Comparison of Sampling Methods for Kriging Models

by

Michaela Claire Beckley

A dissertation submitted in partial fulfillment
of the requirements for the degree

Master of Science (Applied Sciences - Mechanics)

in the

Department of Mechanical and Aeronautical Engineering
Faculty of Engineering, Built Environment and Information Technology

University of Pretoria
Pretoria

June 19, 2014

ABSTRACT

Title: Comparison of Sampling Methods for Kriging Models

Author: Michaela Claire Beckley

Supervisors: Prof. S Kok

Department: Department of Mechanical and Aeronautical Engineering

Degree: Master of Science (Applied Sciences-Mechanics)

Keywords: Kriging; Updated Kriging Variance Algorithm; Maximum Likelihood Estimation; Oceanic Carbon Dioxide Flux; Kriging Variance

This study aims to generate from a three-dimensional data set of carbon dioxide flux in the Southern Ocean, a sample set for use with Kriging in order to generate estimated carbon dioxide flux values across the complete three-dimensional data set. In order to determine which sampling strategies are to be used with the three-dimensional data set, a number of *a-priori* and *a-posteriori* sampling methods are tested on a two-dimensional subset. These various sampling methods are used to determine whether or not the estimated error variance generated by Kriging is a good substitute for the true error as a measure of error.

Carbon dioxide is a well known "greenhouse gas" and is partially responsible for climate change. However, some anthropogenic carbon dioxide is absorbed by the oceans and as such, the oceans currently play a mitigating role in climate change by acting as a sink for carbon dioxide. It has been suggested that if the production of carbon dioxide continues unabated that the oceans may become a source rather than a sink for carbon dioxide. This would mean that the oceanic carbon dioxide flux (exchange of carbon dioxide between the atmosphere and the surface of the ocean) would invert. As such, modelling of the carbon dioxide flux is of clear importance. Additionally as the Southern Ocean is highly undersampled, a sampling strategy for this ocean which would allow for high levels of accuracy with small sample sizes would be ideal.

Kriging is a geostatistical weighted interpolation technique. The weights are based on the covariance structure of the data and the distances between points. In addition to an estimate at a point, Kriging also produces an estimated error variance which can be used as an indication of uncertainty. This study made use of model data for carbon dioxide flux in the Southern

Ocean. This data covers a year by making use of averaged data for 5 day intervals. This results in a three-dimensional data set covering latitude, longitude and time. This study used this data to generate a covariance structure for the data after the removal of trend and using this covariance structure, tested various sampling strategies in two dimensions, sampling approximately 10% of the two-dimensional data subset. These sampling strategies made use of either the estimated error variance or the true error and included two simple heuristics, genetic algorithms, the Updated Kriging Variance Algorithm and Random Sampling. Two of the genetic algorithms tested were selected to maximise the error measure of interest, in order to determine the full range of errors that could be generated. The percentage absolute errors obtained across these methods ranged from 2.1% to 64.4%.

Based on these strategies, the estimated error variance was determined to not be an accurate surrogate for true error and that in cases where absolute error is available, such as data minimisation, absolute error should be used as the measure of error. However, if no data is available then it does provide an easy to calculate measure of error. This study also concluded that Addition of a Point at Point of Maximum Absolute Error does provide a good validation sampling method to which other methods may be compared.

Additionally, based on true errors and computational requirements, three methods were selected to be implemented on a three-dimensional subset of the data. These methods were Random Sampling, Addition of a Point at Point of Maximum Absolute Error and Addition of a Point at Point of Maximum Estimated Error Variance. Each of these methods for sampling were performed twice on the data, sampling up to approximately 5% of the data. Random Sampling produced percentage absolute errors of 21.02% and 20.98%, Addition of a Point at Point of Maximum Estimated Error Variance produced errors of 18.54% and 18.55% while Addition of a Point at Point of Maximum Absolute Error was able to produce percentage absolute errors of 14.33% and 14.32%.

ACKNOWLEDGEMENTS

It is with immense gratitude that I acknowledge and thank my supervisor, Professor Schalk Kok, for all his help and guidance throughout this study.

Also my husband (Linton Ritchie), my parents (Eben and Loryn Beckley), my brother (James Beckley) and all of my grandparents (Cyril and Stella Carter, Athol and Frances Beckley) for their continued support and encouragement over the years.

Additionally, I wish to thank Doctor Kobie Smit for her guidance throughout the process.

A last thanks must be made to Doctor Onno Ubbink and the remainder of Advanced Mathematical Modelling, Modelling and Digital Science, Council for Scientific and Industrial Research for both their support and funding of this work, without your assistance this thesis would not exist.

"Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less." ∼ Marie Curie

CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

Symbol	Description of symbol
	Chemical Symbols
Ca^{2+}	Calcium Ion
CaC_3	Calcium Carbonate
$CO_{2(atmosp)}$	Atmospheric carbon dioxide
$CO_{2(aq)}$	Aqueous carbon dioxide
CO_3^{2-}	Carbonate Ion
H^+	Hydrogen Ion
HCO_3^-	Bicarbonate Ion
H_2CO_3	Carbonic acid
H_2O	Water
	Latin Alphabet
a	Range parameter
a_{lat}	Range parameter for latitude
a_{lon}	Range parameter for longitude
a_{time}	Range parameter for time
$\mathbf{C}$	Matrix of covariances between the sample point residuals
$C(\mathbf{0})$	Covariance between two points with a zero distance vector between them
$C(\cdot)$	Covariance function expressed using distance measure between two points
$C(\cdot,\cdot)$	Covariance function expressed using locations of two points
C_δ	Covariance function of δ_K
$C^{(s)}$	Spatial covariance function
$C^{(t)}$	Temporal covariance function
$\mathbf{c}_0$	Vector of covariances between target location and sample points
c_0	Sill Variance (Equivalent to $C(\mathbf{0})$)
D	$C(\mathbf{s}_{i+1},\mathbf{s}_{i+1}) - \mathbf{\Lambda}_i\mathbf{H}_i^{-1}\mathbf{\Lambda}_i^T$
$D_{\Re}$	Domain in which points may occur
$E[\ldots]$	Expected value
$g(\mathbf{S})$	Function to be minimised
$\mathbf{H}_i$	Appended matrix of covariance between i sample points
$\mathbf{H}_n$	Final appended matrix of covariance between sample points
$\mathbf{h}$	Vector of distance between two points
h	Distance between two points (One dimensional)
h_{lat}	Distance in latitude between two points
h_{lon}	Distance in longitude between two points
h_{time}	Distance in time between two points

Symbol	Description of symbol
$\mathbf{K}$	Matrix of covariances between sample points
K_ν	Bessel function of order ν
$\mathbf{k}$	Vector of covariances between sample points and point to be estimated
$\mathbf{L}$	Cholesky decomposition of R
l	Log-likelihood function
m	Total number of available sampling locations also referred to as the total number of points
m^*	Estimated trend in Regression Kriging
$m(\cdot)$	True trend in Regression Kriging
n	Number of Observed Samples
p	Number of terms in linear trend polynomial
$\mathbf{q}$	Matrix of additional variables at sample points
$\mathbf{q}_0$	Vector of additional variables at target point
q_k	Value of kth additional variable in trend model
$\mathbf{R}$	Correlation matrix
$\mathbf{R}^*$	Correlation matrix generated using estimated covariance parameters
R^2	Coefficient of determination
r	Power parameter
r_{i+1}	Covariance between target point and test sample point
$\hat{r}_{i+1}$	$\mathbf{\Lambda}_i\mathbf{H}_i^{-1}\mathbf{V}_i$
$\mathbf{S}$	Matrix of locations of sample points
$\mathbf{S}_0$	Initial set of sample locations
$\mathbf{S}_i$	Set of sample locations containing i locations
$\mathbf{S}_n$	Final set of sample locations
$\mathbf{s}$	Location/grid reference of point
$\mathbf{s}_i$	Location/grid reference of ith point
$\mathbf{s}_{\text{lat}}$	Grid reference/location in latitude of point
$\mathbf{s}_{\text{lon}}$	Grid reference/location in longitude of point
$\mathbf{s}_{\text{time}}$	Grid reference/location in time of point
$\mathbf{V}_i$	Appended vector of covariances between target point and i sample points
$\mathbf{V}_n$	Final appended vector of covariances between target point and sample points
$\mathbf{Z}$	Vector of values of variable of interest at sample points
Z	Variable of interest (in this study carbon dioxide flux)
Z^*	Kriging estimate of Z
Z^*_{RK}	Regression Kriging estimate for Z
Z^*_{SK}	Simple Kriging Estimate of Z
$Z^*(\mathbf{s}\|\mathbf{S})$	Kriging estimate for point $\mathbf{s}$ given sample locations $\mathbf{S}$
	Greek Alphabet
α	Power parameter with $\alpha < 2$
$\boldsymbol{\beta}$	Trend model coefficients
$\boldsymbol{\beta}^*$	Estimated trend model coefficients
$\boldsymbol{\beta}^*_{GLS}$	Generalised Least Squares estimate for trend model coefficients
$\boldsymbol{\beta}^*_{OLS}$	Ordinary Least Squares estimate for trend model coefficients
β_i	ith trend model coefficient
β^*_k	kth trend model coefficient
Γ	Gamma function
γ	Semivariance
δ	Kronecker delta function
δ_K	Stationary Gaussian process with zero mean
δ_{RK}	True residual in Regression Kriging
δ^*_{RK}	Interpolated residual in Regression Kriging

Symbol	Description of symbol
δ_{SK}	True Residual in Simple Kriging
δ^*_{SK}	Simple Kriging estimate for the Residual
$\boldsymbol{\theta}$	Covariance parameters
$\boldsymbol{\theta}^*$	Estimated covariance function parameters
$\boldsymbol{\theta}_2$	Covariance parameters excluding constant variance
$\boldsymbol{\Lambda}_i$	Appended vector of covariances between test sample point and sample points
$\boldsymbol{\lambda}^{SK}$	Vector of Simple Kriging weights
λ_i	ith Kriging weight
λ_i^{RK}	ith Regression Kriging Weight
λ_i^{SK}	ith Simple Kriging weight
μ	Constant and known mean value
$\mu(\cdot)$	Trend which is a function of $\cdot$
ν	Smoothness parameter
σ^2	Constant variance
σ^{2*}	Estimated constant variance
σ^2_E	Estimated error variance
σ^2_{RK}	Regression Kriging estimated error variance
σ^2_{SK}	Simple Kriging estimated error variance
σ^2_{lat}	Constant variance parameter for latitude covariance function
σ^2_{lon}	Constant variance parameter for longitude covariance function
σ^2_{time}	Constant variance parameter for time covariance function
$\sigma^{2*}[\delta^*_{RK}]$	Estimated error variance of residual
$\sigma^{2*}[m^*]$	Estimated error variance of trend
$\sigma^2(Z^*(\mathbf{s}\|\mathbf{S}))$	Estimated error variance for point $\mathbf{s}$ given sample locations $\mathbf{S}$
φ	Lagrangian parameter

CHAPTER 1

INTRODUCTION

1.1 Background

Kriging is a geostatistical interpolation method [13] which has long been used in mining [73]. Kriging was developed by a South African mining engineer D.G. Krige. Krige developed the method to improve his estimations of the gold ore grades in mining blocks by considering the ore grades in other blocks which are close by [73]. Kriging is different from classical interpolation schemes, in that it not only predicts a function value at any unsampled location, but it also provides an estimate of the uncertainty of the predicted value at this unsampled location (the Kriging variance). Another attractive feature of Kriging is that it can produce highly complex functional landscapes using relatively few sample locations. This last feature of Kriging makes it particularly popular as a surrogate model. Since the quality of a Kriging based surrogate model depends on the location of the samples, it remains an open question how to determine the sampling locations. Techniques that attempt to answer this question are referred to as strategic sampling schemes. In order to demonstrate the relevance of research in Kriging and/or strategic sampling schemes, consider the following list of prior work:

- Aerodynamic Optimisation - Liu *et al.* [43], Rosenbaum and Schulz [66], Laurenceau and Sagaut [41] and Paul-Dubois-Taine and Nadarajah [57] have all made use of Kriging surrogate models in aerodynamic optimisation. These authors have considered various sampling schemes in order to lower the computational cost of the computational fluid dynamics simulations by means of these surrogate models.
- Agricultural Soil Science - Pereira *et al.* [58] made use of a strategic sampling scheme which concerns Kriging as recently as 2013 in order to determine the sampling scheme which should be used to map physical and chemical properties of soil in agriculture.
- Weed Mapping - Cousens *et al.* [12] made use of Kriging in order to produce weed maps. Their article focused on the grid spacing, quadrant size and starting point for the sampling required.
- Contamination Classification - Both Juang, Lee and Chen [37] and more recently Juang, Lee and Teng [38] have made use of Kriging in the classification and estimation of heavy metal contamination of soils.
- Water Quality Monitoring - Hedger *et al.* [29] made use of a systematic sampling scheme in order to reduce the redundancy in observations for the estimation of water quality in lakes when remote sensing and Kriging are used.

- Groundwater Water Level Prediction - Varouchakis and Hristopulos [72] have made use of Kriging to improve the groundwater level prediction in the Mires basin in Crete.
- Underwater Autonomous Vehicle Sampling - Ho and Saripalli [33] made use of Kriging as the interpolation function for various sampling strategies for use with an underwater autonomous sampling vehicle.
- Reliability Studies - Liu *et al.* [44] made use of Kriging combined with Importance Sampling in 2012 in the area of structural reliability.
- Estimation of Environmental Variables - Brus and Heuvelink [9] have considered the optimisation of sampling schemes for various Kriging models as well as a Multiple Linear Regression model.
- Large Scale Ocean Sampling - Zhu *et al.* [75] has made use of Kriging in large scale ocean sampling. They made use of a method called the Updated Kriging Variance Algorithm in order to systematically select a set of sample points.
- Data Minimisation - Brodkin [8] made use of a systematic sampling strategy using Kriging in order to minimise a large data set to lessen the storage space required.
- Terrain Modelling for Flight Simulation - Duckett [22] made use of the same method as Brodkin in order to reduce the sampled gridded terrain elevation data for use in terrain modelling for flight simulators [22].

In the case of data minimisation, the variable of interest is known at all of the data points, while for large scale ocean modelling and terrain modelling this may not be the case. For large scale ocean modelling, it is possible that a model may be available which would give a predicted value at each point in the set of possible points. This model is, however, unlikely to perfectly predict the real world values. Based on the availability of values for the variable of interest, *a-posteriori* sampling may be allowable or only *a-priori* sampling may take place in some instances. In this study, a large model data set is available. Therefore, both *a-priori* and *a-posteriori* methods will be implemented and compared.

With the increasing availability of data especially, spatio-temporal data, the expansion of sampling methods for use with Kriging in three dimensions is becoming increasingly important. However, due to computational costs, there is the possibility that some sampling methods may not be feasible in three dimensions. While in general data is becoming more available, certain variables and/or areas are vastly undersampled, the Southern Ocean being an example of such an area [52]. Within this area, the carbon dioxide flux (that is the exchange of carbon dioxide between the ocean and the atmosphere at the surface interface) is also undersampled [52].

Carbon dioxide is well known as a prominent so called “green-house gas” and plays a role in climate change. The estimates for the percentage of anthropogenic carbon dioxide absorbed by the oceans ranges from $25 - 48\%$ [10]. Thus, it is clear that the oceans play a large role in mitigating climate change. However, some studies have suggested that if the production of anthropogenic carbon dioxide was to continue to increase as it has been, that the oceans could become a source of carbon dioxide rather than a sink [18]. The absorption or release of carbon dioxide is known as the carbon dioxide flux. This flux or exchange of carbon dioxide is driven by the concentration gradients of carbon dioxide at the sea-air surface interface [3]. The absorption of carbon dioxide by the oceans also has an effect on their state and the lifeforms which inhabit in them [21].

According to Lenton *et al.* [42], the Southern Ocean alone may absorb as much as 25% of anthropogenic carbon dioxide. However, as previously stated it is highly undersampled,

making the understanding of the absorption and effects of carbon dioxide in the Southern Ocean difficult to quantify [65]. Thus, sampling the Southern Ocean is clearly a necessary exercise. However, there are many difficulties associated with sampling in the Southern Ocean. These include the distances from ports and shipping routes [52], the presence of seasonal sea ice [70] and the extreme weather conditions [65]. This implies that it would be ideal if an accurate map of the carbon dioxide flux could be produced with few sample points.

1.2 Thesis Objectives

Kriging produces both a prediction and a Kriging variance [46] (also known as an estimated error variance [17]) for each point in the set of possible points. Some authors such as Rosenbaum and Schulz [66] refer to this variance as the mean squared error. This variance has been used by some authors in order to obtain more optimal sample sets. Zhu *et al.* [75] have made use of the estimated error variance in the Updated Kriging Variance Algorithm. This method adds points incrementally and bases the addition of the points on the maximum decrease in total estimated error variance. Brodkin [8] and Duckett [22] make use of the estimated error variance by adding a single point at a time to the sample set. They select this point as the point with the maximum estimated error variance. They continue to add points until the maximum estimated error variance falls below a pre-determined threshold.

Brus and Heuvelink [9] make use of simulated annealing to minimise the spatially averaged Universal Kriging Variance (the spatially averaged estimated error variance for Universal Kriging) in order to determine an optimised sampling strategy. This estimated error variance incorporates the trend estimation error. The spatially averaged estimated error variance for Ordinary Kriging is also minimised to determine an optimised sampling strategy for Ordinary Kriging. Pereira *et al.* [58] also make use of spatial simulated annealing to minimise the averaged estimated error variance from Ordinary Kriging. Hedger *et al.* [29] make use of grid configurations for sampling which produce lower estimated error variances than those produced by random samplings.

Liu *et al.* [43] test various infilling sampling strategies for aerodynamic optimisation, one of which involves infilling based on maximum estimated error variance while others are designed to maximise expected improvements or the probability of improvement. These infilling methods are used in conjunction with genetic algorithms to determine the new sample locations. Laurenceau and Sagaut [41] and Rosenbaum and Schulz [66] make use of a sampling method which adds points at the point of maximum absolute error purely as a validation method for the other methods they test. Both articles consider sample sensitivity error sampling, while Rosenbaum and Schulz also consider the addition of a point at the point of maximum estimated error variance.

However, there remains a question as to whether or not the estimated error variance is a good measure of true error. Makhnin [46] is very clear that the estimated error variance is dependent only on the spatial position of the sample points and the point at which prediction is occurring, and the covariance structure of the variable of interest. Juang *et al.* [37] state that the estimated error variance cannot be used as a measure of estimation accuracy in general because it is independent of the data values. This suggests that the estimated error variance may not provide an accurate surrogate for the true error. However, as described above, there are still many methods which rely solely on this estimated error variance as it allows for a sampling strategy to be selected without having available all values for the variable of interest. Thus, this study wishes to investigate the value of the estimated error variance as a surrogate measure of error for the true absolute error. Given the model data set available for this study, which yields values for the variable of interest at all possible sample points, this study is able to accurately compare these two measures of error.

While the sample sensitivity error sampling is useful with aerodynamic optimisation, it is not considered in this study. This study wishes to focus on the apparent contradiction between author recommendations regarding the use of estimated error variance in the design of sampling strategies. Additionally, this study considers a large 3D spatio-temporal set in its later stages and this method of sampling is not particularly suited to spatio-temporal sampling. This method bases the location of the next sample point on the distances between points and the current sample values. However, this implies that one must sample the next point before determining the following sample point location. This implies that one could be required to wait long periods of time before one is able to sample the next chosen location due to its position in time. Only in a case such as ours, where values are available for the variable at all locations, could this method be used for spatio-temporal data but in such a case, one would prefer to work with the true errors. This method could, however, be considered as a future extension to this study.

Given a large full data set of carbon dioxide flux values for the Southern Ocean generated from a model, this study aims to meet the following objectives:

- Compare various *a-priori* and *a-posteriori* methods for sampling on a two-dimensional set when Kriging is used as the method of prediction. This includes the comparison of both errors and computational costs in order to determine which methods may be feasible on a large three-dimensional data set.
- Compare the true errors and estimated error variances in order to determine if the estimated error variance is an accurate surrogate for the true error. Determine whether or not the addition of a point at the point of maximum absolute error is a good validation function for future studies.
- Comparing selected methods on a three-dimensional data set in order to determine the accuracy attainable on that set by means of Kriging using the sampled sets of a preselected size.

1.3 Delineation and Limitations

A very large data set is available. The complete set (with land and ice removed) consists of 419093 points. This complete data set is too large for many of the methods employed in this study, based on the available computational power. Thus, some sub-sampled sets are to be generated. Some of these sets will be randomly selected from the complete set, while others will be selected by certain criteria. The two-dimensional subset to be used for comparison of all the tested methods, will simply be taken as a portion of the larger Southern Ocean area for one time step, while for the three-dimensional subset, every second grid reference in every direction will be considered admissible to the subset. The randomly chosen subsets will be used in the estimation of the covariance function and in the validation of the functions which are to be fitted.

In order to adequately compare the various sampling methods, a predetermined covariance structure is required. Before a covariance structure can be fitted, it needs to be determined if there is a trend present in the data. The determination of presence of trend is to be completed by means of histograms and Ordinary Least Squares estimation. Ordinary Least Squares is not optimal when the data are not independent, but based on some studies may be sufficient [31]. While any structure may be used for the trend, this study considers only three linear polynomial trends. The set of possible trend structures is too vast to consider every possibility and thus, three common polynomial trends (a constant, linear and quadratic trend) are to be

considered. These trends will be taken in terms of the grid references for latitude, longitude and time, as no auxiliary information is available. While many other methods are available to assess the trend, Ordinary Least Squares is chosen as it is computationally inexpensive and thus, provides a fast method to determine which of the trend structures is more optimal.

Many techniques are available for the fitting of covariance structures. These include but are not restricted to, Maximum Likelihood Estimation, Restricted Maximum Likelihood Estimation, Method of Moments Estimation, Generalised Least Squares Fitting and Cross-Validation [15; 40; 50]. While any of these methods would be sufficient, Maximum Likelihood Estimation allows for the estimation of both covariance and trend parameters simultaneously and additionally may perform well even in the presence of non-normal data [50]. However, Maximum Likelihood Estimation is computationally expensive and as such, the method will be performed on randomly chosen smaller subsets in order to help alleviate some of the computational burden. Additionally, due to computational expense, only three covariance structures will be fitted and randomly chosen validation subsets are to be utilised to determine which of the three covariance structures is to be used as the pre-determined covariance structure for the comparison of sampling methods.

While many covariance structures are available, only three will be fitted along with the previously determined trend structure. Amongst the available covariance structure properties are isotropic, anisotropic, separable and non-separable options [15; 73]. There is also the choice between including or excluding the nugget in the structure [15; 73]. Many covariance functions are available for use within these conditions on the structure. Some of these covariance functions include the Gaussian covariance function, exponential covariance function, spherical covariance function, pentaspherical covariance function and the Matérn covariance function. This list is not comprehensive. In order to maintain a simple covariance structure for fitment, the structure was chosen to be separable. Anisotropy will not be considered due to the separability and the nugget will not be included. This general structure will be used with three popular covariance functions, namely the exponential, Gaussian and pentaspherical covariance functions.

Of these covariance structures estimated by Maximum Likelihood Estimation, the one which provides the lowest errors on the randomly chosen validation subsets will be used further in this study. Once the covariance and trend structures and parameters have been determined, the trend will be treated as deterministic if the parameters are deemed to have converged sufficiently. This would reduce the case from Regression Kriging to Simple Kriging on the residuals as the trend would be deemed to be fully known.

Once this has all been completed, the various sampling methods may be implemented on the two-dimensional set and their errors compared. While the range of sampling techniques for use with the method of Kriging continues to expand and become more complex, the methods chosen for use in this study, reflect some of the more simple available methods. Two of these methods are simple heuristics which allow them to be computationally inexpensive. These two heuristics will decide which single point to add to the sample set per iteration, based on the simple criteria of maximum absolute error or maximum estimated error variance. As the optimal sample set is computationally infeasible to locate, genetic algorithms will be implemented in order to attempt to approximate a good possible solution and the associated errors. Additionally, genetic algorithms will be used to approximate some of the worst possible errors attainable in order to determine where within the range of possible errors the techniques tested lie. These genetic algorithms will be used to both maximise and minimise the total absolute error and the total estimated error variance. Additionally, for the minimisation of total absolute error and total estimated error variance, hybridised genetic algorithms will be implemented which take an initial solution from one of the two simple heuristic techniques.

In addition to the simple heuristics and genetic algorithms, both Random Sampling and the

Updated Kriging Variance Algorithm (used by Zhu *et al.* [75]) will be implemented. These methods encompass *a-priori* and *a-posteriori* methods which rely on either the estimated error variance or the absolute error respectively. The errors and placement of the sample points for each of these methods will then be compared. Also compared will be the computational requirements of these methods. As the values are known at all data points, this data set provides an opportunity to compare not only the errors produced by these methods, but also the effectiveness of the estimated error variance as a substitute for the absolute error on a non-ideal data set. Based on these various factors, certain methods will then be implemented on the three-dimensional subset and the errors and predictions presented. In all the methods tested, it is assumed that sampling can occur at any time and location without any restrictions.

1.4 Significance of Study

This study aims to demonstrate whether or not the estimated error variance is indeed a good surrogate for the true error. Brodkin [8] makes use of the estimated error variance for data minimisation. However, in such a case, the absolute error is available for use. This study aims to determine whether or not true error should rather be used in cases of data minimisation.

Additionally this study aims to find methods of reasonable computational expense (in comparison to other methods) which can be used on a large three-dimensional data set. This study also aims to show that a three-dimensional data set can be sub-sampled and have the predicted values provide some level of accuracy. The data set of which this sampling occurs is a spatio-temporal model for the carbon dioxide flux in the Southern Ocean and given the need for sampling in this area, it may provide an easily obtainable starting point for more accurate methods of selecting sampling points.

1.5 Overview of Chapters

Chapter 2 gives a brief review of the effects of carbon dioxide in both the oceans as a whole and specifically the Southern Ocean. Chapter 2 also explains the importance of and need for sampling in the Southern Ocean. Thus, it provides the necessary oceanic background for this study.

Chapter 3 provides the mathematical background to Kriging as well as the necessary information regarding covariance functions as required for this study. This chapter in no way encompasses the full spectrum of available literature for Kriging but merely explores the mathematics surrounding it, which are used in this study. The theory of Maximum Likelihood Estimation for the prediction of trend and covariance parameters necessary for Kriging is also addressed in this chapter.

Chapter 4 addresses the literature concerning the sampling methods to be implemented. While sampling strategies for use with Kriging is a field which is continuously growing and new methods are being explored, only a few methods have been tested in this study. This chapter also covers the mathematical preliminaries which are required in order to implement the sampling strategies used in this study.

Chapter 5 describes the origins of the data used in this study and explains the references used to describe the data, while Chapter 6 describes the processes followed in order to determine the presence of trend and the resulting decisions in the removal of trend from the

data. Chapter 7 documents the Maximum Likelihood Estimation fitments of three covariance structures making use of the trend structure selected in Chapter 6. This chapter also covers the comparison of these covariances structures and the choice between them.

Chapter 8 covers the basic information relating to the two-dimensional sampling performed in this study. The various sampling methods which are used on the two-dimensional set described in this chapter are addressed in Chapters 9 to 13. Chapter 9 addresses the Random Sampling method as used in this study and also contains the results from this sampling method. Chapter 10 covers the Updated Kriging Variance Algorithm and its results while Chapters 11 and 12 cover the Addition of a Point at Point of Maximum Estimated Error Variance and Addition of a Point at Point of Maximum Absolute Error and their results respectively. The genetic algorithms and their results can all be found in Chapter 13.

The comparison of the methods found in Chapters 9 to 13 can be found in Chapter 14. This chapter addresses not only a comparison of the errors and placement of sample points but also a comparison of the computational requirements of each of the methods. Also in this chapter, a decision is made as to which methods should be implemented for the large three-dimensional data set based on the comparisons of error and computational requirements.

Chapter 15 describes the three-dimensional set on which three-dimensional sampling takes place and additionally also provides the results of the sampling in three dimensions for the sampling methods selected in Chapter 14.

Chapter 16 contains the conclusions of this study as well as suggestions for future work and a short summary of contributions.

Appendix B contains the Ordinary Least Squares Coefficients for the various trends which are tested in Chapter 6. The Maximum Likelihood Estimation results for the various runs across the various covariance structures tested in Chapter 7 are given in Appendix C. This Appendix also contains the histograms of the parameter estimations obtained across the various runs of Maximum Likelihood Estimation of the tested covariance structures.

Appendix D contains additional graphs of errors and additional maps of predicted carbon dioxide flux and estimated error variances from the various two-dimensional sampling methods implemented across Chapters 9 to 13.

Appendix E provides tables of computational requirements for the methods implemented on the two-dimensional data sub-set. Finally, the additional graphs and histograms of errors as well as the maps of predicted carbon dioxide flux and estimated error variance for the three-dimensional sampling methods implemented in Chapter 15 can be found in Appendix F.

CHAPTER 2

OCEANIC OVERVIEW

2.1 Carbon Dioxide in the Oceans

2.1.1 The Importance of Carbon Dioxide in the Oceans

Carbon dioxide is well known as a so called Greenhouse gas and as such is partially responsible for climate change. Prentice *et al.* [59] suggest that 40% of anthropogenic carbon dioxide (that is carbon dioxide generated by human activities) remains in the atmosphere, while 60% is absorbed by oceanic and terrestrial systems and that this uptake occurs in equal percentages. That suggests that the oceans may absorb 30% of anthropogenic carbon dioxide. Others suggest that the percentage of anthropogenic carbon dioxide absorbed by the oceans is 25-48% [10], 40% [16],17-39% [3] or 33% [51]. Based on these figures, it is clear that the ocean plays a significant role in carbon dioxide absorption and, thus, climate change. It was explained by Kaiser and Barnes [39] that the concentration of carbon dioxide in ice cores from the Canadian Islands, Greenland and Antarctica is higher now than at any other time in 800 thousand years of Earth's history and that this concentration is rising at a unprecedented rate, giving extra urgency to the attempts to accurately model carbon dioxide in the oceans.

The exchange of carbon dioxide between the oceans and atmosphere is governed by physical transport at their interface and it is well known that this exchange is driven by the concentration gradients of carbon dioxide [3]. The suggestion has been made by some researchers such as Deng and Chen [18] that these fluxes could invert causing the oceans to become a source of carbon dioxide rather than a sink, while others such as Doney *et al.* [20] believe that the ocean is weakening as a sink.

Once atmospheric carbon dioxide is absorbed by the oceans, it reacts with the water to form carbonic acid. This carbonic acid can dissociate by the loss of hydrogen ions into bicarbonate and carbonate ions [21]. This chemical equation is given by [21]

$$CO_{2(atmosp)} \rightleftharpoons CO_{2(aq)} + H_2O \rightleftharpoons H_2CO_3 \rightleftharpoons H^+ + HCO_3^- \rightleftharpoons 2H^+ + CO_3^{2-}. \tag{2.1}$$

The above chemical reactions are reversible and have been found to be near equilibrium [21]. For a pH of approximately 8.1, the inorganic carbon occurs in the following percentages per form: 90% bicarbonate ions, 9% carbonate ion and 1% dissolved carbon dioxide [21].

When carbon dioxide is added to the oceans, the concentrations of aqueous carbon dioxide, bicarbonate and hydrogen ions increase. This increase in hydrogen ions lowers the pH causing

acidification of the oceans [21]. With respect to ocean acidification, the following authors agree that carbon dioxide uptake by the oceans has the result of surface water acidification: Maier-Reimer *et al.* [45], McNeil [51], Meredith *et al.* [52] and Rintoul *et al.* [65]. In addition Meredith *et al.* [52] holds the opinion that this ocean acidification could have implications of the food webs of the oceans. The above mentioned increase in hydrogen ions also lowers the concentration of carbonate ions [21].

The ability of the ocean to absorb atmospheric carbon dioxide over long time frames is dependent on the dissolution of calcium carbonate in water columns and sediment [21]. This calcium carbonate is found in the shells and skeletons of many marine organisms, and it is believed that ocean acidification could effect these organisms [21]. The equation for the dissociation of calcium carbonate is [21]:

$$CaCO_3 \rightleftharpoons CO_3^{2-} + Ca^{2+}. \tag{2.2}$$

According to Rahmstrof [62] climate change could effect thermohaline circulation. Thermohaline circulation is driven by freshwater and heat fluxes and one of its key features is deep water formation. If the thermohaline effect is weakened (as is suggested will happen if climate change continues), it would influence the global climate as thermohaline circulation is responsible for large scale heat transport [62]. Rahmstorf [62] also suggests there is a risk that the changes to thermohaline circulation may be abrupt and/or irreversible.

Thus accurate modelling is clearly necessary to give clarity on the severity of the environmental effects of carbon dioxide in the ocean, such as those mentioned above. In addition McNeil [51] is of the opinion that there are possible socio-economic implications that could arise from accurate carbon dioxide modelling for the oceans. He has suggested that if the EEZs (exclusive economic zones) of countries were to be added to their carbon dioxide budgets, the budgets for some countries could be dramatically changed and this could potentially have positive influences on some countries' economies [51].

2.1.2 The Southern Ocean and Carbon Dioxide

Chen *et al.* [10] stated that the Southern Ocean accounts for 20% of all ocean area, while Lenton *et al.* [42] states that although the Southern Ocean accounts for less than 20% of the global ocean area, it may account for 25% of the oceanic uptake of anthropogenic carbon dioxide. Gille [25] has stated that the Southern Ocean is expected to be a prime region for the uptake of carbon dioxide although this does not necessarily apply to the storage of carbon dioxide. Metzl *et al.* [53] suggest that the Southern Ocean is a summer sink and winter source and while Takahashi *et al.* [70] agreed with the findings of a summer sink and winter source in the ice free zone of the Southern Ocean, both results were calculated using data which contained more data points for the summer months than for the winter months. Thus, the Southern Ocean is clearly an important factor in the calculation of global oceanic carbon uptake.

It is well known that the Southern Ocean connects the major ocean basins and is the link between the upper and lower layers of oceanic circulation [52; 65]. These and other processes that take place within the Southern Ocean have a great influence on both regional and global climate [65], global oceanic circulation [65] and biogeochemical cycles [65; 52]. Meredith *et al.* [52] state that the Southern Ocean plays a fundamental role in the global Earth system.

In linking the upper and lower layers of global oceanic circulation, the Southern Ocean is responsible for the movement of carbon, heat, nutrients and other oceanic properties between the surface and deep ocean via the upwelling and downwelling branches of the global

overturning circulation found in the Southern Ocean [65]. This also means that the Southern Ocean is responsible for the ventilation of a large portion of the global ocean and may regulate the ocean's capacity in terms of heat and carbon storage [64].

Given the global influence of the Southern Ocean, changes in its waters could have significant and far reaching consequences [65] and models suggest that the Southern Ocean will respond rapidly to climate change [25]. In fact, based on the existing observations, it has been established that the Southern Ocean influences the mean state of not only the global ocean but also the global climate [64]. However, given the limited observations available, the extent of its influence has not been clearly identified [64]. It should be noted, however, that the limited existing observations suggest various changes are already occurring in the Southern Ocean. These changes are of concern given the influence of the Southern Ocean [52]. Some of these changes are listed below:

- Hotspots of extreme warming appear to be occurring [52]
- Regional warming is occurring more rapidly than the global average [65]
- Salinity changes due to changes in precipitation and the melting sea ice and ice shelves [65]
- Rapid changes in the regional sea ice and ice shelves which has effects on the global sea level [52]
- Ocean acidification appears to be occurring, while the uptake of carbon appears to be slowing [65]
- There are indications that ecosystems within the Southern Ocean are changing [65]

Thus, it appears that the Southern Ocean is a definite area of interest when modelling carbon dioxide in the oceans.

Sampling of the Southern Ocean

As previously stated there are only limited observations in the Southern Ocean and it is in fact regarded as being severely under-sampled [52; 70; 64; 25; 42]. Some of the reasons for this undersampling are:

- The presence of partial ice cover in winter [70]
- The Southern Ocean covers a large area [25]
- The Southern Ocean has very few ports for access [25] and is distant from shipping routes and densely populated areas of land making it a very remote ocean [64; 52]
- The Southern Ocean is well known for it's harsh environmental conditions including strong winds, ice and freezing temperatures [25; 52; 64]
- Winter cloud cover which inhibits the use of satellites to gather remotely sensed data [20]

The lack of existing observations and the availability of only short and incomplete time series make the determination of cause and consequences of changes in the Southern Ocean difficult [65]. In addition to the under-sampling, there is currently a lack of data sharing amongst

various institutions and countries [52]. There is a critical need for observation in the Southern Ocean in order to better understand its effects on global environmental changes, to increase our ability to become resilient to these changes and to discover methods for sustainable resource usage [52].

In order to meet the need for additional observations in the Southern Ocean, various institutions and organisations have come together to develop a Southern Ocean Observing System (SOOS) [52]. The SOOS was brought about in order to help achieve the following 6 goals [52]:

- To understand the role of the Southern Ocean in global freshwater and heat balance.
- To investigate the stability of the overturning circulation in the Southern Ocean.
- To understand the role the Southern Ocean plays in the stability of the Antarctic ice-sheet and its contribution to future global sea level rises.
- To determine the future of carbon uptake in the Southern Ocean and the consequences of said carbon uptake.
- To determine the future of Antarctic sea ice.
- To investigate the impacts that global changes may have on the Southern Ocean and its ecosystems.

The traditional methods of data collection in the oceans are labour intensive and have relied on research ships undertaking scientific expeditions [52]. Thus, other methods of data collection need to be considered in order to greatly increase the number of observations in the Southern Ocean. Various additional sampling techniques have been suggested for use in addition to ship based samples and it is hoped to integrate autonomously generated data with human generated data [52]. A short summary of additional sampling techniques is given below.

- Moored Instruments could potentially be very important as an additional data source [20; 52]
- Remotely sensed measurements from satellites [52].
- Profiling floats (such as the Argo floats) and drifters [20]. The Argo floats are robotic profiling floats and recently there have been some advancements in sending some of these floats into polar ice regions [52].
- Tagged animals. Marine animals are often not limited by the ice and thus, tagged animals would allow for some measurements in the ice regions that may previously have been unavailable [52]
- There is also gathering interest and momentum for the possible use of ocean gliders and other autonomous underwater vehicles for data collection [52].

In addition to increasing the observations in the Southern Ocean, it falls within the SOOS's vision to develop methods of giving gliders automated instructions in order to keep spatial data coverage at an optimum [52]. However, this relies on models becoming more sophisticated [52]. Meredith *et al.* [52] also note that real time minimisation of model errors when used by adaptive autonomous sampling could offer a "step-change" in the optimal monitoring of the Southern Ocean.

It is clear that sampling of the Southern Ocean is both necessary and urgent. In addition sampling techniques which allow for point by point selection for sampling may be of further use in conjunction with autonomous vehicles and/or gliders.

CHAPTER 3

MATHEMATICAL OVERVIEW:COVARIANCE FUNCTIONS AND KRIGING

3.1 The Method of Kriging

3.1.1 Background

Kriging is a well known Geostatistical technique. It was first developed in a very rudimentary emperical form by D.G Krige in the 1950s [13]. Krige was a South African mining engineer who developed the technique to improve his estimations of the gold ore grades in mining blocks by considering the ore grades in other blocks which are close by [73]. The name Kriging was given to the method by G.Matheron who was a mathematician in the Paris Mining School [73]. It was Matheron who was able to provide the theory behind the method and was able to add the missing components to Krige's method which have transformed it into Kriging as we know it today [13].

Cressie [13] makes use of the following quote from Matheron to describe Kriging in his 1990 paper entitled the "Origins of Kriging", "It consists [of predicting] the grade of a panel by computing the weighted average of available samples .... The suitable weights a_i... are determined by .... $\Sigma a_i = 1$ ... [and the prediction] variance ... should take the smallest possible value."

The aim of Kriging is to estimate the value of a variable of interest (Z) at a point or points which have not been sampled [73]. Kriging in its simpler linear forms is merely a weighted average of the available values. The weights are, however, dependent of the distances between points and the covariance structure of the variable [73]. In addition to the estimation of a value at given coordinates, Kriging also allows for the calculation of an estimated variance of the prediction error at the unsampled points [30]. This is known as the error variance [7], prediction variance [67] or Kriging variance [46]. This Kriging variance is also called the estimated error variance by De-Vitry [17] and this term for the Kriging variance is used throughout the remainder of this document as it provides an accurate description of the Kriging variance as it is described by Hengl [30]. This estimated error variance gives an indication of how precise the prediction is [30]. The weights are also estimated in such a way as to minimise this error variance and to ensure that the estimates are unbiased [73]. This leads to the Kriging estimator being referred to as a B.L.U.E (best linear unbiased estimator) [34]. This name is used as it is a linear predictor (the estimates are a weighted average of known values), it is unbiased because it attempts to make the mean residual error zero and it is known as best because it attempts to minimise the error variance [34].

The Kriging family as it now stands, however, encompasses both linear and non-linear forms of Kriging and has been further developed into various distinct types for use with a variety of problems [73].

Listed below are some of the more common forms of Kriging used today:

- Simple Kriging - It is assumed that there is a constant known mean and prediction is performed by a weighted linear sum of the known values at sample points [73] which minimises the mean squared prediction error [13]. The weights are dependent on the spatial positions of the sample points and the covariance function between them [13].
- Ordinary Kriging - In Ordinary Kriging it is assumed that there is a constant but unknown mean. This unknown mean must then be estimated in addition to the weights [73]. This is the most commonly used form of Kriging [73].
- Lognormal Kriging - Lognormal Kriging is merely Ordinary Kriging of the logarithms of the variable of interest. This method is for use with data which exhibit an approximate log-normal distribution [73].
- Universal Kriging - Also known as Kriging with External Drift or Kriging with Drift [73]. Universal Kriging recognises two components in the variable of interest: i) a non-stationary deterministic component and ii) a random component [73]. This method estimates the trend in the deterministic component and the covariance or variogram from the random component [73]. In this method the covariance matrix of the residuals is extended with the additional predictors in order to predict the trend and random component simultaneously [30]. Hengl [30] suggests that the name Universal Kriging should be used when the trend is only a function of the spatial co-ordinates, and Kriging with external Drift should be used when the trend is a function which includes other measurements. It is noted that the reference to spatial coordinates does not exclude temporal coordinates as time can be viewed as just another dimension for Kriging [14].
- Regression Kriging - Very similar to Universal Kriging except that the trend and random components are predicted separately and summed at the end. Regression and Universal Kriging are equivalent but do, however, follow different computational steps [30]. For a proof of the equivalence of the two methods see Hengl *et al.* (2003) [32]. Many authors do, however, use the names Universal and Regression Kriging interchangeably [32]
- Ordinary Cokriging - Also known as Cokriging. This method is an extension of Ordinary Kriging for use when there are two or more variables [73]. Although only one variable is predicted, the other variables are used in the prediction [47]. The variable should exhibit some coregionalization [73]. This method is particularly useful when there are properties which are cheap to measure and one which is more expensive. The cheaper to measure variables can be used to help predict the more expensive one more accurately [73]. Cross-covariances are, however, required [47].
- Indicator Kriging - This is a non-linear, non-parametric form. In this method continuous variables are converted into binary indicators [73]. This method also allows for the use of qualitative variables [73; 48] and it does not require the assumption of normality, which is common in Kriging [48].
- Disjunctive Kriging - A non-linear but parametric form of Kriging. Best used in decision making as estimates can be made for the probabilities of exceeding specific pre-defined values [73].

This review will only cover the two linear forms of Kriging required for the purposes of this study, namely Simple Kriging and Regression Kriging. The basics required for Simple and

Regression Kriging will be discussed below and then the specific cases will be discussed further in the sections that follow.

3.1.2 Basics for Simple and Regression Kriging

Suppose the variable of interest can be thought of as a random process,

$$\{Z(\mathbf{s}) : \mathbf{s}\,\epsilon\, D_{\Re}\};\ D_{\Re} \subset \Re^d\,, \tag{3.1}$$

and has been sampled at locations $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, ..., \mathbf{s}_n$ [13]. In addition, suppose that Z can be represented in the form:

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \delta_K(\mathbf{s});\ \mathbf{s}\,\epsilon\, D_{\Re}\,, \tag{3.2}$$

where $\delta_K(\mathbf{s})$ is a stationary Gaussian process with a zero mean [26]. Also assume that the covariance function of $\delta_K(\mathbf{s})$ is known and stationary [7].

$\mu(\mathbf{s})$ is a deterministic function which is either known or needs to be estimated. If this trend is known (Simple Kriging) or is estimated outside of the process of Kriging (Regression Kriging), it can be subtracted from the data and the residuals can undergo Kriging under the assumption of a zero mean. In this case, the form of the Kriging estimator is then [7]:

$$Z^*(\mathbf{s}) = \mu(\mathbf{s}) + \sum_{i=1}^{n} \lambda_i (Z(\mathbf{s}_i) - \mu(\mathbf{s}_i)). \tag{3.3}$$

It is necessary to determine the weights λ_i under the constraint of unbiasedness [7]

$$E[Z^*(\mathbf{s}) - Z(\mathbf{s})] = 0\,, \tag{3.4}$$

and which minimise the error variance given by [7]:

$$\sigma_E^2(\mathbf{s}) = \text{Variance}[Z^*(\mathbf{s}) - Z(\mathbf{s})]. \tag{3.5}$$

From the earlier discussion of $\delta_K(\mathbf{s})$, the following is known about the mean of $\delta_K(\mathbf{s})$ [7],

$$E[\delta_K(\mathbf{s})] = 0\,, \tag{3.6}$$

and based on the stationarity of the covariance of $\delta_K(\mathbf{s})$,

$$\text{Covariance}(\delta_K(\mathbf{s}), \delta_K(\mathbf{s}+\mathbf{h})) = E[\delta_K(\mathbf{s}) \cdot \delta_K(\mathbf{s}+\mathbf{h})] = C_\delta(\mathbf{h}). \tag{3.7}$$

The type of Kriging which needs to be applied to a problem is determined by the form of $\mu(\mathbf{s})$ [7].

3.1.3 Simple Kriging

In the case of Simple Kriging, it is assumed that $\mu(\mathbf{s}) = \mu$, a constant and known mean. Given the known mean, the Simple Kriging Estimator then becomes [7]:

$$Z^*_{SK}(\mathbf{s}) = \mu + \sum_{i=1}^{n} \lambda_i^{SK}(Z(\mathbf{s}_i) - \mu) = \mu + \delta^*_{SK}(\mathbf{s}). \tag{3.8}$$

This is equivalent to subtracting the mean from the data, applying Simple Kriging with a zero mean on the residuals and adding back the mean [30]. This estimator is automatically unbiased as $E[(Z(\mathbf{s}_i) - \mu)] = 0$ [7]:

$$\begin{aligned} E[Z^*(\mathbf{s}) - Z(\mathbf{s})] &= E[Z^*(\mathbf{s})] - E[Z(\mathbf{s})] \\ &= E[\mu + \sum_{i=1}^{n} \lambda_i^{SK}(Z(\mathbf{s}_i) - \mu)] - E[Z(\mathbf{s})] \\ &= \mu + E[\sum_{i=1}^{n} \lambda_i^{SK}(Z(\mathbf{s}_i) - \mu)] - \mu \\ &= 0 + \sum_{i=1}^{n} E[\lambda_i^{SK}(Z(\mathbf{s}_i) - \mu)] \\ &= \sum_{i=1}^{n} \lambda_i^{SK} E[(Z(\mathbf{s}_i) - \mu)] \\ &= 0. \end{aligned} \tag{3.9}$$

It is now necessary to minimise the error variance [7]:

$$\begin{aligned} \sigma^2_E(\mathbf{s}) &= \text{Variance}[Z^*(\mathbf{s}) - Z(\mathbf{s})] \\ &= \text{Variance}[\delta^*_{SK}(\mathbf{s})] + \text{Variance}[\delta(\mathbf{s})] - 2\text{Covariance}[\delta^*_{SK}(\mathbf{s}), \delta(\mathbf{s})] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i^{SK}(\mathbf{s})\lambda_j^{SK}(\mathbf{s}) C_\delta(\mathbf{s}_i - \mathbf{s}_j) + C_\delta(\mathbf{0}) - 2\sum_{i=1}^{n} \lambda_i^{SK}(\mathbf{s}) C_\delta(\mathbf{s}_i - \mathbf{s}). \end{aligned} \tag{3.10}$$

This minimisation is achieved by taking the derivative of the above expression with respect to each weight λ_i and setting the resulting expression equal to zero. This results in the following system of equations [7]:

$$\sum_{j=1}^{n} \lambda_j^{SK} C_\delta(\mathbf{s}_i - \mathbf{s}_j) = C_\delta(\mathbf{s}_i - \mathbf{s}) \quad i = 1, 2, 3, ..., n. \tag{3.11}$$

However, as the mean is constant, the covariance for Z is the same as that of the residual δ and, thus, the above system of equations can be rewritten in terms of the covariance of Z [7].

$$\sum_{j=1}^{n} \lambda_j^{SK} C(\mathbf{s}_i - \mathbf{s}_j) = C(\mathbf{s}_i - \mathbf{s}) \quad i = 1, 2, 3, ..., n. \tag{3.12}$$

This can be rewritten into matrix notation as:

$$\mathbf{K}\boldsymbol{\lambda}^{SK} = \mathbf{k}\,, \tag{3.13}$$

where $K_{i,j} = C(\mathbf{s}_i - \mathbf{s}_j)$, $k_i = C(\mathbf{s}_i - \mathbf{s})$ and $\boldsymbol{\lambda}^{SK}$ is the vector of weights [7]. If no two points are co-located then $\mathbf{K}$ is positive-definite and $\mathbf{K}^{-1}$ can be found [7]. In this case, the weights can be found as:

$$\boldsymbol{\lambda}^{SK} = \mathbf{K}^{-1}\mathbf{k}. \tag{3.14}$$

By substitution into the error variance Equation (3.10), the estimated error variance can be found to be [7]:

$$\sigma_E^2(\mathbf{s}) = \sigma_{SK}^2(\mathbf{s}) = C(0) - \boldsymbol{\lambda}^{SK}(\mathbf{s})\mathbf{k} = C(\mathbf{0}) - \sum_{i=1}^{n} \lambda_i^{SK}(\mathbf{s})C(\mathbf{s}_i - \mathbf{s}). \tag{3.15}$$

3.1.4 Regression Kriging

Regression Kriging is known as a hybrid technique, since it combines both interpolation based on regression and spatial interpolation [31]. It has been shown that hybrid methods provide better results than either of the single approaches [31].

In the case of Regression Kriging, we assume that $\mu(s)$ is a deterministic trend which needs to be estimated.

Unlike in Universal Kriging, the trend is estimated before Kriging takes place. The deterministic trend is then removed and Simple Kriging with a mean of zero is used to fit the residuals [31]. In Universal Kriging, the covariance matrix is extended to allow for the simultaneous estimation of trend and spatial interpolation. It has been shown though, that in at least the case of linear trends, that the two methods are equivalent for both the estimate at a point and the estimated error variance at a point given the same inputs [30].

When the trend is a linear polynomial, Regression Kriging can be represented as [31],

$$Z_{RK}^*(\mathbf{s}) = m^*(\mathbf{s}) + \delta_{RK}^*(\mathbf{s}) = \sum_{k=0}^{p} \beta_k^* q_k(\mathbf{s}) + \sum_{i=1}^{n} \lambda_i^{RK}\delta_{RK}(\mathbf{s}_i)\,, \tag{3.16}$$

where $m^*(\mathbf{s})$ is the fitted trend, $\delta_{RK}^*(\mathbf{s})$ is the interpolated residual, β_k^* are the trend model coefficients and $q_k(\mathbf{s})$ are the values of the additional variables (perhaps spatial coordinates) at the target location [31].

In addition, the estimated error variance now must account for both the error in the trend estimation and in the spatial interpolation. This error variance is given by [32]:

$$\begin{aligned} \sigma_E^2(\mathbf{s}) &= \sigma_{RK}^2(\mathbf{s}) \\ &= \sigma^{2*}[m^*(\mathbf{s})] + \sigma^{2*}[\delta_{RK}^*(\mathbf{s})] \\ &= (\mathbf{q}_0 - \mathbf{q}^{\mathbf{T}}\mathbf{C}^{-1}\mathbf{c}_0)^T(\mathbf{q}^T\mathbf{C}^{-1}\mathbf{q})^{-1}(\mathbf{q}_0 - \mathbf{q}^{\mathbf{T}}\mathbf{C}^{-1}\mathbf{c}_0) + C(\mathbf{0}) - \mathbf{c}_0^T\mathbf{C}^{-\mathbf{1}}\mathbf{c}_0\,, \end{aligned} \tag{3.17}$$

where $\mathbf{q}_0$ is the vector of additional variables at the target point, $\mathbf{q}$ is the matrix of additional variables at the sample points, $\mathbf{C}$ is the covariance matrix of the residuals, $\mathbf{c}_0$ is the vector of covariances at between the target location and the sample points and $C(\mathbf{0})$ is the value of the covariance for a zero distance [31].

The issue in Regression Kriging becomes determining the trend structure and estimating the trend coefficients. The trend structure is often taken to be linear and often a polynomial as it is shown in Equation (3.17) [55]. However, arbitrarily complex regression models can be used [31], which may be parametric or non-parametric [55]. In Section 3.1.5 to follow, the estimation of the trend coefficients is discussed under the assumption of a polynomial trend.

3.1.5 Estimating the Trend Coefficients for Regression Kriging

Due to the spatial correlation in the data, Ordinary Least Squares is no longer an optimal technique for the estimation of the trend coefficients [30; 32; 31]. Hengl *et al.* [32] considers the use of Generalised Least Squares. Pardo-Igúzquiza and Dowd [55] acknowledge that the equation for the trend coefficients is identical in both Generalised Least Squares and Maximum Likelihood Estimation but suggest that Maximum Likelihood Estimation is more sensitive to changes in the parameters of the covariance. Both Generalised Least Squares and Maximum Likelihood Estimation are explained below.

Generalised Least Squares

Generalised Least Squares accounts for the correlation between the residuals. The estimator for the coefficients is given by:

$$\boldsymbol{\beta}^*_{GLS} = (\mathbf{q}^T \mathbf{C}_\delta \mathbf{q})^{-1} \mathbf{q}^T \mathbf{C}_\delta^{-1} \mathbf{Z}\,, \tag{3.18}$$

where all terms are defined as they were for Regression Kriging and in addition $\mathbf{Z}$ is the vector of values of interest at the sample points [31]. In order to determine these coefficients, it is first necessary to determine the Ordinary Least Squares coefficients $\boldsymbol{\beta}^*_{OLS} = (\mathbf{q}^T\mathbf{q})^{-1}\mathbf{q}^T\mathbf{Z}$. These coefficients are then used as the trend, and the covariance of the residuals is estimated. This estimated covariance is then used to determine the Generalised Least Squares coefficients which are in turn used as the trend in order to re-estimate the covariance of the residuals. This is used to update the Generalised Least Squares coefficients which are then used to update the covariance of the residuals and so forth until it is determined that sufficiently correct coefficients have been reached [31].

According to Hengl *et al.* [31] it is worth noting that Kitandis in 1994 showed that the covariance function derived from the Ordinary Least Squares coefficients is often satisfactory.

Maximum Likelihood Estimation

Under Maximum Likelihood Estimation, Z is assumed to follow a multivariate Gaussian distribution [49]. This assumption leads to the following log-likelihood function of Z with the parameters $(\boldsymbol{\beta}, \boldsymbol{\theta})$, where $\boldsymbol{\beta}$ are the trend coefficients, $\boldsymbol{\theta}$ are the covariance parameters and n is the number of sample observations [49]:

$$l = l(\mathbf{Z};\boldsymbol{\beta};\boldsymbol{\theta}) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{C}(\boldsymbol{\theta})| - \frac{1}{2}(\mathbf{Z}-\mathbf{q}\boldsymbol{\beta})^T(\mathbf{C}(\boldsymbol{\theta}))^{-1}(\mathbf{Z}-\mathbf{q}\boldsymbol{\beta}). \tag{3.19}$$

The notation ln is is used for the natural logarithm. The covariance parameters will be further discussed in Section 3.2, as these parameters are dependent on the covariance structure chosen. Equation (3.19) leads to the following Maximum Likelihood Estimate for $\boldsymbol{\beta}$:

$$\boldsymbol{\beta}^* = (\mathbf{q}^T\mathbf{C}^{*-1}\mathbf{q})^{-1}\mathbf{q}^T\mathbf{C}^{*-1}\mathbf{Z}\,, \tag{3.20}$$

where $\mathbf{C}^* = \mathbf{C}(\boldsymbol{\theta}^*)$ and $\boldsymbol{\theta}^*$ is the Maximum Likelihood Estimate of the covariance parameters [49]. If further, the covariance matrix can be represented as the product of a constant variance (σ^2) and a correlation matrix, that is $\mathbf{C}(\boldsymbol{\theta}) = \mathbf{C}(\sigma^2,\boldsymbol{\theta}_2) = \sigma^2\mathbf{R}(\boldsymbol{\theta}_2)$, then the Maximum Likelihood Estimates of $\boldsymbol{\beta}$ and σ^2 become [49]:

$$\boldsymbol{\beta}^* = (\mathbf{q}^T\mathbf{R}^{*-1}\mathbf{q})^{-1}\mathbf{q}^T\mathbf{R}^{*-1}\mathbf{Z} \tag{3.21}$$

and

$$\sigma^{2*} = \frac{1}{n}[(\mathbf{Z}-\mathbf{q}\boldsymbol{\beta}^*)^T\mathbf{R}^{*-1}(\mathbf{Z}-\mathbf{q}\boldsymbol{\beta}^*)]\,, \tag{3.22}$$

where $\mathbf{R}^*$ is the correlation matrix as produced using the Maximum Likelihood Estimates for $\boldsymbol{\theta}_2$. Based on the equation for the log-likelihood (Equation (3.19)) and $\mathbf{C}(\boldsymbol{\theta}) = \sigma^2\mathbf{R}(\boldsymbol{\theta}_2)$, it can be seen that [19]:

$$l(\boldsymbol{\beta};\sigma^2;\boldsymbol{\theta}_2) \propto -\frac{1}{2}[\ln(|\sigma^2\mathbf{R}|) + (\mathbf{Z}-\mathbf{q}\boldsymbol{\beta})^T(\sigma^2\mathbf{R})^{-1}(\mathbf{Z}-\mathbf{q}\boldsymbol{\beta})]. \tag{3.23}$$

By substituting in the Maximum Likelihood Estimates for $\boldsymbol{\beta}$ and σ^2 found in Equations (3.21) and (3.22), the reduced log-likelihood is obtained as [19]

$$l(\boldsymbol{\theta}_2) \propto -\frac{1}{2}[n\ln(|\sigma^{2*}(\mathbf{R})|) + \ln|\mathbf{R}|]. \tag{3.24}$$

Thus, it is only necessary to maximise Equation (3.24) and back substitute to obtain $\boldsymbol{\beta}$ and σ^2. This optimisation procedure can be started by an initial guess for $\boldsymbol{\theta}_2$ [19]. It is important to note that in order to optimise the reduced log-likelihood, it is necessary to recalculate $\boldsymbol{\beta}$ and σ^2 at every step.

Haarhoff *et al.* [27] further discuss numerical strategies for solving these equations especially in the presence of ill-conditioned matrices. Haarhoff *et al.* [27] make the suggestion of using Cholesky decomposition to factor $\mathbf{R}$ such that it can be represented by $\mathbf{R} = \mathbf{L}\mathbf{L}^T$. The following representations can then be used [27]:

$$\boldsymbol{\beta}^* = (\mathbf{q}^T\mathbf{D})^{-1}\mathbf{q}^T\mathbf{U} \tag{3.25}$$

$$\mathbf{R}^{-1}\mathbf{q} = \mathbf{D} \tag{3.26}$$

$$\mathbf{R}^{-1}\mathbf{Z} = \mathbf{U} \tag{3.27}$$

D and **U** can be solved for from [27]:

$$\mathbf{RD} = \mathbf{q}\,, \tag{3.28}$$

$$\mathbf{RU} = \mathbf{Z}. \tag{3.29}$$

This also allows σ^{*2} to be represented as [27]:

$$\sigma^{2*} = \frac{(\mathbf{Z} - \mathbf{q}\boldsymbol{\beta}^*)^T(\mathbf{U} - \mathbf{D}\boldsymbol{\beta})}{n}. \tag{3.30}$$

This method also allows for a simplification in the calculation of $\ln|\mathbf{R}|$ as it is shown by Haarhoff *et al.* [27] that it can be calculated as:

$$\ln|\mathbf{R}| = 2\left(\sum_{i=1}^{n}\ln(\mathbf{L}_{ii})\right). \tag{3.31}$$

Haarhoff *et al.* [27] found that this method offered significant improvements in both computational speed and accuracy.

The benefit of Maximum Likelihood is that the trend coefficients and covariance parameters can be estimated in the same process, while Generalised Least Squares swaps between updating the estimation of the covariance parameters and the trend coefficients using different methods. However, as mentioned previously, it has been suggested that that Ordinary Least Squares may be sufficient.

3.2 Covariance Functions and Their Parameters

The covariance between the variable at two locations, $Z(\mathbf{s}_1)$ and $Z(\mathbf{s}_2)$ with respective means of $\mu(\mathbf{s}_1)$ and $\mu(\mathbf{s}_2)$, is defined as [67]:

$$\text{Covariance}[Z(\mathbf{s}_1), Z(\mathbf{s}_2)] = \text{Cov}[Z(\mathbf{s}_1), Z(\mathbf{s}_2)] = E[(Z(\mathbf{s}_1) - \mu(\mathbf{s}_2))(Z(\mathbf{s}_1) - \mu(\mathbf{s}_2))]. \tag{3.32}$$

However, if second order stationarity is assumed, that is [67]

$$E[Z(\mathbf{s}_1)] = \mu \tag{3.33}$$

$$\text{Cov}[Z(\mathbf{s}_1), Z(\mathbf{s}_2)] = \text{Cov}[Z(\mathbf{s}_1 + \mathbf{h}), Z(\mathbf{s}_2 + \mathbf{h})]\;\forall\,\mathbf{h}\,, \tag{3.34}$$

then we can use the following notation:

$$\text{Cov}[Z(\mathbf{s}_1), Z(\mathbf{s}_2)] = \text{Cov}[Z(\mathbf{0}), Z(\mathbf{s}_2 - \mathbf{s}_1)] =: C(\mathbf{s}_2 - \mathbf{s}_1). \tag{3.35}$$

This implies that the covariance between two locations relies only on the distance between the locations and not directly on the locations themselves [67].

In addition, a semivariance(γ) can be defined as [15]:

$$\gamma(\mathbf{s}_1 - \mathbf{s}_2) = \frac{1}{2}\text{Variance}[Z(\mathbf{s}_1) - Z(\mathbf{s}_2)]. \tag{3.36}$$

If the variable is assumed second order stationary, then the semivariance and covariance can be related by [15]

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}). \tag{3.37}$$

It is important to note that it is possible to have a valid semi-variance which does not possess a valid covariance [15].

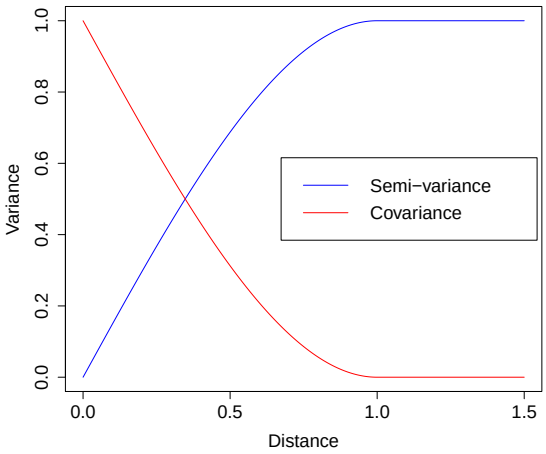


Figure 3.1: A typical covariance and its corresponding semi-variance (generated in R using the gstat package [61]).

Figure 3.1 shows both a typical covariance and its corresponding semi-variance.

In the section to follow some characteristics of covariance functions are discussed as well as some specific forms and their parameters.

3.2.1 Covariance Functions

Covariance functions can display certain characteristics. Some are based on the assumption of stationarity, some are required in order for the function to define a valid covariance function while others may or may not occur based upon the choice of covariance function.

- Symmetry - Based on the assumption of stationarity, it can be shown that the covariance function will exhibit symmetry [73]:

$$\begin{aligned} C(h) &= E[(Z(v) - \mu)(Z(v+h) - \mu)] \\ &= E[Z(v-h)Z(v) - \mu^2] \\ &= E[Z(v)Z(v-h) - \mu^2] \\ &= C(-h). \end{aligned} \tag{3.38}$$

  This allows us to consider only the lags $h \geq 0$.

- Positive Semi-definiteness - In order for a covariance function to be valid, the covariance matrix must be positive semi-definite. That is, the principal minors of the matrix must all be either positive or zero [73], where the covariance matrix is given by:

$$\mathbf{C} = \begin{bmatrix} C(\mathbf{s}_1, \mathbf{s}_1) & C(\mathbf{s}_1, \mathbf{s}_2) & \cdots & C(\mathbf{s}_1, \mathbf{s}_n) \\ C(\mathbf{s}_2, \mathbf{s}_1) & C(\mathbf{s}_2, \mathbf{s}_2) & \cdots & C(\mathbf{s}_2, \mathbf{s}_n) \\ \vdots & \vdots & \cdots & \vdots \\ C(\mathbf{s}_n, \mathbf{s}_1) & C(\mathbf{s}_n, \mathbf{s}_2) & \cdots & C(\mathbf{s}_n, \mathbf{s}_n) \end{bmatrix}. \tag{3.39}$$

- Continuity - Due to most environmental variables being continuous, continuous covariance and semi-variance are required. The covariance should decline from a positive value $C(\mathbf{0}) = \sigma^2$ to smaller values while the semi-variance should climb from $\gamma(\mathbf{0}) = 0$. If one, however, considers the sample semi-variance otherwise known as the sample variogram, it is often the case that $\gamma(\mathbf{0}) > 0$. This is referred to as the nugget effect [73]. According to Cressie and Wikle [14] this nugget effect is composed of two non-negative components, namely the micro-scale variance of the variable without measurement error and the measurement error variance of the variable. However, in geostatistical literature it is commonly assumed that there is no measurement error, that is the that the nugget is only composed of micro-scale variance[14]. The nugget effect is not considered further in this study.

- Monotonically Decreasing - As the lag distance $\mathbf{h}$ increases, $C(\mathbf{h})$ decreases. That is, as the distance between points increases, the points become more dissimilar on average [73]. Conversely, the semi-variance is monotonically increasing [73]. This can be seen in Figure 3.1.

- Sill and Range - Certain semi-variances reach upper bounds at which they remain after a certain distance. This upper bound, if it exists, is known as the sill variance [73]. This occurs if the process is second order stationary [73]. This sill variance is also $C(\mathbf{0})$ [73]. If the sill variance is reached within a finite lag distance, then the lag distance at which the sill is first reached is known as the range or correlation range [73]. This range is also the lag distance at which the covariance reaches zero [73]. Some semi-variances, however, reach their sill in an asymptotic manner and their ranges are then considered effective ranges, defined as the lag distance where they reach 95% of their sill variance [73].

- Unboundedness - If the process is not second order stationary, then the semi-variance increases with increasing lag distances continually and never reaches an upper bound. In this case, the semi-variance is said to be unbounded. For unbounded semi-variances, valid covariances do not exist [73].

- Anisotropy - If the semi-variance can written as a function of

$$\|\mathbf{h}\| \equiv \sqrt{(h_1^2 + h_2^2 + \cdots + h_d^2)} \quad \text{for } \mathbf{h} = (h_1, h_2, \cdots, h_d)\, \epsilon\, \Re^d\,, \tag{3.40}$$

  then the process is isotropic [14]. However, spatial variation is not always the same in all directions. If the process is anisotropic, then the semi-variance will be anisotropic and the covariance, if it exists, will also be anisotropic [73]. Two types of anisotropy are considered below:

  - Geometric Anisotropy - In the case of geometric anisotropy, the initial gradients vary for different directions. If a sill exists (which is the same in all directions), then the ranges or effective ranges will vary for different directions. In this case, a simple transformation of the spatial coordinates will remove the anisotropy [73].

  - Zonal Anisotropy - In the case of zonal anisotropy, a region contains zones with different mean values. In this case, the variance differs with direction and, thus, the sill variance varies with direction [73].

- Separability - In purely spatial coordinates, $\mathbf{h} = (h_1, h_2, \cdots, h_d)\, \epsilon\, \Re^d$, a covariance is said to be separable when it can be represented by [15]:

$$C(\mathbf{h}) = \prod_{i=1}^{d} C_i(h_i). \tag{3.41}$$

  In terms of spatio-temporal coordinates, a process Z is said to have a separable spatio-temporal covariance function if, for all $\mathbf{s}, \mathbf{x}\, \epsilon\, \Re^d; t, r\, \epsilon\, \Re$, the covariance function can be represented as [14]:

$$\text{Cov}(Z(\mathbf{s}, t), Z(\mathbf{x}, r)) = C^{(s)}(\mathbf{s}, \mathbf{x}) C^t(t, r). \tag{3.42}$$

  It is important to note that non-separable covariance structures have been developed more recently, specifically for spatio-temporal data [69]. However, they are not considered in this study.

Some popular covariance functions for use with Kriging are given below in an one-dimensional form for ease of representation. It is noted that the Kriging equations given earlier can be rewritten for use with semi-variances. All the functions and their figures, except for the nugget function, are given for a zero nugget. All of the figures, bar the figure for the Matérn and nugget functions, are scaled so that the range or effective range occurs at one. The nugget function has no range parameter and thus it cannot be scaled to be one, while the Matérn function is difficult to scale due the various parameters which the effective range is dependent on. Furthermore, all figures are shown for a sill of one.

Nugget

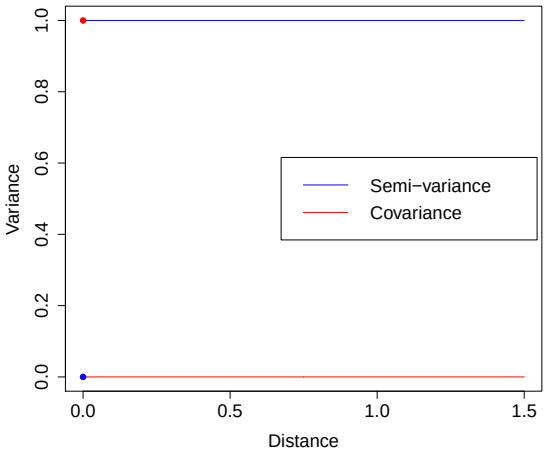


Figure 3.2: Nugget covariance and semi-variance with $c_0 = 1$ (generated in R using the gstat package [61]).

With a nugget function, there is no spatial correlation between points. The semi-variance is given by [73]:

$$\gamma(h) = c_0(1 - \delta(h)) \,, \tag{3.43}$$

where c_0 is the sill variance and $\delta(h)$ is the Kronecker delta function which takes the value of one when $h = 0$ and is zero everywhere else [73].

The covariance is given by:

$$C(h) = c_0\delta(h) \,, \tag{3.44}$$

where c_0 and $\delta(h)$ are defined as above. The nugget covariance and semi-variance with $c_0 = 1$ can be seen in Figure 3.2.

Spherical

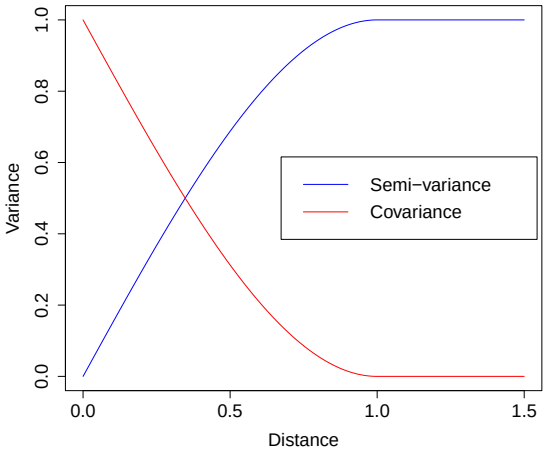


Figure 3.3: Spherical covariance and semi-variance with $c_0 = 1$ and $a = 1$ (generated in R using the gstat package [61]).

The semi-variance for the spherical function is given by:

$$\gamma(h) = \begin{cases} c_0 \left\{ \dfrac{3h}{2a} - \dfrac{1}{2}\left(\dfrac{h}{a}\right)^3 \right\} & \text{for } h \leq a \\ c_0 & \text{for } h \geq a \,, \end{cases} \tag{3.45}$$

where c_0 is the sill variance and a is the range parameter [73].

The covariance is given by:

$$C(h) = \begin{cases} c_0 \left\{ 1 - \dfrac{3h}{2a} + \dfrac{1}{2}\left(\dfrac{h}{a}\right)^3 \right\} & \text{for } h \leq a \\ 0 & \text{for } h \geq a. \end{cases} \tag{3.46}$$

The spherical covariance and semi-variance with $c_0 = 1$ and $a = 1$ can be seen in Figure 3.3. According to Todini [71] the spherical function can lead to problems within the Maximum

Likelihood Estimation framework as it does not have guaranteed continuity of the function as well as its first and second derivatives and it is recommended that the pentaspherical function be used instead.

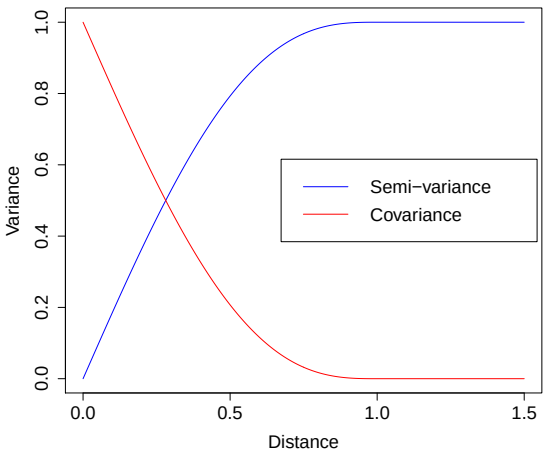


Figure 3.4: Pentaspherical covariance and semi-variance with $c_0 = 1$ and $a = 1$ (generated in R using the gstat package [61]).

The pentaspherical semi-variance is given by [73]:

$$\gamma(h) = \begin{cases} c_0 \left\{ \frac{15h}{8a} - \frac{5}{4}\left(\frac{h}{a}\right)^3 + \frac{3}{8}\left(\frac{h}{a}\right)^5 \right\} & \text{for } h \leq a \\ c_0 & \text{for } h \geq a. \end{cases} \tag{3.47}$$

The covariance is given by:

$$C(h) = \begin{cases} c_0 \left\{ 1 - \frac{15h}{8a} - \frac{5}{4}\left(\frac{h}{a}\right)^3 + \frac{3}{8}\left(\frac{h}{a}\right)^5 \right\} & \text{for } h \leq a \\ c_0 & \text{for } h \geq a. \end{cases} \tag{3.48}$$

The pentaspherical covariance and semi-variance with $c_0 = 1$ and $a = 1$ can be seen in Figure 3.4.

Exponential

The exponential model is one which reaches its sill variance asymptotically and its semi-variance is given by:

$$\gamma(h) = c_0 \left\{ 1 - e^{-h/a} \right\}, \tag{3.49}$$

where c_0 is the sill variance and a is a distance parameter. The exponential model has an effective range of approximately $3a$ [73]. The exponential covariance function is given by:

$$C(h) = c_0 \left\{ e^{-h/a} \right\}. \tag{3.50}$$

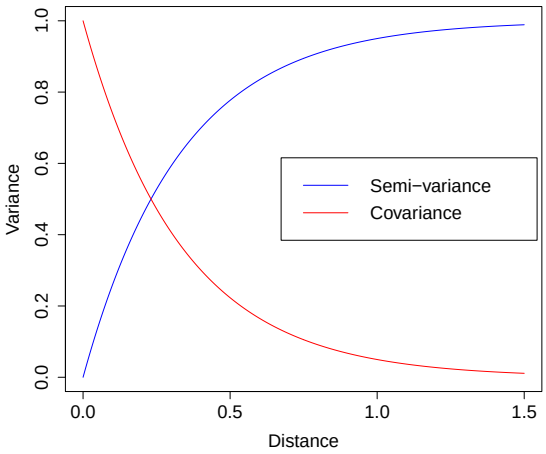


Figure 3.5: Exponential covariance and semi-variance with $c_0 = 1$ and $a = \frac{1}{3}$ (generated in R using the gstat package [61]).

The exponential covariance and semi-variance with $c_0 = 1$ and $a = \frac{1}{3}$ can be seen in Figure 3.5.

Powered Exponential

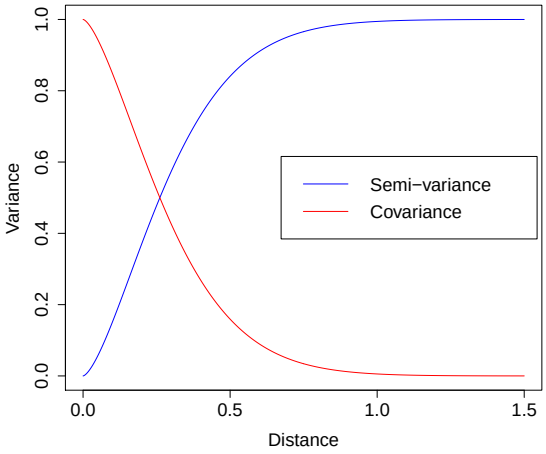


Figure 3.6: Powered exponential covariance and semi-variance with $c_0 = 1$, $r = 1.5$ and $a = 3^1.5$ (generated in R using the gstat package [61]).

For the powered exponential, the semi-variance is given by:

$$\gamma(h) = c_0 \left\{ 1 - e^{(-(h/a)^r)} \right\} , \tag{3.51}$$

where c_0 is the sill variance, a is a distance parameter and r is the power parameter. r is bounded between 0 and 2 ($0 \leq r \leq 2$) [14]. The effective range of the powered exponential model is dependent on r [19]. The exponential semi-variance function is the special case where $r = 1$. The associated covariance function is:

$$C(h) = c_0 \left\{ e^{(-(h/a)^r)} \right\}. \tag{3.52}$$

The powered exponential covariance and semi-variance with $c_0 = 1$, $r = 1.5$ and $a = \frac{1}{3^{1/1.5}}$ can be seen in Figure 3.6.

Gaussian

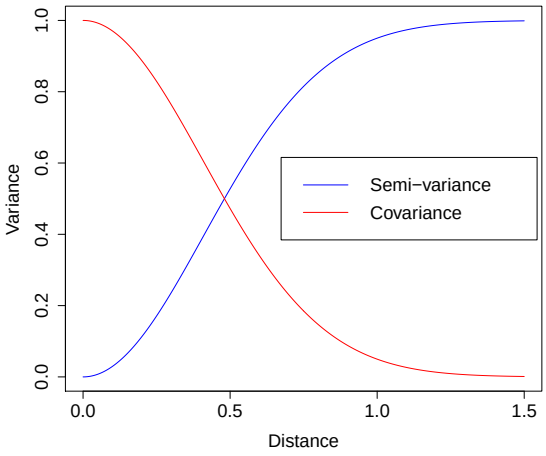


Figure 3.7: Gaussian covariance and semi-variance with $c_0 = 1$ and $a = \frac{1}{\sqrt{3}}$ (generated in R using the gstat package [61]).

The semi-variance for the Gaussian model is given by:

$$\gamma(h) = c_0 \left\{ 1 - e^{\left(-h^2/a^2\right)} \right\}, \tag{3.53}$$

where c_0 is the sill variance and a is a distance parameter. The Gaussian model has an effective range of approximately $\sqrt{3}a$ [73]. The covariance is given by:

$$C(h) = c_0 \left\{ e^{\left(-h^2/a^2\right)} \right\}. \tag{3.54}$$

The Gaussian covariance and semi-variance with $c_0 = 1$ and $a = \frac{1}{\sqrt{3}}$ can be seen in Figure 3.7. The Gaussian model's semi-variance does however approach the origin with a zero gradient and this can lead to unstable Kriging equations. As a result, it is common practice to deprecate the model, that is to replace the exponent of 2 with a slightly different exponent [73]. This yields a semi-variance of:

$$\gamma(h) = c_0 \left\{ 1 - e^{\left(-h^\alpha/a^\alpha\right)} \right\}, \tag{3.55}$$

where the additional parameter α must be such that $\alpha < 2$ [73].

