Battle of Water Networks

Battle of Water Demand Forecasting (BWDF)

(19/09/2023 – rev. 1)

(12/10/2023 – rev. 2)

1. Background and research direction

Population growth, urbanization, and climate change have been raising people's awareness about the impact of human activities on the environment and the available natural sources, such as water resources. In this context, sustainable management of water systems is crucial to avoid water shortage or the depletion of the available sources, and the operational and strategic decisions made by drinking water utilities can take benefit from a reliable and accurate forecast of water demand, which is the main driver of water distribution systems (WDSs).

Drinking water utilities typically need to estimate water demand and its pattern for the near future – such as the upcoming day or week – to operate their treatment plants and wells, or to appropriately regulate the adjustments of pumps, valves, and other controls to ensure water supply. However, water utilities may also need water-demand forecasts for a more distant future (e.g., upcoming years or decades) to rethink water sources, expand their treatment plants and develop new portions of the distribution network (Donkor et al. 2014). These latter forecasts can also be relevant in decision-making processes for urban water management, such as those related to pricing policies or water use restrictions (Herrera et al. 2010).

In the scientific literature, various methods – and their hybrids – have been developed for water demand forecasting. In general, demand-forecasting methods are based on the analysis of the recurring patterns and periodicities existing in water demand, and can be applied at different levels of temporal aggregation (Alvisi et al. 2007). Specifically, two main types of demand-forecasting models were developed, based on forecast horizon and frequency: on the one hand, *short-term* models, providing hourly or daily demand forecasts over time horizons of one day to one month, and mainly used for WDS operation and management; on the other hand, *long-term* models, providing monthly or yearly demand forecasts with a temporal horizon of one or many years, and typically used for resource allocation, system planning, and design. Therefore, water demand forecasting can be performed at different time scales according to the application of the model (Bakker et al. 2013). In general, soft-computing methods – such as Artificial-Neural-Network architectures, unsupervised methods, Deep-Learning methods, etc. – and time series and pattern-based methods are applied for short-term forecasting, whereas econometric models are more often used for longer horizons (Ghalehkhondabi et al. 2017). Indeed, a large set of tools for water demand forecasting are currently available, but the selection of the best technique to successfully solve a given problem is rather difficult (Herrera et al. 2010), and the results of water-demand forecasting practices – undertaken by water utilities and their consultants – may be extensively different based on the methods used (Donkor et al. 2014).

In light of the above, the *Battle of Water Demand Forecasting* (BWDF), organized in the context of the 3rd International WDSA-CCWI Joint Conference in Ferrara, Italy (https://wdsa-ccwi2024.it), aims at comparing the effectiveness of methods for the short-term prediction of urban water demand, relying on Supervisory Control and Data Acquisition (SCADA) measurements, and mass balance calculations, in a set of real District Metered Areas (DMAs). The water-demand forecasting problem described in the following section can be solved by applying different types of methods and approaches, including – but not limited to – engineering judgement, statistical methods, machine learning tools, and signal-processing models.

2. Goal of the challenge, materials, and rules

The challenge proposed in the context of the BWDF is focused on the forecast of the water demands with reference to a case-study WDN located in the North-East of Italy, supplying a variety of areas that are considerably different as far as characteristics, size, and average water demand are concerned. Specifically, forecasting is required for ten DMAs of the WDN concerned with the aim of defining optimal system operation for the near future (i.e., upcoming day and week), and optimizing the energy purchase. The water demand of each DMA is assumed to be represented by the *net inflow* and thus it includes all the types of water consumption and leakages of the DMA.

The main characteristics of each DMA – such as area description, number of users supplied, and average net inflow for the two years 2021 and 2022 – are summarized in Table 1.

Table 1: DMA characteristics.

DMA	Area characteristics	Number of users supplied	Average net inflow (L/s)
A	Hospital district	162	8.4
B	Residential district in the countryside	531	9.6
C	Residential district in the countryside	607	4.3
D	Suburban residential/commercial district	2094	32.9
E	Residential/commercial district close to the city centre	7955	78.3
F	Suburban district including sport facilities and office buildings	1135	8.1
G	Residential district close to the city centre	3180	25.1
H	City centre district	2901	20.8
I	Commercial/industrial district close to the port	425	20.6
J	Commercial/industrial district close to the port	776	26.4

