

UDC 528.854

## Intellectual Approach to Crop Classification

O Kravchenko.

Space Research Institute of NASU-NSAU  
oleksiy.kravchenko@gmail.com

### Abstract

**Kravchenko O. Intellectual Approach to Crop Classification.** Information technology for crop classification and area estimation is presented. It is based on supervised classification of free of charge remote sensing data. To apply technology it is necessary to collect ground data with field boundaries for each crop class in question. Multilayer Perceptron Artificial Neural Network is used as soft classifier. Proposed technology addresses important issue of change of class probabilities in training and real data. Dedicated post processing of classifier outputs that is based on EM algorithm is used to estimate class probabilities in real image data. Such post processing leads to increase of classification accuracy up to 10%. A preliminary step to estimate errors in class probabilities for real data is also performed. Technology is assessed on moderate resolution MODIS NDVI data for 2009 vegetation season for the territory of Northern Ukraine.

### Introduction

Forecasting crop production is one of the main issues in both National and Global food security problems. Such forecasting naturally decomposes to crop acreage estimation and crop yield forecasting. Traditionally it was assumed that total crop acreage and crop proportions remains more or less stable and predictable while crop yield exhibits larger variations due to year to year changes in weather conditions and due to trend component. However now it is widely acknowledged that estimation of crop acreage is as important as yield forecasting [1]. One important reason for this is a global tendency of wide adoption of biodiesel growing and shrinkage of areas indented for food production. It has to be mentioned that these issues are considered by international community, in particular by developers of Global Earth Observation System of Systems ([http://earthobservations.org/geoss\\_ag.shtml](http://earthobservations.org/geoss_ag.shtml)). Crop acreage and yield estimates are the main topics of a number of international projects – GEO PAY, JECAM, CDIGAM and others.

Crop area estimation is of crucial importance for Ukraine for number of reasons. Ukraine is the sixth largest wheat exporter in the world and thus accurate crop area estimates is important to make economical decisions at Governmental level. Also the forms of ownership for the majority of cropped areas have been changed in 1990s. This results in decentralization of management decisions in agriculture and in particular makes crop proportions more unstable, affected by market fluctuation and other factors.

In this paper we present advanced information technology for crop acreage estimation via classification of satellite images. We used neural network approach to classification with dedicated post processing to adjust classifier to unknown a priori crop

probabilities in given image. Such combination of discriminative classification model and post processing is rarely used and to our knowledge it is the first time such technique is applied to crop classification in Ukraine.

The rest of the paper is organized as follows. Satellite and ground data used is described in Section Materials, our classification and post processing approaches are described in Section Methods, results of classification is presented in Section Results while paper conclusion and future works are presented in Section Conclusion.

### Materials

For the purpose of this study we concentrated on the Kyiv oblast (region) that is representative region of Northern Ukraine. We considered vegetation season of year 2009. For this period and the territory we collected remote sensing images and performed survey to gather ground truth crop data.

**Remote sensing data.** The main data for this study is MODIS NDVI timeseries available at LP DAAC USGS (<http://lpdaac.usgs.gov>). NDVI (Normalized Differential Vegetation Index) is simple but effective vegetation index that closely related to vegetation biomass. It is defined as the following:

$$NDVI = \frac{R_{nir} - R_{red}}{R_{nir} + R_{red}}, \quad (1)$$

where  $R_{nir}$  is surface reflectance in Near Infrared band of electromagnetic spectrum and  $R_{red}$  is surface reflectance in red band. Observing NDVI timeseries we can assess vegetation phenology and use this information for crop classification.

We used NDVI composites aggregated for 16 days period with resolution 250 m (product MOD13). This product is produced by USGS from daily cloud-free atmospherically corrected MODIS surface reflectances products. For the composition data with the least viewing angle is used. As a consequence resulting composite contains cloud-free NDVI values with highest possible resolution.

For the purpose of photointerpretation and crop label assignment in ground data acquisition process we also used Landsat5 TM image with 30 m ground resolution acquired 21 April 2009.

**Ground data.** We performed two along the road ground surveys in the Kyiv oblast at 11 and 16 July 2009 when all crops are still present. The GPS tracks of surveys are shown in Fig. 1a. Within survey the starting and end points of each field was marked by GPS. The final field shapes was created by photointerpretation of acquired Landsat5 image and base image data in Google Earth (<http://earth.google.com>). An example of

photointerpreted image is in Fig. 1b. Winter crops are clearly identified on Landsat image due suitable timing. Unfortunately due to cloud problems no Landsat image was available for late season, thus ground data may contain errors due to misinterpretation of summer crops borders.

