AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE

10.1002/2017WR021633

Key Points:

- Distributed hydrological modeling is used as a tool to assess the coherence of precipitation and temperature data sets
- In complex terrain catchments hydrological coherence is strongly influenced by spatial resolution of the gridded data sets
- Gridded data sets of precipitation and temperature show different spatial and temporal patterns in the Southern Alps

Supporting Information:

- Supporting Information S1

Correspondence to:
L. Laiti,
lavinia.laiti@unitn.it

Citation:
Laiti, L., Mallucci, S., Piccolroaz, S., Bellin, A., Zardi, D., Fiori, A., et al. (2018). Testing the hydrological coherence of high-resolution gridded precipitation and temperature data sets. *Water Resources Research*, *54*, 1999–2016. https://doi.org/10.1002/2017WR021633

Received 31 JUL 2017
Accepted 16 FEB 2018
Accepted article online 23 FEB 2018
Published online 23 MAR 2018

Testing the Hydrological Coherence of High-Resolution Gridded Precipitation and Temperature Data Sets

L. Laiti[1], S. Mallucci[1], S. Piccolroaz[1,2], A. Bellin[1], D. Zardi[1], A. Fiori[3], G. Nikulin[4], and B. Majone[1]

[1]Department of Civil, Environmental, and Mechanical Engineering, University of Trento, Trento, Italy, [2]Department of Physics and Astronomy, Institute for Marine and Atmospheric research Utrecht, Utrecht University, Utrecht, The Netherlands, [3]Department of Engineering, University Roma Tre, Rome, Italy, [4]Rossby Centre, Swedish Meteorological and Hydrological Institute, Norrköping, Sweden

Abstract Assessing the accuracy of gridded climate data sets is highly relevant to climate change impact studies, since evaluation, bias correction, and statistical downscaling of climate models commonly use these products as reference. Among all impact studies those addressing hydrological fluxes are the most affected by errors and biases plaguing these data. This paper introduces a framework, coined Hydrological Coherence Test (HyCoT), for assessing the hydrological coherence of gridded data sets with hydrological observations. HyCoT provides a framework for excluding meteorological forcing data sets not complying with observations, as function of the particular goal at hand. The proposed methodology allows falsifying the hypothesis that a given data set is coherent with hydrological observations on the basis of the performance of hydrological modeling measured by a metric selected by the modeler. HyCoT is demonstrated in the Adige catchment (southeastern Alps, Italy) for streamflow analysis, using a distributed hydrological model. The comparison covers the period 1989–2008 and includes five gridded daily meteorological data sets: E-OBS, MSWEP, MESAN, APGD, and ADIGE. The analysis highlights that APGD and ADIGE, the data sets with highest effective resolution, display similar spatiotemporal precipitation patterns and produce the largest hydrological efficiency indices. Lower performances are observed for E-OBS, MESAN, and MSWEP, especially in small catchments. HyCoT reveals deficiencies in the representation of spatiotemporal patterns of gridded climate data sets, which cannot be corrected by simply rescaling the meteorological forcing fields, as often done in bias correction of climate model outputs. We recommend this framework to assess the hydrological coherence of gridded data sets to be used in large-scale hydroclimatic studies.

1. Introduction

Accurate records of climate observations are a prerequisite for reliable analyses of present and past climate, for evaluation of climate models, as well as for many other fundamental applications in climate impact studies in a wide range of fields, including hydrology, agriculture, renewable energy, and ecology. In particular, gridded data sets of precipitation and temperature, obtained from the interpolation of in situ measurements, satellite remote sensing, or atmospheric reanalysis (e.g., Michaelides et al., 2009 for precipitation and Harris et al., 2014 for temperature) are increasingly used for quantifying the sensitivity of natural systems to climate change (e.g., Cooper et al., 2016; Piccolroaz et al., 2018), for statistical downscaling and bias correction of climate models outputs (e.g., Mosier et al., 2014), as well as meteorological forcing of hydrological models (e.g., Ahmadalipour & Moradkhani, 2017; Ledesma & Futter, 2017; Paul et al., 2018; Piccolroaz et al., 2016).

The error associated with gridded data is often unknown, though it may be large depending on the type and amount of information used to obtain the gridded product. Data sets obtained by interpolating measurements taken at meteorological stations are affected by inaccuracies, which are spatially and, in case of precipitation, temporally variable and difficult to quantify (Haylock et al., 2008; Isotta et al., 2015). Measurement errors depend also on local conditions and increase with terrain elevation, as the operational conditions become more extreme (Frei & Schär, 1998). In addition, precipitation and temperature fields are undersampled at high elevations, because, for operational reasons, meteorological stations are preferentially located at low elevations (Hofstra et al., 2010). Furthermore, the interpolation method is an additional source of error, even when sophisticated geostatistical spatial analysis techniques are adopted and auxiliary predictors, such as terrain elevation and morphology, are included (e.g., Bénichou & Le Breton, 1987; Daly et al., 1994). Similar conclusions were reached in recent studies on the Western United States (Henn et al., 2017; Lundquist et al., 2015), where relative differences in the range between 5% and 60% in yearly precipitation totals were observed in six gridded precipitation data sets. On the other hand, the accuracy of gridded climate data derived by remote sensing techniques is closely linked to their spatiotemporal resolution (i.e., pixel size and acquisition frequency of satellite images), as well as to the inherent accuracy of the retrieval algorithms applied to the raw data (Ciabatta et al., 2017). For example, satellite rainfall estimates, either based on empirical calibration or physical modeling, are known to be prone to systematic biases, insensitivity to light precipitation, and failure over snow and ice surfaces (Kidd et al., 2012). Atmospheric reanalyses may also suffer from significant accuracy limitations, mainly due to their typically low spatial resolution and to the poor representation of subgrid processes (e.g., convection and associated precipitation) by the selected model parameterization (Kidd et al., 2013; Prein et al., 2015), both issues being especially relevant in complex terrain (Kotlarski et al., 2014).

A comprehensive review of studies investigating the uncertainty of gridded precipitation data sets at European scale can be found in Prein and Gobiet (2017). Uncertainty, under the form of differences between data sets, was found to be of the same order of magnitude of the precipitation biases derived from an ensemble of eight state-of-the-art Regional Climate Models (RCMs). Differently from the common practice of neglecting this source of uncertainty by selecting a priori a given data set as observational reference, Prein and Gobiet (2017) recommend considering the full ensemble of precipitation data sets in the RCM evaluation process (see also Contractor et al., 2015; Sunyer et al., 2013).

It is well known that uncertainty in the spatial distribution of meteorological variables propagates through the modeling chain adding uncertainty to the controlling variables in impact studies (the "uncertainty cascade" concept introduced by Wilby & Dessai, 2010). In this respect, Sperna Weiland et al. (2015) observed that the meteorological forcing is the major source of uncertainty, with hydrological parametric uncertainty playing a minor role (see also Gosling & Arnell, 2011; Nasonova et al., 2011). At a global scale, the need for accurate precipitation inputs, in order to obtain reliable water balance estimates, was highlighted by Fekete et al. (2004). Nevertheless, many climate studies based their evaluations on a single observational climate data set, without considering the effects of uncertainty possibly associated with that particular choice on the hydrological response (e.g., Yang et al., 2014). On the other hand, with the increasing availability of computational resources, one may be tempted to fully propagate the observational uncertainty, as well as all other sources of uncertainty, along the modeling chain, with the immediate consequence of generating a large ensemble of plausible projections. These projections may artificially inflate uncertainty by including data sets not coherent with the observed hydrological response. The need for reducing uncertainty to generate a plausible range of hydrologic storylines in climate change impact assessments has recently been highlighted in the review by Clark et al. (2016). Furthermore, we argue that the selection of observational gridded data sets should not be "democratic" (using the definition introduced by Knutti, 2010), since not all the data sets may be coherent with hydrological observations.

The idea of using hydrological modeling to assess the accuracy of gridded meteorological data sets has received attention in applications using satellite-based precipitation products over relatively short time periods in the context of flood forecasting applications (see, e.g., Quintero et al., 2016; Wang et al., 2017; Wu et al., 2017), with only a few applications dealing with long-term data sets derived from measurements at meteorological stations (see, e.g., Voisin et al., 2008). A common strategy used in these studies is to avoid the calibration of the hydrological model, since such a procedure may mask biases in the forcing input. In addition, the aim of these studies is to evaluate the sensitivity of streamflow to changes in the forcing, rather than to identify and exclude data sets with large errors. On the other hand, Bitew and Gebremichael (2011) observed that calibrating the model directly to satellite data improves the accuracy of the simulations, with respect to those obtained with a model calibrated by rain gauges data only, at the cost of an erroneous reproduction of evapotranspiration. Not to say that rain gauge data may be themselves affected by significant errors (cf. Estévez et al., 2015).

We contribute to this research topic by proposing an efficient methodology to test the *hydrological coherence* of gridded meteorological data sets with observed data and possibly reject them when their coherence is judged too low. Coherence is evaluated by benchmarking the hydrological variable of interest (streamflow, in the case study examined here) computed with a hydrological model fed by the selected data set and applied in an inverse modeling framework against available measurements. We call this procedure Hydrological Coherence Test (HyCoT). The advantage of this approach is that it condenses in a simple and easy-to-grasp metric the complex benchmarking process of observational meteorological data sets.

We emphasize that the selection of the hydrological model, its level of complexity and the metric to be employed for the comparisons depend on the particular goal at hand (e.g., prediction of streamflow, water content, groundwater flow, or of any other hydrological quantity of interest); hence, the suggested approach is by definition "*goal-oriented*," as for instance recently discussed by Fiori et al. (2016), Savoy et al. (2017), and Guthke (2017) in the context of subsurface hydrology. In the example provided here we focus on the prediction of streamflow at the basin scale. Thus, our contribution is a methodological framework that can easily be extended to other hydrological conceptual models, hydrological quantities, and efficiency metrics.

In our view, inferring the optimal (with respect to the selected metric) set of parameters by comparing the modeling output with available measurements is a necessary step in the procedure. It is in fact impossible, for a specific data set, to provide a better metric by using a set of parameters different from the optimal one, the only bias being in the limits of the metric itself that cannot be exhaustive and grasp all the hydrological processes. On the other hand, considering all the gridded data sets, including those performing poorly in the inversion, would inflate uncertainty, with any possible gain being offset by the poorer reproduction of observational data. In this sense, the metric is used to falsify the hypothesis that a given gridded product is coherent with hydrological observations and not to perform any sort of data sets ranking. In other words, we claim that the best way to marginalize data sets that, despite calibration, are not compatible with streamflow observations is to discard them. The proposed methodology is also flexible enough to allow exploring separately uncertainty associated with precipitation and temperature data sets. This may be particularly relevant in mountain regions, where snow dynamics influence streamflow seasonality, with temperature being the main driver.

The paper is organized as follows: section 2 presents a concise description of the study area, the climate data sets available and the streamflow data used for benchmarking the gridded data sets of climatic forcing; the hydrological modeling framework and the simulation setup are summarized in section 3. Concerning results, section 4 describes the differences among the selected gridded climate data sets, section 5 reports the main results of the hydrological benchmarking exercise, and section 6 analyzes the results from the application of a simple correction scheme to precipitation forcing. The main findings are discussed in section 7, whereas conclusions are finally drawn in section 8.

