Environmental Modelling & Software 26 (2011) 1647–1659

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Application of machine learning methods to spatial interpolation of environmental variables

Jin Li*, Andrew D. Heap, Anna Potter, James J. Daniell

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

A R T I C L E I N F O

Article history:
Received 16 June 2011
Accepted 6 July 2011
Available online 6 August 2011

Keywords:
Geostatistics
Kriging
Random forest
Spatially continuous variable
Spatial prediction
Support vector machine

A B S T R A C T

Machine learning methods, like random forest (RF), have shown their superior performance in various disciplines, but have not been previously applied to the spatial interpolation of environmental variables. In this study, we compared the performance of 23 methods, including RF, support vector machine (SVM), ordinary kriging (OK), inverse distance squared (IDS), and their combinations (i.e., RFOK, RFIDS, SVMOK and SVMIDS), using mud content samples in the southwest Australian margin. We also tested the sensitivity of the combined methods to input variables and the accuracy of averaging predictions of the most accurate methods. The accuracy of the methods was assessed using a 10-fold cross-validation. The spatial patterns of the predictions of the most accurate methods were also visually examined for their validity. This study confirmed the effectiveness of RF, in particular its combination with OK or IDS, and also confirmed the sensitivity of RF and its combined methods to the input variables. Averaging the predictions of the most accurate methods showed no significant improvement in the predictive accuracy. Visual examination proved to be an essential step in assessing the spatial predictions. This study has opened an alternative source of methods for spatial interpolation of environmental properties.

1. Introduction

Spatially continuous data (spatially continuous surfaces) play a significant role in planning, risk assessment and decision making in environmental management and conservation. They are, however, usually not readily available and often difficult and expensive to acquire, especially for mountainous and deep marine regions. Data of environmental properties are usually collected by point sampling. However, scientists and environmental managers often require spatially continuous data over a region of interest to make informed decisions and justified interpretations. As geographic information systems (GIS) and modeling techniques are becoming powerful tools in natural resource management and biological conservation, spatially continuous data of environmental variables become increasingly required. The marine environment in Australia is a typical case, where a range of activities (e.g., seabed mapping and characterization, habitat classification, and prediction of marine biodiversity, essential for marine biodiversity conservation) requires reliable spatially continuous data (Pitcher et al., 2008; Whiteway et al., 2007). For most of the Australian Exclusive Economic Zone (AEEZ), such data are not available, and only sparsely- and unevenly-scattered point samples have been collected. Therefore, spatial interpolation techniques are essential for predicting the spatially continuous data of environmental properties for the unsampled locations using data from limited point observations within a region.

The existing spatial interpolation methods can be largely classified into three groups (Li and Heap, 2008): 1) deterministic or non-geostatistical methods (e.g., inverse distance squared: IDS), 2) stochastic or geostatistical methods (e.g., ordinary kriging: OK) and 3) combined methods (e.g. regression kriging). These methods are, however, often data- or even variable-specific and their performance depends on many factors (Li and Heap, 2011). No consistent findings have been acquired on how these factors affect the performance of spatial prediction methods, making it difficult to select an appropriate method for any given dataset (Li and Heap, 2011). Spatial interpolation methods such as IDS are commonly applied because of their relative simplicity and availability. However, predictions using IDS are usually associated with large predictive errors. Therefore, it is often a challenge to select an appropriate spatial interpolation method for a given study area.

Machine learning methods, like random forest (RF) and support vector machine (SVM), have shown their predictive accuracy in data mining fields and other disciplines in terms of predictive errors (Cutler et al., 2007; Diaz-Uriarte and de Andres, 2006; Drake et al., 2006; Marmion et al., 2009; Shan et al., 2006). However, RF has not been previously applied to the spatial interpolation of

* Corresponding author. Tel.: +61 2 6249 9899; fax: +61 2 6249 9956.
E-mail address: Jin.Li@ga.gov.au (J. Li).

1364-8152/$ – see front matter Crown Copyright © 2011 Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.envsoft.2011.07.004

environmental variables (Li and Heap, 2008) and SVM has been applied to rainfall data in a previous study (Gilardi, 2002). It was argued that RF is not sensitive to non-important variables fed into the model as it selects the most important variable at each node split (Okun and Priisalu, 2007). RF can also deliver good predictive performance even when most predictive variables are noisy (Diaz-Uriarte and de Andres, 2006). However, this assumption has not been tested for spatial interpolation of environmental variables.

The combined methods, such as regression kriging (Asli and Marcotte, 1995; Odeh et al., 1995), are argued to be less sensitive to variation in data and more accurate than other methods (Hengl et al., 2007; Li and Heap, 2008). However, it is not clear whether the combination of machine learning methods with existing methods, such as OK or IDS, can improve the prediction accuracy because they have not been applied in the spatial interpolation of environmental variables (Li and Heap, 2008).

Model averaging (also termed as ensemble mean, ensemble forecasting, consensus method, average of forecasts, combining or combination forecasts etc., see Araújo and New (2007) and Gregory et al. (2001) for more information) have been applied to a number of disciplines including ecology (Araújo and New, 2007; Ellison, 2004), economics (Gregory et al., 2001), biomedicine (Nilsson et al., 2000), climatology (Raftery et al., 2005) and hydrology (Goswami and O'Connor, 2007). Model averaging can often improve the predictive accuracy (Marmion et al., 2009), however, this has not been tested in spatial interpolation of environmental properties using machine learning methods.

In this study, we aim to address the following questions: 1) can machine learning methods like RF and SVM be used for spatial interpolation of environmental variables? 2) can the combination of these machine learning methods with existing spatial interpolation methods like IDS and OK improve the predictive accuracy? 3) are RF and its combined methods sensitive to their input variables? and 4) will model averaging reduce the predictive error? To address these questions, we applied the machine learning methods, a number of existing spatial interpolation methods including IDS and the combined methods (i.e., combination of RF or SVM with IDS or OK respectively) to seabed mud content samples extracted from Geoscience Australia's Marine Samples Database (MARS; www.ga.gov.au/oracle/mars) in February 2008. We examined the effects of the exclusion of a few least important variables on the performance of RF and its combined methods. The accuracy of averaging the predictions of the most accurate methods was tested. Finally, the prediction patterns of the most accurate methods and the control method were analyzed based on their prediction maps.

2. Methods

2.1. Study area and datasets

The study area is located in the southwest region of AEEZ (Fig. 3.1). This region has a north–south orientation, covers an area of 523,000 km^2 and a large range of water depths ranging from 0 to 5539 m, and comprises four geomorphic provinces (Heap and Harris, 2008). The area of geomorphic provinces varies much among geomorphic provinces, with shelf (53,000 km^2), slope (215,000 km^2), rise (52,000 km^2) and abyssal plain/deep ocean floor (203,000 km^2).

In total, 177 samples of seabed mud content are considered in this study following data quality control (Li et al., 2010). The mud content (weight %) is derived by the relative weight proportion of 10–20 g dry sediment passing through a standard mesh size (Li et al., 2010). The grain size of mud content is <63 μm. The spatial distribution of the samples is uneven (Fig. 1), with 65 on the shelf, 101 on the slope, 3 on the rise and 8 on the abyssal plain/deep ocean floor. Sample density is very low, with 0.3 samples per 1000 km^2 on average.

A number of variables can be used as secondary information to improve the performance of spatial interpolation techniques as discussed by Li and Heap (2008). Following a preliminary analysis, geomorphic province and bathymetry data that were available at a resolution of 0.01° were considered as important secondary variables in this study. Bathymetry has previously been used to improve the performance of spatial interpolators of seabed sediments (Verfaillie et al., 2006), so the inclusion of such information was expected 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 with the potential to improve the overall predictions. Since the relationship between the mud content and these secondary variables are non-linear (Li et al., 2010), Spearman's rank correlation rho (ρ) was used measure their correlation: the mud content displays a significant correlation with bathymetry ($\rho = -0.6286$, p-value $= 0.0000$), distance to coast ($\rho = 0.6171$, p-value $= 0.0000$) and slope ($\rho = 0.5522$, p-value $= 0.0000$).

All datasets of secondary variables were generated in ArcGIS and, where necessary, resampled to a 0.01° 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 adjacent major islands) from Geoscience Australia's coastline dataset at 1:250,000 scale, 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 into UTM grids. Slope gradient (in degrees) was calculated separately for each grid, and then each of the slope grids was projected back to World Geodetic System 1984 (WGS84) and merged. The merged grid was finally resampled to 0.01° resolution.

Sample stratification by categorical variables 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 (Fig. 1) mean that sub-setting by feature types, leading to some features without sample, cannot provide adequate sample size for modeling. 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.

A range of other variables could be used as secondary information to improve the spatial interpolation of marine environmental data. These may include topology, substrate type, 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. Consequently, these variables were not used in this study.

