GEOSCIENCE AUSTRALIA

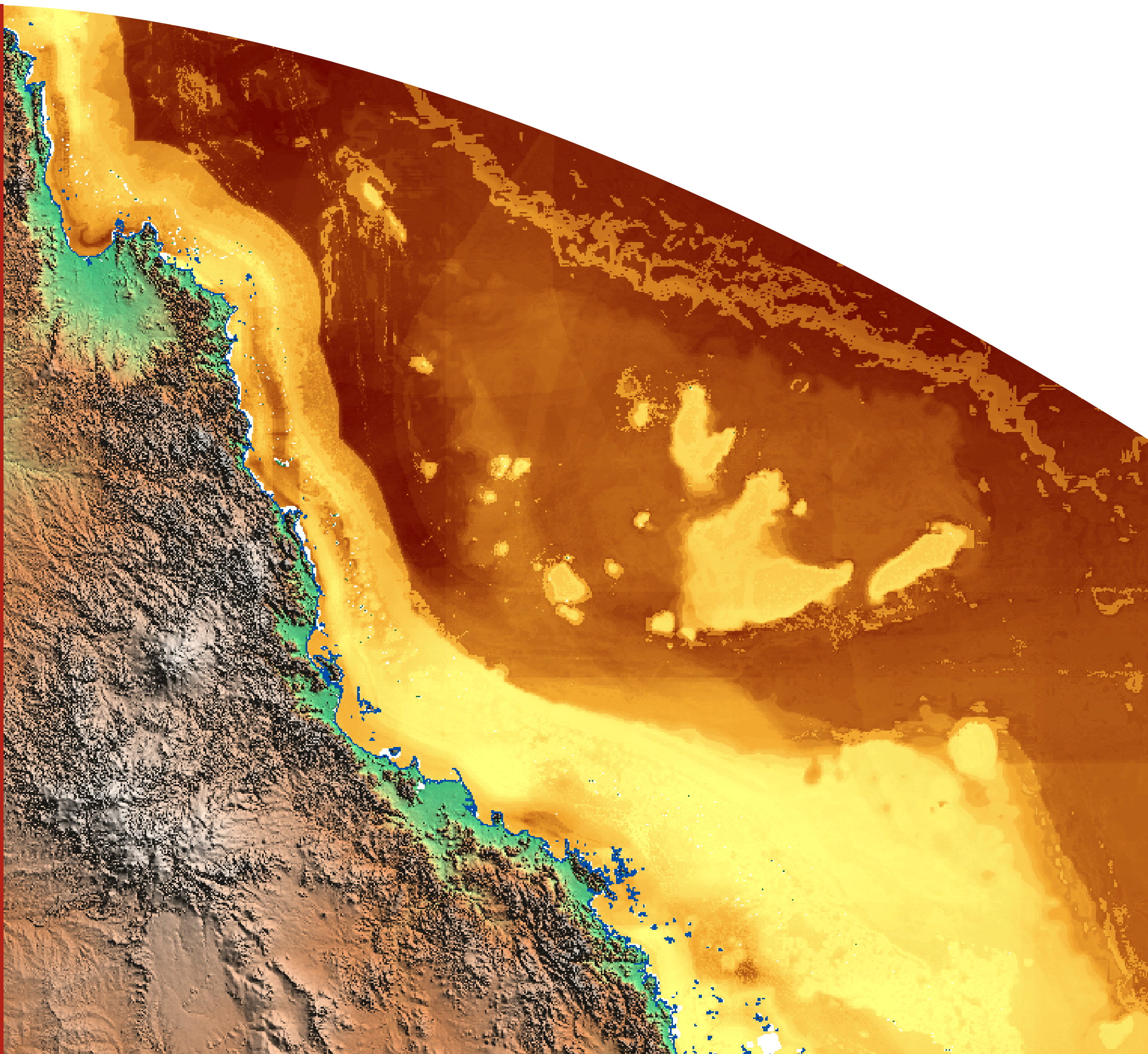
Australian Government
Geoscience Australia

Predicting Seabed Mud Content across the Australian Margin

Comparison of Statistical and Mathematical Techniques using a Simulation Experiment

Jin Li, Anna Potter, Zhi Huang, James J. Daniell, and Andrew D. Heap

Record
2010/11
GeoCat #
70150



APPLYING GEOSCIENCE TO AUSTRALIA'S MOST IMPORTANT CHALLENGES

Geoscience Australia Record 2010/11

Predicting Seabed Mud Content across the Australian Margin: Comparison of Statistical and Mathematical Techniques Using a Simulation Experiment

Jin Li, Anna Potter, Zhi Huang, James J. Daniell, and Andrew D. Heap

Geoscience Australia, GPO Box 378, Canberra, ACT 2601, Australia

Australian Government
Geoscience Australia

Department of Resources, Energy and Tourism
Minister for Resources and Energy: The Hon. Martin Ferguson, AM MP
Secretary: Mr Drew Clarke

Geoscience Australia
Acting Chief Executive Officer: Dr Chris Pigram

IISBN (9781921672781)
ISBN (9781921672774CD-ROM)
ISBN (9781921672767)

GeoCat # 70150

Bibliographic reference: Li, J., Potter, A., Huang, Z., Daniell, J. J. and Heap, A. D., 2010. *Predicting Seabed Mud Content across the Australian Margin: Comparison of Statistical and Mathematical Techniques Using a Simulation Experiment.* Geoscience Australia, Record 2010/11, 146 pp.

Correspondence for feedback:

Sales Centre
Geoscience Australia
GPO Box 378
Canberra
ACT 2601
Australia

sales@ga.gov.au

Executive Summary

Geoscience Australia is supporting the exploration and development of offshore oil and gas resources and establishment of Australia's national representative system of marine protected areas through provision of spatial information about the physical and biological character of the seabed. Central to this approach is using spatially continuous data of physical seabed properties to predict Australia's seabed biodiversity. However, information for these properties is usually collected at sparsely-distributed discrete locations, particularly in the deep ocean. Thus, methods for generating spatially continuous information from point samples become essential tools. Such methods are, however, often data- or even variable- specific and it is difficult to select an appropriate method for any given dataset. Traditionally, simple methods like inverse distance squared (IDS) have been used but its predictions are often associated with large errors.

In this study, we conduct a simulation experiment to identify robust spatial interpolation methods using samples of seabed mud content in Geoscience Australia's Marine Samples database. Due to data noise associated with the samples, criteria are developed and applied for data quality control. Five factors that affect the accuracy of spatial interpolation are considered: 1) regions; 2) statistical methods; 3) sample densities; 4) searching neighbourhoods; and 5) sample stratification. Bathymetry, distance-to-coast and slope are used as secondary variables. Ten-fold cross-validation is used to assess the prediction accuracy. The effects of these factors on the prediction accuracy are analysed using generalised linear models based on the information extracted from 18,350 prediction datasets produced in this simulation experiment.

The prediction accuracy depends on the methods, sample density, sample stratification, search window size, data variation and the study region. No single method performs always best for all scenarios tested. Three methods are more accurate than the control (IDS) in the north and northeast regions respectively; and 12 methods more accurate in the southwest region. A combined method, random forest and ordinary kriging (RKrf), is the most robust method based on the accuracy and the visual examination of prediction maps. This method is novel, with a relative mean absolute error (RMAE) up to 17% less than that of the control. Its RMAE is 15% lower in two regions and 30% lower in the third region than that of the best methods indetified in the previously published studies.

Procedures employed for data quality control and for selecting robust spatial interpolation methods provide guidelines to relevant future studies. This study revealed a new direction for spatial interpolation and opened an alternative source of methods for spatial interpolation. The findings in this study indicate that to achieve optimal predictions in regions with inherent differences in data nature, equal effort is needed to search for robust methods that are tailored to each region. The outcomes of this study can be applied to the modelling of a wide range of physical properties for improved marine biodiversity prediction. A number of suggestions are provided for further studies.

Abbreviations

AEEZ: Australian Exclusive Economic Zone
AMJ: Australian Marine Jurisdiction
GIS: geographic information systems
GLM: generalised linear model
GLS: generalised least squares
GRNN: general regression neural network
IDS: inverse distance squared
IDW: inverse distance weighting
KED: kriging with an external drift
LM: linear regression model
MAE: mean absolute error
MARS: the Marine Samples database
OCK: ordinary cokriging
OK: ordinary kriging
RFidw2: random forest and IDS
RK: regression kriging
RKlm: linear models and OK
RKglm: generalised linear models and OK
RKgls: generalised least squares and OK
RKrf: random forest and ordinary kriging
RMAE: relative mean absolute error
RMSE: root mean squared error
RRMSE: relative root mean square error
RT: regression tree
SVM: support vector machine
TPS: thin plate splines or Laplacian smoothing splines
UK: universal kriging
WGS84: World Geodetic System 1984

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

Prediction of marine biodiversity is important for developing ecosystem-based management strategies. Demand for spatially continuous information of environmental variables has increased with recognition of GIS and modelling techniques as powerful tools for environmental management and biological conservation. Spatially continuous data for a range of variables is required for seabed mapping and characterisation, statistical modelling and surrogacy research that will facilitate prediction of biodiversity. However, spatially continuous data are usually not available and information of many environmental variables is usually collected by point sampling. Spatially continuous data must then be inferred from this often sparse point data. This is particularly true of data from the ocean floor, including the continental Australian Exclusive Economic Zone (AEEZ) (Fig. 1.1), due to the expense and practical limitations of acquiring samples.

Statistical and mathematical techniques for spatial interpolation are essential tools for deriving spatially continuous data from point data to estimate values for unknown locations. These methods are often data and variable specific, and the estimations they produce are affected by a range of factors (Li and Heap, 2008). Existing research provides no consistent findings on how these factors affect the performance of spatial interpolators, making it difficult to select an appropriate method (Li and Heap, 2008).

The inverse distance squared (IDS) method is a commonly applied interpolator because of its relative simplicity. However, predictions using IDS are usually not reliable (Li and Heap, 2008). Currently Geoscience Australia derives raster sediment datasets (texture and composition) for approximately 8 million km² of the continental AEEZ using the IDS based on over 12,000 sparsely and unevenly distributed point samples stored in the Marine Samples Database (MARS, http://www.ga.gov.au/oracle/mars/index.jsp; Fig 1.1). These derived datasets are important for seabed mapping and characterisation, seabed habitat classifications and predictions of marine biodiversity to inform and support ecosystem-based management (*e.g.*, Department of the Environment and Heritage, 2005; Whiteway *et al.*, 2007).

In this study, we aim to identify the most appropriate methods for spatial interpolation of seabed mud content for the continental AEEZ using samples extracted from the MARS database. The performance of 14 statistical techniques for spatial interpolation is compared using a simulation experiment. We examine the effects of sample density and variation in the dataset on the performance of the methods. Samples are stratified using geomorphic provinces (Heap and Harris, 2008), and the effects of sample stratification on the performance of the methods are also examined. The effect of searching neighbourhood size is also tested. The performance of the methods is also visually examined based on their prediction maps.

This study covers several aspects of spatial interpolation, which are presented in the six chapters. Chapter 2 contains a brief description and discussion of data quality control. The experimental design and data analysis including variogram modelling and simulation modelling are described in Chapter 3. We analyse the simulation results, visually compare IDS and a few high performance methods and illustrate their applications, and discuss the findings and their implications in Chapter 4. Finally, in

Chapter 5, we summarise our findings and provide recommendation for the application of interpolation methods.

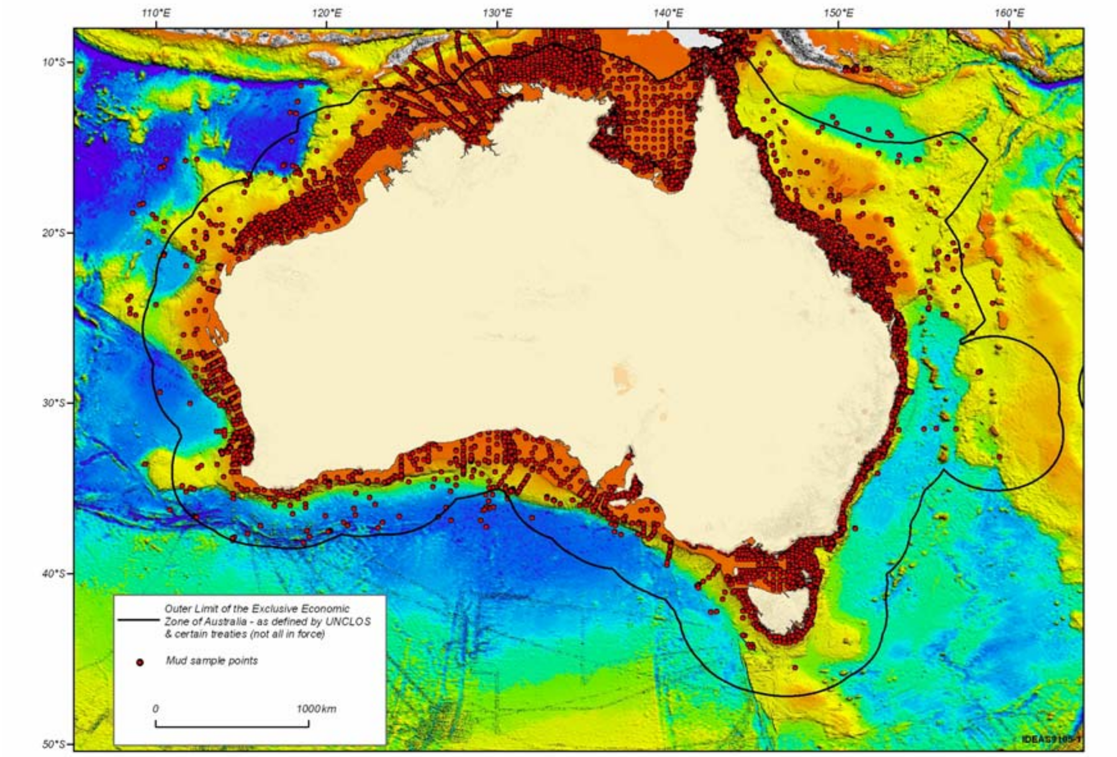


Figure 1.1. Spatial distribution of 12,506 mud samples in the continental AEEZ in World Geodetic System 1984 (WGS84).

This study provides suggestions and guidelines for improving the spatial interpolations of marine environmental data, in general, and results in more accurate mapping and characterisation of seabed in the continental AEEZ. This improved accuracy represents a more reliable and robust physical seabed dataset that will assist in the development of management and conservation strategies within the continental AEEZ. The importance of spatial continuous data and spatial interpolation in geoscience information and knowledge is further illustrated in Figures 1.2 and 1.3.

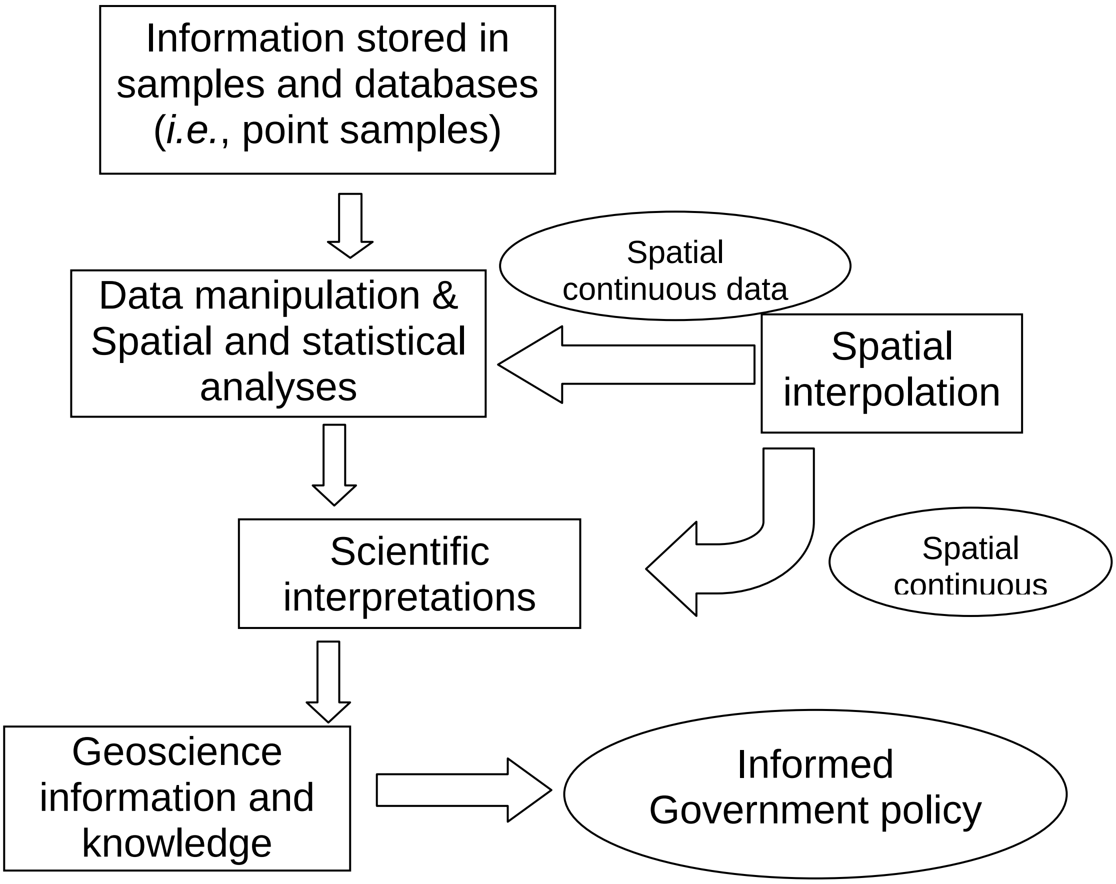


Figure 1.2. The importance of spatial continuous data in geoscience information and knowledge.

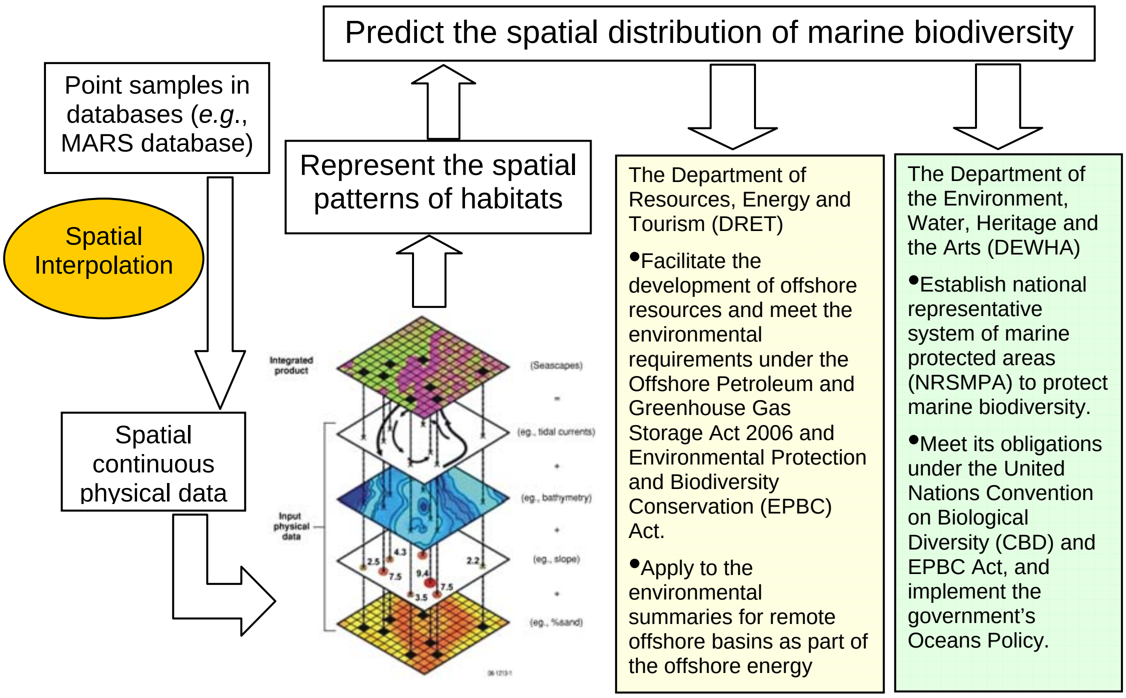


Figure 1.3. The importance of spatial interpolation in geoscience information and knowledge.

Chapter 2. Data Manipulation and Quality Control

The mud content data used in this study was sourced from the MARS database. The accuracy and precision of attributes assigned to sample points in the MARS database varies. At a national scale, this can result in data noise that prevents the identification of real trends. This chapter provides an overview of the MARS database and the factors that influence data quality, followed by a discussion of the sources of noise and the criteria used to 'clean' the data prior to analysis.

2.1. MARS Database

2.1.1. Content and structure

The MARS database was created in 2003 with the vision of collating all existing seabed sediment data for the Australian Marine Jurisdiction (AMJ) into a single publicly accessible database (http://www.ga.gov.au/oracle/mars). Samples may be geological, ecological or oceanographic. Data, including survey metadata, sample metadata, sample analyses and multimedia data (*e.g.*, seafloor images and video) is maintained in a spatial database (Oracle Spatial) and can be searched either by spatial location or data field. The database is consistently updated with new data, as it becomes available.

2.1.2. Data sources

Samples in the MARS database are collected by Geoscience Australia or contributed by external researchers. Analyses may be generated by Geoscience Australia Palaeontology and Sedimentology Laboratory or external researchers. The MARS database contains metadata and assays for seabed samples from the AEEZ collected on more than 300 marine surveys between 1899 and 2009. This includes samples and data contributed by over 275 institutions, both within Australia and from overseas. As a result of ongoing sampling and analysis by Geoscience Australia and external providers, the content of the MARS database is continually updated.

2.1.3 Sediment data

The MARS database facilitates ongoing development of a consistent quantitative sediment dataset for the AMJ. Standard data types generated by Geoscience Australia from MARS include grainsize (Vol%; and gravel/sand/ mud content, Wt%) and carbonate content of the bulk, gravel, sand and mud fractions (Wt%).

- Mean grain size (Vol%; µm): The grain size distribution of the 0.01–2,000 µm fraction of the bulk sediment is determined with a Malvern Mastersizer 2000 laser particle analyser. All samples are wet sieved through a 2,000 µm mesh to remove the coarse fraction. A minimum of 1 g is used for relatively fine material and between 2–3 g for relatively coarse material. Samples are then ultrasonically treated to ensure that good dispersion of the particles occurs. The grain size distribution is then derived from the average of three runs of 30,000 measurements that are divided into 100 particle size bins of equal size.
- Grain size (Wt%): Gravel, sand, and mud contents are determined by passing 10–20 g of bulk sediment through standard mesh sizes (>2,000 µm; Gravel; 63 µm-2,000 µm; Sand; <63 µm Mud). The gravel, sand, and mud contents are derived by the relative weight proportions of dry sediment.

- Carbonate content (Wt%): Bulk, sand and mud carbonate contents are determined on 2–5 g of material using the 'Carbonate bomb' method of (Muller and Gastner, 1971). Carbonate gravel contents are determined by visual inspection.

Where the data generation method is deemed to be compatible, data from external providers is also included. Variable sample ages, physical properties and volumes often result in some analyses being unavailable for some samples.

2.2. Data for this study

2.2.1 Quality control

Samples that failed to meet the minimum metadata standards outlined in Geoscience Australian Data Standards, Validation and Release Handbook, 4th Edition (Geoscience Australia, 2004) were excluded in this study. Only analyses conducted on dredges, grabs or the top 10 cm of a core and where the gravel, sand and mud fractions totalled 100% +/- 1% were included. Core samples that did not include depth measurements were also excluded. This resulted in a total of 12,506 surface sediment data points in the MARS database on 21 April 2008 (Fig. 1.1).

2.2.2 Additional attributes

Additional information, not existing as fields in the MARS database, was added to the samples used in this study so that an analysis of sample stratification could be completed. This information comprises water depth, and position relative to the geomorphic feature and province boundaries of Heap and Harris (2008). Values were obtained by spatial intersection of the point sample data with raster/polygon layers in ArcGIS. Data points were given the attributes of the raster cell or polygon that they occurred within.

2.3. Data noise and data cleaning

Data noise refers to data that does not truly represent the trends that exist in the real world, which may obscure the trends and lead to error. Data noise is inevitable in regional and national scale sediment datasets as changes in sediment properties frequently occur on finer scales than sample densities are adequate to detect. However, some noise results from inaccuracies in collection, analysis or interpretation. Data noise may result from various sources. Where these inaccuracies can be identified, datasets can be cleaned by removing samples that are deemed to be incorrect.

For this experiment, the point data were cleaned by: a) excluding points for which mud contents were deemed to have high uncertainty; and b) removing data points that correspond to areas of high uncertainty in spatial datasets used to stratify sediment data for modelling and analysis of results.

Factors which affect the accuracy of the sediment dataset are identified and listed below. Procedure and criteria are developed for cleaning the dataset based on the factors identified.

2.3.1. 'Within the continental AEEZ'

Origins of data noise: Some samples selected for this experiment were collected outside the research domain (*i.e.*, the continental AEEZ; Fig. 1.1).

Principles and procedure of data cleaning: The sediment data was intersected with the continental AEEZ boundary and samples outside of this boundary were assigned a null value. Samples outside the domain were then excluded. This reduction resulted in a total of 11,082 samples.

2.3.2. Samples with mud content

Origins of data noise: Weight% mud contents are not available for all samples within the continental AEEZ in the MARS database.

Principles and procedure of data cleaning: Samples without mud content information (*i.e.*, missing value) were assigned a null value (*e.g.*, mud = -9999). As this study focuses on mud content, samples without mud content information were deleted. This resulted in a total of 10,825 samples.

2.3.3. Sample type

Origins of data noise: Unlike grab and core samples that are collected at a single point location, dredges represent sediment collected over a transect that is anywhere from a few hundred meters to more than a kilometre in length. This results in uncertainty in the mud content for samples collected by dredge, as it is impossible to accurately assign the mud content to any single point along the transect. Mixing of sediment during the sampling process may be a further confounding factor.

Principles and procedure of data cleaning: Mud content generated from dredge samples were considered less accurate than other data points. Dredges were identified using the sample type field in the MARS database. Samples listed as "Dredge Benthic", "Dredge Chain Bag", "Dredge Pipe", and "Dredge Unspecified" were classified as dredges (0). All other samples (including those for which sampling device was not known) were classified as non-dredges (1). Dredged data samples were excluded because of the uncertainty associated with their location and the inaccuracy in the sampled sediment concentration resulted from mixing during the process of dredging. Exclusion of dredge samples resulted in a total of 7,025 samples.

2.3.4. Bathymetry

Origins of data noise: Grid cells in bathymetry datasets are generally given negative values, representing metres below sea level. However, the methods used to produce bathymetry grids occasionally resulted in grid cells with positive values.

Principles and procedure of data cleaning: Bathymetry data are used as secondary information for the modelling in this study. Positive values are not acceptable and can not be corrected within the timeframes of this study. Therefore sediment samples corresponding to points with positive bathymetry values had been deleted, which led to a further removal of 273 samples. This further reduced the size of the dataset to 6,752 samples.

2.3.5. Surficiality

Origins of data noise: Limited sample volume (particularly for cores) or mixing of sediment during sampling, transport and subsequent storage mean that for many sample locations, sediment collected from the seabed surface is not always available. This is not necessarily a problem for all samples as the effects of homogenisation of the shallow sub-surface seabed sediments (*i.e.*, bioturbation) usually negate having to use the top-most sediments. In the case of cores where the top-most sediments were unavailable, samples were included if they had been collected within 10 cm of the seabed surface (*e.g.*, their base depth was <10cm).

Principles and procedure of data cleaning: Sediments collected from the seabed surface are considered to be the most accurate. These were identified using the Top and base depth fields in MARS database. Samples with a top depth of 0 and base depth of <5 were considered surficial and given a surficiality value of 1. These were distinguished from samples which were collected at all other depths below the seabed (top depth $\neq$ 0), which are likely to exclude uppermost fine pelagic sediments, and were given a surficiality value of 0. Samples with a base depth of >5 cm below the seabed were deemed to be a mixture of surficial and non-surficial sediments and were also given a surficiality value of 0. Sample types for which no depth are generally given (*e.g.*, grabs) were assumed to have a top depth of 0 and consequently assined a surficiality value of 1.

2.3.6. Multiple mud values for a single cell/location

Origins of data noise: When the continental AEEZ is divided into a 0.01 degree grid for modelling, some cells contain multiple samples. These generally represent samples of different types collected from the same station, or replicate samples taken in close proximity to one another. In such cases a range of mud contents may exist for a grid cell.

Principles and procedure of data cleaning: It was necessary for this study to have a single value for mud content in each grid cell. We selected the analysis likely to best represent the actual mud content at that location using a decision tree classification. The best sample was given a value of “1” while all other samples were given a value of “0”.

Decision 1: “multiple samples in the same grid cell?” This was determined by intersecting the sample points and the grid in ArcMap and identifying samples sharing the same cell id. Where only one sample existed for a cell, samples were excluded from decisions and given a value of “1”.

Decision 2: “sample with highest mud content.” This step was necessary to arrive at a single value for each cell and was decided on for two reasons: 1) we are mapping the distribution of mud content, and if mud is present within the grid cell we wanted this captured in the data; and 2) we assume that some percentage of the fine mud was lost during sampling operation, so if multiple samples were collected from the same location within the grid cell, the one that retained the most mud is likely to be the most accurate measure of the actual mud content. The sample with the highest mud content was given a value of “1” and all other samples were given a value of “0”.

This process was repeated for multiple samples at the same location. Latitude and longitude values were sourced from the MARS database. The most appropriate sediment sample for each location was identified based on depth (*i.e.*, closest to the surface) and mud content (*i.e.*, highest). Best samples are given a '1', all other samples are given a '0'. Removal of multiple samples for the same cell/location resulted in a total of 5,281 samples remaining.

2.3.7. Geomorphology

Origins of data noise: Geomorphic province and feature polygons used in this experiment are sourced from the National Marine Bioregionalisation of Australia (Department of the Environment and Heritage, 2005), which divide the continental AEEZ into for four geomorphic provinces, and 21 geomorphic feature types (Heap and Harris, 2008). However, in reality boundaries between some geomorphic features are gradational and therefore can not be always accurately represented by hard boundaries. Also, placement of the boundaries can be subjective as well as being based on other physical datasets containing substantial uncertainty. As a result, seabed characteristics occurring near the feature boundary may be for that feature or possibly for an adjacent feature.

Principles and procedure of data cleaning: Each sample point was allocated a geomorphic province and feature code by intersecting the point shapefile with the geomorphology polygons in ArcMap. The geomorphic provinces comprise four classes (*i.e.*, province code: 1-4) and the geomorphic features comprise 21 classes (*i.e.*, feature code: 1-21). The information was used to stratify the samples. Samples are distributed unevenly across the four provinces (Fig. 2.1) and 21 features (Fig. 2.2). Most geomorphic features contained too few samples for this classification to be used as secondary information in modelling suggesting that only geomorphic provinces could be used in this study.

Samples were allocated to a geomorphic province code (1-4), including a measure of confidence in this classification (0, 1). Confidence was assumed to be lower near boundaries between provinces. To derive the confidence, a 20 km-wide buffer around each boundary (*i.e.*, a distance of 10 km either side) and the boundary at the coast or outer EEZ limit was not buffered unless it represented a transition between provinces. Samples occurring within the buffered area were given a low confidence value of “0” and all other samples were given a high confidence value of “1”. Exclusion of samples with low confidence resulted in a final dataset of 4,817 samples. Therefore, the sample size was gradually reduced from 12,506 to 4,817 (Fig. 2.3). The spatial distribution of sample size was illustrated in Figure 2.4. This final dataset was used in this study.

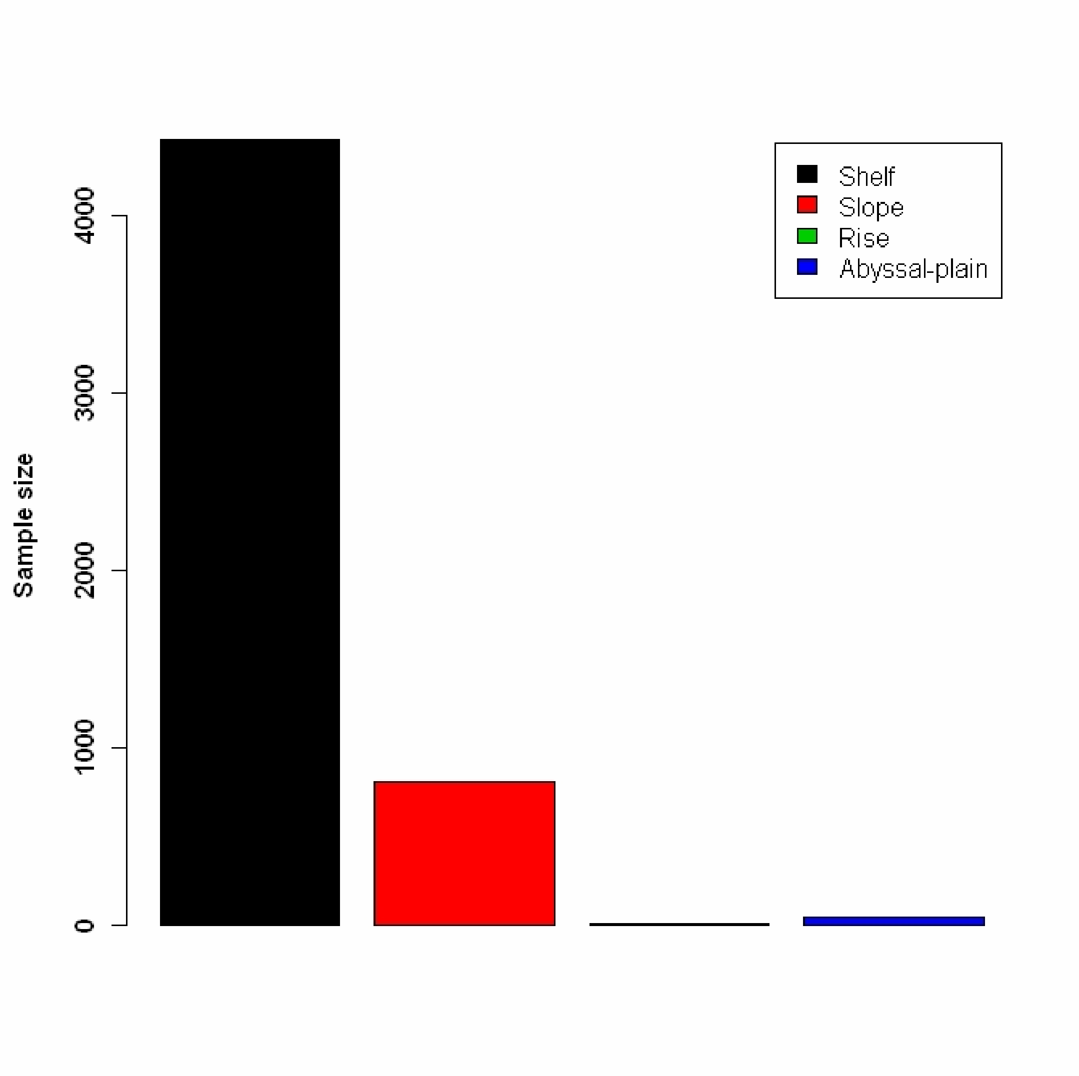


Figure 2.1. Sample size of the four geomorphic provinces in the continental AEEZ.

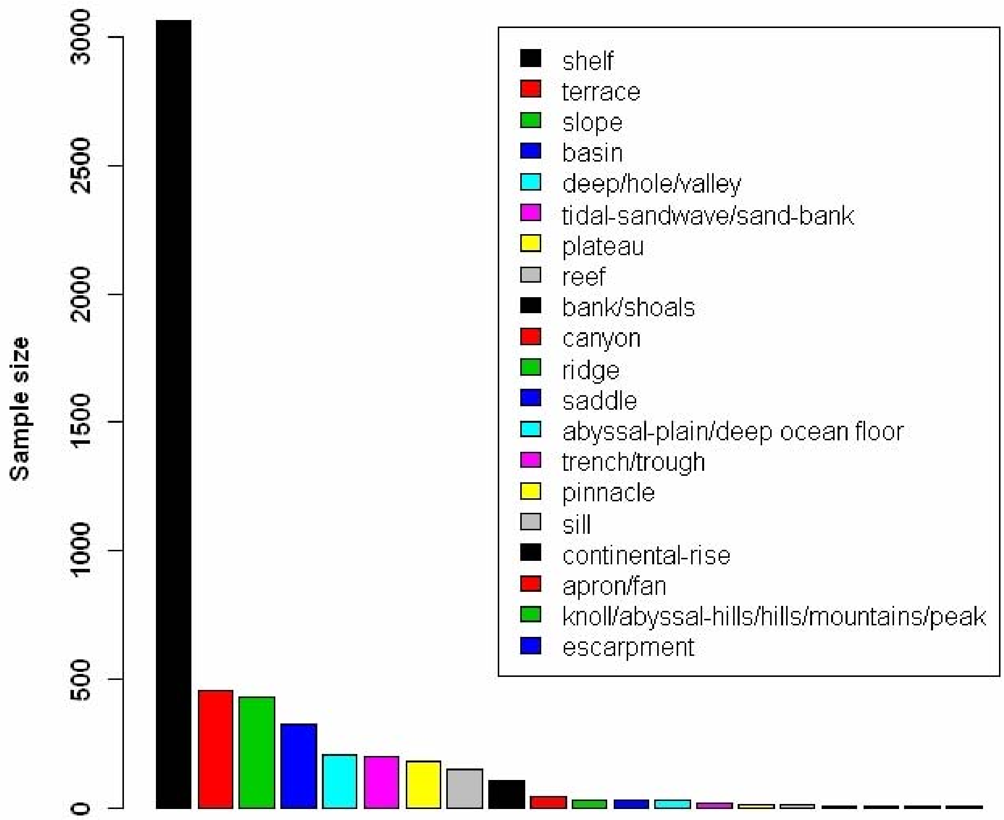


Figure 2.2. Sample size for geomorphic features in the continental AEEZ. Of the total 21 features, 20 contain samples.

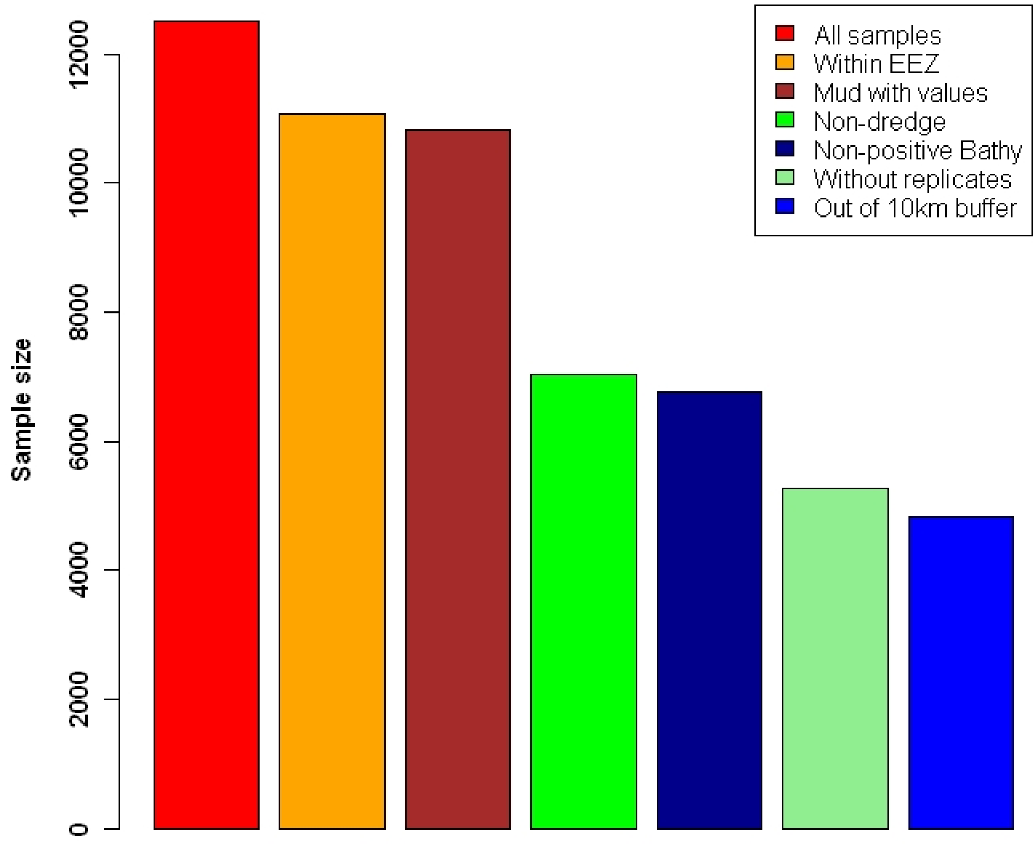


Figure 2.3. Changes of sample size with data cleaning criteria.

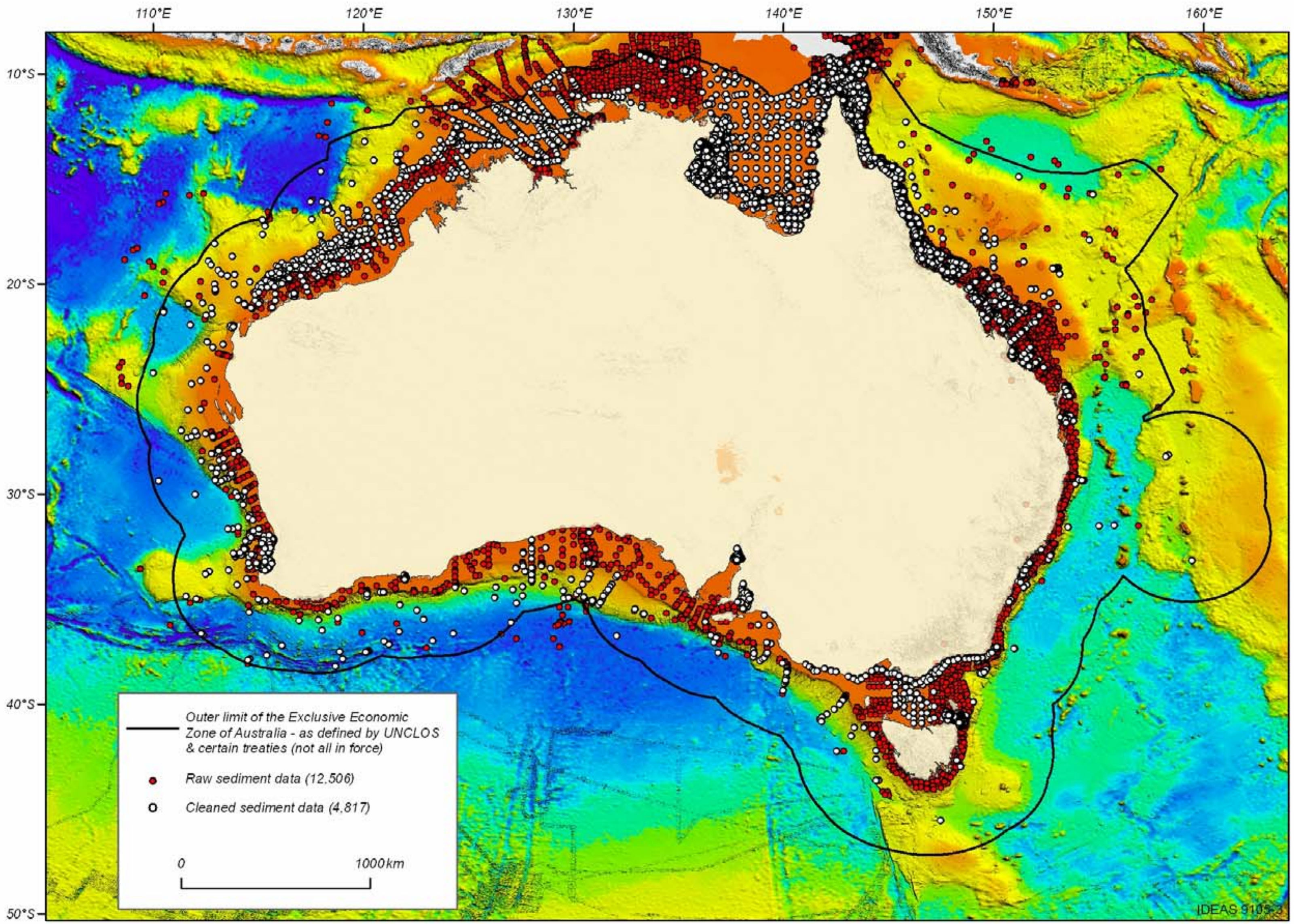


Figure 2.4. Spatial distribution of mud samples in the continental AEEZ, with the original 'raw' (red) and 'cleaned' datasets (white).

Chapter 3. Experimental Design and Data Analysis

Many factors may affect the performance of spatial interpolation methods. These factors may include sampling design, surface type, sample size and density, spatial distribution of samples, quality of samples, spatial structure of data, correlation of primary and secondary variables, and interaction among some of these factors (Li and Heap, 2008). The choice of spatial interpolators is also critical.

Given that the samples have already been collected, little can be done in regard to such as sampling design and quality control in sample collection. Sample density and spatial variations between regions are key factors with respect to data quality for interpolations (Li and Heap, 2008). After an initial analysis, four factors were identified to be important for spatial interpolation: 1) spatial variation (*i.e.*, data from different regions); 2) choice of spatial interpolation methods; 3) sample density; and 4) sample stratification. The results of the effects of sample density and regional variation on the spatial interpolation may provide a foundation for further tests on other factors.

In this chapter, we describe the study area, introduce the spatial interpolation methods, describe the sample density and secondary information, and finally discuss the assessment of the method performance.

3.1. Study area

Samples in three regions (north, northeast and southwest) were selected from the continental AEEZ for this study (Fig. 3.1). These three regions were selected because thay have contrasting physical properties in terms of area, orientation, geomorphic composition, and bathymetry (Tables 3.1 & 3.2; Fig. 3.1), which provide different scenarios for this simulation experiment. The area of geomorphic provinces was also different within each selected region and varied much among these regions (Table 3.2). Sample coverage (sample size and spatial distribution of samples) are also different in the three regions and their spatial distribution was also uneven, with most samples acquired from area near the shore (Fig. 3.2). Sample density was very low, varying from 0.3 to 1.9 samples per 1,000 km^2 (Table 3.1).

The north region covers the continental shelf and slope geomorphic provinces, and is characterised by relatively shallow water depths (Tables 3.1 & 3.2). The region is generally oriented east-west and encompasses the Gulf of Carpentaria and Arafura Sea (Fig. 3.1). Data in the north region is relatively evenly distributed compared to the other regions (Fig. 3.2).

The northeast region comprises all four geomorphic provinces with varying water depths (Tables 3.1 & 3.2). The region is generally oriented southeast-northwest (Fig. 3.1). This region contains the largest number and greatest spatial density of samples, with most samples on the shelf concentrated in the Great Barrier Reef lagoon, but relatively few samples collected from the slope, rise and abyssal plain/deep ocean floor.

The southwest region comprises all four geomorphic provinces and a large range of water depths (Tables 3.1 & 3.2). This region is oriented north-south (Fig. 3.1). The sample density is much lower than the other two regions (Table 3.1; Fig. 3.2).

These three regions provide a good representation of the geomorphic characteristics of the continental AEEZ and different sample coverages, which formed a good base for testing the performance of statistical techniques for spatial interpolation of mud content.

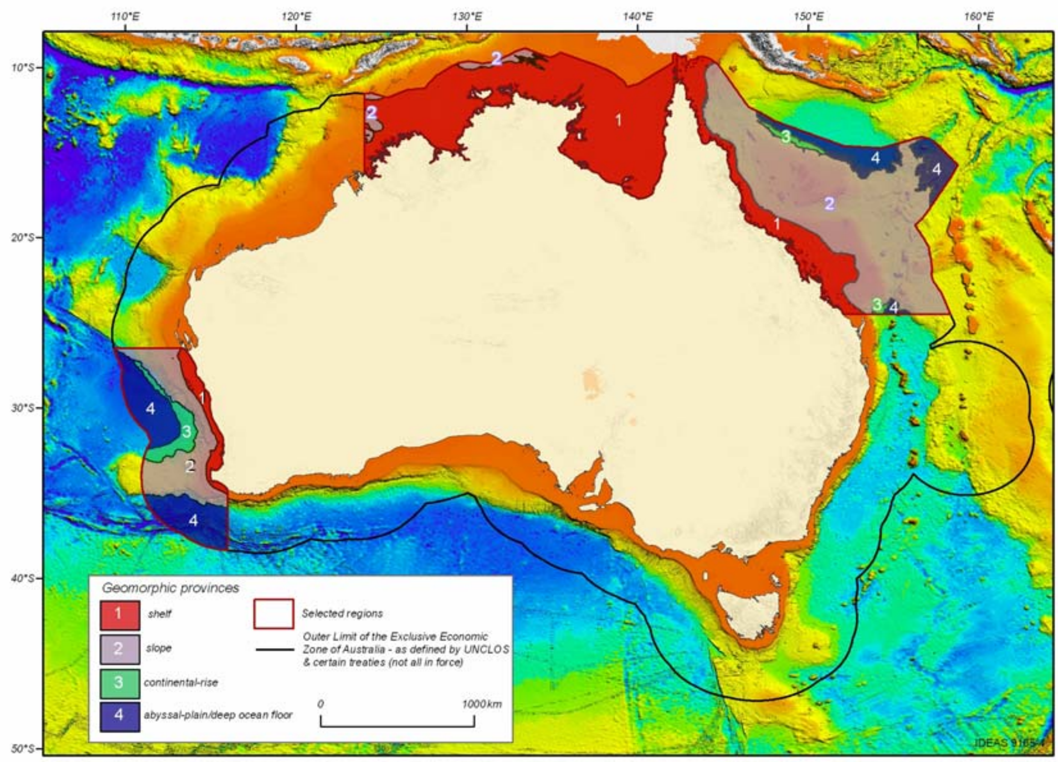


Figure 3.1. Three regions selected for testing the performance of spatial interpolation methods from the continental AEEZ, including spatial distribution of geomorphic provinces.

Table 3.1. Summary of features of each selected region.

Region	Orientation	Bathymetry (m)	Area (km^2)	Sample No	Sample density (per 1000 km^2)
North	W-E	-318	896,700	1,687	1.9
Northeast	NW-SE	-4,150	1,366,100	1,828	1.3
Southwest	N-S	-5,539	523,400	177	0.3

Table 3.2. The area (km^2) of the geomorphic provinces in each region.

Region	Shelf	Slope	Rise	Abyssal plain/ Deep ocean floor
North	855,100	41,600	0	0
Northeast	254,400	930,400	18,600	162,800
Southwest	52,900	214,900	52,200	203,200

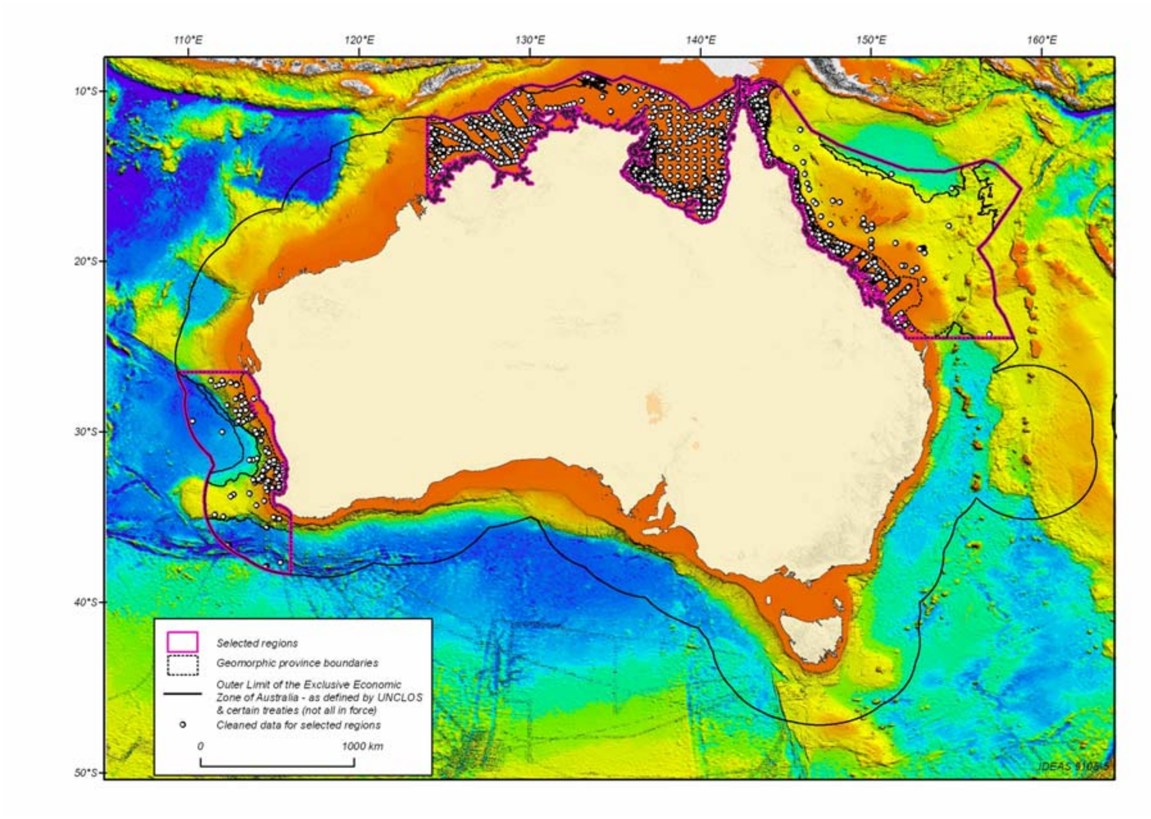


Figure 3.2. Spatial distribution of samples with mud content for the three selected regions, including their occurrence in the geomorphic provinces.

3.2. Statistical and mathematical methods

A total of 14 statistical and mathematical methods for spatial interpolation were compared in this study (Table 3.3). These methods fall into five categories: 1) non-geostatistical spatial interpolation method; 2) geostatistical method; 3) spatial statistical method; 4) machine learning method; and 5) combined method.

3.2.1. Non-geostatistical spatial interpolation methods

Although the inverse distance weighting (IDW) method performs poorly in most cases in terms of prediction accuracy, it provides a good control as it is a widely used spatial interpolation tool at Geoscience Australia and one of the most commonly compared methods in spatial interpolation (Li and Heap, 2008). The thin plate splines (TPS) method was also included in the experiment because of its good performance in some studies (Hartkamp et al., 1999; Jarvis and Stuart, 2001; Laslett et al., 1987).

3.2.2. Geostatistical methods

Kriging with an external drift (KED) and ordinary cokriging (OCK) (Goovaerts, 1997) were compared because they have been proven to be very accurate when appropriate high quality secondary information is available (Li and Heap, 2008). Ordinary kriging (OK) was considered as it is one of the most commonly compared methods in spatial interpolation (Li and Heap, 2008). Universal kriging (UK) was also employed in this experiment as a trend in the data over space was detected in a preliminary analysis.

Table 3.3. Statistical methods compared for spatial interpolation in this study.

No	Method
1	Inverse distance weighting (IDW)
2	Generalised least squares trend estimation (GLS)
3	Kriging with an external drift (KED)
4	Ordinary cokriging (OCK)
5	Ordinary kriging (OK)
6	Universal kriging (UK)
7	Regression tree (RT)
8	Thin plate splines (TPS)
9	General Regression Neural Network (GRNN)
10	Support vector machine (SVM)
11	Linear models and OK (RKlm)
12	Generalised linear models and OK (RKglm)
13	Generalised least squares and OK (RKgls)
14	RandomForest and OK (RKrf)

3.2.3. Spatial statistical method

Generalised least squares trend estimation (GLS) (Bivand et al., 2008) is used in this study as it allows errors to be correlated (Pinheiro and Bates, 2000; Venables and Ripley, 2002) as the samples of mud content in a region are often spatially correlated.

3.2.4. Machine learning methods

Three machine learning approaches were used in this study: regression tree (RT); general regression neural network (GRNN); and support vector machine (SVM). The application of SVM to spatial interpolation has not been reported previously (Li and Heap, 2008). A fourth method, random forest (Breiman, 2001; Strobl et al., 2007) that has not previously been applied to spatial interpolation (Li and Heap, 2008), was used in combination with OK as a combined method below.

3.2.5. Combined methods

Combined methods include linear regression models and OK (RKlm), generalised linear models and OK (RKglm), generalised least squares and OK (RKgls) and random forest and OK (RKrf). These four combined methods are modified versions of regression kriging type C (RK-C) (Asli and Marcotte, 1995; Odeh et al., 1995) that are less sensitive to data variation and more accurate than other methods (Li and Heap, 2008). For these combined methods, linear regression models (lm), generalised linear models (glm), generalised least squares (gls) and random forest models were first applied to the point samples, and OK was then applied to the residuals of each model. Lastly, the predicted values of each model and the corresponding kriged values of residuals were added together to produce the final predictions. RKrf is a new combined method that has not been applied in previous studies.

Most of these methods were briefly described by Li and Heap (2008). The machine learning methods used in this study are briefly introduced in appendix A.

3.3. Sample density

Five sample densities were used to test the performance of statistical techniques for spatial interpolation for each region, namely 20%, 40%, 60%, 80% and 100% of the total samples collected. The resultant sample size for each density (Table 3.4) was different between regions. For densities less than 100% samples were randomly sampled from the full dataset for each region.

Table 3.4. Sample size in each of the three selected regions for their corresponding sample density.

Sample density	20%	40%	60%	80%	100%
North	337	675	1012	1350	1687
Northeast	366	731	1097	1462	1828
Southwest	35	71	106	142	177

Area per sample changed with regions (Table 3.5). The lowest area per sample is in the north region and the highest in the southwest region, with a difference about five-folds. The area per sample in the northeast is about two-thirds of north region. These differences are expected to have some influence on the performance of the methods.

Table 3.5. Area per sample (km^2/sample) in each of the three selected regions for their corresponding sample density.

Sample density	20%	40%	60%	80%	100%
North	2661	1328	886	664	532
Northeast	3733	1869	1245	934	747
Southwest	14953	7371	4937	3685	2957

3.4. Secondary information

A number of variables can be used as the secondary information to improve the performance of spatial interpolation techniques as discussed by Li and Heap (2008). Following a preliminary analysis, geomorphic provinces and bathymetry data that were available at a resolution of 0.01 degree were considered as important secondary information in this study. Bathymetry has previously been used to improve the performance of spatial interpolators (Verfaillie et al., 2006). The relationship between the bathymetry and sediment grain-size depends on the topography, and the substrate type (Verfaillie et al., 2006), so the inclusion of such information was expexted to improve the predictions. Distance-to-coast and seabed slope are likely to have some influence on the transportation of mud from onshore sources and preferential deposition of mud in regions with lower seabed gradient, so they were also considered as important secondary information in this study to improve the overall predictions.

All datasets were generated in ArcGIS and, where necessary, resampled to a 0.01 degree resolution. Distance-to-coast represents the linear distance (in decimal degrees) from any location to the nearest point on the Australian coastline. The data were generated by selecting the Australia coastline (including the mainland, Tasmania and adjecent major islands) from Geoscience Australia's 250k coastline dataset, simplifying features using a 30 km tolerance, and then calculating the Euclidean distance to this line from each grid cell.

Slope was generated by dividing Geoscience Australia's 250 m spatial resolution bathymetry grid into grids for each of the 10 UTM Zones covered by the AEEZ (49–58° S) and re-projecting these appropriately into UTM grids. Slope gradient (in degrees) was calculated separately for each grid, and then each of the slope grids was projected back to WGS84 and merged. The merged grid was finally resampled to 0.01 degree resolution.

Sample stratification can improve the estimation of the spatial interpolators by reducing the variance of the data (Stein et al., 1988; Voltz and Webster, 1990). Geomorphic features for the continental margin of Australia (Heap and Harris, 2008) are expected to provide valuable information for stratifying the samples. However, the small sample size and the uneven spatial distribution of samples (Table 3.6 and Fig. 3.2) mean that sub-setting by feature type did not provide adequate sample points for modelling. Because of this, geomorphic provinces were used in this study. Although an individual province generally covers a greater area than a feature, still few samples occur in rise and abyssal plain/deep ocean floor provinces, which may cause problems in the simulation modelling and influence the outcome of the experiment.

Table 3.6. Sample size in each of the geomorphic provinces by study region.

Geomorphic province	Shelf	Slope	Rise	Abyssal plain/ Deep ocean floor
North	1634	53	0	0
Northeast	1785	41	0	2
Southwest	65	101	3	8

A range of other variables could be used as secondary information to improve the spatial interpolation of marine environmental data. They may include topology, substrate type, slope, sea floor temperature, seabed exposure, disturbance, and those used in Whiteway *et al*. (2007). Oceanographic and sedimentological processes are also known to influence distribution of mud, for instance, combined flow bed shear stress (*i.e.*, a combination of the effects of surface ocean waves, tidal, wind and density driven ocean currents) (Hemer, 2006). However, information for these variables is not available for the whole continental AEEZ at the resolution required for this study. Consequently, these variables were not used in this experiment.

3.5. Simulation Modelling

Simulation modelling is an essential and key section to this study. However, it is a bit too technical for environmental scientists so we put the main body of this section in the Appendix B for interested readers. This Appendix covers a number of issues: identification of appropriate data transformation for mud content and relavent secondary variables; analyses of correlation between mud content and secondary variables; discussion of data projection, variogram model, anisotropy and variogram model selection; and model and parameter specification of 37 sub-methods.

3.6. Assessment of method performance

Based on the experimental design, in this study we applied the 14 methods to 15 mud sediment datasets (five sample densities and three regions). The basic summary statistics for mud samples in these 15 datasets are listed in Table 3.7.

To compare the performance of the statistical methods with different sample densities in different regions, a 10-fold cross-validation was used. The existing cross-validation programs could be used, but due to random sampling, each method may receive different samples for prediction and validation. To avoid such random error, we randomly split each of the 15 datasets into 10 sub-datasets. Nine sub-datasets were combined and used for model development and making predictions and the remaining one was used to validate the predictions. This was repeated, varying the validation dataset, until all 10 sub-datasets had been allocated for validation. Consequently, 150 datasets were generated for model development (*i.e.*, for making predictions) and another 150 datasets for model validation. All methods were applied to the same data in making predictions. Thus, random errors were avoided, though it is much more time consuming than using an existing cross-validation program. Each method generated 150 prediction datasets. A total of 18,350 prediction datasets were produced for assessing the performance of the methods (Fig. 3.3). The performance of each method was then assessed by quantifying the errors in predictions. For each method, predictions in the 150 datasets generated were compared to observed values in the 150 corresponding validation datasets. All data manipulation and computation were conducted in R (R Development Core Team, 2007).

Table 3.7. Summary statistics for mud samples in 15 datasets by 5 sample densities in each of the three study regions.

Region	Sample density (%)	Sample size	Minimum	Mean	Maximum	Std dev
North	20	337	0	32.2	97.76	28.21
North	40	675	0	34.13	98.86	28.45
North	60	1012	0	33.64	99.2	28.1
North	80	1350	0	34.04	99.2	28.12
North	100	1687	0	34.04	99.2	28.15
Northeast	20	366	0	25.33	99.49	20.84
Northeast	40	731	0	25.43	98	21.73
Northeast	60	1097	0	25.38	99.49	21.24
Northeast	80	1462	0	25.19	99.49	21.35
Northeast	100	1828	0	25.09	99.49	21.35
Southwest	20	35	0.01	48.25	97.86	37.84
Southwest	40	71	0.01	48.09	98.25	37.65
Southwest	60	106	0.01	47.73	98.25	37.58
Southwest	80	142	0.01	45.66	98.25	37.1
Southwest	100	177	0.01	46.23	98.25	37.03

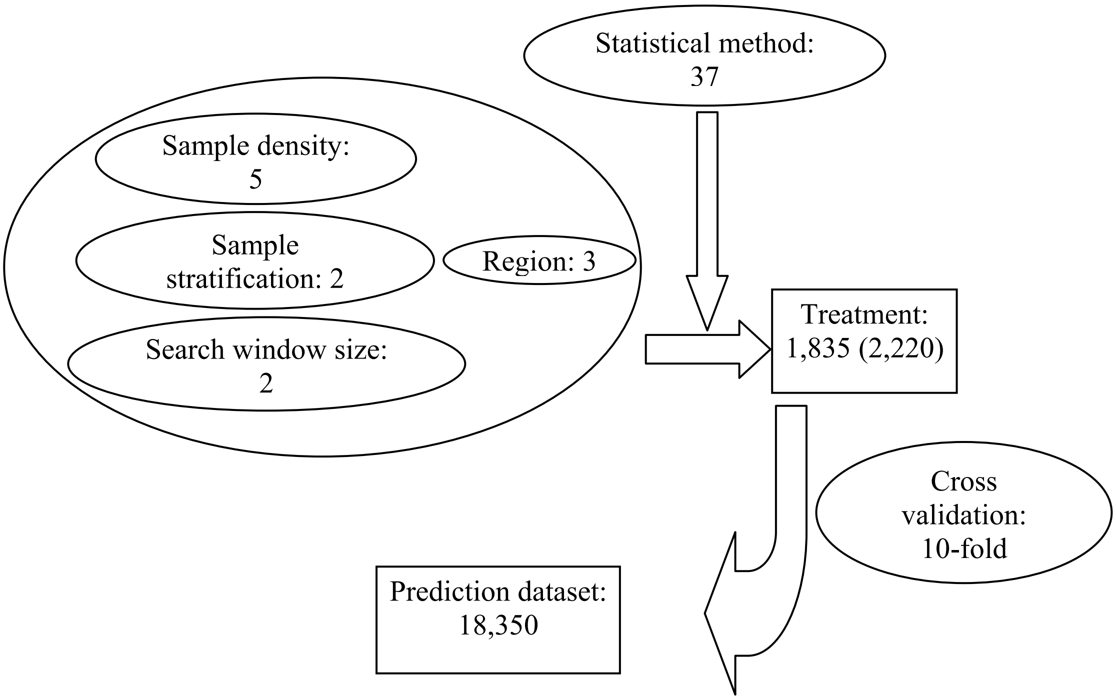


Figure 3.3. Prediction datasets produced fron the simulation experiment. Theorectically 2,220 treatments were expected, but actually 1,835 treatments were obtained because some methods were only applied datasets with 100% sample density or only applied to one of the relevant treatment levels.

