Journal of Hydrology 628 (2024) 130477

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

# Global evapotranspiration models and their performance at different spatial scales: Contrasting a latitudinal gradient against global catchments

Ignacio Fuentes [a,b,*], R. Willem Vervoort [c,d], James McPhee [e]

[a] Facultad de Medicina Veterinaria y Agronomía, Universidad de las Américas, Sede Providencia, Chile
[b] Núcleo en Ciencias Ambientales y Alimentarias (NCAA), Universidad de las Américas, Santiago, Chile
[c] School of Life and Environmental Sciences, The University of Sydney, NSW 2015, Sydney, Australia
[d] ARC ITTC in Data Analytics for Resources and Environments, The University of Sydney, NSW 2006, Sydney, Australia
[e] Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile

ARTICLE INFO

This manuscript was handled by Marco Borga, Editor-in-Chief, with the assistance of Sheng-ping Wang, Associate Editor.

*Keywords:*
Evapotranspiration models
Remote sensing
Validation scales
Total water storage

ABSTRACT

Actual evapotranspiration (AET) is a key variable in the global water balance, driving agricultural production and ecosystem health. It is a complex hydrologic process that depends on vegetation, climate, and available water conditions. Different moderate resolution global AET models have been developed to quantify water resources at large scales. In this work we evaluate five of these products, including MODIS, PML, SSEBop, TerraClimate, and a Synthesis AET using point and catchment-scale datasets based on flux towers. We also contrast water balance changes with total water storage (TWS) products. These comparisons cover different radiation and precipitation regimes over catchments around the world and along a strong climatic gradient in north-central Chile. We rank the models, contrast TWS datasets, and study differences related to scale in validation and the effect of rainfall and radiation on simulated values. Additionally, we use a Budyko framework to evaluate the AET products in terms of their agreement with expected water budgets. At different evaluation scales, AET estimates and observations agreed reasonably well, with the largest mean $R^2$ of about 0.7 and errors of approximately 15% of the magnitude of the observed variables. MODIS and Synthesis AET had the highest $R^2$ at the point (0.62) and at the catchment scales (0.71 and 0.59 for regional and global catchments), respectively, but were closely followed by PML. PML and TerraClimate led to the lowest magnitude errors at the point (RMSE = 0.78 mm day$^{-1}$) and catchment scales (mean RMSE = 1.5 mm day$^{-1}$), respectively. The rainfall gradient is reflected in a performance gradient. PML, MODIS, and TerraClimate gave consistent behaviour based on the Budyko curve, with a few arid catchments exceeding the water limit. The major conclusion is that remotely sensed AET outperforms flux tower AET extrapolation for water balance calculations at the catchment scale, which means that errors in satellite-based AET products tend to cancel out at larger spatial scales, which makes them viable alternatives for regional water balance studies. However, flux data integrated into AET models, such as the FluxCom model, leads to the lowest errors. The assimilation and downscaling of Gravity Recovery and Climate Experiment (GRACE) into the Global Land Data Assimilation System (GLDAS) leads to an improvement in regional results compared with other TWS products.

## 1. Introduction

Remote sensing has increased in popularity given its potential for monitoring surface land processes (Palmer et al., 2015; Fuentes et al., 2019; Weiss et al., 2020; Avtar et al., 2021). However, while gridded data can have advantages over point observations by providing continuous spatial estimates, errors derived from the estimates can accumulate if considered at scales that exceed pixel dimensions (Li and Shao, 2010; Worqlul et al., 2014). Similarly, if land processes quantified in different remote sensing products have biases related to different land features or climates, and different land features and climates characterise the land aggregations under consideration (e.g., catchments, regions, countries), errors may again accumulate or cancel out at these scales (Senay et al., 2020).

Actual evapotranspiration (AET) based on remote sensing has a long development history, resulting in a range of methodologies for quantification (King et al., 2011; Ershadi et al., 2013; Zhang et al., 2019), and different validation techniques (Guerschman et al., 2009; Sriwongsitanon et al., 2020; Elnashar et al., 2021). Thus, multiple AET products

* Corresponding author at: Núcleo en Ciencias Ambientales y Alimentarias (NCAA), Universidad de las Américas, Santiago, Chile.
*E-mail address:* ifuentes@udla.cl (I. Fuentes).

https://doi.org/10.1016/j.jhydrol.2023.130477
Received 27 December 2022; Received in revised form 30 September 2023; Accepted 31 October 2023
Available online 17 November 2023

have been developed that can quantify hydrologic resources (Mu et al., 2011; Zhang et al., 2019; Senay et al., 2020). However, quantifying AET is not a trivial task, since it involves complex processes that include meteorological, soil water availability, and vegetation growth characteristics (Ershadi et al., 2013). Moreover, the dominant hydrologic processes are spatio temporal variables in the landscape, which at regional scales can cause quantification differences. Elnashar et al. (2021), aware of the fact that different AET models perform differently for different types of land cover and climate classes, sought to create an ensemble of different AET products to reduce uncertainties across landscapes.

Most validation techniques to evaluate the performance of AET products make use of point observations such as eddy covariance towers or through known vegetation crop coefficients and the calculation of reference evapotranspiration (Allen et al., 1998; Elnashar et al., 2021). One of the challenges of validation from point observations is that complex processes such as AET cannot be fully extrapolated to larger scales exactly because of land heterogeneity and different processes that take place at catchment or regional scales compared to local observations (Ershadi et al., 2013; Strong and Elliott, 2017). Moreover, in large regions, the sparse spatial distribution of monitoring points in climatic and hydrological monitoring networks and discontinuities in recordings (Hund et al., 2016) hinder potential validation. These factors not only limit the potential for validation of AET, but also the development of gridded products derived from observational data, which may justify the use of global modelled remote sensing datasets.

On the other hand, from the methodologies used for catchment validation of AET, the water balance is probably the most widely used (Guerschman et al., 2009; Senay et al., 2011; Sriwongsitanon et al., 2020). Validation at this scale might be of interest, especially if climate variations occur in studied regions. However, a further difficulty relates to the sources of data to evaluate different components of the water balance at catchment or regional scales. The first problem relates to the common assumption that there is no change in storage over time (Guerschman et al., 2009; Jung et al., 2010; King et al., 2011; Buzacott and Vervoort, 2021). The assumption of no change in the stored water in a catchment is rarely met on the short and medium term, and in some cases not even in the long term. More specifically, this assumption depends on the evaluation time scale, climate aridity, vegetation coverage, and snowfall (Han et al., 2020). This means that for analyses at a monthly scale, different approaches are required.

Different modelled or observed datasets, such as the WaterGAP Global Hydrology Model (Döll et al., 2003) or the use of the NASA Gravity Recovery and Climate Experiment (GRACE) satellites (Tapley et al., 2004; Kornfeld et al., 2019) can be used to close the water balance by allowing the calculation of total water storage (TWS) changes. Nevertheless, these alternatives also present challenges, particularly those caused by the low resolution of total water storage anomalies, such as from GRACE (~3 degree). Sriwongsitanon et al. (2020), for instance, applied GRACE data to validate AET products in Thailand. However, Thailand is strongly dominated by tropical savanna climates, with high annual rainfall and relatively low regional variability (Beck et al., 2018). Therefore, such evaluations must be further applied to a different range of climates to better understand how these products perform under different conditions.

Given the above, the shape of Chile, which crosses many latitudes, and the relatively small size and self-contained characteristics of Chilean catchments, it becomes an interesting region to consider. This is further bolstered by the strong radiation and rainfall gradients that provide different climate and hydrological characteristics (Bonilla and Vidal, 2011; Carretier et al., 2018). Additionally, from a national perspective, there are contrasting views on the dominant processes that are leading to increasing climate variability and water scarcity in central-southern Chile (Garreaud et al., 2020; Fuentes et al., 2021; Madariaga et al., 2021). To estimate water balances under increasing climate variability and water scarcity remains a challenging task considering the limited available data sources. Given the different AET and TWS remote sensing sources, it would be valuable to contrast how these perform under different climate conditions, such as those conditioned by latitudinal gradients. Furthermore, the contrasting Chilean hydrological conditions and the evaluation of AET products can be representative of a wide range of different climates, from desert to temperate oceanic (Sarricolea et al., 2017; Beck et al., 2018), which can be further compared with other global catchments. This evaluation, which combines validation methods at the point and catchment scales, can lead to improvements in water balance calculations by reducing errors associated with the AET component.

The objective of this study is to evaluate different moderate resolution global AET products in a latitudinal gradient of north and central Chile and contrasting these results with other catchments globally. In addition, it will consider a catchment scale validation schema through the use of the water balances based on dynamic storage changes derived from different sources, and through point observations from flux towers. The latitudinal gradient and global catchments are selected to account for different climate and hydrologic variations that may lead to performance differences among AET products. On the basis of these results, a discussion of implications, challenges, and scientific gaps arising from AET products is provided.

## 2. Materials and methods

### 2.1. Study region and datasets

This study was carried out at two different scales: (1) at the regional scale, Northern and Central Chile between parallels −38.5° and −17° (Fig. 1). (2) At the global scale, different regions including North and South America, Africa, Europe, and Australia.

The selected Chilean region represents a latitudinal gradient of increasing rainfall and decreasing radiation to the south; however, it also has a temperature gradient increasing east to west caused by the Andes mountain range. The northern landscapes are characterised by arid conditions and desert climates, which transition to semi-arid and Mediterranean climates in central Chile, up to oceanic climates in the southern extreme of the study region, with annual rainfalls exceeding 2000 mm (Sarricolea et al., 2017).

Additionally, another series of catchments that lie along a latitudinal gradient were included in the global analysis. This consisted of 9 catchments along the Australian east coast (Supplementary materials; Figure S1). These catchments predominantly encompass temperate and arid Köppen climate types (Beck et al., 2018). However, there is no latitudinal rainfall gradient across these catchments.

### 2.2. Total water storage

Monthly TWS and their anomalies were obtained from the Global Land Data Assimilation System (GLDAS) version 2.2 and from the Water – Global Assessment and Prognosis (WaterGAP) v2.2 model. The GLDAS 2.2 model is a land surface model that assimilates GRACE into the Catchment land surface model (Koster et al., 2000). It uses an ensemble Kalman smoother to downscale GRACE and has proven significant regional improvements compared with GRACE given the low spatial resolution of the latter (Li et al., 2019). WaterGAP corresponds to a global hydrological model that accounts for water use, flow and storage. It combines three model components that consider water consumption, the linking model Groundwater-Surface Water Use, and the WaterGAP Global Hydrology Model (Müller Schmied et al., 2021). Additionally, as a reference for comparison, TWS anomalies from the Centre for Space Research (CSR) National Aeronautics and Space Administration (NASA) Gravity Recovery and Climate Experiment (GRACE) and from the NASA Gravity Recovery and Climate Experiment Follow-up (GRACE-FO) RL06 Mascons with all corrections

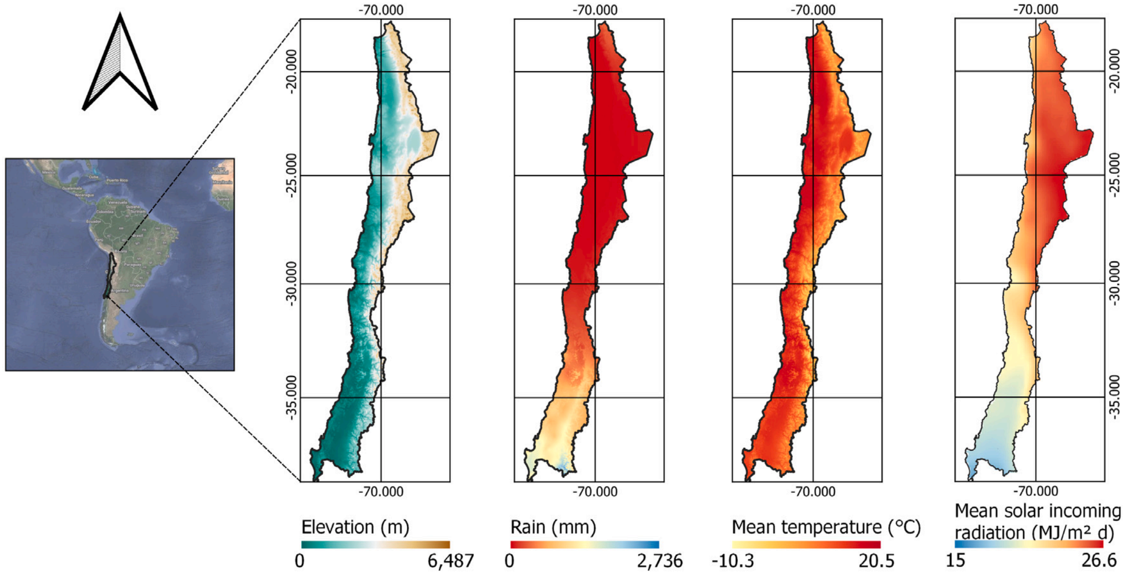


**Fig. 1.** Regional study area and the main characteristics in terms of elevation, mean annual rainfall, temperatures, and solar incoming radiation.