2. Study Area and Data Sets

2.1. Study Area

The area of interest for this study is the upper portion of the Adige river basin (Italy), in the southeastern Alpine region (see Figure 1), closed at the gauging station of Vò Destro (10°57′27.2″E, 45°44′ 06.7″N), accounting for 88% of the 12,100 km^2 total contributing area. Adige is the third largest Italian river basin: it originates close to the Alpine divide, in the Resia Pass area, and ends its course after 410 km in the northern Adriatic Sea. The river has a typical Alpine watershed, with terrain elevations ranging from 185 m a.s.l. in Trento to 3,500 m a.s.l. at the Italian-Austrian border. Morphology is dominated by deep valleys and high-mountain crests.

The climate of the area is characterized by relatively dry and cold winters, followed by humid summers and autumns. Rainfall is abundant in the pre-Alpine reliefs (especially in spring and fall) and over the most elevated peaks in the northern part of the basin (especially in summer, when rainfall is chiefly convective; cf. Panziera et al., 2015, 2016; Weissmann et al., 2005). Streamflow is minimum in winter, when snow falls over most of the catchment, and shows two maxima: one occurring early in summer, due to snowmelt, and the other in autumn, triggered by intense cyclonic storms. Recent analysis of hydroclimatic trends revealed that the Adige river basin is less sensitive to climate changes than other areas in Europe (Lutz et al., 2016). In particular, the projected reduction of snowfall in winter and anticipation of snow-melting, essentially due to rising temperatures associated with global warming (Gampe et al., 2016; Gobiet et al., 2014), will likely affect Adige streamflow regime by the second half of twenty-first century (Bard et al., 2015; Majone et al., 2016). This may have relevant consequences on a range of activities, e.g., hydropower production, which is particularly abundant in this region of the Alps (Zolezzi et al., 2009), especially for those located at the most

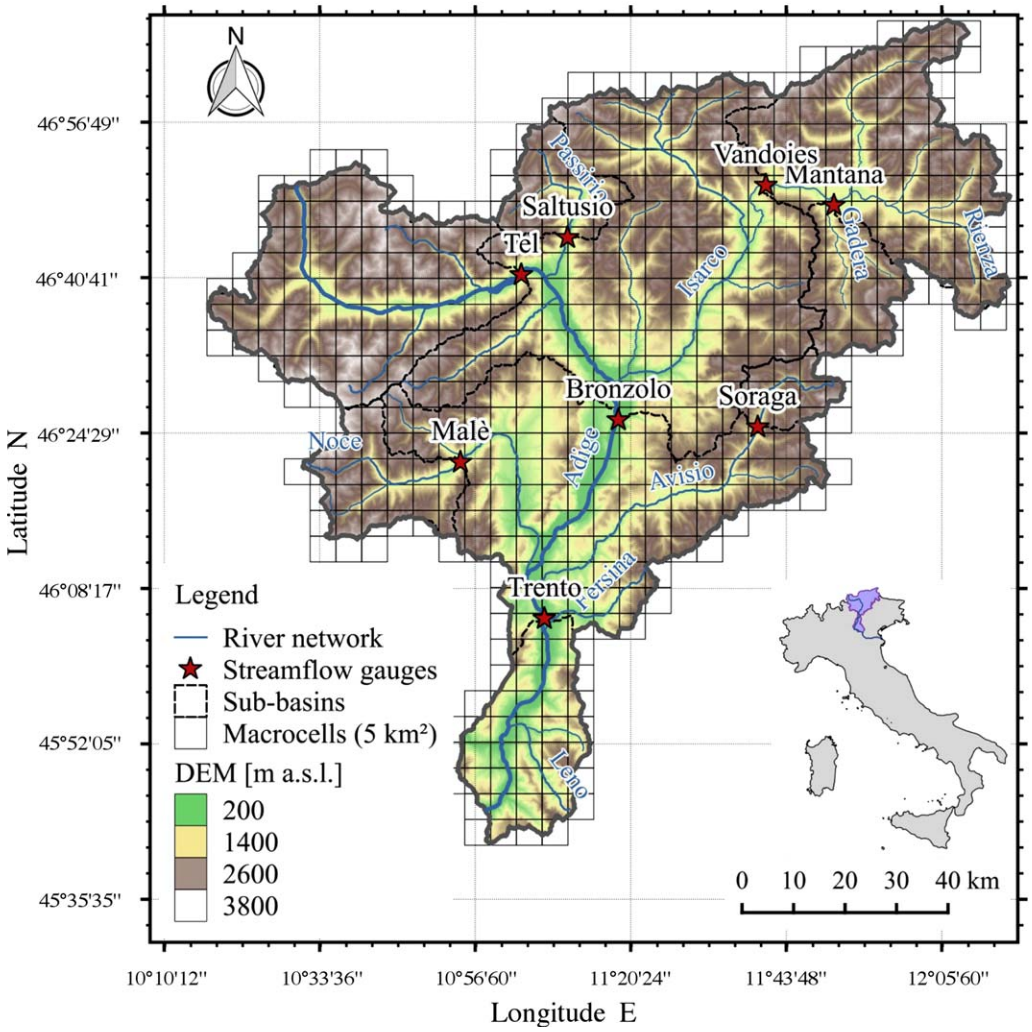

Figure 1. Map of the Adige river basin, representing the Digital Elevation Model (DEM), the river network, the computational grid cells ("macrocells"), the streamflow gauging stations, and the associated subbasins (cf. section 3). The inset shows the location of the Adige river basin within the Italian territory.

elevated sites (Bellin et al., 2016; Majone et al., 2016). See also Chiogna et al. (2016) for a comprehensive review of the hydrological stressors acting in the Adige basin, as well as of its ecological status.

2.2. Climatic Gridded Data Sets

Five gridded climate data sets were selected for the analysis. They are characterized by relatively high nominal spatial resolutions (grid spacing <25 km) and a daily temporal aggregation, and are briefly presented in the ensuing paragraphs (see Table 1 for a summary of their specifications). Almost all data sets have been extensively described in the literature. In particular three of them have been included in the assessment of precipitation data over the Alpine region by Isotta et al. (2015). Based on data availability (cf. Table 1), the 1989–2008 period was identified as common time frame for the analysis. Note that temperature data are not available for MSWEP and APGD data sets.

2.2.1. E-OBS Data Set

E-OBS (Haylock et al., 2008) is a pan-European gridded data set of daily mean, minimum and maximum temperature and total precipitation, available starting from 1950. Interpolation to a 0.25° (~22 km) latitude-longitude regular grid (Hofstra et al., 2009) was performed using observational gauge data from the ECA&D data set (Klok & Klein Tank, 2009). The product was originally developed under the ENSEMBLES EU-FP6 project (www.ensembles-eu.org) and continuously updated under the EURO4M project (www.euro4m.eu) (here v. 13.0 is used). Since its publication, E-OBS represents the state-of-the-art reference data set and has been

Table 1
Characteristics of Climatic Gridded Data Sets

	E-OBS	MSWEP	MESAN	APGD	ADIGE
Variables	P, T	P	P, T	P	P, T
Type of data set	Interpolated station data	Reanalyses, satellite and station data combination	Downscaled reanalyses, assimilated station data	Interpolated station data	Interpolated station data
Type of grid	Regular long.-lat.	Regular long.-lat.	Rotated-pole long.-lat.	ETRS89-LAEA	UTM zone 32
Grid spacing	∼23 km	∼23 km	∼5 km	5 km	1 km
No. of stations	26	Not available	26	185	165
Time resolution	Daily	3 hourly	Daily	Daily	Daily
Time coverage	1950–2016	1979–2015	1989–2010	1971–2008	1956–2014

Note. "P" and "T" stand for "precipitation" and "temperature," respectively. "No. of stations" indicates the number of stations within the Adige river basin used to produce the data set, which is an indicator of the overall observational density (i.e., effective resolution).

widely used for Regional Climate Model evaluation and climate monitoring over Europe (see recent examples in Turco et al., 2013; Kotlarski et al., 2014; Smiatek et al., 2016).

2.2.2. MSWEP Data Set

The recently released Multi-Source Weighted-Ensemble Precipitation (MSWEP) data set (Beck et al., 2017) is a global precipitation data set specifically designed for hydrological modeling, covering the period 1979–2015 on the same grid of E-OBS. Gridded precipitation is obtained as a weighted combination of data from seven data sets: two from interpolated gauge observations, three based on satellite remote sensing and two reanalysis products from numerical weather prediction models. The database includes an a-posteriori correction for gauge under-catch and orographic effects.

2.2.3. MESAN Data Set

The MESoscale ANalysis system (MESAN; Landelius et al., 2016) was developed by the Swedish Meteorological and Hydrological Institute. ERA-Interim global reanalyses (Dee et al., 2011) are first dynamically downscaled to a 0.22° grid to obtain HIRLAM regional reanalyses (Dahlgren et al., 2016), that are further downscaled to a 0.05° (∼5 km) grid and then used as first-guess field for the assimilation of surface station measurements from the ECA&D data set to finally produce the gridded product. For the present work, daily values of temperature (mean, minimum, and maximum) and precipitation, available for the years 1989–2010, have been considered.

2.2.4. APGD Data Set

The Alpine Precipitation Grid Data set (APGD; Isotta et al., 2014) estimates the distribution of daily precipitation over the greater Alpine region for the period 1971–2008, on the basis of measurements from more than 8,500 rain gauges. The unprecedentedly high density of input observations (cf. Table 1) contributes to its very high effective spatial resolution, estimated by the authors in approximately 10–20 km (Isotta et al., 2014, 2015). APGD data (RapdD 1.2 version) are provided on a grid with a nominal resolution of 5 km.

2.2.5. ADIGE Data Set

A regional precipitation and temperature data set (hereafter named ADIGE) has been independently developed for the Adige river basin, on the basis of meteorological station data provided by the Austrian Zentralanstalt für Meteorologie und Geodynamik (www.zamg.ac.at) and the meteorological offices of the Autonomous Provinces of Trento (www.meteotrentino.it) and Bolzano (www.provincia.bz.it/meteo). The data set results from 244 time series of precipitation and 350 of temperature recorded since 1956. The data, whose spatial density (and hence effective resolution) is comparable with that of APGD (cf. Table 1), were interpolated over a 1 km grid at a daily time step by means of a *kriging with external drift* algorithm (Goovaerts, 1997; Journel & Huijbregts, 1978; Journel & Rossi, 1989). The average absolute errors of the daily estimates were 1.32 mm for precipitation and 0.02°C for the temperature. Note that Isotta et al. (2014) reported a cross-validation error larger than 1 mm for APGD precipitation estimates (calculated only for wet days), value that is comparable with the error estimates of ADIGE precipitation fields.

2.3. Streamflow Data

Daily streamflow data collected at eight gauging stations were provided by the Hydrological Offices of the Autonomous Provinces of Trento (www.floods.it) and Bolzano (www.provincia.bz.it/hydro). Stations were selected according to the following criteria: (i) observational period including the 1989–2008 time frame adopted for the comparison of gridded climate data sets; (ii) limited number of gaps; (iii) high distance from storage reservoirs located in headwater catchments; and (iv) broad spatial coverage including the major tributaries of Adige river. The selected gauging stations along the Adige main stem are Trento, Bronzolo and Tel. The other gauging stations are found along the tributaries: Noce at Malè, Avisio at Soraga, Passirio at Saltusio, Gadera at Mantana, and Rienza at Vandoies (see Figure 1). These nodes of the network correspond to subcatchments of different size, elevation, and hydrological regimes (see supporting information Table S3). In particular, the southernmost station (Trento) drains almost the entire study domain (9,852 km^2), while Bronzolo (6,891 km^2) collects all streamflow contributions from the northern portion of the basin.

