



Ex-post assessment of climate and hydrological projections: reliability of CMPI6 outputs in Northern Italy

Flavia Fuso¹ · Giovanni Martino Bombelli¹ · Daniele Bocchiola¹

Received: 30 November 2022 / Accepted: 27 September 2023 / Published online: 13 October 2023
© The Author(s) 2023

Abstract

This paper presents a validation of outputs from some GCMs of the CMIP6 project when used to assess climate projection and hydrological flows at a catchment scale for the case study area of the Lombardy region (Northern Italy). The modeling chain consists of (i) a choice of climatic scenarios from 10 GCMs of the CMIP6, (ii) the application of a stochastic downscaling procedure to make projections usable at the local scale, and (iii) the use of a semi-distributed physically based hydrological model *Poli-Hydro* for the generation of hydrological scenarios. Data on observed precipitation and temperature were collected from automatic weather stations, and the hydrological budget of four target catchments within the study area was assessed using *Poli-Hydro*. An *ex-post* (back-casting) analysis was performed upon the control data series from the GCMs by comparing statistics of relevant climate variables and model-simulated discharges against observed counterparts during the historical period 2002–2014. Then, during 2015–2021, the goodness of projections was assessed using confidence intervals. Our results show that the accuracy of GCMs in representing regional climate is not always reflected in a credible evaluation of local hydrology. The validation of climate patterns provides somewhat poor results; thus, the interaction among climate and hydrology needs to be explored carefully to warrant the credibility of hydrological scenarios. Overall, the spatial and temporal consistency of GCM projections, as explored here climatically and hydrologically, provides a clue about their dependability for basin scale management.

1 Introduction

Nowadays, uncertainty in future climate projections is a main concern since these are widely used to assess the impacts of climate change on the environment and to plan adaptation strategies worldwide (Kundzewicz et al. 2008).

Uncertainty of climate models' outputs arises from the challenge of representing physical processes, e.g., clouds, turbulence, and further from poor understanding of other processes that can be represented only in simplified form, such as the effects upon climate of growing trees and/or presence of buildings (Curry and Webster 2011). These issues are not resolved in global climate models, but indeed, they impact the simulation of wide-scale temperature and precipitation. On top of such complexity, one has the involvement of multiple subsystems and uncertainty associated with the models' structure, e.g.,

parameters, equations, and initial and boundary conditions (Palmer et al. 2005). The Intergovernmental Panel on Climate Change IPCC set several CO₂ concentration pathways as external boundary conditions for (future) climate modeling. In the latest Assessment Report (AR6), it is stated that human activities have been the main cause of present global warming. However, there is still skepticism about the reliability of climate projections (Busch and Judick 2021). Global circulation models (GCMs) provide climate projections on a coarse grid resolution, generally not suitable to represent climatic variability at a local scale. The consequence is a lack of confidence in the future projections (Räisänen 2007) and possible over/underestimation of precipitation projections and the related risks (Moreno-Chamarro et al. 2021). This range of uncertainty increases when assessing hydrological projections since a complex modeling chain, from GCMs scenarios to downscaling (most often necessary to cope with the large-scale jump from climate models to hydrological ones), and then hydrological modeling needs to be applied here (Casale et al. 2021). Every step introduces uncertainty, leading to an overall decrease in the dependability

✉ Daniele Bocchiola
daniele.bocchiola@polimi.it

¹ Department of Civil and Environmental Engineering DICA, Politecnico di Milano, L. da Vinci 32, 20133 Milano, Italy

of future projections (Camici et al. 2017). Although there is no common agreement upon validation methods for climate models (Guillemot 2010), one method to measure/constrain uncertainty is cross-comparison against other simulations and or against observations, and indeed, the use of past observational data is essential in evaluating possible future scenarios (Koutsoyiannis et al. 2009). It seems clear that measuring such uncertainty plays a paramount role in water resource management, not only when applying structural measures but also for economic and social strategies (Knutti 2008). Hydrologists are thus required to assess accurately the impact of climate projections on the hydrological cycle (Kang and Ramírez 2007). Moreover, when assessing the impact of climate change on a regional scale, it is essential to evaluate the ability of GCMs to describe local meteorological variability; otherwise, poorly representative climate projections may be obtained in the study area (Salathe et al. 2007; Eyring et al. 2016).

The two-fold purpose of this study was to validate control data series from some GCMs and to then evaluate the accuracy of the subsequent climate projections at the local scale, also in assessing hydrological discharges at a catchment scale, with a focus upon a case study area in the Lombardy region of Northern Italy. The chosen region is a sensitive target for climate change because (i) it is largely covered by mountains, the (cryosphere dependent) hydrological regimes of which is already suffering from global warming (Beniston et al. 2011; Fuso et al. 2021), and (ii) it hosts two large regulated lakes, crucial for downstream irrigation and hydropower production (Anghileri et al. 2011). Therefore, reliable hydrological projections are essential from an environmental, social, and economic point of view. We used here (i) 10 GCMs and 4 scenarios of projected socioeconomic global changes (shared socioeconomic pathways (SSPs)) of the CMIP6 of IPCC (with a forward projection phase starting in 2015), (ii) a downscaling model to make the climate projections suitable at the local scale, and (iii) the semi-distributed physically based hydrological model *Poli-Hydro* to model hydrology under the obtained climate scenarios. To assess the robustness of the hydrological scenarios, a back-cast analysis has been carried out, both in terms of climatic drivers, i.e., precipitation and temperature, and of hydrological discharges, by comparing projections against observed data for 20 years, 2002–2021. First, we performed the back-cast against the control data series (2002–2014) of the 10 GCMs after downscaling by comparing the related statistics of temperature, precipitation, and discharges against the observed ones. After such a test, the most representative GCMs for the area were evaluated. Then, we assessed the goodness of future projections (here, 2015–2021) by evaluating the confidence interval of the SSPs and verifying whether the observed data are well

contained within the range of uncertainty (i.e., confidence limits) of the projected variables.

The paper is structured as follows. In the “Study area” section, the study area is presented. The methodology is reported in the “Methodology” section. This includes the (choice of) global circulation models, the downscaling procedure, and the hydrological modeling, followed by the back-cast analysis and the projections. The results are shown in the “Results” section. Discussion and conclusions are in the “Discussion” and “Conclusion” sections, respectively.

2 Study area

2.1 Case study

The study area is a part of the Lombardy region, nested within the Ticino-Adda catchment (Fig. 1). The area hosts several mountains and glaciers, with the highest elevation of 4020 m, thus being largely snow/ice fed and rich in both surface and aquifer water (Casale et al. 2021). Two major lakes are located inside the basin, i.e., Lake Maggiore, with a volume of 37 km³, and Lake Como, with a volume of 23.4 km³. The first is watered almost equally from the Piemonte region and Switzerland rivers, whereas the catchment of the second is for 90% in Lombardy and 10% in Switzerland.

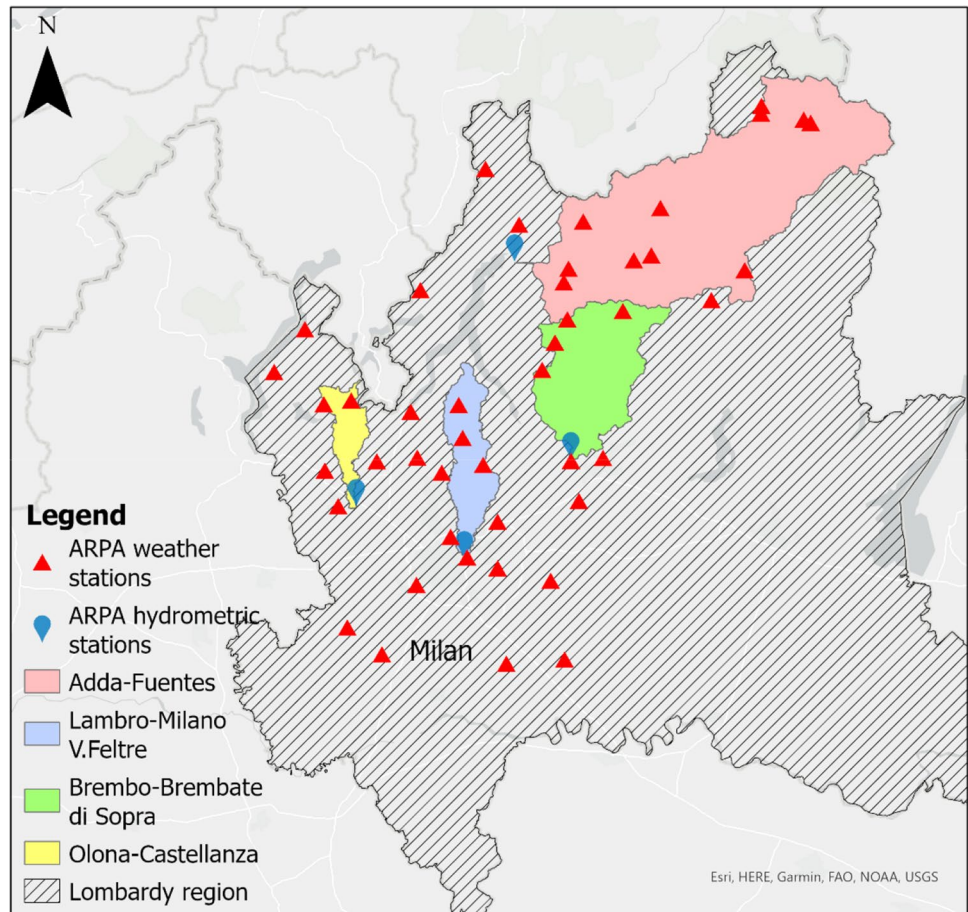
The climate in the region is mainly cold in the lake Como catchment, with hot summers at low altitudes, with temperature decreasing with altitude. Overall, in the high-altitude catchments, temperature varies between -4°C in winter and $+15^{\circ}\text{C}$ in summer. Total precipitation is around 1300 mm/year on average, with peaks in May and November. Downstream into the lowlands, the climate is temperate, with the highest temperature up to $+23^{\circ}\text{C}$ in summer and peaks of precipitation in autumn, with a total of 1800 mm/year on average.

The lowland area is inside the Po valley, the most productive agricultural area of Europe (Bocchiola et al. 2013), intensively exploited for agriculture and hydropower production. In recent years, due to climate change, summer droughts have requested more water to be released from the lake, conflicting with interests from hydropower producers needing water during winter (Denaro et al. 2018). Future projections would depict a potential worsening of these issues, given that the increase in temperature and the decrease in precipitation will necessarily call for a renewal in the management strategies of water resources (Fuso et al. 2021).

2.2 Data

Two observational datasets were employed in this study, i.e., (i) daily series of precipitation and temperature (P, T)

Fig. 1 Automatic weather stations (AWSs) available in the study area. Hydrological sub-basins identified by the hydrometers located at Fuentes, Milano Feltre, Brembate di Sopra, and Castellanza



collected from automatic weather stations (AWS) of ARPA Lombardia for a baseline period (BP) 2002–2021 and (ii) stream flows from four hydrometric stations (daily water level, converted to discharge). In Fig. 1, the 45 AWSs and the catchment area contributing to the four river sections (Adda-Fuentes, Lambro-Milano Via Feltre, Brembo-Brembate di Sopra, and Olona-Castellanza) are shown. The meteorological data available here were used both for GCM validation to feed the hydrological model *Poli-Hydro*. The hydrometric stations were selected in order to investigate basins with different dimensions and features, as shown in Table 1.

