

Technical Note

Local Climate Zone (LCZ) Map Accuracy Assessments Should Account for Land Cover Physical Characteristics that Affect the Local Thermal Environment

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, Ontario K7L 3N6, Canada; shahab.jozdani@queensu.ca

* Correspondence: johnson@iges.or.jp

Received: 2 July 2019; Accepted: 16 October 2019; Published: 18 October 2019



Abstract: Local climate zone (LCZ) maps are increasingly being used to help understand and model the urban microclimate, but traditional land use/land cover map (LULC) accuracy assessment approaches do not convey the accuracy at which LCZ maps depict the local thermal environment. 17 types of LCZs exist, each having unique physical characteristics that affect the local microclimate. Many studies have focused on generating LCZ maps using remote sensing data, but nearly all have used traditional LULC map accuracy metrics, which penalize all map classification errors equally, to evaluate the accuracy of these maps. Here, we proposed a new accuracy assessment approach that better explains the accuracy of the physical properties (i.e., surface structure, land cover, and anthropogenic heat emissions) depicted in an LCZ map, which allows for a better understanding of the accuracy at which the map portrays the local thermal environment.

Keywords: map accuracy; urban heat island; urban climatology; microclimate; LULC mapping

1. Introduction

Mapping and analysis of local climate zones (LCZs), i.e., landscape units with distinct screen-height temperature regimes [1], has become a popular research topic in the fields of remote sensing and urban climatology in recent years. The LCZ classification scheme divides the landscape into 17 classes, each having different surface structure, land cover, and anthropogenic heat emission characteristics (hereafter “physical characteristics”) that affect the local air temperature [2]. LCZ maps have become an important source of data for studies on urban climatology [2–5] and urban planning [6]. In a more general context, the LCZ classification scheme has also been suggested as a standardized framework for the mapping of urban form at the global scale, as it can provide information on the basic physical properties of any urban area [7].

LCZ maps are typically generated by analyzing satellite images using semi-automated image classification algorithms. The most common approach involves performing supervised classification of Landsat satellite imagery using the random forest classification algorithm [8], and this approach has been termed the “world urban database and access portals tools” (WUDAPT) protocol [7]. The WUDAPT protocol has been applied successfully in many LCZ mapping studies in different geographic regions [4–6,9,10]. To build on the WUDAPT approach, some studies have proposed the use of additional remote sensing datasets, e.g., nighttime optical imagery [11], nighttime thermal infrared imagery [12], synthetic aperture radar data [9], or LiDAR data [13]. Other studies have proposed incorporating

ancillary GIS datasets in the image classification workflow, e.g., OpenStreetMap data [11,14], and/or the use of alternative classification algorithms (e.g., convolutional neural networks [11]).

No matter the input data or image classification method used, all LCZ maps contain some degree of classification errors. As with all land use/land cover (LULC) map products, it is important to quantify these errors, e.g., in terms of producer's accuracy (errors of omission), user's accuracy (errors of commission), and overall accuracy [15]. Traditional LULC map accuracy assessment metrics are calculated by comparing the estimated (i.e., mapped) LULC type with the reference (i.e., "ground truth") LULC at sample pixel locations, and making a binary decision as to whether each sample location is correctly/incorrectly classified [16]. This purely thematic approach which penalizes all mapping errors equally, however, does not allow a user to adequately interpret the map's accuracy for modeling specific physical processes (e.g., rainfall runoff [17] or urban microclimate). In the case of LCZ mapping, some LCZ types are quite similar to one another in terms of their physical characteristics (e.g., "LCZ 2: compact mid-rise" and "LCZ 3: compact low-rise"), and a misclassification between two physically similar LCZ types will not pose as great a problem for urban climate studies as a misclassification between two physically dissimilar LCZ types (e.g., "LCZ 1: compact high-rise" and "LCZ A: dense trees"). Evidence of this can be seen in Bechtel et al. [3], where, based on an analysis of LCZ maps and land surface temperature data from 50 cities around the world, LCZ types with more similar physical characteristics were found to have more similar surface urban heat island intensities. Similar findings were also reported in individual case studies [2,10,12].

Based on the above, it is clear that LCZ map accuracy assessments should be conducted in a way that takes into account the physical characteristics of each LCZ type, e.g., by applying a greater penalization to classification errors between more dissimilar LCZ types. To our knowledge, however, only one prior study has attempted to incorporate such an approach. Bechtel et al. [18] proposed to weight the traditional LULC map error matrix using a multiplier that accounts for the similarity between LCZ types (based on the properties of openness, height, surface cover, and thermal inertia). Their idea of weighting the error matrix has merit, but their proposed methodology had two main shortcomings: (1) the determined weights were subjective and not fully transparent (e.g., no equations provided to explain how they were derived); and (2) the weights incorrectly applied a higher penalty to the misclassification of more physically similar LCZ types. The first shortcoming limits the flexibility and transferability of their methodology, as it is not possible to develop locally-optimized weights for a specific study site (i.e., weights based on the typical physical properties of the LCZs in that specific study site) if no equations are provided. The second shortcoming results in the assignment of lower penalizations to map classification errors that have larger impacts on the predicted microclimate (i.e., to misclassifications of physically dissimilar LCZ types), which is the opposite of what is logical. In the literature related to LCZ mapping, most past studies have used the traditional LULC map error matrix to evaluate LCZ mapping accuracy [4,7,12,14], while a few studies have used the method proposed by Bechtel et al. [3,18]. Thus, it is critical to develop more suitable LCZ map accuracy assessment metrics.

