

Geostatistical approach to National Forest Inventory data: a study case in the Province of Trento (Italy)

Maria Rizzo¹, Fabio Pastorella² & Patrizia Gasparini³

¹ Consiglio per la Ricerca e la sperimentazione in Agricoltura - Forest Monitoring and Planning Research Unit (CRA-MPF), p.za Nicolini 6 - 38123 Villazzano di Trento;
email: maria.rizzo@entecra.it.

² Consiglio per la Ricerca e la sperimentazione in Agricoltura - Forest Monitoring and Planning Research Unit (CRA-MPF), p.za Nicolini 6 - 38123 Villazzano di Trento;
email: fabio.pastorella@entecra.it.

³ Consiglio per la Ricerca e la sperimentazione in Agricoltura - Forest Monitoring and Planning Research Unit (CRA-MPF), p.za Nicolini 6 - 38123 Villazzano di Trento;
email: patrizia.gasparini@entecra.it.

Abstract

Geostatistical techniques are widely used in many sectors. In the forestry sector, the use of geostatistical approaches to predict and map forest characteristics is increasing. In this context, the research predicts the forest area in the Province of Trento starting from sampling data collected by the National Inventory of Forests and forest Carbon pools (INFC2005). A set of 6,620 georeferenced sampling points were reclassified into binary data (0 = non forest, 1 = forest and other wooded land) and kriged. To each pixel the probability that the land use/land cover is forest was assigned. Four probability maps were produced differing for pixel size (100 m and 250 m) and probability threshold (0.6 and 0.7). A Land Use map (LU) a Corine Land Cover map (CLC) and a Potential Forest Types map (PFT) were used as reference maps. Classification accuracy of the kriging maps was carried out by analyzing agreement with the reference maps. Forest maps and reference maps were compared at 10,000 randomly selected points producing cross tabulation matrices. The overall concordance coefficient, the Relative Operating Characteristic (ROC) and a McNemar statistical test were calculated for comparison.

The pixel dimension, at the scale and with the large data set used, did not influence the agreement/disagreement with reference maps, while the probability threshold was found to influence the quality of forest/non forest kriging maps. The results indicate that, for the study area, using a pure geostatistical approach, a sufficiently reliable forest/non forest probability map can be derived from inventory data.

Keywords. Geostatistical approach, kriging, forest/non forest map, forest inventories, spatial prediction model, concordance coefficient.

1 - Introduction

Geostatistic provides a set of tools and methods for modeling the spatial distribution and variability of forest attributes. Forest inventories are an important source of data for forest and natural ecosystems management and provide reliable and updated information about the forest resources. Thanks to the possibility of analyzing and predicting values of a spatially distributed variable, geostatistic is also becoming a more widely used technique in forest inventory (Sales et al., 2007; Hernandez and Emery, 2009).

In the last years, sampling data on forest/non forest classification were made available by the second Italian national forest inventory (National Inventory of Forests and forest Carbon pools, INFC2005; www.infc.it). The survey adopted a three-phase sampling design for stratification (Gasparini et al., 2010). Data were collected with reference to a sampling grid of about 300,000 cells of 1x1 km covering the whole country, with one sampling plot randomly selected within each cell of the grid (tessellated sampling; Sarndal et al., 1992). Measurements were taken during 3 different sampling phases: the first one was used to assess the land cover/land use by photo-interpretation of digital orthophotos, the second and the third phases were carried out in the field, on sub-samples of about 30,000 and 7,000 sampling units respectively, to collect qualitative and quantitative data on forest stands. The first phase classification of land cover/land use, particularly the forest/non forest classification, was conducted consistently with the first level of the CORINE Land Cover System and the FAO-Forest Resources Assessment (FRA 2000) definitions. INFC provided the Italian official estimates of forest and other wooded land areas, at the national and regional level, and national statistics on many features of forest ecosystems (total and per area unit estimates of growing stock volume and biomass, annual gross volume increment, deadwood volume and biomass, to list just the main ones).

