www.czasopisma.pan.pl PAN www.journals.pan.pl

GEODESY AND CARTOGRAPHY
Vol. 65, No 2, 2016, pp. 193-218

DOI: 10.1515/geocart-2016-0016

Improving sub-pixel imperviousness change prediction by ensembling heterogeneous non-linear regression models

Wojciech Drzewiecki

AGH University
Faculty of Mining Surveying and Environmental Engineering
Department of Geoinformation, Photogrammetry and Remote Sensing of Environment
al. Mickiewicza 30, 30-059 Kraków, Poland
e-mail: drzewiec@agh.edu.pl

Received: 2 July 2016 / Accepted: 12 September 2016

Abstract: In this work nine non-linear regression models were compared for sub-pixel impervious surface area mapping from Landsat images. The comparison was done in three study areas both for accuracy of imperviousness coverage evaluation in individual points in time and accuracy of imperviousness change assessment. The performance of individual machine learning algorithms (Cubist, Random Forest, stochastic gradient boosting of regression trees, k-nearest neighbors regression, random k-nearest neighbors regression, Multivariate Adaptive Regression Splines, averaged neural networks, and support vector machines with polynomial and radial kernels) was also compared with the performance of heterogeneous model ensembles constructed from the best models trained using particular techniques.

The results proved that in case of sub-pixel evaluation the most accurate prediction of change may not necessarily be based on the most accurate individual assessments. When single methods are considered, based on obtained results Cubist algorithm may be advised for Landsat based mapping of imperviousness for single dates. However, Random Forest may be endorsed when the most reliable evaluation of imperviousness change is the primary goal. It gave lower accuracies for individual assessments, but better prediction of change due to more correlated errors of individual predictions.

Heterogeneous model ensembles performed for individual time points assessments at least as well as the best individual models. In case of imperviousness change assessment the ensembles always outperformed single model approaches. It means that it is possible to improve the accuracy of sub-pixel imperviousness change assessment using ensembles of heterogeneous non-linear regression models.

Keywords: machine learning, model ensembles, sub-pixel classification, impervious areas, Landsat

1. Introduction

The coverage of impervious surface areas (ISA), defined as areas preventing the infiltration of rainwater into the ground, is considered as very important environmental indicator (Arnold and Gibbons, 1996). Growing ISA percentage indicates an increase of anthropogenic impacts on the environment, what is reflected in changes of many environmental functions and processes (e.g. rainfall-runoff transformation, ground water recharge, environmental or even population health) (Arnold and Gibbons, 1996; Shahtahmassebi et al., 2014). For this reason the accurate information about ISA coverage and monitoring of its change plays an important role in many environmental studies, especially urban and hydrological ones (Caldwell et al., 2012; Dams et al., 2013; Shahtahmassebi et al., 2014; Li et al., 2016).

Remote sensing is nowadays commonly used for monitoring changes in land use and land cover (LULC), including changes of imperviousness. Applications of remotely sensed data for mapping ISA have been reviewed by Weng (2012) and Lu et al. (2014a). Recently, very high spatial resolution satellite images, aerial (Nielsen et al., 2011) and even UAV photographs (Tokarczyk et al., 2015) are used for imperviousness mapping. Despite that fact, in regional scale and/or ISA change assessment applications Landsat imagery is the most frequently used (Lu et al., 2014b; Tokarczyk et al., 2015; Li et al., 2016).

LULC changes may take different forms (Turner and Meyer, 1994). In case of conversion one land cover type changes completely into another. However, very often the changes are much more subtle and have a form of modification. What changes is not the LULC category, but the proportions of land cover fractions inside the considered area (pixel, object, mapping unit) still classified to the same class. For example, due to urbanization processes the housing density (and in turn the imperviousness) may increase substantially, nevertheless the area is classified as “discontinuous built-up” or “10-30% impervious” as it was before.

The ISA changes usually takes a form of modifications. Their detection and assessment from high, medium and low-resolution remotely sensed images requires the use of sub-pixel analysis approaches. In case of ISA mapping done for individual cities, the spectral mixture analysis-based methods are preferred (Ridd, 1995; Lu et al., 2014a). However, when the study area is dominated by other than urban types of land cover, mixture analysis may not give the expected accuracy and the regression-based approaches provide an alternative (Lu et al., 2014b; Heremans and Van Orshoven, 2015). These methods range from building regression models with vegetation indices (Bauer et al., 2004), through regression trees applications (Yang et al., 2003) to implementation of other machine learning algorithms, like artificial neural networks (Mohapatra and Wu, 2007; Chormanski et al., 2008) or support vector machines (Walton, 2008; Esch et al., 2008).

Applicability of machine learning algorithms for sub-pixel imperviousness mapping was compared in several studies (Liu and Wu, 2005; Walton, 2008; Mohapatra and Wu, 2010; Bernat and Drzewiecki, 2014; Heremans and. Van Orshoven, 2015). Each

of these studies was focused on evaluating the ISA prediction accuracy achieved for individual point in time. But, imperviousness estimates generated for particular dates are very often subtracted from each other and used for change assessment. The methodology of change detection through sub-pixel percent imperviousness mapping was proposed by Yang et al. (2003). This approach is classified to layer arithmetic change detection techniques (Tewkesbury, 2015). It should be noted however, that the input layers and their difference have semantic meaning. Moreover, the subtracted layers may be created through sophisticated non-linear regression models or even through hybrid approach when classification and regression models are combined (Mountrakis et al., 2009; Bernat and Drzewiecki, 2014).

In case of sub-pixel imperviousness change detection it is commonly assumed that to assure the highest accuracy of change map the accuracies of individual imperviousness maps should be as high as possible. Such assumption is true for post-classification change detection techniques (Hussain, 2013). However, in case of sub-pixel assessment of fractional coverages (eg. imperviousness), although generally true, may fail. As imperviousness change is calculated by subtracting of two values, the standard deviation of ISA change assessment error ($s_{ISA_{ch}}$) may be calculated as (Kircher, 2001; Morgan and Henrion, 1990):

$$s_{ISA_{ch}} = \sqrt{\left(s_{ISA_2}\right)^2 + \left(s_{ISA_1}\right)^2 - 2r_{ISA_1ISA_2}s_{ISA_1}s_{ISA_2}} \tag{1}$$

where s_{ISA_1} and s_{ISA_2} are standard deviations of ISA errors for individual time points and $r_{ISA_1ISA_2}$ is the correlation coefficient between the errors of individual ISA assessments. As the result, due to higher correlation of their prediction errors the approaches less accurate for individual assessments may give lower error of change evaluation.

Drzewiecki (2016) presented the comparison of nine selected machine learning algorithms for mapping the imperviousness and its change in the catchment of Dobczyce Reservoir (south Poland) from Landsat imagery. The Cubist algorithm gave the lowest errors for both individual assessments. Nevertheless, the most accurate change assessment was obtained by subtraction of Random Forest results due to higher correlation of their prediction errors. These findings stimulated the research which are presented in this paper. The first purpose of the present study was to repeat the previous experiment and compare selected machine learning algorithms in the context of sub-pixel imperviousness change detection with new datasets. The second aim of the research was to answer the question if by ensembling prediction algorithms one can find the approach performing better both for individual time points ISA assessments and ISA change evaluation.

In ensemble learning process a set of models is generated and combined to obtain the final prediction (Mendes-Moreira et al., 2012). The approach is used both for classification and regression problems, but some differences in methods exist. A comprehensive survey of ensemble learning approaches for regression is provided

by Mendes-Moreira et al. (2012). Model ensembles are reported to be more robust and accurate than single models (García-Pedrajas et al., 2005). As shown by Perrone and Cooper (1993), it is possible to construct an ensembled model which performs in the mean squared error (MSE) sense as good or better than any of single models in the ensembled set.

Ensemble methods may be divided according to the approach used for generation of ensembled models (Rooney et al., 2004). In homogeneous ones all the models are obtained using the same induction algorithm. If different algorithms are used to create a set of models, the ensemble is called heterogeneous. Wichard et al. (2003) suggest that in regression problems heterogeneous ensembles perform better than homogeneous ones. This is due to less correlated errors of individual models, what results in reduction of ensemble variance.

These advantages of heterogeneous ensemble learning should also be beneficial for ISA change assessment. Predictions of ensembled models for individual time points should have MSE not higher than obtained with the best of ensembled models. At the same time, if their variances are reduced, one may also expect more correlated errors for these predictions. As a result, the error of change evaluation should be reduced as well.

