ORIGINAL ARTICLE

Water consumption time series forecasting in urban centers using deep neural networks

C. G. García-Soto[1] · J. F. Torres[2] · M. A. Zamora-Izquierdo[3] · J. Palma[3] · A. Troncoso[2]

Abstract

The time series analysis and prediction techniques are highly valued in many application fields, such as economy, medicine and biology, environmental sciences or meteorology, among others. In the last years, there is a growing interest in the sustainable and optimal management of a resource as scarce as essential: the water. Forecasting techniques for water management can be used for different time horizons from the planning of constructions that can respond to long-term needs, to the detection of anomalies in the operation of facilities or the optimization of the operation in the short and medium term. In this paper, a deep neural network is specifically designed to predict water consumption in the short-term. Results are reported using the time series of water consumption for a year and a half measured with 10-min frequency in the city of Murcia, the seventh largest city in Spain by number of inhabitants. The results are compared with K Nearest Neighbors, Random Forest, Extreme Gradient Boosting, Seasonal Autoregressive Integrated Moving Average and two persistence models as naive methods, showing the proposed deep learning model the most accurate results.

Keywords Time series forecasting · Deep neural networks · Water demand

Introduction

It is well known that water is an essential resource for economic development, for obtaining food, for the availability of healthy ecosystems and, in short, for the survival of living beings. However, the water availability is becoming increasingly limited due to the rapid growth of the world population. An adequate management, understood as the activity of planning, developing and distributing of resources, is essential to optimize the use of water.

The companies and entities managing the supply of drinking water have the objective of supplying the demand of the consumers every day with the greatest possible efficiency. However, in many cases the operation is only managed to cover the instantaneous water demand without using any advanced technique for predicting consumption. Under these conditions, being able to predict in advance the pattern of demand in the short term (1-48 h) is a valuable tool for optimizing the management of water reserves and the use of associated equipment. Thus, for example, it could allow planning the schedules of the supply pumps to take advantage of the periods with more economic tariffs. Several authors have quantified how operation planning based on demand prediction can lead to energy cost savings, in many cases in excess of 18% (Cembrano et al. 2000; Salomons et al. 2017; Kang et al. 2014).

In this environment, the analysis of water-related time series and, in particular, their prediction, is a tool that can help improve the management of the integral water cycle for drinking water supply or crops irrigation, waste water

✉ A. Troncoso
atrolor@upo.es

C. G. García-Soto
carlosgabriel.garcia@emuasa.es

J. F. Torres
jftormal@upo.es

M. A. Zamora-Izquierdo
mzamora@um.es

J. Palma
jtpalma@um.es

[1] Department of water distribution systems (Emuasa), Murcia, Spain

[2] Data Science and Big Data Lab, Pablo de Olavide University, 41013 Seville, Spain

[3] Information and Communication Technology Oriented Research Institute (INTICO), University of Murcia, 30100 Murcia, Spain

generation, natural sources, etc. Within this field, this work focuses on the analysis and prediction of drinking water consumption in Murcia, which is a city located in southeastern Spain.

In general, the analysis of time series has two main objectives: to identify the nature of the phenomena represented by this time series and its prediction. The forecasting techniques based on time series models (Trull et al. 2019, 2020) have been widely developed and applied in very diverse disciplines, such as economics, meteorology, medicine or resource management.

In the last decades, machine learning techniques (Talavera-Llames et al. 2016, 2019) from the Artificial Intelligence field have been successfully applied to forecasting problems, in particular artificial neural networks (ANN) (Rana et al. 2014; Lin et al. 2019). In recent years the enormous amount of time series measurements collected from smart devices has made deep learning necessary, thus giving birth to deep neural networks (DNN) (Torres et al. 2018, 2019).

In this work, we propose a DNN for the purpose of prediction of the water demand. First, an analysis of the dataset composed of the water consumption measurements collected every 10 min is carried out. Then, the methodology based on a deep feed forward network is presented, providing an improved and robust way to evaluate the learning of a time series model preserving its temporal order. An exhaustive experimentation using real-world water consumption data is provided, obtaining an error of 3% approximately. Finally, a comparison is also performed, making use of the k nearest neighbors, random forest, extreme gradient boosting, a classical time series model and two persistence models based on the real values of the previous day or week.

In summary, the main contributions of this work can be summarized as follows:

1. A deep learning model specifically designed for water consumption forecasting.
2. A robust way to evaluate the learning of a time series model preserving its temporal order.
3. Analysis of the behaviour of the water consumption in a city of Spain.
4. Reported error results of 3% for the real-world water consumption.
5. Comparison of prediction accuracy with other state-of-the-art forecasting methods and statistical test in order to validate the results.

The rest of the paper is structured as follows: the previous researches related to the paper's topics are presented in Sect. 2. Section 3 defines the forecasting problem to be solved. Section 4 summarizes the main characteristics of the water consumption of Murcia and describes how the

proposed algorithm works and the methodology carried out in order to evaluate a time series model. The experimental setting and the results obtained are shown in Sect. 5. Finally, the conclusions and future works are provided in Sect. 6.

State of the art

This section reviews all recently published works related to water demand forecasting.

Classical time series models have shown to be competitive for water consumption forecasting problems. The authors in Anele et al. (2017) obtained predictions of water consumption in southwest Spain by means of AR, MA, ARMA and ARMAX models using water consumption data combined with meteorological information. In Lee and Derrible (2020), Lee et al. applied regression techniques, namely linear, Lasso and Bayesian, to predict daily water consumption using demographic data and housing information. The regression models were compared with other widely used machine learning techniques such as gradient boosting (GB).

The prediction of water demand using machine learning techniques has been intensively studied in recent years due to the increasing availability of easy access to large amounts of data. One of the most widely used approaches has been the tree-based techniques. Nunes-Carvalho et al. trained a random forest (RF) model, among others, using socio-demographic information and historical water consumption data, to predict water demand patterns in the city of Fortaleza, Brazil (Nunes-Carvalho et al. 2021). Bolorinos et al. trained a RF method for detecting changes in consumption (Bolorinos et al. 2020). A RF was also used to forecast daily consumption in southwest China in Chen et al. (2017). A tree-based model, namely a GB, was proposed by Xenochristou et al. in Xenochristou et al. (2020) to predict water demand at different scales and to establish a comparison between the results obtained for each one of them. In Villarin and Rodriguez-Galiano (2019), the authors compared the performance of classification and regression trees (CART) and RF to forecast time series of water demand in the city of Seville in Spain.