During surveys more than 600 km were surveyed and more than 300 fields were inspected. The total number of crop species collected is 16. Each field was rasterized on MODIS grid in Sinusoidal projection. Totally over 2000 ground truth crop labels were obtained. NDVI timeseries for the most frequent crops (winter wheat, spring barley, spring wheat and summer crops) are shown in Fig. 2a. Surprisingly we do not observe much the winter rape, so we excluded it from subsequent consideration. The same is performed for rare spring crops like peat or oats. At last it is observed that summer crops (maize, sunflower, sugar beet, soybean) exhibit very similar NDVI timeseries thus for the purpose of classification they were combined in general class 'summer crops'. To split crop classification further an additional investigation is required, possibly with inclusion of other data, for instance from Synthetic Aperture Radars (SAR).

As we attempted to classify also non crop NDVI timeseries we have collected labels for four broad non-crop classes: water, grassland, forest and buildup. All labels were assigned by images photointerpreation by local experts.



Figure 1. a) The area under consideration (the Kyiv oblast) and location of surveyed fields; b) the result of photointerpretation: field boundaries outlined over base Google Earth image.

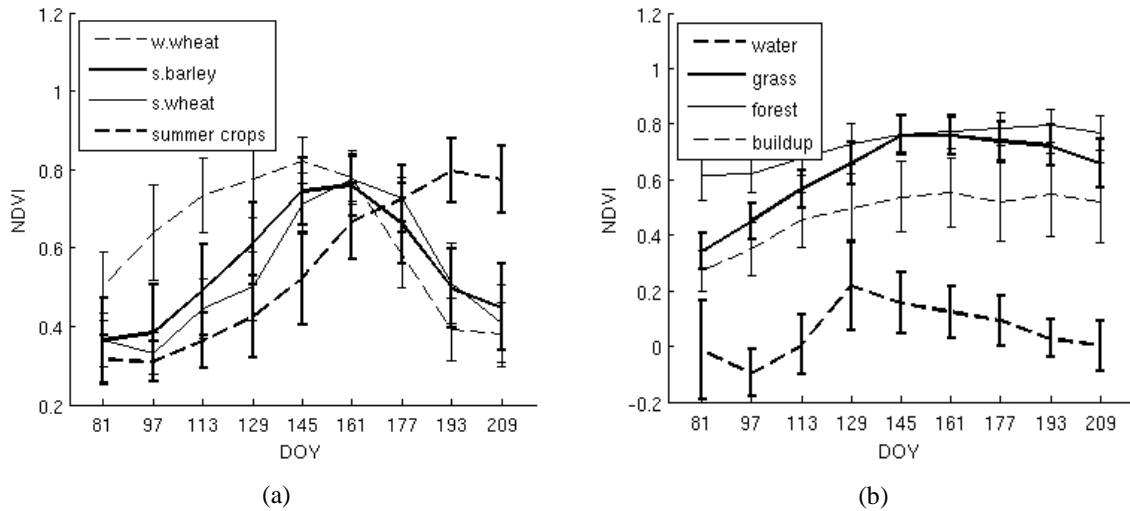


Figure 2. Mean values and error bars (one standard deviation) of NDVI timeseries for main crop and non-crop classes in Kyiv region: a) crop classes; b) non-crop classes.

## Methods

With availability of low-cost daily satellite images the use of time-series of remote-sensing data for land-cover and crop classification is common in remote sensing community. The most used methods are Bayesian that are based on generative or discriminative approach to classification [2]. Within generative approach class-conditional signal density is generally modeled by multivariate Gaussian distribution [3]. Among discriminative methods the most used are decisions trees, which are automatically constructed with algorithms like C4.5 [4] (for instance it is used by MODIS Land Cover group) or constructed in semi-automatic way with manually selected features. The developments in machine learning and pattern recognition begin adopted only recently and mostly for land-cover [5] or hyperspectral classification [6]. In this study we concentrated on established classification method – Artificial Neural Networks.

It is worth to mention that in remote-sensing data classification we have to deal with so called 'dataset shift' that has recently received much attention in machine learning [7]. Here we consider the simplest form of dataset shift. That is difference in a priori class probabilities in labeled and unlabeled data. We approach this problem through adjusting class probabilities by EM algorithm.

