Natural Resources Research, Vol. 20, No. 1, March 2011 (© 2011)
DOI: 10.1007/s11053-010-9132-y

Multiple Regressive Pattern Recognition Technique: An Adapted Approach for Improved Georesource Estimation

Thomas Oommen,[1] Debasmita Misra,[2,5] Anupma Prakash,[3] Sukumar Bandopadhyay,[2] Sathy Naidu,[4] and John J. Kelley[4]

Received 28 July 2010; accepted 18 December 2010
Published online: 14 January 2011

Multiple Regressive Pattern Recognition Technique (MRPRT) is an adapted approach for improved geologic resource estimation. We developed and tested this approach for the Platinum (Pt) bearing region near Goodnews Bay, Alaska, which presents an example of a complex depositional environment. We applied geospatial and pattern recognition methods to assess the spatial distribution of offshore Pt in the Goodnews Bay area from point data collected by various agencies. We used the coefficient of correlation (r) and the Nash–Sutcliffe efficiency (E) to quantitatively assess the degree of accuracy of the estimated Pt distribution. We split the study area, based on trend analysis, into two regions: inside the Bay and outside the Bay. We could not obtain appreciable estimates from the geospatial and pattern recognition methods. Using MRPRT, we were able to improve r from 0.57 to 0.93 and the E from 28.31 to 92.91 inside the Bay. We achieved improvement in r from 0.55 to 0.61 and E from 28.46 to 34.52 outside the Bay. The reasons for a non-significant improvement outside the Bay have been discussed. The results indicate that the proposed MRPRT has wide application potential in georesource estimation where input data is often scarce.

KEY WORDS: Pattern recognition, geology, mining industry, spatial correlation, support vector machine.

INTRODUCTION

The successful development of an offshore mining project requires reliable assessment of the spatial distribution of the mineral resource. The traditional methods of assessing the spatial distribution are mostly empirical in nature and composed of block (triangular, polygonal, and irregular) or cross-sectional (vertical, horizontal and inclined) methods. The advancement in computational capabilities introduced more sophisticated geostatistical assessment techniques such as the use of inverse distance weighting (IDW), global polynomial interpolation (GPI), local polynomial interpolation (LPI), ordinary kriging (OK), radial basis function (RBF) using thin-plate spline, and co-kriging (CK). Several studies have demonstrated the ability of these methods to produce estimates with minimum error variance (Barker and Lamal, 1989; Oommen, 2006). However, assessing the spatial distribution of mineral deposits such as marine placers that result

[1]Department of Geological and Mining Engineering and Sciences, Michigan Technological University, Houghton MI, USA.
[2]Department of Mining and Geological Engineering, University of Alaska Fairbanks, PO Box 755800, Fairbanks AK 99775-5800, USA.
[3]Geophysical Institute, University of Alaska Fairbanks, Fairbanks AK, USA.
[4]Institute of Marine Science, University of Alaska Fairbanks, Fairbanks AK, USA.
[5]To whom correspondence should be addressed; e-mail: debu.misra@alaska.edu

from complex non-linear processes requires the use of non-linear estimation prediction/modeling techniques to improve the confidence in estimation.

Recent research shows that pattern recognition algorithms have the potential for assessing the spatial distribution of mineral deposits resulting from complex processes (Kanevski and others, 2002; Pozdnoukhov, 2005; Dutta, 2006; Twarakavi, Misra, and Bandopadhyay, 2006; Misra and others, 2007). The ability of these algorithms to learn the complex non-linear relationship depends on both the quality and appropriate number/quantity of the input parameters. In the absence of sufficient input parameters that describe the complex underlying depositional processes, even the performance of pattern recognition is compromised, as the algorithms are limited to the use of just the geographic location as the input to model the spatial distribution of the deposit. In this study, we propose and explore a new approach called the Multiple Regressive Pattern Recognition Technique (MRPRT), which has a predictive capability better than an individual geospatial or pattern recognition technique and under certain conditions, overcomes the necessity of having all input parameters that define a complex process being modeled.

WHAT IS MRPRT?

The MRPRT is an ensemble technique that combines the output of several regression techniques. The hypothesis for the MRPRT approach is that if individual regression techniques used to model, e.g., a mineral deposit, make different assumptions leading to complimentary decisions (outputs), then an intelligent combination of these outputs using a pattern recognition algorithm, such as the support vector machines (SVM) or the relevant vector machines (RVM), will provide an improved assessment compared to the individual techniques.

The hypothesis is further illustrated with synthetic data displayed in Figure 1. The x and y axes of the scatter plot represent the observed and predicted values of three models (three individual regression techniques). The coefficients of correlation for the Models 1, 2, and 3 are 0.28, 0.09, and 0.10, respectively. It is evident from Figure 1 that each of the three models individually has overall poor estimative capability.

However, Figure 1 shows that the Model-1 is able to better estimate the low values, Model-2 the median values, and Model-3 the high values; indicating that the decisions of the three models are complimentary to each other. The strengths of these individual models can be combined intelligently using a pattern recognition algorithm to develop an improved model compared to the individual models. In this study, we illustrate the applicability of this new approach for modeling the marine placer Platinum (Pt) deposits of Goodnews Bay and vicinity, in southwest Alaska, which represents an example of a complex mineral deposit. A more detailed discussion on the MRPRT approach to our application data is provided in "The MRPRT Approach" section of this article.

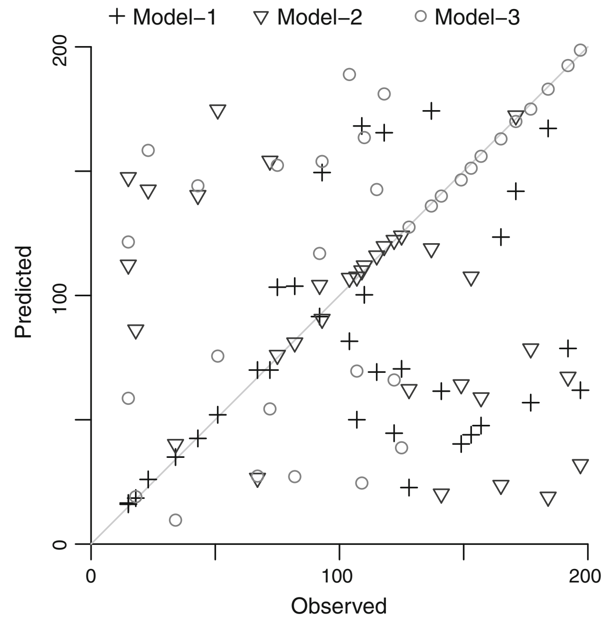

Figure 1. Scatter plot showing the observed vs. predicted values for the three hypothetical data sets represented as models 1, 2, and 3. Model-1 predicts the low values accurately, Model-2 the median, and Model-3 the high values.