Several works published in the last years proposed the support vector regression (SVR) method to obtain accurate predictions of water consumption. Chen et al. designed a model based on SVR to predict hourly water demand using two different data sources in order to optimize pumping operations and to detect anomalies (Candelieri 2017). A least squares SVR was also applied to predict residential, industrial and commercial water demand in the city of Bogotá, Colombia in Peña-Guzmán et al. (2016). Different machine learning models for forecasting water demand were compared in Herrera et al. (2010) using data from an urban area in a city in southeastern Spain. In particular, artificial

neural networks, projection pursuit regression, multivariate adaptive regression splines, random forests and SVR were tested, obtaining the SVR method the best results.

Recently, many architectures of neural networks have been also proposed for water consumption forecasting. Ghiassi et al. (2017) used two neural networks and a model based on nearest neighbors for daily, weekly and monthly forecasting of water demand in the city of Tehran, obtaining highly competitive results. A deep belief network was performed by the authors in Xu et al. (2019) for the prediction of hourly water demand. In Mouatadid and Adamowski (2017) the performance of various machine learning methods was evaluated to forecast urban water demand for one day and three days ahead, with the extreme learning machine (ELM) model having the lowest prediction error.

Finally, ensemble models are booming because they tend to achieve better results than a stand-alone method. A technique based on stacking models, including artificial neural networks and deep learning architectures, to predict daily water demand using real data from United Kingdom was proposed in Xenochristou and Kapelan (2020). Ambrosio et al. combined different models, including the multilayer perceptron, for water demand prediction in Ambrosio et al. (2019). A weighted strategy that gathers the advantages of the different machine learning techniques such as neural networks, random forests, support vector machines and k-nearest neighbors was suggested in Antunes et al. (2018) and compared with an autoregressive integrated moving average (ARIMA).

In addition to forecasting tasks, other studies have also been carried out to analyze the water consumption. For instance, the authors in Coelho et al. (2017) proposed a metaheuristic based on deep learning and graphic processing units (GPU) to analyze time series of water consumption in big data environments. Clustering techniques have been also applied to water consumption data. The application of a mixture of non-homogeneous hidden Markov models to cluster time series that share the same transition dynamics was proposed in Leyli-Abadi et al. (2018). A similar study was carried out in Padulano and Giudice (2018), where first clustering and then classification techniques were applied to data from consumption meters in a household in Soccavo, in the city of Naples (Italy).

A correct selection of the predictive variables is important since a large number of features does not always leads to a significant improvement in the results. Some authors have used climatological, population or even urban mobility data as predictive variables, in addition to the previous values of water consumption (Smolak et al. 2020).

In summary, it can be concluded that water forecasts have been made in many different geographic areas and population centers. Antunes et al. (2018) obtained forecasts of the water demand for two cities in Portugal. Tiwari et al. (2016),

as well as Bougadis et al. (2005); Bata et al. (2020), used several population centers in Canada. Smolak et al. (2020) provided predictions of water consumption for several towns in Poland and (Duerr et al. 2018) for several towns in Florida (USA). Pacchin et al. (2019) and Gagliardi et al. (2017) carried out a comparison of different prediction techniques applied to some places in Italy. Ren and Li (2016) obtained consumption predictions for the city of Shanghai in China. And other works developed the water prediction at the level of individual users, such as households or certain businesses and industries (Rahim et al. 2019; Farah et al. 2019).

With respect to prediction horizons, although most of the predictions are made for short-term, the most common prediction horizon being 24 h, several authors made forecasts for longer periods, such as weeks, months, or even years (Bata et al. 2020; Tian and Xue 2017).

Although all previously cited works present significant differences regarding the models or even the scope in some cases, a summary of results is provided in order to offer a general overview about the performance. Antunes et al. (2018) obtained a mean absolute percentage error (MAPE) between 8.3% and 17.6% for the next 24 h using an ensemble of models. Recently, Bata et al. in Bata et al. (2020) obtained a MAPE of 12.3% for the day-ahead water forecasting and Smolak et al. of 9.6% in Smolak et al. (2020).

After a thorough review of the previously published works, it can be concluded that machine learning techniques have generally provided better results than classical techniques, but also that there is no optimal model of machine learning that is the most appropriate for all cases. On the contrary, several works (Makridakis et al. 2018) concluded that classical prediction methods may have better performance than those based on machine learning in the prediction of certain time series. These two points reinforce the idea that it is necessary to analyze each case with its particularities.

Problem description

In this paper we will analyze and predict the demand of drinking water in the short term in the city of Murcia, located in southeastern Spain, one of the areas of Europe suffering greater water stress. These predictions could be used later for two purposes: the optimization of its management and the detection of anomalies.

The goal of the time series analysis is to obtain mathematical models that allow to explain the behavior observed in a time series and that can be applied to the prediction of future values. To do this, we propose to develop a model of machine learning, based on deep neural networks, as accurately as possible, for the drinking water demand forecasting in the city of Murcia in the short-term, namely, four hours.

As a time horizon of prediction we have considered a value of four hours, since it is a sufficient time to plan some of the main tasks carried out every day in the management systems of a city of these characteristics. To obtain a prediction of the water consumption for the next 4 h, we will need to make a multi-step prediction as the samples are acquired every 10 min. Therefore, the model will provide 24 values in each run. In addition, the required computations to obtain the prediction must be performed every 10 min.

Finally, these predictions would then be used to optimize the operation and to detect anomalous consumption patterns due to breakdowns in the distribution network.

Proposed methodology

Data

The city of Murcia is located in the region of Murcia in southeast Spain and has a population of 453258 inhabitants, with an average annual growth of 0.8% in the last five years. It is the seventh largest city in Spain in terms of population and geographically includes 52 districts covering an area of 882 km^2 as shown in Fig. 1. The network of distribution pipes managed by the municipal water company of Murcia reaches 2203 km, and the consumption of drinking water per inhabitant is approximately 185 ls per day. The use of the water includes mainly domestic, industrial, service and garden irrigation.