3. Methods

Before using them as input in hydrological simulations, the gridded data sets were resampled to the same computational grid with 5 km spacing, covering the Adige river basin (see the grid shown in Figure 1). This was needed because the grids of the data sets were slightly shifted and was done by first resampling the data sets to the same grid used in Adige (1 km spacing), followed by interpolation with the nearest neighbor method, followed by aggregation to the reference 5 km grid resolution by areal averaging. Hydrological inversion (i.e., identification of the optimal parameters) has been performed with each gridded product composed by precipitation and temperature data sets. Since APGD and MSWEP do not include a temperature data set they have been integrated with that of ADIGE. In a second set of simulations, ADIGE and APGD precipitation data sets were paired with temperature from E-OBS and MESAN, in order to explore how differences in the temperature field alone affect the simulation results. The first set of simulations is presented in details in section 5 whereas the remaining simulations are synthetically described in section 7. We anticipate here that no significant differences are observed when exchanging temperature data sets.

Hydrological simulations were performed at a daily time scale with the HYPERstream routing scheme (recently proposed by Piccolroaz et al., 2016), coupled with a continuous SCS-CN module for surface flow generation (Michel et al., 2005) (see Figure 2 for a conceptual scheme of the model). The HYPERstream routing scheme is specifically designed to be used in combination with gridded climate data sets and climate models. In fact, HYPERstream inherits the computational grid from the overlaying meteorological forcing, still preserving geomorphological dispersion caused by the river network (Pilgrim, 1977; Rinaldo et al., 1991, 1995; Rodríguez-Iturbe & Rinaldo, 1997) irrespectively of the grid resolution. This "perfect upscaling" (cf. Piccolroaz et al., 2016) can be achieved by application of suitable transfer functions derived from a high-resolution Digital Elevation Model of the study area. The continuous soil moisture accounting model for surface flow generation, based on the SCS-CN methodology (U.S. Soil Conservation Service, 1964), is here coupled with a nonlinear bucket model for soil moisture depletion (Majone et al., 2010; see also Hargreaves & Samani, 1982; Hock, 2003; Rango & Martinec, 1995). Note that this surface flow generation module was already successfully applied in two previous studies conducted in the Noce (Bellin et al., 2016) and Vermigliana (Piccolroaz et al., 2015) catchments, both subcatchments of the Adige. A detailed description of the hydrological modeling framework is provided in supporting information.

The hydrological model was calibrated against daily streamflow observations with the combination of meteorological forcing described above. The parameters space was explored for optimality, as defined by the Nash-Sutcliffe (NSE) (Nash & Sutcliffe, 1970) and the Kling-Gupta (KGE) (Gupta et al., 2009) efficiency indices, by using the Particle Swarming Optimization algorithm (Kennedy & Eberhart, 1995) in the implementation used by Castagna and Bellin (2009). In doing that, Bronzolo and Trento gauging stations were used in a multisite calibration framework (i.e., NSE and KGE are defined as the average of individual efficiencies obtained at the two stations), whereas the remaining six stations were used only for validation purposes. The first 2 years of the time series, 1989 and 1990, were used as spin up period for the simulations and therefore were excluded from the computation of both NSE and KGE.

NSE and KGE metrics were selected since they are customarily used in hydrological applications to assess the performance of a model in reproducing the observed streamflow. The NSE index ranges from $-\infty$ to 1, with the upper limit indicating perfect reproduction of observations, while NSE $= 0$ indicates that the model does not perform any better than the mean of the observational data. According to Moriasi et al. (2007),

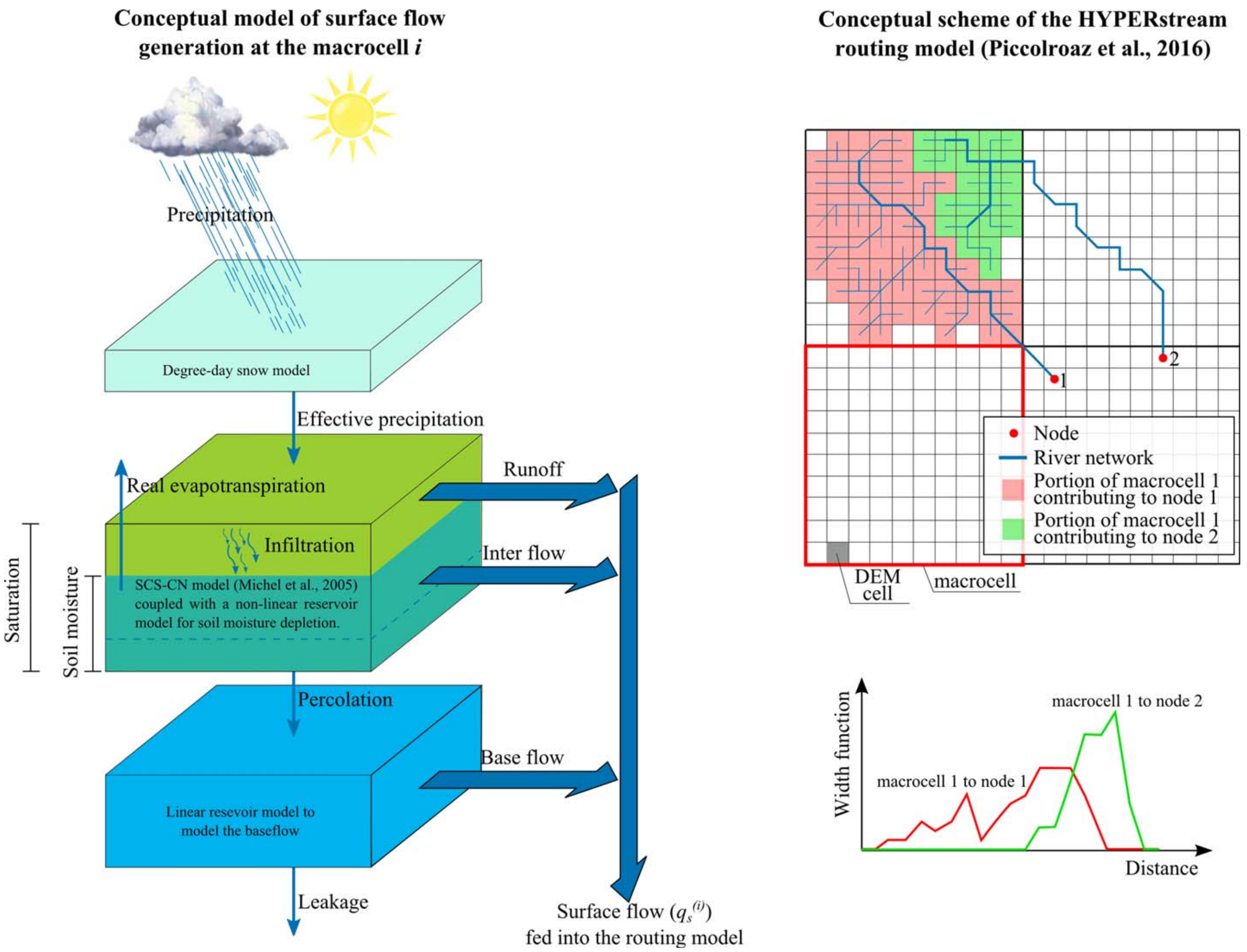

Figure 2. Conceptual scheme of the hydrological modeling framework, coupling the surface flow generation module (left) with the HYPERstream routing scheme (right) applied at each cell ("macrocell") of the computational grid.

NSE can be considered as satisfactorily good when it is larger than 0.5. KGE index computes the Euclidian distance of correlation, bias, and variability between simulated and observed time series. Similarly to NSE, KGE ranges from $-\infty$ to 1, with a value of 1 indicating the best accuracy. According to Gupta et al. (1999), a KGE value greater than 0.5 is considered as satisfactory. KGE is in general a better indicator than NSE for bias and variability, but leads to a reduction of the correlation between observations and simulations (Gupta et al., 2009).

In order to verify whether a simple rainfall correction procedure may improve the performance of biased data sets a Linear Scaling (LS) (Lenderink et al., 2007) transformation of the gridded precipitations was also employed. LS is a simple multiplicative correction method generally applied to correct biases in meteorological forcing as derived from Regional Climate Models (RCMs) simulations. Despite its simplicity, this method is among those used on regular basis in hydrological climate change impact assessments studies (see, e.g., the reviews by Teutschbein & Seibert, 2012; Chen et al., 2013). The LS method consists in rescaling a precipitation data set by a constant multiplicative factor given by the ratio between the precipitation amounts of an observational benchmark data set and those of the data set to be rescaled, both cumulated over the same time interval. In our case, we applied the rescaling by using both annual and monthly accumulation. The resulting multiplicative correction factors were then applied to the entire daily time series, such that the two aggregated time series (i.e., yearly or monthly) were the same. Corrections were applied

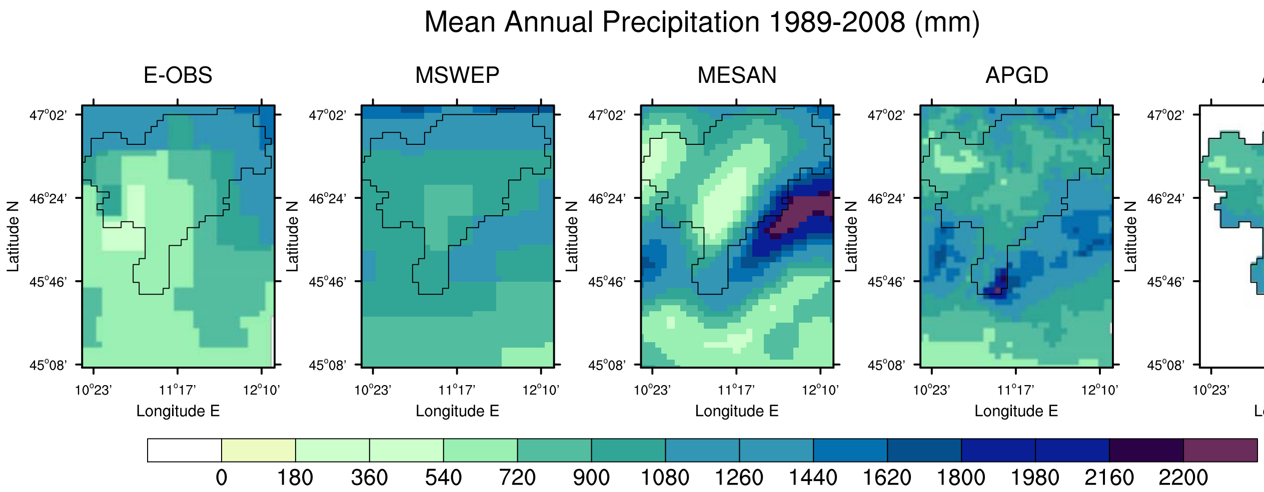

Figure 3. Maps of mean annual precipitation totals (1989–2008 average) according to the five different climate data sets.

directly to the computational 5 km grid, and the modified precipitation input was then fed into HYPERstream in order to carry out new calibration runs. Results of this analysis are presented in section 6.

