Larhyss Journal, ISSN 1112-3680, n°52, Dec 2022, pp. 49-59
© 2022 All rights reserved, Legal Deposit 1266-2002

STOCHASTIC ALTERNATIVES TO MODELS BASED ON THE DALTON'S APPROACH FOR ESTIMATING MONTHLY PAN EVAPORATION FROM RESERVOIR DAMS IN ALGERIA

***BENCHAIBA L.*[1*]*, HOUICHI L.*[2]*, AMARCHI H.*[1]**

[1] Lecturer, Department of Hydraulics, University of Annaba, Algeria
[2] Professor, Department of Hydraulics, University of Batna 2, Algeria

(*) benchaiba@univ-batna2.dz

Research Article – Available at http://larhyss.net/ojs/index.php/larhyss/index
Received July 6, 2022, Received in revised form November 23, 2022, Accepted November 26, 2022

ABSTRACT

In this contribution, we assess the capabilities of monthly pan evaporation estimation using two methods: i) the first is a purely stochastic model designed by ARIMA (autoregressive with integrated moving average), which is based only on endogenous variables, and ii) the second is derived from Dalton's physical approach based on exogenous variables. The areas of application of the two methods are two sites of two dams, Ain Zada and Beni Haroun, with distinct climates of Algeria: semiarid and subhumid. The actual evaporations are measured directly from the pans for seven years. The first six years of observations are used for the calibration of the two models, and the last year is reserved for the test. The performance of the models is evaluated using three commonly used metrics, namely, the correlation coefficient (R), root mean square error (RMSE) and mean absolute percentage error (MAPE). According to the results of this research, it appears that ARIMA models perform better than models derived from Dalton's approach with significantly improved predictive capabilities.

Keywords: ARIMA, pan evaporation, Dalton's approach, monthly, climate.

INTRODUCTION

Estimating the volumes of water evaporated from reservoir dams is of particular importance in water resource management. The two main factors influencing evaporation from an open water surface are the energy supply to provide the latent heat of vaporization and the ability to transport the vapor away from the evaporating surface (Chow et al., 1988). Solar radiation (temperature) is the primary source of thermal energy. The ability

---

to transport vapor away from the evaporating surface depends on the wind speed over the surface and the specific humidity gradient in the air over it (Chow et al., 1988). The amount of evaporation from a water surface is obtained by two methods (Smith, 1909): 1) direct measurements on properly exposed water surfaces, such as the Class A and Colorado evaporation pans. These direct measurements allow the use of models with endogenous variables. 2) Calculation based on the temperature of the water surface and the value of certain meteorological elements. This calculation allows the use of models with exogenous variables. For this second method of determining evaporation in large reservoirs and lakes, the principles of the Dalton formula are used (Dalton, 1802). Indirect methods require reliable data. Otherwise, they suffer from major uncertainties identified by many authors cited in (Alazard et al., 2015).

Knowledge of the quantities of water evaporated from reservoir dams is dependent on the estimation of evaporation from the pans. These quantities contribute directly to the water balance. However, evaporation from pans is greater than that from large lakes (Chow et al., 1988). For this reason, adjustment coefficients less than unity are recommended (Linsey et al., 1982). Records of evaporation from exposed water surfaces (Class A and Colorado pans) are made with great care at Algerian reservoirs under the influence of different climates.

The present work aims to evaluate the capabilities of monthly pan evaporation estimation using two methods: 1) a purely stochastic model designed by ARIMA (autoregressive with integrated moving average), which is based only on endogenous variables, and 2) a physical approach based on exogenous variables derived from Dalton's physical approach as achieved in (Boutoutaou et al., 2020). The two approaches are being used at two dam sites, Ain Zada and Beni Haroun, which have different climates in Algeria: semiarid and subhumid.