The drinking water consumption data consists of measurements in cubic meters per hour (m^3/h) collected by the supervisory, control and data acquisition system of the company that manages water in Murcia. The data are recorded with a frequency of 10 min from January 1, 2019 to June 30, 2020. In short, the starting dataset is composed of 78773 samples and a summary of the main statistical values is shown in Table 1.

Figure 2 shows the water consumption in Murcia from January 2019 to December 2019 divided into quarters. It can be observed high seasonality, as well as that the consumption remains at stable values for most of the year, although it was significantly reduced during the summer period of 2019. This is possibly caused by the decline in the city's population during the holiday period.

Figure 3 presents the values of water consumption for the week comprising the days from Monday 21 to Sunday 27 January 2019. It can be noted that working days from Monday to Friday show a similar pattern. However, weekends and holidays have a different consumption pattern related to the change in activity and schedules.

Figure 4 depicts the water demand for one working day. It follows a pattern according to the activity and habits of the day, that is, consumption is very low in the early morning, with a peak in the early afternoon that decreases, and increases again in the early evening.

Figure 5 shows the water consumption for several weeks. It is not always known in detail what causes the variations that are observed, for example, between working days. They are sure to be very varied from the appearance of breakdowns to specific demands from large consumers or the presence of large events in the city. It should be noted that similar patterns occur in other utilities required by our society such as electricity demand (Galicia et al. 2018; Troncoso et al. 2004) or transportation (Yasdi 1999).

Deep neural networks

There are currently a large number of DNN architectures such as feed forward, convolutional or recurrent networks, each specially designed for a particular type of application

Fig. 1 Location of the city of Murcia in the region of Murcia

Table 1 Descriptive statistics of the water consumption time series

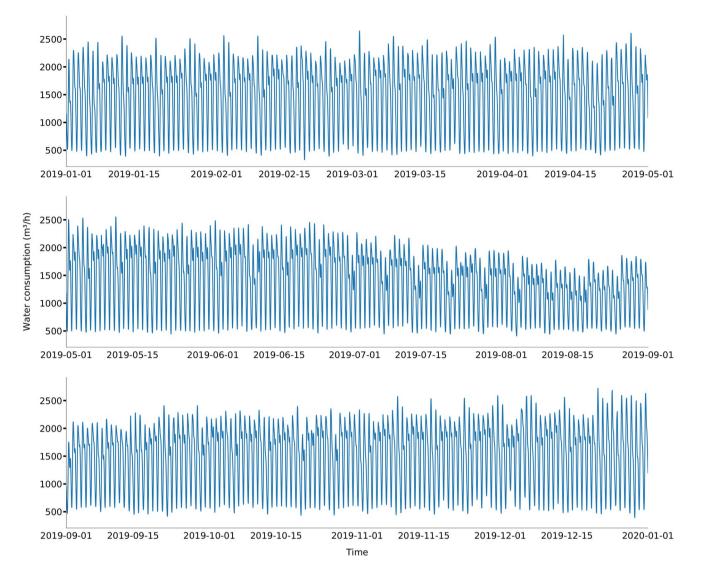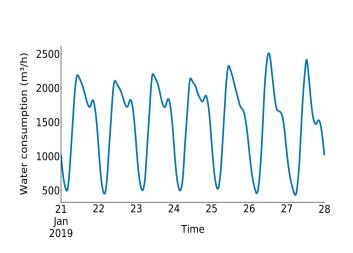| Measures	Value (m^3/h)
Mean	1487
Median	1621
Standard deviation	599
Maximum	2871
Minimum	301
Percentile 25%	934
Percentile 75%	1957

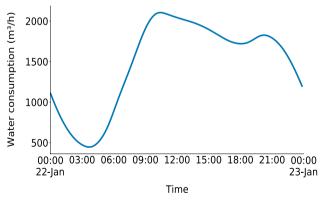**Fig. 2** Water consumption from January 2019 to December 2019 divided into quarters

Fig. 3 Water consumption for a particular week

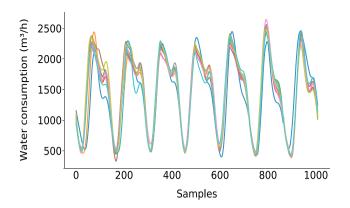**Fig. 4** Water consumption for a particular working day

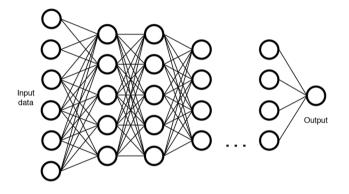**Fig. 5** Water consumption for several weeks

Fig. 6 Basic architecture of a DFFNN for time series forecasting

or data. A full survey of the deep learning for time series forecasting can be found in Torres et al. (2021).

In this work, a Deep Feed Forward Neural Network (DFFNN) has been designed for the water consumption forecasting. Its main advantages are to be able to learn both linear and nonlinear relationships present in the time series, the possibility of making multi-step and multivariate predictions and it needs fewer assumptions in its modeling compared with other techniques. On the other hand, deep learning techniques also have a number of drawbacks such as very poorly interpretable models or a high number of hyper-parameters. Although the explainability or interpretability of the models may be very relevant in other types of applications such as medicine or finance, among others, it is not for the water consumption forecasting.

DFFNN, also called multi-layer perceptron, arose due to the inability of single-layer neural networks to learn certain functions. The architecture of a DFFNN is composed of an input layer, an output layer and different hidden layers as shown in Fig. 6. In addition, each hidden layer has a certain number of neurons to be determined.

