

Reply

Confusion Matrices Help Prevent Reader Confusion: Reply to Bechtel, B., et al. A Weighted Accuracy Measure for Land Cover Mapping: Comment on Johnson et al. Local Climate Zone (LCZ) Map Accuracy Assessments Should Account for Land Cover Physical Characteristics that Affect the Local Thermal Environment, *Remote Sens.* 2019, 11, 2420.

Brian Alan Johnson ^{1,*} and Shahab Eddin Jozdani ²

¹ Natural Resources and Ecosystem Services, Institute for Global Environmental Strategies, 2108-11, Kamiyamaguchi, Hayama, Kanagawa 240-0115, Japan

² Department of Geography and Planning, Queen's University, Kingston, ON K7L3N6, Canada; shahab.jozdani@queensu.ca

* Correspondence: johnson@iges.or.jp; Tel.: +81-45-855-3709

Received: 03 February 2020; Accepted: 06 May 2020; Published: 31 May 2020

Abstract: Land use/land cover (LULC) maps are now being used across disciplines for many different types of applications, e.g., to analyze urban heat islands or rainfall-runoff dynamics. Traditional map accuracy metrics are limited in this regard, as they only assess LULC map thematic accuracy. In reality, some types of misclassification lead to larger estimation errors for these specific applications. In a previous study, we developed a new map accuracy metric (referred to here as “JJ19”) to assess the accuracy of local climate zone maps for urban microclimate analysis. In the previous work, we also attempted to reproduce another metric (weighted accuracy (WA)) proposed for this purpose, but misinterpreted it due to a lack of methodological information available (principally, the lack of a confusion matrix to demonstrate how WA was derived). We sincerely thank the authors of Bechtel et al. 2019 for providing more information on WA in response to our previous study and are happy to report that we found that the metric is now both reproducible and valid. On the other hand, we found some other aspects of Bechtel et al. 2019’s study to be inaccurate, particularly their claims regarding the suitability of the JJ19 metric. Finally, we made a minor improvement to the JJ19 metric based on Bechtel et al.’s comments.

Keywords: accuracy assessment; map validation; land use/land cover; urban heat island; urbanization; microclimate analysis; confusion matrix; image classification

1. Introduction

Land use/land cover (LULC) maps are now commonly used as the basis for many specific applications, e.g., for estimation and modeling of urban heat islands [1,2], forest carbon stocks [3–5], and rainfall-runoff dynamics [6,7]. LULC data is also being incorporated into various environmental and social indicators (e.g., UN Sustainable Development Goal indicators [8,9]) to help monitor progress towards sustainable development at local to global scales. Although this usage of LULC data across scientific disciplines and society is encouraging, there is a danger of misinterpreting the accuracy of LULC maps for these specific applications. For example, the commonly used overall

accuracy (OA) metric does not convey a LULC map's accuracy for a task like above-ground biomass estimation, as some types of misclassifications have greater negative impacts than others, while OA accounts for all types of misclassification equally.

The most basic element for assessing LULC map accuracy is the confusion matrix, or error matrix [10]. From the confusion matrix, traditional LULC map accuracy metrics (OA, producer's accuracy, and user's accuracy) as well as application-specific accuracy metrics can be calculated. However, in many remote sensing studies, no confusion matrix is provided and only the accuracy metrics summarizing the matrix are reported. This lack of a confusion matrix can lead to undue confusion by readers over the accuracy of the LULC map and/or how the reported map accuracy metrics were calculated (particularly when a new accuracy metric is being introduced), as will be shown in this study.

2. Misinterpretation of Weighted Accuracy Metric

In Johnson and Jozdani [11], we misinterpreted a map accuracy metric (weighted accuracy (WA)) originally presented in in Bechtel et al. [12] (hereafter "B17"). The WA metric was intended to convey the accuracy of a local climate zone (LCZ) map [13] (a specific type of LULC map) in portraying the local thermal environment. Thus, WA was given as a substitute (or supplement) for OA to assess overall LULC map accuracy for this specific application. However, the WA metric could not be calculated as the authors had intended, given the information provided in B17.

