A simple framework to quantitatively describe monthly precipitation and temperature climatology

<table>
<thead>
<tr>
<th>Journal:</th>
<th>International Journal of Climatology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>JOC-15-0192.R2</td>
</tr>
<tr>
<td>Wiley - Manuscript type:</td>
<td>Research Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>23-Sep-2015</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Berghuis, Wouter; University of Bristol, Civil Engineering Woods, Ross; University of Bristol, Civil Engineering</td>
</tr>
<tr>
<td>Keywords:</td>
<td>climate, precipitation, temperature, similarity index, analytical framework, classification</td>
</tr>
</tbody>
</table>
A simple framework to quantitatively describe monthly precipitation and temperature climatology

Wouter R. Berghuijs*, Ross A. Woods

*Corresponding author

wb14708@bristol.ac.uk

Department of Civil Engineering, University of Bristol, Bristol, BS8 1TR, United Kingdom

Submitted to: International Journal of Climatology

Keywords: climate; precipitation; temperature; similarity index; analytical framework.

Key points:
- Global monthly precipitation and temperature climatology is described by a simple sinusoidal pattern
- Analytical framework with five indices describes mean monthly climate time-series
- The framework can provide a quantitative basis for climate descriptions among different sciences
Abstract

Climate descriptors and classifications are vital for ordering past, current and future climatic conditions. Yet, these parsimonious descriptors of climatic conditions only capture specific aspects of this climate signal, and lose all other information available in the observations. As a result, climate descriptions are often not physically insightful when they are applied in other studies. In this study, we show that a sinusoidal function with an annual period can adequately describe the vast majority of monthly precipitation and temperature climates around the world. This finding allows us to synthesize intra-annual monthly precipitation and temperature climatology using 5 indices that are easy to interpret. The indices describe (i) the mean precipitation rate ($\bar{P}$), (ii) the mean temperature ($\bar{T}$), (iii) the seasonal precipitation amplitude ($\delta_P$), (iv) the seasonal temperature amplitude ($\Delta_T$), and (v) the phase difference between the precipitation and temperature regimes ($\phi$). The combination of the 5 indices describes the relative time series of precipitation and temperature climatology, in contrast to earlier proposed similarity indices that only capture specific aspects of these time series. We demonstrate how the framework can reproduce many earlier proposed indices and classifications, and provide an example how the framework can be used to classify regions. We argue that the framework provides comprehensive insight into global climatology and can function as a quantitative conceptual basis for climate descriptions among different sciences.

1. Introduction

Climate is defined as the generally prevailing weather conditions of a region, often averaged over a 30-year period [WMO, 1989]. Climate descriptors and classifications summarize characteristics of the climate signal and thereby they help to bring
structure and order to the diversity of climates around the world. Similarity indices and classifications can order climate based on a single climatic characteristic (e.g. precipitation amount), or classify climate based on the combination of several climatic characteristics (e.g. precipitation amount and a temperature condition). Such descriptions help to delineate regions with specific climatic conditions. Because climate influences many factors, these descriptions are vital for understanding, explaining and predicting how regions differ in ecologic, water cycle, landscape and anthropogenic conditions.

Energy availability (temperature, net radiation, potential evaporation) and moisture availability (precipitation) form the core of many widely adopted climate descriptors and classifications [e.g., Köppen, 1936; Thornthwaite, 1931, 1948; Holdridge, 1967; Trewartha, 1968; Budyko, 1974; Alley, 1984; Kottek et al., 2006; Peel et al., 2007]. The mean intra-annual pattern of energy availability and moisture availability has a distinct imprint on a diverse range of factors, such as vegetation type [Köppen, 1936; Holdridge, 1967; Trewartha, 1968; Stephenson, 1992; Gonzalez et al., 2010], ecosystem productivity [Harris et al., 2000; Parton et al., 2012; Robinson et al., 2013], agricultural production [Kurukulasuriya et al., 2003; Deryng et al., 2011], carbon storage and release [Heimann & Reichstein, 2008], dissolved nutrient retention dynamics [Ye et al., 2012], evaporation rates [Wolock & McCabe, 1999; Berghuijs et al., 2014a], soil moisture storage [Milly, 1994; Seneviratne et al., 2010], snowpack and glacial dynamics [Woods, 2009; Bartholomew et al., 2010], droughts [Reynolds et al., 1999; Mishra & Singh, 2010; van Loon et al., 2014], river flow [Thornthwaite, 1931; Budyko, 1974; Petersen et al., 2012; Berghuijs et al., 2014b], aquatic communities [Poff et al., 1997; Kattwinkel et al., 2011], animal activity [Richardson,
Climate descriptors and classifications provide different ways to describe the climatic similarities and differences among places. Providing an overview of climate classification systems is beyond the scope of this study; we refer to Oliver [2005] for a list and description of several climate classification schemes. However, we identify that all classifications and descriptors have in common that they rely on indices that describe only a specific condition of the climate signal (e.g. the number of days a certain temperature is exceeded), which is sufficient to distinguish members of one class from another class. A descriptor is generally chosen because it is strongly linked to another character of interest (e.g. vegetation growing season). Reproduction of the mean monthly climate signal is not possible using solely these classification indices as all other information about the observed climate is not captured within the index. The fact that the indices lose a lot of information makes it difficult to use one set of climate descriptors or a specific classification across different scientific disciplines, as descriptors do not provide enough information to derive other characteristics of the climate. This lack of universal descriptors hinders climate descriptors from being effectively used to support advocated synthesis between different fields of study [Rodriguez-Iturbe, 2000; Harte, 2002; Weart, 2013]. The loss of information is also an important indicator that there is currently no comprehensive concise descriptor of what monthly climate patterns are actually occurring globally, as only some specific
conditions can be inferred from current climate classifications. Additionally, the loss of information prevents climate descriptors from being used to force mechanistic models. Falsifiable models are vital for testing physical understanding of interdependencies between climate and related factors.

