Coherence of Global Hydroclimate Classification Systems

K. L. McCurley Pisarello[1], J.W. Jawitz[1]

[1]Soil and Water Sciences Department, University of Florida, Gainesville, Florida 32611, USA

Correspondence to: James W. Jawitz (jawitz@ufl.edu)

Abstract. Climate classification systems are useful for investigating future climate scenarios, water availability, and even socioeconomic indicators as they relate to climate dynamics. These classification systems typically utilize various forms of water and energy indicators to create zone boundaries. However, there has yet to be a classification framework that includes evapotranspiration (ET) rates as a governing principle, nor has there been an effort to simultaneously compare the structure and function of multiple existing classification schemes. Here, we developed three new classification systems based on ET rates and one new system based on precipitation and potential evapotranspiration, and we compared these four new systems against four previously established climate classification systems. The within-zone similarity, or coherence, of long-term water budget components was evaluated for each system based on the premise that the application of a climate classification framework should correspond to those variables that are most coherent. Additionally, the complexity of zone boundaries in each system was assessed. The most frequently used system, Köppen-Geiger, had high hydroclimate coherence but also high spatial complexity. This study produced classification systems of improved coherence for individual water budget components, lower spatial complexity, and fewer parameters needed for their construction. The Water-Energy Clustering classification system is the primary framework proposed here for future investigations in which regions of interest include zones of differing hydrologic dynamics.

1 Introduction

A variety of classification schemes have been introduced to categorize specific biophysical characteristics of Earth systems, including those based on climatic behavior (Beck et al., 2018; Berghuijs and Woods, 2016; Holdridge, 1967), biodiversity (Olson et al., 2001), plant-climate interactions (Papagiannopoulou et al., 2018), or plant hardiness (Magarey et al., 2008; McKenney et al., 2007). These frameworks classify elements of a system based on common atmospheric or terrestrial characteristics to maximize their within-zone similarity, or coherence, which allows for a transfer of understanding across zones of similar attributes (Lanfredi et al., 2019). This study focuses specifically on climate classification schemes, which have provided a climatic context for a variety of applications, including socioeconomic assessments of human health conditions (Boland et al., 2017; Jagai et al., 2007; Lloyd et al., 2007), economic development (Mellinger et al., 2000; Richards et al., 2019), and evaluating anticipated biophysical and climatic changes (Chen and Chen, 2013; Tapiador et al., 2019).

Different climate classification systems have emerged based on framework-specific suites of hydroclimatic variables used to define the climate zone boundaries. Therefore, users should consider how potential classification system application corresponds to the variables used to create it (Knoben et al., 2018; Meybeck et al., 2013). Climate classification systems are usually based in part on annual and seasonal water-energy budgets (Beck et al., 2018; Berghuijs and Woods, 2016; Holdridge, 1967; Knoben et al., 2018; Meybeck et al., 2013). The Köppen-Geiger (KPG) classification system, the most widely used climate framework (Peel et al., 2007), was developed to regionalize climatic variables (specifically accounting for seasonal precipitation and temperature) and is often employed to compare the output of global climate models (Peel et al., 2007; Tapiador et al., 2019). Another common system is the Holdridge Life Zones scheme, which was created to classify land area with respect to vegetation and soil (Holdridge, 1967). This system subdivides zones based on thresholds of annual precipitation (P), potential evapotranspiration (PET), biotemperature (growing season length and temperature), and latitude and altitude.

Recent work has extended climate classification frameworks to specifically encompass hydrological attributes, since water resources-based analyses should take place within relevant hydrologic boundaries (Knoben et al., 2018; Meybeck et al., 2013). For example, Meybeck et al (2013) proposed a global zoning system that was primarily based on the mean temperatures and average locally generated runoff (Q) of river basins. They compared the resulting boundaries against the Köppen-Geiger and Holdridge frameworks to assess zone overlaps. Those authors also evaluated the within-zone coherence of mean annual temperature, P, and Q, concluding that the latter two were most coherent in dry zones and least coherent in equatorial zones, while temperature was most coherent in equatorial zones. However, Meybeck et al. (2013) did not compare their zone coherence to that of previously established systems. Similarly, Knoben et al. (2018) formed zone boundaries based on climate indices (average aridity, seasonality of aridity, and P as snow) with the objective of minimizing within-zone Q variability (i.e., maximizing Q coherence). Those authors compared their results to the Köppen-Geiger framework and found theirs to be more coherent with respect to Q, but they did not evaluate other water budget components or other climate classification systems.

Although the P and Q components of the long-term water budget have been extensively considered in climate classification schemes (Beck et al., 2018; Berghuijs and Woods, 2016; Holdridge, 1967; Knoben et al., 2018; Meybeck et al., 2013), notably absent is a system that is directly based on actual evapotranspiration (ET) rates. This gap is likely because ET traditionally has been the least empirically identified element of regional to global water budgets (Zhang et al., 2016). Moreover, there has been no comparison of within-zone hydroclimate coherence across climate classification systems, with evaluation particularly lacking in considering ET rates. Furthermore, the spatial complexity of climate classification systems has not been systematically examined. Assessing the spatial structure of a biophysical system is a concept that most notably originates from landscape ecology (O'Neill et al., 1988), which provides a suite of shape metrics that can be cross-disciplinarily applied. Quantifying shape pattern and spatial configuration of climate classification systems is important for understanding the interactions between governing hydroclimatic characteristics.

This work seeks to provide empirical support for application-dependent selection among possible climate classification systems. We suggest that a classification system should have high within-zone coherence for variables that are related to the system's intended use, as well as relatively low shape complexity across zones, which is useful for ease of

interpretation within management and policy contexts. As such, an overarching hypothesis was postulated that for a given climate classification system, within-zone hydrologic coherence and inter-zone shape complexity will be closely related to the organizing principle of that system. For example, the Köppen-Geiger and Meybeck et al. (2013) systems are based in large part on P and Q, respectively, and therefore these systems should show high coherence for these variables. Similarly, zone shape complexity will be lower in classification systems that include spatial contiguity in the organizing criteria (e.g., Meybeck et al., 2013). Given the major gap regarding the inclusion of ET in climate classification systems, we also propose a series of ET-based global classifications that should yield comparatively higher ET coherence than other systems.