**Table 1**
Data sets used and main characteristics.

| Dataset | Spatial resolution | Temporal resolution | Operation period | Access source |
|---|---|---|---|---|
| MOD16A2 | 500 m | 8-days | 2001–now | https://developers.google.com/earth-engine/datasets |
| PML AET | 500 m | 8-days | 2002/07–now | https://developers.google.com/earth-engine/datasets |
| Synthesis AET | 1 km | monthly | 1982–2019 | https://developers.google.com/earth-engine/datasets |
| TerraClimate | ~4 km | monthly | 1958–now | https://developers.google.com/earth-engine/datasets |
| SSEBop | 1 km | monthly | 2003–now | https://earlywarning.usgs.gov/fews/product/460 |
| GRACE/GRACE-FO TWS | 0.25° | monthly | 2002/04–now | http://www2.csr.utexas.edu/grace |
| GLDAS 2.2 | ~0.25° | monthly | 2003/02–now | https://developers.google.com/earth-engine/datasets |
| WaterGAP v2.2d | ~0.5° | monthly | 1980–2016 | https://doi.pangaea.de/10.1594/PANGAEA.918447 |
| CHIRPS | ~5 km | daily | 1981–now | https://developers.google.com/earth-engine/datasets |
| GPM | ~10 km | 3-h/monthly | 2000/06–now | https://developers.google.com/earth-engine/datasets |
| CR2MET | ~5 km | monthly | 1979–2020 | https://www.cr2.cl/datos-productos-grillados/ |
| ERA5-Land | ~10 km | hourly | 1981–now | https://developers.google.com/earth-engine/datasets |
| National discharge | – | daily/monthly | Variable | http://www.dga.cl |
| GRDC data | – | daily/monthly | Variable | https://www.bafg.de/GRDC/ |
| FluxCom | 0.25° | daily/monthly | 2001–2015 | https://www.fluxcom.org/ |
| Fluxnet | – | daily/monthly | Variable | https://fluxnet.org/ |

applied (version 02) (Save et al., 2016; Save, 2020) were also evaluated. These were masked using the CSR land mask from the University of Texas to minimise leakage along the coastline. However, GRACE is used with caution in this case and only as a reference given the small size of studied catchments and the low spatial resolution of Mascons, which may lead to TWS errors. The datasets used were sampled using the filtered catchments to obtain a series of monthly averaged TWS and TWS anomalies per catchment (Table 1).

### 2.3. Discharge, catchment selection, and flux data

At the regional scale, monthly discharge data and catchment boundaries were obtained from the Water Resources Directorate (*Dirección General de Aguas*, http://www.dga.cl). The catchments were filtered based on the presence of a hydrometric station close to the river outlet and a catchment extent that was large enough to contain at least a single WaterGAP data pixel (~0.5°). This resulted in a selection of only 11 catchments (Fig. 2).

At the global scale, 51 catchments from the Global Runoff Data Centre (GRDC; https://www.bafg.de/GRDC/) were filtered based on the availability of discharge recordings between the years 2000 and 2022 and areas ranging from 4130 km$^2$ to 789,162 km$^2$ (Fig. 3 above). Boundaries from these catchments were also obtained from the GRDC dataset (GRDC, 2011). Discharges were aggregated to monthly before further analysis.

Furthermore, to evaluate point-scale validation, 27 flux towers from Fluxnet CC-BY-4.0 (Pastorello et al., 2020) were selected. These were selected because they were within the selected catchments or, in the case of South America, they were located in neighbouring countries of the regional subset (Fig. 3 below). From these, actual evapotranspiration was aggregated to monthly values. Additionally, monthly gridded latent heat from the FluxCom dataset was transformed into evapotranspiration and compared against other evapotranspiration products. FluxCom combines eddy covariance towers from the Fluxnet network with remote sensing and meteorological data through machine learning (Jung et al., 2019).

### 2.4. Rainfall and radiation data

Gridded daily rainfall data from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) at 0.05° resolution was obtained. CHIRPS combines infrared rainfall with *in-situ* station data to produce a good quality rainfall dataset (Funk et al., 2015), which in large regions has strong correlations with observed data, but in some desert climates it has shown to perform poorly (Fuentes et al., 2022), which was further evaluated in this study. Daily rainfall data was aggregated to monthly and averaged to the catchment scale. Additionally, CR2MET gridded rainfall was also obtained (https://www.cr2.cl/). This dataset contains monthly rainfall limited to the Chilean extent, which is based on a statistical regionalisation of ERA interim rainfall

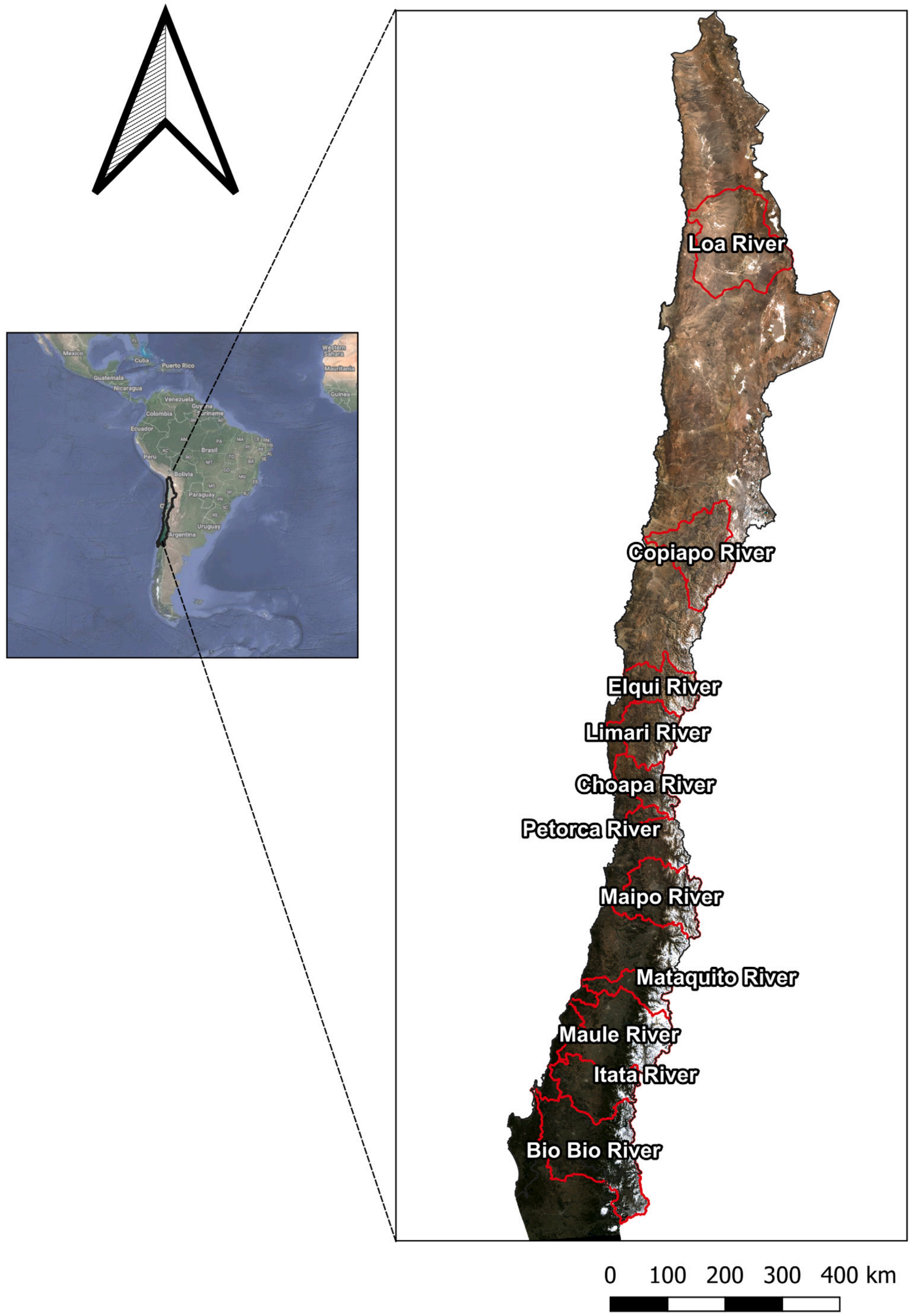


**Fig. 2.** Chilean catchments filtered in the study region.

data. Both datasets were compared with meteorological station data to select the best performing in terms of temporal characterisation (Pearson correlation) and magnitude of errors (root mean squared errors) (Fig. 4). Additionally, rainfall from meteorological stations was averaged by catchment (Fig. 4, right column) and also used as input in the analysis.

Given the better performance of CR2MET, which results in slightly higher correlations and lower errors (root mean squared errors; RMSE), it was selected for further analysis at the regional scale. At the global scale, CHIRPS was used as the main rainfall input. However, since CHIRPS coverage is limited at the northern latitudes, it was complemented for those latitudes with monthly rainfall estimates at 0.1° resolution from the Global Precipitation Measurement (GPM) satellite mission version 6. This dataset applies the Integrated Multi-satellitE Retrievals for GPM (IMERG) algorithm. IMERG processes satellite microwave precipitation, gauge data, and microwave-calibrated infrared (IR) satellite estimates, using monthly observations that lead to research-level estimates (Huffman et al., 2015).

Mean daily downward surface solar radiation was obtained from the fifth-generation European Centre for Medium-Range Weather Forecasts (ECMWF) climate Reanalysis over land (ERA5-Land) collection. This contains grids of global hourly weather and hydrologic data at approximately 10 km (Muñoz-Sabater et al., 2021). The dataset was

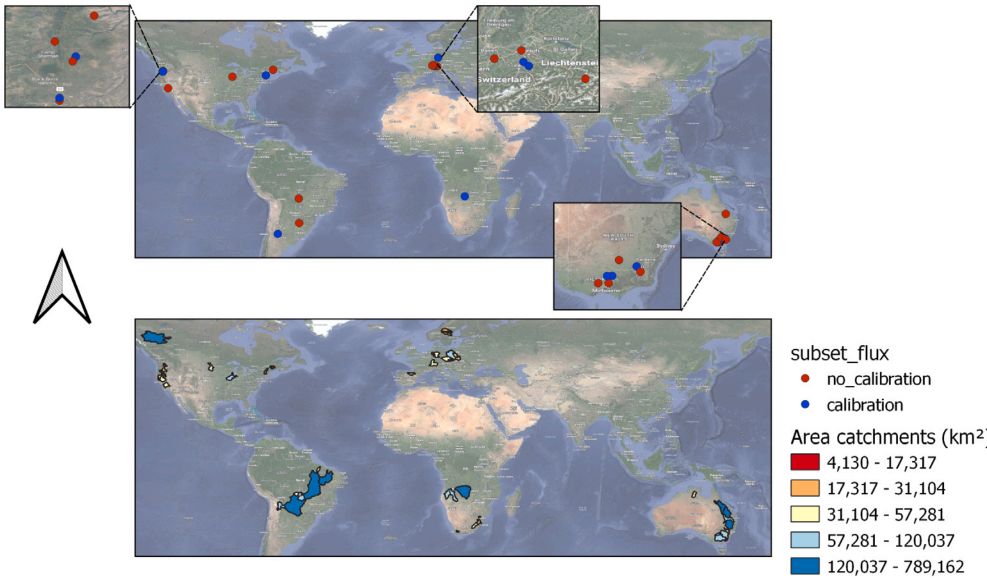


**Fig. 3.** Global GRDC catchments (below) and flux towers (above) selected for additional validation and comparison. Blue points are stations that were used in the calibration of some of the AET products evaluated in this study, while red points were not. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

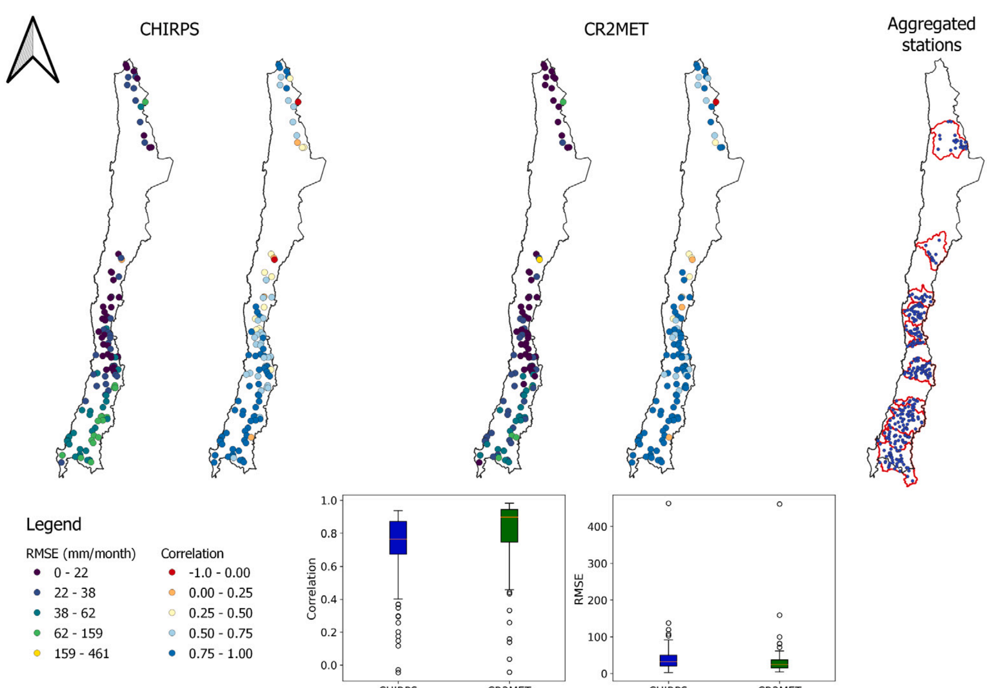


**Fig. 4.** Gridded rainfall comparison between CHIRPS and CR2MET and catchment aggregation of meteorological stations (right). RMSE is in mm month$^{-1}$.

aggregated to daily values and subsequently averaged between the years 1990 and 2020.

