https://doi.org/10.5194/gmd-2024-82
Preprint. Discussion started: 30 May 2024

Geoscientific Model Development Discussions
Open Access

A Unified System for Evaluating, Ranking and Clustering in Diverse Scientific Domains

Zengyun Hu[1,2,3], Xi Chen[2,3,4*], Deliang Chen[5], Zhuo Zhang[2,3], Qiming Zhou[6,7], Qingxiang Li[8]

[1] School of Global Health, Chinese Center for Tropical Diseases Research, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

[2]State Key Laboratory of desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, Xinjiang 830011, China

[3]Research Center for Ecology and Environment of Central Asia,

Chinese Academy of Sciences, Urumqi, Xinjiang 830011, China

[4]University of Chinese Academy of Sciences, Beijing 100049, China

[5]Department of Earth Sciences, University of Gothenburg, Gothenburg 405 30, Sweden

[6]University of Wuhan, Wuhan, China

[7]Department of Geography, Hong Kong Baptist University, Kowloon, Hong Kong

[8]School of Atmospheric Sciences, Sun Yat-Sen University, Zhuhai, China

Correspondence to: Xi Chen, cx@ms.xjb.ac.cn

Abstract:

Evaluating, ranking, and clustering (ERC) stand as fundamental tasks in scientific research, each requiring a mathematical foundation. This study presents an ERC system anchored in the CCHZ-DISO (Chen, Chen, Hu, and Zhou-Distance between Indices of Simulation and Observation) system. Previous research underscores the optimality achieved by the CCHZ-DISO system (Hu et al., 2022). Since the inception of CCHZ-DISO-series research by Hu et al. (2019), DISO has found extensive applications across various domains including geography, hydrology, and economics. Analogous to the CCHZ-DISO system's construction, the ERC system employs the Euclidean distance to perform evaluating, ranking, and clustering tasks. Furthermore, illustrative examples are provided to elucidate the application of the ERC system. In fact, the ERC system unified the evaluating, ranking, and clustering tasks in one simple equation which is more flexible and simpler than the present system. It will have a more widely application than CCHZ-DISO in diverse scientific domains.

Keywords: CCHZ-DISO System; Euclidean Distance; Evaluating, Ranking and Clustering; ERC system.

1 Introduction

Evaluating, ranking, and clustering (ERC) represent essential tools in scientific research, providing qualitative and quantitative assessments of diverse objects within any scientific domain. The advent of big data, artificial intelligence, and extensive model simulation outputs has propelled ERC research into the spotlight, particularly for comprehensive analyses (Hu et al., 2023; Jacox et al., 2022; Velde et al., 2021). Various models and methodologies, including the Taylor diagram (Taylor 2001), Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) and Kling-Gupta efficiency coefficient (KGE) (Gupta et al., 2008), are employed to address ERC research topics. From a mathematical vantage point, an ideal approach would involve a unified or harmonized system for ERC tasks. However, to date, a universally applicable and cohesive system encompassing all three facets of ERC remains elusive.

The CCHZ-DISO system, aptly named after its four major contributors-Chen, Chen, Hu, and Zhou-emerged as a robust system for conducting comprehensive evaluations of models (Hu et al., 2019, 2023; Zhou et al., 2021). Rooted in the concept of 'DISO' denoting the distance between simulation and observation indices, CCHZ-DISO leverages the Euclidean distance, encompassing a dimensional range from 1 to infinity (Hu et al., 2019, 2023). This inclusivity allows for the integration of diverse statistical metrics. Notably, two rigorously derived weighting schemes for statistical metrics, as

https://doi.org/10.5194/gmd-2024-82
Preprint. Discussion started: 30 May 2024

established by Hu et al. (2023), contribute to the adaptability of the CCHZ-DISO system. The flexibility, expansibility, and generality exhibited by CCHZ-DISO distinguish it from alternative approaches, rendering it a versatile choice. Its widespread adoption across domains such as public health (Cui et al., 2020, 2023; Hu et al., 2020; Wang et al., 2021), geography (Ma et al., 2022; Deng et al., 2021), meteorology (Zhuang et al., 2023; Qin et al., 2022; Kalmar et al., 2021), and hydrology (Wu et al., 2023; Longo-Minnolo, et al., 2022; Yin et al., 2022) underscores its applicative prowess.

The extensive utilization of the CCHZ-DISO system has highlighted the need to elucidate various critical aspects. Among these, understanding the intricate relationship between CCHZ-DISO and chosen statistical metrics, comprehending how evaluation outcomes fluctuate with dimension variations and diverse statistical metrics, and ascertaining the significance of these outcomes are of paramount importance. Additionally, recent discoveries pertaining to CCHZ-DISO have amplified its applicability within the realm of ERC research. Notwithstanding these advancements, a comprehensive system that seamlessly unifies the tasks of evaluation, ranking, and clustering remains to be developed.

In light of the aforementioned research requirements, this study endeavors to devise a novel ERC system hinging on the CCHZ-DISO system and the Euclidean distance. Notably, the mathematical foundation for ERC tasks remains consistent, underscoring a unifying system that introduces a fresh research paradigm founded upon straightforward and widely accepted concepts.

The study unfolds in a structured manner: in the second section, a succinct introduction to the CCHZ-DISO system is presented, complemented by an impartial third-party evaluation to underscore its advantages and advancements. Following this, Subsequently, the third section sheds light on newfound insights concerning ranking and clustering, substantiated by illustrative examples showcasing their practical applications. Building upon this, the fourth section establishes and elaborates on the ERC system, portraying its structural components and functioning. In the fifth section, several important characteristics in ERC system are well illustrated for its flexible application. Three examples are provided to well illustrate the evaluating, ranking, and clustering of ERC system in the sixth section. Finally, the last section presents a short summary of the study's key points and potential prospects of the ERC system.

2 A brief introduction of CCHZ-DISO

The CCHZ-DISO system is constructed by comparing observation and simulated data from different models utilizing statistical metrics and the Euclidean distance (Hu et al., 2019, 2022; Zhou et al., 2021). The weights of statistical metrics are introduced in CCHZ-DISO, and the multiple-variable evaluations can be performed.

In our first study (Hu et al., 2019), to address the comprehensive assessment of multiple models/products/datasets, the 3-dimensional DISO system composed by the three statistical metrices of correlation coefficient (CC), absolute error (AE) and root mean square error (RMSE) are constructed with the following equation (1) (Figure 1), which have the significant advantages against the commonly employed Taylor diagram (Kalmar et al., 2021).