In the present study, georeferenced information provided by INFC was used for the first time to produce forest/non forest maps using a geostatistical approach. The aim of the study is to test the possibility to build a spatial prediction model from INFC data to derive an accurate forest map of forest cover without using ancillary information.

2 – Materials and Methods

The study was carried out in the Province of Trento (North-eastern Italy) using the INFC first phase classification data from 6,620 sampling points (Fig. 1) to produce a forest/non forest distribution map. Data processing was carried out in R environment using a set of function from “gstat”, “sp”, “ggplot2” and “automap” libraries. Land use was reclassified into binary data (0 = non forest, 1 = forest and other wooded land), gridded to both a 100 m and 250 m pixel maps and interpolated by Indicator Kriging (IK).

IK essentially consists of applying ordinary kriging to indicator variables and kriging estimated is a weighted average of the surrounding values (Alley, 1993). IK was introduced by Journel (1983) for the estimation of local uncertainty derived by a local cumulative distribution function. IK is one of the few non-parametric techniques available in the literature and it did not rely upon the assumption of a particular distribution model for its results. IK is widely used as an estimation technique over a wide range of subjects: mining industry (Glacken and Blackney, 1998), insect pest management (Nestel et al., 2004), prediction of tree species distribution (Ducette, 2006; Klobucar and Pernar, 2012; Destan et al., 2013), probability of species and forest/non forest and vegetation alliances occurrence (Moeur and Hershey, 1999; Miller and Franklin, 2002), forest management and stand characteristics prediction (Gunnarsson, 1996; Wallerman et al., 2002; Pierce et al., 2009; Destan et al., 2013; Singh and Das, 2014).

