Expert Systems With Applications 247 (2024) 123234

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

# Medium-term water consumption forecasting based on deep neural networks

A. Gil-Gamboa [a], P. Paneque [b], O. Trull [c], A. Troncoso [a,*]

[a] Data Science and Big Data Lab, Universidad Pablo de Olavide, ES-41013, Seville, Spain
[b] Department of Geography, History and Philosophy, Universidad Pablo de Olavide, ES-41013, Seville, Spain
[c] Department of Applied Statistics and Operational Research and Quality, Universitat Politècnica de València, ES-46022, Valencia, Spain

ARTICLE INFO

*Keywords:*
Time series forecasting
Deep learning
Water consumption

ABSTRACT

Water consumption forecasting is an essential tool for water management, as it allows for efficient planning and allocation of water resources, an undervalued but indispensable resource for all living beings. With the increasing demand for accurate and timely water forecasting, traditional forecasting methods are proving to be insufficient. Deep learning techniques, which have shown remarkable performance in a wide range of applications, offer a promising approach to address the challenges of water consumption forecasting. In this work, the use of deep learning models for medium-term water consumption forecasting of residential areas is explored. A deep feed-forward neural network is developed to predict water consumption of a company's customers for the next quarter. First, customers are grouped according to their consumption as these customers include both household consumers and special consumers such as public swimming pools, sports halls or small industries. Then, a deep feed-forward neural network is designed for household customers by obtaining the optimal values for those hyperparameters that have a great influence on the network performance. Results are reported using a real-world dataset composed of the water consumption from 1999 to 2015 on a quarterly basis, corresponding to 3262 clients of a water supply company. Finally, the proposed algorithm is evaluated by comparing it with other reference algorithms including an LSTM network.

## 1. Introduction

Water is a precious resource that plays a critical role in the livability and sustainability of cities. As urban areas continue to grow and face the challenges of climate change and population demands, the need for effective water management practices becomes increasingly vital. Cities around the world recognize the importance of sustainable water management, adopting innovative approaches to conserve water, protect water quality and mitigate the impact of urbanization on water resources. Drought used to be an anomaly that occurred from time to time in our country and very rarely in Europe. However, in recent years it is a phenomenon that has been recurring with assiduity. Thus, efficient water management becomes an imminent necessity to cope with periods of drought (Limones, Vargas Molina, & Paneque, 2022). The development of data monitoring technologies together with the advance of artificial intelligence techniques are crucial for good water resource management. Thus, companies dedicated to water management in cities are applying artificial intelligence techniques to the data collected to know consumption trends, avoid water losses in the distribution system or predict the water consumption in advance.

Time series forecasting is a field of research that has been intensively investigated in recent years due to the large amount of temporal data that is nowadays generated from sensors in diverse domains (Jin, Zeng, Yan, & Ji, 2021; Trull, García-Díaz, & Troncoso, 2020; Yan, Li, Ji, Qi, & Du, 2019). For instance, in Jin et al. (2021), for air quality forecasting, a novel deep learning framework enriched with federated learning outperforms existing models by overcoming intercorrelations and volatile patterns. In Yan et al. (2019), a hybrid model for household energy consumption forecasting excels, mitigating challenges posed by irregular human behaviors and univariate datasets. Additionally, new methods for initializing Holt–Winters models are proposed in Trull et al. (2020), showcasing efficacy in multiple seasonal scenarios.

These advances in smart technology are currently being introduced in the field of water supply such as smart meters that are being used to carry out efficient water management and make data-driven decisions (Faiz & Daniel, 2022). On the other hand, one of the most widely used machine learning techniques for time series forecasting is deep learning, that focuses on learning of complex temporal patterns from the data. Deep learning techniques have been used in a wide

* Corresponding author.
*E-mail addresses:* agilgam1@upo.es (A. Gil-Gamboa), ppansal@upo.es (P. Paneque), otrull@eio.upv.es (O. Trull), atrolor@upo.es (A. Troncoso).

https://doi.org/10.1016/j.eswa.2024.123234
Received 25 September 2023; Received in revised form 29 December 2023; Accepted 14 January 2024
Available online 24 January 2024

variety of applications ranging from finance and economics, to energy or agriculture, with remarkable results. Recently deep learning is being applied to planning and management problems of urban water systems. However, deep learning is still at an early stage of development as most work uses predesigned neural network models, synthetic data or pilot systems to evaluate the performance of the methods. A wide review of the existing deep learning techniques for urban water management can be found in Fu, Jin, Sun, Yuan, and Butler (2022). Currently, companies that manage drinking water supply are faced with a huge volume of data that is impossible to exploit because it is too dirty due to failures in metering equipment, customer fraud, etc. In addition, the data corresponds to customers with very different behaviors. These companies 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 optimize the operation in the medium term is a valuable tool for the management of water reserves and the use of associated equipment. Thus, machine learning methods have the challenge of cleaning and clustering these water consumption data so that deep learning models can obtain water consumption predictions with the best possible accuracy. Therefore, this paper presents a complete methodology in which the first step is the preprocessing, the second step is the recognition of similar patterns using clustering techniques and the final step is the development of a forecasting model based on deep learning.

This paper proposes a prediction methodology based on deep learning for the prediction of the water consumption of a company's customers in the medium term, specifically for the next quarter. The methodology presented is divided into several steps. First, an exploratory analysis of the consumption of all customers has been carried out in order to preprocess the data. Second, a clustering technique has been proposed to obtain a pattern recognition revealing different consumption profiles. Then, focusing on domestic consumers, a deep feed-forward network is developed to obtain a consumption prediction for each customer individually, that is, at the household level. Results using a real-world dataset composed of quarterly water consumption of a company's customers have been reported and compared with isotonic regression, polynomial regression, random forest, gradient boosted tree and a deep long short-term memory neural network.

The remainder of the paper is structured as follows. Section 2 summarizes contributions related to water demand forecasting found in the literature. Section 3 describes the forecasting problem to be solved and an analysis of the water consumption data. Section 4 introduces the proposed methodology and Section 5 presents the results that have been obtained. Finally, Section 6 concludes the paper and points out future work lines.

## 2. Related work

In this section, we review some recently published approaches on water consumption forecasting. Theses approaches have been divided into deep learning, statistical and machine learning methods.

Statistical methods offer simplicity and ease of interpretation, making them accessible for communication of results. They can perform well with smaller datasets and are robust to outliers, providing stability in their predictions. However, their reliance on simplified assumptions limits their ability to capture complex patterns, especially in datasets with nonlinear relationships. Additionally, statistical models may exhibit low predictive capacity compared to more advanced methods like deep learning, and their results can be sensitive to biases when underlying assumptions are not met. Secondly, machine learning methods excel in adaptability to changing data patterns and handling complexity. They can automate prediction processes, saving time and resources. They also are scalable to large datasets. However, some models, particularly complex ones, may be challenging to interpret, posing issues in environments where transparency is crucial. The need for large training datasets, sensitivity to biased data, and the requirements for hyperparameter tuning are notable disadvantages, highlighting challenges in areas with limited data. Finally, deep learning methods provide automatic feature learning and superior performance on complex data, making them effective in capturing patterns. Their scalability and ability to transfer knowledge through pretraining on related tasks are advantageous. However, their success relies on large datasets and significant computational power. The risk of overfitting, especially with small datasets, and sensitivity to noisy data are problems that impact performance in specific scenarios.

### 2.1. Deep learning methods

Deep learning methods have gained significant attention in recent years due to their ability to capture complex patterns in data. For instance, in Bandara, Bergmeir, and Smyl (2020) the main goal was to generate a global prediction model from different time series. As accuracy may degenerate if the time series database is heterogeneous, subgroups of similar time series were previously obtained by time series clustering techniques and a deep learning model was proposed for these subgroups.

In recent years, some works based on deep learning for the prediction of water consumption can be found in the literature. One of the most popular deep learning models used for water consumption forecasting is the Long Short-Term Memory (LSTM) model. In Kühnert, Gonuguntla, Krieg, Nowak, and Thomas (2021) the authors presented a model LSTM for water demand prediction in order to obtain pump schedules for the next 24 h. In this work the authors showed that LSTMs outperform other methods as they can easily integrate additional information such as day of the week or national holidays. In addition, they investigated the capabilities of LSTMs in terms of online learning and transfer and they concluded that LSTMs only need a few days of training data to obtain good results. Two prediction models based on LSTM were proposed in Niknam, Zare, Hosseininasab, and Mostafaeipour (2023) to forecast monthly water consumption in Yazd (Iran), namely, a univariate model and a multivariate model including temperature and humidity. The results showed that the multivariate LSTM model outperforms the univariate LSTM model and other previously developed models in terms of accuracy, as it takes into account the effects of climatic factor. Thus, deep learning models can effectively handle multiple input sequences with different time lengths, allowing for multivariate selection of inputs. In Mu, Zheng, Tao, Zhang, and Kapelan (2020) an LSTM-based model was presented in order to predict short-term urban water demands in Hefei, China. The performance of the LSTM model was computed using data with varying time resolutions. Results indicated that the LSTM model outperforms other methods in accuracy, specially with high-resolution and dynamically changing data. Addressing the complexity and errors at extreme points by introducing virtual data, the authors in Salloom, Kaynak, and He (2021) applied the Gated Recurrent Unit (GRU) and k-means for new feature creation, achieving lower complexity and reduced training times. Results using real data from water treatment plants in China showed a significant reduction in model complexity without sacrificing accuracy, although with an increase in training time.

Some research works published in the last years revealed that the use of other deep learning models such as deep feed-forward neural networks or convolutional neural networks outperforms traditional machine learning models for predicting short-term water demand (Hao, Cominola, & Castelletti, 2022; Kavya, Mathew, Shekar, & P., 2023; Zanfei, Brentan, Menapace, & Righetti, 2022). It should also be noted that the accuracy of prediction models varies depending on the region and the data used. In Antunes, Andrade-Campos, Sardinha-Lourenço, and Oliveira (2018), the authors emphasize the need to move from managing water based on instantaneous demand to a future demand-based approach. They evaluated machine learning methods such as deep feed-forward neural networks of two, five and eight layers, random forest (RF), support vector machine (SVM), and K nearest neighbors to predict short-term water demand from two water utility in Portugal. In addition, factors like climate, seasonality, data volume, and forecast windows were analyzed to achieve accurate results. Three different convolutional networks with attention mechanisms and a data-cleaning algorithm to deal with anomalies and outlier were proposed to extract spatio-temporal dynamic features and to forecast water demand for 1, 3, 6, 12, and 24 h periods in Cao, Yuan, Tian, Xu, and Su (2023). Results were compared with support vector regression and other deep learning models such as LSTM. The graphical convolutional model was the best model and significantly improved prediction accuracy and adaptive spatial–temporal feature extraction.

