Journal of Hydrology: Regional Studies 56 (2024) 102031

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

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

Evaluation of national and international gridded meteorological products for rainfall-runoff modelling in Northern Italy

Gökhan Sarigil, Mattia Neri *, Elena Toth

Department of Civil, Chemical, Environmental and Materials Engineering, University of Bologna, Viale Del Risorgimento 2, Bologna 40136, Italy

A R T I C L E I N F O

Keywords:
Gridded meteorological products
Rainfall-runoff modeling
Indirect validation
E-OBS
ERA5-Land
CHIRPS
Northern Italy

A B S T R A C T

Study region: Northern Italy.
Study focus: Gridded meteorological products provide spatially-distributed meteorological forcings, facilitating hydrological modeling in large-scale experiments. However, their accuracy, in particular as far as precipitation is concerned, varies considerably in space and time, and rigorous validation of these products is essential before their application. This study conducts a large-scale evaluation of five meteorological datasets in Northern Italy through i) a direct comparison of precipitation and temperature estimates and ii) an indirect validation, assessing their ability to reproduce streamflow when used to force the CemaNeige-GR6J hydrological model. The tested datasets include two gauged-based products, namely SCIA (the reference gridded dataset from the Italian Institute for Environmental Protection and Research) and E-OBS, two products based on reanalyses (the global ERA5-Land and the national MERIDA) and a gauged-corrected global satellite precipitation product (CHIRPS).
New hydrological insights for the region: Gauge-based datasets provide the best streamflow simulations when the underlying station density is high: SCIA, based on a uniform and dense gauge network across the entire study area, confirms to be the best choice as the climatic reference dataset, while the use of E-OBS is not recommended in Piedmont due to the low number of stations. In areas with low station density, reanalyses may yield to more accurate results: among reanalysis-based products, the Italian MERIDA dataset outperforms ERA5-Land. Finally, CHIRPS results to be the least accurate precipitation dataset.

1. Introduction

Recent advancements in the availability of gridded meteorological products (GMPs) on both national and international scales have expanded the possibility to apply hydrological models on large sets of catchments at regional, continental or even global domains. GMPs provide spatially continuous estimates of meteorological variables (and in particular precipitation and temperature, that are crucial for hydrological modelling) derived from diverse monitoring sources (ground-based stations, satellite retrievals, and blends of these datasets) or from reanalysis (combining atmospheric dynamical models with data assimilation). However, concerns over the accuracy and suitability of GMPs as forcing to rainfall-runoff models are still significant, especially with regard to precipitation, which exhibits high spatiotemporal variability (Mankin et al., 2024). Factors influencing GMP reliability include the underlying data sources, interpolation methods, spatial and temporal resolutions, and the complexity of the terrain.

* Corresponding author.
E-mail address: mattia.neri5@unibo.it (M. Neri).

https://doi.org/10.1016/j.ejrh.2024.102031
Received 7 August 2024; Received in revised form 11 October 2024; Accepted 15 October 2024
Available online 4 November 2024

The accurate representation of precipitation and temperature through appropriate GMPs is crucial for reliable rainfall-runoff modelling and its subsequent applications in flood forecasting, water resource management, and climate change impact assessments (Fekete et al., 2004). Errors in precipitation inputs, whether in terms of amount, intensity, spatial distribution, or temporal patterns, can propagate through the model, significantly impacting the accuracy of simulated streamflow characteristics such as volume, timing, and peak magnitudes (McMillan et al., 2011; Mei et al., 2016). Additionally, temperature plays a pivotal role in hydrological processes, influencing snow accumulation-melt and evapotranspiration, which in turn affects soil moisture dynamics (Dembélé et al., 2020), further emphasizing the need for GMPs that accurately capture both precipitation and temperature for robust hydrological modelling. Rigorous validation of GMPs is therefore essential to determine their accuracy and suitability for rainfall-runoff modelling. However, it is important to note that the performance of a GMP may vary across regions, necessitating region-specific validation assessments. A substantial body of international research has focused on evaluating the performance of GMPs, employing both direct and indirect (i.e. through the application of hydrological models) validation approaches in different regions of the world.

Direct validation centres on the comparison of GMP estimates against reliable reference datasets, generally with ground-based meteorological observations serving as the benchmark. On the other hand, researchers often face the reality of sparse rain gauge networks (Kidd et al., 2017), especially in regions with limited meteorological infrastructure or areas with challenging terrain: in such cases reanalyses or satellite based GMPs are used as the reference base (e.g. Gehne et al., 2016; Houngnibo et al., 2023).

Indirect validation offers a holistic assessment of how hydrological model performance is influenced by the choice of GMPs (Seibert et al., 2024). By comparing simulations driven by different GMPs against observed streamflow records, researchers gain insights into the complex interactions between meteorological forcing data quality, catchment characteristics, and the hydrological model itself (e.g. Baez-Villanueva et al., 2021; Beck et al., 2017; Mazzoleni et al., 2019; Satgé et al., 2021; Zhang et al., 2021). This approach complements direct validation by revealing how GMP errors translate into simulated hydrological responses, potentially being amplified, dampened, or masked within the integrated model system.

Crucially, indirect validation highlights the context-dependent nature of GMP suitability. A GMP demonstrating relatively low errors in direct validation might lead to poor hydrological simulations due to its inability to capture critical spatiotemporal patterns of meteorological forcings that drive hydrological processes. Studies across diverse geographic regions have underscored this interplay between GMP characteristics and model response (Nijssen and Lettenmaier, 2004; Stephens et al., 2022).

The complex topography and diverse climate regimes of Italy necessitate a particularly careful evaluation of GMPs for hydrological applications. While numerous validation studies exist on a global scale, a comprehensive assessment specific to the Italian context remains limited. Some notable studies have conducted direct validation of specific types of GMPs in Italy, revealing significant performance variation across different regions. For instance, Turco et al. (2013a, 2013b) highlighted the discrepancies between different ground-based precipitation products in northwestern and northern Italy, while Duan et al. (2016), Gentilucci et al. (2022) and Maggioni et al. (2017) studied the reliability of different satellite-based precipitation datasets in the Adige River basin, in the Apennines and in the Eastern Italian Alps respectively, and showed how it may strongly vary in space. Cavalleri et al. (2024a) conducted an assessment of reanalysis products across Italy and Longo-Minnolo et al. (2022) focused on ERA5-Land performance in Sicily, emphasizing the benefits of combining reanalysis data with local observations. On the other hand, comparisons of multiple types of products are more limited (e.g. Cammalleri et al., 2024; Caroletti et al., 2019; Longo-Minnolo et al., 2022; Padulano et al., 2021).

Indirect validation studies have explored GMP impacts on hydrological modelling in Italy. For example, Alfieri et al. (2022) and Camici et al. (2020, 2018) focused on large-scale applications and data-scarce regions, assessing the performance of satellite-based GMPs in the Po River basin and multiple Mediterranean basins. Studies in the Adige catchment and southeastern Alps by Laiti et al. (2018) and Dalla Torre et al. (2024) emphasized the importance of gauge network density and revealed performance issues with some reanalysis products. Research in the Italian Alps by Tuo et al. (2016) demonstrated the significant impact of precipitation input on model performance, with traditional interpolated data often outperforming satellite-based alternatives. Nikolopoulos et al. (2013) examined the use of satellite rainfall products for simulating flash flood events, highlighting biases that impacted flood simulations.

Overall, previous studies have provided valuable insights into the performance of several GMPs in representing different features (extreme events, seasonality, etc.) and their performance when used in hydrological modelling. However, each of them mainly focused on the validation of a specific typology of products, while a more comprehensive assessment of the wide array of available GMPs in Italy is still missing, encompassing datasets of different nature (i.e. gauge-based, reanalysis, satellite-based). This study addresses this need by conducting a large-scale direct and indirect evaluation of GMPs of different typology in Northern Italy, encompassing 158 river catchments. For each category of products, the tested GMPs are selected among those most commonly adopted in practical applications, including both national and international datasets. Our approach involves a two-step validation process. First, we conduct a direct comparison of precipitation and temperature estimates within the study area. Second, we employ an indirect validation method, assessing the ability of each GMP to reproduce streamflow when used as input for a conceptual hydrological model, CemaNeige-GR6J. This dual approach ensures a comprehensive evaluation of GMP performance, highlighting both strengths and limitations.

This paper is structured as follows. Section 2 describes the study region, including geographical characteristics, catchment classification and data collection. Section 3 provides an overview of the GMPs used in the study, details their characteristics and reviews previous comparative studies. Section 4 explains the methods, employed both for direct comparison of GMPs against the reference dataset and for the indirect validation through rainfall-runoff modelling. Section 5 presents and discusses the results. Finally, Section 6 summarizes the key findings and highlights the useful insights of the work for future research and practical application in the region.

2. Study region

The case study is composed of a set of 158 catchments covering a large portion of the Northern Apennines and Western Alps, in Northern Italy (Fig. 1), where sufficiently long time-series of daily streamflow are available as open data from the different regional hydrological services. The basins are classified into three different groups based on the three different Italian administrative regions they belong to: Emilia-Romagna, Piedmont and Tuscany (blue, green and orange shading respectively in Fig. 1c). We decide to keep such classification mainly because Italian hydro-meteorological data are collected and managed by different regional agencies and, as it will be described below, such fragmentation impacts on data availability. In addition, the three regions are representative of a variety of morphological and climatic catchment features. Therefore, the identified large sample of catchments covers a wide spectrum of hydrological behaviours across Northern Italy and allows to evaluate the consistency of different GMPs with the rainfall-runoff response in the study region.

Piedmont, located in the Western Alps and their foothills, has a complex topography that greatly affects its climate. The Alps surround the region on its northern, western and south-western borders, with mountains accounting for 43 % of its territory and hills comprising another 30 % (De Luca et al., 2020). On the other hand, the Emilia-Romagna region includes the North-eastern side of the Apennine mountains and the Southern Po valley. Finally, Tuscany region is characterized by a heterogeneous physical landscape, spanning from coastal plains and river valleys to the southern side of northern Apennines and other chains (Crisci et al., 2002).

The study focuses on mountainous and foothill catchments in the three regions, selected by considering a minimum outlet elevation of 50 m above sea level. Differently from other Italian regions, daily streamflow data are openly accessible and available for a sufficient number of years, and they can be obtained from the three regional agencies managing hydroclimatic data: ARPA-Piemonte for Piedmont, ARPAE for Emilia-Romagna, and SIR for Tuscany (see details in Data Availability Section). Manual validation of the streamflow data was performed, ensuring a minimum observation period of 6 years within 1986 and 2022. Table 1 reports statistics on gauging period lengths and main physical, climatic and hydrological catchment characteristics. Climatic statistics are here derived based on SCIA, i.e. one of the meteorological datasets which will be presented in the Section 3.2 and which will be adopted as reference GMP. It can be noticed that catchments features vary considerably across the study area: basin drainage areas range from 2 km^2 to more than 5000 km^2, including very high elevation and steep watersheds and foothill catchments characterised by gentle reliefs. Mean annual precipitation spans from around 800 mm in the lowlands to over 2000 mm in the highest catchments of the Apennines and Alps. The aridity index varies from 0.3 to 1, meaning that the watersheds are mostly wet or weakly arid (mean annual evapotranspiration never exceeds significantly annual precipitation).

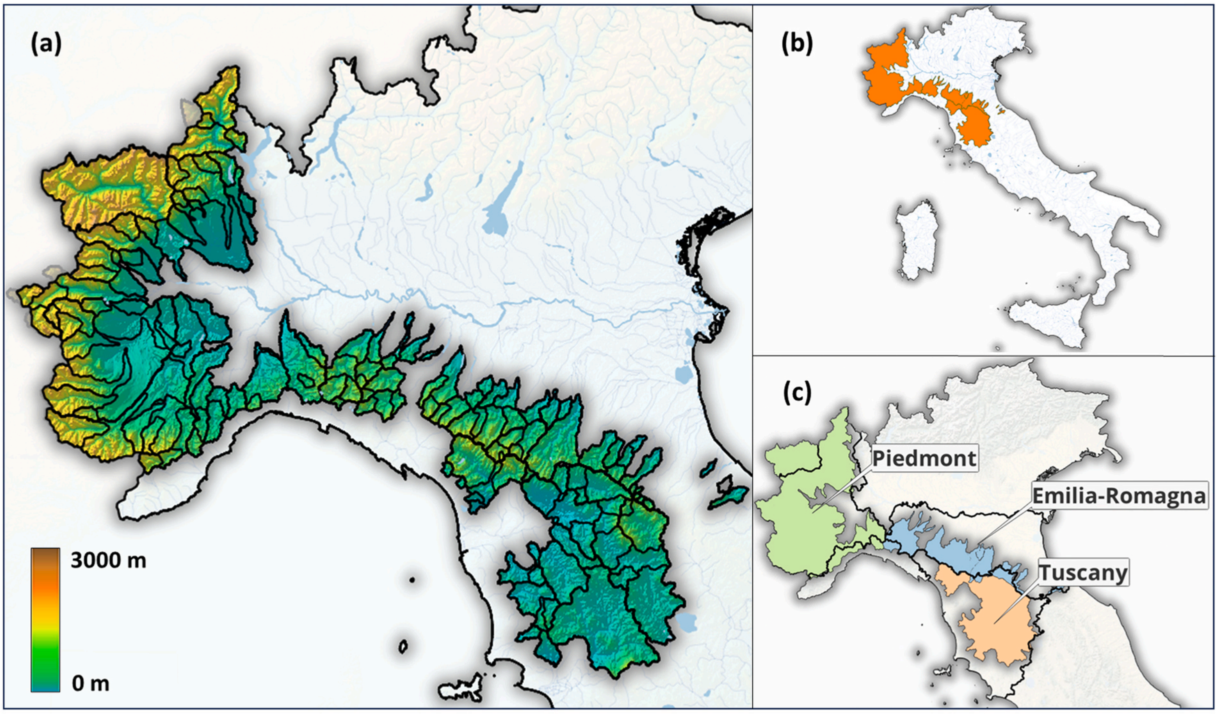

Fig. 1. a) Study region and catchment boundaries; (b) location of the study area within Italy; (c) catchment classification based on the administrative regions they belong to: Piedmont (green), Emilia-Romagna (blue), and Tuscany (orange).

