

# A generic composite measure of similarity between geospatial variables



Yadong Liu <sup>a,\*</sup>, Kwang Soo Kim <sup>a,b</sup>, Robert M. Beresford <sup>c</sup>, David H. Fleisher <sup>d</sup>

<sup>a</sup> Department of Plant Science, Seoul National University, Seoul 08826, Republic of Korea

<sup>b</sup> Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Republic of Korea

<sup>c</sup> The New Zealand Institute for Plant and Food Research Limited, Private Bag 92169, Auckland 1142, New Zealand

<sup>d</sup> USDA-ARS Beltsville Agricultural Research Center, 10300 Baltimore Avenue, Beltsville, MD 20705, USA

## ARTICLE INFO

### Keywords:

Accuracy assessment  
Analogy analysis  
Antianalog identification  
Scale difference  
Similarity metric  
Structural information

## ABSTRACT

The comparison between spatial or temporal patterns is often needed for model evaluation and change detection in ecological studies. The statistics developed for image quality assessment, such as the structural similarity index (SSIM) and the composite similarity measure based on means, standard deviations, and correlation coefficient (CMSC), have been introduced for comparing ecological patterns. However, these measures can be applied only when a positive relationship is expected between patterns having the same scale. We propose a new index, generic composite similarity measure (GCSM), to meet a wide range of potential applications. A set of numerical experiments was performed to illustrate the properties of GCSM in comparison with SSIM and CMSC. Two case studies were conducted examining the (dis)agreement between two products of gross primary production (GPP), and the relative (dis)similarity between GPP and precipitation, respectively. GCSM has advantages over both SSIM and CMSC, including higher sensitivity and the ability to quantify the dissimilarity, which cannot be properly revealed with the latter two indices. The normalization preprocessing constructs universal criteria for assessing the relative (dis)similarity between patterns having unequal scales. The GCSM, overcoming the limitations of preexisting composite measures in quantifying the similarity or dissimilarity between patterns, would aid assessment of heterogeneous relationship between ecological factors over space or time.

## 1. Introduction

Similarity or difference measures are often used to compare spatial and temporal patterns in ecological studies (Legendre and Legendre, 2012; Long and Robertson, 2017). These measures have been employed to analyze geospatial data obtained from remote sensing (Fensholt and Proud, 2012), hydrological models (Wealands et al., 2005), and climate models (Roux et al., 2017). For example, the estimates of gross primary production (GPP) have been compared to identify the agreement and difference among them in terms of the spatial distribution and the interannual variability (Anav et al., 2015; Jiang and Ryu, 2016). The statistics used for this purpose commonly include correlation coefficient (CC), mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE) (Legates and McCabe, 1999; Moriasi et al., 2007).

Mo et al. (2014) and Jackson et al. (2019) suggested that a single measure may have limited capability to appraise the similarity between spatial or temporal patterns. The CC quantifies the strength of joint variability between patterns but is insensitive to the difference between

their magnitudes or variabilities (Legates and McCabe, 1999; McCuen and Snyder, 1975). For example, a CC = 1 does not imply a complete agreement between two patterns if they do not conform to the identical line (Kuhn and Johnson, 2019). In contrast, MAE, MSE, and RMSE are dominated by the amplitude of difference but may not capture the correspondence over space or time (Kufareva and Abagyan, 2012; Wang and Bovik, 2009). In general, different measures could complement in assessing the similarity between patterns.

Composite measures, which considering several aspects regarding the similarity, have been proposed and utilized (Hagen-Zanker, 2006; Long and Robertson, 2017). The structural similarity index (SSIM; Wang and Bovik, 2002; Wang et al., 2004), initially designed for image quality assessment, has been applied in ecological studies. For example, Jones et al. (2016) used the SSIM to compare the predicted occurrence probabilities of different sperm whales in the Mediterranean Sea. The composite similarity measure based on means, standard deviations, and correlation coefficient (CMSC; Palubinskas, 2014, 2017), which is a variant of SSIM, has also been used for assessing the qualities of pan-sharpened satellite images (Palubinskas, 2015).

\* Corresponding author.

E-mail address: liuyadong@snu.ac.kr (Y. Liu).

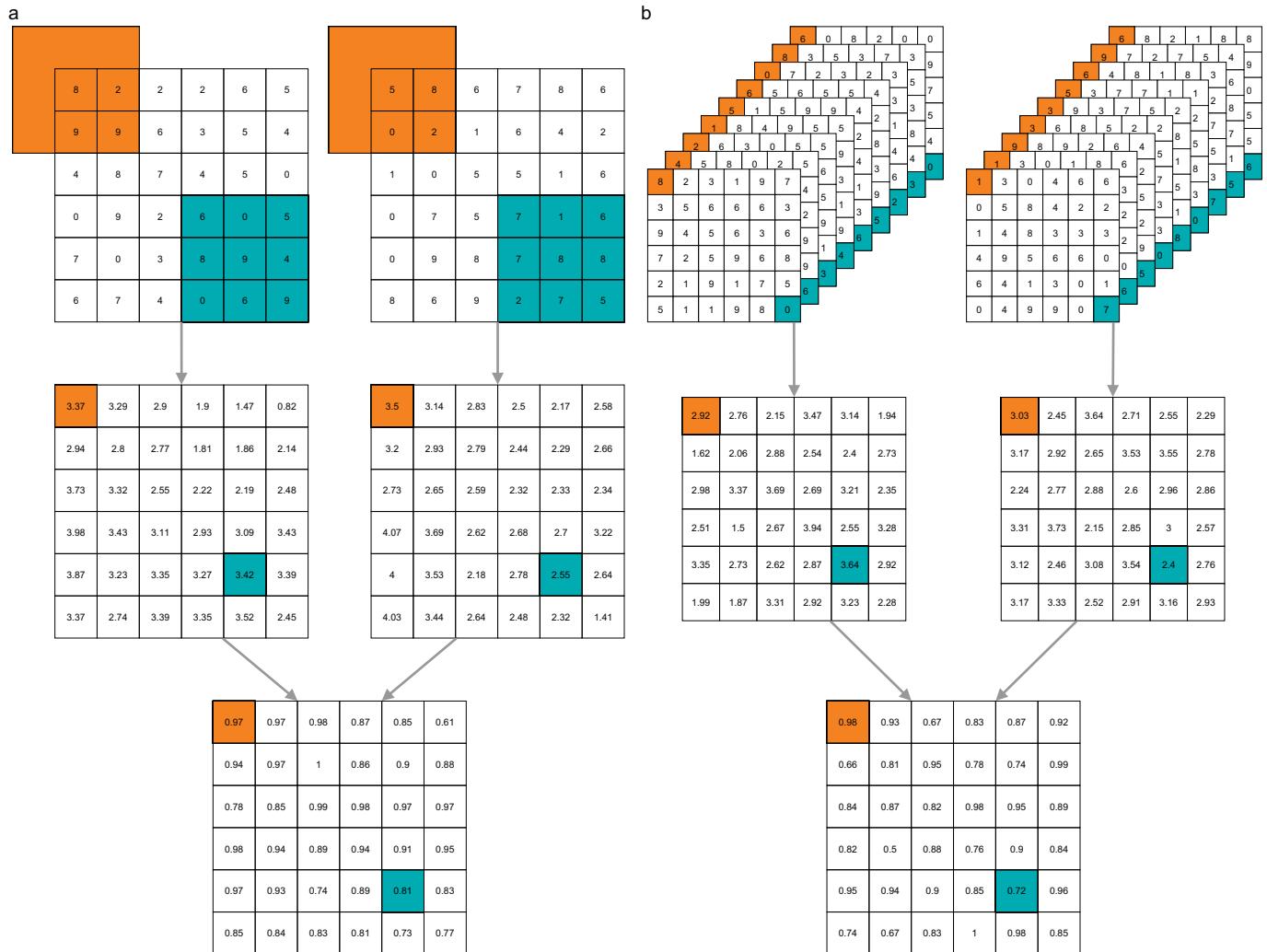
Composite measures such as SSIM and CMSC depend on local statistics calculated within a specific spatial window instead of the entire region, such as local mean and standard deviation (Hagen-Zanker, 2006). Sliding the window cell by cell allows achieving the pattern of similarity over space, representing the local similarity in the global context (Fotheringham et al., 2002; Lobo et al., 2008; Long and Robertson, 2017). Robertson et al. (2014) suggested such indices provide a diagnostic tool for spatial modeling and analysis and would help detect the spatial heterogeneity in ecology as they could map the structure when assessing spatially explicit models.

Ecological studies often examine the relationship between patterns in various contexts. For example, an ecological model could be evaluated by checking the agreement between observations and predictions over a time period (Müller et al., 2017). The relative correspondence between patterns having unequal scales, e.g., GPP and precipitation, may be analyzed (Beer et al., 2010). Quantification of the negative relationship between patterns at the same scale, e.g., climate variability (Scheff et al., 2017), or at different scales, e.g., elevation and temperature (Li et al., 2015), could also be of interest in spatial analysis. However, the composite similarity measures like SSIM and CMSC have limitations dealing with these scenarios. For example, a negative SSIM may not indicate a negative relationship between patterns because it

turns negative if two local means have different signs even when the patterns are positively correlated. Brunet, 2012 and Palubinskas (2017) reported these indices are undesirable to quantify the negative relationship which is of no concern to most image processing applications. Moreover, comparing patterns having unequal scales would lead to invalid outcomes because these measures are scale-dependent, originally proposed for image quality assessment where the scales are fixed, e.g., 0–255. Thus, enhancements to these composite measures are desirable for wide applications in ecological studies.

