# SMALL WATER BODIES IDENTIFICATION BY MEANS OF REMOTE SENSING


**Jitka Komarkova, Pavel Sedlak, Radek Pesek, Ivana Cermakova**

Affiliations:
Assoc. Prof. Jitka Komarkova, Ph.D., MSc. Pavel Sedlak, Ph.D., MSc. Radek Pesek, MSc. Ivana Cermakova;
Faculty of Economics and Administration, University of Pardubice;
Studentska 95, 532 10 Pardubice, the Czech Republic;
Ph.: +420 466036070, fax: +420 466036010, e-mail: jitka.komarkova@upce.cz;



***Abstract***
*Remotely sensed data are frequently used to identify water bodies. In comparison with UAV data, they are limited by resolution and availability. Paper evaluates suitability of Landsat 8, Sentinel 2 and UAV data in the case of identification of shorelines of smaller water bodies. For the study, surrounding of Pardubice city (the Czech Republic) is an area of interest. Three data sources are used: Landsat 8, Sentinel 2 and UAV data. Several algorithms are used for spectral enhancement and data classification: Iso Cluster, Maximum Likelihood, Class Probability, Principal Components, and NDWI. Manual classification is used as a reference method. No post-classification method is used to preserve shapes of small water bodies. Error matrix is used for evaluation of the classification quality. Multi-criteria evaluation shows that Sentinel 2 data classified by means of Iso Cluster provides the best results. NDWI is very close to the best results. Next, we demonstrate that UAV can provide data with a higher spatial resolution on demand for reasonable costs so they are more suitable for small water bodies. Heterogeneity of the data and treetops overlapping the shoreline led to the manual classification based on the results of Iso Cluster classification.*


***Keywords:*** *Small water bodies, Landsat 8, Sentinel 2, UAV, Imagery classification*


## INTRODUCTION

Water is becoming more and more important issue today as far as its sources are more and more exhausted. Urban growth and insufficient management approaches can lead in urban population being in water stress (McDonald, R.I., et al., 2014). Agriculture is another example of water importance and necessity. According to Pimentel et al. (2004), agriculture (both planting and livestock production) consumes about 70 % of fresh water resources. Small water bodies represent an inevitable part of the hydrosphere. They are important for biodiversity and ecosystems functions and for their influence on larger water bodies (Biggs et al., 2017). Various definitions of small water bodies exist. Discussion can be found in (Biggs et al., 2017).

Importance of identification of water bodies and shorelines by remote sensing increases along with increasing water stress. Remotely sensed data has been used for water quality monitoring for a long time – from the 70$^{ties}$ along with Landsat existence (Bukata et al., 1974). Landsat data was successfully used for shoreline and water bodies identification many times, e.g. Bhagat et Sonawane (2011), Jiang et al. (2014), Jones et al. (2017). Verpoorter et al. (2012) proposed an automated extraction of water bodies by means of combination of various methods for Landsat data processing. Sentinel data are suitable too, it was used e.g. for a study of the Nile River (Salama et al., 2012), or identification of urban surface water bodies (Yang et al., 2017). Temporal and spatial resolution of satellite imagery represent their limitation for identification of small water bodies (Verpoorter et al., 2012).

Various water spectral indices have been proposed to identify water and water bodies, e.g. the Normalized Difference Water Index (NDWI) was proposed by McFeeters (1996). Automated Water Extraction Index (AWEI) was proposed by Feyisa et al. (2014) to enhance spectral contrast and reduce an influence of deep shadow caused by the terrain as a source of classification error. Enhanced Water Index (EWI) based on Landsat data was proposed by Wang et al. (2015). Both Landsat and Sentinel data can be used to calculate various spectral indices (Yang, 2017; Zhou et al., 2017). Sentinel 2A MSI imagery was used to identify water bodies on basis of Normalised Difference Water Index (NDWI) (Yang et al., 2017). Fisher et al (2016) tested 7 indices and Landsat TM, ETM+, and OLI data. They showed that all three sensors provided similar accuracy with data for eastern Australia. Comparison of Landsat 7 ETM+, Landsat 8 OLI, and Sentinel 2 MSI sensors done by Zhou (2017) revealed that all sensors and included indices provided reasonable results but Landsat 8 and Sentinel 2, as well as NDWI, provided more accurate results.

Utilization of satellite data is limited by their time and spatial resolution, very high-resolution images are becoming more required. UAV can provide very low elevation very high-resolution data with better temporal resolution, on demand (flexible), of course with respect to limitations given by law regulations and weather conditions (Lejot et al., 2007; Pásler et al., 2015).

Identification of small water inland bodies are pointed out as a specific issue by several authors (Huda et al., 2010; Yang et al., 2010; Pásler et al., 2015; Jones, 2017). Aim of the paper is to evaluate suitability of Landsat 8, Sentinel 2 and UAV data in the case of identification of shorelines of smaller water bodies near the city of Pardubice by means of several classification methods and NDWI.