3 Methodology

In this section, our modeling chain used is presented. First, the GCMs used are listed, each with 4 SSPs of the CMIP6. Given the coarse resolution of the GCMs, a downscaling procedure was necessary. Then, the precipitation and temperature series from GCMs, properly downscaled, were used to feed the hydrological model, previously tuned for the baseline period. To increase the reliability of the climate and hydrological projections and to evaluate the most representative GCMs within the chosen set, a back-cast analysis of GCM control data was performed. Finally, the goodness

Table 1 Catchments’ features, mean discharge, observed during 2002–2021, and bias obs/mod for each sub-basin averaged during the years 2002–2021

Catchment	Area [km ²]	Mean elevation [m a.s.l.]	Max elevation [m a.s.l.]	Mean discharge [m ³ s ⁻¹]	Bias obs /mod
Adda – Fuentes	2526.16	1888.7	3801	85.43	−8.2%
Lambro – Milano Via.Feltre	355.63	370.11	1216	9.76	−1.4%
Brembo – Brembate di Sopra	787.121	1475.5	2572	26.32	+1.9%
Olona – Castellanza	173.81	375.48	876	4.71	+10.6%

of the SSPs projections was assessed, testing whether the observational data occurred within the range of the projected variables.

3.1 Global circulation models

In this study, 10 GCMs were employed, of which scenarios are being used as part of the experiment Coupled Model Intercomparison Project CMIP6 of the IPCC. We used the following models: CNRM-CM6-1 (Voldoire et al. 2019), IPSL-CM6A-LR (Lurton et al. 2020), NorESM2-LM (Seland et al. 2020), INM-CM5-0 (Volodin et al. 2018), EC-EARTH3 (Döscher et al. 2022), CESM2 (Lauritzen et al. 2018), ECHAM6.3 (Mauritsen et al. 2019), CMCC-CM2 (Cherchi et al. 2018), UKESM1 (Sellar et al. 2019), and MIRCOC6 (Tatebe et al. 2019). The forcing scenarios are a combination of socioeconomic scenarios (SSPs) and representative greenhouse emission scenarios (RCPs, expressed in W/m^2) used for the AR5 (Taylor et al. 2012). The SSPs represent several possible evolutions of society, considering investments in education, health, and energy development. SSP 1 and SSP 5 project a positive development of society, but while the latter devise an economy based on fossil fuel, the former conceives a sustainable economy. The SSP 2 scenario depicts a continuation of the historical trend (business-as-usual), while the SSP 3 and SSP 4 foresee a negative (climate-wise) development of societal dynamics worldwide. Four SSP scenarios were used in this study, based on RCP 2.6, 4.5, and 8.5 scenarios, namely, SSP1.2.6, SSP2 4.5, SSP5 8.5, and an intermediate SSP3 7.0 scenario (O'Neill et al. 2016).

3.2 Downscaling procedure

The GCM outputs come with a ~ 100 km cell resolution, much coarser than the resolution of a typical hydrological model (such as *Poli-Hydro*). For this reason, stochastic downscaling is necessary to make the output of the GCM usable at the resolution of the local data or at the resolution of the model (here, $1 \text{ km} \times 1 \text{ km}^2$).

Downscaling of precipitation was pursued through the stochastic space random cascade method (e.g., Groppelli et al. 2011a), applied here in time (e.g., Bocchiola and Rosso 2006) to reproduce the observed precipitation variability (e.g., Groppelli et al. 2011b). The control run CR period 2002–2014 was defined to compare the historical series of GCMs against observed data of precipitation. If one defines R_d^{GAO} as the observed average daily precipitation in a given day d , at a given AWS, while R_d^{GCM} is the daily precipitation simulated by the GCM (in the cell including the AWAs site), the goal of downscaling can be defined as matching (statistically, in terms of mean and variance) the (corrected) value of $R_d^{GCM,corr}$ to the value of R_d^{GAO} .

$$R_d^{GCM,corr} = R_d^{GCM} / Bias_{GAO} B_0 W_0 \quad (1)$$

$$Bias_{GAO} = E[R_{GAO}] / E[R_{GCM}] \quad (2)$$

$$P(B_0 = 1/p_0) = p_0 \quad (3)$$

$$P(B_0 = 0) = 1 - p_0 \quad (4)$$

$$W_0 = \exp(w_0 - \sigma_{w_0}^2/2) \quad (5)$$

With

$$E[B_0] = p_0 \cdot 1/p_0 + 0 \cdot (1 - p_0) = 1$$

$$E[W_0] = 1$$

$$w_0 \sim N(0, \sigma_{w_0}^2)$$

$Bias_{GAO}$, p_0 , and $\sigma_{w_0}^2$ are model parameters to be tuned based on observational data. B_{GAO} is, in practice, a calculated *bias*, forcing the R^{GCM} average to coincide with the R^{GAO} average. B_0 is a β model generator and represents the probability that R^{GAO} is null, conditioned to a positive value of R^{GCM} . W_0 is a positive parameter that gives variability to precipitation. Temperature is downscaled by applying a monthly average temperature *Bias* (ΔT) (Groppelli et al. 2011b). An offset is calculated at the monthly scale between GCM and observed temperature.

$$T_{d,i}^{GCM,corr} = T_{d,i}^{GCM} - T_i^{GCM} - T_i^{obs} \quad (6)$$

Here, $T_{d,i}^{GCM}$ is the temperature of day d in month i , given by the GCM. T_i^{GCM} and T_i^{obs} are the average temperature in month i , given by the GCM and the station, respectively. $T_{d,i}^{GCM,corr}$ is then the corrected temperature of day d , based on the mean temperature shift between the given GCM and observed temperature.

After this procedure, a total of 40 precipitation and temperature scenarios ($4 \text{ SSPs} \times 10 \text{ GCMs}$) were used to feed *Poli-Hydro* for the assessment of hydrological scenarios.

3.3 Hydrological model

We used here the physically based, semi-distributed hydrological model *Poli-Hydro* (Soncini et al. 2017; Aili et al. 2019; Fuso et al. 2021). The model simulates in each grid cell the main physical processes, including (i) snow and ice melt using a mixed degree-day approach based on average daily temperature and short-wave radiation (Aili et al. 2019) and (ii) potential evapotranspiration using Hargreaves' formula, and subsequently actual evapotranspiration, evaluated

based on soil water content, assessed via water budget. The results of model tuning for the involved catchment in terms of the $Bias_{obs/mod}$ are reported in Table 1 for the period BP 2002–2021.

The *Poli-Hydro* model was tested within a large array of areas worldwide and climate conditions, starting from mountainous, ice/snow-fed catchments in Italy (e.g., Soncini et al. 2017), in South America (Bocchiola et al. 2018), the greater Himalayas (Soncini et al. 2015, 2016; Casale et al. 2020), and Caucasus (Baldasso et al. 2019) to semi-arid areas of Europe (e.g., Capolongo et al. 2019) and central Asia (Akbari et al. 2018) and to tropical areas in the Caribbeans (Bozza et al. 2016), Africa (Bombelli et al. 2021), and Indonesia (Stucchi and Bocchiola 2023). *Poly-Hydro* model performed well within all such climatic, topographic, and hydrological setups.

Notice further that the model has specifically been used for hydrological simulation in the area of interest in former contributions, with good outcomes, especially in Adda river and Como lake catchment (Casale et al. 2021), and also in Serio river (not considered here, bordering east of Brembo river in the DEM in Fig. 1; Fuso et al. 2023).

This is a substantially homogeneous area in terms of climate (temperate dry, with hot/warm summer Csa/b, to polar cold and glacial ET, EF at the highest altitudes; Peel et al. 2007), topography (steep mountainous catchments high altitude until ca. 4000 m slm, with large snow/ice feeding), and hydrology (bimodal regime with spring/fall flood season and dry summers; e.g., Bocchiola 2014).

Accordingly, and based on the results here presented for model calibration, we can assume that the *Poli-Hydro* model is flexible (i.e., in terms of parameters tuning) and portable (between catchments) enough that it can be confidently used here for our purpose of GCM–RCM evaluation and ranking (e.g., Dakhlaoui and Djebbi 2021).

3.4 Back-cast analysis

All GCMs provide climate projections with coarse (spatial) resolution, much lower with respect to the resolution of the hydrological model. To make the projections usable for the case study, a downscaling process was performed. The projected period, e.g., when the models start projecting forward future climate variables, starts in 2015 for CMIP6. Anyway, a window of 20 years during 2002–2021 BP was considered, and thus, a period of 13 years before the start of the future projections was taken as a historical validation period.

First, GCM validation is carried out via an ex-post analysis of the control data series. This analysis involves comparing the precipitation and temperature obtained from the GCMs (after downscaling) against the observed values for each year (2002–2014). This analysis has been carried out both in terms of climatic drivers, i.e., precipitation and

temperature, and of hydrological discharges to test the accuracy of GCMs of the downscaling procedure and hydrological modeling.

3.4.1 Climate validation

To assess the suitability of each GCM to represent the local weather, some tests were carried out. A Student's t -test (hypothesis H_0 , same mean value) was performed for the observed (P_{obs}, T_{obs}) and predicted values of the target variables (P_{GCM}, T_{GCM}), and furthermore, a Fisher's F -test (hypothesis H_0 , same variance) was carried out for the variance of precipitation. Temperature variance was not tested because it was observed that (i) once *Bias* corrected, GCMs generally depict sufficiently well periodic changes in precipitation, and (ii) high-frequency (i.e., daily) changes of temperature do not generally affect hydrological behavior, more affected by seasonal patterns. The observed data of cumulated annual precipitation and mean annual temperature were assessed against the observed data at the scale of each catchment. To do so, we proceeded as follows. Within each basin of interest, we calculated for every day d , the spatially averaged value of precipitation, $P_{av,d}$ and of temperature $T_{av,d}$ within the basin. The averaging was carried out by weighting values in each station (including the corrected values of the GCMs precipitation $R_d^{GCM,corr}$ and temperature $T_{d,i}^{GCM,corr}$, downscaled as reported with reference to the AWS stations' sites), according to the corresponding Thiessen polygon, consistently with the fact that Thiessen method is often adopted as a rapid interpolation method, especially for hydrological purposes, and *Poli-hydro* model also operates in this mode. Concerning temperature, to account for the altitudinal thermal shift, we applied a proper lapse rate correction to the AWS temperature within each polygon (Soncini et al. 2017). At the AWS stations, snowfall was identified using snow gauges, and new snow included accounted for in precipitation assessment (with a fresh snow density of 125 kgm^{-3} ; e.g., Bocchiola and Rosso 2007). From the so-obtained spatially average daily values in each catchment, we could calculate several statistics to be tested.