The water utility managing the DMAs concerned provided the hourly net-inflow time series Q_{net} (L/s) for each DMA in relation to the period from 1 January 2021 to 31 March 2023. Net-inflow time series include water consumptions and leakages and are obtained through water balance as shown in Equation (1):

$$Q_{net} = \sum_{i=1}^{n_{in}} Q_{in,i} - \sum_{j=1}^{n_{out}} Q_{out,j} \qquad (1)$$

in which Q_{in} is the flow rate entering the DMA concerned through the i-th inlet point ($i = 1,2, \dots n_{in}$) and acquired by the water utility SCADA system, whereas Q_{out} is the flow rate outgoing from the DMA concerned through the j-th outlet point ($j = 1,2, \dots n_{out}$). It is worth noting that no DMAs with storage facilities are included in Table 1. Moreover, net-inflow data are not post-processed, so they can show some gaps related to SCADA system malfunctioning and other data collection/transmission issues.

Concerning the data provided for the challenge, historical net-inflow data are initially available for the overall period between week 1/2021 and week 29/2022 (see Figure 1a) to set up the forecasting model. Explicit indication of the days belonging to each week is also provided in the file *Calendar_rev1.pdf*.

All historical data can be downloaded from the conference website.

From an operational standpoint, teams will be first asked to forecast the hourly net-inflow time series of the DMAs concerned with reference to week 30/2022 (hereinafter referred to as evaluation week $W1$, as shown in Table 2) and to send their solution to the WDSA-CCWI Organizing Committee by 31 January 2024, along with the code to run the model and all input files to perform water-demand forecasting of evaluation week $W1$. Code must be provided together with a simple instruction file to run it.

Subsequent historical data will then be provided later, in early 2024, and water-demand forecasting will be asked in relation to additional evaluation weeks. In particular:

- Historical net-inflow data for the period between week 31/2022 and week 43/2022 will be made available by 5 February 2024, whereas teams will be asked to forecast the hourly net-inflow time series of the ten DMAs concerned with reference to week 44/2022 (evaluation week $W2$) and to send their solution to the WDSA-CCWI Organizing Committee by 15 February 2024 (see Figure 1b).
- Historical net-inflow data for the period between week 45/2022 and week 2/2023 will be made available by 19 February 2024, whereas teams will be asked to forecast the hourly net-inflow time series of the ten DMAs concerned with reference to week 3/2023 (evaluation week $W3$) and to send their solution to the WDSA-CCWI Organizing Committee by 29 February 2024 (see Figure 1c).
- Historical net-inflow data for the period between week 4/2023 and week 9/2023 will be made available by 4 March 2024, whereas teams will be asked to forecast the hourly net-inflow time series of the ten DMAs concerned with reference to week 10/2023 (evaluation week $W4$) and to send their solution to the WDSA-CCWI Organizing Committee by 15 March 2024 (see Figure 1d).

Figure 1. Historical data publication and data-forecast submission deadlines.

