

Comparative Analysis of Neural Networks and Statistical Approaches to Remote Sensing Image Classification

Nataliya Kussul

Department of Space Information
Technologies and Systems,
Space Research Institute NASU-NSAU,
Glushkov Ave 40, Kyiv, 03650,
Ukraine, E-mail: inform@ikd.kiev.ua

Serhiy Skakun

Department of Space Information
Technologies and Systems,
Space Research Institute NASU-NSAU,
Glushkov Ave 40, Kyiv, 03650,
Ukraine, E-mail: inform@ikd.kiev.ua

Olga Kussul

Physics and Technology Institute,
National Technical
University of Ukraine "KPI",
Peremoga Ave 37, Kyiv, 03056,
Ukraine, E-mail: ok_olga@ukr.net

Abstract – This paper examines different approaches to remote sensing images classification. Included in the study are statistical approach, namely Gaussian maximum likelihood classifier, and two different neural networks paradigms: multilayer perceptron trained with EDBD algorithm, and ARTMAP neural network. These classification methods are compared on data acquired from Landsat-7 satellite. Experimental results showed that to achieve better performance of classifiers modular neural networks and committee machines should be applied.

Keywords - remote sensing image classification, neural networks, statistical methods, Landsat-7 satellite.

I. INTRODUCTION

Recent advances in technologies made it possible to develop new satellite sensors with considerably improved parameters and characteristics. For example, the spectral resolution increased up to 144 channels as in Hyperion sensor; radiometric resolution increased up to 14 bits as in MODIS sensor, etc. In turn, the use of such space-borne satellite sensors enables acquisition of valuable data that can be efficiently used for various applied problems solving in agriculture, natural resources monitoring, land use management, environmental monitoring, and so on.

Land cover classification represent one of the most important and typical applications of remote sensing data. Land cover corresponds to the physical condition of the ground surface, for example, forest, grassland, artificial surfaces etc. To this end, various approaches have been proposed, among which the most popular are neural networks [1] and statistical [2] methods.

In this paper different approaches to remote sensing images classification are examined. The following approaches are included in the study: statistical approach, namely Gaussian maximum likelihood (ML) classifier [2], and two different types of neural networks: feed-forward multilayer perceptron (MLP) and ARTMAP neural network [3]. MLP is trained by means of Extended-Delta-Bar-Delta (EDBD) algorithm [4] which represent a fast modification of standard error

backpropagation algorithm [5]. In turn, ARTMAP belongs to the family of adaptive resonance theory (ART) networks [6], which are characterized by their ability to carry out fast, stable, on-line learning, recognition, and prediction.

Comparative analysis of classification methods is done on data acquired by Enhanced Thematic Mapper Plus (ETM+) sensor of Landsat-7 satellite [7], and land cover data from European Corine project [8].

II. OVERVIEW OF RELATED WORKS

Nowadays, various approaches have been proposed to land cover classification of remote sensing data. In past classification has traditionally been performed by statistical methods (e.g., Bayesian and k-nearest-neighbor classifiers). In recent years, the remote sensing community has become interested in applying neural networks to data classification. Neural networks provide an adaptive and robust approach for the analysis and generalization of data with no need of a priori knowledge on statistical distribution of data. It is particularly important for remote sensing image classification since information is provided by multiple sensors or by the same sensor in many measuring contexts. It is the main problem associated with most statistical models, since it is difficult to define a single model for different types of space-bourn sensors [9]. In this section we give a brief overview of approaches to remote sensing data classification.

In [10] classification of remote sensing data was done using MLP. The main goal was the investigation of applicability of MLP to the classification of terrain radar images. MLP performances were compared with those of a Bayesian classifier, and it was found that significant improvements can be obtained by the MLP classifier.

Benediktsson et al. [9] applied MLP to the classification of multisource remote sensing data. In particular, Landsat MSS and topographic data were considered.

Classification performances were compared with those of a statistical parametric method that takes into account the relative reliabilities of the sources of data. They concluded that the relative performances of the two methods mainly depend on priori knowledge about the statistical distribution of data. MLPs are appropriate for cases where such distributions are unknown, for they are data-distribution-free. The considerable training time required is one of the main drawbacks of MLP, compared with statistical parametric methods.

Bischof et al. [11] reported the application of a three-layer perceptron for classification of Landsat TM data. They compared MLP performances with those of Bayesian classifier. The obtained results showed that the MLP performs better than Bayesian classifier.