Proposed Approach

The objective of this study is to provide an objective and generic method for LCZ map accuracy assessment, taking into account the typical physical characteristics of each LCZ type that affect the local thermal environment. Our approach involves weighting the traditional LULC map error matrix using a "LCZ dissimilarity metric", calculated based on the degree of dissimilarity between the classification result (i.e., mapped LCZ type) and the reference data in terms of the estimated surface geometry, surface cover, and thermal inertia characteristics. The values of this LCZ dissimilarity metric can range from 0 (no penalty for misclassification) to 1 (full penalty for misclassification, which is equivalent to value of the traditional error matrix). The weighted producer's/user's/overall accuracy measures that can be derived from our weighted error matrix indicate the level of similarity between the estimated and reference ("ground truth") landscape physical characteristics depicted in the map, which allows for a better understanding of the map's accuracy in portraying the local thermal environment.

Our proposed methodology is intended to be generic enough to apply in any study site, although site-specific information on the reference physical characteristics of each LCZ type could alternatively be used (in place of the values in Figure 1), if available.

2. Materials and Methods

2.1. Generating the LCZ Dissimilarity Metric

The proposed LCZ dissimilarity metric takes into account nine of the ten physical parameters presented in Stewart and Oke [1] as the basis for LCZ identification:

1. sky view factor (SV), i.e., fraction of sky hemisphere visible from ground level;
2. aspect ratio (AR), i.e., average height-to-width ratio of street canyons/building spacing/tree spacing;
3. mean building/tree height (H);
4. terrain roughness class (TR), based on Davenport et al.'s [19] terrain roughness classification scheme;
5. building surface fraction (BF);
6. impervious surface fraction (IF);
7. surface admittance (SA), i.e., capacity of surface to accept or release heat;
8. surface albedo (A); and
9. anthropogenic heat flux (AH).

The remaining parameter given in Stewart and Oke [1], pervious surface fraction, was not included for our calculations because it is inversely correlated with the impervious surface fraction parameter.

To calculate the LCZ dissimilarity metric, we first determined the average value of each parameter, for each LCZ type, based on the value ranges provided in Stewart and Oke [1] (see Figure 1 for all of these average values). To account for the different minimum–maximum ranges of the nine parameters, we normalized all of the derived average values to a 0–1 range using the following equation:

$$P_{norm} = (P - P_{min}) / (P_{max} - P_{min}), \quad (1)$$

where P is the average value of a parameter, P_{max} and P_{min} are the maximum and minimum P values of the parameter (among all LCZ types), and P_{norm} is the normalized P value. Table A1 shows the P_{norm} values calculated for each LCZ type.

As the second step, we calculated the average (absolute value) difference between the P_{norm} values of each pairwise combination of LCZ types (D_{ij}). D_{ij} is given by:

$$D_{ij} = (|SV_i - SV_j| + |AR_i - AR_j| + |H_i - H_j| + |TR_i - TR_j| + |BF_i - BF_j| + |IF_i - IF_j| + |SA_i - SA_j| + |A_i - A_j| + |AH_i - AH_j|) / 9, \quad (2)$$

where i and j indicate two LCZ types i and j ; and $SV, AR, H, TR, BF, IF, SA, A,$ and AH are the P_{norm} values of the parameters sky view factor, aspect ratio, mean building/tree height, terrain roughness, building surface fraction, impervious surface fraction, surface admittance, surface albedo, and anthropogenic heat flux, respectively. For “LCZ A: Dense trees” the surface admittance parameter is unknown [1], so this parameter was excluded from D_{ij} calculations involving “LCZ A” (and the numerator of Equation (2) was divided by 8 instead of 9). All of the resultant D_{ij} values are reported in Table 1, with higher values indicating a greater dissimilarity between two LCZ classes in terms of their surface structure, land cover, and anthropogenic heat flux properties. These D_{ij} values represent our proposed LCZ dissimilarity metric. From Table 1, it can be seen that the two most similar LCZ types are “LCZ 5: Open mid-rise” and “LCZ 6: Open low-rise” ($D_{ij} = 0.08$), while the two most dissimilar LCZ types are “LCZ 1: Compact high-rise” and “LCZ G: Water” ($D_{ij} = 0.80$).