4. Comparison of Climatological Patterns

Spatial patterns of annual mean precipitation resulting from the five data sets are rather different one from each other, except for APGD and ADIGE, which use similar observational data sets within the study area (Figure 3). Overall, ADIGE shows smaller annual mean precipitation than APGD (Figure 4). According to APGD the mean annual rainfall within the Adige catchment ranges from 550 to 2,300 mm. The largest annual totals are observed in the pre-Alpine reliefs (i.e., southernmost river basin), due to orographic lifting of moist air masses approaching from South (Frei & Schär, 1998), and over the mountains close to the main Alpine divide in the northernmost basin, where summer convective activity is most intense (Weissmann et al., 2005). Instead, the driest areas are the interior of the catchment and, especially, the major valley floors of the upper basin. Despite its high nominal spatial resolution (~5 km), MESAN reveals rather coarse

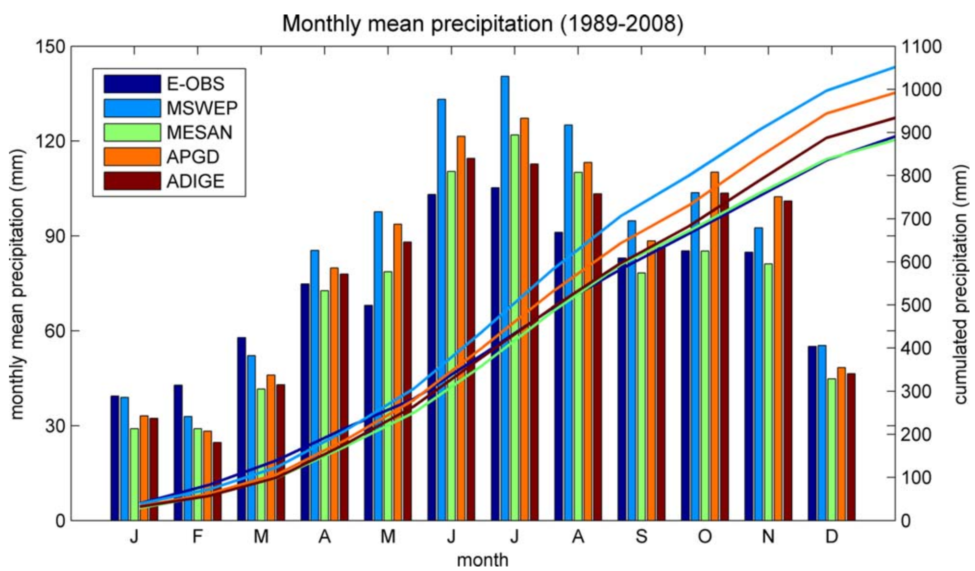

Figure 4. Annual cycle of monthly mean precipitation totals (1989–2008 average), averaged over the entire study area, according to the five different climate data sets (color bars). The associated evolution of cumulated precipitation totals is also shown (color lines).

precipitation field structures. Also, differently from APGD and ADIGE, its annual pattern only slightly resembles the underlying topography (Figure 3). On the one hand, this may be due to the limited number of ground stations used for data assimilation in MESAN, on the other to the grid spacing resolution and smoothed topography adopted in the regional climate modeling (cf. Prein et al., 2015). In general, both ∼23 km resolution maps (E-OBS and MSWEP) are much smoother than the others. However, despite its coarse grid, MSWEP shows spatial patterns closer to APGD and ADIGE than E-OBS.

Catchment-averaged monthly precipitation, shown in Figure 4, exhibits a clear bimodal distribution in all data sets, with a primary peak in summer (mainly associated with intense convection over mountain tops (Weissmann et al., 2005) and a secondary peak in autumn. The largest discrepancies among the analyzed data sets are generally observed in the wetter months (i.e., from May to November), when the range of variation of monthly estimates is within 17–35 mm, while in drier and colder months it typically ranges between 10 and 17 mm. As expected, monthly estimates from APGD and ADIGE agree very well, since they are obtained with a similar density of observations. MESAN estimates are also rather close, except for the lower values (by about 20 mm) observed in October and November. A similar pattern is observed for E-OBS. On the other hand, E-OBS shows higher precipitations (∼10 mm) in winter (i.e., January, February, March, and December). As for MSWEP, its annual cycle is similar to that of APGD and ADIGE with monthly totals being systematically higher. The long-term mean of annual precipitation totals averaged over the Adige basin for the five data sets range between 883 and 1,052 mm (see right *y* axis in Figure 4). In particular, MSWEP cumulates the highest totals, APGD ranks second and ADIGE third, with MESAN and E-OBS providing the lowest rainfall totals.

The added value of using higher-resolution grids is more evident for mean annual air temperature than for annual precipitation, with the spatial distribution of the former following closely the terrain topography (see Figure 5). This is an expected result for the ADIGE data set, since the interpolation scheme considers explicitly the elevation-temperature relationship (i.e., the vertical gradient of temperature) and thus the major morphological features of the catchment are visible in the resulting field. This topographic effect is also accounted for, although to a lesser extent, in MESAN, while it is almost absent in E-OBS, due to the smoothing effect of a lower resolution. In the northwestern part of the basin, ADIGE shows systematically

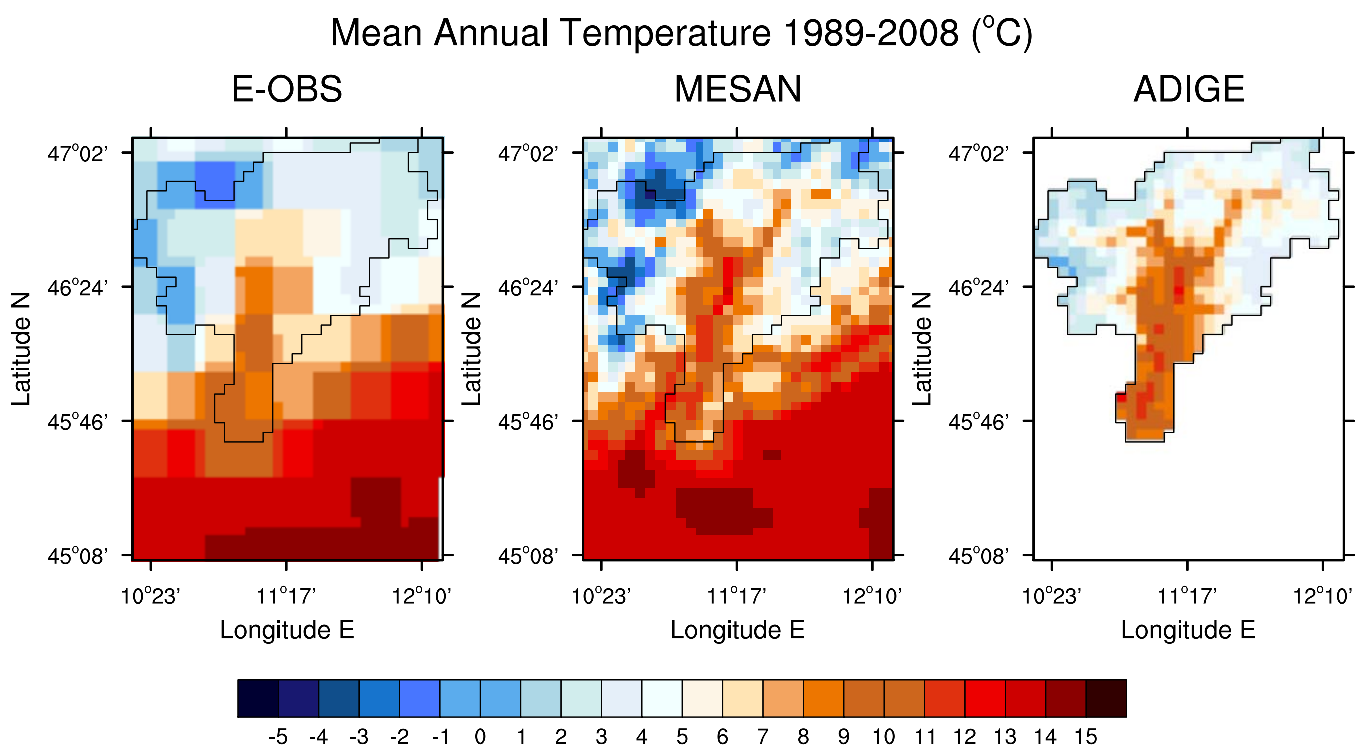

Figure 5. Maps of mean annual temperature (1989–2008 average) according to the three climate data sets including temperature.

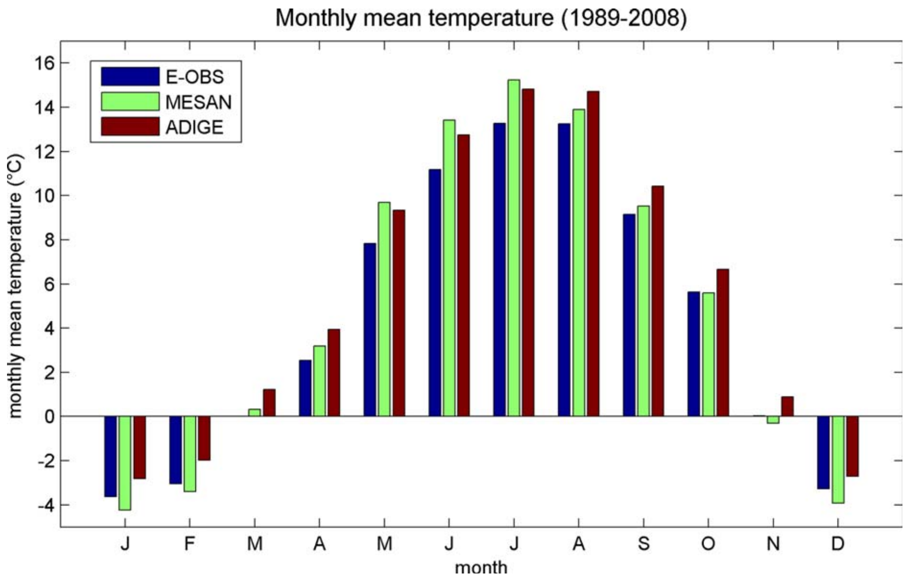

Figure 6. Monthly mean temperature (1989–2008 average), averaged over the entire study area, according to the three climate data sets including temperature.

warmer temperatures at high elevation (cf. in particular the northern and northwestern edges of the basin in Figure 5), which are evident despite small-scale discrepancies between the three temperature maps (i.e., both positive and negative local differences up to 5°C). This can be partially attributed to the different grid resolutions and/or to the different topographic data used in the interpolation scheme adopted in the data sets, as well as to different densities of the meteorological stations used to produce the gridded data sets (cf. Table 1).

Annual cycles of catchment-averaged monthly mean temperatures are shown in Figure 6. Differences among the data sets are between 1.5 and 2.3°C, and are typically larger in summer. Note that ADIGE is the warmest data set, except in late spring and summer (May, June, and July), when MESAN shows slightly higher temperatures.

5. Testing the Hydrological Coherence of the Datasets

In this section, hydrological coherence of the gridded climate data sets is assessed according to their ability to reproduce the observed streamflow time series in the Adige river basin, evaluated in the period 1991–2008 by computing both NSE and KGE indices (see section 3) at all gauging stations (Figure 7).

As specified in section 3, the parameters of the hydrological model were inferred for each data set by maximizing separately the average NSE and KGE at Trento and Bronzolo gauging stations. The data sets Adige and APGD produced the highest NSE values (NSE $>$ 0.81), followed by MSWEP, while E-OBS and MESAN provide smaller NSEs at both gauging stations. The ranking does not change if KGE is considered instead of NSE, as can be seen by comparing Figures 7a and 7b, though differences between the data sets and locations are smaller. As a general rule, by definition KGE values are higher than NSE counterparts. Breakdown of the two metrics obtained at the calibration stations into the three components evidenced that KGE favors parameters providing a better reproduction of variability and slightly smaller bias at the cost of a slight reduction of correlation, with respect to NSE. In particular, variability (defined as the ratio between the standard deviations of simulated and observed streamflows) increased on average from 0.87 in the case of NSE to 0.99 for KGE, bias (defined as the ratio between the mean simulated and mean observed streamflows) decreased on average from 1.03 (for NSE) to 1.02 (for KGE) and correlation (defined as the linear correlation coefficient between simulated and observed streamflows) decreased on average from 0.86 (for NSE) to 0.85 (for KGE). Note that the above values are obtained as average of the individual components among the five independent hydrological inversions (one for each data set).

