African Journal of Biotechnology Vol. 11(91), pp. 15771-15778, 13 November, 2012
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB11.995
ISSN 1684–5315 ©2012 Academic Journals

Full Length Research Paper

Prediction of spatial distribution for some land use allometric characteristics in land use planning models with geostatistic and Geographical Information System (GIS) (Case study: Boein and Miandasht, Isfahan Province, Iran)

Parvaneh Peykanpourfard[1]*, Masoud Monavvari[1], Ali Asghar Alesheikh[2], Atefeh Chamani[3] and Fariba Ghanbari[4]

[1]Department of Environmental Science, Faculty of Environment and Energy, Science and Research Branch, Islamic Azad University, Tehran, Iran.
[2]Faculty of Geomatics Engineering, K. N.Toosi University of Technology, Tehran, Iran.
[3]Department of Environment, Faculty of Agriculture and Natural Resources, Khorasgan (Isfahan) Branch, Islamic Azad University, Isfahan, Iran.
[4]Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran.

Accepted 21 September, 2011

Although traditional census can present unbiased information about different land uses, it is spatial independent and do not present particular information about spatial distribution of studied characteristic. In this study, we used geostatistic and Geographical Information System (GIS) to estimate some different land uses allometric characteristics in Isfahan Province (Iran). Thus, samples information was surveyed considering their geographic position in the studied area. After optimizing variogram parameters, empirical variogram was prepared to investigate spatial structure of different land uses allometric characteristics. Our results confirme that spatial structure for the quantitative characteristics of different land uses has a moderate degree of spatial correlation, except for type variable that has no spatial structure. Nugget effect for variogram obtained from the quantitative characteristics of different land uses was equal to 35 to 64%. We used ordinary Kriging for preparing Kriging map and Kriging standard deviation of different land uses. Also, we used geostatistic and GIS to compare geostatistical and algebraic interpolation methods and nine different interpolation methods (Kriging, local polynomial methods, inverse distance weighted, radial basis functions, global polynomial, moving average weighted, natural neighbor, nearest neighbor and triangulation with Linear Interpolation) were investigated. Spatial distribution of different land uses quantitative characteristics were validated with ordinary Kriging and algebraic methods. Our results confirm that ordinary Kriging has more accuracy than other methods for spatial prediction of different land uses quantitative characteristics.

Key words: Geostatistic, interpolation method, land use allometric characteristics, Kriging.

INTRODUCTION

Since an extensive and continuously sampling for spatial information is necessary but difficult or impossible, generalization of sample information to all different land uses with simple interpolation has low accuracy. Spatial interpolation methods have been applied to many disciplines (Li and Heap, 2010). In classic statistic method, we need extensive sampling to decline variation correlation. Despite high costs spent in this method, the

*Corresponding author. E-mail: parvaneh_pey@yahoo.com.

results are limited only on mean value for specific classes. It is evident that the expression of quantitative and qualitative characteristics of different land uses in the form of a numerical quantity such as the general average is not adequate, even with optimum sampling, because the general average is not able to describe the local changes (from one point to another) of studied characteristic. Different models such as remote sensing; linear regression, Generalized Linear Model (GLM) and Artificial Neural Network (ANN) are used for spatial prediction of biological characteristics of different land uses (Franklin et al., 2000). Inverse Distance Weighting (IDW), Ordinary Kriging (OK), and Ordinary Co-Kriging (OCK) are the most frequently used methods (Li and Heap, 2010). The major difference between geostatistic and classic statistic is that geostatistic can calculate estimation variance and the efficiency of our estimation method with variogram. One of the most priority aspects of geostatistic in spatial prediction of a character is that final estimation of unobserved points has the minimum variance, and also is unbiased. Furthermore, Kriging technique unlike other interpolation methods, can display error map or estimation variance at each point (Goovaerts, 1999).

One of the most popular geostatistic methods that have been used widely in environment, mining exploration, mapping and hydrology is Kriging. The first and so far most significant contribution to forest inventory is due to Guibal (1973), who applied Kriging to terrestrial forest inventory and confirmed that geostatistic are more accurate than classic statistic, particularly for small areas (Mandallaz, 1993). Nieschuleze (2003) showed that geostatistic for forest volume inventory is more suitable than classical method in terms of accuracy and cost. Furthermore, the efficiency of geostatistic in biomass validation has been confirmed (Sales et al., 2007). In this work, we investigated spatial structure and estimated some different land uses characteristics in Boein and Miandasht (Isfahan Province) with geostatistic.