The tests were performed on two different types of samples for each variable, i.e., (i) a sample of cumulated daily precipitation $S_a(P_{obs}), S_a(P_{GCM})$ and mean daily temperature $S_a(T_{obs}), S_a(T_{GCM})$ for each year (2002–2014) and (ii) a sample of cumulated daily precipitation $S_s(P_{obs}), S_s(P_{GCM})$ and mean daily temperature $S_s(T_{obs}), S_s(T_{GCM})$ over the four seasons, considering the entire historical period. The same tests were carried out, also considering monthly aggregation instead of daily aggregation, thus reducing the sample size. We also considered (iii) monthly cumulated precipitation $S_{ca}(P_{GCM})$ and mean monthly temperature $S_{as}(T_{GCM})$ for each year and (iv) cumulated monthly precipitation $S_{cs}(P_{GCM})$ and mean monthly temperature $S_{cs}(T_{GCM})$ over the four seasons,

considering the entire historical period. Concerning precipitation, we did not include a statistical analysis of zero rainfall days (dry spells). Zero rainfall days may reflect internal variability of the phenomenon in time and may influence droughts/flood dynamics and, in general, flow seasonality/regime. However, we decided to test here the precipitation variability because we assume that this will affect more directly stream flow variability. Dry spells will be therein, somewhat reflected by low values of precipitation.

We carried out the tests for each GCMs control series downscaled, for a total of 140 t -tests for P and T and 140 F -tests for P (10 GCMs \times 14 years) and 40 t -tests for P and T and 40 F -tests for P (10 GCMs \times 4 seasons) for each basin. These tests were performed considering both daily (i, ii) and monthly (iii, iv) aggregation. Yearly and seasonal p -values were obtained, measuring the goodness of fit of the (first and second for rainfall) statistics of the simulated variables against the observed ones.

3.4.2 Hydrological validation

To validate the GCMs downstream of the modeling chain, a Student's t -test was carried out to compare the modeled discharges from GCMs (Q_{GCM}) against the modeled discharge from AWS data (Q_{mod}). This comparison was made against the modeled discharges since the performance of the hydrological model was already assessed and deemed acceptable ("Hydrological model") and to avoid including the comparison of the effect of the model's noise. To assess the hydrological variations from GCM projections generally, we focused on the difference between mean discharges over 10 years, i.e., at mid-century and mean discharges over 10 years of the CR (Aili et al. 2019; Fuso et al. 2021; Bombelli et al. 2021; Casale et al. 2021). For this reason, we decided to test the hydrological performance of the GCMs considering 4 mobile windows of 10 years, i.e., 4 T -tests were performed for each GCM (2002–2011, 2003–2012, 2004–2013, 2005–2014), considering a sample of 3650 daily values. Like it was done for climate variables, we carried out the tests also considering the mean monthly discharges, thus reducing the sample size to 120 values.

3.5 Projections of future climate and hydrology

After the GCM validation, we averaged among the four SSPs of the outputs from the 10 GCMs, and we assessed the goodness of climate and hydrological projections by verifying whether the observed (in the future) data are within the range of the projected variables during the years 2015–2021. The range of each SSP is represented by the confidence interval at 95%. For precipitation and temperature, the comparison is carried out between the GCM projections properly downscaled against the observed AWS data, while for the

discharges, we compared the modeled discharges with *Poli-hydro* using the inputted AWS data vs. the climate projections downscaled. In doing so, one may (i) assess changes in future vs. future stream flows by offsetting the hydrological model's error, only considering the effects of the climate, and (ii), specifically here, assess the dependability of outputs from each GCM, independently of the hydrological model interference. Another reason for not using observed stream flows is indeed the potential presence of regulation. Hydrological models generally do not include regulation, and the obtained streamflow series are solely driven by climate variability. Accordingly, when assessing the credibility of future projected streamflow series (against observations), potential differences against the observations may actually depend upon the presence of (even slight) river regulation, which, however, has little to do with the dependability of the tested GCMs. Such potential issues may be actually bypassed using simulated series, both in the control run and in the future. Thereby, any potential effect of regulation would be bypassed, and one would observe only changes/differences as given by the climate input.

4 Results

4.1 Back-cast analysis

4.1.1 Climate models validation

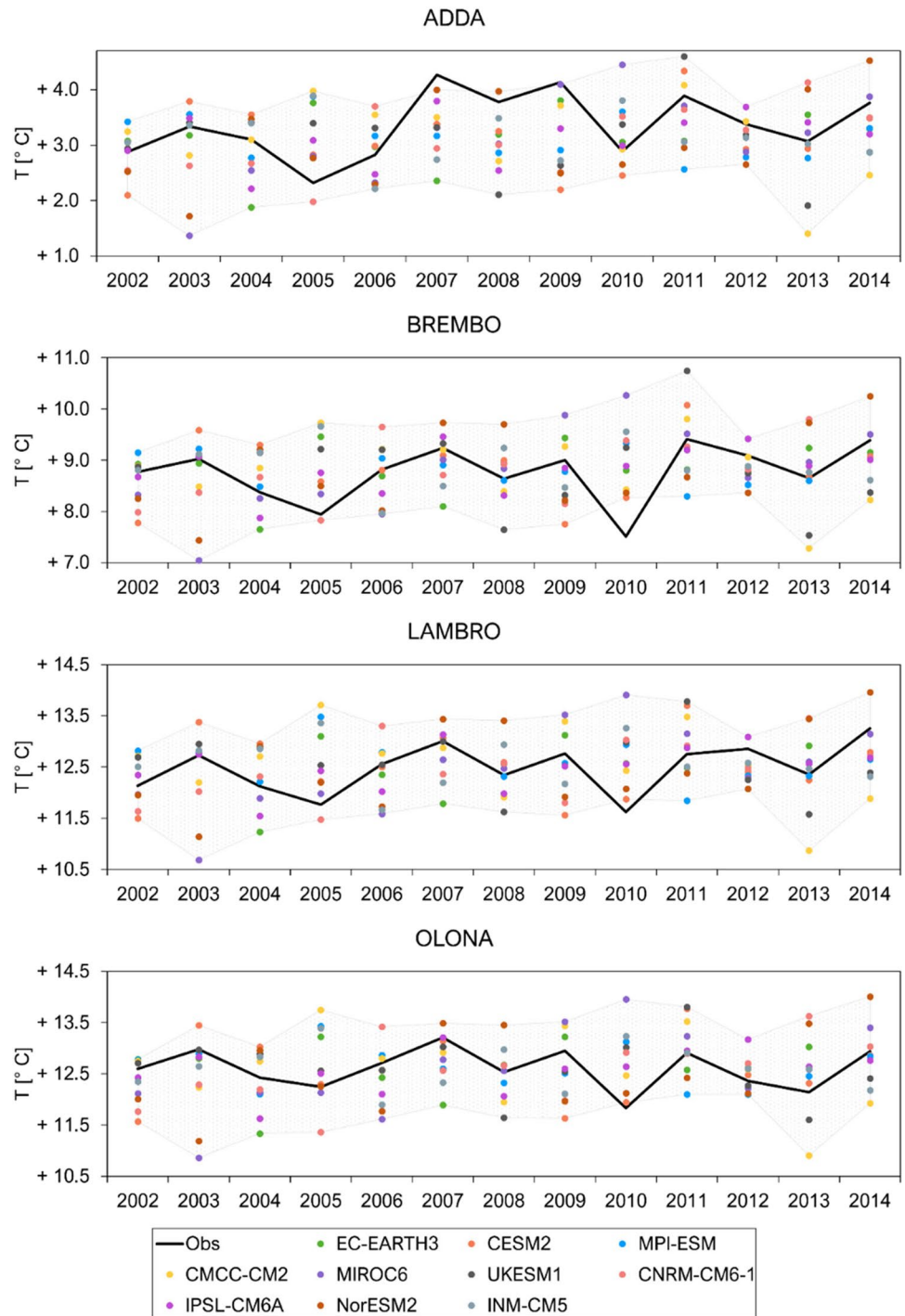
Figures 2 and 3 represent the mean annual temperature and the cumulated annual precipitation observed and modeled for each GCM and over each basin during 2002–2014.

The mean observed annual temperature in the Adda catchment during 2007–2009 is generally underestimated by all GCMs, and none of the models successfully reproduce the lowest observed temperature observed in the year 2010 across all basins. Notice, however, that GCM models, working at the global scale, are expected to represent acceptably general (mean) climate conditions over large scales while being less accurate in depicting more variable (and local) behavior. However, despite such flaws, the historical trend of the GCMs reproduces acceptably well the observed values (Fig. 2).

The observed precipitation displays a similar fluctuating pattern in all basins, with different absolute values, the Adda catchment being the less rainy one (Fig. 3). The historical trend of GCMs is smoother with respect to the observed precipitation, and both overestimation and underestimation can be observed, the former during 2004–2007 for Adda basin and during 2003–2007 for the other catchments, namely, after 2012 for Adda and Brembo and after 2013 for Lambro and Olona.

To aggregate these results, for each model, the number of years (out of 14) when the p -value was significant (i.e.,

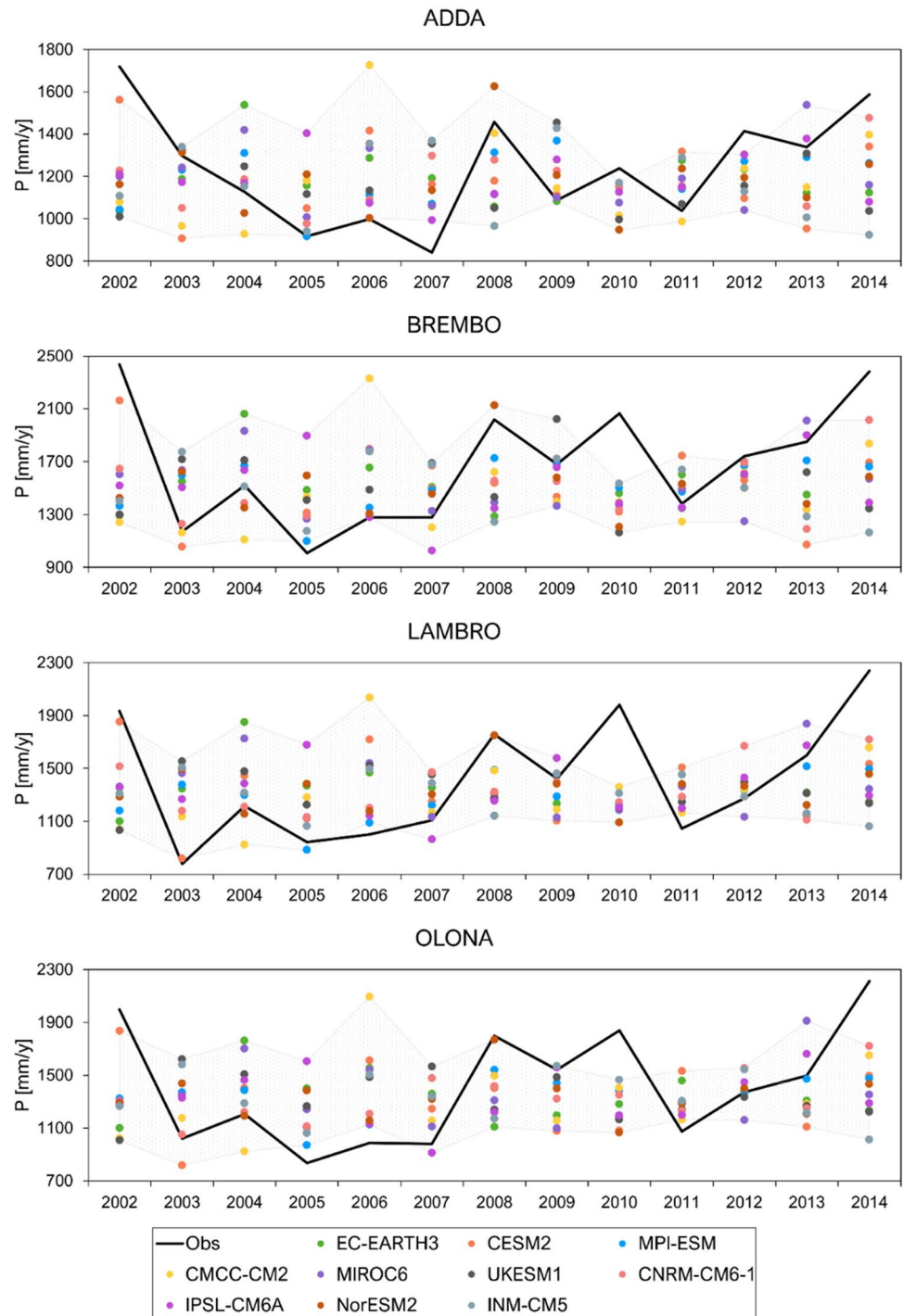
Fig. 2 Mean annual temperature over each basin from 2002 to 2014 for each GCM. The points represent the mean value over each GCM (colored), and the black line represents the mean observed temperature from AWS



