On the use of ARIMA models for short-term water tank levels forecasting

G. Viccione, C. Guarnaccia, S. Mancini and J. Quartieri

ABSTRACT

In this paper a statistical study on the time series of water levels measured, during the 2014, in the water tank of Cesine, Avellino (Italy) is presented. In particular, the ARIMA forecasting methodology is applied to model and forecast the daily water levels. This technique combines the autoregression and the moving average approaches, with the possibility to differentiate the data, to make the series stationary. In order to better describe the trend, over the time, of the water levels in the reservoir, three ARIMA models are calibrated, validated and compared: ARIMA (2,0,2), ARIMA (3,1,3), ARIMA (6,1,6). After a preliminary statistical characterization of the series, the models' parameters are calibrated on the data related to the first 11 months of 2014, in order to keep last month of data for validating the results. For each model, a graphical comparison with the observed data is presented, together with the calculation of the summary statistics of the residuals and of some error metrics. The results are discussed and some further possible applications are highlighted in the conclusions.

Key words | ARIMA models, short-term forecasting, time series analysis, urban water demand

G. Viccione (corresponding author)
C. Guarnaccia
J. Quartieri
Department of Civil Engineering (DICIV),
University of Salerno,
via Giovanni Paolo II, 132, 84084 Fisciano,
Italy
E-mail: gviccion@unisa.it

S. Mancini
Department of Information Engineering, Electric
  Engineering and Applied Mathematics (DIEM),
University of Salerno,
via Giovanni Paolo II, 132, 84084 Fisciano,
Italy

INTRODUCTION

Urban water forecasting allows making predictions of water needs, allowing water distribution system (WDS) operators to handle production, pumps energy and valve regulation. Demand estimation provides a useful predictive tool to water utilities for monitoring and controlling WDSs. Water demand, in the specific context of public water supply, is the total amount of water, including water losses to a certain extent, needed to supply customers, i.e. private households, public buildings, irrigation of public gardens, sewers cleaning, etc., within a time interval. A water demand estimation must ensure nodal demands while satisfying water quality and pressure levels across the network. A number of factors influence water demand, including population growth (residential, fluctuating), economic income, industry (size, technology involved, production types), local climate influencing seasonal demand patterns, price changes. Water demand is strongly related to water tank level, according to the well-known continuity equation. Let $V_i(t)$ be the Volume of water inside the tank at the time t, $h(t)$ be the water level inside the tank and Q_i the generic flow rate (for instance positive if it enters the tank, negative otherwise). The continuity equation coupled with the tank law read respectively:

$$\frac{dV_i(t)}{dt} = \sum_{i=1}^{m} Q_i(t) \tag{1}$$

$$\frac{dV_i(t)}{dt} = A_t \frac{dh_i}{dt}(t) \tag{2}$$

where A_t is the net surface of the tank of constant horizontal section. From the combination of Equations (1) and (2), it follows that there is a straight relation between flowrates and the water level.

In water applications, forecasting, even if complex, is beneficial for many reasons. First of all, investments in water supply systems can be extremely expensive, involving millions to hundreds of millions Euros, thus implying the need of forecasting a scenario which is naturally evolving because of the aging of pipelines, water tank degradation (Viccione *et al.* 2019a), water leakages, sizing, etc. Minimizing investment costs is therefore essential either in a short or long term interval when planning new developments or system expansion. Predictions are also relevant in processes for reviewing prices (Herrera *et al.* 2010). Secondly, as water is a precious resource to preserve for environmental and financial reasons, it is of interest knowing in advance what the water demand is expected to be in the short term, allowing a sustainable exploitation. In addition, short term forecasting of water use helps optimizing day-to-day utility operations and planning maintenance schedules (Shabani *et al.* 2018).

Several forecasting methods can be adopted in water applications, e.g. – per capita (or per customer) model, – end-use models which rely on customer behaviour, – regression models, e.g. Maidment *et al.* (1985), Zhou *et al.* (2000), Bakker *et al.* (2014) and Shabani *et al.* (2017) which are appropriate in the case of time varying water prices, personal income, population growth, - time-series/extrapolation models including averaging, trend analysis, exponential smoothing and autoregressive integrated moving average (ARIMA), in which the prediction is based onto projecting past trends.

The latter forecasting technique is based on the 'Time Series Analysis' (TSA) approach. These models range from straight averages of time series to relatively high elaborated statistical models. The ones based on 'AutoRegressive Integrated Moving Average' techniques, both Seasonal (SARIMA) or not (ARIMA), are powerful models able to predict the slope of data evolving on time (Box & Pierce 1970; Chatfield 1975; Box & Pierce 1976). These techniques have been developed mainly in Statistics, and then applied in Economics, Management, Physics, and other, e.g. (Milanato 2008; Manganelli & Tajani 2015). Some of the

authors applied the TSA deterministic decomposition and the ARIMA models to acoustical noise in urban areas (Guarnaccia *et al.* 2014a, 2014b, 2017) and in proximity of an airport, also in combination with non-homogeneous Poisson distribution (Guarnaccia *et al.* 2016). Other applications by some of the authors to electric consumption by public transportation in Sofia, Bulgaria, and to air pollution in Nuevo Leon, Monterrey, Mexico can be found respectively in Tepedino *et al.* (2015) and in Guarnaccia *et al.* (2014c).

Concerning water demand, water tank levels, inflows and outflows, extrapolation models usually rely on historical data on water use to forecast future trends. Statistical methods such as Multiple Linear Regression (MLR) and ARIMA models are commonly used for short-term urban water demand forecasting. Among the pioneers, Maidment & Miaou (1986) used ARIMA in the daily demand model based on a study of nine American cities. Soon later, Smith (1988) developed a time series model to forecast daily municipal water use in Washington, USA. Wong *et al.* (2010) applied seasonal autoregressive moving series models to analyse the structure of daily urban water consumption in Hong Kong as a function of rainfall. Two autoregressive integrated moving average models, one non-seasonal (daily, ARIMA model) and one seasonal (monthly, SARIMA model), to predict daily and monthly inflows to the Elephant Butte Reservoir, New Mexico, are adopted and discussed in Zamani Sabzi *et al.* (2018). A parallel adaptive weighting strategy of water consumption forecast for the next 24–48 h, using univariate time series of potable water consumption, is proposed in Sardinha-Lourenço *et al.* (2018). Long term forecasting of water consumption in the Hohoe municipality, Ghana, is proposed by Amponsah *et al.* (2015), using ARIMA as well. Suhartono and co-workers (Suhartono *et al.* 2018) proposed a hybrid class of methods based on Singular Spectrum Analysis (SSA) decomposition, Time Series Regression (TSR), and ARIMA, known as hybrid SSA-TSR-ARIMA, for water demand forecasting. A comparison among multiple linear (MLR) and nonlinear regression, autoregressive integrated moving average (ARIMA), artificial neural network (ANN) and hybrid wavelet transforms and artificial neural network methods (WA-ANN) for urban water demand prediction in Montreal, Canada, was made by Adamowski *et al.* (2012). Other techniques comparisons were made by Mohammadi

