

Water Demand Forecast Using Generalized Autoregressive Moving Average Models [†]

Maria Mercedes Gamboa-Medina * and Fabrizio Silva Campos

São Carlos School of Engineering, University of São Paulo, São Carlos 13566-590, SP, Brazil

* Correspondence: mmgamboam@usp.br

[†] Presented at the 3rd International Joint Conference on Water Distribution Systems Analysis & Computing and Control for the Water Industry (WDSA/CCWI 2024), Ferrara, Italy, 1–4 July 2024.

Abstract: Short-time forecasting of the demand on water distribution networks is a challenging task because of the high variability and uncertainty of that demand. Of the different approaches used, we consider the probability modeling of demand time series to be the most interesting, and specifically propose the use of Generalized Autoregressive Moving Average (GARMA) models. The complete proposed model uses a gamma probability density function, variables for weekends, and harmonic functions for daily and weekly seasonality, among other parameters. In the context of the Battle of Water Demand Forecasting, we train and test the model with a demand database for ten District Metered Areas. We obtain high accuracy, with mean absolute error values of around 0.25 L/s to 1.89 L/s.

Keywords: autoregressive models; water supply; smart water



Citation: Gamboa-Medina, M.M.; Campos, F.S. Water Demand Forecast Using Generalized Autoregressive Moving Average Models. *Eng. Proc.* **2024**, *69*, 125. <https://doi.org/10.3390/engproc2024069125>

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

Published: 12 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The efficient management of water distribution systems represents a cornerstone of urban infrastructure operation, and in this context the task of accurately predicting water demand becomes increasingly critical. Reliable forecasts are essential not only for long-term planning, but for the day-to-day operation and decision making of water utility companies. Short-term forecasting strategies should consider the high variability and uncertainty of water consumption, influenced by several physical and socio-economic factors, and should adapt to changing conditions and use the limited quantity/quality of available information to provide precise predictions.

Many different methods have been proposed specifically for short-term demand forecast, and there is no consensus on their strengths and limitations [1]. The Battle of Water Demand Forecasting (BWDF) at the 3rd International WDSA CCWI Joint Conference in 2024 aimed to compare the effectiveness of forecasting methodologies. We propose that forecasting methods based on the probability modeling of time series are theoretically sound, allow the incorporation of problem-specific knowledge, and are diverse enough to be applied to several practice problems. However, careful selection of the models and their features is needed to actually explore their potential. Specifically, we propose that Generalized Autoregressive Moving Average (GARMA) models stand out for their potential to significantly improve demand forecast accuracy.

GARMA models, introduced by Benjamin et al. [1], offer a robust framework for addressing the challenges inherent in water demand forecasting. By accommodating non-normal distribution of data and incorporating complex seasonal patterns, these models provide a nuanced understanding of water usage dynamics. The GARMA framework synthesizes the flexibility of generalized linear models in handling a multitude of covariates and the precision of autoregressive processes in error modeling. It achieves this synthesis by employing distributions from the exponential family to ascertain the likelihood of future observations based on past data [2,3]. The exponential family includes, but is not limited

to, Gaussian, Poisson, gamma, and binomial distributions [2]. The model is articulated through two key equations, detailed as follows:

$$f(y_t|H_t) = \exp\left\{\frac{y_tv_t - b(v_t)}{\varphi} + d(y_t, \varphi)\right\} \quad (1)$$

This equation delineates the conditional density function, which calculates the probability of observing y_t given H_t . It integrates parameters termed v_t and φ alongside the specific functions $b(\cdot)$ and $d(\cdot)$, which are selected in accordance with the intended distribution type.

In the GARMA model, the calculation of the conditional mean μ_t is linked with the linear predictor η_t through a linking function denoted as $g(\cdot)$. This accounts for autoregressive and moving average terms, hence, we have Equation (2):

$$g(\mu_t) = \eta_t = \underline{x}_t'\underline{\beta} + \sum_{j=1}^p \phi_j \{g(y_{t-j}) - \underline{x}_{t-j}'\underline{\beta}\} + \sum_{j=1}^q \theta_j \{g(y_{t-j}) - \eta_{t-j}\} \quad (2)$$

In this configuration, x_t represents a vector of r explanatory variables, and $\underline{\beta}'$ is the coefficient vector $(\beta_1, \beta_2, \dots, \beta_r)$. The autoregressive parameters are denoted as $\underline{\phi}' = (\phi_1, \dots, \phi_p)$, and the moving average parameters as $\underline{\theta}' = (\theta_1, \dots, \theta_q)$. The model also accommodates various types of residuals, including Pearson residuals and residuals measured on both the original data scale and the predictor scale.

