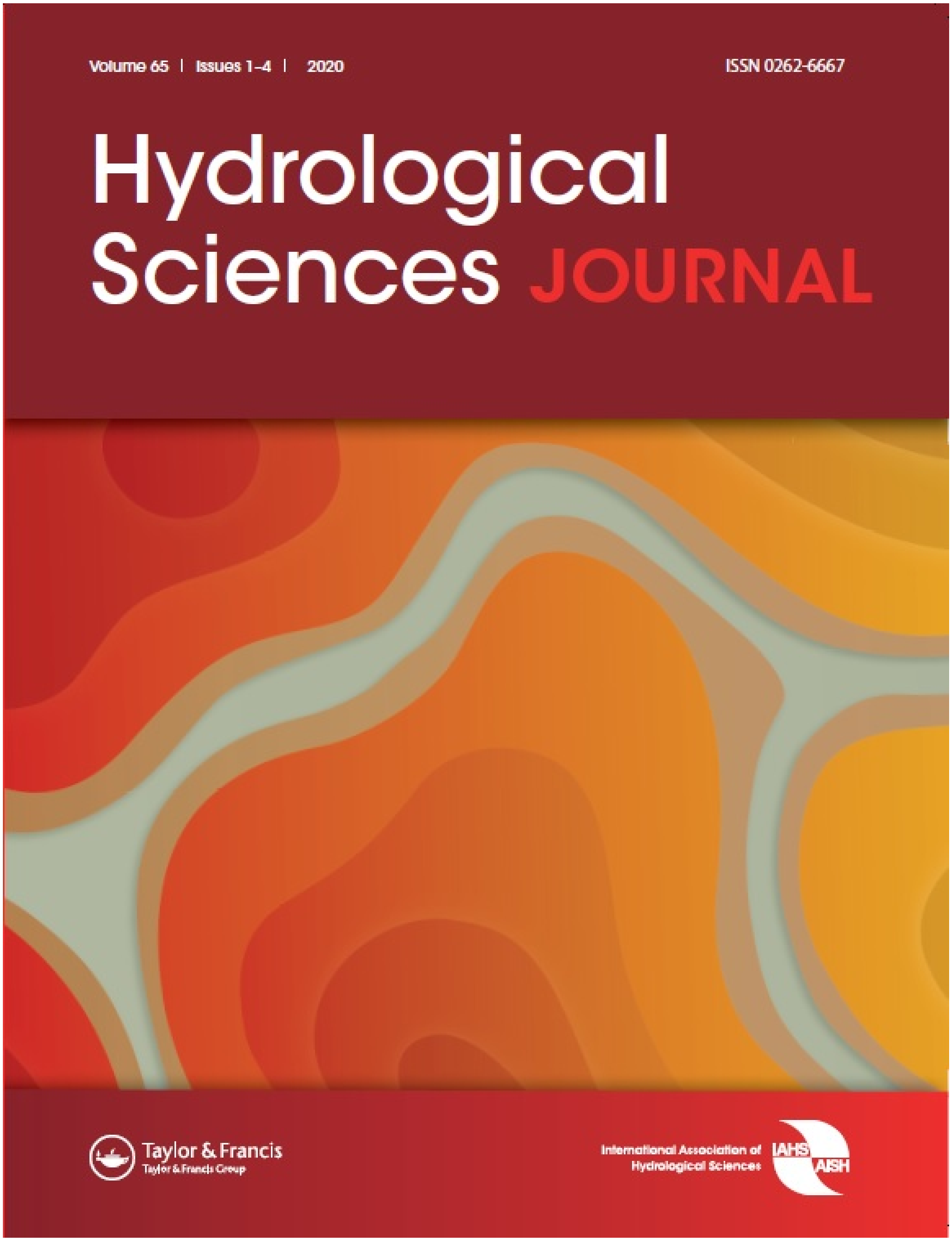

Combining a data-driven approach with seasonal forecasts data to predict reservoir water volume in the Mediterranean

URL: http://mc.manuscriptcentral.com/hsj

area

Journal:	*Hydrological Sciences Journal*
Manuscript ID	Draft
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Francipane, Antonio; University of Palermo Department of Engineering Arnone, Elisa; University of Udine Department of Civil Engineering and Architecture Noto, Leonardo; University of Palermo Department of Engineering
Keywords:	NARX, Mediterranean area, data driven, seasonal forecasts, bias correction, water management in reservoirs

Combining a data-driven approach with seasonal forecasts data to predict reservoir water volume in the Mediterranean area

A. Francipane[a]*, E. Arnone[b] and L.V. Noto[a]

[a] Department of Engineering, University of Palermo - Palermo, Italy; [b] Polytechnic Department of Engineering and Architecture, University of Udine, Udine, Italy

* Corresponding author: Ph.D. Antonio Francipane

Department of Engineering, University of Palermo - Palermo, Italy.

E-mail: antonio.francipane@unipa.it

Telephone number: +39 091238965 20

Combining a data-driven approach with seasonal forecasts data to predict reservoir water volume in the Mediterranean area

In the last years, prolonged droughts and water scarcity have become always more frequent, exacerbating the problem of the artificial reservoirs management in the Mediterranean area. This study proposes a methodology which combines a *Nonlinear AutoRegressive network with eXogenous input* (NARX) data-driven model with Seasonal Forecasts (SFs) data, with the aim to predict the water volume stored in reservoirs at a mid-term scale. The methodology is applied to four Sicilian reservoirs that experienced water scarcity in the recent past. SFs produced at the European Centre for Medium-Range Weather Forecasting are used to force the NARX models. The results show that the NARXs have the capability to reproduce the volumes stored in the considered reservoirs for the investigated period up to four months in advance. The performance of the modeling system strictly depends on: (i) the goodness of climate forecasts and (ii) the strength of the autocorrelation for the water volumes.

Keywords: NARX; Mediterranean area; data driven; seasonal forecasts; bias correction; water management in reservoirs.

1 Introduction

Management of water resources is still a critical issue in Mediterranean areas (Zribi et al. 2020). Droughts and water shortage events frequently put a strain on the water supply systems which serve industrial, civil, and agricultural uses. Jointly, the ever growing water demand (Sanchez et al. 2020) contrasts with an increasing trend of extreme events, such as droughts, heat waves, etc., due to the alteration in climate (Fowler et al. 2021). In this context, it is important to improve the effectiveness and efficiency of the reservoir operation (Ahmad et al. 2014).

A sustainable management of the water supply systems by the water utilities depends on two main capabilities: (i) predicting the future water availability and (ii) adapting promptly and efficiently to the modifications in water resources.

In the case of the Mediterranean area, one of the main multi-purpose water resources is provided by artificial reservoirs. Predicting well in advance future reservoir volume is one of the critical engineering problems to guarantee an efficient water supply planning for water utilities (Awchi 2014, Hassan et al. 2015). Common methods to do this rely on statistical approaches based on condition of stationarity of the climate variables involved. Recently, methodologies based on machine learning (ML) approaches, as optimization algorithms in decision making processes, are more commonly used (Ahmad et al. 2014, El-Shafie et al. 2007, Niu and Feng 2021, Rozos 2019, Yu et al. 2017). More specifically, Artificial Neural Networks (ANNs) have been used to derive operational strategies (Chaves and Chang 2008) for water supply and to identify the optimal sequence of reservoirs water release. Non-linear models like ANNs are capable to detect complex relationships between input and output data series, which are typical of the complex and dynamic nature of hydrological processes. More specifically, to deal with time series, dynamic or recurrent neural networks (RNNs) are preferred to other models, because all layers have feedback connections, preserve and remember the short- and long-past information, i.e., time delays, leading to the perception of temporal pattern of hydrological time series (El-Shafie et al. 2012). Among these, Nonlinear AutoRegressive networks with eXogenous input (NARXs) are widely used for time series forecasting, since their capability in learning long-time dependences among time series, reach faster convergence, and a better ability to generalize (Hadiyan et al. 2020) as compared to other recurrent neural networks. For this reason, NARXs are widely preferred in modeling inflow and outflow reservoir forecasting. In such a context, Yang et al. (2019) provide a comparison among different RNNs, including NARX, that demonstrates the effectiveness of RNNs in reservoir operations. In these studies, precipitation and temperature are usually the two main climate variables involved in the assessment of the reservoir forecast volume.

On the other end, alterations in climate have further stressed the problem of water scarcity in Mediterranean area, more frequently hit by prolonged droughts and short-duration extreme precipitation (Treppiedi et al. 2021, Forestieri et al. 2018). Indeed, increase in intensity and frequency of extreme precipitation events, as well as prolonged droughts and heat waves, has been recognized by different works (Arnone et al. 2020a, WWA 2017, Bonaccorso et al. 2015, Caloiero et al. 2018), thus making even more complex the management of water resources. In a context of probabilistic assessment of the possible climate anomalies, the use of the seasonal forecasts (SFs) data may offer a powerful tool for guiding a strategic planning of the resources across several climate-sensitive sectors (De Felice et al. 2015, Essenfelder et al. 2020, Viel et al. 2016).

More in details, SFs are predictions of climate variables covering up to a 6-month period ahead from initial conditions. As the weather forecasts, they are produced with numerical models of the climate system; conversely, they predict the anomalies with respect to the average, by simulating the processes of both the slow and the fast components of the climate system. Despite the not negligible uncertainty associated with such predictions, it has been demonstrated that they can provide important indications in the fields of drought-risk assessment and in the mid-term reservoir management (Arnone et al. 2020b, Buontempo et al. 2018, Crochemore et al.

52 2017, Viel et al. 2016). Buontempo et al. (2018) provide an overview on potential users of seasonal data, by promoting the use of climate information for decision support in the context of the EUPORIAS projects. Arnal et al. (2018) exploit the advantage of SFs in streamflow forecasting, within the European Flood Awareness System (EFAS). Arnone et al. (2020b) developed an early warning system tested on two Mediterranean islands to assess the probability of drought occurrence in the future based on the seasonal forecast of precipitation data. A further example is given by Peñuela et al. (2020), who evaluate the potential use of a real-time optimization system informed by seasonal forecasts in a water supply system in the UK. Despite all these works, it is worth to highlight, however, that the use of seasonal forecasts in hydrological applications is still rare and yet non-operational.

In this study, the potentiality of both NARX model and SF data are exploited to develop a system able to predict the reservoirs volume and water level at the mid-term scale, using as inputs to the NARX the monthly precipitation and monthly air temperature provided by the SF dataset. The case study is carried out for four Sicilian reservoirs, i.e., Piana degli Albanesi, Poma, Rosamarina, and Scanzano, which are strategic for the water supply of the Metropolitan City of Palermo. At the turn of 2017 and 2018, these reservoirs experienced very low water levels due to the records in extreme heat wave and precipitation anomaly that have hit the South Italy in 2016 and 2017 (ISPRA 2016, SIAS 2016, WWA 2017). The problems experienced by the water supply system of the city of Palermo led the Italian central government to declare the state of emergency.

The aim of this research is then to verify how well and how long in advance is possible to provide a reliable prediction of the reservoirs' volumes, given the forecasts of temperature and precipitation climate variables, thus providing a valuable tool to the interested water utilities which manage the integrated water service of the Metropolitan City of Palermo.

2 Methods

2.1 Nonlinear AutoRegressive network with eXogenous inputs Architecture - NARX

A NARX is a type of ANN based on the linear autoregressive model commonly adopted for input-output modeling of nonlinear systems. According to this model, the value of the dependent variable at the time t, $y(t)$, is calculated by means of a linear regression as follows:

$$y(t) = f(y(t-1), y(t-2), \ldots, y(t-m), x(t-1), x(t-2), \ldots, x(t-n)) \qquad (1)$$

where $y(t-i)$ and $x(t-i)$ are the previous values of the output and the previous values of an independent (exogenous) input, respectively. The indices m and n are the so-called feedback and input delays, respectively, which define the input data that are used to predict output of current time series. If the input delay is i, it implies that the input value at i time steps before is used to predict the current output and the same applies for the feedback delay. This implies that an input delay of 0 is allowable but a feedback delay of 0 is not. Because NARX performances depend also on the values of input and feedback delays, the correct assessment of n and m is extremely important. In this research, the variable $y(t)$ is represented by the reservoir volume, which depends on the climate variables that are involved in the water balance of the reservoir, i.e., precipitation, temperature (that controls the evapotranspiration), and controlled outflows, all represented by the input variable $x(t)$. The input delay, which defines the length of the $x(t-i)$ variable, is representative of the time within which the considered input variables affect the volume at time *t*.

As well as ANN, a NARX consists of some layers, namely an input, an output, and one or more hidden layers, fully connected to each other. Neurons within a layer are connected by weighted links to every neuron of the successive layer; when the neurons in the hidden layers receive the input signals, convert them through an activation (or transfer) function, and then transfer the information to the next layer. The learning ability of a NARX mainly depends on its architecture, the training function, and the number of neurons in the hidden layer (Arnone et al. 2014).

