

Original Research Paper

Modeling Pan Evaporation for Kuwait using Multiple Linear Regression and Time-Series Techniques

Jaber Almedejj

Department of Civil Engineering, Kuwait University, P.O. Box 5969, Safat 13060, Kuwait

Article history

Received: 03-04-2016

Revised: 19-04-2016

Accepted: 30-05-2016

Email: j.almedejj@ku.edu.kw

Abstract: This study attempts to model evaporation for Kuwait under arid conditions by using a wide range of monthly evaporation data, varying from 0.1 to 40 mm/day, from January 1993 to July 2015. Owing to the reason that the well-known theoretical evaporation models presented in the literature have been justified for a much shorter data range, the paper adopts empirical approaches to fit the data. Two evaporation models are presented based on classical statistical methods, one of multiple linear regression and another of time series analysis. The regression model, which is a function of temperature, relative humidity and wind speed, allows different modifications in the independent variables for more natural evaporation data synthesis. The time series model, which is a function of time only, is convenient for producing forecasts. Both evaporation models have been shown to produce results that are in reasonable agreement with observation values. This study advocates that the specific, rather simple, classical procedures performed to model the evaporation data can be effective alternatives to other theoretical and semi-theoretical methods found in the literature.

Keywords: Meteorological Parameter, Multiple Linear Regression, Pan Evaporation, Sinusoidal Pattern, Arid and Semi-Arid Zones

Introduction

Estimation of the water loss by evaporation is important for modeling, survey and management of water resource projects over different time and space scales (e.g., Molina Martinez *et al.*, 2006; Shirsath and Singh, 2010). Evaporation depends on the supply of heat energy and vapor pressure gradient, which in turn depend on meteorological factors such as temperature, relative humidity, wind speed and atmospheric pressure (Xu and Singh, 1998). In practice, potential evaporation data may be measured by using pan evaporimeters and can be used for modeling and analysis purposes due to their simplicity, low cost and availability of long historical data series. The data are useful for general monitoring and assessment of climate conditions, even if actual evaporation is overestimated because of differences in water quality, thermal inertia, advection and edge effect (Shaw *et al.*, 2010; McMahan *et al.*, 2013).

Theoretical evaporation models have been used by many researchers (e.g., Coulomb *et al.*, 2001; Gavin and Agnew, 2004; Chang *et al.*, 2010; Ershadi *et al.*, 2015). The classical Penman (1948) equation may be recommended as a standard method. However, many theoretical models require information that might not be

accessible from weather stations, such as the heat storage within the water body, similar to that found in Penman equation, which requires temperature profile measurements within the water (Brutsaert, 2005). Other models might be applicable to a limited data range or to conditions that are similar to those within which the models were derived. For such models, a calibration process would be necessary to render them applicable to the regional data.

Accordingly, simple empirical models have been developed for regional evaporations (Cahoon *et al.*, 1991; Crago and Brutsaert, 1992; Xu and Singh, 1998; Rotstayn *et al.*, 2006; Shirsath and Singh, 2010; Tabari *et al.*, 2010; Abou El-Magd and Ali, 2012; Kisi, 2015; Malik and Kumar, 2015). For example, Cahoon *et al.* (1991) and Fennessey and Vogel (1996) employed regression methods to develop models for monthly average evaporation in the USA. Tabari *et al.* (2010) estimated evaporation in a semi-arid region of Iran using both techniques of artificial neural network and of multivariate non-linear regression. Kim *et al.* (2015) evaluated combined bootstrap resampling and neural network models for daily evaporation in the Republic of Korea. Kisi (2015) employed a least square support vector machine, multivariate adaptive regression splines

and M5 Model Tree for evaporation data from Mersin and Antalya stations in Turkey. Other similar attempts were made successfully in the arid region of Saudi Arabia (Yassin *et al.*, 2016).

Among the plethora statistical fitting techniques available, regression and time series methods are quite common and have been implemented in many statistical software packages. Regression is a powerful technique for trend behavior estimation. The main advantage of an evaporation model produced by regression is to use explicit meteorological parameters as independent variables. This approach is useful to apply modifications for the adopted meteorological variables that yield more natural evaporation data synthesis. Time series analysis has the capability of simulating repeated data variation patterns. A model developed by this technique can be used to obtain forecasts, because it is a function of time only.

The aim here is to investigate the ability of regression and time series techniques in order to model a wide range of monthly evaporation data obtained from a weather station located in Kuwait in an arid environment. Initially, the meteorological data for the case study will be presented. Then the periodic variation pattern of the meteorological data will be examined in the frequency domain. The evaporation models will then be developed and evaluated.

Meteorological Data

Kuwait, which is about 18,000 km², is a desert country characterized by long, hot and dry summers and short winters. The average depth of annual evaporation is high, approaching 4000 mm, while that of precipitation is low,

varying from 50 to 250 mm. Temperatures during summer, which may vary from average daily temperatures of 43 to 23°C, are considered hot, but they can be worse when hot winds blow from the desert. Owing to the coastal location of the country, the heat is often rendered even more uncomfortable by high humidity, approaching a daily maximum of 97.5%. Winter temperatures, which vary from average daily temperatures of 15 to 5°C, may be classified as mild, but occasionally become cold when northerly or north-westerly winds bring cold air.