Here we propose the generic composite similarity measure (GCSM), which allows comparing patterns in a wide range of contexts. The index was designed suitable for evaluating both positive and negative relationships between spatial or temporal patterns. Meanwhile, it was enabled to assess the relative similarity or dissimilarity between patterns having unequal scales. In particular, we focused on the (dis)agreement between temporal patterns and the relative (dis)similarity between patterns over space. The GCSM would help gain quantitative insight into the heterogeneous relationship between ecological factors over space or time.

In this paper, we first describe composite measures SSIM and CMSC for comparing patterns and the possible limitations in ecological applications, then detail the enhancements leading to the proposal of GCSM.



**Fig. 1.** A schematic view of computing local statistics and  $s_2$  of GCSM within (a) spatial or (b) temporal windows. The spatial window is set as a  $3 \times 3$  area, e.g., the orange or cyan one (a). By sliding over space, the maps of the standard deviation of two patterns are obtained, respectively, then together produce the map of  $s_2$ . The temporal window is set along the time axis of a spatiotemporal cube; by sliding over space, the map of  $s_2$  is finally obtained (b). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

We conduct numerical experiments and two case studies to assess the properties of the proposed measure and demonstrate the suitability for ecological applications. The strength and implications are then discussed.

## 2. Methods

The CMSC was recently developed based on the widely used SSIM. We briefly review them before presenting the potential limitations of CMSC in ecological applications and the GCSM.

### 2.1. SSIM and CMSC

The SSIM was proposed as an alternative to MSE, aiming at assessing image qualities (Wang and Bovik, 2002). It was constructed as a multiplication of three components to quantify the agreement between patterns:

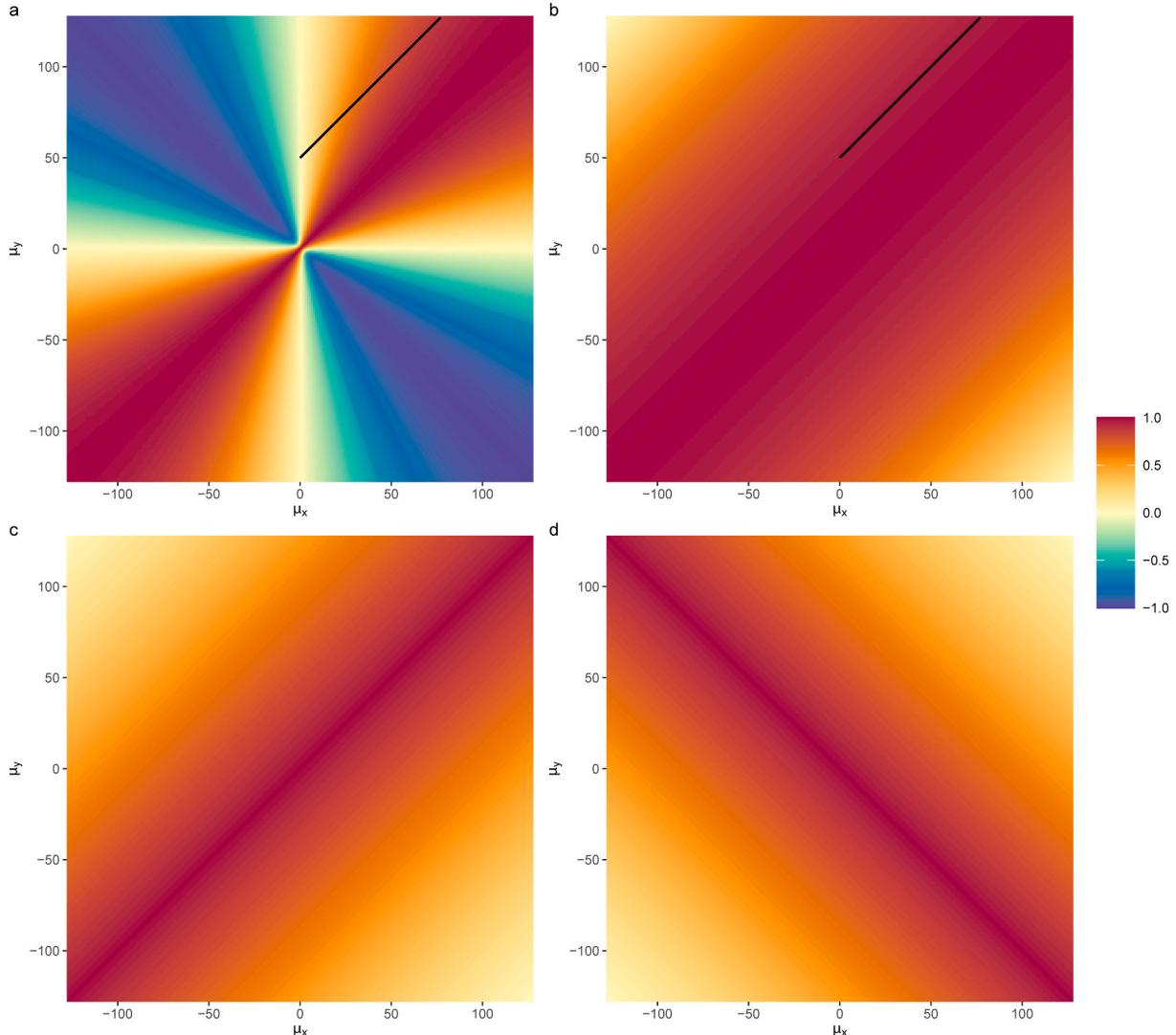
$$S(x, y) = s_1(x, y) \cdot s_2(x, y) \cdot s_3(x, y), \quad (1)$$

where  $x$  and  $y$  are two patterns within a spatial window (Fig. 1a). Both  $s_1$  and  $s_2$  are based on the Dice measure which has a form of  $2ab/(a^2 + b^2)$  (Palubinskas, 2017). They are used to measure the agreement between

local statistics of two patterns, i.e., local means and standard deviations, respectively. The  $s_3$  is close to the Pearson correlation coefficient between the two patterns. The equations for computing these terms are listed in Table 1. SSIM provides an overall evaluation of the agreement between two patterns within a spatial window.

Although showing empirical and analytical properties desirable for comparing patterns (Brunet et al., 2012; Dosselmann and Yang, 2009), the SSIM was suggested applicable only to limited situations where the data meet a set of prerequisites (Pambrun and Noumeir, 2015). The data are restricted to have the same sign. Otherwise, it may produce a negative  $s_1$ , which is beyond the permissible range. Notably, the  $s_1$  would turn negative if two local means have different signs (Fig. 2a). This makes the SSIM challenging to represent the agreement between local means. Besides, Palubinskas (2017) reported that the Dice measure could be altered by the magnitude of local statistics, making  $s_1$  and  $s_2$  less desirable similarity indices. For example, different  $s_1$  could be achieved even when the differences between local means are identical (Fig. 2a).

The CMSC was then developed to overcome the limitations of SSIM with modified equations computing the components (Palubinskas, 2014). In particular,  $s_1$  and  $s_2$  adopted the squared Euclidean distance between local statistics (Table 1). As a consequence, the  $s_1$  would never turn negative even if two local means have different signs. Besides,  $s_1$



**Fig. 2.** The distribution of  $s_1$  in (a) SSIM, (b) CMSC, (c) GCSM if  $s_3 \geq 0$ , and (d) GCSM if  $s_3 < 0$ .  $\mu_x$  and  $\mu_y$  are local means of  $x$  and  $y$  within a window, respectively. The two lines in the first quadrant represent  $\mu_x$  and  $\mu_y$  having a difference of 50.

**Table 1**

Equations for computing components of the structural similarity (SSIM; Wang et al., 2004), the composite similarity measure based on means, standard deviations, and correlation coefficient (CMSC; Palubinskas, 2017), CMSC with the first enhancement (CMSC\_E1), CMSC with the second enhancement (CMSC\_E2), and the generic composite similarity measure (GCSM).

Component	SSIM	CMSC	CMSC_E1	CMSC_E2	GCSM
$s_1$	$1 - \frac{(\mu_x - \mu_y)^2}{\mu_x^2 + \mu_y^2 + c_1}$	$1 - \left( \frac{\mu_x - \mu_y}{\max - \min} \right)^2$	$1 - \frac{ \mu_x - \mu_y }{\max - \min}$	$\begin{cases} 1 - \left( \frac{\mu_x - \mu_y}{\max - \min} \right)^2 & s_3 \geq 0 \\ 1 - \left( \frac{\min + \max - \mu_x - \mu_y}{\max - \min} \right)^2 & s_3 < 0 \end{cases}$	$\begin{cases} 1 - \frac{ \mu_x - \mu_y }{\max - \min} & s_3 \geq 0 \\ 1 - \frac{ \min + \max - \mu_x - \mu_y }{\max - \min} & s_3 < 0 \end{cases}$
$s_2$	$1 - \frac{(\sigma_x - \sigma_y)^2}{\sigma_x^2 + \sigma_y^2 + c_2}$	$1 - \left[ \frac{\sigma_x - \sigma_y}{[(\max - \min)/2]} \right]^2$	$1 - \frac{ \sigma_x - \sigma_y }{(\max - \min)/2}$	$1 - \left[ \frac{\sigma_x - \sigma_y}{(\max - \min)/2} \right]^2$	$1 - \frac{ \sigma_x - \sigma_y }{(\max - \min)/2}$
$s_3$	$\frac{\sigma_{xy} + c_3}{\sigma_x \sigma_y + c_3}$	$\frac{\sigma_{xy}}{\sigma_x \sigma_y}$	$\frac{\sigma_{xy}}{\sigma_x \sigma_y}$	$\frac{\sigma_{xy}}{\sigma_x \sigma_y}$	$\begin{cases} 1 & \sigma_x = \sigma_y = 0 \\ 0 & \sigma_x = 0 \text{ or } \sigma_y = 0 \\ \frac{\sigma_{xy}}{\sigma_x \sigma_y} & \text{otherwise} \end{cases}$

$\mu_x$  and  $\mu_y$  are the local means of patterns  $x$  and  $y$  within a window, respectively.  $\sigma_x$  and  $\sigma_y$  are the corresponding standard deviations;  $\sigma_{xy}$  is the covariation between two patterns. For SSIM, these statistics are computed as weighted averages, while are determined with unweighted approaches in CMSC and GCSM (Table A.1). Both  $c_1$  and  $c_2$  are small positive constants for computation stability. The max and min are the global maximum and minimum of given data, respectively.

and  $s_2$  would not be altered by the magnitude of local statistics (Fig. 2b). In general, the CMSC avoided the problems associated with the Dice measure in SSIM.