## AREA OF INTEREST

Area located north-west of Pardubice (located in the East Bohemia in the Czech Republic,) is used as an area of interest. Its approximate size is 12,6 km (north-south direction) to 16 km (east-west direction), i.e. approx. 201,6 km$^2$ (see Fig. 1). The area is flat, lying at about 220 – 230 m.a.s.l. There are no remarkable hills located. Area is partly wooded but it is mostly used for agriculture purposes. It contains several ponds, mostly as remains of the Vilem of Pernstejn pond network built in at the turn of the 15 and 16$^{th}$ centuries. Several other ponds were created by sand mining (Lazne Bohdanec, 2015). Bohdanecsky pond is the largest one, its water area is 900 000 m$^2$. On the opposite, the smallest water bodies have area about 150 – 200 m$^2$. Some of the small water bodies are of artificial origin. River Elbe is another important water body located in the area of interest. So, area of interest is heterogeneous, it contains various land cover/land use types, mostly fields, trees, greenery, urban areas and water bodies.

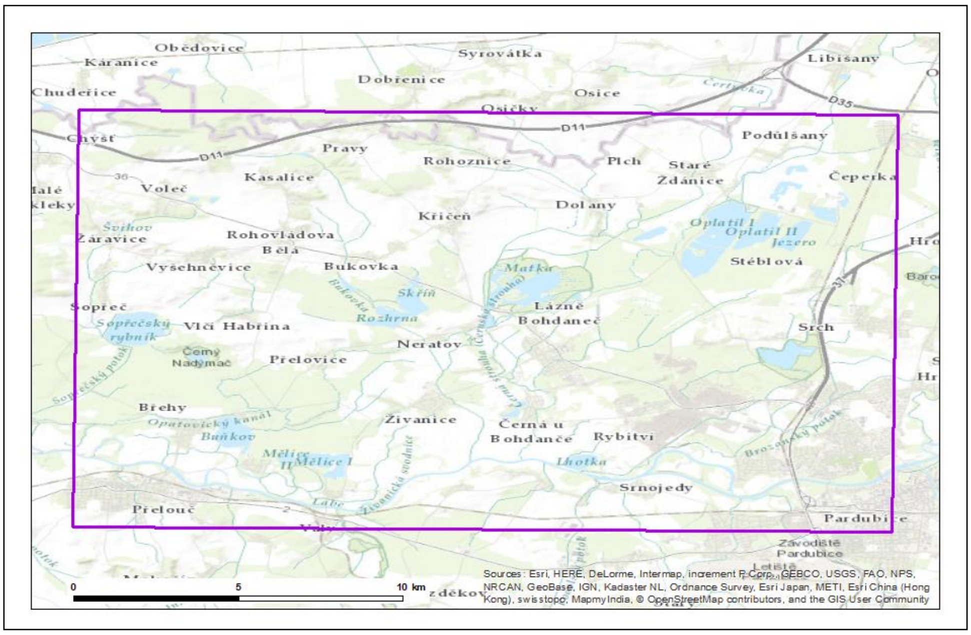


*Figure 1. Area of interest for satellite data (source: authors, data: ESRI)*

## DATA, SOFTWARE AND METHODS

Landsat 8 OLI sensor, Sentinel 2, and own UAV data are used. Both Landsat and Sentinel data was taken on April 1, 2017. Landsat 8, Sentinel 2, and UAV data are used with WGS 1984 UTM Zone 33N coordinate system. Source of Landsat and Sentinel data is USGS EarthExplorer (https://earthexplorer.usgs.gov/). DJI Phantom 3 Pro equipped with a DJI camera was used so only visible wavelengths data were available. Data was taken on April 2. Area of interest for satellite data is defined by polygon shapefile (see Figure 1), which is used to clip all other data.

Both Landsat and Sentinel data was taken during not cloudy weather. Cloud cover of Landsat was 0,01 %, cloud cover of Sentinel was 0 % so no atmospheric correction was necessary. Geometric and radiometric corrections are already done by data providers. All data files and bands are clipped using borders of the area of interest.

Suitable bands are used for analyses. Only 8 bands of Landsat data are used. A new infrared band (No 9) which is used for cirrus detection is not suitable for this work. Thermal bands 10, and 11 are excluded because of their low spatial resolution (100 m). Bands 1, 9, and 10 are excluded from Sentinel data. These bands are used for detection of aerosol, steam, and cirrus clouds. Next reason for exclusion of the bands is their low spatial resolution (60 m) in comparison with other bands.

ArcGIS for Desktop 10.5 is used for data processing. The following methods and tools are used to analyse data (see Lu and Weng, 2007 for methods description):

- Iso Cluster tool – unsupervised classification based on ISODATA algorithm as implemented in ArcGIS
- Maximum Likelihood – supervised, parametric per-pixel classification (Bayes)
- Class Probability – semi-supervised, per-pixel algorithm, based Bayesian statistics, estimates class probabilities
- Principal Components – a statistical method suitable for image transformation to reduce correlation between bands (spectral enhancement)
- NDWI – Normalized difference water index – as proposed by McFeeters (1996)

No post-classification processing is used in any case because it changes shapes of small water streams and bodies. Shape of river Elbe is the most visible example of shape change. Its shape was different or the river disappeared at all.