If the monthly precipitation and temperature climatology consist of patterns that can be described with parsimonious mathematical functions, there is potential to develop descriptors of the monthly climate signal that maintain most of the information that is present in the observations, while using only a few numeric descriptors to characterize the climate. If the indices maintain enough information to describe the within-year patterns of the climate signal, these indices can be used to derive any characteristic of the climate that is a function of the intra-annual precipitation and temperature signal. This would significantly improve our ability to conceptualize what monthly climate patterns are occurring globally, and allow a similar reference framework among different sciences and studies.

The primary factors affecting local temperature are a location’s latitude, and altitude [Fleming et al., 1988]. The latitude strongly influences the seasonality of the mean monthly temperature and the mean temperature. The altitude mainly affects the mean temperature. Other controls on temperature include cloud cover variations [Tsushima & Manabe, 2001], land-cover [Feddema et al., 2005], soil-moisture [Seneviratne et al., 2010], distance from the ocean [Geerts, 2003], air and ocean currents [Jones et al., 2007], among other factors. Yet, due to the dominance of the seasonal change of the inclination of the sun, we hypothesize it is plausible to use a sinusoidal function with an annual period to describe the monthly temperature climatology of locations. The
use of a sinusoidal function to model seasonal temperature variation has been applied
in the contiguous US [Woods, 2009; Berghuijs et al., 2014b], is used for educational
purposes, or to model diurnal temperature variation [Snyder, 1985]. These studies did
not quantify to what degree a simple sinusoidal function is able to describe the mean
monthly temperature signal, nor was it applied on a global scale.

Several studies also used a sinusoidal function with an annual period to describe mean
monthly precipitation. This description for precipitation is used regionally in the
United States [Milly 1994; Woods, 2009; Berghuijs et al., 2014b], in parts of
Australia [Potter et al., 2005; Hickel & Zhang, 2006] and globally [Blöschl et al.,
2013]. Similar to temperature, these studies did not quantify to what degree a
sinusoidal function can describe the mean monthly precipitation signal. Additionally
the sinusoidal functions have been defined such that the seasonal precipitation
amplitude had an upper bound, with the consequence that climates containing several
months without precipitation could not be accurately described. This description
dismisses the possibility of accurately describing bimodal seasonal cycles around the
tropics, but is chosen to maintain parsimony.

In this study we examine to what degree observations of the monthly climate signal of
precipitation and temperature can be described by a sinusoidal function with an
annual period, and no upper bound to the precipitation seasonality. This can
potentially reveal to what degree there is a distinct pattern in the global monthly
climatic signal, which allows synthesizing most of the monthly precipitation and
temperature climatology using 5 indices that are straightforward to interpret
physically. Subsequently we demonstrate how the framework can reproduce other
characteristics of the climate signal and can be used to classify distinct climates. Finally, we discuss how this quantitative conceptualization can improve our understanding of global climatology and can provide a basis for climate similarity schemes among different sciences.

2. Methods

2.1. Data

We use monthly precipitation and surface temperature values for the period 1980-2009 from the Modern-Era Retrospective Analysis for Research and Applications improved set of land surface hydrological fields [MERRA-Land; Reichle et al., 2011]. The data have a 2/3-degrees longitude by 1/2-degrees latitude resolution. A quantitative comparison with the Global Precipitation Climatology Project (GPCP) [Huffman et al., 2009] indicates that MERRA-Land mean annual precipitation rates are lower than GPCP in parts of South America and central Africa, and higher than GPCP in Southeast Asia and along parts of the South American and African coasts [Reichle et al., 2011]. Although research has indicated that small biases of precipitation rates can occur regionally compared to other precipitation products, MERRA-Land reproduces precipitation well over land-surface [Reichle et al., 2011], especially the seasonal cycle [Kim et al., 2014]. We assess precipitation and temperature characteristics for all grid-cells where more than 50% of the cell area is classified as land.

2.2. Sinusoidal functions

The sinusoidal functions used to describe monthly precipitation and temperature are defined in Equation 1 and Equation 2:
where $P$ is the precipitation rate (mm/month), $T$ is the temperature ($^\circ$C), $t$ is the time (year), $\bar{P}$ is the mean precipitation rate (mm/month), $\bar{T}$ is the mean temperature ($^\circ$C), $\delta_P$ is the dimensionless seasonal precipitation amplitude (-), $\Delta_T$ is the seasonal temperature amplitude ($^\circ$C), $\tau$ is the duration of the seasonal cycle, set at 1 year, and the phase shifts (year) of temperature ($s_T$) and precipitation ($s_P$) are time offsets from a reference date, in this study set as Jan 1st. $\Delta_T$ and $\delta_P$ can range from zero to infinity. $s_T$ and $s_P$ can range from zero to one. Figure 1a and 1b illustrates an example climate according to the descriptions in Equation 1 and Equation 2.

