Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v10n1p65

Multivariate time series modeling of monthly rainfall amounts

By Mahdi et al.

Published: 26 April 2017

Electronic Journal of Applied Statistical Analysis
Vol. 10, Issue 01, April 2017, 65-81
DOI: 10.1285/i20705948v10n1p65

Multivariate time series modeling of monthly rainfall amounts

Esam Mahdi*[a], Serge B. Provost[b], Raid B. Salha[a], and Imad I. H. Nashwan[c]

[a]*Islamic University of Gaza, Department of Mathematics, Palestine*
[b]*University of Western Ontario, Department of Statistics and Actuarial Sciences, Canada*
[c]*Al-Quds Open University, Faculty of Technology and Applied Sciences, Palestine*

Published: 26 April 2017

This paper discusses the fitting of suitable models to rainfall observations. Daily rainfall amounts were aggregated to monthly data using the Thiessen polygons method and multivariate seasonal vector integrated autoregressive moving average models (sVARIMA) were fitted to the monthly cumulative rainfall volume. The data were obtained from the 12 Palestinian meteorological gauge stations located across the 5 governorates of the Gaza Strip and incorporated 42 years (from 1973 to 2014) of irregular daily precipitation. It can be concluded that the use of sVARIMA models in the environmental sciences provide a useful approach for forecasting rainfall data as a preliminary guideline toward short and long-term sustainable water resources management.

Keywords: sVARIMA models, exact maximum likelihood estimation, state space representation, Gaza strip.

1 Introduction

Many time series encountered in hydrological, meteorological and environmental studies exhibit strong seasonal behavior with a fixed integer period. Obviously, monthly rainfall amount series usually exhibit strong seasonal behavior. When seasonal correlation exists but is not taken into account then, most likely, the estimators will not be fully efficient. An even more serious problem is that statistical inferences from the model may be

*Corresponding author: emahdi@iugaza.edu.ps

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

incorrect. Clearly the analysis of rainfall amount series impacts many disciplines. For instance, in agricultural planning, one needs to determine the optimum time for planting crops based on the predicted rainfall amount during the precipitation season.

One can argue that the marginal distribution of rainfall as time series data may have a large spike at the origin and positive skewness. For this reason, the combination of these discrete and continuous features makes the use of time series for modelling the rainfall data non accurate. To deal with such a case, Cox (1981) introduced a model of nonseasonal first order autoregressive (AR) character. To handle the seasonal and nonseasonal variations in the rainfall data, Stern and Coe (1984) proposed the fitting and the use of non-stationary Markov chains and gamma distributions to the rainfall occurrence and amounts respectively. They showed that the process of fitting and using these models provides a straightforward and flexible analysis for rainfall records.

In contrast to the univariate time series cases, where the seasonal ARIMA models have received considerable attention, there has been much fewer studies involving multiplicative sVARIMA modeling in multivariate time series. In the literature, Salas et al. (1980); Bras and Rodríguez-Iturbe (1985); Ansley and Kohn (1995) and Soltani et al. (2007) implemented the classical multivariate time series models to preserve the auto and cross correlations of the rainfall data via univariate auto regressive moving average (ARMA) models; however, they did not consider the multivariate seasonal models in details.

Although numerous studies on rainfall data have been carried out with data collected in Middle East countries, none of these utilized seasonal multivariate time series techniques to describe the individual and the possible cross relationships in the rainfall data. In fact, most previously published articles in the field of environmental sciences, especially those in connection with data originating from the Gaza Strip, have consisted simply of a summary of characteristics of rainfall based on the observed data where the time series models did not aim at forecasting future results (El-Nahhal et al., 2013a,b,c; Safi et al., 2014).

Decision making for short and long term water resource management necessitates a clear insight into hydrology and climate volatility, as for instance estimating rainfall amounts during precipitation seasons may provide significant information about fresh water availability in the near and mid-term future (Banu et al., 2016). Generally, rainfall like many other environmental data shows spatial and temporal variation (Cristiano et al., 2016) where the precipitation season displays 12 months of circular stationarity; hence, in many situations, especially within a small geographic area (such as the Gaza Strip), spatial-dependence is almost perfect so that the variability in the rainfall observations can be explained by temporal models. Therefore, in the multivariate analysis of rainfall, the objective in modeling monthly aggregates is to describe the within-year variability, where each variable represents the amount of rainfall at a single location (gauge station).

The main objective of this article is to propose an empirical multivariate temporal model that estimates the amount of monthly cumulative rainfall, which could contribute to the sustainable management of groundwater resources. This research is based on monthly cumulative rainfall datasets obtained from 12 meteorological stations located across the Gaza Strip in the State of Palestine.