STUDY AREA, BACKGROUND, AND DATA USED FOR THE APPLICATION

Platinum resources in the United States (US) are very limited. One of the largest resources is the alluvial placer deposit in the Red Mountain region of Goodnews Bay, in southwest Alaska (Fig. 2). Onshore mining in this region has recovered about 22 tons of Pt (Barker, 1986). The Red Mountain dunite is considered the primary source of Pt in this

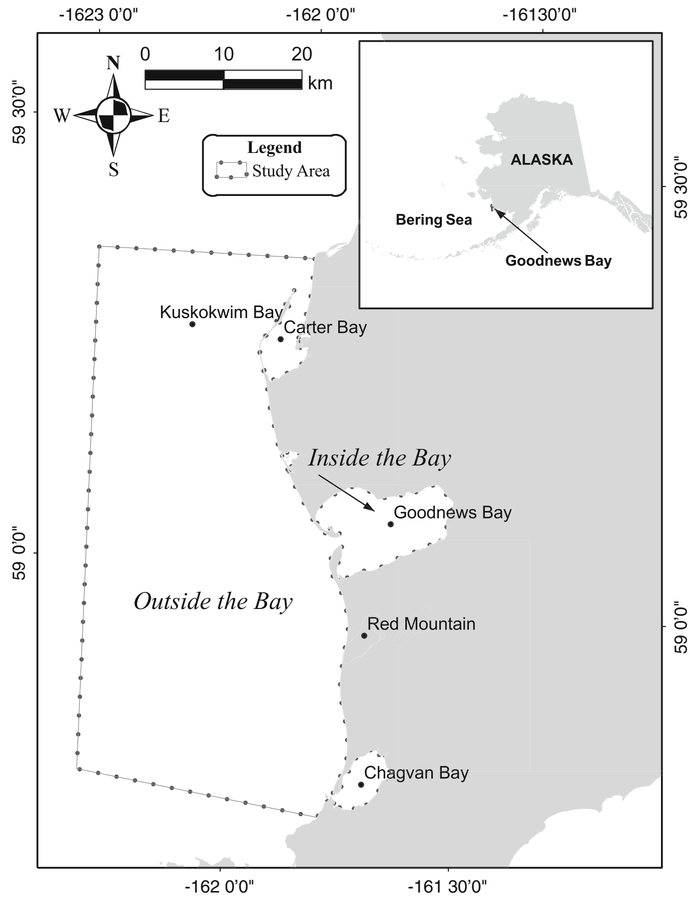

Figure 2. Map showing the study area boundary and the Red Mountain, the primary source of Pt in this region. Inset shows the location of Goodnews Bay in southwest Alaska.

region (Harrington, 1919). However, owing to the depletion of the shallow economic reserves, onshore mining has been discontinued since the 1980s.

The proximity of the Red Mountain to the Bering Sea, the drowned ultramafic rocks and paleo-drainage channels, identified offshore using geophysical survey methods (Barker and Lamal, 1989; Oommen, 2006), and the limited point data for Pt collected by various agencies indicate that the region holds promise for further prospecting.

The limited point data available from this region are from the studies by Inlet Oil Corporation (IOC) and the USGS in the late 1960s and early 1970s, the United States Bureau of Mines (USBM) and the Western Gold Exploration and Mining Limited Partnership (WestGold) in the late 1980s, and the University of Alaska Fairbanks in 2005 (Oommen and others, 2008b). There are a total of 477 offshore point sample locations for which Pt concentrations are known. Out of the 477 data points, 77 samples are collected from inside the (Goodnews) Bay, and the remaining 400 samples are from outside the Bay. The data used for the study are available from http://mms-goodnewsBay.gina.alaska.edu/index.phtml. A detailed discussion on the sampling technique and the geochemical analysis used to determine the Pt concentrations from these point locations is described in Oommen (2006) and Oommen and others (2008b).

The global trend analysis of the entire data set revealed that the data from inside the Bay were free from any spatial trend, while the data from outside the Bay depicted a distinct trend. The difference in these trends emanates from the difference in depositional environments of the regions (Oommen, 2006). Hence, for this study, the data from inside the Bay and those from outside the Bay were analyzed separately.

INITIAL ESTIMATION OF PLATINUM DISTRIBUTION USING GEOSPATIAL AND PATTERN RECOGNITION TECHNIQUES

The spatial distribution of Pt within the study area was first estimated using both classical geospatial techniques and pattern recognition algorithms. The classical geospatial techniques used included IDW, GPI, LPI, RBF, OK, and CK. The input for each method used was the Latitude and Longitude of each point sample location, except for CK where the concentration of Gold (Au) at each sample point was used as an additional input to develop the model for estimating the distribution of Pt concentration. The individual geospatial estimation models were developed using ArcGIS 9.1 geospatial analyst. Besides these geospatial methods, two pattern recognition algorithms, SVM and RVM, were used to develop models for estimation of Pt distribution in the study area. The e1071 and kernlab packages in the R-programming language were used for the pattern recognition algorithms (Karatzoglou, Meyer, and Hornik, 2006; Meyer, 2001). An explanation of the basic principles and the equations used for the various geospatial and pattern recognition techniques is beyond the scope of this article and are well described in the literature (Isaaks and Srivastava, 1989; Vapnik, 1995; Clark and Harper, 2000; Tipping, 2001; Johnston, 2003). The parameters used for these models are also explained in detail in Oommen (2006).

A major measure of the success of any modeling technique is its reliability (estimative capability). Perhaps the most rigorous way to analyze the reliability of a model is to collect additional data and compare those to the model estimations. This process is expensive and subject to logistical limitations in data collection. The alternatives to additional data collections are to use (1) the same data for both training/developing the model and testing, (2) *K*-fold cross validation, or (3) a subset of the data for training and the other subset for testing.

When the model is being developed and tested on the same data, the geospatial and the pattern recognition algorithms suffer from the curse of overfitting, which may allow accurate modeling of the training data, but when verified on an independent testing set, the performance of the model may be extremely poor (Foody and Mathur, 2004; Oommen and others, 2008a). *K*-Fold cross validation is considered to be the most reliable approach; however, it is tedious and expensive (Oommen and Baise, 2010). The use of a subset of the actual data for testing, which was not used for developing/training the model is much more cost effective and provides comparable results to the *K*-fold cross validation in the case of a sufficiently large sample size.

In this study, the reliability of the various models developed using the classical geospatial techniques and the pattern recognition algorithms were evaluated using an independent subset of the data, which was not used for the model development. To achieve this, the available data were split into two random subsets of 80% and 20% before the model development. An 80–20 split was selected based on prior research by Oommen and others (2008a, b) where those authors have shown such a split to be the best in the evaluation of a pattern recognition model. The larger subset was used for developing the models (training data, 80%) while the smaller subset was used to verify the reliability of each model (testing data, 20%). In the case of MRPRT, we additionally verified the predictive capability of the technique for data points inside and outside the Bay using *K*-fold ($K = 5$) cross validation. The additional evaluation using *K*-fold cross validation was performed to verify the reliability of the approach of using an independent subset for training and testing in the case of small samples.

The training of the geospatial and the pattern recognition algorithm models were carried out on the same data set. The optimization of the model parameters for the different techniques were achieved by *K*-fold ($K = 5$) cross validation of the training data. The estimative capability of the various models were compared using the coefficient of correlation (r) (Eq. 1) and coefficient of efficiency (E) (Eq. 2) (Nash and Sutcliffe, 1970).

$$r = \frac{\sum_{i=1}^{N}\left\{\left(O_i - \bar{O}\right)\left(P_i - \bar{P}\right)\right\}}{\left\{\sum_{i=1}^{N}\left(O_i - \bar{O}\right)^2 \sum_{i=1}^{N}\left(P_i - \bar{P}\right)^2\right\}^{0.5}} \quad (1)$$

$$E = \left\{1 - \frac{\sum_{i=1}^{N}\left(O_i - P_i\right)^2}{\sum_{i=1}^{N}\left(O_i - \overline{O}\right)^2}\right\} \times 100 \quad (2)$$

where E, the Nash–Sutcliffe efficiency, varies from $-\infty$ to 100% with values closer to 100% indicating better agreement between the observed and estimated values (better model). When E is less than 0, it indicates that the arithmetic mean of the observed value is a better estimator than the developed model. The same modeling procedures using the geospatial and pattern recognition algorithms were repeated for both Pt data samples inside and outside the Bay.