Dawson and Fung [12] reviewed examples of the use of MLP to classification of remote sensing data. In their study they proposed an interesting combination of clustering algorithms and scattering models to train MLP when no ground truth is available.

Roli et al. [13] proposed a type of structured neural networks (*treelike networks*) to multisource remote sensing data classification. This kind of architecture allows one to interpret the network operations. For example, the roles played by different sensors and by their channels can be explained and quantitatively assessed. The proposed method was compared with fully connected MLP and probabilistic neural networks on images acquired by synthetic aperture radar (SAR) sensor.

Carpenter et al. [14] described the ARTMAP information fusion system. The fusion system uses distributed code representations that exploit the neural network's capacity for one-to-many learning in order to produce self-organizing expert systems that discover hierarchical knowledge structures. The fusion system infers multi-level relationships among groups of output classes, without any supervised labeling of these relationships. The proposed approach was tested on two testbed images, but not limited to the image domain.

In [15] various algorithms are examined in order to estimate mixtures of vegetation types within forest stands based on data from the Landsat TM satellite. The following methods were considered in that study: maximum likelihood classification, linear mixture models, and a methodology based on the ARTMAP neural network. The reported experiments showed that ARTMAP mixture estimation method provides the best estimates of the fractions of vegetation types comparing to others.

Hwang et al. [16] described a structured neural network to classify Landsat-4 TM data. A *one-network one-class* architecture is proposed to improve data separation. Each network is implemented by radial basis function (RBF) neural network. The proposed approach outperformed other methodologies, such as MLP and a Bayesian classifier.

III. METHODOLOGY

In this section we give a brief overview of methodologies that will be compared for remote sensing image classification.

A. MLP trained with EDBD

MLP represent a kind of feed-forward neural networks in which all the connections are unidirectional. MLP consists of an input layer, output layer, and at least one hidden layer of hidden neurons. Unidirectional connections exist from the input layer to hidden layer and from hidden layer to output neurons. There are no connections between any neurons within the same layer.

Error backpropagation algorithm [5] is a popular method for MLP training, i.e. for neural networks weights adjustment. However, despite its widespread use for many applications, it has a drawback of considerable training time required. That is why in this study we use a fast modification of error backpropagation method Extended-Delta-Bar-Delta (EDBD) rule [4]. This algorithm is based on the following heuristics:

- On each step of training process learning rate and momentum factor are automatically estimated for each neural network weight. On the first step initial and maximum values for learning rates and momentum are set, and remain constant during the whole training process.
- If partial derivative of error preserves its sign (positive or negative) within some training steps, then learning rate and momentum for corresponding weight increases.
- If partial derivative of error changes its sign within some training steps, then learning rate and momentum for corresponding weight decreases.

More detailed description of EDBD algorithm can be found in [1, 4]. In this study for EDBD simulations we use MNN CAD software [17].

B. ARTMAP neural networks

ARTMAP belongs to the family of ART networks [6], which are characterized by their ability to carry out fast, stable, on-line learning, recognition, and prediction. These features differentiate ARTMAP from the family of feed-forward MLPs, including backpropagation, which typically require slow learning. ARTMAP systems self-organize arbitrary mappings from input vectors, representing features such as spectral values of remote

sensing images and terrain variables, to output vectors, representing predictions such as vegetation classes or environmental variables. Internal ARTMAP control mechanisms create stable recognition categories of optimal size by maximizing code compression while minimizing predictive error.

ARTMAP is already being used in a variety of application settings, including industrial design and manufacturing, robot sensory motor control and navigation, machine vision, and medical imaging, as well as remote sensing [14, 15]. A more detailed description of ARTMAP neural networks can be found in [3]. For ARTMAP simulations we use ClasserScript v1.1 software [18] from <http://profusion.bu.edu/techlab/>.