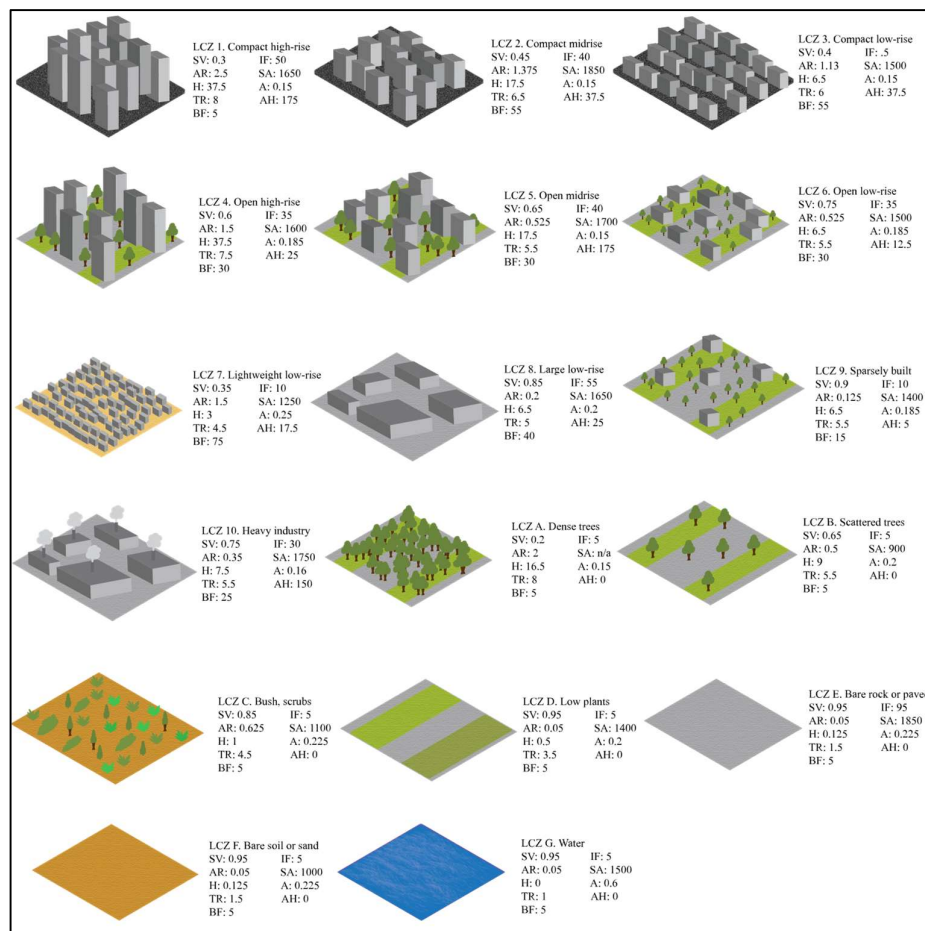


Figure 1. Graphical representation of each local climate zone (LCZ) type; and average values of their surface structure/land cover/anthropogenic heat flux parameters, as derived from the value ranges given in Stewart and Oke [1].

Table 1. D_{ij} value for each pairwise combination of local climate zone (LCZ) types. Higher D_{ij} values indicate that a higher penalty (i.e., multiplier) is applied to misclassifications between the corresponding LCZ types.

| LCZ Class | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | A | B | C | D | E | F | G |
|-----------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 1 | | 0.29 | 0.33 | 0.26 | 0.40 | 0.47 | 0.47 | 0.47 | 0.58 | 0.39 | 0.38 | 0.60 | 0.65 | 0.67 | 0.70 | 0.77 | 0.80 |
| 2 | 0.29 | | 0.11 | 0.19 | 0.17 | 0.24 | 0.27 | 0.27 | 0.35 | 0.28 | 0.27 | 0.38 | 0.43 | 0.44 | 0.45 | 0.54 | 0.57 |
| 3 | 0.33 | 0.11 | | 0.23 | 0.20 | 0.15 | 0.19 | 0.19 | 0.26 | 0.25 | 0.30 | 0.30 | 0.33 | 0.35 | 0.46 | 0.45 | 0.48 |
| 4 | 0.26 | 0.19 | 0.23 | | 0.17 | 0.21 | 0.35 | 0.28 | 0.32 | 0.31 | 0.28 | 0.35 | 0.39 | 0.41 | 0.49 | 0.51 | 0.54 |
| 5 | 0.40 | 0.17 | 0.20 | 0.17 | | 0.08 | 0.33 | 0.15 | 0.19 | 0.17 | 0.31 | 0.21 | 0.27 | 0.28 | 0.32 | 0.38 | 0.41 |
| 6 | 0.47 | 0.24 | 0.15 | 0.21 | 0.08 | | 0.28 | 0.09 | 0.11 | 0.15 | 0.35 | 0.18 | 0.19 | 0.20 | 0.31 | 0.30 | 0.33 |
| 7 | 0.47 | 0.27 | 0.19 | 0.35 | 0.33 | 0.28 | | 0.31 | 0.30 | 0.41 | 0.33 | 0.31 | 0.27 | 0.34 | 0.51 | 0.37 | 0.46 |
| 8 | 0.47 | 0.27 | 0.19 | 0.28 | 0.15 | 0.09 | 0.31 | | 0.14 | 0.21 | 0.45 | 0.26 | 0.23 | 0.21 | 0.26 | 0.31 | 0.34 |
| 9 | 0.58 | 0.35 | 0.26 | 0.32 | 0.19 | 0.11 | 0.30 | 0.14 | | 0.21 | 0.33 | 0.15 | 0.13 | 0.09 | 0.28 | 0.19 | 0.24 |
| 10 | 0.39 | 0.28 | 0.25 | 0.31 | 0.17 | 0.15 | 0.41 | 0.21 | 0.21 | | 0.43 | 0.29 | 0.31 | 0.30 | 0.36 | 0.40 | 0.43 |
| A | 0.38 | 0.27 | 0.30 | 0.28 | 0.31 | 0.35 | 0.33 | 0.45 | 0.33 | 0.43 | | 0.24 | 0.31 | 0.37 | 0.54 | 0.43 | 0.53 |
| B | 0.60 | 0.38 | 0.30 | 0.35 | 0.21 | 0.18 | 0.31 | 0.26 | 0.15 | 0.29 | 0.24 | | 0.10 | 0.18 | 0.38 | 0.18 | 0.33 |
| C | 0.65 | 0.43 | 0.33 | 0.39 | 0.27 | 0.19 | 0.27 | 0.23 | 0.13 | 0.31 | 0.31 | 0.10 | | 0.10 | 0.29 | 0.12 | 0.24 |
| D | 0.67 | 0.44 | 0.35 | 0.41 | 0.28 | 0.20 | 0.34 | 0.21 | 0.09 | 0.30 | 0.37 | 0.18 | 0.10 | | 0.20 | 0.10 | 0.15 |
| E | 0.70 | 0.45 | 0.46 | 0.49 | 0.32 | 0.31 | 0.51 | 0.26 | 0.28 | 0.36 | 0.54 | 0.38 | 0.29 | 0.20 | | 0.22 | 0.25 |
| F | 0.77 | 0.54 | 0.45 | 0.51 | 0.38 | 0.30 | 0.37 | 0.31 | 0.19 | 0.40 | 0.43 | 0.18 | 0.12 | 0.10 | 0.22 | | 0.15 |
| G | 0.80 | 0.57 | 0.48 | 0.54 | 0.41 | 0.33 | 0.46 | 0.34 | 0.24 | 0.43 | 0.53 | 0.33 | 0.24 | 0.15 | 0.25 | 0.15 | |

