# Unveiling the ~~Impact~~ impact of ~~Potential Evapotranspiration Method Selection~~ potential evapotranspiration method selection on ~~Trends~~ trends in ~~Hydrological Cycle Components Across~~ hydrological cycle components across Europe

Vishal Thakur[1], Yannis Markonis[1], Rohini Kumar[2], Johanna Ruth Thomson[1], Mijael Rodrigo Vargas Godoy[1,3], Martin Hanel[1], and Oldrich Rakovec[1]

[1] Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Kamýcká 129, Praha – Suchdol, Czech Republic

[3]River-Coastal Science & Engineering, Tulane University, New Orleans, 70118, Louisiana, USA

[2]Department Computational Hydrosystems, UFZ-Helmholtz Centre for Environmental Research, Leipzig, Germany

**Correspondence:** Vishal Thakur (thakur@fzp.czu.cz)

**Abstract.**

Hydrological models are essential tools for assessing ~~and predicting~~ changes in the hydrological cycle, ~~offering~~ enabling detailed quantification of ~~components like~~ runoff (Q), total water storage (TWS), and actual evapotranspiration (AET). Precipitation (~~PRE~~P) and potential evapotranspiration (PET) are the ~~major required drivers for~~ two major drivers in modeling these components~~. In modeling, the linkage of PRE to changes in these cycle components is well understood compared to PET. Here, we focus on the changes in PET and their influence on hydrological cycle components (~~, with the influence of P more extensively studied than PET. This study evaluates the impact of PET method selection on AET, Q, and TWS ~~). We consider~~ using 12 ~~distinct PET methods from three different categories (~~PET formulations categorized as temperature-based, radiation-based, and ~~combination type) across 553 European catchments. The~~ combinational. We applied the mesoscale Hydrological Model (mHM) ~~was used~~ to simulate 40 years of hydrological components ~~, with a total of 6 636 mHM runs. Comprehensive~~ across 553 European catchments. PET effects were analyzed through trend analysis and ~~data concurrence index~~ the Data Concurrence Index (DCI) ~~based on trend direction were applied to three different catchment categories(~~across three catchment categories: energy-limited, ~~water-limited and mixeddepending on PET method) to assess changes in PET and its influence on~~

~~AET, Q, and TWS. PET methods exhibit diverse~~ mixed, and water-limited. Our results indicate that annual and seasonal trends ~~across catchment categories for PET, AET, Q, and TWS~~are variably sensitive to method choice, depending on each component and catchment category. While PET ~~demonstrate~~ shows strong agreement in trend ~~directions, the~~ direction, trend magnitudes vary ~~depending on the choice of PET method. The findings reveal that the~~ among different PET methods. Jensen-Haise ~~method consistently~~ produces the highest ~~trends for PET on both~~ annual and seasonal ~~scales (summer, spring, and autumn)~~ , PET trend magnitudes, whereas no single ~~PET method consistently represents the lowesttrend~~method consistently yields the lowest. AET trends ~~are similar to those of~~ generally align with PET but are ~~lower in trend magnitude at annual scale, while seasonally,~~ weaker in magnitude on an annual scale. Seasonally, only energy-limited catchments show ~~a trend pattern~~ AET trends similar to PET. ~~Across all PET methods, there is strong agreement in trend direction, except during the winter season. For the majority of European catchments,~~ For Q and TWS~~show strong agreement among different methods, either positive or negative. In the annual trend, the summer season largely contributes to PET. For AET, summer season largely contributes to the annual trend only~~ , most European catchments exhibit strong trend agreement across PET methods. As expected, summer is the primary contributor to annual PET trends, while for AET, its influence is most notable in energy-limited ~~and water-limited catchments. Overall, studies focusing on the directional changes in the hydrological cycle or its components indicate that PET methods have a limited impact. However, when quantifying changes in hydrological cycle components, the choice~~ catchments. Looking at statistically significant trends, there is general agreement for PET and AET, which decline for the other hydroclimatic variables. On an annual scale, varying patterns of hydrological cycle intensification (increases in P, AET, Q, and TWS) are observed across European catchments, highlighting the influence of PET method ~~becomes crucial. Therefore, selecting the appropriate PET method is crucial for studies on~~ selection. Overall, this study highlights how the PET method selection affects the quantification of hydrological trends, emphasizing the importance of method selection for robust assessment of AET, Q, and TWS.

*Copyright statement.* TEXT

## 1 Introduction

~~In 1948, Thornthwaite (1948) introduced the concept of potential~~ Potential evapotranspiration (PET) ~~, which~~ is the potential to evaporate water from the land surface to the atmosphere without any limitation to water availability. ~~PET~~ Although the concept has been in use for centuries, Thornthwaite (1948) was the first to formally introduce the term “potential evapotranspiration” in the scientific literature. A related but distinct concept is “reference crop evapotranspiration”, which is sometimes used interchangeably with PET. However, these terms differ in their conceptual basis and applications. Reference crop evapotranspiration specifically estimates the water requirements of a standardized reference crop under ideal conditions, whereas potential evapotranspiration provides a broader representation of water and energy exchange processes over diverse landscapes and large regions (Xiang et al., 2020) . PET is used in diverse research fields. In agriculture, it is employed for irrigation scheduling and modeling crop water requirements (Xiang et al., 2020). In environmental studies, PET is used for aridification research and investigating extreme events, including meteorological, agricultural, and hydrological droughts (Park et al., 2018; Zhou et al., 2023; Shi et al., 2023a). In hydrology, it is used to determine the long-term states of catchments, such as energy-limited and water-limited catchments, and it plays a key role in the Budyko framework for estimating long-term changes in hydrological components (Reaver et al., 2022). Furthermore, PET is extensively used in hydrological modeling as to define the maximum rate of possible water loss through evaporation and transpiration. It is used as one of the important input variables to simulate key hydrological components, such as actual evapotranspiration (AET), runoff (Q), and total water storage (TWS).

Since ~~the~~ Thornthwaite’s study, more than ~~50~~ 100 empirical PET equations have been developed, ranging from simple to complex types ~~(Lu et al., 2005)~~(Proutsos et al., 2023). They can be classified mainly into three categories based on input data: (1) Temperature-based methods, which utilize temperature as input (Shaw and Riha, 2011). Due to their simplicity and minimal data requirement, these are widely used in the hydrological model (Arnold et al., 1998; Liu et al., 2008). (2) Radiation-based methods require solar radiation (short wave or net radiation) (Xu and Singh, 2000). (3) The combinational type requires temperature, radiation, wind speed, relative humidity, vapor pressure, etc. (Vicente-Serrano et al., 2014; Allen, 1998). Out of these 100+ methods, the majority are temperature-based methods (40+), followed by radiation-based methods (30+) and combination-based methods (10+). Many of these empirical methods were initially developed and tested for particular regional scales or climatic conditions. For instance, the Thornthwaite method is most suitable for humid climates, while the Hargreaves-Samani method is particularly effective in arid and semi-arid regions. Similarly, the Hamon method is suitable for all climates. All methods in these three categories incorporate several assumptions (climatic conditions and data availability) resulting in significant differences in their estimates (Lu et al., 2005).

~~In hydrological models, PET directly influences actual evapotranspiration (AET )~~ PET influences AET and consequently impacts the estimation of infiltration, ~~runoff and total water storage.~~ Q, and TWS in hydrological models. PET can have direct as well as indirect influence on AET. In hydrological models, AET is estimated by either separately determining water surface evaporation, soil evaporation, and vegetation transpiration and then combining these based on land use patterns or by first assessing potential evapotranspiration and subsequently adjusting it to actual evapotranspiration using the soil moisture extraction function (Zhao et al., 2013). ~~AET,~~ The mesoscale Hydrological Model (mHM) explicitly represents interception, where a portion of AET is derived from interception evaporation. This process is estimated as a fraction of PET using a power function derived from Deardorff (1978) and Liang et al. (1994). When the evaporative demand exceeds the intercepted water, the interception storage is fully depleted. Interception storage in the mHM is estimated as a function of leaf area index (LAI) that varies depending on vegetation types and season. AET in mHM is mainly contributed by canpoy evaporation, soil evaporation and open water evaporation. AET, being a key component of the water balance, affects the estimation of other water balance components (Q and TWS). While Q remains relatively insensitive, AET and TWS are more responsive to the choice of PET method (Bai et al., 2016). Hence, uncertainty in PET estimation influences the quantification of change in water cycle components.

Many studies have investigated the sensitivity of the hydrological model output to PET. Oudin et al. (2005) evaluated 27 PET methods with four hydrological models concluding that PET is insensitive to runoff generation, with similar conclusions made by Aouissi et al. (2016); Birhanu et al. (2018). Assessment of four PET methods with two monthly hydrological models reported that runoff is unaffected by the PET method, whereas AET and total water storage depend on the PET method (Bai et al., 2016). The study also concluded that calibration against the runoff is the main cause of PET insensitivity, and AET and total water storage compensate for it. In contrast to previous studies, Ndiaye et al. (2024) compared 21 PET methods for runoff estimation with three conceptual lumped hydrological models (GR4J, GR5J, and GR6J) in the Senegal River Basin, stating that better performance shown by combinational type methods. Similarly, Pimentel et al. (2023) compared three PET methods for their accuracy in simulating runoff and AET in the large-scale hydrological model (HYPE model). They found that Hargreaves-Samani performed best in the Amazonas, central Europe, and Oceania, and Priestley-Taylor in higher latitudes. These studies focus on the sensitivity and choice of PET methods in estimating hydrological components. While these findings reveal how PET methods can impact the magnitude of hydrological components, the impact of PET method selection on changes in these hydrological components is not often investigated. Temporal changes in these hydrological components are crucial for climate change mitigation, water availability, energy availability, and agricultural produce.

Trends in PET and its implication on hydrological components (AET) are examined by Anabalón and Sharma (2017). They compare trends in six PET and AET datasets, mainly estimated by the Penman-Monteith or Priestley-Taylor PET method. They found that PET trends were highly correlated with AET trends in energy-limited regions, while the AET trends were closely correlated with precipitation trends in water-limited regions. Additionally, they reported that PET and AET trends were inversely related in certain cases, mainly due to the prevailing influence of precipitation trends on AET trends. Similarly, Liu et al. (2022) identified a strong positive relationship between PET and AET changes in most global regions and an inverse relationship with total water storage change. The study is limited by using only the Penman-Monteith approach for PET and global datasets for AET and total water storage change. The inconsistency and lack of coherence between existing PET and AET datasets often necessitate using a single PET method compared to various AET datasets. Furthermore, previous studies have primarily focused on one-to-one trend comparisons than comprehensive analysis of all hydrological cycle components, including Q and TWS. Thus, research is needed to explore the impact of changes in PET methods on changes in different hydrological components of hydrological models.

In this study, our objective is to assess the trends of PET using 12 different PET methods and their influence on the trend of hydrological components (~~runoff; Q, AET, and total water storage;~~AET, Q, and TWS) across 553 European catchments. To assess the agreement between changes in different PET methods and corresponding hydrological components. The mesoscale hydrological model (mHM) is used to evaluate the influence of changes in different PET methods, from simple to most advanced approaches, on hydrological components across a range of European catchments. We chose a concurrency index to assess agreement between the PET method and hydrological components at each catchment. The data concurrency index is used to compare directions between different datasets (Anabalón and Sharma, 2017). In our research, we use it to examine directional changes in PET estimates, AET, Q, and TWS across each catchment.

## 2 Methods and data

### 2.1 Study area and catchment classification

This study ~~examines~~ includes 553 ~~catchments across Europe, covering all types of European climates~~European catchments ranging in size from 500 km$^2$ to 252 000 km$^2$. Catchments were selected based on the following criteria: first, a minimum area of 500 km$^2$; second, at least 10 years of observed discharge data based on the GRDC database; and third, a closed water balance condition (i.e., $(P-Q)/P<1$). The selected ~~catchment's sizes vary~~ catchments are divided into three categories based on the aridity index: energy-limited, mixed, and water-limited.

This study includes 553 European catchments ranging in size from 500 km$^2$ to 252 000 km$^2$~~and they~~. Catchments were selected based on the following criteria: first, a minimum area of 500 km$^2$; second, at least 10 years of observed discharge data from GRDC database; and third, a closed water balance condition ($(P-Q)/P<1$). The selected catchments are divided into three categories based on the aridity index: energy-limited, mixed, and water-limited (Figure 1a). This classification is based on the aridity index (AI), estimated as the ratio of mean PET to mean ~~PRE~~precipitation, a widely used metric that quantifies the dry or wet state of the catchment (Zhang et al., 2016; Massari et al., 2022). In our approach, which involves the application of multiple PET methods, a catchment is considered energy-limited if the AI is less than one for all the PET methods. Similarly, a catchment is water-limited if all PET methods report AI greater than one. If AI values appear to be both above and below one, depending on the PET method used, then the catchment is assigned to the mixed category (Figure 1b). Three representative catchments from each category are indicated by dark black lines in (Figure 1a) and are plotted in (Figure 1b). This classification allows us to distinguish the differences in magnitudes of PET and the other key hydrological components among the catchments (Figure 1c). By employing this methodology out of 553 catchments, we find 189 catchments being energy-limited, 34 water-limited, and the rest 330 belong to the mixed category.

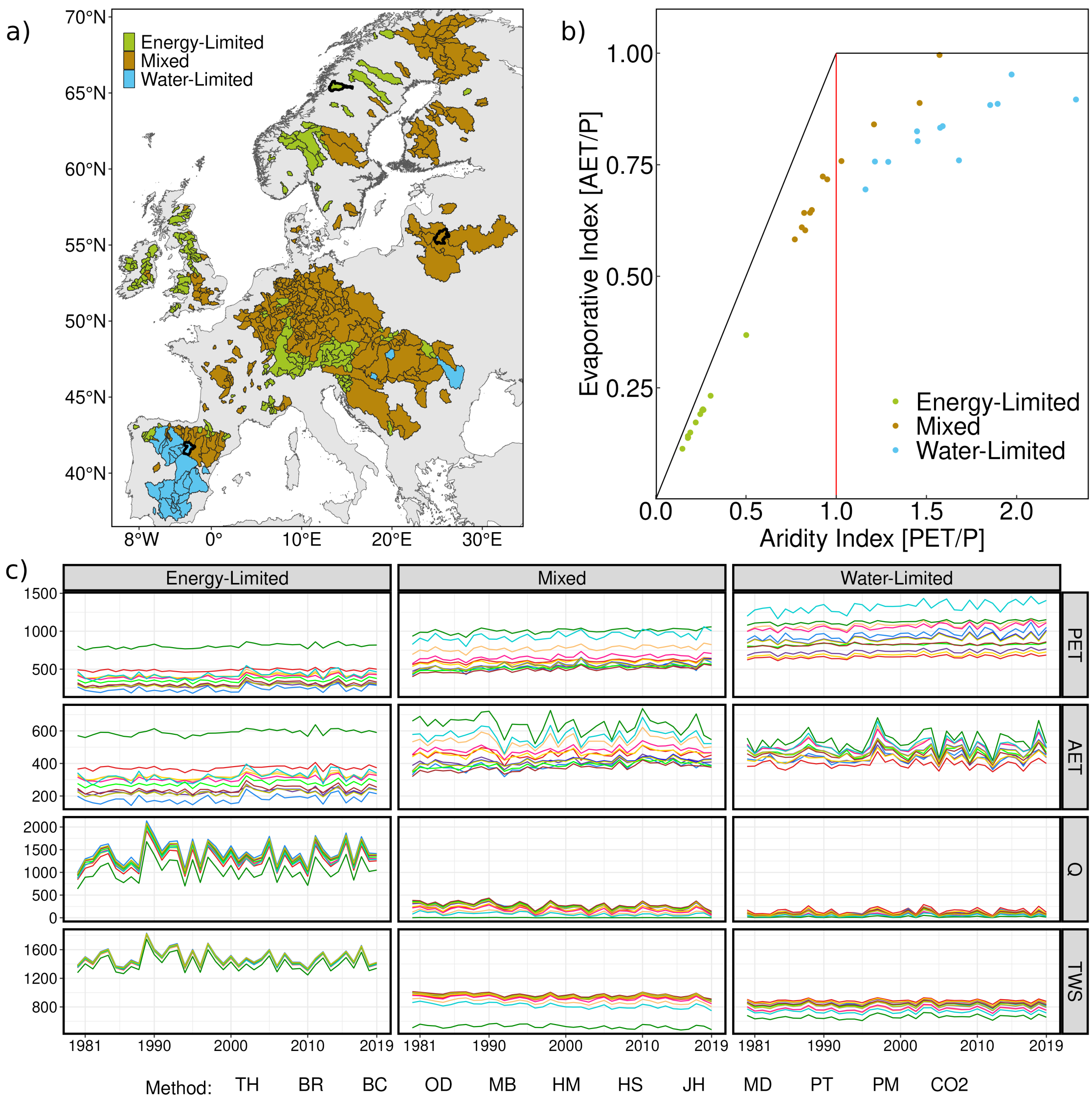


**Figure 1.** Catchment classification to energy-limited, mixed, and water-limited categories. a) ~~Catchment locations~~Spatial location of catchments; black borders indicate a representative catchment of each category. b) Classification example within the Budyko space for the representative catchments. Colored points represent three representative catchments, with each set of 12 points per color corresponding to different PET methods of representative catchments. c) Annual time series of simulated hydrological components corresponding to each representative catchment and PET estimation method (TH: Thornthwaite, BR: Bair-Robertson, BC: Blaney-Criddle, OD: Oudin, MB: McGuinness-Borden, HM: Hamon, HS: Hargreaves-Samani, JH: Jensen-Haise, MD: Milly-Dunne, PT: Priestley-Taylor, PM: Penman-Monteith, $CO_2$: Modified Penman-Monteith accounts $CO_2$.). All units are in mm year$^{-1}$.