The meteorological data used in this study are monthly average measurements of pan evaporation (mm/day), temperature at 2 m height (°C), relative humidity (%) and wind speed at 2 m height (m/s). The effect of precipitation on evaporation rates will not be considered in this study due to the rare rainfall events in the country. The climatological data adopted are readily available from the Meteorological Department of the Directorate of Civil Aviation, collected at a local weather station near Kuwait Airport (Fig. 1). These data are of substantial continuity coverage, within a period of ~23 years between January 1992 and July 2015 and are considered representative of the climate within the urban zone. The reason for such a meteorological point estimate to be considered representative is that the urban zone of the country spans a small area, within latitudes 29°20'N to 29°03'N and longitudes 47°37'E to 48°10'E and is characterized by nearly flat surface elevations. Table 1 presents statistical description for the daily average measurements of the climatological parameters. It can be seen that within the specified time duration, the measurements for each parameter are highly variable reflecting the typical arid climate of the country.

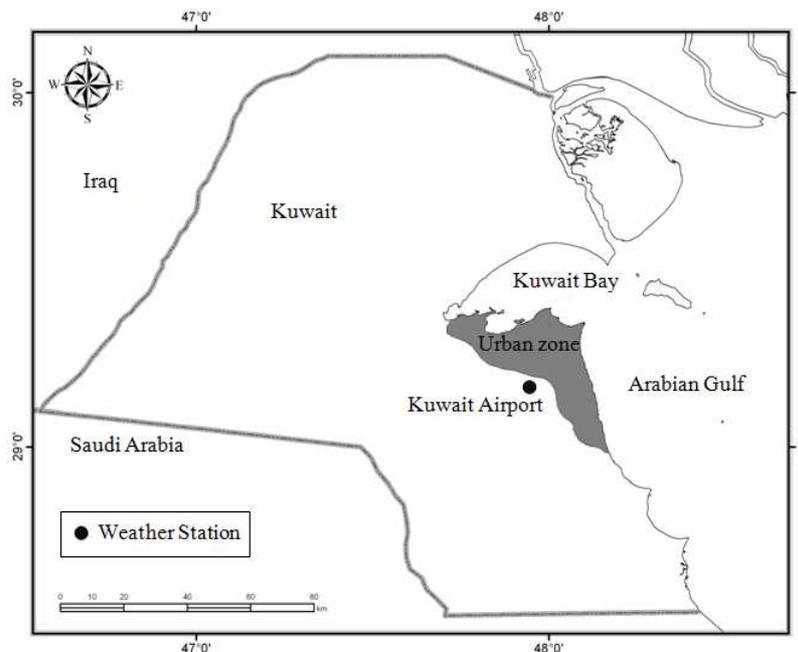


Fig. 1. Location of the weather station in Kuwait

Table 1. Descriptive statistics of daily average meteorological data for Kuwait from 1992 to 2015

	Evaporation (mm/day)	Temperature (°C)	Relative humidity (%)	Wind speed (m/s)
Mean	11.66	26.82	38.54	4.03
Standard deviation	7.85	9.68	20.48	1.86
Sample variance	61.64	93.65	419.37	3.44
Kurtosis	-0.38	-1.35	-0.88	-0.15
Skewness	0.68	-0.16	0.42	0.62
Range	39.90	38.04	92.00	11.36
Minimum	0.10	5.14	5.50	0.10
Maximum	40.00	43.18	97.50	11.46

