Coherence of Global Hydroclimate Classification Systems

Kathryn~~.~~ L. McCurley Pisarello[1], J.W. Jawitz[2~~1~~]

[1]USDA-ARS, Southeast Watershed Research Laboratory, 2316 Rainwater Road, Tifton, Georgia 31793, USA
[2~~1~~]Soil and Water Sciences Department, University of Florida, Gainesville, Florida 32611, USA

Correspondence to: ~~James W. Jawitz (jawitz@ufl.edu)~~ Kathryn L. M. Pisarello (katie.pisarello@usda.gov)

Formatted: Not Superscript/ Subscript

Abstract. Climate classification systems are useful for investigating future climate scenarios, water availability, and even socioeconomic indicators as they relate to climate dynamics. There are several classification systems that apply water and energy variables to create zone boundaries, although there has yet to be a simultaneous comparison of the structure and function of multiple existing climate classification schemes. Moreover, there are presently no classification frameworks that include evapotranspiration (ET) rates as a governing principle. Here, we developed a new system based on precipitation and potential evapotranspiration rates, as well as three systems based on ET rates, which were all compared against four previously established climate classification systems. The within-zone similarity, or coherence, of several long-term hydroclimate variables was evaluated for each system based on the premise that the application and interpretation of classification framework should correspond to variables that are most coherent. Additionally, the shape complexity of zone boundaries was assessed for each system, assuming zone boundaries should be drawn efficiently such that shape simplicity and hydroclimate coherence are balanced for consistent boundary interpretation and application. The most frequently used climate classification system, Köppen-Geiger, had high hydroclimate coherence overall but also high shape complexity. When compared to the Köppen-Geiger framework, the Water-Energy Clustering classification system introduced here showed overall improved or equivalent coherence for hydroclimate variables, yielded lower spatial complexity, and required only two, compared to 24, parameters for its construction.

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), and 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 terrestrial 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 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 classification system, the most widely used climate framework, 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 gauged runoff (Q) of river basins. They compared the resulting boundaries against the Köppen-Geiger and Holdridge frameworks to assess zone boundary overlaps. The 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 flow regime, but they did not compare other water budget components nor additional 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). In addition to the absence of a zoning system that accounts for ET dynamics, there has been no comparison of within-zone hydroclimate coherence across multiple climate classification systems, with evaluation particularly lacking in considering ET rates. Furthermore, the spatial complexity of climate classification systems has not been systematically examined across multiple frameworks, although Guan et al., 2020 quantified the changing spatial structure of the Köppen-Geiger framework over time. Assessing the structure of a biophysical system is a concept that most notably originates from landscape ecology (O'Neill et al., 1988) and provides a suite of shape metrics that can be cross-disciplinarily applied. Quantifying shape pattern and spatial contouring of climate classification systems is important for understanding interactions between governing

hydroclimatic characteristics as well as anticipating socioecological consequences that result from changing atmospheric configurations (Guan et al., 2020).

This work seeks to provide empirical support for application-dependent selection among candidate climate classification systems. We suggest that a successful classification system should have high within-zone coherence for variables that are related to the system's intended use, combined with relatively low shape complexity across zones, which is best for ease of interpretation within management and policy contexts. As such, we postulate 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 created a series of ET-based global classifications that were expected to yield comparatively higher ET coherence than other systems.

We evaluated 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 four new global classification systems and 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 distribution of zone area (km^2), mean zone fragmentation (i.e., mean number of patches comprising each zone), and the number of zones required to effectively form hydroclimate boundaries. 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 values correspond to higher coherence, meaning that regions delineated by zone boundaries that yield low CV values are more spatially homogenous and therefore more hydroclimatically continuous. An additional important component of this analysis is the evaluation of the tradeoff between hydroclimate coherence and the spatial complexity of zone boundaries. It is valuable to consider the structural attributes of zone boundaries because these boundaries are expected to change over time (Beck et al., 2018; Knoben et al., 2018). Building more precise boundaries may better delineate similar 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 primarily based on three principles: 1) Classification systems should consist of a relatively even area 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 Köppen-Geiger 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 areathe number of pixels (km^{2}zone area evenness, CVz) as defined by CV, and the mean number of patches in each zone (zone fragmentation), and the number of needed zones to bound hydroclimatically similar areas. 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 mean number of patches per zone was determined using the R function lsm_c_npClassStat in package *landscapemetricsSDMTools* (Hesselbarth, et al., 2019VanDerWal et al., 2019). For each hydroclimate and complexity variable, statistical differences between the KPG framework and the other classification systems were determined based a series of two-sided Kolmogorov-Smirnov (K-S) tests, which compares probability distributions to a reference distribution.

Formatted: Superscript

2.2 Database construction

SWe everal open access datasets were compiled to create the database used for climate classification system calibration and validation. We evaluated global gridded monthly P and PET and mean annual ET and Q between 1980 and 2014 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 long-term mean monthly values. In this case, long-term mean values muted 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.

Additional ET and Q datasets were used for independent validation purposes. Observation-based monthly Q from 1980-2014 were obtained at 0.5° x 0.5° resolution from monthly global gridded runoff data (GRUN, Ghiggi et al., 2019). The Global Lobal Evaporation Amsterdam Model (GLEAM) produced terrestrial daily ET for 1980-2020 at 0.25° x 0.25° resolution, which was also resampled to 0.5° x 0.5° resolution. Here, we used the updated GLEAM version 3.5a, which is based on ERA5 net radiation (satellite) and air temperature (reanalysis) datasets, downloadable at a monthly timestep (Martens et al., 2017). The GLEAM ET and GRUN Q datasets were shown to be sufficiently different independent from TerraClimate ET and Q datasets both temporally (Figures S1 and S3) and spatially (Figures S2 and S4). The two ET datasets did not have as much variability aswere more similar than the two Q datasets, based on monthly linear models (R^2 ranging from 0.78 to 0.87 for ET and 0.47 and 0.84 for Q), and both ET and Q datasets showed spatially consistent seasonal differences. Hereafter, TerraClimate ET and Q are simply referred to as "ET" and "Q" unless otherwise noted.

Formatted: Superscript

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 60,726 pixels.