Several error measures for assessing method performance have been proposed and have been briefly reviewed by Li and Heap (2008). Mean absolute error (MAE) and root mean square error (RMSE) are argued to be among the best overall measures of model performance as they summarise the mean difference in the units of observed and predicted values (Willmott, 1982). In addition, two newly proposed measures: 1) relative mean absolute error (RMAE); and 2) relative root mean square error (RRMSE), are not sensitive to the changes in unit/scale (Li and Heap, 2008). Therefore these four measurements were used to assess the performance of the various spatial interpolation methods. Their formulae are listed as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |(p_i - o_i)| \qquad (1)$$

$$RMSE = [\frac{1}{n}\sum_{i=1}^{n} ([p_i - o_i])^2]^{1/2} \qquad (2)$$

$$RMAE = MAE/o_m 100 \qquad (3)$$

$$RRMSE = RMSE/o_m 100 \qquad (4)$$

where n is the number of observations or samples, o is the observed value, p is the predicted or estimated value, and o_m is the mean of the observed values.

3.7. Data Analysis

In this study, the accuracy of predicted mud content was analysed using data in Appendix C according to 32 sub-methods, sample density, stratification, searching window size and data variation. The data variation or coefficient of variation (CV) is the ratio of the standard deviation to the mean mud content of the samples used for prediction. The order of the parameters in the model was based on the assumption that the first four parameters are orthogonal and their order should not affect the results. CV was used as the last parameter in the model because it is not a component of the experimental design, but we expect some contribution from it in terms of deviance explained since it may act as a hidden treatment.

The results were similar in terms of MAE, RMAE, RMSE and RRMSE (Appendix C). As we need to assess the effects of sample density and the mean for each density was different, effects of such differences need to be taken into account. Hence we used the results of RMAE, which was developed to remove the effects of such difference (Li and Heap, 2008), for further analysis.

The five sub-methods (GRNN, RT, TPSr, TPSt and SVM) that were only applied to one level of stratification and searching window size were analysed separately because inclusion would make the experimental design unbalanced and cause difficulties in explaining the analysed results. RKgls1, RKgls2, RKgls3, RKgls4 to RKgls5 that were only applied to sample density of 100% were excluded from data analysis in the northeast and southwest regions because their inclusion would also make the experimental design non-orthogonal. The methods excluded in the statistical analyses were plotted against all the other methods for visual comparison.

The data were finally analysed using generalised linear models with a quasi family, a log link and a variance being mean squared in R (R Development Core Team, 2007).

Chapter 4. Results and Discussion

In this chapter, we first present the basic statistics derived from the simulation experiment, which is followed by the results and discussion of the effects of the experimental factors (*i.e.*, statistical methods, sample density, data variation, stratification and searching window size) and their interactions. After a preliminary analysis, the difference in method accuracy between the three regions were found to be considerable and the methods applied were slightly different between regions as discussed below, so the results for each region were analysed and reported separately. Since our aim is to identify the best method for spatial prediction, we first compare the performance of the methods for dataset with 100% sample density against the control method (IDS), and then examine the effects of sample density and compare the possible effects of data variation on the most accurate methods and the control method. Finally, we visually examine the spatial predictions of these methods.

4.1. Statistics of the simulation experiment

4.1.1. Summary statistics

In this study, a total of 14 statistical techniques (37 sub-methods in total) were applied to mud content samples. Two levels of sample stratification and two levels of searching window size were considered in this study. Each sub-method was applied to 5 sample densities in the three study regions, resulting in 15 applications (Table 4.1), with the exception of RT, GRNN, and SVM which were only applied to stratified samples with global search, and both TPSr and TPSt which were applied to non-stratified samples with a local search. RKgls1 to RKgls5 were applied to all sample densities in the north region, but only to sample density of 100% in the northeast and southwest regions because of the heavy demand on computational time. On the basis of the validation of the predictions against validation datasets, statistics measuring the accuary of each method are summarised in Appendix C that formed a base for analysing the performance of the methods in relation to other experimental factors. The basic statistical summaries of predictions of each method are also listed in Appendix C.

Table 4.1. Number of applications for each spatial interpolation method.

Searching window size	Global		Local	
Sample stratification	**non.str**	**str**	**non.str**	**str**
IDW2	15	15	15	15
IDW1	15	15	15	15
IDW3	15	15	15	15
IDW4	15	15	15	15
GLS1	15	15	15	15
GLS2	15	15	15	15
KED1	15	15	15	15
KED2	15	15	15	15
KED3	15	15	15	15
KED4	15	15	15	15
OCK1	15	15	15	15
OCK2	15	15	15	15
OCK3	15	15	15	15
OCK4	15	15	15	15
OK	15	15	15	15
UK	15	15	15	15
RT	0	15	0	0
TPSr	0	0	15	0
TPSt	0	0	15	0
GRNN	0	15	0	0
SVM	0	15	0	0
RKglm1	15	15	15	15
RKglm2	15	15	15	15
RKglm3	15	15	15	15
RKglm4	15	15	15	15
RKglm5	15	15	15	15
RKgls1	7	7	7	7
RKgls2	7	7	7	7
RKgls3	7	7	7	7
RKgls4	7	7	7	7
RKgls5	7	7	7	7
RKlm1	15	15	15	15
RKlm2	15	15	15	15
RKlm3	15	15	15	15
RKlm4	15	15	15	15
RKlm5	15	15	15	15
RKrf	15	15	15	15

4.1.2. Overall effects of the experimental factors

All of the experimental factors have significant impacts on the accuracy of the estimations for all three study regions, and some of their interactions are also significant (Tables 4.2-4.4). Significant higher order interactions are highlighted in these tables. This simulation experiment provides a tremendous amount of information about various experimental factors and their interactions. Since we aim to identify the best statistical method for spatial interpolation in each region, we did not investigate all significant interactive effects in the tables. Instead, we first compare the effects of statistical methods and their interactions with stratification and searching window size for the dataset with 100% sample density within each region. We then examine the effects of sample density and its interaction with the methods. Finally, we compare the possible effects of data variation and its interaction with other factors.

Table 4.2. Effects of method, sample density, stratification, searching window size (sws) and data variation (CV) on the relative absolute mean error (RAME) of predictions of mud content in the north region. The data were analysed using a generalised linear model with a quasi family, a log link and a variance being mean squared in R (R Development Core Team, 2007). The highlighted are the significant higher order interactions.

	Df	Deviance	Resid. Df	Resid. Dev	F-value	p-value
NULL			639	16.16		
method	31	8.41	608	7.76	1531.11	0.0000
sample.density	1	0.67	607	7.09	3777.34	0.0000
stratification	1	0.61	606	6.47	3453.17	0.0000
sws	1	0.00	605	6.47	20.58	0.0000
CV	1	0.07	604	6.41	368.23	0.0000
method:sample.density	31	0.23	573	6.18	41.00	0.0000
method:stratification	31	1.13	542	5.05	206.23	0.0000
method:sws	31	4.43	511	0.62	807.46	0.0000
method:CV	31	0.03	480	0.59	5.41	0.0000
sample.density:stratification	1	0.06	479	0.52	346.98	0.0000
sample.density:sws	1	0.03	478	0.49	175.52	0.0000
sample.density:CV	1	0.18	477	0.32	997.76	0.0000
stratification:sws	1	0.00	476	0.32	1.13	0.2883
stratification:CV	1	0.00	475	0.32	1.26	0.2626
sws:CV	1	0.00	474	0.32	0.65	0.4224
method:sample.density:stratification	**31**	**0.12**	**443**	**0.20**	**21.08**	**0.0000**
method:sample.density:sws	**31**	**0.09**	**412**	**0.11**	**16.87**	**0.0000**
method:sample.density:CV	**31**	**0.02**	**381**	**0.08**	**4.53**	**0.0000**
method:stratification:sws	31	0.00	350	0.08	0.55	0.9761
method:stratification:CV	**31**	**0.01**	**319**	**0.07**	**2.25**	**0.0003**
method:sws:CV	**31**	**0.02**	**288**	**0.05**	**3.03**	**0.0000**
sample.density:stratification:sws	1	0.00	287	0.05	0.05	0.8253
sample.density:stratification:CV	1	0.00	286	0.05	1.27	0.2604
sample.density:sws:CV	1	0.00	285	0.05	1.05	0.3058
stratification:sws:CV	1	0.00	284	0.05	0.14	0.7118

Table 4.3. Effects of method, sample density, stratification, searching window size (sws) and data variation (CV) on the relative absolute mean error (RAME) of predictions of mud content in the northeast region. The data were analysed using a generalised linear model with a quasi family, a log link and a variance being mean squared in R (R Development Core Team, 2007). The highlighted are the significant higher order interactions.

	Df	Deviance	Resid. Df	Resid. Dev	F-value	p-value
NULL			539	15.89		
method	26	7.63	513	8.26	2587.96	0.0000
sample.density	1	1.31	512	6.95	11546.66	0.0000
stratification	1	0.21	511	6.74	1847.71	0.0000
sws	1	0.01	510	6.73	94.05	0.0000
CV	1	0.10	509	6.63	855.02	0.0000
method:sample.density	26	0.33	483	6.30	111.93	0.0000
method:stratification	26	0.59	457	5.71	201.09	0.0000
method:sws	26	5.20	431	0.51	1762.69	0.0000
method:CV	26	0.03	405	0.48	9.14	0.0000
sample.density:stratification	1	0.01	404	0.47	115.04	0.0000
sample.density:sws	1	0.03	403	0.44	279.61	0.0000
sample.density:CV	1	0.20	402	0.24	1778.37	0.0000
stratification:sws	1	0.00	401	0.24	7.15	0.0084
stratification:CV	1	0.00	400	0.24	0.93	0.3371
sws:CV	1	0.00	399	0.23	9.96	0.0020
method:sample.density:stratification	**26**	**0.11**	**373**	**0.12**	**37.03**	**0.0000**
method:sample.density:sws	26	0.06	347	0.07	18.88	0.0000
method:sample.density:CV	26	0.02	321	0.05	5.72	0.0000
method:stratification:sws	**26**	**0.01**	**295**	**0.05**	**2.42**	**0.0006**
method:stratification:CV	**26**	**0.00**	**269**	**0.04**	**1.58**	**0.0485**
method:sws:CV	26	0.01	243	0.03	2.56	0.0003
sample.density:stratification:sws	**1**	**0.00**	**242**	**0.03**	**4.98**	**0.0273**
sample.density:stratification:CV	**1**	**0.00**	**241**	**0.03**	**28.82**	**0.0000**
sample.density:sws:CV	1	0.00	240	0.02	42.97	0.0000
stratification:sws:CV	1	0.00	239	0.02	1.71	0.1930
method:sample.density:stratification:sws	26	0.00	213	0.02	0.32	0.9994
method:sample.density:stratification:CV	26	0.00	187	0.02	0.84	0.6917
method:sample.density:sws:CV	**26**	**0.00**	**161**	**0.02**	**1.61**	**0.0435**
method:stratification:sws:CV	26	0.00	135	0.02	0.16	1.0000
sample.density:stratification:sws:CV	1	0.00	134	0.02	2.00	0.1594

Table 4.4. Effects of method, sample density, stratification, searching window size (sws) and data variation (CV) on the relative absolute mean error (RAME) of predictions of mud content in the southwest region. The data were analysed using a generalised linear model with a quasi family, a log link and a variance being mean squared in R (R Development Core Team, 2007). The highlighted are the significant higher order interactions.

	Df	Deviance	Resid. Df	Resid. Dev	F-value	p-value
NULL			539	105.40		
method	26	47.32	513	58.08	295.50	0.0000
sample.density	1	5.14	512	52.95	833.78	0.0000
stratification	1	13.32	511	39.63	2162.74	0.0000
sws	1	0.40	510	39.23	64.26	0.0000
CV	1	0.04	509	39.19	6.06	0.0145
method:sample.density	26	2.88	483	36.31	17.99	0.0000
method:stratification	26	27.79	457	8.53	173.52	0.0000
method:sws	26	2.36	431	6.17	14.72	0.0000
method:CV	**26**	**0.53**	**405**	**5.64**	**3.33**	**0.0000**
sample.density:stratification	1	0.30	404	5.33	49.38	0.0000
sample.density:sws	1	0.03	403	5.30	5.42	0.0208
sample.density:CV	1	0.02	402	5.29	2.50	0.1150
stratification:sws	1	0.09	401	5.20	13.92	0.0002
stratification:CV	**1**	**0.43**	**400**	**4.77**	**69.08**	**0.0000**
sws:CV	1	0.01	399	4.76	1.63	0.2029
method:sample.density:stratification	**26**	**1.98**	**373**	**2.78**	**12.37**	**0.0000**
method:sample.density:sws	**26**	**0.46**	**347**	**2.33**	**2.86**	**0.0000**
method:sample.density:CV	26	0.22	321	2.10	1.40	0.0997
method:stratification:sws	**26**	**0.33**	**295**	**1.77**	**2.05**	**0.0028**
method:stratification:CV	26	0.19	269	1.58	1.21	0.2237
method:sws:CV	26	0.06	243	1.52	0.39	0.9970
sample.density:stratification:sws	1	0.01	242	1.50	2.22	0.1373
sample.density:stratification:CV	1	0.01	241	1.49	1.61	0.2061
sample.density:sws:CV	1	0.00	240	1.49	0.08	0.7824
stratification:sws:CV	1	0.01	239	1.48	1.66	0.1993

4.2. Effects of methods and the interaction with searching window size and sample stratification

Since statistical methods interact with the other experimental factors (Tables 4.2-4.4), we compare the effects of the methods and the interactive effects with searching window size and sample stratification for datasets with the highest sample density (100%) in each of the three regions (Figs. 4.-4.3).

4.2.1. North region

In the north region, the accuracy of the methods varies with sample stratification and searching widow size (Fig. 4.1). The accuracy of the statistical methods is generally higher when samples are not stratified and the searching neighbourhood is local (*i.e.*, searching window size is small). Sample stratification reduces the accuracy of RKrf considerably, while OK, IDW1-IDW4, and OCK1 to OCK4 are less affected. Global neighbourhood searching significantly reduces the accuracy of IDW1, GLS1 and

GLS2. IDW2 perform better than OK when the searching neighbourhood is local, and vice versa when it is global.

The inclusion of latitude and longitude up to third-order polynomial (i.e., the terms in the Legendre and Legendre's (1998) equation; in this study we refer them as "latitude and logitude terms") as secondary information significantly reduces the accuracy of the predictions of all methods including, KED4, RKlm5, RKglm5 and RKgls5. UK also performs poorly. KED3 that uses the quadratic terms of bathymetry, distance-to-coast and slope also performs quite poorly. However, RKrf performs very well when samples are not stratified although it uses latitude and logitude terms and uses the quadratic terms of bathymetry, distance-to-coast and slope.

As the effects of the methods interactes with the effects of all the other experimental factors, no single method is expected to perform better than other methods in all scenarios in the north region (Fig. 4.1). Despite this, the most accurate method is RKrf when the samples are not stratified and the searching neighbourhood is local. The next most accurate method is RKrf when samples are not stratified and the searching neighbourhood is global. IDW2 and KED1 perform equally well, with a slightly reduction in accuracy when the searching neighbourhood is local. All the remaining methods perform more poorly than the control (*i.e.*, IDW2 for dataset with the 100% sample density, without sample stratification and local searching neighbourhood).

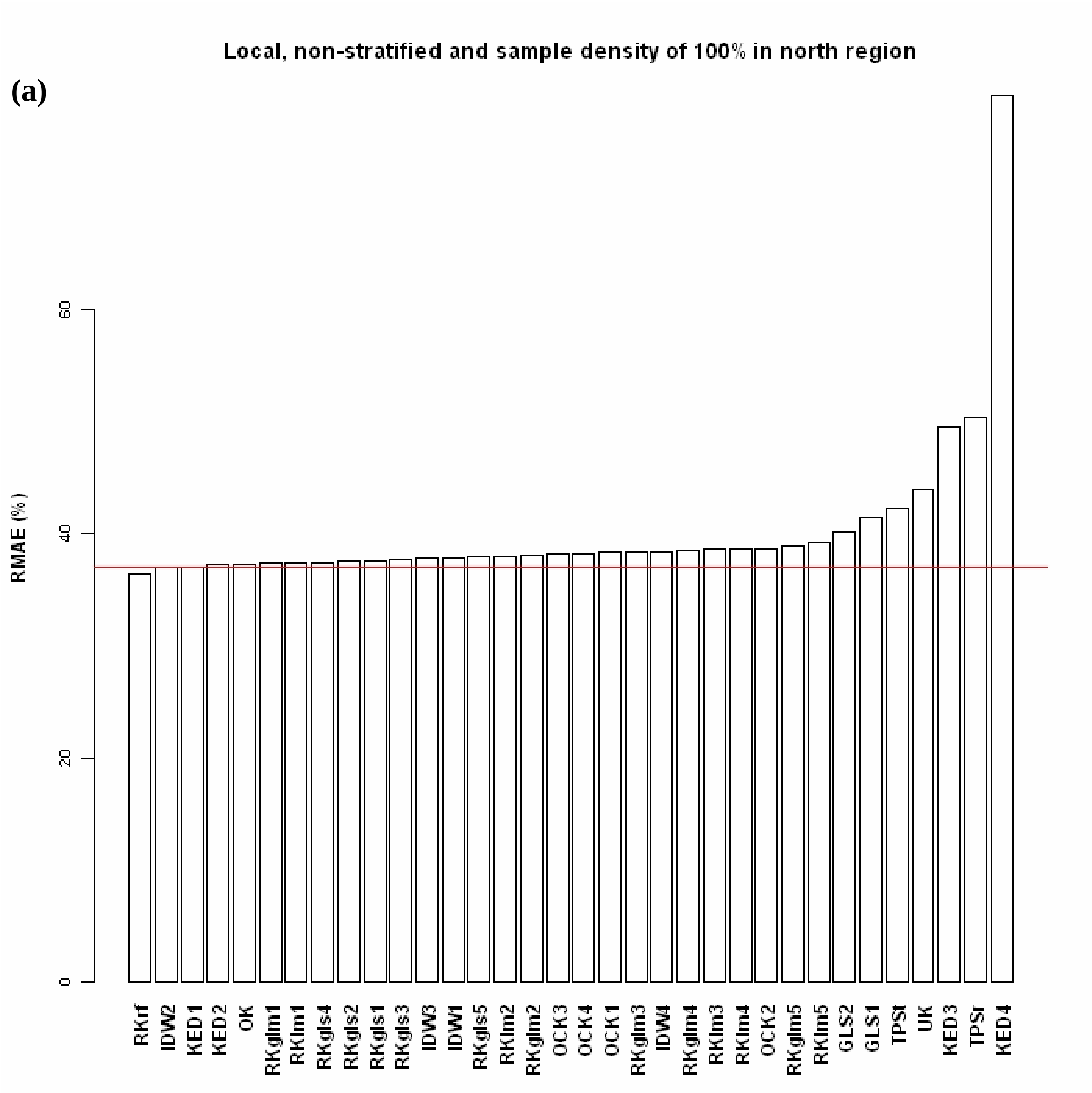


Figure 4.1. The relative absolute mean error (RMAE (%)) of statistical methods for dataset with the 100% sample density in the north region: (a) local and non-stratified, (b) local and stratified; (c) global and non-stratified and (d) global and stratified. The lower the bar the more accurate the method. Horizontal line indicates the accuracy of the control (IDW2).

Fig. 4.1. (cont.):
(b)

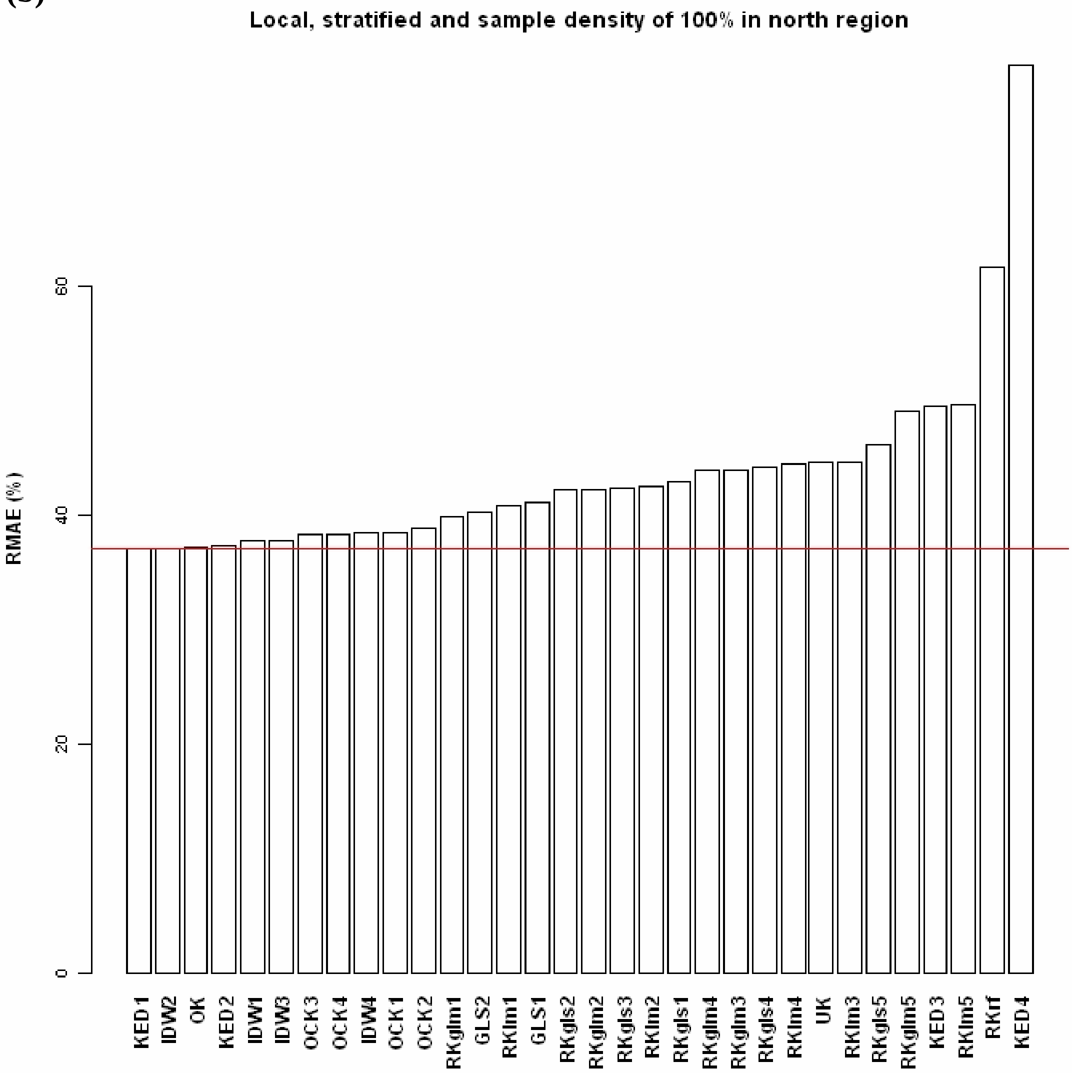

Local, stratified and sample density of 100% in north region
RMAE (%)
0
20
40
60
KED1
IDW2
OK
KED2
IDW1
IDW3
OCK3
OCK4
IDW4
OCK1
OCK2
RKglm1
GLS2
RKlm1
GLS1
RKgls2
RKglm2
RKgls3
RKlm2
RKgls1
RKglm4
RKglm3
RKgls4
RKlm4
UK
RKlm3
RKgls5
RKglm5
KED3
RKlm5
RKrf
KED4

Fig. 4.1. (cont.):

(c)

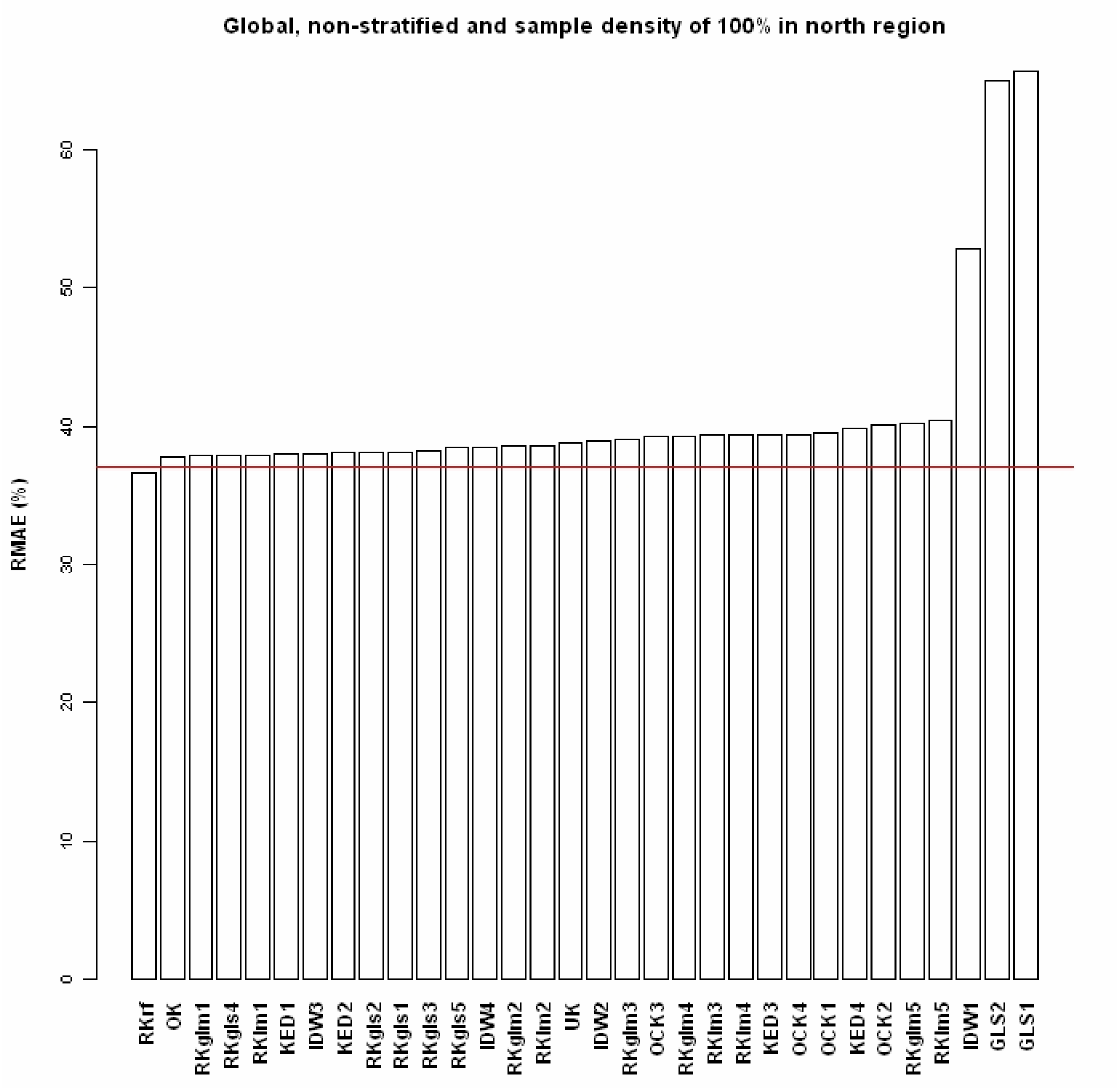
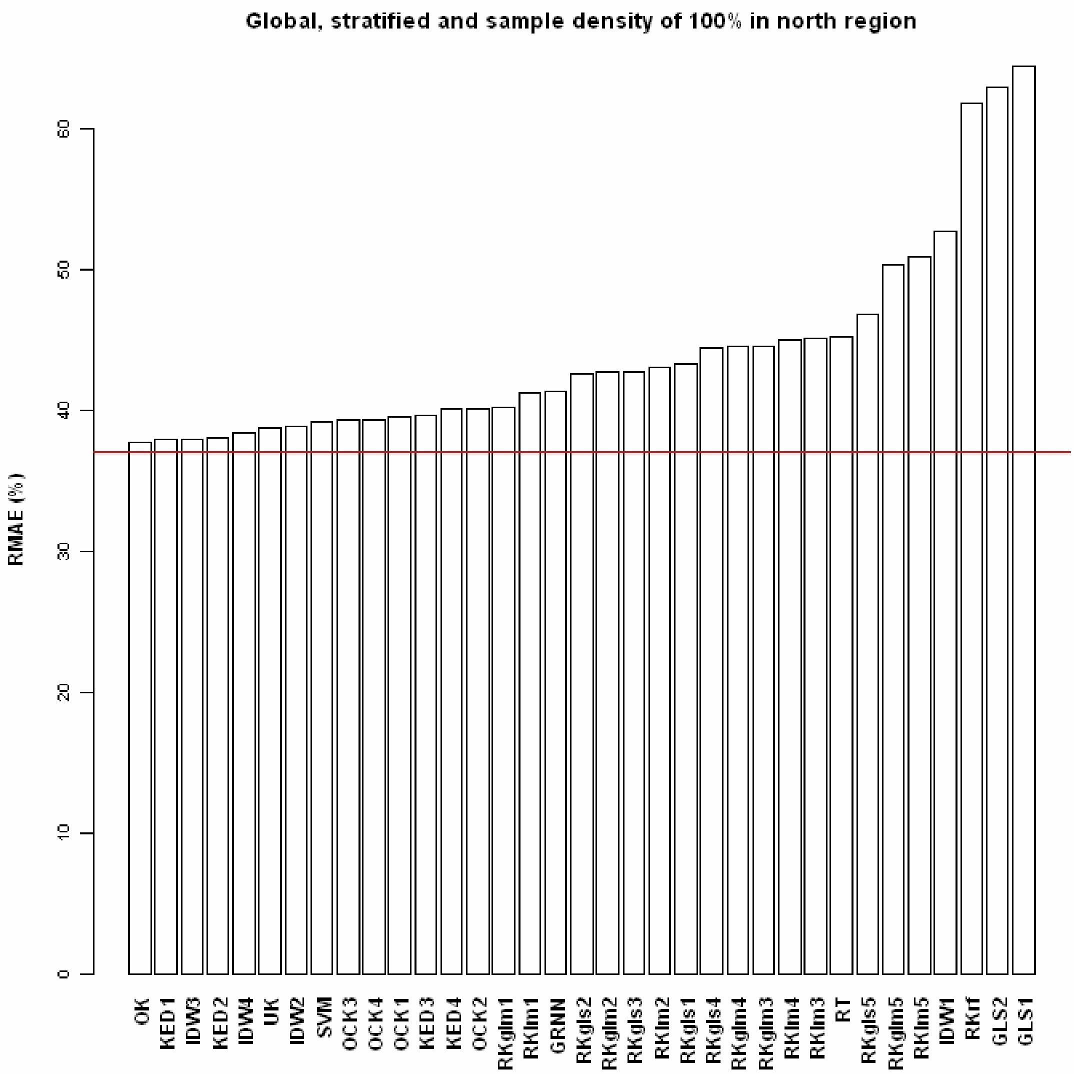

Global, non-stratified and sample density of 100% in north region
RMAE (%)
0
10
20
30
40
50
60
RKrf
OK
RKglm1
RKgls4
RKlm1
KED1
IDW3
KED2
RKgls2
RKgls1
RKgls3
RKgls5
IDW4
RKglm2
RKlm2
UK
IDW2
RKglm3
OCK3
RKglm4
RKlm3
RKlm4
KED3
OCK4
OCK1
KED4
OCK2
RKglm5
RKlm5
IDW1
GLS2
GLS1

Fig. 4.1. (cont.):

(d)


Global, stratified and sample density of 100% in north region
RMAE (%)
0
10
20
30
40
50
60
OK
KED1
IDW3
KED2
IDW4
UK
IDW2
SVM
OCK3
OCK4
OCK1
KED3
KED4
OCK2
RKglm1
RKlm1
GRNN
RKgls2
RKglm2
RKgls3
RKlm2
RKgls1
RKgls4
RKglm4
RKglm3
RKlm4
RKlm3
RT
RKgls5
RKglm5
RKlm5
IDW1
RKrf
GLS2
GLS1

4.2.2. Northeast region

In the northeast region, the accuracy of the statistical methods also depends on sample stratification and searching widow size (Fig. 4.2). Results are similar to those observed in the north region. Prediction accuracy is higher when samples are not stratified and the searching neighbourhood is local. Sample stratification also considerably reduces the accuracy of RKrf, while OK, IDW1-IDW4, OCK1 to OCK4 are less sensitive, and IDW2-IDW4 even perform slightly better. Global searching neighbourhood significantly reduces the accuracy of IDW1, GLS1 and GLS2. IDW2, IDW3 and IDW4 perform better than OK regardless of sample stratification and searching widow size.

When the latitude and logitude terms are used as secondary information, it significantly reduces the prediction accuracy of all methods including KED4, RKlm5, RKglm5 and RKgls5 as in the north region. UK is also performed poorly. KED3, which uses the quadratic terms of bathymetry, distance-to-coast and slope, also performs poorly. However, as in the north region, RKrf performs well when sample stratification is not applied, although it uses the latitude and logitude terms and the quadratic terms of bathymetry, distance-to-coast and slope.

Due to the interactive effects of the methods with the other experimental factors, no single method is expected to perform better than other methods in all scenarios in the northeast region (Fig. 4.2). The most accurate method is IDW2 when samples are stratified and the searching neighbourhood is local; and IDW3 and IDW4 are slightly less accurate. All the remaining methods perform more poorly than the control. Although RKrf (RMAE=37.15) is slightly less accurate, besides IDW, it is the next most accurate method when samples are not stratified and the searching neighbourhood is local.

(a)

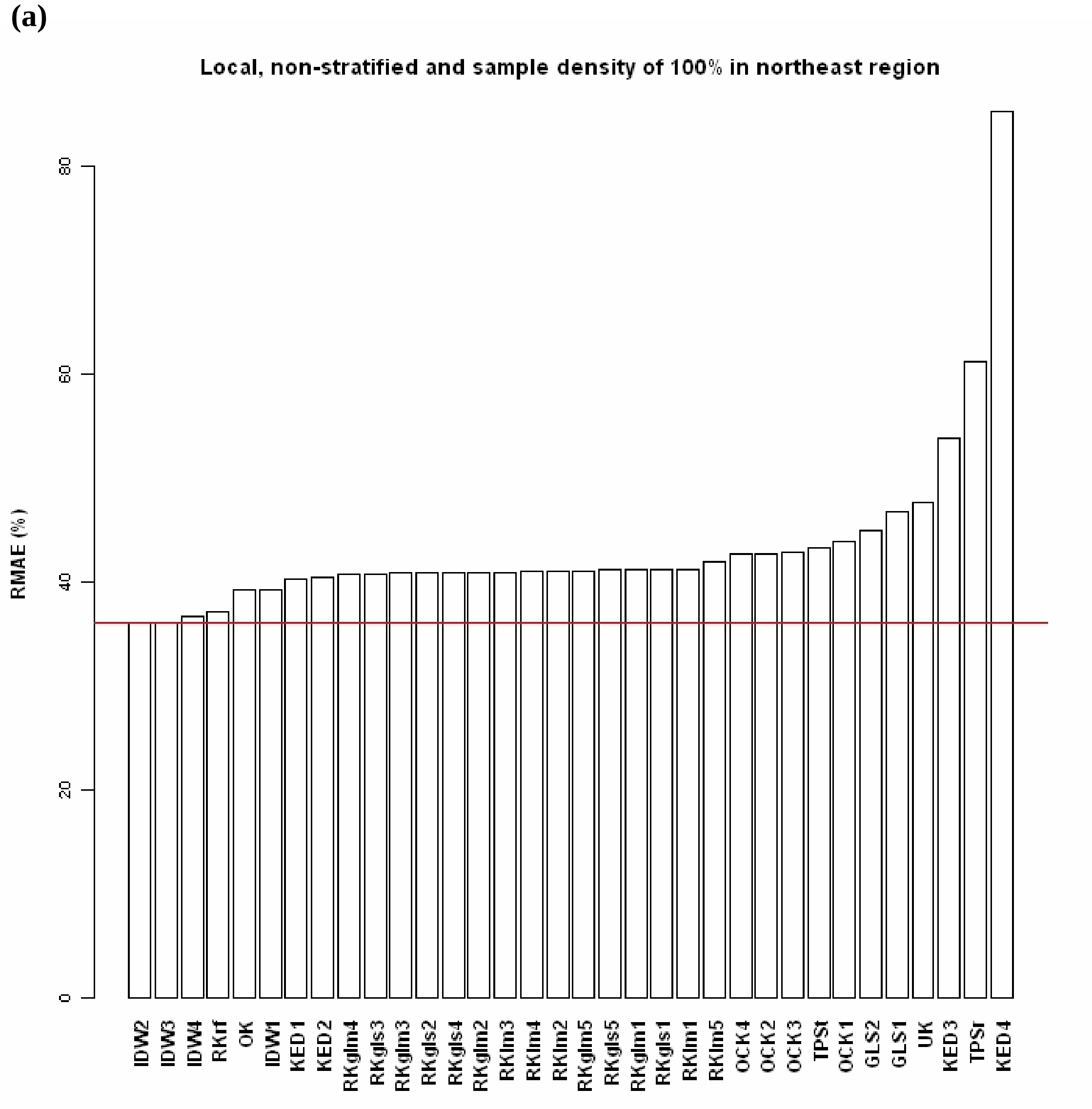


Figure 4.2. The relative absolute mean error (RMAE (%)) of statistical methods for dataset with the 100% sample density in the northeast region: (a) local and non-stratified, (b) local and stratified; (c) global and non-stratified and (d) global and stratified. The lower the bar the more accurate the method. Horizontal line indicates the accuracy of the control (IDW2).

Fig. 4.2. (cont.):

(b)

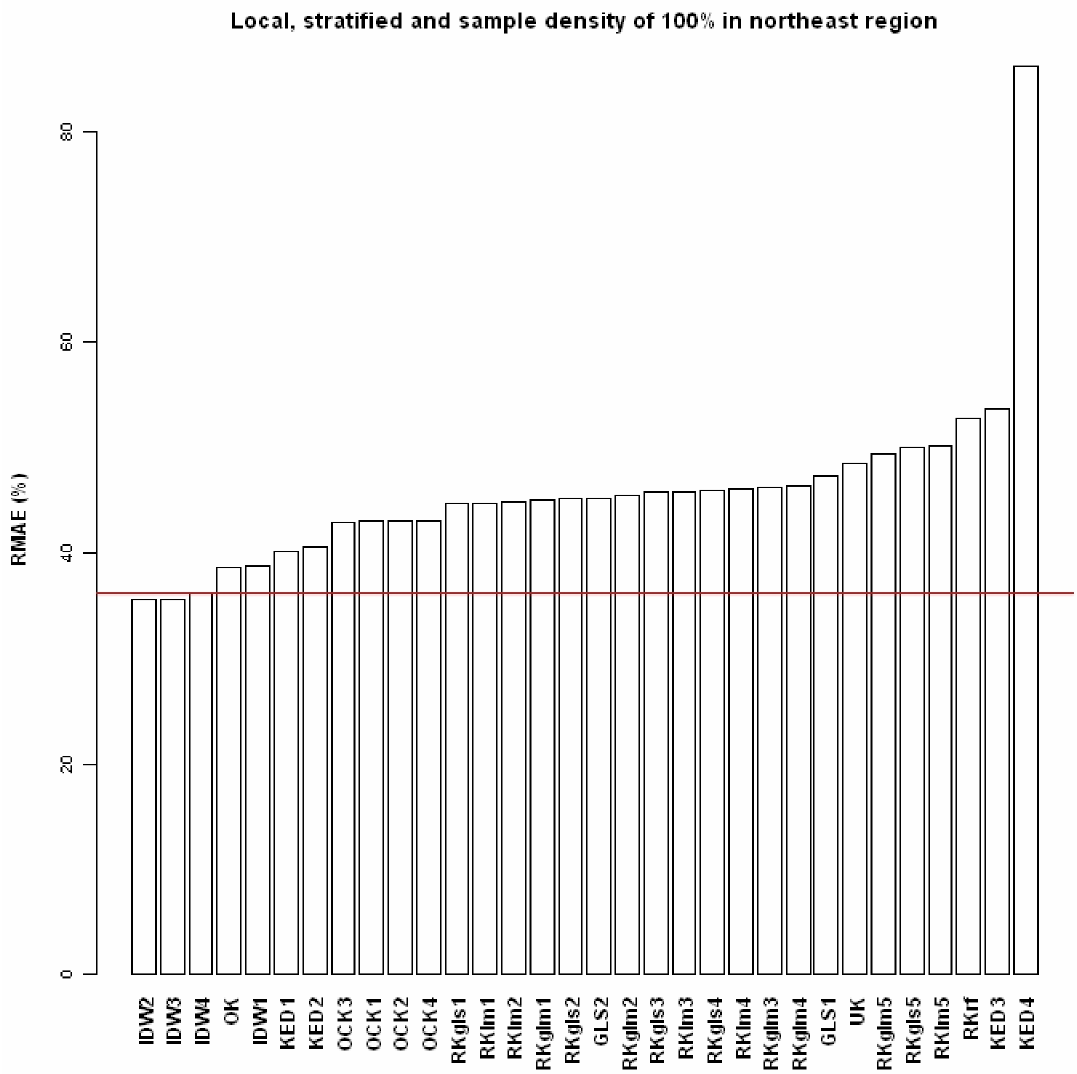

Local, stratified and sample density of 100% in northeast region
RMAE (%)
0
20
40
60
80
IDW2
IDW3
IDW4
OK
IDW1
KED1
KED2
OCK3
OCK1
OCK2
OCK4
RKgls1
RKlm1
RKlm2
RKglm1
RKgls2
GLS2
RKglm2
RKgls3
RKlm3
RKgls4
RKlm4
RKglm3
RKglm4
GLS1
UK
RKglm5
RKgls5
RKlm5
RKrf
KED3
KED4

Fig. 4.2. (cont.):

(c)

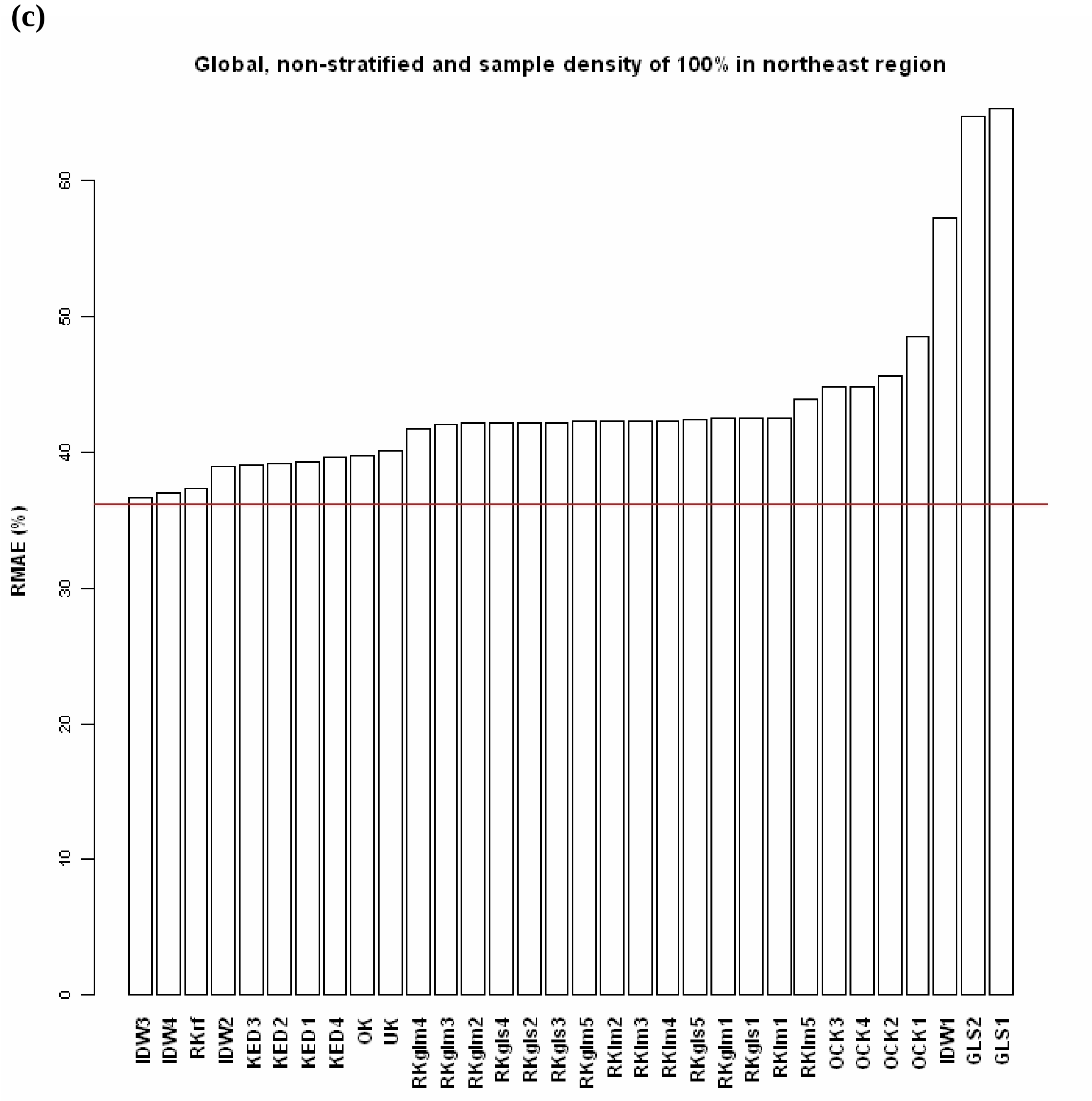
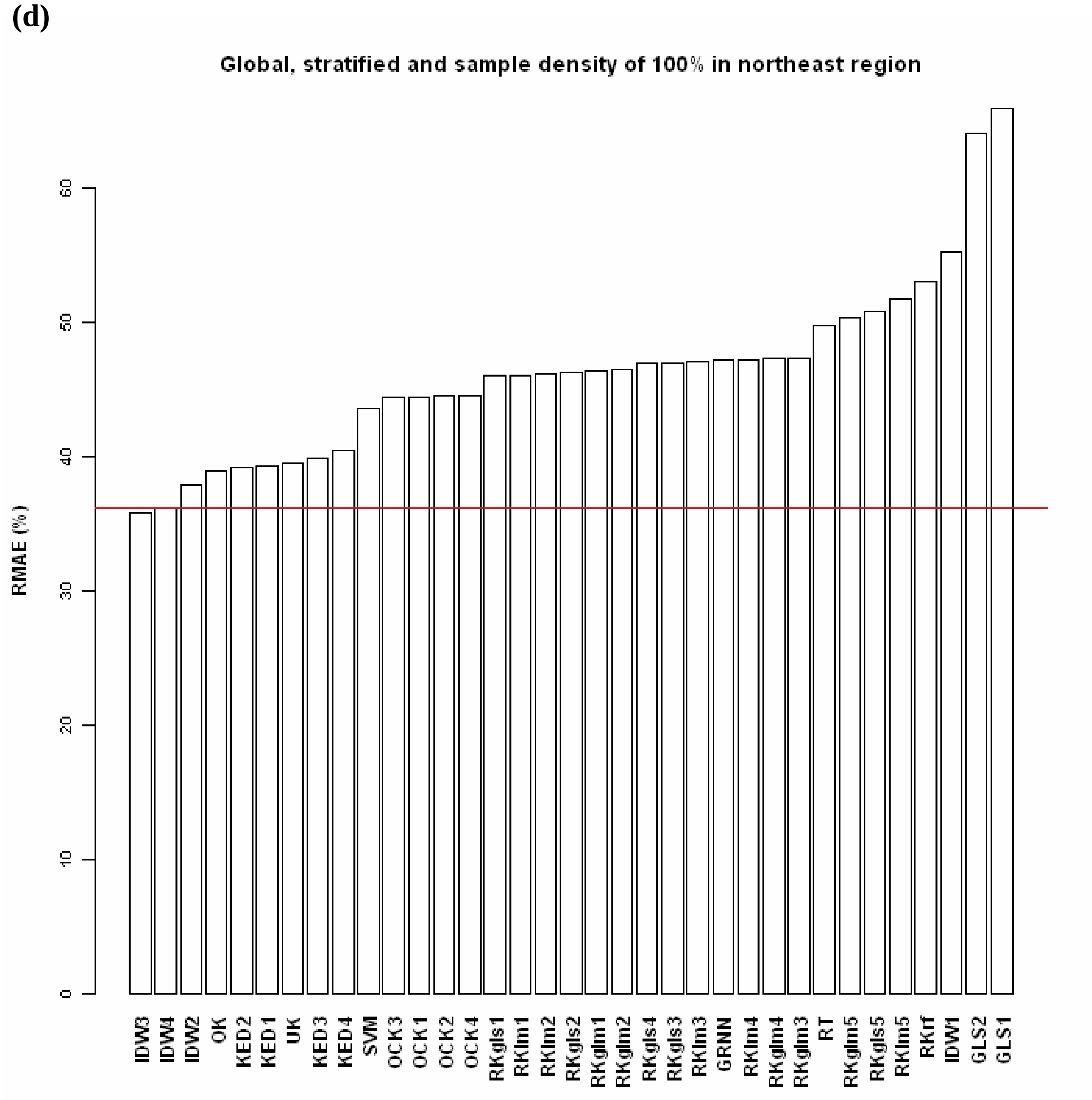

Global, non-stratified and sample density of 100% in northeast region
RMAE (%)
0
10
20
30
40
50
60
IDW3
IDW4
RKrf
IDW2
KED3
KED2
KED1
KED4
OK
UK
RKglm4
RKglm3
RKglm2
RKgls4
RKgls2
RKgls3
RKglm5
RKlm2
RKlm3
RKlm4
RKgls5
RKglm1
RKgls1
RKlm1
RKlm5
OCK3
OCK4
OCK2
OCK1
IDW1
GLS2
GLS1

Fig. 4.2. (cont.):

(d)


Global, stratified and sample density of 100% in northeast region
RMAE (%)
0
10
20
30
40
50
60
IDW3
IDW4
IDW2
OK
KED2
KED1
UK
KED3
KED4
SVM
OCK3
OCK1
OCK2
OCK4
RKgls1
RKlm1
RKlm2
RKgls2
RKglm1
RKglm2
RKgls4
RKgls3
RKlm3
GRNN
RKlm4
RKglm4
RKglm3
RT
RKglm5
RKgls5
RKlm5
RKrf
IDW1
GLS2
GLS1

4.2.3. Southwest region

In the southwest region, the accuracy of statistical method depends on sample stratification and searching widow size (Fig. 4.3). Results are similar to those seen in the north and northeast regions. Prediction accuracy is also generally higher when samples are not stratified and the searching neighbourhood is local in the southwest region (Fig. 4.3). Sample stratification also considerably reduces the accuracy of RKrf, while the accuracy of OK, and OCK-OCK4 is slightly improved by sample stratification. Global neighbourhood searching significantly reduces the accuracy of IDW1, GLS1 and GLS2. IDW3 and IDW4 perform better than OK regardless of the stratification and searching widow size, while IDW2 only performs better than OK when samples are not stratified and the searching neighbourhood is local.

Inclusion of the latitude and logitude terms as secondary information significantly reduces the prediction accuracy of KED4, RKlm5, RKglm5 and RKgls5 when samples are stratified as in other two regions. UK always performs poorly. KED3, which uses the quadratic terms of bathymetry, distance-to-coast and slope, also performs quite poorly. However, RKrf performs well when sample stratification is not applied, although it uses the latitude and logitude terms and the quadratic terms of bathymetry, distance-to-coast and slope.

Due to the interactive effects of the methods with the other experimental factors, no single method is expected to perform better than other methods in all scenarios in the southwest region (Fig. 4.3). The most accurate method is RKrf when samples are not stratified and the searching neighbourhood is local. RKrf, IDW4 and IDW3 are also more accurate than the control when samples are not stratified and the searching neighbourhood is local. OK also performs slightly better than the control when the samples are stratified. All the remaining methods perform more poorly than the control.

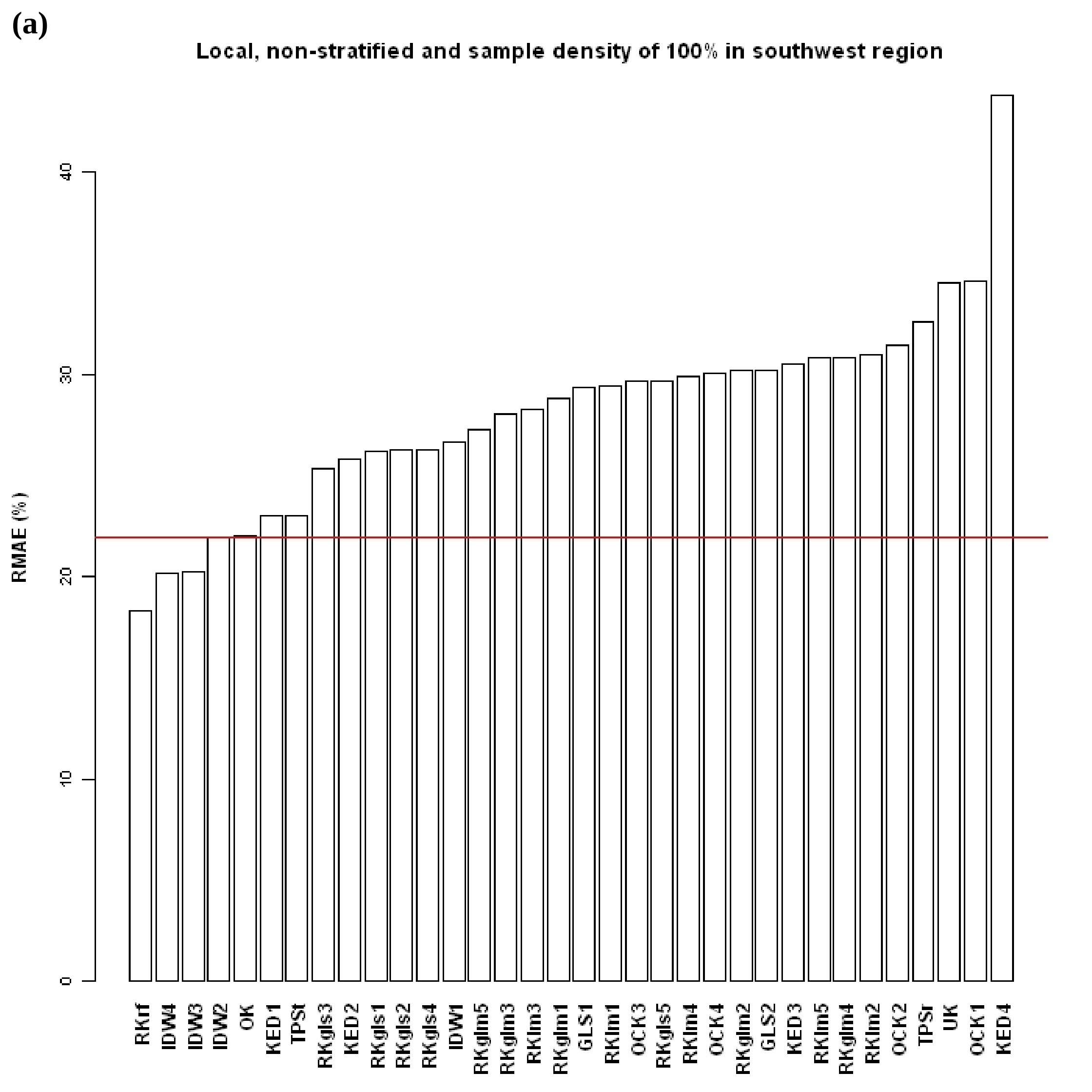


Figure 4.3. The relative absolute mean error (RMAE (%)) of the statistical methods for dataset with the 100% sample density in the southwest region: (a) local and non-stratified, (b) local and stratified; (c) global and non-stratified and (d) global and stratified. The lower the bar the more accurate the method. Horizontal line indicates the accuracy of the control (IDW2).

Fig. 4.3. (cont.):

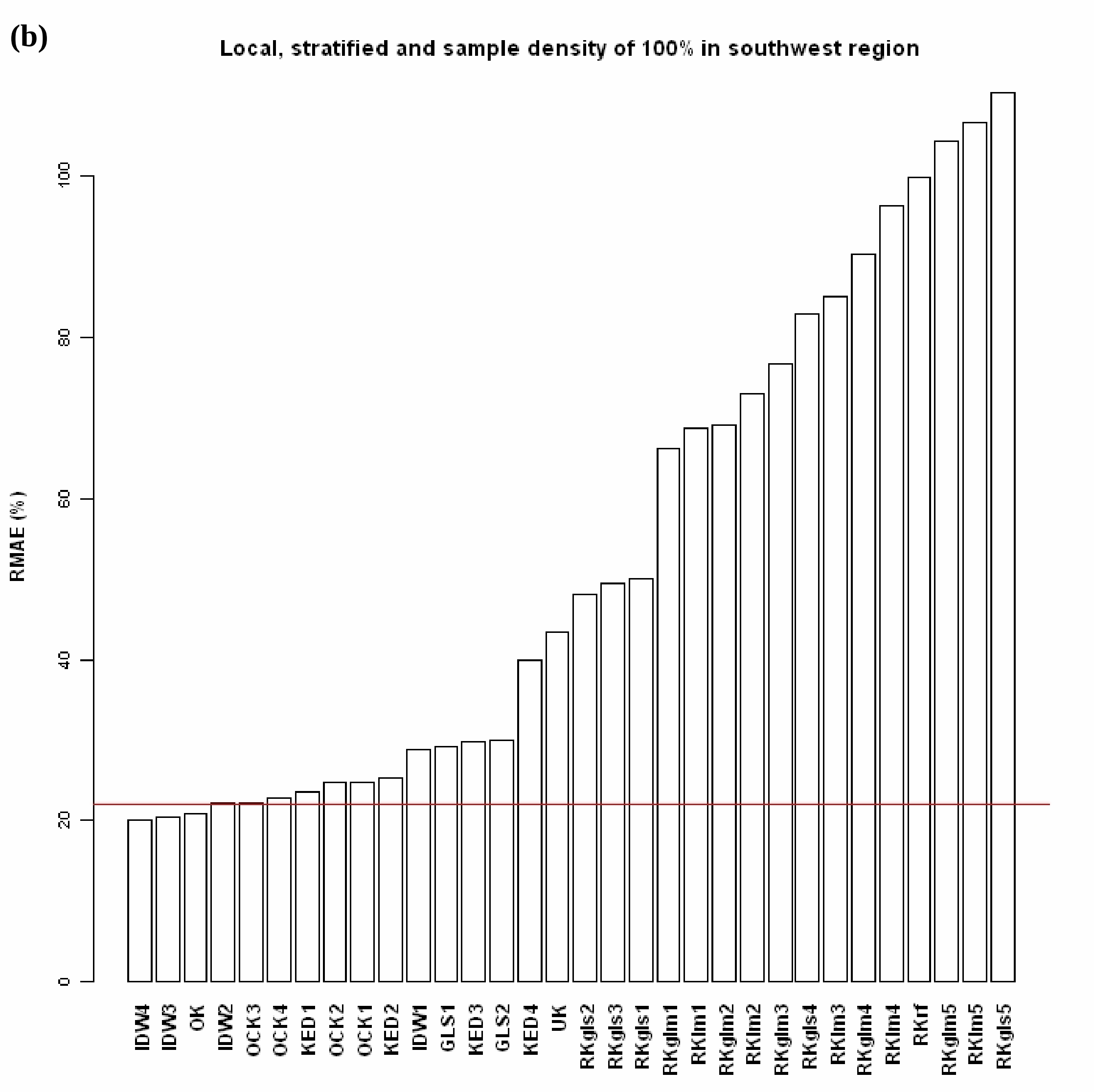

(b)
Local, stratified and sample density of 100% in southwest region
RMAE (%)
0
20
40
60
80
100
IDW4
IDW3
OK
IDW2
OCK3
OCK4
KED1
OCK2
OCK1
KED2
IDW1
GLS1
KED3
GLS2
KED4
UK
RKgls2
RKgls3
RKgls1
RKglm1
RKlm1
RKglm2
RKlm2
RKglm3
RKgls4
RKlm3
RKglm4
RKlm4
RKrf
RKglm5
RKlm5
RKgls5

Fig. 4.3. (cont.):

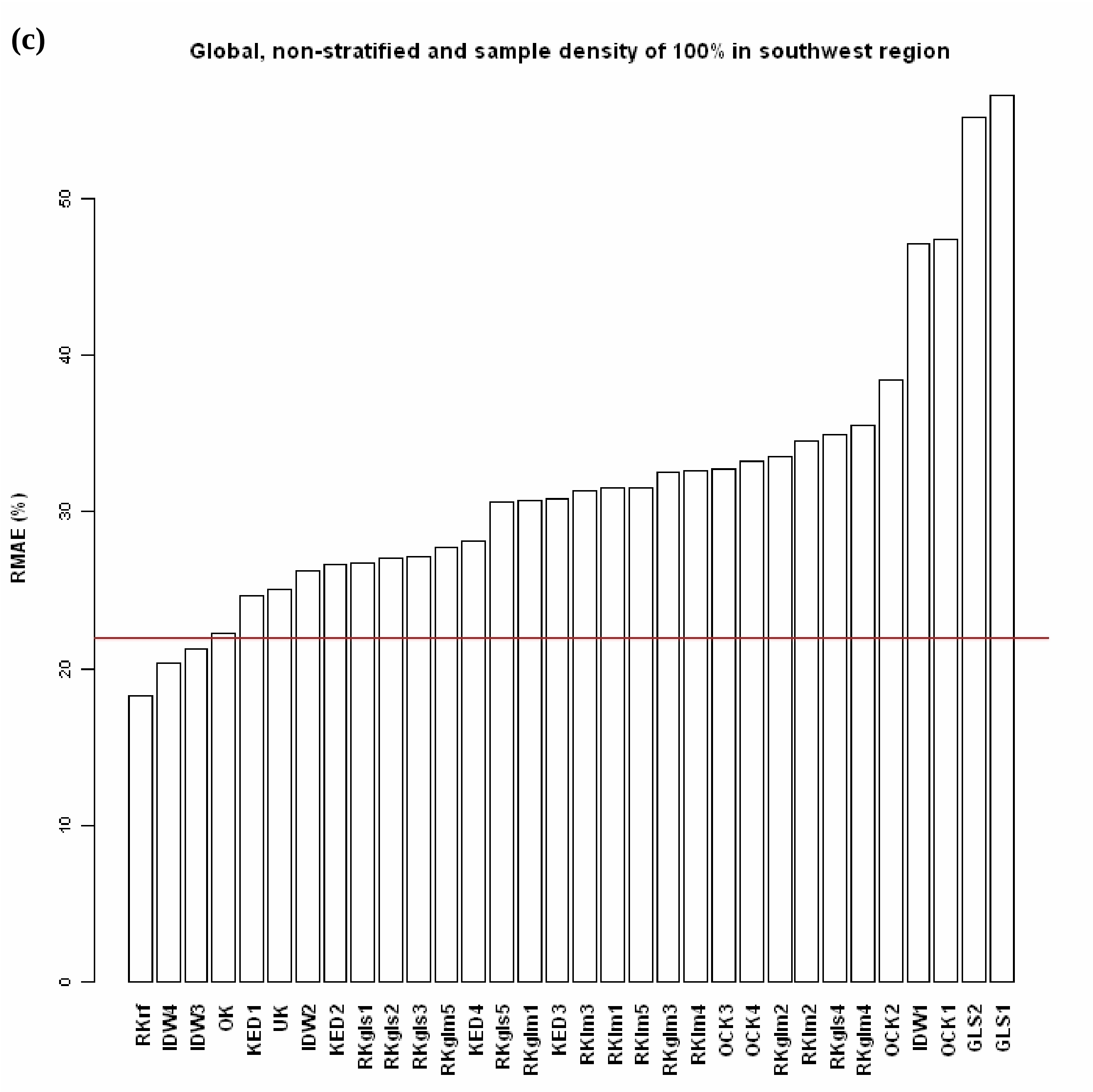
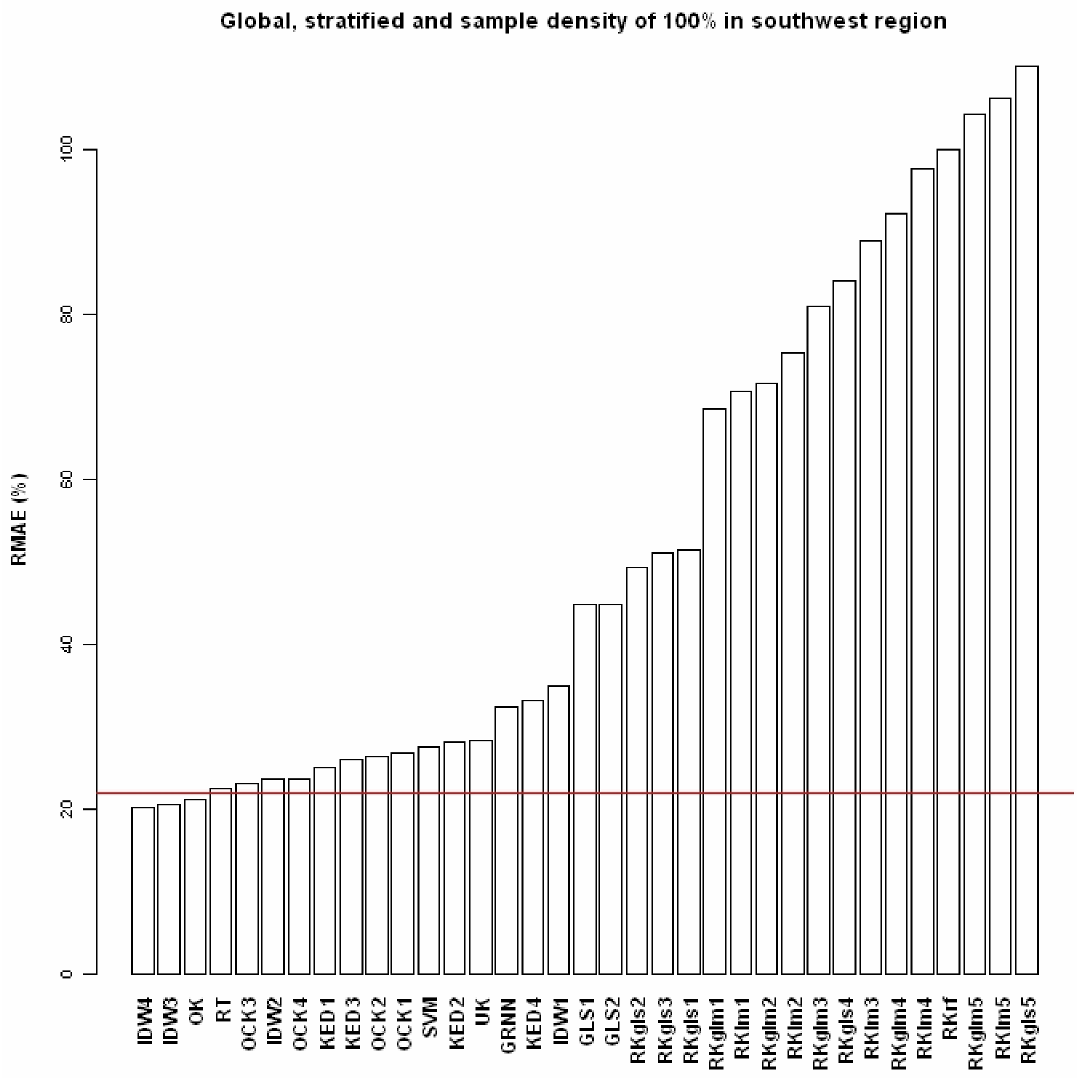

(c)
Global, non-stratified and sample density of 100% in southwest region
RMAE (%)
0
10
20
30
40
50
RKrf
IDW4
IDW3
OK
KED1
UK
IDW2
KED2
RKgls1
RKgls2
RKgls3
RKglm5
KED4
RKgls5
RKglm1
KED3
RKlm3
RKlm1
RKlm5
RKglm3
RKlm4
OCK3
OCK4
RKglm2
RKlm2
RKgls4
RKglm4
OCK2
IDW1
OCK1
GLS2
GLS1

Fig. 4.3. (cont.):

(d)



The most accurate methods and the control in each region are summarised in Table 4.5. In the north region, three sub-methods are more accurate than the control and two are equivalent to the control. In the northeast region, three sub-methods are more accurate than the control and three are equivalent to the control. While in the southwest region, twelve sub-methods are more accurate than the control.

