Agricultural Water Management 217 (2019) 282–291

Contents lists available at ScienceDirect

Agricultural Water Management

journal homepage: www.elsevier.com/locate/agwat

Quantifying the probabilistic divergences related to time-space scales for inferences in water resource management

Clement D.D. Sohoulande[a,*], Kenneth Stone[a], Vijay P. Singh[b]

[a] USDA-ARS Coastal Plain Soil, Water and Plant Conservation Research Center, 2611 West Lucas Street, Florence, SC 29501, USA
[b] Biological and Agricultural Engineering Department & Zachry Department of Civil Engineering, Texas A and M University, Scoates Hall, 2117 TAMU, College Station, TX 77843, USA

ARTICLE INFO

Keywords:
Time-space variability
Scaling
Kernel density modelling
Kullback–Leibler divergence
Water resources management

ABSTRACT

Studies focusing on time-space scaling issues have not yet established guidelines for choosing a time-space scale for water resources management. This study attempts to develop a methodology to understand the deviations induced by the scaling framework on water resource variables. The study addresses eight variables: streamflow, groundwater level, precipitation, air temperature, soil moisture, solar radiation, wind speed, and relative humidity. For each variable, a 10-year (2001 to 2010) daily time-series data were retrieved from weather and hydrologic stations selected across Oklahoma. The methodology entailed three phases, including a variable rescaling, a probability analysis, and a divergence quantification. Hence, daily time-series of the selected variables were rescaled to weekly, biweekly, monthly, bimonthly, and trimonthly scales. To derive the probability density function of each variable, a kernel density estimator was applied to the rescaled time-series. The Kullback–Leibler divergence was used to evaluate the deviations pertaining to the time-space scaling. Results showed meaningful divergences that were a consequence of time-space scale variations. However, the sensitivity of variables to the scaling framework differed. For instance, air temperature was less sensitive to time-space scaling, while wind speed, soil moisture, and precipitation were highly sensitive. Correlation analysis showed that correlations tended to gradually increase with the increase in scale, even though the strength of the relation varied, depending on the variable involved. The methodology was tested via a case study where data obtained from a location outside of the study region were used to address the scaling effect on the Normalized Difference Vegetation Index's response to precipitation and soil moisture. It was concluded that understanding the divergences associated with time-space scales was critical for interpreting results in water resource studies.

1. Introduction

In water resource studies, fine scale data (e.g. daily, hourly) are often needed, because they can be used to rescale or to achieve multiscale water resource assessments (Maurer et al., 2007). However, such datasets generally involve substantial funding, time, equipment, tedium, and labor. Downscaling techniques are often used as inexpensive ways to generate finer scale data (Koutsouris et al., 2017). Even though a variable like air temperature often shows a high downscaling capability (Schoof and Pryor, 2001), most of the techniques used for spatio-temporal downscaling are associated with biases (Malone et al., 2017). Hence, prior to choosing a specific time-space scale in a water resources study, an evaluation of all the constraints and their potential impact on the inferences is needed. However, statistical approaches for choosing time-space scales are not sufficiently developed for water resources variables. This is perhaps due to the paucity of guidelines on scale choices, but often multiscale data are unavailable. Yet, various time-scales are often used somewhat arbitrarily. The decision to use a specific time-space scale induces bias. Several authors used water resources related variables (e.g. surface air temperature, precipitation, soil moisture) as inputs to quantify complex phenomena such as evapotranspiration (Olivera-Guerra et al., 2017) and plant growth (Sohoulande and Singh, 2015). In these cases, it is crucial to understand the bias associated with the input variables, as it may impact interpretation (Perrette et al., 2013; Clark et al., 2004). Especially for water resources management, knowing the types of bias induced by time-space scales is important for quantifying the degradation of freshwater quality and quantity. Because the global distribution of water quantity and quality are uneven (Oki and Kanae, 2006; Vörösmarty et al., 2000), various mitigation strategies have been proposed worldwide (Pires

* Corresponding author.
E-mail addresses: clement.sohoulande@ars.usda.gov (C.D.D. Sohoulande), ken.stone@ars.usda.gov (K. Stone), vsingh@tamu.edu (V.P. Singh).

https://doi.org/10.1016/j.agwat.2019.03.004
Received 30 January 2019; Accepted 1 March 2019
Available online 12 March 2019
0378-3774/

et al., 2017; Simonovic, 2017). These strategies are often based on probabilistic inferences and their effectiveness is dependent on an understanding of water resources dynamics in time and space (Medvigy et al., 2009; Wiens, 1989).

To date, in water resources management, probabilistic tools are often used in decision making at both field and watershed levels (Prada et al., 2017). Therefore, it is important to determine the influence of time-space scales on the probability distributions of water resource related variables in order to better respond to water crises (Clark et al., 2004). The present study presents a method for quantifying the probabilistic divergence related to time-space scaling. The developed method uses kernel density (Qin et al., 2011) and Kullback–Leibler divergence (Polani, 2013; Kullback and Leibler, 1951) to determine the effect of time-space scaling on statistical inferences involving water resources variables. The objectives of the study therefore are to: (i) develop a methodology for choosing a time-space scale for water resource management; (ii) clarify the probabilistic divergence induced by time-space scales in water resource related variables; and (iii) pinpoint the potential effect of time-space scales on statistical inferences and applications in water resources management.

2. Data and method

2.1. Data and study region

To quantify the probabilistic divergences pertaining to the time-space scaling of water resources related variables, the study analyzed data collected at different locations across the State of Oklahoma (see Fig. 1). The assessed variables included streamflow, groundwater level, precipitation, soil moisture (5 cm depth), surface air temperature, solar radiation, relative humidity, and wind speed. Streamflow and groundwater data were retrieved from the USGS database, while the soil moisture and weather data were obtained from the Oklahoma Mesonet database (McPherson et al., 2007) which provides quality soil moisture and weather measurements over several decades (Scott et al., 2013). Over the decade 2001 to 2010, daily time-series were obtained from nine different stations including two groundwater stations, two streamflow stations, and five weather stations. Fig. 1 shows the locations of stations across Oklahoma and Table 1 provides additional details. The stations were selected based on data consistency and their spatial distribution. Five weather stations (i.e. W1, W2, W3, W4, W5) were considered within the Oklahoma Mesonet network. The streamflow stations are located on two different streams, namely the North Canadian River (station S1) and the North Fork Red River (station S2). Likewise, the groundwater wells belong to two different aquifer systems which are the Arbuckle-Simpson (station G1) and the High Plains (station G2).