C. Gaussian Maximum Likelihood Classification

The ML classifier is one of the most popular methods of classification in remote sensing, in which a pixel with the maximum a posteriori probability is classified into the corresponding class. In the case of multivariate Gaussian distribution a posteriori probability is defined as follows:

$$f_i(x|\mu_i, \Sigma_i) = (2\pi)^{-\frac{p}{2}} |\Sigma_i|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(x-\mu_i)^T (\Sigma_i)^{-1} (x-\mu_i)\right] \quad (1)$$

where μ_i and Σ_i are i th class mean vector and covariance matrix, respectively, L is the number of classes and input $x \in \mathbf{R}^p$. Assuming equally likely classes, the ML classification rule then is given by:

$$x \in \hat{i} \Leftrightarrow \hat{i} = \arg \max_{1 \leq i \leq L} d_i(x), \quad (2)$$

where $d_i(x)$ is a discriminate function in the form of:

$$\begin{aligned} d_i(x) &= \ln(f_i(x)) = \\ &= -\frac{1}{2} \left[(x-\mu_i)^T (\Sigma_i)^{-1} (x-\mu_i) + \ln |\Sigma_i| \right] + C. \end{aligned}$$

The ML method has an advantage from the view point of probability theory, but care must be taken with respect to the following items:

- Sufficient ground truth data should be sampled to allow estimation of the mean vector and the variance-covariance matrix of population.

- The inverse matrix of the variance-covariance matrix becomes unstable in the case where there exists very high correlation between two bands or the ground truth data are very homogeneous.

- When the distribution of the population does not follow the Gaussian distribution, the ML method cannot be applied.

IV. DATA DESCRIPTION

An image acquired by ETM+ sensor of Landsat-7 satellite was used for comparative analysis of above-described methods (Fig. 1, a). Parameters of image in World Reference System (WRS) [19] are path=186, row=25. Date of image acquisition is 10.06.2000. Dimensions: 4336x2524 pixels (30 m resolution) = 130x76 km.

ETM+ sensor provides data in 6 visible and infra-red spectral ranges with spatial resolution 30 m (bands 1-5 and 7); in thermal spectral range with spatial resolution 60 m (band 6), and in panchromatic range with spatial resolution 15 m (band 8). In this study we use as input to classification methods the six spectral bands 1-5 and 7.

In raw Landsat-7 images pixel values are digital numbers (DN) ranging from 1 to 255 (8 bits per pixel). Since these values are influenced by solar radiation [20], a procedure of converting DNs to *at-satellite reflectance* was applied according to [21]. In such a case pixel values lie in range [0; 1].

Since in this study we examine methods of *supervised classification* we need to provide so called *ground truth data* (sample pixels) in order to estimate weights and parameters of neural networks and statistical models. Unfortunately, we didn't have a possibility of gathering corresponding independent field data. In this case we use data provided by European Corine project that aims at land cover classification. In particular, we use CLC 2000 version of this project (Fig. 1, b).

Additionally, the following information was also used to distinguish land cover classes on Landsat-7 image.

- Estimated *Normalized Difference Vegetation Index* (NDVI):

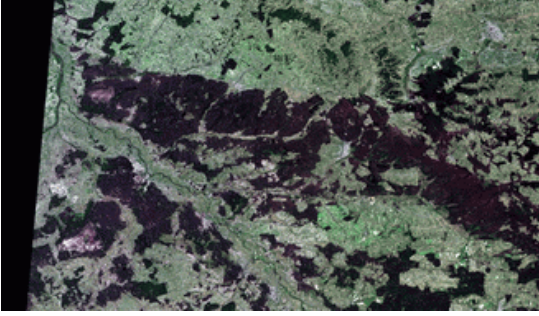
$$NDVI = (ETM4 - ETM3) / (ETM4 + ETM3)$$

where ETM3 and ETM4 are at-satellite reflectance values for spectral bands 3 and 4 respectively;

- Tasseled Cap transformation [20] that is based on principal component analysis (PCA) algorithm [22], and allows one to have decorrelated components. Moreover, in tasseled cap transformation first three major components has the following physical meaning: brightness, greenness, and wetness.

In this study eight target output classes were specified (Table 1).

(a) Landsat-7 image



(b) Corine data

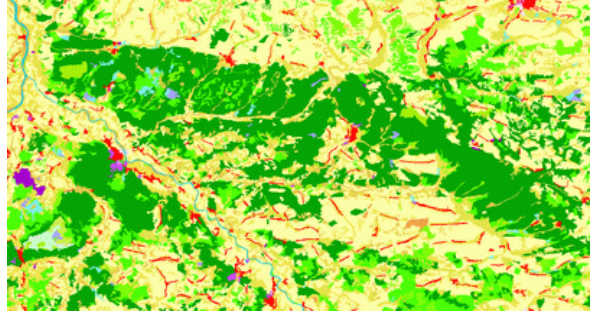


Fig. 1. (a) Image acquired by ETM+ sensor of Landsat-7 satellite (spatial resolution: 30 m). Area covers south-eastern part of Poland that borders with Ukraine. (b) Data for the same area provided by Corine project (spatial resolution: 100 m). The study area is dominated by forests, arable lands, and pastures. © EEA, Copenhagen, 2000.