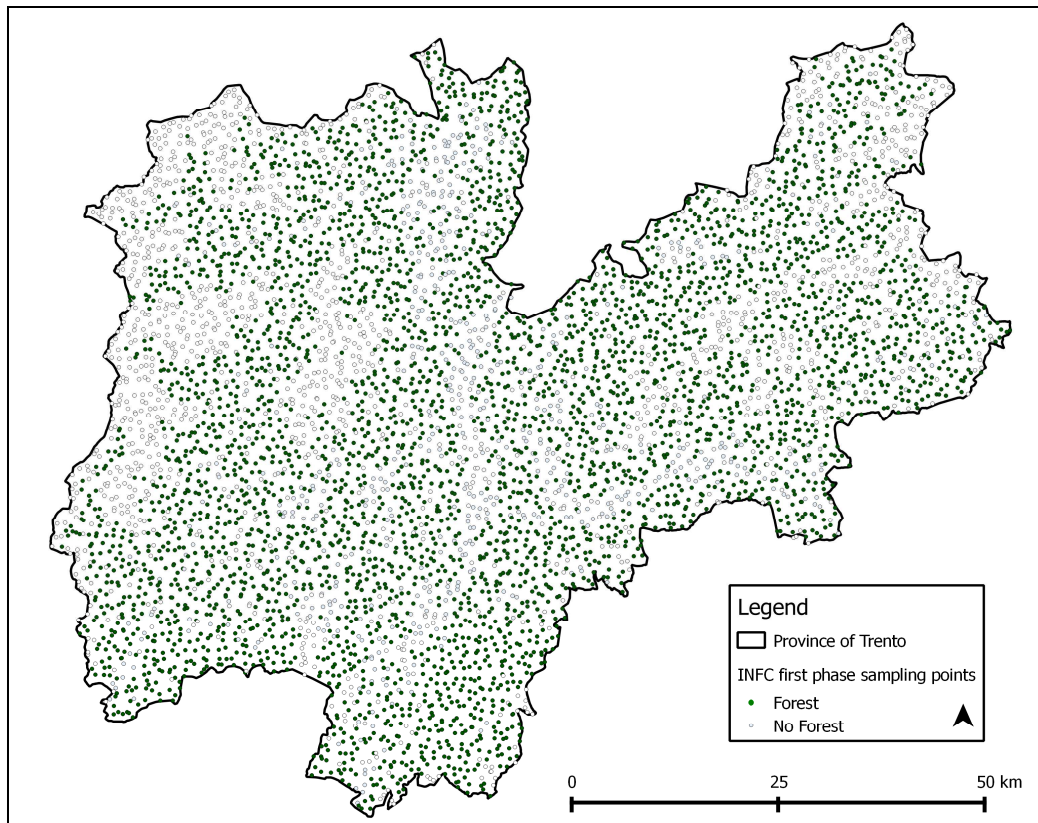


Figure 1: INFC sampling points distribution

Two different pixel sizes (100 m and 250 m) were used to produce the probability maps . The probability that the land use/land cover is forest was assigned to each pixel. Since the same dataset was used for both map elaboration, only one variogram was fitted. The variogram, shown in Fig. 2, was fitted automatically in R environment by the “autoKrige” function using a “Ste” (Matern, M. Stein's parametrization) model (nugget=0.03, psill=0.22, range = 2224 and $k=0.3$).

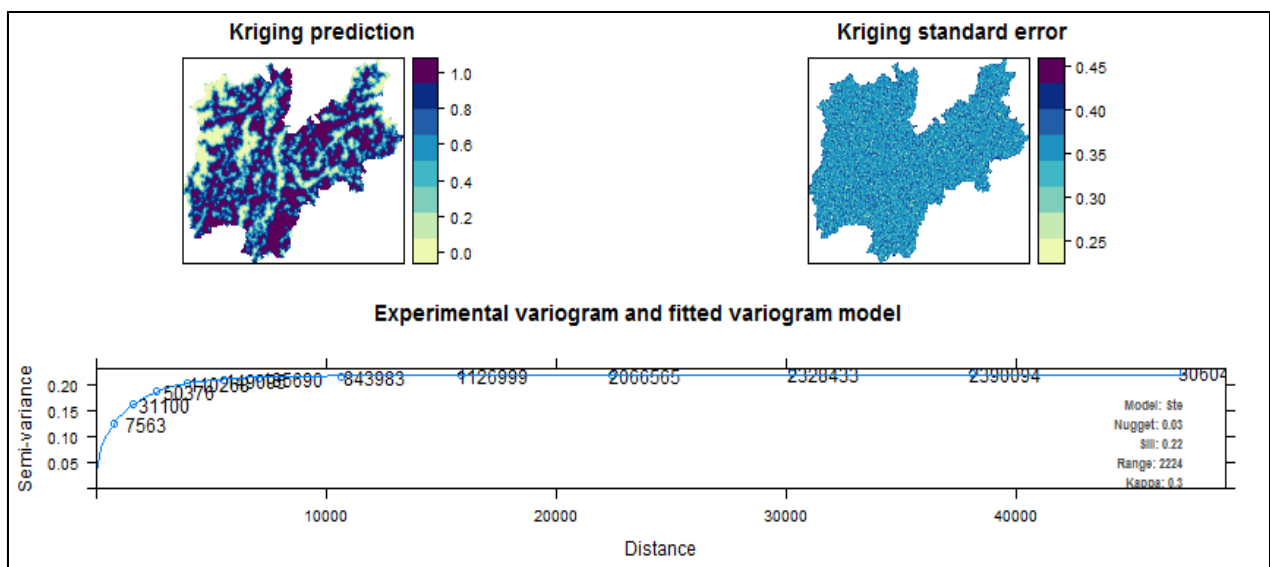


Figure 2: Auto-Kriging results

The maps were validated using three reference maps: a Land Use map (LU) produced for the Province of Trento in the year 2000 at 1:10,000 scale; the Corine Land Cover map (CLC) for the year 2000, at 1:50,000 scale; and a Potential Forest Types map (PFT) published by Forest and Wildlife Service of the Province of Trento in 2008 at 1:10,000 scale. In order to compare the data LU, CLC and PFT classifications have also been transformed into binary data (Tab. 1).