**Artificial Neural Networks.** In this work we used Multilayer Perceptron (MLP), a feed forward network with one of the most elaborated neural network architecture. This network is used to model a

posteriori class probabilities  $P(C_k | x)$ , where  $\{C_k, k = \overline{1, K}\}$  are crop classes,  $K$  is total number of classes,  $x \in \mathbf{R}^m$  is feature vector,  $m$  is total number of features. In this work  $x$  consists of NDVI values for given image pixel measured during the vegetation season.

Here we used MLP with one hidden layer and with softmax output. Thus the vector-valued function  $y_k(x, w)$  represented by neural network is given by:

$$f_k(x, w) = b_2 + \sum_i [u_i \cdot S(x^T v_i + b_1^{(i)})]$$

$$y_k(x, w) = \frac{\exp\{f_k(x, w)\}}{\sum_{k'} \exp\{f_{k'}(x, w)\}} \quad , \quad (2)$$

where  $v_i$  and  $u_i$  are weight vectors of neurons of hidden and output layers,  $b_1$  and  $b_2$  – weight vectors of bias neurons ( $b_1^{(i)}$  –  $i$ -th component of vector  $b_1$ ). Together vectors  $b_1, b_2, v_i$  and  $u_i$  constitute parameters vector of MLP  $w$ . Here  $S$  is a hyperbolic tangent that belongs to the family of sigmoid activation functions. The use of softmax output function ensures that outputs distributed in  $[0;1]$  interval and sums to unity. This important property is required to model a posteriori class probabilities.

The learning procedure is applied to estimate parameters vector  $w$  of neural network. Let  $\{(x_n, t_n), n = \overline{1, N}\}$  is training dataset (that is subset of

available labeled data), where target variables (class labels)  $t_n \in \{0,1\}^K$  has 1-of- $K$  coding scheme. That is each binary variable  $t_n^{(k)}$  indicates membership of sample element  $n$  to class  $C_k$ . Thus  $\sum_k t_n^{(k)} = 1$ . Class conditional distribution  $P(C_k | x)$  is modeled by MLP as follows:

$$P(C_k | x) = P(t^{(k)} = 1 | x) = y_k(x, w).$$

Conditional distribution of targets is usually modeled by generalized Bernoulli distribution:

$$P(t|x) = \prod_k \left\{ P(t^{(k)} = 1 | x) \right\} = \prod_k P(C_k | x)^{t^{(k)}} = \prod_k y_k(x, w)^{t^{(k)}}. \quad (4)$$

Weight vector  $w$  is estimated by standard Maximum Likelihood procedure.  $w$  is obtained by minimizing the following error function (negative log-likelihood of data):

$$E(w) = -\sum_n \log[P(t_n | x_n)] = -\sum_n \sum_k t_n^{(k)} \log[y_k(x_n, w)].$$

It has to be mentioned that direct minimization of (5) leads to overfitting producing classifier that achieves good performance on training data set and exhibits substantially worse performance on test, previously unseen data. To cope with this problem we add noise to feature vectors  $x_n$  in training dataset that is is equivalent to regularization of error function (5) [2].

**Adjusting a priori class probabilities.** Most current machine-learning methods assume that both labeled (used for training and testing of classifier) and unlabeled data is generated from the same probability distribution  $P(x, C)$ . In real situations this assumption does not holds. For instance in our survey, class distribution in collected ground sample will differ from class distribution in the whole population. This sampling bias is introduced due to nonuniformity of ground survey and due to the fact that farmers tend to grow more valuable crops (like vegetables) away from the big roads.

In more general setting we can assume that labeled data is generated from  $P_{lab}(x, C)$  and the whole population is generated from  $P_{unl}(x, C)$ . We assume that  $P_{lab}$  and  $P_{unl}$  have the same conditional signal distribution  $P(x|C)$  for each  $C \in \{C_k, k = \overline{1, K}\}$  and possibly different class distributions:

$$P_{lab}(x, C) = P(x|C)P_{lab}(C), P_{unl}(x, C) = P(x|C)P_{unl}(C). \quad (6)$$

If we know true class probabilities in unlabelled data we can reweight outputs of classifier  $y_k^{lab}(x, w)$

(trained on labelled data) in such a way that it will produce correct estimates of a posteriori class probabilities for each data instance [2]:

$$y_k^{unl}(x, w) = 1/Z \cdot y_k^{lab}(x, w) \frac{P_{unl}(C_k)}{P_{lab}(C_k)}, \quad (7)$$

$$Z = \sum_k y_k^{lab}(x, w) \frac{P_{unl}(C_k)}{P_{lab}(C_k)}. \quad (8)$$

$P_{lab}(C_k)$  can be easily estimated as class proportions in training data set. Unfortunately we do not have estimations of  $P_{unl}(C_k)$ , indeed in crop area estimation problem  $P_{unl}(C_k)$  are our target variables.