RKrf is the most accurate method for datasets in the north and southwest regions when samples are not stratified, and the searching neighbourhood has little impacts on its performance (Table 4.5). In the northeast region the most accurate method is IDW2 when samples are stratified and the searching neighbourhood is local. Although RKrf (RMAE=37.15) is slightly less accurate, besides IDW, it is the next most accurate method when samples are not stratified and the searching neighbourhood is local (Fig. 4.2.a). Methods in red text and the control method (highlighted) are selected for further analyses, along with RKrf for the northeast region. Overall, the most accurate methods identified can reduce the prediction error by 2-17% in comparison with the control.

Table 4.5. The most accurate methods and the control in each region. Methods in red text were selected for further analyses and methods highlighted were the control in each region.

Method	Region	Searching window size	Stratification	RMAE (%)	Order
RKrf	**n**	**Local**	**no**	**36.37**	**1**
RKrf	n	Global	no	36.6	2
KED1	**n**	**Local**	**yes**	**37.02**	**3**
IDW2	n	local	yes	37.04	4
KED1	n	local	no	37.04	5
IDW2	n	local	no	37.04	6
IDW2	**ne**	**local**	**yes**	**35.55**	**1**
IDW3	ne	local	yes	35.59	2
IDW3	ne	global	yes	35.75	3
IDW4	ne	global	yes	36.15	4
IDW4	ne	local	yes	36.15	5
IDW3	ne	local	no	36.15	6
IDW2	ne	local	no	36.15	7
RKrf	**sw**	**global**	**no**	**18.28**	**1**
RKrf	sw	local	no	18.32	2
IDW4	**sw**	**local**	**yes**	**20.1**	**3**
IDW4	sw	global	yes	20.14	4
IDW4	sw	local	no	20.18	5
IDW3	sw	local	no	20.25	6
IDW4	sw	global	no	20.33	7
IDW3	sw	local	yes	20.4	8
IDW3	sw	global	yes	20.61	9
OK	sw	local	yes	20.83	10
OK	sw	global	yes	21.26	11
IDW3	sw	global	no	21.28	12
IDW2	sw	local	no	21.93	13

4.3. Effects of sample density

Sample density interacted with method, stratification and searching window size (Tables 4.2-4.4). In this section we only focus on the interactive effects on the most accurate methods and the control (Table 4.5).

4.3.1. North region

In the north region, the most accurate method, RKrf with local searching neighbourhood and sample non-stratification, displays the best performance when the sample density is 100% (Fig. 4.4). A similar pattern is observed for KED1 with local searching neighbourhood and sample stratification. IDW2 with local searching neighbourhood and sample non-stratification (the control) performs slightly better when the sample density is 20% than when the sample density is 40% and 60%.

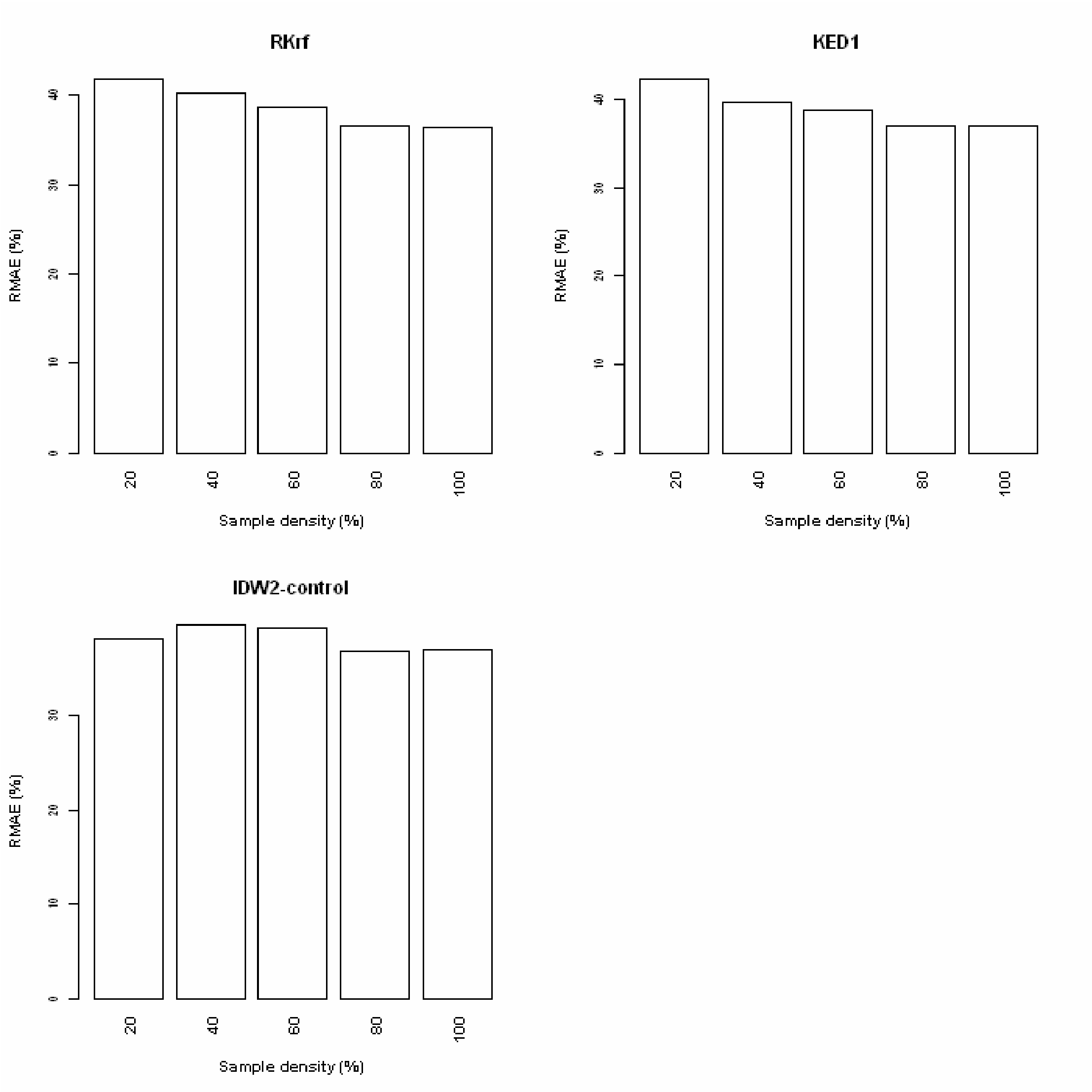


Figure 4.4. The relative absolute mean error (RMAE (%)) of the two best statistical methods and the control in relation to sample density in the north region. Rkrf: local and non-stratified; KED1: local and stratified; and IDW2 (control): local and non-stratified.

4.3.2. Northeast region

In the northeast region, the most accurate method, IDW2 with local searching neighbourhood and sample stratification, displays the best performance when the sample density is 100% (Fig. 4.5). A similar pattern is observed for RKrf with local searching neighbourhood and non-stratified samples, and for IDW2 with local searching neighbourhood and non-stratified samples (*i.e.*, the control).

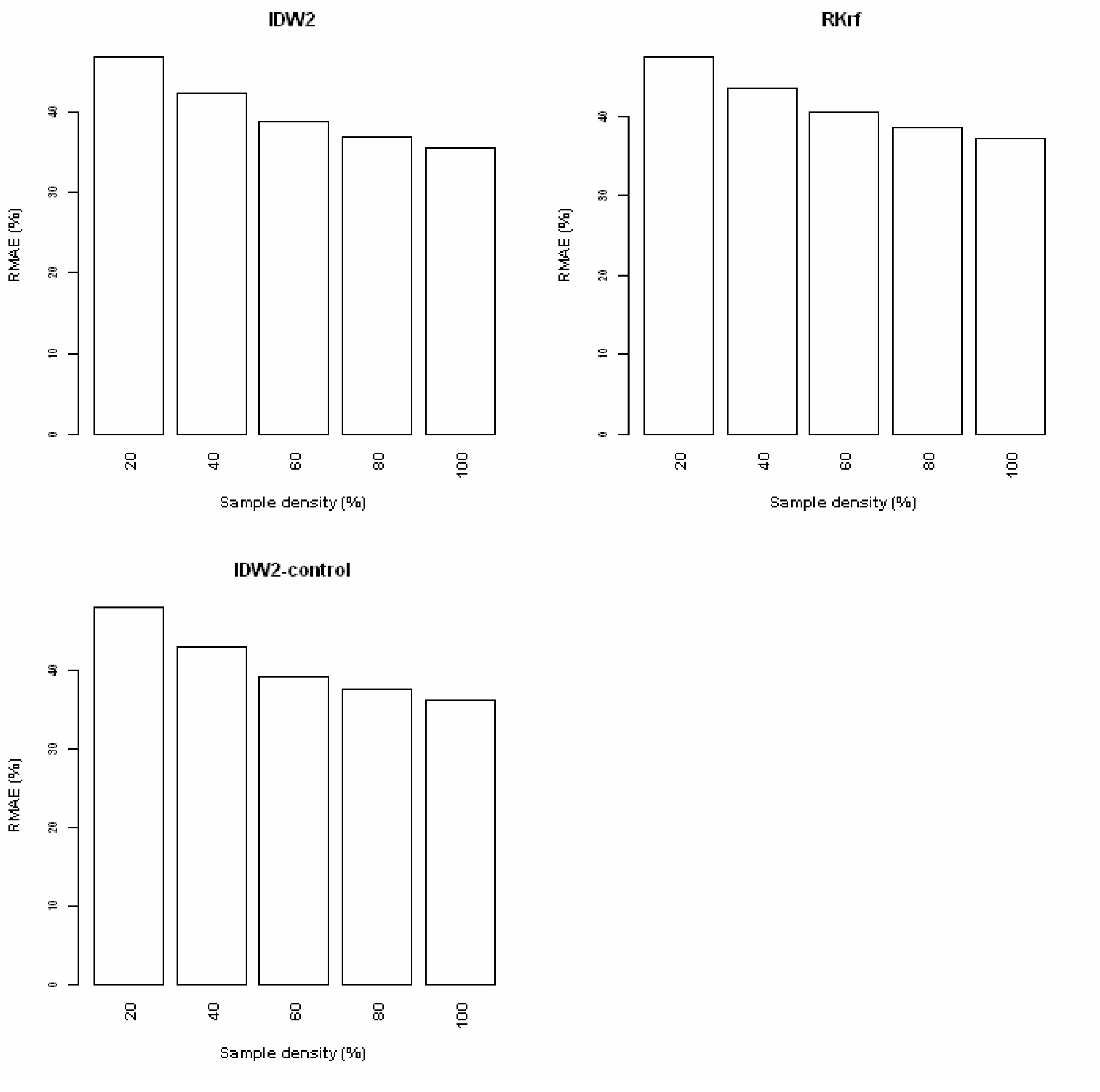


Figure 4.5. The relative absolute mean error (RMAE (%)) of the two best statistical methods and the control in relation to sample density in the northeast region. IDW2: local and stratified; RKrf: local and non-stratified; and IDW2 (control): local and non-stratified.

4.3.3. Southwest region

In the southwest region, the most accurate method, RKrf with global searching neighbourhood and non-stratified samples, displayed the best performance when the sample density was 100%, although it performed slightly better when the sample density is 40% than when it is 60% (Fig. 4.6). A similar pattern was observed for IDW4 with local searching neighbourhood and stratified samples, and IDW2 with local searching neighbourhood and non-stratified samples (*i.e.*, the control).

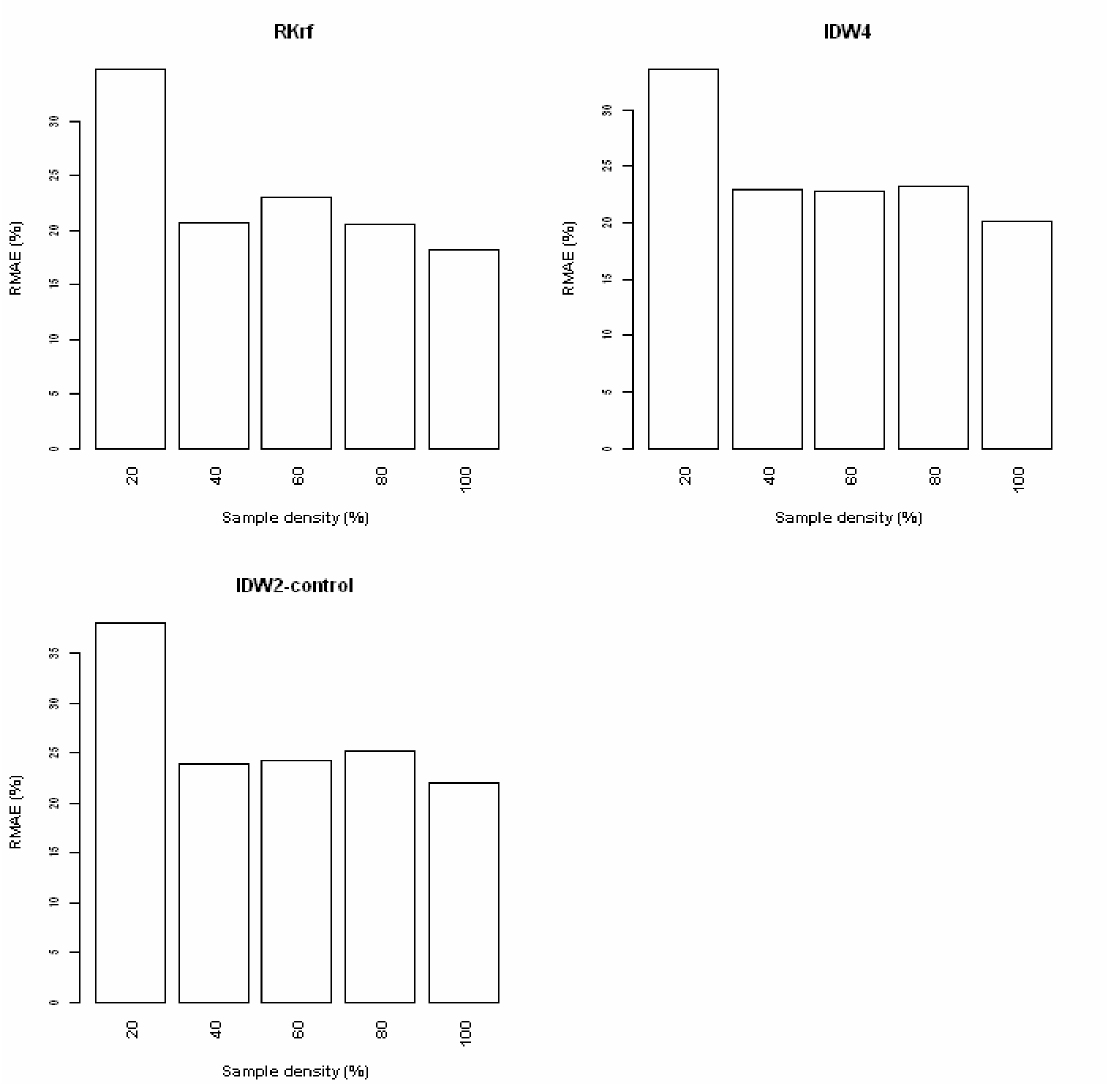


Figure 4.6. The relative absolute mean error (RMAE (%)) of the best two statistical methods and the control in relation to sample density in the southwest region. RKrf: global and non-stratified; IDW4: local and stratified; and IDW2 (control): local and non-stratified.

4.4. Data variation

Data variation interacted with method and stratification in all three study regions and with searching window size in the north and northeast regions (Tables 4.2-4.4). Due to the small size or none sample in some geomorphic provinces when samples in the model development datasets with lower densities were stratified, it is impossible to make any predictions. For some samples, geo-statistical methods may produce negative variogram range and consequently did not make predictions. Thus these methods produced prediction datasets with fewer predictions than other methods, and the mean and CV of the samples for prediction differed from those listed in Table 3.7. This may cause problems in the subsequent comparison. However, it is not the case for the methods selected. So this problem is not further addressed. In this section we focus only on the interactive effects on the selected most accurate methods and the control.

4.4.1. North region

In the north region, the accuracy of RKrf with local searching neighbourhood and non-stratified samples, the most accurate method, decreases with CV (Fig. 4.7). A similar pattern is observed for KED1 with local searching neighbourhood and stratified samples. IDW2 with local searching neighbourhood and non-stratified samples (the control) also performs better when CV is low.

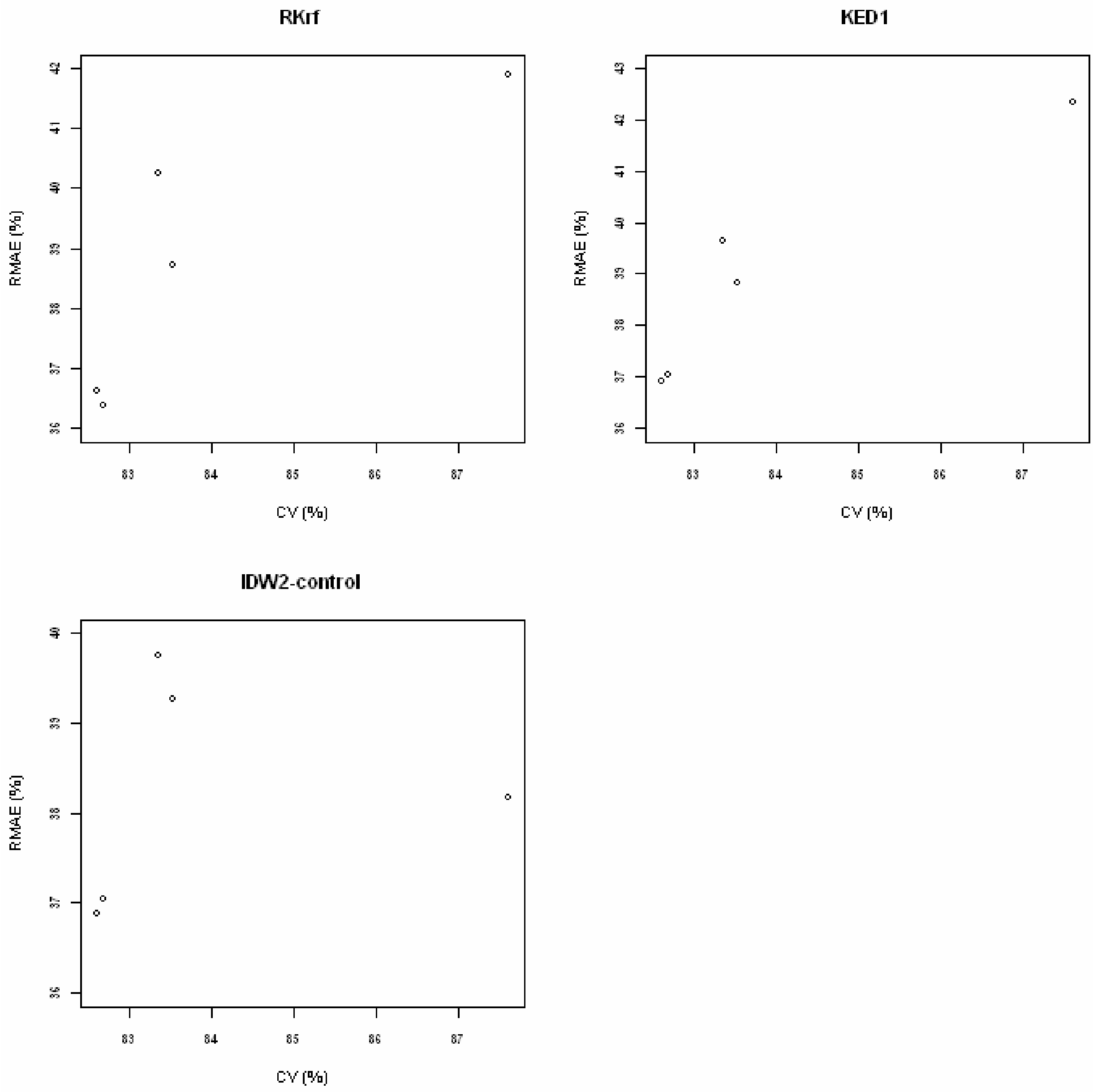


Figure 4.7. The relative absolute mean error (RMAE (%)) of the two best statistical methods and the control in relation to CV (%) in the north region. Rkrf: local and non-stratified; KED1: local and stratified; and IDW2 (control): local and non-stratified.

4.4.2. Northeast region

In the northeast region, all methods (*i.e.*, IDW2 with local searching neighbourhood and stratified samples, the most accurate method, RKrf with local searching neighbourhood and non-stratified samples, and IDW2 with local searching neighbourhood and non-stratified samples) display a similar pattern to CV (Fig. 4.8). As CV increases from about 82 to 85%, the accuracy of all methods increases, and then when CV increases further the accuracy of all methods decrease.

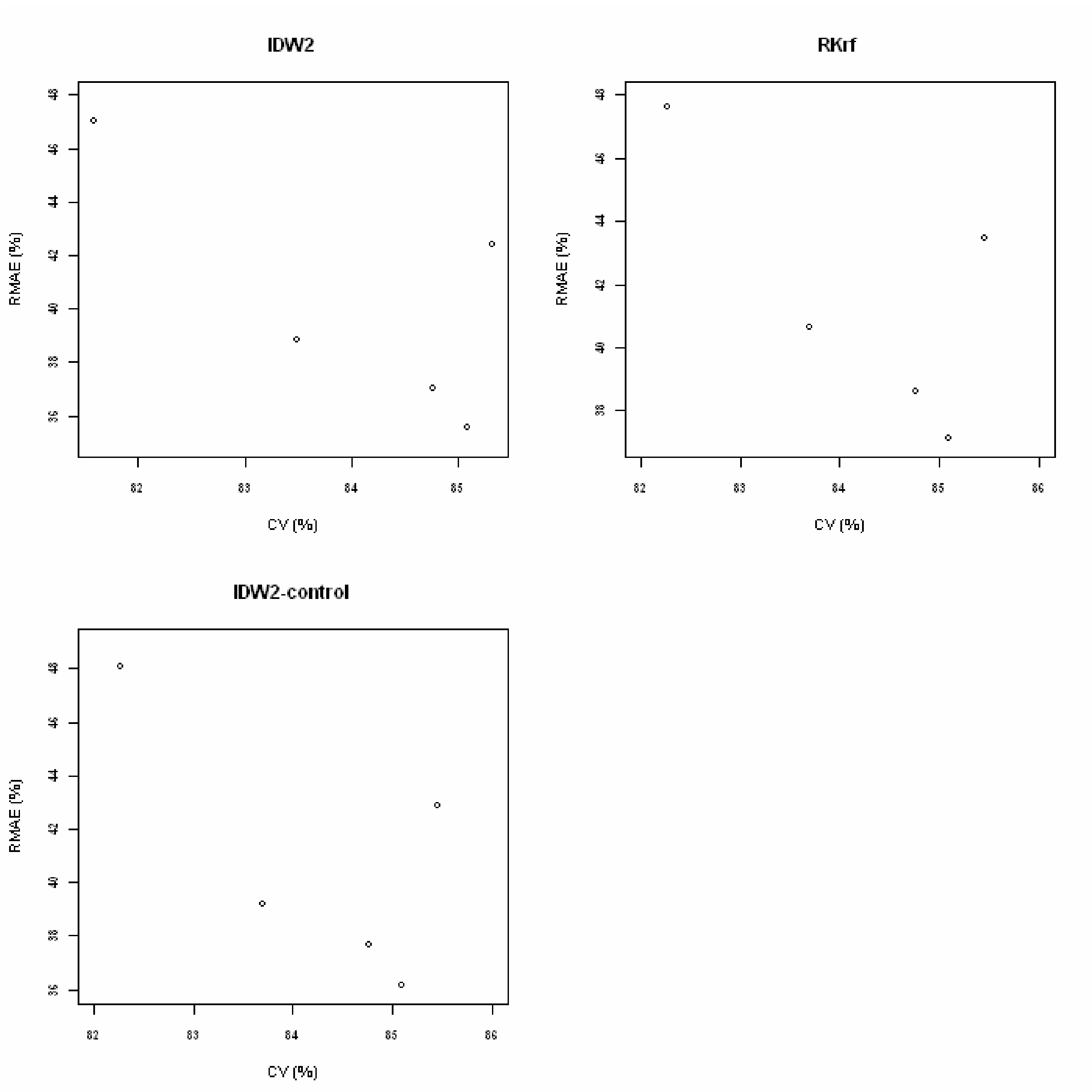


Figure 4.8. The relative absolute mean error (RMAE (%)) of the two best statistical methods and the control in relation to CV (%) in the northeast region. IDW2: local and stratified; RKrf: local and non-stratified; and IDW2 (control): local and non-stratified.

4.4.3. Southwest region

In the southwest region, there is no obvious pattern detected between CV and the performance of the three methods (i.e, RKrf with global searching neighbourhood and non-stratified samples, the most accurate method, IDW4 with local searching neighbourhood and stratified samples, and IDW2 with local searching neighbourhood and non-stratified samples) (Fig. 4.9).

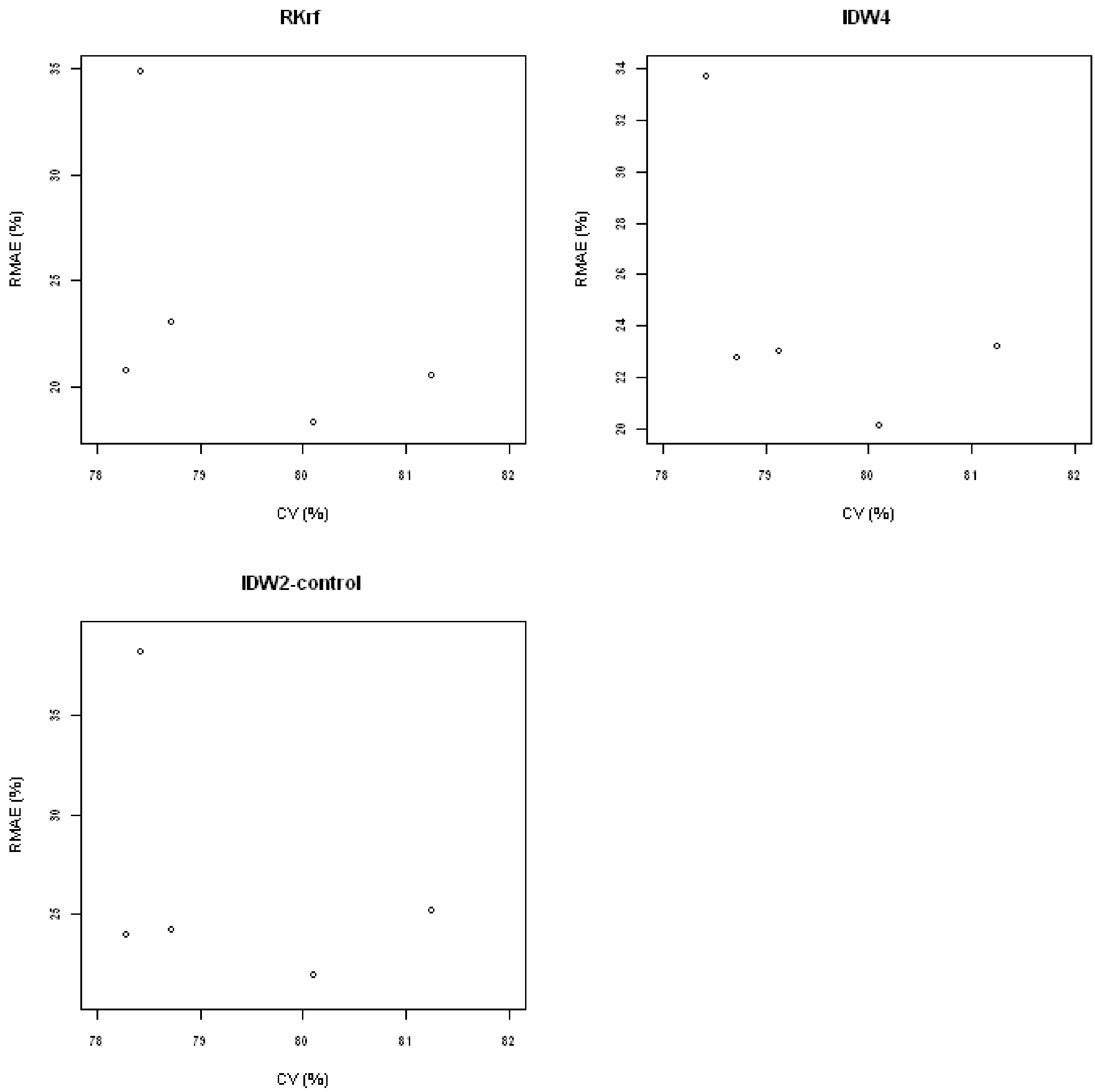


Figure 4.9. The relative absolute mean error (RMAE (%)) of the two best statistical methods and the control in relation to CV (%) in the southwest region. RKrf: global and non-stratified; IDW4: local and stratified; and IDW2 (control): local and non-stratified.

4.5. Visual comparison

In this section, we visually compare IDW2 (the control) with two most accurate methods for each of the three study regions.

4.5.1. North region

The spatial distribution of mud samples in the north region and the predictions of the three methods are illustrated in Figure 4.10. In comparison with the spatial distribution and values of mud samples, all methods capture the major spatial patterns and trend of mud content in the north region. Specifically, in the basins (Heap and Harris, 2008), mud content is relatively higher. The predictions of IDW2, the control, display "bull's eyes" patterns at sample points where samples with either high or low values. The predictions of RKrf, the most accurate method, do not display the artefacts associated with the control method. However, other artefacts are apparent in the predictions, including prominent banding patterns. The effects of geomorphic features on the predictions of RKrf are also apparent (Fig. 4.11). The predictions of KED1, the next most accurate method, mainly reflect the effects of bathymetry or geomorphic features.

The predicted values by RKrf in areas adjacent to land were higher than those predicted by the other two methods. The predictions of KED1 in the most northern portion between 130^{o}Eto 135^{o}E were too low in comparison with the predictions of the other two methods. Clearly, bathymetry and geomorphic features (Fig. 4.11) exert a strong influence on the final predictions. These features include channels, basins and submarine canyons. For other features (*e.g.*, banks and ridges), their correspondence with the predicted mud content is not so clear-cut.

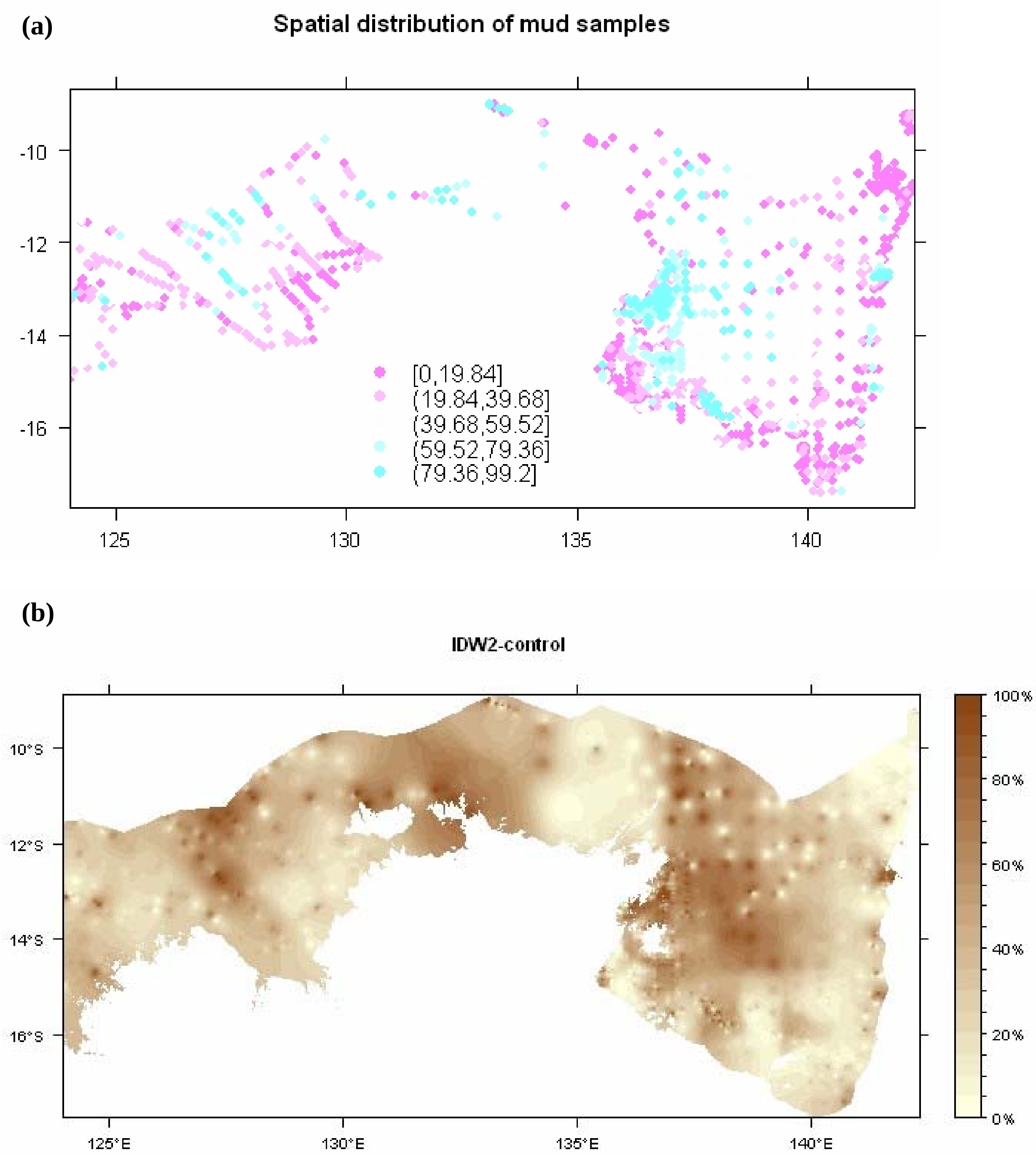
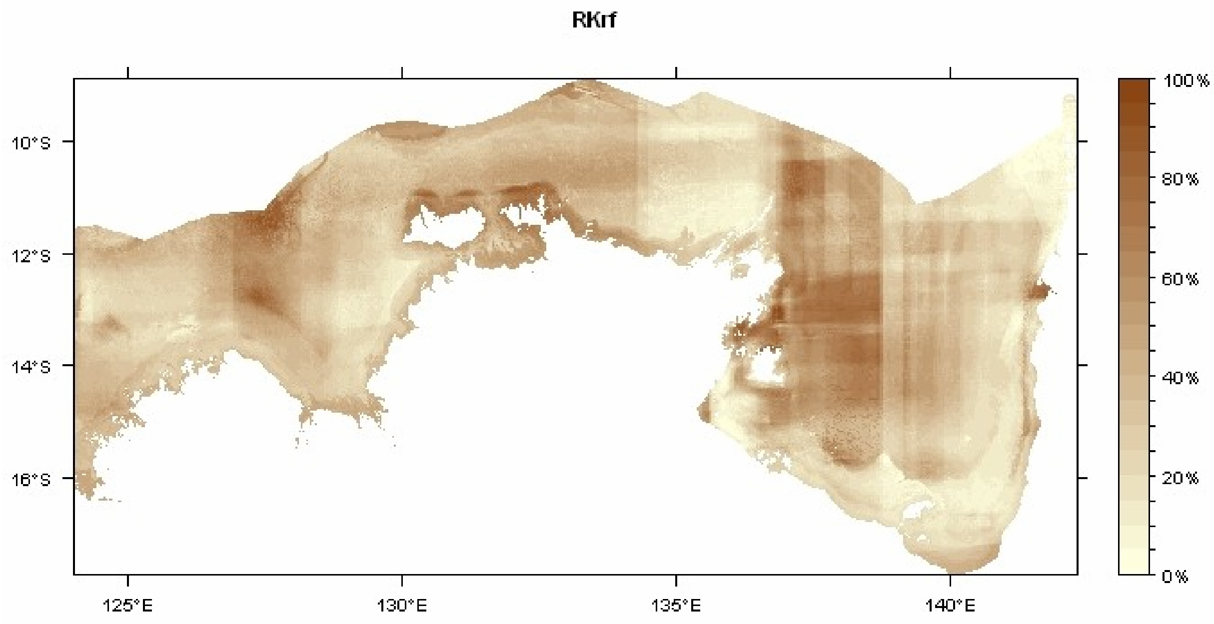


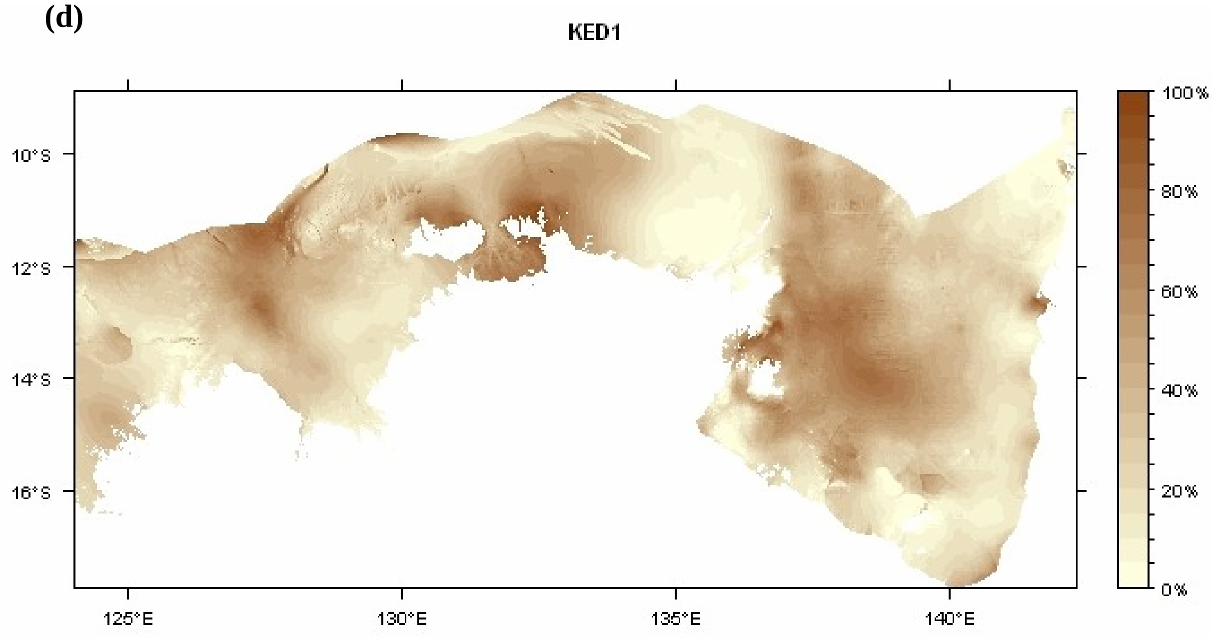
Figure 4.10. Predicted spatial distribution of seabed mud content in the north region: (a) spatial distribution of samples; (b) IDW2 (control): local and non-stratified; (c) Rkrf: local and non-stratified; and (d) KED1: local and stratified.

Fig. 4.10 (cont.):

(c)



(d)



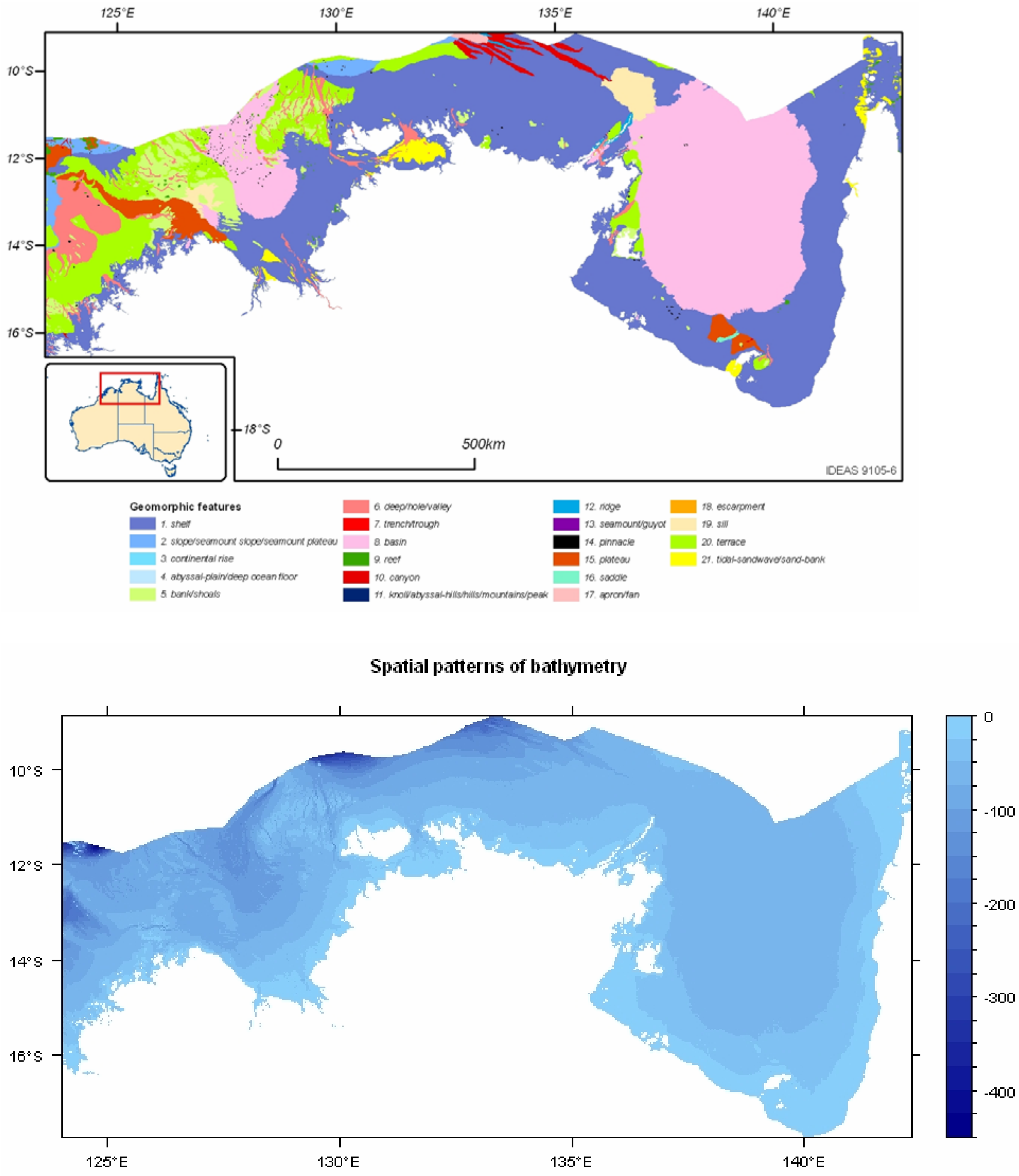


Figure 4.11. The spatial distribution of geomorphic features (above) and the spatial pattern of bathymetry (bottom) in the north region. Note the close correspondence between geomorphic features and bathymetry.

4.5.2. Northeast region

All methods capture the major spatial patterns and trends of seabed mud content in the northeast region (Fig. 4.12). In comparison with the spatial distribution of mud samples and mud content, the predictions of the control are lower than those of IDW2 with local searching neighbourhood and stratified samples, and those of RKrf with local searching neighbourhood and non-stratified samples, although the major spatial patterns are captured.

However, the predictions of IDW2 (the control) and the best performing sub-method (IDW2, local searching neighbourhood and stratified samples) display "bull's eyes" patterns at sample points having either high or low values. The best performing method is unable to make predictions for geomorphic province 3 (*i.e.*, continental rise). It is obvious that the application of sample stratification improved the predictions of IDW2 and captured the influences of geomorphic province (*i.e.*, bathymetry). However, it failed to detect the changes with the secondary information like bathymetry.

The spatial patterns of predictions of RKrf, the next most accurate method, are similar to those of the most accurate one (*i.e.*, IDW2 for local searching neighbourhood and stratified samples), but removed the obvious modelling artefacts generated by the control. We also present the predictions of RFidw2 with local searching neighbourhood and non-stratified samples (Fig. 4.12), a method slightly better than RKrf and the control. The spatial patterns are almost identical to those of RKrf. The muddy north-facing coastal embayments, inner-shelf and Capricorn Channel are accurately portrayed in these predictions. Similarly, the relatively muddy Townsville and Queensland Troughs are also highlighted. Moreover, the sandy and gravelly outer shelf and Queensland and Marion Plateaus (Heap and Harris, 2008) are clearly depicted.

The influence of bathymetry and geomorphology (Fig. 4.13) on controlling the predictions is apparent, especially in the deeper regions, including a prominent N-S trending ship-track. The predicted spatial patterns reflect the effect of geomorphic province 3 through bathymetry (Fig. 4.13) and the predicted values are lower than those in the surrounding areas.

(a)

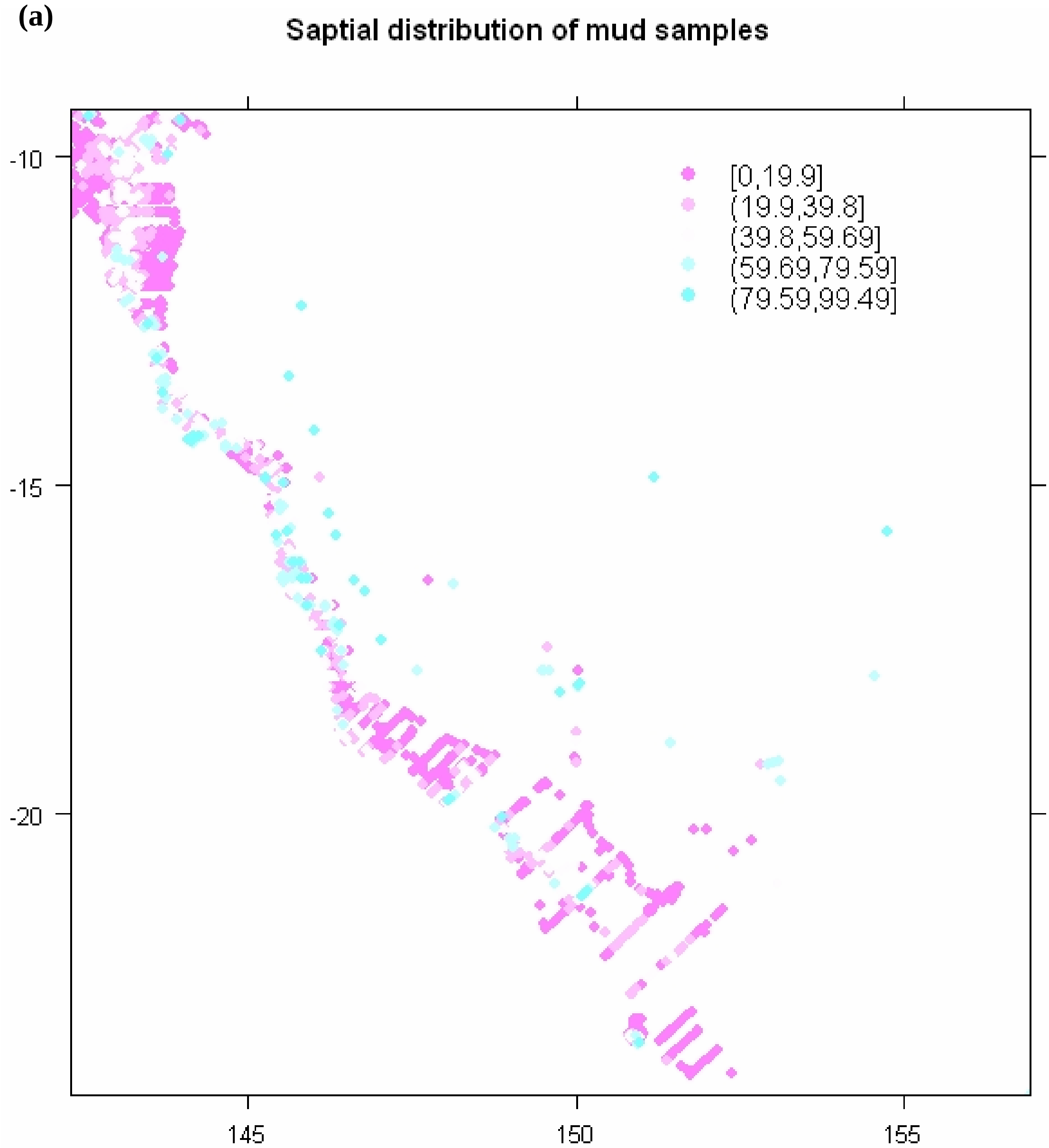


Figure 4.12. Predicted spatial distribution of seabed mud content in the northeast region: (a) spatial distribution of samples; (b) IDW2-control: local and non-stratified; (c) IDW2: local and stratified, the missing portion was due to the lack of samples in geomorphic province 3; (d) RKrf: local and non-stratified and (e) RKidw2: local and non-stratified.

Fig. 4.12. (cont.):

(b)

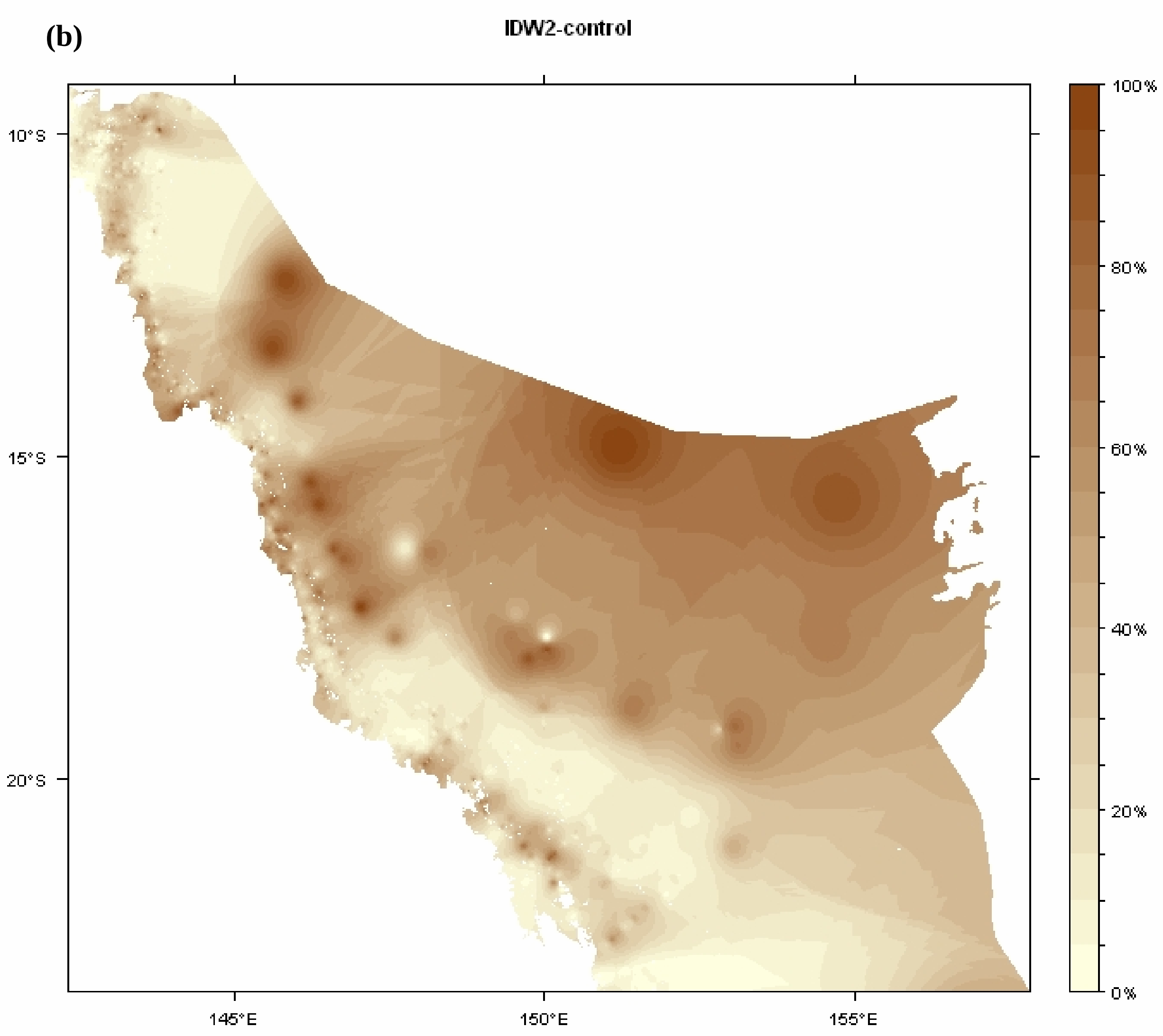


Fig. 4.12. (cont.):

(c)

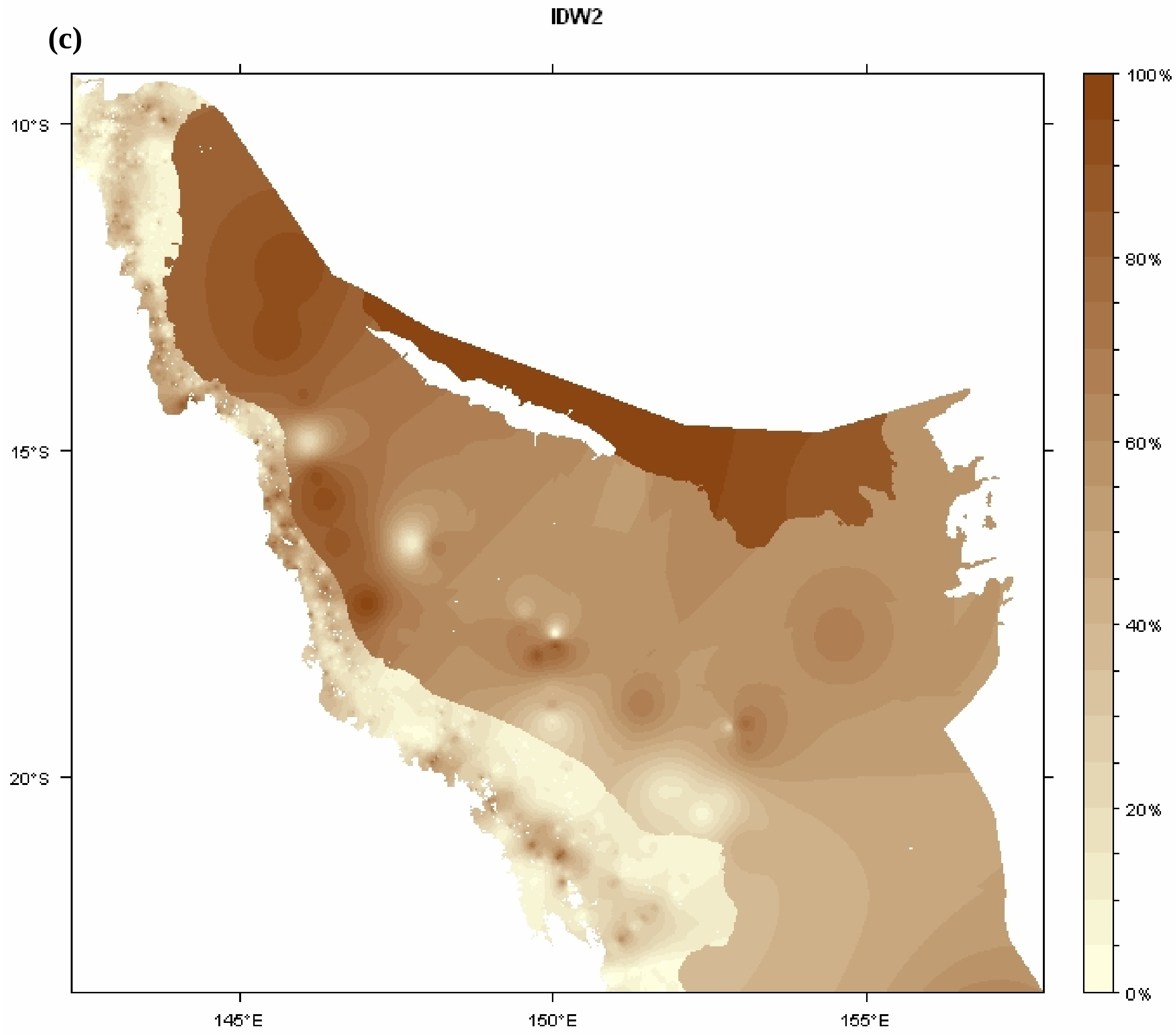


Fig. 4.12. (cont.):

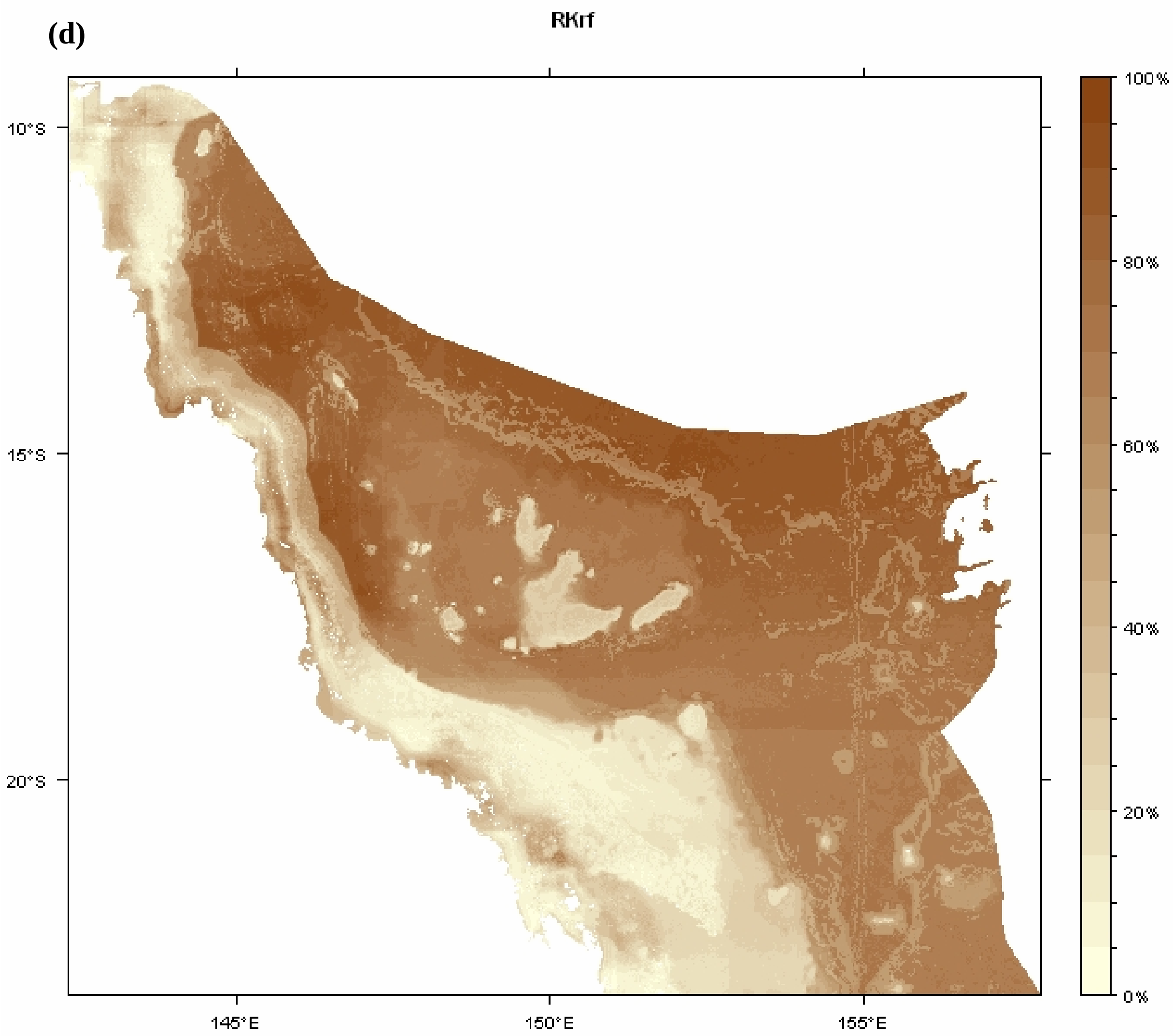
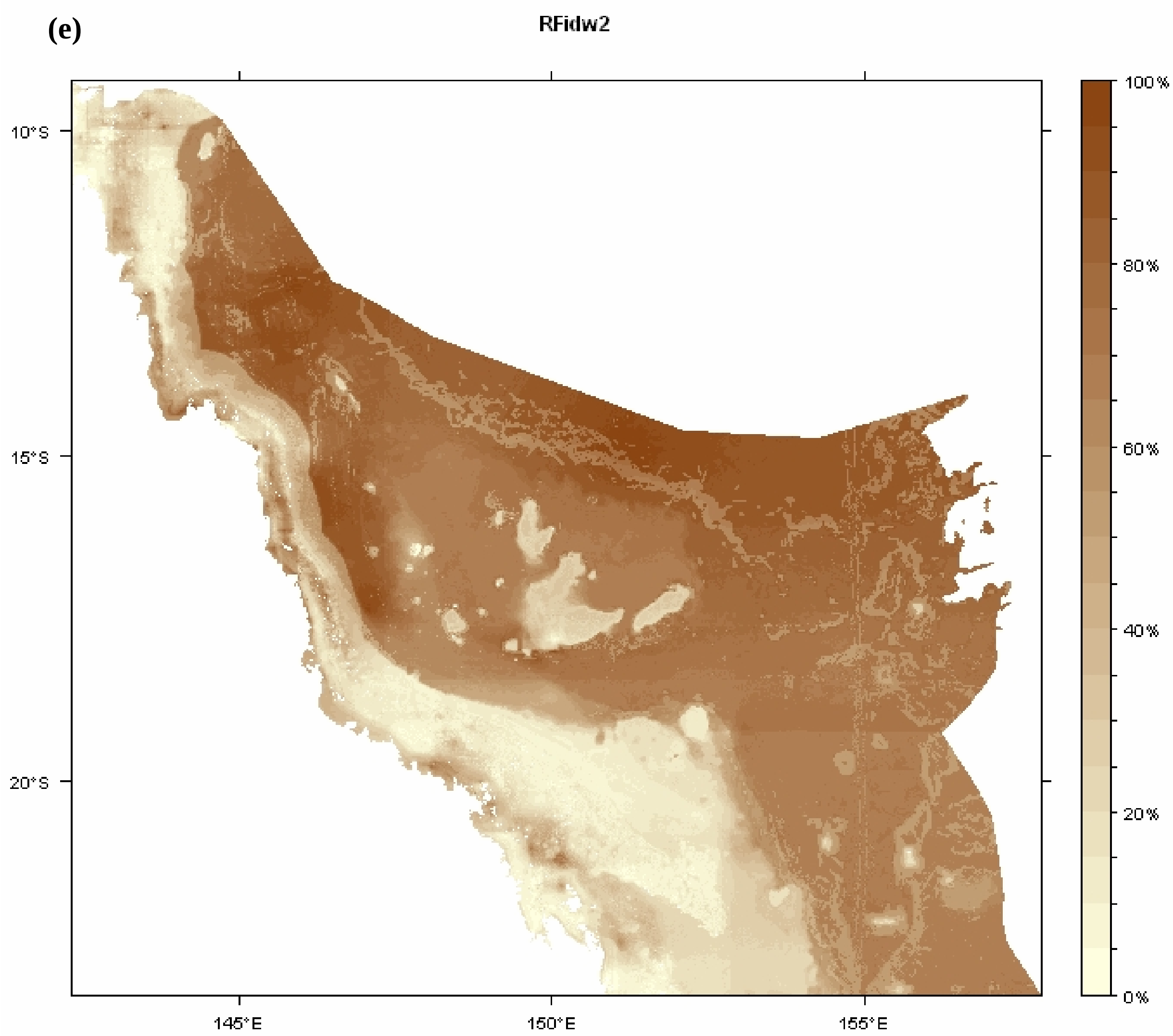


Fig. 4.12. (cont.):



(a)

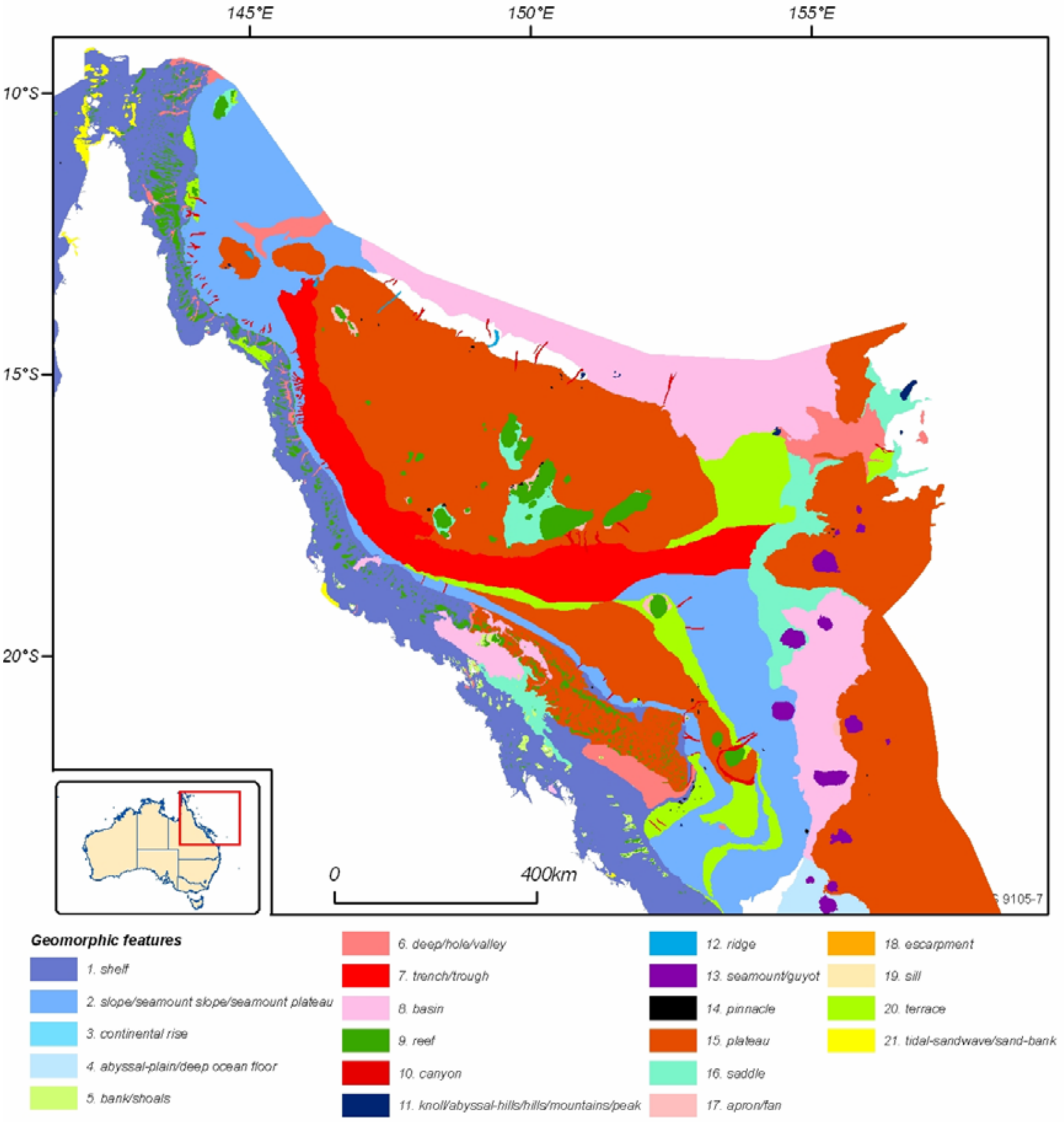
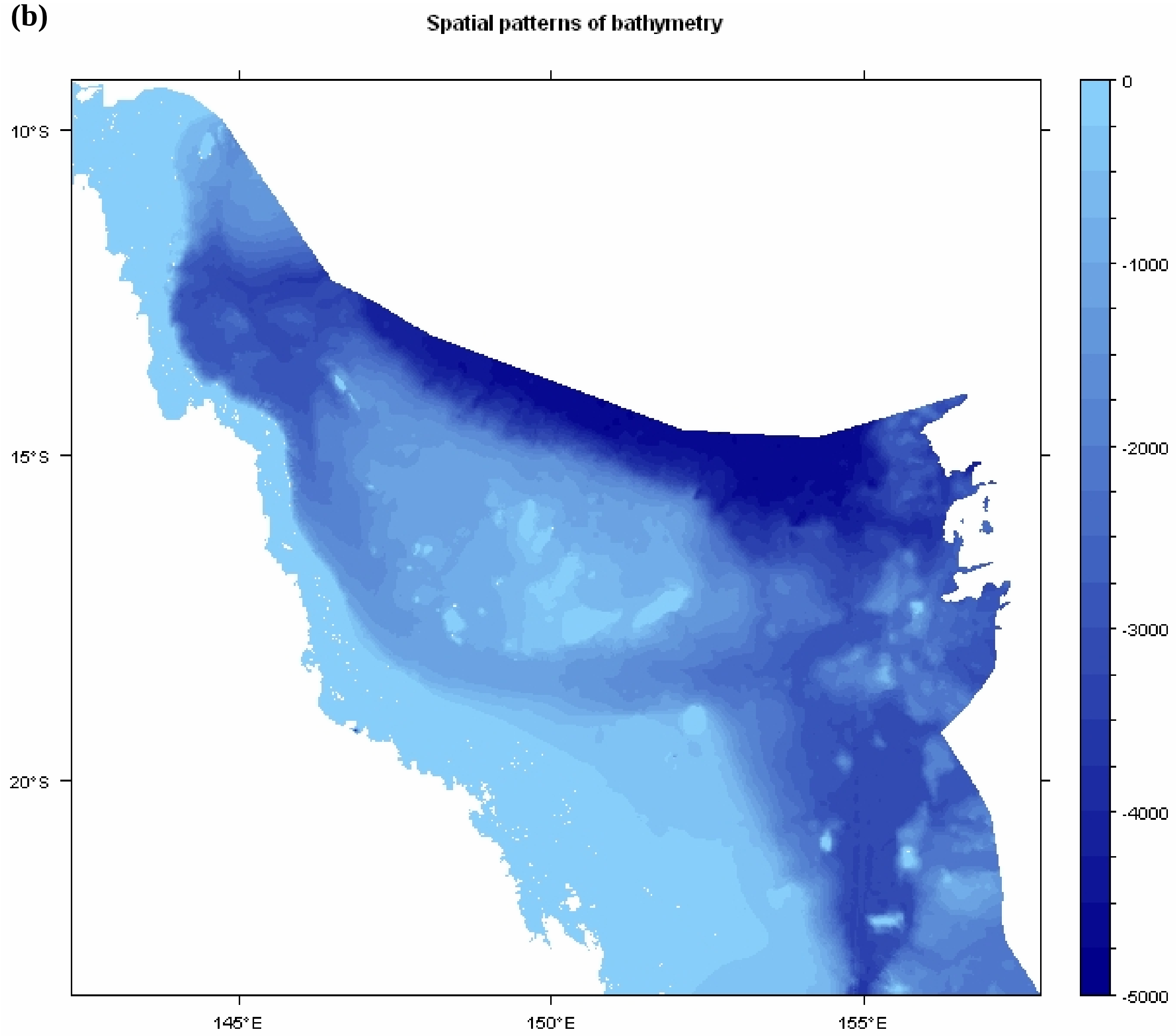


Figure 4.13. Spatial distribution of (a) geomorphic features and (b) spatial pattern of bathymetry in the northeast region.