MATERIALS AND METHODS

Study area

Ain Zada Dam (semiarid climate)

The Ain Zada dam is made of embankment and rockfill with a central clay core (Figure 1). It was put into service in 1986 on the wadi Boussellam, which made an outlet of the subbasin initially intended for irrigation, bypassed for the benefit of the cities of Sétif, Bougaa, El Eulma and Bordj Bou Arreridj with a capacity of 125 Mm^3 (Mebarkia, 2011). The site of the Ain Zada dam is located both on the wilayas of Sétif and Bordj Bou Arreridj, 40 km north of the capital of the wilaya of Bordj Bou Arreridj and 25 km west of the capital of the wilaya of Sétif. It is located 11 km northeast of the village Ain-Taghrout on the Oued Boussellam. The watershed draining the entire site covers 2080 km^2 (LEM, 2013). The Ain Zada dam is taken for the case of a semiarid climate where the average annual rainfall is approximately 350 mm.

Beni Haroun dam (subhumid climate)

The Beni Haroun compacted concrete dam is Algeria's major strategic hydraulic structure (Figure 2). The total volume of its reservoir is 997 Mm^3, which allows the regulation of approximately 435 Mm^3 per year. The site of the dam is located in Mila Wilaya in eastern Algeria on Oued El Kebir. The reservoir generated by the dam is located to the south of the dam. It is located approximately 40 km northwest of the city of Constantine and 350 km east of Algiers. The total basin of Oued Kebir at the site of the dam has an area of 7725 km². Considering that part of the Rhumel basin is mobilized by the dam, the relative area of the Beni Haroun dam is approximately 6595 km² (LEM, 2013). The dam provides drinking water to the cities of Mila, Constantine, Jijel, Oum-El-Bouaghi, Khenchela, and Batna as well as the irrigation of orchards in Mila. The Beni Haroun Dam is taken for the case of a subhumid climate with an average annual rainfall of approximately 550 mm.

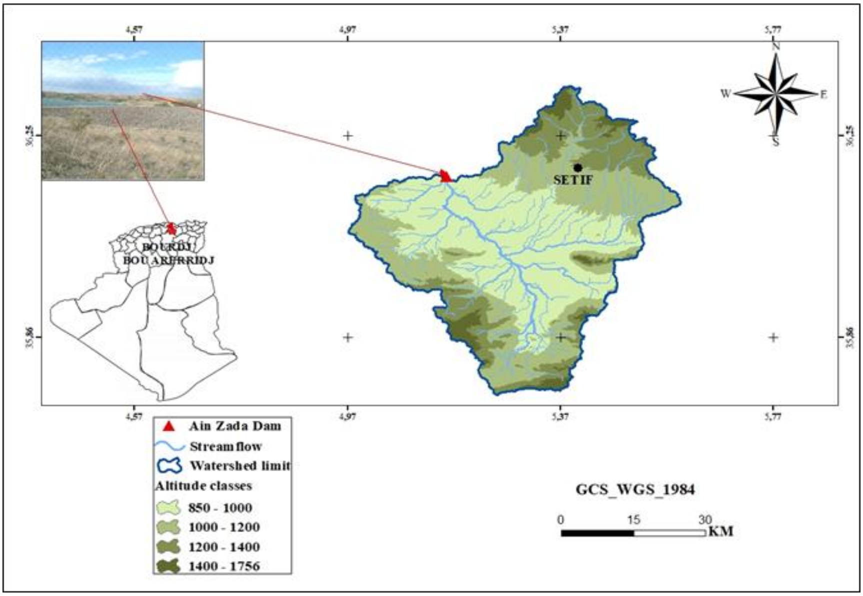

Figure 1: Geographical location of the Ain Zada Dam (semiarid case)