The coordinates of the mud content data were in latitude and longitude based on WGS84 in this study. Geostatistical methods such as those in the gstat (Pebesma, 2004) package in R (R Development Core Team, 2008) expect the input data having been equal distance projected. For data in WGS84, gstat assumes that a unit difference in longitude reflects approximately the same distance in latitude and it ignores the changes in distance along the latitude in computing a distance between two points in variogram modeling and kriging (pers. comm. with E. Pebesma, May 2009). Although we were aware of these limitations in using gstat, we decided to model our data in WGS84, because: 1) the study area spans two UTM zones, 2) equal distance projections do not produce satisfactory projections, and 3) the control method (IDS) has used WGS84.

2.2. Methods for spatial interpolation

To test the performance of machine learning methods and their combination with existing methods for spatial interpolation, we applied 23 methods to the mud content samples (Table 1). These methods largely fall into five categories: 1) non-geostatistical spatial interpolation methods, 2) geostatistical methods, 3) statistical methods, 4) machine learning methods, and 5) combined methods. These methods were selected mainly according to the review of over 40 spatial interpolation methods by Li and Heap (2008) and also on the basis of the applications of machine learning methods in previous studies (Arthur et al., 2010; Drake et al., 2006; Shan et al., 2006).

Non-geostatistical spatial interpolation methods: Although the inverse distance weighting (IDW) method performs poorly in most cases, it can be used as a control because it is a commonly compared method in spatial interpolation studies (Li and Heap, 2011).

Geostatistical methods: Kriging with an external drift (KED) and Ordinary cokriging (OCK) (Goovaerts, 1997) were compared because they have been proven to be able to obtain high accuracy 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 prediction (Li and Heap, 2008). Universal kriging (UK) was also considered.

Statistical methods: Generalized least squares trend estimation (GLS) (Bivand et al., 2008) is used because it allows errors to be correlated (Pinheiro and Bates, 2000; Venables and Ripley, 2002). Regression kriging methods have been applied to a number of disciplines in environmental sciences (Li and Heap, 2008), which usually use linear regression model (LM). Given that this study is to model percentage data, generalized linear model (GLM) was employed. Hence these methods have been considered in combination with OK.

Machine learning methods: Three machine learning approaches were also considered in this study, namely: regression tree (Rpart) (Breiman et al., 1984), random forest (RF) (Breiman, 2001; Strobl et al., 2007) and Support vector machine (SVM) (Cortes and Vapnik, 1995). One application of SVM to spatial interpolation was reported by Gilardi and Bengio, 2000. RF has not been reported previously for spatial interpolation in environmental science (Li and Heap, 2008).

Combined methods: Combined methods include LM and OK (RKlm), GLM and OK (RKglm), GLS and OK (RKgls), Rpart and OK or IDS (RpartOK, RpartIDS), RF and OK or IDS (RFOK, RFIDS), SVM and OK or IDS (SVMOK, SVMIDS). For SVM, four sub-methods were used because of their availability (SVM and KSVM) and of different kernels (radial base and linear). These combined approaches are modified versions of regression kriging type C (RK-C) (Asli and Marcotte, 1995; Odeh et al., 1995) that is less sensitive to data variation and more accurate than other methods (Hengl et al., 2007; Li and Heap, 2008). For these combined methods, firstly LM, GLM, GLS, Rpart, RF, SVM and KSVM were applied to the data, then OK or IDS was applied to the residuals of these models, and finally the predicted values of each model and the corresponding interpolated residual values were added together to produce the final predictions of each combined method. The combination of RF and SVM with OK or IDS are novel and has not been applied in previous studies in environmental sciences (Li and Heap, 2008).

2.3. Data transformation and variogram modeling

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 modeled as if they were the primary variable. Distributions of mud, bathymetry, distant to coast and slope are left-skewed and non-normal, so appropriate transformations were identified for each variable: arcsine for mud content, and double square root for bathymetry, distance-to-coast and slope. Mud content was also square-root transformed for generalized least squares and transformed to between 0 and 1 for generalized linear models.

Variogram modeling: No obvious directional trend was detected in the semivariogram map. There are a number of variogram models that could be employed and different variogram models may lead to different predictions (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 in gstat in R (R Development Core Team, 2008). Of these models, Spherical model was selected for kriging methods as it fitted the data and the residuals of relevant methods better than other variogram models in terms of range, nugget and sill.

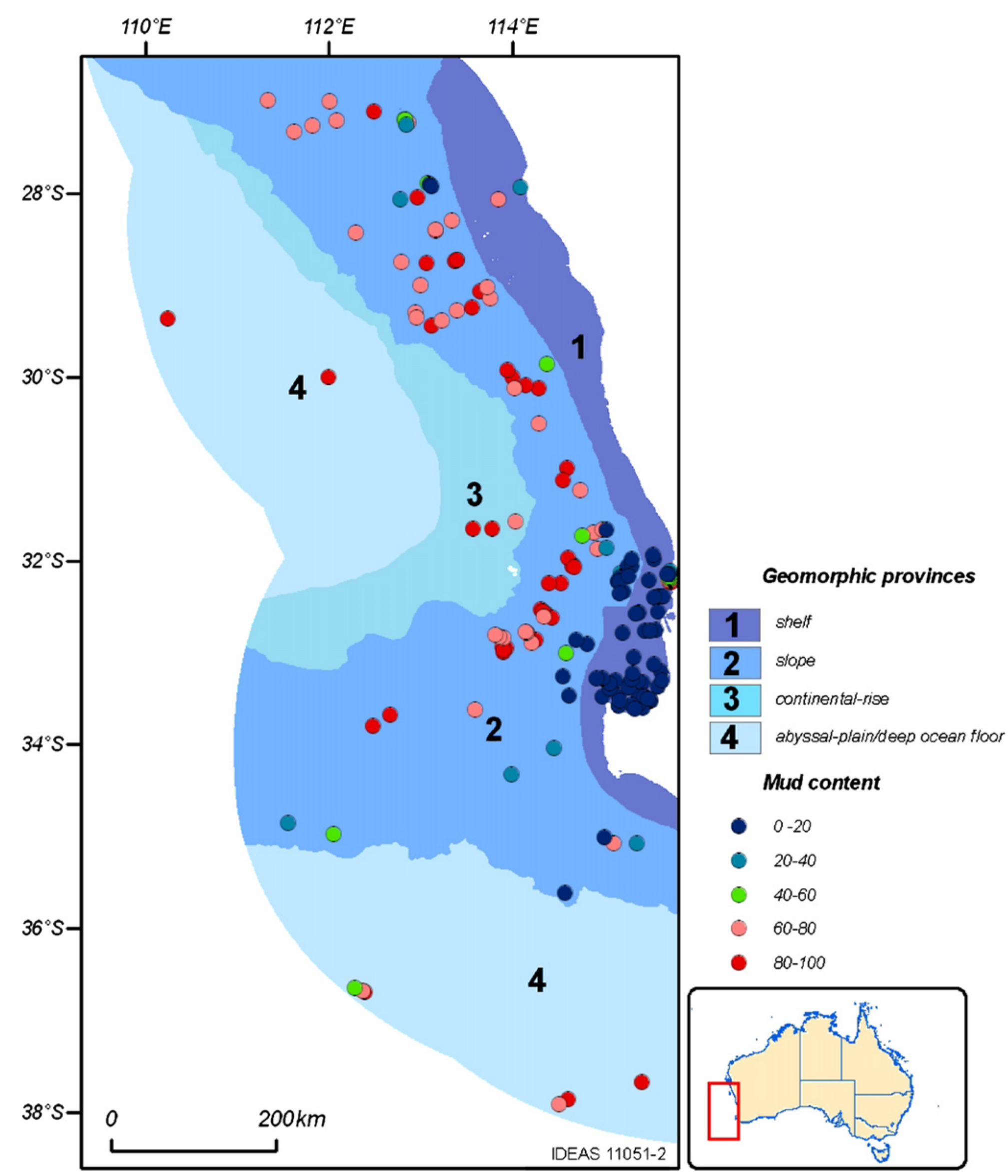

Fig. 1. Spatial distribution of mud content samples, including their occurrence in the geomorphic provinces.

Table 1
Methods compared for predicting mud content in the southwest region of AEEZ.[a]

No	Method	Method type
1	Inverse distance weighting (IDW) with distance power 2 to 4	Non-geostatistical spatial interpolation method (NGSIM)
2	Kriging with an external drift (KED)	Geostatistical method (GM)
3	Ordinary cokriging (OCK)	GM
4	Ordinary kriging (OK)	GM
5	Universal kriging (UK)	GM
6	Regression tree (Rpart)	Machine learning method (MLM)
7	Support vector machine with radial basis kernel (SVM)	MLM
8	SVM with linear kernel (SVMlinear)	MLM
9	Support vector machine (KSVM)	MLM
10	RandomForest (RF)	MLM
11	Linear models and OK (RKlm)	Combination of statistical method and GM (CSMG)
12	Generalized linear models and OK (RKglm)	CSMG
13	Generalized least squares and OK (RKgls)	CSMG
14	Rpart and IDS (RpartIDS)	Combination of MLM and NGSIM (CMN)
15	Rpart and OK (RpartOK)	Combination of MLM and GM (CMG)
16	KSVM and IDS (KSVMIDS)	CMN
17	KSVM and OK (KSVMOK)	CMG
18	SVM and IDS (SVMIDS)	CMN
19	SVM and OK (SVMOK)	CMG
20	SVMlinear and IDS (SVMlinearIDS)	CMN
21	SVMlinear and OK (SVMlinearOK)	CMG
22	RF and IDS (RFIDS)	CMN
23	RF and OK (RFOK)	CMG

[a] For SVM and KSVM, we also used polynomial kernel but no results were produced, hence not reported in this study.

