# An Assessment of Concurrency in Evapotranspiration Trends Across Multiple Global Datasets

Seokhyeon Kim[1], Alfonso Anabalón[1, a] and Ashish Sharma[1 *]

[1] School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW 2052, Australia

[a] Alfonso Anabalón is currently at SGA S.A., Santiago, Chile

* Corresponding author: Ashish Sharma (a.sharma@unsw.edu.au), School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW 2052, Australia

# Abstract

While broad consensus exists that temperatures are increasing, there is uncertainty surrounding the direction of change manifested in actual evapotranspiration (ET) worldwide. This study assessed trends in ET across the land surface using eleven widely used global datasets for a 32-year study period. To demonstrate the agreement and disagreement of trends, the spatial distribution, concurrence, correlation and similitude were estimated. The results showed that while the global average trend in ET is -0.072 mm/month/year, the trends from individual datasets show a wide range of differences in magnitudes and directions. The considerable differences in the trends in each dataset were found to be weakly correlated to each other and highly divergent in their distribution and direction. No single dataset was sufficiently similar to another to offer a fair representation of trends. In a dynamic trend analysis using a 10-year moving window over the study period, high concurrence in the significant trends throughout the datasets was found to be rare for each time period. In general, the global data concurrence became negative by 1997 but rebounded to positive towards the end of the study period. In terms of spatial tendency, some regions were more prone to change the direction of their significant trends within the study period. This shows a high inconsistency in the location and direction of significant ET trends, implying selection of an ET dataset should consider its spatiotemporal uncertainty before use for any water balance study aiming to infer hydrological change over time.

**Keywords**: actual evapotranspiration; concurrency; trend analysis

# 1. Introduction

Actual evapotranspiration (ET) is a major component in the hydrological cycle together with precipitation, along with being one of the major constituents of the surface energy balance (Trenberth et al. 2009; Trenberth et al. 2007). Changes in terrestrial ET can affect precipitation, streamflow and temperature amongst other hydro-climatological variables (Koster et al. 2004; Seneviratne et al. 2010). Additionally, ET plays an important role in the climate system coupling with the water, carbon and energy cycles (Jung et al. 2010).

Being such an essential variable, its global characterization is critical to fully understand and model the main processes behind climate and life cycles on our planet. Despite global efforts to directly measure ET such as the FLUXNET project (Baldocchi et al. 2001), few long time series of ground observations exist. Additionally, while remote sensing has made major strides in recent years, satellites are still incapable of quantifying ET amounts (Zhang et al. 2016a) directly. Instead, remotely sensed data of related variables, such as temperature, radiation and vegetation coverage have allowed the creation of several derived ET products in recent years. Such ET products have been often used as surrogates of ET in water balance studies (Fisher et al. 2008; Jung et al. 2010; Miralles et al. 2011; Mu et al. 2007; Yang et al. 2013; Zhang et al. 2010). Modelled ET products, and derivations using land surface models such as the ones used in the Global Land Data Assimilation (Rodell et al. 2004) have also been developed and are being used for a range of applications.

All of the above-mentioned products rely on different forcing data and methods to estimate ET. This leads to different magnitudes and spatial distributions of ET on our planet. Several studies have investigated such differences in detail, showing high uncertainties in the ET estimations using different models (Chen et al. 2014; Ershadi et al. 2014; Jiménez et al. 2011; Miralles et al. 2016; Mueller et al. 2011), and McMahon et al. (2013) provide an insight of the vast disparities in evapotranspiration assessment using different theoretical methods. This is consistent with uncertainties associated with other global variable datasets, many of which are derived from satellite retrievals (Kim; Sharma 2019; Libertino et al. 2016; Zhang et al. 2019).

The one aspect that has not been looked in detail is the response of ET, as represented in these data, to rising temperatures. The dissimilarity between the trends in ET across different datasets is important especially when studying climate change, as there is consensus across modelling and observational groups worldwide that global temperatures are increasing, leading to speculation that similar increases should be reflected in observed ET as well as projected ET for future conditions (Johnson; Sharma 2010; Stephens et al. 2018). Quicker dryness of catchments following precipitation events is also noted to be occurring, again attributed to higher potential evaporation rates (Fitsum; Ashish 2016; Hettiarachchi et al. 2019). The observational uncertainty and dependence on models are limiting our capacity to detect changes in ET (IPCC 2014), even without considering the intrinsic variations amongst the datasets. This uncertainty also has considerable implications on climate change impact assessments as global ET datasets are used to compare the results of General Circulation

Models (Mueller et al. 2011) and to investigate the drivers behind the trends in ET (Jung et al. 2010; Miralles et al. 2013; Trambauer et al. 2014).

Attempts have been made to assess the uncertainty of the trends in ET at a global scale (Vinukollu et al. 2011; Zhang et al. 2016c). However, their regional variations are less well understood. Moreover, previous works have studied the ET trends over several datasets, but have mostly focused on specific regions (Kiptala et al. 2013; Marshall et al. 2012), global totals (Jung et al. 2010; Zhang et al. 2016c) or merged products (Mueller et al. 2013). Even though the uses of merged products, constructed from a group of observations, provide useful insights over the expected changes in ET in a specific zone, they tend to mask the differences between the datasets. Namely, the trend of the merged product could be overshadowed by the presence of a dataset with considerably higher values of ET or trend magnitudes.

Considering the limitations of existing research on ET trends, this work first aims to assess the spatial differences between the ET trends across global datasets, in order to identify zones where the trends concur and differ. The second aim of this study is to analyze how many of these datasets coincide in terms of magnitude and direction of trends so as to provide assessments on the more consistent ET data to use. Therefore, in the present work, the analysis focuses on the consistency of direction, statistical significance and currency in ET trends amongst the datasets. In this way, we expect to provide a deeper understanding of uncertainty of the trends and identify regions where the direction of change appears inconsistent, and hence, dubious.

# 2. Data and methodology

## 2.1 Selected ET datasets and data processing

The global scale of the problem required a group of global ET datasets, with enough monthly data points to allow a trend analysis to be carried out with reliability. Considering this, as detailed in Table 1, eleven global ET datasets were selected. Based on Mueller et al. (2011), this study classifies the eleven datasets into three categories. These are ‘diagnostic’: derived from combinations of observations or observation-based estimates, together with relatively simple or empirical formulations; ‘land-surface model’: derived from land surface models with observation-based forcing; ‘reanalysis’: models with assimilating observations, respectively.

For the trend analysis in this study, the following preprocessing was carried out for the datasets. First, a monthly temporal scale (mm/month) was used to fully utilize the available data. Second, all datasets at various spatial resolutions were resampled to the global cylindrical 36-km Equal-Area Scalable Earth, version 2 (EASEv2) grid (Brodzik et al. 2012) by using the nearest neighbor algorithm to simplify the comparison being presented, and the trend test was implemented over each cell. Note that Antarctica was not considered in the analysis because of lack of data over the area. Lastly, a 32-year period from January 1981 to December 2012 was selected for the evaluation which is the longest common period covered by all datasets.

[Table 1 about here]

### 2.1.1 CSIRO

The Commonwealth Scientific and Industrial Research Organization (CSIRO) dataset is based on the Penman-Monteith-Leuning model, using the WATCH-Forcing-DATA-ERA-Interim (WFDEI) and the Princeton Global Forcing (PGF) as meteorological forcing data (Zhang et al. 2016b; Zhang et al. 2016c). The leaf area index data was obtained from the Boston University (BU) dataset, meanwhile the Global Land Surface Satellite (GLASS) dataset was used for emissivity and albedo. The land cover was taken from International Geosphere-Biosphere Program (IGBP) data. This dataset has the particularity of being conceived to study trends in evaporation, thus its trends were validated against estimated evaporation trends in several parts of the world.

### 2.1.2 GLEAM

The Global Land Evaporation Amsterdam Model (GLEAM) has two different datasets which were based on reanalysis (v3.3a) and remote sensing data (v3.3b) respectively (Martens et al. 2017; Miralles et al. 2011). In this study, we used the first one, GLEAM v3.3a. GLEAM v3.3a uses various data as; 1) European Centre for Medium-Range Weather Forecasts (ECMWF) ReAnalysis Interim (ERA-Interim) data for radiation and air temperature; 2) Multi-Source Weighted-Ensemble Precipitation (MSWEP) v1.0 for precipitation; 3) Global Snow Monitoring for Climate Research (GLOBSNOW) L3A v2 and National Snow and Ice Data Center (NSIDC) v01 for snow water equivalents; and 4) Land Parameter Retrieval Model (LPRM)-based vegetation optical depth.

### 2.1.3 GLDAS

Four Global Land Data Assimilation System (GLDAS) version 1.0 (GLDAS-1) and two version 2.0 (GLDAS-2) datasets were used in this study, each one based on a different Land Surface Model (LSM) (Rodell et al. 2004). GLDAS-1 datasets share the same forcing data, which change over time. To estimate ET from 1979 to 1993 a bias-corrected ERA dataset was used. From 1994 to 1999, the utilized dataset was a bias-corrected National Center for Atmospheric Research (NCAR) reanalysis; for year 2000, the National Oceanic and Atmospheric Administration (NOAA)/Global Data Assimilation System (GDAS) atmospheric analysis fields were used; and from 2001 to the present, a mixture of NOAA/GDAS atmospheric fields, Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP) precipitation and observation-based downward shortwave and longwave radiation fields derived using the method of the Air Force Weather Agency's AGRicultural METeorological modeling system (AGRMET) were combinedly used to estimate ET amongst other variables. It should be noted that the GLDAS-1 was decommissioned in June 2020, and replaced with GLDAS-2.0, 2.1 and 2.2. While GLDAS-1 used multiple different meteorological datasets as mentioned above, GLDAS-2 is entirely forced with the Princeton meteorological forcing dataset (Sheffield et al. 2006) and therefore provides a temporally consistent series from 1948 through 2014 (Rui; Beaudoing 2018). Taking into account that both GLDAS-1 and -2 have different forcing datasets in terms of type and temporal continuity, the differences in time trends were detailed and discussed in Section 3.