MATERIALS AND METHODS

The studied area (Boien-Miandasht, Fereydonshahr, Isfahan Province) is located in 32° 37′ to 33° 4′ north latitude and 49° 36′ to 50° 19′ east longitude. This region is located in 110 km of Isfahan. Its total area is 3 km^2 and has seven different land uses (Figure 1).The rainfall amount is about 649 mm and is usually cold and wet. The soil is mainly brown and gray and has acidic nature.

Inventory network was designed with dimensions of 250 (east - west) and 200 m (north - south) and placed on the map. In the center of each grid, two to four points were randomly chosen, then, other samples were randomly chosen in different dimensions and intervals (50, 100, 150 and 200 m in eight geographical vectors). Therefore, our sampling is combination of two methods; random systematic and random clustering. Geographical location of sample plots was recorded using the Global Positioning System (GPS). Volume inventory in plot sites was calculated with one agent table. Variogram was used to determine and describe the spatial structure. The spatial patterns of the predictions of the most accurate methods were also visually examined for their validity (Li et al., 2011).

Isotropic and anisotropic conditions

Isotropic and anisotropic conditions of the studied variables were investigated with surface variogram. In Kriging, it is important to recognize anisotropic conditions. Different methods for anisotropic recognition such as surface variogram, was used in this study.

Spatial structure analysis

VARIOWIN 2.2 (Pannatier, 1996) was used to calculate and prepared empirical variogram. After normal test and due to our isotropic data, omni- directional variogram for all variables were plotted. Then, the variogram was fitted to empirical models with visual hybrid and automated interpretation method using VARIOWIN. Goodness of fit index was used as the fit congruence criterion.

Kriging maps and estimation error

Kriging also offers variance of estimation and thus not only calculate the mean value of estimation error, but also estimate error distribution in all the studied range and calculate confidence interval of our estimation. With this unique feature, Kriging can identify parts with high error and more necessary data. Ordinary Kriging maps and estimation error for the studied variables were produced in Arc GIS 9.2.

Variogram validation control

Validation control is the estimation of sampling points in each region using a neighboring sample values (without regard to the sample) with the Kriging. The number of neighbor points used for estimation, was selected based on the effect range of fitted model to variogram, and nonlinear least square method to obtain reduced error mean near zero and its variance to unit. In Jack Knife method, each known point is estimated based on its neighborhood samples. Model validation evaluation was done with mean bias error (MBE), mean square error (MSE) and root mean square error (RMSE) (Equations 1 to 4).

$$MSE = \left({}^{1}\!/_{n} \sum_{i=1}^{n} \left(\hat{y}_i - y_i \right) \right)^2 \quad (1)$$

$$ME = \sum_{i=1}^{n} (\hat{y}_i - y_i) / n \quad (2)$$

$$RMSE = \left(\sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \Big/ n \right)^{1/2} \quad (3)$$