Still, the CMSC shows some shortcomings for wide applications. The triangle inequality has been illustrated as an essential property of difference measures (Brunet et al., 2012; Paul et al., 2000). However, the CMSC depends on the squared Euclidean distance, which does not satisfy the triangle inequality. The squared distance gives unequal emphasis across the scope of differences (Willmott and Matsuura, 2005). Particularly, it deemphasizes the small differences between local statistics of two patterns. As a result,  $s_1$  and  $s_2$  would not proportionately increase when the differences are narrowing down (Fig. 3). Such a property makes the CMSC less comparable across multiple windows and less efficient to identify the similarity between patterns.

The CMSC likely has a limited utility in representing the dissimilarity between patterns. If two patterns are positively correlated, the CMSC exhibits a unique maximum if and only if the data conform to the identity line  $y = x$  (Fig. 4a). This line represents the perfect agreement between two patterns, instinctively passing through the points (min,

min) and (max, max) where min and max are the global minimum and maximum of given data. Symmetrically, the negative diagonal line  $y = -x + \min + \max$  would indicate the perfect antianalog where two patterns tend to be opposite and potentially span the whole range of given data (Fig. 4b; Mo et al., 2014). Therefore, the minimum, i.e.,  $-1$ , is expected to be uniquely achieved from the negative diagonal line. However, the CMSC varies even when the data precisely fit this line because  $s_1$  is not always one (Fig. 2b). Moreover, the meaning of negative CMSC is ambiguous. It would be achieved where the  $s_3$  is negative, which indicates the data are distributed around a regression line with a negative slope (Rodgers and Nicewander, 1988). The  $s_1$ , on the contrary, measures the agreement between the local means of two patterns. Therefore, the multiplication of  $s_1$  and negative  $s_3$  would become less indicative.

## 2.2. Generic composite similarity measure

The generic composite similarity measure (GCSM) is designed to overcome the drawbacks of the CMSC and SSIM. The GCSM still consists of three components following the general structure of these composite measures. The squared Euclidean distances in CMSC are displaced to eliminate the less desirable properties associated with them. Instead, the GCSM utilizes the normalized, non-squared Euclidean distance to specify the difference between local statistics (Fig. 3). For example, the  $s_2$  could be computed as:

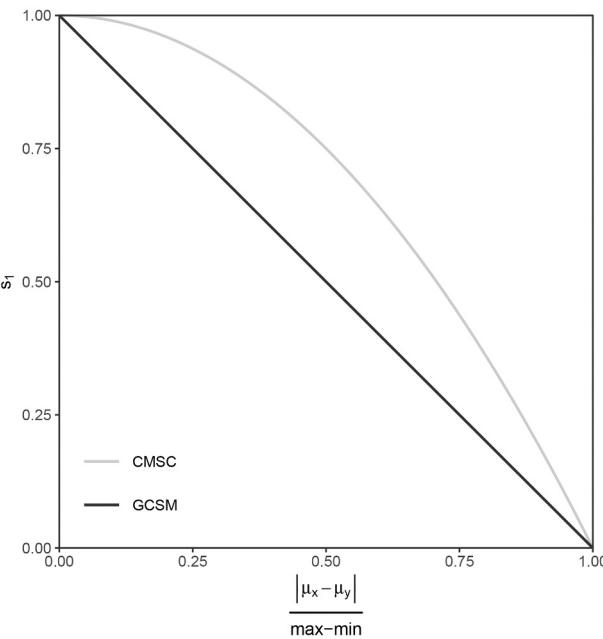
$$s_2 = 1 - \frac{|\sigma_x - \sigma_y|}{(\max - \min)/2}, \quad (2)$$

where max and min refer to the global maximum and minimum, respectively. The denominator was incorporated for normalizing the difference (Palubinskas, 2017).

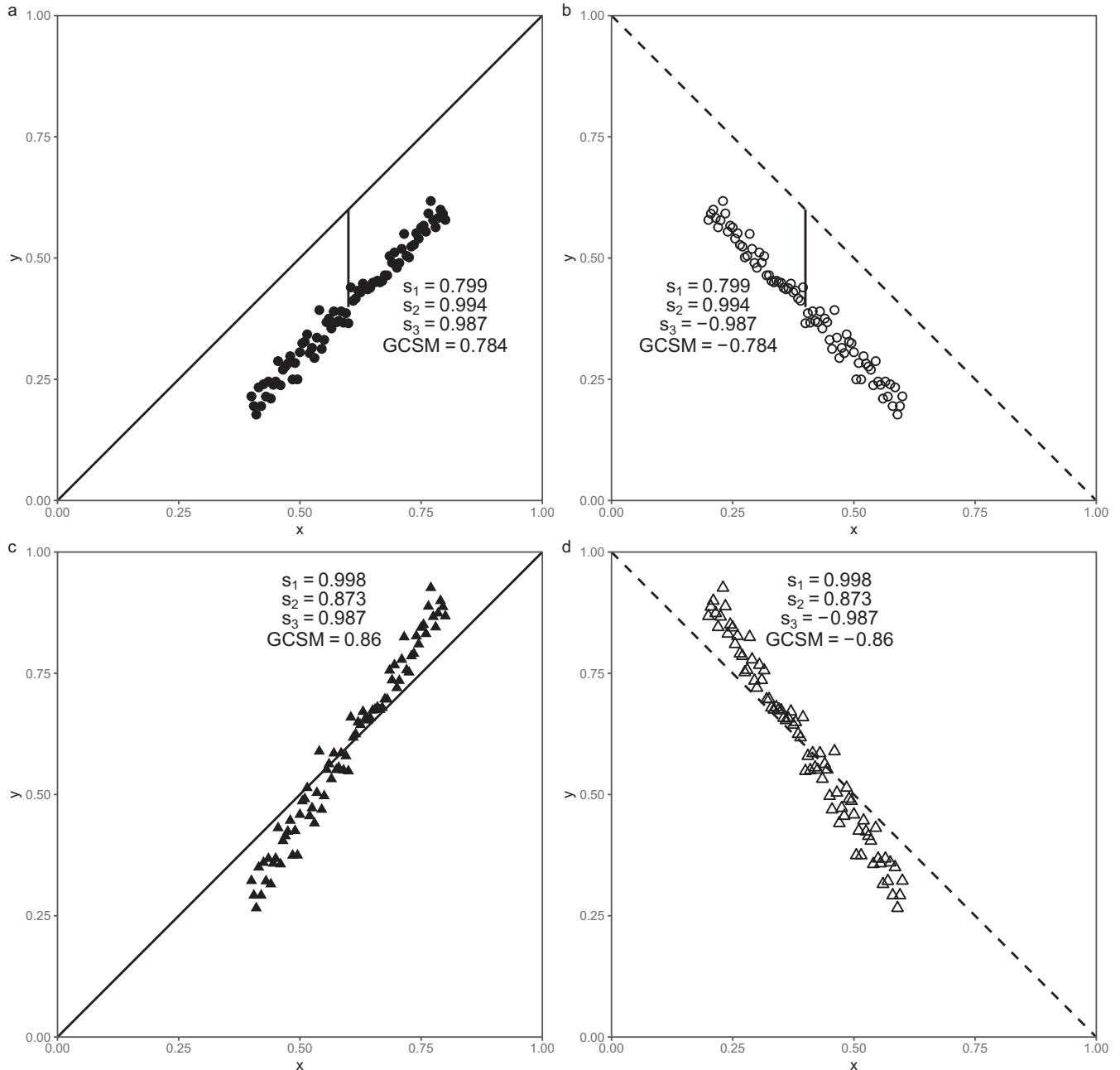
The  $s_1$  is computed according to the type of correlation, e.g., positive or negative. In the situation of a positive  $s_3$ , the normalized absolute difference between local means  $|\mu_x - \mu_y|/(\max - \min)$  specifies the vertical distance from  $(\mu_x, \mu_y)$  to the identity line (Fig. 2c). In negative cases, the vertical distance from  $(\mu_x, \mu_y)$  to the negative diagonal line,  $|\min + \max - \mu_x - \mu_y|/(\max - \min)$ , is adopted (Fig. 2d). The definition of  $s_1$  is given as:

$$s_1 = \begin{cases} 1 - \frac{|\mu_x - \mu_y|}{\max - \min} & \text{if } s_3 \geq 0; \\ 1 - \frac{|\min + \max - \mu_x - \mu_y|}{\max - \min} & \text{if } s_3 < 0. \end{cases} \quad (3)$$

The maximum GCSM would be obtained as 1 if and only if  $x$  and  $y$  conform to the identity line. Conversely, the minimum GCSM, i.e.,  $-1$ , would be uniquely achieved if  $x$  and  $y$  conform to the negative diagonal



**Fig. 3.** The difference between the  $s_1$  in CMSC and GCSM if  $s_3 \geq 0$ .  $\mu_x$  and  $\mu_y$  are local means of  $x$  and  $y$  within a window, respectively. The max and min are the global maximum and minimum, respectively.