2.2. Using the LCZ Dissimilarity Metric to Weight the Traditional Error Matrix

A traditional LULC map error matrix is generated by systematically comparing the remote sensing-derived LULC map and the reference LULC dataset at sample pixel locations. The central part of the matrix is an array of number $k \times k$ in size, where k is equal to the number of LULC types in the map [15]. Rows in the error matrix represent the mapped LULC information at each sample location, columns represent the reference information, and the intersection of rows and columns summarize the number of samples mapped as a particular LULC type relative to the reference data [15]. From this error matrix, overall accuracy is calculated as the fraction of all samples that are correctly classified, while producer's accuracy is calculated for each LULC type as the number of correctly classified samples divided by the total number of reference samples belonging to that LULC type (i.e., column total), and user's accuracy is calculated for each LULC type as the number of correctly classified samples divided by the total number of samples mapped as that LULC type (i.e., row total).

A weighted error matrix, which accounts for the dissimilarity between each pair of LCZ types, can be derived by multiplying the values in the traditional error matrix by the corresponding LCZ dissimilarity metric (D_{ij}) values presented in Table 2. The result of this weighting is that misclassifications between two physically similar LCZ classes receive a lower penalty in the calculation of producer's accuracy/user's accuracy/overall accuracy, as these classification errors will have less impact on the thermal properties that can be estimated from the LCZ map. On the other hand, misclassifications between two dissimilar LCZ types are applied with a higher penalty (up to 1, i.e., complete misclassification, in the case of highly dissimilar LCZ types). For correctly classified sample pixels, no penalty is applied because no misclassification exists, and the values are equivalent to those of the traditional error matrix.

Table 2. Traditional error matrix generated using our synthetic dataset with 12 LCZ classes.

| LCZ Type | Reference Data | | | | | | | | | | | | Sum | UA |
|----------|----------------|------|------|------|------|------|------|------|------|------|------|------|--------|------|
| | 1 | 2 | 3 | 4 | 6 | 8 | 9 | A | B | D | F | G | | |
| 1 | 18 | 1 | 0 | 27 | 25 | 0 | 11 | 0 | 0 | 0 | 2 | 0 | 84 | 0.21 |
| 2 | 8 | 11 | 4 | 52 | 7 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 84 | 0.13 |
| 3 | 0 | 3 | 193 | 15 | 101 | 61 | 33 | 0 | 2 | 4 | 78 | 0 | 490 | 0.39 |
| 4 | 7 | 0 | 0 | 18 | 31 | 0 | 47 | 0 | 10 | 0 | 6 | 0 | 119 | 0.15 |
| 6 | 20 | 0 | 19 | 123 | 193 | 5 | 226 | 0 | 8 | 4 | 17 | 0 | 615 | 0.31 |
| 8 | 28 | 26 | 26 | 45 | 376 | 30 | 34 | 0 | 3 | 4 | 13 | 0 | 585 | 0.05 |
| 9 | 0 | 0 | 0 | 4 | 89 | 1 | 65 | 3 | 5 | 85 | 5 | 0 | 257 | 0.25 |
| A | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 1682 | 0 | 4 | 0 | 0 | 1688 | 1.00 |
| B | 1 | 1 | 0 | 9 | 33 | 0 | 48 | 14 | 5 | 76 | 12 | 0 | 199 | 0.03 |
| D | 0 | 0 | 0 | 0 | 175 | 0 | 15 | 3 | 0 | 592 | 18 | 4 | 807 | 0.73 |
| F | 1 | 0 | 12 | 3 | 171 | 14 | 20 | 0 | 0 | 48 | 26 | 0 | 295 | 0.09 |
| G | 0 | 0 | 0 | 1 | 10 | 0 | 3 | 0 | 0 | 0 | 0 | 4855 | 4869 | 1.00 |
| Sum | 83 | 42 | 254 | 297 | 1213 | 113 | 502 | 1702 | 33 | 817 | 177 | 4859 | 10,092 | |
| PA | 0.22 | 0.26 | 0.76 | 0.06 | 0.16 | 0.27 | 0.13 | 0.99 | 0.15 | 0.72 | 0.15 | 1.00 | | |

OA = 0.76

This proposed approach to derive a weighed error matrix—and corresponding weighted producer's accuracy/user's accuracy/overall accuracy values—is quite similar to the approach used to calculate “weighted accuracy” in Bechtel et al. [18], but the previous study incorrectly applied greater penalties for misclassification between physically similar LCZ types rather than physically dissimilar LCZ types.

3. Results and Discussion

3.1. Comparison of Our LCZ Dissimilarity Metric with a Reference and the Traditional Error Metric

Figure 2 shows several examples of how the misclassification penalties derived by our LCZ dissimilarity metric differ from those of Bechtel et al. [18] and the traditional error matrix. As can be

seen in the figure, compared to the metric of Bechtel et al. [18], our approach resulted in a lower penalty being applied to the misclassification of physically similar LCZ classes, and a higher penalty being applied to the misclassification of physically dissimilar LCZ classes. Thus, our weighted error matrix should give a better indication of how accurately a LCZ map depicts the local thermal environment.