Heterogeneous ensemble approaches were successfully used for regression problems in different applications (Partalas et al., 2008; Kew and Mitchel, 2015; Heinermann and Kramer, 2015). Nevertheless, remote sensing applications are hard to find. On the other hand heterogeneous models were reported as outperforming approaches based on individual machine learning techniques in remote sensing classification tasks (Qi and Huang, 2007; Engler et al., 2013; Gómez-Chova et al., 2013) and Zhang (2010) expects that ensemble learning would be adopted to high-level fusion of multi-source remote sensing data.

2. Study areas and image data

2.1. Study areas

Three watersheds located in South Poland were chosen as study areas (Figure 1). All of them are the catchments of reservoirs. The first one is the part of Sola river watershed till the point of water intake for Dzieckowice reservoir. The second one – the watershed of Dobczyce reservoir – is located on Raba river. These two reservoirs are important sources of drinking water for upper Silesia and Krakow agglomerations. The last one is the watershed of Czorsztyn reservoir located on Dunajec river. Names of the rivers are used as the labels of study areas further in the paper.

All three catchments are hilly and partially even mountain regions with large forest cover (40–50% of area). Urbanized areas covers from ca. 5 (Dunajec) to ca. 8 percent (Raba) of the area. More information about land use and land cover in studied watersheds may be found in Wężyk et al. (2016).

Fig. 1. Location of study areas (S – Sola watershed, R – Raba watershed, D – Dunajec watershed)

2.2. Image data

The time points for ISA assessments were chosen due to availability of high-resolution image data needed for acquiring the reference information about imperviousness. Photogrammetric aerial projects were realized in all study areas in 1996 and 2009. To gather reference data digital color aerial orthophotomaps were used.

Landsat-5 TM images were used for ISA mapping. All cloud-free images available for mid 1990s (1996 ± 1 year) and late 2000s (2009 ± 1 year) were orthorectified using orbital model. As the imperviousness change was evaluated by subtraction of individual ISA maps, no atmospheric correction was applied. An overview of the remote sensing imagery used in this study is given in Table 1.

Table 1. Overview of image data

Time Period	Data Characteristics			
	Data type	Platform	Acquisition dates	Spatial resolution
Raba dataset				
mid 1990s	source	Landsat-5	02.07.1994, 22.10.1994, 24.08.1996, 12.09.1997	30 m
	reference	aerial	1996	0.75 m
late 2000s	source	Landsat-5	28.08.2009, 12.06.2010	30 m
	reference	aerial	2009	0.25 m

Time Period	Data Characteristics			
	Data type	Platform	Acquisition dates	Spatial resolution
Dunajec dataset				
mid 1990s	source	Landsat-5	25.06.1994, 02.07.1994	30 m
	reference	aerial	1996	0.75 m
late 2000s	source	Landsat-5	03.05.2007, 18.06.2009, 21.08.2009, 12.06.2010, 11.10.2010	30 m
	reference	aerial	2009	0.25 m
Sola dataset				
mid 1990s	source	Landsat-5	02.07.1994, 22.10.1994	30 m
	reference	aerial	1996	0.75 m
late 2000s	source	Landsat-5	19.05.2007, 28.08.2009, 12.06.2010	30 m
	reference	aerial	2009	0.25 m

3. Methods

Nine machine learning (ML) regression algorithms were tested in this study. For every study area, each of them was used to predict imperviousness for both mid 1990s and late 2000s. Then, the performance of every ML algorithm for imperviousness change assessment was evaluated based on post-classification comparison. The methodological framework (Figure 2) used in the study for the evaluation of individual machine learning algorithms performance follows (Drzewiecki, 2016). Succeeding subsections present details on data preparation, machine learning algorithms used in the study, tunning and selecting the models for ISA prediction and accuracy assessment. In the last subsection the methodological issues of model ensembling and ensembles accuracy evaluation are described.

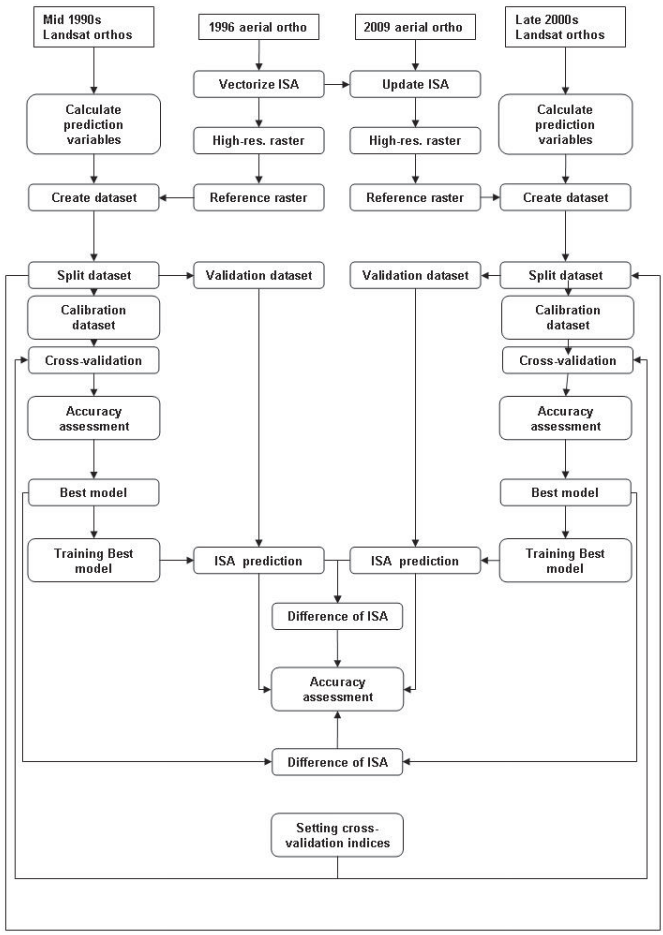

Fig. 2. Framework of ISA and ISA change prediction accuracy assessment

3.1. Preparation of calibration and validation datasets

The reference data on imperviousness were gathered by photo interpretation. In every study catchment test sites (300 m x 300 m or 200 m x 200 m) were spread out all over the area. First, in these reference areas representing various land use categories (urban, suburban, agricultural, industrial, commercial, transport, forest, water) ISA were vectorized for 1996. The reference imperviousness layer was then overlaid on 2009 orthophotomap and updated. Such approach eliminated differences caused by independent drawing of boundaries and interpretation errors (in case of any doubts, the 1996 photos were checked and eventually re-interpreted). Then, from obtained ISA polygons high-resolution (0.1 m) raster was created and used to calculate ISA percentage for 30 m resolution pixels of orthorectified Landsat images which laid

completely inside test sites areas. As the pixel grid for mid 1990s and late 2000s was the same, these data provided also reference for ISA change assessment.

For every Landsat pixel having reference ISA information, the set of input data for imperviousness prediction was created. It consisted of spectral (and thermal) band values (7 variables), their ratios (21 variables) and values of chosen indices (5 variables) derived for every of Landsat images available in given study catchment for individual time period. Following spectral indices were selected from the ones reported in other studies as potentially useful for evaluation of imperviousness:

- Normalized Difference Vegetation Index (NDVI)

$$NDVI = \frac{TM_4 - TM_3}{TM_4 + TM_3} \tag{2}$$

- Modified Normalized Difference Water Index (MNDWI) (Xu, 2006)

$$MNDWI = \frac{TM_2 - TM_5}{TM_2 + TM_5} \tag{3}$$

- Normalized Difference Built-Up Index (NDBI) (Deng and Wu, 2012)

$$NDBI = \frac{TM_5 - TM_4}{TM_5 + TM_4} \tag{4}$$

- Normalized Difference Impervious Index (NDISI) (Deng and Wu, 2012)

$$NDISI = \frac{TM_6 - \frac{(MNDWI + TM_4 + TM_5)}{3}}{TM_6 + \frac{(MNDWI + TM_4 + TM_5)}{3}} \tag{5}$$

where TM1-TM7 are the bands of Landsat TM image.