On the other hand, the development of hybrid methods that combine both different types of deep learning models and deep learning models with other artificial intelligence techniques is one of the current trends in research work. A study of water footprint modeling and forecasting using various artificial intelligence algorithms integrated with a Seasonal Auto-Regressive Integrated Moving Average (SARIMA) model was presented in Tao, Zhang, Xie, and Liang (2023). The forecasting models were developed using the historical data and artificial intelligence algorithms selected were a deep feed-forward neural network, SVM, and RF. The modeling results indicated that the SVM algorithm outperformed the other algorithms for water footprint modeling, while the RF algorithm performed the best for water footprint forecasting. The incorporation of techniques such as wavelet decomposition and principal component analysis improves the predictive performance of deep learning models, especially in capturing complex patterns and peaks in water demand time series. For instance, in Du, Zhou, Guo, Guo, and Wang (2021) a hybrid model which combines a discrete Wavelet transform, an analysis of principal components and a LSTM was proposed. This model was compared with other benchmark models, and it showed that the hybrid model outperformed the others in terms of capturing complex patterns and average prediction accuracy. An innovative approach for short-term water demand prediction is proposed in Shan, Ni, Chen, Lin, and Li (2023). In particular, the maximal information coefficient is used for feature selection, incorporating an Attention-BiLSTM network and the XGBoost algorithm to enhance accuracy. After hyperparameter tuning, the method showed superiority over other models in terms of precision and stability. In Chen et al. (2022) a new framework for short-term water demand forecasting is proposed. In particular, a hybrid model based on convolutional and recurrent neural networks is developed. First, a convolution of one dimension is applied to obtain the important features in an automatic way, and second, a recurrent neural network, namely Gated Recurrent Unit is used to obtain predictions. Different settings for hyperparameters and training strategies were carried out using historical data. The model showed better forecast accuracy and ability for feature extraction when compared to other deep learning models in the literature. A hybrid model combining a convolutional neural network and a bidirectional LSTM neural network (Bi-LSTM) or standard LSTM was adopted in Hu, Tong, Wang, Yang, and Oliveira Turci (2019), Zhou, Guo, Du, Huang, and Guo (2022). The convolutional network is used to extract the features from historical water time series data and weather data, and these features are the input set for the Bi-LSTM. The proposed hybrid approach obtained better accuracy than the corresponding single models. In Liu, Zhou, and Zhang (2023), an ensemble deep learning model was proposed for daily water demand forecasting from two water plants in Yiwu, China. Combining the Seasonal and Trend decomposition using the Loess (STL) method with the AdaBoost algorithm and LSTM model, this approach enhances forecast accuracy, stability, and conciseness. The current water demand forecasting literature focuses on precise point predictions, but increased data uncertainty poses reliability challenges. To address this, a model combining LSTM to kernel density estimation (KDE) optimized through particle swarm optimization (PSO), is proposed for predicting water demand intervals in Du et al. (2022). The model's innovative strategies included error level splitting, PSO-optimized KDE fitting, and confidence-window shifting. Experimental results showed the proposal outperforming other models, offering reliable decision support for urban water supply management.

### 2.2. Non-deep learning methods

For many years the methods that have been used for water consumption prediction have been non-deep leaning methods such as classical statistical methods based on time series models (Troncoso, Riquelme-Santos, Riquelme, Gómez-Expósito, & Martínez-Ramos, 2002) or machine learning methods (Melgar-García, Gutiérrez-Avilés, Rubio-Escudero, & Troncoso, 2023).

#### 2.2.1. Statistical methods

A recently published work in Mohammad and Mousavi-Mirkalaei (2019) explored the use of non-deep learning methods, such as an Auto-Regressive Integrated Moving Average (ARIMA) model and a Nonlinear Auto-Regressive Exogenous (NARX) model. These models were extended by using climatic variables to predict urban water consumption data. The results showed that the ARIMA model that includes sunshine hours outperformed the traditional ARIMA that only uses water consumption data for forecasting. Similarly, the NARX model that takes into account the population feature improved performance of the baseline NARX model. If we focus on NARX models and geographically warm locations, a NARX model for monthly water demand forecasting in Kuwait can be found in Alsumaiei (2020). In this case, a linear trend is performed on the water consumption data to remove the influence of population growth. The results showed that the model is efficient and robust in predicting short-term water demand, especially during the summer season. Overall, the NARX model establishes itself as a promising tool to assist in the development of resilient water supply plans in areas with limited water resources and arid climates.

#### 2.2.2. Machine learning methods

In recent years there has been a growing interest in combining different no-deep learning techniques as well as ensembles techniques that highlight most of the strengths of each technique to obtain water consumption forecasts. For instance, tree-based regression (RT) models and Self-Organizing Map (SOM) models have been commonly used for short-term demand forecasting and for clustering large amounts of unlabeled data, respectively. In Bata, Carriveau, and Ting (2020), a RT model was combined with a SOM model and compared with a stand-alone RT model and a SARIMA model for short-term water demand forecasting. The results showed that the inclusion of the SOM model as input to the RT model significantly improved its performance, doubling on average the accuracy of water demand forecasting. Furthermore, the hybrid model outperformed both the stand-alone RT model and the SARIMA model in terms of forecast accuracy at different periodicity. In Pandey, Bokde, Dongre, and Gupta (2021), two hybrid approaches for water demand forecasting were presented. These approaches combined the three following methods: the ensemble empirical mode decomposition (EEMD), which is an adaptive signal processing method used to decompose nonlinear time series, a method based on differences between sequences of patterns (DPSF) and ARIMA models. The EEMD-DPSF approach is found to outperform other methods in terms of prediction accuracy. The comparison between the two proposed models shows that the EEMD-DPSF approach provides better results, while the EEMD-DPSF-ARIMA approach requires less computational time. Furthermore, the results were compared with those obtained using Pattern Sequence Forecasting (Pérez-Chacón, Asencio-Cortés, Martínez-Álvarez, & Troncoso, 2020), ARIMA, DPSF, and several other non-deep neural networks models. The authors in Lu, Matthews, and Han (2020) used a hybrid model based on various evolutionary algorithms and SVM to improve the accuracy of monthly water demand prediction. The prediction results were compared with another hybrid model combining a different evolutionary algorithm with SVM and a backpropagation neural network. The results showed that the mean absolute percentage error of proposed hybrid method was the smallest. Pesantez, Berglund, and Kaza (2020) focuses on the utilization of machine learning methods such as RF, non-deep neural networks, and SVM. Lagged demand, seasonality, climate, and household characteristics are considered as features for making predictions. Temporal clustering is also applied to enhance the model's performance. Results show that RF and non-deep neural networks perform better than SVM in all scenarios. In Karamaziotis, Raptis, Nikolopoulos, Litsiou, and Assimakopoulos (2020) seven forecasting models and ensemble techniques were proposed to predict mid-term water consumption. The ensemble techniques combined the prediction methods according to a single average, weighted average, trimmed average by removing a percentage of best and worst predictions or a median. The baseline models to be combined were models based on time series such ARIMA and exponential smoothing and a neural network of a single-hidden layer. The ensemble using a trimmed average and the ARIMA were the methods that achieved the most accurate predictions. The forecast results enabled water utilities to detect potential spikes in demand in specific areas and take preventive actions to mitigate stress on the distribution system. In Rezaali, Quilty, and Karimi (2021) the authors explored the combination of wavelets decomposition with some machine learning algorithms and their probabilistic versions. In addition, different ways of dataset partitioning were proposed to reduce overfitting. The results showed that the probabilistic version of RF was the one that obtained the most accurate predictions. The study made in Guo, Liu, Dai, and Xu (2020) introduced three forecasting models and proposed an enhanced Whale Optimization Algorithm (WOA) with social learning and wavelet mutation. The algorithm achieved outstanding performance. The hybrid model, incorporating linear, exponential, and logarithmic functions with WOA, produced the best results. This research (Smolak et al., 2020), compared classical and adapted machine learning algorithms, incorporating human mobility data. The top-performing algorithm was RF, outperforming classic ARIMA. Human mobility data significantly improved prediction accuracy, showcasing potential applications in smart water supply management. The study recommends incorporating human mobility data for localized pressure adjustments based on predicted user relocations. This recommendation is extrapolated to other investigations, as Koo et al. (2021), where smart water grid technology is evaluated for efficient water management in cities, using real-time data from sensors on YeongJong Island. Short-term water demand forecasting with ARIMA, non-deep neural networks, a predictor based on moving average and LSTM models showed limitations, highlighting the importance of considering factors like usage time and customer behavior for accurate system management. Finally, in Niknam, Zare, Hosseininasab, Mostafaeipour, and Herrera (2022) a review of 100 articles from 2010 to 2021 on urban water supply challenges and water demand forecasting can be found. Traditional time series and artificial neural network methods prevail. The study concludes that forecasting method superiority remains unclear, emphasizing considerations like temporal scope and data availability. Notably, interpretable methods succeed, while artificial neural networks and regression dominate, signaling potential for increased hybrid model adoption. The scarcity of deep learning applications suggests a research avenue. Additionally, it is noted that LSTM and hybrid algorithms are popular for water demand forecasting, offering improved temporal dependency capture, primarily for short-term forecasts of less than one day.

## 3. Definition of the problem

In this section a definition of the forecasting problem and the dataset to be used are presented. We firsly introduce the problem to be solved in Section 3.1. Second, a description of the dataset is made along with a brief exploration of the data in Section 3.2.

### 3.1. Definition of the problem

The problem consists of obtaining the water consumption predictions of the customers of a company in a particular area of the city of Seville in Spain. Specifically, the aim is to predict the consumption of the next quarter of each customer from the historical quarterly consumption data of the previous years.

Let $N$ be the number of customers and $C_i(t)$ the water consumption of the $i$th customer at time t. The problem can be formulated as obtaining the prediction model $f$ such that:

$$[\hat{C}_i(t+1), \dots, \hat{C}_i(t+h)] = f(C_i(t), \dots, C_i(t-(W-1))) \tag{1}$$

where $\hat{C}_i(t)$ is the forecast of the consumption for the $i$th customer at time $t$, $h$ is the prediction horizon and $W$ is a window composed of past values. In this work, the objective is to predict the next quarter consumption from the three previous years, that is, $h = 1$ and $W = 12$.

This problem is an one-step time series forecasting problem. Thus, this work falls under the category of time-series models. Essentially, a time series $y(t)$ represents a collection of measurements for a variable, $y$, recorded as time progresses. Typically, it is assumed that these data points are acquired at spaced time intervals denoted by the time index $t$, and the resulting samples $y(1)$, $y(2)$, …, $y(n)$ form a discrete-time series. Time series methods are designed to handle the inherent time-dependency in the data, while mitigating any potential issues arising from inappropriate time labeling (Pérez-Chacón et al., 2020). Deep learning, specifically, is well-suited for analyzing time-indexed data, making it an effective approach for time series forecasting. Further details on deep learning techniques for time series forecasting can be found in Lara-Benítez, Carranza-García, and Riquelme (2020).

### 3.2. Data

In this work, the data are collected from a public water supply company corresponding to water consumption of a residential area of the city of Seville in Spain. The dataset consists of quarterly water consumption of $N$ customers of the utility from 1999 to 2015. That is, the dataset $D$ contains $N$ time series, each corresponding to the consumption of one of the company's customers.

$$D = [C_1(t), \dots, C_N(t)] \tag{2}$$

This dataset contains, in addition to the water consumption from 1999 to 2015, other values such as address, street, number of block, etc. for each customer. Also, we can find values, such as the policy, the municipality code or the ID code as an identifying value of the water consumption for each customer.

Fig. 1 shows the mean volume of water supplied per capita to the public network in the region of Andalusia from 2002 to 2016. The region of Andalusia is made up of eight provinces, one of which is Seville. These data have been obtained from the National Statistics Institute of Spain. It can be noted that consumption decreases over the years. This could be due to changes in water management policy, laws may have been implemented to promote more efficient water use or to reduce water demand in certain sectors, such as agriculture or industry. Another factor may be climate change, climatic conditions may influence the amount of water available in the region and therefore water consumption, Andalusia is particularly dry, especially Seville. Finally, it could be changes in customer behavior in order to save water in their daily lives.

A brief exploratory analysis of the actual data used in this work is made below. Fig. 2 presents the range of the water consumption values corresponding to the customers of the company. The histogram is notable for being decreasing, so that moderate values tend to be more common than values further away. As can be seen in the last vertical bar, an overflow has been performed from the value 350 on wards, offering slightly more than 3000 values from this overflow, which are considered very high consumption.

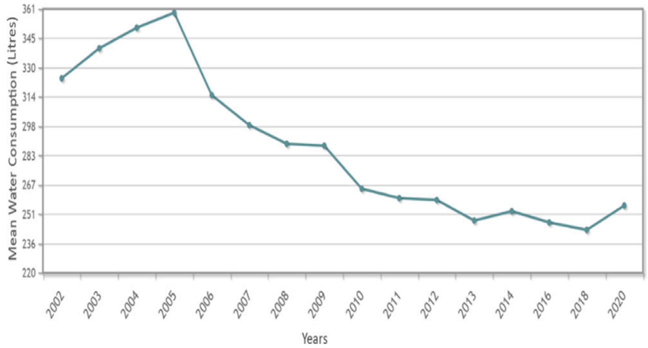


**Fig. 1.** Consumption of water in Andalusia from 2002 to 2016.

Next, we will focus on the water consumption values found in the first three intervals from 0 to 61 litres, which are where the majority of customers' consumption are found. Fig. 3 presents the distribution of these values. It is observed exactly the same behavior as in the histogram that contained all the values but in this case the decrease in the number of water consumption values as the range of the interval increases is much smoother.