**Fig. 4.** The GCSM and its components between four pairs of  $x$  and  $y$  at  $9 \times 9$  spatial windows. The  $x$  was evenly derived from (a, c) [0.4, 0.8] and (b, d) [0.2, 0.6], respectively. The  $y$  was determined with different functions: (a)  $y = x - 0.2 + \varepsilon_1$ , (b)  $y = -x + 0.8 + \varepsilon_1$ , (c)  $y = 1.5x - 0.3 + \varepsilon_2$ , and (d)  $y = 1.5x + 1.2 + \varepsilon_2$ , where  $\varepsilon_1$  and  $\varepsilon_2$  are random noises. The vertical lines represent the vertical distances from  $(\mu_x, \mu_y)$  to (a) the identity line, and (b) the negative diagonal line, respectively. The global minimum and maximum were assumed as 0 and 1, respectively.

line.

The GCSM could be regarded as a comprehensive (dis)similarity measure between patterns. The penalty is given to the correlation coefficient through  $s_1$  due to the (dis)agreement between the local means. The  $s_2$  quantifies the agreement between their standard deviations. Fig. 4 illustrates the impacts of  $s_1$  and  $s_2$  with some examples. The normalized Euclidean distances in Fig. 4a and b are 0.201, so the  $s_1$  are both obtained as 0.799, indicating the strengths of agreement and disagreement between local means, respectively. The relatively low  $s_2$  in Fig. 4c and d are caused by the relatively small strengths of agreement between the standard deviations of  $x$  and  $y$ .

Enhancements of GCSM to CMSC can be grouped into two types. The

first enhancement proposes the use of non-squared Euclidean distance instead of the squared one for  $s_1$  and  $s_2$ . The second one introduces the dependency of  $s_1$  on the sign of  $s_3$ . To thoroughly investigate their effects, the CMSC was separately modified with one type of enhancement and subjected to further analysis. For convenience, we called the CMSC incorporating non-squared Euclidean distance as CMSC\_E1, and the other one with  $s_1$  being dependent on the sign of  $s_3$  as CMSC\_E2. The equations are listed in Table 1.

### 2.3. Normalization preprocessing

Another limit for utilizing composite measures would be the

difference in the scales of patterns, e.g., GPP and precipitation. In such a case, the  $s_3$  could measure the strength of the correlation. However, it alone could not provide a universal criterion for all windows where the underlying regression lines may be diverse considerably. Moreover, both  $s_1$  and  $s_2$  are scale-dependent assuming two patterns are at the same scale, thus could be dominated by the pattern which has larger values while being insensitive to the other one (Armstrong and Collopy, 1992). As a result, the composite measures may not correctly capture the relative (dis)similarity between patterns having unequal scales.

A min-max normalization is applied to rescale patterns to the same scale before computing composite measures. The formula is as follows:

$$f(v) = \frac{v - \min_v}{\max_v - \min_v}, \quad (4)$$

where  $v$  refers to the original value of a pattern  $v$ . The  $\min_v$  and  $\max_v$  are custom parameters setting the target range of interest. By selecting these parameters for both patterns, the positive diagonal line passing through  $(\min_x, \min_y)$  and  $(\max_x, \max_y)$  and the negative one passing through  $(\min_x, \max_y)$  and  $(\max_x, \min_y)$  are consequently constructed. The two lines define the universal criteria of relative similarity and dissimilarity, respectively, for all windows in the entire region of interest. These parameters can also be extracted from the ranges of given data if no prior knowledge is available. For patterns having the same scale, this normalization would not affect the results of composite measures considered in this paper, thus can be skipped.

#### 2.4. Spatial and temporal GCSM

The map of GCSM can be achieved by sliding the spatial window cell by cell (Fig. 1a). It presents the distribution of the relationship between spatial patterns over the entire region of interest. Likewise, the window could be set along the time axis to compare temporal patterns (Fig. 1b). For convenience, the GCSM obtained from two types of windows are called spatial and temporal GCSM, respectively. Both types of windows could slide over space or time to reveal the spatial or temporal distribution of the relationship, respectively. Besides, a spatiotemporal cube can be regarded as a window without distinguishing the space and time axes.

### 3. Sensitivity test and case studies

We conducted a series of numerical experiments to assess the benefits of our enhancements. The applicability of GCSM was shown with two case studies comparing two GPP products and exploring the relationship between GPP and precipitation, respectively.

#### 3.1. Sensitivity test

The sensitivities of GCSM to mean shift and random noise were examined with numerical experiments in comparison to SSIM and CMSC. The sensitivities of SSIM and CMSC have been conducted for some types of distortion, such as mean shift, contrast stretching, random noise, and blurring in the image processing field (Palubinskas, 2017; Wang and Bovik, 2009). In the present study, we focused on mean shift and random noise, which may often occur in ecological studies, such as the residual of a spatial model. A gridded product of terrestrial GPP was used and subjected to noises to create distorted variants for comparison. GPP, which is the amount of carbon photosynthetically assimilated by vegetation over per unit space and time (Monteith, 1972), is usually used to represent the productivity of a terrestrial ecosystem (Beer et al., 2010). The GPP is often estimated because there is no direct observation at a global scale. Several products of GPP have been generated with large deviations among them (Anav et al., 2015; Piao et al., 2013).

The GPP product from the Max Planck Institute for Biogeochemistry was utilized, referred to as **MPI**. It was produced inputting remote

sensing and climate data to an empirical model, which was trained against flux tower observations with a machine learning method (Jung et al., 2009, 2011). The monthly data, which are available from 1982 to 2011 at the spatial resolution of  $0.5^\circ$ , were collected from the data portal (<https://www.bgc-jena.mpg.de/geodb/BGI/Home>). The data were then aggregated to a gridded map representing the mean annual GPP at each cell. Constant values and random noises were added to this gridded map to create 100 distorted variants, denoted by **MPI<sub>shift</sub>** and **MPI<sub>noise</sub>** respectively. The constants varied from  $-1.62$  to  $1.62$ , which representing  $-50\%$  to  $50\%$  of the range of values in the gridded map. The noises were derived from normal distributions  $N(0, \sigma^2)$ , with the  $\sigma$  ranging from 0 to  $0.1 \text{ kg C m}^{-2}$  indicating different distortion levels.

The SSIM, CMSC, CMSC\_E1, CMSC\_E2, and GCSM between the **MPI** and 100 variants were computed, respectively. The global minimum and maximum in the **MPI**, i.e., 0 and  $3.24 \text{ kg C m}^{-2}$ , were fixed as normalization parameters to rescale all datasets before computing these measures. In fact, these two parameters were used to specify a unique reference for all 100 pairs of comparisons, normalizing or not would not alter the outcomes because all datasets are at the same scale. The spatial window was set to be  $9 \times 9$  cells, which represents an area of  $4.5^\circ \times 4.5^\circ$ . By sliding the window cell by cell, the gridded maps of these measures were produced. The global average of each measure was then determined for each comparison.

Besides, an arbitrary variable **MPI'** was created as the perfect anti-analog of **MPI** by reversing its low and high values:

$$\mathbf{MPI}' = -\mathbf{MPI} + \min(\mathbf{MPI}) + \max(\mathbf{MPI}) \quad (5)$$

Then the same procedure as above was performed between **MPI'** and 100 **MPI<sub>shift</sub>** and **MPI<sub>noise</sub>** variants to examine the validity of these measures for quantifying the dissimilarity between patterns.

#### 3.2. Case study I: temporal GCSM between MOD and MPI

The annual GPP derived from **MPI** was compared with that from another product, MOD17A3 (**MOD**), to assess the temporal (dis)similarity between two products. The ability to reproduce the temporal variability has been regarded as an essential characteristic of GPP products (Anav et al., 2015; Jiang and Ryu, 2016). The **MOD** was generated with a light use efficiency model incorporating photosynthetically active radiation, leaf area index, and climatic variables (Zhao et al., 2005; Zhao and Running, 2010). Annual GPP data during the period of 2000–2015 at a spatial resolution of 1 km were obtained from the Numerical Terradynamic Simulation Group at the University of Montana (<http://files.ntsg.umt.edu>). The data were averaged to the corresponding cell of **MPI** for comparison. The temporal GCSM between annual **MPI** and **MOD** was computed at each cell, setting 2000–2011 as the temporal window.

#### 3.3. Case study II: spatial GCSM between GPP and precipitation

The relative (dis)similarity between GPP and precipitation over space was evaluated with spatial GCSM. The spatial distribution of GPP primarily depends on climatic conditions (Anav et al., 2015). Thus, assessing the relationships between GPP and climatic variables would help in understanding the terrestrial ecosystem (Yao et al., 2017). The Land-surface precipitation product at  $0.5^\circ$  resolution from the Global Precipitation Climatology Centre was used in the present study (Schneider et al., 2018). Similar to the GPP product from the Max Planck Institute for Biogeochemistry, the precipitation data during 1982–2011 were aggregated to obtain the mean annual precipitation. Due to the huge variability of precipitation across the globe, the global regression line between GPP and precipitation was derived, based on which the normalization parameters were determined for two variables. After rescaling, the spatial GCSM was computed with a window of  $4.5^\circ \times 4.5^\circ$ , which is the same as Beer et al. (2010) in which the partial correlation

coefficients were calculated between GPP and climatic variables.

## 4. Results

### 4.1. The sensitivity to mean shift

When comparing **MPI** with **MPI<sub>shift</sub>**, all measures decreased with the rising magnitude of shift (Fig. 5a). The GCSM linearly declined from 1 to 0.5 as the **MPI** shifted by 50%, either positively or negatively. The CMSC, in contrast, reduced with a slower rate, especially when the shift magnitude was small. It was achieved as 0.75 at the 50% shift. The CMSC\_E1 and CMSC\_E2 were identical to GCSM and CMSC, respectively, since the second enhancement takes no effect in this case. Besides, the SSIM had the highest rate if adding positive values to **MPI**, but inordinately declined when shifted to the left. The average SSIM started to be negative with the -22% shift.