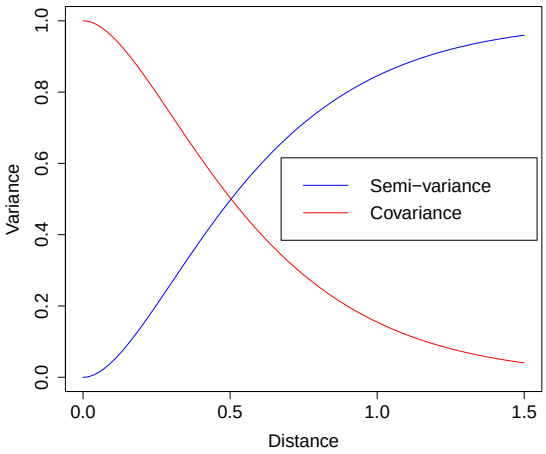


Figure 3.8: Matérn covariance and semi-variance with $c_0 = 1$, $a = 0.3$ and $\nu = 1.5$ (generated in R using the gstat package [61]).

Matérn

The Matérn model is a generalization of several of the earlier mentioned models. Its semi-variance is given by:

$$\gamma(h) = c_0 \left\{ 1 - \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{h}{a}\right)^{\nu} K_\nu \left(\frac{h}{a}\right) \right\} , \tag{3.56}$$

where c_0 is the sill variance, a is a distance parameter, ν is a smoothness parameter and K_ν is a modified Bessel function of order ν [73]. The covariance is given by:

$$\gamma(h) = c_0 \left\{ \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{h}{a}\right)^{\nu} K_\nu \left(\frac{h}{a}\right) \right\} . \tag{3.57}$$

The Matérn covariance and semi-variance with $c_0 = 1$, $a = 0.3$ and $\nu = 1.5$ can be seen in Figure 3.8.

3.2.2 Estimating the Covariance Parameters

There are various ways in which to estimate the covariance parameters. Some involve the semi-variance and others the covariance. Some of these methods are listed below:

- Maximum Likelihood Estimation [15]
- Restricted Maximum Likelihood Estimation [15]
- Method of Moments Estimation [40]
- Least Squares Fitting [15]
- Generalised Least Squares Fitting [15]
- Cross Validation [50]

Martin and Simpson [50] concluded that Maximum Likelihood Estimation performs better than Cross Validation even when the process is non-Gaussian, while Lark [40] found both advantages and disadvantages to using Maximum Likelihood Estimation in comparison to Method of Moments Estimation. The Least Squares and Generalised Least Squares Fitting work with the semi-variance function [15]. Restricted Maximum Likelihood requires the population of an additional matrix which allows the method to apply Maximum Likelihood Estimation to error contrasts in place of the data [15]. Additionally, the method of Maximum Likelihood Estimation allows us to estimate both the trend parameters and covariance parameters, as was explained in Section 3.1.5.

The ability of Maximum Likelihood Estimation to estimate both the trend coefficients and the covariance parameters, as well as its ability to perform more favourably than cross-validation even when the data is non-normal, leads to the use of Maximum Likelihood Estimation in this study.

CHAPTER 4

MATHEMATICAL OVERVIEW: SAMPLING METHODS

We aim to determine the best sample of a specific size in order to attain knowledge about the field variable, as a whole, as accurately as possible. In doing so we wish to find the n locations chosen out of a total of m locations which minimise a function $g(\mathbf{S})$ [75]. There are approximately $\binom{m}{n} = \frac{m!}{n!\,(m-n)!}$ different sample combinations. Thus, using an exhaustive search algorithm, it would be infeasible for large scale sampling [75]. Due to the infeasibility of an exhaustive search, other sampling techniques must be considered.

4.1 Random Sampling

It is possible to perform Random Sampling. However, it is known to be rather inconsistent, sometimes yielding sets which produce very good interpolation results and sometimes yielding sets which produce poor interpolation results. This method effectively consists of randomly generating numbers in the range of the available number of points and then selecting the locations associated with those numbers as the sample set. Although this method could be considered both unreliable and unpredictable, it is easy and very quick to implement [68]. It is however, advisable to also consider some more predictable and reliable methods for selecting sample sets. For this study, it is considered as a comparator for the other methods.

4.2 Genetic Algorithms

Genetic algorithms were first developed in the 1960s by John Holland. However, they were only popularised by one of his students, David Goldberg, in 1989 [28]. Genetic algorithms are optimisation and search methods which are based on the evolutionary process [74] and natural selection or"survival of the fittest" theory [68]. In "survival of the fittest" theory, the better performing individuals are more likely to attract a mate and to reproduce and thus, produce offspring which perform just as well if not better [68].

Holland suggested that within a genetic pool of possible solutions to a problem, the optimal solution or a better solution exists. This solution may, however, currently have not been found as it is some genetic combination of some of the individual solutions that are currently available [68]. Holland was of the opinion that evolutionary techniques, such as genetic algorithms, could allow the optimal or at least a better solution to be found. Holland made

use of mutation and reproduction in order to alter the possible solutions to obtain a better one [68]. These processes closely mimic the process of evolution in nature.

Genetic Algorithms do not promise the optimal solution but rather an acceptably good solution [68]. They are broad based optimisation techniques [24] but because of this they are not very problem specific, leading to slower computation than problem specific solvers [28].

A basic genetic algorithm consists of the following components [24; 28; 4]

- A genetic representation of possible solutions to the optimisation problem (Sometimes known as encoding)
- A method of generating an initial population from the possible solutions to the optimisation problem
- A method to evaluate the fitness of solutions in order to determine which solutions are more optimal
- Genetic operators which allow changes in genetics due to reproduction and/or mutation
- Choice of values for parameters of the the genetic operators
- Repeated applications of the evolutionary process until convergence is reached

Once the genetic representation of possible solutions, fitness function and parameters of the genetic operators are defined, the initial population must be selected. The initial population should ideally offer a wide diversity and be as large as possible, so that solutions across the whole solution space can be generated [68]. This initial population is often selected randomly [68; 63]. After generation of the initial population, an iterative process takes place to imitate the evolutionary process. This iterative process continues until the convergence criterion is met [68]. Each iteration of a genetic algorithm consists of the following steps [68]:

- Selection
- Evaluation
- Reproduction and Mutation
- Replacement and Elitism

Each of the above steps as well as the convergence criterion and encoding will be considered in more detail in the sections to follow.

4.2.1 Encoding

This is the process of representing a possible solution in its genetic form, i.e. transforming a possible solution into an individual for use in the genetic algorithm. Some of the more common encodings are:

- Binary Encoding - Solutions are represented by binary (bit) strings. Each position in the string is 0 or 1 [24]. Each bit can represent a characteristic of the solution or the string may represent a number as a whole [68].

- Real Number Encoding - Solutions are represented by a string of real numbers, often used in optimisation problems [24].
- Permutation/Integer Encoding - Solutions are represented by a string of integers. This type of encoding is the natural choice for ordering problems [68].
- Value Encoding - Solutions are represented by a string of values relevant to the problem. These values can be numbers and even characters amongst other things. This method of encoding often requires specialised genetic operators [68].

4.2.2 Selection

This step relates to choosing individuals as "parents" for reproduction. The methods are aimed at allowing the individuals with highest fitness to reproduce or undergo crossover more readily than those with a lower fitness, in the hope that the "children" or products of the crossover will have an even higher fitness [68]. The parents are selected and added to a mating pool from which they are randomly selected for breeding [68]. Individuals may be selected more than once in an iteration for reproduction [4]. There are various ways of selecting individuals for reproduction, which include:

- Roulette Wheel Selection - Individuals are selected from the population with a probability which is proportional to fitness. A linear search is conducted through a roulette wheel where each portion of the wheel representing one individual, is weighted based on the fitness of individuals [68]. A random proportion of the sum of all the individual's fitnesses is chosen and the method then advances one individual at a time until that proportion is reached or exceeded [68]. The individual at which it is reached or first exceeded is then selected for reproduction [68].
- Tournament Selection - M individuals are selected randomly from the population. The individual with the highest fitness amongst these M individuals is then selected and added to the mating pool. This process is repeated until there are enough individuals in the mating pool [68]. Due to the tournament process, the mating pool has a higher average fitness than the general population [68].
- Rank Selection - Individuals are ranked according to their fitness. The individual with the lowest fitness receives a rank of 1 and the best performing individual receives a rank of N (where N is the number of individuals in the population) [68]. A tournament style selection then occurs to determine which individuals should be selected as parents [68].
- Random Selection - Individuals are randomly selected from the population to serve as parents [68].

4.2.3 Evaluation

This step comprises of evaluating the fitness function of each individual. The fitness function is determined by the problem and in optimisation it is often some function we want to optimise. However, it may also be a transformed version of the optimisation function [63]. The fitness function determines how well a particular solution performs. The higher the value of the fitness function for an individual, the "fitter" that individual is deemed to be and the more optimal that solution is [68]. The above evaluation of the higher the fitness function, the more optimal the solution is for the case of maximisation. In the case of minimisation, we either consider the negative of the maximisation fitness function as the new fitness function or we regard a lower fitness function value as more optimal.

4.2.4 Reproduction and Mutation

Reproduction and mutation are genetic operators which alter the genetic representation of solutions [68]. They are based on natural processes [68] and are explained in more detail in the sections which follow.

Reproduction

Reproduction is also known as recombination and the most common form is crossover [68]. It is the process whereby a child or children, which are also individual solutions, are created by the mixing of the genes from two selected parents [68]. The most common form of reproduction creates two children from two parents [28]. The steps in basic crossover are [68]:

- Two individuals are randomly chosen from the mating pool for crossover.
- A single cross site or multiple cross sites are randomly selected along the length of an individual's string.
- A portion or portions of the strings are swapped based on the crossover site(s).

While more complex crossover techniques are available, three of the more basic crossover techniques which are commonly used are explained below:

- One Point Crossover - This is the traditional crossover [68]. A single cross site is selected and the portions of the strings after this site are swapped. This leads to the production of two children. One point crossover is illustrated on binary encoded solutions in Figure 4.1

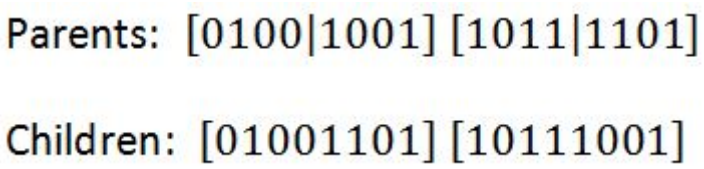


Figure 4.1: One point crossover.

- Two Point Crossover - Two cross points are selected and the portions of the strings between these points is swapped [68]. This is illustrated on binary encoded solutions in Figure 4.2. Two point crossover may negatively influence the performance of the

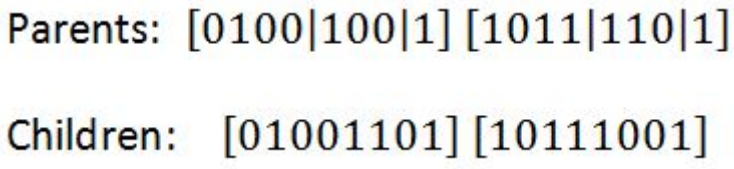


Figure 4.2: Two point crossover.

genetic algorithm in comparison to one point crossover as adding more cross sites can lead to the disruption of building blocks. It is, however, beneficial to have more cross sites, as this allows the solution space to be searched more effectively [68].

- Three Parent Crossover - This crossover technique yields only one child as opposed to the two yielded by one and two point crossover. In this technique three parents are chosen. The bits of the first parent are compared to those of the second parent. If the corresponding bits match, then the child receives those bits in their corresponding positions [68]. However, if they do not match, then the child receives the bit which is in the corresponding position of the third parent. This is illustrated for binary encoded solutions in Figure 4.3.

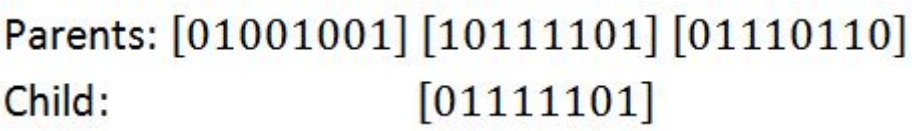


Figure 4.3: Three parent crossover.

Mutation

Mutation helps the genetic algorithm escape local maxima/minima. It also helps ensure that the population remains diverse [68]. In effect, mutation alters a single bit, some bits or all bits of individuals [68]. Mutation occurs with a chosen probability. This probability must not be too high, otherwise the genetic algorithm becomes a random search algorithm [68]. Several types of mutation are available, of which three basic forms of mutation are:

- Flipping - Flipping works in binary encoding of solutions and involves the changing of a randomly selected bit. If the selected bit is 0, it is changed to a 1 and vice versa [68].
- Interchanging - In this technique, two bits from an individual are randomly chosen and the values of these bits are swapped [68].
- Reversing -A bit is selected by random choice and the bits next to it are then reversed [68].

4.2.5 Replacement and Elitism

Replacement and elitism define how children and parents advance to the next generation.

Replacement

After reproduction and mutation, there are both children and parents and it must be decided which of these individuals continue on to become the next population [68]. Clearly not all children and parents can be kept, as the number of individuals to be returned is equal to the number of parents chosen [68].

In general there are two types of replacement which can occur, generational replacement and steady-state replacement [68]. In generational replacement, one generation is replaced at the end of the iteration with either only the children or with some combination of the children and parents [68]. In steady-state replacement, as soon as a child is created it is inserted into the population. This does, however, mean that a current member of the population must be removed [68]. There are various methods in order to determine which individual is removed. Some of these methods call for the removal of the oldest member, the weakest member or

the most similar individual in the population [68]. Another method allows for a form of tournament selection to take place between the child and the selected member [68].

Three specific replacement methods are considered here:

- Random Replacement - Here two individuals from the current population (including the parents) are randomly selected to be replaced by the children [68].
- Weak Parent Replacement - Here only the fittest two individuals out of the set of four (two parents and two children) advance to the next generation [68]. This method can improve the overall fitness of the population but only when it is combined with a selection technique that allows for the selection of both fit and less fit individuals [68]. If only strong individuals are selected for breeding, then this method is never able to replace the worst solutions [68].
- Both Parent Replacement - The two children replace the two parents [68]. Each individual only has the opportunity to reproduce once [68]. This does move around genetic material but it can lead to problems if it is combined with a method of selection that firmly favours the fittest individuals [68].

Elitism

In elitism, the single best or a few top performing individuals are selected to be carried through into the next generation [68]. This allows them to be placed into the new population even when they have been destroyed by mutation and/or crossover [68]. This process has a positive influence on the genetic algorithm's performance [68].

4.2.6 Convergence Criterion

The convergence criterion determines when the genetic algorithm should be terminated [68]. Some of the various options for the convergence criterion are listed below:

- Maximum Generations - The genetic algorithm is terminated when the evolutionary process reached a specified maximum number of generations [68].
- Elapsed Time - The genetic algorithm is terminated after a specified amount of time has elapsed [68].
- No Change in Fitness - Termination occurs when the best fitness remains unchanged for a specified number of generations [68].
- Stall Generations - Termination occurs when the objective function undergoes no improvement for a specified number of generations [68].
- Stall Time Limit - Termination occurs when the objective function undergoes no improvement over a specified time limit [68].

Once the genetic operators to be used and their parameters have been chosen, the genetic algorithm can be implemented. A basic genetic algorithm follows the scheme shown below [68]:

Start: Generate randomly n starting individuals to form the first population

Fitness: Evaluate the fitness of the current population

New Population: Create the next generation/population by performing the following steps:

Selection: Select parents from current generation using the chosen selection method

Crossover: With crossover probability, generate offspring from crossover procedure

Mutation: With mutation probability, mutate new offspring

Acceptance: Place offspring or a portion of them into the new population

Replace: Replace the old population with the new one

Test: Test for convergence. If the criterion is met stop and output the solution, else proceed to the loop step

Loop: Return to evaluation step and continue the process

Genetic algorithms like any other method have both advantages and disadvantages, some of which are:

Advantages

- Based on simple principles [68].
- Genetic algorithms have a very wide applicability as they can be used on any problem for which a fitness function can be generated [68].
- Can be written to run on multiple processors. This ability to run in parallel can greatly reduce computational time [68].
- They can be hybridised with other techniques such as traditional optimisation methods to improve optimisation [74].
- They can be used when there is a multi objective function [68].
- Genetic algorithms can potentially escape local maxima/minima [68].
- No knowledge of derivatives is required [68].
- They are easily modified for different problems [68].
- Capable of handling large and not well defined search spaces [68].
- Capable of solving non-linear problems [74].

Disadvantages

- The fitness function can be hard to identify [68].
- Parameter selection for the genetic operators can be difficult [68].
- Premature convergence may occur [68].
- Gradient information cannot be used [68].
- Requires a large number of fitness evaluations [68].
- Does not always find the exact global optimum [68].

4.3 Updated Kriging Variance Algorithm

Zhu *et al.* [75] make use of the Updated Kriging Variance Algorithm (UKVA) in their paper on the optimisation of large scale ocean sampling. They make use of Ordinary Kriging within the Updated Kriging Variance Algorithm. However, it is easy to adapt the method for both Simple and Universal Kriging.

The estimated error variance gives a quantitative measure of the impact a set of points has on the error in the sampling scheme [75]. Since the estimated error variance is dependent only on the distances between the points and the covariance function, the estimated error variance can be calculated even before samples are measured [75]. For a given covariance function, the total estimated error variance ($g(\mathbf{S})$) is given by:

$$g(\mathbf{S}) = \sum_{i=1}^{m} \sigma^2(Z^*(\mathbf{s}_i|\mathbf{S}))\,, \tag{4.1}$$

where m is the total number of points, $\mathbf{S}$ are the locations of the sampled points, $Z^*(\mathbf{s}_i|\mathbf{S})$ is the Kriging estimate for point $\mathbf{s}_i$ given a set of sampled locations $\mathbf{S}$ and $\sigma^2(Z^*(\mathbf{s}_i|\mathbf{S})$ is the estimated error variance or Kriging variance for the point $\mathbf{s}_i$ [75]. In the case of Ordinary Kriging,

$$\sigma^2(Z^*(\mathbf{s}|\mathbf{S})) = C(\mathbf{s},\mathbf{s}) - \sum_{i=1}^{n} \lambda_i C(\mathbf{s},\mathbf{s}_i) + \varphi\,, \tag{4.2}$$

where φ is the Lagrangian parameter associated with the estimation of the constant mean, λ_i is the ith weight and $C(\mathbf{s},\mathbf{s})$ is equivalent to $C(\mathbf{0})$ discussed in Chapter 3. The estimated error variance can also be represented by [75]:

$$\sigma^2(Z^*(\mathbf{s}|\mathbf{S})) = C(\mathbf{s},\mathbf{s}) - \mathbf{V}_n^T \mathbf{H}_n^{-1} \mathbf{V}_n\,, \tag{4.3}$$

where

$$\mathbf{V}_n = [C(\mathbf{s},\mathbf{s}_1), C(\mathbf{s},\mathbf{s}_2), \cdots, C(\mathbf{s},\mathbf{s}_n), 1]^T$$

and

$$\mathbf{H}_n = \begin{bmatrix} C(\mathbf{s}_1,\mathbf{s}_1) & \cdots & C(\mathbf{s}_1,\mathbf{s}_n) & 1 \\ \vdots & \vdots & \vdots & \vdots \\ C(\mathbf{s}_n,\mathbf{s}_1) & \cdots & C(\mathbf{s}_n,\mathbf{s}_n) & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}.$$

In order to select the best sample, we aim to minimise the total estimated error variance [75]. That is, the objective function is given by :

$$\text{obj}(S^*) = \text{Argmin}[g(\mathbf{S})]. \tag{4.4}$$

Since an exhaustive search is infeasible for large scale sampling, the Greedy Searching Algorithm (GSA) is used [75]. This algorithm often starts from an empty set of locations $\mathbf{S}_0 = \emptyset$ and adds locations until $\mathbf{S}_n = [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_n]$ is reached [75]. The greedy rule used to add each location is given by [75]:

$$\text{obj}(\mathbf{s}_i^*) = \text{Argmax}[g(\mathbf{S}_{i-1}) - g(\mathbf{S}_i)]\,,\ i > 1. \tag{4.5}$$

The Greedy Search Algorithm does not ensure the optimal set but rather one which is near optimal [75]. In order to reduce computational time, Zhu *et al.* [75] recommend the use of the updated Kriging variance equation. If one applies the updated Kriging variance equation as derived for Ordinary Kriging by Barnes and Watson [5], one is able to obtain [75]:

$$\sigma^2(Z^*(\mathbf{s}|\mathbf{S}_{i+1})) = \sigma^2(Z^*(\mathbf{s}|\mathbf{S}_i)) - \frac{(r_{i+1} - \hat{r}_{i+1})^2}{D}\,, \tag{4.6}$$

where

$$r_{i+1} = C(\mathbf{s}_{i+1}, \mathbf{s}), \qquad \hat{r}_{i+1} = \mathbf{\Lambda}_i \mathbf{H}_i^{-1} \mathbf{V}_i, \qquad D = C(\mathbf{s}_{i+1}, \mathbf{s}_{i+1}) - \mathbf{\Lambda}_i \mathbf{H}_i^{-1} \mathbf{\Lambda}_i^T$$

and

$$\mathbf{\Lambda}_i = [C(\mathbf{s}_{i+1}, \mathbf{s}_1), C(\mathbf{s}_{i+1}, \mathbf{s}_2), \cdots, C(\mathbf{s}_{i+1}, \mathbf{s}_i)].$$

Equation (4.6) in conjunction with the Greedy Search Algorithm (Equation (4.5)) leads to the following objective function [75]:

$$\text{obj}(\mathbf{s}_{i+1}^*) = \text{Argmax}\left\{\sum_{i=1}^{m} \frac{(r_{i+1} - \hat{r}_{i+1})^2}{D}\right\}. \tag{4.7}$$

This method allows the user to calculate $\mathbf{H}_i^{-1}$ only once per iteration instead of calculating m different inverses, which greatly reduces computational time [75].

For the case of Simple Kriging, the basis function set is void as opposed to the single basis function ($f_1 = 1$) in Ordinary Kriging and for Universal Kriging the basis function set contains monomials of the spatial coordinates [23]. These changes influence the matrices which are used in the equations and affect the matrix sizes, but they do not change the equations in general.

4.4 Addition of a Point at Point of Maximum Error Variance

Brodkin [8] made use of this method in order to minimise large data sets in a controlled manner while Duckett made use of it, citing Brodkin, in the field of terrain modelling [22]. This method consists of the addition of a single point at a time to an existing set. The additional point is selected as the point at which the estimated error variance was greatest [8]. The algorithm for the development of this minimal set is described below [8]:

Step 1: Selection of an initial set of points

Step 2: Calculation of values for the variable of interest and estimated error variance by the method of Kriging

Step 3: Find the location and the value of the largest estimated error variance

Step 4: Compare the maximum estimated error variance to a predefined value. If the maximum estimated error variance exceeds this value, add this point and return to Step 2. If the maximum estimated error variance is less than this value then stop the algorithm.

Brodkin [8] advises that this method will not produce the optimal minimal set but that the sub-optimal set produced will not be significantly different and that the difference in computational time between this method and the branch and bound method to search every possible set makes the use of this method worth the difference in optimality.

In addition, Brodkin [8] proposes that to calculate the inverse of the correlation matrix ($\mathbf{R}$) after the addition of a point, the following method which makes use of the inverse calculated before the addition of the point should be used in order to reduce the computational time.

4.4.1 Calculating the Inverse of the Correlation Matrix

Suppose that two matrices $\mathbf{A}$ and $\mathbf{B}$ are partitioned into submatrices as follows [8]:

$$\mathbf{A} = \begin{bmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{S} & \mathbf{T} \end{bmatrix}, \tag{4.8}$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{bmatrix}, \tag{4.9}$$

where $\mathbf{AB} = \mathbf{I}$ ($\mathbf{I}$ is an identity matrix) and $\mathbf{P}^{-1}$ is known. The aim is to find $\mathbf{B} = \mathbf{A}^{-1}$. From the multiplication, it is found that [8]:

$$\mathbf{PE} + \mathbf{QG} = \mathbf{I}, \tag{4.10}$$

$$\mathbf{PF} + \mathbf{QH} = \mathbf{0}, \tag{4.11}$$

$$\mathbf{SE} + \mathbf{TG} = \mathbf{0}, \tag{4.12}$$

$$\mathbf{SF} + \mathbf{TH} = \mathbf{I}. \tag{4.13}$$

Solving Equations (4.10) to (4.13) for $\mathbf{E}$, $\mathbf{F}$, $\mathbf{G}$ and $\mathbf{H}$ yields the following relations [8]:

$$\mathbf{E} = \mathbf{P}^{-1} - (\mathbf{P}^{-1}\mathbf{Q})\mathbf{G}, \tag{4.14}$$

$$\mathbf{F} = -\mathbf{P}^{-1}\mathbf{QH}, \tag{4.15}$$

$$\mathbf{G} = -\mathbf{HSP}^{-1}, \tag{4.16}$$

$$\mathbf{H} = (\mathbf{T} - \mathbf{SP}^{-1}\mathbf{Q})^{-1}. \tag{4.17}$$

In the case of inverting the correlation matrix, $\mathbf{P}^{-1}$ is equivalent to the inverse of the correlation matrix from the previous iteration, $\mathbf{A}$ is equivalent to the new correlation matrix ($\mathbf{R}$) and $\mathbf{B}$ is equivalent to the inverse of the new correlation matrix ($\mathbf{R}^{-1}$). $\mathbf{Q}$, $\mathbf{S}$ and $\mathbf{T}$ are the matrices which are appended to the previous correlation matrix due to the addition of a point.

CHAPTER 5

DATA

5.1 The Model

Carbon dioxide sea-air flux data in the Southern Ocean were used in this study. The data provided for this study was generated using the biogeochemical PISCES (Pelagic Interaction Scheme for Carbon and Ecosystem Studies) model embedded within ORCA2-LIM ocean model [2]. The ORCA2-LIM model is based on an ORCA2 configuration of version 8.2 of the OPA (Océan Parallélisé) model [2] which is a general ocean circulation model [6]. The data used in this study was obtained from the Council for Scientific and Industrial Research Ocean Systems and Climate Research Group.

The PISCES model's development can be traced back to the P3ZD model released in 1997 and combines some aspects of a quota model with a Monod model [1]. It has been found that this model data does under-estimate the sea-air carbon dioxide flux especially in summer months [60]. However, it is recognised that this model is a simplified representation of the real world problem [2]. For the purposes of this study, however, the model data are taken to be correct and without error.

5.2 The Data Used

The units in which the data are represented is mol(C)/m^{2}s [60]. The data covers the area from approximately 180°W to approximately 180°E (that is from approximately $-180°$ to approximately 180°) and from approximately 39°S to 78°S (that is from approximately $-78°$ to approximately $-39°$). The resolution in the W-E (longitudinal) direction is approximately 2°, while the resolution in the N-S (latitudinal) direction is 2cos(latitude)$^\circ$ [2]. In addition, the data used in this study occurs in 73 time slices, the data at each time slice representing the mean data for a five day period [60]. The approximate area of the world covered by this data can be seen as the area found within the red block in Figure 5.1.

The data is given on a grid which has reference numbers of 0-72 in the direction of time, $0-44$ in the direction of latitude and $0-179$ in the direction of longitude. It is these grid references which will be used in the calculation of trend and covariance parameters as well as for prediction. Rectangular grids are quite common and as such, it was decided to make use of the grid references.

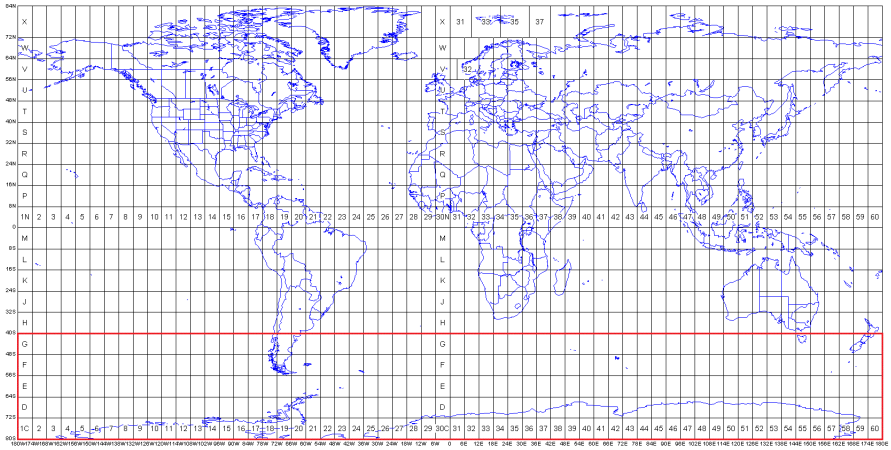


Figure 5.1: World map with area for this study indicated by red block (map obtained from www.dmap.co.uk/utmworld.gif [54]).

5.3 Excluded Data

Locations for which the carbon dioxide flux value was less than 1×10^{-10} mol(C)/m^{2}s were removed from the data set to be further used in this study. These removals resulted in a data set containing 419093 locations. All the data points removed were checked and found to have values of exactly 0 mol(C)/m^{2}s and thus, for the purposes of this study were deemed either to be land or areas of severe ice coverage. The removal of such points also resulted in the removal of all points with a latitude grid reference of 0, as the line of latitude associated with this grid reference is purely land and/or ice cover. This can be seen for a single time step in Figure 5.2.

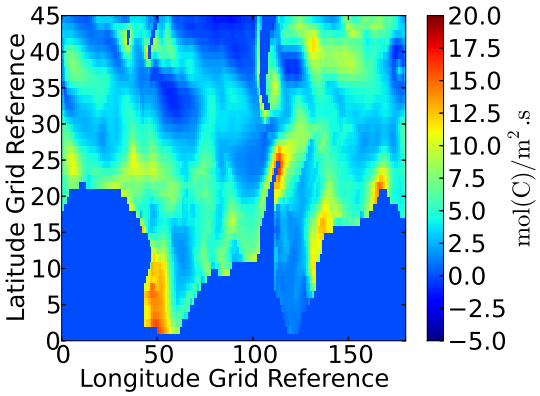


Figure 5.2: Map of carbon dioxide flux across the whole southern ocean for time slice 0.

5.4 Limitations

It is acknowledged that in the directions of time and longitude, the last grid reference would in reality occur next to the first grid reference. This occurs as the time coverage is for a whole year, that is, the last point in time would occur just before the first point in time in the next year. For the case of longitude, as the data covers the full 360° when grid size is taken into account, the first longitudinal grid reference must occur next to that of the last. This, however, has not being taken into account in this study. Rather, the data has been treated as if it does not wrap around to meet itself again in these directions.

CHAPTER 6

DETERMINING THE PRESENCE OF TREND

6.1 Determing the Presence of Trend

In order to determine whether trend is present a histogram of the data is considered as well as the coefficient of determination (R^2) values of various trends. In addition, a normal (Gaussian) distribution is fitted by means of Python's stats package [35] (See Appendix A for a short description of the stats package) and this is also plotted onto the histogram of the data. If the data possesses a non-zero mean, then at least a constant trend is present. Ideally for Kriging, the data should possess no trend, have a zero mean and follow a normal distribution.

Figure 6.1(a) shows the histogram of the original data. It can clearly be seen that the mean is non-zero. The data also does not appear to follow a normal distribution perfectly, this is corroborated by a skewness value of 0.8628 and a kurtosis value of 4.7108. This leads to the conclusion that a trend is indeed present in the data.

In this study, only a constant trend, first order trend and second order trend in the spatial and temporal co-ordinate references are considered. The reference co-ordinates were used to provide a representation of the actual time, longitude and latitude which would be easy to use. These three trend structures can be seen as:

- Constant Trend - $\mu(\mathbf{s}) = \beta_0$
- First Order Trend - $\mu(\mathbf{s}) = \beta_0 + \beta_1 s_{\text{time}} + \beta_2 s_{\text{lat}} + \beta_3 s_{\text{lon}}$
- Second Order Trend - $\mu(\mathbf{s}) = \beta_0 + \beta_1 s_{\text{time}} + \beta_2 s_{\text{lat}} + \beta_3 s_{\text{lon}} + \beta_4 s_{\text{time}}^2 + \beta_5 s_{\text{lat}}^2 + \beta_6 s_{\text{lon}}^2 + \beta_7 s_{\text{time}} s_{\text{lat}} + \beta_8 s_{\text{time}} s_{\text{lon}} + \beta_9 s_{\text{lat}} s_{\text{lon}}$

where $\mathbf{s} = [s_{\text{time}}, s_{\text{lat}}, s_{\text{lon}}]$ is the position on the grid, s_{time} is the grid reference(0-72) for time, s_{lat} is the grid reference (1-44) for latitude and s_{lon} is the grid reference(0-179) for longitude.

In order to determine which of these trends should be used, Ordinary Least Squares is applied on the whole data set. Ordinary Least Squares is computationally very fast and is thus easy to implement on the whole set. While Ordinary Least Squares is no longer optimal, it should provide a good indication of the effect of the trend removal, since Hengl *et al.* [31] mention that Ordinary Least Squares is often deemed sufficient. These Ordinary Least Squares trends are then removed from the data and the residuals are once again plotted as histograms along

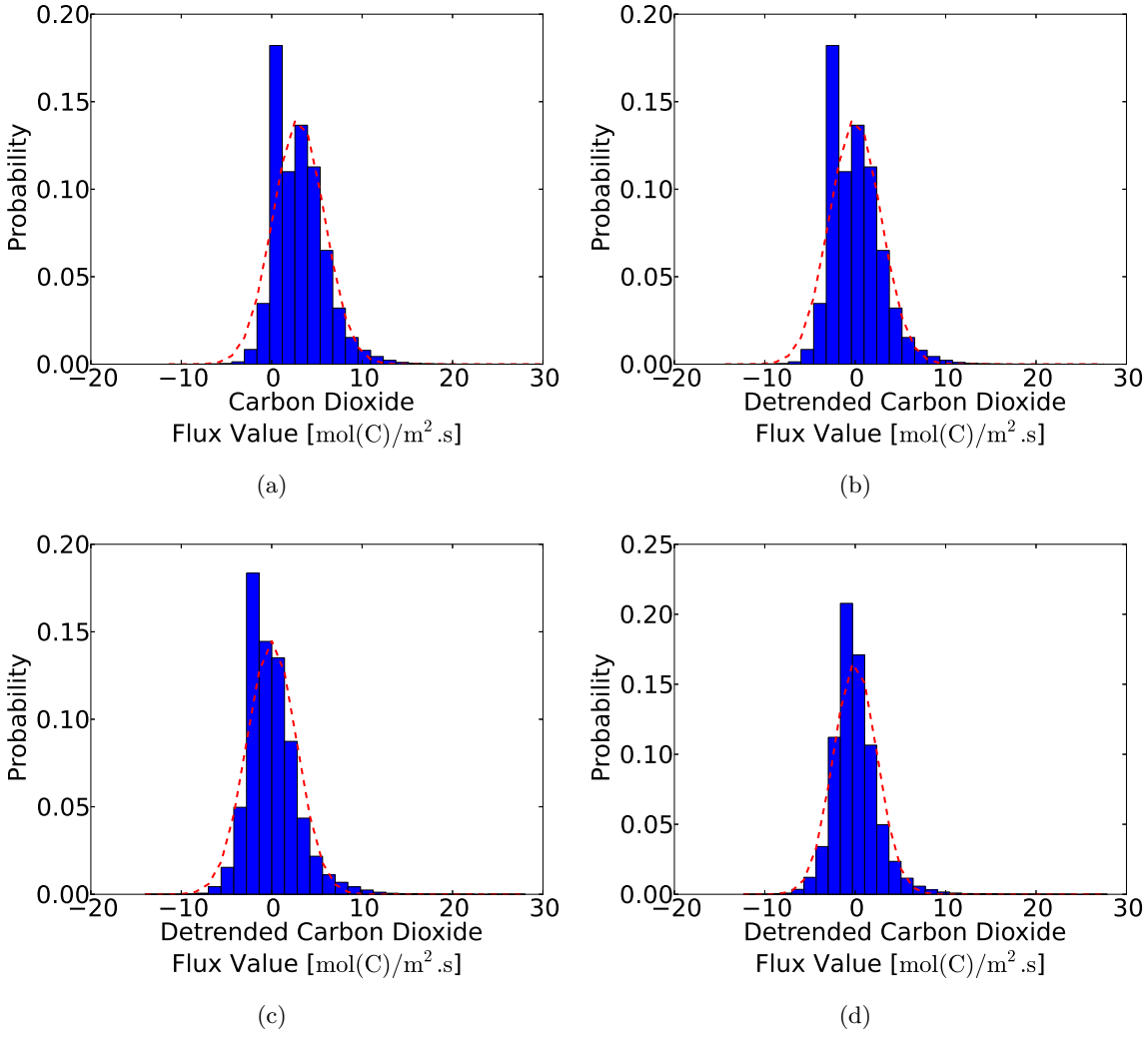


Figure 6.1: Histograms of the residuals after the removal of (a) no trend, (b) constant trend, (c) linear trend and (d) quadratic trend along with their respective fitted normal distributions.

with fitted normal distributions. We consider these histograms (Figures 6.1(b), 6.1(c) and 6.1(d)) and check once again for zero mean and normality. The Ordinary Least Squares fitted trend coefficients are presented in Appendix B.

All three trends produce means which appear to be near zero. None of the three trends perfectly matches a normal distribution, this is corroborated by the skewness and kurtosis values for the residuals found in Table 6.1.

Trend	Skewness	Kurtosis
Constant	0.8628	4.7108
Linear	1.1273	5.9409
Quadratic	0.9989	6.6819

Table 6.1: Skewness and kurtosis values for tested trends.

In order to determine which trend should be used, the coefficient of determination (R^2) was calculated for each of the trends.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (Z_i - Z_i^*)^2}{\sum_{i=1}^{n} (Z_i - \text{mean}(Z))^2}, \tag{6.1}$$

where Z is the variable of interest, Z_i is the variable of interest at location i and Z_i^* is the Ordinary Least Squares estimate for the variable of interest at location i.

It is acknowledged that due to the correlation between the data points, the true interpretation of this statistic as the proportion of the sample variation in Z which can be explained by the using the independent variables to predict Z in the linear model is no longer valid. However, the premise of the closer the R^2 value is to one the better the model still holds. These values can be seen in Table 6.2.

Trend	R^2
Constant	2.220×10^{-16}
Linear	0.0614
Quadratic	0.285

Table 6.2: Coefficient of determination for tested trends.

While even the quadratic trend has a low coefficient of determination, it is significantly higher than those for the constant and linear trends. Thus, the second order (quadratic) trend is chosen as the trend structure to be used further in this study. This low coefficient of determination also suggests the need for the use of another method in conjunction with the quadratic trend.

6.2 Limitations

The trend structure is chosen based on Ordinary Least Squares trends. These trends are not optimal and do not take into account the spatial and temporal correlation between points. Thus, the resulting residuals when a more optimal technique is used to fit the trend may vary in both value and distribution when compared to the Ordinary Least Squares residuals. As mentioned in Section 6.1 Ordinary Least Squares is often sufficient. More optimal techniques

require much greater computational time and thus, Ordinary Least Squares was chosen to determine the trend structure.

Only three linear polynomial trend structures were considered. Regression Kriging allows for the choice of arbitrarily complex trend structures and as such, this study does not claim that the optimal trend structure was chosen, but merely that the one chosen was deemed to be the best of the three considered.

CHAPTER 7

DETERMINING TREND COEFFICIENTS AND COVARIANCE PARAMETERS

7.1 Covariance Structures

As explained in Section 6.1 a quadratic trend structure was chosen. This trend structure was fitted in conjunction with three separable covariance structures using Maximum Likelihood Estimation. The three separable covariance structures fitted were:

- Exponential - $C(\mathbf{h}) = \sigma^2 e^{(-h_{\text{time}}/a_{\text{time}})} e^{(-h_{\text{lat}}/a_{\text{lat}})} e^{(-h_{\text{lon}}/a_{\text{lon}})}$
- Pentaspherical -

$$C(h_{\text{time}}) = \begin{cases} \sigma^2_{\text{time}} \left\{ 1 - \dfrac{15h_{\text{time}}}{8a_{\text{time}}} - \dfrac{5}{4}\left(\dfrac{h_{\text{time}}}{a_{\text{time}}}\right)^3 + \dfrac{3}{8}\left(\dfrac{h_{\text{time}}}{a_{\text{time}}}\right)^5 \right\} & \text{for } h_{\text{time}} \leq a_{\text{time}} \\ \sigma^2_{\text{time}} & \text{for } h_{\text{time}} \geq a_{\text{time}} \end{cases}$$

$$C(h_{\text{lat}}) = \begin{cases} \sigma^2_{\text{lat}} \left\{ 1 - \dfrac{15h_{\text{lat}}}{8a_{\text{lat}}} - \dfrac{5}{4}\left(\dfrac{h_{\text{lat}}}{a_{\text{lat}}}\right)^3 + \dfrac{3}{8}\left(\dfrac{h_{\text{lat}}}{a_{\text{lat}}}\right)^5 \right\} & \text{for } h_{\text{lat}} \leq a_{\text{lat}} \\ \sigma^2_{\text{lat}} & \text{for } h_{\text{lat}} \geq a_{\text{lat}} \end{cases}$$

$$C(h_{\text{lon}}) = \begin{cases} \sigma^2_{\text{lon}} \left\{ 1 - \dfrac{15h_{\text{lon}}}{8a_{\text{lon}}} - \dfrac{5}{4}\left(\dfrac{h_{\text{lon}}}{a_{\text{lon}}}\right)^3 + \dfrac{3}{8}\left(\dfrac{h_{\text{lon}}}{a_{\text{lon}}}\right)^5 \right\} & \text{for } h_{\text{lon}} \leq a_{\text{lon}} \\ \sigma^2_{\text{lon}} & \text{for } h_{\text{lon}} \geq a_{\text{lon}} \end{cases}$$

$$C(\mathbf{h}) = C(h_{\text{time}})C(h_{\text{lat}})C(h_{\text{lon}})$$

  Within Maximum Likelihood Estimation, $\sigma^2 = \sigma^2_{\text{time}}\sigma^2_{\text{lat}}\sigma^2_{\text{lon}}$ is estimated, not the individual terms.

- Approximate Gaussian - $C(\mathbf{h}) = \sigma^2 e^{\left(-h^{1.99}_{\text{time}}/a^{1.99}_{\text{time}}\right)} e^{\left(-h^{1.99}_{\text{lat}}/a^{1.99}_{\text{lat}}\right)} e^{\left(-h^{1.99}_{\text{lon}}/a^{1.99}_{\text{lon}}\right)}$

In each of these three structures, h_{time} represents the lag in time grid references, h_{lat} represents the lag in latitudinal grid references and h_{lon} represents the lag in longitudinal grid references. The range parameters then represent ranges in the respective directions for the grid references. Due to the manner is which the latitudinal grid is calculated, it is acknowledged that a lag distance (h_{lat}) of 1 may not always represent the same distance in true latitude. However, all calculations will be completed using the grid references. In essence, the changing latitudinal

resolution will merely imply that the conversion of grid reference range to true range will differ based on latitudinal position. That is, in areas of higher resolution, the range will cover a smaller true latitudinal range than in areas of lower resolution.

Only the three separable covariance structures given above are considered. It is not implied in any way that these structures will best fit the data out of all possible structures. This study does however, attempt to select the best of these three structures for use in comparing the various sampling methods.

7.2 Details of Covariance Structure and Trend Fitment

The covariance structures described above were fitted along with the quadratic trend using 20 randomly chosen sets each containing 10000 data points. The structures were fitted as previously mentioned using Maximum Likelihood Estimation. The likelihood was optimised using Scipy's built in optimise.minimise function [35]. Within this function, Sequential Least Squares Programming was selected as the method of optimisation and lower bounds of 1 were given for the range or distance parameters. The lower bounds of 1 are given as a range of 1 already implies that the data are independent as the data has minimum grid reference distances of 1 between points.

In order to determine reasonable starting values for the optimisation across the sets of 10000 points described above, a single set of 1000 points was chosen randomly. This set of 1000 points was used to fit the parameters for all three covariance structures using the same method as was later applied to the sets of 10000 points. a_{time}, a_{lat} and a_{lon} were all set to a value of 5 in order to provide a starting point. The run across this set converged for all three covariance structures, and the end values from this run were used to start the runs across the 20 sets of 10000 data points for their respective covariance structures.

Each of the Maximum Likelihood Estimations reached convergence and the resulting parameter values for all 20 random sets across the three covariance structures can be seen in Appendix C.

The trend coefficients and covariance parameters were averaged with equal weighting across the 20 sets. The averaged (mean) trend and covariance parameters for the three covariance structures can be seen in Table 7.1, while the standard deviations for these parameters can be seen in Table 7.2. Histograms of these parameters across the 20 randomly selected sets can be seen in Appendix C.

It is acknowledged that absolute convergence of the parameters may not have been reached in terms of the values of all parameters. That is, some parameters exhibit standard deviations which are large in relation to their means. This is especially true when the parameters have obtained both positive and negative values across the 20 random sets.

However, for the purposes of this study the averaged coefficients are taken to be the true exact parameters as it is deemed that these averages will provide a good representation of the true parameters. As such, the trend is considered to be deterministic and without error. Therefore, the error variance associated with the trend for Regression Kriging is considered to be zero. That is, the estimated error variance is reduced to the Simple Kriging error variance.

Covariance Structure	σ^2	β		a	
Exponential	4.9273	β_0	8.9266	a_{time}	23.5807
		β_1	-1.9649×10^{-1}	a_{lat}	6.2075
		β_2	-1.4053×10^{-1}	a_{lon}	14.8855
		β_3	-2.7621×10^{-2}		
		β_4	1.6380×10^{-3}		
		β_5	1.4693×10^{-4}		
		β_6	7.9951×10^{-5}		
		β_7	2.4664×10^{-3}		
		β_8	2.1888×10^{-5}		
		β_9	5.2173×10^{-4}		
Pentaspherical	5.1382	β_0	7.8040	a_{time}	50.8998
		β_1	-1.7662×10^{-1}	a_{lat}	13.4278
		β_2	-1.1027×10^{-1}	a_{lon}	33.0400
		β_3	-2.2152×10^{-2}		
		β_4	1.5025×10^{-3}		
		β_5	4.6895×10^{-5}		
		β_6	5.3942×10^{-5}		
		β_7	2.1620×10^{-3}		
		β_8	1.2873×10^{-5}		
		β_9	5.2266×10^{-4}		
Approximate Gaussian	4.6081	β_0	9.2587	a_{time}	11.6437
		β_1	-2.6595×10^{-1}	a_{lat}	8.3595
		β_2	-2.0545×10^{-1}	a_{lon}	20.7621
		β_3	-4.0304×10^{-3}		
		β_4	1.6972×10^{-3}		
		β_5	1.1091×10^{-3}		
		β_6	-3.2337×10^{-5}		
		β_7	4.7296×10^{-3}		
		β_8	4.3576×10^{-5}		
		β_9	3.9944×10^{-4}		

Table 7.1: Average trend and covariance parameters for various covariance structures.

Covariance Structure	σ^2		β		a
Exponential	0.212820	β_0	0.534993	a_{time}	1.192262
		β_1	0.011179	a_{lat}	0.168571
		β_2	0.0192506	a_{lon}	0.648767
		β_3	0.006238		
		β_4	6.861867×10^{-5}		
		β_5	0.000250		
		β_6	1.861939×10^{-5}		
		β_7	0.000208		
		β_8	5.3086112×10^{-5}		
		β_9	0.0001125		
Pentaspherical	0.259799	β_0	0.554664	a_{time}	1.970660
		β_1	0.011705	a_{lat}	0.399957
		β_2	0.020361	a_{lon}	1.404853
		β_3	0.005817		
		β_4	7.133221×10^{-5}		
		β_5	0.000253		
		β_6	1.568352×10^{-5}		
		β_7	0.000211		
		β_8	5.117492×10^{-5}		
		β_9	0.000109		
Approximate Gaussian	0.109099	β_0	0.291157	a_{time}	2.408253
		β_1	0.005037	a_{lat}	0.849498
		β_2	0.010475	a_{lon}	3.096664
		β_3	0.003913		
		β_4	5.576261×10^{-5}		
		β_5	0.000183		
		β_6	9.915505×10^{-6}		
		β_7	0.000114		
		β_8	2.356401×10^{-5}		
		β_9	6.418317×10^{-5}		

Table 7.2: Standard deviations for the trend and covariance parameters for various covariance structures.

7.3 Validation of Covariance Structures and Trend

In order to select the best covariance structure of the three tested using the averaged parameters from Maximum Likelihood Estimation, 10 sets of 50000 data points were randomly chosen to act as validation sets. For each of these 10 sets, 5000 points were randomly selected as sampled points to be used for prediction. This is equivalent to 10% of each of the data sets. Given the conditions found in the Southern Ocean and its vast area, it is unrealistic to expect that a very large percentage of the possible data points will eventually be sampled. As such, 10% is chosen here and later in this study, 10% and 5% will be used as limits for sampling. The trend was removed from all 50000 points in each set and using the de-trended 5000 sub-sampled points, prediction by means of Simple Kriging with a zero mean was performed on the residuals. The trend was then added back to the predictions and various error measures were calculated. The error measures calculated were:

- Maximum absolute error -

$$\text{Maximum Absolute Error} = \text{Argmax}\{\mid Z^*(\mathbf{s}) - Z(\mathbf{s}) \mid\}$$

- Total absolute error -

$$\text{Total Absolute Error} = \sum_{i=1}^{n}\{\mid Z^*(\mathbf{s}_i) - Z(\mathbf{s}_i) \mid\}$$

- Total estimated error variance -

$$\text{Total Estimated Error Variance} = \sum_{i=1}^{n}\{\sigma^2(Z^*(\mathbf{s}_i))\}$$

- Percentage absolute error - this is calculated as the total absolute error divided by the total absolute value multiplied by 100. The total absolute value is the sum across all points of the absolute value of the true value at that point.

$$\text{Percentage Absolute Error} = \frac{\text{Total Absolute Error}}{\text{Total Absolute Value}} \times 100$$

The error measurements mentioned above for the 10 validation sets can be found in Table 7.3 for the exponential covariance structure, Table 7.4 for the spherical covariance structure and Table 7.5 for the approximate Gaussian covariance structure.

In addition histograms of the residuals were generated as described in Section 6.1. These histograms can be seen in Figures 7.1(a), 7.1(b) and 7.1(c). From these histograms, it can be seen that none of the resulting sets follow a strictly normal distribution. However, they all appear to have means close to zero. The mean values for the residuals from the different covariance structures can be seen in Table 7.6. It is clear that the approximate Gaussian distribution gives a mean closest to zero of the 3 covariance structures. However, the final decision on which covariance structure is to be used further in this study is based on the errors calculated and presented in Tables 7.3 to 7.5.

The error measurements were considered and in particular, the total absolute error and its scaled counterpart (the percentage absolute error) were considered as the deciding factors in the process of deciding which covariance structure would be the best of the three structures for further use. Based on the exponential covariance structure having the lowest total absolute error and thus, the lowest percentage absolute error, it was chosen for use in the comparison of sampling methods.

Random Set	Maximum Absolute Error	Total Absolute Error	Total Absolute Value	Percentage Absolute Error	Total Estimated Error Variance
1	15.67	27280.05	156135.25	17.47%	39391.24
2	14.52	27336.75	157030.55	17.41%	39297.66
3	12.16	27546.57	157385.71	17.50%	39365.18
4	16.70	27516.50	156604.59	17.57%	39028.43
5	12.87	27093.81	156823.91	17.28%	39123.10
6	13.71	26889.47	156109.29	17.22%	39282.29
7	17.39	27066.43	156591.10	17.28%	39382.37
8	12.25	27884.11	157136.77	17.75%	39145.60
9	13.80	27509.44	156292.28	17.60%	39339.25
10	13.55	26905.15	156867.59	17.15%	39108.94
Average	14.26	27302.83	156697.70	17.42%	39246.41

Table 7.3: Errors associated with the exponential covariance structure over 10 random sets.

Random Set	Maximum Absolute Error	Total Absolute Error	Total Absolute Value	Percentage Absolute Error	Total Estimated Error Variance
1	16.26	35782.71	156135.25	22.92%	38813.49
2	14.75	36037.39	157030.55	22.95%	38711.56
3	11.96	36275.73	157385.71	23.05%	38794.21
4	16.81	36427.29	156604.59	23.26%	38422.53
5	13.92	35825.17	156823.91	22.84%	38530.54
6	13.81	35535.46	156109.29	22.76%	38706.98
7	17.29	35887.49	156591.10	22.92%	38802.45
8	12.50	36034.78	157136.77	22.93%	38540.01
9	13.92	35673.65	156292.28	22.83%	38745.67
10	14.18	35937.23	156867.59	22.91%	38507.85
Average	14.54	35941.69	156697.70	22.94%	38657.53

Table 7.4: Errors associated with the pentaspherical covariance structure over 10 random sets.

Random Set	Maximum Absolute Error	Total Absolute Error	Total Absolute Value	Percentage Absolute Error	Total Estimated Error Variance
1	16.94	51703.65	156135.25	33.12%	77754.77
2	31.00	51583.90	157030.55	32.85%	77779.37
3	20.37	51597.24	157385.71	32.78%	77801.43
4	16.87	51204.00	156604.59	32.70%	77108.03
5	15.52	51582.56	156823.91	32.89%	76843.86
6	13.36	51519.86	156109.29	33.00%	77326.22
7	18.56	52405.01	156591.10	33.47%	77390.96
8	13.69	52197.04	157136.77	33.22%	77440.21
9	16.04	52028.80	156292.28	33.29%	77729.77
10	16.16	51015.80	156867.59	32.52%	77127.03
Average	17.85	51683.79	156697.70	32.98%	77430.16

Table 7.5: Errors associated with the approximate Gaussian covariance structure over 10 random sets.

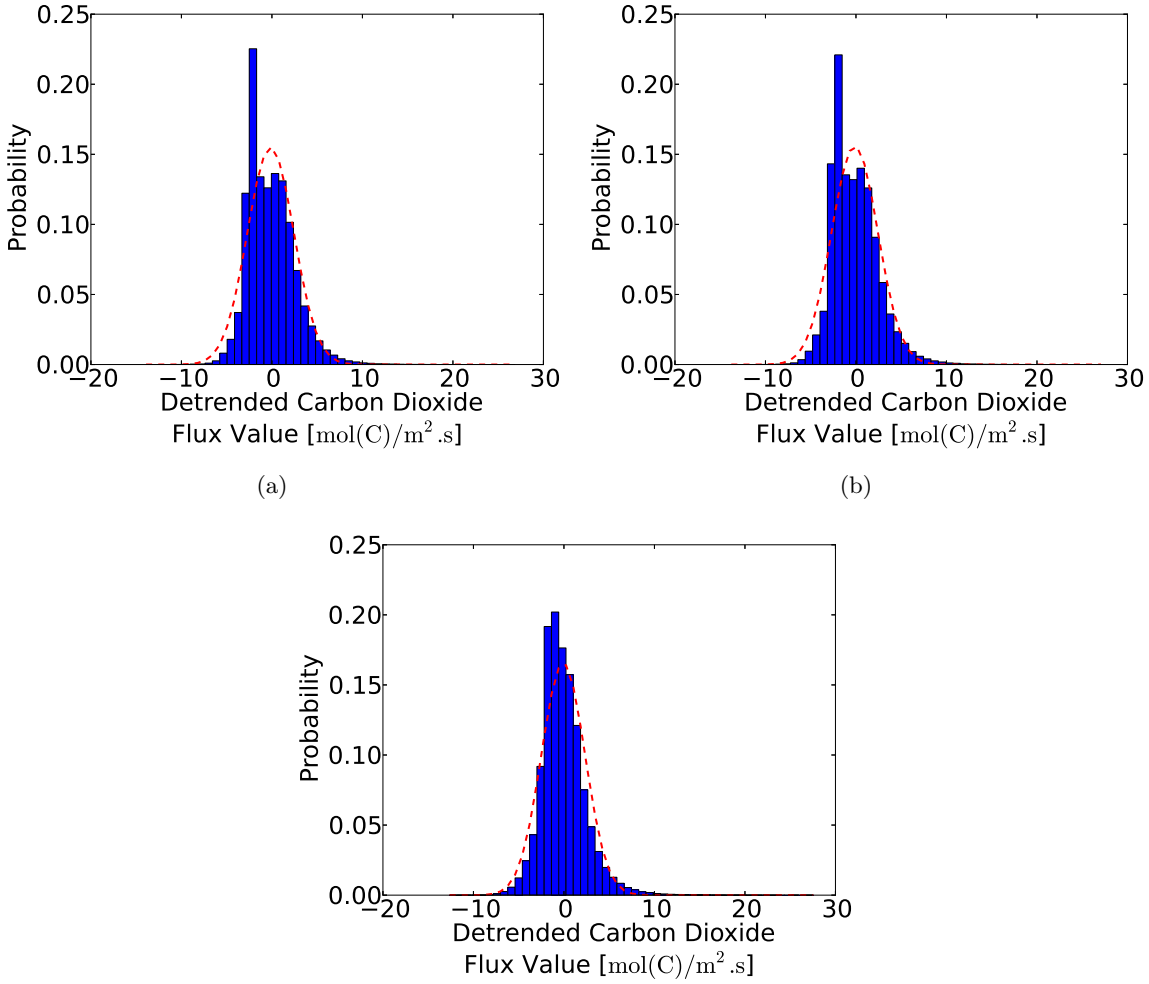


Figure 7.1: Histograms of the residuals after the removal of quadratic trend with co-efficients (as shown in Table 7.1) from the (a) exponential covariance structure, (b) pentaspherical covariance structure and (c) Gaussian covariance structure along with their respective fitted normal distributions.

Covariance Structure	Mean
Exponential	−0.1426
Spherical	−0.1356
Approximate Gaussian	−0.0807

Table 7.6: Mean value of the residuals for various covariance structures.

7.4 Final Covariance Structure and Trend

As the exponential covariance structure was chosen, the following trend and separable covariance structure will be considered as the true trend and covariance structure for the remainder of this study (across all methods for both two-dimensional and three-dimensional sampling).

$$\begin{aligned}\mu(\mathbf{s}) = 8.9266 - 1.9649 \times 10^{-1} s_{\text{time}} - 1.4053 \times 10^{-1} s_{\text{lat}} - 2.7621 \times 10^{-2} s_{\text{lon}} + \\ 1.6380 \times 10^{-3} s_{\text{time}}^2 + 1.4693 \times 10^{-4} s_{\text{lat}}^2 + 7.9951 \times 10^{-5} s_{\text{lon}}^2 + 2.4664 \times 10^{-3} s_{\text{time}} s_{\text{lat}} + \\ 2.1888 \times 10^{-5} s_{\text{time}} s_{\text{lon}} + 5.2173 \times 10^{-4} s_{\text{lat}} s_{\text{lon}}\end{aligned} \tag{7.1}$$

$$C(\mathbf{h}) = 4.9273 e^{(-h_{\text{time}}/23.5807)} e^{(-h_{\text{lat}}/6.2075)} e^{(-h_{\text{lon}}/14.8855)} \tag{7.2}$$

This study takes these trend and covariance structures to be constant across the entire sampling process. This is done to ensure that these parameters are suitable for the final number of sample points selected. Some of the methods implemented such as the genetic algorithms, do not build up from a smaller set to the required size but rather always select sets of the required size. Thus, they require a suitable covariance structure for the required size data set from the offset. Additionally, for the implementation of the Updated Kriging Variance Algorithm, Zhu *et al.* [75] maintain a constant covariance structure across the sampling algorithm, while for Addition of a Point at Point of Maximum Estimated Error Variance Brodkin's [8] algorithm holds both the trend and covariance structure constant across the sampling procedure.

7.5 Limitations

Only three covariance structures were fitted. These structures were all separable. There are many covariance structures which could be considered including ones which take into account anisotropy and non-separability. One of these many structures could outperform the final exponential covariance structure chosen. Thus, it is not claimed the the chosen covariance structure is optimal, rather that it is the best of the three tested.

Additionally, only Maximum Likelihood Estimation is used to fit the three chosen covariance structures. Many other methods are available but for the reasons given in Section 3.2.2 maximum likelihood was used.

It is acknowledged that some of the parameters have a standard deviation which is large in relation to the mean for those parameters which could indicate that they have not fully reached convergence. However, it was deemed that using 10000 points should give values close to the true values. As such, the values were averaged across the 20 random sets. This study does not claim that the values for the parameters given are the true values. However, for the purposes of this study, they are considered close enough to the true values to be used as such.

CHAPTER 8

TWO-DIMENSIONAL SAMPLING - INTRODUCTION

The following chapters cover various sampling strategies which were conducted on a smaller sub-sampled two-dimensional set of data. The subset consisted of 671 sub-sampled points with a time reference of 72, latitude references between 25 and 35 and longitude references between 40 and 100. The time step and area chosen were selected as they showed large amounts of variability. The variability was required in order to determine if the sampling methods could in fact reproduce large amounts of variability. The sampling strategies tested consisted of both *a-priori* and *a-posteriori* methods and were used to select 70 of the 671 sub-sampled points. That is, approximately 10% of the sub-sampled points were selected as sample points to be used for estimation purposes. The sum of the absolute values associated with this set is 1733.32 mol(C)/m^{2}s, while the sum of the values associated with this set is 1711.04 mol(C)/m^{2}s.

The prediction process used with these samples can be seen as follows:

Step 1: Remove the trend as found in Equation (7.1) from the data including the sample points

Step 2: Perform Simple Kriging using the residuals calculated in step one using the covariance structure as found in Equation (7.2). Calculate both the estimated value and the estimated error variance at each point.

Step 3: Add the trend from Step 1 back to the estimated values from Step 2 to form the final predicted values.

Step 4: Calculate errors between the original data and the final predicted values obtained in Step 3. The estimated error variance is solely from Simple Kriging.

Seven of the chosen sampling strategies were designed to minimise errors, one was random and the remaining two were designed to maximise errors in an attempt to gauge the worst performance possible. These sampling strategies and their results are explained and discussed in the following chapters. In addition, the strategies were compared in order to determine which of them should be implemented on a larger three-dimensional data set.

The various sampling methods tested were:

- Random Sampling -Chapter 9
- Updated Kriging Variance Algorithm [*a-priori*] -Chapter 10

- Addition of a Point at Point of Maximum Error Variance [*a-priori*] -Chapter 11
- Addition of a Point at Point of Maximum Absolute Error [*a-posteriori*] -Chapter 12
- Genetic Algorithm designed to Minimise Total Absolute Error [*a-posteriori*] -Chapter 13
- Genetic Algorithm designed to Minimise Total Error Variance [*a-priori*] -Chapter 13
- Hybridised Genetic Algorithm designed to Minimise Total Absolute Error [*a-posteriori*] -Chapter 13
- Hybridised Genetic Algorithm designed to Minimise Total Error Variance [*a-priori*] -Chapter 13
- Genetic Algorithm designed to Maximise Total Absolute Error [*a-posteriori*] -Chapter 13
- Genetic Algorithm designed to Maximise Total Error Variance [*a-priori*] -Chapter 13

For each of these methods, the following errors were calculated on each set/run:

- Maximum absolute error -

$$\text{Maximum Absolute Error} = \text{Argmax}\{\mid Z^*(\mathbf{s}) - Z(\mathbf{s}) \mid\}$$

- Total absolute error -

$$\text{Total Absolute Error} = \sum_{i=1}^{n}\{\mid Z^*(\mathbf{s}_i) - Z(\mathbf{s}_i) \mid\}$$

- Percentage absolute error - this is calculated as the total absolute error divided by the total absolute value multiplied by 100. The total absolute value is the sum across all points of the absolute value of the true value at that point.

$$\text{Percentage Absolute Error} = \frac{\text{Total Absolute Error}}{\text{Total Absolute Value}} \times 100$$

- Total integrated error - this is the difference between the sum of true values and the sum of the predicted values.

$$\text{Total Integrated Error} = \text{Total Value} - \text{Total Predicted Value}$$

  The total integrated error can be both positive or negative. A positive total integrated error implies that the total value has been under-estimated while a negative total integrated error implies over-estimation.

- Maximum estimated error variance -

$$\text{Maximum Estimated Error Variance} = \text{Argmax}\{\sigma^2(Z^*(\mathbf{s}))\}$$

- Total estimated error variance -

$$\text{Total Estimated Error Variance} = \sum_{i=1}^{n}\{\sigma^2(Z^*(\mathbf{s}_i))\}$$

In addition Chapter 14 addresses the comparison of the above mentioned methods.

CHAPTER 9

TWO-DIMENSIONAL SAMPLING - RANDOM SAMPLING

Sample sets of 70 points are chosen randomly without replacement, to prevent repetition of points. This ensures that the covariance matrix is positive definite. Each of the 671 points is assigned a number ranging from 0 to 670. The random function in Python is used to generate a value between 0 and 1. This value is then multiplied by 670 and the resulting value is passed through the int function to ensure it is an integer. The point assigned to this number is then added to the sample set and deleted from the set from which selection takes place. The remaining points are then assigned numbers in the range of 0 to 669. Again the random function is used to generate a value between 0 and 1 which is multiplied by 669 and then passed through the int function. The point associated with this integer is then added to the sample set and deleted from the selection set. This continues, reducing the number of points in the selection set by one each iteration until 70 points are present in the sample set.

This process was repeated 10 times. The 10 resulting sample sets were then used to predict the values of carbon dioxide at all 671 points.

9.1 Results and Discussion

The errors associated with these predictions can be found in Tables 9.1 and 9.2. The units associated with the maximum absolute error, total absolute error and total integrated difference are mol(C)/m^2s.

Considering the errors for sets one and two in Table 9.1, it can be seen that the total integrated error for two sets can be very different even when the two sets have very similar total absolute errors. This is due to the fact the the total integrated error is a sum of negative and positive components which can cancel each other out. This means that predicted values which are very different to the true values can sum to a value very similar to that of the sum of the true values. As such, the total integrated error was not minimised in any of the methods which follow in this study, as minimising this error does not guarantee predicted values which are similar to the true values.

Of interest from Tables 9.1 and 9.2 is that although set four has the highest total absolute error, it does not possess the highest total estimated error variance. The highest total estimated error variance belongs to set three which is neither the best nor the worst performing set. In addition the two very similar total absolute errors of sets one and two produced total estimated error variances with a difference of 48 between them. This demonstrates that a range of total absolute errors can be produced for the same total estimated error variance.

Random Set	Maximum Absolute Error	Total Absolute Error	Percentage Absolute Error	Total Integrated Error
1	1.5442	136.8507	7.8953%	6.4844
2	1.4520	137.4879	7.9321%	−39.2448
3	1.2438	139.9847	8.0761%	−32.1355
4	2.3398	157.3586	9.0784%	3.4931
5	1.7705	145.2019	8.3771%	−25.6501
6	1.7841	141.2688	8.1502%	24.5299
7	1.1492	132.4753	7.6429%	18.5520
8	1.3412	122.0329	7.0404%	24.9801
9	1.9810	153.5618	8.8594%	48.0460
10	1.4709	127.6217	7.3628%	−15.5297

Table 9.1: Maximum absolute, total absolute, percentage absolute and total integrated errors associated with Random Sampling in 2D.

Random Set	Maximum Estimated Error Variance	Total Estimated Error Variance
1	2.4731	510.2283
2	2.8782	558.1351
3	2.1022	576.3621
4	3.3111	562.5471
5	2.8670	546.3766
6	2.6058	539.1881
7	2.0994	524.7515
8	2.3992	525.0600
9	2.0062	551.5516
10	2.7088	570.7298

Table 9.2: Maximum estimated error variances and total estimated error variances associated with Random Sampling in 2D.

This also implies that one could see the same total absolute error with different total estimated error variances. The reason for this is that the estimated error variance relies solely on the distances between points and the covariance structure. Since the covariance structure is fixed, it is purely dependent on the distances between points. In order to minimise the estimated error variance, one needs to minimise the distances between the sampled and unsampled points. However, it is possible to have the same estimate error variance at two points, while having two different absolute errors (based on the two points having different flux values). This is why it is possible to have a higher total estimated error variance but still a lower total absolute error and vice versa.

Maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux, the estimated error variance and the true error across the chosen two-dimensional region for the best and worst performing sets (in terms of total absolute error minimisation) can be seen in Figures 9.1 and 9.2 respectively. The maps for the remaining sets can be seen in Appendix D.1.

From Figure 9.1, it is clear that even the best performing randomly selected sample set did not perform particularly well. There are very clear visual differences between the maps of the true values (Figure 9.1(a)) and predicted values (Figure 9.1(b)). The worst performing set produces an even more inaccurate map as can be seen from its map of true error, Figure

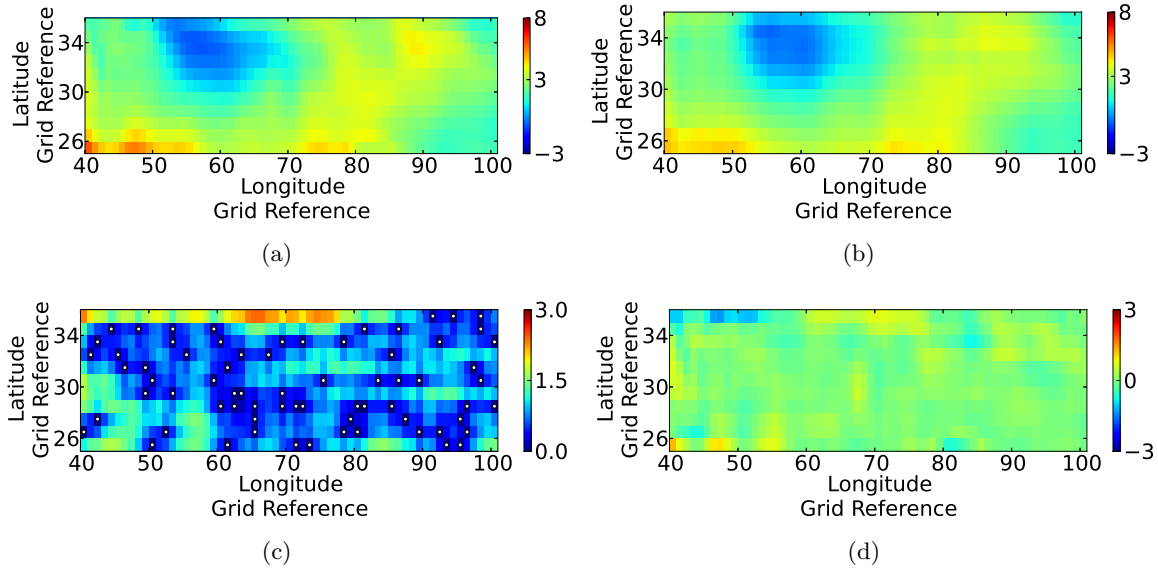


Figure 9.1: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for the best performing set, set eight, generated using Random Sampling. The white dots in (c) indicate the sampling points.

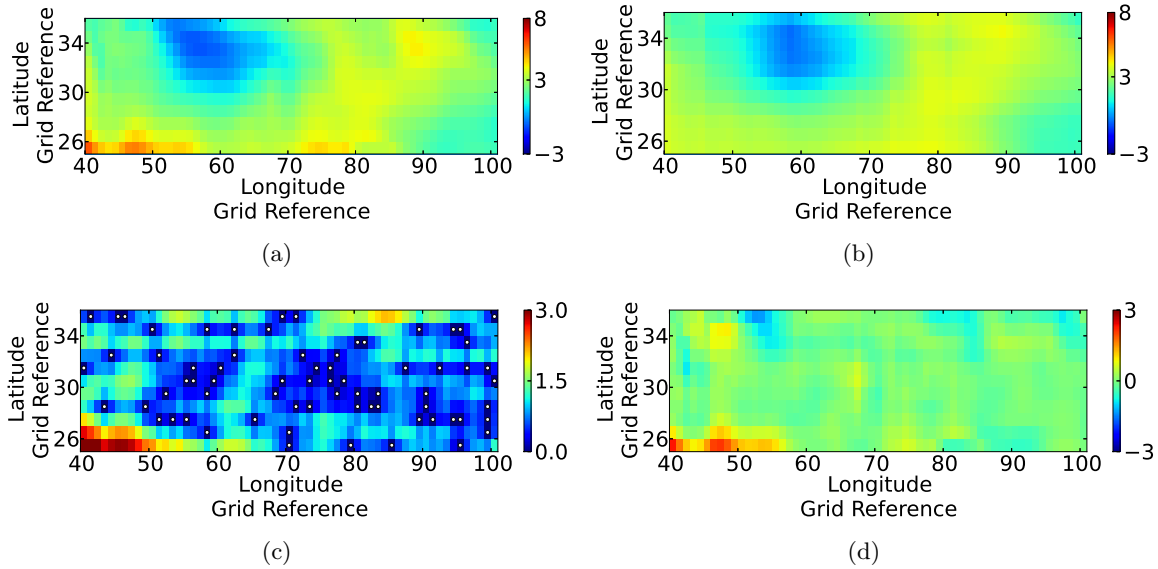


Figure 9.2: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for the worst performing set, set four, generated using Random Sampling. The white dots in (c) indicate the sampling points.

9.2(d). The location of the sample points is indicated with white dots found on the maps of estimated error variance (Figures 9.1(c) and 9.2(c)).

9.2 Limitations

It is acknowledged that this method of random selection excludes the last point in the data set from selection due to the nature of the random() and int() functions.

Only ten instances of Random Sampling were considered. Given the vast number of possible combinations, it is possible that if more instances had been considered, that the range of errors would have been larger.

CHAPTER 10

TWO-DIMENSIONAL SAMPLING - UPDATED KRIGING VARIANCE ALGORITHM

The Updated Kriging Variance Algorithm is applied as described in Section 4.3 for the addition of a single point at a time, with the only change being that it is applied for Simple Kriging on the residuals. The algorithm is stopped when 70 points are in the sample set. In addition, prediction is performed at each iteration in order to calculate errors. This method was implemented 10 times using 10 different randomly chosen starting sets. The starting sets for this method consisted of 4 points. These same starting points were used in the methods of Addition of Point at Point of Maximum Error Variance and Addition of Point at Point of Maximum Absolute Error found in Chapters 11 and 12 respectively.

The predictions performed made use of the method detailed in Chapter 8.

10.1 Results and Discussion

The errors associated with the predictions using the 70 points can be found in Tables 10.1 and 10.2. The units associated with the maximum absolute error, total absolute error and total integrated difference are mol(C)/m^{2}s.

Set	Maximum Absolute Error	Total Absolute Error	Percentage Absolute Error	Total Integrated Error
1	1.3216	115.0895	6.6398%	−6.7870
2	1.6273	108.0042	6.2311%	6.8906
3	1.4081	108.0028	6.2310%	−1.3657
4	1.8229	126.4596	7.2958%	−0.0612
5	1.7040	108.5669	6.2635%	8.1235
6	1.4878	114.2765	6.5929%	−24.6735
7	1.3809	105.4375	6.0830%	14.8140
8	1.6046	114.4195	6.6012%	−6.5234
9	0.9288	107.2994	6.1904%	7.6449
10	1.4300	107.2945	6.1901%	−10.2925

Table 10.1: Final maximum absolute, total absolute, percentage absolute and total integrated errors associated with Updated Kriging Variance Algorithm sampling in 2D.

Set	Maximum Estimated Error Variance	Total Estimated Error Variance
1	1.7657	409.3044
2	1.5228	409.8325
3	1.6568	408.6157
4	1.6382	408.9705
5	1.4785	413.4339
6	1.4660	405.4529
7	1.4395	408.0583
8	1.4044	407.4754
9	1.4393	409.6080
10	1.5134	410.5291

Table 10.2: Final maximum estimated error variances and total estimated error variances associated with Updated Kriging Variance Algorithm sampling in 2D.

Tables 10.1 and 10.2 illustrate that even when the total estimated error variance is at it's lowest for this method (set six), it does not guarantee that the total absolute error will be at it's lowest. Once again it appears as if the same total estimated error variance can produce different total absolute errors, as was explained in Chapter 9. It is noted, however, that the total estimated error variances for this method fall within a relatively small range.

Maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux, the estimated error variance and the true error across the chosen two-dimensional region for the best and worst performing sets (in terms of total absolute error minimisation) can be seen in Figures 10.1 and 10.2 respectively. Maps for the remaining sets can be seen in Appendix D.2.

Both Figures 10.1 and 10.2 show some similarities between the true and predicted flux maps although some visual differences are still distinctly seen. These differences also are apparent in the maps of true error (Figures 10.1(d) and 10.2(d) respectively). It can be seen from the estimate error variance maps where the sample points have been placed as they are indicated by the white dots. In both Figures 10.1(c) and 10.2(c), it can be seen that the points have been relatively evenly distributed over the region of interest. This occurs as to minimise estimated error variance. Hence, the distances between points that have not been sampled and the sampled points are minimised.

Additionally, Figure 10.3 shows how the errors progressed with the addition of points using the Updated Kriging Variance Algorithm across the best and worst performing sets. Figures of the progression of the errors across the algorithm for the remaining sets can be found in Appendix D.2.

From Figure 10.3(a) it can be seen that the addition of points to take the sample size from 4 to 70 results in a large drop in the maximum absolute error. It is important to note though, that this decrease is not smooth and the values do increase again with the addition of certain points. In fact, the end values are not the lowest values that are taken on throughout the algorithm. However, a lower maximum absolute error merely implies a lower upper bound value on the total absolute error and does not guarantee a lower total absolute error. If one considers Figure 10.3(b) in conjunction with Figure 10.3(a), it can be seen that increases and decreases in maximum absolute values do not always translate like for like in total absolute error.

Figure 10.3(b) also shows a dramatic decrease in total absolute error with the addition of the 66 points to the starting set of 4. However, once again this is not a smooth decrease. In

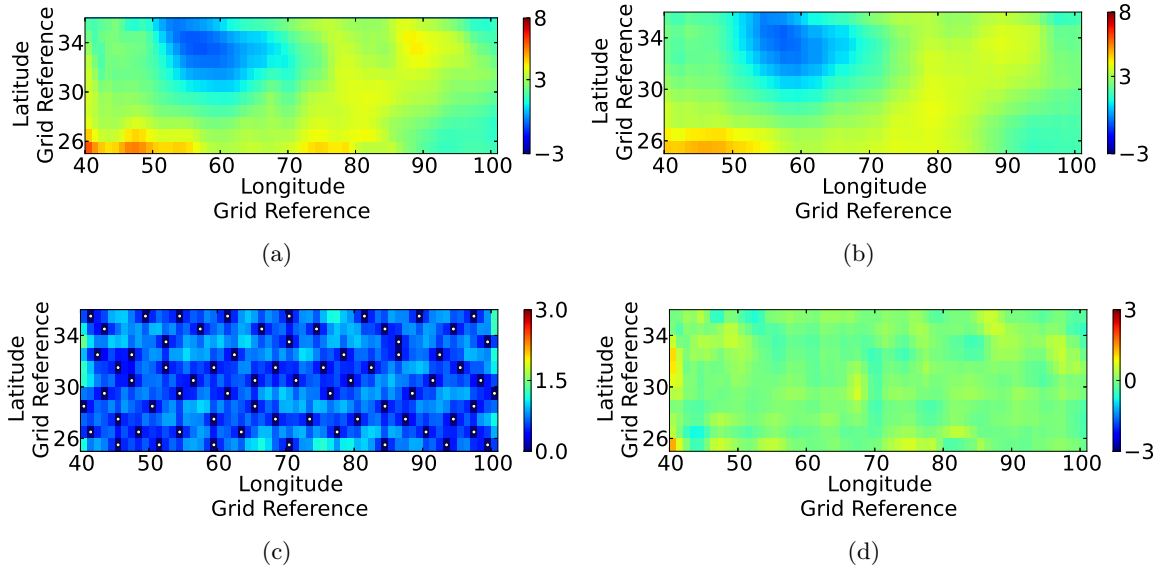


Figure 10.1: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for the best performing set, set seven, generated using Updated Kriging Variance Algorithm sampling. The white dots in (c) indicate the sampling points.

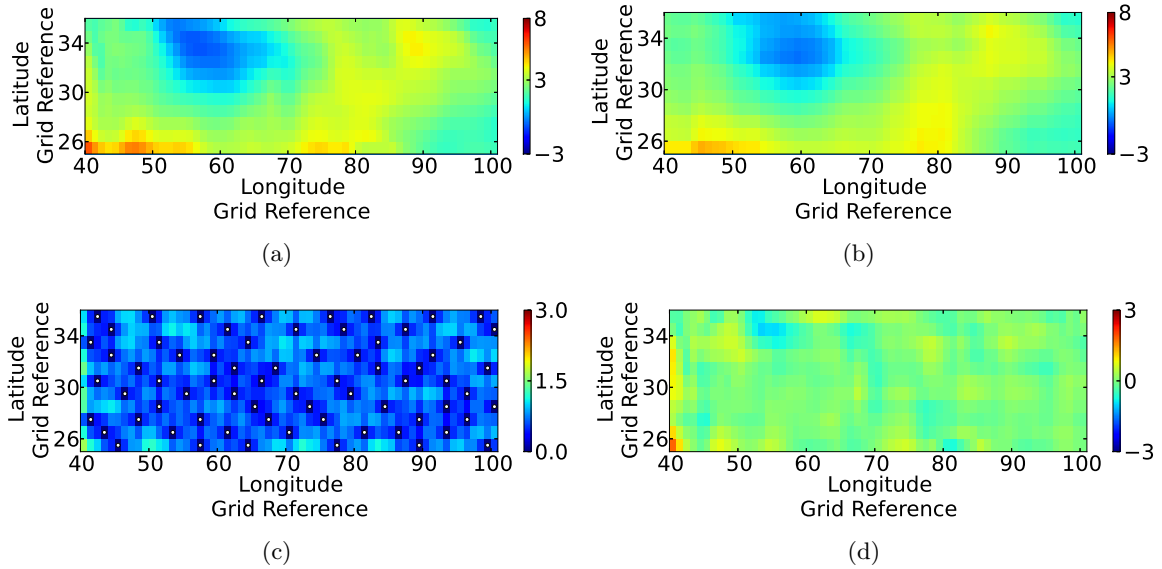


Figure 10.2: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for the worst performing set, set four, generated using Updated Kriging Variance Algorithm sampling. The white dots in (c) indicate the sampling points.

