

Reply

**Response to Johnson B.A. Scale Issues Related to the Accuracy Assessment of Land Use/Land Cover Maps Produced Using Multi-Resolution Data: Comments on “The Improvement of Land Cover Classification by Thermal Remote Sensing”.**  
*Remote Sens.* 2015, 7, 8368–8390

Liya Sun <sup>1,2,\*</sup> and Karsten Schulz <sup>2</sup>

<sup>1</sup> Department of Geography, Ludwig Maximilian University of Munich, Munich 80333, Germany

<sup>2</sup> Institute for Water Management, Hydrology and Hydraulic Engineering (IWHW), University of Natural Resources and Life Sciences, Vienna 1180, Austria; E-Mail: karsten.schulz@boku.ac.at

\* Author to whom correspondence should be addressed; E-Mail: liya.sun@boku.ac.at;  
Tel.: +43-1-47654-5501; Fax: +43-1-47654-5549.

Academic Editors: Ruiliang Pu and Prasad S. Thenkabail

Received: 24 August 2015 / Accepted: 10 October 2015 / Published: 15 October 2015

---

**Abstract:** Following the suggestion made by Johnson (Johnson B.A., 2015), a polygon-based cross validation (CV) method is compared to the pixel-based CV method to classify different levels of land cover categories using a single-date Landsat 8 image and time series of Landsat TM images. Also, different variants of band combinations, with and without the thermal bands, were considered. The results demonstrate that the inclusion of thermal information into the classification process will improve the classification performance, as was already shown in our original study (Sun and Schulz, 2015). However, it is also demonstrated that the polygon-based CV method produced lower overall accuracy values when compared to the pixel-based CV method. This confirms the argument made by Johnson that a correlation of calibration and validation data due to random sampling of multi-scale data will overestimate the performance of the classifier, and independent polygon-based CV methods have to be applied instead.

**Keywords:** thermal remote sensing; land cover classification; polygon-based cross validation; multi-resolution images; Landsat 8 image; calibration/validation data correlation

---

## 1. Introduction

We very much appreciate Johnson's comments [1] on our recently published results [2] on the improvement by thermal remote sensing for land cover classification. In our original paper, we made use of single-date Landsat 8 and time series of Landsat 4/5 images to investigate the potential of thermal information for an improved land cover classification in the Attert Catchment in the Grand Duchy of Luxembourg [2]. The classification results were assessed by a 10-fold pixel based cross-validation (CV) method where pixel were randomly selected and the overall accuracy (OA) taken as an evaluation measure. We found that the inclusion of thermal bands can improve the accuracy of the land cover classification when added to the multispectral bands. Based on the accuracy assessment, we also concluded that the time series of thermal images alone produced similar classification results when compared to all other VIS/NIR and TIR band combinations.

In his comment, Johnson pointed out that the high accuracy data produced by the thermal images were likely caused by the overestimation of the pixel-based CV method [1]. He also recommended performing the CV at the region-of-interest (ROI)/polygon level, thereby avoiding the spatial autocorrelation between training and validation pixels given multi-resolution data.

We acknowledge the various arguments around the accuracy assessment of the multi-scale remotely sensed images [3–5] and the numerous data fusion methods [6,7]. While our experiments focused on the evaluation of the thermal bands compared to the other multispectral variables, the assessment problems of the random sampling procedure related to the scale issues of the resampled thermal images were neglected. In this response, we have carefully taken Johnson's suggestion into account and evaluated all our original land cover classification with the polygon-based CV method.

In the following sections, we briefly introduce the new assessment procedure and show the classification results for both the single-date and time series of image applications. While we will briefly introduce the data used, a detailed description of the images and the classification algorithm can be found in [2].

## 2. Polygon-Based Cross Validation Method

Based on the suggestion of [1], we added a polygon-based CV analysis to our study. Training and validation samples were collected using the area of interest tool of the drawing toolbox in the ERDAS software (ERDAS, Inc., Atlanta, GA, USA); the polygons here are the same as in the original paper [2]. However, in order to make sure the pixels from the polygon as pure as possible, here the pixels were kept only when their central points are inside the polygon; the polygons were then refined. This leads to the slightly smaller sample sizes compared to the original paper. As we are not aiming at evaluating the effects of the training size on the classification accuracy, here the comparison between the pixel-based CV method and the polygon-based CV method was conducted using the same size of samples (pixels with central point inside the polygon).

