engineering proceedings

Proceeding Paper

From Deterministic to Probabilistic Forecasts of Water Demand †

Annalaura Gabriele [1,*], Daniela Biondi [2], Rudy Gargano [1] and Ezio Todini [3]

1 Dipartimento di Ingegneria Civile e Meccanica, University of Cassino and Southern Lazio, 03043 Cassino, Italy; gargano@unicas.it
2 Department of Informatics, Modelling, Electronics and System Engineering, University of Calabria, 87036 Rende, Italy; daniela.biondi@unical.it
3 Italian Hydrological Society, 40127 Bologna, Italy; eziotodini@gmail.com
* Correspondence: annalaura.gabriele@unicas.it
† Presented at the 3rd International Joint Conference on Water Distribution Systems Analysis & Computing and Control for the Water Industry (WDSA/CCWI 2024), Ferrara, Italy, 1–4 July 2024.

Abstract: This work presents the application of the MCP as the integrator of two NN-type models, to provide probabilistic forecasts of water demand within the frame of the WDSA 2024 Battle. The proposed approach allows assessing the probability distribution of future values of water demand conditional to the deterministic model forecasts.

Keywords: model conditional processor; data-driven; ANN; NARX; water demand

check for updates

Citation: Gabriele, A.; Biondi, D.; Gargano, R.; Todini, E. From Deterministic to Probabilistic Forecasts of Water Demand. *Eng. Proc.* **2024**, *69*, 122. https://doi.org/10.3390/engproc2024069122

Academic Editors: Stefano Alvisi, Marco Franchini, Valentina Marsili and Filippo Mazzoni

Published: 12 September 2024

1. Introduction

Although the WDSA 2024 Battle considers only deterministic forecasts, we aim to show that even if we limit the forecasts to the median values derived from the predictive density, probabilistic forecasts may lead to improved estimates. To this end, we used the MCP approach [1] to build the probability distribution of a future value conditional to two deterministic forecasts provided by the artificial neural network (ANN) and non-linear auto-regressive with external inputs (NARX) neural network models [2].

Probabilistic forecasts allow combining several deterministic forecasts and provide more comprehensive information, highly improving the reliability and robustness of decisions based on their information, as shown in several environmental disciplines, such as hydrology, meteorology and climatology. Although, early attempts were made by Alvisi and Franchini and Anele et al. [3–5], the probabilistic approaches have not yet gained popularity and widely spread use in WDN demand forecasting applications.

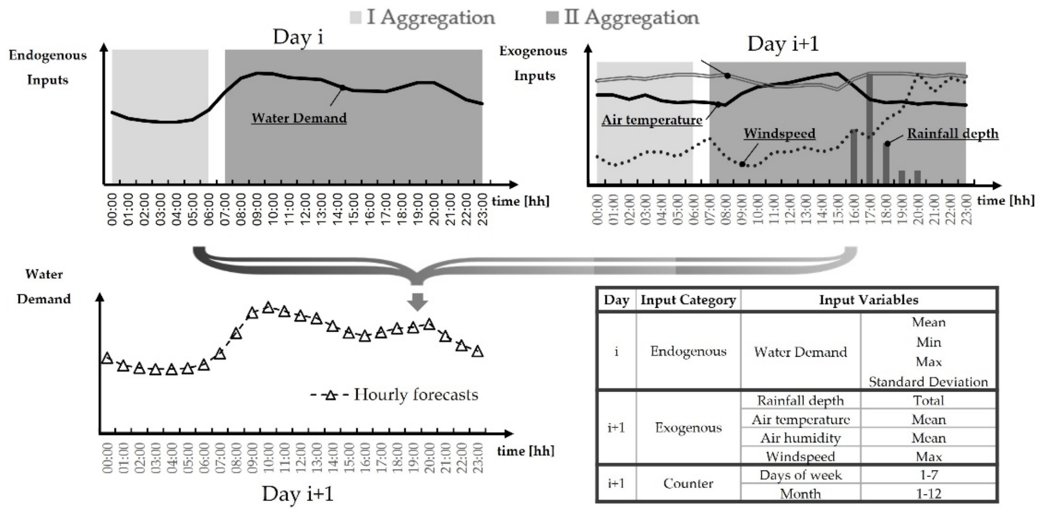
2. Deterministic Model Components

The deterministic component of the proposed approach was obtained by means of two different data-driven models, both based on artificial neural networks, the feed-forward ANN and the NARX.

The time series were initially adjusted to overcome missing data, where this loss affected time intervals lasting less than 8 h. For this aim, the data preprocessing involved the determination of the average monthly patterns, which were also classified into weekdays and holidays.

2.1. Feed-Forward ANN with Multiple Inputs and Outputs

The classical feed-forward ANN for regression was used for future daily demand modeling. Forecasting the four weeks without data needed seven runs of the proposed ANN model. The sketch and data in Figure 1 specify model inputs and outputs.



