Hourly Discharge Modelling and Forecast for a Run-of-river Dam

JOANA SEABRA-SILVA[1], PAULA MILHEIRO-OLIVEIRA[2,3],
PAULO AVILEZ-VALENTE[2,4]

[1] Sonae SGPS,
Lugar do Espido, Via Norte, 4471-909 Maia,
PORTUGAL

[2] Faculdade de Engenharia,
Universidade do Porto
Rua Dr Roberto Frias s/n, 4200-465 Porto,,
PORTUGAL

[3] Centro de Matemática da Universidade do Porto,
Rua do Campo Alegre 687, 4169-007 Porto,
PORTUGAL

[4] Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto,
Terminal de Cruzeiros do Porto de Leixões, Av. General Norton de Matos s/n, 4450-208 Matosinhos,
PORTUGAL

*Abstract:-*Water resources have become a growing concern in society. This is largely due to the scarcity of this natural asset and the realisation that increasing demand could lead to future conflicts. Sometimes, human action limits access to water or alters natural flows. Run-of-river hydropower schemes manage river flows on a short-term basis, altering the natural flow of rivers according to the energy needs of consumers or the risk of flooding. The aim of this work is to show how to model and predict the hourly flow in a run-of-river reservoir, using the Crestuma-Lever dam on the river Douro (Portugal) as a case study. Data collected from 1998 to 2020 will be used. The study focuses on the use of time series models capable of dealing with multiple periodicities, such as the TBATS model. The findings show that the model can be used for 48-hour to weekly forecasting. In general, it captures the large fluctuations in the turbine discharges and most peak discharges. However, it does not capture most zeros and has difficulty in dealing with low flow values. The results of the time-series model are also compared with those obtained using three machine learning algorithms: the Seasonal Naïve, the Neural Network, and the Random Forest.

Key-Words: Time series, Multiple periodicity, TBATS model, Forecasting, Water resources.

Received: June 17, 2024. Revised: December 16, 2024. Accepted: February 11, 2025. Published: March 18, 2025.

1 Introduction

The management of water resources plays a fundamental role in society today. Awareness of the finite nature of natural resources has led to increased interest in studies and research, leading to new developments in this field. The construction of dams is a source of hydroelectric power, producing renewable and non-polluting energy. Strategic and efficient planning of reservoirs and dam operations can reduce water losses and pollution, help prevent floods and droughts and provide adequate access to water. Adequate management of dams and reservoirs requires, at the very least, knowledge of the inflows and outflows involved, which may or may not be measured and must, therefore, be estimated based on an assumed model. On the other hand, proper management also requires forecasting inflows or outflows into the future, most often within short-term horizons. Release flows in dams depend mainly on inflows, but the dependency is complex because management decisions that condition releases use both data on electrical power demand, essentially, and optimised strategies for the use of combined energy sources. These data and strategies are kept confidential by the utilities.

The scenario that we consider in this paper is relevant to water resources management. We consider a dam for which daily data are available as time series for the inflow or for the instantaneous water level in the reservoir (see e.g., [1], [2], [3], [4], [5], for related problems). The goal of the present work is to analyse the ability to predict hourly discharges at the dam. Of particular practical interest are the one-week-ahead release forecasts. We will rely on a data-driven model to achieve this objective.

The time series of dam releases are multi-periodic, including a natural hydrological annual period, and weekly and daily periods both resulting from energy demand. At first sight, SARIMA (Seasonal Autoregressive Integrated Moving Average) models would be natural candidates for modelling the time series. However, these models have virtually no success in dealing with multiple seasonalities, especially when large periods are combined with small periods and large data sets are involved, as is the case here. Modelling and forecasting with multi-periodic time series can be challenging in terms of statistical methodology. Indeed, the literature on this subject is not very prolific. A key reference is the work of Hydman and co-workers, [6], [7], [8]; see also, [9]. These works are notable for their new approaches to the problem and their consideration of practical aspects. They propose a new time series model named TBATS (Trigonometric seasonality, Box-Cox transform, ARMA errors, Trend, and Seasonal components). The TBATS model can be seen as a generalisation of earlier state-space models for time series with multiple seasonality, incorporating ARMA error corrections and Box-Cox transformations.