Fig. 4.13. (cont.):

(b)



4.5.3. Southwest region

All methods capture similar major spatial patterns and trends of seabed mud content in the northeast region (Fig. 4.14). In comparison with the spatial distribution of mud samples and the values of mud content, the predictions for the area adjacent to the coast region by the control are higher than those by the other two methods. However, the predictions of IDW2 (the control) and IDW4 display "bull's eyes" patterns at sample points having either high or low values, although such patterns are more apparent for IDW4.

The spatial patterns of predictions of RKrf, the most accurate method, are similar to those of IDW4, but with the removal of the obvious artefacts evident in the predictions derived using IDW2 and IDW4, specifically, 'bulls-eyes' and relatively sharp boundaries. The predicted spatial patterns mainly reflect the effect of bathymetry (Fig. 4.15). On the shelf, predicted seabed mud content using this method captures the relatively sandy nature of this geomorphic province and the accumulations of coarse sand and gravel around the shelf edge Houtman-Abrolhos Reefs. In the deeper water, the influence of bathymetry and geomorphology (Fig. 4.15) is also apparent; the influence is greatest where the features are topographically prominent (*e.g.*, Perth Canyon). Linear tracks are apparent in the predictions and banding patterns are also apparent.

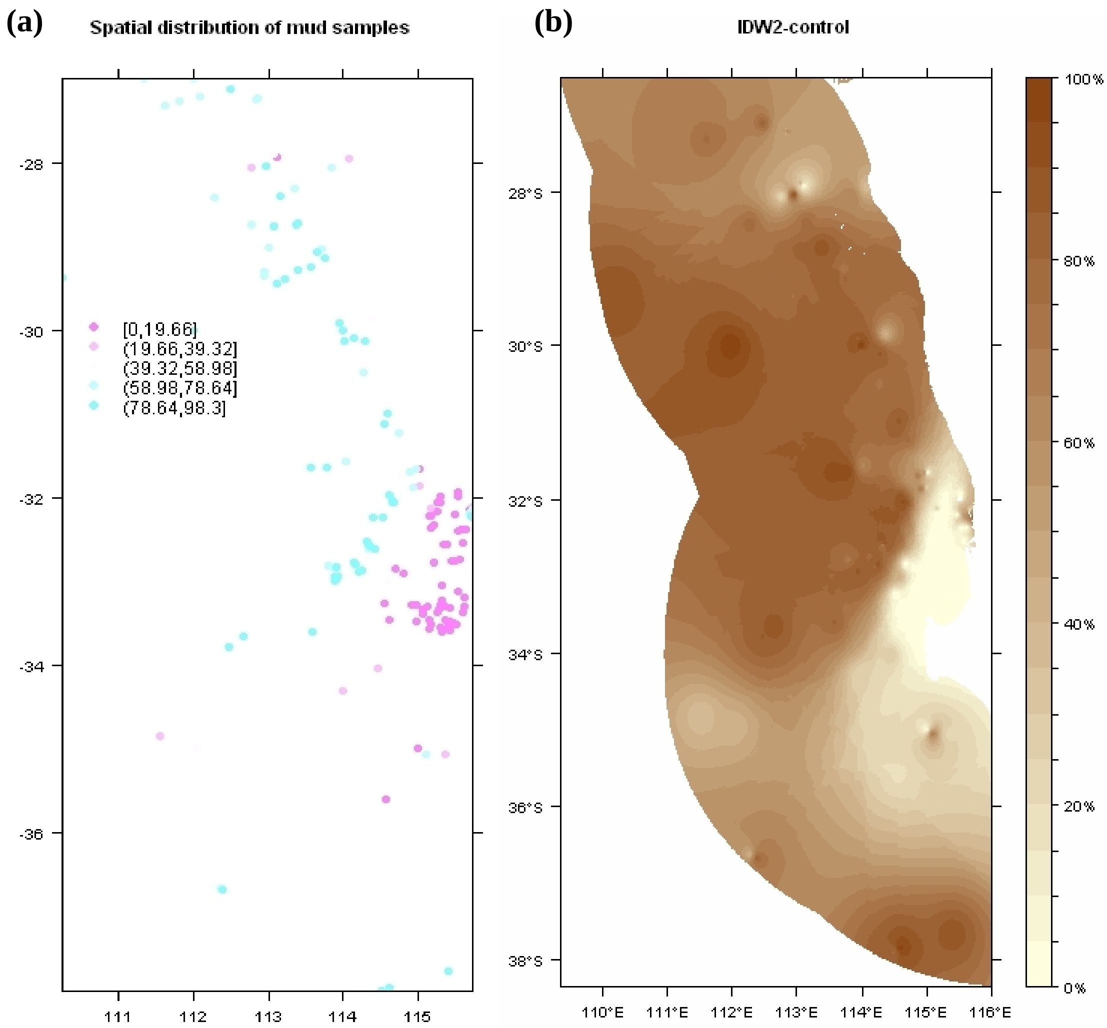
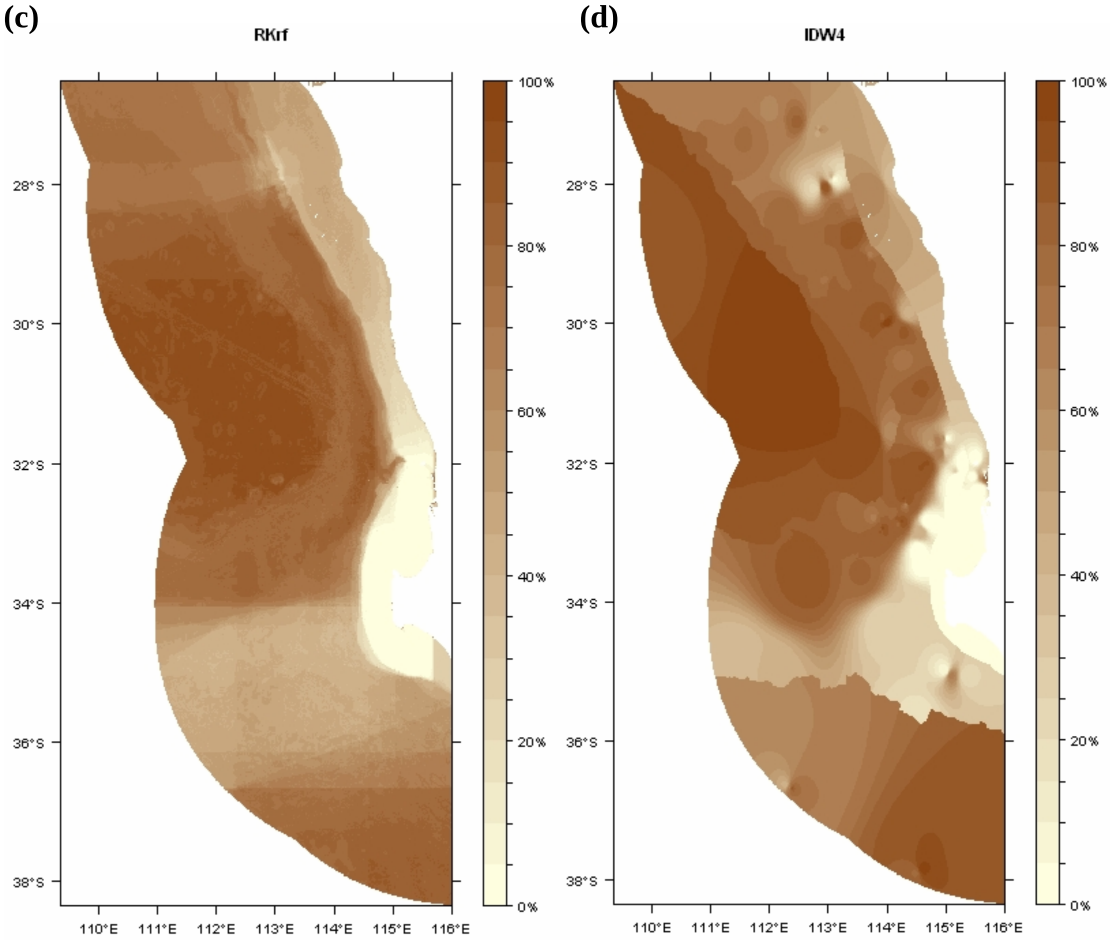


Figure 4.14. Predicted spatial distribution of seabed mud content in the southwest region: (a) spatial distribution of samples; (b) IDW2-control: local and non-stratified; (c) RKrf: global and non-stratified; and (d) IDW4: local and stratified.

Fig. 4.14. (cont.):


(c)
RKrf
(d)
IDW4
28°S
30°S
32°S
34°S
36°S
38°S
110°E
111°E
112°E
113°E
114°E
115°E
116°E
100%
80%
60%
40%
20%
0%

(a)

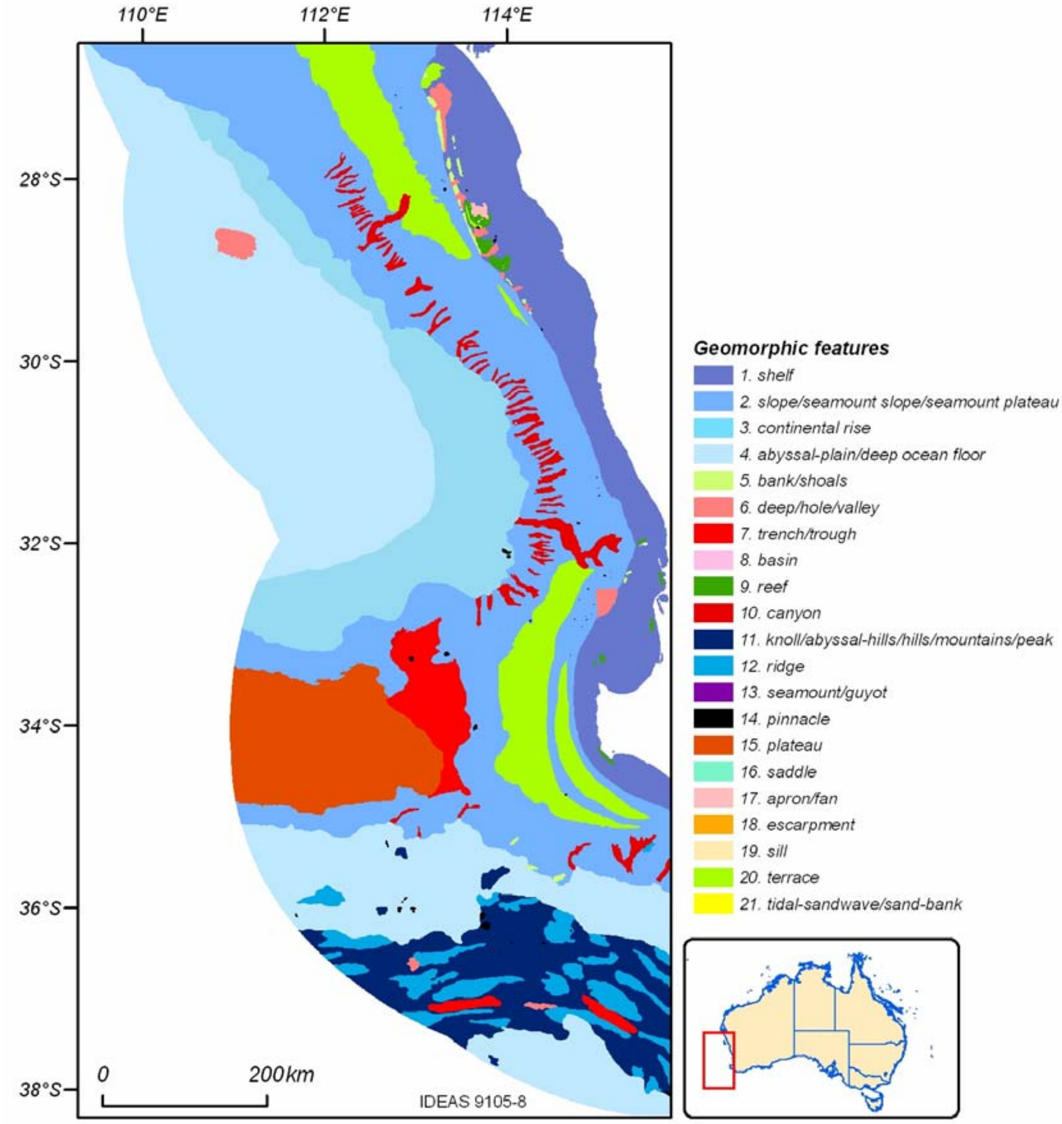
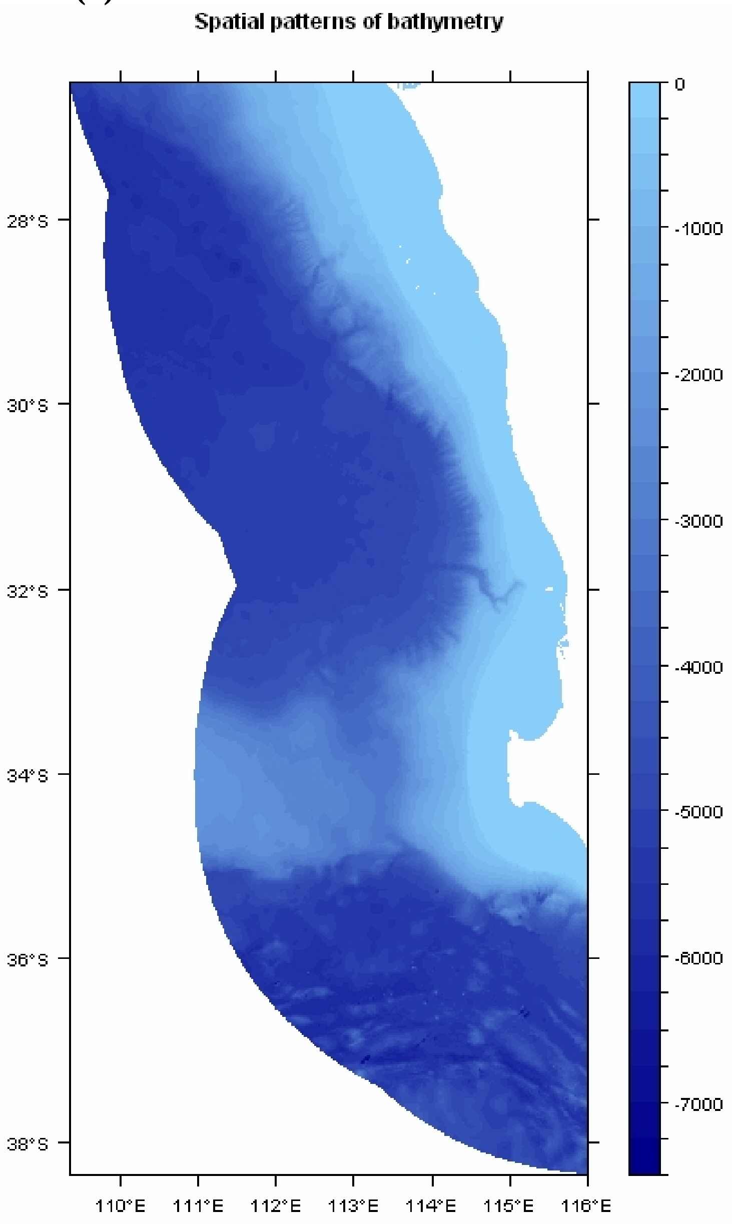


Figure 4.15. Spatial distribution of (a) geomorphic features and (b) spatial pattern of bathymetry in the southwest region.

Fig. 4.15. (cont.):

(b)



4.6. Discussion

4.6.1. Effects of methods and the interaction with searching window size and sample stratification

Overall, the performance of the statistical methods for spatial interpolation of mud content depends on the other experimental factors as evidenced by the significant interactive effects between these factors in all three study regions (Tables 4.2-4.4). No single method performed better than the other methods in all scenarios. This outcome is consistent with the 'no free lunch theorems' in optimisation (Wolpert and Macready, 1997).

The high performance of RKrf in the north and southwest regions could be attributed to the method itself and the high correlation of mud content to the secondary variables, especially in the southwest region (Tables 4.2 & 4.3). This is because the accuracy of regression modelling depends on how well the data are sampled and how significant the correlation is between the primary variable and secondary variable (Hengl, 2007). However, the latter explanation may not be true in this study because all methods using secondary information like OCK, KED and RK (except RKrf) performed more poorly than IDW and OK. This contrasts the findings of previous studies that show stronger correlations result in more accurate predictions by CK and OCK (Goovaerts, 1997), by OCK over OK and RK-C (Martínez-Cob, 1996) and by SKlm, KED and OCCK (Goovaerts, 2000). It was also argued that a threshold exists because for a correlation >0.4 SCK and OCK performed better than other methods (SK, OK, LM) (Asli and Marcotte, 1995). This can not explain the results observed in this study since the correlation coefficients are above this threshold in the southwest region (Tables 4.2 & 4.3). Therefore, it is the method itself (RKrf) that is attributed to its superior performance.

The performance of RKrf was slightly less accurate in the northeast region. This is likely due to the lack of spatial structure of the residuals of random forest as a singular model in variogram fit was observed. Because of this, we applied random forest and IDW2 (RFidw2) for local searching size and non-stratified in the northeast region. The observed RMAE for RFidw2 was 35.75 (MAE: 8.97), which demonstrates that RFidw2 is more accurate than the control.

OK usually performed better than IDW and is superior at least in theory (Li and Heap, 2008). However, OK generally performed more poorly than IDW in this study. Collins and Bolstad (1996) reported a similar finding where Optimal IDW (OIDW) was found to be superior over kriging when the data were isotropic and the primary variable was not correlated with secondary variable(s). The poor performance of OK in this study could be attributed to a number of reasons: 1) the data were not projected in this study; 2) the data stationarity required by OK was not fully satisfied although relevant transformation was employed, and the data transformation was based on full dataset in each region and might not be the most appropriate one for sub-datasets; and 3) the variogram model selected was based on the full dataset and might also not be the most appropriate one for sub-datasets used for prediction. The selection of a data-transformation and variogram model for each sub-dataset was not practical for such a comprehensive simulation experiment, which is perhaps a disadvantage of simulation automation.

The backtransformation of the predicted values could be another reason for the poor performance of OK, but it was excluded because the backtransformation method we used in this study is not less accurate than other backtransformation methods in terms of RMAE as demonstrated in Appendix D. In addition, RKrf is still the most accurate methods no matter what kind of backtransformation methods were applied to the other kriging methods.

A singular model in variogram fit may also contribute to the poor performance of KED3, KED4 and RKgls as this was observed for datasets in the southwest region for both non-stratified and stratified samples and for both global and local modelling, which may contribute to their poor performance. Overfitting may also contribute to the poor performance of some regression kriging methods such as RKglm4 and RKlm5 because the models were not simplified in this study.

TPS performed poorly in all cases. This is consistent with the findings of previous studies where OK and IDW appeared more accurate than TPS (Brus et al., 1996; Schloeder et al., 2001) and for precipitation(Hartkamp et al., 1999). However, TPS performed slightly better than IDW and OCK for temperature (Hartkamp et al., 1999), and than IDW and OK (Jarvis and Stuart, 2001; Laslett et al., 1987). These findings indicate that the performance of these methods varies between studies and may depend on other variables such as the nature of primary variables.

This study has showed that three machine learning methods, like RT and SVM, performed more poorly than IDW and OK, but that another machine learning method, random forest, outperformed all other methods when it was combined with OK. A combined method of RT and OK has been shown to be more accurate than IDW and OK (Martínez-Cob, 1996) and SVM was found to be superior to random forest in other disciplines (Statnikov et al., 2008). Therefore, machine learning methods may have great potential for spatial predictions if they are combined with other methods, despite the fact that they were only applied to one scenario (global and stratified). Further studies are needed to test their performance in different scenarios and in combination with other methods like OK.

Comparing the accuracy of methods using secondary information (*e.g.*, KED, OCK, RK and UK) and those not using secondary information like IDW and OK (Figs. 4.1-4.3) indicates that the effects of inclusion of secondary information are method-dependent. This suggests that inclusion of secondary information does not always improve the prediction accuracy. However, for RKrf, secondary information is essential.

Geomorphic province information was used in this study to stratify the samples so that the variance within each geomorphic feature was expected to be reduced and the effects of non-stationary could be avoided (Voltz and Webster, 1990), then the accuracy of the spatial interpolation of the environmental variables was expected to be improved. However, this is not supported by findings in this study because the performance of most statistical methods were generally more poorly, RKrf in particular, when samples were stratified by geomorphic province. Whereas IDW, OK and OCK were less sensitive to sample stratification in two regions, and IDW, OK and OCK achieved a small improvement in one region. It seems that there are no consistent findings about the effects of sample stratification as contradictory results were also observed in previous studies (Brus et al., 1996; Voltz and Webster, 1990).

Geomorphic province was largely a categorical expression of bathymetry (Heap and Harris, 2008), which explains why geomorphic province was the least important factor in random forest, and why stratification reduced the accuracy of RKrf. This is because the relevant information had already been explained by bathymetry and the residuals of random forest have little relation to geomorphic province. Ultimately, our study indicates that the effects of sample stratification are method-dependent.

The negative effects of sample stratification on most methods using bathymetry as secondary information could be explained by the fact that bathymetry has already explained relevant information in the variation of mud content. In such cases geomorphic provinces have nothing to do with the residuals of relevant models (lm, glm and gls).

The negative effects of including the latitude and logitude terms as secondary information on the prediction accuracy of KED and all combined methods (except RKrf) probably result from the un-projected coordinates of mud content as explained previously. For data in longitude and latitude, the *gstat* package in R assumes a unit difference in longitude axis reflects approximately the same distance in latitude axis. Ignoring changes in distance along latitude will certainly reduce the accuracy of the distance estimated and thus may reduce the reliability of the predictions.

Prediction accuracy displayed tremendous variations between regions. Overall the accuracy of the methods was much higher in the southwest region compared to the other two regions (Table 4.5; Figs. 4.1-4.3). The accuracy of the control method and RKrf (local and non-stratified), two methods comparable between regions, in the southwest region was 15.11% and 18.05% higher in terms of RMAE than in the north region, and 14.22% and 18.83% higher than in the northeast region. However, the sample size and density in the southwest region is much lower than those in the north and northeast regions (Tables 3.4 & 3.5). Given that the control method did not use secondary information, it seems the difference between regions was independent of the statistical methods and availability of secondary information. The the effects of regions on the accuracy of the predictions remains difficult to explain.

4.6.2. Effects of sample density

The prediction error is theoretically expected to display three response patterns to sample sendity: 1) the prediction error is contantly low when sample density changes; 2) the prediction error is contantly high when sample density changes; and 3) the prediction error decreases when sample density increases, and then the sample density reaches a threshold, further increase in sample density will not lead to the decrease in prediction error (Figure 4.16). The results show that as sample density increased from 20% to 100%, the accuracy of the three methods considered increased by 5.34-5.52% in the north region, 10.45-11.44% in the northeast region and 13.6-16.66% in the southwest region in terms of RMAE. On the basis of these findings, we can conclude that as sample density increases, the accuracy of the statistical methods for spatial interpolation increases and no threshold of sample density has been reached in any of these regions. This finding is consistent with results from previous studies (Englund et al., 1992; Isaaks and Srivastava, 1989; Stahl et al., 2006). The general trend and observed interactive effects at lower densities also agree with findings in studies of Li *et al*. (2007) and Wang *et al*. (2005).

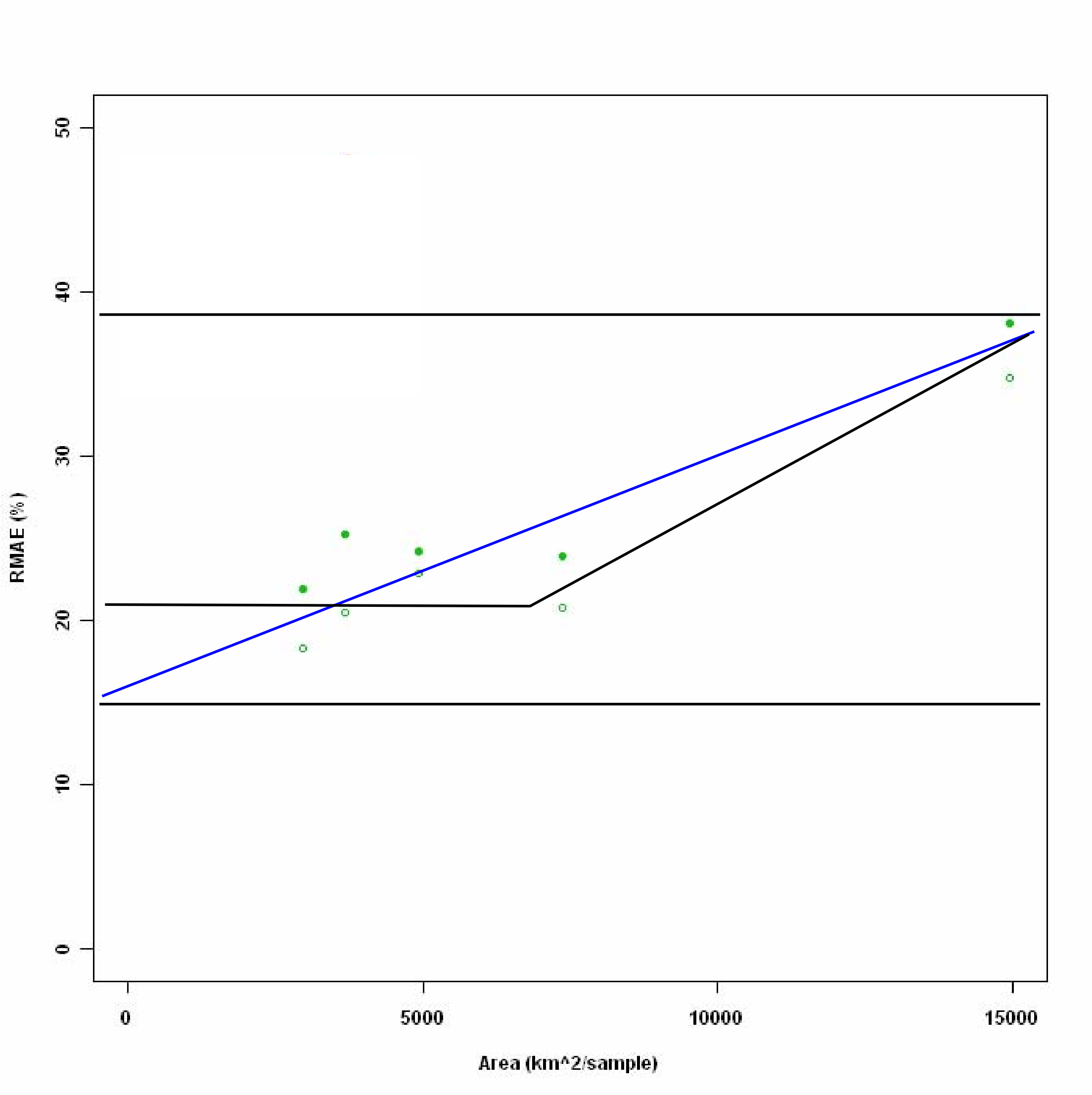


Figure 4.16. Effects of sample density on relative absolute mean error (RAME) as illustrated using two most accurate methods in the southwest region: the relative absolute mean error (RMAE (%)) of RKrf: local and non-stratified (open circle); and IDW2: local and non-stratified (closed circle).

The findings in this study indicate that an increase in sample density (or sample size) can considerably improve the prediction accuracy of statistical methods for spatial interpolation. No threshold in sample density was found in any region, which suggests that the number of samples collected so far is still below any threshold of sample density, if one exists, to capture the variability in actual mud content.

Despite the fact that the accuracy of the selected methods increases with sample density in all three regions, this finding cannot be generalised to compare studies from different regions. This is because the accuracy of a method may not be necessarily higher for a region with higher sample density than that for a region with a lower sample density (Tables 3.4 & 3.5). This is further demonstrated by the performance of RKrf and IDW2 along sample density in the three study regions (Fig. 4.17). This suggests that the effects of sample size or sample density are region-dependent.

Moreover, this helps to explain why no patterns were detected in relation to sample density when studies from different disciplines were compared (2008).

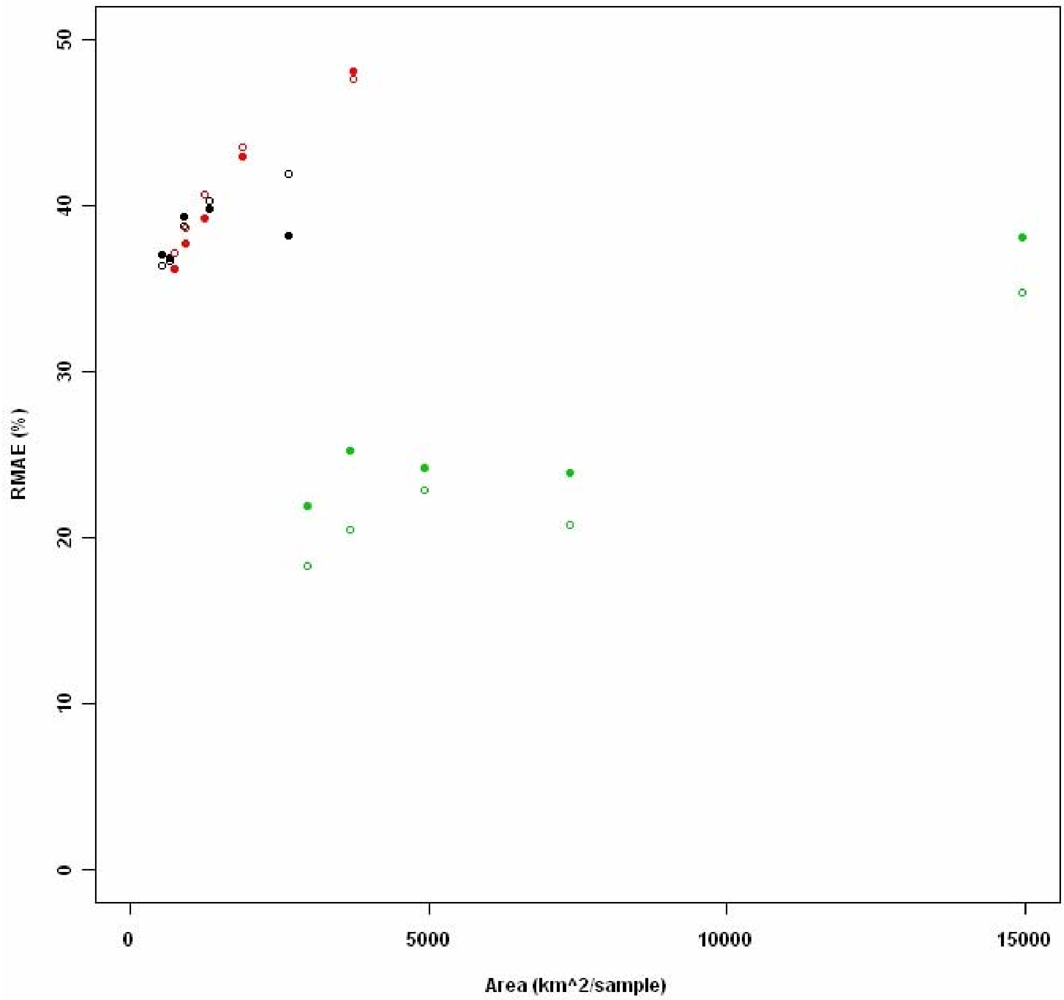


Figure 4.17. The relative absolute mean error (RMAE (%)) of RKrf: local and non-stratified (open circle); and IDW2: local and non-stratified (closed circle) in the north (black), northeast (red) and southwest (green) regions in relation to sample density (km^2/sample).

4.6.3. Data variation

The observed patterns of method performance with CV are different in the three study regions. This result suggests that there is no consistent relationship between the accuracy of the statistical methods for spatial interpolation and data variation between regions, which is inconsistent with the patterns observed by Li and Heap (2008), Collins and Bolstad (1996), Martínez-Cob (1996) and Schloeder *et al*. (2001). This phenomenon might be attributed to the very short gradient of data variation for all three regions. The gradient has a range of 5% in the north region and <3% in the northeast and southwest region.

A second important finding is that the overall performance of the two most accurate methods is much higher than that reported in previous studies (Li and Heap, 2009) (Fig. 4.18). The data CV is 83% and 85% in the north and northeast regions respectively, but the accuracy of the best two methods for these two regions increased by approximately 15% in terms of RMAE in comparison with the performance of the best methods reported in other tudies. The data CV is 80% in the southwest region, but the accuracy of two most accurate methods for this region increased by approximately 30% in terms of RMAE in comparison with the performance of the best methods reported in other studies. This finding demonstrates that the methods selected in this study are far more robust than the methods identified in previously published studies (Li and Heap, 2009).

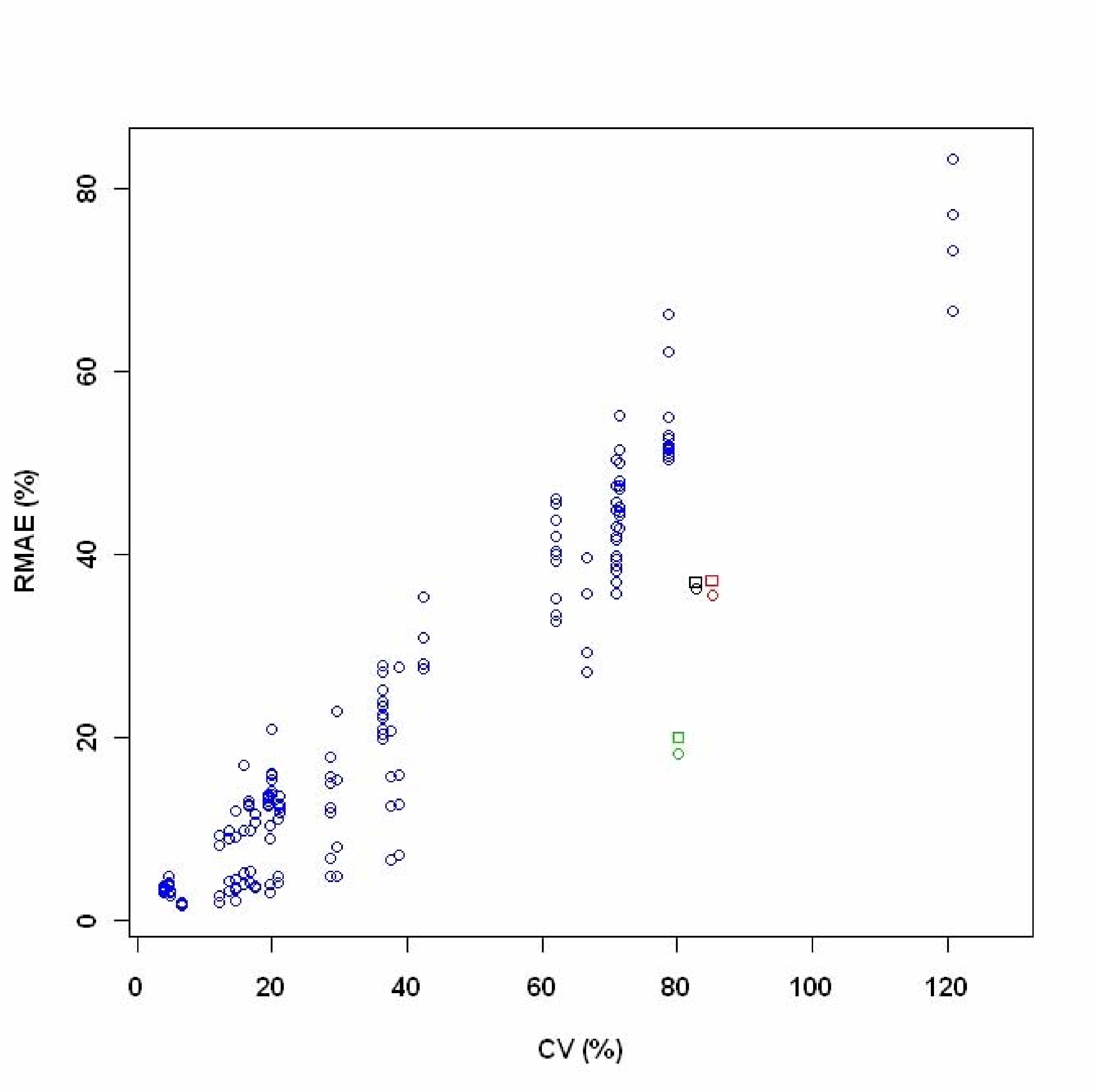


Figure 4.18. The relative absolute mean error (RMAE (%)) of the best two statistical methods (the most accurate: circle; the next accurate: square) in relation to CV (%) in the north (black), northeast (red) and southwest(green) regions compared to results (blue) of other studies in several disciplines (Li and Heap, 2009).

4.6.4. Visual comparison

The relatively high mud content in the basins in the north region is consistent to findings in the previously published studies (*e.g.*, Van Andel and Veevers, 1967). The banding patterns in the predictions of RKrf reflect the influence of longitude and latitude as secondary information. This suggests that in these areas mud content was more related to longitude and latitude than other secondary variables including bathymetry. This may also be caused by the coincidence of the spatial patterns of samples along latitude and longitude. The apparent effects of geomorphic features on the predictions of RKrf is because geomorphic features were mainly based on bathymetry (Heap and Harris, 2008). The predictions of KED1 mainly reflect the effects of bathymetry, thus were the mostly visually pleasant.

The higher predicted values by RKrf than the other two methods in areas adjacent to land in the north region may more likely reflect the real situation as soil deposit may occur due nearby land erosion. The low predictions of KED1 in the most northern portion between 130^{o}Eto 135^{o}E in comparison with the other two methods may result from the changes in bathymetry, and therefore geomorphic features. The strong influence of geomorphic features such as channels and basins this is not necessarily a problem, as they are associated with different mud contents because of their contrasting environments compared with the surrounding shelf. Interestingly, the strong influence on the predictions occurs for those features that are topographically prominent on the generally low-gradient shelf (*i.e.*, submarine canyons). This implies that the added effects of geomorphology and bathymetry are contributing significantly to the final predictions.

In the northeast region, the lack of predictions for geomorphic province 3 (*i.e.*, continental rise) for the best performing method was due to the method used geomorphic provinces to stratify the samples and geomorphic province 3 contains no samples for prediction. When there is no sample in a geomorphic province, there is no prediction. Both RKrf and RFidw2 produced results that are broadly consistent with previously published work on the sedimentology of the northeast Australian margin (*e.g.*, Maxwell and Swinchatt, 1970; Belperio, 1983; Keene *et al*., 2008).

The prominent N-S trending ship-track in the predictins of RKrf and Rkidw2 in the northest region are due to where high-resolution multibeam sonar data were collected. The relatively strong influence of the combined effects of bathymetry and geomorphology at these deeper depths is mainly due to the lack of samples, compared with the numerous samples for the shelf. The lower predicted values for geomorphic province 3 than those in the surrounding areas are pretty convincing given its shallower water depth. Also apparent in both the RKrf and RFidw2 predictions is the faint banding associated with the influence of latitude and longitude as secondary information. Despite these artefacts, which are restricted to regions of limited measured data, the predictionss derived from RKrf and RFidw2 are generally more consistent with existing knowledge of the broad-scale sedimentology of the northeast margin—certainly they capture more spatial details—and are more visually appealing than those produced using the relatively simple IDW method (*i.e.*, the control). This suggests that both methods could be used for spatial interpolation, but RFidw2 is slightly preferred because its predictions are slightly more accurate.

In the southwest region , The predicted spatial patterns by RKrf mainly reflect the effect of bathymetry related variables as the patterns were similar to those identified by Heap and Harris (2008).The influence of bathymetry and geomorphology is beacue they were used as secondary information to derive the predictions. For submarine canyons, this may not necessarily be a problem as samples from the Perth Canyon indicate that at shallow depths (<1,000 m) the canyon is slightly more muddy than the surrounding slope environments (Heap *et al*., 2008). The linearship tracks were apparent where multibeam sonar data were collected and the banding patterns reflects the influence of longitude and latitude as they were used as secondary infromation. Like the north and northeast margins, these artefacts derived from secondary information are prominent due to the relatively few samples collected from the deeper parts of the margin.

4.7. Implications, limitations and uncertainty

4.7.1. Implications

This study found that the accuracy of spatial predictions is affected by all experimental factors considered including spatial interpolation methods, sample stratification, search window size, sample density, data variation and the study region and their interactions. Although no single method is superior in all scenarios, a combined method (*i.e.*, random forest and ordinary kriging; RKrf) is the most robust in all three regions based on both the accuracy and the visual examination. Another combined method (*i.e.*, random forest and IDS; RFidw2) has a superior performance in the northeast region. Both RKrf and RKidw2 are newly developed methods in this study. Their RMAE is 15% lower in two regions and 30% lower in the remaining region than that of the best methods in the previously published studies, highlighting the robustness of these methods. They are new methods to predict seabed mud content. This is a major step forward from the control (IDW2) that will have application for modelling spatial data across Geoscience Australia. No threshold in sample density was found in any region in relation to prediction accuracy. This suggests that more samples need to be collected.

The ‘no free lunch theorems’ in optimisation (Wolpert and Macready, 1997) are confirmed in the application of statistical methods to spatial interpolation because the performance of the methods tested changes with regions and other experimental factors. We would therefore argue that regarding the spatial interpolation of environmental variables, it is predicable; however, once you predict, whatever the prediction is, it is inaccurate. Both of these arguments suggest that to achieve optimal prediction, equal efforts are needed to develop and search robust methods in any given study region. This activity should continue until the desired accuracy is achieved.

4.7.2. Limitations and uncertainty

Data sources: The samples in MARS database were collected on over 300 surveys between 1899 and 2009. In this study, the temporal changes in mud content were not considered. This limitation needs to be taken into account in applying resultant predictions.

Data projection: In this study, we used datum WGS84 without any projection. As discussed in Appendix B.3.1, this may reduce the reliability of the predictions. Data in this study was in latitude and longitude, and not projected, which would reduce the accuracy of geostatistical methods and all other methods using latitude and longitude as secondary information. Since RKrf used such secondary information, its accuracy might also be affected. This limitation needs to be taken into account in assessing the performance of statistical methods using such secondary information or using variogram modelling.

Data sampling: To test the effect of sample density (or sample size), we randomly selected a subset of samples from the full dataset. The subsets of higher sample density do not necessarily contain all samples in subsets of lower sample density, which could result in that the estimations using a dataset with lower sample density may be more accurate than those using a dataset with higher sample density. This phenomenon is mainly caused by the difference in the mean and CV of samples, in two datasets. Differences in the spatial distribution of samples in two datasets may also be crucial (Isaaks and Srivastava, 1989; Laslett, 1994; Zimmerman et al., 1999). Therefore, to avoid such random errors, in future studies testing the effect of sample density, samples in the dataset with lower sample density should be included in the dataset with higher sample density. For example, the estimation accuracy of the same method (*e.g.*, IDW2, IDW3) was higher for datasets with lowest sample density (20%) than those with higher sample densities in terms of MAE. This phenomenon is possibly due to the changes in the mean and range of mud content in datasets, because the maximum value and mean value of mud content in the dataset with lowest sample density is the smallest. Including extra samples (*e.g.*, dataset with sample density of 100%) increased the range and mean. Using RMAE reduced the effects of such changes, indicating that these changes have contributed to the observed phenomenon.

Automation in the simulation experiment: To satisfy the data stationarity requirement by geo-statistical methods and linear models, the data were transformed. The data transformations in this study were based on the full dataset in each region, because selection of the data transformation method for each sub-dataset was not practical for such a comprehensive simulation experiment. This introduces uncertainty in the subsequent predictions.

The selection of the variogram model in this study was also based on the full dataset in each region. The selected variogram model might not be the most appropriate one for the subsets used for prediction. This may affect the observed accuracy as it was found that the first-order trend OK performed better with Gaussian semi-variogram model than with spherical and exponential models (Hu et al., 2004).

Searching neighbourhood: The local search widow size was specified based on previous studies for methods computed in R or on the default value for methods computed in ArcGIS. This size may not lead to an optimised prediction, although the prediction was more accurate when the searching window size is local. The search window size needs to be optimised.

Chapter 5. Summary and recommendations

In this chapter, we summarise the important findings from the simulation experiment and offer recommendations for future study. The findings in this study have wide application, practically for any spatial data set that must be interpolated from relatively sparse point data, and may be useful for a variety of environmental scientists.

5.1. Important findings

Regional differences: The three regions are different in terms of orientation (*i.e.*, aspect); sample density, stationarity and variogram model of mud data; properties of secondary information like bathymetry, geomorphology; and correlation between primary variable and secondary variables.

The most robust statistical methods: No single method was superior in all scenarios tested in this study. Three sub-methods were more accurate than the control and two were equivalent to the control (IDW2) in the north region. In the northeast region, three sub-methods were more accurate than the control and three were equivalent to the control. In the southwest region, 12 sub-methods were more accurate than the control. Overall, RKrf was the most accurate method for datasets in the north and southwest regions when samples are not stratified, and the searching neighbourhood has little impact on its performance. In the northeast region, the most accurate method was IDW2 when samples were stratified and the searching neighbourhood was local. Despite this, IDW2 failed to detect the influence of secondary information like bathymetry. Although RKrf was slightly less accurate than the control, besides IDW, it was the next most accurate method when samples were not stratified and the searching neighbourhood was local. Considering both the accuracy and the visual examination, we found that RKrf is the most robust method for spatial interpolation of Australian marine mud samples in all three regions. RFidw2 is an alternative in the northeast region.

New direction and new source of alternative methods: The combination of existing spatial interpolation methods with machine learning methods revealed a new direction for searching spatial interpolation methods. Such combination with machine learning methods opened a new source of alternative tools for spatial interpolation.

Sample stratification: Sample stratification reduces the accuracy of the methods that use bathymetry as secondary information and its effects are method-dependent.

Searching neighbourhood: Searching window size has significant influence on the accuracy of the statistical methods. A local searching neighbourhood is preferred to a global one. Its effects are method-dependent.

Sample density: Increasing the sample density (or sample size) significantly improves the prediction accuracy of the statistical methods for spatial interpolation. No threshold in sample density was found in any region in relation to prediction accuracy. The effects of sample size are region-dependent because the accuracy was not necessarily higher for a region with higher sample density than for a region with lower sample density.

Data variation: No consistent patterns in the performance of the methods were observed in relation to data variation in the three study regions. However, in terms of RMAE, the accuracy of two most accurate methods increased by approximately 15% in the north and northeast regions and by approximately 30% in the southwest region in comparison with performance the best methods in the previously published studies (Li and Heap, 2008). This highlights the robustness of the methods selected in this study.

Secondary information: The effects of the inclusion of secondary information are method-dependent, suggesting that inclusion of secondary information does not always improve the prediction accuracy.

Regional difference in the performance of statistical methods: The accuracy of the methods displayed tremendous variation between the three regions studied. The difference in method performance between the regions is independent of statistical methods and availability of secondary information.

Visual examination: Visual (subjective) examination is equally important to assessing the performance of each method. It is an essential step in assessing the appropriateness of the final spatial predictions from a statistical method.

Performance of statistical methods and modelling input: The '*no free lunch theorems*' in optimisation (Wolpert and Macready, 1997) are confirmed in the application of statistical methods to spatial interpolation. Regarding the spatial interpolation of environmental variables, it is predicable; however, once you predict, whatever the prediction is, it is inaccurate. Both of these arguments suggest that to achieve optimal prediction in regions with inherent difference in sample size, aspect, ect., equal effort is needed to develop and search robust methods that are tailored to each region. This should continue until the desired accuracy is achieved for each region.

Criteria for quality control of MARS database: The samples in the MARS database were found to contain a lot of noise. Criteria directly related to the database were developed for quality control of the samples in the MARS database.

Guidelines for future studies: The procedures employed and criteria developed for data quality control and for searching robust spatial interpolation methods provide guidelines to relevant future studies.

5.2. Recommendations for future study

Data projection: We strongly recommend that a variety of projections, such as equal distance projections and Lambert conformal conic (a projection used by Geoscience Australia for large-scale mapping), should be tested in the future to quantify their effects on the prediction accuracy.

Sample stratification: Sample stratification by geomorphic provinces generally reduced the accuracy of the methods. Geomorphic provinces should not be used as secondary information or as a factor to stratify the samples if a method uses bathymetry as its secondary variable.

Searching neighbourhood: Given that we have only considered two levels of searching window size in this study, we recommend that more levels should be tested to find an optimal size.

Sample density: No threshold in sample density (or sample size) was found in relation to the accuracy of the method in any region, suggesting that the number of samples collected so far is still below the threshold (if it exists). More samples should be collected to improve future spatial interpolation efforts.

Generalision of the effects of sample density: Within each region, as sample density increases, the accuracy of the statistical methods for spatial interpolation increases. This finding can not be generalised to compare studies from the different regions, because the accuracy may not be necessarily higher for a region with higher sample density than that for a region with lower sample density.

Secondary information: Secondary information considered in this study is only limited to bathymetry, distance-to-coast, slope, latitude and longitude. Other correlated secondary variables should be identified and employed, and thus the artefacts associated with influence of latitude and longitude could be removed from the predictions.

Resolution: The predictions in this study were based on secondary information at 0.01^{o} spatial resolution grid cells. As secondary information like bathymetry becomes available at higher resolution, such information should be used.

Machine learning methods: The combination of machine learning methods, like random forest, with other methods for spatial interpolation is novel. The findings in this study demonstrate that there is great potential in applying machine learning methods in spatial interpolation. These approachs warrant further study.

Visual examination: For spatial predictions, visual examination is an essential step to assess the predictions of a statistical method.

Quality control of MARS database: As the number of samples in the MARS database continues to increase, the quality control criteria proposed in this study should be applied to the new samples. In future studies, for methods not using bathymetry as secondary information and using geomorphic provinces to stratify samples, the criteria should be applied in the following order: within the continental AEEZ; mud content with values between 0 to 100% (of course the sum of mud, sand and gravel should equal to 100%); non-dredge; without replicates; out of 10 km buffer; and non-positive bathymetry. For methods using bathymetry as secondary information, sample stratification should not be used and the criteria 'out of 10 km buffer' should not be used.

Computational requirement: Sophisticated spatial interpolations of large datasets have significant computational requirements. Even for a section of the continental AEEZ like northeast region we had to use servers with Linux operation system to accomplish the computation for some methods. Ensuring that sufficient computational resources are available for completing the modelling is essential, particularly when applying to a data set that spans an entire continent, as was the case in this experiment. Adoption of dual operating systems (*i.e.*, Windows and Linux) on high performance computers would help to address such computational constraints.

Directions and requirements for future studies: Although this simulation experiment has identified a few more robust statistical methods for spatial interpolation than the method applied widely to Geoscience Australia's datasets, it is just a beginning in the search for statistical methods to optimise spatial interpolation. This study has laid a solid scientific foundation for future studies. A number of issues and potential new combinations have been identified for future study. To optimise the predictions, further experiments need to be carried out to examine the effects of the factors identified. The combined appraoches of machine learning methods and other existing spatial interpolation techniques should be one direction of future studies selecting statistical methods for spatial interpolation. Machine learning methods, like RT, random forest and SVM, have great potential in spatial interpolation, especially in combination with other methods. Further studies are warranted for testing their performance in different scenarios. To achieve the optimised spatial interpolation of environmental variables for the continental AEEZ, more than one experiment is needed.

Division of the continental AEEZ: As we have discussed previously, regions are different in many aspects. Consequently, the most robust statistical method for spatial interpolation in each region is also different as we have observed in this study. Therefore, we suggest that region specific factors need to be taken into account in spatial interpolation of marine environmental variables in the continental AEEZ. Heap and Harris (2008) divided the Australian margin in the continental AEEZ into eight regions based on geomorphology. Physical variables display different response patterns along either latitude or longitude in regions with different oritations. For example, in the north region, as latitude increases from north to south, seabed mud content is expected to decrease but there is little relation with longitude; while in the southwest region, the seabed mud content decreases as longitude increases, but little relation is expected with latitude. In the northeast region, the relation of seabed mud content with both latitude and longitude is expected. Taking this into account we suggest that the Australian margin in the continental AEEZ should at least be divided into 12 regions in future study of statistical methods for spatial interpolation (Fig. 5.1).

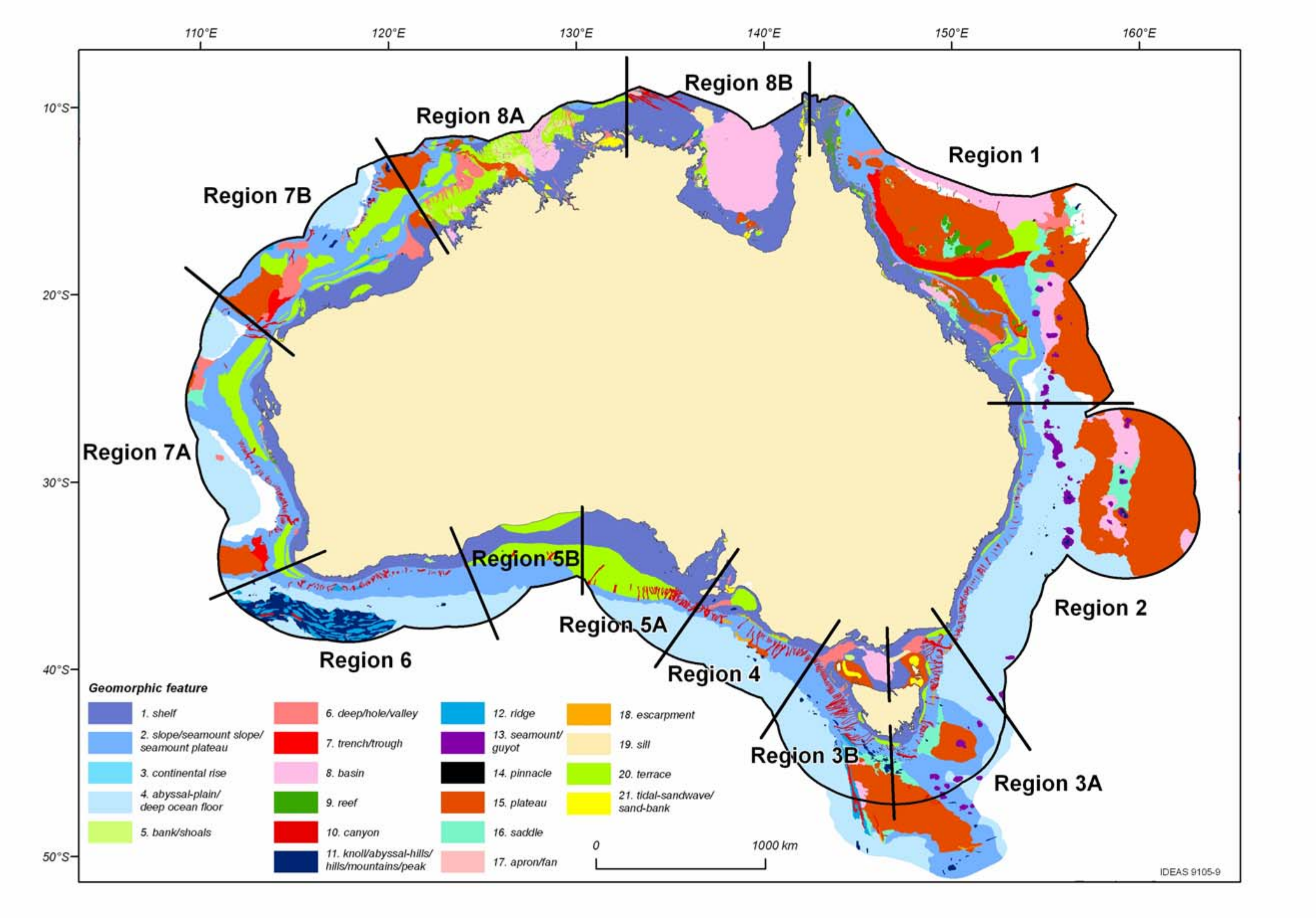


Figure 5.1. Twelve regions are defined for future study in the application of statistical methods for spatial interpolation. This figure was modified based on Heap and Harris (2008).

Acknowledgements

We would like to thank Peter Harris and Augusto Sanabria for their critical comments on this record. Christina Baker, Shoaib Burq, Mark Webster, and Tanya Whiteway are appreciated for preparing datasets. Chris Lawson is particularly appreciated for providing bathymetry data and producing several maps. Scott Nichol, David Ryan, and Frederic Saint-Cast are appreciated for providing suggestions and comments on the experimental design. We also thank Roger Bivand (Norwegian School of Economics and Business Administration), Paul Hiemstra (University of Utrecht), Edzer Pebesma (University of Münster) and Michael Sumner (University of Tasmania) for their help in using various functions in *gstat*, *maptools* and *sp* packages in *R*. This record is published with permission of the Acting Chief Executive Officer, Geoscience Australia.

References:

Asli, M. and Marcotte, D., 1995. Comparison of approaches to spatial estimation in a bivariate context. Mathematical Geology, 27: 641-658.

Bivand, R.S., Pebesma, E.J. and Gómez-Rubio, V., 2008. Applied Spatial Data Analysis with R. Use R! Springer, New York, 374 pp.

Breiman, L., 2001. Random forests. Machine Learning, 45: 5-32.

Brus, D.J. et al., 1996. The performance of spatial interpolation methods and choropleth maps to estimate properties at points: a soil survey case study. Environmetrics, 7: 1-16.

Burrough, P.A. and McDonnell, R.A., 1998. Principles of Geographical Information Systems. Oxford University Press, Oxford, 333 pp.

Collins, F.C. and Bolstad, P.V., 1996. A comparison of spatial interpolation techniques in temperature estimation, Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, NM. Santa Barbara, CA: National Center for Geographic Information and Analysis, Santa Barbara.

Cortes, C. and Vapnik, V., 1995. Support-vector networks. Machine Learning, 20: 273-297.

Dambolena, I.G., Eriksen, S.E. and Kopcso, D.P., 2009. Logarithmic transformations in regression: do you transform back correctly? Primus, 19: 280-290.

Englund, E., Weber, D. and Leviant, N., 1992. The effects of sampling design parameters on block selection. Mathematical Geology, 24: 329-343.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York, 483 pp.

Goovaerts, P., 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. Journal of Hydrology, 228: 113-129.

Hartkamp, A.D., De Beurs, K., Stein, A. and White, J.W., 1999. Interpolation Techniques for Climate Variables, CIMMYT, Mexico, D.F.

Heap, A.D. and Harris, P.T., 2008. Geomorphology of the Australian margin and adjacent seafloor. Australian Journal of Earth Sciences, 55: 555-585.

Hemer, M.A., 2006. The magnitude and frequency of combined flow bed shear stress as a measure of exposure on the Australian continental shelf. Continental Shelf Research, 26: 1258-1280.

Hengl, T., 2007. A Practical Guide to Geostatistical Mapping of Environmental Variables, Luxembourg: Office for Official Publication of the European Communities.

Department of the Environment and Heritage, 2005. National marine bioregionalisation of Australia 2005 (National Oceans Office), Hobart.

Hernandez-Stefanoni, J.L. and Ponce-Hernandez, R., 2006. Mapping the spatial variability of plant diversity in a tropical forest: comparison of spatial interpolation methods. Environmental Monitoring and Assessment, 117: 307-334.

Hu, K., Li, B., Lu, Y. and Zhang, F., 2004. Comparison of various spatial interpolation methods for non-stationary regional soil mercury content. Environmental Science, 25: 132-137.

Isaaks, E.H. and Srivastava, R.M., 1989. Applied Geostatistics. Oxford University Press, New York, 561 pp.

Jarvis, C.H. and Stuart, N., 2001. A comparison among strategies for interpolating maximum and minimum daily air temperature. Part II: the interaction between number of guiding variables and the type of interpolation method. Journal of Applied Meteorology, 40: 1075-1084.

Journel, A.G. and Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, London, 600 pp.

Laslett, G.M., 1994. Kriging and splines: an empirical comparison of their predictive performance in some applications. Journal of the American Statistical Association, 89: 391-400.

Laslett, G.M., McBratney, A.B., Pahl, P.J. and Hutchinson, M.F., 1987. Comparison of several spatial prediction methods for soil pH. Journal of Soil Science, 38: 325-341.

Legendre, P. and L., L., 1998. Numerical Ecology. ELSEVIER, Amsterdam, 853 pp.

Li, J. and Heap, A., 2008. A Review of Spatial Interpolation Methods for Environmental Scientists. Record 2008/23, Geoscience Australia, Canberra.

Li, J. and Heap, A., 2009. A review of comparative studies of spatial interpolation methods: performance and impact factors. Ecological Informatics (submitted).

Li, Y., Shi, Z., Wu, C., Li, H. and Li, F., 2007. Improved prediction and reduction of sampling density for soil salinity by different geostatistical methods. Agricultural Science in China, 6: 832-841.

Martínez-Cob, A., 1996. Multivariate geostatistical analysis of evapotranspiration and precipitation in mountainous terrain. Journal of Hydrology, 174: 19-35.

Odeh, I.O.A., McBratney, A.B. and Chittleborough, D.J., 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma, 67: 215-226.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Computer & Geosciences, 30: 683-691.