### 2.5. Actual evapotranspiration (AET)

In this study, five different actual evapotranspiration products were evaluated. These were selected because they correspond to publicly available global datasets, some of them identified as the "state of the art" evapotranspiration algorithms. These include:

*MODIS AET (MOD16A2)*

The MOD16A2 product is based on the Penman Monteith evapotranspiration equation (Mu et al., 2011). This product integrates land cover (MOD12Q1), albedo (MCD43), leaf area index (LAI), and fraction of photosynthetically active radiation (MOD15A2), and reanalysis of meteorological data (Mu et al., 2013) to obtain 8-day actual and potential evapotranspiration at a resolution of 500 m. This product was originally validated using 46 flux towers.

*PML AET*

The Penman–Monteith–Leuning Evapotranspiration V2 (PML_V2) consists of the sum of three components: the evaporation from the soil, the transpiration of vegetation and the evaporation of rainfall intercepted by vegetation (Zhang et al., 2019). PML uses the Penman–Monteith equation, but is improved by Leuning et al. (2008) by modifying the surface conductance formulation considering soil water and canopy losses, and using a biophysical canopy conductance model that couples gross primary production with canopy transpiration (Gan et al., 2018). Unlike other products, PML AET uses 95 flux tower stations for calibration of parameters associated with plant functional types, some of which are depicted as blue points in Fig. 3.

*SSEBop AET*

The Operational Simplified Surface Energy Balance (SSEBop) combines the formulation of potential evapotranspiration and the use of land surface temperature through the development of an energy balance. For a reference crop, the standardised Penman–Monteith equation applies. Then, actual evapotranspiration is calculated by using land surface temperatures through an evapotranspiration fraction, which varies spatially depending on water availability and vegetation health. This means anchor pixels for wet and dry conditions (cold and hot, respectively) need to be defined, and then evapotranspiration can be calculated based on these extremes in proportion to LST (Savoca et al., 2013).

*Synthesis AET*

The synthesis of global evapotranspiration was carried out by Elnashar et al. (2021). This algorithm consists of an ensemble of different global products to produce a continuous actual evapotranspiration from 1982 with a low uncertainty level irrespective of different climate and landscape conditions. Among the algorithms used in the ensemble are MODIS AET, PML AET, SSEBop, Surface Energy Balance System (SEBS), GLEAM, and TerraClimate.

*TerraClimate*

TerraClimate contains monthly variables of the climatic water balance and climate. It combines information from WorldClim, CRU Ts4.0, and the 55-year Japanese reanalysis datasets. In TerraClimate a one-dimensional modified Thornthwaite–Mather climatic water-balance model is used, which consists of a single bucket model applied to the land surface. Then, actual evapotranspiration is a result of this balance and can be expressed as the sum of liquid water supply and the soil water used (Abatzoglou et al., 2018).

We used Google Earth Engine (Gorelick et al., 2017) and Python libraries for data acquisition and processing. Subsequent methodologies and analyses were carried out using Google Colabs.

### 2.6. Catchment water balance

A monthly water balance (WB) was estimated using the following relationship:

$$\frac{\delta S}{\delta t} = P - AET - Q_{out} \tag{1}$$

where $\frac{\delta S}{\delta t}$ corresponds to the change in storage on a specific time period, $P$ corresponds to the lumped catchment precipitation, $AET$ is the lumped actual evapotranspiration and $Q_{out}$ is the discharge at the outlet station of the river. Although $\frac{\delta S}{\delta t}$ has been frequently assumed to be zero, which could apply in the very long-term, transforming Eq. (1) into $AET = P - Q_{out}$. However, this is not necessarily the case in the medium-term.

To cope with a scenario of changing storage in time, TWS and TWS anomalies from models and gravitational changes are used. For this reason, GRACE mission data have been widely used to assist the calculations of water balance (Syed et al., 2008; Sriwongsitanon et al., 2020; Wong et al., 2021). In this case, the change in storage can be formulated as:

$$\frac{\delta S}{\delta t} = \frac{TWS_{t+1} - TWS_{t-1}}{2\Delta t} \tag{2}$$

being $\Delta t$ equal to 1 month. In this case, the change in storage is considered as the difference in TWS or TWS anomalies before and after the observation.

Additionally, a simplified balance without including AET was also defined as:

$$\frac{\delta S}{\delta t} = P - E_p \tag{3}$$

being $E_p$ potential evapotranspiration. This way, AET and streamflow are excluded from the analysis.

### 2.7. Validation of AET products

Estimated catchment storage change from GLDAS, WaterGAP, and GRACE can then be compared against storage changes from the water balance using different AET models, and taking into account the latitudinal, radiative, and precipitation gradients at the regional scale. The same can then be applied to the global scale to contrast the results. Additionally, point actual evapotranspiration observations from flux towers were compared against pixel estimates of different AET products (single pixels at the flux tower location using AET products at their original resolution and single pixels of resampled AET products at the coarsest resolution of TerraClimate) and extrapolated to the catchment scale, which results in a multiscale validation that allows a comparison between validation scales (i.e., at the catchment and point scales). Flux data assimilated into gridded models (FluxCom) was also compared with AET products, both aggregated at the catchment scale.

From these results, different AET products can be ranked based on different metrics, such as:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \tag{4}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (x_i - y_i)^2}{n}} \tag{5}$$

$$NRMSE = \frac{RMSE}{x_{max} - x_{min}} \tag{6}$$

$$MAE = \frac{\sum_{i=1}^{n} |x_i - y_i|}{n} \tag{7}$$

$$bias = \frac{\sum_{i=1}^{n} (y_i - x_i)}{n} \tag{8}$$

where $r$ is the Pearson correlation, $RMSE$ is the root mean squared error, $NRMSE$ is the normalisation of $RMSE$, $MAE$ is the mean absolute error, $x_i$ are AET observations or reference TWS changes, $x_{max}$ and $x_{min}$ are the maximum and minimum observed values, $\bar{x}$ is the mean of AET or reference TWS changes, while $y_i$ are AET or water balance storage change estimations, and $\bar{y}$ is the mean of AET or water balance storage change estimations.

Statistically significant differences between AET models and TWS products were evaluated based on the residuals and for the catchment correlation coefficients after applying a Fisher z-transformation. This evaluation was done either by using one-way analysis of variance (ANOVA) or through the non-parametric Kruskal Wallis test if the ANOVA assumptions were not met. These tests were complemented by a Tukey's range test or the non-parametric Dunn's test to obtain pairwise comparisons, respectively.

Additionally, Budyko's framework of analysis (Budyko, 1974) was used for different catchments to evaluate the consistency of the AET products. This analysis restricts atmospheric water fluxes from the land surface based on energy and water availability limits. In this case, Yang et al. (2008) approach was used:

$$\frac{AET}{P} = \left[1 + \left(\frac{E_p}{P}\right)^{-n}\right]^{\frac{-1}{n}} \tag{9}$$

where the ratio $\frac{AET}{P}$ can also be expressed as evaporative index, while the ratio between $E_p$ and $P$ is also defined as the aridity or dryness index, and $n$ corresponds to a parameter that considers vegetation and catchments characteristics. In this formulation, long-term annual averages are used for the different variables used, and the parameter $n$ was obtained from a nonlinear least squares curve fit optimisation for each AET product (Virtanen et al., 2020).

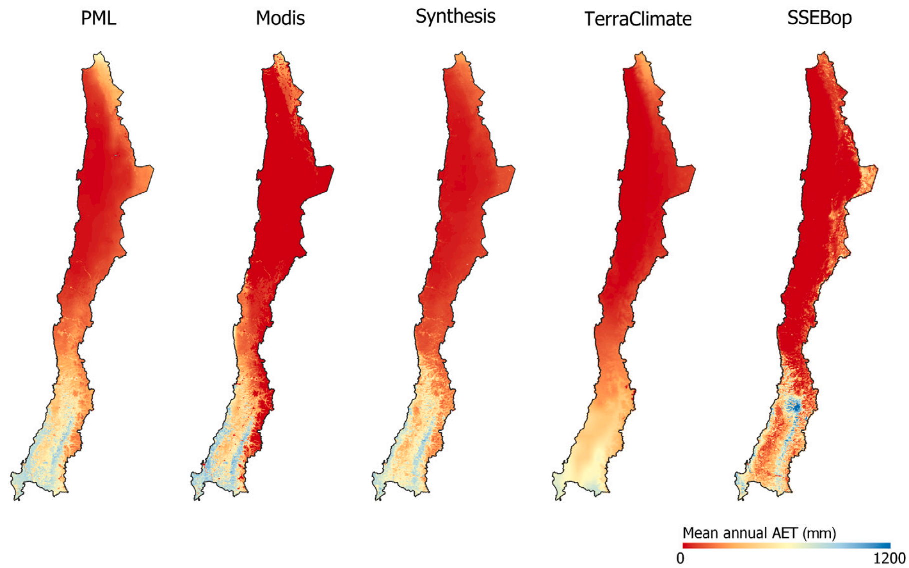


**Fig. 5.** Mean annual AET from different products in the regional scale area.

## 3. Results

Mean annual AET for different products are in Fig. 5. While all products show a similar spatial pattern, local differences can be seen among products. For instance, PML indicates larger AET in the northeastern region, possibly caused by plateau monsoons. In contrast, the central-southern region AETs have the largest variation between products, where the Mediterranean climate is related to Sclerophyll vegetation types and widespread agricultural land cover.

### *3.1. Regional catchment aggregated AET and TWS changes*

Fig. 6 indicates the 95% confidence interval of storage changes estimated through the water balance using different AET products, i.e. the median plus/minus two standard deviation, and the time series of TWS changes. This demonstrates that GLDAS, WaterGAP, and GRACE storage changes are mostly within the 95% confidence interval of the storage change estimates from the water balance and follow similar seasonal patterns, which up to some degree supports the potential of AET products. However, among TWS change products, GLDAS seems to fall more consistently within the AET confidence intervals. In the northern catchments the pattern of storage change is quite chaotic up to the Limari River catchment, where seasonal patterns start to be more evident. This behaviour is further analysed in Fig. 7 using monthly boxplots of TWS changes for the different Chilean catchments. This confirms a seasonal behaviour in storage changes that increases southwards. Northern catchments mainly occupy desert climates with very low rainfall, leading to small storage changes. This also highlights small positive changes during summer in the Loa River catchment, which are associated with plateau monsoons that occur in the Andes mountain range. Monthly deviation of TWS changes are larger in northern catchments using GRACE but increase comparatively southwards using other TWS sources.

### *3.2. Point scale AET validation*

The spatial distribution of the performance of the PML AET evaluated against flux towers is in Fig. 8. Determination coefficients ($R^2$), RMSE, and NRMSE are quite variable depending on the location of the flux tower. Larger $R^2$ and lower normalised errors are found in Europe. Some stations in western North America, in South America, and southeastern Australia are associated with low $R^2$ and high normalised errors. A similar spatial distribution of performance metrics and magnitudes are obtained using other AET products (Supplementary materials, Figure S2).