Considering that each version of GLDAS datasets uses the same forcing variables, the results of these LSMs are expected to be similar and highly correlated, as their differences arise from their LSM configuration. The Mosaic LSM estimates evaporation through an energy and water balance on the surface layer of the model (Koster; Suarez 1996; Rodell et al. 2007d); meanwhile the Variable Infiltration Capacity (VIC) model uses the Penman-Monteith equation for the potential evaporation as the basis of its estimation, limiting its effects through water balances in every layer of the model (Liang et al. 1996; Liang et al. 1994; Rodell et al. 2007a). The Common Land Model (CLM) uses a modified version of the Biosphere-Atmosphere Transfer Scheme (BATS) (Dickinson et al. 1993) and the Philip (1957) method to calculate the ground evaporation (Dai et al. 2003; Rodell et al. 2007c). The basis for the evaporation estimation in the Noah model is the Penman equation (Chen et al. 1996; Koren et al. 1999; Rodell et al. 2007b). The Catchment Land Surface (CLS) model (Koster et al. 2000) uses the Mosaic model's energy balance calculation and canopy interception reservoir formulation.

### 2.1.4 ERA-Interim

This study also used the ERA-Interim (Dee et al. 2011) dataset that represents the synoptic monthly means of evaporation. The revised land surface Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land (HTESSEL) is adopted for the land component which interacts with the atmosphere component. The ET is estimated based on an energy balance using the ERA-Interim forecasts of near-surface conditions (i.e., temperature, humidity, pressure and wind speed) and downward energy and water fluxes (i.e., precipitation and solar/thermal radiation) (Balsamo et al. 2009; Trambauer et al. 2014).

### 2.1.5 MERRA2

We also used the monthly Evaporation land (EVLAND) reanalysis data from the second Modern-Era Retrospective analysis for Research and Applications (MERRA2) products (Gelaro et al. 2017). MERRA2 is successor of the original MERRA reanalysis (Rienecker et al. 2011), produced by NASA's Global Modeling and Assimilation Office (GMAO) using an upgraded version of the Goddard Earth Observing System Model, Version 5 (GEOS-5) data assimilation system.

### 2.1.6 JRA55

Lastly, we used monthly Evaporation (EVP) reanalysis data from the Japanese 55-year Reanalysis (JRA55) produced by the Japan Meteorological Agency (JMA) (Kobayashi et al. 2015). JRA55 data is an advanced version of its predecessor, JRA25, which considerably mitigated the major biases reported in JRA25 and also improved the temporal consistency of temperature.

## 2.2 Methods

### 2.2.1 Seasonal Kendall test

The Mann-Kendall test (Hirsch et al. 1982; Kendall 1948; Mann 1945) is used to reject the null hypothesis ($\mathrm{H}_0$) or accept the alternative hypothesis ($\mathrm{H}_1$) of a two-sided test, which are; $\mathrm{H}_0$: the data ($x_1 \ldots x_n$) are a sample of *n* independent and identically distributed random variables, and $\mathrm{H}_1$: the distribution of $x_k$ and $x_j$ are not identical for all $k, j$ with $k \neq j$, respectively.

The test statistic S is defined as

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \mathrm{sgn}(\mathrm{x}_j - \mathrm{x}_k)$$

where (1)

$$sgn(\theta) = \begin{cases} 1 & \text{if } \theta > 0 \\ 0 & \text{if } \theta = 0 \\ -1 & \text{if } \theta < 0 \end{cases}$$

ET generally exhibits a strong seasonality (Huizhi; Jianwu 2012; Vourlitis et al. 2002) which should be incorporated to the trend analysis but the original MK test does not explicitly consider this. In this study, the Seasonal Kendall (SK) test (Helsel; Hirsch 1992; Hirsch et al. 1982) using the Theil-Sen slope (Sen 1968; Theil 1992) was used with consideration of the seasonal nature of monthly evaporation data. Here, the Theil–Sen slope is a robust nonparametric method to fit a line to data points which is the median of the slopes of all paired points. When a large-sample (n) approximation is available (n>20), confidence intervals of the calculated Theil–Sen slope can be calculated as upper (U) and lower (L) limits using the following equations (Helsel; Hirsch 1992).

$$\mathrm{U} = \frac{N + Z_{\alpha/2}\sqrt{\frac{n(n-1)(2n+5)}{18}}}{2} + 1 \quad (2)$$

$$\mathrm{L} = \frac{N - Z_{\alpha/2}\sqrt{\frac{n(n-1)(2n+5)}{18}}}{2} \quad (3)$$

where, $n$ is the number of data points; $N$ is the number of pairwise slopes, i.e., $N = n(n-1)/2$; $Z_{\alpha/2}$ is the critical value having a p-value nearest to the half of a significance level (α), i.e., α/2, which can be obtained from a table of standard normal quantiles. These equations are used for presenting the confidence intervals of the trends in Table 2.The SK test accounts for seasonality by computing the MK test for each season, and then combining the results as

$$\mathrm{S}_k = \sum_{i=1}^{\mathrm{m}} \mathrm{S}_i \tag{4}$$

where $\mathrm{S}_i$ is the Kendall's S statistic for each season ($i$=1, …, m) and $\mathrm{S}_k$ is the overall statistic $\mathrm{S}_k$. In this study, we used each month throughout the study period as representing a season and therefore m is 12.

According to Hirsch et al. (1982), when the product of the number of seasons and number of years is more than approximately 25, the distribution of $\mathrm{S}_k$ can be approximated as a normal distribution. Here, $\mathrm{S}_k$ is standardized by subtracting its expectation $\mu_{\mathrm{S}_k} = 0$ and dividing by its standard deviation $\sigma_{\mathrm{S}_k}$. Then +1 or -1 is continuously added for the statistic to be closer to a smooth curve as per Equation (5). The estimated $\mathrm{Z}_{\mathrm{S}_k}$ now can be evaluated by a table representing a standard normal distribution.

$$Z_{S_k} = \begin{cases} \frac{S_k - 1}{\sigma_{S_k}} & \text{if } S_k > 0 \\ 0 & \text{if } S_k = 0 \\ \frac{S_k + 1}{\sigma_{S_k}} & \text{if } S_k < 0 \end{cases}$$

where (5)

$${\sigma_{S_k}}^2 = \sum_{i=1}^{m} \frac{n_i(n_i - 1)(2n_i + 5)}{18}$$

$n_i$=number of data in the season $i$

For the two-sided SK test, the null hypothesis ($H_0$) is rejected at a significance level (α) if $|Z_{S_k}| > Z_{crit}$ where $Z_{crit}$ is the critical value in the standard normal distribution at a probability of exceedance of α/2.

### 2.2.2 Data Concurrence Index

For assessing similitude across the datasets, trends of the eleven ET datasets at each grid cell were first calculated as described in 2.2.1. Then, the number of significant trends (NS) at a significance level (α) of 0.05, ranging from 0 to 11, was divided by number of datasets (i.e. eleven). In this way, zones with few or no significant trends can be easily identified, as well as zones where most of the datasets presents significant trends. However, to study these higher orders of concurrence, not only the NS is important, but also their direction, since opposing trends cannot be considered as consistent. Thus, the Data Concurrence Index (DCI) (Anabalón;

Sharma 2017) was also used which is an index to measure differences among results of the datasets and formulated as

$$\mathrm{DCI} = \frac{1}{\mathrm{ND}} \sum_{i=1}^{\mathrm{ND}} \frac{\mathrm{h}_i \times \mathrm{ST}_i}{\mathrm{abs}(\mathrm{ST}_i)} \tag{6}$$

where ND is the number of datasets with information, and $\mathrm{ST}_i$ is the magnitude of the trend, and $\mathrm{h}_i$ is a binary variable that indicate if there is a significant trend (1) or not (0) for dataset $i$ and abs(x) represents the absolute value of x. Therefore, a positive DCI indicates that there is a predominance of positive significant trends through the datasets, meanwhile a negative value represents a dominance of negative trends. A DCI equal to 0 could imply the lack of significant trends or inconsistent trends from different datasets. This index was calculated for each cell with information.

The DCI shows how these datasets behave as a group. By definition, the upper limit (in absolute terms) of the DCI is NS/ND, when all the significant trends found throughout the datasets in a specific location have the same sign. This implies that a difference between the value of NS/ND and the DCI in a specific location must be caused by inconsistencies on the directions of the detected trends. In addition, to investigate how the DCI distribution is temporally varied, a dynamic trend analysis with a 10-year moving window (120 months) was also carried over all the datasets over the 32-year study period (e.g., the first 10-year moving window includes monthly datasets from Jan 1981 to Dec 1990).

### 2.2.3 Pearson correlation, CSI and bias across trends of compared datasets

To deepen the analysis of how the spatially-distributed trends differ from each other, the Pearson correlation coefficient, the Critical Success Index (CSI), and bias between the trends of the datasets were calculated.

The Pearson correlation measures how much the datasets look alike in term of magnitudes of estimated trends. These correlation coefficients were calculated only in the common area of the two datasets analyzed, and they consider only the cells with significant trends in both of the compared datasets. For further comparisons, we also added Table S1 in Supplementary Material, presenting the Pearson correlation coefficients for all calculated trends regardless of their statistical significance.

We calculated the Critical Success Index (CSI) (Donaldson et al. 1975; Hogan; Mason 2011) to measure the similarity between the two datasets, A and B, formulated as

$$\mathrm{CSI} = \frac{a}{a + b + c} \tag{7}$$

where a is the number of significant trends with same sign presented in the same location on both datasets, b is the number of significant trends in the second dataset (B) which are not presented in or have the opposite sign with the first dataset (A), and c is the number of significant trends in the first dataset (A) which are not presented or have different sign in the second one (B). To illustrate the classification of a, b, and c, we presented a contingency table as Table S2 in Supplementary Material. Here, each dataset has four cases considered by statistical significance (i.e., significant or insignificant) and sign (+ or -) of the calculated trends, the contingency table is sized as 4×4.

CSI represents the number of significant trends present in the same location with the same sign on both datasets, divided by the total number of significant trends detected on each dataset. Consequently, the CSI measures the similitude of two datasets in terms of trend detection.

Similarly, the Bias shows if one dataset is generating more or less significant trends than the other (Donaldson et al. 1975; Hogan; Mason 2011), based on following formula:

$$\mathrm{Bias} = \frac{a+b}{a+c} \tag{8}$$

where the symbols are the same used for CSI. These indices were calculated only over the common coverage of the pairs of data analyzed. Therefore, the results for CSI are symmetrical, while the results of the bias are the inverse of their symmetrical counterpart, i.e., $a_{ij} = {a_{ji}}^{-1}$ for every $i$ and $j$ dataset.

# 3. Results and discussion

## 3.1 Consistency of trends in global mean ET

As a more general assessment of trends amongst the datasets, the monthly global mean ET time series were calculated and analyzed. The results are summarized in Table 2.