et al. (2005), de Lima *et al.* (2016), between ARIMA and ANN, and by Antunes *et al.* (2018) between Machine learning and ARIMA. The latter paper concludes that it is difficult to affirm that one of the two techniques is always better than the other, since in some cases ARIMA performs better than the machine learning and viceversa. The main conclusion is that in order to determine which strategy performs better, it should be taken into account the peculiarities of the data of the case under study. Moreover, one of the most interesting features of ARIMA techniques is that they require as input only the measured data of the parameter that it is going to be forecasted. This is a strong simplification with respect to machine learning approaches, that usually require several parameters in input.

This large literature about water demand and related parameters forecasting shows the large interest of the scientific community. Anyway, the reservoir or tank water level prediction is very important in reservoir management, in order to support the decision system. For instance, Ashaary *et al.* (2015), Rani & Parekh (2014) and Valizadeh & El-Shafie (2013) underline the importance of predicting water reservoir levels, implementing interesting models based on ANN models. Nwobi-Okoye & Igboanugo (2013) perform daily water level predictions in a reservoir, comparing ANN and ARIMA techniques, concluding that the accuracy of the ANN model increase with increasing inputs, while ARIMA model yields the best prediction in the considered dataset.

In this paper, ARIMA models are introduced as predictive tool for short term daily average water tank level forecasting in a rural area. This study is motivated by the fact that, in Italy, especially in rural towns, a single water tank can be used to serve the small area. For larger towns, the methodology used in this paper can be easily extended to any number of interconnected tanks, opening the way to studies about the relations and the influence between the models applied in each tank. Very often the tanks are equipped with water level meters and more rarely with flow rate meters (Viccione *et al.* 2019b). Thus, the potential interest of water managers can be to take a decision on the basis of the post processing of recorded water levels in tanks. By predicting short term water levels, it is possible to prevent overflow discharges in water tanks and to

optimize management plans for the water distribution, e.g. during drought events.

METHODOLOGY

The Box-Jenkins/ARIMA forecasting model is here adopted. It is amongst the most popular procedures for time series analysis and forecasting application. The order of an ARIMA model is usually denoted by the notation ARIMA(p,d,q), where p is the order of the autoregressive part, d is the order of the differencing and q is the order of the moving-average process. The general source formula is:

$$\phi_p(B)(1 - B)^d Y_t = \theta_q(B) e_t \tag{3}$$

in which Y_t is the value of the series observed at the time t, B is the delay operator, Φ are the autoregressive polynomials, θ are the moving average polynomials and e_t is the difference between the observed value Y_t and the forecast $\hat{Y}_t$ at the time t.

Error metrics

To have a numerical comparison on the effectiveness of the proposed models, some error metrics can be adopted. In this paper, the authors use the Mean Percentage Error (MPE), the Coefficient of Variation of the Error (CVE) and the Mean Absolute Scaled Error (MASE), defined as follow:

The MPE measures the error distortion, i.e. it is able to describe if the model overestimates or underestimates the actual data:

$$MPE = \frac{\sum_{t=1}^{n} ((Y_t - \hat{Y}_t)/(Y_t))100}{n} \tag{4}$$

The variation from the actual data in absolute value is given by the CVE, that provides the error dispersion:

$$CVE = \frac{\sqrt{(\sum_{t=1}^{n} (e_t)^2)/(n - 1)}}{\bar{Y}} \tag{5}$$

where $\bar{Y}$ is the mean value of the actual data in the considered time range. The MASE for seasonal time series is

computed according to the following formula (Franses 2016):

$$MASE = \frac{1}{n}\sum_{t=1}^{n}\frac{|e_t|}{(1)/n-k\sum_{i=k+1}^{n}|Y_i - Y_{i-k}|} \qquad (6)$$

In this case, according to the features of the dataset that will be presented in the following sections, a simple naïve model based on the assumption that the data of today is the same as yesterday ($\hat{Y}_t = Y_{t-1}$) is applied, choosing a lag $k = 1$ in the above formula (Hyndman & Athanasopoulos 2013).

DATASET ANALYSIS

This statistical study was applied on the time series of the levels observed in one of the reservoirs of the water supply system of the town of Avellino, Italy, located in the area of Cesine. Data refer to daily average water levels, expressed in meters, measured in 2014, from January to December. The original dataset was divided into two subsets: the first 333 values were used to perform the calibration and, therefore, the estimation of the model parameters (calibration dataset), the remaining ones were used for the validation phase (validation dataset).

The calibration dataset runs from 2 January 2014 to 30 November 2014. It shows some missing value, due to the occurrence of holidays and other, not identified, events. So, in the first 333 days of the year only 280 measurements were acquired. The last month of 2014, from the 1st to the 31st of December, was used for validation of the model. This subset is made of 31 data: 26 measured and 5 imputed by calculating the average of the two closest available measurements as already explained above.

The summary statistics of the 280 calibration data (Table 1) show a low value of the standard deviation, if compared to the average of the levels. Thus, the series seems

fairly regular over time, without frequent changes and great entity. Furthermore, since the values of asymmetry and kurtosis are very close to zero, the data can be considered normally distributed, as extracted from a normal, or Gaussian, population.

Then, the calibration dataset was reconstructed by imputing the missing values with the average of the immediately preceding and the immediately following one. Generally, for not-seasonal time series such the one under study, the imputation technique offers good results since the linear interpolation between two contiguous values takes into account the temporal location of the omitted value with respect to the contiguous ones.

The result of this elaboration is a calibration dataset that consists of 280 observed and 53 reconstructed values. In Figure 1(a) and 1(b) the complete reconstructed time series has been presented, together with the 'duration plot' of the data (sorted from the highest to the lowest).

The autocorrelation plot of the reconstructed series, Figure 1(c), shows that the function decreases as the lag increases. So, a periodicity in the data cannot be identified. In these cases, the forecasting through seasonal deterministic model does not seem plausible but the scenario suggests to use a stochastic technique as the ARIMA model.