To highlight the variability water consumption data contain, it will be necessary to rely on Fig. 4, which shows the quarterly average and standard deviation of water consumption when the consumption of all customers has been aggregated. In particular, the average water consumption is the central curve indicating the average amount of water consumed. It represents the break-even point between the highest and lowest values of water consumption in the data set. The standard deviation is a measure of the dispersion or variability of the data. By subtracting the standard deviation from the mean water consumption, we obtain a value that represents an estimate of the lower limit within which most of the data lie. By adding the standard deviation to the average water consumption, we obtain a value that represents an estimate of the upper limit within which most of the data lie.

By displaying these three curves together on a graph, a more comprehensive view of how water consumption data is distributed can be obtained in addition to gain a better understanding of its statistical behavior. In this case, the standard deviations are quite broad, indicating high variability in the data. Additionally, this representation helps to identify outlier values for each quarter. For instance, based on the graph, four noteworthy outliers could be considered which correspond to the highest values of the standard deviation. These outliers all correspond to the third quarter of the years 2000, 2001, 2010, and 2011.

In 2000, two records exceeding 13000 litres were recorded for the third quarter. Both records correspond to a covered pavilion and municipal swimming pool. In addition, there are more than five records with values above 5000 litres corresponding to sports centers, parks and schools. In 2001, the standard deviation is even wider, due to the existence of an indoor pavilion with a value of almost 25000 litres of water consumption. In addition, two infrastructures such as a municipal swimming pool and a soccer field exceed 10000 litres. In 2010, we again find a value close to 23000 litres corresponding to a school and two values of 9000 and 16000 litres corresponding to a covered pavilion and a sports center, respectively. Finally, in 2011, where the peak in the graph stands out the most, it is mainly due to three values close to tens of thousands. The first of them, over 30000 litres, corresponds to a high school, the next one is close to 20000 litres and corresponds to an indoor pavilion and the last one corresponds to a municipal swimming pool.

Once the outliers values of these quarters have been analyzed, it is clear that they correspond to large infrastructures and precisely all of them correspond to the third quarter of the year. These are the hottest months and the months with the highest water consumption, whether for recreational or health purposes. In spite of this, they are still really abnormal and high values.

The consumption of two customers randomly selected from the whole dataset is included in Fig. 5. The graph shows a somewhat high consumption of the customer 1 for the first two years and a rather irregular consumption in all quarters, but following a pattern throughout the years. Also noteworthy is the decrease on both sides in water consumption over the years and the regularity of consumption for the customer 2.

## 4. Methodology

In this Section we describe the proposed methodology in order to obtain an estimate of water consumption for all customers of a particular company. In particular, Section 4.1 presents a general description of the methodology. Section 4.2 presents the preprocessing performed on the data. A clustering process to obtain patterns in data is made in Section 4.3, and finally, the prediction model based on deep neural networks is described in Section 4.4.

### *4.1. Description of the methodology*

In order to provide an overview of the proposed methodology, all the steps that have been carried out are described in this Section.

### *4.2. Data pre-processing*

Most data analysis tasks require data pre-processing as a first step. In a large percentage of real-world applications, the data often contain incomplete or missing values, among other problems that must be solved. Therefore, if we want to obtain a good quality prediction whatever the data context, we must have the most uniform and cleanest data possible.

For this case, and after analyzing the data, it is concluded that the data needs a high level of pre-processing. For this reason, it is decided to make the modifications to the dataset as described below.

It is noted that there are duplicated instances. That is, the quarterly water consumption data of customers are sometimes separated in different records. To solve this, instances with the same ID code are joined into only one.

The next step is eliminate instances or customers that contain a big number of missing values between their quarterly water consumption data. Specifically, the historical water consumption data range from the third quarter of 1999 to the last quarter of 2015. Thus, if we calculate the number of the total quarters, we obtain 66 quarters for each customer. Those customers who have more than four quarters (or in other words more than one year) without data have been removed. In this way, customers with recent water supply contracts are discarded.

For customers with four or fewer quarters without data, these missing values have been modified by using the average of the water consumption of that customer.

It is also observed in the dataset that there are customers with some extremely small water consumption close to zero, which may be due to consumption measurement errors. Consequently, a further step is added for the values that are very small compared to the rest values of the same customer. These values fulfill the following condition:

$$value < (mean_i - -\sigma_i) \tag{3}$$

where $mean_i$ and $\sigma_i$ are the mean and the standard deviation of the water consumption for the $i$th customer. It can be noticed that mean and standard deviation is different for each customer. For each customer, these values are detected using the Eq. (3) and they are replaced by the zero value. Next, those customers with more than four zeros in their quarterly water consumption values are eliminated. Basically, with

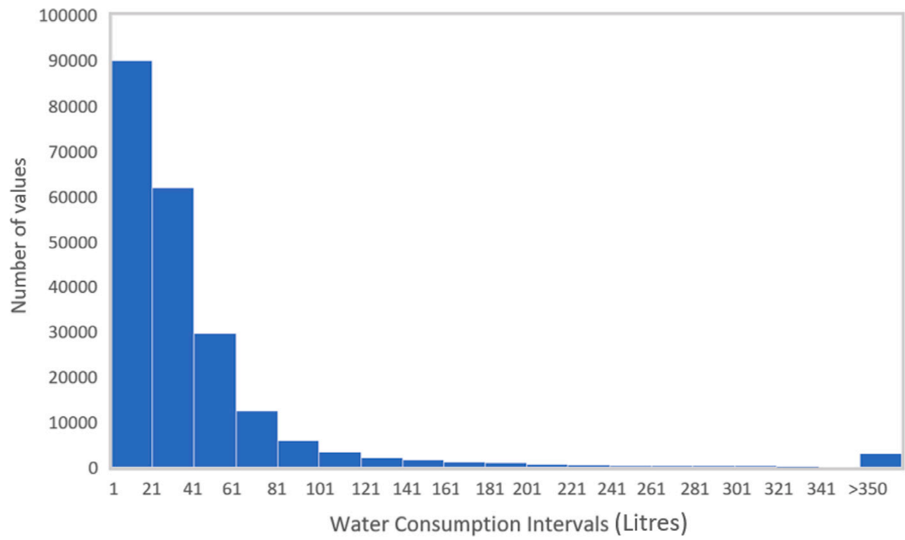


**Fig. 2.** Distribution of water consumption values.

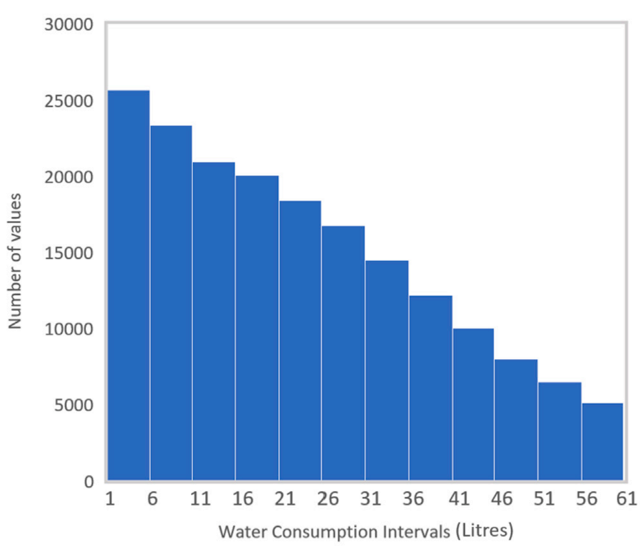


**Fig. 3.** Distribution of water consumption values less than 61 litres.

these steps, we eliminate customers with anomalous water consumption for many quarters.

Finally, for customers with four or fewer quarters with zero values, these values are replaced by the average of the water consumption of that customer, in order to provide us more useful and practical information.

Table 1 represents how the number of instances or customers has been decreasing according to the above mentioned steps. Step 1 consists of joining duplicated customers, step 2 means eliminating customers with a very short historical data and step 3 corresponds with the elimination of customers with many quarterly water consumptions zero or close to zero.

After these steps, the dataset consists of quarterly water consumptions from the third quarter of 1999 to the last quarter of 2015 (66 values) for a total of 3262 customers.

### 4.3. Pattern recognition

During the stage of the data exploratory analysis, a very wide difference in terms of water consumption is observed between different customers. Therefore, it is decided to analyze the customer ID with the type of infrastructure it is, and it is observed that the data range from residential flats to municipal swimming pool or irrigation services. It is obvious that the ranges of water consumption in the above examples will be very different, which would negatively affect the accuracy of the prediction models.

**Table 1**
Pre-processing values.

| | Start | Step 1 | Step 2 | Step 3 |
|---|---|---|---|---|
| Clients | 162 174 | 128 838 | 57 574 | 3262 |
| Percentage | 100% | 79.44% | 35.44% | 2.01% |

**Table 2**
Clustering values.

| | Cluster 0 | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|---|
| Average | 37.96 | 267.45 | 1351.58 | 3335.14 |
| Clients | 3159 | 83 | 15 | 5 |
| Percentage | 96.84% | 2.54% | 0.46% | 0.16% |

For this reason, it is decided to apply a clustering to obtain a characterization of customers in order to develop a predictive model with the highest possible accuracy. Thus, the k-means clustering algorithm is applied together the elbow method to determine the optimal number of clusters. This method calculates the average cluster distortion, i.e., the average distance from the centroid to all points forming the cluster. The aim of this method is to find the appropriate number of clusters according to the data provided. This value is shown graphically with the “elbow” that is generated in the graph when the dispersion value starts to decrease slowly. Furthermore, it means that it is the right value of clusters. Otherwise, if the right value of clusters has not been found, the dispersion value decreases sharply, showing in the graph an almost vertical drop, which is very noticeable.

After launching several executions of the elbow method and observing the graph and the values shown by this method, four clusters are considered to be the best option. Table 2 shows the average consumption value, number of clients and percentage of clients obtained for each cluster. It can be seen that cluster 0 contains 3159 instances out of 3262 (96.84%) that make up the instances extracted after the pre-processing. Specifically, cluster 0 contains most of the customers and is characterized by residential water consumption. The remaining clusters are dominated by infrastructures with high water consumption, such as schools, swimming pools, sports centers, football pitches, etc.

