**Published by AIT
the Italian Society of Remote Sensing**

K-NN FOREST: a software for the non-parametric prediction and mapping of environmental variables by the *k*-Nearest Neighbors algorithm

Gherardo Chirici[1]*, Piermaria Corona[2], Marco Marchetti[1], Alessandro Mastronardi[3], Fabio Maselli[4], Lorenzo Bottai[5] and Davide Travaglini[6]

[1] Dipartimento di Bioscienze e Territorio, Università degli Studi del Molise, Contrada F.te Lappone snc, I-86090 Pesche, Isernia (Italy)
[2] Dipartimento per l'Innovazione nei Sistemi Biologici, Agroalimentari e Forestali, Università degli Studi della Tuscia, Via San Camillo de Lellis snc, I-01100 Viterbo (Italy)
[3] GM WEB STUDIO, Via Modigliani 23, I-50127 Firenze (Italy)
[4] IBIMET-CNR, Via Madonna del Piano 10, I-50123 Sesto Fiorentino, Firenze (Italy)
[5] Lamma Regione Toscana, Via Madonna del Piano 10, I-50123 Sesto Fiorentino, Firenze (Italy)
[6] Dipartimento di Economia, Ingegneria, Scienze e Tecnologie Agrarie e Forestali, Università degli Studi di Firenze, Via S. Bonaventura 13, I-50145 Firenze (Italy)
*Corresponding author, e-mail address: gherardo.chirici@unimol.it

Abstract

In the last decades researchers investigated the possibility of extending the information collected in sampling units during a field survey to wider geographical areas through the use of remotely sensed images. One of the most widely adopted approaches is based on the non-parametric *k*-Nearest Neighbors (*k*-NN) algorithm. This contribution describes the software K-NN FOREST we developed to provide a complete tool for the implementation of the *k*-NN technique to generate spatially explicit estimations (maps) of a response variable acquired in the field by sampling units through the use of remotely sensed data or other ancillary variables. K-NN FOREST is designed to guide the user through a graphic user interface in the different phases of the process. K-NN FOREST is freely available for download and it is designed to run under Windows environment in conjunction with the GIS software IDRISI.

Keywords: Environmental Inventory and Mapping, Prediction, Remote Sensing, *k*-Nearest Neighbors.

Introduction

Managing natural resources requires a large amount of georeferenced information [Corona et al., 2011]. The acquisition of thematic information regarding biophysical attributes of terrestrial ecosystems is traditionally based on two distinct approaches: inventorying or

mapping [Corona, 2010].

A given geographical area may be divided in a finite number of geographical units of finite dimension constituting a statistical population of N units. The investigated variable is called response variable Y.

Inventories investigate the response variable in a small number (n) of units extracted in the form of a sample from the population N. Then, the information from the sample is inferred to the population through the use of estimators. The result of an inventory is an estimation of a given statistical parameter of the response variable Y. In contrast, maps investigate all units of the population and can thus be considered similar to a geographical census. Inventories are mainly carried out on the basis of field work while thematic maps are usually based mainly on the classification of remotely sensed images acquired by aerial or satellite platforms.

In the last decades researchers investigated the possibility of extending the information collected in the field by the n sampling units during an environmental (e.g. forest) inventory to the remaining N-n geographical units through the use of remotely sensed images. This approach can be interpreted as a bridging tool to move from inventories to mapping. For such a purpose the geographic units N are usually the pixels of remotely sensed digital images.

A vast literature exists describing different approaches developed on the basis of several types of remotely sensed data, using different algorithms and with the intent of estimating a number of different response variables measured in the field.

All studies underline the need for the existence of significant relationships between the variables Xi acquired by remote sensing (called feature space variables or covariate variables) and the response variable of interest Y [McRoberts and Tomppo, 2007]. Similar approaches may be based on the thematic information contained in digital maps instead of, or together with, remotely sensed images; in these cases the covariate variables are sometimes called ancillary variables.