Table 1
Summary of catchment characteristics and daily streamflow statistics for 158 catchments across three regions.

	Piedmont (68 catchments)			**Emilia-Romagna** (55 catchments)			**Tuscany** (35 catchments)		
	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
Catchment Area (km^2)	44	895	5362	2	232	1050	29	985	5345
Mean Elevation (m)	241	1161	2221	177	807	1410	183	467	886
Mean Slope (°)	3.3	20.4	33	8.4	14.7	21.1	6.6	15	33.6
Gauging Period (years)	15	20	28	6	15.5	19	13	23	35
Average Annual Runoff (mm/year)	138.5	652.4	1381.1	200.4	710.0	1998.0	119.9	475.1	1736.7
Runoff Coefficient of Variation (CV)	0.77	1.86	5.17	1.23	2.01	4.04	1.30	2.20	3.47
Mean Annual Precipitation (mm)	744	1110	1785	826	1256	2085	785.1	1020	2038
Aridity	0.31	0.57	1.04	0.28	0.57	0.88	0.34	0.71	0.98

3. Gridded meteorological products

In this study we evaluate the accuracy of the five GMPs presented in Table 2, which summarizes their key characteristics, including temporal coverage, temporal resolution, spatial coverage, spatial resolution, and underlying methodology/source. They include two datasets derived through spatial interpolation of ground observations (SCIA and E-OBS), two products based on reanalysis (ERA5-Land and MERIDA), and one gauge-corrected satellite-based product (CHIRPS).

In between gauge-based products, the SCIA ("Sistema nazionale per la raccolta, elaborazione e diffusione di dati Climatologici di Interesse Ambientale") dataset (Desiato et al., 2011, 2007), curated by the Italian Institute for Environmental Protection and Research (ISPRA), encompasses daily gridded precipitation data at 10 km resolution from 1961 onwards and minimum/maximum temperature data at 5 km resolution across Italy from 1981. It is sourced from a high-density network of regional meteorological services, though it's worth noting that only a subset of these data points is available as individual station time series. Precipitation is derived using Inverse Distance Weighting of gauge measurements, while temperatures are interpolated through Gaussian Process Regression. It must be acknowledged that ISPRA offers an additional product, named BIGBANG (Braca et al., 2021), which provides monthly estimates of hydrological water balance components (including precipitation and temperature) at the higher resolution of 1 km from 1951. While it may be extremely useful for water resource management studies, it does not meet the required fine temporal resolution of our analysis.

The E-OBS (European Climate Assessment & Dataset) dataset (Cornes et al., 2018), developed by the ECA&D initiative, provides a comprehensive suite of climatic variables across Europe at 0.1° spatial resolution (approximately 9 km in the study region). These variables include daily mean, minimum, and maximum temperatures, daily precipitation depths, mean sea level pressure, mean wind speed, mean relative humidity, and global radiation. The E-OBS dataset employs a three-stage interpolation process (Cornes et al., 2018). First, monthly means of precipitation and temperature are interpolated using thin plate splines. Second, daily anomalies are interpolated using different kriging techniques: indicator kriging for precipitation and elevation-based for temperature. Finally, these high-resolution grids are aggregated to produce various E-OBS grid resolutions.

Reanalysis products assimilate observational data with numerical weather prediction models to produce a complete and consistent historical record of atmospheric conditions. ERA5-Land (Muñoz-Sabater et al., 2021), a global reanalysis dataset starting in 1950, is produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). This dataset provides hourly data at a 0.1° (approximately 9 km) resolution, generated through the interpolation of ERA5 reanalysis data (that have 0.25° resolution) using land surface modelling techniques. Therefore, while this produces higher resolution output, ERA5-Land should not be considered a true reanalysis at 9 km. ERA5-Land encompasses estimates of temperature, precipitation, and a multitude of other land surface variables. Given its global coverage and widespread application, ERA5-Land has been subject to extensive evaluation in various studies worldwide, revealing contrasting performances in precipitation and temperature estimation across different regions, as will be discussed below.

Table 2
Main characteristics of the gridded meteorological products used in this study. Variables include P: Precipitation, Tmin, Tmax and Tmean: minim, maximum and average temperature, respectively.

Product Name	**Starting Year**	**Temporal resolution**	**Spatial coverage**	Spatial resolution	Methodology/Source	Variables
SCIA (Desiato et al., 2011, 2007)	1961 (P) 1981 (T)	Daily	Italy	10 km (P) 5 km (T)	Gauge-based	P, Tmin, Tmax
E-OBS (v28.0) (Cornes et al., 2018)	1950	Daily	Europe	9 km	Gauge-based	P, Tmin, Tmax, Tmean
ERA5-Land (Muñoz-Sabater et al., 2021)	1950	Hourly	Global	9 km	Reanalysis Dataset	P, T
MERIDA- HRES (Bonanno et al., 2019)	1986	Hourly	Italy	4 km	Reanalysis Dataset	P, T
CHIRPS (v2.0) (Funk et al., 2014)	1981	Daily	Quasi-Global (50S°–50°N)	5 km	Satellite adjusted with in-situ station data	P

MERIDA (MEteorological Reanalysis Italian DAtaset), developed by Bonanno et al. (2019), is an Italian reanalysis dataset that offers meteorological estimates for the entire country starting from 1986 at hourly temporal resolution. MERIDA variables include temperature, precipitation, wind, and radiation. This dataset integrates diverse meteorological observations into a numerical weather prediction model, providing a detailed representation of atmospheric conditions, crucial for hydrological applications. The original dataset is the product of a dynamic downscaling of the global ERA5 reanalysis by means of the Weather Research and Forecasting (WRF) model from 31 to 7 km resolution. This model is specifically configured to capture the typical weather conditions of Italy at a finer spatial resolution. Bonanno et al. (2019) demonstrated that this regional focus enables MERIDA to more accurately capture local weather patterns compared to ERA5. For this study, we utilized MERIDA-HRES, a higher-resolution version of MERIDA obtained through a dynamical downscaling of ERA5 at 4 km resolution, which better suites the purpose of this study, given the reduced size of some of the study catchments. However, for the sake of brevity we will refer to the GMP with the acronym MERIDA.

Gauge-corrected satellite rainfall products leverage the extensive spatial coverage of satellite observations while incorporating ground-based measurements from rain gauges for calibration and bias correction. CHIRPS (Climate Hazards group InfraRed Precipitation with Station data) exemplifies this approach, offering a quasi-global dataset (1981–2024) covering latitudes from 50°S to 50°N (Funk et al., 2014). CHIRPS integrates satellite imagery with in-situ station data to provide daily and pentadal precipitation estimates at a 5 km resolution. The dataset's long record and fine spatial resolution have made it particularly valuable for drought monitoring and climate studies, especially in regions with limited ground-based observations (e.g., Lemma et al., 2022; Pandey et al., 2021; Rivera et al., 2019). While validation studies across various regions revealed contrasting performance, CHIRPS has been widely applied in climatic contexts similar to the Mediterranean region (e.g. Aksu et al., 2022). For these reasons and given the purpose of the present study, we consider CHIRPS the most suitable dataset among bias-corrected satellite precipitation products at daily scale, with a sufficiently fine spatial resolution and 40-years long observation period.

Overall, the present study deliberately focuses on a select group of widely used products. The selection was carried out by considering European and national products of different nature with the longest daily or hourly records. However, it is important to acknowledge other notable datasets that, while not included in this study, still contribute significantly to the field. Among others, EMO (European Meteorological Observations; Gomes et al., 2020), a Copernicus Emergency Management Service product derived from the interpolation of ground meteorological observations, provides sub-daily precipitation and temperatures from 1990 across the entire continent. Also, the CERRA (Copernicus European Regional ReAnalysis; Ridal et al., 2024), developed by the Copernicus Climate Change Service, is a European reanalysis dataset that assimilates both observational data and lateral boundary conditions sourced from ERA5, providing sub-daily atmospheric and meteorological estimates at 5.5 km spatial coverage from 1984 to near real-time. In the Italian context, the VHR-REA_IT dataset (Raffa et al., 2021), curated by CMCC (Euro-Mediterranean Center on Climate Change), offers hourly data at 2.2 km resolution for the entire Italian territory from 1989 to 2023, achieved through dynamical downscaling of ERA5 reanalysis using the COSMO Regional Climate Model (Rockel et al., 2008). Such datasets were not included in the present analysis due to their shorter temporal coverage, compared respectively to E-OBS, ERA5-Land and MERIDA.

3.1. State of the art on the performance and limitation of the selected GMPs in Italy

Some studies have already compared the performance and applicability of some of the selected GMPs in different contexts, in Italy.

Among nation-wide applications, Padulano et al. (2020) compared the spatial pattern of annual and monthly rainfall derived from SCIA, E-OBS, ERA5, and ERA5-Land products in order to assess the historical rainfall regime across Italy. The study indicates that ERA5 and ERA5-Land tend to overestimate precipitation over the Alps and Southern Italy compared to SCIA, while E-OBS underestimates it. In addition, Padulano et al. (2021) evaluated the accuracy of SCIA, E-OBS, ERA5, ERA5-Land datasets for assessing rainfall erosivity across Italy: in particular, they tested the reliability of their monthly precipitation estimates when used as input to a number of empirical models. The results revealed significant differences among the products, with ERA5 and ERA5-Land generally overestimating rainfall erosivity, while E-OBS tended to underestimate it, particularly in mountainous regions. Overall, the SCIA was found to provide the most accurate results. Rianna et al. (2023) analysed the reliability of E-OBS and ERA5-Land temperature estimates compared to the reference dataset SCIA, for updating Italian thermal load maps. E-OBS aligned better with SCIA than ERA5-Land for both maximum and minimum temperatures. The authors concluded that both E-OBS and ERA5-Land temperature datasets are suitable for standardizing climatic actions in European structural design standards, with E-OBS showing closer agreement to local data. As far as reanalysis products are concerned, Cavalleri et al. (2024a) conducted a comprehensive validation of eight reanalysis GMPs, including two versions of MERIDA (at 7 and 4 km resolution), focusing on their ability to reproduce precipitation fields across different spatial scales in Italy. Their analysis revealed similar spatial patterns of bias for both MERIDA versions, with wet biases primarily in the Po Valley and Central Alps, though MERIDA-HRES showed improvements in reducing dry biases across southern regions.

Focusing on specific regions within Italy, Turco et al. (2013a) assessed three gauge-based products, including E-OBS, over the Great Alpine Region. The results show good agreement between two high-resolution regional datasets, but highlight the limitations of EOBS, particularly in reproducing spatial patterns of extreme precipitation indices in the northwestern Italian Alps. In a separate study, Turco et al. (2013b) compared these products over the Po basin, finding good temporal agreement between datasets but cautioning against the use of E-OBS for assessing climate extremes. In Eastern Sicily, Longo-Minnolo et al. (2022) evaluated the performance of ERA5-Land reanalysis and various spatial interpolation methods in estimating precipitation in Eastern Sicily. They found that ERA5-Land generally underestimated precipitation compared to interpolated estimates, but performance improved significantly when local observations were used for bias correction, highlighting the value of combining ERA5-Land with local data for enhanced accuracy. Cammalleri et al. (2024) conducted a comprehensive comparison of E-OBS and ERA5 datasets against ground observations in southern Italy, focusing on their ability to capture rainfall trends and inter-annual variability. Their findings reveal that while both datasets generally capture major trends observed in ground-based data, ERA5 tends to produce flatter results compared to observations, whereas E-OBS shows larger range of variability in capturing trends and interannual variability.

Satellite-based precipitation datasets have also been subject to validation studies on specific regions in Italy. Duan et al. (2016) examined eight such datasets, including CHIRPS, in the Adige River basin (South-eastern Alps), assessing their performance against ground-based observations across multiple temporal and spatial scales. Their study highlighted significant seasonal variations in estimation errors. Similarly, Caroletti et al. (2019) evaluated CHIRPS, E-OBS, and climate model outputs in reproducing monthly precipitation over Calabria, Southern Italy, finding that CHIRPS performances were among the best ones within non-observational datasets.