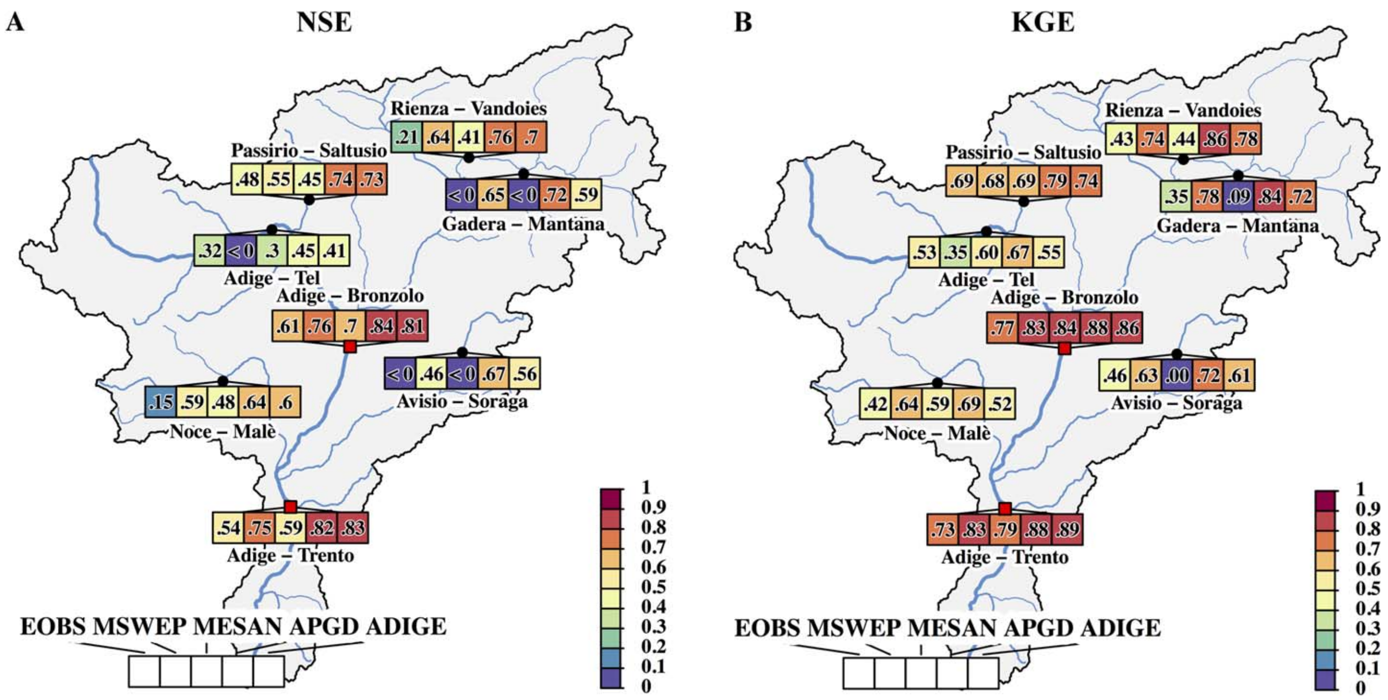

Figure 7. (a) Map of the NSE values calculated at the gauging stations of the Adige river basin. (b) As in Figure 7a but for KGE values. Simulated streamflows are obtained with HYPERstream model's parameters selected such as to maximize the average NSE and KGE at the gauging stations of Trento and Bronzolo. The model has been calibrated separately for each data set and performance metric.

At the validation stations, APGD and ADIGE generally provide the best performance in terms of NSE, with APGD performing slightly better than ADIGE. Conversely, the other data sets produce significantly smaller values in almost all the validation subcatchments (Figure 7a). In particular, MESAN and E-OBS produce the worst results in small subcatchments (i.e., lower than 2,000 km^2, see supporting information Table S3), with low and even negative NSE found in the northeastern part of the simulation domain. Conversely, MSWEP shows a single negative value at a gauging station in the northwestern part of the basin (Tel), where also APGD and ADIGE are not satisfactory (i.e., NSE < 0.5). Investigation of KGE at validation stations confirms the generally better performances of ADIGE and APGD in reproducing observed data (KGE > 0.5 threshold everywhere), though MSWEP performs better than ADIGE at the gauging stations of Malè, Mantana, and Soraga (see Figure 7b). The reduction of performances from calibration to validation sites is more pronounced in simulations performed by using NSE as metric with respect to KGE, which assumes values smaller than 0.5 only at Mantana and Soraga with MESAN. At most locations, these two metrics are in agreement, though at a few gauging stations values of NSE < 0.5 are paired with values of KGE > 0.5 (see, e.g., Tel for four out of five data sets), accordingly to the fact that KGE values are generally higher than NSE counterparts.

6. Analysis of a Simple Correction Scheme Applied to Precipitation Forcing

Here we assess the role of spatiotemporal patterns in shaping the hydrological response and analyze the impact of the differences between the data sets. In particular, we applied a simple LS correction (cf. section 3) to the MESAN data set, here selected as representative of underperforming data sets, using APGD as observational benchmark due to its higher accuracy in reproducing the observed streamflow (see section 5). These data sets were also chosen because they are characterized by almost the same nominal resolution (see Table 1), thereby avoiding possible scale effects due to the original grid size. Note that calibrations were performed only with reference to NSE efficiency and that modified precipitation inputs were coupled with unmodified MESAN temperature data. This correction was applied by matching annual and monthly totals of the mean precipitation computed over the entire river basin (identified with MESAN_1 and MESAN_2 scenarios for the annual and monthly rescaling factors, respectively) and separately at each

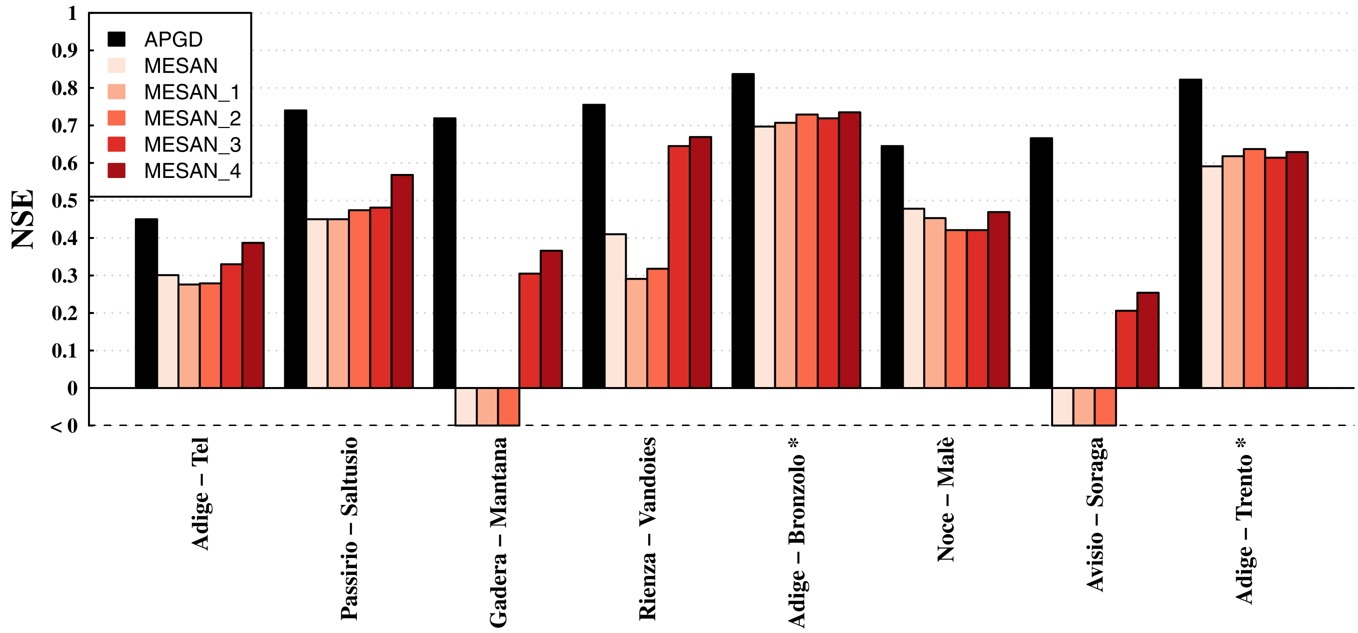

Figure 8. Comparison of NSE indices associated with the four LS scenarios (MESAN_1 to MESAN_4, in red scale), together with the values obtained from the calibration of the hydrological model using original APGD (black) and MESAN (light pink) as input meteorological forcing.

cell (identified with MESAN_3 and MESAN_4 scenarios, respectively), thereby resulting in a total of four possible rescaling scenarios. In MESAN_1 and MESAN_2 scenarios, the rescaling coefficients are constant in space (uniform) and vary temporally at monthly time scale only in the latter case. On the other hand, in MESAN_3 and MESAN_4 scenarios, the matching is imposed at single cell level, such that the rescaling coefficients are variable in space and are constant in time in MESAN_3 while they vary monthly in MESAN_4. The numbering of the scenarios increases with the degree of alteration with respect to the original time series.

Due to the coherent results obtained using alternatively NSE or KGE (see section 5 and Figure 7), for simplicity we performed this exercise by using the former. The comparison of NSE indices at all calibration and validation points is shown in Figure 8 in the four LS scenarios described above. We emphasize here that HYPERstream has been calibrated on each one of the four scenarios, such that the resulting model parameters are optimal for each of them. For the sake of clarity, we also show again the NSE indices obtained using APGD and MESAN as input meteorological forcing. The most striking result is that none of the LS scenarios is able to match or even get close to the performance of APGD. This is valid for all the investigated locations. A general and moderate improvement of NSE index with respect to the original MESAN simulation is however achieved, although not for all the LS scenarios. In particular, scenarios MESAN_1 and MESAN_2 (i.e., those not altering the spatial distribution of the precipitation) often lead to a small deterioration of the NSE index with respect to the original MESAN data set (e.g., at Tel, Vandoies, and Malè). On the other hand, scenarios MESAN_3 and MESAN_4, i.e., cell-by-cell corrections, lead to a general improvement of the accuracy, which is a reasonable result, given the pronounced spatial variability of the precipitation field in the study area (cf. section 4), which is an important determinant of the differences in the unitary contribution (i.e., the water discharge divided by the area of the catchment) at the stream gauges. At some locations such improvement is extremely relevant: for instance, at Mantana and Soraga NSE indices shift from negative (i.e., complete model failure) to small positive values (0.20–0.35 range), while at Vandoies NSE increases from 0.3 to 0.65. As a general result, corrections applied at the monthly scale (scenarios MESAN_2 and MESAN_4) perform slightly better than at corrections applied at the annual scale (scenarios MESAN_1 and MESAN_3), as they allow a better reproduction of the seasonal rainfall pattern of the benchmark data set. As expected, the scenario MESAN_4 shows the largest improvement in performance, with the exceptions of Malè, Bronzolo, and Trento stations. In particular, at Malè NSE does not show any appreciable improvement with respect to the original MESAN performance. We attribute this fact to the balancing and smoothing effects entailed by both the hydrological model calibration process (independently carried out for each data set) and the spatiotemporal averaging effect of the catchment, which intensifies as the size of the catchment increases.

