

Consistency Assessment and Uncertainty Analysis of Spatial-temporal Characteristics of Evaporation Data in the Greater Mekong Subregion

Shuaichen He,^{a,b} Qidong Yang,^{a,b} Liqian Zhang,^c Zilin Shi,^{a,b} Xiuzhi Wang,^{a,b} Binrong Lv,^{a,b}

^a *Department of Atmospheric Sciences, Yunnan University, Kunming, Yunnan Province, China*

^b *Yunnan Key Laboratory of Meteorological Disasters and Climate Resources in the Greater Mekong Subregion, Yunnan University, Kunming, Yunnan Province, China*

^c *Institutes of Science and Development, Chinese Academy of Science, Beijing, China*

Corresponding author: Qidong Yang, yangqd@ynu.edu.cn

ABSTRACT

To optimize the use of evaporation products in hydro-meteorological applications across the Greater Mekong Subregion (GMS), it is essential to evaluate the spatiotemporal consistency of datasets and quantify their uncertainty. This study employs the Comparison Map Profile (CMP) and Three Corner Hat (TCH) methods to evaluate four widely available evaporation datasets—ERA5-Land, FLUXCOM, GLDAS, and GLEAM—over the period 1980-2013. The CMP method evaluates the spatiotemporal consistency, while the TCH method estimates the relative uncertainty of dataset across different land cover types. The results show that the annual average evaporation across these datasets ranges from 809.47 to 920.66 mm/yr, with spatial similarity exceeding 0.76. ERA5-Land exhibits slightly lower consistency in the southern GMS. However, the datasets show limited consistency in evaporation variability; while FLUXCOM has the lowest variability and lacks a significant trend, the other datasets reveal consistent upward trends. Spatially, FLUXCOM exhibits lower similarity in the eastern GMS, whereas GLEAM shows weaker correlation with other datasets in the western GMS. The uncertainty assessment highlights notable regional differences. Across the entire GMS, ERA5-Land has the highest relative uncertainty, while GLEAM consistently maintains the lowest. Regionally, uncertainty is highest in grasslands within the northern GMS (N-GMS), while in the middle (M-GMS) and southern GMS (S-GMS), it primarily stems from rainfed cropland. Among datasets, ERA5-Land has the highest uncertainty in N-GMS, FLUXCOM in M-GMS, and GLDAS in S-GMS. By integrating the CMP and TCH results, this study explores potential causes of dataset spatial inconsistencies and the uncertainty across different land cover types in the GMS. These findings offer valuable insights for selecting and applying evaporation datasets in future hydro-meteorological research and applications in GMS.

1. Introduction

The Lancang-Mekong River (LMR) originates in the Tibetan Plateau, flows through southwestern China, Myanmar, Laos, Thailand, Cambodia, and Vietnam, and eventually empties into the South China Sea, forming the Greater Mekong Subregion (GMS). Spanning a total length of 4,880 km and covering an area of 795,000 km², the Lancang-Mekong River Basin (LMRB) is one of the world's most significant transboundary river basins (Liu et al. 2022). For transnational river basins, basin cooperation mechanisms are essential for achieving integrated governance and international collaboration, with GMS cooperation being one of the most active regional management mechanisms in the LMRB region. The GMS was established

in 1992 and includes five countries from the Association of Southeast Asian Nations (ASEAN)—Myanmar, Laos, Thailand, Cambodia, and Vietnam—as well as two provinces of China (Yunnan and Guangxi).

The LMRB is home to over 70 million people and has supported large populations with its abundant water resources since the Angkor period (approximately the 9th to 15th centuries) and even earlier. Tibetan Plateau, the source region of the LMBR, is highly sensitive to climate change (Kuang and Jiao 2016), and has had a significant impact on the hydrology of the LMBR in recent decades (Lyon et al. 2017; Hoang et al. 2016). Significantly influenced by the tropical monsoon climate, precipitation distribution within the basin is highly uneven throughout the year. About 80% of the precipitation and 75% of the annual runoff occur during the rainy season (Liu et al. 2020, 2022). Affected by distinct wet and dry climates, as well as climate change and El Niño, floods and droughts of varying degrees frequently occur in the LMRB (Liu et al. 2020). Since the 1980s, drought and flood disasters have become more frequent in the LMRB due to the interaction of multiple monsoon systems (Hundertmark 2008).

In recent decades, rapid industrialization and agricultural development in the region has significantly increased water demand, intensifying competition for water resources and heightening water security challenges (Grumbine et al. 2012). Therefore, studying evaporation in the GMS is crucial for providing valuable insights to support water resource management decisions across riparian countries and regions.

Terrestrial evaporation refers to the process by which the evaporation of water from the land surface, including plant leaf, into the atmosphere (Wang and Dickinson 2012). This process consumes more than half of the solar energy absorbed by the land surface and returns approximately 60-70% of the global annual precipitation to the atmosphere (Trenberth et al. 2009; Oki and Kanae 2006). As a critical indicator of land-atmosphere interactions, evaporation plays a crucial role in assessing global and regional climate changes (Liu et al. 2012; Wang and Dickinson 2012). Under global warming, these water cycle processes are projected to intensify (Huntington 2006), highlighting the importance of studying land surface evaporation to improve our understanding of climate change, water cycle dynamics, and ecosystem evolution.

Accurate and reliable evaporation data are essential for numerous hydro-meteorological applications, including water resource management, irrigation scheduling, crop yield estimation, and drought prediction (Liaqat and Choi 2015; Baik and Choi 2015). Although various methods are available for estimating land surface evaporation and analyzing its patterns, quantifying

evaporation remains a major challenge due to uncertainties across multiple scales. These uncertainties arise from complex climate feedback mechanisms, diverse land surface characteristics, changing environmental conditions, and their spatial and temporal Variability (Ferguson et al. 2010; Long et al. 2014). Existing various methods for evaporation estimation primarily include site observations, land surface modeling, machine learning, and remote sensing.

Among these methods, the eddy covariance technique is currently the most widely used approach for site-based observations (Yu et al. 2006). Although evaporation observations based on the flux gradient method began in the 1970s, technical limitations hindered the acquisition of reliable data until the 1990s. With continuous efforts from the scientific community, various countries and regions gradually established long-term flux observation sites, and in 1998, NASA funded the global-scale FLUXNET project to integrate and standardize these observations (Balocchi 2001). Site-based evaporation observation provide essential data for climate, hydrology, and ecological research. However, in the GMS, site observations are constrained by short observation periods and limited spatial coverage, which makes them insufficient for capturing long-term variations, such as decadal and multi-decadal climate change impacts.

Remote sensing observations offer superior spatial resolution and broader coverage compared to regional data upscaled from station observations. By utilizing meteorological data and surface conditions observed by satellites, evaporation can be estimated through various inversion methods. For instance, the SEBS model can estimate evaporation based on land surface temperature (Su et al. 2001), or from surface temperature and vegetation index using the Ts-VI method (Prince et al. 1998). However, the accuracy of remotely sensed evaporation estimates is largely influenced by the choice of inversion methods, the quality of reference data, and alternative solutions in areas lacking site observations (Tran et al. 2023). These factors introduce considerable uncertainties, making it challenging to capture long-term evaporation patterns and often resulting in notable discrepancies among different datasets due to the inherent complexity of evaporation processes.

Numerical simulation serves as a valuable complement to existing observations. Since the 1970s, Manabe (1969) introduced the “Bucket” model to demonstrate the impact of land surface evaporation on climate, highlighting its critical role in the climate system. This foundational work spurred the development of numerous land surface models(LSMs) aimed at representing land-atmosphere interactions, with the parameterization of land surface

evaporation emerging as a key component (Dickinson et al. 1986). Third-generation LSMs , such as CLM and Noah-MP, incorporate advanced evaporation parameterization schemes that account for carbon and nitrogen cycles, dynamic vegetation, and other complex land surface processes, significantly improving the accuracy of evaporation simulations. These models can provide essential tools for simulating long-term global and regional variations in land surface evaporation, contributing to a better understanding of climate dynamics (Niu et al. 2011; Oleson et al. 2010). However, differences in the parameterization schemes across models introduce substantial discrepancies in simulated evaporation estimates, posing challenges for accurately assessing evaporation patterns.

The use of machine learning methods offers a novel approach for estimating evaporation. Machine learning models are trained using meteorological and flux data from site observations and then used to estimate large-scale fluxes from remotely sensed observations. Additionally, machine learning models can utilize site-level data to upscale evaporation estimates to representative regional scales. For example, Jung et al. (2019) combined multiple machine learning techniques to obtain global land surface carbon, water, and heat flux data, demonstrating high accuracy in temperate regions and effectively capturing the spatiotemporal variations of global-scale evaporation. With the continuous growth of observational datasets, machine learning models now offer higher spatial resolution than site observations and longer time series than satellite data, significantly advancing research on the water cycle and climate change (Pan et al. 2020). However, current machine learning methods are remains constrained by the quality, quantity, and completeness of training data and the lack of model interpretability. These limitations necessitate careful selection of machine learning methods, particularly for regional-scale applications and interannual variability analysis (Jung et al. 2019).

The four aforementioned methods can provide global land surface evaporation data. However, no single evaporation product consistently outperforms the others across different land surface types and conditions (Elnashar et al. 2021). For instance, Wartenburger et al. (2018) utilized cluster analysis to examine the spatiotemporal variability of global evaporation from different datasets, finding that the choice of evaporation model or algorithm had the most significant impact on the results. Pan et al. (2020) analyzed global land surface evaporation variations using remote sensing, machine learning algorithms, and land surface process models, identifying high uncertainty areas in the Amazon Basin and arid/semiarid regions. They reported that global evaporation estimates varied from 589.6 to 617.1 mm/yr across different models, with all models consistently showing an increasing trend in global evaporation.

However, the magnitude of evaporation differed significantly, as remote sensing-based estimates were approximately three times higher than those from land surface models (Pan et al. 2020). Additionally, Kim et al. (2021) noted that the choice of evaporation datasets in East Asia requires careful consideration, as results can vary significantly depending on the dataset used. Khan et al. (2018) compared land surface evaporation data uncertainty among remote sensing-based GLEAM and MOD16 and land surface model-based GLDAS in Asia, revealing considerable differences between datasets. These findings underscore the importance of evaluating the performance and uncertainty of various evaporation products, especially in regions with complex land surface and climate conditions.

In the GMS, ground-based measurements from eddy covariance (EC) instruments at a point scale can be used to evaluate evaporation data. However, this approach is not feasible for assessing evaporation products across regional scales. The limited number and duration of EC site observations make it challenging to capture the spatiotemporal patterns of evaporation and assess the accuracy of various evaporation datasets in the GMS” for brevity and clarity. Given the constraints of current evaporation data generation methods, inconsistencies may arise between different evaporation datasets, impacting the reliability of conclusions in evaporation-related studies across countries and regions within the GMS. To address these challenges, the Comparison Map Profile (CMP) method has been introduced to measure spatial similarities across images at different scales (Gaucherel et al. 2007; Li et al. 2018), this non-point-to-point quantitative similarity method allows us to consider the similarity of data within different spatial ranges between two images at each grid point, enabling the assessment of the spatial consistency of evaporation datasets. Meanwhile, the Three Corner Hat (TCH) method, a relative uncertainty estimation approach, has been widely applied to evaluate the uncertainty of various geophysical datasets, such as precipitation and soil moisture, particularly in the absence of ground-based observations (Awange et al. 2016; Miralles et al. 2010). TCH assumed that each dataset consists of the true value plus an error. The relative uncertainty is then estimated by iteratively calculating the variance of the error between each dataset and the assumed true value based on the covariance of the differences between datasets. In the absence of large-scale reference data, this study employs the two aforementioned methods to evaluate four GMS evaporation datasets from 1980 to 2013. The CMP method is used to assess the spatial similarity of mean values, variability, and trends among the datasets, while the TCH method estimates the error magnitude between each dataset and the assumed true value. Furthermore, the TCH method provides error estimates for the major land cover types in the northern, middle, and

southern regions of the GMS. The study will provide insights and data selection references for evaporation estimation, water resource management, and hydro-meteorological applications across different regions of the GMS.