A GARMA(p, q) model is, hence, defined by the above equations. The model parameters, namely $\underline{\beta}'$, $\underline{\phi}'$, and $\underline{\theta}'$, are estimated through the method of conditional maximum likelihood within an iterative process of weighted least squares.

This paper seeks to contribute to the body of knowledge in this area by applying a GARMA model to forecast water demand in real-world water supply networks. Through detailed analysis and model application, and by focusing on the patterns of water usage and leveraging advanced statistical methods, we aim to demonstrate the effectiveness of this approach, offering insights and practical tools for urban water utility companies to enhance their forecasting capabilities.

2. Materials and Methods

Following the BWDF conditions, this study focuses on forecasting water demands within ten (10) District Metered Areas (DMAs) of a water distribution network in Northeast Italy, catering to diverse areas with varying characteristics, sizes, and water demands. We used hourly net inflow data (in L/s), representing all water consumption and leakages, for each DMA from 1 January 2021 to 31 March 2023.

During the calibration of the GARMA model, the time series to be modeled and the exogenous variables were defined [4]. In this instance, dummy variables were included to numerically represent the weekends within the model, allowing for the nuanced incorporation of day-specific effects on the modeled time series. These dummies consist of two binary variables to represent Saturday and Sunday, respectively.

We also incorporated harmonic components, using Fourier series, to model the complex seasonality of the demand series, where multiple seasonal patterns occur simultaneously [5–7]. Consequently, the methodology involves defining the number of sine and cosine pairs to account for daily and weekly seasonality, with 12 pairs for daily patterns and 48 pairs for weekly patterns.

The appropriate probability density function was determined [8], and was found to be the gamma distribution. For the ARMA(p, q) component of the model, various combinations were tested, with orders ranging from 0 to 3. Additionally, different periods of model training were explored to identify the optimal setup that minimizes forecast error. The standard forecast horizon was set at 7 days, or 168 observations, as specified in the BWDF conditions. Following the model definition and selection stages, parameter

estimation was carried out, with the model being trained for each dataset corresponding to each DMA.

3. Results and Discussion

The assessment of various combinations of ARMA parameter orders, coupled with an analysis of residuals, led to the selection of a GARMA(2,0) model. This choice was substantiated by the observed fit of the model to the data, as well as its ability to accurately represent the series' behavior. Figure 1 illustrates the net flow values (q_{net}) observed (depicted in black) alongside those adjusted by the model (shown in red) for the DMA A series. This comparison clearly demonstrates the model's robustness and its effectiveness in capturing the dynamics of the series. The congruence between the observed and model-adjusted values underscores the GARMA model's aptitude for accurately forecasting water demand.

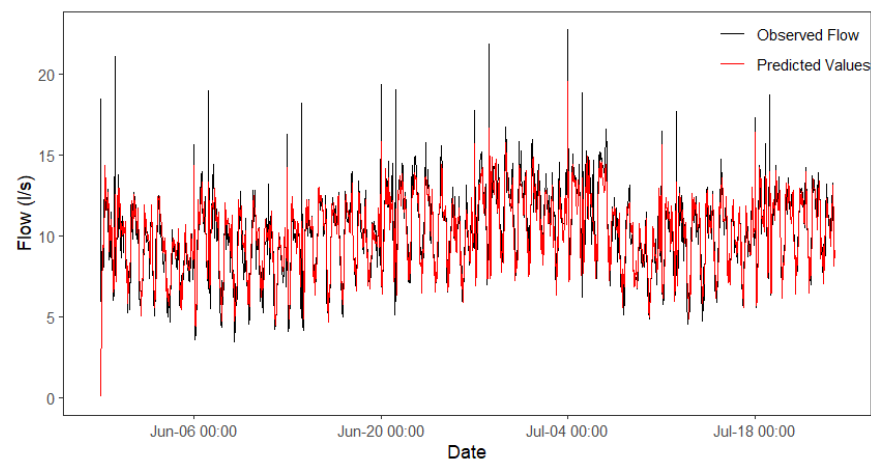


Figure 1. Comparison of observed (black) and GARMA model-adjusted (red) net flow values for DMA A.

The residual analysis suggests that the residuals can be considered independent, with no significant autocorrelation detected. The normal curve resemblance in the density distribution further supports the assumption of normality in the residuals.