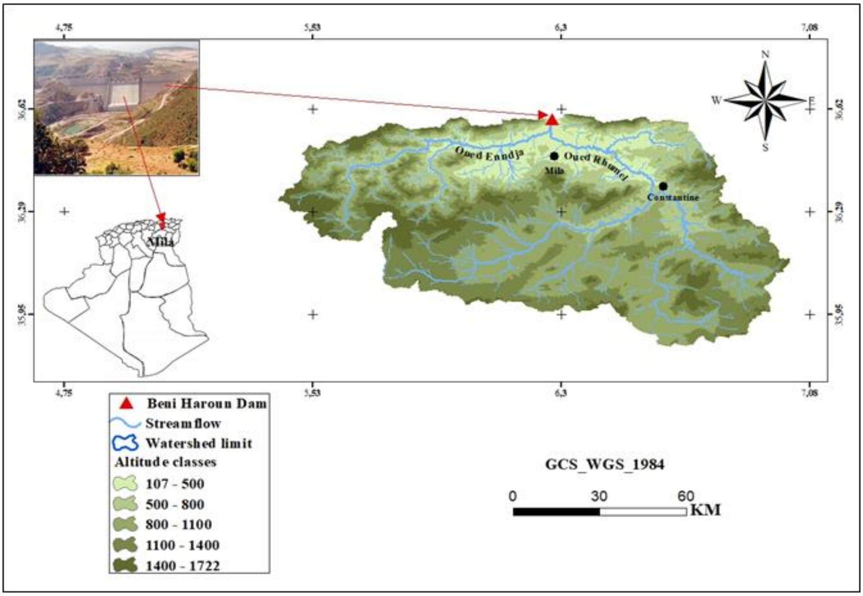

Figure 2: Geographical location of the Beni Haroun dam (subhumid case)

Data

The database used was collected from the National Agency for Dams and Transfers (ANBT). It consists of a time series of climatic data (temperature, air humidity, wind speed) and observations of the evaporation of the water body from pan class A or Colorado.

The period at the monthly scale, thus serving as a basis for the rest of this contribution, is spread over a common period from 01/01/2009 to 31/12/2015, i.e., seven years of observation for the two dams Ain Zada and Beni Haroun. The sizes of the two series are identical (84 months). The numerical values of the monthly pan evaporations at the sites of the two dams vary from 19.60 to 330.20 mm and 30.20 to 376.31 mm for Ain Zada and Beni Haroun, respectively. For the models considered in this study, the training phase is spread over the first six years, and the test phase concerns the last year for each case studied separately.

Evaluated models

Stochastic model ARIMA type

One of the most popular means of stochastic modeling is the autoregressive integrated moving average (ARIMA) introduced by Box and Jenkins to forecast time series (Box and Jenkins, 1976). The ARIMA or Box–Jenkins models are relatively easy to implement (Hyndman and Khandakar, 2008). A seasonal ARIMA model is denoted ARIMA (autoregressive integrated moving average, (*p, d, q*) (*P, D, Q*)*s*, and the data depend on the previous values (not seasonal part) and on the values for the same period of previous years (seasonal part), where s is the seasonal period, which is equal to 12 for the monthly case. An ARIMA (*p*, *d*, *q*) (nonseasonal part) model can take into account time dependence in several ways. First, the time series is differentiated to make it stationary. If d = 0, the observations are modeled directly, and if d = 1, the differences between consecutive observations are modeled. Second, the time dependence of the stationary process y_t is modeled by including p autoregressive models. Third, q are terms of the moving average models. The process supports the observation of previous errors. Finally, by combining these three models, we obtain the ARIMA model. Thus, the general form of ARIMA models is given by:

$$\mathrm{y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}} \tag{1}$$

where y_t and y_{t-i} are stationary stochastic processes at time t and t-i, c is the constant that determines the level of the time series, ε_t is the error or white noise term, ϕ_i are the autoregressive coefficients and θ_j are the moving average coefficients.

For a seasonal part, these steps can be repeated according to the cycle period, regardless of the time interval. The process is the same for the seasonal part as for $D \neq 0$.

Models derived from Dalton's approach

In the contribution of (Boutoutaou et al., 2020), the authors proposed the following formulas, which are based on Dalton's law (Dalton, 1802), where pan evaporation (class A or Colorado) is determined from the air saturation deficit denoted by D.

For the arid and semiarid regions of Algeria:

$$E = 0.403 n D^{0.73}(1 + 0.39 V_2) \tag{2}$$

For the humid and subhumid regions of Algeria:

$$E = 0.342 n D^{0.8}(1 + 0.39 V_2) \tag{3}$$

where:

E: pan evaporation (mm),

n: number of days in the considered month (n = 30 or 31 depending on the month, with *n*= 28.5 for February),

V_2: average monthly wind speed in m/s at 2 m from the ground,

D: air saturation deficit in millibars (mb), given by the following relationship:

$$D = 0.0632(100 - H)e^{0.0632t} \tag{4}$$

H: average monthly air humidity (%),

t: average monthly air temperature (°C).