### 2.2 Meteorological and geomorphological data

The Ensemble Meteorological Dataset for Planet Earth (EM-Earth; Tang et al., 2022) and ~~ERA5-land (Hersbach et al., 2020)~~ ERA5-Land (Muñoz-Sabater et al., 2021) were used to calculate different PET estimates and run the mesoscale Hydrological Model (mHM; Samaniego et al., 2010; Kumar et al., 2013b). The EM-Earth dataset ~~is generated based on the~~ (Tang et al., 2022) is derived from observed station data ~~(~~SC-Earth ~~)~~ (Tang et al., 2021) and ERA5 data (Hersbach et al., 2020). It incorporates a novel optimal interpolation technique and considers the temporal inconsistencies between the station and reanalysis data (Tang et al., 2022). ~~ERA5-land~~ ERA5-Land dataset is a reanalysis data product created by the European Centre for Medium-Range Weather Forecasts (ECMWF) and has been widely used in numerous hydrological modeling studies (Muñoz-Sabater et al., 2021). Both datasets are available at 0.1° × 0.1° spatial resolution, but EM-Earth has hourly as well as daily time step, while ~~ERA5-land~~ ERA5-Land is at hourly scale. Meteorological forcings, particularly precipitation and temperature from EM-Earth, have been found to be more accurate based on comparisons with several datasets in data-rich regions (Tang et al., 2022).

In our analysis, we use daily temperature and precipitation from EM-Earth, and radiation (long- and short-wave), surface pressure, and wind components (U and V) from ERA5-Land for the period 1980–2019 (Table 1). ~~They~~ The EM-Earth dataset provides high-quality precipitation and temperature data and has been shown to perform well over Europe (Tang et al., 2022). It has undergone climatology-based bias correction and accounts for precipitation undercatch. However, since EM-Earth does not include all necessary variables for PET estimation, we utilize ERA5-Land as a complementary dataset. ERA5-Land has been demonstrated to perform better than other reanalysis datasets, including ERA5 and ERA-Interim (Muñoz-Sabater et al., 2021). Nonetheless, its limitations in hydrological modeling have been acknowledged by Clerc-Schwarzenbach et al. (2024); Tarek et al. (2020). Several recent global studies follow a similar strategy, combining precipitation and temperature from EM-Earth with radiation, wind speed, and other meteorological variables from ERA5-Land (Tang et al., 2023; Yin et al., 2024; Rakovec et al., 2023). These meteorological data combinations, along with the simulated hydrological components derived from them, demonstrate lower uncertainty across Europe (Tang et al., 2023). These are homogenized to daily temporal scale and 0.125° × 0.125° spatial scale to be compatible with the previous simulations run by mHM (Pohl et al., 2023; Fang et al., 2024). Homogenization using the nearest neighbor technique and necessary mathematical operations (appropriate unit conversion of datasets) are performed using the Climate Data Operator (CDO; Schulzweida, 2022).

Morphological data such as Leaf Area Index (LAI), soil properties, and terrain characteristics (such as flow direction, flow accumulation, slope, and aspect) are sourced from mHM European database (Rakovec et al., 2016). This database originally utilized data from different sources, such as soil properties from the International Soil Reference and Information Centre (ISRIC), terrain characteristics from the U.S. Geological Survey (USGS) and the National Geospatial-Intelligence Agency (NGA), LAI from Global Inventory Modeling and Mapping Studies (GIMMS) and Land cover from Global Land Cover (GlobCover) by European Space Agency (ESA). $CO_2$ concentration is sourced from Cheng et al. (2022), which is reconstructed from the Carbon Dioxide Information Analysis Center (CDIAC) data.

**Table 1.** Summary of meteorological and morphological data. ~~PRE~~ P is Precipitation, $T_{avg}$ is average air temperature, $T_{range}$ is the temperature range, which is the difference between maximum and minimum air temperature, $T_{dew}$ is dew point temperature of air, SW is Short wave radiation, LW is ~~longwave~~ Long wave radiation, U is eastward component of wind speed at 10 m, V is northward component of wind speed at 10 m, $Con_{CO_2}$ is $CO_2$ concentration

| Variable | Temporal Scale | Spatial Scale | Record length | Source | Reference |
|---|---|---|---|---|---|
| | | | Meteorological data | | |
| ~~PRE~~ P | Hourly/Daily | $0.1° \times 0.1°$ | 1950–2019 | EM-Earth | Tang et al. (2022) |
| $T_{avg}$ | Hourly/Daily | $0.1° \times 0.1°$ | 1950–2019 | EM-Earth | Tang et al. (2022) |
| $T_{range}$ | Hourly/Daily | $0.1° \times 0.1°$ | 1950–2019 | EM-Earth | Tang et al. (2022) |
| $T_{dew}$ | Hourly/Daily | $0.1° \times 0.1°$ | 1950–2019 | EM-Earth | Tang et al. (2022) |
| SW | Hourly | $0.1° \times 0.1°$ | 1950-2022 | ~~ERA5-land~~ ERA5-Land | Muñoz-Sabater et al. (2021) |
| LW | Hourly | $0.1° \times 0.1°$ | 1950-2022 | ~~ERA5-land~~ ERA5-Land | Muñoz-Sabater et al. (2021) |
| U | Hourly | $0.1° \times 0.1°$ | 1950-2022 | ~~ERA5-land~~ ERA5-Land | Muñoz-Sabater et al. (2021) |
| V | Hourly | $0.1° \times 0.1°$ | 1950-2022 | ~~ERA5-land~~ ERA5-Land | Muñoz-Sabater et al. (2021) |
| | | | Other data | | |
| $Con_{CO_2}$ | Annual | $0.1° \times 0.1°$ | 1950-2022 | – | Cheng et al. (2022) |
| LAI | monthly | 1/512° | static | GIMMS | Tucker, Pinzon, and Brown (2004) |
| Soil properties | – | 1/512° | – | SoilGrids | ISRIC - World SoilInformation (2017) |
| Land cover | static | 1/512° | static | GlobCover | Arino et al. (2012) |
| DEM (+ derivatives) | static | 1/512° | static | GMTED2010 | USGS and NGA (2018) |
| Geology | static | 1/512° | static | GLiM | Hartmann and Moosdorf (2012) |

## 2.3 Methodology

### 2.3.1 PET methods/formulations

We incorporate 12 PET methods at a daily scale from all three categories of estimation: temperature, radiation, and combinational methods (Table 2). Temperature-based methods require temperature data, which can include average temperature, minimum temperature, or maximum temperature. Additionally, PET methods that incorporate extraterrestrial radiation are also considered under this category. Combinational-based methods require a larger number of variables compared to temperature- and radiation-based methods to estimate various physical terms, such as wind speed and surface pressure (Table 2). Most temperature-based methods use only daily average temperature (Thornthwaite, Oudin, Hamon, Jensen-Haise, ~~Mcguinnes-Bordne,~~ McGuinness-Bordne, and Blaney-Criddle), while Baier-Robertson employs both minimum and maximum daily ~~temperature and Hargreaves-Samani uses minimum, maximum, and average daily temperature~~temperatures. Some of ~~them~~ these methods also include an extraterrestrial radiation term in their formulation. However, since this ~~extraterrestrial~~ radiation term is calculated based on ~~latitudinal information~~latitude and follows a consistent annual cycle varying only with the calendar date, only temperature data is ~~required to calculate PET~~ needed for PET calculation. We utilize only one radiation-based method, Milly-

Dunne PET~~that~~ , which requires only net radiation data to estimate PET. The combinational ~~type includes the PET methodswith~~ methods, such as Penman-Monteith and Priestley-Taylor, have a stronger physical basis~~(Penman-Monteith, Priestley-Taylor). They employ more variables than the temperature- and radiation-based methods to estimate various physical terms such as relative humidity, vapor pressure, saturation vapor pressure, slope of vapor pressure curve, etc.~~ . In our analysis, all these physical terms are estimated ~~according to~~ following Allen (1998). Additionally, ~~in~~ within the combinational category, we ~~use~~ employ the modified Penman-Monteith (~~$CO_2$~~CO2) method, which accounts for temporal ~~variation~~ variations in changing carbon dioxide concentrations. Formulation details~~(~~, including mathematical equations and associated constants ~~) of~~ for each PET method, are provided in Table A1.

**Table 2.** List of PET methods and required input data. $T_{max}$ is maximum air temperature (°C), $T_{min}$ is minimum air temperature (°C), Pr is surface pressure (pa), $R_n$ is net radiation ($J/m^2$), $u_2$ is the wind speed at 2m from the surface (m/s), $Con_{CO_2}$ is $CO_2$ concentration (ppm)

| Type | Method name | Method abbreviation | Required input | References |
|---|---|---|---|---|
| Temperature | Hargreaves-Samani | HS | $T_{max}$, $T_{min}$, $T_{avg}$ | ~~George H. Hargreave~~ Hargreaves and Sama |
| | Thornthwaite | TH | $T_{avg}$ | Thornthwaite (1948) |
| | Oudin | OD | $T_{avg}$ | Oudin et al. (2005) |
| | Hamon | HM | $T_{avg}$ | Hamon (1961) |
| | Baier-Robertson | BR | $T_{max}$, $T_{min}$ | Bai et al. (2016) |
| | Jensen-Haise | JH | $T_{avg}$ | Jensen and Haise (19 |
| | McGuinness-Borden | MB | $T_{avg}$ | McGuinness and Bor (1972) |
| | Blaney-Criddle | BC | $T_{avg}$ | Blaney (1952) |
| Radiation | Milly-Dunne | MD | $R_n$ | Milly and Dunne (20 |
| Combinational | Priestley-Taylor | PT | $T_{avg}$, Pr, $R_n$ | Priestley and Ta (1972) |
| | Penman-Monteith | PM | $T_{max}$, $T_{min}$, $T_{avg}$, $T_{dew}$, Pr, $u_2$, $R_n$ | Penman (1948) |
| | ~~Penman Monteith~~Penman-Monteith[$CO_2$] | $CO_2$ | $T_{max}$, $T_{min}$, $T_{avg}$, $T_{dew}$, Pr, $u_2$, $R_n$, $Con_{co2}$ | Yang et al. (2019) |

### 2.3.2 mesoscale Hydrological Model (mHM)

mHM is a hydrological model which explicitly accounts for sub-grid variability of hydrological processes (Samaniego et al., 2010; Kumar et al., 2013b; Thober et al., 2019). mHM has been successfully applied and tested in more than 1000 European basins ranging in size from 4 $km^2$ to more than 100 000 $km^2$ at various spatial resolutions or grid cell size (1-100 km) (Samaniego et al., 2010; Kumar et al., 2013b; Rakovec et al., 2016, 2019; Shrestha et al., 2024). Additionally, the model is currently applied at the global scale with comparable and sometimes even improved model performance with respect to other large-scale hydrological models (Samaniego et al., 2019). mHM demonstrates robust performance and applicability across

Europe (Kumar et al., 2020). It is also one of the several ~~large scale~~ large-scale hydrological models, which were used by the WMO for their annual State of Global Water Resources reports (World Meteorological Organization (WMO), 2023).

We run mHM(v5.12.0) over 553 European catchments, using the meteorological data from EM-Earth and the 12 different PET estimation methods. Overall, 6 636 (12 × 553) mHM simulations are performed for all the study basins. The ~~basins were not calibrated for the each PET method to access their true response in hydrological cycle components~~model was set up for each catchment at a daily temporal resolution and 0.125° × 0.125° spatial scale. All meteorological forcings were kept constant with only varying PET ~~estimate~~estimates. To calculate TWS, we aggregate soil moisture at different layers, canopy interception storage, snowpack, groundwater levels, sealed area reservoirs, and unsaturated zone reservoirs at each grid cell and time step. The hydrological components (AET, Q, and TWS) and PET are averaged over the catchment area and monthly time steps. In this research default parametrization is used for model setup. The default parameterization of mHM has been shown to perform well in previous studies (Kumar et al., 2013a; Rakovec et al., 2016). Furthermore, it has been demonstrated as one of the best-performing configurations compared to other large-scale hydrological models (Samaniego et al., 2019). Additionally, our assessment of model performance is consistent with the findings of Samaniego et al. (2019). Our model evaluation against discharge shows that the median Kling–Gupta Efficiency (KGE) ranges from 0.60 to 0.75 across most PET methods (Figure S1).

#### 2.3.3 Trend analysis

We use Theil-Sen's slope method to calculate the magnitude and direction of linear change in PET, AET, Q, and TWS (Sen, 1968). The Mann-Kendall trend test is used to test the statistical significance of the observed trend (Kendall, 1948). A trend is considered statistically significant at the 5% level ($p \leq 0.05$). Sen's slope is non-parametric and insensitive to outliers and types of distribution. Due to its robust application, this method is widely used in hydrology, climate, and environmental-related studies (Anabalón and Sharma, 2017; Thackeray et al., 2022). It accounts for all possible pairs of data points from a time series and finds the median value as the slope magnitude. Eq. 1 and 2 shows the calculation steps of Sen's slope:

$$S_k = \frac{X_j - X_i}{t_j - t_i} \quad \text{where } 1 \leq i < j \leq n \tag{1}$$

$$S_{med} = \begin{cases} S_{\left[\frac{n+1}{2}\right]} & \text{if } n \text{ is odd} \\ \frac{S_{\left[\frac{n}{2}\right]} + S_{\left[\frac{n+2}{2}\right]}}{2} & \text{if } n \text{ is even,} \end{cases} \tag{2}$$

where $S_k$ is the linear slope for pair $X_i$ and $X_j$, $S_{med}$ is the median slope, $X_i$ and $X_j$ are data points from periods $t_i$ and $t_j$, $n$ is the number of data points in time series. Positive $S_{med}$ represents a positive trend, with the magnitude indicating the rate of increase. Similarly negative $S_{med}$ represents a negative trend, with the magnitude indicating the rate of decrease.

Here, we use the *trend* R package to estimate Sen's slope over a 40-year period from 1980 to 2019 at annual and seasonal (winter, spring, summer, and autumn) scales for each catchment. The units of trend at the annual scale are expressed as mm year$^{-1}$, while at the seasonal scale, they are represented as mm seas$^{-1}$ year$^{-1}$. For instance, a summer season trend of 1 mm seas$^{-1}$ year$^{-1}$ indicates that each year, an additional 1 mm is added to the summer season. The trend of each PET method were analyzed exclusively within seasons (for eaxmple winter season compared with winter season), without cross-seasonal comparisons.

#### 2.3.4 ~~Modified~~ Data Concurrence Index (DCI)

The Data Concurrence Index quantifies the level of concurrence between the significant trends in different datasets of the same variable ~~Anabalón and Sharma (2017). We modified this index by considering only directional information of corresponding slopes irrespective of trend significancei.e., counting the overall positive and negative slope occurrences. The adjustment was made to include all trend estimates from the PET methods . The Modified Data Concurrence Index (DCI) can be described as~~ (Anabalón and Sharma, 2017). DCI is applied in two forms in this study. The first is the original formulation, which includes only statistically significant trends. It is used to evaluate agreement among PET methods that yield significant changes in hydrological cycle components. The second is a modified version that considers all detected trends, regardless of significance. This captures the overall agreement across all PET methods for each catchment and each hydrological component. The detailed formulation for both cases are described in Eq. 3:

$$\cancel{\mathrm{DCI}}\,\mathrm{DCI} = \cancel{\frac{1}{\mathrm{ND}}\sum_{i=1}^{\mathrm{ND}}\frac{S_i}{abs(S_i)},}\begin{cases} \frac{1}{\mathrm{N}}\sum\limits_{i=1}^{\mathrm{N}}\delta_i\cdot\frac{S_i}{|S_i|} & \text{For statistically significant trends} \\ \frac{1}{\mathrm{N}}\sum\limits_{i=1}^{\mathrm{N}}\frac{S_i}{|S_i|} & \text{For all trends}\end{cases} \quad \text{with} \quad \delta_i = \begin{cases} 1, & \text{if } S_i \text{ is statistically significant} \\ 0, & \text{otherwise}\end{cases} \tag{3}$$

where DCI is the ~~Modified~~ Data Concurrence Index, ~~ND~~ N denotes the number of datasets, and $S_i$ is the magnitude of the slope.

The positive DCI represents a higher number of positive slopes than negative slopes and vice-versa. For instance, a DCI of 1 for AET and Q implies positive change for all the PET methods. Similarly, a DCI of -1 for AET and Q implies a negative change for all the PET methods. A DCI of 0.5 indicates that nine out of 12 methods, or 75% of the methods, show a positive change, and similarly, a DCI of -0.5 indicates that nine out of 12 methods show a negative change, or 75% of the methods. A DCI of zero denotes an equal number of positive and negative slopes (six positive and six negative). Our analysis estimates DCI from PET, AET, Q, and TWS slopes at annual as well as seasonal scales.