p -value ≥ 0.05) was lumped (Table 2). We report here the number of years considering both the daily and monthly values. The results show that considering the daily values may result in lower p -values and so into a smaller number of years when we can consider the mean and variance of the climate variables from GCMs to be consistent with the observed values.

When considering a monthly aggregation, all GCMs present acceptable performance in reproducing the observed climate patterns in all catchments. On the other hand, by increasing the temporal resolution at a daily scale, the models show different results. For temperature overall, the best models seem to be the MPI-ESM, the MIROC6, and the IPSL. The mean value of precipitation is well represented

Fig. 3 Cumulated annual precipitation over each basin from 2002 to 2014 for each GCM. The points represent the mean value over each GCM (colored), and the black line represents the mean observed value from AWS



overall by MPI-ESM and CNRM, whereas in terms of variability, the results are less satisfactory.

Beyond average annual values, one wants to assess the changing seasonal patterns of climate variables since their seasonal behavior directly impacts hydrological response. In Table 3, the mean seasonal error between the temperature from the GCMs and the observed values is reported for each basin. For each GCM and basin, we performed 4 Student's

T-tests, one for each season for all the historical periods 2002–2014, considering both the daily and the monthly aggregation. The results are shown in Table 3. For the mean daily temperature, the results of the *T*-tests were significant for all the GCMs for all the basins, except for the Adda catchment, where all the GCMs gave poor performances in spring and summer (Table 3). This is possibly caused by many missing (*no-data*) temperature values in the AWSs of the Adda

Table 2 Number of years in which the *p*-value of the *t*-test (for *T* and *P*) and *F*-test (for *P*) is significant for each GCM

		EC-EARTH3	CESM2	MPI-ESM	CMCC-CM2	MIROC6	UKESM1	CNRM-CM6-1	IPSL-CM6A	NotESM2	INM-CM5		
T	ADDA	<i>t</i> -test	D	6	5	4	5	6	6	5	6		
		M	10	12	11	11	11	12	12	13	11		
	BREMBO	<i>t</i> -test	D	6	6	9	5	5	7	6	0	4	
		M	11	12	11	12	11	11	12	12	13	11	
	LAMBRO	<i>t</i> -test	D	6	5	7	3	4	7	8	3	5	
		M	11	12	11	11	11	12	12	13	11	11	
	OLONA	<i>t</i> -test	D	8	8	8	5	7	6	7	4	7	
		M	11	12	12	11	10	12	10	13	11	12	
	P	ADDA	<i>t</i> -test	D	7	9	12	11	8	9	8	9	
			M	10	12	13	12	13	12	12	13	11	
			<i>F</i> -test	D	4	4	5	3	5	5	3	4	3
			M	10	12	11	11	12	11	12	11	12	10
BREMBO		<i>t</i> -test	D	8	9	10	8	7	6	8	7	6	
		M	12	12	13	12	11	12	12	9	10	11	
		<i>F</i> -test	D	4	1	5	5	1	6	3	4	1	
		M	9	11	10	11	11	10	12	11	10	10	
LAMBRO		<i>t</i> -test	D	4	7	9	8	7	6	7	8	6	
		M	9	11	11	11	7	9	13	9	10	9	
		<i>F</i> -test	D	3	4	4	5	3	4	5	5	4	
		M	9	11	12	12	11	9	11	11	10	12	
OLONA	<i>t</i> -test	D	6	10	11	10	5	8	8	9	8		
	M	12	12	13	12	11	11	13	10	10	10		
	<i>F</i> -test	D	3	4	2	1	0	3	2	2	4		
	M	8	9	10	11	10	10	12	11	9	11		

Daily (D) and monthly (M) aggregation is considered

Table 3 Mean error on the mean seasonal temperature from GCMs $S_s(T_{GCM})$ against the mean seasonal observed value $S_s(T_{obs})$ over each basin for the years 2002–2014 for each GCM

		EC-EARTH3		CESM2		MPI-ESM		CMCC-CM2		MIROC6		UKESM1		CNRM-CM6-1		IPSL-CM6A		NorESM2		INM-CM5				
		D	M	D	M	D	M	D	M	D	M	D	M	D	M	D	M	D	M	D	M	D	M	
T ADDA	Winter	<i>t</i> -test	-0.0°C	Y	-0.0°C	Y	+0.1°C	Y	+0.0°C	Y	-0.1°C	Y	+0.1°C	Y	-0.0°C	Y	-0.0°C	Y	-0.1°C	Y	-0.1°C	Y	+0.0°C	
			Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Spring	<i>t</i> -test	-0.5°C	Y	-0.5°C	Y	-0.5°C	Y	-0.5°C	Y	-0.5°C	Y	-0.6°C	Y	-0.5°C	Y	-0.5°C	Y	-0.5°C	Y	-0.5°C	Y	-0.5°C	Y
			Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
BREMBO	Summer	<i>t</i> -test	-0.5°C	Y	-0.5°C	Y	-0.5°C	Y	-0.4°C	Y	-0.4°C	Y	-0.5°C	Y	-0.4°C	Y	-0.4°C	Y	-0.5°C	Y	-0.5°C	Y	-0.4°C	Y
			Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Fall	<i>t</i> -test	-0.0°C	Y	-0.0°C	Y	-0.0°C	Y	+0.0°C	Y	-0.0°C	Y	+0.0°C	Y	-0.1°C	Y	+0.0°C	Y	-0.0°C	Y	-0.0°C	Y	+0.0°C	Y
			Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
LAMBRO	Winter	<i>t</i> -test	+0.1°C	Y	+0.2°C	Y	+0.3°C	Y	+0.2°C	Y	+0.1°C	Y	+0.3°C	Y	+0.2°C	Y	+0.1°C	Y	+0.1°C	Y	+0.1°C	Y	+0.2°C	Y
			Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Spring	<i>t</i> -test	+0.0°C	Y	-0.1°C	Y	-0.0°C	Y	-0.1°C	Y	-0.0°C	Y	+0.0°C	Y	-0.0°C	Y	-0.1°C	Y	-0.1°C	Y	-0.1°C	Y	+0.0°C	Y
			Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
T ADDA	Summer	<i>t</i> -test	+0.1°C	Y	+0.1°C	Y	+0.2°C	Y	+0.2°C	Y	+0.1°C	Y	+0.2°C	Y	+0.1°C	Y	+0.2°C	Y	+0.1°C	Y	+0.1°C	Y	+0.1°C	Y
			Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Fall	<i>t</i> -test	-0.1°C	Y	-0.1°C	Y	-0.0°C	Y	-0.1°C	Y	-0.0°C	Y	+0.0°C	Y	-0.0°C	Y	-0.1°C	Y	-0.1°C	Y	-0.1°C	Y	+0.1°C	Y
			Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Table 3 (continued)

		EC-EARTH3		CESM2		MPI-ESM		CMCC-CM2		MIROC6		UKESM1		CNRM-CM6-1		IPSL-CM6A		NorESM2		INM-CM5					
		M	D	M	D	M	D	M	D	M	D	M	D	M	D	M	D	M	D	M	D	M	D		
OLONA	Winter	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		
		<i>t</i> -test	-0.0°C	-0.0°C	+0.0°C	+0.0°C	-0.1°C	-0.1°C	+0.1°C	+0.1°C	-0.1°C	-0.1°C	-0.1°C	-0.1°C	-0.1°C	-0.1°C	-0.1°C	-0.1°C	-0.1°C	-0.1°C	-0.1°C	-0.1°C	-0.1°C	+0.0°C	
		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
	Spring	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
		<i>t</i> -test	+0.0°C	-0.0°C	+0.0°C	-0.1°C	-0.0°C	-0.0°C	-0.0°C	-0.0°C	-0.1°C	-0.1°C	-0.1°C	-0.1°C	-0.1°C	-0.1°C	-0.1°C	-0.1°C	-0.0°C	-0.0°C	-0.0°C	-0.0°C	-0.1°C	-0.1°C	-0.1°C
		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Summer	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
		<i>t</i> -test	+0.1°C	+0.0°C	+0.0°C	+0.0°C	-0.0°C	-0.0°C	-0.0°C	-0.0°C	+0.0°C	+0.0°C	+0.0°C	+0.0°C	+0.0°C	+0.0°C	+0.0°C	+0.0°C	-0.0°C	-0.0°C	-0.0°C	-0.0°C	+0.1°C	+0.1°C	+0.1°C
		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Fall	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
		<i>t</i> -test	+0.0°C	+0.0°C	+0.1°C	+0.1°C	-0.1°C	-0.1°C	+0.0°C	+0.0°C	+0.0°C	+0.0°C	+0.0°C	+0.0°C	+0.0°C	+0.0°C	+0.0°C	+0.0°C	-0.1°C	-0.1°C	-0.1°C	-0.1°C	+0.0°C	+0.0°C	+0.0°C
		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Student's *t*-test is used, and the significant values are in bold. Results of *t*-test for temperature for each GCM. Daily (D) and monthly (M) aggregation is considered. Significant tests are marked with Y ($\alpha = 5\%$)

basin in the summer and spring months, which were not filled. However, considering the mean monthly temperature, such an issue is masked since the use of the monthly scale seemingly hides the issue of the daily *no-data* records.