Some of the most promising results for this purpose were achieved by the k-Nearest Neighbors (k-NN) method. In 2006 and 2007 two workshops were organized to discuss methods and approaches for k-NN predictions: proceedings are available on line at http://knn.gis.umn.edu. From those discussions it was clear that a major limitation to the practical application of k-NN methods for supporting environmental management and strategic planning lied in the unavailability of a specific scientific software.

As a matter of fact, the k-NN technique is well known and is implemented in most commercial statistical packages such as SPSS [SPSS, 2007] and in Geographic Information Systems such as IDRISI [Clark Labs, 2012]. In the case of statistical software packages such as SPSS, however, the application of the k-NN in a geographic environment is difficult because input geocoded images cannot be directly handled. In the case of GIS, such as IDRISI, the k-NN is instead mainly implemented to solve classification problems of a qualitative response variables, i.e. typically to derive classified maps on the basis of remotely sensed images.

Source codes implementing the general k-NN algorithm are also available on-line and distributed within open-source frameworks in C++ [Gulli, 2012], in Java [Ferreira, 2012; Thisara, 2012], in Visual Basic [Lammertsma, 2004] and even on the basis of Excel [Teknomo, 2012]. More advanced and complete suites for the use of k-NN method exist

too. Li [2003] developed a suite in C code for microarray and proteomic applications. Weka-knn is instead an implementation of the *k*-NN algorithm in Weka, a collection of machine learning algorithms for data mining tasks [Hall et al., 2009]. Tilburg Memory Based Learner (TiMBL) is an open source software package implementing several memory-based learning algorithms, including *k*-NN [Daelemans et al., 2010]. A special mention should be given to yaImpute, an R package for nearest neighbor search and imputation which offers a number of different tools facilitating the comparison among different nearest neighbor search algorithms and subsequent imputation techniques [Crookston and Finley, 2008].

A brief review indicates that, after two decades of scientific experiments, the *k*-NN method is ready to become an advanced tool for supporting the operational work of environmental resource managers and planners, who need practical and easy to use software [McRoberts et al., 2011]. In spite of this none of the currently available *k*-NN software applications have at the same time the characteristics of completeness and usage easiness granted by an intuitive graphic user interface (GUI).

This contribution describes the software K-NN FOREST that was recently developed by our research group. This is a complete tool which implements the *k*-NN technique to create the spatially explicit predictions (maps) of a response variable acquired in field sampling units through the use of remotely sensed or ancillary data. The software is designed to run under Windows environment. It has a complete GUI to guide the user in the different phases of the procedure. K-NN FOREST is based on the use of some modules from the IDRISI GIS software, both software packages must be installed on the same PC. All data required for using the software must be available in standard GIS formats. K-NN FOREST is provided free of charge for non-profit utilization together with a manual and a tutorial with sample data.

The k-*NN algorithm*

The first original extensive tests of the *k*-NN algorithm for predicting environmental reference variables on the basis of remote sensing feature space variables were probably developed within the Finnish National Forest Inventory [Tomppo, 1991; Tokola et al., 1996]. Here we refer to the algorithm configuration implemented in K-NN FOREST. The unknown value of the response variable Y_t for each target geographical unit t (i.e. a pixel or a group of pixels like that obtained by image segmentation) can be predicted using the values Y_i of the same variable measured in the field in locations corresponding to the k nearest neighbor units:

$$\hat{Y}_t = \frac{\sum_{i=1}^{k} W_{t,i} Y_i}{\sum_{i=1}^{k} W_{t,i}}$$

where the weight W is inversely proportional to the distance between the units t and i measured in the multidimensional feature space created on the basis of the feature space variables.

The multidimensional distance can be calculated by several measures.

The simplest is the Euclidean Distance (ED) calculated as:

$$ED = \sum_{j=1}^{m} (Xr_j - Xt_j)^2$$

where m is the number of feature space variables, Xr_j is the value of j-th feature space variable in one of the n geographic units (pixels or group of pixels) where the response variable is available and Xt_j is the value of j-th feature space variable in one of the N-n target geographic units (pixels or group of pixels) where the response variable is not available.