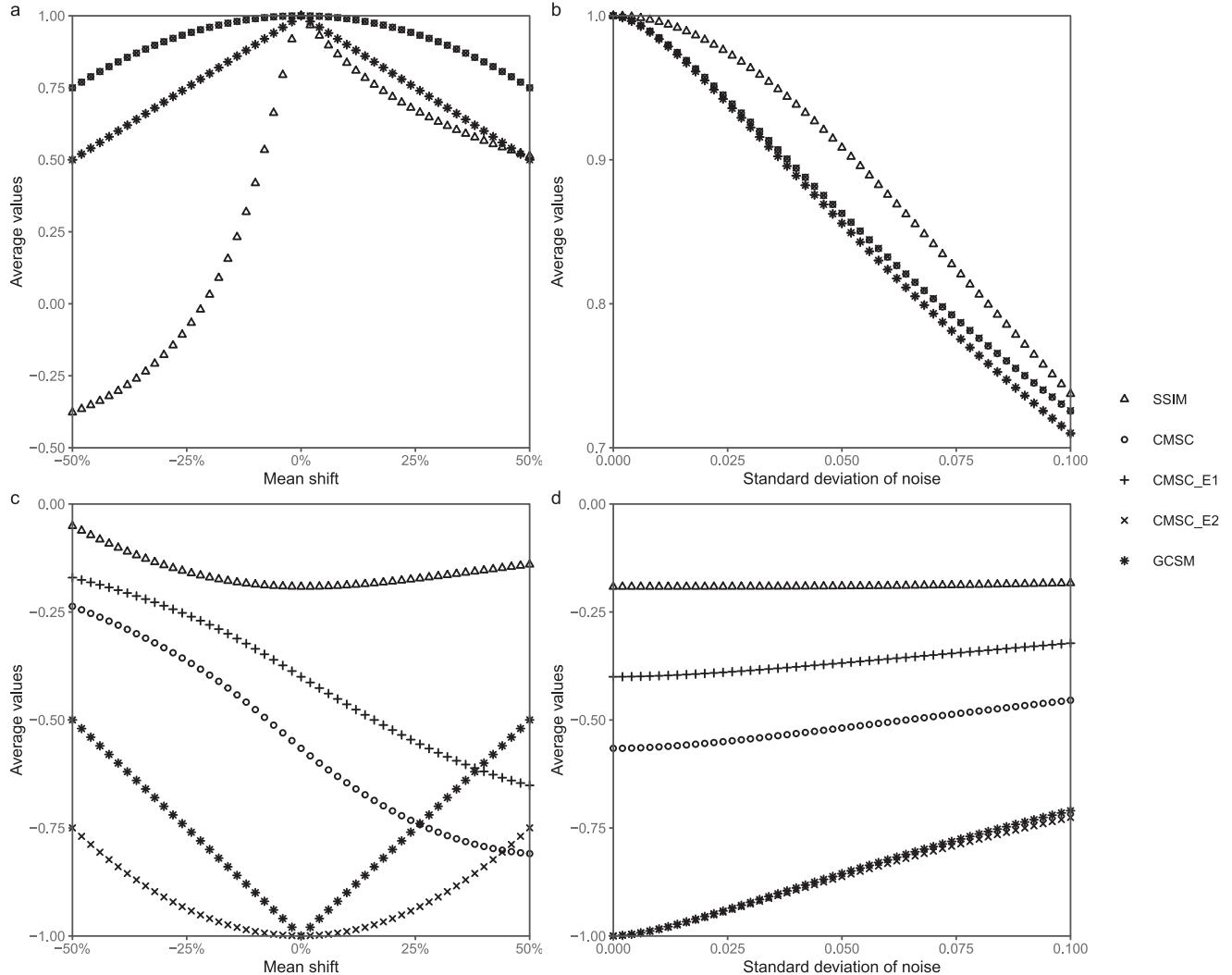
For comparisons between **MPI'** and **MPI<sub>shift</sub>**, the GCSM and CMSC\_E2 varied symmetrically to the cases between **MPI** and **MPI<sub>shift</sub>** (Fig. 5c). The signs were changed and the absolute values were identical at the same shift magnitude. Both CMSC and CMSC\_E1 showed monotonic increases when shifting from left to right. The average CMSC, CMSC\_E1,

and SSIM between **MPI'** and **MPI** were -0.57, -0.40, and -0.19, respectively, rather than -1.

### 4.2. The sensitivity to random noise

All measures between **MPI** and **MPI<sub>noise</sub>** decreased with the rising size of random noise (Fig. 5b). The GCSM had the most considerable decreasing rate among them. It was obtained as 1 when no noise was applied. When the standard deviation of noise was set to be 0.1 kg C m<sup>-2</sup>, the average GCSM declined to 0.71. Besides, the CMSC was slightly higher than the GCSM. For example, the difference between them was only 0.015 at the maximum size of noise. The average CMSC\_E1 was nearly identical to GCSM, and the average CMSC\_E2 was nearly identical to CMSC. In contrast, the SSIM showed a slower falling rate, especially when the size of the noise was small.

For the dissimilarity between **MPI'** and **MPI<sub>noise</sub>**, the GCSM increased at the highest rate with the rising noise size (Fig. 5d). The absolute values of negative GCSM and CMSC\_E2 between **MPI'** and **MPI<sub>noise</sub>** were identical to the positive ones between **MPI** and **MPI<sub>noise</sub>** under the same distortion level. In contrast, the other three measures increased at considerably low rates, especially the SSIM.



**Fig. 5.** The average SSIM, CMSC, CMSC\_E1, CMSC\_E2, and GCSM between (a) **MPI** and **MPI<sub>shift</sub>**, (b) **MPI** and **MPI<sub>noise</sub>**, (c) **MPI'** and **MPI<sub>shift</sub>**, (d) **MPI'** and **MPI<sub>noise</sub>**. **MPI** is the mean annual GPP derived from the Max Planck Institute for Biogeochemistry. The **MPI<sub>shift</sub>** and **MPI<sub>noise</sub>** are generated by adding constant values and random noise, respectively. **MPI'** is an arbitrary variable reversing the low and high values of **MPI**.

#### 4.3. Temporal (dis)similarity between MPI and MOD

The temporal (dis)similarity between annual GPP from **MPI** and **MOD** differed by region (Fig. 6). The majority, e.g., about 88% of cells, had positive GCSM with an average of 0.612. Negative GCSM mostly occurred in tropical regions such as the Amazon rainforest, the Conga Basin, and Southeast Asia. In these areas, the average of negative GCSM was  $-0.11$ , indicating the strength of dissimilarity was low between two products. Such a result was mainly contributed by the relatively small values of  $s_1$  and  $s_2$ . For example, the  $s_2$  in Southeast Asia was low compared with the global average 0.986. The  $s_1$  was also not high in cells having negative GCSM with an average of 0.492. In comparison, the average  $s_1$  in cells having positive GCSM was 0.959.

#### 4.4. Relative spatial (dis)similarity between GPP and precipitation

The relative (dis)similarity between GPP and precipitation showed a heterogeneous pattern over space (Fig. 7). About 82.9% of cells had positive GCSM with an average of 0.495. Other cells showed negative values, mostly distributed in tropical and boreal forests. The absolute value of GCSM tended to be less than that of  $s_3$ , especially for negative cases. The  $s_3$  varied from  $-1$  to  $1$ , whereas the minimum GCSM was only  $-0.774$ . The differences between GCSM and  $s_3$  were mainly produced by the small values of  $s_1$ . For example, the average of  $s_1$  in cells having negative GCSM was only 0.332. Besides, the agreement between the standard deviations of GPP and precipitation varied over space, indicated by the broad range of  $s_2$ , i.e., 0.302 to 1. Relatively low  $s_2$  mainly occurred in edges of the mountainous area.

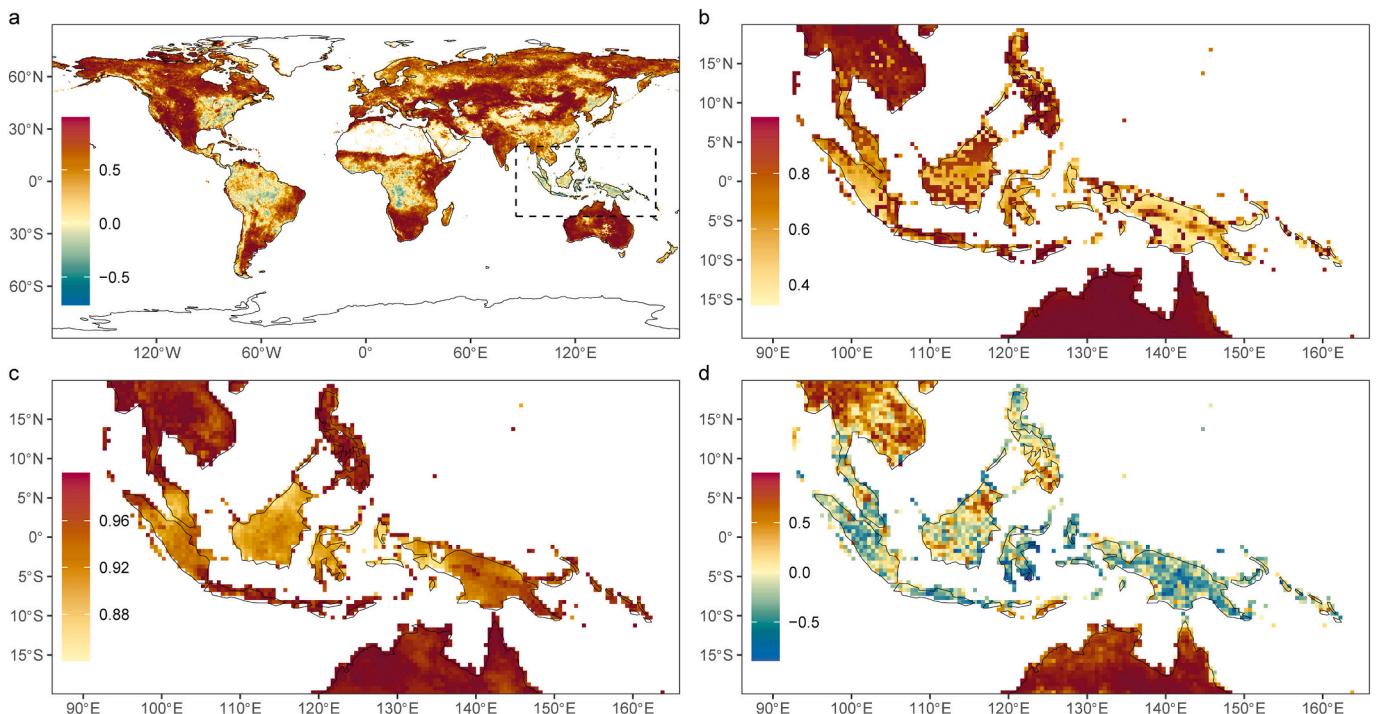
### 5. Discussion

Our results demonstrated that the generic composite similarity measure (GCSM) has advantages in evaluating the relationship between ecological variables compared with preexisting composite measures. The GCSM shows higher sensitivity to small differences than SSIM and