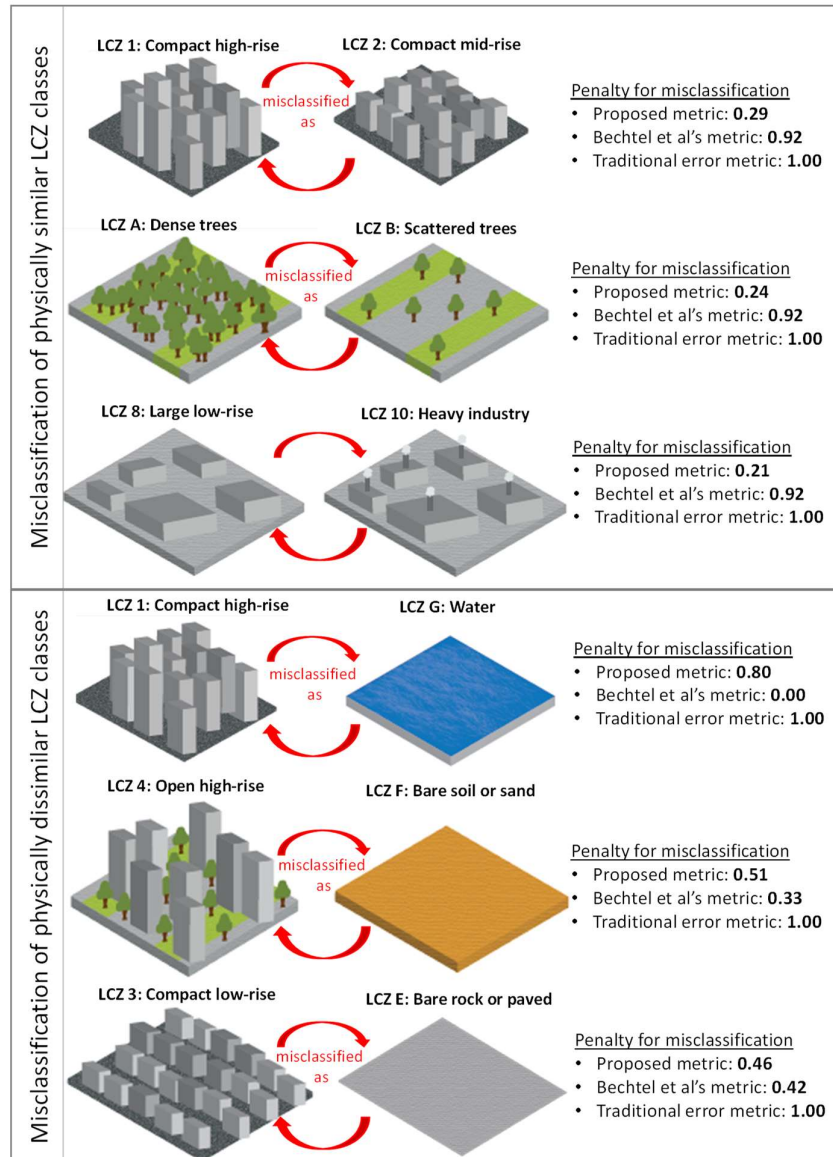


Figure 2. Comparison of misclassification penalties derived using our proposed LCZ dissimilarity metric, the metric proposed by Bechtel et al. [18], and the traditional error metric (i.e., equal penalty applied to all types of misclassifications).

3.2. Demonstration of Proposed Approach Using a Synthetic Dataset

In this section, we used a synthetic dataset to demonstrate the processes of generating a weighted error matrix, and calculating weighted overall/producer's/user's accuracy from the weighted error matrix. We elected to use a synthetic dataset here rather than real data to illustrate the metric calculations in a simpler way, i.e., without delving into explanations of the image data, training data, classification algorithm, etc., which would be needed if a real remote sensing dataset was used.

Table 2 shows a traditional (unweighted) error matrix for this synthetic dataset. The matrix contains 12 LCZ types, with 5 LCZ types assumed to not be present (missing LCZs is a common case, in practice). Table 3 shows the weighted error matrix, after the values in Table 2 have been

multiplied by the LCZ dissimilarity metric values in Table 1. From the traditional error matrix, the overall accuracy (OA) of the LCZ map is calculated to be 0.76, while producer's accuracy (PA) values range from 0.06–1.0 and user's accuracy (UA) values range from 0.03–1.0. For the weighted error matrix, the weighted overall accuracy (wOA) value is calculated as 0.92, while the weighted producer's accuracy (wPA) values range from 0.22–1.0 and the weighted user's accuracy (wUA) values range from 0.14–1.0. Comparing the cell values of the two error matrices, it is evident that the differences between the PA/ wPA values and UA/ wUA values were greater for LCZ types that had been mostly misclassified as other physically similar LCZ types (e.g., misclassifications among the urban LCZ types; LCZ 1–10). This is because the misclassification penalties (i.e., D_{ij} values) for these types of errors were relatively low under our proposed approach. From the comparison of the two matrices is also clear that the values in the weighted error matrix are always less than or equal to those of the traditional error matrix, and this is true because no pairwise combination of LCZs has a D_{ij} value > 1 .