The relationships between the neurons of two consecutive layers are modelled by weights, which are calculated during the training phase of the network. In particular, the weights

are computed by minimizing a cost function by means of gradient descent optimization methods. Then, the back-propagation algorithm is used to calculate the gradient of the cost function. Once the weights are computed, the values of the output neurons of the network are obtained using a feed forward process defined by the following equation:

$$a^l = g(W_a^l a^{l-1} + b_a^l) \tag{1}$$

where a^l are the activation values in the l-th layer, that is, a vector composed of the values of the neurons of the l-th layer, W_a^l and b_a^l are the weights and bias corresponding to the l-th layer, and g is the activation function. Therefore, the a^l values are computed using the activation values of the $l-1$ layer, a^{l-1}, as input. In time series forecasting, the rectified linear unit function (ReLU) is commonly used as activation function for all layers, except for the output layer to obtain the predicted values which generally uses the hyperbolic tangent function (*tanh*).

For all network architectures, the values of some hyper-parameters have to be chosen in advance. These hyper-parameters, such as the number of layers and the number of neurons, define the network architecture, and other hyper-parameters, such as the learning rate, the momentum, number of iterations or minibatch size, among others, have a great influence on the convergence of the gradient descent methods. The optimal choice of these hyper-parameters is important as these values greatly influence the prediction results obtained by the network. The hyper-parameters will be discussed in more detail in Sect. 5.3.

Model evaluation

Classical techniques for the selection and evaluation of machine learning models have limitations when applied to time series forecasting. Thus, the hold-out technique with a single training and test set involves arbitrarily selecting a set of test. This set of test will correspond only to the final temporal range of the available values of the time series. Thus, an error measure that is not very representative of the model's predictive performance can be obtained when applied at any other timestamp of the time series. However, the classical k-fold cross-validation implies not respecting the temporal order of the samples, an essential feature in time series (Bergmeir and Benítez 2012).

In this work, a nested cross-validation technique is used (Varma and Simon 2006). With this evaluation technique, the water consumption time series is studied in different time ranges, repeating the training and testing process for each of these ranges. Finally, a more robust and representative final error is obtained. This error is the average of the errors obtained for each aforementioned time range, as is depicted in Fig. 7. For the proposed DFFNN model, the re-training
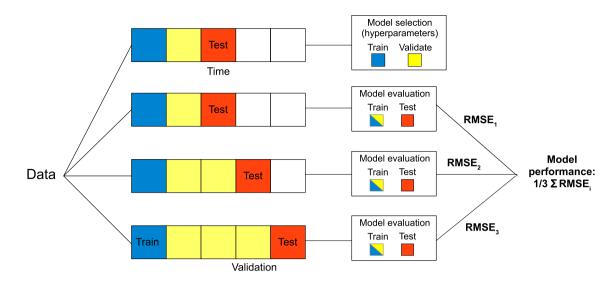
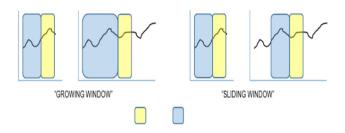**Fig. 7** Nested cross-validation procedure for model evaluation (adaptation of Cochrane et al. 2021)

Fig. 8 Growing window versus sliding window

process is repeated for 10 different periods, using the datasets composed of the first 6, 10, 11, 12, 13, 14, 15, 16, 17 and 18 months, respectively.

The proposed model was periodically re-trained with all available data as the DFFNN model can obtain better results using a larger amount of data. Thus, a growing window strategy is applied instead of the typical sliding window, as shown in Fig. 8. The historical window of values used for each forecast is the number of neurons for the input layer of the DFFNN model and it is one of the parameters to be optimized. In this work, the percentage distribution of the data for the training, validation and test sets are 60%, 15% and 25%, respectively.

Results

Quality measures

Four well-established metrics in the context of time series have been chosen in order to evaluate the performance of the DFFNN model proposed in this work.

The mean absolute percentage error (MAPE) is a relative error expressed as a percentage. It is used as a guideline to measure the goodness of the prediction method when comparing to other models:

$$\text{MAPE}(\%) = \frac{100}{n} \sum_{t=1}^{n} \frac{|y_t - \widehat{y}_t|}{y_t} \tag{2}$$

The mean absolute error (MAE), expressed in m^3/h, indicates the average deviation between actual and predicted values:

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^{n} |y_t - \widehat{y}_t| \tag{3}$$

The root mean squared error (RMSE), expressed in m^3/h, is the square root of the average of squared differences between predicted and actual values. By using the squared values, all of them are forced to have a positive value and the errors of greater magnitude have, proportionally, a higher weight in the result.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^{n} |y_t - \widehat{y}_t|^2} \tag{4}$$

Finally, the coefficient of determination R^2 provides a measure of the accuracy with which predictions match actual values. Its value is between 0 and 1, indicating poor fit or perfect fit, respectively.

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - \widehat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \tag{5}$$

For all the equations above, y_t represents the actual value of the time series, $\hat{y}_t$ represents the forecasted value, n represents the number of points included in the prediction and $\bar{y}$ denotes the mean of the time series values.

Preprocessing

The quality of the input data is essential for any deep learning model to obtain accurate predictions. Therefore, an analysis of the performance of the DFFNN when applying different preprocessing techniques has been carried out.

The time series has a total of 33 missing values and no values equal to 0, null, or negative. In order to determine which technique is the most appropriate for the imputation of missing values, some values from the time series have been randomly removed. The assignation of these missing values has been performed using different methods such as forward fill, backward fill, linear interpolation, linear fill, cubic fill, mean of k nearest neighbors and seasonal mean. Then, the mean square error (MSE) is computed for a training set and the method providing the lowest MSE is selected. The best results have been obtained using linear interpolation.

In a time series the presence of statistically anomalous values or outliers is common. Some outliers can be simply due to the presence of errors in the system for measuring and recording the water consumption data. However, other outliers may be caused by real variations in consumption as undesired punctual situations (breakdowns in the transport and distribution networks), or occasional demands from large consumers (municipal swimming pools, industries, large events, etc.), which cannot always be known in advance. It is recommended to keep the outliers corresponding to high consumption that may occur periodically and that are not caused by failures for model learning. However, both types of outliers are indistinguishable and considering that our DFFNN model has a significant tolerance to the presence of these anomalous values, no special treatment for outliers has been considered.

The time series includes consumption values ranging from 301 to 2871 m^3/h. It is known that the gradient descent technique used by the DFFNN model in the training phase works better if the variables are in a smaller range, being able to converge more quickly to its solution. The effect of standardization and scaling transformations to the ranges [0, 1] and [−1, 1] has been tested. It was observed the range [−1, 1] provided the best results.