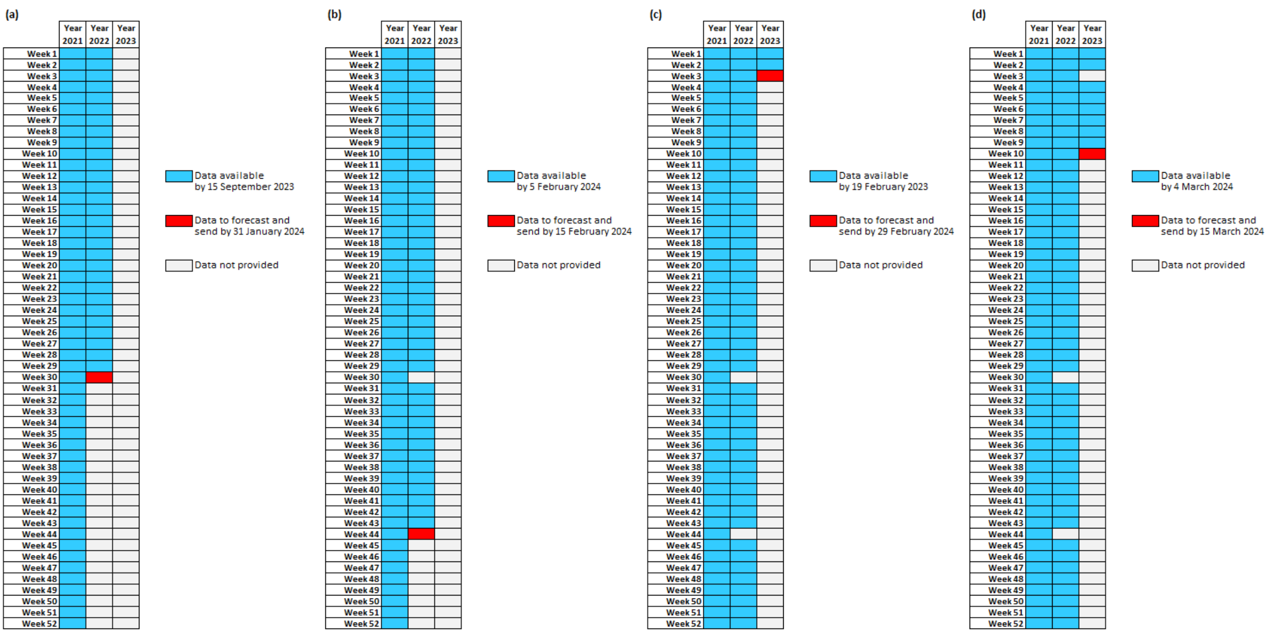

Table 2: Evaluation weeks and submission deadlines.

Evaluation week	Week number	Initial-final day	Deadline for forecast submission
W1	30/2022	25-31 July 2022	31 January 2024, 23:59 CET
W2	44/2022	31 October-6 November 2022	15 February 2024, 23:59 CET
W3	3/2023	16-22 January 2023	29 February 2024, 23:59 CET
W4	10/2023	6-12 March 2023	15 March 2024, 23:59 CET

It is worth noting that, in addition to net-inflow data, raw weather data (i.e. air temperature, rainfall depth, air humidity, and wind speed) observed at a weather station located within the case-study WDN will be progressively made available in relation to the periods considered. Like net-inflow data, raw weather data can be downloaded from the conference website.

Moreover, all days for which a peculiar pattern in the water consumption time series may be observed as opposed to working days (i.e., Sundays, holidays, local-event days) are marked in the “Calendar.pdf” file provided on the website of the conference, along with all the time changes from Central European Time (CET) to Central European Summer Time (CEST) and vice versa.

The final goal of the BWDF is to develop a new water-demand forecasting model – or apply/adapt one or more existing models – to accurately forecast water demands for all the DMAs selected.

In particular, the performance of the models will be evaluated with reference to:

a) the forecast of water demand (i.e. net inflow) at 1-hour time resolution in relation to the 24 hours making up to the initial day of each evaluation week;
b) the forecast of water demand (i.e. net inflow) at 1-hour time resolution in relation to the period between the second and the final day of each evaluation week.

Competition rules

- Participating teams can use net-inflow data and/or weather data and/or calendar as input for their water-demand forecasting models.
- With reference to weather data, it is worth noting that data referring to the evaluation week to forecast can be used as an input for the forecasting model together with data referring to the time windows preceding the evaluation week to forecast. Indeed, even though these observed data are actually not available when the forecast is performed, in the context of the BWDF they can be used assuming that they represent a “perfect” forecast of future weather data.
- Each participating team has to submit the forecasted demand (i.e. net inflow) time series for each DMA and evaluation week by filling in the dedicated solution template. The template file features ten Microsoft Excel® datasheets, one per DMA, where the team has to report the results of the forecasting for: *a*) the 24 hours making up to the initial day of the evaluation week considered; and *b*) to the period between the second and the final day of each evaluation week – for the selected evaluation week.

- The forecasted hourly net-inflow time series of each evaluation week must be submitted to the WDSA-CCWI Organizing Committee by the dates indicated in Table 2. If the solution of one or more evaluation weeks is not received until the respective deadline, the team(s) will be assigned to the lowest position(s) as far as the water demand-forecast performance for that evaluation week is concerned (see the subsequent section for further details about the evaluation process).
- In addition to forecasted results, each team has to send the code of its forecasting model – along with all input files to run the model to forecast water demand of evaluation week $W1$ – in the form of a *.zip* file by 31 January 2024. The code may be run by the BWDF Organizing Committee during the evaluation process to test the reproducibility of the method. However, the BWDF Organizing Committee reserves the right to ask teams for clarifications and/or run the water demand forecasting code also for subsequent evaluation weeks (i.e., $W2 - W4$) in the event of doubts as to the reproducibility of the method.
- All files (*.zip* folder including the code file and all input data to run water-demand forecast of evaluation week $W1$, individual solutions for evaluation weeks $W1 - W4$) have to be sent to *info@wdsa-ccwi2024.it* along with the all the information concerning the participating team (i.e., team name, participant names, and participant affiliations) by their respective deadlines.
- The submission of an abstract (maximum 500 words) is required to participate in the BWDF. The abstract has to be submitted through the EasyChair platform by December 15, 2023, as detailed in the "Abstract Call" page of the conference website. For all the accepted abstracts, the submission of a full paper will also be required by April 1, 2024, details of which will be provided in early 2024.