In order to complete the characterization of the time series, some common statistical tests have been carried out. In particular, the Ljung-Box (LB) and Box-Pierce (BP) tests were used to verify a possible autocorrelation, while Lee-White-Granger and Terasvirta-Lin Granger tests were used to check linearity. Results (Table 2) show that the autocorrelation can be considered statistically significant in this case and that a linear process is present. Thus, it seems reasonable to use linear models like those of the ARIMA class.

Finally, the Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests (Table 3), performed to verify the stationarity, show the tendency to reject the null hypothesis, i.e. the presence of unit roots and the non-stationarity of the process. Thus, the analysed series

Table 1 | Summary statistics of the measured data (not imputed) of the calibration dataset referred to the water levels observed in the reservoir of Cesine

Sample Size	Mean [m]	Std. Dev. [m]	Median [m]	Min [m]	Max [m]	Skewness	Kurtosis
280	2.73	0.64	2.7	1.4	4	−0.16	−0.89

Uncorrected Proof

5   G. Viccione *et al.* │ ARIMAs for short-term urban water tank levels forecasting   **Water Supply** │ in press │ 2019
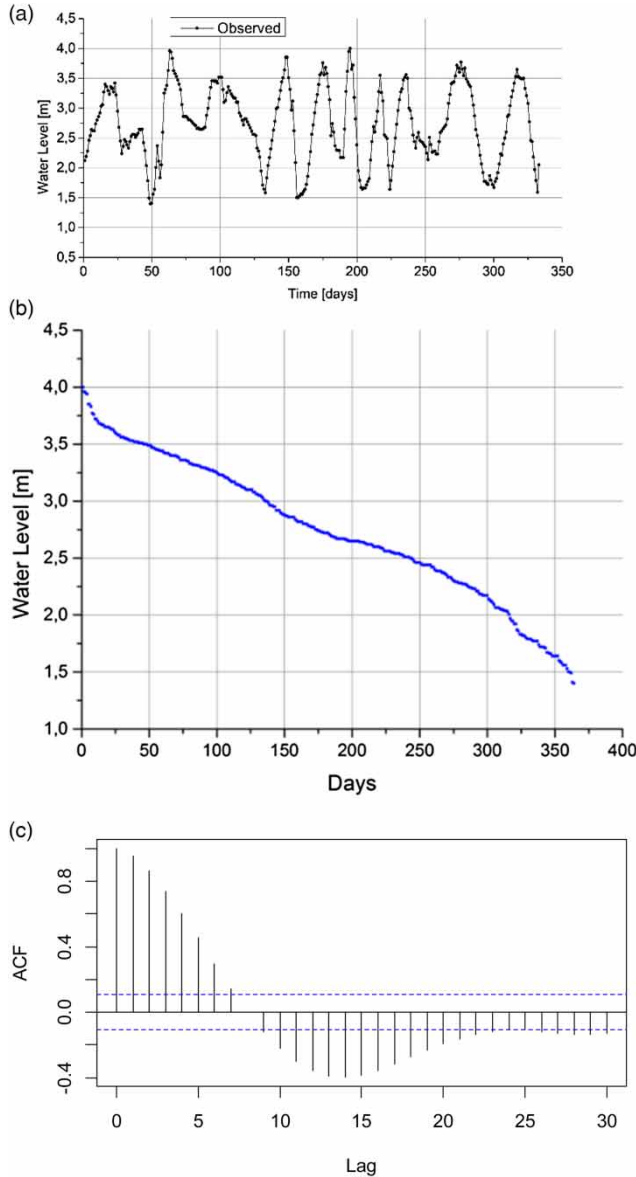
Figure 1 │ Water levels of the reservoir during the 2014, (a) observed, versus time (b) observed and reconstructed, sorted from the highest to lowest level (c) Autocorrelation plot of the reconstructed series, i.e. autocorrelation function plotted as a function of the lag.

Table 2 │ Results of Ljung-Box and Box-Pierce autocorrelation tests, and of Lee-White-Granger and Terasvirta-Lin Granger linearity tests, performed on the calibration dataset

Test	χ^2	d	p-value
Ljung-Box	1366.32	20	<2.2e-16
Box-Pierce	1389.20	30	<2.2e-16
Lee-White-Granger	1.9282	2	0.3813
Terasvirta-Lin-Granger	2.2767	2	0.3203

Table 3 │ Augmented Dickey-Fuller and Phillips-Perron Unit Root tests for stationarity of the time series performed on the 333 data of the water levels

Test	Statistic of the test	Lag	p-value
Augmented Dickey-Fuller	−6.4257	6	0.01
Phillips-Perron	−4.4054	5	0.01

presents a stationary trend: the mean, variance and autocorrelation does not significantly change during the overall observation time.

CALIBRATION OF THE MODEL

In order to better describe the trend, over the time, of the water levels in the reservoir, three ARIMA models were performed. Coefficients were estimated on the first 333 calibration periods by means of the statistical program 'R' and using the likelihood maximization as technique of parameter estimation.

Model ARIMA (2,0,2)

Because of the stationarity of the data, ARIMA model (2,0,2) does not include a differentiation process. The prediction of the water level in the tank in a given period t, provided by the model, is described by Equation (7).

$$\hat{Y}_t = \mu + \varphi_1(Y_{t-1} - \mu) + \varphi_2(Y_{t-2} - \mu) + \vartheta_1(e_{t-1}) + \vartheta_2(e_{t-2}) \qquad (7)$$

Table 4, instead, reports the estimated value of the coefficients of the model and their relative standard errors. The value of the intercept (about 2.7 m) can be considered the average value of the water level in the tank during the calibration period.

The graphical comparison between the observed data and the ARIMA model (2,0,2) (Figure 2(a)) shows how the curve of the predicted data (red curve) follows very closely the one of the observed ones (black curve). The time horizon of the forecast is one day. Only the fastest fluctuations of the level are not well foreseen since the model needs, at least, one period of time (in this case one day) to settle down after a sharp change in the slope of the observed signal.

