Of The American Meteorological Society*, 79, 2019 - 2031, DOI: 10.1175/1520-0477(1998)079<2019:APRMAI>2.0.CO;2.

Simmons, A. J., and Coauthors, 2004: Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature. *Journal of Geophysical Research*, 109, DOI: 10.1029/2004JD005306.

Smith, S. D., 1980: Wind stress and heat flux over the ocean in gale force winds. *Journal of Physical Oceanography*, 10, 709 - 726, DOI: 10.1175/1520-0485(1980)010<0709:WSAHFO>2.0.CO;2.

Smith, S. D., 1988: Coefficients for sea surface wind stress, heat flux and wind profiles as a funtion of wind speed and temperature. *Journal of Geophysical Research*, 93, 15467 - 15472.

Smith, S. R., D. M. Legler, and K. V. Verzone, 2001: Quantifying Uncertainties in NCEP Reanalyses Using High-Quality Research Vessel Observations. *Journal of Climate*, 14, 4062 - 4072, DOI: 10.1175/1520-0442(2001)014<4062:QUINRU>2.0.CO;2.

Smith, T. M., R. E. Livezey, and S. S. Shen, 1998: An Improved Method for Analyzing Sparse and Irregularly Distributed SST Data on a Regular Grid: The Tropical Pacific Ocean. *Journal of Climate*, 11, 1717 - 1729, DOI: 10.1175/1520-0442(1998)011<1717:AIMFAS>2.0.CO;2.

Stendel, M., J. R. Christy, and L. Bengtsson, 2000: Assessing levels of uncertainty in recent temperature time series. *Climate Dynamics*, 16, 587-601, DOI: 10.1007/s003820000064.

Sterl, A., 2001: On the impact of gap-filling algorithms on variability patterns of reconstructed oceanic surface fields. *Geophysical Research Letters*, 28, 2473 - 2476.

Sterl, A., 2004: On the (In)Homogeneity of Reanalysis Products. *Journal of Climate*, 17, 3866 - 3873, DOI: 10.1175/1520-0442(2004)017<3866:OTIORP>2.0.CO;2.

Stull, R. B., 1988: *An Introduction to Boundary Layer Meteorology* Klewer Academic Publishers, 666 pp.

Taylor, J. R., 1997: *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. 2nd ed. University Science Books, 327 pp.

Taylor, P. K., 1995: A program to calculate fluxes using the bulk formulae - BFORM. James Rennell Centre or Ocean Circulation, Southampton, 43 pp.

Taylor, P. K., and M. J. Yelland, 2001: The dependance of sea surface roughness on the height and steepness of the waves. *Journal of Physical Oceanography*, 31, 572 - 590, DOI: 10.1175/1520-0485(2001)031<0572:TDOSSR>2.0.CO;2.

Thejll, P., and T. Schmith, 2005: Limitations on regression analysis due to serially correlated resiudals: Application to climate reconstructions from proxies. *Journal of Geophysical Research*, 110, DOI: 10.1029/2005JD005895.

Thomas, B. R., E. C. Kent, and V. R. Swail, 2005: Methods to homogenize wind speeds from ships and buoys. *International Journal of Climatology*, 25, 979 - 995, DOI: 10.1002/joc.1176.

Thomas, B. R., E. C. Kent, V. R. Swail, and D. I. Berry, 2008: Trends in ship wind speeds adjusted for observation method and height. *International Journal of Climatology*, 28, 747 - 763, DOI: 10.1002/joc.1570.

UK Meteorological Office, 1964a: *Hygrometric Tables Part II: Stevenson Screen Readings Degrees Celsius*. Second ed. Her Majesty's Stationary Office, 80 pp.

UK Meteorological Office, 1964b: *Hygrometric Tables Part III: Aspirated Psychrometer Readings Degrees Celsius*. Second ed. Her Majesty's Stationary Office, 80 pp.

Uppala, S. M., and Coauthors, 2005: The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, 131, 2961-3012, DOI: 10.1256/qj.04.176.

Wentz, F. J., 1997: A well-callibrated ocean algorithm for special sensor microwave / imager. *Journal of Geophysical Research*, 102, 8703 - 8718.

Wentz, F. J., L. A. Mattox, and S. Peteherych, 1986: New Algorithms for Microwave Measurements of Ocean Wind: Applications to SeaSat and the Special Sensor Microwave Imager. *Journal of Geophysical Research*, 91, 2289 - 2307.

WGASF, 2000: Intercomparison and validation of ocean-atmosphere energy flux fields - Final report of the Joint WCRP/SCOR Working Group on Air-Sea Fluxes. WCRP-112, WMO/TD-1036. P. K. Taylor, Ed., 306 pp.

Willett, K. M., P. D. Jones, N. P. Gillett, and P. W. Thorne, 2008: Recent Changes in Surface Humidity: Development of the HadCRUH Dataset. *Journal of Climate*, 21, 5364 - 5383, DOI: 10.1175/2008JCLI2274.1.

Willis, J. K., D. Roemmich, and B. Cornuelle, 2004: Interannual variability in upper ocean heat content, temperature and thermosteric expansion on global scales. *Journal of Geophysical Research*, 109, DOI: 10.1029/2003JC002260.

Wolter, K., 1997: Trimming Problems and Remedies in COADS. *Journal of Climate*, 10, 1980 - 1997, DOI: 10.1175/1520-0442(1997)010<1980:TPARIC>2.0.CO;2.

Worley, S. J., S. D. Woodruff, R. W. Reynolds, S. J. Lubker, and N. Lott, 2005: ICOADS Release 2.1 Data and Products. *International Journal of Climatology*, 25, 823 - 842, DOI: 10.1002/joc.1166.

Yelland, M. J., B. I. Moat, R. W. Pascal, and D. I. Berry, 2002: CFD Model Estimates of the Airflow Distortion over Research Ships and the Impact on Momentum Flux Measurements. *Journal of Atmospheric and Oceanic Technology*, 19, 1477 - 1499, DOI: 10.1175/1520-0426(2002)019<1477:CMEOTA>2.0.CO;2.

Yevjevich, V., 1972: *Probability and Statistics in Hydrology*. Water Resources Publications, 302 pp.

Yu, L., and R. A. Weller, 2007: Objectively Analyzed Air - Sea Heat Fluxes for the Global Ice Free Oceans (1981 - 2005). *Bulletin Of The American Meteorological Society*, 88, 527 - 539, DOI: 10.1175/BAMS-88-4-527.

Yu, L., R. A. Weller, and B. Sun, 2004a: Improving Latent and Sensible Heat Flux Estimates for the Atlantic Ocean (1988–99) by a Synthesis Approach. *Journal of Climate*, 17, 373 - 393, DOI: 10.1175/1520-0442(2004)017<0373:ILASHF>2.0.CO;2.

Yu, L., R. A. Weller, and B. Sun, 2004b: Mean and Variability of the WHOI Daily Latent and Sensible Heat Fluxes at In Situ Flux Measurement Sites in the Atlantic Ocean. *Journal of Climate*, 17, 2096 - 2118, DOI: 10.1175/1520-0442(2004)017<2096:MAVOTW>2.0.CO;2.

Yu, L., X. Jin, and R. A. Weller, 2008: Multidecade Global Flux Datasets from the Objectively Analyzed Air-sea Fluxes (OAFlux) Project: Latent and Sensible Heat Fluxes, Ocean Evaporation, and Related Surface Meteorological Variables. Woods Hole Oceanographic Institution OAFlux Project Technical Report (OA-2008-01), 64 pp.