To take into account the multicollinearity of feature space variables two modifications of the Euclidean distance are implemented in K-NN FOREST: the Mahalanobis distance (MD) and the fuzzy distance (FD).

The Mahalanobis Distance (MD) considers the variance–covariance matrix of the feature space variables, C [Richards, 1993]. MD between reference and target pixels is computed as:

$$MD = (Xr - Xt)' C^{-1} (Xr - Xt)$$

where Xr and Xt are the n-dimensional feature space vectors of the reference and target pixels, respectively, and C can be derived from the reference data set.

As can be easily understood, neither ED or MD emphasizes the relationship of the feature space variables to the response variable. In order to perform this task, modified forms of multidimensional distances have been proposed aiming at giving preferential consideration to the most informative feature space variables [Holmström et al., 2001; Maselli et al., 2005].

The distance weighted with fuzzy weights or Fuzzy Distance (FD) is a modification of MD where the variance–covariance matrix is computed via a fuzzy approach. The fuzzy variance–covariance matrix C^* is obtained by weighting the reference pixels through a fuzzy membership function that gives preferential consideration to the reference pixels having values of the response variable close to the mean [Maselli, 2001]. This is obtained deriving the membership grade of each reference pixel i, Fg_i, from the Gaussian probability density function of the response variable, through [Anderson, 1984]:

$$Fg_i = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(Y_i - \mu)^2}{2\sigma^2}}$$

where: Yi is the value of the response variable at reference pixel i; μ is the mean value of the response variable; σ^2 is the variance of the response variable.

This modification results in a reduction of the variance–covariances for the feature space variables most correlated with the response variable, and thus most informative [Wang, 1990; Maselli, 2001]. Consequently these variables become the most important in the calculation of FD between reference and target pixels through:

$$FD = (Xr - Xt)' C^{*-1} (Xr - Xt)$$

Validation of k-*NN predictions*

Several approaches can be used to assess the accuracy of *k*-NN predictions, but they are all based on the comparison between the response variable measured in the field and the *k*-NN predictions. These approaches can differ for two possible reasons: *i*) the accuracy may be computed on the same dataset used to derive the predictions or on an independent dataset, and *ii*) the validation may be performed at pixel level or for regions made of aggregated pixels.

Since external datasets for validation are not always easily available, bootstrapping techniques are often applied in which the observations from a subset of the reference set are used to calculate predictions for the other subset [McRoberts, 2008]. Accuracy is assessed by comparing the observations and predictions for the second subset. The leave-one-out (LOO) technique is a special case of bootstrapping in which a prediction is iteratively calculated for each individual element of the reference set using the remaining reference set observations [Lachenbruch and Mickey, 1986]. When dealing with continuous variables accuracy is assessed by comparing the observations and predictions for all reference set elements using the Root Mean Squared Errors or an error similar measure [Franco-Lopez et al., 2001].

K-NN FOREST

The K-NN FOREST software is intended to be used in conjunction with the IDRISI GIS software [Clark Labs, 2012]. The user should therefore first prepare all input datasets within the IDRISI environment before starting to use K-NN FOREST. The implementation of the *k*-NN prediction procedure may be subdivided in four phases.

Data preparation

This part of the work must be accomplished before using K-NN FOREST in order to have input datasets ready for the process. Four input datasets can be used, two of them are optional. The first is the set of feature space variables that are usually constituted of the different bands of a multispectral remotely sensed image. All the bands in IDRISI file format must be geographically coregistered with the same number of rows and columns. The software accepts a maximum of 256 input feature space variables.

The second dataset is the response variable in IDRISI raster format which must have the same geographic extent and geometric resolution of the feature space variables files. The response variable is usually constituted of the rasterized sampling units that are acquired and georeferenced in the field by means of a GPS. The response variable file must be accurately coregistered with the feature space variables.