[Table 2 about here]

During the study period between 1981 and 2012, the trends of the datasets presented a wide range of differences in their values and signs. The two diagnostic datasets, CSIRO and GLEAM v3.3a, presented positive global trends during the period, while the four GLDAS-1 products and MERRA2 showed negative trends whereas VIC presented the most negative value among them. In contrast to the GLDAS-1 datasets, the two GLDAS datasets, CLSv2 and Noahv2, presented positive global trends and confidence intervals that are similar to those of the two diagnostic datasets. The two remaining reanalysis datasets, ERA-Interim and JRA55 displayed insignificant trends. The series of the mean ET across the datasets showed a negative significant trend, clearly influenced by the values of the GLDAS-1 group (Mosaic, VIC, CLM and Noah).

### 3.2 Spatial distribution of concurrent trends

Figure 1 shows the spatial distribution of the results obtained from the DCI analysis and Table 3 presents a summary of the DCI analysis.

[Figure 1 about here]

[Table 3 about here]

The first interesting result in Table 3 is that more than 40% of the land surface has a ratio of NS/ND higher than 0.6. This suggests that over a large part of the world, there is a high concurrence among the datasets indicating significant evaporation trends during the studied period. This phenomenon is globally well represented except for Canada/Alaska, Northern Eurasia, North Africa and Western Australia (Figure 1a). For example, as presented in Figure 1b (marked with two yellow squares), noteworthy is the presence of two defined zones in the world with high values of DCI, implying a high consistency in both the magnitude and the sign of significant trends. The first zone is around the region joining USA and Mexico, where the vast majority of the datasets show significant downward trends of the actual evaporation (i.e., negative DCI). The second zone is located in China, where the datasets coincide in significant increasing evaporation trends (i.e., positive DCI). Even where the DCI is not as high as in the two zones described above, a positive or negative value implies some concurrence of significant trends amongst the datasets. Generally speaking, as summarized in Table 3, the significant trends found across the datasets globally during the studied period resulted in a positive DCI covering 48.8% of the land area (excluding zero-DCI regions) and negative DCI covered 43.6%. The remaining 9.4% of the continental surface has a DCI of zero, which is larger than the area covered with no significant trends found in the datasets (i.e., NS/ND=0, 2.1%).

Positive DCI values are widely spread around the globe (Figure 1b), a pattern that is greater and more diverse than what was expected from Miralles et al. (2013). It seems that the datasets were more prone towards positive trends (DCI>0.5) over these 32 years, in East Europe, Western Sahara, South Africa, Tibetan Plateau, Central and Eastern China and Northern Eurasia. Meanwhile, zones in North and South America, Middle Africa/Arabian Peninsula and Northern China present negative DCI (<-0.5), indicating a general agreement between the datasets on decreasing trends.

High coincidences of significant trends throughout the datasets are scarce, covering only 11.3% of the surface area with NS/ND higher than 0.8 as shown in Table 3. Meanwhile, DCI over 0.6 is present over just 10.5% of the land area, mostly concentrated in the areas described above. On the other hand, the two lowest quintiles in Table 3 cover most of the world for DCI, with a value of 73.7%. This concentration of high DCI scores contrasted with the widespread low scores implying that some phenomena affecting the trends in these places are being modelled consistently by all the datasets, but for most of the world, the main drivers behind the evaporation trends are not congruently represented. The reduction of the area covered by the highest quintile of DCI and the increase of the two lowest quintiles is caused by inconsistent significant trends that lower the score. This problem, while present all over the globe, is more acute in some areas (NS/ND-abs(DCI)>0.5), as presented in Figure 1c. Northeastern United States, Central Amazonia, Middle Africa and Middle/East Asia are the most affected zones with contradictory trends throughout the datasets. The selection of a specific dataset to work in these zones should be done carefully considering that the results may be substantially sensitive to the selected dataset.

In the zones with high (negative or positive) DCI scores, the selection of a specific dataset is less problematic since the trends are consistent. However, as shown in Figure 2, differences appear when the magnitudes are compared. The box plots in Figure 2a and b present of magnitudes of trends over two 10°×10° tiles showing strong negative and positive DCI scores, respectively. As marked with two squares in Figure 1b, the first tile covers grid cells with strong negative DCI scores located at North America (30°N – 40°N, 100°W – 110°W), and the second tile is for positive DCI scores at China (25°N – 35°N, 105°E – 115°E).

[Figure 2 about here]

For the decreasing trends in Figure 2a, all datasets generally show negative trend values but with different distributions of trend magnitudes. It is noticeable that the four GLDAS-1 datasets show strong negative values with similar interquartile ranges to each other, the remaining datasets present different distributions of trend magnitudes, but the strong negative trends considerably ease in the two GLDAS-2 datasets by showing similar distribution shapes with the two diagnostic datasets, CSIRO and GLEAM. The three reanalysis datasets show different shapes of distributions. Especially, in the case of JRA55, it produces greater and more disperses trends than the others, followed by ERA-Interim and MERRA2. On contrary to the decreasing trend zone, as shown in in Figure 2b, the datasets produce considerably different patterns from each other over the increasing trend zone although they generally show positive trend values. In general terms, the diagnostic and reanalysis datasets have more conservative and less dispersed significant trends than the GLDAS-1 and -2 datasets. Especially GLDAS-1 produces greater and more dispersed trends than GLDAS-2. These results suggest that even when the directions are consistent, the selection of a dataset can still have a great impact on the results of a study conducted in the area.

The spatial distribution of the trends in the mean ET throughout the datasets (Figure 1d) shows some resemblance with the DCI: zones with positive (negative) DCI generally fit zones with positive (negative) trends, in the mean series. While the mid-latitude area, 30°N – 30°S, generally presents negative trends except for Arid zones like Sahara Desert and Arabian Peninsula, the high-latitude area, positive trends. The steepest ET trends are observed over Southern Africa/Tibetan Plateau with the positives (>0.3 mm/month/year) and Middle western Africa/Southeast Asia with the negatives (<-0.7 mm/month/year). Regions as Middle America, Eastern Africa and Eastern Australia also present strong negative trends, even when none of these places show high concurrence between the datasets (DCI > 80%). In fact, the zones with high DCI scores do not show particularly great trends in the mean series. Therefore, high trends in the ET mean series do not necessarily represent high agreement amongst the datasets in terms of trends direction or detection and vice versa.

### 3.3 Correlation, CSI and bias amongst datasets

As described in Section 2.2.3, we calculated the Pearson correlation, CSI and bias for the datasets in Tables 4, 5 and 6, respectively.

[Table 4 about here]

The Pearson correlation between the trends of the datasets is generally low except for the GLDAS datasets that belong to the same data version. The main reasons behind these high correlations are the shared forcing variables and methods used. It is interesting to note the moderate correlations between JRA55 and ERA-Interim (i.e., 0.439). The rest of the correlations are low, implying that there is no substantial relationship between the location of the trends and their value across the datasets. Table S1 in Supplementary Material shows the Pearson correlation for all trends regardless of their statistical significance. Here, the pattern of the values in Table S1 is similar to that of Table 4, but their average value is about 0.05 lesser than that of Table 4.

Even if the Pearson correlation gives a good idea about how much the significant trends on two datasets are alike in terms of magnitude, it does not clearly reflect that trends with different signs separated by a constant value represent opposite phenomena. This aspect is better represented using the CSI and Bias results in Tables 5 and 6. It should, however, be noted that these results of CSI and Bias are influenced by their various spatial resolutions and the resampling method that we applied. Given that the limitations are, an indication of the similarities and differences across the datasets on a regional basis is provided in the results as follows.

[Table 5 about here]

[Table 6 about here]

Looking at the bias in Table 6, the datasets that generally produced more significant trends (i.e., mean bias>100) are GLEAM, GLDAS-1 except for CLM, GLDAS-2 CLS, and ERA-Interim. The biases of the trends for CSIRO, GLDAS-1 CLM, and MERRA2 datasets were relatively lower (i.e., mean bias<100). Especially, the trends of GLDAS-1 MOSAIC and MERRA2 are contrasted with their notably high and low bias values, respectively. Lastly, GLDAS-2 NOAH and JRA55, in particular, were less biased amongst their pairs (i.e., mean bias≈100). The bias also echoes on the CSI scores (Table 5), lowering it in those comparisons when the bias was more pronounced. This is particularly visible in the scores obtained between the GLDAS datasets in each version, where the number of significant trends between the datasets is the most relevant reason behind the CSI scores. In other cases, the main driver is the difference in the directions of the trends.

### 3.4 Variation of trends with the time period considered

Figure 3a presents the distribution in time of DCI over the land surface, showing that high concurrence (DCI >0.8 or <-0.8) among the datasets is rare throughout the entire available period. The major changes in coverage happen in the range between 0.2 and 0.6, for both positive and negative DCI. Probably this is an effect of the correlation between the GLDAS datasets, changing at the same time in relative discordance with the remaining datasets. Also interesting is that the global DCI is getting negative by 1997 but the pattern rebounds since that time, positive DCI increases till the end of the period.

[Figure 3 about here]

The general shape of the series in Figure 3a shows a very dynamic DCI distribution. These changes in distribution make the selection of the time period critical to the analysis of evaporation trends. To illustrate where these variations are more relevant, the total number of changes of sign on each cell was calculated and presented in Figure 3b. In there, it can be seen that certain points of the world are more prone to present important variations in the direction of the trends, as Southern Africa, India, North America and Europe. Studies of evaporation trends in these regions should be particularly cautious regarding the studied time period. We further discuss this point in the next section.

### 3.5 How to interpret and use DCI

For investigating the dynamics of DCI, we noted the refraction of the ET trend between two periods: 1982 to 1997 and 1998 to 2008 (Jung et al. 2010). The change in DCI toward both positive and negative directions in the period of 1998–2007 (Period 2) in Figure 3a seems to correspond to the cessation in the ET increase from 1998 to 2008 as reported in Jung et al. (2010). However, the strong negative DCI in the period of 1988–1997 (Period 1) is contradictory to the increasing global ET trend from 1982 to 1997. For further examining this contradiction and similarity, we investigated the distributions of ET trends over three 10°×10° tiles as presented in Figure 4. Those represent Russia (60°N – 70°N, 60°E – 70°E), South Africa (17°S – 27°S, 10°W – 20°W) and Western Australia (20°S – 30°S, 120°E – 130°E) showing contrasting changes in the ET trends, also highlighted by Jung et al. (2010).