Table 1: Conversion of land use classification into binary data (CLC = Corine Land Cover map; PFT = Potential Forest Types map; LU = Land Use map; INFC2005= National Inventory of Forests and forest Carbon pools-2005)

CLC	Forest value	PFT	Forest value	LU	Forest value	INFC2005	Forest value
Airports	0	Carpinus spp. forest	1	Annual or permanent crops	0	Artificial areas	0
Annual crops associated with permanent crops	0	Quercus ilex forest	1	Broad-leaved forest	1	Agricultural areas	0
Bare rocks	0	Fraxinus ornus and Ostrya spp. forest	1	Dense Coniferous forest	1	Forest and other wooded land	1
Beaches, dunes, sands	0	Pinus mugo forest	1	Glaciers and perpetual snow	0	Grassland, pastures, uncultivated land	0
Broad-leaved forest	1	Ostrya spp. and Quercus spp. forest	1	Inland wetlands	0	Open areas with little or no vegetation	0
Complex cultivation patterns	0	Abies alba forest	1	Larix decidua and/or Pinus cembra forest	1	Wetlands	0
Coniferous forest	1	Acer spp. and Fraxinus spp. forest	1	Open spaces with little or no vegetation	0	Water bodies	0
Continuous urban fabric	0	Larix decidua forests	1	Pastures	0		
Discontinuous urban fabric	0	Acer spp. and Tilia spp. forest	1	Pinus nigra and Pinus sylvestris forest	1		
Fruit trees and berry plantations	0	Alnus spp. forest	1	Poor quality Coniferous forest	1		
Glaciers and perpetual snow	0	Pinus cembra forest	1	Shrubs and shrubs mountain pine	1		
Industrial or commercial units	0	Picea abies forest	1	Ski areas	0		
Inland marshes	0	Quercus spp. and Carpinus spp. forest	1	Sparse Coniferous forest	1		
Land principally occupied by agric., with significant areas of nat veg.	0	Pinus spp. forest	1	Sparsely vegetated areas	0		
Mineral extraction sites	0	Fagus sylvatica forest	1	Urban area	0		
Mixed forest	1	Larix decidua and/or Pinus cembra forest	1	Water courses and water bodies	0		
Moors and heathland	1	Quercus petraea forest	1	Wooded rocks	1		
Natural grasslands	0	Fraxinus spp. forest	1				
Non-irrigated arable land	0						
Pastures	0						
Sparsely vegetated areas	0						
Transitional woodland-shrub	1						
Vineyards	0						
Water bodies	0						
Water courses	0						

Results from IK and data from reference maps were compared at 10,000 randomly selected points (a random point = one pixel). The probability data resulting from kriging were converted into categorical data using as thresholds the 0.6 and 0.7 probability values (value 1 – is forest – for probability above the threshold, 0 otherwise). The probability threshold was set taking into account references (a value of 0.5 was indicated by Martinez, 2013) and empirical considerations.

Classification accuracy of maps was carried out by analysing agreement with reference maps. Based on cross tabulation matrices, the overall concordance coefficient and the Relative Operating Characteristic (ROC) defined by Pontius and Schneider (2001) were calculated. Coefficients vary between 0 and 100 indicating the proportion of pixels that are correctly (agreement) or incorrectly (disagreement) classified. ROC coefficients (allocation and quantity agreement and disagreement) were calculated using a spreadsheet tool available at <http://www.clarku.edu/~rpontius/>. Quantity disagreement and allocation disagreement were preferred instead of kappa that, according to Pontius and Millones (2011), does not give guidance on how to improve classification. As highlighted by Sarmento et al. (2012) kappa measures how much the agreement is better than random, quantity disagreement and allocation disagreement measure how much the agreement is less than perfect. In this sense the latter provide additional information that helps to explain errors. In particular, as described by Celio et al. (2014) quantity disagreement comes from a non-perfect match between the modelled and the reference proportions of the different land use categories in the case study region. Allocation disagreement is the result of a non-perfect match between the spatial distribution of the modelled and the reference land use categories given these proportions. As indicated by Estoque et al. (2012) the two parameters can be used in determining whether the achieved accuracy levels of the land-cover maps strongly support, however a specific baseline on what proportion of quantity or allocation disagreement is acceptable for a particular type of analysis, is yet to be established. In addition, the statistical significance of differences between pairs of kriging maps and between each of them and one of the reference maps was tested with McNemar test for paired nominal data using the observations in the 10,000 random points.