2.3 Sinusoidal functions as descriptors of seasonality

The seasonal dynamics of monthly P and PET were additionally 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)

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

where y is either P or PET (mm month^{-1}) for each month, m, with overall monthly mean (i.e., mean of the 12 long term monthly means) denoted by the overbar, r is dimensionless amplitude, and t is the p. Phase ~~angle, θ, is the~~ offset (months) from the reference time, January ($m = 1$). Phase difference, Δt, measures the synchronization of P and PET throughout the year, is determined as the difference $t_{PET} - t_P$, and , ~~is with absolute value of phase difference $|\Delta\theta| \leq 6$~~). Phase difference (constrained $-6 \leq \Delta t \leq 6$ ~~(as~~ more detail ~~described in the~~ SI).) ~~measures the synchronization of P and PET throughout the year Phase difference, $\Delta\theta$, describes the synchronization of P and PET throughout the year as~~

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

Figure 1A shows the overall global distribution of Δt (Equation S1), where some banding around the Tropics as well as the Middle East can be seen. Equation 1 yielded overall good fits to the long-term mean monthly distributions of P and PET, with R^2=0.67±0.28 and 0.84±0.18, respectively (mean±standard deviation across all pixels). These sine fits to monthly PET were statistically significant (p-value $\leqslant$ 0.05) in 97% of pixels, while fits to monthly P were statistically significant in 85% of pixels (Figure 1B-C). Cumulative distribution functions of both R^2 and p-value for PET and P sine fits can be seen in Figure S5. Similar to Berghuijs and Woods (2016), P fits were good in South America, and not as good in parts of the Sahara (Figure 1C). Our P 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, compared to the performance of our PET fits shown in Figure 1B, the temperature fits of Berghuijs and Woods (2016) were overall much less spatially homogeneous than ours.

~~To constrain $-6 \leq \Delta\theta \leq 6$, 12 was either added to or subtracted from $\Delta\theta$ 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~~

Formatted: Font: Italic

Formatted: Indent: First line: 0"

Formatted: Superscript

~~zone, the Δθ distributions were normalized by centering around the mode and correcting to contain only positive values (0 to 12 months).~~

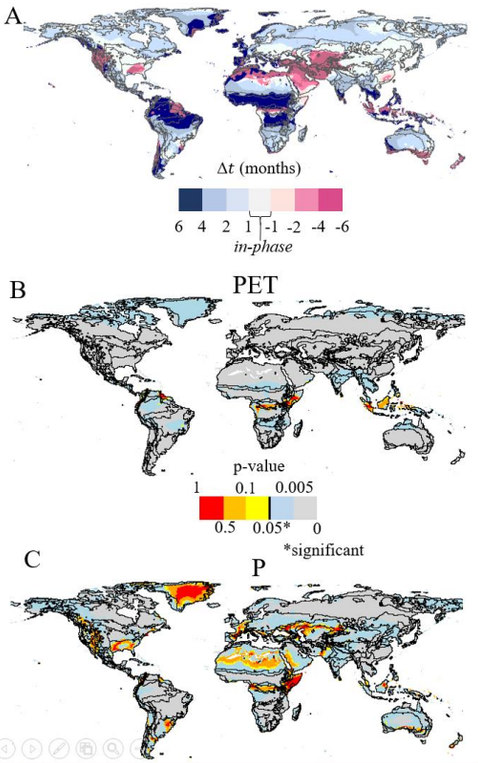

Figure 1: Global spatial distributions of Δt (A) and of performance of monthly P (B) and PET (C) sine fits represented by p-value.

Formatted: Indent: First line: 0"

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 pixels with a corresponding probability continuum of belonging to a zone. However, Knoben et al., 2018 chose to bound 18 zones using their provided climate indices (aridity index,

seasonal aridity index, and precipitation as snow) for inter-system comparison purposes. In the present study, 18 KHC boundaries were re-created using a clustering approach similar to the clustering methodology of Knoben et al., 2018. 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, 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 Novel~~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 developed 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 novel~~proposed~~ univariate classification systems were based on the global ET empirical cumulative distribution function (CDF). The first classification system, ET Area-optimizing (ETA), was created with the additional condition 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 S10B~~1~~), resulting in highly variable relevance for regional analyses. A classification system that is more spatially uniform can better 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 seeks to additionally 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," although this is not a physical reality (Meybeck et al., 2013). The cumulative probability interval [0,1] was divided into 15 equal parts, each corresponding to a separate zone, and the 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 S~~6~~1-A). The number of ETA zones was chosen based on the number of zones in previously established systems and the relative improvement of ET coherence with the addition of more zones (Figure S~~6~~1 B).

The second proposed classification system, ET Variability-optimizing (ETV), was based on the principle of maximizing within-zone ET coherence subject to the tradeoff of 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 S~~7~~2), we fitted a uniform distribution, which is characterized

Formatted: Not Highlight

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

$$a_i = \frac{b_i(1-\sqrt{3}\mathrm{CV}_{\mathrm{min}})}{1+\sqrt{3}\mathrm{CV}_{\mathrm{min}}} \quad (2)$$

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 $\mathrm{CV}_{\mathrm{min}}$ = 0.075 was chosen based on marginal CV decrease with increasing number of zones (Figure S7~~2~~-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 S7~~2~~-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 the previously described analytical ET-based zoning frameworks. The final number of 20 clustering centers (i.e., zones) was selected based on the smallest number of zones with CV of mean annual ET below a low threshold, selected here as 0.1 (Figure S8~~3~~).

2.6 Novel~~Proposed~~ multivariate climate classification system~~s~~

The final developed 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 from multiple variables. This final system is herein referred to as the Water-Energy Clustering (WEC) climate classification system.