[Figure 4 about here]

The negative tendency of DCI in Period 1 is well depicted in space as shown in Figure 4a, where the world shows a negative DCI in general except for some regions such as Northwest America, Arabia peninsular and Western Australia. This overall negative DCI is contradicting the increasing ET trend over the period of 1982 to 1997 found in Jung et al. (2010) while the spatial pattern of DCI over Period 2 (Figure 4d) is relatively well matched to that of Jung et al. (2010) over the period of 1998 to 2008. When it comes to the DCI distributions over the three regions, the contradiction in Period 1 is well presented. Jung et al. (2010) reported that the trend change in Russia is not considerable. However, as shown in Figure 4b and f, the GLDAS, especially GLDAS-1, datasets present a clear trend reversal from negative to positive while the diagnostic and reanalysis datasets are holding small trend changes. On the other hand, the reported trend change over South Africa (negative to positive) and Western Australia (positive to negative) is clearly reflected through all datasets as presented in Figure 4c and g; Figure 4d and h.

As a possible reason for the discrepancy in the ET trend, we noticed that the temporal pattern of DCI (the bottom panel of Figure 3a) is well correlated with the spatial variation of precipitation (in the right y-axis of the top panel of Figure 3a) along with the moving windows. For this, we used Global Precipitation Climatology Project (GPCP) version2.3 monthly precipitation data (Pendergrass et al. 2020) over the study period, from January 1981 to December 2012. Here, the spatial variation of precipitation is represented by the coefficient of variation ($CV_P$) of temporally averaged monthly precipitation data in each 10-year period. In addition to this, Figure 5 shows temporal changes of the ET trend-precipitation correlation in space ($r_{Trend-mean\,P}$) for the 11 ET datasets. Here, for simplicity, the 11 datasets were aggregated into four groups: Diagnostic (2); GLDASv1 (4); GLDASv2 (2); Reanalysis (2), and then the temporal trend lines and confidence intervals were simply represented by medians and standard errors, respectively, at each time step.

As presented in Figure S1 in Supplementary Material, these changes are separately plotted for two regions with weak (abs(DCI)<0.3) and strong (abs(DCI)>0.7) DCI scores referred from Figure 1b.

[Figure 5 about here]

Over the weak DCI regions (Figure 5a and b), interestingly, strong inverse behaviors of $r_{Trend-mean\,P}$ for the four GLDAS-1 datasets compared to other datasets are observed, especially for the CSIRO, GLDAS-2 and JRA55 datasets. This tendency exhibits the greatest contrast during Period 1 (1988–1997) which showed the ET trend contradiction in Figure 4a and tends to be somewhat eased in Period 2 (1998–2007). In case of the strong DCI regions (Figure 5c and d), these inverse behaviors reverse or subside over the strong DCI regions in both periods (Figure 5c and d). Interestingly, the high DCI tends to appear over the supply limited regions identified in Jung et al. (2010) where ET is more positively correlated with precipitation than temperature, but the low DCI is likely to be distributed over the demand limited regions in which ET is more positively correlated with temperature than precipitation. Given that the tendency of the four GLDAS-1 datasets seems an important reason for the contradiction noted in Period 1, it can be inferred that the performance of GLDAS-1 over the demand limited regions tends to be sensitively affected.

Consistent with this, DCI locally agrees with the real phenomena or not depending on the spatiotemporal behaviors of the datasets considered. Therefore, it should be noted that DCI is not to be used for indicating how the datasets are close to the truth but how the datasets are different from each other, especially when the number of datasets used is not statistically sufficient.

## 4. Summary and Conclusions

The present work analyzed significant trends in monthly evaporation from eleven global datasets, during their common period, between 1981 and 2012. In this period, the global ET trends of the mean throughout the datasets decreased with a pace of -0.072 mm/month/year, and the datasets showed a wide range of differences in magnitudes and directions their global trends. While CSIRO and GLEAM v3.3a showed positive trends during the period, the four GLDAS products and MERRA2 presented negative trends and ERA-Interim and JRA55 displayed insignificant trends. The scales of the global trends throughout the datasets were considerably different. The trends found in the GLDAS group were substantially greater than the rest, while the remaining datasets presented smaller magnitudes.

Only 2.1% of the world presented a high coincidence of no significant trends. For around half of the world, there was a general disagreement amongst the trends of the datasets in the presence and location of significant trends, represented by low to moderate DCI scores (0.2 – 0.6). The datasets tend to be positive trends (DCI>0.5) over East Europe, Western Sahara, South Africa, Tibetan Plateau, China and Northern Eurasia. Over North and South America, Middle Africa/Arabian Peninsula and Northern China, negative trends, where DCI <-0.5, were observed. Also, it was found that the magnitudes of the trends on the series composed by the mean ET across the datasets, were not representative of the agreement between them in term of direction and detection of trends. The significant trends between the datasets were weakly correlated, except for those belonging to the same group, as GLDAS. However, even when highly correlated, their general behavior was considerably different, as the low CSI scores demonstrate. The main reasons behind these differences are the tendency to present more or less significant trends (bias) and the presence of contradictory trends in a specific location. To study the similarities of the trends in the datasets through time, a dynamic trend analysis was performed, using a 10-year window moving over the datasets. The results show a very dynamic system, with important changes over time. However, high concurrences are rare in every point in time, and the most significant variations can be attributed to the correlated GLDAS datasets, changing simultaneously in relative discrepancy with the rest. Some places like Southern Africa, India, North America and Europe are more prone to revert the direction of the significant trends in many datasets with time.

The results of this work show how different approaches to estimate the evaporation can lead to important inconsistencies in the location and direction of its trends. These inconsistencies are a big concern for climate change studies that require accuracy to evaluate models or estimate actual and future impacts. Therefore, it is important to address these discrepancies and generate more reliable global datasets with validated trends, to improve the performance of our models and provide better information for decision makers. Studies in regions with highly inconsistent trends should be especially careful when selecting an evaporation dataset to work with and be aware that its results may be greatly sensitive to the dataset selected. The same can be said of the places with tendency to revert the direction of their trends. In these locations, the selected time period could also change the results of a trend study drastically. Finally, no single dataset can be used as a general representation of the rest to analyze trends in evaporation. Even within the groups of datasets that use similar methods to estimate ET, the differences are big enough to recommend the use of several datasets instead of just one as a representative choice.

## Acknowledgements

This work has been supported with a Linkage Project (LP160100620) funded by the Australian Research Council, WaterNSW, and Sydney Water. The authors of this work want to express their gratitude to the authors of the datasets used here for providing such important information.

# References

Anabalón, A., and A. Sharma, 2017: On the divergence of potential and actual evapotranspiration trends: An assessment across alternate global datasets. *Earth's Future*, **5,** 905-917.

Baldocchi, D., and Coauthors, 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,** 2415-2434.

Balsamo, G., A. Beljaars, K. Scipal, P. Viterbo, B. v. d. Hurk, M. Hirschi, and A. K. Betts, 2009: A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System. *J Hydrometeorol*, **10,** 623-643.

Brodzik, M. J., B. Billingsley, T. Haran, B. Raup, and M. H. Savoie, 2012: EASE-Grid 2.0: Incremental but significant improvements for Earth-gridded data sets. *ISPRS International Journal of Geo-Information*, **1,** 32-45.

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.

Chen, Y., and Coauthors, 2014: Comparison of satellite-based evapotranspiration models over terrestrial ecosystems in China. *REMOTE SENS ENVIRON*, **140,** 279-293.

Dai, Y., and Coauthors, 2003: The Common Land Model. *Bulletin of the American Meteorological Society*, **84,** 1013-1024.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137,** 553-597.

Dickinson, E., A. Henderson-Sellers, and J. Kennedy, 1993: Biosphere-atmosphere transfer scheme (BATS) version 1e as coupled to the NCAR community climate model.

Donaldson, R., R. M. Dyer, and M. J. Kraus, 1975: Objective evaluator of techniques for predicting severe weather events. *Bulletin of the American Meteorological Society*, AMER METEOROLOGICAL SOC 45 BEACON ST, BOSTON, MA 02108-3693, 755-755.

Ershadi, A., M. McCabe, J. P. Evans, N. W. Chaney, and E. F. Wood, 2014: Multi-site evaluation of terrestrial evaporation models using FLUXNET data. *Agricultural and Forest Meteorology*, **187,** 46-61.

Fisher, J. B., K. P. Tu, and D. D. Baldocchi, 2008: Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. *REMOTE SENS ENVIRON*, **112,** 901-919.

Fitsum, W., and S. Ashish, 2016: Should flood regimes change in a warming climate? The role of antecedent moisture conditions. *Geophys Res Lett*, **43,** 7556-7563.

Gelaro, R., and Coauthors, 2017: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *J Clim*, **30,** 5419-5454.

Helsel, D. R., and R. M. Hirsch, 1992: *Statistical methods in water resources.* Vol. 49, Elsevier.

Hettiarachchi, S., C. Wasko, and A. Sharma, 2019: Can antecedent moisture conditions modulate the increase in flood risk due to climate change in urban catchments? *J Hydrol*, **571,** 11-20.

Hirsch, R. M., J. R. Slack, and R. A. Smith, 1982: Techniques of trend analysis for monthly water quality data. *Water Resour Res*, **18,** 107-121.

Hogan, R. J., and I. B. Mason, 2011: Deterministic Forecasts of Binary Events. *Forecast Verification*, 2nd ed., I. T. Jolliffe, and D. B. Stephenson, Eds., John Wiley & Sons, Ltd, 31-59.

Huizhi, L., and F. Jianwu, 2012: Seasonal and Interannual Variations of Evapotranspiration and Energy Exchange over Different Land Surfaces in a Semiarid Area of China. *Journal of Applied Meteorology and Climatology*, **51,** 1875-1888.

IPCC, 2014: *Climate Change 2013: The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press.

Jiménez, C., and Coauthors, 2011: Global intercomparison of 12 land surface heat flux estimates. *Journal of Geophysical Research: Atmospheres*, **116**.

Johnson, F., and A. Sharma, 2010: A Comparison of Australian Open Water Body Evaporation Trends for Current and Future Climates Estimated from Class A Evaporation Pans and General Circulation Models. *J Hydrometeorol*, **11,** 105-121.

Jung, M., and Coauthors, 2010: Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, **467,** 951.