We tested our hypothesis by evaluating within-zone coherence of long-term water budget components (mean annual ET, P, and Q) and synchronous P and PET seasonality, as well as zone shape complexity for our four new global classification systems, and further compared these against four previously established systems (Beck et al., 2018; Holdridge, 1967; Knoben et al., 2018; Meybeck et al., 2013). The primary zone shape complexity metrics were zone area and zone fragmentation (i.e., number of patches comprising each zone). This work presents novel approaches to determine appropriate applications and boundary complexities of classification frameworks. Understanding the relevance of a climate classification system is important since such frameworks are used in multi-disciplinary contexts to examine hydrological, ecological, and societal phenomena.

2 Methods

2.1 Coherence and complexity metrics

Variable coherence is defined by within-zone variability, represented by the intra-zone coefficient of variation (CV) of the hydroclimate variable of interest. Lower CV corresponds to higher coherence, meaning zone delineation bounded more spatially homogenous hydroclimate rates and thereby capturing more hydroclimatically similar regions. An important component of this analysis is the evaluation of the tradeoff between hydroclimate coherence and the spatial complexity of zone boundaries. Building more precise boundaries may better delineate hydroclimate processes, but overly precise geographic specificity may compromise ease of interpretation, communication, and relevant application for management purposes (Knoben et al., 2018).

Classification system complexity metrics were based on three principles: 1) Classification systems should consist of a relatively even distribution of pixels across zones, avoiding disproportionately large or small zones, 2) Zones should be as hydrologically continuous as possible (Meybeck et al., 2013), minimizing patchiness or fragmentation, and 3) Classification systems should comprise less than or equal to the number of zones in the KPG framework, which is used here as the standard to which other systems are compared. Therefore, complexity was assessed based on the inter-zone distribution of the number of pixels (zone area evenness, CV_z) and the number of patches in each zone (zone fragmentation). Note that the subscript *z* is added to differentiate between-zone complexity from the above metrics which emphasize intra-zone coherence. The number of patches is the only primary coherence or complexity metric in which CV is not used, since the objective here was to minimize

the degree of fragmentation, and not the similarity of fragmentation across zones. The number of patches was determined using the R function ClassStat in package *SDMTools* (VanDerWal et al., 2019). For each hydroclimate and complexity variable, differences between the KPG framework and the other classification systems were determined based on the Kolmogorov-Smirnov (K-S) test.

2.2 Database construction

We evaluated global gridded monthly P and PET and mean annual ET and Q between 1980 and 2018 at a 0.5° x 0.5° spatial resolution. The Climate Research Unit TimeSeries V4.04 supplied monthly P and PET (Harris et al., 2020), while mean annual ET and Q were constructed from aggregated TerraClimate monthly data (Abatzoglou et al., 2018) by summing the long-term mean monthly values. Long-term mean values were used to mute interannual variability. Annual ET and Q were resampled from their original 1/24° x 1/24° resolution to the 0.5° x 0.5° resolution of P and PET. Spatial analysis R packages *raster* (Hijmans, 2017), *sp* (Bivand et al., 2013) and *ncdf4* (Pierce, 2017) were used to build the database of long-term monthly and annual averages. The spatial extent of this study comprised all global land areas, excluding Antarctica, which resulted in a total of 61,701 pixels. To ensure a fair comparative assessment, all metrics were evaluated after stacking variable and classification system layers.

2.3 Sinusoidal functions as a descriptor of seasonality

The seasonal dynamics of monthly P and PET were also considered in this analysis, as they are also included in the Köppen-Geiger framework, which considers temperature as a general proxy for PET (Beck et al., 2018). Sine functions, and their corresponding parameters, can be used to describe intra-annual climate behavior. Sine functions were fitted to the long-term monthly distribution (following Berghuijs and Woods, 2016)

$$y_m = \bar{y}\left[1 + r_y \sin\left(\frac{2\pi(m-\theta_y)}{12}\right)\right] \quad (1)$$

where y is P or PET (mm month^{-1}) for each month, m, with overall mean denoted by the overbar, and r is dimensionless amplitude. Phase angle, θ, is the offset (months) from the reference time, January ($m = 1$), with absolute value of phase difference $|\Delta\theta| \leq 6$). Phase difference, Δθ, describes the synchronization of P and PET throughout the year as

$$\Delta\theta = \begin{cases} \theta_{PET} - \theta_P, & -6 \leq \theta_{PET} - \theta_P \leq 6 \\ \theta_{PET} - \theta_P - 12, & \theta_{PET} - \theta_P > 6 \\ \theta_{PET} - \theta_P + 12, & \theta_{PET} - \theta_P < 6 \end{cases} \quad (2)$$

Equation 1 showed overall good fits to the long-term mean monthly distributions of P and PET, with R^2=0.67±0.28 and 0.85±0.17, respectively (mean±standard deviation across all pixels). Sine fits to monthly PET were statistically significant (p-value ≤ 0.05) in 98% of pixels, while fits to monthly P were statistically significant in 89% of pixels. Similar to Berghuijs and Woods (2016), P fits were best in South America, and not as good in parts of the Sahara. Our fits were good in East Asia and not as good in the southern United States, while Berghuijs and Woods (2016) had more error in East Asia and less error in the United States. Lastly, the performance of our PET fits was overall much more spatially homogenous than their temperature fits.

To constrain $-6 \leq \Delta\theta \leq 6$, 12 was either added to or subtracted from Δθ values outside of these bounds (e.g., $\Delta\theta =$ 8 months is translated to -4 months). Because a constant reference time of January does not describe the water year for each zone, the Δθ distributions were normalized by centering around the mode and correcting to contain only positive values (0 to 12 months).