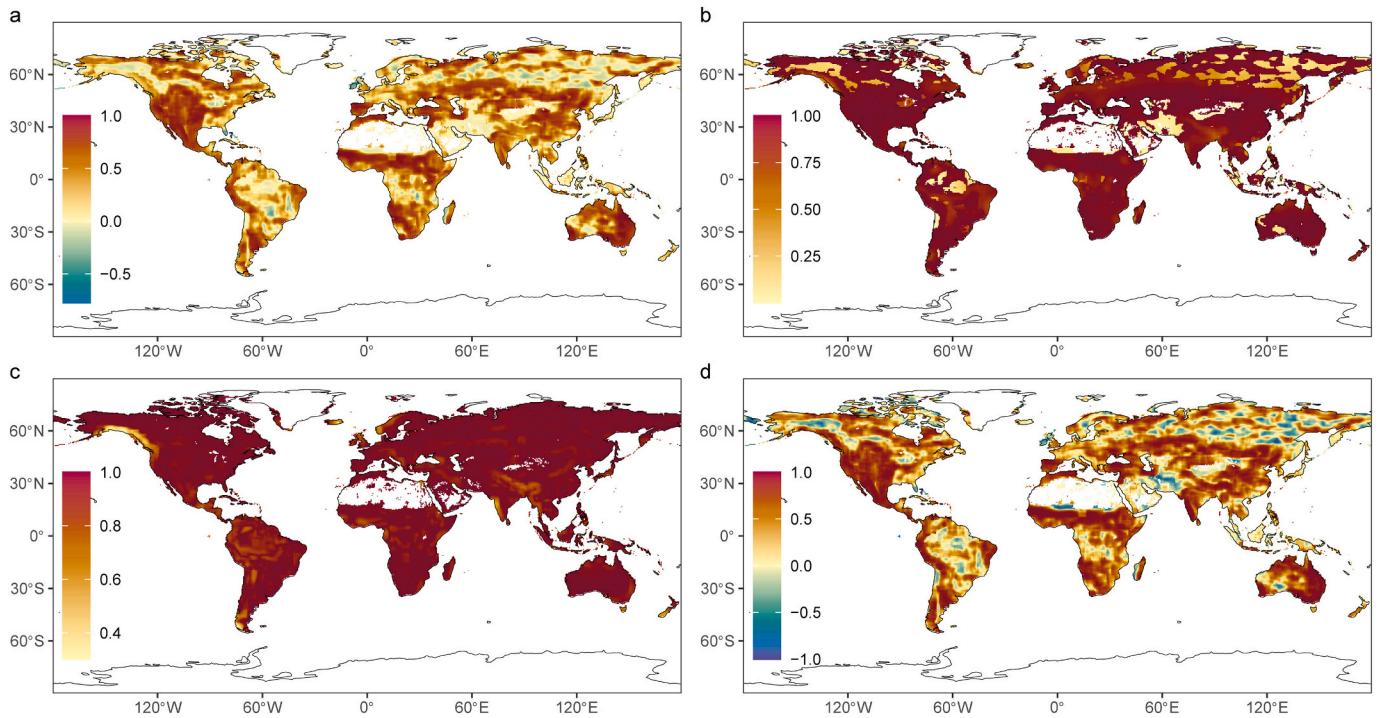
CMSC. It is capable of quantifying the dissimilarity between patterns. The normalization preprocessing allows assessing the relative (dis)similarity between patterns having unequal scales. In general, the GCSM would introduce new opportunities for assessing the heterogeneous relationship in ecology. The evaluation can be extended to a broad range of scenarios, such as similarity and dissimilarity, absolute and relative (dis)similarity, spatial and temporal (dis)similarity.

The sensitivity test indicated the GCSM was more sensitive to small differences than SSIM and CMSC. When the shift magnitude rises from zero, the non-squared Euclidean distance used in GCSM increases more substantially than its square in CMSC (Fig. 3), leading to the slightly higher sensitivity of GCSM (Fig. 5a). More importantly, the non-squared distance gives equal emphasis over the scope of differences, making the GCSM more comparable across the entire region of interest. The GCSM also showed higher sensitivity to small random noise than SSIM (Fig. 5b). Meanwhile, inheriting from the CMSC, the  $s_1$  and  $s_2$  in GCSM depend only on the difference between local statistics (Fig. 2c). These properties result in the advantage of GCSM when assessing the absolute similarity over CMSC and SSIM.

The GCSM could also accurately quantify the dissimilarity between patterns. A negative correlation coefficient indicates two patterns are negatively correlated. In such a case, the  $s_1$  needs to be adapted to give a reasonable composite dissimilarity measurement. Symmetrical to positive cases, the vertical distance  $|\min + \max - \mu_x - \mu_y| / (\max - \min)$  represents the difference from the pair of local means to the perfect antianalog (Fig. 2d). Therefore, the GCSM would have a unique minimum  $-1$  if and only if two patterns fit the negative diagonal line  $y = -x + \min + \max$ . The absolute value of negative GCSM provides a reasonable quantification of the strength of dissimilarity. In contrast, the CMSC could not correctly measure the dissimilarity, although it increases with the distortion level (Fig. 5d). For example, the  $s_2$  and  $s_3$  in CMSC between **MPI** and **MPI'** are 1 and  $-1$ , respectively, but  $s_1$  is not always 1 and would vary by the difference between local means (Fig. 2b). Therefore, the CMSC is unknown even between one pattern and its perfect antianalog.



**Fig. 6.** Distributions of the temporal GCSM and its components between two GPP products, **MPI** and **MOD**, which are derived from the Max Planck Institute for Biogeochemistry and the Numerical Terradynamic Simulation Group at the University of Montana, respectively. The (a) shows the GCSM over the globe, and (b-d) are the  $s_1$ ,  $s_2$ , and  $s_3$ , respectively, in Southeast Asia.



**Fig. 7.** Distributions of the spatial GCSM and its components between mean annual GPP and precipitation, which are derived from the Max Planck Institute for Biogeochemistry and the Global Precipitation Climatology Centre, respectively. The (a) GCSM is determined by multiplying (b)  $s_1$ , (c)  $s_2$ , and (d)  $s_3$  obtained in a  $9 \times 9$  spatial window.

The comparison between **MPI** and **MOD** showed consistent results with previous studies (Anav et al., 2015; Jiang and Ryu, 2016). In cells having positive GCSM, considerable large  $s_1$  was obtained with an average of 0.959, indicating the high agreement between the means in two products, as shown by Jiang and Ryu (2016). The interannual variabilities of **MPI** and **MOD** showed differences in tropical regions and the Southern United States, which was affirmed by the negative GCSM and relatively low  $s_2$  in these areas (Fig. 6c). In particular, the standard deviation in **MPI** tended to be less than that in **MOD** (Fig. A.1c), most likely since the interannual variability of GPP was weakly reproduced by **MPI**, as suggested by Anav et al. (2015). Besides, the absolute value of negative GCSM was small, indicating the strength of dissimilarity between **MPI** and **MOD** was not strong (Fig. 6a). Although two products showed negative correlations in some areas, their means, however, were still close to each other (Fig. A.1b). In such a case, the GCSM could give a reasonable evaluation of the dissimilarity by setting the negative diagonal line as the criterion. Otherwise, the absolute value of the composite measure would be considerably higher.

The distribution of the relationship type between GPP and precipitation, i.e., positive or negative, was similar to a previous study (Beer et al., 2010). In most temperate regions, the GCSM was positive, indicating the positive impact of precipitation on GPP (Fig. 7a). While in some regions, such as tropical and boreal forests, GPP tended to have negative correlations with precipitation (Fig. 7d). The negative GCSM may reflect the indirect effect of precipitation on GPP by influencing other variables. GPP is the integrated result of several processes in a terrestrial ecosystem, affected by a broad range of factors. In areas showing negative GCSM, either solar radiation or temperature has a strong correlation with GPP, thus could be the dominating climatic factor (Anav et al., 2015). Therefore, the negative GCSM could be caused by the negative correlations between precipitation and other climatic variables. For example, the growth of tropical rainforests is mainly restricted by low solar radiation due to high cloudiness (Zhao and Running, 2010). Still, the absolute values of GCSM were less than those of the correlation coefficient because of the differences between local

regression lines and the global one (Fig. A.2).