Table 1. Class titles, Corine code levels, and number of sample pixels for each class.*

#	Class Title	Corine Code Level	Number of pixels
1	Broad-leaved forest	311	17890
2	Coniferous forest	312	20025
3	Mixed forest	313	10110
4	Non-irrigated arable land	211	25588
5	Pastures	231	9177
6	Inland waters	51x	7379
7	Artificial surfaces	1xx	12369
8	Open spaces with little or no vegetation	33x	2799
Total			105337

* x symbol is used to denote lower level classes that cannot be discriminated on Landsat-7 images. For example, it is hardly possible to distinguish water courses (e.g. rivers) from water bodies (e.g. lakes), or different types of artificial surfaces since their spectral characteristics do not differ. For this purpose, additional information should be provided.

V. RESULTS OF EXPERIMENTS

A. Performance Measures and Training and Testing Protocols

For comparative analysis of neural networks and statistical models for Landsat-7 images classification we use the same measure and the same training and testing sets. Performance of classification methods was evaluated in terms of classification rate. Both overall classification rate for all sample pixels and classification rate for each class separately were estimated.

Training and testing was done using *five-fold cross-validation* procedure [1, 23] as statistical design tool for methods assessment. According to this procedure available set of sample pixels is divided into five disjoint subsets; i.e. each subset consists of 20% of data. Models are trained on all subsets except for one, and classification rate is estimated by testing it on subset left out. All reported results reflect values averaged across 5

training/testing runs. So, this procedure produces robust performance measures while ensuring that no test sample pixels were ever used in training.

From table 1 it can be seen that number of sample pixels among target classes varies considerably. For example, there are 25588 sample pixels labeled “Non-irrigated arable land”, and 7379 sample pixels labeled “Inland waters”. In order for neural networks models to prevent imbalances of exemplars, we copied existing sample pixels for each class to be the same size. Such procedure allows one to “generate” training sets of the same size.

B. Input and Output Representation

Six channels from ETM+ sensor, namely 1-5 and 7, were selected to form feature vector for each pixel. Components of such vector represent at-satellite reflectance values lying in the range [0; 1].

Considering output coding for neural networks models, both MLP and ARTMAP have 8 output neurons corresponding to 8 target classes. During training target output is set to 1 for pixels belonging to such a class; otherwise, they are set to 0.

C. Classification with MLP

Five-fold cross-validation procedure was repeated at different MLP architectures: with 5, 15, 20, 25, 35, and 45 hidden neurons. Only one hidden layer was used in this study. For MLPs training EDBD algorithm was applied. Training was stopped after 500 epochs. *Save best* mode was applied during training process. Within this mode training and testing are sequentially applied to neural network. After each test the current classification rate is compared with previous results, and neural network is saved as the *best one* if current result is better than previous.

In all simulations initial values for learning rate and momentum factor in EDBD algorithm were set to 0.7 and 0.5 respectively.

Table 2 shows averaged classifications rates on testing sets for different MLP architectures.

Table 2. Averaged cross-validation results for MLP trained with EDBD algorithm.*

Class no.	MLP Architecture					
	6-5-8	6-15-8	6-20-8	6-25-8	6-35-8	6-45-8
1	97.63	98.78	98.99	99.02	99.15	98.97
2	80.95	83.57	83.99	84.20	84.64	85.67
3	67.09	68.70	68.12	68.38	68.00	67.37
4	85.44	87.72	88.24	89.03	89.84	89.56
5	86.16	90.42	91.55	90.41	91.01	91.43
6	97.14	97.71	97.66	97.75	97.63	97.64
7	69.09	83.45	84.09	83.99	83.46	83.56
8	95.57	96.82	96.28	96.53	96.79	96.52
Total	84.88	88.40	88.62	88.68	88.81	88.85

* the best estimates are indicated in boldface type.

The best value of classification rate was obtained for MLP with 45 hidden neurons.