Third, optionally one ancillary variable may be used: again it must be in the IDRISI raster format, and it must have the same geographic characteristics of the feature space variables. The ancillary variable (for example a Digital Elevation Model, DEM, see below) may be used to restrict the selection of the *k* nearest sampling units.

Finally, the *k*-NN prediction may be limited to a subset (defined as target set) of the geographic area represented by the feature space variables. For this reason the user may optionally consider a target boolean raster file where the pixels excluded from the target set are identified. This target file must also have the same geographic characteristics of the feature space variables.

Feature space creation

K-NN FOREST first requires a few general information which serve to save the project coherently with the IDRISI project structure. The user is then asked for the files to be used as input feature space variables, the response variable file, the statistical parameters to be extracted from the feature space variables for each sampling unit of the response variables, and, optionally, the ancillary file (Fig. 1). For each sampling unit of the response variable file the software extracts the selected statistics from the feature space variables and, optionally, from the ancillary file. This information is stored in a database in DBF format having a number of records equal to the number of sampling units in the response variable file and the fields filled with the response variable, the statistical parameters extracted from the feature space variables and, optionally, the statistical parameter extracted from the ancillary file.
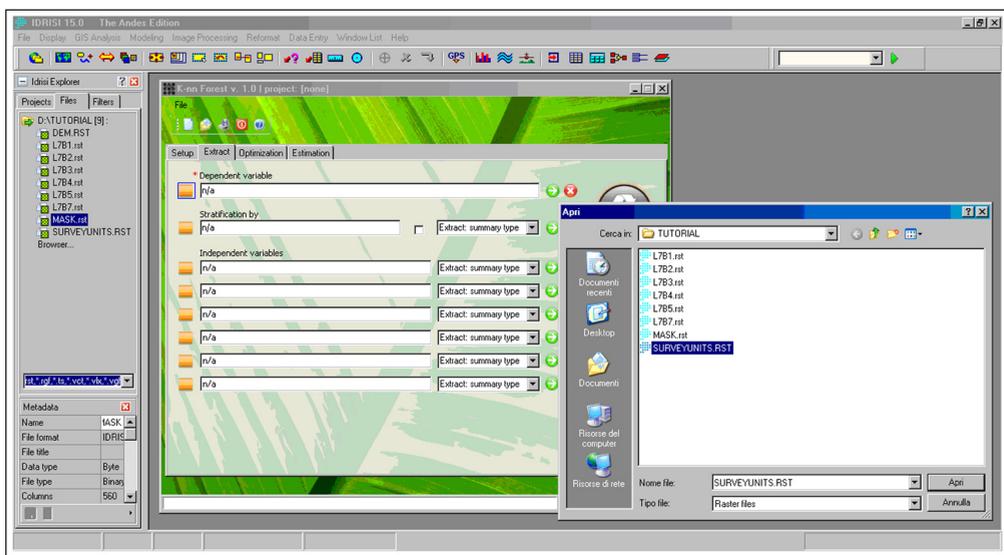

Figure 1 - The graphic user interface of K-NN FOREST in the input data phase, with the IDRISI software environment on the background.

k-*NN optimization*

On the basis of the database created in the previous phase K-NN FOREST estimates the response variable for all sampling units through a leave-one-out (LOO) cross validation procedure; the user may set the type of the multidimensional distance (ED, MD or FD) and the maximum number of k to be considered. For each value of k from 1 to the maximum value set by the user the software calculates the Spearman coefficient of correlation (r) and the Root Mean Square Error (RMSE) of the predictions comparing the k-NN outputs obtained by LOO to the observed values of the response variable.

The software also includes two optional possibilities for restricting the searching of the k units. The first is based on a geographic distance (or horizontal reference area). The user may set a maximum search distance, within which the software will search for the nearest k units. The second option is based on searching the k units within a thematic distance

measured on the basis of the ancillary variable. For example using a DEM as ancillary variable it is possible to restrict the search for the nearest *k* units within a defined vertical distance [Katila and Tomppo, 2001].