Table 3. Weighted error matrix generated using our synthetic dataset with 12 LCZ classes.

| LCZ Type | Reference Data | | | | | | | | | | | | Sum | wUA |
|----------|----------------|-------|--------|-------|--------|-------|--------|---------|-------|--------|-------|---------|---------|--------------|
| | 1 | 2 | 3 | 4 | 6 | 8 | 9 | A | B | D | F | G | | |
| 1 | 18.00 | 0.29 | 0.00 | 6.96 | 10.05 | 0.00 | 5.20 | 0.00 | 0.00 | 0.00 | 1.40 | 0.00 | 41.90 | 0.43 |
| 2 | 2.30 | 11.00 | 0.43 | 10.12 | 1.16 | 0.54 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 25.55 | 0.43 |
| 3 | 0.00 | 0.32 | 193.00 | 3.45 | 20.05 | 11.46 | 6.38 | 0.00 | 0.61 | 1.34 | 35.67 | 0.00 | 272.28 | 0.71 |
| 4 | 1.80 | 0.00 | 0.00 | 18.00 | 5.22 | 0.00 | 13.32 | 0.00 | 2.85 | 0.00 | 2.95 | 0.00 | 44.15 | 0.41 |
| 6 | 8.04 | 0.00 | 3.77 | 20.72 | 193.00 | 1.64 | 34.88 | 0.00 | 2.48 | 1.08 | 5.51 | 0.00 | 271.11 | 0.71 |
| 8 | 13.19 | 7.06 | 4.88 | 15.79 | 123.26 | 30.00 | 10.50 | 0.00 | 0.99 | 1.09 | 6.69 | 0.00 | 213.45 | 0.14 |
| 9 | 0.00 | 0.00 | 0.00 | 1.13 | 13.73 | 0.31 | 65.00 | 0.62 | 2.25 | 19.52 | 1.32 | 0.00 | 103.89 | 0.63 |
| A | 0.00 | 0.00 | 0.00 | 0.00 | 0.34 | 0.00 | 0.00 | 1682.00 | 0.00 | 1.25 | 0.00 | 0.00 | 1683.59 | 1.00 |
| B | 0.38 | 0.27 | 0.00 | 2.56 | 10.23 | 0.00 | 21.64 | 6.03 | 5.00 | 23.82 | 6.49 | 0.00 | 76.43 | 0.07 |
| D | 0.00 | 0.00 | 0.00 | 0.00 | 47.22 | 0.00 | 3.44 | 0.94 | 0.00 | 592.00 | 5.22 | 0.46 | 649.28 | 0.91 |
| F | 0.70 | 0.00 | 5.49 | 1.48 | 55.39 | 7.20 | 5.28 | 0.00 | 0.00 | 13.92 | 26.00 | 0.00 | 115.45 | 0.23 |
| G | 0.00 | 0.00 | 0.00 | 0.51 | 3.76 | 0.00 | 0.92 | 0.00 | 0.00 | 0.00 | 0.00 | 4855.00 | 4860.19 | 1.00 |
| Sum | 44.42 | 18.93 | 207.57 | 80.73 | 483.42 | 51.15 | 166.55 | 1689.59 | 14.19 | 654.02 | 91.25 | 4855.46 | 8357.27 | |
| wPA | 0.41 | 0.58 | 0.93 | 0.22 | 0.40 | 0.59 | 0.39 | 1.00 | 0.35 | 0.91 | 0.28 | 1.00 | | |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | $wOA = 0.92$ |

3.3. Potential, Limitations, and Alternative Implementations of Proposed Approach

The weighted accuracy measures calculated using our proposed approach, wOA in particular, can serve as the basis for comparing multiple classified LCZ maps of the same study site—e.g., using different input datasets, different classification algorithms, and/or different classification variables—to identify the most accurate classified map. In this way, a LCZ map that best portrays the local thermal environment of a study site can be selected, allowing for more accurate simulations of urban microclimate to be conducted using other spatially-explicit models (e.g., the Weather Research Forecasting model [20]), and (ideally) for more climate-sensitive urban planning practices to be adopted.

One potential limitation of our approach, however, is that the values of the LCZ dissimilarity metric can only range from 0–1 (or more specifically, from 0.08–0.80 using our generic approach). The significance is that the weighted accuracy values cannot be lower than their unweighted counterparts, because the multiplier for misclassification is always assumed to be 1 in the traditional error matrix. Thus, our error matrix weighting approach is somewhat akin to the process of merging of LCZs into fewer, more general LULC categories (e.g., either “built-up” or “natural” [9,18]) for accuracy assessment, but our approach is more of a partial merging (unless $D_{ij} = 0$).

A second limitation of our approach is that some of the weighted accuracy metrics may be more difficult to interpret than the traditional metrics. wOA can be generally understood as the degree of accuracy at which the generated LCZ map estimates the physical characteristics of the landscape relevant to the local thermal environment. The wPA and wUA values, however, are more difficult to interpret and should be used with caution. Additionally, the LCZ dissimilarity metric weighted accuracy metrics are based the typical physical characteristics (surface structure, land cover, and anthropogenic heat emissions) of each LCZ type, as identified in Stewart and Oke [1]. In reality, however, these physical characteristics can vary from region to region [11]. Our approach is intended

to be generic enough to use anywhere, but the collection and usage of locally appropriate parameter values would be beneficial, when feasible.

Finally, we should point out that for some applications (e.g., LULC change assessment or urban planning), an LCZ map creator or map user may be concerned with both the map's thematic accuracy and the accuracy of the physical characteristics depicted in the map. Considering this, rather than simply use OA or w OA to evaluate an LCZ map's accuracy, an alternative would be to combine these two metrics, e.g., based on the mean average accuracy of the two metrics $[(OA + wOA)/2]$ or the harmonic mean (F1-measure) $[(2 \times OA \times wOA)/(OA + wOA)]$ [21]. In this way, multiple classified LCZ maps could be compared to identify one that achieves the desired trade-off between thematic and physical accuracy. Indeed, it is relatively common to interpret and/or combine multiple measures of accuracy in remote sensing studies, e.g., for image classification [18,22,23], image segmentation [24–26], or image enhancement [27]. In the case of our synthetic dataset, the mean average accuracy of the two metrics would be equal to 0.840, while the F1-measure would be equal to 0.837.