The study period 2001–2010 was considered, based on time-space data consistency at all the selected stations. For this period, all the stations had continuous daily records with minor gaps, as the missing values were on average below 1%. The missing values were filled using either the closest daily value for variables displaying high temporal autocorrelations (i.e. streamflow, groundwater, soil moisture) or the inverse distance weighted averages (Di Luzio et al., 2008) for variables with high spatial correlations (i.e. precipitation, temperature, wind speed, relative humidity, and solar radiation). Note, the Oklahoma mesonet network counts 120 stations widely distributed across the State, and each of the selected weather station has several neighboring stations which data are correlated. The daily time-series were used as a basis for a scaling framework where the data were rescaled to generate weekly, biweekly, monthly, bimonthly, and trimonthly time-series. The rescaling technique averaged daily observations. As a result, daily time-series can be easily compared with weekly, biweekly, monthly, bimonthly, and trimonthly time-series. The rescaled data were analyzed separately and for each variable and location, the divergences due to the scaling were quantified probabilistically.

2.2. Methodology

This study used a non-parametric kernel density function to derive probability distributions and deviations caused by different scales were estimated using the Kullback–Leibler divergence.

2.2.1. Kernel density estimator

The Kernel density function (Silverman, 1986) was used to derive probability distributions of the eight different variables (Botev et al., 2010). The kernel density provides a smooth representation of the real data distribution (Qin et al., 2011). Applying a gaussian kernel to the variables (Jann, 2007; Seaman and Powell, 1996; Silverman, 1986), this study estimated the density functions. Given the time-series of a random variable X with n elements, the corresponding kernel density estimator $\hat{f}(.)$ was defined as:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{1}$$

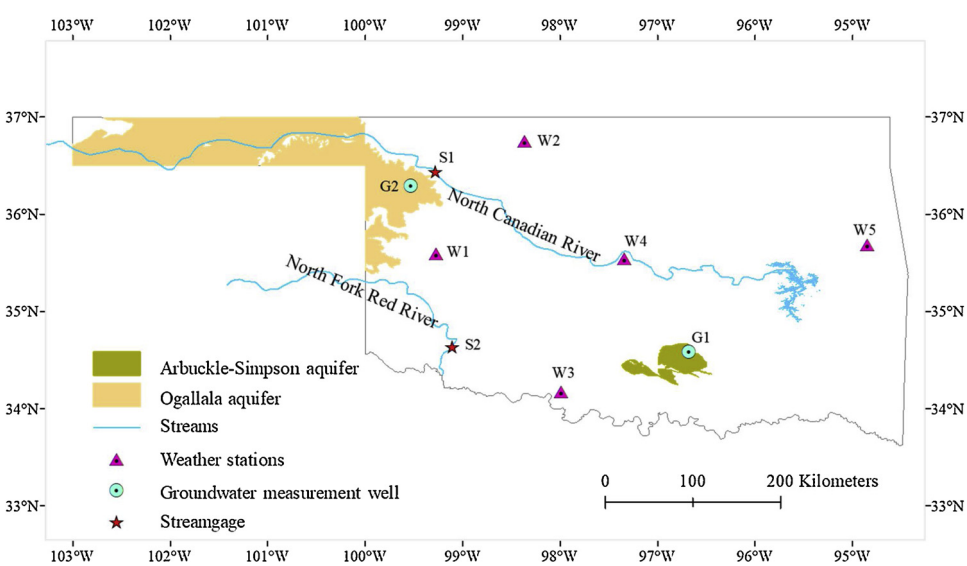


Fig. 1. Weather, streamflow, and groundwater measurement sites in Oklahoma.

Table 1
Streamflow, groundwater, and weather stations used in the study. The ID in the table is indicated in Fig. 1.

Station	Feature	Location	ID	Latitude (°N)	Longitude (°W)	Elevation (m above NGVD29)
Streamflow	HUC 11100301	Woodward	S1	36.44	99.28	558
	HUC 11120303	Tillman	S2	34.64	99.10	395
Groundwater	Arbuckle-Simpson aquifer	Pontotoc	G1	34.58	96.68	352
	High Plains aquifer	Woodward	G2	36.29	99.53	712
Weather	–	Butler	W1	35.59	99.27	520
	–	Cherokee	W2	36.75	98.36	362
	–	Waurika	W3	34.17	97.99	283
	–	Spencer	W4	35.54	97.34	373
	–	Cookson	W5	35.68	94.85	299

$$K(u) = \frac{1}{\sqrt{2\pi}} Exp\left(-\frac{u^2}{2}\right); \; with\ u = \frac{x - X_i}{h} \tag{2}$$

$$\int K(u)\, du = 1 \tag{3}$$

$$\int \hat{f}(x)\, dx = 1 \tag{4}$$

where h represents the band width or smoothing parameter, n is the size of the time-series; X_i is any value of the variable X within the range of the variation of X, x denotes the center of any bin with h width, and $K(.)$ is the gaussian kernel. In the kernel density approach, the choice of band width h is critical because it determines the degree of smoothness of the distribution. Silverman (1986) proposed a simple formula to identify the optimum band width h_{opt}. In the case of a gaussian kernel, the formula proposed by Silverman (1986) is a function of the series size n and its standard deviation σ as given by:

$$h_{opt} = 1.06\, \sigma\, n^{-1/5} \tag{5}$$

Using the estimated density function of each variable, the discrete probability values were derived by multiplying the kernel estimator value by the incremental bin value.

2.2.2. Kullback–Leibler divergence

The Kullback–Leibler divergence D_{KL}, also known as Kullback–Leibler distance, is a measure of dissimilarity between two probability distributions (Polani, 2013; Kullback and Leibler, 1951). D_{KL} does not require any parametric assumption on variables (Cover and Thomas, 2006). Two types of divergence were quantified in this study including the temporal divergence D_{KL} and the spatial divergence D'_{KL}. The values of D_{KL} were estimated by coupling daily probability distributions with distributions corresponding to rescaled data (i.e. weekly, biweekly, monthly, bimonthly, and trimonthly probability distributions). Hence, for each variable, the temporal divergences were evaluated in comparison with daily time scale data at station S1 for streamflow, station G1 for groundwater, station W1 for soil moisture, and weather variables. The spatial divergences D'_{KL} were evaluated between stations using monthly time scale. Thus, the D'_{KL} values were estimated for the couples (S1, S2), (G1, G2), (W1, W2), (W1, W3), (W1, W4), and (W1, W5).

The D_{KL} and D'_{KL} values were computed using the discrete probability values derived from the kernel density functions. Given P the daily discrete probability distribution of a target variable X, and Q an alternate probability distribution of the same variable (i.e. weekly, biweekly, monthly, bimonthly or trimonthly probability distributions), the temporal Kullback–Leibler divergence D_{KL} estimate was given by:

$$D_{KL}(P||Q) = -\sum_{i=1}^{m} P(i)\, \log \frac{P(i)}{Q(i)} \tag{6}$$

$$\sum_{i=1}^{m} P(i) \approx 1 \quad ; \quad \sum_{i=1}^{m} Q(i) \approx 1 \tag{7}$$

In Eqs. (6) and (7), m is the number of discrete intervals defined within the range of the variable X, $P(i)$ and $Q(i)$ are, respectively, the discrete probability values associated with an interval i within the range of X. For consistency, the same number of bins $m = 100$ was considered, but the ranges depend on the variable. The spatial Kullback–Leibler divergence D'_{KL} was defined as

$$D'_{KL}(P||P') = -\sum_{i=1}^{m} P(i)\, \log \frac{P(i)}{P'(i)} \tag{8}$$

$$\sum_{i=1}^{m} P(i) \approx 1 \quad ; \quad \sum_{i=1}^{m} P'(i) \approx 1 \tag{9}$$

The formulation of both D_{KL} and D'_{KL} is based on the same D'_{KL} is defined to measure the divergence between the probability distributions P and P' of two different stations given the same variable X at a monthly time scale. Thus, $P(i)$ and $P'(i)$ are, respectively, the discrete probability values of each of the two stations given a bin i within the range of X.