Finally, transformations have been performed to make the time series of water consumption stationary, without obtaining any improvement in the accuracy of the predictions.

Table 2 shows a summary of the different techniques applied to preprocess the water consumption data regarding missing values, feature scaling and transformation to

Table 2 Summary of preprocessing techniques

	Preprocessing	Optimal
Missing values	Forward fill	Linear interpolation
	Backward fill	
	Linear interpolation	
	Linear fill	
	Cubic fill	
	Mean of k nearest neighbors	
	Seasonal mean	
Feature scaling	Normalization [-1,1]	Normalization [-1, 1]
	Normalization [0, 1]	
	Standardization	
Stationary	Yes, No	No

stationary time series and the technique selected according to the lowest mean square error.

Hyper-parameters

Most machine learning algorithms require the selection of several parameters, which are not directly learned by the model. These are called hyper-parameters. The hyper-parameters to be optimized for the DFFNN model are shown in Table 3. With the objective of minimizing the MSE, a grid search strategy was used to find the best values for the hyper-parameters. Thus, once the best parameters have been obtained from different possible combinations, the final model is trained. For the rest of the methods used in the comparison, all hyperparameters were optimized following the same grid search strategy. The most widespread search thresholds in the literature were established.

Analysis of results

Figure 9 illustrates a comparison between the original and predicted values by the DFFNN. It can be seen how the actual and predicted values are quite similar and how the forecast has been able to capture the seasonal component of the original series and differentiate the behavior between weekdays and weekends.
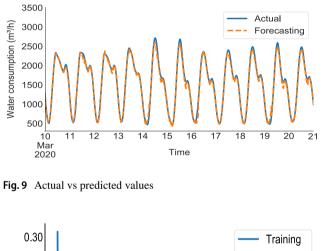
Furthermore, the evolution of the MSE loss function in the training and validation phases indicates that the model obtained does not have significant overfitting or underfitting, as illustrated in Fig. 10.

Figure 11 shows the correlation between the actual values and forecasted values obtained by the DFFNN model for the test set. A R^2 value of 0.987 is displayed, showing how good the predictions are.

Table 4 presents the largest errors obtained by the DFFNN model for the test set, ordered from largest to smallest. It can be observed that three errors correspond to the

Table 3 Hyper-parameter search for DFFNN model

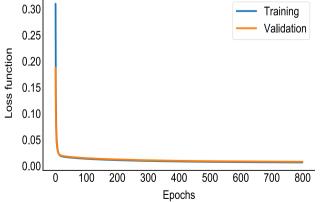| Hyper-parameter	Values	Optimal
Number of layers	2, 3, 4, 5	4
Number of neurons for input layer (historical window)	72 (12 h) 144 (24 h) 216 (36 h)	24 h
Number of neurons for hidden layers	50, 100, 150, 200, 250, 300	250
Activation function for hidden layers	ReLU, Leaky ReLU tanh, sigmoid, linear	tanh
Activation function for output layer	ReLU, tanh, linear	linear
Number of epochs	400, 800, 1600	800
Mini-batch size	256	256
Learning rate	0.001, 0.01, 0.1, 0.5, 1 (fixed/dynamic reduction)	0.5 (dynamic reduction)
Optimization algorithm	SGD, RMSprop Adam, Adagrad	SGD
Loss function	MSE, MAE	MSE
Weights initialization	Glorot uniform Glorot normal uniform, normal	Glorot uniform
Regularization	With/Without dropout	Without dropout
Batch normalization	Yes/No	No

Fig. 9 Actual vs predicted values

Fig. 10 Evolution of the loss function versus the number of epochs for the training and validation sets

Fig. 11 Correlation between actual and predicted values for the test set

early morning of 24 March 2020. However, water consumption was higher than usual for those hours, possibly due to a breakdown or some other incident, as shown in Fig. 12.

The residuals are the difference between the time series and the predictions obtained by the forecasting model for the training set. An uncorrelated residual with a mean of zero indicates that the forecasting method is able to

Table 4 Maximum absolute errors for the DFFNN model

Timestamp	Real value	Predicted value	Absolute error
2020-03-24 02:50	944.24	535.95	408.30
2020-06-05 10:50	2488.82	2095.90	392.92
2020-03-24 02:40	946.46	557.96	388.50
2020-03-24 02:30	948.62	567.02	381.60
2020-03-28 10:50	2433.17	2054.26	378.90

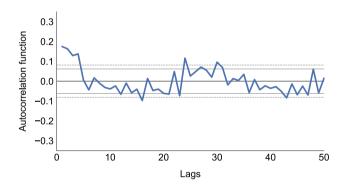**Fig. 12** Actual vs predictions for the days from March 22 to March 25, 2020

Fig. 13 Autocorrelation function of the residuals for the DFFNN model

model most of the information available in the original data. This does not ensure that the model has a good performance when predicting the test set, but it suggests that there is little room for improvement with the available information. On the other hand, if these conditions are not met, it is important to clarify that the model can still provide predictions that satisfy the expectations according to the errors metrics depending on the application under study. Figure 13 shows the residual errors obtained by the DFFNN model. From the autocorrelation function, it can be observed that the residuals of the predictions model a

white noise. Most of the values have a low value, below the 95% (solid line) and 99% (dotted line) confidence band.

Comparison with benchmarking methods

In order to compare the performance of the proposed DFFNN model to other possible forecasting techniques, six methods are considered such as *K* Nearest Neigbors (KNN), Random Forest (RF), Extreme Gradient Boosting (XGBoost), a Seasonal Autoregressive Integrated Moving Average (SARIMA) model and two baseline models.

The KNN has been successfully applied to obtain predictions of energy consumption in recent years and the prediction is based on the weighted linear combination of the time series values following in time order to the nearest neighbors, where the weights are determined depending on the distance of the neighbors to the past values (Talavera-Llames et al. 2019). In this work, the distance for the calculation of the neighbors has been the Manhattan distance and a single close neighbor has been considered.