2.4. Parameter specification and secondary variable selection

The specification of parameters was based on the requirements of the methods and data nature. For GLS, spherical spatial correlation was specified. For GLM, a quasibinomial family with a logit link was used. For RF, the number of trees was specified as 2000 to generate reliable predictions and the number of candidate variables randomly sampled at each split was specified as 6 according to tuneRF (i.e., tune randomForest). To prune the regression tree generated by rpart with complexity parameter of 0.001, the complexity parameter of 0.006 was specified on the basis of the plot of 10-fold cross-validation results. In SVM, the gamma parameter for kernel and the cost of constraints violation were selected based on the mean squared error produced by 10-fold cross-validation, and the best choice of cost was 0.5 and the optimal gamma was 1. In KSVM, the sigma parameter was assigned 5 according to the results of 10-fold cross-validation. A distance power of 2, 3 and 4 was used in IDW. Of which IDS, a commonly used method (Li and Heap, 2011), was used as the control.

The secondary variables used for each method are summarized in Table 2. Given that bathymetry was the most strongly correlated variable with mud content based on the Spearman's rank correlation coefficient, it was used as a secondary variable in all methods that consider secondary information. Distance-to-coast and slope were also used in relevant models. Latitude and longitude up to third-order polynomial (i.e., the terms in Legendre and Legendre's (1998) equation) were used in UK and machine learning methods. For machine learning methods and their combinations with IDS or OK, we used all possible secondary variables, their second and third power.

Table 2
Parameters or secondary variable used for each of the 23 methods: bathy, bathymetry; dist.coast, distance-to-coast; lat, latitude; lon, longitude; prov, geomorphic province.[a] The secondary variables for KED, OCK, RKglm, RKgls and RKlm were determined based on a previous study by Li et al. (2010).

Method	Secondary variables
IDW	na
KED	bathy
OCK	sqrt(sqrt(bathy*(−1))), sqrt(sqrt(dist.coast)), sqrt(sqrt(slope))
OK	na
UK	lat, lat2, lon, lon2, lat*lon, lat*lon2, lon*lat2, lat3, lon3
RKglm, RKgls, RKlm	bathy, dist.coast, slope, prov
RF, RFIDS, RFOK, Rpart, RpartIDS, RpartOK, SVM, SVMIDS, SVMOK, SVMlinear, SVMlinearIDS, SVMlinearOK, KSVM, KSVMIDS, KVMOK	bathy, dist.coast, slope, bathy2, bathy3, dist.coast2, dist.coast3, slope2, slope3, lat, lat2, lon, lon2, lat*lon, lat*lon2, lon*lat2, lat3, lon3

[a] prov was initially considered as a secondary variable for RF, but it was excluded because of the small sample size in two geomorphic provinces and of its least importance.

The modeling was implemented in R (R Development Core Team, 2008), using packages e1071, gstat, kernlab, nlme, rpart and randomForest, with a searching neighborhood size of 20 if applicable. Predictions were corrected by resetting the faulty estimates to the nearest bound of the data range (i.e., 0 or 100%) if applicable (Goovaerts, 1997).

2.5. Exclusion of the least important secondary variables

We then dropped slope related variables to test the sensitivity of RF and its combined methods to the changes in their input variables. Slope related variables were dropped because they were the least important variable in terms of the results of RF (see Fig. 8):

- RF without slope;
- RFOK without slope; and
- RFIDS without slope.

2.6. Model averaging

We tested if averaging the predictions of the best performing methods could further improve the prediction accuracy. In this study, we averaged the predictions of two methods (RFOK and RFIDS) or three methods (RFOK, RFIDS and RF) to produce the final predictions; hence we used the term, "model averaging". This method was found to be one of the most accurate methods for averaging the predictions from different modeling methods (Marmion et al., 2009).

2.7. Assessment of method performance

To compare the performance of these methods, a 10-fold cross-validation was used. An existing cross-validation program could be used to do this task, but due to random sampling, each method may receive different samples for prediction and validation. To avoid this random error, we randomly split the mud sample dataset into 10 sub-datasets. Nine sub-datasets were combined and used for model development and the remaining one was used to validate the predictions of the model developed. This process was repeated, varying the validation dataset, until all 10 sub-datasets had been allocated for validation. Consequently, all methods were applied to the same training and test datasets. Each method generated 10 prediction datasets.

The performance of these methods was assessed by identifying the error in the predictions. For each method, the predictions from the 10 sub-datasets were compared to the observed values in the 10 corresponding validation sub-datasets. Relative mean absolute error (RMAE) and relative root mean square error (RRMSE), which are not sensitive to the changes in unit/scale (Li and Heap, 2011), were used to assess the performance of the methods tested and subsequently to compare with findings in previous studies from various disciplines in environmental sciences:

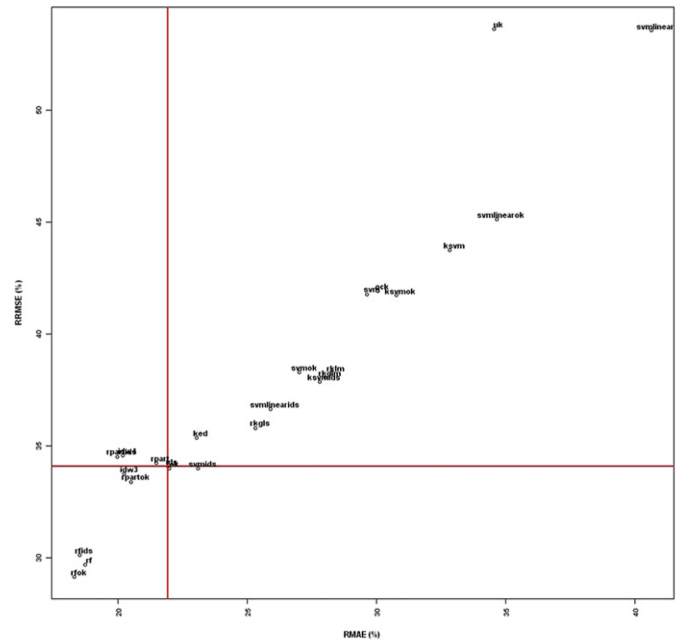

Fig. 2. The relative absolute mean error (RMAE (%)) and relative root mean square error (RRMSE (%)) of modeling methods for mud content dataset in the southwest region. The horizontal and vertical lines (red) indicate the accuracy of the control (IDS) (Li et al., 2010). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$RMAE = \frac{1}{n}\sum_{i=1}^{n}|(p_i - o_i)|/o_m 100 \quad (1)$$

$$RRMSE = \left[\frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2\right]^{1/2}/o_m 100 \quad (2)$$

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

The data manipulation and computation were implemented in R (R Development Core Team, 2008).

3. Results

3.1. Performance of the methods

The predictive error varies with the methods compared (Fig. 2). RFOK, RFIDS and RF are the three most accurate methods, RPARTOK and IDW3 also outperformed the control (IDS). All the remaining methods performed relatively poorly and the accuracy of various support machine learning methods varies considerably. The combination of various support vector machine methods (SVM, SVMlinear and KSVM) with either OK or IDS considerably improved their prediction accuracy, although they are less accurate than IDS and OK. The results of paired t-test indicate that RFOK, RKIDS and RF significantly outperformed the control (Table 3).

Table 3
Comparison of the most accurate five methods with the control method (IDS) based on the results of the 10-fold cross-validation. p-values derived from paired t-test of the predictive errors (i.e., RMAE and RRMSE) of IDS with the most accurate methods to show whether the predictive errors of IDS are greater than those of the most accurate methods.

Method	RMAE	RRMSE
RFOK	0.0147	0.0374
RFIDS	0.0106	0.0339
RF	0.0298	0.0654
RpartOK	0.1780	0.3919
IDW3	0.0092	0.3302

Table 4
Effects of slope exclusion on the prediction error of random forest and its combined methods. Paired t-test was used to examine if the predictive errors (i.e., RMAE and RRMSE) of methods with slope are greater than those without slope based on the results of the10-fold cross-validation.