Since the KPG is the standard framework to which other systems were compared, a main objective was to create a classification scheme that was at least as good as KPG, while also using fewer biophysical parameters to draw zone boundaries. The final number of proposed WEC zones was chosen based on the common “elbow method” for visually determining the optimal number of clusters, or zones (Syakur et al., 2018). Within the context of the presently applied k-means clustering method, the elbow method seeks to efficiently minimize the total within-zone sum of squares (TSS), such that the optimal number of zones exists where the rate of TSS change starts to decrease with the addition of more zones. According to the goal of efficiently minimizing TSS, about 5 zones would be best (Figure S9). However, the aim of this study was to optimize zones based on hydroclimate coherence and zone shape complexity. Based on this premise, the “elbow” of hydroclimate coherence (i.e., low CV values) with respect to number of zones is around 10 to 20 zones for all hydroclimate variables, except for Δt

(Figure 2). Similarly, the elbow denoting the efficient minimization of mean number of patches across zones is approximately ~~around~~ 15 to 25 WEC zones, but CV of zone area was relatively constant between 10 and 30 zones (Figure S10). A WEC system consisting of at least 10 zones yielded mean coherence values that were better than those mean coherence values of KPG for PET, P and Q, as denoted by dark blue dots in Figure 3. It should also be noted that although there was no number of WEC zones that provided a lower mean number of patches than KPG, most possible numbers of WEC zones yielded mean values that were within one standard deviation of the KPG mean (Figure S10A). Also, all possible numbers of WEC zones allowed for a more equal distribution of zone areas compared to the CV of zone area for KPG (Figure S10B). We chose to include both 15 and 20 possible WEC zones (compared to KPG's 30 zone system) for further evaluation across all climate classification systems. However, the results for 15 WEC zones will be presented henceforth, since the K-S test showed no statistical difference in coherence nor complexity between 15 and 20 zones (the coherence and complexity results for 20 WEC zones can be seen in Table S1).

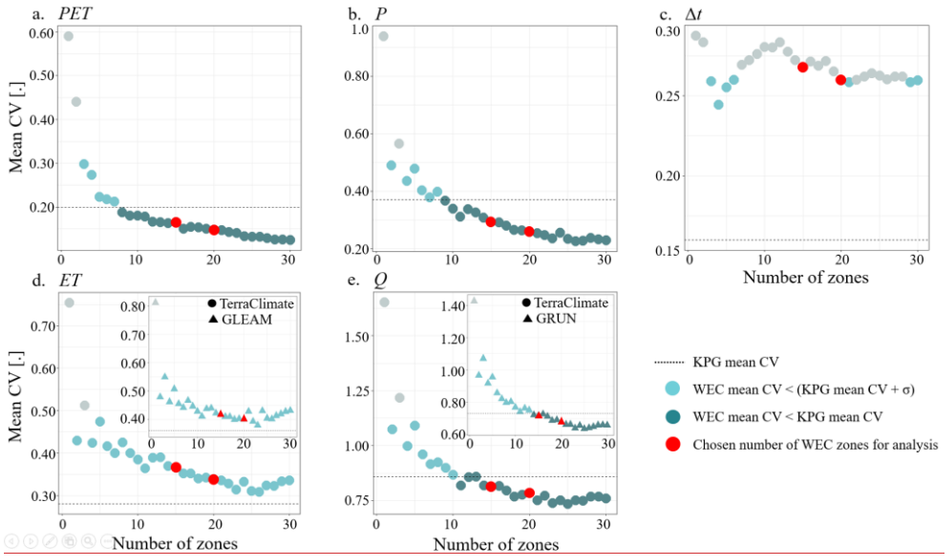

Figure 2: Hydroclimate coherence with respect to number of possible zones within the WEC framework for all hydroclimate variables: PET (A), P (B), Δt (C), ET (D), and Q (E). For independent validation, ET and Q are also included from secondary gridded data sources, GLEAM and GRUN, respectively, and are differentially illustrated by triangles. Number of zones that yielded a mean CV value lower than that of KPG (gridded horizontal line) are shown in dark blue, number of zones that yielded a mean CV value that was lower than that of KPG plus one standard deviation (σ) are shown in light blue, number of zones that yielded a mean CV value that was higher than that of KPG plus σ are shown in light grey, and the final number of zones chosen for further evaluation are in red. ~~from a~~A ~~suite of potential candidate multivariate systems that wasas generated using mean annual~~

Formatted: Line spacing: single

P, PET, Δθ, ET, and Q. These potential candidate classification systems were evaluated based on elimination thresholds that. were primarily 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 multivariateProposed systems were eliminated if their complexity metrics exceeded the mean KPG complexityvalues (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 the KPG water budget coherencevalues (Table 1)Since the KPG system is the most widely used classification system, the specific thresholds of elimination corresponded to the mean values of KPG water budget coherence and complexity (KPG values listed in Table 1). There was some flexibility with the water budget coherence thresholds, where the new system was allowed mean coherence values that did not exceed 50% those of KPG. Some allowance for reduced water budget coherence was made in favor ofs 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 best performing One multivariate classification scheme was chosen from the batch of potential systems by evaluating criteria corresponding to KPG system characteristics. The primary criteria considered were water budget coherence (mean CV of ET, P, and Q) and zone complexity. Some water budget coherence flexibility was allowed for the final multivariate system, with water budget coherence constrained to be so conditions were based on maintaining at least a within 50% range in relation to of KPG coherence values. The specific 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 Tables S1 and S2, where Table S1 indicates the full suite of assessed multivariate systems, and with Table S2 shows the final two candidate multivariate systems in Table S2.

The final two compared systems were chosen, because 1) they only required only two input variables and 2) they met all conditions of coherence and complexity as described by the 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 Ultimately, the climate classification system formed from clustering mean annual P and mean annual PET was chosen as the representative multivariate framework proposed here, and is herein referred to which. This classification scheme was named as the Water-Energy Clustering (WEC) climate classification system.

Formatted: Indent: First line: 0", Line spacing: single

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 of hydroclimate variables, PET, P, Δt, ET (TerraClimate), and Q (TerraClimate) and complexity metrics for the KPG systemfor each evaluated climate classification system is, the standard used in this study, are shown in Figure 3 for all zones.. Figure S11 illustrates the coherence of GLEAM ET and GRUN Q, which were variables used to augment independent validation. A two-sided K-S test was conducted to determine statistical differences across systems, which compared between the cumulative distributions of CVs compared to a reference system values for each variable, using WEC_{20} as the reference distribution system. The WEC system was used as the reference system, since WEC is the novel multivariate system proposed in this study (SIxxx).