## 3 Results

### 3.1 Trend comparison of PET methods at annual scales

~~To quantify the long-term changes in hydrological cycle components, we used the~~ By applying the Theil-Sen slope method~~. Changes~~ , we observe that changes in PET depend on the choice of PET ~~method selection~~ estimation formulation (Figure 2). Considerable variability is observed among the PET methods, with median ~~trends~~ slopes ranging from slightly positive to 6 mm year$^{-1}$ during the 1980–2019 period. The Jensen-Haise method shows the highest ~~trend~~ change among all methods across different ~~catchments. In~~ catchment categories and also has the highest absolute average PET across European catchments (Figure S21). Generally, changes in PET are higher in water-limited than in energy-limited ~~catchments, a positive trend in PET is observed across all methods , except Penman-Monteith and Penman-Monteith$CO_2$The median trend for each PET method is positive, ranging from 1 mm year$^{-1}$ to 2 mm year$^{-1}$.In mixed catchments, most PET methods reflect a positive trend, though a few catchments using the Penman-Monteith$CO_2$method exhibit a slight negative trend. All PET methods exhibit a positive trend in water-limited catchments , except the energy-based Milly-Dunne method. Overall variability in trend estimates of~~ catchments. This difference arises since temperature-based methods depend on temperature changes. Conversely, combinational methods are influenced by more than one meteorological variable (temperature, wind speed, radiation, etc.). When we consider only statistically significant trends ($p < 0.05$), we observe that temperature-based PET methods consistently demonstrate statistically significant positive trends for most of the catchments (Table S2). Radiation-based and combinational methods account for fewer catchments than temperature-based PET methods (Table S2), implying that they generate weaker trends compared to the temperature-based PET methods. In addition, overall trend variability among the PET methods decreases from energy-limited to ~~mixed and~~ water-limited catchments, irrespective of trend significance.

AET trends ~~closely follow PET trends across all catchment categories~~ generally align with PET trends in energy-limited catchments but with smaller magnitudes (Figure 2). In ~~energy-limited~~ these catchments, all PET methods ~~show a positive~~ leads to a positive AET trend in terms of median values. However, a few catchments in this category reveal a slight negative change for the Blaney-Criddle, Jensen-Haise, Milly-Dunne, and Priestley-Taylor methods. For mixed catchments, the median AET trend is positive for all PET methods except Blaney-Criddle. The negative AET trends are similar to those in energy-limited catchments. Overall, the ~~trend patterns~~ AET trend patterns (high and low trends) for energy-limited and mixed catchments are similar to the trends in PET for these catchments, regardless of trend magnitude, with a few exceptions such as Blaney-Criddle and Jensen-Haise (Figure ~~S10~~2). In water-limited catchments, both positive and negative trends in AET are observed. The pattern remains similar to PET trends with a few exceptions, such as Blaney-Criddle.

~~The long-term trends in Q are relatively insensitive compared to that of the PET for~~ Across all PET methods, statistically significant positive AET trends were found in 162 energy-limited, 217 mixed, and 2 water-limited catchments (Figure S2). Among these significant trends, Jensen-Haise yields the highest AET trend estimates in energy-limited and mixed ~~catchment categories~~ catchments. These statistically significant trends closely follow the overall AET trend patterns in these catchment categories.

Q trends exhibit lower sensitivity to PET methods in energy-limited and mixed catchments compared to PET trends when considering all trends (Figure 2). Despite the positive median, a substantial fraction of catchments exhibit a negative trend in energy-limited catchments. In contrast, ~~all methods show negative~~ for mixed catchments, most PET methods produce negative Q trends, though some catchments maintain positive trends~~in mixed catchments with numerous catchments maintaining a positive trend~~. In water-limited catchments, there is variability in PET methods; for instance, Milly-Dunne has a larger trend, whereas Blaney-Criddle shows the lowest trend. Even though PET methods are insensitive in Q, variability exists among the PET methods within each catchment category. Overall, one-third of the catchments (183) show statistically significant Q trends: 28 energy-limited, 154 mixed, and one water-limited. All energy-limited catchments show positive Q trends, except those estimated with Jensen-Haise. In mixed catchments, all statistically significant trends are negative, with the exception of Blaney-Criddle. Despite fewer catchments with statistically significant trends, the variability in Q trends across PET methods persists, particularly in energy-limited and mixed catchments.

TWS ~~shows sensitivity~~ trends are sensitive to PET methods in water-limited and mixed catchments, ~~whereas~~ but energy-limited catchments remain ~~rather insensitive~~ largely unaffected (Figure 2). ~~In~~ Despite a consistent negative median trend across PET methods in energy-limited ~~catchments, all PET methods have a negative trend corresponding to the median ; however, a substantial fraction of catchments exhibit a positive trend~~catchments, few catchments still exhibit positive trends. Mixed catchments ~~show similar results. Temperature-based methodsexhibit larger~~ display a similar pattern. Among PET methods, temperature-based approaches show greater variability than radiation ~~and combination~~ or combinational types in mixed catchments. In water-limited catchments, ~~PET methods show both positive and negative trends , with negative trends close to zero. The~~ trends span positive to near-zero-negative, with Blaney-Criddle, Hargreaves-Samani, and Milly-Dunne ~~methods show large~~ showing notable positive trends. ~~Trend estimates based on~~ The Blaney-Criddle method ~~demonstrate large variability as compared to other PET methods in the~~ also yields higher variability in trend estimates, especially in mixed and water-limited ~~categories.~~ catchments. When focusing on statistically significant trends, TWS has a similar distribution to Q (30/172/1). Most PET methods show decreasing trends in energy-limited and mixed catchments, with the exception again of Blaney-Criddle (Figure S2). Compared to the full trend set, statistically significant trends reveals stronger inter-method differences, with Jensen-Haise showing the steepest decline and Thornthwaite the weakest.

~~Each PET method results in different trends in hydrological components (~~Hydrological trend estimates for AET, ~~Q, and TWS) for all catchment categories, except for~~ Q, and TWS vary across PET methods and catchment types, regardless of whether all trends or only statistically significant trends are considered. Statistically significant trends reveal greater divergence between methods in energy-limited ~~catchments where Qand TWS show insensitivity to PET methods . The trend magnitude of PET is reduced in that of AET trend, but the overall pattern of PET methods matches well with PET in all categories. In energy-limited catchments , PET methods are insensitive to~~ and mixed catchments. AET trends, though weaker in magnitude than PET trends, show similar spatial patterns, particularly in energy-limited and mixed regions. No single PET method stands out as consistently dominant across all components. For the trends beyond the statistically significant threshold a stronger pattern emerges. Approximately 70% of the catchments exhibit statistically significant AET trends, compared to only 33% for Q and 36% and TWS. Despite widespread statistical insignificance for Q and ~~TWS, while in mixed and water-limited~~

~~categories, Q and~~ TWS~~exhibit varying trends. There is no single PET method that shows a consistently higher or lower trend in all the hydrological components~~, distinct regional patterns appear. For instance, northern catchments display a mix of increasing and decreasing trends, while southern regions, especially the Iberian Peninsula, consistently demonstrate positive Q trends.

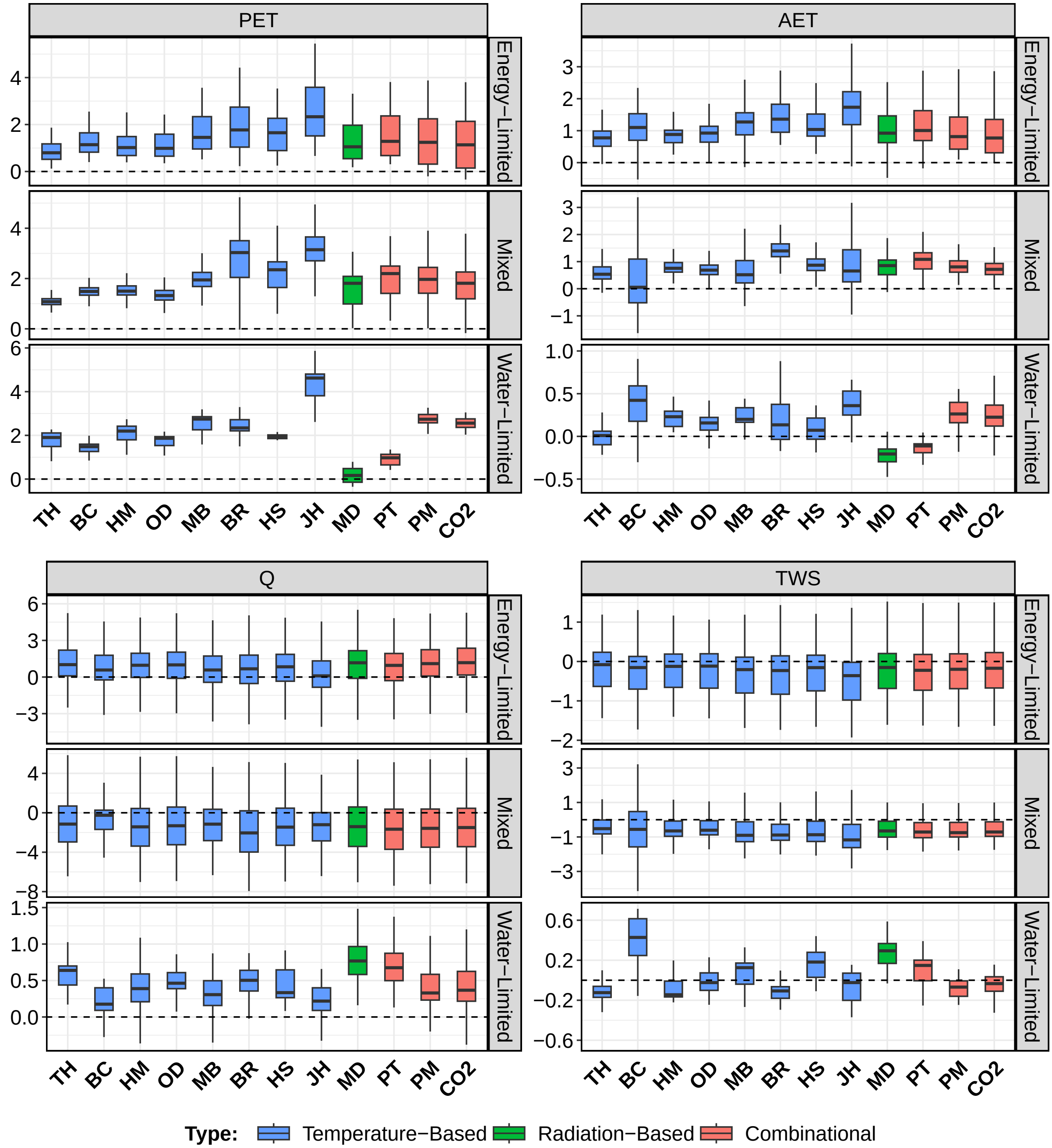


**Figure 2.** Boxplots represent the annual trends (mm $year^{-1}$) of different PET methods for PET, AET, Q, and TWS across various categories of ~~catchment~~catchments. The whiskers represent the 10th and 90th percentiles, and the box encompasses the 25th and 75th percentiles, with the median represented by middle line of the box. Abbreviations used for different PET methods are TH: Thornthwaite, BR: Bair-Robertson, BC: Blaney-Criddle, OD: Oudin, MB: McGuinness-Borden, HM: Hamon, HS: Hargreaves-Samani, JH: Jensen-Haise, MD: Milly-Dunne, PT: Priestley-Taylor, PM: Penman-Monteith, $CO_2$: Modified Penman-Monteith accounts $CO_2$.

### 3.2 Trend comparison of PET methods at seasonal scales

Hydrological cycle components exhibit considerable ~~variability in trends across different seasons. In the summer season~~ seasonal variability in trend magnitude. During summer (JJA), nearly all PET methods ~~demonstrate a positive trend for PET across all catchment categories, except for the~~ exhibit positive trends, with the exception of Milly-Dunne ~~method~~ in water-limited catchments (Figure 3). ~~The~~ Jensen-Haise ~~method has~~ consistently shows the highest trend and ~~the greatest variability among all PET methods across each catchment category~~greatest variability across all catchments. Notable variability is observed among PET methods in water-limited catchments. ~~The winter season exhibits the lowest trends, whereas the autumn season demonstrates comparable trendswith summer for PET~~ These patterns persist in statistically significant trends, despite a smaller number of catchments per category (Figure S3). Winter records the weakest PET trends, while spring trends are comparable to those in summer (Text S2). The summer season is the primary contributor to the annual trends in PET across all catchment categories (Figure ~~S12~~S22). The ~~trend pattern of AET in the winter season is similar to PET, while slight differences are observed in the spring and summer seasons (Text S2). AET in the summerseason, energy-limited and mixed catchments typically exhibit~~ distribution of statistically significant trends aligns with these findings, with the highest catchment count in spring (460) and the lowest in winter (66).

In summer, AET exhibits an overall positive trend ~~. The trend pattern among PET methods is consistent with the PET trends observed~~ in energy-limited ~~catchments~~and mixed catchments for both all trends and statistically significant trends (Figure 3, S3). The Jensen-Haise method exhibits greater variability for ~~both catchment types. However, despite the positive trend in PET ,~~ energy-limited and mixed catchments. In water-limited catchments, despite positive PET trends, AET trends are negative across all PET methods~~reveal a negative trend in AET, with Blaney-Criddle displaying the highest negative trend~~, with Bair-Robertson showing the strongest decline, followed by Jensen-Haise ~~. For AET , the summer season significantly contributes to the annual trends in both~~ (Figure 3). In spring, AET increases across methods, with the highest trends observed for Jensen-Haise in energy-limited and ~~water-limited catchments. In contrast, for mixed catchments, both~~ Bair-Robertson in mixed catchments. Combinational methods show consistent trends across both types (Figure S6). This pattern is maintained for statistically significant trends (Figure S7), with all PET methods showing positive median trends in autumn (Figure S8, S9). Summer season primarily drives annual trends in energy- and water-limited catchments, while spring and summer ~~seasons play a more substantial role,~~ contribute jointly in mixed catchments depending on the PET method (Figure ~~S12~~S22).

~~In summer, Q remains largely insensitive to changes in PET method trends across all catchment categories, with slight variability among~~ Q remains generally insensitive to PET method variation across all seasons and catchment types, with minor variability among methods within each ~~catchment~~ category (Figure 3~~). A similar result is observed for the winter, spring, and autumn seasons are discussed in the supplementary information (Text S2). For~~, S4, S6, S8). In energy-limited catchments, ~~the trend in Q across different PET methods remains~~ trends remain close to zero ~~. For mixed catchments , almost all methods demonstrate a negative trend with varying trend magnitude~~across PET methods. Mixed catchments show broadly negative Q trends across methods, with Blaney-Criddle ~~registering the least negative trend (Figure 3). Similarly, for~~ exhibiting the weakest decline. In water-limited ~~catchments, Q shows insensitivity to PET method selection. However, radiation and combination~~

~~methods reflect a positive trend corresponding to PET methods, whereas~~ regions, Q trends are similarly insensitive, though radiation and combinational methods tend to positive median trends, unlike temperature-based methods~~exhibit both positive and negative trends in Q (Figure 3). For Q, distinct patterns are observed in the primary contributors to annual trends. In both~~ , which exhibit mixed results. Statistically significant Q trends reveal limited seasonality in water-limited ~~and mixed catchments~~catchments, with notable exceptions like Blaney-Criddle, which contributes fewer catchments (Figures S3, ~~spring is the most contributing season. In contrast, for~~ S5, S7, S9). In energy-limited ~~catchments, the December and summer seasons are the main contributors, with their impact varying~~ and mixed catchments, statistically significant Q trends demonstrate higher seasonal magnitude as weaker trends fail to surpass the significance threshold. No single PET method consistently dominates in trend magnitude, though Jensen-Haise and Blaney-Criddle frequently yield the highest number of statistically significant catchments. Spring emerges as the dominant contributor to annual Q trends in mixed and water-limited catchments, while in energy-limited areas, winter and summer are most influential, depending on the ~~selected~~ PET method (Figure ~~S13). For TWS in summer,~~ S22).

TWS trends across seasons show minimal sensitivity to PET method in energy-limited and mixed catchments ~~are largely insensitive to the PET method, except for Jensen-Haise, which shows a negative trend. Although there is slight variation~~ under all-trend conditions, though slight variability exists among PET methods ~~within each category, mixed behavior is observed in water-limited catchments~~ (Figure 3). ~~Some temperature-based methods, including~~ Water-limited catchments, however, display a mixed response. Median trend patterns are generally stable, but Blaney-Criddle ~~, McGuinness-Bordne, and Hargreaves-Samani, indicate~~ consistently shows greater variability than other methods across all seasons and catchment categories. Statistically significant trends are uncommon in water-limited catchments, but trend magnitudes increase notably in energy-limited and mixed catchments during spring, summer, and autumn. In winter, temperature-based methods demonstrate positive trends, ~~while others show negative trends. All combination-type methods exhibit negative trends for TWS in summer.~~ with Blaney-Criddle and Jensen-Haise showing the lowest values (Figure S5). Slight variability across methods also seen in non-winter seasons (Figures S3, S7, S9). Seasonal contributions to annual trends vary by catchment category: spring contributes most in energy-limited catchments, summer in mixed catchments, and in water-limited catchments. The dominant season depends on the PET method—summer leads for several (TH, BR, HM, PM, and CO2), while spring dominates for the remaining methods.