Error matrix is created for each classification to evaluate quality of classification. Classified images are transformed into polygon vector file by Raster to Polygon tool. Create Random Points tool is used to randomly place 50 points to each class. Extract Values to Points tool is used to assign values from raster pixels to random points. Frequency tool calculation is used to calculate number of classified points in each class.

Two water bodies (Soprecsky and Pohranovsky ponds) are classified manually from the data to obtain reference data to evaluate quality of classification by means of comparison of water bodies' sizes. Area of Soprecsky pond is 85.6 ha and of Pohranovsky pond is 45,48 ha based on manual classification.

A short part of shoreline of the pond Skrin was chosen as an area of interest in the case of UAV-based remote sensing. Its size is 0,02 $km^2$. It covers heterogeneous land cover types, mainly consisting of water surface, trees, bushes, grass, and seasonally flooded greenery. This area was chosen with respect to accessibility for flying UAV.

## RESULTS AND DISCUSSION

Various methods (listed above) are used to process data into two final classes. More classes are used for the classification in the beginning but later they are integrated into two final output classes: water and all other land cover types.

### Iso Cluster

Classification into 5 and 20 classes is used. The number of classes was chosen on experimental base. All classes are finally integrated by means of reclassification into 2 classes: water surfaces, and all other surfaces.

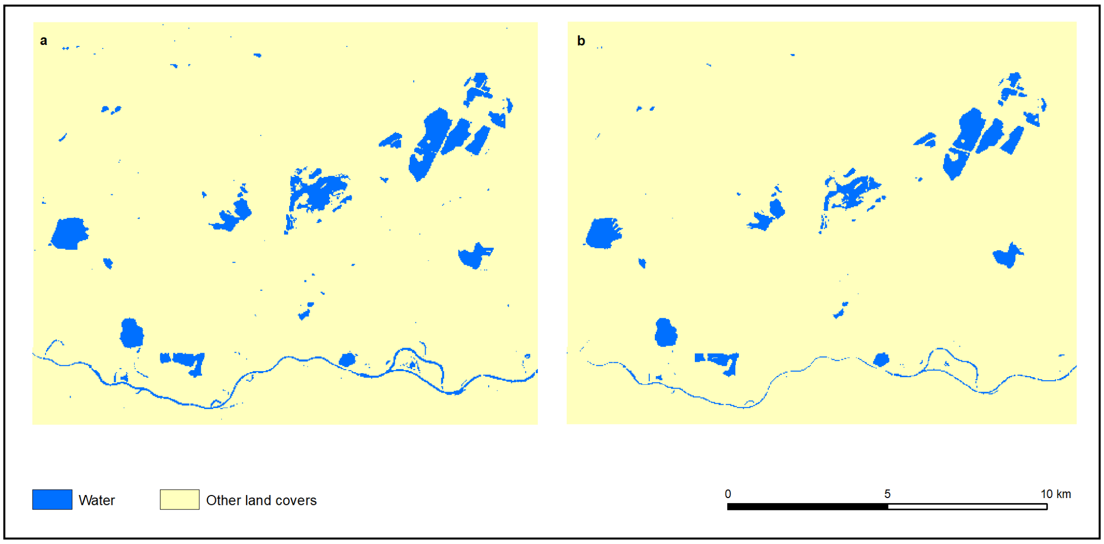


*Figure 2. Landsat 8 data classification by means of Iso Cluster into 5 (a) and 20 (b) classes, (source: authors, data: USGS)*

Classification of Landsat image into 20 classes resulted into slightly worse results (see Figure 2). Error matrix calculation revealed 4 not correctly classified points representing water.

Classification of Sentinel data is one in the same way, into 5 and 20 classes, which are then reclassified (see Figure 3). Error matrix provided the same result – classification in 20 classes was slightly worse, with 4 correctly classified points representing water.