$$\%RMSE = RMSE \Big/ \bar{\hat{y}}_i \quad (4)$$

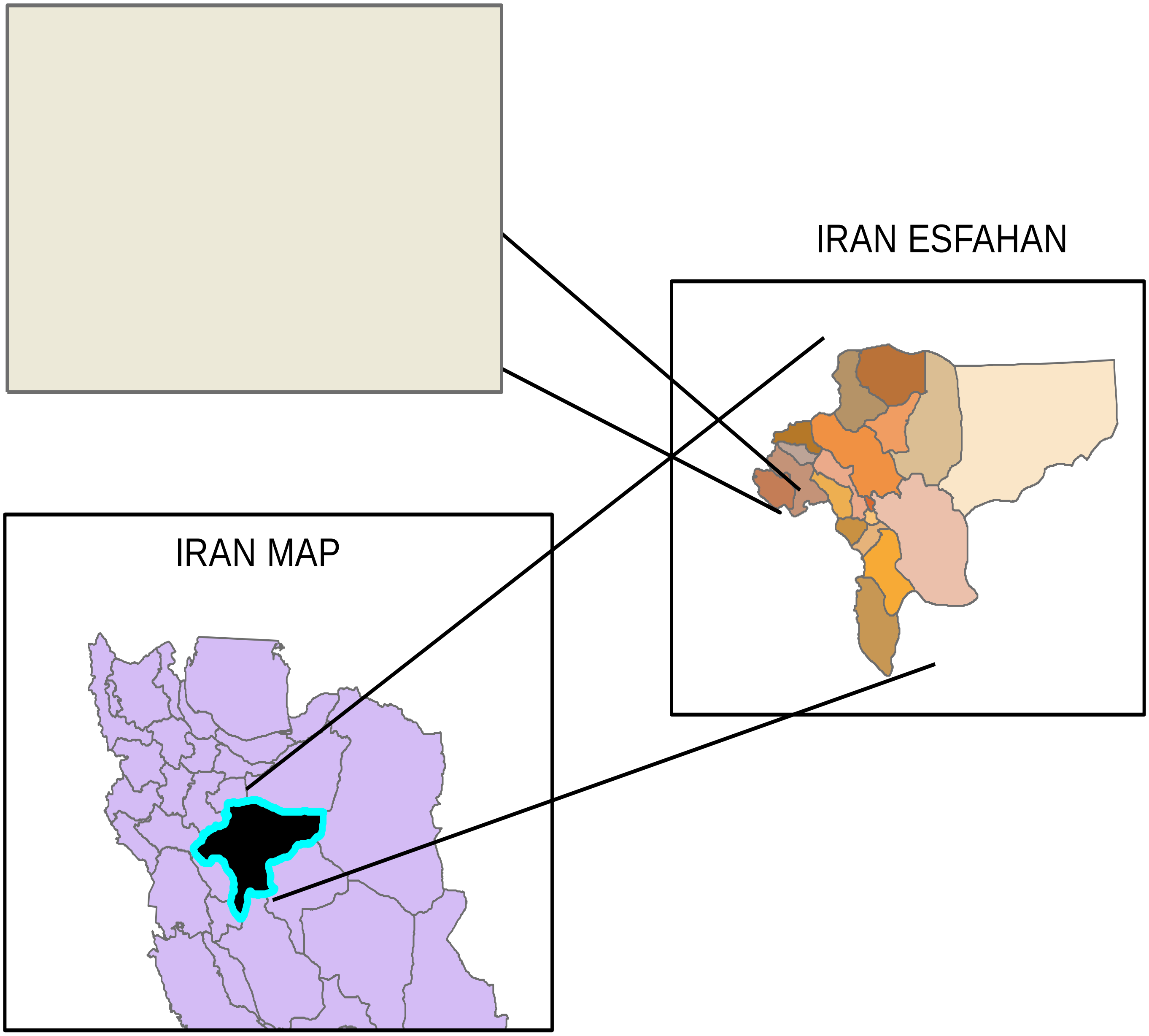

Figure 1. The tudied area (Boien-Miandasht, Fereydonshahr, Isfahan Province).

Where, $\bar{\hat{y}}$ = mean of estimates; y = observed value; n = number of observations; $\hat{y}$ = estimate value.

In this method, initially the model of variogram for studied variables was estimated and then the resulting model was used for estimation. Therefore, the accuracy of fitted model was investigated with estimation error analysis. The fitting model should not cause systematic error; in other words, the estimation error average should be zero and mean square error should be minimal. RMSE index represents the degree of accuracy for estimation and must be minimum for an unbiased estimate as much as possible. In this study, investigation was carried out with tried and error method and the most appropriate variogram model for studied variables were determined (Ramesh et al., 2005).

Different interpolation methods in prediction of quantitative characteristics

Algebraic interpolation techniques used in this study include local polynomial methods, inverse distance weighted, radial basis functions, global polynomial, moving average weighted, natural neighbor, nearest neighbor and triangulation with linear Interpolation. Furthermore, for geostatistic interpolation, ordinary Kriging was used. Finally, to produce the resulting maps, Surfer 8 and Arc GIS 9.2 were used. To determine the accuracy of estimators, ME and MSE criteria were used. The best estimate

Table 1. Descriptive statistic of land uses volume inventory characteristic.

kurtosis	Skewness	Variation coefficient	Variance	Standard deviation	Mean	Variation range	Maximum	Minimum
-0.13	0.476	0.41	22350.76	149.5	363.84	733.3	797.4	64.1