D. Classification with ML

Mean vectors and covariance matrixes were estimated for each class using each of five training sets. For this purpose we use the following standard estimates

$$\hat{\mu}_i = \frac{1}{M_i} \sum_{j=1}^{M_i} x_i^j, \hat{\Sigma}_i = \frac{1}{M_i - 1} \sum_{j=1}^{M_i} (x_i^j - \hat{\mu}_i)(x_i^j - \hat{\mu}_i)^T$$

where x_i^j is j th sample of i th class, and M_i is number of sample pixels in i th class.

Averaged classifications rates on testing sets for Gaussian ML classifier are shown in Table 3.

Table 3. Averaged cross-validation results for ML classifier.

Class no.	
1	98.73
2	83.68
3	67.68
4	89.66
5	92.82
6	96.57
7	82.18
8	96.75
Total	88.02

E. Classification with ARTMAP

Five-fold cross-validation procedure was repeated for different vigilance parameters of ARTMAP network: 0.1, 0.2, 0.3, 0.5, and 0.95. The obtained results are shown in Table 4.

The best value of classification rate was obtained for ARTMAP with vigilance parameter set to 0.95.

Table 4. Averaged cross-validation results for ARTMAP neural network.*

Class no.	Vigilance parameters				
	0.1	0.2	0.3	0.5	0.95
1	98.92	99.68	99.56	98.52	99.88
2	79.58	80.86	80.34	79.16	80.88
3	69.14	68.16	68.66	69.36	68.14
4	81.50	81.50	81.72	81.88	83.50
5	76.48	74.26	75.34	74.10	78.94
6	96.70	96.60	96.76	97.40	93.76
7	79.38	77.28	78.32	77.12	76.78
8	96.42	97.36	97.00	97.54	98.24
Total	83.68	83.80	83.74	83.24	84.22

* the best estimates are indicated in boldface type.

F. Comparison of classification methods

The comparative analysis of best results obtained by neural networks models with ML classifier show no preferences of one method on others (Table 5).

The best overall classification rate of 88.85% (on all sample pixels) was achieved by using MLP.

Considering classification rates obtained for classes separately, different methods performed better on different classes. For class no. 2, 6, and 7 MLP outperformed ARTMAP and ML classifier. In turn, ARTMAP neural network was better for classes 1, 3, 8, and ML classifier was better for classes 4 and 5.

The worst performance of all classification methods was for class no. 3, ‘‘Mixed forest’’ (maximum 68.14%). This is due to the fact that mixed forests (class 3) consist of both broad-leaved (class 1) and coniferous forests (class 2), and its corresponding spectral properties mix up.

Table 5. Comparison of classification methods.*

Class no.	Method		
	MLP	ML	ARTMAP
1	98.97	98.73	99.88
2	85.67	83.68	80.88
3	67.37	67.68	68.14
4	89.56	89.66	83.50
5	91.43	92.82	78.94
6	97.64	96.57	93.76
7	83.56	82.18	76.78
8	96.52	96.75	98.24
Total	88.85	88.02	84.22

* the best estimates are indicated in boldface type.

VI. CONCLUSIONS AND FUTURE WORKS

In this paper we examined different neural networks models, namely MLP and ARTMAP networks, and statistical approach, namely maximum likelihood method,

for classification of remote sensing images. For comparative analysis of these methods data acquired by ETM+ sensor of Landsat-7 satellite and land cover data from European Corine project were used. The best overall classification rate for all classes (88.85%) was achieved by using MLP. While considering classification rates obtained for classes separately, different methods performed better on different classes. This, probably, is due to the complex topology of data that were used in this paper, and, thus, for different classes different classification methods are appropriate. The analysis of available data set represents a separate task, and is not covered in this article.

In order to improve performance of methods for remote sensing image classification future works should be directed to the use of *modular neural networks* and *committee machines*. It envisages the use of different models within a single architecture (e.g. neural networks with various parameters, or neural networks combined with statistical methods) allowing one to exploit advantages of different classification methods.

ACKNOWLEDGEMENT

This research was jointly supported by the Science and Technology Center in Ukraine (STCU) and the National Academy of Sciences of Ukraine (NASU) within project "GRID technologies for environmental monitoring using satellite data", no. 3872.