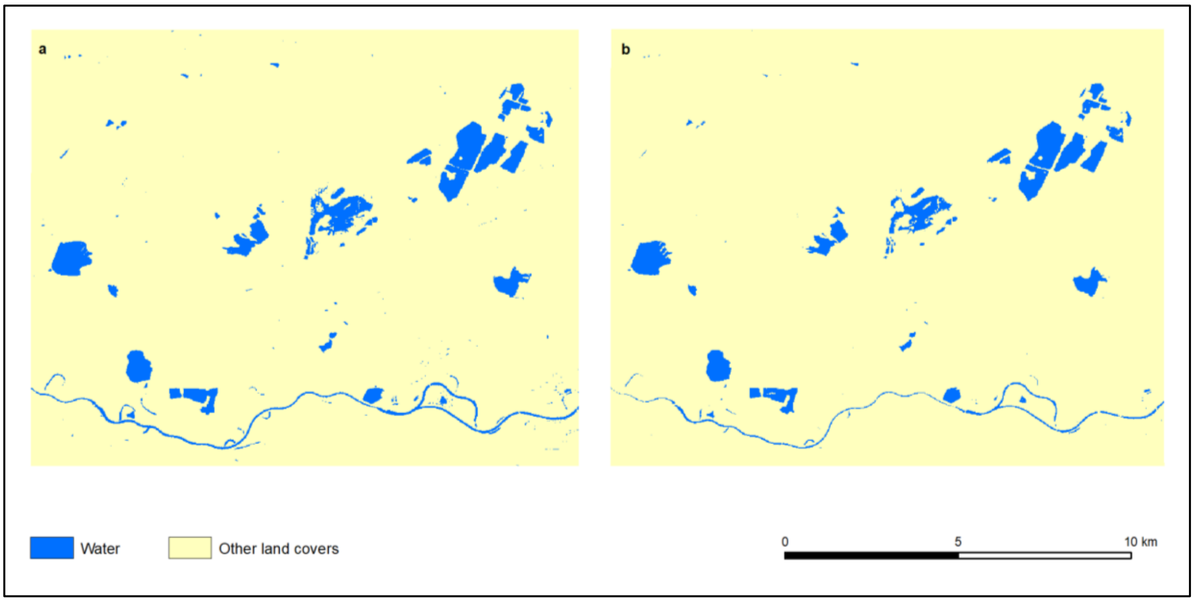


*Figure 3. Sentinel 2 data classification by means of Iso Cluster into 5 (a) and 20 (b) classes, (source: authors, data: USGS)*

### Maximum Likelihood

Four classes are used in the case of maximum likelihood: water, vegetation, urban areas, and bare ground. Training areas are prepared at first. Next, they are used for classification itself. Results are again reclassified into 2 final classes.

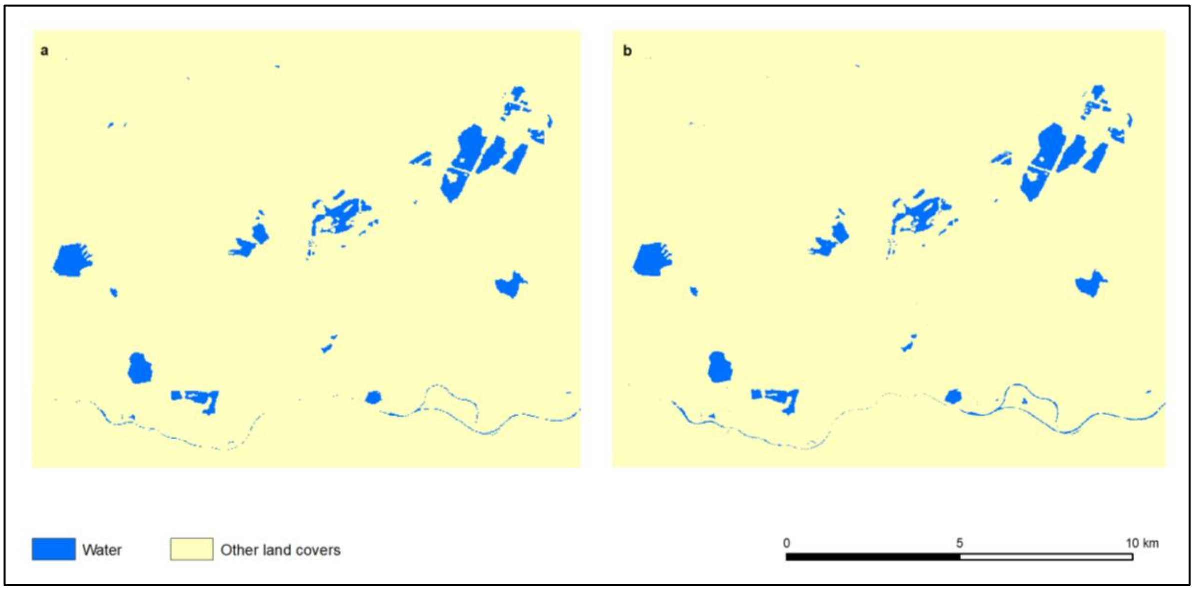


*Figure 4. Landsat 8 (a) and Sentinel 2 (b) data classification by means of Maximum Likelihood, (source: authors, data: USGS)*

River Elbe is not well classified as it can be clearly seen in the Figure 4. Error matrix revealed in both cases 12 wrongly classified points representing water.

### Class Probability

Training areas created for maximum likelihood are used in this method as well. This classification provided very similar results to the maximum likelihood classification with 12 wrongly classified points representing water in both cases. River Elbe is not well classified as in previous case – classification by maximum likelihood.

### Principal Components

Determining number of principal components is the first step of this method. At first, 10 components for Sentinel and 8 components for Landsat is used to respect number of bands. Results reveal that 3 components explain 98 % of variance. The 3 main components (bands) are used for the next classification by means of Iso Cluster. Image is classified into 5 classes, which are integrated into 2 final classes by reclassification. Error matrix contained no wrongly classified points for Landsat data and 1 wrongly classified water point for Sentinel data.

