Final report of

Geoscientific analysis of urban and rural environment issues with remote sensing techniques

TNN 123457 project

In this project we aimed to explore the possibilities of application of geoinformatics and remote sensing data in urban and periurban/rural environments. We had the following research questions:

- (1) How efficient is the application of remote sensing data in mapping of urban objects: trees, buildings, and roofing materials?
- (2) How successful can be the landscape change modelling?
- (3) Can we determine the dust-adsorption capacity of urban trees?
- (4) How efficient is the crowd-sourced noise mapping using cellphones?
- (5) Can spectral data reflect soil properties?

1. Remote sensing data and urban applications

1.1. Asbestos roofings

We established the roof material cadaster of 2500 houses. Using these data as a reference, we performed several analyses with a WorldView-2 image, and an orthophotograph of an aerial survey with RGB+NIR bands. WV-2 image was pan-sharpened (from 2 m to 0.5 m). Red tiles, dark tiles and asbestos roofing materials were aimed to identify. Asbestos represents serious health risk, thus, an inventory of possible sources can help local authorities in future project planning.

Using the WV-2 image, we delineated a study of the asbestos identification and compared three types of machine learning algorithms (Linear Discriminant Analysis, Quadratic Discriminant Analysis and Random Forest). We also evaluated the effects of spatial resolution, the aspect of roof planes (sunny or shadowed) and the resolution. We found that:

- discrimination of sunny and shadowed roof planes did not improve the classification accuracy;
- pan-sharpening improved the classification accuracy with 2-3%;
- Random Forest was the most efficient algorithm which helped to gain the best outcome, but Quadratic Discriminant Analysis was also efficient, especially with fewer roof types (i.e. when planes were nit divided into sunny and shadowed planes);
- thus, the pan-sharpened images were the best input with fewer classes using the Random Forest classifier, overall accuracy reached more than 90%.

Results were published in the Hungarian Geographical Bulletin journal.

We also analyzed the possibilities of hyperspectral imagery and conducted an investigation with an AISA Eagle image. We tested three algorithms (Support Vector Machine, Spectral Angular Mapper [SAM] and Maximum Likelihood) in two study areas with PCA principal components (raw data is not useful due to high level of multicollinearity, and algorithms have increased demand on training data with large number of predictors, i.e. bands). We found that:

- SAM algorithm was the least efficient, provided 10-20% worse (51-71% overall accuracy) related to the best performing Maximum Likelihood (75-81%);
- when number of asbestos roofs was low, the accuracy was at least 10% worse.

Results are published in a conference proceedings (Térinformatikai Konferencia és Szakkiállítás, Geophysical Research Abstracts)

Our applied methodology involved the machine learning with repeated k-fold cross-validation, usually with the algorithm of Random Forest. Although this procedure is more reliable than the simple train-test method, we also have to take care of the independency of the data. However, using pixel-based methods, remote sensing data is not spatially independent when not only a single pixel is used to train the model due to spatial autocorrelation (i.e. adjacent pixels are similar). We realized this issue, and developed a method in Python programming environment to overcome. Several tests and approaches had been tested with the WV-2 image, the aerial photograph and using an ISPRS benchmark dataset, and we found that:

- a traditional training was the reference with a spatially independent testing dataset;
- simple resampling of the training dataset and running 50 repetition with randomly selected data had high level of autocorrelation as well as with the k-fold cross-validation;
- segmentation, as a possible solution is not appropriate in this case, because the roofs have limited extent, and a roof is more or less can be regarded as a segment, segments inside a roof also can have spatial autocorrelation;
- using a one-pixel selection (e.g. centroid pixel) method reduces the element number to low, which has influence on the final results (decreases it);
- keeping a buffer distance among pixels (sampling from the training data) was tested with 2 m and 10 m and while 2 m did not diminished the autocorrelation, 10 m was optimal;
- spatial resolution had strong effect on the level of autocorrelation.

Findings were summarized in a manuscript which is under review at ISPRS Journal of Photogrammetry and Remote Sensing.