In B17, the authors explained that WA was calculated by applying (multiplying) the cells of a confusion matrix to a corresponding set of weights, representing the similarity between each pair of LCZ types. However, no explanation was given on how WA was derived from the resultant weighted confusion matrix. Because WA is a proxy for the map's general accuracy, we logically followed the same procedure as is used to calculate OA based on the confusion matrix, i.e., summing the values of the principal diagonal cells in the matrix (i.e., the correctly classified reference samples) and dividing the result by the sum of all the values in the matrix [10] (Table 1). The only difference was that we used the weighted confusion matrix for this rather than the original unweighted confusion matrix (Table 1, "WA; our interpretation"). Indeed, this is a common practice for incorporating other types of weights, e.g., the proportional area of different LULC classes within a study site, for the calculation of overall map accuracy [10,14]. However, we found that this was not what the authors of B17 had intended, as Bechtel et al. [15] revealed that in B17 they had used the sum of *all* cells in the matrix (including the off-diagonal cells), divided by the total number of reference samples, to calculate WA (Table 1, "WA; actual"). This is a rather unconventional interpretation of the confusion matrix, because the off-principal diagonal cells represent the misclassified reference samples in the matrix. Although a more detailed explanation would certainly have helped to convey their calculation method, the simplest reason for the irreproducibility of the metric was that no confusion matrix had been provided in B17 to demonstrate how WA was derived from the weighted confusion matrix.

Table 1. Method used to calculate weighted accuracy (WA). Values in (a) are multiplied by weights in (b) to generate a weighted confusion matrix (c). WA is calculated from the values in (c). Local climate zone (LCZ); overall accuracy (OA).

	(a) Original confusion matrix				(b) Class similarity weights			(c) Weighted confusion matrix					
	LCZ1	LCZ2	LCZ3	Sum	LCZ1	LCZ2	LCZ3	LCZ1	LCZ2	LCZ3	Sum		
LCZ1	50	10	5	65	LCZ1	1	0.92	0.83	LCZ1	50	9.2	4.15	63.35
LCZ2	15	40	6	61	LCZ2	0.92	1	0.92	LCZ2	13.8	40	5.52	59.32
LCZ3	20	5	50	75	LCZ3	0.83	0.92	1	LCZ3	16.6	4.6	50	71.2
Sum	85	55	61	201					Sum	80.4	53.8	59.67	193.87

$OA = (50 + 40 + 50)/201 = 0.696$
WA; our interpretation $= (50 + 40 + 50)/193.87 = 0.722$
WA; actual $= (50 + 9.2 + 4.14 + 13.8 + 40 + 5.52 + 16.6 + 4.6 + 50)/201 = 0.965$

We would like to thank the authors of Bechtel et al. [15] for providing more information on the WA metric in response to our previous study [11]. They clearly explained how WA could be calculated from the weighted confusion matrix, using several example confusion matrices to

demonstrate this. It is now possible for readers to better understand how WA is calculated and to reproduce the results. We were also very happy to find that their WA metric was not calculated illogically (i.e., not improperly applying greater penalties to misclassifications between more similar land use/land cover features), as we had previously reported [11]. Thus, we sincerely welcome this contribution of Bechtel et al. [15]. A few issues remain with the WA metric (e.g., the limited transparency of the point scheme used), but a critique of the metric is outside the scope of this current work.

3. Responses to the Claims of Bechtel et al. [15] Regarding the JJ19 Metric

3.1. Points of Disagreement

Although we acknowledge several contributions of Bechtel et al. [15], some aspects of the study were found to be inaccurate, including most of the claims regarding our own proposed map accuracy metric [11], which they termed the “JJ19” metric. We previously acknowledged some limitations of the JJ19 metric, and indeed *all* metrics have their own unique pros and cons. However, many of the claims made were untrue or unsubstantiated, and, as we explain, the JJ19 metric has several benefits for LCZ map accuracy assessment. Most notably, it provides great flexibility, allowing for local optimization of its parameters. The full details of the JJ19 metric are provided in Johnson and Jozdani [11].