The accuracy metrics presented in Table 1 evaluate the GARMA model's performance throughout a four-week forecasting challenge involving different District Metered Areas (DMAs), based on a training dataset. DMAs B and C stand out, as they consistently registered the lowest RMSE (root mean square error), MAE (mean absolute error), and MAPE (mean absolute percentage error) values, which indicates a notably reliable model fit for these areas. In contrast, DMAs D and E reported the highest error metrics, suggesting that these models' forecasts are less precise and might benefit from additional refinement and analysis for enhanced accuracy.

Table 1. The average accuracy metrics for the GARMA model across the four-week challenge period, detailing the RMSE, MAE, and MAPE for each DMA.

DMA	RMSE (L/s)	MAE (L/s)	MAPE (%)
A	1.32	0.90	12.10
B	0.48	0.31	3.05
C	0.35	0.25	6.34
D	2.46	1.89	6.40
E	3.19	1.74	2.14
F	0.94	0.69	7.99
G	1.36	0.91	3.27
H	1.34	0.95	4.31
I	1.55	1.08	4.54
J	1.50	1.04	3.89

4. Conclusions

It was shown that a demand time series could be satisfactorily modeled with a Generalized Autoregressive Moving Average (GARMA) model aimed towards demand forecasting. Based on the features of the time series and on its generator process, the model uses the gamma probability density function and includes additional variables to identify weekends, as well as Fourier series to account for seasonal (daily and weekly) fluctuations. The model was successfully calibrated with data from ten DMAs, each with unique demand patterns, showcasing the model's efficacy and flexibility. High accuracy was obtained when using the trained model to forecast demand for all the DMAs, with variations among them. The proposed methodology has good potential as a robust method for short-term demand forecasting within various urban water systems, and specific definition of model features could improve the results.

Author Contributions: Conceptualization, F.S.C. and M.M.G.-M.; methodology, F.S.C. and M.M.G.-M.; software, F.S.C. and M.M.G.-M.; validation, F.S.C. and M.M.G.-M.; writing—original draft preparation, F.S.C.; writing—review and editing, M.M.G.-M.; visualization, F.S.C.; supervision, M.M.G.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All datasets are available at <https://wdsa-ccwi2024.it/battle-of-water-networks/> (accessed on 10 January 2024). The processed data that support the findings of this study are available from the corresponding author upon reasonable request.

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

References

1. Donkor, E.A.; Mazzucchi, T.A.; Soyer, R.; Robertson, J.A. Urban Water Demand Forecasting: Review of Methods and Models. *J. Water Resour. Plan. Manag.* **2014**, *140*, 146–159. [\[CrossRef\]](#)
2. Benjamin, M.; Rigby, R.A.; Stasinopoulos, D.M. Generalized autoregressive moving average models. *J. Am. Statist. Assoc.* **2003**, *98*, 214–223. [\[CrossRef\]](#)
3. Albarracin, O.Y.E.; Alencar, A.P.; Ho, L.L. Generalized autoregressive and moving average models: Multicollinearity, interpretation and a new modified model. *J. Stat. Comput. Simul.* **2019**, *89*, 1819–1840. [\[CrossRef\]](#)
4. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*, 3rd ed.; OTexts: Melbourne, Australia, 2021. Available online: www.OTexts.com/fpp3 (accessed on 9 April 2022).
5. Brentan, B.; Luvizotto, E., Jr.; Izquierdo, J.; Pérez-García, R. Fourier series and Chebyshev polynomials applied to real-time water demand forecasting. *Acta Univ.* **2016**, *26*, 74–81. [\[CrossRef\]](#)
6. Odan, F.K.; Reis, L.F.R. Hybrid water demand forecasting model associating artificial neural network with Fourier series. *J. Water Resour. Plan. Manag.* **2012**, *138*, 245–256. [\[CrossRef\]](#)
7. Bakker, M.; Vreeburg, J.H.G.; van Schagen, K.M.; Rietveld, L.C. A fully adaptive forecasting model for short-term drinking water demand. *Environ. Model. Softw.* **2013**, *48*, 141–151. [\[CrossRef\]](#)
8. Rigby, R.A.; Stasinopoulos, M.D.; Heller, G.Z.; De Bastiani, F. *Distributions for Modeling Location, Scale, and Shape: Using GAMLSS in R*, 1st ed.; Chapman and Hall/CRC: New York, NY, USA, 2019. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.