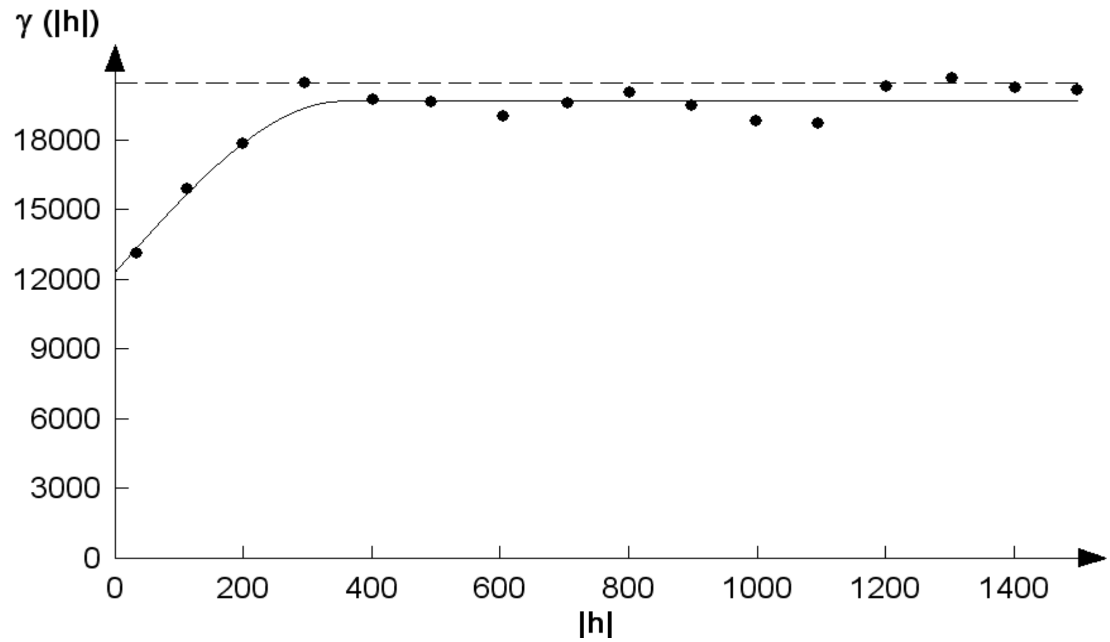

Figure 2. Omni-directional variogram to volume inventory variable. h= distance (m) and $\gamma(|h|)$ = paired sample variance.

should be unbiased and with minimum variance of error. ME show the bias value and ideally should be zero. Significant positive or negative values respectively indicate overestimate or underestimate of the true value (Wakernagel, 2002). MSE indicates the mean standard deviation of the estimated value and observed value and whatever is closer to zero is better (Allexandra and Bullock, 1999). Finally, significant differences between estimated and actual values were calculated with paired t-test to identify the best estimator. Although a reconnaissance sampling is necessary and basic requirements of geostatistics have to be met, Kriging has the advantage of giving estimates with a minimized error (Van Beurden and Riezebos, 2003).

RESULTS

Normal distribution of data was confirmed with Kolmogorov-smirnov test. Data distribution were studied with statistic indices such as average, maximum, minimum, standard deviation, variation coefficient and variation range (Table 1). High variation coefficient in the table indicates much variability between the data. Variogram analysis showed the studied variable as isotropic; therefore omni- directional variogram was used for Kriging. Among various empirical models, spherical model had the best fit to omni- directional variogram. The hybrid method and goodness of fit were performed in Variowin 2.21 (Figure 2).

Non-structured part (nugget effect) explained 63% and structured part 37% of the total variance. The variogram value had small effect range, thus reducing the allowed range which we can use available data for variable estimation in unknown point or block. It is evident that the smaller range effect indicated the more limited spatial structure. It is notable that this variable has silled model and thus, spatial structure. The profile of fitted model on empirical variogram, are given in Table 2. Structure ratio in the table, was calculated via nugget-to-sill [C0/(C+C0)] that is defined as spatial dependence index. So, the values less than 25% indicated strong spatial dependence, values of 25 to 75% as moderate spatial dependence, and more than 75% represented weak spatial dependence (Sumfleth and Duttmann, 2008). Thus, spatial structure of the studied variables was moderate.