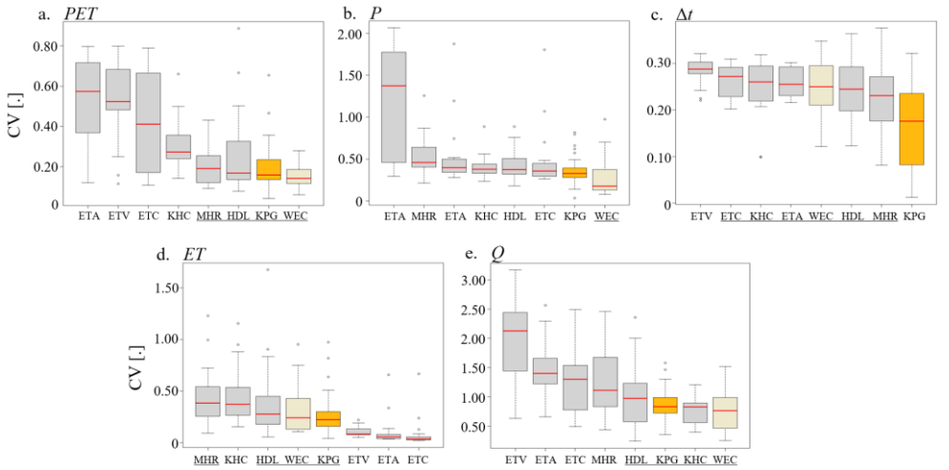

Figure 3: Boxplots of hydroclimate coherence, quantified as intra-zone CV (A-E), for each assessed climate classification system, with KPG shown in gold and WEC in light beige. The results of the K-S test were used to determine statistical difference of distributions compared to WEC. Systems whose distributions were not statistically different from that of WEC are underlined.

Figure 4 showed MHR was the least fragmented system overall, although the KPG system also was characterized by low patchiness when compared to the distributions of the other systems (Figure 4A). The KPG system also had relatively high hydroclimate coherence for most variables, including validation datasets GLEAM ET and GRUN Q (Figure S11), appearing as the best system (i.e., low CV values) for Δt and not statistically different from the best system for PET and Q ~~, in the case of P coherence, except Q~~ (Figure 3B~~1~~). However, it did not have the highest P or ET coherence. The high Δt coherence of KPG is sensible, because KPG zones are built using intra-annual P and temperature (i.e., PET) dynamics. While t~~T~~he KPG system showed overall high coherence, which supports its status as the most widely used climate classification system, it was not the highest for all variables, and it ~~but~~ also exhibited high complexity with respect to zone area distribution and number of zones used in its framework (Figure 4B-C). Lastly, the number of biophysical parameters required to construct KPG zone boundaries (monthly P and temperature, n =24) is much higher than th~~os~~e novel systems presented here (n=2 for WEC and n=1, mean annual ET, for ETA-based systems).

The coherence of hydroclimate drivers, PET, P, and Δt, as well as hydroclimate response variables, ET and Q, were~~as~~ variable across systems (Figure 3). The variable that was overall least coherent was Q (Figures 3E and S11B), with both TerraClimate and GRUN Q CV values ranging beyond 1.50 for most classification systems, while the variable that was generally most coherent was Δt (Figure 3C), with CV values generally between 0.10 and 0.30 for all assessed classification systems. Of all variables, Δt yielded the greatest number of systems not statistically different from WEC, based on the K-S

Formatted: Font: 9 pt, Bold
Formatted: Indent: First line: 0", Line spacing: single
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt
Formatted: Font: 9 pt
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold

test for differences in CV distributions (Figure 3C). Additionally, the three novel ETA, ETV, and ETC systems had better TerraClimate ET coherence than the other classification systems (Figure 3D). The ETA system, along with WEC, also had the fewest number of zones and provided the most uniform zone size distribution but was not as coherent with respect to hydroclimate variables apart from ET.

The WEC system had the lowest median CV ~~value~~ for PET, P, and both TerraClimate Q (Figure 3A- B and E) and GRUN Q (Figure S11B). It is reasonable that the WEC system is the most PET and P coherent, since these were the variables used to form the zone boundaries of the system. The high coherence of GRUN Q serves as an independent validation of the WEC framework, such that it can be concluded that the WEC system most effectively bounds zones that capture water availability drivers. Although the ET-based systems were best at bounding within-zone ET similarities and yielding high coherence, WEC did not perform worse than KPG in ET coherence, according to the K-S test (Figure 3D). The WEC system was also relatively less complex compared to most other systems, including KPG, with respect to zone area distribution and number of zones required to draw hydroclimatically coherent boundaries (Figure 4B-C). The WEC distribution of the mean number of patches in each~~per~~ zone was statistically different from that of KPG, and the WEC system had the next lowest median value following KPG (Figure 4A). Since the proposed WEC system had similar or better performance than the KPG system in most coherence and complexity metrics (except for Δt coherence and mean number of patches), and required 2 compared to 24 parameters to construct, the evaluated WEC framework was selected as the overall best hydroclimate classification system.

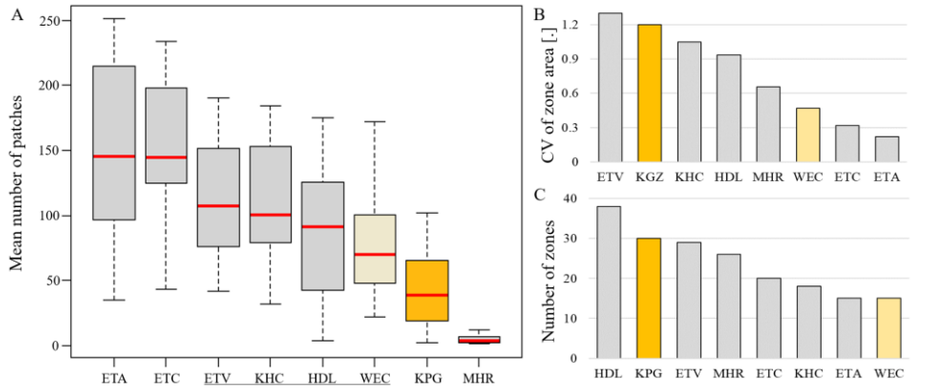

Figure 4: Boxplots of mean number of patches per zone (A), barplots of CV of zone areas (B), and barplots of number of constructed zones (C) for each assessed climate classification system, with KPG shown in gold and WEC in light beige. The results of the K-S test were used to determine statistical difference of distributions compared to WEC. Systems whose distributions of mean number of patches were not statistically different from that of WEC are underlined. ~~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).~~ all and