Intuitively if our classifier is consistent estimates of a posteriori class probabilities  $P^*(C_k) = 1/N \sum_n y_k(x_n, w)$  have to be similar to a priori class probabilities  $P(C_k)$ . The idea is to choose a priori class probabilities  $P^*(C_k)$  for unlabeled data that will produce the same a posteriori class probabilities:

$$P^*(C_k) = \frac{1}{N} \sum_n y_k^{lab}(x_n, w) = \frac{1}{N} \sum_n \left\{ \frac{1}{Z_n} y_k^{lab}(x_n, w) \frac{P^*(C_k)}{P_{lab}(C_k)} \right\}, \quad (9)$$

where  $Z_n$  are normalization variables (8) for each data vector  $x_n$  in unlabeled data set. In fact equation (9) suggests the following iterative algorithm to find required  $P^*(C_k)$ :

1. Choose initial values  $P^0(C_k)$ , for instance equal;

2. For each  $l \in \mathbf{N}$ , calculate  $P^l(C_k)$  by:

$$P^l(C_k) = \frac{1}{N} \sum_n \left\{ \frac{1}{Z_n} y_k^{lab}(x_n, w) \frac{P^{l-1}(C_k)}{P_{lab}(C_k)} \right\}; \quad (10)$$

3. Stop if a convergence criterion is met.

Such algorithm can be viewed as particular instance of famous Expectation-Maximization (EM) algorithm [8] for estimation of density  $P_{unl}(x, C)$ . In our case it is assumed that estimation of parameters of signal conditional distribution  $P_{unl}(x|C)$  has already performed and thus maximization step is omitted. We have to mention that within this approach training of classifier is performed before adjusting class probabilities and therefore the whole procedure is not time-consuming. It becomes possible because MLP classifier provides probabilistic output. In the case of discriminative models (such as SVMs) classifier produces only class labels and thus we have to retrain classifier for each iteration (with reweighed data set).

As a consequence such procedure can be very time consuming.

To our knowledge in context of remote sensing such approach was applied in [9] at first time and than in [10] for land-cover classification. Here we apply this method for crop classification.

### Results

As described in previous section we have performed two ground surveys and use these data for supervised learning of MLP classifier. Each polygon with field boundaries has been rasterized onto MODIS Land grid. This grid has Sinusoidal projection with 250 m grid spacing. Sinusoidal projection is equal-area projection that simplifies area calculation as each grid cell in such projection has the same area. As a result from the both surveys we have collected 1763 crop labels and 11862 non-crop labels. 2/3 of data is used as training dataset and 1/3 for the testing of classifier. As described in Materials section we attempt to classify NDVI timeseries into 8 classes (4 crop and 4 non-crop classes).

To create more robust classifier training and test datasets were resampled to equal class probabilities. This is required step as in the most cases learning algorithms produce worse classifiers using unbalanced data. After resampling each class in training dataset has 4000 labels (32000 in total) and each class in training

dataset has 500 labels (4000 in total). To prevent overfitting of classifier each input  $x_n$  from training dataset were corrupted by additive Gaussian noise with sigma 0.05. The value of sigma is chosen as a priory known noise level which can be introduced by variations in NDVI due to variations in satellite viewing geometry, crop varieties, soil properties and other factors.

Training dataset was used to train MLP classifier. We gradually increased number of hidden neurons, beginning with 10 neurons and stopped when classification error is stabilized. Resulting classifier has 30 hidden neurons. For minimization of error function (5) we used quasi-Newton algorithm implemented in Netlab software [11].