3 – Results and discussion

Four probability maps of forest/non forest area, differing for the probability threshold (0.6 and 0.7) at each pixel size (100 m and 250 m), were produced using the same IK procedure (Tab. 2). The overall concordances derived by comparing the set of 10,000 points of reference maps and kriging maps are shown in Table 3. The overall concordance consists of chance agreement, allocation agreement and quantity agreement. While the overall disagreement is composed by quantity disagreement and allocation disagreement.

Table 2: Kriging maps characteristics

Map	Pixel dimension (m)	Probability threshold
A	100	0.6
B	100	0.7
C	250	0.6
D	250	0.7

Kriging maps calculated with a 0.6 probability threshold show higher overall concordance than the ones calculated with a 0.7 threshold. The first ones reach overall concordance values of about 80 – 81 % while the second ones show a concordance of about 78 – 79 %. The highest agreement is reached with the PFT map in all comparisons.

Analysing the disagreement distribution between allocation and quantity components, differences between kriging maps emerge. The allocation disagreement is slightly higher for the map produced with a probability threshold of 0.6 (17 – 19 %) than for that with a 0.7 threshold (14 - 16%). The quantity disagreement is much higher for 0.7 threshold maps (about 6 - 7%) than for the 0.6 ones (about 0 – 1%). Quantity and allocation disagreement results are coherent with Estoque et al. (2012), Sloan and Pelletier (2012) and Mallinis et al. (2014) which classified forest/non forest areas from satellite images. They gained respectively values between 4 and 9 % in quantity disagreement and between 6 and 17 % in allocation disagreement.

Table 3: Comparison of concordance coefficients (CLC = Corine Land Cover map; PFT = Potential

Forest Types map; LU = Land Use map; for maps characteristics see table 2)

Map	Index of concordance	CLC	LU	PFT
A	overall concordance	80.29	80.52	81.45
	chance agreement	50.00	50.00	50.00
	quantity agreement	3.56	3.95	3.60
	allocation agreement	26.73	26.57	27.85
	allocation disagreement	18.48	19.32	17.44
	quantity disagreement	1.23	0.16	1.11
B	overall concordance	77.36	77.77	78.84
	chance agreement	50.00	50.00	50.00
	quantity agreement	1.64	1.82	1.66
	allocation agreement	25.72	25.95	27.18
	allocation disagreement	16.34	14.54	14.74
	quantity disagreement	6.30	7.69	6.42
C	overall concordance	80.20	80.35	81.30
	chance agreement	50.00	50.00	50.00
	quantity agreement	3.54	3.92	3.57
	allocation agreement	26.66	26.43	27.73
	allocation disagreement	18.68	19.38	17.70
	quantity disagreement	1.12	0.27	1.00
D	overall concordance	77.43	77.80	78.91
	chance agreement	50.00	50.00	50.00
	quantity agreement	1.65	1.83	1.67
	allocation agreement	25.78	25.97	27.24
	allocation disagreement	16.30	14.54	14.70
	quantity disagreement	6.27	7.66	6.39

As shown in Tab.4, comparing maps with different pixel sizes (A-C and B-D comparisons) we cannot reject the null hypothesis of no differences (p-values of McNemar test above the 0.05 level); on the other hand, the comparisons between A-B and C-D show p-values smaller than 0.05, indicating significant differences between kriging maps obtained with different probability thresholds.

Table 4: Results of McNemar tests for kriging maps (for maps characteristics see table 2)

Maps comparison	Variable	p-value
A - C	pixel size	0.55
B - D	pixel size	0.88
A - B	threshold	< 2.2e-16
C - D	threshold	< 2.2e-16

McNemar test was used also for paired comparisons of kriging maps and reference maps. Results confirm that threshold values have a major influence (Tab. 5). Maps produced at 0.6 threshold (A and C maps) show p-values higher than maps at 0.7 threshold (B and D maps). All comparisons show significant statistical differences, at 0.05 probability level, except A-LU and C-LU.