At the European scale, Bandhauer et al. (2022) evaluated E-OBS and ERA5 datasets by comparing them to high-resolution regional datasets in three subregions: the Alps, the Carpathians, and Fennoscandia. Their findings indicated that both datasets reproduced mesoscale precipitation patterns well, but E-OBS underestimated high quantiles in data-sparse mountain regions (in the Alps and the Carpathians), highlighting the importance of station density on its accuracy. Conversely, they also showed how ERA5 overall overestimated mean precipitation in all the subregions.

Indirect validation studies in Italy have also explored the impact of GMP choices on hydrological model performance. Laiti et al. (2018) examined five GMPs, including E-OBS, using a physically-based hydrological model in the Adige catchment. They confirmed the crucial role of gauge network density for gauge-based products in GMP accuracy for two regional products, while satellite products and E-OBS showed lower performance, especially in smaller catchments. Dalla Torre et al. (2024) tested the suitability of the ERA5-Land reanalysis product for streamflow simulation with the semi-distributed ICHYMOD model in the same region, revealing an overestimation of both streamflow and snow water equivalent. Tuo et al. (2016) assessed the impact of four different precipitation datasets, including CHIRPS, on SWAT model simulations in Adige River basin, finding that precipitation input significantly affects model performance, with interpolated data yielding the best results. CHIRPS satellite data also performed well, demonstrating its potential for use in data-scarce regions.

Hagemann and Stacke (2023) compared ERA5 and E-OBS datasets for generating high-resolution river discharge data across Europe, focusing on areas of 1500 km^2 or larger, including several Italian catchments. Both datasets indicated increasing maximum and mean discharges in southern Italy, with E-OBS exhibiting more pronounced trends. However, the study found that E-OBS experiments showed lower performance due to biases from data gaps in the recent years, underscoring the crucial role of gauge density in determining E-OBS accuracy.

The regional performances of these datasets underscore the importance of careful selection and validation when adopting them for practical applications in hydrology, highlighting the need for a more comprehensive assessment of the wide array of currently available GMPs, especially when used with the purpose of rainfall-runoff modelling.

3.2. Rain gauge network density and distribution

The accuracy and reliability of gauge-based SCIA and E-OBS products are strongly linked to the density and spatial arrangement of the underlying meteorological station network. An uneven distribution or sparse network is in fact acknowledged to lead to interpolation errors and inaccuracies in the resulting GMPs (Girons Lopez et al., 2015). A comprehensive assessment of the gauge networks underpinning the SCIA and E-OBS datasets is therefore essential for understanding their potential reliability and limitations in representing the actual meteorological conditions. Here, we focus exclusively on the density of the rain gauges used by the two gauge-based products, since the impact of the density of temperature gauges is generally less significant compared to rain gauges, due to the lower spatial variability of temperature.

While the E-OBS website includes the list of all the stations coordinates with the actual working periods for each one, ISPRA has provided on our request, for the purpose of the present study, the average monthly regional station density (based on actual working

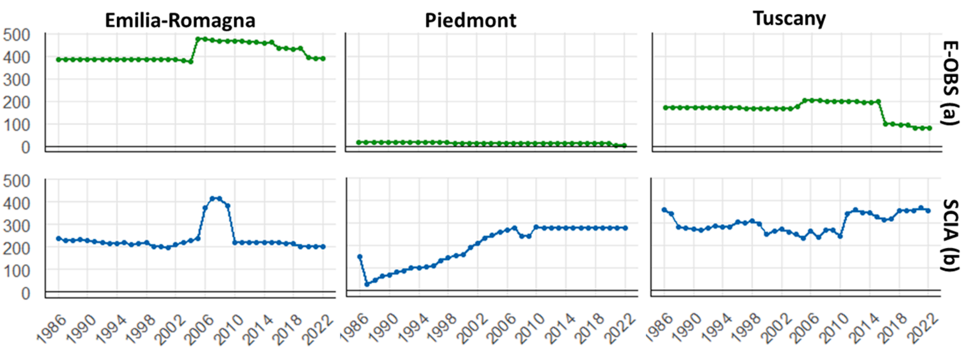

Fig. 2. Temporal evolution of number of rain gauge stations used by E-OBS (panel a) and SCIA (panel b) in the study region (1986–2021).

stations) between 1986 and 2020. The spatial distribution of E-OBS stations in the study regions between 1986 and 2021 is shown in Supplementary material (Figure S1). During that time window, average station densities in Emilia-Romagna for E-OBS and SCIA are both high: 17.6 and 10.5 stations per 1000 km^2 respectively. On the other hand, in Piedmont E-OBS has a very low density of only 0.5 stations per 1000 km^2, while SCIA maintains a density of 7.9 stations per 1000 km^2. Tuscany exhibits densities of 7.0 and 13.2 stations per 1000 km^2 for E-OBS and SCIA, respectively.

Fig. 2 provides a visual representation of the evolution of the gauge networks utilized by SCIA and E-OBS over time. It is evident that the SCIA network exhibits a much higher station counts compared to E-OBS in Piedmont and Tuscany. Although the number of SCIA stations in Piedmont was lower in the first two decades, it was still much higher than that of E-OBS, indicating potentially greater reliability in those areas.

With the exception of Piedmont during 80 s and 90 s, the density of the SCIA gauges over the three regions is relatively uniform, ranging between 10 and 15 station per 1000 km^2, whereas the number of stations for which rainfall data are conferred to the ECA&D database is very different, and in particular extremely high for Emilia-Romagna and very low for Piedmont. In fact, even if meteorological data are collected (and publicly accessible as single stations data) for a very dense network of sensors in all the three regions, E-OBS uses much less stations than SCIA in Tuscany and especially in Piedmont. In Emilia-Romagna, EOBS uses all the stations from the regional network, whereas SCIA estimates are based on a smaller number of constant 'reference stations' (around 200) - a part from a few years, 2006–2009, where the data from all the Emilia-Romagna stations (more than 400) were used in SCIA gridded product.

In data-rich regions, observational GMPs based on high-density gauge networks are generally considered the more reliable estimates and used as reference for assessing reanalysis or satellite-based products (see, among the many others, Gomis-Cebolla et al., 2023). Considering the two datasets based on ground observations, the denser and more homogeneous distribution of the SCIA gauge network suggests a greater capacity for accurately capturing local meteorological variability compared to E-OBS, as confirmed also by the results of the indirect validation presented later on. In fact, SCIA has been already adopted as reference meteorological product in various studies. For instance, Manco et al. (2023) employed SCIA as observational dataset to evaluate the performance of three high-resolution atmospheric models (WRF, COSMO, and ICON) in reproducing temperature and precipitation fields over the Italian Peninsula. Similarly, De Lucia et al. (2022) conducted a sensitivity study of the ICON-LAM model over Italy, using SCIA as the primary reference dataset to evaluate model performance for temperature and precipitation. Therefore, despite its relatively coarse resolution, SCIA is here considered as the reference dataset.

4. Methods

4.1. Direct comparison of GMPs against SCIA reference dataset

The initial phase of the study involves a direct validation of the performance of four of the GMPs –E-OBS, MERIDA, ERA5-Land, and CHIRPS – in representing precipitation and temperature patterns across the study area, assuming SCIA as the reference meteorological product. For this purpose, areal average values of daily precipitation (MAP) and temperature (MAT) for each GMP are considered. Henceforth, all references to precipitation and temperature values will refer to such areal-averaged values.

The analysis began with examining the long-term mean annual precipitation (annual MAP) and mean annual temperature (annual MAT) estimated by each GMP. We focused on the agreement between each GMP and the reference SCIA dataset, paying particular attention to the spatial patterns of discrepancies across the study area. SCIA, characterized by its highly dense and uniform gauge network, served as the benchmark reference dataset.

In order to assess the ability of each GMP to capture the intra-annual variability of the meteorological variables of interest, we also aim to identify biases in the representation of seasonal patterns. Therefore, long-term mean monthly precipitation and temperature values of the different GMPs are compared at regional scale.

Moreover, the accuracy of daily precipitation estimates is also assessed by means of the well-known Kling-Gupta Efficiency (KGE, Eq. (1), Gupta et al., 2009) between each of the GMPs and the reference SCIA dataset.

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

KGE measures the agreement between reference daily precipitation with the evaluated ones, decomposing the total performance into three components: correlation coefficient (r), mean bias (β), and deviation bias (γ). r is the Pearson product–moment correlation coefficient; β measures the ratio between the mean of the evaluated values and the mean of the reference ones, indicating the average tendency of the precipitation to underestimate ($\beta < 1$) or overestimate ($\beta > 1$); γ measures the ratio between the standard deviation of the evaluated values and the standard deviation of the reference ones, the relative dispersion between the gridded and the ground-based measurements. Even if Kling-Gupta Efficiency was initially conceived to evaluate the accuracy of rainfall-runoff model simulations, it has been also widely used to evaluate the performance of precipitation products both at daily scale (e.g. Beck et al., 2019; Hafizi and Sorman, 2022; Saemian et al., 2021) and at coarser time scales (e.g. Centella-Artola et al., 2020; Khan et al., 2023; Valencia et al., 2023). In this study, we will refer to KGE scores for precipitation with the acronym KGE-P. In order to identify different source of error, the error components of KGE-P are also analysed singularly.

4.2. Indirect validation through rainfall-runoff model

As already introduced, in order to assess the reliability of the different GMPs for hydrological impact studies, an indirect validation approach based on rainfall-runoff modelling is employed. The rationale behind an indirect validation consists in running a rainfall-runoff model on each of the study catchments by forcing it with meteorological inputs computed from each of the GMPs and evaluating model performances against observed streamflow time series. Then, we assume that rainfall-runoff model performances are representative of GPMs accuracy and reliability.

The rainfall-runoff model used in this experiment is the CemaNeige-GR6J model (Coron et al., 2017b). It is the combination of the CemaNeige snow accounting routine (Valéry et al., 2014) with the GR6J model (Pushpalatha et al., 2011), a daily lumped continuously simulating rainfall-runoff model, developed at INRAE (Antony, France) by the Équipe Hydrologie des Bassins versants. The software is freely available in the “airGR” R package (Coron et al., 2017b, 2017a).

The model is forced with spatially averaged daily air temperature, precipitation, and potential evapotranspiration (PET) for each catchment. Here, PET was estimated for each GMP using the simplified Blaney-Criddle method (Blaney and Criddle, 1962), which relies solely on average daily temperature data and daylight duration. The PET data included in ERA5-Land were not used due to concerns regarding unrealistically high values if employed for watershed water balance, as highlighted by few previous studies (e.g. Clerc-Schwarzenbach et al., 2024). For the evaluation of CHIRPS, which provides precipitation data exclusively, temperature and PET estimates from SCIA are used.

The CemaNeige snow accounting routine is based on a degree-day concept, where the thermal inertia of the snowpack is also taken into account. It is ruled by two parameters: the snowmelt factor (θ_{G1}) and the cold-content factor (θ_{G2}). Although the module uses daily spatially-lumped inputs, in order to better simulate snow accumulation and melting, it allows for dividing the catchment into elevation zones of equal area by means of the hypsometric curve. While the module functions are applied with a lumped set of calibrated parameters, internal states are instead allowed to vary over each elevation layer. The total liquid output at the catchment scale is the average of the outputs from each elevation zone. Here, we decided to use, as a default, five elevation layers. Detailed descriptions of the CemaNeige routines can be found in Valéry et al. (2014).

The GR6J model receives the total liquid output from CemaNeige and potential evapotranspiration as inputs. It is ruled by six parameters and comprises a soil moisture reservoir and a conceptual groundwater exchange function. The routing procedure includes two flow components routed by two-unit hydrographs, a non-linear store, and an exponential store. A detailed description of the GR6J model is available in Pushpalatha et al. (2011).

All eight parameters of the combined model (two for CemaNeige, six for GR6J) are calibrated. Lower and upper bounds of the parameter space are kept as the default (note that the parameters are normalised in the calibration procedure). Table 3 reports brief parameters description and boundaries. For simplicity, we will refer to the CemaNeige-GR6J model just with the acronym GR6J, even if it will always include the CemaNeige snow module.

For each catchment, the GR6J model is calibrated against daily streamflow observations for each of the five GMPs, i.e. using daily precipitation, temperature, and potential evapotranspiration time series from each different meteorological product to force the model. Such approach, widely adopted in the literature (e.g. Andréassian et al., 2004; Beck et al., 2017; Oudin et al., 2006; Tarek et al., 2020), allows to compare the relative differences in efficiency between various forcing datasets and their corresponding calibrated model outputs serves as a valuable indicator of temperature and precipitation representation accuracy. In fact, while finding universally applicable parameter sets across all forcing datasets remains a challenge, by calibrating the GR6J model for each GMP we acknowledge explicitly that the calibrated watershed parameters are dependent on climatic input data and we aim to enhance the accuracy and reliability of streamflow simulations, thereby facilitating a more robust evaluation of temperature and precipitation data representation.

The model is calibrated automatically with the Dynamically Dimensioned Search (DDS) algorithm (Tolson and Shoemaker, 2007), using as objective function the Kling-Gupta Efficiency (KGE) between simulated and observed daily streamflow time series.