Kendall, M. G., 1948: *Rank correlation methods.* Griffin.

Kim, S., and A. Sharma, 2019: The Role of Floodplain Topography in Deriving Basin Discharge Using Passive Microwave Remote Sensing. *Water Resour Res*, **55,** 1707-1716.

Kiptala, J. K., Y. Mohamed, M. L. Mul, and P. Van der Zaag, 2013: Mapping evapotranspiration trends using MODIS and SEBAL model in a data scarce and heterogeneous landscape in Eastern Africa. *Water Resour Res*, **49,** 8495-8510.

Kobayashi, S., and Coauthors, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, **93,** 5-48.

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.

Koster, R. D., and M. J. Suarez, 1996: Energy and Water Balance Calculations in the Mosaic LSM.

Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, 2000: A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure. *Journal of Geophysical Research: Atmospheres*, **105,** 24809-24822.

Koster, R. D., and Coauthors, 2004: Regions of Strong Coupling Between Soil Moisture and Precipitation. *Science*, **305,** 1138-1140.

Liang, X., D. P. Lettenmaier, and E. F. Wood, 1996: One-dimensional statistical dynamic representation of subgrid spatial variability of precipitation in the two-layer variable infiltration capacity model. *Journal of Geophysical Research: Atmospheres*, **101,** 21403-21422.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research: Atmospheres*, **99,** 14415-14428.

Libertino, A., A. Sharma, V. Lakshmi, and P. Claps, 2016: A global assessment of the timing of extreme rainfall from TRMM and GPM for improving hydrologic design. *Environmental Research Letters*, **11,** 054003.

Mann, H. B., 1945: Nonparametric Tests Against Trend. *Econometrica*, **13,** 245-259.

Marshall, M., C. Funk, and J. Michaelsen, 2012: Examining evapotranspiration trends in Africa. *Clim Dynam*, **38,** 1849-1865.

Martens, B., and Coauthors, 2017: GLEAM v3: satellite-based land evaporation and root-zone soil moisture. *Geosci. Model Dev.*, **10,** 1903-1925.

McMahon, T. A., M. C. Peel, L. Lowe, R. Srikanthan, and T. R. McVicar, 2013: Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: a pragmatic synthesis. *Hydrol. Earth Syst. Sci.*, **17,** 1331-1363.

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.

Miralles, D. G., and Coauthors, 2013: El Niño–La Niña cycle and recent trends in continental evaporation. *Nature Climate Change*, **4,** 122.

Miralles, D. G., and Coauthors, 2016: The WACMOS-ET project – Part 2: Evaluation of global terrestrial evaporation data sets. *Hydrol. Earth Syst. Sci.*, **20,** 823-842.

Mu, Q., F. A. Heinsch, M. Zhao, and S. W. Running, 2007: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *REMOTE SENS ENVIRON*, **111,** 519-536.

Mueller, B., and Coauthors, 2013: Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis. *Hydrol. Earth Syst. Sci.*, **17,** 3707-3720.

Mueller, B., and Coauthors, 2011: Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. *Geophys Res Lett*, **38**.

Pendergrass, A., J.-J. Wang, and National Center for Atmospheric Research Staff (Eds), 2020: The Climate Data Guide: GPCP (Monthly): Global Precipitation Climatology Project. v2.3 ed.

Philip, J. R., 1957: EVAPORATION, AND MOISTURE AND HEAT FIELDS IN THE SOIL. *Journal of Meteorology*, **14,** 354-366.

Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's modern-era retrospective analysis for research and applications. *J Clim*, **24,** 3624-3648.

Rodell, M., H. K. Beaudoing, and NASA/GSFC/HSL, 2007a: GLDAS VIC Land Surface Model L4 Monthly 1.0 x 1.0 degree V001. G. E. S. D. a. I. S. C. G. DISC), Ed.

——, 2007b: GLDAS Noah Land Surface Model L4 Monthly 1.0 x 1.0 degree V001. G. E. S. D. a. I. S. C. G. DISC), Ed.

——, 2007c: GLDAS CLM Land Surface Model L4 Monthly 1.0 x 1.0 degree V001. G. E. S. D. a. I. S. C. G. DISC), Ed.

——, 2007d: GLDAS Mosaic Land Surface Model L4 Monthly 1.0 x 1.0 degree V001. G. E. S. D. a. I. S. C. G. DISC), Ed.

Rodell, M., and Coauthors, 2004: The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*, **85,** 381-394.

Rui, H., and H. Beaudoing, 2018: README Document for NASA GLDAS Version 2 Data Products. *Goddart Earth Sciences Data and Information Services Center (GES DISC): Greenbelt, MD, USA.*

Sen, P. K., 1968: Estimates of the Regression Coefficient Based on Kendall's Tau. *Journal of the American Statistical Association*, **63,** 1379-1389.

Seneviratne, S. I., and Coauthors, 2010: Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews,* **99,** 125-161.

Sheffield, J., G. Goteti, and E. F. Wood, 2006: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J Clim*, **19,** 3088-3111.

Stephens, C. M., T. R. McVicar, F. M. Johnson, and L. A. Marshall, 2018: Revisiting Pan Evaporation Trends in Australia a Decade on. *Geophys Res Lett*, **45,** 11,164-111,172.

Theil, H., 1992: A Rank-Invariant Method of Linear and Polynomial Regression Analysis. *Henri Theil's Contributions to Economics and Econometrics: Econometric Theory and Methodology*, B. Raj, and J. Koerts, Eds., Springer Netherlands, 345-381.

Trambauer, P., E. Dutra, S. Maskey, M. Werner, F. Pappenberger, L. P. H. van Beek, and S. Uhlenbrook, 2014: Comparison of different evaporation estimates over the African continent. *Hydrol. Earth Syst. Sci.*, **18,** 193-212.

Trenberth, K. E., J. T. Fasullo, and J. Kiehl, 2009: Earth's Global Energy Budget. *Bulletin of the American Meteorological Society*, **90,** 311-324.

Trenberth, K. E., L. Smith, T. Qian, A. Dai, and J. Fasullo, 2007: Estimates of the Global Water Budget and Its Annual Cycle Using Observational and Model Data. *J Hydrometeorol*, **8,** 758-769.

Vinukollu, R. K., R. Meynadier, J. Sheffield, and E. F. Wood, 2011: Multi-model, multi-sensor estimates of global evapotranspiration: climatology, uncertainties and trends. *Hydrological Processes*, **25,** 3993-4010.

Vourlitis, G. L., N. P. Filho, M. M. S. Hayashi, J. de S. Nogueira, F. T. Caseiro, and J. H. Campelo Jr., 2002: Seasonal variations in the evapotranspiration of a transitional tropical forest of Mato Grosso, Brazil. *Water Resour Res,* **38,** 30-31-30-11.

Yang, Y., D. Long, and S. Shang, 2013: Remote estimation of terrestrial evapotranspiration without using meteorological data. *Geophys Res Lett*, **40,** 3026-3030.

Zhang, K., J. S. Kimball, and S. W. Running, 2016a: A review of remote sensing based actual evapotranspiration estimation. *Wiley Interdisciplinary Reviews: Water*, **3,** 834-853.

Zhang, R., S. Kim, and A. Sharma, 2019: A comprehensive validation of the SMAP Enhanced Level-3 Soil Moisture product using ground measurements over varied climates and landscapes. *REMOTE SENS ENVIRON*, **223,** 82-94.

Zhang, Y., R. Leuning, L. B. Hutley, J. Beringer, I. McHugh, and J. P. Walker, 2010: Using long-term water balances to parameterize surface conductances and calculate evaporation at 0.05° spatial resolution. *Water Resour Res,* **46**.

Zhang, Y., J. Pena Arancibia, T. McVicar, F. Chiew, J. Vaze, H. Zheng, and Y. P. Wang, 2016b: Monthly global observation-driven Penman-Monteith-Leuning (PML) evapotranspiration and components. v2., v4 ed., D. C. Commonwealth Scientific and Industrial Research Organisation (CSIRO), Ed.

Zhang, Y., and Coauthors, 2016c: Multi-decadal trends in global terrestrial evapotranspiration and its components. *Sci Rep*, **6,** 19124.

# Tables and Figures

Table 1. Global ET datasets used in the study and their main characteristics of which data links are presented below the table. The temporal resolutions are the ones used in this study and finer resolutions are available for some datasets. The coverage indicates the extension in latitude with evaporation data. Acronyms used in this table are given in Section 2.1.

| **Category** | **Dataset** | **Reference** | **Temp./Spa. Resolution** | **Temporal coverage** | **Spatial Coverage (Lat)** |
|---|---|---|---|---|---|
| **Diagnostic (2)** | CSIRO 3) | Zhang et al. (2016c) | Monthly/0.5° | Jan 1981-Dec 2012 | 90°N - 60°S |
| | GLEAM v3.3a 2) | Martens et al. (2017) | Daily/0.25° | Jan 1980-Sep 2018 | 90°N - 90°S |
| **Land-surface model (6)** | GLDAS Mosaic 4) | Koster; Suarez (1996) | Monthly/1° | Jan 1979-June 2020 | 90°N - 60°S |
| | GLDAS VIC 4) | Liang et al. (1996);<br>Liang et al. (1994) | Monthly/1° | | |
| | GLDAS CLM 4) | Rodell et al. (2004) | Monthly/1° | | |
| | GLDAS Noah 4) | Rodell et al. (2004) | Monthly/1° | | |
| | GLDAS CLSv2 4) | Rodell et al. (2004) | Daily/0.25° | Jan 1948-Dec 2014 | |
| | GLDAS Noahv2 4) | Rodell et al. (2004) | Monthly/0.25° | | |
| **Reanalysis (3)** | ERA-Interim 5) | Dee et al. (2011);<br>Balsamo et al. (2009);<br>Trambauer et al. (2014) | Monthly/0.75° | Jan 1979-Aug 2019 | 90°N - 90°S |
| | MERRA2 2) | Gelaro et al. (2017) | Monthly /0.5°×0.625° | Jan 1980-Present | 90°N - 90°S |

| | | | | | |
|---|---|---|---|---|---|
| | JRA55 6) | Kobayashi et al. (2015) | Monthly /1.25°×1.25° | Jan 1958- Present | 90°N - 90°S |