Uncorrected Proof

6   G. Viccione *et al.* │ ARIMAs for short-term urban water tank levels forecasting   **Water Supply** │ in press │ **2019**

Table 4 │ Estimated coefficients of the ARIMA (2,0,2) model and related standard errors

	Intercept, μ [m]	AR1, φ_1	AR2, φ_2	MA1, ϑ_1	MA2, ϑ_2
Estimated Value	2.7314	1.8491	−0.8926	−0.4743	−0.1794
Standard Error	0.0634	0.0389	0.0354	0.0697	0.0659

Model ARIMA (3,1,3)

Model ARIMA (3,1,3) contemplates an order one of differentiation of the data. Moreover, both an autoregressive and a moving average term were added. The level in the reservoir, predicted by the model at a generic period t, is described by Equation (8).

$$\hat{Y}_t = Y_{t-1} + \varphi_1(Y_{t-1} - Y_{t-2}) + \varphi_2(Y_{t-2} - Y_{t-3}) \\ + \varphi_3(Y_{t-3} - Y_{t-4}) + \vartheta_1(e_{t-1}) + \vartheta_2(e_{t-2}) + \vartheta_3(e_{t-3}) \quad (8)$$

This model, as the previous one, has a forecasting time horizon of one day.

Table 5 shows the estimated value of the coefficients of the model and their relative standard errors.

Model ARIMA (6,1,6)

In order to provide a three-day forecasting horizon a model ARIMA (6,1,6) was developed. In the phase of estimation of parameters, the values of the first three autoregressive parameters (φ_1, φ_2, φ_3) and the first three moving average parameters (ϑ_1, ϑ_2, ϑ_3) were manually set to zero, performing the likelihood maximization with the new function. A differentiation of order one of the series was also carried out. The model forecasting formula is reported in Equation (9).

$$\hat{Y}_t = Y_{t-4} + \varphi_4(Y_{t-4} - Y_{t-5}) + \varphi_5(Y_{t-5} - Y_{t-6}) \\ + \varphi_6(Y_{t-6} - Y_{t-7}) + \vartheta_4(e_{t-4}) + \vartheta_5(e_{t-5}) + \vartheta_6(e_{t-6}) \quad (9)$$

Table 6 shows the model coefficients with the relative standard errors.

The graphical comparison, Figure 2, shows how the curve of ARIMA model (6,1,6) (green line) follows quite precisely the one related to the observed data (black line) but the extension of the forecasting horizon from one to three day is paid with a loss of precision. Furthermore, it is possible to observe the typical delay of the expected signal with respect to the observed one. As could be expected this model needs three periods to settle any rapid variations in the series.

VALIDATION OF THE MODEL

In order to verify the forecasting abilities of the three proposed models, a validation phase was carried out by using, as already explained, the values of the daily water level registered in the tank of Cesine in the month of December 2014.

The summary statistics of the forecast error are reported in Table 7.

Figure 3(a) shows the comparison between the level value observed in December (continuous black line with circles) and the level predicted by the ARIMA model (2,0,2) (dashed red line with squares). In the first days, the model's delay in anticipating the sharp change in slope in the observed series is manifested. Starting from the period 337 the forecast manages to follow very well the trend of the data which present very progressive and not rapid variations in value.

Model (3,1,3), instead, averagely overestimates the data observed throughout the validation period, as can be noticed in Figure 3(b).

Figure 3(c) shows the comparison between the observed water level and the forecast level provided by the ARIMA model (6,1,6). The forecast shows the typical delay, since the information extracted to predict the series is that of the three previous days.

RESULTS AND DISCUSSION

In order to quantitatively compare the proposed models, the analysis of the forecast errors in the calibration phase
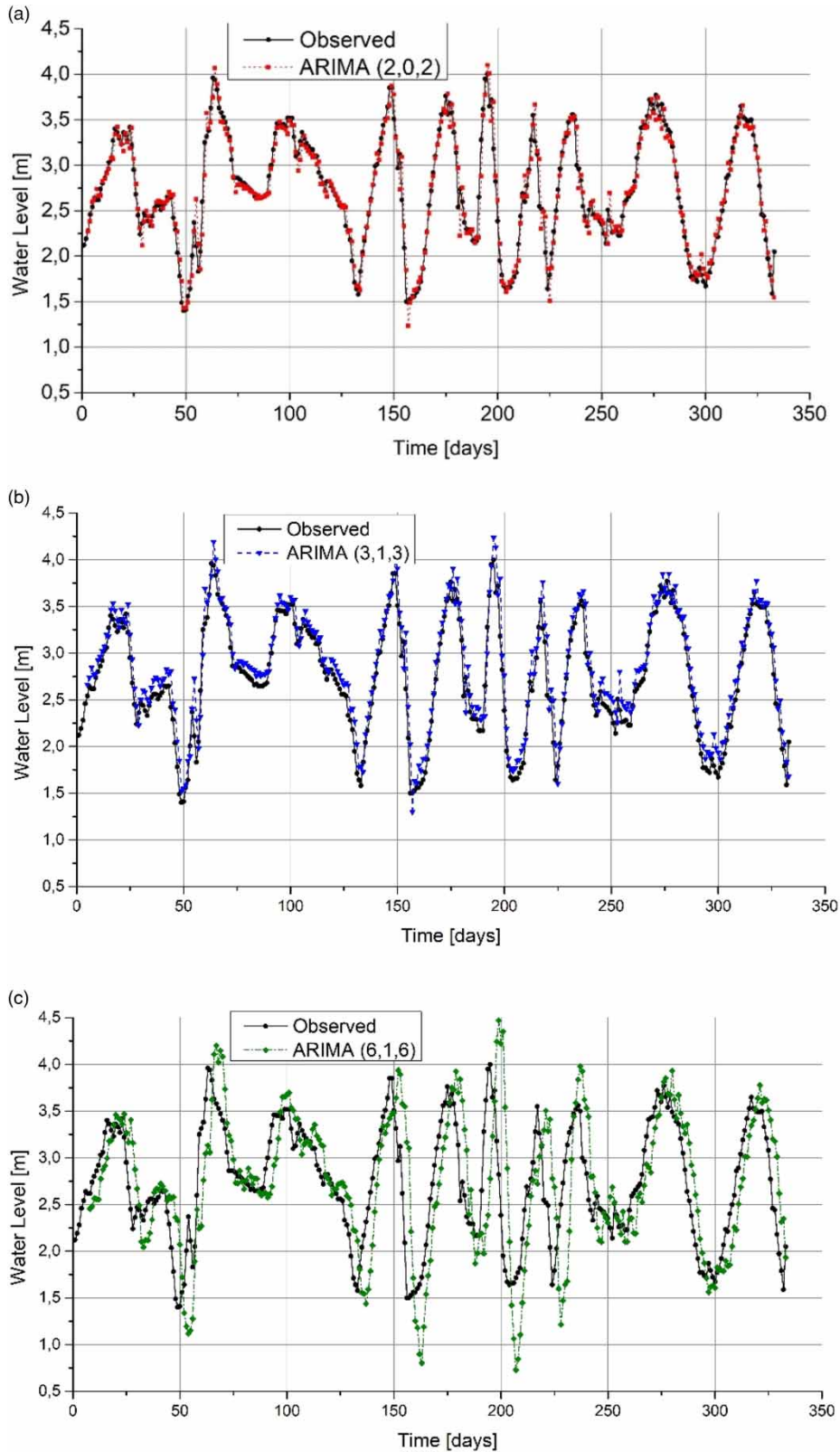
Figure 2 | Observed and forecasted water levels of the reservoir of Cesine in the calibration time range. The black line is the observed series, the coloured lines are the forecasts of the following models: (a) ARIMA (2,0,2), red line; (b) ARIMA (3,1,3), blue line; (c) ARIMA (6,1,6), green line.