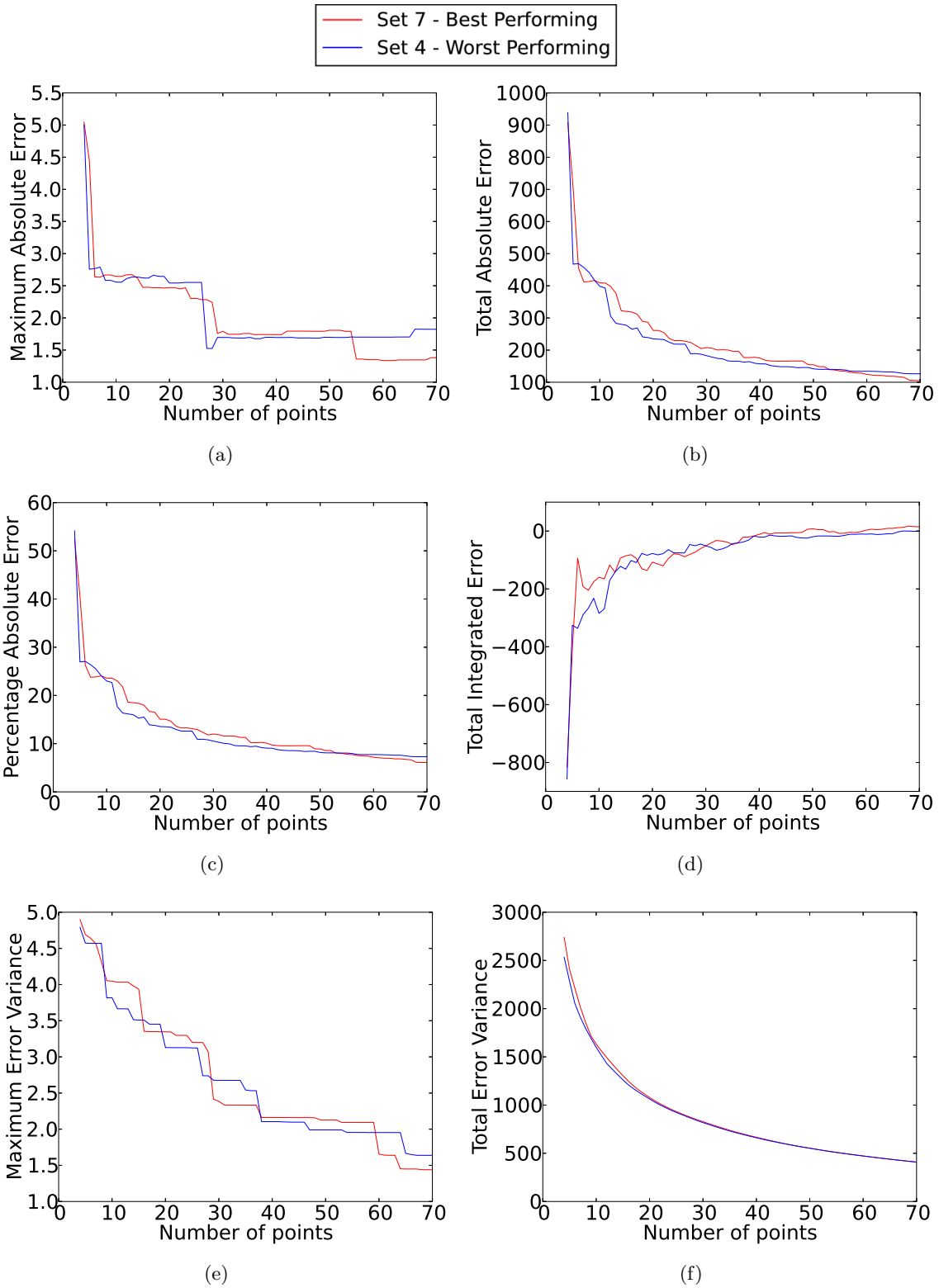


Figure 10.3: Progression of (a) maximum absolute error, (b) total absolute error, (c) percentage absolute error, (d) total integrated error, (e) maximum estimated error variance and (e) total estimated error variance across the best and worst performing sets generated using the Updated Kriging Variance Algorithm.

fact, if one considers the worst performing set, one can clearly see that there are times where the addition of a point actually increases the total absolute error (while the total estimated error variance does continue to decrease). However, over the total 66 additions, the overall behaviour is definitely decreasing.

Figure 10.3(c) illustrates the percentage absolute error as it changes over the addition of points. This follows the same pattern as the total absolute error. This figure does however show that even the worst performing set of the ten generated still has an error of below 10% with a sample size of only 70 points. If one considers Table 10.1 it can be seen that this error is in fact just less than 7.3%. It is also important to note that the percentage absolute error with the initial random sample of 4 points is somewhere in the region of 53 - 55% and thus, there is a very large decrease in percentage absolute error across this algorithm.

The total integrated error does undergo a large improvement across the algorithm (Figure 10.3(d)). The total absolute error provides a upper bound for the total integrated error. However, as for reason explained in Chapter 9 it is not a good indicator of overall performance.

Figure 10.3(e) shows the dramatic yet non-smooth decrease in maximum estimated error variance associated with this algorithm, while Figure 10.3(f) shows the smooth decrease in total estimated error variance which this algorithm provides. This smooth decrease is expected as the algorithm tests all possible additions for the highest decrease before adding the point which provides that decrease. From Table 10.2 and Figure 10.3(f), it can be seen that these two sets produce very similar total estimated error variance values. Yet, they produce the best and worst total absolute errors which translates to a percentage absolute error difference of around 1.2%.

10.2 Limitations

Only ten instances of this algorithm have been run. The starting set consists of 4 points and it is possible that these 4 points would not belong in the truly optimal set of 70 points. The conclusions and discussions drawn up about this algorithm are based solely on the ten sets which have been produced for this study.

CHAPTER 11

TWO-DIMENSIONAL SAMPLING - ADDITION OF A POINT AT POINT OF MAXIMUM ERROR VARIANCE

This method was performed as described in Section 4.4, with the exception of the stopping criterion. Instead of stopping when a lower bound for the maximum estimated error is reached, the method is stopped when a sample size of 70 is reached. This algorithm is summarised as:

Step 1: Selection of an initial set of points

Step 2: Calculation of values of variable of interest and estimated error variance by the method of Kriging

Step 3: Find the location and the value of the largest estimated error variance

Step 4: If the sample size is less than a pre-specified number, add this point to the sample set and return to step 2, else stop the algorithm.

The Kriging used within this sampling method is Simple Kriging on the residuals. For any given sample set size, predictions are performed using the method described in Chapter 8.

11.1 Results and Discussion

The errors associated with the predictions using the 70 points can be found in Tables 11.1 and 11.2. The units associated with the maximum absolute error, total absolute error and total integrated difference are mol(C)/m^{2}s.

Table 11.2 indicates that for this method, the total estimated error variance values fall within a relatively small range. However, once again the lowest total estimated error variance does not guarantee the lowest total absolute error. If Table 11.1 is also considered it can be seen that the set with the highest total estimated error variance, set eight, also has the lowest total absolute error.

From Table 11.1 it can also be seen that all percentage absolute errors fall within the range of 5.6 to 6.7%.

Maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux, the estimated error variance and true error across the chosen two-dimensional region for the

Set	Maximum Absolute Error	Total Absolute Error	Percentage Absolute Error	Total Integrated Error
1	1.3485	109.2782	6.3046%	−4.5915
2	0.8540	106.9973	6.1730%	−4.4147
3	0.9016	113.3509	6.5395%	−2.6408
4	1.0593	99.9819	5.7682%	4.5817
5	1.3766	109.4288	6.3132%	−2.3647
6	0.9336	113.5421	6.5506%	−27.2000
7	0.8161	108.3488	6.2509%	4.0762
8	0.7439	97.6324	5.6327%	−12.4819
9	0.7841	115.5823	6.6683%	−2.6802
10	1.1838	111.2495	6.4183%	8.3153

Table 11.1: Final maximum absolute, total absolute, percentage absolute and total integrated errors associated with Addition of a Point at Point of Maximum Estimated Error Variance sampling in 2D,

Set	Maximum Estimated Error Variance	Total Estimated Error Variance
1	1.1180	411.5144
2	1.1274	417.9051
3	1.0897	415.8471
4	1.1224	414.8858
5	1.0791	414.1715
6	1.1016	412.2581
7	1.1357	416.7959
8	1.1330	418.2989
9	1.1149	411.5168
10	1.1250	418.0174

Table 11.2: Final maximum estimated error variances and total estimated error variances associated with Addition of a Point at Point of Maximum Estimated Error Variance sampling in 2D.

best and worst performing sets (in terms of total absolute error minimisation) can be seen in Figures 11.1 and 11.2 respectively. The maps for the remaining sets can be found in Appendix D.3.

Figure 11.1 shows some similarities between the predicted and true carbon dioxide flux maps, while in Figure 11.2 the similarities are slightly less clear. The points are fairly evenly distributed over the area of interest and this is due to the use of the maximum estimated error variance as point of placement of the next point. The location of the sample points in the two sets is indicated by white dots in Figures 11.1(c) and 11.2(c).

Figures 11.3 shows how the errors progressed with the addition of points using the method of Addition of a Point at Point of Maximum Estimated Error Variance across the best and worst performing sets. The figures illustrating the progression of the errors for the remaining sets can be found in Appendix D.3.

From Figure 11.3(a) the maximum absolute errors do not exhibit a smooth decrease. The addition of a point at times causing an increase in the maximum absolute error. However, the overall drop in maximum absolute error across the algorithm is significant.

The decrease in total absolute error across the algorithm is not smooth as can be seen in

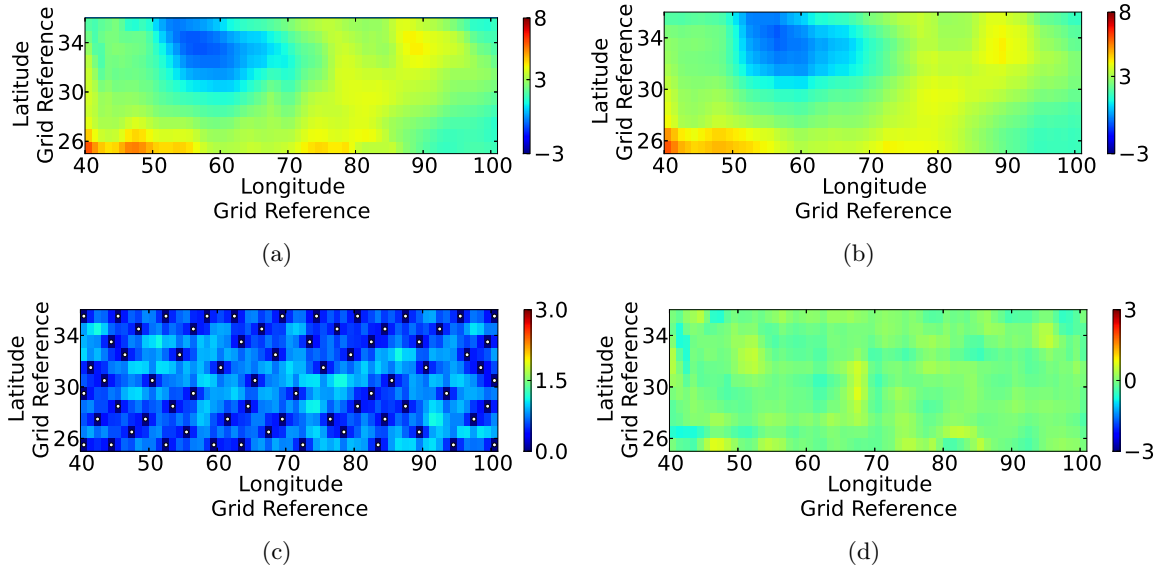


Figure 11.1: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for the best performing set, set eight, generated using Addition of a Point at Point of Maximum Estimated Error Variance sampling. The white dots in (c) indicate the sampling points.

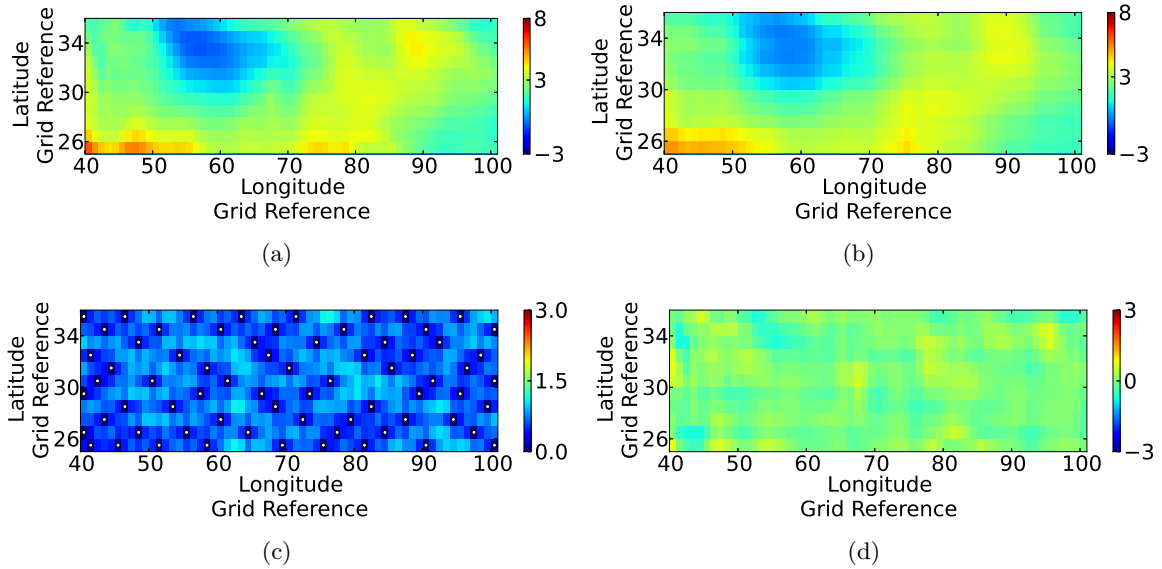


Figure 11.2: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for the worst performing set, set nine, generated using Addition of a Point at Point of Maximum Estimated Error Variance sampling. The white dots in (c) indicate the sampling points.

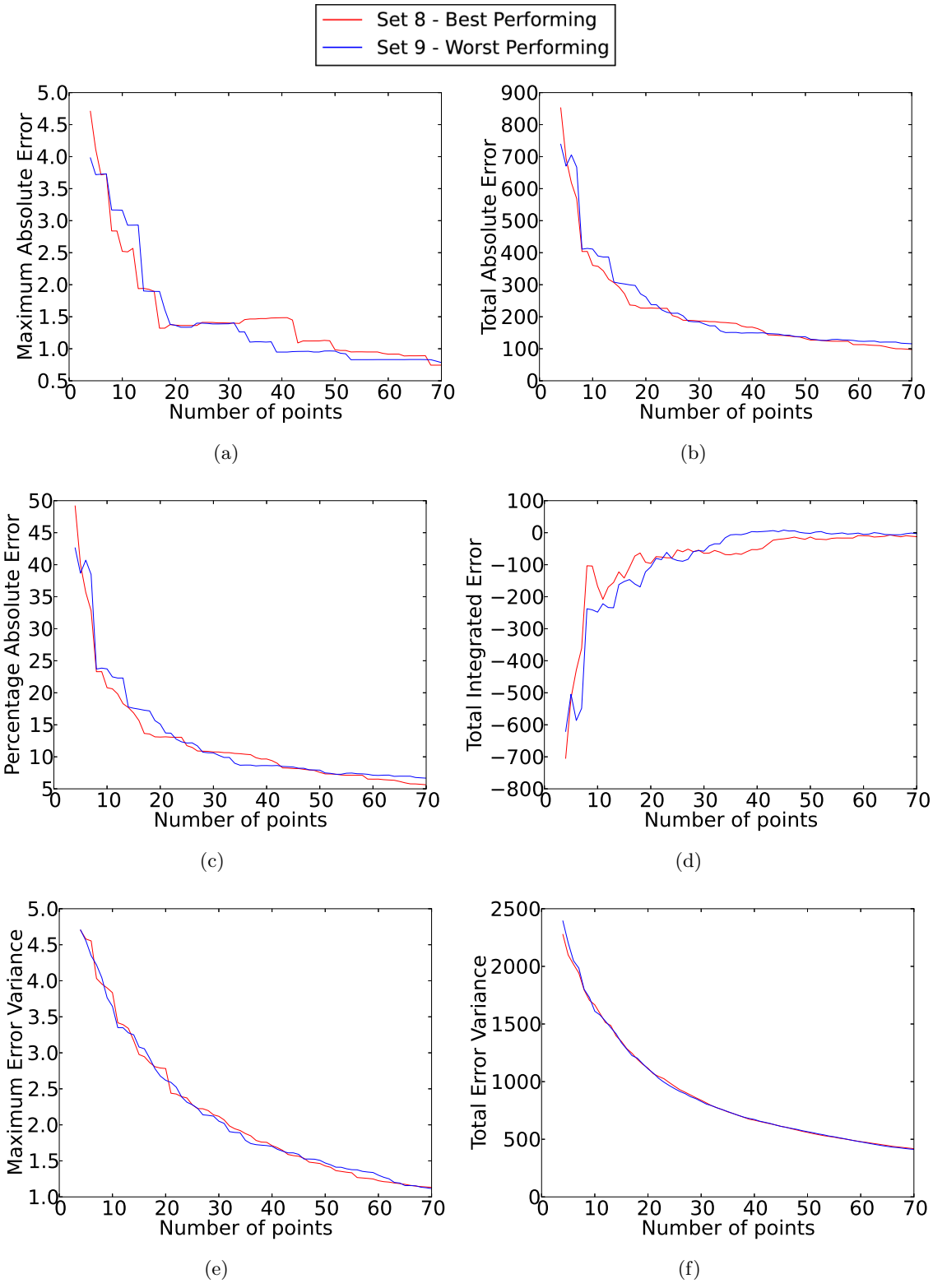


Figure 11.3: Progression of (a) maximum absolute error, (b) total absolute error, (c) percentage absolute error, (d) total integrated error, (e) maximum estimated error variance and (e) total estimated error variance across the best and worst performing sets generated using the Addition of a Point at Point of Maximum Estimated Error Variance.

Figure 11.3(b). However, over the course of the algorithm, a large drop in the total absolute error is seen. From Figure 11.3(c) it is clear that the drop in percentage absolute error is large. However, the best and worst sets are not always the best and worst sets respectively across the algorithm. Although these sets end with the highest and lowest total absolute errors, there are periods in the algorithm (from just before 30 points to just after 40 points) where the worst performing set outperforms the best performing set.

Figure 11.3(d) illustrates once again why the total integrated error is not considered a good indication of performance as the worst performing set has a total integrated error nearer to zero than the best performing set.

Illustrated in Figure 11.3(e) is the non smooth decrease in maximum estimated error variance which occurs when the point added is placed at the point of maximum estimated error variance. The total estimated error variance undergoes a non-smooth decrease as the number of points is increased (illustrated in Figure 11.3(f)) following the algorithm. The curve appears to become more smooth as the number of points increases.

11.2 Limitations

Only ten sets are considered. They are considered using only the trend and covariance structure chosen earlier in this study. All conclusions are based on only the ten sets generated using this method.

CHAPTER 12

TWO-DIMENSIONAL SAMPLING - ADDITION OF A POINT AT POINT OF MAXIMUM ABSOLUTE ERROR

This method is very similar to the method in Chapter 11. The algorithm for this method is presented below:

Step 1: Selection of an initial set of points

Step 2: Calculation of values of variable of interest and estimated error variance by the method of Kriging

Step 3: Find the location and the value of the largest absolute error

Step 4: If the sample size is less than a pre-specified number, add this point to the sample set and return to step 2, else stop the algorithm.

The Kriging used in this algorithm is Simple Kriging of the residuals. Once again, for any given sample set size, predictions are performed using the method described in Chapter 8.

12.1 Results and Discussion

The errors associated with the predictions using the 70 points can be found in Tables 12.1 and 12.2. The units associated with the maximum absolute error, total absolute error and total integrated difference are mol(C)/m^{2}s.

Table 12.1 shows the all the sets achieve a percentage absolute error between 3.4 and 4.2%. This table also shows that this method results in a very low maximum absolute error while Table 12.2 shows that in spite of the low total absolute error, the total estimated error variance is quite high. Considering sets five and seven shows once again that very similar total absolute errors can exhibit very different total estimated error variances, as has been explained in Chapter 9.

Maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux and the estimated error variance across the chosen two-dimensional region for the best and worst performing sets (in terms of total absolute error minimisation) can be seen in Figures 12.1 and 12.2 respectively. Maps for the remaining sets are given in Appendix D.4.

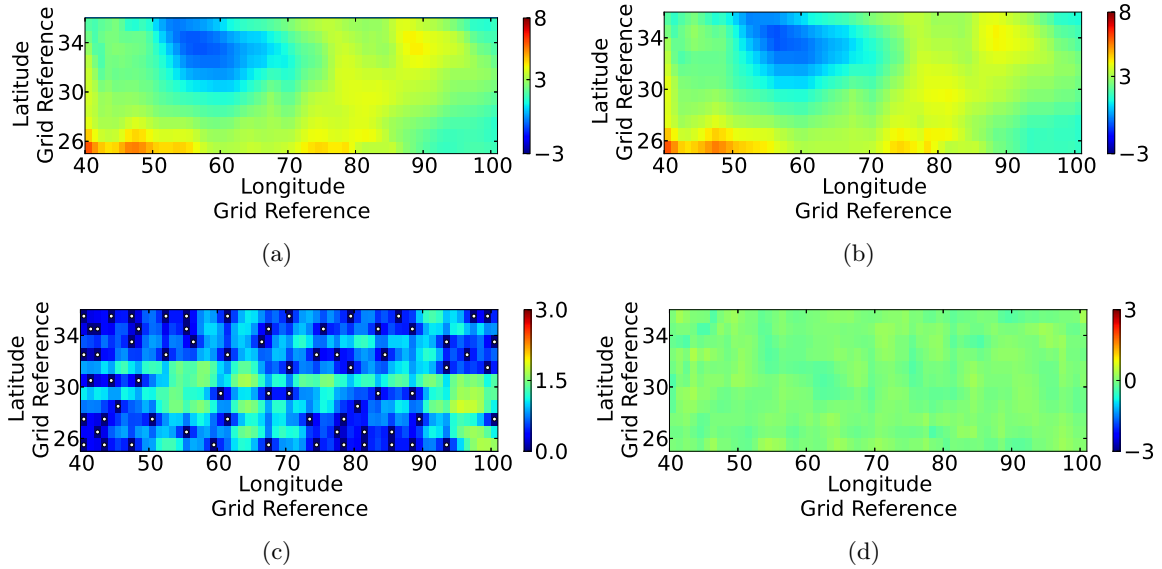


Figure 12.1: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for the best performing set, set eight, generated using Addition of a Point at Point of Maximum Absolute Error sampling. The white dots in (c) indicate the sampling points.

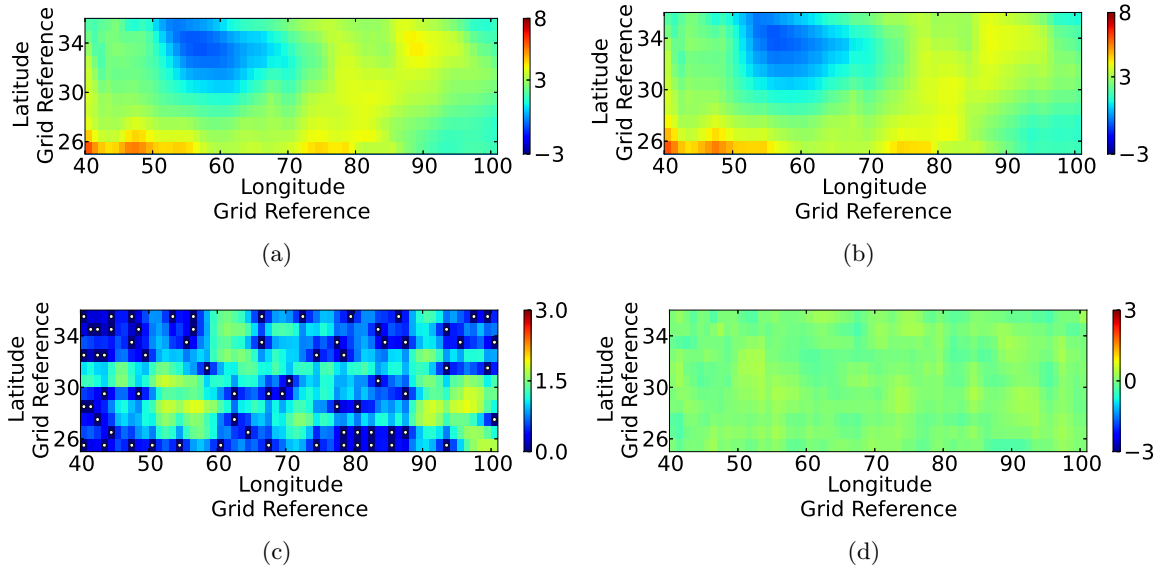


Figure 12.2: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for the worst performing set, set six, generated using Addition of a Point at Point of Maximum Absolute Error sampling. The white dots in (c) indicate the sampling points.

Set	Maximum Absolute Error	Total Absolute Error	Percentage Absolute Error	Total Integrated Error
1	0.4777	70.4430	4.0640%	−20.9385
2	0.3752	70.5894	4.0725%	−1.05180
3	0.3966	66.1099	3.8141%	1.8664
4	0.4248	69.6738	4.0197%	−19.3797
5	0.3575	64.3786	3.7142%	6.2411
6	0.3647	72.0442	4.1564%	16.6088
7	0.3932	62.8166	3.6241%	−11.8342
8	0.3712	58.4807	3.3739%	1.5703
9	0.3969	69.1577	3.9899%	10.7022
10	0.4017	63.4083	3.6582%	−0.1515

Table 12.1: Final maximum absolute, total absolute, percentage absolute and total integrated errors associated with Addition of a Point at Point of Maximum Absolute Error sampling in 2D.

Set	Maximum Estimated Error Variance	Total Estimated Error Variance
1	2.0006	554.7707
2	2.0937	544.9054
3	1.9079	531.8254
4	1.8729	542.6504
5	2.0711	566.5581
6	1.8971	564.4201
7	1.7525	512.4042
8	1.8085	512.3191
9	1.8090	546.9972
10	1.9031	532.5952

Table 12.2: Final maximum estimated error variances and total estimated error variances associated with Addition of a Point at Point of Maximum Absolute Error sampling in 2D.

Both Figures 12.2 and 12.1 show very strong similarities between the true and predicted flux maps although some differences can still be found visually. Of interest in these figures is the location of the sample points as can be seen from the white dots on Figures 12.1(c) and 12.2(c). It appears as if this method is not providing as even a distribution as the previous methods, but is rather placing fewer points where there is less variability and more points where the flux is more variable. Placing more points where there is more variability allows for more of this variability to be captured by the Kriging predictions and thus, the errors are lower.

Additionally, Figure 12.3 shows how the errors progressed with the additional of points using the method of Addition of a Point at Point of Maximum Absolute Error across the above mentioned two sets. Figures showing the progression of these errors over the remaining sets can be seen in Appendix D.4.

Figure 12.3(a) illustrates that in spite of the fact that the point is added at the point of maximum absolute error, there is occasionally an increase in the maximum absolute error. This maximum absolute error does, however, decrease to a very low value over the course of the algorithm. Shown in Figure 12.3(b) is a similar trend for total absolute error. The addition of a point at the point of maximum absolute error does not guarantee a decrease in total absolute error. While the addition of a point does sometimes cause an increase in

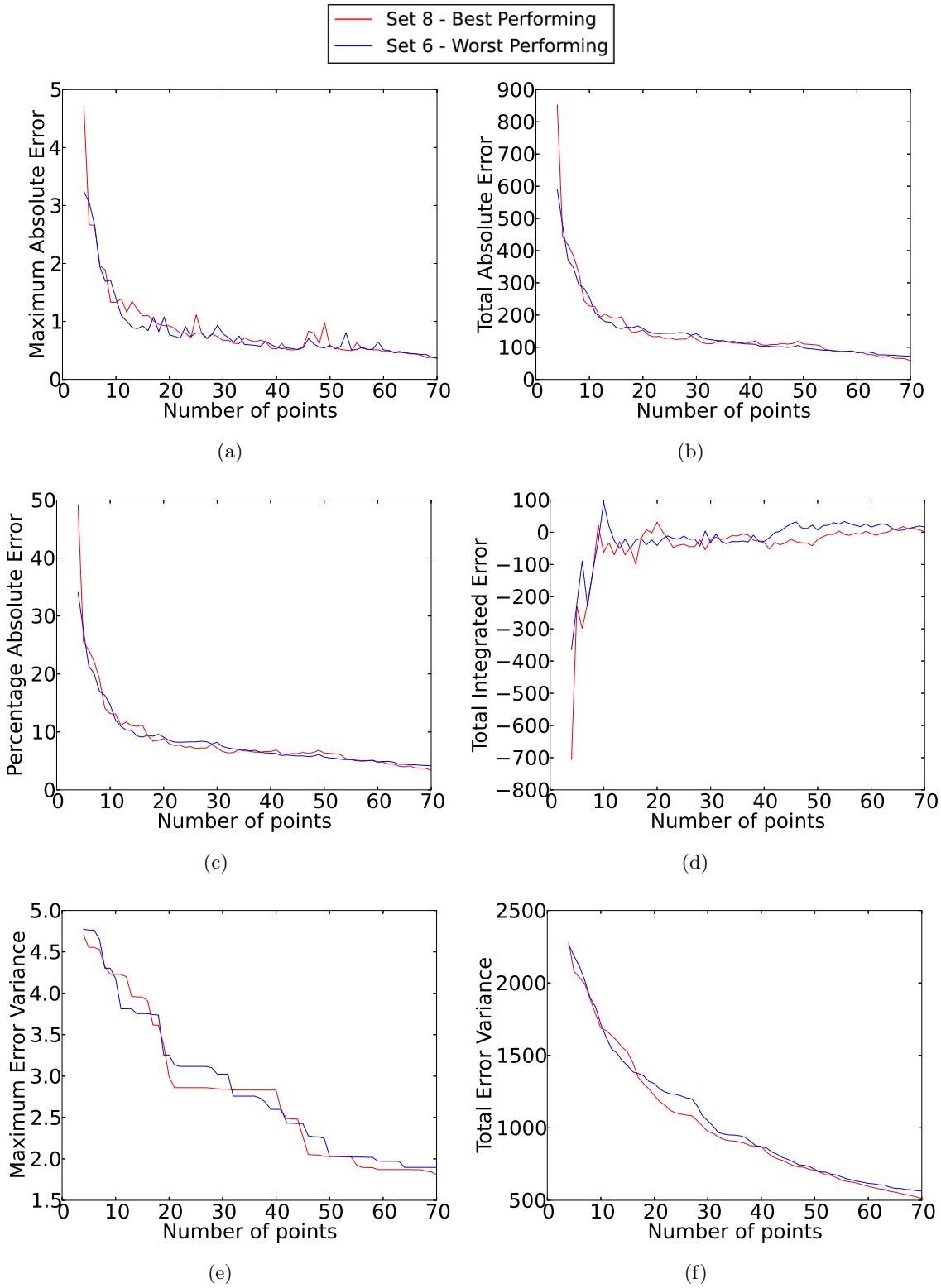


Figure 12.3: Progression of (a) maximum absolute error, (b) total absolute error, (c) percentage absolute error, (d) total integrated error, (e) maximum estimated error variance and (e) total estimated error variance across the best and worst performing sets generated using the Addition of a Point at Point of Maximum Absolute Error.

the total absolute error, the total absolute error decreases significantly over the algorithm for both the best and worst performing sets.

Figure 12.3(c) which gives the change in the percentage absolute error over the course of the algorithm naturally reflects the trend shown in Figure 12.3(b). While 12.3(d) indicates how the total integrated error changes across the algorithm and in these sets ends close to zero.

Figure 12.3(e) shows the non-smooth decrease in maximum estimated error variance. This graphs shows periods where the maximum error variance appears relatively stable. Figure 12.3(f) demonstrates the non smooth decrease seen in total estimated error variance due to this method. This figure also shows that even when the maximum estimated error variance remains quite stable, the total estimated error variance can experience changes. This can occur as the maximum estimated error variance provides an indication only of the upper bound for total estimated error variance.

12.2 Limitations

The method was applied to sample only up to 10.5% of the data. All conclusions were drawn based only on the 10 sets generated using this method. Only one covariance structure and trend were applied.

CHAPTER 13

TWO-DIMENSIONAL SAMPLING - GENETIC ALGORITHMS

In this section, the genetic algorithms used in this study are discussed and the results obtained from them presented. The genetic algorithms used all made use the of the same genetic features apart from how the initial population is selected (two of the six use a solution obtained from another method as an individual in the starting set, referred to as the hybrid methods in this study), the fitness function and the convergence criterion. Two of the genetic algorithms try to find the worst sample set, while four (including the two hybrid methods) try to find the optimal sample set. Two different fitness functions are considered, the sum of the absolute error and the sum of the estimated error variance. Each of the six genetic algorithms tested were run 5 times only. This is due to the computational time required in the running of a genetic algorithm in comparison to the other sampling strategies.

The common genetic features are discussed next and this is followed by a section on each genetic algorithm used in which details for the choice of initial population, fitness function and convergence criterion are presented. The results for each genetic algorithm are both presented and discussed in their respective sections.

13.1 Encoding

Once again, a sample set of 70 distinct points is desired. As a result, the encoding chosen uses integers in the range of 0 to 670 to represent each point. Each individual consists of a string of 70 integers from the given range. No repetition of integers is allowed within an individual. Each individual represents a single sample set and the terms individual and set will be used interchangeably. Thus, in order to avoid confusion, the final sample set generated will be referred to as a run (as opposed to the term set which has been used in previous chapters).

13.2 Selection

Selection for reproduction across all six genetic algorithms takes place by means of tournament selection between two individuals. The two individuals are selected randomly from the possible twenty without replacement. Only one of these is added to the reproduction pool based on the criteria discussed below. This process is completed twice, allowing for twenty individuals to be placed into the reproduction pool. Individuals may be selected for the reproduction pool once, twice or not at all. Each individual competes in two tournaments.

In the case of the four genetic algorithms attempting to select the optimal sample set, the individual with the lower fitness function is chosen, as we wish to minimise the errors. For the two genetic algorithms attempting to generate the worst performing sample set, the individual with the higher fitness function is chosen.

13.3 Reproduction

Single point crossover is the reproduction method across all six genetic algorithms. The crossover point is randomly chosen for every reproduction. The first two individuals placed into the reproduction pool are paired together, the next two together and so forth.

13.4 Mutation

Every new individual (child) is subject to the mutation section of the algorithm. For each integer within the individual's string, a random number between 0 and 1 is generated. If this number is less than the mutation probability, then that integer is exchanged for a randomly generated one between 0 and 670. The mutation probability for all six genetic algorithm was chosen to be 0.001.

13.5 Replacement

In all six genetic algorithms, all 20 new individuals (children) replace all 20 individuals of the previous generation.

13.6 Replacement of Duplicate Points within an Individual

No repetition of points within a sample set is allowed in order to ensure positive-definiteness of the covariance matrix. As such, the new generation of individuals must be checked for repeated integers within an individual. If an integer is found to occur more than once, the latter occurrences are replaced with randomly chosen integers which are not already present within the individual.

13.7 Elitism

The individual with the highest fitness (when aiming to find the worst performing sample set) or the lowest fitness (when attempting to find the optimal sample set) is carried through to the next generation. This individual replaces a randomly chosen individual of the new generation. This occurs across all six genetic algorithms implemented.

13.8 Genetic Algorithm designed to Minimise Total Absolute Error

13.8.1 Selection of Starting Population

The starting population consists of 20 individuals. Each individual is randomly chosen without repetition and consists of 70 points.

13.8.2 Evaluation - Fitness Function

The evaluation process calculates the sum of the absolute errors from prediction using the individual/sample set. This is used in conjunction with a lower fitness function being considered more optimal.

13.8.3 Convergence Criterion

A maximum number of iterations of 30000 is chosen as the convergence criterion. Each iteration leads to a new generation of individuals. If the initial generation of individuals is referred to as generation zero, then the convergence criterion could also be expressed as a maximum number of generations of 30000.

13.8.4 Results and Discussion

The errors associated with the predictions using the 70 points can be found in Tables 13.1 and 13.2. The units associated with the maximum absolute error, total absolute error and total integrated difference are mol(C)/m^2s.

Run	Maximum Absolute Error	Total Absolute Error	Percentage Absolute Error	Total Integrated Error
1	0.8338	51.9996	2.1000%	−0.6576
2	0.8220	50.9861	2.9415%	−6.7661
3	0.5791	53.8685	3.1078%	0.5883
4	0.5522	56.7306	3.2729%	−3.8904
5	0.5790	60.2845	3.4780%	6.1781

Table 13.1: Final maximum absolute, total absolute, percentage absolute and total integrated errors associated with the Genetic Algorithm designed to Minimise Total Absolute Error in 2D.

From Tables 13.1 and 13.2, it can be seen that the percentage absolute errors from across the 5 runs all fall within the small range of 2.94-3.48%. They also show the low total absolute errors obtained by the runs. These low total absolute errors result in a low upper bound on the total integrated error which the total integrated errors fall well within. The maximum absolute errors are also quite low, all falling below 0.85. The total estimated error variances all fall above 500 with relatively high maximum estimated error variances.

Maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux,the estimated error variance and the true error across the chosen two-dimensional region associated with the best and worst performing runs (in terms of total absolute error minimisation)

Run	Maximum Estimated Error Variance	Total Estimated Error Variance
1	2.0092	522.4028
2	2.1127	531.7697
3	2.3053	539.4515
4	2.0056	506.9597
5	1.8584	523.0464

Table 13.2: Final maximum estimated error variances and total estimated error variances associated with the Genetic Algorithm designed to Minimise Total Absolute Error in 2D.

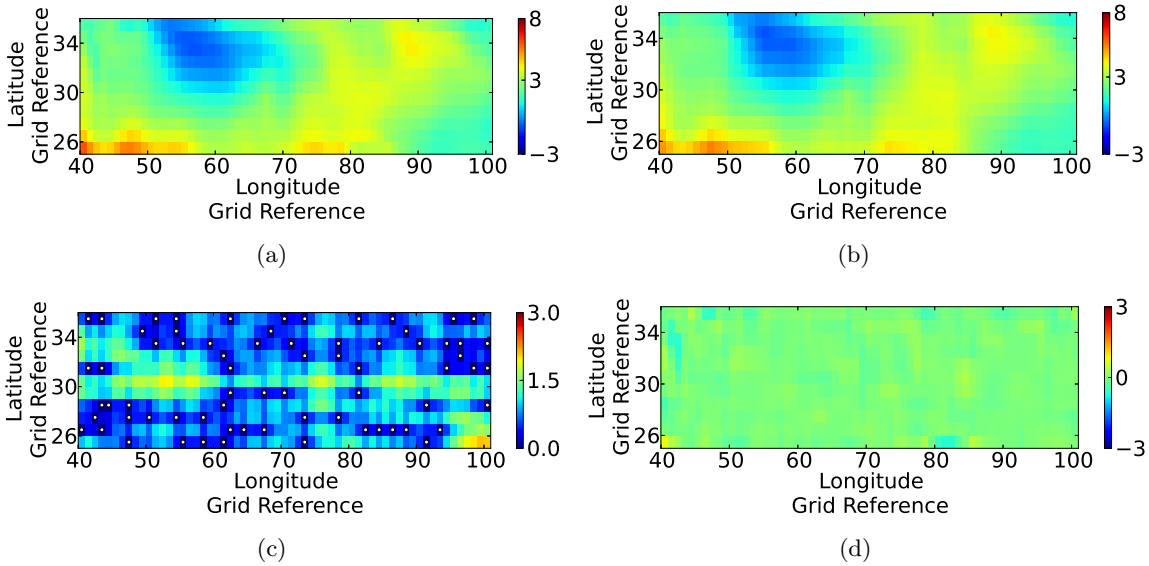


Figure 13.1: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for the best performing run, run two, generated using the Genetic Algorithm designed to Minimise Total Absolute Error. The white dots in (c) indicate the sampling points.

can be seen in Figures 13.1 and 13.2 respectively. The maps for the remaining runs can be seen in Appendix D.5.

Considering Figures 13.1 and 13.2, it can be see the maps of the true flux and predicted flux are visually very similar. There are still a few visual differences but the predictions and true values appear to compare favourably. The location of the sample points is indicated by the white dots on the maps of the estimated error variance (Figures 13.1(c) and 13.2(c)).

Additionally, Figure 13.3 shows how the errors progressed across the best and worst performing runs as the number of iterations increased. The progression of these errors across the remaining runs can be seen in Appendix D.5.

Figure 13.3(a) illustrates that the maximum error variance become relatively stable after 10000 iterations, while Figure 13.3(b) shows the step decreases in total absolute error. These step decreases are expected as this genetic algorithm is designed to minimise total absolute error while carrying the best solution through from one iteration to the next. The worst performing run's total absolute error does not undergo much improvement from roughly 17500 iterations, while the best performing run is still showing good improvement at relatively short intervals until very near the end of the 30000 iterations. Naturally the changes in percentage absolute error found in Figure 13.3(c) mimic those of the total absolute error.

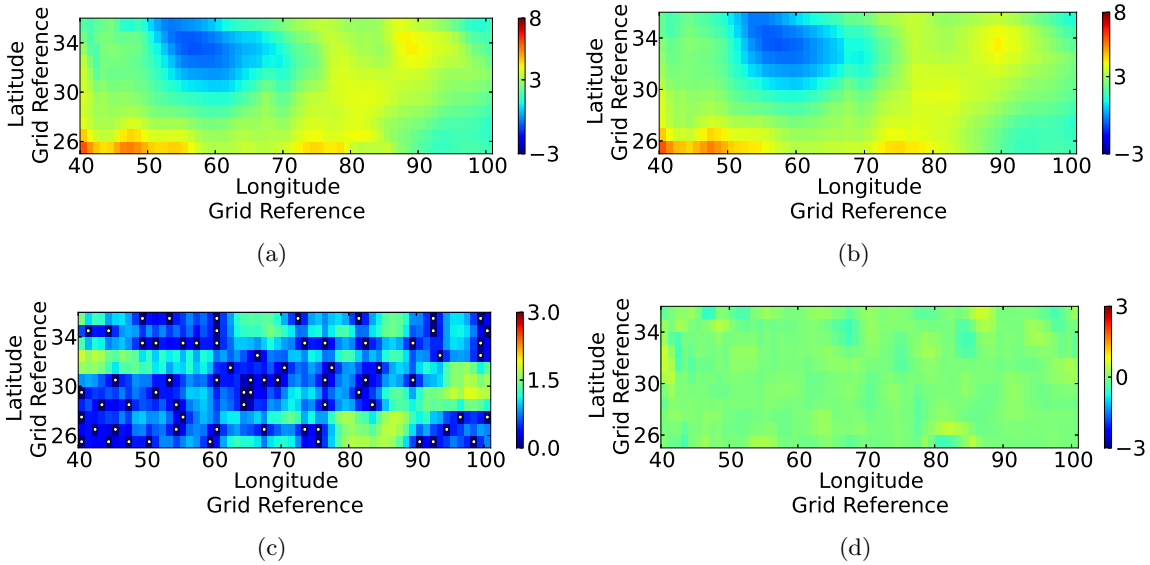


Figure 13.2: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for the worst performing run, run five, generated using the Genetic Algorithm designed to Minimise Total Absolute Error. The white dots in (c) indicate the sampling points.

Figure 13.3(d) shows that even with a decreasing total absolute error, the final total integrated error does not approach zero. The maximum estimated error variance in both cases becomes stable for at least the last 5000 iterations as can be seen in Figure 13.3(e). In Figures 13.3(a) and 13.3(e), it can be seen that the best performing set has a higher maximum absolute error and maximum estimated error variance. In fact, if Figure 13.3(f) is considered, it can be seen that the best performing set which has the lowest total absolute error of the 5 runs actually has a higher total estimated error variance than the worst performing run. If one further considers this figure along with Figure 13.3(b), it can be seen that as the total absolute error decreases, the total estimated error variance can increase, decrease or remain stable.

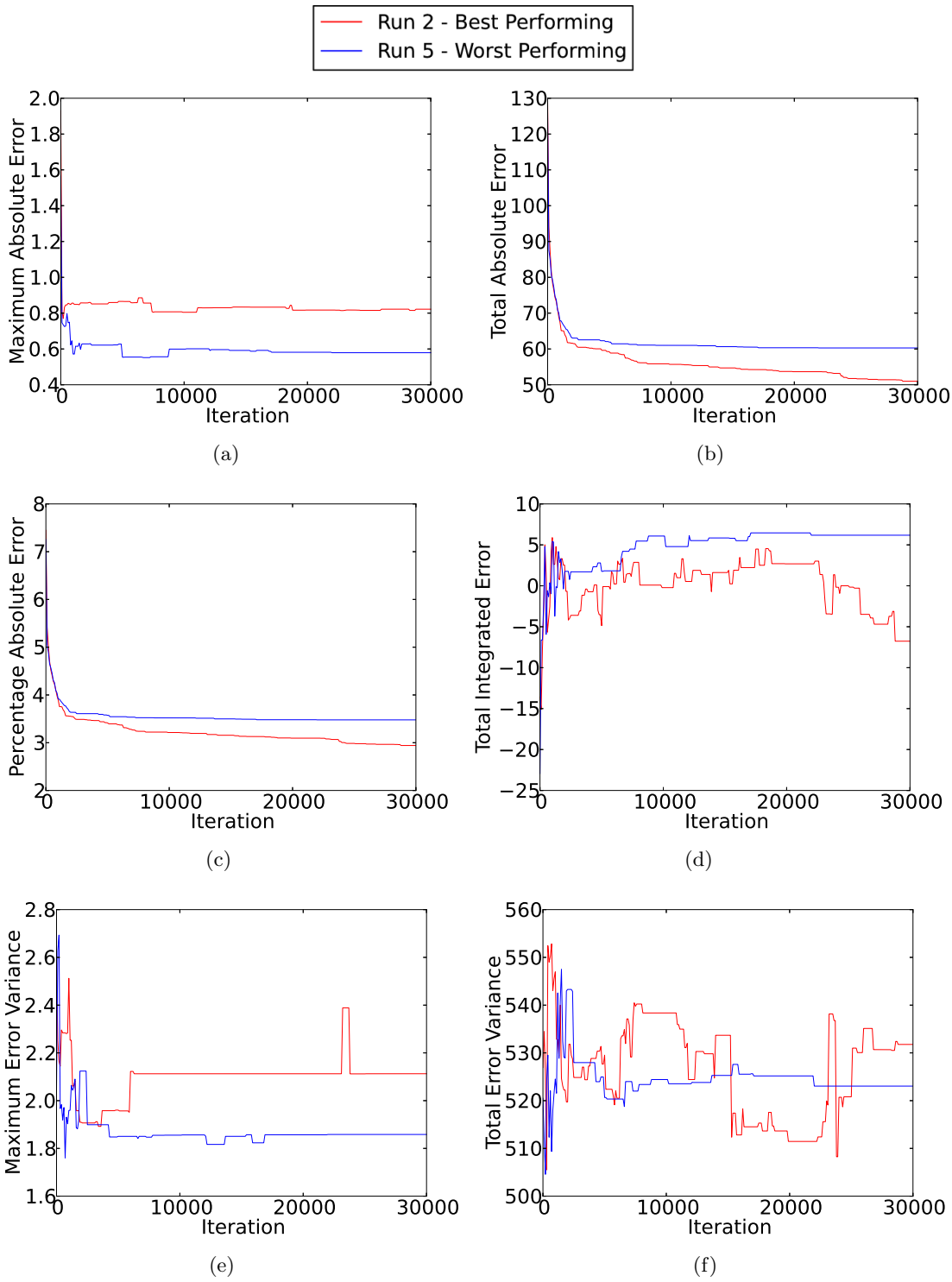


Figure 13.3: Progression of (a) maximum absolute error, (b) total absolute error, (c) percentage absolute error, (d) total integrated error, (e) maximum estimated error variance and (e) total estimated error variance across the best and worst performing runs of the Genetic Algorithm designed to Minimise Total Absolute Error.

13.9 Genetic Algorithm designed to Minimise Total Estimated Error Variance

13.9.1 Selection of Starting Population

The starting population consists of 20 individuals. Each individual is randomly chosen without repetition and consists of 70 points.

13.9.2 Evaluation - Fitness Function

The evaluation process calculates the sum of the estimated error variances from prediction using the individual/ sample set. This is used in conjunction with a lower fitness function being considered more optimal.

13.9.3 Convergence Criterion

A maximum number of iterations of 30000 is chosen as the convergence criterion.

13.9.4 Results and Discussion

The errors associated with the predictions using the 70 points can be found in Tables 13.3 and 13.4. The units associated with the maximum absolute error, total absolute error and total integrated difference are mol(C)/m^{2}s.

Run	Maximum Absolute Error	Total Absolute Error	Percentage Absolute Error	Total Integrated Error
1	1.3294	103.6156	5.9779%	−2.5667
2	0.9616	103.8526	5.9915%	−2.2675
3	0.9990	107.6070	6.2081%	3.6387
4	1.2127	103.7934	5.9881%	−15.7027
5	0.9286	105.9694	6.1137%	−18.4238

Table 13.3: Final maximum absolute, total absolute, percentage absolute and total integrated errors associated with the Genetic Algorithm designed to Minimise Total Estimated Error Variance in 2D.

Run	Maximum Estimated Error Variance	Total Estimated Error Variance
1	1.2762	401.9557
2	1.6344	401.5645
3	1.4355	400.9536
4	1.4551	403.9386
5	1.2369	403.0735

Table 13.4: Final maximum estimated error variances and total estimated error variances associated with the Genetic Algorithm designed to Minimise Total Estimated Error Variance in 2D.

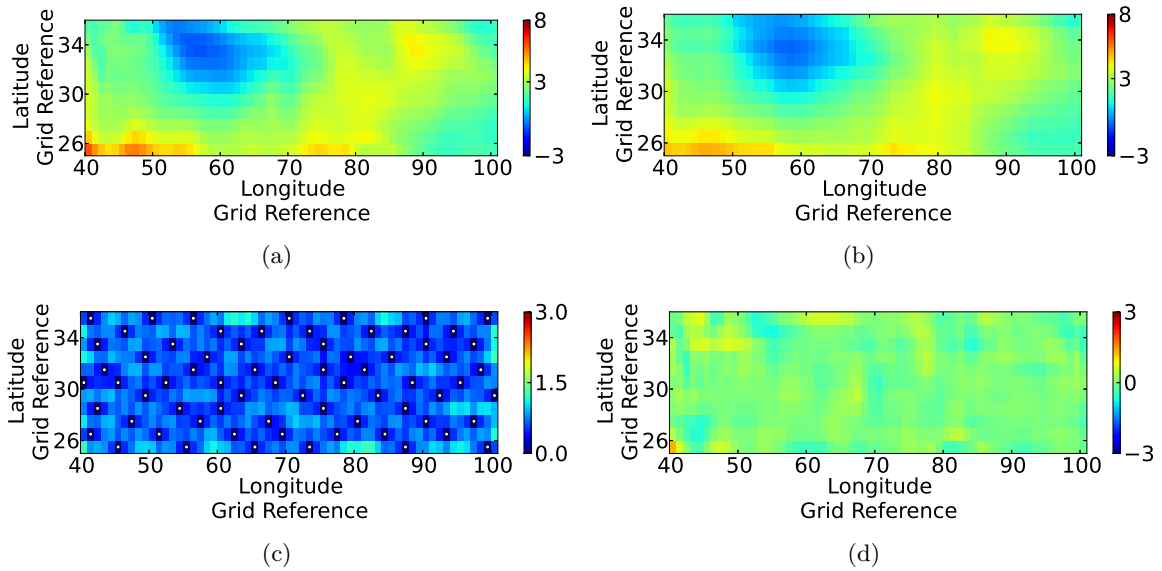


Figure 13.4: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for the best performing run, run one, generated using the Genetic Algorithm designed to Minimise Total Estimated Error Variance. The white dots in (c) indicate the sampling points.

From Tables 13.3 and 13.4 it can be seen that all percentage absolute errors are found in a small range between 5.97 and 6.21%. Total estimated error variance also falls within a small range of $400 - 404$.

Maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux, the estimated error variance and the true error across the chosen two-dimensional region associated with the best and worst performing runs (in terms of total absolute error minimisation) can be seen in Figures 13.4 and 13.5 respectively. The maps for the remaining runs can be seen in Appendix D.6.

Figures 13.4 and 13.5 both show visual differences between the predicted flux and true flux maps. The estimated error variance maps (Figures 13.4(c) and 13.5(c)) show a relatively evenly distribution of sample points. The sample point locations are indicated by means of the white dots.

Additionally, Figure 13.6 shows how the errors progressed across the best and worst performing runs as the number of iterations increased. The progression of these errors across the remaining runs can be seen in Appendix D.6.

Figure 13.6(a) demonstrates the changes in maximum absolute error over the course of the genetic algorithm. It can be noted that the best performing run has a higher maximum absolute error than the poorer performing one. The change in the total absolute error across the genetic algorithm can be seen in Figure 13.6(b) and it can clearly be seen that the worst performing run obtained a lower total absolute error earlier in the algorithm than its final value. This can occur as the goal of this algorithm is not the minimisation of the total absolute error. If one considers Figure 13.6(f), it is clear that the total estimated error variance decreases in steps as would be expected. As has been seen with many of the other methods previously discussed, a decrease in total estimated error variance does not guarantee a decrease in total absolute error. It can also be seen that the total estimated error variance drops off very rapidly at first but then begins to decrease at a decreasing rate.

Figure 13.6(c) shows the changes in the percentage absolute error which mimic the changes shown in Figure 13.6(b). It can be seen that there is a period in the algorithm where the

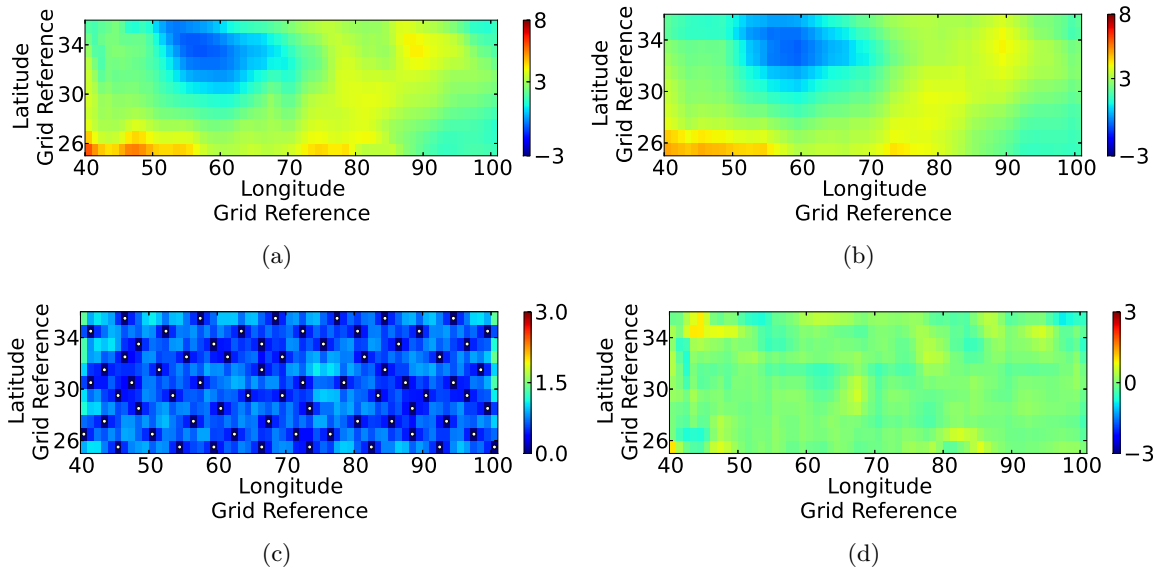


Figure 13.5: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for the worst performing run, run three, generated using the Genetic Algorithm designed to Minimise Total Estimated Error Variance. The white dots in (c) indicate the sampling points.

worst performing run does outperform the best performing run. Figure 13.6(d) illustrates the changes in the total integrated error across the algorithm. Referring to Figure 13.6(e) it can be seen that the maximum estimated error variance does not show a continual decrease across the algorithm but rather increases and decreases across the algorithm to end at lower values than the initial values, but, not necessarily at the lowest value obtained across the algorithm.

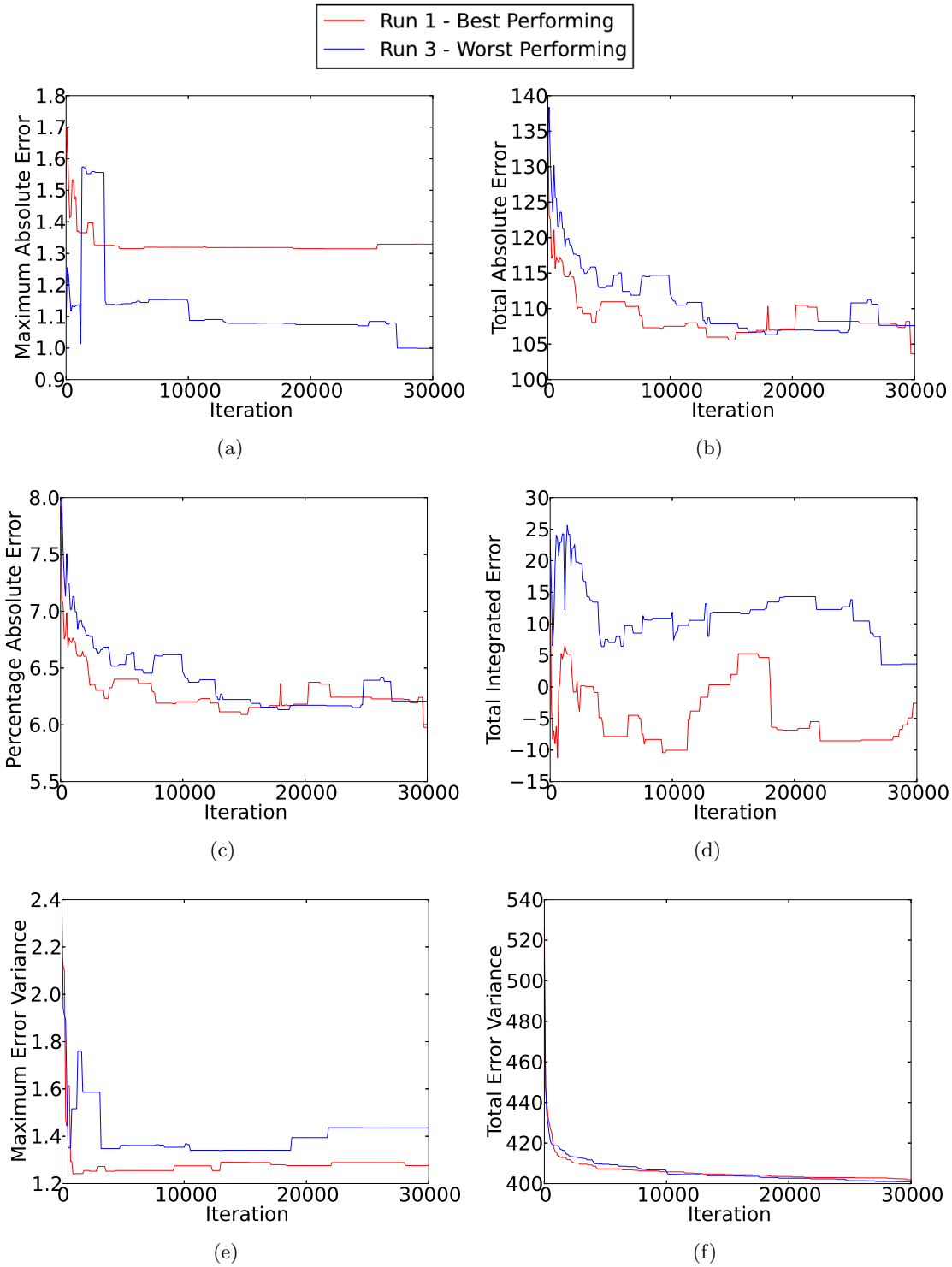


Figure 13.6: Progression of (a) maximum absolute error, (b) total absolute error, (c) percentage absolute error, (d) total integrated error, (e) maximum estimated error variance and (e) total estimated error variance across the best and worst performing runs generated using the Genetic Algorithm designed to Minimise Total Estimated Error Variance.

13.10 Hybridised Genetic Algorithm designed to Minimise Total Absolute Error

13.10.1 Selection of Starting Population

The starting population consists of 20 individuals. 19 individuals are randomly chosen without repetition and consist of 70 points each, while a single individual of 70 points is taken as the solution set from a run of the method of Addition of Point at Point of Maximum Absolute Error (see Chapter 12). The solution set selected from the method of Addition of Point at Point of Maximum Absolute Error for each run of this algorithm can be found in Table 13.5.

Run	Solution Set from Addition of Point at Point of Maximum Absolute Error
1	1
2	3
3	5
4	7
5	9

Table 13.5: Solution sets from Addition of a Point at Point of Maximum Absolute Error associated with each run of the Hybridised Genetic Algorithm designed to Minimise Total Absolute Error.

13.10.2 Evaluation - Fitness Function

The evaluation process calculates the sum of the absolute errors from prediction using the individual/sample set. This is used in conjunction with a lower fitness function being considered more optimal.

13.10.3 Convergence Criterion

A maximum number of iterations of 30000 is chosen as the convergence criterion.

13.10.4 Results and Discussion

The errors associated with the predictions using the 70 points can be found in Tables 13.6 and 13.7. The units associated with the maximum absolute error, total absolute error and total integrated difference are mol(C)/m^{2}s.

From Tables 13.6 and 13.7 it can be seen that while the percentage absolute error falls within the very narrow range of 2.68 to 3.05% and the total absolute error falls within the small range of 46.5 to 52.9, the total estimated error variances takes on the larger range of 498 to 547.

Maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux, the estimated error variance and the true error across the chosen two-dimensional region for the best and worst performing runs (in terms of total absolute error minimisation) can be seen in Figures 13.7 and 13.8 respectively. The maps for the remaining runs can be seen in Appendix D.7.

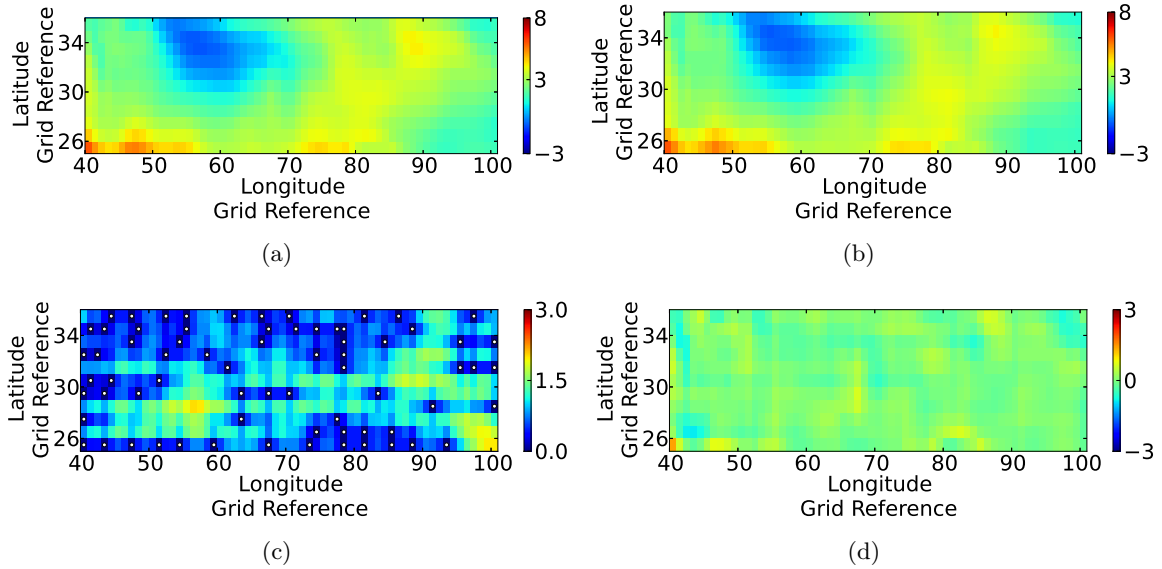


Figure 13.7: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for the best performing run, run three, generated using the Hybridised Genetic Algorithm designed to Minimise Total Absolute Error. The white dots in (c) indicate the sampling points.

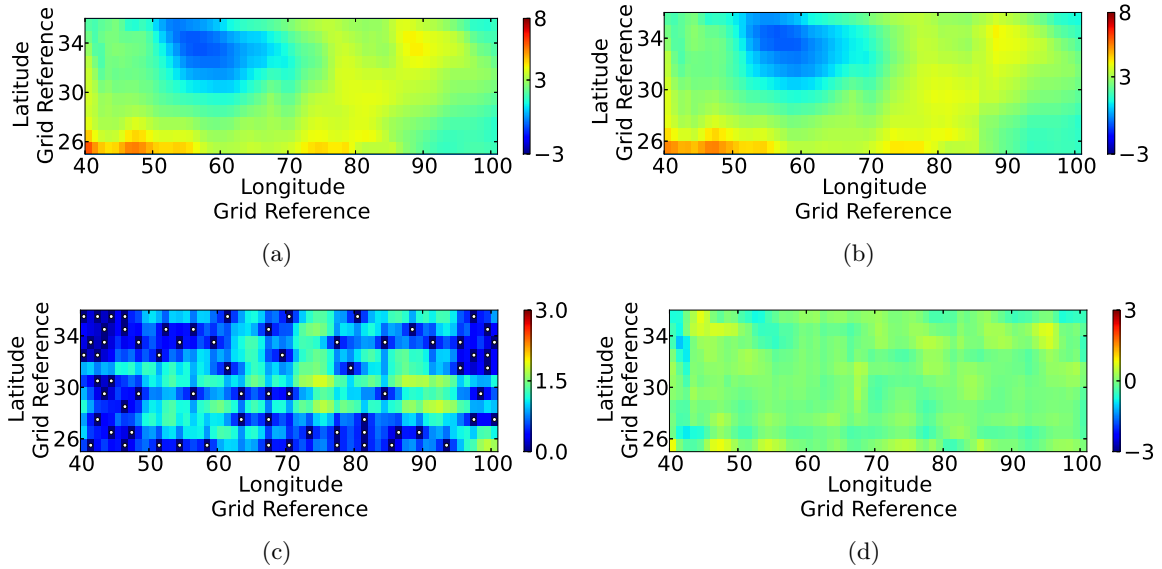


Figure 13.8: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for the worst performing run, run five, generated using the Hybridised Genetic Algorithm designed to Minimise Total Absolute Error. The white dots in (c) indicate the sampling points.

Run	Maximum Absolute Error	Total Absolute Error	Percentage Absolute Error	Total Integrated Error
1	0.5920	51.3219	2.9609%	−2.1085
2	0.5301	51.1355	2.9502%	−0.8203
3	0.5382	46.5260	2.6842%	−5.7738
4	0.7808	52.0173	3.0010%	2.0533
5	0.6214	52.8374	3.0483%	−0.1058

Table 13.6: Final maximum absolute, total absolute, percentage absolute and total integrated errors associated with the Hybridised Genetic Algorithm designed to Minimise Total Absolute Error in 2D.

Run	Maximum Estimated Error Variance	Total Estimated Error Variance
1	2.0828	544.1025
2	2.3056	528.7569
3	2.0180	546.5626
4	1.7465	498.0167
5	1.8196	525.3167

Table 13.7: Final maximum estimated error variances and total estimated error variances associated with the Hybridised Genetic Algorithm designed to Minimise Total Absolute Error in 2D.

The true flux and predicted flux maps in Figures 13.7 and 13.8 both show remarkable similarity. However, neither of them captures the values of the variable perfectly as is seen on the maps of true error (Figures 13.7(d) and 13.8(d)). Considering the location of sample points, seen as the white dots on the estimated error variance maps, it is clear that this hybridised genetic algorithm maintains grouping rather than evenly spacing the sample points.

Additionally, Figure 13.9 shows how the errors progressed across the best and worst performing runs as the number of iterations increased. The progression of these errors across the remaining runs can be seen in Appendix D.7.

From Figure 13.9(a) it can be seen that the maximum error variance ends the algorithm at a higher value than when it started. However, when Figure 13.9(b) is considered, it can be seen that this increase in maximum estimated error was actually taking place while the total absolute error decreased. This decrease occurred in steps due to the nature of the algorithm and the fitness function used.

From Figure 13.9(c), the percentage absolute error dropped across the 30000 iterations for both the best and worst cases. Majority of the drop occurred within the first 5000 iterations.

Figure 13.9(d) shows the changes in the total integrated error across the algorithm. Total integrated error for reasons previously mentioned is not considered a good performance indicator.

Figures 13.9(e) and 13.9(f) illustrate the changes in maximum estimated error variance and total estimated error variance respectively. Examining these graphs shows that the maximum estimated error variance changed quite dramatically at some points in the algorithm and was stable for long stretches at other times. When considering the total estimated error variance, it shows that although the total estimated error variance was lower at the end than at the start of the algorithm, the end values were not the lowest values attained across the length of the algorithm. This is of interest as the total absolute error was never allowed to increase so

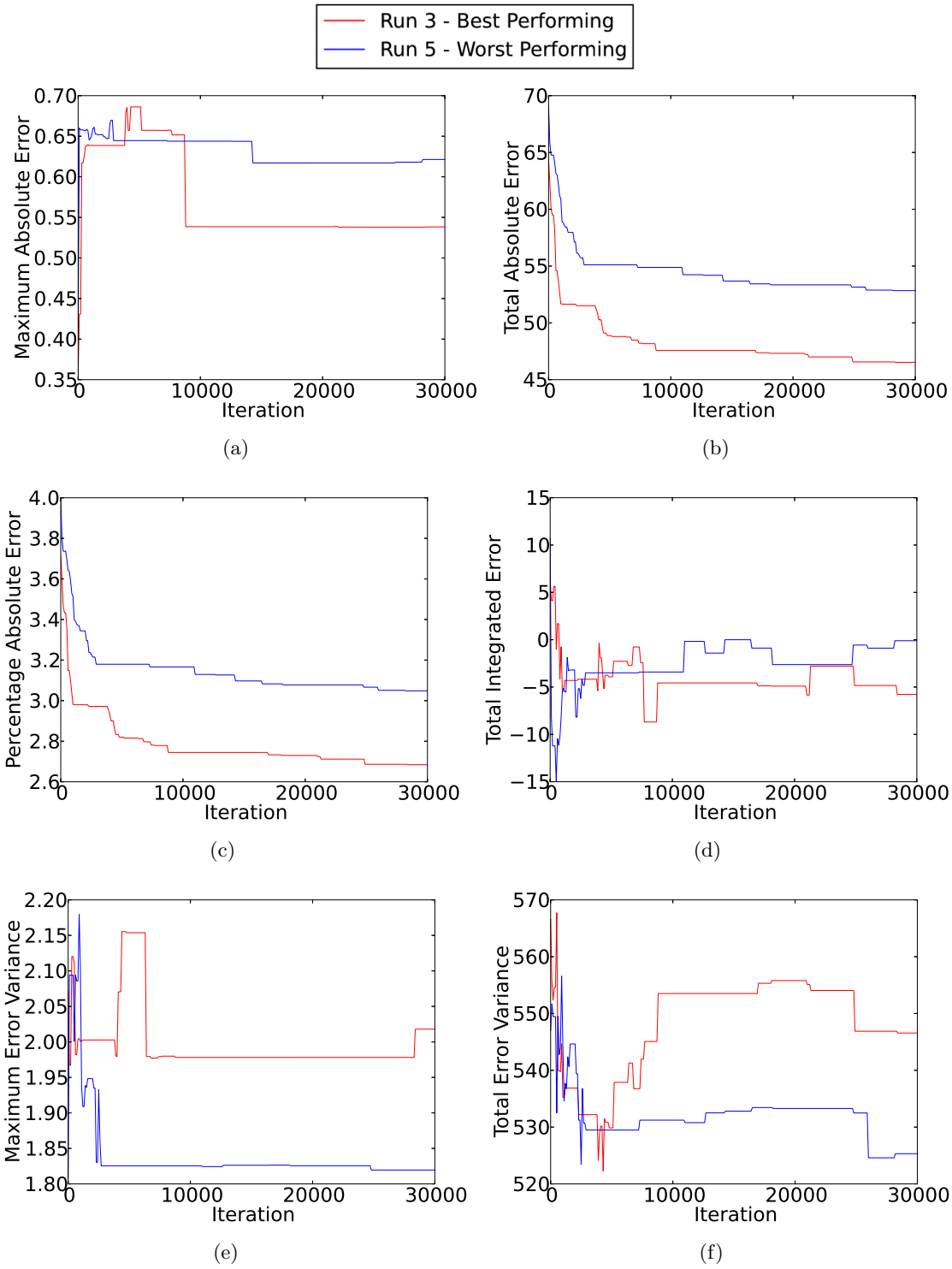


Figure 13.9: Progression of (a) maximum absolute error, (b) total absolute error, (c) percentage absolute error, (d) total integrated error, (e) maximum estimated error variance and (e) total estimated error variance across the best and worst performing runs generated using the Hybridised Genetic Algorithm designed to Minimise Total Absolute Error.

these graphs illustrate clearly that a for decreasing total absolute error it is possible to have an increasing total estimated error variance.

13.11 Hybridised Genetic Algorithm designed to Minimise Total Estimated Error Variance

13.11.1 Selection of Starting Population

The starting population consists of 20 individuals. 19 individuals are randomly chosen without repetition and consist of 70 points each, while a single individual of 70 points is taken as the solution set from a run of the method of Addition of Point at Point of Maximum Estimated Error Variance (see Chapter 11). The solution set selected from the method of Addition of Point at Point of Maximum Estimated Error Variance for each run of this algorithm can be found in Table 13.8.

Run	Solution Set from Addition of Point at Point of Maximum Estimated Error Variance
1	1
2	3
3	5
4	7
5	9

Table 13.8: Solution sets from Addition of a Point at Point of Maximum Estimated Error Variance associated with each run of the Hybridised Genetic Algorithm designed to Minimise Total Estimated Error Variance.

13.11.2 Evaluation - Fitness Function

The evaluation process calculates the sum of the estimated error variances from prediction using the individual/sample set. This is used in conjunction with a lower fitness function being considered more optimal.

13.11.3 Convergence Criterion

A maximum number of iterations of 30000 is chosen as the convergence criterion.

13.11.4 Results and Discussion

The errors associated with the predictions using the 70 points can be found in Tables 13.9 and 13.10. The units associated with the maximum absolute error, total absolute error and total integrated difference are mol(C)/m^{2}s.

The percentage absolute errors for this genetic algorithm can be found in found in Table 13.9 and show a narrow range of 6.0-6.4% error. The total estimated error variance found in Table 13.10 also covers a narrow range of 400.0-404.3.

Run	Maximum Absolute Error	Total Absolute Error	Percentage Absolute Error	Total Integrated Error
1	1.6963	104.8695	6.0502%	−3.8154
2	1.3990	110.7658	6.3904%	−4.4000
3	1.4130	106.8266	6.1631%	−0.6051
4	1.5500	107.6192	6.2088%	−4.9350
5	1.0783	108.5775	6.2641%	−8.9593

Table 13.9: Final maximum absolute, total absolute, percentage absolute and total integrated errors associated with the Hybridised Genetic Algorithm designed to Minimise Total Estimated Error Variance in 2D.

Run	Maximum Estimated Error Variance	Total Estimated Error Variance
1	1.2703	400.0009
2	1.3588	404.2867
3	1.1641	403.3292
4	1.3552	403.4786
5	1.5076	400.9658

Table 13.10: Final maximum estimated error variances and total estimated error variances associated with the Hybridised Genetic Algorithm designed to Minimise Total Estimated Error Variance in 2D.

Maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux, the estimated error variance and true error across the chosen two-dimensional region associated with the best and worst performing runs (in terms of total absolute error minimisation) can be seen in Figures 13.10 and 13.11 respectively. The maps for the remaining runs can be seen in Appendix D.8.

Both Figures 13.10(c) and 13.11(c) reflect relatively evenly distributed sample points (sample point locations are seen as the white dots). However, by comparing the maps of the true flux to the predicted flux in Figures 13.10 and 13.11, there are very clear visual differences in spite of some similarities.

Additionally, Figure 13.12 shows how the errors progressed across the best and worst performing runs as the number of iterations increased. The progression of these errors across the remaining runs can be seen in Appendix D.8.

Considering Figure 13.12(a) it can be seen that the genetic algorithm ended with a higher maximum absolute error than it started with in both the best and worst cases. When the graph of percentage absolute error is examined (Figure 13.12(c)), it is clear that for the worst performing run, the end percentage absolute error was not the lowest taken across the iterations in spite of the total estimate error variance continuing to drop across the remainder of the iterations. From this figure it is also clear that the percentage absolute error from the start to end of the algorithm did not undergo massive decreases. The differences between the starting and ending percentage absolute errors were only 0.25 and 0.15% respectively. Even if the algorithm had ended for the worst run when the percentage absolute error was at its lowest, it would only have achieved approximately a 0.39% decrease.

Figure 13.12(d) is not considered but is included for the sake of completeness as it is clear from Chapter 9 that total integrated error is not a good performance measure. From considering

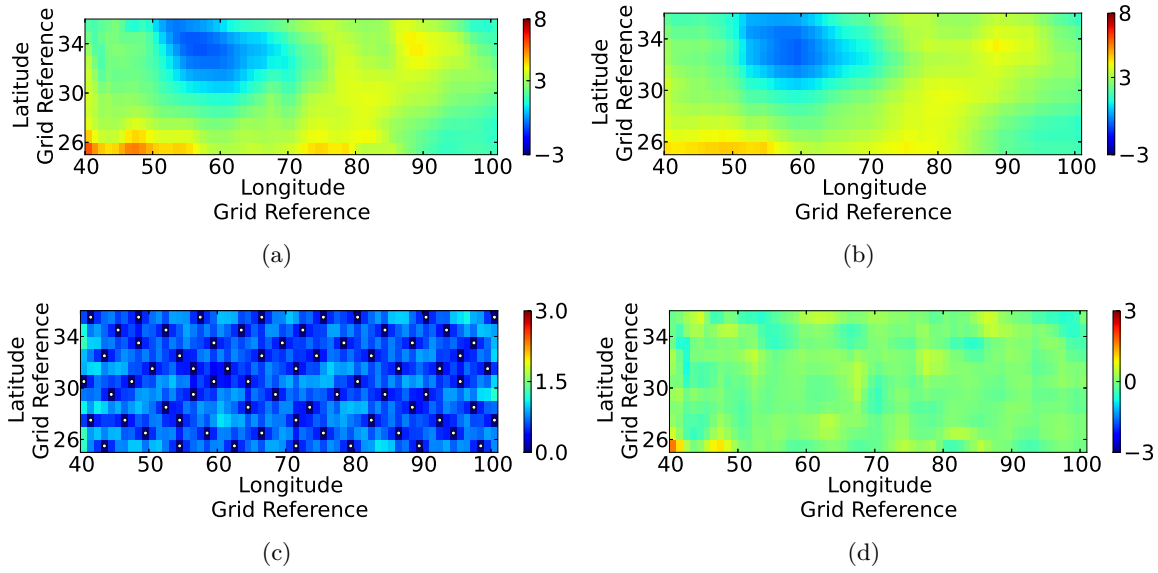


Figure 13.10: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for the best performing run, run one, generated using the Hybridised Genetic Algorithm designed to Minimise Total Estimated Error Variance. The white dots in (c) indicate the sampling points.