The fifth index used to derive the input variable was the ZABUD1 index proposed in (Lewiński, 2006) for improved recognition of urbanized areas on Landsat-7 images. In this case the panchromatic channel value in original ZABUD1 formula was substituted by average of blue, green and red bands values and the index calculated as:

$$ZABUD1 = \sqrt{(TM_2 - TM_3)^2 + (TM_3 - TM_4)^2 + (TM_4 - TM_5)^2 + (TM_5 - TM_7)^2 + (TM_7 - \frac{TM_1 + TM_2 + TM_3}{3})^2} \tag{6}$$

To split created datasets into calibration (80 percent of data) and validation (20 percent of data) datasets the random approach stratified according to ISA percentage in mid 1990s was used. Characteristics of final datasets is presented in Table 2.

Table 2. Datasets characteristics

Dataset	Time period	No of predictors	No of pixels in calibration dataset	No of pixels in validation dataset
Raba	mid 1990s	132	2310	578
	late 2000s	66		
Dunajec	mid 1990s	66	1382	346
	late 2000s	165		
Sola	mid 1990s	66	1507	376
	late 2000s	99		

3.2. Machine learning algorithms

In presented study following nine machine learning regression algorithms were compared: Cubist (Quinlan, 1993), Random Forest (RF) (Breiman, 2001), stochastic gradient boosting of regression trees (GBM) (Friedman, 2002), k-nearest neighbors (kNN), random k-nearest neighbors (rkNN) (Li et al., 2011), Multivariate Adaptive Regression Splines (MARS) (Friedman, 1991), averaged neural networks (avNN) (Ripley, 1996), support vector machines (Smola and Schölkopf, 2004) with polynomial (SVMp) and radial (SVMr) kernels.

Random Forest uses an ensemble of classification or regression trees, usually built with CART (Breiman et al., 1984) algorithm. Each of these trees is built on randomly selected subsets of training data and using a random subset of classification features at each split. Final value is based on the averaged responses of each tree. This approach was previously used for sub-pixel imperviousness assessment based on Landsat (Walton, 2008); (Bernat and Drzewiecki, 2014) and MODIS images (Deng and Wu, 2013; Tsutsumida, 2016). It was also chosen by U.S. Forest Service to produce the 2011 NLCD percent tree canopy cover dataset (Freeman et al., 2016).

GBM is based on the principle of boosting algorithms, in which weak classifiers (having prediction accuracy marginally better than chance) are ensembled and their predictions combined. Gradient boosting machines may be used both for classification and regression. In case of regression the algorithm starts from computing the average response. Then the residuals from observed values are calculated and used to fit a regression tree. The model is added to the previous one till the user-specified number of iterations is reached. In case of stochastic gradient boosting, the models in each iteration are based only on the randomly selected fraction of training data (Kuhn and

Johnson, 2013). This method was used to estimate aboveground biomass based on regression with Landsat and SPOT (Güneralp et al., 2014) or ALOS PALSAR derived predictors (Carreiras, 2013) and predicting tree canopy cover based on Landsat-5 images (Freeman, 2016). To my best knowledge, it has not been used for sub-pixel imperviousness evaluation so far.

Cubist algorithm differs from CART in criteria used for constructing the trees and the pruning approach. Models can be also boosted, but in a different way than in GBM. Training sample values are increased for underpredicted and decreased for overpredicted points, new models fitted and the final prediction obtained as the simple average of individual model predictions (Kuhn and Johnson, 2013). This prediction may be further adjusted based on neighboring samples from the training set. Cubist is quite often used in sub-pixel ISA assessment (Yang et al., 2003; Walton, 2008; Mohapatra and Wu, 2010; Bernat and Drzewiecki, 2014). This algorithm was also chosen for ISA mapping in case of US National Land Cover Database (Homer et al., 2007).

In kNN approach to predict the value for considered sample its k nearest neighbors in training data are identified first. Then the predicted value is calculated as the weighted mean of their responses. The distance between samples can be calculated using different metrics. Different weighting methods may be used as well. The random kNN algorithm was developed as randomized kNN generalization for high dimensional datasets (Li et al., 2011). In this approach an ensemble of kNN models is created using random subsets of input features. The kNN approach is widely used for forest biomass and carbon estimations based on remotely sensed images (Sun et al., 2015; Galeana-Pizaña et al., 2016). However, I have not found any application of kNN regression for sub-pixel imperviousness mapping. In case of rkNN no application to remotely sensed data was found.

MARS is another non-linear regression model. In this algorithm the input features are divided into two intervals and for each of them a separate piecewise linear regression model is created. The final model is built from these functions and their products (Hastie et al., 2009). This method was used to estimate soil properties and biomass through regression with variables derived from satellite imagery (Fillipi et al., 2014; Nawar et al., 2014; Nawar et al., 2015). Again, no example of ISA evaluation has been found.

Apart from regression trees, artificial neural networks are non-linear regression techniques most often used for sub-pixel imperviousness assessment. From many existing kinds of neural networks, the multi-layer perceptron (MLP) is the most frequently chosen (Mohapatra and Wu, 2007; Chormanski et al., 2008; Shao and Lunetta, 2011; Bernat and Drzewiecki, 2014). MLP structure is build of layers of nodes (called neurons). Each node in input layer represents the input variable. In case of regression, the output layer consists of one neuron (Hastie et al., 2009). Between the input and output layers, there is one or more hidden layers of neurons (hidden units). Each neuron in hidden layer is a linear combination of input variables (or hidden units from the previous layer), usually transformed using a non-linear function

(typically sigmoidal). The final prediction is obtained as linear combination of the hidden units from the last hidden layer (Hastie et al., 2009; Kuhn and Johnson, 2013). To assure the best performance of MLP model the input variables should be centered and scaled and highly correlated predictors should be removed (Kuhn and Johnson, 2013). As large number of parameters have to be estimated when fitting the MLP model, the solution is very often only locally optimal. As a result, different models giving comparable performance can be obtained. To increase the stability of the MLP prediction one may create several models using different initial values of parameters and average their results (Perrone and Cooper, 1993; Kuhn and Johnson, 2013).

Support vector machines is a non-parametric statistical learning technique developed originally for classification (Vapnik and Chervonenkis, 1971; Vapnik, 2010). Initially linear classifier was extended to non-linear classification boundaries thanks to implementation of kernel function (Boser et al., 1992). The most frequently used are linear, polynomial, sigmoid and radial basis function kernels (Qian et al., 2015). SVMs were adapted for regression by Smola and Schölkopf (2004). In this approach the samples that fit with residuals within the defined threshold are ignored and do not influence the regression equation. On the other hand, the effect of large outliers is mitigated, as the remaining samples contribute to the regression fit in linear manner (Hastie et al., 2009; Kuhn and Johnson, 2013). SVM regression was tested for sub-pixel imperviousness assessment by Walton (2008), Esch et al. (2008) and Xi et al. (2011).

3.3. Model tunning and ISA predictions

Each of tested machine learning algorithms was used to predict imperviousness for mid 1990s and late 2000s. All predictive models were trained using the caret package (Kuhn, 2008) in R environment. In every case the regression models were trained and evaluated on calibration dataset using 10-times repeated 5-fold cross-validation. For kNN, rkNN, neural networks and SVM algorithms input data were preprocessed by centering and scaling. In case of neural network the correlated predictors were removed using the approach proposed in (Kuhn and Johnson, 2013).

As for all models, the training and test data indexes were kept the same in each cross-validation fold, it was possible to compare the performances of the models trained with different sets of parameters. Model parameters tuned in cross-validation approach are shown in Table 3. The combination of parameters resulting in the lowest average RMS error was then used in final model. These models were fitted to the whole calibration dataset and used for ISA prediction. Then ISA change was calculated by subtracting the assessments obtained with the same method. As a result, nine change maps were created for every study area.