Figure 1 shows a generic architecture of a NARX, where it is possible to observe the input, hidden, and output layers. Generally, in a neural network, the input $x(t)$ is weighted with an appropriate weight (w within the square); the sum of the weighted inputs and the bias (b within the square) forms the input to a transfer function (f within the box), which produces the neuron output as $f(wx + b)$. As compared to a classical ANN, in a NARX comes into play also a regressive component by means of the input (1:n within the circle) and feedback (1:m within the circle) delays, in which the estimated output can be fed back within the NARX and connected to the appropriate input to estimate the next output value.

The creation of a NARX generally consists of two phases. Since the true output is available, in the first phase, a series-parallel architecture, also called *open loop network*, is used to train the NARX; in this phase the true output is used instead of feeding back the estimated output. In the second phase, the NARX is converted from the series-parallel configuration to a parallel configuration, also called closed loop network (see Figure 1 for an example), which is useful for multi-step-ahead prediction. In this phase, each estimated output is fed back within the NARX and connected to the appropriate input to estimate the next output value. Generally, all the training is done within an open loop, including the validation and testing steps, and only when the NARX has been trained it is transformed into a closed loop for multistep-ahead prediction.

2.2 Assessment of model performance

The performances of the NARX models here developed are assessed by means of the Nash–Sutcliffe efficiency (NSE) coefficient (Nash and Sutcliffe 1970) and the Root Mean Square Error (RMSE) coefficient calculated for the observed and simulated volumes.

126 ### 2.3 *Seasonal Forecast - SF*

127 Seasonal forecasts (SFs) are gridded data provided by climate numerical models (Hoskins 2013),
128 which allow to obtain ensembles of forecasts of climate variables covering a time window up to
129 6-months ahead in time starting from an initial date and initial conditions.

130 The slowly varying components of the climate system provide the sources of
131 predictability at seasonal time scales; these components can act as boundary forcings for the
132 troposphere and subsequently affect local weather and climate after some time lag (Pyrina et al.
133 2021). Starting from slightly different initial states, several predictions are performed by the
134 numerical model and their results are the members of an ensemble of predictions. The
135 uncertainty in the forecasts is implied in the ensemble, whose members might differ significantly
136 in the future ahead, due to the different initialization.

137 A peculiarity of the seasonal climate data is the lead time, which is the time distance, in
138 months, between the release of the forecast and the occurrence of the predicted phenomena. It
139 ranges between 0 and 6 months and will be indicated hereinafter as LT_i, where i is the target
140 month (i.e., the i-th month after the release); LT_0 will indicate the month of the release (present
141 time).

142 Datasets are produced and released by different climate centers. In this study, the last
143 generation of SF system, the System 5 (SEAS5), released by the European Centre for Medium-
144 Range Weather Forecasts (ECMWF) is used, which can be retrieved through the data access
145 system of Copernicus Climate Data Store (CDS). The dataset has a global coverage, with a
146 spatial resolution of 1°x1°, and includes forecasts in real-time (hereinafter SRT) and hindcasts
147 (hereinafter SH); SRTs start from 2017, while SH are initialized in the period 1986-2016. SRTs
148 consist of 51-member ensembles, generated at different atmospheric initial conditions, whereas
149 hindcasts have a 25-member ensemble. SH data are useful to calibrate and correct the dataset, as
150 in the case of the bias correction discussed in section 5.2.1. An accurate description of the dataset
151 can be read in Johnson et al. (2019).

152 Skill in the seasonal predictions depends on the type of climate variables, location, and season. In particular, temperature is one of the climate variable most successfully reproduced (Clark et al. 2017, Doblas-Reyes et al. 2013).

155 ### *2.4 Bias correction methods*

156 SFs are typically affected by systematic and random model errors. This poses a problem for using these data as input for hydrological impact studies. One possible solution is to apply a bias correction to the SFs by means of observed data. Several bias correction methods have already been applied in weather forecasting under the name of model output statistics (MOS) about five decades ago (Glahn and Lowry 1972, Klein and Glahn 1974).

161 Typical correction approaches aim at correcting the systematic error (bias) in SF variables by applying a transformation algorithm and are therefore named bias correction methods. The concept is based on the identification of possible biases between observed and simulated climate variables. A common assumption of most bias correction methods is stationarity, or time invariance, of the model errors. This implies that the empirical relationships in the correction algorithm and its parametrization for current climate conditions do not change over time and are also valid for future conditions. This assumption is, however, likely not met under changing climate conditions (Ehret et al. 2012, Maraun 2012, Maraun et al. 2010, Vannitsem and Nicolis 2008).

170 More in details, transformations attempt to find a function h that maps a modeled variable, P_m, such that its new distribution equals the distribution of the observed variable, P_o. Following Piani et al. (2010), this transformation can in general be formulated as $P_o = h(P_m)$.

173 The quantile–quantile relation of observed and modeled precipitation (or temperature) can be modeled using parametric or non-parametric transformations. In the first case, theoretical distributions are used to achieve a statistical transformation, while in the second case, the empirical cumulative distribution function of observed and modeled variables are usually used

instead of assuming parametric distributions. Here a parametric transformation and three different non-parametric transformations are used for the bias correction of the SF data. Excluding the parametric transformation, which performed worse than the non-parametric methods, for both the precipitation and the air temperature, there is not a method that performed significantly better than another one. Therefore, for the sake of brevity, only the analyses conducted by using a non-parametric method for the bias correction of the SF dataset are reported here. The method is a quantile mapping method that fits a smoothing spline to the quantile-quantile plot of observed and modeled time series (hereinafter referred to as SSPLIN); the method then uses the spline function to adjust the distribution of the modeled data to match the distribution of the observations (Kouhestani et al. 2016). For more information about the method the reader can refer to Gudmundsson et al. (2012).

All the methods have been implemented in the R language by means of the package *qmap*, which is available on the Comprehensive R Archive Network (http://www.cran.r-project.org/).

3 Study Area and Datasets

3.1 Case studies: the Piana degli Albanesi, Poma, Rosamarina, and Scanzano reservoirs

The Piana degli Albanesi, Poma, Rosamarina, and Scanzano reservoirs are strategic for the water supply to about one million citizens of 23 municipalities of the Metropolitan City of Palermo (Sicily, Italy).

The Piana degli Albanesi reservoir originates from a dam that interrupts the natural flow of the Belice Destro river. It has a total volume of 32.80 Mm^3 and is mainly used for energy production but also for irrigation and water supply of the city of Palermo. The Poma reservoir has been created by means of the barrage of the Jato river and has a total volume of 72.50 Mm^3. It is mainly used for irrigation (about 10 Mm^3/year) and to supply water to the city of Palermo (about 25 Mm^3/year). The Rosamarina reservoir, with a total volume of 101.20 Mm^3, is the biggest of the four reservoirs considered here. The reservoir has been realized interrupting the San Leonardo river with the Rosamarina dam and supplies about 30 Mm^3/year of water to the city of Palermo. The Scanzano reservoir has been created by interrupting the Scanzano and Rossella rivers with a couple of dams, namely the Scanzano and Rossella dams, and has a total volume of 18 Mm^3.

All the previous information has been provided by the integrated water service company AMAP S.p.A., which manages the integrated water service in 35 municipalities of the Metropolitan City of Palermo, the *Autorità di Bacino della Regione Sicilia* (Basin Authority of the Sicilian Region; hereinafter referred to as AdB), which manages the Poma, Rosamarina, Scanzano reservoirs, and the company ENEL S.p.A., which manages the Piana degli Albanesi reservoir for energy production purposes.

All the reservoirs are in between the parallels 37°N and 38°N, very close to the 38°N parallel, and the meridians 13°E and 14°E (Figure 2), within the area here defined as cell of interest (COI) #6 with reference to the SF grid (see inset at the top right in Figure 2). Blue contours in the insets at the bottom of the Figure 2 show the extension of the reservoirs when the maximum volume is reached; red contours, on the contrary, indicate the extension reached by the reservoirs in February 2018, after the drought occurred in 2016 and 2017.

3.2 Climate dataset: the AdB and SF datasets

3.2.1 *Reference network: AdB dataset*

Precipitation, temperature, and volume data for assessing the goodness of the SF dataset and calibrating the NARXs for the four examined reservoirs have been collected from the dataset of the AdB. The meteorological gauges of the network are reported in Figure 2 as blue points.

The AdB network includes 195 stations equipped with a tipping-bucket rain gauge and a thermometer. The stations are rather homogeneously distributed over the entire island with an average density equal to about 130 km^2/gauge. Data are retrieved with a time resolutions of 10 or 30 min. For each reservoir, the AdB has provided the monthly precipitation and the mean monthly air temperature spatially averaged over the studied area. For the sake of simplicity, previous variables will be hereinafter referred to as monthly precipitation and monthly temperature. In addition, the AdB has provided the monthly volume stored within each reservoir as well, which represents the reservoir volume to be estimated by means of the NARX. Monthly precipitation and monthly temperature data cover a period of 32 years, from January 1988 through December 2020, while the stored volumes range between April 1995 and December 2020 (i.e., about 25 years) for the reservoirs of Piana degli Albanesi, Poma, and Scanzano and between February 2002 and December 2020 (i.e., about 18 years) for the Rosamarina reservoir.

Water withdrawals to supply water system should be included in the modeling system as input data as well; however, this type of data is present in the AdB dataset only for a small period. Nevertheless, this limitation is overcome within the NARX structure itself that allows to take indirectly into account the water withdrawals.

Figure 3 shows the time series of normalized monthly volumes stored within the four reservoirs; in the gray shaded box, it is possible to notice the effects of the drought that affected Sicily in between 2017 and 2018 on the reservoirs' volumes. In addition, data in Figure 3 show two more critical droughts events occurred in late 2002 and 2009.

3.2.2 *Seasonal Forecast dataset*

The SF dataset used here consists of the monthly total precipitation and monthly average air temperature retrieved for the twelve cells that cover the entire Sicily (see inset at the top right in Figure 2). Specifically, among the twelve cells, the COI #6 (Figure 2) is the one that covers the analyzed reservoirs and is characterized by a more homogenous climate system given that it overlaps mostly the terrain system.

Figure 4 reports an example of ensemble of SF for the selected COI #6 released in January and June 2019 and predicting the six months ahead in time for both precipitation and air temperature. A comparison with the observed series from the AdB dataset is reported as well. As it is possible to observe from the boxplots, the precipitation (Figure 4a and b) shows a higher variability as compared to the air temperature (Figure 4c and d). From the comparison with observations, it is possible to notice that the capability of the SFs to predict the observed data depends on the case study. Additionally, referring to the case studies, the SFs are capable to predict better the observed data for the summer (i.e., June, July, and August), since the Sicilian climate is usually characterized by summers almost rainless and with high air temperatures.

4 Setup of NARX model

The phases involved in the development of a NARX mainly consist in the definition of the network design, in terms of variables, structure and algorithms, and the network training.

Using the observed data provided by the AdB, the NARX has been trained to reproduce the historical responses of the four reservoirs in terms of stored volumes. The use of the NARX model along with the SF data allows for a probabilistic forecast of reservoir volumes in the mid-period. The forecasted volumes can be verified by means of a comparison with the historical data of the volumes. The variables are assessed at the monthly scale, which defines the time step of the NARX.

The NARX models have been developed within the neural networks toolbox included in the software Matlab 2021a.

4.1 Design of NARX architecture

Figure 5 shows a flowchart of all the phases that have been followed to implement the NARX models, one for each reservoir.

274 As previously mentioned, the involved variables are the monthly volume stored within
275 each reservoir as target variable, $y(t)$, and the monthly precipitation and monthly temperature as
276 exogenous input variables, $x(t)$. The use of an autoregressive model allows to implicitly
277 consider the water withdrawals provided to the water supply system, which affect the reservoir
278 water balance and, in turn, the volume at time t; by considering the previous values of water
279 volume as input, the NARX is capable to learn during the training for any variation in the stored
280 volume that occurs over the time of the feedback delay. Clearly, the procedure is valid under the
281 assumption of no significant changes in water withdrawals, i.e., under the assumption of
282 Business-as-Usual (BaU) (Fei and Shuang-Qing 2012), i.e., there are no significant changes in
283 people's attitudes and priorities, agriculture, and energy industry scenarios, and thus in water
284 demand and supply (EIA 2010).

285 The collected data have been first controlled to identify possible missing data, duplicates,
286 outliers, and errors such as negative values in precipitation and/or volumes. The input delay has
287 been defined by means of an input/target cross-correlation analysis between monthly
288 precipitation and monthly stored volumes by means of the Pearson correlation (Pearson 1895).
289 Specifically, it has been defined as the time-lag value associated to the highest cross-correlation
290 value. Since the input delay can be strictly influenced by several factors, such as the soil
291 characteristics of the basin and its specific hydrological response, it is site dependent. For the
292 feedback delay, instead, a target-target auto-correlation analysis has been applied to the monthly
293 stored volumes. In this case, it has been used the auto-correlogram test. The characteristics of the
294 NARX models for the four analyzed reservoirs are summarized in Table 1.

295 The NARX architecture uses a tan-sigmoid transfer function in between the input and
296 hidden layers and a linear transfer function in between the hidden and output layers.