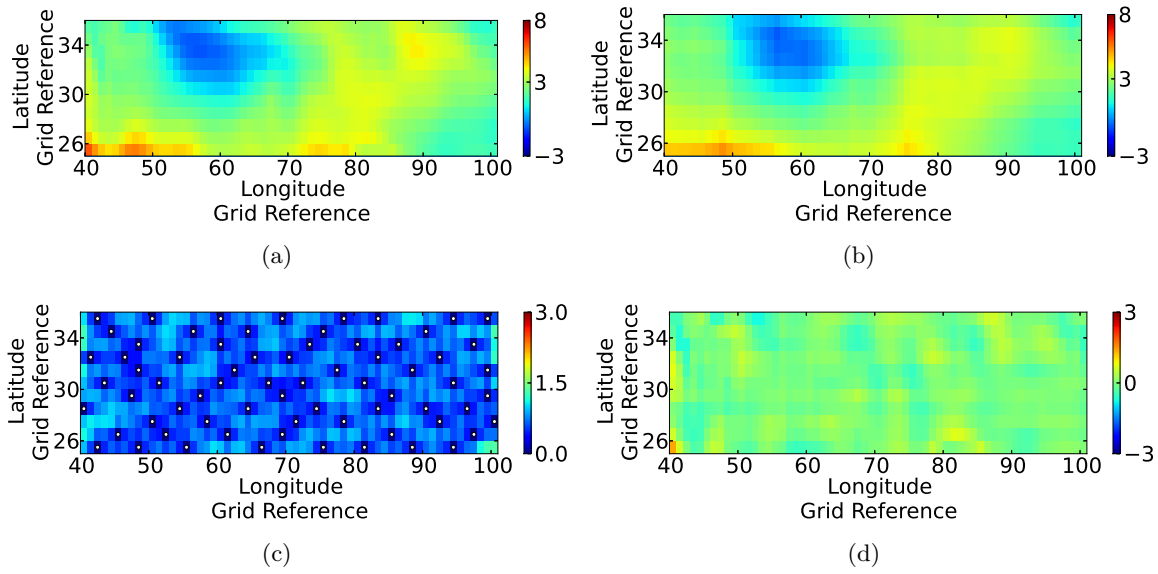


Figure 13.11: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for the worst performing run, run two, generated using the Hybridised Genetic Algorithm designed to Minimise Total Estimated Error Variance. The white dots in (c) indicate the sampling points.

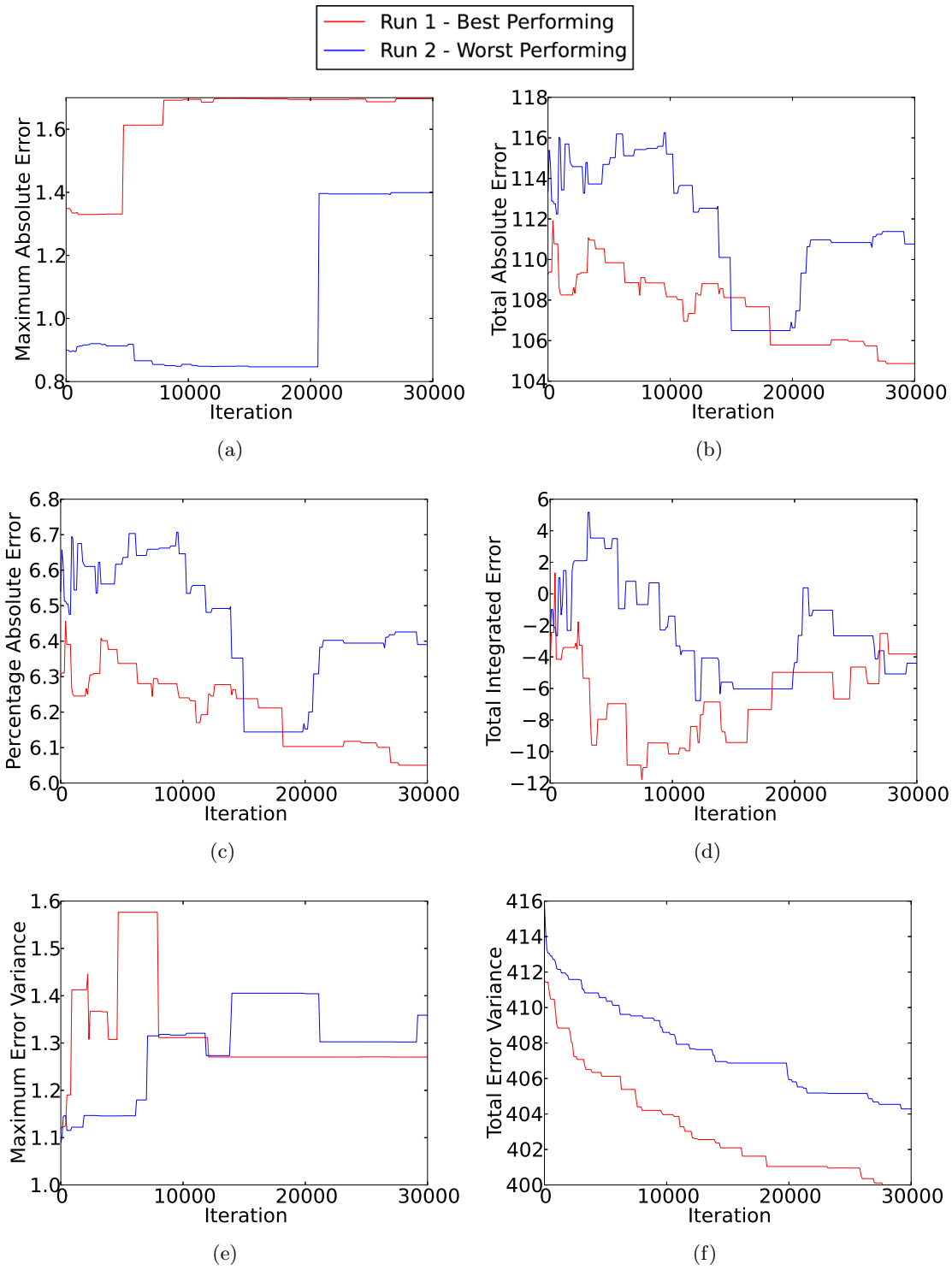


Figure 13.12: Progression of (a) maximum absolute error, (b) total absolute error, (c) percentage absolute error, (d) total integrated error, (e) maximum estimated error variance and (e) total estimated error variance across the best and worst performing runs generated using the Hybridised Genetic Algorithm designed to Minimise Total Estimated Error Variance.

Figures 13.12(e) and 13.12(f) it is clear that there are occasions where the total estimated error variance continues to decrease in steps while the maximum estimated error variance remains constant. Like the maximum absolute error, the maximum estimated error variance also ends on higher values than it started on. The step decreases in total estimated error variance yield only a decrease of approximately 10-12 units. When compared to the values they start and end with, this is not a large difference.

13.12 Genetic Algorithm designed to Maximise Total Absolute Error

13.12.1 Selection of Starting Population

The starting population consists of 20 individuals. Each individual is randomly chosen without repetition and consists of 70 points.

13.12.2 Evaluation - Fitness Function

The evaluation process calculates the sum of the absolute errors from prediction using the individual/sample set. This is used in conjunction with a higher fitness function being considered more optimal.

13.12.3 Convergence Criterion

A maximum number of iterations of 10000 is chosen as the convergence criterion.

13.12.4 Results and Discussion

The errors associated with the predictions using the 70 points can be found in Tables 13.11 and 13.12. The units associated with the maximum absolute error, total absolute error and total integrated difference are mol(C)/m^{2}s.

Run	Maximum Absolute Error	Total Absolute Error	Percentage Absolute Error	Total Integrated Error
1	4.3916	846.1220	48.8151%	−713.1526
2	5.1548	1115.7378	64.3700%	−1081.0975
3	4.9996	1062.2979	61.2869%	−1059.1415
4	4.1759	779.0798	44.9473%	−620.7446
5	5.0600	961.9681	55.4986%	−922.0639

Table 13.11: Final maximum absolute, total absolute, percentage absolute and total integrated errors associated with the Genetic Algorithm designed to Maximise Total Absolute Error in 2D.

From Tables 13.11 and 13.12 it can be seen that all the error measures are high for this genetic algorithm which maximises total absolute error. The percentage absolute error is very high falling within the range of 48.8-64.4%. This algorithm appears to produce a large

Run	Maximum Estimated Error Variance	Total Estimated Error Variance
1	4.4908	1348.1241
2	4.5862	1648.4166
3	4.7114	1630.3684
4	4.2919	1164.8040
5	4.5426	1409.9580

Table 13.12: Final maximum estimated error variances and total estimated error variances associated with the Genetic Algorithm designed to Maximise Total Absolute Error in 2D.

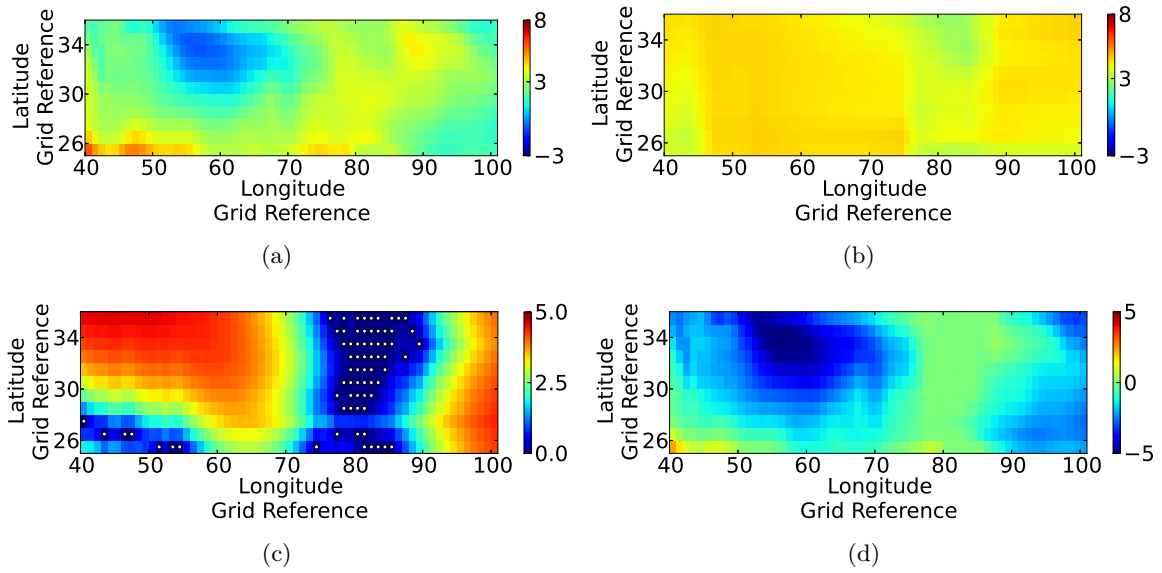


Figure 13.13: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for the best performing run, run two, generated using the Genetic Algorithm designed to Maximise Total Absolute Error. The white dots in (c) indicate the sampling points.

range across the summed errors and this is not surprising as the algorithm was only run for 10000 iterations and as such may not have converged yet.

Maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux, the estimated error variance and true error across the chosen two-dimensional region associated with the best and worst performing runs (in terms of total absolute error maximisation) can be seen in Figures 13.13 and 13.14 respectively. The maps for the remaining runs can be seen in Appendix D.9.

Considering Figures 13.13 and 13.14, it is clear that the maps of true flux and predicted flux are unrecognisable as representing the same variable. It also appears from the white dots present on the estimated error variance maps (Figures 13.13(c) and 13.14(c)) that these algorithm groups sample points in areas of similar values with lower variability.

Additionally, Figure 13.15 shows how the errors progressed across the best and worst performing runs as the number of iterations increased. The progression of these errors across the remaining runs can be seen in Appendix D.9.

Considering Figure 13.15(a) it is seen that the maximum absolute error more than doubles across the 10000 iterations. This allows for a dramatic increase in the upper bound of the total absolute error. Figure 13.15(b) demonstrates how the total absolute error undergoes

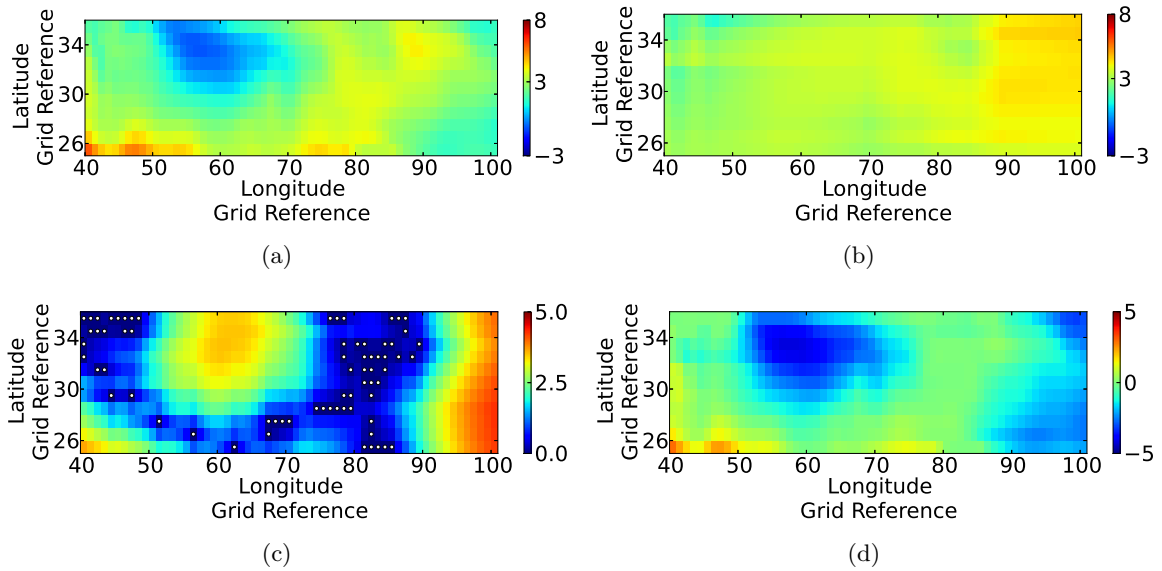


Figure 13.14: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for the worst performing run, run four, generated using the Genetic Algorithm designed to Maximise Total Absolute Error. The white dots in (c) indicate the sampling points.

step increases across the 10000 iterations. In the case of the run producing the set with the highest total absolute error, the total absolute error increase by a factor of more than five across the algorithm.

Figure 13.15(d) shows how with a dramatically increasing total absolute error, the total integrated error does diverge from zero. However, if one examines the curve carefully, there are still times where the total integrated error moves towards zero slightly even though the total absolute error must be stable or decreasing. This shows very well how unsuitable total integrated error is as a performance indicator.

Figures 13.15(e) and 13.15(f) indicate the changes in maximum estimated error variance and total estimated error variance respectively. It is once again illustrated that the total estimated error variance can increase while the maximum estimated error variance remains stable and that there is the occasional drop in total estimated error variance value while the total absolute error value remains stable or increases. In spite of this, it is clear that over such a large increase in total absolute error, that total estimated error variance also exhibits an overall increasing trend.

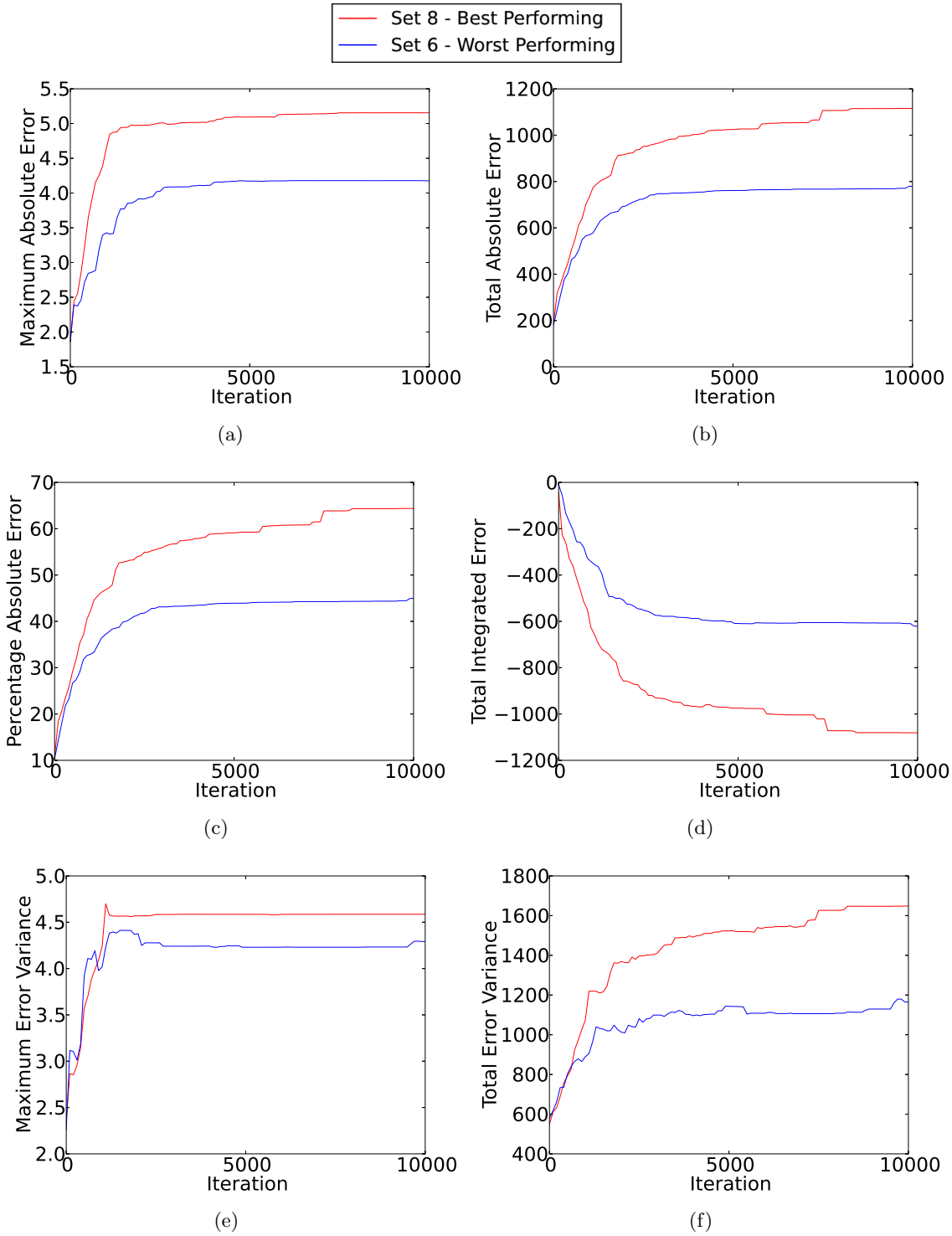


Figure 13.15: Progression of (a) maximum absolute error, (b) total absolute error, (c) percentage absolute error, (d) total integrated error, (e) maximum estimated error variance and (e) total estimated error variance across the best and worst performing runs generated using the Genetic Algorithm designed to Maximise Total Absolute Error.

13.13 Genetic Algorithm designed to Maximise Total Estimated Error Variance

13.13.1 Selection of Starting Population

The starting population consists of 20 individuals. Each individual is randomly chosen without repetition and consists of 70 points.

13.13.2 Evaluation - Fitness Function

The evaluation process calculates the sum of the estimated error variances from prediction using the individual/sample set. This is used in conjunction with a higher fitness function being considered more optimal.

13.13.3 Convergence Criterion

A maximum number of iterations of 10000 is chosen as the convergence criterion.

13.13.4 Results and Discussion

The errors associated with the predictions using the 70 points can be found in Tables 13.13 and 13.14. The units associated with the maximum absolute error, total absolute error and total integrated difference are mol(C)/m^{2}s.

Run	Maximum Absolute Error	Total Absolute Error	Percentage Absolute Error	Total Integrated Error
1	5.1003	890.2971	51.3637%	−756.6382
2	4.0670	681.1351	39.2966%	−522.1721
3	3.1320	423.9432	24.4584%	−205.7628
4	3.5888	660.0253	38.0787%	−545.4286
5	4.9063	842.2705	48.5929%	−718.5922

Table 13.13: Final maximum absolute, total absolute, percentage absolute and total integrated errors associated with the Genetic Algorithm designed to Maximise Total Estimated Error Variance in 2D.

Run	Maximum Estimated Error Variance	Total Estimated Error Variance
1	4.9253	2493.6870
2	4.9118	2242.4910
3	4.9254	2389.8840
4	4.7398	1962.9460
5	4.9249	2382.0271

Table 13.14: Final maximum estimated error variances and total estimated error variances associated with the Genetic Algorithm designed to Maximise Total Estimated Error Variance in 2D.

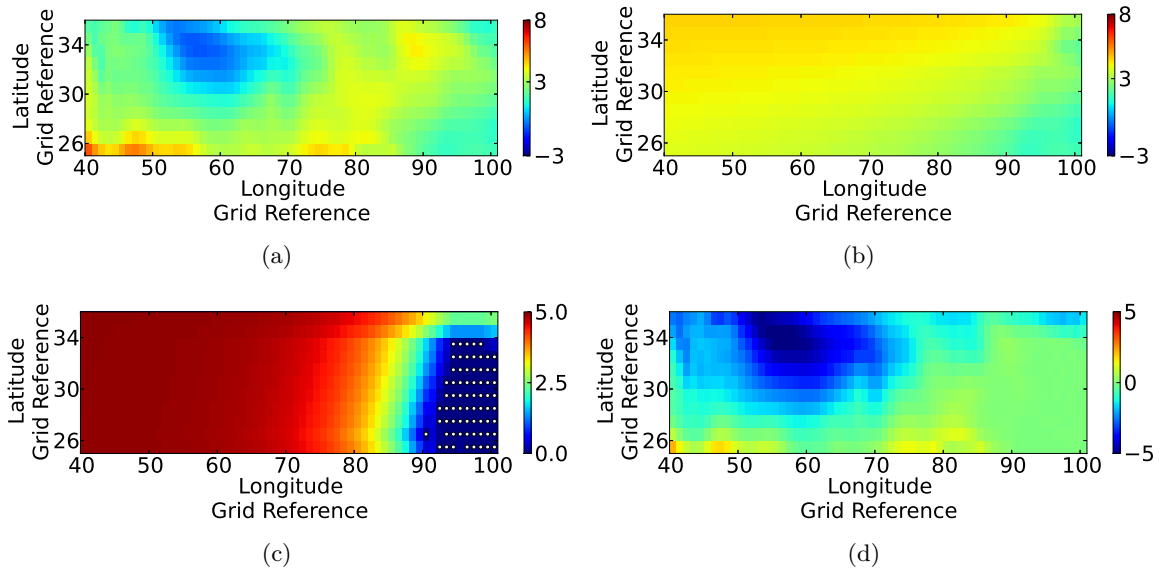


Figure 13.16: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for the best performing run, run one, generated using the Genetic Algorithm designed to Maximise Total Estimated Error Variance. The white dots in (c) indicate the sampling points.

Of interest from Tables 13.13 and 13.14 is that in spite of the high total estimated error variance values, the total absolute errors and percentage absolute errors are not as high.

Maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux, the estimated error variance and the true error across the chosen two-dimensional region for the best and worst performing runs (in terms of total absolute error maximisation) can be seen in Figures 13.16 and 13.17 respectively. The maps for the remaining runs can be seen in Appendix D.10.

From Figures 13.16(c) and 13.17(c) it is clear that the points are grouped very tightly in order to maximise distances between sampled and unsampled points. The maps of the true and predicted fluxes are vastly different except for the areas in which the sample points are clustered.

Additionally, Figure 13.18 shows how the errors progressed across the best and worst performing runs as the number of iterations increased. The progression of these errors across the remaining runs can be seen in Appendix D.10.

From Figure 13.18(a) it can be seen that there is a dramatic difference in the maximum absolute error increase in the best and worst performing runs. The same can be said of the total absolute error (Figure 13.15(b)) where the worst performing set increases by a factor of less than two while the best performing set increases by a factor of approximately four. Figure 13.18(c) illustrates that at the end of the algorithm there is approximately a 25% difference in the percentage absolute error values.

Considering Figures 13.18(e) and 13.18(f) it can be seen that the best and worst performing runs have almost identical maximum estimated error variances although the total estimated error variances differ by approximately 200. This difference of 200 translates, however, into a 25% difference in percentage absolute error.

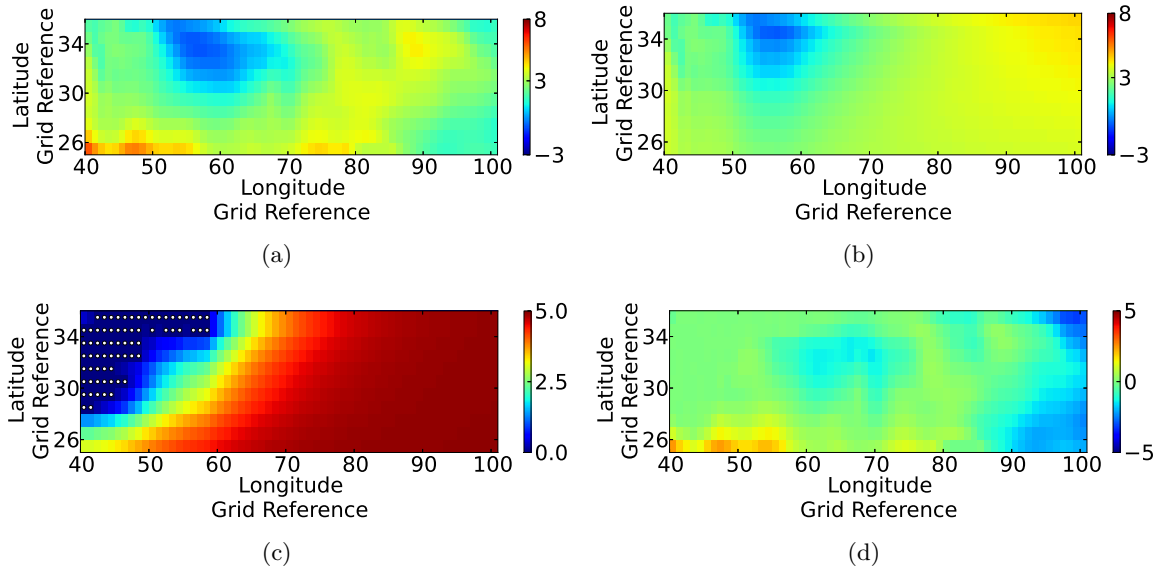


Figure 13.17: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for the worst performing run, run three, generated using the Genetic Algorithm designed to Maximise Total Estimated Error Variance. The white dots in (c) indicate the sampling points.

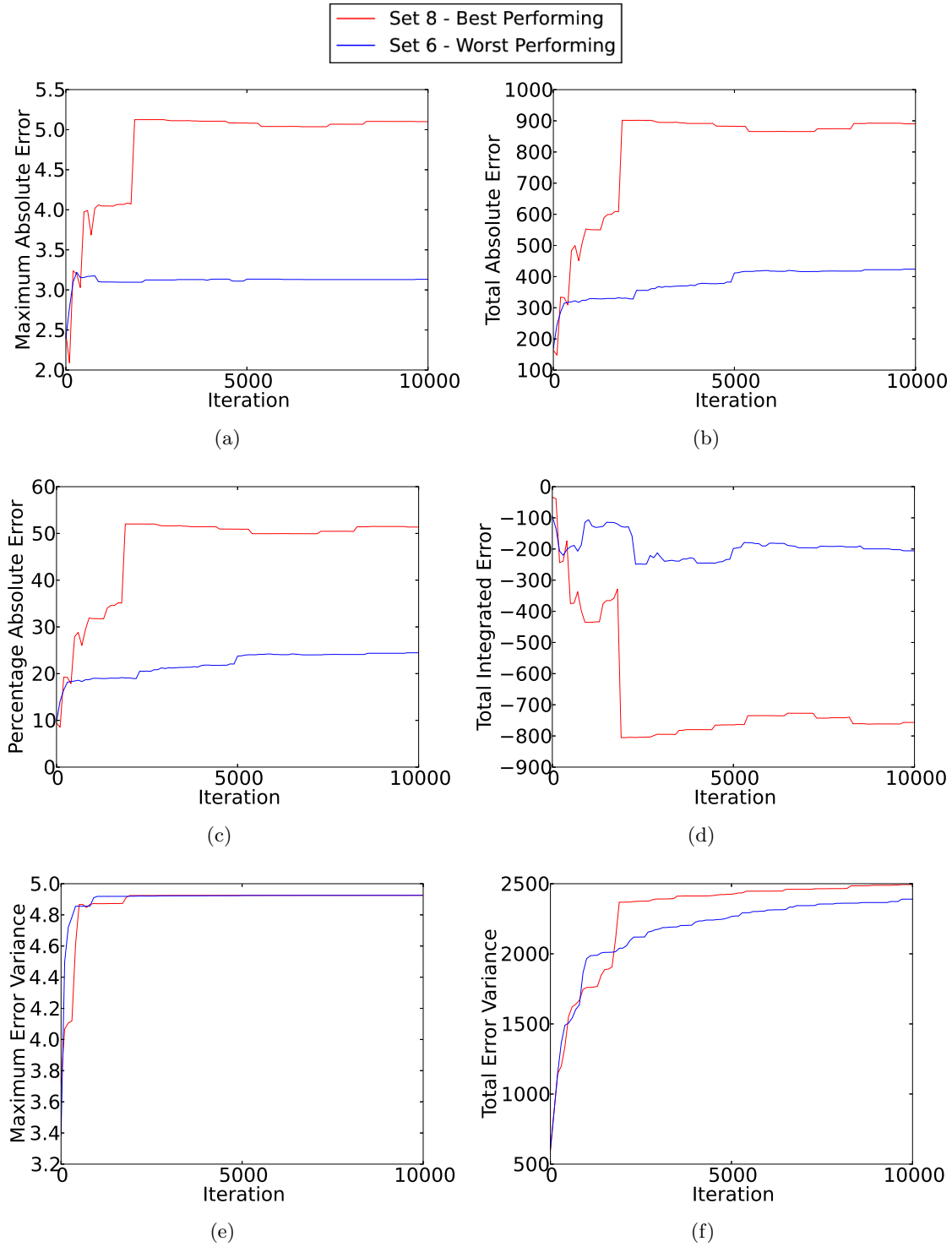


Figure 13.18: Progression of (a) maximum absolute error, (b) total absolute error, (c) percentage absolute error, (d) total integrated error, (e) maximum estimated error variance and (e) total estimated error variance across the best and worst performing runs generated using the Genetic Algorithm designed to Maximise Total Estimated Error Variance.

13.14 Limitations

All runs of all the genetic algorithms are stopped on a maximum number of iterations criteria and thus, do not guarantee the absolute optimal solution.

Only simple genetic operations have been used in the algorithm and as such, more advanced methods may reach optimal solutions in a more timely manner.

Only 5 runs of each genetic algorithm were performed using a single covariance and trend structure.

CHAPTER 14

TWO-DIMENSIONAL SAMPLING - COMPARISON OF SAMPLING METHODS

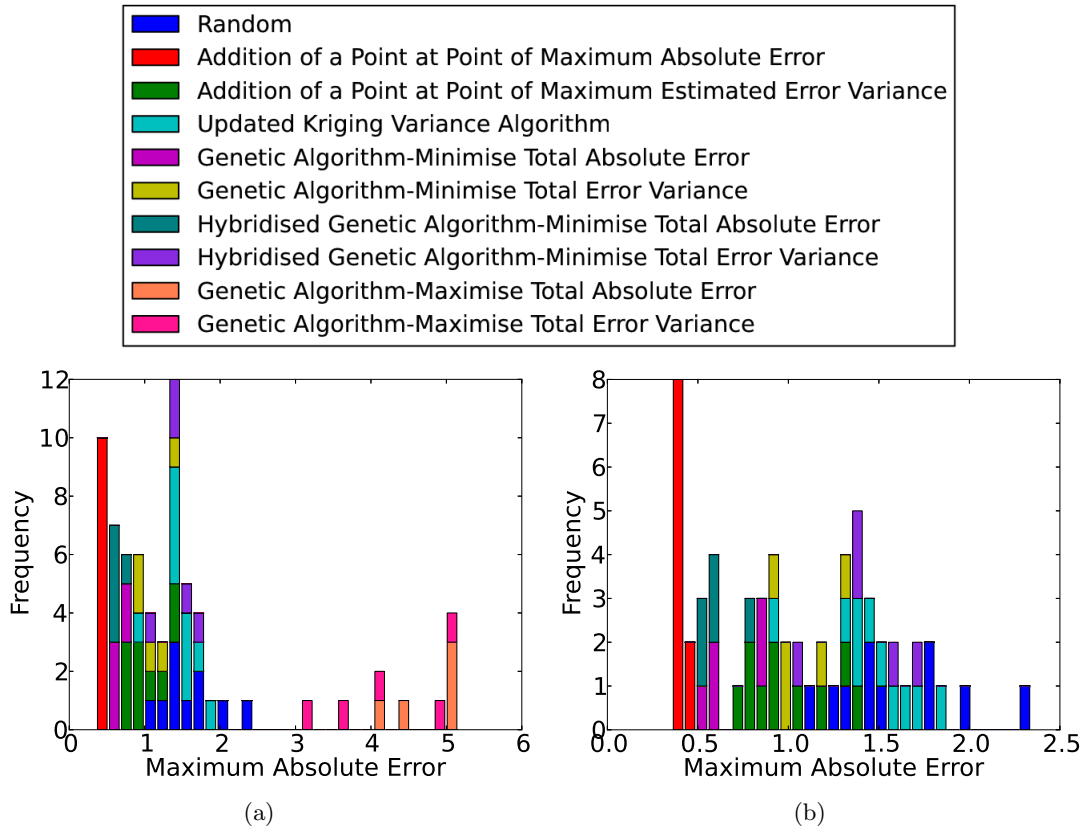
14.1 Comparison of Errors Across All Methods

In order to compare the errors produced by the methods described and implemented in Chapters 9 to 13 histograms of each error were produced. Each histogram shows one of the resulting errors of all the implemented methods. These can be seen in the (a) sub-figures of Figures 14.1 to 14.6. Additionally, in order to compare only the methods which aim to minimise the errors, a histogram was produced per error containing the results only of methods aiming to minimise errors. The two genetic algorithms which aimed to maximise total absolute error and total error variance were not included in these histograms. These histograms can be seen in the (b) sub-figures of Figures 14.1 to 14.6.

Figure 14.1(a) indicates a clear gap in maximum absolute error between the methods designed to minimise and those designed to maximise errors. Figures 14.2(a) and 14.3(a) show the total absolute errors and percentage absolute errors obtained across all the methods tested. It is clear that the two genetic algorithms designed to maximise errors generate these errors across a very wide range but that is not unusual given that the method was only allowed to run for 10000 iterations. This does, however, imply that the true worst possible sample set for prediction by means of Kriging using the chosen covariance and trend structures and parameter values may in fact produce higher errors than those shown in these figures. It is also important to note that although the 10 runs of Random Sampling performed produced errors significantly lower than the genetic algorithms designed to maximise errors, due to the nature of Random Sampling it is possible to select a set which performs anywhere across the range of errors. It is also clear that the worst performing of the methods designed to minimise errors has a total absolute error less than half of that obtained by the worst performing method designed to maximise errors.

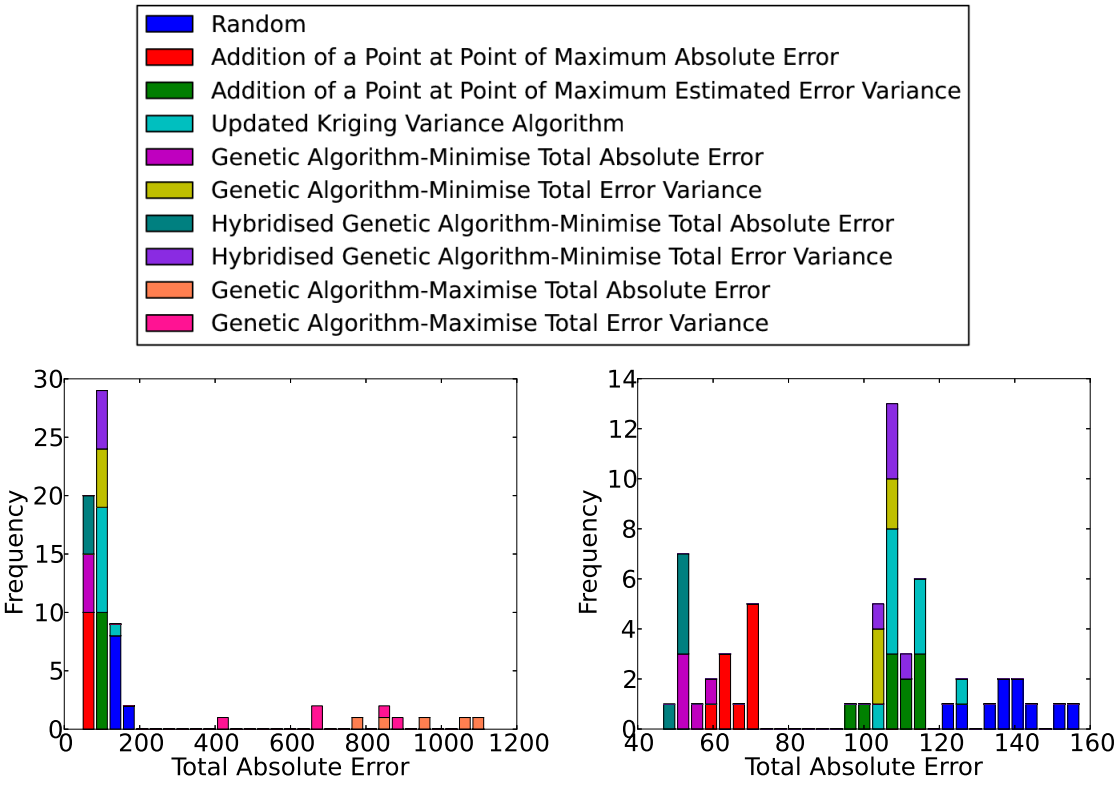
Considering Figure 14.4(a) it can be seen that in terms of total integrated error the Random Sampling and methods designed to minimise errors have much smaller errors than those of the methods designed to maximise errors. It is also noted that while the majority of methods both under- and over-predict, the methods designed to maximise errors all dramatically over predict the integrated carbon dioxide flux.

Figure 14.5(a) indicates a large gap in maximum estimated error variance values between the methods designed to maximise errors and the remaining methods. These higher maximum estimated error variance values place a higher upper bound on the total estimated error variance.



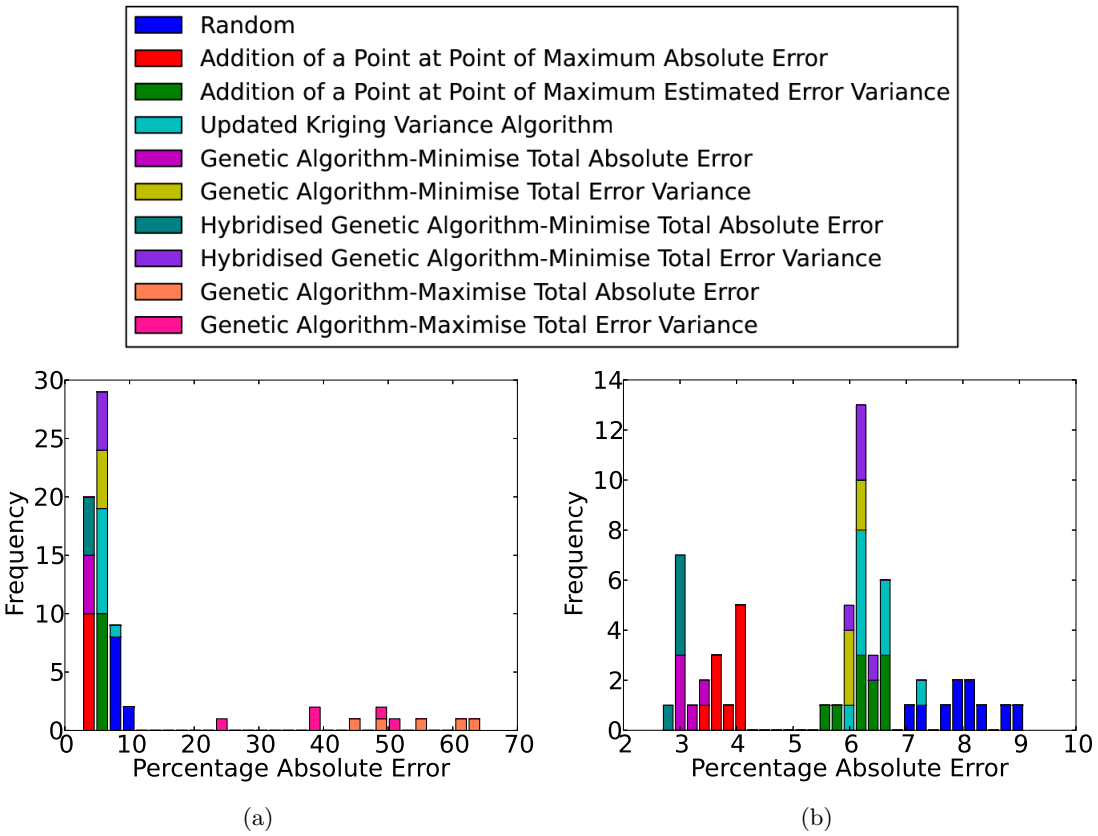
(a) (b)

Figure 14.1: Maximum absolute errors obtained across (a) all methods and (b) all methods designed to minimise errors.



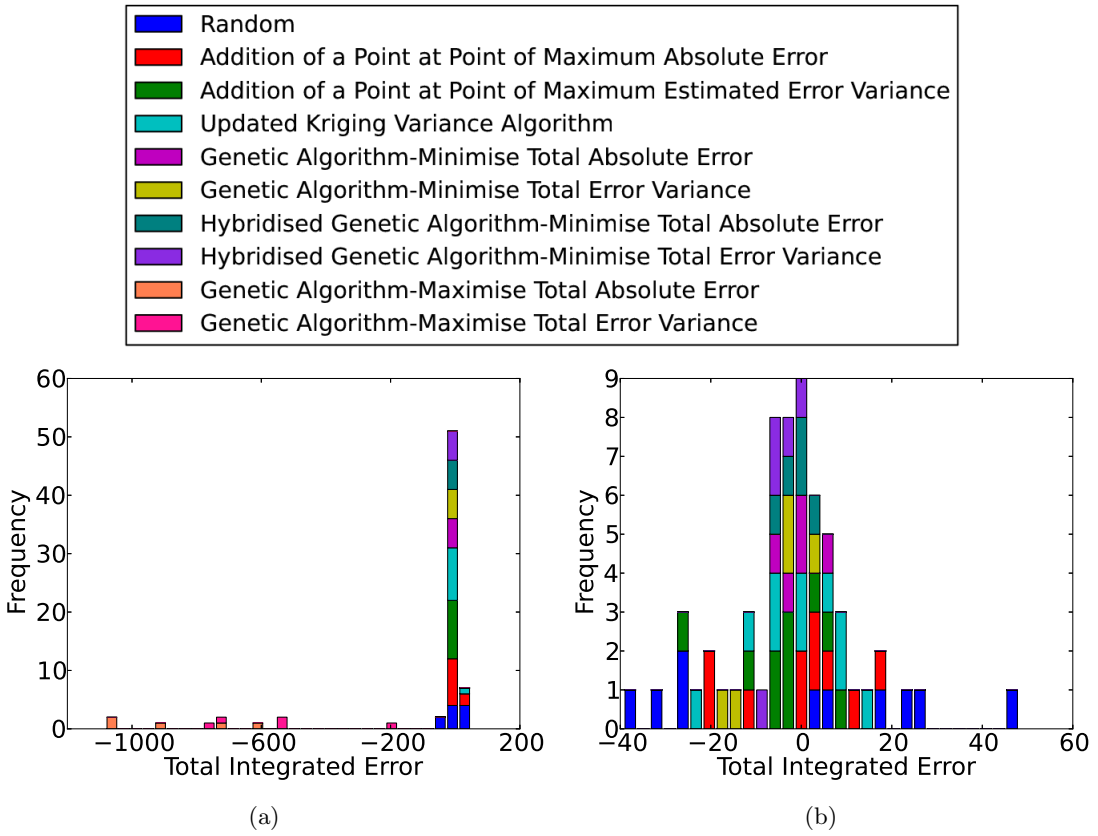
(a) (b)

Figure 14.2: Total absolute errors obtained across (a) all methods and (b) all methods designed to minimise errors.



(a) (b)

Figure 14.3: Percentage absolute errors obtained across (a) all methods and (b) all methods designed to minimise error.



(a) (b)

Figure 14.4: Total integrated errors obtained across (a) all methods and (b) all methods designed to minimise errors.

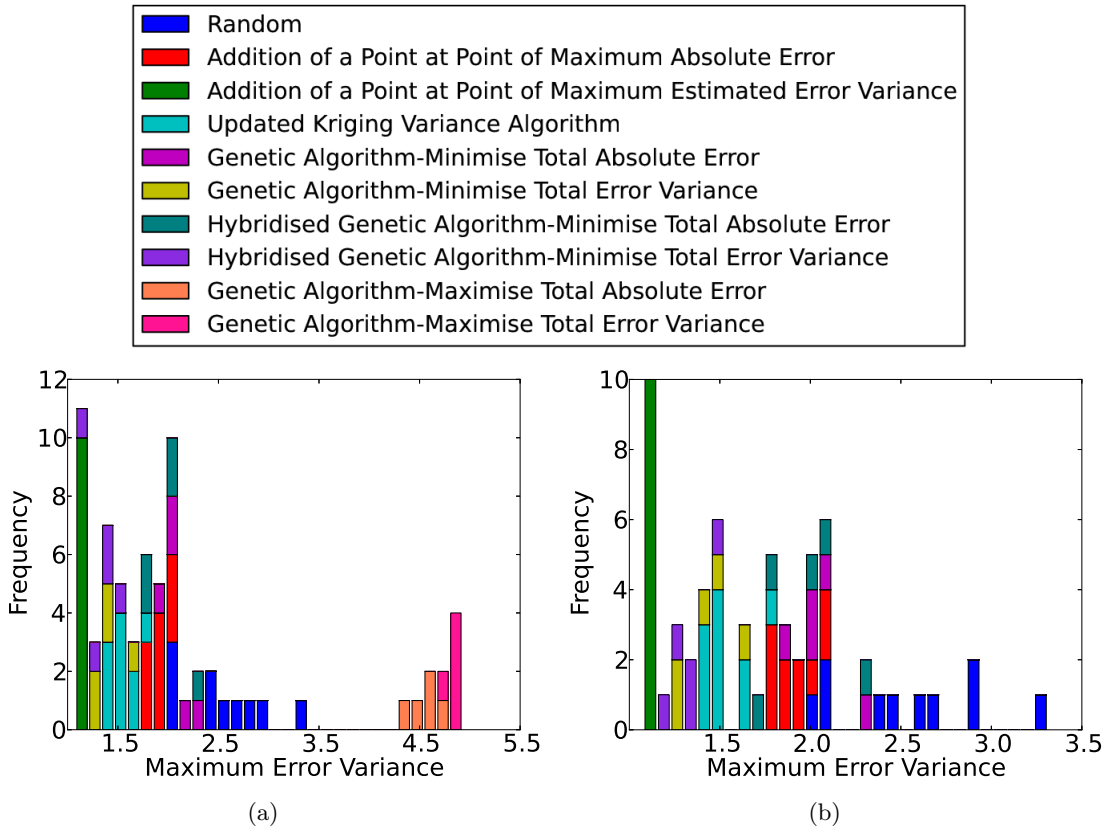


Figure 14.5: Maximum estimated error variances obtained across (a) all methods and (b) all methods designed to minimise errors.

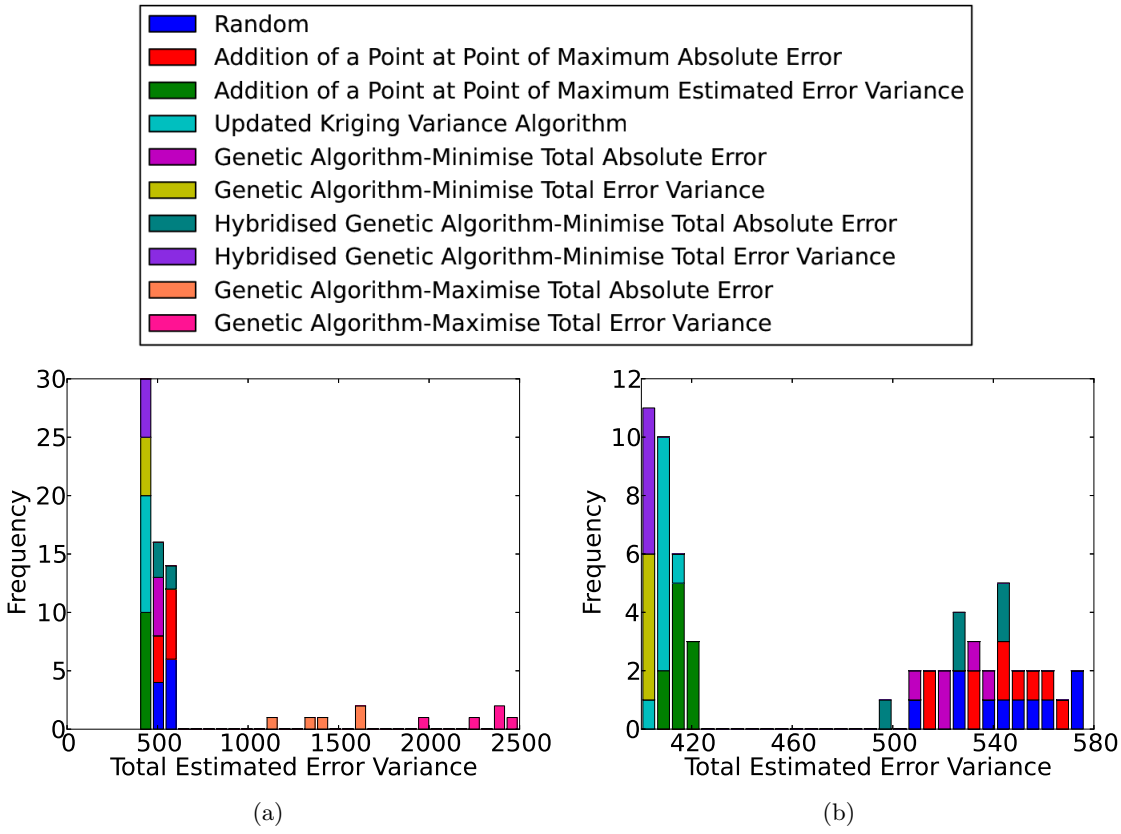


Figure 14.6: Total estimated error variances obtained across (a) all methods and (b) all methods designed to minimise errors.

The total estimated error variances for the various methods can be seen in Figure 14.6(a) from which it is clear that the Genetic Algorithm designed to Maximise Total Estimated Error Variance yields the largest total estimated error variances. The next highest total estimated error variances are generated by the Genetic Algorithm designed to Maximise Total Absolute Error. If Figure 14.2(a) is considered again it can be seen that the method which yields the highest total absolute error is the Genetic Algorithm designed to Maximise Total Absolute Error, while the next highest is the Genetic Algorithm designed to Maximise Total Estimated Error Variance. This implies that although the Genetic Algorithm designed to Maximise Total Estimated Error Variance produces far larger total estimated error variances, it fails to produce larger total absolute errors. The explanation for why this is possible can be found in Chapter 9.

From Figure 14.1(b) it is clear that of the methods designed to find the more optimal sets (minimise errors); Random Sampling and the Updated Kriging Variance Algorithm fall across a wider range than the other methods. Addition of a Point at Point of Maximum Absolute Error clearly produces the lowest maximum absolute errors. However, it does not produce the lowest total absolute errors, as can be seen in Figure 14.2(b).

Figure 14.3(b) shows the definite gap in percentage absolute error between the *a-priori* and *a-posteriori* methods. Also of interest is the wider range of percentage absolute errors that can be taken on the Random Sampling and Updated Kriging Variance Algorithm. None of the optimising *a-priori* methods appear to have a clear advantage over each other, although they do show some advantage over Random Sampling. The *a-posteriori* methods also appear to perform similarly, although the Genetic Algorithm designed to Minimise Total Absolute Error and the hybridised genetic algorithm with the same goal do show a slight improvement over the Addition of a Point at Point of Maximum Absolute Error. If one considers the method of Addition of a Point at Point of Maximum Absolute Error and the Genetic Algorithm designed to Minimise Total Absolute Error along with its hybridised counterpart, it can be seen that the genetic algorithms produced slightly lower total absolute errors and percentage absolute errors than the more simple method of Addition of a Point at Point of Maximum Absolute Error.

Figure 14.4(b) illustrates that Random Sampling takes the largest range of total integrated error values, while the remaining methods designed to minimise errors fall within a relatively small range on either side of zero.

From Figure 14.5(b) it can be seen that the Genetic Algorithm designed to Minimise Total Estimated Error Variance, its hybridised version and the Addition of a Point at Point of Maximum Estimated Error Variance produce the lowest maximum estimated error variances. The same can be said of the total estimated error variances which can be seen in Figure 14.6(b). While a large difference is present between the *a-priori* and remaining methods in terms of total estimated error variance, it is of interest to note that the *a-posteriori* methods which exhibit higher total estimated error variances than the *a-priori* methods exhibit lower total absolute errors than the *a-priori* methods. This suggests that the estimated error variance does not always provide a good indication of true error.

Additionally it can be seen that while the Genetic Algorithm designed to Minimise Total Estimated Error Variance and its corresponding hybridised method do produce total estimated error variance that are slightly lower than those produced by the Updated Kriging Variance Algorithm and the method of Addition of a Point at Point of Maximum Estimated Error Variance, the total absolute errors produced by all of the above mentioned methods are very similar.

14.2 Comparison of Methods Which Perform the Addition of a Single Sample Point Per Iteration

In order to compare how the addition of points across the methods of Updated Kriging Variance Algorithm, Addition of a Point at Point of Maximum Error Variance and Addition of a Point at Point of Maximum Absolute Error perform from the same starting sets, each of the recorded errors has been plotted against the number of points in the sample across the three methods for set two. This set was chosen as the results it gave across all of the three methods where good but neither the best nor the worst for that method. These can be seen in Figure 14.7.

From Figure 14.7(a) it can be seen that the Addition of a Point at Point of Maximum Absolute Error results in the largest decrease in the maximum absolute error value. This is followed by the method of Addition of a Point at Point of Maximum Estimated Error Variance, with the Updated Kriging Variance Algorithm yielding the smallest decrease. It is also clear that the decrease is not smooth across any of these three methods.

Figure 14.7(b) illustrates the changes in total absolute error as points are added by the three methods. It is clear that the Addition of a Point at Point of Maximum Absolute Error performs best in this regard, although up until 8 points, there are times where the Addition of a Point at Point of Maximum Estimated Error Variance performs better. This is however, still a sample size of approximately 1.2% of the total number of points. After this point, the Addition of a Point at Point of Maximum Absolute Error remains the best performing method for the remainder of the additions. From 12 points, the Updated Kriging Variance Algorithm and the Addition of a Point at Point of Maximum Estimated Error Variance experience very similar total absolute errors ending on values which differ by less than 2 units. In terms of percentage absolute error (seen in Figure 14.7(c)), it can be seen that the difference between the method of Addition of a Point at Point of Maximum Absolute Error and the other two single addition per iteration methods is only around 2%. However, this difference places the method of Addition of a Point at Point of Maximum Absolute Error below 5% absolute error. In addition it can be seen that the method of Addition of a Point at Point of Maximum Absolute Error's percentage absolute error falls below 10% using far fewer sample points than the other two methods.

Figure 14.7(d) has been included for the sake of completeness but as it has been explained previously, total integrated error does not provide a good measure of performance or accuracy.

Figure 14.7(e) demonstrates the changes in the maximum estimated error variance. It is clear that this value never increases as by adding another point it is impossible to increase the distances between the previously chosen sample points and the unsampled points. It is however, possible for the maximum estimated error variance to remain constant for a period. Additionally it can be seen that the largest drop in maximum estimated error variance belongs to the method of Addition of a Point at Point of Maximum Estimated Error Variance, although this decrease is in no way smooth. The Updated Kriging Variance Algorithm and the method of Addition of a Point at Point of Maximum Absolute Error trade places in terms of lower maximum estimated error variance until approximately 31 points, from which point the Updated Kriging Variance Algorithm clearly produces a lower maximum estimated error variance.

Figure 14.7(f) shows the changes in total estimated error variance with the addition of points. The Updated Kriging Variance Algorithm exhibits a smooth decrease which is expected due to the testing portion of this algorithm. The method of Addition of a Point at Point of Maximum Estimated Error Variance shows a less smooth curve which yields higher values until approximately 25 points, after which the curve follows the curve of the Updated Kriging

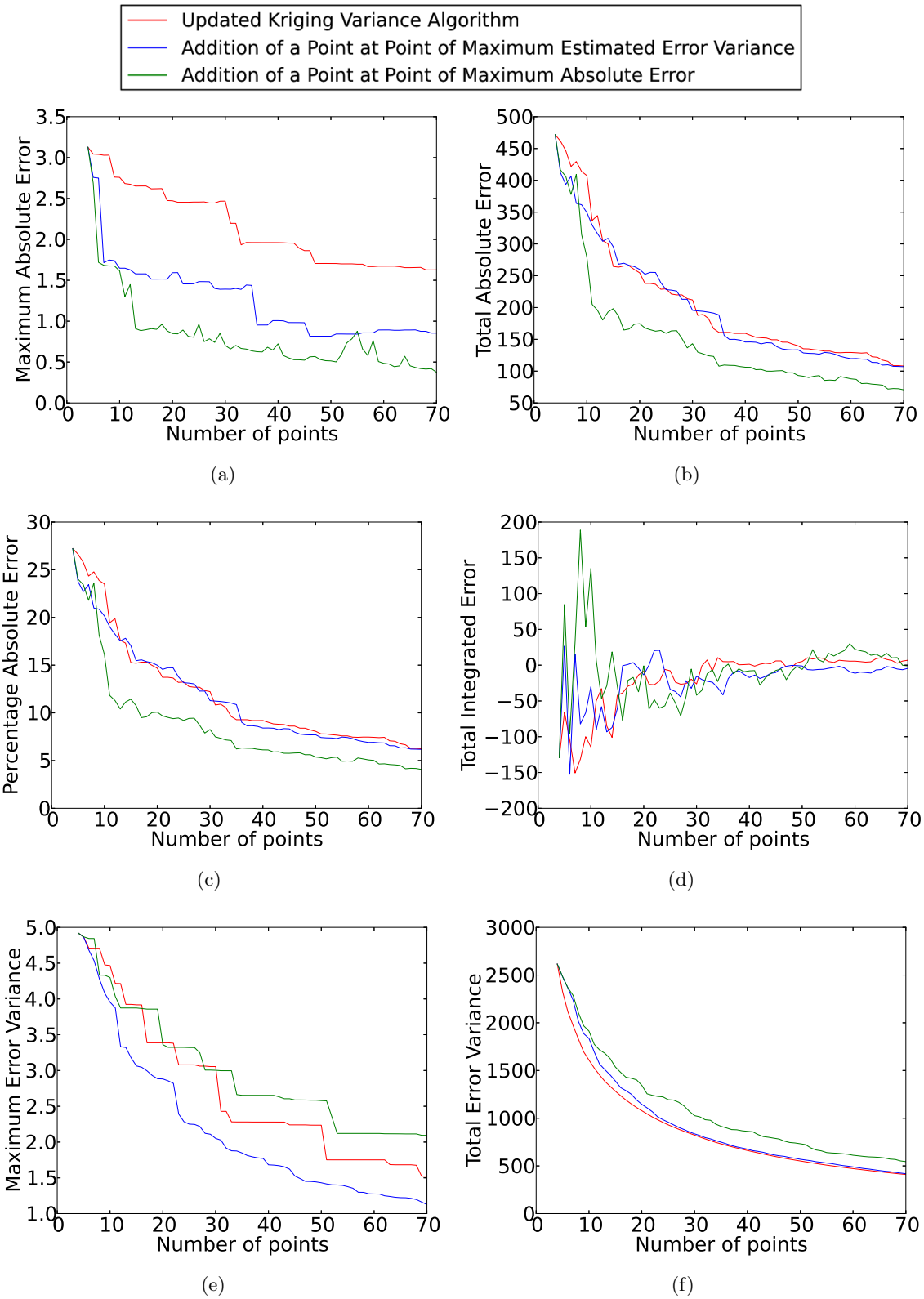


Figure 14.7: Progression of (a) maximum absolute error, (b) total absolute error, (c) percentage absolute error, (d) total integrated error, (e) maximum estimated error variance and (e) total estimated error variance across set two generated using the Updated Kriging Variance Algorithm, Addition of a Point at Point of Maximum Error Variance and Addition of a Point at Point of Maximum Absolute Error.

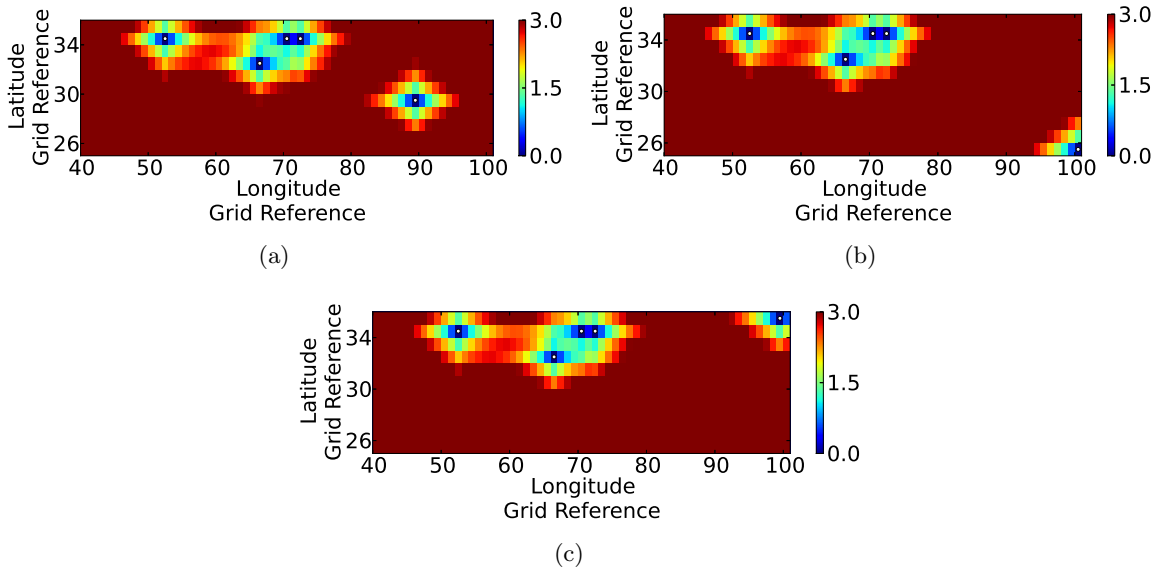


Figure 14.8: Maps of estimated error variance for a sample size of 5 for the sampling methods of (a)Updated Kriging Variance Algorithm, (b) Addition of a Point at Point of Maximum Estimated Error Variance and (c) Addition of a Point at Point of Maximum Absolute Error with sample points indicated by white dots.

Variance Algorithm very closely. The Addition of a Point at Point of Maximum Absolute Error has the least smooth curve and exhibits the highest total estimated error variance values in spite of this method exhibiting the lowest total absolute error values. This clearly shows that the estimate error variance is not an accurate measure of true performance. However, if there are no values for the variable of interest and no models available for this parameter, then it is the only reasonable measure available.

In order to consider the differences in sample point placement, Figures 14.8 to 14.12 show the estimated error variances across the Updated Kriging Variance Algorithm, Addition of a Point at Point of Maximum Estimated Error Variance and Addition of a Point at Point of Maximum Absolute Error methods for 5, 10, 20, 40 and 70 points. The sample points can be seen by the white dots on the estimated error variance maps.

The first four points are randomly chosen and these can be seen as the common four points in Figure 14.8. The only different point is the single point which has been added by each method at this point in the algorithms. It can be seen that the three methods considered place this single point in very different positions. All three methods do, however, place the point in the right hand side of the map, as there are no points here, while there is one randomly chosen point on the left hand side of the map. It is also noted that as would be expected, the estimated error variances are high and increase as one moves further away from the sample point locations.

Figure 14.9 shows a drop in the individual estimated error variances as 5 more points are added. This drop is most significant in the method of Addition of a Point at Point of Maximum Estimated Error Variance, as would be expected. This does not, however, guarantee that this method also produces the largest drop in the total estimated error variance. The Updated Kriging Variance Algorithm attempts to distribute the sample points relatively evenly across the map. The method of Addition of a Point at Point of Maximum Estimated Error Variance attempts a similar distribution. However, the placement is not identical due to the different criteria for placement. Nevertheless, since the estimated error variance is dependent only on the distances between sampled and unsampled points and the covariance structure, the relatively even distribution is expected. This distribution is not entirely even,

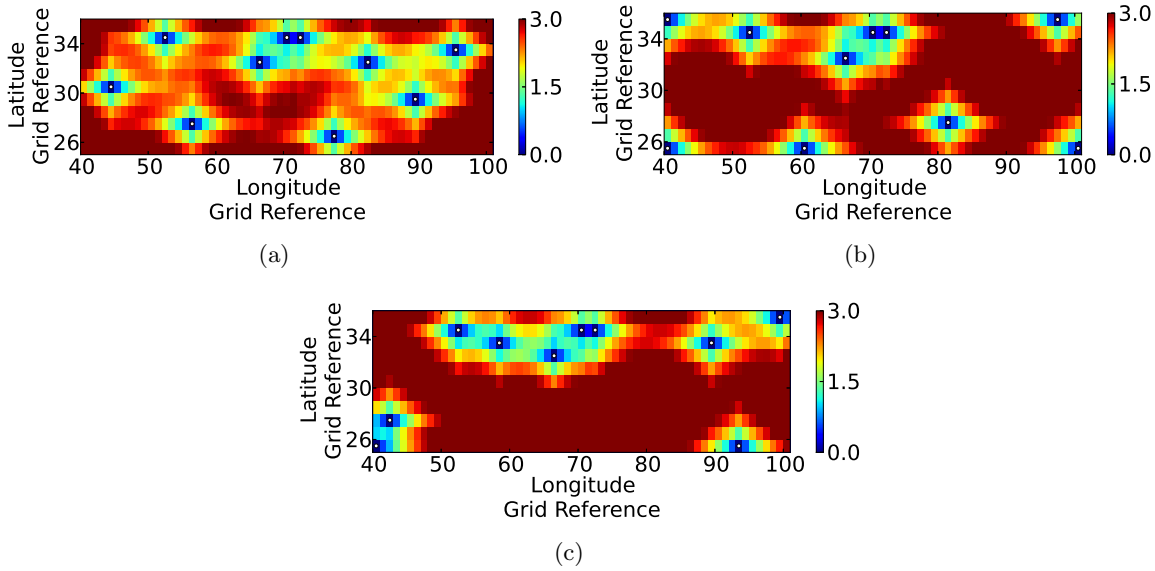


Figure 14.9: Maps of estimated error variance for a sample size of 10 for the sampling methods of (a)Updated Kriging Variance Algorithm, (b) Addition of a Point at Point of Maximum Estimated Error Variance and (c) Addition of a Point at Point of Maximum Absolute Error with sample points indicated by white dots.

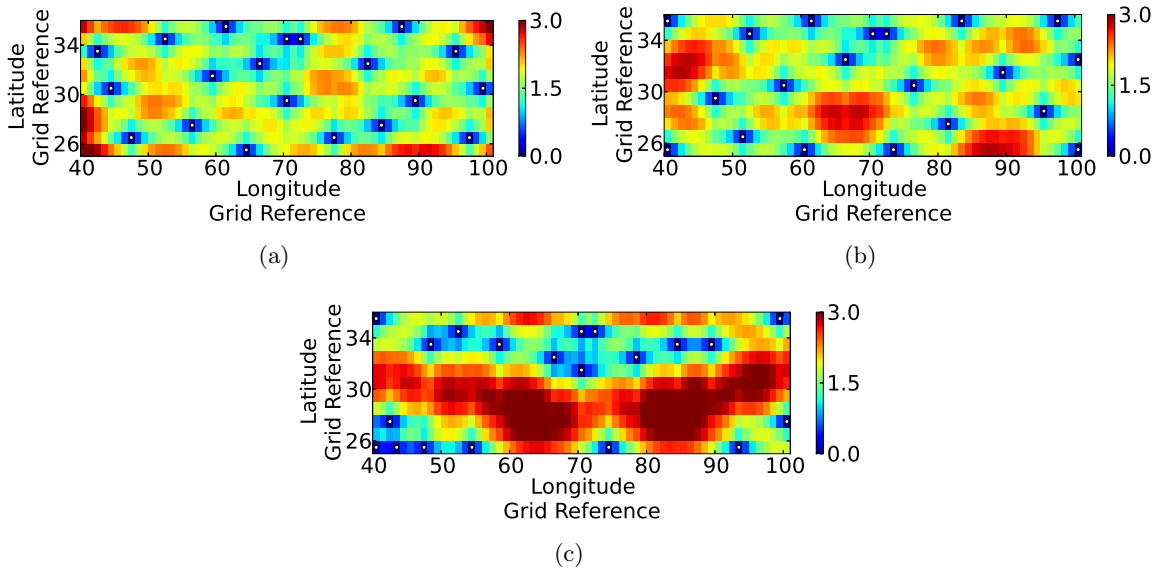


Figure 14.10: Maps of estimated error variance for a sample size of 20 for the sampling methods of (a)Updated Kriging Variance Algorithm, (b) Addition of a Point at Point of Maximum Estimated Error Variance and (c) Addition of a Point at Point of Maximum Absolute Error with sample points indicated by white dots.

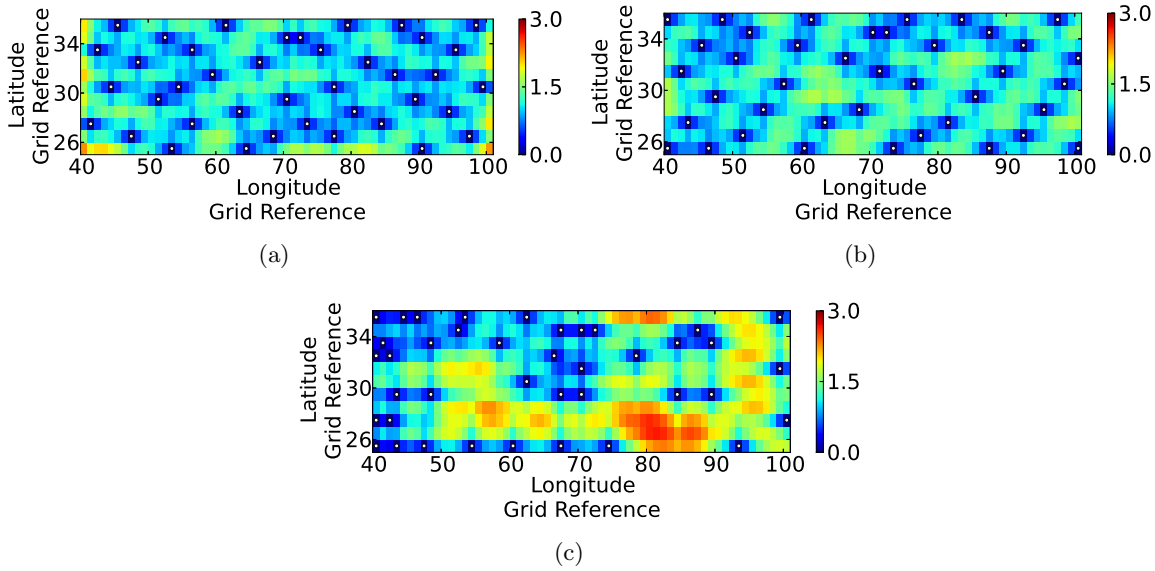


Figure 14.11: Maps of estimated error variance for a sample size of 40 for the sampling methods of (a)Updated Kriging Variance Algorithm, (b) Addition of a Point at Point of Maximum Estimated Error Variance and (c) Addition of a Point at Point of Maximum Absolute Error with sample points indicated by white dots.

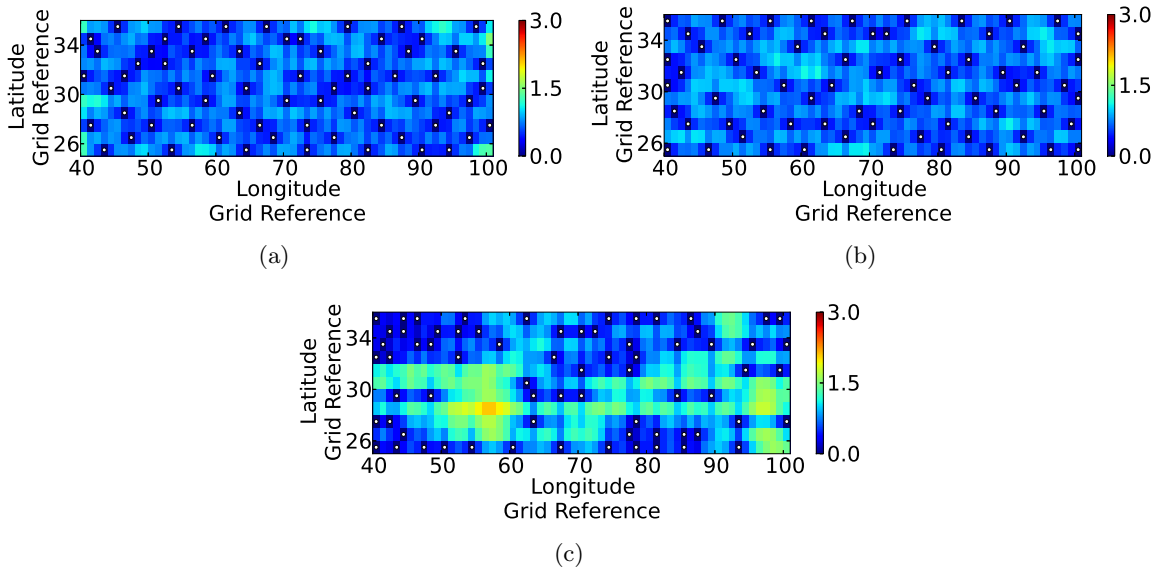


Figure 14.12: Maps of estimated error variance for a sample size of 70 for the sampling methods of (a)Updated Kriging Variance Algorithm, (b) Addition of a Point at Point of Maximum Estimated Error Variance and (c) Addition of a Point at Point of Maximum Absolute Error with sample points indicated by white dots.

as the distribution would change slightly for the various directions, based on the range in that direction. The method of Addition of a Point at Point of Maximum Absolute Error produces sample points which are more grouped than the other two methods. It is noted that this method has placed two points in the lower left corner where, if one considers a map of the true fluxes (e.g. Figure 13.17) the flux is quite variable.

Figure 14.11 illustrates the estimated error variances when the three methods under consideration have generated sample sets of 40 points. From these maps, it can be seen that both the Updated Kriging Variance Algorithm and the method of Addition of a Point at Point of Maximum Estimated Error Variance are distributing points fairly evenly across the map, although the placement of points differs and thus, the unsampled areas differ. However, the method of Addition of a Point at Point of Maximum Absolute Error clearly groups the sample points. This grouping appears to occur around areas that exhibit more change in flux values. This method also has higher estimated error variances as the grouping generates larger distances between sampled and unsampled points than the more even distributions of the other two methods.

Figure 14.12 indicates the estimated error variances produced by the three methods when they were completed, that is they had reached a sample size of 70 points. It is clear once again, that the Updated Kriging Variance Algorithm and the method of Addition of a Point at Point of Maximum Estimated Error Variance both produce relatively evenly distributed sampled points although the positioning differs. Also clear is that the method of Addition of a Point at Point of Maximum Absolute Error continues to group sample points is areas where the change in flux is more sudden (where the flux is more variable). This method thus also produces larger areas which remain unsampled. However, in spite of this, the total absolute error produced by this method is lower as it is able to more accurately model the changing flux.

14.3 Comparison of the Genetic Algorithms

If one considers the genetic algorithms presented in Chapter 13 and the histograms of errors presented in Section 14.1, it can be seen that the Genetic Algorithm designed to Minimise Total Absolute Error and the Hybridised Genetic Algorithm designed to Minimise Total Estimated Error Variance yield the lowest maximum absolute errors and lowest maximum estimated error variances respectively. The Hybridised Genetic Algorithm designed to Minimise Total Absolute Error, however, produces higher maximum absolute errors in spite of the fact that it ends on a lower total absolute error.

In terms of the total absolute error, the Genetic Algorithm designed to Minimise Total Absolute Error and the hybridised genetic algorithm with the same fitness function both experience step decreases in total absolute error across their 30000 iterations due to elitism and their fitness function. Similarly the Genetic Algorithm designed to Maximise Total Absolute Error experiences step increases in the total absolute error across its 10000 iterations. The Genetic Algorithm designed to Minimise Total Absolute Error and its respective hybridised method both show large decreases in total absolute error at the start of the algorithm but this rate of decrease drops as the number of iterations increases. The Genetic Algorithm designed to Minimise Total Estimated Error Variance and its hybridised version show continued stability or decrease in total estimated error variance. However, this does not always imply a decrease in total absolute error. If one considers Figures 13.12(b) and 13.12(f), it can be seen that for the worst performing set of the Hybridised Genetic Algorithm to Minimise Total Estimated Error Variance, the lowest total absolute error is obtained around 15000 iterations and this continues to undergo increases and decreases (ending on a higher value), while the total

estimated error variance remains stable or decreases. Once again this shows that the total estimated error variance is not always an accurate substitute for true error.

The Hybridised Genetic Algorithm designed to Minimise Total Estimated Error Variance does not produce total estimated error variances that are lower than those produced by the Genetic Algorithm designed to Minimise Total Estimated Error Variance. The total estimated error variance for this method does, however, appear to still be stepping down although the values often remain stable for many iterations, as seen in Figure 13.12(f).

14.4 Comparison of Computational Requirements

For some of the methods used in this study, it is not necessary to generate either the Kriging predictions or the estimated error variances, and in the case of the Updated Kriging Variance Algorithm, neither needs to be calculated. For this study, however, all errors were required at all iterations and as such their calculation (implying the calculation of Kriging predictions and estimated error variances) was performed at every iteration. This is unnecessary for producing the samples and as such could be done at a later stage if needed. Thus, the unnecessary calculations that were performed merely to generate the errors have been excluded in the following discussion on the computational requirements of the methods.

In order to compare the computational requirements of the methods, the algorithms have been summarised in Tables E.1 to E.10. These tables have been broken down into the description of the first iteration, the number of further iterations and the description of the remaining iterations and can be found in Appendix E. In each of these descriptions, Z is regarded to be the value of the residuals rather than that of the variable itself.

By considering these tables, it is clear that Random Sampling is the least computationally expensive. The hybridised genetic algorithms are obviously the most computationally expensive as they require the computational resources in order to run the Addition of a Point at Point of Maximum Absolute Error or Addition of a Point at Point of Maximum Estimated Error Variance, along with the resources to perform the 30000 iterations of the genetic algorithm, all of which require calculations across 20 sets of 70 points. Additionally no update algorithm is available for the covariance matrix requiring the solution of a linear system of 70 equations for each set at each iteration.

The Addition of a Point at the Point of Maximum Absolute Error is more computationally expensive than Random Sampling but less computationally expensive than the Addition of a Point at Point of Maximum Estimated Error Variance, the Updated Kriging Variance Algorithm and the genetic algorithms. It requires less matrix-vector multiplications than the Addition of a Point at Point of Maximum Estimated Error Variance, as the matrix-vector multiplication for the generation of predicted values can be done once per sample set and then merely undergo one dot-product multiplication with the covariance vector generated for a point to generate the Kriging prediction at that point, while the estimated error variance requires a matrix-vector and a dot-product multiplication at every point, as the vector of covariances at a point occurs twice in its formula.

The method of Addition of a Point at Point of Maximum Estimated Error Variance is the third least computationally expensive. It requires slightly more resources than the method of Addition of a Point at Point of Maximum Absolute Error as has been explained. However, it also requires less resources than the Updated Kriging Variance Algorithm. This is due to that fact that the Updated Kriging Variance Algorithm requires a matrix-vector and a dot-product multiplication per point not yet in the sample set and then an additional matrix-vector and dot-product multiplication for every point in the data set. This occurs as the Updated Kriging

Variance Algorithm tests each possible point for the maximum decrease in total estimated error variance before selecting a point. Notably, the Updated Kriging Variance Algorithm, Addition of a Point at Point of Maximum Absolute Error and Addition of a Point at Point of Maximum Estimated Error Variance, all make use of the updating algorithm for the inverse of the covariance matrix. While this matrix grows by a single point every iteration, it ends on a size of 69×69, as this allows the 70th point to be added. For the genetic algorithms, the size of the covariance matrix is always 70×70.