1) MOD16: http://files.ntsg.umt.edu/data/NTSG_Products/MOD16/, 2) GLEAM: https://www.gleam.eu/, 3) CSIRO: https://data.csiro.au/dap/landingpage?pid=csiro%3A17375, 4) GLDAS and MERRA2: https://earthdata.nasa.gov/, 5) ERA-Interim: https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim 6)JRA55: https://jra.kishou.go.jp/JRA-55/index_en.html

Table 2. ET global trends by dataset. The trends were estimated using the series of monthly global evaporation on each dataset.

| **Dataset** | **Monthly Mean (mm/month)** | **Trend (mm/month/year)** | **Confidence Interval (α=0.05) (mm/month/year)** | **Significant** |
|---|---|---|---|---|
| **CSIRO** | 41.6 | 0.028 | 0.019 – 0.038 | Yes |
| **GLEAM v3.3a** | 39.3 | 0.067 | 0.058 – 0.075 | Yes |
| **GLDAS Mosaic** | 49.6 | -0.187 | -0.220 – -0.151 | Yes |
| **GLDAS VIC** | 47.6 | -0.306 | -0.344 – -0.272 | Yes |
| **GLDAS CLM** | 38.0 | -0.195 | -0.221 – -0.167 | Yes |
| **GLDAS Noah** | 43.3 | -0.187 | -0.212 – -0.158 | Yes |
| **GLDAS CLSv2** | 49.4 | 0.044 | 0.033 – 0.055 | Yes |
| **GLDAS Noahv2** | 41.9 | 0.036 | 0.027 – 0.044 | Yes |
| **ERA-Interim** | 45.1 | 0.001 | -0.008 – 0.011 | No |
| **MERRA2** | 50.7 | -0.056 | -0.074 – -0.038 | Yes |
| **JRA55** | 43.4 | 0.001 | -0.009 – 0.011 | No |
| **Mean** | 44.5 | -0.072 | -0.086 – -0.058 | Yes |

Table 3. Summary of DCI analysis by land area percentages in range classes. Where, NS=number of significant trends; ND=number of datasets; DCI= Data Concurrence Index.

| **Range class (absolute value)** | **Area covered by each metric in each range class (%)** | | | |
|---|---|---|---|---|
| | NS/ND | Positive DCI | Negative DCI | Total DCI |
| **0.8 – 1.0** | 11.3 | 1.3 | 1.4 | 2.8 |
| **0.6 – 0.8** | 32.0 | 3.9 | 3.8 | 7.8 |
| **0.4 – 0.6** | 29.4 | 7.1 | 8.6 | 15.8 |
| **0.2 – 0.4** | 13.7 | 11.1 | 13.3 | 24.4 |
| **0 – 0.2** | 11.5 | 25.3 | 14.6 | 39.9 |
| **0** | 2.1 | 9.4 | 0.0 | 9.4 |
| **Sum** | 100.0 | 58.3 | 41.7 | 100.0 |

Table 4. Comparison of Pearson correlation coefficients between significant trends of the datasets over their common areas. Where, CS: CSIRO; GL: GLEAM v3.3a; MOS: GLDAS Mosaic; VIC: GLDAS VIC; CLM: GLDAS CLM; NOAH: GLDAS Noah; CLv2: GLDAS CLSv2; NOv2: GLDAS Noahv2; EI: ERA-Interim; ME: MERRA2; JRA: JRA55.

| **Data** | **CS** | **GL** | **MOS** | **VIC** | **CLM** | **NOAH** | **CLv2** | **NOv2** | **EI** | **ME** | **JRA** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **CS** | 1.000 | – | – | – | – | – | – | – | – | – | – |
| **GL** | 0.218 | 1.000 | – | – | – | – | – | – | – | – | – |
| **MOS** | 0.284 | 0.125 | 1.000 | – | – | – | – | – | – | – | – |
| **VIC** | 0.210 | 0.064 | 0.839 | 1.000 | – | – | – | – | – | – | – |
| **CLM** | 0.217 | 0.139 | 0.859 | 0.847 | 1.000 | – | – | – | – | – | – |
| **NOAH** | 0.245 | 0.114 | 0.894 | 0.875 | 0.871 | 1.000 | – | – | – | – | – |
| **CLv2** | 0.596 | 0.259 | 0.139 | 0.250 | 0.163 | 0.257 | 1.000 | – | – | – | – |
| **NOv2** | 0.718 | 0.326 | 0.316 | 0.250 | 0.238 | 0.327 | 0.850 | 1.000 | – | – | – |
| **EI** | 0.133 | 0.329 | 0.105 | 0.070 | 0.059 | 0.185 | 0.266 | 0.275 | 1.000 | – | – |
| **ME** | 0.302 | 0.167 | 0.138 | 0.242 | 0.165 | 0.178 | 0.218 | 0.228 | 0.150 | 1.000 | – |
| **JRA** | 0.096 | 0.352 | 0.089 | 0.156 | 0.091 | 0.211 | 0.298 | 0.273 | 0.439 | 0.189 | 1.000 |

Table 5. Comparison of CSI (%) between significant trends of the datasets over their common areas. Higher values of CSI imply similar higher similitudes in the number of the significant trends, their location and direction.

| **Data** | **CS** | **GL** | **MOS** | **VIC** | **CLM** | **NOAH** | **CLv2** | **NOv2** | **EI** | **ME** | **JRA** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **CS** | 100.00 | – | – | – | – | – | – | – | – | – | – |
| **GL** | 33.65 | 100.00 | – | – | – | – | – | – | – | – | – |
| **MOS** | 27.77 | 25.62 | 100.00 | – | – | – | – | – | – | – | – |
| **VIC** | 23.22 | 17.46 | 62.69 | 100.00 | – | – | – | – | – | – | – |
| **CLM** | 23.72 | 19.48 | 65.47 | 69.58 | 100.00 | – | – | – | – | – | – |
| **NOAH** | 25.59 | 21.45 | 68.11 | 69.74 | 69.94 | 100.00 | – | – | – | – | – |
| **CLv2** | 37.00 | 34.30 | 28.04 | 24.50 | 24.93 | 27.84 | 100.00 | – | – | – | – |
| **NOv2** | 42.46 | 35.69 | 25.83 | 20.82 | 21.52 | 25.43 | 62.28 | 100.00 | – | – | – |
| **EI** | 28.68 | 29.06 | 33.47 | 29.86 | 28.28 | 31.68 | 29.24 | 27.59 | 100.00 | – | – |
| **ME** | 26.26 | 28.34 | 31.44 | 31.13 | 31.13 | 30.92 | 33.28 | 27.87 | 27.17 | 100.00 | – |
| **JRA** | 32.62 | 33.06 | 30.55 | 26.05 | 27.74 | 29.32 | 34.45 | 32.91 | 32.18 | 31.61 | 100.00 |

Table 6. Comparison of Significant trends with focus on bias (%) between datasets. Biases lower than 100% indicate that the datasets in the columns present less significant trends than the ones in the rows, while values greater than 100% imply the contrary.

| **Data** | **CS** | **GL** | **MOS** | **VIC** | **CLM** | **NOAH** | **CLv2** | **NOv2** | **EI** | **ME** | **JRA** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **CS** | 100.00 | 109.69 | 112.56 | 101.45 | 95.92 | 106.02 | 109.16 | 104.74 | 108.60 | 89.01 | 104.69 |
| **GL** | 91.17 | 100.00 | 107.26 | 90.56 | 89.60 | 99.57 | 98.05 | 95.38 | 105.73 | 82.87 | 99.80 |
| **MOS** | 88.84 | 93.23 | 100.00 | 92.99 | 89.55 | 96.96 | 91.57 | 87.36 | 94.06 | 78.13 | 90.71 |
| **VIC** | 98.57 | 110.42 | 107.54 | 100.00 | 94.27 | 100.85 | 101.97 | 96.45 | 99.83 | 84.21 | 98.47 |
| **CLM** | 104.26 | 111.61 | 111.67 | 106.08 | 100.00 | 105.51 | 109.30 | 103.67 | 110.75 | 90.75 | 107.75 |
| **NOAH** | 94.32 | 100.43 | 103.14 | 99.15 | 94.78 | 100.00 | 97.79 | 92.93 | 98.75 | 81.00 | 95.40 |
| **CLv2** | 91.61 | 101.99 | 109.21 | 98.07 | 91.49 | 102.26 | 100.00 | 95.24 | 101.67 | 81.37 | 96.38 |
| **NOv2** | 95.47 | 104.85 | 114.46 | 103.68 | 96.46 | 107.61 | 105.00 | 100.00 | 105.41 | 83.17 | 99.16 |
| **EI** | 92.08 | 94.58 | 106.31 | 100.17 | 90.29 | 101.27 | 98.36 | 94.87 | 100.00 | 80.74 | 95.92 |
| **ME** | 112.35 | 120.67 | 127.99 | 118.75 | 110.19 | 123.46 | 122.90 | 120.24 | 123.85 | 100.00 | 114.54 |
| **JRA** | 95.52 | 100.20 | 110.24 | 101.55 | 92.81 | 104.82 | 103.76 | 100.85 | 104.25 | 87.31 | 100.00 |