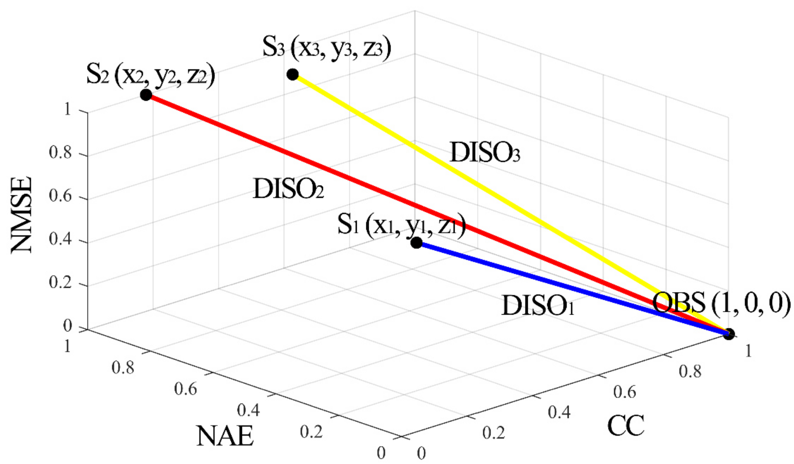

Figure 1. 3-dimensional DISO system composed by the CC, NAE (normalized absolute error) and NRMSE (normalized root mean square error) statistical metrics for three models S_1, S_2 and S_3 against OBS.

In figure 1, it shows that the three models S_1, S_2 and S_3 are comprehensively evaluated

https://doi.org/10.5194/gmd-2024-82
Preprint. Discussion started: 30 May 2024

Geoscientific Model Development Discussions
Open Access

by the 3-dimensional DISO system with the values of $DISO_1$, $DISO_2$, and $DISO_3$, respectively. $DISO_1$ has the smallest value than the DISO values of models S_2 and S_3 which show the highest overall performance among the three models based on the 3-dimensional DISO system.

$$DISO_i = \sqrt{(CC_i - CC_0)^2 + (NAE_i - NAE_0)^2 + (NRMSE_i - NRMSE_0)^2} \quad (1)$$

Since CC_0, NAE_0 and $NRMSE_0$ are the statistical metric values of the OBS against itself, they have the values of 1, 0, and 0. Then, equation (1) has the final form

$$DISO = \sqrt{(CC - 1)^2 + NAE^2 + NRMSE^2} \quad (2).$$

After adding new principles of DISO (e.g., more datasets, more statistical metrics and adding the weights), the 3-dimensional DISO system is developed to the general form with the n-dimensions (Zhou et al., 2021; Hu et al., 2022). For n-dimensional CCHZ-DISO, the number of models included is assumed as m and the number of selected statistical metrics is n. (s_i^1, s_i^2,..., s_i^n) is the statistical metric set for model *i,* where i = 0, 1,..., *m.* (s_0^1, s_0^2,..., s_0^n) is the statistical metric set for OBS against itself. The n-dimensional CCHZ-DISO equation is as follows (Hu et al., 2022; Zhou et al., 2021).

$$DISO_i = \sqrt{(nors_i^1 - nors_0^1)^2 + (nors_i^2 - nors_0^2)^2 + \cdots + \left(nors_i^n - nors_0^n\right)^2} \ . \quad (3)$$

In equation (3), $nors_i^j$ $(i = 0,1,2,\dots,m; j = 1,2,3,\dots,n)$ is the normalized statistical index, and the DISO values could be larger than 1. If the dimension n of the DISO model is normalized in equation (3), the DISO values will fall in the interval of [0, 1] based on the following equation:

$$DISO_i = \frac{1}{n}\sqrt{(nors_i^1 - nors_0^1)^2 + (nors_i^2 - nors_0^2)^2 + \cdots + \left(nors_i^n - nors_0^n\right)^2} \ . \quad (4)$$

The wide application and the third-party evaluations of CCHZ-DISO can be found in the Supporting Information.

3 Extension of CCHZ-DISO to Ranking and Clustering Applications

Following the construction process of CCHZ-DISO as outlined by Hu et al. (2019, 2022) and Zhou et al. (2021), it becomes evident that CCHZ-DISO is not limited to evaluation alone; it can effectively extend its application to ranking and clustering tasks. Indeed,

https://doi.org/10.5194/gmd-2024-82
Preprint. Discussion started: 30 May 2024

the elements within equation (3) can be broadened to encompass variables across diverse research domains, such as population, urban area, or gross domestic product of various cities, each aligned with the unique sustainable development goals (SDGs) of different countries. Incorporating variables into equation (3) transforms CCHZ-DISO into a versatile system, serving as both an evaluation tool and a robust system for ranking and clustering.

3.1 CCHZ-DISO application in ranking scenarios

In this section, we present a practical illustration of CCHZ-DISO's capabilities in ranking and clustering using data pertaining to the first two sustainable development goals (SDGs) from 15 randomly selected countries. The SDG datasets were sourced from https://sdgtransformationcenter.org/online-library. Each SDG's score is normalized within the range [0, 100], with 100 denoting the highest score and 0 the lowest. We focus on the two SDGs aligning with the 2-dimensional CCHZ-DISO approach. For ranking, we utilize a set (100, 100) as the observed data (OBS) and consider the SDG scores of the selected countries with the number of N as the simulated data (SIM). The closeness of the selected countries' OBS values directly correlates with their ranking based on CCHZ-DISO—lower distances signify a higher rank for each country within the CCHZ-DISO system. The total score is computed as follows:

$$DISO_i = \sqrt{(SDG1_i - 100)^2 + (SDG2_i - 100)^2} \quad (5)$$

where i=1, 2, …, N. According to equation (11) and the first two SDG scores, the total SDG scores are computed to rank the selected countries.

Moreover, if the smallest score of (0, 0) is defined as the OBS, then the total score is calculated as

$$DISO_i = \sqrt{(SDG1_i - 0)^2 + (SDG2_i - 0)^2} = \sqrt{SDG1_i^2 + SDG2_i^2} \quad (6)$$

where i=1, 2, …, N. In this case, a large distance of the selected countries indicates a better ranking result.

https://doi.org/10.5194/gmd-2024-82
Preprint. Discussion started: 30 May 2024

In the CCHZ-DISO system, the distance calculated through equation (3) serves as a crucial metric for determining the relative rank among various datasets. When the observed data (OBS) or the reference point scores are high, a smaller distance signifies a superior rank, highlighting a better overall performance among the datasets. Conversely, in cases where the OBS scores are low, a larger distance can indicate a better rank among the datasets, underlining the nuanced interpretation of the distance metric within the CCHZ-DISO system.

3.2 CCHZ-DISO application in clustering scenarios

The foundation of CCHZ-DISO, as expressed in equation (1), is fundamentally akin to the equation used in the K-means clustering method. K-means clustering, an unsupervised technique introduced by Lloyd (1982), serves to partition unlabeled data into 'k' distinct groups, with 'k' being a predefined number. This algorithm efficiently assigns 'm' observations to one of the 'k' clusters, each characterized by a centroid.