Usually, a standard model parameterisation would involve the use of a split-sample procedure, in which the available streamflow observation record is separated into two different sub-periods used respectively to train (i.e. calibrate parameters) and test (i.e. validate) the model. Here instead, given the limited length of the observed streamflow time series for some of the study catchments (see Table 1), all the available observation records were entirely used for both model calibration and validation. Even if the risk of overfitting cannot be excluded, such an approach is commonly used to indirectly validate meteorological products for large samples of catchments (e.g., Beck et al., 2017; Satgé et al., 2021; Tarek et al., 2020), since it allows to use the entire information content of the available meteorological and hydrological time series and to take into account a greater variety of hydrological phenomena when testing the reliability of the different GMPs.

Once the model is calibrated with each of the GMPs, resulting performances in simulating the daily streamflow time series are

Table 3
Cemaneige-GR6J model parameters and their transformed real value ranges.

Parameter	Units	Range	Description
θ_{G1}	mm/(°C*day)	0–109	Snowmelt (degree-day) factor
θ_{G2}	-	0–1	Cold content factor
X1	mm	0–21807	Non-linear production storage capacity
X2	mm/day	−1903–1903	Groundwater exchange coefficient
X3	mm	0–21807	Non-linear routing store capacity
X4	days	0–22	Time parameter for unit hydrographs routing
X5	-	−1.998–1.998	Threshold parameter for water exchange with groundwater
X6	mm	0–21807	Exponential routing store capacity

analysed in terms of KGE metric and its decomposed components. The KGE scores for streamflow simulations is here named as KGE-Q.

5. Results and discussion

5.1. Direct comparison of the precipitation and temperature

The initial phase of our analysis involved a comparative assessment of catchment-averaged precipitation and temperature values derived from various GMPs. We specifically focused on the mean areal temperature (MAT) and mean areal precipitation (MAP) estimates produced by each GMP and compared them against the SCIA reference dataset. The temporal window selected for this comparison spans from 1986 to 2021, ensuring a consistent period for all datasets.

The panels in the left column of Figs. 3 and 4 illustrate the spatial distributions of long-term annual MAT and MAP respectively across the study catchments. To have a first order estimate of the accuracy of the GMPs, we use the differences between the GMP-derived annual MAT and MAP values and those from the SCIA dataset, which are reported in the middle column panels of Figs. 3 and 4. This analysis highlights the magnitude of deviations and identifies specific regions where the GMPs exhibit significant discrepancies from the reference data. The right column panels of Figs. 3 and 4 provide scatter plots that illustrate the agreement between the GMPs' annual MAT and MAP estimates and the SCIA reference values (each dot refers to a single catchment).

Fig. 3 reveals a broad agreement between datasets in capturing the general spatial patterns of annual MAT, with increasing temperatures from north to south and from high to lower altitudes. However, subtle deviations are apparent in specific regions. Notably, the E-OBS dataset (panel b) aligns closely with the reference SCIA dataset (panel a), displaying minimal discrepancies across catchments. This close agreement is further supported by the low mean absolute differences (panel e) and the tight clustering of data points around the 1:1 line in the scatterplot (panel h), underscoring the high accuracy of E-OBS in estimating annual MAT. In contrast,

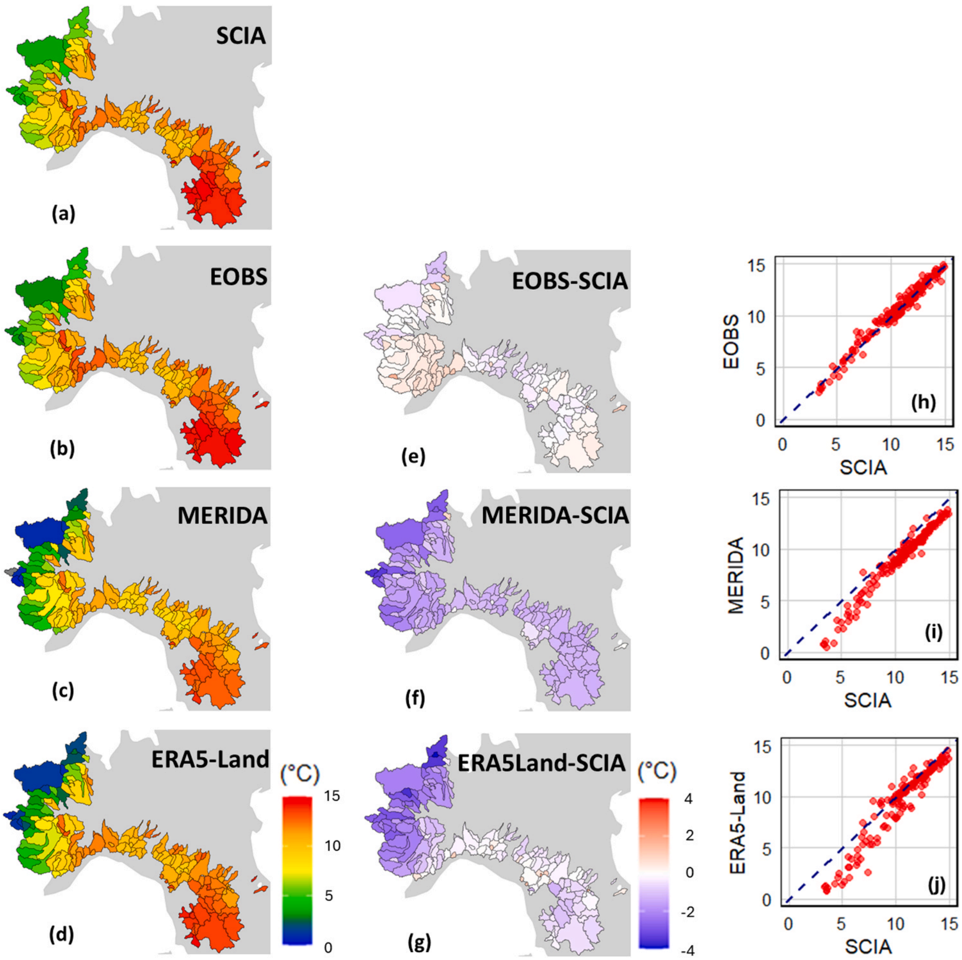

Fig. 3. Panels a-d: annual MAT estimated for all datasets: (a) SCIA, (b) E-OBS, (c) MERIDA, and (d) ERA5-Land. Panels e-g: differences in annual MAT between the reference SCIA dataset and E-OBS (e), MERIDA (f), and ERA5-Land (g), respectively. Panels h-j: scatter plots between SCIA and single GMPs annual MAT.

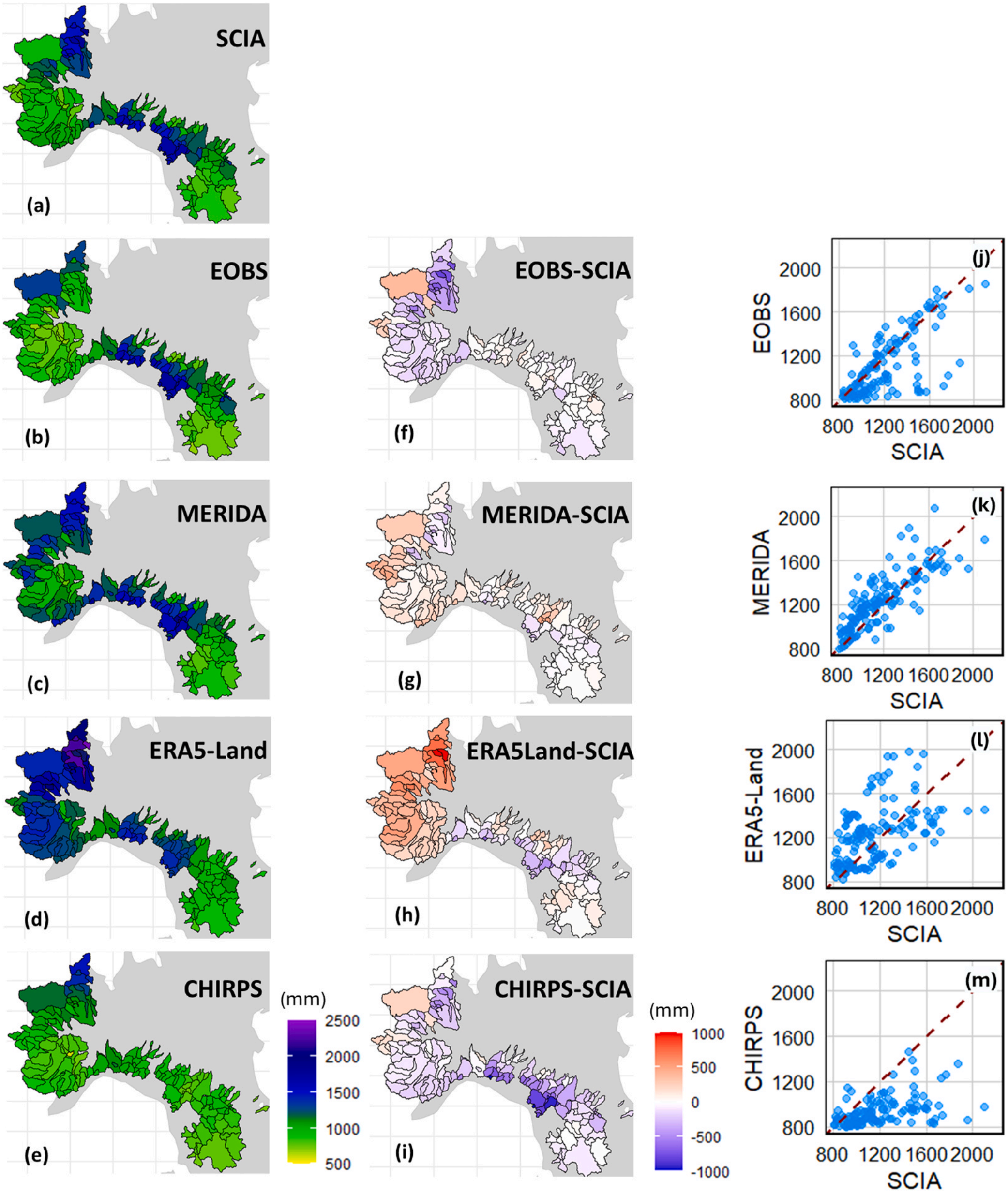

Fig. 4. Panels a-e: annual MAP estimated for all datasets: (a) SCIA, (b) E-OBS, (c) MERIDA, (d) ERA5-Land and (e) CHIRPS. Panels f-i: differences in annual MAP between the reference SCIA dataset and E-OBS (f), MERIDA (g), ERA5-Land (h) and (i) respectively. Panels j-m: scatter plots between SCIA and single GMPs annual MAP.

the MERIDA dataset (panel c) underestimates temperatures across most catchments (panel i), with more pronounced deviations observed in the northern Piedmont region (panel f). This suggests a potential systematic bias in MERIDA temperature estimation, particularly in areas with complex orography. The ERA5-Land dataset (panel d) exhibits the worst underestimation of annual MAT in Piedmont, while in the other two regions the values overall agree with SCIA estimates (panel g). The temperature bias of reanalyses datasets over the Italian mountainous areas aligns with the findings of Cavalleri et al. (2024b), who noted the difficulties of models in reproducing physical processes at high elevations, even with the help of high-resolution models.

The analysis of annual MAP in Fig. 4 reveals more pronounced discrepancies compared to annual MAT. This is expected due to the inherent spatial and temporal variability of precipitation, which makes it more challenging to capture accurately. The E-OBS dataset (panel b) demonstrates a good agreement with the SCIA dataset in Emilia-Romagna and Tuscany, while notable deviations with underestimation are observed in Piedmont (see also panels f and j). This aligns with the findings of Turco et al. (2013a), who compared E-OBS with other gridded datasets across the Great Alpine Regions, concluding that E-OBS does not reproduce reliably the climatology over the study area (Piedmont included) and that its use in these regions should be done with caution. The MERIDA dataset (panels c, g and k) displays the better agreement with SCIA between the non-observational products, especially in Tuscany. In fact, ERA5-Land (panel d) display larger differences, in particular in mountainous areas of Piedmont where it significantly overestimates precipitation (panels h and l). Such comparison is consistent with the findings of Capecchi et al. (2023) who studied ERA5-Land's accuracy across Alps and Apennines. The CHIRPS dataset (panel e) also shows significant deviations, overall underestimating annual precipitation, especially in Emilia-Romagna.

Once GMP long-term bias has been assessed, it is interesting to analyse differences in the seasonality of the meteorological variables. While temperature analysis is here omitted due to the general agreement in the intra-annual pattern of MAT between the products, Fig. 5 reports a comparison of long-term mean monthly precipitation estimates from the various GMPs across Piedmont (a), Emilia-Romagna (b), and Tuscany (c), during the period 1986–2021 at regional scale, i.e. spatially averaging precipitation values over the entire catchment domain within each region.

E-OBS (blue line) demonstrates strong alignment with SCIA (yellow) in Emilia-Romagna (panel 5b) and Tuscany (panel 5c), overall accurately capturing the seasonality of precipitation with peaks in Spring and Autumn. In Piedmont (panel 5a), there is a noticeable underestimation, especially in the wetter Spring and Autumn seasons, i.e. smoothing intra-annual variability of precipitation.