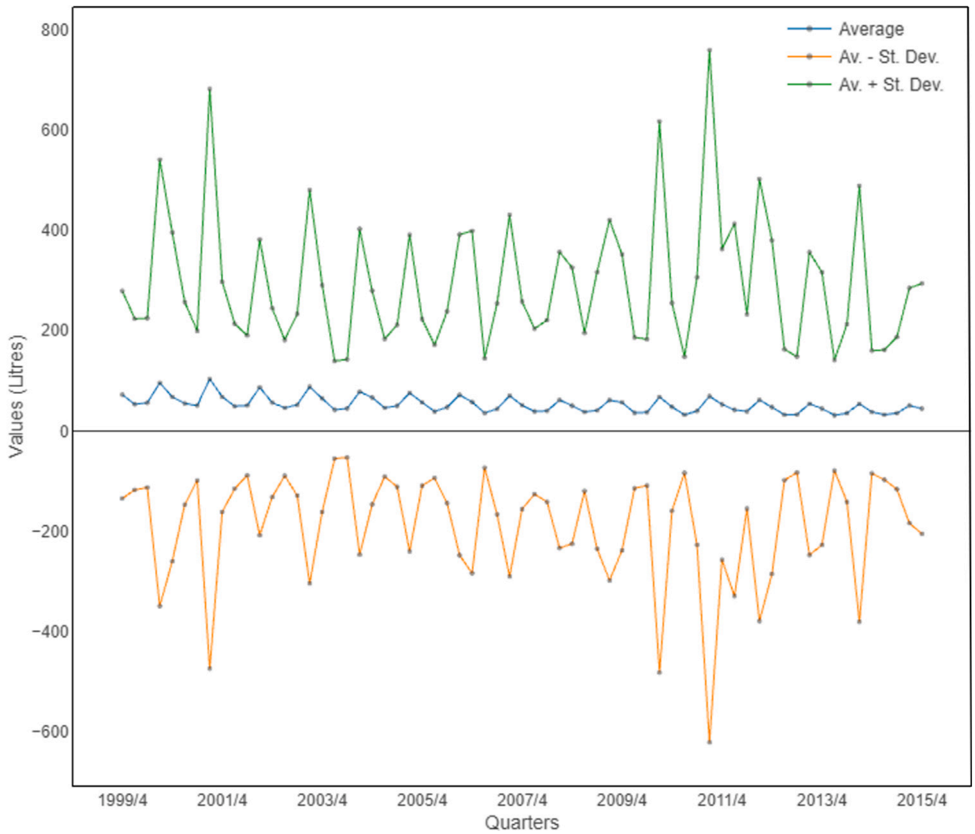


**Fig. 4.** Quarterly average of the water consumption data.

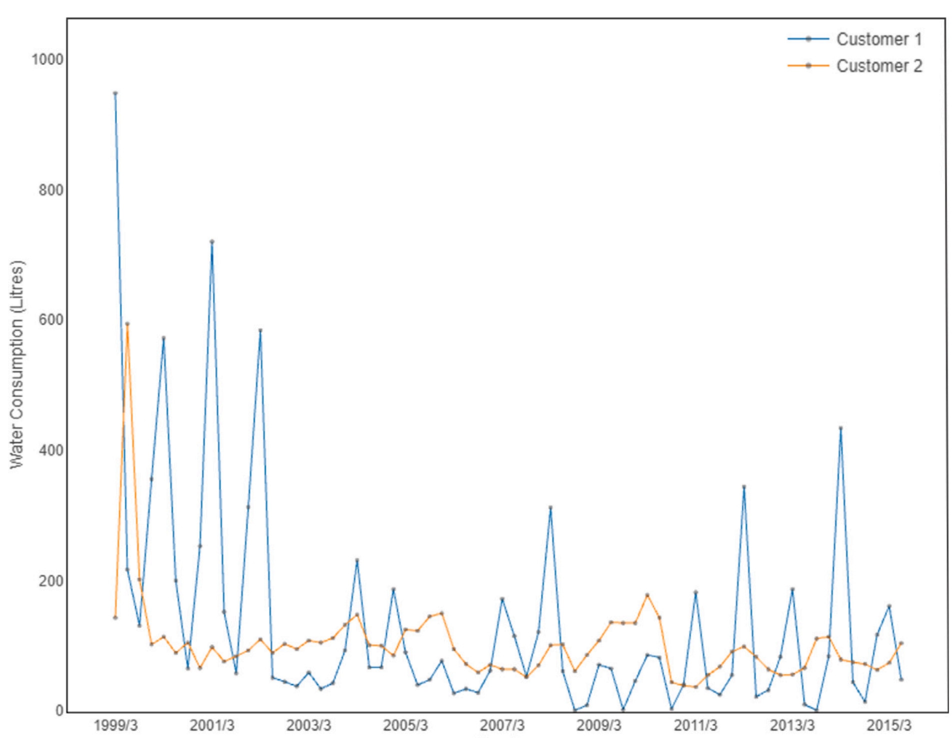


**Fig. 5.** Water consumption of two customers.

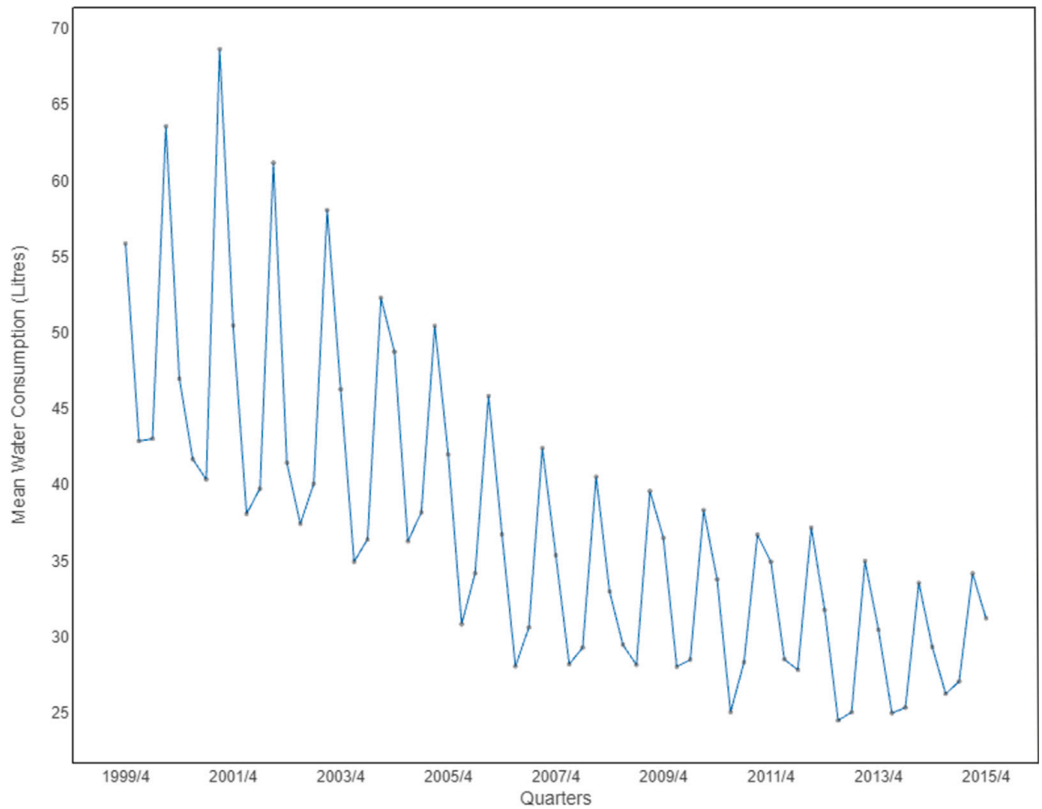


**Fig. 6.** Centroid of the cluster 0.

Figs. 6, 7 and 8 show the centroids of the four clusters. Due to the large differences in water consumption of customers in each cluster, the centroids have had to be represented separately. These differences range from about 50 litres in cluster 0 to some buildings with more than 10000 litres in cluster 3.

Then, in this work our objective is obtain a prediction model to forecast the water consumption for residential customers, that is, household consumption. Thus, from now on we will focus on the consumption data of cluster 0.

### 4.4. Deep neural network

Fully connected neural networks, also known as feed-forward neural networks, are widely used for various problem domains. While convolutional neural networks are commonly applied to image processing and recurrent neural networks are suitable for sequential data analysis, fully connected neural networks have shown their effectiveness even in domains beyond their initial design.

In the context of our research, a deep feed-forward neural network (DFFNN) has been proposed because this type of deep neural network has not been sufficiently explored for water consumption forecasting as the deep learning models most commonly used in this problem have been LSTMs.

The DFFNN is characterized by its feed-forward structure, where information flows only in one direction, from the input layer through the hidden layers to the output layer. This architecture is particularly suitable for capturing complex relationships in data, making it applicable to forecasting tasks. For water consumption forecasting, the DFFNN is utilized to predict future water consumption based on historical data. This network excel in modeling patterns and dependencies in sequential data, which is crucial for accurate forecasting. By incorporating historical water consumption data as input features, the DFFNN learns to recognize patterns and make predictions for future time steps.

The structure of the DFFNN consists of layers of interconnected neurons. Each neuron in a layer is connected to every neuron in the subsequent layer. The neurons in the input layer receive the historical water consumption data as input, while the neurons in the output layer generate the predicted water consumption values. The hidden layers in between perform complex calculations and transformations on the input data, allowing the network to learn and extract relevant features for forecasting.

The training of the proposed DFFNN involves adjusting the weights of the connections between neurons to minimize the difference between the predicted outputs and the actual water consumption values. This process, known as backpropagation, utilizes optimization algorithms such as gradient descent to iteratively update the weights based on the error calculated during training. The network learns to adjust the weights in such a way that it improves its predictions over time.

To make predictions, the DFFNN processes the input data through the layers by applying weights to the connections between neurons. Each neuron receives weighted inputs from the previous layer, applies an activation function to compute its output, and passes the result to the neurons in the next layer. This process continues until the output layer produces the final water consumption prediction.

The choice of activation function for each neuron depends on the specific requirements of the problem. Commonly used activation functions include the sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$, the rectified linear unit (ReLU) function $ReLU(x) = \max(0, x)$, and the hyperbolic tangent function $\tanh(x)$. These functions introduce non-linearities into the network, enabling it to capture complex relationships between the input features and the target variable.

The DFFNN is implemented with API Keras 2.10.0 and TensorFlow framework 2.10.1 under Python language 3.9.

#### 4.4.1. Input matrix

In order to feed the proposed DFFNN we must build the input data matrix. In supervised learning methods, this matrix consisting of