The output of each K-NN FOREST optimization phase is an ASCII file reporting the Spearman correlation coefficient and the RMSE determined by the LOO cross validation procedure for each *k* value set by the user. In general the accuracy of the *k*-NN predictions increases with increasing *k* values. On the basis of this optimization phase, the user has to decide the best configuration of *k*-NN in terms of feature space variables to be used, type of feature space distance, *k* value, inclusion of an ancillary variable or a geographic search criterion. These settings are critical since different *k*-NN configurations may led to very different results in terms of prediction accuracy [Chirici et al., 2008].

k-*NN prediction*
The user is finally asked to run K-NN FOREST with the optimal configuration empirically defined in the previous phase. K-NN FOREST performs the prediction of the response variable for each pixel of the target set. The output is a raster image in IDRISI file format. The accuracy of the *k*-NN predictions can be calculated at pixel level or for larger aggregated areas on the basis of a number of different methodologies provided by standard GIS environments [Kim and Tomppo, 2006; Chirici et al., 2008].

Conclusions
In this paper we presented the software called K-NN FOREST, which is implemented under Windows environment developed for the spatial prediction of a quantitative response variable on the basis of a set of feature space variables. The software is based on the non-parametric *k*-NN algorithm and it is specifically designed to spatialize the variables observed in field sampling units over a large territory through the use of multispectral remotely sensed images and ancillary data.

The software is driven through a GUI which guides the user through the different phases of the method, from input data formatting, to algorithm optimization by LOO cross validation, and to final spatial prediction of the response variable. The software K-NN FOREST was extensively tested in several research applications [Chirici et al., 2008; Chirici et al., 2010; Lasserre et al., 2011; Mattioli et al., 2012] and it is now freely distributed on-line.

Private or public stakeholders involved in environmental monitoring and assessment may benefit of this software to derive spatial predictions of environmental attributes useful for quantifying the conditions and trends of environmental resources [Corona, 2010].

K-NN FOREST users should be conscious of the theoretical framework consolidated in recent literature concerning the use of the *k*-NN algorithm for predicting and mapping environmental response variables. Before deciding to use K-NN FOREST preliminary tests should be carried out for ascertaining the validity of the assumptions underlying this technique [McRoberts, 2009; Baffetta et al., 2011]. For example, the feature space variables are usually derived from remotely sensed acquisitions while the response variable is collected in field sampling units georeferenced through GPS. The two datasets must be coregistered with an error that should be lower than the geometric resolution of the analysis, usually coincident with the pixel dimension of the feature space variables. Halme and Tomppo [2001] demonstrated that this prerequisite is essential for obtaining accurate

predictions of the response variable.

Finally, the results obtained with *k*-NN can be reported together with model based [Magnussen et al., 2009] or design based [Baffetta et al., 2009] estimates of the uncertainty contained in pixel-level or spatially aggregated *k*-NN predictions.

We distinctively hope that K-NN FOREST will help reducing the distance between forest mapping and inventorying enabling more close cooperation among specialists of different disciplines. More broadly, the software should promote the operational use of the *k*-NN method by environmental researchers, technicians, professionals and managers who may not have specific programming skills.

References

Anderson T.W. (1984) - *An Introduction to Multivariate Statistical Analysis*. New York, John Wiley & Sons.

Baffetta F., Fattorini L., Franceschi S., Corona P. (2009) - *Design-based approach to k-nearest neighbours technique for coupling field and remotely sensed data in forest surveys*. Remote Sensing of Environment, 113 (3): 463-475. doi: http://dx.doi.org/10.1016/j.rse.2008.06.014.

Baffetta F., Corona P., Fattorini L. (2011) - *Design-based diagnostics for k-NN estimators of forest resources*. Canadian Journal of Forest Research, 41: 59-72. doi: http://dx.doi.org/10.1139/X10-157.

Chirici G., Barbati A., Corona P., Marchetti M., Travaglini D., Maselli F., Bertini R. (2008) - *Non-parametric and parametric methods using satellite images for estimating growing stock volume in alpine and mediterranean forest ecosystems*. Remote Sensing of Environment, 112 (5): 2686-2700. doi: http://dx.doi.org/10.1016/j.rse.2008.01.002.