2. Data and Methods

a. Study area and its main land cover types

The GMS (2°N - 32°N , 90°E - 112°E) and its land cover types are shown in Fig. 1. The GMS features four main vegetation types (Fig. S2): evergreen broadleaf forest, evergreen needleleaf forest, cropland, and grassland. To facilitate detailed analysis, we divided the region into three sections—North GMS (N-GMS), Middle GMS (M-GMS), and South GMS (S-GMS)—with complete details available in the supplementary material. The N-GMS is dominated by grassland, the M-GMS is characterized by evergreen broadleaf forest, and cropland prevails in the S-GMS. For a comprehensive overview of the major land cover types in each region, refer to Fig. S2.

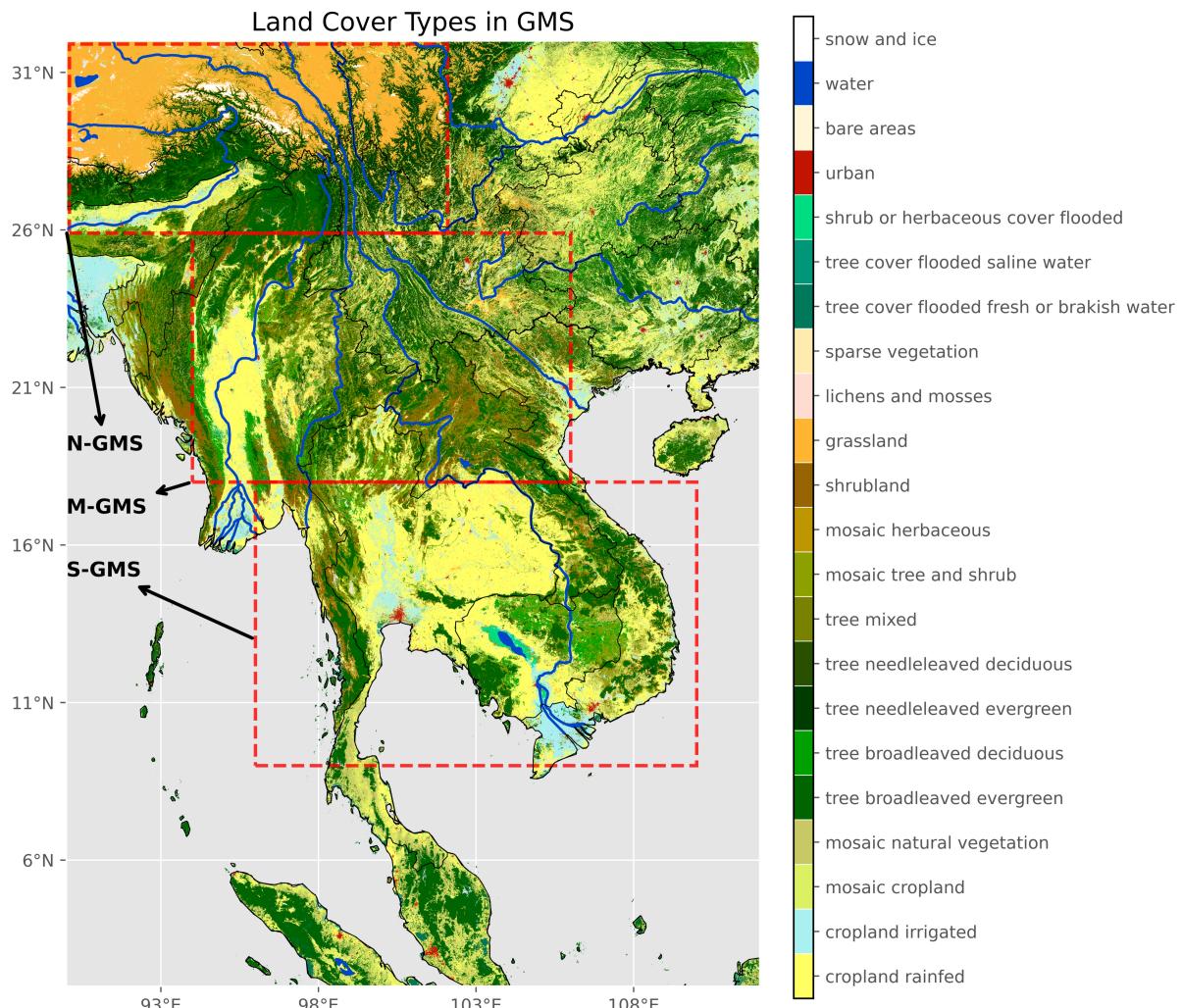


Fig. 1. GMS Land Cover Type(2013). Data from Copernicus Climate Change Service (C3S), Climate Data Store (CDS), (2019): Land cover classification gridded maps from 1992 to present derived from satellite observation. The three dashed box regions are divided from north to south into the northern GMS (N-GMS), middle GMS (M-GMS), and southern GMS (S-GMS), respectively.

b. Datasets

This study selected monthly average data of land surface evaporation from four different sources in the GMS from 1980 to 2013. The sources include ERA5-Land (ECMWF Re-Analysis 5), FLUXCOM, GLDAS (Global Land Data Assimilation System), and GLEAM (Global Land Evaporation Amsterdam Model). All datasets have a temporal resolution of one month, with additional information provided in Table 1. The four sets of land surface evaporation data were interpolated bilinearly to a resolution of $0.5^\circ \times 0.5^\circ$.

ERA5-Land data is the land component of the fifth generation of European ReAnalysis (ERA5), produced by the European Centre for Medium-Range Weather Forecasts (ECMWF).

It is derived from the ECMWF global high-resolution land surface model using meteorological data-driven models obtained by downscaling ERA5 climate reanalysis data (Muñoz-Sabater et al. 2021). FLUXCOM employs machine learning methods to merge flux observations from the FLUXNET eddy covariance towers with remote sensing and meteorological data to estimate global carbon, water, and energy fluxes (Jung et al. 2019). The FLUXCOM latent heat flux data are estimated based on different energy closure corrections, machine learning methods, and climate forcing data. In this study, we use the ensemble data from all energy closure corrections and machine learning methods. To obtain a sufficiently long time series, we chose the set trained on the WFDEI climate forcing data. The Global Land Data Assimilation System (GLDAS) is a high-resolution, offline terrestrial modeling system that incorporates satellite and ground-based observations to produce optimal fields of land surface states and fluxes. GLDAS utilizes ground and space-based observation systems that provide data to constrain the modeled land surface states (Rodell et al. 2004). The GLDAS monthly data used in this study are v2.1, based on the Noah model. The basis for evaporation estimation in the Noah model is the Penman equation (Chen et al. 1996; Koren et al. 1999). The Global Land Evaporation Amsterdam Model (GLEAM) is a set of algorithms designed to estimate terrestrial evaporation from satellite data. Evaporation resistance factors are calculated from observed vegetation optical thickness, soil moisture, and multilayer infiltration models, while canopy interception evaporation is computed using the Gash analytical model in conjunction with precipitation observations. Finally, snow and ice evaporation is calculated using the Priestley-Taylor equation for supercooled water conditions (Miralles et al. 2011; Martens et al. 2017). Considering the time span of the study, GLEAM data with potential evaporation constraints (v3.7a), calculated based on reanalysis information with a longer time series, will be used in this study.

Due to the lack of large-scale evaporation reference data in the GMS, we compared the four evaporation datasets with the only two publicly available station-based evaporation datasets in Section C of the Supplementary Materials (Fig. S3, S4). The evaluation, based on regression coefficients, correlation coefficients, and root mean square error, shows that all four datasets demonstrate good absolute accuracy at these two stations. Furthermore, to address the challenge of limited large-scale evaporation data in the GMS, this study employs the CMP and TCH methods to compare the four widely used datasets. By analyzing spatial distribution similarity and relative uncertainty, this study provides insights into evaporation patterns and serves as a reference for understanding dataset differences in GMS-related research.

Land cover data used in this study were obtained from the Copernicus Climate Change Service, Climate Data Store (2019), which provides land cover classification gridded maps derived from satellite observations from 1992 to present (Copernicus Climate Change Service (C3S) Climate Data Store (CDS). DOI: [10.24381/cds.006f2c9a](https://doi.org/10.24381/cds.006f2c9a)). This dataset, updated yearly since 1992, offers global maps that categorize land surfaces into 22 classes based on the United Nations Food and Agriculture Organization's (UN FAO) Land Cover Classification System (LCCS). For our analysis, we used the 2013 land cover data with a 300m spatial resolution. To assess relative uncertainty across different land cover types, we first interpolated the land cover data to a global 0.5-degree resolution. For each 0.5-degree grid cell, we selected the most frequent land cover type from all 300m pixels as the representative type (Fig. S1). We defined major land cover types based on their proportional coverage and calculated their distributions (Fig. S2). To study relative uncertainty in this research, we selected the top four major land cover types in each region for analysis. In the GMS and N-GMS, the primary land cover types consist of grassland, rainfed cropland, evergreen broadleaf forest, and needleleaf forest. In M-GMS, the main land cover types include evergreen broadleaf forest, rainfed cropland, mosaic tree and shrub, and mosaic natural vegetation. In S-GMS, the predominant land cover types are two varieties of cropland (rainfed and irrigated), evergreen broadleaf forest, and mosaic tree and shrub (Fig. S2).

Data sets	Spatial resolution	Temporal resolution	Citation
ERA5-Land	$0.1^\circ \times 0.1^\circ$	1980 to 2013.	(Muñoz-Sabater et al. 2021)
FLUXCOM	$0.5^\circ \times 0.5^\circ$	Monthly data.	(Jung et al. 2019)
GLDAS Noah v2.0	$0.25^\circ \times 0.25^\circ$	Total 408 month per dataset.	(Rodell et al. 2004)
GLEAM 3.7a	$0.25^\circ \times 0.25^\circ$		(Martens et al. 2017)

Table 1. Evaporation data sets information.

c. Comparison Map Profile Method

The Comparison Map Profile (CMP) method, proposed by Gaucherel et al. (2007) quantifies the magnitude and distribution of differences between image data. The cross-correlation coefficient (CC) is used to measure the similarity of spatial features at each grid point of the image(Gaucherel et al. 2007). Using a sliding window approach, data from the

region around each grid point with an aperture size R are sequentially acquired, and the CC is calculated for the sequence between two images. The aperture size is then varied to obtain cross-correlation coefficient images at different apertures. For each grid point in the original image, spatial similarities at different aperture sizes are considered. The average of these similarities across all apertures is used to comprehensively assess the spatial distribution similarity between the two images at each grid point and its surrounding area.

The aperture size R refers to the number of grid points from the image grid point to the window boundary. The relationship between the number of grid points N in the sliding window and the aperture size R is given by the formula: $N = (2 \times R + 1)^2$. When $R = 1$, the CC is calculated using the current grid point as the center, along with the eight surrounding grid points. Similarly, when $R = 15$, the CC is calculated considering the 960 surrounding grid points.