3. Results

The methodology was applied to the target variables. The effect of scaling on the probability distributions and the implication on inferences in water resources were analyzed.

3.1. Scaling and kernel density estimation

Using the kernel density method, the probability distributions of streamflow data from station S1, groundwater data from station G1, soil moisture and weather data from station W1 were assessed using different time-scales, including daily, weekly, monthly, bimonthly, and trimonthly scales. It is to note that the initial 10-year daily time-series size changed as the data was rescaled from daily (3652 daily observations) to weekly (522 weekly observations), biweekly (261 biweekly observations), monthly (120 monthly observations), bimonthly (60 bimonthly observations), trimonthly (40 trimonthly observations). For the kernel density method, Seaman et al. (1999) recommend a minimum sample size of 30, thus the data size used in this study allowed modelling up to a trimonthly time scale. Figs. 2 and 3 show the effect of scaling on the kernel density distribution of each variable. The optimal band widths, estimated based on Eq. (7), are reported in Table 2. Note the band widths have the same units as the variables. For each of the variables, the changes caused by time-scales were noticeable from the kernel density curves and disparities were notable, depending on the variable. For instance, the scaling effect on relative humidity (Fig. 3a) and wind speed (Fig. 3b) seemed comparable as the distribution's peak tended to increase with increasing time scale. Likewise, similarities can be seen between air temperature (Fig. 3c) and solar radiation (Fig. 3d) for which the kernel density distributions appeared bimodal. However, these visual resemblances can hardly be defined among the other four variables, including precipitation (Fig. 2a),

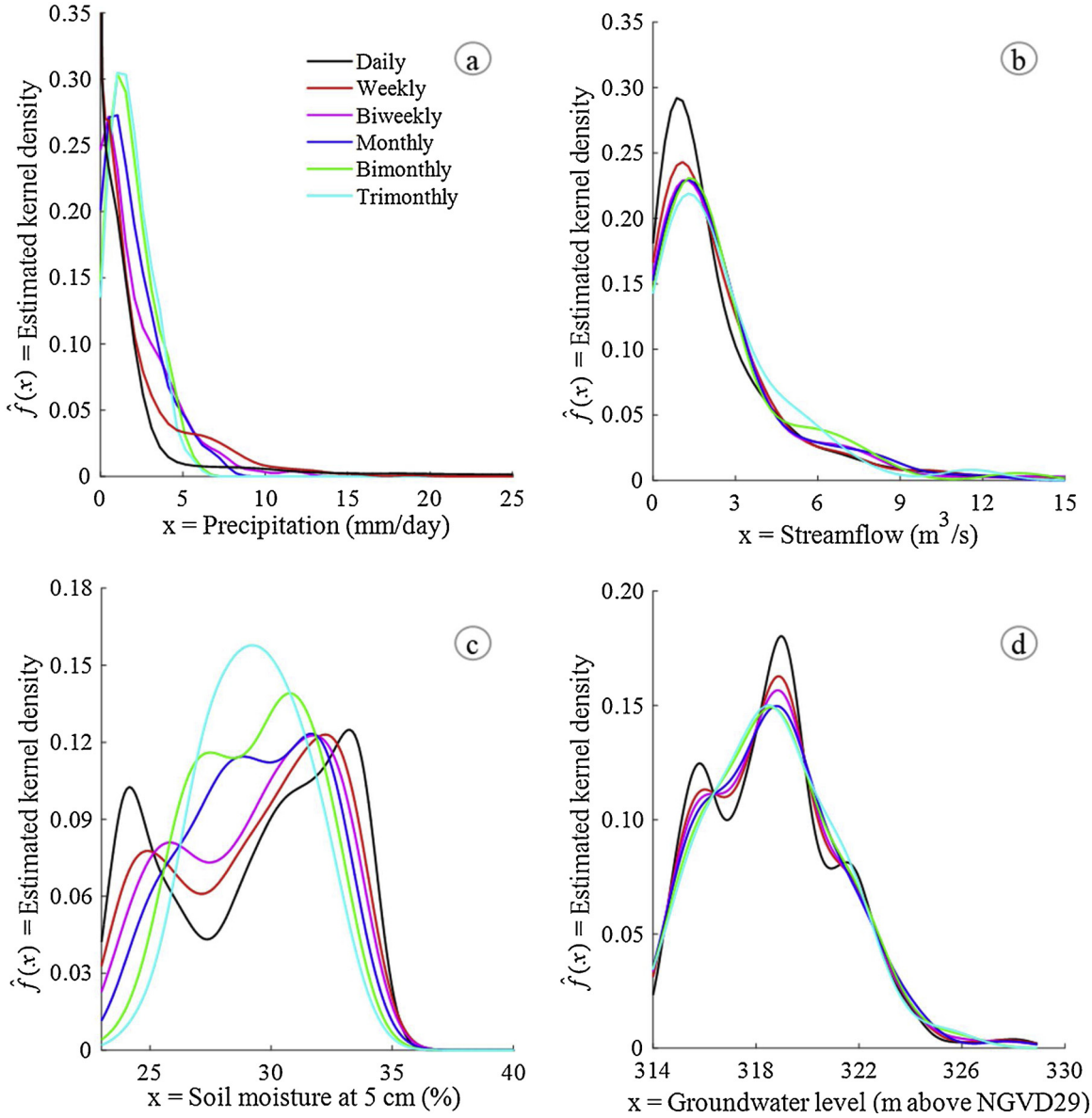


Fig. 2. Scaling effect on the shapes of the kernel density distribution of precipitation at station W1 (a), streamflow at station S1 (b), 5 cm depth soil moisture at station W1 (c), and groundwater level at station G1 (d).

streamflow (Fig. 2b), soil moisture (Fig. 2c), and groundwater (Fig. 2d).

In summary, the kernel density analysis brought out critical signals of scaling interference in the distribution shapes of the variables. Shifts were noticeable, as one moved from a fine to a larger time-scale. For some variables (e.g. relative humidity, wind speed), the change was apparently expressed as a shift of the kurtosis (peak), while for other variables the changes seemed more complex to define. The discrete probability values obtained from the kernel density were employed to determine the divergence pertaining to the temporal scaling framework.