Pebesma, E.J. and Kwaadsteniet, D., 1997. Mapping groundwater quality in the Netherlands. Journal of Hydrology, 200: 364-386.

Pinheiro, J.C. and Bates, D.M., 2000. Mixed-Effects Models in S and S-PLUS. Springer, New York, 528 pp.

Robertson, G.P. et al., 1997. Soil resources, microbial activity, and primary production across an agricultural ecosystem. Ecological Applications, 7: 158-170.

Schloeder, C.A., Zimmerman, N.E. and Jacobs, M.J., 2001. Comparison of methods for interpolating soil properties using limited data. Soil Science Society of American Journal, 65: 470-479.

Schuurmans, J.M., Bierkens, M.F.P. and Pebesma, E.J., 2007. Automatic prediction of high-resolution daily rainfall fields for multiple extents: the potential of operational radar. Journal of Hydrometeorology, 8: 1204-1224.

Shortridge, A., 2008. nscore.R. http://www.msu.edu.

Stahl, K., Moore, R.D., Floyer, J.A., Asplin, M.G. and McKendry, I.G., 2006. Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density. Agricultural and Forest Meteorology, 139: 224-236.

Statnikov, A., Wang, L. and Aliferis, C.F., 2008. A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. BMC Bioinformatics, 9: 319.

Stein, A., Hoogerwerf, M. and Bouma, J., 1988. Use of soil map delineations to improve (co-)kriging of point data on moisture deficits. Geoderma, 43: 163-177.

Strobl, C., Boulesteix, A., Zeileis, A. and Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics, 8: 25.

R Development Core Team, R.D.C., 2007. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Venables, W.N. and Ripley, B.D., 2002. Modern Applied Statistics with S-Plus. Springer-Verlag, New York, 495 pp.

Verfaillie, E., van Lancker, V. and van Meirvenne, M., 2006. Multivariate geostatistics for the predictive modelling of the surficial sand distribution in shelf seas. Continental Shelf Research, 26: 2454-2468.

Voltz, M. and Webster, R., 1990. A comparison of kriging, cubic splines and classification for predicting soil properties from sample information. Journal of Soil Science, 41: 473-490.

Wang, H., Liu, G. and Gong, P., 2005. Use of cokriging to improve estimates of soil salt solute spatial distribution in the Yellow River delta. Acta Geographica Sinica, 60: 511-518.

Webster, R. and Oliver, M., 2001. Geostatistics for Environmental Scientists. John Wiley & Sons, Ltd, Chichester, 271 pp.

Whiteway, T. et al., 2007. Seascapes of the Australian Margin and Adjacent Sea Floor: Methodology and Results, Geoscience Australia.

Willmott, C.J., 1982. Some comments on the evaluation of model performance. Bulletin American Meteorological Society, 63: 1309-1313.

Wolpert, D. and Macready, W., 1997. No free lunch theorems for optimization. IEEE Transactions on Evolutionary Computation, 1: 67-82.

Zimmerman, D., Pavlik, C., Ruggles, A. and Armstrong, M.P., 1999. An experimental comparison of ordinary and universal kriging and inverse distance weighting. Mathematical Geology, 31: 375-390.

Appendix A. Description of machine learning methods

A GRNN model, a version of Probabilistic Neural Network (Specht, 1990) for regression problems, has an input layer, a hidden layer, a pattern layer and output layer. The basic idea is that a predicted target value of a point is likely to be about the same as some input points that have similar values of the predictor variables. The distance is computed from the point being evaluated to each of the other points, and a radial basis function (RBF) is applied to the distance to compute the weight (influence) for each point. The further an input point is from the new point, the less influence it has. The best predicted value for the new point is found by summing the values of the other points weighted by the RBF function. It is much faster to train a GRNN network than a traditional Multilayer Perceptron neural network. It is relatively insensitive to outliers, but slower than Multilayer Perceptron neural network at classifying new cases and requires more computer memory.

A SVM performs classification by constructing an N-dimensional hyperplane that optimally separates the data into two categories (Cortes and Vapnik, 1995). The goal of SVM modeling is to find the optimal hyperplane that separates clusters of vectors in such a way that cases with one category of the target variable are on one side of the plane and cases with the other category are on the other size of the plane. The vectors near the hyperplane are the support vectors. An SVM analysis finds the hyperplane that is oriented so that the margin between the support vectors is maximized. SVM can also be used for regression problems.

A RT is a binary tree which consists of internal and terminal nodes. It relies on measures of “goodness of split” (splitting rules) to select an appropriate predictor variable on each internal node to split the branches until all nodes are terminals. After building the initial tree, pruning algorithms are often applied to simplify the tree for better prediction capability. RT is able to produce hyperplanes for discriminating classes through the path described from the root to a terminal node. This study used the regression mode of the CART (classification and regression tree) decision tree (Breiman *et al.*, 1984) for the regression problem.

Random forest is an ensemble method that combines many individual classification trees in the following way: From the original sample many bootstrap samples are drawn, and an unpruned RT is fit to each bootstrap sample. From the complete forest the status of the response variable is predicted as an average of the predictions of all trees (Breiman, 2001; Strobl et al., 2007).

Appendix B. Simulation Modelling

In this appendix we discuss a number of issues in relation to variogram modelling and simulation modelling. As indicated by Li and Heap (2008), geostatistics has a few assumptions and limitations. Geostatistical methods assume stationarity of data, which is usually not true, although this assumption can be relaxed with specific forms of kriging. The definition of the required variogram model is time consuming and somewhat subjective; and definition of neighbourhoods is also required and is difficult to do objectively. Geostatistical methods also assume that the data are isotropic.

B.1. Data transformation

Geostatistical methods and linear regression models (lm) assume data stationarity of the primary variable. This assumption is also necessary for secondary variables when OCK is applied because in OCK secondary variables are modelled as if they were primary variable. Distributions of mud content, bathymetry, distance-to-coast and slope are left skewed and non-normal (Figs. B.1 and B.2), so several transformations were tested and the appropriate transformation was identified for each variable in each region (Table B.1). The distributions of the transformed data for most variables are approximately normal, but not for all (Figs. B.1 and B.3). However, that is the best transformation we can get. In the application of OCK, bathymetry and slope were double-square rooted, because the selected transformation, square root and log, would yield wrong results for variogram modelling. Mud content was also transformed to between 0 and 1 for generalised linear models.

Table B.1. Data transformation of mud content, bathymetry, distance-to-coast and slope in the three study regions.

Region	Mud content	Bathymetry	Distance-to-coast	Slope
North	square root	double square root	double square root	square root and log
Northeast	square root	square root and log	double square root	double square root
Southwest	arcsine	double square root	double square root	square root and log

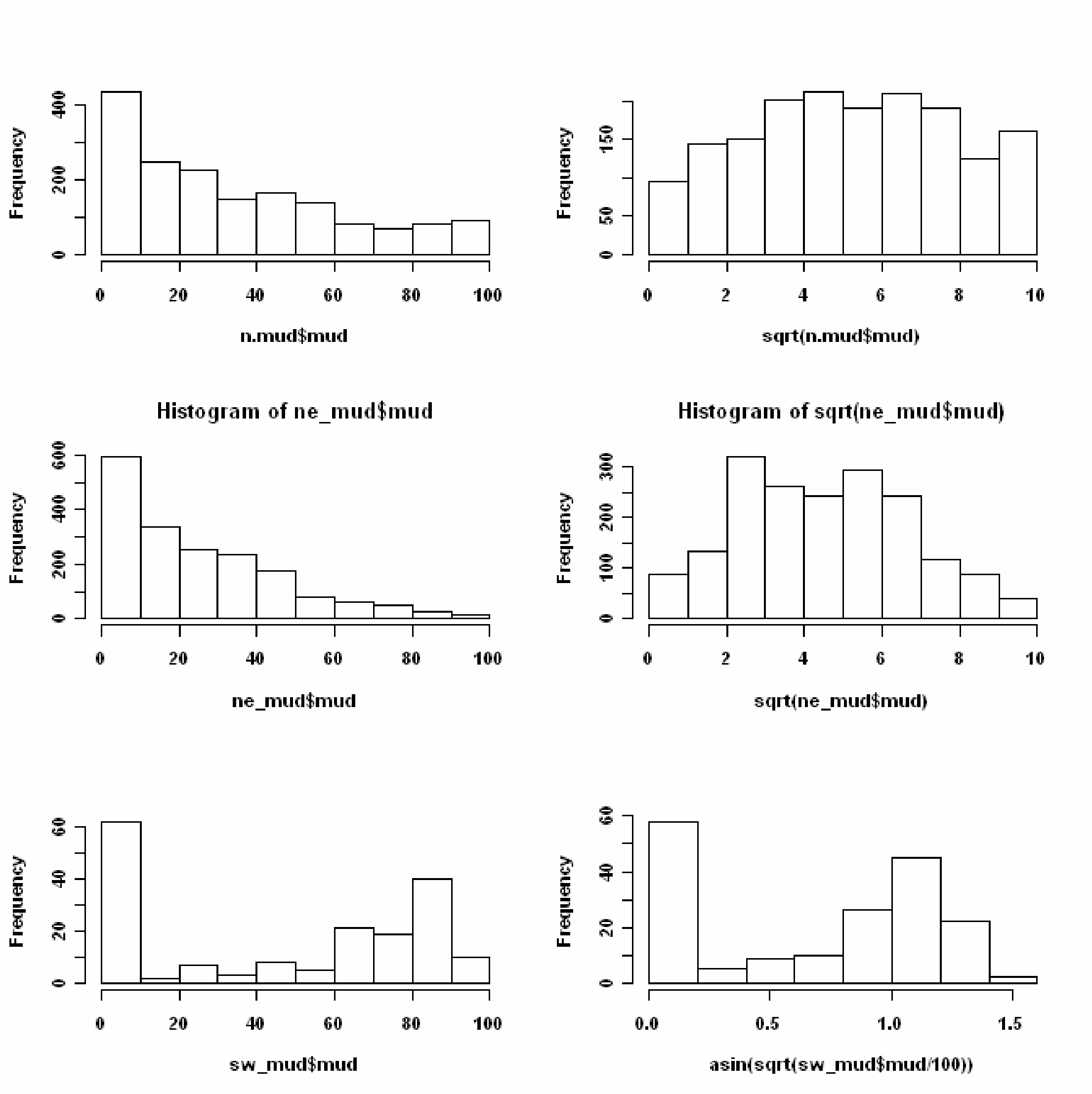


Figure. B.1. Data distribution of mud content in the three study regions before and after transformation.

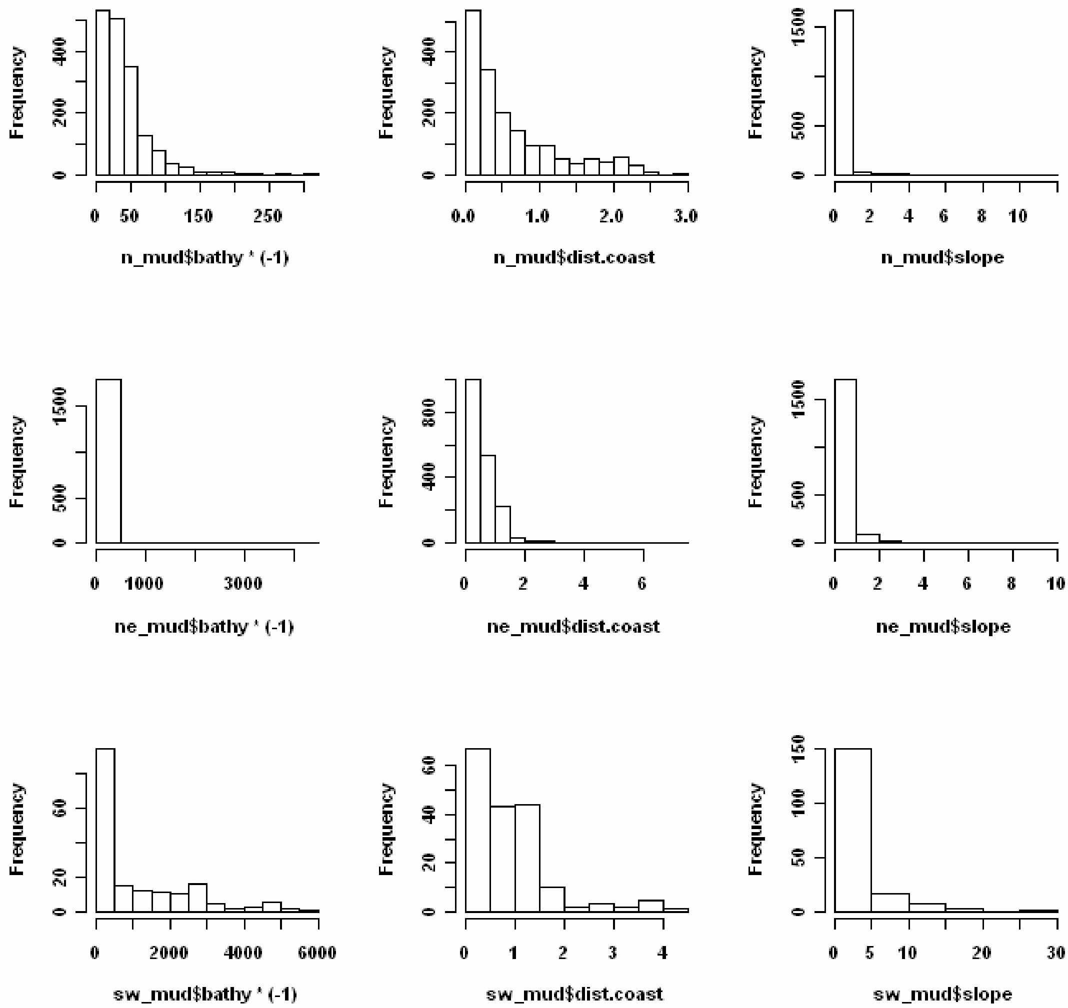


Figure. B.2. Data distribution of bathymetry, distance-to-coast and slope in the three study regions.

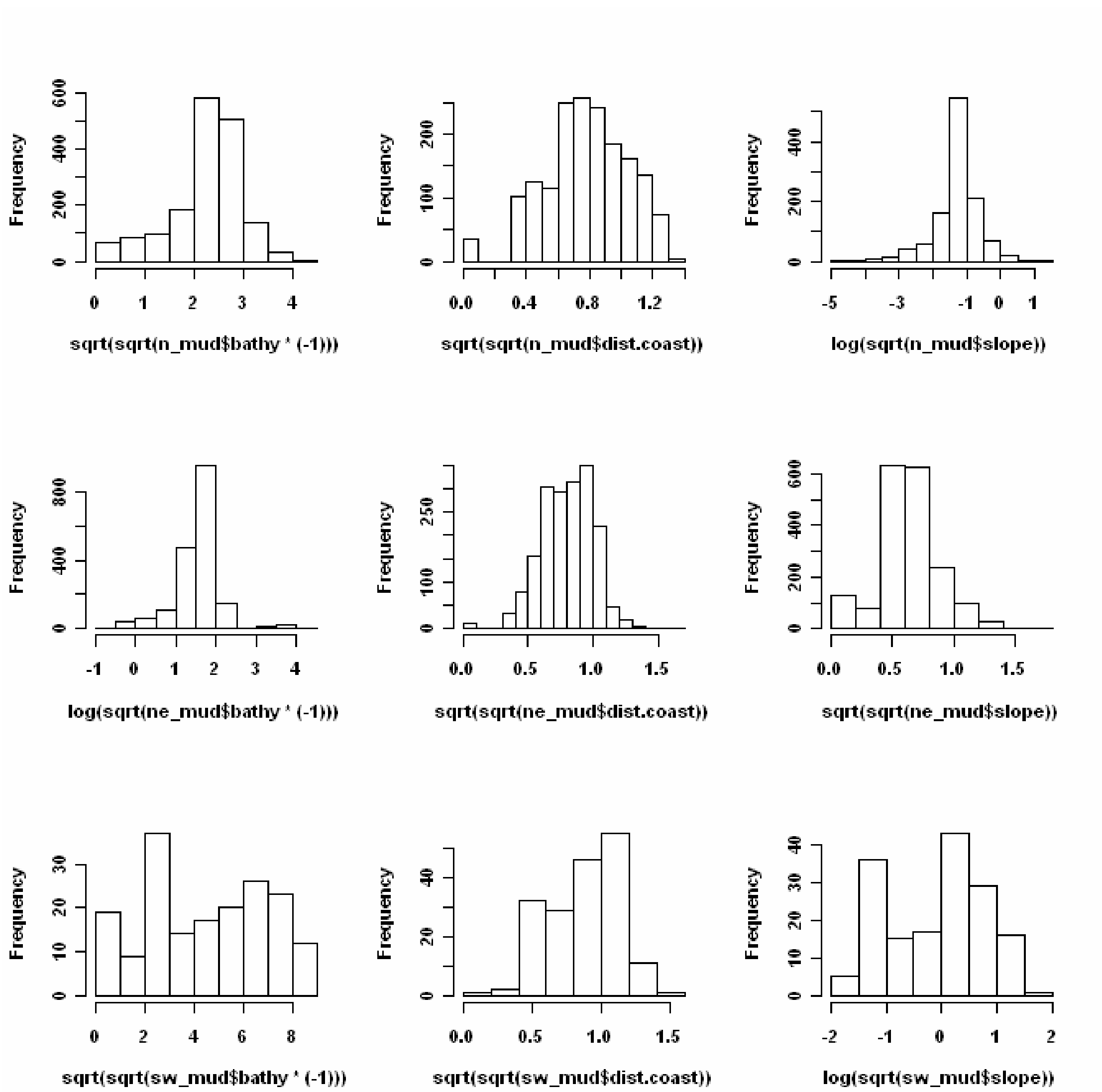


Figure. B.3. Data distribution of bathymetry, distance-to-coast and slope in the three study regions after data transformation.

B.2. Correlation between mud content and secondary variables

Correlation between the primary and secondary variables is critical for the spatial interpolation methods that use auxiliary information. As the correlation increases, the information brought from the secondary variable on to the primary value increases (Goovaerts, 1997). In this study, the correlation of mud content and the secondary variables changes with variables and regions in terms of Pearson's product-moment correlation (*r*) (Table B.2) and Spearman's rank correlation rho (ρ) (Table B.3). Bathymetry has a high correlation with mud content, particularly in the southwest region (Tables B.2 and B.3, and Fig. B.4). Distance-to-coast highly correlated with mud content in the southwest region (Tables B.2 and B.3, and Fig. B.5) in terms of r and in all three regions in terms of ρ. Slope displayed a similar pattern as distance-to-coast in terms of *r* and ρ, with an exception of ρ in the northeast region (Tables B.2 and B.3, and Fig. B.6). In most cases, the relationship between mud content and the secondary variables are non-linear. In both variogram and simulation modelling, either bathymetry or all three secondary variables were used depending on statistical method.

Table B.2. Pearson's product-moment correlation of mud content with bathymetry, distance-to-coast and slope in the three study regions. The test for correlation between paired samples was conducted in R (R Development Core Team, 2007).

Variable	Region	t value	Degree of freedom	Correlation coefficient	p-value
Bathymetry	North	-8.445	1685	-0.2015	0
	Northeast	-12.5107	1826	-0.2810	0
	Southwest	-10.6435	175	-0.6268	0
Distance-to-coast	North	1.4775	1685	0.0360	0.1397
	Northeast	1.2844	1826	0.0300	0.1992
	Southwest	8.3233	175	0.5325	0
Slope	North	-3.1408	1685	-0.0763	0.0017
	Northeast	0.6062	1826	0.0142	0.5444
	Southwest	6.1715	175	0.4228	0

Table B.3. Spearman's rank correlation rho of mud content with bathymetry, distance-to-coast and slope in the three study regions. The test for correlation between paired samples was conducted in R (R Development Core Team, 2007).

Variable	Region	rho	p-value
Bathymetry	North	-0.3360	0
	Northeast	-0.0914	0.0001
	Southwest	-0.6286	0
Distance-to-coast	North	0.0864	0.0004
	Northeast	-0.1108	0
	Southwest	0.6171	0
Slope	North	-0.1573	0
	Northeast	0.0356	0.1277
	Southwest	0.5522	0

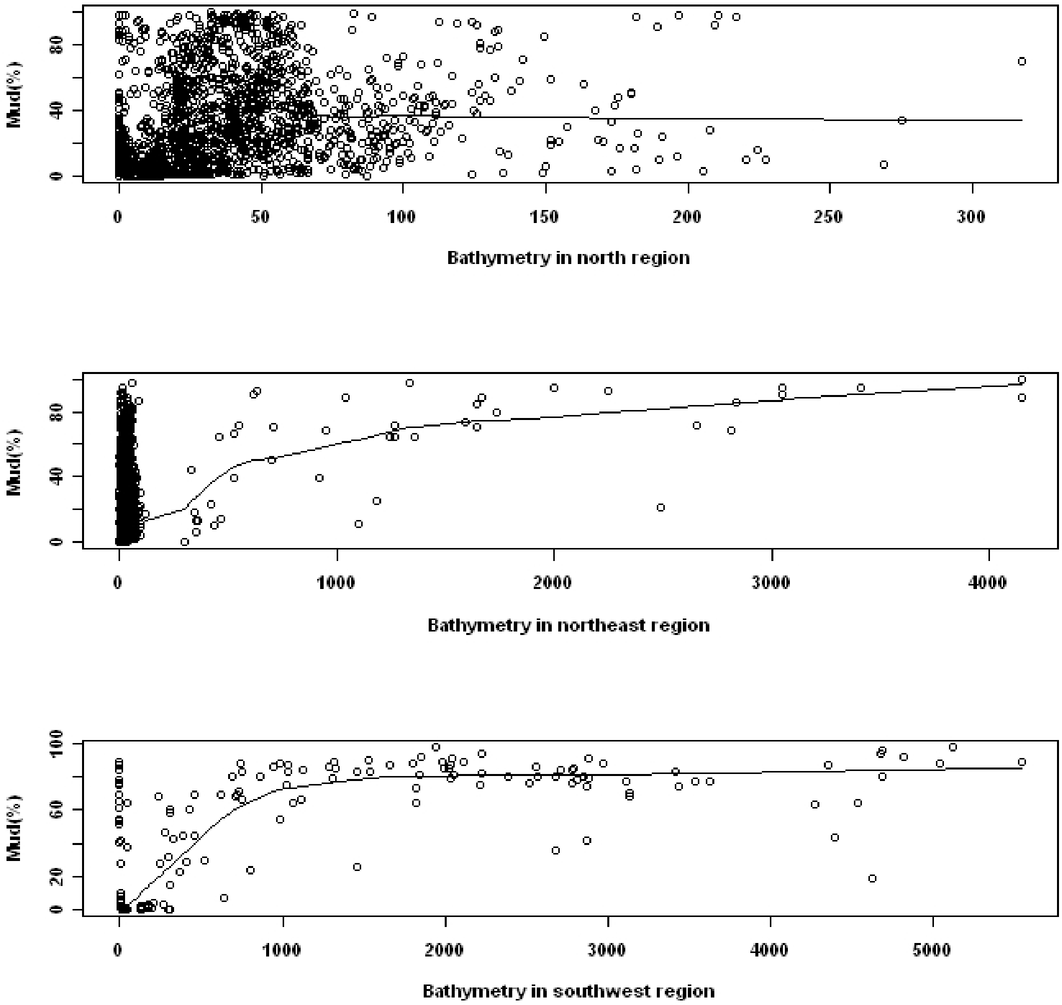


Figure. B.4. Relation between mud data and bathymetry in the three study regions and the curve was fitted using lowess in *R* (R Development Core Team, 2007).

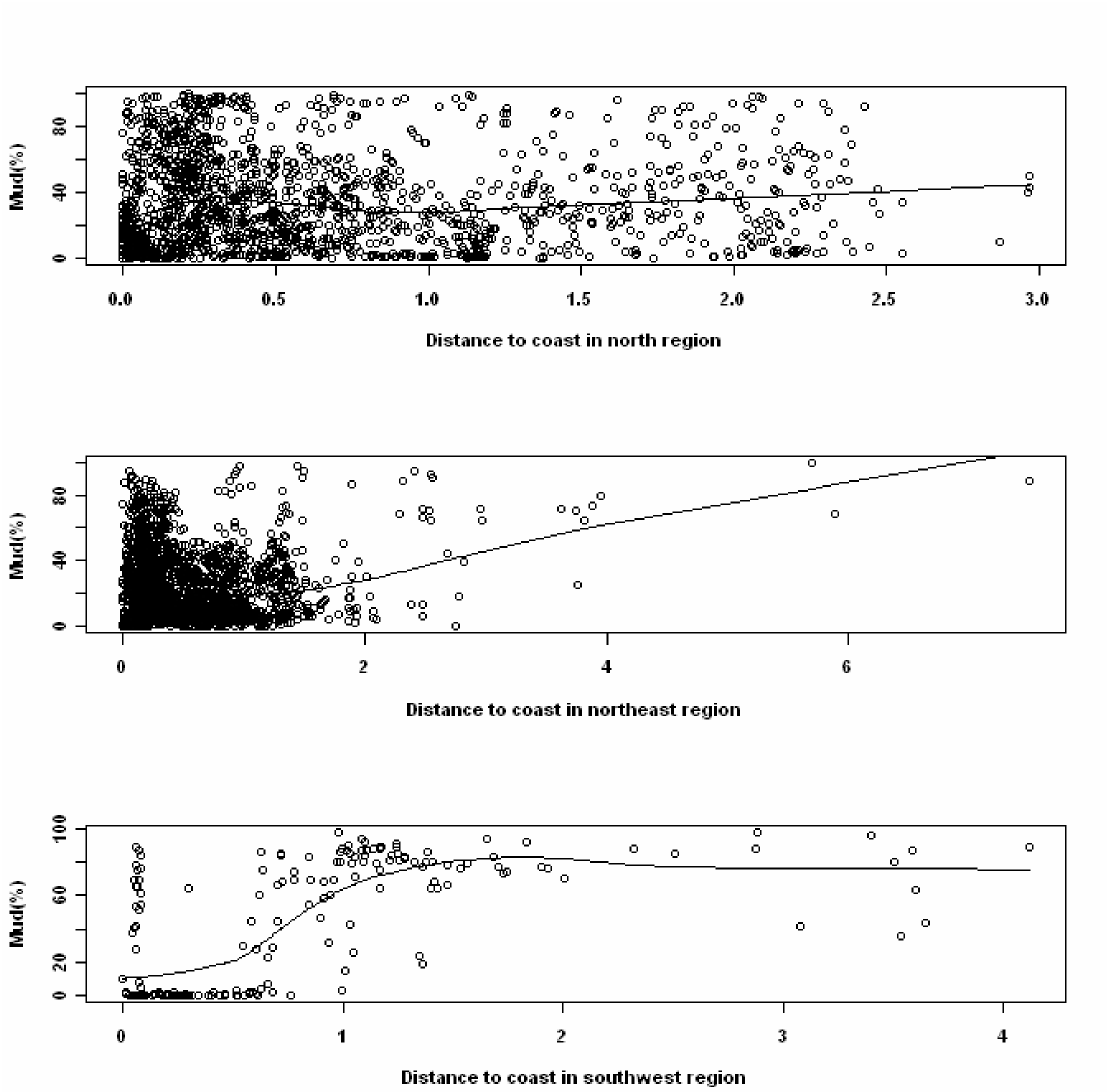


Figure. B.5. Relation between mud data and distance-to-coast in the three study regions and the curve was fitted using lowess in *R* (R Development Core Team, 2007).

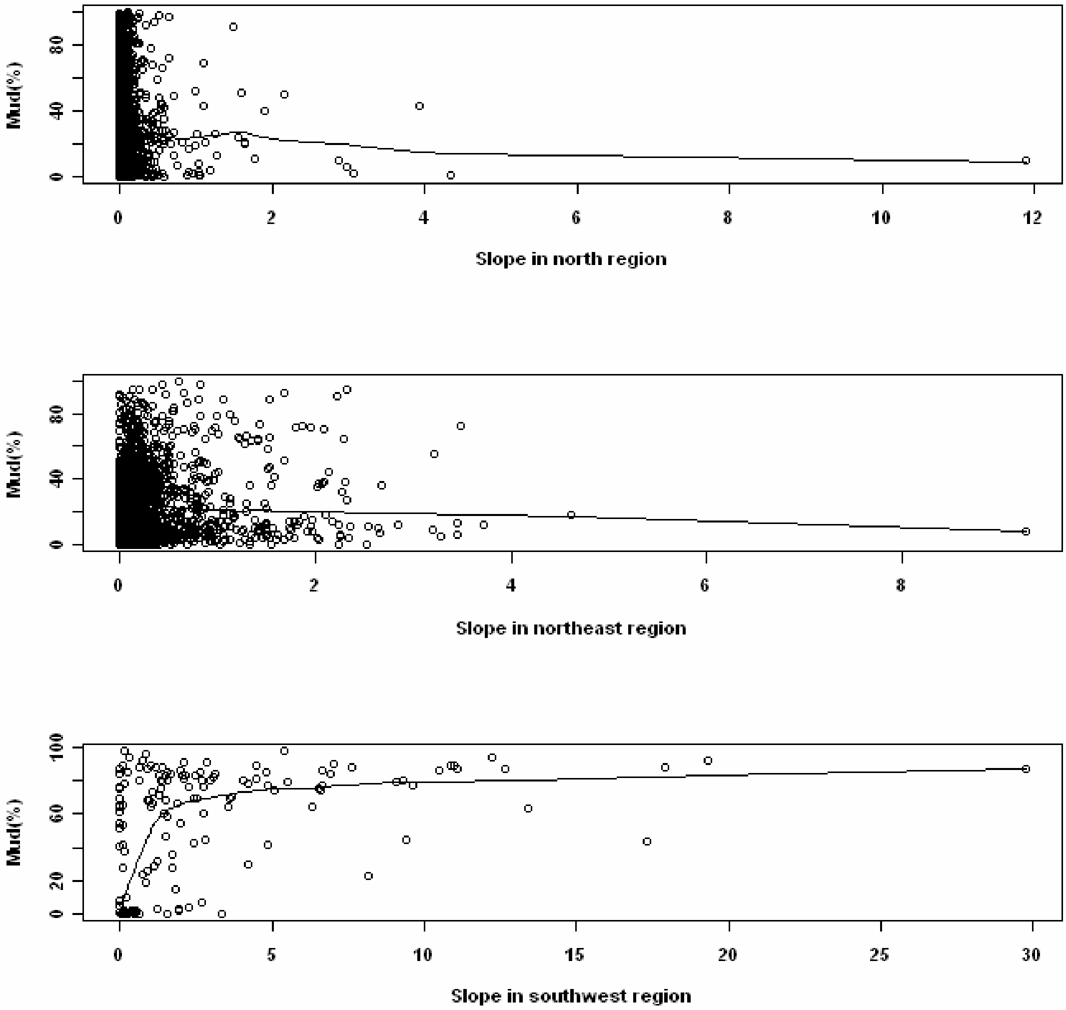


Figure. B.6. Relation between mud data and slope in the three study regions and the curve was fitted using lowess in *R* (R Development Core Team, 2007).

B.3. Variogram Modelling

Variogram modelling is an essential component in geostatistics. The result of variogram modelling depends on the selection of data projection, variogram models and data nature.

B.3.1. Data projection

The coordinates of mud data were in latitude and longitude based on WGS84 without any projection in this study. Geostatistical methods such as those in the gstat package (Pebesma, 2004) in *R* (R Development Core Team, 2007) expect the input data having been appropriately projected. That is the data should be projected using equal distance projection, and projections are set up for regions on the earth to go from longitude and latitude to new coordinates for which a unit difference in x axis reflects approximately the same distance in y axis. Without knowing coordinates are in longitude and latitude, gstat assumes they are in some planar system and a unit difference in longitude reflects approximately the same distance in latitude and ignores the changes in distance along the latitude, meaning that it can use Pythagoras to compute a distance between two points in variogram modelling and kriging (personal communication with Edzer Pebesma in May 2009). Although gstat does use spherical distances when data are in geographical coordinates, the usual variogram models are typically not non-negative definite on the sphere, and no appropriate models are available (personal communication with Edzer Pebesma in May 2009).

For longitude and latitude, interpolating with gstat we have the following choices:
1) interpolate in longlat, but don't set proj4string; gstat will assume one unit long equals one unit lat (use Pythagoras);
2) interpolate in longlat, but set proj4string to CRS("+proj=longlat") (CRS is coordinate reference system); gstat will use great circle distances, but doesn't have the right covariance functions for this; and
3) reproject the data to a new CRS, and interpolate; this is usually done. It was recommended to work in a projected system (option 3), and not in longitude and latitudet as long as we don't have good variogram models implemented in gstat, or somewhere else (personal communication with Edzer Pebesma in May 2009).

Although we were aware of the above mentioned limitations in using gstat, we still decided to model our data in WGS84 by using option 1. The reasons are as bellow: firstly we have to make interpolation for an area spanning many UTM zones, and secondly we have tried some equal distance projections to convert our data into meters for the continental AEEZ, but it seems that none of them can produce satisfactory projection for the purpose of this study, although sinusoidal gave slightly better results in comparison with UTM projection that was used as a control. From ten-fold cross validation results, MAE for most of the methods that used data (with mud content ranging from 0 to 100%) in WGS84 varies between 9-13(%). Lastly, we compared the first two options by applying OK to 50 datasets in the north region, an area spanning ca. 20 degrees in terms of longitude from 9 to 18 degrees south in latitude. OK with CRS specified gave 'Warning: singular model in variogram fit' for three datasets and the consequent predictions are poor. After removing the results for these datasets, the average MAE and RMSE for the remaining 47 datasets were 12.986 and 18.335 respectively for OK with CRS specified and 12.977 and 18.326 respectively for OK without CRS specification. It seems that the difference in terms of MAE and RMSE is really marginal. Based on above reasons, we decided to make interpolation using the first option. However, we strongly recommend that projections, such as Lambert conformal conic, should be tested in the further research.

B.3.2. Variogram model

The semivariance can be estimated from the data, as follows:

$$\hat{\gamma}(h) = \frac{1}{2n}\sum_{i=1}^{n}(z(x_i) - z(x_i + h))^2 \qquad (7)$$

where n is the number of pairs of sample points (x_i, x_j) separated by distance h (Burrough and McDonnell, 1998). Variogram modelling and estimation is extremely important for structural analysis and spatial interpolation (Burrough and McDonnell, 1998). A plot of $\hat{\gamma}(h)$ against h is known as the experimental variogram (Fig. B.7), which displays several important features. The first is the “nugget”, a positive value of $\hat{\gamma}(h)$ at h close to 0, which is the residual reflecting the variance of sampling errors and the spatial variance at shorter distance than the minimum sample spacing. The "range" is a value of distance at which the "sill" is reached. Samples separated by a distance larger than the range are spatially independent because the estimated semivariance of differences will be invariant with sample separation distance. If the ratio of sill to nugget is close to 1, then most of the variability is non-spatial (Hartkamp et al., 1999). The range provides information about the size of a search window used in the spatial interpolation methods (Burrough and McDonnell, 1998). The variogram models may consist of simple models, including: Nugget, Exponential, Spherical, Gaussian, Linear, and Power model or the nested sum of one or more simple models (Burrough and McDonnell, 1998; Li and Heap, 2008; Pebesma, 2004; Webster and Oliver, 2001).

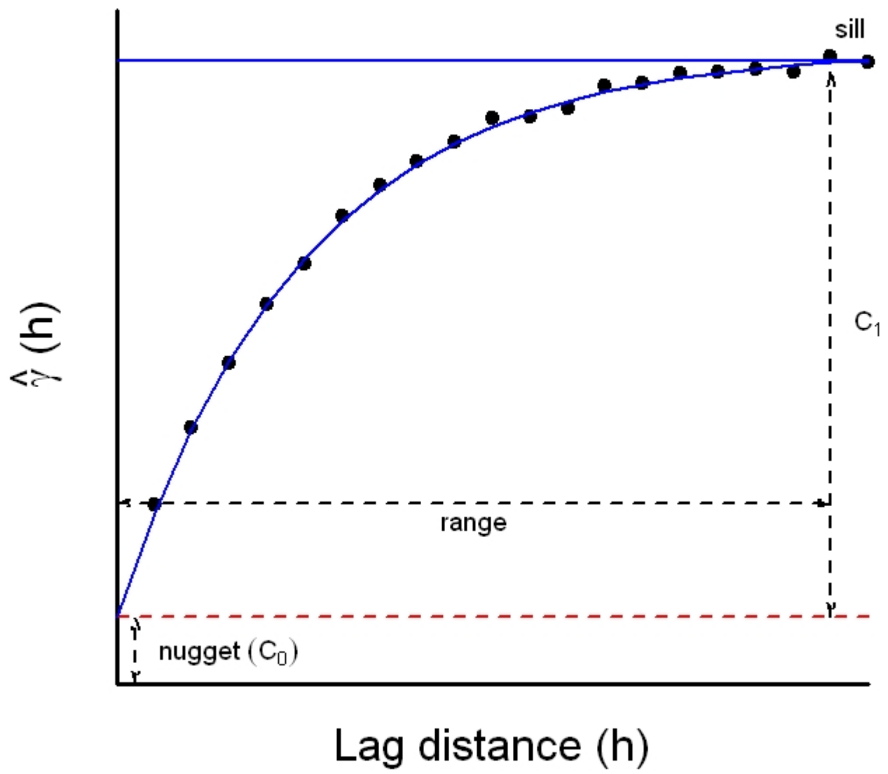


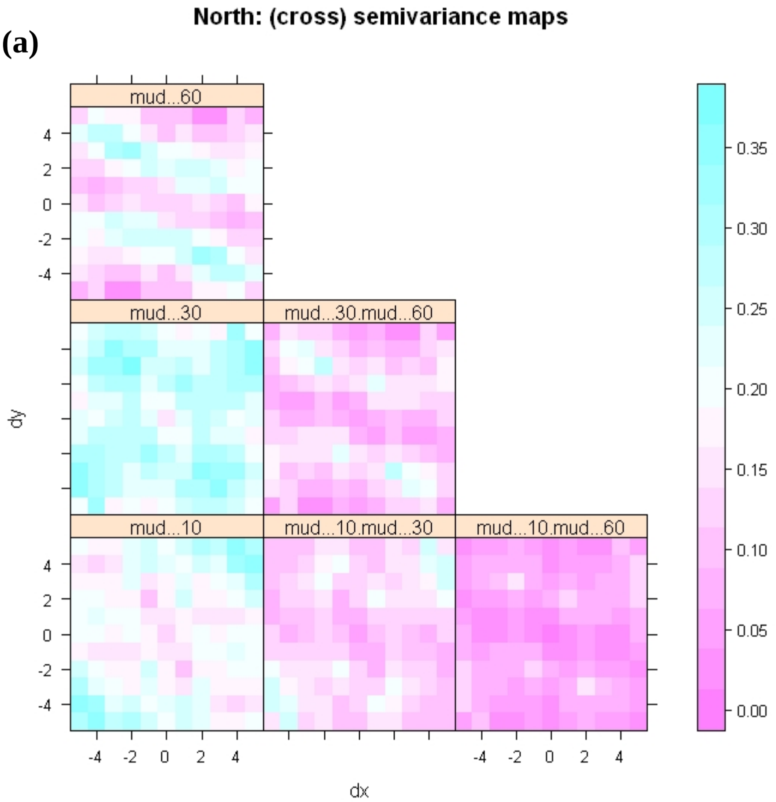
Figure B.7. An example of a semivariogram as illustrated by an exponential model, with range, nugget (C_0) and sill (C_0+C_1) (Li and Heap, 2008).

The structural variance, which determines the variance due to spatial dependence explained by the variogram model, is calculated as the difference of total variance and nugget variance divided by the total variance (Hernandez-Stefanoni and Ponce-Hernandez, 2006). The relative "noisy" nature of the spatial variability is represented by the values of the nugget variance (Hernandez-Stefanoni and Ponce-Hernandez, 2006). The nugget variance may result from 1) the sampling error, and 2) the spatial dependence that may exist at finer scales than the minimum separation distance between samples (Hernandez-Stefanoni and Ponce-Hernandez, 2006). The ratio of nugget to sill reflects the spatial heterogeneity of the data (Robertson et al., 1997; Wang et al., 2005). If the ratio is big, the spatial variation is mainly resulted from the random process and the measurement error is high, and if it is small, the variation is mainly due to the spatial structure.

B.3.3 Variogram anisotropy

The spatial structure of the data affects the performance of geostatistical interpolators. To test for anisotropy, semivariogram maps were generated for each of the three regions using gstat in R (R Development Core Team, 2007) (Fig. B.8). There is no obvious trend in the north and southwest regions. For data in the northeast region the maps showed that there is a weak directional changes at 45^{o}. The range at 45^{o} is about 2.5 and at 135^{o} is mostly around 2.5, so the ratio is about 1 (Fig. B.9), suggesting a weak anisotropy. This weak anisotropy was even weaker when a trend (*e.g.*, bathymetry was used as secondary variable) was considered (Figs. B.10 and B.11). Therefore, mud content was modelled as isotropic in all three regions in this study.

(a)



(b)

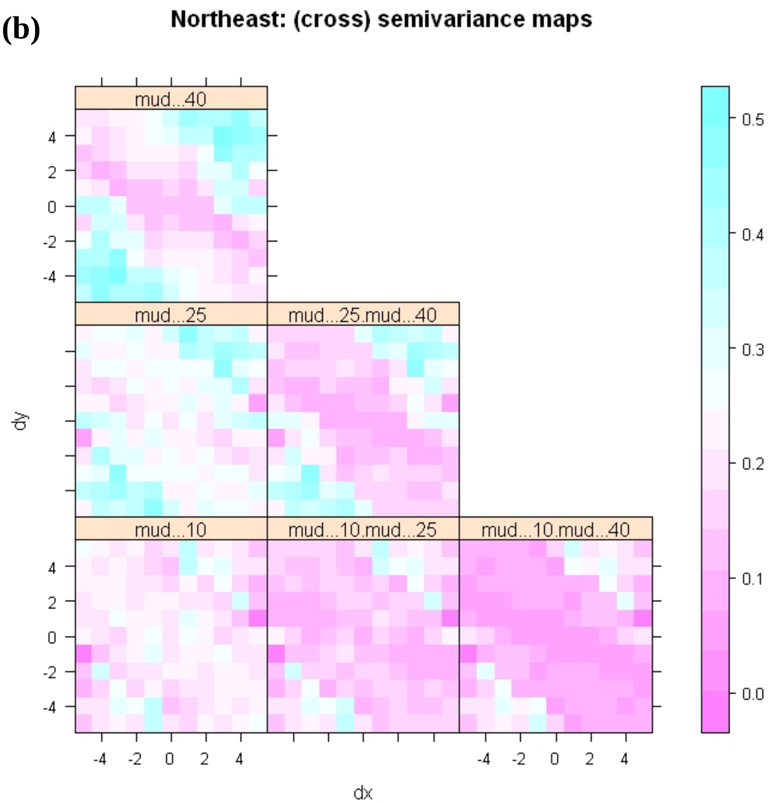
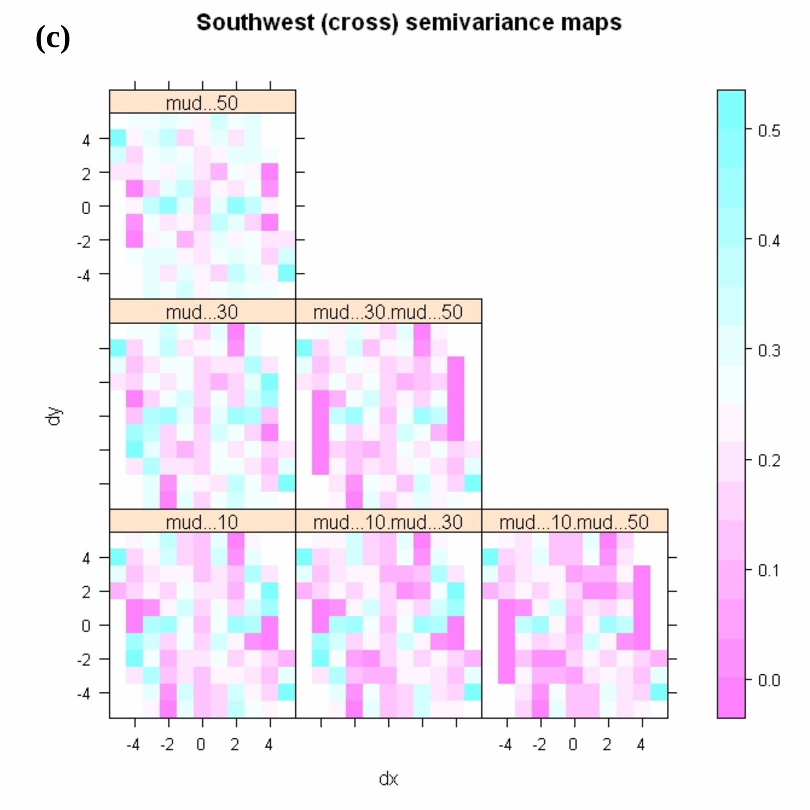


Figure B.8. Variogram maps: a) north, b) northeast and c) southwest.

Fig. B.8. (cont.):

(c)



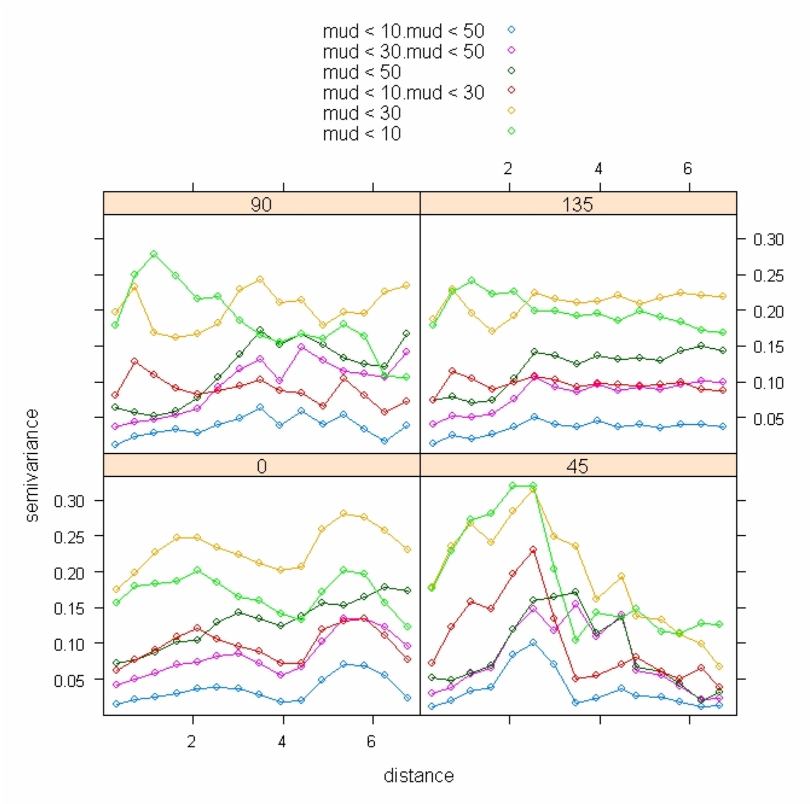


Figure B.9. Semivariance of mud data in the northeast region at different directions.

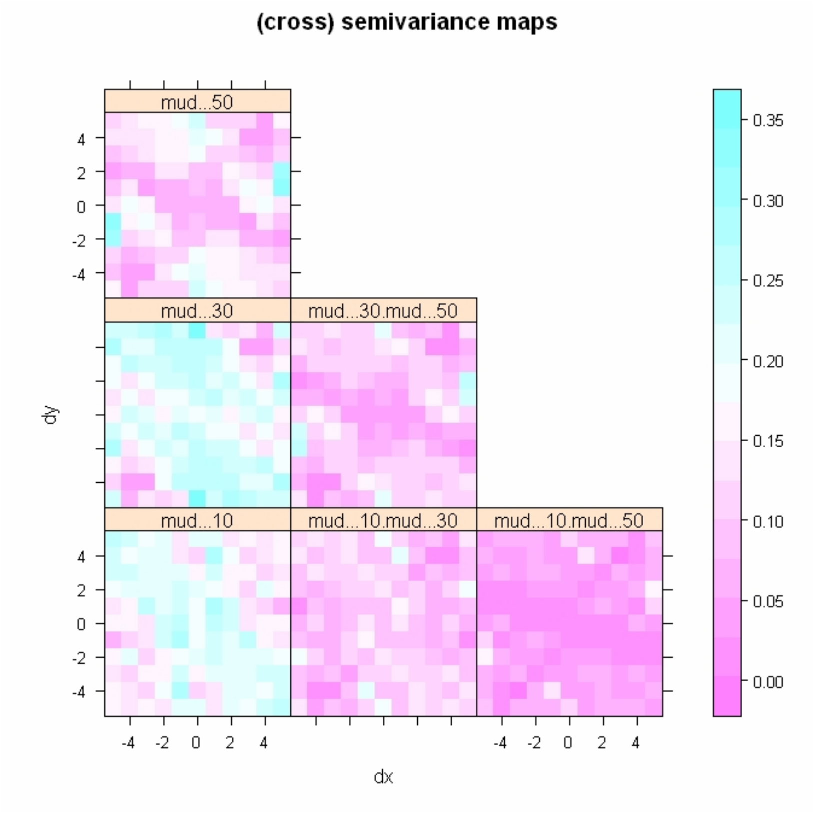


Figure B.10. Variogram map of mud content in the northeast with bathymetry as secondary information.

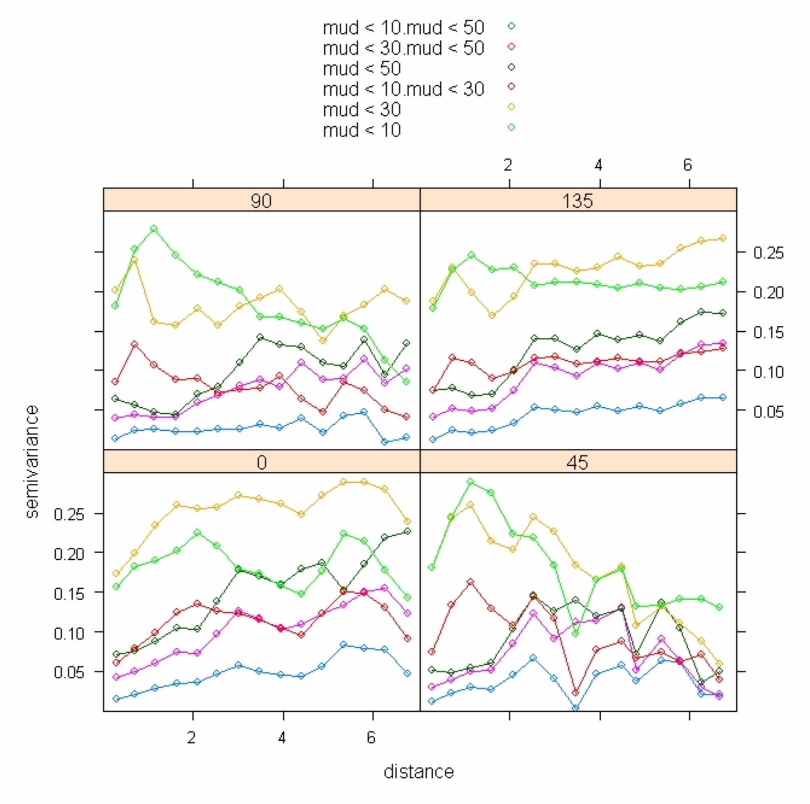
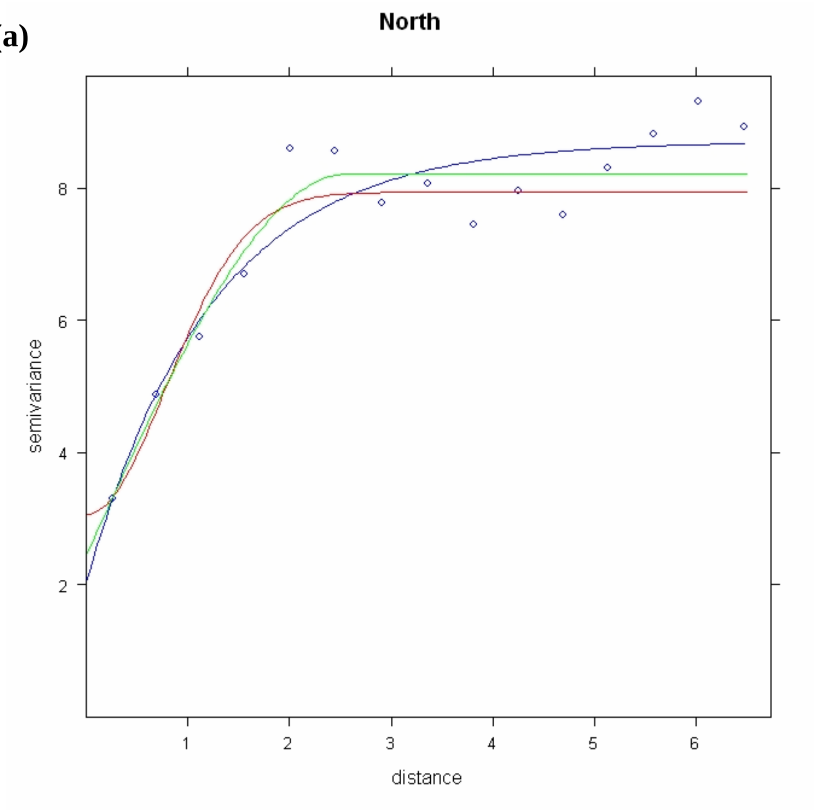


Figure B.11. Semivariance of mud data in the northeast region at different directions with bathymetry as secondary information.

B.3.4. Variogram model selection

There are a number of variogram models that can be employed; and different variogram models may lead to different interpolations (Li and Heap, 2008). Thus selecting an appropriate model to capture the features of the data is critical. In this study, variogram model was selected based on the fitted values of range, nugget and sill from a range of models including Bessel, Circular, Exponential, Exponential class, Gaussian, Linear, Logarithmic, Pentaspherical, Periodic and Spherical using gstat in R (R Development Core Team, 2007). Of these models, three models performed better than the others and then were compared in terms of range, nugget and sill (Figure B.12 and Table B.4). The Spherical model was selected for north and southwest region and the Exponential model for northeast region. After the inclusion of secondary variable(s) in the northeast region, the Spherical model fitted the data better in terms of range, nugget and sill, so it was used for methods using secondary information.

(a)



(b)

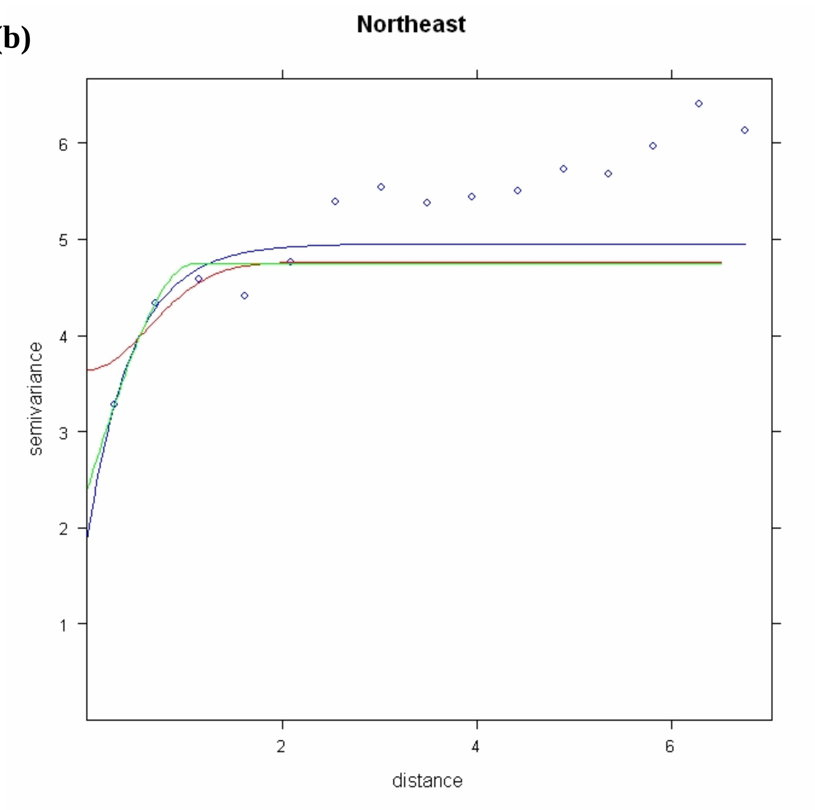


Figure B.12. Variogram models (Exponential: blue, Gaussian: red, and Spherical: green) compared and selected for each region: a) Spherical for north, b) Exponential for northeast and c) Spherical for southwest.

Fig. B.12. (cont.):

(c)

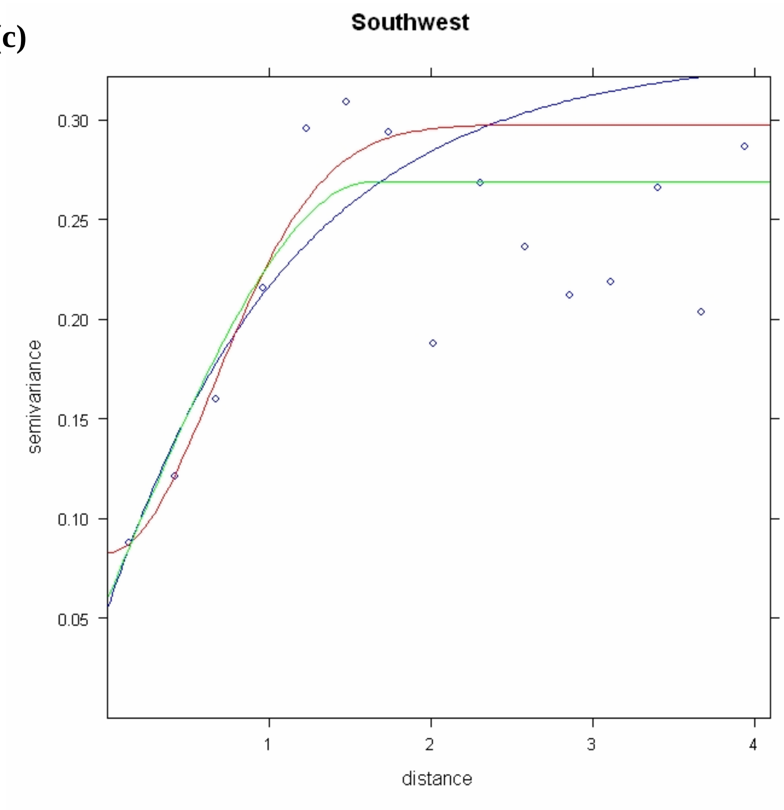


Table B.4. Fitted values by three variogram models for each region.

Region	Variogram model	Nugget	Partial sill	Range
North	Exponential	2.0386	6.667	1.2345
	Gaussian	3.0637	4.8698	1.1067
	Spherical	2.4502	5.7572	2.5639
Northeast	Exponential	1.8576	3.0964	0.46
	Gaussian	3.6398	1.1227	0.8915
	Spherical	2.3836	2.3596	1.0958
Southwest	Exponential	0.0548	0.2781	1.1453
	Gaussian	0.0828	0.2149	0.9317
	Spherical	0.0599	0.209	1.6286

B.4. Simulation modelling

B.4.1. Model specification

For each statistical method, data transformation, variogram model, size of searching neighbourhood and application of sample stratification change with study regions (Table B.5). For IDW, TPS and machine learning methods, data were not transformed; and for geostatistical methods, linear models and generalised least square methods, relevant transformation was applied to normalise the data. Data was also transformed for generalised linear model to rescale data range to 0 and 1.

A Spherical (Sph) variogram model was used for all geostatistical methods and their combinations with other methods, with an exception of OK in the northeast region.

Searching window size, that is the number of nearest samples used for making prediction, was set at two levels for most methods. One is global that means all samples were used for interpolation. Another is local, that is only 20 nearest samples were used for interpolation. However, for TPS, the window size was 12, a default value; for RT, SVM and GRNN, local searching option is not applicable.

Sample stratification by geomorphic provinces, that is samples stratified versus non-stratified, was employed by all methods with an exception of TPS, RT, SVM and GRNN. TPS did not use geomorphic province information and was non-stratified, while the other three methods used geomorphic province information and were stratified only.

In the southwest and northeast regions, some datasets contains too few samples for stratified modelling, *e.g.*, OCK, RKlm, RKrf, RKglm, especially in stratified cases. One dataset in the north region resulted in an error in GLS trend estimation and in KED3 and KED4. No predictions were generated for these datasets, thus for these methods in relevant regions, the validation was based on predictions from samples smaller than expected for each sample density.

A singular model in variogram fit was observed for datasets in the southwest region for both non-stratified and stratified samples and for both global and local modelling by KED3 and KED4.

RKgls was applied to all datasets in the north region and to all datasets with 100% sample density in the northeast and southwest regions. A singular model in variogram fit was also observed for datasets in the southwest region for both non-stratified and stratified samples and for both global and local modelling. This was considered in results and discussion below.

A few specific model specifications were used in the following methods. In RKrf , number of variables randomly sampled as candidates at each split was six as determined by tuning random forest. In RKgls, the within-group correlation structure was specified using spherical spatial correlation and nugget was also specified. In RKglm, a quasi-binomial distribution was used and a logit link was specified.

Table B.5. Data transformation, variogram model, searching window and stratification (str vs non.str) for each spatial interpolation method in each region.

Method	Region	Data transformation	Variogram model	Global/local	Stratification (str vs. non-str)
IDW2	all	no	na	both	yes
IDW1	all	no	na	both	yes
IDW3	all	no	na	both	yes
IDW4	all	no	na	both	yes
GLS1	n	square root	Sph	both	yes
	ne	square root	Sph	both	yes
	sw	arcsine	Sph	both	yes
GLS2	n	square root	Sph	both	yes
	ne	square root	Sph	both	yes
	sw	arcsine	Sph	both	yes
KED1	n	square root	Sph	both	yes
	ne	square root	Sph	both	yes
	sw	arcsine	Sph	both	yes
KED2	n	square root	Sph	both	yes
	ne	square root	Sph	both	yes
	sw	arcsine	Sph	both	yes
KED3	n	square root	Sph	both	yes
	ne	square root	Sph	both	yes
	sw	arcsine	Sph	both	yes
KED4	n	square root	Sph	both	yes
	ne	square root	Sph	both	yes
	sw	arcsine	Sph	both	yes
OCK1	n	square root	Sph	both	yes
	ne	square root	Sph	both	yes
	sw	arcsine	Sph	both	yes
OCK2	n	square root	Sph	both	yes
	ne	square root	Sph	both	yes
	sw	arcsine	Sph	both	yes
OCK3	n	square root	Sph	both	yes
	ne	square root	Sph	both	yes
	sw	arcsine	Sph	both	yes
OCK4	n	square root	Sph	both	yes
	ne	square root	Sph	both	yes
	sw	arcsine	Sph	both	yes
OK	n	square root	Sph	both	yes
	ne	square root	Exp	both	yes
	sw	arcsine	Sph	both	yes
UK	n	square root	Sph	both	yes
	ne	square root	Sph	both	yes
	sw	arcsine	Sph	both	yes
RT	all	no	na	na	str
TPSr	all	no	na	local	non.str only
TPSt	all	no	na	local	non.str only
GRNN	all	no	na	na	str only

Method	Region	Data transformation	Variogram model	Global/local	Stratification (str vs. non-str)
SVM	all	no	na	na	str only
RKglm1	all	1/100	Sph	both	yes
RKglm2	all	1/100	Sph	both	yes
RKglm3	all	1/100	Sph	both	yes
RKglm4	all	1/100	Sph	both	yes
RKglm5	all	1/100	Sph	both	yes
RKgls1	n	square root	Sph	both	yes
	ne	square root	Sph	both	yes
	sw	arcsine	Sph	both	yes
RKgls2	n	square root	Sph	both	yes
	ne	square root	Sph	both	yes
	sw	arcsine	Sph	both	yes
RKgls3	n	square root	Sph	both	yes
	ne	square root	Sph	both	yes
	sw	arcsine	Sph	both	yes
RKgls4	n	square root	Sph	both	yes
	ne	square root	Sph	both	yes
	sw	arcsine	Sph	both	yes
RKgls5	n	square root	Sph	both	yes
	ne	square root	Sph	both	yes
	sw	arcsine	Sph	both	yes
RKlm1	n	square root	Sph	both	yes
	ne	square root	Sph	both	yes
	sw	arcsine	Sph	both	yes
RKlm2	n	square root	Sph	both	yes
	ne	square root	Sph	both	yes
	sw	arcsine	Sph	both	yes
RKlm3	n	square root	Sph	both	yes
	ne	square root	Sph	both	yes
	sw	arcsine	Sph	both	yes
RKlm4	n	square root	Sph	both	yes
	ne	square root	Sph	both	yes
	sw	arcsine	Sph	both	yes
RKlm5	n	square root	Sph	both	yes
	ne	square root	Sph	both	yes
	sw	arcsine	Sph	both	yes
RKrf	all	no	Sph	both	yes

B.4.2. Parameters specification

The parameters specified for each of the 14 methods are summarised in Table B.6, which lead to 37 sub-methods compared in this study. A distance power of 1, 2, 3 and 4 were used in IDW. Of which IDS (IDW with power 2), a standard method used in GA, was used as the control. Given that bathymetry (bathy) was the most strongly correlated variable with mud content in all three regions, it was used as a secondary variable in all methods that consider secondary information. Although slope correlated with mud content weakly in the northeast region, it was used in relevant models in all three regions. So was distance-to-coast. The latitude and logitude terms were used in UK, KED4, RKlm5, RKglm5, and RKgls5. Appropriate data transformation of secondary information was used for OCK3 and OCK4 as discussed in section B.1. In RKrf, KED3, KED4, RKlm5, RKglm5 and RKgls5, we used all possible secondary variables, their second and third power, and the latitude and logitude terms in the Legendre and Legendre's (1998) equation. All these modelling work was implemented using *gstat*, *randomForest* and/or *nlme* in R (R Development Core Team, 2007) and the searching neighhood size is 20, with an exception of of TPS, RT, SVM and GRNN as further discussed in the next section B.4.3. Predictions were corrected by resetting the faulty estimate to the nearest bound of the data range (*i.e.*, 0 or 100%) (Goovaerts, 1997).

Table B.6. Parameters used for each spatial interpolation methods.