297 The neurons within the hidden layer have been set by choosing the number of neurons
298 that returns the best calibration of the NARX models in terms of NSE and RMSE (Table 1). For
299 further details, the reader is referred to section 4.2.

As it is possible to observe from the values in Table 1, for all the NARX models, the correlation analysis has returned a feedback delay of one month, which means that the monthly stored volume strongly depends on the value that was stored on the previous month; the input delay is instead equal to five months for the Rosamarina reservoir and to four months for the other three cases.

4.2 *NARX calibration*

The period of calibration ranges between February 2002 and December 2020 for the Rosamarina reservoir and between April 1995 and December 2020 for the other three reservoirs. For each reservoir, the calibration has made possible to identify the best NARX architecture to be used during the forecast phase with the seasonal forecasts. Each calibration is made of a different combination of neurons in the hidden layer (from 1 through 50 neurons for a total of 50 combinations) and training function (2 combinations) for a total of 100 combinations.

For each reservoir, during the calibration phase, the original dataset has been separated into three subsets: the 70%, 15%, and 15% of the original dataset have been used for the training, validation, and test of the NARX, respectively. Two different training functions, namely the Bayesian Regularization backpropagation and the Levenberg-Marquardt backpropagation, have been considered. Moreover, to guarantee the replicability of the results, for each calibration round, the control random number generator has been initialized always with the same seed.

Figure 6 reports an example of the performances returned by the NARX model during a calibration round for the Piana degli Albanesi reservoir and using the Levenberg-Marquardt backpropagation. Figures 6a through 6d show the results of calibration in terms of regression analysis during the training, validation, test, and the entire calibration, respectively; the y-axis shows the simulated volumes, the x-axis reports the targets (i.e., observed volumes), the dotted line (Y = T) is the perfect agreement line, and the blue, green, red, and black solid lines are the fit lines obtained during the training (Figure 6a), validation (Figure 6b), test (Figure 6c), and the entire calibration (Figure 6d), respectively. As it is possible to notice, in all the phases, also the Pearson's correlation coefficient, R, is always close to 1, thus indicating a strong positive relationship and a good performance of the NARX during all the calibration phases and in the overall.

For each reservoir, the best performing NARX has been obtained by calibrating the number of neurons in the hidden layer. Table 1 shows the calibrated hidden neurons corresponding to the best performing NARX model for each reservoir, along with the computed model performance metrics (i.e., NSE and RMSE). In all the case studies, the best calibration has been obtained considering the Bayesian Regularization backpropagation training function.

Results obtained considering the calibrated NARXs, for each reservoir, are reported in Figure 6e, where it is possible to notice how the calibrated models can efficiently predict volumes stored within each reservoir.

5 Results

For each reservoir, the best calibrated NARX model has been forced with the SF data to estimate, starting from an initial month, the stored volumes for the six months ahead (i.e., forecasted volumes). Specifically, once chosen the release month of the SF, it is used to run the simulations and to assess the target variable up to six months after the release month, by considering the six lead times of the SF release (i.e., from LT_0 to LT_6). As an example, this means that the water volume stored in July can be predicted throughout the LT_6 of the SF released in January, or the LT_5 of the SF released in the February, or LT_4 of the SF released in the March, and so on until June, with LT_1.

The performances of the forecast models have been assessed in two different configurations: (i) by using the SF dataset as it is and (ii) by applying to the SF data the bias correction method described in section 2.4. For each simulation, the initial conditions are defined

349 by forcing the NARX with the observed monthly data up to the month at which the forecast starts.

5.1 *Forecast modeling with original SFs*

Simulations with the original uncorrected SF data have been run for the entire period ranging from January 2017 through April 2020.

Since the performances of the NARX models depend on the reliability of the SFs, the capability of the original SFs to reproduce the observed data is here qualitatively analyzed. As previously mentioned, Figure 4 shows the comparison between the SFs (with lead times from LT_0 to LT_6) and the observed monthly precipitation (Figure 4a and b) and monthly temperature (Figure 4c and d) for the January 2019 (Figure 4a and c) and June 2019 (Figure 4b and d) for all the study cases. The plots show that the SFs are capable to better reproduce both the values of precipitation and temperature for the summer months, also in the case in which these are the results of a projection of five or six months forward in time (e.g., forecasts at LT_5 and LT_6 in Figure 4a and c, which represent the forecasts for months of June and July 2019). In this case, indeed, the variability of the ensemble members is very tiny and the climate models correctly forecast the low precipitations and high air temperature which are typical of the summer season in Sicily. Additionally, it can be observed that the uncertainty of the ensemble increases either with the lead time or over the rainy months, as previously discussed.

Figure 7 shows the results of the NARX model for the 2019 and for the Rosamarina reservoir, while similar plots for the remaining reservoirs are shown in Figure SM1 through SM3 in the supplementary material. Each subplot shows, for the period that ranges from the indicated *i*-th month to the next six months, the observed monthly volume (blue solid line), the monthly volume simulated forcing the NARX with the observed input data provided by the AdB (red dashed line), and the ensemble of monthly volumes (boxplots) obtained forcing the NARX with the SF data of the *i*-th release month and all its lead times (i.e., from LT_0 to LT_6). The ensemble of the 51 models of SF data is represented by means of boxplots where the red plus markers indicate outliers, and the red horizontal lines and the blue plus markers represent the median and the mean of the volumes forecasted with the trained NARX and the SF data, respectively.

Simulated volumes returned by the NARX forced with observed data (red dashed lines) reproduce accurately the observed stored volumes (blue solid lines) over the entire shown period; only two significant exceptions can be noticed in October 2019 and May 2020, when the stored volumes are slightly underestimated, with an error of about 5 Mm^3 in both the cases. The volumes forecasted by the NARX combined with the SF data (boxplots) are characterized by the variability inherited from the input data ensembles. It can be observed that the interquartile range varies across both (i) the lead times, from LT_0 to LT_6, starting from the release month, and (ii) the *i*-th month of the year, given a fixed lead time. More in details, the variability is significant (up to ~70 Mm^3) when $LT > LT_3$ and the release month falls in the autumn and winter periods, i.e., from September to February. When the release month of the forecast falls in between March and August, the interquartile range is less variable across the LT (see March and April) and tinier (see May through July) with few outliers. This means that forecasts released in summer are affected by a less uncertainty. The overall trend of the forecasted volumes is synthesized by means of the ensemble mean (blue plus markers) or median (red horizontal lines) values. The most significant differences between observed and simulated stored volumes can be observed in February and March for $LT \geq LT_3$, October and December for $LT \geq LT_5$ and November for $LT = LT_6$.

To assess the overall performances of the NARXs in forecasting the dynamic of stored volumes as a function of the LT value over the entire period of simulations, Figure 8 shows an overview of the RMSE of SFs for the precipitation (Figure 8a) and air temperature (Figure 8b), and NARX model output (Figure 8c) for all the possible combinations of starting months of simulation and LTs for the Rosamarina reservoir. Specifically, the x axes report the release month of the SF, the y axes indicate the LTs values, while each cell denotes the relative performance using a graduated color scale.

As it is possible to observe from both the overviews of precipitation (Figure 8a) and air temperature (Figure 8b), regardless of the considered LT, the SFs are capable to better reproduce the observed values for the spring/summer as compared to the remaining months of the year. This means that, during the winter period (from November to February), SFs of precipitation and air temperature realize reliable predictions of summer values up to five or six months in advance.

Looking at the overview for the simulated volumes (Figure 8c), instead, it is possible to notice that for LT_0 the NARX is capable to successfully reproduce the stored volumes dynamics. As the LT increases, the NARX performs worse, especially during the autumn/winter months. For such months the NARX reproduces the reliable stored volumes only for lower LTs (i.e., $LT < LT_3$ from September to November; $LT < LT_2$ from December to April). The upper-left and upper-right corners correspond with the worst performances of the model. On the contrary, during the summer, the NARX is capable to reproduce well the volumes within the reservoir up to five months in advance. This different behavior, as compared to the overviews of the SF performances, is mainly due to the autoregressive component of the model, which exploits the observed volumes at the previous month, thus making possible to always have a good prediction of the volumes at the lower LTs. Additionally, during the spring/summer time the good performances in forecasting the volumes extend to the higher LTs as well, since the low values of RMSE at the lower LTs causes less propagation of error in the following months.

From the analyses of simulations carried out for the other reservoirs (shown in Figures SM4 through SM6 in the supplementary material), it is possible to affirm that the NARX model performs better in some case studies than in others, although the models are forced always with the same SF data. This can mainly depend on the fact that the SF data, although all the reservoirs lie within the same SF cell, can be more representative of the real climatic conditions of some reservoirs than others, as it is possible to notice from the example reported in Figure 4, as well.

5.2 *Forecast modeling with bias corrected SFs*

5.2.1 *Bias correction of the SF dataset*

The SF dataset of monthly precipitation and air temperature has been bias corrected through the SSPLIN method presented in section 2.4. Specifically, it has been used a quantile step equal to 0.01; with reference to the precipitation, the smoothing spline is only fit to the fraction of the CDF corresponding to observed wet days ($P_o > 0$) and modeled values below this are set to zero.

The values of monthly precipitation and monthly air temperature of each model of the SF ensemble (i.e., 25 models for the SH and 51 for the SRT) are corrected by using the AdB dataset as reference and then the mean of each ensemble is ultimately considered to force the NARX model. The correction covers the period 1995-2020, by using the SH from 1995 to 2016 and the SRT from January 2017 to April 2020.

For all the case studies, Figure 9 shows the quantile-quantile plots (q-q plots hereinafter) for the monthly precipitation (Figure 9a) and monthly temperature (Figure 9b) after the application of the SSPLIN. The gray circles in Figure 9 indicate the q-q plot of observed (i.e., AdB) and modeled (i.e., SH and SRT) data for a quantile step equal to 0.01.

Looking at the q-q plots of the uncorrected SF data (gray circles), it is possible to notice that the monthly precipitation (Figure 9a) is always underestimated for all the case studies, showing the inability of the models in correctly reproducing the precipitation, especially for the higher values. On the contrary, monthly temperature (Figure 9b) is always overestimated by the SF data, especially at the lower temperatures, for all the reservoirs. Only for the higher temperatures (higher than about 25 °C) of the case study of the Poma, the SF data slightly underestimate the observed ones. Very likely, this combination between underestimation of precipitation and overestimation of air temperature is one of the reasons of the underestimation of volumes stored within the reservoirs often observed in the previous analyzed cases (see Figures 7).

After the application of the SSPLIN method, the q-q plot (red lines) for both the monthly precipitation and the monthly temperature approaches the perfect agreement line (Figure 9), thus indicating a better description of observed data by the SF dataset. Moreover, for all the four reservoirs, the q-q plots seem to show a better agreement for the monthly temperature as compared with the monthly precipitation, as demonstrated by other studies (Clark et al. 2017, Doblas-Reyes et al. 2013).

5.2.2 *Forecasts of NARX with bias corrected SFs*

The results of the NARX model forced with the bias corrected SFs for the Rosamarina reservoir are shown in the plots reported in Figures 7, where black dashed lines refer to the NARX forecasts of stored volumes obtained after correcting the SF data. One can observe an improvement of the reproduction of the volumes stored within the reservoirs. In most cases, the black dashed lines approach and follow the line of the observed volumes also when the results obtained with the uncorrected SF data were not that good. Only for the months of March and April 2019, obtained as LT_4 and LT_5 of the release of November 2019 and LT_3 and LT_4 of the release of December 2019, respectively, the results worsen as compared to those obtained with uncorrected SFs. The results depend on the degree of the LT considered as well, showing an agreement between the simulated and observed volumes as stronger as lower is the LT considered.

For the sake of brevity, Figure 10 provides a resume of the results obtained for the Rosamarina reservoir when the NARX is forced with the bias corrected data. Starting from the top, each subplot shows the comparison between the observed volumes (tick blue solid line) and the simulated volumes (red dotted line with x marker).

As it is possible to observe, the model catches the behavior of the observed volumes, especially for the lower LTs. Even during the drought period in between December 2017 and April 2018 (gray boxes in Figures 3 and 10), the model performs well, making a reliable forecast of the stored volumes in the reservoir for the medium LTs (e.g., LT_3 - LT_4). For complete information, the remaining case studies, which are shown in Figures SM7 through SM9 in the supplementary material, for the same drought period, performed less well than the Rosamarina reservoir, probably because of a faster growth in the registered volumes after the drought period that is not correctly caught by the autoregressive component of the model. Nonetheless, all the models are still capable to reproduce the water scarcity period due to the drought.

Looking at the performances of the NARXs over the entire simulated period, the bias correction of the data makes it possible to obtain higher performances especially for the higher LTs (e.g., $LT > LT_3$). In this regard, Figure 11 shows the NSE values for different LTs when the NARXs are forced with the mean values of SFs for both the uncorrected and bias corrected data. For $LT < LT_3$, except for the Scanzano reservoir, the NSE of the bias corrected data is always higher than 0.6, thus indicating that the fit between the observed and simulated data is “acceptable” to “good”, according to the criteria provided in Moriasi et al. (2007). Table SM1 in the supplementary material reports the values of NSE shown in Figure 11.