The K-means clustering methodology identifies common attributes among observations and leverages them to form cohesive clusters (Lloyd 1982). Generally, the Euclidean distance equation is employed to gauge the similarity between samples in this approach:

$$d(m, C_i) = \sqrt{(m_1 - C_{i1})^2 + (m_2 - C_{i2})^2 + \cdots + (m_n - C_{in})^2} \quad (7)$$

where m denotes the data points, C_i is the i-th cluster center, n is the dimensionality of the dataset, and m_j and C_{ij} are the j-th values of m and C_{ij}, respectively. The main steps of the K-means approach are as follows: (1) randomly select k cluster centers C_i $(1 \leq i \leq k)$; (2) compute the Euclidean distance from the remaining data points to each cluster center C_i to find the data with the smallest distance to each cluster center; (3) assign the data point to the corresponding cluster C_i; and (4) calculate the mean of the data points in each cluster to generate new cluster centers. The above process is iterated until stable cluster centers are obtained, or the specified maximum number of iterations is reached (Hamerly and Elkan 2003; Pham et al., 2005).

Drawing from the CCHZ-DISO equation (3), in the context of the K-means clustering

https://doi.org/10.5194/gmd-2024-82
Preprint. Discussion started: 30 May 2024

Geoscientific Model Development Discussions

methodology, the cluster center C_i corresponds to the observed data (OBS), while the data points mmm represent the simulated data originating from various models. In the K-means approach, a small distance between samples signifies a high degree of similarity. Therefore, within the CCHZ-DISO system, a small distance between simulated (SIM) and observed (OBS) data points indicate an accurate simulation model.

The K-means method, renowned for its excellent performance and straightforward conceptualization, finds extensive applications across diverse study domains, including decision support, image segmentation, data mining, and machine learning (Huang et al., 2021; Shi et al., 2021). Hence, in addition to its roles in evaluation and ranking, CCHZ-DISO extends its utility to clustering tasks, akin to the K-means method. Notably, the weighting scheme employed in CCHZ-DISO mirrors that in K-means clustering.

4 A system for evaluating, ranking, and clustering (ERC)

According to the above analysis, it can be concluded that the CCHZ-DISO system can be used to perform the ERC tasks in diverse subjects of science. Hence, the new ERC system is constructed as follows.

ERC system: For $n+1$ variables $X_i\ (i=1,2,\dots,n)$ and X_0, m quantified characteristics denoted as $x_i^j (i=1,2,\dots,n,\ j=1,2,\dots,m)$ exist. The ERC values between $X_i\ (i=1,2,\dots,n)$ and X_0 can be calculated as

$$ERC(X_i, X_0) = \sqrt{(x_i^1 - x_0^1)^2 + (x_i^2 - x_0^2)^2 + \cdots + (x_i^m - x_0^m)^2}. \quad (8)$$

Case 1: If the characteristics of $X_i\ (i=1,2,\dots,n)$ are quantified by statistical metrics and X_0 is the OBS, then ERC can be used to evaluate the performance of X_i against X_0. The ERC values in equation (8) are the CCHZ-DISO values from our previous studies (Hu et al., 2019, 2022; Zhou et al., 2021).

Case 2: If the characteristics of $X_i\ (i=1,2,\dots,n)$ are quantified and X_0 is the reference point, then ERC can be used to rank the X_i data points. When X_0 is the best point, the small ERC values in equation (8) correspond to the highest ranks. When X_0

https://doi.org/10.5194/gmd-2024-82
Preprint. Discussion started: 30 May 2024

is the worst point, large ERC values correspond to the highest ranks.

Case 3: If the characteristics of X_i $(i = 1,2, \dots, n)$ are quantified and X_0 is the cluster center, then equation (8) is the same as the basic equation in K-means clustering, and the ERC concept can be used to cluster the X_i data points.

Moreover, the weights of the quantified characteristics can also be added in equation (8) when the ERC tasks are addressed in application if necessary. Let $c_i^j = |x_i^j - x_0^j|$, then equation (9) essentially follows.

$$ERC_i = \sqrt{(c_i^1)^2 + (c_i^2)^2 + \cdots + (c_i^m)^2}. \quad (9)$$

For $c_i^j, i = 1,2, \dots, n, j = 1,2, \dots, m$, the weight w_i^j can be computed in two approaches:

$$w_i^j = \frac{c_i^j}{\sum_{i=1}^{n} c_i^j} \quad (10)$$

and

$$w_i^j = \frac{c_i^j}{\sum_{j=1}^{m} c_i^j} \quad (11).$$

5 Several important characteristics in ERC system

To have a wide and flexible application of ERC system in different scientific domains, there are several important characteristics in ERC system should be notified in its application.

5.1 Principled Selection of characteristics for the variable X in ERC system

While employing ERC system, researchers across different domains possess the autonomy to set characteristics (e.g., statistical metrics for CCHZ-DISO) based on the specific requirements of their field, underscoring the system's inherent flexibility. However, this flexibility does not imply a random selection of characteristics. Instead, two guiding principles should be adhered to when choosing characteristics: (1) when multiple characteristics exhibit a significant similarity, employing a single

https://doi.org/10.5194/gmd-2024-82
Preprint. Discussion started: 30 May 2024

Geoscientific Model Development
Discussions
Open Access

representative characteristic suffices, and (2) contradictory characteristics should not be utilized concurrently with ERC, as the CCHZ-DISO established by Hu et al. (2022).

Taking the CCHZ-DISO for instance, if the absolute error(AE) and relative error(RE) demonstrate a proportional relationship, the application of either one suffices within the CCHZ-DISO methodology. Furthermore, it is imperative to avoid combining contradictory statistical metrics within the CCHZ-DISO approach. In many research scenarios, the 3-dimensional CCHZ-DISO method, incorporating correlation coefficient (CC), absolute error (AE), and root mean square error (RMSE) metrics, offers an effective means to quantify model performance.

5.2 Each single quantified characteristic is a special form of ERC

To address this important characteristic of ERC, we take the 3-dimensional CCHZ-DISO as example. As in Section 2, there are three models (m = 3) and three statistical metrics (n = 3). CC, AE, and RMSE are the three metrics of statistics. OBS represents the observed data, and three models are defined as (S_1, S_2, S_3). Moreover, CC_i, AE_i, and $RMSE_i$, where i = 0, 1, 2, and 3, are the CC, AE and RMSE values for the OBS and results of the three models.

For the first dimension in equation (1), with CC as an example, $s_i^1 = CC_i$, and i = 0, 1, 2, and 3. Equation (1) is transformed into

$$DISO_i = \sqrt{(CC_i - 1)^2} = 1 - CC_i, \qquad (12)$$

where i = 0, 1, 2, and 3 and $DISO_0 = 0$. It is very important to note that CC_i requires a special $DISO_i$ form in equation (2). Additionally, when equation (1) is one-dimensional, each statistical metric is associated with a special form of CCHZ-DISO.

In other words, each statistical metric in this study is a special form of our CCHZ-DISO. For example, NSE and KGE are also special forms of CCHZ-DISO with 1 dimension in equation (1).