Table 5 | Estimated values and standard error of the coefficients of the ARIMA (3,1,3) model

	AR1, φ_1	AR2, φ_2	AR3, φ_3	MA1, ϑ_1	MA2, ϑ_2	MA3, ϑ_3
Estimated Value	1.2350	0.2090	−0.5201	−0.8703	−0.5839	0.4542
Standard Error	0.2225	0.3822	0.1782	0.2154	0.2689	0.0950

Table 6 | Estimated values and standard error of the coefficients of the ARIMA (6,1,6) model

	AR4, φ_4	AR5, φ_5	AR6, φ_6	MA4, ϑ_4	MA5, ϑ_5	MA6, ϑ_6
Estimated Value	0.2594	−0.0225	−0.4463	−0.1590	0.2224	0.3933
Standard Error	0.3138	0.3347	0.1301	0.3416	0.3163	0.1532

Table 7 | Summary statistics of the errors during the validation phase

ARIMA Model	Mean [m]	Std. Dev. [m]	Median [m]	Min [m]	Max [m]	Skewness	Kurtosis
(2,0,2)	0.02	0.20	0.04	−0.92	0.44	−2.75	11.89
(3,1,3)	−0.10	0.20	−0.09	−1.03	0.27	−2.84	12.04
(6,1,6)	0.11	0.46	0.05	−0.56	1.36	0.65	−0.14

(residuals) and in the validation phase (errors) has been executed. The ARIMA model (2,0,2) has both an average of residuals and a median equal to zero and the standard deviation is particularly low, too (Table 8). The model (3,1,3) confirms the tendency to overestimate the real value of the level showing a negative average and median of the residuals. The asymmetry and kurtosis indices suggest a normal distribution of the residues for all the three proposed models. The ARIMA model (2,0,2) has a minimum and maximum error minor than the other proposed ones. The residuals distribution of model (6,1,6) exhibits the highest standard deviation, even though the mean residual is very low. This worsening in the performances is balanced by the advantage to give a broader forecasting horizon.

Residuals of the ARIMA model (2,0,2) appear to be symmetrically and regularly distributed (Figure 4(a)). As regards the other two models there are slight asymmetries in the histograms, which appear of modest magnitude, also when evaluating the quantile-quantile diagrams of Figure 5.

Figure 5 shows the quantile-quantile diagrams (Q-Q plot) of the residuals for all the three proposed models. This graph compares the cumulative distribution of the observed variable with the cumulative distribution of a normal distribution. If the observed variable has a normal distribution, the points of this joint distribution gather on a diagonal directed from the bottom to the top and from left to right. In the Q-Q plots shown we can see that all three models provide normally distributed forecast errors, i.e. they do not differ significantly from this desired distribution.

Thus, in general the proposed ARIMA models guarantee a forecast error apparently due only to random fluctuations well described by a normal distribution.

As for the autocorrelation of the residuals (Figures 6), the models (2,0,2) and (3,1,3) present a very low autocorrelation. The model (6,1,6), instead, seems not to have been able to exploit all the autocorrelation of the series. This result was predictable due to the fact that the lowest lag data (1, 2 and 3) could not be used in order to extend the forecast horizon to the next three days.

In Table 9, the summary statistics of the errors calculated in the validation phase are reported. As expected, the error distributions tend to get worse with respect to the calibration phase. Anyway, the mean errors and the standard deviations still give good results.

Finally, in Table 10 the values of the average percentage error (MPE), of the error variation coefficient (CVE) and of the MASE are reported, in both the phases. The model (2,0,2) presents the lower average error (MPE) and the
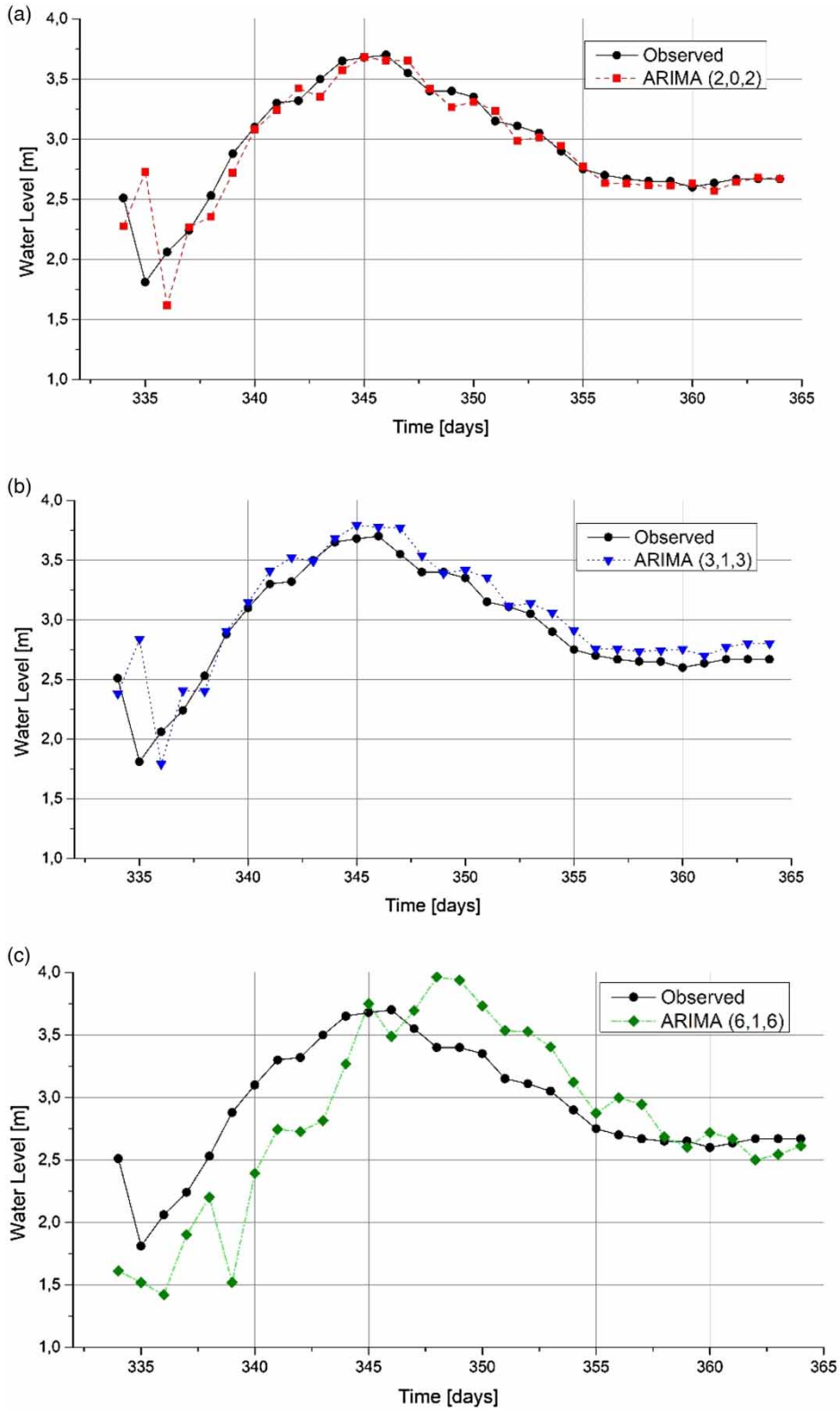
Figure 3 | Water levels (observed and forecasted) of the Cesine reservoir during the validation period (December 2014). The black line is the actual series, the red line is the forecast of the ARIMA (2,0,2) model (a), the blue line is the forecast of the ARIMA (3,1,3) model (b), the green line is the ARIMA (6,1,6) model (c).