7. Discussion

In the present work, we propose a procedure, coined as HyCoT, which allows to identify gridded data sets not coherent with hydrological observations. Their inclusion in hydrological simulations should be carefully evaluated, since they may inflate uncertainty. The procedure falsifies the hypothesis that a gridded data set is hydrologically coherent if, after calibration, the reproduction of hydrological observational data are not satisfying according to a selected metric, compensation effects of hydrological models, notwithstanding. In other words, we are not trying to identify here the best data set, rather exclude those that after calibration are still not able to reproduce observational data, while other data sets do. Hydrological modeling has been performed with HYPERstream routing scheme coupled with a continuous SCS-CN module for surface flow generation (see supporting information for a detailed description), but any other hydrological model can be used, better if it is the one that will be used in the applications.

We found that applying the hydrological model at several streamflow gauging stations within the study area (in particular those not included in the computation of efficiency metrics) allowed verifying how the hydrological coherence of the data sets varies in space. This is particular evident for stations with small drainage areas (see, e.g., NSE and KGE efficiencies at Tel, Mantana, and Soraga in Figures 7a and 7b) where the limited averaging effect of the catchment highlights the importance of accurate and spatially well resolved precipitation and temperature fields (Heistermann & Kneis, 2011). This entails that the accuracy in the reproduction of the spatiotemporal patterns of precipitation and temperature inputs, which is especially hard to achieve in the domain of analysis and in regions with complex terrain, is crucial for obtaining accurate estimates of local-scale hydrological processes (see also Tuo et al., 2016). The much smaller reduction of NSE and KGE metrics obtained for APGD and ADIGE moving from the gauging stations used for calibration (i.e., Trento and Bonzolo) to the nested stations (not used in the calibration) shows that the compensating effects between not fully constrained hydrological fluxes does not influence the identification of hydrologically noncoherent data sets.

Simulation results evidenced that the HyCoT response does not change by considering two alternative efficiency metrics, though at a few locations NSE and KGE suggest different level of accuracy. This is not surprising and is in accordance with the general understanding that different alternative metrics have their own peculiarities, limitations, and trade-offs (see, e.g., Gupta et al., 2009; Schaefli & Gupta, 2007). Ultimately, our point of view is that the selection of the conceptual model itself, its level of complexity, and the efficiency criteria (as well as the definition of suitable thresholds) should be made in a "*goal-oriented*" framework, as stated in section 1.

In addition, HyCoT allows for investigating the possible interactions of biases in precipitation and temperature data sets, which to our best knowledge has not been investigated so far. To this end, precipitation data sets from ADIGE and APGD, which provided globally the better coherence with streamflow observations, were paired with the temperature extracted from E-OBS and MESAN data sets. Similarly to the previous analysis, these inversions were performed in a multisite framework adopting NSE as efficiency metric. The analysis revealed that at the calibration sites of Trento and Bronzolo NSE obtained with these combinations of precipitation and temperature data sets present negligible differences among each other (results not shown here), and are practically identical to the values obtained using the precipitation data set of the gridded product. At the validation sites, small deviations were observed depending on the selected temperature data set. In this application, HyCoT is insensitive to switching temperature data sets chiefly because of the marginal differences in temperature distribution among the selected data sets, in relation to their influence in hydrological processes. However, HyCoT is not limited to particular conditions and can be used to evidence inconsistencies in the temperature distribution as well.

Results of this analysis are also relevant for climate change impact assessments in regions characterized by low observational density (i.e., low effective spatial resolution and accuracy of gridded climate data sets; cf. also Duan et al., 2016; Tuo et al., 2016). In particular, the framework allows to highlight the deficiencies of hydrological simulations driven by inaccurate input meteorological forcing, i.e., data sets which do not pass hydrological validation at relevant spatial scales. Following the concept introduced by Knutti (2010), we argue that the selection of gridded data sets should not be "*democratic*," since not all are equally performing. In general, we agree with the recommendations of Prein and Gobiet (2017) not to neglect this source of uncertainty by selecting a priori a single data set as reference. However, we disagree when it is suggested to consider the full ensemble of precipitation data sets in both RCM evaluation process and climate-hydrology modeling chain. HyCoT can be seen in the wider Bayesian Model Averaging (BMA) framework (Duan et al., 2007; Hoeting et al., 1999; Roy et al., 2017; Vrugt et al., 2006) aimed at providing a probabilistic weighted average of alternative models and meteorological forcing. However, the negative impact of a hydrologically incoherent data set, i.e., a data set not able to provide an acceptable reproduction of streamflow when coupled with a hydrological model, is only reduced, yet not eliminated, by any kind of Bayesian weighting, being preferable its elimination from the collections of data sets. From a Bayesian point of view, this elimination can be considered as part of prior information to be included in the updating process (Majone et al., 2016; Sadegh & Vrugt, 2014).

Results obtained with the LS approach discussed in section 6 show that the correct reproduction of the spatiotemporal distribution of the precipitation field is fundamental for an accurate reproduction of hydrological processes in climate change impact studies. Such finding is particularly relevant when assessments are performed at small spatial scales and in areas characterized by complex topography, where the spatiotemporal variability of the precipitation field is increased by topographic uplift (see, e.g., Majone et al., 2012, 2016). The LS exercise revealed how bias correction methods used in climate change impact assessment studies may not be able to counterbalance the inadequate representation of spatial patterns in meteorological forcing that are crucial for hydrological application. In this sense, we show that, besides the exclusion of incoherent data sets, also the selection of the most appropriate bias correction approach can be achieved through hydrological benchmarking. We limited our analysis here to the simplest possible bias correction methodology, which is often used in applications, and did not analyze more sophisticated approaches (see, e.g., Berg et al., 2012; Teutschbein & Seibert, 2012), because testing the effectiveness of a number of bias correction techniques is beyond the scope of the present work.

In addition to the general results illustrated above, additional remarks on specific data sets can also be drawn. The first is that embedding high-resolution model reanalyses in MESAN does not bring much improvement in the region of interest. In fact, hydrological simulations driven by MESAN show performances similar to those of the hydrological model driven by E-OBS, which shares the same observational database but on a coarser grid (see section 2.2). The density of observational points (i.e., the meteorological stations) is a crucial factor in determining the effective resolution (i.e., the accuracy) of a gridded data set. The second consideration concerns the MSWEP data set. According to the present study, which represents one of the first hydrological applications of such data set in the Alpine region, MSWEP is suboptimal with respect to products obtained exclusively with interpolation of ground observations (i.e., APGD and ADIGE). This may be due to its low resolution, which blurs relevant spatial patterns of the precipitation field. On the other hand, MSWEP performs better than E-OBS and MESAN, despite the higher nominal resolution of the latter. This indicates that the merging of multiple data sources might be a convenient approach for improving the accuracy of gridded climate products. Note, however that, differently from the other data sets, MSWEP already includes a correction for precipitation under-catch based on streamflow observations (cf. section 2.2.2 and Beck et al., 2017).

8. Conclusions

We proposed a method, coined here as Hydrological Coherence Test (HyCoT), that consists in the use of a physically based hydrological model as a tool to test the "*hydrological coherence*" of precipitation and temperature data sets, i.e., their compatibility with the physical processes and the hydrological variables of interest.

In the present study, HyCoT is applied to the analysis of streamflow at the basin scale, comparing high-resolution gridded daily data sets of precipitation and temperature available in the Alpine catchment of the Adige river (northeastern Italy) for the period 1989–2008. Given the large variability of the precipitation provided by these data sets, we evaluated their accuracy with an integrated approach through numerical simulation of hydrological processes within the basin. This was done by applying a distributed hydrological model with climate inputs extracted from each data set and by validating the results against streamflow observations.

The main results of this work are the following:

1. The proposed screening method allowed for testing the coherence of the five gridded climate data sets according to their accuracy in simulating hydrological processes in the study area. The results identified APGD and ADIGE as the best candidates for hydrological applications in the Adige catchment. On the contrary, despite acceptable results at the larger basin scale, E-OBS, MESAN, and MSWEP were found unable to correctly reproduce the observed streamflow at the smaller nested subcatchments, notwithstanding the compensating effect introduced by calibration; we showed that HyCoT is a powerful tool to exclude poorly performing data sets.
2. The results show sensitivity to the particular efficiency metric adopted (i.e., NSE and KGE in our case); this somehow expected result is consistent with the goal-oriented nature of HyCoT.
3. The fundamental role of observational density in determining the effective accuracy of gridded climate products was thus clearly confirmed, emphasizing the key importance of an accurate representation of spatiotemporal patterns of precipitation and temperature (cf. also Hofstra et al., 2010; Isotta et al., 2015). This is especially true in complex-topography areas, which typically show large spatiotemporal variability of the meteorological forcing.
4. The temperature data sets associated to the selected gridded products can be used interchangeably, since their differences have a negligible impact on the simulation results and the data set selection, thus reinforcing the conclusion that biases in precipitation are responsible for most of the biases in modeling results, at least in the mountain environment considered here.
5. It was found that correction methods based on linear rescaling of rainfall data, which are conceptually similar (although not identical) to simple bias correction approaches, may not be able to alleviate the aforementioned problems plaguing the coarser data sets. More sophisticated bias correction approaches may perform differently; still we surmise that the proposed HyCoT benchmarking procedure can be effectively employed also for assessing the performances of such approaches.

Summarizing, the proposed HyCoT methodology is a useful tool to screen available gridded meteorological data sets to be used as climatic forcing in hydrological applications. The method is particularly powerful in the identification of climatic data sets that are not coherent with hydrological observations, suggesting their exclusion in hydrological analyses. HyCoT is also useful in the identification of biases and inaccuracies of climate data in regional climate studies, e.g., to consistently reduce the uncertainty of projections that employ multiple data sets.

The HyCoT methodology is neither model nor metric dependent, and any modeling approach can be used, as well as any type of metric to assess modeling accuracy. The selection of the hydrological model, its level of complexity, and the metric to be employed for the comparisons depend on the particular goal at hand (e.g., prediction of streamflow, water content, groundwater flow, or of any other hydrological quantity of interest). Under such conditions, the method relies on the judicious model and metric selection by the modeler.

Acknowledgments
This research has been partially supported by the European Union under the 7th Framework Programme (grant agreement 603629-ENV-2013-6.2.1-Globaqua), and by the CLIMAWARE (CLIMatic change impacts on future Availability of WAter REsources and hydrogeological risks) project, funded by the University of Trento. The authors acknowledge the E-OBS data set from the EU-FP6 project ENSEMBLES (ensembles-eu.metoffice.com) and the data providers of the ECA&D project (www.ecad.eu), the Swiss Federal Office of Meteorology and Climatology MeteoSwiss for providing the Alpine Precipitation Gridded Data set (http://www.meteoswiss.admin.ch/home/search.subpage.html/en/data/products/2015/alpine-precipitation.html), and the Swedish Meteorological Hydrological Institute for providing the MESAN data. MSWEP data set was downloaded from www.gloh2o.org. The authors also thank the Austrian Zentralanstalt für Meteorologie und Geodynamik, the meteorological service of the Autonomous Province of Trento Meteotrentino (www.meteotrentino.it), and the Hydrographic Office of the Autonomous Province of Bolzano (www.provincia.bz.it/meteo), for providing the weather station data forming the basis for the ADIGE data set. Streamflow data were kindly provided by the Hydrographic Office of the Autonomous Province of Bolzano (www.provincia.bz.it/hydro) and by the Service for Hydraulic Works of the Autonomous Province of Trento (www.floods.it). Finally, we would like to thank Andrew Newman, Shervan Gharari, and an anonymous Reviewer for their helpful and constructive comments.