For precipitation, Student's *T*-tests and *F*-tests were carried out for each season for the years 2002–2014, and the mean error for the cumulated seasonal precipitation from GCMs against the cumulated seasonal observed values is reported in Table 4. The GCMs depict well the mean seasonal variability of precipitation, but they fail to represent its changes as quantified by variance, and the noise therein is the highest. By looking at Table 4, *T*-tests are significant for all the basins, considering both the daily and the monthly aggregation. *F*-tests, contrarily, are not significant for the daily samples but improve at the monthly scale. However, in spring and autumn, the GCMs still poorly represent the variance of precipitation. This occurs mostly in the rainiest months when the intrinsic variability of the precipitation pattern is higher, thus making it more challenging its representation.

Overall, the best GCMs in representing the seasonal variability of our target climate variables are the CESM2, CMCC, UKESM2, and IPSL.

4.1.2 Hydrological validation

The uncertainty associated with hydrological modeling was also detected, and the benchmarking of hydrological discharges is reported in Fig. 4. We compared the projected discharges from each GCM against the discharges modeled by *Poli-Hydro* using the observed AWS weather data.

Since the main driver of the hydrological model (and of the hydrological cycle) is precipitation, the annual discharge trend is correlated with precipitation patterns. Thus, GCM discharges overestimated AWS discharges during 2003/2004–2007 and during the most recent years of the historical period.

In Table 5, the mean errors in hydrological modeling using the GCMs for the 4 decades, i.e., 2002–2011, 2003–2012, 2004–2013, and 2005–2014, are assessed. The lowest errors are found in the Adda catchment, while the highest are in the Brembo basin. Overall, the errors are quite small in all decades, except for the most recent one, since in 2014, precipitation was higher than the average value, resulting in very high modeled discharges when using AWS data (Figs. 4 and 5). The significance of the *t*-tests is investigated in Table 5, where the results are like those previously shown for the climate variables. By reducing the sample size, i.e., from daily to monthly aggregation, the *p*-values increase.

However, it is interesting that small mean errors in the order of 2–3% are considered large in the tests when using daily samples, likely representing an effect of the large

Table 4 Mean error on the cumulated seasonal precipitation from GCMs S_i (PGCM) against the cumulated seasonal observed value S_i (Pobs) for the years 2002–2014 for each GCM

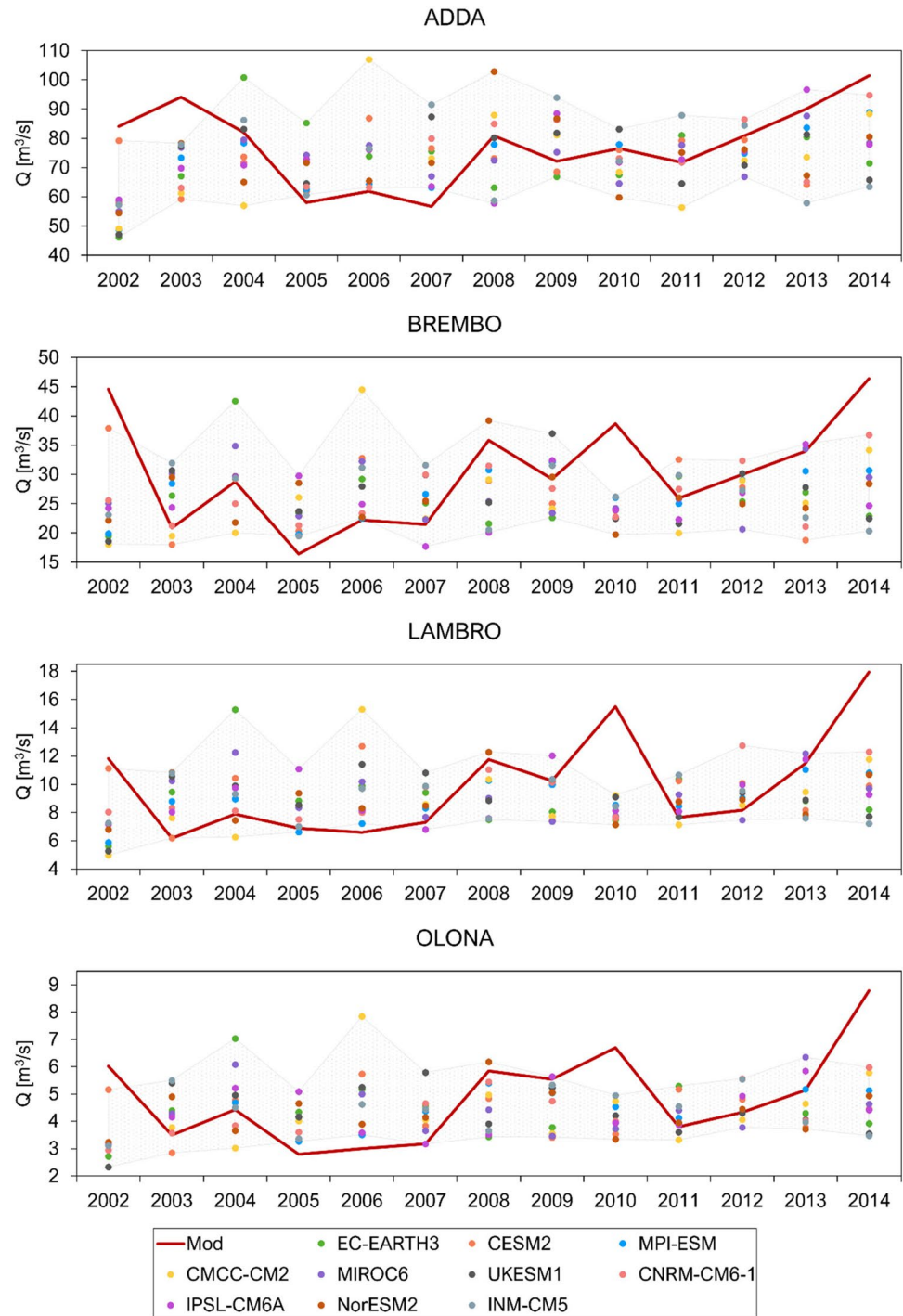
P	ADDA	Winter	EC-EARTH3		CESM2		MP-ESM		CMCC-CM2		MIROC6		UKESM1		CNRM-CM6-1		IPSL-CM6A		NorESM2		INM-CM5	
			D	M	D	M	D	M	D	M	D	M	D	M	D	M	D	M	D	M	D	M
		<i>t</i> -test	-4.90%	Y	-5.10%	Y	-4.50%	Y	-5.30%	Y	-6.00%	Y	-6.00%	Y	-4.30%	Y	-4.00%	Y	-3.70%	Y	-4.10%	Y
		<i>F</i> -test	-17.90%	-	-29.20%	-	-14.80%	-	-21.70%	-	-30.70%	-	-23.80%	-	-31.20%	-	-31.20%	-	-34.10%	-	-20.90%	-
		<i>t</i> -test	-3.20%	Y	-2.40%	Y	-2.20%	Y	-4.60%	Y	-3.40%	Y	-4.70%	Y	-3.10%	Y	-2.50%	Y	-3.60%	Y	-6.20%	Y
		<i>F</i> -test	-30.40%	Y	-26.80%	Y	-17.30%	Y	-27.70%	Y	-26.60%	Y	-21.60%	Y	-27.50%	Y	-34.10%	Y	-17.90%	Y	-19.10%	Y
		<i>t</i> -test	-1.90%	Y	-3.30%	Y	-2.30%	Y	-4.80%	Y	-2.00%	Y	-3.40%	Y	-3.00%	Y	-1.50%	Y	-2.80%	Y	-6.90%	Y
		<i>F</i> -test	-18.80%	Y	-14.80%	Y	-8.10%	Y	-14.10%	Y	-19.40%	Y	-10.90%	Y	-25.90%	Y	-34.00%	Y	-15.60%	Y	-13.10%	Y
		<i>t</i> -test	-6.10%	Y	-5.90%	Y	-5.50%	Y	-5.10%	Y	-4.20%	Y	-6.50%	Y	-4.60%	Y	-7.30%	Y	-5.80%	Y	-4.30%	Y
		<i>F</i> -test	-8.90%	Y	-10.60%	Y	-16.70%	Y	-6.90%	Y	-25.20%	Y	-13.90%	Y	-28.40%	Y	-33.80%	Y	-23.00%	Y	-8.10%	Y
		<i>t</i> -test	-4.50%	Y	-7.00%	Y	-3.40%	Y	-14.00%	Y	-9.10%	Y	-4.20%	Y	-4.80%	Y	-2.50%	Y	-1.60%	Y	-1.40%	Y
		<i>F</i> -test	-7.20%	Y	-5.00%	Y	-8.10%	Y	2.90%	Y	-14.50%	Y	4.80%	Y	-2.00%	Y	-1.70%	Y	-24.60%	Y	-6.90%	Y
		<i>t</i> -test	-9.10%	Y	-8.10%	Y	-7.60%	Y	-13.40%	Y	-10.00%	Y	-9.10%	Y	-12.20%	Y	-9.10%	Y	-11.70%	Y	-11.40%	Y
		<i>F</i> -test	-32.90%	Y	-21.80%	Y	-18.90%	Y	-20.70%	Y	-23.50%	Y	-19.20%	Y	-14.40%	Y	-30.20%	Y	-22.10%	Y	-26.90%	Y
		<i>t</i> -test	-8.60%	Y	-6.20%	Y	-10.80%	Y	-13.90%	Y	-5.80%	Y	-4.00%	Y	-14.10%	Y	-5.90%	Y	-9.80%	Y	-16.40%	Y
		<i>F</i> -test	0.10%	Y	14.20%	Y	-3.50%	Y	8.80%	Y	-0.30%	Y	22.40%	Y	8.60%	Y	-6.20%	Y	4.20%	Y	0.30%	Y
		<i>t</i> -test	-13.50%	Y	-12.90%	Y	-9.20%	Y	-12.10%	Y	-8.40%	Y	-13.90%	Y	-10.20%	Y	-19.40%	Y	-11.60%	Y	-10.10%	Y
		<i>F</i> -test	-16.70%	Y	-8.90%	Y	-22.10%	Y	-5.50%	Y	-25.60%	Y	-15.90%	Y	-16.80%	Y	-35.50%	Y	-30.00%	Y	-18.60%	Y

Table 4 (continued)