The variogram parameters were estimated with Kriging Jack Knife. Model validation was calculated with the MBE, MSE and RMSE. According to Table 3 (biased estimates and high MAE and RMSE), the fitted model, could not present accurate estimate for rangeland and forest volume inventory due to the high nugget effect and

Table 2. Fitted model to omni-directional variogram characteristics.

Effect range (m)	Structure ratio $C_0/(C+C_0)$	Sill (C_0+C)	Nugget effect (C_0)	Structure variability (C)	Fitted model
288.55	64	19757.5	12656	7101.5	Spherical

Table 3. Variogram validation control results.

% RMSE (m^3/ha)	RMSE (m^3/ha)	MSE (m^3/ha)	MAE (m^3/ha)	MBE (m^3/ha)
37.85	135.83	18450.27	109.11	-0.183

weak spatial structure. Figure 3 also confirms these results. After determining the optimal model and its parameters, the interpolation was performed in Arc GIS 9.2 with this hypothesis that the average is unknown (ordinary Kriging). Estimation errors map indicated that despite irregular sampling network, estimation error decreased in areas with more sampling and increased in areas with low sampling and also in marginal zones with lower points.

Comparison between different interpolation methods

Accuracy

Selection of an optimal interpolation technique to estimate the studied characteristics of non-sampling points plays an important role in data management. In some interpolation methods, it is necessary to determine the optimal value. Hence, optimal value of each method was initially determined and then, interpolation was performed. Optimal value was determined via calculation of the minimum values of RMSPE; that is minimum square of prediction error. The value is best where RMSPE is least. Points in the curve (Figure 4), shows error prediction square with different value. The minimum RMSE in the curves specifies the best value in each model. Finally, interpolation methods were carefully validated with the mean error and mean square error indicators. The result of different interpolation method, for volume variable is presented in Figure 4.

Volume in hectare

The results of inverse distance weighted (IDW), local polynomial and global polynomial for optimal value is unit. In other words, linear relationships are better than other types of relationships. Also, first degree equations had the best results to 2 or 3 degrees (Figure 4). Model validation for volume in hectare was calculated using different interpolation methods, whose results are presented in Tables 4 and 5. Significantly high levels (greater than 0.05) in 95% confidence level for paired t test showed no significant differences between real and predicted values.

DISCUSSION

Averaging the predictions of the most accurate methods showed no significant improvement in the predictive accuracy (Li et al., 2011). Results show that the spatial structure of rangeland and forest quantitative characteristics revealed a moderate degree of spatial dependence but forest types did not have spatial structure. Variograms for quantitative attribute revealed 35 to 64% nugget effect (nugget-to-sill). Ordinary Kriging was used for Kriging and Kriging standard deviation maps allometric attribute. Geostatistics, coupled with a GIS were used to test nine different techniques of interpolation ordinary Kriging (OK), inverse distance weighting, local polynomial, radial basis functions, global polynomial, moving average weighted, natural neighbor, nearest neighbor and triangulation with linear interpolation to compare geostatistical and deterministic interpolation methods. The spatial distribution of quantitative characteristics was estimated by ordinary Kriging and other deterministic procedure.

The cross-validation analysis showed that in the case of spatially prediction of these forest quantitative characteristics, OK had more precision than the deterministic methods. To evaluate possibility of spatial prediction of allometric characteristics using terrain analysis and linear regression model, digital elevation models with 10 × 10 m resolution was used. The primary topographic attributes (slope, aspect, profile curvature and plan curvature, tangential curvature, specific catchment area and shade relief and secondary attributes) wetness index, solar radiation, relative stream power and sediment transport capacity index) and elevation from sea level, were derived from the digital elevation model and terrain analysis software package. The relationships between the allometric (training site) and terrain attributes were analyzed then modeled using multiple linear regression models by stepwise approach. The developed models were validated using test data. Adjusted R2 and RMSE were determined to validate accuracy of predictors.