3.2. Temporal divergence analysis

The actual probability distributions of variables were diversely affected by time-scales even though some resemblances were noted. The Kullback-Leibler divergence was applied to quantify the divergences pertaining to time scale changes. Table 3 shows the temporal divergence estimations. Since the Kullback-Leibler divergence is unitless, the values are comparable among variables. Hence, based on the D_{KL} values, the scaling appeared to have more effect on wind speed [$D_{KL(daily,\ trimonthly)}$ = -0.88], relative humidity [$D_{KL(daily,\ trimonthly)}$ = -0.88], precipitation [$D_{KL(daily,\ trimonthly)}$ = -0.24], and soil moisture [$D_{KL(daily,\ trimonthly)}$ = -0.20]. In contrast, the air temperature was less sensitive to temporal scaling [$D_{KL(daily,\ trimonthly)}$ = -0.01]. These observations are concordant with Figs. 2 and 3, as one compares daily kernel density curves with trimonthly curves. The large values of Kullback-Leibler divergence reported in Table 3 correspond to larger deviations (i.e. gap) between the kernel density curves displayed by Figs. 2 and 3. This finding supports the consistency of the methodology developed.

From an analytical point of view, it is normal to suspect that changes affecting the data probability distributions are likely to influence inferences based on these data. In water resources management, probabilistic tools are often involved in decision making. Therefore, it is important to understand how the time scale changes may affect statistical inferences. To illustrate this point, correlation analysis was done to evaluate the effect of time scale change on statistical relationships between variables. Fig. 4 shows examples of correlation changes due to

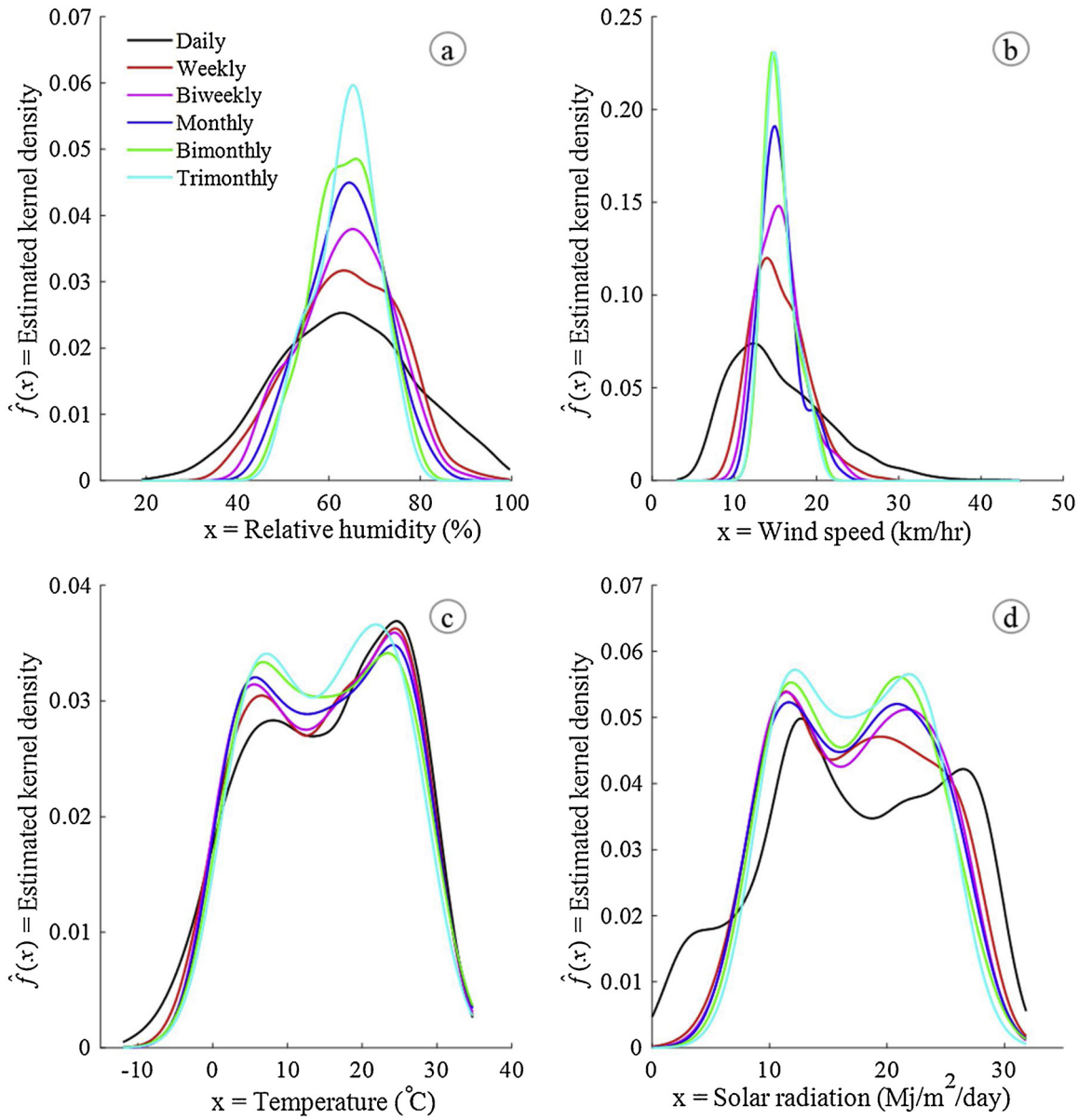


Fig. 3. Scaling effect on the shapes of the kernel density distribution of relative humidity (a), wind speed (b), air temperature (c), and solar radiation (d) at station W1.

the time scales variation. The patterns illustrated are relevant, as they pinpoint the scaling effect on statistical relationships. Subsequently, inferences made on correlation analysis may mislead interpretations unless the time scales used are emphasized.

3.3. Spatial divergence analysis

The spatial divergences D'_{KL} were evaluated using the same time scale (i.e. monthly) at individual stations, and the kernel density estimator corresponding to each variable was considered to obtain the

Table 2
Estimated optimal band widths for the studied variables at different time-scales.

Variable	Time Scales					
	Daily	Weekly	Biweekly	Monthly	Bimonthly	Trimonthly
Precipitation (mm/day)	1.42	0.89	0.75	0.67	0.59	0.60
Streamflow (m^3/s)	0.69	0.94	1.04	1.06	1.10	1.13
Groundwater level (m)	0.52	0.77	0.88	1.01	1.12	1.19
Soil moisture at 5 cm (%)	0.73	0.98	1.04	1.06	1.07	1.02
Relative humidity (%)	3.06	3.35	3.32	3.25	3.17	3.20
Wind speed (km/hr)	1.24	1.00	0.92	0.91	0.85	0.89
Air temperature (°C)	2.05	2.86	3.19	3.63	4.02	4.16
Solar radiation (Mj/m^2/day)	1.58	1.90	2.07	2.33	2.53	2.59

Table 3
Kullback-Leibler divergence estimates between daily probability distribution and alternate distributions (weekly, biweekly, monthly, bimonthly or trimonthly) for the target variables.