Figure 12 shows the overall performances of the SF and NARX model for the Rosamarina reservoir when the SF data are bias corrected, for all months and LT_{S}. Specifically, by comparison with Figure 8a and 8b, it can be noticed a general improvement in the prediction of precipitation and air temperature (Figure 12a and 12b, respectively).

With reference to the volumes (Figure 12c), the comparison with the uncorrected SF forcing data (Figure 8c) highlights a decreasing in the RMES during the late winter and spring at LT higher than $LT_{3,}$ thus indicating that the bias correction improves the predictability of the stored volumes, while during the summer the performances are still good. Moreover, there is also an increase in performances during the autumn and winter, even if less considerable than the previous one. Generally, for each month, performances increase for all the LTs. This effect is obviously due to the correction of monthly precipitation and monthly temperature with the observed data. Overall, it is possible to assert that in the case of bias corrected SFs, the results provided by the NARX are more or less reliable up to LT_3.

Looking at the results for the other reservoirs (shown in Figures SM4 through SM6 in the supplementary material), it is still possible to observe the same patterns than in Figure 12, even though there is not a best combination of starting month and *LT*; this mainly depends on the specific case. In this perspective, Figure 12 provides a complete characterization of the capability of the NARX in predicting the volumes within a reservoir.

5.3 Comparison among all models

Overall, NARX models developed for the analyzed reservoirs have been forced with three different climate time series, i.e., the observations provided by the AdB, and the mean of uncorrected and bias corrected SF data. Figure 13 summarizes the overall results by means of normalized Taylor diagrams at different *LT*s. The diagram summarizes the distance between observed and modeled time series in terms of normalized standard deviation, correlation coefficient (CC), and normalized root mean square difference (RMSD). The green square in Figure 13 refers to the observed values for which the normalized standard deviation is equal to 1, the radial distance from the green square quantifies the centered RMSD normalized by the standard deviation of observed data (i.e., monthly volumes), the azimuth and the radial distance from the origin quantify CC and standard deviation normalized by the standard deviation of observations, respectively.

Looking at the first panel of Figure 13, it is clearly shown that, for all the reservoirs, the NARX models are capable to reproduce very well the observed volumes when forced with observed precipitation and air temperature provided by the AdB (red markers). Only in the Scanzano case (x marker) performances in terms of RMSD are slightly worsen (i.e., RMSD greater than 0.15). These results are independent from the *LT* and thus they are repeated in each panel. In all the cases, indeed, the red symbols are very close to the green square that indicates the observed data. Moving from lower to higher *LT*s, generally, it is possible to notice that all the symbols shift upwards, moving away from the green square, thus indicating a lower performance of the NARX models, as previously observed.

All the blue markers, which denote the results of NARX model forced with the uncorrected SFs, are shifted upwards, moving away from the green square, thus indicating general lower performances of the NARX models. The distance increases as the *LT* increases, i.e., moving from the first to the last panel. Additionally, looking at the different marker types, which denote the case study, it is possible to observe that the Rosamarina case study (+) performs better than other models, having always lower values of RMSD (except for $LT = LT_1$) and higher values of CC, while the model for the Scanzano (x) seems to be the worst. As the *LT* increases, the Poma case study shows the lowest correlation coefficient (Δ). Finally, the use of the bias corrected data (black markers), in most of the cases leads to an improvement of the three analyzed metrics. For the Piana degli Albanesi (o) and Poma (Δ) reservoirs, the positive effects of the bias correction are more evident at $LT > LT_1$; the bias corrected results, in this case, besides being closer to the green square, have a standard deviation very close to that of the observed data, thus indicating that the observed and simulated time series have the same variability and that the model is capable to also reproduce the extreme values of the stored volumes.

6 Discussion

A methodology which combines a NARX model with SF climate data to predict the water volume stored in four Sicilian reservoirs at a mid-term scale is presented in this study. Correlation analysis (Table 1) has shown that the stored volumes exhibit a strong dependency on the value that was stored on the previous month, while for all the cases, except the Rosamarina reservoir, the volumes show a strong dependency on the value of precipitation collected in the previous four months. The calibrated NARX models are capable to reproduce very well the observed volumes when forced with observed precipitation and air temperature (Figure 6e).

A first experiment on the four reservoirs using the seasonal forecast data without applying any correction (Figure 7) has demonstrated that the capability of the NARX models in reproducing the volumes depends on the specific season, as also obtained by other studies (Arnone et al. 2020a, Buontempo et al. 2018, Crochemore et al. 2017). This depends on the combination of two factors: (i) the reliability of SFs in correctly predicting the climate variables of input and (ii) the strength of the autocorrelation of the dependent variable, i.e., the stored volume. In fact, on one hand, results are as much better as the SF data are closer to the values of observed monthly precipitation and monthly temperature, as shown in summer months at low lead times (Figure 4). This implies that the models are capable to better reproduce the fluctuations in volumes during the summer period, when it is clearly easier to correctly forecast the precipitation and air temperature for a region as the Sicily, which is characterized by a summer season almost rainless and with high air temperatures. However, although in winter months and at high lead times ($LT > LT_4$) the SFs reliability is good (see Figure 9), the model performances in predicting the stored volume are low due to the uncertainty in reproducing the volumes in the previous months, i.e., at lower lead times. Additionally, for $LT > LT_4$, the volume is auto-correlated only with previous predictions (i.e., any observations) and thus it is more affected by the absence of the water withdrawals fluxes.

The variability of results, generally, increases as the LT increases, especially during the winter months because of the high variability in forecasting the precipitation for those months even at low LTs (Crochemore et al. 2017, Viel et al. 2016). However, a good predictability has been shown up to LT_3. This allows to have an acceptable time in advance of forecasts for the operational needs of the water utilities (Arnone et al. 2020a). So, for example, eventual anomalies in water reservoirs in August can be predict in May and thus give time to the water manager to take actions in advance.

574 To assess the effect of the systematic and random model errors that generally affect the
575 seasonal forecasts, the first experiment has been replicated by bias correcting the seasonal
576 forecasts by means of observed data. By comparing the uncorrected seasonal forecast dataset
577 with the observed one, it has been possible to notice that the monthly precipitation is always
578 underestimated for all the case studies, especially for higher values, while on the contrary the
579 monthly temperature is always overestimated, especially at the lower temperatures (Figure 9).
580 After the bias correction, both the monthly precipitation and air temperature show a better
581 description of observed dataset.

582 In most cases, the volumes simulated forcing the calibrated NARX models with the bias
583 corrected seasonal forecast data have demonstrated to better reproduce the observed volumes
584 stored within the reservoirs as compared to the case in which the model is forced with
585 uncorrected data (Figure 11). Also in this case, results have shown the dependence on the lead
586 time, showing an agreement between the simulated and observed volumes as stronger as lower is
587 the lead time considered and acceptable up to LT_3 (Figure 12). Finally, with regard to the
588 different case studies, overall results have demonstrated that the stored volumes are better
589 forecasted in some reservoirs than other as a consequence of the different reliability of the SF
590 data as compared to the observations. In fact, even the bias correction leads to a different rate of
591 improvement.

592 There are certain limitations to the approach undertaken in this study. First, the observed
593 data of monthly precipitation and monthly temperature are spatially averaged over the entire
594 contributing area of the considered reservoir, while the seasonal forecast values are
595 representative of the average climatic conditions of a very extended area (i.e., about 10,000 km^2).
596 Consequently, although all the reservoirs are within the same cell of the seasonal forecasts, these
597 can be more representative of the real climatic conditions of some reservoirs than others. A
598 second limit in the study is that, as previously mentioned, the effective water withdrawals from
599 the reservoirs have not been considered here since information about those is partially known.

The knowledge of this component could further improve the performance of the NARX models, especially for the summer period, where this component is significant in Sicily.

Despite these limitations, results are encouraging, providing reliable information about the real fluctuations of water levels within the reservoirs. Even during the drought period in between 2017 and 2018, the NARX models have demonstrated to make a reliable forecast of the stored volumes in the reservoirs for the medium *LT*s when forced with the bias corrected dataset.

7 Conclusions

In 2016 and 2017, a very severe drought hit the Mediterranean area and particularly the Sicily (ISPRA, 2016; SIAS, 2016; WWA, 2017), causing a critical water shortage with the consequent problems regarding the water supply for the cities, the industry, and the agriculture sector at the turn of 2017 and 2018. Already in the recent past (e.g., in late 2002 and 2009), the area had experienced some droughts that had led to even more consequences. In such an environment, artificial reservoirs are one of the main water supply resources. Since their management can be strongly affected by the problems of drought, predicting well in advance reservoir volumes is critical for a correct planning of the water usage at the short- and mid-term scales.

The reservoir water level is the result of the hydrological processes occurring in the upstream catchment, which, in turn, depend on meteorological variables, such as rainfall and temperature. It follows that a reliable forecast model of the meteorological forcing, along with a reliable water balance model, could enhance the correct management of a reservoir. Regarding the rainfall/temperature forecast model, the use of forecast climate data at the mid-term may provide further support for the future water level estimation of reservoirs.

From the perspective of the water balance model, instead, among the approaches used to predict the water levels for the next future, those based on data-driven methods have been demonstrated to be particularly capable of correctly reproducing the correlation between a dependent variable and some climate covariates.

This study presents the results of a novel application that exploits the seasonal forecast data, produced at the ECMWF, within a data-driven model aimed to predict the reservoir water volume at mid-term scale, up to six months ahead in time, in four reservoirs of the Sicily. For each case, a NARX was calibrated to reproduce the monthly stored water volume starting from the monthly precipitation and mean monthly air temperature variables. Results show the capability of the NARXs to reproduce the water levels in the investigated period, including the variations during dry periods. Indeed, the proposed methodology allows to predict well in advance the probable stored volumes, thanks to the use of seasonal forecast data, thus providing a reliable tool for the management of reservoirs. Moreover, the methodology would be particularly suitable for considering the water withdrawals as well, where available.

Acknowledgements and data availability

The dataset of observed data has been provided by the *Autorità di Bacino della Regione Sicilia* (Basin Authority of the Sicilian Region). More information about the dataset is available at the following link:
https://pti.regione.sicilia.it/portal/page/portal/PIR_PORTALE/PIR_LaStrutturaRegionale/PIR_PresidenzadellaRegione/PIR_AutoritaBacino

The seasonal forecast dataset is freely available through the data access system of Copernicus Climate Data Store at the following link: https://cds.climate.copernicus.eu/#!/home.

Authors would like to thank the *Autorità di Bacino della Regione Sicilia* for providing its dataset.

References

AHMAD, A., *et al.* 2014. Reservoir Optimization in Water Resources: a Review. *Water Resources Management,* 28(11), 3391-3405.

ARNAL, L., *et al.* 2018. Skilful seasonal forecasts of streamflow over Europe? *Hydrol. Earth Syst. Sci.,* 22(4), 2057-2072.

ARNONE, E., *et al.* 2020a. Droughts Prediction: a Methodology Based on Climate Seasonal Forecasts. *Water Resources Management,* 34(14), 4313-4328.

ARNONE, E., *et al.* 2014. Strategies investigation in using artificial neural network for landslide susceptibility mapping: Application to a Sicilian catchment. *Journal of Hydroinformatics,* 16(2), 502-515.

ARNONE, E., *et al.* 2020b. The drought-alert decision support system for water resources management. *Desalination and Water Treatment,* 194, 304-314.

AWCHI, T. A. 2014. River Discharges Forecasting In Northern Iraq Using Different ANN Techniques. *Water Resources Management,* 28(3), 801-814.

BONACCORSO, B., CANCELLIERE, A. and ROSSI, G. 2015. Probabilistic forecasting of drought class transitions in Sicily (Italy) using Standardized Precipitation Index and North Atlantic Oscillation Index. *Journal of Hydrology,* 526, 136-150.

BUONTEMPO, C., *et al.* 2018. What have we learnt from EUPORIAS climate service prototypes? *Climate Services,* 9, 21-32.

CALOIERO, T., *et al.* 2018. Drought Analysis in Europe and in the Mediterranean Basin Using the Standardized Precipitation Index. *Water,* 10(8), 1043.

CHAVES, P. and CHANG, F.-J. 2008. Intelligent reservoir operation system based on evolving artificial neural networks. *Advances in Water Resources,* 31(6), 926-936.

CLARK, R. T., *et al.* 2017. Skilful seasonal predictions for the European energy industry. *Environmental Research Letters,* 12(2), 024002.

CROCHEMORE, L., *et al.* 2017. Seasonal streamflow forecasting by conditioning climatology with precipitation indices. *Hydrol. Earth Syst. Sci.,* 21(3), 1573-1591.

DE FELICE, M., ALESSANDRI, A. and CATALANO, F. 2015. Seasonal climate forecasts for medium-term electricity demand forecasting. *Applied Energy,* 137(C), 435-444.

DOBLAS-REYES, F. J., *et al.* 2013. Seasonal climate predictability and forecasting: status and prospects. *WIREs Climate Change,* 4(4), 245-268.