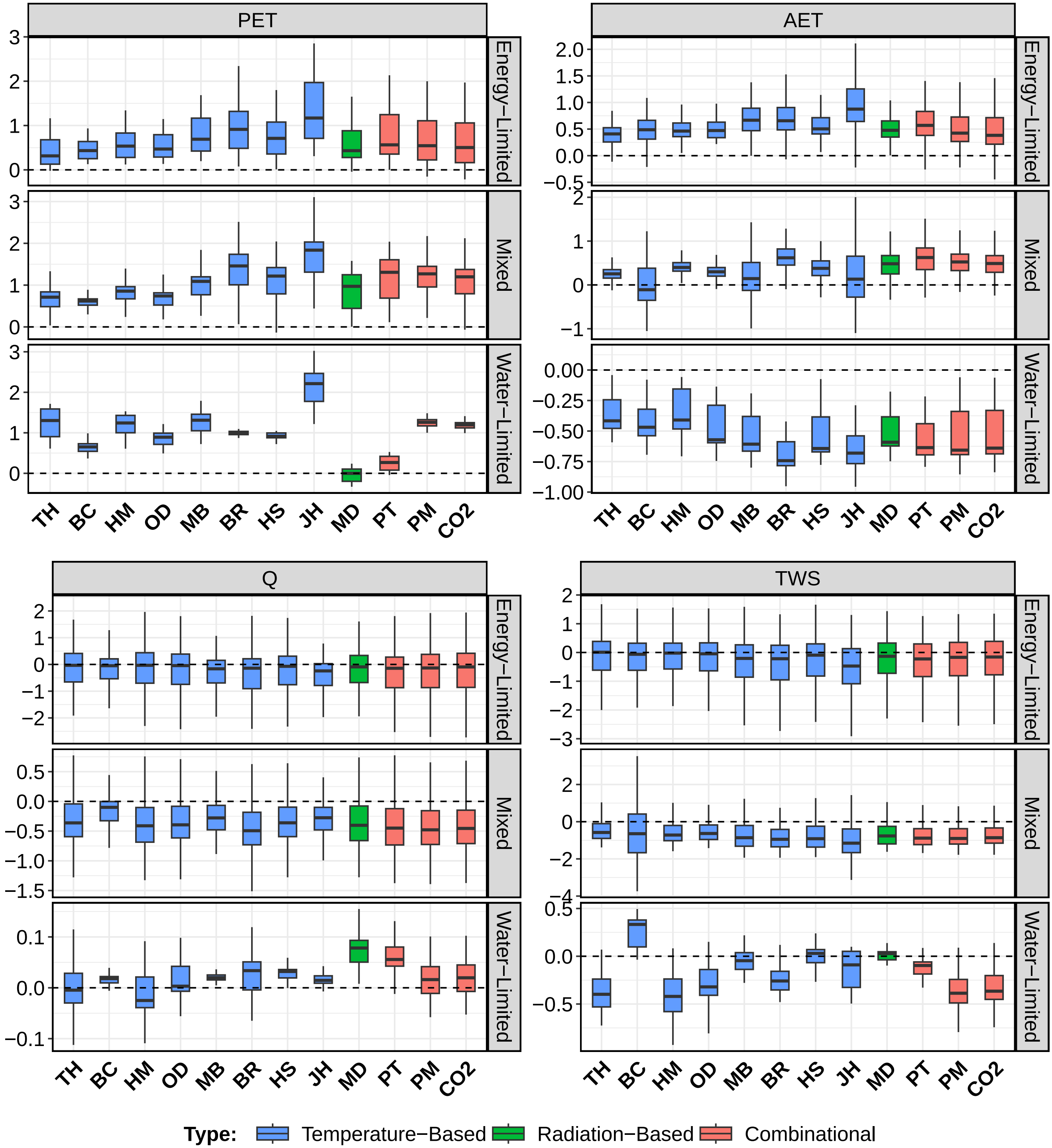


**Figure 3.** Boxplot represents ~~the seasonal (~~summer season (JJA) ~~)~~ trends of different PET methods for AET, and Q across three categories of ~~catchment~~catchments: energy-limited, mixed, and water-limited. The whiskers represent the 10th and 90th percentiles, and the box encompasses the 25th and 75th percentiles, with the median represented by the black line within the box. Abbreviations used for different PET methods are TH: Thornthwaite, BR: Bair-Robertson, BC: Blaney-Criddle, OD: Oudin, MB: ~~Meguinnes-Bordne~~Mcguinness-Bordne, HM: Hamon, HS: Hargreaves-Samani, JH: Jensen-Haise, MD: Milly-Dunne, PT: Priestley-Taylor, PM: Penman-Monteith, $CO_2$: Modified Penman-Monteith accounts $CO_2$. Trend units are in mm seas$^{-1}$ year$^{-1}$.

### 3.3 Catchment-wise DCI distribution across annual and seasonal scales

Even though there is ~~an~~ strong agreement across different PET methods in annual PET and AET trends, ~~the response~~ substantial variation exists in Q and TWS ~~varies considerably~~ responses (Figure 4). ~~For PET, all catchments demonstrate~~ All catchments exhibit strong positive DCI for PET, indicating that ~~a minimum of~~ at least 75% of ~~the PET methods exhibit~~ methods report a positive trend. ~~A similar patternis observed for AET, where higher positive DCI is noted in northern~~ AET follows a similar pattern, with high positive DCI values across northern (Scandinavia), central, ~~and a few southern European catchments. Conversely, few catchments in southern Europe exhibit lower DCI values, with very few showing strong~~ western, eastern, southeastern (Balkans), and parts of southern (Iberian Peninsula) Europe, except for a few southern catchments with low as well as negative DCI. ~~Q reflect~~ In contrast, Q reflects strong positive agreement in southern ~~catchments, with very few northern catchments showing positive DCI. The majority of northern catchments exhibit strong negative DCI . In contrast, central~~ regions but mostly negative DCI in northern and eastern Europe. Central and western European catchments are marked by both strong positive and negative DCI, with few catchments ~~with~~ showing disagreement among PET methods. TWS shows ~~disagreement among PET methods for southern, few central~~widespread disagreement, especially in southern, western, central, southeastern, and northern European catchments. Most ~~northern and centralEuropean catchments register~~ central, eastern, and northern catchments show strong negative agreement for ~~PET methods~~TWS. Overall, ~~strong positive concurrence is observed between the PET methods for the directional agreement in the~~ PET methods show high directional consistency for PET and AET~~trends, whereas mixed concurrency is seen~~, but diverge notably for Q and TWS. ~~Despite the higher positive DCI in PET , Q and TWS show higher negative DCI.~~

The above findings appear consistent when the selection of PET methods becomes balanced to eight methods: four temperature-based, one radiation-based, and three combinational methods, with the latter two fixed across all 70 possible method combinations. All the 70 combination results are consistent with the previous findings. For TWS, only a few catchments in southern and western Europe show strong negative DCI, consistent with earlier areas of disagreement (Figure S10). To assess the impact of weaker trends, we applied the DCI using only statistically significant trends. PET shows similar agreement patterns to the all-trend analysis, with minor disagreement in western Europe. For AET, disagreement emerges in most southern catchments and some in western, central, and eastern Europe, which previously showed agreement. Q and TWS exhibit widespread disagreement, largely due to weak trends across PET methods. Catchments with strong agreement under statistically significant trends align well with those under all-trend analysis. PET and AET trends remain predominantly positive in both statistically significant and all-trend evaluations.

To better understand the annual changes of the level of concurrence between ~~the~~ all PET method trends, we decompose them into sub-seasonal values. Figure 5 shows ~~that PET and AET demonstrate~~ strong positive agreement for PET and AET across all seasons, ~~whereas~~ while Q and TWS ~~predominantly show strong~~ exhibit predominantly negative agreement in central Europe, with evident regional ~~variations. During spring, summer, and autumn, PET demonstrates a higher positive DCI across most catchments , reflecting an overall increasing trend among PET methods~~variation. PET demonstrates high positive DCI in most catchments during spring, summer, and autumn, indicating consistent upward trends. In winter, ~~central and southern European~~

~~catchments have a high agreement, while others exhibit lower consistency. AET in central European catchments shows a higher positive DCI for~~ agreement is highest in central, eastern, and southern Europe, but weaker elsewhere. AET shows strong seasonal agreement in central Europe, particularly in winter, spring, and summer~~. However, northern European catchments reflect disagreement for AET across PET methods in winter, with a similar trend noted in some~~ , while disagreement emerges in northern and western regions during winter and in several southern and central catchments during autumn. ~~In southern Europe, AET demonstrates a strong negative DCI~~ AET also exhibits strong negative agreement in southern Europe during summer. ~~Across~~ Q consistently reflects a negative DCI across central Europe in all seasons, ~~Q shows a strong decreasing trend for all~~ although southern catchments show positive agreement in spring and autumn. TWS trends show persistent negative DCI in central Europe across seasons, while southern Europe sees a shift from strong positive agreement in spring to strong negative in summer, with weak agreement in winter and autumn. Northern Europe shows a mix of agreement and disagreement across all seasons.

Comparing sub-seasonal concurrence based on statistically significant trends (Figure S11) highlights both consistencies and divergences compared to the all-trend results. Strong positive DCI for PET and AET is observed in most catchments during spring, which aligns with the findings from all trends. In summer, PET ~~methods in most central European catchments. During springand autumn, however, most southern catchments exhibit positive agreement among PET methods. In~~ shows strong agreement in central and southern regions. In contrast, Q and TWS exhibit substantial disagreement across all seasons, ~~central European catchments generally display strong negative DCI~~ with most catchments showing low concurrence among PET methods~~for the TWS component. Southern European catchments exhibit a~~. Similar inconsistencies are observed for PET and AET in winter. Some central and eastern catchments consistently exhibit strong negative DCI ~~in summer and a strong positive DCI in spring, while showing poor agreement in winter and autumn. Northern European catchments also experience a combination of positive and negativeconcurrence.~~ for Q and TWS in both spring and summer, reflecting similar patterns in the all-trend case. Applying statistical significance thresholds often shifts catchments from strong agreement (positive or negative) to disagreement, primarily because trends are no longer significant. Notably, no catchments switch from strong positive to strong negative DCI (or vice versa), confirming the consistency between statistically significant and all-trend results.

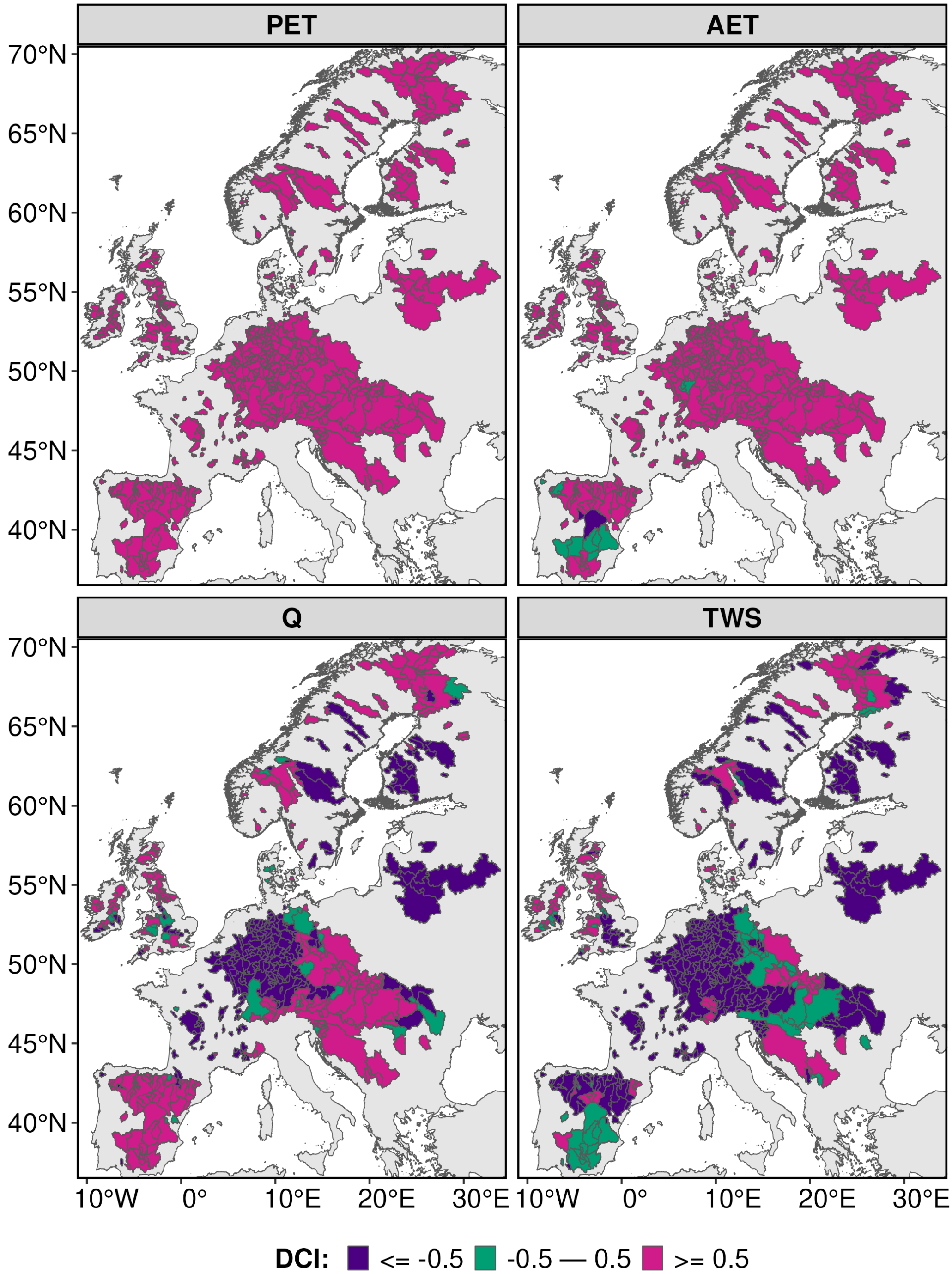


**Figure 4.** Spatial distribution of annual scale data concurrence index (DCI) for PET, AET, Q, and TWS. PET represents potential evapotranspiration, AET represents actual evapotranspiration, Q represents runoff at the outlet of the catchment and TWS represents total water storage.

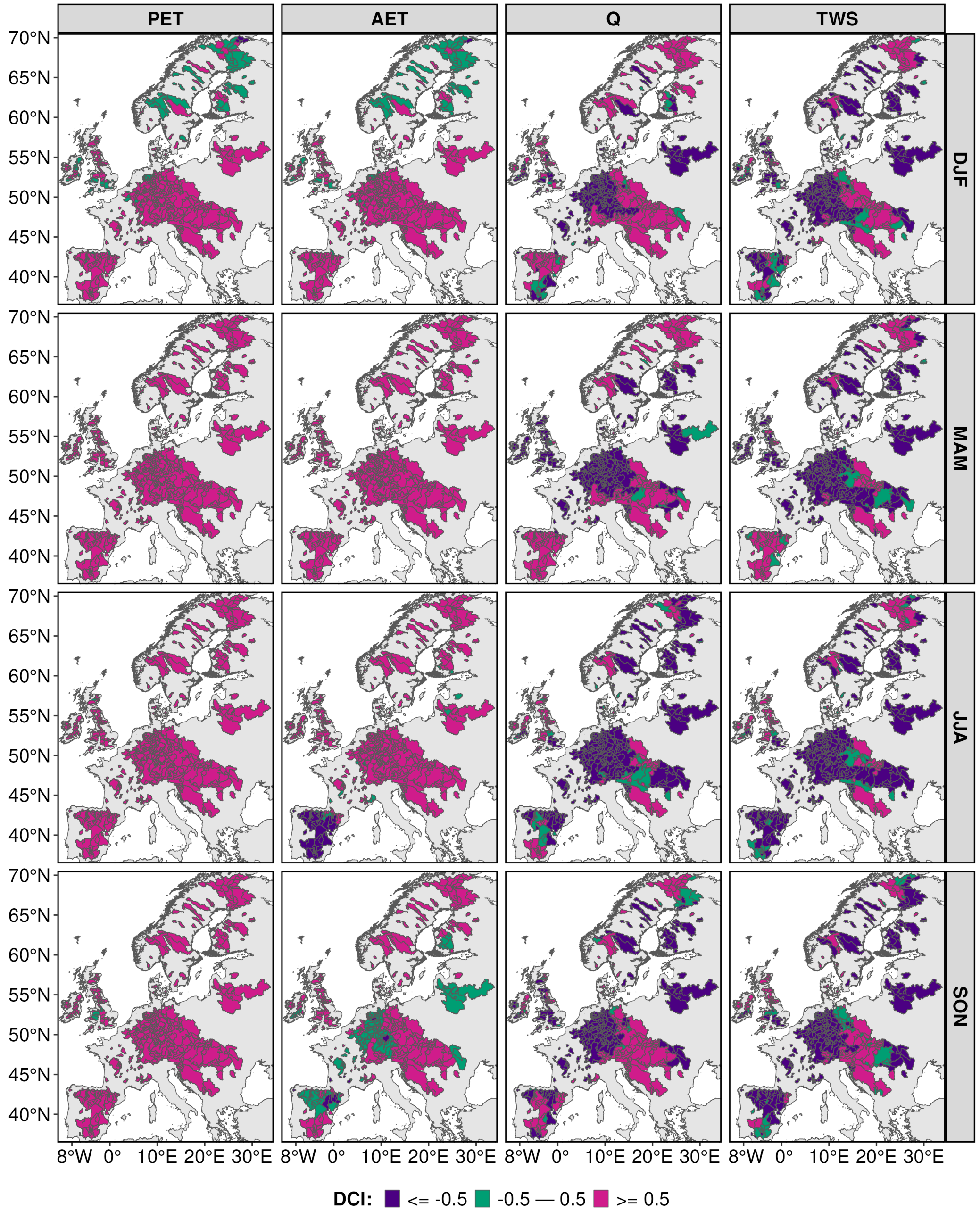


**Figure 5.** Spatial distribution of seasonal scale (winter (DJF), spring (MAM), summer (JJA), and autumn (SON)) DCI for PET, AET, Q, and TWS. Where DCI represents data concurrence index, PET represents potential evapotranspiration, AET represents actual evapotranspiration, Q represents runoff at the outlet of the catchment and TWS represents total water storage.