The performance of each product aggregated across time at their original resolution and flux towers is in the left panel of Fig. 9. MODIS AET has the best performance in terms of $R^2$, which means that it captures the temporal and spatial pattern best, but it is also the only product that tends to overestimate AET. The lowest errors are obtained using PML AET. The worst performance is obtained using SSEBop, which consistently underestimates AET. All products have non trivial errors, with the best $R^2$ being only 0.62, and the best RMSE being 0.78 (mm $d^{-1}$), implying about 15% of the maximum observed AET. Very small performance differences compared to the original resolution occur when AET products are resampled to the coarsest resolution of TerraClimate (right panel of Fig. 9). Therefore, resolution differences do not seem to play a significant effect on the performance of models at a spatial resolution between 500 m and about 4 km.

### *3.3. Catchment scale AET validation*

At the Chilean regional scale, the performance of predictions of $\delta S/\delta t$ per catchment using PML AET are in Fig. 10. Similar magnitudes and spatial patterns for performance metrics are obtained using all AET products and averaged rainfall from meteorological stations (Supplementary materials, Table S1). For the correlations and errors (RMSE) a clear gradient is observed. Errors tend to increase southwards even though correlations increase. This is related to the magnitude of seasonal changes in TWS and storage changes estimated through the water balance that can be observed in Figs. 6 and 7, and may be also related to the rainfall gradient shown in Fig. 1. Larger TWS changes may lead to larger errors, but also lead to a better representation of the temporal pattern. TWS changes depend on differences between water inputs and outputs in the catchments. High rainfall leads to large changes in storage, while small water inputs lead to small water losses due to the water limit in Budyko's approach, also causing small TWS changes. Therefore, the magnitude of prediction errors increases with increasing water availability, although seasonal patterns are more easily captured by models. For this reason, by normalising errors, the latitudinal error pattern changes. Evaluating the TWS change references, GLDAS leads to better results, increasing the correlation in northern catchments and reducing the errors compared with WaterGAP and GRACE.

The regional and global catchment validations per AET and TWS product are in Fig. 11. For Chilean catchments, results using gridded

 

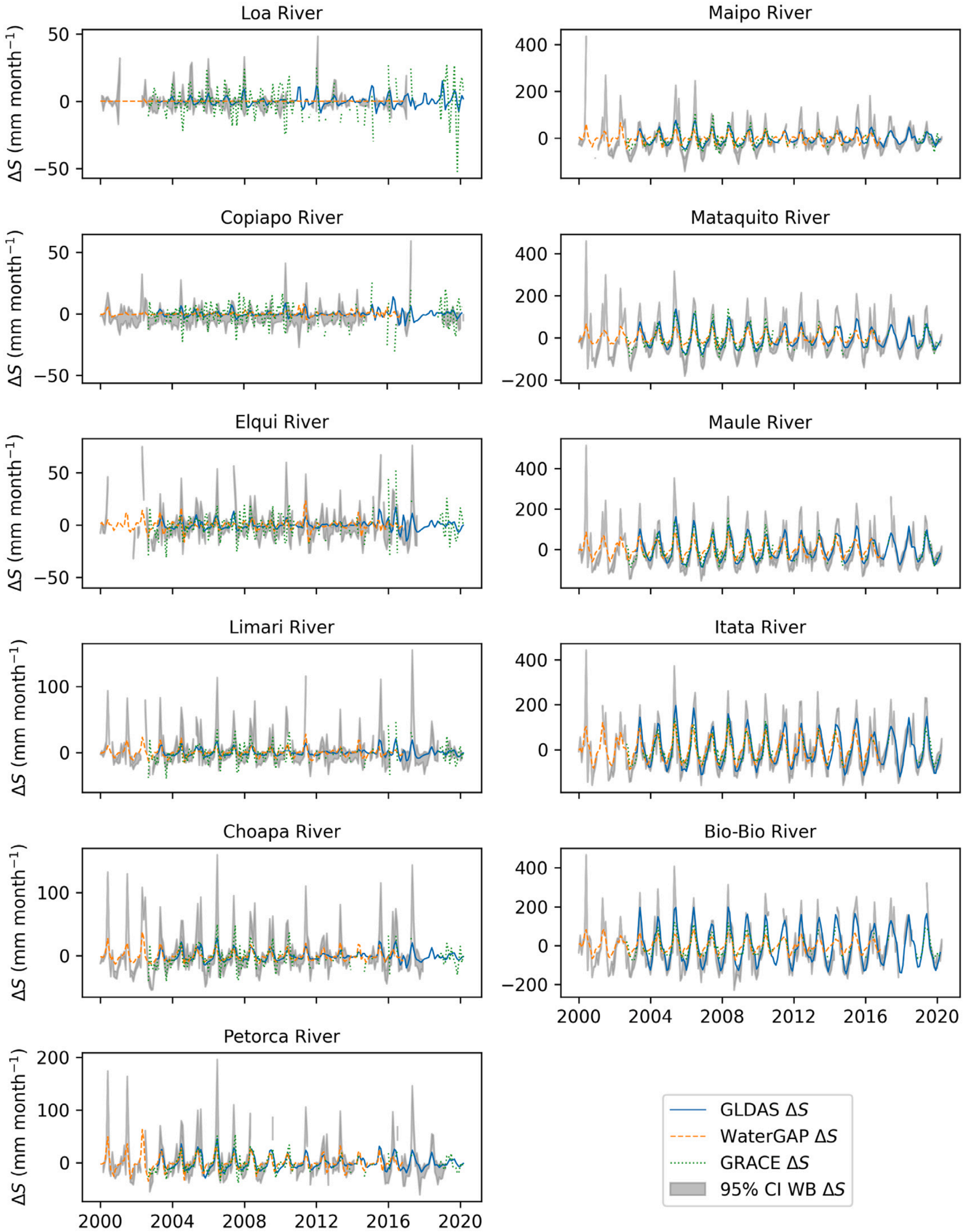


**Fig. 6.** Storage change time series using GLDAS, WaterGAP, and GRACE and 95% confidence interval of storage change using water balances for different Chilean catchments.

rainfall data from CR2MET are similar to those obtained averaging rainfall from meteorological stations per catchment (Supplementary materials, Figure S3), but leading to a slight increase in $R^2$ and errors. Small differences between products can be observed. At the Chilean regional scale, which covers a smaller range of observations compared to the global scale, the largest $R^2$ occurs using GLDAS together with Synthesis and MODIS AET (0.71), followed closely by PML (0.69) and SSEBop (0.68) AET. Non significant differences ($p$-value > 0.05) between Fisher's z-transformation of correlations were found between AET models for Chilean catchments, but statistically significant differences were found between GLDAS and WaterGAP TWS products for SSEBop and Synthesis models (Supplementary materials, Table S2). Using AET models in the water balance significantly improved results compared to the simplified balance ($P - E_p$), reducing the RMSE to almost a third and leading to a significant increase in $R^2$. On the other hand, significant differences were found for the residuals between AET models (Supplementary materials, Table S3). The lowest errors occur using TerraClimate (RMSE = 36.56) despite its lower spatial resolution (~5 km), followed by PML AET (RMSE = 38.25). Comparing TWS products, GLDAS represents a significant gain in performance with respect to WaterGAP and GRACE for several of the AET models (Supplementary materials, Table S4). Interestingly, even though Mascons solutions from GRACE should be used with caution given its low native resolution, they still lead to a relatively good performance, even outperforming WaterGAP.

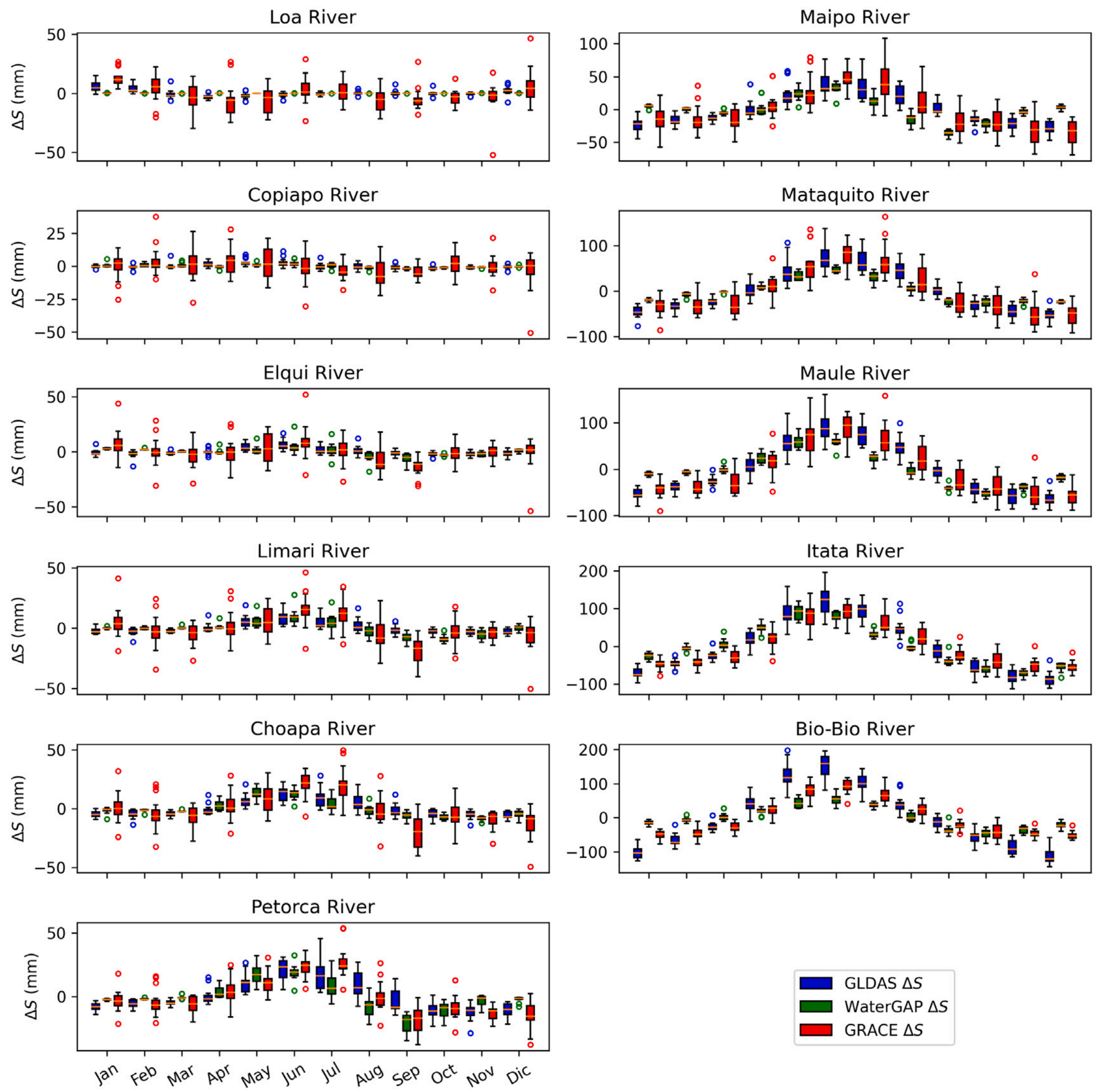


**Fig. 7.** Monthly distribution of GLDAS, WaterGAP, and GRACE TWS changes per Chilean catchment.

Different results are observed for the Australian latitudinal gradient (Figure S4), showing variable performances. For the AET models, Synthesis, SSEBop, and PML have higher $R^2$ values (0.52, 0.51, and 0.5, respectively). Conversely, Synthesis, PML, and TerraClimate AET resulted in the lowest errors, with RMSE values of 26.4, 27.0, and 28.5 mm month$^{-1}$, respectively. Notably, the use of GLDAS TWS for Australian catchments also improves the results compared to other TWS products.

At the global scale, the AET products are more homogeneous in performance and lead to a smaller but still significant improvement of results compared to the simplified balance ($P - E_p$). Maximum $R^2$ (0.59) occurs by combining GLDAS with MODIS, followed by Synthesis AET. Significant differences in Fisher's z-transformed correlations between AET models only occur using GLDAS TWS and only between TerraClimate and the rest of AET models (Supplementary materials, Table S5). Again, GLDAS TWS changes lead to the best performance regardless of the AET product used (Supplementary materials, Table S6). Significant differences between AET models were observed based on the residuals (Supplementary Materials, Table S7). The lowest errors are obtained using the FluxCom model (RMSE = 43.91), that combines flux, meteorological and satellite inputs, followed by TerraClimate (RMSE = 46.0) and Synthesis AET (RMSE = 46.72). Lowest errors at the regional and global scale are about 13% and 11% of the range of TWS changes, respectively, and tend to be similar to the magnitudes of the errors from the point scale validation.

Evaluating the extrapolation of flux towers AET in the water balance (lowest panel of Fig. 11) leads to larger errors compared with remote sensing AET models in terms of storage changes except using WaterGAP. However, this is also variable depending on the catchment (Fig. 12). The comparison of performances constrained to catchments that contain flux towers is in the Supplementary materials (Figure S5). Metrics are similar to those used in all catchments.