RF and XGBoost are two methods based on ensembles of trees, but the training processes are very different. XGBoost train one tree at a time, while RF can train multiple trees in parallel. After extensive experimentation, 18 and 200 trees of maximum depth 14 and 2 have been used for RF and XGBoost, respectively.

The baseline models are based on a persistence algorithm, i.e., the prediction for a future time instant has the same value as in previous instants, so it represents a high correlation. In some works, this approximation is also known as seasonal naive (Livera et al. 2011). From the consumption patterns and correlation plots, a similarity between the measurement at instant t and the same instant of the previous day or week can be seen. Mathematically, the prediction is computed as follows:

$$\hat{y}_t = y_{t-144} \tag{6}$$

$$\hat{y}_t = y_{t-1008} \tag{7}$$

where $\hat{y}_t$ denotes the predicted value at instant t and y_{t-144} and y_{t-1008} the actual values of the time series at same time instant of the previous day or week, respectively.

On the other hand, the performance of the DFFNN model has been compared with a classical time series model, in particular, the SARIMA model. This model has been successfully used in a large number of practical problems and offers a high interpretability of the results, being also able to obtain well-defined confidence intervals in the predictions (Arunraj et al. 2016). As for the disadvantages, it can only extract the linear relationships present in the time series. SARIMA is an extension of the ARIMA model for univariate time series, which also includes a seasonal component.

For this reason, SARIMA is of special interest in time series that exhibit periodic characteristics such as the time series of water consumption. The SARIMA model has 7 hyperparameters: p, d and q for the autoregressive, differential and moving average components, respectively, and P, D and Q for these same components of the seasonal part, and finally, a value m including the number of samples for a single seasonal period. As in the case of the DFFNN model, a grid search has been used to find the best SARIMA model configuration. The metric used has been the Akaike information criterion (AIC), which allows to compare the performance of different statistical models. The AIC value is lower as the model output has a higher similarity to the data, but it also adds a penalty term depending on the number of hyperparameters in the model in order to avoid overfitting. Therefore, a lower value of the AIC indicates a better model fit.

Table 5 shows the optimal values for the hyperparameters of the SARIMA model.

Figure 14 shows the prediction made by the SARIMA model for the week of October 8–13, 2019, including the 95% confidence interval (shaded in grey colour). A certain similarity with the real series can be observed, but the error is significant at some time points. Even so, the model has been able to capture a good part of the seasonality of the water consumption.

For the SARIMA model, the mean of the residuals is practically zero, but the residuals of the predictions show significant correlations, as shown in the correlogram and histogram in Fig. 15. Therefore, very accurate predictions are not expected by the SARIMA model.

Table 6 shows the average of the MAPE, MAE, RSME and R^2 errors when predicting the test set for a total of 10 runs. The DFFNN model provides the best performance. The second best method is the RF, although it is 0.7% above the DFFNN. The persistence model has the advantage of its great simplicity, although it obtains greater errors than the DFFNN model and all other methods, except the SARIMA model. The SARIMA model does not improve the performance of the persistence models, which confirms, once again, that a more complex model does not necessarily always give better results. In addition, the predictions
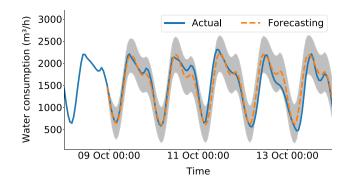

Fig. 14 Actual versus predicted values by the SARIMA for the week from October 8 to October 13, 2019

obtained by the DFFNN model range within a small interval as the standard deviation is low for the water consumption.

In order to increase the confidence in the results, a statistical significance test has been used, in particular, the Wilcoxon test (García et al. 2010). The Wilcoxon test is non-parametric, i.e. it does not assume a specific distribution of the data and is suitable for use with paired results. The null hypothesis in the Wilcoxon test consists of assuming that the results being compared come from the same population, and that, therefore, they have the same statistical parameters. In this study, a value of 0.05 has been considered as the level of significance α. If the p-value obtained from the test set is less than α, it can be concluded that the distributions of the results are different, and therefore, the observed differences are not random, i. e. the differences between forecasting methods are statistically significant. Table 7 shows
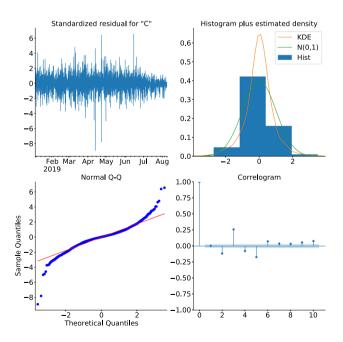
Table 5 Hyperparameter search for SARIMA model	Hyperparameter	Values	Optimal
	p	0, 1, 2, 3	3
	d	0, 1, 2, 3	0
	q	0, 1, 2, 3	1
	P	0, 1, 2, 3	3
	D	0, 1, 2, 3	1
	Q	0, 1, 2, 3	1
	m	24, 1268	24

Fig. 15 Diagnosis of the residuals for the SARIMA model

Table 6 Errors for DFFNN, KNN, RF, XGBoost, SARIMA and persistence models for the test set

		DFFNN	KNN	RF	XGBoost	SARIMA	Baseline 1 day	Baseline 1 week
MAPE	Mean	2.99	5.20	3.69	4.85	7.94	7.79	5.62
	St. dev	0.21	0.79	0.29	0.25	0.76	0.81	0.42
MAE	Mean	45.15	78.36	56.22	73.96	119.60	118.70	86.75
	St. dev	2.32	10.39	4.38	4.36	12.05	13.14	5.92
RMSE	Mean	66.75	111.55	85.85	107.21	174.50	172.14	127.80
	St. dev	2.95	16.29	6.01	5.60	16.45	17.57	11.10
R^2	Mean	0.987	0.961	0.979	0.966	0.905	0.911	0.952
	St. dev	0.002	0.015	0.002	0.005	0.024	0.024	0.010

Table 7 Statistical tests for the MAPE for all algorithms

	DFFNN	KNN	RF	XGBoost	SARIMA	Baseline 1 day	Baseline 1 week
DFFNN		0.002	0.492	0.002	0.002	0.375	
RF			0.002	0.002	0.002	0.002	
XGBoost				0.002	0.002	0.002	
SARIMA					1.000	0.002	
Baseline 1 day						0.002	

the p-values obtained for the MAPE in the Wilcoxon test. The p-values have been adjusted using the Holm procedure. Similar results were obtained regarding the MAE, RSME and R^2. Note that since a multiple testing has been applied, the Bonferroni correction is necessary, being a statistically significant difference if α is less than 0.0024 as 21 comparisons of paired samples is made. It can be observed that the DFFNN presents significant differences with all forecasting methods according to the p-values. KNN does not present significant differences with XGBoost and the 1-week based baseline, and SARIMA with the 1-day based baseline either.