In contrast to earlier studies [Milly, 1994; Potter et al., 2005; Hickel & Zhang, 2006; Woods, 2009; Blöschl et al., 2013; Berghuijs et al., 2014b], we remove the restriction that the maximum seasonality of precipitation has an upper bound of $\delta_P = 1$. This change allows description of climates where there are multiple months without precipitation. In the case that $\delta_P$ exceeds 1, equation 1 is generalized with a correction factor ($C_r$) to ensure that the average precipitation rate remains $\bar{P}$:

$$P(t) = \max\left(0, \bar{P} \cdot \left[1 + C_R + \delta_P \sin(2\pi(t - s_p)/\tau)\right]\right)$$ (3)

where,

$$C_r = -0.001 \cdot \delta_P^6 + 0.026 \cdot \delta_P^5 - 0.245 \cdot \delta_P^4 + 0.2432 \cdot \delta_P - 0.038$$ (4)

Figure 1c gives an overview of several precipitation regimes for a range of seasonal precipitation amplitudes. Figure 1d displays how the correction factor ($C_r$) varies as a function of the seasonal precipitation amplitude ($\delta_P$). The time-averaged value of $P(t)$
can deviate from $\bar{P}$ because $C_r$ is numerically approximated (see Supplementary Material, Figure S1).

To reduce the number of indices needed to characterize the climate, we introduce the phase difference between the precipitation and temperature regime ($s_d$). $s_d$ expresses to what degree the precipitation and temperature patterns are in phase, by quantifying how much earlier temperature peaks compared to the precipitation regime:

$$s_d = s_p - s_T \quad , \text{for } |s_p - s_T| \leq 0.5 \quad (5a)$$

$$s_d = -1 + (s_p - s_T) \quad , \text{for } (s_p - s_T) > 0.5 \quad (5b)$$

$$s_d = 1 + (s_p - s_T) \quad , \text{for } (s_p - s_T) < -0.5 \quad (5c)$$

$s_d$ can range from -0.5 (completely out of phase, P peaks before T), to 0 (completely in phase), to 0.5 (completely out of phase, P peaks after T). For the climate displayed in Figure 1a,b $s_d$ equals -0.40 [year].

The 5 indices needed to characterize the relative time series of mean monthly precipitation and temperature now are: (i) the mean precipitation rate ($\bar{P}$), (ii) the mean temperature ($\bar{T}$), (iii) the seasonal precipitation amplitude ($\delta_p$), (iv) the seasonal temperature amplitude ($\Delta_T$), and (v) the phase difference between the precipitation and temperature regimes ($s_d$).

**2.3. Derivation of other climate characteristics**

The 5 indices can be used to derive any climate characteristic that is a function of the mean within-year pattern of precipitation and temperature. Derived characteristics
can, for example, consist solely of temperature characteristics such as the duration that the temperature is above a certain threshold temperature ($T_c$):

$$t_T = \frac{-2 \sin^{-1} \left( \frac{T_c - \bar{T}}{\Delta T} \right) + \pi}{2\pi}$$  \hspace{1cm} (6)

Similarly, the duration that the seasonal precipitation is above a certain threshold precipitation ($P_c$) can be approximated by:

$$t_p = \frac{-2 \sin^{-1} \left( \frac{P_c - \bar{P}}{\bar{P} \cdot \delta_p} \right) + \pi}{2\pi}, \text{ for } \delta_p \leq 1$$  \hspace{1cm} (7)

These equations can be used to derive climate characteristics such as the number of frost days [Easterling, 2002], number of tropical days [Nastos & Matzarakis, 2008], number of dry months [Trejo & Dirzo, 2002], number of wet months [Trejo & Dirzo, 2002]. Similar expressions can be derived for indices such as precipitation seasonality [Walsh & Lawler, 1981], precipitation concentration index [Oliver, 1980], degree-day factor [Hock, 2003], and cooling degree month [Sturm et al., 1995].

Temperature and precipitation characteristics can be combined to express how much precipitation falls while a certain temperature condition is met. Examples are annual snowfall, the fraction of precipitation that falls as snow [Woods, 2009; Berghuijs et al. 2014b], and the precipitation in the growing season [Ylhäisi et al., 2010]. Woods [2009] showed how the fraction of precipitation falling below a certain temperature threshold ($T_0$) is calculated as follows:

$$f_s = f_s(\bar{T}, \delta_p) = \frac{1}{2} - \frac{\sin^{-1}(\bar{T})}{\pi} - \frac{\delta_p}{\pi} \sqrt{1 - \bar{T}^2}, \text{ for } \delta_p \leq 1$$  \hspace{1cm} (8a)

where,

$$\delta_p^* = \delta_p \cdot \text{sgn} \left( \Delta_T \right) \cdot \cos(2\pi \cdot s_d)$$  \hspace{1cm} (8b)
Because the indices describe the character of widely used sinusoidal functions, analytical solutions can be derived for other precipitation, temperature or combined characteristics. The widely adopted classifications of [Köppen, 1936; Thornthwaite, 1931, 1948; Holdridge, 1967; Trewartha, 1968; Budyko, 1974; Peel et al., 2007] can also be reproduced, but this requires more laborious expressions, sometimes including calculation of potential evaporation based on mean monthly temperature [e.g. Hamon, 1961].

2.4. Calibration and evaluation

To test the adequacy of the sinusoidal function with an annual period for the description of the precipitation and temperature climate we define two objective functions that express the goodness of fit for the temperature and precipitation approximations:

\[
X_T = \sum_{t=1}^{12} \frac{|T(t) - T_t|}{\Delta_T} \quad (9)
\]

\[
X_P = \sum_{t=1}^{12} \frac{|P(t) - P_t|}{P} \quad (10)
\]

where \(X_P\) expresses the mean monthly precipitation error normalized by the average precipitation rate (\(\bar{P}\)). When the error, \(X_P\), is 0 the sinusoidal function is a perfect fit to the observed precipitation value \(P_t\).