1.2. Mapping urban green infrastructure

We performed city-scale classification of green areas focusing on trees, and we also managed to classify tree species with multispectral and hyperspectral data. We built two cadaster of reference data: 1500 trees in the northern part of Debrecen, and further 900 trees in the city. We had the following results:

- Hyperspectral data (368 bands, 400-2500 nm range) provided accurate data to discriminate tree species for 7 species with 1500 reference trees. *Results are published in a conference proceedings.*
- WV-2 image was used to identify 6 common species of a city parks and alleys with 700 reference trees. We applied minimum noise fraction (MNF) for dimension reduction, and used two classifiers, the Support Vector Machine and the Maximum Likelihood. We found that four MNF components ensured the highest overall accuracies (71-75%), and with the confusion matrix we identified species having similar spectral

characteristics (lower class level accuracies, such as Silver maple and Black locust). *Results are published in the European Journal of Geography.*

- Finally, we extended the analysis to Debrecen using the WV-2 image with four common species of city parks and alleys were in the focus (see at dust adsorption capacity, section 3.4.). Maximum Likelihood, Support Vector Machine and Random Forest were used as classifiers, and the Maximum Likelihood had the best performance (81%, i.e. 6% better than the others). We found that, in this case, the Norway maple had the lowest class level accuracy metrics. *The results are part of the PhD dissertation of Molnár Vanda Éva and will be published in an international journal.*
- WV-2 and WV-3 images were compared in efficacy of species-level idientification of trees. 700 tree units served as reference data. We found that WV2 was more appropriate for classifications related to the classification accuracy than the WV-3. The reason was the capture date (WV-3 was from September), and sun angle (off-nadir), and the fact that the availability of WV images is very limited. Results showed that the most common trees (6 species) of an urban environment, on the example of Debrecen, can be ~75%. MNF transformation (i.e. a special type of PCA) improved the accuracy by 4%. *Results were published in a conference proceedings.*

We determined the built-in area and green area ratios per resident plots for Hajdúböszörmény, Nagyhegyes and Nádudvar. In this analysis, we used aerial imagery of 2011 (freely available MADOP images). We segmented the areas into spectrally homogenous small parts, calculated the means for all segments by bands and performed Random Forest classification. We then overlapped the map with the map of residential plots and calculated the ratio of built-in and green areas. We provided the maps in html format to the local authorities, and they use it in the urban planning. We always seek the connection with the possible users of our work and this outcome was found to be valuable for the practice. Next step will be the publication in an international journal.

We conducted two UAV surveys in Debrecen, Úrrétje and Csapó-kert districts. We produced orthophotos and DSMs and analyzed the potential applications of the drone based photogrammetry focusing to extract the buildings and the vegetation. Original bands, spectral indices and texture indices were determined and an oversegmented vector layer was generated to reduce the salt and pepper effect of the pixel-based classifications. Using the segments, we identified 300 polygons per land cover classes to build a database for reference data. Two approaches were applied: (i) discrimination of urban green areas from the 'other' classes, (ii) discrimination of herbaceous vegetation and arboreal, and 'other' land covers. We used 4 classification algorithms, neural network (NN), Partial Least Squares (PLS), Support Vector Machine (SVM), Multiple Adaptive Regression Splines (MARS), and Random Forest (RF) with 10-fold cross-validation and hyperparameter tuning. Results showed that the simple RGB images were eligible to discriminate the green areas from all other land cover types, the accuracy was even 95%. However, the discrimination of herbaceous and arboreal vegetation resulted an ambiguous outcome: overall accuracies indicated high accuracy, but the visual validation showed serious omission errors. It means that green areas can be identified with an acceptable rate, but sub-classes cannot be distinguished. Results are under publication.

1.3. Ecological evaluation

City level green infrastructure classified from the WV-2 image, and also the high resolution Copernicus layers were involved in the analysis of the landscape connectivity. We applied the Graphab and Conefor Sensinode software with a graph-based approach both with binary and probabilistic indices. We delineated the key habitat patches and possible corridors according to the following options:

- trees/urban forest patches;
- grasslands/urban grassy areas;
- urban areas were mapped with higher resolution WV-2 image;
- surroundings (rural/agricultural) areas were mapped with the Copernicus layers.

Habitat patches were evaluated with Carabid species, with daily movement parameter of 14 m 100% abundance. Species' presence outside the green areas decreased with the increasing distances between the patch pairs.

2. Landscape change modelling

2.1. Land change modelling and evaluation

One of our main task was to evaluate the possibilities of landscape change modelling. We did not intend to provide future scenarios but to find the limits of the modelling. The main issue with these models is that hundreds of published papers in the topic of future projections, stated with high accuracy, but the validation procedure is not appropriate. Most authors collect reference data (mostly) from the non-changing areas and only a few are from the changing parts, e.g. 5 out of 100 reference data, and they find 95% accuracy in the prediction and it can be possible that none of predicts on the changes had been correct. We performed an Intensity Analysis, and determined the Figure of Merit to evaluate the performance of a Markov chain cellular automaton simulation model. Our approach introduced a novel way to interpret the land change models. We found that even that the best prediction had 4% accuracy. We concluded with recommendations that apply generally.