OUTCOME OF GEOSPATIAL AND PATTERN RECOGNITION MODELS

Tables 1 and 2 present the estimative capability of the various geospatial and pattern recognition algorithms validated using the testing data for sample points inside and outside the Bay. Table 1 shows that of all the different techniques used to analyze the spatial distribution of Pt inside the Bay, the RBF has the highest *r* and *E* (0.57 and 28.31). Three techniques (GPI, LPI and OK) have *E* values less than 0, which indicates that the arithmetic average of the observed value is a better estimator than these techniques. Table 2 shows that for sample points outside the Bay, the SVM model has the highest *r* and *E* values (0.56 and 28.46).

The *r* and *E* values listed in Tables 1 and 2 for the sample points both inside and outside the Bay reveal that none of the techniques is able to reliably estimate the entire spatial variability of the Pt concentrations within the study area. The reason for the failure of these techniques could be simply due to the limitation of the method used or the method's inability to capture the complexity of the factors governing the accumulation and redistribution of marine Pt. The factors that govern the accumulation and redistribution of marine Pt in the offshore region of Goodnews Bay include, but are not limited to, the general geology, glaciation history, littoral currents, bathymetry, sea-level transgressions and regressions, and paleo-drainage processes (Oommen and others, 2008b). The historic records of some of these processes are incomplete and, the outcome of the processes is difficult to quantify.

Table 1. The Predictive Capability of the Various Geospatial and Pattern Recognition Algorithms Verified on the Testing Data for the Pt Samples Inside the Bay Using *r* and *E*

Models	*r*	*E* (%)
IDW	0.48	21.07
GPI	0.36	−137.86
LPI	0.37	−121.58
RBF	**0.57**	**28.31**
OK	−0.18	−25.13
CK–Pt–Au	0.32	10.23
SVM	0.53	26.37
RVM	0.50	6.74

The model with best results (RBF) is shown in bold

Table 2. The Predictive Capability of the Various Geospatial and Pattern Recognition Algorithms Verified on the Testing Data for the Pt Samples Outside the Bay Using *r* and *E*

Models	*r*	*E* (%)
IDW	0.53	13.76
GPI	0.52	24.48
LPI	0.46	18.38
RBF	0.55	27.11
OK	0.50	22.87
CK–Pt–Au	0.48	8.52
SVM	**0.56**	**28.46**
RVM	0.53	23.30

The model with best results (SVM) is shown in bold

Since none of these techniques was able to reliably estimate the entire spatial variability in the Pt concentrations, we performed a split analysis to verify if any of these techniques have a better capability of estimation within a particular range or subset of Pt concentrations. The training and testing estimations were combined, sorted, and divided into quartiles. Further, the *r* of each of the techniques was analyzed separately within the first quartile (25th percentile), second quartile (25th–50th percentile), third quartile (50th–75th percentile), and the fourth quartile (75th–100th percentile).

The variability in *r* within the specified ranges for the sample points both inside and outside the Bay was assessed using the box plot (Fig. 3). The *r* values for the different techniques vary largely in the case of the sample points inside the Bay compared to those outside the Bay (Fig. 3a, b). The larger variation in *r* values for the Pt values inside the Bay (Fig. 3a) indicate that for the given data, the different techniques have different assumptions leading to estimations that are complimentary (as illustrated in the hypothetical example of Fig. 1). While, for the values outside the Bay, the low variation in *r* (Fig. 3b) indicates that, for the given data, all the techniques have similar assumptions leading to similar estimation capability (non-complimentary).

The complimentary nature of the different techniques in predicting the Pt values inside the Bay is also presented using a scatter plot in Figure 4. It is observed that the GPI (Fig. 4a), and OK (Fig. 4c) better predict the low and high values whereas each method contributed differently to the prediction of median values. A similar comparison of the predictive capability of the different techniques in

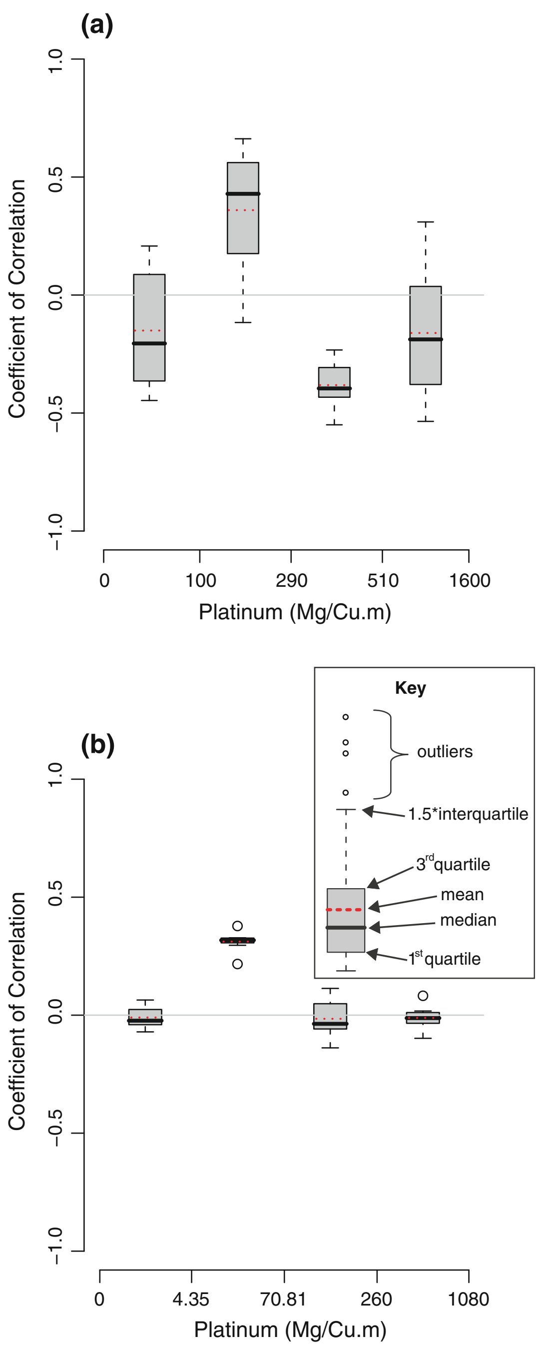

Figure 3. Box plot illustrating the variability in the coefficient of correlation within the quartiles. (a) Inside the Bay; (b) outside the Bay.

predicting the Pt values outside the Bay is presented in Figure 5. It is observed from Figure 5a–c that most of the techniques have very similar predictive capability, which is contrary to what was observed in Figure 4, and that none of the techniques have any reasonable predictive power for Pt concentration values >400 mg/m^3.