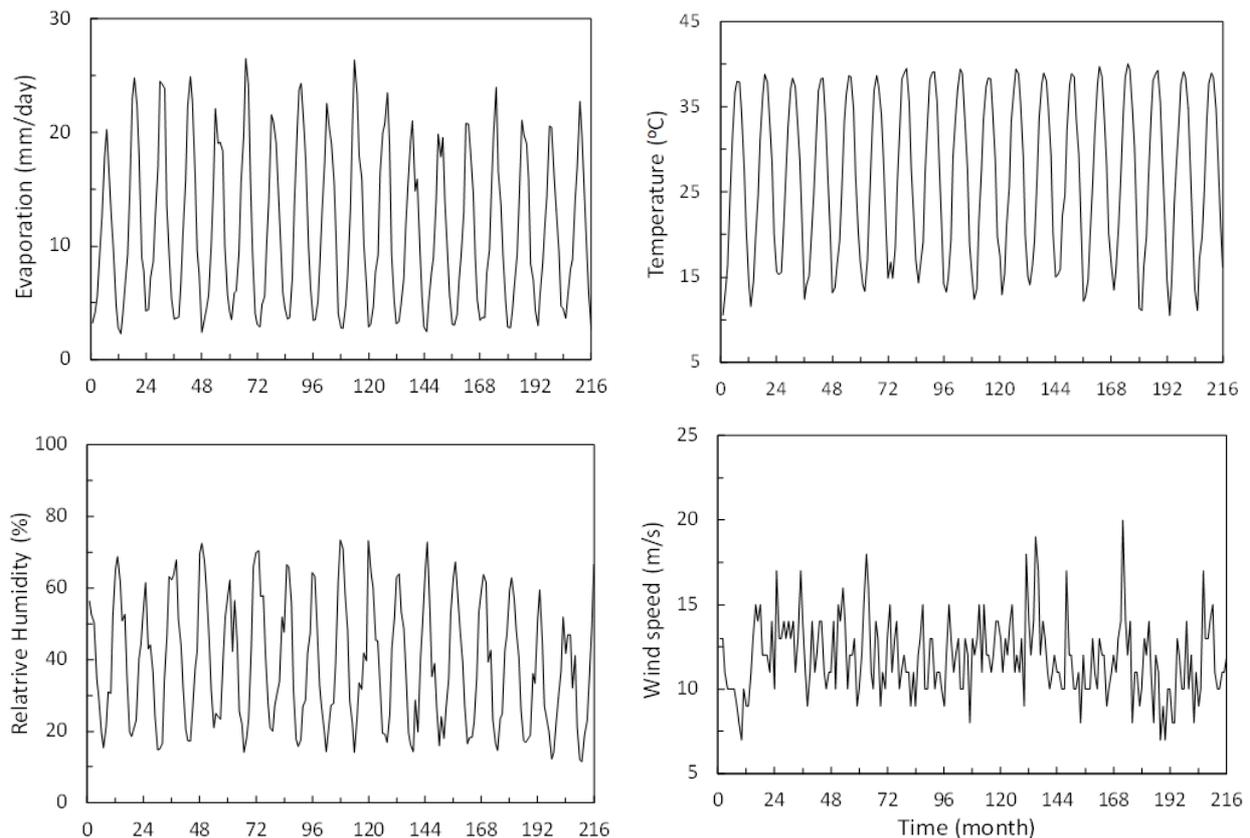


Fig. 2. Historical wind speed for the data of Kuwait

Variation Pattern

The monthly variation patterns for the meteorological data of evaporation, temperature, relative humidity and wind speed are plotted in Fig. 2. The time series patterns can be considered deterministic of a periodic nature with no obvious trend. An apparent correlation can be observed such that the variation of evaporation is directly proportional to temperature and wind speed, but inversely proportional to relative humidity. The correlations can better be described by considering statistics on a seasonal basis. The seasonal mean for the four datasets is obtained by using the expression:

$$\bar{j} = \frac{1}{N} \sum_{\tau=1}^N j_{v,\tau}, \tau = 1, \dots, 12 \quad (1)$$

Where:

j = Seasonal time series

\bar{j} = Seasonal mean

N = Total number of years

The results are shown in Fig. 3.

The estimated correlation coefficients for the seasonally-averaged evaporation with temperature, relative humidity and wind speed are $r_T = 0.97$, $r_H = -0.98$ and $r_u = 0.89$, respectively. The high positive correlation with temperature is a characteristic of desert environments. The high negative correlation with relative humidity reflects the coastal location of the country along the Arabian Gulf. Relative humidity refers to the amount of water in the air as a fraction of the total amount saturated air can hold.