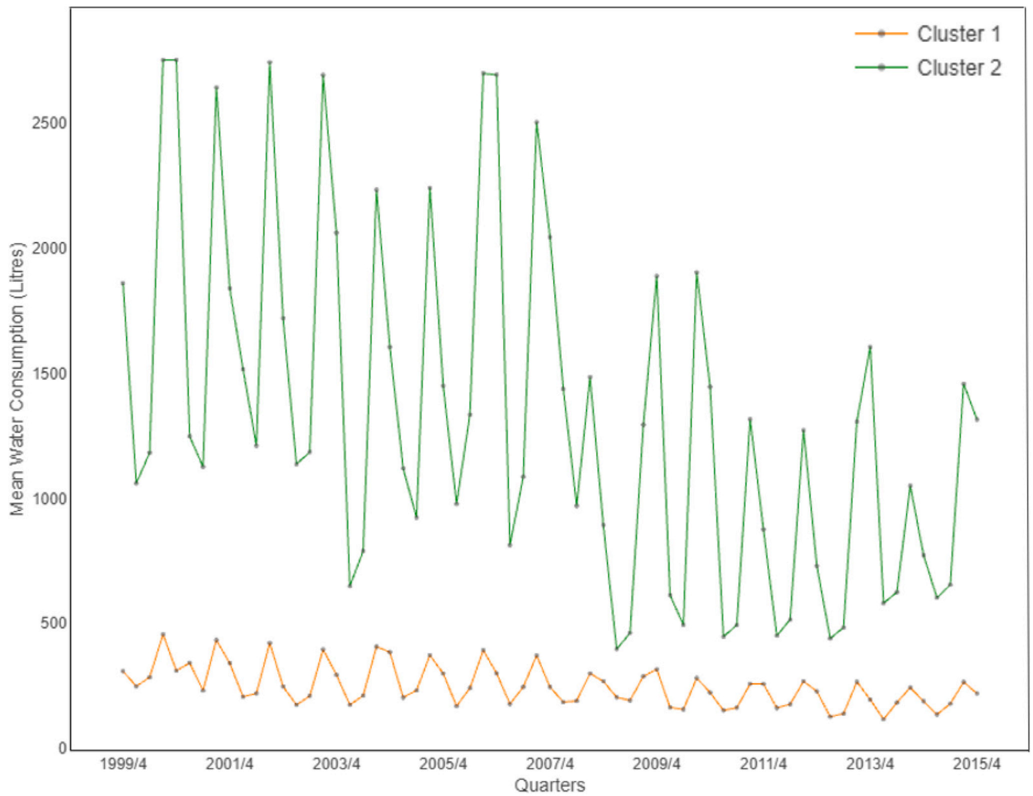


**Fig. 7.** Centroids of the cluster 1 *y* 2.

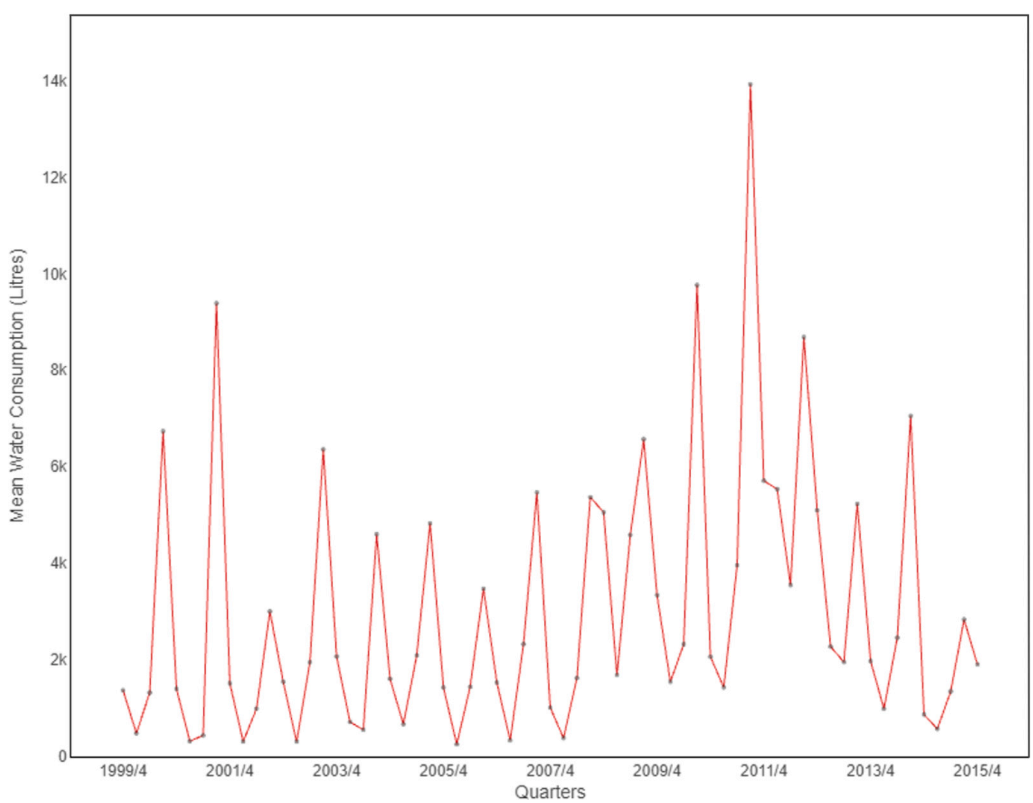


**Fig. 8.** Centroid of the cluster 3.

instances (rows) and attributes (columns), the last attribute being the class to be predicted. In this work, the attributes are a window of $W$ past water consumption values and the class is the next quarterly water consumption value to be predicted. For the construction of instances in this matrix we use a sliding window that moves one value from one instance to another. Thus, we have a target column, which will be every quarter to be predicted and the previous quarters are used for the DFFNN learning. Finally, the dataset we obtain is a window of twelve quarters ($W = 12$) for learning and one quarter as a target class or horizon ($h = 1$), which means to the next quarter, spanning from the third quarter of 2002 to the last quarter of 2015 for each customer. The input matrix has a total of 170586 rows (54 rows × 3159 customers) and 13 columns. This matrix can be represented as follows:

$$\left[\begin{array}{ccc|c} \longleftarrow & W & \longrightarrow & h \\ \hline C_1(1) & \dots & C_1(12) & C_1(13) \\ \vdots & & \vdots & \vdots \\ C_1(54) & \dots & C_1(65) & C_1(66) \\ \hline & & & \\ \vdots & & \vdots & \vdots \\ & & & \\ \hline C_N(1) & \dots & C_N(12) & C_N(13) \\ \vdots & & \vdots & \vdots \\ C_N(54) & \dots & C_N(65) & C_N(66) \end{array}\right] \tag{4}$$

#### 4.4.2. Design of the network architecture

The choice of the proposed DFFNN architecture and of its hyperparameters is not straightforward and the results given by the DFFNN are very susceptible to small variations of the hyperparameter values (Torres, Gutiérrez-Avilés, Troncoso, & Martínez-Álvarez, 2019).

For this reason, to evaluate and validate the effectiveness and accuracy of the DFFNN, the Keras Tuner tool, a scalable and easy to use hyperparameter optimization framework (Joshi, Owens, Shah, & Munasinghe, 2021), is applied to solve one of the main problems related to deep neural networks such as determining the value of hyperparameters.

With the objective of calculating the optimal values of the hyperparameters, a validation set has been used. Thus, the data are divided in 70% for training and the remaining 30% for the test. 30% of the training data are used for validation. In addition, in order to ensure that all customers are represented in the training, test and validation sets, these sets are generated by stratified selecting instances. Thus, 70% of each customer's instances are obtained for training and 30% for testing. The choice of these instances is made randomly, discarding options such as older or more recent registrations.

In our case, a grid search is carried out with the possibility of two, three or four hidden layers with the ReLU activation function and a variable number of neurons from 30 to 90 by 10 for 30 iterations. In turn, a random learning rate between 10–5 and 10–3 is also set. The optimizer is modified manually between Nadam and RMSProp optimizers, and loss function values is given by mean squared error (MSE).

After 100 training epochs, many iterations with different values and a precise study of the values obtained respected to the number of parameters included in the DFFNN, it is concluded that the best architecture for the DFFNN is two hidden layers with sixty neurons each. On the other hand, in relation to the parameters, Nadam is the chosen optimizer with a learning rate of 4 × 10-4.

Fig. 9 shows the evolution of the best model over 100 epochs in the training process. In particular, the MSE loss function for the training and validation sets is presented. In the training phase, it can be observed how the MSE decreases as the number of epochs increases, thus showing the convergence of the model and the absence of overfitting. From about epoch 15 onwards, the decrease in the validation set is no longer noticeable, although it continues, but in a much milder form than in the previous epochs, until it ends at epoch number 100 with a MSE value for validation of 1729.

The computational cost for hyperparameter optimization was 30 min for each optimizer (Nadam and RMSProp) approximately. Therefore, the total time invested was around one hour. Finally, the training time for the DFFNN was 20 s.

#### 4.4.3. Evaluation metrics

To evaluate the performance of the proposed DFFNN model, the following three performance metrics, namely Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE) are used. These selected metrics indicate the forecasting accuracy of the model. They are expressed in Eqs. (5), (6) and (7), respectively.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{C}_i(t+1) - C_i(t+1)| \tag{5}$$

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\frac{|\hat{C}_i(t+1) - C_i(t+1)|}{C_i(t+1)} \tag{6}$$

$$RMSE = \sqrt{\left(\frac{1}{N}\sum_{i=1}^{N}(\hat{C}_i(t+1) - C_i(t+1))^2\right)} \tag{7}$$

where $C_i(t+1)$ denotes the actual consumption at time t+1 for the $i$th customer, $\hat{C}_i(t+1)$ denotes the estimated consumption for the $i$th customer at time t+1 and $N$ is total number of customers considered.

## 5. Results

In this section, the results provided by the proposed DFFNN will be analyzed and the error metrics will be compared with that of other prediction methods such as Isotonic Regression (IR), Polynomial Regression (PR), Random Forest (RF), Gradient Boosted Tree (GBT) and Long Short-Term Memory (LSTM).

Figs. 10 and 11 depict the actual water consumption and predicted water consumption for each prediction method when predicting the test set. Once the water consumption has been predicted for each customer, all consumption has been aggregated for each quarter. Note that the predictions of the different algorithms have been plotted separately in two graphs to improve visualization.

From this graphically comparison, it can said that PR never reaching the peaks that are observed in the real water consumption data. In relation to DFFNN and IR methods, the predicted values are very similar, although it can be highlighted that the quarters in the first years with high consumption are better predicted by DFFNN, as the quarters progress this difference starts to stabilize. It is worth noting the poor predictions obtained by the GBT, which are always much lower than actual values.

In order to improve the visualization, Figs. 12 and 13 show the actual water consumption and predicted water consumption for each prediction method when both actual and predicted water consumption have been aggregated for all customers for each year. It should be noted in Fig. 12, that the low value produced in 2002 is due to the fact that the first quarter to be predicted in the test data is the third quarter of 2002. On the other hand, the prediction by shortage of IR and PR until 2006 where the prediction changes by excess and is almost always worse than DFFNN is shown in this Figure. In Fig. 13, it can be observed the poor prediction of GBT due to shortage and the prediction obtained by the RF always by excess. It should be noted, that the most accurate predictions are those obtained by the DFFNN and LSTM neural networks, with the DFFNN network obtaining the best results in most years. This same information, where the numerical values corresponding both actual and predicted water consumption can be also seen, is shown in Table 3. Note that all the values are expressed in litres.

Table 4 shows the MAE obtained by the proposed DFFNN and IR, PR, RF, GBT an LSTM forecasting models when aggregated errors for all customers and for each year. In the last row, the average of the MAE for all years is shown. It can be observed that the highest errors correspond to the predictions obtained by the GBT algorithm. On the other hand, it can be shocking the high errors reported in the year 2002 considering that this year has only two quarters in the test set. This is because there are some excessively high water consumption values in those quarters