### Normalized difference water index

Normalized difference water index (NDWI) and NDWI-based indices are understood as suitable indices for water bodies identification (Yang, 2017; Zhou et al., 2017). According to Jiang et al. (2014), NDWI may be inefficient at identifying pixels containing mixture of water bodies and vegetation.

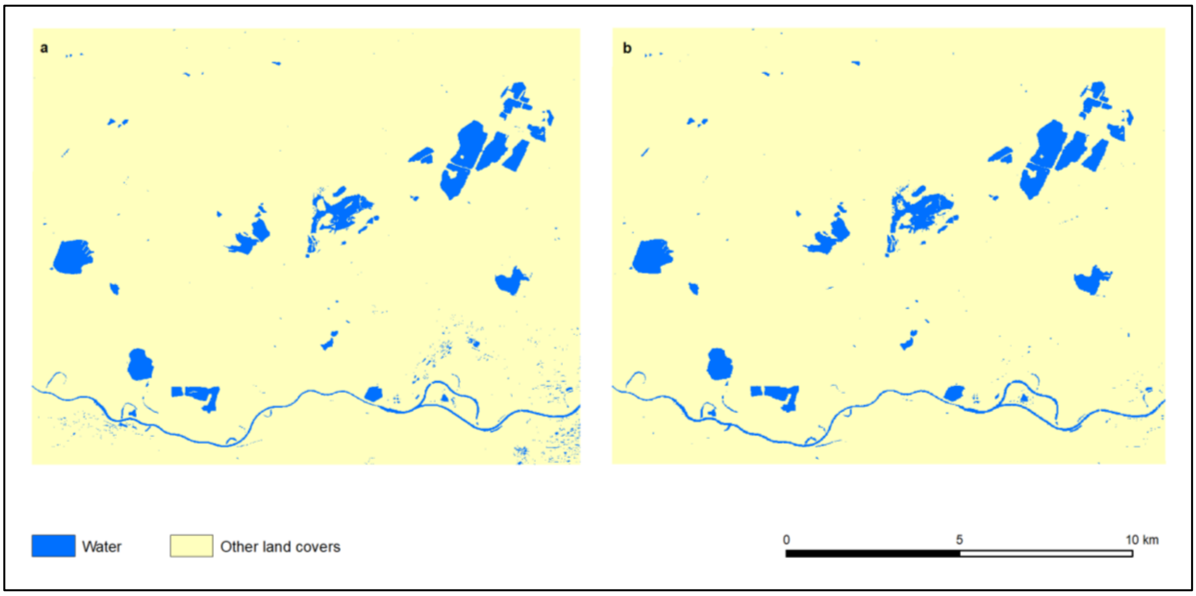


*Figure 4. NDWI calculation based on Sentinel 2 data (a) and the most successful classification method – Iso Cluster classified into 5 classes (b), (source: authors, data: USGS)*

### UAV's data

DJI camera allows collection of data only in the visible part of spectrum. Thus, NDWI cannot be calculated. Iso Cluster was used as the classification method. Data set was classified into 5, 6, 7, and 8 classes to find suitable parameters of the classification. In all classifications, one class represents no data, all other classes represent different land cover types. Results of classification into 6 and 8 classes are compared with results of Sentinel data classification by the same method and with NDWI calculated from Sentinel data (see Figure 5). Blue colours represent water, green colours trees and greenery in general, and beige colours represents seasonally flooded greenery.

UAV images are very heterogeneous, so classification results were used to support manual identification of the shoreline. Manual classification was a reasonable solution because of available data (only the visible part of spectrum) and a small size of the area of interest. Next reason supporting manual shoreline identification is the fact that some parts of the shoreline are covered with treetops. Manual interpretation allows to follow the direction of the shoreline, which was examined in terrain.