Method	Slope	RMAE (%)	p-value	RRMSE (%)	p-value
RF	Yes	18.73	0.0051	29.70	0.0033
RF	No	18.04		28.90	
RFOK	Yes	18.32	0.0324	29.16	0.0212
RFOK	No	17.87		28.68	
RFIDS	Yes	18.52	0.0171	30.13	0.0717
RFIDS	No	18.02		29.79	

3.2. Effects of the exclusion of slope related variables

The prediction errors of all three methods (i.e., RF, RFOK and RFIDS) are reduced after the exclusion of slope related variables in terms of RMAE and RRMSE, although p-values change with the methods and with predictive error measurements (Table 4). Overall, the methods without slope are relatively more accurate than those with slope.

3.3. Effects of averaging the predictions of the most accurate methods

The effects of averaging the predictions of two or three most accurate methods (i.e., RFOK and RFIDS or RFOK, RFIDS and RF, with and without slope) are summarized in Table 5. The averaging of the predictions of two or three most accurate methods produces negligible effects on the prediction accuracy in comparison with RFOK, the most accurate method in terms of both RMAE and RRMSE whether or not slope is included.

3.4. Visual examination

3.4.1. With slope related variables

The spatial predictions of two most accurate methods (i.e., RFIDS and RFOK) with slope, their average, and IDS are illustrated in Fig. 3. The spatial patterns are similar between the two methods (i.e., RFIDS and RFOK) capturing similar major spatial patterns and trends of seabed mud content, but 'bull's eyes' patterns are evident for RFIDS. Linear tracks, sharp transitions and banding patterns are apparent in the predictions. The average of these two methods produced a map similar to that of RFIDS, but with the 'bull's eyes' patterns slightly weakened. The predictions of IDS displayed similar major patterns, but failed to predict the changes in mud content from shelf to deep ocean floor and displayed strong 'bull's eyes' patterns at sample points having either high or low values.

3.4.2. Without slope related variables

The spatial predictions of the best two methods (i.e., RFIDS and RFOK) without slope and their average reveal similar patterns as those with slope (Fig. 4). The major spatial patterns are largely

Table 5
The accuracy of model averaging (i.e., averages of the predictions of RFOK and RFIDS or RFOK, RFIDS and RF, with and without slope). p-values are from paired t-test of their predictive errors (i.e., RMAE and RRMSE) against the performance of RFOK with or without slope based on the results of the10-fold cross-validation to test if model averaging reduces the predictive error.

Method	Slope	RMAE	p-value	RRMSE	p-value
RFOK and RFIDS	Yes	18.34	0.4023	29.46	0.6282
RFOK and RFIDS	No	17.85	0.3599	29.05	0.7080
RF, RFOK and RFIDS	Yes	18.43	0.5368	29.42	0.6570
RF, RFOK and RFIDS	No	17.87	0.3923	28.88	0.6568

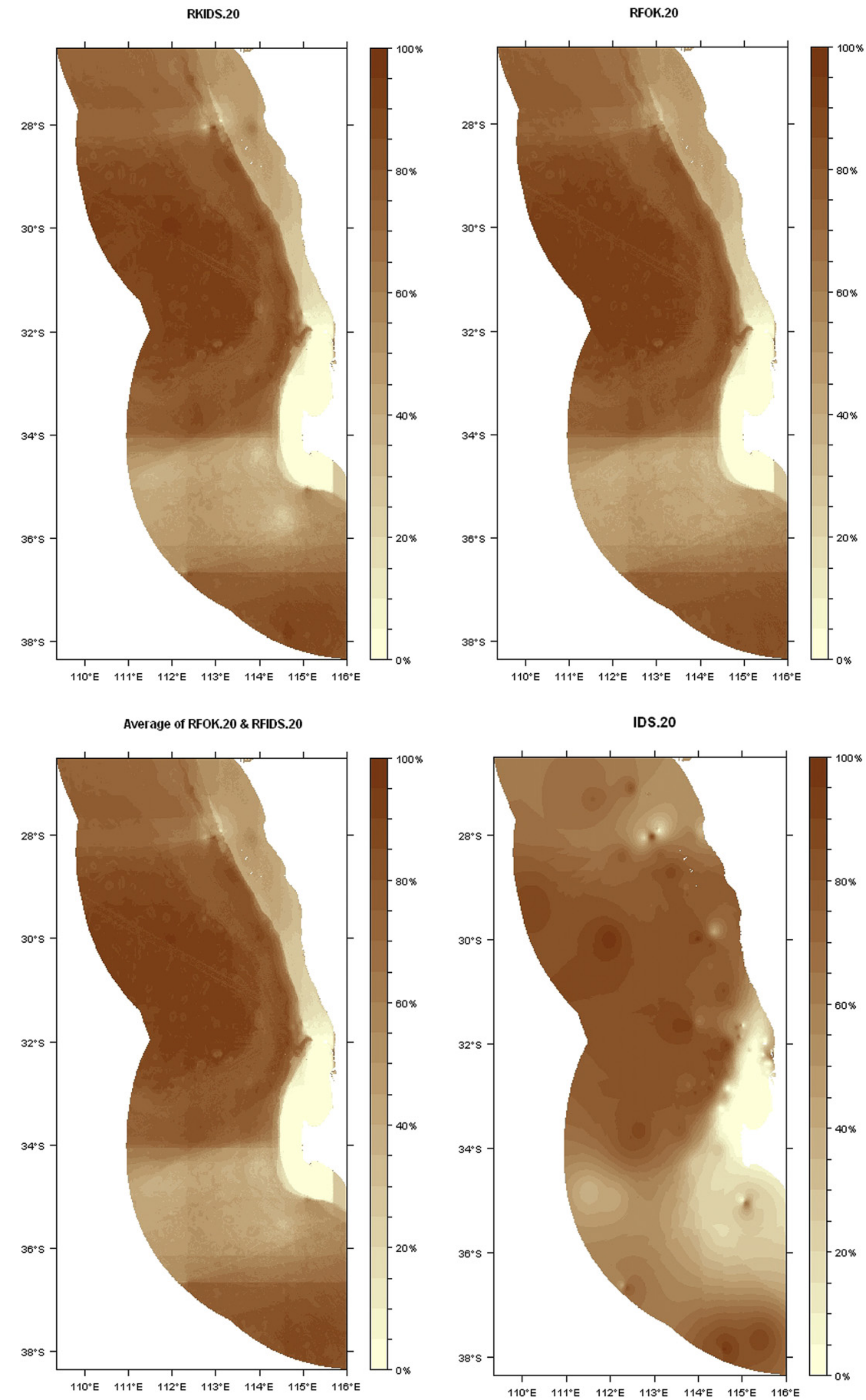

Fig. 3. The predictions of two most accurate methods (RFIDS.20 and RFOK.20) with slope, their averaging and IDS (the control) in the southwest region.