To illustrate the proposed method, we consider a study case of a dam located on the Douro river in Portugal. Daily inflow and hourly outflow data are available for about two decades and are used to fit a TBATS model. The model is then used to forecast the hourly discharge of the dam over a one-week time horizon. The results are compared with those obtained using Seasonal Naïve, Neural Network and Random Forest algorithms.

2 Multi-periodicity Modelling and Forecasting. The TBATS Model

The time series of dam releases simultaneously exhibit at least daily, weekly, and annual periods; therefore modelling should take into account this multiplicity of periods. In the literature on time series modelling, one can find a diversity of proposals that can be used in our problem. As mentioned in the introduction, TBATS, a model proposed by [8], is one such model. It is a powerful tool in that it allows for multiple seasonal non-integer periods, trends, and Box-Cox transformations to induce non-linearity as well as ARMA error correction.

In this work, the time series of dam releases is denoted by Q_{OUT} and the inflow time series is denoted by Q_{IN}.

The TBATS model is described by the following set of equations:

$$Q^{(\lambda)}_{\mathrm{OUT},t} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{T} S^{(i)}_{t-m_i} + d_t \,, \tag{1}$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t \,, \tag{2}$$

$$b_t = \phi b_{t-1} + \beta d_t \,, \tag{3}$$

$$d_t = \sum_{i=1}^{p} \varphi_i d_{t-i} + \sum_{i=1}^{q} \theta_i e_{t-i} + e_t \,, \tag{4}$$

where $Q^{(\lambda)}_{\mathrm{OUT},t}$ is the time series $Q_{\mathrm{OUT},t}$ after a possible Box-Cox transformation with parameter λ, $S^{(i)}_t$ is the seasonal component of the series with period m_i, l_t consists of the so-called local level of the series, b_t represents the (stochastic) trend of the series and d_t is the associated ARMA(p, q) process. The parameters ϕ, α and β must be estimated when fitting the model to the available data, here constituted by the observed time series Q_{OUT}. In this model, each seasonal component $S^{(i)}$ appearing in equation (1) is modelled by a trigonometric representation based on Fourier series, although the remaining components are allowed a more elaborate stochastic structure than in a simple ARMA or ARIMA model with a Fourier term (see, [10], [11], for application examples). TBATS models are flexible enough to allow the seasonal components to change slowly over time. Additionally, by incorporating ARMA errors, these models can effectively capture existing autocorrelations in the time series data. TBATS models can accommodate linear as well as nonlinear trends. The R language $tbats()$ function can be used to perform the necessary computations, [12].

2.1 Stationarity of the Residuals

As usual in times series modelling, it is necessary to analyse the stationarity of the residuals and to carry out a diagnostic step. For this analysis, we use the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, [13], and the Augmented Dickey-Fuller (ADF) test. The model fit assumption of non-autocorrelation of the residuals is also assessed using the Ljung-Box test. The usual significance level of 5 % is used in the testing procedures.

2.2 Error Metrics

The quality of the predictions provided by the different models can be assessed by computing some error metrics. The Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Error (ME) metrics will be used in the present work. They are defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}\,, \tag{5}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|\ , \tag{6}$$

$$\text{ME} = \frac{1}{n}\sum_{i=1}^{n}y_i - \hat{y}_i\,, \tag{7}$$

where y_i is the i-th measured value and $\hat{y}_i$ is the corresponding forecast, and n is the length of the time series considered in the computation. Recall that these metrics, when applied to the training set, allow the goodness of fit to be assessed. When applied to the test set, they assess the ability of the model to produce forecasts, [14].

3 Case study: Crestuma-Lever Dam

We are particularly interested in the forecast of the hourly discharge of the Crestuma-Lever dam on the river Douro in Portugal.

3.1 Study area and Available Data

The basin of river Douro covers an area of 97603 km^2, of which 18643 km^2 are in Portugal (Fig. 1). The estuary of the river Douro is 21 km long and is limited upstream by the Crestuma-Lever dam (Fig. 2). This dam houses a run-of-river hydroelectric power station with a useful capacity of 22.5×10^6 m^3 and fully controls the flow of the river into the estuary. Its turbines and weir gates can be operated simultaneously, and the instantaneous release of water from the dam is $Q_{\text{OUT}} = Q_{\text{HP}} + Q_{\text{W}}$, where Q_{HP} is the turbine discharge, and Q_{W} is the flow rate through the weirs.