- Scientists must compare visually and quantitatively the changes during three intervals: (i) reference change during the calibration interval, (ii) simulation change during the validation interval, and (iii) reference change during the validation interval. Comparison between (i) and (ii) relates the calibration patterns to the subsequent simulation.
- Comparison between (ii) and (iii) distinguishes between simulation and reference changes during the validation interval. Comparison between (i) and (iii) shows the degree to which the reference patterns are stationary through time. For each comparison, Intensity Analysis reveals various levels of information concerning quantity disagreement.
- The Figure of Merit's components distinguish quantity disagreement from allocation disagreement during the validation interval. Our recommended collection of metrics generate insights that are deeper than any single metric can communicate.

Results are published in the Ecological Indicators journal.

2.2. Land change modelling and the number/aggregation of land cover classes

Accuracy of land change modelling is also a function of the number of land cover/land use classes. Especially, when a class has only a small area related to the total area, predictions can be misleading; therefore, aggregation is suggested. However, merging of land cover classes also bias the accuracy of the models. Accordingly, we conducted a study on the effects of aggregation and examined the Corine Land Cover (CLC) maps using the Level 3 classes (without aggregation, 20 classes in the area), Level 2 data (10 classes), Level 1 (5 classes), behavioral-based aggregation (BB), and threshold-based aggregation using three types of zoom level (i.e. scale). We analyzed 8 study areas with different intensity of changes, and applied the Markov chain model for the predictions. Altogether 114 CA-Markov simulation model had been run. We analyzed the effects of aggregation methods on changes in the study areas, the Figure of Merit (FOM) and FOM's components: Wrong Hits, False Alarms, Misses, Hits. We had five main conclusions.

- L1 and BB aggregations produced the fewest categories.
- BB aggregation maintained the largest sizes of changes.
- L1 had generally lower sizes of changes in the calibration, validation and simulation intervals.
- L1 medians of change were considerably lower, specifically, half of the medians were zero.
- L1 had generally lower values interms of all FOM components.

Based on the results and the aggregation rules of various aggregation methods, we warn users that the Corine standard level aggregation rules can eliminate sizeable changes. We recommended users apply aggregation methods that reduce the number of categories while maintaining changes and not reducing the correctly simulated changes in the area. In our analysis, the behavior-based aggregation method met these goals. We also recommended that users calculate FOM and FOM's components to gain important insights concerning the interaction of the simulation model performance and changes in the reference data.

Results had been published in the Remote Sensing journal.

2.3. Corine Land Cover as reference data

We also aimed to analyze the Corine Land Cover maps from the aspect of using them in validation process, i.e. when satellite images are classified can we use the CLC as reference data. Accordingly, we analyzed the images of Landsat-8, the Sentinel-2 and the PlanetScope satellites captured in the same dates (within 3 days). Besides, CLC is ideal input data in simulation studies, but the question is how the polygons reflect the land cover considering its limitations (25 ha minimal mapping unit). We aimed to reveal if the satellite data can reflect Level 1 CLC classes (5 classes) using Linear Discriminant Analysis and Random Forest. We applied a segment (CLC unit) based analysis aggregating the surface reflectance as mean, median, standard deviation, minimum and maximum values, which were the input data in classifications by satellite bands. We found the followings:

- Medians of CLC polygons provided the least mixture among the LC classes, while the maximums were the worst input parameters without significant differences. Wetlands and water bodies categories were the most frequently mixing categories of CLC based on reflectance values.
- Bivariate statistical tests cannot provide enough information to conclude on the spectral separability of LC classes, but classification algorithms involving several variables can be efficient techniques. Generally, LDA and RF classifiers had similar OAs, but in the

case of coarser resolutions (Sentinel and Landsat), RF outperformed the LDA. Data derived from PlanetScope provided 7% better OAs (78%) than those of Landsat (71%) regarding the model medians; thus, better spatial resolution ensured better classification performance. >80% OA was gained with using all available bands of the Sentinel-2; accordingly, more spectral information in the infra-red range can counterbalance the coarser geometric resolution.