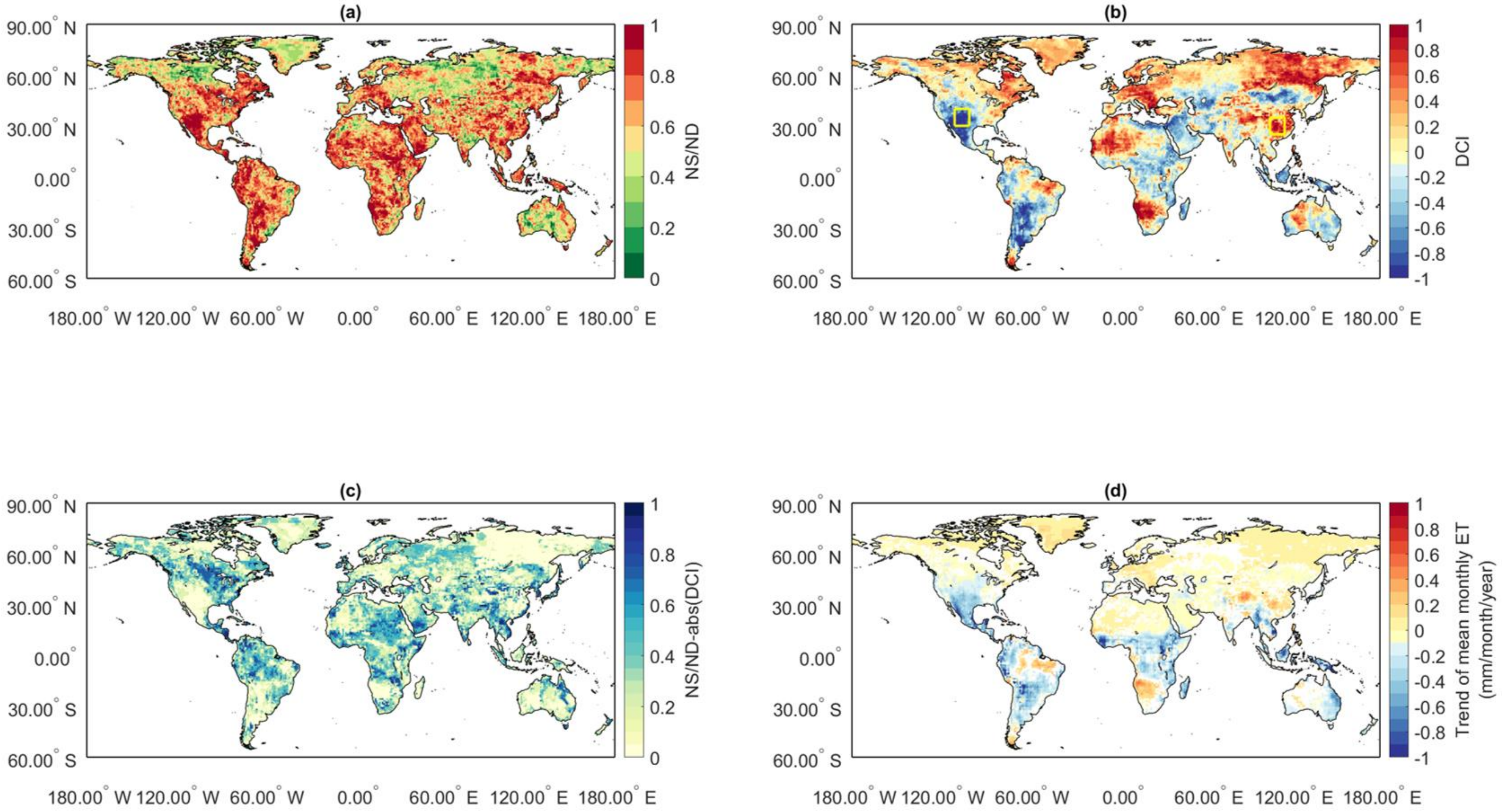


Figure 1. Spatial distributions of (a) NS/ND; (b) DCI; (c) NS/ND minus DCI; (d) estimated trends of mean monthly ET throughout eleven datasets

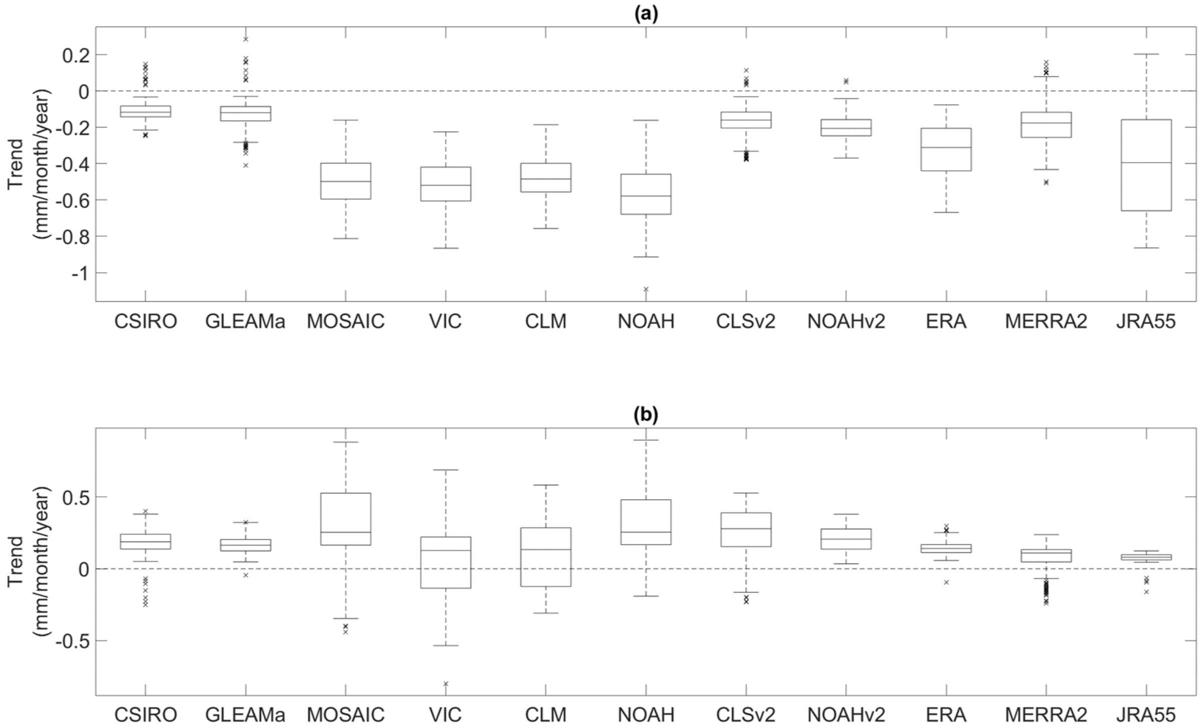


Figure 2. Boxplots of estimated trends across datasets in (a) a decreasing (North America; 30°N – 40°N, 100°W – 110°W) and (b) an increasing trend zone (China; 25°N – 35°N, 105°E – 115°E), of which locations are marked as yellow squares in Figure 1b.

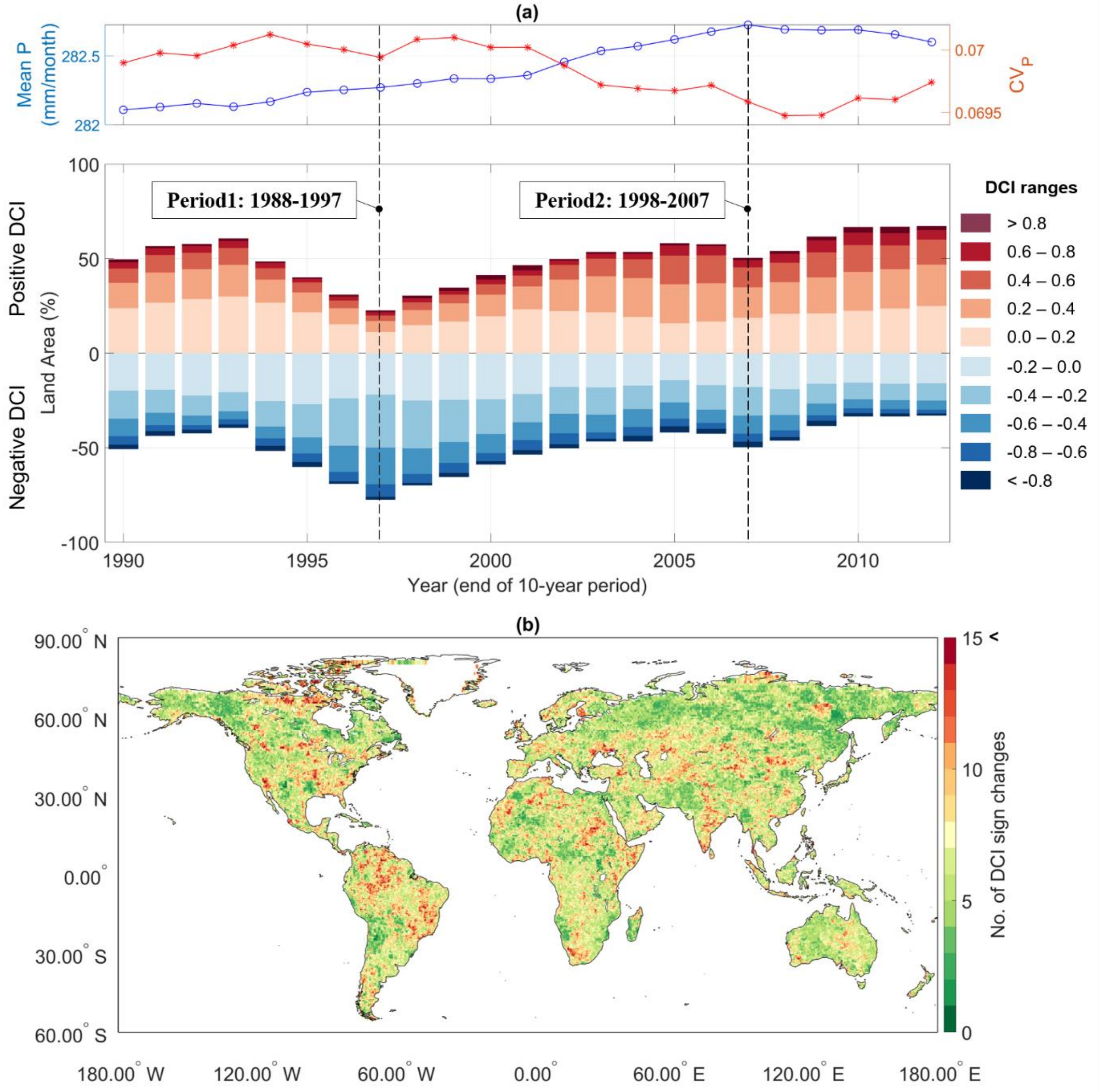


Figure 3. Changes of DCI in time using a dynamic trend analysis with a 10-year moving window. (a) Top panel: time series of mean precipitation (mean P: mm/month) and coefficient of variation of precipitation ($CV_P$); bottom panel: time series of the land covered by percentile of DCI, differentiating positive and negatives values of DCI. The x-axis represents the end year of each 10-year period. The land covered with a DCI of zero was not considered. Here, Periods 1 and 2 were selected for further investigations to be compared with Jung et al. (2010). (b) Number of times DCI changed sign between 1990 and 2012.