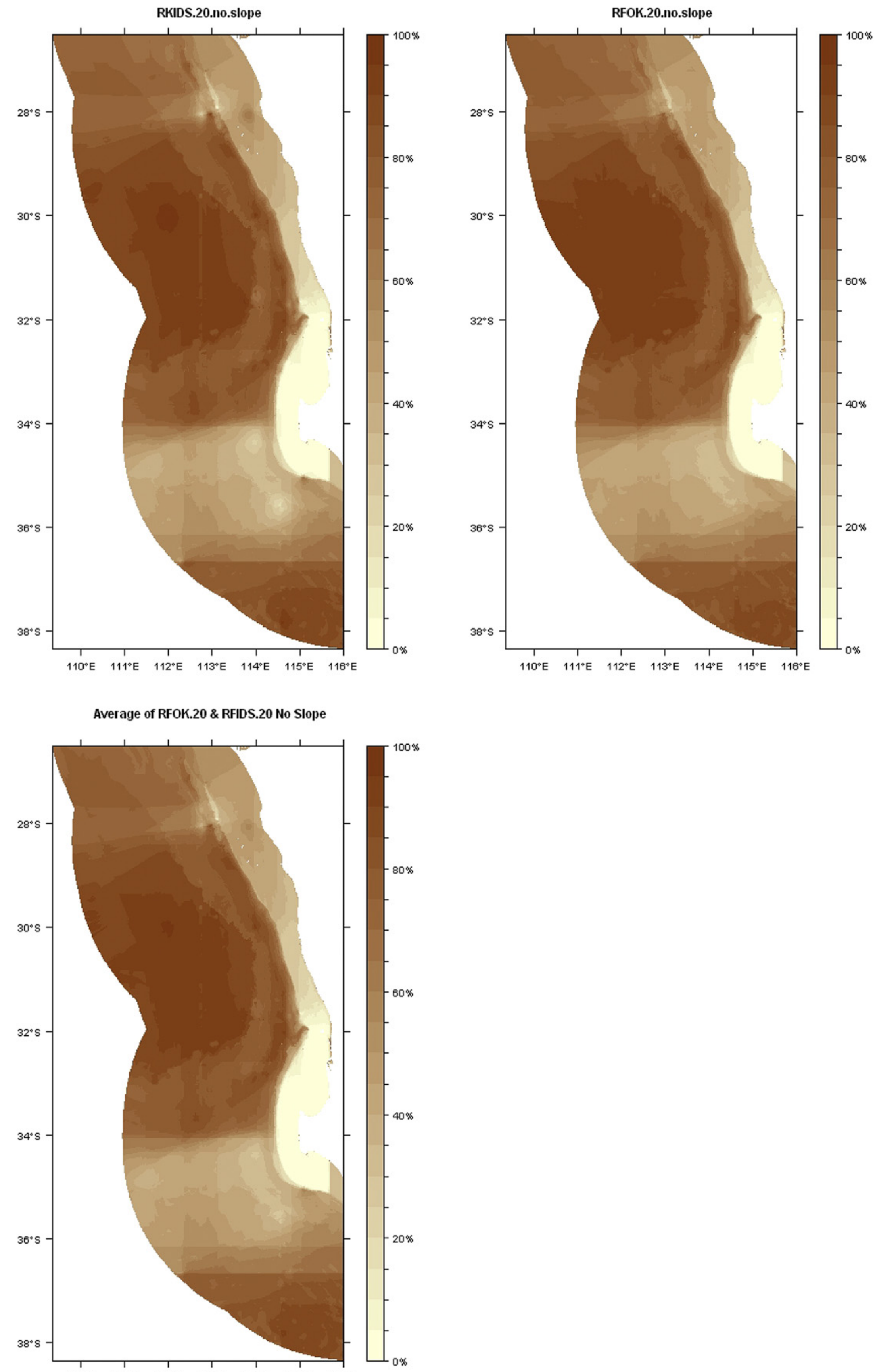

Fig. 4. The predictions of two most accurate methods (RFIDS.20 and RFOK.20) without slope and the predictions of their averaging in the southwest region.

similar between these two methods, and the weak 'bull's eye' patterns are evident for RFIDS (Fig. 4). However, when slope was excluded some details disappeared including some linear features and the effects of longitude and latitude are more prominent than those with slope. The average of these two methods produced a map similar to that of RFIDS, but the 'bull's eyes' patterns are slightly weakened.

4. Discussion

4.1. Can machine learning be used for spatial interpolation?

Overall, the predictive errors for spatial interpolation of mud content depend on the methods. On average, RFOK, RFIDS and averaging their predictions are the best performing methods and their accuracies are similar. This suggests that RF successfully predicted the general trend, and OK and IDS made similar contribution to the observed patterns of the combined methods. The most accurate method, RFOK, reduces RMAE relatively by up to 17% in comparison with the control (IDS). The accuracy of the two most accurate methods (i.e., RFOK.20 with or without slope in Table 4) is much higher than those reported in previous studies in environmental sciences (Li and Heap, 2011) (Fig. 5). The coefficient variation (CV) of the data in this study is 81%, and the RMAE of these two best methods was approximately 30% lower than that of the best methods reported in previous publications (Li and Heap, 2011), which demonstrates that the best methods identified in this study are far more reliable than those methods identified in previously published studies.

OK usually outperforms IDW and is superior at least in theory (Li and Heap, 2008). However, in this study OK performed similarly to IDW2, IDW3 and IDW4 or more poorly. A similar finding was also reported previously where optimal IDW (OIDW) was found to be superior over kriging when data were isotropic and the primary variable was not correlated with the secondary variable (Collins and Bolstad, 1996). The poor performance of OK in this study

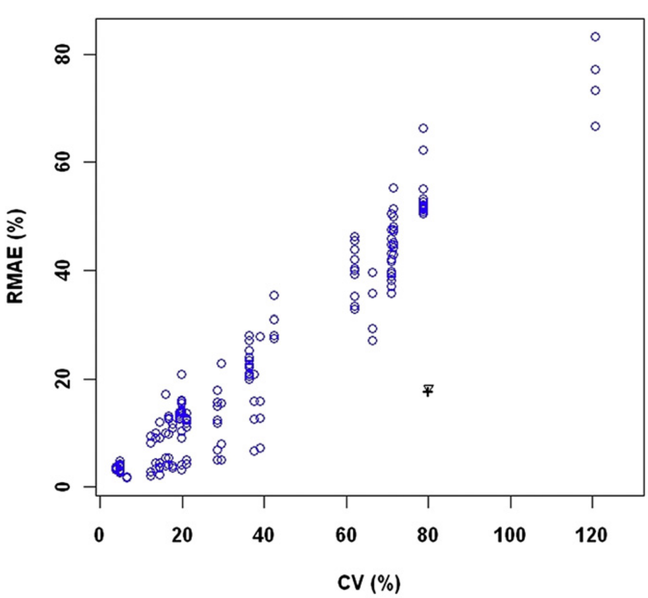

Fig. 5. The relative mean absolute error (RMAE (%)) of two most accurate methods (RFOK.20. with slope: triangle point down; RFOK.20.without slope: +) in relation to coefficient variation (CV (%)) in comparison with the results of previous studies (blue open circles) in a number of disciplines in environmental sciences (Li and Heap, 2011). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

could be attributed to: 1) the fact that the data were not equal distance projected; 2) the data stationarity required by OK was not fully satisfied although relevant transformation was employed; 3) the data transformation was selected based on full dataset this might not be the most appropriate for all of the sub-datasets; and 4) the type of variogram model selected was also based on the full dataset, which might also not be the most appropriate for all of the sub-datasets used for prediction. Selection of data transformation and of the type of variogram model for each sub-dataset was not practical, which is perhaps a disadvantage of the simulation automation.

A combined method of regression tree and OK has been shown to be more accurate than IDW and OK (Martínez-Cob, 1996), which is consistent with our finding in this study where RpartOK outperformed IDS.

SVM was found to be superior to RF in a previous study (Statnikov et al., 2008), but our results showed that the machine learning methods like SVM and KSVM performed more poorly than RF and were even less accurate than IDW in this study. This suggests that not all machine learning methods should be applied to spatial interpolation of environmental variables, mud content at least.

Comparing the accuracy of methods that use secondary information (e.g., KED, OCK, RK (except with RF), UK) and methods that do not use secondary information (e.g., IDW and OK) shows that methods using the secondary information are less accurate than those without using secondary information, despite of the fact that the correlation between the primary variable and the secondary variables are strong in this study, suggesting a strong spatial trend. This finding is not consistent with 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 argue 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). However, this still cannot explain the results observed in this study as the correlation coefficients are above this threshold. This finding suggests that the inclusion of secondary information does not always improve the prediction accuracy, which does not support the argument regarding the role of secondary variables in spatial interpolation (Hengl, 2007). This is probably because these methods assume a linear relationship between the primary and the secondary variables, but the primary variable is not linearly correlated to the secondary variable in this study (Li et al., 2010).

The high accuracy of RFOK could be attributed to the method itself and the high correlation of mud content to the secondary variables, because the accuracy of regression modeling 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 as discussed above. Therefore, it is the method itself (RF) that attributed to its superior performance observed and it can effectively predict the primary variable of a non-linear correlation with the secondary variables. The excellent performance of RF and its combined methods with OK or IDS is largely supported by the finding of Diaz-Uriarte and de Andres (Diaz-Uriarte and de Andres, 2006). The superior performance of RF, RFOK and RFIDS in this study may be attributed to the following factors associated with the methods.

- RF uses both bagging (bootstrap aggregation), a successful approach for combining unstable learners, and random variable selection for tree building; and it thus yields an ensemble of low correlation trees. So RF can achieve both low bias and low variance from averaging over a large ensemble of low bias, high variance but low correlation trees (Breiman, 1996; Breiman et al., 1984; Diaz-Uriarte and de Andres, 2006).

- It was even argued that the performance of RF is not much influenced by parameter choices (Diaz-Uriarte and de Andres, 2006; Liaw and Wiener, 2002; Okun and Priisalu, 2007).
- The predictions of RF are more reasonable for extrapolation and more accurate than regression tree (RT), bagging trees (BT) and multivariate adaptive regression splines (MARS) (Prasad et al., 2006). Extrapolation occurred in spatial predictions of mud content in this study.
- Although the unpruned individual trees will result in overfitted models (Prasad et al., 2006), our results of RF based on unpruned trees suggest that the overfitting of individual trees may not be a big concern to RF.
- RF does not overfit with respect to the source data because of the law of large numbers (Breiman, 2001; Maindonald and Xie, 2008; Okun and Priisalu, 2007).