The window traverses each grid point in the image, calculating the CC for two data sets at each grid point, considering the spatial similarity of the surrounding region within the defined aperture. The larger the value of the spatial similarity, the higher the similarity in the spatial distribution of the two datasets.

$$CC = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \frac{(x_{ij} - \bar{x})(y_{ij} - \bar{y})}{\sigma_x \times \sigma_y} \quad (13)$$

$$\sigma_x = \sqrt{\frac{1}{N^2 - 1} \sum_{i=1}^N \sum_{j=1}^N (x_{ij} - \bar{x})^2} \quad (14)$$

$$\sigma_y = \sqrt{\frac{1}{N^2 - 1} \sum_{i=1}^N \sum_{j=1}^N (y_{ij} - \bar{y})^2} \quad (15)$$

Where N represents the number of grid points in each sliding window. x_{ij} , y_{ij} represent the grid values in the i-th row and j-th column of the sliding window for two sets of comparative data images. \bar{x} , \bar{y} represent the mean values of all grid points in the sliding window for the two sets of comparative data. Standard deviation of the sliding window data at σ_x and σ_y respectively.

In this study, sliding windows were calculated from 3×3 ($R=1$) to 31×31 ($R=15$) to obtain similarity indices. The CC value decreases to small values when the window size reaches 31×31 ($R=15$), and in our study area (62×46), no additional information can be captured with larger window sizes. The images at $R=15$ were averaged to obtain the multiscale average value. A high cross-correlation coefficient between the two sets of data indicates a high similarity between them.

d. Three-Cornered Hat Method

The TCH method, proposed by Tavella and Premoli in 1993 (Premoli and Tavella 1993), allows for comparing the relative uncertainty of three or more time series without requiring a real reference field (Tavella and Premoli 1994).

The detailed methodology is as follows:

Assume that the time series to be compared is represented by:

$$X_i = X_{true} + \varepsilon_i, \quad i = 1, 2, \dots, N \quad (1)$$

In the equation, N represents the number of evaporation datasets, which in this paper is 4, X_i represents i-th time series, X_{true} represents the true value, and ε_i is the error between the i-th dataset and the true value.

To begin, any dataset is selected as a reference sequence X_R , and no matter which dataset was used as a reference, its uncertainties is not affected. The difference between the other time series and the reference sequence is denoted as y_i .

$$y_i = X_i - X_R = \varepsilon_i - \varepsilon_R, \quad i = 1, 2, \dots, N - 1 \quad (2)$$

X_R is the reference sequence, and ε_R represents the error between the reference dataset and the true value. M denotes the length of time series, which in this study corresponds to 408, given that monthly data is used. All y_i can be combined into a $M \times (N-1)$ matrix called Y .

$$Y = \begin{bmatrix} y_{11} & \cdots & y_{1(N-1)} \\ \vdots & \ddots & \vdots \\ y_{M1} & \cdots & y_{M(N-1)} \end{bmatrix} \quad (3)$$

Y is represented by S as the covariance matrix. S is the covariance matrix, $cov(\bullet)$ is the covariance operator, $s_{ij} = s_{ji}$ ($i, j = 1, 2, \dots, N-1$) is the variance ($i = j$) or covariance ($i \neq j$) between y_i and y_j .

$$S = cov(Y) = \begin{bmatrix} s_{11} & \cdots & s_{1(N-1)} \\ \vdots & \ddots & \vdots \\ s_{(N-1)1} & \cdots & s_{(N-1)(N-1)} \end{bmatrix} \quad (4)$$

Introducing an $N \times N$ covariance matrix Q ($q_{ij} = cov(\varepsilon_i, \varepsilon_j)$). J is an $N-1 \times N$ matrix. S can be decomposed as follows:

$$\mathbf{Q}_{N, N} = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1(N-1)} & q_{1N} \\ q_{21} & q_{22} & \cdots & q_{2(N-1)} & q_{2N} \\ \vdots & \ddots & & \vdots & \vdots \\ q_{N1} & \cdots & & q_{N(N-1)} & q_{NN} \end{bmatrix} \quad (5)$$

$$\mathbf{J}_{N-1, N} = \begin{bmatrix} 1 & 0 & \cdots & 0 & -1 \\ 0 & 1 & \cdots & 0 & -1 \\ \vdots & \ddots & & \vdots & \vdots \\ 0 & \cdots & 1 & -1 \end{bmatrix} \quad (6)$$

$$S = J \cdot Q \cdot J^T = J \cdot \begin{bmatrix} \hat{Q} & \hat{\mathbf{q}} \\ \hat{\mathbf{q}}^T & q_{NN} \end{bmatrix} \cdot J^T \quad (7)$$

Since the equations are underdetermined with $N \times (N+1)/2$ unknowns but only $(N-1) \times N/2$ equations (where \mathbf{Q} is a symmetric matrix), these unknowns cannot be solved directly based on equation(7). The values $\hat{\mathbf{q}}$ and q_{NN} are necessary to solve this underdetermined system. Once the N free parameters have been estimated, the solution for the unknown elements $\hat{\mathbf{q}}$ is given by:

$$\hat{\mathbf{q}} = \mathbf{S} - \mathbf{q}_{NN} \mathbf{U} \mathbf{U}^T + \mathbf{U} \hat{\mathbf{q}}^T + \hat{\mathbf{q}} \mathbf{U}^T \quad (8)$$

The N free parameters (q_{1N}, \dots, q_{NN}) can be defining an objective function \mathbf{F} based on the Kuhn-Tucker theorem with specific constraints (Galindo and Palacio 1999):

$$\mathbf{F}(q_{1N}, \dots, q_{NN}) = \frac{1}{K^2} \cdot \sum_{i < j}^N q_{ij}^2 \quad (9)$$

where $K = \sqrt[N-1]{\det(S)}$ and \det computes the determinant value of \mathbf{S} . Since the covariance matrix \mathbf{Q} is a symmetric matrix and positive *definite* if and only if $\det(\mathbf{Q}) > 0$, as demonstrated by Galindo and Palacio (1999), the Kuhn-Tucker theorem indicates that the utilization of a constraint function should be appropriately modified to enable the application of the Kuhn-Yucker theorem. Consequently, a constraint \mathbf{H} for solving \mathbf{F} is adopted:

$$\mathbf{H}_{q_{1N}, \dots, q_{NN}} = -\frac{\det(Q)}{\det(S) \cdot K} < 0 \quad (10)$$

The initial conditions of iteration fulfill the constraints:

$$q_{iN}^{(0)} = 0, \quad i < N \text{ and } q_{NN}^{(0)} = (2 \cdot \mathbf{U}^T \cdot \mathbf{S}^{-1} \cdot \mathbf{U})^{-1} \quad (11)$$

By minimizing the objective function in equation (9), the N free parameters (where $N = 4$) are obtained. The i -th diagonal element of \mathbf{Q} , denoted as q_{ii} , represents the error variances of the observations.

To make the TCH uncertainties comparable, the relative uncertainty (RU) is introduced, which is defined as the ratio between the square root of the error variance ($\sqrt{q_{ii}}$) and the mean evaporation value (\bar{X}_i) of the i-th time series:

$$RU = \frac{\sqrt{q_{ii}}}{\bar{X}_i}, \quad i = 1, 2, \dots, N \quad (12)$$

In this study, each grid point of the GMS data contains a 34-year time series. The relative uncertainty is calculated for the same grid point across all four datasets. For each dataset, the relative uncertainty is then averaged over all grid points.

3. Results

a. Analysis of the multiyear mean of land surface evaporation

Fig. 2 shows that annual evaporation ranges from 809.47 mm to 920.66 mm, with ERA5-Land having the highest value, followed by FLUXCOM, while GLEAM and GLDAS exhibit similar values. FLUXCOM has the smallest standard deviation, while GLDAS has the largest.

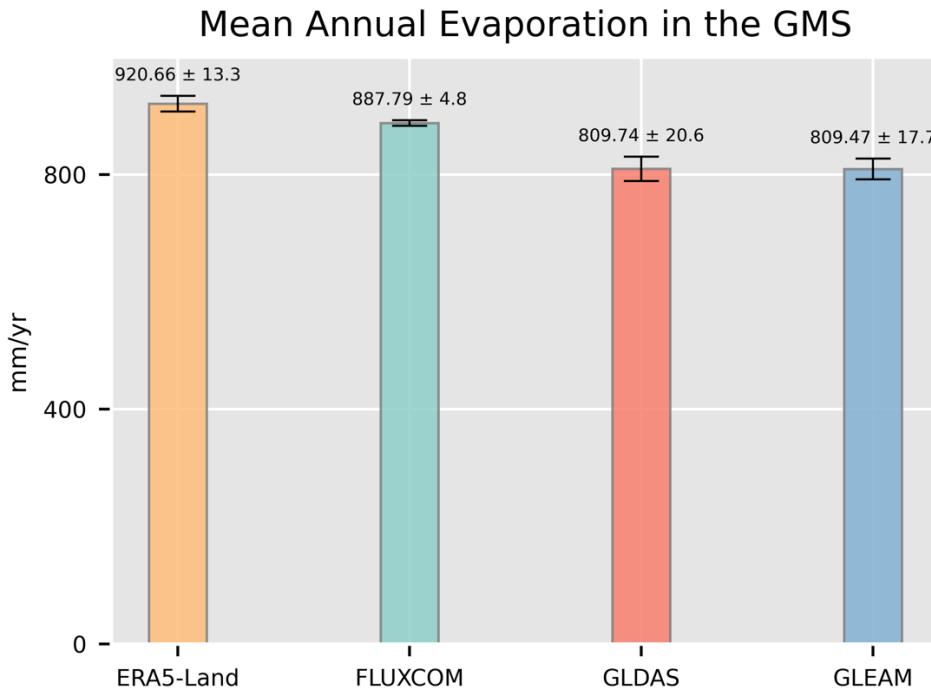


Fig. 2. Annual mean of four sets of land surface evaporation data from 1980 to 2013. The numbers above the bars represent the mean and one standard deviation. The error bars indicate one standard deviation of the annual mean evaporation values.

Fig. 3 illustrates the spatial distribution of the mean values of the four sets of evaporation data, the CC values between the two datasets, and the corresponding scatterplot for each grid point. The four evaporation datasets in the GMS region exhibit a consistent decreasing pattern from southeast to northwest, with high evaporation values primarily observed in the southeastern regions of Vietnam, Laos, and Cambodia. However, as shown in Fig. 3(b, c, and d), the ERA5 data differs noticeably from the other datasets in these high-evaporation areas. In terms of the similarity in the spatial distribution of mean evaporation values, the primary differences between the datasets are observed in the southern peninsula region (Fig. 3b, c and d), the northeastern part of the GMS and the southern part of Vietnam. (Fig. 3d, h and l), and the northwestern part of Thailand (Fig. 3c, d, g and h). The results from the lower triangular scatterplots indicate that the linear proportional relationship between grid points in the four evaporation datasets is close to 1, with correlation coefficients exceeding 0.78. When considering spatial similarity, the correlation coefficients decrease but remain within the range of 0.86 to 0.92. Among them, GLDAS and GLEAM exhibit the strongest linear proportional relationship(Fig. 3o), while ERA5-Land and FLUXCOM have the highest correlation coefficient(Fig. 3e). When considering spatial distribution similarity, GLDAS and ERA5-Land show the highest average similarity. However, for specific regions, it is essential to refer to the spatial similarity distribution in the upper triangular plots for a more detailed assessment. For example, while GLDAS and ERA5-Land exhibit the highest multi-scale average similarity, Fig. 3c shows that this pattern is primarily concentrated in the northern GMS. In contrast, the similarity gradually decreases in the southern region.