If an additional statistical metric is constructed, it can be added to CCHZ-DISO in equation (1). For example, if a new statistical metric s^* is constructed, we set $s^*=s^{n+1}$ in equation (1). Then, CCHZ-DISO takes the new form of

https://doi.org/10.5194/gmd-2024-82
Preprint. Discussion started: 30 May 2024

$$DISO_i = \sqrt{(nors_i^1 - nors_0^1)^2 + (nors_i^2 - nors_0^2)^2 + \cdots + \left(nors_i^{n+1} - nors_0^{n+1}\right)^2}\ , \quad (13)$$

If other statistical metrics are developed in the future, they can also be included in CCHZ-DISO. Moreover, each statistical metric in the future is a special form of our CCHZ-DISO with 1 dimension.

For the ranking and clustering tasks, the single quantified characteristic is the special form of ERC system with 1-dimension. In other words, the evaluating, ranking, and clustering can be addressed from single quantified characteristic for different variables.

5.3 Changes in the evaluating, ranking, and clustering results are normal

In the construction of CCHZ-DISO, the dimension and statistical metrics are determined based on the research objectives of different studies. Thus, CCHZ-DISO can have different dimensions, and different combinations of metrics can have the same dimension. For example, in this case, with three models S_1, S_2, and S_3, the observed time series is OBS, and three statistical metrics, namely, CC, AE and RMSE, are selected to construct the CCHZ-DISO model. If weighting is not considered, there will be seven possible forms of CCHZ-DISO with dimensions from 1 to 3:

1-dimensional CCHZ-DISO forms:

$$DISO_i = \sqrt{(CC_i - 1)^2} = 1 - CC_i \quad (14)$$

$$DISO_i = \sqrt{(norAE_i)^2} = |norAE_i| \quad (15)$$

$$DISO_i = \sqrt{(norRMSE_i)^2} = norRMSE_i \quad (16)$$

2-dimensional CCHZ-DISO forms:

$$DISO_i = \sqrt{(CC_i - 1)^2 + (norAE_i - 0)^2} \quad (17)$$

$$DISO_i = \sqrt{(CC_i - 1)^2 + (norRMSE_i - 0)^2} \quad (18)$$

$$DISO_i = \sqrt{(norAE_i - 0)^2 + (norRMSE_i - 0)^2} \quad (19)$$

3-dimensional CCHZ-DISO form:

$$DISO_i = \sqrt{(CC_i - 1)^2 + (norAE_i - 0)^2 + (norRMSE_i - 0)^2} \quad (20)$$

https://doi.org/10.5194/gmd-2024-82
Preprint. Discussion started: 30 May 2024

where i = 0, 1, 2, and 3 and $DISO_0 = 0$.

Different CCHZ-DISO values are calculated from equation (14) to equation (20). The evaluation results are classified into two groups: changed or unchanged.

Case 1-Changes in the evaluation results are normal: Two or more CCHZ-DISO forms in equations (14)-(20) are selected, and the evaluation results of the S_1, S_2, and S_3 models vary. For example, if equation (14) is selected, the accuracy of the three models versus OBS ranks as follows $S_1>S_2>S_3$, and if equation (15) is used, the accuracy of the three models versus OBS ranks as $S_1<S_2<S_3$, as shown in Figure 1 of Hu et al. (2022).

Case 2-Unchanged evaluation results are reasonable: When different models are employed to simulate the same OBS, there can be two types of results. The first type of result is very accurate, and the corresponding model performs better than the other models. The second type of result is inaccurate, and the corresponding model performs poorly compared to other models. Therefore, regardless of what statistical metrics are used in CCHZ-DISO, models with different performance levels may yield unchanged evaluation results, which is a reasonable scenario.

In fact, the dimension of CCHZ-DISO can decrease from n to m (smaller than n), which is essentially the coordinate projection from n-dimensional space to m-dimensional space with the scenario number of $C_n^m = \frac{n!}{m!(n-m)!}$. We consider a shift from a 3-dimensional space to a 2-dimensional surface and a 2-dimensional planar system to a 1-dimensional coordinate system to illustrate the coordinate projection process using the same data from Figure 1 of Hu et al. (2022).

Figure 2A and Figure 2B display the coordinate projection of CCHZ-DISO from 3-dimensional space to 2-dimensional surface, and they illustrate that the 3-dimensional DISO can be converted to 2-dimensional DISO. Figure 2B1, Figure 2B2 and Figure 2B3 are the results in Figure 2A projected to the XY (CC-NAE), XZ (CC-NRMSE) and YZ (NAE-NRMSE) plane coordinate systems, respectively. Figure 2C shows the results in Figure 2B projected in the X, Y, and Z directions, a DISO transformation from 2 dimensions to 1 dimension. The evaluation results vary with changes in the DISO criteria from Figure 2.

https://doi.org/10.5194/gmd-2024-82
Preprint. Discussion started: 30 May 2024

Geoscientific Model Development
Discussions
Open Access

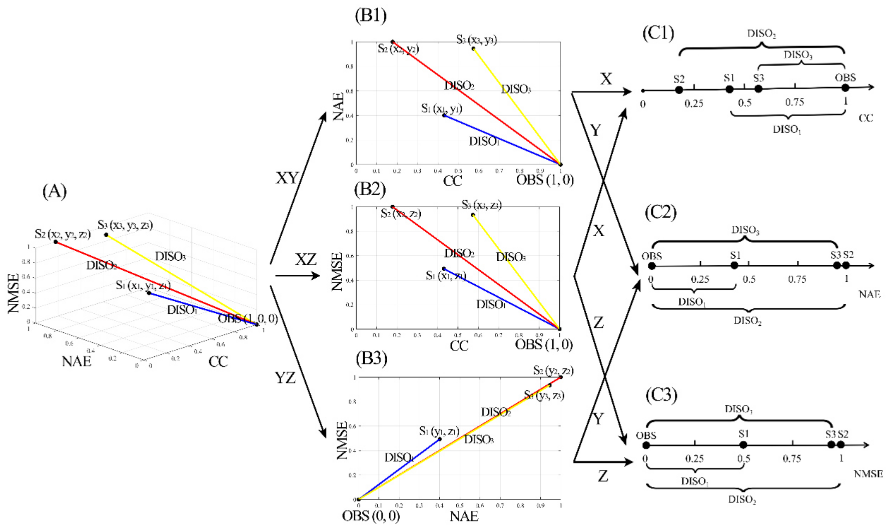

Figure 2. DISO results from a 3-dimensional space to a 2-dimensional surface and a 2-dimensional planar system to a 1-dimensional coordinate system. The dataset is same as Figure 1 in Hu et al (2022).

The above analysis suggests that these different evaluation results are related to the use of different CCHZ-DISO equations, which are set based on the research objectives. For the ranking and clustering tasks of ERC system, the results are also absolutely depended by the different characteristics and different variables selected.

5.4 Significance test of models with very small differences in ERC values

When evaluating various models, it's not uncommon for multiple models to produce similar ERC (i.e., CCHZ-DISO) values. However, to draw meaningful conclusions, it's imperative to objectively quantify the significance of the differences between or among these models. Taking the example of two models, S1 and S2, with corresponding CCHZ-DISO values of 0.2 and 0.205, a robust method is needed to determine whether the comprehensive performance of model S1 genuinely surpasses that of model S2, or if the result can be attributed to randomness.