It is evident from these observations (Figs. 3, 4) that each technique has its own advantages and limitations. An important question that ensued from the observations obtained from split analysis (Fig. 3) was, "Could the strengths of each of these techniques be captured to obtain an improved estimate of the spatial distribution of Pt?" Ancillary questions such as, "Should the split analysis be carried out on quartiles, deciles, vingtiles or percentiles, or should it be done intelligently on a case specific basis?" and "At what scale should these strengths be captured?" also stemmed from the observation. In this study, we limit our discussion to addressing the first question only, which was the basis of the development of MRPRT approach. As discussed before, outcomes of models with complimentary estimation capabilities (e.g., the sample point estimates of data from inside the Bay) could be efficiently combined to create a robust estimator using a pattern learning algorithm.

In the case of data classification problems, there have been several approaches to combine classifiers (outputs of individual classifiers), which have provided improved performance [e.g., ensemble classifiers (Drucker and others, 1994), multiple classifier system (Kittler and others, 1998; Fumera and Roli, 2005), mixture of experts (Gutta and others, 2000), committees of neural networks (Chee and Harrison, 2003), neuro-fuzzy fusion (Meher and others, 2006), etc.]. These classifier combinations are very popular, and are considered to be one of the most promising current research directions in pattern recognition and machine learning (Kittler and Roli, 2001; Kuncheva, 2004). In the case of pattern recognition techniques, Wolpert (1992) introduced stacked generalization where the biases of multiple pattern recognition techniques with respect to a training set was reduced by generalizing in a second space where the inputs are the outputs of the original space. Stacked generalization is a powerful approach to combine multiple techniques and has not been widely used other than in the machine learning community. In the case of regression problems, ensemble techniques such as multiple linear regression to obtain the final outputs have been demonstrated (e.g., Hansen and Salamon, 1990; Perrone and Cooper, 1993; Jacobs, 1995; Dutta and others, 2006). We utilized a similar approach by combining complimentary estimations of various geospatial and pattern recognition models to form the basis of the MRPRT application as discussed in

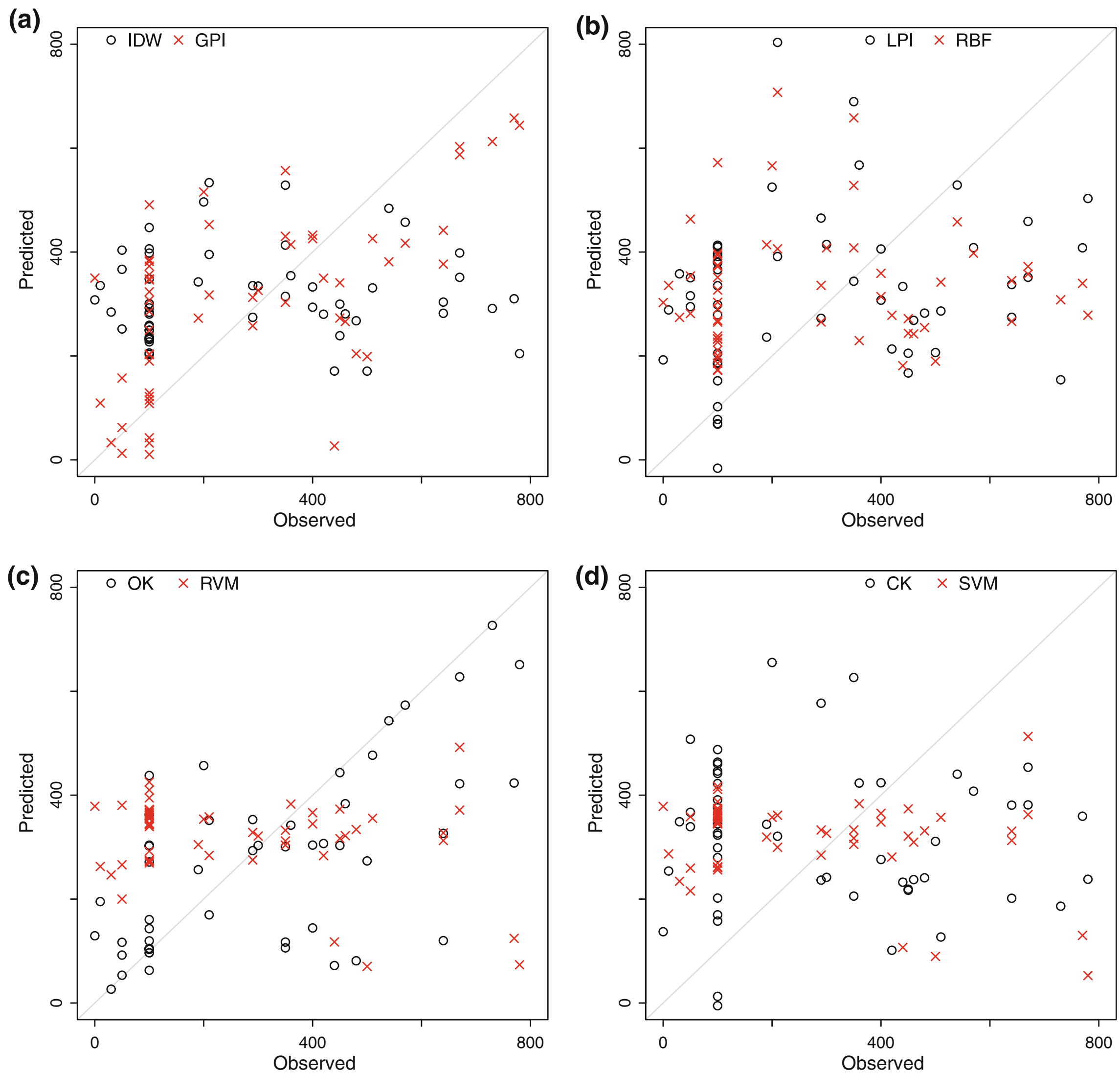

Figure 4. Observed vs. predicted Pt concentration values in mg/m^3 using the individual regression technique for Pt concentration values inside the Bay.

the next section. However, comparison of the MRPRT method with other contemporary methods discussed here is beyond the scope of this article.

THE MRPRT APPROACH

The MRPRT approach to model the Pt sample data obtained from inside and outside the Goodnews Bay is summarized below:

1. In the case of the sample points inside the Bay, eight input variables were used for the final estimation of Pt concentrations using the MRPRT model. The workflow of the MRPRT approach considering the sample points inside the Bay is presented in Figure 6. These input variables were the estimated Pt concentrations obtained using IDW, GPI, LPI, RBF, OK, CK, SVM, and RVM. However, in the case of sample points outside the Bay, three input variables gave the best improved estimation using the MRPRT. These input variables were the estimated Pt concentrations using IDW, RBF, and SVM. The reason for the fewer input variables