	EC-EARTH3		CESM2		MPL-ESM		CMCC-CM2		MIROC6		UKESM1		CNRM-CM6-1		IPSL-CM6A		NorESM2		INM-CM5				
	D	M	D	M	D	M	D	M	D	M	D	M	D	M	D	M	D	M	D	M			
LAMBRO	Winter	<i>r</i> -test	-3.00%	-2.20%	-8.00%	-9.80%	-3.90%	-2.20%	-1.10%	-2.20%	-2.20%	-1.10%	-1.10%	-1.10%	-1.10%	0.00%	-1.80%						
		<i>F</i> -test	-19.00%	-9.90%	-16.30%	-16.40%	-25.50%	-14.10%	-14.10%	-13.40%	-14.10%	-13.00%	-13.00%	-13.00%	-13.00%	-13.00%	-25.60%	-15.70%					
	Spring	<i>r</i> -test	-5.50%	-6.00%	-4.00%	-4.10%	-1.40%	-4.10%	-1.40%	-3.70%	-3.70%	-3.70%	-5.50%	-5.50%	-5.50%	-6.50%	-8.70%						
		<i>F</i> -test	-21.20%	-15.00%	-11.80%	-10.20%	-15.60%	-10.20%	-15.60%	-8.80%	-8.80%	-8.80%	-5.70%	-5.70%	-13.80%	-14.90%	-18.70%						
	Summer	<i>r</i> -test	-3.30%	-3.40%	-14.00%	-4.30%	-2.10%	-4.30%	-2.10%	-5.40%	-5.40%	-5.40%	-4.40%	-4.40%	-3.70%	-3.40%	-9.30%						
		<i>F</i> -test	-14.40%	-8.30%	-25.30%	-6.40%	-20.80%	-6.40%	-20.80%	-6.30%	-6.30%	-6.30%	-10.50%	-10.50%	-16.20%	-9.00%	-23.30%						
	Fall	<i>r</i> -test	-5.80%	-9.00%	-9.00%	-9.10%	-7.20%	-9.10%	-7.20%	-7.50%	-7.50%	-7.50%	-5.90%	-5.90%	-8.70%	-6.70%	-7.30%						
		<i>F</i> -test	-7.60%	1.10%	-21.00%	-5.00%	-14.40%	-5.00%	-14.40%	-0.60%	-0.60%	-0.60%	-8.10%	-8.10%	-12.70%	-18.50%	-10.60%						
	OLONA	Winter	<i>r</i> -test	3.90%	-0.80%	4.10%	-3.10%	0.40%	-3.10%	4.10%	2.60%	2.60%	3.10%	3.10%	3.50%	3.70%	4.20%						
			<i>F</i> -test	-19.20%	-18.20%	-21.90%	-16.20%	-32.20%	-16.20%	-32.20%	-15.50%	-15.50%	-15.50%	-23.20%	-23.20%	-15.70%	-22.60%	-19.70%					
		Spring	<i>r</i> -test	-4.80%	-6.20%	-3.00%	-4.20%	-2.40%	-4.20%	-2.40%	-3.80%	-3.80%	-3.80%	-7.50%	-7.50%	-3.80%	-6.40%	-7.10%					
			<i>F</i> -test	-28.60%	-26.20%	-23.70%	-23.10%	-21.10%	-23.10%	-21.10%	-25.40%	-25.40%	-25.40%	-21.00%	-21.00%	-23.20%	-20.10%	-23.20%					
Summer		<i>r</i> -test	-7.50%	-7.30%	-7.30%	-8.00%	-6.30%	-8.00%	-6.30%	-7.30%	-7.30%	-7.30%	-7.40%	-7.40%	-4.60%	-7.70%	-9.90%						
		<i>F</i> -test	-21.30%	-14.70%	-23.60%	-11.70%	-23.00%	-11.70%	-23.00%	-8.90%	-8.90%	-8.90%	-17.80%	-17.80%	-20.60%	-17.00%	-25.10%						
Fall		<i>r</i> -test	-7.40%	-10.50%	-9.20%	-9.70%	-7.70%	-9.70%	-7.70%	-8.70%	-8.70%	-8.70%	-8.80%	-8.80%	-13.00%	-8.00%	-6.10%						
		<i>F</i> -test	-16.80%	-25.30%	-32.90%	-17.20%	-26.40%	-17.20%	-26.40%	-17.60%	-17.60%	-17.60%	-26.30%	-26.30%	-33.50%	-32.10%	-19.50%						

Student's *t*-test and Fisher's *F*-test are used, and the significant values are in bold. Results of *t*-tests and *F*-test for precipitation for each GCM. Daily (D) and monthly (M) aggregation is considered. Significant tests are marked with Y ($\alpha = 5\%$)

Fig. 4 Mean annual discharge over each basin from 2002 to 2014 for each GCM. The points represent the mean value over each GCM (colored), and the red line represents the mean modeled values from AWS



sample size (3652 ca. values). To extend the analysis, we combined the results of all t -tests of the basins by representing the relation between p -values and the mean errors (Fig. 5). The graph shows that the acceptable limit here for p -value (≥ 0.05) does not visibly identify a threshold in the mean error, and even p -values of the order of 2% gave relatively small errors. The choice of p -values < 0.05 (for test

failure) is quite customary, and it was taken here as a reference. However, Fig. 5 seems to demonstrate that, below the p -value ≤ 0.01 or so, considerably higher noise would be seen. Accordingly, one may hypothesize that a more discriminating power (i.e., to highlight projections displaying large inaccuracy) may be obtained by setting a threshold of p -value ≤ 0.01 . Such a topic may deserve further verification.

Table 5 Mean error on the yearly discharge from GCMs (Q_{GCM}) against its modeled value from AWS (Q_{obs}) for each moving window and GCM. Student's *t*-test is used, and the significant values are in bold. Results of *t*-test for discharge for each GCM. Daily (D) and monthly (M) aggregation is considered. Significant tests are marked with Y ($\alpha = 5\%$)

Q	EC-EARTH3	CESM2		MPI-ESM		CMCC-CM2		MIROC6		UKESM1		CNRM-CM6-1		IPSL-CM6A		NorESM2		INM-CM5			
		D	M	D	M	D	M	D	M	D	M	D	M	D	M	D	M	D	M	D	M
Q ADDA	2002–2011	-1.40%		-0.40%		-3.30%		-2.40%		1.10%		-3.00%		-4.70%		-0.90%		3.40%			
		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	2003–2012	3.10%		0.00%		0.30%		-0.30%		4.70%		1.30%		-2.00%		2.50%		7.60%			
		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	2004–2013	5.40%		1.20%		2.50%		1.70%		5.90%		2.20%		2.20%		1.50%		5.40%			
BREMBO	2005–2014	-1.20%		-0.70%		4.00%		-1.10%		0.80%		2.60%		0.50%		1.00%		-0.40%			
		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	2002–2011	-7.00%		-3.30%		-12.60%		-6.10%		-6.00%		-9.80%		-12.10%		-6.70%		-3.20%			
		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	2003–2012	0.30%		-1.90%		-3.80%		-2.70%		3.40%		-2.40%		-6.40%		-0.60%		3.60%			
LAMBRO	2004–2013	-4.20%		-6.20%		-6.30%		-5.70%		-2.40%		-7.00%		-7.00%		-7.10%		-4.40%			
		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	2005–2014	-16.40%		-11.90%		-7.10%		-13.00%		-10.50%		-8.60%		-14.00%		-10.30%		-13.00%			
		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	2002–2011	0.10%		-1.90%		-6.70%		-2.70%		0.50%		-4.60%		-5.40%		-2.30%		-0.90%			
OLONA	2003–2012	8.10%		1.00%		1.20%		1.70%		9.40%		4.70%		1.70%		4.20%		5.80%			
		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	2004–2013	1.30%		-2.60%		-2.60%		-2.10%		1.40%		-2.00%		-0.10%		-4.90%		-3.60%			
		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	2005–2014	-15.40%		-12.60%		-6.70%		-14.00%		-10.60%		-7.40%		-10.30%		-11.00%		-15.00%			
	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	

Table 5 (continued)

	EC-EARTH3		CESM2		MPI-ESM		CMCC-CM2		MIROC6		UKESM1		CNRM-CM6-1		IPSL-CM6A		NorESM2		INM-CM5	
	M	D	M	D	M	D	M	D	M	D	M	D	M	D	M	D	M	D	M	D
2004–2013	2.90%	-3.50%	Y	Y	-0.50%	-1.20%	Y	Y	0.70%	1.60%	-2.30%	0.10%	-3.80%	0.40%	Y	Y	Y	Y	Y	Y
2005–2014	-12.60%	-12.70%	Y	Y	-8.40%	-4.30%	Y	Y	-11.20%	-10.20%	-6.60%	-10.40%	-9.70%	-10.60%	Y	Y	Y	Y	Y	Y
			Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

4.2 Projected scenarios

We used the GCMs previously tested to project forward climate variables during 2015–2021 to further assess the goodness of the different SSP scenario projections. We calculated the average value and confidence limit for the 10 GCMs for each SSP, and we evaluated whether the observed data of climatic variables and modeled data of hydrological variables were within the confidence intervals of the GCM projections. We represented the mean observed values and the mean value averaged on the 10 GCMs for each SSP with their interval of confidence at 95%.

Supplementary Figs. A3 and A4 represent the average annual temperature and the cumulated annual precipitation over each of the four basins, calculated as the mean value for each SSP, averaged on the GCMs vs. the mean values assessed from ARPA weather stations. The results show that the mean observed temperature in all basins is well represented by the average value of GCMs for each SSP scenario. The observed precipitation presents a similar fluctuating pattern in all the basins with different absolute values, with the Adda catchment being the less rainy catchment. Similarly to temperature, the observed precipitation falls within the interval confidence of each SSP, but with a slight difference between the catchments and the year due to the erratic behavior of precipitation.

The uncertainty associated with the hydrological projections is reported in Supplementary Fig. A5. We compare the modeled projected discharges for each GCM with the modeled discharges with *Poli-Hydro* using the observed AWS weather data. Since the main driver of the hydrological model is clearly precipitation (Casale et al. 2021), the annual discharge trend is somewhat correlated with the precipitation pattern presented in each basin, and obviously, the lowest, least snowy watersheds display a stricter correlation.

5 Discussion

In this study, we provided a validation of the ability of GCMs to capture local climate and hydrological variability. Although several studies have attempted to evaluate the goodness of GCMs in representing climate change at a regional scale (Ruane and McDermid 2017; Prein et al. 2019), there are still concerns on how to classify whether a model is adequate or not (Knutti and Knutti 2010; Palmer et al. 2022). We used here an objective statistical approach to validate the control data series of GCMs (Groppelli et al. 2011a). The results of the statistical tests are aligned with previous studies, where a large number of values may result in considerably low *p-values*, leading, therefore, to possibly less statistically significant results. Indeed, by reducing the sample size, i.e., with monthly aggregation, most of the tests

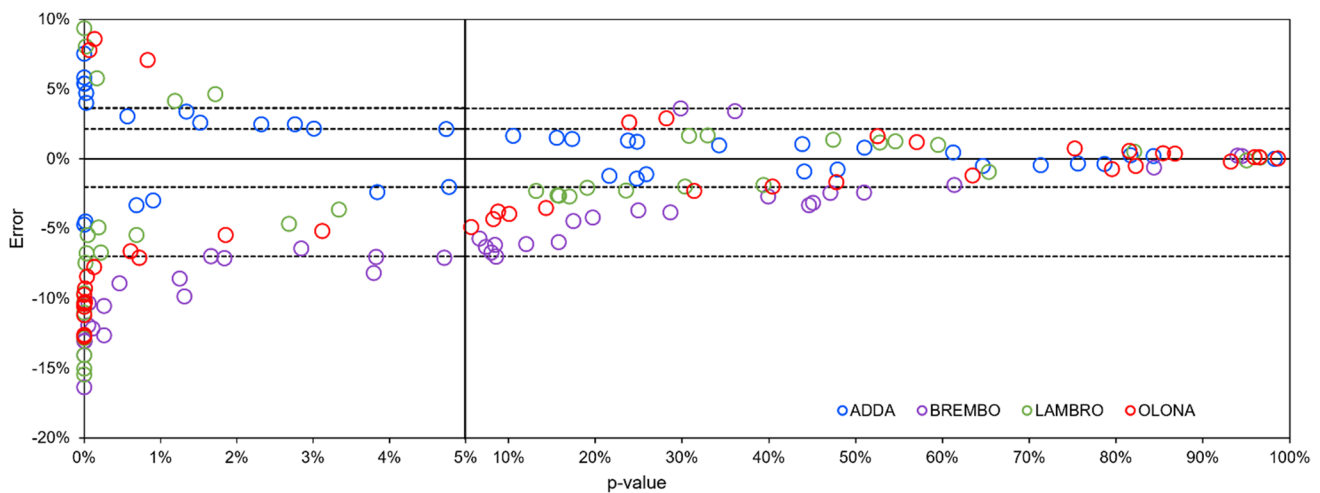


Fig. 5 Mean annual discharges. Representation of p -value vs. mean error, calculated for the t -tests carried out with a daily resolution for each combination of GCM and year. The results for each basin are highlighted with different colors

gave acceptable results. This seems legitimate, given the small noise that was generally found here. For climate variables, this is true more in terms of average value, whereas the variance of the precipitation is harder to reproduce.