DISO values derived from repeated sampling (e.g., utilizing moving windows) or multiple simulations can effectively address this concern. Assuming the length of the observed data (OBS) is denoted by 'n,' which is generally larger than 30 for both models S1 and S2, a 2-dimensional DISO approach employing correlation coefficient (CC) and absolute error (AE) statistical metrics is utilized. In this method, a single DISO value is calculated for each model based on the entire OBS series. However, to provide a comprehensive assessment, a moving window of size 'm' (where 'm' is less than 'n') is employed. For instance, with a moving window of 15, a total of 17 DISO values are obtained for each model (i.e., n – m+1). Subsequently, the t-test is applied to ascertain the significance of the overall performance difference between the two models.

For ranking and clustering tasks, the significance test can also be applied to test the ranking and clustering result based on the multiple times ranking and clustering.

6 Three Examples to Illustrate the Application of the ERC System

Three examples are discussed in this section to demonstrate the implementation of the ERC system in different scientific domains. The first example illustrates the evaluation of different global temperature products for geographic applications. We use the China China Merged global Surface Temperature 2.0 (CMST 2.0) as the OBS (Sun et al., 2022) to evaluate the overall performance of the 23 CMIP (Coupled Model Intercomparison Project) 6 models (https://aims2.llnl.gov/search/cmip6/). The global surface air temperature of the CMST 2.0 has the period of 1850-2022 (Sun et al., 2022). For the 23 CMIP 6 models, the history period is from 1850 to 2014, and the data of period 2015-2022 is from the SSP (shared socioeconomic pathway) 585 scenario. Three statistical metrics CC, AE and RMSE are employed to construct the 3-dimensions ERC system (i.e., CCHZ-DISO system). The second example and the third example are used to illustrate the ranking and clustering applications of ERC system with the same SDG1 (no poverty) and SDG2 (zero hunger) data from 148 countries at 2023.

Example 1. In this example, CMST 2.0 is X_0 and 23 CMIP 6 models are the $X_i, i = 1,2, \ldots, 23$. The 3-dimensions ERC system is constructed by CC, normalized AE (NAE), and normalized RMSE (NRMSE) as x_i^1, x_i^2 and $x_i^3, i = 1,2, \ldots, 23$, respectively. Then, equation (14) is:

$$ERC_i = \sqrt{(CC_i - CC_0)^2 + (NAE_i - NAE_0)^2 + (NRMSE_i - NRMSE_0)^2} \quad (18)$$

Since $CC_0 = 1, NAE_0 = 0, NRMSE_0 = 0,$ the above equation becomes

$$ERC_i = \sqrt{(CC_i - 1)^2 + (NAE_i)^2 + (NRMSE_i)^2} \quad (19)$$

After the computation, the corresponding statistical metrics: CC, NAE, NRMSE and ERC of the 23 models are obtained in Figure 2. $1 - CC_i, |NAE_i|, NRMSE_i$ are the three cases with 1-dimension ERC system. The small ERC values in 3-dimensions and 1-dimension indicate the better performance of the 23 models.

For the 3-dimensions ERC system, the ERC values (DISO values), it suggests that MPI-ESM1-2-LR has the most accurate overall performance with the smallest 3-D ERC value of 0.30 and CanESM5-CanOE has the worst overall performance with the largest 3-D ERC value of 0.73.

For the 1-D ERC system, there are three cases: $1 - CC_i, |NAE_i|$ and $NRMSE_i$. For the 1-D ERC of CC, it has the same evaluation result as the DISO, which is only used to measure the strength and direction of the linear association between the OBS and CMIP 6 models. For the other two 1-D ERC condition of NAE and RMSE, the evaluation results are similar as the DISO and CC (Figure 3).

This example is well illustrated the evaluating task of ERC system in 3-dimensions and 1-dimension. In this example, three statistical metrics CC, NAE and NRMSE are employed to construct the 3-D ERC system. Other statistical metrics can also be applied in the ERC system in 3-D or other dimensions. The weights of the statistical metrics can be added according to the equations (16) and (17). Moreover, this example is only used to illustrate the evaluating task of ERC application in earth sciences. In fact, the evaluating task of ERC application can be well used in any other research fields.

https://doi.org/10.5194/gmd-2024-82
Preprint. Discussion started: 30 May 2024

Geoscientific Model Development Discussions

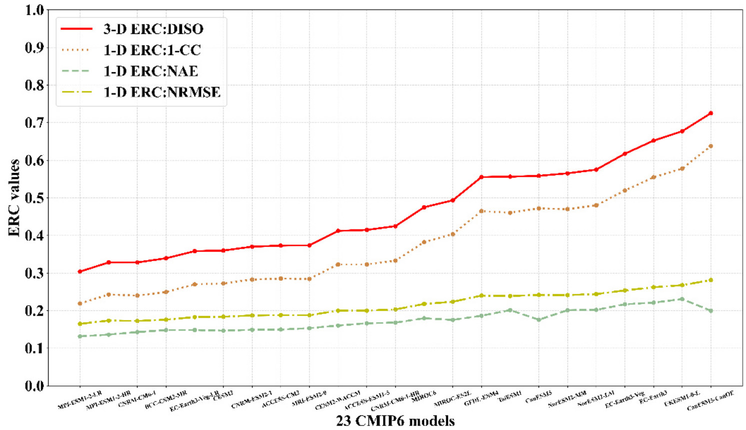

Figure 3. Evaluating result of the 23 CMIP 6 models using the ERC system based on three statistical metrics: CC, NAE and NRMSE. 3-dimensions ERC system is the DISO system in Hu et al (2019). 1-dimension ERC system includes the 1-CC, NAE and NRMSE.

Example 2. In this example, the ranking application of ERC system is illustrated. The 166 countries are ranked by 2-dimensions ERC system constructed by SDG 1 and SDG 2 as the equations (11) and (12) in flowing forms:

$$ERC_i = \sqrt{(SDG1_i - 100)^2 + (SDG2_i - 100)^2} \qquad (20)$$

and

$$ERC_i = \sqrt{SDG1_i^2 + SDG2_i^2} \quad i = 1,2,\dots,166 \qquad (21)$$

The ranking results based on equations (20) and (21) are noted as Rank 1 and Rank 2 which are displayed in Figure 4.