Our results suggested that efforts are needed to explore the relationships between GPP and climatic variables over small areas. In about 13% of land area, the GPP showed negative relationships with the precipitation. However, their local averages were still positively correlated when putting in the global context (Fig. A.1b). Surprisingly, the agreement between their averages showed no clear differences with the cells showing positive correlation coefficients. Also, the regions with complex topography tended to have less strength of agreement in terms of the standard deviation. Thus, it would be helpful to examine the spatial pattern of the dominating factor of GPP for further improving GPP estimates.

The GCSM could give a comprehensive evaluation of (dis)similarity between patterns at the same scale with three components. The correlation coefficient measures the strength of the joint variability of two patterns within a window. The  $s_1$  and  $s_2$  quantify the deviations between the local regression line and the global positive or negative diagonal line by comparing their local features, i.e., the mean and standard deviation. By multiplying three terms, the GCSM provides a stricter assessment of the agreement or disagreement (Fig. 4). For example, the temporal correlation coefficient between **MPI** and **MOD** ranged from -0.904 to 0.997, whereas the minimum and maximum GCSM were -0.755 and 0.984, respectively.

The normalization preprocessing makes the GCSM feasible to assess the relative (dis)similarity between patterns having unequal scales. Normalizing the patterns to the same scale allows to obtain reasonable  $s_1$  and  $s_2$  because the mean and standard deviation are scale-dependent. By selecting the normalization parameters, the global positive and negative diagonal lines are constructed as the universal criteria of relative similarity and dissimilarity, respectively. Therefore, the obtained values of GCSM are comparable over the entire region of interest. In the case study between GPP and precipitation, the normalization parameters were determined based on the global regression line. In such a case, the positive GCSM would be less than the correlation coefficient due to the differences between local regression lines and the global one (Fig. A.2).

Still, the normalization parameters can be customized based on the purpose of research and domain knowledge. For example, in the sensitivity test, the MPI was the reference to compare with, thus the minimum and maximum values were used to rescale all datasets.

The GCSM provides a unified framework for (dis)similarity assessment that could be applied in ecology and other disciplines. If both patterns are at the same scale, the GCSM assesses the similarity or dissimilarity between them. For patterns having unequal scales, the relative (dis)similarity could be quantified under universal criteria. Besides, by setting the window over space, time, or both, it would capture the relationships between spatial, temporal, or combined spatiotemporal patterns. And sliding the window over space or time allows obtaining the spatial or temporal distribution of the relationship, respectively.

## Author contributions

YL and KSK conceived and designed the algorithm; YL collected and analyzed the data. YL and KSK led the writing of the manuscript. RB and DF revised the manuscript. All authors contributed critically to the draft and gave final approval for publication.

## Declaration of Competing Interest

None

## Acknowledgments

This study was supported by Rural Development Administration, Republic of Korea, under “Cooperative Research for Agriculture Science & Technology Development (PJ010115022016)” program.

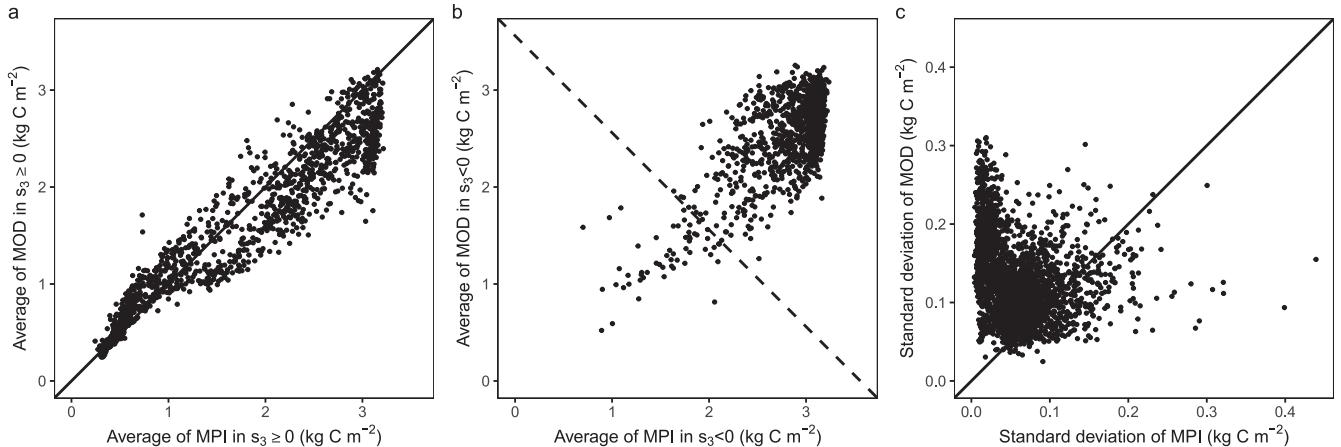
## Appendix A

**Table A.1**

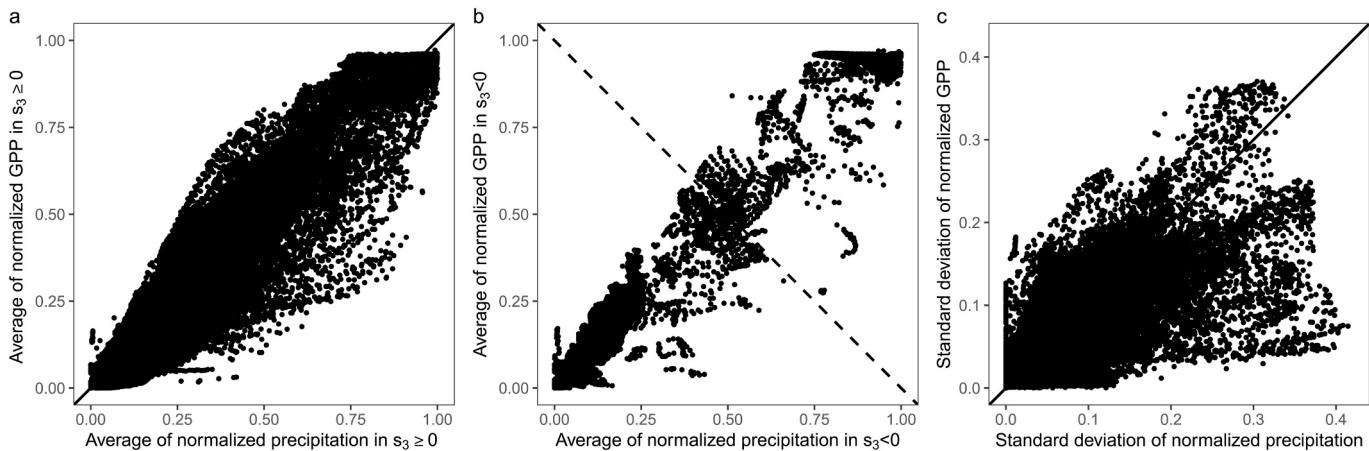
Local statistics used in the structural similarity (SSIM; Wang et al., 2004), composite similarity measure based on means, standard deviations, and correlation coefficient (CMSC; Palubinskas, 2017), and generic composite similarity measure (GCSM).

Statistic	SSIM	CMSC and GCSM
$\mu_x$	$\sum_{i=1}^n w_i x_i$	$\frac{1}{n} \sum_{i=1}^n x_i$
$\mu_y$	$\sum_{i=1}^n w_i y_i$	$\frac{1}{n} \sum_{i=1}^n y_i$
$\sigma_x^2$	$\sum_{i=1}^n w_i (x_i - \mu_x)^2$	$\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu_x)^2$
$\sigma_y^2$	$\sum_{i=1}^n w_i (y_i - \mu_x)^2$	$\frac{1}{n-1} \sum_{i=1}^n (y_i - \mu_y)^2$
$\sigma_{xy}$	$\sum_{i=1}^n w_i (x_i - \mu_x)(y_i - \mu_y)$	$\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y)$

n is the number of cells in a window. w is the weight function based on the distance between cell i and the center of the window.



**Fig. A.1.** The relationship between statistics of two GPP products, MPI and MOD, obtained in the 2000–2011 temporal window over the Southeast Asia. (a) and (b) show the means in cells having positive and negative  $s_3$ , respectively; (c) shows the standard deviations. MPI and MOD are derived from the Max Planck Institute for Biogeochemistry and the Numerical Terradynamic Simulation Group at the University of Montana, respectively.



**Fig. A.2.** The relationship between statistics of mean annual GPP and precipitation obtained in  $9 \times 9$  spatial windows. (a) and (b) show the means in cells showing positive and negative  $s_3$ , respectively; (c) shows the standard deviations. GPP and precipitation are derived from the Max Planck Institute for Biogeochemistry and the Global Precipitation Climatology Centre, respectively.