Considering the reanalysis products, MERIDA shows overall better agreement with the observational products, consistently with the findings of Bonanno et al. (2019). In Piedmont, the excess rainfall of both reanalysis products is concentrated in the hotter months, even if with different magnitude, with a very strong overestimation by ERA5-Land for all the months from March to November. In Emilia-Romagna, both reanalysis products overestimate again the summer regional MAP, but ERA5-Land exhibits also a significant underestimation of monthly precipitation in Autumn and Winter, differently from MERIDA, which matches well with SCIA in such period of the year. In general, the lower performance of ERA5-Land during summer and autumn months aligns with the outcomes of Bandhauer et al. (2022) who showed that ERA5, from which ERA5-Land is derived, overestimates precipitation systematically in North-eastern Italian Alps, due to overestimation of wet days, with a stronger discrepancy in high mountain catchments in the convective summer period. Furthermore, summer overestimation trend of ERA5-derived products was also observed in other regions of the world (e.g. Gomis-Cebolla et al., 2023; Jiao et al., 2021; Longo-Minnolo et al., 2022).

Finally, the analysis highlights a significant underestimation of precipitation observed for CHIRPS during all the months apart the summer, dry ones. This can be attributed to CHIRPS' reliance on a fixed infrared precipitation threshold for cloud-top temperature. Such threshold may be too conservative for detecting precipitation from warm-top stratiform clouds and orographic processes, which are common in these regions, resulting in underrepresentation of rainfall events (Paredes-Trejo et al., 2023). The concentration of CHIRPS' underestimation in wetter months suggests it is systematically missing a significant portion of precipitation events associated with warm-top clouds and orographic processes, which are common during rainy seasons in such Italian regions. This observation contrasts with Rivera et al. (2018), who concluded that CHIRPS data more accurately capture seasonal precipitation totals during the rainy season compared to the dry seasons in the Central Andean Area of Argentina. However, their study focused on a very different geographical and climatic area.

To comprehensively evaluate the performance of various GMPs in capturing precipitation patterns, we employed the Kling-Gupta Efficiency (KGE-P) metric, along with its components: correlation coefficient, mean bias, and deviation bias.

Fig. 6 presents the KGE-P scores between each of the GMP and the reference SCIA dataset. In particular, the top panel illustrates the KGE-P scores (higher values for better performance) for each product and region, while the bottom panels display, in a similar fashion, the different error components: the correlation coefficients (panel b, higher values for better performance), reflecting the temporal accuracy of the precipitation estimates, the mean bias (panel c, i.e. the volume error) and the deviation bias (panel d), which represents the variability component of the error (with the two biases indicating better performance for values closer to 1).

E-OBS demonstrates the highest overall performance in Emilia-Romagna and Tuscany, as evidenced by high KGE-P values, exceeding 0.80. This superior performance is supported by consistently high correlation coefficients ($R > 0.80$) and mean and deviation bias values approaching the optimum unity value, indicating both accuracy and precision in the estimations. Confirming the previous results, E-OBS underperforms in Piedmont, likely due to the very sparse rain gauge network in this region. This observation aligns with Bandhauer et al. (2022), who found larger discrepancies in catchments with poor station density and complex topography. Notably, the mean bias and deviation bias statistics in Piedmont indicates an underestimation of precipitation daily amounts and lower

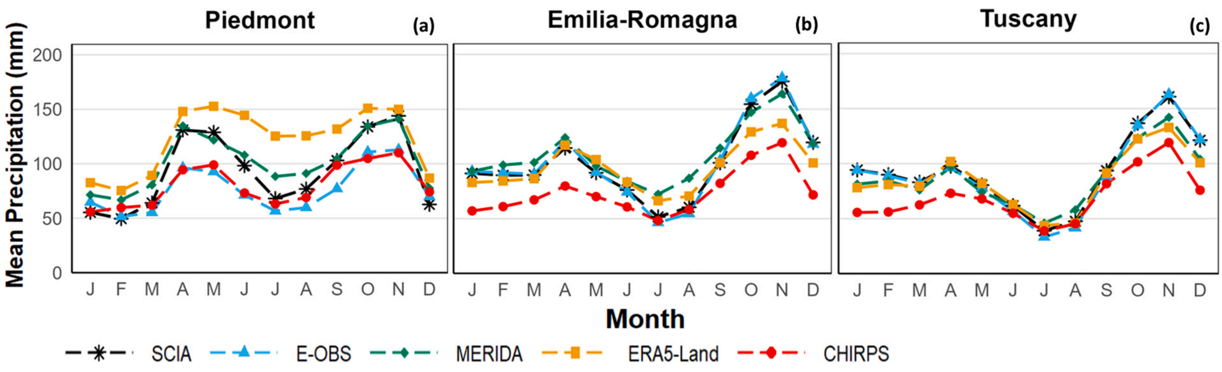

Fig. 5. Long-term mean monthly regional precipitation (mm) for each of the different GMPs.

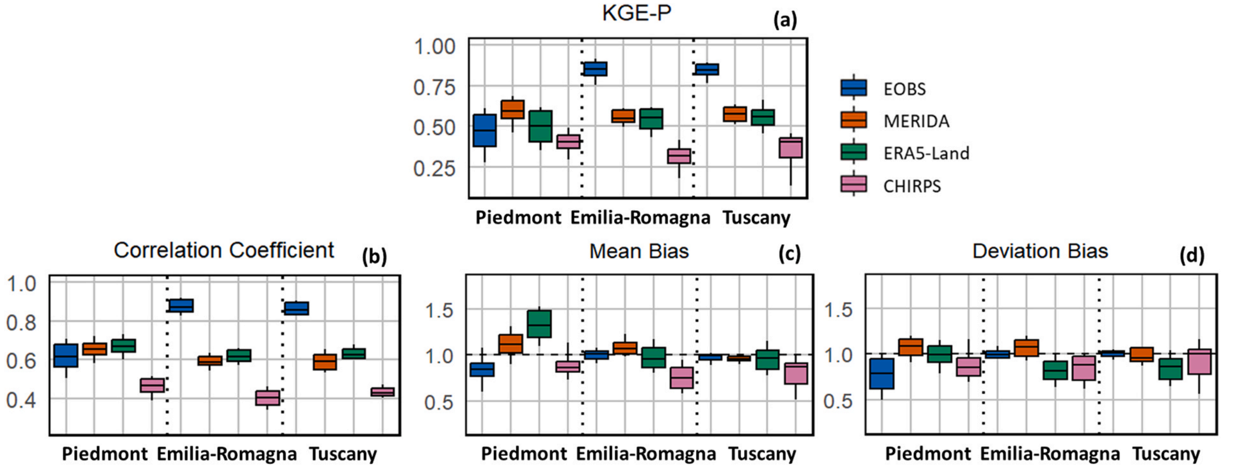

Fig. 6. Performances of monthly precipitation estimated against the SCIA reference estimates across three Italian regions. Boxplots show the distribution of four performance metrics: Kling-Gupta Efficiency, Correlation Coefficient, Mean Bias, and Deviation Bias. Vertical dotted lines separate the results by region. Boxes refer to 25 % and 75 % quantiles, whiskers refer to 10 % and 90 % quantiles, and black lines indicate the median.

variability in E-OBS estimates in respect to SCIA for this region. In general, E-OBS reliability is expected to be strongly related to the density and homogeneity of ground meteorological stations used in the interpolation process, which varies significantly across Europe (e.g., Hofstra et al., 2009, Kostopoulou et al., 2012; Krauskopf and Huth, 2020; Mavromatis and Voulanas, 2021).

MERIDA exhibits consistent performance across all regions, with median KGE-P scores around 0.55, signifying quite good agreement with the reference dataset, resulting to be the best alternative to SCIA in Piedmont. The correlation coefficients (R > 0.60) observed across all regions suggest that MERIDA captures the general temporal patterns of precipitation, a notable achievement considering the challenges of predicting daily precipitation in orographically complex areas like Piedmont. The mean bias exceeding 1 in Piedmont and Emilia-Romagna suggests a tendency toward overestimation of precipitation values. Volume errors (mean bias) reflect what already observed in Fig. 4, while the variability of precipitation aligns well with the reference estimates (i.e. deviation bias close to 1).

ERA5-Land performances are very similar to those of MERIDA across Emilia-Romagna and Tuscany; looking at the single error components, ERA5-Land shows slightly better agreement in timing (correlation) with SCIA but higher errors in terms of rainfall volume and variability. In Piedmont instead, ERA5-Land performs worse than MERIDA, mainly due to the volume component of the error. Such results are in line with those by Cavalleri et al. (2024a), who conducted a comprehensive assessment of high-resolution regional reanalyses over Italy, including MERIDA, highlighting both its strengths and limitations: their findings indicate that MERIDA performs well in capturing precipitation patterns over Italy, particularly in comparison to the global reanalysis reference ERA5. In general, the non-uniform performances of ERA5-Land in the different regions confirm the need to test its reliability locally, as already highlighted by previous studies all around the world (i.e., Xu et al. 2022 in China, De Andrade et al. 2022 in North-eastern Brazil).

CHIRPS provides the lowest Kling-Gupta efficiency (KGE-P) against SCIA compared to the other datasets. As already seen, this dataset notably underestimated precipitation (mean bias < 1), but also its temporal variability, compared to the reference SCIA (deviation bias < 1). These findings suggest that CHIRPS struggles to accurately capture both the timing and magnitude of precipitation events within the study region. Other direct validation studies conducted in different Mediterranean areas found better performance for CHIRPS, but they considered coarser time scales exclusively. For example, Caroletti et al. (2019) compared CHIRPS monthly precipitation to ground observations and found good correlation of monthly precipitation values. Similarly, Katsanos et al. (2016) reported good agreement of monthly rainfall derived from CHIRPS and reference in-situ datasets in Cyprus.

5.2. Indirect comparison through rainfall-runoff model

In order to test the hydrological coherency of the GMPs in northern Italy, we evaluated the GMPs indirectly through their impact on hydrological modelling performance. Specifically, we assessed the performance of the GR6J (always coupled with CemaNeige) rainfall-runoff model when driven by precipitation, temperature, and potential evapotranspiration data derived from each GMP over the entire period of available streamflow data. The performance of the streamflow simulations obtained when forcing the model with the different GMPs are assessed computing the Kling-Gupta Efficiency for daily streamflow (KGE-Q), in reference to the streamflow observations, for each study catchment.

Fig. 7 illustrates both the spatial distribution and the scatterplots of the streamflow Kling-Gupta Efficiency (KGE-Q) across the study catchments, comparing the performances of the streamflow simulations obtained with SCIA and with the other GMPs. The SCIA reference dataset generally yields the most accurate results, with the highest KGE-Q (darker colors in panel 7a), confirming its reliability as a benchmark. The performance of E-OBS, MERIDA, and ERA5-Land varies regionally, suggesting their potential utility under

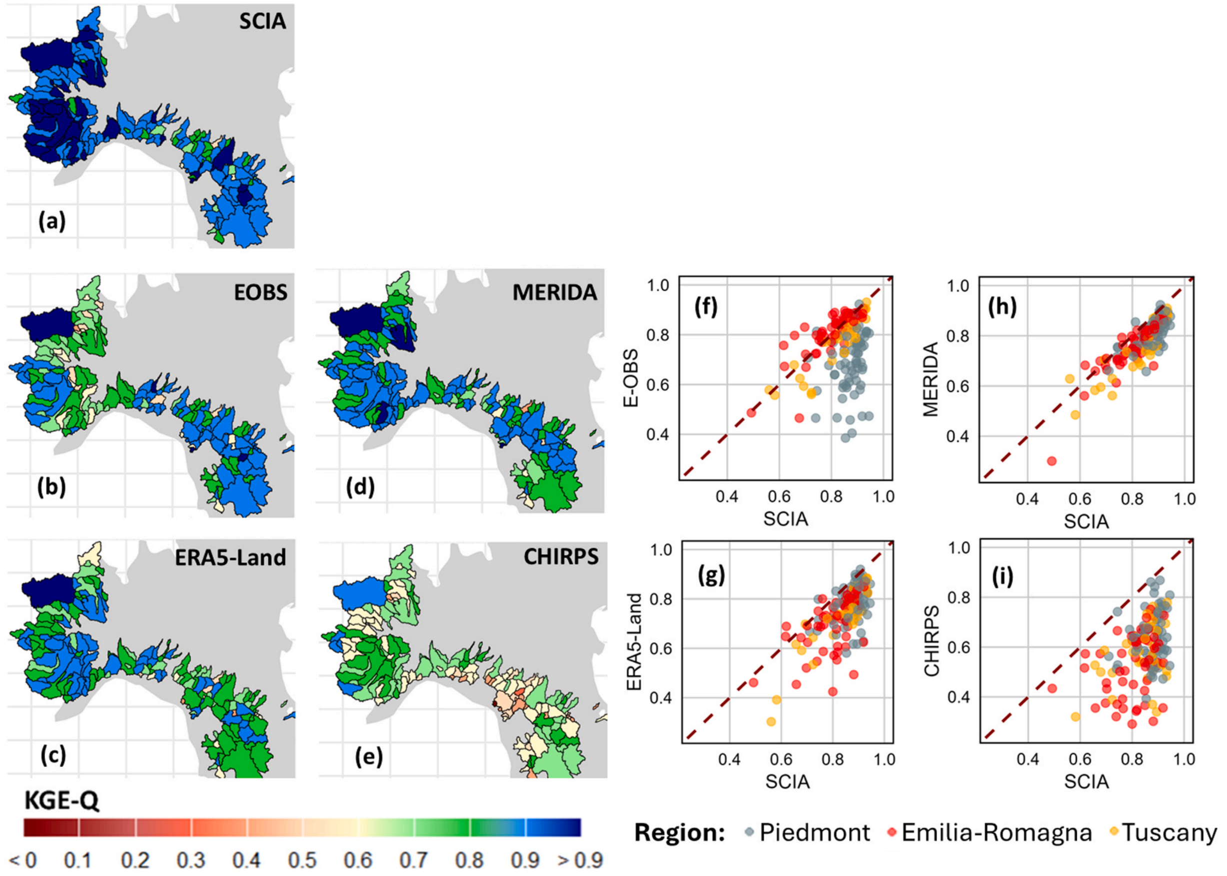

Fig. 7. Spatial distribution (panels a-e) and scatter plots (panels f-i) of the KGE-Q scores of the CemaNeige-GR6J rainfall-runoff model driven by the different GMPs.

specific conditions, while CHIRPS demonstrates limited performance over all the regions (see panels 7e and 7i).