Day	Input Category	Input Variables	
i	Endogenous	Water Demand	Mean Min Max Standard Deviation
i+1	Exogenous	Rainfall depth	Total
		Air temperature	Mean
		Air humidity	Mean
		Windspeed	Max
i+1	Counter	Days of week	1-7
		Month	1-12

Figure 1. The *Feed-Forward ANN* model input and output variables.

A preliminary analysis was developed to define the best input dataset to forecast the demand trend in the missing data of the four weeks. This involved dividing the variables into two sets:

- endogenous variables (flow demand)
- exogenous variables (rain, temperature, humidity, wind)

In addition, aggregating the input data for several hours in a day improved the deterministic forecasts. More precisely, the solution required accumulating the hourly data into two inputs to distinguish the night water demand (0:00–6:00) from the diurnal one (7:00–23:00). The aggregation in time reduced the number of the input data of each run: 8 statistics of the endogenous variables of day *i*; 8 statistics of the exogenous variables of day *i* + 1 (see Figure 1).

A counter, determining the days of the week (1–7) and the month of the year (1–12), also contributed to forecasting the water demand of day *i* + 1.

The final resulting neural network comprises a total of 18 input neurons, 24 output neurons, and 15 neurons for the hidden layer.

2.2. NARX Neural Networks

The second approach, the NARX neural networks, represents a type of dynamic recurrent ANN commonly used for time series modeling [2,6]. Therefore, hourly time series were used to forecast the hourly flow demand of a future week.

This model relates the current value of the forecast variable to both previous endogenous values (demand) and current or previous values of exogenous data series (rain, temperature, humidity, wind).

The NARX model initially uses a *series-parallel* architecture for forecasting one hour step ahead. After training, it shifts to a *parallel* architecture, utilizing previously calculated values into the following time step, thus enabling multi-step predictions. Although a promising approach, training results did not guarantee successful multi-step predictions; a preliminary analysis was then required to define the most appropriate input dataset for weekly demand trend forecasting. Adding additional exogenous variables to the available time series, such as observation time and month, proved crucial for robust forecasts at the weekly scale.

The NARX model for demand forecasting is clarified in Figure 2, summarizing the set of all considered input variables.

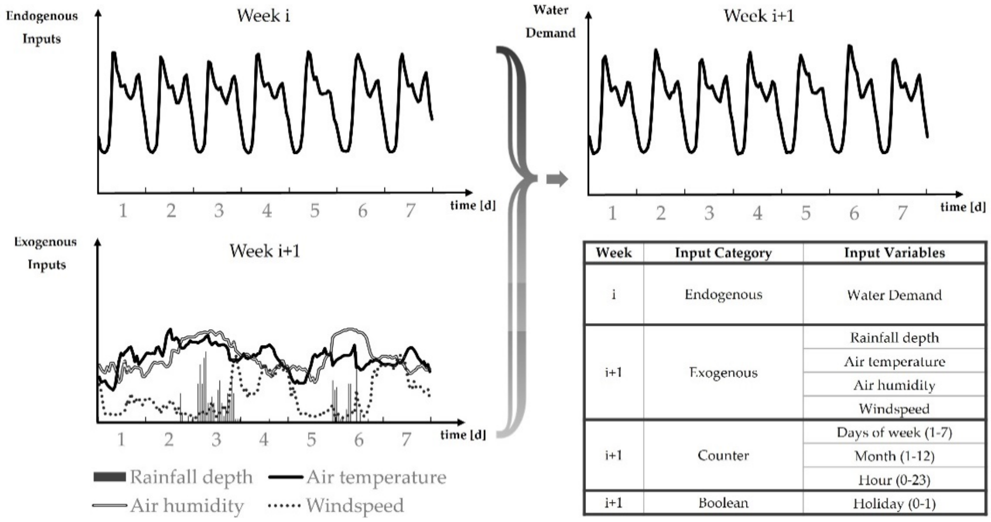


Figure 2. The chosen NARX model input and output variables.

Finally, the adopted NARX parameters were five neurons for the hidden layer with a delay equal to 168 for the endogenous variable and equal to 0 for the exogenous variables.

2.3. General Consideration on Models' Calibration

Neural network models vary with each training due to factors like random data subdivision and initial weight values; therefore, both feed-forward ANN and NARX net were trained multiple times. The final model was selected by minimizing a linear combination of the performance indexes (the mean absolute deviation (P1), the maximum absolute deviation (P2), both for the first day of the forecasted week, and the mean absolute deviation from the second to the last day (P3)).

3. Probabilistic Forecasts Component

The MCP approach, widely described in the hydrological literature, is mostly used for hydrological forecasting [1,7]. In this paper, we extend its early applications to water demand forecasting [3,4]. The relevance of the approach lies in its ability to combine, in an extremely simple manner, one or, as in this study, several deterministic forecasts into a single predictive probability distribution. The parameters of the MCP predictive distributions at hourly time steps were recalibrated every new set of observed data, and deterministic forecasts became available.