The genetic algorithms designed to maximise errors are the least expensive computationally of the genetic algorithms but purely due to the lower number of iterations (10000). These are followed by the genetic algorithms to minimise errors which run for 30000 iterations, while as explained earlier, the hybridised genetic algorithms are the most computationally expensive. The genetic algorithms are made further computationally expensive by the various genetic operators surrounding each iteration, in addition to having to calculate the absolute error or estimated error variance at every point in the data set 20 times per iteration.

In each of these classes of the genetic algorithms, the one maximising or minimising total absolute error is less computationally expensive as once again part of the formula for the prediction at each point can be solved once for a set, meaning that for each point only one dot-product multiplication is necessary.

Thus, to summarise the methods are presented in the order of ascending computationally intensity:

- Random Sampling
- Addition of a Point at Point of Maximum Absolute Error
- Addition of a Point at Point of Maximum Estimated Error Variance
- Updated Kriging Variance Algorithm
- Genetic Algorithm to Maximise Total Absolute Error
- Genetic Algorithm to Maximise Total Estimated Error Variance
- Genetic Algorithm to Minimise Total Absolute Error
- Genetic Algorithm to Minimise Total Estimated Error Variance
- Hybridised Genetic Algorithm to Minimise Total Absolute Error
- Hybridised Genetic Algorithm to Minimise Total Estimated Error Variance

14.5 Decision Regarding Three-dimensional Sampling

In each of the methods implemented, parallel computing could be implemented. However, in this study it was not used. As such, some of these methods could become time intensive on a larger and/or three-dimensional set. Also available for the Updated Kriging Variance Algorithm in two dimensions are computational speed ups. However, these are not easily adapted to three dimensions.

The Genetic Algorithm designed to Minimise Total Absolute Error and its hybridised counterpart outperform the Addition of a Point at Point of Maximum Absolute Error. However, they are much more computationally expensive and provide no order of preference amongst

the points. Thus, when considering the *a-posteriori* methods, in spite of the additional accuracy given by the genetic algorithms, the Addition of a Point at Point of Maximum Absolute Error was chosen to be implemented on the three-dimensional set. The increase in accuracy does not warrant the dramatically increased computation requirements of the Genetic Algorithm designed to Minimise Total Absolute Error or that of its hybridised version. Additionally the method of Addition of a Point at Point of Maximum Absolute Error, gives a preferential order to the points. This means that if the algorithm is run to reach a certain size but later it is determined that fewer points can be sampled in reality, then the points added up until the new size should be sampled. The genetic algorithm does not offer this feature and given the cost of sampling in the Southern Ocean, this feature may be of great value to those performing the sampling.

The Genetic Algorithm designed to Minimise Total Estimated Error Variance and its hybridised counterpart do provide a slightly lower total absolute error than the other *a-priori* methods (Addition of a Point at the Point of Maximum Estimated Error Variance and the Updated Kriging Variance Algorithm). However, they do not provide a better total absolute error. All four of the *a-priori* methods to minimise errors related to the estimated error variance provide very similar results but have varying computational requirements. Since none of the *a-priori* methods stands out in terms of total absolute error minimisation, the method with the lowest computational cost of the four is selected. Thus, the method of Addition of a Point at Point of Maximum Estimated Error Variance was chosen to be implemented on the three-dimensional set.

CHAPTER 15

THREE-DIMENSIONAL SAMPLING

Various methods of sampling were compared in Chapter 14. However, only three of those methods were implemented on the three-dimensional set. These methods were implemented on a three-dimensional set merely to prove the capability of these methods to sample large data sets. However, as the original data with the zero values removed consists of 419093 points, the decision was taken to sub-sample this set in order to reduce the computational requirements in implementing the chosen methods for a three-dimensional set. The data set was sub-sampled such that only every second point in time, latitude and longitude was taken. This sub-sampling resulted in a data set of 53502 points. The aim of this chapter is to show that these methods can be used in large scale three-dimensional sampling and the sub-sampled set is large enough to indicate that. While prediction on the original three-dimensional set would be of great value, it would also be computationally expensive and as such, would be better implemented using parallel computing on a cluster which was not available for this study.

The following methods were implemented on this sub-sampled three-dimensional set:

- Random Sampling - This is a quick and easy method of sampling although not very reliable
- Addition of a Point at Point of Maximum Absolute Error - This is the chosen *a-posteriori* method from the two-dimensional sampling
- Addition of a Point at Point of Maximum Estimated Error Variance - This is the chosen *a-priori* method from the two-dimensional sampling

The methods implemented follow the same algorithms as for the two-dimensional sampling. The only differences are the total number of points and the final sample size. The prediction procedure is also identical to that found in Chapter 8.

Each of these methods selects 2675 points from the data set, this is approximately 5% of the data. For the chosen *a-posteriori* and *a-priori* methods, a starting set of 4 points is randomly selected. These same starting points are used by both methods, that is set one for both of these methods starts using the same 4 points and similarly for set two.

15.1 Implementation of Sampling Methods on Three-dimensional Data

The errors for the three methods implemented for 2675 points can be seen in Tables 15.1 and 15.2. The units associated with the maximum absolute error, total absolute error and total integrated difference are mol(C)/m^{2}s.

Set	Maximum Absolute Error	Total Absolute Error	Percentage Absolute Error	Total Integrated Error
		Random Sampling		
1	12.74	35549.38	21.02%	−1160.83
2	12.99	35481.89	20.98%	−1096.53
	Addition of a Point at Point of Maximum Absolute Error			
1	2.35	24230.39	14.33%	−1907.75
2	1.80	24207.60	14.32%	−183.50
Addition of a Point at Point of Maximum Estimated Error Variance				
1	10.70	31355.20	18.54%	−123.36
2	9.40	31373.88	18.55%	−554.12

Table 15.1: Final maximum absolute, total absolute, percentage absolute and total integrated errors associated with sampling in 3D.

Set	Maximum Estimated Error Variance	Total Estimated Error Variance
	Random Sampling	
1	3.84	52812.12
2	4.16	53241.83
	Addition of a Point at Point of Maximum Absolute Error	
1	4.30	74641.82
2	4.46	74806.62
Addition of a Point at Point of Maximum Estimated Error Variance		
1	1.31	43615.67
2	1.30	43642.56

Table 15.2: Final maximum estimated error variances and total estimated error variances associated with sampling in 3D.

From the errors in Tables 15.1 and 15.2, it is clear that Addition of a Point at Point of Maximum Absolute Error provides the lowest maximum absolute errors, total absolute errors and thus, percentage absolute errors. It does however, provide the highest maximum estimated error variances and total estimated error variances. Addition of a Point at Point of Maximum Estimated Error Variance provides the lowest estimated error variance related errors. However, absolute errors fall in between those of Random Sampling and Addition of a Point at Point of Maximum Absolute Error. Random Sampling provides the worst percentage absolute errors of three methods implemented. Once again, it can be seen that the total integrated error is a poor indicator of performance.

Additionally for each set run, the final number of sample points per time slice can be seen in Table 15.3. Naturally for the two sets generated by Random Sampling, no preference or emphasis is placed on any time slice. Addition of a Point at Point of Maximum Estimated Error Variance places more sample points/more emphasis on the first and last time slice and

Sampling Method	Random Sampling		Addition of a Point at Point of Maximum Absolute Error		Addition of a Point at Point of Maximum Estimated Error Variance	
Time Slice	Set 1	Set 2	Set 1	Set 2	Set 1	Set 2
0	78	81	180	195	177	178
2	79	68	130	125	78	79
4	62	78	114	127	68	72
6	63	74	101	104	69	71
8	71	63	103	91	70	64
10	67	80	90	99	69	70
12	66	61	89	84	64	66
14	58	73	74	75	67	70
16	65	93	71	74	61	60
18	60	73	83	81	63	66
20	83	67	99	84	64	67
22	66	70	49	57	73	63
24	78	77	81	76	65	62
26	87	64	59	56	64	67
28	80	85	48	44	64	67
30	87	66	55	46	61	65
32	82	70	58	46	66	60
34	77	80	50	46	65	62
36	82	85	29	25	62	65
38	67	53	40	46	63	62
40	84	73	28	27	70	63
42	75	74	32	26	64	70
44	70	73	34	34	64	60
46	73	73	41	45	62	62
48	73	62	42	41	60	67
50	69	65	39	34	66	64
52	63	77	52	48	65	61
54	78	55	44	52	69	68
56	51	68	81	72	65	62
58	47	73	45	56	65	73
60	89	64	63	60	68	61
62	82	69	58	68	66	63
64	71	75	74	80	63	75
66	86	93	83	75	70	66
68	72	82	105	118	64	73
70	60	73	106	105	85	73
72	74	65	145	153	176	178

Table 15.3: Sample points per time slice in 3D sampling.

distributes the points relatively evenly over the remainder of the time slices with a slight emphasis on time slices closer to the first and last time slices. This emphasis on the first and last time slices occurs as this study did not allow the data to wrap around temporally. This makes the first and last time slices extremes for the data. The method tries to distribute the sample points relatively evenly based on the temporal range parameter, as can be seen by the relatively even number of sample points per time slice across the majority of the time slices. However, the distances to the points in the first and last time slices will clearly be larger in general and this implies larger estimated error variances in general. Thus, more points are placed in the first and last time slices as they tend in general to have larger estimated error

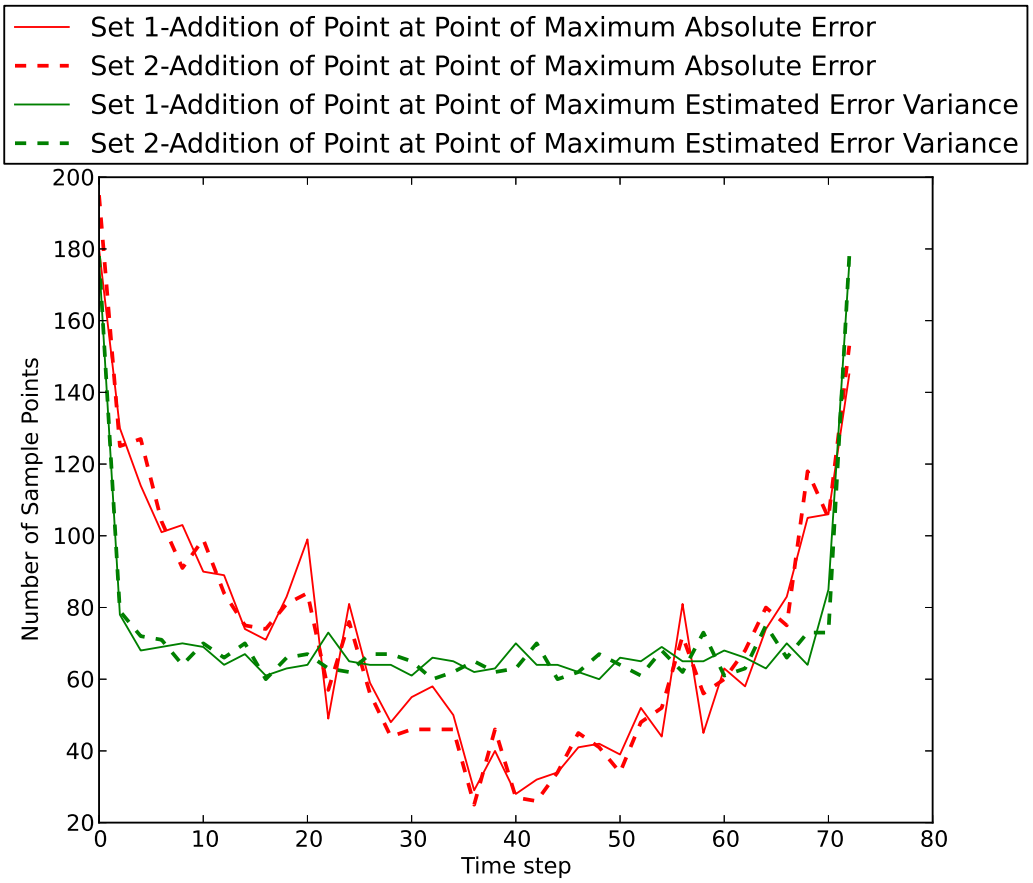


Figure 15.1: Sample points per time slice for sets one and two for Addition of a Point at Point of Maximum Absolute Error and Addition of a Point at Point of Maximum Estimated Error Variance in 3D sampling.

variances.

Addition of a Point at Point of Maximum Absolute Error places extreme emphasis on some time slices such as the first, second, fourth, sixty-eighth, seventieth and last time slices while very little emphasis is placed on time slices such as thirty-sixth, fortieth, forty-second and forty-fourth time slices. The predictions at a point are dependent on both the distances between the sample points and that point (through the weights) and the values of the sample points. This dependence on the distances helps account for why the starting and ending time slices have more emphasis. Points at the start and end tend to be further away temporally resulting in larger distances. These larger distances imply that the covariances will be smaller. A smaller covariance implies less correlation between the sample points and the point at which estimation occurs. This results in higher errors. Additionally, those maps per time slice which show more variability, tend to have more sample points placed in that time slice as the more variable the values are, the higher the errors tend to be unless a lot of information regarding the variability is found in the sample point values. Those time slices where very few sample points are allocated, tend to be the least variable of the time slices.

The final number of sample points per time slice for Addition of a point at Point of Maximum Absolute Error and Addition of a Point at Point of Maximum Estimated Error Variance for both set one and two can be seen in Figure 15.1. It is clear from this figure that as previously discussed, the first and last time slices are favoured. However, it is possible that for the method of Addition of a Point at Point of Maximum Absolute Error, some of the emphasis on these time slices may also be linked to increased activity in the oceans at these time slices. In the summer months, the Southern Ocean exhibits more activity than in the winter

months and this may account for the placement of some of the points in these time slices. Additionally, this figure also demonstrates clearly how evenly distributed the sample points are when using the method of Addition of a Point at Point of Maximum Absolute Error. Addition of a Point at Point of Maximum Absolute Error, clearly has time slices with more sample points located within them and this is due to the activity and variability at those time slices.

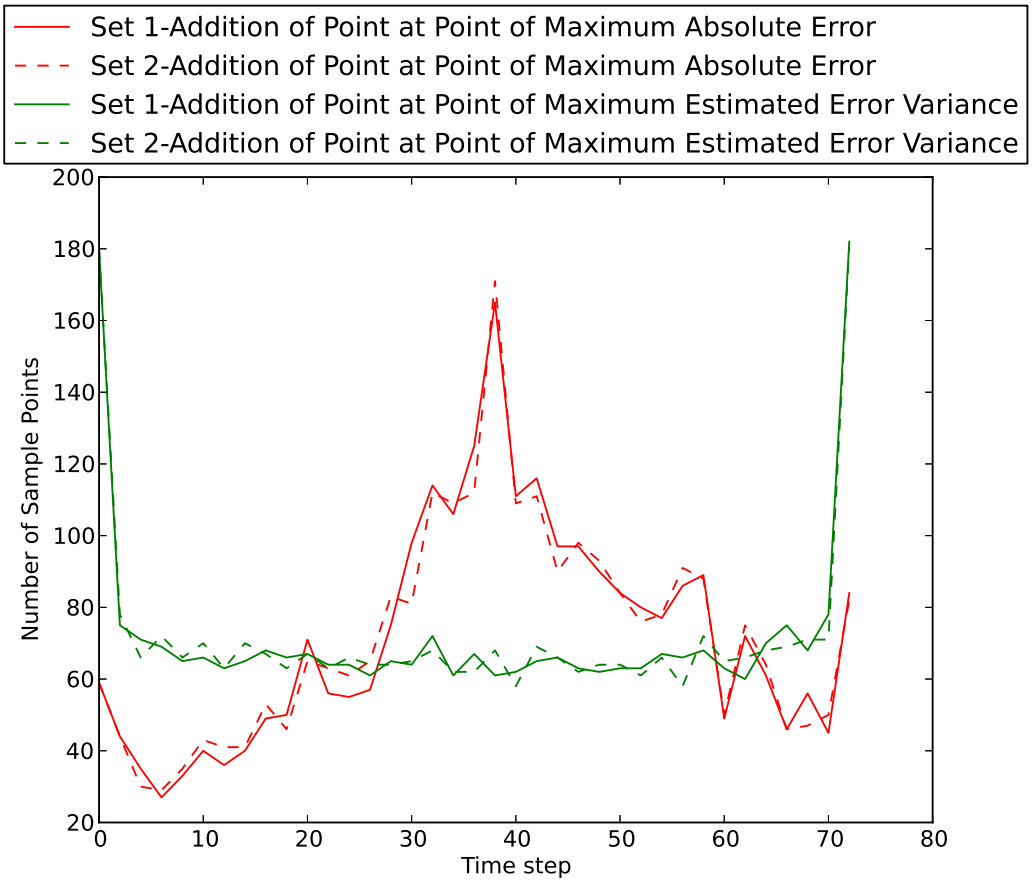


Figure 15.2: Sample points per time slice for sets one and two for Addition of a Point at Point of Maximum Absolute Error and Addition of a Point at Point of Maximum Estimated Error Variance in 3D sampling with the modified time step labels.

In order to clarify the affect of higher activity/higher variability in carbon dioxide flux values on the placement of sample points, the time step reference values were modified such that the most active period of the year now occurs in the centre of the temporal range. The starting set of sample points used for the original runs were also modified to reflect the new time step labelling and used as the starting sets once again. The resulting errors for the new time step labelling are similar to those obtained under the original labelling. However, the new number of points per time slice can be seen in Figure 15.2.

For the original time step labels, the high variability period coincides with the temporal extremes of the data set, thus, we cannot distinguish between the emphasis placed on these time steps due to the variability in the data or due to extreme distances within the covariances. Considering Figure 15.2, it is seen that the emphasis on the extremes of the data set is much less prevalent when the high variability period falls in the temporal centre of the data set. This high variability period does still receive emphasis for the Addition of a Point at Point of Maximum Absolute Error method. This suggests that for Addition of a Point at Point of Maximum Absolute Error that the variability of the data values has more prominent influence than the extreme distances within the covariances. The placement of sample points by Addition of a Point at Point of Maximum Estimated Error Variance remains largely unaltered

with the emphasis occurring at the temporal extremes of the data set and a relatively even distribution of points over the remaining of the time steps.

Figure 15.3 demonstrates how the various errors change with the addition of points within the methods of Addition of a Point at Point of Maximum Absolute Error and Addition of a Point at Point of Maximum Estimated Error Variance.

Figures 15.3(a) and 15.3(e) show the changes in maximum absolute error and maximum estimated error variance respectively. The notable feature here is that the differences in these errors between the methods is larger than it was in two-dimensional sampling. However, as there is now a third dimension and greater distances are possible between points, this is not totally unexpected. Figure 15.3(c) illustrated that it requires many more points for Addition of a Point at Point of Maximum Estimated Error Variance to fall below 20% absolute error than it does for the method of Addition of a Point at Point of Maximum Absolute Error. The figures of the changes of the remaining errors, as points are added to the sample set, show similar trends to the results for the same methods in two dimensions.

Figures 15.4 to 15.12 are maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux, the estimated error variance and the true error across the full latitude and longitude grid reference ranges for time slices 0, 40 and 72 for the three different methods for set two. Time slices 0 and 72 were chosen due to emphasis (in terms of number of sample points) placed on them by the methods which perform the addition of one point per iteration. Time slice 40 was chosen as it was given an average emphasis by the Random Sampling and Addition of a Point at Point of Maximum Estimated Error Variance and a very low emphasis by the method of Addition of a Point at Point of Maximum Absolute Error. Appendix F contains similar maps for time slices 20 and 60 for set two and for time slices 0, 20, 40, 60 and 72 for set one.

If one considers Figures 15.4 to 15.6 which show time slice zero, it is clear that this time slice is quite variable. Additionally, it can be seen that Addition of a Point at Point of Maximum Absolute Error best predicts this time slice, while Random Sampling worst predicts it. Considering the sample point locations, indicated by white dots in the estimated error variance maps, it can be seen that Random Sampling has no preferential placement. Addition of a Point at Point of Maximum Estimated Error Variance places the points with a relatively even distribution, while Addition of a Point at Point of Maximum Absolute Error groups the sample points in more variable regions. The same can be seen in Figures 15.10 to 15.12 which are the maps for the last time slice.

If the maps in Figures 15.7 to 15.9 are considered, all three methods make decent predictions for this time slice, although none are perfectly matched to the true flux values. Random Sampling is the worst predictor of the three methods, while visually it is difficult to decide whether the Addition of a Point at Point of Maximum Absolute Error or the Addition of a Point at Point of Maximum Estimated Error Variance provides a better prediction. However, it is noted that the Addition of a Point at Point of Maximum Absolute Error produces a very similar map with 36 less sample points in this time slice. The locations of sample points follows the previously mentioned trends once again.

Further considering Figures 15.8(c) and 15.8(d), it is clear that in the areas of highest estimated error variance, the true error is not at its largest. This indicates that the estimated error variance is not a good surrogate for the true error or the absolute value of the true error.

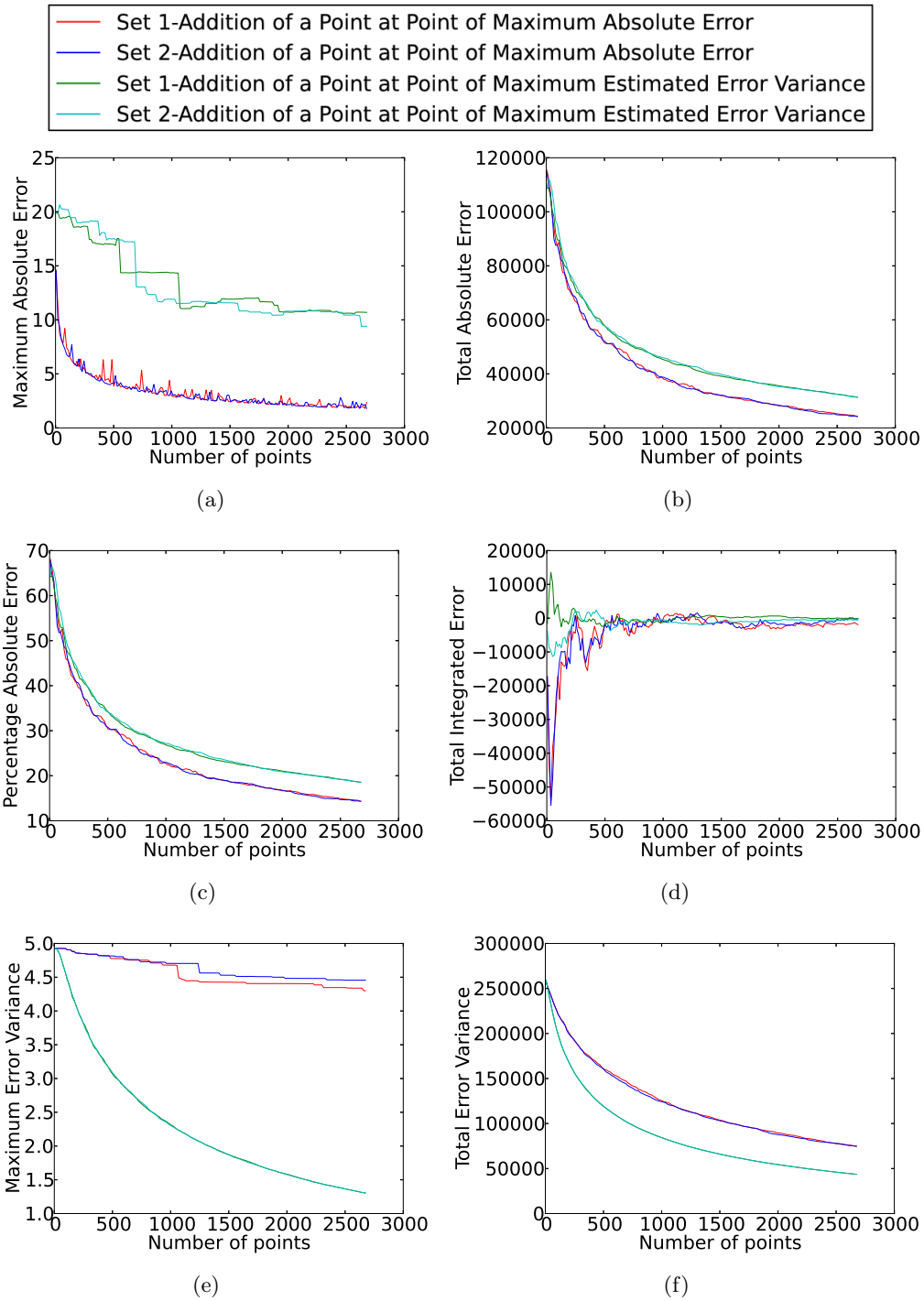


Figure 15.3: Progression of (a) maximum absolute error, (b) total absolute error, (c) percentage absolute error, (d) total integrated error, (e) maximum estimated error variance and (e) total estimated error variance across sets one and two generated using Addition of a Point at Point of Maximum Absolute Error and Addition of a Point at Point of Maximum Estimated Error Variance.

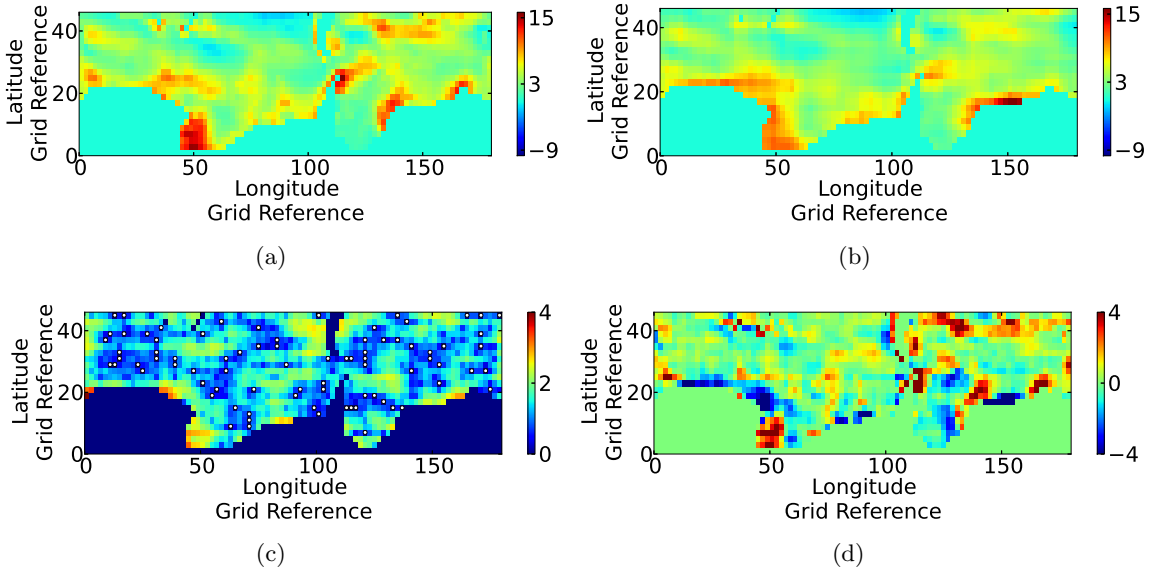


Figure 15.4: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 0 of set two generated using Random Sampling. The white dots in (c) indicate the sampling points.

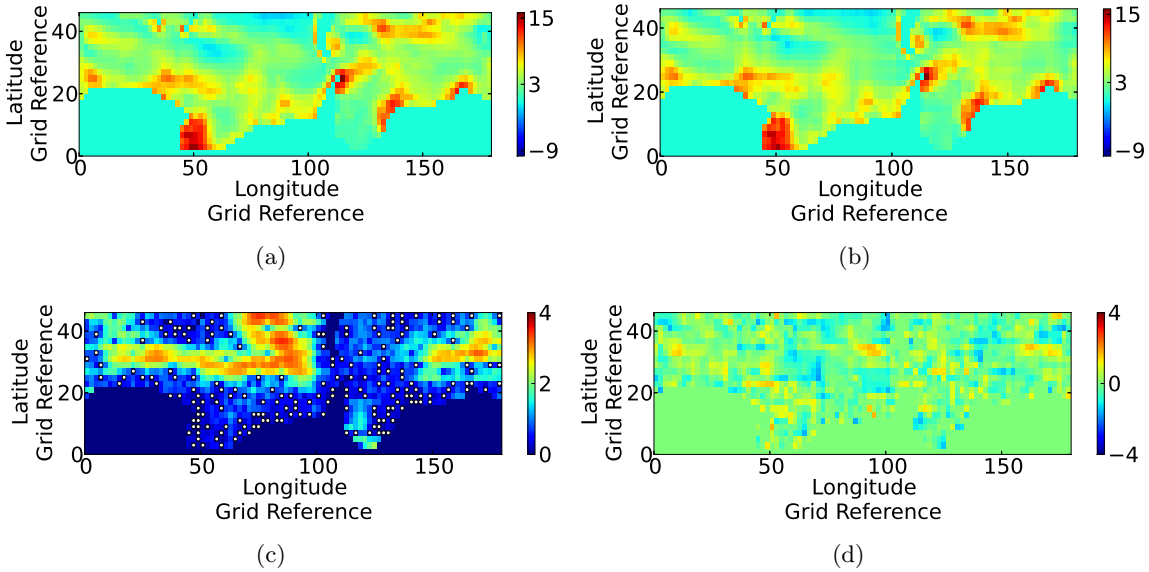


Figure 15.5: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 0 of set two generated using Addition of a Point at Point of Maximum Absolute Error sampling. The white dots in (c) indicate the sampling points.

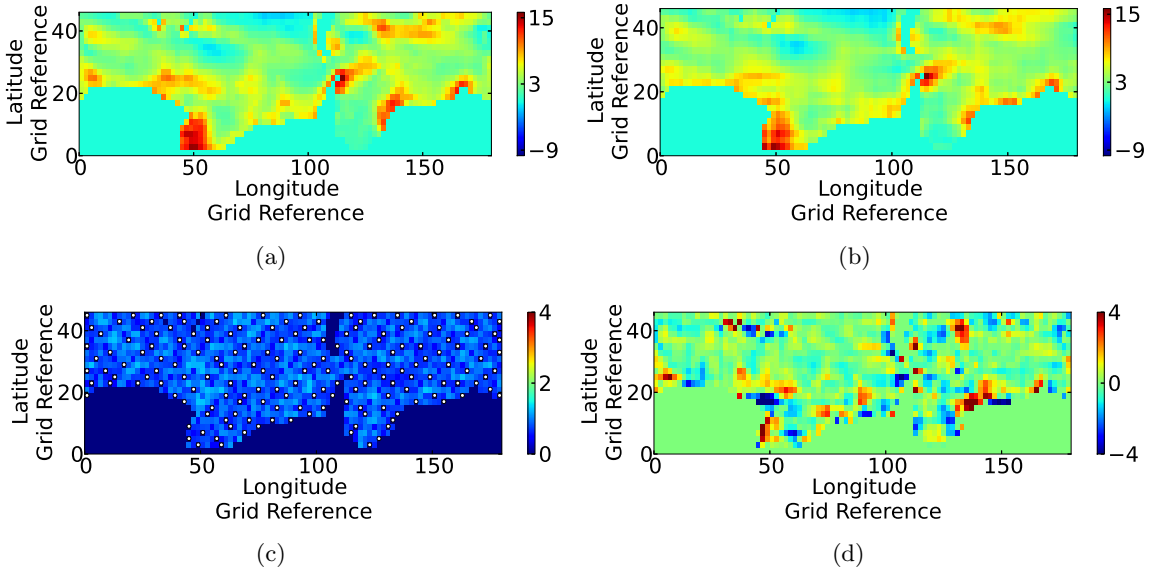


Figure 15.6: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 0 of set two generated using Addition of a Point at Point of Maximum Estimated Error Variance sampling. The white dots in (c) indicate the sampling points.

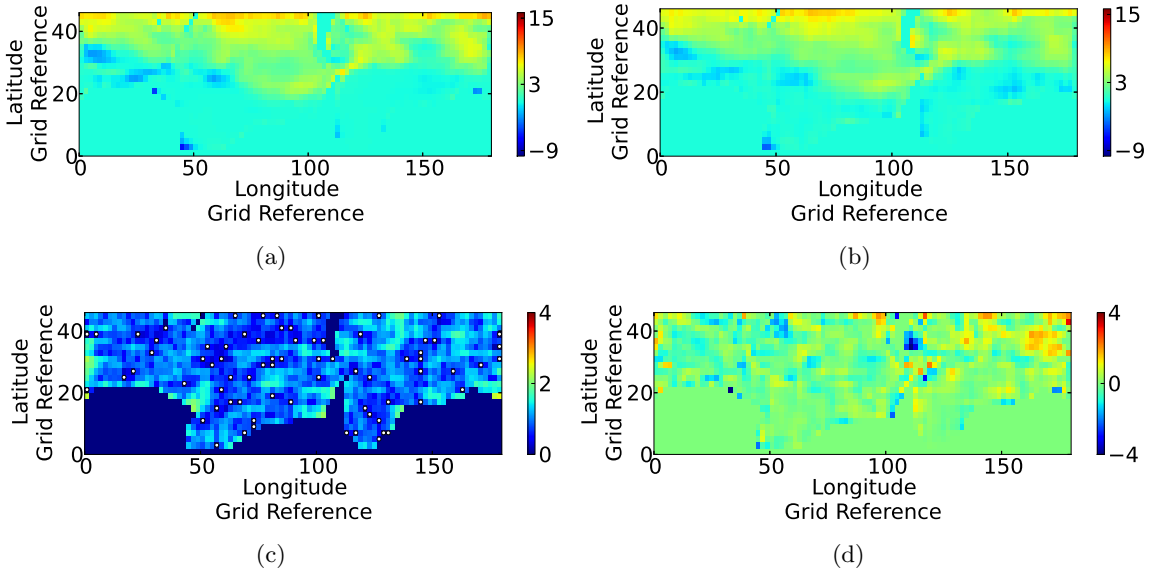


Figure 15.7: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 40 of set two generated using Random Sampling. The white dots in (c) indicate the sampling points.

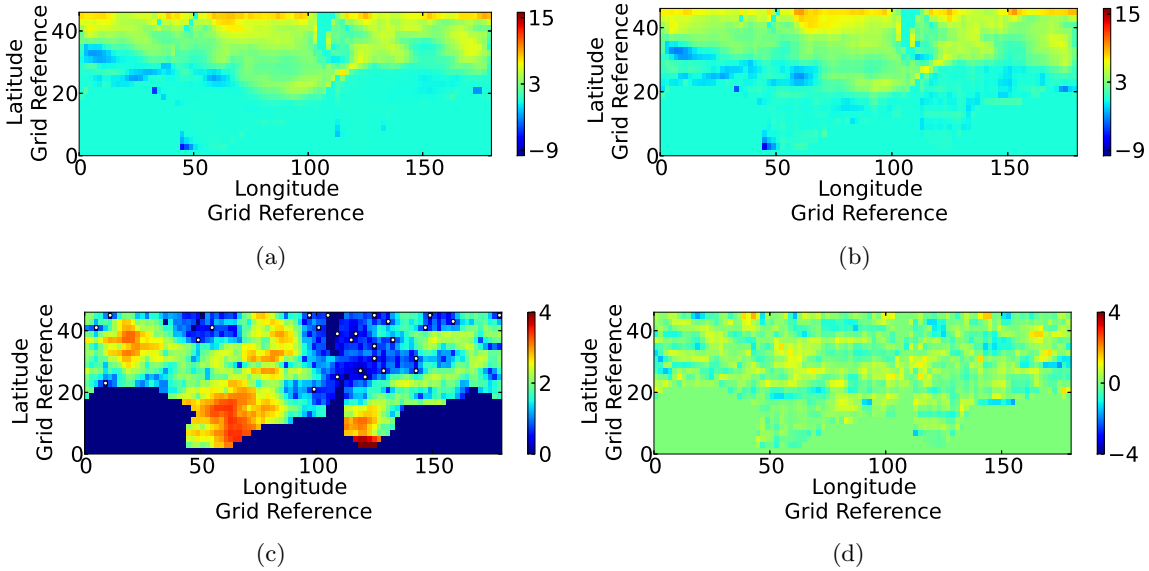


Figure 15.8: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 40 of set two generated using Addition of a Point at Point of Maximum Absolute Error sampling. The white dots in (c) indicate the sampling points.

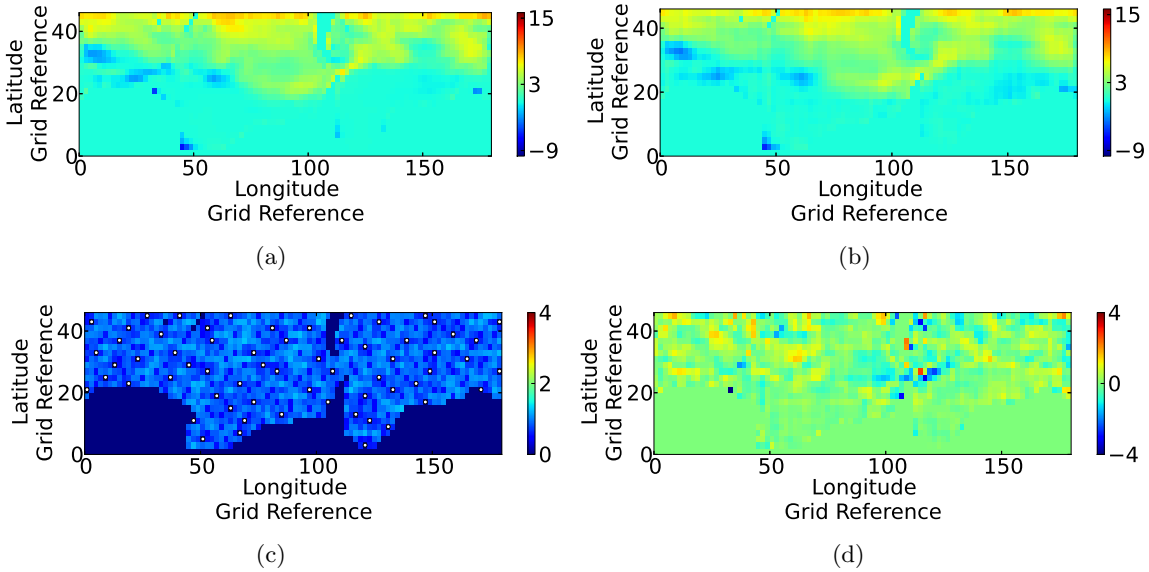


Figure 15.9: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 40 of set two generated using Addition of a Point at Point of Maximum Estimated Error Variance sampling. The white dots in (c) indicate the sampling points.

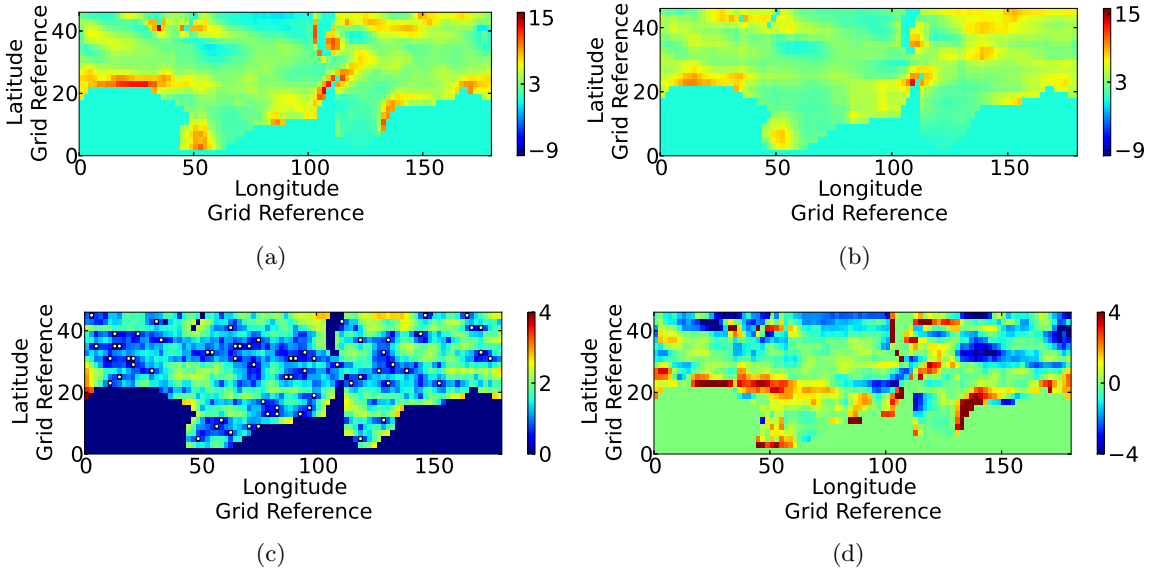


Figure 15.10: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 72 of set two generated using Random Sampling. The white dots in (c) indicate the sampling points.

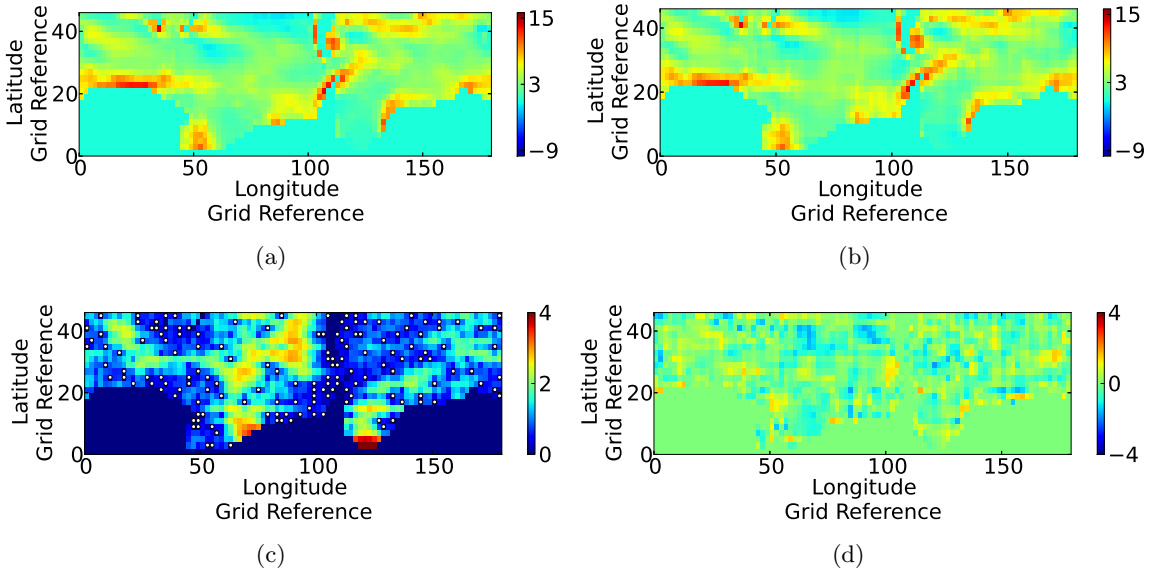


Figure 15.11: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 72 of set two generated using Addition of a Point at Point of Maximum Absolute Error sampling. The white dots in (c) indicate the sampling points.

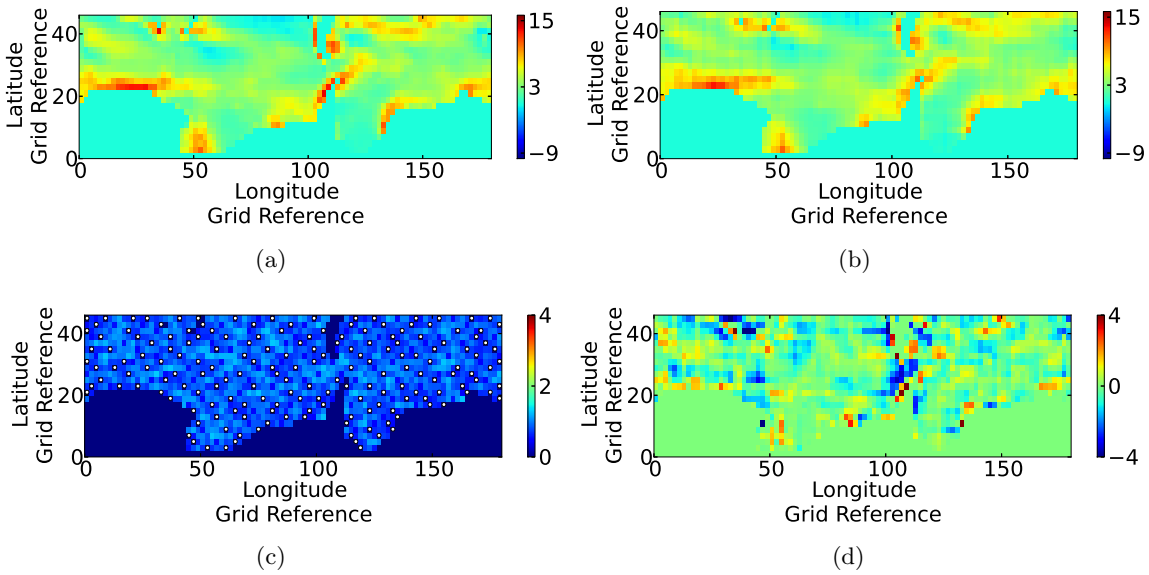


Figure 15.12: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 72 of set two generated using Addition of a Point at Point of Maximum Estimated Error Variance sampling. The white dots in (c) indicate the sampling points.

15.2 Limitations

Each method was only implemented twice. The methods were implemented using only the single chosen trend and covariance structure. This covariance structure is separable and a non-separable covariance structure may yield better results. The methods were only allowed to choose approximately 5% of the data and results would improve if more data was selected.

CHAPTER 16

CONCLUSIONS

When the sampling methods tested on the two-dimensional subset (which were used to select approximately 10% of the subset) are considered, it is clear that the *a-posteriori* methods outperform the *a-priori* methods. It was found that the Hybridised Genetic Algorithm designed to Minimise Total Absolute Error and its respective non-hybridised version yielded the lowest total absolute errors. The percentage absolute errors produced by Hybridised Genetic Algorithm designed to Minimise Total Absolute Error were between 2.68% and 3.05%. The non-hybridised genetic algorithm with the same fitness function, produced percentage absolute errors between 2.10% and 3.48%. The Hybridised Genetic Algorithm designed to Minimise Total Absolute Error was the most computationally expensive of the *a-posteriori* methods, while the non-hybridised version was only slightly less expensive.

The remaining *a-posteriori* method designed to reduce errors, Addition of a Point at Point of Maximum Absolute Error, produces percentage absolute errors of between 3.37% and 4.16% while remaining significantly less computationally expensive than the other *a-posteriori* methods. This lead to the conclusion that the Addition of a Point at Point of Maximum Absolute Error method produced the best trade-off between computationally intensity and minimisation of total absolute error amongst the *a-posteriori* methods.

Random Sampling achieved percentage absolute errors in the range of 7.04% to 9.08% and while this method was also the least computationally expensive, it is known to be unpredictable in that it may produce errors across the full range from the best possible to the worst possible errors, making it unreliable. This range covers at the very least from the best attained percentage error of 2.10% to the worst attained percentage error of 64.37%, attained by the Genetic Algorithm designed to Maximise Total Absolute Error.

Of the *a-priori* methods tested which relied solely on the estimated error variance, the Addition of a Point at Point of Maximum Estimated Error Variance was the least computationally expensive, followed by the Updated Kriging Variance Algorithm, then the Genetic Algorithm designed to Minimise Total Estimated Error Variance. Finally the most computationally expensive *a-priori* method was the Hybridised Genetic Algorithm designed to Minimise Total Estimated Error Variance. Addition of a Point at Point of Maximum Estimated Error Variance achieved percentage absolute errors in the range of 5.63% to 6.67%, while those for the computationally more expensive Updated Kriging Variance Algorithm fell within the range of 6.08% to 7.30%. The Genetic Algorithm designed to Minimise Total Estimated Error Variance produced percentage absolute errors in the range of 5.98% to 6.21% while its hybridised counterpart generated percentage absolute errors in the range of 6.05% to 6.39%. Given the very small difference in errors produced by the various *a-priori* methods, the computationally least expensive Addition of a Point at Point of Maximum Estimated Error Variance was deemed to be the better of the methods for use in three-dimensional sampling.

Of interest is that the *a-posteriori* methods attaining the lower percentage absolute errors, select sample points such that the areas of higher variability are more densely sampled than those of lower variability, while the *a-priori* methods designed to minimise errors which rely solely on the estimated error variance, select the sample points such that there is a relatively even distribution across the area of interest.

When the total absolute errors and total estimated error variances were compared, it was noted that a lower total estimated error variance did not guarantee a lower total absolute error. When the results from the Genetic Algorithm designed to Maximise Total Estimated Error Variance were considered, it was seen that a run with a resulting percentage absolute error of 24.46% had a total estimated error variance of 2389.88, while a run with a resulting percentage error of 48.59% had a total estimated error variance of 2382.03. In this case, a difference of approximately 0.33% in total estimated error variance resulted in difference of approximately 24% in percentage absolute error. In addition, the run with the lower total estimated error variance produced a higher percentage absolute error. However, in the case of the Genetic Algorithm designed to Maximise Total Absolute Error, the order of runs in terms of lower total estimated error is also the order of the runs in terms of lower total absolute error. This appears to be an exception rather than a rule in general though as in many other instances, this is not the case.

In fact, the results suggest that one can incur a range of total absolute errors for the same total estimated error variance. This is not unexpected, however, as estimated error variances relies solely on the distances between points and the covariance structure. Thus, one could experience the same estimated error variance at two unsampled points (if they are the same distances from the sampled points) and they would be assigned the same predicted value. However, it is possible that they would have quite different true values, resulting in a different absolute error. This explains not only why the same total estimated error may be seen with different total absolute errors but also why conversely the same total absolute error may be found to have different total estimated error variances. Thus, it is concluded that the estimated error variance is not an accurate surrogate for the true absolute error. However, if the sampling is to be done *a-priori*, then the estimated error variance is an easily obtainable error measure. This study does, however, recommend that if the sampling is being performed *a-posteriori* (such as for data minimisation), that the absolute error be used to obtain more accurate estimations. Additionally, if modelled data is available for the parameter to be sampled, one could consider making use of that data for *a-posteriori* sampling strategies to obtain points for sampling in the real world.

One additional benefit of Addition of a Point at Point of Maximum Absolute Error, Addition of a Point at Point of Maximum Estimated Error Variance and the Updated Kriging Variance Algorithm is that these methods have an order of preference within the sampling. This order of preference is not available within the genetic algorithms. This order of preference allows one to sample a smaller percentage of the data than the algorithm was run up until by merely selecting the points in the order they were chosen by the algorithm until the new smaller percentage is reached instead of having to run the algorithm again for the smaller percentage. Additionally, if a larger percentage is to sampled, the algorithm can be started with the current percentage's sample locations as the starting set with the end number of sample locations determined by the new larger percentage.

The results for sampling from the three-dimensional subset up to approximately 5% of the subset can be summarised as follows. The Random Sampling produced percentage absolute errors of 21.02% and 20.98% and total estimated error variances of 52812.12 and 53241.83 respectively. Addition of a Point at Point of Maximum Estimated Error Variance produced percentage absolute errors of 18.54% and 18.55% with respective total estimated error variances of 43615.67 and 43642.56. While Addition of a Point at Point of Maximum Absolute

Error produced the lowest percentage absolute errors of 14.33% and 14.32% with total estimated error variances of 74641.82 and 74806.62 respectively. It is clear that as with the two-dimensional sampling, the *a-posteriori* method outperforms the *a-priori* method and that a lower total estimated error variance does not imply a lower total absolute error as, the *a-posteriori* method also generates the highest total estimated error variances. Thus, using the covariance and trend structures and parameters selected for this study, the *a-priori* method selected for use on the three-dimensional subset achieved a percentage absolute error of below 20% when selecting approximately 5% of the data, while the *a-posteriori* method was able to bring the percentage absolute error to below 15% for the same sample size.

16.1 Summary of Contributions

This study has shown that at least for this data set, the estimated error variance does not provide a good surrogate for absolute error. The results also clearly indicate that in cases where the true error is available, such as data minimisation, the absolute error should be used as the measure of error. While this does not follow the decision of Brodkin [8], he did not conduct a comparison of error measures before making use of Addition of a Point at Point of Maximum Estimated Error Variance. The conclusions of this study regarding estimated error variance as a surrogate for absolute error are as follows:

- If the variable is known at all points, in cases such as data minimisation, absolute error should be used as the error measure.
- If a model of the data is available, or an expert on the variable of interest is available, the knowledge gained from these sources should be used to determine areas of higher variability. These areas should be more densely sampled than the areas of lower variability.
- If there are minimal or no available observations, then the estimated error variance may be used as a surrogate for absolute error. While the estimated error variance does not provide a good surrogate for absolute error, if there are no data available, it provides an easy to calculate error measure.

This study concludes that based on the 2D sampling, that the genetic algorithms designed to minimise total absolute error have shown that the Addition of a Point at Point of Maximum Absolute Error would prove to be a good validation sampling method for future studies.

Additionally, this study has shown that there are methods that are relatively computationally inexpensive (both *a-priori* and *a-posteriori*) which can be used to sample a large three-dimensional data set and produce reasonably accurate results which can be seen from the predicted maps for three-dimensional sampling in Chapter 15. As the data set used is modelled data for the carbon dioxide flux in the Southern Ocean, it is hoped that these methods may help inform new methods or provide starting points for sampling in the Southern Ocean.

16.2 Suggestions for Further Research

It is suggested that the following be completed in order to further this research:

- Data Set - It is suggested that the three-dimensional portion of this study be repeated on the complete data set. This would require more computational power but given that the methods could be adapted for parallel computing, the time taken for completion could be reduced.
- Parallel Computing - Regardless of size of the data set, it is advised that the methods presented in this study be coded for parallel computing to reduce the computational time of the methods.
- Data Wrap Around - It is suggested that in terms of time and longitude, the code is reproduced such that the first and last points are considered to be sequential.
- Trend - More complex trends should be considered.
- Covariance Structure - It is suggested that the analysis be performed for various other conditions on the covariance structure. These include but are not restricted to the inclusion of a nugget term, the use of non-separable covariance functions and the use of anisotropy.
- Tapering - In order to reduce the computational burden, it is suggested that a tapering function (such as the Wendland taper) be used in order to reduce the covariance matrix to a sparse matrix. This would allow sparse solvers to be used and would reduce storage for the matrix.
- Genetic Operators - If more complex genetic operators are considered, there is the possibility that the sample set will converge to a near-optimal set in a more timely manner.
- Estimation Updating - There are update equations available for the Kriging estimations and estimated error variances which if used could reduce the computational requirements of some of the methods used. Only the update equations for the covariance matrix were used in this study.
- Moving Window Technique - This technique allows for the model and parameters to be adapted for a local area based on the data within a window [56]. This technique has the advantage of an expected increase in accuracy. As one would not expect the model and parameters to be constant over the entire Southern Ocean (as has been assumed in this study), this would be the next logical step in the modelling of this data. Use of the moving window technique would however, increase the complexity of the work. Additionally, Journel and Rossi [36] provide recommendations for trend removal in this scenario.
- Sample Sensitivity Error Sampling - It may be worthwhile testing the sample sensitivity error method of sampling against the methods used in this study. This method requires that every point in the data set be estimated as many times as there are points in the sample set, as it must be estimated using the sample set with one sample point removed. This must be done for each sample point in the sample set [66]. These values must then be averaged before the next sample point can be selected [66]. This suggests it may be computationally expensive on large data sets. As the focus of this study was to determine if the estimated error variance is a good surrogate error for the absolute error, this method was not tested. The only drawback of this method for sampling in reality is that it would require the new sample value to be acquired before the next sample point could be selected. This would make this method infeasible in some sampling environments, especially if the data under consideration is spatio-temporal. However, as this method does not require the true values at all locations, only those already sampled, it would be of value to compare its performance and computational expense to the that of the methods tested in this study.

BIBLIOGRAPHY

[1] O. Aumont. Pisces biogeochemical model, 2005.

[2] O. Aumont and L. Bopp. Globalizing results from ocean in situ iron fertilization studies. *Global Biogeochemical Cycles*, 20, 2006.

[3] D. Bakker, H. De Baar, and U. Bathmann. Changes of carbon dioxide in surface waters during spring in the southern ocean. *Deep Sea Research II*, 44:91–127, 1996.

[4] S. Bandyopadhyay and S. K. Pal. *Classification and Learning using Genetic Algorithms: Applications in Bioinformatics and Web Intelligence*. Springer, 2007.

[5] R. J. Barnes and A. G. Watson. Efficient updating of kriging estimations and variances. *Mathematical Geology*, 24:129–133, 1992.

[6] S. Baum. Ocean/atmosphere circulation modeling projects. `http://stommel.tamu.edu/~baum/ocean_models.html#OPA`, 2004. last accessed: 18 February 2014.

[7] G. Bohling. Hydrogeophysics:theory, methods, and modeling. `http://people.ku.edu/~gbohling/BoiseGeostat`, 2007. last accessed: 09 January 2012.

[8] C. Brodkin. The application of kriging for the controlled minimization of large data sets. Master's thesis, Airforce Institute of Technology Air University, 1991.

[9] D. J. Brus and G. B. M. Heuvelink. Optimization of sample patterns for universal kriging of environmental variables. *Geoderma*, 138:86–95, 2007.

[10] L. Chen, S. Xu, Z. Gao, H.Chem, Y. Zhang, and J. Zhan. Estimation of monthly air-sea CO˙2 flux in the southern atlantic and indian ocean using in-situ and remotely sensed data. *Remote Sensensing of Environment*, 115, 2011.

[11] Scipy community. Statistical function (scipy.stats). `http://docs.scipy.org/doc/scipy-0.13.0/reference/stats.html`, 2013. last accessed: 22 May 2014.

[12] R. D. Cousens, R. W. Brown, A. B. McBratney, and M. Moerkerk. Sampling strategy is important for producing weed maps: a case study using kriging. *Weed Science*, 50:542–546, 2002.

[13] N. Cressie. The origins of kriging. *Mathematical Geology*, 22:239–252, 1990.

[14] N. Cressie and C. K. Wikle. *Statistics for Spatio-Temporal Data*. John Wiley and Sons Inc, 2011.

[15] N. A. C. Cressie. *Statistics for Spatial Data*. John Wiley and Sons Inc, revised edition, 1993.

[16] H. J. W. de Baar and E. Suess. Ocean carbon cycle and climate change - an introduction to the interdisciplinary union symposium. *Global and Planetary Change*, 8:7–11, 1993.

[17] C. De-Vitry. Resource classification - a case study from the joffre-hosted iron ore of bhp billiton's mount whaleback operations. *Mining Technology*, 112:185–196, 2003.

[18] F. Deng and J. Chen. Recent global CO˙2 flux inferred from atmospheric CO˙2 observations and its regional analyses. *Biogeosciences Discussions*, 8:3497–3536, 2011.

[19] P. J. Diggle, P. J. Jr. Ribeiro, and O. F. Christensen. An introduction to model-based geostatistics, 2003.

[20] S. Doney, B. Tilbrook, S. Roy, N. Metzl, C. Le Quéré, and M. Hood. Surface-ocean CO2 variability and vunerability. *Deep Sea Research II*, 56:504–511, 2009.

[21] S. C. Doney, V. J. Fabry, R. A. Feely, and J. A. Kleypas. Ocean acidification: The other CO˙2 problem. *Annual Review of Marine Science*, 1:169–192, 2009.

[22] D. P. Duckett. The application of statistical estimation techniques to terrain modeling. Master's thesis, Airforce Institute of Technology Air University, 1991.

[23] X. Emery. The kriging update equations and their application to the selection of neighbouring data. *Computers and Geoscience*, 13:269–280, 2009.

[24] M. Gen and R. Cheng. *Genetic Algorithms and Engineering Optimization*. John Wiley and Sons Inc, 2000.

[25] S. T. Gille. Decadal-scale temperature trends in the southern hemisphere ocean. *Journal of Climate*, 21:4749–4765, 2008.

[26] D. Ginsbourger, D. Dupuy, A. Badea, L. Carraro, and O. Roustant. A note on the choice and the estimation of kriging models for the analysis of deterministic computer experiments. *Applied Stochastic Models in Business and Industry*, 25:115–131, 2009.

[27] L. J. Haarhoff, S. Kok, and D. N. Wilke. Numerical strategies to reduce the effect of ill-conditioned correlation matrices and underflow errors in kriging. *Journal of Mechanical Design*, 2013.

[28] R. L. Haupt and S. E. Haupt. *Practical Genetic Algorithms*. John Wiley and Sons Inc, second edition, 2004.

[29] R. D. Hedger, P. M. Atkinson, and T. J. Malthus. Optimizing sampling strategies for estimating mean water quality in lakes using geostatistical techniques with remote sensing. *Lakes and Reservoirs: Research and Management*, 6:279–288, 2001.

[30] T. Hengl. *A Practical Guide to Geostatistical Mapping of Environmental Variables*. European Commission,Joint Research Centre,Institute for Environment and Sustainability, 2007.

[31] T. Hengl, G. B. M. Heuvelink, and D. G. Rossiter. About regression-kriging: From equations to case studies. *Computers and Geosciences*, 33:1301–1315, 2007.

[32] T. Hengl, G. B. M. Heuvelink, and A. Stein. Comparison of kriging with external drift and regression-kriging. `http://www.itc.nl/library/Papers_2003/misca/hengl_comparison.pdf`, 2003.

[33] C. Ho and S. Saripalli. Where do you sample? - an autonomous underwater vehicle story, 2011.

[34] E. H. Isaaks and R. Mohan Srivastava. *An Introduction to Applied Geostatistics*. Oxford University Press, 1989.

[35] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–.

[36] A. G. Journel and M. E. Rossi. When do we need a trend model in kriging? *Mathematical Geology*, 21:715–739, 1989.

[37] K. Juang, D. Lee, and Z. Chen. Geostatistical cross-validation for the design of additional sampling regimes in heavy-metal contaminated soils. *Journal of the Chinese Institute of Environmental Engineering*, 9:89–96, 1999.

[38] K. Juang, D. Lee, and Y. Teng. Adaptive sampling based on the cumulative distribution function of order statistics to delineate heavy-metal contaminated soils using kriging. *Environmental Pollution*, 138:268–277, 2005.

[39] S. Kaiser and D. K. A. Barnes. Southern ocean deep-sea biodiversity: Sampling strategies and predicting responses to climate change. *Climate Research*, 37:165–179, 2008.

[40] R. M. Lark. Estimating variograms of soil properties by the method-of-moments and maximum likelihood. *European Journal of Soil Science*, 51:717–728, 2000.

[41] J. Laurenceau and P. Sagaut. Building efficient response surfaces of aerodynamic functions with kriging and co-kriging, 2008.

[42] A. Lenton, B. Tilbrook, R. M. Law, D. Bakker, S.C. Doney, N. Gruber, M. Ishii, M. Hoppema, N. S. Lovenduski, R. J. Matear, B. I. McNeil, N. Metzl, S. E. Mikaloff Fletcher, P. M. S. Monteiro, C. Rödenbeck, C. Sweeney, and T. Takahashi. Sea-air fluxes in the southern ocean for the period 1990–2009. *Biogeosciences*, 10:4037–4054, 2013.

[43] J. Liu, Z. Han, and W. Song. Comparison of infill sampling criteria in kriging-based aerodynamic optimization. In *28th International Congress of the Aeronautical Sciences*, 2012.

[44] Z. Liu, J. Zhan, and C. Tan. Improved relability approximate method combining kriging and importance sampling. In *2012 Prognostics and System Health Management Conference*, 2012.

[45] E. Maier-reimer, U. Mikolajewicz, and A. Winguth. Future ocean uptake of CO˙2:interaction between ocean circulation and biology. *Climate Dynamics*, 12:711–721, 1996.

[46] O. Makhnin. Lecture 10: Introduction to kriging. `http://infohost.nmt.edu/~olegm/586/HYD10.pdf`, 2012. last accessed: 09 January 2012.

[47] O. Makhnin. Lecture 12: Universal kriging, cross validation, co-kriging. `http://infohost.nmt.edu/~olegm/586/HYD12.pdf`, 2012. last accessed: 09 January 2012.

[48] O. Makhnin. Lecture 13: Indicator kriging. `http://infohost.nmt.edu/~olegm/586/HYD13.pdf`, 2012. last accessed: 09 January 2012.

[49] K. V. Mardia. Should geostatics be model-based? In *12th Conference of International Association for Mathematical Geology*, 2007.

[50] J. D. Martin and T. W. Simpson. On the use of kriging models to approximate deterministic computer models. In *ASME 2004 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 2004.

[51] B. I. McNeil. Oceanic implications for climate change policy. *Environmental Science and Policy*, 9:595–606, 2006.

[52] M. P. Meredith, O. Schofield, L. Newman, E. Urban, and M. Sparrow. The vision for a southern ocean observing system. *Current Opinion in Environmental Sustainability*, 5:1–8, 2013.

[53] N. Metzl, C. Brunet, A. Jabaud-Jan, A. Poisson, and B. Schauer. Summer and winter air-sea CO2 fluxes in the southern ocean. *Deep Sea Research II*, 53:1548–1563, 2006.

[54] A. Morton. Distribution mapping software. `http://www.dmap.co.uk/utmworld.gif`, 2013.

[55] E. Pardo-Igúzquiza and P. A. Dowd. The second-order stationary universal kriging model revisited. *Mathematical Geology*, 30:347–378, 1998.

[56] E. Pardo-Igúzquiza, P. A. Dowd, and D. I. F. Grimes. An automatic moving window approach for mapping meteorological data. *Internation Journal of Climatology*, 25:665–678, 2005.

[57] A. Paul-Dubois-Taine and S. Nadarajah. A sequential sampling approach to build efficient kriging response surfaces for aerodynamics data, 2012.

[58] G. T. Pereira, Z. M. de Souza, D. De Bortoli Teixeira, R. Montanari, and J. Marques Jnr. Optimization of the sampling scheme for maps of physical and chemical properties estimated by kriging. *R. Bras. Ci. Solo*, 37:1128–1135, 2013.

[59] I. C. Prentice, G. D. Farquhar, M. J. R. Fasham, M. L. Goulden, M. Heimann, V. J. Jaramillo, H. S. Kheshgi, C. Le Quéré, R. J. Scholes, and D. W. R. Wallace. The carbon cycle and atmospheric carbon dioxide, 2001.

[60] N. Pringle. Personal Communication, October 2013.

[61] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. ISBN 3-900051-07-0.

[62] S. Rahmstorf. Thermohaline ocean circulation. *Encyclopedia of Quaternary Sciences*, 2006.

[63] C. R. Reeves and J. E. Rowe. *Genetic Algorithms - Principles and Perspectives: A Guide to GA Theory*. Kluwer Academic Publishers, 2003.

[64] S. R. Rintoul. Monitoring and understanding southern ocean variability and its impact on climate: A strategy for sustained observations, 2001.

[65] S. R. Rintoul, K. Speer, M. Sparrow, M. Meredith, E. Hofmann, E. Fahrbach, T. de Bruin, A. Naveira Garabato, K. Alverson, V. Ryabinin, H. C. Shin, and S. Gladyshev. Southern ocean observing system(SOOS):rationale and strategy for sustained observations of the southern ocean, 2009.

[66] B. Rosenbaum and V. Schulz. Comparing sampling strategies for aerodynamic kriging surrogate models. *Z. Angew. Math. Mech.*, 92:852–868, 2012.

[67] M. Sherman. *Spatial Statistics and Spatio-Temporal Data - Coavariance Functions and Directional Properties*. John Wiley and Sons Inc, 2011.

[68] S. N. Sivanandam and S. N. Deepa. *Introduction to Genetic Algorithms*. Springer, 2008.

[69] L. Spadavecchia. Spatio-temporal geostatistical methods: Edinburgh space-time geostatistics users guide, 2008.

[70] T. Takahashi, S. Sutherland, R. Wanninkhof, C. Sweeney, R. Feely, and D. Chipman. Climatological mean and decadal change in surface ocean pCO2 , and net sea-air CO2 flux over the global oceans. *Deep Sea Research II*, 56:554–577, 2009.

[71] E. Todini. Influence of parameter estimation uncertainty in kriging: Part 1-theorectical development. *Hydrology and Earth System Sciences*, 5:215–223, 2001.

[72] E. A. Varouchakis and D. T. Hristopulos. Improvement of groundwater leevel prediction in sparsely gauged basins using physical laws and local geographic features as auxiliary variables. *Advances in Water Resources*, 52:34–49, 2013.

[73] R. Webster and M. A. Oliver. *Geostatistics for Environmental Scientists*. John Wiley and Sons Inc, second edition, 2009.

[74] D. Whitley. A genetic algorithm tutorial, 1994.

[75] X. Zhu, J. Yu, and X. Wang. Optimization of large scales ocean sampling for minimization of the kriging variance. In *Proceedings of the 8th World Congress on Intelligent Control and Automation*, 2010.

APPENDIX A

PYTHON'S STATS PACKAGE

This package is found within Scipy and contains many probability distributions and statistical functions [11]. A short and incomplete list of these distributions and functions is [11]:

- expon - A continuous exponential variable
- norm - A continuous normal (Gaussian) variable
- binom - A discrete binomial variable
- skew() - Returns skewness of data set
- kurtosis() - Returns kurtosis of data set
- f˙oneway() - Performs a oneway ANOVA

For more information and the remaining distributions and functions see [11].

APPENDIX B

ORDINARY LEAST SQUARES TRENDS

In Table B.1, the Ordinary Least Squares fitted trend coefficients are given for the constant, linear and quadratic trends presented in 6.1.

Trend Structure		β
Constant	β_0	2.9782
Linear	β_0	1.2221
	β_1	-0.0076
	β_2	0.0626
	β_3	0.0035
Quadratic	β_0	9.0659
	β_1	-2.7548×10^{-1}
	β_2	-2.1838×10^{-1}
	β_3	2.1173×10^{-3}
	β_4	1.7457×10^{-3}
	β_5	1.4499×10^{-3}
	β_6	-6.4410×10^{-5}
	β_7	4.9714×10^{-3}
	β_8	6.5585×10^{-5}
	β_9	3.6007×10^{-4}

Table B.1: Ordinary Least Squares Trend Coefficients for Various Trend Structures.

APPENDIX C

TREND AND COVARIANCE PARAMETER ESTIMATION

Any additional figures and values from the determination of the trend and covariance parameters will be presented in this Appendix.

C.1 Maximum Likelihood Estimation Results - Exponential Covariance Structure

Tables C.1 to C.3 give the values of the quadratic trend and exponential covariance parameters as fitted to the 20 randomly generated sets, as discussed in Section 7.1. The histograms of these parameters can be seen in Figures C.1 to C.2.

Random Set	σ^2	a_{time}	a_{lat}	a_{lon}
1	5.0573	22.7835	6.1693	15.0702
2	4.7250	21.4301	6.3511	15.2178
3	5.0345	24.3857	6.2375	14.9917
4	4.8854	23.1719	5.9747	14.8331
5	4.7097	22.5230	6.0012	14.5867
6	4.9937	22.7475	6.4561	13.9191
7	5.3576	24.9536	6.1280	15.5901
8	4.7088	22.2955	5.9718	14.8174
9	4.6233	22.4939	6.2066	14.5908
10	4.6292	23.0525	6.2542	15.0794
11	5.1644	22.8773	6.3991	16.1252
12	4.8642	24.5844	6.2979	14.2846
13	4.6353	22.7629	6.1049	14.2013
14	4.7695	25.0176	6.0136	14.9733
15	5.0303	24.2962	6.4479	15.3430
16	5.0141	24.2188	6.1094	15.0203
17	5.2401	26.3618	5.9868	16.1845
18	5.0765	24.0453	6.3878	13.6712
19	5.0103	24.4827	6.4369	14.9731
20	5.0172	23.1294	6.2146	14.2381

Table C.1: σ^2, a_{time}, a_{lat} and a_{lon} across the 20 random sets for the exponential covariance structure.

Random Set	β_0	β_1	β_2	β_3	β_4
1	8.6935	−0.1895	−0.1189	−0.0284	0.0016
2	9.0787	−0.1912	−0.1462	−0.0321	0.0016
3	8.6656	−0.1909	−0.1299	−0.0278	0.0016
4	9.4797	−0.1942	−0.1422	−0.0377	0.0016
5	8.4042	−0.1893	−0.1407	−0.0190	0.0017
6	9.3431	−0.2058	−0.1651	−0.0301	0.0018
7	8.6701	−0.1757	−0.1349	−0.0307	0.0015
8	8.2232	−0.1981	−0.1002	−0.0247	0.0016
9	10.2740	−0.2246	−0.1932	−0.0312	0.0017
10	8.3086	−0.2063	−0.1191	−0.0182	0.0017
11	9.1265	−0.1901	−0.1366	−0.0335	0.0016
12	9.0890	−0.1972	−0.1517	−0.0277	0.0017
13	9.2247	−0.1964	−0.1579	−0.0294	0.0016
14	9.4169	−0.1986	−0.1463	−0.0387	0.0016
15	9.2661	−0.2169	−0.1395	−0.0300	0.0018
16	8.6081	−0.1932	−0.1317	−0.0274	0.0016
17	8.7676	−0.1973	−0.1486	−0.01668	0.0016
18	9.3616	−0.2046	−0.1485	−0.02942	0.0016
19	8.4690	−0.1868	−0.1291	−0.02222	0.0016
20	8.0629	−0.1834	−0.1301	−0.01759	0.0016

Table C.2: β_0, β_1, β_2, β_3 and β_4 across the 20 random sets for the exponential covariance structure.

Random Set	β_5	β_6	β_7	β_8	β_9
1	-3.2178×10^{-4}	5.6456×10^{-5}	0.0023	3.9701×10^{-5}	0.0006
2	2.3698×10^{-4}	8.1300×10^{-5}	0.0023	4.3924×10^{-5}	0.0006
3	4.0169×10^{-5}	8.0612×10^{-5}	0.0022	5.6514×10^{-5}	0.0005
4	-7.1953×10^{-5}	9.2234×10^{-5}	0.0023	5.3165×10^{-5}	0.0007
5	3.0287×10^{-4}	4.9960×10^{-5}	0.0024	-4.6870×10^{-5}	0.0005
6	5.6232×10^{-4}	9.2522×10^{-5}	0.0026	1.2528×10^{-5}	0.0005
7	-1.1252×10^{-5}	8.8858×10^{-5}	0.0023	-3.7337×10^{-5}	0.0006
8	-3.0318×10^{-4}	7.5290×10^{-5}	0.0025	8.0508×10^{-5}	0.0004
9	6.3191×10^{-4}	6.8071×10^{-5}	0.0030	5.3935×10^{-5}	0.0007
10	3.5894×10^{-5}	5.8092×10^{-5}	0.0025	4.7480×10^{-5}	0.0004
11	-3.2187×10^{-5}	8.2777×10^{-5}	0.0023	3.3315×10^{-5}	0.0006
12	2.2741×10^{-4}	7.7569×10^{-5}	0.0025	-9.0743×10^{-6}	0.0006
13	4.3675×10^{-4}	1.1935×10^{-4}	0.0026	-6.5640×10^{-5}	0.0005
14	2.3701×10^{-4}	1.0706×10^{-4}	0.0022	1.4438×10^{-4}	0.0006
15	1.4786×10^{-4}	1.0400×10^{-4}	0.0027	7.6378×10^{-5}	0.0004
16	1.2081×10^{-4}	8.7931×10^{-5}	0.0025	2.8806×10^{-6}	0.0005
17	2.4956×10^{-4}	7.2970×10^{-5}	0.0027	-7.1636×10^{-5}	0.0003
18	5.5715×10^{-5}	8.9070×10^{-5}	0.0028	8.0947×10^{-6}	0.0005
19	9.7067×10^{-6}	5.4581×10^{-5}	0.0024	2.6492×10^{-5}	0.0005
20	3.8409×10^{-4}	6.0311×10^{-5}	0.0023	-1.0986×10^{-5}	0.0003

Table C.3: β_5, β_6, β_7, β_8 and β_9 across the 20 random sets for the exponential covariance structure.

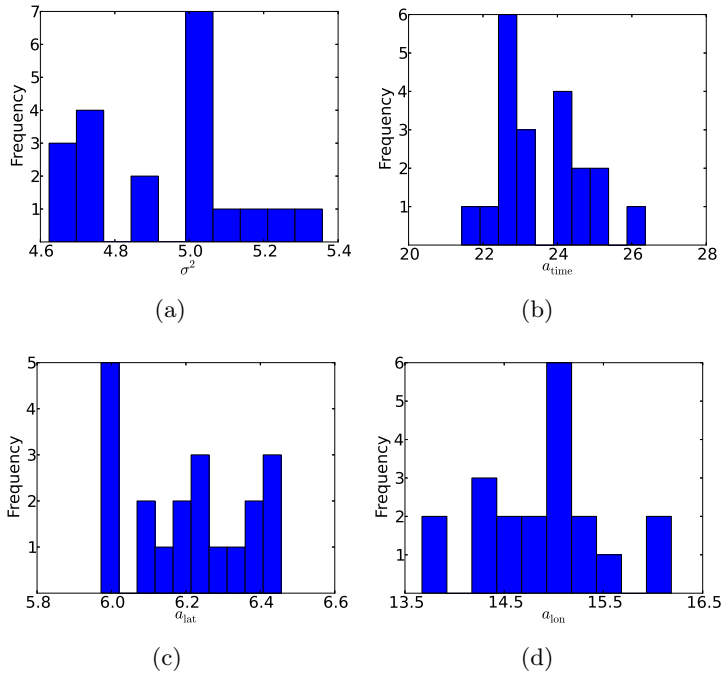


Figure C.1: Histogram of (a) σ^2, (b) a_{time}, (c) a_{lat} and (d) a_{lon} values from the 20 random sets for the exponential covariance structure.

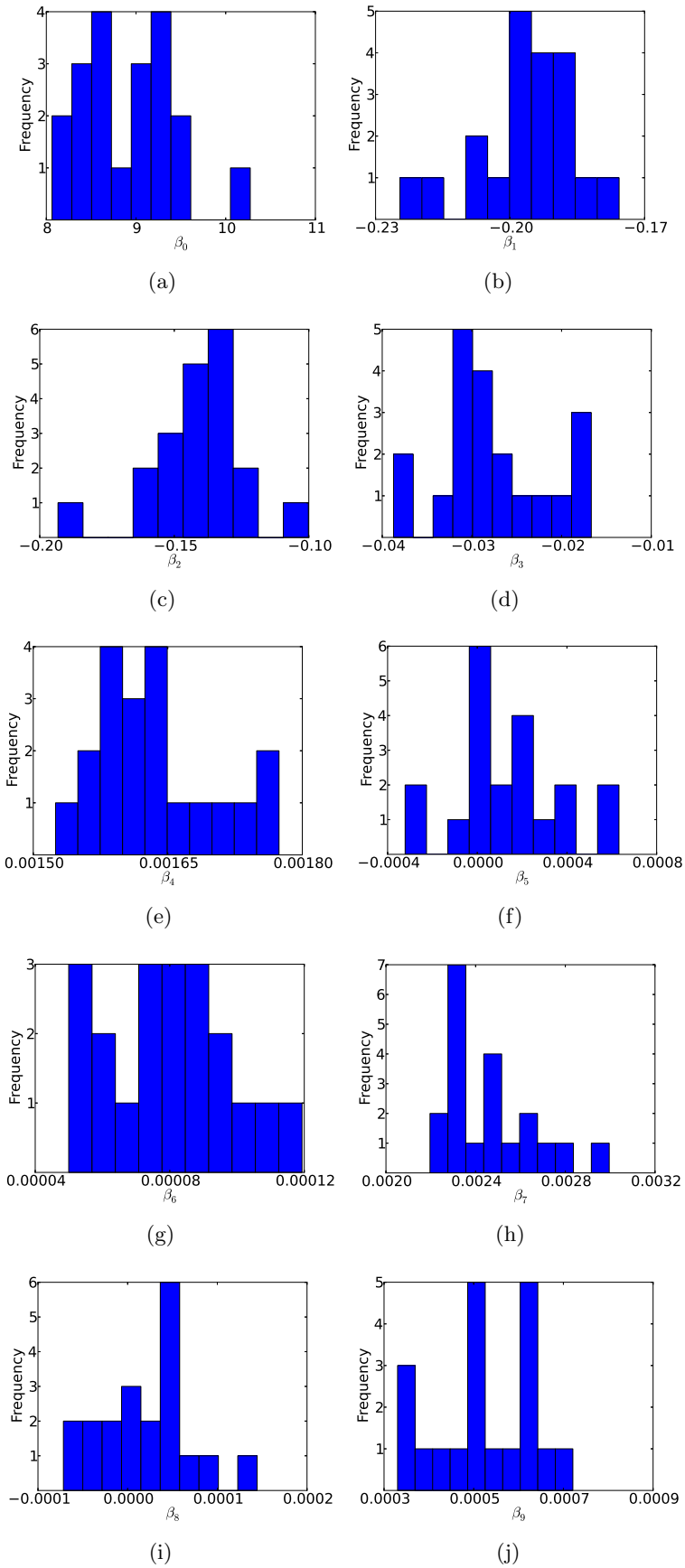


Figure C.2: Histogram of (a) β_0, (b) β_1, (c) β_2, (d) β_3, (e) β_4, (f) β_5, (g) β_6, (h) β_7, (i) β_8 and (j) β_9 values from the 20 random sets for the exponential covariance structure.