Overall, the four datasets exhibit strong consistency in both average evaporation values and spatial distribution patterns, except in the regions of northern Thailand and Laos within the M-GMS and S-GMS areas of the GMS.

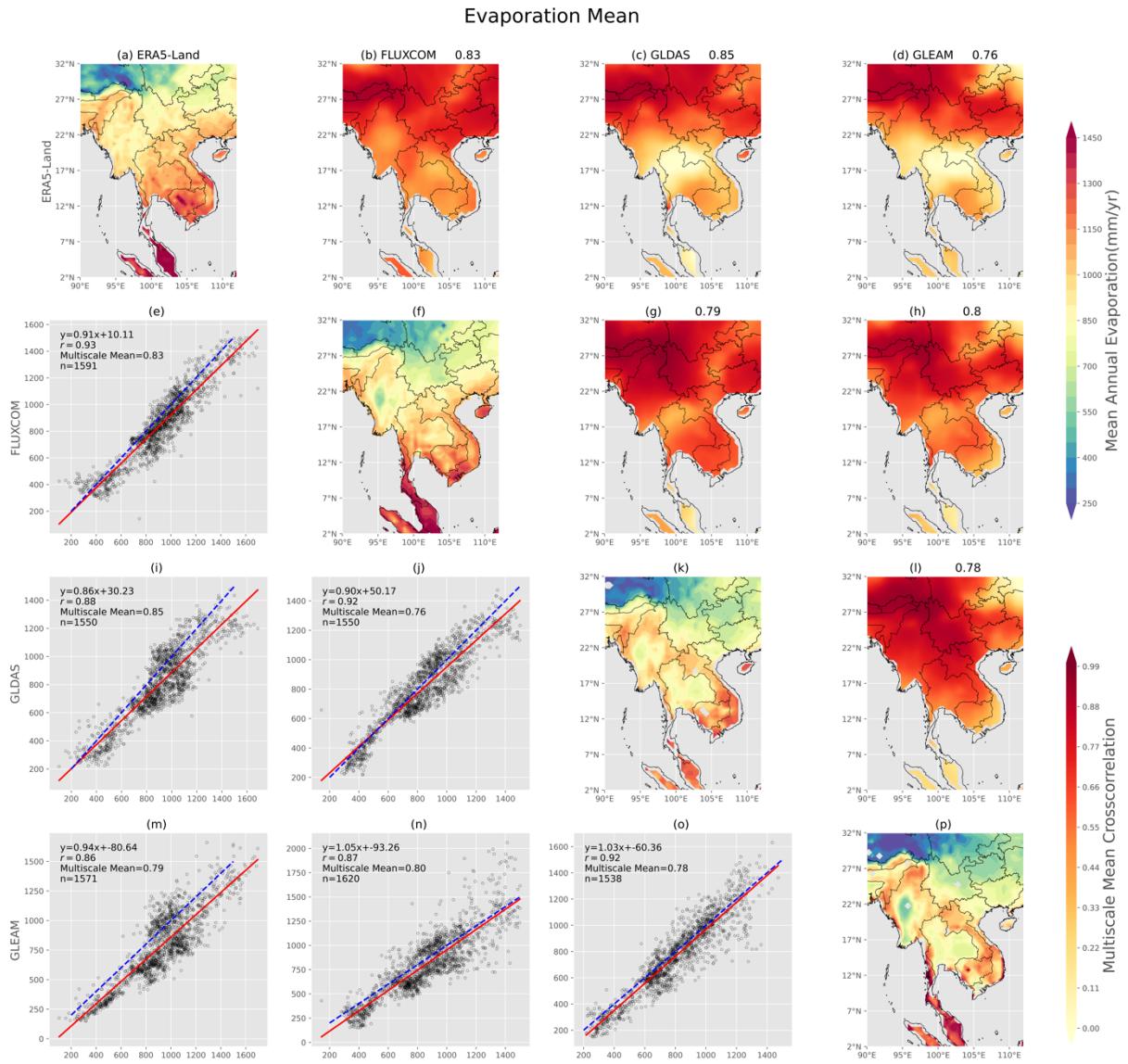


Fig. 3. Spatial distribution of the annual mean values for four sets of land surface evaporation datasets and the multiscale mean CC values calculated using the CMP method. The diagonal plots represent the spatial distributions of the mean values for each of the four datasets. The upper triangular matrix plots display the spatial distribution of the CC values calculated using the CMP method, with the numbers on the right side of each title indicating the multiscale mean CC values between dataset pairs, illustrating the similarity in the spatial distribution of the evaporation datasets. The lower triangular matrix plots are scatter plots of the corresponding grid points in two-by-two comparisons between the datasets, with the red line representing a fitted line of the scatter points and the 1:1 relationship indicated by the blue dashed line. Each scatterplot subplot is annotated with the univariate linear regression coefficient, correlation coefficient, multiscale mean CC, and the number of grid points.

b. Analysis of the interannual variability of land surface evaporation

Fig. 4 illustrates the spatial distribution of interannual variability across four land surface evaporation datasets, along with scatter plots of the corresponding grid points and the spatial distribution of variability CC values. The scatter plots reveal notable differences in the interannual variability of evaporation among the datasets. FLUXCOM data exhibit low and uniformly distributed interannual variability across the GMS, as shown in the Fig. 4f. In contrast, ERA5-Land displays higher variability in the northwestern and southern regions of the GMS(Fig. 4a). GLDAS shows greater variability in the eastern GMS, as well as in central Myanmar and Thailand(Fig. 4k). GLEAM data highlight significant variability primarily in central Myanmar, western Thailand, and southern Vietnam(Fig. 4p). High evaporation variability is predominantly observed in central Myanmar, western Thailand, and southern Vietnam. As shown in the upper triangular area of Fig. 4, we focus on three datasets—ERA5-Land, GLDAS, and GLEAM—that offer more insightful data on evaporation variability in the GMS (Fig. 4c, d, and l), excluding FLUXCOM due to its low variability. The spatial distribution of evaporation variability differs among these datasets, particularly in the northeastern parts of the GMS(Fig. 4b, g and h), M-GMS(Fig. 4g), and S-GMS(Fig. 4c), as well as in central Yunnan, China(Fig. 4c and l), Myanmar and Thailand(Fig. 4g). Across the GMS, the four datasets reveal discrepancies in interannual variability. Notably, although the correlation coefficients for variability are generally low, GLEAM exhibits the highest correlation with the other datasets(Fig. 4m, n and o). Additionally, GLDAS maintains its correlation even after accounting for multi-scale spatial similarity(Fig. 4i, j and n).

Evaporation Interannual Variability

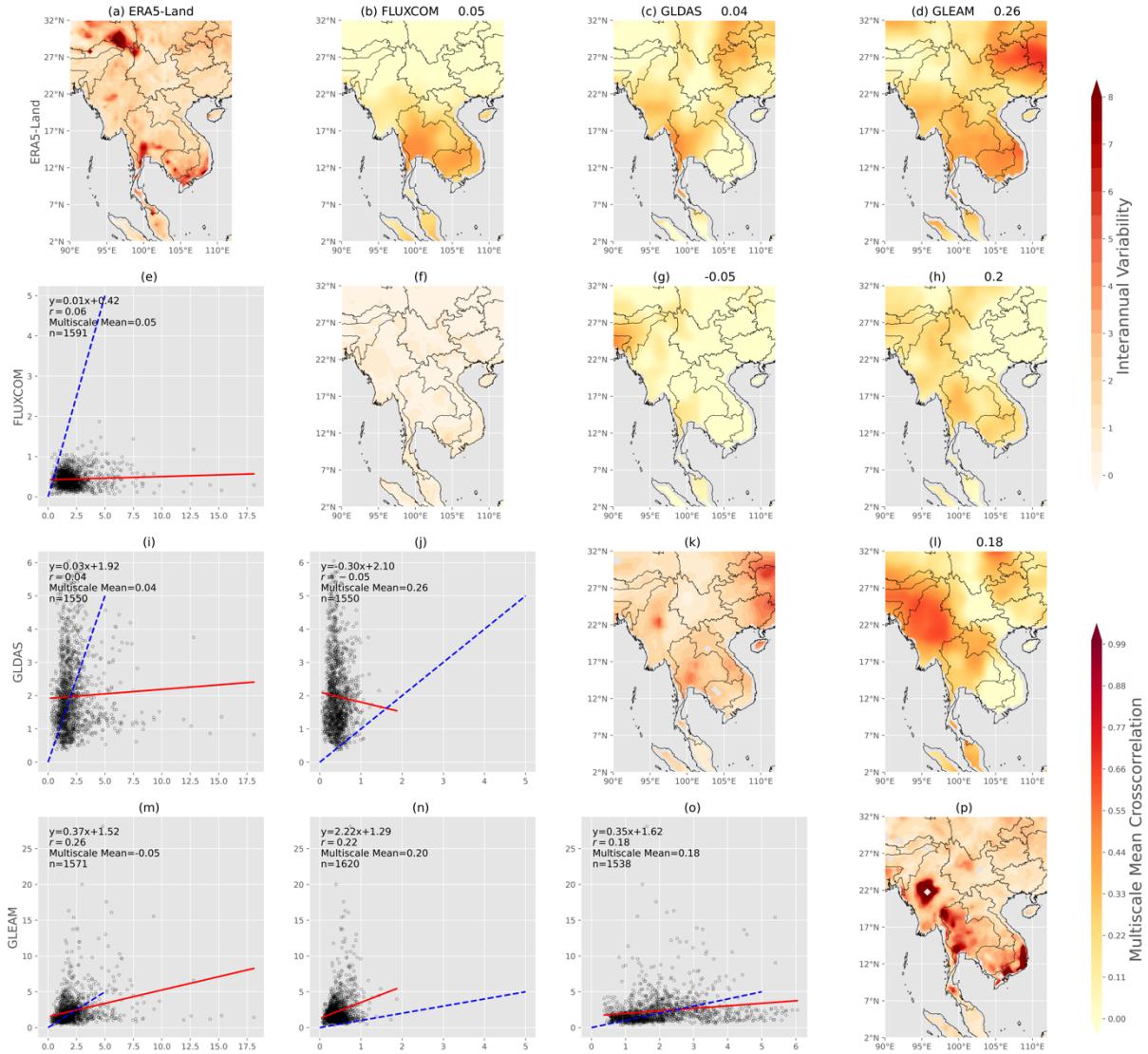


Fig. 4. Spatial distribution of the interannual variability for the four evaporation datasets, along with scatter plots of the corresponding grid points and the spatial distribution of the CC values of the variability between dataset pairs. The diagonal plots represent the spatial distributions of the interannual variability for each dataset. The upper triangular matrix plots show the spatial distribution of CC values obtained using the CMP method, with the numbers on the right side of each title indicating the multiscale mean CC values between dataset pairs, reflecting the similarity in the spatial distribution of evaporation. The lower triangular matrix contains scatter plots comparing the grid points of two datasets. The red line represents the fitted line of the scatter points, and the 1:1 line is shown as a blue dashed line. Each scatterplot subplot is annotated with the univariate linear regression coefficient, correlation coefficient, multi-scale mean CC, and the number of grid points.

c. Analysis of the long-term trend of land surface evaporation

Fig. 5 shows the evolution of the annual mean of the four sets of evaporation data from 1980 to 2013. The Mann-Kendall (MK) non-parametric test, commonly used to assess trend significance in climate and environmental studies (Machiwal and Jha, 2006), was applied to the annual means of the four datasets. The MK test was conducted on 34 years of monthly mean evaporation values, with each dataset containing a total of 408 data points. The results reveal a significant increasing trend in evaporation across three datasets: ERA5-Land (0.645 mm/yr^2 , p-value: 0.0048), GLDAS (1.421 mm/yr^2 , p-value: 3.313×10^{-5}) and GLEAM (1.261 mm/yr^2 , p-value: 1.715×10^{-5}), while the FLUXCOM dataset shows a non-significant decrease (-0.103 mm/yr^2 , p-value: 0.097). Among these, GLDAS exhibits the highest trend value, and notably, GLDAS and GLEAM display consistent evaporation trends.