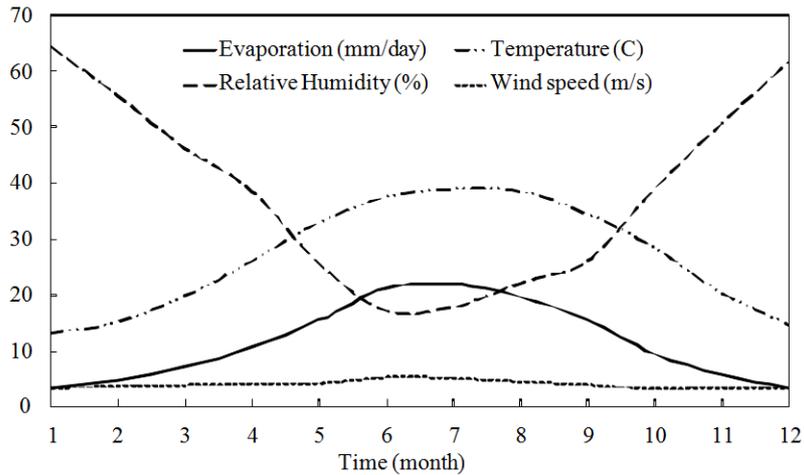


Fig. 3. Seasonal mean variation for monthly meteorological data of Kuwait

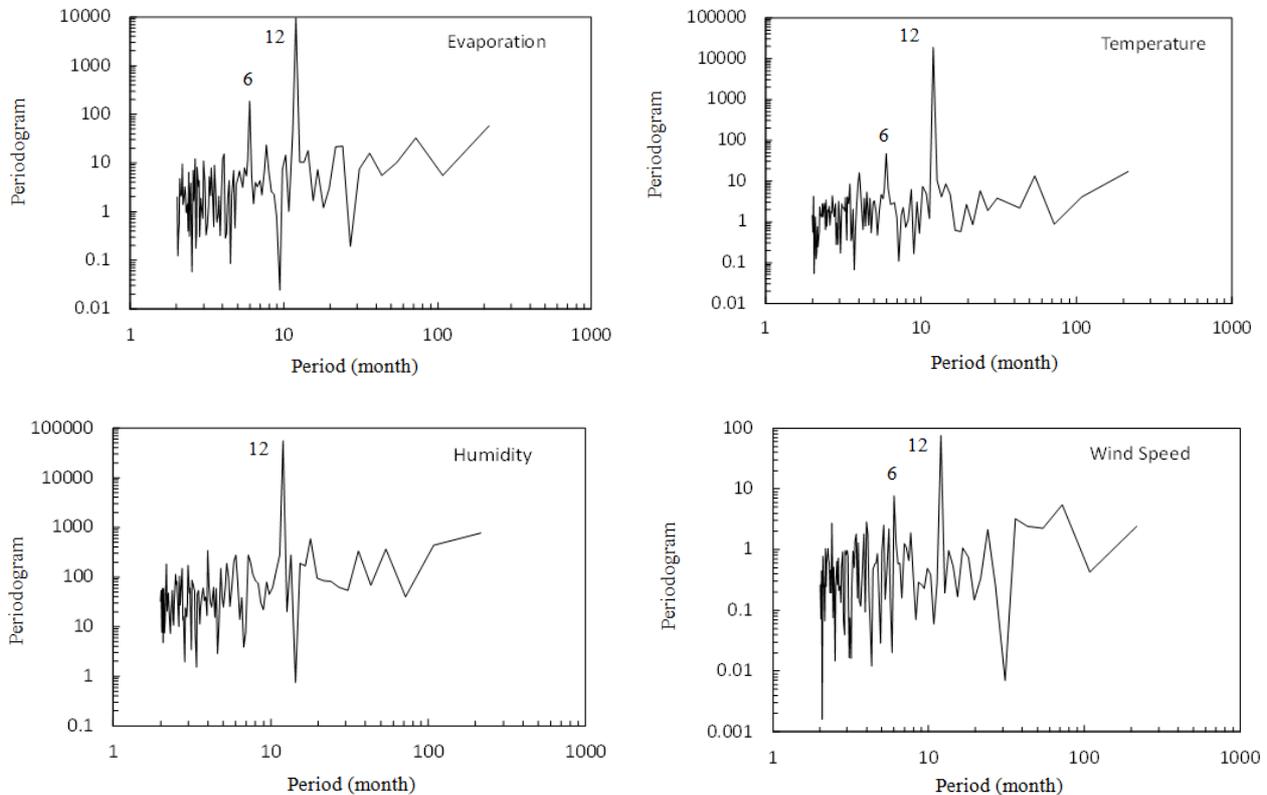


Fig. 4. Periodograms for monthly meteorological data of Kuwait

The more humidity in the air, the less space will be available for evaporation. Once the air reaches an upper relative humidity limit of 100%, it is no longer able to hold additional water molecules. The positive correlation with wind speed is also sufficiently high. Wind speed affects the rate of evaporation by sweeping away water particles that are in the air, allowing more particles to evaporate in the space above the water surface.

The periodic behavior of the meteorological data can be investigated using the periodogram technique, which is a Fourier transform of the autocovariance function representing an unsmoothed spectral plot for examining the cyclic structure in the frequency domain (Box *et al.*, 2015). This technique is used to reduce the effect of the measurement noise and, thus, detect which frequencies within the range of time are most responsible for the data

pattern. Typically, a large peak value shown in a periodogram corresponds to a period that is strongly represented in the time series.

The periodograms for the meteorological data are plotted in Fig. 4. A dominant annual periodicity of 12 months is present in all meteorological data implying that 6 months of the year possess considerably lower evaporation, temperature, relative humidity and wind speed than the other 6 months. Another, but less significant, periodicity of 6 month can be found in the data of evaporation, temperature and wind speed. This period can be related to the typical four seasons during the year. Though other periodicities are present in all periodograms, they can be considered insignificant.

Model Development

Two models are developed here for the evaporation data of Kuwait, one based on regression technique and the other on time series analysis.

Regression Model

Regression technique can be used to model pan evaporation data of Kuwait in terms of temperature, relative humidity and wind speed. Those parameters are important to consider because they have direct influence on the evaporation process. Temperature is correlated to solar radiation, which is the main source of heat energy required for vaporization. The ability to transport the vapor away from the evaporative surface is correlated to relative humidity and wind speed.