According to equations (20) and (21), two ranking approaches have the same ranking results for almost the 148 countries. The first 5% ranking countries of the two ranking approaches are same listing as Korea, China, Serbia, Croatia, Japan, Austria, and France (Figure 4). This example indicates the ranking task application of ERC system. Moreover, the weights also can be added for the two variables of SDG1 and SDG2 in equations (20) and (21), and more variables can be induced in ERC system for a higher dimension for more ranking objects.

https://doi.org/10.5194/gmd-2024-82
Preprint. Discussion started: 30 May 2024

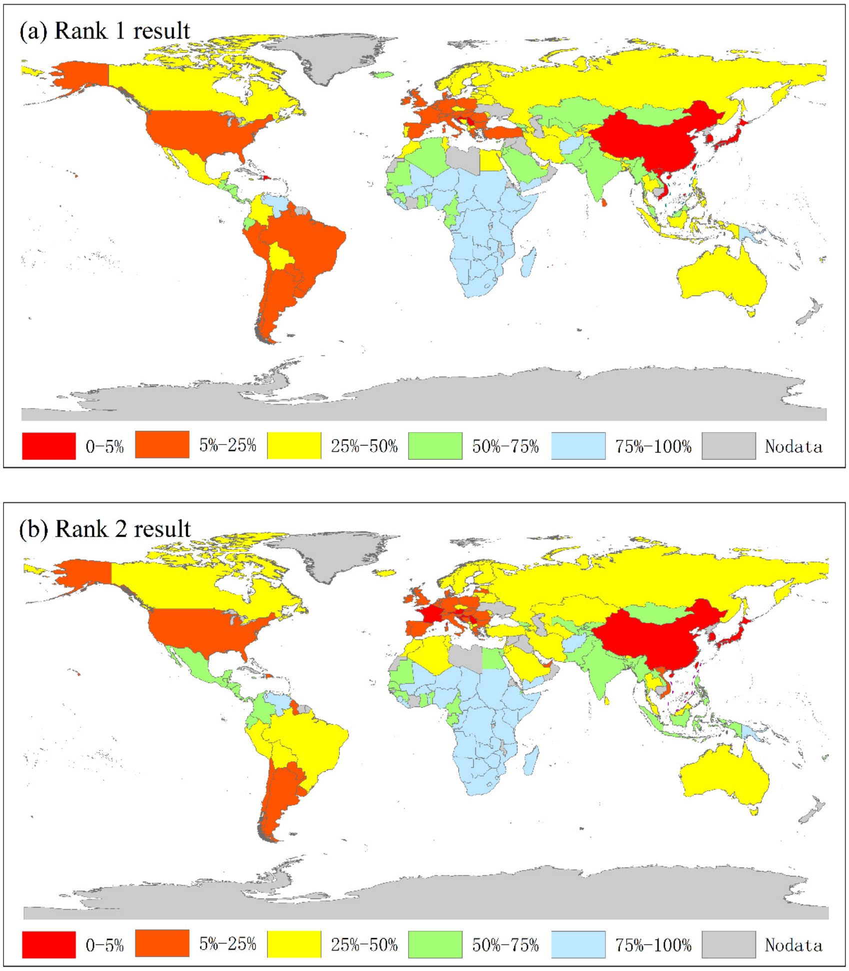

Figure 4 Ranking results of the 2-D ERC system based on the SDG1 and SDG2 data from 148 countries at 2023.

Example 3. This example is the clustering application of ERC system with the same SDG data as in Example 2. In this example, the essential of ERC system is the K-means clustering. The clustering is based on the SDG1 and SDG2 of the cluster center points, which has the following equation

$$ERC_i = \sqrt{(SDG1_i - SDG1_j^*)^2 + (SDG2_i - SDG2_j^*)^2} \quad (22)$$

The clustered data points are noted as P_i=$(SDG1_i, SDG2_i)$, ($i = 1,2,3, \ldots ,148$), and the cluster center points are noted as Q_j^*=$(SDG1_j^*, SDG2_j^*), (\,j = 1,2,3, \dots, q),\ \ q < 148$. For convenience, the ranking 20%, 40%, 60%, and 80% countries based on the equation (22) are chosen as the four cluster center points which include GUYANA (99.1295, 66.678625), JORDAN (98.6615, 5688525), MEXICO (86.266, 61.12975) and KENYA (49.1385, 58.452125). The clustering result is displaying in Figure 5.

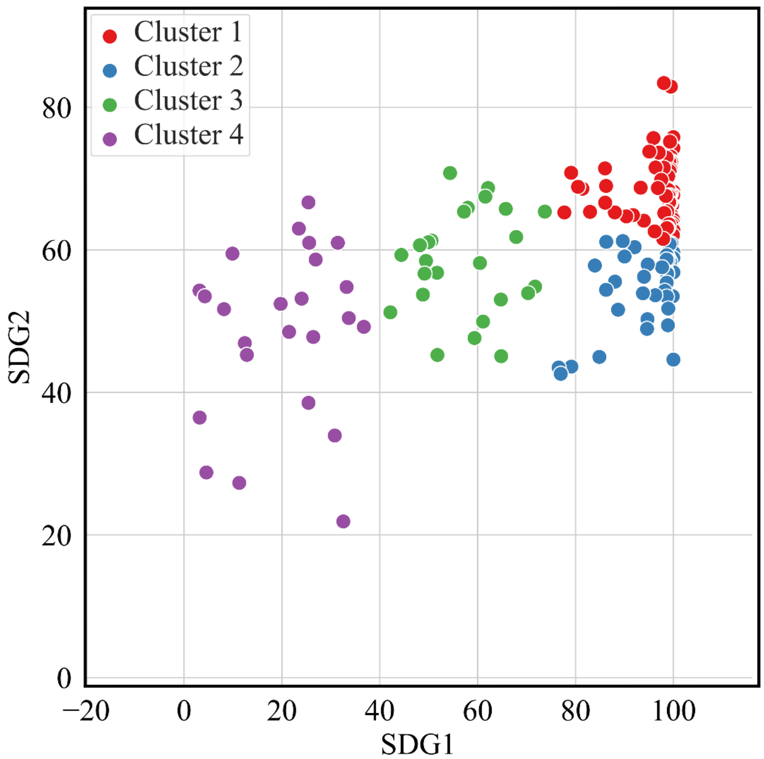

Figure 5 Clustering results of the 148 countries based on the SDG1 and SDG2 data using the equation (22).

From the above three examples, it is clearly shown that ERC system can well address the evaluating, ranking, and clustering tasks in diverse scientific domains.

7 Discussion

7.1 Theoretical Base and Application of ERC System

https://doi.org/10.5194/gmd-2024-82
Preprint. Discussion started: 30 May 2024

Generally, for any scientific subject, when some approaches (or methods, models, and systems) are employed to solve the same scientific questions, the simplest one will be chosen if it has equivalent results to the others. In other words, it follows the principle of the Chinese philosopher Lao Zi's Da Dao Zhi Jian, which means that the most basic truth is very simple (Hu et al., 2022).

Numerous approaches are proposed by large scientists to solve each task of evaluating, ranking, and clustering in their own research areas. For example, NSE, KGE, and Taylor diagram are used to evaluate the models' performance. For data clustering, there are multiple clustering techniques, such as hierarchical clustering algorithms, nearest neighbor clustering, and fuzzy clustering (Jain et al., 1999). Whether a common or general approach exists derived from the above multiple approaches to address the three tasks, the principle of constructing the common approach should be simple. In fact, evaluating, ranking, and clustering have the same essence in comparing the data characteristics. Therefore, inspired by our previous CCHZ-DISO system, the ERC system is assembled with the Euclidean distance, which is the theoretical base.