Application: Anomaly detection

The predictions obtained with DFFNN can be used for water consumption anomalies detection. The methodology consists of analyzing which values of the time series differ significantly from the prediction made by DFFNN. For this purpose a band is defined through a lower and upper margin of the prediction obtained by the DFFNN. In particular, when the values of the original series fall outside this band, the possible presence of anomalous values or outliers can be predicted.

Figure 16 shows the prediction of the DFFNN model along with a certain upper and lower margin of 15%. It can be seen how this methodology points to the water consumption values occurring in the early morning of March 24, 2020 as possible outliers as shown also in Fig. 12.

Fig. 16 Detection of anomalous water consumption

Conclusions

In this paper the DFFNN deep learning approach based on feed-forward neural networks has been proposed to forecast water consumption in the short-term. A grid search has been carried out in order to tune the multiple hyperparameters involved in the performance of the DFFNN and an evaluation methodology based on growing windows is introduced in order to preserve the temporal order of the time series. Prediction results have been reported using a dataset of water consumption in the city of Murcia in Spain. The proposed DFFNN method has been evaluated according to the MAE, MAPE, RMSE and R^2, yielding an average error close to 3%. The comparison results show that the DFFNN model improves significantly the

forecasting performance compared with the KNN, RF, XGBoost, SARIMA seasonal method and two persistence models. The statistical significance of the DFFNN model developed has been assessed through the Wilcoxon signed-rank test, showing p-values smaller than 0.05 for all the paired combinations.

Future work will be directed towards developing other types of deep neural networks, applying learning transfer from other fields such as electricity consumption as well as making predictions for medium or long-term horizons.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval The present research did not involve any human or animal participants. Hence, there was no issue of ethical standards.

Consent to participate There were no human subjects, and therefore no informed consent was needed.

Consent for publication All authors express their consent for the publication of this article.

References

Ambrosio JK, Brentan BM, Herrera M, Luvizotto E, Ribeiro L, Izquierdo J (2019) Committee machines for hourly water demand forecasting in water supply systems. Math Probl Eng 2019 Article ID 9765468

Anele AO, Hamam Y, Abu-Mahfouz AM, Todini E (2017) Overview, comparative assessment and recommendations of forecasting models for short-term water demand prediction. Water 9(11):887

Antunes A, Andrade-Campos A, Sardinha-Lourenço A, Oliveira MS (2018) Short-term water demand forecasting using machine learning techniques. J Hydroinform 1(6):1343–1366

Arunraj NS, Ahrens D, Fernandes M (2016) Application of sarimax model to forecast daily sales in food retail industry. Int J Oper Res Inf Syst 7(2):1–21

Bata MH, Carriveau R, Ting DS-K (2020) Short-term water demand forecasting using nonlinear autoregressive artificial neural networks. J Water Resour Plan Manag 146(3):04020008

Bergmeir C, Benítez JM (2012) On the use of cross-validation for time series predictor evaluation. Inf Sci 191:192–213

Bolorinos J, Ajami NK, Rajagopal R (2020) Consumption change detection for urban planning: monitoring and segmenting water customers during drought. Water Resour Res 56(3):e2019WR025812

Bougadis J, Adamowski K, Diduch R (2005) Short-term municipal water demand forecasting. Hydrol Process 19(1):137–148

Candelieri A (2017) Clustering and support vector regression for water demand forecasting and anomaly detection. Water 9(3):224

Cembrano G, Wells G, Quevedo J, Pérez R, Argelaguet R (2000) Optimal control of a water distribution network in a supervisory control system. Control Eng Pract 8(10):1177–1188

Chen G, Long T, X J et al (2017) Multiple random forests modelling for urban water consumption forecasting. Water Resour Manag 31:4715–4729

Cochrane C, Ba D, Klerman EB, St. Hilaire MA (2021) An ensemble mixed effects model of sleep loss and performance. J Theor Biol 509:110497

Coelho IM, Coelho VN, da S. Luz EJ, Ochi LS, Guimarães FG, Rios E (2017) A GPU deep learning metaheuristic based model for time series forecasting. Appl Energy 201:412–418

Duerr I, Merrill HR, Wang C, Bai R, Boyer M, Dukes MD, Bliznyuk N (2018) Forecasting urban household water demand with statistical and machine learning methods using large space-time data: A comparative study. Environ Model Softw 102:29–38

Farah E, Abdallah A, Shahrour I (2019) Prediction of water consumption using artificial neural networks modelling (ANN). MATEC Web Conf 295:01004

Gagliardi F, Alvisi S, Franchini M, Guidorzi M (2017) A comparison between pattern-based and neural networks short-term water demand forecasting models. Water Sci Technol 17(5):1426–1435

Galicia A, Torres J, Martínez-Álvarez F, Troncoso A (2018) A novel spark-based multi-step forecasting algorithm for big data time series. Inf Sci 467:800–818

García S, Fernández A, Luengo J, Herrera F (2010) Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. Inf Sci 180(10):2044–2064

Ghiassi M, Fa'al F, Abrishamchi A (2017) Large metropolitan water demand forecasting using DAN2, FTDNN, and KNN models: a case study of the city of Tehran, Iran. Urban Water J 14(6):655–659

Herrera M, Torgo L, Izquierdo J, Pérez-Garcia R (2010) Predictive models for forecasting hourly urban water de-mand. J Hydrol 387:141–150