For all land cover categories (except for water bodies) more than 10 polygons of ground truth data exists and there is little difference among the size of each polygon. For the 10-fold polygon-based cross validation, all polygons were split into 10 smaller sets and reorganized into 10 groups of data. The classification model was trained using pixels from nine groups of polygons (training data);

the resulting model is then validated on the remaining group (validation data), and this procedure is repeated 10-times so that each group is used for validation once. In this way the correlation of multi-resolutions calibration and validation is avoided.

### 3. Results and Discussion

In this section, the accuracy statistics for a land cover classification applying both the original pixel based CV method and the polygon-based CV method are compared. Input data (a single-date Landsat 8 image on 21 July 2013, as well as time series of images from Landsat 4/5) and classification methods (Random Forest and the  $k$ -NN algorithms) are kept as in the original paper [2].

#### 3.1. Three-Level Classification Based on the Single-Date Landsat 8 Image (S1)

In a first part, three variants of spectral band combinations of the single Landsat 8 image are used as input data and include: (i) Bands4, only considering bands 2 to 5 in the VIS/NIR spectral region, without thermal bands; (ii) Bands6T, as Bands4, but adding the two thermal bands 10 and 11; and (iii) Bands10T, including all bands, except the panchromatic band.

Table 1 summarizes the accuracy evaluation results for both the polygon-based and the pixel-based CV methods. The OA values for the classification of Level 1 land cover category obtained by the polygon-based CV are almost identical to the pixel-based evaluation for the three variants (97% to 98%). The OA values of Level 2 and Level 3 categories decreased about 5% and 9%, respectively, when switching to the polygon-based CV method. However, the increase in performance when adding the thermal bands into the classification is still pronounced and our conclusions drawn in the original paper still hold.

**Table 1.** The mean values of overall accuracy (OA) calculated by a polygon-based 10-fold cross validation (CV) method for the three variants from Landsat 8 in 2013 classified by  $k$ -NN and Random Forest: Bands4, only considering bands 2 to 5 in the visible and near infrared spectral region without thermal bands; Bands6T, as Bands4, but adding the two thermal bands 10 and 11, and Bands10T, including all bands, except the panchromatic band (see [2] for a detailed channel/band description);  $k$ -NN5 and RF represents the nearest neighbor method with  $k = 5$  and Random Forest, respectively.

Image Classification Accuracy in 2013	Level 1 (%)		Level 2 (%)		Level 3 (%)	
	$k$ -NN5	RF	$k$ -NN5	RF	$k$ -NN5	RF
<b>Assessed by Polygon-Based CV</b>						
Bands4	97.0	97.5	83.3	82.5	67.5	67.7
Bands6T	96.5	96.7	88.7	88.8	79.1	79.2
Bands10T	96.9	97.6	88.0	87.5	80.8	78.0
<b>Assessed by Pixel-Based CV</b>						
Bands4	98.4	98.5	86.2	87.2	76.5	78.2
Bands6T	98.2	98.4	92.1	92.9	87.7	88.1
Bands10T	98.7	98.6	93.5	93.5	89.3	89.2

### 3.2. Two-Level Classification Based on Time Series of Images (TS1 and TS2)

The second part of the analysis focus on the differences between polygon-based and pixel-based CV methods when using time series of Landsat 4/5 images as input in the land cover classification system (Tables 2 and 3). The time series of images consist of two groups: TS1, including 7 images between 1984 and 1990, and TS2, including 6 images from 2006 to 2011. Five variants of the times series were analyzed including the following different band combinations: B3B4, the combinations of band 3 and band 4 of Landsat TM; 3PC, the first three principal components of all VIS/NIR bands; 6Bands, all bands except the thermal band; Thermal only the single thermal band; and 7Bands, the combination of all 7 Landsat 4/5 bands.