Fig. 8 presents the variability of GR6J model performance in the three regions when driven by precipitation and temperature estimates from each GMP, in terms of KGE-Q scores (panel a) and its components: correlation coefficient (panel b), mean bias (panel c), and deviation bias (panel d).

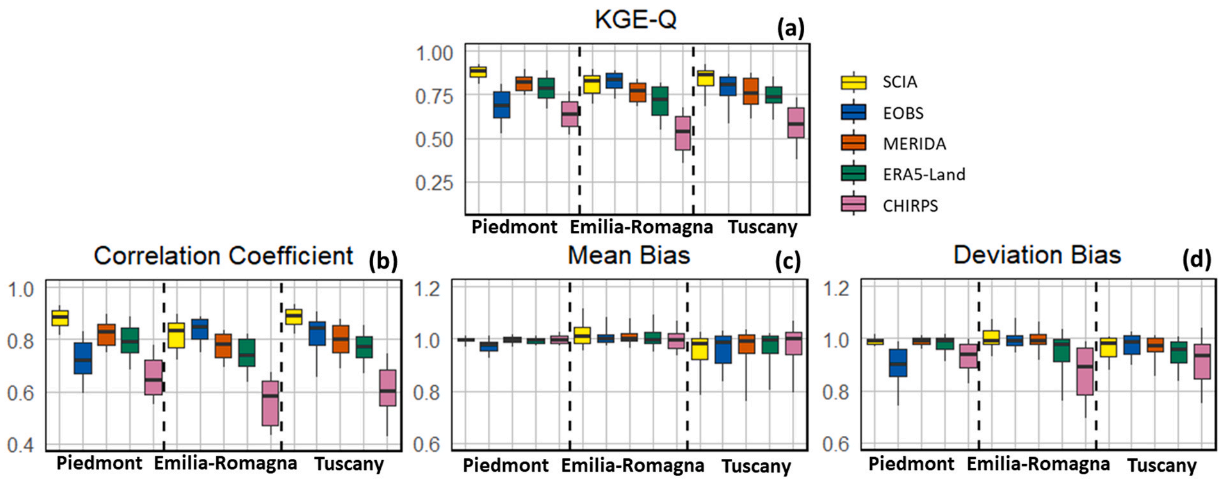

Fig. 8. Calibration performance of the CemaNeige-GR6J rainfall-runoff model using GMPs across three Italian regions. Boxplots display the distribution of four performance metrics: Kling-Gupta Efficiency, Correlation Coefficient, Mean Bias, and Deviation Bias. Vertical dotted lines separate results by region. Boxes represent 25 % and 75 % quantiles, whiskers extend to 10 % and 90 % quantiles, and black lines indicate the median.

Fig. 8 demonstrate, as expected, that the rainfall-runoff model parameterisation, repeated for each product, is able to compensate the volume error in the precipitation: the mean bias (panel 8c) is in fact very similar for all the GMPs and very close to 1, except for a few catchments in Tuscany, where all the GMPs lead to streamflow underestimation. Therefore, differences in KGE-Q are mainly attributed to the remaining two error components on timing and variability of streamflow.

SCIA consistently provides the most robust results, with median KGE-Q scores exceeding 0.75 across all regions, reaffirming its high quality and reliability in driving hydrological models.

E-OBS demonstrates performances comparable to SCIA in regions with dense rain gauge networks, i.e. in Emilia-Romagna and Tuscany. However, its performance deteriorates in Piedmont, where the rain gauge network is sparse. This aligns with the results of Hagemann and Stacke (2023), who observed a widespread low bias in E-OBS-simulated discharge, translated from the underestimated precipitation, particularly in data-sparse regions like Southeastern Europe. Similarly, Laiti et al. (2018) reported lower NSE and KGE values for models fed by E-OBS compared to other datasets in the Adige River Basin, particularly for catchments with sparse rain gauge density.

MERIDA exhibits overall good performance, with median KGE-Q scores above 0.75 across all regions, indicating its potential as a viable alternative to SCIA, particularly in data-scarce areas. However, its performance does not match that of observational products in data-rich regions, suggesting potential biases or limitations in its precipitation estimates.

ERA5-Land, despite exhibiting overall good performance with median KGE-Q scores above 0.68 across all regions, tends to perform worse than MERIDA in the entire study area. Comparing KGE-Q (Fig. 8a) to KGE-P (Fig. 6a) scores, we can expect such results in Piedmont, where MERIDA demonstrates to be significantly more reliable in the direct validation. On the other hand, KGE-P scores in the other two regions are very similar: a better KGE-Q score for MERIDA may be justified by the overall smaller volume errors (Fig. 6c) and better representation of rainfall variability in time (Fig. 6d). In addition, in Emilia-Romagna the better hydrological coherency of MERIDA may be also due to the better representation of rainfall seasonality (Fig. 5b): its seasonal precipitation amount is significantly closer to the reference in the autumn months, which are also the wetter along the year.

CHIRPS consistently underperforms in driving the rainfall-runoff model, as reflected by the lowest KGE-Q scores among the evaluated datasets. While the GR6J model partially offsets the mean bias error, the poor correlation in streamflow simulation ultimately resulted in very low KGE-Q scores. Notably, the very poor performance of CHIRPS in our study contrasts with the findings of Silva et al. (2022), who reported superior accuracy of CHIRPS over ERA5-Land in streamflow simulations for northeast Brazil. This discrepancy underscores the significant regional variability in the performance of satellite precipitation datasets. Our findings suggest that while CHIRPS may excel in certain hydroclimatic contexts, its global applicability without regional adjustments remains limited. As Gebrechorkos et al. (2024) emphasize, the effectiveness of CHIRPS is highly dependent on the specific region and its unique climatic and hydrological characteristics. Therefore, CHIRPS may not be the most suitable choice for hydrological modelling in regions like Northern Italy.

In-depth analysis of the error metrics from Figs. 6 and 8 shows a notable phenomenon known as error compensation in hydrological modelling. In general, the low KGE-P scores observed in all regions, attributed to the challenges of capturing spatial and temporal variability of precipitation (especially in areas with complex orography), correspond to higher values in the streamflow simulations (KGE-Q), indicating the model ability to compensate for meteorological input errors.

In order to better capture the role of single model parameters in compensating GMPs errors and biases, the differences of calibrated parameter values obtained when forcing the model with the five GMPs are also investigated. Detailed discussion of such results, which are not the main focus of the work, are reported in the Supplementary material for the sake of brevity, but it is worth mentioning that the analysis highlights how differences in CemaNeige parameter values regulating snow accumulation and melting are able to compensate MERIDA and ERA5-Land tendency to underestimate temperature, while model parameters controlling flow exchange with groundwater can partially compensate CHIRPS precipitation deficit.

Overall, the model is capable of compensating systematic under/overestimation error, while it is more sensitive to timing errors (as can be seen comparing panels b to d of Figs. 6 and 8). This model adaptability to volume and variability errors in precipitation, up to a certain threshold, aligns with previous findings (e.g. Jiang et al., 2024; Oudin et al., 2006; Sivasubramaniam et al., 2020). This suggests that even GMPs with known biases can still be valuable for hydrological modelling in data-scarce regions.

6. Conclusions

This study evaluated five Gridded Meteorological Products across Northern Italy through both direct comparison of rainfall and temperature spatial averages and indirect validation using a rainfall-runoff model and assessing the performances in streamflow simulation. The direct validation of GMPs against the SCIA reference gauge-based dataset revealed an underestimation of temperatures in the reanalysis products and varying performance across the study regions for the precipitation datasets: E-OBS showed high accuracy in regions with dense rain gauge networks (Emilia-Romagna and Tuscany), but underperformed in Piedmont where rain gauge density is low. MERIDA demonstrated consistent performance across all regions, while ERA5-Land showed variable accuracy, particularly overestimating precipitation in mountainous areas. CHIRPS underestimated precipitation across the entire study area.

The indirect validation through hydrological modelling corroborated the importance of considering local climate and catchment characteristics when selecting a suitable dataset for hydrological modelling, as already emphasized by a number of previous studies such as those by Gampe and Ludwig (2017) and Bitew and Gebremichael (2011).

SCIA consistently provided the streamflow simulations closest to the observations, reaffirming its reliability as a reference dataset. Its steady performance across diverse regions emphasizes the importance of dense rain gauge networks for accurate hydrological modelling, as highlighted by previous studies such as those by Beck et al. (2017) and Sun et al. (2018). E-OBS performed comparably to SCIA in data-rich regions but struggled in Piedmont, where its underlying rain gauge network is sparse and it does not allow it to capture the spatial and temporal variability of meteorological variables in such a orographically complex region. MERIDA and ERA5-Land showed fairly good overall performance in driving the hydrological model, with MERIDA outperforming ERA5-Land across all regions. CHIRPS consistently underperformed in hydrological simulations. Notably and as expected, the study highlighted the phenomenon of error compensation in hydrological modelling, where the rainfall-runoff model was able to partially mitigate biases in precipitation inputs. This suggests that even GMPs with known biases can still be valuable for hydrological applications in data-sparse regions.

Based on these findings, the results of the study suggest some key recommendations. Firstly, E-OBS should expand its network of stations for precipitation product development, particularly in areas with complex topography or sparse coverage, especially where meteorological data are publicly accessible for a dense network of sensors: with regards to the study region, the entire set of available ground measures is conferred to ECA&D only for Emilia-Romagna. Secondly, while reanalysis products such as ERA5-Land and MERIDA show potential as alternatives in data-scarce regions, ERA5-Land should be used cautiously, particularly in mountainous regions, and requires local validation before implementation. The good performance of MERIDA across all regions underscores the added value of its finer spatial resolution and of the use of local meteorological observations in the regional dynamic downscaling, highlighting its ability to better represent rainfall patterns across complex terrain, compared to ERA5-Land, and its potential as a reliable alternative to global reanalysis products in the Italian context. Finally, there is a pressing need for further research to enhance the applicability of CHIRPS in Northern Italy, as its performance in this study has been suboptimal.

Future research, currently not possible due to the lack of data, may replicate the indirect validation procedure on sub-daily time series, which would be particularly interesting for analysing the diurnal cycle of specific rainfall-runoff dynamics, as actual evapotranspiration and snow processes.

Overall, this study underscores the need for comprehensive evaluation of GMPs to enhance their reliability when applied for hydrological applications across diverse geographical contexts. When selecting GMPs for hydrological applications, it is beneficial to consider both direct and indirect validation results, as this dual approach can lead to more informed decisions.

CRediT authorship contribution statement

Gökhan Sarigil : Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Mattia Neri:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Data curation, Conceptualization. **Elena Toth:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the writing of this manuscript, the authors used ChatGPT in order to enhance readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank Giovanni Braca, Emanuela Piervitali and Walter Perconti (ISPRA) for their valuable support during the analysis and for providing additional SCIA metadata. This study was carried out within the RETURN Extended Partnership and received funding from the European Union Next-GenerationEU (National Recovery and Resilience Plan – NRRP, Mission 4, Component 2, Investment 1.3 – D.D. 1243 2/8/2022, PE0000005).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejrh.2024.102031.

Data availability

Streamflow data are publicly available at regional agencies websites: ARPA-Piemonte (Agenzia Regionale per la Protezione Ambientale del Piemonte, https://www.arpa.piemonte.it/rischi_naturali/snippets_arpa_graphs/), ARPAE (Agenzia Regionale Prevenzione Ambiente Energia dell'Emilia Romagna, https://simc.arpae.it/dext3r/) and SIR (Servizio Idrologico della Regione Toscana, https://www.sir.toscana.it/consistenza-rete). Original gridded data from the five GMPs are freely available as described in the corresponding references. Catchment boundaries and areal averaged meteorological forcings can be provided upon request to the corresponding author.

References

Aksu, H., Cavus, Y., Aksoy, H., Akgul, M.A., Turker, S., Eris, E., 2022. Spatiotemporal analysis of drought by CHIRPS precipitation estimates. Theor. Appl. Climatol. 148, 517–529. https://doi.org/10.1007/s00704-022-03960-6.