Budyko curves for regional and global catchments are in Fig. 13. Differences in the scale of the $x$-axis (dryness index) between regional and global Budyko curves are strong due to northern Chilean catchments, which are the driest places on Earth (Bozkurt et al., 2016). PML, MODIS and TerraClimate AET indicate consistent behaviour at both scales, with data in a few severely dry catchments exceeding the water limit. SSEBop, on the other hand, shows a large variation across catchments. Differences between models that result in some of them exceeding the water limit can be attributed to algorithm differences. Models relying on energy budgets, such as SSEBop and algorithms derived from Penman–Monteith equations (PML and MODIS), might be influenced by the calculation of net radiation and soil heat flux, particularly in arid conditions. Water balance-based models like TerraClimate tend to better preserve water volumes, thus avoiding significant exceedance of the water limit. Additionally, algorithms that integrate multiple sources

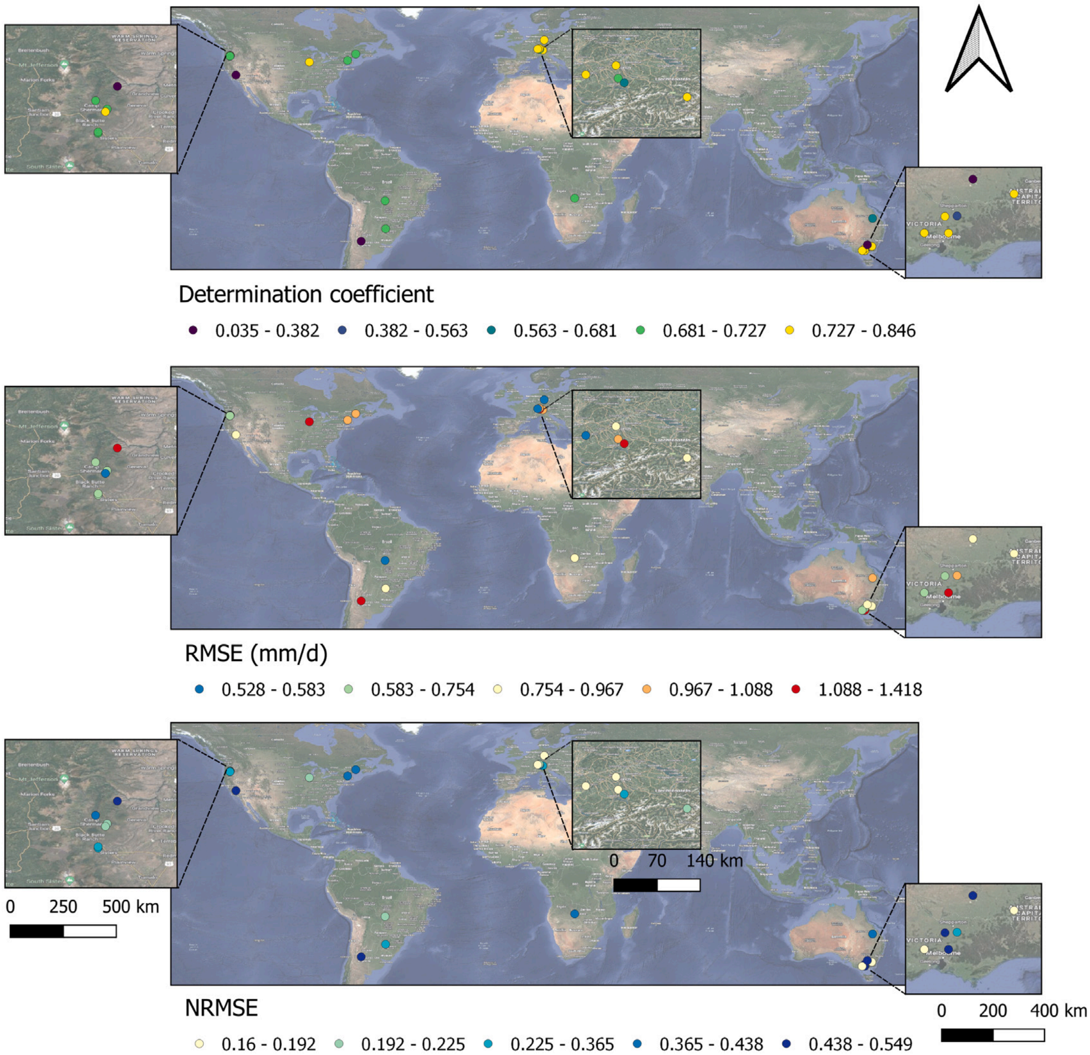


**Fig. 8.** Performance of PML AET against selected flux towers.

**Table 2**
Optimised *n* parameter from Budyko curve.

| AET product | Regional | Global |
|---|---|---|
| PML AET | 1.47 | 2.24 |
| MODIS AET | 1.32 | 1.17 |
| Synthesis AET | 0.75 | 2.05 |
| TerraClimate AET | 0.66 | 1.84 |
| SSEBop AET | 0.44 | 1.33 |

of information (such as Synthesis AET) may lead to a smoothing or attenuation of differences. At the global scale, no catchments within arid ($5 <$ Dryness index $\leq 20$) and hyperarid climates (Dryness index $> 20$) were used (Atlas, 1992), and only MODIS AET derived points did not exceed the water limit. The $n$ value from the curve optimisation varies depending on the AET product and the analysis scale (i.e., regional and global) as shown in Table 2.

The range of values associated with Budyko's $n$ parameter are close to the ranges empirically described for instance in Gunkel and Lange (2017) for $\omega$ in Fu's equation (Zhang et al., 2004), even though $n$ in Yang et al. (2008) ranges from 0 while $\omega$ in Fu's formulation ranges from 1, and values are quite variable. Global scale catchments studied lead to larger $n$ values except using MODIS AET, suggesting in general more rainfall becoming AET compared with regional catchments. This may be associated with vegetation types adapted to arid/hyperarid conditions in the regional gradient of hydroclimatic conditions while vegetation from global catchments studied might be more prone to transpiration. Variation of $n$ between models is also quite significant, indicating that PML has the highest values and therefore a larger ratio AET-rainfall than other models. On the other hand, only $n$ values from MODIS AET decrease from the regional to the global scale, implying lower evapotranspiration with respect to precipitation at the global scale.

At the regional scale there is a clear relationship between latitude and the performance of catchment scale storage changes (Fig. 14), and similarly a relationship between rainfall-radiation and catchment storage changes. This is not directly visible for global catchments. However, for the global catchments selected the precipitation range starts from about 40 mm month$^{-1}$, while for Chilean regional catchments this value leads to a good performance. More generally, lower rainfall tends to decrease the accuracy of the AET predictions. Similarly, solar incoming radiation from global catchments selected ranges from about 10 MJ m$^{-2}$ d$^{-1}$ to 22 MJ m$^{-2}$ d$^{-1}$, while Chilean catchments selected range from 18 MJ m$^{-2}$ d$^{-1}$ to about 26 MJ m$^{-2}$ d$^{-1}$ which may explain the differences. Across the latitude gradient, the regional and global catchments tend to agree. In both cases, a drop in performance is observed from latitudes −40° to −22°, region that includes the southern hemisphere subtropical ridge with very little rainfall.

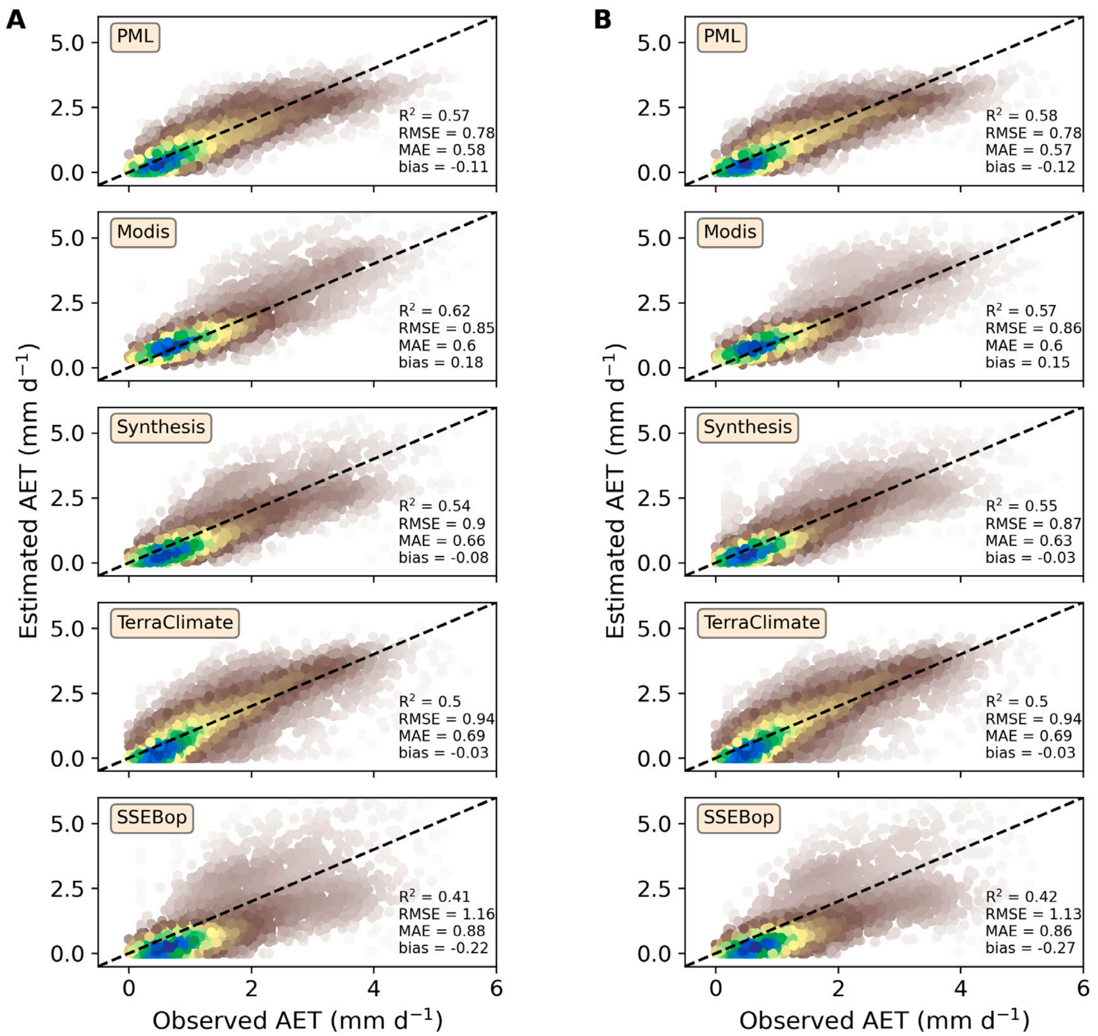


**Fig. 9.** Density scatter plots of the time series of global AET products against flux towers (point observations) at the original resolution (left panel) and resampled at the coarsest resolution of TerraClimate (right panel).

## 4. Discussion

Several studies have developed and validated AET models using point observations (Zhang et al., 2019; Mu et al., 2013; Elnashar et al., 2021; Senay et al., 2020; Guerschman et al., 2009). These highlight different performances depending on the flux towers used. For instance, Zhang et al. (2019) used 95 flux towers to validate PML AET and found an $R^2$ of 0.72 and a RMSE of 0.69. Mu et al. (2013) validated their results (MODIS AET) against 46 flux towers with a mean $R^2$ of 0.65 and an average MAE of 0.33, but with a range of $R^2$ from 0.11 to 0.91. Senay et al. (2020) used only 12 flux towers to assess SSEBop and found a mean $R^2$ of 0.44 and a mean RMSE of 0.72, but with a range of $R^2$ from 0.01 to 0.74. In Elnashar et al. (2021) different models are compared with 645 flux towers. The monthly PML AET had the best $R^2$ of 0.58 and a RMSE of 0.87 mm day$^{-1}$, followed by the Synthesis AET with a $R^2$ of 0.58 and a RMSE of 0.9 mm day$^{-1}$. In this study, 27 flux towers were used and point validation performances were within the range of results found in other studies. For instance, MODIS AET had the highest $R^2$ of 0.62 with a RMSE of 0.85 mm day$^{-1}$ and was followed by PML AET with a $R^2$ of 0.57 and a RMSE of 0.78 mm day$^{-1}$. Clearly AET products evaluated at the point scale have limitations. Moreover, there is a gap in the knowledge of how the performance of these models changes if aggregated to a larger scale. Therefore, some studies have used the water balance as an alternative to address this issue assuming a null change in storage (Guerschman et al., 2009; King et al., 2011), but this limits the study to at least annual time scales. Another validation alternative at the catchment scale was used by Sriwongsitanon et al. (2020) in Thailand, but it was limited only to subtropical climates. However, as seen in Elnashar et al. (2021) and in Salazar-Martínez et al. (2022), global AET model performance varies depending on climate and land cover types, and highest uncertainties in AET models occur in dry regions in South America (Sörensson and Ruscica, 2018).