2.4 Established climate classification systems

Four previously established climate classification schemes were assessed in this analysis. We included two legacy schemes, Köppen-Geiger (KPG, Beck et al., 2018) and Holdridge Life (HDL, Holdridge, 1967) zoning systems, and two recently proposed frameworks, here referred to as Meybeck Hydroregion (MHR, Meybeck et al., 2013) and Knoben Hydroclimate (KHC, Knoben et al., 2018) systems. Note that the original KHC zones created by Knoben et al. (2018) were not delineated by discrete boundaries but were instead represented as a probability continuum of belonging to a zone. In the present study, KHC boundaries were re-created using a clustering approach similar to that of Knoben et al., 2018, in which they delineated 18 zones using their provided climate indices (aridity index, seasonal aridity index, and precipitation as snow). Here we applied a k-means, multi-start clustering method (n=80 starts), which was also used to form boundaries in two of our proposed frameworks described below. This k-means clustering approach was based on the Hartigan and Wong (1979) algorithm was employed using the kmeans function in the R package *stats* (R Core Team, 2018). Note that the very small KPG zones "Csc" and "Cwc" did not appear in the 0.5° x 0.5° resolution KPG output created by Beck et al. (2018) that was used in this study, resulting in 28 KPG presently analyzed zones. As in other climate classification studies (Knoben et al., 2018; Meybeck et al., 2013), KPG was considered here to be the standard to which other systems are primarily compared and evaluated for performance.

2.5 Proposed univariate ET climate classification systems

This study establishes and verifies ET-relevant climate classification frameworks by creating zones primarily based on ET rates and comparing ET coherence between systems. Three of the four systems proposed in this study were univariate (formatted from global mean annual ET rates) and uni-conditional (incorporating an additional system-dependent single condition). The additional conditions were included to emphasize a specific optimization goal.

The first two proposed univariate classification systems were based on the global ET empirical cumulative distribution function (CDF) and sought to minimize within-zone ET variability. The first classification system, ET Area-optimizing (ETA), was created with the objective of having an equal number of pixels in each ET-based zone. This was motivated by the first complexity principle described in Section 2.1, which states zones should not be meaninglessly small nor disproportionately large. The KPG system has relatively high spatial non-uniformity (Figure 1), resulting in highly variable relevance for regional analyses. A classification system that is more spatially uniform could more easily inform large spatial scale understanding as well as the application of regional to semi-continental management strategies. Additionally, it is useful to have a simple baseline framework upon which to compare the other ET-based systems. In this case, ETA is a system that maximizes area efficiency. This type of spatial condition is similar to the prioritizations of the MHR framework that state zones should ideally be "delineated in one piece" (Meybeck et al., 2013). The cumulative probability interval [0,1] was divided into 15 equal parts, and the corresponding upper and lower bounds of ET thresholds for each zone were determined from the CDF of mean annual ET for all global land pixels (Figure S1-A). The number of ETA zones was chosen based on the number of zones in previously established systems, the relative improvement of ET coherence with the addition of more zones (Figure S1-B), and the objective of having equal or fewer zones than the standard KPG framework.

The second proposed classification system, ET Variability-optimizing (ETV), was based on the principle of maximizing within-zone ET coherence subject to the tradeoffs with increasing complexity by adding zones. By fitting the empirical CDF with a continuous distribution, zone boundaries can be determined analytically for the minimum desired CV_{min}. For simplicity, and also supported by empirical evidence (Figure S2), we fitted a uniform distribution, which is characterized by lower and upper bounds a and b, with $\mathrm{CV} = (b-a)/[\sqrt{3}(b+a)]$. The ET limits defining each zone, i, were then determined directly from this relation as

$$a_i = \frac{b_i(1-\sqrt{3}\mathrm{CV}_{\mathrm{min}})}{1+\sqrt{3}\mathrm{CV}_{\mathrm{min}}} \quad (3)$$

where the upper and lower limits of sequential zones are shared (i.e., $b_{i-1} = a_i$). The largest value of b = 1,454 mm yr^{-1} was based on the maximum ET for all pixels, and CV_{min} = 0.075 was chosen based on marginal CV decrease with increasing number of zones (Figure S2-B), which resulted in 29 zones. This method produces nearly equal CV in all zones. Corresponding ET limits for each zone are shown in Figure S2-A.

The third univariate scheme proposed here is the ET Clustering (ETC) classification system, in which the k-means clustering approach was applied. This is justified by previous analyses that have used clustering techniques for climate classification purposes (Knoben et al., 2018; Tapiador et al., 2019). Zones were built using a multi-start framework (n=80 starts) by forming clustering centers iteratively until the within-zone sum of squares of mean annual ET, based on Euclidean distances, was reduced. This method encompasses aspects of both ETA and ETV, in which ET variability and area distribution are considered. The ETC approach also compares a clustering methodology against analytical ET-based zoning frameworks.

The final number of clustering centers (i.e., zones) was 20, which was chosen because it is the fewest number of zones with CV of mean annual ET below a low threshold, selected here as 0.1 (Figure S3).

2.6 Proposed multivariate climate classification system

The final proposed system in this study is a multivariate climate clustering framework, which was created from the same k-means clustering method as the ETC framework. This new climate classification system included two hydroclimate variables (mean annual P and PET) and was designed for comparison against the univariate ET classification frameworks, as well as previously established systems that were similarly formed with multiple variables. This system was selected from a suite of candidate multivariate systems that was generated using mean annual P, PET, $\Delta\theta$, ET, and Q. These candidate classification systems were evaluated based on water budget coherence (mean CV of ET, P, and Q) and zone complexity (number of zones, number of input variables, pixel distribution, and mean zone patchiness). Since KPG is the standard framework to which other systems are compared, these potential candidate multivariate systems were eliminated if their complexity metrics exceeded mean KPG complexity (excluding mean number of patches, in which a 50% threshold was allowed) or if their water budget coherence metrics were more than 50% greater than KPG water budget coherence (Table 1). Some allowance for reduced water budget coherence was made in favor of reduced complexity. Here, the tradeoffs between precise hydroclimate bounding (i.e., maximum coherence) and minimizing zone shape complexity (e.g., fewest possible number of zones) are assessed for efficient balance. The values for these elimination thresholds are listed in Supporting Information, and the resulting eligibility of the full suite of multivariate climate classification systems is shown in Table S1, with the final two candidate multivariate systems in Table S2.

The final two compared systems required only two input variables and met all coherence and complexity elimination thresholds (Table S1). Ultimately, the final system was chosen primarily based its relatively higher mean water budget coherence (CV of ET, P, and Q), and the fewest number of zones within that system was selected (n=22, Table S2). Therefore, the climate classification system formed from clustering mean annual P and mean annual PET was the representative multivariate framework proposed here, and is herein referred to as the Water-Energy Clustering (WEC) climate classification system.