In Portugal, the hydrological year begins on 1 October and ends on 30 September the following year. Reservoir inflow varies throughout the year and is highly correlated with rainfall, with large intra- and inter-annual variability. The hydrological year has a wet period from October to April, when the highest flows are usually recorded, and a dry period from May to September, when flows are usually lower. The inter-annual variability is also considerable, leading to the existence of wet hydrological years with a large annual inflow and dry hydrological years with reduced annual inflows.

Two sets of data are available, from SNIRH (Sistema Nacional de Informação de Recursos Hídricos) and from EDP (Energias de Portugal), from the year 1998 to the year 2020, although the second set of data is not published. The following time series are available: average hourly dam releases, Q_{OUT}; average daily reservoir inflows, Q_{IN}; average hourly weir releases, Q_{W}; average hourly hydropower releases, Q_{HP}. These data start on 13 January 1998 and end on 30 September 2020. Additionally, the following time series have been provided by SNIRH from September 1998 to March 2020: cumulative (volumetric) reservoir inflow, cumulative (volumetric) dam release, instantaneous water volume in the reservoir, instantaneous water level in the reservoir, z_{R}. Table 1 summarises general information on the relevant available data. The data were first checked for gaps, consistency (using the volume and water level data) and outliers. The analysis dictated the exclusion of data prior to 3 May 2001 in the modelling procedure.

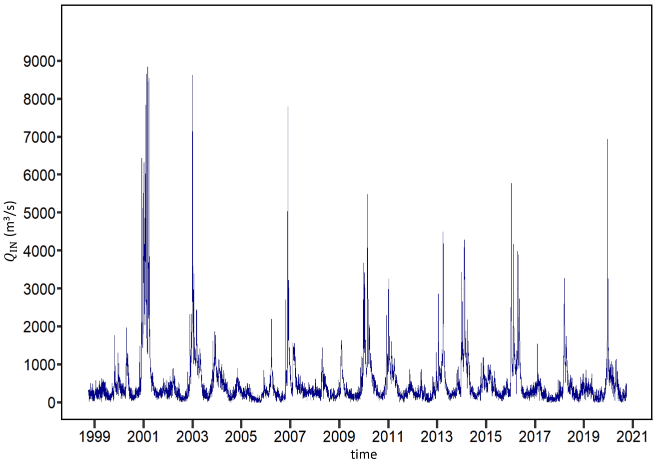
In the present study, the dataset recorded from 4 May 2001 to 30 Sep 2019 is used as a training dataset, while the dataset recorded in the hydrological year of 2019/20, i.e. from 1 Oct 2019 to 30 Sep 2020,

Fig. 1: The basin of the river Douro in the Iberian Peninsula

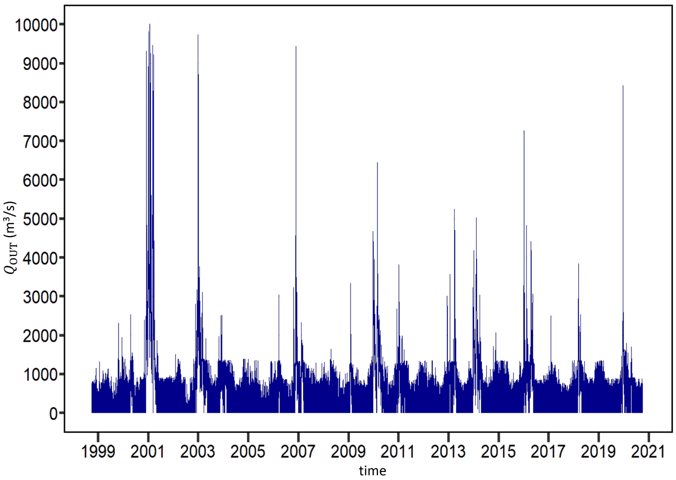
Fig. 2: Upstream view of the Crestuma-Lever dam (photo by Francisco Piqueiro)

Table 1. Structure of the database

Time series	Source	Type of data
Q_{OUT}	EDP	hourly
Q_{W}	EDP	hourly
Q_{HP}	EDP	hourly
Q_{IN}	SNIRH	daily



(a) Reservoir inflow.