Finally, some restrictions for teams and their participants:

- Participants must not belong to more than one team.
- Teams are not allowed to submit more than one solution.
- As in the case of all the other WDSA-CCWI technical sessions, conference registration is needed for at least one member/author of each BWDF participating team.

3. Solution evaluation and team ranking

The effectiveness and accuracy of the forecasting models will be quantitatively assessed by considering the water demand (i.e. net inflow) forecasted for each DMA d ($d = 1, \dots, 10$) and evaluation week w ($w = 1, \dots, 4$). This is done through the assessment of three Performance Indicators (PI):

1. The mean-absolute error in relation to the 24 hours making up to the initial day of each evaluation week w (i.e., $PI1_w^d$, as shown in Equation 2)
2. The maximum-absolute error in relation to the 24 hours making up to the initial day of each evaluation week w (i.e., $PI2_w^d$, as shown in Equation 3)
3. The mean-absolute error in relation to the period between the second and the final day of each evaluation week (i.e., $PI3_w^d$, as shown in Equation 4)

$$PI1_w^d = \frac{1}{24}\sum_{h=1}^{24}\left|O_{d,w,h} - F_{d,w,h}\right| \tag{2}$$

$$PI2_w^d = \max\{\left|O_{d,w,1} - F_{d,w,1}\right|, \left|O_{d,w,1} - F_{d,w,1}\right|, \dots, \left|O_{d,w,24} - F_{d,w,24}\right|\} \tag{3}$$

$$PI3_w^d = \frac{1}{144}\sum_{h=25}^{168}\left|O_{d,w,h} - F_{d,w,h}\right| \tag{4}$$

where $O_{d,w,h}$ is the true value of the target variable (i.e. observed net-inflow in DMA d at time h of evaluation week w) and $F_{d,w,h}$ is the respective predicted value (i.e., forecasted net-inflow). Note that Performance Indicators $PI1$, $PI2$ and $PI3$ are expressed in the same units as the target variable (i.e., L/s).

For each DMA $d = 1, \dots, 10$, evaluation week $w = 1, \dots, 4$, and performance indicator $PI = 1, \dots, 3$, the solutions of the nt-teams will be ranked (assigning a rank-value equal to 1 to the team providing the lowest PI_w^d-value and a rank-value equal to nt to the team providing the highest PI_w^d-value). The final score of each team will be the sum of the rank values obtained for each DMA, week, and Performance Indicator. The team with the lowest score will be the winner of the challenge.

References

Alvisi, S., Franchini, M. and Marinelli, A. **2007**. "A short-term, pattern-based model for water-demand forecasting." *Journal of Hydroinformatics*, 9(19): 39-50. https://doi.org/10.2166/hydro.2006.016.

Bakker, M., Vreeburg, J. H. G., van Schagen, K. M. and Rietveld, L. C. **2013**. "A fully adaptive forecasting model for short-term drinking water demand." *Environmental Modelling & Software*, 48: 141-151. http://dx.doi.org/10.1016/j.envsoft.2013.06.012

Donkor, E. A., Mazzucchi T. A., Soyer, R., and Robertson, J. A. **2014**. "Urban Water Demand Forecasting: Review of Methods and Models." *Journal of Water Resources Planning and Management*, 140(2): 146-159. https://doi.org/10.1061/%28ASCE%29WR.1943-5452.0000314.

Ghalehkhondabi, I., Ardjmand, E., Young II, W. A., and Weckman, G.R. **2017**. "Water demand forecasting: review of soft computing methods." *Environmental Monitoring and Assessment*, 189: 313. https://doi.org/10.1007/s10661-017-6030-3.

Herrera, M., Torgo, L., Izquierdo, J., and Pérez-García, R. **2010**. "Predictive models for forecasting hourly urban water demand." *Journal of Hydrology*, 387: 141-150. https://doi.org/10.1016/j.jhydrol.2010.04.005.