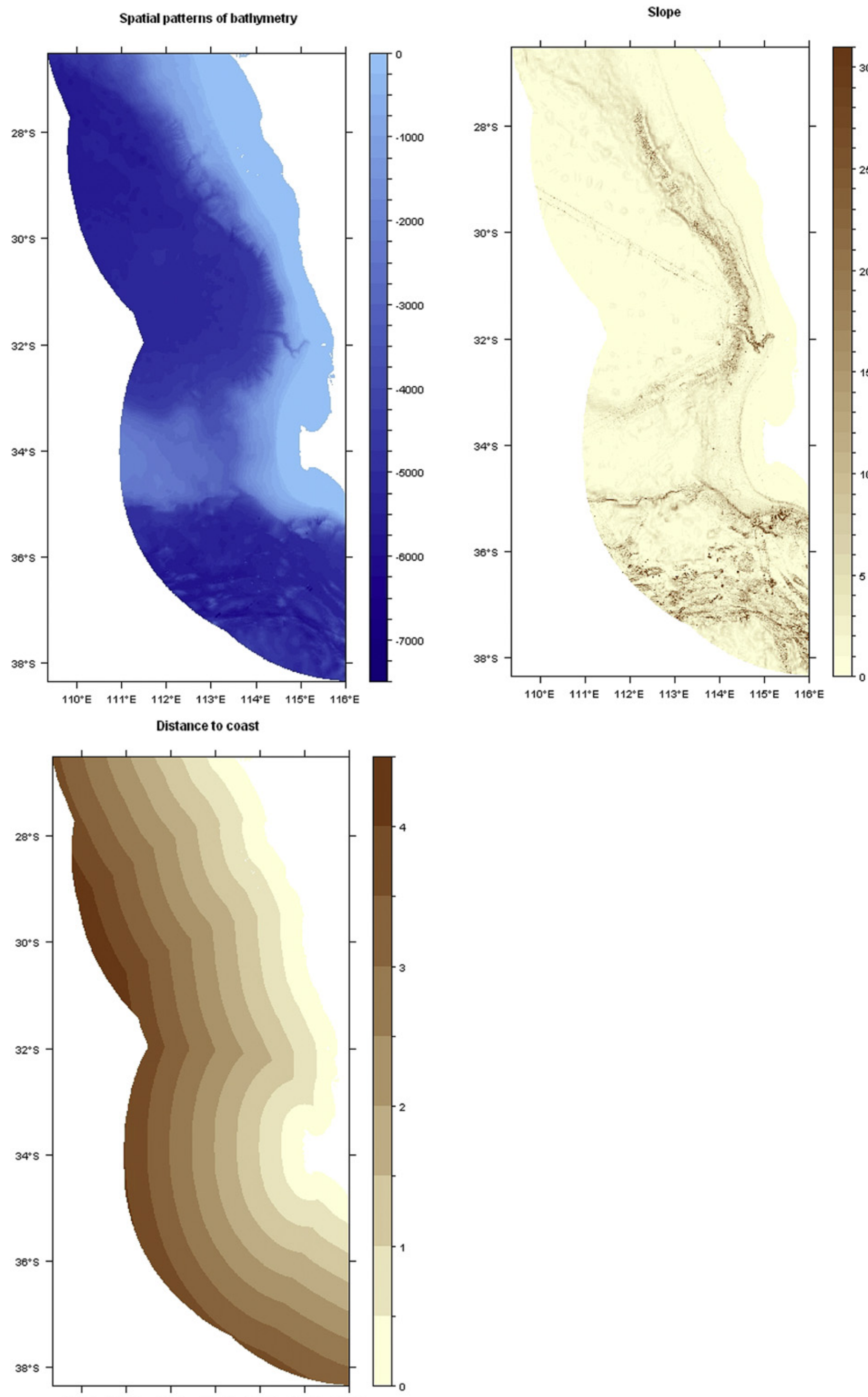

Fig. 6. Spatial pattern of bathymetry, slope and distance to coast in the southwest region. To display the patterns of slope in the majority area, values over 30 were converted to 31.

- RF is relatively robust to outliers and noise (Breiman, 2001).
- Although decision trees, neural networks, support vector machines are able to deal with poorly predictable data (i.e., primary variable) (Shan et al., 2006), they are outperformed by RF and its combined methods in both this study and our previous study (Li et al., 2010). This suggests that RF may be able to deal with poorly predictable data because the data used in this study are sparsely- and unevenly- spatially distributed, which is also supported by findings in a previous study (Marmion et al., 2009).
- Each tree is unpruned, so as to obtain low-bias trees; and it can also deliver good predictive performance even when most predictive variables are noise (Diaz-Uriarte and de Andres, 2006).
- RF can model complex interactions among predictive variables (Cutler et al., 2007; Diaz-Uriarte and de Andres, 2006; Okun and Priisalu, 2007).
- Finally, the prediction residuals of RF are interpolated using OK or IDS and the interpolated values are added to the predictions of random forest, which further reduces the prediction error as evidenced in this study.

4.2. Are noisy predictive variables a problem for RF?

RF selects the most important variable to split the samples at each node split for each individual trees, thus it is argued to implicitly perform variable selection (Okun and Priisalu, 2007). Consequently, it is not sensitive to non-important and many irrelevant variables. This statement is not supported by the results of RF and RFOK in this study. When the least important variables are excluded, the predictive errors have been significantly reduced for these two methods, which is supported by findings regarding feature selection for machine learning methods in a previous study (Guyon et al., 2009). This phenomenon could be explained by that the inclusion of noisy predictive variables can reduce the probability of the inclusion of good predictive variables at each node split when a portion of predictive variables are randomly selected for each individual tree, thus the chance of the contribution of good predictive variables to the tree is reduced. Consequently, the tree developed produces less accurate predictions. This suggests that pre-selection of variables for RF is an important step and should not be ignored in using random forest, although RF can also deliver

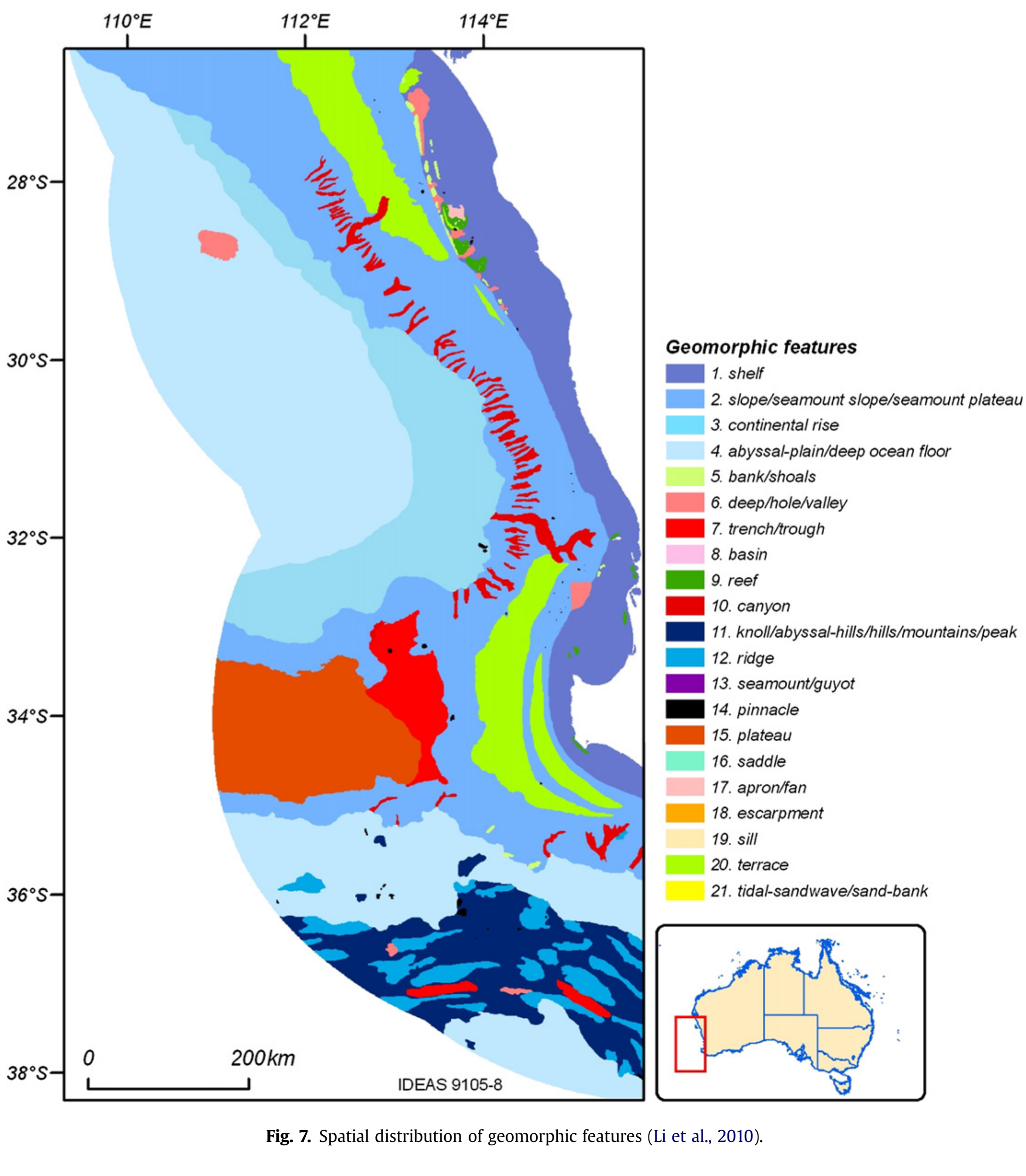

Fig. 7. Spatial distribution of geomorphic features (Li et al., 2010).

good predictive performance even when most predictive variables are noise (Diaz-Uriarte and de Andres, 2006).