- We developed a randomization-based technique to gain 10 repetitions of class-level metrics (UA and PA), which showed that satellites had no direct effect on the accuracy. UAs were the lowest in agricultural areas, while PAs were the lowest among wetlands.
- Variable importance of statistical parameters showed that usually the medians were the most important statistical layers, and the green, red and near-infrared bands were the first three most important bands.
- We provided an approach to prove the possibility of the generalization of the results with multiple randomized subsampling and found that the results of Landsat and Sentinel data can be generalized, but in the case of PlanetScope, a larger area with more CLC polygons would be desirable.
- Generally, using the overlapping bands (RGB + NIR) of Landsat-8, Sentinel-2 with the PlanetScope, the best OAs were >70% OAs, but the most accurate was the PlanetScope with the highest spatial resolution (78.5%). Higher OAs (~80%) have also been acquired with the higher spectral accuracy of the Sentinel-2, which means a cost-efficient solution in spite of the coarser spatial resolution.
- As we found several studies where CLC maps were used as ground truth data to quantify thematic accuracy, 70–80% OAs do not seem satisfactory. Nevertheless, our experiment was performed with the CLC L1 classes; further investigations can reveal if CLC is more appropriate for ground truth with the more detailed L2 or L3 nomenclature.
- CLC maps have many features limiting their use as validation data. The 25 ha MMU size creates several issues, which influence validation by satellite images: land cover polygons do not follow the exact borders of patches, smaller objects are merged into larger ones to reach the 25 ha area, and even the supposedly homogenous patches also have a nonhomogeneous texture (e.g., lakes, rivers).

Results are published in the Remote Sensing journal.

3. Dust adsorption by trees and the extrapolation possibilities of remote sensing

The workflow was designed to (i) test a general index to indicate dust loadings, (ii) test tree species' leaves as dust adsorption traps, (iii) analyze the element concentrations, (iv) extrapolate the results with remote sensing data and determine the dust adsorption capacity of the green areas.

3.1. APTI as an indicator of air pollution

We aimed to assess the role of eight tree species in urban green area planning from the aspect of pollution tolerance and bioindication of air pollution using the APTI (Air Pollution Tolerance Index). We demonstrated that *R. pseudoacacia, T. × europaea, A. platanoides, F. excelsior, B. pendula, and C. occidentalis* were sensitive indicator species of air pollution. Tolerance was moderate for *A. saccharinum*, while *P. acerifolia* was intermediate, based on the APTI value. There was a significant difference among species based on leaves for the ascorbic acid content, for the pH of the leaf, and for the total chlorophyll content of leaves. The ascorbic acid content was the highest in *B. pendula* leaves, and the pH was the highest on the leaves of *R. pseudoacacia* and *C. occidentalis*. The total chlorophyll content was the highest on the leaves of *A. platanoides*. There were significant differences in the Al, Ba, Ca, Fe, Mg, Ni, S, Sr, and Zn concentrations of leaves among the species. We found that tree leaves are reliable bioindicators of urban air pollution. APTI is useful in selecting pollution-tolerant species and can be used for urban green infrastructure planning in the phase of species selection. Based on the APTI, *A. saccharinum* and *P. acerifolia*, and based on the PM, *A. saccharinum* and B. pendula, are recommended as pollutant-accumulator species, while other studied species, especially those with lower APTI values, are useful bioindicators of air pollution and proxies of urban health. It was a first step in finding a pollution indicator which fits to identify depositing dust on tree leaves.

Results are published in the Plants journal.

We performed a thorough meta-analysis of the APTI in high accordance with the dust sedimentation. We started to measure the depositing dust with plastic dust collectors, and also performed in Debrecen, and, as a part of another campaign, a parallel collection both in Debrecen and Bangkok, with our partner at the Asian Institute of Technology. Beside the results gained from our samples, a meta-analysis were carried out using the findings of previous studies with a careful selection from the Scopus database. Accordingly, involved 9 cities of the world and, as a meta-analysis, we re-analysed the pollution results of previous studies where the APTI was also used. We applied random forest regression and found that population, PM10 and land use type are important influencing factors of dust pollution, and the model had an R^2 of 49%. Main findings were:

- We concluded that the characteristics of anthropogenic activities influence the biochemical composition of plants. Different pollution levels in industrial, roadside and urban areas were reflected in APTI values of regionally typical plant species, as plant tolerance levels adjusted to a certain environment.
- We observed that under extreme conditions, the tolerance of plants drops drastically rather than increasing along with pollution levels. APTI values also correlated with the population of cities, which is related to the size of inhabited areas and the emission loads from households, vehicles and industries.
- Our model showed the separation of cities by APTI; heavily polluted areas had more tolerant plant species with a relatively high APTI than less polluted cities. Through statistical model building, APTI can be predicted using the population of cities, the annual mean PM10 and land use type. However, as PM10 did not have a significant relationship with the APTI, it is recommended to assess the effects of other pollutants to determine which ones influence the value of APTI.