The principal claims of Bechtel et al. [15] were that “the JJ19 paper was based on wrong assumptions” and “the JJ19 method is not as innovative as claimed”. The JJ19 paper had two main objectives. The first was to present the rationale for why LCZ map accuracy assessments should take into account the physical characteristics of each LCZ type. This argument, which had not been made before, was reiterated in Bechtel et al. [15], helping to confirm our point. The second purpose of the JJ19 paper was to present a transparent and adaptable approach for LCZ map accuracy assessment. For this, we presented the JJ19 metric, which was calculated based on the typical physical and land-cover properties of each LCZ type (e.g., the average building height, impervious surface fractional cover, and anthropogenic heat flux values of each LCZ type), according to the parameter values given in Stewart and Oke [13] (the paper that provided the basis for most subsequent LCZ studies). Additionally, we fully acknowledged that the parameter values for each LCZ type may vary from one geographic region to another, and the JJ19 metric was designed so that these values could be easily adjusted according to the local conditions. As was reported by Bechtel et al. [15], “...a weighted accuracy is always related to a specific purpose, and hence its appropriateness requires expert judgement”. The flexibility of our method allows expert judgement (local knowledge of LCZ characteristics) and/or local field measurements of these parameters to be easily incorporated, and thus fulfills this criteria as well. Finally, we followed standard protocol to calculate overall map accuracy (wOA) and producer’s/user’s accuracy (wPA and wPA) values from the weighted confusion matrix [10,14]. Although we never claimed that the JJ19 method was particularly innovative, for the above reasons we argue that it is both supported by sound assumptions and useful in practice. It is also clear that the JJ19 method differs substantially from the WA method.

Most other issues with the JJ19 metric noted by Bechtel et al. [15] were relatively minor. For example, they pointed out that the JJ19 metric shows less variation in weights than the WA metric because we did not normalize our final derived values to a 0–1 range. Although this is of course easily done, it is not clear whether or not it is beneficial. The JJ19 metric was also criticized of “failing basic requirements of a weighted accuracy scheme” because it produces a value of 0 if all reference samples are misclassified. Actually, it results in an undefined value (0/0) if all reference samples are misclassified, but we agree that the JJ19 metric should not be reported if all reference data is misclassified (and the LULC map should probably be improved before using it as the basis for any subsequent analysis).

3.2. Point of Agreement and Suggestion for Improvement of JJ19 Metric

There is, of course, the potential to further improve the JJ19 metric, e.g., by incorporating additional parameters or modifying the existing parameters. As correctly noted by Bechtel et al. [15], one potential problem with the metric is that it does not discriminate between water-dominant (LCZ G) and land-dominant LCZs (all other LCZs) as clearly as the WA metric does. To alleviate this, here we have suggested adding a new parameter containing the “water fractional cover” of each LCZ type.

Assuming all of the land-dominant LCZs contain roughly the same water fractional cover (e.g., ~10% or less), which is much less than that of LCZ G (e.g., ~75% or more water fractional cover), the normalized parameter values (P_{norm}) of LCZ 1-F and LCZ G correspond to 0 and 1, respectively (Supplementary Tables S1 and S2). The resultant LCZ class dissimilarity weights (D_{ij}) for our generic calculation approach are shown in Table 2. For lack of a better term, here we refer to this improved version of the JJ19 metric as JJ20. The full set of LCZ parameter values and equations necessary to calculate JJ20 are included in Supplementary Tables S1–S3, so the metric can be calculated automatically after the user inputs the values of an unweighted confusion matrix into Table S3. Users can also modify the parameters based on local field measurements or expert knowledge to generate location specific D_{ij} values.

Table 2. LCZ class dissimilarity weights (D_{ij}) for the JJ20 metric.