4. Results and Discussion

The median of the MCP predictive distributions was considered a deterministic forecast in the calculation of performance indices. Table 1 shows the average indices obtained over the entire simulated period for all the District Metered Areas (DMAs) investigated, highlighting how in almost all cases, MCP leads to an improvement in performance compared to the single deterministic components. In addition, Figure 3 shows the median, 50%, and 90% confidence bands for hourly water demand probabilistic predictions derived from MCP, together with the deterministic forecasts provided by the ANN and NARX models, compared with observations from one of the analyzed districts. Although not discussed in this paper, operational real-time applications can highly benefit from forecast information provided in the form of a predictive density [4].

Table 1. Average performance indices obtained over the entire simulated period.

	MCP Median			ANN			NARX		
	P1	**P2**	**P3**	**P1**	**P2**	**P3**	**P1**	**P2**	**P3**
DMAA	**0.784** [1]	2.573	**1.012**	0.849	2.949	1.143	0.827	**2.394**	1.089
DMAB	**0.298**	0.967	**0.377**	0.305	**0.958**	0.419	0.356	1.114	0.486
DMAC	**0.261**	0.772	**0.346**	0.266	**0.770**	0.400	0.320	0.925	0.449
DMAD	**1.740**	**4.951**	**1.913**	1.833	5.100	2.087	1.857	5.281	2.052
DMAE	**1.093**	**3.559**	**1.399**	1.265	4.319	1.662	1.249	3.983	1.585
DMAF	**0.658**	**2.053**	**0.725**	0.697	2.189	0.806	0.698	2.130	0.773
DMAG	**0.750**	**2.330**	**0.884**	0.799	2.481	1.009	0.846	2.603	1.011
DMAH	**0.636**	**1.938**	**0.773**	0.717	2.138	0.931	0.714	2.161	0.859
DMAI	1.083	3.264	**1.246**	**1.077**	**3.189**	1.506	1.241	3.590	1.409
DMAJ	**0.977**	**2.829**	**1.070**	1.045	2.921	1.265	1.068	3.032	1.174

[1] The values in bold highlight the best performance.

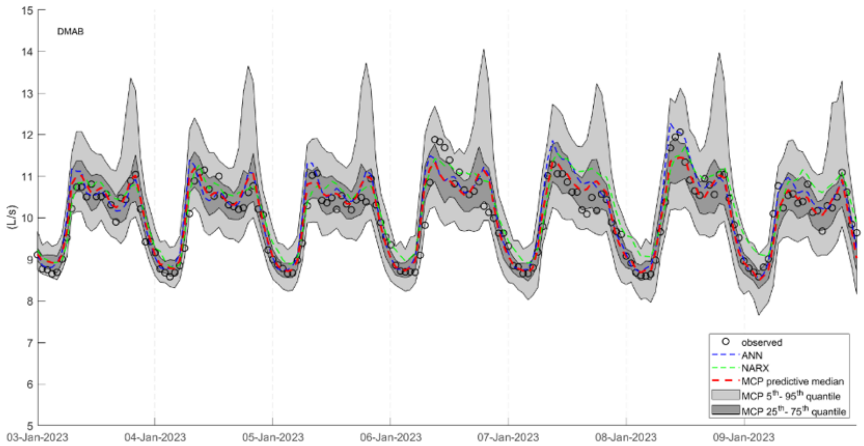


Figure 3. Comparison between observed values and deterministic forecasts (ANN, NARX, MCP median) of water demand for the week starting on 3 January 2023 in the DMAB district. The predictive probability distribution bands provided by MCP are also displayed.

Author Contributions: Conceptualization, A.G. and D.B.; methodology, E.T.; writing—original draft preparation, A.G. and D.B.; writing—review and editing, R.G. and E.T; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable: no new data were created or analyzed in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Todini, E. A Model Conditional Processor to Assess Predictive Uncertainty in Flood Forecasting. *Int. J. River Basin Manag.* **2008**, *6*, 123–137. [CrossRef]
2. Aggarwal, C.C. *Neural Networks and Deep Learning: A Textbook*; Springer: Cham, Switzerland, 2018.
3. Anele, A.O.; Hamam, Y.; Abu-Mahfouz, A.M.; Todini, E. Overview, Comparative Assessment and Recommendations of Forecasting Models for Short-Term Water Demand Prediction. *Water* **2017**, *9*, 887. [CrossRef]
4. Alvisi, S.; Franchini, M. Assessment of Predictive Uncertainty within the Framework of Water Demand Forecasting Using the Model Conditional Processor (MCP). *Urban Water J.* **2015**, *14*, 1–10. [CrossRef]
5. Anele, A.O.; Todini, E.; Hamam, Y.; Abu-Mahfouz, A.M. Predictive Uncertainty Estimation in Water Demand Forecasting Using the Model Conditional Processor. *Water* **2018**, *10*, 475. [CrossRef]

6. Gabriele, A.; Di Nunno, F.; Granata, F.; Gargano, R. Data-Driven Approaches for Quantitative and Qualitative Control of Urban Drainage Systems (Preliminary Results). In Proceedings of the Environmental Sciences Proceedings, Naples, Italy, 12–15 July 2022.
7. Coccia, G.; Todini, E. Recent Developments in Predictive Uncertainty Assessment Based on the Model Conditional Processor Approach. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 3253–3274. [CrossRef]