Table 8 | Summary statistics of the residuals (calibration phase)

ARIMA Model	Mean [m]	Std. Dev. [m]	Median [m]	Min [m]	Max [m]	Skewness	Kurtosis
(2,0,2)	0.00	0.15	0.00	−0.52	0.52	−0.22	2.18
(3,1,3)	−0.12	0.15	−0.12	−0.63	0.38	−0.15	2.03
(6,1,6)	−0.01	0.57	0.02	−2.40	1.56	−0.44	1.30
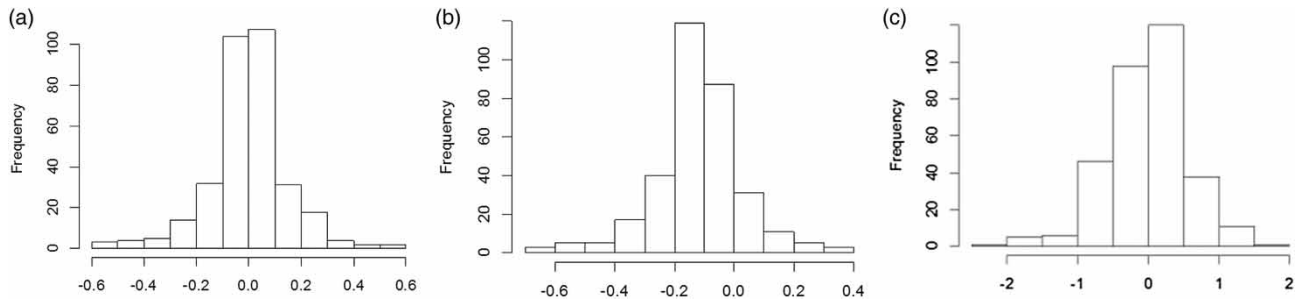
Figure 4 | Histograms of the residuals during the calibration phase. In (a) residuals of the ARIMA (2,0,2); in (b) residuals of the ARIMA (3,1,3); in (c) residuals of the ARIMA (6,1,6).
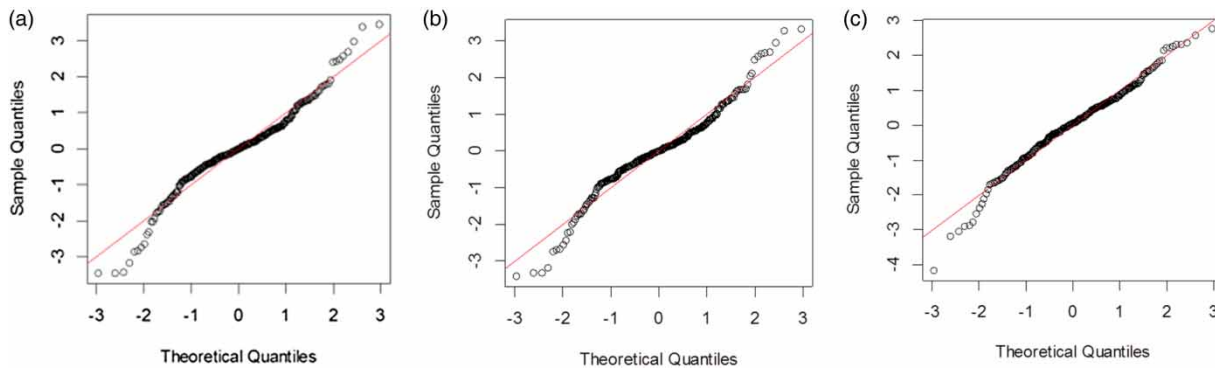

Figure 5 | Q-Q plots of the residuals during the calibration phase. In (a) residuals of the ARIMA (2,0,2); in (b) residuals of the ARIMA (3,1,3); in (c) residuals of the ARIMA (6,1,6).
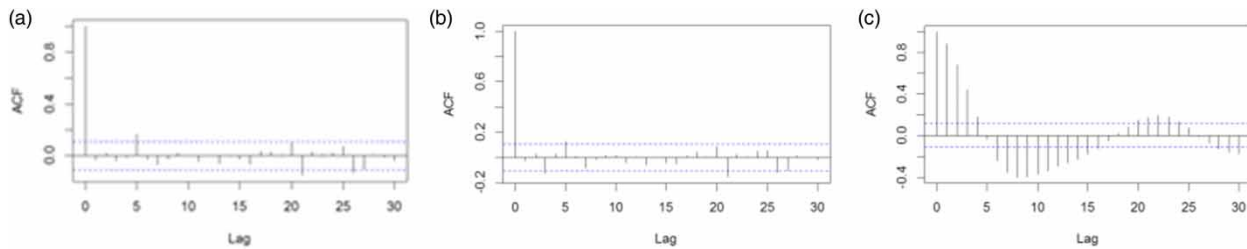

Figure 6 | Autocorrelation plots of the residuals during the calibration phase. In (a) residuals of the ARIMA (2,0,2); in (b) residuals of the ARIMA (3,1,3); in (c) residuals of the ARIMA (6,1,6).