EHRET, U., *et al.* 2012. HESS Opinions "Should we apply bias correction to global and regional climate model data?". *Hydrol. Earth Syst. Sci.,* 16(9), 3391-3404.

EIA, 2010. *International energy outlook 2010.* Energy Information Administration.

EL-SHAFIE, A., *et al.* 2012. Dynamic versus static neural network model for rainfall forecasting at Klang River Basin, Malaysia. *Hydrology and Earth System Sciences,* 16(4), 1151-1169.

EL-SHAFIE, A., TAHA, M. R. and NOURELDIN, A. 2007. A neuro-fuzzy model for inflow forecasting of the Nile river at Aswan high dam. *Water Resources Management,* 21(3), 533-556.

ESSENFELDER, A. H., *et al.* 2020. Smart Climate Hydropower Tool: A Machine-Learning Seasonal Forecasting Climate Service to Support Cost-Benefit Analysis of Reservoir Management. *Atmosphere,* 11(12).

FEI, T. and SHUANG-QING, X. 2012. Definition of Business as Usual and Its Impacts on Assessment of Mitigation Efforts. *Advances in Climate Change Research,* 3(4), 212-219.

FORESTIERI, A., *et al.* 2018. The impact of climate change on extreme precipitation in Sicily, Italy. *Hydrological Processes,* 32(3), 332-348.

FOWLER, H. J., *et al.* 2021. Anthropogenic intensification of short-duration rainfall extremes. *Nature Reviews Earth & Environment,* 2(2), 107-122.

GLAHN, H. R. and LOWRY, D. A. 1972. The Use of Model Output Statistics (MOS) in Objective Weather Forecasting. *Journal of Applied Meteorology and Climatology,* 11(8), 1203-1211.

GUDMUNDSSON, L., *et al.* 2012. Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations - a comparison of methods. *Hydrology and Earth System Sciences,* 16(9), 3383-3390.

HADIYAN, P. P., MOEINI, R. and EHSANZADEH, E. 2020. Application of static and dynamic artificial neural networks for forecasting inflow discharges, case study: Sefidroud Dam reservoir. *Sustainable Computing-Informatics & Systems,* 27.

HASSAN, M., *et al.* 2015. Predicting streamflows to a multipurpose reservoir using artificial neural networks and regression techniques. *Earth Science Informatics,* 8(2), 337-352.

HOSKINS, B. 2013. The potential for skill across the range of the seamless weather-climate prediction problem: a stimulus for our science. *Quarterly Journal of the Royal Meteorological Society,* 139(672), 573-584.

ISPRA, 2016. *Bollettino Siccità, Italia.* Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA).

JOHNSON, S. J., *et al.* 2019. SEAS5: the new ECMWF seasonal forecast system. *Geoscientific Model Development,* 12(3), 1087-1117.

KLEIN, W. H. and GLAHN, H. R. 1974. forecasting local weather by means of model output statistics. *Bulletin of the American Meteorological Society,* 55(10), 1217-1227.

KOUHESTANI, S., *et al.* 2016. Projection of climate change impacts on precipitation using soft-computing techniques: A case study in Zayandeh-rud Basin, Iran. *Global and Planetary Change,* 144, 158-170.

MARAUN, D. 2012. Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums. *Geophysical Research Letters,* 39.

MARAUN, D., *et al.* 2010. Precipitation Downscaling under Climate Change: Recent Developments to Bridge the Gap between Dynamical Models and the End User. *Reviews of Geophysics,* 48.

MORIASI, D. N., *et al.* 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the Asabe,* 50(3), 885-900.

NASH, J. E. and SUTCLIFFE, J. V. 1970. River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology,* 10(3), 282-290.

NIU, W.-J. and FENG, Z.-K. 2021. Evaluating the performances of several artificial intelligence methods in forecasting daily streamflow time series for sustainable water resources management. *Sustainable Cities and Society,* 64, 102562.

PEARSON, K. 1895. Note on Regression and Inheritance in the Case of Two Parents. *Proceedings of the Royal Society of London,* 58, 240-242.

PEÑUELA, A., HUTTON, C. and PIANOSI, F. 2020. Assessing the value of seasonal hydrological forecasts for improving water resource management: insights from a pilot application in the UK. *Hydrol. Earth Syst. Sci.,* 24(12), 6059-6073.

PIANI, C., *et al.* 2010. Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *Journal of Hydrology,* 395(3-4), 199-215.

PYRINA, M., *et al.* 2021. Statistical Seasonal Prediction of European Summer Mean Temperature Using Observational, Reanalysis, and Satellite Data. *Weather and Forecasting,* 36(4), 1537-1560.

ROZOS, E. 2019. Machine Learning, Urban Water Resources Management and Operating Policy. *Resources,* 8(4), 173.

SANCHEZ, G. M.*, et al.* 2020. Forecasting water demand across a rapidly urbanizing region. *Science of the Total Environment,* 730.

SIAS, 2016. *Siccità invernale 2015-2016: Siccità periodo Dicembre 2015 - Febbraio 2016 Temperature anomale Febbraio 2016.* Servizio Informativo Agrometereologico Siciliano (SIAS) - Regione Siciliana.

TREPPIEDI, D.*, et al.* 2021. Detecting precipitation trend using a multiscale approach based on quantile regression over a Mediterranean area. *International Journal of Climatology,* 41(13), 5938-5955.

VANNITSEM, S. and NICOLIS, C. 2008. Dynamical Properties of Model Output Statistics Forecasts. *Monthly Weather Review,* 136(2), 405-419.

VIEL, C.*, et al.* 2016. How seasonal forecast could help a decision maker: an example of climate service for water resource management. *Adv. Sci. Res.,* 13, 51-55.

WWA, 2017. *Euro-Mediterranean heat - summer 2017.* World Weather Attribution (WWA).

YANG, S.*, et al.* 2019. Real-time reservoir operation using recurrent neural networks and inflow forecast from a distributed hydrological model. *Journal of Hydrology,* 579, 124229.

YU, P. S.*, et al.* 2017. Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. *Journal of Hydrology,* 552, 92-104.

ZRIBI, M.*, et al.*, 2020. Introduction. *In:* ZRIBI, M.*, et al.* eds. *Water Resources in the Mediterranean Region.* Elsevier, xv-xix.

Statements & Declarations

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

The authors have no relevant financial or non-financial interests to disclose.

All authors contributed to the study conception and design. Material preparation and data collection were performed by Elisa Arnone, while analyses were performed by Antonio Francipane. All the previous activities were coordinated by Prof. Leonardo V. Noto. The first draft of the manuscript was written by Antonio Francipane. All authors commented on previous versions of the manuscript, read, and approved the final manuscript.

Tables with captions

Table 1. Characteristics of NARX models architecture for each of the considered reservoirs.

Reservoir	*Input delay* [months]	*Feedback delay* [months]	*N. of neurons in the hidden layer*	*NSE for calibration*	*RMSE for calibration*
P. degli Albanesi	4	1	11	0.985	0.040
Poma	4	1	19	0.988	0.105
Rosamarina	5	1	14	0.982	0.207
Scanzano	4	1	21	0.971	0.034

Figures captions

- Figure 1. NARX architecture. The scheme refers to a closed loop configuration. In this case, the network uses information from exogenous input variables along with the target series itself and the feedbacks past predicted and observed values delayed (e.g., $1{:}n$ delays for the input variables and $1{:}m$ delays for the feedbacks).

- Figure 2. Aerial view of the city of Palermo and the Piana degli Albanesi, Poma, Rosamarina, and Scanzano reservoirs. The blue points indicate the gauge networks of the AdB. Inset at the top right indicates the SF grid for the Sicily and highlights the COI #6 with a red box. Blue and red contours in the insets at the bottom indicate the extension of the reservoirs when the maximum volume is reached and during the month of February 2108, respectively. Source: © Google Maps Satellite basemap available within the QuickMapServices plugin of Quantum GIS.

- Figure 3. Normalized monthly stored volumes within the four reservoirs. The gray shaded box highlights the drought period that affected the reservoirs in between 2017 and 2018.

- Figure 4. Example of SF released in January (a and c) and June (b and d) 2019 and predicting the six months ahead in time for the monthly precipitation (a and b) and monthly air temperature (c and d).

- Figure 5. Flowchart of the phases followed to define the NARX model to be used with the SF data to forecast future stored volumes within a reservoir.

- Figure 6. NARX model's performances returned by one of the calibration rounds for the Piana degli Albanesi reservoir. Performances are evaluated in terms of regression analysis for the a) training, b) validation, c) test, and d) overall calibration. Subplot e) shows NARX calibration for the four reservoirs; blue solid lines are referred to the observed volumes stored within the reservoirs, while red dash-dotted lines are the volumes simulated with the calibrated NARX. The NSE values in the subplots refer to the calibration phase.

- Figure 7. Rosamarina reservoir: monthly stored volumes at each *i*-th month of 2019 and the six months of *LT*s ahead. Blue solid lines denote the observed volumes, while red dashed lines are the volumes returned by the NARX forced with observed monthly precipitation and monthly temperature data. Boxplots describe the ensemble of the forecasted volumes obtained with the NARX forced by the SF data of the *i*-th release month and all its lead times (i.e., from LT_0 to LT_6).

- Figure 8. Performances of SF for the a) precipitation and b) air temperature and of c) NARX model for the Rosamarina reservoir for volumes forecasted at different months and *LT*s over the entire period of simulations. The numeric values reported within each cell is equal to the RMSE.

- Figure 9. q-q plots for the a) monthly precipitation and b) monthly air temperature bias corrected with the SSPLIN method and for all the reservoirs. The gray circles indicate the q-q plot of observed (i.e., AdB) and modeled (i.e., SF) data with quantile step equal to 0.01. The dashed line indicates the perfect agreement line.

- Figure 10. Observed and simulated stored volumes for the Rosamarina reservoir at different *LT*s obtained by the NARX forced with the bias corrected SF data.

- Figure 11. NSE values obtained running the NARXs with the mean values of both uncorrected and bias corrected SF data for the four reservoirs and the entire dataset.

- Figure 12. As the Figure 8 for bias corrected SF data and NARX model forced with them.

- Figure 13. Normalized Taylor diagram for the results obtained forcing the calibrated NARX models with observed data provided by the AdB and with uncorrected and bias corrected SF data, at different *LT*s. The marker indicates the reservoir case study. The green square refers to the observed value, where the normalized standard deviation is equal to 1; the radial distance from the green square quantifies the centered RMSD normalized by the standard deviation of observed volumes, while the azimuth and the radial distance from the origin quantify the CC and the normalized standard deviation, respectively.

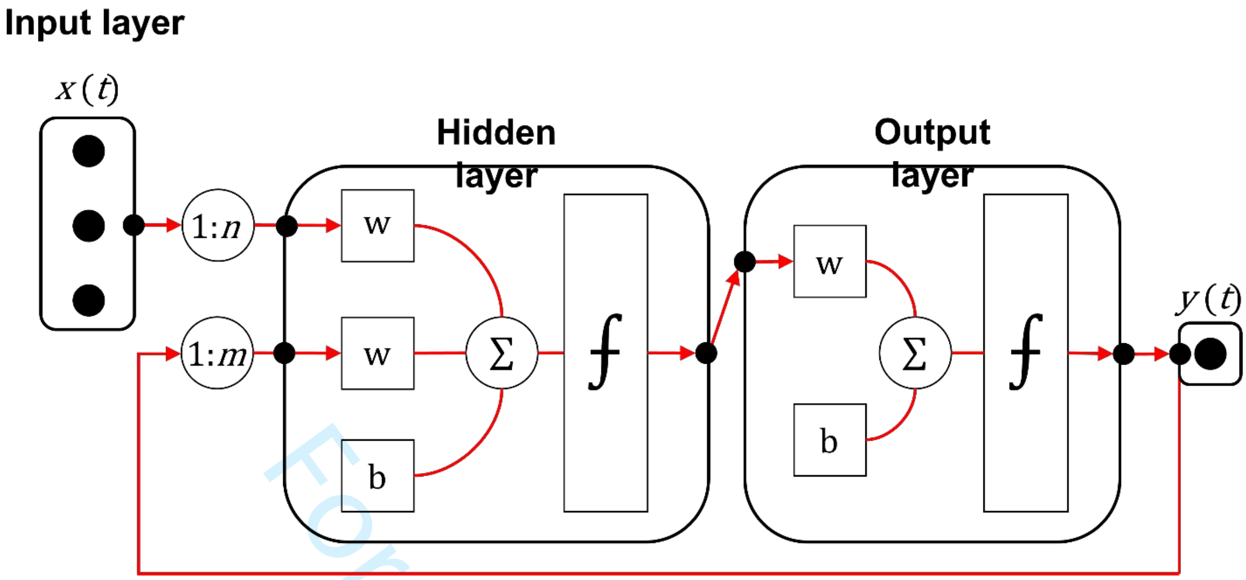

Figure 1. NARX architecture. The scheme refers to a closed loop configuration. In this case, the network uses information from exogenous input variables along with the target series itself and the feedbacks past predicted and observed values delayed (e.g., 1:n delays for the input variables and 1:m delays for the feedbacks).