Performance indices

Various statistical measures have been developed and used in the literature. To assess the fit and predictive accuracy of the models, in this contribution, the datasets were mathematically evaluated by calculating and limiting the following three performance criteria: the correlation coefficient (*R*), the root mean square error (*RMSE*), and the mean absolute error in percentage (*MAPE*) expressed by the following equations:

$$R = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(M_i - M_m\right)\left(P_i - P_m\right)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(M_i - M_m\right)^2}\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(P_i - P_m\right)^2}} \tag{5}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i - P_i)^2} \tag{6}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{(M_i - P_i\ )}{M_i}\right| * 100 \tag{7}$$

where *N* is the number of data points (size of the time series), M_i are the measured values and P_i are the corresponding predicted values. M_m and P_m are the average values of M_i and $P_{i,}$ respectively.

One of the most frequently used metrics to evaluate the accuracy of the model's predictions is MAPE, which is the mean absolute error in percent. This performance indicator is easy to interpret. For example, a x% value of MAPE means that the average variance between the predicted and measured values is x%. The MAPE metric is adopted for the final selection of the best model in this study.

The best ARIMA forecasting models are those that meet the most recognized quality criterion known as the Akaike information criterion (*AIC*) (Akaike, 1974). This last criterion leads to the choice of the model with the smallest mean square error by applying a penalty that depends on the quantity of unknown parameters that must be estimated. This criterion favors parsimonious models and is calculated by the following relationship:

$$AIC = 2k - 2\log(L) \tag{8}$$

where *k* is the number of coefficients estimated in the model and *L* denotes the maximum value of the likelihood function for the model.

RESULTS AND INTERPRETATIONS

Ain Zada Dam (semiarid climate)

The assessment of the monthly pan evaporation at the site of the Ain Zada dam using ARIMA models needs to split its time series into a training phase (from January 1, 2009, to December 31, 2015) and a test phase (from January 1, 2016, to December 31, 2016). The function auto.arima (R Core Team, 2021) compares several models and gives the one that minimizes the *AIC*. The application of the candidate models gives the optimal result designated by ARIMA (1,0,0) (2,1,0) *(12)*. With the same data in the training and test phases, Eq. (2) (Boutoutaou et al., 2020), deduced from the physical formulation of Dalton for the evaporation of pans in semiarid areas of Algeria (case of the Ain Zada dam), was applied. The results of the temporal evolutions of the observed and predicted evaporations are illustrated graphically in Figure 3. Figure 3 shows that the ARIMA model and the model derived from Dalton reproduce the patterns of real values at the monthly scale. However, the ARIMA model faithfully adjusts the reality for better performance than the model derived from Dalton's approach according to Eq. (2). The results summarized in Table 1 affirm this finding in terms of numerical values of the criteria (*R*, *RMSE*, and *MAPE*) for the stochastic ARIMA model. These criteria far exceed those of the Dalton-derived model. Taking into account the *MAPE* value, the predictive capacity of the ARIMA model is then recommended in this case study (*MAPE* = 16.34%), which concerned a region with a semiarid climate in Algeria.

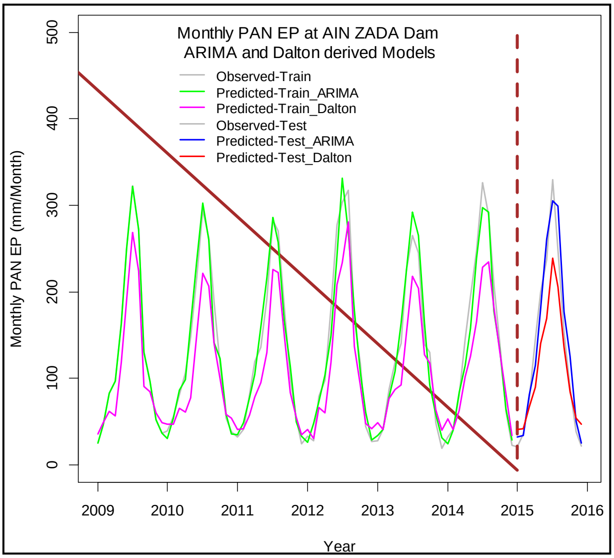

Figure 3: Monthly pan evaporations evaluated by the stochastic ARIMA model and the model derived from Dalton's approach compared with the observed values at the Ain Zada Dam