No	Method	Distance power/trend/secondary variables
1	**IDW2**	2
2	IDW1	1
3	IDW3	3
4	IDW4	4
5	GLS1	bathy
6	GLS2	bathy, dist.coast, slope
7	KED1	bathy
8	KED2	bathy, dist.coast, slope
9	KED3	bathy, dist.coast, slope, bathy^2, bathy^3, dist.coast^2, dist.coast^3, slope^2, slope^3
10	KED4	bathy, dist.coast, slope, bathy^2, bathy^3, dist.coast^2, dist.coast^3, slope^2, slope^3, lat, lat^2, lon, lon^2, lat*lon, lat*lon^2, lon*lat^2, lat^3, lon^3
11	OCK1	bathy
12	OCK2	bathy, dist.coast, slope
13	OCK3	sqrt(sqrt(bathy*(-1)))
14	OCK4	sqrt(sqrt(bathy*(-1))), sqrt(sqrt(dist.coast)), sqrt(sqrt(slope))
15	OK	na
16	UK	lat, lat^2, lon, lon^2, lat*lon, lat*lon^2, lon*lat^2, lat^3, lon^3
17	RT	lat, lon, bathy, slope, dist
18	TPSr	Spline.regularised
19	TPSt	Spline.tension
20	GRNN	lat, lon, bathy, slope, dist
21	SVM	lat, lon, bathy, slope, dist
22	RKglm1	bathy
23	RKglm2	bathy, dist.coast, slope
24	RKglm3	bathy, dist.coast, slope, Prov_code
25	RKglm4	bathy, dist.coast, slope, lat , lon , Prov_code

No	Method	Distance power/trend/secondary variables
26	RKglm5	bathy, dist.coast, slope, bathy^2, bathy^3, dist.coast^2, dist.coast^3, slope^2, slope^3, lat, lat^2, lon, lon^2, lat*lon, lat*lon^2, lon*lat^2, lat^3, lon^3
27	RKgls1	bathy
28	RKgls2	bathy, dist.coast, slope
29	RKgls3	bathy, dist.coast, slope, Prov_code
30	RKgls4	bathy, dist.coast, slope, lat , lon , Prov_code
31	RKgls5	bathy, dist.coast, slope, bathy^2, bathy^3, dist.coast^2, dist.coast^3, slope^2, slope^3, lat, lat^2, lon, lon^2, lat*lon, lat*lon^2, lon*lat^2, lat^3, lon^3
32	RKlm1	bathy
33	RKlm2	bathy, dist.coast, slope
34	RKlm3	bathy, dist.coast, slope, Prov_code
35	RKlm4	bathy, dist.coast, slope, lat , lon , Prov_code
36	RKlm5	bathy, dist.coast, slope, bathy^2, bathy^3, dist.coast^2, dist.coast^3, slope^2, slope^3, lat, lat^2, lon, lon^2, lat*lon, lat*lon^2, lon*lat^2, lat^3, lon^3
37	RKrf	bathy, dist.coast, slope, bathy^2, bathy^3, dist.coast^2, dist.coast^3, slope^2, slope^3, lat, lat^2, lon, lon^2, lat*lon, lat*lon^2, lon*lat^2, lat^3, lon^3

B.4.3. Model and parameter specification of TPS, RT, SVM and GRNN

RT, SVM and GRNN are non-parametric predictive models. For the purpose of this experiment, they use the following secondary variables: latitude, longitude, geomorphic provinces, bathymetry, slope, and distance-to-coast. The TPS technique does not usethe secondary information.

RT was implemented in DTREG software and modelling parameters used for each region and sample density are detailed in Table B.7. As the target variable is continuous, regression tree modelling was used. We varied two parameters for the experiments and kept other constant. The "smooth minimum spikes" parameter was kept as 3. The model used 10-fold cross-validation for validating and pruning the tree. The trees were pruned to minimum cross-validated error. The two varied parameters are the numbers of "Minimum size node to split" and "Maximum tree level", which have effect on the final tree size. The two parameters from all experiments are listed in the following table.

Table B.7. Modelling parameters used for RT in each region.

Region/sample density	Minimum size node to split	Maximum tree levels
$CV_{1\text{-}10}$n.0.2	10	7
$CV_{1\text{-}10}$n.0.4	20	7
$CV_{1\text{-}10}$n.0.6	40	7
$CV_{1\text{-}10}$n.0.8	50	7
$CV_{1\text{-}10}$n.1	100	7
$CV_{1\text{-}10}$ne.0.2	10	8
$CV_{1\text{-}10}$ne.0.4	20	8
$CV_{1\text{-}10}$ne.0.6	50	8
$CV_{1\text{-}10}$ne.0.8	60	8
$CV_{1\text{-}10}$ne.1	80	8
$CV_{1\text{-}10}$sw.0.2	4	4
$CV_{1\text{-}10}$sw.0.4	5	4
$CV_{1\text{-}10}$sw.0.6	6	4
$CV_{1\text{-}10}$sw.0.8	8	4
$CV_{1\text{-}10}$sw.1	10	4

SVM was implemented in DTREG software. All experiments using this technique applied the same set parameters. They are as follows.

- They of SVM model: Epsilon-SVR regression
- Kernel function: RBF
- Stopping criteria: 0.001
- Cache Size (MB): 1000.0
- Use shrinking heuristics: Yes
- Calculate importance of variables: Yes
- Model testing and validation: 10 fold cross-validation
- How to handle missing values: Replace missing values with medians
- Parameter optimization search control:
    - Do grid search for optimal parameters: Yes, Intervals: 5, 2

    - Do pattern search for optimal parameters: Yes, Intervals: 200, tolerance: 0.00000001
    - % rows to use for search: 100
    - Cross validate: 5 folds
    - Optimize: Minimum total error
- Model parameters
    - C: 1 – 100
    - Gamma: 0.001 – 400
    - P: 0.0001 – 100

GRNN was implemented in DTREG software. The "Model optimization and simplification" parameter was applied differently for the experiments. All other parameters were kept constant and they are:

- Sigma values for model: Sigma for each variable
- Starting sigma search control:
    - Minimum Sigma: 0.0001
    - Maximum Sigma: 50
    - Search steps: 20
- Model testing and validation: leave-one-out validation
- How to handle missing values: Replace missing predictors with medians
- Type of kernel function: Gaussian
- Compute importance of variables: Yes
- Conjugate gradient parameters
    - Maximum total iterations: 200
    - Iterations without improvement: 10
    - Minimum improvement delta: 0.00001
    - Absolute convergence tolerance: 0.00000001
    - Relative convergence tolerance: 0.0001

For the "Model optimization and simplification" parameter, it was decided to remove unnecessary neurons and to retrain after removing neurons. The removal of unnecessary neurons depends on either the "Minimize error" option, which lets the model decide how many neurons to remove, or the "number of neurons" option which you decide how many neurons to remove. The parameter setting varied with experiments and list as follows:

- CV_{1-10}.n.0.2: using the "Minimize error" option
- CV_{1-10}.n.0.4: remove 30 neurons
- CV_{1-10}.n.0.6: remove 10 neurons
- CV_{1-10}.n.0.8: remove 5 neurons
- CV_{1-10}.n.1: remove 4 neurons
- CV_{1-10}.ne.0.2: using the "Minimize error" option
- CV_{1-10}.ne.0.4: remove 30 neurons
- CV_{1-10}.ne.0.6: remove 10 neurons
- CV_{1-10}.ne.0.8: remove 5 neurons
- CV_{1-10}.ne.1: remove 5 neurons
- CV_{1-10}.sw.0.2: using the "Minimize error" option
- CV_{1-10}.sw.0.4: using the "Minimize error" option
- CV_{1-10}.sw.0.6: using the "Minimize error" option
- CV_{1-10}.sw.0.8: using the "Minimize error" option
- CV_{1-10}.sw.1: using the "Minimize error" option

TPS was implemented in ArcGIS desktop with two sets of parameters and a seaching neighbourhood size of 12. The first set used the spline type of “REGULARIZED”, a weight of 0.1, and the second set used the spline type of “TENSION”, a weight of 5.

Appendix C. Basic statistical summaries of predictions and statistics measuring the performance of each statistical method.

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
1	IDW2	n	20	non.str	global	1.34	31.88	89.5	12.79	17.23	39.72	53.51
1	IDW2	n	20	str	global	1.34	31.83	89.51	12.85	17.35	39.91	53.88
1	IDW2	n	20	non.str	local	1.34	32.53	92.05	12.29	16.69	38.17	51.83
1	IDW2	n	20	str	local	1.34	32.49	92.05	12.32	16.78	38.26	52.11
1	IDW2	n	40	non.str	global	0.02	33.83	94.12	14.13	19.27	41.40	56.46
1	IDW2	n	40	str	global	0.02	33.88	94.12	14.17	19.3	41.52	56.55
1	IDW2	n	40	non.str	local	0.02	34.51	95.53	13.57	19.31	39.76	56.58
1	IDW2	n	40	str	local	0.02	34.55	95.53	13.61	19.33	39.88	56.64
1	IDW2	n	60	non.str	global	0.63	33.79	91.55	13.72	19.09	40.78	56.75
1	IDW2	n	60	str	global	0.63	33.8	91.55	13.74	19.13	40.84	56.87
1	IDW2	n	60	non.str	local	0.29	34.22	93.32	13.21	19.14	39.27	56.90
1	IDW2	n	60	str	local	0.29	34.21	93.32	13.24	19.17	39.36	56.99
1	IDW2	n	80	non.str	global	1.03	34	90.3	13.19	18.24	38.75	53.58
1	IDW2	n	80	str	global	1.03	34.01	90.3	13.17	18.22	38.69	53.53
1	IDW2	n	80	non.str	local	0.71	34.4	93.23	12.55	18.26	36.87	53.64
1	IDW2	n	80	str	local	0.71	34.4	93.23	12.53	18.22	36.81	53.53
1	IDW2	n	100	non.str	global	0.08	33.84	89.9	13.27	18.45	38.98	54.20
1	IDW2	n	100	str	global	0.08	33.85	89.9	13.25	18.44	38.92	54.17
1	IDW2	n	100	non.str	local	0.07	34.19	93.5	12.61	18.36	37.04	53.94
1	IDW2	n	100	str	local	0.07	34.18	93.5	12.61	18.35	37.04	53.91
1	IDW2	ne	20	non.str	global	2.86	25.5	77.96	12.35	16.94	48.76	66.88
1	IDW2	ne	20	str	global	2.86	25.94	92.06	11.98	16.01	47.67	63.71
1	IDW2	ne	20	non.str	local	2.39	25.31	78.44	12.18	17.01	48.09	67.15
1	IDW2	ne	20	str	local	2.39	25.8	92.06	11.82	16.09	47.04	64.03
1	IDW2	ne	40	non.str	global	3.57	25.44	94.47	11.53	15.69	45.34	61.70
1	IDW2	ne	40	str	global	3.57	25.8	94.9	11.2	15.12	44.20	59.67
1	IDW2	ne	40	non.str	local	1.37	25.43	94.83	10.91	15.21	42.90	59.81
1	IDW2	ne	40	str	local	1.37	25.7	94.9	10.75	15	42.42	59.19
1	IDW2	ne	60	non.str	global	1.23	25.32	94.28	10.44	14.66	41.13	57.76
1	IDW2	ne	60	str	global	1.23	25.62	94.89	10.26	14.41	40.54	56.93
1	IDW2	ne	60	non.str	local	0.72	25.42	94.83	9.95	14.45	39.20	56.93
1	IDW2	ne	60	str	local	0.72	25.61	94.89	9.83	14.24	38.84	56.26
1	IDW2	ne	80	non.str	global	0.67	24.76	94.06	10.12	14.38	40.17	57.09
1	IDW2	ne	80	str	global	0.67	25.2	99.49	9.87	13.89	39.18	55.14
1	IDW2	ne	80	non.str	local	0.36	24.95	94.84	9.49	13.92	37.67	55.26
1	IDW2	ne	80	str	local	0.36	25.24	99.49	9.33	13.63	37.04	54.11
1	IDW2	ne	100	non.str	global	0.74	24.62	93.84	9.77	13.92	38.94	55.48
1	IDW2	ne	100	str	global	0.74	25.08	99.49	9.51	13.34	37.90	53.17
1	IDW2	ne	100	non.str	local	0.3	24.83	94.83	9.07	13.36	36.15	53.25
1	IDW2	ne	100	str	local	0.3	25.11	99.49	8.92	13.01	35.55	51.85
1	IDW2	sw	20	non.str	global	0.43	46.46	87.61	18.73	23.87	38.82	49.47
1	IDW2	sw	20	str	global	0.19	52.53	97.53	16.63	22.35	34.47	46.32
1	IDW2	sw	20	non.str	local	0.4	46.72	87.62	18.39	23.74	38.11	49.20
1	IDW2	sw	20	str	local	0.19	52.53	97.53	16.63	22.35	34.47	46.32
1	IDW2	sw	40	non.str	global	1.79	45.61	94.66	13.19	17.13	27.43	35.62
1	IDW2	sw	40	str	global	0.7	47.53	96.23	12.55	17.94	26.22	37.48
1	IDW2	sw	40	non.str	local	0.62	46.03	96.24	11.51	15.83	23.93	32.92
1	IDW2	sw	40	str	local	0.62	47.63	96.41	12.34	17.69	25.78	36.96
1	IDW2	sw	60	non.str	global	1.09	47.36	93.93	13.54	18.59	28.37	38.95
1	IDW2	sw	60	str	global	0.39	49.76	97.48	12.75	18.82	26.71	39.43
1	IDW2	sw	60	non.str	local	0.08	48.22	95.16	11.54	17.3	24.18	36.25
1	IDW2	sw	60	str	local	0.08	49.85	97.48	12.3	18.5	25.77	38.76
1	IDW2	sw	80	non.str	global	1.5	44.75	88.39	13.74	18.11	30.09	39.66
1	IDW2	sw	80	str	global	0.88	47.32	94.4	12.4	17.44	27.16	38.20
1	IDW2	sw	80	non.str	local	0.14	45.19	92.33	11.51	16.7	25.21	36.57
1	IDW2	sw	80	str	local	0.14	47.59	94.4	11.91	17.35	26.08	38.00
1	IDW2	sw	100	non.str	global	1.47	45.75	91	12.15	16.89	26.28	36.53
1	IDW2	sw	100	str	global	0.67	48.13	94.78	10.91	16.28	23.60	35.22
1	IDW2	sw	100	non.str	local	0.1	46.41	91.83	10.14	15.76	21.93	34.09
1	IDW2	sw	100	str	local	0.09	48.11	94.78	10.3	16.22	22.28	35.09
2	IDW1	n	20	non.str	global	1.8	31.7	71.07	17.9	22.36	55.59	69.44
2	IDW1	n	20	str	global	1.79	31.69	71.13	17.97	22.46	55.81	69.75
2	IDW1	n	20	non.str	local	1.42	32.39	82.38	14.15	18.72	43.94	58.14
2	IDW1	n	20	str	local	1.42	32.37	82.38	14.23	18.84	44.19	58.51
2	IDW1	n	40	non.str	global	4.14	34.07	65.94	18.11	22.37	53.06	65.54
2	IDW1	n	40	str	global	4.1	33.92	66.03	18.05	22.38	52.89	65.57
2	IDW1	n	40	non.str	local	0.81	34.3	85.89	14.08	19.31	41.25	56.58
2	IDW1	n	40	str	local	0.81	34.33	85.89	14.11	19.36	41.34	56.72
2	IDW1	n	60	non.str	global	12.55	33.69	66.12	18.17	22.34	54.01	66.41
2	IDW1	n	60	str	global	12.47	33.7	66.2	18.18	22.4	54.04	66.59
2	IDW1	n	60	non.str	local	3.07	34.1	88.31	13.44	18.75	39.95	55.74
2	IDW1	n	60	str	local	3.07	34.1	88.31	13.46	18.8	40.01	55.89

Appendix

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
2	IDW1	n	80	non.str	global	11.41	34.06	63.42	18.05	22.16	53.03	65.10
2	IDW1	n	80	str	global	11.34	34.02	63.5	18.04	22.18	53.00	65.16
2	IDW1	n	80	non.str	local	2.33	34.48	91.19	12.95	18.16	38.04	53.35
2	IDW1	n	80	str	local	2.33	34.44	91.19	12.91	18.11	37.93	53.20
2	IDW1	n	100	non.str	global	3.49	34.04	65.66	18	22.16	52.88	65.10
2	IDW1	n	100	str	global	3.47	34	65.76	17.96	22.13	52.76	65.01
2	IDW1	n	100	non.str	local	1.51	34.24	91.74	12.87	18.13	37.81	53.26
2	IDW1	n	100	str	local	1.51	34.21	91.74	12.85	18.11	37.75	53.20
2	IDW1	ne	20	non.str	global	15.23	25.42	49.6	14.56	18.25	57.48	72.05
2	IDW1	ne	20	str	global	15.04	25.59	77.22	14.14	17.36	56.27	69.08
2	IDW1	ne	20	non.str	local	5.71	25.42	64.74	12.67	16.81	50.02	66.36
2	IDW1	ne	20	str	local	5.71	25.86	77.22	12.31	15.89	48.99	63.23
2	IDW1	ne	40	non.str	global	12.06	25.11	51.37	15.08	18.75	59.30	73.73
2	IDW1	ne	40	str	global	11.5	25.43	94.23	14.45	17.79	57.02	70.21
2	IDW1	ne	40	non.str	local	2.98	25.09	87.03	11.78	15.96	46.32	62.76
2	IDW1	ne	40	str	local	2.98	25.58	94.23	11.55	15.46	45.58	61.01
2	IDW1	ne	60	non.str	global	13.68	25.06	49.43	14.43	18.04	56.86	71.08
2	IDW1	ne	60	str	global	13.55	25.34	93.1	14.01	17.42	55.35	68.83
2	IDW1	ne	60	non.str	local	2.38	25.21	87.63	10.85	14.93	42.75	58.83
2	IDW1	ne	60	str	local	2.38	25.54	93.18	10.71	14.65	42.32	57.88
2	IDW1	ne	80	non.str	global	13.28	24.68	46.11	14.4	18.14	57.17	72.01
2	IDW1	ne	80	str	global	13.12	25.12	99.49	13.84	17.26	54.94	68.52
2	IDW1	ne	80	non.str	local	1.54	24.7	88.52	10.23	14.37	40.61	57.05
2	IDW1	ne	80	str	local	1.54	25.09	99.49	10.08	14.06	40.02	55.82
2	IDW1	ne	100	non.str	global	14.16	24.6	47	14.38	18.13	57.31	72.26
2	IDW1	ne	100	str	global	13.61	25.02	99.49	13.84	17.25	55.16	68.75
2	IDW1	ne	100	non.str	local	1.76	24.6	87.65	9.86	13.99	39.30	55.76
2	IDW1	ne	100	str	local	1.76	24.98	99.49	9.72	13.65	38.74	54.40
2	IDW1	sw	20	non.str	global	10.1	44.5	76.69	24.78	27.87	51.36	57.76
2	IDW1	sw	20	str	global	2.89	51.43	96.35	19.16	23.32	39.71	48.33
2	IDW1	sw	20	non.str	local	8.24	44.82	77.08	23.18	26.66	48.04	55.25
2	IDW1	sw	20	str	local	2.89	51.43	96.35	19.16	23.32	39.71	48.33
2	IDW1	sw	40	non.str	global	17.17	45.03	79.84	23.29	25.68	48.43	53.40
2	IDW1	sw	40	str	global	6.45	48.61	93.75	17.68	21.34	36.94	44.59
2	IDW1	sw	40	non.str	local	2.9	44.63	82.42	16.65	20.83	34.62	43.31
2	IDW1	sw	40	str	local	5.34	48.93	93.75	16.95	20.7	35.42	43.25
2	IDW1	sw	60	non.str	global	15.5	44.93	69.74	22.98	25.44	48.15	53.30
2	IDW1	sw	60	str	global	3.96	49.75	96.35	16.91	21.71	35.43	45.49
2	IDW1	sw	60	non.str	local	0.17	46.16	87.23	13.91	18.78	29.14	39.35
2	IDW1	sw	60	str	local	0.14	50.04	96.35	14.91	20.38	31.24	42.70
2	IDW1	sw	80	non.str	global	12.92	44.1	74.95	23.08	25.51	50.55	55.87
2	IDW1	sw	80	str	global	4.84	47.98	94.4	17.83	21.56	39.05	47.22
2	IDW1	sw	80	non.str	local	0.21	44.65	86.62	14.09	19.35	30.86	42.38
2	IDW1	sw	80	str	local	0.19	48.6	94.4	15.09	20.35	33.05	44.57
2	IDW1	sw	100	non.str	global	12.6	44.48	82.77	21.79	24.24	47.13	52.43
2	IDW1	sw	100	str	global	3.96	48.59	93.71	16.2	20.21	35.04	43.72
2	IDW1	sw	100	non.str	local	0.12	45.93	86.56	12.31	17.89	26.63	38.70
2	IDW1	sw	100	str	local	0.12	49.29	93.71	13.29	18.74	28.75	40.54
3	IDW3	n	20	non.str	global	0.24	32.32	96.5	12.23	16.66	37.98	51.74
3	IDW3	n	20	str	global	0.24	32.29	96.51	12.25	16.71	38.04	51.89
3	IDW3	n	20	non.str	local	0.23	32.55	96.8	12.27	16.68	38.11	51.80
3	IDW3	n	20	str	local	0.23	32.52	96.8	12.29	16.71	38.17	51.89
3	IDW3	n	40	non.str	global	0.01	34.52	97.1	13.89	19.95	40.70	58.45
3	IDW3	n	40	str	global	0.01	34.55	97.1	13.92	19.97	40.79	58.51
3	IDW3	n	40	non.str	local	0.01	34.71	97.3	13.81	20.1	40.46	58.89
3	IDW3	n	40	str	local	0.01	34.73	97.3	13.83	20.12	40.52	58.95
3	IDW3	n	60	non.str	global	0.07	34.22	96.37	13.61	19.85	40.46	59.01
3	IDW3	n	60	str	global	0.07	34.21	96.37	13.63	19.87	40.52	59.07
3	IDW3	n	60	non.str	local	0.07	34.28	96.41	13.55	19.99	40.28	59.42
3	IDW3	n	60	str	local	0.07	34.27	96.41	13.57	20.01	40.34	59.48
3	IDW3	n	80	non.str	global	0.33	34.26	96.4	12.76	18.75	37.49	55.08
3	IDW3	n	80	str	global	0.33	34.27	96.4	12.75	18.73	37.46	55.02
3	IDW3	n	80	non.str	local	0.3	34.32	96.46	12.74	18.96	37.43	55.70
3	IDW3	n	80	str	local	0.3	34.33	96.46	12.72	18.92	37.37	55.58
3	IDW3	n	100	non.str	global	0.01	34.07	96.11	12.93	18.95	37.98	55.67
3	IDW3	n	100	str	global	0.01	34.08	96.11	12.92	18.93	37.96	55.61
3	IDW3	n	100	non.str	local	0.01	34.11	96.24	12.85	19.07	37.75	56.02
3	IDW3	n	100	str	local	0.01	34.11	96.24	12.85	19.05	37.75	55.96
3	IDW3	ne	20	non.str	global	0.24	25.17	81.3	12.26	17.49	48.40	69.05
3	IDW3	ne	20	str	global	0.24	25.64	94.52	11.92	16.64	47.43	66.22
3	IDW3	ne	20	non.str	local	0.23	25.1	81.31	12.28	17.55	48.48	69.29
3	IDW3	ne	20	str	local	0.23	25.59	94.52	11.92	16.69	47.43	66.41
3	IDW3	ne	40	non.str	global	0.4	25.42	94.89	11.05	15.75	43.45	61.93
3	IDW3	ne	40	str	global	0.4	25.73	94.9	10.83	15.48	42.74	61.09
3	IDW3	ne	40	non.str	local	0.21	25.51	94.9	10.89	15.51	42.82	60.99
3	IDW3	ne	40	str	local	0.21	25.71	94.9	10.79	15.51	42.58	61.21
3	IDW3	ne	60	non.str	global	0.14	25.43	94.89	10	14.78	39.40	58.23
3	IDW3	ne	60	str	global	0.14	25.65	94.9	9.9	14.58	39.11	57.61
3	IDW3	ne	60	non.str	local	0.13	25.48	94.9	10	14.84	39.40	58.47
3	IDW3	ne	60	str	local	0.13	25.65	94.9	9.9	14.64	39.11	57.84

Appendix

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
3	IDW3	ne	80	non.str	global	0.05	24.91	94.89	9.58	14.33	38.03	56.89
3	IDW3	ne	80	str	global	0.05	25.29	99.49	9.37	13.89	37.20	55.14
3	IDW3	ne	80	non.str	local	0.05	25.06	94.9	9.46	14.14	37.55	56.13
3	IDW3	ne	80	str	local	0.05	25.33	99.49	9.31	13.89	36.96	55.14
3	IDW3	ne	100	non.str	global	0.06	24.78	94.89	9.19	13.72	36.63	54.68
3	IDW3	ne	100	str	global	0.06	25.15	99.49	8.97	13.22	35.75	52.69
3	IDW3	ne	100	non.str	local	0.05	24.93	94.9	9.07	13.54	36.15	53.97
3	IDW3	ne	100	str	local	0.05	25.17	99.49	8.93	13.25	35.59	52.81
3	IDW3	sw	20	non.str	global	0.07	47.22	93.26	16.98	23.51	35.19	48.73
3	IDW3	sw	20	str	global	0.06	52.58	97.79	16.11	22.9	33.39	47.46
3	IDW3	sw	20	non.str	local	0.07	47.33	93.45	16.88	23.58	34.98	48.87
3	IDW3	sw	20	str	local	0.06	52.58	97.79	16.11	22.9	33.39	47.46
3	IDW3	sw	40	non.str	global	0.21	46.68	98	10.14	14.67	21.09	30.51
3	IDW3	sw	40	str	global	0.15	47.88	98.08	11.44	17.49	23.90	36.54
3	IDW3	sw	40	non.str	local	0.1	46.9	98.08	10.07	14.47	20.94	30.09
3	IDW3	sw	40	str	local	0.14	47.94	98.09	11.34	17.3	23.69	36.15
3	IDW3	sw	60	non.str	global	0.09	48.49	96.29	11.12	17.61	23.30	36.90
3	IDW3	sw	60	str	global	0.07	49.39	97.78	11.24	18.2	23.55	38.13
3	IDW3	sw	60	non.str	local	0.06	48.69	96.32	10.69	17.39	22.40	36.43
3	IDW3	sw	60	str	local	0.06	49.46	97.78	11.12	18.03	23.30	37.77
3	IDW3	sw	80	non.str	global	0.3	45.05	95.33	11.12	16.67	24.35	36.51
3	IDW3	sw	80	str	global	0.15	46.62	96.93	10.86	16.79	23.78	36.77
3	IDW3	sw	80	non.str	local	0.1	45.39	95.55	10.5	16.3	23.00	35.70
3	IDW3	sw	80	str	local	0.1	46.83	96.93	10.77	16.81	23.59	36.82
3	IDW3	sw	100	non.str	global	0.17	46.25	96.01	9.84	15.79	21.28	34.16
3	IDW3	sw	100	str	global	0.1	47.46	97.03	9.53	15.78	20.61	34.13
3	IDW3	sw	100	non.str	local	0.07	46.54	96.3	9.36	15.61	20.25	33.77
3	IDW3	sw	100	str	local	0.07	47.41	97.03	9.43	15.82	20.40	34.22
4	IDW4	n	20	non.str	global	0.13	32.44	97.31	12.66	17.2	39.32	53.42
4	IDW4	n	20	str	global	0.13	32.44	97.31	12.67	17.21	39.35	53.45
4	IDW4	n	20	non.str	local	0.13	32.55	97.32	12.71	17.23	39.47	53.51
4	IDW4	n	20	str	local	0.13	32.54	97.32	12.71	17.24	39.47	53.54
4	IDW4	n	40	non.str	global	0.01	34.77	97.94	14.18	20.76	41.55	60.83
4	IDW4	n	40	str	global	0.01	34.78	97.94	14.19	20.77	41.58	60.86
4	IDW4	n	40	non.str	local	0.01	34.81	97.95	14.16	20.81	41.49	60.97
4	IDW4	n	40	str	local	0.01	34.81	97.95	14.16	20.82	41.49	61.00
4	IDW4	n	60	non.str	global	0.01	34.31	97.42	13.89	20.63	41.29	61.33
4	IDW4	n	60	str	global	0.01	34.29	97.42	13.91	20.64	41.35	61.36
4	IDW4	n	60	non.str	local	0.01	34.33	97.42	13.89	20.68	41.29	61.47
4	IDW4	n	60	str	local	0.01	34.32	97.42	13.9	20.69	41.32	61.50
4	IDW4	n	80	non.str	global	0.24	34.26	97.65	13.02	19.51	38.25	57.31
4	IDW4	n	80	str	global	0.24	34.27	97.65	13.01	19.48	38.22	57.23
4	IDW4	n	80	non.str	local	0.24	34.28	97.67	13.03	19.59	38.28	57.55
4	IDW4	n	80	str	local	0.24	34.29	97.67	13.01	19.55	38.22	57.43
4	IDW4	n	100	non.str	global	0.01	34.04	96.83	13.11	19.59	38.51	57.55
4	IDW4	n	100	str	global	0.01	34.05	96.83	13.1	19.57	38.48	57.49
4	IDW4	n	100	non.str	local	0.01	34.04	96.84	13.08	19.66	38.43	57.76
4	IDW4	n	100	str	local	0.01	34.05	96.84	13.07	19.63	38.40	57.67
4	IDW4	ne	20	non.str	global	0.02	24.97	81.83	12.38	17.88	48.87	70.59
4	IDW4	ne	20	str	global	0.02	25.44	94.85	12.03	17.07	47.87	67.93
4	IDW4	ne	20	non.str	local	0.02	24.94	81.83	12.39	17.9	48.91	70.67
4	IDW4	ne	20	str	local	0.02	25.42	94.85	12.03	17.09	47.87	68.01
4	IDW4	ne	40	non.str	global	0.04	25.42	94.9	11.14	16.06	43.81	63.15
4	IDW4	ne	40	str	global	0.04	25.67	94.9	10.99	15.94	43.37	62.90
4	IDW4	ne	40	non.str	local	0.03	25.52	94.9	11.03	15.86	43.37	62.37
4	IDW4	ne	40	str	local	0.03	25.67	94.9	10.98	15.96	43.33	62.98
4	IDW4	ne	60	non.str	global	0.03	25.45	94.9	10.2	15.25	40.19	60.09
4	IDW4	ne	60	str	global	0.03	25.64	94.9	10.1	15.01	39.91	59.30
4	IDW4	ne	60	non.str	local	0.03	25.47	94.9	10.2	15.24	40.19	60.05
4	IDW4	ne	60	str	local	0.03	25.64	94.9	10.11	15.05	39.94	59.46
4	IDW4	ne	80	non.str	global	0.01	25.01	94.9	9.66	14.59	38.35	57.92
4	IDW4	ne	80	str	global	0.01	25.35	99.49	9.45	14.18	37.51	56.29
4	IDW4	ne	80	non.str	local	0.01	25.11	94.9	9.58	14.4	38.03	57.17
4	IDW4	ne	80	str	local	0.01	25.37	99.49	9.44	14.19	37.48	56.33
4	IDW4	ne	100	non.str	global	0.01	24.86	94.9	9.27	13.98	36.95	55.72
4	IDW4	ne	100	str	global	0.01	25.18	99.49	9.07	13.55	36.15	54.01
4	IDW4	ne	100	non.str	local	0.01	24.97	94.9	9.19	13.79	36.63	54.96
4	IDW4	ne	100	str	local	0.01	25.19	99.49	9.07	13.57	36.15	54.09
4	IDW4	sw	20	non.str	global	0.06	47.41	94.27	16.74	23.9	34.69	49.53
4	IDW4	sw	20	str	global	0.06	52.5	97.84	16.26	23.4	33.70	48.50
4	IDW4	sw	20	non.str	local	0.06	47.45	94.29	16.69	23.99	34.59	49.72
4	IDW4	sw	20	str	local	0.06	52.5	97.84	16.26	23.4	33.70	48.50
4	IDW4	sw	40	non.str	global	0.04	47.42	98.23	9.54	14.06	19.84	29.24
4	IDW4	sw	40	str	global	0.03	48.25	98.23	11.09	17.49	23.17	36.54
4	IDW4	sw	40	non.str	local	0.03	47.49	98.23	9.62	14.15	20.00	29.42
4	IDW4	sw	40	str	local	0.03	48.31	98.23	11	17.34	22.98	36.23
4	IDW4	sw	60	non.str	global	0.05	48.85	97.02	10.65	17.65	22.31	36.98
4	IDW4	sw	60	str	global	0.04	49.24	97.84	10.93	18.19	22.90	38.11
4	IDW4	sw	60	non.str	local	0.05	48.88	97.02	10.57	17.62	22.15	36.92
4	IDW4	sw	60	str	local	0.04	49.3	97.84	10.86	18.04	22.75	37.80

Appendix

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
4	IDW4	sw	80	non.str	global	0.09	45.35	97.01	10.52	16.74	23.04	36.66
4	IDW4	sw	80	str	global	0.09	46.33	97.63	10.57	17.07	23.15	37.39
4	IDW4	sw	80	non.str	local	0.08	45.54	97.02	10.44	16.69	22.86	36.55
4	IDW4	sw	80	str	local	0.08	46.51	97.63	10.6	17.1	23.22	37.45
4	IDW4	sw	100	non.str	global	0.05	46.49	97.65	9.4	15.96	20.33	34.52
4	IDW4	sw	100	str	global	0.05	47.11	97.66	9.31	16	20.14	34.61
4	IDW4	sw	100	non.str	local	0.04	46.59	97.66	9.33	15.99	20.18	34.59
4	IDW4	sw	100	str	local	0.05	47.09	97.66	9.29	16.04	20.10	34.70
5	GLS1	n	20	non.str	global	16.28	23.95	88.98	21.64	29.12	67.20	90.43
5	GLS1	n	20	str	global	12.68	24.42	98.47	21.43	28.67	66.55	89.04
5	GLS1	n	20	non.str	local	0.52	29.95	88.58	18.14	23.93	56.34	74.32
5	GLS1	n	20	str	local	0.52	29.94	98.47	18.22	24.36	56.58	75.65
5	GLS1	n	40	non.str	global	17.34	24.64	79.41	22.84	29.74	66.92	87.14
5	GLS1	n	40	str	global	1.44	25.04	61.15	22.06	28.93	64.64	84.76
5	GLS1	n	40	non.str	local	0.7	31.75	86.61	16.45	22.48	48.20	65.87
5	GLS1	n	40	str	local	0.7	31.76	86.61	16.33	22.26	47.85	65.22
5	GLS1	n	60	non.str	global	14.02	23.31	100	22.2	29.68	65.99	88.23
5	GLS1	n	60	str	global	7.38	23.75	84.92	21.88	29.23	65.04	86.89
5	GLS1	n	60	non.str	local	0	31.35	93.4	15.03	21.1	44.68	62.72
5	GLS1	n	60	str	local	0	31.52	95.85	14.92	20.79	44.35	61.80
5	GLS1	n	80	non.str	global	14.36	23.96	100	22.29	29.66	65.48	87.13
5	GLS1	n	80	str	global	8.24	24.87	100	21.99	29	64.60	85.19
5	GLS1	n	80	non.str	local	0	31.47	100	14.3	20.44	42.01	60.05
5	GLS1	n	80	str	local	0	31.5	100	14.39	20.35	42.27	59.78
5	GLS1	n	100	non.str	global	15.17	24.37	100	22.38	29.55	65.75	86.81
5	GLS1	n	100	str	global	6.58	24.63	100	21.96	29.1	64.51	85.49
5	GLS1	n	100	non.str	local	0	31.67	100	14.26	20.16	41.47	58.62
5	GLS1	n	100	str	local	0	31.46	100	14.12	19.96	41.06	58.04
5	GLS1	ne	20	non.str	global	17.74	19.31	100	15.98	20.72	63.09	81.80
5	GLS1	ne	20	str	global	12.66	18.96	100	16.14	20.9	64.23	83.17
5	GLS1	ne	20	non.str	local	0.01	23.01	100	13.71	18.22	54.13	71.93
5	GLS1	ne	20	str	local	0.01	22.81	100	13.83	18.25	55.03	72.62
5	GLS1	ne	40	non.str	global	18.73	20.57	100	16.78	21.38	65.99	84.07
5	GLS1	ne	40	str	global	15.21	19.4	96.58	16.83	21.66	66.42	85.48
5	GLS1	ne	40	non.str	local	0.38	22.65	100	13.15	18.08	51.71	71.10
5	GLS1	ne	40	str	local	0.38	22.59	96.58	13.34	18.08	52.64	71.35
5	GLS1	ne	60	non.str	global	18.9	20.35	100	16.51	21.23	65.05	83.65
5	GLS1	ne	60	str	global	16.19	19.5	97.59	16.62	21.51	65.67	84.99
5	GLS1	ne	60	non.str	local	0.4	23.33	100	12.79	17.66	50.39	69.58
5	GLS1	ne	60	str	local	0.4	23.09	97.66	12.84	17.69	50.73	69.89
5	GLS1	ne	80	non.str	global	19.42	20.77	100	16.41	20.93	65.14	83.09
5	GLS1	ne	80	str	global	16.57	19.91	100	16.41	21.14	65.14	83.92
5	GLS1	ne	80	non.str	local	0.59	23.21	100	11.87	16.79	47.12	66.65
5	GLS1	ne	80	str	local	0.59	23.14	100	11.91	16.81	47.28	66.73
5	GLS1	ne	100	non.str	global	18.84	20.22	100	16.41	21.1	65.40	84.10
5	GLS1	ne	100	str	global	4.88	19.27	100	16.55	21.5	65.96	85.69
5	GLS1	ne	100	non.str	local	0.22	22.94	100	11.74	16.61	46.79	66.20
5	GLS1	ne	100	str	local	0.22	22.74	100	11.85	16.86	47.23	67.20
5	GLS1	sw	20	non.str	global	19.38	44.07	99.98	22.28	26.81	46.18	55.56
5	GLS1	sw	20	str	global	0.18	42.1	98.76	20.89	28.57	43.30	59.21
5	GLS1	sw	20	non.str	local	10.23	44.5	100	18.79	24.65	38.94	51.09
5	GLS1	sw	20	str	local	0.18	42.1	98.76	20.89	28.57	43.30	59.21
5	GLS1	sw	40	non.str	global	19.77	43.41	98.44	27.61	30.53	57.41	63.49
5	GLS1	sw	40	str	global	0.32	41.98	95.89	22.16	28.22	46.30	58.96
5	GLS1	sw	40	non.str	local	0.34	38.62	94.43	20.26	28.72	42.13	59.72
5	GLS1	sw	40	str	local	0.08	43.86	95.89	17.39	23.28	36.34	48.64
5	GLS1	sw	60	non.str	global	24.53	47.23	98.74	26.97	29.72	56.51	62.27
5	GLS1	sw	60	str	global	0	43.39	95.9	22.03	26.91	46.16	56.38
5	GLS1	sw	60	non.str	local	0.04	39.43	94.82	17.38	25.56	36.41	53.55
5	GLS1	sw	60	str	local	0.01	43.22	99.48	16.22	22.43	33.98	46.99
5	GLS1	sw	80	non.str	global	22.29	42.24	93.15	26.82	30.1	58.74	65.92
5	GLS1	sw	80	str	global	0	40.61	99.89	21.7	26.5	47.53	58.04
5	GLS1	sw	80	non.str	local	0.03	42.24	97.62	15.44	21.22	33.82	46.47
5	GLS1	sw	80	str	local	0	42.65	99.89	14.83	20.94	32.48	45.86
5	GLS1	sw	100	non.str	global	25.07	45.24	93.5	26.18	28.94	56.63	62.60
5	GLS1	sw	100	str	global	0.07	41.11	99.78	20.71	25.11	44.80	54.32
5	GLS1	sw	100	non.str	local	0.01	42.99	97.59	13.59	19.53	29.40	42.25
5	GLS1	sw	100	str	local	0	43.44	99.78	13.46	19.59	29.12	42.38
6	GLS2	n	20	non.str	global	11.44	24.21	85.79	21.65	28.9	67.24	89.75
6	GLS2	n	20	str	global	8.42	24.73	92.59	20.82	27.92	64.66	86.71
6	GLS2	n	20	non.str	local	0.12	29.63	100	18.46	25.32	57.33	78.63
6	GLS2	n	20	str	local	0.12	29.21	100	17.82	24.87	55.34	77.24
6	GLS2	n	40	non.str	global	2.49	27.72	64.47	22.23	28.18	65.13	82.57
6	GLS2	n	40	str	global	1.25	28.13	65.18	21.53	27.53	63.08	80.66
6	GLS2	n	40	non.str	local	0.01	32.01	100	16.04	22.02	47.00	64.52
6	GLS2	n	40	str	local	0.01	32.02	100	16.19	22.13	47.44	64.84
6	GLS2	n	60	non.str	global	11.51	25.63	100	21.63	28.3	64.30	84.13
6	GLS2	n	60	str	global	2.55	25.98	75.73	20.98	27.56	62.37	81.93
6	GLS2	n	60	non.str	local	0.02	31.56	100	15.38	21.71	45.72	64.54
6	GLS2	n	60	str	local	0.02	32	100	15.63	21.85	46.46	64.95

Appendix

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
6	GLS2	n	80	non.str	global	13.13	25.55	100	21.99	28.74	64.60	84.43
6	GLS2	n	80	str	global	8.02	26.11	100	21.61	28.16	63.48	82.73
6	GLS2	n	80	non.str	local	0.03	31.72	100	14.25	20.44	41.86	60.05
6	GLS2	n	80	str	local	0.03	32.27	100	14.48	20.65	42.54	60.66
6	GLS2	n	100	non.str	global	11.65	25.49	100	22.13	28.82	65.01	84.67
6	GLS2	n	100	str	global	8.6	25.98	78.34	21.43	28.1	62.96	82.55
6	GLS2	n	100	non.str	local	0	31.6	100	13.84	19.73	40.24	57.37
6	GLS2	n	100	str	local	0.01	31.56	100	13.83	19.85	40.22	57.72
6	GLS2	ne	20	non.str	global	9.96	19.64	100	15.8	20.57	62.38	81.21
6	GLS2	ne	20	str	global	9.56	19.26	100	15.82	20.6	62.95	81.97
6	GLS2	ne	20	non.str	local	0.95	22.48	100	13.62	18.84	53.77	74.38
6	GLS2	ne	20	str	local	0.95	22.43	100	13.33	18.14	53.04	72.18
6	GLS2	ne	40	non.str	global	11.98	20.98	100	16.78	21.42	65.99	84.23
6	GLS2	ne	40	str	global	4.82	20.33	100	16.89	21.44	66.65	84.61
6	GLS2	ne	40	non.str	local	0.01	23.21	100	12.82	18.17	50.41	71.45
6	GLS2	ne	40	str	local	0.01	23.28	100	13.06	18.3	51.54	72.22
6	GLS2	ne	60	non.str	global	13.04	20.93	100	16.42	21.08	64.70	83.06
6	GLS2	ne	60	str	global	6.32	20.43	100	16.28	21.08	64.32	83.29
6	GLS2	ne	60	non.str	local	0.01	23.77	100	12.61	17.9	49.68	70.53
6	GLS2	ne	60	str	local	0.01	23.69	100	12.59	17.84	49.74	70.49
6	GLS2	ne	80	non.str	global	12.61	21.28	100	16.28	20.75	64.63	82.37
6	GLS2	ne	80	str	global	5.95	20.71	100	16.09	20.7	63.87	82.18
6	GLS2	ne	80	non.str	local	0	23.35	100	11.56	17	45.89	67.49
6	GLS2	ne	80	str	local	0	23.38	100	11.52	16.95	45.73	67.29
6	GLS2	ne	100	non.str	global	8.69	20.95	100	16.25	20.84	64.77	83.06
6	GLS2	ne	100	str	global	5.36	20.44	100	16.07	20.75	64.05	82.70
6	GLS2	ne	100	non.str	local	0.02	23.33	100	11.27	16.44	44.92	65.52
6	GLS2	ne	100	str	local	0.02	23.36	100	11.34	16.62	45.20	66.24
6	GLS2	sw	20	non.str	global	14.62	41.81	97.18	21.06	25.85	43.65	53.58
6	GLS2	sw	20	str	global	0.06	39.05	98.05	20.07	30.22	41.60	62.63
6	GLS2	sw	20	non.str	local	5.7	40.63	99.2	18.21	25.44	37.74	52.73
6	GLS2	sw	20	str	local	0.06	39.05	98.05	20.07	30.22	41.60	62.63
6	GLS2	sw	40	non.str	global	18.69	43.5	99.87	27.43	30.65	57.04	63.73
6	GLS2	sw	40	str	global	0.03	43.43	100	21.01	27.31	43.90	57.06
6	GLS2	sw	40	non.str	local	0.01	44.63	99.2	15.38	20.62	31.98	42.88
6	GLS2	sw	40	str	local	0	43.39	100	15.75	21.47	32.91	44.86
6	GLS2	sw	60	non.str	global	14.58	42.76	100	25.09	28.97	52.57	60.70
6	GLS2	sw	60	str	global	0.01	40.46	99.54	23.3	31.26	48.82	65.49
6	GLS2	sw	60	non.str	local	0.05	42.3	99.98	15.88	24.04	33.27	50.37
6	GLS2	sw	60	str	local	0	43.13	99.81	16.9	26.43	35.41	55.37
6	GLS2	sw	80	non.str	global	16.64	42.74	99.8	26.03	29.9	57.01	65.48
6	GLS2	sw	80	str	global	0	39.12	99.78	23.11	29.63	50.61	64.89
6	GLS2	sw	80	non.str	local	0	42.14	98.53	16.49	22.87	36.11	50.09
6	GLS2	sw	80	str	local	0.01	40.17	98.53	16.64	24.35	36.44	53.33
6	GLS2	sw	100	non.str	global	23.34	45.08	97.23	25.5	28.43	55.16	61.50
6	GLS2	sw	100	str	global	0.03	40.33	99.03	20.77	26.63	44.93	57.60
6	GLS2	sw	100	non.str	local	0	42.91	96.63	13.98	20.64	30.24	44.65
6	GLS2	sw	100	str	local	0.01	41.78	99.16	13.89	21.46	30.05	46.42
7	KED1	n	20	non.str	global	2.36	28.66	85.08	14.05	19.13	43.63	59.41
7	KED1	n	20	str	global	2.18	28.66	89.43	13.83	18.79	42.95	58.35
7	KED1	n	20	non.str	local	0.31	29.43	90.68	13.66	18.38	42.42	57.08
7	KED1	n	20	str	local	0.31	29.28	90.68	13.64	18.3	42.36	56.83
7	KED1	n	40	non.str	global	2.58	30.78	75.65	14	19.45	41.02	56.99
7	KED1	n	40	str	global	0.07	30.83	75.61	13.96	19.51	40.90	57.16
7	KED1	n	40	non.str	local	0.09	31.41	86.27	13.66	19.27	40.02	56.46
7	KED1	n	40	str	local	0.01	31.49	92.5	13.53	19.17	39.64	56.17
7	KED1	n	60	non.str	global	2.1	30.42	94.3	13.42	18.98	39.89	56.42
7	KED1	n	60	str	global	1.21	30.47	80.28	13.42	19.01	39.89	56.51
7	KED1	n	60	non.str	local	0	31.4	84.99	13.11	18.97	38.97	56.39
7	KED1	n	60	str	local	0	31.4	90.57	13.06	18.87	38.82	56.09
7	KED1	n	80	non.str	global	1.96	31.11	100	12.79	18.24	37.57	53.58
7	KED1	n	80	str	global	1.97	31.16	100	12.83	18.31	37.69	53.79
7	KED1	n	80	non.str	local	0.01	31.74	100	12.49	18.31	36.69	53.79
7	KED1	n	80	str	local	0.01	31.82	100	12.56	18.38	36.90	54.00
7	KED1	n	100	non.str	global	0.65	30.93	92.24	12.92	18.41	37.96	54.08
7	KED1	n	100	str	global	0.11	31	100	12.91	18.41	37.93	54.08
7	KED1	n	100	non.str	local	0	31.63	100	12.61	18.4	37.04	54.05
7	KED1	n	100	str	local	0	31.67	100	12.6	18.37	37.02	53.97
7	KED1	ne	20	non.str	global	1.35	22.67	100	11.61	16.01	45.83	63.21
7	KED1	ne	20	str	global	1.47	22.59	100	11.71	16.12	46.60	64.15
7	KED1	ne	20	non.str	local	1.28	23.3	100	11.9	16.67	46.98	65.81
7	KED1	ne	20	str	local	1.28	23.15	100	12.03	16.7	47.87	66.45
7	KED1	ne	40	non.str	global	0.96	22.65	100	11.33	15.64	44.55	61.50
7	KED1	ne	40	str	global	0.81	22.63	95.63	11.34	15.68	44.75	61.88
7	KED1	ne	40	non.str	local	0.29	23.2	100	11.3	15.89	44.44	62.49
7	KED1	ne	40	str	local	0.29	23.08	95.63	11.32	15.79	44.67	62.31
7	KED1	ne	60	non.str	global	1.65	22.72	100	10.62	14.93	41.84	58.83
7	KED1	ne	60	str	global	1.48	22.69	96.14	10.66	15.01	42.12	59.30
7	KED1	ne	60	non.str	local	0.23	23.46	100	10.92	15.59	43.03	61.43
7	KED1	ne	60	str	local	0.23	23.23	95.38	10.93	15.58	43.18	61.56

Appendix

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
7	KED1	ne	80	non.str	global	1.09	22.6	100	10.24	14.54	40.65	57.72
7	KED1	ne	80	str	global	1.01	22.66	100	10.24	14.57	40.65	57.84
7	KED1	ne	80	non.str	local	0.09	23.24	100	10.42	15.17	41.37	60.22
7	KED1	ne	80	str	local	0.09	23.18	100	10.43	15.13	41.41	60.06
7	KED1	ne	100	non.str	global	0.73	22.6	100	9.87	14.05	39.34	56.00
7	KED1	ne	100	str	global	0.66	22.66	100	9.87	14.07	39.34	56.08
7	KED1	ne	100	non.str	local	0.07	23.2	100	10.09	14.65	40.22	58.39
7	KED1	ne	100	str	local	0.07	23.16	100	10.08	14.66	40.18	58.43
7	KED1	sw	20	non.str	global	0.41	45.82	99.99	18.86	23.44	39.09	48.58
7	KED1	sw	20	str	global	0.04	44.47	98.76	19.92	28.04	41.28	58.11
7	KED1	sw	20	non.str	local	0.26	46.15	100	16.33	21.14	33.84	43.81
7	KED1	sw	20	str	local	0.04	44.47	98.76	19.92	28.04	41.28	58.11
7	KED1	sw	40	non.str	global	0.09	43.62	98.44	14.04	19.3	29.20	40.13
7	KED1	sw	40	str	global	0.01	45.47	95.89	13.04	18.96	27.25	39.62
7	KED1	sw	40	non.str	local	0.06	45.41	95.06	11.49	15.73	23.89	32.71
7	KED1	sw	40	str	local	0	46.24	96.2	11.69	17.47	24.43	36.50
7	KED1	sw	60	non.str	global	0	46.92	98.74	12.71	18.77	26.63	39.33
7	KED1	sw	60	str	global	0	47.69	96.03	12.81	19.5	26.84	40.85
7	KED1	sw	60	non.str	local	0	47.15	98.52	12.22	18.6	25.60	38.97
7	KED1	sw	60	str	local	0	47.83	99.78	12.46	19.47	26.11	40.79
7	KED1	sw	80	non.str	global	0.24	42.94	93.98	13.88	19.22	30.40	42.09
7	KED1	sw	80	str	global	0	44.16	96.63	13.56	18.88	29.70	41.35
7	KED1	sw	80	non.str	local	0.03	44.23	97.66	12.31	17.88	26.96	39.16
7	KED1	sw	80	str	local	0	44.68	99.47	12.31	18.26	26.96	39.99
7	KED1	sw	100	non.str	global	0	44.3	96.38	11.38	16.61	24.62	35.93
7	KED1	sw	100	str	global	0	44.76	99.78	11.58	17.09	25.05	36.97
7	KED1	sw	100	non.str	local	0	45.05	97.59	10.65	16.35	23.04	35.37
7	KED1	sw	100	str	local	0.03	45.37	99.78	10.93	16.8	23.64	36.34
8	KED2	n	20	non.str	global	1.23	28.63	85.12	14.12	19.23	43.85	59.72
8	KED2	n	20	str	global	1.36	28.64	91.6	13.49	18.44	41.89	57.27
8	KED2	n	20	non.str	local	0.08	29.77	100	14.91	21.28	46.30	66.09
8	KED2	n	20	str	local	0.08	29.64	100	14.39	20.65	44.69	64.13
8	KED2	n	40	non.str	global	0.1	30.82	75.68	14.02	19.5	41.08	57.13
8	KED2	n	40	str	global	0.01	30.95	80.1	14.06	19.68	41.20	57.66
8	KED2	n	40	non.str	local	0.03	32.15	100	14.22	19.97	41.66	58.51
8	KED2	n	40	str	local	0.03	32.15	100	14.17	19.97	41.52	58.51
8	KED2	n	60	non.str	global	1.11	30.45	100	13.46	18.98	40.01	56.42
8	KED2	n	60	str	global	1.4	30.48	86.93	13.47	19.08	40.04	56.72
8	KED2	n	60	non.str	local	0	31.86	100	14.01	20.46	41.65	60.82
8	KED2	n	60	str	local	0	31.97	100	14.11	20.65	41.94	61.39
8	KED2	n	80	non.str	global	1.99	31.15	99.22	12.86	18.25	37.78	53.61
8	KED2	n	80	str	global	1.57	31.19	100	12.91	18.37	37.93	53.97
8	KED2	n	80	non.str	local	0.08	32.24	100	13.1	19.24	38.48	56.52
8	KED2	n	80	str	local	0.08	32.39	100	13.16	19.4	38.66	56.99
8	KED2	n	100	non.str	global	0.97	30.94	88.15	12.97	18.42	38.10	54.11
8	KED2	n	100	str	global	1.51	31	86.37	12.97	18.45	38.10	54.20
8	KED2	n	100	non.str	local	0	31.94	100	12.8	18.69	37.22	54.35
8	KED2	n	100	str	local	0	32	100	12.84	18.86	37.34	54.84
8	KED2	ne	20	non.str	global	1.59	22.69	100	11.61	16.05	45.83	63.36
8	KED2	ne	20	str	global	1.87	22.71	100	11.74	16.17	46.72	64.35
8	KED2	ne	20	non.str	local	0.19	23.15	100	12.42	17.83	49.03	70.39
8	KED2	ne	20	str	local	0.19	23.15	100	12.13	17.09	48.27	68.01
8	KED2	ne	40	non.str	global	0.94	22.71	100	11.35	15.78	44.63	62.05
8	KED2	ne	40	str	global	0.63	22.71	100	11.38	15.85	44.91	62.55
8	KED2	ne	40	non.str	local	0.06	23.69	100	11.49	16.7	45.18	65.67
8	KED2	ne	40	str	local	0.06	23.71	100	11.59	16.65	45.74	65.71
8	KED2	ne	60	non.str	global	1.64	22.73	100	10.64	14.97	41.92	58.98
8	KED2	ne	60	str	global	1.26	22.76	100	10.66	15.16	42.12	59.90
8	KED2	ne	60	non.str	local	0.04	23.87	100	11.2	16.52	44.13	65.09
8	KED2	ne	60	str	local	0.04	23.78	100	11.18	16.48	44.17	65.11
8	KED2	ne	80	non.str	global	1.07	22.63	100	10.23	14.56	40.61	57.80
8	KED2	ne	80	str	global	1.03	22.74	100	10.24	14.72	40.65	58.44
8	KED2	ne	80	non.str	local	0.01	23.47	100	10.57	16.05	41.96	63.72
8	KED2	ne	80	str	local	0.01	23.46	100	10.53	15.93	41.80	63.24
8	KED2	ne	100	non.str	global	0.73	22.64	100	9.83	14.05	39.18	56.00
8	KED2	ne	100	str	global	0.65	22.75	100	9.82	14.14	39.14	56.36
8	KED2	ne	100	non.str	local	0.02	23.56	100	10.16	15.28	40.49	60.90
8	KED2	ne	100	str	local	0.02	23.61	100	10.19	15.39	40.61	61.34
8	KED2	sw	20	non.str	global	0.29	44.06	96.85	18.58	23.07	38.51	47.81
8	KED2	sw	20	str	global	0.13	40.7	98.58	20.67	30.38	42.84	62.96
8	KED2	sw	20	non.str	local	0.26	43.82	98.37	16.33	21.78	33.84	45.14
8	KED2	sw	20	str	local	0.13	40.7	98.58	20.67	30.38	42.84	62.96
8	KED2	sw	40	non.str	global	0	43.76	99.03	14.34	20.04	29.82	41.67
8	KED2	sw	40	str	global	0	45.73	100	12.07	18.5	25.22	38.65
8	KED2	sw	40	non.str	local	0.03	46.39	99.96	11.16	15.62	23.21	32.48
8	KED2	sw	40	str	local	0.06	45.42	100	10.49	15.31	21.92	31.99
8	KED2	sw	60	non.str	global	0	45.65	100	15.53	22.05	32.54	46.20
8	KED2	sw	60	str	global	0	43.06	99.87	18.2	27.93	38.13	58.52
8	KED2	sw	60	non.str	local	0	47.26	99.92	13.45	21.39	28.18	44.81
8	KED2	sw	60	str	local	0	44.01	99.87	16.41	26.77	34.38	56.09

Appendix

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
8	KED2	sw	80	non.str	global	0.35	42.65	99.8	15.09	20.69	33.05	45.31
8	KED2	sw	80	str	global	0	41.57	99.45	16.58	24.18	36.31	52.96
8	KED2	sw	80	non.str	local	0	43.97	99.39	14.64	21.14	32.06	46.30
8	KED2	sw	80	str	local	0.01	41.72	99.45	15.46	23.47	33.86	51.40
8	KED2	sw	100	non.str	global	0.01	43.96	97.19	12.34	17.6	26.69	38.07
8	KED2	sw	100	str	global	0	43.7	99.03	13.05	19.66	28.23	42.53
8	KED2	sw	100	non.str	local	0	44.53	98.44	11.92	18.52	25.78	40.06
8	KED2	sw	100	str	local	0.02	44.01	99.79	11.69	18.22	25.29	39.41
9	KED3	n	20	non.str	global	1.95	28.66	92.01	13.98	19.24	43.42	59.75
9	KED3	n	20	str	global	1.97	28.8	100	14.43	19.91	44.81	61.83
9	KED3	n	20	non.str	local	0	34.96	100	20.06	29.66	62.30	92.11
9	KED3	n	20	str	local	0	35.3	100	20.2	29.99	62.73	93.14
9	KED3	n	40	non.str	global	0.71	30.76	100	14.82	20.67	43.42	60.56
9	KED3	n	40	str	global	0.2	31.14	100	15.16	21.24	44.42	62.23
9	KED3	n	40	non.str	local	0	35.96	100	19.04	28.08	55.79	82.27
9	KED3	n	40	str	local	0	35.91	100	19.31	28.5	56.58	83.50
9	KED3	n	60	non.str	global	0.01	30.38	100	13.88	19.52	41.26	58.03
9	KED3	n	60	str	global	0.55	30.53	100	14.17	20.04	42.12	59.57
9	KED3	n	60	non.str	local	0	37.06	100	18.9	28.92	56.18	85.97
9	KED3	n	60	str	local	0	36.77	100	18.94	28.85	56.30	85.76
9	KED3	n	80	non.str	global	1.32	31.05	100	13.34	18.86	39.19	55.41
9	KED3	n	80	str	global	0.33	31.13	100	13.48	19.09	39.60	56.08
9	KED3	n	80	non.str	local	0	36.02	100	17.76	26.67	52.17	78.35
9	KED3	n	80	str	local	0	36.09	100	17.73	26.56	52.09	78.03
9	KED3	n	100	non.str	global	0.27	30.81	100	13.42	18.96	39.42	55.70
9	KED3	n	100	str	global	0.95	30.97	100	13.52	19.1	39.72	56.11
9	KED3	n	100	non.str	local	0	36.35	100	17.03	25.47	49.52	74.06
9	KED3	n	100	str	local	0	36.35	100	17.03	25.47	49.52	74.06
9	KED3	ne	20	non.str	global	1.83	23.08	100	12.22	17.16	48.24	67.75
9	KED3	ne	20	str	global	0.03	22.78	100	12.94	19.02	51.49	75.69
9	KED3	ne	20	non.str	local	0.01	29.83	100	17.37	26.76	68.57	105.65
9	KED3	ne	20	str	local	0.01	28.8	100	17.61	27.02	70.08	107.52
9	KED3	ne	40	non.str	global	0.8	22.91	100	11.37	16.03	44.71	63.04
9	KED3	ne	40	str	global	0.63	22.88	100	11.39	16	44.95	63.14
9	KED3	ne	40	non.str	local	0	28.16	100	14.83	22.88	58.32	89.97
9	KED3	ne	40	str	local	0	27.81	100	14.78	22.87	58.33	90.25
9	KED3	ne	60	non.str	global	1.29	22.91	100	10.61	15.05	41.80	59.30
9	KED3	ne	60	str	global	1.19	23.07	100	10.85	15.67	42.87	61.91
9	KED3	ne	60	non.str	local	0	29.2	100	15.8	25.68	62.25	101.18
9	KED3	ne	60	str	local	0	29.09	100	15.72	25.41	62.11	100.40
9	KED3	ne	80	non.str	global	1.07	22.66	100	10.2	14.61	40.49	58.00
9	KED3	ne	80	str	global	1.12	22.85	100	10.29	14.86	40.85	58.99
9	KED3	ne	80	non.str	local	0	27.6	100	14.38	22.95	57.09	91.11
9	KED3	ne	80	str	local	0	27.56	100	14.38	22.88	57.09	90.83
9	KED3	ne	100	non.str	global	0.76	22.78	100	9.79	14.21	39.02	56.64
9	KED3	ne	100	str	global	0.78	22.89	100	9.99	14.77	39.82	58.87
9	KED3	ne	100	non.str	local	0	27.2	100	13.49	21.76	53.77	86.73
9	KED3	ne	100	str	local	0	26.89	100	13.48	21.82	53.73	86.97
9	KED3	sw	20	non.str	global	0.45	47.18	98.11	16.89	24.08	35.01	49.91
9	KED3	sw	20	str	global	0.06	53.44	99.4	27.37	37.01	56.73	76.70
9	KED3	sw	20	non.str	local	0.47	35.45	99.98	26.35	36.2	54.61	75.03
9	KED3	sw	20	str	local	0.06	53.44	99.4	27.37	37.01	56.73	76.70
9	KED3	sw	40	non.str	global	0.04	45.95	99.8	12.63	18.82	26.26	39.13
9	KED3	sw	40	str	global	0.02	43.25	99.91	16.08	27.75	33.60	57.98
9	KED3	sw	40	non.str	local	0.01	45.26	99.91	14.6	24.09	30.36	50.09
9	KED3	sw	40	str	local	0	44.77	99.21	15.92	25.73	33.26	53.76
9	KED3	sw	60	non.str	global	0.03	47.15	99.98	12.57	19.16	26.34	40.14
9	KED3	sw	60	str	global	0.03	45.68	98.64	14.64	24.33	30.67	50.97
9	KED3	sw	60	non.str	local	0	48.67	99.97	14.35	24.64	30.06	51.62
9	KED3	sw	60	str	local	0	47.21	99.45	15.64	25.84	32.77	54.14
9	KED3	sw	80	non.str	global	0.02	42.73	99.93	15.03	21.88	32.92	47.92
9	KED3	sw	80	str	global	0	42.79	95.24	12.61	19.77	27.62	43.30
9	KED3	sw	80	non.str	local	0	48.09	98.81	14.86	23.16	32.54	50.72
9	KED3	sw	80	str	local	0	47.69	99.82	14.92	24.01	32.68	52.58
9	KED3	sw	100	non.str	global	0	42.65	100	14.26	21.59	30.85	46.70
9	KED3	sw	100	str	global	0	44.22	99.95	11.99	19.65	25.94	42.50
9	KED3	sw	100	non.str	local	0	48.13	99.83	14.12	23.64	30.54	51.14
9	KED3	sw	100	str	local	0	47.29	99.95	13.78	22.29	29.81	48.22
10	KED4	n	20	non.str	global	0.97	28.6	86.74	14.19	19.44	44.07	60.37
10	KED4	n	20	str	global	1.12	29.64	100	15.13	21.11	46.99	65.56
10	KED4	n	20	non.str	local	0	45.22	100	30.34	42.18	94.22	130.99
10	KED4	n	20	str	local	0	46.6	100	30.99	42.98	96.24	133.48
10	KED4	n	40	non.str	global	0.02	30.79	100	14.88	20.8	43.60	60.94
10	KED4	n	40	str	global	0.18	31.87	100	15.9	22.94	46.59	67.21
10	KED4	n	40	non.str	local	0	44.68	100	28.26	39.51	82.80	115.76
10	KED4	n	40	str	local	0	44.41	100	28.77	40.01	84.30	117.23
10	KED4	n	60	non.str	global	0	30.36	100	14.09	19.76	41.88	58.74
10	KED4	n	60	str	global	0.29	30.88	100	14.46	20.52	42.98	61.00
10	KED4	n	60	non.str	local	0	44.91	100	27.9	39	82.94	115.93
10	KED4	n	60	str	local	0	45.33	100	28.17	39.34	83.74	116.94