C.2 Maximum Likelihood Estimation Results - Spherical Covariance Structure

Tables C.4 to C.6 give the values of the quadratic trend and exponential covariance parameters as fitted to the 20 randomly generated sets, as discussed in Section 7.1. The histograms of these parameters can be seen in Figures C.3 to C.4.

Random Set	σ^2	a_{time}	a_{lat}	a_{lon}
1	5.3605	50.2239	13.5137	33.5363
2	4.9507	46.8068	13.5578	34.0753
3	5.1489	51.5012	13.4025	32.9962
4	5.1807	50.6921	13.1981	33.1790
5	4.8423	48.7064	12.8241	32.2104
6	5.2940	50.2454	13.9823	31.3071
7	5.6875	53.7203	13.2934	35.0900
8	4.9454	49.0319	13.2168	32.5935
9	4.7653	48.6569	13.2669	32.1951
10	4.7766	50.4847	13.6215	32.3945
11	5.3898	50.1465	13.4757	35.7351
12	5.0728	52.2925	13.8196	31.9248
13	4.7778	48.7663	13.1575	31.4395
14	4.8526	52.0244	13.1592	32.7050
15	5.2070	52.0300	13.8161	34.1887
16	5.1957	52.1298	12.8605	33.7832
17	5.4008	55.4858	12.7633	35.9469
18	5.3714	52.1224	13.7318	30.9927
19	5.1523	51.5264	14.3448	32.2174
20	5.3925	51.4034	13.5510	32.2885

Table C.4: σ^2, a_{time}, a_{lat} and a_{lon} across the 20 random sets for the spherical covariance structure.

Random Set	β_0	β_1	β_2	β_3	β_4
1	7.4508	−0.1691	−0.0842	−0.0220	0.0015
2	8.0658	−0.1751	−0.1183	−0.0259	0.0015
3	7.6596	−0.1719	−0.1039	−0.0236	0.0015
4	8.4769	−0.1730	−0.1171	−0.0336	0.0014
5	7.2334	−0.1717	−0.1079	−0.0133	0.0016
6	8.2105	−0.1870	−0.1349	−0.0238	0.0016
7	7.4027	−0.1512	−0.1005	−0.0246	0.0014
8	7.1067	−0.1790	−0.0698	−0.0192	0.0015
9	9.2158	−0.2041	−0.1660	−0.0264	0.0016
10	7.3445	−0.1869	−0.0941	−0.0147	0.0016
11	7.9131	−0.1718	−0.1068	−0.0262	0.0015
12	7.9370	−0.1789	−0.1173	−0.0226	0.0015
13	8.1781	−0.1766	−0.1344	−0.0235	0.0015
14	8.2472	−0.1777	−0.1193	−0.0313	0.0015
15	8.1469	−0.1986	−0.1108	−0.0231	0.0016
16	7.3146	−0.1721	−0.0913	−0.0209	0.0015
17	7.5546	−0.1759	−0.1103	−0.0115	0.0015
18	8.2598	−0.1827	−0.1172	−0.0258	0.0015
19	7.3650	−0.1667	−0.1026	−0.0173	0.0014
20	6.9976	−0.1626	−0.0986	−0.0136	0.0014

Table C.5: β_0, β_1, β_2, β_3 and β_4 across the 20 random sets for the spherical covariance structure.

Random Set	β_5	β_6	β_7	β_8	β_9
1	-4.4154×10^{-4}	3.0785×10^{-5}	0.0020	2.2034×10^{-5}	0.0006
2	1.0821×10^{-4}	5.1543×10^{-5}	0.0021	3.9252×10^{-5}	0.0006
3	-1.2891×10^{-5}	5.7303×10^{-5}	0.0019	4.7674×10^{-5}	0.0005
4	-1.2319×10^{-4}	7.0993×10^{-5}	0.0020	4.7282×10^{-5}	0.0007
5	2.1009×10^{-4}	2.9548×10^{-5}	0.0021	-4.9937×10^{-5}	0.0005
6	4.4629×10^{-4}	6.7326×10^{-5}	0.0023	2.5324×10^{-6}	0.0005
7	-1.2799×10^{-4}	6.4796×10^{-5}	0.0019	-6.1851×10^{-5}	0.0006
8	-3.9689×10^{-4}	5.1374×10^{-5}	0.0022	6.6396×10^{-5}	0.0004
9	5.5258×10^{-4}	4.9387×10^{-5}	0.0027	2.7709×10^{-5}	0.0007
10	4.5999×10^{-6}	3.6275×10^{-5}	0.0022	5.2902×10^{-5}	0.0004
11	-8.1359×10^{-5}	5.4497×10^{-5}	0.0020	1.5312×10^{-5}	0.0006
12	6.6551×10^{-5}	5.3248×10^{-5}	0.0021	-1.4301×10^{-5}	0.0006
13	4.1127×10^{-4}	8.4283×10^{-5}	0.0023	-7.5208×10^{-5}	0.0005
14	2.0002×10^{-4}	7.1026×10^{-5}	0.0020	1.2499×10^{-4}	0.0006
15	5.3971×10^{-5}	6.9113×10^{-5}	0.0024	6.8436×10^{-5}	0.0004
16	-1.0579×10^{-4}	5.2883×10^{-5}	0.0021	1.1019×10^{-5}	0.0005
17	8.4962×10^{-6}	4.5845×10^{-5}	0.0024	-7.4101×10^{-5}	0.0003
18	-9.3147×10^{-5}	7.0132×10^{-5}	0.0025	2.3467×10^{-7}	0.0005
19	-5.0068×10^{-6}	2.9259×10^{-5}	0.0020	1.3512×10^{-5}	0.0005
20	2.6364×10^{-4}	3.9226×10^{-5}	0.0020	-6.4328×10^{-6}	0.0003

Table C.6: β_5, β_6, β_7, β_8 and β_9 across the 20 random sets for the spherical covariance structure.

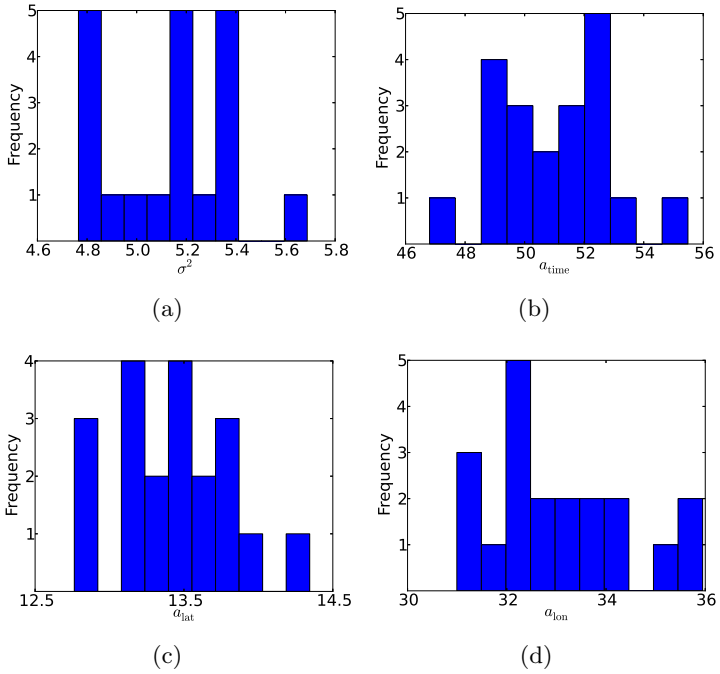


Figure C.3: Histogram of (a) σ^2, (b) a_{time}, (c) a_{lat} and (d) a_{lon} values from the 20 random sets for the spherical covariance structure.

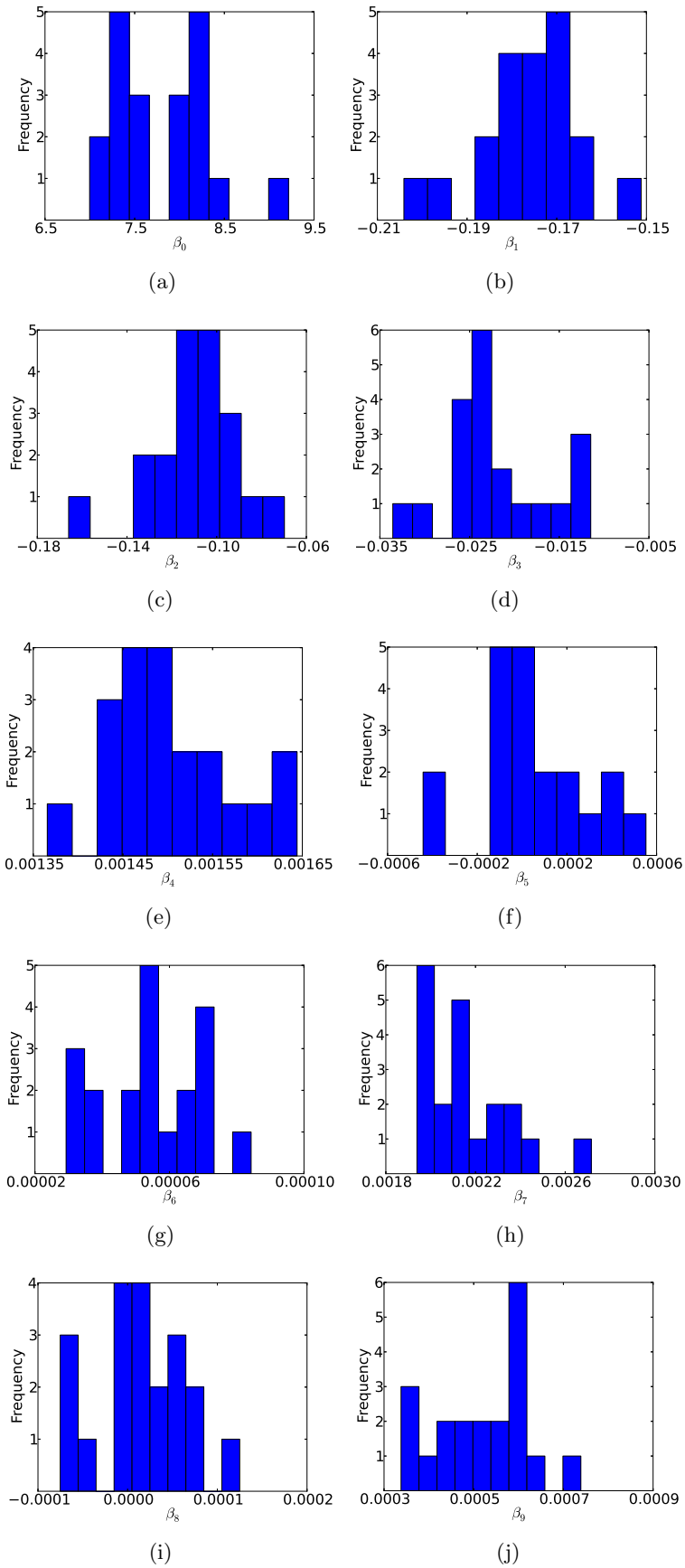


Figure C.4: Histogram of (a) β_0, (b) β_1, (c) β_2, (d) β_3, (e) β_4, (f) β_5, (g) β_6, (h) β_7, (i) β_8 and (j) β_9 values from the 20 random sets for the spherical covariance structure.

C.3 Maximum Likelihood Estimation Results - Approximate Gaussian Covariance Structure

Tables C.7 to C.9 give the values of the quadratic trend and exponential covariance parameters as fitted to the 20 randomly generated sets, as discussed in Section 7.1. The histograms of these parameters can be seen in Figures C.5 to C.6.

Random Set	σ^2	a_{time}	a_{lat}	a_{lon}
1	4.7796	9.2314	8.8221	22.6119
2	4.5548	12.1954	8.5727	18.7210
3	4.6101	11.0939	8.9878	18.3619
4	4.6874	15.5184	7.3557	16.6821
5	4.4889	8.5797	8.6546	26.1962
6	4.6780	8.3240	9.0031	25.7619
7	4.8524	9.4512	9.7783	20.8382
8	4.5543	14.4621	7.6881	19.4464
9	4.4619	11.5512	7.8470	21.6766
10	4.4459	12.1568	8.5923	20.9701
11	4.6518	8.6465	9.7745	24.0278
12	4.5603	9.8959	9.1478	22.5086
13	4.4567	10.2680	6.8393	24.0173
14	4.5229	16.9987	7.6082	16.9866
15	4.6737	13.4340	7.6743	20.2086
16	4.7404	10.8897	7.1012	24.9786
17	4.6165	12.2321	8.5475	18.4114
18	4.6428	14.5745	7.5564	18.2987
19	4.5807	11.6347	9.0561	18.3679
20	4.6031	11.7356	8.5834	16.1706

Table C.7: σ^2, a_{time}, a_{lat} and a_{lon} across the 20 random sets for the approximate Gaussian covariance structure.

Random Set	β_0	β_1	β_2	β_3	β_4
1	9.7171	−0.2722	−0.2113	−0.0076	0.0017
2	9.3739	−0.2606	−0.2111	−0.0069	0.0017
3	9.6350	−0.2627	−0.2160	−0.0094	0.0017
4	9.3554	−0.2631	−0.1968	−0.0078	0.0016
5	9.3745	−0.2722	−0.2229	−0.0009	0.0018
6	8.9531	−0.2695	−0.1907	−0.0012	0.0018
7	9.4328	−0.2665	−0.2108	−0.0063	0.0017
8	8.9550	−0.2614	−0.1882	−0.0045	0.0016
9	9.0957	−0.2746	−0.1969	0.0003	0.0017
10	8.9580	−0.2610	−0.1958	−0.0019	0.0017
11	9.1857	−0.2665	−0.2024	−0.0049	0.0017
12	9.3080	−0.2696	−0.2111	−0.0037	0.0018
13	9.4095	−0.2684	−0.2241	−0.0014	0.0017
14	9.5485	−0.2562	−0.2161	−0.0110	0.0016
15	8.6102	−0.2651	−0.1904	0.0044	0.0017
16	9.1963	−0.2695	−0.1995	−0.0033	0.0018
17	9.4841	−0.2698	−0.2103	−0.0044	0.0017
18	9.3592	−0.2625	−0.2012	−0.0067	0.0016
19	9.4626	−0.2691	−0.2044	−0.0057	0.0017
20	8.7601	−0.2583	−0.2089	0.0025	0.0017

Table C.8: β_0, β_1, β_2, β_3 and β_4 across the 20 random sets for the approximate Gaussian covariance structure.

Random Set	β_5	β_6	β_7	β_8	β_9
1	0.0009	-3.7430×10^{-5}	0.0048	4.2380×10^{-5}	0.0005
2	0.0012	-3.4543×10^{-5}	0.0045	4.8651×10^{-5}	0.0005
3	0.0011	-2.4213×10^{-5}	0.0047	5.9154×10^{-5}	0.0005
4	0.0009	-2.1008×10^{-5}	0.0047	8.2915×10^{-5}	0.0004
5	0.0014	-3.6299×10^{-5}	0.0049	6.8285×10^{-6}	0.0004
6	0.0010	-3.7360×10^{-5}	0.0047	3.3042×10^{-5}	0.0003
7	0.0012	-2.2367×10^{-5}	0.0048	4.0224×10^{-5}	0.0004
8	0.0009	-2.4403×10^{-5}	0.0047	5.4336×10^{-5}	0.0004
9	0.0009	-4.6820×10^{-5}	0.0049	4.9074×10^{-5}	0.0004
10	0.0010	-4.1390×10^{-5}	0.0046	3.5091×10^{-5}	0.0004
11	0.0011	-2.0143×10^{-5}	0.0047	2.4988×10^{-5}	0.0004
12	0.0013	-3.1080×10^{-5}	0.0047	2.6555×10^{-5}	0.0004
13	0.0014	-4.6627×10^{-5}	0.0048	1.7058×10^{-5}	0.0004
14	0.0014	-1.3837×10^{-5}	0.0045	1.0148×10^{-4}	0.0004
15	0.0011	-4.9898×10^{-5}	0.0048	2.3672×10^{-5}	0.0003
16	0.0010	-3.8155×10^{-5}	0.0047	6.2773×10^{-5}	0.0004
17	0.0011	-2.4394×10^{-5}	0.0048	4.1435×10^{-5}	0.0004
18	0.0009	-2.8415×10^{-5}	0.0047	4.7165×10^{-5}	0.0004
19	0.0009	-3.0733×10^{-5}	0.0048	6.4641×10^{-5}	0.0004
20	0.0014	-3.7618×10^{-5}	0.0047	1.0046×10^{-5}	0.0003

Table C.9: β_5, β_6, β_7, β_8 and β_9 across the 20 random sets for the approximate Gaussian covariance structure.

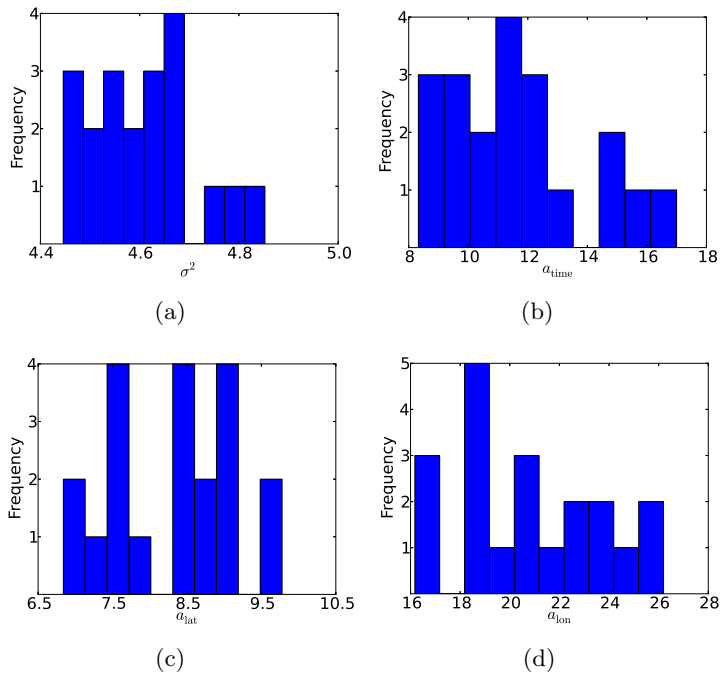


Figure C.5: Histogram of (a) σ^2, (b) a_{time}, (c) a_{lat} and (d) a_{lon} values from the 20 random sets for the approximate Gaussian covariance structure.

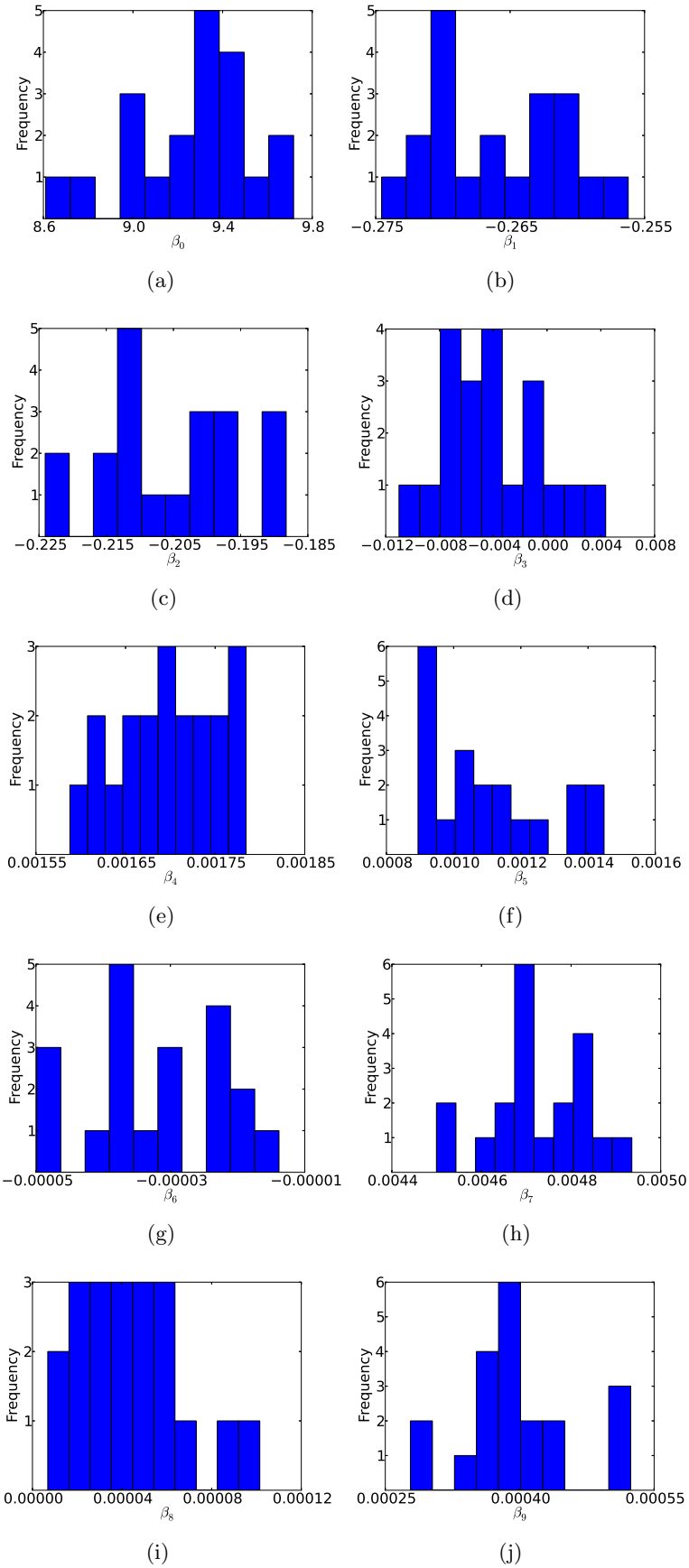


Figure C.6: Histogram of (a) β_0, (b) β_1, (c) β_2, (d) β_3, (e) β_4, (f) β_5, (g) β_6, (h) β_7, (i) β_8 and (j) β_9 values from the 20 random sets for the approximate Gaussian covariance structure

APPENDIX D

TWO-DIMENSIONAL SAMPLING

Any additional figures from the two-dimensional sampling methods will be presented in this appendix.

D.1 Random Sampling

Maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux, the estimated error variance and the true error across the chosen two-dimensional region for sets one, two, three, five, six, seven, nine and ten can be seen in Figures D.1 to D.8. These maps are given for the full sample set of size 70 generated using Random Sampling (See Chapter 9).

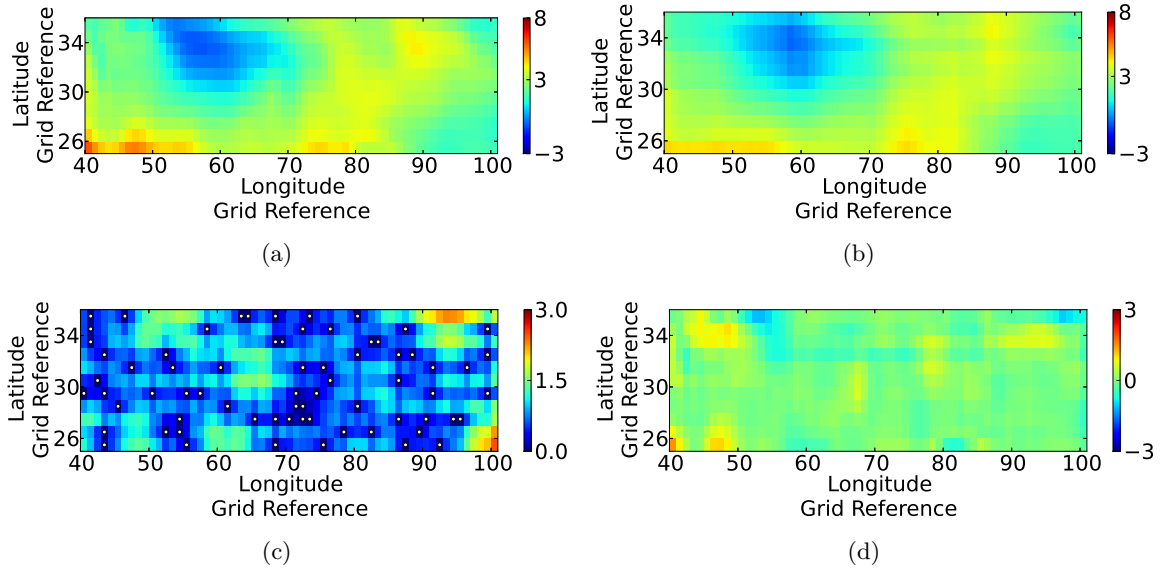


Figure D.1: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set one generated using Random Sampling. The white dots in (c) indicate the sampling points.

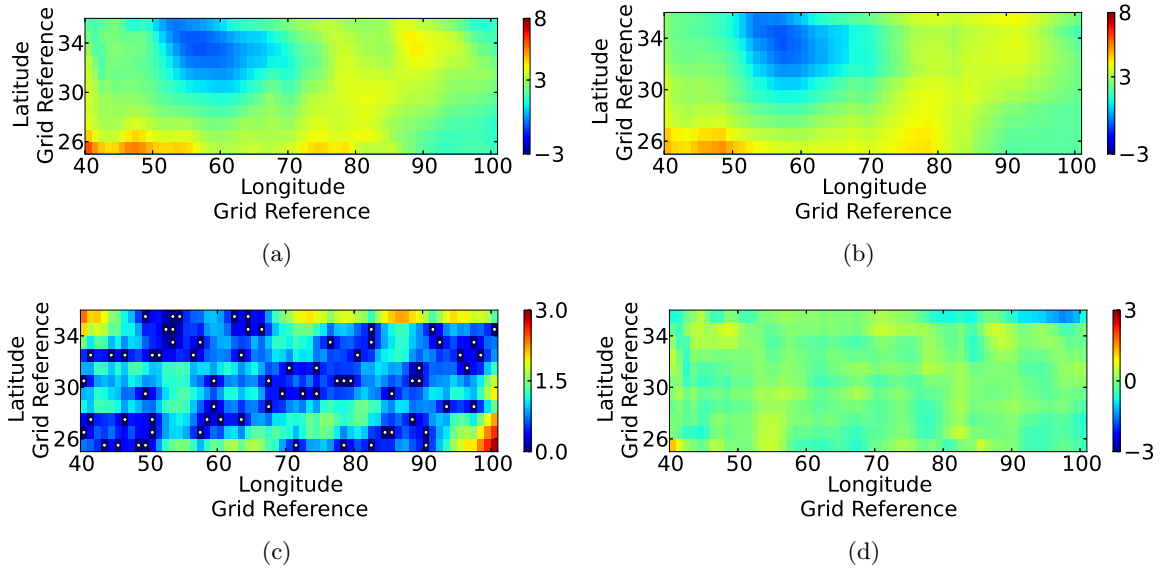


Figure D.2: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set two generated using Random Sampling. The white dots in (c) indicate the sampling points.

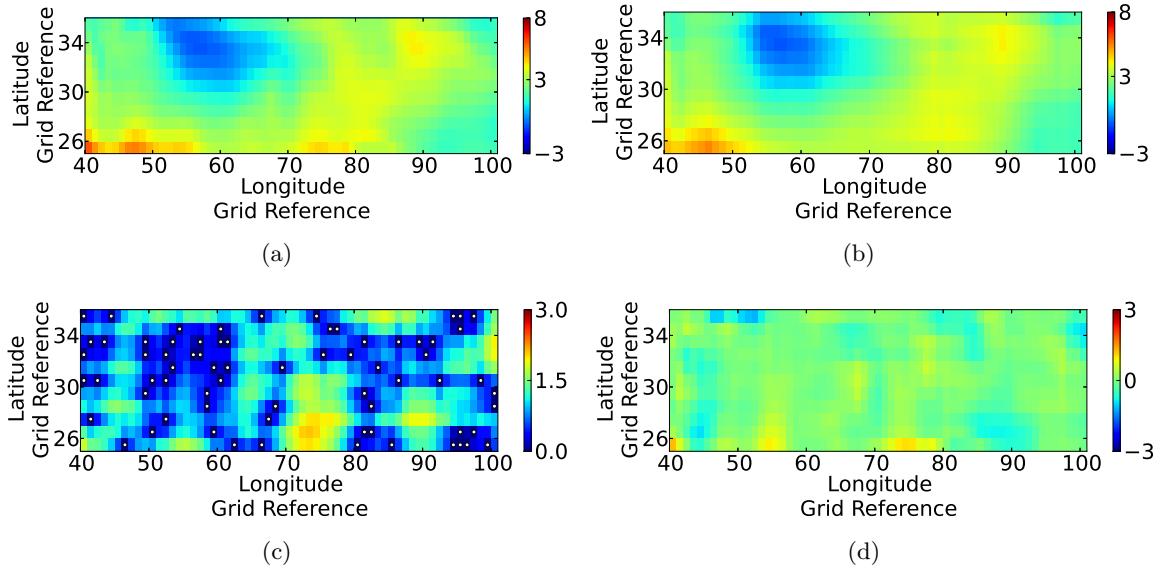


Figure D.3: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set three generated using Random Sampling. The white dots in (c) indicate the sampling points.

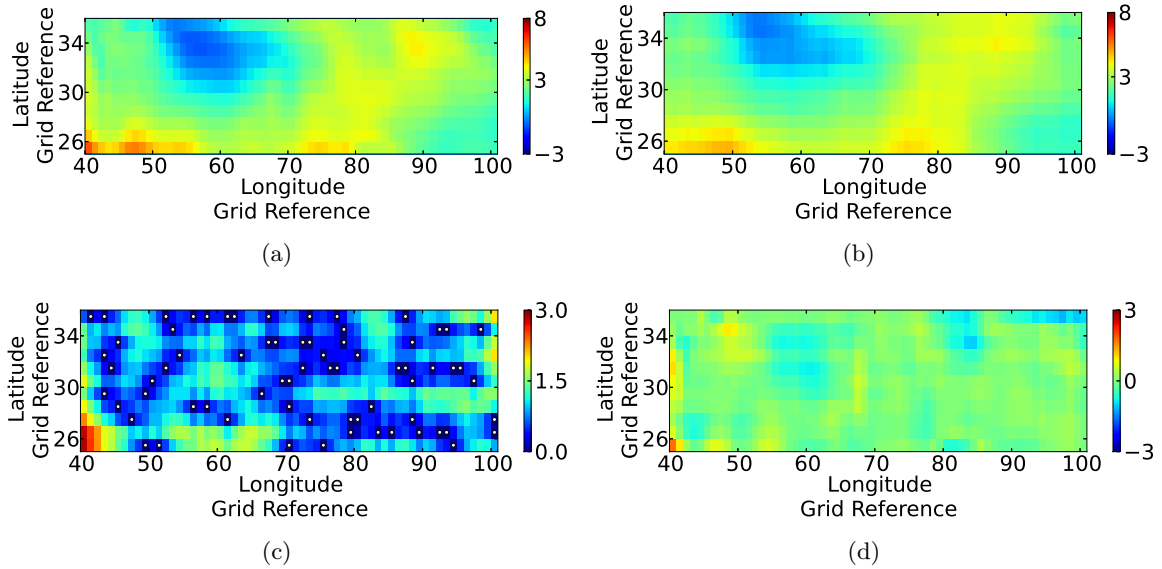


Figure D.4: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set five generated using Random Sampling. The white dots in (c) indicate the sampling points.

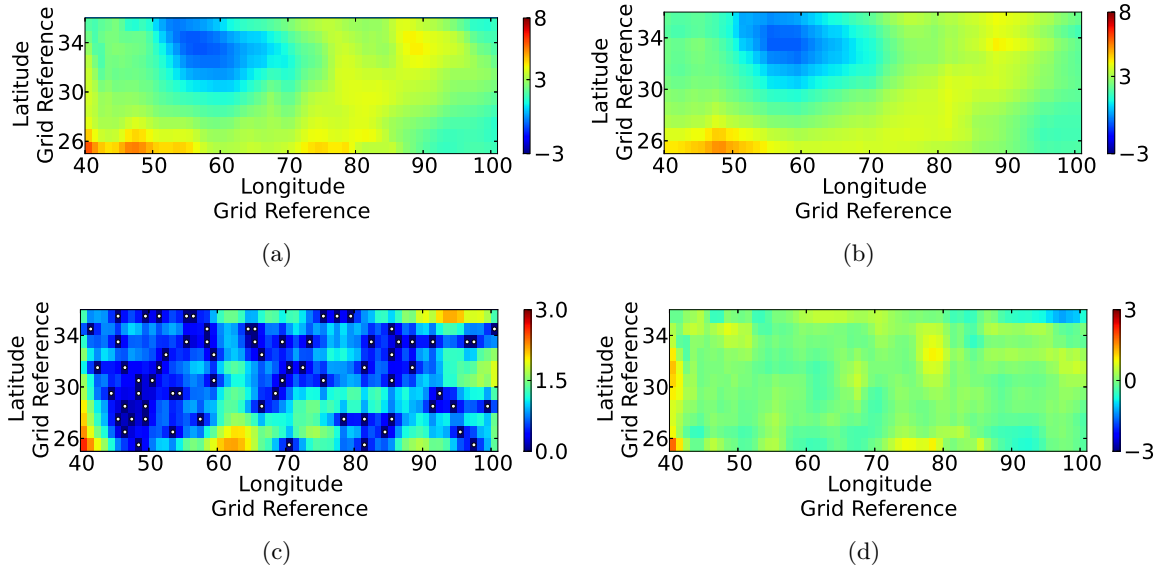


Figure D.5: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set six generated using Random Sampling. The white dots in (c) indicate the sampling points.

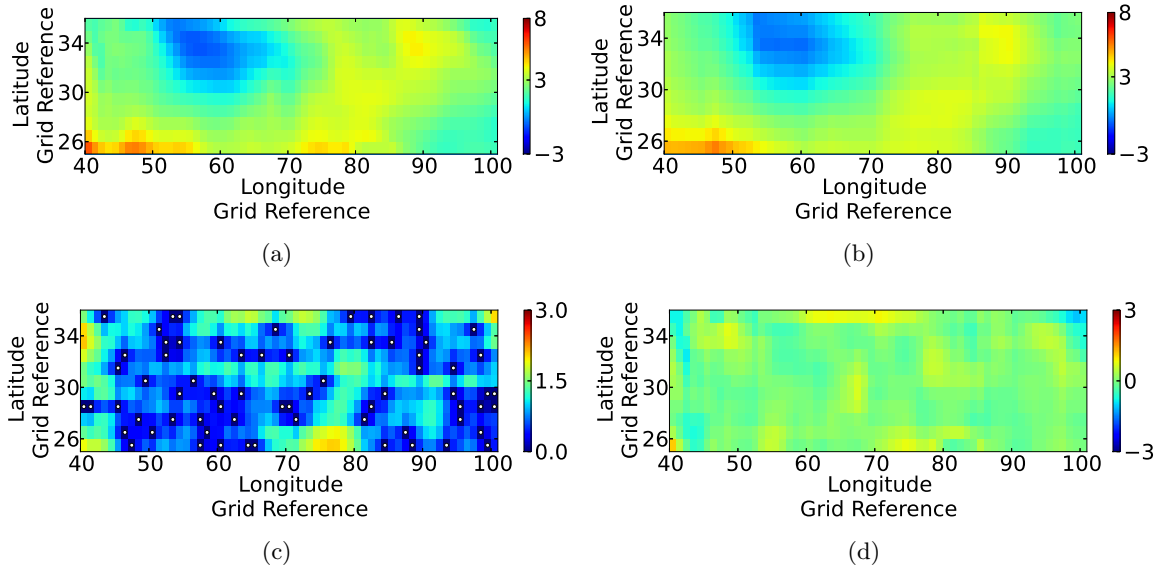


Figure D.6: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set seven generated using Random Sampling. The white dots in (c) indicate the sampling points.

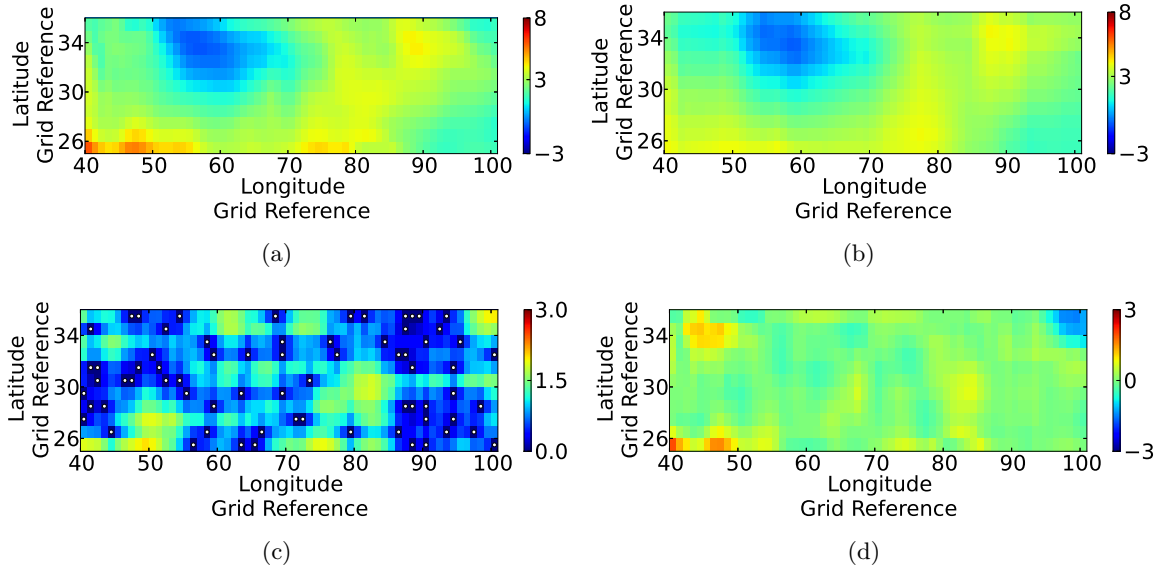


Figure D.7: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set nine generated using Random Sampling. The white dots in (c) indicate the sampling points.

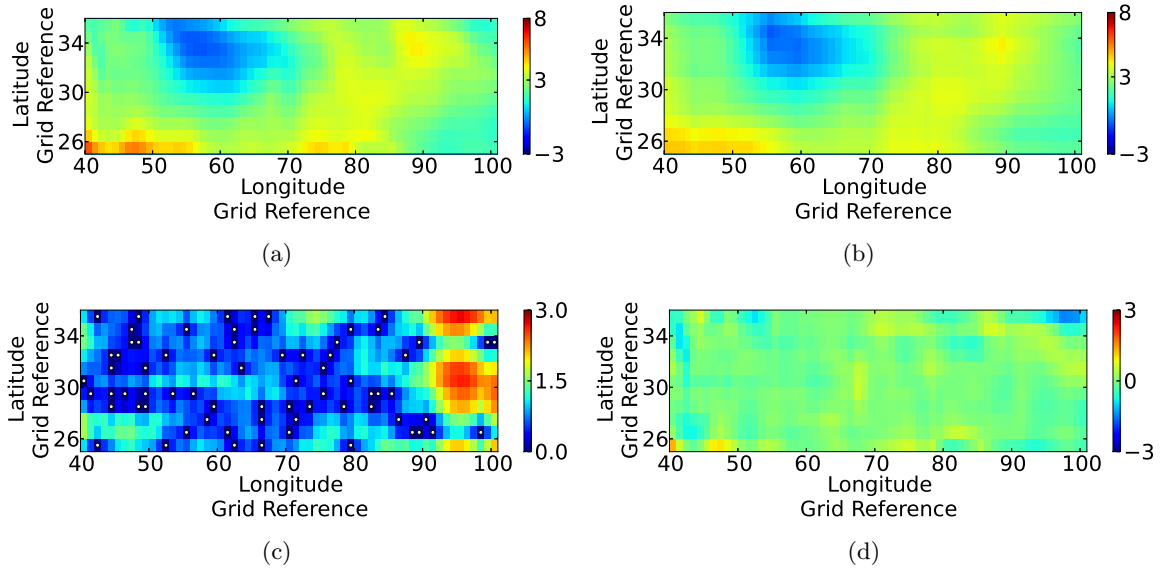


Figure D.8: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set ten generated using Random Sampling. The white dots in (c) indicate the sampling points.

D.2 Updated Kriging Variance Algorithm

Maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux, the estimated error variance and the true error across the chosen two-dimensional region for sets one, two, three, five, six, eight, nine and ten can be seen in Figures D.9 to D.16. These maps are given for the full sample set of size 70 generated using the Updated Kriging Variance Algorithm (See Chapter 10).

Figure D.17 shows how the errors progressed with the addition of points using the Updated Kriging Variance Algorithm across sets one, two, three, five, six, eight, nine and ten.

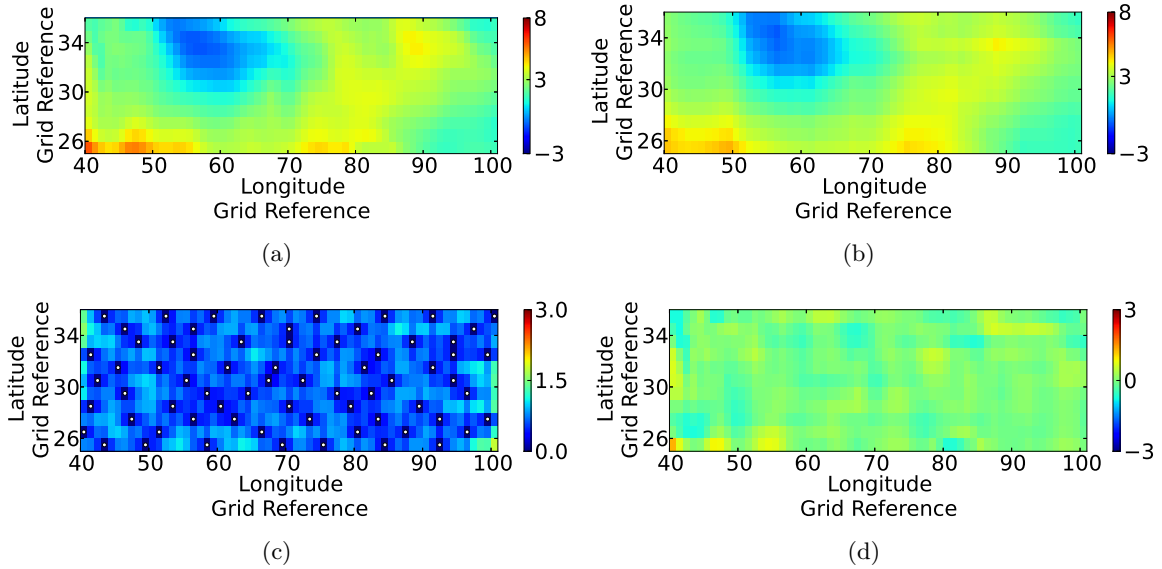


Figure D.9: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set one generated using Updated Kriging Variance Algorithm sampling. The white dots in (c) indicate the sampling points.

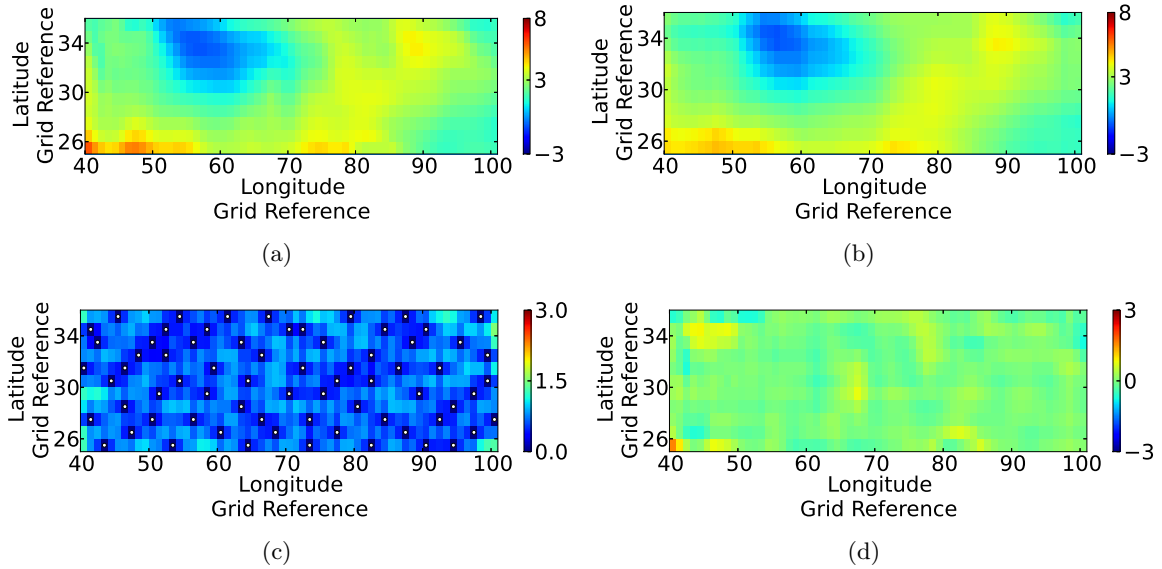


Figure D.10: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set two generated using Updated Kriging Variance Algorithm sampling. The white dots in (c) indicate the sampling points.

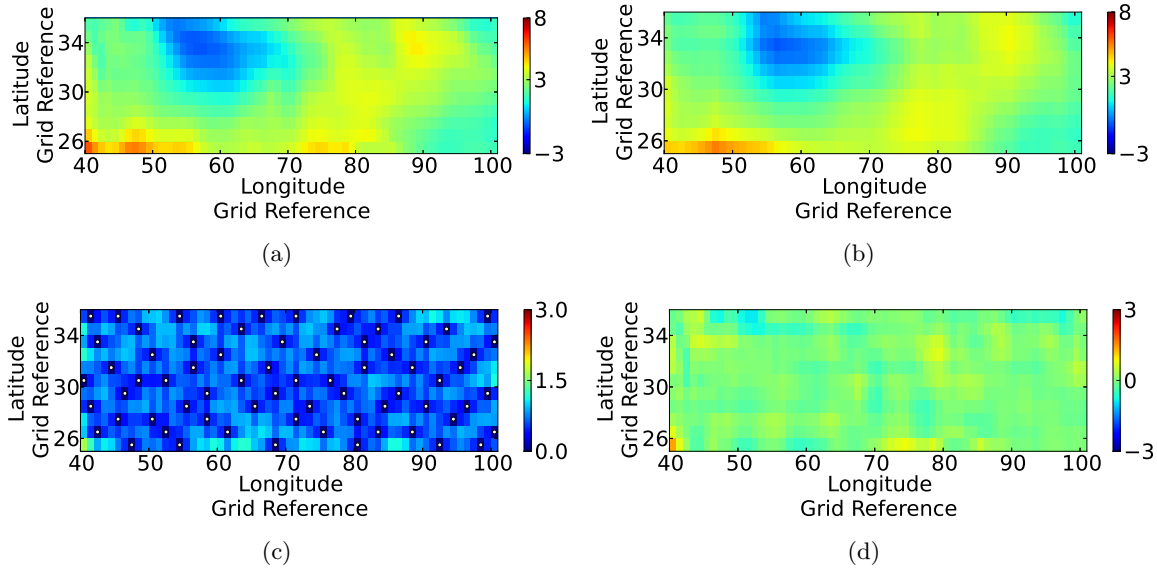


Figure D.11: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set three generated using Updated Kriging Variance Algorithm sampling. The white dots in (c) indicate the sampling points.

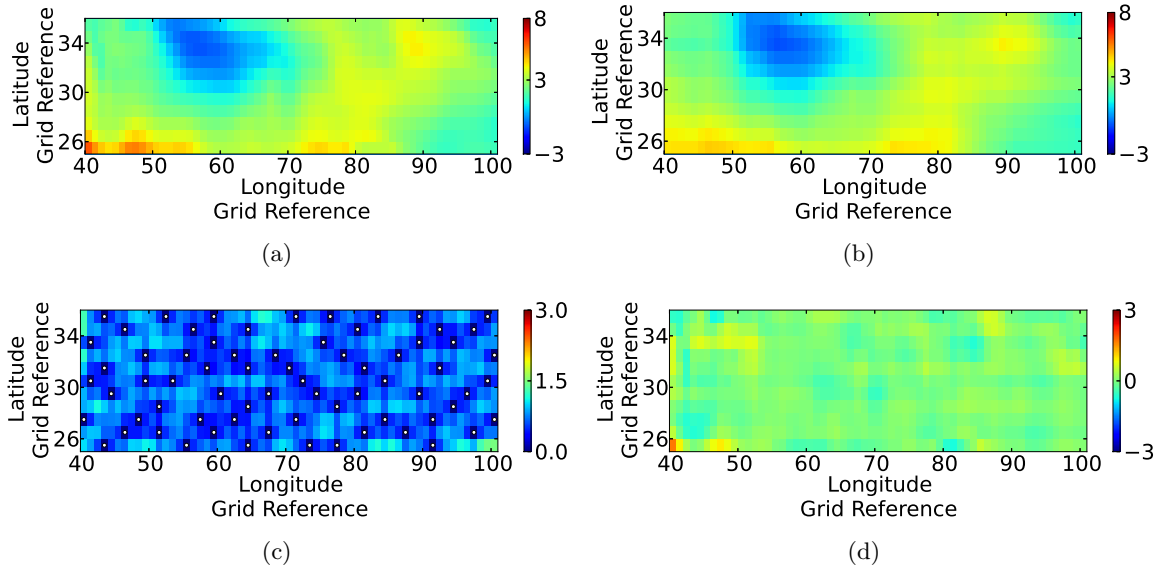


Figure D.12: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set five generated using Updated Kriging Variance Algorithm sampling. The white dots in (c) indicate the sampling points.

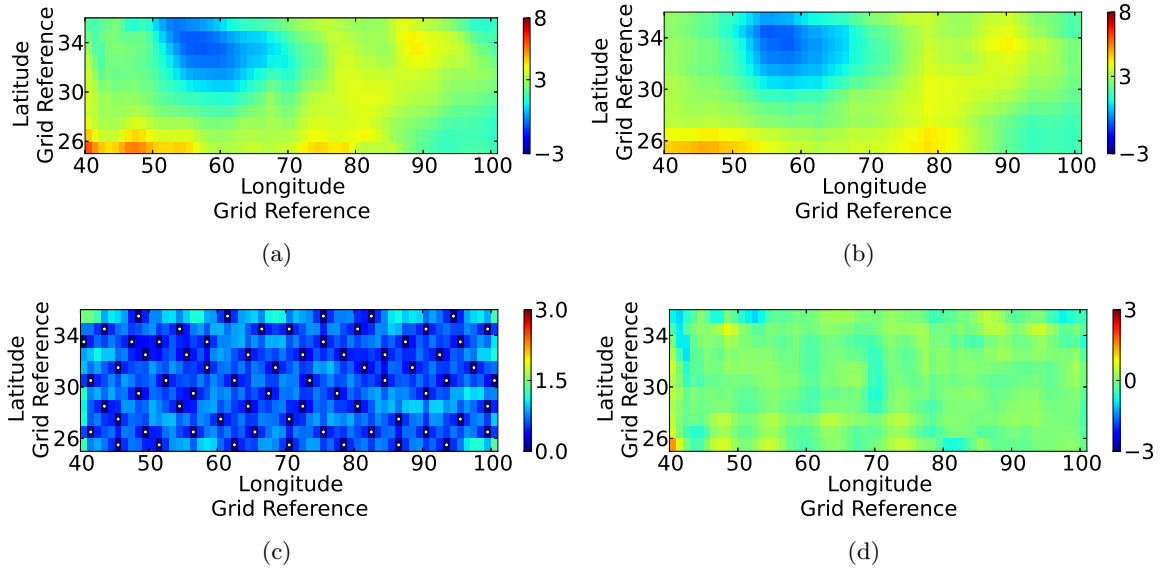


Figure D.13: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set six generated using Updated Kriging Variance Algorithm sampling. The white dots in (c) indicate the sampling points.

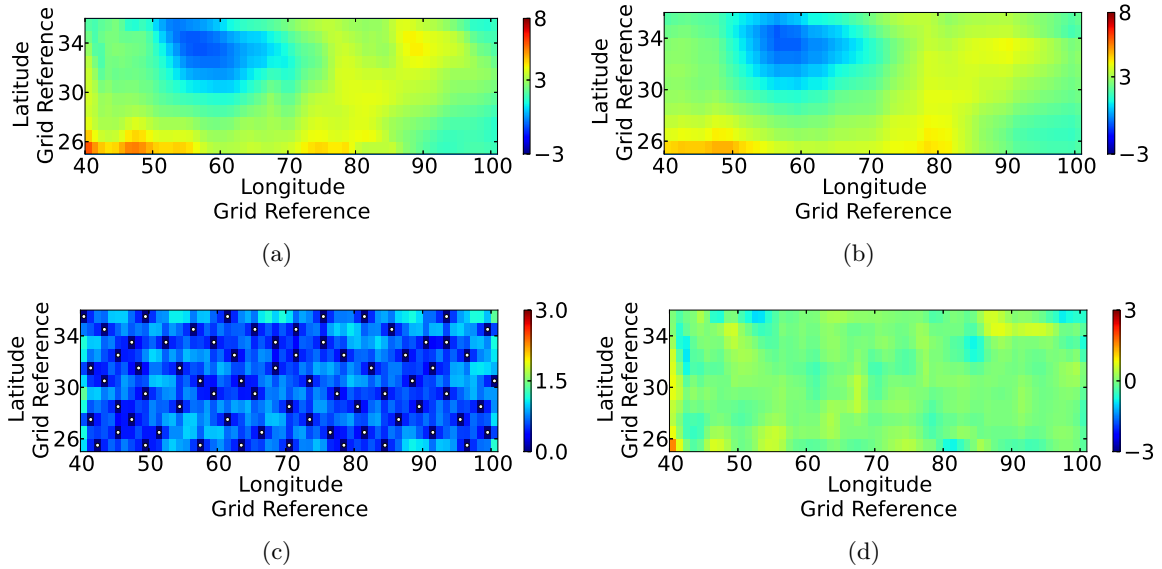


Figure D.14: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set eight generated using Updated Kriging Variance Algorithm sampling. The white dots in (c) indicate the sampling points.

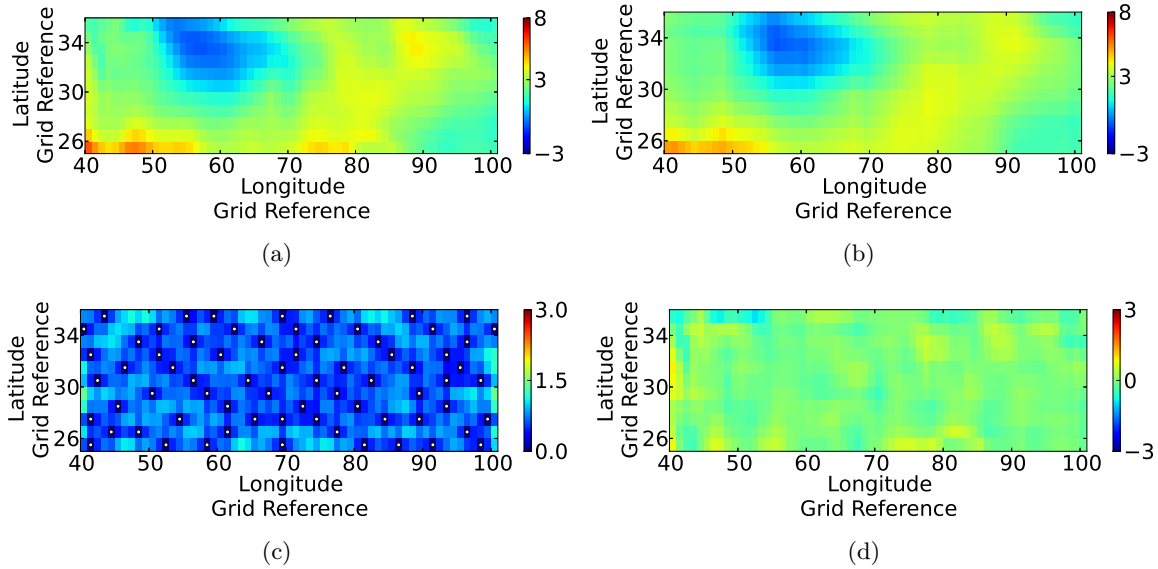


Figure D.15: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set nine generated using Updated Kriging Variance Algorithm sampling. The white dots in (c) indicate the sampling points.

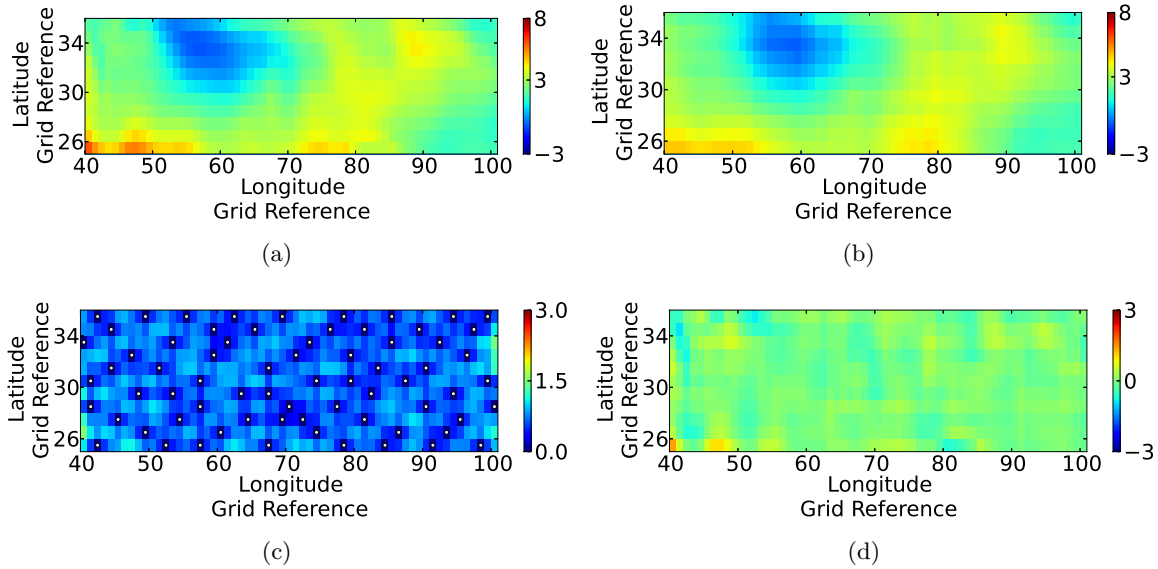


Figure D.16: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set ten generated using Updated Kriging Variance Algorithm sampling. The white dots in (c) indicate the sampling points.

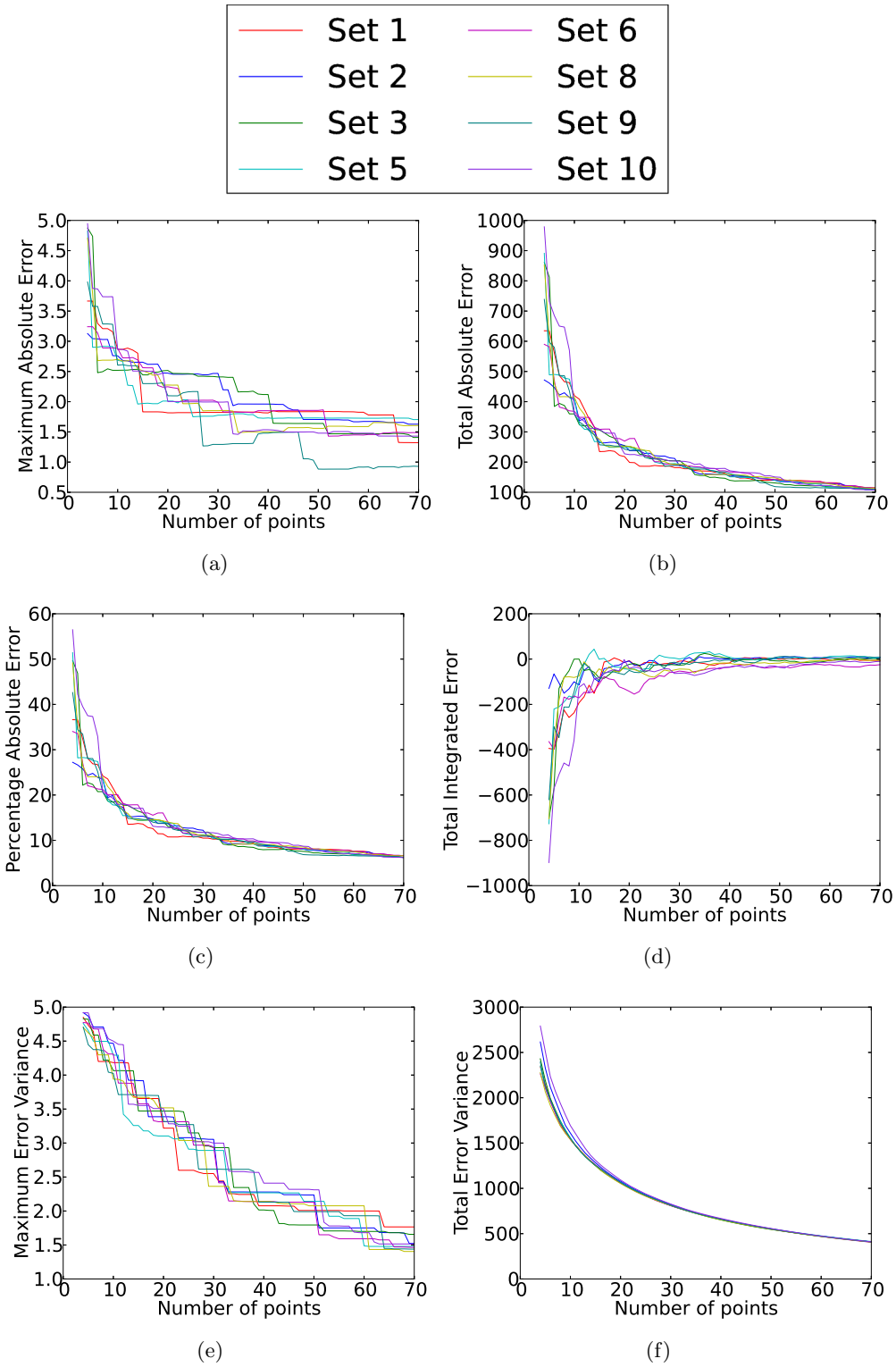


Figure D.17: Progression of (a) maximum absolute error, (b) total absolute error, (c) percentage absolute error, (d) total integrated error, (e) maximum estimated error variance and (e) total estimated error variance across various sets generated using the Updated Kriging Variance Algorithm.

D.3 Addition of a Point at Point of Maximum Estimated Error Variance

Maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux, the estimated error variance and the true error across the chosen two-dimensional region for sets one, two, three, four, five, six, seven and ten can be seen in Figures D.18 to D.25. These maps are given for the full sample set of size 70 generated using the Addition of a Point at Point of Maximum Estimated Error Variance (See Chapter 11).

Figure D.26 shows how the errors progressed with the additional of points using the method of Addition of a Point at Point of Maximum Estimated Error Variance across sets one, two, three, four, five, six, seven and ten.

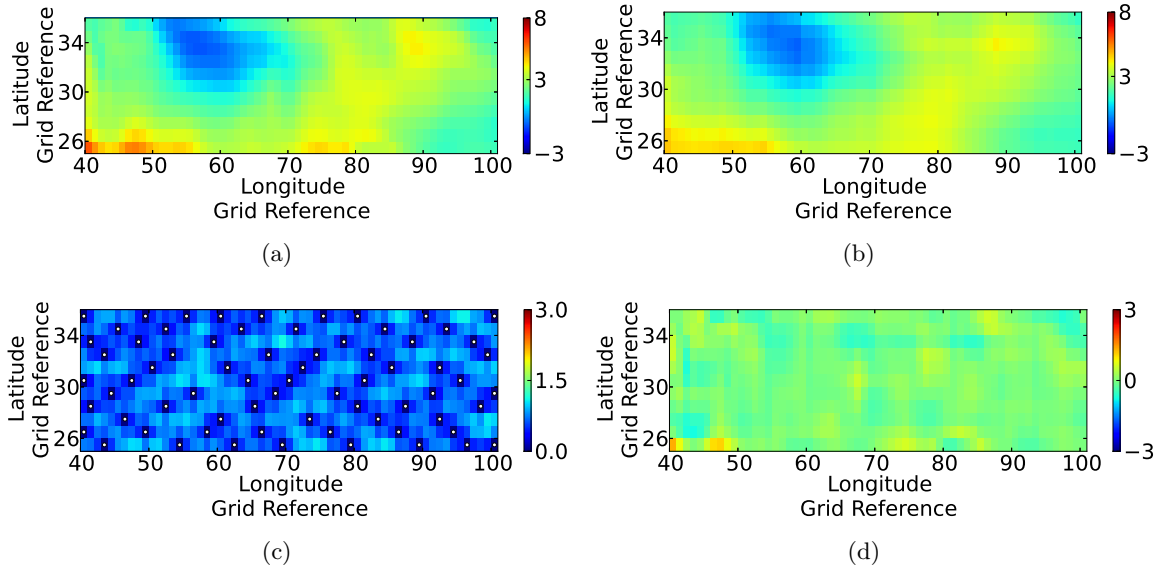


Figure D.18: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set one generated using Addition of a Point at Point of Maximum Estimated Error Variance sampling. The white dots in (c) indicate the sampling points.

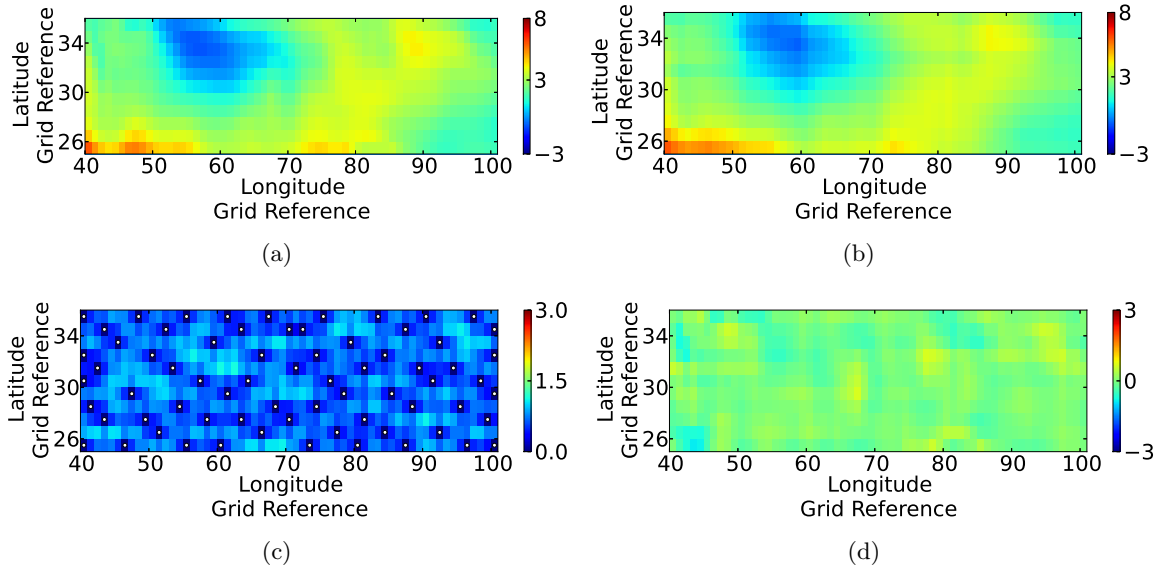


Figure D.19: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set two generated using Addition of a Point at Point of Maximum Estimated Error Variance sampling. The white dots in (c) indicate the sampling points.

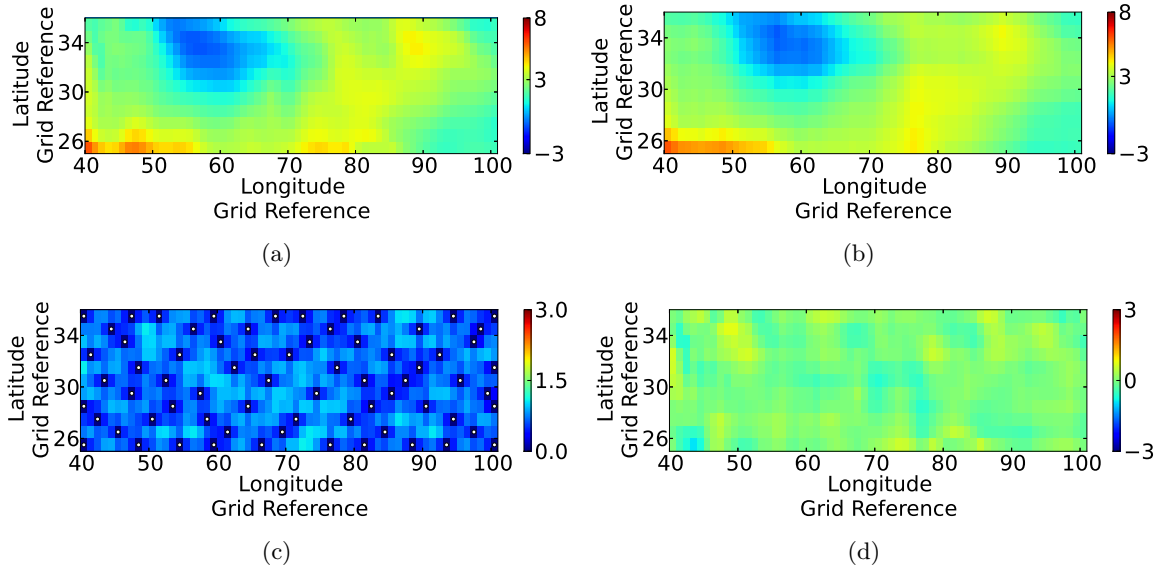


Figure D.20: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set three generated using Addition of a Point at Point of Maximum Estimated Error Variance sampling. The white dots in (c) indicate the sampling points.

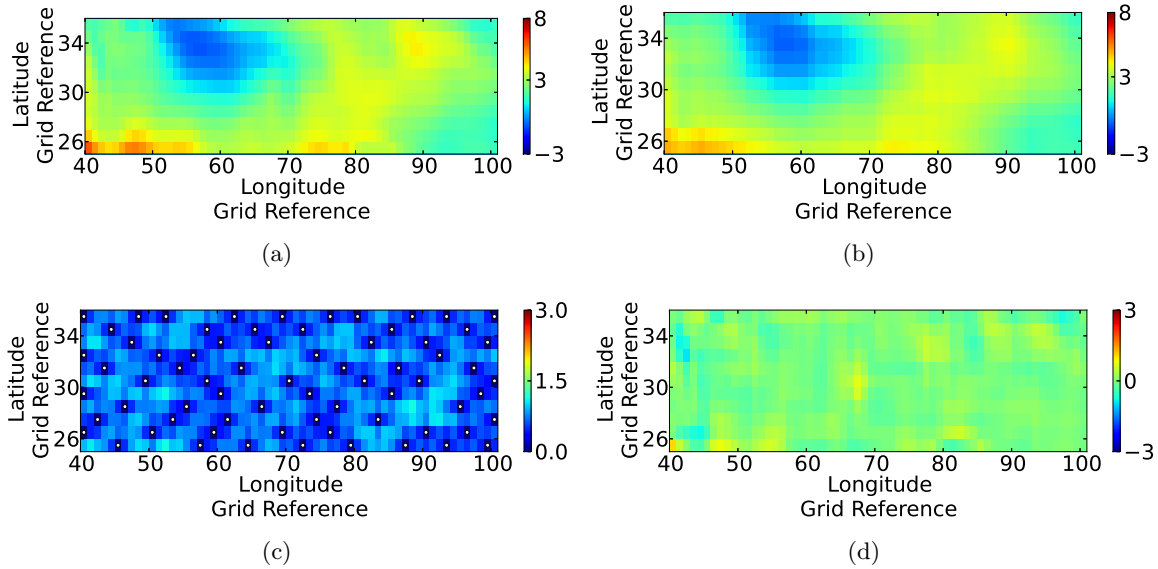


Figure D.21: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set four generated using Addition of a Point at Point of Maximum Estimated Error Variance sampling. The white dots in (c) indicate the sampling points.

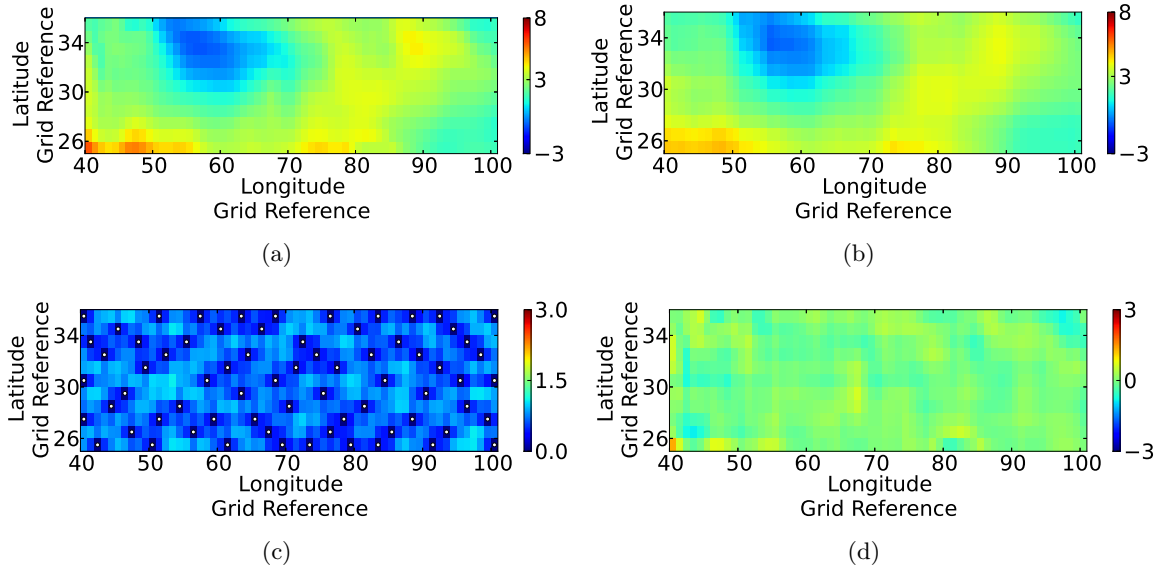


Figure D.22: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set five generated using Addition of a Point at Point of Maximum Estimated Error Variance sampling. The white dots in (c) indicate the sampling points.

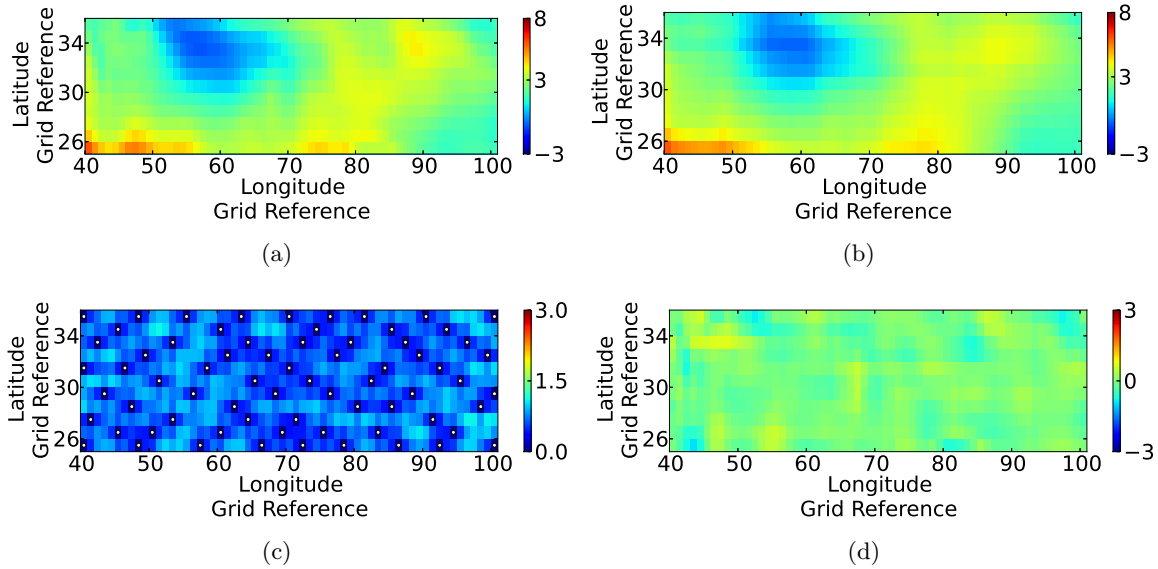


Figure D.23: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set six generated using Addition of a Point at Point of Maximum Estimated Error Variance sampling. The white dots in (c) indicate the sampling points.

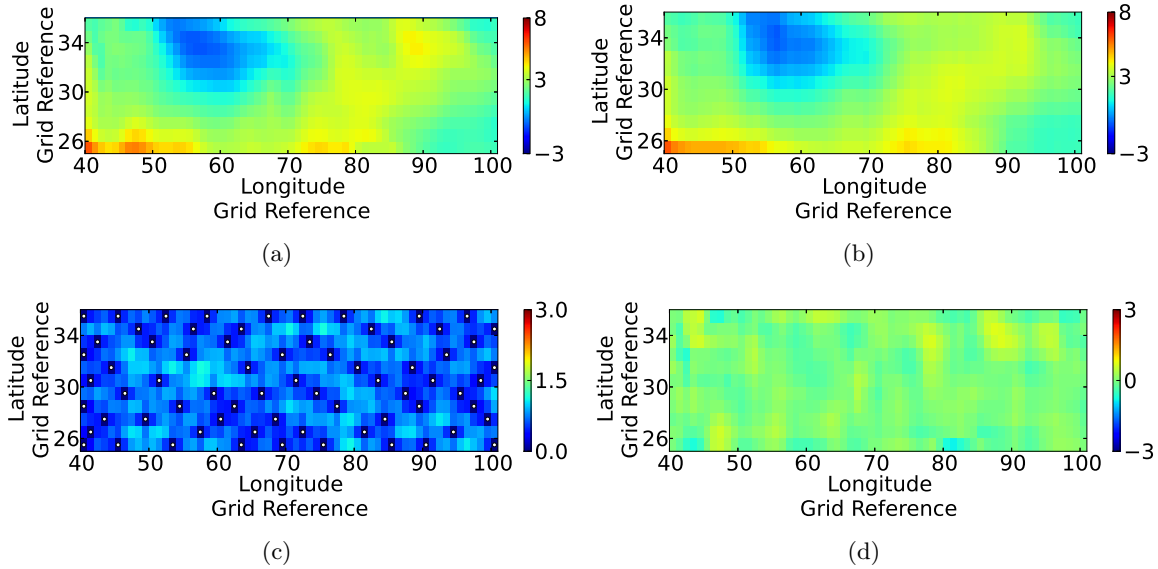


Figure D.24: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set seven generated using Addition of a Point at Point of Maximum Estimated Error Variance sampling. The white dots in (c) indicate the sampling points.

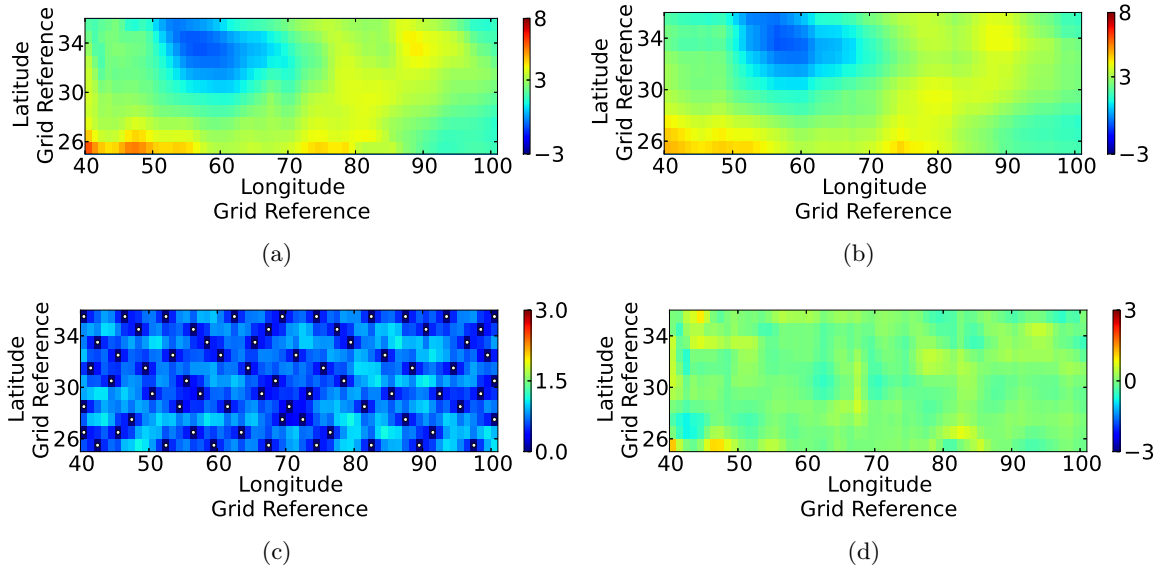


Figure D.25: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set ten generated using Addition of a Point at Point of Maximum Estimated Error Variance sampling. The white dots in (c) indicate the sampling points.

