

Comment

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(7), 8368–8390

Brian A. Johnson

Institute for Global Environmental Strategies, 2108-11 Kamiyamaguchi, Hayama, Kanagawa 240-0115, Japan; E-Mail: johnson@iges.or.jp; Tel.: +81-46-855-3875; Fax: +81-46-855-3809

Academic Editors: Ruiliang Pu and Prasad Thenkabail

Received: 2 July 2015 / Accepted: 10 October 2015 / Published: 15 October 2015

Abstract: Much remote sensing (RS) research focuses on fusing, *i.e.*, combining, multi-resolution/multi-sensor imagery for land use/land cover (LULC) classification. In relation to this topic, Sun and Schulz [1] recently found that a combination of visible-to-near infrared (VNIR; 30 m spatial resolution) and thermal infrared (TIR; 100–120 m spatial resolution) Landsat data led to more accurate LULC classification. They also found that using multi-temporal TIR data alone for classification resulted in comparable (and in some cases higher) classification accuracies to the use of multi-temporal VNIR data, which contrasts with the findings of other recent research [2]. This discrepancy, and the generally very high LULC accuracies achieved by Sun and Schulz (up to 99.2% overall accuracy for a combined VNIR/TIR classification result), can likely be explained by their use of an accuracy assessment procedure which does not take into account the multi-resolution nature of the data. Sun and Schulz used 10-fold cross-validation for accuracy assessment, which is not necessarily inappropriate for RS accuracy assessment in general. However, here it is shown that the typical pixel-based cross-validation approach results in non-independent training and validation data sets when the lower spatial resolution TIR images are used for classification, which causes classification accuracy to be overestimated.

Keywords: image fusion; cross-validation; multi-resolution; multi-sensor; Landsat 8

Fusion of multi-resolution and/or multi-sensor remote sensing (RS) imagery has been shown to result in higher classification accuracy in many past studies [2–15], so classification-oriented image fusion is an important research topic. Satellite data from the Landsat series is commonly-used for land use/land cover (LULC) classification in RS, and Landsat 4/5/7/8 have image bands that vary in terms of spatial resolution, as detailed in [16]. In a recent study using Landsat 4/5/8 data, Sun and Schulz [1] combined the lower spatial resolution thermal infrared (TIR) image bands (120 m for Landsat 4/5, 100 m for Landsat 8) with the higher spatial resolution visible-to-near infrared (VNIR) image bands (30 m) for LULC classification, and found that the combined result led to higher overall classification accuracy. They also found that using multi-temporal TIR data alone for classification resulted in comparable (and in some cases higher) LULC classification accuracies to the use of multi-temporal VNIR data, which contrasts with the results of another recent study [2] and is particularly surprising given the lower spatial resolutions of the TIR bands. In general, I agree with the authors' sentiment that more research on to the combined use of VNIR-TIR data for classification is needed. However, given the details provided in the manuscript, the authors' very encouraging results seem to be due to the use of an improper accuracy assessment procedure rather than the utility of VNIR-TIR data fusion for LULC classification. The aim of this comment is not to dismiss the work of Sun and Schulz, which was actually quite interesting, but rather to highlight the importance of considering scale issues for accuracy assessment, particularly when multi-resolution imagery is used for classification.

Sun and Schulz used a 10-fold cross-validation procedure for accuracy assessment [17], which means 10% of the training pixels are withheld for accuracy assessment in each fold. While the use of cross-validation is not uncommon in RS, a problem with it in their study comes from the fact that the TIR image bands are resampled from their original resolutions of 100 m (Landsat 8) or 120 m (Landsat 4/5) to 30 m to match the VNIR bands [16], meaning that each original TIR pixel is represented by roughly nine (3×3) resampled pixels in the case of Landsat 8, or 16 pixels (4×4) in the case of Landsat 4/5. The implication of this resampling is that in the cross-validation process, a single original TIR pixel is very likely to be represented in both the training and validation sets, as shown in the example in Figure 1. The training and validation sets should be spatially independent to ensure reliable estimation of LULC classification accuracy [18], as the inclusion of the same data in the training and validation data sets will lead to overestimation of classification accuracy. This non-independence of the training and validation data sets would explain why the TIR bands performed as well as the VNIR bands for LULC classification, despite their lower spatial resolutions, and it would also explain why the classification accuracy they achieved using the combined VNIR-TIR was so high (>99% overall accuracy for the classifications that used multi-temporal imagery).

A correct way to perform cross-validation taking into account the multi-resolution nature of the data would be to do it at the region-of-interest (ROI), *i.e.*, polygon, level rather than at the individual pixel level. This would involve holding out all of the pixels within 10% of the ROIs in each cross-validation fold, so individual pixels would still be the base units for accuracy assessment. This procedure would ensure training/validation data independence as long as the distance between ROIs is significantly larger than the pixel size of the lowest spatial resolution image. The only caveat is that there needs to be at least as many ROIs for each LULC class as there are cross-validation folds (e.g., at least 10 ROIs for 10-fold cross-validation), which may in some cases require gathering additional ground truth data. A ROI-based cross-validation approach would also reduce spatial autocorrelation between training and

validation samples caused by their close proximity to one another, which also compromises the assumption of training/validation data set independence even for the higher resolution images [18].