Fig. 6 illustrates the calculated trends for the four sets of evaporation data on a grid-point-by-grid-point basis, along with scatter plots of the corresponding grid points and the spatial distribution of the trend CC. The evaporation trends for the four datasets consistently increase in the southeastern part of the GMS. Except for FLUXCOM, the remaining three datasets show a significant increase in the eastern part of the GMS. In terms of the magnitude of trend changes, there is less consistency between the datasets across the GMS. Regarding the spatial characteristics of trend changes, ERA5-Land and GLDAS exhibit high spatial similarity, with a multi-scale average CC value of 0.57 (Fig. 6i). The inconsistency between the ERA5-Land data and the other datasets is primarily observed in the central and western parts of the GMS, while FLUXCOM shows greater inconsistency in the eastern part of the GMS (Fig. 6b, g and h).. Additionally, the inconsistency between GLDAS and GLEAM is mainly found in the western part of the GMS (Fig. 6i).

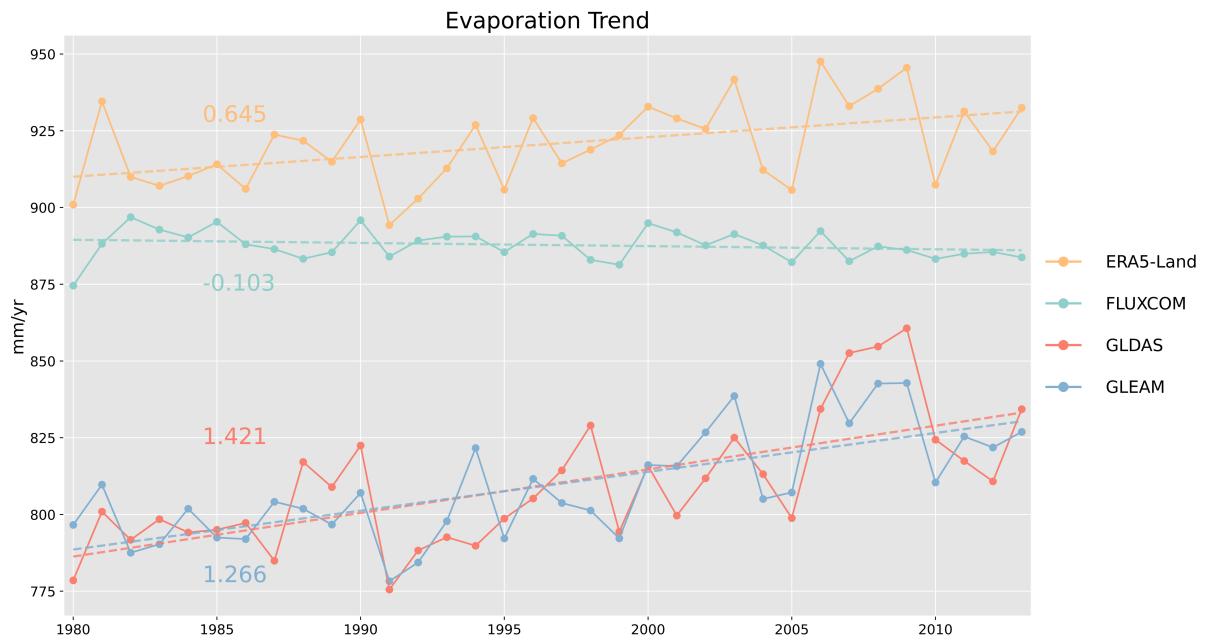


Fig. 5. Changes in the annual average values of evaporation for the four datasets from 1980 to 2013. Dots represent the mean annual evaporation, and dashed lines indicate the linearly fitted trend lines. The colored numbers show the slope of the linear trends for each dataset.

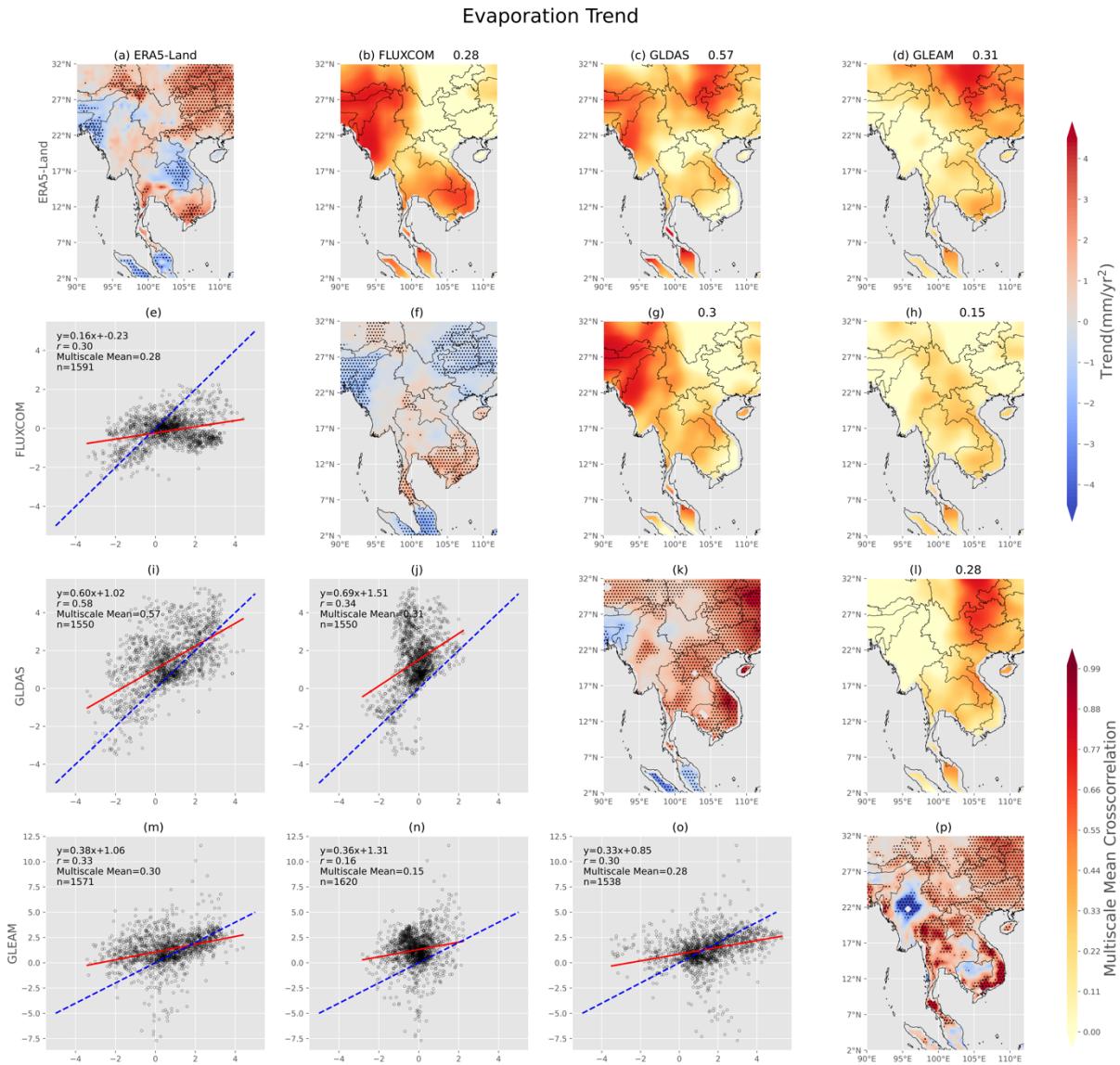


Fig. 6. The spatial distribution of linear trends for the four sets of evaporation data, along with scatter plots of the corresponding grid points and the spatial distribution of linear trend CC values between dataset pairs. The plots on the diagonal display the spatial distribution of the linear trends calculated grid point by grid point for each dataset. The dotted areas indicate regions where the linear regression results passed the 95% significance t-test, suggesting a significant increasing or decreasing trend in evaporation from 1980 to 2013. The upper triangular matrix shows the spatial distributions of CC values obtained using the CMP method, with the numbers next to each title representing the multiscale mean CC values, indicating the similarity in spatial distribution between the datasets. The lower triangular matrix contains scatter plots of corresponding grid points in pairwise comparisons, with the red line representing the fitted line and the blue dashed line indicating the 1:1 line. Each scatterplot subplot is annotated with the univariate linear regression coefficient, correlation coefficient, multi-scale mean CC, and the number of grid points.

d. Analysis of Relative Uncertainty in Evaporation Data

Fig. 7a to d illustrates the spatial distribution of relative uncertainty for the four sets of evaporation data, calculated using the TCH method. As shown, the relative uncertainty of ERA5-Land is highest in the northwestern region of the GMS, followed by the southwestern peninsula, and lowest in the eastern part of China. FLUXCOM data show low uncertainty in the central GMS. The GLDAS data exhibit high uncertainty in the northwestern part of the GMS, followed by the southwestern region, with the lowest uncertainty in the eastern part. GLEAM data demonstrates low relative uncertainty across most of the GMS, except for the high uncertainty observed in the northwestern area.

Fig. 7e illustrates the relative uncertainty of the four evaporation datasets across major land cover types in the GMS and its three subregions. Across the GMS, the average relative uncertainty of the four datasets is 0.15, with GLEAM exhibiting the lowest uncertainty at 0.11, while the other three datasets exceed the average. ERA5-Land shows the highest uncertainty across the entire GMS region. In terms of specific subregions, N-GMS has higher uncertainty than the average, whereas M-GMS and S-GMS have lower uncertainty, with M-GMS showing the lowest uncertainty. The dataset with the highest uncertainty varies across regions: ERA5-Land has the highest uncertainty in N-GMS, FLUXCOM exhibits the highest uncertainty in M-GMS, and GLDAS has the highest uncertainty in S-GMS, where ERA5-Land and FLUXCOM display similar levels of uncertainty. Despite these regional differences in the dataset with the highest uncertainty, GLEAM consistently maintains the lowest relative uncertainty across both the GMS and its subregions.