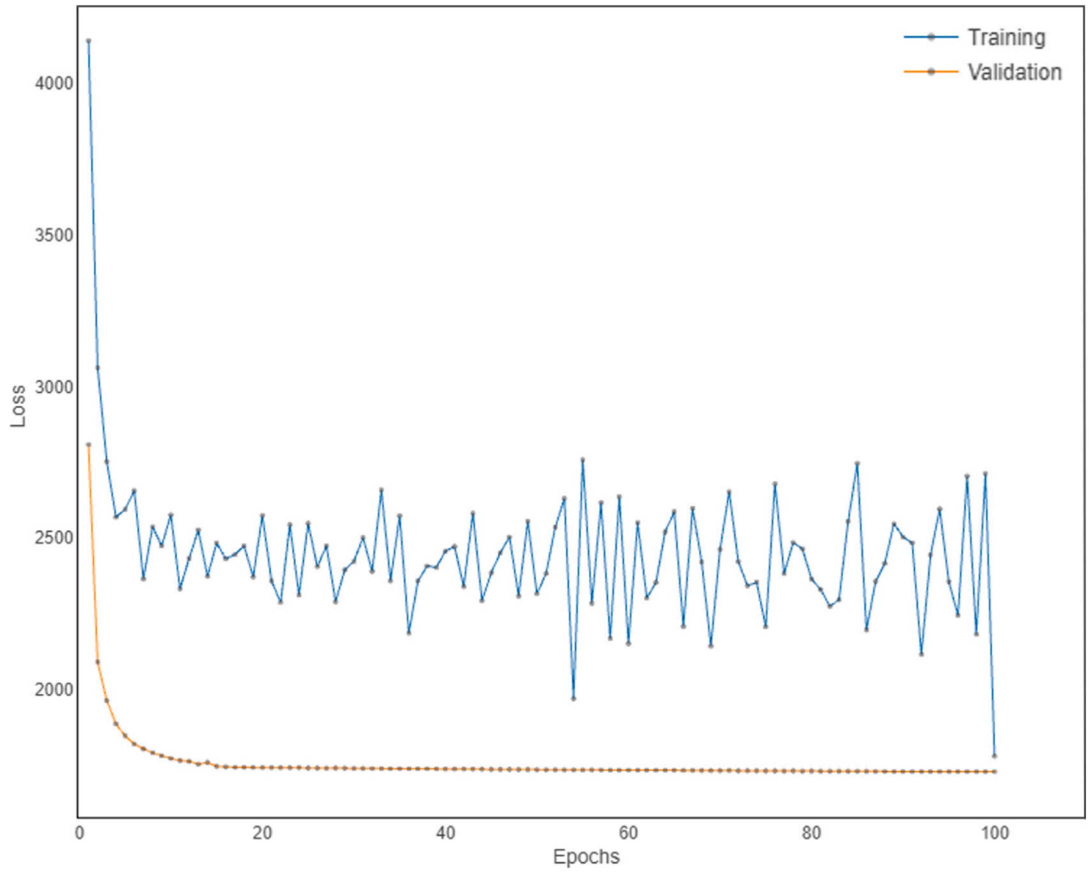


**Fig. 9.** Loss function through 100 epochs for training and validation.

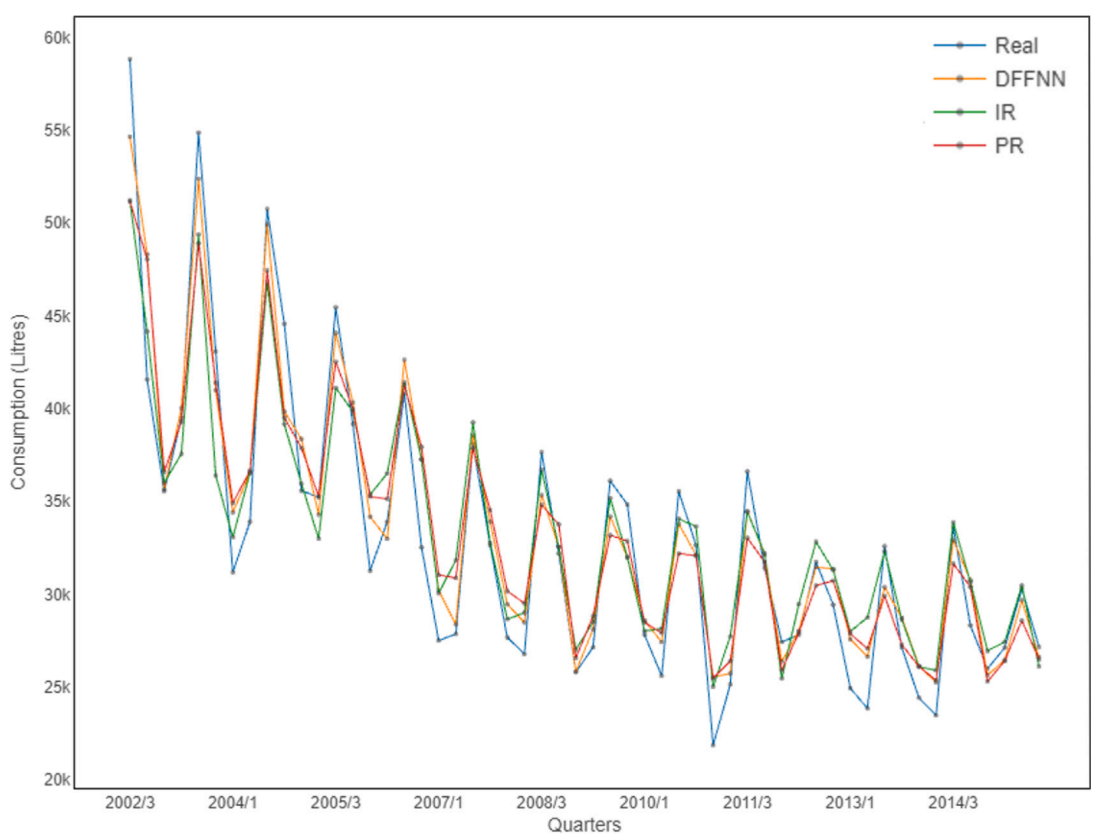


**Fig. 10.** Actual and predicted values for each quarter.

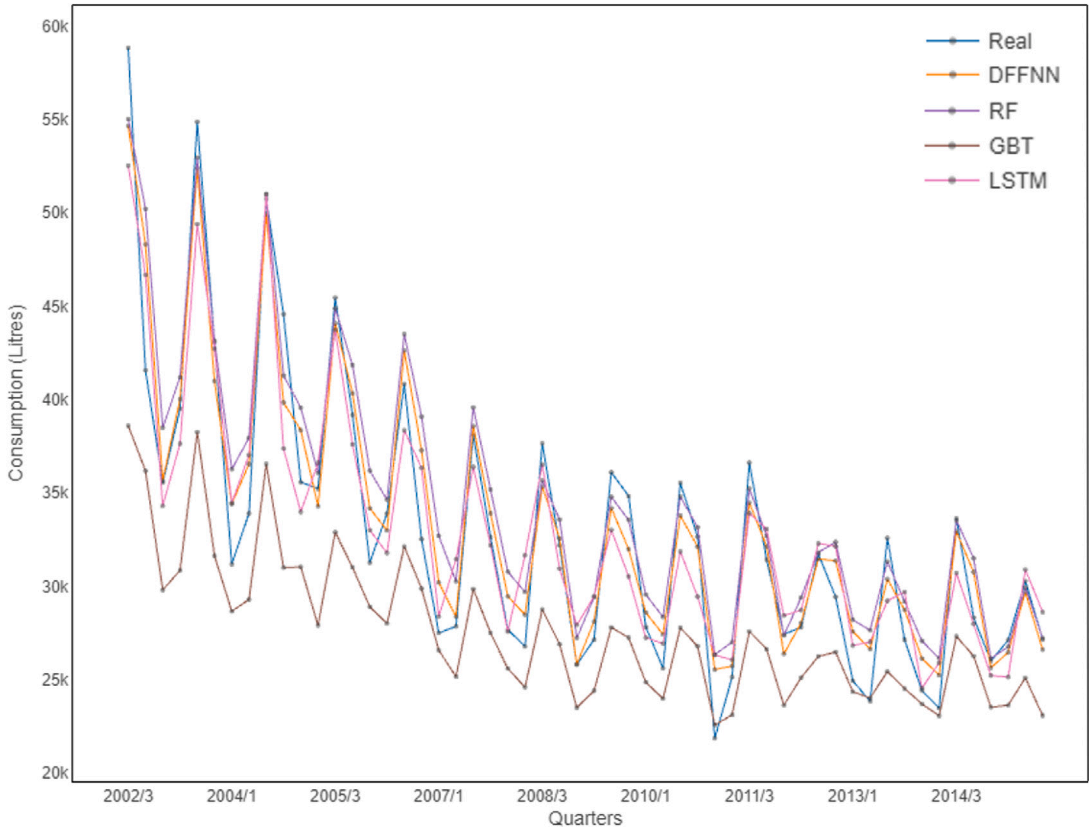


**Fig. 11.** Actual and predicted values for each quarter.

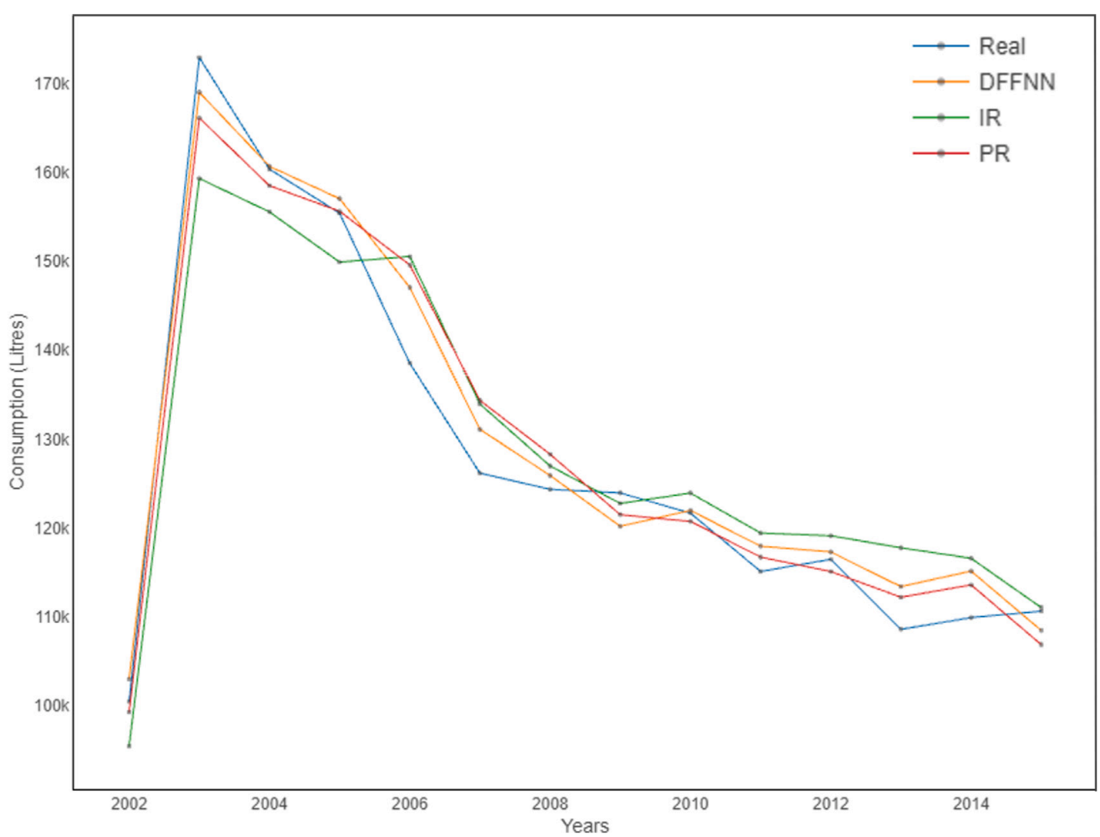


**Fig. 12.** Actual and predicted values for each year.

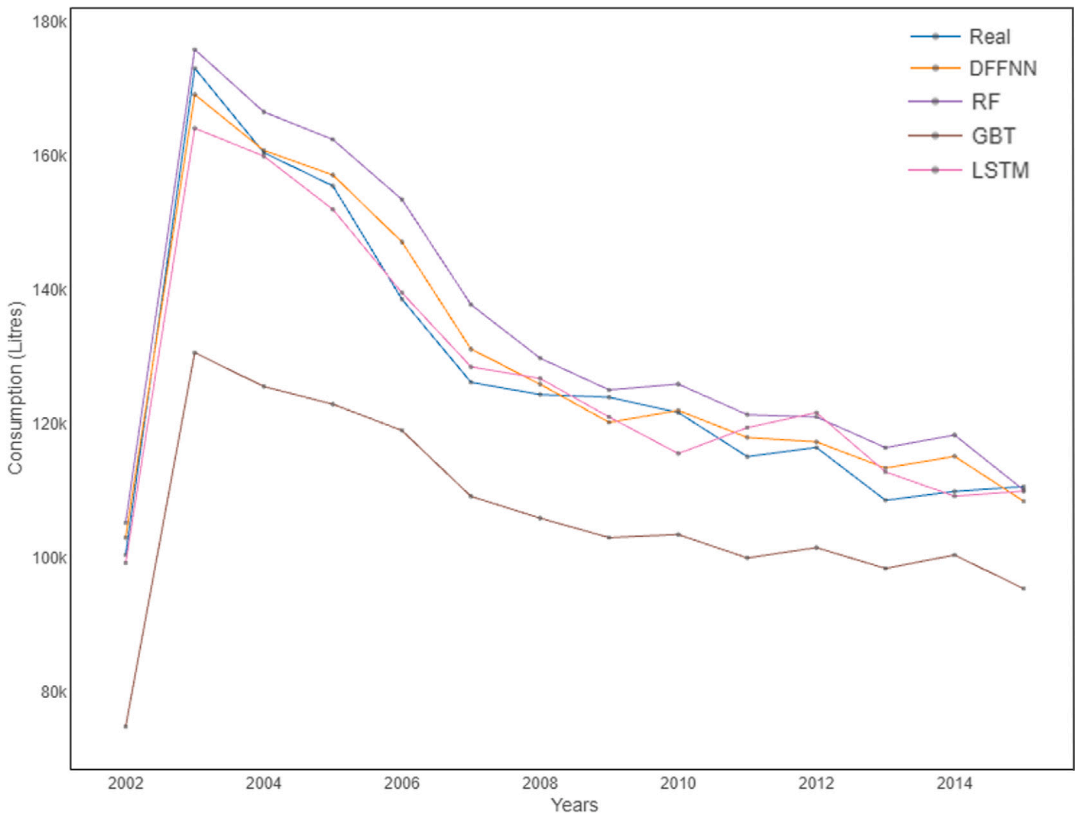


**Fig. 13.** Actual and predicted values for each year.

**Table 3**
Actual and predicted values for all forecasting algorithms for each year.

| Year | Real | DFFNN | IR | PR | RF | GBT | LSTM |
|---|---|---|---|---|---|---|---|
| 2002 | 100 408 | 102 962 | 95 383 | 99 220 | 105 232 | 74 794 | 99 197 |
| 2003 | 173 023 | 169 101 | 159 381 | 166 208 | 175 815 | 130 587 | 164 073 |
| 2004 | 160 429 | 160 754 | 155 612 | 158 566 | 166 503 | 125 540 | 159 905 |
| 2005 | 155 471 | 157 098 | 149 950 | 155 690 | 162 378 | 122 902 | 151 954 |
| 2006 | 138 552 | 147 104 | 150 586 | 149 625 | 153 443 | 118 972 | 139 525 |
| 2007 | 126 161 | 131 103 | 133 952 | 134 362 | 137 753 | 109 147 | 128 472 |
| 2008 | 124 337 | 125 895 | 126 972 | 128 294 | 129 761 | 1 005 903 | 126 745 |
| 2009 | 123 936 | 120 179 | 122 750 | 121 477 | 125 031 | 103 008 | 120 745 |
| 2010 | 121 658 | 121 963 | 123 923 | 120 724 | 125 903 | 103 465 | 115 537 |
| 2011 | 115 083 | 117 927 | 119 415 | 116 697 | 121 328 | 99 956 | 119 402 |
| 2012 | 116 454 | 117 285 | 119 104 | 115 079 | 120 996 | 101 508 | 121 639 |
| 2013 | 108 563 | 113 378 | 117 749 | 112 177 | 116 416 | 98 391 | 112 806 |
| 2014 | 109 884 | 115 127 | 116 565 | 113 569 | 118 303 | 100 398 | 109 179 |
| 2015 | 110 607 | 108 428 | 111 007 | 106 836 | 110 098 | 95 386 | 109 924 |

of 2002 that can be considered outliers. In fact, the annual average of water consumption for the test set is 33660 litres while the average of water consumption for 2002 is 50204 litres, a value much higher than the average. This issues is repeated for all error metrics used to assess quality of the prediction methods. Finally, the proposed DFFNN obtains the lowest MAE or to be among the lowest in almost all the years, and for this reason the average of the MAE for all years is the best when compared with other prediction methods including the deep LSTM network as can also be seen in Fig. 14.