**Table 2.** Overall Accuracy of Level 1 classification assessed by the polygon-based CV method using five variants of time series of images: B3B4, the combination of band 3 and band 4; 3PC, the first three principal components of all bands; 6Bands, the combination of six bands; Thermal, the thermal band; and 7Bands, the combination of all seven bands.

Image Number	B3B4 (%)		3PC (%)		6Bands (%)		Thermal (%)		7Bands (%)	
	<i>k</i> -NN	RF	<i>k</i> -NN	RF	<i>k</i> -NN	RF	<i>k</i> -NN	RF	<i>k</i> -NN	RF
TS1-1	86.6	87.7	91.6	91.2	86.6	87.7	87.8	90.5	96.2	95.5
TS1-2	91.9	90.9	94.8	94.4	95.6	96.2	88.4	88.8	96.5	97.0
TS1-3	94.2	94.9	95.9	95.8	96.4	96.1	88.4	90.9	96.8	96.3
TS1-4	95.6	96.2	96.8	96.7	96.4	97.1	88.7	87.0	96.8	96.7
TS1-5	96.1	96.2	96.8	96.2	97.3	97.0	91.7	91.3	96.8	97.3
TS1-6	96.1	95.9	97.2	96.3	96.8	97.4	91.7	89.3	96.9	97.1
TS1-7	96.0	96.6	97.4	97.2	97.7	97.6	93.2	94.6	97.7	97.0
TS2-1	91.4	91.5	95.6	94.2	95.8	95.5	80.0	86.0	95.9	95.1
TS2-2	95.1	95.9	97.2	96.9	97.5	97.0	84.5	89.3	97.8	97.0
TS2-3	97.1	97.2	98.0	97.6	98.4	97.8	92.6	92.7	98.3	98.2
TS2-4	97.0	97.8	97.8	97.9	97.8	98.2	95.0	92.6	98.7	98.0
TS2-5	98.1	97.9	98.7	98.3	98.0	99.0	94.2	94.4	98.3	98.3
TS2-6	98.5	98.2	98.6	98.1	98.6	98.4	95.4	95.4	98.3	98.4

**Table 3.** Overall Accuracy of Level 2 classification assessed by the polygon-based CV method using five variants of time series of images: B3B4, the combination of band 3 and band 4; 3PC, the first three principal components of all bands; 6Bands, the combination of six bands; Thermal, the thermal band; and 7Bands, the combination of all seven bands.

Image Number	B3B4 (%)		3PC (%)		6Bands (%)		Thermal (%)		7Bands (%)	
	<i>k</i> -NN	RF	<i>k</i> -NN	RF	<i>k</i> -NN	RF	<i>k</i> -NN	RF	<i>k</i> -NN	RF
TS1-1	63.3	62.0	73.3	72.2	79.4	80.3	39.2	51.6	79.7	79.4
TS1-2	72.6	73.3	78.7	77.9	81.5	82.0	50.4	52.9	80.4	81.8
TS1-3	77.9	76.9	79.3	81.9	83.4	82.8	55.4	57.9	81.9	83.3
TS1-4	80.8	81.5	85.0	85.5	86.1	84.2	62.2	60.7	81.7	83.2
TS1-5	82.7	83.7	84.1	84.5	84.6	86.1	66.3	66.3	83.3	85.2
TS1-6	83.2	83.4	86.0	85.9	85.2	84.7	66.8	68.8	85.1	86.2

Table 3. Cont.