For this reason, we also evaluated the performance of global remote sensing based AET products at the catchment scale using a water balance and comparing storage changes with TWS changes as in Sriwongsitanon et al. (2020), but including different sources with the purpose to evaluate how performances vary across scales. At both scales, AET products have limitations that result in a maximum $R^2$ of about 0.7, and RMSEs that are between 10% and 15% of the range of observations. While at both scales performances are similar in terms of $R^2$, storage changes at the catchment scale have a slightly larger RMSE compared with AET at the point scale, with mean magnitudes being 0.4 and 0.8 times larger at the regional and global scales using the best performing TWS change product (GLDAS), respectively. Additionally, gridded remote sensing AET products lead in general to lower errors than the extrapolation and use of AET from flux towers in the water balance at the catchment scale. Other validation approaches can also be included. For instance, here we demonstrated the utility of Budyko's approach, and obtained consistent results at the catchment scale, specifically using PML, MODIS and TerraClimate AET, even though some few

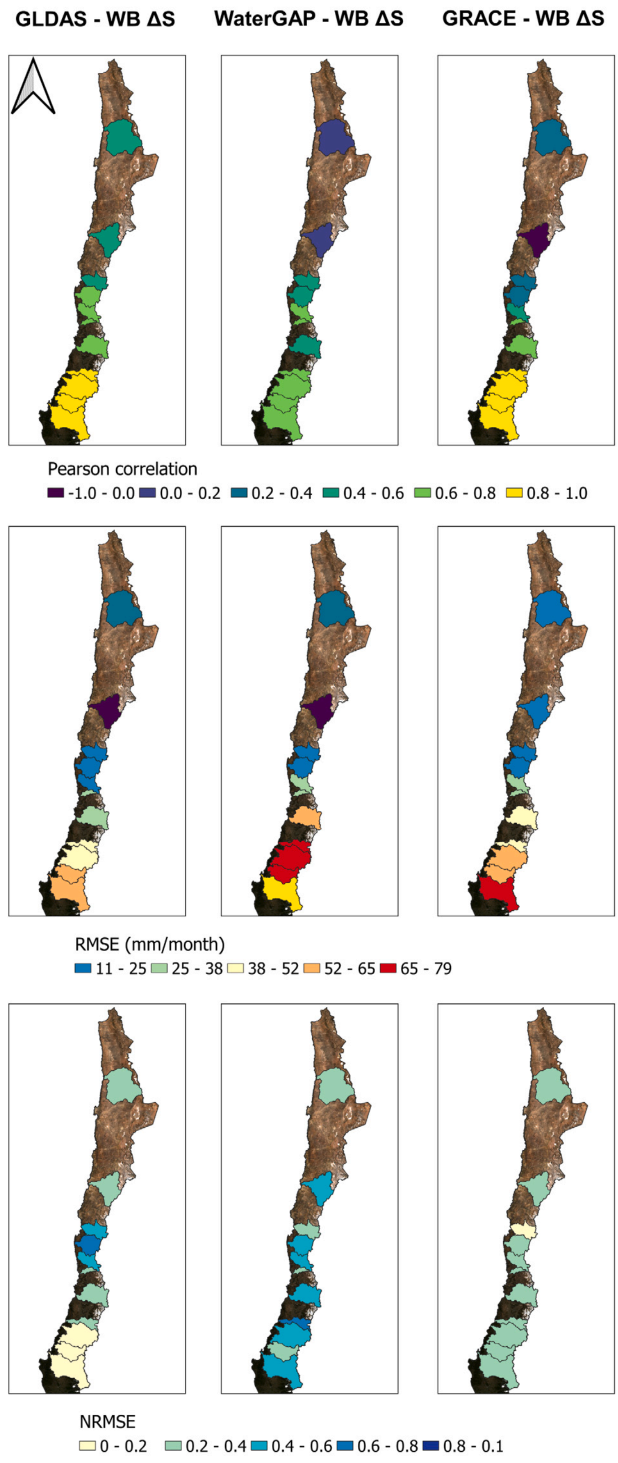


**Fig. 10.** Performance of storage changes contrasting water balances using PML AET and GLDAS, WaterGAP, and GRACE TWS in Chilean regional catchments.

catchments slightly exceeded the water limit. This has been similarly observed using global AET products in some water limited regions of South America (Sörensson and Ruscica, 2018). Budyko's *n* parameter was larger for global catchments than for regional catchments selected from the hydroclimatic gradient because the regional catchments include arid and hyperarid climate conditions, which might foster vegetation types adapted to the lack of water. Budyko's approach has been mentioned as a method to validate both rainfall and evapotranspiration remote sensing data (Koppa and Gebremichael, 2017; Mianabadi et al., 2020), and can be particularly useful in data scarce scenarios (Gunkel and Lange, 2017). AET modelling, regardless of the method, is a complex process to study (Izadifar and Elshorbagy, 2010), since it is well known that it depends on vegetation characteristics, soil water availability, and climate conditions (Allen et al., 1998; Stephenson, 1998).

Water availability is an important factor that affects AET (Jung et al., 2010) and in this study, it also seems to impact the performance of AET modelling. Similarly, in other studies climate types have led to performance differences in AET models compared to flux towers (Elnashar et al., 2021; Salazar-Martínez et al., 2022). The latitudinal gradient in AET performance in this current study is strongly influenced by rainfall and solar incoming radiation. Low rainfall relates to lower performance at the regional scale (Chile) and is also related to large variability in the performance of AET products at the global scale, particularly with respect to the temporal patterns as highlighted by Pearson correlations in Fig. 14. For the Chilean regional catchments, rainfall over 30 mm month$^{-1}$ leads to strong to very strong correlations between TWS datasets and water balance estimated storage changes (correlations greater than 0.6), but for global catchments values below 60 mm month$^{-1}$ create high variability. This can be explained by an increase in the variability of rainfall in arid and desert climates with reduced seasonality, which is evident in the low seasonality of TWS changes from Fig. 7. This also causes water available to be relatively unpredictable for evapotranspiration, leading to a poor performance of global AET models (Crawford and Gosz, 1982) and also leads to large AET uncertainties in some arid and semiarid regions from South America (Sörensson and Ruscica, 2018). Another way to conceive this may imply the "noise to signal ratio" concept. In dry regions, the signal of water balance components is very small, presenting a reduced seasonality, and therefore is very hard to capture its temporal behaviour. As a consequence, the noise to signal ratio in these regions is quite significant. On the contrary, in wet regions the seasonality signal is strong and can be easily detected. Therefore, the noise to signal ratio in these regions is small. Thus, in contrast to Sriwongsitanon et al. (2020) study carried out in Thailand, the performance of AET products was spatially variable due to different climate types. Overall, in terms of modelling the temporal patterns of AET, all products perform regular to poorly under arid conditions. However, given the low precipitation under such conditions, errors tend to be relatively small. For the latitudinal gradient of Chile this is particularly evident. When we overlay the Köppen Geiger climate types from Beck et al. (2018) with the correlations between storage changes from GLDAS and water balance changes in Chilean catchments some insights emerge. Firstly, catchments dominated by an arid desert climate in the northern region show moderate correlations. Secondly, catchments characterised by arid steppe and temperate climates with dry and hot summers in the central region exhibit strong correlations. Moving southwards, catchments dominated by a temperate climate with dry and warm summers demonstrate very strong correlations. It is clear from these climate differences that in general the accuracy of the overall water balance closure also decreases with increasing aridity. This may also account for the water limit exceedance observed in northern catchments under an arid desert climate. This deficiency in terms of AET modelling in more arid climates creates a potential for improvement, which is important since the signals of water balance components in arid and semi-arid regions are difficult to identify, indicating a priority for future research. This is specially important for water management because these regions sustain significant agricultural lands reliant on water resources (Golla, 2021), and for the protection of environmental functioning since fragile water dependant ecosystems occur in these regions (Cui and Shao, 2005).

Different factors may contribute to the uncertainties of storage changes derived from the water balance calculation. Gridded rainfall products are one of these sources. As discussed in Section 2.4, errors in rainfall are propagated into the water balance. CHIRPS rainfall

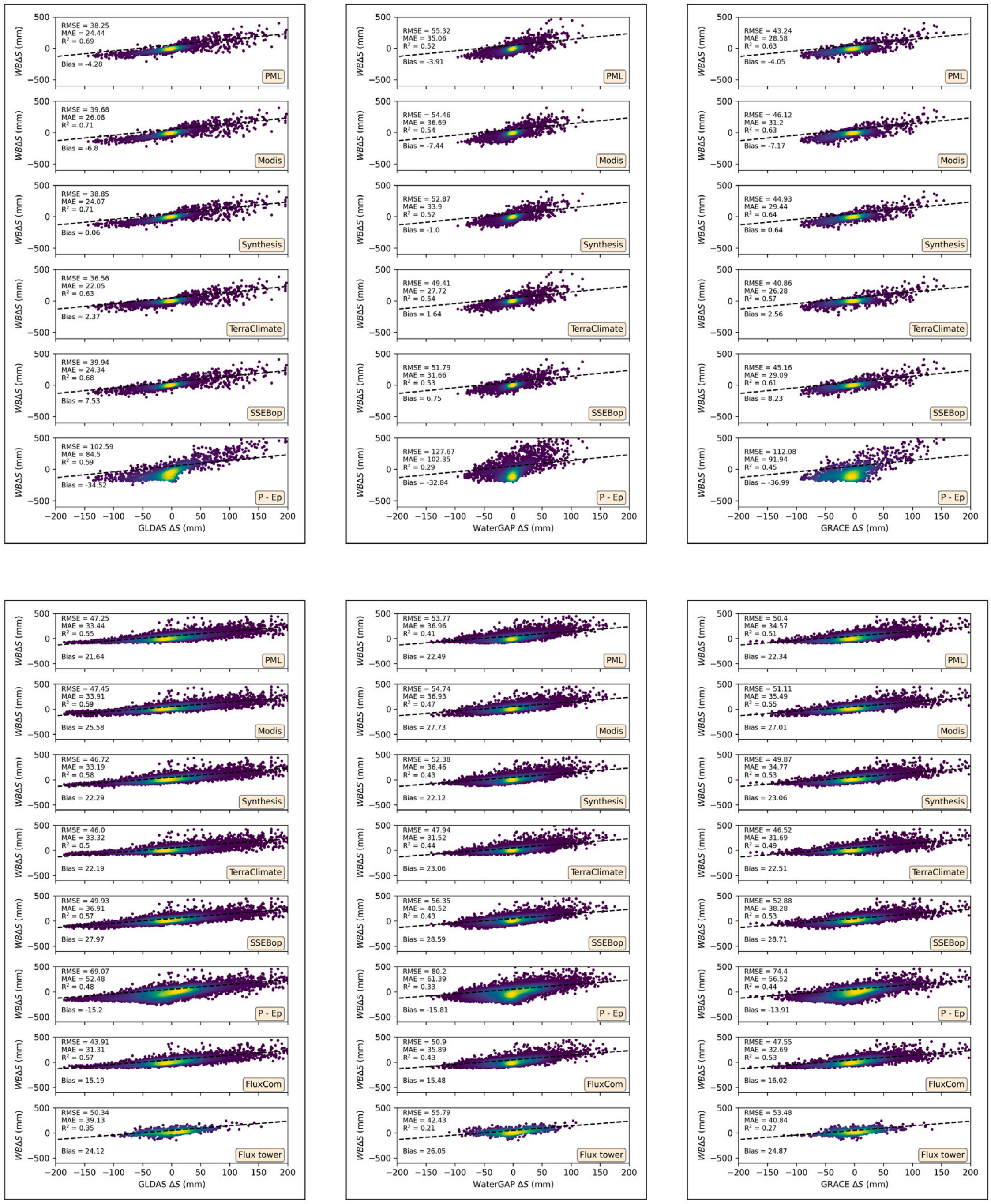


**Fig. 11.** Density scatter plots between monthly GLDAS (left column panel), WaterGAP (middle column panel), and GRACE (right column panel) TWS changes and water balance storage changes by product for Chilean (above) and GRDC (below) catchments. In the GRDC catchments, the lowest panel shows the performance using the AET from flux towers in the water balance within 11 catchments.