4.3. Can model averaging reduce predictive error?

In this study, we simply averaged the predictions of two modeling methods to produce the final predictions. The model averaging had little influence on the prediction accuracy. This result does not support previous findings about ensemble mean approach (Goswami and O'Connor, 2007; Hoeting et al., 1999; Marmion et al., 2009; Raftery et al., 2005). Suggesting that model averaging may not necessarily always deliver improved predictions.

Model averaging has been applied to a number of disciplines (Araújo and New, 2007; Ellison, 2004; Goswami and O'Connor, 2007; Hoeting et al., 1999), but it has not been applied to geostatistics; and to our limited knowledge averaging the predictions of combined method of machine learning method and IDS or OK is the first in this discipline. Given the improved results of model averaging in other disciplines (Goswami and O'Connor, 2007; Hoeting et al., 1999; Marmion et al., 2009; Raftery et al., 2005), further tests on the effects of such model averaging should not be discouraged in future studies.

4.4. Visual examination of the predictions

4.4.1. With slope related variables

The spatial patterns of predictions of RFOK, RFIDS and their average, mainly reflect the effect of bathymetry and its related variables like geomorphic features as the patterns were similar to those identified by Heap and Harris (2008) (Figs. 6 and 7). This can be explained by the fact that bathymetry is the most important variable for random forest (Fig. 8). The influences of geomorphic features (shelf, slope, continental rise, abyssal plain, canyons and Naturaliste Plateau) (Heap and Harris, 2008) were obvious, despite the fact that the effects of latitude were also apparent in area between $-34°$ to $-37°$. On the shelf, predicted patterns of seabed mud content capture the relatively sandy nature of this geomorphic province and the accumulations of coarse sand and gravel around the shelf edge of Houtman-Abrolhos Reefs. In the deeper water, the influence of bathymetry and geomorphology is also apparent; the influence is greatest where the features are topographically

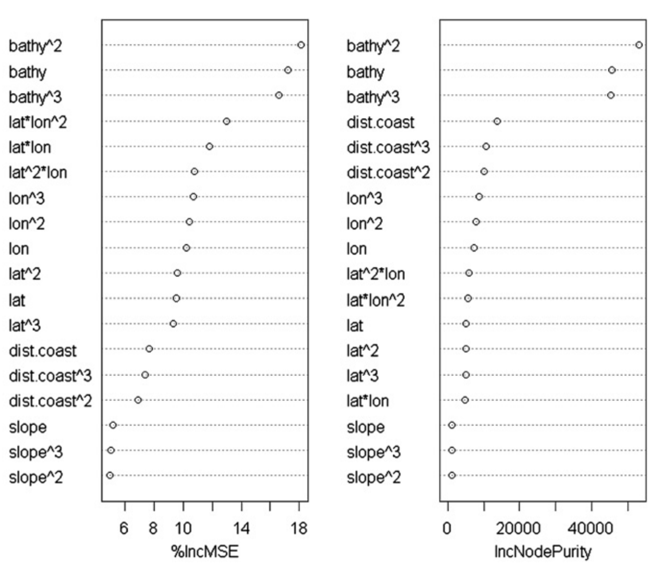

Fig. 8. Variable importance measured by random forest in the randomForest package in R (R Development Core Team, 2008).

prominent (e.g., Perth Canyon). The effects of geomorphic features were discussed in detail by Li et al. (2010). However, IDW2 failed to capture these patterns. Therefore, the predictions of RF, RFOK, RFIDS and their average are more realistic than those of IDW2 based on visual examination.

The influence of bathymetry is apparent because it was 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 (<1000 m) the canyon is slightly more muddy than the surrounding slope environments (Heap et al., 2008). The banding patterns reflect the influence of longitude and latitude because they were used as secondary information. This suggests that in these areas mud content was more related to longitude and latitude than other secondary variables including bathymetry. These artifacts derived from secondary information are prominent due to the relatively few samples collected from the deeper parts of the margin. 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 RFOK, RFIDS and their average are because geomorphic features were largely the categorized bathymetry (Heap and Harris, 2008).

The weak 'bull's eyes' patterns are evident for RFIDS which is typical phenomenon of IDW (Fig. 3), although it was only applied to the residuals of RF.

A prominent N–S trending ship-track feature and detailed small-scale patterns in the predicted spatial patterns of mud by RFOK, RFIDS and their average mainly reflect the influence of slope (Fig. 6), because after excluding slope from the models, such features are disappeared or considerably weakened. These features also reflect the transition between geomorphic features. While the contributions of distance to coast to the observed patterns are largely invisible, despite of the fact that it is more important than slope (Fig. 8).

4.4.2. Without slope related variables

In general, the spatial patterns are similar between the methods (i.e., RFIDS, RFOK and their average), but the weak 'bull's eyes' patterns are still evident for RFIDS. The predictions of RFOK and the averaged method are more visually pleasant than that of RFIDS. The predictions of the methods with the exclusion of slope showed band patterns with sharp transition and failed to pick up the changes in mud content associated with transitions between geomorphic features and promoted the effects of latitude and longitude, although some specific linear features associated slope are weakened or disappeared. Consequently, the predicted patterns are less visually pleasant.

Overall, RFIDS, RFOK and their averaged methods with slope are preferred visually to those without slope. Therefore, the inclusion of slope should be considered for producing the spatial predictions although it is the least important variable and its inclusion may slightly reduce the prediction accuracy. Visual examination proved to be an essential step in accessing the performance of prediction methods and is as important as the predictive error assessments, because two methods with similar prediction error may produce different spatial patterns as evidenced in this study.

4.5. Limitations

Data in this study was in latitude and longitude (i.e., WGS84), and not equal distance projected, which could reduce the accuracy of geostatistical methods and all other methods using latitude and longitude as secondary information. Since RFOK, the selected method, used the secondary information, its accuracy might also be affected. This limitation needs to be taken into account in assessing the performance of the methods using such secondary information or using variogram modeling. In this study, IDS is believed to be relatively less sensitive to projection issue due to its local neighborhood searching window. Since RFOK outperformed IDS, we can confidently conclude that the effects of projection issue on the performance of RFOK are limited.

Secondary information considered was only limited to bathymetry, distance-to-coast, slope, and geomorphic province. Other correlated secondary variables should be identified and employed and more samples need to be collected in future studies, and thus the artifacts associated with influence of latitude and longitude could be removed from the predictions.

The samples used in this and previous studies were collected over a period of time as described by Li et al. (2010), the temporal change in seabed sediments has not been considered in this study as all data at each sample location were assumed unchanged over time in the modeling. This assumption is obviously questionable, but due to the limitation in the availability of samples we have to take such assumption, otherwise it is impossible to make any predictions because of the very limited sample size collected at each survey. Despite of this limitation, the predictions are more accurate than those of the control method and the methods compared in previous studies in environmental sciences (Li and Heap, 2008). Therefore, predictions with improved accuracy are surely expected from this study. However, uncertainties associated with this limitation should be taken into account in applying the findings of this study and in using the subsequently produced predictions of mud content, although this limitation is shared by all methods including the control.

The direct back-transformation approach was adopted study, which may result in a biased estimation of the primary variable because similar phenomena have been reported for lognormal and square-root transformed data (Dambolena et al., 2009; Schuurmans et al., 2007; Yamamoto, 2007). However, for the transformations used in this study, the unbiased procedures of back-transformation are not readily available. Obviously, this is a field worth further research.

As we stated above that the existing spatial interpolation methods are often data- or even variable- specific and their performance depends on many factors (Li and Heap, 2011). This may also apply to the findings in this study. Since the combined methods developed in this study are novel, their performance is unknown for other data and variables. Since their success in this study, machine learning methods in general, and these combined methods in particular, warrant further test across different primary variables in future.

5. Conclusions

The application of machine learning methods like RF to spatial interpolation is novel. The combined method, RFOK, is also new, and its prediction error was relatively up to 19% less than the control (IDS). The prediction error (RMAE) of this method is also up to 30% lower than that of the best methods published in the environmental sciences, highlighting the superiority of this method. Since all methods were tested under the same conditions, and RF, RFOK and RFIDS are observed to be the most accurate methods, we can conclude that their superior performance is attributed to the methods themselves rather than to any other factors. In this study we only tested a few machine learning methods. Given the excellent performance of RF, other machine learning methods may worth testing in future studies.

This study confirmed the effectiveness of the combination of RF with OK or IDS and opened an alternative source of methods for the discipline of spatial statistics. The methods identified can be applied to the modeling of a range of both marine and terrestrial environmental variables. The combined approaches of machine learning methods and other existing spatial interpolation techniques provided a direction for future studies to select statistical methods for spatial interpolation.