The value of \(X_P\) expresses to what degree the monthly precipitation deviates relative to the mean monthly value observed at that location. \(X_T\) expresses the mean monthly temperature error (\(^{0}\)C), which is the mean absolute error in the temperature approximation (\(T_t\) is the observed temperature). The coefficients of Equation 1 and 2
are obtained by the Simplex search method [Nelder & Mead, 1965] of MATLAB’s fminsearch to minimize $X_T$ and $X_P$ [Lacouture & Cousineau, 2008]. For both the optimizations $P$ and $T$ are fixed according to the long-term average observed values; only the seasonal amplitude and phase shift are calibrated. The objective functions are chosen because they have the same units as the observed and described signal, and they can be interpreted without information on the variance in the observations.

3. Results

We first provide an overview of the global monthly climatology according to the description by the sinusoidal functions. Subsequently we evaluate in more detail the appropriateness of the sinusoidal function to describe the monthly precipitation and temperature climatology. Finally we assess the correspondence of characteristics of the climate derived from the 5 indices and characteristics of the climate directly derived from the observations.

3.1. Global monthly climatology

Figure 2 displays the global occurrence of the mean temperature ($\bar{T}$), the seasonal amplitude of temperature ($\Delta_T$), the phase shift of the temperature regime compared to January 1st ($s_T$), and the temperature error ($X_T$) in approximating the observed data by a sinusoidal function. The mean temperature for the assessed grid cells varies between -28.1 and 37.1°C. The seasonal temperature amplitude also varies strongly across the grid cells with a maximum $\Delta_T$ of 32.5°C. The approximation of the monthly temperature signal gives an average temperature error ($X_T$) of 0.85°C, with a standard deviation of 0.44°C. This error is relatively small compared to the mean seasonal amplitude of temperature, $\Delta_T$, of 12.8°C (median = 12.8°C). The regions where the
temperature error is large coincide with the regions where the seasonal temperature amplitude ($\Delta T$) is also large or with regions with a highly seasonal precipitation regime. In the areas with a seasonal precipitation regime the seasonal change in soil moisture can be a strong control on the surface energy balance, thereby affecting the intra-annual temperature pattern; this is one possible cause of the larger errors.

Figure 3 displays the global occurrence of the mean precipitation rate ($\bar{P}$), the seasonal precipitation amplitude ($\delta_p$), the phase shift ($s_p$), and the precipitation error ($X_p$). The precipitation rate ranges from a minimum of 4 mm/y, to a maximum of 10561 mm/y. The global mean precipitation rate is 706 mm/y (median = 501 mm/y). The seasonality of the precipitation varies regionally; $\delta_p$ has an average value of 0.80 (-) (median = 0.63), but can locally be as high as 4.7 (-). The approximation of the seasonal precipitation signal, on average, leads to a mean absolute error of the monthly precipitation of $X_p = 0.17$ (-), with a standard deviation of 0.12. With a mean seasonality of precipitation ($\delta_p$) equal to 0.80 this suggests that, on average, the within-year seasonality of precipitation is largely captured by the sinusoidal description.

Figure 4 displays the phase difference between the precipitation and temperature regimes ($s_d$). This phase difference is for most regions relatively close to 0 indicating that precipitation amounts are the highest during the warmer months at the given location. In some regions of all the continents the precipitation amounts are highest during the cool season.

3.2. Assessment of errors
To improve understanding of the ability of the sinusoidal function to describe the precipitation regime we highlight how well the description works as a function of precipitation characteristics, and how the errors vary between regions.

The regional differences in errors indicate that the sinusoidal function is not always an informative description of the monthly precipitation regime as the approximation can show relatively high error values (see map of $X_P$ in Fig 3). The percentage of grid cells where $X_P$ is larger than 0.30 (-) is 12.6%. Of these grid cells 69.0% are located in dry regions with annual precipitation below 300 mm/y. The regions with very low precipitation rates (<300 mm/y) sometimes have too few precipitation events to identify a smooth seasonal pattern. Other regions where high precipitation errors are observed are mostly in highly seasonal precipitation regimes ($\delta_p > 1.0$). Figure S2 (Supplementary Material) delineates the grid cells in discrete classes based on the $X_P$, $\delta_p$, and $P$ values.

The grid cells where $X_P$ is larger than 0.3 are only located in a limited number of regions (See Figure S2). Reasons for these high $X_P$ values vary regionally. Table S1 gives a point wise description per region that shows high ($X_P > 0.3$) values. These descriptions indicate the regional reasons for the higher error value and should improve understanding of the regional adequacy of the hypothesis that the monthly precipitation pattern can be described with the sinusoidal function. The sinusoidal approximation is not informative in regions with a bimodal rainfall pattern such as southwestern United States and the Horn of Africa.
Figure 5 gives an overview of the measured and modeled temperature and precipitation regimes, to give qualitative understanding how well the approximations describe the observed regimes. For different ranges of precipitation seasonality we have selected individual grid-cells whose error value is the 25th percentile, median and 75th percentile for that category, in order to view seasonal regimes where the sinusoidal functions produce high, medium and low errors. For the temperature regimes the 75th percentile and better fits all have a very good correspondence between the sinusoidal function and the actual observations. Hence the sinusoidal functions also visually appear very suitable for describing the monthly temperature pattern. For the precipitation patterns the correspondence between the sinusoidal function and the actual observations is lower. Although we visually inspected the measurements of all grid cells, we were not able to identify a more suitable simple mathematical function to describe the measured precipitation regime in a similar parsimonious manner.