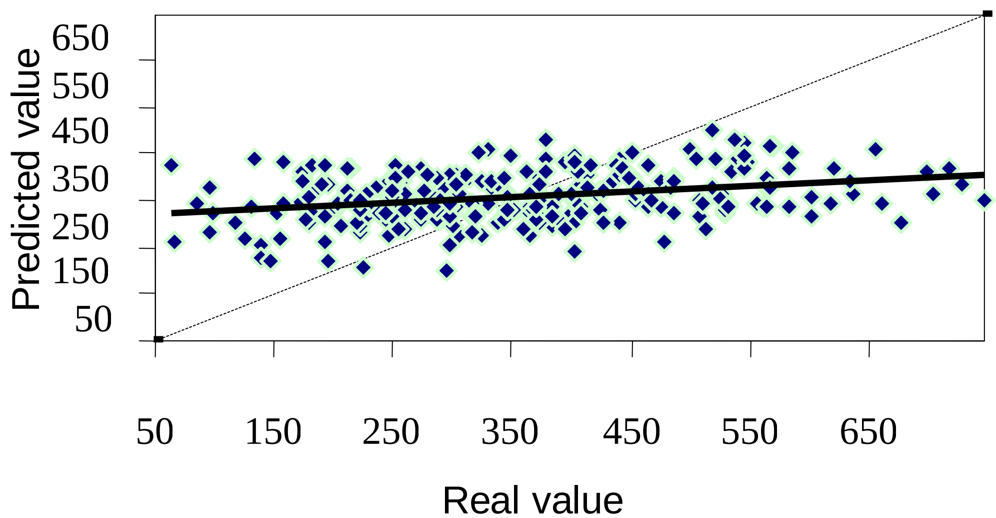

Figure 3. Real values versus predicted values.