Table 9 | Summary statistics of the errors (validation phase)

ARIMA Model	Mean [m]	Std. Dev. [m]	Median [m]	Min [m]	Max [m]	Skewness	Kurtosis
(2,0,2)	0.02	0.20	0.04	−0.92	0.44	−2.75	11.89
(3,1,3)	−0.10	0.20	−0.09	−1.03	0.27	−2.84	12.04
(6,1,6)	0.11	0.46	0.05	−0.56	1.36	0.65	−0.14

Table 10 | Value of the error metrics computed for the three proposed models during the two phase of the modelling process

Phase	ARIMA Model	MPE	CVE	MASE
Calibration	(2,0,2)	−0.306	0.055	0.764
Calibration	(3,1,3)	−4.959	0.070	1.123
Calibration	(6,1,6)	−1.781	0.209	3.145
Validation	(2,0,2)	0.278	0.070	0.858
Validation	(3,1,3)	−4.035	0.077	1.105
Validation	(6,1,6)	4.471	0.161	2.922

lower error variation coefficient (CVE), both in calibration and in validation, compared to the other models. This result confirms that for this time series, it does not seem useful to carry out the differentiation of order one and therefore the preferable model among those proposed to obtain a forecast with a time horizon of one day is the ARIMA (2,0,2). The model (6,1,6) shows a worsening of the values of the error metrics offset, basically due to the extension of the forecast horizon.

CONCLUSIONS

In this paper the problem of modelling the time series of water levels in a reservoir connected to the supply system of a city has been discussed. In particular, a methodology to model and forecast the daily levels was applied.

The univariate time series is related to the daily measurements of the water levels, observed during the year 2014, in the reservoir of the Cesine area connected to the network for the drinking water use of the city of Avellino (Italy). Some measurements, in correspondence of public holidays, were missing and, so, reconstructed imputing the average value of the two temporally closest measurements, one previous and one following the missing period. In this way a complete time series of 364 periods was obtained.

Once statistically characterized the time series, a stochastic modelling of the ARIMA class was proposed. The statistical analysis showed that the series of levels does not present evident periodicity or seasonality, which could be used for the construction of a deterministic or seasonal model. Furthermore, the series was linear and stationary. Then, the dataset was divided into two subgroups: 333

calibration periods, for the estimation of the coefficients of the proposed models, and 31 validation periods, corresponding to the measurements of the levels in December 2014.

The results of the modelling were very encouraging both by comparing, graphically, the observed level with the expected one, and by analysing the distribution of the forecast error from a quantitative point of view. The best forecast results both in the calibration phase, i.e. during the parameter estimation, and in the validation phase, were obtained with an ARIMA model (2,0,2). A model with differentiation of the first-order series, ARIMA (3,1,3), was also tested, however it did not provide clear improvements in the forecast. Both of these models are characterized by a one-day forecasting horizon. To obtain a longer time horizon, an ARIMA model (6,1,6) was implemented. It provided slightly worse forecasts but with a three-day forecast horizon. Therefore, a possible and novel approach could be to implement two models in parallel: the ARIMA (6,1,6) model could be used for 'medium term' (three days ahead) rough predictions, combined with the ARIMA (2,0,2) model, used to provide short term (one step ahead, daily) forecast. The use of these techniques on water data showed excellent potential, presenting a possible way for the integration between the current water service monitoring systems and the forecasts obtained with this methodology. The contribution of this paper is to extend the studies on the adoption of ARIMA techniques on this kind of data, dealing with a particular case study and focusing on the methodology to choose the best TSA model for the data under study. In the authors' opinion, the proposed approach could give water managers a tool to operate, especially in cases where flow rate data are not available.

Further studies will include a similar approach to the flow rate, in order to explore the potentiality of these techniques to a different dataset. In addition, in other time series in which the hourly data are available, different time bases can be considered, so that seasonal patterns could arise and could be implemented in the models.

REFERENCES

Adamowski, J., Fung Chan, H., Prasher, S. O., Ozga-Zielinski, B. & Sliusarieva, A. 2012 Comparison of multiple linear and

nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resour. Res.* **48** (1), W01528, 14 pages. doi:10.1029/2010WR009945.

Amponsah, S. K., Otoo, D. & Todoko, C. A. K. 2015 Time series analysis of water consumption in the Hohoe municipality of the Volta region. *Ghana. International Journal of Applied Mathematical Research* **4** (2), 393–403. doi:10.14419/ijamr.v4i2.3629.

Antunes, A., Andrade-Campos, A., Sardinha-Lourenço, A. & Oliveira, M. S. 2018 Short-term water demand forecasting using machine learning techniques. *Journal of Hydroinformatics* **20** (6), 1343–1366. doi:10.2166/hydro.2018.163.

Ashaary, N. A., Wan Ishak, W. H. & Ku-Mahamud, K. R. 2015 Forecasting model for the change of reservoir water level stage based on temporal pattern of reservoir water level. In: *Proc. of the 5th Int. Conf. on Computing and Informatics, ICOCI*, Istanbul, Turkey, pp. 692–697.

Bakker, M., Van Duist, H., Van Schagen, K., Vreeburg, J. & Rietveld, L. 2014 Improving the performance of water demand forecasting models by using weather input. *Procedia Engineering* **70**, 93–102. doi:10.1016/j.proeng.2014.02.012.

Box, G. E. P. & Jenkins, G. M. 1976 *Time Series Analysis: Forecasting and Control*. Holden-Day series in time series analysis and digital processing, Holden-Day, Inc., Oakland, USA.

Box, G. E. P. & Pierce, D. A. 1970 Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* **65**, 1509–1526.

Chatfield, C. 1975 *The Analysis of Time Series: an Introduction*. Chapman & Hall, New York.

de Lima, J. D., Adamczuk, G. O., Trentin, M. G., Batistus, D. R. & Pozza, C. B. 2016 A study of the performance of individual techniques and their combinations to forecast urban water demand. *Rev. Espac.* **37** (22), 5–28.