LCZ	1	2	3	4	5	6	7	8	9	10	A	B	C	D	E	F	G
1		0.26	0.30	0.23	0.36	0.42	0.42	0.43	0.52	0.35	0.34	0.54	0.59	0.60	0.63	0.69	0.82
2	0.26		0.10	0.18	0.15	0.22	0.24	0.25	0.32	0.25	0.24	0.34	0.38	0.40	0.41	0.49	0.61
3	0.30	0.10		0.21	0.18	0.14	0.17	0.17	0.23	0.23	0.27	0.27	0.30	0.32	0.41	0.41	0.53
4	0.23	0.18	0.21		0.15	0.19	0.32	0.26	0.29	0.28	0.25	0.31	0.35	0.37	0.44	0.46	0.58
5	0.36	0.15	0.18	0.15		0.07	0.30	0.14	0.17	0.15	0.28	0.19	0.24	0.25	0.29	0.34	0.47
6	0.42	0.22	0.14	0.19	0.07		0.25	0.08	0.10	0.13	0.31	0.16	0.17	0.18	0.28	0.27	0.40
7	0.42	0.24	0.17	0.32	0.30	0.25		0.28	0.27	0.37	0.29	0.27	0.24	0.30	0.46	0.34	0.52
8	0.43	0.25	0.17	0.26	0.14	0.08	0.28		0.13	0.19	0.40	0.24	0.21	0.19	0.24	0.28	0.40
9	0.52	0.32	0.23	0.29	0.17	0.10	0.27	0.13		0.19	0.29	0.13	0.12	0.08	0.25	0.17	0.32
10	0.35	0.25	0.23	0.28	0.15	0.13	0.37	0.19	0.19		0.38	0.26	0.28	0.27	0.33	0.36	0.49
A	0.34	0.24	0.27	0.25	0.28	0.31	0.29	0.40	0.29	0.38		0.21	0.28	0.33	0.48	0.38	0.58
B	0.54	0.34	0.27	0.31	0.19	0.16	0.27	0.24	0.13	0.26	0.21		0.09	0.16	0.34	0.17	0.40
C	0.59	0.38	0.30	0.35	0.24	0.17	0.24	0.21	0.12	0.28	0.28	0.09		0.09	0.26	0.10	0.31
D	0.60	0.40	0.32	0.37	0.25	0.18	0.30	0.19	0.08	0.27	0.33	0.16	0.09		0.18	0.09	0.24
E	0.63	0.41	0.41	0.44	0.29	0.28	0.46	0.24	0.25	0.33	0.48	0.34	0.26	0.18		0.20	0.33
F	0.69	0.49	0.41	0.46	0.34	0.27	0.34	0.28	0.17	0.36	0.38	0.17	0.10	0.09	0.20		0.23
G	0.82	0.61	0.53	0.58	0.47	0.40	0.52	0.40	0.32	0.49	0.58	0.40	0.31	0.24	0.33	0.23	

4. Conclusions

New metrics are required to describe the accuracy of land use/land cover maps for specific applications. Along these lines, Bechtel et al. [15] and Johnson and Jozdani [11] presented new metrics for conveying the accuracy at which local climate zone (LZC) maps (a specific type of land use/land cover map) depict the local thermal environment. Here, we responded to comments made by Bechtel et al. [15] and suggested an improved version of the Johnson and Jozdani [11] method (called “JJ20” here).

Bechtel et al. [15] correctly noted that we misinterpreted their weighted accuracy (WA) metric in Johnson and Jozdani [11]. We explained the reason for this misunderstanding, and noted that it could have been avoided if a confusion matrix had been provided to show how the WA metric was calculated in the original work [12]. Based on this experience, we would like to stress the importance of including the confusion matrix in any LULC mapping study (either as a table in the text or supplementary file) to help avoid these types of misunderstandings in the future.

Both the JJ20 method and the WA method presented by Bechtel et al. [15] were found to provide valid options for the task of LCZ map accuracy assessment, as they apply greater penalization to

misclassification between more physically dissimilar LCZ classes. We recommend future LCZ mapping studies to report both metrics in addition to traditional map accuracy metrics, such as overall accuracy, producer's accuracy, and user's accuracy. That said, further improvements to these new metrics, and other alternative metrics for this task, are still needed. Finally, similar approaches could be used to develop metrics that convey the accuracy of land use/land cover maps for other specific applications. We thank the journal *Remote Sensing* for providing us a platform for this discussion.

Supplementary Materials: The following are available online at www.mdpi.com/2072-4292/12/11/1771/s1, Table S1: Parameter values, Table S2: D_{ij} values, Table S3: JJ20 metric calculation.