### 3.4 PET methods and ~~combination~~ patterns of hydrological cycle ~~components~~ at annual and seasonal scales

In the previous section, we compared ~~different PET method estimates and their impact on each hydrological component (~~PET methods and their influence on individual hydrological components (P, AET, Q, and TWS), ~~as well as the agreement between them~~including agreement among PET methods. Here, we ~~demonstrate the overall impact~~ assess the influence of PET methods on ~~possible combinations of hydrological cycle component changes. Most catchmentsfall within the first five hydrological cycle combinations on annual scale (~~patterns of key hydrological components across European catchments, identifying the most prevalent trend patterns where components (P, AET, Q, and TWS) concurrently increase or decrease. For instance, one pattern involves all components showing positive trends. Figure 6 ~~). The Blaney-Criddle method has the highest catchment count for combinations featuring positive trends across all hydrological cycle components, while temperature-based methods account for more catchmentsthan combinational methods (Figure 6). For combination featuring negative~~ summarizes these patterns, presenting the average and total number of catchments associated with each PET method. The analysis includes only the five most frequent patterns, covering the majority of catchments, excluding those patterns with minimal representation.

Most European catchments exhibit increasing trends across all hydrological ~~cycle componentsexcept AET, the Blaney-Criddle method and Bair-Robertson exhibit the lowest and highest catchment count respectively (Figure 6). For combinations showing positive trends in PRE and AET and negative trends in Q~~components. The second most common pattern involves a decrease in P, Q, and TWS, ~~temperature-based methods generally represent fewer catchments than~~ with an increase in AET. Temperature-based PET methods generally account more number of catchments for these patterns than radiation and combinational methods, ~~with the Blaney-Criddle and Bair-Robertson methods exhibiting the lowest and highest counts, respectively~~ though exceptions exist, such as Baier-Robertson and Blaney-Criddle. Catchment distributions are spatially consistent across methods for these two patterns (Figure ~~6). In cases with negative trends across all components,~~ S14), with Blaney-Criddle ~~and Bair-Robertson again demonstrate the highest and lowest catchment counts~~ showing the highest counts for patterns involving uniform increases or decreases. In case of statistically significant trends, only the all-positive pattern is prominent (Figure S13), with an average of 80 catchments. ~~The last five combinations of hydrological cycle in Figure 6 contribute very limited number of catchments, though the Blaney-Criddle method tends to capture more catchments within these combinations. Across combinations with positive AET and negative TWS trends, the Blaney-Criddle method accounts for the fewest catchments , while the Bair-Robertson method has the highest count~~This pattern reveals notable differences between temperature-based and combinational methods. Overall, PET methods differ in the number of catchments assigned to each pattern; however, combinational methods consistently demonstrate similar catchment counts across most patterns.

~~Furthermore, we examine the seasonal patterns of PET method selection across different hydrological cycle component combinations. The initial five combinations remain consistent across seasons, while the last five are different due to the absence of certain hydrological combinations. In summer, for the combination featuring positive trends across all hydrological cycle components, temperature-based methods capture more catchments than radiation-based and combinational methods, with the Blaney-Criddle method accounting for the largest number (Figure S6~~Seasonal analyses reveal distinct PET method preferences across hydrological component patterns. During winter, spring, and autumn, the prevailing pattern involves decreasing P, Q, and TWS with increasing AET (Figures S15, S16, S18). In ~~cases where all components except AET show decreasing trends, combinational and radiation-based methods dominate, covering more catchments than temperature-based methods. For catchments with positive PRE and AET but negative~~ contrast, summer is characterized by increased P and AET, and decreased Q and TWS~~trends, the~~. Substantial variation is observed among PET methods for each hydrological cycle pattern across all seasons. For instance, a pattern where all components exhibit positive trends, Baier-Robertson captures the fewest catchments, while Blaney-Criddle ~~method demonstrates the least count.~~ captures the most. The pattern reverses for the combination of decreased P, Q, and TWS with increased AET. Blaney-Criddle consistently represents the highest number of catchments in all-positive and all-negative patterns during spring, summer, and autumn. Combinational methods generally ~~exhibit less variability in~~ show stable catchment counts, ~~while~~ whereas temperature-based methods ~~demonstrate more pronounced variations across hydrological cycle component combinations. In spring,~~ exhibit greater variability. Statistically significant trends, however, are associated with very few catchments, and no single pattern dominates.

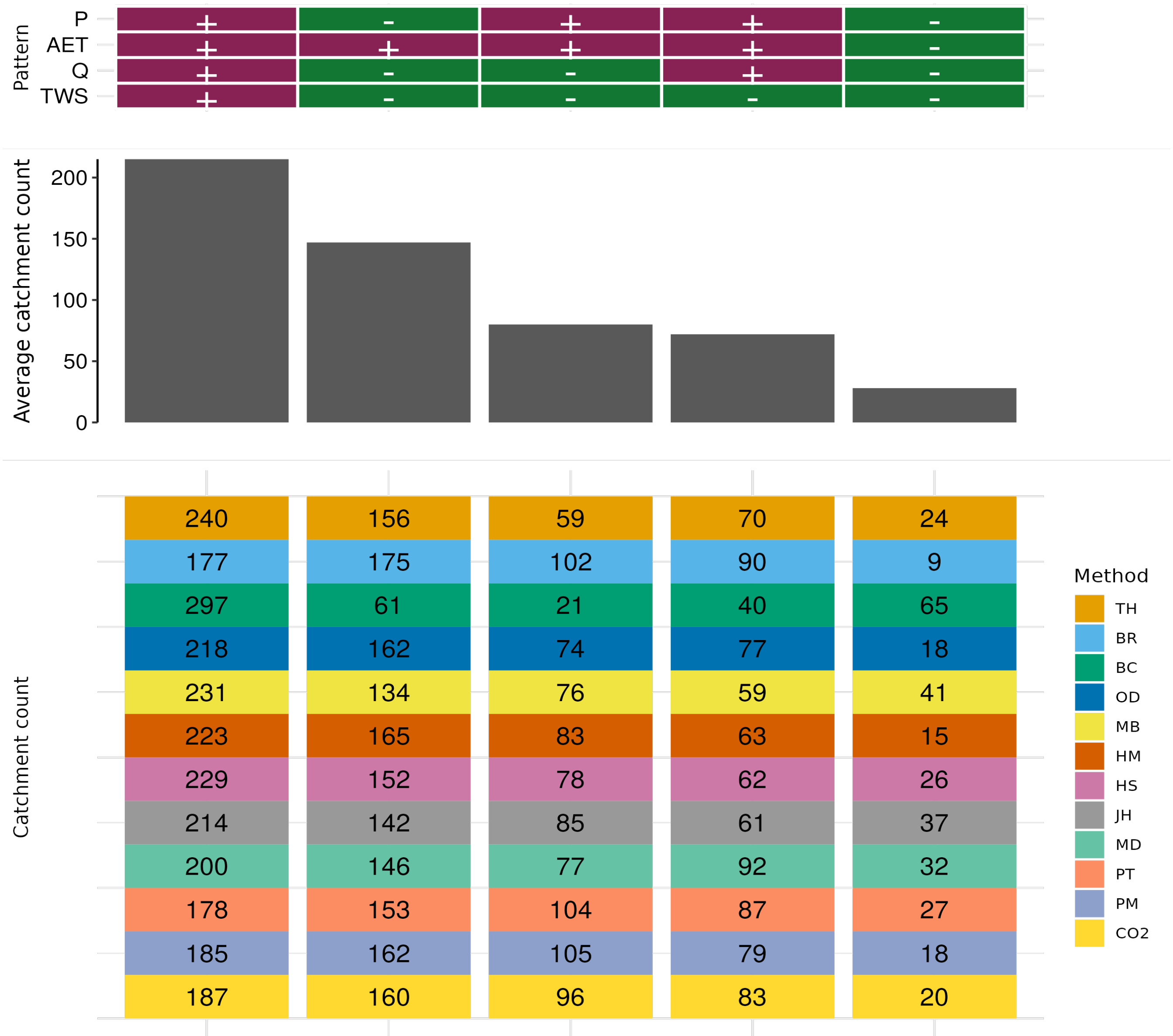


**Figure 6.** Pattern of different hydrological cycle components and the corresponding influence of PET methods on an annual scale. The first panel represents different patterns of hydrological cycle components. Each vertical column in this table corresponds to one pattern of hydrological cycle components. For example, the first column is filled with '+' signs, indicating that all hydrological components (P, AET, Q, and TWS) exhibit positive changes. The '+' and '-' signs denote positive and negative changes in the respective components. In the second panel, each bar represents the average number of catchments for each hydrological cycle pattern. The third panel shows the number of catchments associated with each PET method for the corresponding hydrological cycle patterns. The color of each cell represents a specific PET method, and each column aligns with the hydrological cycle pattern represented in the corresponding column of Panel One. Where P is precipitation, AET is actual evapotranspiration, Q is runoff, and TWS is total water storage. Abbreviations used for different PET methods are TH: Thornthwaite, BR: Bair-Robertson, BC: Blaney-Criddle, OD: Oudin, MB: McGuinness-Borden, HM: Hamon, HS: Hargreaves-Samani, JH: Jensen-Haise, MD: Milly-Dunne, PT: Priestley-Taylor, PM: Penman-Monteith, $CO_2$: Modified Penman-Monteith accounts $CO_2$.

### 3.5 Relationship of PET and precipitation

Precipitation is an important component of the ~~Blaney-Criddle method follows a similar pattern as of summer season, being the only method that shows decreasing trends in all hydrological cyclecomponents (Figure S5). In the autumn (Figure S7), temperature-based methods again capture the largest number of catchments with positive trends in all components, while combinational methods have the fewest. For catchments with negative trends across all hydrological cycle components except AET, Baier-Robertson method shows the most catchments while the Blaney-Criddle method shows the least. Across~~ hydrological cycle. Figure 7 shows the changes in precipitation (P) without considering statistical significance. Annually, positive P trends dominate northern, western, southern, and southeastern catchments, while central Europe shows mixed positive and negative trends. In western Europe, a few catchments exhibit decreasing P trends (Figure 7A). Seasonally, southern catchments experience increased P in winter, spring, and autumn but declines in summer. Southeastern Europe shows consistent P increases across all seasons, ~~most catchments in summer exhibit positive trends for PRE and AET but negative trends for~~ while eastern Europe exhibits negative P trends in summer and autumn and positive trends in winter and spring (Figure 7A). P demonstrates a higher correlation with Q and TWS across all catchment categories (Figure ~~S6). In contrast, winter (Figure S4), spring~~ 7C). This suggests that P has a greater influence on Q and TWS than PET. In energy-limited catchments, AET is mainly driven by PET. In mixed catchments, both P and PET contribute to AET. In water-limited catchments, AET is mainly influenced by P. When we consider statistically significant P trends, only 129 catchments show significant trends at the annual scale. Across seasons, the number of statistically significant catchments varies from 20 to 61 (Figure ~~S5), and autumn (Figure S7) generally exhibit decreasing trends in all components except AET~~S18). Despite the limited number of statistically significant catchments, our findings regarding the influence of P and PET on AET, Q, and TWS remain consistent with all trends.

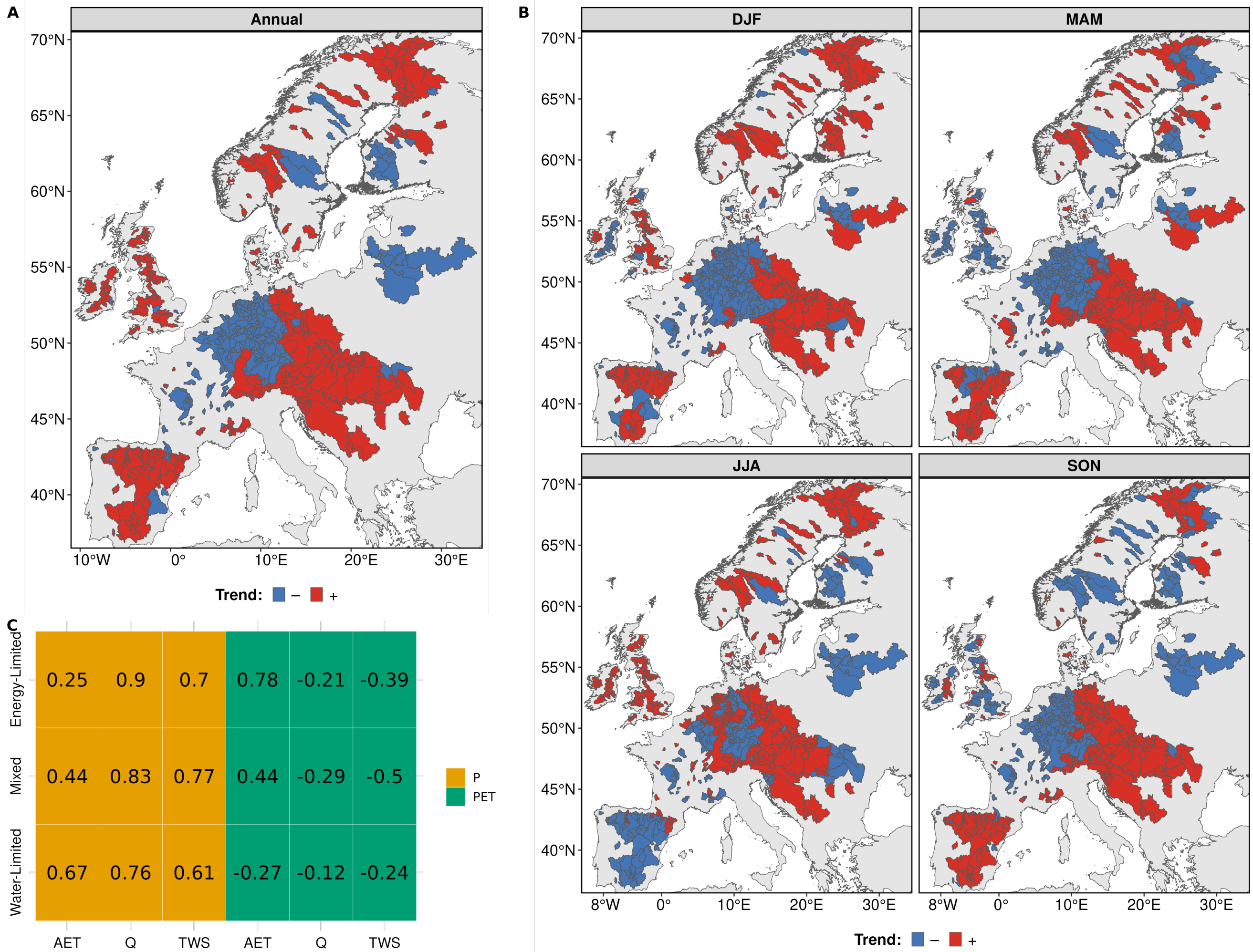


**Figure 7.** ~~Combination~~ Spatial variation of ~~different~~ precipitation (P) trends and their relationship with other hydrological ~~cycle~~ components ~~and~~ across catchment categories. Panel A shows the ~~corresponding influence~~ spatial distribution of ~~PET methods on an~~ increasing and decreasing annual ~~scale~~P trends. ~~PRE+, AET+, Q+, and TWS+ represent an~~ Panel B illustrates the seasonal variation in increasing ~~trend for PRE,~~ and decreasing P trends. Panel C represents the median correlation between P and AET, Q, and TWS~~respectively. Similarly~~, ~~PRE-, AET-, Q-~~ as well as between PET and ~~TWS- represent a decreasing trend. Where PRE is precipitation,~~ AET~~is actual evapotranspiration~~, Q~~is runoff~~ , and TWS~~is total water storage. Abbreviations used~~ , for ~~different~~ each catchment category across all PET methods~~are TH: Thornthwaite, BR: Bair-Robertson, BC: Blaney-Criddle, OD: Oudin, MB: McGuinness-Borden, HM: Hamon, HS: Hargreaves-Samani, JH: Jensen-Haise, MD: Milly-Dunne, PT: Priestley-Taylor, PM: Penman-Monteith,~~ ~~$CO_2$: Modified Penman-Monteith accounts $CO_2$~~.