Franses, P. H. 2016 A note on the mean absolute scaled error. *International Journal of Forecasting* **32**, 20–22.

Guarnaccia, C., Quartieri, J., Rodrigues, E. R. & Tepedino, C. 2014a Acoustical noise analysis and prediction by means of multiple seasonality time series model. *International Journal of Mathematical Models and Methods in Applied Sciences* **8**, 384–393.

Guarnaccia, C., Quartieri, J., Mastorakis, N. E. & Tepedino, C. 2014b Development and application of a time series predictive model to acoustical noise levels. *WSEAS Transactions on Systems* **13**, 745–756.

Guarnaccia, C., Cerón Bretón, J. G., Quartieri, J., Tepedino, C. & Cerón Bretón, R. M. 2014c An application of time series analysis for forecasting and control of carbon monoxide concentrations. *International Journal of Mathematical Models and Methods in Applied Sciences* **8**, 505–515.

Guarnaccia, C., Quartieri, J., Tepedino, C. & Rodrigues, E. R. 2016 A time series analysis and a non-homogeneous Poisson model with multiple change-points applied to acoustic data. *Applied Acoustics* **114**, 203–212.

Guarnaccia, C., Mastorakis, N. E., Quartieri, J., Tepedino, C. & Kaminaris, S. D. 2017 Development of Seasonal ARIMA Models for Traffic Noise Forecasting. In: *Proc. of the 21st Int. Conf. on Circuits, Systems, Communications and Computers, CSCC 2017*, Heraklion, Crete, Greece. *MATEC Web of Conferences*, **125**.

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

Hyndman, R. J. & Athanasopoulos, G. 2013 *Forecasting: Principles and Practice*. OTexts.

Maidment, D. R. & Miaou, S. P. 1986 Daily water use in nine cities. *Water Resources Research* **22** (6), 845–851. doi:10.1029/WR022i006p00845.

Maidment, D., Miaou, S. & Crawford, M. 1985 Transfer function models of daily urban water use. *Water Resour. Res.* **21** (4), 425–432. doi:10.1029/WR021i004p00425.

Manganelli, B. & Tajani, F. 2015 Macroeconomic variables and real estate in Italy and USA. *Italian Journal of Regional Science* **14** (3), 31–48.

Milanato, D. 2008 *Demand Planning. Processi, metodologie e modelli matematici per la gestione della domanda commerciale*. Springer, Milano, in Italian.

Mohammadi, K., Eslami, H. R. & Dardashti, S. D. 2005 Comparison of regression, ARIMA and ANN models for reservoir inflow forecasting using snowmelt equivalent (a Case study of Karaj). *J. Agri. Sci. Technol.* **7**, 17–30.

Nwobi-Okoye, C. C. & Igboanugo, A. C. 2013 Predicting water levels at Kainji Dam using artificial neural networks. *Nigeria Journal of Technology* **32** (1), 129–136.

Rani, S. & Parekh, F. 2014 Application of artificial neural network (ANN) for reservoir water level forecasting. *International Journal of Science and Research (IJSR)* **7** (3), 1078–1082.

Sardinha-Lourenço, A., Andrade-Campos, A., Antunes, A. & Oliveira, M. S. 2018 Increased performance in the short-term water demand forecasting through the use of a parallel adaptive weighting strategy. *Journal of Hydrology* **558**, 392–404. doi:10.1016/j.jhydrol.2018.01.047.

Shabani, S., Yousefi, P. & Naser, G. 2017 Support vector machines in urban water demand forecasting using phase space reconstruction. *Procedia Engineering* **186**, 537–543. doi:10.1016/j.proeng.2017.03.267.

Shabani, S., Candelieri, A., Archetti, F. & Naser, G. 2018 Gene expression programming coupled with unsupervised learning: a two-stage learning process in multi-scale, short-termwater demand forecasts. *Water (Switzerland)* **10** (2), 142. doi:10.3390/w10020142.

Smith, J. 1988 A model of daily municipal water use for short-term forecasting. *Water Resour. Res.* **24** (2), 201–206. doi:10.1029/WR024i002p00201.

Suhartono, S., Isnawati, S., Salehah, N. A., Prastyo, D. D., Kuswanto, H. & Lee, M. H. 2018 Hybrid SSA-TSR-ARIMA for water demand forecasting. *International Journal of Advances in Intelligent Informatics* **4** (3), 238–250. doi:10.26555/ijain.v4i3.275.

Tepedino, C., Guarnaccia, C., Iliev, S., Popova, S. & Quartieri, J. 2015 A forecasting model based on time series analysis applied to electrical energy consumption. *International Journal of Mathematical Models and Methods in Applied Sciences* **9**, 432–445.

Valizadeh, N. & El-Shafie, A. 2013 Forecasting the level of reservoirs using multiple input fuzzification in ANFIS. *Water Resources Management* **27** (9), 3319–3331. http://doi.org/10.1007/s11269-013-0349-5.

Viccione, G., Ingenito, L., Evangelista, S. & Cuozzo, C. 2019a Restructuring a water distribution network through the reactivation of decommissioned water tanks. *Water (Switzerland)* **11**, 1740. doi:10.3390/w11091740.

Viccione, G., Pellecchia, V. & Parente, G. 2019b Una proposta per la riduzione delle portate di sfioro nei serbatoi di testata. In: *VIII Seminario Tecnologie e Strumenti Innovativi per le Infrastrutture Idrauliche 'TeSI'*, Naples, 8th–9th July 2019 (in italian).

Wong, J. S., Zhang, Q. & Chen, Y. D. 2010 Statistical modeling of daily urban water consumption in Hong Kong: trend, changing patterns, and forecast. *Water Resour. Res.* **46**, W03506, 10 pages. doi:10.1029/2009WR008147.

Zamani Sabzi, H., Abudu, S., Alizadeh, R., …, Dilekli, N. & King, J. P. 2018 Integration of time series forecasting in a dynamic decision support system for multiple reservoir management to conserve water sources. *Energy Sources, Part A: Recovery, Utilization and Environmental Effects* **40** (11), 1398–1416. doi:10.1080/15567036.2018.1476934.

Zhou, S. L., McMahon, T. A., Walton, A. & Lewis, J. 2000 Forecasting daily urban water demand: a case study of Melbourne. *Journal of Hydrology* **236** (3–4), 153–164. doi:10.1016/S0022-1694(00)00287-0.