Author Contributions: Conceptualization, B.A.J. and S.E.J.; methodology, B.A.J.; software, n/a; validation, B.A.J.; formal analysis, B.A.J. and S.E.J.; investigation, B.A.J. and S.E.J.; resources, B.A.J.; data curation, n/a; writing—original draft preparation, B.A.J.; writing—review and editing, B.A.J. and S.E.J.; visualization, B.A.J.; supervision B.A.J.; project administration, B.A.J.; funding acquisition, B.A.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Environment Research and Technology Development Fund (S-15-1(4) Predicting and Assessing Natural Capital and Ecosystem Services (PANCES)) of the Ministry of the Environment, Japan.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Maimaitiyiming, M.; Ghulam, A.; Tiyyip, T.; Pla, F.; Latorre-Carmona, P.; Halik, Ü.; Sawut, M.; Caetano, M. Effects of green space spatial pattern on land surface temperature: Implications for sustainable urban planning and climate change adaptation. *ISPRS J. Photogramm. Remote Sens.* **2014**, *89*, 59–66.
- Wang, R.; Cai, M.; Ren, C.; Bechtel, B.; Xu, Y.; Ng, E. Detecting multi-temporal land cover change and land surface temperature in Pearl River Delta by adopting local climate zone. *Urban Clim.* **2019**, *28*, 100455.
- Sinha, S.; Santra, A.; Das, A.K.; Sharma, L.K.; Mohan, S.; Nathawat, M.S.; Mitra, S.S.; Jeganathan, C. Accounting tropical forest carbon stock with synergistic use of space-borne ALOS PALSAR and COSMO-SkyMed SAR sensors. *Trop. Ecol.* **2019**, *60*, 83–93.
- Johnson, B.A.; Dasgupta, R.; Mader, A.D.; Scheyvens, H. Understanding national biodiversity targets in a REDD+ context. *Environ. Sci. Policy* **2019**, *92*, 27–33.
- Ministry of the Environment. Brazil's submission of a Forest Reference Emission Level (FREL) for Reducing Emissions from Deforestation in the Amazonia Biome for REDD+ Results-Based Payments under the UNFCCC from 2016 to 2020. Ministry of the Environment, Brazil: Brasilia, Brazil, 2018.
- Weng, Q. Modeling urban growth effects on surface runoff with the integration of remote sensing and GIS. *Environ. Manag.* **2001**, *28*, 737–748.
- Hong, Y.; Adler, R.F. Estimation of global SCS curve numbers using satellite remote sensing and geospatial data. *Int. J. Remote Sens.* **2008**, *29*, 471–477.
- Kussul, N.; Lavreniuk, M.; Shumilo, L.; Kolotii, A. Nexus Approach for Calculating SDG Indicator 2.4.1 Using Remote Sensing and Biophysical Modeling. In Proceedings of the IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 2 August 2019; pp. 6425–6428.
- Melchiorri, M.; Pesaresi, M.; Florczyk, A.J.; Corbane, C.; Kemper, T. Principles and applications of the global human settlement layer as baseline for the land use efficiency indicator—SDG 11.3.1. *ISPRS Int. J. Geo-Information* **2019**, *8*, 96.
- Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 35–46.
- Johnson, B.A.; Jozdani, S.E. Local climate zone (LCZ) map accuracy assessments should account for land cover physical characteristics that affect the local thermal environment. *Remote Sens.* **2019**, *11*, 2420.
- Bechtel, B.; Demuzere, M.; Sismanidis, P.; Fenner, D.; Brousse, O.; Beck, C.; Van Coillie, F.; Conrad, O.; Keramitsoglou, I.; Middel, A.; et al. Quality of Crowdsourced Data on Urban Morphology—The Human Influence Experiment (HUMINEX). *Urban Sci.* **2017**, *1*, 15.

13. Stewart, I.D.; Oke, T.R. Local climate zones for urban temperature studies. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 1879–1900.
14. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good Practices for Assessing Accuracy and Estimating Area of Land Change. *Remote Sens. Environ.* **2014**, *148*, 42–57.
15. Bechtel, B.; Demuzere, M.; Stewart, I.D. A Weighted Accuracy Measure for Land Cover Mapping: Comment on Johnson et al. Local Climate Zone (LCZ) Map Accuracy Assessments Should Account for Land Cover Physical Characteristics that Affect the Local Thermal Environment, *Remote Sens.* 2019, *11*, 2420. *Remote Sens.* **2020**, *12*, 1769.



© 2020 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 (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).