For multiple linear regression, the dependent variable y is assumed to be a function of k independent variables $x_1, x_2, x_3, \dots, x_k$. The model is expressed in the form:

$$y_i = b_0 + b_1x_{1,i} + \dots + b_kx_{k,i} + e_i \quad (2)$$

where, b_0, b_1, \dots, b_k are fitting constants; $y_i, x_{1,i}, \dots, x_{k,i}$ represent the i th observations of the variables y, x_1, \dots, x_k , respectively; and e_i is a random error term representing the remaining effects on y of variables not explicitly included in the model. For simple regression models, e_i can be assumed to be an uncorrelated variable with zero mean. The most common procedure for estimating b_0, b_1, \dots, b_k is to employ the least squares criterion with the minimum sum of squares of error terms (S); that is, to find b_0, b_1, \dots, b_k to minimize:

$$\begin{aligned} S &= \sum_{i=1}^n (y_{i\text{observed}} - b_0 - b_1x_{1,i} - \dots - b_kx_{k,i})^2 \\ &= \sum_{i=1}^n (y_{i\text{observed}} - y_{i\text{calculated}})^2 \\ &= \sum_{i=1}^n e_i^2. \end{aligned} \quad (3)$$

As a result, b_0, b_1, \dots, b_k must satisfy:

$$\frac{\partial S}{\partial b_j} = 2 \sum_{i=1}^n e_i \frac{\partial e_i}{\partial b_j} = 0, \quad j = 0, 1, \dots, k \quad (4)$$

And since $e_i = y_{i\text{observed}} - y_{i\text{calculated}}$, the Equation 4 becomes:

$$\frac{\partial S}{\partial b_j} = -2 \sum_{i=1}^n e_i \frac{\partial y_{i\text{calculated}}}{\partial b_j} = 0, \quad j = 0, 1, \dots, k \quad (5)$$

The meteorological data of Kuwait can be examined for the suitability of fitting this type of regression model. The range of meteorological data from January 1992 to December 2009 can be used to perform model fitting. The remaining data until July 2015 can be used later for verification. Figure 5 presents the relation between evaporation and the other climatological parameters of temperature, relative humidity and wind speed. A general multiple regression model expressing the evaporation in terms of those climatological parameters can initially be assumed as:

$$E_i = b_1T_i + b_2H_i + b_3u_i \quad (6)$$

Where:

E = Pan evaporation (mm/day)

T = Temperature ($^{\circ}\text{C}$)

H = Relative humidity (%)

u = Wind speed (m/s)

Based on the classical assumptions of multiple regression modeling, Equation 6 suggests linear correlations between the evaporation and the independent variables. However, Fig. 5 shows a definite curvilinear appearance for the relation between evaporation and both temperature and relative humidity. It is seen that those relations are best expressed correspondingly as power and exponential functions and the above multiple regression equation can thus be linearized by transforming the independent variables of temperature and relative humidity as:

$$E_i = b_1T_i^* + b_2H_i^* + b_3u_i \quad (7)$$

Where:

$$T^* = T^{1.69} \quad (8)$$

$$H^* = e^{-0.038H} \quad (9)$$

The general regression model with the fitted parameters b_1, b_2 and b_3 becomes:

$$E_i = 0.024T_i^{1.69} + 18e^{-0.038H_i} + 0.033u_i \quad (10)$$

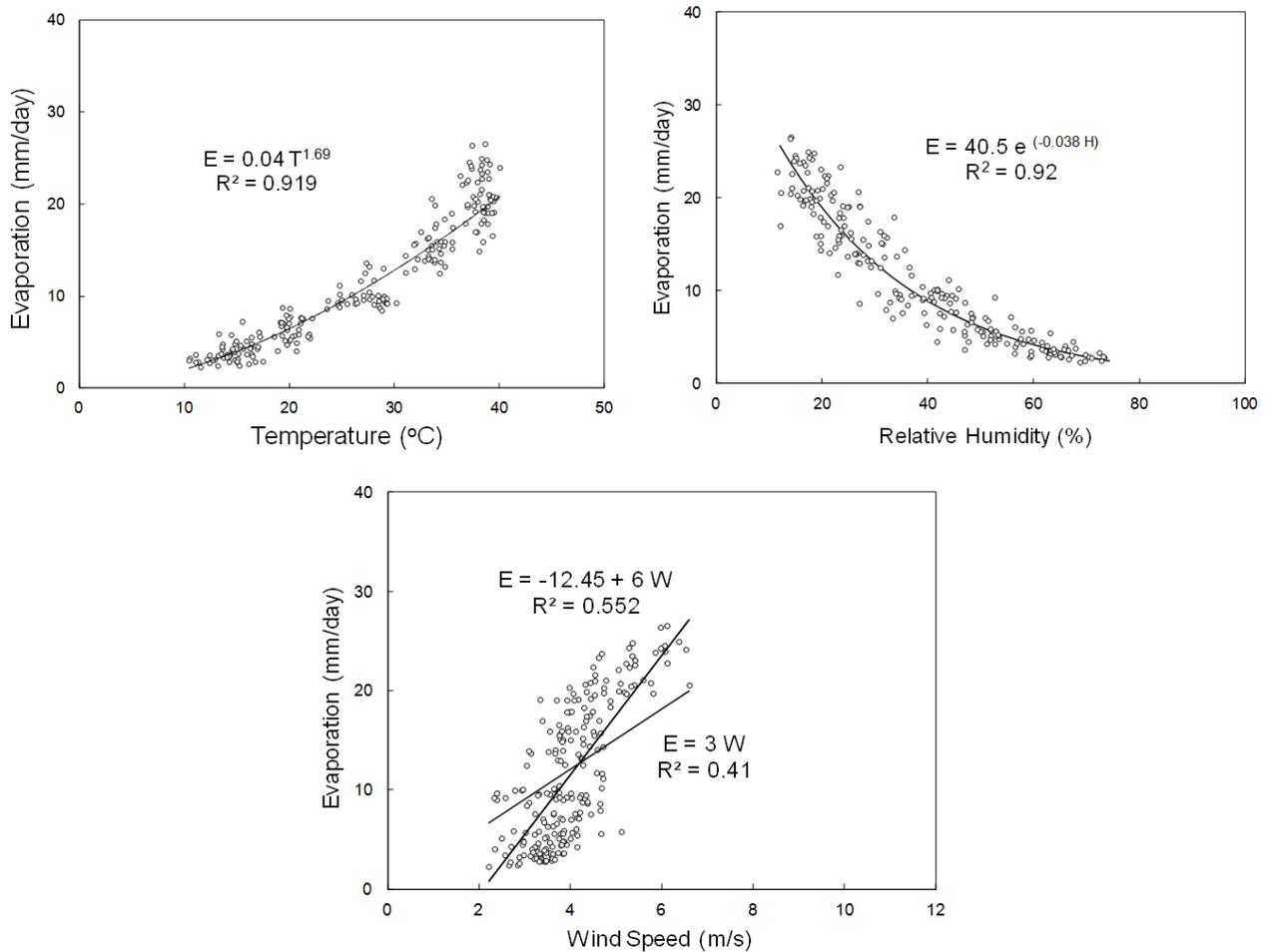


Fig. 5. Evaporation as a function of temperature and relative humidity for the data of Kuwait from 1992 to 2009. The solid curves represent suggested relations fitted by regression

It is worth mentioning that Equation 10 neglects the influence of other meteorological parameters on evaporation rates, as it considers an intercept equal to zero, i.e., $b_0 = 0$. Regarding the effect of wind speed, although Fig. 5 suggests that a model with a non-zero intercept can account for a linear correlation with evaporation of the form:

$$E = b_0 + bu \quad (11)$$

Regression might not result with a significant fit for the coefficients because of the considerable data scatter. However, it is possible to assume that the intercept of this correlation is equal to zero:

$$E = b u \quad (12)$$

Although this assumption results with a fitting accuracy less significant for the average data presented on a monthly basis, Fig. 5 shows that the non-zero intercept trend constitutes a possible relation fitted by eye.

Time Series Model

A possible application for examining the periodic behavior of the meteorological data is to employ the detected periodicities in order to provide model forecasts. In general, a time series containing a periodic sinusoidal component with a known wavelength can be modeled using:

$$s(t) = \sum_{i=1}^k R_i \cos(2\pi f_i t + \theta_i) \quad (13)$$

Where:

- s = Periodic sinusoidal component
- R_i = Amplitude of variation
- f_i = Frequency, equal to the inverse of period
- θ_i = Phase angle
- k = Total number of periodicities

The term $(2\pi f_i t + \theta_i)$ is measured in radians. That is, the k value for the temperature data is equal to two and

the values of f_1 and f_2 are set as 1/6 and 1/12 cycles per month, respectively. The phase angle θ is necessary to adjust the model so that the cosine function crosses the mean at the appropriate time, t . The values of θ and R_i can be determined by means of numerical optimization. The fitted periodic sinusoidal model for the evaporation data is found to be:

$$E(t) = 11.33 - 1.2 \cos\left(\frac{2\pi}{6}t + 1.85\right) + 9.17 \cos\left(\frac{2\pi}{12}t + 2.7\right) \quad (14)$$

Model Evaluation

Figure 6 presents the two evaporation models of regression (Equation 10) and of time series (Equation 14). For both models, there is a reasonable agreement between observed and calculated values for the data within the time from January 1992 to December 2009, used to perform model calibration. The remaining data from January 2010 to July 2015 were used for model verification. The figure shows that both models may be considered successful in representing most of the seasonal variation pattern for the data from January 1992 to July 2015.