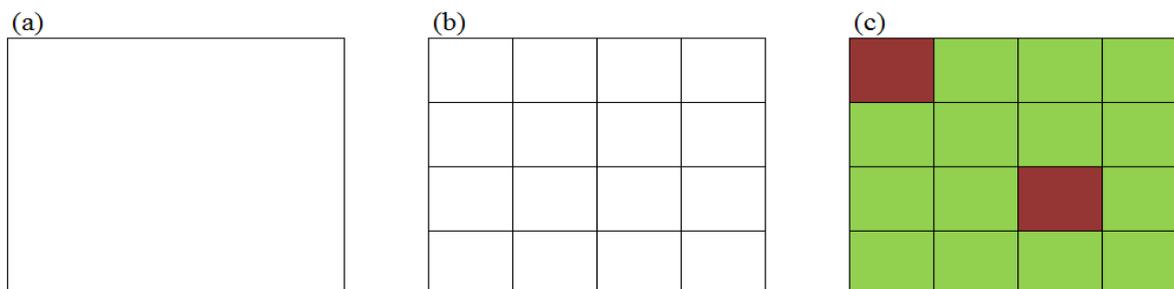


Figure 1. (a) One pixel from an original spatial resolution Landsat 4/5 thermal infrared (TIR) band; (b) original TIR pixel resampled to 4×4 pixels to match the pixels of the higher spatial resolution visible-to-near infrared (VNIR) bands; (c) resampled TIR training pixels (green cells) and validation pixels (red cells) in one fold of a cross-validation, assuming approximately 10% of pixels are held out for validation.

It should be noted that, although Sun and Schulz simply used the resampled lower resolution TIR pixels for classification, meaning that the TIR images were not “sharpened” using the higher resolution imagery (as in some other studies on classification-oriented image fusion [5,6,9,11–15]), the scale issues pointed out here also apply when the lower resolution imagery is “sharpened” using the higher resolution imagery prior to classification, as the spatial resolution of the lower resolution image is only artificially increased, and its pixel values are still derived in part from the original lower resolution image.

I hope that Sun and Schulz can respond to this comment by providing additional information on their cross-validation procedure (in the case that their accuracy assessment did not suffer from the problems pointed out here), or to submit a correction to their manuscript using a proper accuracy assessment method so that the RS community can have a better understanding of the utility of VNIR-TIR data fusion for LULC classification.

Conflicts of Interest

The author declares no conflict of interest.

References

1. Sun, L.; Schulz, K. The improvement of land cover classification by thermal remote sensing. *Remote Sens.* **2015**, *7*, 8368–8390.
2. Eisavi, V.; Homayouni, S.; Maleknezhad, Y.; Alimohammadi, A. Land cover mapping based on random forest classification of multitemporal spectral and thermal images. *Environ. Monit. Assess.* **2015**, *187*, 1–14.
3. Hoan, N.T.; Tateishi, R.; Alsaaidh, B.; Ngigi, T.; Alimuddin, I.; Johnson, B. Tropical forest mapping using a combination of optical and microwave data of ALOS. *Int. J. Remote Sens.* **2013**, *34*, 139–153.

4. Reiche, J.; Verbesselt, J.; Hoekman, D.; Herold, M. Fusing Landsat and SAR time series to detect deforestation in the tropics. *Remote Sens. Environ.* **2015**, *156*, 276–293.
5. 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.
6. Jia, K.; Liang, S.; Zhang, N.; Wei, X.; Gu, X.; Zhao, X.; Yao, Y.; Xie, X. Land cover classification of finer resolution remote sensing data integrating temporal features from time series coarser resolution data. *ISPRS J. Photogramm. Remote Sens.* **2014**, *93*, 49–55.
7. Zhang, C.; Xie, Z. Data fusion and classifier ensemble techniques for vegetation mapping in the coastal Everglades. *Geocarto Int.* **2014**, *29*, 228–243.
8. Zhang, C. Applying data fusion techniques for benthic habitat mapping and monitoring in a coral reef ecosystem. *ISPRS J. Photogramm. Remote Sens.* **2015**, *104*, 213–223.
9. Jia, K.; Liang, S.; Wei, X.; Yao, Y.; Su, Y.; Jiang, B.; Wang, X. Land cover classification of Landsat data with phenological features extracted from time series MODIS NDVI data. *Remote Sens.* **2014**, *6*, 11518–11532.
10. Reiche, J.; Souza, C.M.; Hoekman, D.H.; Verbesselt, J.; Persaud, H.; Herold, M. Feature level fusion of multi-temporal ALOS PALSAR and Landsat data for mapping and monitoring of tropical deforestation and forest degradation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 2159–2173.
11. Johnson, B.A.; Scheyvens, H.; Shivakoti, B.R. An ensemble pansharpener approach for finer-scale mapping of sugarcane with Landsat 8 imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *33*, 218–225.
12. Johnson, B.A.; Tateishi, R.; Hoan, N.T. A hybrid pansharpener approach and multiscale object-based image analysis for mapping diseased pine and oak trees. *Int. J. Remote Sens.* **2013**, *34*, 6969–6982.
13. Lu, D.; Li, G.; Moran, E.; Dutra, L.; Batistella, M. A comparison of multisensor integration methods for land cover classification in the Brazilian Amazon. *GISci. Remote Sens.* **2011**, *48*, 345–370.
14. Palsson, F.; Sveinsson, J.R.; Benediktsson, J.A.; Aanaes, H. Classification of pansharpened urban satellite images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 281–297.
15. Shackelford, A.K.; Davis, C.H. A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1920–1932.
16. Frequently Asked Questions about the Landsat Missions. Available online: http://landsat.usgs.gov/band_designations_landsat_satellites.php (accessed on 13 October 2015).
17. Kohavi, R. A Study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th International Joint Conference on Artificial Intelligence, Montreal, QC, Canada, 20–25 August 1995.
18. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 35–46.