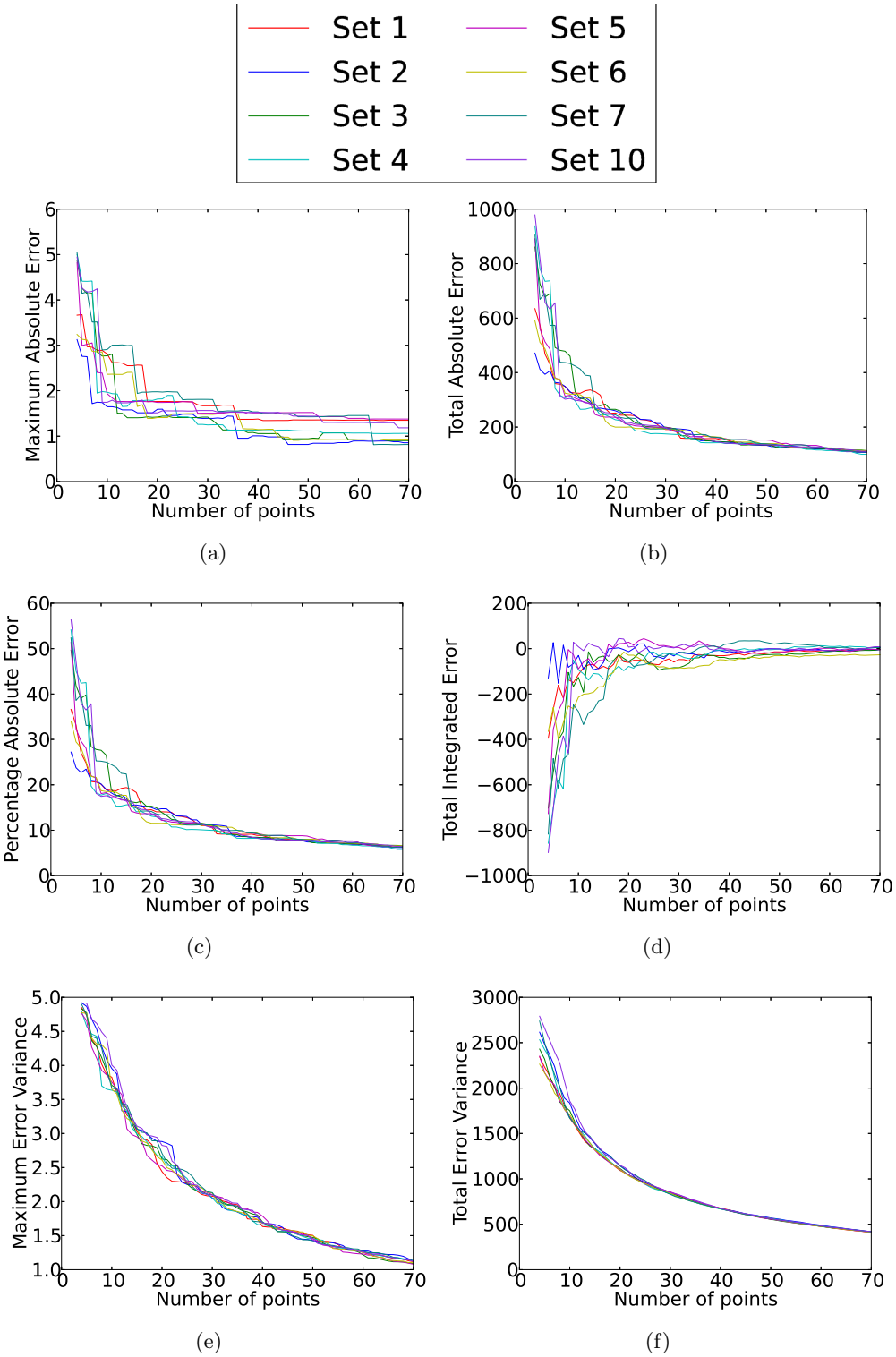


Figure D.26: Progression of (a) maximum absolute error, (b) total absolute error, (c) percentage absolute error, (d) total integrated error, (e) maximum estimated error variance and (e) total estimated error variance across various sets generated using the Addition of a Point at Point of Maximum Estimated Error Variance.

D.4 Addition of a Point at Point of Maximum Absolute Error

Maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux, the estimated error variance and the true error across the chosen two-dimensional region for sets one, two, three, four, five, seven, nine and ten can be seen in Figures D.27 to D.34. These maps are given for the full sample set of size 70 generated using the Addition of a Point at Point of Maximum Absolute Error (See Chapter 12).

Figure D.35 shows how the errors progressed with the additional of points using the method of Addition of a Point at Point of Maximum Absolute Error across sets one, two, three, four, five, seven, nine and ten.

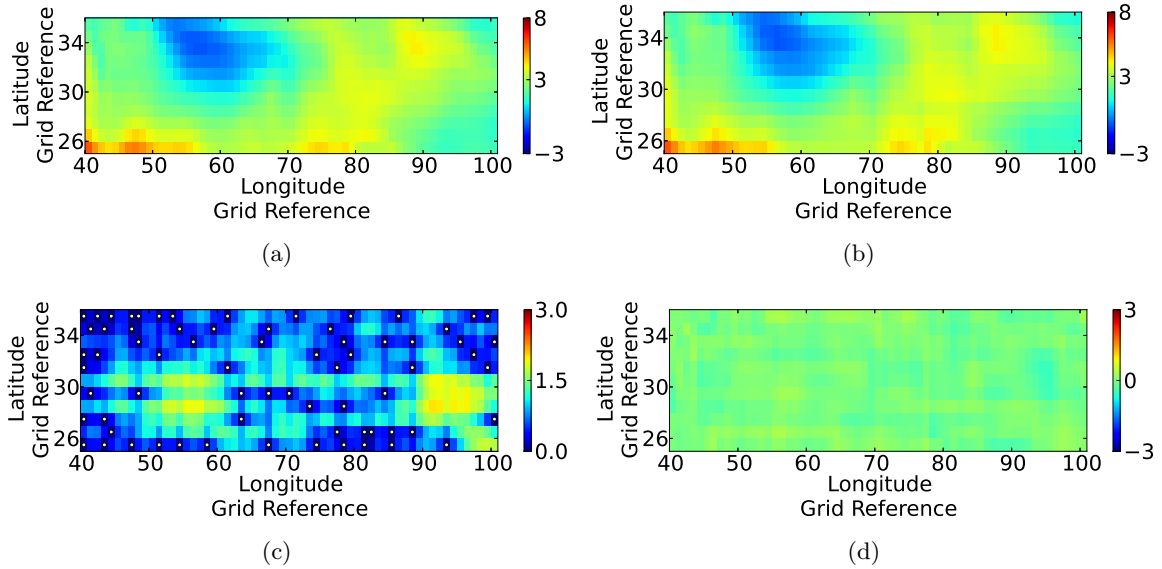


Figure D.27: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set one generated using Addition of a Point at Point of Maximum Absolute Error sampling. The white dots in (c) indicate the sampling points.

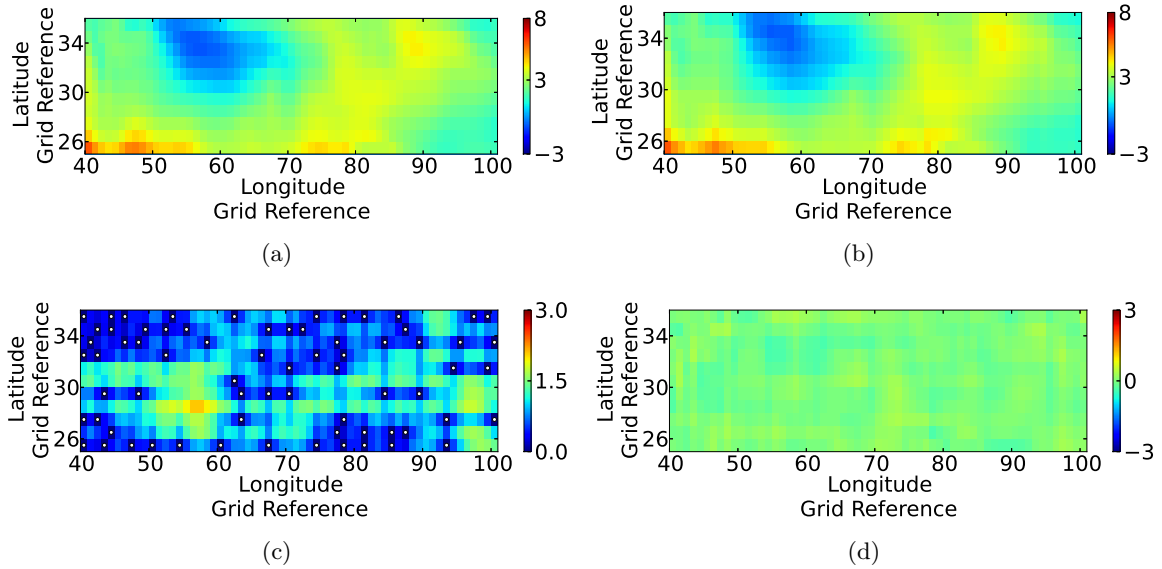


Figure D.28: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set two generated using Addition of a Point at Point of Maximum Absolute Error sampling. The white dots in (c) indicate the sampling points.

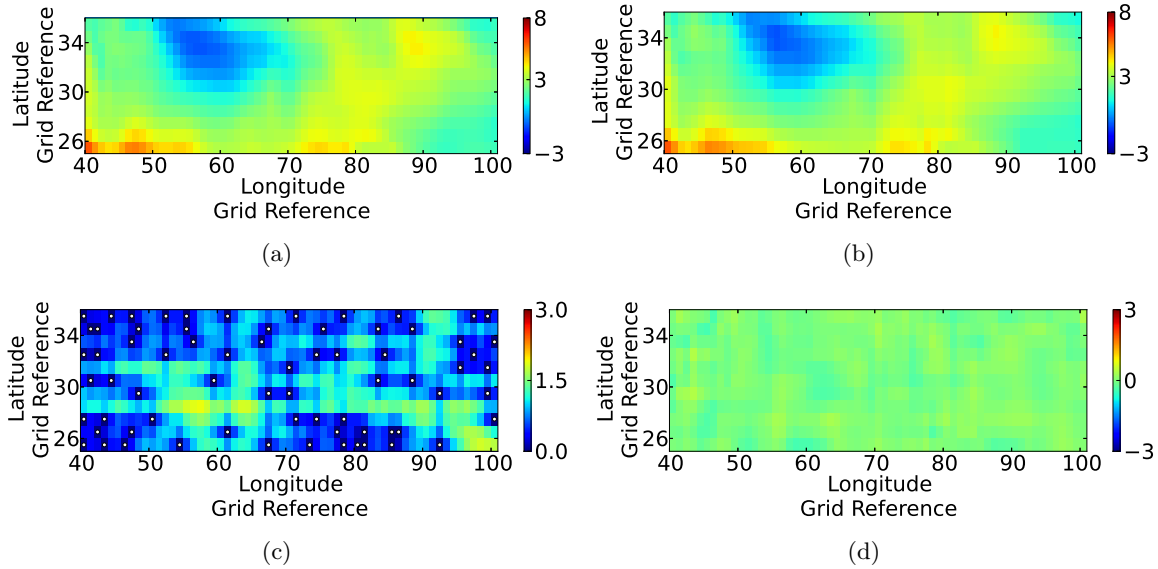


Figure D.29: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set three generated using Addition of a Point at Point of Maximum Absolute Error sampling. The white dots in (c) indicate the sampling points.

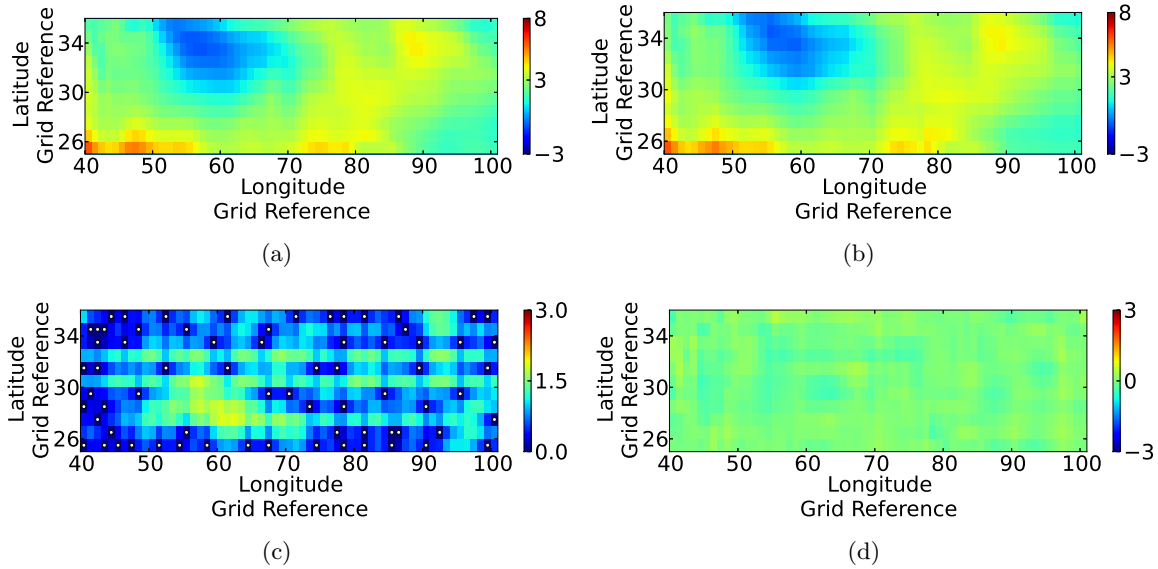


Figure D.30: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set four generated using Addition of a Point at Point of Maximum Absolute Error sampling. The white dots in (c) indicate the sampling points.

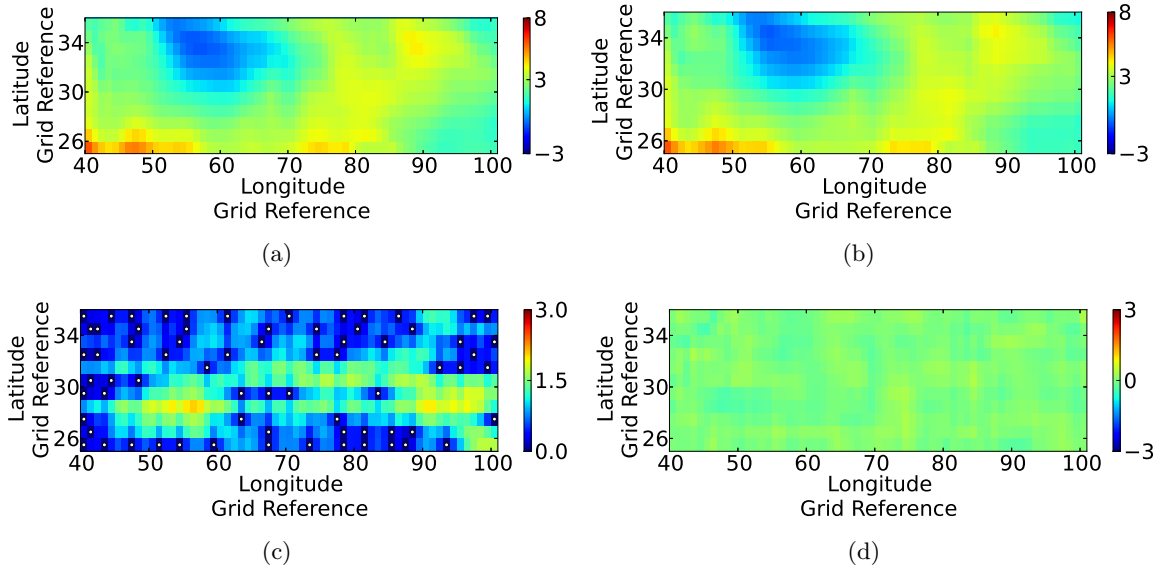


Figure D.31: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set five generated using Addition of a Point at Point of Maximum Absolute Error sampling. The white dots in (c) indicate the sampling points.

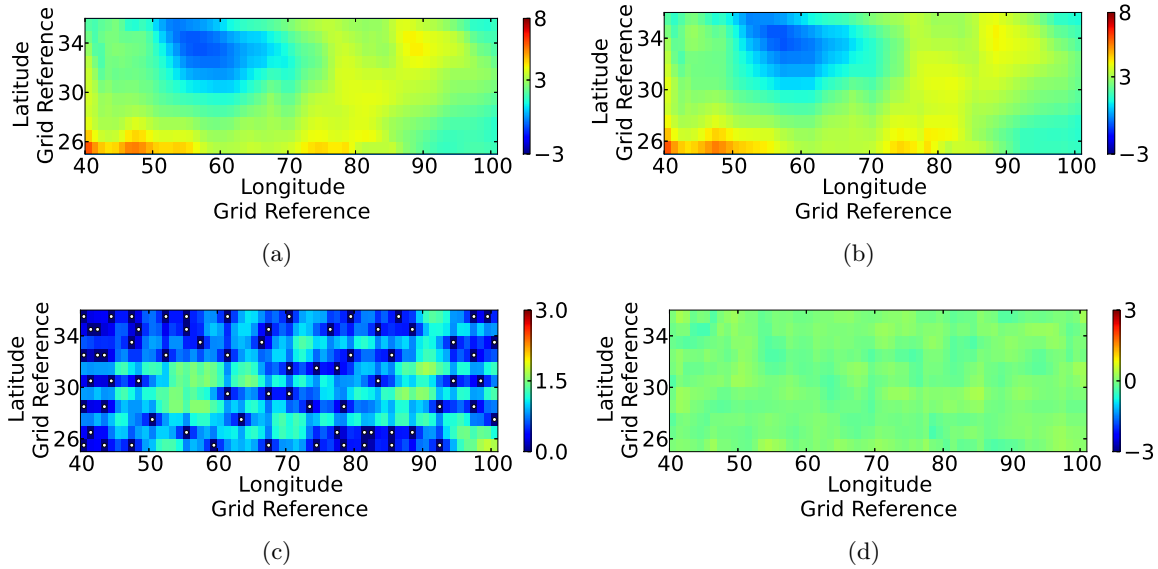


Figure D.32: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set seven generated using Addition of a Point at Point of Maximum Absolute Error sampling. The white dots in (c) indicate the sampling points.

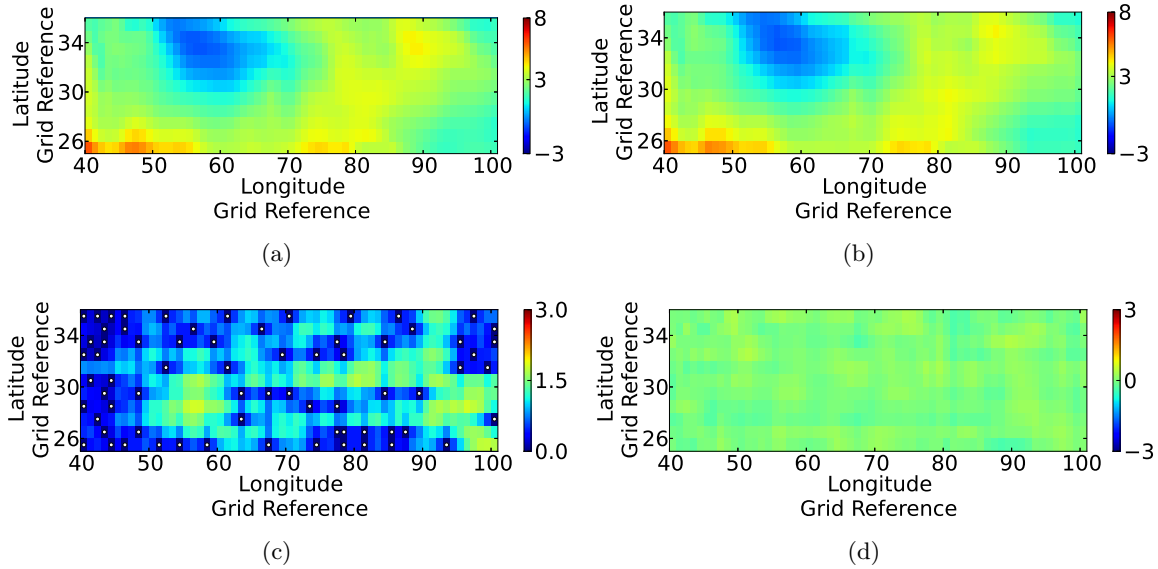


Figure D.33: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set nine generated using Addition of a Point at Point of Maximum Absolute Error sampling. The white dots in (c) indicate the sampling points.

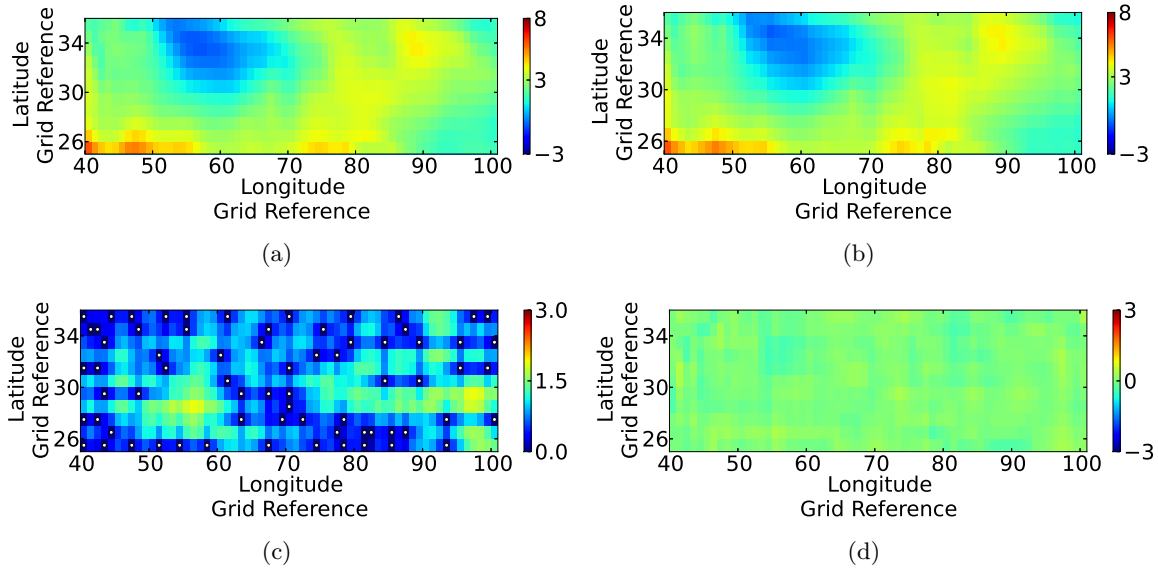


Figure D.34: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for set ten generated using Addition of a Point at Point of Maximum Absolute Error sampling. The white dots in (c) indicate the sampling points.

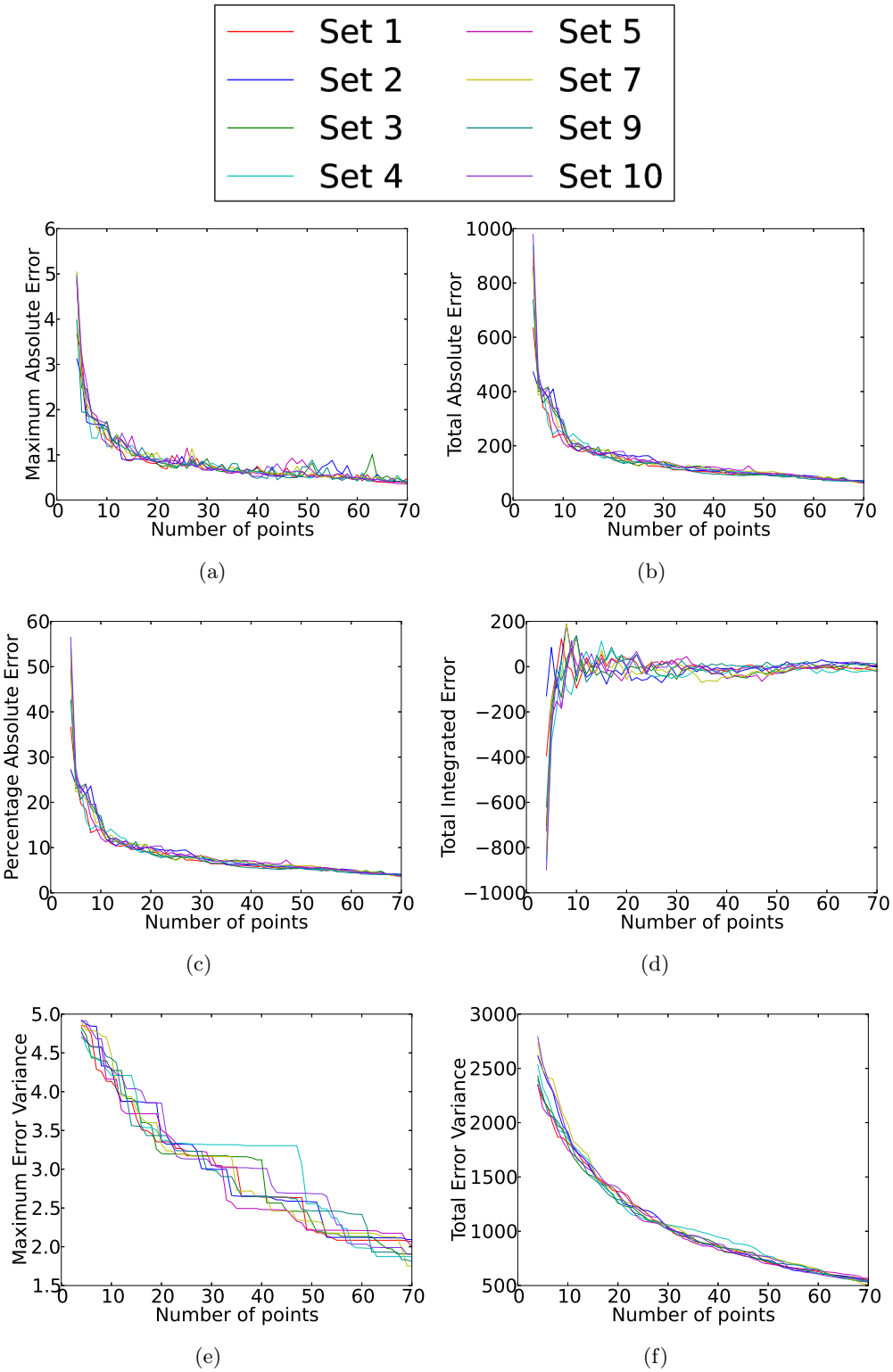


Figure D.35: Progression of (a) maximum absolute error, (b) total absolute error, (c) percentage absolute error, (d) total integrated error, (e) maximum estimated error variance and (e) total estimated error variance across various sets generated using the Addition of a Point at Point of Maximum Absolute Error.

D.5 Genetic Algorithm Designed to Minimise Total Absolute Error

Maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux, the estimated error variance and the true error across the chosen two-dimensional region for runs one, three and four can be seen in Figures D.36 to D.38. These maps are given for the full sample set of size 70 generated using the Genetic Algorithm designed to Minimise Total Absolute Error (See Chapter 13.8).

Figure D.39 shows how the errors progressed across the iterations using the Genetic Algorithm designed to Minimise Total Absolute Error across runs one, three and four.

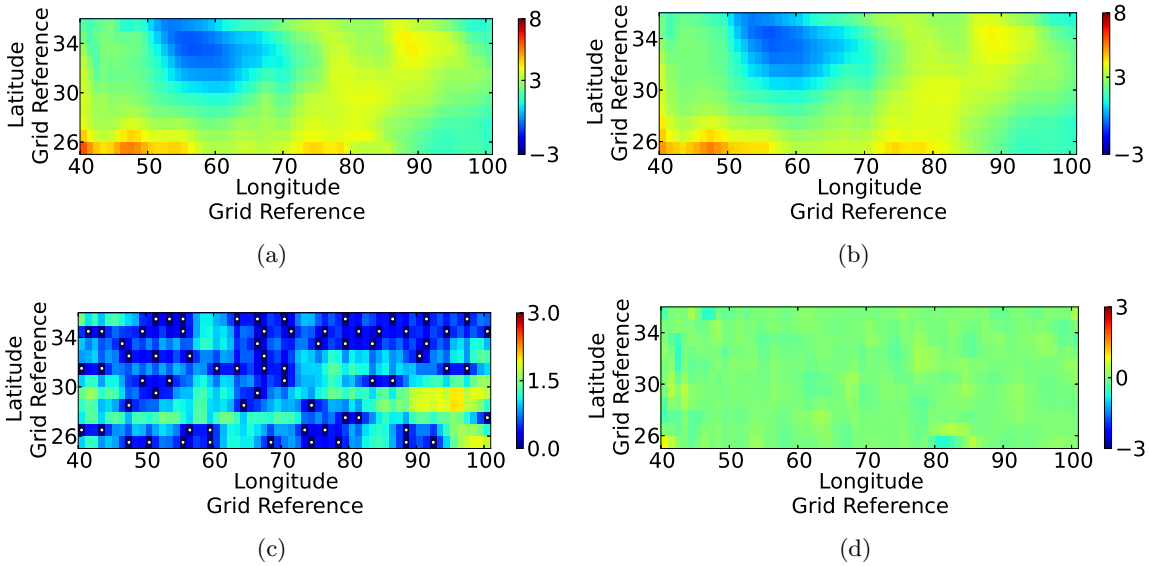


Figure D.36: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for run one generated using the Genetic Algorithm designed to Minimise Total Absolute Error. The white dots in (c) indicate the sampling points.

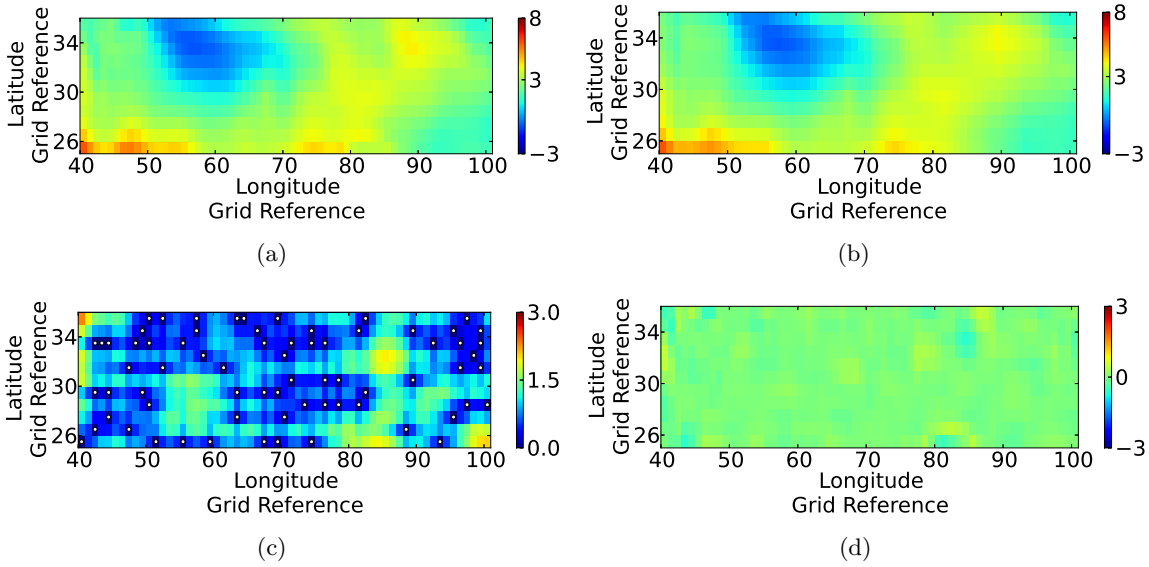


Figure D.37: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for run three generated using the Genetic Algorithm designed to Minimise Total Absolute Error. The white dots in (c) indicate the sampling points.

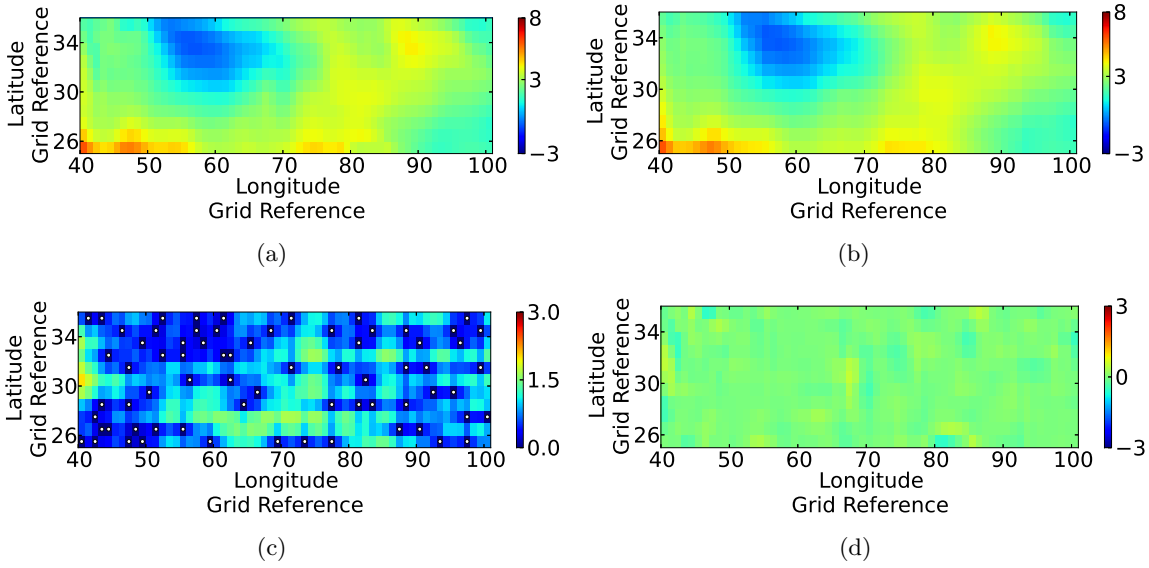


Figure D.38: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for run four generated using the Genetic Algorithm designed to Minimise Total Absolute Error. The white dots in (c) indicate the sampling points.

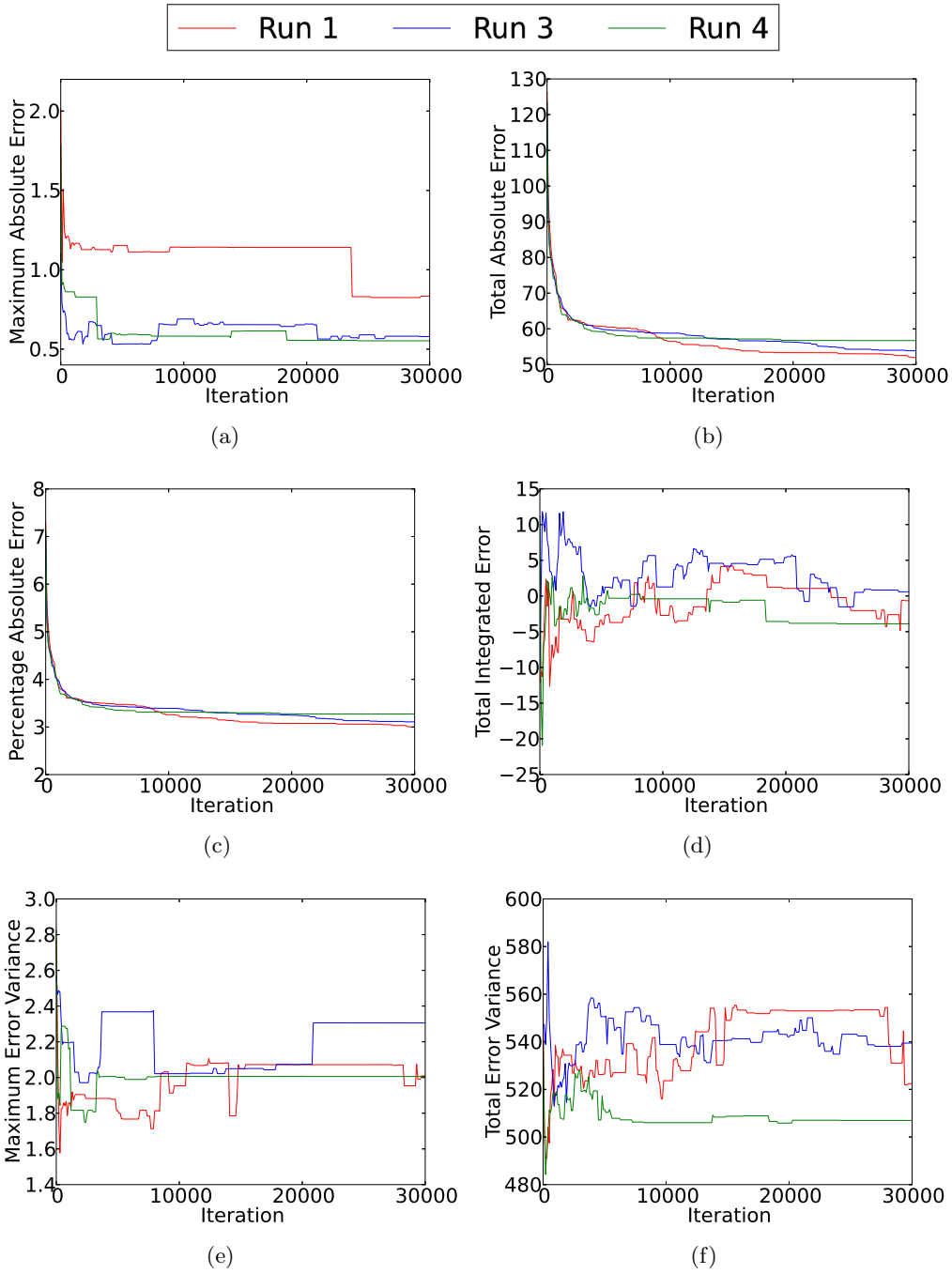


Figure D.39: Progression of (a) maximum absolute error, (b) total absolute error, (c) percentage absolute error, (d) total integrated error, (e) maximum estimated error variance and (e) total estimated error variance across all iterations for various runs of the Genetic Algorithm designed to Minimise Total Absolute Error.

D.6 Genetic Algorithm Designed to Minimise Total Estimated Error Variance

Maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux, the estimated error variance and the true error across the chosen two-dimensional region for runs two, four and five can be seen in Figures D.40 to D.42. These maps are given for the full sample set of size 70 generated using the Genetic Algorithm designed to Minimise Total Estimated Error Variance (See Chapter 13.9).

Figure D.43 shows how the errors progressed with the number of iterations using the Genetic Algorithm designed to Minimise Total Estimated Error Variance across runs two, four and five.

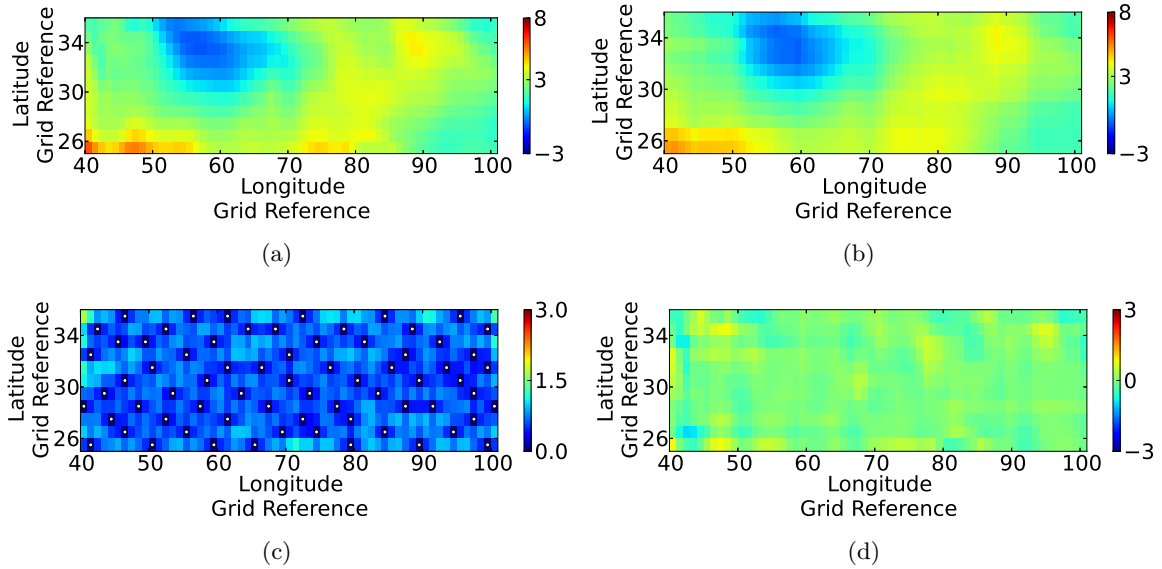


Figure D.40: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for run two generated using the Genetic Algorithm designed to Minimise Total Estimated Error Variance. The white dots in (c) indicate the sampling points.

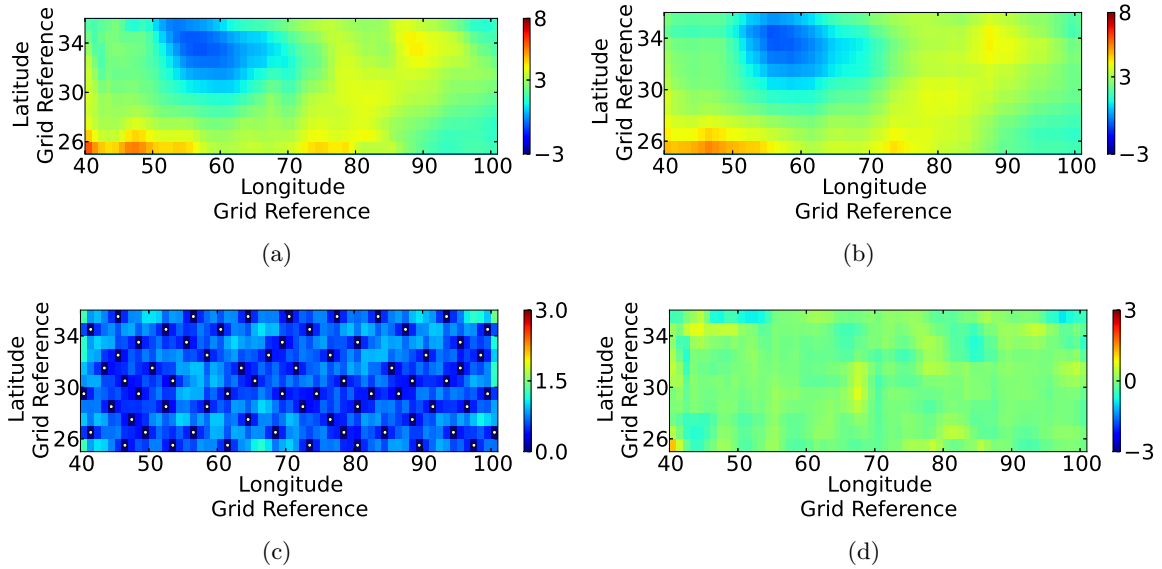


Figure D.41: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for run four generated using the Genetic Algorithm designed to Minimise Total Estimated Error Variance. The white dots in (c) indicate the sampling points.

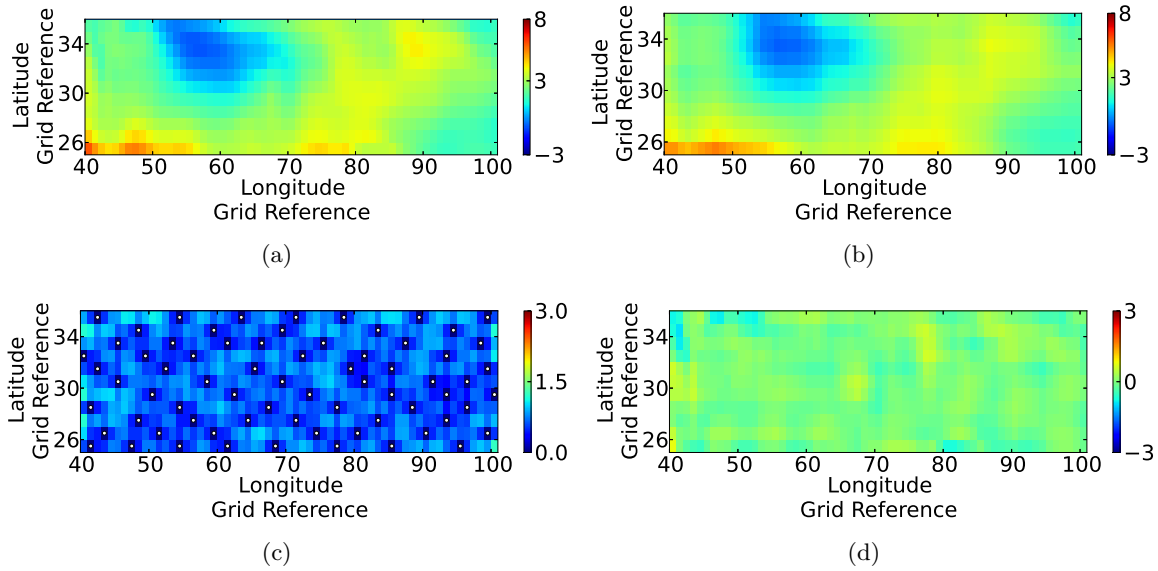


Figure D.42: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for run five generated using the Genetic Algorithm designed to Minimise Total Estimated Error Variance. The white dots in (c) indicate the sampling points.

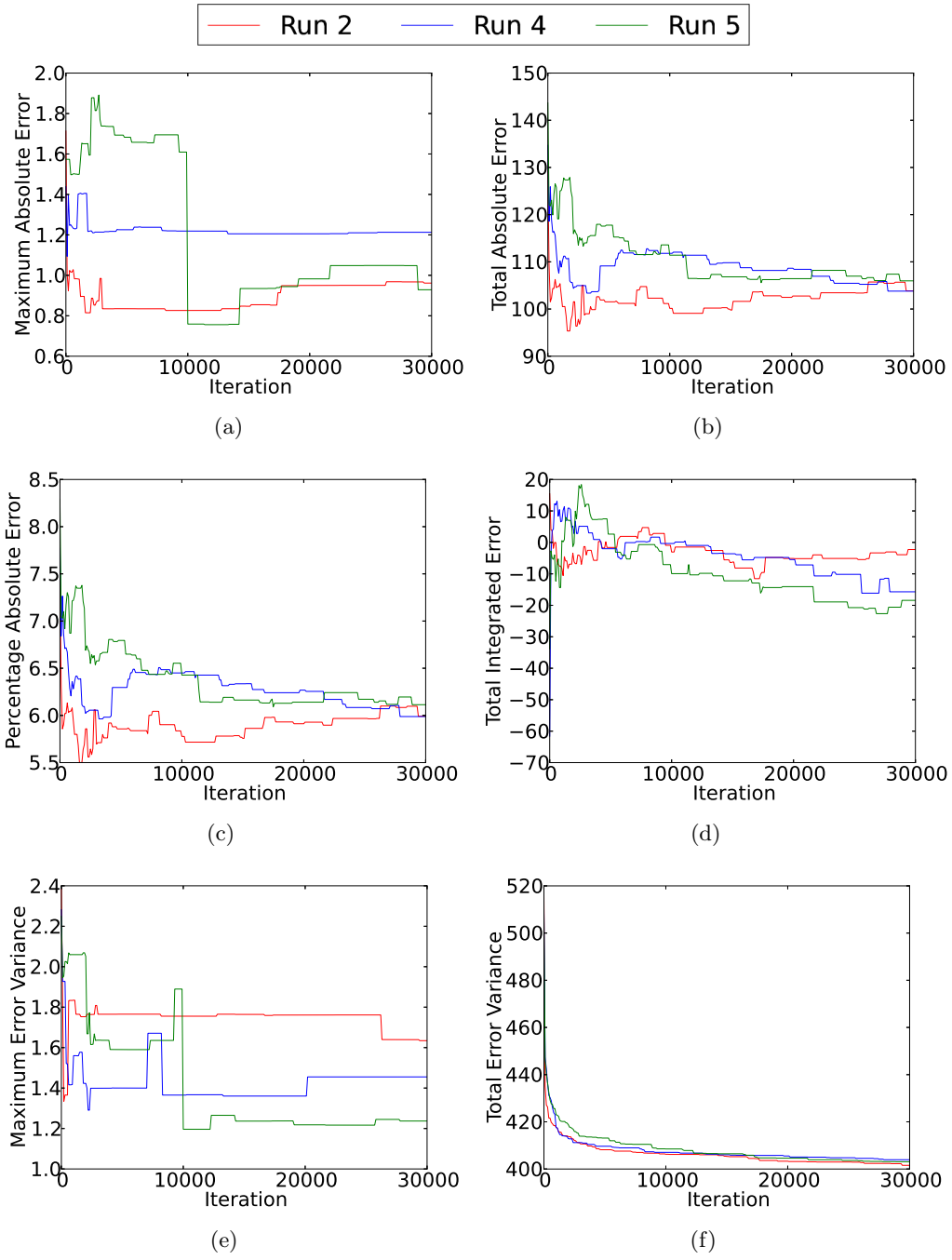


Figure D.43: Progression of (a) maximum absolute error, (b) total absolute error, (c) percentage absolute error, (d) total integrated error, (e) maximum estimated error variance and (e) total estimated error variance across all iterations for various runs of the Genetic Algorithm designed to Minimise Total Estimated Error Variance.

D.7 Hybridised Genetic Algorithm Designed to Minimise Total Absolute Error

Maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux, the estimated error variance and the true error across the chosen two-dimensional region for runs one, two and four can be seen in Figures D.44 to D.46. These maps are given for the full sample set of size 70 generated using the Hybridised Genetic Algorithm designed to Minimise Total Absolute Error (See Chapter 13.10).

Figure D.47 shows how the errors progressed with the number of iterations using the Hybridised Genetic Algorithm designed to Minimise Total Absolute Error across runs one, two and four.

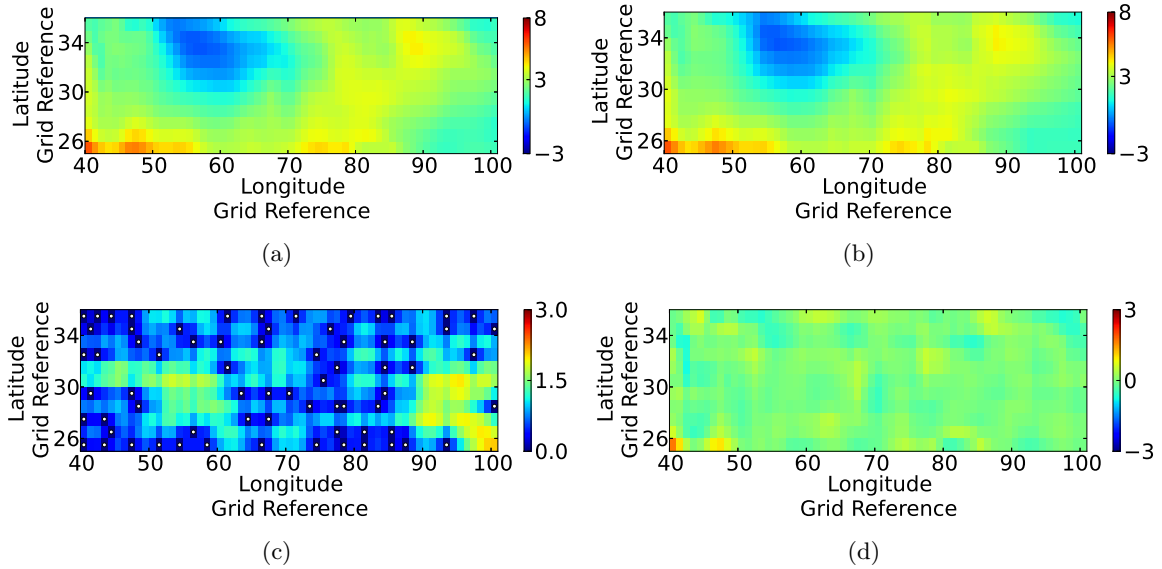


Figure D.44: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for run one generated using the Hybridised Genetic Algorithm designed to Minimise Total Absolute Error. The white dots in (c) indicate the sampling points.

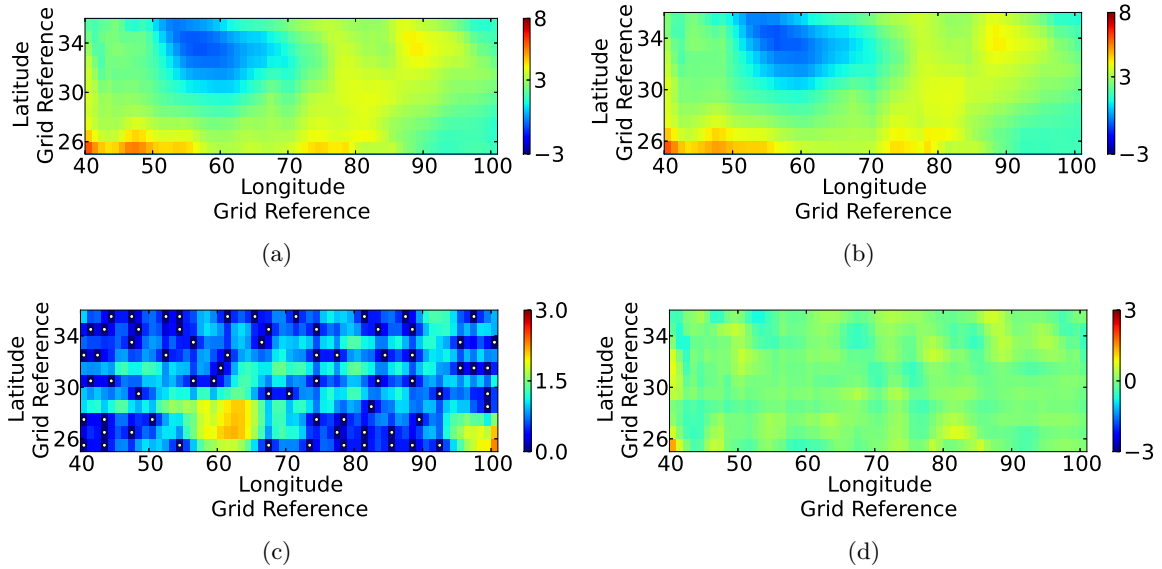


Figure D.45: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for run two generated using the Hybridised Genetic Algorithm designed to Minimise Total Absolute Error. The white dots in (c) indicate the sampling points.

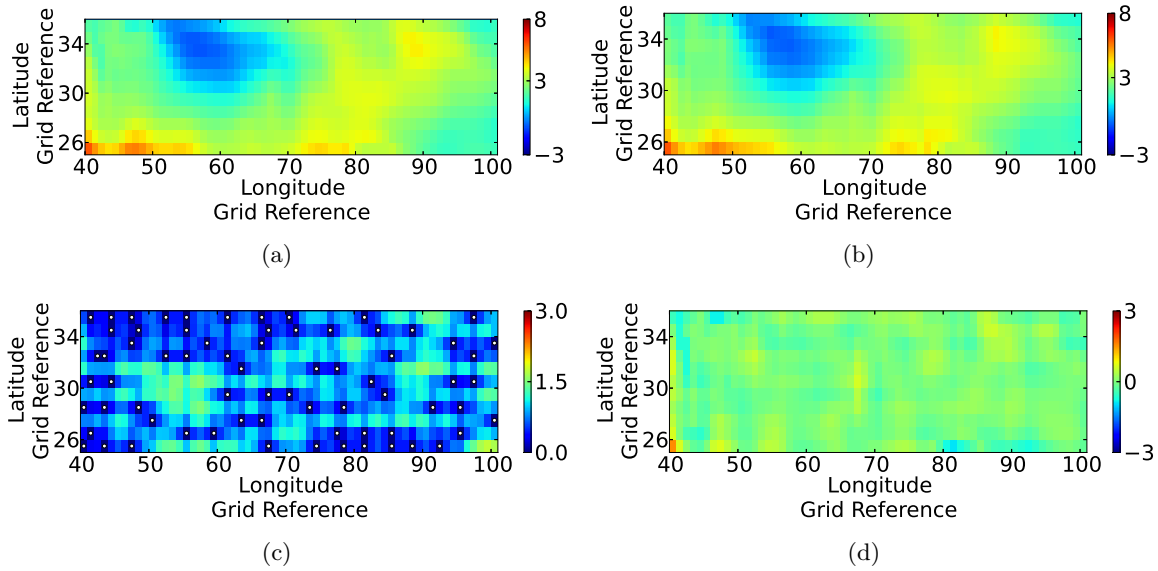


Figure D.46: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for run four generated using the Hybridised Genetic Algorithm designed to Minimise Total Absolute Error. The white dots in (c) indicate the sampling points.

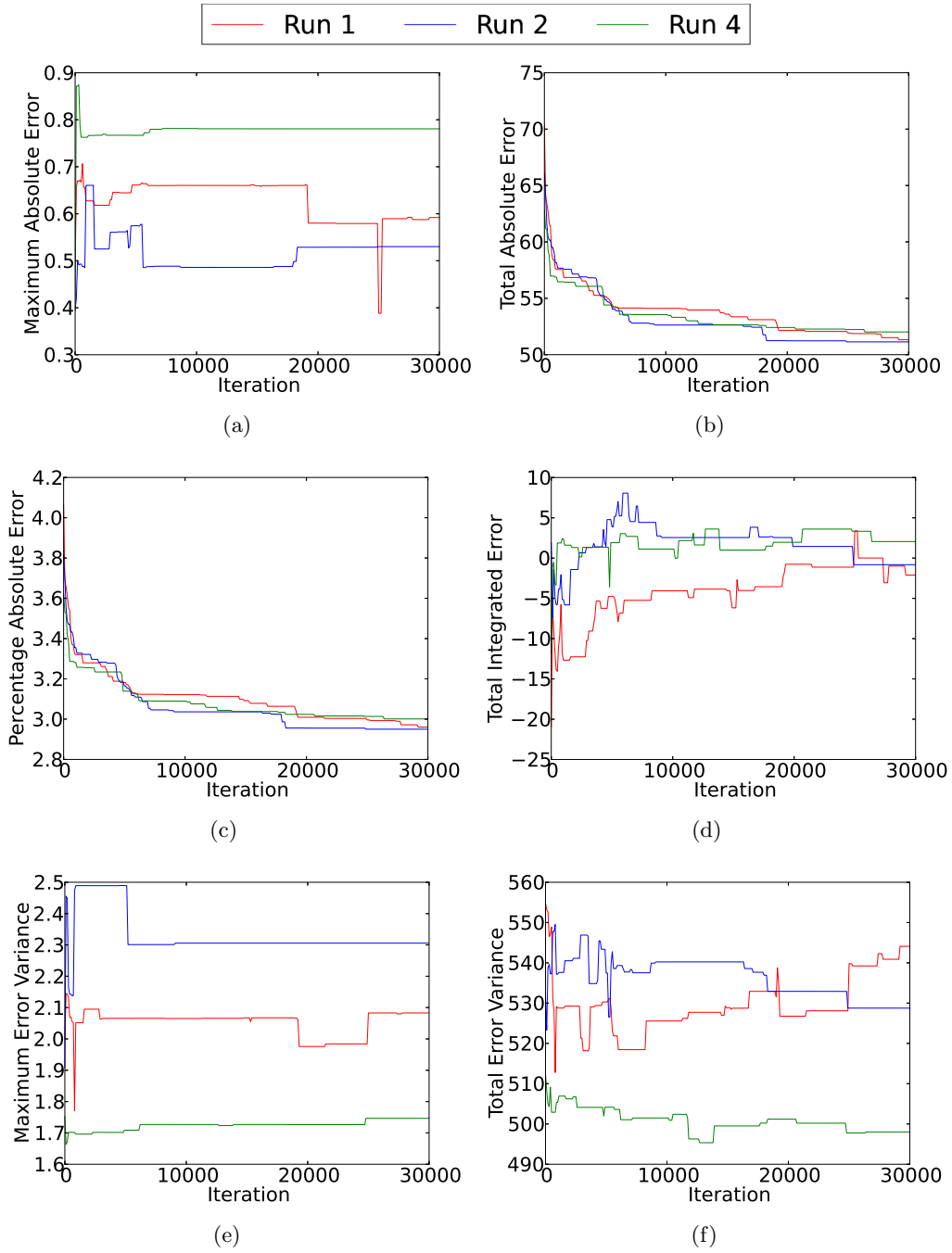


Figure D.47: Progression of (a) maximum absolute error, (b) total absolute error, (c) percentage absolute error, (d) total integrated error, (e) maximum estimated error variance and (e) total estimated error variance across all iterations for various runs of the Hybridised Genetic Algorithm designed to Minimise Total Absolute Error.

D.8 Hybridised Genetic Algorithm Designed to Minimise Total Estimated Error Variance

Maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux, the estimated error variance and the true error across the chosen two-dimensional region for runs three, four and five can be seen in Figures D.48 to D.50. These maps are given for the full sample set of size 70 generated using the Hybridised Genetic Algorithm designed to Minimise Total Estimated Error Variance (See Chapter 13.11).

Figure D.51 shows how the errors progressed with the number of iterations using the Hybridised Genetic Algorithm designed to Minimise Total Estimated Error Variance across runs three, four and five.

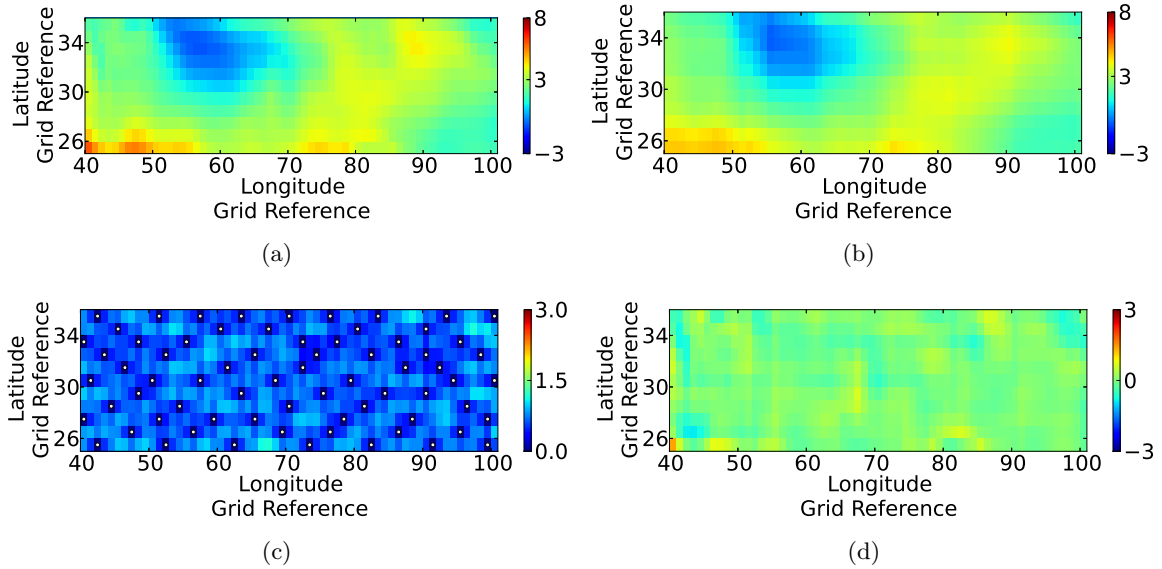


Figure D.48: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for run three generated using the Hybridised Genetic Algorithm designed to Minimise Total Estimated Error Variance. The white dots in (c) indicate the sampling points.

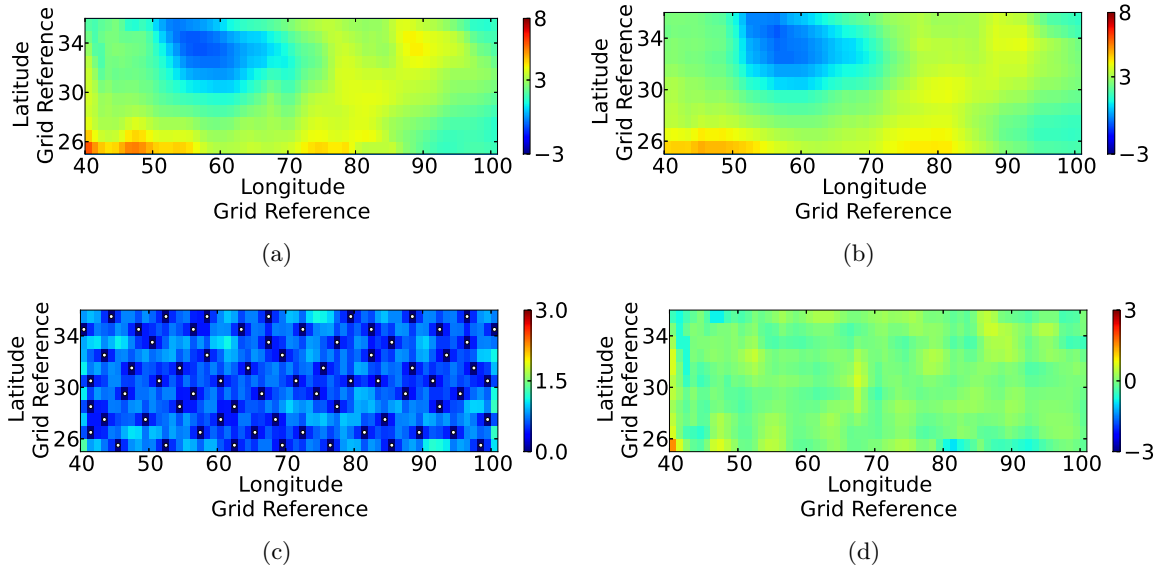


Figure D.49: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for run four generated using the Hybridised Genetic Algorithm designed to Minimise Total Estimated Error Variance. The white dots in (c) indicate the sampling points.

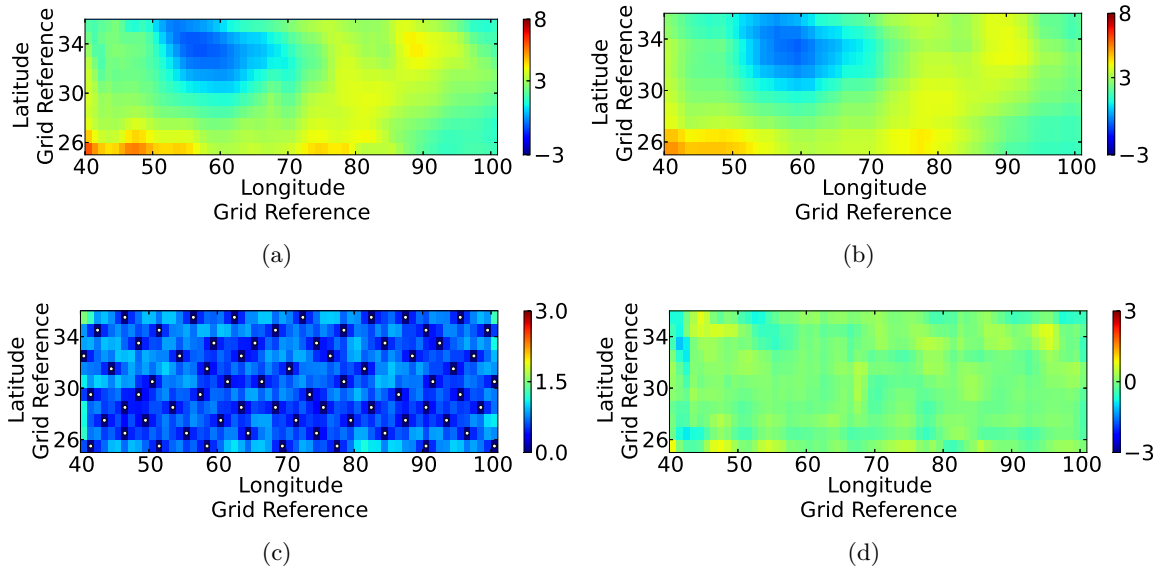


Figure D.50: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for run five generated using the Hybridised Genetic Algorithm designed to Minimise Total Estimated Error Variance. The white dots in (c) indicate the sampling points.

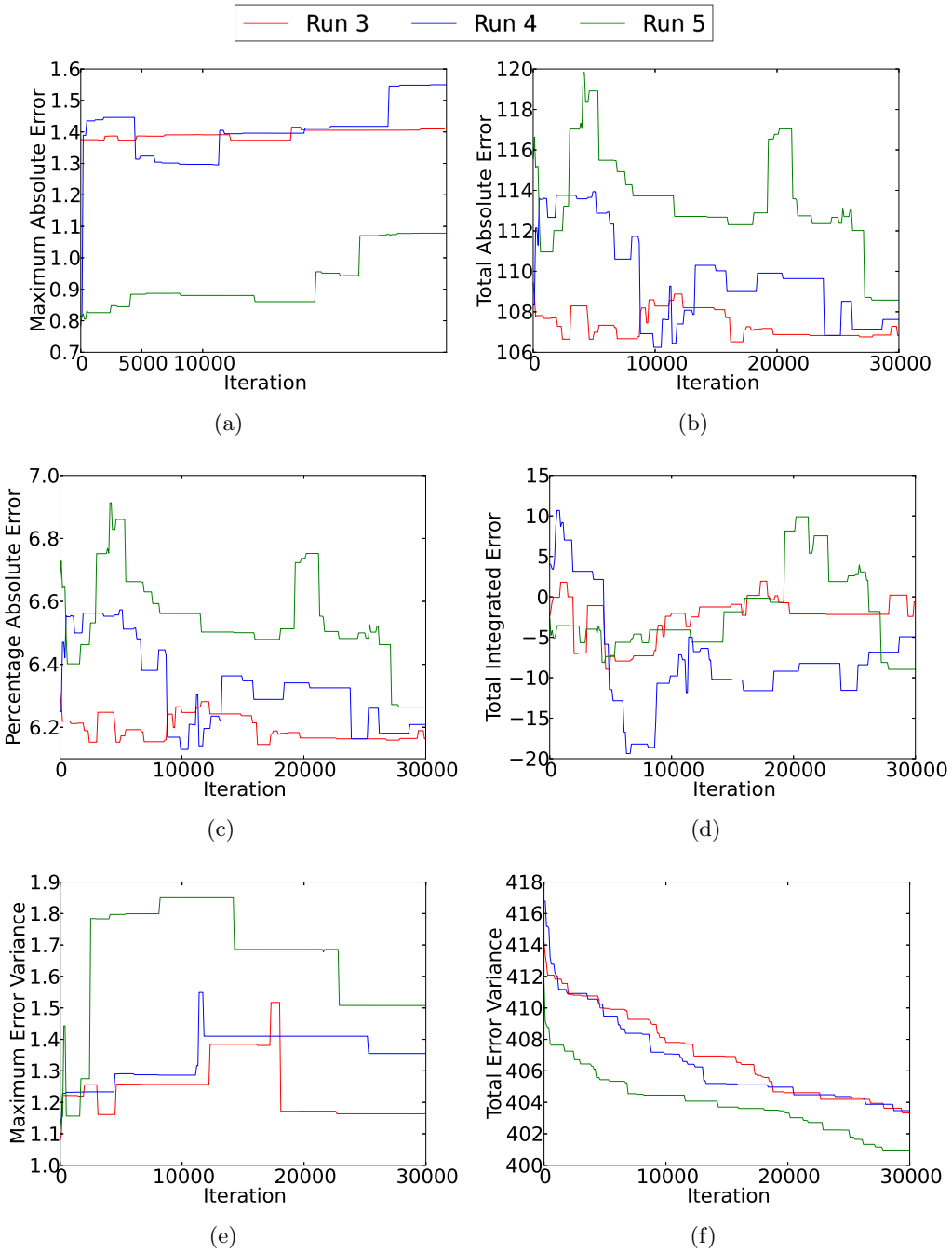


Figure D.51: Progression of (a) maximum absolute error, (b) total absolute error, (c) percentage absolute error, (d) total integrated error, (e) maximum estimated error variance and (e) total estimated error variance across all iterations for various runs of the Hybridised Genetic Algorithm designed to Minimise Total Estimated Error Variance.

D.9 Genetic Algorithm Designed to Maximise Total Absolute Error

Maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux, the estimated error variance and the true error across the chosen two-dimensional region for runs one, three and five can be seen in Figures D.52 to D.54. These maps are given for the full sample set of size 70 generated using the Genetic Algorithm designed to Maximise Total Absolute Error (See Chapter 13.12).

Figure D.55 shows how the errors progressed with the number of iterations using the Genetic Algorithm designed to Maximise Total Absolute Error across runs one, three and five.

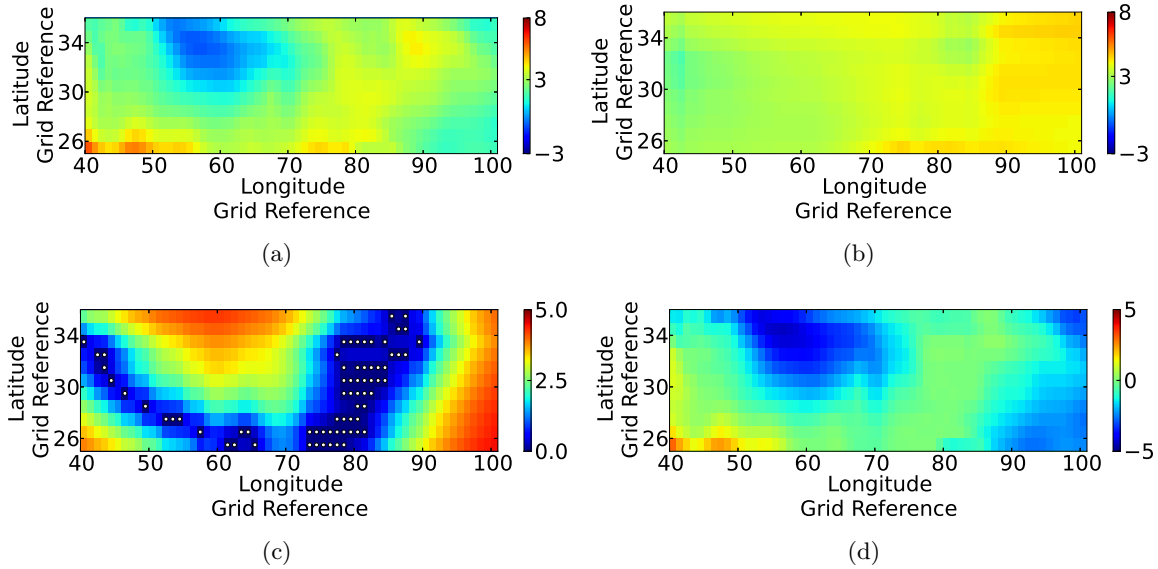


Figure D.52: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for run one generated using the Genetic Algorithm designed to Maximise Total Absolute Error. The white dots in (c) indicate the sampling points.

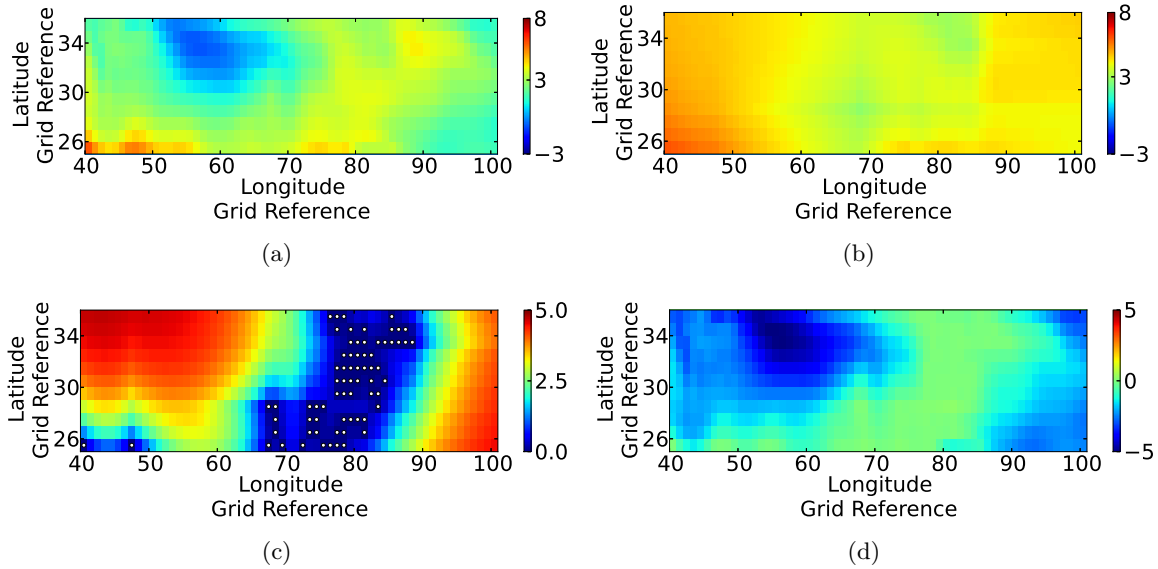


Figure D.53: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for run three generated using the Genetic Algorithm designed to Maximise Total Absolute Error. The white dots in (c) indicate the sampling points.

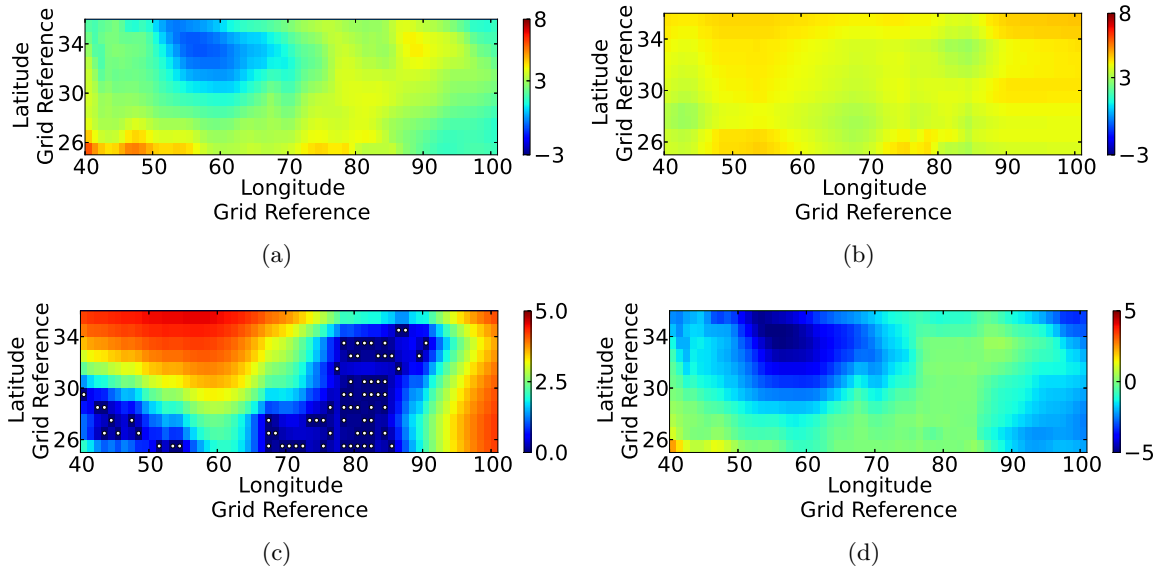


Figure D.54: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for run five generated using the Genetic Algorithm designed to Maximise Total Absolute Error. The white dots in (c) indicate the sampling points.