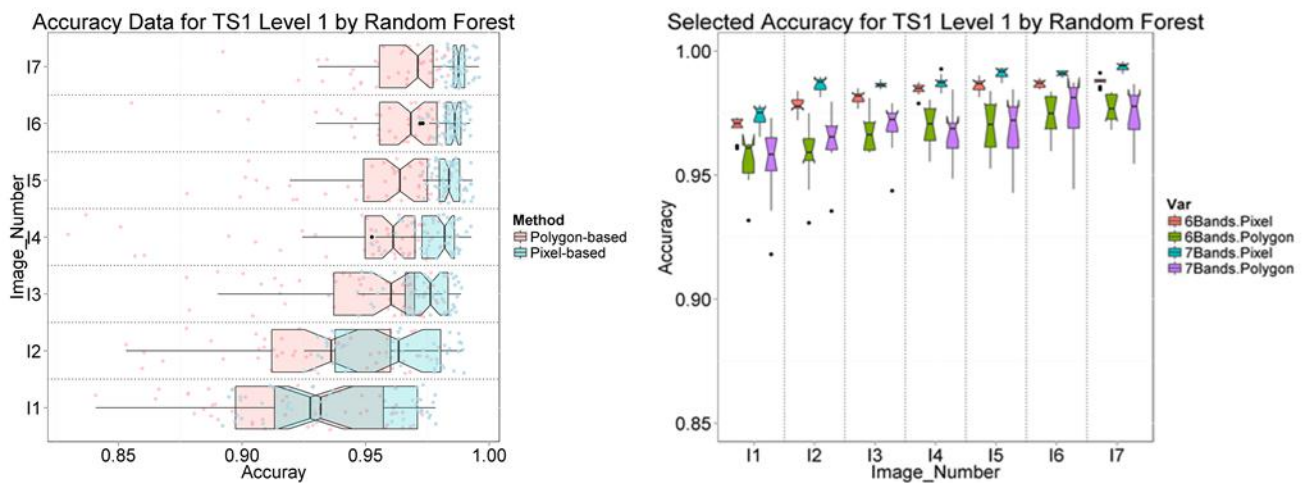
Image Number	B3B4 (%)		3PC (%)		6Bands (%)		Thermal (%)		7Bands (%)	
	<i>k</i> -NN	RF	<i>k</i> -NN	RF	<i>k</i> -NN	RF	<i>k</i> -NN	RF	<i>k</i> -NN	RF
TS1-7	82.9	83.0	84.4	85.0	85.1	86.3	69.9	72.0	85.4	86.6
TS2-1	68.8	69.6	83.7	82.4	85.7	86.0	42.0	48.7	83.6	84.5
TS2-2	79.4	80.3	88.4	89.8	90.1	90.9	57.3	58.5	90.6	89.8
TS2-3	86.8	86.2	89.9	92.4	88.7	91.0	60.9	63.9	91.2	91.0
TS2-4	87.8	89.6	90.6	91.6	91.8	91.5	65.7	64.9	92.7	91.6
TS2-5	87.7	90.1	92.0	92.6	91.1	92.3	70.3	71.5	92.2	93.0
TS2-6	89.1	91.3	92.1	92.3	90.2	91.9	74.5	70.4	92.0	93.3

The best OA value from the pixel-based CV is almost the same as in the original paper, the details can be found in the Supplementary materials (Tables S1 and S2). For simplification, we here only list the OA values of the polygon-based CV for Level 1 and Level 2 land cover categories classification. Table 2 shows OA values of the polygon-based CV for the classification of Level 1 land cover categories, whereby images within each group (TS1, TS2) are added subsequently as input. The classification performance increases with increasing numbers of images for all variants, a behavior that is already shown in our original study using a random pixel-based CV. Also, the 7Bands variant including the thermal band still achieved the best overall performance, especially for smaller number of images included. Using the full set of available images all variants performed almost equally well, with OA values of 96%–98.5%. When only using the thermal band, classification performance is reduced by 4.5% compared to the pixel-based CV in our original paper, nicely demonstrating the overestimation of performance when correlation of multi-resolution calibration and validation data are existent, as commented by Johnson [1].