(b) Dam release.

Fig. 3: Inflow and outflow chronograms from 1998/99 to 2019/20

is used as a test dataset. The chronograms of the time series of reservoir inflow Q_{IN} (daily data) and reservoir outflow Q_{OUT} (hourly data) are shown in Fig. 3. There is clearly a relationship between the two time series.

In fact, the hourly discharge of the dam depends, on the one hand, on the reservoir inflow and, on the other hand, on the decisions taken by the hydropower plant manager, which can be seen mainly in the hourly hydropower release time series, which is usually kept confidential. Fig. 4 illustrates the variability of the hydropower release time series along one full hydrological year and among years. Two instances of the time series of hydropower release along two full hydrological years are shown in the figure.

On the other hand, since our aim in this paper is to forecast the hourly discharge of the dam and this depends on the reservoir inflow, the daily time series Q_{IN} was disaggregated into an hourly time series. Several disaggregation methods were tried but without success due to unresolved physical inconsistencies in the results. Therefore, a linear interpolation method was used to perform the disaggregation.

Table 2. Estimates of parameters for the TBATS model (1)–(4)

TBATS		ARMA	
$\hat{\phi}$	-8.6×10^{-2}	$\hat{\theta}_1$	$+1.14$
$\hat{\alpha}_1$	$+3.8 \times 10^{-3}$	$\hat{\theta}_2$	-0.35
$\hat{\alpha}_2$	$+8.7 \times 10^{-4}$	$\hat{\gamma}_1$	-0.21
$\hat{\alpha}_3$	-2.0×10^{-3}	$\hat{\gamma}_2$	-0.03
$\hat{\beta}_1$	-1.0×10^{-3}		
$\hat{\beta}_2$	$+1.3 \times 10^{-3}$		
$\hat{\beta}_3$	$+9.5 \times 10^{-4}$		

Table 3. Basic statistics of the training dataset and fitted model

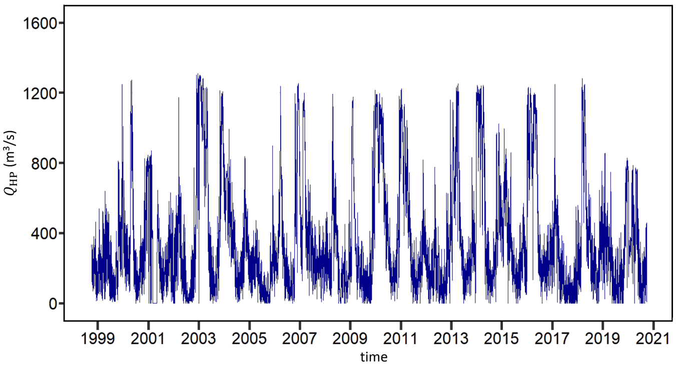
	Training dataset	TBATS model (1)–(4)
mean	447	481
median	257	484
standard deviation	632	276
max	9727	1141
min	0	0
1st quartil	0	264
3rd quartil	695	690
5 % quantil	0	0
10 % quantil	0	0
90 % quantil	1159	1133
95 % quantil	1367	1378