Results are published in the Ecological Indicators journal.

3.2. Dust in tree leaves and sediment traps

When we aimed to monitor the dust adsorption of the tree leaves, we also wanted to know if the amounts can be compared to the traditional measurements with dust traps (i.e. patents). Accordingly, we measured the sedimented dust in traps and collected by the tree leaves: Common lime (*Tilia europaea*) was examined at two heights (1.8 and 3.6 m) in three months.

After several trying we chose filter papers with vacuum filter machine to separate the dust fractions, so finally we determined the PM5 and PM10 fractions as the available filters made it possible. Our findings demonstrated that dust trapping by leaves is an effective and eco-friendly way to alleviate dust and particulate material pollution in urban areas. Leaves are a useful indicator of PM deposition for biomonitoring studies and settlement plans of cities. We justified that there was no difference in the concentration of PM between dust traps and leaves, and there are differences in concentration of PM based on studied months. There were no differences based on the sampling heights and there was a correlation between the results of used gravimetric and laser diffraction methods based on the concentration of PM as our predicted hypothesis. The correlation between leaves and dust traps based on the concentration of PM5 indicated that the efficiency of two collection methods (leaf and dust trapping) was similar in the case of fine particle material. These findings also confirmed that trees play an important role in the mitigation of air pollution in urban habitats, and that leaves are not perfect dust traps, but the analysis can be performed without previous installation of traps. Leaves can be good indicators of dust deposition in spite of all the issues.

Results are published in the Atmosphere journal.

3.3. Metal Concentration of tree leaves in Debrecen and Bangkok

We performed soil sample and tree leaves collection in an area near to Bangkok with the help of colleagues of Asian Institute of Technology (Bangkok, Thailand). Soil and tree leave samples (*Ficus religiosa* and *Mimusops elengi*) had been collected in a gradient of industrial-urban-rural areas. Metal content was determined by ICP-OES and we performed a statistical analysis.

- We found that the distribution of heavy metals in the environment is relevantly affected by biotic and abiotic factors. We revealed that *F. religiosa* accumulated significantly higher metal content than *M. elengi* in the leaves in the cases of Ba, Ca, Cu, K, Mg, and Zn, while *M. elengi* had higher metal content in the cases of Al, Cr, Fe, Mn, Na, Ni, and Pb.
- According to the multivariate analysis, the species had a significantly different accumulation pattern. The location, the industrial–urban–rural gradient, had not significant difference regarding the metal uptake, i.e., the intensity of land-use had no discriminating effect which can be in connection with the polluted Chao Phraya River near the rural sampling site.
- Bioaccumulation (BAF) was significantly larger in *F. religiosa* especially in the industrial and rural sites. We revealed that the multiple linear regression was not as efficient in finding the relationship between metals SOM and the humus quality as the MARS regression which has no assumptions regarding linearity, distribution, and outliers. The applied k-fold cross-validation provided more reliable results related to a correlation because it uses a training and an independent testing dataset, and the repetitions provided 50 R^2-values which showed the range of possible realizations instead of a single value.
- We proved the connection between bioaccumulation and humus in case of Al, Ba, and Cu. Our findings demonstrated that trees are a useful indicator of air pollutants for biomonitoring studies of cities.

Results are published in Environmental Science and Pollution Research journal.