Alfieri, L., Avanzi, F., Delogu, F., Gabellani, S., Bruno, G., Campo, L., Libertino, A., Massari, C., Tarpanelli, A., Rains, D., Miralles, D.G., Quast, R., Vreugdenhil, M., Wu, H., Brocca, L., 2022. High-resolution satellite products improve hydrological modeling in northern Italy. Hydrol. Earth Syst. Sci. 26, 3921–3939. https://doi.org/10.5194/hess-26-3921-2022.

Andréassian, V., Perrin, C., Michel, C., 2004. Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models. J. Hydrol. 286, 19–35. https://doi.org/10.1016/j.jhydrol.2003.09.030.

Baez-Villanueva, O.M., Zambrano-Bigiarini, M., Mendoza, P.A., McNamara, I., Beck, H.E., Thurner, J., Nauditt, A., Ribbe, L., Thinh, N.X., 2021. On the selection of precipitation products for the regionalisation of hydrological model parameters. Hydrol. Earth Syst. Sci. 25, 5805–5837. https://doi.org/10.5194/hess-25-5805-2021.

Bandhauer, M., Isotta, F., Lakatos, M., Lussana, C., Båserud, L., Izsák, B., Szentes, O., Tveito, O.E., Frei, C., 2022. Evaluation of daily precipitation analyses in E-OBS (v19.0e) and ERA5 by comparison to regional high-resolution datasets in European regions. Int. J. Climatol. 42, 727–747. https://doi.org/10.1002/joc.7269.

Beck, H.E., Vergopolan, N., Pan, M., Levizzani, V., Van Dijk, A.I.J.M., Weedon, G.P., Brocca, L., Pappenberger, F., Huffman, G.J., Wood, E.F., 2017. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. Hydrol. Earth Syst. Sci. 21, 6201–6217. https://doi.org/10.5194/hess-21-6201-2017.

Beck, H.E., Pan, M., Roy, T., Weedon, G.P., Pappenberger, F., Van Dijk, A.I.J.M., Huffman, G.J., Adler, R.F., Wood, E.F., 2019. Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. Hydrol. Earth Syst. Sci. 23, 207–224. https://doi.org/10.5194/hess-23-207-2019.

Bitew, M.M., Gebremichael, M., 2011. Evaluation of satellite rainfall products through hydrologic simulation in a fully distributed hydrologic model. Water Resour. Res. 47, 2010WR009917. https://doi.org/10.1029/2010WR009917.

Blaney, H.F., Criddle, W.D., 1962. Determining Consumptive Use and Irrigation Water Requirements. USDA Technical Bulletin 1275, US Department of Agriculture, Beltsville.

Bonanno, R., Lacavalla, M., Sperati, S., 2019. A new high-resolution Meteorological Reanalysis Italian Dataset: MERIDA. Q. J. R. Meteorol. Soc. 145, 1756–1779. https://doi.org/10.1002/qj.3530.

Braca G., Bussettini M., Lastoria B., Mariani S., Piva F., 2021, Il Bilancio Idrologico Gis BAsed a scala Nazionale su Griglia regolare – BIGBANG: metodologia e stime. Rapporto sulla disponibilità naturale della risorsa idrica. Istituto Superiore per la Protezione e la Ricerca Ambientale, Rapporti 339/21, Roma.

Camici, S., Ciabatta, L., Massari, C., Brocca, L., 2018. How reliable are satellite precipitation estimates for driving hydrological models: A verification study over the Mediterranean area. J. Hydrol. 563, 950–961. https://doi.org/10.1016/j.jhydrol.2018.06.067.

Camici, S., Massari, C., Ciabatta, L., Marchesini, I., Brocca, L., 2020. Which rainfall score is more informative about the performance in river discharge simulation? A comprehensive assessment on 1318 basins over Europe. Hydrol. Earth Syst. Sci. 24, 4869–4885. https://doi.org/10.5194/hess-24-4869-2020.

Cammalleri, C., Sarwar, A.N., Avino, A., Nikravesh, G., Bonaccorso, B., Mendicino, G., Senatore, A., Manfreda, S., 2024. Testing trends in gridded rainfall datasets at relevant hydrological scales: A comparative study with regional ground observations in Southern Italy. J. Hydrol. Reg. Stud. 55, 101950. https://doi.org/10.1016/j.ejrh.2024.101950.

Capecchi, V., Pasi, F., Gozzini, B., Brandini, C., 2023. A convection-permitting and limited-area model hindcast driven by ERA5 data: precipitation performances in Italy. Clim. Dyn. 61, 1411–1437. https://doi.org/10.1007/s00382-022-06633-2.

Caroletti, G.N., Coscarelli, R., Caloiero, T., 2019. Validation of Satellite, Reanalysis and RCM Data of Monthly Rainfall in Calabria (Southern Italy). Remote Sens 11, 1625. https://doi.org/10.3390/rs11131625.

Cavalleri, F., Viterbo, F., Brunetti, M., Bonanno, R., Manara, V., Lussana, C., Lacavalla, M., Maugeri, M., 2024b. Inter-comparison and validation of high-resolution surface air temperature reanalysis fields over Italy. Int. J. Climatol. 44 (8), 2681–2700. https://doi.org/10.1002/joc.8475.

Cavalleri, F., Lussana, C., Viterbo, F., Brunetti, M., Bonanno, R., Manara, V., Lacavalla, M., Sperati, S., Raffa, M., Capecchi, V., Cesari, D., Giordani, A., Cerenzia, I., Maugeri, M. 2024a. Multi-Scale Assessment of High-Resolution Reanalysis Precipitation Fields Over Italy. Available at SSRN: https://doi.org/10.2139/ssrn.4896721.

Centella-Artola, A., Bezanilla-Morlot, A., Taylor, M.A., Herrera, D.A., Martinez-Castro, D., Gouirand, I., Sierra-Lorenzo, M., Vichot-Llano, A., Stephenson, T., Fonseca, C., Campbell, J., Alpizar, M., 2020. Evaluation of sixteen gridded precipitation datasets over the caribbean region using gauge observations. Atmosphere 11, 1334. https://doi.org/10.3390/atmos11121334.

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

Cornes, R.C., Van Der Schrier, G., Van Den Besselaar, E.J.M., Jones, P.D., 2018. An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets. J. Geophys. Res. Atmospheres 123, 9391–9409. https://doi.org/10.1029/2017JD028200.

Coron, L., Perrin, C., Michel, C., 2017a. airGR: suite of GR hydrological models for precipitation-runoff modelling. R. N.

Coron, L., Thirel, G., Delaigue, O., Perrin, C., Andréassian, V., 2017b. The suite of lumped GR hydrological models in an R package. Environ. Model. Softw. 94, 166–171. https://doi.org/10.1016/j.envsoft.2017.05.002.

Crisci, A., Gozzini, B., Meneguzzo, F., Pagliara, S., Maracchi, G., 2002. Extreme rainfall in a changing climate: regional analysis and hydrological implications in Tuscany. Hydrol. Process. 16, 1261–1274. https://doi.org/10.1002/hyp.1061.

Dalla Torre, D., Di Marco, N., Menapace, A., Avesani, D., Righetti, M., Majone, B., 2024. Suitability of ERA5-Land reanalysis dataset for hydrological modelling in the Alpine region. J. Hydrol. Reg. Stud. 52, 101718. https://doi.org/10.1016/j.ejrh.2024.101718.

De Andrade, J.M., Ribeiro Neto, A., Bezerra, U.A., Moraes, A.C.C., Montenegro, S.M.G.L., 2022. A comprehensive assessment of precipitation products: Temporal and spatial analyses over terrestrial biomes in Northeastern Brazil. Remote Sens. Appl. Soc. Environ. 28, 100842. https://doi.org/10.1016/j.rsase.2022.100842.

De Luca, D.A., Lasagna, M., Debernardi, L., 2020. Hydrogeology of the western Po plain (Piedmont, NW Italy). J. Maps 16, 265–273. https://doi.org/10.1080/17445647.2020.1738280.

De Lucia, C., Bucchignani, E., Mastellone, A., Adinolfi, M., Montesarchio, M., Cinquegrana, D., Mercogliano, P., Schiano, P., 2022. A Sensitivity Study on High Resolution NWP ICON—LAM Model over Italy. Atmosphere 13, 540. https://doi.org/10.3390/atmos13040540.

Dembélé, M., Schaefli, B., Van De Giesen, N., Mariéthoz, G., 2020. Suitability of 17 gridded rainfall and temperature datasets for large-scale hydrological modelling in West Africa. Hydrol. Earth Syst. Sci. 24, 5379–5406. https://doi.org/10.5194/hess-24-5379-2020.

Desiato, F., Lena, F., Toreti, A., 2007. SCIA: a system for a better knowledge of the Italian climate. Boll. Geofis. Teor. Ed. Appl. 48, 351–358.

Desiato, F., Fioravanti, G., Fraschetti, P., Perconti, W., Toreti, A., 2011. Climate indicators for Italy: calculation and dissemination. Adv. Sci. Res. 6, 147–150. https://doi.org/10.5194/asr-6-147-2011.

Duan, Z., Liu, J., Tuo, Y., Chiogna, G., Disse, M., 2016. Evaluation of eight high spatial resolution gridded precipitation products in Adige Basin (Italy) at multiple temporal and spatial scales. Sci. Total Environ. 573, 1536–1553. https://doi.org/10.1016/j.scitotenv.2016.08.213.

Fekete, B.M., Vörösmarty, C.J., Roads, J.O., Willmott, C.J., 2004. Uncertainties in precipitation and their impacts on runoff estimates. J. Clim. 17 (2), 294–304. https://doi.org/10.1175/1520-0442(2004)0172.0.CO;2.

Funk, C.C., Peterson, P., Landsfeld, M.F., Pedreros, D.H., Verdin, J.P., Rowland, J.D., Romero, B.E., Husak, G.J., Michaelsen, J.C., Verdin, A.P., 2014. A quasi-global precipitation time series for drought monitoring (USGS Numbered Series No. 832), Data Series. USGS.

Gampe, D., Ludwig, R., 2017. Evaluation of Gridded Precipitation Data Products for Hydrological Applications in Complex Topography. Hydrology 4, 53. https://doi.org/10.3390/hydrology4040053.

Gebrechorkos, S.H., Leyland, J., Dadson, S.J., Cohen, S., Slater, L., Wortmann, M., Ashworth, P.J., Bennett, G.L., Boothroyd, R., Cloke, H., Delorme, P., Griffith, H., Hardy, R., Hawker, L., McLelland, S., Neal, J., Nicholas, A., Tatem, A.J., Vahidi, E., Liu, Y., Sheffield, J., Parsons, D.R., Darby, S.E., 2024. Global-scale evaluation of precipitation datasets for hydrological modelling. Hydrol. Earth Syst. Sci. 28, 3099–3118. https://doi.org/10.5194/hess-28-3099-2024.

Gehne, M., Hamill, T.M., Kiladis, G.N., Trenberth, K.E., 2016. Comparison of Global Precipitation Estimates across a Range of Temporal and Spatial Scales. J. Clim. 29, 7773–7795. https://doi.org/10.1175/JCLI-D-15-0618.1.

Gentilucci, M., Barbieri, M., Pambianchi, G., 2022. Reliability of the IMERG product through reference rain gauges in Central Italy. Atmos. Res 278, 106340. https://doi.org/10.1016/j.atmosres.2022.106340.

Girons Lopez, M., Wennerström, H., Nordén, L., Seibert, J., 2015. Location and density of rain gauges for the estimation of spatial varying precipitation. Geogr. Ann. Ser. Phys. Geogr. 97, 167–179. https://doi.org/10.1111/geoa.12094.

Gomes, G., Thiemig, V., Skøien, J., Ziese, M., Rauthe-Schöch, A., Rustemeier, E., Rehfeldt, K., Walawender, J., Kolbe, Christine, Pichon, D., Schweim, C., Salamon, P., (2020). EMO: A high-resolution multi-variable gridded meteorological data set for Europe. European Commission, Joint Research Centre (JRC) [Dataset] doi: 10.2905/0BD84BE4-CEC8-4180-97A6-8B3ADAAC4D26.

Gomis-Cebolla, J., Rattayova, V., Salazar-Galán, S., Francés, F., 2023. Evaluation of ERA5 and ERA5-Land reanalysis precipitation datasets over Spain (1951–2020). Atmos. Res 284, 106606. https://doi.org/10.1016/j.atmosres.2023.106606.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol. 377, 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.

Hafizi, H., Sorman, A.A., 2022. Assessment of 13 Gridded Precipitation Datasets for Hydrological Modeling in a Mountainous Basin. Atmosphere 13, 143. https://doi.org/10.3390/atmos13010143.

Hagemann, S., Stacke, T., 2023. Complementing ERA5 and E-OBS with high-resolution river discharge over Europe. Oceanologia 65, 230–248. https://doi.org/10.1016/j.oceano.2022.07.003.

Hofstra, N., Haylock, M., New, M., Jones, P.D., 2009. Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature. J. Geophys. Res.: Atmospheres 114 (D21). https://doi.org/10.1029/2009JD011799.

Houngnibo, M.C.M., Minoungou, B., Traore, S.B., Maidment, R.I., Alhassane, A., Ali, A., 2023. Validation of high-resolution satellite precipitation products over West Africa for rainfall monitoring and early warning. Front. Clim. 5, 1185754. https://doi.org/10.3389/fclim.2023.1185754.