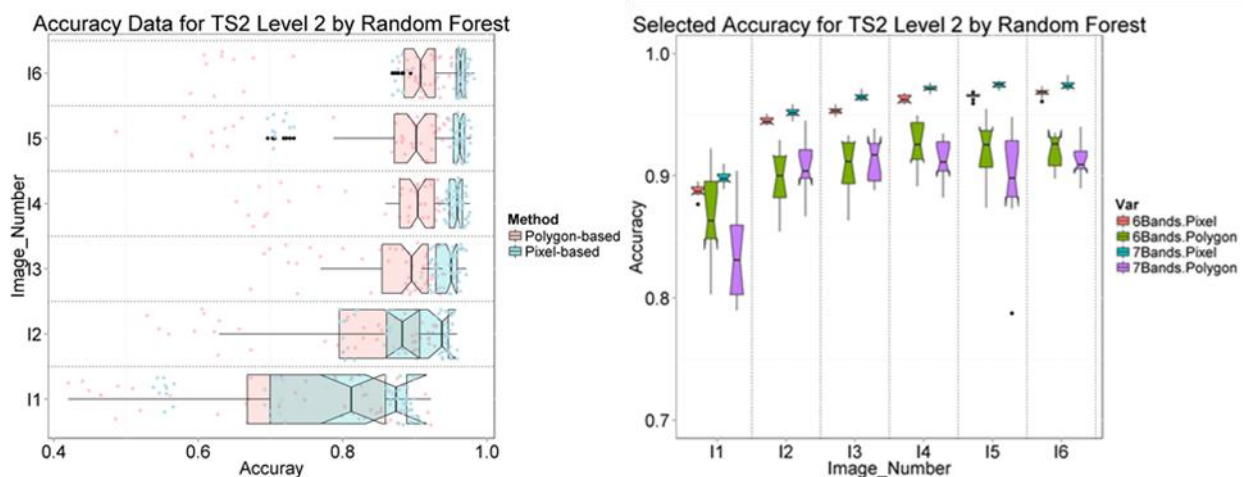
The differences between the pixel-based based and a polygon-based 10-fold CV method are summarized in Figure 1. Here, OA values of all five variants from TS1 are summarized in a single box-plot for each time-step in the left image of Figure 1. It is clearly seen that on average the polygon-based CV method produced significant lower OA for both time series (TS1, TS2), again supporting the issue raised by Johnson’s comment. Two variants of 6Bands and 7Bands were selected to show the detailed variation for the two methods in the right image of Figure 1. Besides the lower OA in comparison to the pixel-based CV method, the polygon-based CV still produced higher average OA for the 7Bands compared to the 6Bands without thermal band (the right image in Figure 1).

Repeating this analysis for the classification of Level 2 land cover categories, the differences in the performance measure (OA) between both CV methods is even more pronounced. Table 3 summarizes the classification results. The average OA values are in general lower, as we analyze more specific land cover categories. The best OA for TS1 and TS2 are 86.6% and 93.3% when including the full set of images. The 7Bands variant including the thermal band still achieved the best OA value of 86.6% for TS1, which is 10.9% lower than 97.5% from the pixel-based CV method. The best OA of TS2 from polygon-based CV method is about 5% lower than the corresponding value for the pixel-based CV method. Again, the Thermal variant including only the single thermal band, showed only a relatively weak performance with OA values of 72% and 74.5% for the both time series, compared to 96.8% 95.8% for the pixel-based CV method. Figure 2, taking TS2 as an example, summarizes the differences between

the pixel-based and the polygon-based CV methods in the left image and displays the variation of the selected 6Bands and 7Bands variants, again supporting the issue raised by [1].



**Figure 1.** The distribution of OA values for the Level 1 land cover category classification using times series TS1, the polygon-based and pixel-based 10-fold cross validation methods and the Random Forest methods. **(Left):** all variants (Table 2) are summarized in a single box-plot; **(Right):** OA comparison of selected variants of 6Bands without thermal band and 7Bands with thermal band.



**Figure 2.** The distribution of OA values for the Level 2 land cover category classification using times series TS2, the polygon-based and pixel-based 10-fold cross validation methods and the Random Forest methods. **(Left):** all variants (Table 3) are summarized in a single box-plot; **(Right):** OA comparison of selected variants of 6Bands without thermal band and 7Bands with thermal band.

#### 4. Conclusions

In this response to [1], a polygon-based CV method, as suggested by [1], was applied to evaluate a land cover classification for three different levels of land cover categories. The classification was based on (i) a single-date Landsat 8 image; and (ii) time series of Landsat 4/5 images. The performance of

classification results using the polygon-based CV were compared to a pixel-based CV method as applied in our original application [2].