Table 5: Results of McNemar tests comparing kriging maps and reference maps (for maps characteristics see table 2)

Map	CLC	LU	PFT
A	0.01	0.7	0.01
B	< 2.2e-16	< 2.2e-16	< 2.2e-16
C	0.01	0.56	0.02
D	< 2.2e-16	< 2.2e-16	< 2.2e-16

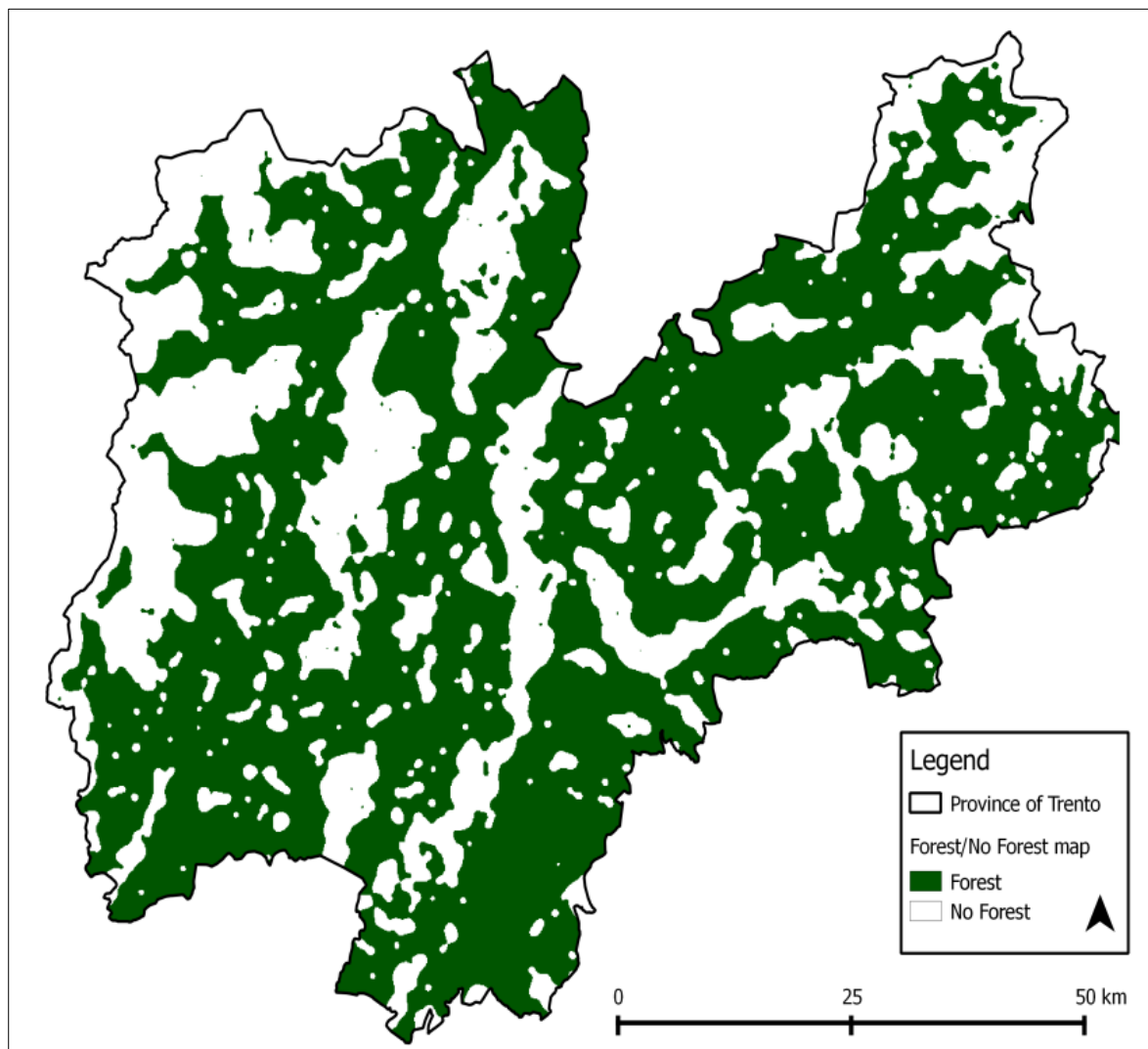


Figure 3: Kriging map (probability threshold: 0.6; pixel size: 100)

4 – Conclusions

The study is focused on the production of forest/non forest maps using a geostatistical approach applied to forest inventory data. The application of the technique could be useful to derived spatialized informations from sampling data of large scale forest inventories.