Time scale	Precipitation	Streamflow	Ground water level	Soil moisture at 5 cm	Relative humidity	Wind speed	Air temperature	Solar radiation
Daily	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Weekly	−0.023	−0.019	−0.005	−0.012	−0.052	−0.222	−0.004	−0.044
Biweekly	−0.125	−0.030	−0.008	−0.025	−0.120	−0.432	−0.008	−0.064
Monthly	−0.169	−0.028	−0.011	−0.061	−0.216	−0.600	−0.009	−0.069
Bimonthly	−0.227	−0.032	−0.015	−0.126	−0.355	−0.848	−0.010	−0.088
Trimonthly	−0.241	−0.037	−0.016	−0.202	−0.416	−0.876	−0.013	−0.115

Note: Kullback-Leibler divergence estimate is unitless.

discrete probabilities using m intervals defined within the range of the variable. Coupling the discrete probability distributions between stations, the spatial divergence D'_{KL} was quantified. Tables 4 and 5 summarize the spatial divergence for each variable. The pictorial changes are also reported. To ease comparison, the ranges of the variables were unit-normalized. Changes of probability distribution curves are noticeable among the compared locations, but the extent of these changes seems to depend on the type of variable. Amongst all the variables, groundwater showed the most important spatial divergence ($D'_{KL} = -0.18$). This is probably due to the nature of the groundwater wells G1 and G2 which belong to two different aquifer systems (see Table 1 and Fig. 1 for details). However, the most interesting aspect of the result is probably noticed with the curves presenting the pictorial changes of solar radiation and air temperature, as they tend to display similar shapes regardless of the locations. The D'_{KL} values corresponding to these two variables are very low, suggesting a minor spatial interference in the probabilistic behavior of the variables. Hence, for air temperature and solar radiation, the site-specific probability distributions could be likened using the same parent model. For soil moisture, precipitation, and wind speed, the spatial divergences are noticeable

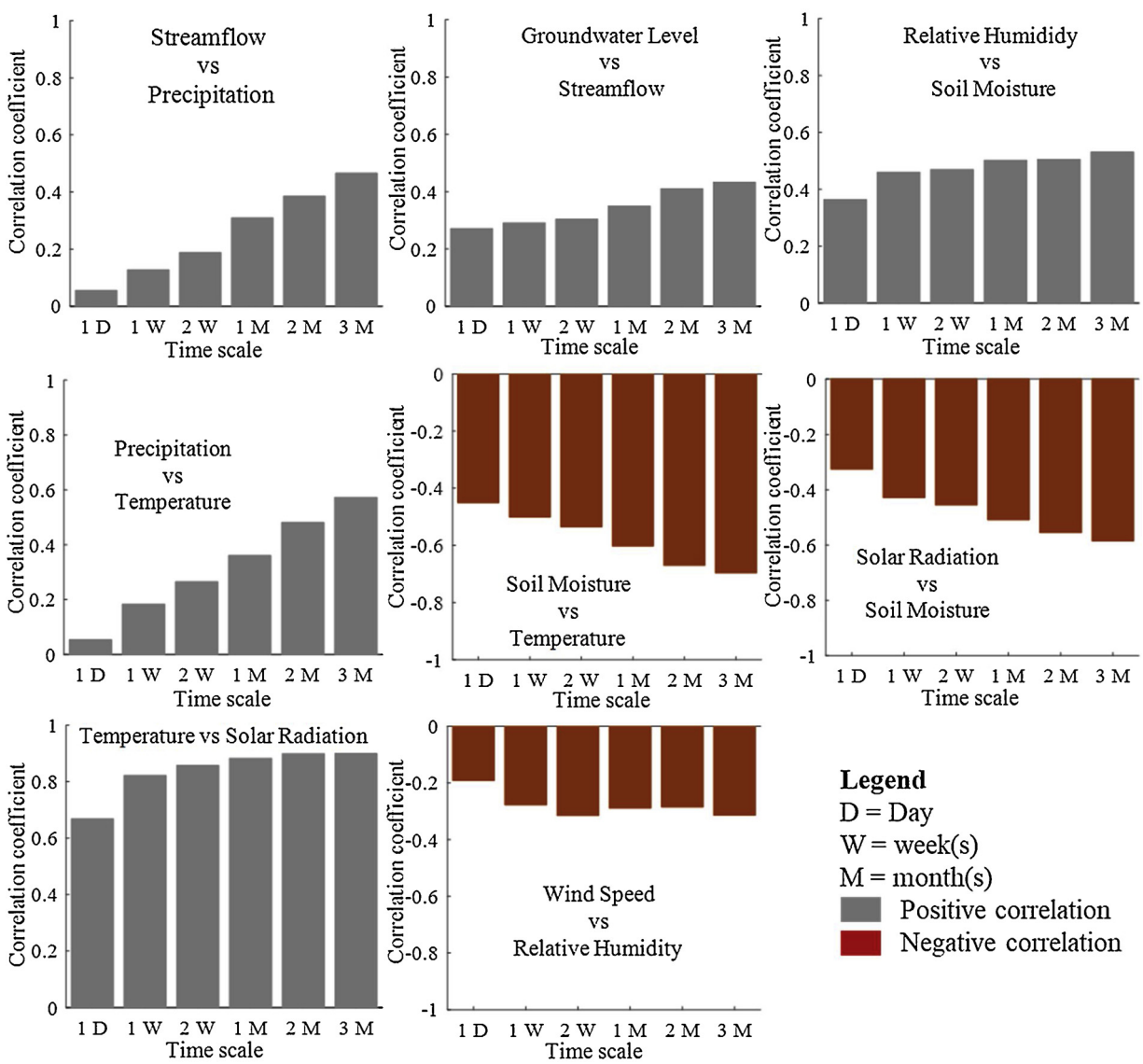
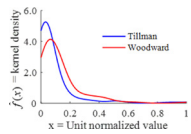
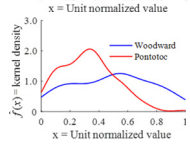
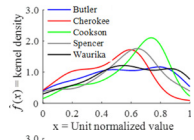
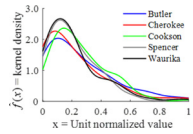


Fig. 4. Scaling effect on statistical inferences explained using a correlation analysis of pared variables.

Table 4
Spatial Kullback-Leibler divergence estimates for streamflow, groundwater, precipitation, and soil moisture using a monthly time-scale. D'_{KL} $(P||P')$ is estimated by coupling two stations associated with the discrete probability distribution P and P'.