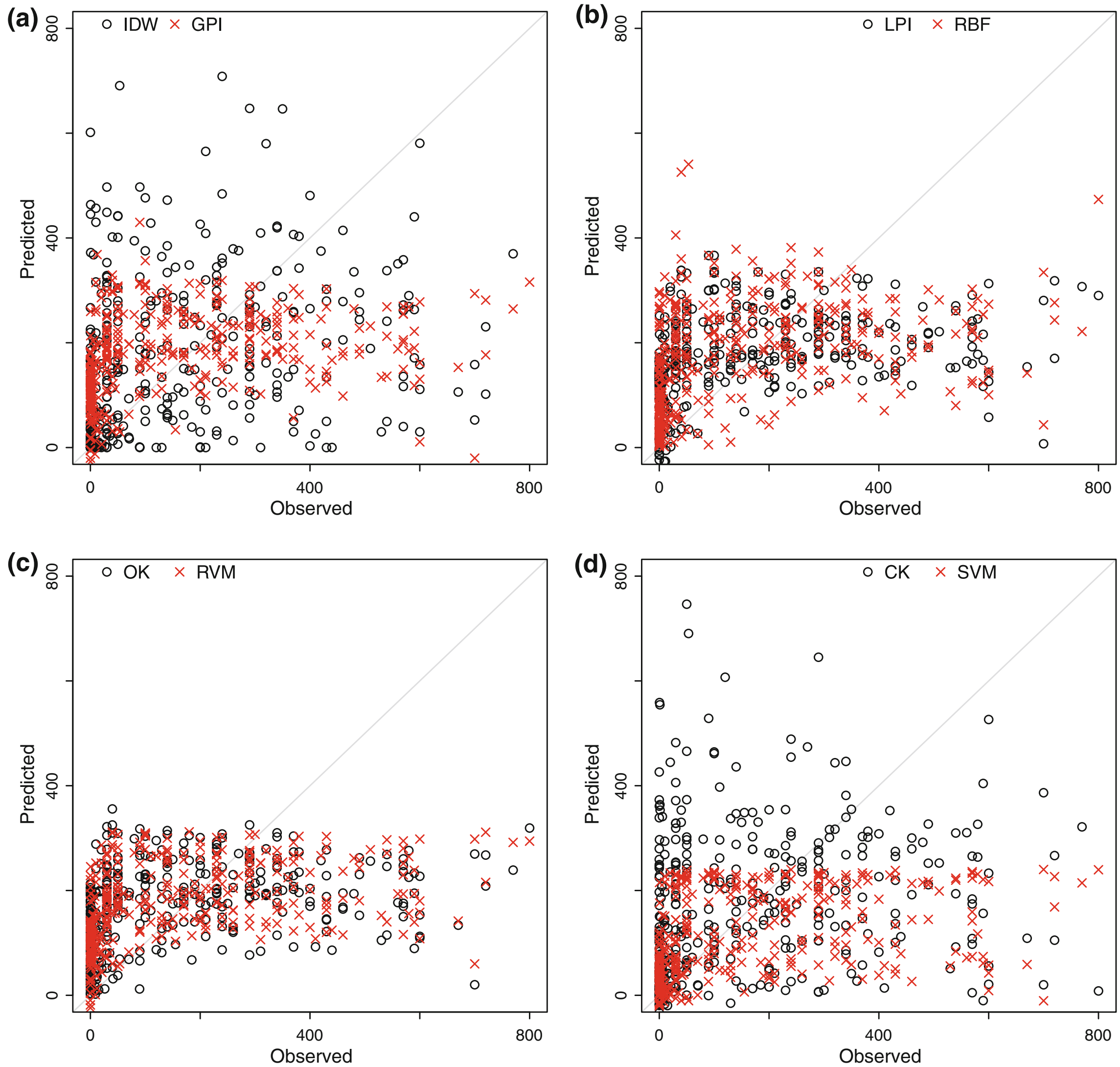

Figure 5. Observed vs. predicted Pt concentration values in mg/m^3 using the individual regression technique for Pt concentration values outside the Bay.

improving the MRPRT model outside the Bay is the non-complimentary nature of the different individual regression techniques (Figs. 3b, 5) which leads to data redundancy when all of them are used together.

2. In both cases (sample points inside and outside the Bay), 80% of the estimated data from the individual regression models were used for training and the remaining 20% for testing. In the case of data from inside the Bay, the reliability of evaluating the predictive capability using an independent subset for training and testing in the case of small samples was verified using K-fold ($K = 5$) cross validation.
3. Two pattern learning algorithms, the SVM (MRPRT–SVM) and the RVM (MRPRT–RVM), were used to verify the MRPRT.
4. In the case of MRPRT–SVM, when the radial basis kernel was used, there were three parameters to be optimized during training: they were the Gaussian radial basis function parameter γ, the magnitude of penalty term C, and the width of the error margin ε (Hastie and Friedman, 2003). However, in the case of MRPRT–RVM, the only parameter

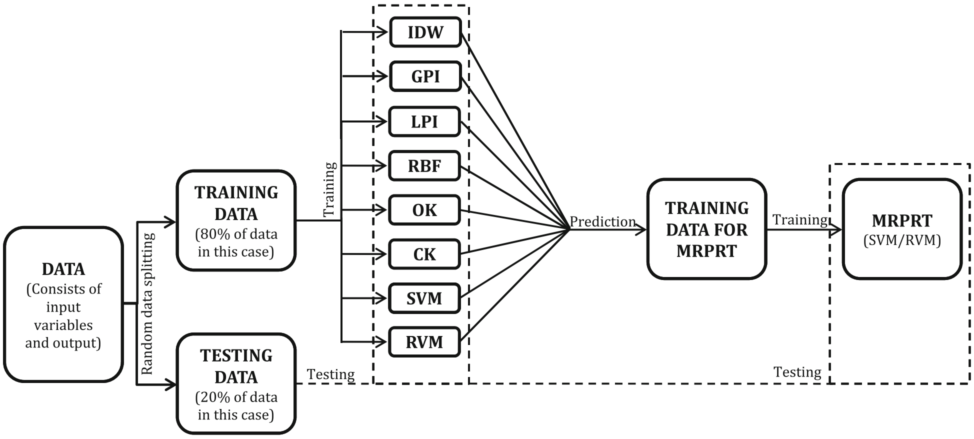

Figure 6. Workflow of the MRPRT approach considering the sample points inside the Bay.

to be optimized was the $\sum$, which was the inverse kernel width for the radial basis kernel function (Tipping, 2001).

5. The training data were used to obtain the optimal MRPRT–SVM and MRPRT–RVM parameters by K-fold cross validation ($K = 5$) and evaluated using E and r. The optimal MRPRT–SVM parameters obtained for the sample points inside the Bay were $\gamma = 0.001$, $C = 15$, and $\varepsilon = 0.15$ and for outside the Bay were $\gamma = 0.00011$, $C = 100$, and $\varepsilon = 0.15$. The optimal MRPRT–RVM parameter was $\sum = 0.91\text{E}{-}08$ for the sample points inside the Bay, and $\sum = 0.8\text{E}{-}08$ for the points outside the Bay.

The estimative capability of the developed MRPRT model was evaluated on the remaining 20% testing data.

RESULTS AND DISCUSSION

Figure 7 illustrates the estimative capability of the MRPRT approach validated using the testing data, and it was compared to the best individual regression technique. Table 3 summarizes the improvement in the estimative capability obtained using the MRPRT approach compared to the best individual regression technique for the sample points both inside and outside the Bay. For sample points inside the Bay, MRPRT approach using both SVM and RVM considerably improved the estimative capability compared with the RBF technique (Fig. 7a). The r improved from 0.57 (RBF) to 0.89 using MRPRT–SVM and to 0.98 using MRPRT–RVM. Similarly, the E improved from 28.31(RBF) to 70.78 using MRPRT–SVM and to 95.69 using MRPRT–RVM. The reliability of the approach of using independent subset for training and testing in the case of small samples was verified using K-fold cross validation. The results from the K-fold cross validation are presented in Table 4. It is observed that the variation of r and E for small samples when a single subset is used for testing is about 10% compared to the more robust K-fold cross validation. In the case of MRPRT–SVM, r varied from 0.77 to 0.89 with a mean of 0.82 whereas, E varied from 63.91 to 70.78 with a mean of 65.99. Similarly, in the case of MRPRT–RVM, r varied from 0.87 to 0.98 with a mean of 0.93, whereas E varied from 90.06 to 95.69 with a mean of 92.91.