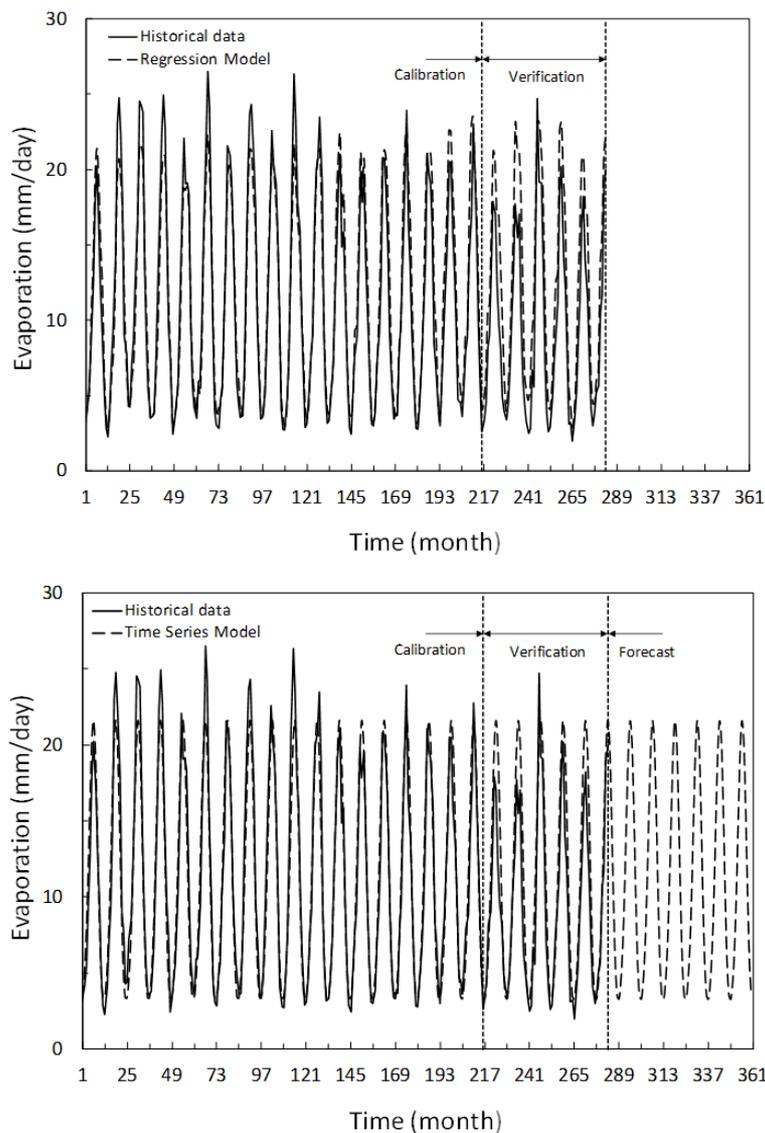


Fig. 6. Evaporation as a function of time in months for the two models. Calibration is performed for the duration from January 1992 (corresponding to month number 1) to December 2009 (month number 216); verification is performed for the duration from January 2010 (month 217) to July 2015 (month 283); and forecast is performed for the duration from August 2015 (month 284) to December 2021 (month 360)

Table 2. Fitting accuracy for the model

Model	Technique	Range	Accuracy measure		
			MAPE	RMSE (mm/day)	NSE
Equation (10)	Regression	Calibration	13.2	1.69	0.94
		Verification	38.8	3.48	0.64
Equation (14)	Time series	Calibration	11.6	1.69	0.94
		Verification	23.9	2.76	0.77

The performance of the two models can be evaluated quantitatively. The following statistical error tests can be adopted, which are the Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and Nash-Sutcliffe Equation (NSE). These indicators are calculated as follows:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_{i\text{calculated}} - y_{i\text{observed}}}{y_{i\text{observed}}} \right| \times 100 \quad (15)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{i\text{calculated}} - y_{i\text{observed}})^2} \quad (16)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (y_{i\text{observed}} - y_{i\text{calculated}})^2}{\sum_{i=1}^n (y_{i\text{observed}} - \bar{y}_{i\text{observed}})^2} \quad (17)$$

Respectively, where \bar{y} is the mean value of the dependent variable. For better data modeling, MAPE and RMSE statistics should be closer to zero. The NSE criterion compares the model performance to the use of the mean value of the dependent variable as an estimate. Given a perfect fit, the NSE criterion is equal to 1.0; if the model is worse than the mean value of the dependent variable, the NSE statistic will be negative. Table 2 presents the accuracy measures, which show that the performances of the two models are nearly identical to each other.

Given that the conditions used to derive the model remain the same, Fig. 6 presents the forecasts provided by the time series model for the times pan from August 2015 to December 2021. Based on the previous verification range, the mean forecasting error is assumed to be nearly within accuracy measures of MAPE, RMSE and NSE equal to 23.9, 2.76 and 0.77, respectively. The produced forecasts fall within the same range of monthly average historical data and exhibit an apparent seasonal variation pattern.

Conclusion

The two derived evaporation models, one based on the technique of multiple linear regression and the other on time series analysis, were shown to be successful in describing most of the variation pattern for the data obtained from Kuwait. For the former model, the transformation of both temperature and relative humidity variables, which was considered by using corresponding

mathematical functions of power and exponential forms, improved the correlation results. Although the employed mathematical functions are applicable locally for the meteorological data of Kuwait, it is hoped that this will prompt others to examine whether such correlations exist universally in data collected from other locations. The time series model, which exploited two typical periods of seasonal and annual cycles, produced monthly forecasts for the times pan from August 2015 to December 2021 that can be of great importance for water resources management applications in Kuwait.

Acknowledgment

The author is grateful to the Metrological Department of the Directorate General of Civil Aviation of Kuwait, Climatological Division, for providing the relevant data of pan evaporation, temperature, relative humidity and wind speed.

Ethics

The author declares that there is no conflict of interests regarding the publication of this paper.