data provides a good representation of the observations, but in desert climates the accuracy reduces (Fuentes et al., 2022). At the regional scale, CR2MET provides more correct rainfall patterns. However, it still causes non-negligible errors compared with meteorological stations (mean bias of 7 mm month$^{-1}$ and 3 mm month$^{-1}$ for CR2MET and CHIRPS, respectively). These values are offset by AET errors in the storage change estimates for regional catchments, leading to biases that range from −7.4 mm month$^{-1}$ to 8.23 mm month$^{-1}$, but increase to over 20 mm month$^{-1}$ using global catchments. However, some global catchments are considerably larger than regional ones, which might mean that propagated errors might not cancel out, but grow with catchment size, leading to increasing errors. On the other hand, errors in observational data are not uncommon, which may apply to both rainfall observations and river discharge (McMillan et al., 2012). Di Baldassarre and Montanari (2009), for instance, conclude that discharge observation uncertainties are far from negligible. On the other hand, TWS change products were contrasted. The coarse spatial resolution of GRACE data regardless of the mascon solution resolution (Save et al., 2016) still leads to relatively good results, even outperforming WaterGAP, but it might be a result of low regional variability in areas

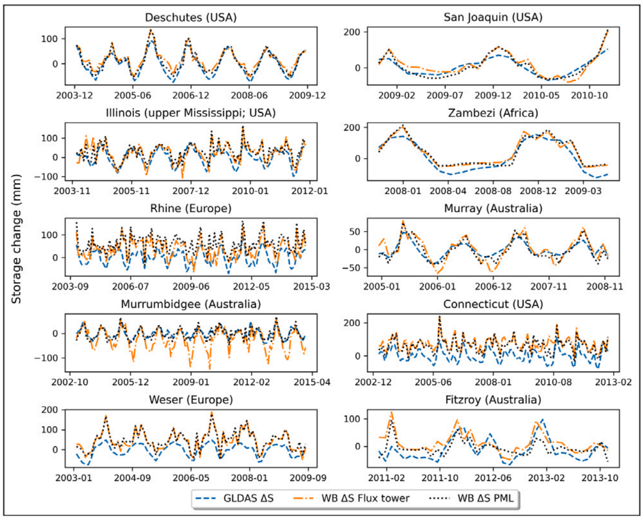


**Fig. 12.** Time series comparison in 10 catchments between GLDAS TWS changes and water balance storage changes using AET from flux towers and the PML model.

with large rainfall and reduced rainfall in highly variable arid regions. However, it is outperformed by its assimilation into GLDAS, especially on arid regions, which reflects the regional improvements described in Li et al. (2019).

Overall, the highest average correlation between observed and predicted variables is related to MODIS AET, followed closely by the Synthesis, SSEBop, and PML AET. However, the lowest errors at the point scale of validation are by using PML AET, which also has the third lowest average errors across the global catchments. On the other hand, the lowest errors at the catchment scale are obtained using FluxCom, followed by TerraClimate, although the last gives the second largest errors at the point scale validation. These results may explain some of the differences between remote sensing and hydrologic models. Here, remote sensing products are better in general at capturing spatial variability of AET compared to hydrologic models, due to their higher spatial resolution and the direct observation of landscape parameters. Process-based products, although constrained to maintain the mass balance which leads to low errors at the catchment scale in TerraClimate, are limited in representing spatial variability. SSEBop, by its part, led to a large dispersion in the Budyko curve, which may be related to the regional biases described in Senay et al. (2020). Although Elnashar et al. (2021) sought to develop an AET product through an ensemble of other AET datasets with the purpose of reducing uncertainties regardless of land cover or climate conditions, in the studied catchments it did not outperform all other products. PML AET was described as performing as well as the "state of the art" AET models (Zhang et al., 2019). In this study, this statement is partially confirmed, but it requires further attention in arid climates. Related to this, in Salazar-Martínez et al. (2022), the GLEAM model depicted good performance in arid climates against flux towers, which may be considered in future studies. Finally, since all components of the water budget have uncertainties (Levin et al., 2023), further analysis is needed to evaluate the weight and potential propagation of these uncertainties into water storage change errors. However, this requires future research.

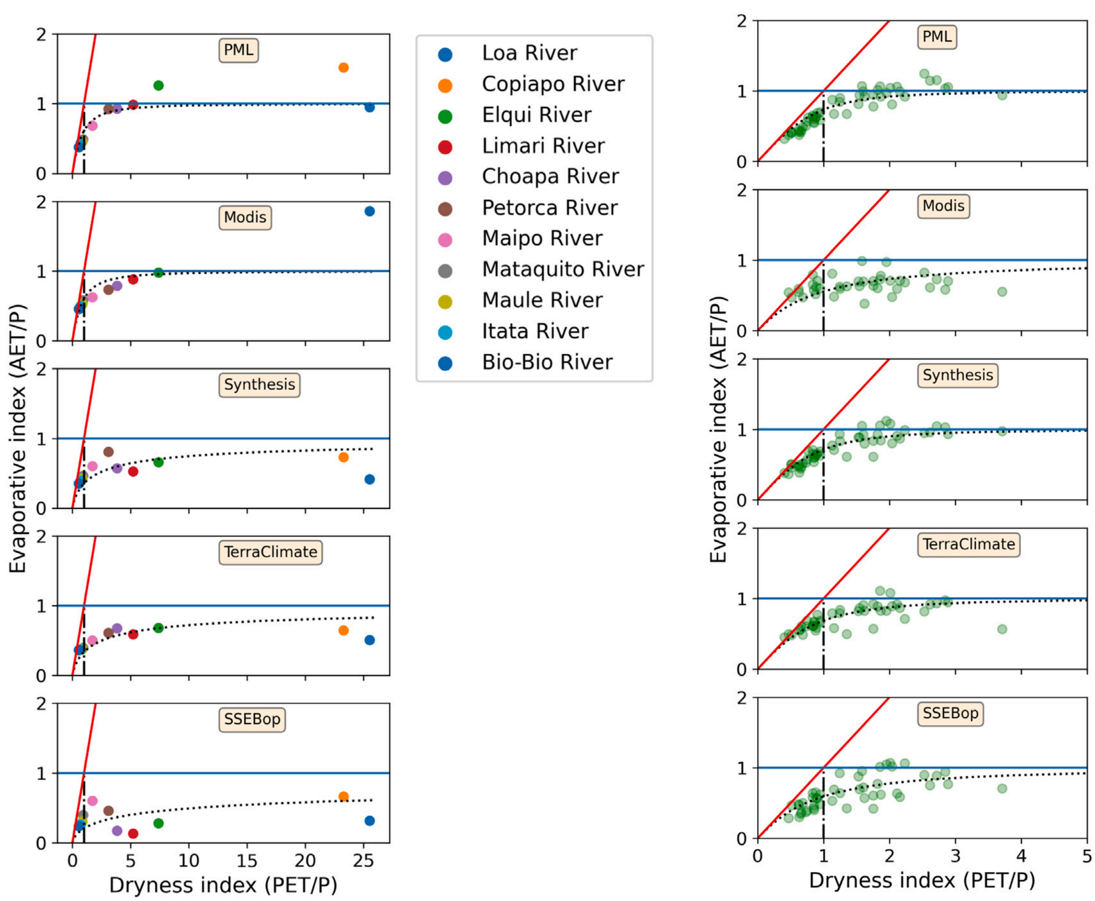


**Fig. 13.** Budyko curves applied to Chilean (left) and global (right) catchments. The blue and red continuous lines represent the water and energy limits to evapotranspiration. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

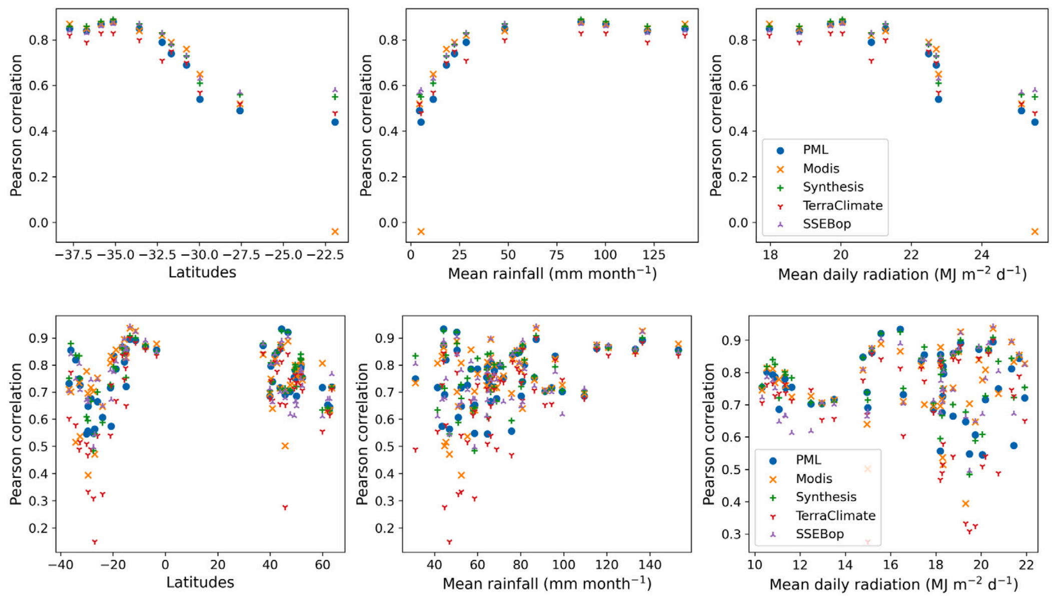

**Fig. 14.** Latitudinal (left column panel), rainfall (middle column panel), and incoming solar radiation (right column panel) gradients and their relationship with Pearson correlations in storage changes (between GLDAS TWS and catchment water balance storage changes) for Chilean (above) and global (below) catchments.

## 5. Conclusions

Different moderate resolution global AET models were evaluated using point and catchment scale validations on a latitudinal gradient in northern-central Chile at the regional scale and at different world catchments. Point validation was evaluated using flux towers, while catchment scale validation was done comparing water budget storage changes against TWS change products. Additionally, the Budyko's approach was also applied to evaluate AET products. Point and catchment scale validation tended to converge, with best performances of on average $R^2 \sim 0.7$, and errors that account for about 15% of the range of observed variables. Arid conditions led to a degradation of performances, specially in terms of temporal pattern modelling. While MODIS AET led to highest $R^2$, but closely followed by Synthesis and PML AET, PML and TerraClimate AET led to lowest errors at the point and catchment validation scales, respectively. However, these results also vary across regions. PML, MODIS, and TerraClimate AET indicate a consistent pattern in the Budyko curve, with few arid catchments exceeding the water limit. Data Assimilation of GRACE into GLDAS leads to a better regional performance compared against other TWS datasets.

Although the evaluated AET models are defined for large scale studies, having a moderate resolution, local scale remote sensing AET models at high resolution (<100 m; e.g., METRIC from EEFlux: https://eeflux-level1.appspot.com/; SEBAL: https://github.com/et-brasil/geesebal; TSEB: https://github.com/kaust-halo/geeet) may also be studied, since these can facilitate the quantification of water demand to improve irrigation management. The performance of high scale AET resolution models can also be compared against moderate resolution AET models to investigate if better resolution leads to better results. However, this is considered future research.

## CRediT authorship contribution statement

**Ignacio Fuentes:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Funding acquisition. **R. Willem Vervoort:** Supervision, Writing – review & editing. **James McPhee:** Supervision, Writing – review & editing.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Ignacio Fuentes reports financial support was provided by National Agency for Research and Development.

## Data availability

Data is publicly available and data sources are properly cited.

## Acknowledgements

This study was supported by the ANID FONDECYT Postdoctoral Project No. 3220317.

## Code availability

The code associated with this project is available in the following repository: https://github.com/IFuentesSR/AET_moderate_eval.

## Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jhydrol.2023.130477.

## References

Abatzoglou, J.T., Dobrowski, S.Z., Parks, S.A., Hegewisch, K.C., 2018. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. Sci. Data 5 (1), 1–12.

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., et al., 1998. Crop evapotranspiration-guidelines for computing crop water requirements-FAO irrigation and drainage paper 56. Fao Rome 300 (9), D05109.

Atlas, U., 1992. World atlas of desertification, vol. 80. UNEP and E. Arnold Ltd, Kent, UK.

Avtar, R., Kouser, A., Kumar, A., Singh, D., Misra, P., Gupta, A., Yunus, A.P., Kumar, P., Johnson, B.A., Dasgupta, R., et al., 2021. Remote sensing for international peace and security: Its role and implications. Remote Sens. 13 (3), 439.

Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. Sci. Data 5 (1), 1–12.

Bonilla, C.A., Vidal, K.L., 2011. Rainfall erosivity in central Chile. J. Hydrol. 410 (1–2), 126–133.

Bozkurt, D., Rondanelli, R., Garreaud, R., Arriagada, A., 2016. Impact of warmer eastern tropical Pacific SST on the March 2015 Atacama floods. Mon. Weather Rev. 144 (11), 4441–4460.