Jiang, C., Parteli, E.J.R., Xia, Q., Shao, Y., 2024. Evaluation of precipitation reanalysis products for regional hydrological modelling in the Yellow River Basin. Theor. Appl. Climatol. 155, 2605–2626. https://doi.org/10.1007/s00704-023-04758-w.

Jiao, D., Xu, N., Yang, F., Xu, K., 2021. Evaluation of spatial-temporal variation performance of ERA5 precipitation data in China. Sci. Rep. 11, 17956. https://doi.org/10.1038/s41598-021-97432-y.

Katsanos, D., Retalis, A., Michaelides, S., 2016. Validation of a high-resolution precipitation database (CHIRPS) over Cyprus for a 30-year period. Atmos. Res 169, 459–464. https://doi.org/10.1016/j.atmosres.2015.05.015.

Khan, M.W., Ahmad, S., Dahri, Z.H., Syed, Z., Ahmad, K., Khan, F., Azmat, M., 2023. Development of high resolution daily gridded precipitation and temperature dataset for potohar plateau of indus basin. Theor. Appl. Climatol. 154, 1179–1201. https://doi.org/10.1007/s00704-023-04626-7.

Kidd, C., Becker, A., Huffman, G.J., Muller, C.L., Joe, P., Skofronick-Jackson, G., Kirschbaum, D.B., 2017. So, How Much of the Earth's Surface Is Covered by Rain Gauges? Bull. Am. Meteorol. Soc. 98, 69–78. https://doi.org/10.1175/BAMS-D-14-00283.1.

Kostopoulou, E., Giannakopoulos, C., Hatzaki, M., Tziotziou, K., 2012. Climate extremes in the NE Mediterranean: assessing the E-OBS dataset and regional climate simulations. Clim. Res. 54, 249–270. https://doi.org/10.3354/cr01110.

Krauskopf, T., Huth, R., 2020. Temperature trends in Europe: comparison of different data sources. Theor. Appl. Climatol. 139, 1305–1316. https://doi.org/10.1007/s00704-019-03038-w.

Laiti, L., Mallucci, S., Piccolroaz, S., Bellin, A., Zardi, D., Fiori, A., Nikulin, G., Majone, B., 2018. Testing the Hydrological Coherence of High-Resolution Gridded Precipitation and Temperature Data Sets. Water Resour. Res. 54, 1999–2016. https://doi.org/10.1002/2017WR021633.

Lemma, E., Upadhyaya, S., Ramsankaran, R., 2022. Meteorological drought monitoring across the main river basins of Ethiopia using satellite rainfall product. Environ. Syst. Res. 11, 7. https://doi.org/10.1186/s40068-022-00251-x.

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

Maggioni, V., Nikolopoulos, E.I., Anagnostou, E.N., Borga, M., 2017. Modeling Satellite Precipitation Errors Over Mountainous Terrain: The Influence of Gauge Density, Seasonality, and Temporal Resolution. IEEE Trans. Geosci. Remote Sens. 55, 4130–4140. https://doi.org/10.1109/TGRS.2017.2688998.

Manco, I., De Lucia, C., Repola, F., Fedele, G., Mercogliano, P., 2023. A Comparative Performance Study of WRF, COSMO and ICON Atmospheric Models for the Italian Peninsula at Very High Resolution. Tethys J. Weather Clim. West. Mediterr. https://doi.org/10.3369/tethys.2023.20.01.

Mankin, K.R., Mehan, S., Green, T.R., Barnard, D.M., 2024. Review of Gridded Climate Products and Their Use in Hydrological Analyses Reveals Overlaps, Gaps, and Need for More Objective Approach to Model Forcings. https://doi.org/10.5194/hess-2024-58.

Mavromatis, T., Voulanas, D., 2021. Evaluating ERA-Interim, Agri4Cast, and E-OBS gridded products in reproducing spatiotemporal characteristics of precipitation and drought over a data poor region: The Case of Greece. Int. J. Climatol. 41, 2118–2136. https://doi.org/10.1002/joc.6950.

Mazzoleni, M., Brandimarte, L., Amaranto, A., 2019. Evaluating precipitation datasets for large-scale distributed hydrological modelling. J. Hydrol. 578, 124076. https://doi.org/10.1016/j.jhydrol.2019.124076.

McMillan, H., Jackson, B., Clark, M., Kavetski, D., Woods, R., 2011. Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models. J. Hydrol. 400, 83–94. https://doi.org/10.1016/j.jhydrol.2011.01.026.

Mei, Y., Nikolopoulos, E.I., Anagnostou, E.N., Borga, M., 2016. Evaluating Satellite Precipitation Error Propagation in Runoff Simulations of Mountainous Basins. J. Hydrometeorol. 17, 1407–1423. https://doi.org/10.1175/JHM-D-15-0081.1.

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., Thépaut, J.-N., 2021. ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. Earth Syst. Sci. Data 13, 4349–4383. https://doi.org/10.5194/essd-13-4349-2021.

Nijssen, B., Lettenmaier, D.P., 2004. Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the Global Precipitation Measurement satellites. J. Geophys. Res. Atmospheres 109, 2003JD003497. https://doi.org/10.1029/2003JD003497.

Nikolopoulos, E.I., Anagnostou, E.N., Borga, M., 2013. Using High-Resolution Satellite Rainfall Products to Simulate a Major Flash Flood Event in Northern Italy. J. Hydrometeorol. 14, 171–185. https://doi.org/10.1175/JHM-D-12-09.1.

Oudin, L., Perrin, C., Mathevet, T., Andréassian, V., Michel, C., 2006. Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models. J. Hydrol. 320, 62–83. https://doi.org/10.1016/j.jhydrol.2005.07.016.

Padulano, R., Lama, G.F.C., Rianna, G., Santini, M., Mancini, M., Stojiljkovic, M., 2020. Future rainfall scenarios for the assessment of water availability in Italy, in: 2020 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor). Presented at the 2020 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor), IEEE, Trento, Italy, pp. 241–246. https://doi.org/10.1109/MetroAgriFor50201.2020.9277599.

Padulano, R., Rianna, G., Santini, M., 2021. Datasets and approaches for the estimation of rainfall erosivity over Italy: A comprehensive comparison study and a new method. J. Hydrol. Reg. Stud. 34, 100788. https://doi.org/10.1016/j.ejrh.2021.100788.

Pandey, V., Srivastava, P.K., Singh, S.K., Petropoulos, G.P., Mall, R.K., 2021. Drought Identification and Trend Analysis Using Long-Term CHIRPS Satellite Precipitation Product in Bundelkhand, India. Sustainability 13, 1042. https://doi.org/10.3390/su13031042.

Paredes-Trejo, F., Olivares, B.O., Movil-Fuentes, Y., Arevalo-Groening, J., Gil, A., 2023. Assessing the spatiotemporal patterns and impacts of droughts in the Orinoco River basin using earth observations data and surface observations. Hydrology 10 (10), 195. https://doi.org/10.3390/hydrology10100195.

Pushpalatha, R., Perrin, C., Le Moine, N., Mathevet, T., Andréassian, V., 2011. A downward structural sensitivity analysis of hydrological models to improve low-flow simulation. J. Hydrol. 411, 66–76. https://doi.org/10.1016/j.jhydrol.2011.09.034.

Raffa, M., Reder, A., Marras, G.F., Mancini, M., Scipione, G., Santini, M., Mercogliano, P., 2021. VHR-REA_IT dataset: very high resolution dynamical downscaling of ERA5 reanalysis over Italy by COSMO-CLM. Data 6 (8), 88. https://doi.org/10.3390/data6080088.

Rianna, G., Reder, A., Sousa, M.L., Dimova, S., 2023. Harmonised procedure to update thermal loads in the Eurocodes. Case study for Italy. Clim. Serv. 30, 100391. https://doi.org/10.1016/j.cliser.2023.100391.

Ridal, M., Bazile, E., Le Moigne, P., Randriamampianina, R., Schimanke, S., Andrae, U., Berggren, L., Brousseau, P., Dahlgren, P., Edvinsson, L., El-Said, A., 2024. CERRA, the Copernicus European Regional Reanalysis system. Q. J. R. Meteorol. Soc. DOI: 10.1002/qj.4764.

Rivera, J.A., Marianetti, G., Hinrichs, S., 2018. Validation of CHIRPS precipitation dataset along the Central Andes of Argentina. Atmos. Res 213, 437–449. https://doi.org/10.1016/j.atmosres.2018.06.023.

Rivera, J.A., Hinrichs, S., Marianetti, G., 2019. Using CHIRPS Dataset to Assess Wet and Dry Conditions along the Semiarid Central-Western Argentina. Adv. Meteorol. 2019, 1–18. https://doi.org/10.1155/2019/8413964.

Rockel, B., Will, A., Hense, A., 2008. The regional climate model COSMO-CLM (CCLM). Meteorol. Z. 17 (4), 347. https://doi.org/10.1127/0941-2948/2008/0309.

Saemian, P., Hosseini-Moghari, S.-M., Fatehi, I., Shoarinezhad, V., Modiri, E., Tourian, M.J., Tang, Q., Nowak, W., Bárdossy, A., Sneeuw, N., 2021. Comprehensive evaluation of precipitation datasets over Iran. J. Hydrol. 603, 127054. https://doi.org/10.1016/j.jhydrol.2021.127054.

Satgé, F., Pillot, B., Roig, H., Bonnet, M.-P., 2021. Are gridded precipitation datasets a good option for streamflow simulation across the Juruá river basin, Amazon? J. Hydrol. 602, 126773. https://doi.org/10.1016/j.jhydrol.2021.126773.

Seibert, J., Clerc-Schwarzenbach, F.M., (Ilja) Van Meerveld, H.J., 2024. Getting your money's worth: Testing the value of data for hydrological model calibration. Hydrol. Process. 38, e15094. https://doi.org/10.1002/hyp.15094.

Silva, E.H.D.L., Silva, F.D.D.S., Junior, R.S.D.S., Pinto, D.D.C., Costa, R.L., Gomes, H.B., Júnior, J.B.C., De Freitas, I.G.F., Herdies, D.L., 2022. Performance Assessment of Different Precipitation Databases (Gridded Analyses and Reanalyses) for the New Brazilian Agricultural Frontier: SEALBA. Water 14, 1473. https://doi.org/10.3390/w14091473.

Sivasubramaniam, K., Alfredsen, K., Rinde, T., Sæther, B., 2020. Can model-based data products replace gauge data as input to the hydrological model? Hydrol. Res. 51, 188–201. https://doi.org/10.2166/nh.2020.076.

Stephens, C.M., Pham, H.T., Marshall, L.A., Johnson, F.M., 2022. Which Rainfall Errors Can Hydrologic Models Handle? Implications for Using Satellite-Derived Products in Sparsely Gauged Catchments. Water Resour. Res. 58, e2020WR029331. https://doi.org/10.1029/2020WR029331.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., Hsu, K., 2018. A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons. Rev. Geophys. 56, 79–107. https://doi.org/10.1002/2017RG000574.

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

Tolson, B.A., Shoemaker, C.A., 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resour. Res. 43, 2005WR004723. https://doi.org/10.1029/2005WR004723.

Tuo, Y., Duan, Z., Disse, M., Chiogna, G., 2016. Evaluation of precipitation input for SWAT modeling in Alpine catchment: A case study in the Adige river basin (Italy). Sci. Total Environ. 573, 66–82. https://doi.org/10.1016/j.scitotenv.2016.08.034.

Turco, M., Zollo, A.L., Ronchi, C., De Luigi, C., Mercogliano, P., 2013a. Assessing gridded observations for daily precipitation extremes in the Alps with a focus on northwest Italy. Nat. Hazards Earth Syst. Sci. 13, 1457–1468. https://doi.org/10.5194/nhess-13-1457-2013.

Turco, M., Zollo, A.L., Vezzoli, R., Ronchi, C., Mercogliano, P., 2013b. Daily precipitation statistics over the Po Basin: observation and post-processed RCM results, in: Climate Change and Its Implications on Ecosystem and Society. Presented at the First SISC Conference, Lecce.

Valencia, S., Marín, D.E., Gómez, D., Hoyos, N., Salazar, J.F., Villegas, J.C., 2023. Spatio-temporal assessment of Gridded precipitation products across topographic and climatic gradients in Colombia. Atmos. Res 285, 106643. https://doi.org/10.1016/j.atmosres.2023.106643.

Valéry, A., Andréassian, V., Perrin, C., 2014. As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 2 – Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments. J. Hydrol. 517, 1176–1187. https://doi.org/10.1016/j.jhydrol.2014.04.058.

Xu, J., Ma, Z., Yan, S., Peng, J., 2022. Do ERA5 and ERA5-land precipitation estimates outperform satellite-based precipitation products? A comprehensive comparison between state-of-the-art model-based and satellite-based precipitation products over mainland China. J. Hydrol. 605, 127353. https://doi.org/10.1016/j.jhydrol.2021.127353.

Zhang, Y., Ye, A., Nguyen, P., Analui, B., Sorooshian, S., Hsu, K., 2021. Error characteristics and scale dependence of current satellite precipitation estimates products in hydrological modeling. Remote Sens 13, 3061. https://doi.org/10.3390/rs13163061.