However, for sample points outside the Bay, the estimative capability of MRPRT was only marginally superior to both the SVM and RVM techniques (Fig. 7b). The r improved from 0.56 (SVM) to 0.59 using MRPRT–RVM and to 0.61 using MRPRT–SVM. Similarly, the E improved from 28.46 (SVM) to 29.37 using MRPRT–RVM and to 34.52 using MRPRT–SVM.

As mentioned before, MRPRT is an approach that uses outcomes of several poorly estimating but complimentary models to provide an improved estimate using a pattern recognition method that captures the estimative strengths of these models. We observed that inside the Bay, both MRPRT–RVM and MRPRT–SVM provided significantly better accuracy than the individual geospatial techniques. However, outside the Bay, both MRPRT–SVM and MRPRT–RVM did not show

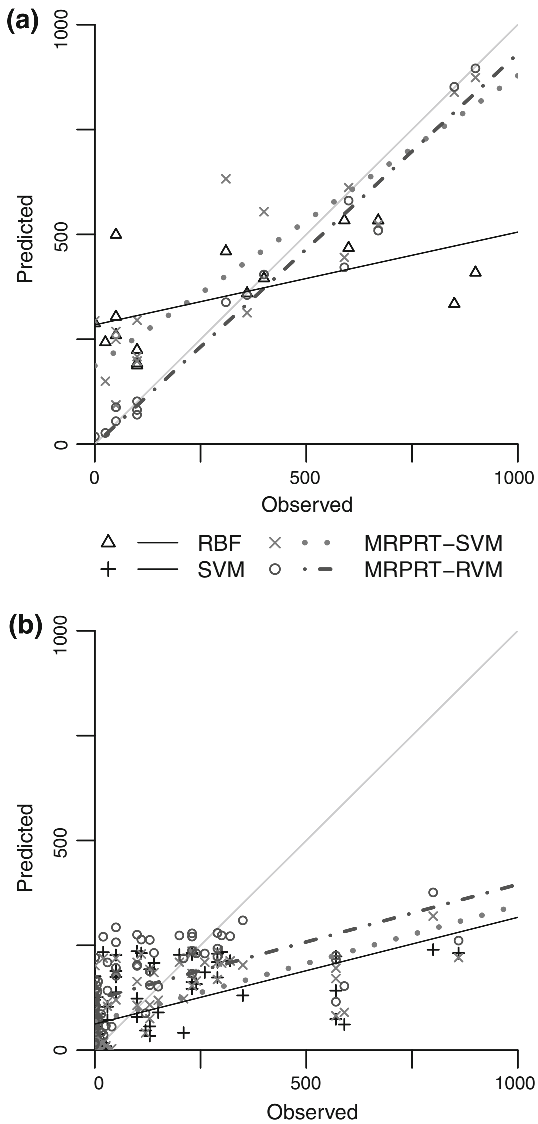

Figure 7. Observed vs. predicted Pt concentration values in mg/m^3 using the MRPRT–SVM, MRPRT–RVM and the best individual regression technique for Pt concentration values, (a) inside the Bay, (b) outside the Bay. The linear models are included to summarize the general trends in the predictions of the different models.

Table 3. The Predictive Capability of the MRPRT Approach Compared with the Best Individual Regression Techniques Verified on the Testing Data for the Pt Samples Both Inside and Outside the Bay Using *r* and *E*

Models	*r*	*E* (%)
Inside		
RBF	0.57	28.31
MRPRT–SVM	0.89	70.78
MRPRT–RVM	0.98	95.69
Outside		
SVM	0.56	28.46
MRPRT–SVM	0.59	29.37
MRPRT–RVM	0.61	34.52

Table 4. The Predictive Capability of the MRPRT Using SVM and RVM Approaches Verified Using *K*-fold Cross Validation for the Pt Samples Inside the Bay Using *r* and *E*

K-Fold	*r*	*E* (%)
MRPRT–SVM		
Fold-1	0.89	70.78
Fold-2	0.79	64.18
Fold-3	0.82	65.06
Fold-4	0.86	66.00
Fold-5	0.77	63.91
Mean (Fold 1–5)	0.82	65.99
MRPRT–RVM		
Fold-1	0.94	92.94
Fold-2	0.96	93.02
Fold-3	0.98	95.69
Fold-4	0.87	90.06
Fold-5	0.91	92.87
Mean (Fold 1–5)	0.93	92.91

any considerable improvement in comparison with the individual geospatial techniques. As noted before, the global trend analysis of the available Pt values revealed that the data from inside the Bay were free from any spatial trend, while the data from outside the Bay depicted strong trends in the north–south and east–west directions. In the north–south direction, it is found that the values of platinum increase toward the north, whereas in the east–west direction, the values of platinum increase in the center and decrease toward both east and west (Oommen, 2006). The difference in the spatial trend of Pt values within and outside the Bay indicates the difference in the accumulation and depositional processes of Pt in these regions. This could be a potential reason for the difference in the MRPRT performance inside and outside the Bay.

In addition, inside, the Bay acts as a well-constrained system for which we have reasonable amount of well-distributed data points that capture the variability of the process (geospatially interpolation is the major factor in play for analysis). Hence, the suites of geospatial and pattern recognition models were able to provide complimentary estimations, albeit with poor individual estimative capability. In such situations, an MRPRT approach could considerably improve the predictive capability as observed in Table 3.