Compared to other recent studies (Eyring et al. 2019; Brunner et al. 2019), we tried to take a step forward from the analysis of models' climatic outputs to also validate hydrological projections, widely used in future water resource management (Ravazzani et al. 2016; Gianinetto et al. 2017; Bombelli et al. 2019; Stucchi et al. 2019; Aili et al. 2019).

Our results show that the goodness of GCMs in representing the local climate variability will not always reflect in a good representation of hydrology variables of the catchment area (using the same hydrological model as here). The analysis of climate patterns may give poor results, i.e., with a poor capacity of representing the water cycle of an area. When assessing hydrological projections, a seasonal validation of climate patterns is crucial since changes in the seasonality of precipitation have a paramount effect on stream flow discharge (Christensen and Lettenmaier 2007). Our results, both in terms of seasonality of climate patterns and hydrological projections, showed that the GCMs outputs are consistent with the observed data.

One may discuss whether the forcing scenarios adopted here in the control period (2002–2014) were coherent with actual environmental conditions as observed therein, which may (partly) explain the observed differences in Figs. 2 and 3. The GCM projections are, in practice, based on assumptions of GHG concentration (normally a bulk value valid overall on the earth's surface), evolving in time according to the hypothesized SSP scenarios. Considering that Figs. 2 and 3 refer to the recent past, the GCMs most likely assumed levels of GHGs fairly equivalent to those now (e.g., for CO_2 , ca. 400 ppm, and so on for other gasses). Accordingly, one

may assume that the reference (environmental) conditions as set out within the different SSPs would overlap in practice these values in the control period (however, with more and more different values moving toward the end of the century).

Accordingly, such an aspect would not largely affect the results of the simulation if not for the model's inherent noise.

To broaden our knowledge of the uncertainty range, more hydrological models could be employed (Krysanova et al. 2020). However, a recent study in the same area demonstrated that uncertainty in flow projection is likely to be affected more by climate projections and internal variability than by the choice of the hydrological model (Casale et al. 2021). This result is coherent with other research that identifies the largest source of uncertainty in the short term, the choice of a given GCM rather than the choice of any hydrological models (Aryal et al. 2019).

The GCMs provide global projections; thus, to determine their validity in representing local climate, a downscaling procedure is most often needed. We used here a statistical downscaling method widely used in this area with acceptable results (Aili et al. 2019; Fuso et al. 2021; Bombelli et al. 2021). The methodology proposed here is space-independent, but the main findings are strictly correlated to the study area, underlying the importance of a local analysis in the evaluation of climate change effects. In the future, more downscaling procedures may be tested.

6 Conclusion

The study here presented demonstrates that the climate projections provided by models within the CMIP6 of the IPCC are (i) acceptably consistent, spatially, and temporally with observed data collected from AWSs in the study area of the

Lombardy region and (ii) amenable to provide acceptable hydrological conjectures in the area.

Our comparison here was carried out for a period of 20 years, where 14 years covered a control period (it is therefore assessed the reliability of the GCM against historical observations available already in the simulation phase), and 7 covered a projection period (thus, *ex-post* validation was pursued). In both cases, the reliability of the GCMs (all subjected to an equivalent downscaling process) was evaluated. Subsequently, to introduce an additional measure of uncertainty and to also validate the results from a hydrological point of view, stream flows were also compared. The elaborations carried out aimed at assessing the general consistency of the GCM climate and hydrological scenarios on a wide spatial scale. It is clear that these scenarios represent projections valid in a purely statistical sense, which are useful to provide an array of potential conditions for planning and management at a region/basin scale. Moreover, the projection horizon considered here, of 7 years, is particularly short. Therefore, a statistical validation such as the one investigated here can be carried out in the years to come as new observed data become available. Given, however, the evidently very difficult (if even possible) task of projecting the future of climate and hydrology acceptably beyond short (e.g., seasonal) time scales, we provide here at least an indication that GCMs outputs, whatsoever noisy, could be at least considered acceptably accurate in a broad average sense and used for (whatsoever uncertain) water policy-making, which hitherto was rarely (if ever) demonstrated.

One may wonder whether outputs of CMPI6 models, as a new generation of GCMs, may/will bring improvements in model adequacy when using stochastic downscaling. Such a question may point seemingly to a specific comparison between CMIP5 (and former) and CMIP6 models. This was not the scope of our work, and indeed, we focused on CMIP6, given that, likely, in the future, most projection studies will be based on such models.

In the personal experience of the authors that have worked in the recent past in the field of hydrological modeling and projections using outputs from CMIP4-5-6, the latter GCM outputs from CMIP6 seemingly provide an improving depiction of climate and especially of precipitation cycle (likely the most critical, and yet important variable when dealing with hydrological conjectures), in terms of amount and timing. However, such qualitative perception is not backed up by specific testing.

Notice that stochastic downscaling provides a certain (noticeable) degree of (mathematical) manipulation of the GCM data, especially in order to provide a proper (time/space) variability of the precipitation process. Accordingly, after such modifications, the initial (GCM) data's accuracy/credibility may be less relevant in terms of the final result. Such a topic seems worth some investigation in the future.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00704-023-04698-5>.

Author contributions All authors contributed to the study's conception and design. Material preparation, data collection, and analysis were performed by Flavia Fuso and Giovanni Martino Bombelli. The first draft of the manuscript was written by Flavia Fuso, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by Politecnico di Milano within the CRUI-CARE Agreement.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Code availability The code generated during and/or analyzed during the current study is available from the corresponding author upon reasonable request.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aili T, Soncini A, Bianchi A et al (2019) Assessing water resources under climate change in high-altitude catchments: a methodology and an application in the Italian Alps. *Theor Appl Climatol* 135:135–156. <https://doi.org/10.1007/s00704-017-2366-4>
- Akbari H, Soncini A, Dinpashoh Y, Fakheri-Fard A, Bocchiola D (2018). Operation of two major reservoirs of Iran under IPCC scenarios during XXI Century. *Hydrolog Process* 32:3254–3271
- Anghileri D, Pianosi F, Soncini-Sessa R (2011) A framework for the quantitative assessment of climate change impacts on water-related activities at the basin scale. *Hydro Earth Syst Sci* 15:2025–2038. <https://doi.org/10.5194/hess-15-2025-2011>
- Aryal A, Shrestha S, Babel MS (2019) Quantifying the sources of uncertainty in an ensemble of hydrological climate-impact projections. *Theor Appl Climatol* 135:193–209. <https://doi.org/10.1007/s00704-017-2359-3>
- Baldasso V, Soncini A, Azzoni RS, Diolaiuti G, Bocchiola D (2019) Recent evolution of glaciers in Western Asia in response to global warming: the case study of Mount Ararat, Turkey. *Theor Appl Climatol* 137:1–15
- Beniston M, Stoffel M, Hill M (2011) Impacts of climatic change on water and natural hazards in the Alps: can current water governance cope with future challenges? Examples from the European