REFERENCES

- [1] S. Haykin (1999). "Neural Networks: a comprehensive foundation", Upper Saddle River, New Jersey: Prentice Hall, 842 p.
- [2] G.M. Foody, N.A. Campbell, N.M. Trodd, T.F. Wood (1992). "Derivation and applications of probabilistic measures of class membership from maximum likelihood classification", *Photogramm. Eng. Remote Sens.*, 58(9), pp. 1335-1341.
- [3] G.A. Carpenter, S. Grossberg, J.H. Reynolds (1991). "ARTMAP: Supervised Real-Time Learning and Classification of Nonstationary Data by a Self-Organizing Neural Network", *Neural Networks*, Vol. 4, pp. 565-588.
- [4] A.A. Minai, R.J. Williams (1990). "Back-propagation heuristics: A study of the extended delta-bar-delta algorithm", *IEEE International Joint Conference on Neural Networks*, 1990, vol. I, pp. 595-600.
- [5] P.J. Werbos (1994). "The roots of backpropagation: from ordered derivatives to neural networks and political forecasting", John Wiley & Sons, Inc., New York, 319 p.
- [6] G.A. Carpenter, S. Grossberg (1987). "ART 2: Stable selforganization of pattern recognition codes for analog input patterns", *Applied Optics*, vol. 26, pp. 4919-4930.
- [7] NASA Landsat 7, <http://landsat.gsfc.nasa.gov>.
- [8] European Topic Centre on Terrestrial Environment, <http://terrestrial.eionet.eu.int/CLC2000>.
- [9] J.A. Benediktsson, P.H. Swain, and O.K. Ersoy (1990). "Neural Network Approaches versus Statistical Methods in Classification of MultiSource Remote sensing Data", *IEEE Trans. On Geoscience and Remote Sensing*, Vol. 28, no. 4, pp. 540-552.
- [10] S.E. Decatur (1989). "Applications of Neural Networks to Terrain Classification," *Proceedings of the International Joint Conference on Neural Networks*, vol. 1, pp. 283-288.
- [11] H. Bischof, W. Schneider, and A.J. Pinz (1992). "Multispectral Classification of Landsat-Images Using Neural Networks", *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 30, no. 3, pp. 482-490.
- [12] M.S. Dawson, and A.K. Fung (1993). "Neural Networks and Their Applications to Parameter Retrieval and Classification," *IEEE Geoscience and Remote Sensing Society Newsletter*, pp. 6-14.
- [13] F. Roli, S.B. Serpico, and G. Vernazza (1996). "Neural Networks for Classification of Remotely-Sensed Images", In C.H. Chen, "Fuzzy Logic and Neural Networks Handbook", McGraw-Hills.
- [14] G.A. Carpenter, S. Martens, O.J. Ogas (2005). "Self-organizing information fusion and hierarchical knowledge discovery: a new framework using ARTMAP neural network", *Neural Networks*, 18, pp. 287-295.
- [15] G.A. Carpenter, S. Gopal, S. Macomber, S. Martens, C.E. Woodcock (1999). "A Neural Network Method for Mixture Estimation for Vegetation Mapping", *Remote Sens. Environ*, 70, pp. 138-152.
- [16] J.N. Hwang, S.R. Lay, and R. Kiang (1993). "Robust Construction Neural Networks for Classification of Remotely Sensed Data", *Proceedings of World Congress on Neural Networks*, vol. 4, pp. 580-584.
- [17] M. Kussul, A. Riznyk, E. Sadovaya, A. Sitchov, T.Q. Chen (2002). "A visual solution to modular neural network system development", *Proc. of the 2002 International Joint Conference on Neural Networks (IJCNN'02)*, Honolulu, HI, USA, vol. 1, pp. 749-754.
- [18] S. Martens (2005). "ClasserScript v1.1 User's Guide", *Technical Report CAS/CNS-TR-05-009*, 51 p.
- [19] The Worldwide Reference System (WRS), <http://landsat.gsfc.nasa.gov/documentation/wrs.html>.
- [20] C. Huang, B. Wylie, L. Yang, C. Homer, G. Zylstra (2002). "Derivation of a Tasseled Cap Transformation Based on Landsat 7 At-Satellite Reflectance", *International Journal of Remote Sensing*, v. 23, no. 8, pp. 1741-1748.

- [21] Landsat-7 Science Data User's Handbook, http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook_toc.html.
- [22] I.T. Jolliffe (1986). "Principal Component Analysis", New York: Springer-Verlag, 487 p.
- [23] M. Stone (1974). "Cross-validated choice and assessment of statistical predictions", Journal of the Royal Statistical Society, vol. B36, pp. 111-133.