Variable	Station (P)	Station (P')	D'_{KL} $(P\|\|P')$	Pictorial distributions
Streamflow	Woodward	Woodward	0.000	
	"	Tillman	0.005	
Groundwater	Woodward	Woodward	0.000	
	"	Pontotoc	−0.176	
Soil Moisture at 5 cm	Butler	Butler	0.000	
	"	Cherokee	−0.057	
	"	Cookson	−0.011	
	"	Spencer	−0.015	
	"	Waurika	−0.011	
Precipitation	Butler	Butler	0.000	
	"	Cherokee	−0.018	
	"	Cookson	0.006	
	"	Spencer	−0.020	
	"	Waurika	−0.012	

Note: Kullback-Leibler divergence estimate is unitless.

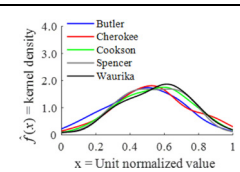
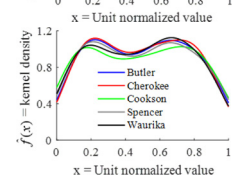
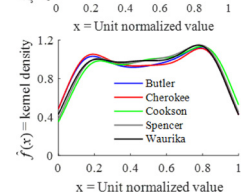
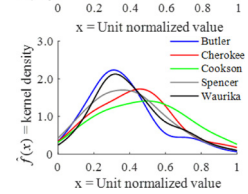
from the pictorial distribution curves. Based on the D'_{KL} values and the pictorial distributions, it clearly appears that the spatial variability interferes diversely in the probability distributions of the variables. Among all the variables, air temperature and solar radiation showed only minor sensitivity to space scaling.

3.4. Case study

The purpose of this case study was to evaluate the methodology on a different set of data outside the study region. Note the concept of time-space scaling is transversal for earth science disciplines as most biophysical phenomena are dynamic in time and space (White et al., 2018; Clark et al., 2004). To describe these phenomena, different categories of variables are often used. In a multiscale context, a variable, such as Normalized Difference Vegetation Index (NDVI), is considered practical for monitoring land-cover and water stress (Bauer et al., 2015; Sohoulande and Singh, 2015). Hence, the case study employed the methodology developed to address NDVI responses to precipitation and soil moisture.

The NDVI, soil moisture (5 cm depth), and precipitation data have been described by Sohoulande et al. (2015) who used the same data to address the large-scale vegetation response to precipitation across the southwestern United States. The NDVI data is a biweekly generated time-series corresponding to a Landsat's scene spanning approximately 170 km north-southward and 185 km east-westward, centered at the latitude 31.75 °N and longitude 100.58 °W. The target is located in Texas and corresponds to the Worldwide Reference System-2's path 29 and row 38. The data are processed and released for different vegetation covers at a biweekly time-scale by the United States Geological Survey (https://eros.usgs.gov). The case study only focused on the NDVI of two types of vegetation covers which are deciduous forest and grassland. Along with the vegetation index data, soil moisture and precipitation data were obtained from stations encompassed by the target Landsat's scene (Sohoulande and Singh, 2015). Twenty-three year (1989 to 2011) biweekly time-series were considered, then rescaled to 0.5, 1, 1.5, 2, 3, and 6 months' time scales. The probability distributions of the variables were approached using the kernel density modelling. The scaling effect on the data was analyzed using the Kullback–Leibler formulation. The implication of the scaling framework on statistical inferences was examined using a correlation analysis of NDVI response to precipitation and soil moisture given different time scales.

Table 5
Spatial Kullback-Leibler divergence estimates for relative humidity, solar radiation, temperature, and wind speed using a monthly time-scale. D'_{KL} $(P||P')$ is estimated by coupling two stations associated with the discrete probability distributions P and P'.

Variable	Station (P)	Station (P')	D'_{KL} $(P\|\|P')$	Pictorial distributions
Relative Humidity	Butler	Butler	0.000	
	"	Cherokee	−0.015	
	"	Cookson	−0.012	
	"	Spencer	−0.016	
	"	Waurika	−0.025	
Solar Radiation	Butler	Butler	0.000	
	"	Cherokee	0.003	
	"	Cookson	−0.014	
	"	Spencer	−0.005	
	"	Waurika	−0.003	
Air temperature	Butler	Butler	0.000	
	"	Cherokee	0.001	
	"	Cookson	−0.001	
	"	Spencer	0.002	
	"	Waurika	0.001	
Wind Speed	Butler	Butler	0.000	
	"	Cherokee	−0.042	
	"	Cookson	−0.076	
	"	Spencer	−0.021	
	"	Waurika	−0.009	

Note: Kullback-Leibler divergence estimate is unitless.

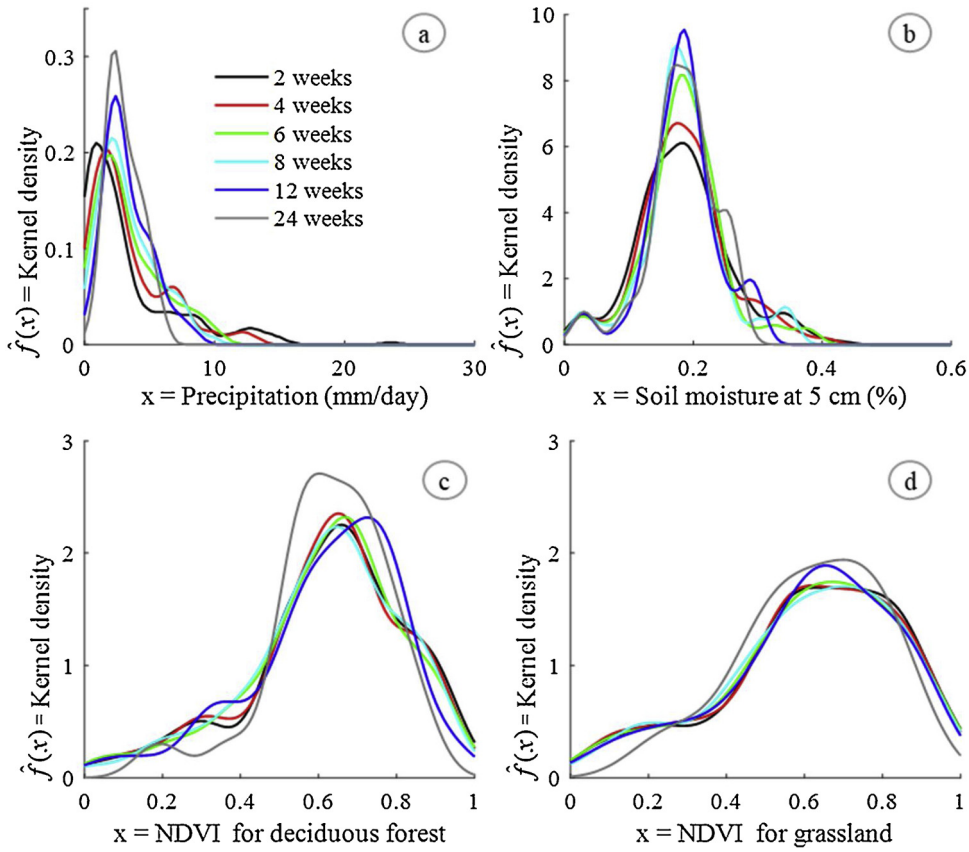


Fig. 5. Scaling effect on the kernel density distribution of precipitation (a), soil moisture at 5 cm (b), NDVI of deciduous forest (c), and NDVI of grassland (d).