- “ACQWA” project. *Environ Sci Policy* 14:734–743. <https://doi.org/10.1016/j.envsci.2010.12.009>
- Bocchiola D (2014) Long term (1921–2011) changes of Alpine catchments regime in Northern Italy. *Adv Water Resour* 70:51–64
- Bocchiola D, Nana E, Soncini A (2013) Impact of climate change scenarios on crop yield and water footprint of maize in the Po valley of Italy. *Agric Water Manag* 116:50–61. <https://doi.org/10.1016/j.agwat.2012.10.009>
- Bocchiola D, Rosso R (2006) The use of scale recursive estimation for short term Quantitative Precipitation Forecast. *Phys Chem Earth* 31(18):1228–1239
- Bocchiola D, Rosso R (2007) The distribution of daily Snow Water Equivalent in the Central Italian Alps. *Adv Water Resour* 30:135–147
- Bocchiola D, Soncini A, Senese A, Diolaiuti G (2018) Modelling hydrological components of the Rio Maipo of Chile, and their prospective evolution under climate change. *Climate* 6(3):57
- Bombelli GM, Soncini A, Bianchi A, Bocchiola D (2019) Potentially modified hydropower production under climate change in the Italian Alps. *Hydrol Process* 33:2355–2372. <https://doi.org/10.1002/hyp.13473>
- Bombelli GM, Tomiet S, Bianchi A, Bocchiola D (2021) Impact of prospective climate change scenarios upon hydropower potential of Ethiopia in GERD and GIBE dams. *Water* 13(5):716. <https://doi.org/10.3390/W13050716>
- Bozza A, Durand A, Confortola G, Allenbach B, Bocchiola D (2016) Potential of remote sensing and open street data for flood mapping in poorly gauged areas: a case study in Gonaives, Haiti. *Appl Geomat* 8:117–131
- Brunner L, Lorenz R, Zumwald M, Knutti R (2019) Quantifying uncertainty in European climate projections using combined performance-independence weighting. *Environ Res Lett* 14:124010. <https://doi.org/10.1088/1748-9326/AB492F>
- Busch T, Judick L (2021) Climate change—that is not real! A comparative analysis of climate-sceptic think tanks in the USA and Germany. *Clim Change* 164:18. <https://doi.org/10.1007/s10584-021-02962-z>
- Camicci S, Brocca L, Moramarco T (2017) Accuracy versus variability of climate projections for flood assessment in central Italy. *Clim Change* 141:273–286. <https://doi.org/10.1007/s10584-016-1876-x>
- Capolongo D, Refice A, Bocchiola D, D’Addabbo A, Vouvalidis K, Soncini A, Stamatopoulos L (2019) Coupling multitemporal remote sensing with geomorphology and hydrological modeling for post flood recovery in the Strymonas dammed river basin (Greece). *Sci Total Environ* 651(2):1958–1968
- Casale F, Bombelli GM, Monti R, Bocchiola D (2020) Hydro-power potential in the Kabul river of Afghanistan in the era of climate change. *Theor Appl Climatol* 139:1415–1434
- Casale F, Fuso F, Giuliani M et al (2021) Exploring future vulnerabilities of subalpine Italian regulated lakes under different climate scenarios: bottom-up vs top-down and CMIP5 vs CMIP6. *J Hydrol Reg Stud* 38:100973. <https://doi.org/10.1016/j.ejrh.2021.100973>
- Cherchi A, Fogli PG, Lovato T et al (2018) Global mean climate and main patterns of variability in the CMCC-CM2 coupled model. *J Adv Model Earth Syst* 11:2018MS001369. <https://doi.org/10.1029/2018MS001369>
- Christensen NS, Lettenmaier DP (2007) A multimodel ensemble approach to assessment of climate change impacts on the hydrology and water resources of the Colorado River Basin. *Hydrol Earth Syst Sci* 11:1417–1434. <https://doi.org/10.5194/HESS-11-1417-2007>
- Curry JA, Webster PJ (2011) Climate science and the uncertainty monster. *Bull Am Meteorol Soc* 92:1667–1682. <https://doi.org/10.1175/2011BAMS3139.1>
- Dakhlaoui H, Djebbi K (2021) Evaluating the impact of rainfall–runoff model structural uncertainty on the hydrological rating of regional climate model simulations. *J Water Clim Change* 12(8):3820–3838. <https://doi.org/10.2166/wcc.2021.004>
- Denaro S, Castelletti A, Giuliani M, Characklis GW (2018) Fostering cooperation in power asymmetrical water systems by the use of direct release rules and index-based insurance schemes. *Adv Water Resour* 115:301–314. <https://doi.org/10.1016/j.advwatres.2017.09.021>
- Döscher R, Acosta M, Alessandri A et al (2022) The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6. *Geosci Model Dev* 15:2973–3020. <https://doi.org/10.5194/gmd-15-2973-2022>
- Eyring V, Bony S, Meehl GA et al (2016) Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci Model Dev* 9:1937–1958. <https://doi.org/10.5194/GMD-9-1937-2016>
- Eyring V, Cox PM, Flato GM et al (2019) Taking climate model evaluation to the next level. *Nat Clim Chang* 9:102–110. <https://doi.org/10.1038/s41558-018-0355-y>
- Fuso F, Casale F, Giudici F, Bocchiola D (2021) Future hydrology of the cryospheric driven lake como catchment in Italy under climate change scenarios. *Climate* 9:1–24. <https://doi.org/10.3390/cli9010008>
- Fuso F, Stucchi L, Bonacina L, Fornaroli R, Bocchiola D (2023) Evaluation of water temperature under changing climate and its effect on river habitat in a regulated Alpine catchment. *J Hydrol* 616:128816
- Gianinetto M, Polinelli F, Frassy F et al (2017) Analysis of changes in crop farming in the Dudh Koshi (Nepal) driven by climate changes. *Proceedings of SPIE - The International Society for Optical Engineering* 10428:64–71. <https://doi.org/10.1117/12.2278637>
- Groppelli B, Bocchiola D, Rosso R (2011a) Spatial downscaling of precipitation from GCMs for climate change projections using random cascades: a case study in Italy. *Water Resour Res* 47:3519. <https://doi.org/10.1029/2010WR009437>
- Groppelli B, Soncini A, Bocchiola D, Rosso R (2011b) Evaluation of future hydrological cycle under climate change scenarios in a mesoscale Alpine watershed of Italy. *Nat Hazards Earth Syst Sci* 11:1769–1785. <https://doi.org/10.5194/nhess-11-1769-2011>
- Guillemot H (2010) Connections between simulations and observation in climate computer modeling. Scientist’s practices and “bottom-up epistemology” lessons. *Stud Hist Philos Sci Part B - Stud Hist Philos Mod Phys* 41:242–252. <https://doi.org/10.1016/j.shpsb.2010.07.003>
- Kang B, Ramírez JA (2007) Response of streamflow to weather variability under climate change in the Colorado Rockies. *J Hydrol Eng* 12:63–72. [https://doi.org/10.1061/\(asce\)1084-0699\(2007\)12:1\(63\)](https://doi.org/10.1061/(asce)1084-0699(2007)12:1(63))
- Knutti R (2008) Should we believe model predictions of future climate change? *Philos Trans R Soc A Math Phys Eng Sci* 366:4647–4664. <https://doi.org/10.1098/rsta.2008.0169>
- Knutti R, Knutti R (2010) The end of model democracy? *Clim Chang* 102:395–404. <https://doi.org/10.1007/S10584-010-9800-2>
- Koutsoyiannis D, Montanari A, Lins HF, Cohn TA (2009) DISCUSSION OF “The implications of projected climate change for freshwater resources and their management”: climate, hydrology and freshwater: towards an interactive incorporation of hydrological experience into climate research. *Hydrol Sci J* 54:394–405
- Krysanova V, Zaherpour J, Didovets I et al (2020) How evaluation of global hydrological models can help to improve credibility of river discharge projections under climate change. *Clim Change* 163:1353–1377. <https://doi.org/10.1007/S10584-020-02840-0/TABLES/8>
- Kundzewicz ZW, Mata LJ, Arnell NW et al (2008) The implications of projected climate change for freshwater resources and their management. *Hydrol Sci J* 53:3–10. <https://doi.org/10.1623/hysj.53.1.3>
- Lauritzen PH, Nair RD, Herrington AR et al (2018) NCAR release of CAM-SE in CESM2.0: a reformulation of the spectral element

- dynamical core in dry-mass vertical coordinates with comprehensive treatment of condensates and energy. *J Adv Model Earth Syst* 10:1537–1570. <https://doi.org/10.1029/2017MS001257>
- Lurton T, Balkanski Y, Bastrikov V et al (2020) Implementation of the CMIP6 forcing data in the IPSL-CM6A-LR model. *J Adv Model Earth Syst* 12:e2019MS001940. <https://doi.org/10.1029/2019MS001940>
- Mauritsen T, Bader J, Becker T et al (2019) Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO₂. *J Adv Model Earth Syst* 11:998–1038. <https://doi.org/10.1029/2018MS001400>
- Moreno-Chamarro E, Caron L-P, Ortega P et al (2021) Can we trust CMIP5/6 future projections of European winter precipitation? *Environ Res Lett* 16:054063. <https://doi.org/10.1088/1748-9326/abf28a>
- O'Neill BC, Tebaldi C, van Vuuren DP et al (2016) The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geosci Model Dev* 9:3461–3482. <https://doi.org/10.5194/gmd-9-3461-2016>
- Palmer TE, McSweeney CF, Booth BBB et al (2022) Performance based sub-selection of CMIP6 models for impact assessments in Europe. *Earth Syst Dyn Discuss* 2022:1–45. <https://doi.org/10.5194/esd-2022-31>
- Palmer TN, Shutts GJ, Hagedorn R et al (2005) Representing model uncertainty in weather and climate prediction. *Annu Rev Earth Planet Sci* 33:163–193. <https://doi.org/10.1146/annurev.earth.33.092203.122552>
- Peel MC, Finlayson BL, McMahon TA (2007) Updated world map of the Köppen-Geiger climate classification. *Hydrol Earth Syst Sci* 11:1633–1644
- Prein AF, Bukovsky MS, Mearns LO et al (2019) Simulating North American weather types with regional climate models. *Front Environ Sci* 7:36. <https://doi.org/10.3389/FENV.2019.00036/BIBTEX>
- Räisänen J (2007) How reliable are climate models? *Tellus A Dyn Meteorol Oceanogr* 59:2. <https://doi.org/10.1111/j.1600-0870.2006.00211.x>
- Ravazzani G, Curti D, Gattinoni P et al (2016) Assessing groundwater contribution to streamflow of a large alpine river with heat tracer methods and hydrological modelling. *River Res Appl* 32:871–884. <https://doi.org/10.1002/rra.2921>
- Ruane AC, McDermid SP (2017) Selection of a representative subset of global climate models that captures the profile of regional changes for integrated climate impacts assessment. *Earth Perspect* 4:1–20. <https://doi.org/10.1186/S40322-017-0036-4>
- Salathe EP, Mote PW, Wiley MW (2007) Review of scenario selection and downscaling methods for the assessment of climate change impacts on hydrology in the United States pacific northwest. *Int J Climatol* 27:1611–1621. <https://doi.org/10.1002/joc.1540>
- Seland Ø, Bentsen M, Olivé D et al (2020) Overview of the Norwegian Earth System Model (NorESM2) and key climate response of CMIP6 DECK, historical, and scenario simulations. *Geosci Model Dev* 13:6165–6200. <https://doi.org/10.5194/gmd-13-6165-2020>
- Sellar AA, Jones CG, Mulcahy JP et al (2019) UKESM1: description and evaluation of the U.K. Earth System Model. *J Adv Model Earth Syst* 11:4513–4558. <https://doi.org/10.1029/2019MS001739>
- Soncini A, Bocchiola D, Azzoni R, Diolaiuti G (2017) A methodology for monitoring and modeling of high altitude Alpine catchments. *Prog Phys Geogr Earth Environ* 41:393–420. <https://doi.org/10.1177/0309133317710832>
- Soncini A, Bocchiola D, Confortola G, Bianchi A, Rosso R, Mayer C, Lambrecht A, Palazzi E, Smiraglia C, Diolaiuti G (2015) Future hydrological regimes in the upper Indus basin: a case study from a high altitude glacierized catchment. *J Hydrometeorol* 16(1):306–326
- Soncini A, Bocchiola D, Confortola G, Minora U, Vuillermoz E, Salerno F, Viviano G, Shrestha D, Senese A, Smiraglia C, Diolaiuti G (2016) Future hydrological regimes and glacier cover in the Everest region: the case study of the Dudh Koshi basin. *Sci Total Environ* 565:1084–1101
- Stucchi L, Bocchiola D (2023) Environmental flow assessment using multiple criteria: a case study in the Kumbih river, West Sumatra (Indonesia). *Sci Total Environ* 901:166516
- Stucchi L, Bombelli GMGM, Bianchi A, Bocchiola D (2019) Hydropower from the alpine cryosphere in the era of climate change. The case of the Sabbione storage plant in Italy. *Water (Switzerland)* 11. <https://doi.org/10.3390/w11081599>
- Tatebe H, Ogura T, Nitta T et al (2019) Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geosci Model Dev* 12:2727–2765. <https://doi.org/10.5194/gmd-12-2727-2019>
- Taylor KE, Stouffer RJ, Meehl GA (2012) An Overview of CMIP5 and the Experiment Design. *Bull Am Meteorol Soc* 93:485–498. <https://doi.org/10.1175/BAMS-D-11-00094.1>
- Voltaire A, Saint-Martin D, Sénési S et al (2019) Evaluation of CMIP6 DECK Experiments With CNRM-CM6-1. *J Adv Model Earth Syst* 11:2177–2213. <https://doi.org/10.1029/2019MS001683>
- Volodin EM, Mortikov EV, Kostykin SV et al (2018) Simulation of the modern climate using the INM-CM48 climate model. *Russ J Numer Anal Math Model* 33:367–374. <https://doi.org/10.1515/rnam-2018-0032>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.