4. Conclusions

In this study, we presented a new approach for evaluating the accuracy of local climate zone (LCZ) maps. The proposed approach takes into account the dissimilarity between different LCZ types in terms of nine physical characteristics (sky view factor, aspect ratio, mean building/tree height, terrain roughness, building surface fraction, impervious surface fraction, surface admittance, surface albedo, and anthropogenic heat flux), and allows for a better understanding of the accuracy at which LCZ maps depict the local thermal environment of the landscape. We have presented a generic approach based on the typical physical parameters of each LCZ type (as identified in [1]), but it is also possible to optimize the approach for a specific location, considering the locally-appropriate physical characteristics of each LCZ type, by substituting locally-specific values for the values given in Figure 1. In contrast, the previous approach developed by Bechtel et al. does allow for this kind of optimization, as equations for calculating the error matrix weights are not provided. Although our proposed approach has some limitations, it is more suitable for evaluating the accuracy of LCZ maps than the traditional land use/land cover accuracy assessment approach that considers only map thematic accuracy, and is also more suitable than the existing physically-based LCZ map accuracy assessment approach (i.e., that proposed by Bechtel et al.).

Finally, although in this study we focused only on LCZ map accuracy assessment, there is a need to develop other application-specific LULC map accuracy assessment approaches, as LULC maps are now being used as input data to model various physical processes. Our proposed approach could be slightly modified for use in other application domains. For example, to better evaluate the accuracy of LULC maps used as inputs for hydrological models, it would be necessary to account for the dissimilarity between LULC classes in terms of their hydrological properties (e.g., typical rainfall runoff and/or groundwater infiltration rates of different LULC types) rather than their thermal properties.

Author Contributions: Conceptualization, B.A.J.; methodology, B.A.J. and S.E.J.; validation, B.A.J. and S.E.J.; formal analysis, B.A.J.; investigation, B.A.J.; resources, B.A.J.; data curation, B.A.J. and S.E.J.; writing—original draft preparation, B.A.J.; writing—review and editing, S.E.J.; visualization, B.A.J.

Funding: This research was supported 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.

Appendix A

Table A1. P_{norm} values for each LCZ type. Note: SV, AR, H, TR, BF, IF, SA, A, and AH columns show the P_{norm} values of the parameters sky view factor, aspect ratio, mean building/tree height, terrain roughness, building surface fraction, impervious surface fraction, surface admittance, surface albedo, and anthropogenic heat flux, respectively.

| LCZ Type | SV | AR | H | TR | BF | IF | SA | A | AH |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1 | 0.133 | 1.000 | 1.000 | 1.000 | 0.643 | 0.500 | 0.789 | 0.000 | 1.000 |
| 2 | 0.333 | 0.541 | 0.467 | 0.786 | 0.714 | 0.389 | 1.000 | 0.000 | 0.214 |
| 3 | 0.267 | 0.439 | 0.173 | 0.714 | 0.714 | 0.333 | 0.632 | 0.000 | 0.214 |
| 4 | 0.533 | 0.592 | 1.000 | 0.929 | 0.357 | 0.333 | 0.737 | 0.078 | 0.143 |
| 5 | 0.600 | 0.194 | 0.467 | 0.643 | 0.357 | 0.389 | 0.842 | 0.078 | 0.071 |
| 6 | 0.733 | 0.194 | 0.173 | 0.643 | 0.357 | 0.333 | 0.632 | 0.078 | 0.071 |
| 7 | 0.200 | 0.592 | 0.080 | 0.500 | 1.000 | 0.056 | 0.368 | 0.222 | 0.100 |
| 8 | 0.867 | 0.061 | 0.173 | 0.571 | 0.500 | 0.556 | 0.632 | 0.111 | 0.143 |
| 9 | 0.933 | 0.051 | 0.173 | 0.643 | 0.143 | 0.056 | 0.526 | 0.078 | 0.029 |
| 10 | 0.733 | 0.122 | 0.200 | 0.643 | 0.286 | 0.278 | 0.895 | 0.022 | 0.857 |
| A | 0.000 | 0.796 | 0.440 | 1.000 | 0.000 | 0.000 | n/a | 0.000 | 0.000 |
| B | 0.600 | 0.184 | 0.240 | 0.643 | 0.000 | 0.000 | 0.000 | 0.111 | 0.000 |
| C | 0.867 | 0.235 | 0.027 | 0.500 | 0.000 | 0.000 | 0.211 | 0.167 | 0.000 |
| D | 1.000 | 0.000 | 0.013 | 0.357 | 0.000 | 0.000 | 0.526 | 0.111 | 0.000 |
| E | 1.000 | 0.000 | 0.003 | 0.071 | 0.000 | 1.000 | 1.000 | 0.167 | 0.000 |
| F | 1.000 | 0.000 | 0.003 | 0.071 | 0.000 | 0.000 | 0.105 | 0.278 | 0.000 |
| G | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.632 | 1.000 | 0.000 |