The exclusion of least important variables for RF may increase the prediction accuracy. This suggests that it is necessary to test whether the exclusion of the least important variables improve the predictive accuracy for RF in future studies. This process should be stopped when no improvement is observed.

Averaging the best methods does not always improve the prediction accuracy. Given it is not computationally expensive, it is recommended for inclusion in future studies.

Visual examination proved to be an essential step in accessing the performance of prediction methods and is as important as the predictive error assessments, because methods with similar prediction error may produce different spatial patterns. Visual examination can display any abnormal patterns in the predictions.

This study provides suggestions for improving the spatial interpolations of environmental data, in general, and results in more accurate mapping and characterization of seabed mud content in the AEEZ in particular. The selection of the most robust method for given study region is an endless process as we may not have enough resources to test all possible combinations of impact factors. The 'best' methods identified are conditioned on the resources allocated, which really imply that what is the acceptable prediction error for a study area. This must be taken into account in making the final decision. To acquire the highest prediction accuracy, we would recommend that the most accurate method should be identified by considering all relevant impact factors, such as those factors recommended in previous studies (Li and Heap, 2008, 2011) and model diagnostics decisions suggested by Bivand et al. (2008), for each study area.

Acknowledgments

Christina Baker, Shoaib Burq, Mark Webster, and Tanya Whiteway are appreciated for preparing datasets. Chris Lawson and Zhi Huang are particularly appreciated for providing bathymetry, slope and distance to coast 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 Chief Executive Officer, Geoscience Australia.

References

Araújo, M.B., New, M., 2007. Ensemble forecasting of species distribution. TREE 22 (1), 42–47.

Arthur, A.D., Li, J., Henry, S., Cunningham, S.A., 2010. Influence of woody vegetation on pollinator densities in oilseed Brassica fields in an Australian temperate landscape. Basic Appl. Ecol. 11, 406–414.

Asli, M., Marcotte, D., 1995. Comparison of approaches to spatial estimation in a bivariate context. Math. Geol. 27, 641–658.

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

Breiman, L., 1996. Bagging predictors. Mach Learn. 24, 123–140.

Breiman, L., 2001. Random forests. Mach Learn. 45, 5–32.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth, Belmont.

Collins, F.C., Bolstad, P.V., 1996. A Comparison of Spatial Interpolation Techniques in Temperature Estimation. In: 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., Vapnik, V., 1995. Support-vector networks. Mach Learn. 20, 273–297.
Cutler, D.R., Edwards, T.C.J., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecography 88 (11), 2783–2792.
Dambolena, I.G., Eriksen, S.E., Kopcso, D.P., 2009. Logarithmic transformations in regression: do you transform back correctly? Primus 19 (3), 280–290.
Diaz-Uriarte, R., de Andres, S.A., 2006. Gene selection and classification of micro-array data using random forest. BMC Bioinform. 7 (3), 1–13.
Drake, J.M., Randin, C., Guisan, A., 2006. Modelling ecological niches with support vector machines. J. Appl. Ecol. 43, 424–432.
Ellison, A.M., 2004. Bayesian inference in ecology. Ecol. Lett. 7, 509–520.
Gilardi, N., 2002. Machine Learning for Spatial Data Analysis, p. 73.
Gilardi, N., Bengio, S., 2000. Local machine learning models for spatial data analysis. J. Geogr. Inf. Decis. Anal. 4 (1), 11–28.
Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York.
Goovaerts, P., 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. J. Hydrol. 228, 113–129.
Goswami, M., O'Connor, K.M., 2007. Real-time flow forecasting in the absence of quantitative precipitation forecasts: a multi-model approach. J. Hydrol 334, 125–140.
Gregory, A.W., Smith, G.W., Yetman, J., 2001. Testing for forecast consensus. J. Bus Econ. Stat. 19, 34–43.
Guyon, I., Lemaire, V., Boullé, M., Dror, G., Vogel, D., 2009. Analysis of the KDD Cup 2009: fast scoring on a large Orange customer database. In: Lawrence, N. (Ed.), JMLR: Workshop and Conference Proceedings, pp. 1–22.
Heap, A.D., Harris, P.T., 2008. Geomorphology of the Australian margin and adjacent seafloor. Aust. J. Earth Sci. 55, 555–585.
Heap, A.D., Edwards, J., Fountain, L., Spinnocia, M., Hughes, M., Mathews, E., Griffin, J., Borissova, I., Blevin, J., Mitchell, C., Krassay, A., 2008. Geomorphology, Sedimentology and Stratigraphy of Submarine Canyons on the SW Australian Slope. Post Survey Report. Geoscience Australia, Canberra, p. 138.
Hemer, M.A., 2006. The magnitude and frequency of combined flow bed shear stress as a measure of exposure on the Australian continental shelf. Cont Shelf Res. 26, 1258–1280.
Hengl, T., 2007. A Practical Guide to Geostatistical Mapping of Environmental Variables. Office for Official Publication of the European Communities, Luxembourg, p. 143.
Hengl, T., Heuvelink, G.B.M., Rossiter, D.G., 2007. About regression-kriging: from equations to case studies. Comput. Geosci. 33, 1301–1315.
Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: a tutorial. Stat. Sci. 14 (4), 382–417.
Legendre, P., Legendre, L., 1998. Numerical Ecology, second ed.. Elsevier, Amsterdam.
Li, J., Heap, A., 2008. A Review of Spatial Interpolation Methods for Environmental Scientists. Geoscience Australia, Canberra, p. 137.
Li, J., Heap, A., 2011. A review of comparative studies of spatial interpolation methods: performance and impact factors. Ecol. Inform., 228–241.
Li, J., Potter, A., Huang, Z., Daniell, J.J., Heap, A., 2010. Predicting Seabed Mud Content Across the Australian Margin: Comparison of Statistical and Mathematical Techniques Using a Simulation Experiment. Geoscience Australia, Canberra, p. 146.
Liaw, A., Wiener, M., 2002. Classification and regression by radomForest. R News 2 (3), 18–22.
Maindonald, J., Xie, Y., 2008. Data Mining with R. http://cos.name/wp-content/uploads/2008/12/data-mining-with-r-by-john-maindonald.pdf.
Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K., Thuiller, W., 2009. Evaluation of consensus methods in predictive species distribution modelling. Divers. Distrib 15, 59–69.
Martínez-Cob, A., 1996. Multivariate geostatistical analysis of evapotranspiration and precipitation in mountainous terrain. J. Hydrol. 174, 19–35.
Nilsson, J., Persson, B., von Heijne, G., 2000. Consensus prediction of membrane protein topology. FEBS Lett. 486, 267–269.
Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma 67, 215–226.
Okun, O., Priisalu, H., 2007. Random forest for gene expression based cancer classification: overlooked issues. In: Martí, J., Benedí, J.M., Mendonça, A.M., Serrat, J. (Eds.), Pattern Recognition and Image Analysis: Third Iberian Conference. IbPRIA 2007 Lecture Notes in Computer Science, Girona, Spain, pp. 4478, 4483–4490.
Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci. 30, 683–691.
Pinheiro, J.C., Bates, D.M., 2000. Mixed-Effects Models in S and S-PLUS. Springer, New York.
Pitcher, C.R., Doherty, P.J., Anderson, T.J., 2008. Seabed environments, habitats and biological assemblages. In: Hutchings, P., Kingsford, M., Hoegh-Guldberg, O. (Eds.), The Great Barrier Reef: Biology, Environment and Management. CSIRO Publishing, Collingwood, p. 377.
Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosyst 9, 181–199.
R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.
Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian model averaging to calibrate forecast ensembles. Mon Weather Rev. 133, 1155–1174.
Schuurmans, J.M., Bierkens, M.F.P., Pebesma, E.J., 2007. Automatic prediction of high-resolution daily rainfall fields for multiple extents: the potential of operational radar. J. Hydrometeorol 8, 1204–1224.
Shan, Y., Paull, D., McKay, R.I., 2006. Machine learning of poorly predictable ecological data. Ecol. Modell. 195, 129–138.
Statnikov, A., Wang, L., Aliferis, C.F., 2008. A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. BMC Bioinform. 9, 319.
Stein, A., Hoogerwerf, M., 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., Hothorn, T., 2007. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinform. 8 (25), 25.
Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S-Plus, fourth ed. Springer-Verlag, New York.
Verfaillie, E., van Lancker, V., van Meirvenne, M., 2006. Multivariate geostatistics for the predictive modelling of the surficial sand distribution in shelf seas. Cont Shelf Res. 26, 2454–2468.
Voltz, M., Webster, R., 1990. A comparison of kriging, cubic splines and classification for predicting soil properties from sample information. J. Soil Sci. 41, 473–490.
Whiteway, T., Heap, A., Lucieer, V., Hinde, A., Ruddick, R., Harris, P.T., 2007. Seascapes of the Australian Margin and Adjacent Sea Floor: Methodology and Results. Geoscience Australia, p. 133.
Yamamoto, J.K., 2007. On unbiased backtransform of lognormal kriging estimates. Comput. Geosci. 11, 219–234.