3.3. Comparison of framework and data-derived climate characteristics

We evaluate the ability of the framework to reproduce specific climatic characteristics. This gives an indication of the suitability of the framework to provide a common reference for studies that are interested in specific climate characteristics. We compare characteristics of the climate as assessed by the 5 similarity indices and characteristics of the climate directly derived from the data. The derived indices include temperature-based, precipitation-based and combined temperature and precipitation characteristics. The characteristics of the climate assessed, and their definitions are listed in Table 1. Given the large number of grid cells involved, the correspondence between the analytically derived and the data-derived values is
summarized by the slope of a linear regression (indication of accuracy), and the $R^2$-value of the linear regression (indication of precision). The analytically derived value is used as the explanatory variable. The combination of the linear regression slope and the $R^2$-value expresses how all the information contained in these similarity indices can be reproduced with the reference framework.

The slopes of the linear regression approach one for most climate indices, with $R^2$-values also approaching one (see Table 1). This indicates enough information is captured within the framework to accurately and relatively precisely reproduce a variety of widely used climate indices. Temperature indices (duration frost season, duration growing season, cooling degree month) have the highest $R^2$-value, which is also expected considering the good fit between temperature observations and descriptions. The $R^2$-value for precipitation characteristics (dry period [Peel et al., 2007], wet period [Peel et al., 2007] and precipitation seasonality [Walsh & Lawler, 1981]) decrease slightly, but slopes still are close to one with $R^2$ also close to one.

One variable to highlight is the precipitation seasonality index as defined by Walsh & Lawer [1985]. The slope of the linear regression gives a value of 0.90 which confirms that most of the precipitation variability is captured by the sinusoidal function. For combined characteristics (fraction of precipitation falling as snowfall [Woods, 2009], growing season precipitation [Ylhäisi et al., 2010], Holdridge aridity index [Holdridge, 1969; Shen et al., 2011]) the performance decreases again, but still $R^2$-values are around 0.90 and the slope of the linear regression still approaches one. The correspondence with the Köppen main class according to the definitions used in Peel et al. [2007] gives a 99.81% correspondence between derived classes, indicating that this widely used classification scheme can be reproduced as well.
4. The framework as a classification tool

The framework can be used as a classification tool to characterize or cluster climate based on the five indices using the notation: $[P, \bar{T}, \delta_p, \Delta_T, s_d]$. An example grid-cell in New Zealand [43.5300°S, 172.6203°E] has the characteristics [662.9, 6.8, 0.30, 6.86, -0.01]. When regions with comparable climates are defined, the single values can be replaced by the associated minimum and maximum value, e.g. [600/800, 5/10, 0.1/0.4, 4/8, -0.25/0.25]. Another type of classification can make the different components dependent on another, e.g. $[(600+30\bar{T})/(800+30\bar{T}), 5/10, 0.1/0.4, 4/8, -0.25/0.25]$.

As an example, we classify the land surface into different climatic regions. The four indices $[P, \bar{T}, \delta_p, \Delta_T]$ are divided into tertiles with an equal number of grid-cells per group; per index there is a group of low, medium and high values. The 5th index ($s_d$) is divided into a group of small and large phase differences, again with an equal number of grid-cells. Climate classes are constructed based on the combination of the above-mentioned groups, leading to $3^4 \times 2 = 162$ climate classes. However, not all combinations of groups occur, resulting in 120 classes with grid-cells assigned. Figure 6 displays the class boundary conditions (bottom right), and the spatial distribution of classes with more than 250 grid-cells. Although the current example classification does not have a specific purpose beyond providing an example, the framework allows classifying climate groups quantitatively, while maintaining the qualitatively easy to interpret character (e.g. cold, wet, high rainfall seasonality, medium temperature seasonality, out of phase). Table S2 (Supplementary Material) provides an overview of all classes and the number of grid-cells assigned per class.
5. Discussion

5.1. Is the sinusoidal function suitable to describe monthly climatology?

We aimed to develop descriptors of the intra-annual precipitation and temperature climate that maintain most of the monthly information that is present in the observed signal, while using a limited number of descriptors to characterize the climate. By identifying that most of the climates around the world can be described by a sinusoidal pattern with an annual period, both for monthly precipitation and temperature, simple analytical functions appear to be very suitable for this purpose. The most parsimonious description that still acknowledges intra-annual variation of precipitation and temperature consist of 5 indices: here described by $\bar{P}$, $\bar{T}$, $\delta_p$, $\Delta_T$ and $s_d$. More parsimonious descriptors integrate these dimensions and therefore by definition lose information.

The systematic comparison of the analytical model performance with the observed data indicates regional differences in the adequacy of the sinusoidal function for describing the observed monthly regimes. For the temperature climatology, Figure 5 shows that the seasonal pattern is well described by the sinusoidal function, as the mean absolute error ($X_T$) is much smaller than the within-year variability of the temperature regime ($\Delta_T$). Considering that the climatic descriptors should be parsimonious and easily understandable, we have not identified an opportunity to improve on the sinusoidal description to describe the monthly temperature pattern, while still maintaining the parsimony and simplicity of the current sinusoidal description.
The goodness of fit ($X_P$) of the precipitation regimes indicates that the sinusoidal function for most regions provides a reasonable approximate for the precipitation regimes. High errors, with few exceptions, occur either in the very dry places ($P<300$ mm/y), or in places with hyperseasonal precipitation ($\delta_P > 1$). The significant percentage of grid cells with a hyper seasonal precipitation regime indicates that previous characterizations with an upper bound of 1.0 for the seasonality [Milly, 1994; Potter et al., 2005; Hickel & Zhang, 2006; Woods, 2009; Blöschl et al., 2013; Berghuijs et al., 2014b] are not suitable for characterizing the global monthly precipitation climatology, though it can be applied in some regions.