## References

- Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., Murray-Tortarolo, G., Papale, D., Parazoo, N.C., Peylin, P., Piao, S., Sitch, S., Viovy, N., Wiltshire, A., Zhao, M., 2015. Spatiotemporal patterns of terrestrial gross primary production: a review. *Rev. Geophys.* 53, 785–818. <https://doi.org/10.1002/2015rg000483>.
- Armstrong, J.S., Collopy, F., 1992. Error measures for generalizing about forecasting methods: empirical comparisons. *Int. J. Forecast.* 8, 69–80. [https://doi.org/10.1016/0169-2070\(92\)90008-w](https://doi.org/10.1016/0169-2070(92)90008-w).
- Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rodenbeck, C., Arain, M.A., Baldocchi, D., Bonan, G.B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K.W., Roupsard, O., Veenendaal, E., Viovy, N., Williams, C., Woodward, F.I., Papale, D., 2010. Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. *Science* 329, 834–838. <https://doi.org/10.1126/science.1184984>.
- Brunet, D., 2012. A study of the structural similarity image quality measure with applications to image processing (PhD thesis). In: UWSpace.
- Brunet, D., Vrscay, E.R., Wang, Z., 2012. On the mathematical properties of the structural similarity index. *IEEE Trans. Image Process.* 21, 1488–1499. <https://doi.org/10.1109/tip.2011.2173206>.
- Dossmann, R., Yang, X.D., 2009. A comprehensive assessment of the structural similarity index. *SIVIP* 5, 81–91. <https://doi.org/10.1007/s11760-009-0144-1>.
- Fensholt, R., Proud, S.R., 2012. Evaluation of earth observation based global long term vegetation trends comparing GIMMS and MODIS global NDVI time series. *Remote Sens. Environ.* 119, 131–147. <https://doi.org/10.1016/j.rse.2011.12.015>.
- Fotheringham, A.S., Brunsdon, C., Charlton, M., 2002. *Geographically Weighted Regression*. John Wiley & Sons.
- Hagen-Zanker, A., 2006. Comparing Continuous Valued Raster Data: A Cross Disciplinary Literature Scan.
- Jackson, E.K., Roberts, W., Nelsen, B., Williams, G.P., Nelson, E.J., Ames, D.P., 2019. Introductory overview: error metrics for hydrologic modelling a review of common practices and an open source library to facilitate use and adoption. *Environ. Model. Softw.* 119, 32–48. <https://doi.org/10.1016/j.envsoft.2019.05.001>.
- Jiang, C., Ryu, Y., 2016. Multi-scale evaluation of global gross primary productivity and evapotranspiration products derived from breathing earth system simulator (BESS). *Remote Sens. Environ.* 186, 528–547. <https://doi.org/10.1016/j.rse.2016.08.030>.
- Jones, E.L., Rendell, L., Pirotta, E., Long, J.A., 2016. Novel application of a quantitative spatial comparison tool to species distribution data. *Ecol. Indic.* 70, 67–76. <https://doi.org/10.1016/j.ecolind.2016.05.051>.
- Jung, M., Reichstein, M., Bondeau, A., 2009. Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model. *Biogeosciences* 6, 2001–2013. <https://doi.org/10.5194/bg-6-2001-2009>.
- Jung, M., Reichstein, M., Margolis, H.A., Cescatti, A., Richardson, A.D., Arain, M.A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B.E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E.J., Papale, D., Sottocornola, M., Vaccari, F., Williams, C., 2011. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *J. Geophys. Res.* 116 <https://doi.org/10.1029/2010jg001566>.
- Kufareva, I., Abagyan, R., 2012. In: Orry, A.J.W., Abagyan, R. (Eds.), *Methods of protein structure comparison*. Human Press, pp. 231–257. [https://doi.org/10.1007/978-1-61779-588-6\\_10](https://doi.org/10.1007/978-1-61779-588-6_10).
- Kuhn, M., Johnson, K., 2019. *Feature Engineering and Selection: A Practical Approach for Predictive Models*. CRC Press LLC, Milton.
- Legates, D.R., McCabe, G.J., 1999. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* 35, 233–241. <https://doi.org/10.1029/1998wr900018>.
- Legendre, P., Legendre, L., 2012. *Spatial Analysis*, in: Numerical Ecology. Elsevier LTD, Oxford, pp. 785–858. <https://doi.org/10.1016/b978-0-444-53868-0.50013-7>.
- Li, Y., Zeng, Z., Zhao, L., Piao, S., 2015. Spatial patterns of climatological temperature lapse rate in mainland China: a multi-time scale investigation. *J. Geophys. Res.-Atmos.* 120, 2661–2675. <https://doi.org/10.1002/2014jd022978>.
- Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. *Glob. Ecol. Biogeogr.* 17, 145–151. <https://doi.org/10.1111/j.1466-8238.2007.00358.x>.
- Long, J., Robertson, C., 2017. Comparing spatial patterns. *Geogr. Compass* 12, e12356. <https://doi.org/10.1111/gec3.12356>.
- McCuen, R.H., Snyder, W.M., 1975. A proposed index for comparing hydrographs. *Water Resour. Res.* 11, 1021–1024. <https://doi.org/10.1029/wr011i006p01021>.
- Mo, R., Ye, C., Whitfield, P.H., 2014. Application potential of four nontraditional similarity metrics in hydrometeorology. *J. Hydrometeorol.* 15, 1862–1880. <https://doi.org/10.1175/jhm-d-13-0140.1>.
- Monteith, J.L., 1972. Solar radiation and productivity in tropical ecosystems. *J. Appl. Ecol.* 9, 747. <https://doi.org/10.2307/2401901>.
- Moriasi, D.N., Arnold, J.G., Liew, M.W.V., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* 50, 885–900. <https://doi.org/10.13031/2013.23153>.
- Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Izumi, T., Izaurralde, R.C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T.A.M., Ray, D.K., Reddy, A., Rosenzweig, C., Ruane, A.C., Sakurai, G., Schmid, E., Skalsky, R., Song, C.X., Wang, X., de Wit, A., Yang, H., 2017. Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications. *Geosci. Model Dev.* 10, 1403–1422. <https://doi.org/10.5194/gmd-10-1403-2017>.
- Palubinskas, G., 2014. Mystery behind similarity measures MSE and SSIM. In: 2014 IEEE International Conference on Image Processing (ICIP). IEEE. <https://doi.org/10.1109/icip.2014.7025115>.
- Palubinskas, G., 2015. Joint quality measure for evaluation of pansharpening accuracy. *Remote Sens.* 7, 9292–9310. <https://doi.org/10.3390/rs70709292>.
- Palubinskas, G., 2017. Image similarity/distance measures: what is really behind MSE and SSIM? *Int. J. Image Data Fusion* 8, 32–53. <https://doi.org/10.1080/19497983.2016.1273259>.
- Pambrun, J.-F., Noumeir, R., 2015. Limitations of the SSIM quality metric in the context of diagnostic imaging. In: 2015 IEEE International Conference on Image Processing (ICIP). IEEE. <https://doi.org/10.1109/icip.2015.7351345>.
- Paul, W., Mielke, J., Berry, K.J., 2000. Euclidean distance based permutation methods in atmospheric science. *Data Min. Knowl. Disc.* 4, 7–27. <https://doi.org/10.1023/a:1009824531876>.
- Piao, S., Sitch, S., Ciais, P., Friedlingstein, P., Peylin, P., Wang, X., Ahlström, A., Anav, A., Canadell, J.G., Cong, N., Huntingford, C., Jung, M., Levis, S., Levy, P.E., Li, J., Lin, X., Lomas, M.R., Lu, M., Luo, Y., Ma, Y., Myeni, R.B., Poultier, B., Sun, Z., Wang, T., Viovy, N., Zaehle, S., Zeng, N., 2013. Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO<sub>2</sub> trends. *Glob. Chang. Biol.* 19, 2117–2132. <https://doi.org/10.1111/gcb.12187>.
- Robertson, C., Long, J.A., Nathoo, F.S., Nelson, T.A., Plouffe, C.C.F., 2014. Assessing quality of spatial models using the structural similarity index and posterior predictive checks. *Geogr. Anal.* 46, 53–74. <https://doi.org/10.1111/gean.12028>.
- Rodgers, J.L., Nicewander, W.A., 1988. Thirteen ways to look at the correlation coefficient. *Am. Stat.* 42, 59. <https://doi.org/10.2307/2685263>.
- Roux, R.L., de Rességuier, L., Corpetti, T., Jégou, N., Madelin, M., van Leeuwen, C., Quénol, H., 2017. Comparison of two fine scale spatial models for mapping

- temperatures inside winegrowing areas. *Agric. For. Meteorol.* 247, 159–169. <https://doi.org/10.1016/j.agrformet.2017.07.020>.
- Scheff, J., Seager, R., Liu, H., Coats, S., 2017. Are glacials dry? Consequences for paleoclimatology and for greenhouse warming. *J. Clim.* 30, 6593–6609. <https://doi.org/10.1175/jcli-d-16-0854.1>.
- Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., 2018. GPCC full data monthly version 2018.0 at 0.5°: monthly land-surface precipitation from rain-gauges built on gts-based and historic data. [https://doi.org/10.5676/dwd\\_gpcc/fd\\_m\\_v2018\\_050](https://doi.org/10.5676/dwd_gpcc/fd_m_v2018_050).
- Wang, Z., Bovik, A.C., 2002. A universal image quality index. *IEEE Signal Process. Lett.* 9, 81–84. <https://doi.org/10.1109/97.995823>.
- Wang, Z., Bovik, A.C., 2009. Mean squared error: love it or leave it? A new look at signal fidelity measures. *IEEE Signal Process. Mag.* 26, 98–117. <https://doi.org/10.1109/msp.2008.930649>.
- Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 600–612. <https://doi.org/10.1109/tip.2003.819861>.
- Wealands, S.R., Grayson, R.B., Walker, J.P., 2005. Quantitative comparison of spatial fields for hydrological model assessment—some promising approaches. *Adv. Water Resour.* 28, 15–32. <https://doi.org/10.1016/j.advwatres.2004.10.001>.
- Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* 30, 79–82. <https://doi.org/10.3354/cr030079>.
- Yao, Y., Wang, X., Li, Y., Wang, T., Shen, M., Du, M., He, H., Li, Y., Luo, W., Ma, M., Ma, Y., Tang, Y., Wang, H., Zhang, X., Zhang, Y., Zhao, L., Zhou, G., Piao, S., 2017. Spatiotemporal pattern of gross primary productivity and its covariation with climate in China over the last thirty years. *Glob. Chang. Biol.* 24, 184–196. <https://doi.org/10.1111/gcb.13830>.
- Zhao, M., Running, S.W., 2010. Drought-induced reduction in global terrestrial net primary production from 2000 through 2009. *Science* 329, 940–943. <https://doi.org/10.1126/science.1192666>.
- Zhao, M., Heinsch, F.A., Nemani, R.R., Running, S.W., 2005. Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote Sens. Environ.* 95, 164–176. <https://doi.org/10.1016/j.rse.2004.12.011>.