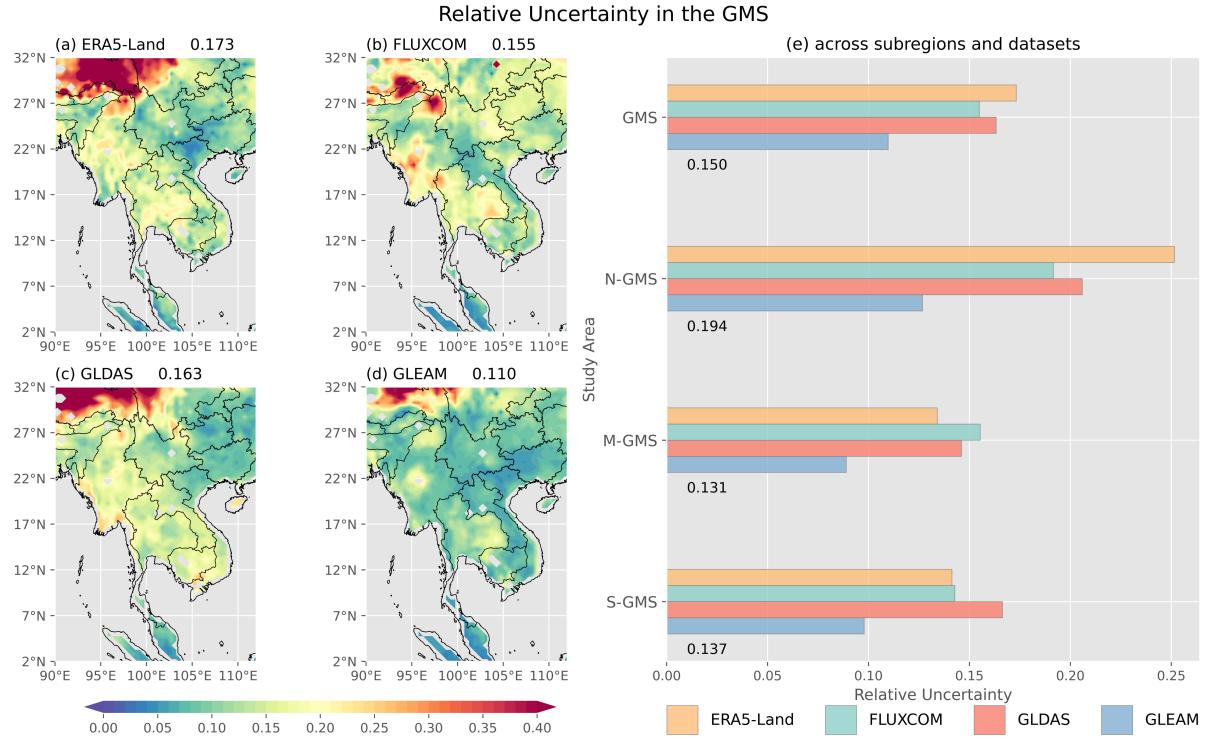


Fig. 7. (a) to (d): Spatial distribution of relative uncertainties for the four datasets based on TCH calculations. The numbers next to the subplot titles indicate the regional averages of relative uncertainty. (e): Relative uncertainty for the four datasets across the GMS, N-GMS, M-GMS, and S-GMS regions, with the numbers below each group representing the regional average relative uncertainty.

Fig. 8a illustrates the relative uncertainty across major land cover types in the GMS. Among these land cover types, grassland areas exhibit the highest average relative uncertainty at 0.294, with ERA5-Land showing the highest uncertainty and FLUXCOM the lowest. FLUXCOM data demonstrates the highest relative uncertainty in evergreen broadleaf forest and rainfed cropland, with GLEAM showing the lowest. In contrast, ERA5-Land data exhibits the highest uncertainty in grassland and evergreen needleleaf forest.

Fig. 8b illustrates the relative uncertainty for the N-GMS. Among the land cover types, grassland exhibits the highest uncertainty, with an average of 0.291 across the datasets, with ERA5-Land showing the highest uncertainty and FLUXCOM the lowest. In contrast, rainfed cropland demonstrates the lowest relative uncertainty at 0.116, where GLDAS shows the highest and GLEAM the lowest values. For evergreen forests, ERA5-Land data presents the highest relative uncertainty in needleleaf varieties, while GLEAM shows the lowest. In broadleaf forests, FLUXCOM displays the highest relative uncertainty, and GLEAM the lowest. In the M-GMS (Fig. 8c), rainfed cropland exhibits the highest relative uncertainty in evaporation data, averaging 0.156. For the major land cover types, GLDAS demonstrates the

highest uncertainty in evergreen broadleaf forests and mosaic tree and shrub areas, while FLUXCOM shows the highest uncertainty in rainfed cropland and mosaic natural vegetation areas. Fig. 8d illustrates the relative uncertainty for the S-GMS. Rainfed cropland demonstrates the highest relative uncertainty in evaporation data, averaging 0.152. FLUXCOM shows the highest relative uncertainty among the four datasets for rainfed cropland, while GLDAS exhibits the highest relative uncertainty for other major land cover types in the S-GMS region.

In summary, the relative uncertainty results of the four evaporation datasets across different land cover types in the GMS indicate that GLEAM shows the lowest relative uncertainty across almost all land cover types. N-GMS exhibits higher relative uncertainty compared to the other two regions, with grassland cover type showing the highest uncertainty. ERA5 data shows the highest relative uncertainty in grassland areas, while FLUXCOM has the highest relative uncertainty in M-GMS, and GLDAS shows the lowest relative uncertainty in the S-GMS.

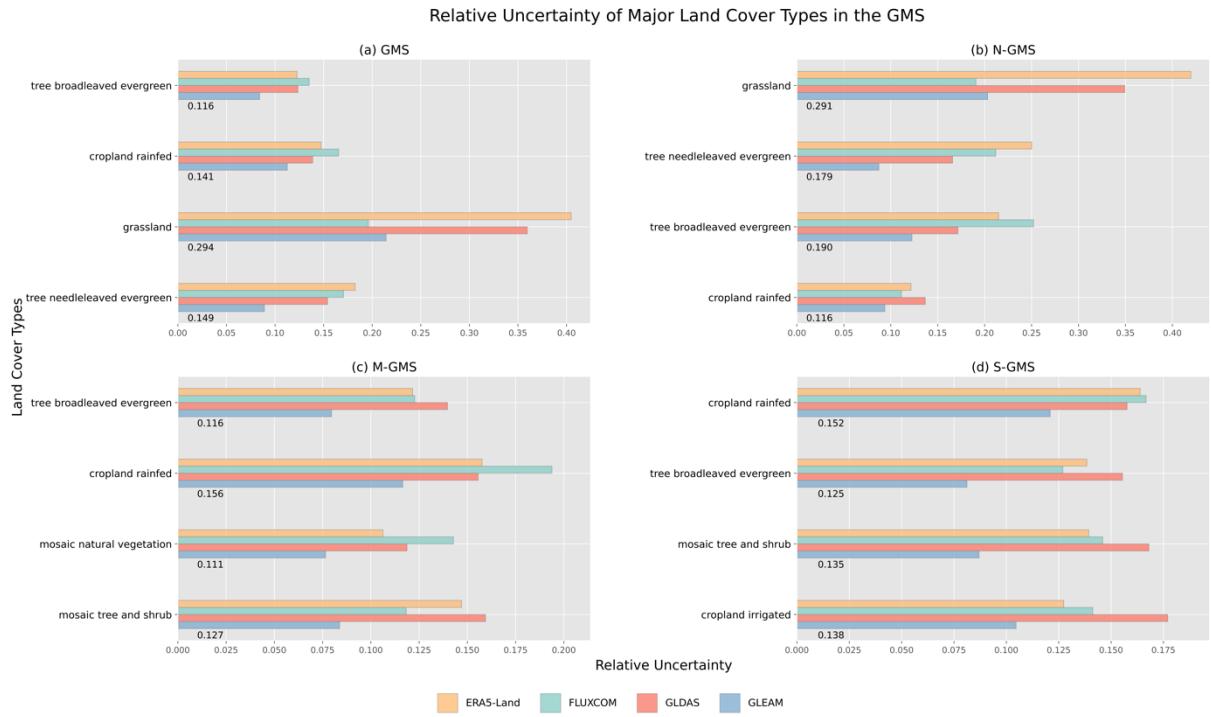


Fig. 8. Relative uncertainty of the four evaporation datasets across the four major land cover types in the GMS, N-GMS, M-GMS, and S-GMS regions. The numbers below each group represent the average relative uncertainty of the four datasets for each corresponding land cover type.

4. Discussion

In this study, we used the CMP method and scatter plots between datasets to evaluate the annual mean, interannual variability, and trends of evaporation in the GMS region across

the four datasets. Compared to previous studies on evaporation data evaluation (Elnashar et al. 2021; Tang et al. 2024), the annual evaporation in the GMS is approximately 1.5 times the global average and twice the Asian average. Zhang et al. (2016) suggested that the smaller variability of evaporation compared to precipitation is primarily due to the buffering effect of soil and vegetation on evaporation. In this study, the relatively weak spatial consistency of evaporation variability among the four datasets may stem from differences in how vegetation biochemical and physical processes, as well as soil evaporation, are handled across the datasets. Jung et al. (2019) highlighted that remote sensing data offer high spatiotemporal resolution biophysical variables, and machine learning models specifically trained on these data could help mitigate such limitations. Notably, the FLUXCOM dataset in this study exhibits the smallest interannual variability, yet in theory, evaporation data derived from machine learning should display greater variability since vegetation buffering processes are not considered; however, the observed results contradict this expectation, suggesting that other factors may contribute to the lack of pronounced interannual variability. Wang et al. pointed out that in the FLUXNET data used for training machine learning models, the primary drivers of evaporation variability are the seasonal and spatial variations of meteorological factors, particularly air temperature. In tropical rainforest regions, surface incoming solar radiation serves as the dominant control on evaporation, and while air temperature and surface solar radiation exhibit similar seasonal patterns, making temperature an effective proxy for simulating seasonal evaporation variations in tropical regions, the capacity of site-based upscaling machine learning models to accurately capture evaporation variability over longer timescales, such as interannual or decadal scales, remains to be assessed (Wang and Dickinson 2012). Jung et al. (2020) further emphasized that the machine learning methods employed in FLUXCOM assume training data to be independently and identically distributed, thereby failing to leverage the temporal structure within the data. Moreover, machine learning-based datasets exhibit reduced sensitivity to temporal variations when extrapolated beyond their original training domains (Pan et al., 2020), and the scarcity of long-term site observations in the GMS region may help explain the distinctive spatial patterns and limited interannual variability observed in FLUXCOM data. In the GMS region, we validated the only two available FLUXNET site datasets in the supplementary material (Fig. S1 and S2), where FLUXCOM data demonstrated strong correlations at the site scale, yet across other areas of the GMS region(Fig. 4, 5 and 6), their ability to effectively simulate interannual variability and trends could not be sustained. Jung et al. (2010) pointed out that the reversal of the global evaporation trend from an increase to a decrease in 1997, observed between 1982 and 2008, was primarily attributed to a decline in soil

moisture in the Southern Hemisphere. In this study, except for the FLUXCOM dataset, the other three datasets exhibit a significant increasing trend in evaporation across the GMS region, without any apparent stagnation in evaporation growth. Zhang et al. (2016) suggested that the overall increasing trend in evaporation was driven by enhanced vegetation transpiration and canopy interception evaporation, while soil evaporation showed a decreasing trend due to vegetation growth, which increased transpiration and canopy interception while reducing soil evaporation. In this study, the FLUXCOM evaporation trend differs from the other datasets mainly in the northeastern GMS region, which may be due to the fact that FLUXCOM data do not explicitly account for vegetation's influence on evaporation and that the spatial representativeness of flux tower sites in the GMS region is limited. Therefore, future research on evaporation estimation in the GMS region should focus on the integration of vegetation process information, cross-validation of multiple datasets to understand bias sources and uncertainties, and the exploration of climate drivers' contributions to evaporation at different temporal scales. Jung et al. (2010) pointed out that the shift from an increasing to a decreasing global evaporation trend in 1997, observed between 1982 and 2008, was primarily driven by declining soil moisture in the Southern Hemisphere. Unlike this global trend, in the present study, all three datasets—except for FLUXCOM—exhibit a significant upward trend in evaporation across the GMS region. Zhang et al. (2016) investigated global evaporation trends from 1981 to 2012 and concluded that the overall increase in evaporation was mainly attributed to the rise in vegetation transpiration and canopy interception evaporation, whereas soil evaporation exhibited a declining trend due to enhanced vegetation growth, which increased transpiration and interception while reducing soil evaporation. In this study, FLUXCOM data show an evaporation trend in the northeastern GMS region that contrasts with the other datasets, and the overall trend during the study period is not significant. Considering the discussion on variability, we speculate that this discrepancy may stem from FLUXCOM's lack of direct consideration of vegetation influences on evaporation, as well as the limited spatial representativeness of FLUXNET sites in the GMS region. Therefore, future research on evaporation estimation in the GMS should focus on incorporating vegetation process information, conducting cross-validation of datasets to better understand bias sources and uncertainties, and exploring the contributions of climate factors to evaporation across different temporal scales.