For the application of the ERC system, it should be noted that the ERC system aims to solve the evaluating, ranking, and clustering for diverse scientific domains, not for special ones. Hence, the ERC system can address the three tasks in any scientific subject. In addition, the ERC system objectively solves the evaluating, ranking, and clustering without considering the data characteristics, such as outliers. In general, before any analysis of the data is carried out, we will check correct the error data, discontinuity points, and outliers; therefore, whether the ERC results are impacted by the outliers is dependent on the users.

7. 2 Comparison between NSE, KGE, and CCHZ-DISO

In previous literature (Hu et al., 2019, 2022, Zhou et al., 2021), we comprehensively compared the CCHZ-DISO and Taylor diagram to reveal the advantages of CCHZ-DISO and disadvantages of Taylor diagram, which has been well confirmed by prominent scientists (Deng et al., 2021; Kalmar et al. 2021). There also exist other widely used statistical metrics for the calibration and evaluation of models, such as

https://doi.org/10.5194/gmd-2024-82
Preprint. Discussion started: 30 May 2024

Nash-Sutcliffe efficiency (NSE, defined by Nash and Sutcliffe, 1970) and KGE (Gupta et al., 2009). In this section, the systematic comparison between NSE, KGE, and CCHZ-DISO is addressed to display the differences and advantages of CCHZ-DISO.

For the observed time series $x_{o,t}, t = 1,2,3, \dots, n$, the simulated time series is $x_{s,t}, t = 1,2,3, \dots, n$. μ_o and μ_s are the mean values of $x_{o,t}$ and $x_{s,t}$, respectively. σ_o and σ_s are the standard deviation of $x_{o,t}$ and $x_{s,t}$, respectively, with the following forms

$\sigma_o = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_{o,t} - \mu_o)^2}$ and $\sigma_s = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_{s,t} - \mu_s)^2}$.

The mean squared error (MSE) criterion and its related normalization, the NSE, are the two criteria most widely used for the calibration and evaluation of hydrological models with observed data.

$$NSE = 1 - \frac{\sum_{t=1}^{n}(x_{s,t} - x_{o,t})^2}{\sum_{t=1}^{n}\left(x_{o,t} - \mu_o\right)^2} = 1 - \frac{MSE}{\sigma_O^2}$$

where $MSE = \frac{1}{n}\sum_{t=1}^{n}(x_{s,t} - x_{o,t})^2$ is the mean squared error (MSE). The close value of NSE to 1 indicates the high preformation of the model.

According to the decomposition of NSE, a new statistical metric named KGE is established as the following form:

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2}$$

where r is the correlation coefficient between $x_{o,t}$ and $x_{s,t}$, $\alpha = \sigma_s/\sigma_o$ and $\beta = (\mu_s - \mu_o)/\sigma_o$ (Gupta et al., 2009). The close value of KGE to 1 indicates the high preformation of the model as well as NSE.

Since CCHZ-DISO can be constructed by any statistical metrics, which also includes NSE and KGE, based on the construction principle of CCHZ-DISO and the equation (1), we establish a 2-dimensional CCHZ-DISO using NSE and KGE as

$$DISO = \sqrt{(NSE - 1)^2 + (KGE - 1)^2}.$$

The corresponding 1-dimensional CCHZ-DISO is

https://doi.org/10.5194/gmd-2024-82
Preprint. Discussion started: 30 May 2024

Geoscientific Model Development
Discussions

$$DISO = \sqrt{(NSE-1)^2}$$

or

$$DISO = \sqrt{(KGE-1)^2}$$

The above equations display that NSE and KGE are only the special forms of DISO with the 1-dimension. In fact, DISO has multiple forms determined by the different statistical metrics and dimensions. Moreover, the weights are included in DISO, while they are not considered in NSE and KGE. Therefore, it is meaningless to compare DISO, NSE, and KGE as well as other statistical metrics.

8 Conclusion

Evaluating, ranking, and clustering (ERC) tasks are pervasive across diverse scientific domains. Although these tasks may seem disparate, they share a common mathematical framework. The need to amalgamate these tasks into a unified, cohesive system based on simple and widely accepted concepts and methodologies is paramount. Nevertheless, a comprehensive system that seamlessly integrates ERC processes is conspicuously absent.

A new ERC system is proposed in this study, and the key points are summarized as follows:

(1) Clarity on CCHZ-DISO Applications: Important topics related to CCHZ-DISO applications are elucidated, including the selection principle for statistical metrics, evaluating result variations across different statistical metrics and dimensions, and conducting significance testing for models with minute differences in CCHZ-DISO values. Notably, a single statistical metric represents a specialized form of CCHZ-DISO characterized by one dimension.

(2) A Unified ERC System: We construct an ERC system grounded in a simple equation sourced from the CCHZ-DISO system. This system proves to be adaptable and versatile, making ERC tasks across diverse scientific domains accessible. Practical examples are employed to showcase its applicability to three distinct research topics.

https://doi.org/10.5194/gmd-2024-82
Preprint. Discussion started: 30 May 2024

This ERC system stands as a potent and straightforward approach for conducting ERC tasks in research. Leveraging the construction principles of the CCHZ-DISO system and the ERC system based on Euclidean distance, this approach simplifies complex problems through uncomplicated methodologies. The proposed ERC system exhibits immense potential for widespread application across the scientific landscape.

9 Code and Data Availability

The data analyzed and figures are obtained by MATLAB. The SDG data used in this manuscript is from https://www.unsdsn.org/sdg-index-and-monitoring. CMIP 6 data are from https://aims2.llnl.gov/search/cmip6/. The data and code used in this paper are permanently archived at https://zenodo.org/records/11216889 (Hu, 2024)

10 Author contribution

Study design: Zengyun Hu, Deliang Chen, Xi Chen, Qiming Zhou; Conceptualization: Zengyun Hu, Deliang Chen, Xi Chen,Qingxiang Li, and Zhuo Zhang; Visualization: Zengyun Hu, Xi Chen, Deliang Chen, Qingxiang Li, and Zhuo Zhang; Writing: Zengyun Hu; Review and editing: Zengyun Hu, Deliang Chen, Xi Chen

11 Competing interests

The authors declare no competing interests

12 Acknowledgments

This study was supported by the National Natural Science Foundation of P.R. China (Grant No. 42230708, 42361144887), the Alliance of International Science Organizations (Grant No. ANSO-CR-KP- 2021-02), the Third Xinjiang Scientific Expedition Program (Grant No.2021xjkk1300), Western Scholars of the Chinese Academy of Sciences (2020- XBQNXZ-010), and the Shenzhen science and technology innovations Committee (JCYJ20210324101406019).

References:

Cui, Q., Hu, Z., Li, Y., Han, J.,Teng, Zh., and Qian, J.: Dynamic variations of the

COVID-19 disease at different quarantine strategies in Wuhan and mainland China, Journal of Infection and Public Health.,13, 849-855, https://doi.org/10.1016/j.jiph.2020.05.014, 2020.