For the precipitation pattern the error in the sinusoidal approximation can be regionally relatively high, and there is more room for a refined mathematical description, especially in regions with a clear bimodal monthly precipitation regime. In dry regions the monthly precipitation rates are based on a limited number of precipitation events, so there is often no smooth mean monthly pattern. Improvement of the parsimonious precipitation description will consequently be very difficult for regions with low precipitation rates. The data we used for the fitting of our framework are interpolated, which may impact the performance of the framework. This may be particularly important in arid data poor regions, where there is the possibility of poor performance due to inaccurate data interpolation.

The balance between providing an appropriate and detailed description of the climate and providing a simple parsimonious understandable description depends on the purpose of the frameworks. Earlier studies used more detailed sinusoidal functions to
describe regional climatic gradients [Horn & Bryson, 1960], or suggested to regionally change the period of the seasonal cycle to half a year [Milly, 1994]. Although such refinements may improve the correspondence of the analytical function and the observed climate signal, they also require more indicators to describe the climate and are physically less easy to interpret. The most detailed description of monthly precipitation and temperature values, are the actual observed values. However, description of this information requires two numbers for every month to characterize the climate, and thus is inappropriate to characterize the climate in a quickly understandable way when the climate of many different locations needs to be characterized or compared.

Whether the errors introduced by the approximation are problematic completely depends on the purpose the framework is used for. In context of studies that use other climate indices or climate classes, the suitability of the mathematical approximation is underpinned by the high correspondence between derived climate characteristics with the framework and climate characteristics based on measurements. This indicates the amount of information lost by summarizing the monthly climate with the 5-indices is very small as the reproduction of other variables is well maintained. Comparison with the precipitation seasonality index of Walsh & Lawler [1985] indicates that on average most of the variability of mean intra-annual precipitation is captured within the description. However, some information (the error) is lost and not available for detailed assessments when only the 5 climate descriptors are used.

Evaluation of the descriptors of the monthly climate is only performed for grid-scale precipitation and temperature, which does not take into account sub-grid variability.
Hence the hypothesis is not tested at sub-grid scales. Further testing and mapping for sub-grid variability is left for future work. Yet as the hypothesis originates from applications at local sites, it is not expected that at sub-grid scales the performance will change significantly. The proposed description is scale-independent in its application and hence a potentially useful way to characterize any place at any scale or to characterize the variability or mean of a single unit, at other than grid-scales (e.g. a river basin).

5.2. What insight can the similarity indices give?

By identifying that the mean monthly climatology in many parts of the world can be described by a sinusoidal pattern we simplified the mean climate signal into five dimensions, which has multiple uses, and limitations. A clear limitation of the framework is the loss of detail available in the observed signal, such as between year variability, short-term variability etc. A description of mean seasonal climate does not incorporate, but can be expanded by, descriptors that characterize precipitation characteristics such as storminess and inter-annual variability. The 5 indices are thus currently not adequate for forcing mechanistic models or studies that require detailed data (e.g. daily) of the temporal climate conditions. Additionally the error of precipitation and/or temperature can be too large to highlight climatic differences in regional studies that compare climatologically almost equivalent sites. Therefore the descriptors will not always be suitable for local assessments that require as much detailed information as possible. These limitations are intrinsic properties of any climate classification and climate descriptors.
The framework is rather intended as a tool to order global dominant features of monthly precipitation and temperature climatology. Because our description provides a good approximation to the time series of observed climatology, our framework can provide a much more comprehensive understanding on what monthly climate patterns are occurring globally compared to earlier parsimonious climate descriptors. This more comprehensive way of describing monthly climatology has multiple distinct advantages compared to the classifications and indices that describe only specific characteristics of the monthly climatology but lose all other information obtained in the observations.

The framework makes it conceptually much easier to describe the actual physical gradients of monthly climatology between two places. The similarity indices we propose all have a well-defined, unambiguously interpretable definition. Many previous similarity indices and classifications [e.g., Köppen, 1936; Kottek et al., 2006; Peel et al., 2007] are rather a combination of numerical indices where the physical gradient between places cannot be expressed within a quantitative manner or sometimes even conceptual manner. Expressing these physical gradients between places in a conceptually easy manner is not only valuable for education purposes but also can assist in exposing physical gradients that underpin differences and similarity between places for research purposes.

Sanderson [1999] advocated for a novel classification of the world climates. “Modern textbooks continue to use the 100-year old Köppen classification of climates [Köppen, 1936], which is based on de Candolle’s vegetation groups, themselves based on the five climatic zones of the ancient Greeks.” The limited physical information contained
in the similarity indices systems remains a barrier to give insight into the climatic
similarity and differences between places. Additionally, because all classification
systems have their specific purpose (e.g. cluster vegetation similarity) it is difficult to
use the indices across different studies and sciences. For example, the limited
quantitative information on the intra-annual climate conditions that is contained in
Köppen’s classification makes it unsuitable as a common reference framework for
many different studies. Because of the much smaller loss of information in our
framework and the ability to reproduce previous classifications we argue our
framework can generate a conceptual step forward in characterizing the within year
variations of the climate, where climatic differences between places are easily
expressed.