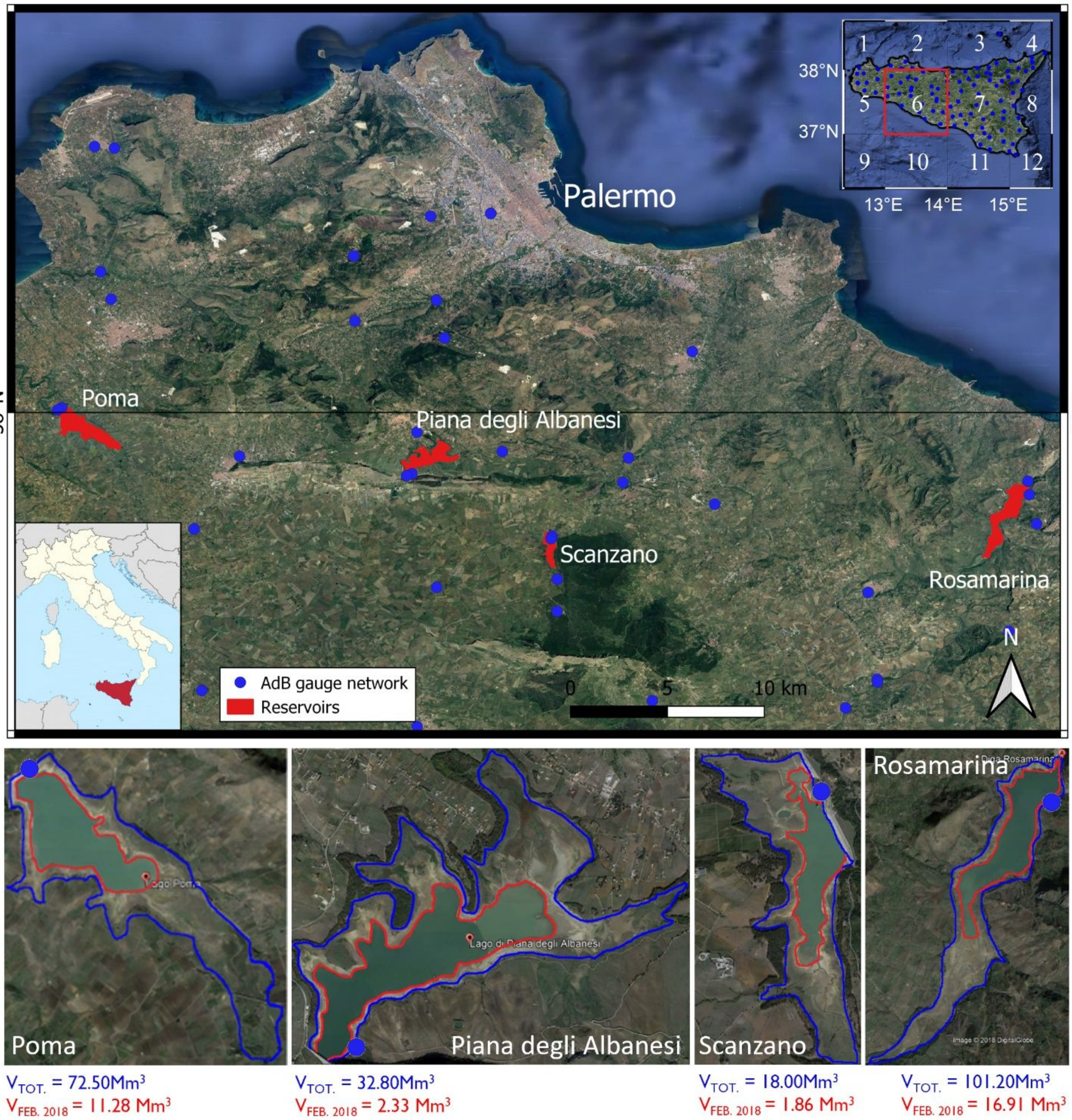

Figure 2. Aerial view of the city of Palermo and the Piana degli Albanesi, Poma, Rosamarina, and Scanzano reservoirs. The blue points indicate the gauge networks of the AdB. Inset at the top right indicates the SF grid for the Sicily and highlights the COI #6 with a red box. Blue and red contours in the insets at the bottom indicate the extension of the reservoirs when the maximum volume is reached and during the month of February 2108, respectively. Source: © Google Maps Satellite basemap available within the QuickMapServices plugin of Quantum GIS.

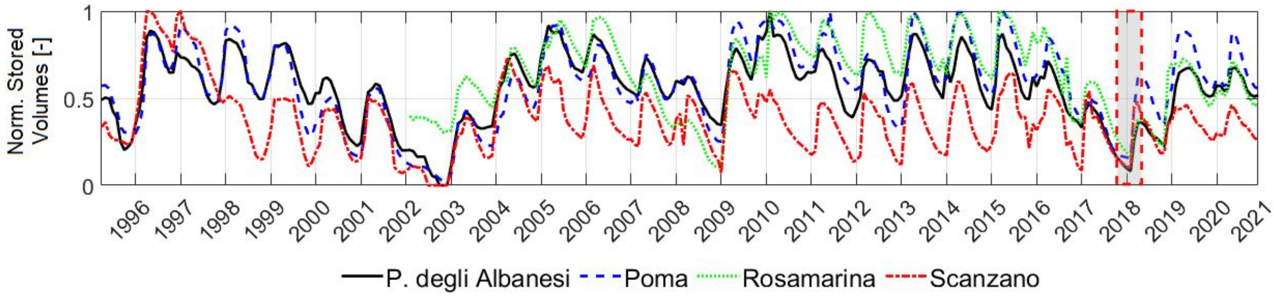

Figure 3. Normalized monthly stored volumes within the four reservoirs. The gray shaded box highlights the drought period that affected the reservoirs in between 2017 and 2018.

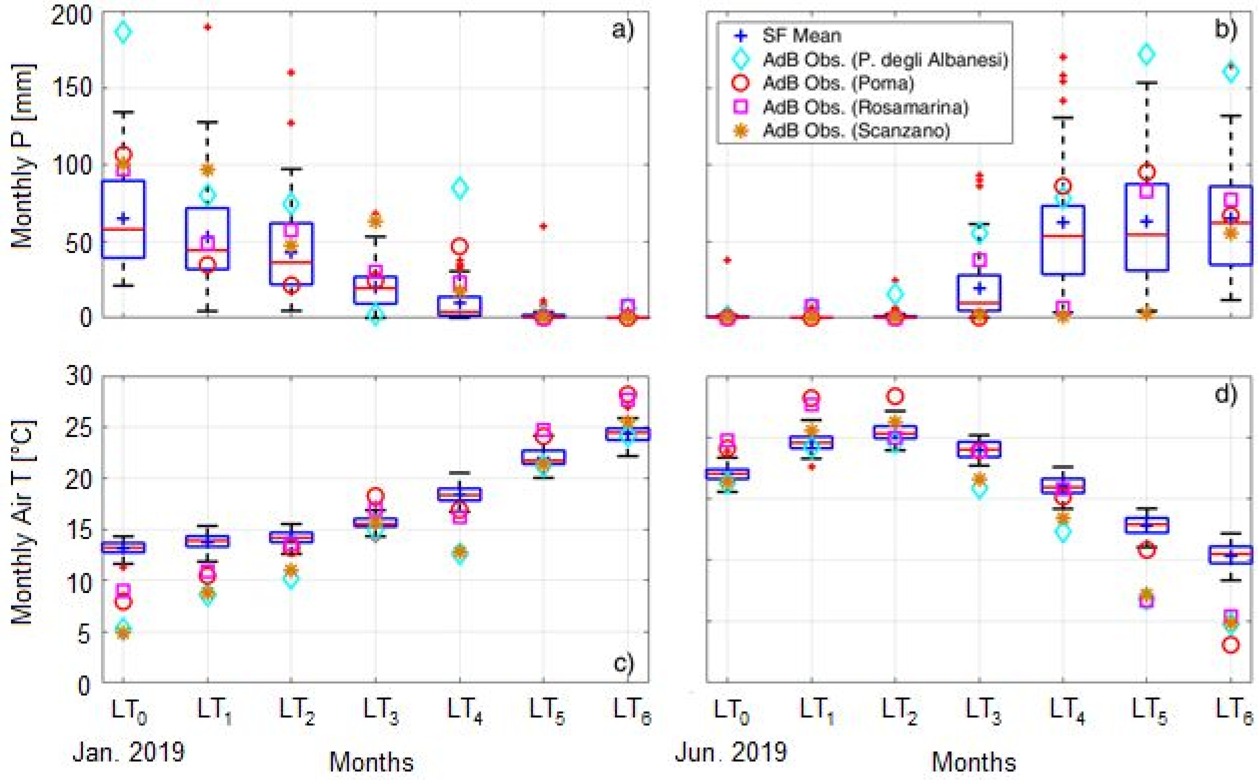

Figure 4. Example of SF released in January (a and c) and June (b and d) 2019 and predicting the six months ahead in time for the monthly precipitation (a and b) and monthly air temperature (c and d).

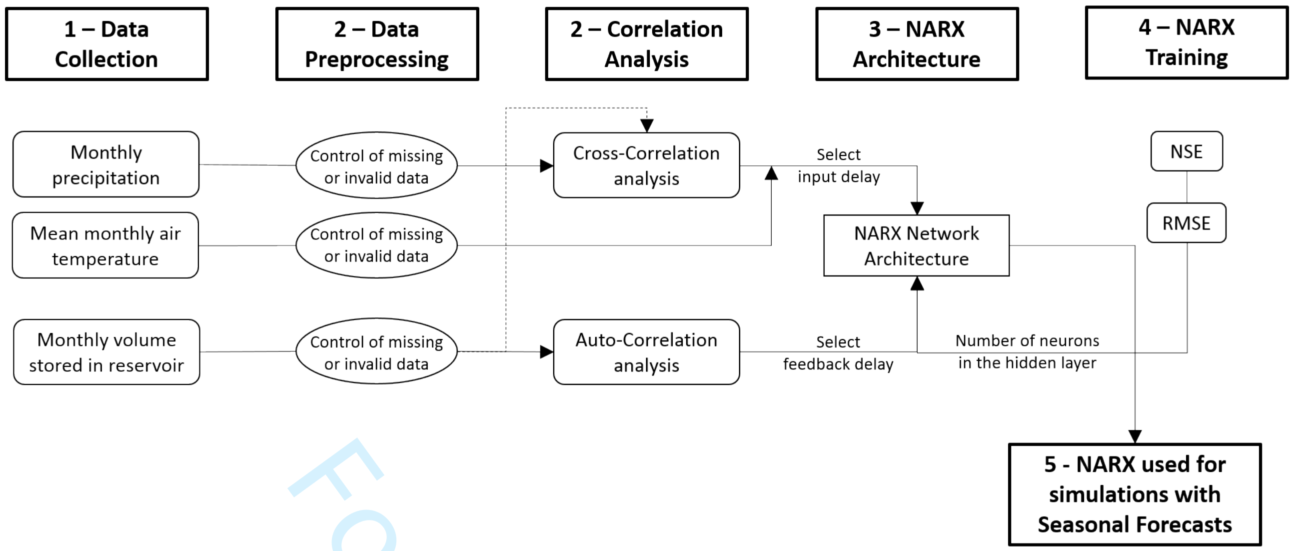

Figure 5. Flowchart of the phases followed to define the NARX model to be used with the SF data to forecast future stored volumes within a reservoir.