Cui, Q., Shi, Z. and Yimamaidi, D., Hu, B., Zhang, Zhuo., Saqib, M., Zohaib, A., Gulnara, B., Yersyn, M., Hu, Z., and Li, Sh., Dynamical variations of the COVID-19 with the SARS-CoV-2 Omicron of Kazakhstan and Pakistan, Infectious Diseases of Poverty, 12,18, https://doi.org/10.1186/s40249-023-01072-5, 2023.

Deng, M., Meng, X. Lu, Y., Li, Zh., Zhao, L., Hu, Z., Chen, H., Shang, L., Wang, Sh., and Li, Q., Impact and Sensitivity Analysis of Soil Water and Heat Transfer Parameterizations in Community Land Surface Model on the Tibetan Plateau, Journal of Advances in Modeling Earth Systems, 13, e2021MS002670, https://doi. org/10.1029/2021MS002670, 2021.

Gupta, H., Kling, H., Yilmaz, K., and Martinez, G., Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80-91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

Hamerly, G., Elkan, C., Learning the k in k-means, Advances in Neural Information Processing Systems (MIT Press), 2003.

Hu, Z., Chen, D., Chen, X., Zhou, Q., Peng, Y., Li, J. and Sang, Y., CCHZ-DISO: A Timely New Assessment System for Data Quality or Model Performance From Da Dao Zhi Jian. Geophysical Research Letters, https://doi. org/10.1029/2022GL100681, 49(23), 2022.

Hu, Z., Chen, X. Zhou, Q., Chen, D., and Li, J., DISO: A rethink of Taylor diagram. International Journal of Climatology, 39(5), 2825-2832, https://doi.org/10. 1002/joc.5972, 2019.

Hu, Z., Cui, Q. Han, J., Wang, X., Sha, W., and Teng Zh., Evaluation and prediction of the COVID-19 variations at different input population and quarantine strategies, a case study in Guangdong province, China, International Journal of Infectious Disease, 95,

https://doi.org/10.5194/gmd-2024-82
Preprint. Discussion started: 30 May 2024

231-240,https://doi.org/10.1016/j.ijid.2020.04.010, 2020.

Jacox, M., Alexander, M. Amaya, D.,Becker, E.m, Bograd, S. Brodie, S. Hazen, E., Buil, M. and Tommasi, D., Global seasonal forecasts of marine heatwaves, Nature, 604, 486-490, https://doi.org/10.1038/s41586-022-04573-9, 2022.

Jain, A., Murty, M. and Flynn, P., Data Clustering: A Review, ACM Computing Surveys, 31, 264-323. 1999.

Kalmar, T., Pieczka, H. and Pongracz, R., A sensitivity analysis of the different setups of the RegCM 4.5 model for the Carpathian region, International Journal of Climatology, 41, E1180-E1201, https://doi.org/10.1002/joc.6761, 2021.

Liu, Z., Huang, J. Xiao, X., and Tong, X., The capability of CMIP6 models on seasonal precipitation extremes over Central Asia, Atmospheric Research, 278, 106364, https://doi.org/10.1016/j.atmosres.2022.106364, 2022

Lloyd, S., Least squares quantization in PCM, IEEE Transactions on Information Theory 28, 129-137, 1982.

Longo-Minnolo, G., Vanella, D. Consoli, S., Pappalardo, S., and Ramrez-Cuesta, J., Assessing the use of ERA5-Land reanalysis and spatial interpolation methods for retrieving precipitation estimates at basin scale, Atmospheric Research, 271, 106131, https://doi.org/10.1016/j.atmosres.2022.106131, 2022

Ma, R., Xiao, J., Liang, S., Ma, H., He, T., Guo, D., Liu, X., and Lu H., Pixel-level parameter optimization of a terrestrial biosphere model for improving estimation of carbon fluxes with an efficient model-data fusion method and satellite-derived LAI and GPP data, Geoscientific Model Development, 15, 6637-6657, https://doi.org/10.5194/gmd-15-6637-2022, 2022.

Nash, J.E., Sutcliffe, J.V., River flow forecasting through. Part I. A conceptual models discussion of principles. Journal of Hydrology. 10, 282-290, 1970.

Pham, D. T., Dimov, S. S., Nguyen, C. D., Selection of K in K-means clustering,

https://doi.org/10.5194/gmd-2024-82
Preprint. Discussion started: 30 May 2024

Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 219, 103-119, DOI: 10.1243/095440605X8298, 2005.

Qin, J., Pan, W., He, M., Lu, N., Yao, L., Jiang, H. and Zhou, Ch., A 60-year (1961-2020) near-surface air temperature dataset over the glaciers of the Tibetan Plateau, Earth System Science Data, https://doi.org/10.5194/essd-2022-278, 2022.

Sun, W., Yang, Y., Chao, L. Dong, W., Huang, B., Jones, P., and Li, Q., Description of the China global Merged Surface Temperature version 2.0, Earth Syst. Sci. Data, 14, 1677-1693, https://doi.org/10.5194/essd-14-1677-2022, 2022.

Taylor, K., Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research, 106, 7183-7192, 2001.

Velde, I., Werf, G., Houweling, S., Maasakkers, J., Borsdorff, T., Landgraf, J., Tol, P., Kempen, T., Hees, R., Hoogeveen, R. Veefkind, P., and Aben L., Vast CO2 release from Australian fires in 2019 – 2020 constrained by satellite, Nature, 597, 366-369, https://doi.org/10.1038/s41586-021-03712-y, 2021.

Wang, X., Yin, G., Hu, Z., He, D., Cui, Q., Feng, X., Teng, Zh., Hu, Q., Li, J., and Zhou, Q., Dynamical variations of the Global COVID-19 Pandemic based on a SEICR disease model: a new approach of Yi Hua Jie Mu, GeoHealth, 5, 2021GH000455, https://doi.org/10.1029/2021GH000455, 2021.

Wu, F., Jiao, D., Yang, X., Cui, Zh., Zhang, H., and Wang, Y., Evaluation of NEX-GDDP-CMIP6 in simulation performance and drought capture utility over China-based on DISO, Hydrology Research, doi: 10.2166/nh.2023.140, 2023.

Yin, W., Yang, S., Hu, L.,Tian, S., Wang, X., Zhao R., and Li, P., Improving understanding of spatiotemporal water storage changes over China based on multiple datasets, Journal of Hydrology, 612, 128098, https://doi.org/10.1016/j.jhydrol.2022.128098, 2022.

Zhou, Q., Chen, D., Hu, Z., and Chen, X., Decompositions of Taylor diagram and DISO performance criteria. International Journal of Climatology, 41, 5726-5732, https://doi.org/10.1002/joc.7149, 2021.

Zhuang, S., Fang, S., Goto, D., Dong, X., Xu, Y., and Sheng, L., Model behavior regarding in- and below-cloud 137Cs wet scavenging following the Fukushima accident using 1-km-resolution meteorological field data, Science of the Total Environment, 872, 162165, https://doi.org/10.1016/j.scitotenv.2024.171558, 2023.