Table 3. Model parameters tuned

Machine learning model	Model parameter	Tested values
Cubist	number of committees	1; 5; 10; 50; 75; 100
	neighbors	0; 1; 3; 5; 7; 9
RF (500 trees)	number of randomly selected predictors	15 values evenly distributed between 2 and maximum number of predictors
GBM	number of iterations (trees)	from 100 to 1500 by 100
	complexity of the tree (tree depth)	from 1 to 9 by 1
	learning rate	0.001; 0.01; 0.1
	minimum number of training set samples in a node	from 2 to 20 by 2
kNN	number of neighbors	from 1 to 15 by 1
	kernel functions to weight the neighbors	triangulal; rectangular; epanechnikov; optimal
	parameter of Minkowski distance	1; 2; 3
rkNN (500 kNNs)	number of randomly selected predictors	15 values evenly distributed between 2 and maximum number of predictors
	number of neighbors	from 1 to 15 by 1
MARS	maximum degree of interaction	1; 2
	maximum number of terms	from 1 to 50 by 1
avNN 5 feed-forward neural networks with a single hidden layer (maximum number of iterations = 500)	number of neurons in the hidden layer	from 1 to 12 by 1
	weight decay	0; 0.1; 0.01; 0.001; 0.0001
SVMp	degree	1; 2; 3
	cost	0.25; 0.5; 1; 2; 4; 8
	scale factor	0.001; 0.005; 0.01; 0.05; 0.1
SVMr	cost	0.25; 0.5; 1; 2; 4; 8; 16; 32; 64; 128; 256; 512; 1024; 2048; 4096

3.4. Accuracy assessment

The root mean squared error (RMSE) and the mean absolute error (MAE) were used to assess the performance of final models built with particular tested algorithms. The evaluation was done for both individual time points as well as for the prediction of ISA change.

As in each cross-validation fold individual models were trained and tested using identically resampled datasets, their accuracies could be compared with approach proposed by Hothorn et al. (2005) and Eugster et al. (2008). A paired t-test with Bonferroni correction for multiple comparisons (Dunn, 1961) was used to determine if the differences between accuracies of models trained using different algorithms are statistically significant. The same approach was also used to compare the results obtained for validation dataset.

3.5. Model ensembles

According to Mendes-Moreira et al. (2012) the process of ensemble learning can be divided into three stages: ensemble generation, ensemble pruning and ensemble integration. In the first step a set of models for ensemble is created. During the pruning stage some of them are eliminated. The last step is to define a way of combining the models remaining in the set after the pruning. In some cases a strategy adopted for models combination have to be chosen before pruning as decision about selection or elimination of particular model is taken based on the performance of the ensemble (Coelho and Von Zuben, 2006). A wide range of methods have been proposed for each of ensemble learning stages. Mendes-Moreira et al. (2012) reviews the ones used for regression model ensembles.

In presented research the initial set of models for individual time points consisted of the best models generated with particular algorithms. Then model ensembles were created using forward and backward selection search schemas (FSS and BSS, respectively) proposed by Perrone and Cooper (1993) and Coelho and Von Zuben (2006). All candidate models were ranked according to their performance (RMSE) during cross-validation. The FSS approach starts from inserting the best model into the ensemble. Then the next one (the second best) is added and the performance of created ensemble evaluated. If it improves, the added model stays in the ensemble. Otherwise, it is removed and the next one is tested. The process is repeated till the ensemble performance can not be improved by adding subsequent model. In BSS approach all models are initially inserted into the ensemble. Then, the model which performed the worst is removed. If the performance of the ensemble is improved, the model is dropped. Otherwise, the model returns into the ensemble set and the second worst is tested. The process is repeated till the ensemble performance does not improve after removal of any model. In cases when different sets of model were obtained from FSS and BSS approaches, the model ensemble having lower RMS error was chosen as the final one.

As the approach adopted for ensemble pruning requires ensemble performance evaluation, the ensemble integration method has to be defined as well. In this study the Basic Ensemble Method (BEM) (Perrone and Cooper, 1993) was used and the prediction of the ensemble calculated simply as the mean of the predictions of the models in the ensemble set.

The accuracies of ISA and ISA change predictions obtained using model ensembles were assessed in the same manner as for single models and compared to the best of them.

4. Results

4.1. Performance of individual machine learning algorithms

This section presents the results obtained using individual tested machine learning approaches for sub-pixel ISA and ISA change mapping in researched study areas. Table 4 presents the values of model parameters for particular machine learning algorithms established in model tuning procedure described in Sect. 3.3.

The performance measures (RMSE and MAE) of these models calculated in cross-validation approach are shown in Tab. 5–7 and for validation dataset in Tab. 8–10. The lowest errors are bolded. Values of errors for models which performed equally well according to paired t-test are also indicated and information on statistical significance is provided. The results for Raba dataset were published earlier in Drzewiecki (2016). However, they are included here as they are necessary both for comparison with results obtained in other study areas and for presentation of model ensembling effect.

Table 4. Values of model parameters

Machine learning model	Model parameter	Raba		Dunajec		Sola	
		Mid 1990s	Late 2000s	Mid 1990s	Late 2000s	Mid 1990s	Late 2000s
Cubist	number of committees	100	100	100	100	75	75
	neighbors	7	0	9	7	8	7
RF (500 trees)	number of randomly selected predictors	11	29	43	83	38	64
GBM	number of iterations (trees)	1500	1200	1500	1500	1500	1000
	complexity of the tree (tree depth)	9	8	9	9	9	9
	learning rate	0.01	0.01	0.01	0.01	0.01	0.01
	minimum number of training set samples in a node	6	2	6	4	4	6
kNN	number of neighbors	11	12	7	8	9	8
	kernel functions to weight the neighbors	triangulal	triangulal	triangulal	triangulal	triangulal	optimal
	parameter of Minkowski distance	3	2	3	2	1	1

Machine learning model	Model parameter	Raba		Dunajec		Sola	
		Mid 1990s	Late 2000s	Mid 1990s	Late 2000s	Mid 1990s	Late 2000s
rkNN (500 kNNs)	number of randomly selected predictors	29	15	34	48	20	22
	number of neighbors	5	5	5	5	5	5
MARS	maximum degree of interaction	1	2	2	2	2	2
	maximum number of terms	19	29	11	14	27	38
avNN 5 feed-forward neural networks with a single hidden layer (maximum number of iterations = 500)	number of neurons in the hidden layer	12	12	12	8	8	12
	weight decay	0.1	0.001	0.001	0.01	0.01	0.1
SVMp	degree	2	2	2	3	3	3
	cost	0.25	8	0.25	1	4	0.5
	scale factor	0.001	0.05	0.05	0.005	0.005	0.01
SVMr	cost	4	4	8	4	2	2

Table 5. ISA models performance (values averaged from cross-validation on calibration dataset) – Raba dataset

Method	Mid 1990s		Late 2000s		Correlation coefficient of prediction errors	ISA Change	
	RMSE	MAE	RMSE	MAE		RMSE	MAE
avNN	0.1149	0.0680	0.1154	0.0675	0.5813	0.1053	0.0626
RF	0.1099	0.0653	0.1116[b]	0.0650	0.6680[b]	**0.0902**[a]	0.0520
Cubist	0.1074[b]	0.0608[c]	**0.1115**[a]	**0.0619**[a]	0.6246	0.0953	0.0531
GBM	**0.1073**[a]	0.0617	0.1119[b]	0.0640	0.6268	0.0947	0.0540
kNN	0.1080[b]	**0.0599**[a]	0.1166	0.0656	0.6226	0.0978	0.0543
rkNN	0.1074[b]	0.0620	0.1165	0.0664	**0.6680**[a]	0.0917[b]	**0.0505**[a]
SVMp	0.1267	0.0734	0.1132[c]	0.0643	0.5893	0.1095	0.0640
SVMr	0.1085[b]	0.0613	0.1153	0.0647	0.6269	0.0967	0.0575
MARS	0.1191	0.0731	0.1179	0.0680	0.5840	0.1081	0.0656

[a] The best performed model
[b] No significant difference to the best performed model at p-value>0.1
[c] No significant difference to the best performed model at p-value>0.001

Table 6. ISA models performance (values averaged from cross-validation on calibration dataset) – Dunajec dataset

Method	Mid 1990s		Late 2000s		Correlation coefficient of prediction errors	ISA Change	
	RMSE	MAE	RMSE	MAE		RMSE	MAE
avNN	RMSE	0.1098	0.1414	0.0961	0.5384	0.1467	0.1043
RF	0.1623[c]	0.1089[d]	0.1390	0.0956	**0.6344[a]**	**0.1286[a]**	**0.0889[a]**
Cubist	**0.1590[a]**	0.1070[b]	**0.1312[a]**	**0.0876[a]**	0.5632	0.1389	0.0954
GBM	0.1609[b]	0.1068[b]	0.1324[b]	0.0893[d]	0.5865	0.1346	0.0923
kNN	0.1594[b]	0.1063[b]	0.1434	0.0932	0.5292	0.1516	0.1004
rkNN	0.1668	0.1131	0.1464	0.0979	0.5644	0.1497	0.1015
SVMp	0.1716	0.1160	0.1462	0.0978	0.5884	0.1538	0.1032
SVMr	0.1839	**0.1062[a]**	0.1350[c]	0.0927	0.5548	0.1474	0.1000
MARS	0.1698	0.1245	0.1616	0.1113	0.5350	0.1655	0.1158