References

Ahmadalipour, A., & Moradkhani, H. (2017). Analyzing the uncertainty of ensemble-based gridded observations in land surface simulations and drought assessment. *Journal of Hydrology*, *555*, 557–568. https://doi.org/10.1016/j.jhydrol.2017.10.059

Bard, A., Renard, B., Lang, M., Giuntoli, I., Korck, J., Koboltschnig, G., et al. (2015). Trends in the hydrologic regime of Alpine rivers. *Journal of Hydrology*, *529*, 1823–1837. https://doi.org/10.1016/j.jhydrol.2015.07.052

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., et al. (2017). MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite and reanalysis data. *Hydrology and Earth System Sciences*, *21*, 589–615. https://doi.org/10.5194/hess-21-589-2017

Bellin, A., Majone, B., Cainelli, O., Alberici, D., & Villa, F. (2016). A continuous coupled hydrological and water resources management model. *Environmental Modelling & Software*, *75*, 176–192. https://doi.org/10.1016/j.envsoft.2015.10.013

Bénichou, P., & Le Breton, O. (1987). AURELHY: une méthode d'analyse utilisant le relief pour les besoins de l'hydrométéorologie. In *Deuxièmes journées hydrologiques de l'ORSTOM à Montpellier* (Vol. 2, pp. 299–304). Paris, FRA: ORSTOM. ISBN 2–7099-0865-4

Berg, P., Feldmann, H., & Panitz, H. J. (2012). Bias correction of high resolution regional climate model data. *Journal of Hydrology*, *448–449*, 80–92. https://doi.org/10.1016/j.jhydrol.2012.04.026

Bitew, M. M., & Gebremichael, M. (2011). Evaluation of satellite rainfall products through hydrologic simulation in a fully distributed hydrologic model. *Water Resources Research*, *47*, W06526. https://doi.org/10.1029/2010WR009917

Castagna, M., & Bellin, A. (2009). A Bayesian approach for inversion of hydraulic tomographic data. *Water Resources Research*, *45*, W04410. https://doi.org/10.1029/2008WR007078

Chen, J., Brissette, F. P., Chaumont, D., & Braun, M. (2013). Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. *Water Resources Research*, *49*, 4187–4205. https://doi.org/10.1002/wrcr.20331

Chiogna, G., Majone, B., Cano Paoli, K., Diamantini, E., Stella, E., Mallucci, S., et al. (2016). A review of hydrological and chemical stressors in the Adige basin and its ecological status. *Science of the Total Environment*, *540*, 429–443. https://doi.org/10.1016/j.scitotenv.2015.06.149

Ciabatta, L., Marra, A. C., Panegrossi, G., Casella, D., Sanò, P., Dietrich, S., et al. (2017). Daily precipitation estimation through different microwave sensors: Verification study over Italy. *Journal of Hydrology*, *545*, 436–450. https://doi.org/10.1016/j.jhydrol.2016.12.057

Clark, M. P., Wilby, R. L., Gutmann, E. D., Vano, J. A., Gangopadhyay, S., Wood, A. W., et al. (2016). Characterizing uncertainty of the hydrologic impacts of climate change. *Current Climate Change Reports*, *2*, 55–64. https://doi.org/10.1007/s40641-016-0034-x

Contractor, S., Alexander, L. V., Donat, M. G., & Herold, N. (2015). How well do gridded datasets of observed daily precipitation compare over Australia? *Advances in Meteorology*, *2015*, 325718. https://doi.org/10.1155/2015/325718

Cooper, M. G., Nolin, A. W., & Safeeq, M. (2016). Testing the recent snow drought as an analog for climate warming sensitivity of Cascades snowpacks. *Environmental Research Letters*, *11*(8), 084009. https://doi.org/10.1088/1748-9326/11/8/084009

Dahlgren, P., Landelius, T., Kållberg, P., & Gollvik, S. (2016). A high-resolution regional reanalysis for Europe. Part 1: Three-dimensional reanalysis with the regional HIgh-Resolution Limited-Area Model (HIRLAM). *Quarterly Journal of the Royal Meteorological Society*, *142*, 2119–2131. https://doi.org/10.1002/qj.2807

Daly, C., Neilson, R. P., & Phillips, D. L. (1994). A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology*, *33*, 140–158. https://doi.org/10.1175/1520-0450(1994)033 < 0140:ASTMFM>2.0.CO;2

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*, 553–597. https://doi.org/10.1002/qj.828

Duan, Q., Ajami, N. K., Gao, X., & Sorooshian, S. (2007). Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Advances in Water Resources*, *30*(5), 1371–1386. https://doi.org/10.1016/j.advwatres.2006.11.014

Duan, Z., Liu, J. Z., 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. *Science of the Total Environment*, *573*, 1536–1553. https://doi.org/10.1016/j.scitotenv.2016.08.213

Estévez, J., Gavilán, P., García-Marín, A. P., & Zardi, D. (2015). Detection of spurious precipitation signals from automatic weather stations in irrigated areas. *International Journal of Climatology*, *35*, 1556–1568. https://doi.org/10.1002/joc.4076

Fekete, B. M., Vörösmarty, C. J., Roads, J. O., & Willmott, C. J. (2004). Uncertainties in precipitation and their impacts on runoff estimates. *Journal of Climate*, *17*, 294–304. https://doi.org/10.1175/1520-0442(2004)017<0294:UIPATI>2.0.CO;2

Fiori, A., Cvetkovic, V., Dagan, G., Attinger, S., Bellin, A., Dietrich, P., et al. (2016). Debates-stochastic subsurface hydrology from theory to practice: The relevance of stochastic subsurface hydrology to practical problems of contaminant transport and remediation. What is characterization and stochastic theory good for? *Water Resources Research*, *52*, 9228–9234. https://doi.org/10.1002/2015WR017525

Frei, C., & Schär, C. (1998). A precipitation climatology of the Alps from high-resolution rain-gauge observations. *International Journal of Climatology*, *18*, 873–900. https://doi.org/10.1002/(SICI)1097-0088(19980630)18:8<873::AID-JOC255>3.0.CO;2-9

Gampe, D., Nikulin, G., & Ludwig, R. (2016). Using an ensemble of regional climate models to assess climate change impacts on water scarcity in European river basins. *Science of the Total Environment*, *573*, 1503–1518. https://doi.org/10.1016/j.scitotenv.2016.08.053

Gobiet, A., Kotlarski, S., Beniston, M., Heinrich, G., Rajczak, J., & Stoffel, M. (2014). 21st century climate change in the European Alps: A review. *Science of the Total Environment*, *493*, 1138–1151. https://doi.org/10.1016/j.scitotenv.2013.07.050

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation* (483 pp.). Oxford, UK: Oxford University Press.

Gosling, S. N., & Arnell, N. W. (2011). Simulating current global river runoff with a global hydrological model: Model revisions, validation and sensitivity analysis. *Hydrological Processes*, *25*, 1129–1145. https://doi.org/10.1002/hyp.7727

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez-Barquero, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modeling. *Journal of Hydrology*, *377*, 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1999). Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *Journal of Hydrologic Engineering*, *4*(2), 135–143. https://doi.org/10.1061/(ASCE)1084-0699(1999)4:2(135)

Guthke, A. (2017). Defensible model complexity: A call for data-based and goal-oriented model choice. *Groundwater*, *55*(5), 646–650. https://doi.org/10.1111/gwat.12554

Hargreaves, G. H., & Samani, Z. A. (1982). Estimating potential evapotranspiration. *Journal of the Irrigation and Drainage Division*, *108*, 225–230.

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 dataset. *International Journal of Climatology*, *34*, 623–642. https://doi.org/10.1002/joc.3711

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., & New, M. (2008). A European daily high-resolution gridded dataset of surface temperature and precipitation. *Journal of Geophysical Research*, *113*, D20119. https://doi.org/10.1029/2008JD10201

Heistermann, M., & Kneis, D. (2011). Benchmarking quantitative precipitation estimation by conceptual rainfall-runoff modeling. *Water Resources Research*, *47*, W06514. https://doi.org/10.1029/2010WR009153

Henn, B., Newman, A. J., Livneh, B., Daly, C., & Lundquist, J. D. (2017). An assessment of differences in gridded precipitation datasets in complex terrain. *Journal of Hydrology*, *556*, 1205–1219. https://doi.org/10.1016/j.jhydrol.2017.03.008

Hock, R. (2003). Temperature index melt modelling in mountain areas. *Journal of Hydrology*, *282*, 104–115. https://doi.org/10.1016/S0022-1694(03)00257-9

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, *14*(4), 382–417. https://doi.org/10.1214/ss/1009212519

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. *Journal of Geophysical Research*, *114*, D21101. https://doi.org/10.1029/2009JD011799

Hofstra, N., New, M., & McSweeney, C. (2010). The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. *Climate Dynamics*, *35*, 841–858. https://doi.org/10.1007/s00382-009-0698-1

Isotta, F. A., Frei, C., Weilguni, V., Perčec Tadić, M., Lassègues, P., Rudolf, B., et al. (2014). The climate of daily precipitation in the Alps: Development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. *International Journal of Climatology*, *34*, 1657–1675. https://doi.org/10.1002/joc.3794

Isotta, F. A., Vogel, R., & Frei, C. (2015). Evaluation of European regional reanalyses and downscalings for precipitation in the Alpine region. *Meteorologische Zeitschrift*, *24*(1), 15–37. https://doi.org/10.1127/metz/2014/0584