3.4. Tree leaves, dust adsorption and remote sensing

We utilized the experiences of the previous researches in this extrapolation step. We determined the most common species of city parks and alleys in Debrecen (*Acer platanoides, Tilia europaea, Platanus acerifolia, Celtis occidentalis*) and then collected leave samples to quantify the adsorbed dust by species. Next, we performed image classification with Maximum Likelihood, Support Vector Machine and Random Forest classifiers. After identifying the species, we determined the Leaf Area Index (LAI) and then we quantified the dust volumes. Our findings are the following:

- Tree species were identified with the accuracy of >75% regarding the class level metrics except the *Acer platanoides* (Norway maple), which was 68%.
- Areas of tree species had been quantified, and we also elaborated the LAI map of the green infrastructure.
- LAI distribution (regarding the data by pixels) reflected the potential surface of the dust adsorption of leaves. Platanus and maple species had the largest capacities whilst the Tilia had lowest values. Quantification of the sedimented dust showed the following data for dust adsorbed by the given species: *Celtis occidentalis* (9.94 kg/ha), *Tilia europaea* (4.00 kg/ha), *Platanus acerifolia* (10.65 kg/ha) and *Acer pltanoides* (7.44 kg/ha).
- Adsorbed dust by leaf surfaces can be estimated between 4196-4247 kg projected to the green infrastructure of parks and alleys of Debrecen (excluding the Great Forest which requires further measurements).

Results currently are the parts of PhD dissertation of Molnár Vanda, can be shared on demand.

4. Crowd-sourced noise-mapping with cell phones

We performed field surveys in order to map urban noise pollution using smartphones. We analyzed 14 smart phones and an official noise meter, 3 software, and two sapling methods (many repetitions at few locations vs. many measurements without repetitions). We came to the conclusion that smartphones can help in producing a noise map, but a careful selection is needed with data validation and that many (i.e. 20) repetitions do not help in getting reliable information about the accuracy, instead, more in situ measurements is advisable. We found that:

- R^2 was >0.9, and the differences between data pairs usually stayed in the range of 2±SD. Consequently, smartphones can effectively help noise mapping, but a careful smartphone selection, with comparing the noise measurements with a calibrated device, is needed to verify the accuracy.
- Our suggestion is the Noise Meter for the software, but smartphones develop quickly; our selection, the Sony Xperia Z3 is outdated now; thus, our work serves as a methodological guide;
- As a first step is a selection of a proper cell phone, a validation with a calibrated professional device, and all measurements should be performed with the selected phone, we came to conclusion that crowd-sourced noise mapping is not a good option, but smart phones can mean good help in the mapping with suggested methodology.

Results are published in the Environment and Planning B journal.

5. Soil analysis and geoinformatical analyses

We collected 50 soil samples from Debrecen, and 104 from Tokaj. In Debrecen, the primary aim was to survey the garden soils, while in Tokaj, we aimed to find correlation between the spectral profile of the soils and their heavy metal content. We found that spectral profiles can help in a quick estimation of the organic matter (out of 12 soil properties), the pseudo-R^2 between the predicted and measured values was 0.51. This result was gained by feature selection method of the HypDA modul developed by us for the MS Excel.

Regarding the peri-urban agricultural area/garden transformation analysis, we analyzed the peri-urban regions of Debrecen and four neighboring smaller settlements using the CLC Land Cover 1990, 2000, 2006 and 2012. Totally 296 ha were delineated, which are mostly new residential areas with small gardens, in one case a recreational area was constructed. These land use patches were dispersed at the edges of settlements, but we also found clustered patches at the quickly developing areas of the cities. Large changes were observed between 1990-2000 and 2000-2006 periods. The identification of these areas allows a 'space for time substituion' designed study of soil transformations in the future. *Results are published in a conference proceedings.*

A main project was the UAV-based monitoring/mapping of the soil water content. This study was performed in a small area close to Debrecen, at parcel level with a UAV, multispectral and thermal camera. We compared the efficacy of the estimations of the multispectral and thermal cameras using 4 machine learning algorithms and 2 approaches regarding the input data: (i) involving the soil water content of the single survey dates, and (ii) merging all data into one database. We found that multispectral cameras ensured better input data than thermal cameras: R^2s were 0.97 vs. 0.71, normalized root mean square errors were 10 vs. 25%, respectively. Best models were gained by the Random Forest regression, and in case of multispectral data elastic net regression was also efficient with 0.88 R^2. Relationship between soil water content and thermal data was exponential, which was incorrectly handled by the general linear model (robust linear model and elastic net regression was not working with only one variable), thus, thermal data was acceptable only with the Random Forest. Single pixel extraction provided the best input for the estimations, mean of buffered areas did not perform better in the models. Maps provided soil water content estimations in details, with high spatial resolution, which can be a great help to farmers to increase the efficacy of planning irrigation in precision agriculture.

Results are submitted and is under review at Computers and Electronics in Agriculture Journal.