Table 6
Kullback-Leibler divergence estimates between biweekly (0.5 month) probability distribution and the alternate distributions (1, 1.5, 2, 3, and 6 months) for NDVI, precipitation, and soil moisture data.

Time scale	Precipitation	Soil moister at 5 cm	NDVI for deciduous forest	NDVI for grassland
0.5 month	0	0	0	0
1 month	0.024	0.005	0.001	0.000
1.5 month	0.103	0.019	0.001	0.001
2 months	0.143	0.037	0.001	0.001
3 months	0.217	0.104	0.010	0.002
6 months	0.431	0.162	0.039	0.015

Fig. 5 presents the curves of kernel density estimates which also portray the changes of probability distributions, a consequence of scaling. The Kullback–Leibler divergence estimates are summarized in Table 6 which shows that precipitation and soil moisture are highly affected by scaling, but the Kullback–Leibler divergence is much higher for precipitation than soil moisture. In contrast, the NDVI of both vegetation types (deciduous forest, grassland) is less affected by the scale changes and the Kullback–Leibler divergence values are relatively negligible. Meanwhile, analysis of the scaling effect on NDVI response to precipitation and soil moisture seems to bring out a relevant paradigm. Indeed, Fig. 6 shows a gradual increase of the correlation between NDVI and precipitation, while the correlation between NDVI and soil moisture is likely stable as the time scale changes. This contrast is concordant with the result reported by Fig. 4 which also highlights different effects of time-space scaling on statistical inferences.

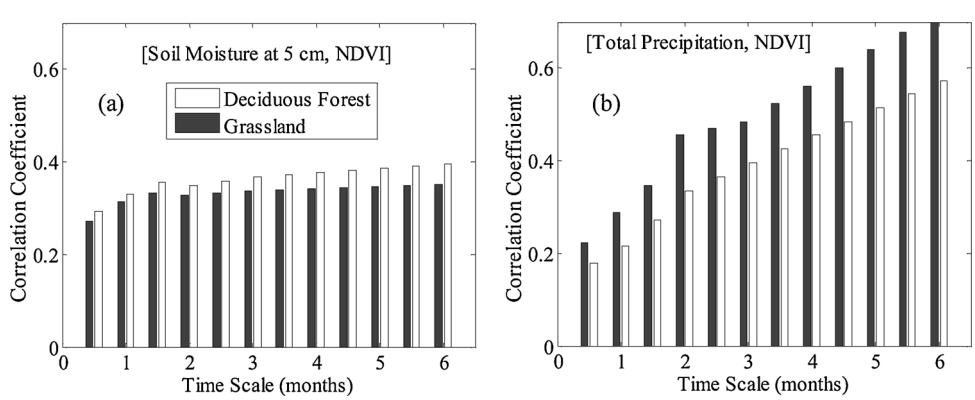


Fig. 6. Examples of marginal effect of time-scaling on the correlation analyses between soil moisture and NDVI (a); precipitation and NDVI (b). Note: all the correlations are significant at p-value = 0.05.

4. Synthesis and discussion

The results revealed critical discrepancies among the probability distributions of rescaled data. However, the magnitudes of divergences differed, depending on the variable. For instance, the distributions of wind speed, relative humidity, precipitation, and temperature changed remarkably, as the time scale changed. In contrast, the changes observed with air temperature seemed negligible as the scale changed. Interestingly, the spatial divergence analysis also indicated a negligible scaling effect on the probability distribution of air temperature which appeared less sensitive to time-space scaling. This observation is likely to have critical implications in water resources in view of the important role of air temperature in evapotranspiration (Olivera-Guerra et al., 2017; Trajkovic, 2005). For instance, the fact that the time-space scaling has a minor effect on the probability distribution of air temperature in a given region (e.g. Oklahoma) could leverage the accuracy of atmospheric water demand estimate. In opposite to air temperature, wind speed, relative humidity, precipitation, and soil moisture were found highly sensitive to time-space scales. Therefore, the inferences made using these variables would require more perspicacity to achieve wise water resources management.

Regardless of the regional considerations involved in the study, the outcomes highlighted seem consistent with conclusions drawn from previous studies related to water resources. For instance, Schoof and Pryor (2001) downscaled temperature and precipitation and concluded a high predictability of temperature in comparison to precipitation. Indeed, this conclusion may be sustained by the low impact of time-space scaling on the probability distribution of temperature. Thus, there is a possibility to anticipate the high downscaling capability of temperature by just referring to the probabilistic divergences pertaining to time-space scaling. From this standpoint, the methodology reported in this paper may be considered as a supporting tool to evaluate the probabilistic divergences pertaining to time-space scaling and indirectly help choose appropriate scales. However, the method can also be used to improve input quality in water resources modelling. For example, Olivera-Guerra et al. (2017) reported a method of quantifying actual evapotranspiration using surface air temperature. Likewise, Sohoulande and Singh (2015) reported a method to quantify plant growth using precipitation, soil moisture, and temperature. For each of these methods, a prior understanding of the scaling effect on the input variables is desirable, as it may improve the interpretation of outcomes.

From correlation analyses in Fig. 4, it appears that scaling has a noticeable effect on statistical inferences. Indeed, large scales tend to provide stronger statistical relationships (e.g. higher correlations). However, the statistics seem either moderate or substantial, depending on the variables involved. This tendency is confirmed by the case study which evaluated the NDVI response to precipitation and soil moisture at different time scales (i.e. 0.5, 1, 1.5, 2, 3, 6 months). The analyses were summarized by Fig. 6 which showed a gradual increase of correlation between NDVI and precipitation as the time scale changed. At the same time, the correlation between NDVI and soil moisture seemed moderate and visually stable as the time scale changed. From a statistical perspective, rescaling a variable is often associated with a loss of variance (Peleg et al., 2018; Sohoulande and Singh, 2016). This loss of variance is likely to affect the statistics involving the variables. This is probably the explanation of the paradigm of correlation changes observed with the analyses in both Figs. 4 and 6. Nevertheless, the magnitudes of time-space scaling on the probabilistic divergences of the variables are signals of the potential biases in statistical inferences.

5. Conclusion

The key findings of the study are as follows:

- Water resource related variables are diversely affected by time-space scaling.
- Air temperature is less sensitive to time-space scaling.
- Wind speed, relative humidity, soil moisture, and precipitation are much sensitive to time-space scaling.
- Statistical inferences are moderately or highly affected by time-space scales, depending on the variables involved. However, the correlation values tend to increase with larger scales.
- Inferences made in water resources must emphasize the time-space scales, since for certain variables statistical correlations gradually change with scaling.