References

- Abou El-Magd, I.H. and E.M. Ali, 2012. Estimation of the evaporative losses from lake Nasser, Egypt using optical satellite imagery. *Int. J. Digital Earth*, 5: 133-146.
- Box, G.E., G.M. Jenkins and G.C. Reinsel, 2015. *Time Series Analysis: Forecasting and Control*. 5th Edn., John Wiley and Sons, Hoboken, ISBN-10: 1118674928, pp: 712.
- Brutsaert, W., 2005. *Hydrology: An Introduction*. 1st Edn., Cambridge University Press, Cambridge, ISBN-10: 0521824796, pp: 605.
- Cahoon, J.E., T.A. Costello and J.A. Ferguson, 1991. Estimating pan evaporation using limited meteorological observations. *Agric. Forest Meteorol.*, 55: 181-190.
DOI: 10.1016/0168-1923(91)90061-T
- Chang, F.J., L.C. Chang, H.S. Kao and G.R. Wu, 2010. Assessing the effort of meteorological variables for evaporation estimation by self-organizing map neural network. *J. Hydrol.*, 384: 118-129.
DOI: 10.1016/j.jhydrol.2010.01.016

- Coulomb, C.V., D. Legesse, F. Gasse, Y. Trivic and T. Chernetd, 2001. Lake evaporation estimates in tropical Africa (Lake Ziway, Ethiopia). *Hydrol. Process.*, 245: 1-18.
DOI: 10.1016/S0022-1694(01)00341-9
- Crago, R.D. and W. Brutsaert, 1992. A comparison of several evaporation equations. *Water Resources Res.*, 28: 951-954. DOI: 10.1029/91WR03149
- Ershadi, A., M.F. McCabe, J.P. Evans and E.F. Wood, 2015. Impact of model structure and parameterization on Penman-Monteith type evaporation models. *J. Hydrol.*, 525: 521-535.
DOI: 10.1016/j.jhydrol.2015.04.008
- Fennessey, N.M. and R.M. Vogel, 1996. Regional models of potential evaporation and reference evapotranspiration for the northeast USA. *J. Hydrol.*, 184: 337-354. DOI: 10.1016/0022-1694(95)02980-X
- Gavin, H. and C.A. Agnew, 2004. Modelling actual, reference and equilibrium evaporation from a temperate wet grassland. *Hydrol. Process.*, 18: 229-246. DOI: 10.1002/hyp.1372
- Kim, S., Y. Seo and V.P. Singh, 2015. Assessment of pan evaporation modeling using bootstrap resampling and soft computing methods. *J. Comput. Civil Eng.*
DOI: 10.1061/(ASCE)CP.1943-5487.0000367
- Kisi, O., 2015. Pan evaporation modeling using least square support vector machine, multivariate adaptive regression splines and M5 model tree. *J. Hydrol.*, 528: 312-320.
DOI: 10.1016/j.jhydrol.2015.06.052
- Malik, A. and A. Kumar, 2015. Pan evaporation simulation based on daily meteorological data using soft computing techniques and multiple linear regression. *Water Resources Manage.*, 29: 1859-1872.
DOI: 10.1007/s11269-015-0915-0
- McMahon, T.A., M.C. Peel, L. Lowe, R. Srikanthan and T.R. McVicar, 2013. Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: A pragmatic synthesis. *Hydrol. Earth Syst. Sci.*, 17: 1331-1363.
DOI: 10.5194/hess-17-1331-2013
- Molina Martinez, J.M., V. Martinez Alvarez, M.M. Gonzalez-Real and A. Baille, 2006. A simulation model for predicting hourly pan evaporation from meteorological data. *J. Hydrol.*, 318: 250-261. DOI: 10.1016/j.jhydrol.2005.06.016
- Penman, H.L., 1948. Natural evaporation from open water, bare soil and grass. *Proc. Royal Society A*, 193: 120-145. DOI: 10.1098/rspa.1948.0037
- Rotstain, L.D., M.L. Roderick and G.D. Farquar, 2006. A simple pan-evaporation model for analysis of climate simulations: Evaluation over Australia. *Geophys. Res. Lett.* DOI: 10.1029/2006GL027114
- Shaw, E.M., K.J. Beven, N.A. Chappell and R. Lamb, 2010. *Hydrology in Practice*. 4th Edn., CRC Press, ISBN-10: 0203030230, pp: 546.
- Shirsath, P.B. and A.K. Singh, 2010. A comparative study of daily pan evaporation estimation using ANN, regression and climate based models. *Water Resources Manage.*, 24: 1571-1581.
DOI: 10.1007/s11269-009-9514-2
- Tabari, H., S. Marofi and A. Sabziparvar, 2010. Estimation of daily pan evaporation using artificial neural network and multivariate non-linear regression. *Irrigat. Sci.*, 28: 399-406.
DOI: 10.1007/s00271-009-0201-0
- Xu, C.Y. and V.P. Singh, 1998. Dependence of evaporation on meteorological variables at different time-scales and intercomparison of estimation methods. *Hydrol. Process.*, 12: 429-442. DOI: 10.1002/(SICI)1099-1085(19980315)12:3<429::AID-HYP581>3.0.CO;2-A
- Yassin, M.A., A.A. Alazba and M.A. Mattar, 2016. Artificial neural networks versus gene expression programming for estimating reference evapotranspiration in arid climate. *Agric. Water Manage.*, 163: 110-124.
DOI: 10.1016/j.agwat.2015.09.009