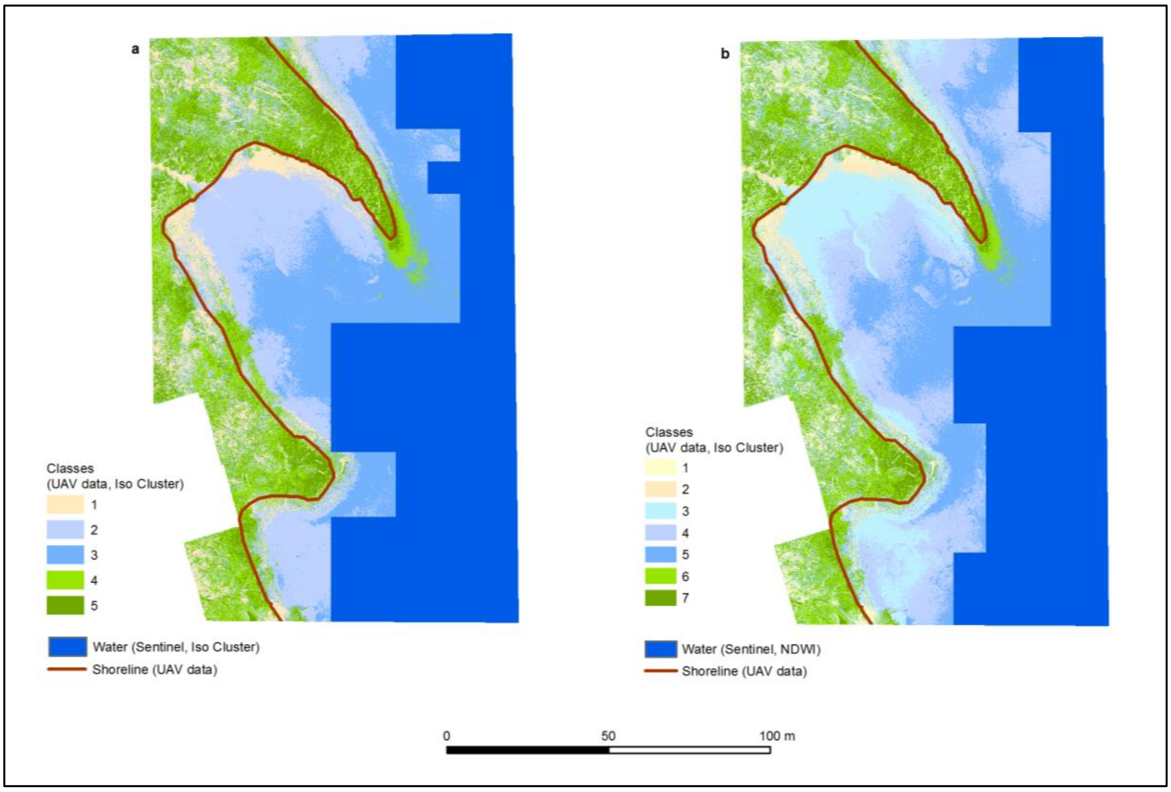


*Figure 5. UAV data classification by Iso Cluster into: (a) 5 classes, results of Sentinel 2 classified by Iso Cluster into 5 classes are included; and (b) into 8 classes, results of NDWI calculated from Sentinel 2 are included (source: authors, data: USGS)*

## Comparison of Used Data and Methods

Landsat and Sentinel data have been frequently used for many purposes, including water bodies identification, e.g. Bhagat et Sonawane (2011), Jiang et al. (2014), Jones et al. (2017), Salama et al. (2012), Yang et al. (2017). Their time and spatial resolution and presence of cloud cover can represent a limitation of their utilization (Pásler et al. 2015). UAV is available on demand, with respect only to legal and weather limitations.

Various criteria are used to compare used classification methods: Wrongly classified points; difference in total area of two reference ponds; difference of total area of all classified water bodies; complexity of a tool (i.e. number of necessary steps); and time necessary to calculation. Classification of UAV data is not included because of significant differences. Table 1 shows comparison of the used methods and data.

*Table 1. Comparison of used methods*

| Classification tool/Method | Data | No. of Classes | Wrongly Classified Points | Comple-xity | Difference in Total Area of Two Reference Ponds [ha] | Difference in Total Area of Water Surfaces [ha] | Time of Calculation [s] |
|---|---|---|---|---|---|---|---|
| Iso Cluster | Sentinel | 5 | 0 | 1 | 0.04 | 0 | 7.04 |
| Iso Cluster | Landsat | 5 | 0 | 1 | 1.00 | 17 | 3.98 |
| Principal Components + Iso Cluster | Sentinel | 10 + 5 | 0 | 3 | 13.28 | 20 | 84.52 |
| Principal Components + Iso Cluster | Landsat | 3 + 5 | 0 | 3 | 1.68 | 20 | 7.38 |
| Principal Components + Iso Cluster | Landsat | 8 + 5 | 0 | 3 | 21.46 | 37 | 11.88 |
| Principal Components + Iso Cluster | Sentinel | 3 + 5 | 1 | 3 | 1.37 | 9 | 30.44 |
| NDWI | Sentinel | - | 2 | | | | |
| Iso Cluster | Sentinel | 20 | 4 | 1 | 1.17 | 177 | 11.83 |
| Iso Cluster | Landsat | 20 | 4 | 1 | 11.75 | 219 | 4.74 |
| Maximum Likelihood | Sentinel | 4 | 12 | 2 | 1.18 | 314 | 4.54 |
| Class Probability | Sentinel | 4 | 12 | 2 | 19.86 | 316 | 6.64 |
| Class Probability | Landsat | 4 | 12 | 2 | 19.72 | 331 | 3.83 |
| Maximum Likelihood | Landsat | 4 | 12 | 2 | 21.46 | 331 | 2.03 |

*Source: authors*

The best obtained result of satellite data processing, i.e Sentinel 2 data classified by Iso Cluster into 5 classes, is visualised together with 8A band of the Sentinel image to demonstrate the quality of classification (see Figure 6).