Confusion matrices for training (without noise) and test datasets are shown in Fig. 3. The correspondence of class labels and indexes is shown in Table 1. We can see that in average we achieve good classification accuracy 93% on training data and 89% on test data. However 2 classes, namely spring barley (#2) and spring wheat (#3) can not be discriminate with high accuracy. The precisions for spring barley and spring wheat are 80.2% and 77.9%, while recalls are 65.8% and 84.8%. Such result is expected, because both crops grow in the same period and have similar NDVI timeseries

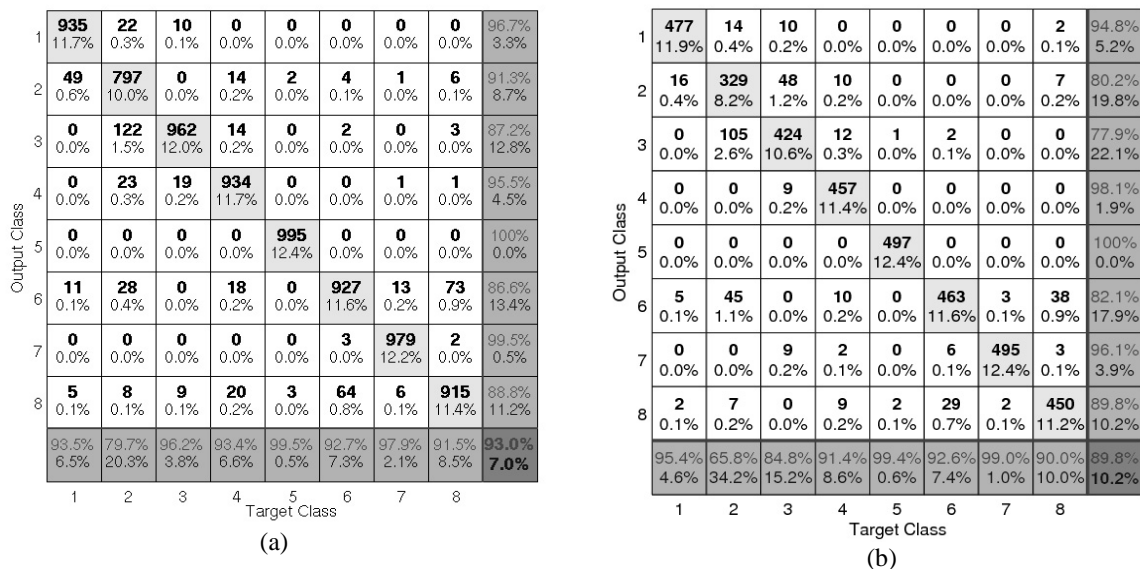


Figure 3. Confusion matrix on training set (a) and test set (b).

In our case if we have classes that are not discriminated perfectly it become important to take into account a priory class probabilities that were assumed

during learning of classifier and prescribed in training dataset. For instance if we increase proportion of spring barley (#2) in training dataset and learn MLP the

accuracy of classifier on spring barley will increase, but accuracy on other classes, fist of all for spring wheat (#3), will decrease. So it becomes important to estimate class probabilities in real data and adjust classifier to this new probabilities by equations (7), (8).

However class probabilities estimation procedure, described in previous section, can introduce errors in estimates due to finite data sets, unavoidable errors in approximation of class conditional distribution etc. Here we attempt to estimate upper bound of such errors considering the following experiment. We created artificially shifted dataset by resampling original test dataset with new specified class probabilities. We artificially increased a priory class probabilities of classes with the biggest classification error (they are spring barley (class #2) and spring wheat (class #3). The probability of spring barley is 64.5%, spring wheat is 22.9%, while other classes have the same probability 2%. Thus we considered the extreme situation when class proportions in real data is severely shifted from class proportions in training data (in our case they are equal).

We run iterative procedure (10) for 10 iterations until RMS difference of class probability vector between iterations became less then 0.1%. The subsequent iterations do not change the class probability vector significantly (RMS difference remains less then 0.1%). Let  $P_{sh}(C_k)$  and  $P_{sh}^*(C_k)$  will be true and estimated class probabilities in shifted dataset.  $P_{sh}(C_k)$ ,  $P_{sh}^*(C_k)$  and relative errors  $|P_{sh}^*(C_k) - P_{sh}(C_k)| / P_{sh}(C_k)$  are shown in Table 1. It can be seen that accuracy for dominant classes is rather good, relative errors are less than 10% (which is the requirement of State Statistics Committee of Ukraine), while for small classes relative errors can be rather large (up to 90% for grasslands).