3 Results

This study compared four previously established climate classification systems (KPG, HDL, MHR, KHC) and four potential new climate classification systems (ETA, ETV, ETC, and WEC) to assess for hydroclimate coherence as well as zone boundary complexities. The coherence and complexity metrics for the KPG system, the standard used in this study, are shown in Figure 1 for all zones. The KPG system had relatively high hydroclimate coherence, as all zones had $CV < 1$ for all variables except Q (Figure 1). However, the KPG spatial complexity was also relatively high. Zones in the KPG system had high

variability in the number of pixels, with several Boreal zones less than 5% the size of the largest zone, polar tundra zone "ET" (Figure 1E).

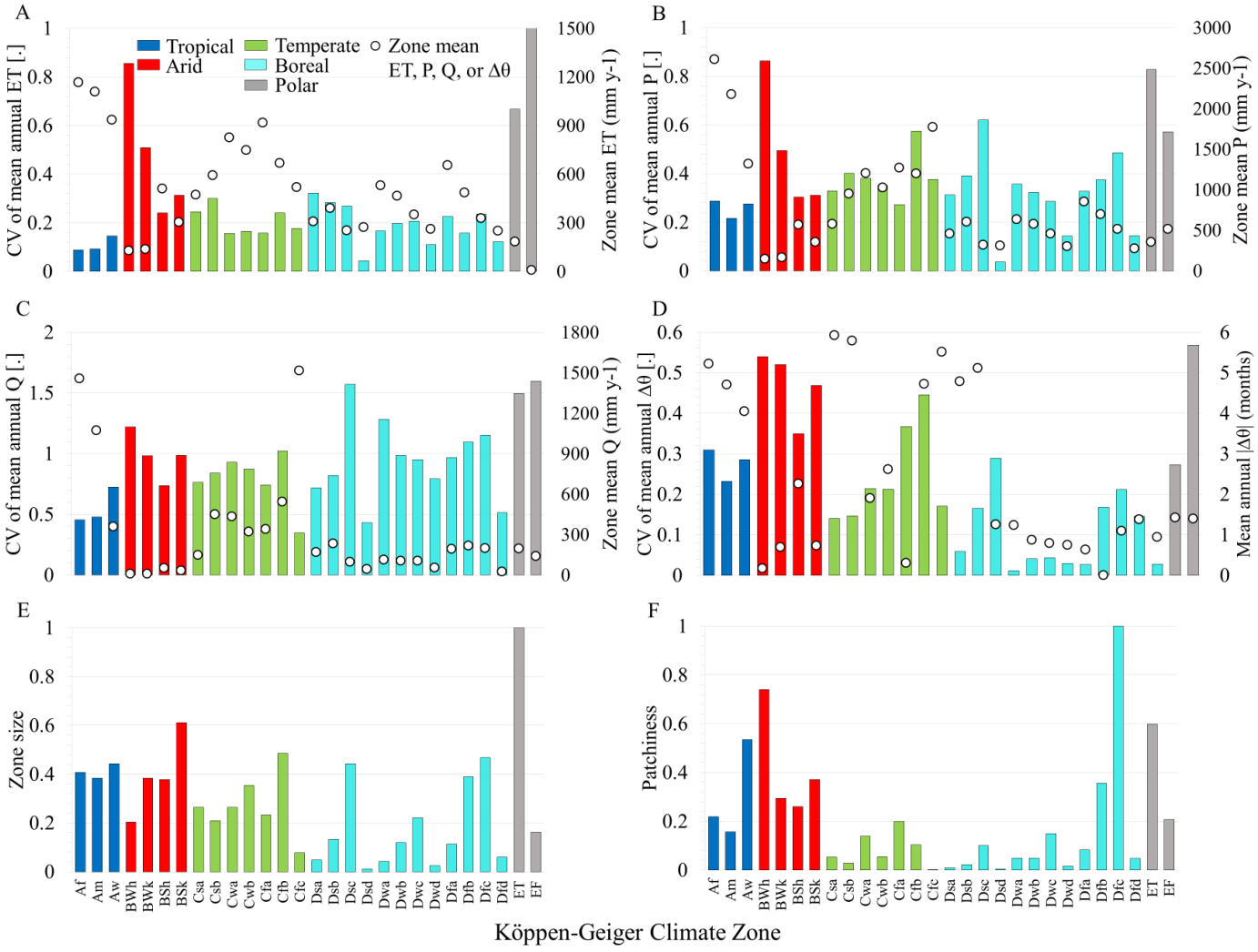

Figure 1: Water budget coherence, quantified as intra-zone CV (A-D), and complexity metrics of zone size and patchiness (E-F) for each zone for the Köppen-Geiger climate classification system. Mean hydroclimate values for each zone are also shown (circles, A-D). Zone size (number of pixels) and patchiness (number of patches) are normalized to their maximum values.

Coherence and complexity results for the six best performing of the eight climate classification systems are shown in Table 1. The performance compared to KPG, in order of worst-to-best, was generally HDL, ETV, KHC, MHR, ETC, ETA, and WEC. The HDL and ETV systems performed either worse than or not statistically different from KPG in all metrics, except higher ET coherence for ETV, and these results are therefore not shown in Table 1 (see Table S3). The established

KHC and MHR systems were also poor-performing, with overall worse or similar hydroclimate coherence compared to KPG. However, KHC and MHR had improved pixel coherence and fewer patches, respectively (Table 1). The proposed ETA and ETC systems had much better ET coherence than KPG and similar P coherence to KPG (based on the K-S test). The ETA system also had the fewest number of zones and very high coherence for zone size (by far the lowest CV_z), but both ETA and ETC had lower coherence than KPG for the remaining hydroclimate variables. Finally, the proposed WEC system had similar or better performance than the KPG system in all coherence and complexity metrics, except for $\Delta\theta$ coherence. The WEC framework is therefore selected as the overall best hydroclimate classification system.