Table 1: Results of the application of the stochastic ARIMA model and the model derived from Dalton's approach for forecasting monthly pan evaporation in the training and test phases at the Ain Zada Dam

Models	**Stochastic model ARIMA (1,0,0)(2,1,0)*(12)***		**Model derived from the Dalton's approach**	
Performances	Training	Test	Training	Test
R (-)	0.98	0.97	0.98	0.98
RMSE (mm)	16.81	24.44	39.69	43.84
MAPE (%)	10.83	16.34	25.79	31.57

Beni Haroun dam (subhumid climate)

Evaluation of monthly pan evaporation at the Beni Haroun dam site with ARIMA models required its series to be divided into a learning phase (from January 1, 2009, to December 31, 2015) and a test period (from January 1, 2016, to December 31, 2016). The function auto.arima (R Core Team, 2021) compares several models and gives the one that minimizes the *AIC*. The application of the competitor models gives in the final result the optimal result designated by ARIMA(0,1,0)(0,1,2)*(12)*. With the same data in the training and test phases, Eq. (3) (Boutoutaou et al., 2020), deduced from the physical formulation of Dalton for the evaporation of pans in semiarid areas of Algeria (case of the Beni Haroun dam), was applied. The results of the temporal evolutions of the observed and predicted evaporations are illustrated graphically in Figure 4. Figure 4 shows that the ARIMA model and the model derived from Dalton reproduce the patterns of real values. However, the ARIMA model adjusts the reality for improved performance compared to the model derived from Dalton's approach according to Eq. (3). The results in Table 2 support this outcome in terms of numerical values of the criteria (*R*, *RMSE*, and *MAPE*) for the stochastic ARIMA model. These criteria far exceed those of the Dalton-derived model. Taking into account the *MAPE* value, the predictive capacity of the ARIMA model is also suggested in this second case study for *MAPE* = 21.84%, which is related to a subhumid climate region of Algeria.

Table 2: Results of the application of the stochastic ARIMA model and the model derived from Dalton's approach for forecasting monthly pan evaporation in the training and test phases at the Beni Haroun Dam

Models	**Stochastic model ARIMA (0,1,0)(0,1,2)*(12)***		**Model derived from the Dalton's approach**	
Performances	Training	Test	Training	Test
R (-)	0.96	0.95	0.80	0.77
RMSE (mm)	32.27	37.59	84.72	95.62
MAPE (%)	19.15	21.84	37.98	71.69

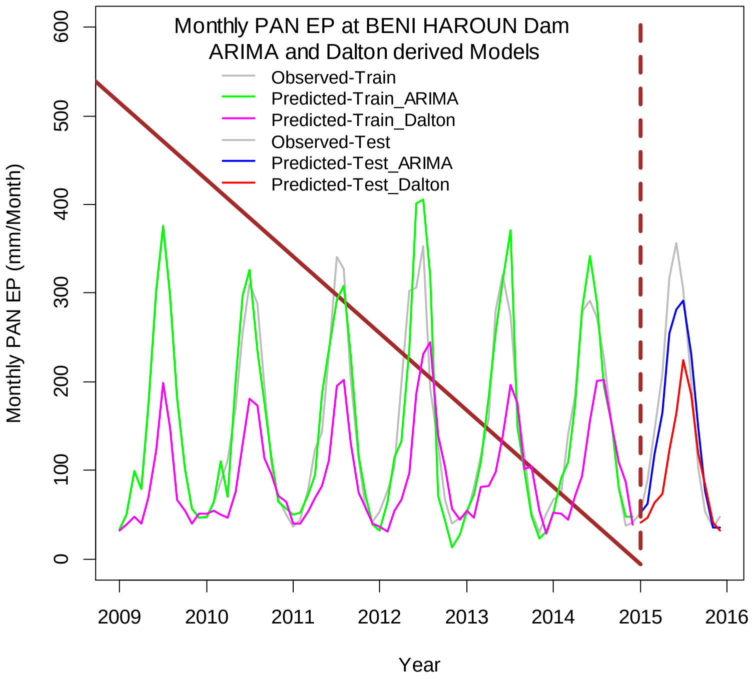

Figure 4: Monthly pan evaporation evaluated by the stochastic ARIMA model and the model derived from Dalton's approach compared with the observed values at Beni Haroun Dam