[a] The best performed model
[b] No significant difference to the best performed model at p-value>0.1
[c] No significant difference to the best performed model at p-value>0.01
[d] No significant difference to the best performed model at p-value>0.001

Table 7. ISA models performance (values averaged from cross-validation on calibration dataset) – Sola dataset

Method	Mid 1990s		Late 2000s		Correlation coefficient of prediction errors	ISA Change	
	RMSE	MAE	RMSE	MAE		RMSE	MAE
avNN	0.1338	0.0822	0.1360	0.0861	0.5416	0.1290	0.8340
RF	0.1257	0.0787	0.1210	0.0756	**0.6185[a]**	**0.1077[a]**	**0.0667[a]**
Cubist	0.1213[b]	**0.0720[a]**	**0.1178[a]**	**0.0703[a]**	0.5564	0.1124	0.0695
GBM	**0.1205[a]**	0.0728[b]	0.1196[b]	0.0723	0.5820	0.1096[b]	0.0677[b]
kNN	0.1289	0.0740[d]	0.1276	0.0755	0.5206	0.1257	0.0760
rkNN	0.1293	0.0778	0.1288	0.0799	0.5936[d]	0.1163	0.0704
SVMp	0.1279	0.0776	0.1254	0.0768	0.6179[b]	0.1107[b]	0.0685[b]
SVMr	0.1255	0.0762	0.1262	0.0767	0.5450	0.1199	0.0746
MARS	0.1346	0.0836	0.1310	0.0799	0.5492	0.1290	0.0791

[a] The best performed model
[b] No significant difference to the best performed model at p-value>0.1
[c] No significant difference to the best performed model at p-value>0.01
[d] No significant difference to the best performed model at p-value>0.001

Table 8. ISA models performance on validation dataset – Raba dataset

Method	Mid 1990s		Late 2000s		Correlation coefficient of prediction errors	ISA Change	
	RMSE	MAE	RMSE	MAE		RMSE	MAE
avNN	0.1218[c]	0.0736	0.1131[b]	0.0679[d]	0.6410	0.0998[c]	0.0629
RF	0.1128[b]	0.0678[c]	0.1120[b]	0.0647[b]	0.7124	**0.0853**[a]	0.0504[b]
Cubist	**0.1086**[a]	**0.0624**[a]	**0.1084**[a]	**0.0611**[a]	0.6787	0.0875[b]	0.0505[b]
GBM	0.1117[b]	0.0639[b]	0.1116[b]	0.0639[b]	0.6939	0.0873[b]	0.0505[b]
kNN	0.1117[b]	0.0636[b]	0.1182[b]	0.0680[c]	0.6704	0.0938[b]	0.0534[c]
rkNN	0.1098[b]	0.0640[b]	0.1167[b]	0.0684	0.7118	0.0863[b]	**0.0486**[a]
SVMp	0.1334	0.0783	0.1091[b]	0.0625[b]	0.6656	0.1012[c]	0.0620
SVMr	0.1154[b]	0.0665[b]	0.1120[b]	0.0654[b]	**0.7174**[a]	0.0854[b]	0.0522[b]
MARS	0.1248[d]	0.0768	0.1110[b]	0.0629[b]	0.7050	0.0915[b]	0.0594

[a] The best performed model
[b] No significant difference to the best performed model at p-value>0.1
[c] No significant difference to the best performed model at p-value>0.01
[d] No significant difference to the best performed model at p-value>0.001

Table 9. ISA models performance on validation dataset – Dunajec dataset

Method	Mid 1990s		Late 2000s		Correlation coefficient of prediction errors	ISA Change	
	RMSE	MAE	RMSE	MAE		RMSE	MAE
avNN	**0.1496**[a]	0.1041[b]	0.1396[b]	0.0954[b]	0.5458	0.1383[b]	0.0987[b]
RF	0.1500[b]	0.1051[b]	0.1368[b]	0.0951[c]	0.5793	0.1316[b]	0.0889[b]
Cubist	0.1538[b]	0.1008[b]	0.1441[b]	0.0948[b]	0.4966	0.1494[b]	0.0954[b]
GBM	0.1509[b]	0.1039[b]	0.1335[b]	**0.0895**[a]	0.5825	**0.1307**[a]	**0.0873**[a]
kNN	0.1670[b]	0.1046[b]	0.1478[b]	0.0961[b]	0.5505	0.1493[b]	0.0952[b]
rkNN	0.1638[b]	0.1101[c]	0.1452[b]	0.0977[b]	**0.5867**[a]	0.1403[b]	0.0929[b]
SVMp	0.1543[b]	0.0987[b]	0.1384[b]	0.0941[b]	0.5707	0.1366[b]	0.0920[b]
SVMr	0.1533[b]	**0.0962**[a]	**0.1332**[a]	0.0917[b]	0.5571	0.1364[b]	0.0921[b]
MARS	0.1662[b]	0.1176	0.1534[b]	0.1104	0.5348	0.1542[b]	0.1069[d]

[a] The best performed model
[b] No significant difference to the best performed model at p-value>0.1
[c] No significant difference to the best performed model at p-value>0.01
[d] No significant difference to the best performed model at p-value>0.001

Table 10. ISA models performance on validation dataset – Sola dataset

Method	Mid 1990s		Late 2000s		Correlation coefficient of prediction errors	ISA Change	
	RMSE	MAE	RMSE	MAE		RMSE	MAE
avNN	0.1156[b]	0.0702[c]	0.1386[c]	0.0818	0.4103	0.1396	0.0856
RF	0.1062[b]	0.0645[b]	0.1113[b]	0.0676[b]	0.5266	0.1060[b]	0.0632[b]
Cubist	0.1074[b]	0.0608[b]	0.1158[b]	0.0670[b]	0.4489	0.1176[b]	0.0692[b]
GBM	**0.1002[a]**	**0.0592[a]**	0.1171[b]	0.0692[b]	0.4431	0.1157[b]	0.0688[b]
kNN	0.1203[b]	0.0641[b]	0.1104[b]	**0.0633[a]**	0.4372	0.1227[b]	0.0687[b]
rkNN	0.1139[b]	0.0656[b]	0.1113[b]	0.0670[b]	**0.5604 [a]**	**0.1057[a]**	**0.0618[a]**
SVMp	0.1206[b]	0.0689[b]	**0.1104[a]**	0.0648[b]	0.5510	0.1097[b]	0.0634[b]
SVMr	0.1061[b]	0.0617[b]	0.1154[b]	0.0690[b]	0.4751	0.1139[b]	0.0682[b]
MARS	0.1199[b]	0.0736[d]	0.1245[b]	0.0745[b]	0.4085	0.1330[b]	0.0809[d]

[a)] The best performed model
[b)] No significant difference to the best performed model at p-value>0.1
[c)] No significant difference to the best performed model at p-value>0.01
[d)] No significant difference to the best performed model at p-value>0.001

Following observations can be made when looking into presented results:

- In cross-validation based evaluation of imperviousness mapping accuracy for individual time points, the best result according to RMSE measure were obtained using regression trees algorithms. For late 2000s the Cubist algorithm performed the best in every study area. For mid 1990s GBM gave the highest accuracy in Raba and Sola catchments, and RF performed the best for Dunajec dataset. It should be noted that in every case the Cubist and GBM methods were the best or did not differ significantly from the best approaches.
- When MAE measure is considered the set of the best algorithms for individual time point ISA mapping differs. The k-nearest neighbors algorithm is the best for mid 1990s Raba assessment, SVM approach with radial basis function kernel – for mid 1990s in Dunajec catchment and Cubist for the four remaining evaluations. The last one (Cubist) is the only one which is the best or not significantly different from the best one in every case of individual time point assessments.
- When looking from the ISA change mapping point of view the image is a little bit different. According to cross-validation results the Random Forest approach gave the lowest RMS errors in every area and the lowest MAE for Dunajec and Sola datasets. In case of Raba dataset the lowest mean absolute error was obtained with random k-NN algorithm. The Random Forest method gave the second best result, however significantly worse than rkNN.
- In cross-validation the best ISA change assessment was obtained using algorithms which gave the most correlated errors of individual time point evaluations. In case

of Random Forest algorithm in each study area the average correlation coefficient was the highest or did not differ significantly from the highest one.