This study employs the TCH method to assess the relative uncertainty of four evaporation datasets across the GMS region. The results indicate that within GMS, N-GMS

exhibits the highest uncertainty. In N-GMS, ERA5-Land data show the greatest uncertainty, particularly in grassland-dominated areas, primarily due to the sparse and uneven distribution of observational stations in high-altitude regions and the lack of assimilation corrections based on actual observational data (Muñoz-Sabater et al. 2021). Cui et al. (2024) also found that ERA5-Land has higher uncertainty over the Tibetan Plateau and southern India, likely due to the scarcity of assimilated observational data in these regions. In M-GMS, FLUXCOM data exhibit the highest uncertainty. This region features diverse and fragmented land cover types (Fig. 1 and S2), and FLUXCOM has the lowest spatial resolution among the four datasets, which may contribute to its higher uncertainty. Additional factors influencing FLUXCOM's high uncertainty include the insufficient representativeness of flux tower sites across different global land cover types and the difficulty in distinguishing surface types at coarser spatial resolutions (Zhang et al. 2019). In M-GMS, where cropland rainfed and evergreen broadleaf forests are the dominant vegetation cover types, FLUXCOM uncertainty is particularly high in rainfed croplands compared to other land cover types. FLUXCOM is trained using FLUXNET site data, including AmeriFlux (Jung et al. 2019), and a study by Yi et al. (2024) found that agricultural areas account for only 22% of AmeriFlux sites as of July 2024, whereas agriculture consumes approximately 70% of global freshwater resources (Zhang et al. 2022). Expanding the spatial coverage of in situ observations in agricultural ecosystems is therefore crucial for evaluating and improving evaporation models (Volk et al. 2024). In S-GMS, GLDAS data exhibit the highest uncertainty. The uncertainty of GLDAS data is particularly pronounced over croplands and grasslands, consistent with previous findings (Wang et al. 2011). Since GLDAS is generated using offline land surface models, its evaporation data uncertainty is influenced by the parameterization of land cover types and the uncertainty in meteorological forcing (Badgley et al. 2015). Among the four datasets, GLEAM demonstrates the lowest relative uncertainty across GMS and nearly all land cover types. GLEAM integrates a soil moisture constraint algorithm based on satellite observations, and as more remotely sensed soil moisture data become available, incorporating these data is expected to further enhance the accuracy of evaporation estimates (Xu et al. 2018). Therefore, future research should focus on leveraging soil observational data and refining soil evaporation-related parameters to improve evaporation simulations.

Finally, although the TCH method can serve as a tool for uncertainty assessment in the absence of extensive reference data, its limitations must be acknowledged. A detailed assessment of the reliability and applicability of TCH in the GMS region can be found in

Section C of the Supplementary Materials. In the TCH framework, relative uncertainty refers to the covariance of errors between each dataset and the assumed true value—based on the assumption that each evaporation dataset captures a portion of the actual evaporation—and is iteratively computed using the covariance of errors among datasets. The method exhibits sensitivity to data characteristics, particularly when strong error correlations exist between datasets. For example, if both site-level observations and a machine learning model trained on those site data are included in the TCH calculation, the error covariance may decrease simultaneously. Although the TCH method has some tolerance for such cases (Sjoberg et al. 2021), they should be avoided as much as possible to ensure that TCH can effectively serve its intended role in uncertainty analysis when reference data are limited. Moreover, due to its methodological constraints, TCH cannot directly identify the specific causes of high or low uncertainty. Therefore, when applying the TCH method, it is essential to conduct multidimensional comparative analyses—such as the CMP method used in this study to evaluate the spatial consistency of datasets—and integrate multiple sources of information to better understand uncertainties and guide future improvements.

5. Conclusion

This study evaluates the spatial consistency and relative uncertainty of four evaporation datasets (ERA5-Land, FLUXCOM, GLDAS, and GLEAM) in the GMS region from 1980 to 2013 using the CMP and TCH methods. The results show that the annual average evaporation across these datasets ranges from 809.47 to 920.66 mm/yr. While the spatial consistency of mean evaporation is generally high, ERA5-Land exhibits slightly lower consistency in the southern GMS. However, the spatial consistency of evaporation variability is poor. Among the datasets, FLUXCOM has the lowest variability with a uniform spatial distribution, while GLEAM shows weak correlation with the other datasets. The remaining datasets exhibit almost no correlation with each other, although GLDAS demonstrates some similarity to the other datasets when analyzed using the CMP method for multi-scale spatial similarity. Regarding annual trends, three datasets (ERA5-Land, GLDAS, and GLEAM) exhibit significant increasing trends, while FLUXCOM shows no significant interannual variation or trend. Spatially, FLUXCOM is inconsistent with the other datasets in the eastern GMS, whereas GLDAS has the highest consistency with ERA5-Land, with a correlation coefficient of 0.57. The relative uncertainty assessment indicates that GLEAM has the lowest relative uncertainty across all regions and nearly all land cover types in the GMS.

To further investigate spatial variations, we divided the GMS into three subregions and analyzed the relative uncertainty of each dataset across different land cover types. The results show that in the overall GMS, ERA5-Land has the highest relative uncertainty, while GLEAM has the lowest. In the N-GMS, ERA5-Land exhibits the highest uncertainty, mainly due to the high uncertainty in grassland, the dominant land cover type in this region. In the M-GMS, FLUXCOM shows the highest uncertainty, primarily due to high uncertainty in rainfed cropland. In the S-GMS, GLDAS exhibits the highest uncertainty, largely driven by uncertainty in irrigated cropland.

By integrating the CMP and TCH method results, we further discuss the potential causes of inconsistencies in the mean values, variability, and trends among the four datasets, as well as the spatial variations in relative uncertainty across different land cover types in the GMS. Additionally, we assess the reliability and applicability of the TCH method in the GMS, highlighting its limitations related to data sensitivity and error correlation. When applying the TCH method for data analysis, multiple dimensions of information should be considered comprehensively. Future research on evaporation estimation should focus on incorporating vegetation process information, using cross-validation to understand the sources of bias and uncertainty among datasets, and exploring the influence of climatic factors on evaporation across different timescales to enhance the ability of models to capture temporal variations in evaporation. In studies of evaporation components, utilizing soil observation data and adjusting soil evaporation-related parameters will be crucial for improving evaporation simulations.

Acknowledgments.

This study was jointly funded by the National Key Research and Development Program of China (2023YFF0805504), National Natural Science Foundation of China (42065002), the Natural Science Foundation of Yunnan Province (Grant No. 202401BC070004, 202401AT070447), and the Youth Innovation Team of China Meteorological Administration “Climate change and its impacts in the Tibetan Plateau” (CMA2023QN16).

Data Availability Statement.

Datasets available from public repository that assigns persistent identifiers to datasets: All latent heat and evaporation data come from: ERA5-Land data as cited in Muñoz-Sabater et al. (2021), FLUXCOM data as cited in (Jung et al. 2019), GLDAS-Noah data as cited in Rodell

et al. (2004), GLEAM data as cited in Martens et al. (2017). Land cover types data come from: Copernicus Climate Change Service, Climate Data Store, (2019): Land cover classification gridded maps from 1992 to present derived from satellite observation. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). DOI: 10.24381/cds.006f2c9a

REFERENCES

- Awange, J. L., V. G. Ferreira, E. Forootan, Khandu, S. A. Andam-Akorful, N. O. Agutu, and X. F. He, 2016: Uncertainties in remotely sensed precipitation data over Africa: UNCERTAINTIES IN REMOTELY SENSED PRECIPITATION DATA OVER AFRICA. *Int. J. Climatol.*, 36, 303–323, <https://doi.org/10.1002/joc.4346>.
- Badgley, G., J. B. Fisher, C. Jim, K. P. Tu, and R. Vinukollu, 2015: On Uncertainty in Global Terrestrial Evapotranspiration Estimates from Choice of Input Forcing Datasets. *JOURNAL OF HYDROMETEOROLOGY*, 16.
- Baik, J., and M. Choi, 2015: Evaluation of geostationary satellite (COMS) based Priestley–Taylor evapotranspiration. *Agricultural Water Management*, 159, 77–91, <https://doi.org/10.1016/j.agwat.2015.05.017>.
- Baldocchi, D., 2001: FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*, 82, [https://doi.org/10.1175/1520-0477\(2001\)082<2415:FANTTS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2).
- Chen, F., and Coauthors, 1996: Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *Journal of Geophysical Research: Atmospheres*, 101, 7251–7268, <https://doi.org/10.1029/95JD02165>.
- Cui, Z., Y. Zhang, A. Wang, J. Wu, and C. Li, 2024: Uncertainty Analysis and Data Fusion of Multi-Source Land Evapotranspiration Products Based on the TCH Method.
- Dickinson, E., A. Henderson-Sellers, J. Kennedy, and F. Wilson, 1986: Biosphere-atmosphere Transfer Scheme (BATS) for the NCAR Community Climate Model. <https://doi.org/10.5065/D6668B58>.
- Elnashar, A., L. Wang, B. Wu, W. Zhu, and H. Zeng, 2021: Synthesis of global actual evapotranspiration from 1982 to 2019. *Earth Syst. Sci. Data*, 13, 447–480, <https://doi.org/10.5194/essd-13-447-2021>.
- Ferguson, C. R., J. Sheffield, E. F. Wood, and H. Gao, 2010: Quantifying uncertainty in a remote sensing-based estimate of evapotranspiration over continental USA. *International Journal of Remote Sensing*, 31, 3821–3865, <https://doi.org/10.1080/01431161.2010.483490>.