Journel, A. G., & Huijbregts, C. (1978). *Mining geostatistics* (600 pp.). New York, NY: Academic Press.
Journel, A. G., & Rossi, M. E. (1989). When do we need a trend model in kriging? *Mathematical Geology*, *21*, 715–739. https://doi.org/10.1007/BF00893318
Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of IEEE international conference on neural networks*. Perth, WA: Institute of Electrical & Electronics Engineering, University of Western Australia. https://doi.org/10.1109/ICNN.1995.488968
Kidd, C., Bauer, P., Turk, J., Huffman, G., Joyce, R., Hsu, K., et al. (2012). Intercomparison of high-resolution precipitation products over Northwest Europe. *Journal of Hydrometeorology*, *13*, 67–83. https://doi.org/10.1175/JHM-D-11-042.1
Kidd, C., Dawkins, E., & Huffman, G. (2013). Comparison of precipitation derived from the ECMWF operational forecast model and satellite precipitation datasets. *Journal of Hydrometeorology*, *14*, 1463–1482. https://doi.org/10.1175/JHM-D-12-0182.1
Klok, E. J., & Klein Tank, A. M. G. (2009). Updated and extended European dataset of daily climate observations. *International Journal of Climatology*, *29*, 1182–1191. https://doi.org/10.1002/joc.1779
Knutti, R. (2010). The end of model democracy? *Climate Change*, *102*, 395–404. https://doi.org/10.1007/s10584-010-9800-2
Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., et al. (2014). Regional climate modeling on European scales: A joint standard evaluation of the EURO-CORDEX RCM ensemble. *Geoscientific Model Development*, *7*, 1297–1333. https://doi.org/10.5194/gmd-7-1297-2014
Landelius, T., Dahlgren, P., Gollvik, S., Jansson, A., & Olsson, E. (2016). A high-resolution regional reanalysis for Europe. Part 2: 2D analysis of surface temperature, precipitation and wind. *Quarterly Journal of the Royal Meteorological Society*, *142*, 2132–2142. https://doi.org/10.1002/qj.2813
Ledesma, J. L. J., & Futter, M. N. (2017). Gridded climate data products are an alternative to instrumental measurements as inputs to rainfall-runoff models. *Hydrological Processes*, *31*, 3283–3293. https://doi.org/10.1002/hyp.11269
Lenderink, G., Buishand, A., & van Deursen, W. (2007). Estimates of future discharges of the river Rhine using two scenario methodologies: Direct versus delta approach. *Hydrology and Earth System Sciences*, *11*, 1145–1159. https://doi.org/10.5194/hess-11-1145-2007
Lundquist, J. D., Hughes, M., Henn, B., Gutmann, E. D., Livneh, B., Dozier, J., et al. (2015). High-elevation precipitation patterns: Using snow measurements to assess daily gridded datasets across the Sierra Nevada, California. *Journal of Hydrometeorology*, *16*, 1773–1792. https://doi.org/10.1175/JHM-D-15-0019.1
Lutz, S. R., Mallucci, S., Diamantini, E., Majone, B., Bellin, A., & Merza, R. (2016). Hydroclimatic and water quality trends across three Mediterranean river basins. *Science of the Total Environment*, *571*, 1392–1406. https://doi.org/10.1016/j.scitotenv.2016.07.102
Majone, B., Bertagnoli, A., & Bellin, A. (2010). A non-linear runoff generation model in small Alpine catchments. *Journal of Hydrology*, *385*, 300–312. https://doi.org/10.1016/j.jhydrol.2010.02.033
Majone, B., Bovolo, C. I., Bellin, A., Blenkinsop, S., & Fowler, H. J. (2012). Modeling the impacts of future climate change on water resources for the Gallego river basin (Spain). *Water Resources Research*, *48*, W01512. https://doi.org/10.1029/2011WR010985
Majone, B., Villa, F., Deidda, R., & Bellin, A. (2016). Impact of climate change and water use policies on hydropower potential in the south-eastern Alpine region. *Science of the Total Environment*, *543*(B), 965–980. https://doi.org/10.1016/j.scitotenv.2015.05.009
Michaelides, S., Levizzani, V., Anagnostou, E., Bauer, P., Kasparis, T., & Lane, J. E. (2009). Precipitation: Measurement, remote sensing, climatology and modeling. *Atmospheric Research*, *94*, 512–533. https://doi.org/10.1016/j.atmosres.2009.08.017
Michel, C., Andreassian, V., & Perrin, C. (2005). Soil conservation service curve number method: How to mend a wrong soil moisture accounting procedure? *Water Resources Research*, *41*, W02011. https://doi.org/10.1029/2004WR003191
Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*, 885–900.
Mosier, T. M., Hill, D. F., & Sharp, K. V. (2014). 30-Arcsecond monthly climate surfaces with global land coverage. *International Journal of Climatology*, *34*, 2175–2188. https://doi.org/10.1002/joc.3829
Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models Part I. A discussion of principles. *Journal of Hydrology*, *10*, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6
Nasonova, O. N., Gusev, Y. M., & Kovalev, Y. E. (2011). Impact of uncertainties in meteorological forcing data and land surface parameters on global estimates of terrestrial water balance components. *Hydrological Processes*, *25*, 1074–1090. https://doi.org/10.1002/hyp.7651
Panziera, L., Giovannini, L., Laiti, L., & Zardi, D. (2015). The relation between circulation types and regional Alpine climate. Part I: Synoptic climatology of Trentino. *International Journal of Climatology*, *35*, 4655–4672. https://doi.org/10.1002/joc.4314
Panziera, L., Giovannini, L., Laiti, L., & Zardi, D. (2016). The relation between circulation types and regional Alpine climate Part II: The dependence of the predictive skill on the level of the classification method for Trentino. *International Journal of Climatology*, *36*, 2189–2199. https://doi.org/10.1002/joc.4487
Paul, P. K., Kumari, N., Panigrahi, N., Mishra, A., & Singh, R. (2018). Implementation of cell-to-cell routing scheme in a large scale conceptual hydrological model. *Environmental Modelling Software*, *101*, 23–33. https://doi.org/10.1016/j.envsoft.2017.12.003
Piccolroaz, S., Di Lazzaro, M., Zarlenga, A., Majone, B., Bellin, A., & Fiori, A. (2016). HYPERstream: A multi-scale framework for streamflow routing in large-scale hydrological model. *Hydrology and Earth System Sciences*, *20*, 2047–2061. https://doi.org/10.5194/hess-20-2047-2016
Piccolroaz, S., Healey, N. C., Lenters, J. D., Schladow, S. G., Hook, S. J., Sahoo, G. B., et al. (2018). On the predictability of lake surface temperature using air temperature in a changing climate: A case study for Lake Tahoe (U.S.A.). *Limnology and Oceanography*, *63*, 243–261. https://doi.org/10.1002/lno.10626
Piccolroaz, S., Majone, B., Palmieri, F., Cassiani, G., & Bellin, A. (2015). On the use of spatially distributed, time-lapse microgravity surveys to inform hydrological modeling. *Water Resources Research*, *51*, 7270–7288. https://doi.org/10.1002/2015WR016994
Pilgrim, D. H. (1977). Isochrones of travel time and distribution of flood storage from a tracer study on a small watershed. *Water Resources Research*, *13*, 587–595. https://doi.org/10.1029/WR013i003p00587
Prein, A. F., & Gobiet, A. (2017). Impacts of uncertainties in European gridded precipitation observations on regional climate analysis. *International Journal of Climatology*, *37*, 305–327. https://doi.org/10.1002/joc.4706
Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., et al. (2015). A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Reviews of Geophysics*, *53*, 323–361. https://doi.org/10.1002/2014RG000475
Quintero, F., Krajewski, W. F., Mantilla, R., Small, S., & Seo, B. (2016). A spatial-dynamical framework for evaluation of satellite rainfall products for flood prediction. *Journal of Hydrometeorology*, *17*, 2137–2154. https://doi.org/10.1175/JHM-D-15-0195.1
Rango, A., & Martinec, J. (1995). Revisiting the degree-day method for snowmelt computations. *Water Resources Bulletin*, *31*, 657–669. https://doi.org/10.1111/j.1752-1688.1995.tb03392.x
Rinaldo, A., Marani, A., & Rigon, R. (1991). Geomorphological dispersion. *Water Resources Research*, *27*, 513–525. https://doi.org/10.1029/90WR02501

Rinaldo, A., Vogel, G. K., Rigon, R., & Rodriguez-Iturbe, I. (1995). Can one gauge the shape of a basin? *Water Resources Research*, *31*, 1119–1127. https://doi.org/10.1029/94WR03290

Rodríguez-Iturbe, I., & Rinaldo, A. (1997). *Fractal river basins: Chance and self-organization* (547 pp.). Cambridge, UK: Cambridge University Press.

Roy, T., Serrat-Capdevila, A., Gupta, H., & Valdes, J. (2017). A platform for probabilistic multimodel and multiproduct streamflow forecasting. *Water Resources Research*, *53*, 376–399. https://doi.org/10.1002/2016WR019752

Sadegh, M., & Vrugt, J. A. (2014). Approximate Bayesian computation using Markov Chain Monte Carlo simulation: DREAM(ABC). *Water Resources Research*, *50*, 6767–6787. https://doi.org/10.1002/2014WR015386

Savoy, H., Kalbacher, T., Dietrich, P., & Rubin, Y. (2017). Geological heterogeneity: Goal-oriented simplification of structure and characterization needs. *Advances in Water Resources*, *109*, 1–13. https://doi.org/10.1016/j.advwatres.2017.08.017

Schaefli, B., & Gupta, H. V. (2007). Do Nash values have value? *Hydrological Processes*, *21*, 2075–2080. https://doi.org/10.1002/hyp.6825

Smiatek, G., Kunstmann, H., & Senatore, A. (2016). EURO-CORDEX regional climate model analysis for the Greater Alpine Region: Performance and expected future change. *Journal of Geophysical Research: Atmospheres*, *121*, 7710–7728. https://doi.org/10.1002/2015JD024727

Sperna Weiland, F. C., Vrugt, J. A., van Beek, R. L. P. H., Weerts, A. H., & Bierkens, M. F. P. (2015). Significant uncertainty in global scale hydrological modeling from precipitation data errors. *Journal of Hydrology*, *529*, 1095–1115. https://doi.org/10.1016/j.jhydrol.2015.08.061

Sunyer, M. A., Sørup, H. J. D., Christensen, O. B., Madsen, H., Rosbjerg, D., Mikkelsen, P. S., et al. (2013). On the importance of observational data properties when assessing regional climate model performance of extreme precipitation. *Hydrology and Earth System Sciences*, *17*, 4323–4337. https://doi.org/10.5194/hess-17-4323-2013

Teutschbein, C., & Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology*, *456*, 12–29. https://doi.org/10.1016/j.jhydrol.2012.05.052

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). *Science of the Total Environment*, *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. (2013). Assessing gridded observations for daily precipitation extremes in the Alps with a focus on northwest Italy. *Natural Hazards and Earth System Sciences*, *13*, 1457–1468. https://doi.org/10.5194/nhess-13-1457-2013

U.S. Soil Conservation Service (1964). *SCS national engineering handbook*. Washington, DC: U.S. Soil Conservation Service.

Voisin, N., Wood, A. W., & Lettenmaier, D. P. (2008). Evaluation of precipitation products for global hydrological prediction. *Journal of Hydrometeorology*, *9*, 388–407. https://doi.org/10.1175/2007JHM938.1

Vrugt, J. A., Clark, M. P., Diks, C. G. H., Duan, Q., & Robinson, B. A. (2006). Multi-objective calibration of forecast ensembles using Bayesian model averaging. *Geophysical Research Letters*, *33*, L19817. https://doi.org/10.1029/2006 GL027126

Wang, Z., Zhonga, R., Laia, C., & Chen, J. (2017). Evaluation of the GPM IMERG satellite-based precipitation products and the hydrological utility. *Atmospheric Research*, *196*, 151–163. https://doi.org/10.1016/j.atmosres.2017.06.020

Weissmann, M., Braun, F. J., Gantner, L., Mayr, G. J., Rahm, S., & Reitebuch, O. (2005). The Alpine Mountain-Plain circulation: Airborne Doppler Lidar measurements and numerical simulations. *Monthly Weather Review*, *133*, 3095–3109. https://doi.org/10.1175/MWR3012.1

Wilby, R. L., & Dessai, S. (2010). Robust adaptation to climate change. *Weather*, *65*, 180–185. https://doi.org/10.1002/wea.543

Wu, H., Adler, R. F., Tian, Y., Gu, G., & Huffman, G. J. (2017). Evaluation of quantitative precipitation estimations through hydrological modeling in IFloodS River Basins. *Journal of Hydrometeorology*, *18*, 529–553. https://doi.org/10.1175/JHM-D-15-0149.1

Yang, Y., Wang, G., Wang, L., Yu, J., & Xu, Z. (2014). Evaluation of gridded precipitation data for driving SWAT model in area upstream of three gorges reservoir. *PLoS One*, *9*(11), e112725.

Zolezzi, G., Bellin, A., Bruno, M. C., Maiolini, B., & Siviglia, A. (2009). Assessing hydrological alterations at multiple temporal scales: Adige River, Italy. *Water Resources Research*, *45*, W12421. https://doi.org/10.1029/2008WR007266