- Evaluation based on validation dataset do not give clear results as for most of the evaluated algorithms the accuracy measures do not differ significantly. Nevertheless, it should be stressed that the values of both performance measures are at similar levels as in cross-validation approach. The influence of correlation between the individual time point assessments errors on the ISA change accuracy is also noticeable.

4.2. Model ensembles

Table 11 presents the best model ensembles obtained for individual time point assessments using forward and backward selection search schemas (as described in section 3.5) together with their average RMSE and MAE values from cross-validation. In every case the best (giving the lowest RMSE) ensambles were constructed using BSS approach. Accuracy parameters of ISA change evaluation based on these models were calculated as well.

Table 11. Performance of model ensembles
(values averaged from cross-validation on calibration dataset)

Dataset	Mid 1990s				Late 2000s				Correlation coefficient of prediction errors	ISA Change	
	Ensembled models	Selection search schema	RMSE	MAE	Ensembled models	Selection search schema	RMSE	MAE		RMSE	MAE
Raba	GBM, CUB, kNN, SVMr, avNN	BSS	0.1039	0.0597	CUB, RF, GBM, avNN, kNN, SVMp	FSS/BSS	0.1094	0.0626	0.6607	0.0879	0.0498
Dunajec	avNN, RF, GBM, kNN	BSS	0.1544	0.1036	CUB, GBM, SVMr	FSS/BSS	0.1279	0.0870	0.6041	0.1275	0.0879
Sola	GBM, CUB, SVMr, kNN	FSS/BSS	0.1185	0.0713	CUB, GBM, kNN, SVMp	BSS	0.1167	0.0710	0.6032	0.1047	0.0646

In every case model ensembles resulted in lower RMSE values than obtained using the best models trained with individual algorithms. Only for late 2000s Sola assessment no significant difference to the best performed individual model (Cubist) was observed (p-value = 0.022). The remaining five assessments gave significantly lower RMS errors. However, MAE value is significantly lower only for mid 1990s Dunajec evaluation. Moreover, for late 2000s Raba assessment the individual model (Cubist) gave even significantly better result.

When ISA change assessment is considered, the model ensembles gave better results. In every study area both, RMSE and MAE values are lower than for the best individual algorithms. Statistical significance is observed however only in three cases (Raba RMSE as well as Sola RMSE and MAE). It should be noted that correlation between errors of individual assessments giving the highest ISA change prediction accuracy is higher than for model ensembles. But their final performance is worse due to lower individual time points prediction accuracies.

Model ensembles were also used to predict imperviousness for validation datasets (Table 12). In this case no statistically significant differences to the best individual models were observed. Nevertheless, what should be noted are lower errors (both RMSE and MAE) of ISA change prediction in every study area. Also RMSE values obtained for individual time point ISA evaluations are better for model ensembles, except for mid 1990s Sola assessment.

Table 12. Performance of model ensembles on validation datasets

Dataset	Mid 1990s		Late 2000s		Correlation coefficient of prediction errors	ISA Change	
	RMSE	MAE	RMSE	MAE		RMSE	MAE
Raba	0.1082	0.0629	0.1081	0.0630	0.7272	0.0799	0.0470
Dunajec	0.1471	0.1007	0.1322	0.0880	0.5873	0.1272	0.0841
Sola	0.1013	0.0578	0.1068	0.0630	0.5169	0.1025	0.0614

5. Discussion

Regarding the accuracy of imperviousness maps made for individual time points, the Cubist algorithm outperformed the other ones. It was the only one constantly present in the group of the best models, irrespectively of the dataset and the measure used for performance evaluation (RMSE or MAE). This finding is consistent with the previously reported in the literature. Cubist was compared with other machine learning approaches to imperviousness mapping in studies presented by Walton (2008) and Mohapatra and Wu (2010). Walton (2008) mapped imperviousness in the city of Syracuse (USA) from Landsat-7 ETM+ images. He compared Cubist, Random Forest and support vector regression algorithms and found Cubist the best of them.

Mohapatra and Wu (2010) evaluated ISA in Grafton (USA) using prediction variables derived from Ikonos and Landsat TM images. In their study linear regression was compared to Cubist and artificial neural network (MLP with a single hidden layer). Cubist gave the best results, however the differences between this method and MLP were not statistically significant.

Despite the best results yielded by Cubist algorithm for individual time points evaluations, this method was not the best one for ISA change assessment. When comparing individual algorithms for such application, Random Forest seems to be better than others. Having lower accuracies for individual time point assessments, this method allowed for more accurate ISA change evaluation in every study area. These results cannot be confronted with other findings as any comparison of machine learning algorithms for sub-pixel ISA change assessment was published before. One should remember, however, that as performance of machine learning algorithms is problem-depended, the best method may differ depending on application. The findings of Walton (2008) may be recalled as the example – in his study the Cubist algorithm was the best for imperviousness mapping, but support vector regression was found the best for sub-pixel forest canopy cover assessment using the same image dataset.

The explanation of differences in the most accurate methods for sub-pixel imperviousness mapping in individual time points and ISA change evaluation can be found in differences of correlation of individual time points prediction errors, as illustrated in Eq.1. The implication of this formula, confirmed experimentally in this research, is that the most accurate assessment of change not necessarily is based on the most accurate assessments of individual states. It should be stressed that these rule, although proved here on the example of imperviousness change evaluation, is universal and valid for all studies where the change is predicted as the difference of two values obtained from independent non-linear regression models.

As expected, when RMSE is considered ensembles of heterogeneous non-linear regression models outperformed the best single approaches for mapping imperviousness in individual time points. They always gave lower RMS errors in cross-validation and in five of six cases the difference was statistically significant. For validation datasets the differences between RMS errors of the best individual method and model ensembles were not significant. Nevertheless, model ensembles were more accurate in five of six assessments.

When MAE measure is considered, performance of model ensembles for individual time points ISA assessment is also usually better than performance of single models. However, in some cases single models gave lower mean absolute errors. This can be explained by the fact that the ensembles were constructed to minimize RMS errors.

From ISA change assessment point of view, model ensembles always outperformed single model approaches. In every study area, both in cross-validation and using validation dataset, values of RMSE and MAE were lower for ensembled models. It means that by constructing ensembles based on model performances for individual

time points solely, one can obtain ensembled model which is at least as good as the best of individual models both for individual time points imperviousness assessment and ISA change evaluation. To construct such ensembles we do not need reference data for ISA change. Reference information about imperviousness in single dates is enough, whereas existence of ISA change reference data is necessary to determine the best performing individual model. This is very important as in many cases it is not possible to verify the accuracy of ISA change evaluation due to the lack of the reference areas (Yang et al., 2003; Dams et al. 2013).

6. Conclusions

The first aim of the presented study was to compare selected non-linear regression algorithms in the context of sub-pixel imperviousness mapping from Landsat images, both for individual time points assessments and for ISA change evaluation. Comparison was done for nine machine learning methods: Cubist, Random Forest, stochastic gradient boosting of regression trees, k-nearest neighbors regression, random k-nearest neighbors regression, Multivariate Adaptive Regression Splines, averaged neural networks, and support vector machines with polynomial and radial kernels. Obtained results confirmed the ones presented by Drzewiecki (2016). Cubist algorithm outperformed the other techniques for imperviousness evaluation in single time steps. However, when the goal is to assess imperviousness change Random Forest would be better choice. Despite lower accuracies for individual time point predictions, it allowed for more accurate change evaluation thanks to more correlated errors of individual assessments.

Presented study was also designed to answer a question if ensembling of heterogeneous non-linear regression models may increase the accuracy of both, individual time points ISA assessments and ISA change evaluation. Obtained results are encouraging. Using backward selection schema approach it was possible to construct from the best models trained using individual algorithms model ensembles which performed at least not worse than best individual models for single time points. These ensembles outperformed individual algorithms in ISA change evaluation. It is important as it means that one can possibly improve sub-pixel imperviousness change assessment even without reference change information.