## 4 Discussion

### 4.1 Trends in P, PET, AET, Q, and TWS across Europe

~~The estimates from~~ PET methods consistently ~~show an upward trend over European catchments annually~~ exhibit upward annual trends across European catchments (Figure 2), ~~which is in agreement with the work of Anabalón and Sharma (2017), who also reported such an increase across Europe~~ aligning with findings by Anabalón and Sharma (2017), who reported similar increases using diverse PET datasets. ~~The~~ Among the methods, Jensen-Haise ~~method consistently yields higher absolute valuesfor European catchments, in agreement with Hanselmann et al. (2024) research in Poland. The Jensen-Haise method shows the highest trend in all categories. The Jensen-Haise method~~ consistently produces the highest absolute values, a pattern observed in various regions by other studies (Kingston et al., 2009; Hanselmann et al., 2024; Seiller and Anctil, 2016). This method relies on temperature and extraterrestrial radiation data~~. Extraterrestrial radiation remains constant each year. Therefore~~, with the latter remaining constant annually. Thus, the observed trends are primarily ~~driven by temperature variations~~temperature-driven. Shi et al. (2023b) found that Jensen-Haise trends align closely with Penman-Monteith, which is used as a benchmark for evaluating other PET methods. ~~We observed~~ In our study, we observe a notable distinction between the Jensen-Haise and Penman-Monteith PET methods~~. The differences in trends between these methods stem mainly from~~ , primarily driven by differences in meteorological forcings. ~~In the Jensen-Haise method, temperature is a significant factor in the trend. For Penman-Monteith, however, radiation is the main factor~~While temperature is the dominant factor in Jensen-Haise, Penman-Monteith is more influenced by radiation, followed by temperature, vapor pressure deficit, and wind speed ~~(Maček et al. (2018)). The~~ (Maček et al., 2018) . Conversely, the Milly-Dunne method consistently ~~demonstrates~~ shows the lowest trends in water-limited catchments~~. This trend, influenced by net radiation, remains lower in the south of Europe (Pfeifroth et al. (2018)). In the presented work, we do not perform sensitivity~~ , with lower net radiation-driven trends in southern Europe (Pfeifroth et al., 2018). Sensitivity analyses of each PET method with meteorological forcings ~~, as this falls outside~~ are beyond the scope of ~~the paper~~this research.

The AET trend is proportional to the PET trend in energy-limited and mixed catchments. In water-limited regions, there is not enough water to evaporate, and it is mainly governed by available water (~~PRE) Bruno and Duethmann (2024)~~P) (Bruno and Duethmann, 20 . This results in a notable decline in AET when compared to PET. ~~As reported by Anabalón and Sharma (2017) , AET tends to correlate more closely with~~ Anabalón and Sharma (2017) similarly reported stronger correlations between AET and PET in energy-limited regions~~and with precipitation~~ , while AET in water-limited ~~regions. However, their analysis did not consider different~~ areas aligned more closely with precipitation. Notably, their study did not differentiate between PET methods but ~~used various existing datasetsinstead.~~ relied on existing datasets. In our analysis, AET exhibits the same directional changes as precipitation at both annual and seasonal scales in water-limited catchments (Figure ~~S9~~7). Despite differences among catchment categories, PET methods demonstrate strong positive or negative agreement in AET trends.

~~In general, the runoff (Q) trend varies based on the PET method across~~ Runoff trends vary with PET method selection in water-limited catchments ~~, whereas~~ but remain largely insensitive to PET methods in energy-limited and mixed ~~catchment show insensitivity to PET method selection. This is in line with earlier work, e.g., Bai et al. (2015); Oudin et al. (2005); Seiller and Anctil (2016) who reported that runoff was generally insensitive~~ catchments. This aligns with previous findings (Bai et al., 2015; Oudin et al., 2005; Seille

, which reported insensitivity of runoff to PET formulations. This insensitivity is ~~primarily attributed to the calibration of hydrological models, where the impact of PET models is offset by the parameterizationof the hydrological model. Surprisingly, even though we did not calibrate the hydrological model for individual PET methods, we observed insensitivity to PET methods~~ often attributed to hydrological model calibration, where PET impacts are counterbalanced by parameterization. Notably, despite the absence of individual PET method calibration in our study, runoff in energy-limited and mixed catchments remained insensitive to PET variation. This is ~~primarily because the trend in precipitation is strongly correlated~~ likely due to the strong correlation between precipitation and runoff trends (Figure ~~S11)with the trend in runoff (Q) (Berghuijs et al., 2017)~~7c), which often outweighs the impact of PET ~~(Anabalón and Sharma, 2017)~~(Berghuijs et al., 2017; Anabalón and Sharma, 2017). The strong negative agreement ~~between~~ among PET methods for runoff ~~(Q) across~~ in central European catchments ~~was due~~ is similarly related to their strong correlation with precipitation (Figure ~~S9~~7a).

Total water storage (TWS) appears to be insensitive in energy-limited and mixed catchments, while it exhibits variability in the trend of different PET methods in water-limited catchments. Bai et al. (2016) observed that TWS in energy-limited catchments is more strongly impacted by PET than in water-limited catchments, though their study focused solely on Chinese catchments. More recently, Boeing et al. (2024) reported a decline in TWS over Germany, consistent with our findings that TWS decreases in energy-limited and mixed catchments for all PET methods. In hydrological models, TWS compensates for long-term changes, i.e., higher PET results in lower TWS in energy-limited catchments, whereas water-limited catchments are primarily governed by precipitation (Bai et al., 2016). This is in line of our findings on the opposite TWS trends relative to AET and a stronger agreement among PET methods in energy-limited and mixed catchments.

~~A large number of catchmentsacross Europe exhibit positive changes in the hydrological cycle; however, variations exist depending on the PET methodsemployed. Teuling et al. (2019) found increasing PRE, AET and Q in central-west Europe, while these hydrological cycle components decreased in the Mediterranean region using the Penman method for PET estimation~~Total water storage appears to be insensitive in energy-limited and mixed catchments, while it exhibits variability in the trend of different PET methods in water-limited catchments. Bai et al. (2016) observed that TWS in energy-limited catchments is more strongly impacted by PET than in water-limited catchments, though their study focused on Chinese catchments. More recently, Boeing et al. (2024) reported a decline in TWS over Germany, consistent with our findings that TWS decreases in energy-limited and mixed catchments across all PET methods. In hydrological models, TWS often compensates for long-term changes, with higher PET leading to lower TWS in energy-limited catchments, while precipitation governs TWS in water-limited regions (Bai et al., 2016). Our results ~~confirm that the Penman method identifies fewer catchments showing these trends compared to temperature-based PET methods . This variation will cause discrepancies in the conclusions of hydrological cycle intensification studies. Similarly, many catchmentsdemonstrate a drying hydrological cycle (positive change in AET and negative changes in Q, TWS, and PRE). The number of catchments demonstrating these trends varies depending on the choice of PET method. Massari et al. (2022) found that over Europe runoff deficit are more pronounced in water-limited regions due to increased AET, whereas~~ further underscore these patterns, revealing inverse TWS trends relative to AET and stronger agreement among PET methods in energy-limited ~~catchments exhibit smaller deficits. During these drying conditions, AET is further influenced by reductions in TWS (Massari et al., 2022).~~ and mixed catchments.

Precipitation generally increases across most catchments, with annual and seasonal patterns largely consistent. Exceptions occur in summer for the southern catchment and spring for eastern catchments, consistent with findings by Caloiero et al. (2018); Markonis . Precipitation has a stronger influence on the hydrological cycle than AET in low and mid-latitude regions compared to higher latitudes (Zhang et al., 2019). This spatial distribution explains the stronger correlations observed in mixed and energy-limited catchments, predominantly located in mid and high latitudes. Globally, Q is more sensitive to P than PET, supporting the observed decline in Q despite rising PET across many catchments (Berghuijs et al., 2017).

### 4.2 Methodological sensitivity of PET estimation methods

PET methods vary in both absolute values and trends, even within the same category, due to structural and empirical differences in their formulations. For instance, while Jensen–Haise and Hargreaves–Samani both incorporate extraterrestrial radiation and air temperature, Hargreaves–Samani includes a diurnal temperature range term absent in Jensen–Haise. Thornthwaite, in contrast, relies solely on a heat index derived from monthly temperature. Among combinational methods, Penman–Monteith and its modified version differ by the inclusion of a $CO_2$ concentration term. Structurally similar methods like McGuinness–Bordne, Oudin, and Jensen–Haise diverge primarily due to empirically derived constants tailored to specific regions or climates (Proutsos et al., 2023 . Moreover, catchment size does not significantly influence the overall findings related to PET and hydrological components (Figure S20), consistent with Tang et al. (2023), who observed that spatial averaging over larger catchments reduces uncertainty and enhances reliability.

Even though our study's experimental design varies from the global analysis of Pimentel et al. (2023), both utilize large-scale hydrological models at the basin scale. Pimentel et al. (2023) ~~compare three different~~ evaluated three PET methods to ~~choose the best PETmethod to estimate PET~~identify the optimal approach for estimating PET, AET, and Q ~~in the hydrological model~~across the Globe. Across Europe, they reported that Jensen-Haise in northern Europe (energy-limited), Hargreaves in central (mixed), and Priestley-Taylor alongside Hargreaves-Samani in southern Europe (water-limited) perform better to estimate PET. For AET estimation, ~~the~~ Jensen-Haise ~~method outperforms~~ remains the preferred method in northern Europe~~(energy-limited), while~~ , with Hargreaves-Samani ~~leads~~ leading in central Europe ~~(mixed)~~ and Priestley-Taylor in southern Europe~~(water-limited). They also argue that~~ . Priestley-Taylor ~~is predominantly~~ was also deemed the most effective ~~method~~ for runoff estimation. In ~~our study, we found a consistently notable~~ contrast, our findings indicate a consistent distinction between Jensen-Haise and Penman-Monteith for PET across all catchment types~~for PET~~. Similarly, for AET, Jensen-Haise consistently shows higher trends than Penman-Monteith in both energy-limited and water-limited catchments. ~~We observed that changes~~ For runoff, Priestley-Taylor and Penman-Monteith exhibit similar patterns in energy-limited and mixed catchments~~are similar for runoff (Q) when using the Priestley-Taylor and Penman-Monteith methods. However, in~~ ; however, in water-limited catchments, Priestley-Taylor ~~leads to greater changes than ones estimated by the Penman-Monteith method~~yields more pronounced changes than Penman-Monteith.

### 4.3 Patterns of hydrological cycle components

When we look at the combination of changes among the hydrological cycle components (P, AET, Q, TWS) across European catchments, two dominant patterns of changes are observed: a water cycle intensification pattern of simultaneous increase in all components and an aridification pattern of simultaneous decline in all except AET. These two patterns can be seen across over 60% of European catchments. However, when focusing only on significant trends, the pattern of intensification becomes more dominant, declining approximately to 15% (80 catchments) compared to 0.4% (2 catchments) aridification. The intensification pattern aligns with findings by Teuling et al. (2019), who reported rising P, AET, and Q in central-western Europe and declines in these components in the Mediterranean. Their analysis, based on the Penman–Monteith method, also indicated fewer catchments classified under this pattern compared to temperature-based methods, which show stronger responses. This is a vivid example of how PET method selection ~~impacts the hydroclimatic state of the catchment. The hydroclimatic state of a catchment is commonly classified into two categories~~ could amplify or dampen our estimates about hydrological cycle intensification.

On the contrary, the aridification pattern (decrease in P, Q, and TWS and increase in AET) suggests that water reserves are being depleted to sustain evapotranspiration, a mechanism particularly evident in water-limited ~~and~~ regions. Bruno and Duethmann (2024) noted that rising atmospheric demand, without sufficient water supply, results in reduced Q and TWS. Massari et al. (2022) similarly reported that increasing AET contributes to Q reductions in water-limited regions. Even with decreasing P and Q, continued declines in TWS appear to support increases in AET (Massari et al., 2022). For Europe, this is also very relevant to the compound warm season droughts that have been reported to increase since the beginning of the century Markonis et al. (2021) , as well as with the conditions that favor the onset and propagation of flash droughts Shah et al. (2023). Since the evaporative demand is expected to further increase in the future Rakovec et al. (2022), it is essential to acknowledge the uncertainties due to PET method selection.

### 4.4 Implications of PET method selection to hydroclimatic regime classification

The selection of PET methods significantly influences the hydroclimatic classification of catchments. Traditionally, catchments are categorized as water-limited or energy-limited ~~catchments~~ based on the aridity index. ~~In our research, we introduce a third catchment category~~However, our study introduces a third category, termed "mixed", which ~~does not represent any physical basis . However, it underscores the crucial importance~~ lacks a physical basis but highlights the critical role of PET method selection ~~specifically for the mixed catchment category. For instance, using~~ in defining catchment types. This is when, a PET method that consistently ~~generates a higher PET estimate may change~~ estimates higher values may shift a catchment from energy-limited to water-limited, whereas a method ~~that produces a lower PET estimate can shift it from water-limited to energy-limited~~with lower estimates could reverse this shift. This underscores the importance of method selection, where variations in PET estimates can alter the hydroclimatic classification. Similarly, Zhang et al. (2016) introduced a ~~slightly different and not well-known classification called "~~less common classification termed "equitant", which applies a single PET method to calculate the aridity index. Such ~~variations could cause discrepancies in the results. For example~~methodological differences can lead to inconsistencies in catchment classification. For instance, Kuentz et al. (2017) uses the Jensen-Haise method, Ajami et al. (2017) ~~uses~~ utilizes the Priestley-Taylor method, and Zhang et al. (2016) ~~uses~~ adopt the Penman method

~~to calculate the aridity index to classify catchments. We observed catchments shifting from the mixed to the energy-limited category after~~ for aridity index estimation. In our analysis, excluding PET methods that consistently ~~overestimate PET . For instance, removing Blaney-Cridle~~estimate higher PET values, such as Blaney-Criddle, Jensen-Haise, and McGuinness-Bordne, resulted in 42% of ~~the total catchments shift from mixed category to~~ catchments transitioning from the mixed to the energy-limited ~~catchment~~ category (Figure ~~S8). Shifts in catchments,~~ S23). Detailed catchment shifts based on various combinations ~~,~~ ~~are presented~~ are outlined in Table S1.

### 4.5 Limitations and future research

Our study comes with certain limitations that pave the way for future research. ~~First of all, it is limited by~~ One key limitation is the uncertainty associated with the input data used to calculate ~~the~~ PET methods. ~~However, previous detailed studies have investigated the uncertainties related to these data (Hua et al., 2020; Guo et al., 2017). They reported that PET~~ Previous studies have shown that temperature-based methods are sensitive to ~~their input data: temperature-based methods to~~ temperature, radiation-based methods to radiation, and ~~combination methods predominantly to both temperatureand radiation, as well as to relative humidity~~combinational methods to multiple variables, including temperature, radiation, relative humidity, and wind speed ~~. Additionally, we~~ (Hua et al., 2020; Guo et al., 2017). In addition, we use a monthly time step, which tends to mask the influence of PET method selection on AET extremes. These extremes can behave differently from the mean state of AET, potentially leading to different implications for changes in hydrological cycle fluxes compared to those based on mean AET values (Markonis, 2025). Finally, we limited our analysis to one precipitation product to isolate the specific impact of the PET method. However, ~~it is well known that precipitation is~~ precipitation is widely recognized as the most sensitive meteorological input, ~~Voisin et al. (2008); Mazzoleni et al. (2019) have extensively studied the uncertainties related to precipitation.~~ with extensive studies highlighting its uncertainties (Mazzoleni et al., 2019; Markonis et al., 2024). Our findings also confirm its dominant influence over PET in certain catchment categories (Figure 7). This identifies a potential gap for exploring the combination of precipitation with PET for ~~accurate simulation~~ more accurate simulations of hydrological cycle components. Although we selected meteorological datasets with comparatively lower uncertainty, data quality, whether from observational or reanalysis sources, remains an issue for hydrological assessments. While our focus was on methodological comparisons among PET methods, future research could benefit from multi-source assessments to enhance the robustness and reliability of hydrological modeling.

~~Often large-scale hydrological modelsuse~~ Large-scale hydrological models, including mHM, typically rely on default parameterization. ~~This research can be further extended by incorporating calibration for all three hydrological components (AET, Q, and TWS) in areas with data availability. This study~~ In mHM, these parameters were initially developed using German basins, as outlined by Samaniego et al. (2010); Kumar et al. (2013a). Since then, mHM has been extensively tested across various basins and hydrological variables (Rakovec et al., 2016; Samaniego et al., 2019; Boeing et al., 2024). For instance, Rakovec et al. (2016) evaluated discharge simulations across 400 European catchments using 36 parameter sets, demonstrating consistent model performance regardless of parameterization, thereby reinforcing confidence in the model's reliability. Similar approaches have been applied in global water models; Beck et al. (2017) employed ensemble parameters derived from ten catchments, while Kumar et al. (2013a) tested default parameters from European basins across 80 American catchments with varying climatic conditions. While these studies demonstrate the robustness of default parameterization, investigating how PET-specific calibration affects hydrological trends could provide valuable insights for future research. Notably, the Hargreaves-Samani PET method, used in developing these parameters in mHM, demonstrated best model performance in this study but did not consistently stand out compared to other PET methods in trend analysis across hydrological components. Moreover, the study is confined to temperate ~~climate~~ European catchments, ~~and there is a potential gap in extending this research to~~ leaving a gap in assessing arid and tropical climates, ~~which could yield different and interesting results. Although~~ where distinct patterns may emerge. While the Modified Penman-Monteith method accounts for $CO_2$, it did not ~~show~~ exhibit substantial differences compared to the ~~Penman method. Further~~ Penman-Monteith method, indicating the need for further exploration of this method~~, along with others,~~. It would be interesting to assess their impact under changing climate ~~condtions~~conditions and their implications for trend assessments.