- Galindo, F., and J. Palacio, 1999: Estimating the Instabilities of N Correlated Clocks. 13.
- Gaucherel, C., S. Alleaume, and C. Hély, 2007: The Comparison Map Profile method: a strategy for multiscale comparison of quantitative and qualitative images. 2007,
- Grumbine, R. E., J. Dore, and J. Xu, 2012: Mekong hydropower: drivers of change and governance challenges. *Frontiers in Ecol & Environ*, 10, 91–98, <https://doi.org/10.1890/110146>.
- Hoang, L. P., and Coauthors, 2016: Mekong River flow and hydrological extremes under climate change. *Hydrol. Earth Syst. Sci.*, 20, 3027–3041, <https://doi.org/10.5194/hess-20-3027-2016>.
- Hundertmark, W., 2008: Building drought management capacity in the Mekong River basin. *Irrigation and Drainage*, 57, 279–287, <https://doi.org/10.1002/ird.435>.
- Huntington, T. G., 2006: Evidence for intensification of the global water cycle: Review and synthesis. *Journal of Hydrology*, 319, 83–95, <https://doi.org/10.1016/j.jhydrol.2005.07.003>.
- Jung, M., and Coauthors, 2010: Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, 467, 951–954, <https://doi.org/10.1038/nature09396>.
- , and Coauthors, 2019: The FLUXCOM ensemble of global land-atmosphere energy fluxes. *Sci Data*, 6, 74, <https://doi.org/10.1038/s41597-019-0076-8>.
- , and Coauthors, 2020: Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach.
- Khan, M. S., U. W. Liaqat, J. Baik, and M. Choi, 2018: Stand-alone uncertainty characterization of GLEAM, GLDAS and MOD16 evapotranspiration products using an extended triple collocation approach. *Agricultural and Forest Meteorology*, 252, 256–268, <https://doi.org/10.1016/j.agrformet.2018.01.022>.
- Kim, S., A. Anabal, and A. Sharma, 2021: An Assessment of Concurrency in Evapotranspiration Trends across Multiple Global Datasets. *JOURNAL OF HYDROMETEOROLOGY*, 22.
- Koren, V., J. Schaake, K. Mitchell, Q.-Y. Duan, F. Chen, and J. M. Baker, 1999: A parameterization of snowpack and frozen ground intended for NCEP weather and climate models. *Journal of Geophysical Research: Atmospheres*, 104, 19569–19585, <https://doi.org/10.1029/1999JD900232>.
- Kuang, X., and J. J. Jiao, 2016: Review on climate change on the Tibetan Plateau during the last half century. *JGR Atmospheres*, 121, 3979–4007, <https://doi.org/10.1002/2015JD024728>.

Li, X., and Coauthors, 2018: Spatiotemporal pattern of terrestrial evapotranspiration in China during the past thirty years. *Agricultural and Forest Meteorology*, 259, 131–140, <https://doi.org/10.1016/j.agrformet.2018.04.020>.

Liaqat, U. W., and M. Choi, 2015: Surface energy fluxes in the Northeast Asia ecosystem: SEBS and METRIC models using Landsat satellite images. *Agricultural and Forest Meteorology*, 214–215, 60–79, <https://doi.org/10.1016/j.agrformet.2015.08.245>.

Liu, J., D. Chen, G. Mao, M. Irannezhad, and Y. Pokhrel, 2022: Past and Future Changes in Climate and Water Resources in the Lancang–Mekong River Basin: Current Understanding and Future Research Directions. *Engineering*, 13, 144–152, <https://doi.org/10.1016/j.eng.2021.06.026>.

Liu, L., P. Bai, C. Liu, W. Tian, and K. Liang, 2020: Changes in Extreme Precipitation in the Mekong Basin. *Advances in Meteorology*, 2020, 1–10, <https://doi.org/10.1155/2020/8874869>.

Liu, Y., T. Hiyama, T. Yasunari, and H. Tanaka, 2012: A nonparametric approach to estimating terrestrial evaporation: Validation in eddy covariance sites. *Agricultural and Forest Meteorology*, 157, 49–59, <https://doi.org/10.1016/j.agrformet.2012.01.012>.

Long, D., L. Longuevergne, and B. R. Scanlon, 2014: Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. *Water Resour. Res.*, 50, 1131–1151, <https://doi.org/10.1002/2013WR014581>.

Lyon, S. W., K. King, O. Polpanich, and G. Lacombe, 2017: Assessing hydrologic changes across the Lower Mekong Basin. *Journal of Hydrology: Regional Studies*, 12, 303–314, <https://doi.org/10.1016/j.ejrh.2017.06.007>.

Manabe, S., 1969: Climate and the ocean circulation, I, The atmospheric circulation and the hydrology of the Earth's surface. *mwr*, 97, [https://doi.org/10.1175/1520-0493\(1969\)09760;0739:catoc62;2.3.co;2](https://doi.org/10.1175/1520-0493(1969)09760;0739:catoc62;2.3.co;2).

Martens, B., and Coauthors, 2017: GLEAM v3: satellite-based land evaporation and root-zone soil moisture. *Geosci. Model Dev.*, 10, 1903–1925, <https://doi.org/10.5194/gmd-10-1903-2017>.

Miralles, D. G., W. T. Crow, and M. H. Cosh, 2010: Estimating Spatial Sampling Errors in Coarse-Scale Soil Moisture Estimates Derived from Point-Scale Observations. *Journal of Hydrometeorology*, 11, 1423–1429, <https://doi.org/10.1175/2010JHM1285.1>.

Miralles, D. G., T. R. H. Holmes, R. A. M. De Jeu, J. H. Gash, A. G. C. A. Meesters, and A. J. Dolman, 2011: Global land-surface evaporation estimated from satellite-based observations. *Hydrol. Earth Syst. Sci.*, 15, 453–469, <https://doi.org/10.5194/hess-15-453-2011>.

Muñoz-Sabater, J., and Coauthors, 2021: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data*, 13, 4349–4383, <https://doi.org/10.5194/essd-13-4349-2021>.

Niu, G.-Y., and Coauthors, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research-Atmospheres*, <https://doi.org/10.1029/2010JD015139>.

Oki, T., and S. Kanae, 2006: Global Hydrological Cycles and World Water Resources. *Science*, 313, 1068–1072, <https://doi.org/10.1126/science.1128845>.

Oleson, W., and Coauthors, 2010: Technical Description of version 4.0 of the Community Land Model (CLM). <https://doi.org/10.5065/D6FB50WZ>.

Pan, S., and Coauthors, 2020: Evaluation of global terrestrial evapotranspiration using state-of-the-art approaches in remote sensing, machine learning and land surface modeling. *Hydrol. Earth Syst. Sci.*, 24, 1485–1509, <https://doi.org/10.5194/hess-24-1485-2020>.

Premoli, A., and P. Tavella, 1993: A revisited three-cornered hat method for estimating frequency standard instability. *IEEE Trans. Instrum. Meas.*, 42, 7–13, <https://doi.org/10.1109/19.206671>.

Prince, S. D., S. J. Goetz, R. O. Dubayah, K. P. Czajkowski, and M. Thawley, 1998: Inference of surface and air temperature, atmospheric precipitable water and vapor pressure deficit using Advanced Very High-Resolution Radiometer satellite observations: comparison with field observations. *Journal of Hydrology*, 212–213, 230–249, [https://doi.org/10.1016/S0022-1694\(98\)00210-8](https://doi.org/10.1016/S0022-1694(98)00210-8).

Rodell, M., and Coauthors, 2004: The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*, 85, 381–394, <https://doi.org/10.1175/BAMS-85-3-381>.

Sjoberg, J. P., R. A. Anthes, and T. Rieckh, 2021: The Three-Cornered Hat Method for Estimating Error Variances of Three or More Atmospheric Datasets. Part I: Overview and Evaluation. *Journal of Atmospheric and Oceanic Technology*, 38, 555–572, <https://doi.org/10.1175/JTECH-D-19-0217.1>.

Su, Z., T. Schmugge, W. P. Kustas, and W. J. Massman, 2001: An Evaluation of Two Models for Estimation of the Roughness Height for Heat Transfer between the Land Surface and the Atmosphere. *J. Appl. Meteor.*, 40, 1933–1951, [https://doi.org/10.1175/1520-0450\(2001\)040<1933:AEOTMF>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<1933:AEOTMF>2.0.CO;2).

Tang, R., and Coauthors, 2024: Spatial-temporal patterns of land surface evapotranspiration from global products. *Remote Sensing of Environment*, 304, 114066, <https://doi.org/10.1016/j.rse.2024.114066>.

Tavella, P., and A. Premoli, 1994: Estimating the Instabilities of N Clocks by Measuring Differences of their Readings. *Metrologia*, 30, 479–486, <https://doi.org/10.1088/0026-1394/30/5/003>.

Tran, B. N., J. Van Der Kwast, S. Seyoum, R. Uijlenhoet, G. Jewitt, and M. Mul, 2023: Uncertainty assessment of satellite remote-sensing-based evapotranspiration estimates: a systematic review of methods and gaps. *Hydrol. Earth Syst. Sci.*, 27, 4505–4528, <https://doi.org/10.5194/hess-27-4505-2023>.

Trenberth, K. E., J. T. Fasullo, and J. Kiehl, 2009: Earth's Global Energy Budget. *Bull. Amer. Meteor. Soc.*, 90, 311–324, <https://doi.org/10.1175/2008BAMS2634.1>.

Volk, J. M., and Coauthors, 2024: Assessing the accuracy of OpenET satellite-based evapotranspiration data to support water resource and land management applications. *Nat Water*, 2, 193–205, <https://doi.org/10.1038/s44221-023-00181-7>.

Wang, F., L. Wang, T. Koike, H. Zhou, K. Yang, A. Wang, and W. Li, 2011: Evaluation and application of a fine-resolution global data set in a semiarid mesoscale river basin with a distributed biosphere hydrological model: GLDAS/NOAH EVALUATION AND APPLICATION. *J. Geophys. Res.*, 116, <https://doi.org/10.1029/2011JD015990>.

Wang, K., and R. E. Dickinson, 2012: A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Reviews of Geophysics*, 50, <https://doi.org/10.1029/2011RG000373>.

Wartenburger, R., and Coauthors, 2018: Evapotranspiration simulations in ISIMIP2a—Evaluation of spatio-temporal characteristics with a comprehensive ensemble of independent datasets. *Environ. Res. Lett.*, 13, 075001, <https://doi.org/10.1088/1748-9326/aac4bb>.

Xu, T., and Coauthors, 2018: Evaluating Different Machine Learning Methods for Upscaling Evapotranspiration from Flux Towers to the Regional Scale. *JGR Atmospheres*, 123, 8674–8690, <https://doi.org/10.1029/2018JD028447>.

Yi, K., and Coauthors, 2024: Challenges and Future Directions in Quantifying Terrestrial Evapotranspiration. *Water Resources Research*, 60, e2024WR037622, <https://doi.org/10.1029/2024WR037622>.

Yu, G.-R., X.-F. Wen, X.-M. Sun, B. D. Tanner, X. Lee, and J.-Y. Chen, 2006: Overview of ChinaFLUX and evaluation of its eddy covariance measurement. *Agricultural and Forest Meteorology*, 137, 125–137, <https://doi.org/10.1016/j.agrformet.2006.02.011>.

Zhang, K., X. Li, D. Zheng, L. Zhang, and G. Zhu, 2022: Estimation of Global Irrigation Water Use by the Integration of Multiple Satellite Observations. *Water Resources Research*, 58, e2021WR030031, <https://doi.org/10.1029/2021WR030031>.

Zhang, Y., and Coauthors, 2016: Multi-decadal trends in global terrestrial evapotranspiration and its components. *Sci Rep*, 6, 19124, <https://doi.org/10.1038/srep19124>.

—, D. Kong, R. Gan, F. H. S. Chiew, T. R. McVicar, Q. Zhang, and Y. Yang, 2019: Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017. *Remote Sensing of Environment*, 222, 165–182, <https://doi.org/10.1016/j.rse.2018.12.031>.