CONCLUSION

In the present study, the search for a more efficient stochastic alternative to predict the monthly pan evaporation was completed within two reservoir dams located in distinct climatic regions in Algeria: the Ain Zada dam (for a semiarid zone) and Beni Haroun dam (for a subhumid zone). The alternative model evaluated ranks among the well-explored linear stochastic types in the literature. The physical model to be replaced is derived from Dalton's theoretical formulation according to two equations proposed by Boutoutaou et al. (2020), namely, Eq. (2) for the semiarid case and Eq. (3) for the subhumid case.

The time series of 84 months (07 years) is analyzed with a single splitting, which exploits the first 72 months (6 years) in the training phase of the two models, and the last 12 months will be used for the testing phase of the models. The evaluation results of the two considered models demonstrate that the ARIMA models adequately reproduce the variability of the monthly pan evaporation for the two configurations (semiarid and subhumid). It follows from this research that ARIMA models are more efficient than models derived from Dalton's approach with much higher predictive capabilities in the test phase: $MAPE_{ARIMA}$ = 16.34%, $MAPE_{DALTON}$ = 31.57%, and $MAPE_{ARIMA}$ = 37.98%, $MAPE_{DALTON}$ = 71.69% for the semiarid case in the Ain Zada dam and the subhumid case in the Beni Haroun dam. Consequently, stochastic ARIMA models presented as alternatives are highly recommended to overcome the limits of the predictive capabilities of approaches derived from physical formulations, as established in the work of (Boutoutaou et al., 2020).

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.

REFERENCES

AKAIKE H. (1974). A New-Look at the Statistical Model Identification. IEEE Transactions on Automatic Control, AC- 19, Vol. 19, Issue 6, pp. 716-723.

ALAZARD M., LEDUC C., TRAVI Y., BOULET G., SALEM A.B. (2015). Estimating evaporation in semiarid areas facing data scarcity: Example of the El Haouareb dam (Merguellil catchment, Central Tunisia), Journal of Hydrology: Regional Studies, Vol. 3, pp. 265-284.

BOUTOUTAOU D., ZEGGANE H., SAGGAI S. (2020). Evaporation from the water surface of lakes and reservoirs of the arid zone of the Mediterranean: case of Algeria, LARHYSS Journal, No 43, pp. 91-101.

BOX G.E., JENKINS G.M. (1976). Time Series analysis: forecasting and control, Revised Edition, Holden-Day, San Francisco, USA, 575p.

CHOW V.T., MAIDMENT D.R., MAYS L.W. (1988). Applied Hydrology. McGraw-Hill, New York, USA.

DALTON J. (1802). Experimental essays on the constitution of mixed gases; on the force of steam or vapor from water and other liquids at different temperatures, both in a Torricellian vacuum and in air; on evaporation; and on the expansion of gases by heat. Mem. Lit. Philos. Soc. Manch, Vol. 5, Issue 11, pp. 535-602.

HYNDMAN R.J., KHANDAKAR Y. (2008). Automatic time series forecasting: the forecast package for R. Journal of Statistical Software, Vol. 26, Issue 3, pp. 1-2.

LINSLEY R.K., KOHLER M.A., PAULHUS J.L. (1982). Hydrology for engineers, McGraw-Hill, New York, USA.

MARITIME STUDIES LABORATORY (LEM). (2013). Bathymetric surveys of eleven dams in operation: Béni Haroun measurement campaign, Report December 2013, Djenane El Malik, Hydra, Algiers, In French.

MEBARKIA A., BOUFEKANE A. (2020). Human activity impact on surface water quality in semiarid regions: a case study of Aïn Zada Lake (North–East Algeria). Water Supply, Vol. 20, Issue 5, pp. 1726-1744.

SOFTWARE R (2021). Language and environment for statistical computing, Available at: https://www.R-project.org/