Table 5 shows the MAPE (in percentage) obtained by the proposed DFFNN and IR, PR, RF, GBT an LSTM forecasting models when aggregated errors for all customers and for each year. As expected, considering the relationship between MAE and MAPE, results similar to those obtained with the MAE are reported. Again, the best MAPE is reached with the proposed DFFNN with an average of 5.59%, lower than LSTM with 7.02% and IR with 6.61%, which are the second and third best prediction methods as is also shown in Fig. 15. In general, the errors are low, except for the years 2004, 2006, 2011 and especially

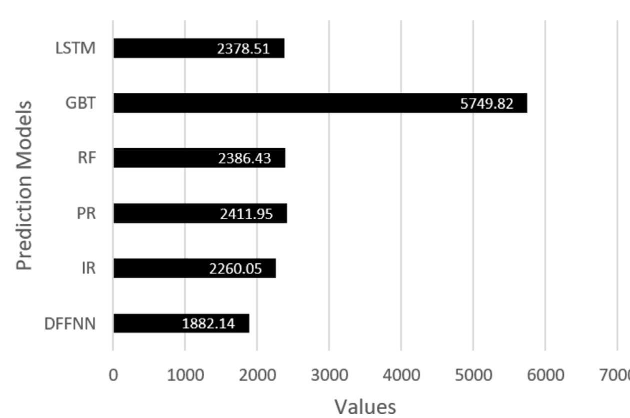


**Fig. 14.** Comparison of the MAE for the proposed DFFNN and other forecasting methods.

**Table 4**
The MAE obtained by all forecasting algorithms for each year.

| Year | DFFNN | IR | PR | RF | GBT | LSTM |
|---|---|---|---|---|---|---|
| 2002 | 5465.32 | 5101.46 | 7070.75 | 6227.88 | 12 806.83 | 5710.30 |
| 2003 | 1299.56 | 3641.11 | 2231.23 | 1651.23 | 10 609.24 | 2237.66 |
| 2004 | 2850.95 | 3471.89 | 3707.56 | 3161.37 | 8722.12 | 3464.40 |
| 2005 | 1565.60 | 1899.90 | 1518.48 | 2013.81 | 8142.37 | 1571.39 |
| 2006 | 2591.98 | 3008.50 | 2768.19 | 3722.86 | 4894.93 | 2536.66 |
| 2007 | 1235.49 | 1947.87 | 2173.97 | 2897.91 | 4253.43 | 1657.42 |
| 2008 | 1535.40 | 1128.25 | 2406.05 | 2343.61 | 4608.55 | 1839.66 |
| 2009 | 1438.97 | 1583.43 | 1831.75 | 1573.25 | 5231.95 | 2957.10 |
| 2010 | 1216.31 | 1301.75 | 1719.80 | 1430.22 | 4548.22 | 2196.41 |
| 2011 | 1782.25 | 2182.03 | 2196.63 | 2252.03 | 4137.99 | 2431.79 |
| 2012 | 867.00 | 1643.00 | 1054.08 | 1164.73 | 3736.46 | 1296.42 |
| 2013 | 2303.82 | 2433.31 | 2238.73 | 2591.61 | 2628.8 | 2744.40 |
| 2014 | 1652.71 | 1670.03 | 1907.24 | 2149.62 | 2371.52 | 1452.43 |
| 2015 | 544.60 | 628.17 | 942.88 | 229.90 | 3805.09 | 1202.98 |
| Average | 1882.14 | 2260.05 | 2411.95 | 2386.43 | 5749.82 | 2378.50 |

**Table 5**
The MAPE obtained by all forecasting algorithms for each year.

| Year | DFFNN | IR | PR | RF | GBT | LSTM |
|---|---|---|---|---|---|---|
| 2002 | 11.67 | 9.58 | 14.3 | 13.63 | 23.66 | 11.50 |
| 2003 | 2.75 | 7.93 | 4.58 | 4.05 | 23.71 | 4.76 |
| 2004 | 7.58 | 8.44 | 9.48 | 9.01 | 20.02 | 9.09 |
| 2005 | 4.12 | 4.65 | 3.77 | 5.41 | 20.49 | 4.06 |
| 2006 | 7.75 | 9.2 | 8.53 | 11.16 | 13.58 | 7.39 |
| 2007 | 4.19 | 6.73 | 7.53 | 9.77 | 12.61 | 5.48 |
| 2008 | 5.00 | 3.87 | 7.90 | 7.91 | 13.92 | 6.36 |
| 2009 | 4.29 | 5.10 | 5.71 | 5.30 | 15.94 | 9.41 |
| 2010 | 4.10 | 4.45 | 5.63 | 5.14 | 14.18 | 6.85 |
| 2011 | 6.83 | 8.3 | 8.09 | 8.94 | 12.80 | 9.16 |
| 2012 | 3.00 | 5.72 | 3.60 | 4.03 | 12.72 | 4.47 |
| 2013 | 8.71 | 9.79 | 8.48 | 10.10 | 8.65 | 10.14 |
| 2014 | 6.30 | 6.54 | 7.04 | 8.44 | 7.67 | 5.17 |
| 2015 | 1.96 | 2.33 | 3.34 | 0.82 | 13.62 | 4.40 |
| Average | 5.59 | 6.62 | 7.00 | 7.41 | 15.25 | 7.02 |

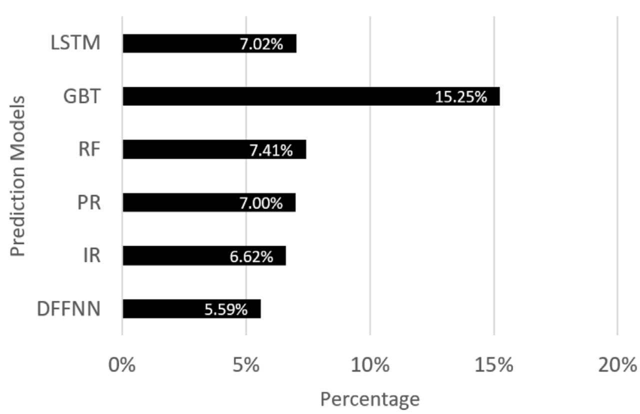


**Fig. 15.** Comparison of the MAPE for the proposed DFFNN and other forecasting methods.

in 2013, where a MAPE of more than 10% is obtained by the RF as the worst case and higher than 8.5% is obtained by the DFFNN as the best case. The reason for these MAPE values is the high variability in water consumption among the company's different customers.

Finally, Table 6 and Fig. 16 show the RMSE obtained by the proposed DFFNN and IR, PR, RF, GBT an LSTM forecasting models when aggregated errors for all customers and for each year. Before analyzing the results, it is considered very important to explain the differences between the metrics analyzed previously (MAE and MAPE) and the metric that is going to be analyzed next (RMSE). Since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors. This means the RMSE should be more useful when large errors are particularly undesirable.

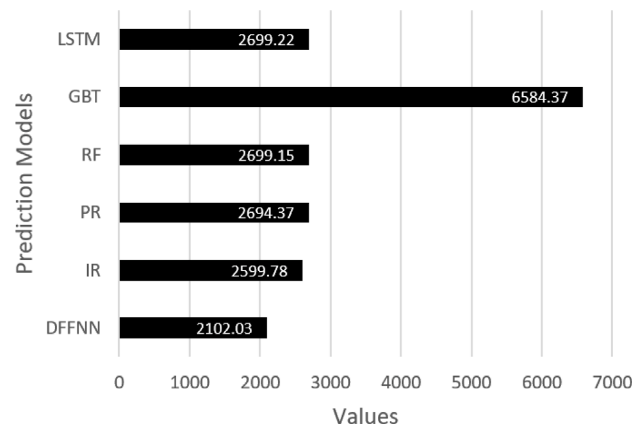


**Fig. 16.** Comparison of the MAPE for the proposed DFFNN and other forecasting methods.

Fig. 17 shows the actual and predicted water consumption for the customer with which the worst predictions were obtained, namely, a MAPE of 8.89%. There is a notorious difference between most quarters, but it is undoubtedly the fourth quarter of 2015 (2015/4) where the prediction soars with a difference of more than 120 litres. The difference that occurs in the quarters can be due to the fact that the range of values in water consumption for this customer is between 0 and 80, a fairly wide range of data that does not follow any pattern.

Fig. 18 shows the actual and predicted water consumption for the customer with which the most accurate predictions were obtained, namely, a MAPE of 0.008%. The differences between real and predicted values are minimal, with most of them being differences of less than 5 litres.

## 6. Conclusions

In this paper a model based on deep learning has been proposed for water consumption forecasting, contributing thus significantly in the field of deep learning applied to the water management domain. The main advances of this research have been the development of a model for the medium term that being trained with data from a set of customers is very accurate to obtain the predictions of each customer individually. Thus, a deep feed-forward network architecture has been proposed to predict the domestic consumption of a residential area at the household level and in the medium term. In addition, a grid search to find the best values of the hyperparameters such as number of layers, number of neurons for layer and learning rate have been carried out. Once these optimal values are determined, the best neural

**Table 6**
The RMSE obtained by all forecasting algorithms for each year.

| Year | DFFNN | IR | PR | RF | GBT | LSTM |
|---|---|---|---|---|---|---|
| 2002 | 5612.53 | 5686.57 | 7095.68 | 6678.56 | 14 805.80 | 5742.32 |
| 2003 | 1641.27 | 4434.64 | 3145.12 | 1947.32 | 11 339.85 | 2964.49 |
| 2004 | 3176.48 | 3720.28 | 3803.47 | 3640.75 | 10 161.91 | 4251.16 |
| 2005 | 1722.27 | 2465.77 | 1902.48 | 2447.37 | 8636.56 | 1576.05 |
| 2006 | 2963.97 | 3404.01 | 3422.32 | 4327.85 | 5540.48 | 2655.39 |
| 2007 | 1528.11 | 2432.03 | 2512.01 | 3213.01 | 5065.61 | 2051.69 |
| 2008 | 1692.70 | 1317.38 | 2457.57 | 2445.81 | 5383.73 | 2581.73 |
| 2009 | 1780.52 | 1747.83 | 1989.30 | 1624.02 | 5900.57 | 3078.70 |
| 2010 | 1344.89 | 1553.51 | 2074.50 | 1683.54 | 5143.50 | 2542.31 |
| 2011 | 2180.13 | 2350.34 | 2627.90 | 2604.57 | 5221.46 | 2770.20 |
| 2012 | 1111.45 | 1677.83 | 1188.53 | 1668.34 | 3892.12 | 1541.10 |
| 2013 | 2349.77 | 2989.3 | 2551.08 | 2776.84 | 3817.73 | 2805.93 |
| 2014 | 1767.57 | 1882.69 | 1912.25 | 2465.44 | 3341.17 | 1905.44 |
| 2015 | 556.75 | 734.80 | 1039.04 | 264.71 | 3930.68 | 1322.64 |
| Average | 2102.03 | 2599.78 | 2694.37 | 2699.15 | 6584.37 | 2699.22 |