Table 1: Hydroclimate coherence (intra-zone CV for mean annual ET, P, Q, $\Delta\theta$, and PET) and complexity (inter-zone CV_z for pixels, and numbers of patches and zones) in established and proposed climate classification systems. Mean(standard deviation), with significantly higher ($^{\uparrow}$) or lower ($^{\downarrow}$) values than KPG determined based on K-S tests. Bold indicates the best overall system for each metric (more than one system had statistically similar results for some metrics).

Metric	Established systems			Proposed systems		
	KPG	MHR	KHC	ETA	ETC	WEC
CV(ET)	0.27(0.21)	0.44(0.26)$^{\uparrow}$	0.48(0.30)$^{\uparrow}$	**0.12(0.17)$^{\downarrow}$**	**0.08(0.15)$^{\downarrow}$**	0.31(0.23)
CV(P)	0.38(0.20)	0.55(0.24)$^{\uparrow}$	0.41(0.16)	0.56(0.42)	0.47(0.35)	**0.25(0.23)$^{\downarrow}$**
CV(Q)	**0.88(0.30)**	1.30(0.57)$^{\uparrow}$	**0.78(0.26)**	1.45(0.50)$^{\uparrow}$	1.29(0.53)$^{\uparrow}$	**0.75(0.38)**
CV($\Delta\theta$)	**0.24(0.18)**	0.31(0.14)$^{\uparrow}$	0.36(0.16)$^{\uparrow}$	0.38(0.08)$^{\uparrow}$	0.37(0.09)$^{\uparrow}$	**0.33(0.14)**
CV(PET)	0.20(0.13)	0.20(0.09)	0.31(0.12)$^{\uparrow}$	0.53(0.24)$^{\uparrow}$	0.42(0.25)$^{\uparrow}$	**0.14(0.06)$^{\downarrow}$**
CV_z(zone areas)	1.24	0.78	0.54$^{\downarrow}$	**0.10$^{\downarrow}$**	0.47 $^{\downarrow}$	0.55 $^{\downarrow}$
patches	46(36)	**10(9)$^{\downarrow}$**	97(41)$^{\uparrow}$	153(70)$^{\uparrow}$	151(52)$^{\uparrow}$	59(36)
zones	28	27	18	**15**	20	22

The coherence and complexity metrics for the WEC system are shown in Figure 2 for all zones. Like KPG, coherence was high ($CV < 1$) in all zones for each hydroclimate variable apart from Q, but WEC had even higher hydroclimate coherence than KPG overall. Zone area was more equally distributed in WEC, but with similar patchiness to KPG (Table 1). The KPG system qualitatively groups 30 zones into 5 categories, and here the WEC zones were also divided into 5 groups. These zone groupings were organized around 1) zone mean aridity index ($\bar{\phi}$ = P/PET) and 2) distributing zones such that each group, denoted G1-G5, had a similar number of total pixels (Figure 3). Zone aridity indices were arranged in decreasing order, with G1 to G5 $\bar{\phi}$ = {2.4, 1.1, 0.83, 0.46, 0.12}, resulting in 11,367 to 13,895 pixels in each group. While WEC groups represented similar total areas, they comprised different numbers of zones, from G3 with only two large zones, to G1 with 7 zones.