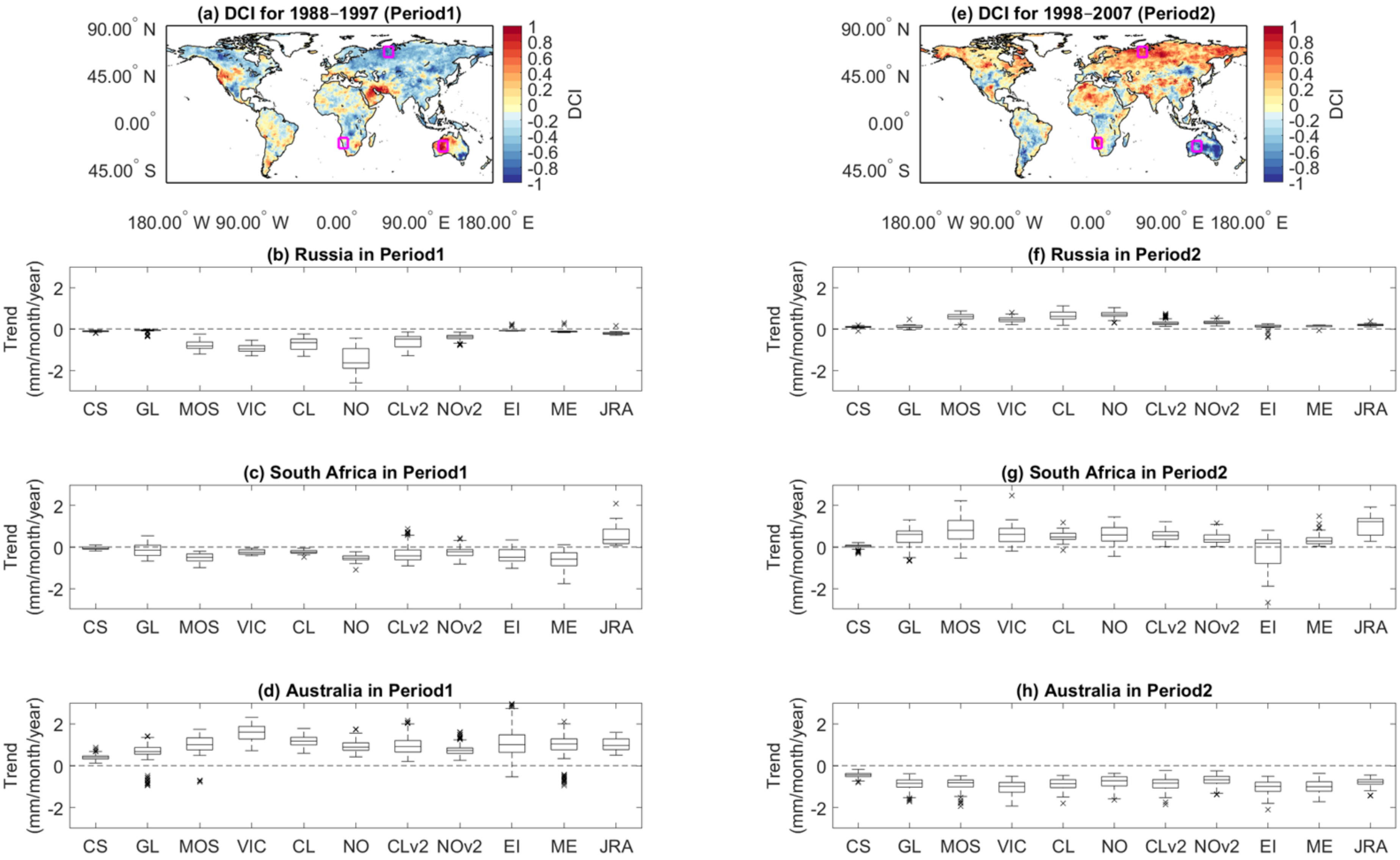
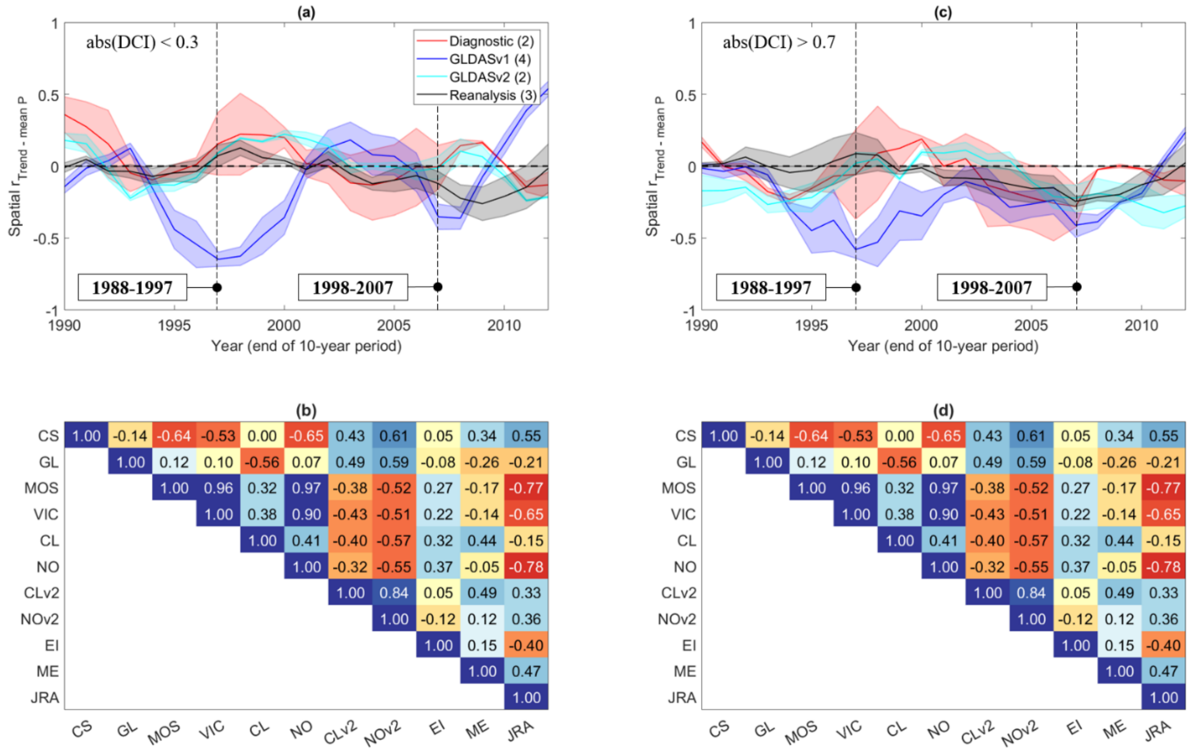

(a) DCI for 1988–1997 (Period1)
(e) DCI for 1998–2007 (Period2)
90.00° N
45.00° N
0.00°
45.00° S
180.00° W 90.00° W 0.00° 90.00° E 180.00° E
DCI
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
(b) Russia in Period1
(f) Russia in Period2
(c) South Africa in Period1
(g) South Africa in Period2
(d) Australia in Period1
(h) Australia in Period2
Trend (mm/month/year)
2
0
-2
CS
GL
MOS
VIC
CL
NO
CLv2
NOv2
EI
ME
JRA

Figure 4. Spatial distributions of DCI over two 10-year periods: (a) 1998–1997 (Period1); 1998–2007 (Period2). Boxplots of estimated trends across datasets in both periods: (b) and (f) for Russia (60°N – 70°N, 60°E – 70°E); (c) and (g) for South Africa (17°S – 27°S, 10°W – 20°W); (d) and (h) for Western Australia (20°S – 30°S, 120°E – 130°E), as marked with three squares in (a) and (e).



(b)

| | CS | GL | MOS | VIC | CL | NO | CLv2 | NOv2 | EI | ME | JRA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CS | 1.00 | -0.14 | -0.64 | -0.53 | 0.00 | -0.65 | 0.43 | 0.61 | 0.05 | 0.34 | 0.55 |
| GL | | 1.00 | 0.12 | 0.10 | -0.56 | 0.07 | 0.49 | 0.59 | -0.08 | -0.26 | -0.21 |
| MOS | | | 1.00 | 0.96 | 0.32 | 0.97 | -0.38 | -0.52 | 0.27 | -0.17 | -0.77 |
| VIC | | | | 1.00 | 0.38 | 0.90 | -0.43 | -0.51 | 0.22 | -0.14 | -0.65 |
| CL | | | | | 1.00 | 0.41 | -0.40 | -0.57 | 0.32 | 0.44 | -0.15 |
| NO | | | | | | 1.00 | -0.32 | -0.55 | 0.37 | -0.05 | -0.78 |
| CLv2 | | | | | | | 1.00 | 0.84 | 0.05 | 0.49 | 0.33 |
| NOv2 | | | | | | | | 1.00 | -0.12 | 0.12 | 0.36 |
| EI | | | | | | | | | 1.00 | 0.15 | -0.40 |
| ME | | | | | | | | | | 1.00 | 0.47 |
| JRA | | | | | | | | | | | 1.00 |

(d)

| | CS | GL | MOS | VIC | CL | NO | CLv2 | NOv2 | EI | ME | JRA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CS | 1.00 | -0.14 | -0.64 | -0.53 | 0.00 | -0.65 | 0.43 | 0.61 | 0.05 | 0.34 | 0.55 |
| GL | | 1.00 | 0.12 | 0.10 | -0.56 | 0.07 | 0.49 | 0.59 | -0.08 | -0.26 | -0.21 |
| MOS | | | 1.00 | 0.96 | 0.32 | 0.97 | -0.38 | -0.52 | 0.27 | -0.17 | -0.77 |
| VIC | | | | 1.00 | 0.38 | 0.90 | -0.43 | -0.51 | 0.22 | -0.14 | -0.65 |
| CL | | | | | 1.00 | 0.41 | -0.40 | -0.57 | 0.32 | 0.44 | -0.15 |
| NO | | | | | | 1.00 | -0.32 | -0.55 | 0.37 | -0.05 | -0.78 |
| CLv2 | | | | | | | 1.00 | 0.84 | 0.05 | 0.49 | 0.33 |
| NOv2 | | | | | | | | 1.00 | -0.12 | 0.12 | 0.36 |
| EI | | | | | | | | | 1.00 | 0.15 | -0.40 |
| ME | | | | | | | | | | 1.00 | 0.47 |
| JRA | | | | | | | | | | | 1.00 |

Figure 5. Temporal changes of spatial correlation between ET trends of the 11 datasets and mean precipitation ($r_{Trend-mean\ P}$) over regions with (a) weak and (b) strong DCI scores. Here, strong DCI: abs(DCI) >0.7; weak DCI: abs(DCI) <0.3. (b) and (d) represent cross-correlation between time series in (a) and (c), respectively.