A confusion matrix of original MLP classifier for artificially shifted dataset is presented in Fig. 4a. The total accuracy is 74% that is 15% less than in case with non-shifted data (Fig. 3b). Such drop of total accuracy is largely determined by dominant class (spring barley, #2), which in this case has the worst accuracy 65%. Such example emphasizes the importance of determining correct estimations of a priory class probabilities instead of setting them to some predetermined values (like equal ones, that is rather common). In Fig. 4b we present a confusion matrix for the same data, but after adjustment of classifier outputs by equations (7),(8) using estimated class probabilities  $P_{sh}^*(C_k)$ . The total accuracy is 85% which is much higher than before adjustment. It has to mentioned, that in adjustment procedure we do not use labels from test data. Labels are used solely for final

accuracy calculation. Thus we can perform the same adjustment procedure for unlabeled data.

Finally we present results of classification of the whole territory of the Kyiv oblast. We chose rectangular area in MODIS Land Sinusoidal projection which fully covers the Kyiv oblast. The size of result image is 1000x1000 pixels. We applied trained MLP classifier to the data, than perform class probability estimation procedure (10) and finally adjust output of classifier using these estimates (equations (7), (8)). Obtained thematic map and legend with estimates of class probabilities are shown in Fig. 5 (color version of figure can be obtained from author). It is interesting to note that estimated spring wheat class probability is rather low (1.7%) that corresponds to the growing practice in Northern Ukraine, while the proportion of this crop was unusually high in collected ground data sample.

Unfortunately at the moment of publication the official statistical information about crop acreage in the Kyiv oblast in 2009 is not available so it is not possible to compare our result with official one.

## Conclusion

We considered crop area estimation problem and propose to solve it with the help of remote-sensing data. Our approach consists in using free of charge satellite images with moderate resolution and ground data to perform image classification. We used supervised approach to classification and apply Multilayer Perceptron with regularization as a soft classifier. One important question is considered. It is the difference in class probabilities in training dataset and in real data, which can be seen as simplest case of dataset shift. The solution proposed is post processing of classifier outputs based on particular instance of EM algorithm.

Table 1. Class probabilities for artificially shifted test dataset

Class labels	True probabilities	Estimated probabilities	Relative error, %
1. Winter wheat	0.0208	0.0208	0.00
2. Spring barley	0.6458	0.6088	5.73
3. Spring wheat	0.2292	0.2459	7.29
4. Summer crops	0.0208	0.0171	17.79
5. Water bodies	0.0208	0.0204	1.92
6. Grassland	0.0208	0.0395	89.90
7. Forest	0.0208	0.0247	18.75
8. Buildup	0.0208	0.0228	9.61