The IK was tested in a mountainous region in the North-eastern Italy characterized by an high forest area index (60.5%) with a low forest fragmentation.

The maps obtained using IK and land use/land cover classifications of the national forest inventory show a high overall concordance when compared with three different land use/land cover maps. The highest level of agreement was reached in all comparisons with the PFT map (Tab. 3), which is a map specifically focused on the forest cover type.

The probability threshold was found to influence the quality of forest/non forest kriging maps, while the pixel dimension did not influence the agreement/disagreement with reference maps at the scale and with the large data set used.

The allocation disagreement is higher for the A-C maps (0.6 probability threshold) than for the B-D maps (0.7 threshold), while the quantity disagreement is much higher in the maps obtained with a higher probability threshold.

The results demonstrate that, for the study area, a sufficiently reliable forest/non forest probability map could be derived from INFC sampling data using a pure geostatistical approach. The dataset used was large (one sampling point per km²) and the number of classes to be distinguished very small (two classes, forest – including other wooded land – and non forest). Some tests performed with a smaller data set of more detailed data on vegetation classification, derived from the INFC

second phase field survey, did not produce enough accurate maps. For the purpose of producing forest/non forest probability maps the technique used, although promising, needs to be tested in other areas characterized by different features (for example different altitude pattern, smaller proportion of forest area or higher forest fragmentation). Further analysis aimed at obtaining the best variogram model fitting and using ancillary data may increase map accuracy as well as usability of inventory field data on different variables.

Acknowledgements

Data used in the study were provided by the National Forest Service (Corpo Forestale dello Stato - CFS) that funded and carried out the Italian national forest inventory, with the scientific and technical support of CRA-MPF. The study was funded by a research grant from CRA aimed at improving the use of NFI spatial information.

References

- Alley, W. M. (1993), Geostatistical Models. In: Regional Ground-Water Quality, edited by William M. Alley, Van Nostrand Reinhold, New York, USA, 613-622.
- Arsanjani, J.J. (2012), Dynamic land use/cover change modelling: Geosimulation and multiagent-based Modelling, Springer-Verlag Berlin Heidelberg, Germany.
- Celio, E., Koellner, T., Grêt-Regamey, A. (2014), Modeling land use decisions with Bayesian networks: Spatially explicit analysis of driving forces on land use change, *Environmental Modelling and Software*, 52, 222–233.
- Chen, D. (2008), A Standardized Probability Comparison Approach for Evaluating and Combining Pixel-based Classification Procedures, *Photogrammetric Engineering & Remote Sensing*, 74, 601–609.
- Destan, S., Yilmaz, O., Sahin, A. (2013), Making objective forest stand maps of mixed managed forest with spatial interpolation and multi-criteria decision analysis, *iForest - Biogeosciences and Forestry*, 6, 268–277.
- Ducette, J. S. (2006), A Rules-based Approach to Predicting Eastern Hemlock Abundance Using Multispectral Imagery and Topographic Measures, Master of Science Degree Thesis, State University of New York.
- Estoque, R.C., Estoque, R.S., Murayama, Y. (2012), Prioritizing Areas for Rehabilitation by Monitoring Change in Barangay-Based Vegetation Cover. *ISPRS International Journal of Geo-Information*, 1, 46–68.
- Gasparini, P., Tosi, V., Di Cosmo, L. (2010), Country Report Italy. In: National Forest Inventories - Pathways for Common Reporting (eds.: Tomppo E., Gschwantner T., Lawrence M., McRoberts R.E.). Springer Science, ISBN 978-90-481-3232-4, 311-331.
- Gunnarsson, F. (1996), On the potential of Kriging for forest management planning, Second cycle, A1E. Umeå: SLU, Dept. of Forest Resource Management.
- Glacken, I. and Blackney, P. (1998), A practitioners implementation of Indicator Kriging, Proceedings of Symposium on Beyond Ordinary Kriging, Perth, Western Australia, on line:

<http://www.gaa.org.au/pdf/bok%20glacken.pdf>.