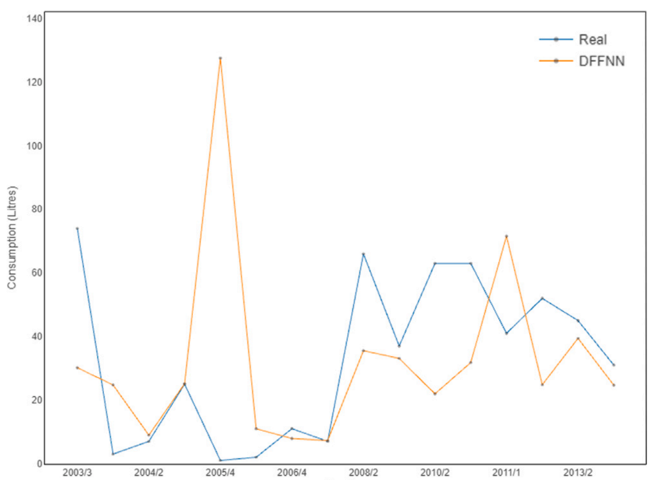


**Fig. 17.** Client with worst MAPE.


**Fig. 18.** Client with best MAPE.

network is applied to the water demand from 1999 to 2005 with a quarterly frequency to obtain forecasts for the next quarter. Results using a real-world water consumption corresponding the customers of a water supply company have been reported, reaching an average error of 5.5%. The performance of the proposed DFFNN network has obtained the smallest errors when compared with a isotonic regression, polynomial regression, random forest, gradient boosted tree and a deep LSTM network, with LSTM being the most popular deep learning model used to predict short-term water consumption. It is recommended to apply the forecasting model developed in this work when the available water consumption data are disaggregated by customer.

Future work will be based on transfer learning from the energy sector to the water sector, using deep learning models trained to predict electricity demand. Also, hybrid deep neural network models will also be designed in order to improve the results.

## CRediT authorship contribution statement

**A. Gil-Gamboa:** Conceptualization, Methodology, Investigation, Software, Data curation, Visualization, Writing – original draft, Writing – review & editing. **P. Paneque:** Methodology, Investigation, Writing – original draft, Writing – review & editing, Supervision. **O. Trull:** Conceptualization, Methodology, Investigation, Visualization, Writing – original draft, Writing – review & editing, Supervision. **A. Troncoso:** Methodology, Investigation, Visualization, Writing – original draft, Writing – review & editing, Supervision.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The data that has been used is confidential.

## Acknowledgments

Funding for open access publishing: Universidad Pablo de Olavide/CBUA. The authors would like to thank the Spanish Ministry of Science and Innovation for the support within the projects PID2020-117954RB-C21 and TED2021-131311B-C22. In addition, the authors would also like to thank the water supply company Aljarafesa for providing anonymized water consumption data of its customers for this research.

## References

Alsumaiei, A. A. (2020). A nonlinear autoregressive modeling approach for forecasting groundwater level fluctuation in urban aquifers. *Water*, *12*(3), 820.
Antunes, A., Andrade-Campos, A., Sardinha-Lourenço, A., & Oliveira, M. (2018). Short-term water demand forecasting using machine learning techniques. *Journal of Hydroinformatics*, *20*(6), 1343–1366.
Bandara, K., Bergmeir, C., & Smyl, S. (2020). Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. *Expert Systems with Applications*, *140*, Article 112896.
Bata, M., Carriveau, R., & Ting, D. (2020). Short-term water demand forecasting using hybrid supervised and unsupervised machine learning model. *Smart Water*, *5*.

Cao, L., Yuan, X., Tian, F., Xu, H., & Su, Z. (2023). Forecasting of water consumption by integrating spatial and temporal characteristics of short-term water use in cities. *Physics and Chemistry of the Earth, Parts A/B/C*, *130*, Article 103390.

Chen, L., Yan, H., Yan, J., Wang, J., Tao, T., Xin, K., et al. (2022). Short-term water demand forecast based on automatic feature extraction by one-dimensional convolution. *Journal of Hydrology*, *606*, Article 127440.

Du, B., Huang, S., Guo, J., Tang, H., Wang, L., & Zhou, S. (2022). Interval forecasting for urban water demand using PSO optimized KDE distribution and LSTM neural networks. *Applied Soft Computing*, *122*, Article 108875.

Du, B., Zhou, Q., Guo, J., Guo, S., & Wang, L. (2021). Deep learning with long short-term memory neural networks combining wavelet transform and principal component analysis for daily urban water demand forecasting. *Expert Systems with Applications*, *171*, Article 114571.

Faiz, M., & Daniel, A. (2022). Wireless sensor network based distribution and prediction of water consumption in residential houses using ANN. In *Proceedings of the international conference on internet of things and connected technologies* (pp. 107–116).

Fu, G., Jin, Y., Sun, S., Yuan, Z., & Butler, D. (2022). The role of deep learning in urban water management: A critical review. *Water Research*, *223*, Article 118973.

Guo, W., Liu, T., Dai, F., & Xu, P. (2020). An improved whale optimization algorithm for forecasting water resources demand. *Applied Soft Computing*, *86*, Article 105925.

Hao, W., Cominola, A., & Castelletti, A. (2022). Comparing predictive machine learning models for short- and long-term urban water demand forecasting in Milan, Italy. *IFAC-PapersOnLine*, *55*(33), 92–98.

Hu, P., Tong, J., Wang, J., Yang, Y., & Oliveira Turci, L. d. (2019). A hybrid model based on CNN and Bi-LSTM for urban water demand prediction. In *Proceedings of the IEEE congress on evolutionary computation* (pp. 1088–1094).

Jin, N., Zeng, Y., Yan, K., & Ji, Z. (2021). Multivariate air quality forecasting with nested long short term memory neural network. *IEEE Transactions on Industrial Informatics*, *17*(12), 8514–8522.

Joshi, S., Owens, J. A., Shah, S., & Munasinghe, T. (2021). Analysis of preprocessing techniques, keras tuner, and transfer learning on cloud street image data. In *Proceedings of the IEEE international conference on big data* (pp. 4165–4168).

Karamaziotis, P. I., Raptis, A., Nikolopoulos, K., Litsiou, K., & Assimakopoulos, V. (2020). An empirical investigation of water consumption forecasting methods. *International Journal of Forecasting*, *36*(2), 588–606.

Kavya, M., Mathew, A., Shekar, P. R., & P., S. (2023). Short term water demand forecast modelling using artificial intelligence for smart water management. *Sustainable Cities and Society*, *95*, Article 104610.

Koo, K.-M., Han, K.-H., Jun, K.-S., Lee, G., Kim, J.-S., & Yum, K.-T. (2021). Performance assessment for short-term water demand forecasting models on distinctive water uses in Korea. *Sustainability*, *13*(11), 6056.

Kühnert, C., Gonuguntla, N. M., Krieg, H., Nowak, D., & Thomas, J. A. (2021). Application of LSTM networks for water demand prediction in optimal pump control. *Water*, *13*(5), 644.

Lara-Benítez, P., Carranza-García, M., & Riquelme, J. (2020). An experimental review on deep learning architectures for time series forecasting. *International Journal of Neural Systems*, *31*(3), Article 2130001.

Limones, N., Vargas Molina, J., & Paneque, P. (2022). Spatiotemporal characterization of meteorological drought: A global approach using the Drought Exceedance Probability Index (DEPI). *Climate Research*, *88*, 137–154.

Liu, J., Zhou, X.-L., & Zhang, L.-Q. (2023). Forecasting short-term water demands with an ensemble deep learning model for a water supply system. *Water Resources Management*, *37*, 2991–3012.

Lu, H., Matthews, J., & Han, S. (2020). A hybrid model for monthly water demand prediction: A case study of Austin, Texas, 2. Article e1175.

Melgar-García, L., Gutiérrez-Avilés, D., Rubio-Escudero, C., & Troncoso, A. (2023). A novel distributed forecasting method based on information fusion and incremental learning for streaming time series. *Information Fusion*, *95*, 163–173.

Mohammad, E. B., & Mousavi-Mirkalaei, P. (2019). Extended linear and non-linear auto-regressive models for forecasting the urban water consumption of a fast-growing city in an arid region. *Sustainable Cities and Society*, *48*, Article 101585.

Mu, L., Zheng, F., Tao, R., Zhang, Q., & Kapelan, Z. (2020). Hourly and daily urban water demand predictions using a long short-term memory based model. *Journal of Water Resources Planning and Management*, *146*(9), Article 05020017.

Niknam, A., Zare, H. K., Hosseininasab, H., & Mostafaeipour, A. (2023). Developing an LSTM model to forecast the monthly water consumption according to the effects of the climatic factors in Yazd, Iran. *Journal of Engineering Research*, *11*(1), Article 100028.

Niknam, A., Zare, H. K., Hosseininasab, H., Mostafaeipour, A., & Herrera, M. (2022). A critical review of short-term water demand forecasting tools–what method should I use? *Sustainability*, *14*(9), 5412.

Pandey, P., Bokde, N. D., Dongre, S., & Gupta, R. (2021). Hybrid models for water demand forecasting. *Journal of Water Resources Planning and Management*, *147*(2), Article 04020106.

Pérez-Chacón, R., Asencio-Cortés, G., Martínez-Álvarez, F., & Troncoso, A. (2020). Big data time series forecasting based on pattern sequence similarity and its application to the electricity demand. *Information Sciences*, *540*, 160–174.

Pesantez, J. E., Berglund, E. Z., & Kaza, N. (2020). Smart meters data for modeling and forecasting water demand at the user-level. *Environmental Modelling and Software*, *125*, Article 104633.

Rezaali, M., Quilty, J., & Karimi, A. (2021). Probabilistic urban water demand forecasting using wavelet-based machine learning models. *Journal of Hydrology*, *600*, Article 126358.

Salloom, T., Kaynak, O., & He, W. (2021). A novel deep neural network architecture for real-time water demand forecasting. *Journal of Hydrology*, *599*, Article 126353.

Shan, S., Ni, H., Chen, G., Lin, X., & Li, J. (2023). A machine learning framework for enhancing short-term water demand forecasting using attention-biLSTM networks integrated with XGBoost residual correction. *Water*, *15*(20), 3605.

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 Journal*, *17*(1), 32–42.

Tao, M., Zhang, T., Xie, X., & Liang, X. (2023). Water footprint modeling and forecasting of cassava based on different artificial intelligence algorithms in Guangxi, China. *Journal of Cleaner Production*, *382*, Article 135238.

Torres, J. F., Gutiérrez-Avilés, D., Troncoso, A., & Martínez-Álvarez, F. (2019). Random hyper-parameter search-based deep neural network for power consumption forecasting. In *Proceedings of the international work-conference on artificial neural networks* (pp. 259–269).

Troncoso, A., Riquelme-Santos, J., Riquelme, J., Gómez-Expósito, A., & Martínez-Ramos, J. L. (2002). A comparison of two techniques for next- day electricity price forecasting. In *Proceedings of the 3th international conference on intelligent data engineering and automated learning* (pp. 384–390).

Trull, O., García-Díaz, J. C., & Troncoso, A. (2020). Initialization methods for multiple seasonal holt–winters forecasting models. *Mathematics*, *8*(2), 268.

Yan, K., Li, W., Ji, Z., Qi, M., & Du, Y. (2019). A hybrid LSTM neural network for energy consumption forecasting of individual households. *IEEE Access*, *7*, 157633–157642.

Zanfei, A., Brentan, B. M., Menapace, A., & Righetti, M. (2022). A short-term water demand forecasting model using multivariate long short-term memory with meteorological data. *Journal of Hydroinformatics*, *24*(5), 1053–1065.

Zhou, S., Guo, S., Du, B., Huang, S., & Guo, J. (2022). A hybrid framework for multivariate time series forecasting of daily urban water demand using attention-based convolutional neural network and long short-term memory network. *Sustainability*, *14*(17).