## 5 Summary and conclusions

Twelve PET methods were used to evaluate their impacts on changes in the components of the hydrological cycle using the mesoscale Hydrological Model (mHM). These methods were applied across 553 European catchments, which vary in size and include different European climate types. These catchments were classified as water-limited, energy-limited, and mixed catchments based on their aridity index. ~~To analyze changes~~ Changes in PET and hydrological components ~~, we employed Sen's slope estimate, and to~~ were assessed using Sen's slope for trend magnitude and the Mann–Kendall test for statistical significance. We analysed our results under two cases: first, using all trends; and second, using only statistically significant trends. To assess the agreement between different PET methods, we used the ~~modified~~ data concurrence index for the period 1980 to 2019. The results demonstrate that the choice of PET method can substantially affect changes in AET, Q, and TWS, especially in water-limited and mixed catchments, with smaller changes and greater variability observed in water-limited catchments on an annual scale. Seasonal variations in changes and agreement between PET methods were also observed, as discussed in detail in sections 3.2 and 3.3. In general, there is agreement among the different methods that, since ~~1981~~1980, PET and AET are increasing over Europe, while runoff and total water storage exhibit mixed fluctuations depending on the method used and the catchment latitude. The key findings of our study are summarized as follows:

1. PET is increasing across European catchments. The majority of the PET methods indicate a positive trend in all categories of catchments, but the increase rates differ among the methods employed.

2. At the annual scale for all trends, the Jensen-Haise PET method stands out by consistently showing the highest trends for PET and AET across all catchment categories. The Milly-Dunne (energy-based) method is notable as the only one to exhibit a negative trend for water-limited catchments. Regarding Q and TWS, the PET methods display different changes and variability, with no method consistently showing either the lowest or highest trends for these hydrological components. Conclusions remain consistent when considering only statistically significant trends.

3. The trend patterns between PET and AET are similar across all methods for the hydrological cycle components. However, Q and TWS do not exhibit the same pattern and appear to be less sensitive to choice of PET methods. Most PET methods agree on the trend direction for PET and AET, but in a few catchments, the trends for Q and TWS show an opposite direction. The negative trends in Q and TWS are primarily due to negative precipitation trends, which have a stronger impact on these components in all catchment categories.

4. At the seasonal scale, PET methods reveal different trends for PET, with no method consistently showing the highest or lowest trends across all seasons. However, the Jensen-Haise method shows the highest PET trends during spring, autumn, and summer. AET trends follow a similar pattern to PET in all seasons. The PET methods show strong agreement on trend direction for central and southern European catchments, especially for PET, but there is less agreement for northern catchments in winter. Strong negative agreement is found for Q and TWS in summer and spring, while disagreement is observed for AET in central and southern catchments during autumn.

5. The summer season contributes more to the annual PET trends than any other season across all catchment categories. Similarly, for AET, the summer season has a higher contribution to the annual AET trend in energy-limited and water-limited catchments. For runoff (Q), the spring season contributes more in mixed and water-limited catchments. For TWS, the spring season has a higher contribution in energy-limited catchments.

6. Overall, ~~The~~ the magnitude of trends varied between PET methods for PET and the hydrological components (Q, AET, and TWS). The use of a specific PET method in a hydrological model can notably affect studies focused on the hydrological cycle.

7. Precipitation ~~is the primary factor influencing the trends of~~ primarily governs trends in all hydrological components and ~~is responsible for the overall direction of hydrological cycle components (AET, Q, and TWS)~~catchment types, except for AET in energy-limited catchments, which is largely influenced by the choice of PET variations.

8. The choice of PET method substantially influences hydrological patterns across European catchments on both annual and seasonal scales. Combinational methods generally account for fewer catchments than temperature-based methods in dominant hydrological patterns. This observation remains consistent for statistically significant trends as well.

9. In the case of statistically significant trends, the conclusions remain consistent with those from the all-trend analysis. The only difference is the reduced number of catchments, primarily omitted due to weaker trends.

Our research demonstrates the critical role of PET method selection and its implications for quantifying fluctuations in the hydrological cycle. Our findings reveal that two methods notably deviate from the others. Specifically, the Jensen-Haise method shows higher trend values, while the Milly-Dunne method exhibits lower trends in water-limited catchments. Consequently, we recommend exercising caution when applying these methods as they appear to be outliers. Despite these variations, the PET methods generally agree that atmospheric moisture demand is increasing across Europe, reflecting recent shifts in temperature and radiation. ~~Given the differences~~ The observed variability in trend magnitudes ~~across methods, we encourage the use an~~

~~ensemble of PET formulations in the assessment of changes in the water cycle components. It allows us to capture a more comprehensive and reliable representation~~ emphasizes the importance of careful PET method selection to ensure robust and representative assessments of hydrological trends.

*Code and data availability.*

The necessary data required to reproduce the final analysis and figures are available in the zenodo repository (https://doi.org/10.5281/zenodo.14008649). Codes used to analyze the results are publicly available in the GitHub repository (https://github.com/imarkonis/ithaca/tree/main/projects/pet_europe).

*Author contributions.*

VT: conceptualization, PET calculation, hydrological model simulations, postprocessing, formal analysis, and writing (original draft), YM: conceptualization, supervision, writing (review and editing) RK: conceptualization, writing (review and editing), JRT: writing (review and editing), MRVG: writing (review and editing), MH: writing (review and editing), OR: setup of hydrological model, supervision, conceptualization, writing (review and editing).

*Competing interests.*

At least one of the (co)-authors is a member of the editorial board of Hydrology and Earth System Sciences.

*Acknowledgements.* This work was carried out within the project "Investigation of Terrestrial HydrologicAl Cycle Acceleration (ITHACA)" funded by the Czech Science Foundation (Grant 22-33266M). Computational resources were provided by the e-INFRA CZ project (ID:90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic. VT was also funded by the Internal Grant Agency (Project no: 2023B0008), Czech University of Life Sciences. We sincerely thank editor Elena Toth, referee Franziska Clerc-Schwarzenbach, and other two referee for their valuable feedback and constructive suggestions, which significantly contributed to improving the quality of this manuscript.

## Appendix A: Potential Evapotranspiration formulation

**Table A1.** Formulations of PET methods. Where R$_e$ is extraterrestrial radiation (MJ m$^{-2}$ d$^{-1}$), $\lambda$ is the latent heat of vaporization (MJ kg$^{-1}$), $\rho$ is water density (= 1000 kg m$^{-3}$), $d_a$ is air density (kg m$^{-3}$), $T_a$ is air temperature (°C), $T_d$ is dew point temperature (°C), $T_{\max}$ is maximum air temperature (°C), $T_{\min}$ is minimum air temperature (°C), $\Delta$ is the slope of the vapor pressure curve (kPa °C$^{-1}$), $\gamma$ is the psychrometric constant (kPa °C$^{-1}$), $e_s$ is saturation vapour pressure (kPa), $e_a$ is actual vapour pressure (kPa), $u_2$ is wind speed 2 m above the soil surface (m s$^{-1}$), $R_s$ is net short-wave radiation (MJ m$^{-2}$ d$^{-1}$), $R_n$ is net incoming solar radiation (MJ m$^{-2}$ d$^{-1}$), $G$ is soil heat flux (MJ m$^{-2}$ d$^{-1}$), $RH$ is relative humidity (%), DL is day length (h d$^{-1}$), $I$ is annual heat index, and $CO_2$ is carbon dioxide concentration (ppm).

| Method | Formulation | Reference |
|---|---|---|
| ~~Hargreaves Samani~~ Hargreaves-Samani | $0.0023 \times \frac{R_e}{\lambda \times \rho} \times \sqrt{t_{max} - t_{min}} \times (t_{avg} + 17.8) \times 1000$ | Hargreaves and Samani (1985) |
| McGuinness-Bordne | $1000 \times \frac{R_e}{\lambda \times \rho} \times \frac{T_a+5}{68}$ | McGuinness and Bordne (1972) |
| Hamon | $k \times 0.165 \times 216.7 \times \frac{DL}{12} \times \frac{e_s}{t_{avg}+273.3}$ | Hamon (1961) |
| Oudin | $1000 \times \frac{R_e}{\lambda \times \rho} \times \frac{t_{avg}+5}{100}$ | Oudin et al. (2005) |
| ~~Baier and Robertson~~ Baier-Robertson | $0.157 \times t_{max} + 0.158(t_{max} - t_{min}) + 0.109 \times R_e - 5.39$ | Bai et al. (2016) |
| Blaney-Criddle | $0.825 \times (0.46 \times t_{avg} + 8.13) \times \frac{100 \times DL}{365 \times 12}$ | Blaney (1952) |
| Thornthwaite | $16 \times \frac{DL}{360} \times \left(\frac{10 \times t_{avg}}{I}\right)^k$ | Thornthwaite (1948) |
| Jensen-Haise | $1000 \times \frac{R_e}{\lambda \times \rho} \times \frac{T_{avg}}{40}$ | Jensen and Haise (1963) |
| ~~Priestley and Taylor~~ Priestley-Taylor | $\frac{1.26 \times \Delta \times (R_n - G)}{\lambda \times \rho \times (\Delta + \gamma)}$ | Priestley and Taylor (1972) |
| ~~Milley-Dunne~~ Milly-Dunne | $0.8 \times (R_n - G)$ | Milly and Dunne (2016) |
| Penman-Monteith | $\frac{0.408 \times \Delta \times (R_n - G) + \gamma \times \left(\frac{900}{T_{avg}+273}\right) \times u_2 \times (e_s - e_a)}{\Delta + \gamma \times (1 + 0.34 \times u_2)}$ | Penman (1948) |
| ~~Modified~~ Penman-Monteith[$CO_2$] | $\frac{0.408 \times \Delta \times (R_n - G) + \gamma \times \left(\frac{900}{T_{avg}+273}\right) \times u_2 \times (e_s - e_a)}{\Delta + \gamma \times (1 + 0.34 \times (u_2 + 2 \times 10^{-4} \times ([CO_2] - 300)))}$ | Yang et al. (2019) |

## References


Ajami, H., Sharma, A., Band, L. E., Evans, J. P., Tuteja, N. K., Amirthanathan, G. E., and Bari, M. A.: On the non-stationarity of hydrological response in anthropogenically unaffected catchments: an Australian perspective, Hydrology and Earth System Sciences, 21, 281–294, https://doi.org/10.5194/hess-21-281-2017, 2017.

Allen, R. G., ed.: Crop evapotranspiration: guidelines for computing crop water requirements, no. 56 in FAO irrigation and drainage paper, Food and Agriculture Organization of the United Nations, Rome, ISBN 978-92-5-104219-9, 1998.

Anabalón, A. and Sharma, A.: On the divergence of potential and actual evapotranspiration trends: An assessment across alternate global datasets, Earth's Future, 5, 905–917, https://doi.org/10.1002/2016EF000499, 2017.

Aouissi, J., Benabdallah, S., Lili Chabaâne, Z., and Cudennec, C.: Evaluation of potential evapotranspiration assessment methods for hydrological modelling with SWAT—Application in data-scarce rural Tunisia, Agricultural Water Management, 174, 39–51, https://doi.org/10.1016/j.agwat.2016.03.004, 2016.

Arino, O., Ramos Perez, J. J., Kalogirou, V., Bontemps, S., Defourny, P., and Van Bogaert, E.: Global Land Cover Map for 2009 (GlobCover 2009), https://doi.org/10.1594/PANGAEA.787668, artwork Size: 40 data points Pages: 40 data points, 2012.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R.: LARGE AREA HYDROLOGIC MODELING AND ASSESSMENT PART I: MODEL DEVELOPMENT, Journal of the American Water Resources Association, 34, 73–89, https://doi.org/10.1111/j.1752-1688.1998.tb05961.x, 1998.

Bai, P., Liu, X., Liang, K., and Liu, C.: Comparison of performance of twelve monthly water balance models in different climatic catchments of China, Journal of Hydrology, 529, 1030–1040, https://doi.org/10.1016/j.jhydrol.2015.09.015, 2015.

Bai, P., Liu, X., Yang, T., Li, F., Liang, K., Hu, S., and Liu, C.: Assessment of the Influences of Different Potential Evapotranspiration Inputs on the Performance of Monthly Hydrological Models under Different Climatic Conditions, Journal of Hydrometeorology, 17, 2259–2274, https://doi.org/10.1175/JHM-D-15-0202.1, 2016.

Beck, H. E., Van Dijk, A. I. J. M., De Roo, A., Dutra, E., Fink, G., Orth, R., and Schellekens, J.: Global evaluation of runoff from 10 state-of-the-art hydrological models, Hydrology and Earth System Sciences, 21, 2881–2903, https://doi.org/10.5194/hess-21-2881-2017, 2017.

Berghuijs, W. R., Larsen, J. R., van Emmerik, T. H. M., and Woods, R. A.: A Global Assessment of Runoff Sensitivity to Changes in Precipitation, Potential Evaporation, and Other Factors, Water Resources Research, 53, 8475–8486, https://doi.org/10.1002/2017WR021593, 2017.

Birhanu, D., Kim, H., Jang, C., and Park, S.: Does the Complexity of Evapotranspiration and Hydrological Models Enhance Robustness?, Sustainability, 10, 2837, https://doi.org/10.3390/su10082837, 2018.

Blaney, H. F.: Determining water requirements in irrigated areas from climatological and irrigation data, 1952.

Boeing, F., Wagener, T., Marx, A., Rakovec, O., Kumar, R., Samaniego, L., and Attinger, S.: Increasing influence of evapotranspiration on prolonged water storage recovery in Germany, Environmental Research Letters, 19, 024 047, https://doi.org/10.1088/1748-9326/ad24ce, 2024.

Bruno, G. and Duethmann, D.: Increases in Water Balance-Derived Catchment Evapotranspiration in Germany During 1970s–2000s Turning Into Decreases Over the Last Two Decades, Despite Uncertainties, Geophysical Research Letters, 51, e2023GL107 753, https://doi.org/10.1029/2023GL107753, 2024.

Caloiero, T., Caloiero, P., and Frustaci, F.: Long-term precipitation trend analysis in Europe and in the Mediterranean basin, Water and Environment Journal, 32, 433–445, https://doi.org/10.1111/wej.12346, 2018.

Cheng, W., Dan, L., Deng, X., Feng, J., Wang, Y., Peng, J., Tian, J., Qi, W., Liu, Z., Zheng, X., Zhou, D., Jiang, S., Zhao, H., and Wang, X.: Global monthly gridded atmospheric carbon dioxide concentrations under the historical and future scenarios, Scientific Data, 9, 83, https://doi.org/10.1038/s41597-022-01196-7, 2022.

Clerc-Schwarzenbach, F., Selleri, G., Neri, M., Toth, E., van Meerveld, I., and Seibert, J.: Large-sample hydrology – a few camels or a whole caravan?, Hydrology and Earth System Sciences, 28, 4219–4237, https://doi.org/10.5194/hess-28-4219-2024, 2024.

Deardorff, J. W.: Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation, Journal of Geophysical Research: Oceans, 83, 1889–1903, https://doi.org/10.1029/JC083iC04p01889, 1978.

Fang, B., Bevacqua, E., Rakovec, O., and Zscheischler, J.: An increase in the spatial extent of European floods over the last 70 years, https://doi.org/10.5194/egusphere-2023-2890, 2024.

George H. Hargreaves and Zohrab A. Samani: Reference Crop Evapotranspiration from Temperature, Applied Engineering in Agriculture, 1, 96–99, https://doi.org/10.13031/2013.26773, 1985.

Guo, D., Westra, S., and Maier, H. R.: Sensitivity of potential evapotranspiration to changes in climate variables for different Australian climatic zones, Hydrology and Earth System Sciences, 21, 2107–2126, https://doi.org/10.5194/hess-21-2107-2017, 2017.

Hamon, W. R.: Estimating Potential Evapotranspiration, Journal of the Hydraulics Division, 87, 107–120, https://doi.org/10.1061/JYCEAJ.0000599, 1961.

Hanselmann, N., Osuch, M., Wawrzyniak, T., and Alphonse, A. B.: Evaluating potential evapotranspiration methods in a rapidly warming Arctic region, SW Spitsbergen (1983–2023), Journal of Hydrology: Regional Studies, 56, 101 979, https://doi.org/10.1016/j.ejrh.2024.101979, 2024.

Hargreaves and Samani: Reference Crop Evapotranspiration from Temperature, Applied Engineering in Agriculture, 1, 96–99, https://doi.org/10.13031/2013.26773, 1985.

Hartmann, J. and Moosdorf, N.: The new global lithological map database GLiM: A representation of rock properties at the Earth surface, Geochemistry, Geophysics, Geosystems, 13, 2012GC004 370, https://doi.org/10.1029/2012GC004370, 2012.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hua, D., Hao, X., Zhang, Y., and Qin, J.: Uncertainty assessment of potential evapotranspiration in arid areas, as estimated by the Penman-Monteith method, Journal of Arid Land, 12, 166–180, https://doi.org/10.1007/s40333-020-0093-7, 2020.

Jensen, M. E. and Haise, H. R.: Estimating Evapotranspiration from Solar Radiation, Journal of the Irrigation and Drainage Division, 89, 15–41, https://doi.org/10.1061/JRCEA4.0000287, 1963.

Kendall, M.: Rank correlation methods., Rank correlation methods., Griffin, Oxford, England, 1948.

Kingston, D. G., Todd, M. C., Taylor, R. G., Thompson, J. R., and Arnell, N. W.: Uncertainty in the estimation of potential evapotranspiration under climate change, Geophysical Research Letters, 36, 2009GL040 267, https://doi.org/10.1029/2009GL040267, 2009.