Chirici G., Corona P., Marchetti M., Tonti D., Travaglini D. (2010) - *Biomass estimation by satellite data and ground measurements*. In: Miranda D., Suarez J., Rafael C. (Eds.), Forestsat2010, Lugo (Spain), pp. 42-45.

Corona P. (2010) - *Integration of forest inventory and mapping to support forest management*. iForest - Biogeosciences and Forestry 3: 59-64.

Corona P., Chirici G., McRoberts R.E., Winter S., Barbati A. (2011) - *Contribution of large-scale forest inventories to biodiversity assessment and monitoring*. Forest Ecology and Management, 262 (11): 2061-2069. doi: http://dx.doi.org/10.1016/j.foreco.2011.08.044.

Clark Labs (2012) - *IDRISI Selva GIS and Image Processing Software*. Available at: http://www.clarklabs.org (last accessed 14/04/2012).

Crookston N.L., Finley A.O. (2008) - *yaImpute: an R package for kNN imputation*. Journal of Statistical Software, 23: 62-72.

Daelemans W., Zavrel J., Van der Sloot K., Van den Bosch A. (2010) - *TiMBL: Tilburg Memory Based Learner, version 6.3, Reference Guide*. ILK Research Group Technical Report Series no. 10-01.

Franco-Lopez H., Ek A.R., Bauer M.E. (2001) - *Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method*. Remote Sensing of Environment, 77: 251-274. doi: http://dx.doi.org/10.1016/S0034-4257(01)00209-7.

Ferreira L.N. (2012) - *K-Nearest Neighbor (KNN) Classifier*. Avaiable at: http://www.leonardonascimento.com/en/knn.html.s.

Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I.H. (2009) - *The WEKA Data Mining Software: An Update*. SIGKDD Explorations, 11 (1). doi: http://dx.doi.org/10.1145/1656274.1656278.

Halme M., Tomppo E. (2001) - *Improving the accuracy of multisource forest inventory estimates to reducing plot location error, a multicriteria approach*. Remote Sensing of Environment, 78: 321-327. doi: http://dx.doi.org/10.1016/S0034-4257(01)00227-9.

Gulli A. (2012) - *Nearest Neighbour on KD-Tree in C++ and Boost*. Available at: http://codingplayground.blogspot.it/2010/01/nearest-neighbour-on-kd-tree-in-c-and.html (last accessed 10/04/2012).

Holmstrom H., Nilsson M., Ståhl G. (2001) - *Simultaneous estimations of forest parameters using aerial photograph-interpreted data and the k nearest neighbor method*. Scandinavian Journal of Foresr Research, 16: 67-78. doi: http://dx.doi.org/10.1080/028275801300004424.

Katila M., Tomppo E. (2001) - *Selecting estimation parameters for the Finnish multisource National Forest Inventory*. Remote Sensing of Environment, 76: 16-32. doi: http://dx.doi.org/10.1016/S0034-4257(00)00188-7.

Kim H.-J., Tomppo E. (2006) - *Model-based prediction error uncertainty estimation for k-nn method*. Remote Sensing of Environment, 104: 257-263. doi: http://dx.doi.org/10.1016/j.rse.2006.04.009.

Lachenbruch P.A., Mickey M.R. (1986) - *Estimation of error rates in discriminant analysis*. Technometrics, 10: 1-11. doi: http://dx.doi.org/10.1080/00401706.1968.10490530.

Lammertsma P. (2004) - *K-nearest-neighbor algorithm*. Available at: http://paul.luminos.nl/documents/show_document.php?d=197 (last accessed 6/05/2012).

Lasserre B., Chirici G., Chiavetta U., Garfì V., Tognetti R., Drigo R., DiMartino P., Marchetti M. (2011) - *Assessment of potential bioenergy from coppice forests trough the integration of remote sensing and field surveys*. Biomass and Bioenergy, 35(1): 716-724. doi: http://dx.doi.org/10.1016/j.biombioe.2010.10.013.