Kang HS, Kim H, Lee J, Lee I, Kwak BY, Im H (2014) Optimization of pumping schedule based on water demand forecasting using a combined model of autoregressive integrated moving average and exponential smoothing. Water Supply 15(1):188–195

Lee D, Derrible S (2020) Predicting residential water demand with machine-based statistical learning. J Water Resour Plan Manag 146(1):04019067

Leyli-Abadi M, Samé A, Oukhellou L, Cheifetz N, Mandel P, Féliers C, Chesneau O (2018) Mixture of non-homogeneous hidden markov models for clustering and prediction of water consumption time series. In: Proceedings of the 2018 IEEE international joint conference on neural networks (IJCNN), pp 1–8

Lin Y, Koprinska I, Rana M, Troncoso A (2019) Pattern sequence neural network for solar power forecasting. In: Neural information processing, pp 727–737

Livera AMD, Hyndman RJ, Snyder RD (2011) Forecasting time series with complex seasonal patterns using exponential smoothing. J Am Stat Assoc 106:1513–1527

Makridakis S, Spiliotis E, Assimakopoulos V (2018) Statistical and machine learning forecasting methods: concerns and ways forward. PLOS ONE 13(3):e0194889

Mouatadid S, Adamowski J (2017) Using extreme learning machines for short-term urban water demand forecasting. Urban Water J 14(6):630–638

Nunes-Carvalho TM, Souza-Filho FA, Costa-Porto V (2021) Urban water demand modeling using machine learning techniques: case study of Fortaleza, Brazil. J Water Resour Plan Manag 147(1):05020026

Pacchin E, Gagliardi F, Alvisi S, Franchini M (2019) A comparison of short-term water demand forecasting models. Water Resour Manag 33:1481–1497

Padulano R, Giudice GD (2018) A mixed strategy based on self-organizing map for water demand pattern profiling of large-size smart water grid data. Water Resour Manag 32:3671–3685

Peña-Guzmán C, Melgarejo J, Prats D (2016) Forecasting water demand in residential, commercial, and industrial zones in Bogotá, Colombia, using least-squares support vector machines

Rahim MS, Nguyen KA, Stewart RA, Giurco D, Blumenstein M (2019) Predicting household water consumption events: towards a personalised recommender system to encourage water-conscious behaviour. In: Proceedings of the 2019 IEEE international joint conference on neural networks (IJCNN), pp 1–8

Rana M, Koprinska I, Troncoso A (2014) Forecasting hourly electricity load profile using neural networks. In: Proceedings of the 2014 IEEE international joint conference on neural networks (IJCNN), pp 824–831

Ren Z, Li S (2016) Short-term demand forecasting for distributed water supply networks: a multi-scale approach. In: Proceedings of the 2016 12th World congress on intelligent control and automation (WCICA), pp 1860–1865

Salomons E, Goryashko A, Shamir U, Rao Z, Alvisi S (2017) Optimizing the operation of the Haifa-A water-distribution network. J Hydroinform 9(1):51–64

Smolak K, Kasieczka B, Fialkiewicz W, Rohm W, Siła-Nowicka K, Kopańczyk K (2020) Applying human mobility and water consumption data for short-term water demand forecasting using classical and machine learning models. Urban Water J 17(1):32–42

Talavera-Llames RL, Pérez-Chacón R, Martínez-Ballesteros M, Troncoso A, Martínez-Álvarez F (2016) A nearest neighbours-based algorithm for big time series data forecasting. In: Proceedings of the 2016 hybrid artificial intelligent systems (HAIS), vol 9648, pp 174–185

Talavera-Llames R, Pérez-Chacón R, Troncoso A, Martínez-Álvarez F (2019) MV-kWNN: a novel multivariate and multi-output weighted nearest neighbours algorithm for big data time series forecasting. Neurocomputing 353:56–73

Tian T, Xue H (2017) Prediction of annual water consumption in Guangdong province based on Bayesian neural network. IOP Conf Ser Earth Environ Sci 69:012032

Tiwari M, Jan A, Kazimierz A (2016) Water demand forecasting using extreme learning machines. J Water Land Dev 28(1):37–52

Torres JF, Troncoso A, Koprinska I, Wang Z, Martínez-Álvarez F (2018) Deep learning for big data time series forecasting applied to solar power. In: Proceedings of the 13th international conference on soft computing models in industrial and environmental applications (SOCO), pp 123–133

Torres JF, Troncoso A, Koprinska I, Wang Z, Martínez-Álvarez F (2019) Big data solar power forecasting based on deep learning and multiple data sources. Expert Syst 36:e12394

Torres JF, Hadjout D, Sebaa A, Martínez-Álvarez F, Troncoso A (2021) Deep learning for time series forecasting: a survey. Big Data 9(1):3–21

Troncoso A, Riquelme-Santos JM, Riquelme JC, Gómez-Expósito A, Martínez-Ramos JL (2004) Time-series prediction: application to the short-term electric energy demand. Lect Notes Comput Sci 3040:577–586

Trull O, García-Díaz JC, Troncoso A (2019) Application of discrete-interval moving seasonalities to spanish electricity demand forecasting during easter. Energies 12(6):1083

Trull O, García-Díaz JC, Troncoso A (2020) Initialization methods for multiple seasonal holt-winters forecasting models. Mathematics 8(2):268

Varma S, Simon R (2006) Bias in error estimation when using cross-validation for model selection. BMC Bioinform 7:91

Villarin MC, Rodriguez-Galiano VF (2019) Machine learning for modeling water demand. J Water Resour Plan Manag 145(5):04019017

Xenochristou M, Kapelan Z (2020) An ensemble stacked model with bias correction for improved water demand forecasting. Urban Water J 17(3):212–223

Xenochristou M, Hutton C, Hofman J, Kapelan Z (2020) Water demand forecasting accuracy and influencing factors at different spatial scales using a gradient boosting machine. Water Resour Res 56(8):e2019WR026304

Xu Y, Zhang J, Long Z, Tang H, Zhang X (2019) Hourly urban water demand forecasting using the continuous deep belief echo state network. Water 11(2):351

Yasdi R (1999) Prediction of road traffic using a neural network approach. Neural Comput Appl 8(2):135–142