Kuentz, A., Arheimer, B., Hundecha, Y., and Wagener, T.: Understanding hydrologic variability across Europe through catchment classification, Hydrology and Earth System Sciences, 21, 2863–2879, https://doi.org/10.5194/hess-21-2863-2017, 2017.

Kumar, R., Livneh, B., and Samaniego, L.: Toward computationally efficient large-scale hydrologic predictions with a multiscale regionalization scheme: LARGE-SCALE HYDROLOGIC PREDICTIONS, Water Resources Research, 49, 5700–5714, https://doi.org/10.1002/wrcr.20431, 2013a.

Kumar, R., Samaniego, L., and Attinger, S.: Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations, Water Resources Research, 49, 360–379, https://doi.org/10.1029/2012WR012195, 2013b.

Kumar, R., Heße, F., Rao, P. S. C., Musolff, A., Jawitz, J. W., Sarrazin, F., Samaniego, L., Fleckenstein, J. H., Rakovec, O., Thober, S., and Attinger, S.: Strong hydroclimatic controls on vulnerability to subsurface nitrate contamination across Europe, Nature Communications, 11, 6302, https://doi.org/10.1038/s41467-020-19955-8, 2020.

Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, Journal of Geophysical Research: Atmospheres, 99, 14 415–14 428, https://doi.org/10.1029/94JD00483, 1994.

Liu, C., Wang, Z., Zheng, H., Zhang, L., and Wu, X.: Development of Hydro-Informatic Modelling System and its application, Science in China Series E: Technological Sciences, 51, 456–466, https://doi.org/10.1007/s11431-008-0040-x, 2008.

Liu, Y., Jiang, Q., Wang, Q., Jin, Y., Yue, Q., Yu, J., Zheng, Y., Jiang, W., and Yao, X.: The divergence between potential and actual evapotranspiration: An insight from climate, water, and vegetation change, Science of The Total Environment, 807, 150 648, https://doi.org/10.1016/j.scitotenv.2021.150648, 2022.

Lu, J., Sun, G., McNulty, S. G., and Amatya, D. M.: A COMPARISON OF SIX POTENTIAL EVAPOTRANSPIRATION METHODS FOR REGIONAL USE IN THE SOUTHEASTERN UNITED STATES, Journal of the American Water Resources Association, 41, 621–633, https://doi.org/10.1111/j.1752-1688.2005.tb03759.x, 2005.

Markonis, Y.: On the Definition of Extreme Evaporation Events, Geophysical Research Letters, 52, e2024GL113 038, 2025.

Markonis, Y., Papalexiou, S. M., Martinkova, M., and Hanel, M.: Assessment of Water Cycle Intensification Over Land using a Multisource Global Gridded Precipitation DataSet, Journal of Geophysical Research: Atmospheres, 124, 11 175–11 187, https://doi.org/10.1029/2019JD030855, 2019.

Markonis, Y., Kumar, R., Hanel, M., Rakovec, O., Máca, P., and AghaKouchak, A.: The rise of compound warm-season droughts in Europe, Science Advances, 7, eabb9668, https://doi.org/10.1126/sciadv.abb9668, 2021.

Markonis, Y., Vargas Godoy, M. R., Pradhan, R. K., Pratap, S., Thomson, J. R., Hanel, M., Paschalis, A., Nikolopoulos, E., and Papalexiou, S. M.: Spatial partitioning of terrestrial precipitation reveals varying dataset agreement across different environments, Communications Earth & Environment, 5, 217, 2024.

Massari, C., Avanzi, F., Bruno, G., Gabellani, S., Penna, D., and Camici, S.: Evaporation enhancement drives the European water-budget deficit during multi-year droughts, Hydrology and Earth System Sciences, 26, 1527–1543, https://doi.org/10.5194/hess-26-1527-2022, 2022.

Mazzoleni, M., Brandimarte, L., and Amaranto, A.: Evaluating precipitation datasets for large-scale distributed hydrological modelling, Journal of Hydrology, 578, 124 076, https://doi.org/10.1016/j.jhydrol.2019.124076, 2019.

Maček, U., Bezak, N., and Šraj, M.: Reference evapotranspiration changes in Slovenia, Europe, Agricultural and Forest Meteorology, 260-261, 183–192, https://doi.org/10.1016/j.agrformet.2018.06.014, 2018.

McGuinness, J. L. and Bordne, E. F.: A comparison of lysimeter-derived potential evapotranspiration with computed values, US Department of Agriculture, ISBN 0082-9811, issue: 1452, 1972.

Milly, P. C. D. and Dunne, K. A.: Potential evapotranspiration and continental drying, Nature Climate Change, 6, 946–949, https://doi.org/10.1038/nclimate3046, 2016.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth System Science Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

Ndiaye, P. M., Bodian, A., Dezetter, A., Ogilvie, A., and Goudiaby, O.: Sensitivity of global hydrological models to potential evapotranspiration estimation methods in the Senegal River Basin (West Africa), Journal of Hydrology: Regional Studies, 53, 101 823, https://doi.org/10.1016/j.ejrh.2024.101823, 2024.

Oudin, L., Michel, C., and Anctil, F.: Which potential evapotranspiration input for a lumped rainfall-runoff model?, Journal of Hydrology, 303, 275–289, https://doi.org/10.1016/j.jhydrol.2004.08.025, 2005.

Park, C.-E., Jeong, S.-J., Joshi, M., Osborn, T. J., Ho, C.-H., Piao, S., Chen, D., Liu, J., Yang, H., Park, H., Kim, B.-M., and Feng, S.: Keeping global warming within 1.5 °C constrains emergence of aridification, Nature Climate Change, 8, 70–74, https://doi.org/10.1038/s41558-017-0034-4, 2018.

Penman, H. L.: Natural evaporation from open water, bare soil and grass, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 193, 120–145, number: 1032 ISBN: 0080-4630 Publisher: The Royal Society London, 1948.

Pfeifroth, U., Sanchez-Lorenzo, A., Manara, V., Trentmann, J., and Hollmann, R.: Trends and Variability of Surface Solar Radiation in Europe Based On Surface- and Satellite-Based Data Records, Journal of Geophysical Research: Atmospheres, 123, 1735–1754, https://doi.org/10.1002/2017JD027418, 2018.

Pimentel, R., Arheimer, B., Crochemore, L., Andersson, J. C. M., Pechlivanidis, I. G., and Gustafsson, D.: Which Potential Evapotranspiration Formula to Use in Hydrological Modeling World-Wide?, Water Resources Research, 59, e2022WR033 447, https://doi.org/10.1029/2022WR033447, 2023.

Pohl, F., Rakovec, O., Rebmann, C., Hildebrandt, A., Boeing, F., Hermanns, F., Attinger, S., Samaniego, L., and Kumar, R.: Long-term daily hydrometeorological drought indices, soil moisture, and evapotranspiration for ICOS sites, Scientific Data, 10, 281, https://doi.org/10.1038/s41597-023-02192-1, 2023.

Priestley, C. H. B. and Taylor, R. J.: On the assessment of surface heat flux and evaporation using large-scale parameters, Monthly weather review, 100, 81–92, iSBN: 1520-0493, 1972.

Proutsos, N., Tigkas, D., Tsevreni, I., Alexandris, S. G., Solomou, A. D., Bourletsikas, A., Stefanidis, S., and Nwokolo, S. C.: A Thorough Evaluation of 127 Potential Evapotranspiration Models in Two Mediterranean Urban Green Sites, Remote Sensing, 15, 3680, https://doi.org/10.3390/rs15143680, 2023.

Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., Attinger, S., Schäfer, D., Schrön, M., and Samaniego, L.: Multiscale and Multivariate Evaluation of Water Fluxes and States over European River Basins, Journal of Hydrometeorology, 17, 287–307, https://doi.org/10.1175/JHM-D-15-0054.1, 2016.

Rakovec, O., Mizukami, N., Kumar, R., Newman, A. J., Thober, S., Wood, A. W., Clark, M. P., and Samaniego, L.: Diagnostic Evaluation of Large-Domain Hydrologic Models Calibrated Across the Contiguous United States, Journal of Geophysical Research: Atmospheres, 124, 13 991–14 007, https://doi.org/10.1029/2019JD030767, 2019.

Rakovec, O., Samaniego, L., Hari, V., Markonis, Y., Moravec, V., Thober, S., Hanel, M., and Kumar, R.: The 2018–2020 Multi-Year Drought Sets a New Benchmark in Europe, Earth's Future, 10, e2021EF002 394, https://doi.org/10.1029/2021EF002394, 2022.

Rakovec, O., Kumar, R., Shrestha, P. K., and Samaniego, L.: Global assessment of hydrological components using a seamless multiscale modelling system, https://doi.org/10.5194/egusphere-egu23-11945, 2023.

Reaver, N. G. F., Kaplan, D. A., Klammler, H., and Jawitz, J. W.: Theoretical and empirical evidence against the Budyko catchment trajectory conjecture, Hydrology and Earth System Sciences, 26, 1507–1525, https://doi.org/10.5194/hess-26-1507-2022, 2022.

Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resources Research, 46, https://doi.org/10.1029/2008WR007327, 2010.

Samaniego, L., Thober, S., Wanders, N., Pan, M., Rakovec, O., Sheffield, J., Wood, E. F., Prudhomme, C., Rees, G., Houghton-Carr, H., Fry, M., Smith, K., Watts, G., Hisdal, H., Estrela, T., Buontempo, C., Marx, A., and Kumar, R.: Hydrological Forecasts and Projections for Improved Decision-Making in the Water Sector in Europe, Bulletin of the American Meteorological Society, 100, 2451–2472, https://doi.org/10.1175/BAMS-D-17-0274.1, 2019.

Schulzweida, U.: CDO User Guide, https://doi.org/10.5281/ZENODO.7112925, publisher: Zenodo Version Number: 2.1.0, 2022.

Seiller, G. and Anctil, F.: How do potential evapotranspiration formulas influence hydrological projections?, Hydrological Sciences Journal, 61, 2249–2266, https://doi.org/10.1080/02626667.2015.1100302, 2016.

Sen, P. K.: Estimates of the Regression Coefficient Based on Kendall's Tau, Journal of the American Statistical Association, 63, 1379–1389, https://doi.org/10.1080/01621459.1968.10480934, 1968.

Shah, J., Kumar, R., Samaniego, L., Markonis, Y., Hanel, M., Attinger, S., Hari, V., and Rakovec, O.: On the role of antecedent meteorological conditions on flash drought initialization in Europe, Environmental Research Letters, 18, 064 039, https://doi.org/10.1088/1748-9326/acd8d3, 2023.

Shaw, S. B. and Riha, S. J.: Assessing temperature-based PET equations under a changing climate in temperate, deciduous forests, Hydrological Processes, 25, 1466–1478, https://doi.org/10.1002/hyp.7913, 2011.

Shi, H., Luo, G., Hellwich, O., He, X., Kurban, A., De Maeyer, P., and Van de Voorde, T.: Global dryland aridity changes indicated by atmospheric, hydrological, and vegetation observations at meteorological stations, Hydrology and Earth System Sciences, 27, 4551–4562, https://doi.org/10.5194/hess-27-4551-2023, 2023a.

Shi, L., Wang, B., Liu, D. L., Feng, P., Cleverly, J., Li, L., Zhang, G., and Yu, Q.: Performance of potential evapotranspiration models across different climatic stations in New South Wales, Australia, Journal of Hydrology: Regional Studies, 50, 101 573, https://doi.org/10.1016/j.ejrh.2023.101573, 2023b.

Shrestha, P. K., Samaniego, L., Rakovec, O., Kumar, R., Mi, C., Rinke, K., and Thober, S.: Toward Improved Simulations of Disruptive Reservoirs in Global Hydrological Modeling, Water Resources Research, 60, e2023WR035 433, https://doi.org/10.1029/2023WR035433, 2024.

Tang, G., Clark, M. P., and Papalexiou, S. M.: SC-Earth: A Station-Based Serially Complete Earth Dataset from 1950 to 2019, Journal of Climate, 34, 6493–6511, https://doi.org/10.1175/JCLI-D-21-0067.1, 2021.

Tang, G., Clark, M. P., and Papalexiou, S. M.: EM-Earth: The Ensemble Meteorological Dataset for Planet Earth, Bulletin of the American Meteorological Society, 103, E996–E1018, https://doi.org/10.1175/BAMS-D-21-0106.1, 2022.

Tang, G., Clark, M. P., Knoben, W. J. M., Liu, H., Gharari, S., Arnal, L., Beck, H. E., Wood, A. W., Newman, A. J., and Papalexiou, S. M.: The Impact of Meteorological Forcing Uncertainty on Hydrological Modeling: A Global Analysis of Cryosphere Basins, Water Resources Research, 59, e2022WR033 767, https://doi.org/10.1029/2022WR033767, 2023.

Tarek, M., Brissette, F. P., and Arsenault, R.: Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America, Hydrology and Earth System Sciences, 24, 2527–2544, https://doi.org/10.5194/hess-24-2527-2020, 2020.

Teuling, A. J., de Badts, E. A. G., Jansen, F. A., Fuchs, R., Buitink, J., Hoek van Dijke, A. J., and Sterling, S. M.: Climate change, reforestation/afforestation, and urbanization impacts on evapotranspiration and streamflow in Europe, Hydrology and Earth System Sciences, 23, 3631–3652, https://doi.org/10.5194/hess-23-3631-2019, 2019.

Thackeray, C. W., Hall, A., Norris, J., and Chen, D.: Constraining the increased frequency of global precipitation extremes under warming, Nature Climate Change, 12, 441–448, https://doi.org/10.1038/s41558-022-01329-1, 2022.

Thober, S., Cuntz, M., Kelbling, M., Kumar, R., Mai, J., and Samaniego, L.: The multiscale routing model mRM v1.0: simple river routing at resolutions from 1 to 50 km, Geoscientific Model Development, 12, 2501–2521, https://doi.org/10.5194/gmd-12-2501-2019, 2019.

Thornthwaite, C. W.: An Approach toward a Rational Classification of Climate, Geographical Review, 38, 55, https://doi.org/10.2307/210739, 1948.

Vicente-Serrano, S. M., Azorin-Molina, C., Sanchez-Lorenzo, A., Revuelto, J., López-Moreno, J. I., González-Hidalgo, J. C., Moran-Tejeda, E., and Espejo, F.: Reference evapotranspiration variability and trends in Spain, 1961–2011, Global and Planetary Change, 121, 26–40, https://doi.org/10.1016/j.gloplacha.2014.06.005, 2014.

Voisin, N., Wood, A. W., and Lettenmaier, D. P.: Evaluation of Precipitation Products for Global Hydrological Prediction, Journal of Hydrometeorology, 9, 388–407, https://doi.org/10.1175/2007JHM938.1, 2008.

World Meteorological Organization (WMO): State of Global Water Resources 2022, Tech. rep., Geneva, Switzerland, https://wmo.int/publication-series/state-of-global-water-resources-2022, 2023.

Xiang, K., Li, Y., Horton, R., and Feng, H.: Similarity and difference of potential evapotranspiration and reference crop evapotranspiration – a review, Agricultural Water Management, 232, 106 043, https://doi.org/10.1016/j.agwat.2020.106043, 2020.

Xu, C.-Y. and Singh, V. P.: Evaluation and generalization of radiation-based methods for calculating evaporation, Hydrological Processes, 14, 339–349, https://doi.org/10.1002/(SICI)1099-1085(20000215)14:2<339::AID-HYP928>3.0.CO;2-O, 2000.

Yang, Y., Roderick, M. L., Zhang, S., McVicar, T. R., and Donohue, R. J.: Hydrologic implications of vegetation response to elevated CO2 in climate projections, Nature Climate Change, 9, 44–48, https://doi.org/10.1038/s41558-018-0361-0, 2019.

Yin, Z., Lin, P., Riggs, R., Allen, G. H., Lei, X., Zheng, Z., and Cai, S.: A synthesis of Global Streamflow Characteristics, Hydrometeorology, and Catchment Attributes (GSHA) for large sample river-centric studies, Earth System Science Data, 16, 1559–1587, https://doi.org/10.5194/essd-16-1559-2024, 2024.

Zhang, D., Liu, X., Zhang, Q., Liang, K., and Liu, C.: Investigation of factors affecting intra-annual variability of evapotranspiration and streamflow under different climate conditions, Journal of Hydrology, 543, 759–769, https://doi.org/10.1016/j.jhydrol.2016.10.047, 2016.

Zhang, Y., He, B., Guo, L., Liu, J., and Xie, X.: The relative contributions of precipitation, evapotranspiration, and runoff to terrestrial water storage changes across 168 river basins, Journal of Hydrology, 579, 124 194, https://doi.org/10.1016/j.jhydrol.2019.124194, 2019.

Zhao, L., Xia, J., Xu, C.-y., Wang, Z., Sobkowiak, L., and Long, C.: Evapotranspiration estimation methods in hydrological models, Journal of Geographical Sciences, 23, 359–369, https://doi.org/10.1007/s11442-013-1015-9, 2013.

Zhou, J., Li, Q., Ye, A., Xu, S., Yuan, Y., Xu, S., Zhang, D., Zhao, X., Zhu, Y., Zhao, Y., Xue, D., Dou, J., Liu, C., Shi, W., Wei, W., and Yang, X.: An improved methodology for quantifying the impact of human activities on hydrological drought change, Journal of Hydrology: Regional Studies, 50, 101 603, https://doi.org/10.1016/j.ejrh.2023.101603, 2023.