Li L. (2003) - *Variable selection and sample classification using a genetic algorithm and k-nearest neighbors (GA/KNN) method*. Available at: http://www.niehs.nih.gov/research/resources/software/gaknn (last accessed 1/05/2012).

Magnussen S., McRoberts R.E., Tomppo E.O. (2009) - *Model-based mean square error estimators for k-nearest neighbour predictions and applications using remotely sensed data for forest inventories*. Remote Sensing of Environment, 113: 476-488. doi: http://dx.doi.org/10.1016/j.rse.2008.04.018.

Maselli F. (2001) - *Extension of environmental parameters over the land surface by improved fuzzy classification of remotely sensed data*. International Journal of Remote Sensing, 22: 3597-3610. doi: http://dx.doi.org/10.1080/01431160010006458.

Maselli F., Chirici G., Bottai L., Corona P., Marchetti M. (2005) - *Estimation of Mediterranean forest attributes by the application of k-NN procedures to multitemporal Landsat ETM+ images*. International Journal of Remote Sensing, 17: 3781-3796. doi: http://dx.doi.org/10.1080/01431160500166433.

Mattioli W., Quatrini V., Di Paolo S., Di Santo D., Giuliarelli D., Angelini A., Portoghesi P., Corona P. (2012) - *Experimenting the design-based k-NN approach for mapping and estimation under forest management planning*. iForest - Biogeosciences and Forestry, 5: 26-30.

McRoberts R.E. (2008) - *Using satellite imagery and the k-nearest neighbors technique as a bridge between strategic and management forest inventories*. Remote Sensing of Environment, 112: 2212-2221. doi: http://dx.doi.org/10.1016/j.rse.2007.07.025.

McRoberts R.E. (2009) - *Diagnostic tools for nearest neighbors techniques when used with satellite imagery.* Remote Sensing of Environment, 113: 489-499. doi: http://dx.doi.org/10.1016/j.rse.2008.06.015.

McRoberts R.E., Tomppo E.O. (2007) - *Remote sensing support for national forest inventories.* Remote Sensing of Environment, 110: 412-419. doi: http://dx.doi.org/10.1016/j.rse.2006.09.034.

McRoberts R.E., Magnussen S., Tomppo E.O., Chirici G. (2011) - *Parametric, bootstrap, and jackknife variance estimators for the k-Nearest Neighbors technique with illustrations using forest inventory and satellite image data.* Remote Sensing of Environment, 115: 3165-3174. doi: http://dx.doi.org/10.1016/j.rse.2011.07.002.

Richards J.A. (1993) - *Remote sensing digital image analysis: An introduction (2nd ed.).* Heilderberg: Springer-Verlag, 340 pp.

SPSS (2007) - *SPSS Statistics Base 17.0. User's Guide*. SPSS Inc. Available at: www.spss.com (last accessed 31/03/2012).

Teknomo K. (2012) - *K Nearest Neighbors Tutorial.* Available at: http://people.revoledu.com/kardi/tutorial/KNN/index.html (last accessed 28/03/2012).

Thisara F. (2012) - *Optimized Nearest Neighbor Framework.* Available at: http://code.google.com/p/thisara/people/list (last accessed 1/05/2012).

Tokola T.J., Pitkänen S., Muinonen E. (1996) - *Point accuracy of a nonparametric method n estimation of forest characteristics with different satellite materials.* International Journal of Remote Sensing, 17 (12): 2333-2351. http://dx.doi.org/10.1080/01431169608948776.

Tomppo E. (1991) - *Satellite image-based national forest inventory of Finland*. International Archives of Photogrammetry and Remote Sensing, 28 (7-1): 419-424.

Wang F. (1990) - *Fuzzy supervised classification of remote sensing images.* IEEE Transactions on Geoscience and Remote Sensing, 28: 194–201. doi: http://dx.doi.org/10.1109/36.46698.