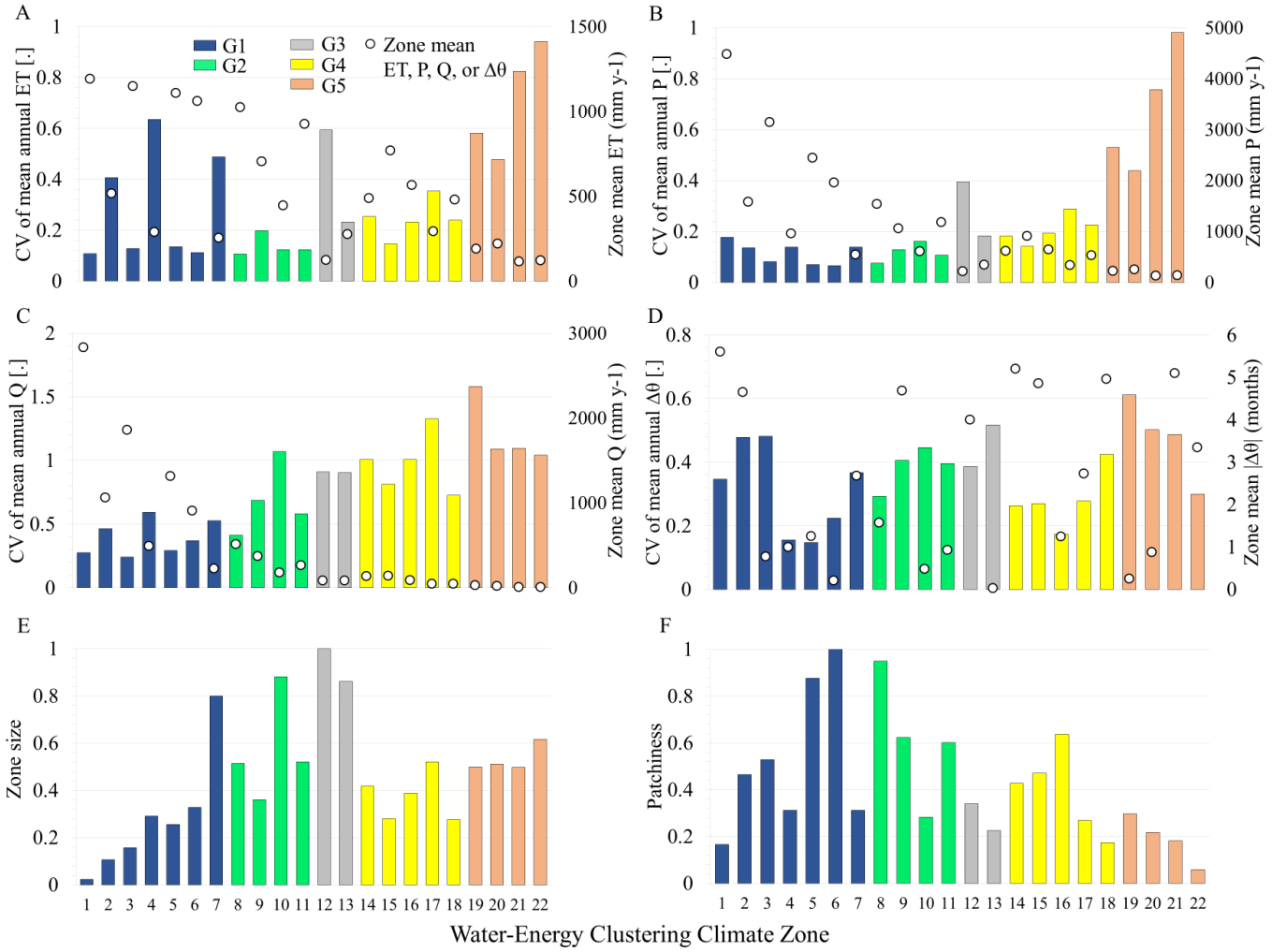

Figure 2. Water budget coherence (intra-zone CV, A-D) and complexity metrics (zone size and patchiness, E-F) for each zone for the Water-Energy Clustering (WEC) climate classification system. Mean hydroclimate values for each zone are also shown (circles, A-D). Zone size (number of pixels) and patchiness (number of patches) are normalized to their maximum values.

Maps of the boundaries for the proposed WEC system and the standard KPG framework are compared in Figure 3. While there were some similarities (e.g., see the Iberian Peninsula in Figure 3), most regions are divided differently. For example, parts of northern Europe are divided into three KPG zones but five WEC zones. Similarly, the southeastern United States, excluding south Florida, is mostly one KPG Temperate zone, but is separated in the WEC system into two distinct G2 zones. The KPG framework conversely divides eastern Europe from Russia in respective temperate and boreal zones, while WEC consolidates much of this area along with parts of western Europe into one G2 zone. Clustering centers for the WEC climate classification system are listed in Table S4.

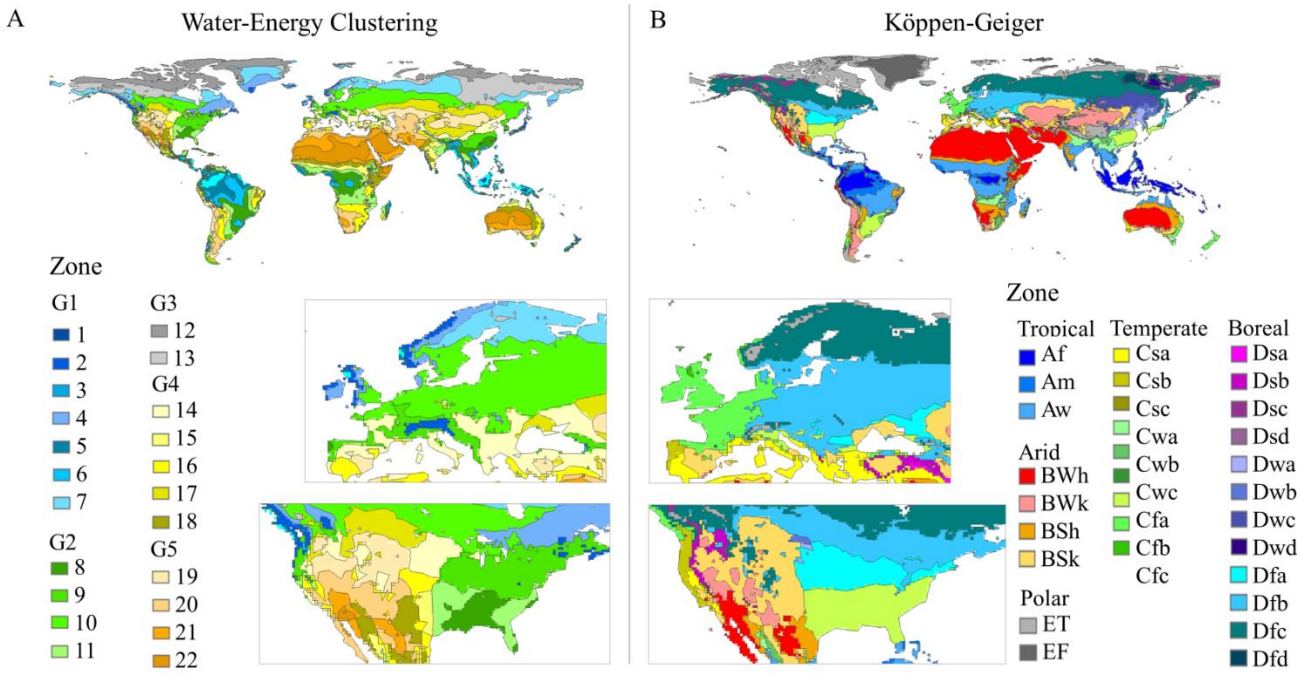

Figure 3. Spatial distribution of WEC (A) and KPG (B) classification systems. Europe and North America are magnified.

4 Discussion

We hypothesized that variable coherence and zone shape complexity would be related to the governing principles of the classification systems, which was mostly supported by the results of this study. Of the four previously established systems, KPG was the most hydroclimatically coherent, but had high variability in pixel distribution across zones (Table 1), even with the omission of the two small KPG zones that were removed when resampled to 0.5° x 0.5° resolution. The KPG and WEC frameworks had the overall highest $\Delta\theta$ coherence of the eight total compared systems, which is reasonable since KPG was the only system that accounted for monthly variability of water (P) and energy (temperature) that resulted in 24 parameters, or input variables (Beck et al., 2018). Similarly, the KHC framework accounted for long term mean monthly ratio P and PET. The WEC system was also based on water (P) and energy (PET), but from a mean annual perspective, thus requiring only two parameters. It is important to highlight that the number of required input variables, a notable aspect of system complexity, was lower for all systems proposed here (between 1 and 3 parameters) compared to the KPG system (24 parameters).