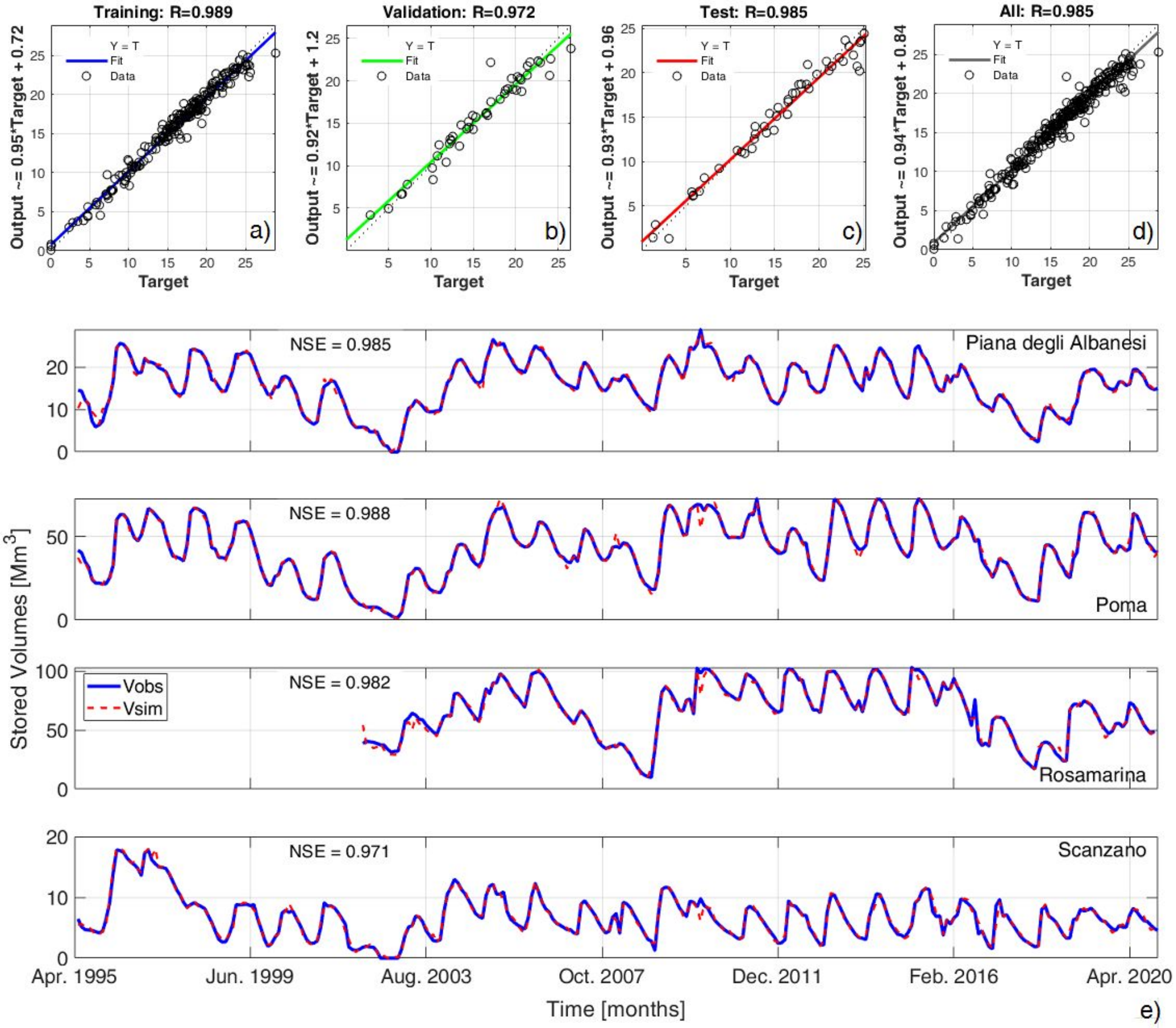

Figure 6. NARX model's performances returned by one of the calibration rounds for the Piana degli Albanesi reservoir. Performances are evaluated in terms of regression analysis for the a) training, b) validation, c) test, and d) overall calibration. Subplot e) shows NARX calibration for the four reservoirs; blue solid lines are referred to the observed volumes stored within the reservoirs, while red dash-dotted lines are the volumes simulated with the calibrated NARX. The NSE values in the subplots refer to the calibration phase.

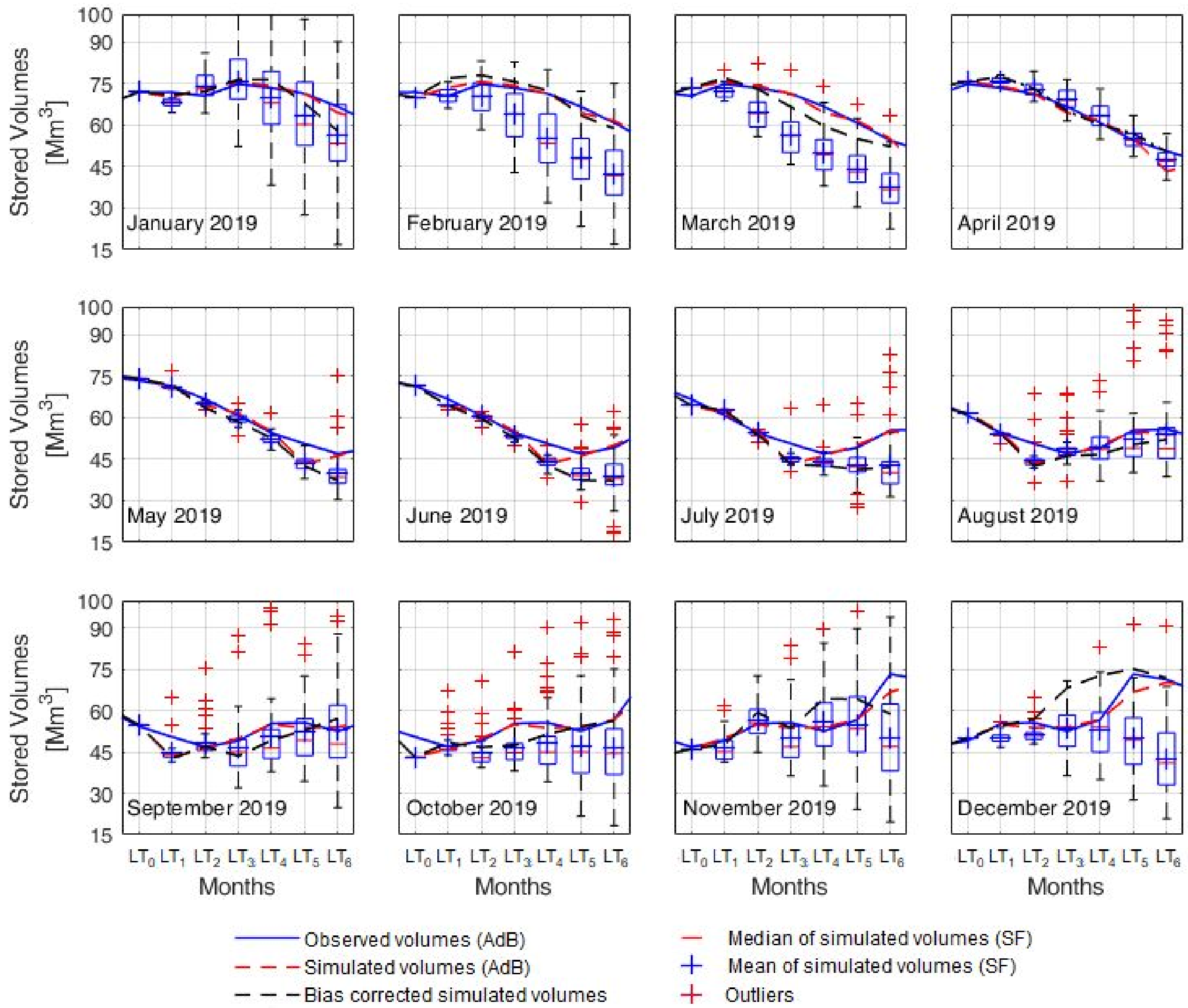

Figure 7. Rosamarina reservoir: monthly stored volumes at each *i*-th month of 2019 and the six months of *LT*s ahead. Blue solid lines denote the observed volumes, while red dashed lines are the volumes returned by the NARX forced with observed monthly precipitation and monthly temperature data. Boxplots describe the ensemble of the forecasted volumes obtained with the NARX forced by the SF data of the *i*-th release month and all its lead times (i.e., from LT_0 to LT_6).

Figure 8. Performances of SF for the a) precipitation and b) air temperature and of c) NARX model for the Rosamarina reservoir for volumes forecasted at different months and *LT*s over the entire period of simulations. The numeric values reported within each cell is equal to the RMSE.

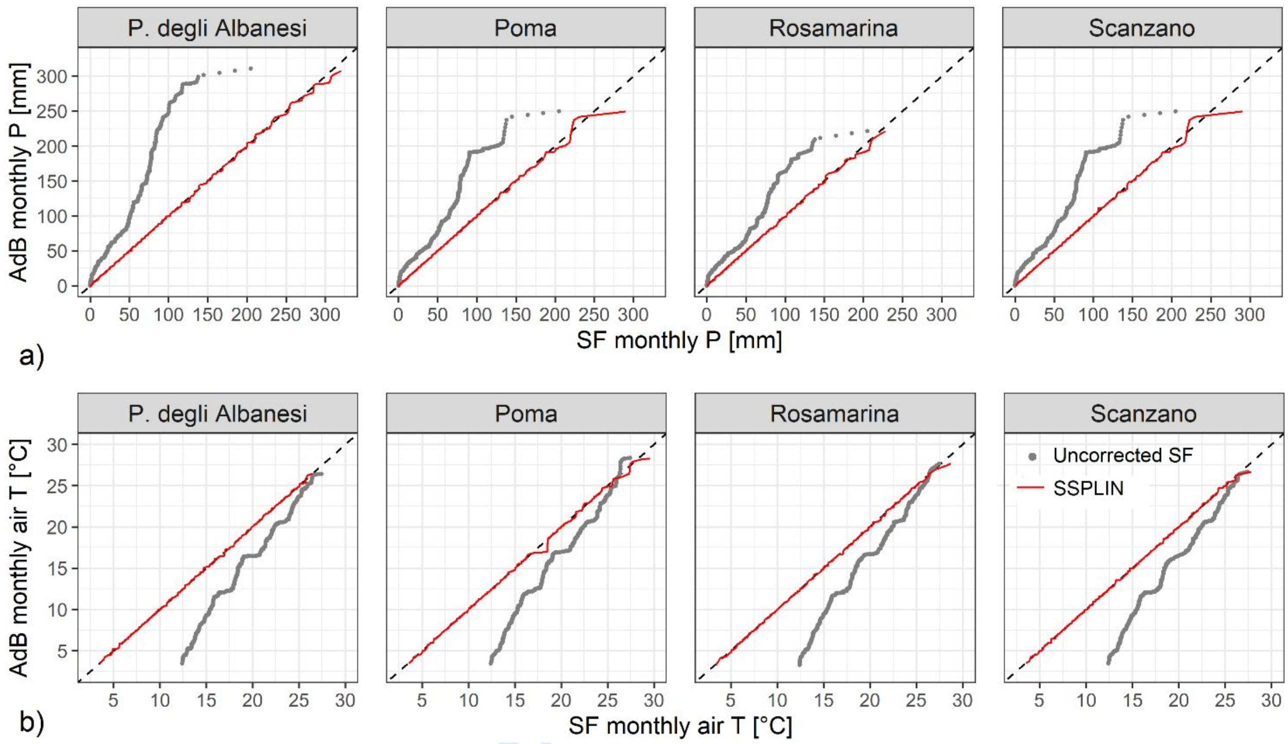

Figure 9. q-q plots for the a) monthly precipitation and b) monthly air temperature bias corrected with the SSPLIN method and for all the reservoirs. The gray circles indicate the q-q plot of observed (i.e., AdB) and modeled (i.e., SF) data with quantile step equal to 0.01. The dashed line indicates the perfect agreement line.

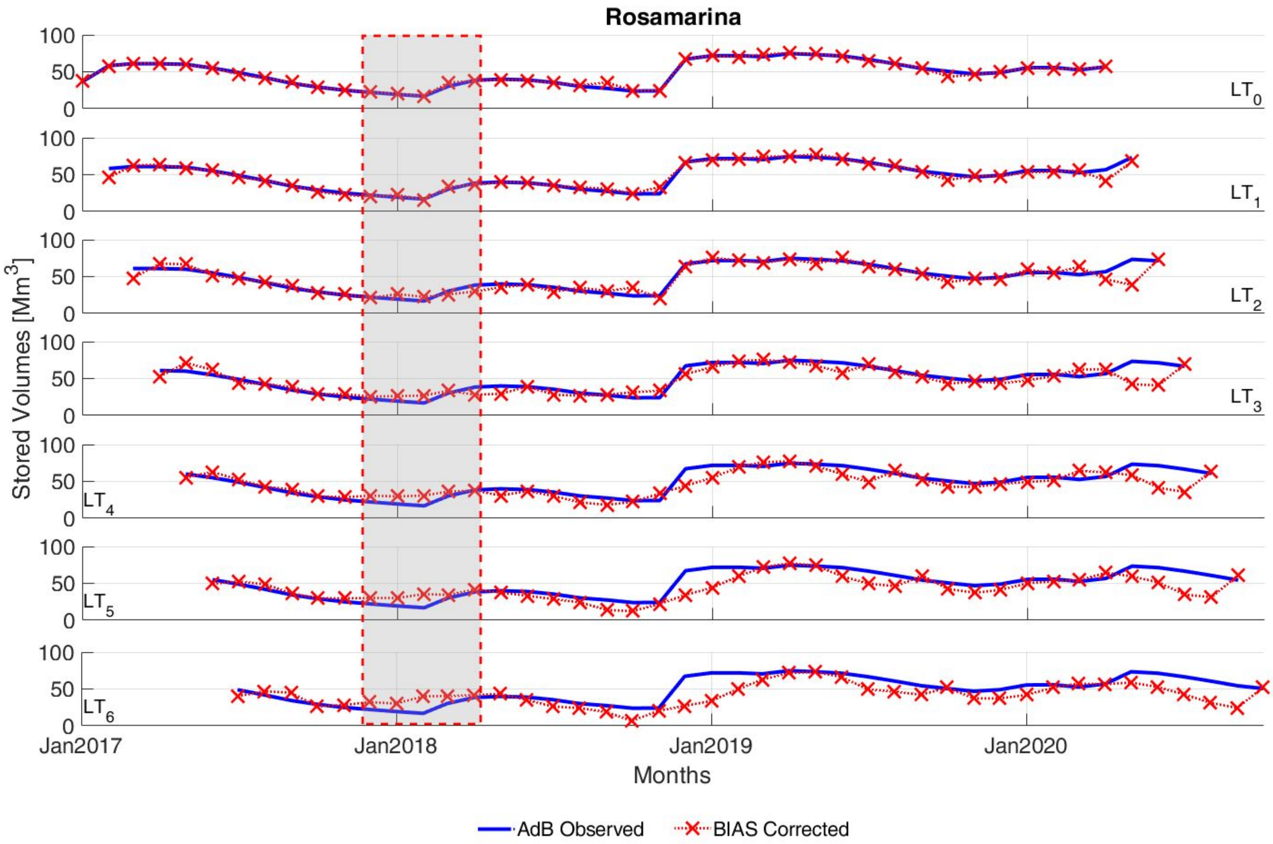

Figure 10. Observed and simulated stored volumes for the Rosamarina reservoir at different *LT*s obtained by the NARX forced with the bias corrected SF data.