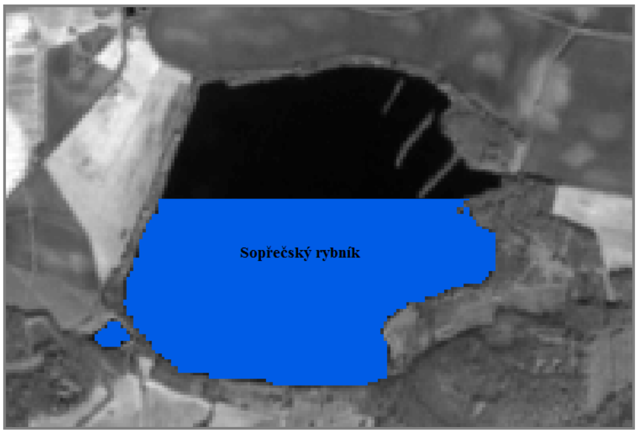


*Figure 6. Result of Sentinel 2 data by Iso Cluster Classification into 5 classes, 8A band of the Sentinel image is used as a background (source: authors, data: USGS)*

Figure 7 provides visualisation of the best method of satellite data processing for small water bodies identification in the whole area of interest. Namely, result of the Iso Cluster classification of Sentinel 2 data into 5 classes is shown. Figure 5 provides comparison of the results of the Iso Cluster classification of Sentinel 2 data into 5 classes with NDWI calculated from the Sentinel data. It can be seen that NDWI provides slightly worse result. UAV data classification and observation in terrain reveal that there is a mixture of water and vegetation.  According to Jiang et al. (2014), NDWI may be inefficient at identifying this type of land cover.

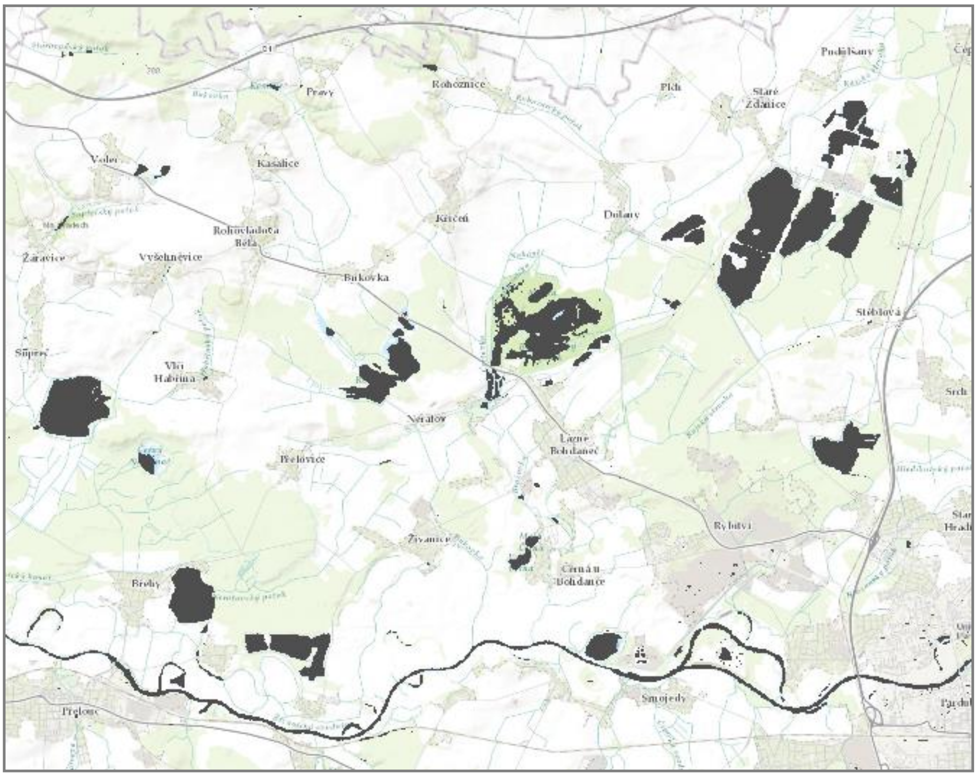

*Figure 7. NDWI calculation based on Sentinel 2 data.*

## CONLCUSION

Identification of small water bodies and their shorelines based on remotely sensed data belong to important tasks today. Landsat and Sentinel data have been frequently used for water bodies identification but they are still partly limited by spatial and temporal resolution. Higher spatial and temporal resolution can be obtained using UAVs for data collection.

The case study compared Landsat 8, Sentinel 2 and UAV data suitability for identification of smaller water bodies and their shorelines near the city of Pardubice, the Czech Republic, by means of several classification methods and NDWI. Sentinel data classified by means of Iso Cluster provided the best results, followed by NDWI based on Sentinel data too.