When comparing all eight systems, WEC had the highest P and PET coherence and similar ET, $\Delta\theta$, and Q coherence to KPG. It is not surprising that the WEC classification system yielded the highest P and PET coherence, given these were the variables used to draw its zone boundaries, but WEC also had much more uniform pixel distribution, similar zone fragmentation, fewer zones, and required substantially fewer parameters than KPG. The MHR system used long term mean Q

as a governing principle (Meybeck et al., 2013), while the KHC framework considered the Q regime as independent validation for their zones (Knoben et al., 2018). Although the MHR system used mean annual Q in their framework, it was not comparatively high in mean annual Q coherence, perhaps because their relatively larger zones did not reduce as much within-zone Q variability, or because the present analysis considers locally-generated Q (P – ET) and not gaged streamflow as they did. However, the KHC framework used gaged streamflow data for system evaluation (Knoben et al., 2018), and this system was one of the three highest for Q coherence. Additionally, the principle of contiguity in the MHR system led to the lowest patchiness of all systems evaluated, so this system could be useful when continuous boundaries are important for ease of implementation or interpretation purposes. Lastly, concordant with the objectives of each ET-based framework, the three univariate ET-based classification systems had the highest ET coherence, while ETA (which additionally optimized equal zone area) also had the most uniform pixel distribution across zones.

Evaluating hydroclimate coherence is important for understanding water availability distribution within a group of related zones and within individual zones to make informed management decisions. While land cover and land use impact the hydrologic cycle (Sterling et al., 2013), hydroclimate factors are the primary water budget drivers, especially at larger spatial scales where land cover effects are more muted (Sanford and Selnick et al., 2013). For KPG, hydroclimate dynamics are most uniform in Tropical zones and least coherent in Polar and Arid zones (Table S5). Arid zones had the most uniform pixel distribution, while the Boreal group was least fragmented with the lowest mean number of patches, suggesting these zones are interrupted neither by other zones nor by water bodies. For WEC, group G1 had the highest mean P and Q coherence, but also the lowest mean PET coherence. The latter is likely because of the temperature variation across G1 zones, which encompass both equatorial and subarctic regions (Figure 3). Group G5, comprising the most arid zones, had the lowest water budget (ET, P, and Q) and $\Delta\theta$ coherence but highest PET coherence, indicating relative uniformity in (low) rainfall and (high) temperature. Pixel distribution was most uniform in G3 and G5, while G5 was least fragmented. It is valuable to note the structural attributes of zone boundaries because these boundaries are expected to change over time (Beck et al., 2018; Knoben et al., 2018).

Of the water budget components, Q was the least coherent while ET was the most coherent across all systems except KHC and WEC (Table 1). This overall high coherence suggests that the variability of the drivers of ET (water and energy budget components) are mostly captured, even if ET itself is not a governing principle in the framework. However, there was still room for improvement within the established classification systems with respect to optimizing ET variability. Based on the climate classification system comparison presented here, it can be concluded that water and energy ET drivers are important considerations for broad hydroclimate analyses, since water budget coherence is mostly achieved when P and PET are included as governing principles. However, when specifically evaluating ET dynamics, using an ET framework is most appropriate. Depending on which spatial complexity metric is favored, MHR (mean zone patches = 10) and ETA (CV_z = 0.02) were the least complex systems. Based on both ET coherence and spatial complexity, the ETA system established here is suggested for ET-focused questions such as large-scale assessments of crop productivity (Howell et al., 2015).

This study is limited by a few factors. First, distinct climate zone boundaries, although useful in practice, do not exist in the physical system (Knoben et al., 2018). Second, this study compared averaged metrics that were applied across zones

within each classification system, although some individual zones were better or worse than others with respect to coherence and complexity. Third, the focus on long-term mean annual hydroclimate attributes for zone formation does not account for interdecadal climate dynamics. Last, the TerraClimate ET and Q data used to assess the suite of classification systems was in part formed using the same CRU climate data used here to create the WEC boundaries (Abatzoglou et al., 2018). However, the use of this data is justifiable in the context of this study for two primary reasons: 1) the spatial scope of this analysis is sufficiently large such that calibrated water budget rates are regionally representative (Abatzoglou et al., 2018), and 2) similarly, long term hydrologic dynamics are not as subject to interannual variability, since these effects are more muted across longer timescales. In this way, the broad spatiotemporal nature of this analysis makes it reasonable that all available P, ET, Q, and PET data are appropriate metrics for forming more robust hydroclimate boundaries and subsequently assessing the water (and energy) budget therein. It is also important to recall that Q here is based on locally generated runoff (P – ET), rather than the accumulated runoff from upstream contributing areas, which would be representative of gaged streamflow.

5 Conclusions

It is widely accepted that water and energy, chiefly in the form of rainfall and solar radiation, govern long term socioecological water availability at large spatiotemporal scales (Berghuijs and Woods, 2016; Knoben et al., 2018; Sanford and Selnick, 2013). It is important that boundaries representative of this water-energy interaction are drawn in order to encompass similar hydroclimatic sensitivities (Meybeck et al., 2013). The KPG system is the most widely used climate classification system, and this analysis revealed that it indeed has high hydroclimatic coherence. However, it was concluded that WEC was either better than or not statistically different from the KPG framework in all assessed metrics.

This study proposes WEC as a new framework for large-scale water budget inquiries, based on overall increased within-zone water budget coherence and reduced complexity (i.e., more even zone area distribution and fewer required input variables), which allows for more direct management interpretation and application. The WEC system is robust, based on long-term mean annual rates that have low susceptibility to interannual and seasonal variability. This framework is thus useful for regional to national scale management strategies to account for potential hydroclimate zone-dependent responses to climate and land cover change.

Data availability

The data used in this manuscript can be obtained from the Climate Research Unit TimeSeries V4.04 (monthly P and PET, Harris et al., 2020) and TerraClimate (monthly ET and Q aggregated to mean annual, Abatzoglou et al., 2018). Shape files for proposed climate classification system boundaries, including WEC, are available at the following repository: *(URL pending publication)*.

Author Contribution

KLMP performed the analyses and led the manuscript preparation. JWJ conceived and directed the study.

Competing Interests

The authors declare that they have no conflict of interest.

Acknowledgements

This research was supported in part by USDA National Institute of Food and Agriculture Hatch project FLA-SWS-005461.

References

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., and Hegewisch, K. C.: TerraClimate, a high resolution global dataset of monthly climate and climatic water balance from 1958-2015. *Scientific Data*, 5, 170191, 2018.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F.: Present and future Köppen Geiger climate classification maps at 1 km resolution. *Scientific Data*, 5,180214, 2018.