The methodology employed helped to understand potential divergences caused by time space scales on water resources variables. Upon data availability, the method can be conveniently considered as a data appraisal tool in a situation where the time-space scaling effect on statistical inferences need to be understood.

Disclaimer

Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

Conflict of interest

None.

References

Bauer, P.J., Stone, K.C., Andrae, J.G., Busscher, W.J., Millen, J.A., Strickland, E.E., Evans, D.E., 2015. Water deficit and nitrogen fertilizer effects on NDVI of 'Tifton 85' Bermudagrassduring regrowth. Appl. Eng. Agric. 31 (1), 97–102.

Botev, Z.I., Grotowski, J.F., Kroese, D.P., 2010. Kernel density estimation via diffusion. Ann. Stat. 38 (5), 2916–2957.

Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., Wilby, R., 2004. The Schaake shuffle: a method for reconstructing space–time variability in forecasted precipitation and temperature fields. J. Hydrometeorol. 5 (1), 243–262.

Cover, T.M., Thomas, J.A., 2006. Elements of Information Theory. John Wiley & Sons.

Di Luzio, M., Johnson, G.L., Daly, C., Eischeid, J.K., Arnold, J.G., 2008. Constructing retrospective gridded daily precipitation and temperature datasets for the conterminous United States. J. Appl. Meteorol. Climatol. 47 (2), 475–497.

Jann, B., 2007. Univariate Kernel Density Estimation. Statistical Software Component. , S456410.

Koutsouris, A., Seibert, J., Lyon, S.W., 2017. Statistical Downscaling of Global Precipitation Datasets in Data Limited Regions: a Case Study of Kilombero Valley. Tanzania. .

Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Ann. Math. Stat. 22 (1), 79–86.

Malone, B.P., Styc, Q., Minasny, B., McBratney, A.B., 2017. Digital soil mapping of soil carbon at the farm scale: a spatial downscaling approach in consideration of measured and uncertain data. Geoderma 290, 91–99.

Maurer, E.P., Brekke, L., Pruitt, T., Duffy, P.B., 2007. Fine-resolution climate projections enhance regional climate change impact studies. Eos Trans. Am. Geophys. Union 88 (47), 504.

McPherson, R.A., Fiebrich, C., Crawford, K.C., Elliott, R.L., Kilby, J.R., Grimsley, D.L., Martinez, J.E., Basara, J.B., Illston, B.G., Morris, D.A., Kloesel, K.A., Stadler, S.J., Melvin, A.D., Sutherland, A.J., Shrivastava, H., 2007. Statewide monitoring of the mesoscale environment: a technical update on the Oklahoma Mesonet. J. Atmos. Oceanic Technol. 24, 301–321 (link to article).

Medvigy, D., Wofsy, S.C., Munger, J.W., Hollinger, D.Y., Moorcroft, P.R., 2009. Mechanistic scaling of ecosystem function and dynamics in space and time: ecosystem demography model version 2. J. Geophys. Res. Biogeosci. (G1), 114.

Oki, T., Kanae, S., 2006. Global hydrological cycles and world water resources. Science 313 (5790), 1068–1072.

Olivera-Guerra, L., Mattar, C., Merlin, O., Durán-Alarcón, C., Santamaría-Artigas, A., Fuster, R., 2017. An operational method for the disaggregation of land surface temperature to estimate actual evapotranspiration in the arid region of Chile. Isprs J. Photogramm. Remote. Sens. 128, 170–181.

Peleg, N., Marra, F., Fatichi, S., Paschalis, A., Molnar, P., Burlando, P., 2018. Spatial variability of extreme rainfall at radar subpixel scale. J. Hydrol. 556, 922–933.

Perrette, M., Landerer, F., Riva, R., Frieler, K., Meinshausen, M., 2013. A scaling approach to project regional sea level rise and its uncertainties. Earth Syst. Dyn. 4 (1), 11–29.

Pires, A., Morato, J., Peixoto, H., Botero, V., Zuluaga, L., Figueroa, A., 2017. Sustainability assessment of indicators for integrated water resources management. Sci. Total Environ. 578, 139–147.

Polani, D., 2013. Kullback-leibler divergence. Encyclopedia of Systems Biology. Springer, New York, NY, pp. 1087–1088.

Prada, A.F., Chu, M.L., Guzman, J.A., Moriasi, D.N., 2017. Evaluating the impacts of agricultural land management practices on water resources: a probabilistic hydrologic modeling approach. J. Environ. Manage. 193, 512–523.
Qin, Z., Li, W., Xiong, X., 2011. Estimating wind speed probability distribution using kernel density method. Electr. Power Syst. Res. 81 (12), 2139–2146.
Schoof, J.T., Pryor, S.C., 2001. Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks. Int. J. Climatol. 21 (7), 773–790.
Scott, B.L., Ochsner, T.E., Illston, B.G., Fiebrich, C.A., Basara, J.B., Sutherland, A.J., 2013. New soil property database improves Oklahoma Mesonet soil moisture estimates. J. Atmos. Oceanic Technol. 30 (11), 2585–2595.
Seaman, D.E., Powell, R.A., 1996. An evaluation of the accuracy of kernel density estimators for home range analysis. Ecology 77 (7), 2075–2085.
Seaman, D.E., Millspaugh, J.J., Kernohan, B.J., Brundige, G.C., Raedeke, K.J., Gitzen, R.A., 1999. Effects of sample size on kernel home range estimates. J. Wildl. Manage. 739–747.
Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis 26 CRC press.
Simonovic, S.P., 2017. Bringing future climatic change into water resources management practice today. Water Resour. Manag. 31 (10), 2933–2950.
Sohoulande, D.D.C., Singh, V.P., 2015. Retrieving vegetation growth patterns from soil moisture, precipitation and temperature using maximum entropy. Ecol. Modell. 309, 10–21.
Sohoulande, D.D.C., Singh, V.P., 2016. Entropy-based index for spatiotemporal analysis of streamflow, precipitation, and land-cover. J. Hydrol. Eng. 21 (11), 05016024.
Sohoulande, D.D.C., Singh, V.P., Frauenfeld, O.W., 2015. Vegetation response to precipitation across the aridity gradient of the southwestern United states. J. Arid Environ. 115, 35–43.
Trajkovic, S., 2005. Temperature-based approaches for estimating reference evapotranspiration. J. Irrig. Drain. Eng. 131 (4), 316–323.
Vörösmarty, C.J., Green, P., Salisbury, J., Lammers, R.B., 2000. Global water resources: vulnerability from climate change and population growth. science 289 (5477), 284–288.
White, H.J., Montgomery, W.I., Pakeman, R.J., Lennon, J.J., 2018. Spatiotemporal scaling of plant species richness and functional diversity in a temperate semi-natural grassland. Ecography 41 (5), 845–856.
Wiens, J.A., 1989. Spatial scaling in ecology. Funct. Ecol. 3 (4), 385–397.