Appendix

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
10	KED4	n	80	non.str	global	1.2	30.97	100	13.53	19.05	39.75	55.96
10	KED4	n	80	str	global	0.38	31.3	100	13.81	19.49	40.57	57.26
10	KED4	n	80	non.str	local	0	44.87	100	28.04	39.04	82.37	114.69
10	KED4	n	80	str	local	0	44.7	100	27.78	38.75	81.61	113.84
10	KED4	n	100	non.str	global	0.11	30.75	100	13.56	19.09	39.84	56.08
10	KED4	n	100	str	global	0.02	31	100	13.65	19.14	40.10	56.23
10	KED4	n	100	non.str	local	0	44.13	100	27.23	38.61	79.18	112.27
10	KED4	n	100	str	local	0	44.13	100	27.26	38.65	79.27	112.39
10	KED4	ne	20	non.str	global	2.14	22.88	100	12.34	17.27	48.72	68.18
10	KED4	ne	20	str	global	0.34	23.21	100	12.65	18.12	50.34	72.11
10	KED4	ne	20	non.str	local	0.04	43.27	100	29.9	41.85	118.04	165.22
10	KED4	ne	20	str	local	0.04	42.74	100	29.63	41.6	117.91	165.54
10	KED4	ne	40	non.str	global	1.18	22.84	100	11.52	16.31	45.30	64.14
10	KED4	ne	40	str	global	0.37	23.47	100	12.13	17.62	47.87	69.53
10	KED4	ne	40	non.str	local	0	37.92	100	26.81	38.65	105.43	151.99
10	KED4	ne	40	str	local	0	38.45	100	27.04	38.92	106.71	153.59
10	KED4	ne	60	non.str	global	1.63	22.86	100	10.81	15.34	42.59	60.44
10	KED4	ne	60	str	global	0.39	23.04	100	11.25	16.44	44.45	64.95
10	KED4	ne	60	non.str	local	0	37.64	100	25.24	36.83	99.45	145.11
10	KED4	ne	60	str	local	0	37.17	100	25.41	36.95	100.40	145.99
10	KED4	ne	80	non.str	global	1.45	22.58	100	10.35	14.85	41.09	58.95
10	KED4	ne	80	str	global	1.38	22.99	100	10.48	15.28	41.60	60.66
10	KED4	ne	80	non.str	local	0	36.44	100	23.31	34.51	92.54	137.00
10	KED4	ne	80	str	local	0	36.28	100	23.5	34.63	93.29	137.48
10	KED4	ne	100	non.str	global	0.01	22.73	100	9.96	14.41	39.70	57.43
10	KED4	ne	100	str	global	0	23.08	100	10.16	15.1	40.49	60.18
10	KED4	ne	100	non.str	local	0	34.5	100	21.4	32.39	85.29	129.10
10	KED4	ne	100	str	local	0	34.65	100	21.65	32.74	86.29	130.49
10	KED4	sw	20	non.str	global	0.01	39.1	99.83	26.11	34.94	54.11	72.41
10	KED4	sw	20	str	global	0.04	49.85	99.78	37.99	48.88	78.74	101.31
10	KED4	sw	20	non.str	local	0.01	40.74	99.95	35.63	44.98	73.84	93.22
10	KED4	sw	20	str	local	0.04	49.85	99.78	37.99	48.88	78.74	101.31
10	KED4	sw	40	non.str	global	0.01	43.29	98.66	16.84	24.78	35.02	51.53
10	KED4	sw	40	str	global	0.03	52.87	99.02	24.34	34.92	50.86	72.96
10	KED4	sw	40	non.str	local	0.05	49.01	100	22.33	32.22	46.43	67.00
10	KED4	sw	40	str	local	0.01	45.75	99.97	24.79	35.87	51.80	74.95
10	KED4	sw	60	non.str	global	0.05	45.5	96.53	12.75	19.55	26.71	40.96
10	KED4	sw	60	str	global	0	45.38	99.08	14.54	22.9	30.46	47.98
10	KED4	sw	60	non.str	local	0	50.54	99.99	28.12	38.5	58.91	80.66
10	KED4	sw	60	str	local	0	48.68	99.97	23.49	34.72	49.21	72.74
10	KED4	sw	80	non.str	global	0	43.37	97.09	13.26	19.64	29.04	43.01
10	KED4	sw	80	str	global	0	45.52	99.07	15.49	24.93	33.92	54.60
10	KED4	sw	80	non.str	local	0.01	44.02	99.99	26.93	38.01	58.98	83.25
10	KED4	sw	80	str	local	0	44.41	99.76	24.14	35.44	52.87	77.62
10	KED4	sw	100	non.str	global	0	43.82	99.73	13	19.62	28.12	42.44
10	KED4	sw	100	str	global	0	48.62	99.85	15.38	25.74	33.27	55.68
10	KED4	sw	100	non.str	local	0	43.32	100	20.27	29.78	43.85	64.42
10	KED4	sw	100	str	local	0	44.17	100	18.48	26.96	39.97	58.32
11	OCK1	n	20	non.str	global	2.31	28.37	76.89	14.48	19.5	44.97	60.56
11	OCK1	n	20	str	global	2.3	28.64	89.84	14.43	19.38	44.81	60.19
11	OCK1	n	20	non.str	local	2	28.48	77.47	14.47	19.43	44.94	60.34
11	OCK1	n	20	str	local	1.96	28.72	89.84	14.43	19.31	44.81	59.97
11	OCK1	n	40	non.str	global	2.91	30.6	74.75	14.51	20.02	42.51	58.66
11	OCK1	n	40	str	global	2.83	30.59	78.76	14.5	20.02	42.48	58.66
11	OCK1	n	40	non.str	local	2.34	30.78	79.7	14.32	19.9	41.96	58.31
11	OCK1	n	40	str	local	2.39	30.79	79.72	14.3	19.89	41.90	58.28
11	OCK1	n	60	non.str	global	2.33	30.18	76.73	13.95	19.38	41.47	57.61
11	OCK1	n	60	str	global	2.34	30.17	78.39	13.87	19.37	41.23	57.58
11	OCK1	n	60	non.str	local	2.33	30.58	84.84	13.55	19.04	40.28	56.60
11	OCK1	n	60	str	local	2.27	30.6	84.58	13.5	19.03	40.13	56.57
11	OCK1	n	80	non.str	global	2.49	30.84	81.65	13.45	18.89	39.51	55.49
11	OCK1	n	80	str	global	2.55	30.9	81.53	13.4	18.89	39.37	55.49
11	OCK1	n	80	non.str	local	1.93	31.26	89.43	13.12	18.71	38.54	54.96
11	OCK1	n	80	str	local	1.88	31.34	89.6	13.06	18.67	38.37	54.85
11	OCK1	n	100	non.str	global	2.31	30.7	82.42	13.44	18.93	39.48	55.61
11	OCK1	n	100	str	global	2.27	30.68	82.23	13.47	18.96	39.57	55.70
11	OCK1	n	100	non.str	local	1.24	31.26	90.57	13.04	18.61	38.31	54.67
11	OCK1	n	100	str	local	1.12	31.21	90.72	13.09	18.64	38.45	54.76
11	OCK1	ne	20	non.str	global	8.07	21.03	63.71	13.66	18.06	53.93	71.30
11	OCK1	ne	20	str	global	4.81	22.25	100	12.56	17.02	49.98	67.73
11	OCK1	ne	20	non.str	local	5.12	22.51	83.67	12.72	17.05	50.22	67.31
11	OCK1	ne	20	str	local	3	22.8	100	12.41	16.84	49.38	67.01
11	OCK1	ne	40	non.str	global	6.86	21.34	100	13.77	18.2	54.15	71.57
11	OCK1	ne	40	str	global	2.91	22.06	100	12.31	16.8	48.58	66.30
11	OCK1	ne	40	non.str	local	1.87	22.19	100	12.71	17.37	49.98	68.31
11	OCK1	ne	40	str	local	2.05	22.29	100	12.15	16.7	47.95	65.90
11	OCK1	ne	60	non.str	global	5.93	21.69	100	12.62	17.16	49.72	67.61
11	OCK1	ne	60	str	global	3.72	22.24	82.05	11.77	16.28	46.50	64.32
11	OCK1	ne	60	non.str	local	2.04	22.55	100	11.94	16.56	47.04	65.25
11	OCK1	ne	60	str	local	2.26	22.6	80.23	11.56	16.12	45.67	63.69

Appendix

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
11	OCK1	ne	80	non.str	global	5.31	21.52	100	12.53	17.05	49.74	67.69
11	OCK1	ne	80	str	global	2.88	21.99	85.34	11.36	15.92	45.28	63.45
11	OCK1	ne	80	non.str	local	1.68	22.3	100	11.43	16.14	45.38	64.07
11	OCK1	ne	80	str	local	1.51	22.29	84.77	11.02	15.61	43.92	62.22
11	OCK1	ne	100	non.str	global	4.74	21.47	100	12.17	16.67	48.51	66.44
11	OCK1	ne	100	str	global	2.38	21.94	100	11.12	15.5	44.44	61.95
11	OCK1	ne	100	non.str	local	1.24	22.23	100	11.02	15.66	43.92	62.42
11	OCK1	ne	100	str	local	1.35	22.26	100	10.75	15.25	42.97	60.95
11	OCK1	sw	20	non.str	global	8.39	40.74	99.13	21.05	27.61	43.63	57.22
11	OCK1	sw	20	str	global	1.09	33.25	77.47	20.69	25.92	52.21	65.40
11	OCK1	sw	20	non.str	local	7.92	42.65	97.94	20.18	27.5	41.82	56.99
11	OCK1	sw	20	str	local	1.09	33.25	77.47	20.69	25.92	52.21	65.40
11	OCK1	sw	40	non.str	global	9.03	39.82	99.85	23.62	30.13	49.12	62.65
11	OCK1	sw	40	str	global	0	43.67	93.54	19.89	27.7	41.56	57.88
11	OCK1	sw	40	non.str	local	0.75	42.18	99.58	21.04	26.88	43.75	55.90
11	OCK1	sw	40	str	local	0	43.58	90.06	19.1	26.72	39.91	55.83
11	OCK1	sw	60	non.str	global	7.89	39.41	99.99	21.4	27.38	44.84	57.36
11	OCK1	sw	60	str	global	0.07	45.88	99.79	13.59	20.12	29.58	43.79
11	OCK1	sw	60	non.str	local	0.14	42.74	99.95	17.29	23.37	36.22	48.96
11	OCK1	sw	60	str	local	0.1	45.46	99.79	13.35	20.11	29.05	43.76
11	OCK1	sw	80	non.str	global	8.28	37.7	99.93	22.73	29.34	49.78	64.26
11	OCK1	sw	80	str	global	0.02	41.6	96.78	14.14	19.69	33.74	46.98
11	OCK1	sw	80	non.str	local	0.14	43.6	99.96	17.6	23.06	38.55	50.50
11	OCK1	sw	80	str	local	0.01	42.48	96.78	13.82	19.4	32.98	46.29
11	OCK1	sw	100	non.str	global	7.53	37.96	99.99	21.92	28.55	47.42	61.76
11	OCK1	sw	100	str	global	0	46.71	95.56	12.49	18.25	26.81	39.17
11	OCK1	sw	100	non.str	local	0.12	45.47	100	16.02	21.82	34.65	47.20
11	OCK1	sw	100	str	local	0.02	47.19	95.56	11.57	17.49	24.83	37.54
12	OCK2	n	20	non.str	global	2.55	28.66	74.11	14.69	19.75	45.62	61.34
12	OCK2	n	20	str	global	2.21	28.87	91.42	14.84	19.83	46.09	61.58
12	OCK2	n	20	non.str	local	2.52	28.7	77.23	14.63	19.68	45.43	61.12
12	OCK2	n	20	str	local	2.14	28.61	91.42	14.81	19.77	45.99	61.40
12	OCK2	n	40	non.str	global	2.52	30.68	78.15	14.71	20.21	43.10	59.21
12	OCK2	n	40	str	global	2.48	30.65	81.2	14.71	20.24	43.10	59.30
12	OCK2	n	40	non.str	local	2.25	30.73	79.61	14.49	20.06	42.46	58.78
12	OCK2	n	40	str	local	2.2	30.59	80.24	14.43	20.03	42.28	58.69
12	OCK2	n	60	non.str	global	2.09	30.24	75.64	14.14	19.59	42.03	58.23
12	OCK2	n	60	str	global	1.9	30.19	76.7	14.08	19.66	41.85	58.44
12	OCK2	n	60	non.str	local	1.75	30.51	84.84	13.74	19.23	40.84	57.16
12	OCK2	n	60	str	local	1.47	30.43	84.59	13.67	19.26	40.64	57.25
12	OCK2	n	80	non.str	global	2.64	30.83	80.82	13.68	19.11	40.19	56.14
12	OCK2	n	80	str	global	2.79	30.86	80.26	13.62	19.14	40.01	56.23
12	OCK2	n	80	non.str	local	1.66	31.15	89.26	13.31	18.91	39.10	55.55
12	OCK2	n	80	str	local	1.55	31.11	89.39	13.23	18.86	38.87	55.41
12	OCK2	n	100	non.str	global	2.25	30.69	81.5	13.64	19.13	40.07	56.20
12	OCK2	n	100	str	global	2.14	30.66	80.8	13.67	19.23	40.16	56.49
12	OCK2	n	100	non.str	local	1.05	31.15	90.36	13.18	18.76	38.72	55.11
12	OCK2	n	100	str	local	0.88	31.02	90.45	13.23	18.81	38.87	55.26
12	OCK2	ne	20	non.str	global	6.72	21.44	71.8	12.97	17.52	51.20	69.17
12	OCK2	ne	20	str	global	5.03	22.23	100	12.55	17.01	49.94	67.69
12	OCK2	ne	20	non.str	local	4.94	22.36	93.28	12.36	16.85	48.80	66.52
12	OCK2	ne	20	str	local	3.19	22.77	100	12.4	16.84	49.34	67.01
12	OCK2	ne	40	non.str	global	4.89	21.59	100	12.93	17.39	50.85	68.38
12	OCK2	ne	40	str	global	3.06	21.98	100	12.31	16.88	48.58	66.61
12	OCK2	ne	40	non.str	local	2.06	22	100	12.22	16.83	48.05	66.18
12	OCK2	ne	40	str	local	2	22.21	100	12.14	16.77	47.91	66.18
12	OCK2	ne	60	non.str	global	5.58	21.86	100	12.03	16.59	47.40	65.37
12	OCK2	ne	60	str	global	3.83	22.26	89.55	11.78	16.35	46.54	64.60
12	OCK2	ne	60	non.str	local	2.22	22.41	100	11.59	16.15	45.67	63.63
12	OCK2	ne	60	str	local	2.23	22.61	88.5	11.57	16.19	45.71	63.97
12	OCK2	ne	80	non.str	global	4.99	21.71	100	11.79	16.37	46.80	64.99
12	OCK2	ne	80	str	global	2.88	21.98	72.49	11.38	15.97	45.36	63.65
12	OCK2	ne	80	non.str	local	1.57	22.28	100	11.09	15.76	44.03	62.56
12	OCK2	ne	80	str	local	1.47	22.27	76.24	11.03	15.63	43.96	62.30
12	OCK2	ne	100	non.str	global	4.58	21.66	100	11.44	15.94	45.60	63.53
12	OCK2	ne	100	str	global	2.52	21.94	100	11.13	15.58	44.48	62.27
12	OCK2	ne	100	non.str	local	1.29	22.23	100	10.71	15.28	42.69	60.90
12	OCK2	ne	100	str	local	1.33	22.28	100	10.76	15.32	43.01	61.23
12	OCK2	sw	20	non.str	global	6.7	41.39	99.94	21.37	28.84	44.29	59.77
12	OCK2	sw	20	str	global	0.52	33.01	75.74	20.71	26.31	52.26	66.39
12	OCK2	sw	20	non.str	local	2.59	40.85	99.92	20.34	29.04	42.16	60.19
12	OCK2	sw	20	str	local	0.52	33.01	75.74	20.71	26.31	52.26	66.39
12	OCK2	sw	40	non.str	global	1.54	43.05	99.29	16.5	22.98	34.31	47.79
12	OCK2	sw	40	str	global	0	47.92	99.77	16.2	22.85	33.85	47.74
12	OCK2	sw	40	non.str	local	0.02	41.36	99.98	15.26	22.35	31.73	46.48
12	OCK2	sw	40	str	local	0	47.02	99.79	15.99	22.47	33.41	46.95
12	OCK2	sw	60	non.str	global	4.07	42.12	100	16.97	22.49	35.55	47.12
12	OCK2	sw	60	str	global	0.07	45.76	99.88	13.56	20.1	29.51	43.74
12	OCK2	sw	60	non.str	local	0.17	42.56	99.96	14.71	20.65	30.82	43.26
12	OCK2	sw	60	str	local	0.01	45.16	99.88	13.32	20.02	28.99	43.57

Appendix

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
12	OCK2	sw	80	non.str	global	4.48	39.52	98.99	18.4	24.45	40.30	53.55
12	OCK2	sw	80	str	global	0	41.44	96.01	14.11	19.54	33.67	46.62
12	OCK2	sw	80	non.str	local	0.14	43.05	100	15.32	20.88	33.55	45.73
12	OCK2	sw	80	str	local	0.02	42.03	95.97	13.79	19.25	32.90	45.93
12	OCK2	sw	100	non.str	global	3.68	40.22	99.78	17.75	23.5	38.39	50.83
12	OCK2	sw	100	str	global	0.02	46.92	98.28	12.33	18.02	26.46	38.68
12	OCK2	sw	100	non.str	local	0.05	44.14	100	14.56	19.94	31.49	43.13
12	OCK2	sw	100	str	local	0.04	47.26	98.28	11.48	17.42	24.64	37.39
13	OCK3	n	20	non.str	global	2.29	28.41	77.32	14.3	19.28	44.41	59.88
13	OCK3	n	20	str	global	2.29	28.39	77.38	14.26	19.21	44.29	59.66
13	OCK3	n	20	non.str	local	1.76	28.39	77.06	14.3	19.22	44.41	59.69
13	OCK3	n	20	str	local	1.74	28.35	77.12	14.27	19.15	44.32	59.47
13	OCK3	n	40	non.str	global	2.8	30.55	74.21	14.38	19.81	42.13	58.04
13	OCK3	n	40	str	global	2.76	30.6	74.47	14.27	19.73	41.81	57.81
13	OCK3	n	40	non.str	local	2.44	30.6	79.77	14.22	19.71	41.66	57.75
13	OCK3	n	40	str	local	2.43	30.65	79.83	14.13	19.65	41.40	57.57
13	OCK3	n	60	non.str	global	2.39	30.13	77.1	13.83	19.31	41.11	57.40
13	OCK3	n	60	str	global	2.38	30.16	77.29	13.77	19.23	40.93	57.16
13	OCK3	n	60	non.str	local	2.32	30.46	85.74	13.46	18.99	40.01	56.45
13	OCK3	n	60	str	local	2.29	30.51	85.75	13.44	18.94	39.95	56.30
13	OCK3	n	80	non.str	global	2.49	30.8	82.4	13.34	18.78	39.19	55.17
13	OCK3	n	80	str	global	2.48	30.83	82.57	13.27	18.68	38.98	54.88
13	OCK3	n	80	non.str	local	2.04	31.14	89.36	13.02	18.6	38.25	54.64
13	OCK3	n	80	str	local	2.03	31.22	89.39	12.98	18.48	38.13	54.29
13	OCK3	n	100	non.str	global	2.44	30.66	82.73	13.37	18.84	39.28	55.35
13	OCK3	n	100	str	global	2.42	30.67	82.91	13.37	18.84	39.28	55.35
13	OCK3	n	100	non.str	local	1.5	31.15	90.35	13	18.54	38.19	54.47
13	OCK3	n	100	str	local	1.47	31.12	90.38	13.02	18.55	38.25	54.49
13	OCK3	ne	20	non.str	global	5.02	21.73	56.91	12.73	17.34	50.26	68.46
13	OCK3	ne	20	str	global	4.81	21.93	55.38	12.56	16.97	49.98	67.53
13	OCK3	ne	20	non.str	local	3.02	22.28	62.21	12.51	16.94	49.39	66.88
13	OCK3	ne	20	str	local	3	22.47	55.94	12.41	16.79	49.38	66.81
13	OCK3	ne	40	non.str	global	3.22	21.92	75.57	12.45	16.94	48.96	66.61
13	OCK3	ne	40	str	global	2.89	22.09	96.52	12.2	16.74	48.15	66.06
13	OCK3	ne	40	non.str	local	1.59	22.12	87.79	12.15	16.69	47.78	65.63
13	OCK3	ne	40	str	local	1.96	22.32	96.52	12.03	16.64	47.47	65.67
13	OCK3	ne	60	non.str	global	3.5	22.12	71.36	11.84	16.4	46.65	64.62
13	OCK3	ne	60	str	global	3.67	22.23	79.88	11.84	16.42	46.78	64.88
13	OCK3	ne	60	non.str	local	2.08	22.53	87.47	11.54	16.08	45.47	63.36
13	OCK3	ne	60	str	local	2.22	22.58	79.86	11.62	16.26	45.91	64.24
13	OCK3	ne	80	non.str	global	2.4	21.99	80.09	11.49	16.04	45.61	63.68
13	OCK3	ne	80	str	global	2.89	21.96	85.01	11.4	15.99	45.44	63.73
13	OCK3	ne	80	non.str	local	1.38	22.29	93.23	11.06	15.65	43.91	62.13
13	OCK3	ne	80	str	local	1.51	22.23	85.43	11.07	15.69	44.12	62.53
13	OCK3	ne	100	non.str	global	2.35	21.95	72.11	11.24	15.71	44.80	62.61
13	OCK3	ne	100	str	global	2.37	21.96	88.14	11.1	15.49	44.36	61.91
13	OCK3	ne	100	non.str	local	1.04	22.21	82.81	10.73	15.29	42.77	60.94
13	OCK3	ne	100	str	local	1.34	22.26	94.87	10.73	15.23	42.89	60.87
13	OCK3	sw	20	non.str	global	0.09	44.05	94.42	20.09	27.87	41.64	57.76
13	OCK3	sw	20	str	global	0.01	37.77	86.86	18.53	24.02	46.76	60.61
13	OCK3	sw	20	non.str	local	0.07	45.14	96.04	19.91	28.36	41.26	58.78
13	OCK3	sw	20	str	local	0.01	37.77	86.86	18.53	24.02	46.76	60.61
13	OCK3	sw	40	non.str	global	0.07	46.1	94.87	13.45	18.38	27.97	38.22
13	OCK3	sw	40	str	global	0	45.78	94.45	13.23	19.34	27.64	40.41
13	OCK3	sw	40	non.str	local	0.04	46.09	95.43	13.55	18.23	28.18	37.91
13	OCK3	sw	40	str	local	0.01	45.32	94.12	13.27	19.35	27.73	40.43
13	OCK3	sw	60	non.str	global	0.06	45.66	96.2	13.63	19.46	28.56	40.77
13	OCK3	sw	60	str	global	0.03	45.28	95.95	13.43	20.44	29.23	44.48
13	OCK3	sw	60	non.str	local	0.01	45.46	93.69	13.54	19.32	28.37	40.48
13	OCK3	sw	60	str	local	0	44.95	95.95	13.25	20.46	28.84	44.53
13	OCK3	sw	80	non.str	global	0.35	42.39	96.31	16.12	22.34	35.30	48.93
13	OCK3	sw	80	str	global	0	40.23	92.62	12.06	17.56	28.78	41.90
13	OCK3	sw	80	non.str	local	0	44.11	94.68	14.97	20.42	32.79	44.72
13	OCK3	sw	80	str	local	0.01	40.9	92.63	11.89	17.26	28.37	41.18
13	OCK3	sw	100	non.str	global	0.11	43.3	96.18	15.13	21.86	32.73	47.29
13	OCK3	sw	100	str	global	0	45.99	92.3	10.73	16.55	23.03	35.52
13	OCK3	sw	100	non.str	local	0	45.7	95.28	13.71	19.23	29.66	41.60
13	OCK3	sw	100	str	local	0	46.28	91.96	10.39	16.13	22.30	34.62
14	OCK4	n	20	non.str	global	2.2	28.55	75.52	14.44	19.43	44.84	60.34
14	OCK4	n	20	str	global	2.22	28.51	76.18	14.41	19.36	44.75	60.12
14	OCK4	n	20	non.str	local	2.15	28.2	75.62	14.46	19.4	44.91	60.25
14	OCK4	n	20	str	local	2.12	28.14	75.71	14.44	19.35	44.84	60.09
14	OCK4	n	40	non.str	global	3.4	30.58	73.67	14.35	19.78	42.05	57.95
14	OCK4	n	40	str	global	3.38	30.62	73.89	14.25	19.71	41.75	57.75
14	OCK4	n	40	non.str	local	2.61	30.49	79.55	14.18	19.67	41.55	57.63
14	OCK4	n	40	str	local	2.6	30.53	79.61	14.09	19.62	41.28	57.49
14	OCK4	n	60	non.str	global	2.15	30.15	76.98	13.85	19.36	41.17	57.55
14	OCK4	n	60	str	global	2.14	30.15	77.18	13.78	19.3	40.96	57.37
14	OCK4	n	60	non.str	local	1.94	30.38	85.71	13.49	19.02	40.10	56.54
14	OCK4	n	60	str	local	1.91	30.43	85.73	13.47	18.99	40.04	56.45

Appendix

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
14	OCK4	n	80	non.str	global	2.4	30.79	82.02	13.4	18.86	39.37	55.41
14	OCK4	n	80	str	global	2.38	30.81	82.18	13.31	18.74	39.10	55.05
14	OCK4	n	80	non.str	local	1.87	31.05	89.22	13.06	18.65	38.37	54.79
14	OCK4	n	80	str	local	1.86	31.12	89.23	13.01	18.52	38.22	54.41
14	OCK4	n	100	non.str	global	2.41	30.64	82.32	13.42	18.93	39.42	55.61
14	OCK4	n	100	str	global	2.41	30.66	82.5	13.41	18.94	39.39	55.64
14	OCK4	n	100	non.str	local	1.37	31.06	90.16	13.03	18.58	38.28	54.58
14	OCK4	n	100	str	local	1.34	31.03	90.19	13.05	18.6	38.34	54.64
14	OCK4	ne	20	non.str	global	5.62	21.72	54.54	12.72	17.31	50.22	68.34
14	OCK4	ne	20	str	global	5.07	21.91	52.39	12.59	17.01	50.10	67.69
14	OCK4	ne	20	non.str	local	3.26	22.31	65.56	12.5	16.92	49.35	66.80
14	OCK4	ne	20	str	local	3.03	22.46	54.68	12.44	16.84	49.50	67.01
14	OCK4	ne	40	non.str	global	3.52	21.93	78.25	12.43	16.92	48.88	66.54
14	OCK4	ne	40	str	global	3.22	22.06	85.4	12.22	16.76	48.22	66.14
14	OCK4	ne	40	non.str	local	1.59	22.13	89.27	12.13	16.66	47.70	65.51
14	OCK4	ne	40	str	local	1.88	22.3	85.4	12.04	16.63	47.51	65.63
14	OCK4	ne	60	non.str	global	3.77	22.12	73.54	11.85	16.38	46.69	64.54
14	OCK4	ne	60	str	global	3.85	22.23	78.45	11.83	16.35	46.74	64.60
14	OCK4	ne	60	non.str	local	2.36	22.55	90.14	11.53	16.05	45.43	63.24
14	OCK4	ne	60	str	local	2.42	22.6	78.56	11.6	16.18	45.83	63.93
14	OCK4	ne	80	non.str	global	2.81	21.99	83.41	11.5	16.06	45.65	63.76
14	OCK4	ne	80	str	global	3.15	21.97	79.45	11.43	16.03	45.56	63.89
14	OCK4	ne	80	non.str	local	1.4	22.32	94.59	11.05	15.65	43.87	62.13
14	OCK4	ne	80	str	local	1.47	22.24	80.18	11.08	15.71	44.16	62.61
14	OCK4	ne	100	non.str	global	2.79	21.95	75.07	11.24	15.71	44.80	62.61
14	OCK4	ne	100	str	global	2.71	21.94	100	11.15	15.59	44.56	62.31
14	OCK4	ne	100	non.str	local	1.18	22.24	85.27	10.7	15.26	42.65	60.82
14	OCK4	ne	100	str	local	1.45	22.28	100	10.76	15.32	43.01	61.23
14	OCK4	sw	20	non.str	global	0.79	44.52	99.04	18.49	27.35	38.32	56.68
14	OCK4	sw	20	str	global	0.01	37.5	88.45	17.64	23.45	44.51	59.17
14	OCK4	sw	20	non.str	local	2.1	46.3	99.64	18.67	27.68	38.69	57.37
14	OCK4	sw	20	str	local	0.01	37.5	88.45	17.64	23.45	44.51	59.17
14	OCK4	sw	40	non.str	global	0.24	45.13	99.32	13.81	19.1	28.72	39.72
14	OCK4	sw	40	str	global	0.03	46.11	96.8	12.92	18.95	27.00	39.59
14	OCK4	sw	40	non.str	local	0.19	45.89	99.67	13.52	18.53	28.11	38.53
14	OCK4	sw	40	str	local	0.04	45.85	96.27	12.99	18.92	27.14	39.53
14	OCK4	sw	60	non.str	global	0.95	44.61	99.69	14.09	19.99	29.52	41.88
14	OCK4	sw	60	str	global	0	45.21	95.95	13.42	20.54	29.21	44.70
14	OCK4	sw	60	non.str	local	0.09	45.39	99.55	13.64	19.3	28.58	40.44
14	OCK4	sw	60	str	local	0	44.81	95.95	13.29	20.57	28.92	44.77
14	OCK4	sw	80	non.str	global	1.26	41.62	99.97	16.36	22.8	35.83	49.93
14	OCK4	sw	80	str	global	0	40.08	92.19	11.79	17.29	28.13	41.26
14	OCK4	sw	80	non.str	local	0.02	44.02	99.99	15	20.6	32.85	45.12
14	OCK4	sw	80	str	local	0	40.67	92.15	11.64	17	27.77	40.56
14	OCK4	sw	100	non.str	global	0.58	42.34	99.47	15.35	22.32	33.20	48.28
14	OCK4	sw	100	str	global	0	46.32	98.55	11	16.65	23.61	35.74
14	OCK4	sw	100	non.str	local	0	45.4	98.9	13.89	19.39	30.05	41.94
14	OCK4	sw	100	str	local	0	46.72	98.55	10.6	16.12	22.75	34.60
15	OK	n	20	non.str	global	1.72	28.62	80.86	13.5	18.28	41.93	56.77
15	OK	n	20	str	global	1.72	28.59	80.86	13.49	18.26	41.89	56.71
15	OK	n	20	non.str	local	1.94	28.66	80.27	13.58	18.27	42.17	56.74
15	OK	n	20	str	local	1.94	28.6	80.27	13.58	18.25	42.17	56.68
15	OK	n	40	non.str	global	2.25	30.83	75.63	13.89	19.32	40.70	56.61
15	OK	n	40	str	global	2.26	30.84	75.63	13.89	19.33	40.70	56.64
15	OK	n	40	non.str	local	2.01	30.9	80.76	13.87	19.43	40.64	56.93
15	OK	n	40	str	local	2.01	30.9	80.76	13.88	19.45	40.67	56.99
15	OK	n	60	non.str	global	1.74	30.44	80.28	13.34	18.88	39.66	56.12
15	OK	n	60	str	global	1.74	30.45	80.28	13.37	18.89	39.74	56.15
15	OK	n	60	non.str	local	1.65	30.64	86.46	13.16	18.77	39.12	55.80
15	OK	n	60	str	local	1.65	30.65	86.46	13.17	18.78	39.15	55.83
15	OK	n	80	non.str	global	1.61	31.13	87.3	12.8	18.35	37.60	53.91
15	OK	n	80	str	global	1.61	31.13	87.3	12.78	18.29	37.54	53.73
15	OK	n	80	non.str	local	1.05	31.31	89.99	12.7	18.36	37.31	53.94
15	OK	n	80	str	local	1.05	31.33	89.99	12.69	18.31	37.28	53.79
15	OK	n	100	non.str	global	1.54	30.96	87.54	12.87	18.38	37.81	54.00
15	OK	n	100	str	global	1.54	30.96	87.54	12.86	18.33	37.78	53.85
15	OK	n	100	non.str	local	1.13	31.27	91.14	12.69	18.28	37.28	53.70
15	OK	n	100	str	local	1.13	31.27	91.14	12.67	18.23	37.22	53.55
15	OK	ne	20	non.str	global	0.67	23.25	68.85	12.44	18.02	49.11	71.14
15	OK	ne	20	str	global	0.68	23.45	68.65	12.21	17.39	48.59	69.20
15	OK	ne	20	non.str	local	0.69	23.25	70.43	12.37	17.92	48.84	70.75
15	OK	ne	20	str	local	0.69	23.64	70.43	12.19	17.4	48.51	69.24
15	OK	ne	40	non.str	global	0.71	22.54	68.73	11.65	16.27	45.81	63.98
15	OK	ne	40	str	global	0.68	22.76	83.16	11.29	15.65	44.55	61.76
15	OK	ne	40	non.str	local	0.55	22.51	73.19	11.52	16.13	45.30	63.43
15	OK	ne	40	str	local	0.55	22.95	83.16	11.25	15.57	44.40	61.44
15	OK	ne	60	non.str	global	1.37	22.72	67.56	10.66	15.24	42.00	60.05
15	OK	ne	60	str	global	1.35	22.86	78.53	10.49	14.79	41.45	58.44
15	OK	ne	60	non.str	local	1.21	22.83	74.45	10.55	15.09	41.57	59.46
15	OK	ne	60	str	local	1.21	23.07	77.59	10.41	14.76	41.13	58.32

Appendix

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
15	OK	ne	80	non.str	global	1.1	22.5	70.37	10.42	15.01	41.37	59.59
15	OK	ne	80	str	global	1.09	22.74	99.49	10.19	14.48	40.45	57.48
15	OK	ne	80	non.str	local	0.99	22.56	77.75	10.26	14.75	40.73	58.55
15	OK	ne	80	str	local	0.99	22.87	99.49	10.1	14.39	40.10	57.13
15	OK	ne	100	non.str	global	0.34	22.52	70.8	9.98	14.48	39.78	57.71
15	OK	ne	100	str	global	0.34	22.75	99.49	9.76	13.93	38.90	55.52
15	OK	ne	100	non.str	local	0.21	22.55	74.94	9.84	14.28	39.22	56.92
15	OK	ne	100	str	local	0.21	22.84	99.49	9.69	13.91	38.62	55.44
15	OK	sw	20	non.str	global	0.3	46.87	86.27	20.67	25.66	42.84	53.18
15	OK	sw	20	str	global	0.1	48.36	94.4	17.14	22.7	35.52	47.05
15	OK	sw	20	non.str	local	0.26	46.23	86.24	20.07	25.18	41.60	52.19
15	OK	sw	20	str	local	0.1	48.36	94.4	17.14	22.7	35.52	47.05
15	OK	sw	40	non.str	global	0	46.68	96.26	10.98	15.61	22.83	32.46
15	OK	sw	40	str	global	0.01	46.47	96.35	12.11	18.06	25.30	37.74
15	OK	sw	40	non.str	local	0.02	46.23	96.61	10.45	14.67	21.73	30.51
15	OK	sw	40	str	local	0.01	46.76	96.61	11.7	17.55	24.45	36.67
15	OK	sw	60	non.str	global	0	47.47	95.19	10.89	16.72	22.82	35.03
15	OK	sw	60	str	global	0	47.39	96.22	10.87	17.19	22.77	36.02
15	OK	sw	60	non.str	local	0	47.24	94.12	10.72	16.54	22.46	34.65
15	OK	sw	60	str	local	0.01	47.52	96.22	10.65	17.02	22.31	35.66
15	OK	sw	80	non.str	global	0	44.05	92.35	11.81	16.96	25.87	37.14
15	OK	sw	80	str	global	0.04	44.19	94.4	11.6	16.95	25.41	37.12
15	OK	sw	80	non.str	local	0.09	43.25	90.2	11.35	16.81	24.86	36.82
15	OK	sw	80	str	local	0.09	44.44	94.4	11.35	16.65	24.86	36.47
15	OK	sw	100	non.str	global	0	45.04	94.01	10.31	15.68	22.30	33.92
15	OK	sw	100	str	global	0.01	45.26	93.77	9.83	15.28	21.26	33.05
15	OK	sw	100	non.str	local	0.03	44.59	91.01	10.17	15.72	22.00	34.00
15	OK	sw	100	str	local	0.06	45.44	93.77	9.63	15.06	20.83	32.58
16	UK	n	20	non.str	global	1.95	28.44	77.13	14.31	19.13	44.44	59.41
16	UK	n	20	str	global	1.92	29.01	100	14.52	19.86	45.09	61.68
16	UK	n	20	non.str	local	0.01	33.1	100	15.8	22.91	49.07	71.15
16	UK	n	20	str	local	0.01	33.4	100	16.19	23.65	50.28	73.45
16	UK	n	40	non.str	global	1.72	30.76	79.68	13.99	19.23	40.99	56.34
16	UK	n	40	str	global	0.98	31	100	14.13	19.62	41.40	57.49
16	UK	n	40	non.str	local	0.04	35.96	100	18.08	27.48	52.97	80.52
16	UK	n	40	str	local	0.03	35.34	100	17.79	27.05	52.12	79.26
16	UK	n	60	non.str	global	0.48	30.19	77.97	13.71	19.2	40.76	57.07
16	UK	n	60	str	global	0.35	30.25	100	13.9	19.66	41.32	58.44
16	UK	n	60	non.str	local	0	34.18	100	16.21	24.55	48.19	72.98
16	UK	n	60	str	local	0	34.11	100	16.29	24.8	48.42	73.72
16	UK	n	80	non.str	global	2.47	30.89	85.22	13.21	18.66	38.81	54.82
16	UK	n	80	str	global	0.95	31.03	100	13.22	18.78	38.84	55.17
16	UK	n	80	non.str	local	0	34.24	100	15.37	23.31	45.15	68.48
16	UK	n	80	str	local	0	34.45	100	15.29	23.31	44.92	68.48
16	UK	n	100	non.str	global	2.29	30.74	84.87	13.21	18.68	38.81	54.88
16	UK	n	100	str	global	0.24	30.82	100	13.2	18.72	38.78	54.99
16	UK	n	100	non.str	local	0.01	34.06	100	14.94	22.45	43.89	65.95
16	UK	n	100	str	local	0.01	34.32	100	15.18	22.88	44.59	67.22
16	UK	ne	20	non.str	global	1.3	22.67	100	11.93	16.77	47.10	66.21
16	UK	ne	20	str	global	0.06	22.92	100	12.05	17.1	47.95	68.05
16	UK	ne	20	non.str	local	0.01	28.06	100	16.34	24.92	64.51	98.38
16	UK	ne	20	str	local	0.01	27.5	100	16.31	24.87	64.90	98.97
16	UK	ne	40	non.str	global	1.16	22.56	100	11.44	15.78	44.99	62.05
16	UK	ne	40	str	global	1.15	22.71	100	11.47	16.04	45.26	63.30
16	UK	ne	40	non.str	local	0	26.9	100	14.53	22.16	57.14	87.14
16	UK	ne	40	str	local	0	26.47	100	14.23	21.49	56.16	84.81
16	UK	ne	60	non.str	global	1.59	22.67	100	10.76	15.2	42.40	59.89
16	UK	ne	60	str	global	0.91	22.82	100	10.76	15.35	42.51	60.65
16	UK	ne	60	non.str	local	0	26.33	100	13.46	20.94	53.03	82.51
16	UK	ne	60	str	local	0	26.03	100	13.49	21.08	53.30	83.29
16	UK	ne	80	non.str	global	1.12	22.54	100	10.4	14.88	41.29	59.07
16	UK	ne	80	str	global	1.11	22.79	100	10.29	14.81	40.85	58.79
16	UK	ne	80	non.str	local	0	26.04	100	13.31	20.96	52.84	83.21
16	UK	ne	80	str	local	0	25.8	100	13.02	20.32	51.69	80.67
16	UK	ne	100	non.str	global	0.79	22.52	100	10.05	14.4	40.06	57.39
16	UK	ne	100	str	global	0.77	22.72	100	9.93	14.3	39.58	56.99
16	UK	ne	100	non.str	local	0	25.76	100	11.95	19.08	47.63	76.05
16	UK	ne	100	str	local	0	25.59	100	12.16	19.32	48.47	77.00
16	UK	sw	20	non.str	global	0	41.54	97.37	23.79	32.85	49.31	68.08
16	UK	sw	20	str	global	0.01	57.91	99.92	35.74	47.31	74.07	98.05
16	UK	sw	20	non.str	local	0.05	44.32	99.9	22.63	32.07	46.90	66.47
16	UK	sw	20	str	local	0.01	57.91	99.92	35.74	47.31	74.07	98.05
16	UK	sw	40	non.str	global	0.01	46.37	99.59	14.61	19.9	30.38	41.38
16	UK	sw	40	str	global	0.06	49.31	99.95	16.81	26.35	35.12	55.06
16	UK	sw	40	non.str	local	0.03	45.48	98.93	17.8	26.48	37.01	55.06
16	UK	sw	40	str	local	0	47.2	99.84	16.83	25.68	35.17	53.66
16	UK	sw	60	non.str	global	0	46.78	99.91	12.86	19.6	26.94	41.06
16	UK	sw	60	str	global	0	47.55	100	16.45	25.84	34.46	54.14
16	UK	sw	60	non.str	local	0	45.56	99.98	15.71	24.23	32.91	50.76
16	UK	sw	60	str	local	0	43.37	100	18.5	28.46	38.76	59.63

Appendix

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
16	UK	sw	80	non.str	global	0.23	42.22	98.59	14.63	21.02	32.04	46.04
16	UK	sw	80	str	global	0	46.55	96.51	17.59	26.1	38.52	57.16
16	UK	sw	80	non.str	local	0.01	41.16	99.99	15.92	24.58	34.87	53.83
16	UK	sw	80	str	local	0	44.07	99.99	17.22	26.92	37.71	58.96
16	UK	sw	100	non.str	global	0.01	43.99	99.99	11.58	17.53	25.05	37.92
16	UK	sw	100	str	global	0	44	96.87	13.1	21.34	28.34	46.16
16	UK	sw	100	non.str	local	0	44.23	99.98	15.97	24.8	34.54	53.64
16	UK	sw	100	str	local	0	42.33	99.98	20.05	31.24	43.37	67.58
17	RT	n	20	str	global	4.3	32.31	93.6	18.08	23.81	56.15	73.94
17	RT	n	40	str	global	3.14	34.17	97.05	16.72	22.59	48.99	66.19
17	RT	n	60	str	global	1.26	33.9	97.38	15.73	21.18	46.76	62.96
17	RT	n	80	str	global	2.96	34.06	87.93	15.72	21.1	46.18	61.99
17	RT	n	100	str	global	0.66	33.88	92.52	15.39	20.73	45.21	60.90
17	RT	ne	20	str	global	1.32	25.06	90.29	15.4	19.81	60.80	78.21
17	RT	ne	40	str	global	4.09	25.23	75.83	13.91	18.5	54.70	72.75
17	RT	ne	60	str	global	3.5	25.84	76.29	13.5	17.88	53.19	70.45
17	RT	ne	80	str	global	3.36	25.24	74.81	13	17.31	51.61	68.72
17	RT	ne	100	str	global	2.13	25.1	87	12.48	16.91	49.74	67.40
17	RT	sw	20	str	global	1.3	50.97	86.49	19.7	28.38	40.83	58.82
17	RT	sw	40	str	global	0.73	46.56	83.63	10.32	17.02	21.46	35.39
17	RT	sw	60	str	global	1.57	48.96	84.51	13.21	21.42	27.68	44.88
17	RT	sw	80	str	global	3.3	46.7	85.04	11.61	17.8	25.43	38.98
17	RT	sw	100	str	global	3.3	47.69	86.9	10.37	15.98	22.43	34.57
18	TPSr	n	20	non.str	local	0	32.37	100	16.71	24.37	51.89	75.68
18	TPSr	n	40	non.str	local	0	36.3	100	18.57	27.24	54.41	79.81
18	TPSr	n	60	non.str	local	0	36.05	100	18.96	28.49	56.36	84.69
18	TPSr	n	80	non.str	local	0	35.09	100	17.49	26.53	51.38	77.94
18	TPSr	n	100	non.str	local	0	35.16	100	17.14	25.64	50.35	75.32
18	TPSr	ne	20	non.str	local	0	28.39	100	17.24	24.94	68.06	98.46
18	TPSr	ne	40	non.str	local	0	26.68	100	14.38	21.28	56.55	83.68
18	TPSr	ne	60	non.str	local	0	27.49	100	14.11	21.48	55.59	84.63
18	TPSr	ne	80	non.str	local	0	25.92	100	13.89	21.97	55.14	87.22
18	TPSr	ne	100	non.str	local	0	29.08	100	15.34	24.55	61.14	97.85
18	TPSr	sw	20	non.str	local	0	52.54	100	29.39	38.86	60.91	80.54
18	TPSr	sw	40	non.str	local	0	44.67	100	16.45	25.23	34.21	52.46
18	TPSr	sw	60	non.str	local	0	45.49	100	17.47	27.38	36.6	57.36
18	TPSr	sw	80	non.str	local	0	43.06	100	16.51	25.71	36.16	56.31
18	TPSr	sw	100	non.str	local	0	45.72	100	15.06	25.02	32.58	54.12
19	TPSt	n	20	non.str	local	0	32.39	100	13.75	18.92	42.7	58.76
19	TPSt	n	40	non.str	local	0	34.94	100	14.74	21.85	43.19	64.02
19	TPSt	n	60	non.str	local	0	34.77	100	15.41	23.2	45.81	68.97
19	TPSt	n	80	non.str	local	0	34.26	100	14.64	22.13	43.01	65.01
19	TPSt	n	100	non.str	local	0	34.45	100	14.39	21.64	42.27	63.57
19	TPSt	ne	20	non.str	local	0	25.4	100	13.86	19.64	54.72	77.54
19	TPSt	ne	40	non.str	local	0	25.96	100	11.96	17.49	47.03	68.78
19	TPSt	ne	60	non.str	local	0	25.49	100	11.08	16.78	43.66	66.12
19	TPSt	ne	80	non.str	local	0	24.99	100	10.96	16.64	43.51	66.06
19	TPSt	ne	100	non.str	local	0	26.04	100	10.88	16.5	43.36	65.76
19	TPSt	sw	20	non.str	local	0	47.89	100	19.82	27.17	41.08	56.31
19	TPSt	sw	40	non.str	local	0	47.16	100	12.7	19.25	26.41	40.03
19	TPSt	sw	60	non.str	local	0	47.56	100	14.11	24.15	29.56	50.6
19	TPSt	sw	80	non.str	local	0	45.45	100	12.18	20.28	26.68	44.42
19	TPSt	sw	100	non.str	local	0	47.05	100	10.65	18.17	23.04	39.3
20	GRNN	n	20	str	global	0	31.87	97.38	14.56	20.84	45.22	64.72
20	GRNN	n	40	str	global	0	34.73	97.34	14.43	20.43	42.28	59.86
20	GRNN	n	60	str	global	0	33.77	97.72	14.56	20.76	43.28	61.71
20	GRNN	n	80	str	global	0	34.53	97.72	14.35	20.34	42.16	59.75
20	GRNN	n	100	str	global	0	34.29	97.67	14.07	20.34	41.33	59.75
20	GRNN	ne	20	str	global	0	24.99	71	15.08	19.75	59.53	77.97
20	GRNN	ne	40	str	global	0	24.31	94.9	15.81	20.59	62.17	80.97
20	GRNN	ne	60	str	global	0	24.79	94.9	11.95	17.36	47.21	68.59
20	GRNN	ne	80	str	global	0	24.51	94.9	11.4	17.26	45.26	68.52
20	GRNN	ne	100	str	global	0	23.77	97.29	11.83	17.84	47.15	71.10
20	GRNN	sw	20	str	global	0	47.86	94.45	22.3	31.76	46.22	65.82
20	GRNN	sw	40	str	global	0	43.24	90.58	18.71	29.32	38.91	60.97
20	GRNN	sw	60	str	global	0	47.65	96.36	13.86	23.52	29.04	49.28
20	GRNN	sw	80	str	global	0	43.96	94.4	15.45	24.15	33.84	52.89
20	GRNN	sw	100	str	global	0	46.27	94.37	14.96	24.62	32.36	53.26
21	SVM	n	20	str	global	0	32.57	96.7	15.81	21.58	49.1	67.02
21	SVM	n	40	str	global	0	34.55	99.33	14.68	20.42	43.01	59.83
21	SVM	n	60	str	global	0	33.43	95.64	14.14	20.12	42.03	59.81
21	SVM	n	80	str	global	0	34.28	95.68	13.4	18.95	39.37	55.67
21	SVM	n	100	str	global	0	33.85	99.06	13.36	19.12	39.25	56.17
21	SVM	ne	20	str	global	0	23.01	68.13	13.09	17.53	52.09	69.76
21	SVM	ne	40	str	global	0	24.84	85.71	13.05	17.7	51.5	69.85
21	SVM	ne	60	str	global	0	24.61	87.06	11.71	16.18	46.27	63.93
21	SVM	ne	80	str	global	0	23.9	87.72	11.84	16.38	47	65.03
21	SVM	ne	100	str	global	0	24.13	90.68	10.93	15.73	43.56	62.69
21	SVM	sw	20	str	global	0	47.8	100	20.07	28.18	41.6	58.4
21	SVM	sw	40	str	global	0	48.37	99.27	13.68	18.73	28.58	39.13

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
21	SVM	sw	60	str	global	0	48.74	100	13.57	19.66	28.43	41.19
21	SVM	sw	80	str	global	0	46.45	98.57	15.75	21.38	34.49	46.82
21	SVM	sw	100	str	global	0	46.94	100	12.78	18.95	27.64	40.99
22	RKglm1	n	20	non.str	global	3.77	32.3	83.93	13.49	17.7	41.89	54.97
22	RKglm1	n	20	str	global	0	32.17	100	13.86	18.08	43.04	56.15
22	RKglm1	n	20	non.str	local	3.49	32.2	83.53	13.48	17.67	41.86	54.88
22	RKglm1	n	20	str	local	0	32.02	100	13.92	18.1	43.23	56.21
22	RKglm1	n	40	non.str	global	1.72	34.14	77.63	14.01	19.02	41.05	55.73
22	RKglm1	n	40	str	global	0	34.2	100	14.56	19.92	42.66	58.37
22	RKglm1	n	40	non.str	local	1.36	34.13	82.36	13.93	19.08	40.81	55.90
22	RKglm1	n	40	str	local	0	34.18	100	14.45	19.98	42.34	58.54
22	RKglm1	n	60	non.str	global	2.02	33.6	81.2	13.42	18.49	39.89	54.96
22	RKglm1	n	60	str	global	0	33.61	100	14.11	19.54	41.94	58.09
22	RKglm1	n	60	non.str	local	2.71	33.7	86.79	13.21	18.4	39.27	54.70
22	RKglm1	n	60	str	local	0	33.7	100	13.96	19.47	41.50	57.88
22	RKglm1	n	80	non.str	global	2.57	34.05	87.91	12.79	17.87	37.57	52.50
22	RKglm1	n	80	str	global	0	34.07	100	13.57	18.82	39.86	55.29
22	RKglm1	n	80	non.str	local	2.26	34.13	90.64	12.69	17.87	37.28	52.50
22	RKglm1	n	80	str	local	0	34.14	100	13.47	18.83	39.57	55.32
22	RKglm1	n	100	non.str	global	0	33.94	87.89	12.89	17.97	37.87	52.79
22	RKglm1	n	100	str	global	0	33.97	100	13.71	19.07	40.28	56.02
22	RKglm1	n	100	non.str	local	0	34.1	91.65	12.73	17.92	37.40	52.64
22	RKglm1	n	100	str	local	0	34.13	100	13.57	19.04	39.86	55.93
22	RKglm1	ne	20	non.str	global	3.48	25.08	88.77	11.84	15.9	47.06	63.20
22	RKglm1	ne	20	str	global	3.52	25.06	94.23	12.44	17.2	49.44	68.36
22	RKglm1	ne	20	non.str	local	2.75	25.18	95.65	11.84	15.96	47.06	63.43
22	RKglm1	ne	20	str	local	2.75	25.13	96.15	12.44	17.28	49.44	68.68
22	RKglm1	ne	40	non.str	global	2.57	25.31	89.06	11.97	15.84	47.50	62.86
22	RKglm1	ne	40	str	global	2.49	25.32	100	12.85	17.59	50.99	69.80
22	RKglm1	ne	40	non.str	local	2.63	25.24	94.26	11.81	15.79	46.87	62.66
22	RKglm1	ne	40	str	local	2.63	25.27	100	12.73	17.56	50.52	69.68
22	RKglm1	ne	60	non.str	global	4.74	25.24	94.75	11.42	15.3	45.64	61.15
22	RKglm1	ne	60	str	global	2.76	25.21	100	12.08	16.95	48.28	67.75
22	RKglm1	ne	60	non.str	local	4.46	25.29	99.65	11.27	15.28	45.04	61.07
22	RKglm1	ne	60	str	local	1.93	25.23	100	11.94	16.91	47.72	67.59
22	RKglm1	ne	80	non.str	global	3.22	25.2	98.68	10.93	14.86	43.39	58.99
22	RKglm1	ne	80	str	global	3.19	25.15	100	11.95	16.99	47.44	67.45
22	RKglm1	ne	80	non.str	local	2.43	25.12	100	10.62	14.7	42.16	58.36
22	RKglm1	ne	80	str	local	2.43	25.06	100	11.66	16.81	46.29	66.73
22	RKglm1	ne	100	non.str	global	3.03	25.14	100	10.68	14.53	42.57	57.91
22	RKglm1	ne	100	str	global	3	25.08	100	11.64	16.61	46.39	66.20
22	RKglm1	ne	100	non.str	local	2.63	25	100	10.33	14.36	41.17	57.23
22	RKglm1	ne	100	str	local	2.63	24.94	100	11.31	16.42	45.08	65.44
22	RKglm1	sw	20	non.str	global	2.42	49.51	98.19	19.82	24.28	41.76	51.16
22	RKglm1	sw	20	str	global	1.38	49.76	100	33.25	41.2	70.06	86.81
22	RKglm1	sw	20	non.str	local	2.11	49.91	100	18.43	22.9	38.83	48.25
22	RKglm1	sw	20	str	local	1.38	49.76	100	33.25	41.2	70.06	86.81
22	RKglm1	sw	40	non.str	global	0.98	46.47	96.93	16.19	19.85	33.44	41.00
22	RKglm1	sw	40	str	global	0.97	47.2	100	26.08	31.08	53.87	64.20
22	RKglm1	sw	40	non.str	local	0	45.56	100	14.05	17.67	29.02	36.50
22	RKglm1	sw	40	str	local	0.75	47.08	100	25.56	30.68	52.80	63.38
22	RKglm1	sw	60	non.str	global	0	49.07	100	15.35	20.68	32.16	43.33
22	RKglm1	sw	60	str	global	0	48.25	100	28.87	35.13	60.49	73.60
22	RKglm1	sw	60	non.str	local	0	48.72	100	13.99	19.27	29.31	40.37
22	RKglm1	sw	60	str	local	0	48.4	100	28.01	34.18	58.68	71.61
22	RKglm1	sw	80	non.str	global	1.71	47.16	98.04	18.43	22.66	39.62	48.71
22	RKglm1	sw	80	str	global	2.1	46.79	100	31.2	38.05	67.07	81.79
22	RKglm1	sw	80	non.str	local	0.1	47.21	100	15.94	20.26	34.26	43.55
22	RKglm1	sw	80	str	local	0	47.97	100	29.53	36.86	63.48	79.23
22	RKglm1	sw	100	non.str	global	0	46.21	100	14.19	18.74	30.69	40.54
22	RKglm1	sw	100	str	global	0	46.05	100	31.66	38.53	68.48	83.34
22	RKglm1	sw	100	non.str	local	0	46.51	100	13.34	17.85	28.86	38.61
22	RKglm1	sw	100	str	local	0	47.18	100	30.64	37.66	66.28	81.46
23	RKglm2	n	20	non.str	global	0	32.26	86.18	13.85	18.15	43.01	56.37
23	RKglm2	n	20	str	global	0	32.1	100	14.51	19	45.06	59.01
23	RKglm2	n	20	non.str	local	0	32.14	85.31	13.82	18.13	42.92	56.30
23	RKglm2	n	20	str	local	0	31.92	100	14.58	19.03	45.28	59.10
23	RKglm2	n	40	non.str	global	0	34.27	80.79	14.34	19.55	42.02	57.28
23	RKglm2	n	40	str	global	0	34.3	100	15.87	21.3	46.50	62.41
23	RKglm2	n	40	non.str	local	0	34.2	82.47	14.23	19.56	41.69	57.31
23	RKglm2	n	40	str	local	0	34.21	100	15.76	21.34	46.18	62.53
23	RKglm2	n	60	non.str	global	0	33.69	99.05	13.65	18.62	40.58	55.35
23	RKglm2	n	60	str	global	0	33.7	100	14.99	20.57	44.56	61.15
23	RKglm2	n	60	non.str	local	0	33.75	99.25	13.42	18.53	39.89	55.08
23	RKglm2	n	60	str	local	0	33.75	100	14.79	20.5	43.97	60.94
23	RKglm2	n	80	non.str	global	0	34.09	88.33	13.02	17.95	38.25	52.73
23	RKglm2	n	80	str	global	0	34.18	100	14.38	19.72	42.24	57.93
23	RKglm2	n	80	non.str	local	0	34.14	91.57	12.89	17.95	37.87	52.73
23	RKglm2	n	80	str	local	0	34.24	100	14.24	19.71	41.83	57.90
23	RKglm2	n	100	non.str	global	0	34.01	88.41	13.15	18.17	38.63	53.38

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
23	RKglm2	n	100	str	global	0	34.15	100	14.56	20.19	42.77	59.31
23	RKglm2	n	100	non.str	local	0	34.15	92.57	12.97	18.12	38.10	53.23
23	RKglm2	n	100	str	local	0	34.29	100	14.37	20.12	42.22	59.11
23	RKglm2	ne	20	non.str	global	3.94	25.13	96.56	11.87	15.94	47.18	63.35
23	RKglm2	ne	20	str	global	3.73	25.11	98.78	12.5	17.23	49.68	68.48
23	RKglm2	ne	20	non.str	local	2.92	25.16	99.05	11.87	16	47.18	63.59
23	RKglm2	ne	20	str	local	2.92	25.17	100	12.5	17.29	49.68	68.72
23	RKglm2	ne	40	non.str	global	2.55	25.32	93.22	11.87	15.78	47.10	62.62
23	RKglm2	ne	40	str	global	2.47	25.3	100	12.77	17.58	50.67	69.76
23	RKglm2	ne	40	non.str	local	2.84	25.25	96.14	11.76	15.76	46.67	62.54
23	RKglm2	ne	40	str	local	2.84	25.25	100	12.68	17.57	50.32	69.72
23	RKglm2	ne	60	non.str	global	4.07	25.24	98.88	11.37	15.31	45.44	61.19
23	RKglm2	ne	60	str	global	0	25.2	100	12.1	16.99	48.36	67.91
23	RKglm2	ne	60	non.str	local	3.36	25.27	100	11.26	15.27	45.00	61.03
23	RKglm2	ne	60	str	local	0	25.2	100	11.98	16.96	47.88	67.79
23	RKglm2	ne	80	non.str	global	3.21	25.19	100	10.79	14.76	42.83	58.59
23	RKglm2	ne	80	str	global	1.62	25.12	100	11.94	17.01	47.40	67.53
23	RKglm2	ne	80	non.str	local	2.49	25.14	100	10.53	14.61	41.80	58.00
23	RKglm2	ne	80	str	local	0.47	25.04	100	11.66	16.85	46.29	66.89
23	RKglm2	ne	100	non.str	global	3.04	25.14	100	10.57	14.44	42.13	57.55
23	RKglm2	ne	100	str	global	0.53	25.07	100	11.68	16.69	46.55	66.52
23	RKglm2	ne	100	non.str	local	2.4	25.02	100	10.26	14.29	40.89	56.95
23	RKglm2	ne	100	str	local	1.47	24.93	100	11.39	16.52	45.40	65.84
23	RKglm2	sw	20	non.str	global	2.96	49.94	97.63	19.28	24.36	40.62	51.33
23	RKglm2	sw	20	str	global	1.97	49.86	100	33.38	40.91	70.33	86.20
23	RKglm2	sw	20	non.str	local	2.33	49.61	100	17.88	22.9	37.67	48.25
23	RKglm2	sw	20	str	local	1.97	49.86	100	33.38	40.91	70.33	86.20
23	RKglm2	sw	40	non.str	global	1.45	46.52	95.54	16.85	20.7	34.81	42.76
23	RKglm2	sw	40	str	global	0	47.23	100	26.04	31.16	53.79	64.37
23	RKglm2	sw	40	non.str	local	0	45.32	100	14.59	18.22	30.14	37.64
23	RKglm2	sw	40	str	local	0	46.7	100	25.68	31.02	53.05	64.08
23	RKglm2	sw	60	non.str	global	0	48.99	100	17.41	22.44	36.48	47.01
23	RKglm2	sw	60	str	global	0	48.59	100	32.16	39.4	67.38	82.55
23	RKglm2	sw	60	non.str	local	0	48.5	100	14.8	20.2	31.01	42.32
23	RKglm2	sw	60	str	local	0	48.35	100	30.77	37.99	64.47	79.59
23	RKglm2	sw	80	non.str	global	3.22	47.37	100	19.16	23.44	41.19	50.39
23	RKglm2	sw	80	str	global	3.02	47.12	100	32.55	38.82	69.97	83.45
23	RKglm2	sw	80	non.str	local	0	47.57	100	16.51	21.1	35.49	45.36
23	RKglm2	sw	80	str	local	0	48.25	100	30.91	37.71	66.44	81.06
23	RKglm2	sw	100	non.str	global	0	46.23	99.32	15.49	19.93	33.51	43.11
23	RKglm2	sw	100	str	global	0	46.08	100	33.11	40.1	71.62	86.74
23	RKglm2	sw	100	non.str	local	0	46.61	100	13.97	18.63	30.22	40.30
23	RKglm2	sw	100	str	local	0	47.45	100	31.93	39.12	69.07	84.62
24	RKglm3	n	20	non.str	global	0.04	32.3	90.28	14.24	18.79	44.22	58.35
24	RKglm3	n	20	str	global	0	32.22	98.92	14.66	19.04	45.53	59.13
24	RKglm3	n	20	non.str	local	0.33	32.13	91.15	14.26	18.81	44.29	58.42
24	RKglm3	n	20	str	local	0	32	100	14.73	19.11	45.75	59.35
24	RKglm3	n	40	non.str	global	0	34.21	81.32	14.55	19.78	42.63	57.95
24	RKglm3	n	40	str	global	0	34.17	99.6	16.59	21.95	48.61	64.31
24	RKglm3	n	40	non.str	local	0	34.1	82.17	14.37	19.75	42.10	57.87
24	RKglm3	n	40	str	local	0	34.07	100	16.49	22.01	48.32	64.49
24	RKglm3	n	60	non.str	global	0	33.7	92.71	13.82	18.88	41.08	56.12
24	RKglm3	n	60	str	global	0	33.7	100	15.53	21.07	46.17	62.63
24	RKglm3	n	60	non.str	local	0	33.77	93.43	13.57	18.78	40.34	55.83
24	RKglm3	n	60	str	local	0	33.78	100	15.33	20.98	45.57	62.37
24	RKglm3	n	80	non.str	global	0	34.1	89.74	13.16	18.13	38.66	53.26
24	RKglm3	n	80	str	global	0	34.16	100	14.68	19.74	43.13	57.99
24	RKglm3	n	80	non.str	local	0	34.16	91.42	13	18.1	38.19	53.17
24	RKglm3	n	80	str	local	0	34.24	100	14.54	19.73	42.71	57.96
24	RKglm3	n	100	non.str	global	0	33.99	87.33	13.29	18.35	39.04	53.91
24	RKglm3	n	100	str	global	0	34.07	100	15.16	20.5	44.54	60.22
24	RKglm3	n	100	non.str	local	0	34.16	92.33	13.07	18.26	38.40	53.64
24	RKglm3	n	100	str	local	0	34.22	100	14.94	20.42	43.89	59.99
24	RKglm3	ne	20	non.str	global	3.99	25.11	97.8	11.94	16.04	47.46	63.75
24	RKglm3	ne	20	str	global	0	25.15	97.65	12.51	17.15	49.72	68.16
24	RKglm3	ne	20	non.str	local	1.87	25.15	99.96	11.93	16.1	47.42	63.99
24	RKglm3	ne	20	str	local	0	25.21	98.88	12.5	17.21	49.68	68.40
24	RKglm3	ne	40	non.str	global	1.88	25.34	93.29	11.79	15.7	46.79	62.30
24	RKglm3	ne	40	str	global	1.36	25.31	100	13.03	18.01	51.71	71.47
24	RKglm3	ne	40	non.str	local	1.28	25.26	95.94	11.7	15.7	46.43	62.30
24	RKglm3	ne	40	str	local	1.36	25.25	100	12.93	18	51.31	71.43
24	RKglm3	ne	60	non.str	global	2.44	25.24	96.49	11.36	15.29	45.40	61.11
24	RKglm3	ne	60	str	global	0	25.22	100	12.29	17.2	49.12	68.75
24	RKglm3	ne	60	non.str	local	2.14	25.29	100	11.27	15.29	45.04	61.11
24	RKglm3	ne	60	str	local	0	25.22	100	12.18	17.17	48.68	68.63
24	RKglm3	ne	80	non.str	global	3.3	25.19	100	10.73	14.68	42.60	58.28
24	RKglm3	ne	80	str	global	0	25.14	100	12.11	17.29	48.07	68.64
24	RKglm3	ne	80	non.str	local	2.55	25.15	100	10.49	14.57	41.64	57.84
24	RKglm3	ne	80	str	local	0	25.03	100	11.86	17.14	47.08	68.04
24	RKglm3	ne	100	non.str	global	3.07	25.14	100	10.55	14.42	42.05	57.47