Formatted: Font: 9 pt
Formatted: Font: 9 pt
Formatted: Left, Indent: First line: 0", Line spacing: single
Formatted: Font: 9 pt
Formatted: Font: 9 pt
Formatted: Font: 9 pt

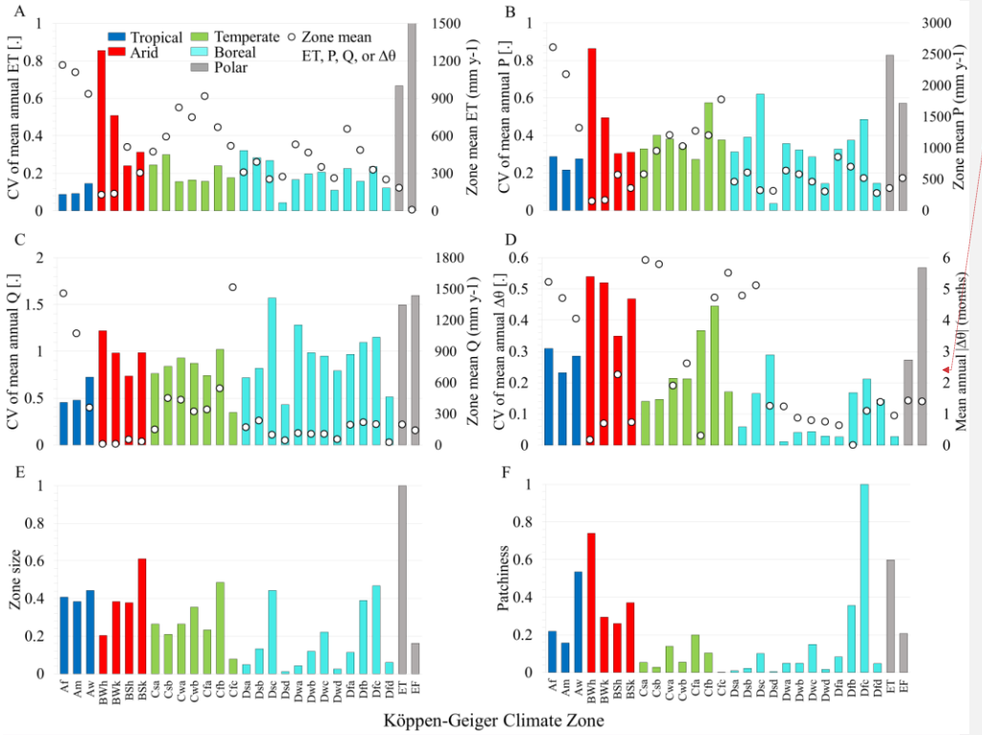

Figure 1: Water budget coherence, quantified as intra-zone CV (A-D), and complexity metrics of zone size and patchiness (E-F) for each zone for the Köppen-Geiger climate classification system. Mean hydroclimate values for each zone are also shown (circles, A-D). Zone size (number of pixels) and patchiness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

Formatted: Line spacing: single

Formatted: Justified, Line spacing: single

Formatted: Normal

Formatted: Line spacing: single

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 Δθ coherence. The WEC framework is, and thereforecan be concluded selected as a strong contender for the overall best hydroclimate classification system.

Table 1: Hydroclimate coherence (intra-zone CV for mean annual ET, P, Q, Δθ, and PET) and complexity (inter-zone CV, for pixels, and numbers of patches and zones) in established and proposed climate classification systems. Mean(standard deviation), with significantly higher (↑) or lower (↓) 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)↑	0.48(0.30)↑	**0.12(0.17)**↓	**0.08(0.15)**↓	0.31(0.23)
CV(P)	0.38(0.20)	0.55(0.24)↑	0.41(0.16)	0.56(0.42)	0.47(0.35)	**0.25(0.23)**↓
CV(Q)	**0.88(0.30)**	1.30(0.57)↑	**0.78(0.26)**	1.45(0.50)↑	1.29(0.53)↑	**0.75(0.38)**
CV(Δθ)	**0.24(0.18)**	0.31(0.14)↑	0.36(0.16)↑	0.38(0.08)↑	0.37(0.09)↑	**0.33(0.14)**
CV(PET)	0.20(0.13)	0.20(0.09)	0.31(0.12)↑	0.53(0.24)↑	0.42(0.25)↑	**0.14(0.06)**↓
CV_a(zone areas)	1.241	0.78	0.54↓	**0.1002**↓	0.478↓	0.55↓
patches	46(36)	**10(9)**↓	97(41)↑	153(70)↑	15210(52)↑	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 primary categories (Tropical, Arid, Temperate, Boreal, and Polar), and here the 15 WEC zones were also divided into 5 primary groups by ranking zones based on increasing zone mean aridity index ($\varphi = \frac{\overline{PET}}{\bar{P}}$). The zones were then evenly grouped, resulting in five primary categories: Superhumid ($\varphi = 0.39$), Humid ($\varphi = 0.59$), Temperate ($\varphi = 1.10$), Arid ($\varphi = 2.548$), and Hyperarid ($\varphi = 62$). Note that the single WEC zone with highest aridity encompasses the Sahara and parts of Saudi Arabia and western Australia, for which $\varphi = 121$. (Figure 3). These zone groupings were, organized around 1) zonedered by based on their mean aridity index ($\bar{\phi}\phi = P/PET$) and 2) distributing zones such that each group, denoted G1-G5, had a similar number of total pixelsminimizing the variability of zone areas (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 inGroups were