Acknowledgments

Research founded by AGH University grant No. 11.11.150.949. Datasets used in this research were prepared by Ms. Katarzyna Bernat under Author's supervision for the needs of SaLMaR project co-financed by the Ministry of Science and Higher Education of Poland and the German Federal Ministry for Education and Research.

References

Arnold, C.L. Jr. and Gibbons, C.J. (1996). Impervious surface coverage: the emergence of a key environmental indicator. *Journal of the American Planning Association*, 62, 243–258.

Bauer, M.E., Heiner, N.J., Doyle, J.K. and Yuan, F. (2004). Impervious surface mapping and change monitoring using Landsat remote sensing. ASPRS Annual Conference Proc., unpaginated CD-ROM.

Bernat, K. and Drzewiecki W. (2014). Two-stage subpixel impervious surface coverage estimation: comparing classification and regression trees and artificial neural networks. In: Proc. SPIE Vol. 9244, 92441l, Image and Signal Processing for Remote Sensing. DOI: 10.1117/12.2067308

Boser, B., Guyon, I. and Vapnik, V.A. (1992). Training Algorithm for Optimal Margin Classifiers. In: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 144–152.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32.

Breiman, L, Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984). *Classification and Regression Trees*. CRC Press.

Caldwell, P.V., Sun, G., McNulty, S.G., Cohen, E.C. and Moore Myers, J.A. (2012). Impacts of impervious cover, water withdrawals, and climate change on river flows in the conterminous US, *Hydrol. Earth Syst. Sci.*, 16, 2839–2857. DOI: 10.5194/hess-16-2839-2012

Carreiras, J.M.B., Melo, J.B. and Vasconcelos, M.J. (2013). Estimating the Above-Ground Biomass in Miombo Savanna Woodlands (Mozambique, East Africa) Using L-Band Synthetic Aperture Radar Data. *Remote Sens.*, 5, 1524-1548. DOI: 10.3390/rs5041524

Chormański, J., Van de Voorde, T., De Roeck, T., Batelaan, O. and Canters, F. (2008). Improving distributed runoff prediction in urbanized catchments with remote sensing based estimates of impervious surface cover. *Sensors*, 8, 910–932.

Coelho, G.P. and Von Zuben, F.J. (2006). The influence of the pool of candidates on the performance of selection and combination techniques in ensembles. In: Proceedings of the International Joint Conference on Neural Networks, 10588–10595.

Dams, J., Dujardin, J., Reggers, R., Bashir, I., Canters, F. and Batelaan, O. (2013). Mapping impervious surface change from remote sensing for hydrological modeling. *Journal of Hydrology*, 485, 84–95. DOI: 10.1016/j.jhydrol.2012.09.045

Deng, C. and Wu, C. (2012). BCI: A biophysical composition index for remote sensing of urban environments. *Remote Sensing of Environment*, 127, 247–259.

Deng, C. and Wu, C. (2013). The use of single-date MODIS imagery for estimating large-scale urban impervious surface fraction with spectral mixture analysis and machine learning techniques. *ISPRS Journal of Photogrammetry and Remote Sensing*, 86, 100–110. DOI: 10.1016/j.isprsjprs.2013.09.010

Drzewiecki, W. (2016). Comparison of Selected Machine Learning Algorithms for Sub-Pixel Imperviousness Change Assessment. In: *2016 Baltic Geodetic Congress (Geomatics)*, 67–72. DOI: 10.1109/BGC.Geomatics.2016.21

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, 56, 293, 52– 64.

Engler, R., Waser L.T., Zimmermann, N.E., Schaub, M., Berdos, S., Ginzler, C. and Psomas, A. (2013). Combining ensemble modeling and remote sensing for mapping individual tree species at high spatial resolution. *Forest Ecology and Management*, 310, 64–73. DOI: 10.1016/j.foreco.2013.07.059

Esch, T., Conrad, C., Schorcht, G., Thiel, M. Wehrmann, T. and Dech, S. (2008). Model-based estimation of impervious surface by application of support vector machines. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* Vol. XXXVII, Part B8, 41–44.

Eugster, M., Hothorn, T. and Leisch, F. (2008). Exploratory and Inferential Analysis of Benchmark Experiments. Ludwigs-Maximilians-Universität München, Department of Statistics, Tech. Rep, 30.

Fillipi, A.M., Güneralp, İ. and Randal, J. (2014). Hyperspectral remote sensing of aboveground biomass on a river meander bend using multivariate adaptive regression splines and stochastic gradient boosting. *Remote Sensing Letters*, 5, 5, 432–441. DOI: 10.1080/2150704X.2014.915070

Freeman, E.A., Moisen, G.G., Coulston, J.W. and Wilson, B.T. (2016). Random forest and stochastic gradient boosting for predicting tree canopy cover: comparing tuning processes and model performance. *Can. J. For. Res.*, 46, 323–339. DOI: 10.1139/cjfr-2014-0562

Friedman, J. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1), 1–141.

Friedman, J. (2002). Stochastic Gradient Boosting. *Computational Statistics and Data Analysis*, 38(4), 367–378

Galeana-Pizaña, J.M., Núñez Hernández, J.M. and Romero, N.C. (2016). Remote Sensing-Based Biomass Estimation. In: M. Marghany (ed.), *Environmental Applications of Remote Sensing*, InTech, DOI: 10.5772/61813. Available online at: http://www.intechopen.com/books/environmental-applications-of-remote-sensing/remote-sensing-based-biomass-estimation

García-Pedrajas, N., Hervás-Martínez, C. and Ortiz-Boyer, D. (2005). Cooperative coevolution of artificial neural network ensembles for pattern classification. *IEEE Trans. Evolut. Comput*., 9, 3, 271–302.

Gómez-Chova, L., Muñoz-Marí, J., Amorós-López, J., Izquierdo-Verdiguier, E. and Camps-Valls, G. (2013). Advances in synergy of AATSR-MERIS sensors for cloud detection. In: I*EEE Int. Geos. Rem. Sensing Symposium (IGARSS)* 2013, Jul 2013, pp. 4391–4394.

Güneralp, İ, Filippi, A.M. and Randall, J. (2014). Estimation of floodplain aboveground biomass using multispectral remote sensing and nonparametric modeling. *International Journal of Applied Earth Observation and Geoinformation*, 33, 119–126. DOI: 10.1016/j.jag.2014.05.004

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Minig, Inference, and Prediction*. Springer.

Heinermann, J. and Kramer, O. (2015). On heterogeneous machine learning ensembles for wind power prediction. In: Computational Sustainability: Papers from the 2015 AAAI Workshop, pp. 54–59.

Heremans, S. and Van Orshoven, J. (2015). Machine learning methods for sub-pixel land-cover classification in the spatially heterogeneous region of Flanders (Belgium): a multi-criteria comparison. *International Journal of Remote Sensing*, 36, 11, 2934–2962. DOI: 10.1080/01431161.2015.1054047

Homer, C., Dewitz, J., Fry, J., Coan M., Hossain, N., Larson, C., Herold, N., McKerrow, A., Van Driel, J.N. and Wickham, J. (2007). Completion of the 2001 National Land Cover Database for the Conterminous United States. *Photogrammetric Engineering and Remote Sensing*, 73, 337–341.

Hothorn, T., Leisch F., Zeileis A. and Hornik K., (2005): The design and analysis of benchmark experiments. *Journal of Computational and Graphical Statistics*, 14, 3, 675–699.

Hussain, M., Chen, D., Cheng, A., Wei, H. and Stanley, D. (2013). Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, 80, 91–106. DOI: 10.1016/j.isprsjprs.2013.03.006

Kew, W. and Mitchel, J.B.O. (2015). Greedy and linear ensembles of machine learning methods outperform single approaches for QPSR regression problems. *Molecular Informatics*, 34, 634–647.

Kircher, J. (2001). Data Analysis Toolkit #5: Uncertainty Analysis and Error Propagation. University of California Berkeley Seismological Laboratory. Available online at: *http://seismo.berkeley.edu/~kirchner/eps_120/Toolkits/Toolkit_05.pdf*

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28, 5, 1–26.

Kuhn, M. and Johnson K. (2013). *Applied Predictive Modeling.* Springer.

Lewiński, S. (2006). Rozpoznanie form pokrycia i użytkowania ziemi na zdjęciu satelitarnym Landsat ETM+ metodą klasyfikacji obiektowej. *Roczniki Geomatyki*, 4, 3, 139–150.