A short part of one pound inside the area of interest was sensed by means of UAV in a visible part of the spectrum as well to obtain more detailed data. The same classification method (Iso Cluster) was used. Manual vectorization of the shoreline was used as another method. Data sets classified by mean of Iso Cluster into 5 – 8 classes were used to support manual classification. It resulted into shoreline identification at the high level of resolution. Indented shoreline and shallow water were probably reasons for slightly worse classification of satellite imagery in the particular case of comparison with the UAV data. Treetops represent one of the reason why manual interpretation of the shoreline was chosen. It allowed to follow the direction in the cases when the shoreline was overlapped by treetops. Other methods of classification will be tested in future, e.g. neural networks or object-oriented classification are methods.


## ACKNOWLEDGEMENT

This research is supported by University of Pardubice, SGS_2018_19 project.


## REFERENCES


Bhagat, V.S., Sonawane, K.R., (2011). Use of Landsat ETM plus data for delineation of water bodies in hilly zones. *Journal of Hydroinformatics*, 13(4), pp. 661-671.

Biggs, J. von Fumetti, S., Kelly-Quinn, M., (2017). The importance of small waterbodies for biodiversity and ecosystem services: implications for policy makers. *Hydrobiologia*, 793(1), pp. 3-39.

Bukata, R.P., Harris, G.P., Bruton J.E., (1974). The detection of suspended solids and chlorophyll-a utilizing digital multispectral ERTS-1 data. In proceedings of the 2nd Can. Symp. Remote Sensing, pp. 552–564.

Feyisa, G.L., Meilby, H., Fensholt, R., Proud, S.R., (2014). Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sensing of Environment*, 140, pp. 23-35.

Fisher, A., Flood, N., Danaher, T., (2016). Comparing Landsat water index methods for automated water classification in eastern Australia. *Remote Sensing of Environment*, 175, pp. 167-182.

Huda, K.M.S., Atkins, P.J., Donoghue, D.N.M., Cox, N.J. (2010). Small water bodies in Bangladesh. *Area* 42(2), pp. 217-227.

Jiang, H., et al., (2014). An Automated Method for Extracting Rivers and Lakes from Landsat Imagery. *Remote Sensing*, 6(6), pp. 5067-5089.

Jones, S. et al., (2017). Big Data and Multiple Methods for Mapping Small Reservoirs: Comparing Accuracies for Applications in Agricultural Landscapes. *Remote Sensing*, 9(12), article number 1307.

Lazne Bohdanec: Short Introduction, (2015), [online]. URL:<http://lazne.bohdanec.cz/english<

Lejot et al., (2007). Very high spatial resolution imagery for channel bathymetry and topography from an unmanned mapping controlled platform. *Earth Surface Processes and Landforms*, 32(11), pp. 1705–1725.

Lu. D., Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5), pp. 823-870.

McDonald, R.I., et al. (2014). Water on an urban planet: Urbanization and the reach of urban water infrastructure. *Glob. Environ. Chang.*, 27, pp. 96–105.

McFeeters, S.K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), 1425–1432.

Pásler, M., Komárková, J., Sedlák, P., (2015). Comparison of possibilities of UAV and Landsat in observation of small inland water bodies. In International Conference on Information Society, i-Society, pp. 45-49.

Pimentel, D. et al. (2004). Water resources: Agricultural and environmental issues. *BioScience*, 54(10), pp. 909–918.

Salama, M., Radwan, M., van der Velde, R., (2012). A hydro-optical model for deriving water quality variables from satellite images (HydroSat): A case study of the Nile River demonstrating the future Sentinel-2 capabilities. *Physics and Chemistry of the Earth, Parts A/B/C*, 50 – 52, pp. 224-232.

Verpoorter, C., Kutser, T., Tranvik, L., (2012). Automated mapping of water bodies using Landsat multispectral data. *Limnology and Oceanography: Methods*, 10, pp. 1037–1050

Wang S.D. et al., (2015). A Simple Enhanced Water Index (EWI) for Percent Surface Water Estimation Using Landsat Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8, pp. 90-97.

Yang, S., Xue, C., Liu, T., Li, Y., (2010). A method of small water information automatic extraction from TM remote sensing images. *Acta Geodaetica et Cartographica Sinica*, 39(6), pp. 611-617.

Yang X, Zhao S, Qin X, et al., (2017). Mapping of Urban Surface Water Bodies from Sentinel-2 MSI Imagery at 10 m Resolution via NDWI-Based Image Sharpening. *Remote Sensing*, 9(6), article number 596.

Zhou, Y., et al., (2017). Open Surface Water Mapping Algorithms: A Comparison of Water-Related Spectral Indices and Sensors. *Water* 9(4), article number 256.

## BIOGRAPHY

Jitka Komarkova works for the Institute of System Engineering and Informatics, Faculty of Economics and Administration, University of Pardubice, the Czech Republic as an Associate Professor and GIS specialist. She is major in the following topics: a) web-based GIS solutions including their usability; b) spatial data analyses in GIS; c) utilization of UAVs as a way how to remotely sense the Earth’s surface.