Formatted Table
Formatted: Font: (Default) +Headings (Times New Roman)
Formatted: Font: (Default) +Headings (Times New Roman), Not Bold
Formatted: Font: (Default) +Headings (Times New Roman)
Formatted: Space After: 6 pt, Line spacing: single
Formatted: Font: (Default) +Headings (Times New Roman)
Formatted: Space After: 6 pt, Line spacing: single
Formatted: Font: (Default) +Headings (Times New Roman)
Formatted: Space After: 6 pt, Line spacing: single
Formatted: Font: (Default) +Headings (Times New Roman), Bold
Formatted: Font: (Default) +Headings (Times New Roman)
Formatted: Font: (Default) +Headings (Times New Roman)
Formatted: Space After: 6 pt, Line spacing: single
Formatted: Font: (Default) +Headings (Times New Roman)
Formatted: Space After: 6 pt, Line spacing: single
Formatted: Font: (Default) +Headings (Times New Roman)
Formatted: Space After: 6 pt, Line spacing: single
Formatted: Font: (Default) +Headings (Times New Roman), Not Bold
Formatted: Font: (Default) +Headings (Times New Roman)
Formatted: Space After: 6 pt, Line spacing: single
Formatted: Font: (Default) +Headings (Times New Roman)
Formatted: Space After: 6 pt, Line spacing: single
Formatted: Font: (Default) +Headings (Times New Roman), Not Bold
Formatted: Font: (Default) +Headings (Times New Roman)
Formatted: Indent: First line: 0.5"
Formatted: Not Highlight

organized to, comprise near-equal pixel distributions (11,367 to 13,895 pixels in each group,) across zones of decreasing aridity index, with G1 to G5 $\bar{\phi}$ = {2.4, 1.1, 0.83, 0.46, 0.12}. In WEC zones, there were positive linear relationships ($p<0.05$) between aridity index and CV of each water budget variable: ET (R^2=0.25), P (R^2=0.53), and Q (R^2=0.69), indicating higher spatial variability (lower coherence) with increasing aridity suggesting that as zones become drier, hydrologic regimes become more spatially variable, or less coherent. 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.

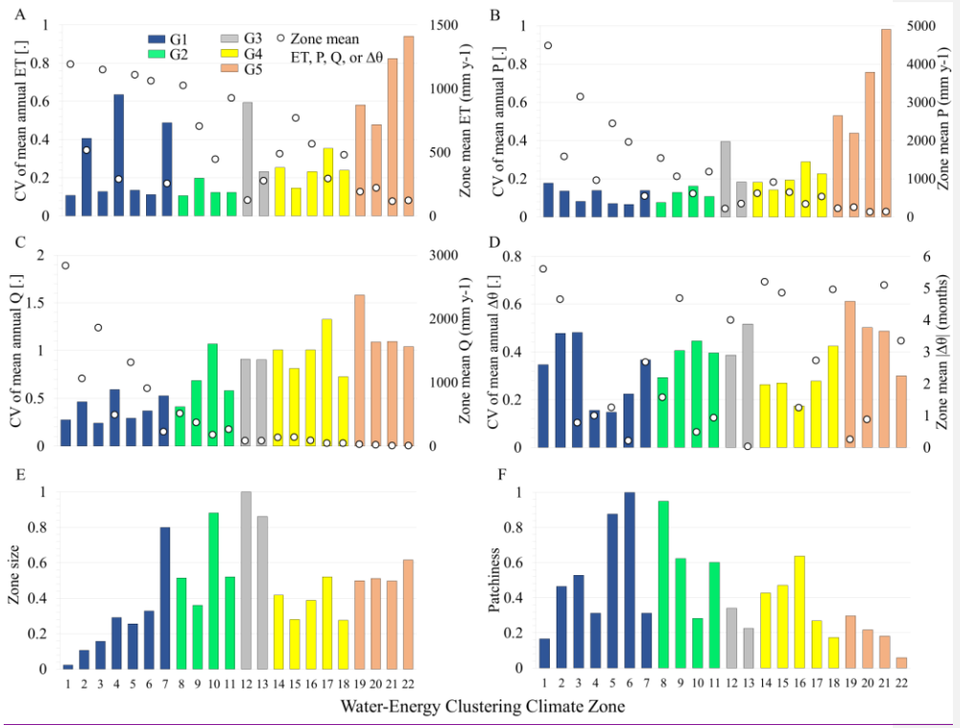

Figure 2. **Water budget coherence (**intra-zone CV, **A-D) and complexity metrics (**zone size and patchiness, **E-F) for each zone for the Water-Energy Clustering (WEC) climate classification system.** Mean hydroclimate values for each zone are also shown (circles, A-D). **Zone size (number of pixels) 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 5~~3~~. While there are~~were~~ some spatial similarities (e.g., see the Iberian Peninsula in Figure 5~~3~~), most regions are~~are~~ divided differently. For example, parts of northern Europe are mainly divided into three KPG zones but four~~five~~ WEC zones. Similarly, the southeastern United States, excluding south Florida, is mostly one KPG ~~Temperate~~ zone, but is separated in the WEC~~EC~~ system into two distinct ~~G2~~ zones. The KPG framework conversely divides eastern and western Europe ~~from Russia~~ in respective temperate and boreal zones, while WEC treats western Europe as more heterogeneous~~consolidates much of this area along with parts of western Europe into one G2 zone~~. Clustering centers, which are the arithmetic means of each of the clusters, for the WEC climate classification system are listed in Table S24.