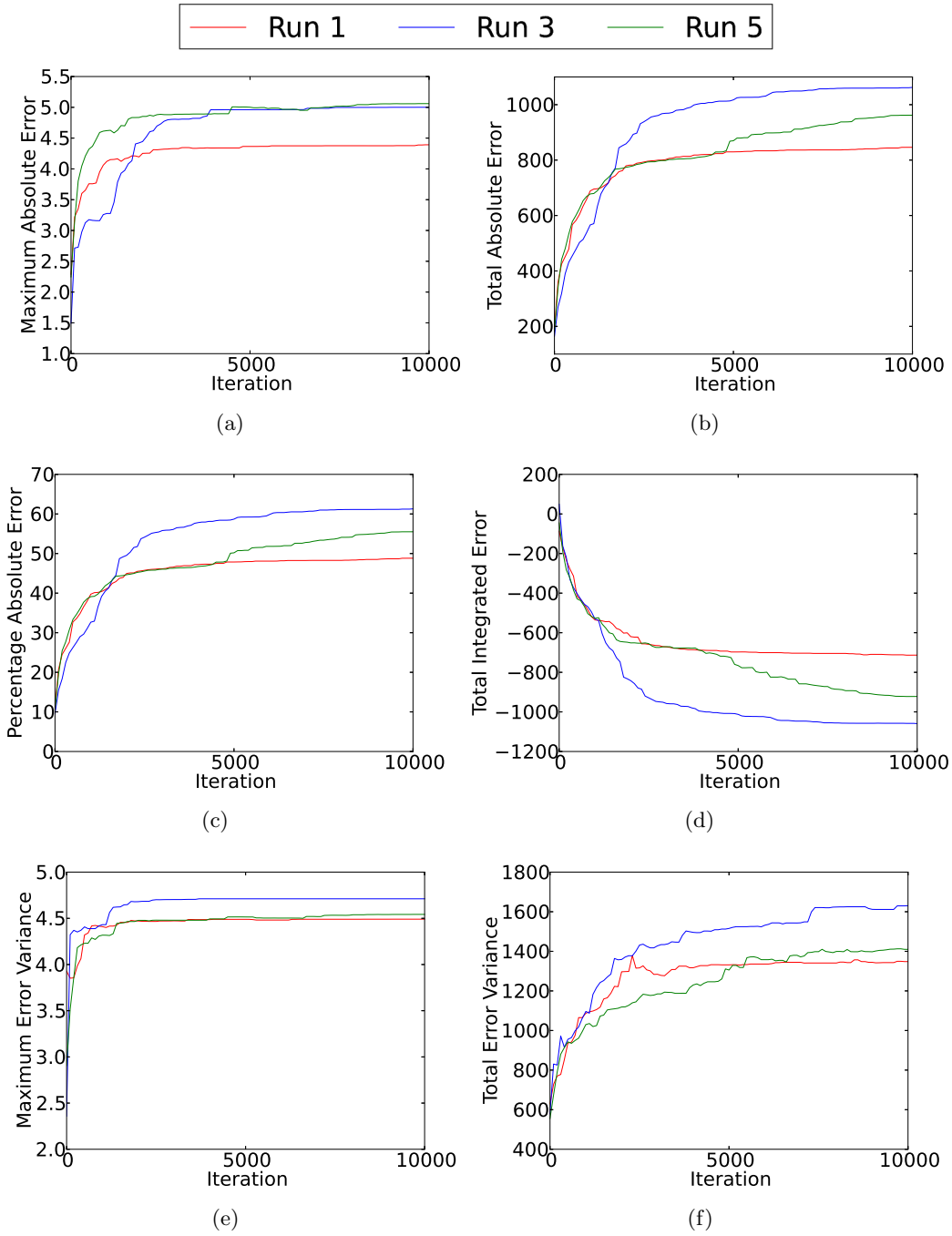


Figure D.55: Progression of (a) maximum absolute error, (b) total absolute error, (c) percentage absolute error, (d) total integrated error, (e) maximum estimated error variance and (e) total estimated error variance across all iterations for various runs of the Genetic Algorithm designed to Maximise Total Absolute Error.

D.10 Genetic Algorithm Designed to Maximise Total Estimated Error Variance

Maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux, the estimated error variance and the true error across the chosen two-dimensional region for runs two, four and five can be seen in Figures D.56 to D.58. These maps are given for the full sample set of size 70 generated using the Genetic Algorithm designed to Maximise Total Estimated Error Variance (See Chapter 13.13).

Figure D.59 shows how the errors progressed with the number of iterations using the Genetic Algorithm designed to Maximise Total Estimated Error Variance across runs two, four and five.

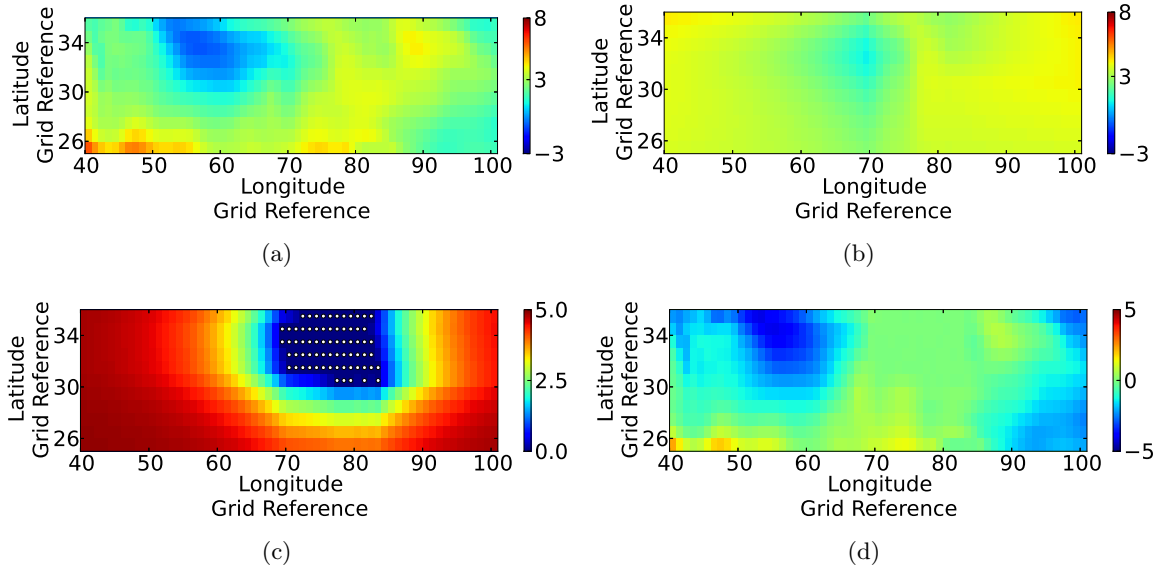


Figure D.56: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for run two generated using the Genetic Algorithm designed to Maximise Total Estimated Error Variance. The white dots in (c) indicate the sampling points.

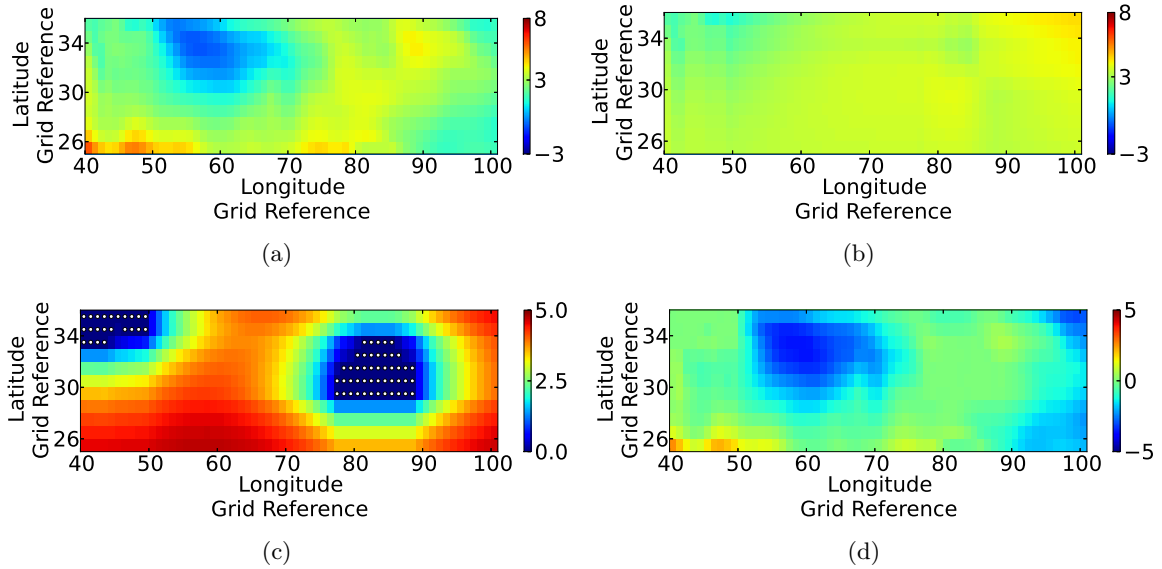


Figure D.57: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for run four generated using the Genetic Algorithm designed to Maximise Total Estimated Error Variance. The white dots in (c) indicate the sampling points.

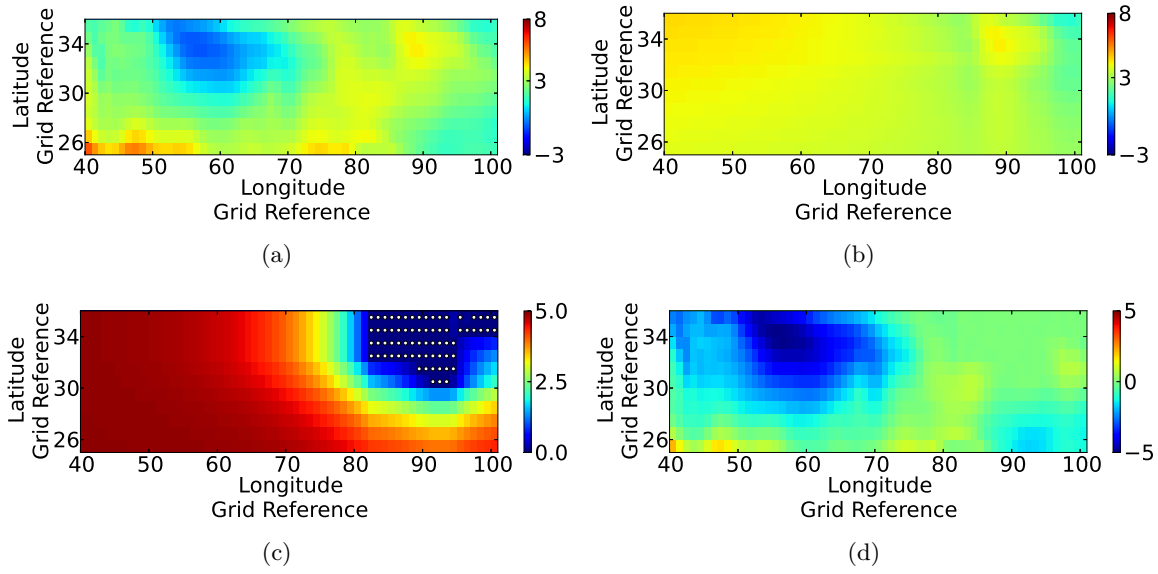


Figure D.58: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for run five generated using the Genetic Algorithm designed to Maximise Total Estimated Error Variance. The white dots in (c) indicate the sampling points.

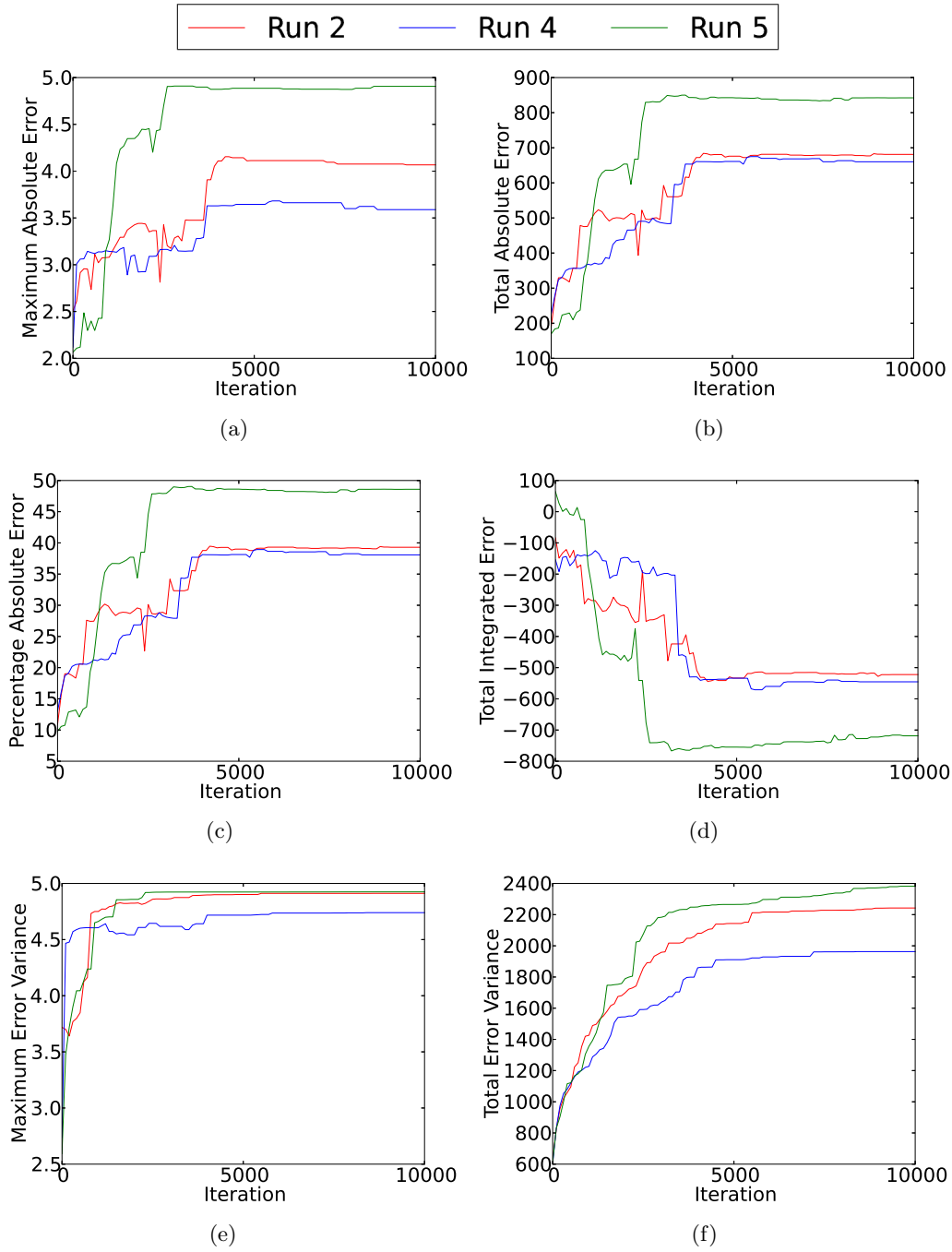


Figure D.59: Progression of (a) maximum absolute error, (b) total absolute error, (c) percentage absolute error, (d) total integrated error, (e) maximum estimated error variance and (e) total estimated error variance across all iterations for various runs of the Genetic Algorithm designed to Maximise Total Estimated Error Variance.

APPENDIX E

COMPARISON OF COMPUTATIONAL REQUIREMENTS

The computational requirements of each of the methods implemented can be found in Tables E.1 to E.10. These tables have been broken down into the description of the first iteration, the number of further iterations and the description of the remaining iterations. In each of these descriptions, Z is regarded to be the value of the residuals rather than that of the variable itself.

Method	Random Sampling
First Iteration $i = 1$	1. Random selection of 70 sample points with repetition.
Number of Remaining Iterations	0
Iteration $i > 1$	N/A

Table E.1: Summary of Random Sampling

Method	Updated Kriging Variance Algorithm
First Iteration $i = 1$	1. Removal of trend. 2. Random selection of 4 points without repetition 3. Generation of covariance matrix $\mathbf{R}$ equivalent to $\mathbf{H}$ described in Section 4.3. $\mathbf{R}$ is a 4×4 matrix. 4. Invert $\mathbf{R}$ to find $\mathbf{R}^{-1}$. This is not computationally intensive as it is a 4×4 matrix only. 5. For each point which is not yet in the sample ($671 - 4 = 667$ points): (a) Generate $\mathbf{\Lambda}_i$ which is the covariance vector between the sample points and the point selected at 5. It has length 4. (b) Generate $D = C(0) - \mathbf{\Lambda}_i \mathbf{R}^{-1} \mathbf{\Lambda}_i^T$. This involves a matrix-vector and then a dot-product multiplication. Vectors of length 4 and matrix of size 4×4. (c) For each point in the data set (671 points): i. Generate $\mathbf{V}_i$, the covariance between the point selected at 5c and the current sample set. Vector of length 4. ii. Calculate $\hat{r}_{i+1} = \mathbf{\Lambda}_i \mathbf{R}^{-1} \mathbf{V}_i$. This involves a matrix-vector and a dot-product multiplication. Vectors of length 4 and matrix size 4×4. iii. Generate r_{i+1} which is the covariance between the points selected at 5 and 5c. iv. Calculate $\frac{(r_{i+1} - \hat{r}_{i+1})^2}{D}$ and add it to the variable holding the sum of these values across all 671 points for each point selected at 5. This means at this iteration there will be 667 of these sums in total. 6. Find the maximum of the sums described in 5(c)iv. Select the point associated with this maximum and add to the sample set. Also, remove this point from the set of points which are not in the sample set.
Number of Remaining Iterations	65
Iteration $i > 1$	1. Update $\mathbf{R}^{-1}$ for new point using method described in Section 4.4.1. $\mathbf{T} = C(0)$, $\mathbf{Q}$ is the covariance vector between added point from previous iteration and previous sample set (needs to be generated) and has length $i + 3$ and $\mathbf{S} = \mathbf{Q}^T$. Due to same combinations of multiplications being used more than once, only 2 matrix-vector, 2 dot-product and 2 constant-vector multiplications need to take place. Additionally the inversion required to calculate $\mathbf{H}$ is reduced to the inversion of a constant as only one point has been added. Thus, merely a division, additions, subtractions, 2 matrix-vector, 2 dot-product and 2 constant-vector multiplications are needed to generate the new $\mathbf{R}^{-1}$. 2. Repeat steps 5 to 6 for the new $\mathbf{R}^{-1}$. The new vectors' lengths will be $(i + 3)$, the new matrix sizes will be $(i+3) \times (i+3)$ and there will be $671 - (i+3)$points at step 5 and the same number of sums in step 5(c)iv.

Table E.2: Summary of the Updated Kriging Variance Algorithm.

Method	Addition of Point at Point of Maximum Estimated Error Variance
First Iteration $i = 1$	1. Removal of trend. 2. Random selection of 4 points without repetition. 3. Generate covariance matrix $\mathbf{R}$, equivalent to $\mathbf{K}$ in Section 3.1.3. $\mathbf{R}$ has size of 4×4 4. Invert $\mathbf{R}$ to obtain $\mathbf{R}^{-1}$. This is not computationally expensive as $\mathbf{R}$ is only of size 4×4. 5. For each point in the data set (671 points): (a) Generate $\mathbf{k}$, the vector of covariances between the sample points and the point chosen at 5. $\mathbf{k}$ has length 4. (b) Calculate the estimated error variance at point chosen at 5 by $\sigma_E^2 = C(0) - \mathbf{k}^T\mathbf{R}^{-1}\mathbf{k}$. This involves a matrix-vector and a dot-product multiplication. The vectors are of length 4 and the matrix is of size 4×4. 6. Find the maximum estimated error variance of the 671 generated and add point associated with this maximum to the sample set.
Number of Remaining Iterations	65
Iteration $i > 1$	1. Update $\mathbf{R}^{-1}$ for new point using method described in Section 4.4.1. $\mathbf{T} = C(0)$, $\mathbf{Q}$ is the covariance vector between added point from previous iteration and previous sample set (needs to be generated) and has length $i + 3$ and $\mathbf{S} = \mathbf{Q}^T$. Due to same combinations of multiplications being used more than once, only 2 matrix-vector, 2 dot-product and 2 constant-vector multiplications need to take place. Additionally the inversion required to calculate H is reduced to the inversion of a constant as only one point has been added. Thus, merely a division, additions, subtractions, 2 matrix-vector, 2 dot-product and 2 constant-vector multiplications are needed to generate the new $\mathbf{R}^{-1}$. 2. Repeat steps 5 to 6 from the first iteration. The vectors' lengths will be $(i + 3)$ and the matrix sizes $(i + 3) \times (i + 3)$.

Table E.3: Summary of Addition of a Point at Point of Maximum Estimated Error Variance.

Method	Addition of Point at Point of Maximum Absolute Error
First Iteration $i = 1$	1. Removal of trend. 2. Random selection of 4 points without repetition. 3. Generate covariance matrix $\mathbf{R}$, equivalent to $\mathbf{K}$ in Section 3.1.3. $\mathbf{R}$ has size of 4×4 4. Invert $\mathbf{R}$ to obtain $\mathbf{R}^{-1}$. This is not computationally expensive as $\mathbf{R}$ is only of size 4×4. 5. Matrix multiply $\mathbf{R}^{-1}$ and the vector of flux values at the sample points ($\mathbf{Z}$). $\mathbf{W} = \mathbf{R}^{-1}\mathbf{Z}$. Only one matrix-vector multiplication is involved. The vector is of length 4 and the matrix of size 4×4. 6. For each point in the data set(671 points): (a) Generate $\mathbf{k}$, the vector of covariances between the sample points and the point chosen at 6. This vector is of length 4. (b) Calculate the predicted value at the point chosen at 6, $Z^*(\mathbf{s}) = \mathbf{k} \cdot \mathbf{W}$. This involves one dot-product multiplication. The vectors have length 4. (c) Take absolute value of difference between predicted and true flux value at point selected at 6. 7. Find the maximum absolute error of the 671 generated and add the point associated with this maximum to sample set.
Number of Remaining Iterations	65
Iteration $i > 1$	1. Update $\mathbf{R}^{-1}$ for new point using method described in Section 4.4.1. $\mathbf{T} = C(0)$, $\mathbf{Q}$ is the covariance vector between added point from previous iteration and previous sample set (needs to be generated) and has length $i + 3$ and $\mathbf{S} = \mathbf{Q}^T$. Due to same combinations of multiplications being used more than once, only 2 matrix-vector, 2 dot-product and 2 constant-vector multiplications need to take place. Additionally the inversion required to calculate H is reduced to the inversion of a constant as only one point has been added. Thus, merely a division, additions, subtractions, 2 matrix-vector, 2 dot-product and 2 constant-vector multiplications are needed to generate the new $\mathbf{R}^{-1}$. 2. Repeat steps 5 to 7 from the first iteration. The vectors' lengths will be $(i + 3)$ and the matrix sizes $(i + 3) \times (i + 3)$.

Table E.4: Summary of Addition of a Point at Point of Maximum Absolute Error.

Method	Genetic Algorithm designed to Minimise Total Absolute Error
First Iteration $i = 1$	1. Removal of trend. 2. 20 sets of 70 points each randomly selected without repetition from the 671 points in the data set. Each of the 20 sets forms an individual. 3. Evaluation of each set/individual (20 sets): (a) Generate the matrix of covariances, $\mathbf{R}$, equivalent to the matrix $\mathbf{K}$ in Section 3.1.3. $\mathbf{R}$ has size 70×70. (b) Solve the linear system of equations $\mathbf{RW} = \mathbf{Z}$ for $\mathbf{W}$, where $\mathbf{Z}$ is the vector of flux values at the sample points. $\mathbf{Z}$ has length 70 and $\mathbf{R}$ has size 70×70. (c) For each point in the data set (671 points): i. Generate $\mathbf{k}$, the vector of covariances between the sample points in the set selected at 3 and the point selected at 3c. $\mathbf{k}$ has length 70. ii. Calculate the predicted value at the point chosen at 3c, $Z^*(\mathbf{s}) = \mathbf{k} \cdot \mathbf{W}$. One dot-product multiplication is involved. The vectors have length 70. iii. Take absolute value of difference between the predicted and the true flux value at the point selected at 3c. iv. Sum the absolute errors across the 671 points for the set selected at 3. 4. Find the set with the lowest total absolute error and save it for use later in elitism. 5. In order to select sets to be added to the mating pool, tournament selection takes place. This tournament selection takes place as described in Section 13.2. 6. The sets in the mating pool undergo single point crossover as described in Section 13.3. 7. Each new set must be given the chance to undergo mutation as described in Section 13.4. 8. Replacement and elitism are implemented as per Sections 13.5 and 13.7 respectively. 9. Each new set must be checked for repetition of points. If repetition is found, the later of the instances of the point are replaced with a point which is randomly chosen from all possible points which are not in the set.
Number of Remaining Iterations	29999
Iteration $i > 1$	1. Repeat steps 3 to 9 from the first iteration.

Table E.5: Summary of Genetic Algorithm designed to Minimise Total Absolute Error.

Method	Genetic Algorithm designed to Minimise Total Estimated Error Variance
First Iteration $i = 1$	1. Removal of trend. 2. 20 sets of 70 points each randomly selected without repetition from the 671 points in the data set. Each of the 20 sets forms an individual. 3. Evaluation of each set/individual (20 sets): (a) Generate the matrix of covariances, $\mathbf{R}$, equivalent to the matrix $\mathbf{K}$ in Section 3.1.3. $\mathbf{R}$ has size 70×70. (b) For each point in the data set (671 points): i. Generate $\mathbf{k}$, the vector of covariances between the sample points in the set selected at 3 and the point selected at 3b. $\mathbf{k}$ has length 70. ii. Solve linear system of equations($\mathbf{R}\boldsymbol{\lambda} = \mathbf{k}$) for $\boldsymbol{\lambda}$. Perform dot-product multiplication $\mathbf{k} \cdot \boldsymbol{\lambda}$. Subtract this value from $C(0)$ to give the estimated error variance at the point. This involves one dot-product multiplication and the solving of one linear system of 70 equations. iii. Sum the estimated error variances across the 671 points for the set selected at 3. 4. Find the set with the lowest total estimated error variance and save it for use later in elitism. 5. In order to select sets to be added to the mating pool, tournament selection takes place. This tournament selection takes place as described in Section 13.2. 6. The sets in the mating pool undergo single point crossover as described in Section 13.3. 7. Each new set must be given the chance to undergo mutation as described in Section 13.4. 8. Replacement and elitism are implemented as per Sections 13.5 and 13.7 respectively. 9. Each new set must be checked for repetition of points. If repetition is found, the later of the instances of the point are replaced with a point which is randomly chosen from all possible points which are not in the set.
Number of Remaining Iterations	29999
Iteration $i > 1$	1. Repeat steps 3 to 9 from the first iteration.

Table E.6: Summary of Genetic Algorithm designed to Minimise Total Estimated Error Variance.

Method	Hybridised Genetic Algorithm designed to Minimise Total Absolute Error
First Iteration $i = 1$	1. Removal of trend. 2. 19 sets of 70 points each randomly selected without repetition from the 671 points in the data set. One additional set is selected as the solution from the method of Addition of a Point at Point of Maximum Absolute Error, adding all the computational requirements of that algorithm. Each of the 20 sets forms an individual. 3. Evaluation of each set/individual (20 sets): (a) Generate the matrix of covariances, $\mathbf{R}$, equivalent to the matrix $\mathbf{K}$ in Section 3.1.3. $\mathbf{R}$ has size 70×70. (b) Solve the linear system of equations $\mathbf{RW} = \mathbf{Z}$ for $\mathbf{W}$, where $\mathbf{Z}$ is the vector of flux values at the sample points. $\mathbf{Z}$ has length 70 and $\mathbf{R}$ has size 70×70. (c) For each point in the data set (671 points): i. Generate $\mathbf{k}$, the vector of covariances between the sample points in the set selected at 3 and the point selected at 3c. $\mathbf{k}$ has length 70. ii. Calculate the predicted value at the point chosen at 3c, $Z^*(\mathbf{s}) = \mathbf{k} \cdot \mathbf{W}$. One dot-product multiplication is involved. The vectors have length 70. iii. Take absolute value of difference between the predicted and the true flux value at the point selected at 3c. iv. Sum the absolute errors across the 671 points for the set selected at 3. 4. Find the set with the lowest total absolute error and save it for use later in elitism. 5. In order to select sets to be added to the mating pool, tournament selection takes place. This tournament selection takes place as described in Section 13.2. 6. The sets in the mating pool undergo single point crossover as described in Section 13.3. 7. Each new set must be given the chance to undergo mutation as described in Section 13.4. 8. Replacement and elitism are implemented as per Sections 13.5 and 13.7 respectively. 9. Each new set must be checked for repetition of points. If repetition is found, the later of the instances of the point are replaced with a point which is randomly chosen from all possible points which are not in the set.
Number of Remaining Iterations	29999
Iteration $i > 1$	1. Repeat steps 3 to 9 from the first iteration.

Table E.7: Summary of Hybridised Genetic Algorithm designed to Minimise Total Absolute Error.

Method	Hybridised Genetic Algorithm designed to Minimise Total Estimated Error Variance
First Iteration $i = 1$	1. Removal of trend. 2. 19 sets of 70 points each randomly selected without repetition from the 671 points in the data set. One additional set is selected as the solution from the method of Addition of a Point at Point of Maximum Estimated Error Variance, adding all the computational requirements of that algorithm. Each of the 20 sets forms an individual. 3. Evaluation of each set/individual (20 sets): (a) Generate the matrix of covariances, $\mathbf{R}$, equivalent to the matrix $\mathbf{K}$ in Section 3.1.3. $\mathbf{R}$ has size 70×70. (b) For each point in the data set (671 points): i. Generate $\mathbf{k}$, the vector of covariances between the sample points in the set selected at 3 and the point selected at 3b. $\mathbf{k}$ has length 70. ii. Solve linear system of equations($\mathbf{R}\boldsymbol{\lambda} = \mathbf{k}$) for $\boldsymbol{\lambda}$. Perform dot-product multiplication $\mathbf{k} \cdot \boldsymbol{\lambda}$. Subtract this value from $C(0)$ to give the estimated error variance at the point. This involves one dot-product multiplication and the solving of one linear system of 70 equations. iii. Sum the estimated error variances across the 671 points for the set selected at 3. 4. Find the set with the lowest total estimated error variance and save it for use later in elitism. 5. In order to select sets to be added to the mating pool, tournament selection takes place. This tournament selection takes place as described in Section 13.2. 6. The sets in the mating pool undergo single point crossover as described in Section 13.3. 7. Each new set must be given the chance to undergo mutation as described in Section 13.4. 8. Replacement and elitism are implemented as per Sections 13.5 and 13.7 respectively. 9. Each new set must be checked for repetition of points. If repetition is found, the later of the instances of the point are replaced with a point which is randomly chosen from all possible points which are not in the set.
Number of Remaining Iterations	29999
Iteration $i > 1$	1. Repeat steps 3 to 9 from the first iteration.

Table E.8: Summary of Hybridised Genetic Algorithm designed to Minimise Total Estimated Error Variance.

Method	Genetic Algorithm designed to Maximise Total Absolute Error
First Iteration $i = 1$	1. Removal of trend. 2. 20 sets of 70 points each randomly selected without repetition from the 671 points in the data set. Each of the 20 sets forms an individual. 3. Evaluation of each set/individual (20 sets): (a) Generate the matrix of covariances, $\mathbf{R}$, equivalent to the matrix $\mathbf{K}$ in Section 3.1.3. $\mathbf{R}$ has size 70×70. (b) Solve the linear system of equations $\mathbf{RW} = \mathbf{Z}$ where $\mathbf{Z}$ is the vector of flux values at the sample points. $\mathbf{Z}$ has length 70 and $\mathbf{R}$ has size 70×70. (c) For each point in the data set (671 points): i. Generate $\mathbf{k}$, the vector of covariances between the sample points in the set selected at 3 and the point selected at 3c. $\mathbf{k}$ has length 70. ii. Calculate the predicted value at the point chosen at 3c, $Z^*(\mathbf{s}) = \mathbf{k} \cdot \mathbf{W}$. One dot-product multiplication is involved. The vectors have length 70. iii. Take absolute value of difference between the predicted and the true flux value at the point selected at 3c. iv. Sum the absolute errors across the 671 points for the set selected at 3. 4. Find the set with the highest total absolute error and save it for use later in elitism. 5. In order to select sets to be added to the mating pool, tournament selection takes place. This tournament selection takes place as described in Section 13.2. 6. The sets in the mating pool undergo single point crossover as described in Section 13.3. 7. Each new set must be given the chance to undergo mutation as described in Section 13.4. 8. Replacement and elitism are implemented as per Sections 13.5 and 13.7 respectively. 9. Each new set must be checked for repetition of points. If repetition is found, the later of the instances of the point are replaced with a point which is randomly chosen from all possible points which are not in the set.
Number of Remaining Iterations	9999
Iteration $i > 1$	1. Repeat steps 3 to 9 from the first iteration.

Table E.9: Summary of Genetic Algorithm designed to Maximise Total Absolute Error.

Method	Genetic Algorithm designed to Maximise Total Estimated Error Variance
First Iteration $i = 1$	1. Removal of trend. 2. 20 sets of 70 points each randomly selected without repetition from the 671 points in the data set. Each of the 20 sets forms an individual. 3. Evaluation of each set/individual (20 sets): (a) Generate the matrix of covariances, $\mathbf{R}$, equivalent to the matrix $\mathbf{K}$ in Section 3.1.3. $\mathbf{R}$ has size 70×70. (b) For each point in the data set (671 points): i. Generate $\mathbf{k}$, the vector of covariances between the sample points in the set selected at 3 and the point selected at 3b. $\mathbf{k}$ has length 70. ii. Solve linear system of equations($\mathbf{R}\boldsymbol{\lambda} = \mathbf{k}$) for $\boldsymbol{\lambda}$. Perform dot-product multiplication $\mathbf{k} \cdot \boldsymbol{\lambda}$. Subtract this value from $C(0)$ to give the estimated error variance at the point. This involves one dot-product multiplication and the solving of one linear system of 70 equations. iii. Sum the estimated error variances across the 671 points for the set selected at 3. 4. Find the set with the highest total estimated error variance and save it for use later in elitism. 5. In order to select sets to be added to the mating pool, tournament selection takes place. This tournament selection takes place as described in Section 13.2. 6. The sets in the mating pool undergo single point crossover as described in Section 13.3. 7. Each new set must be given the chance to undergo mutation as described in Section 13.4. 8. Replacement and elitism are implemented as per Sections 13.5 and 13.7 respectively. 9. Each new set must be checked for repetition of points. If repetition is found, the later of the instances of the point is replaced with a point which is randomly chosen from all possible points which are not in the set.
Number of Remaining Iterations	9999
Iteration $i > 1$	1. Repeat steps 3 to 9 from the first iteration.

Table E.10: Summary of Genetic Algorithm designed to Maximise Total Estimated Error Variance.

APPENDIX F

THREE-DIMENSIONAL SAMPLING

Any additional figures from Chapter 15 relating to the three-dimensional sampling can be found in this appendix. The maps of the true carbon dioxide flux as well as maps of the predicted carbon dioxide flux, the estimated error variance and the true error across the full latitude and longitude grid reference ranges for time slices 0, 20, 40, 60 and 72 for the three different methods for set one are presented in Figures F.1 to F.15. While the maps for time slices 20 and 60 for set two across the three different methods can be seen in Figures F.16 to F.21. Only selected time slices have been included as the total number of maps produced is too large to include all of them.

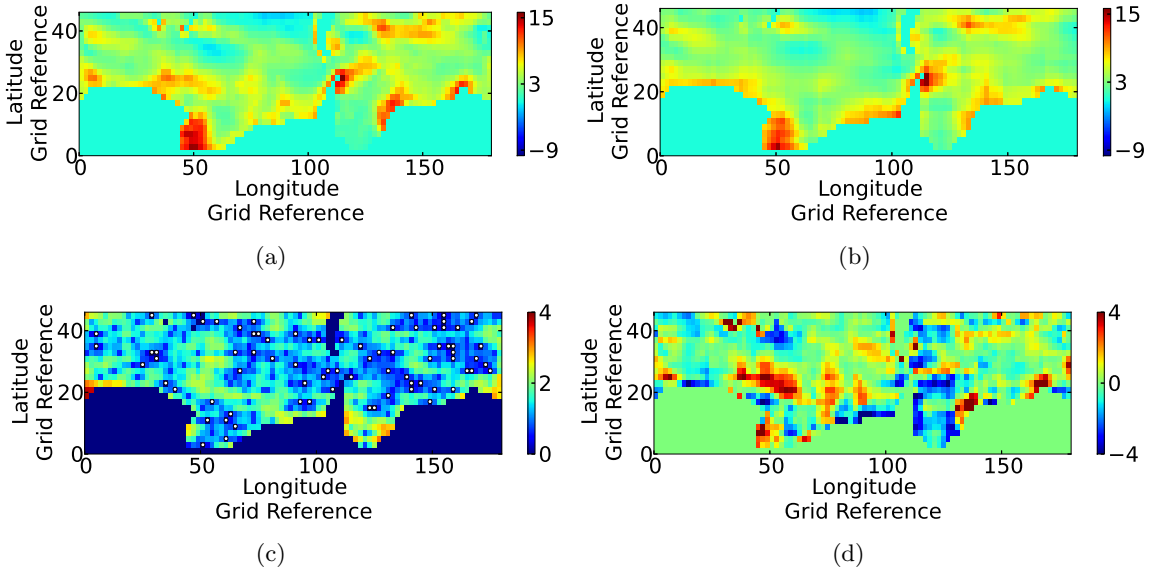


Figure F.1: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 0 of set one generated using Random Sampling. The white dots in (c) indicate the sampling points.

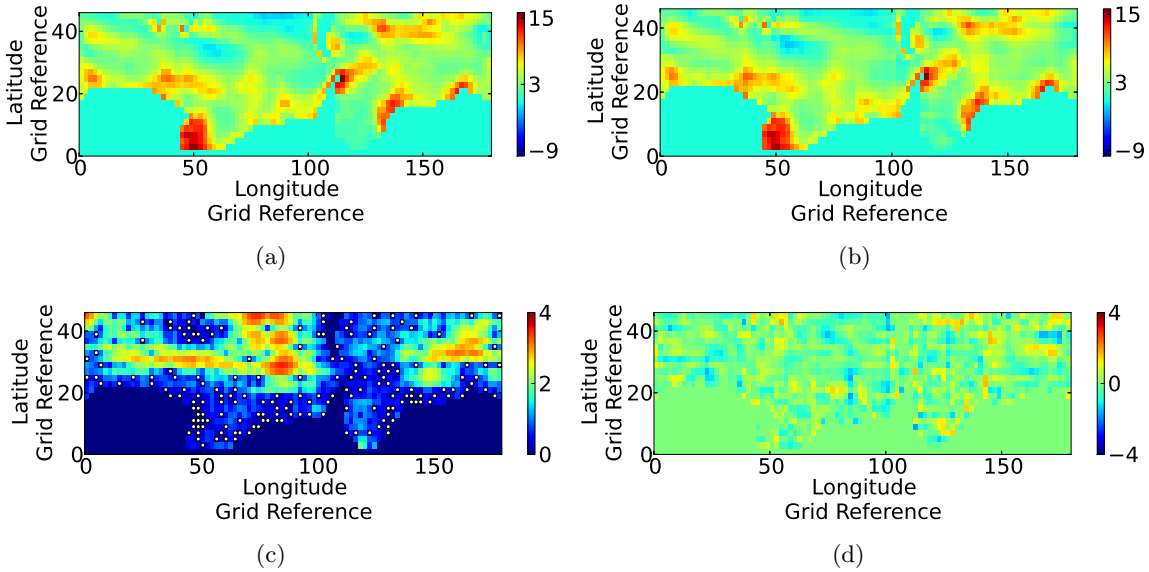


Figure F.2: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 0 of set one generated using Addition of a Point at Point of Maximum Absolute Error sampling. The white dots in (c) indicate the sampling points.

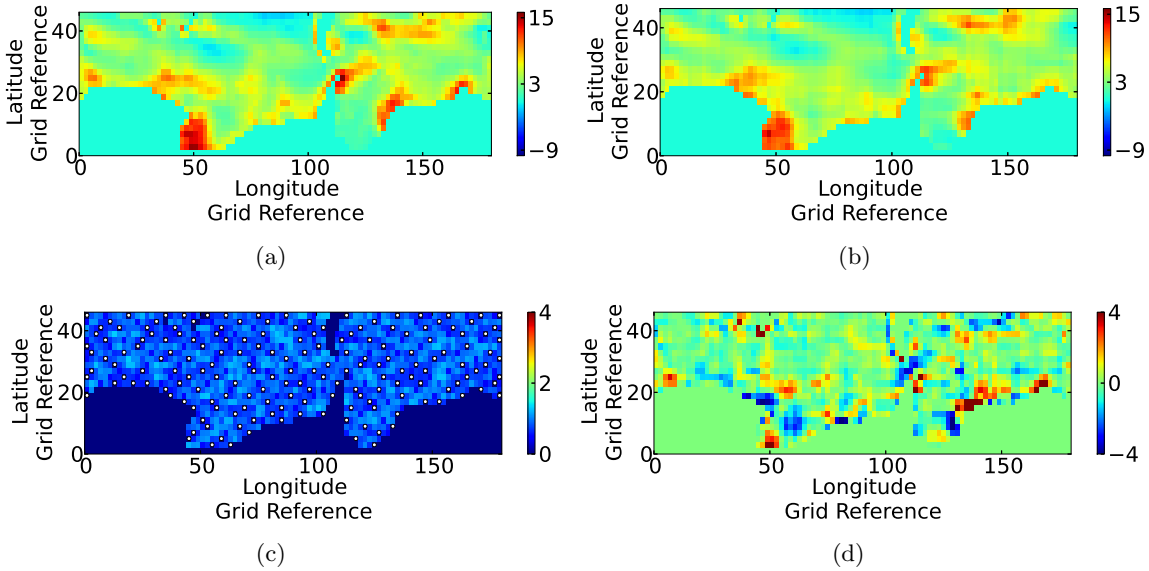


Figure F.3: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 0 of set one generated using Addition of a Point at Point of Maximum Estimated Error Variance sampling. The white dots in (c) indicate the sampling points.

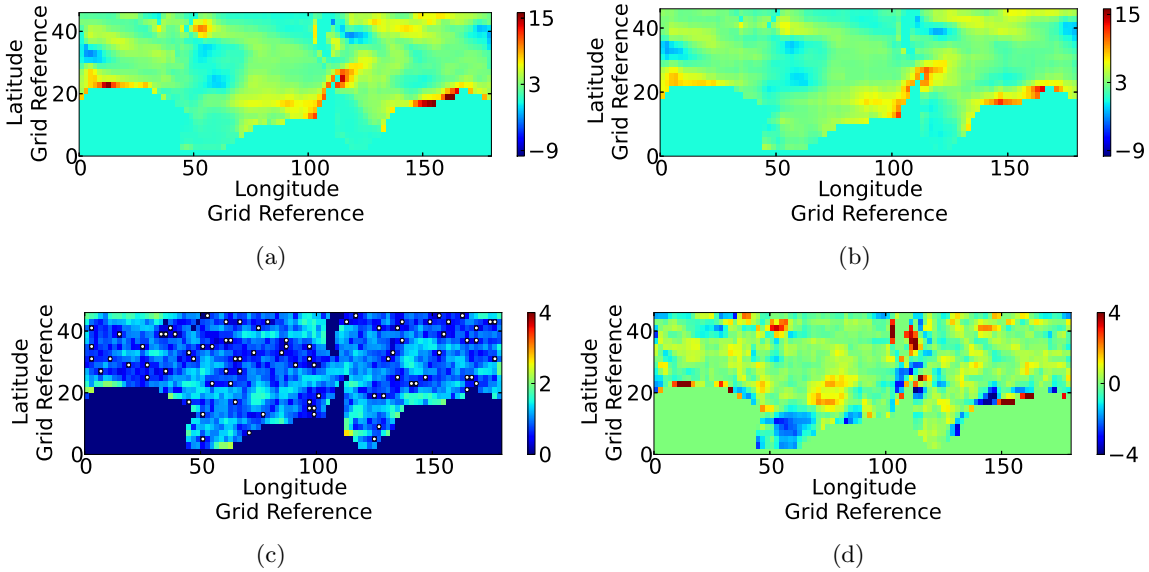


Figure F.4: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 20 of set one generated using Random Sampling. The white dots in (c) indicate the sampling points.

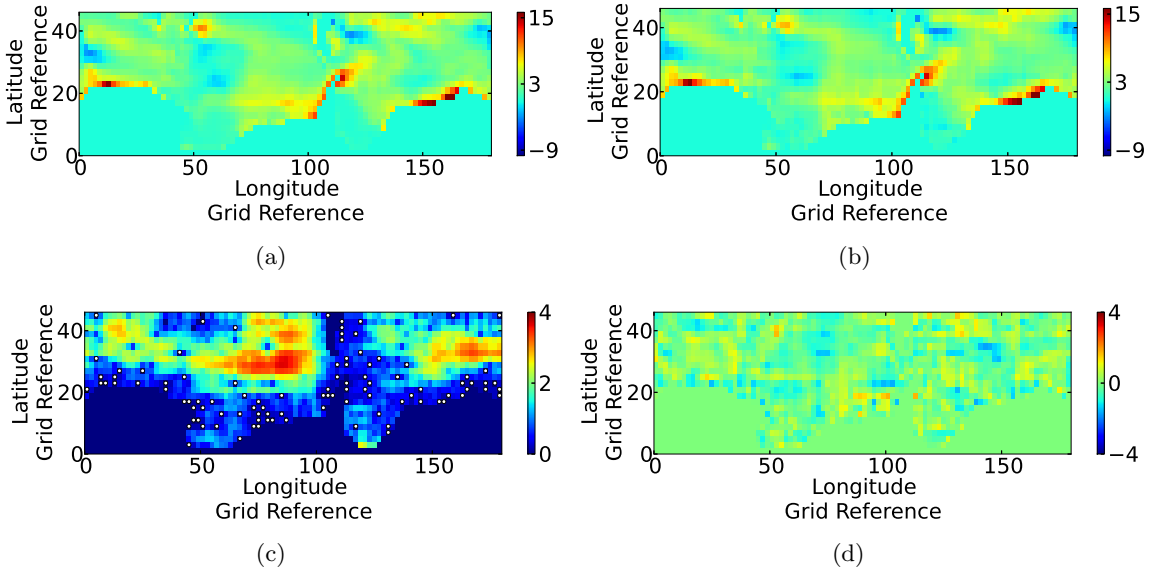


Figure F.5: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 20 of set one generated using Addition of a Point at Point of Maximum Absolute Error sampling. The white dots in (c) indicate the sampling points.

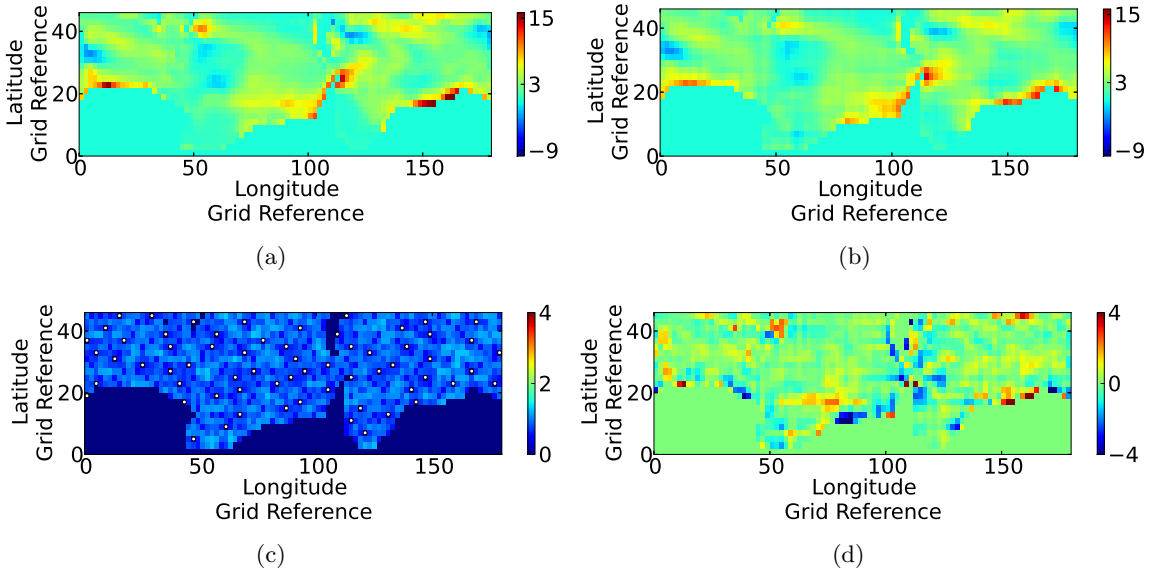


Figure F.6: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 20 of set one generated using Addition of a Point at Point of Maximum Estimated Error Variance sampling. The white dots in (c) indicate the sampling points.

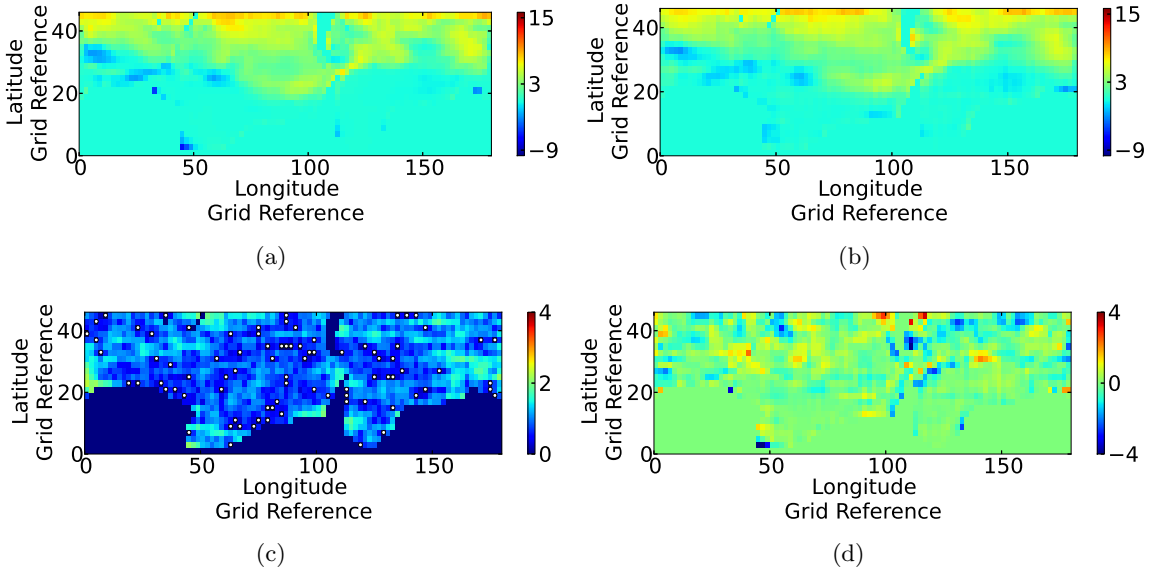


Figure F.7: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 40 of set one generated using Random Sampling. The white dots in (c) indicate the sampling points.

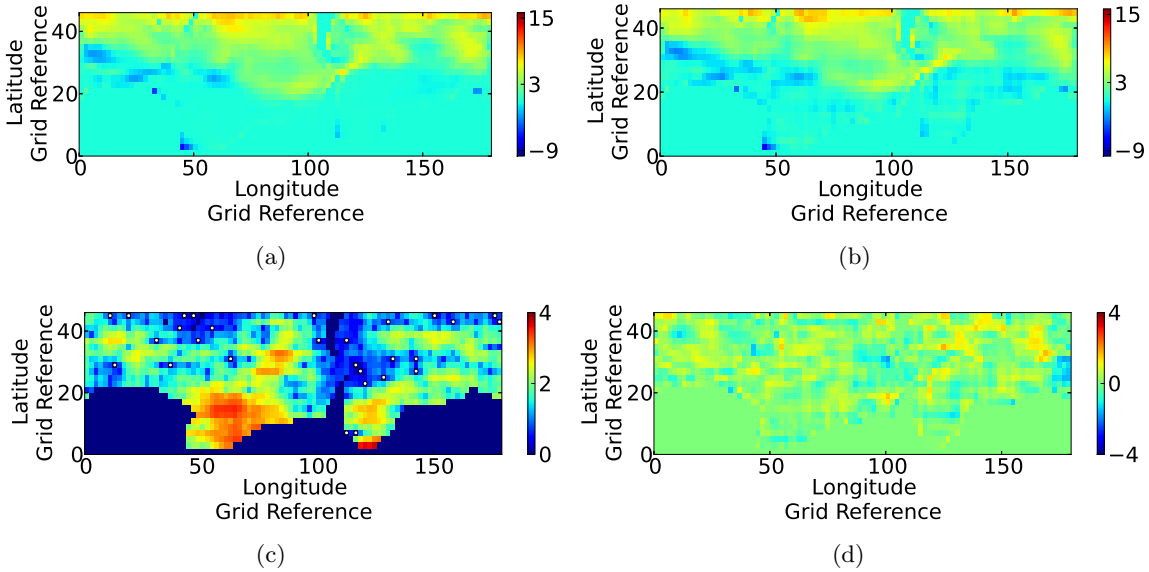


Figure F.8: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 40 of set one generated using Addition of a Point at Point of Maximum Absolute Error sampling. The white dots in (c) indicate the sampling points.

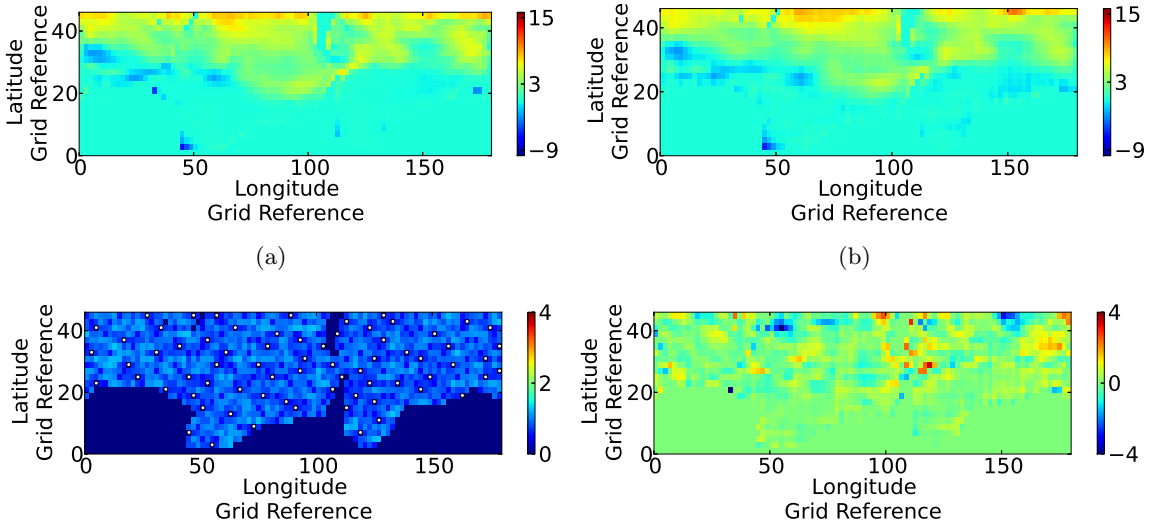


Figure F.9: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 40 of set one generated using Addition of a Point at Point of Maximum Estimated Error Variance sampling. The white dots in (c) indicate the sampling points.

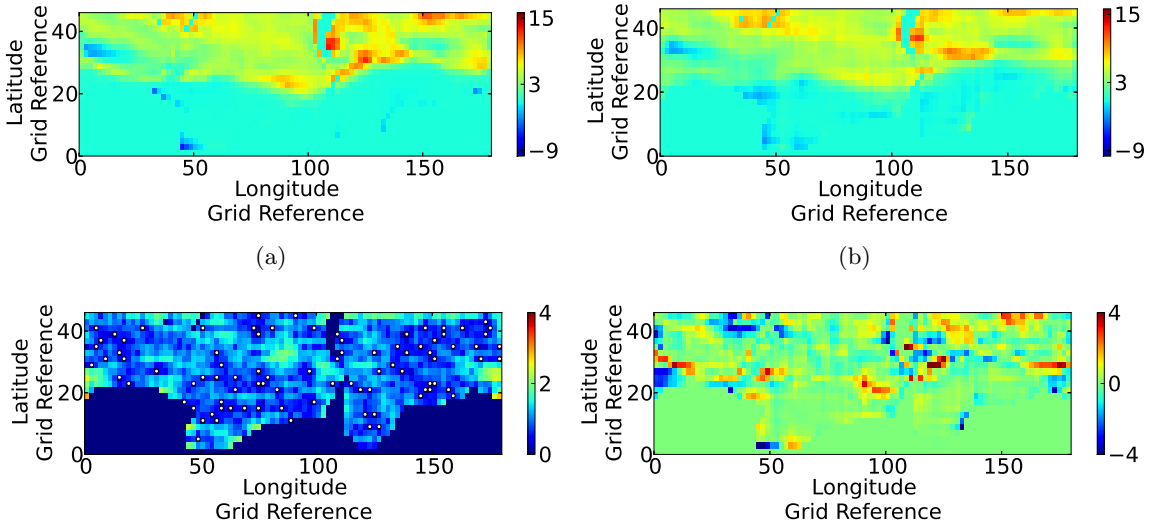


Figure F.10: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 60 of set one generated using Random Sampling. The white dots in (c) indicate the sampling points.

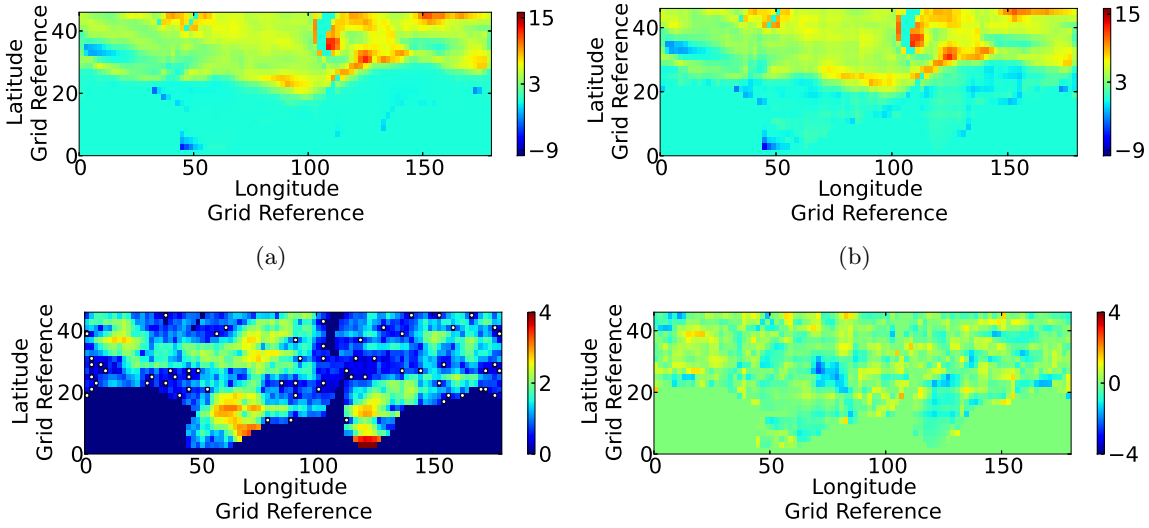


Figure F.11: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 60 of set one generated using Addition of a Point at Point of Maximum Absolute Error sampling

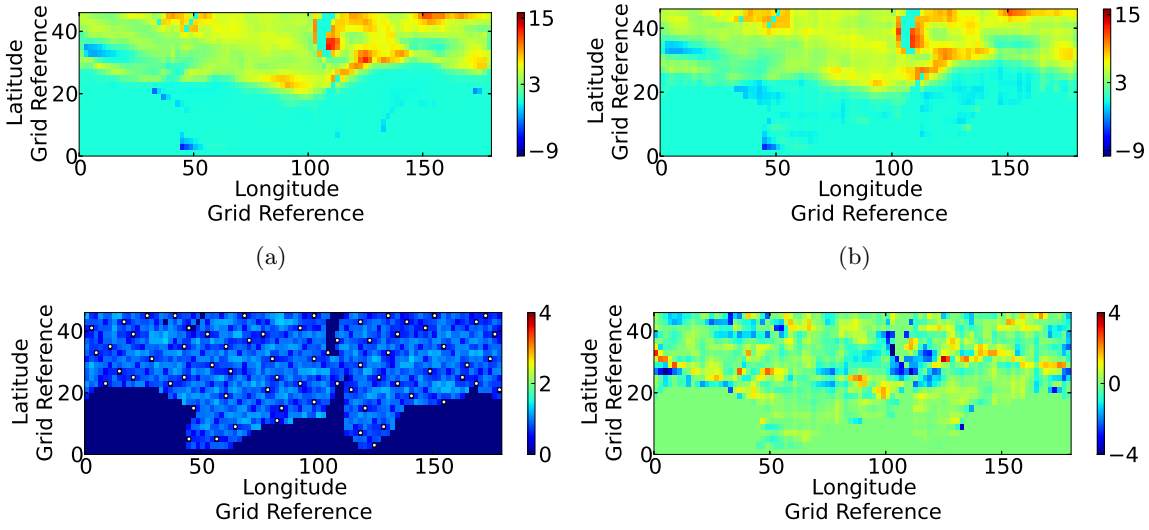


Figure F.12: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 60 of set one generated using Addition of a Point at Point of Maximum Estimated Error Variance sampling. The white dots in (c) indicate the sampling points.

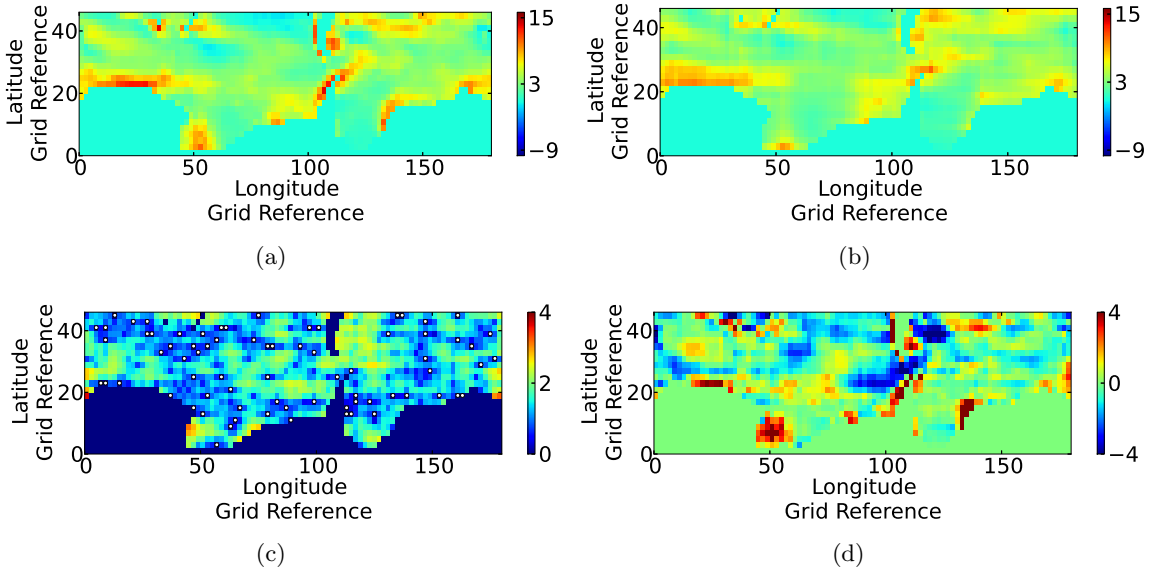


Figure F.13: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 72 of set one generated using Random Sampling. The white dots in (c) indicate the sampling points.

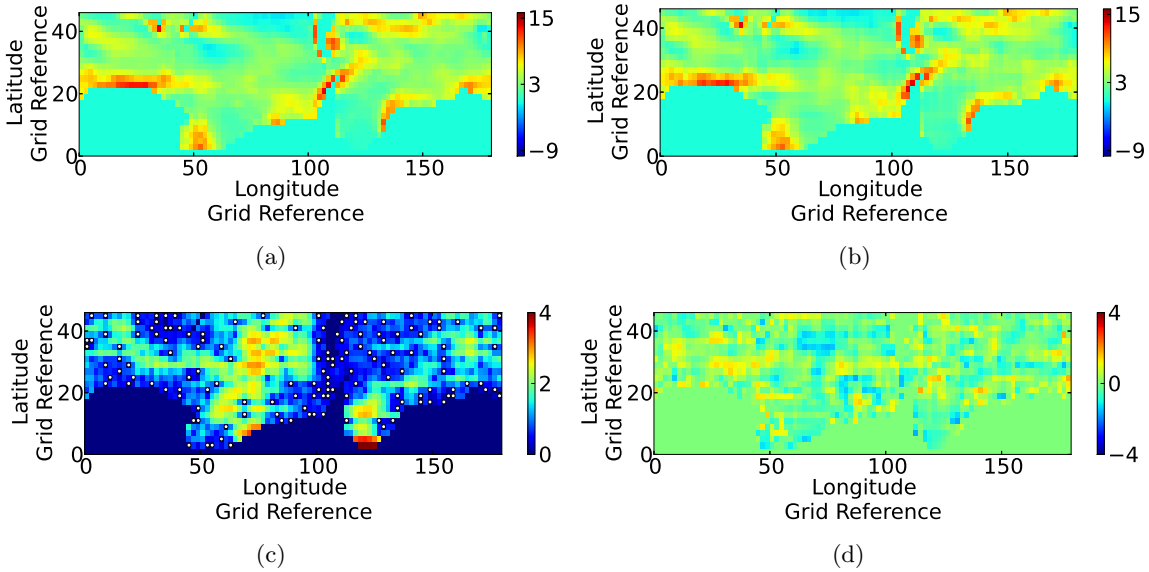


Figure F.14: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 72 of set one generated using Addition of a Point at Point of Maximum Absolute Error sampling. The white dots in (c) indicate the sampling points.

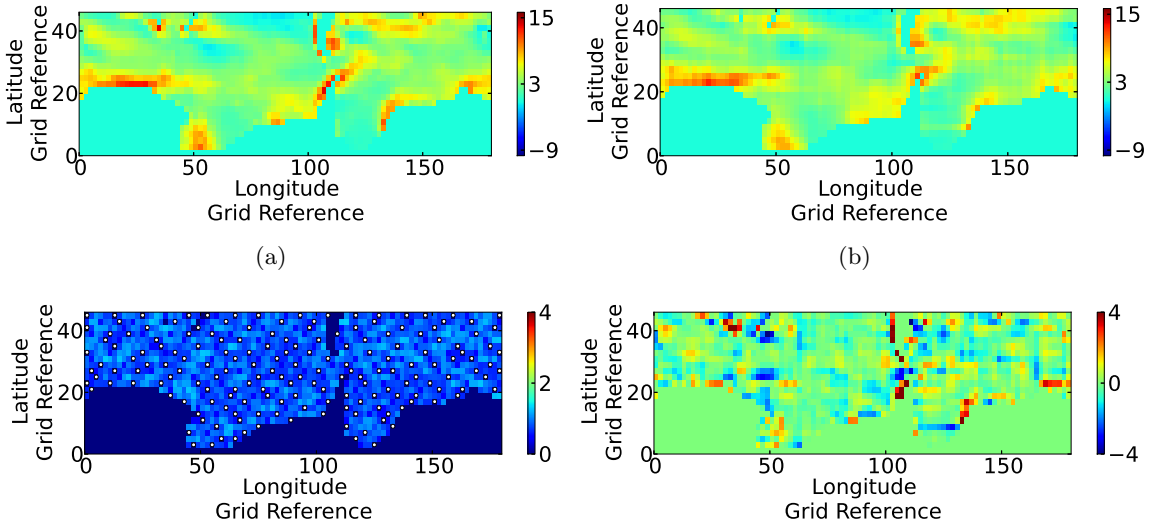


Figure F.15: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 72 of set one generated using Addition of a Point at Point of Maximum Estimated Error Variance sampling. The white dots in (c) indicate the sampling points.

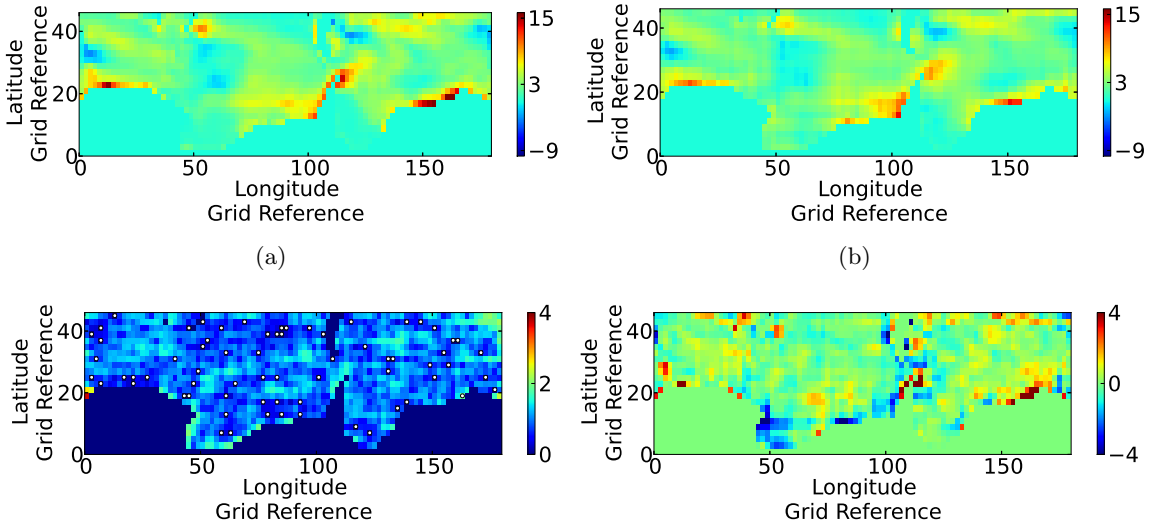


Figure F.16: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 20 of set two generated using Random Sampling. The white dots in (c) indicate the sampling points.

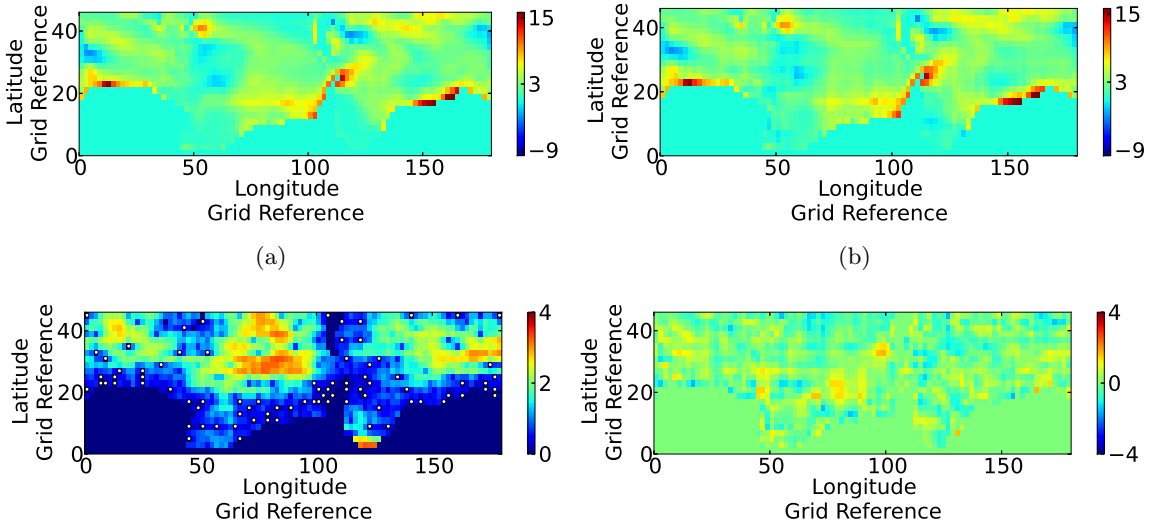


Figure F.17: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 20 of set two generated using Addition of a Point at Point of Maximum Absolute Error sampling. The white dots in (c) indicate the sampling points.

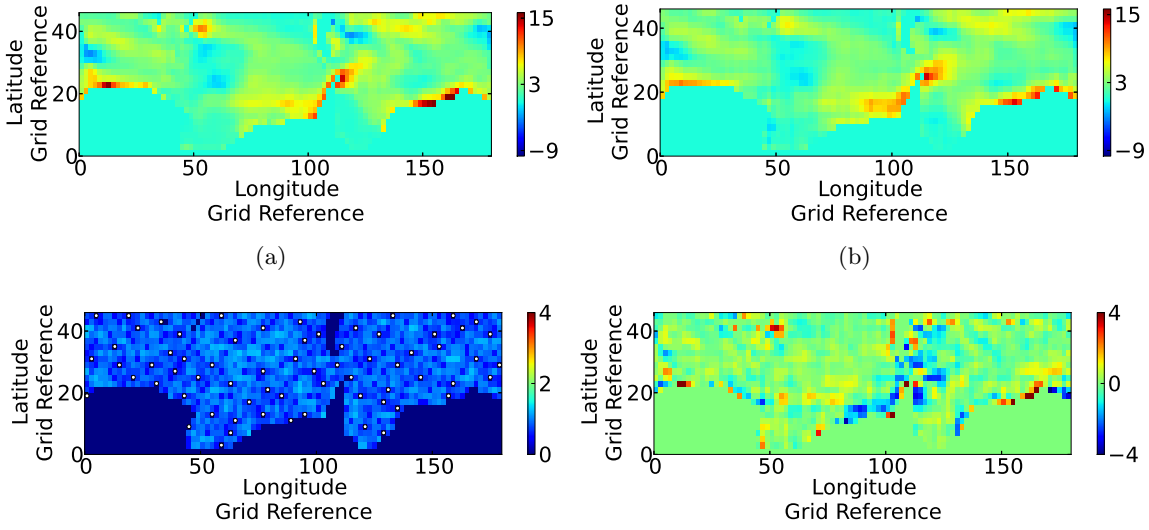


Figure F.18: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 20 of set two generated using Addition of a Point at Point of Maximum Estimated Error Variance sampling. The white dots in (c) indicate the sampling points.

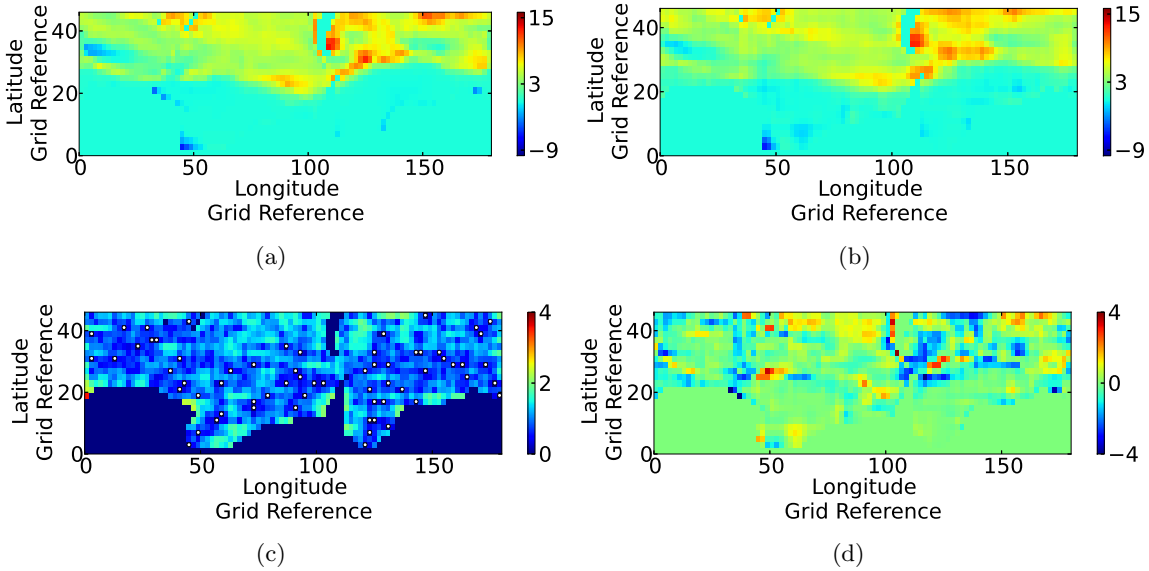


Figure F.19: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 60 of set two generated using Random Sampling

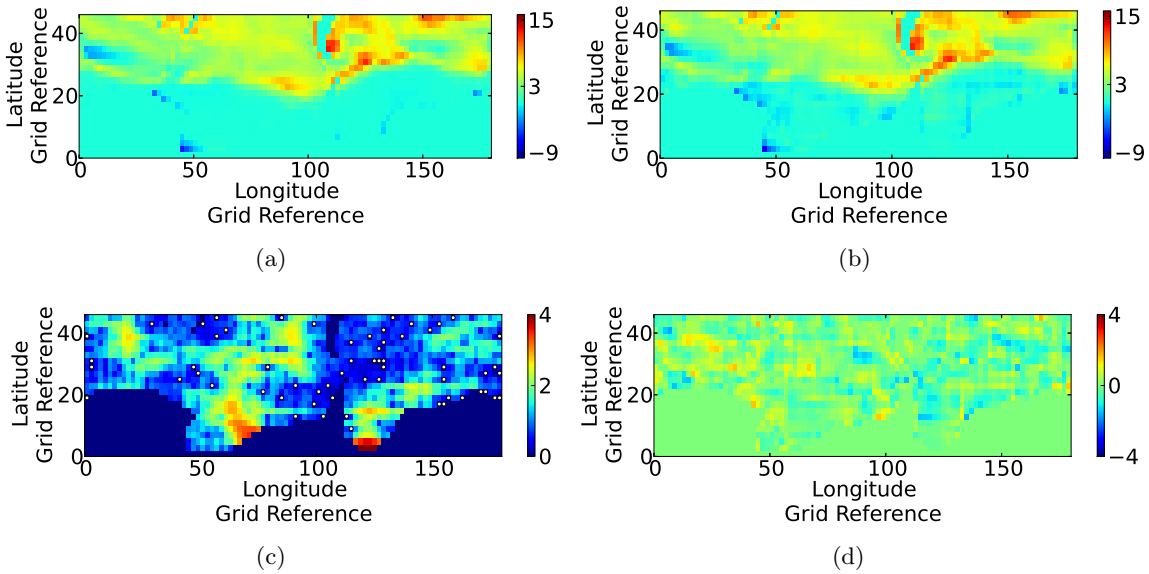


Figure F.20: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 60 of set two generated using Addition of a Point at Point of Maximum Absolute Error sampling. The white dots in (c) indicate the sampling points.

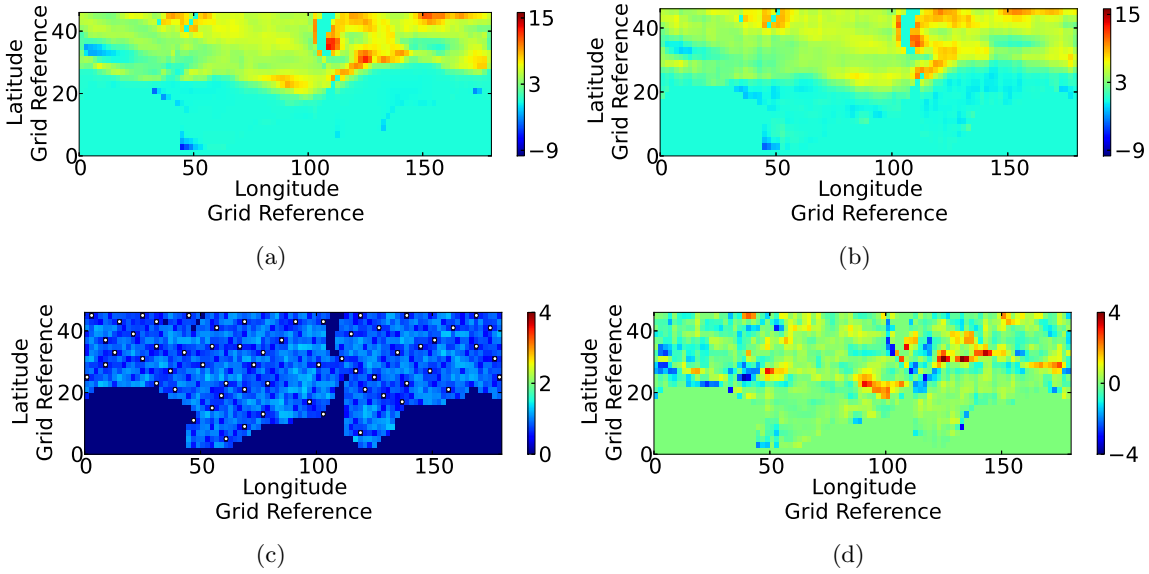


Figure F.21: Maps of (a) model data, (b) predicted data, (c) estimated error variance and (d) true error for time slice 60 of set two generated using Addition of a Point at Point of Maximum Estimated Error Variance sampling. The white dots in (c) indicate the sampling points.