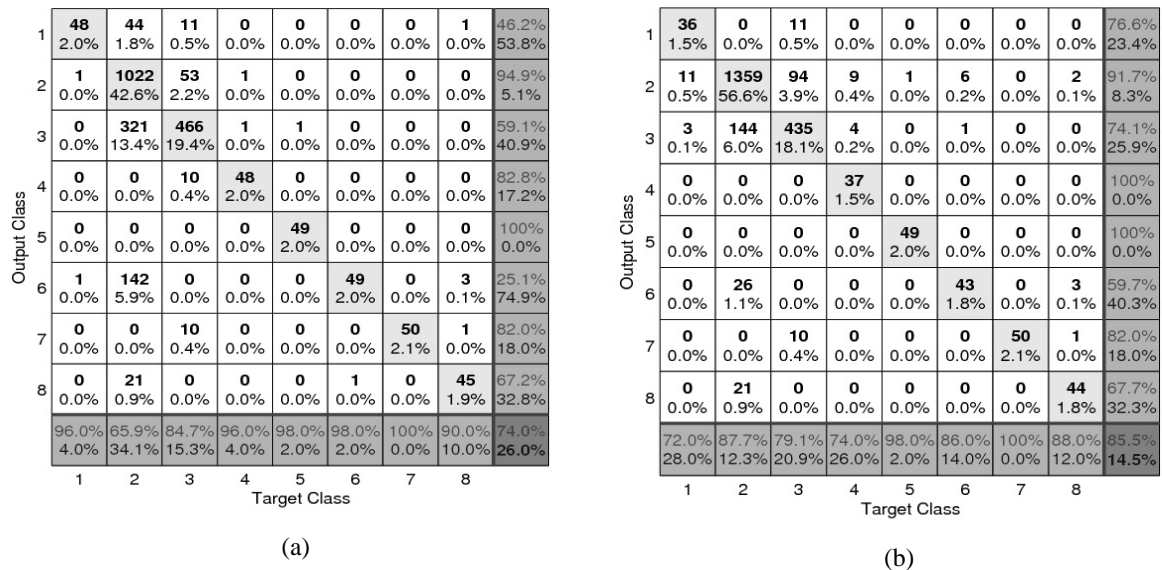


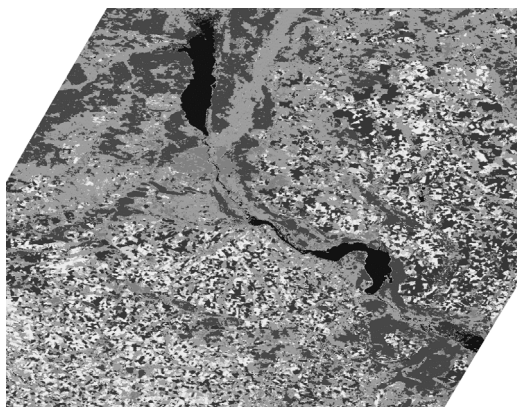
Figure 4. Confusion matrices for artificially shifted dataset: original MLP (a), MLP with adjusted output by EM algorithm.

As reasonably mentioned in [12] such post processing procedure is not perfect and assumes precise knowledge of conditional class probabilities while it has to be estimated from finite and possible biased sample. Our attempts of estimation of resulting errors can be seed as preliminary step in this direction and further work is required. However the increase of performance of adjusted classifier on test data up to 10% is showed the potential of this way.

We archived good classification results with accuracy 89% and obtained good accuracy for crop acreage estimates (<10% for dominant classes) which meets accuracy standards of State Statistics Committee of Ukraine.

Further works include profound treatment of errors in classifier adjustment procedure, explicit modelling of mixed pixels that are unavoidable at moderate resolution and adaptation of acreage estimation approach to other regions.

This work is partially supported by joint NASU-STCU grant #4928 "Grid technologies for multi-source data integration".



(a)

Classes	Estimated class prob., %
1. Winter wheat	8.74
2. Spring barley	7.27
3. Spring wheat	1.74
4. Summer crops	16.37
5. Water bodies	2.10
6. Grassland	38.21
7. Forest	16.54
8. Buildup	9.03

(b)

Figure 5 Thematic map, reprojected in geographic projection (a), legend with estimated class probabilities (b)

### Reference

1. GEOSS Best practices for crop area estimation with Remote Sensing, 2008 // Workshop GEOSS - Crop Area Estimates with Remote Sensing. — Ispra, Italy. — 5-6 June 2008.
2. Bishop C. Pattern Recognition and Machine Learning // Springer, 2006. — 740 p.
3. Wardlow B, Egbert S. Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U.S. Central Great Plains // Remote Sensing of Environment. — **112**. — N 3. — 2008. — P. 1096-1116.
4. Quinlan J. R. C4.5: Programs for Machine Learning // Morgan Kaufmann Publishers, 1993. — 302 p.
5. Foody G. M. SVM-based multi-class classification of remotely sensed data // Int. J. of Remote Sensing. — **29**. — N 6. — 2008. — P. 1817-1823.
6. Bazi Y., Melgani F. Toward an optimal SVM classification system for hyperspectral remote sensing images // IEEE Trans. on Geoscience and Remote Sensing. — **44**. — N 1. — 2006. — P. 3374-3385.
7. Quiñero-Candela J. et al. Dataset Shift in Machine Learning // The MIT Press, 2009. — 248 p.
8. Шлезингер М.И. Десять лекций по статистическому и структурному распознаванию / М.И. Шлезингер, В. Главач // К.: Наукова думка, 2004. — 535 с.
9. Gorte and Stein, 1998 Bayesian classification and class area estimation of satellite images using stratification // IEEE Trans. on Geoscience and Remote Sensing. — **36**. — 1998. — P. 803-812.
10. Saerens M., Latinne P., Decaestecker C. Adjusting the Outputs of a Classifier to New a Priori Probabilities: A Simple Procedure // Neural Computation. — 14. — N 1. — 2002. — P. 21-41.
11. Nabney I.T. Netlab: Algorithms for Pattern Recognition // London: Springer Verlag, 2002. — 410 p.
12. F.J. Gallego. Remote sensing and land cover area estimation // Int. J. Remote Sensing. — **25**. — N. 15. — P. 3019-304

Received 30.03.2010