Formatted: Not Superscript/ Subscript

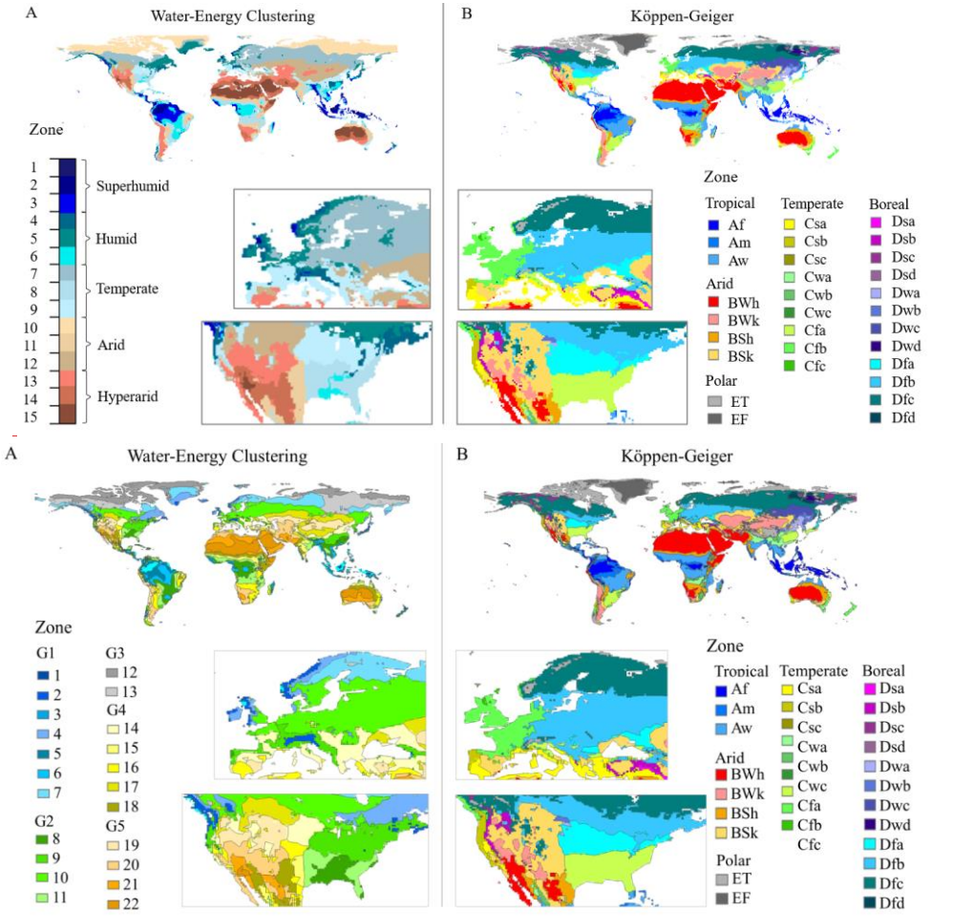

Figure 53. Spatial distribution of WEC (A) and KPG (B) classification systems. Europe and North America are magnified.

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 zone area variability ~~in pixel distribution across zones~~ (Figure S10~~Table 1~~), even with the omission of the two small KPG zones when resampled to 0.5° x 0.5° resolution in this study. The KPG ~~and WEC~~ framework~~s~~ had the overall highest Δ~~θ~~t coherence of the eight~~eight~~ total compared systems, which is reasonable since KPG was the only system that accounted for monthly variability of water (P) and energy (temperature), resulting in 24 biophysical parameters (Beck et al., 2018). The KHC framework similarly accounted for the long term mean monthly ratio of P and PET, but it was not particularly high in Δt coherence (Figure 3C). The WEC system was also based on water (P) and energy (PET), but from a mean annual perspective, thus requiring only two biophysical parameters as input variables. It is important to highlight that all novel systems presented here required fewer input variables, a notable aspect of system complexity, than any other evaluated previously established climate classification system, and substantially fewer than KPG.

When comparing all eight systems, WEC had the highest P, ~~and~~ PET, and Q coherence and similar ET~~, Δθ, and Q~~ coherence to KPG. Although it was not surprising that the WEC classification systems yielded the highest P and PET coherence, given these were the variables used to draw its zone boundaries, WEC also had much more uniform zone area~~pixel~~ distribution~~, similar zone fragmentation~~, half the number of~~fewer~~ zones, and required substantially fewer parameters when compared to KPG. Areas of similar water availability rates, as defined by low CV of Q, was best delineated by WEC, given that this system yielded the highest coherence for both TerraClimate Q and the independent validation source, GRUN Q. 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 within-zone Q variability as much, or because the present analysis considers locally-generated Q (P – ET) and not gauged streamflow as they did. However, the KHC framework used gauged streamflow data for system evaluation (Knoben et al., 2018), and this system yielded a distribution of CV of Q~~was~~ not statistically different than that of WEC, which yielded the lowest median CV of Q. ~~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 area~~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 Δθ coherence butand 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 was a previously unconsidered objective in creating and validating climate classification schemes. The climate classification system comparison presented here supportssBased on the climate classification system comparison presented here, theit can be longstanding assertionconclu ded that the primary mean ET drivers, water and energy (i.e., P and PET), ET drivers are important considerations for broad hydroclimate analyses. To delineate the landscape based on ET dynamics, the Budyko framework is a longstanding, well-vetted mechanism for estimating the evaporative index (ET/P) using the primary drivers of the water budget, PET and P, as represented by the aridity index (Budyko, 1974; Milly, 1994; Reaver et al., 2020a; Reaver et al., 2020b; Zhang et al., 2004). We conclude that, since hydroclimatewater budget coherence is bemostly achieved when P and PET are theincluded as governing principles of a zoning framework. However, when specifically evaluating ET dynamics, applyingusing an ET-based delineationframework could be usefulis most, especially if the objective of such a study is to distinctively evaluate factors that influence ET. It should be noted that boundaries created by ET drivers and not ET rates may influence the determined importance of such drivers, since intra-zone driver variability is likely to be reduced 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 ET drivers or 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 and did not distinguish between individual zones, which could be evaluated in a follow-up studysubsequent studies. 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, GLEAM ET and GRUN Q were also used as independent datasets and did not yield different results, which is likely due tothe 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 rates for all hydroclimate variables 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 ET), rather than the accumulated runoff from upstream contributing areas, which would be representative of gaged streamflow.~~

5 Conclusions

Formatted: Heading 1

The KPG system is the most widely used climate classification system, and this analysis revealed that it indeed has relatively high hydroclimatic coherence with respect to several variables, but it also has high spatial complexity as evidenced by multiple metrics, in addition to its 24-parameter requirement. It was concluded that WEC was either better than or not statistically different from all other previously established systems, including the KPG framework, in all assessed coherence metrics apart from Δt. Moreover, compared to KPG, WEC builds half the number of zones using only two parameters as input variables and delineates a more uniform zone area distribution to better facilitate meaningful spatial interpretations.