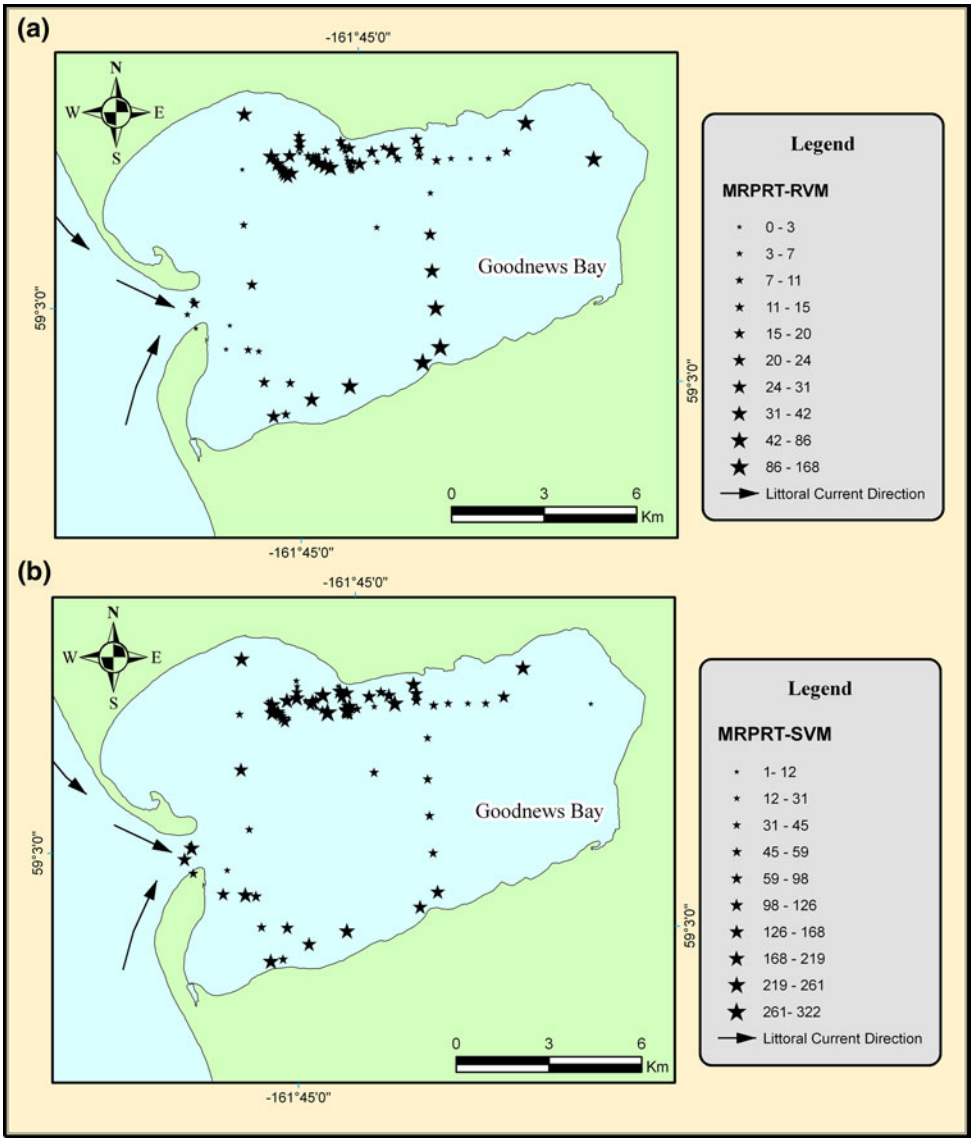

Figure 8. Map showing the difference in the observed and predicted Pt concentration values in mg/m^3 using both (a) MRPRT–RVM and (b) MRPRT–SVM for data inside the Bay. The size of the star symbol correlates to the difference in the values with larger symbols representing higher differences between the observed and predicted Pt concentration values.

On the other hand, the open water outside the Bay provides an open-ended depositional environment where we have insufficient data capturing process information further offshore. We envisage the lack of significant improvement outside the Bay by both MRPRT–SVM and MRPRT–RVM over the individual SVM estimations, and that the two MRPRT methods provided almost comparable accuracy was because of the non-complimentary estimative ability of the suite of individual geospatial and pattern recognition model outcomes used (Fig. 5). Being non-complimentary in nature, the support vectors and the relevant vectors obtained due to each method for the MRPRT models were redundant, thus rendering the improvement in estimative capability of the MRPRT difficult.

Figures 8 and 9 present the absolute difference in the observed and predicted Pt concentration values (mg/m^3) using both MRPRT–RVM and MRPRT–SVM for data inside the Bay and outside

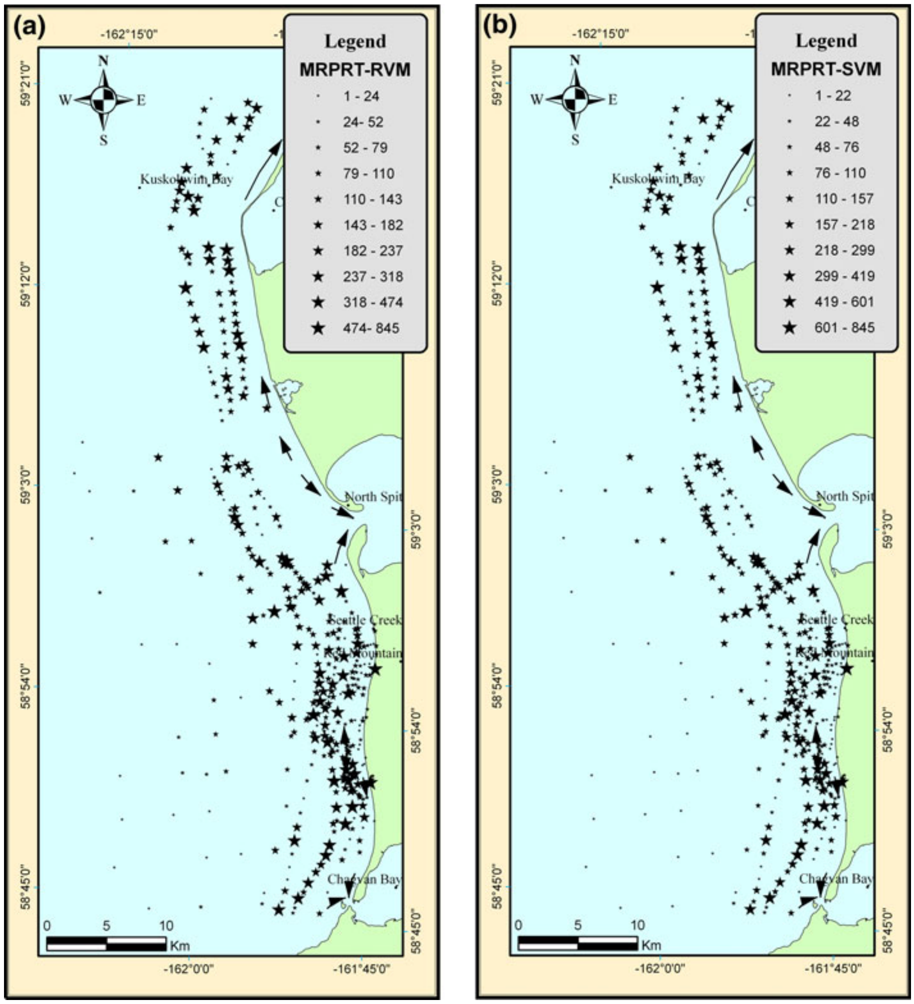

Figure 9. Map showing the difference in the observed and predicted Pt values using both (a) MRPRT–RVM and (b) MRPRT–SVM for data outside the Bay. The size of the star symbol correlates to the difference in the values with larger symbols representing higher differences between the observed and predicted Pt values.

the Bay, respectively. The size of the star symbol correlates to the difference in the values with larger symbols representing higher differences between the observed and predicted Pt concentration values. In the case of data inside the Bay, the difference in the observed to the predicted values is much lower in the case of MRPRT–RVM (Fig. 8a) compared to MRPRT–SVM (Fig. 8b). However, Figure 8a and b demonstrates that the uncertainty between the observed and predicted Pt values increase towards the coast. In the case of data outside the Bay, both MRPRT–RVM (Fig. 9a) and MRPRT–SVM (Fig. 9b) have similar differences in the observed compared to the predicted Pt values. In general, it is observed that values north of Goodnews Bay have higher uncertainty between the observed and predicted Pt values compared to the values south.

CONCLUSIONS AND RECOMMENDATIONS

The proposed MRPRT technique improved the Pt predictions in the case of values both inside and outside the Bay. This improvement was far more significant inside the Bay where processes were constrained and the prediction results by individual analysis techniques were complimentary. These results demonstrate that the developed MRPRT approach holds promise, not only for marine Pt exploration but also for wider applications in georesource estimation, where data paucity is common.