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
24	RKglm3	ne	100	str	global	0	25.07	100	11.88	16.95	47.35	67.56
24	RKglm3	ne	100	non.str	local	2.55	25.02	100	10.25	14.27	40.85	56.88
24	RKglm3	ne	100	str	local	0	24.92	100	11.61	16.81	46.27	67.00
24	RKglm3	sw	20	non.str	global	2.65	49.75	97.34	21.22	26.01	44.71	54.80
24	RKglm3	sw	20	str	global	1.95	49.63	100	33.68	41.62	70.97	87.69
24	RKglm3	sw	20	non.str	local	2.02	49.28	98.15	19.66	24.96	41.42	52.59
24	RKglm3	sw	20	str	local	1.95	49.63	100	33.68	41.62	70.97	87.69
24	RKglm3	sw	40	non.str	global	9.23	48.87	95.23	18.72	21.99	38.67	45.42
24	RKglm3	sw	40	str	global	2	48.21	100	33.29	39.63	68.77	81.86
24	RKglm3	sw	40	non.str	local	0	46.93	100	15.6	19.42	32.22	40.12
24	RKglm3	sw	40	str	local	1.08	47.36	100	33.09	39.71	68.35	82.03
24	RKglm3	sw	60	non.str	global	0.3	48.51	97.54	16.56	21.75	34.70	45.57
24	RKglm3	sw	60	str	global	0	48.4	100	34.66	43.06	72.62	90.22
24	RKglm3	sw	60	non.str	local	0	48.31	100	14.29	19.64	29.94	41.15
24	RKglm3	sw	60	str	local	0	48.42	100	33.88	41.71	70.98	87.39
24	RKglm3	sw	80	non.str	global	7.21	46.63	94.9	18.46	22.36	39.68	48.07
24	RKglm3	sw	80	str	global	0	46.9	100	33.66	40.95	72.36	88.03
24	RKglm3	sw	80	non.str	local	0	46.76	100	15.06	19.5	32.37	41.92
24	RKglm3	sw	80	str	local	0	48.11	100	31.98	39.11	68.74	84.07
24	RKglm3	sw	100	non.str	global	1.65	46.48	96.41	15.04	19.2	32.53	41.53
24	RKglm3	sw	100	str	global	0	46.37	100	37.46	45.08	81.03	97.51
24	RKglm3	sw	100	non.str	local	0	46.64	100	12.96	17.59	28.03	38.05
24	RKglm3	sw	100	str	local	0	47.68	100	35.5	43.48	76.79	94.05
25	RKglm4	n	20	non.str	global	0	32.3	89.2	14.26	18.76	44.29	58.26
25	RKglm4	n	20	str	global	0	32.24	95.89	14.55	19	45.19	59.01
25	RKglm4	n	20	non.str	local	0	32.13	89.97	14.26	18.73	44.29	58.17
25	RKglm4	n	20	str	local	0	32.02	97.16	14.62	19.02	45.40	59.07
25	RKglm4	n	40	non.str	global	0	34.2	82.05	14.7	19.97	43.07	58.51
25	RKglm4	n	40	str	global	0	34.18	97.89	16.72	22.08	48.99	64.69
25	RKglm4	n	40	non.str	local	0	34.08	83.4	14.47	19.88	42.40	58.25
25	RKglm4	n	40	str	local	0	34.06	99.23	16.58	22.1	48.58	64.75
25	RKglm4	n	60	non.str	global	0	33.71	96.38	13.92	19.01	41.38	56.51
25	RKglm4	n	60	str	global	0	33.72	100	15.57	21.06	46.28	62.60
25	RKglm4	n	60	non.str	local	0	33.79	97.28	13.65	18.88	40.58	56.12
25	RKglm4	n	60	str	local	0	33.81	100	15.35	20.95	45.63	62.28
25	RKglm4	n	80	non.str	global	0	34.11	91.19	13.24	18.2	38.90	53.47
25	RKglm4	n	80	str	global	0	34.14	100	14.67	19.77	43.10	58.08
25	RKglm4	n	80	non.str	local	0	34.17	92.12	13.06	18.16	38.37	53.35
25	RKglm4	n	80	str	local	0	34.23	100	14.53	19.73	42.69	57.96
25	RKglm4	n	100	non.str	global	0	33.99	86.77	13.38	18.44	39.31	54.17
25	RKglm4	n	100	str	global	0	34.05	100	15.15	20.49	44.51	60.19
25	RKglm4	n	100	non.str	local	0	34.16	92.31	13.13	18.34	38.57	53.88
25	RKglm4	n	100	str	local	0	34.21	100	14.91	20.4	43.80	59.93
25	RKglm4	ne	20	non.str	global	4.04	25.09	99.27	11.91	16.08	47.34	63.91
25	RKglm4	ne	20	str	global	0	25.12	100	12.5	17.16	49.68	68.20
25	RKglm4	ne	20	non.str	local	3.25	25.18	99.89	11.95	16.14	47.50	64.15
25	RKglm4	ne	20	str	local	0	25.21	100	12.55	17.22	49.88	68.44
25	RKglm4	ne	40	non.str	global	0	25.35	93.92	11.73	15.67	46.55	62.18
25	RKglm4	ne	40	str	global	1.3	25.33	100	13.12	18	52.06	71.43
25	RKglm4	ne	40	non.str	local	0.66	25.3	96.24	11.68	15.7	46.35	62.30
25	RKglm4	ne	40	str	local	1.16	25.28	100	13.08	18.01	51.90	71.47
25	RKglm4	ne	60	non.str	global	2.5	25.24	98.47	11.3	15.24	45.16	60.91
25	RKglm4	ne	60	str	global	0.66	25.22	100	12.17	17.04	48.64	68.11
25	RKglm4	ne	60	non.str	local	2.41	25.3	100	11.22	15.23	44.84	60.87
25	RKglm4	ne	60	str	local	0	25.22	100	12.09	17.04	48.32	68.11
25	RKglm4	ne	80	non.str	global	2.68	25.19	100	10.59	14.57	42.04	57.84
25	RKglm4	ne	80	str	global	0	25.12	100	12	17.13	47.64	68.00
25	RKglm4	ne	80	non.str	local	2.31	25.16	100	10.42	14.49	41.37	57.52
25	RKglm4	ne	80	str	local	0	25.04	100	11.8	17.05	46.84	67.69
25	RKglm4	ne	100	non.str	global	2.48	25.14	100	10.48	14.35	41.77	57.19
25	RKglm4	ne	100	str	global	0	25.08	100	11.87	16.88	47.31	67.28
25	RKglm4	ne	100	non.str	local	2.71	25.04	100	10.22	14.23	40.73	56.72
25	RKglm4	ne	100	str	local	0	24.93	100	11.63	16.77	46.35	66.84
25	RKglm4	sw	20	non.str	global	2.42	48.62	98.42	19.26	24.75	40.58	52.15
25	RKglm4	sw	20	str	global	2.42	48.36	100	41.42	49.72	87.27	104.76
25	RKglm4	sw	20	non.str	local	3.86	48.85	98.44	19.15	24.62	40.35	51.88
25	RKglm4	sw	20	str	local	2.42	48.36	100	41.42	49.72	87.27	104.76
25	RKglm4	sw	40	non.str	global	2.48	46.87	97.57	17.38	21.68	35.90	44.78
25	RKglm4	sw	40	str	global	1.7	46.34	100	41.25	48.1	85.21	99.36
25	RKglm4	sw	40	non.str	local	0	45.22	100	15.75	21.15	32.53	43.69
25	RKglm4	sw	40	str	local	1.28	46.43	100	41.8	48.64	86.35	100.48
25	RKglm4	sw	60	non.str	global	0.94	48.55	97.66	15.7	20.39	32.89	42.72
25	RKglm4	sw	60	str	global	0	48.13	100	38.87	47.31	81.44	99.12
25	RKglm4	sw	60	non.str	local	0	47.81	100	14.56	19.95	30.50	41.80
25	RKglm4	sw	60	str	local	0	47.7	100	37.86	46.32	79.32	97.05
25	RKglm4	sw	80	non.str	global	3.15	46.42	99.48	18.88	23.09	40.58	49.63
25	RKglm4	sw	80	str	global	0	46.21	100	39.08	46.96	84.01	100.95
25	RKglm4	sw	80	non.str	local	0	46.78	100	16.75	21.49	36.01	46.20
25	RKglm4	sw	80	str	local	0	47.69	100	38.59	46.7	82.95	100.39
25	RKglm4	sw	100	non.str	global	2.55	46.13	95.75	16.43	20.57	35.54	44.49

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
25	RKglm4	sw	100	str	global	0	46.05	100	42.61	50.56	92.17	109.37
25	RKglm4	sw	100	non.str	local	0	46.62	98.64	14.27	19.26	30.87	41.66
25	RKglm4	sw	100	str	local	0	47.71	100	41.81	49.74	90.44	107.59
26	RKglm5	n	20	non.str	global	2.62	32.29	89.63	14.55	18.98	45.19	58.94
26	RKglm5	n	20	str	global	0	32.35	89.58	15.73	20.08	48.85	62.36
26	RKglm5	n	20	non.str	local	3.19	32.05	88.39	14.51	18.96	45.06	58.88
26	RKglm5	n	20	str	local	0	32.13	87.81	15.71	20.14	48.79	62.55
26	RKglm5	n	40	non.str	global	0	34.16	82.61	15.27	20.39	44.74	59.74
26	RKglm5	n	40	str	global	0	34.21	96	18.77	24.46	55.00	71.67
26	RKglm5	n	40	non.str	local	0	34.12	83.51	14.78	20.07	43.31	58.80
26	RKglm5	n	40	str	local	0	34.22	96.86	18.47	24.21	54.12	70.93
26	RKglm5	n	60	non.str	global	0	33.71	95.46	14.43	19.49	42.90	57.94
26	RKglm5	n	60	str	global	0	33.81	99.54	17.26	22.51	51.31	66.91
26	RKglm5	n	60	non.str	local	0	33.9	96.13	13.99	19.21	41.59	57.10
26	RKglm5	n	60	str	local	0	34.01	99.1	16.83	22.2	50.03	65.99
26	RKglm5	n	80	non.str	global	0	34.16	99.47	13.71	18.72	40.28	54.99
26	RKglm5	n	80	str	global	0	34.31	100	16.72	22	49.12	64.63
26	RKglm5	n	80	non.str	local	0	34.26	100	13.3	18.5	39.07	54.35
26	RKglm5	n	80	str	local	0	34.44	100	16.4	21.8	48.18	64.04
26	RKglm5	n	100	non.str	global	0	34	84.67	13.67	18.68	40.16	54.88
26	RKglm5	n	100	str	global	0	34.2	100	17.16	22.58	50.41	66.33
26	RKglm5	n	100	non.str	local	0	34.22	92.38	13.25	18.45	38.92	54.20
26	RKglm5	n	100	str	local	0	34.41	100	16.71	22.32	49.09	65.57
26	RKglm5	ne	20	non.str	global	0.3	25.4	99.61	12.69	17.23	50.44	68.48
26	RKglm5	ne	20	str	global	0.14	25.4	98.08	13.34	18.28	53.02	72.66
26	RKglm5	ne	20	non.str	local	0.23	25.54	98.25	12.7	17.34	50.48	68.92
26	RKglm5	ne	20	str	local	0.26	25.56	98.25	13.35	18.34	53.06	72.89
26	RKglm5	ne	40	non.str	global	0.63	25.53	99.57	12.13	16.44	48.13	65.24
26	RKglm5	ne	40	str	global	0	25.56	100	13.82	18.99	54.84	75.36
26	RKglm5	ne	40	non.str	local	1.58	25.53	100	12.02	16.38	47.70	65.00
26	RKglm5	ne	40	str	local	0	25.56	100	13.76	18.94	54.60	75.16
26	RKglm5	ne	60	non.str	global	0	25.2	92.16	11.52	15.46	46.04	61.79
26	RKglm5	ne	60	str	global	0	25.37	100	13.29	18.1	53.12	72.34
26	RKglm5	ne	60	non.str	local	0	25.28	92.77	11.4	15.43	45.56	61.67
26	RKglm5	ne	60	str	local	0	25.44	100	13.21	18.11	52.80	72.38
26	RKglm5	ne	80	non.str	global	0	25.11	100	10.8	14.99	42.87	59.51
26	RKglm5	ne	80	str	global	0	25.2	100	12.7	17.88	50.42	70.98
26	RKglm5	ne	80	non.str	local	0	25.14	100	10.58	14.84	42.00	58.91
26	RKglm5	ne	80	str	local	0	25.21	100	12.52	17.79	49.70	70.62
26	RKglm5	ne	100	non.str	global	1.13	25.22	100	10.6	14.59	42.25	58.15
26	RKglm5	ne	100	str	global	0	25.29	100	12.62	17.78	50.30	70.86
26	RKglm5	ne	100	non.str	local	0.66	25.15	100	10.3	14.45	41.05	57.59
26	RKglm5	ne	100	str	local	0	25.19	100	12.39	17.69	49.38	70.51
26	RKglm5	sw	20	non.str	global	0	33.48	100	38.25	50.15	80.59	105.67
26	RKglm5	sw	20	str	global	0	33.47	100	39.89	51.66	84.05	108.85
26	RKglm5	sw	20	non.str	local	0	33.46	100	38.26	50.15	80.62	105.67
26	RKglm5	sw	20	str	local	0	33.47	100	39.89	51.66	84.05	108.85
26	RKglm5	sw	40	non.str	global	0	46.24	100	14.29	23.02	29.52	47.55
26	RKglm5	sw	40	str	global	0	46.27	100	40.47	50.05	83.60	103.39
26	RKglm5	sw	40	non.str	local	0	46.37	100	14.45	23.06	29.85	47.63
26	RKglm5	sw	40	str	local	0	46.53	100	40.7	50.34	84.07	103.99
26	RKglm5	sw	60	non.str	global	0	48.01	100	13.08	19.09	27.40	40.00
26	RKglm5	sw	60	str	global	0	47.97	99.92	44.55	54.52	93.34	114.23
26	RKglm5	sw	60	non.str	local	0	48.05	99.75	13.13	19.13	27.51	40.08
26	RKglm5	sw	60	str	local	0	48.05	99.91	44.37	54.29	92.96	113.74
26	RKglm5	sw	80	non.str	global	0	46.4	100	13.01	19.08	27.97	41.01
26	RKglm5	sw	80	str	global	0	46.39	100	42.81	53.06	92.02	114.06
26	RKglm5	sw	80	non.str	local	0	46.51	100	12.88	18.99	27.69	40.82
26	RKglm5	sw	80	str	local	0	46.73	100	42.87	53.06	92.15	114.06
26	RKglm5	sw	100	non.str	global	0	45.94	100	12.83	19.17	27.75	41.47
26	RKglm5	sw	100	str	global	0	45.97	100	48.17	56.8	104.20	122.86
26	RKglm5	sw	100	non.str	local	0	46.39	100	12.61	19.11	27.28	41.34
26	RKglm5	sw	100	str	local	0	46.7	100	48.29	56.91	104.46	123.10
27	RKgls1	n	20	non.str	global	3.59	32.32	85.3	13.59	17.97	42.20	55.81
27	RKgls1	n	20	str	global	0	32.18	100	14.02	18.4	43.54	57.14
27	RKgls1	n	20	non.str	local	3.49	32.24	84.92	13.58	17.93	42.17	55.68
27	RKgls1	n	20	str	local	0	32.03	100	14.07	18.41	43.70	57.17
27	RKgls1	n	40	non.str	global	0	34.16	77.64	14.07	19.08	41.22	55.90
27	RKgls1	n	40	str	global	0	34.31	100	14.84	20.35	43.48	59.62
27	RKgls1	n	40	non.str	local	0	34.15	82.32	13.99	19.14	40.99	56.08
27	RKgls1	n	40	str	local	0	34.29	100	14.72	20.41	43.13	59.80
27	RKgls1	n	60	non.str	global	0	33.65	100	13.46	18.69	40.01	55.56
27	RKgls1	n	60	str	global	0	33.72	100	14.83	20.67	44.08	61.44
27	RKgls1	n	60	non.str	local	0	33.77	100	13.25	18.58	39.39	55.23
27	RKgls1	n	60	str	local	0	33.84	100	14.68	20.61	43.64	61.27
27	RKgls1	n	80	non.str	global	0	34.04	100	12.85	17.89	37.75	52.56
27	RKgls1	n	80	str	global	0	34.25	100	14.92	20.69	43.83	60.78
27	RKgls1	n	80	non.str	local	0	34.14	100	12.72	17.89	37.37	52.56
27	RKgls1	n	80	str	local	0	34.36	100	14.82	20.69	43.54	60.78
27	RKgls1	n	100	non.str	global	0	33.95	100	12.98	18.13	38.13	53.26

Appendix

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
27	RKgls1	n	100	str	global	0	34.13	100	14.73	20.61	43.27	60.55
27	RKgls1	n	100	non.str	local	0	34.12	100	12.8	18.08	37.60	53.11
27	RKgls1	n	100	str	local	0	34.31	100	14.58	20.59	42.83	60.49
27	RKgls1	ne	100	non.str	global	3.07	25.12	100	10.68	14.53	42.57	57.91
27	RKgls1	ne	100	str	global	0	25.03	100	11.56	16.58	46.07	66.08
27	RKgls1	ne	100	non.str	local	2.71	24.99	100	10.34	14.37	41.21	57.27
27	RKgls1	ne	100	str	local	0	24.9	100	11.23	16.39	44.76	65.32
27	RKgls1	sw	100	non.str	global	0	46.03	100	12.21	16.91	26.71	36.99
27	RKgls1	sw	100	str	global	0	46.38	100	23.75	29.23	51.37	63.23
27	RKgls1	sw	100	non.str	local	0.03	46.42	100	12.11	16.82	26.20	36.38
27	RKgls1	sw	100	str	local	0	47.01	100	23.09	28.66	49.95	61.99
28	RKgls2	n	20	non.str	global	1.96	32.27	84.3	13.64	18.03	42.36	55.99
28	RKgls2	n	20	str	global	0	32.14	100	14.07	18.44	43.70	57.27
28	RKgls2	n	20	non.str	local	2.04	32.17	83.89	13.62	17.98	42.30	55.84
28	RKgls2	n	20	str	local	0	31.98	100	14.1	18.44	43.79	57.27
28	RKgls2	n	40	non.str	global	0	34.16	77.22	14.06	19.12	41.20	56.02
28	RKgls2	n	40	str	global	0	34.18	100	14.98	20.33	43.89	59.57
28	RKgls2	n	40	non.str	local	0	34.11	82.52	13.98	19.18	40.96	56.20
28	RKgls2	n	40	str	local	0	34.11	100	14.86	20.39	43.54	59.74
28	RKgls2	n	60	non.str	global	0	33.66	100	13.48	18.66	40.07	55.47
28	RKgls2	n	60	str	global	0	33.67	100	14.46	20.19	42.98	60.02
28	RKgls2	n	60	non.str	local	0	33.75	100	13.25	18.57	39.39	55.20
28	RKgls2	n	60	str	local	0	33.75	100	14.3	20.13	42.51	59.84
28	RKgls2	n	80	non.str	global	0	34.08	100	12.84	17.87	37.72	52.50
28	RKgls2	n	80	str	global	0	34.21	100	14.48	20.09	42.54	59.02
28	RKgls2	n	80	non.str	local	0	34.17	100	12.71	17.85	37.34	52.44
28	RKgls2	n	80	str	local	0	34.3	100	14.36	20.09	42.19	59.02
28	RKgls2	n	100	non.str	global	0	33.96	100	12.97	18.07	38.10	53.08
28	RKgls2	n	100	str	global	0	34.08	100	14.5	20.26	42.60	59.52
28	RKgls2	n	100	non.str	local	0	34.13	100	12.78	18.02	37.54	52.94
28	RKgls2	n	100	str	local	0	34.26	100	14.34	20.24	42.13	59.46
28	RKgls2	ne	100	non.str	global	2.11	25.09	100	10.58	14.46	42.17	57.63
28	RKgls2	ne	100	str	global	0	24.97	100	11.61	16.57	46.27	66.04
28	RKgls2	ne	100	non.str	local	1.93	24.98	100	10.25	14.28	40.85	56.92
28	RKgls2	ne	100	str	local	0	24.87	100	11.33	16.39	45.16	65.32
28	RKgls2	sw	100	non.str	global	0	46.1	100	12.34	17.17	27.00	37.56
28	RKgls2	sw	100	str	global	0	46.53	100	22.82	28.1	49.36	60.78
28	RKgls2	sw	100	non.str	local	0	46.46	100	12.13	16.98	26.24	36.73
28	RKgls2	sw	100	str	local	0	47.11	100	22.23	27.5	48.09	59.49
29	RKgls3	n	20	non.str	global	2.27	32.28	84.26	13.74	18.12	42.67	56.27
29	RKgls3	n	20	str	global	0	32.2	100	13.88	18.19	43.11	56.49
29	RKgls3	n	20	non.str	local	2.3	32.17	83.85	13.72	18.09	42.61	56.18
29	RKgls3	n	20	str	local	0	32.04	100	13.93	18.21	43.26	56.55
29	RKgls3	n	40	non.str	global	0	34.14	76.95	14.07	19.16	41.22	56.14
29	RKgls3	n	40	str	global	0	34.11	87.78	15.03	20.17	44.04	59.10
29	RKgls3	n	40	non.str	local	0	34.09	82.38	13.96	19.19	40.90	56.23
29	RKgls3	n	40	str	local	0	34.06	90.22	14.92	20.25	43.72	59.33
29	RKgls3	n	60	non.str	global	0	33.67	100	13.54	18.72	40.25	55.65
29	RKgls3	n	60	str	global	0	33.67	100	14.46	20.01	42.98	59.48
29	RKgls3	n	60	non.str	local	0	33.77	100	13.31	18.62	39.57	55.35
29	RKgls3	n	60	str	local	0	33.76	100	14.3	19.95	42.51	59.30
29	RKgls3	n	80	non.str	global	0	34.09	100	12.89	17.9	37.87	52.59
29	RKgls3	n	80	str	global	0	34.18	100	14.38	19.74	42.24	57.99
29	RKgls3	n	80	non.str	local	0	34.18	100	12.76	17.88	37.49	52.53
29	RKgls3	n	80	str	local	0	34.28	100	14.25	19.74	41.86	57.99
29	RKgls3	n	100	non.str	global	0	33.98	100	13.01	18.09	38.22	53.14
29	RKgls3	n	100	str	global	0	34.04	100	14.57	20.02	42.80	58.81
29	RKgls3	n	100	non.str	local	0	34.16	100	12.82	18.04	37.66	53.00
29	RKgls3	n	100	str	local	0	34.22	100	14.41	19.99	42.33	58.73
29	RKgls3	ne	100	non.str	global	1.79	25.1	100	10.58	14.45	42.17	57.59
29	RKgls3	ne	100	str	global	0	24.98	100	11.78	16.83	46.95	67.08
29	RKgls3	ne	100	non.str	local	1.71	25.01	100	10.24	14.26	40.81	56.84
29	RKgls3	ne	100	str	local	0	24.87	100	11.5	16.65	45.83	66.36
29	RKgls3	sw	100	non.str	global	0	46.31	100	12.42	17.08	27.17	37.37
29	RKgls3	sw	100	str	global	0	46.39	100	23.63	29.49	51.11	63.79
29	RKgls3	sw	100	non.str	local	0	46.68	100	11.71	16.55	25.33	35.80
29	RKgls3	sw	100	str	local	0	47.04	100	22.87	28.69	49.47	62.06
30	RKgls4	n	20	non.str	global	2.39	32.33	85.03	13.62	18.05	42.30	56.06
30	RKgls4	n	20	str	global	1.86	32.17	100	14.06	18.41	43.66	57.17
30	RKgls4	n	20	non.str	local	2.41	32.2	84.64	13.6	17.99	42.24	55.87
30	RKgls4	n	20	str	local	0.3	32	100	14.12	18.42	43.85	57.20
30	RKgls4	n	40	non.str	global	0	34.15	77.67	13.94	19.06	40.84	55.85
30	RKgls4	n	40	str	global	0	34.13	90.35	15.01	20.14	43.98	59.01
30	RKgls4	n	40	non.str	local	0	34.09	82.47	13.87	19.11	40.64	55.99
30	RKgls4	n	40	str	local	0	34.06	92.48	14.92	20.22	43.72	59.24
30	RKgls4	n	60	non.str	global	0	33.64	100	13.39	18.63	39.80	55.38
30	RKgls4	n	60	str	global	0	33.67	100	15.14	20.78	45.01	61.77
30	RKgls4	n	60	non.str	local	0	33.74	100	13.2	18.52	39.24	55.05
30	RKgls4	n	60	str	local	0	33.77	100	15	20.7	44.59	61.53
30	RKgls4	n	80	non.str	global	0	34.12	100	12.71	17.78	37.34	52.23

Appendix

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
30	RKgls4	n	80	str	global	0	34.29	100	15.37	20.93	45.15	61.49
30	RKgls4	n	80	non.str	local	0	34.18	100	12.65	17.82	37.16	52.35
30	RKgls4	n	80	str	local	0	34.36	100	15.29	20.95	44.92	61.55
30	RKgls4	n	100	non.str	global	0	33.99	100	12.9	18.03	37.90	52.97
30	RKgls4	n	100	str	global	0	34.16	100	15.13	20.65	44.45	60.66
30	RKgls4	n	100	non.str	local	0	34.15	100	12.75	17.99	37.46	52.85
30	RKgls4	n	100	str	local	0	34.33	100	15.01	20.61	44.10	60.55
30	RKgls4	ne	100	non.str	global	0.63	25.11	100	10.57	14.45	42.13	57.59
30	RKgls4	ne	100	str	global	0	24.99	100	11.77	16.74	46.91	66.72
30	RKgls4	ne	100	non.str	local	1.43	25	100	10.25	14.28	40.85	56.92
30	RKgls4	ne	100	str	local	0	24.86	100	11.52	16.6	45.91	66.16
30	RKgls4	sw	100	non.str	global	0	43.37	100	15.96	23.72	34.92	51.89
30	RKgls4	sw	100	str	global	0	46.22	100	38.86	46.4	84.06	100.37
30	RKgls4	sw	100	non.str	local	0	46.93	100	12.16	17.17	26.30	37.14
30	RKgls4	sw	100	str	local	0	47.41	100	38.31	45.74	82.87	98.94
31	RKgls5	n	20	non.str	global	0	32.11	100	13.57	17.93	42.14	55.68
31	RKgls5	n	20	str	global	0	32.08	100	14.71	19.29	45.68	59.91
31	RKgls5	n	20	non.str	local	0.56	31.91	100	13.59	18	42.20	55.90
31	RKgls5	n	20	str	local	0	31.85	100	14.78	19.44	45.90	60.37
31	RKgls5	n	40	non.str	global	0	34.35	100	14.34	19.8	42.02	58.01
31	RKgls5	n	40	str	global	0	34.3	100	15.69	20.93	45.97	61.32
31	RKgls5	n	40	non.str	local	0	34.28	100	14.29	19.86	41.87	58.19
31	RKgls5	n	40	str	local	0	34.23	100	15.66	21.04	45.88	61.65
31	RKgls5	n	60	non.str	global	0	33.75	100	13.83	19.23	41.11	57.16
31	RKgls5	n	60	str	global	0	33.94	100	16.81	22.69	49.97	67.45
31	RKgls5	n	60	non.str	local	0	33.88	100	13.58	19.05	40.37	56.63
31	RKgls5	n	60	str	local	0	34.09	100	16.55	22.5	49.20	66.88
31	RKgls5	n	80	non.str	global	0	34.18	100	12.95	18.17	38.04	53.38
31	RKgls5	n	80	str	global	0	34.32	100	16.03	21.59	47.09	63.43
31	RKgls5	n	80	non.str	local	0	34.24	100	12.86	18.17	37.78	53.38
31	RKgls5	n	80	str	local	0	34.39	100	15.9	21.58	46.71	63.40
31	RKgls5	n	100	non.str	global	0	33.97	100	13.09	18.28	38.45	53.70
31	RKgls5	n	100	str	global	0	34.23	100	15.93	21.79	46.80	64.01
31	RKgls5	n	100	non.str	local	0	34.16	100	12.9	18.21	37.90	53.50
31	RKgls5	n	100	str	local	0	34.43	100	15.7	21.65	46.12	63.60
31	RKgls5	ne	100	non.str	global	2.5	25.19	100	10.63	14.69	42.37	58.55
31	RKgls5	ne	100	str	global	0	25.18	100	12.73	17.88	50.74	71.26
31	RKgls5	ne	100	non.str	local	1.08	25.1	100	10.32	14.55	41.13	57.99
31	RKgls5	ne	100	str	local	0	25.08	100	12.57	17.78	50.10	70.86
31	RKgls5	sw	100	non.str	global	0	44.38	96.54	13.99	20.45	30.61	44.74
31	RKgls5	sw	100	str	global	0	44.25	100	50.92	60.24	110.14	130.30
31	RKgls5	sw	100	non.str	local	0	45.9	97.35	13.72	20.8	29.68	44.99
31	RKgls5	sw	100	str	local	0	44.95	100	51.08	60.39	110.49	130.63
32	RKlm1	n	20	non.str	global	3.88	32.33	83.95	13.57	17.91	42.14	55.62
32	RKlm1	n	20	str	global	0	32.15	100	14.02	18.38	43.54	57.08
32	RKlm1	n	20	non.str	local	3.48	32.23	83.52	13.58	17.88	42.17	55.53
32	RKlm1	n	20	str	local	0	31.98	100	14.09	18.4	43.76	57.14
32	RKlm1	n	40	non.str	global	0	34.15	77.64	14.07	19.06	41.22	55.85
32	RKlm1	n	40	str	global	0	34.29	100	14.78	20.28	43.31	59.42
32	RKlm1	n	40	non.str	local	0	34.14	82.33	13.98	19.12	40.96	56.02
32	RKlm1	n	40	str	local	0	34.27	100	14.67	20.35	42.98	59.62
32	RKlm1	n	60	non.str	global	1.98	33.6	81.13	13.41	18.52	39.86	55.05
32	RKlm1	n	60	str	global	0	33.62	100	14.35	19.96	42.66	59.33
32	RKlm1	n	60	non.str	local	1.71	33.71	86.73	13.2	18.43	39.24	54.79
32	RKlm1	n	60	str	local	0	33.71	100	14.2	19.89	42.21	59.13
32	RKlm1	n	80	non.str	global	2.05	34.07	94.42	12.79	17.85	37.57	52.44
32	RKlm1	n	80	str	global	0	34.09	100	13.87	19.31	40.75	56.73
32	RKlm1	n	80	non.str	local	2.29	34.15	94.89	12.68	17.85	37.25	52.44
32	RKlm1	n	80	str	local	0	34.18	100	13.76	19.33	40.42	56.79
32	RKlm1	n	100	non.str	global	0	33.95	87.74	12.9	18	37.90	52.88
32	RKlm1	n	100	str	global	0	34.02	100	14.03	19.56	41.22	57.46
32	RKlm1	n	100	non.str	local	0	34.11	91.63	12.74	17.95	37.43	52.73
32	RKlm1	n	100	str	local	0	34.18	100	13.88	19.53	40.78	57.37
32	RKlm1	ne	20	non.str	global	3.48	25.03	85.21	11.86	15.91	47.14	63.24
32	RKlm1	ne	20	str	global	3.51	25.07	90.78	12.34	16.94	49.05	67.33
32	RKlm1	ne	20	non.str	local	2.83	25.14	93.56	11.82	15.96	46.98	63.43
32	RKlm1	ne	20	str	local	2.83	25.15	92.67	12.34	17.02	49.05	67.65
32	RKlm1	ne	40	non.str	global	2.61	25.27	96.73	11.99	15.85	47.58	62.90
32	RKlm1	ne	40	str	global	2.55	25.29	100	12.77	17.48	50.67	69.37
32	RKlm1	ne	40	non.str	local	2.67	25.22	99.25	11.83	15.8	46.94	62.70
32	RKlm1	ne	40	str	local	2.67	25.24	100	12.66	17.46	50.24	69.29
32	RKlm1	ne	60	non.str	global	4.73	25.22	93.11	11.43	15.31	45.68	61.19
32	RKlm1	ne	60	str	global	0.49	25.19	100	12.02	16.83	48.04	67.27
32	RKlm1	ne	60	non.str	local	4.51	25.28	99.29	11.3	15.29	45.16	61.11
32	RKlm1	ne	60	str	local	0	25.21	100	11.88	16.8	47.48	67.15
32	RKlm1	ne	80	non.str	global	3.19	25.17	100	10.94	14.87	43.43	59.03
32	RKlm1	ne	80	str	global	0	25.11	100	11.86	16.85	47.08	66.89
32	RKlm1	ne	80	non.str	local	2.46	25.1	100	10.63	14.71	42.20	58.40
32	RKlm1	ne	80	str	local	0	25.03	100	11.56	16.68	45.89	66.22
32	RKlm1	ne	100	non.str	global	3.05	25.12	100	10.68	14.54	42.57	57.95

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
32	RKlm1	ne	100	str	global	0	25.04	100	11.56	16.54	46.07	65.92
32	RKlm1	ne	100	non.str	local	2.7	24.99	100	10.34	14.38	41.21	57.31
32	RKlm1	ne	100	str	local	0	24.91	100	11.23	16.35	44.76	65.17
32	RKlm1	sw	20	non.str	global	2.56	46.1	100	20.17	23.48	44.80	52.15
32	RKlm1	sw	20	str	global	0	45.21	100	30.38	38.39	67.48	85.27
32	RKlm1	sw	20	non.str	local	2.27	43.14	100	20.49	23.93	48.89	57.10
32	RKlm1	sw	20	str	local	1.44	43.44	100	26.73	32.66	63.78	77.93
32	RKlm1	sw	40	non.str	global	0.9	46.45	100	16.47	20.5	34.02	42.35
32	RKlm1	sw	40	str	global	0.94	46.98	100	26.94	32.2	55.65	66.52
32	RKlm1	sw	40	non.str	local	0	45.75	100	14.38	18.1	29.70	37.39
32	RKlm1	sw	40	str	local	0.73	46.8	100	26.45	31.8	54.64	65.69
32	RKlm1	sw	60	non.str	global	0	48.85	100	15.36	20.86	32.18	43.70
32	RKlm1	sw	60	str	global	0	47.92	100	29.62	36.29	62.06	76.03
32	RKlm1	sw	60	non.str	local	0	48.76	100	14.18	19.66	29.71	41.19
32	RKlm1	sw	60	str	local	0	48.05	100	28.88	35.47	60.51	74.31
32	RKlm1	sw	80	non.str	global	1.35	45.49	100	17.95	22.19	39.31	48.60
32	RKlm1	sw	80	str	global	0	45.35	100	30.45	37.89	66.69	82.98
32	RKlm1	sw	80	non.str	local	0.07	45.79	100	15.63	20.29	34.23	44.44
32	RKlm1	sw	80	str	local	0	46.51	100	29.13	36.94	63.80	80.90
32	RKlm1	sw	100	non.str	global	0	46.24	100	14.58	19.22	31.54	41.57
32	RKlm1	sw	100	str	global	0	45.89	100	32.7	40.36	70.73	87.30
32	RKlm1	sw	100	non.str	local	0	46.71	100	13.6	18.25	29.42	39.48
32	RKlm1	sw	100	str	local	0	46.99	100	31.8	39.6	68.79	85.66
33	RKlm2	n	20	non.str	global	0.34	32.27	100	14.02	18.61	43.54	57.80
33	RKlm2	n	20	str	global	0	32.11	100	14.74	19.47	45.78	60.47
33	RKlm2	n	20	non.str	local	0.63	32.16	100	14	18.6	43.48	57.76
33	RKlm2	n	20	str	local	0	31.9	100	14.8	19.52	45.96	60.62
33	RKlm2	n	40	non.str	global	0	34.3	93.01	14.45	19.8	42.34	58.01
33	RKlm2	n	40	str	global	0	34.37	100	15.93	21.54	46.67	63.11
33	RKlm2	n	40	non.str	local	0	34.24	92.78	14.33	19.8	41.99	58.01
33	RKlm2	n	40	str	local	0	34.28	100	15.81	21.58	46.32	63.23
33	RKlm2	n	60	non.str	global	0	33.67	100	13.61	18.9	40.46	56.18
33	RKlm2	n	60	str	global	0	33.75	100	14.93	20.83	44.38	61.92
33	RKlm2	n	60	non.str	local	0	33.74	100	13.39	18.82	39.80	55.95
33	RKlm2	n	60	str	local	0	33.79	100	14.77	20.75	43.91	61.68
33	RKlm2	n	80	non.str	global	0	34.11	100	12.96	18.08	38.07	53.11
33	RKlm2	n	80	str	global	0	34.27	100	14.39	20.04	42.27	58.87
33	RKlm2	n	80	non.str	local	0	34.17	100	12.82	18.06	37.66	53.06
33	RKlm2	n	80	str	local	0	34.33	100	14.27	20.03	41.92	58.84
33	RKlm2	n	100	non.str	global	0	33.97	100	13.15	18.41	38.63	54.08
33	RKlm2	n	100	str	global	0	34.15	100	14.66	20.6	43.07	60.52
33	RKlm2	n	100	non.str	local	0	34.12	100	12.94	18.34	38.01	53.88
33	RKlm2	n	100	str	local	0	34.29	100	14.47	20.55	42.51	60.37
33	RKlm2	ne	20	non.str	global	4.03	25.11	100	11.87	15.93	47.18	63.31
33	RKlm2	ne	20	str	global	3.49	25.12	100	12.46	17.18	49.52	68.28
33	RKlm2	ne	20	non.str	local	3.03	25.15	100	11.85	15.99	47.10	63.55
33	RKlm2	ne	20	str	local	2.83	25.19	100	12.45	17.23	49.48	68.48
33	RKlm2	ne	40	non.str	global	2.54	25.28	100	11.97	15.93	47.50	63.21
33	RKlm2	ne	40	str	global	0	25.2	100	12.74	17.63	50.56	69.96
33	RKlm2	ne	40	non.str	local	2.68	25.24	100	11.82	15.88	46.90	63.02
33	RKlm2	ne	40	str	local	0	25.14	100	12.65	17.62	50.20	69.92
33	RKlm2	ne	60	non.str	global	4.51	25.24	100	11.42	15.35	45.64	61.35
33	RKlm2	ne	60	str	global	0	25.14	100	12.06	16.98	48.20	67.87
33	RKlm2	ne	60	non.str	local	4.22	25.28	100	11.29	15.3	45.12	61.15
33	RKlm2	ne	60	str	local	0	25.14	100	11.94	16.94	47.72	67.71
33	RKlm2	ne	80	non.str	global	3.21	25.18	100	10.85	14.81	43.07	58.79
33	RKlm2	ne	80	str	global	0	25.03	100	11.85	16.99	47.04	67.45
33	RKlm2	ne	80	non.str	local	2.54	25.13	100	10.58	14.65	42.00	58.16
33	RKlm2	ne	80	str	local	0	24.95	100	11.55	16.82	45.85	66.77
33	RKlm2	ne	100	non.str	global	3.02	25.12	100	10.6	14.48	42.25	57.71
33	RKlm2	ne	100	str	global	0	24.98	100	11.57	16.64	46.11	66.32
33	RKlm2	ne	100	non.str	local	2.61	25.01	100	10.29	14.32	41.01	57.07
33	RKlm2	ne	100	str	local	0	24.86	100	11.26	16.45	44.88	65.56
33	RKlm2	sw	20	non.str	global	2.14	46.57	100	18.46	21.97	41.00	48.80
33	RKlm2	sw	20	str	global	0	46.01	100	30.46	38.15	67.66	84.74
33	RKlm2	sw	20	non.str	local	1.7	42.51	100	19.34	22.74	46.15	54.26
33	RKlm2	sw	20	str	local	1.78	44.3	100	26.87	32.35	64.11	77.19
33	RKlm2	sw	40	non.str	global	1.55	46.48	100	16.62	21.2	34.33	43.79
33	RKlm2	sw	40	str	global	0	46.88	100	27.85	33.6	57.53	69.41
33	RKlm2	sw	40	non.str	local	0.42	45.44	100	14.41	18.4	29.77	38.01
33	RKlm2	sw	40	str	local	0	46.2	100	27.68	33.63	57.18	69.47
33	RKlm2	sw	60	non.str	global	0	48.68	100	17.6	23.16	36.87	48.52
33	RKlm2	sw	60	str	global	0	48.11	100	32.67	40.46	68.45	84.77
33	RKlm2	sw	60	non.str	local	0	48.42	100	15.4	20.95	32.26	43.89
33	RKlm2	sw	60	str	local	0	47.88	100	31.29	39.06	65.56	81.84
33	RKlm2	sw	80	non.str	global	3.04	45.21	100	19.21	23.56	42.07	51.60
33	RKlm2	sw	80	str	global	3.85	45.23	100	32.42	39.38	71.00	86.25
33	RKlm2	sw	80	non.str	local	0	45.75	100	16.35	21.29	35.81	46.63
33	RKlm2	sw	80	str	local	0.25	46.35	100	30.95	38.58	67.78	84.49
33	RKlm2	sw	100	non.str	global	0	46.11	100	15.98	20.6	34.57	44.56

Appendix

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
33	RKlm2	sw	100	str	global	0	45.85	100	34.8	42.53	75.28	92.00
33	RKlm2	sw	100	non.str	local	0	46.56	100	14.31	19.1	30.95	41.32
33	RKlm2	sw	100	str	local	0	47.15	100	33.73	41.71	72.96	90.22
34	RKlm3	n	20	non.str	global	1.82	32.32	93.56	14.31	19.03	44.44	59.10
34	RKlm3	n	20	str	global	0	32.24	100	14.73	19.24	45.75	59.75
34	RKlm3	n	20	non.str	local	0.81	32.14	94.4	14.35	19.06	44.57	59.19
34	RKlm3	n	20	str	local	0	32.01	100	14.81	19.33	45.99	60.03
34	RKlm3	n	40	non.str	global	0	34.21	84.67	14.54	19.89	42.60	58.28
34	RKlm3	n	40	str	global	0	34.19	100	16.47	22.07	48.26	64.66
34	RKlm3	n	40	non.str	local	0	34.11	82.73	14.36	19.84	42.07	58.13
34	RKlm3	n	40	str	local	0	34.09	100	16.36	22.14	47.93	64.87
34	RKlm3	n	60	non.str	global	0	33.71	100	13.99	19.41	41.59	57.70
34	RKlm3	n	60	str	global	0	33.73	100	15.73	21.44	46.76	63.73
34	RKlm3	n	60	non.str	local	0	33.79	100	13.73	19.3	40.81	57.37
34	RKlm3	n	60	str	local	0	33.8	100	15.54	21.36	46.20	63.50
34	RKlm3	n	80	non.str	global	0	34.14	100	13.22	18.34	38.84	53.88
34	RKlm3	n	80	str	global	0	34.2	100	14.76	19.95	43.36	58.61
34	RKlm3	n	80	non.str	local	0	34.21	100	13.05	18.3	38.34	53.76
34	RKlm3	n	80	str	local	0	34.28	100	14.62	19.92	42.95	58.52
34	RKlm3	n	100	non.str	global	0	33.98	100	13.39	18.61	39.34	54.67
34	RKlm3	n	100	str	global	0	34.07	100	15.36	21	45.12	61.69
34	RKlm3	n	100	non.str	local	0	34.15	100	13.15	18.52	38.63	54.41
34	RKlm3	n	100	str	local	0	34.23	100	15.18	20.95	44.59	61.55
34	RKlm3	ne	20	non.str	global	4.04	25.1	100	11.9	15.97	47.30	63.47
34	RKlm3	ne	20	str	global	0.33	25.1	100	12.5	17.27	49.68	68.64
34	RKlm3	ne	20	non.str	local	3.07	25.16	100	11.88	16.02	47.22	63.67
34	RKlm3	ne	20	str	local	0	25.17	100	12.49	17.32	49.64	68.84
34	RKlm3	ne	40	non.str	global	2.2	25.31	100	11.89	15.84	47.18	62.86
34	RKlm3	ne	40	str	global	0	25.2	100	13.04	18.09	51.75	71.79
34	RKlm3	ne	40	non.str	local	1.07	25.26	100	11.78	15.82	46.75	62.78
34	RKlm3	ne	40	str	local	0	25.15	100	12.94	18.07	51.35	71.71
34	RKlm3	ne	60	non.str	global	3.58	25.26	100	11.41	15.36	45.60	61.39
34	RKlm3	ne	60	str	global	0	25.18	100	12.21	17.16	48.80	68.59
34	RKlm3	ne	60	non.str	local	3.24	25.31	100	11.3	15.31	45.16	61.19
34	RKlm3	ne	60	str	local	0	25.18	100	12.08	17.12	48.28	68.43
34	RKlm3	ne	80	non.str	global	3.29	25.19	100	10.84	14.81	43.03	58.79
34	RKlm3	ne	80	str	global	0	25.06	100	12.09	17.32	48.00	68.76
34	RKlm3	ne	80	non.str	local	2.61	25.15	100	10.55	14.64	41.88	58.12
34	RKlm3	ne	80	str	local	0	24.96	100	11.79	17.15	46.80	68.08
34	RKlm3	ne	100	non.str	global	3.18	25.13	100	10.6	14.49	42.25	57.75
34	RKlm3	ne	100	str	global	0	25	100	11.8	16.95	47.03	67.56
34	RKlm3	ne	100	non.str	local	2.7	25.02	100	10.27	14.31	40.93	57.03
34	RKlm3	ne	100	str	local	0	24.86	100	11.5	16.78	45.83	66.88
34	RKlm3	sw	20	non.str	global	1.67	47.88	100	17.7	22.08	39.32	49.04
34	RKlm3	sw	20	str	global	1.79	47.82	100	31.61	39.36	70.21	87.43
34	RKlm3	sw	20	non.str	local	1.21	44.49	100	18.1	22.56	43.19	53.83
34	RKlm3	sw	20	str	local	1.79	45.69	100	30.2	38.26	72.06	91.29
34	RKlm3	sw	40	non.str	global	7.82	49.28	100	18.2	21.87	37.60	45.18
34	RKlm3	sw	40	str	global	3.76	47.89	100	37.42	44.69	77.30	92.32
34	RKlm3	sw	40	non.str	local	0	47.73	100	15.53	19.57	32.08	40.43
34	RKlm3	sw	40	str	local	0.33	46.92	100	37.06	44.75	76.55	92.44
34	RKlm3	sw	60	non.str	global	0.43	48.53	100	16.51	22.05	34.59	46.20
34	RKlm3	sw	60	str	global	0	48.18	100	37.95	47.4	79.51	99.31
34	RKlm3	sw	60	non.str	local	0	48.78	100	14.78	20.31	30.97	42.55
34	RKlm3	sw	60	str	local	0	48.37	100	37.21	45.94	77.96	96.25
34	RKlm3	sw	80	non.str	global	4.21	45.79	100	17.52	21.61	38.37	47.33
34	RKlm3	sw	80	str	global	0	45.48	100	35.48	43.91	77.70	96.17
34	RKlm3	sw	80	non.str	local	0	46.07	100	15.26	19.52	33.42	42.75
34	RKlm3	sw	80	str	local	0	46.65	100	34.08	42.45	74.64	92.97
34	RKlm3	sw	100	non.str	global	1.2	46.46	100	14.5	18.97	31.36	41.03
34	RKlm3	sw	100	str	global	0	46.21	100	41.07	49.82	88.84	107.77
34	RKlm3	sw	100	non.str	local	0	46.9	100	13.06	17.69	28.25	38.27
34	RKlm3	sw	100	str	local	0	47.45	100	39.33	48.2	85.07	104.26
35	RKlm4	n	20	non.str	global	1.03	32.33	91.49	14.22	18.8	44.16	58.39
35	RKlm4	n	20	str	global	0	32.24	100	14.47	18.97	44.94	58.91
35	RKlm4	n	20	non.str	local	1.45	32.14	92.77	14.23	18.77	44.19	58.29
35	RKlm4	n	20	str	local	0	32	100	14.56	19	45.22	59.01
35	RKlm4	n	40	non.str	global	0	34.22	85	14.56	19.91	42.66	58.34
35	RKlm4	n	40	str	global	0	34.2	100	16.39	21.98	48.02	64.40
35	RKlm4	n	40	non.str	local	0	34.11	82.92	14.37	19.85	42.10	58.16
35	RKlm4	n	40	str	local	0	34.09	100	16.27	22.04	47.67	64.58
35	RKlm4	n	60	non.str	global	0	33.73	100	14.02	19.45	41.68	57.82
35	RKlm4	n	60	str	global	0	33.74	100	15.68	21.39	46.61	63.59
35	RKlm4	n	60	non.str	local	0	33.8	100	13.75	19.33	40.87	57.46
35	RKlm4	n	60	str	local	0	33.82	100	15.49	21.31	46.05	63.35
35	RKlm4	n	80	non.str	global	0	34.14	100	13.24	18.36	38.90	53.94
35	RKlm4	n	80	str	global	0	34.19	100	14.71	19.9	43.21	58.46
35	RKlm4	n	80	non.str	local	0	34.21	100	13.07	18.31	38.40	53.79
35	RKlm4	n	80	str	local	0	34.27	100	14.57	19.87	42.80	58.37
35	RKlm4	n	100	non.str	global	0	33.98	100	13.4	18.62	39.37	54.70

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
35	RKlm4	n	100	str	global	0	34.07	100	15.32	20.94	45.01	61.52
35	RKlm4	n	100	non.str	local	0	34.16	100	13.16	18.53	38.66	54.44
35	RKlm4	n	100	str	local	0	34.23	100	15.12	20.89	44.42	61.37
35	RKlm4	ne	20	non.str	global	4.38	25.07	100	11.86	15.99	47.14	63.55
35	RKlm4	ne	20	str	global	2.28	25.07	100	12.49	17.27	49.64	68.64
35	RKlm4	ne	20	non.str	local	3.35	25.19	100	11.89	16.04	47.26	63.75
35	RKlm4	ne	20	str	local	3.09	25.17	100	12.54	17.34	49.84	68.92
35	RKlm4	ne	40	non.str	global	1.09	25.32	100	11.83	15.79	46.94	62.66
35	RKlm4	ne	40	str	global	1.77	25.23	100	13.18	18.16	52.30	72.06
35	RKlm4	ne	40	non.str	local	0.47	25.3	100	11.74	15.79	46.59	62.66
35	RKlm4	ne	40	str	local	1.92	25.2	100	13.13	18.16	52.10	72.06
35	RKlm4	ne	60	non.str	global	3.64	25.24	100	11.37	15.32	45.44	61.23
35	RKlm4	ne	60	str	global	0	25.16	100	12.16	17.11	48.60	68.39
35	RKlm4	ne	60	non.str	local	3.44	25.32	100	11.29	15.3	45.12	61.15
35	RKlm4	ne	60	str	local	0	25.18	100	12.07	17.11	48.24	68.39
35	RKlm4	ne	80	non.str	global	2.98	25.17	100	10.82	14.78	42.95	58.67
35	RKlm4	ne	80	str	global	0	25.05	100	12.09	17.32	48.00	68.76
35	RKlm4	ne	80	non.str	local	2.22	25.16	100	10.55	14.63	41.88	58.08
35	RKlm4	ne	80	str	local	0	24.97	100	11.83	17.2	46.96	68.28
35	RKlm4	ne	100	non.str	global	2.62	25.13	100	10.6	14.48	42.25	57.71
35	RKlm4	ne	100	str	global	0	25	100	11.85	16.96	47.23	67.60
35	RKlm4	ne	100	non.str	local	2.82	25.03	100	10.28	14.31	40.97	57.03
35	RKlm4	ne	100	str	local	0	24.86	100	11.58	16.83	46.15	67.08
35	RKlm4	sw	20	non.str	global	3.68	47.85	100	17.18	21.33	38.16	47.38
35	RKlm4	sw	20	str	global	0	47.47	100	34.73	44.44	77.14	98.71
35	RKlm4	sw	20	non.str	local	4.2	46.37	100	18.04	22.01	43.04	52.52
35	RKlm4	sw	20	str	local	0	45.11	100	31.81	40.75	75.90	97.23
35	RKlm4	sw	40	non.str	global	1.87	46.36	100	17.16	21.29	35.45	43.98
35	RKlm4	sw	40	str	global	0	45.77	100	43.58	50.28	90.02	103.86
35	RKlm4	sw	40	non.str	local	0	45	100	16	20.82	33.05	43.01
35	RKlm4	sw	40	str	local	0	45.72	100	44.07	50.79	91.03	104.92
35	RKlm4	sw	60	non.str	global	0.59	48.49	100	15.75	20.6	33.00	43.16
35	RKlm4	sw	60	str	global	0	47.81	100	41.5	50.63	86.95	106.08
35	RKlm4	sw	60	non.str	local	0	48.1	100	15.24	20.43	31.93	42.80
35	RKlm4	sw	60	str	local	0	47.56	100	40.53	49.72	84.92	104.17
35	RKlm4	sw	80	non.str	global	3.42	45.14	100	18.28	23	40.04	50.37
35	RKlm4	sw	80	str	global	0	44.95	100	40.96	49.56	89.71	108.54
35	RKlm4	sw	80	non.str	local	0	45.42	100	16.6	21.68	36.36	47.48
35	RKlm4	sw	80	str	local	0	45.77	100	39.68	48.63	86.90	106.50
35	RKlm4	sw	100	non.str	global	1.6	46.09	100	15.11	19.45	32.68	42.07
35	RKlm4	sw	100	str	global	0	45.68	100	45.13	53.52	97.62	115.77
35	RKlm4	sw	100	non.str	local	0	46.61	100	13.83	18.78	29.92	40.62
35	RKlm4	sw	100	str	local	0	47.09	100	44.6	52.84	96.47	114.30
36	RKlm5	n	20	non.str	global	2	32.25	89.32	14.43	18.82	44.81	58.45
36	RKlm5	n	20	str	global	0	32.29	89.31	15.55	19.95	48.29	61.96
36	RKlm5	n	20	non.str	local	2.99	32.01	88.39	14.41	18.81	44.75	58.42
36	RKlm5	n	20	str	local	0	32.06	89.18	15.57	20.03	48.35	62.20
36	RKlm5	n	40	non.str	global	0	34.28	100	15.45	20.72	45.27	60.71
36	RKlm5	n	40	str	global	0	34.28	100	18.46	23.95	54.09	70.17
36	RKlm5	n	40	non.str	local	0	34.17	100	15.01	20.45	43.98	59.92
36	RKlm5	n	40	str	local	0	34.22	100	18.17	23.73	53.24	69.53
36	RKlm5	n	60	non.str	global	0	33.86	100	14.57	19.81	43.31	58.89
36	RKlm5	n	60	str	global	0	33.89	100	17.24	22.58	51.25	67.12
36	RKlm5	n	60	non.str	local	0	33.98	100	14.15	19.54	42.06	58.09
36	RKlm5	n	60	str	local	0	34.02	100	16.82	22.28	50.00	66.23
36	RKlm5	n	80	non.str	global	0	34.24	100	13.85	18.98	40.69	55.76
36	RKlm5	n	80	str	global	0	34.37	100	16.79	22.07	49.32	64.84
36	RKlm5	n	80	non.str	local	0	34.32	100	13.44	18.76	39.48	55.11
36	RKlm5	n	80	str	local	0	34.46	100	16.48	21.9	48.41	64.34
36	RKlm5	n	100	non.str	global	0	34.05	100	13.75	18.86	40.39	55.41
36	RKlm5	n	100	str	global	0	34.21	100	17.34	22.82	50.94	67.04
36	RKlm5	n	100	non.str	local	0	34.23	100	13.34	18.62	39.19	54.70
36	RKlm5	n	100	str	local	0	34.39	100	16.89	22.58	49.62	66.33
36	RKlm5	ne	20	non.str	global	2.87	25.15	100	12.63	17.16	50.20	68.20
36	RKlm5	ne	20	str	global	1.66	25.16	100	13.32	18.39	52.94	73.09
36	RKlm5	ne	20	non.str	local	2.23	25.32	100	12.7	17.35	50.48	68.96
36	RKlm5	ne	20	str	local	1.5	25.33	100	13.36	18.51	53.10	73.57
36	RKlm5	ne	40	non.str	global	1.63	25.54	100	12.4	16.8	49.21	66.67
36	RKlm5	ne	40	str	global	0	25.53	100	14.15	19.47	56.15	77.26
36	RKlm5	ne	40	non.str	local	1.53	25.59	100	12.27	16.72	48.69	66.35
36	RKlm5	ne	40	str	local	0	25.53	100	14.06	19.41	55.79	77.02
36	RKlm5	ne	60	non.str	global	1.58	25.23	95.41	11.8	15.76	47.16	62.99
36	RKlm5	ne	60	str	global	0	25.39	100	13.41	18.23	53.60	72.86
36	RKlm5	ne	60	non.str	local	0	25.34	97.67	11.6	15.65	46.36	62.55
36	RKlm5	ne	60	str	local	0	25.46	100	13.27	18.17	53.04	72.62
36	RKlm5	ne	80	non.str	global	1.35	25.11	100	11.11	15.22	44.10	60.42
36	RKlm5	ne	80	str	global	0	25.19	100	13.06	18.34	51.85	72.81
36	RKlm5	ne	80	non.str	local	1.74	25.14	100	10.75	14.96	42.68	59.39
36	RKlm5	ne	80	str	local	0	25.18	100	12.76	18.15	50.66	72.05
36	RKlm5	ne	100	non.str	global	2.65	25.2	100	11.02	15.06	43.92	60.02

Appendix

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
36	RKlm5	ne	100	str	global	0	25.25	100	12.98	18.22	51.73	72.62
36	RKlm5	ne	100	non.str	local	0.84	25.13	100	10.54	14.75	42.01	58.79
36	RKlm5	ne	100	str	local	0	25.15	100	12.61	17.99	50.26	71.70
36	RKlm5	sw	20	non.str	global	0.2	42.76	100	22.74	30.36	50.51	67.44
36	RKlm5	sw	20	str	global	0	42.84	100	44.41	55.08	98.65	122.35
36	RKlm5	sw	20	non.str	local	0.24	39.12	97.33	23.39	31.61	55.81	75.42
36	RKlm5	sw	20	str	local	0	39.2	97.12	42.49	53.64	101.38	127.99
36	RKlm5	sw	40	non.str	global	0	47.67	99.72	12.4	18.46	25.61	38.13
36	RKlm5	sw	40	str	global	0	47.62	100	39.72	48.93	82.05	101.07
36	RKlm5	sw	40	non.str	local	0	47.79	99.51	12.53	18.52	25.88	38.26
36	RKlm5	sw	40	str	local	0	47.94	100	40.02	49.26	82.67	101.76
36	RKlm5	sw	60	non.str	global	0	47.31	100	13.25	19.8	27.76	41.48
36	RKlm5	sw	60	str	global	0	47.27	100	43.72	53.77	91.60	112.65
36	RKlm5	sw	60	non.str	local	0	47.37	100	13.33	19.91	27.93	41.71
36	RKlm5	sw	60	str	local	0	47.3	100	43.49	53.56	91.12	112.21
36	RKlm5	sw	80	non.str	global	0	44.3	100	14.21	20.56	31.12	45.03
36	RKlm5	sw	80	str	global	0	44.24	100	41.66	51.74	91.24	113.32
36	RKlm5	sw	80	non.str	local	0	44.44	100	14.06	20.54	30.79	44.98
36	RKlm5	sw	80	str	local	0	44.8	100	41.8	51.92	91.55	113.71
36	RKlm5	sw	100	non.str	global	0	44.29	100	14.58	22.16	31.54	47.93
36	RKlm5	sw	100	str	global	0	44.16	100	49.1	57.37	106.21	124.10
36	RKlm5	sw	100	non.str	local	0	44.94	100	14.26	22.11	30.85	47.83
36	RKlm5	sw	100	str	local	0	45.05	100	49.35	57.63	106.75	124.66
37	RKrf	n	20	non.str	global	2.48	32.67	89.07	13.49	17.83	41.89	55.37
37	RKrf	n	20	str	global	2.48	32.66	89.59	15.91	21.35	49.41	66.30
37	RKrf	n	20	non.str	local	1.83	32.62	89.15	13.49	17.82	41.89	55.34
37	RKrf	n	20	str	local	1.83	32.6	90.38	15.91	21.34	49.41	66.27
37	RKrf	n	40	non.str	global	1.46	34.36	94.98	13.86	19.13	40.61	56.05
37	RKrf	n	40	str	global	1.81	34.36	95.15	20.93	27.13	61.32	79.49
37	RKrf	n	40	non.str	local	0.46	34.35	95.07	13.74	19.03	40.26	55.76
37	RKrf	n	40	str	local	1.17	34.35	93.57	20.76	26.96	60.83	78.99
37	RKrf	n	60	non.str	global	0.89	33.84	91.33	13.12	18.64	39.00	55.41
37	RKrf	n	60	str	global	0.6	33.83	91.38	20.08	26.29	59.69	78.15
37	RKrf	n	60	non.str	local	0.69	33.82	93.1	13.03	18.59	38.73	55.26
37	RKrf	n	60	str	local	0	33.83	92.75	19.97	26.19	59.36	77.85
37	RKrf	n	80	non.str	global	0.63	34.15	94.58	12.56	17.92	36.90	52.64
37	RKrf	n	80	str	global	0.76	34.15	94.58	20.16	26.91	59.22	79.05
37	RKrf	n	80	non.str	local	0.25	34.2	94.5	12.46	17.88	36.60	52.53
37	RKrf	n	80	str	local	0.92	34.2	95.13	20.06	26.8	58.93	78.73
37	RKrf	n	100	non.str	global	0.43	34.13	92.95	12.46	17.86	36.60	52.47
37	RKrf	n	100	str	global	0.34	34.12	93.05	21.07	27.8	61.90	81.67
37	RKrf	n	100	non.str	local	0	34.2	95.4	12.38	17.83	36.37	52.38
37	RKrf	n	100	str	local	0	34.19	94.46	20.95	27.68	61.55	81.32
37	RKrf	ne	20	non.str	global	2.92	25.28	71.27	12.07	16.16	47.65	63.80
37	RKrf	ne	20	str	global	2.96	25.27	68.42	12.45	16.9	49.48	67.17
37	RKrf	ne	20	non.str	local	2.09	25.3	71.26	12.06	16.17	47.61	63.84
37	RKrf	ne	20	str	local	2.07	25.38	68.41	12.48	16.9	49.60	67.17
37	RKrf	ne	40	non.str	global	2.1	25.43	85.55	11.11	15.23	43.69	59.89
37	RKrf	ne	40	str	global	2.27	25.35	84.57	14.11	19	55.99	75.40
37	RKrf	ne	40	non.str	local	1.93	25.39	86.09	11.06	15.15	43.49	59.58
37	RKrf	ne	40	str	local	1.68	25.35	84.74	14.06	18.96	55.79	75.24
37	RKrf	ne	60	non.str	global	1.56	25.45	82.7	10.36	14.44	40.82	56.90
37	RKrf	ne	60	str	global	1.76	25.25	84.22	13.95	19	55.76	75.94
37	RKrf	ne	60	non.str	local	1.41	25.48	83.83	10.32	14.41	40.66	56.78
37	RKrf	ne	60	str	local	1.58	25.29	83.77	13.86	18.91	55.40	75.58
37	RKrf	ne	80	non.str	global	1.1	25.24	88.74	9.79	13.98	38.86	55.50
37	RKrf	ne	80	str	global	1.36	25.27	87.98	14.09	19.59	55.93	77.77
37	RKrf	ne	80	non.str	local	0.41	25.23	90.67	9.73	13.94	38.63	55.34
37	RKrf	ne	80	str	local	1.18	25.25	89.42	14	19.51	55.58	77.45
37	RKrf	ne	100	non.str	global	1.12	25.14	88.57	9.37	13.37	37.35	53.29
37	RKrf	ne	100	str	global	0.95	25.16	88.51	13.31	18.83	53.05	75.05
37	RKrf	ne	100	non.str	local	0.53	25.12	89.83	9.32	13.35	37.15	53.21
37	RKrf	ne	100	str	local	0.29	25.14	88.93	13.23	18.75	52.73	74.73
37	RKrf	sw	20	non.str	global	0	47.86	92.58	16.81	21.96	34.84	45.51
37	RKrf	sw	20	str	global	1.15	47.79	90.26	36.98	46.79	76.64	96.97
37	RKrf	sw	20	non.str	local	0	47.9	93.77	16.78	21.98	34.78	45.55
37	RKrf	sw	20	str	local	1.15	47.79	90.26	36.98	46.79	76.64	96.97
37	RKrf	sw	40	non.str	global	0	47.76	90.34	9.96	14.79	20.71	30.75
37	RKrf	sw	40	str	global	0	47.63	90.94	38.85	49.64	80.25	102.54
37	RKrf	sw	40	non.str	local	0	47.79	90.54	9.98	14.8	20.75	30.78
37	RKrf	sw	40	str	local	0	47.64	90.95	38.84	49.62	80.23	102.50
37	RKrf	sw	60	non.str	global	0	48.5	90.89	10.99	17.67	23.03	37.02
37	RKrf	sw	60	str	global	0	48.65	92.5	43.05	53.93	90.19	112.99
37	RKrf	sw	60	non.str	local	0	48.33	91.19	10.92	17.61	22.88	36.90
37	RKrf	sw	60	str	local	0	48.63	92.5	43.01	53.88	90.11	112.88
37	RKrf	sw	80	non.str	global	0	45.42	88.98	9.36	13.98	20.50	30.62
37	RKrf	sw	80	str	global	0	45.53	91.37	39.35	50.46	86.18	110.51
37	RKrf	sw	80	non.str	local	0	45.2	88.89	9.35	13.95	20.48	30.55
37	RKrf	sw	80	str	local	0	45.52	91.4	39.36	50.5	86.20	110.60
37	RKrf	sw	100	non.str	global	0	46.38	90.65	8.45	13.47	18.28	29.14

Appendix

No	Method	Region	Sample density (%)	Sample stratification	Local.global	Min.	Mean	Max.	MAE	RMSE	RMAE (%)	RRMSE (%)
37	RKrf	sw	100	str	global	0	46.55	89.96	46.2	56.01	99.94	121.16
37	RKrf	sw	100	non.str	local	0	46.29	90.71	8.47	13.48	18.32	29.16
37	RKrf	sw	100	str	local	0	46.56	89.98	46.22	56.04	99.98	121.22

Appendix D. Test the effects of different backtransformations on the performance of OK using the dataset in the north region.

Back transformation of predictions based on transformed data is a critical issue in model prediction. Although this issue has long been known in the scientific community and the correct procedures for some transformations have been well documented for such as logarithmic transformations (Journel and Huijbregts, 1978; Webster and Oliver, 2001), incorrect or misleading procedures have been used in many literatures (Dambolena *et al*., 2009). The usually incorrect back transformation procedures used are such as taking the square of the prediction based on the square root-transformed data, and the exponent of the prediction based on the logarithmic transformed data. Such back transformations ignore that if a Gaussian distribution is back transformed (squared or antilogarithm), it becomes positively skewed, resulting in a mean larger than the median. These incorrect procedures resulted in a median rather than mean, thus underestimate the conditional mean of the response variable (Schuurmans *et al*., 2007).

However, for some transformations like square-root transformation, correct procedures of back transformation are not readily available. To solve this problem, there is a lot of money going around in processing meteorological data in the north side of the sphere (personal communication with Edzer Pebesma, 2 October 2009). In this study, normal score transformation (Goovaerts, 1997) was employed using the R script by Ashton Shortridge (2008), and an approach for backtransformation of square root data by (Schuurmans *et al*., 2007) was adopted using the R script kindly provided by Hanneke Schuurmans.

In this simulation experiment, we used square root-transformation and arcsine-transformation of primary variable to meet the requirement of data stationary. Given that lack of readily available computing program and time limit to the simulation work, we directly squared or anti-arcsine the predictions without considering the possible influence of such back transformation on the mean. Theorectically these simple back transformation yields correct prediction of median (Pebesma and Kwaadsteniet, 1997), an underestimation of mean.

To examine how these back transformation affect the prediction, we tested a number of scenarios: 1) using the original data to make prediction; 2) the square of the prediction based on the square root-transformed data; and 3) normal score transformation. We used mud content samples in the north Australian margin. Based on 10-fold cross validation the RMAE and RRMSE of OK using different datasets are summarised in Table D1.

Table D1. The accuracy of OK using datasets of mud content in north Australian margin under different transformations and back transformations. The method highlighted is the most accurate.

Methods	Transformation	Back transformation	RMAE (%)
IDW2-control	Original data	NA	37.04
RKrf	Original data	NA	36.37
OK	Original data	NA	37.49
OK	Normal score	Linear interpolation	37.28
OK	Square root	Schuurmans	37.72
OK	Square root	Squared	37.28

It is clearly shown that for OK, application of normal score transformation performed no better than square root after relevant back-transformation. Overall, RKrf and IDW2-control performed much better that OK no matter what kind of data transformation being employed. Therefore, we can conclude that 1) Square the predictions based on the square-root transformation performed slightly better that other transformation and back-transformation approaches; and 2) conclusions about the methods we identified in the experiment are still valid, that is RKrf is the most accurate method and the control is more accurate than OK in terms of RMAE.

There are a few major issues with normal transformation (http://www.gstat.org/gstat-info/msg00064.html). If there are few sample data observations the risk of underestimating the length of the tails or simply mis-specifying the population distribution is quite large. The second problem with normal scores is tied data. Back transformation of these tied data will not give the original data. In the transformed space, the first tied data is treated as a lower value than the second tied data. In addition, it is also hard to transform back values outside the observed range if rank order transformation was employed.