We argue that our framework can provide a quantitative conceptual basis for climate
descriptions among different sciences. Because the analysis of section 3.3 indicates
the approximated regimes can accurately reproduce other climate descriptors, the
framework can provide a holistic picture of the monthly precipitation and temperature
climatology. Goals of previous climate indices [e.g. Walsh & Lawler, 1985;
Easterling, 2002; Trejo & Dirzo, 2002; Nastos & Matzarakis, 2008] and
classifications have been to organize the climate such that specific climate-dependent
characteristics occur in a region [e.g. Köppen, 1936; Holdridge, 1967; Trewartha,
1968]. In contrast, our framework provides five climate dimensions that in a simple
manner can characterize under which monthly precipitation and temperature
climatology the case specific assessments occur. Many climate indices and climates
of classification schemes are derived from the mean intra-annual precipitation and
temperature pattern. Consequently, these indicators can all be expressed in terms of
the 5 proposed climate indices. The framework can thus provide a common reference scheme to describe climatic conditions, and thereby better highlight climatic similarity and differences between places.

Classification, the delineation of groups with similar characteristics, is always purpose specific, except when there are discrete differences between the observed items, such as classes in Linnaean taxonomy [Linnaeus, 1788], elements of the periodic table [Mendeleev, 1869], and turbulent and laminar flow in fluid mechanics [Belanger, 1828]. Our framework rather uses continuous numbers to describe the character of climate where discrete classes are based on more purpose-specific conditions. The 5-dimensions form a continuum in which we can only subdivide by putting in artificial boundaries. We provided an example based on arbitrarily chosen class boundaries, which classified the land surface into different climatic regions. Although this classification does not have a specific purpose beyond providing an example, it shows how the framework allows classifying climate groups quantitatively, while maintaining a qualitatively easy to interpret character.

The fact that the full within-year climatology is described using the indices means that the indices can force mechanistic models [e.g. Woods, 2003, 2009; Potter et al., 2005]. This characteristic, combined with the notion that the indices can express the climatic gradients between several places, make it potentially a powerful tool to combine simple mechanistic and falsifiable models and large scale climate classifications. Additionally, the framework may provide a useful tool to characterize past or future climatic change or variations in a holistic, physically easily interpretable way compared to using changes in the discrete Köppen climate classes [Rubel &
Kottek, 2010; Chen & Chen, 2013], changes in speed of change of Köppen climate
classes [Mahlstein et al., 2013], changes in precipitation concentration [Luis et al.,
2011], and changes in mean-annual climatology [Greve et al., 2014].

6. Conclusions

Climate is a key factor in many sciences and determines the diversity of many biotic
and abiotic factors around the world. Climate descriptors and climate classifications
are widely used tools to synthesize climatic conditions in a parsimonious manner and
are vital for understanding, ordering and describing the global climatic diversity. The
diversity of climates around the world makes it difficult to produce parsimonious
descriptors of climatic conditions that still maintain most of the information present in
the observed signal. Consequently, climate descriptors and classifications only
describe a specific aspect of the climate signal, or they have a qualitative character.
As a result, climate descriptions are often physically not very insightful when they are
applied in other sciences or studies.

In this study we showed that a sinusoidal function with an annual period can describe
most of the monthly precipitation and temperature patterns. The mean absolute
temperature error of the sinusoidal function is 0.85 (°C), which is an order of
magnitude smaller than the mean intra-annual variation of temperature. Similarly, the
mean monthly error of precipitation is on average below 0.18 [-]; high error values
mainly occur in regions with low precipitation rates or in regions with a very seasonal
precipitation regime.
This finding allows us to synthesize most of the monthly precipitation and temperature patterns using 5 indices that are physically easy to interpret. The indices describe (i) the mean precipitation rate ($\overline{P}$), (ii) the mean temperature ($\overline{T}$), (iii) the seasonal precipitation amplitude ($\delta_p$), (iv) the seasonal temperature amplitude ($\Delta_T$), and (v) the phase difference between the precipitation and temperature regime ($s_d$). The combination of the 5 indices summarizes the relative time series of mean monthly precipitation and temperature. Quantitative comparison of characteristics of the climate as assessed by the 5 similarity indices and directly derived from the original climatic data shows good correspondence. This indicates the framework is able to give a holistic picture of climatic conditions, but also indicates its ability to provide a common reference framework for studies that are interested in more specific climate characteristics. As an example, we classify the land surface into different climatic regions based on the five indices. Although this classification does not have a specific purpose beyond providing an example, it shows how the framework allows classifying climate groups quantitatively, while maintaining a qualitatively easy to interpret character.

Hence the proposed framework provides a basis to summarize the global diversity of monthly precipitation and temperature climatology within a 5-dimensional space. This allows expressing the climatic diversity in a simple and understandable manner, while the quantitative character of the monthly climate signal is maintained. Because a wide range of climatic classification and similarity indices can be brought back to the 5-dimensional space the framework can be used as a common reference scheme among different sciences.
Acknowledgements

W.R.B. was supported by the Queen’s School of Engineering (University of Bristol) Ph.D. scholarship. We thank Murray Peel and one anonymous reviewer for their constructive comments that helped to improve the manuscript.
References


Holdridge, L.R., 1967. Life Zone Ecology, Tropical Science Center, San José, Costa Rica.


http://mc.manuscriptcentral.com/joc


Petersen, T., N. Devineni, and A. Sankarasubramanian (2012), Seasonality of monthly runoff over the continental United States: Causality and relations to mean annual and mean monthly distributions of moisture and energy, J. Hydrol., 468–469, 139–150.


relationships. Journal of Applied Meteorology and Climatology, 49(12), 2559-2573.


http://mc.manuscriptcentral.com/joc
List of Tables

Table 1: Correspondence of the climate characteristics calculated with the analytical model and climate characteristics calculated with MERRA-Land data.

List of Figures

Figure 1: Conceptual description of monthly climate according the framework; example of a precipitation regime (Fig. 1a), a temperature regime (Fig. 1b), several precipitation regimes for a range of seasonal precipitation amplitudes (Fig. 1c), and correction factor \( C_r \) as a function of the seasonal precipitation amplitude (\( \delta_p \)) (Fig. 1d).