Li, L., Lu, D. and Kuang, W. (2016). Examining urban impervious surface distribution and its dynamic change in Hangzhou metropolis. *Remote Sensing*, 8, 3, 265. DOI: 10.3390/rs8030265

Li, S., Harner, E.J. and Adjeroh, D.A. (2011): Random KNN feature selection – a fast and stable alternative to Random Forests. *BMC Bioinformatics*, 12:450.

Liu, W. and Wu E.Y. (2005): Comparison of non-linear mixture models: sub-pixel classification. *Remote Sensing of Environment*, 94, 145–154.

Lu, D., Li, G., Kuang, W. and Moran, E. (2014). Methods to extract impervious surface areas from satellite images. *International Journal of Digital Earth*, 7, 2, 93–112. DOI: 10.1080/17538947.2013.866173

Lu, D., Li, G., Kuang, W. and Moran, E. (2014). Current situation and needs of change detection techiques. *International journal of Image and Data Fusion*, 5, 1, 13–38. DOI: 10.1080/19479832.2013.868372

Mendes-Moreira, J., Soares, C., Jorge, A.M. and de Sousa, J.F. (2012). Ensamble approaches for regression: A survey. *ACM Comp. Surv,* 45, 1, Article 10.

Mohapatra, R.P. and Wu, C. (2010). High resolution impervious surface estimation: An integration of Ikonos and Landsat-7 ETM+ imagery. *Photogrammetric Engineering and Remote Sensing*, 76, 12, 1329–1341.

Morgan, M.G. and Henrion, M. (1990). *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis.* Cambridge University Press.

Mountrakis, G., Watts R., Luo, L. and Wang, J. (2009). Developing collaborative classifiers using an expert-based model. *Photogrammetric Engineering and Remote Sensing*, 75, 7, 831–843.

Nawar, S., Buddenbaum, H., Hill, J. and Kozak, J. (2014). Modeling and Mapping of Soil Salinity with Reflectance Spectroscopy and Landsat Data Using Two Quantitative Methods (PLSR and MARS). *Remote Sens.*, 6, 10813–10834. DOI: 10.3390/rs61110813

Nawar, S., Buddenbaum, H. and Hill, J. (2015). Digital Mapping of Soil Properties Using Multivariate Statistical Analysis and ASTER Data in an Arid Region. *Remote Sens.*, 7, 1181–1205. DOI: 10.3390/rs70201181

Nielsen, N.H., Joergensen, A. and Larsen, A. (2011). Use of spectral analysis in urban drainage modelling. In: International Conference on Urban Drainage, 11–16 September 2011, Porto Alegre, Brazil. Available online at: https://web.sbe.hw.ac.uk/staffprofiles/bdgsa/temp/12th%20ICUD/PDF/PAP005251.pdf

Partalas, I., Tsoumakas G., Hatzikos E.V. and Vlahavas I. (2008). Greeedy regression ensemble selection: Theory and application to water quality prediction. *Information Sciences*, 178, 3867–3879.

Perrone, M.P. and Cooper L.N. (1993). When networks disagree: Ensemble methods for hybrid neural networks. In: R. Mammone (ed.), *Neural Networks for Speech and Image Processing*. Chapman-Hall.

Qi, H. and Huang, M. (2007). Research on SVM ensemble and its application to remote sensing classification. In: Proceedings of the International Conference on Intelligent Systems and Knowledge Engineering (ISKE 2007), DOI: 10.2991/iske.2007.102

Qian, Y., Zhou, W., Yan, J., Li, W. and Han, L. (2015). Comparing Machine Learning Classifiers for Object-Based Land Cover Classification Using Very High Resolution Imagery. *Remote Sensing*, 7, 153–168. DOI: 10.3390/rs70100153

Quinlan, R. (1993). Combining instance–based and model–based learning. In: Proceedings of the Tenth International Conference on Machine Learning, pp. 236–243.

Ridd, M.K. (1995). Exploring a V-I-S (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote sensing: Comparative anatomy for cities. *Int. J. Remote Sens.*, 16, 2165–2185.

Ripley, B. (1996). *Pattern Recognition and Neural Networks.* Cambridge University Press.

Rooney, N., Patterson, D., Anand, S. and Tsymbal, A. (2004). Dynamic integration of regression models. In: Proceedings of the International Workshop on Multiple Classifier Systems. Lecture Notes in Computer Science, vol. 3181. Springer, pp. 164–173.

Shahtahmassebi, A.R., Song, J., Zheng, Q., Blackburn, G.A., Wang, K., Huang, L.Y., Pan, Y., Moore, N., Shahtahmassebi, G., Haghighi, R.S. and Deng, J.S. (2016). Remote sensing of impervious surface growth: A framework for quantifying urban expansion and re-densification mechanisms. *International Journal of Applied Earth Observation and Geoinformation*, 46, 94–112. DOI: 10.1016/j.jag.2015.11.007

Shao, Y. and Lunetta, R.S. (2011). Sub-pixel mapping of tree canopy, impervious surfaces, and cropland in the Laurentian Great Lakes Basin using MODIS time-series data. *IEEE Journal of*

Selected Topics in Applied Earth Observation and Remote Sensing, 4, 2, 336–347. DOI: 10.1109/JSTARS.2010.2062173

Smola, A.J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14, 199–222.

Sun, H., Qie, G., Wang, G., Tan, Y., Li, J., Peng, Y., Ma, Z. and Luo, C. (2015). Increasing the accuracy of mapping urban forest carbon density by combining spatial modeling and spectral unmixing analysis, *Remote Sensing*, 7, 15114–15139. DOI: 10.3390/rs71115114

Tewkesbury, A.P., Comber, A.J., Tate, N.J., Lamb, A. and Fisher, P.F. (2015). A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sensing of Environment*, 160, 1–14. DOI: 10.1016/j.rse.2015.01.006

Tokarczyk, P., Leitao, J.P., Rieckermann, J., Schindler, K. and Blumensaat, F. (2015). High-quality observation of surface imperviousness for urban runoff modelling using UAV imagery. *Hydrol. Earth Syst. Sci.*, 19, 4215–4228.

Tsutsumida, N., Comber, A., Barret, K., Saizen, I. and Rustiadi, E. (2016). Sub-pixel classification of MODIS EVI for annual mappings of impervious surface areas. *Remote Sensing*, 8, 143. DOI: 10.3390/rs8020143

Turner, II B.L. and Meyer, W.B. (1994). Global Land Use and Land Cover Change: An Overview. In: Meyer W.B. and Turner II B.L. (eds.), Changes in Land Use and Land Cover: A Global Perspective. Cambridge University Press, pp. 3–10.

Vapnik, V. (2010). *The Nature of Statistical Learning Theory*. Springer.

Vapnik, V., and Chervonenkis, A. (1971): On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.*, 16, 264–280.

Walton, J.T. (2008). Subpixel urban land cover estimation: comparing Cubist, Random Forest and Support Vector Regression. *Photogrammetric Engineering and Remote Sensing*, 75, 10, 1213–1222.

Weng, Q. (2012). Remote sensing of impervious surface in the urban areas: Requirements, methods and trends. *Remote Sensing of Environment*, 117, 34–49.

Wężyk, P., Hawryło P., Szostak, M., Pierzchalski, M. and de Kok, R. (2016). Using geobia and data fusion approach for land use and land cover mapping. Quaestiones Geographicae, 35, 1, 93–104.

Wichard, J., Merkwirth,C. and Ogorzałek, M. (2003). Building ensembles with heterogeneous models. In: Course of the International School on Neural Nets, IIASS, 22–28 Sep. 2002, Salerno, Italy. Available online at: http://www.j-wichard.de/publications/salerno_lncs_2003.pdf

Xi, C., Jiancheng L., Zhanfeng, S., Changming, Z., Xin, Z. and Liegang, X. (2011). Estimation of impervious surface based on integrated analysis of classification and regression by using SVM. In: Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International, pp. 2809–2812 .

Xu, H. (2006). Modification of normalized difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, 27, 14, 3025–2033.

Yang, L., Xian, G., Klaver, J. M. and Deal, B. (2003). Urban land-cover change detection through sub-pixel imperviousness mapping using remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, 69, 9, 1003–1010.

Zhang, J. (2010). Multi-source remote sensing data fusion: status and trends. *International Journal of Image and Data Fusion*, 1, 1, 5–24.