For the single-date Landsat 8 image, the polygon-based method achieved almost similar accuracy values when compared to the pixel-based method, for all three-levels of land cover categories and for both classification methods used. When using time series of images, five different variants of band combinations (see Table 2) with and without thermal information have been considered.

The accuracy of the Level 1 classification decreased but to a very acceptable and still useful level when compared to the commonly recommended standard of 85% [8] (the best OA of Thermal is 94.6% and the best OA of 7Bands is 98.4%). The most obvious decline in performance is observed in the classification results for the Level 2 category, of which the best OA among the five variants is 93.3% and only 74.5% when using only the thermal bands.

Consistent with our former findings, the inclusion of the thermal bands still improved the land cover classification in comparison to only using the VIS/NIR bands, also when assessing the classification results with a polygon type of CV approach. This has also been shown by other researcher, Eisavi *et al.* [9] applied the random forest classifier to the multi-temporal spectral and thermal features in land cover classification and found that the contribution of multi-temporal thermal information led to a considerable increase in the accuracy data. When using time series of thermal images to classify land cover at the Level 2 category, the performance and OA values were significantly lower for the polygon based CV when compared to the pixel-based evaluation for all band combinations considered. Again, the inclusion of thermal information improved the classification results on various levels.

In summary, a clear effect of correlation in the samples for calibration and validation due to multi-resolution data could be observed here. Classification accuracy (OA) was highly overestimated when ignoring correlation effects in the selection of calibration and validation data using time series of images as input. We therefore strongly support the comments made by Johnson and can only support his recommendation concerning the appropriate choice of CV methods. However, the choice of the CV method did not change our original conclusions, in that the inclusion of thermal data into the classification process, can significantly improve classification results.

## Acknowledgments

The authors would like to thank Brian A. Johnson for his comment on our original paper which initialized a very intensive and still ongoing discussion about the effects of correlated multi-scale data within calibration-validation exercises.

We would also like to sincerely thank the German Research Foundation (DFG) for funding this research through the CAOS (Catchments as Organized Systems) Research Unit (FOR 1598, SCHU1271/5-1), and the China Scholarship Committee for the support of the research and USGS for providing all the Landsat images.

## Author Contributions

Liya Sun and Karsten Schulz designed this experiment and analyzed the data together. The results were compared and summarized in this manuscript by Liya Sun. Karsten Schulz has spent the most effort in revising the manuscript.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

1. Johnson, B.A. Scale issues related to the accuracy assessment of land use/land cover maps produced using multi-resolution data: Comments on “the improvement of land cover classification by thermal remote sensing”. *Remote Sens.* 2015, 7, 8368–8390. *Remote Sens.* **2015**, 7, 13436–13439.
2. Sun, L.; Schulz, K. The improvement of land cover classification by thermal remote sensing. *Remote Sens.* **2015**, 7, 8368–8390.
3. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*; CRC Press/Taylor & Francis: Boca Raton, FL, USA, 2009.
4. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, 37, 35–46.
5. Foody, G.M. Status of land cover classification accuracy assessment. *Remote Sens. Environ.* **2002**, 80, 185–201.
6. Colditz, R.R.; Wehrmann, T.; Bachmann, M.; Steinnocher, K.; Schmidt, M.; Strunz, G.; Dech, S. Influence of image fusion approaches on classification accuracy: A case study. *Int. J. Remote Sens.* **2006**, 27, 3311–3335.
7. Zhu, X.; Chen, J.; Gao, F.; Chen, X.; Masek, J.G. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sens. Environ.* **2010**, 114, 2610–2623.
8. Anderson, J.R. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*; U.S. Government Printing Office (GPO): Washington, DC, USA, 1976.
9. Eisavi, V.; Homayouni, S.; Yazdi, A.; Alimohammadi, A. Land cover mapping based on random forest classification of multitemporal spectral and thermal images. *Environ. Monit. Assess.* **2015**, 187, 1–14.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).