Figure 2: The mean temperature (\( T \)), the seasonal temperature amplitude (\( \Delta_T \)), the phase shift (\( s_T \)), and the monthly temperature error (\( X_T \)).

Figure 3: The mean precipitation rate (\( P \)), the dimensionless seasonal temperature amplitude (\( \delta_p \)), the phase shift (\( s_p \)), and the monthly precipitation error (\( X_P \)).

Figure 4: The phase difference between the precipitation and temperature regime (\( s_d \)).

Figure 5: MERRA-Land observation (bar) and analytically approximated (dashed line) monthly temperature and precipitation climatology of the individual grid-cells that fall at 25\(^{th}\) percentile, median, and 75\(^{th}\) percentile values of different \( \delta_p \) and \( \Delta_T \) intervals.

Figure 6: Climate classification of the world based on the five climatic indices: \( [\bar{P}, T, \delta_p, \Delta_T, s_d] \). Observations are divided into tertiles with equal membership for every index. The figure shows class boundary conditions (bottom right), and the spatial distribution of classes. The phase shift \( s_d \) is delineated into two classes: high (H) and...
low (L) values are merged into one group (L) because of the cyclic behavior of $s_d$.

Classes with less than 250 grid-cells are not presented.
Figure 1. Conceptual description of monthly climate according the framework; example of a precipitation regime (Fig. 1a), a temperature regime (Fig. 1b), several precipitation regimes for a range of seasonal precipitation amplitudes (Fig. 1c), and correction factor $C_r$ as a function of the seasonal precipitation amplitude ($\delta_p$) (Fig. 1d).

193x193mm (300 x 300 DPI)
Figure 2. The mean temperature ($T$), the seasonal temperature amplitude ($\Delta T$), the phase shift ($s_t$), and the monthly temperature error ($X_t$).
Figure 3: The mean precipitation rate (P), the dimensionless seasonal temperature amplitude (δP), the phase shift (sP), and the monthly precipitation error (XP).
Figure 4. The phase difference between the precipitation and temperature regime ($s_d$).
Figure 5. MERRA-Land observation (bar) and analytically approximated (dashed line) monthly temperature and precipitation climatology of the individual grid-cells that fall at 25th percentile, median, and 75th percentile values of different δₚ and Δₜ-intervals.
<table>
<thead>
<tr>
<th>Climate descriptor</th>
<th>Description</th>
<th>Definition</th>
<th>Slope linear regression</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration frost season</td>
<td>Period that mean temperature is below freezing point [Easterling, 2002].</td>
<td>$\sum t(T(t)&lt;0)/\sum t$</td>
<td>0.9970</td>
<td>0.9890</td>
</tr>
<tr>
<td>Duration growing season</td>
<td>Period that mean temperature is above a certain threshold, here set at 8 °C.</td>
<td>$\sum t(T(t)&gt;8)/\sum t$</td>
<td>1.0115</td>
<td>0.9899</td>
</tr>
<tr>
<td>Cooling degree month</td>
<td>Time-accumulated winter temperature exceeding a temperature threshold [Sturm et al., 1995].</td>
<td>$\sum(Tc - T(t))$, if $T(t) &lt; Tc$</td>
<td>0.9981</td>
<td>0.9999</td>
</tr>
<tr>
<td>Dry period</td>
<td>Period that the mean precipitation rate is lower than 60 (mm/month) [Peel et al., 2007].</td>
<td>$\sum t(P(t)&lt;60)/\sum t$</td>
<td>1.0137</td>
<td>0.9592</td>
</tr>
<tr>
<td>Wet period</td>
<td>Period that the mean precipitation rate is higher than 60 (mm/month) [Peel et al., 2007].</td>
<td>$\sum t(P(t)&gt;60)/\sum t$</td>
<td>0.9672</td>
<td>0.9688</td>
</tr>
<tr>
<td>Precipitation seasonality</td>
<td>Mean deviation of monthly precipitation compared to the mean annual precipitation [Walsh &amp; Lawler, 1981].</td>
<td>$(\sum</td>
<td>P(t) - P(t)/12</td>
<td>)/\sum P(t)$</td>
</tr>
<tr>
<td>Fraction of precipitation falling as snowfall</td>
<td>Precipitation falling as snowfall (as derived by a temperature threshold) divided by the total amount of precipitation [Woods, 2009].</td>
<td>$(\sum P(T(t)&lt;1))/\sum P$</td>
<td>1.0463</td>
<td>0.8997</td>
</tr>
<tr>
<td>Growing season precipitation</td>
<td>Annual amount of precipitation falling when growing season conditions ($T &gt; 8 °C$) [Ylhäisi et al., 2010].</td>
<td>$\sum P(T(t)&gt;8)$</td>
<td>1.0367</td>
<td>0.9262</td>
</tr>
<tr>
<td>Holdridge aridity index</td>
<td>Climatic water availability in each part of the year, defined as the ratio of the temperature to the annual precipitation [Holdridge, 1969; Shen et al., 2011].</td>
<td>$(58.93 \sum T(t(T&gt;0))/P$</td>
<td>1.0121</td>
<td>0.9845</td>
</tr>
<tr>
<td>Köppen-Geiger main-class</td>
<td>Percentage of grid-cells that are assigned to the correct Köppen-class according to the definitions of Peel et al [2007].</td>
<td>See Table 1 in Peel et al [2007].</td>
<td>-</td>
<td>99.81%</td>
</tr>
</tbody>
</table>