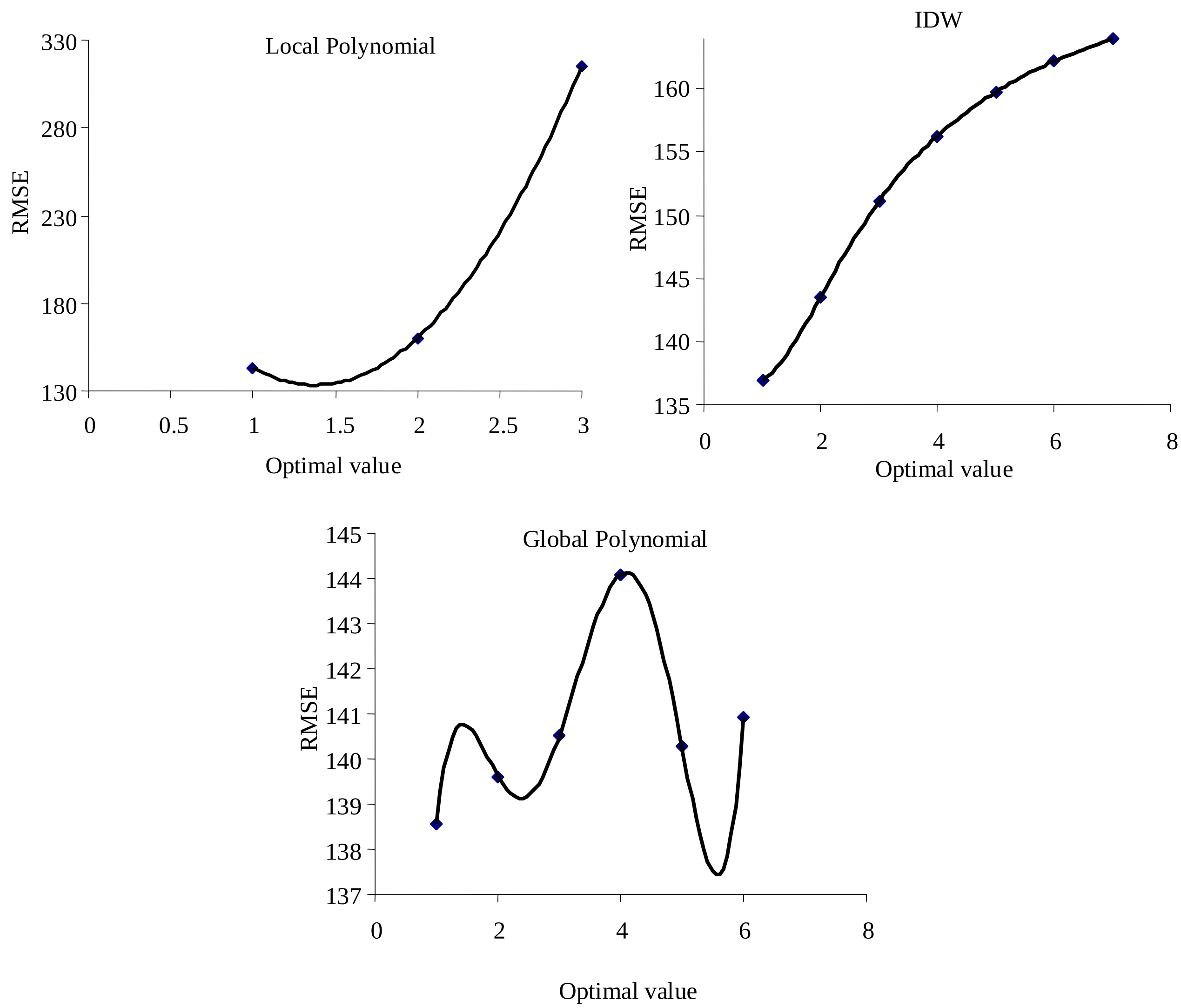

Figure 4. Optimal value calculation results with some interpolation methods for volume in per hectare.

Table 4. Different interpolation methods validation results for volume per hectare variable.

Variable	Interpolation method	Type	ME	MSE	RMSE	RMSE%
Volume per hectare	Kriging	Ordinary Kriging	-0.183	18450.27	135.83	37.85
	Inverse distance weighted	Value 1	-2.54	20585.99	143.48	40
	Natural neighbor	-	-0.889	23085.96	151.94	42.3
	Nearest neighbor	-	-1.28	30946.29	175.92	49
		Inverse multiquadratic	-0.31	18501.6	136.02	37.9
		Multiquadratic	-3.21	23095.45	151.97	42.4
	Radial basis functions	Completely regularized spline	-0.58	19044.11	138	38.5
		Spline with tension	-0.45	18920.12	137.55	38.33
		Thin plan spline	-6.96	30277.11	174	48.5
	Triangulation with Linear Interpolation	-	-1.32	24723.9	157.24	43.8
	Moving average weighted	-	0.67	19802.64	140.72	39
	Local polynomial	Value 1	-2.15	20480.29	143.11	39.88
	Global polynomial	Value 1	0.0102	19203.37	138.58	38.6

Table 5. Validation results for comparison of estimated values to real values with paired t test.

Significance	Significant correlation between real and predicted values	Mean estimation	Mean observation	The best estimator	Mean volume (m^3/ha)
0.983	0.318	358.64	358.83	Kriging	

Results show that elevation, solar radiation and aspect were the most significant terrain attributes that determined the spatial distribution of allometric forest.

Results also show that characteristics could be predicted about 0.04 to 45% of variation by these linear models. Logistic regression technique was implemented for forest type's predictive modeling from terrain variables. The results also show that altitude, solar radiation potential and aspect was the main factor controlling land use types, respectively. However, to date, the selection of image object sets to represent landscape patterns has been largely subjective. Changes in observation scale affect the variance and spatial dependence of measured variables, and may be useful in determining which levels of image segmentation are most appropriate for a given purpose (Karl and Maurer, 2010). We found that the segmentation level whose regression predictions had spatial dependence that most closely matched the spatial dependence of the field samples also had the strongest predicted-to-observed correlations (Karl and Maurer, 2010).

Conclusion

In this study, we tried to predict spatially land use inventory particularly forest and rangeland volume inventory as dominant regional land uses. In this method, variogram analysis was used to study spatial correlation of nugget effect and then, this correlation was used to site selection models and nugget effect prediction in the entire area (Kriging). The fitted model to studied variable variogram had distinct sill and thus, the general trend in this variable was rejected (Nieschuleze, 2003). Considering that pure nugget effect model does not fit into variogram, weak spatial structure for the studied variable was proved. Weak validation results are also due to weak spatial structure and therefore, estimates values are not so similar to actual values. Some high nugget effects may be due to site selection and measured errors or it may be that the studied variables have variability in distances less than the minimum distance between sampling plots. High variability in volume inventory variogram is due to high nugget effect that is an indicator for forest and rangeland data (Jost, 1993). Furthermore, the structure studied areas had high variability and due to human interference, the spatial structure had been weaker over time (Biondi et al., 1994).