Hernandez, J. and Emery, X. (2009), A geostatistical approach to optimize sampling designs for local forest inventories, *Canadian Journal of Forest Research*, 39, 1465–1474.

Journel, A. G. (1983), Nonparametric estimation of spatial distributions, *Mathematical Geology*, 15, 445-468.

Klobučar, D. and Pern, R. (2012), Geostatistical approach to spatial analysis of forest damage, *Periodicum Biologorum*, 114, 103-110.

Mallinis, G., Galidaki, G., Gitas, I. (2014), A Comparative Analysis of EO-1 Hyperion, Quickbird and Landsat TM Imagery for Fuel Type Mapping of a Typical Mediterranean Landscape, *Remote Sensing*, 6, 1684–1704.

Moeur, M. and Hershey, R. (1999), Preserving Spatial and Attribute Correlation in the Interpolation of Forest Inventory Data. In: Lowell, K. and Jatón, A. (eds.), *Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources*, Ann Arbor Press, Chelsea, Michigan, USA, 419-430.

Magnussen, S. and de Bruin, S. (2003), Updating cover type maps using sequential indicator simulation, *Remote Sensing and Environment*, 87, 161–170.

Miller, J. and Franklin, J. (2002), Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence, *Ecological Modelling*, 157, 227–247.

Nestel, D., Carvalho, J., Nemny-Lavy, E. (2004), The Spatial Dimension in the Ecology of Insect Pests and Its Relevance to Pest Management. In: Horowitz, A. R. and Ishaaya, I., *Insect Pest Management: Field and Protected Crops*, Springer-Verlag Berlin Heidelberg, Germany, 45-64.

Pierce, K.B., Ohmann, J.L., Wimberly, M.C., Gregory, M.J., Fried, J.S. (2009), Mapping wildland fuels and forest structure for land management: a comparison of nearest neighbor imputation and other methods, *Canadian Journal of Forest Research*, 39, 1901–1916.

Pontius, R.G., and Schneider, L.C. (2001), Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts , USA, *Agriculture, Ecosystems and Environment*, 85, 239–248.

Pontius, R.G. and Millones, M. (2011), Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment, *International Journal of Remote Sensing*, 15, 4407-4429.

Sales, M.H., Souza, C.M., Kyriakidis, P.C., Roberts, D.A., Vidal, E. (2007), Improving spatial distribution estimation of forest biomass with geostatistics: A case study for Rondônia, Brazil, *Ecological Modelling*, 205, 221–230.

Sarmiento, E.C., Giasson, E., Weber, E., Flores, C.A., Hasenack, H. (2012), Prediction of soil orders with high spatial resolution: response of different classifiers to sampling density, *Pesquisa Agropecuária Brasileira*, 47, 1395–1403.

Sarndal, C.E., Swensson, B., Wretman, J. (1992), *Model assisted survey sampling*. Springer-Verlag, New York.

Singh, T. P., and Das, S. (2014), Predictive Analysis for Vegetation Biomass Assessment in Western Ghat Region (WG) Using Geospatial Techniques, *Journal of the Indian Society of Remote Sensing*, doi:10.1007/s12524-013-0335-7.

Sloan, S., Pelletier, J. (2012), How accurately may we project tropical forest-cover change? A validation of a forward-looking baseline for REDD, *Global Environmental Change*, 22, 440–453.

Solaimani, K., Arekhi, M., Tamartash, R., Miryaghobzadeh, M. (2010), Land use/cover change detection based on remote sensing data (A case study; Neka Basin), *Agriculture and Biology Journal of North America*, 1, 1148–1157.

Wallerman, J., Joyce, S., Vencatasawmy, C.P., Olsson, H. (2002), Prediction of forest stem volume using kriging adapted to detected edges, *Canadian Journal of Forest Research*, 32, 509–518.