Further research is required to establish criteria for determining the optimal number of individual regression techniques that should be included in the MRPRT approach, and establishing the minimum variation in the *r* required within the different quartiles (or other statistical segregation of the data) for obtaining a significant improvement in the estimative capability when MRPRT is used. Another aspect that needs attention in the future is the utilization of a trend removal model for the enhanced predictability of the MRPRT method.

ACKNOWLEDGMENTS

Part of this study was done by the first author when he was a graduate student at the University of Alaska Fairbanks, who acknowledges the financial support that was provided by the Minerals Management Service (MMS) of the US Department of Interior under the cooperative agreement number 1435-01-02-CA-85124. The authors are grateful to all those who helped them to gather data from the various research studies done in this region during the last five decades.

REFERENCES

Barker, J. C., 1986, Platinum-group metals, gold and chromium resource potential offshore of Platinum, Alaska Presented at the 17th annual underwater mining conference, Biloxi, Missouri, available upon request from files at Bureau of Mines, Alaska Field Operations Center: Anchorage, Alaska.

Barker, J. C., and Lamal, K., 1989, Offshore extension of platiniferous bedrock and associated sedimentation of the Goodnews Bay ultramafic complex, Alaska: Mar. Min., v. 8, no. 4, p. 65.

Chee, L. P., and Harrison, R. F., 2003, Online pattern classification with multiple neural network systems: an experimental study: IEEE Trans. Syst. Man Cybern., v. 23, no. 2, p. 235–247.

Clark, I., and Harper, W. V., 2000, Practical geostatistics: Ecosse North American LLC, Columbus, OH, 411 p.

Drucker, H., Cortes, C., Jackel, L. D., Lecun, Y., and Vapnik, V., 1994, Boosting and other ensemble methods: Neural Comput., v. 6, no. 6, p. 1289–1301.

Dutta, S., 2006, Predictive performance of machine learning algorithms for ore reserve estimation in sparse and imprecise data: University of Alaska, Fairbanks, p. 189.

Dutta, S., Misra, D., Ganguli, R., Samanta, B., and Bandopadhyay, S., 2006, A hybrid ensemble model of kriging and neural network for ore grade estimation: Int. J. Surf. Min. Reclamat. Environ., v. 20, no. 1, p. 33–45.

Foody, G. M., and Mathur, A., 2004, A relative evaluation of multiclass image classification by support vector machines: IEEE Trans. Geosci. Remote Sens., v. 42, no. 6, p. 1335–1343.

Fumera, G., and Roli, F., 2005, A theoretical and experimental analysis of linear combiners for multiple classifier systems: IEEE Trans. Pattern Anal. Mach. Intell., v. 27, no. 6, p. 942–956.

Gutta, S., Huang, J. R. J., Jonathon, P., and Wechsler, H., 2000, Mixture of experts for classification of gender, ethnic origin, and pose of human faces: IEEE Trans. Neural Netw., v. 11, no. 4, p. 948–960.

Hansen, L. K., and Salamon, P., 1990, Neural network ensembles: IEEE Trans. Pattern Anal. Mach. Intell., v. 12, no. 10, p. 993–1001.

Harrington, G. L., 1919, Mineral resources of the Goodnews Bay region: U.S. Geol. Surv. Bull., v. 714, p. 207–228.

Hastie, T. R. T., and Friedman, J., eds., 2003, The elements of statistical learning: data mining, inference and prediction: Springer-Verlag, New York, p. 533.

Isaaks, E. H., and Srivastava, R. M., 1989, An introduction to applied geostatistics: Oxford University, New York, 561 p.

Jacobs, R. A., 1995, Methods for combining experts probability assessments: Neural Comput., v. 7, no. 5, p. 867–888.

Johnston, K., 2003, Using ArcGIS geostatistical analyst: ESRI Press, 316 p.

Kanevski, M., Pozdnoukhov, A., Canu, S., and Maignan, M., 2002, Advanced spatial data analysis and modeling with support vector machines: Int. J. Fuzzy Syst., v. 4, no. 1, p. 606–616.

Karatzoglou, A., Meyer, D., and Hornik, K., 2006, Support vector machines in R: J. Stat. Softw., v. 15, no. 9, p. 25.

Kittler, J., Hatef, M., Duin, R. P. W., and Matas, J., 1998, On combining classifiers: IEEE Trans. Pattern Anal. Mach. Intell., v. 20, no. 3, p. 226–239.

Kittler, J., and Roli, F., 2001, Lecture notes in computer science, *in* Proceedings of 2nd International Workshop on Multiple Classifier Systems, Cambridge, UK.

Kuncheva, L. I., 2004, Combining pattern classifiers: methods and algorithms: Wiley–Interscience.

Meher, S. K., Ghosh, A., Shankar, B. U., and Bruzzone, L., 2006, Neuro-fuzzy fusion: a new approach to multiple classifier system, *in* ICIT 2006: 9th International Conference on Information Technology, Proceedings, p. 209–212.

Meyer, D., 2001, Support vector machines: R News, v. 1/3. http://cran.r-project.org/doc/Rnews/Rnews_2001-3.pdf.

Misra, D., Samanta, B., Dutta, S., and Bandopadhyay, S., 2007, Evaluation of artificial neural network and kriging for the prediction of arsenic in alaskan bedrock derived sediments using gold concentration data: Int. J. Min. Reclamat. Environ., v. 21, no. 4, p. 282–294.

Nash, J. E., and Sutcliffe, J. V., 1970, River flow forecasting through conceptual models. Part I. A discussion of principles: J. Hydrol., v. 10, no. 3, p. 282–290.

Oommen, T., 2006, Geodatabase development and GIS based analysis for resource assessment of placer platinum in the offshore region of Goodnews Bay, Alaska: University of Alaska, Fairbanks, 198 p.

Oommen, T., and Baise, L. G., 2010, Model validation for intelligent data collection in geotechnical earthquake engineering: J. Comput. Civil Eng., v. 24, no. 6, p. 467–477.

Oommen, T., Misra, D., Twarakavi, N. K. C., Prakash, A., Sahoo, B., and Bandopadhyay, S., 2008a, An objective analysis of support vector machine based classification for remote sensing: Math. Geosci., v. 40, no. 4, p. 409–424.

Oommen, T., Prakash, A., Misra, D., Kelley, J. J., Naidu, S. A., and Bandopadhyay, S., 2008b, GIS based marine platinum exploration, Goodnews Bay Region, Southwest Alaska: Mar. Georesour. Geotechnol., v. 26, no. 1, p. 1–18.

Perrone, M., and Cooper, L. N., eds., 1993, Neural Network for Speech and Image. When networks disagree: ensemble methods for hybrid neural networks: Chapman and Hall, London.
Pozdnoukhov, A., 2005, Support vector regression for automated robust spatial mapping of natural radioactivity: J. Appl. GIS, v. 1, no. 2, p. 2101–2110.
Tipping, M. E., 2001, Sparse Bayesian learning and the relevance vector machine: J. Mach. Learn. Res., v. 1, no. 3, p. 211–244.
Twarakavi, N. K. C., Misra, D., and Bandopadhyay, S., 2006, Prediction of arsenic in bedrock derived stream sediments at a gold mine site under conditions of sparse data: Nat. Resour. Res., v. 15, no. 1, p. 15–26.
Vapnik, V., 1995, The nature of statistical learning theory: Springer-Verlag, New York.
Wolpert, D. H., 1992, Stacked generalization: Neural Netw., v. 5, no. 2, p. 241–259.