3.2 Results

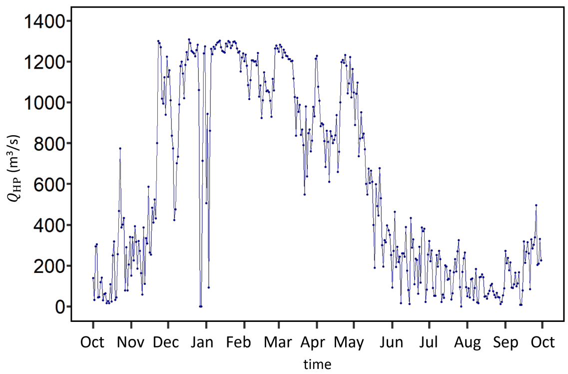
A $\text{TBATS}(1, 2, 2, -, \langle 24, 4\rangle, \langle 168, 5\rangle, \langle 8766, 5\rangle)$ model was fitted to the training dataset. Within the class of TBATS models, this model provided the best fit. No Box-Cox transform needs to be applied (i.e. parameter λ equals 1), the autoregressive component is an $\text{ARMA}(2, 2)$, and three seasonal terms were included: a daily period (24 h) with 4 harmonics, a weekly period (168 h) with 5 harmonics, and an annual period (8766 h) with 5 harmonics. The estimates of the parameters obtained are shown in Table 2.

The residuals obtained in the model fitting present some challenges as they cannot be strictly considered as random white noise (see Fig. 5), but no bias is detected. The AIC is 3550918. The basic statistics are presented in Table 3.

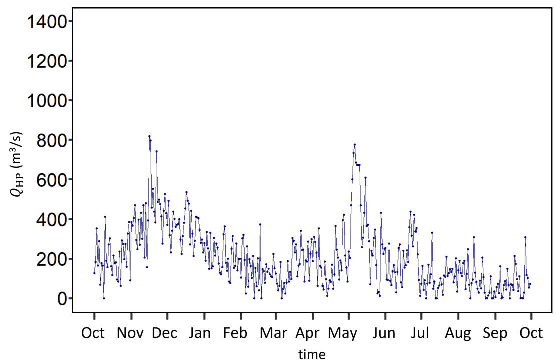
Short-term forecasts one week ahead have been made. As mentioned above, the data for the



(a) From 1998/99 to 2019/20.

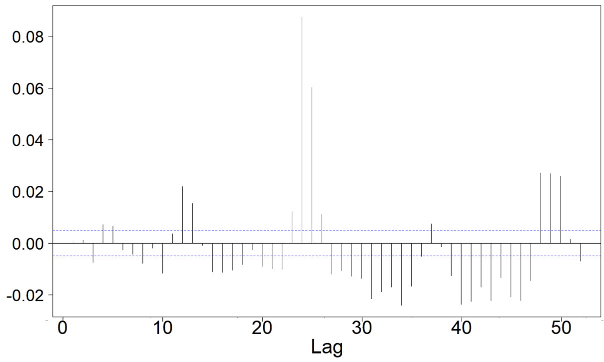


(b) Hydrological year 2002/03.

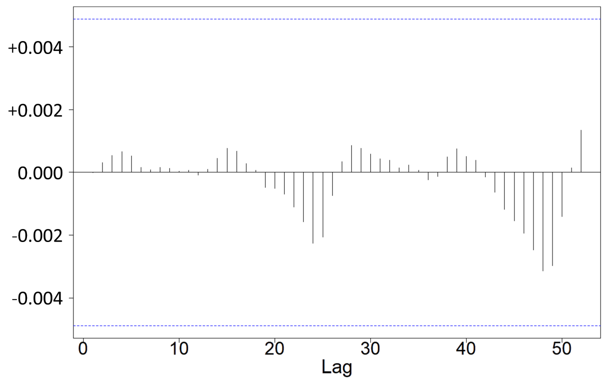


(c) Hydrological year 2011/12.

Fig. 4: Chronograms of turbine flow release

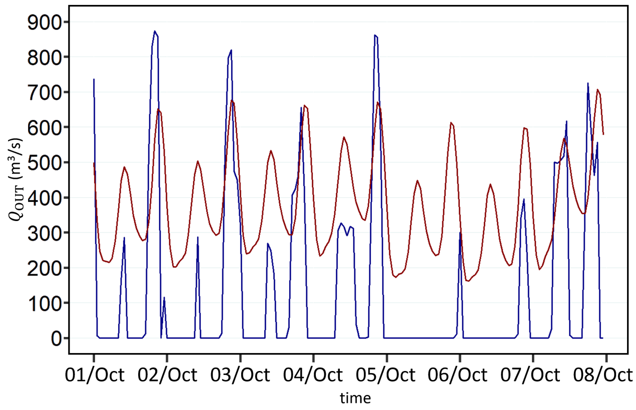


(a) ACF.