It is widely accepted that water and energy, chiefly in the form of ~~rainfall~~ precipitation and solar radiation, govern long term socioecological water availability at large spatiotemporal scales (Budyko, 1974; Berghuijs and Woods, 2016; Knoben et al., 2018; Sanford and Selnick, 2013). Several previous c~~C~~limate classification systems~~schemes~~ aimed~~seek~~ to represent ~~It is important that boundaries representative of~~ this water-energy interaction within bounded zones that ~~are drawn in order to~~ encompass similar hydroclimatic sensitivities (Knoben et al., 2018; 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, but xxx. However, it was concluded that WEC was either better than or not statistically different from the KPG framework in all assessed metrics.~~It was concluded here that WEC, using water and energy in the form of P and PET rates, was the best overall system for building zones that encompass similar Q rates. This suggests that the WEC scheme is valuable for assessing and predicting water availability changes given changes in water and energy. Therefore, WEC is the most relevant system for direct management understanding and application as it relates to hydroclimate dynamics.

This study proposes WEC as a new framework for large ~~-~~scale hydroclimate~~water budget~~ inquiries, and other large spatial scale research endeavors that may be influenced by hydroclimate systems that vary across the landscape. ~~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, since it is based on long ~~-~~term mean annual rates that have low ~~are less~~ susceptibi~~l~~eity 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.~~ This work is a promising pathway to regionalization within many different biophysical and socioeconomic contexts, clustering drivers to form zones of similar response variable sensitivities in order to more accurately extrapolate locally derived results and regional impacts of local management practices.

Formatted: Font: 10 pt
Formatted: Font: 10 pt
Formatted: Font: 10 pt
Formatted: Font: 10 pt
Formatted: Font: 10 pt
Formatted: Font: 10 pt

The WEC framework can thus inform regional to national scale management strategies in the effort to account for potential hydroclimate zone-dependent responses to climate and land cover changes.

Formatted: Indent: First line: 0.5"

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.

Budyko, M.I.: Climate and Life. Academic Press, 1974.

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.

Ghiggi, G., Humphrey, V., Seneviratne, S. I., and Gudmundsson, L.: GRUN: an observation-based global gridded runoff dataset from 1902 to 2014. *Earth System Science Data*, 11(4), 1655-1674, 2019.

Formatted: Indent: First line: 0.5"

Formatted: Font: Italic

Guan, Y., Lu, H., He, L., Adhikari, H., Pellikka, P., Maeda, E., and Heiskanen, J.: Intensification of the dispersion of the global climatic landscape and its potential as a new climate change indicator. *Environmental Research Letters*, 15(11), 114032, 2020.

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.

Hesselbarth MH, Sciaini M, With KA, Wiegand K, and Nowosad J. "landscapemetrics: an open-source R tool to calculate landscape metrics." *Ecography*, 42, 1648-1657, 2019.

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.

Li, D., Pan, M., Cong, Z., Zhang, L., and Wood, E.: Vegetation control on water and energy balance within the Budyko framework. *Water Resources Research*, 49(2), 969-976, 2013.

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.

Martens, B., Miralles, D. G., Lievens, H., Schalie, R. V. D., De Jeu, R. A., Fernández-Prieto, D., Beck, Hylke E., Dorigo, Wouter A., and Verhoest, N. E.: GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*, 10(5), 1903-1925, 2017.

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.

Formatted: Font: Italic
Formatted: Indent: Left: 0.5"
Formatted: Font: Italic
Formatted: Indent: First line: 0.5"
Formatted: Indent: First line: 0.5"
Formatted: Font: Italic
Formatted: Font: Italic
Formatted: Indent: First line: 0.5"

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.

Milly, P. C. D.: Climate, soil water storage, and the average annual water balance. *Water Resources Research*, 30(7), 2143 2156, 1994.

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.

[a]Reaver, N. G., Kaplan, D. A., Klammler, H., and Jawitz, J. W.: Reinterpreting the Budyko Framework. *Hydrology and Earth System Sciences Discussions*, 1-31, 2020.

[b]Reaver, N. G., Kaplan, D. A., Klammler, H., and Jawitz, J. W.: Analytical Inversion of the Parametric Budyko Equations. *Hydrology and Earth System Sciences Discussions*, 1-19, 2020.

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.

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., and Satoto, B. D.: Integration k-means clustering method and elbow method for identification of the best customer profile cluster. In *IOP Conference Series: Materials Science and Engineering* (Vol. 336, No. 1, p. 012017). IOP Publishing, 2018.

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~~

Formatted: Font: Italic
Formatted: Indent: First line: 0.5"
Formatted: Superscript
Formatted: Superscript
Formatted: Font: Italic
Formatted: Indent: First line: 0.5"
Formatted: Font: Italic
Formatted: Indent: First line: 0.5"
Formatted: Font: (Default) +Headings (Times New Roman)
Formatted: Font: (Default) +Headings (Times New Roman)
Formatted: Font: (Default) +Headings (Times New Roman)
Formatted: Font: (Default) +Headings (Times New Roman)
Formatted: Font: (Default) +Headings (Times New Roman)
Formatted: Indent: First line: 0.5"
Formatted: Font: (Default) +Headings (Times New Roman)

~~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, L., Hickel, K., Dawes, W. R., Chiew, F. H., Western, A. W., and Briggs, P. R.: A rational function approach for estimating mean annual evapotranspiration. *Water resources research*, *40*(2), 2004.

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.

Zhang, L., Potter, N., Hickel, K., Zhang, Y., and Shao, Q.: Water balance modeling over variable time scales based on the Budyko framework–Model development and testing. *Journal of Hydrology*, 360(1-4), 117-131, 2008.

Formatted: Font: (Default) +Headings (Times New Roman)

Formatted: Font: (Default) +Headings (Times New Roman)

Formatted: Font: (Default) +Headings (Times New Roman)

Formatted: Font: (Default) +Headings (Times New Roman)

Formatted: Indent: First line: 0.5"

Formatted: Font: (Default) +Headings (Times New Roman)

Formatted: Font: Italic