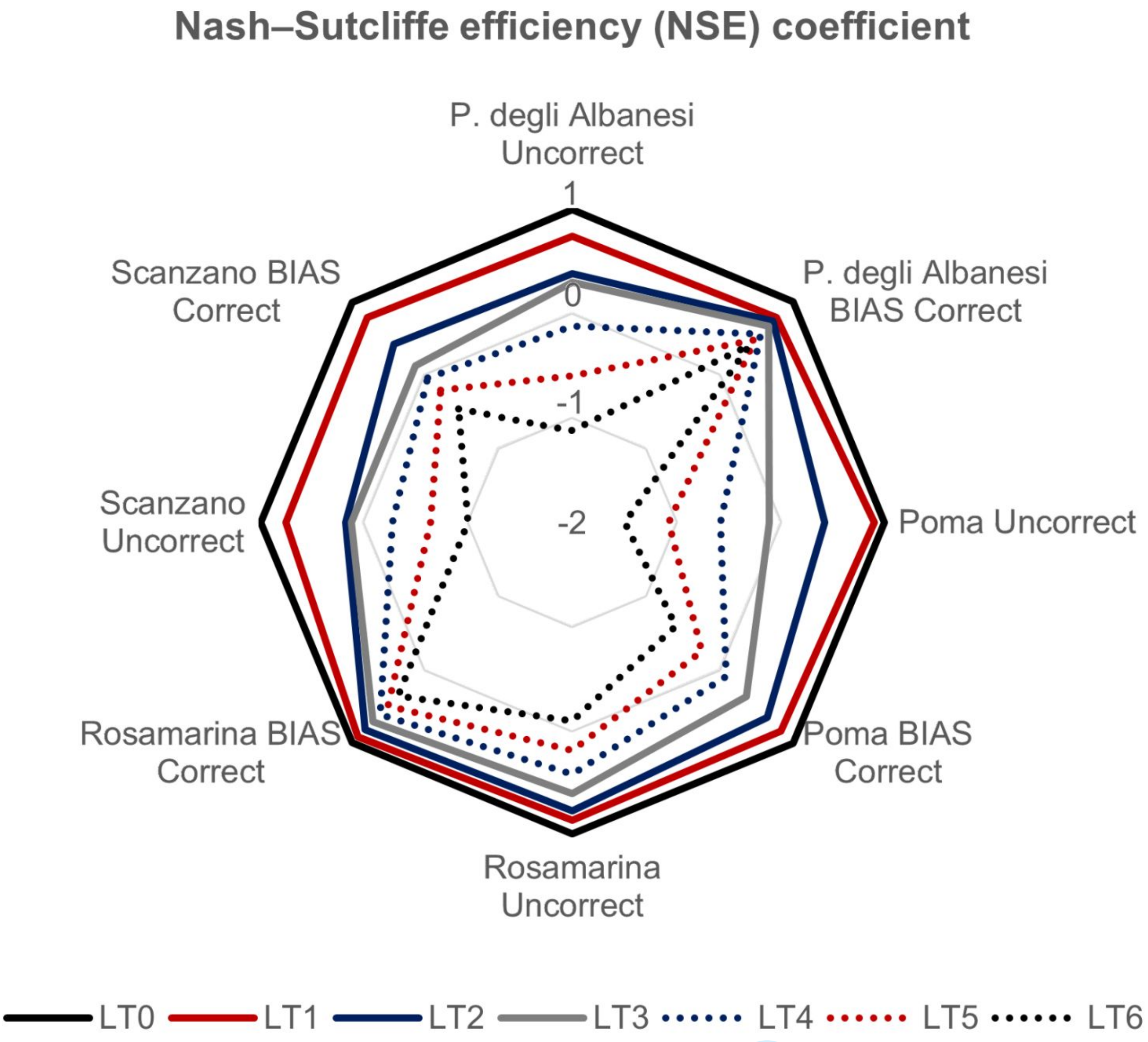

Figure 11. NSE values obtained running the NARXs with the mean values of both uncorrected and bias corrected SF data for the four reservoirs and the entire dataset.

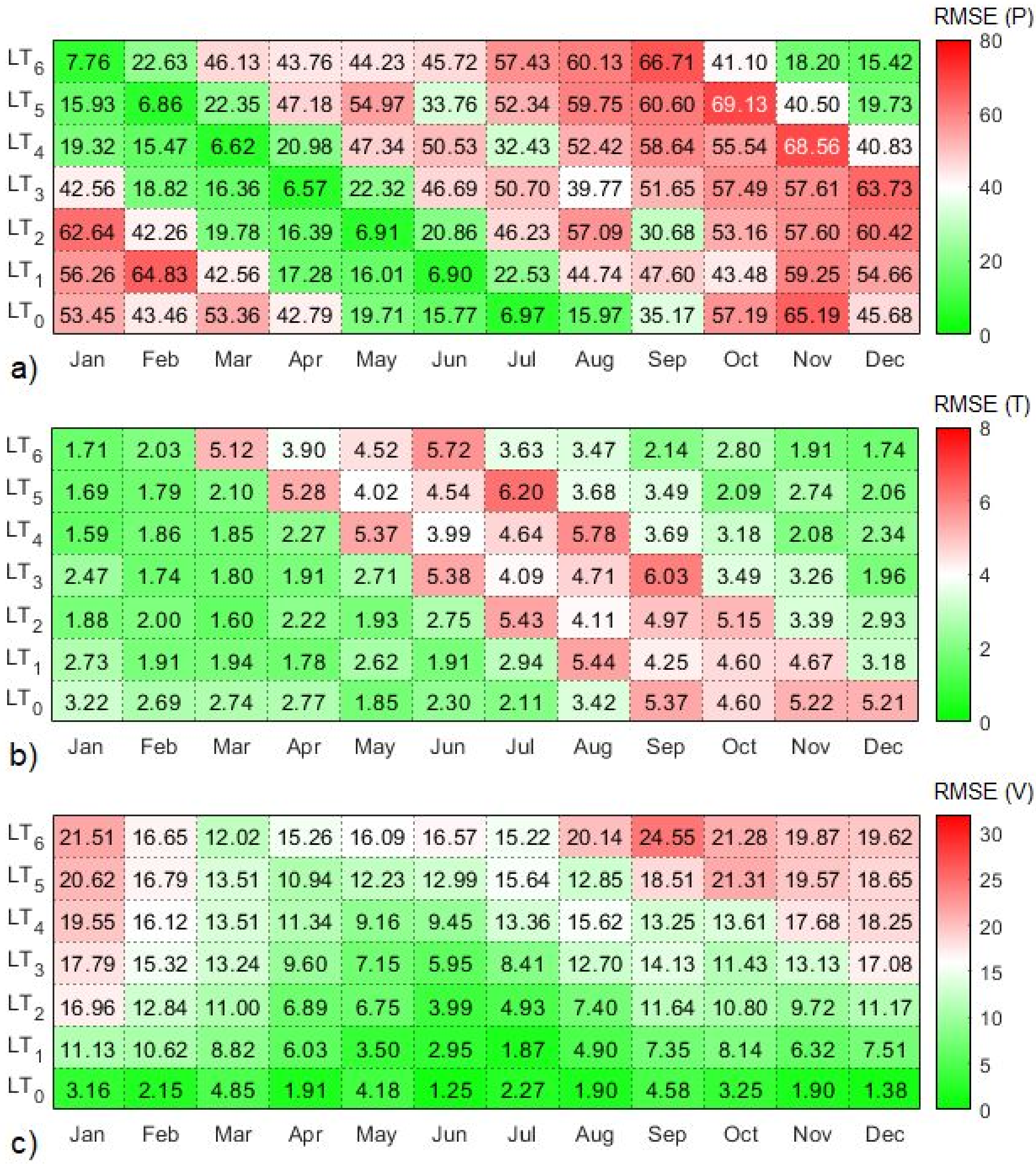

Figure 12. As the Figure 8 for bias corrected SF data and NARX model forced with them.

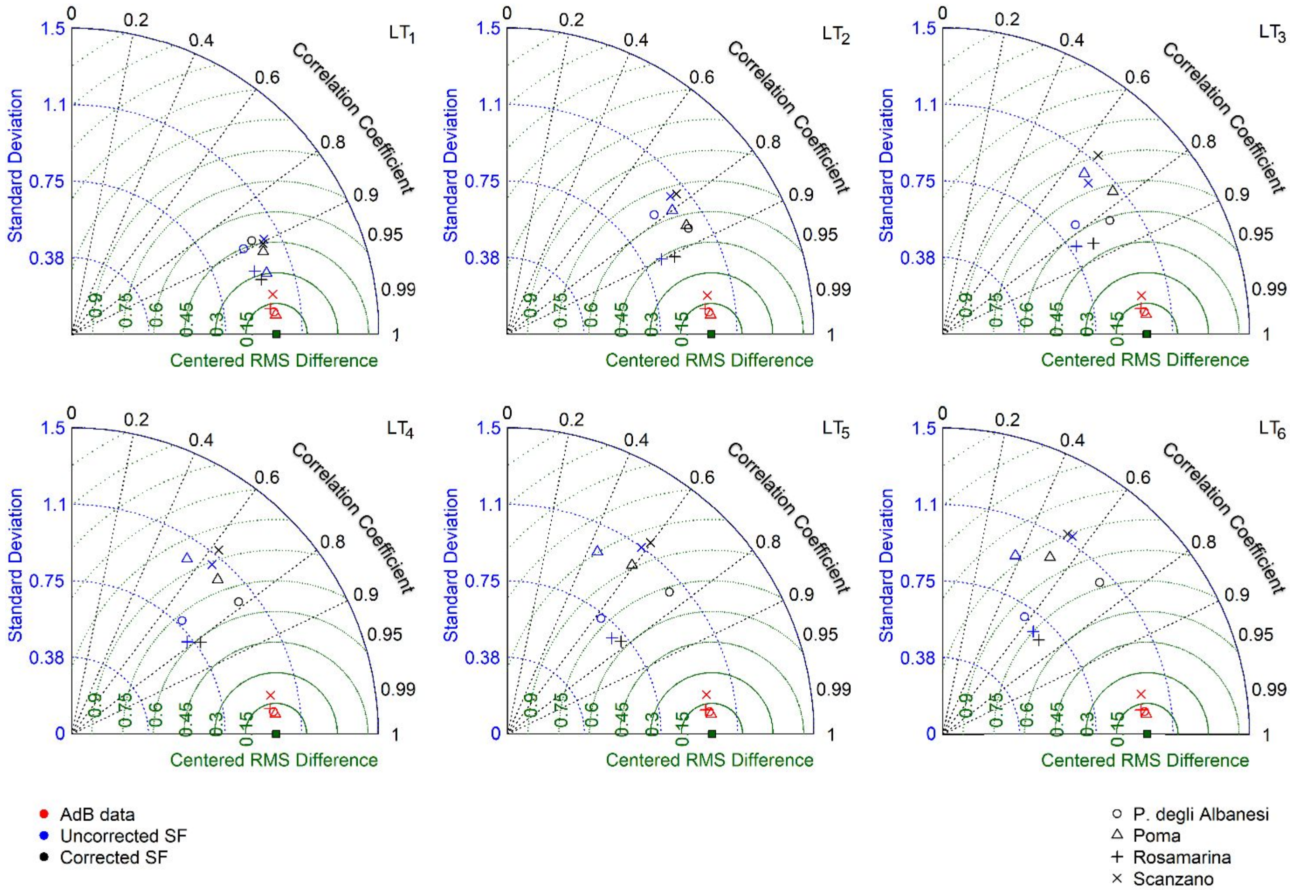

Figure 13. Normalized Taylor diagram for the results obtained forcing the calibrated NARX models with observed data provided by the AdB and with uncorrected and bias corrected SF data, at different *LT*s. The marker indicates the reservoir case study. The green square refers to the observed value, where the normalized standard deviation is equal to 1; the radial distance from the green square quantifies the centered RMSD normalized by the standard deviation of observed volumes, while the azimuth and the radial distance from the origin quantify the CC and the normalized standard deviation, respectively.