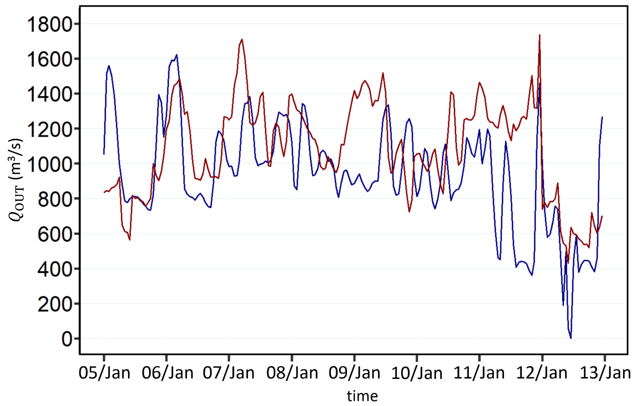


(b) PACF.

Fig. 5: ACF and PACF of the residuals (model (1)–(4))



(a) 1–7 Oct 2019.



(b) 5–12 Jan 2020.

Fig. 6: Dam release forecasts

hydrological year 2019/20 (1 October 2019 to 30 September 2020) are used as a test dataset. The ME, RMSE and MAE metrics for these forecasts are presented in Table 4. As an example, the week from 1 to 7 October 2019 (the first week of the test dataset) is analysed in more detail here. Fig. 6 shows the dam release forecasts for this first week, as well as the forecasts for a week in January, which is generally characterised by larger flows. The predictions are shown in red, and the observed data (test dataset) are shown in blue. Table 4 shows the forecast metrics for the week from 1 to 7 October 2019 as well as the metrics for the entire 2019/20 hydrological year.

We have found that the predictors may produce negative forecasts at some points in time when very low flows occurred. As such values are not physically possible, these predictions have been replaced by zeros. Such a procedure introduces some unwanted bias.

We found that the TBATS model is generally not good at tracking fast decreases in discharge values and most zeros, and exhibits significantly poor forecast metrics (see, for instance, the forecast error results

Table 4. Forecast metrics (in m^3 /s) for TBATS, Seasonal Naïve, Neural Network and Random Forest models

Model	ME	RMSE	MAE
Forecast	01 Oct 2019 to 07 Oct 2019		
TBATS	-243	310	282
Seasonal Naïve	16.8	281	165
Neural Network	-103	228	174
Random Forest	-39	320	268
Forecast	01 Oct 2019 to 30 Sep 2020		
TBATS	41.1	679	345
Seasonal Naïve	270	745	398
Neural Network	306	810	473
Random Forest	58	685	240

for the first week of October 2019 in Table 4). However, similar conclusions were reached when a Seasonal Naïve model (see, [6], for model details) was applied. The Seasonal Naïve model clearly showed its inability to predict the hourly discharge, which is not surprising as it is not supposed to deal with time series with multiple periodicities. Predictions based on a Random Forest algorithm (refer to the *randomForest* library in the *R* software) considering Q_{OUT} as explained by Q_{IN}, as well as hour of the day, day of the week, month and year, showed no improvement, even when different sets of years were used in the training phase. The same happened when using the Neural Network algorithm implemented in the function *nnetar*, also in the *R* software. In the latter case, the entire training dataset was used, as this Neural Network clearly gave the lowest prediction errors. Despite of the considerable computational effort required to run these algorithms, the resulting errors were not significantly reduced.

4 Final Comments

The TBATS model can be used to produce reasonable quality hourly forecasts of dam releases based on daily reservoir inflow data. Specifically, it demonstrates the ability to track the peaks and large variations in hourly dam releases.

Future research should explore alternative time series models that incorporate multiple periodicities and that may be of non-linear nature, as well as other machine learning algorithms applied to the problem of forecasting dam releases. Some investment in methodologies that address the disaggregation of daily inflow data into hourly data will also be required.