The comparison between spatial structure in this study with other studies shows that spatial distribution pattern of each land use characteristic is unique and considered as function of some factors such as soil, topography, region microclimate, field management and utilization manner. Finally, we suggest that:

1) Before sampling, homogeneous surfaces should be

marked with suitable stratification method and inside them, geostatistic methods should be used. In other words, the field should be categorized with update aerial photos and satellite images and Kriging can be used in these categories.
2) Geostatistic could be used in forest and rangeland which covers more homogeneous surfaces and their changes are gradual and soil dependent.
3) In geostatistic analysis, precise plot geographic location is very important.
4) Larger plots increase the accuracy of geostatistic models, and because the sample size is smaller, the variance was greater. Geostatistical tools are recommended for mapping statistically estimated hot spots of vectors and pathogens (Schröder, 2006).

The variance of the predictions tended to underestimate that of the observations. The correlation of the predictions with the observations was weak at relatively fine scales, but strong at relatively coarse scales. There is evidence that the correlation of the predictions with the observations was not uniform across the transect at relatively fine scales (Pringle et al., 2008).

Abbreviations

GIS, Geographical Information System; **GLM,** Generalized Linear Model; **ANN,** Artificial Neural Network; **IDW,** Inverse Distance Weighting; **OK,** Ordinary Kriging; **OCK,** Ordinary Co-Kriging; **MBE,** mean bias error; **MSE,** mean square error; **RMSE,** root mean square error; **RMSPE,** minimum square of prediction error.

REFERENCES

Allexandra K, Bullock G (1999). A comparative study of interpolation methods for mapping soil properties. Agron. J. 91:393-400.
Biondi F, Myers DE, Avery CC (1994). Geostatistically modeling stems size and increment in an old growth forest. Can. J. For. Res. 24(7):1354-1368.
Franklin J, McCullough P, Gray C (2000). Terrain variables used for predictive mapping of vegetation communities in southern California. In: Wilson JP and Gallant JC (eds.), Terrain Analysis, Principles and Applications. John Wiley & Sons. Inc. pp. 331-353.
Goovaerts P (1999). Geostatistic in soil science: State of the art and perspective. Geoderma 38:45-93.
Guibal D (1973). The estimation of oukoumés Gabon, internal memo 333, Center of Mathematical Morphology, Fontainebleau, France.
Jost A (1993). Geostatistische analyse des Stichprobenfehlers systematischer Stichproben, Ph.D. thesis, University of Freiburg in Breisgau, p. 90.
Karl JW, Maurer BA (2010). Spatial dependence of predictions from image segmentation: A variogram-based method to determine appropriate scales for producing land-management information. J. Ecol. inform. 5(3):194-202.
Li J, Heap AD (2010). A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. J. Ecol inform. 6(3-4):228-241.
Li J, Heap AD, Potter A, Daniell JJ (2011). Advances in spatially environmental modeling: Application of machine learning methods to spatial interpolation of environmental variables. J. Envsoft. (in Press).
Mandallaz D (1993). Geostatistical methods for double sampling schemes: application to combined forest inventory. Technical report, Ph.D. thesis, ETH Zürich, chair of forest inventory and planning, p. 133.
Pannatier Y (1996). VARIOWIN software for spatial data analysis in 2D. Springer- Verlag, New York Inc. p. 91.
Pringle MJ, Marchant BP, Lark RM (2008). Analysis of two variants of a spatially distributed crop model, using wavelet transforms and geostatistics. J. Agric. Syst. 98(2):135-146.
Ramesh S, Teegavarapu V, Chandramouli V (2005). Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. J. Hydrol. 312:191-206.
Schröder W (2006). GIS, geostatistics, metadata banking, and tree-based models for data analysis and mapping in environmental monitoring and epidemiology. Int. J. Med. Microbiol. 296(1):23-36.
Sumfleth K, Duttmann R (2008). Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators. Ecol. Indic. 8:485-501.
Van Beurden SAHA, Riezebos HTh (2003). The application of geostatistics in erosion hazard mapping. Soil Technol. 1(4): 349-364.
Wakernagel H (2002). Multivariate geostatistics. Springer Press, p. 387.