Budyko, M.I., 1974. Climate and Life. Academic Press.

Buzacott, A.J., Vervoort, R.W., 2021. Storage in south-eastern Australian catchments. Water Resour. Res. 57 (12), e2021WR029799.

Carretier, S., Tolorza, V., Regard, V., Aguilar, G., Bermúdez, M.A., Martinod, J., Guyot, J.L., Hérail, G., Riquelme, R., 2018. Review of erosion dynamics along the major NS climatic gradient in Chile and perspectives. Geomorphology 300, 45–68.

Crawford, C.S., Gosz, J.R., 1982. Desert ecosystems: their resources in space and time. Environ. Conserv. 9 (3), 181–195.

Cui, Y., Shao, J., 2005. The role of ground water in arid/semiarid ecosystems, Northwest China. Groundwater 43 (4), 471–477.

Di Baldassarre, G., Montanari, A., 2009. Uncertainty in river discharge observations: a quantitative analysis. Hydrol. Earth Syst. Sci. 13 (6), 913–921.

Döll, P., Kaspar, F., Lehner, B., 2003. A global hydrological model for deriving water availability indicators: model tuning and validation. J. Hydrol. 270 (1–2), 105–134.

Elnashar, A., Wang, L., Wu, B., Zhu, W., Zeng, H., 2021. Synthesis of global actual evapotranspiration from 1982 to 2019. Earth Syst. Sci. Data 13 (2), 447–480.

Ershadi, A., McCabe, M.F., Evans, J.P., Walker, J.P., 2013. Effects of spatial aggregation on the multi-scale estimation of evapotranspiration. Remote Sens. Environ. 131, 51–62.

Fuentes, I., Fuster, R., Avilés, D., Vervoort, W., 2021. Water scarcity in central Chile: the effect of climate and land cover changes on hydrologic resources. Hydrol. Sci. J. 66 (6), 1028–1044.

Fuentes, I., Padarian, J., Van Ogtrop, F., Vervoort, R.W., 2019. Spatiotemporal evaluation of inundated areas using MODIS imagery at a catchment scale. J. Hydrol. 573, 952–963.

Fuentes, I., Padarian, J., Vervoort, R.W., 2022. Spatial and temporal global patterns of drought propagation. Front. Environ. Sci. 140.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., et al., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Sci. Data 2 (1), 1–21.

Gan, R., Zhang, Y., Shi, H., Yang, Y., Eamus, D., Cheng, L., Chiew, F.H., Yu, Q., 2018. Use of satellite leaf area index estimating evapotranspiration and gross assimilation for Australian ecosystems. Ecohydrology 11 (5), e1974.

Garreaud, R.D., Boisier, J.P., Rondanelli, R., Montecinos, A., Sepúlveda, H.H., Veloso-Aguila, D., 2020. The central Chile mega drought (2010–2018): a climate dynamics perspective. Int. J. Climatol. 40 (1), 421–439.

Golla, B., 2021. Agricultural production system in arid and semi-arid regions. Int. J. Agric. Sci. Food Technol. 7, 234–244.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.

GRDC, 2011. Watershed boundaries of GRDC stations / global runoff data centre.

Guerschman, J.P., Van Dijk, A.I., Mattersdorf, G., Beringer, J., Hutley, L.B., Leuning, R., Pipunic, R.C., Sherman, B.S., 2009. Scaling of potential evapotranspiration with MODIS data reproduces flux observations and catchment water balance observations across Australia. J. Hydrol. 369 (1–2), 107–119.

Gunkel, A., Lange, J., 2017. Water scarcity, data scarcity and the budyko curve—An application in the Lower Jordan River Basin. J. hydrol.: Reg. Stud. 12, 136–149.

Han, J., Yang, Y., Roderick, M.L., McVicar, T.R., Yang, D., Zhang, S., Beck, H.E., 2020. Assessing the steady-state assumption in water balance calculation across global catchments. Water Resour. Res. 56 (7), e2020WR027392.

Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R., Xie, P., Yoo, S.-H., 2015. NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG). In: Algorithm Theoretical Basis Document (ATBD) Version, 4(26).

Hund, S.V., Johnson, M.S., Keddie, T., 2016. Developing a hydrologic monitoring network in data-scarce regions using open-source arduino dataloggers. Agric. Environ. Lett. 1 (1), 160011.

Izadifar, Z., Elshorbagy, A., 2010. Prediction of hourly actual evapotranspiration using neural networks, genetic programming, and statistical models. Hydrol. Process. 24 (23), 3413–3425.

Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., Reichstein, M., 2019. The FLUXCOM ensemble of global land-atmosphere energy fluxes. Sci. Data 6 (1), 74.

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S.I., Sheffield, J., Goulden, M.L., Bonan, G., Cescatti, A., Chen, J., De Jeu, R., et al., 2010. Recent decline in the global land evapotranspiration trend due to limited moisture supply. Nature 467 (7318), 951–954.

King, E., Van Niel, T., Van Dijk, A., Wang, Z., Paget, M., Raupach, T., Guerschman, J., Haverd, V., McVicar, T., Miltenburg, I., et al., 2011. Actual evapotranspiration estimates for Australia: Intercomparison and evaluation. In: Water Information Research and Development Alliance Science Symposium, Melbourne, Australia. pp. 1–5.

Koppa, A., Gebremichael, M., 2017. A framework for validation of remotely sensed precipitation and evapotranspiration based on the budyko hypothesis. Water Resour. Res. 53 (10), 8487–8499.

Kornfeld, R.P., Arnold, B.W., Gross, M.A., Dahya, N.T., Klipstein, W.M., Gath, P.F., Bettadpur, S., 2019. GRACE-FO: the gravity recovery and climate experiment follow-on mission. J. Spacecr. Rocket. 56 (3), 931–951.

Koster, R.D., Suarez, M.J., Ducharne, A., Stieglitz, M., Kumar, P., 2000. A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure. J. Geophys. Res.: Atmos. 105 (D20), 24809–24822.

Leuning, R., Zhang, Y., Rajaud, A., Cleugh, H., Tu, K., 2008. A simple surface conductance model to estimate regional evaporation using MODIS leaf area index and the Penman-Monteith equation. Water Resour. Res. 44 (10).

Levin, S.B., Briggs, M.A., Foks, S.S., Goodling, P.J., Raffensperger, J.P., Rosenberry, D.O., Scholl, M.A., Tiedeman, C.R., Webb, R.M., 2023. Uncertainties in measuring and estimating water-budget components: Current state of the science. Wiley Interdiscip. Rev.: Water e1646.

Li, B., Rodell, M., Kumar, S., Beaudoing, H.K., Getirana, A., Zaitchik, B.F., de Goncalves, L.G., Cossetin, C., Bhanja, S., Mukherjee, A., et al., 2019. Global GRACE data assimilation for groundwater and drought monitoring: Advances and challenges. Water Resour. Res. 55 (9), 7564–7586.

Li, M., Shao, Q., 2010. An improved statistical approach to merge satellite rainfall estimates and raingauge data. J. Hydrol. 385 (1–4), 51–64.

Madariaga, A., Maillet, A., Rozas, J., 2021. Multilevel business power in environmental politics: The avocado boom and water scarcity in Chile. Environ. Polit. 30 (7), 1174–1195.

McMillan, H., Krueger, T., Freer, J., 2012. Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. Hydrol. Process. 26 (26), 4078–4111.

Mianabadi, A., Davary, K., Pourreza-Bilondi, M., Coenders-Gerrits, A., 2020. Budyko framework; towards non-steady state conditions. J. Hydrol. 588, 125089.

Mu, Q., Zhao, M., Running, S.W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sens. Environ. 115 (8), 1781–1800.

Mu, Q., Zhao, M., Running, S.W., 2013. MODIS global terrestrial evapotranspiration (ET) product (NASA MOD16A2/A3). Algorithm Theor. Basis Doc. Collect. 5, 600.

Müller Schmied, H., Cáceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., Peiris, T.A., Popat, E., Portmann, F.T., Reinecke, R., et al., 2021. The global water resources and use model waterGAP v2. 2d: Model description and evaluation. Geosci. Model Dev. 14 (2), 1037–1079.

Muñoz-Sabater, J.n., Dutra, E., Agustí Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., et al., 2021. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. Earth Syst. Sci. Data 13 (9), 4349–4383.

Palmer, S.C., Kutser, T., Hunter, P.D., 2015. Remote sensing of inland waters: Challenges, progress and future directions. Remote Sens. Environ. 157, 1–8.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., et al., 2020. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Sci. Data 7 (1), 1–27.

Salazar-Martínez, D., Holwerda, F., Holmes, T.R., Yépez, E.A., Hain, C.R., Alvarado-Barrientos, S., Ángeles-Pérez, G., Arredondo-Moreno, T., Delgado-Balbuena, J., Figueroa-Espinoza, B., et al., 2022. Evaluation of remote sensing-based evapotranspiration products at low-latitude eddy covariance sites. J. Hydrol. 610, 127786.

Sarricolea, P., Herrera-Ossandon, M., Meseguer-Ruiz, O., 2017. Climatic regionalisation of continental Chile. J. Maps 13 (2), 66–73.

Save, H., 2020. CSR GRACE and GRACE-FO RL06 Mascon Solutions v02. 2020, Vol. 10. University of Texas [data set].

Save, H., Bettadpur, S., Tapley, B.D., 2016. High-resolution CSR GRACE RL05 mascons. J. Geophys. Res.: Solid Earth 121 (10), 7547–7569.

Savoca, M.E., Senay, G.B., Maupin, M.A., Kenny, J.F., Perry, C.A., et al., 2013. Actual Evapotranspiration Modeling Using the Operational Simplified Surface Energy Balance (SSEBop) Approach. US Department of the Interior, US Geological Survey.

Senay, G.B., Kagone, S., Velpuri, N.M., 2020. Operational global actual evapotranspiration: Development, evaluation, and dissemination. Sensors 20 (7), 1915.

Senay, G., Leake, S., Nagler, P., Artan, G., Dickinson, J., Cordova, J., Glenn, E., 2011. Estimating basin scale evapotranspiration (ET) by water balance and remote sensing methods. Hydrol. Process. 25 (26), 4037–4049.

Sörensson, A.A., Ruscica, R.C., 2018. Intercomparison and uncertainty assessment of nine evapotranspiration estimates over south america. Water Resour. Res. 54 (4), 2891–2908.

Sriwongsitanon, N., Suwawong, T., Thianpopirug, S., Williams, J., Jia, L., Bastiaanssen, W., 2020. Validation of seven global remotely sensed ET products across thailand using water balance measurements and land use classifications. J. Hydrol.: Reg. Stud. 30, 100709.

Stephenson, N., 1998. Actual evapotranspiration and deficit: biologically meaningful correlates of vegetation distribution across spatial scales. J. Biogeogr. 25 (5), 855–870.

Strong, J.A., Elliott, M., 2017. The value of remote sensing techniques in supporting effective extrapolation across multiple marine spatial scales. Mar. Pollut. Bull. 116 (1–2), 405–419.

Syed, T.H., Famiglietti, J.S., Rodell, M., Chen, J., Wilson, C.R., 2008. Analysis of terrestrial water storage changes from GRACE and GLDAS. Water Resour. Res. 44 (2).

Tapley, B.D., Bettadpur, S., Watkins, M., Reigber, C., 2004. The gravity recovery and climate experiment: Mission overview and early results. Geophys. Res. Lett. 31 (9).
Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods 17, 261–272.
Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. Remote Sens. Environ. 236, 111402.
Wong, J.S., Zhang, X., Gharari, S., Shrestha, R.R., Wheater, H.S., Famiglietti, J.S., 2021. Assessing water balance closure using multiple data assimilation–and remote sensing–based datasets for canada. J. Hydrometeorol. 22 (6), 1569–1589.
Worqlul, A.W., Maathuis, B., Adem, A.A., Demissie, S.S., Langan, S., Steenhuis, T.S., 2014. Comparison of rainfall estimations by TRMM 3b42, MPEG and CFSR with ground-observed data for the lake tana basin in ethiopia. Hydrol. Earth Syst. Sci. 18 (12), 4871–4881.
Yang, H., Yang, D., Lei, Z., Sun, F., 2008. New analytical derivation of the mean annual water-energy balance equation. Water Resour. Res. 44 (3).
Zhang, L., Hickel, K., Dawes, W., Chiew, F.H., Western, A., Briggs, P., 2004. A rational function approach for estimating mean annual evapotranspiration. Water Resour. Res. 40 (2).
Zhang, Y., Kong, D., Gan, R., Chiew, F.H., McVicar, T.R., Zhang, Q., Yang, Y., 2019. Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017. Remote Sens. Environ. 222, 165–182.