Acknowledgment:

The authors would like to thank Dr Isabel Iglesias (CIIMAR, Portugal) for her help, Dr Francisco Piqueiro for providing the photograph of the dam, and Energias de Portugal (EDP) for providing part of the data used in this research. The first author carried out much of the research as a student at the Faculty of Sciences of the University of Porto, Portugal, and is therefore grateful for this research opportunity.

References:

[1] Sharifi, F., Haddad, O.B., Naderi, M., Alimohammadi, S.: Continuous decision making in optimal reservoir operation using dp-ann. WSEAS Transactions on Information Science and Applications **2**(5), 630 – 636 (2005)

[2] Feloni, E., Baltas, E.: Impact of population growth and climate change on the resilience of florina's hydrosystem. WSEAS Transactions on Environment and Development **16**, 132 – 140 (2020). DOI 10.37394/232015.2020.16.13

[3] Peng, F., Shi, X., Li, K., Wang, Y., Feng, J., Li, R., Liang, R.: How to comprehensively evaluate river discharge under the influence of a dam. Ecological Informatics **69**, 101,637 (2022). DOI https://doi.org/10.1016/j.ecoinf.2022.101637

[4] Nikas-Nasioulis, I., Bertsiou, M.M., Baltas, E.: Investigation of energy, water, and electromobility through the development of a hybrid renewable energy system on the island of kos. WSEAS Transactions on Environment and Development **18**, 543 – 554 (2022). DOI 10.37394/232015.2022.18.53

[5] Nisha, C.M., Thangarasu, N.: A comparative analysis of deep-learning models for dam discharge forecasting for disaster management. Water Supply **24**(12), 4024–4045 (2024). DOI 10.2166/ws.2024.248

[6] Hyndman, R.J., Athanasopoulos, G.: Forecasting: Principles and Practice. OTexts (2018)

[7] Hyndman, R., Khandakar, Y.: Automatic time series forecasting: the forecast package for R. Journal of Statistical Software **27**(1), 1–22 (2008)

[8] De Livera, A., Hyndman, R., Snyder, R.: Forecasting time series with complex seasonal patterns using exponential smoothing. Journal of the American Statistical Association **106**(496), 1513–1527 (2011)

[9] Hurd, H., Pipiras, V.: Modeling periodic autoregressive time series with multiple periodic effects. In: F. Chaari, J. Leskow, R. Zimroz, A. Wyłomańska, A. Dudek (eds.) Cyclostationarity: Theory and Methods – IV, pp. 1–18. Springer International Publishing (2020)

[10] Godahewa, R., Bergmeir, C., Webb, G.I., Montero-Manso, P.: An accurate and fully-automated ensemble model for weekly time series forecasting. International Journal of Forecasting **39**(2), 641–658 (2023). DOI https://doi.org/10.1016/j.ijforecast.2022.01.008

[11] Karabiber, O., Xydis, G.: Electricity price forecasting in the Danish day-ahead market using the TBATS, ANN and ARIMA methods. Energies **12**(5), 928 (2019)

[12] R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2017). URL https://www.R-project.org/

[13] Kwiatkowski, D., Phillips, P.C., Schmidt, P., Shin, Y.: Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? Journal of Econometrics **54**(1-3), 159–178 (1992)

[14] Lee, Y., Tay, K., Choy, Y.: Forecasting electricity consumption using time series model. International Journal of Engineering and Technology **7**(4), 218–223 (2018)

Contribution of Individual Authors to the Creation of this Scientific Article (Ghostwriting Policy)

The 1st author contributed with data curation, implementation of methodologies, data visualisation and software code. The 2nd author contributed with the conceptualisation, methodology and validation. The 3rd author contributed with validation, analysis and interpretation of results. The 2nd and 3rd authors contributed with the supervision and writing of the manuscript. All authors contributed with references to the state of the art and to the discussion.

Sources of Funding

This research was partially supported by Strategic Funding (UIDB/00144/2020 and UIDP/00144/2020) through national funds made available by the FCT—Foundation for Science and Technology—and the European Regional Development Fund (ERDF) under the Portugal 2020 partnership agreement, and by the EsCo-Ensembles project (PTDC/ECI-EGC/30877/2017), co-funded by NORTE 2020, Portugal 2020 and the European Union through the ERDF, and by FCT through national funds.

Conflicts of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.