References

1. Stewart, I.D.; Oke, T.R. Local climate zones for urban temperature studies. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 1879–1900. [\[CrossRef\]](#)
2. Stewart, I.D.; Oke, T.R.; Krayenhoff, E.S. Evaluation of the “local climate zone” scheme using temperature observations and model simulations. *Int. J. Climatol.* **2014**, *34*, 1062–1080. [\[CrossRef\]](#)
3. Bechtel, B.; Demuzere, M.; Mills, G.; Zhan, W.; Sismanidis, P.; Small, C.; Voogt, J. SUHI analysis using Local Climate Zones—A comparison of 50 cities. *Urban Clim.* **2019**, *28*, 100451. [\[CrossRef\]](#)
4. 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. [\[CrossRef\]](#)
5. Wang, R.; Ren, C.; Xu, Y.; Lau, K.K.L.; Shi, Y. Mapping the local climate zones of urban areas by GIS-based and WUDAPT methods: A case study of Hong Kong. *Urban Clim.* **2018**, *24*, 567–576. [\[CrossRef\]](#)
6. Perera, N.G.R.; Emmanuel, R. A “Local Climate Zone” based approach to urban planning in Colombo, Sri Lanka. *Urban Clim.* **2018**, *23*, 188–203. [\[CrossRef\]](#)
7. Bechtel, B.; Alexander, P.; Böhner, J.; Ching, J.; Conrad, O.; Feddema, J.; Mills, G.; See, L.; Stewart, I. Mapping Local Climate Zones for a Worldwide Database of the Form and Function of Cities. *ISPRS Int. J. Geo Inf.* **2015**, *4*, 199–219. [\[CrossRef\]](#)
8. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
9. Demuzere, M.; Bechtel, B.; Middel, A.; Mills, G. Mapping Europe into local climate zones. *PLoS ONE* **2019**, *14*, e0214474. [\[CrossRef\]](#)
10. Shi, Y.; Lau, K.K.L.; Ren, C.; Ng, E. Evaluating the local climate zone classification in high-density heterogeneous urban environment using mobile measurement. *Urban Clim.* **2018**, *25*, 167–186. [\[CrossRef\]](#)
11. Qiu, C.; Schmitt, M.; Mou, L.; Ghamisi, P.; Zhu, X.X. Feature importance analysis for local climate zone classification using a residual convolutional neural network with multi-source datasets. *Remote Sens.* **2018**, *10*, 1572. [\[CrossRef\]](#)
12. Cai, M.; Ren, C.; Xu, Y.; Lau, K.K.L.; Wang, R. Investigating the relationship between local climate zone and land surface temperature using an improved WUDAPT methodology—A case study of Yangtze River Delta, China. *Urban Clim.* **2018**, *24*, 485–502. [\[CrossRef\]](#)

13. Zhao, C.; Jensen, J.; Weng, Q.; Currit, N.; Weaver, R. Application of airborne remote sensing data on mapping local climate zones: Cases of three metropolitan areas of Texas, U.S. *Comput. Environ. Urban Syst.* **2019**, *74*, 175–193. [[CrossRef](#)]
14. Fonte, C.C.; Lopes, P.; See, L.; Bechtel, B. Urban Climate Using OpenStreetMap (OSM) to enhance the classification of local climate zones in the framework of WUDAPT. *Urban Clim.* **2019**, *28*, 100456. [[CrossRef](#)]
15. Jensen, J.R. *Introductory Digital Image Processing—A Remote Sensing Perspective*, 3rd ed.; Pearson Prentice Hall: Upper Saddle River, NJ, USA, 2005.
16. Congalton, R.G. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*; CRC Press: Boca Raton, FL, USA, 2009.
17. Johnson, B.A.; Endo, I.; Onishi, A.; Bragais, M.; Magcale-Macandog, D.B. A Land cover map accuracy metric for hydrological studies. In *Proceedings of the 16th World Lake Conference*; Maghfiroh, M., Dianto, A., Jasalesmana, T., Melati, I., Samir, O., Kurniawan, R., Eds.; Research Center for Limnology, Indonesian Institute of Sciences: Bali, Indonesia, 2017; pp. 191–195.
18. 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. [[CrossRef](#)]
19. Davenport, A.G.; Grimmond, C.S.B.; Oke, T.R.; Wieringa, J. Estimating the roughness of cities and sheltered country. In *Proceedings of the 12th Conference on Applied Climatology*; American Meteorological Society, Boston: Asheville, NC, USA, 2000; pp. 96–99.
20. Brousse, O.; Martilli, A.; Foley, M.; Mills, G.; Bechtel, B. WUDAPT, an efficient land use producing data tool for mesoscale models? Integration of urban LCZ in WRF over Madrid. *Urban Clim.* **2016**, *17*, 116–134. [[CrossRef](#)]
21. Witten, I.H.; Frank, E.; Hall, M. *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed.; Morgan Kaufmann: Burlington, VT, USA, 2011; Volume 31.
22. Gautam, S.; Li, Y.; Johnson, T.G. Do alternative spatial healthcare access measures tell the same story? *GeoJournal* **2014**, *79*, 223–235. [[CrossRef](#)]
23. Johnson, B.A.; Scheyvens, H.; Shivakoti, B.R. An ensemble pansharpening approach for finer-scale mapping of sugarcane with Landsat 8 imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *33*, 218–225. [[CrossRef](#)]
24. Zhang, X.; Feng, X.; Xiao, P.; He, G.; Zhu, L. Segmentation quality evaluation using region-based precision and recall measures for remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2015**, *102*, 73–84. [[CrossRef](#)]
25. Johnson, B.A.; Bragais, M.; Endo, I.; Magcale-Macandog, D.B.; Macandog, P.B.M. Image Segmentation Parameter Optimization Considering Within- and Between-Segment Heterogeneity at Multiple Scale Levels: Test Case for Mapping Residential Areas Using Landsat Imagery. *ISPRS Int. J. Geo Inf.* **2015**, *4*, 2292–2305. [[CrossRef](#)]
26. Clinton, N.; Holt, A.; Scarborough, J.; Yan, L.; Gong, P. Accuracy assessment measures for object-based image segmentation goodness. *Photogramm. Eng. Remote Sens.* **2010**, *76*, 289–299. [[CrossRef](#)]
27. Amro, I.; Mateos, J.; Vega, M.; Molina, R.; Katsaggelos, A.K. A survey of classical methods and new trends in pansharpening of multispectral images. *EURASIP J. Adv. Signal Process.* **2011**, *79*, 1–22. [[CrossRef](#)]