Berghuijs, W. R., and Woods, R. A.: A simple framework to quantitatively describe monthly precipitation and temperature climatology. *International Journal of Climatology*, 36(9), 3161 3174, 2016.

Bivand, R., Pebesma, E., and Gomez-Rubio, V.: *Applied Spatial Data Analysis with R*, Second edition. Springer, NY, 2013.

Boland, M. R., Parhi, P., Gentine, P., and Tatonetti, N. P.: Climate classification is an important factor in assessing quality-of-care across hospitals. *Scientific Reports*, 7(1), 1-6, 2017.

Chen, D., and Chen, H. W.: Using the Köppen classification to quantify climate variation and change: An example for 1901–2010. *Environmental Development*, 6, 69-79, 2013.

FAO (Food and Agriculture Organization of the United Nations – with UNESCO and WMO). *World Map of Desertification*. Food and Agricultural Organization (FAO), Rome, 1977.

Harris, I., Osborn, T. J., Jones, P., and Lister, D.: Version 4 of the CRU TS monthly high resolution gridded multivariate climate dataset. *Scientific Data*, 7(1), 1-18, 2020.

Hartigan, J. A., and Wong, M. A.: Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), 100-108, 1979.

Hijmans, R.: Raster: Geographic Data Analysis and Modeling. R package version 2.6-7, 2017.

Holdridge, L. R.: "Life zone ecology", *Life Zone Ecology*. Tropical Science Center, San Jose, Costa Rica, 1967.

Howell, T. A., Evett, S. R., Tolk, J. A., Copeland, K. S., and Marek, T. H.: Evapotranspiration, water productivity and crop coefficients for irrigated sunflower in the US Southern High Plains. *Agricultural Water Management*, 162, 33 46, 2015.

Jagai, J. S., Castronovo, D. A., and Naumova, E. N.: The use of Köppen climate classification system for public health research. *Epidemiology*, 18(5), S30, 2007.

Knoben, W. J., Woods, R. A., and Freer, J. E.: A Quantitative Hydrological Climate Classification Evaluated With Independent Streamflow Data. *Water Resources Research*, 54(7), 5088-5109, 2018.

Lanfredi, M., Coluzzi, R., Imbrenda, V., Macchiato, M., and Simoniello, T.: Analyzing Space–Time Coherence in Precipitation Seasonality across Different European Climates. *Remote Sensing*, 12(1), 171, 2020.

Lloyd, S. J., Kovats, R. S., and Armstrong, B. G.: Global diarrhea morbidity, weather and climate. *Climate Research*, 34(2), 119-127, 2007.

Magarey, R. D., Borchert, D. M., and Schlegel, J. W.: Global plant hardiness zones for phytosanitary risk analysis. *Scientia Agricola*, 65(SPE), 54-59, 2008.

McKenney, D. W., Pedlar, J. H., Lawrence, K., Campbell, K., and Hutchinson, M. F.: Beyond traditional hardiness zones: using climate envelopes to map plant range limits. *BioScience*, 57(11), 929-937, 2007.

Mellinger, A. D., Sachs, J. D., and Gallup, J. L.: Climate, coastal proximity, and development. The Oxford handbook of economic geography, 169, 194, 2000.

Meybeck, M., Kummu, M., and Dürr, H. H.: Global hydrobelts and hydroregions: improved reporting scale for water related issues? *Hydrology and Earth System Sciences*, 17(3), 1093 1111, 2013.

O'Neill, R. V., Krummel, J. R., Gardner, R. E. A., Sugihara, G., Jackson, B., and DeAngelis, D. L.: Indices of landscape pattern. *Landscape Ecology*, *1*(3), 153-162, 1988.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V., and Underwood: Terrestrial Ecoregions of the World: A New Map of Life on Earth, A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, 51(11), 933-938, 2001.

Papagiannopoulou, C., Gonzalez Miralles, D., Demuzere, M., Verhoest, N., and Waegeman, W.: Global hydro-climatic biomes identified via multitask learning. *Geoscientific Model Development*, 11(10), 4139-4153, 2018.

Peel, M. C., Finlayson, B. L., and McMahon, T. A.: Updated world map of the Köppen Geiger climate classification. *Hydrology and Earth System Sciences Discussions*, 4(2), 439-473, 2007.

Pierce, D.: ncdf4: Interface to Unidata netCDF (Version 4 or Earlier) Format Data Files. R package version 1.16, 2017.

R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/, 2018.

Richards, D., Masoudi, M., Oh, R. R., Yando, E. S., Zhang, J., and Friess, D. A.: Global Variation in Climate, Human Development, and Population Density Has Implications for Urban Ecosystem Services. *Sustainability*, 11(22), 6200, 2019.

Sanford, W. E., and Selnick, D. L. Estimation of evapotranspiration across the conterminous United States using a regression with climate and land-cover data 1. *JAWRA Journal of the American Water Resources Association*, 49(1), 217-230, 2013.

Sterling, S. M., Ducharne, A., and Polcher, J. The impact of global land-cover change on the terrestrial water cycle. Nature Climate Change, 3(4), 385-390. 2013.

Van der Ent, R. J., Savenije, H. H., Schaefli, B., and Steele-Dunne, S. C.: Origin and fate of atmospheric moisture over continents. *Water Resources Research*, *46*(9), 2010.

VanDerWal, J., Falconi, L., Januchowski, S., Shoo, L., and Storlie, C.: SDMTools: Species Distribution Modelling Tools: Tools for processing data associated with species distribution modelling exercises. R package version 1.1 221.1. https://CRAN.Rproject.org/package=SDMTools, 2019.

Wang-Erlandsson, L., Fetzer, I., Keys, *P*. W., Van Der Ent, R. J., Savenije, H. H., and Gordon, L. J.: Remote land use impacts on river flows through atmospheric teleconnections. *Hydrology and Earth System Sciences*, 22(8), 4311 4328, 2018.

Willmott, C. J., and Feddema, J. J.: A more rational climatic moisture index. *The Professional Geographer*, 44(1), 84 88, 1992.

Zhang, K., Kimball, J. S., and Running, S. W.: A review of remote sensing based actual evapotranspiration estimation. *Wiley Interdisciplinary Reviews: Water*, 3(6), 834-853, 2016.