The next section introduces the study and describes the dataset being analyzed. In Section 3, we provide some background on the multivariate seasonal ARIMA models. We discuss the exact maximum likelihood methodology of multivariate processes with missing values in Section 4. The Akaike and Schwarz Bayesian information criteria for model selection and multivariate portmanteau diagnostic seasonal and non-seasonal test statistics are discussed in Sections 5 and 6, respectively. In Section 7, we fit the seasonal vector ARIMA model to the monthly rainfall amounts (measured in Million Cubic Meter, MCM) data. Then we assess and summarize the results in Section 8.

2 Study area and data set

The Gaza Strip is a small region with a total area of 365 square kilometers, which is located in arid and semi-arid areas on the eastern coast of the Mediterranean Sea at longitude $34^{o}200''$ East and latitude $31^{o}250''$ North. It is well known that the Gaza Strip is one of the world's most densely populated areas, and its unique coastal plain aquifer is not meeting the population needs due to the unsustainably high rate of extraction of this aquifer (Melloul and Collin, 2000). This aquifer extends over a distance of 120 km starting from South of mount Carmel in Haifa and ending in the Gaza Strip; it has a width of 7-20 km and disappears near the foothills of the mountains of the West Bank (Al-Najar, 2011; Shomer et al., 2004). The aquifer in the Gaza Strip receives an average annual recharge of 50-60 Million Cubic Meter per Year (MCM /y) mainly from rainfall; rainfall being the main source of groundwater and surface water resources as it provides more than 90% of all water supplies in the Gaza Strip (Palestinian Water Authority, 2013). The precipitation season in the Gaza Strip extends from October to March. Rain rarely falls in September, April, or May. Since October 31, 1972, rain never fell during the summer months (June-August) except for only one day, namely, June 15, 1992. This short winter season cannot replenish the groundwater systems as needed, which has led to the lowering of the groundwater level and sea water intrusion. The contamination from sea water intrusion and the lack of sufficient refilling of the aquifer of the Gaza Strip has resulted in brackish irrigation, and consequently, drinking water that is not complying with the standards proposed by the Food and Agriculture Organization of the United Nations, FAO (Al-Khatib and Al-Najar, 2011; Palestinian Water Authority, 2013). Problems associated with the lack of quality and quantity of fresh water resources in Palestine and many other countries in the Middle East may be alleviated and controlled if decision makers in these countries support scientific research, including statistical models for rainfall precipitation, and adopt the recommendations.

The data used in this research was obtained from the 12 Palestinian meteorological stations located across the five governorates of the Gaza Strip[1] and incorporated 42 years (from 1973 to 2014) of 1622 irregular daily observations (data available at the $n = 1622$ integer times $t_1 < t_2 < \cdots < t_n$ are not equally spaced, having been taken during precipitation seasons). In general, rain gauge data were collected by the Palestinian

[1]The five governorates of the Gaza Strip are: Northern, Gaza, Deir el-Balah, Khan Yunis, and Rafah.

Ministry of Agriculture from October of a given year to March of the following year[2], so that the amount of the rainfall can be considered to be null during the non-precipitation months.

Table 1: Summary statistics for rainfall gauge stations

Governorate	Station	Longitude	Latitude	Area (Km^2)	
Northern	Beit Hanon	$35^o13'$E	$32^o56'$N	29.0	58.8
	Beit Lahia	$35^o21'$E	$31^o37'$N	14.3	
	Jabalia	$34^o29'$E	$31^o32'$N	15.5	
Gaza	Shati	$35^o16'$E	$32^o21'$N	2.3	73.5
	Gaza City	$34^o27'$E	$31^o30'$N	13.0	
	Tuffah	$35^o23'$E	$31^o76'$N	23.3	
	South Gaza	$34^o24'$E	$31^o28'$N	35.0	
Middle	Nussirat	$34^o23'$E	$31^o26'$N	29.5	68.0
	Deir el-Balah	$35^o27'$E	$31^o51'$N	38.5	
Khan Yunis	Khan Yunis	$35^o18'$E	$32^o28'$N	83.5	126.0
	Khuzaa	$34^o21'$E	$31^o18'$N	42.5	
Rafah	Rafah	$34^o16'$E	$31^o14'$N	38.8	38.8

Figure 1 shows the locations of the 12 rainfall gauge stations within the Gaza Strip and Table 1 summarizes some descriptive geographic information about these stations. It is worth noting that the data has missing series from October 31, 1972 till November 10, 1998 for both Jabalia and Tuffah stations, and from October 30, 1980 till October 12, 1991 for the South Gaza (Mughraka) station, and from October 31, 1972 till November 5, 1999 for Khuzaa station. To analyze such a process involving an unusual pattern of missing data, the exact maximum likelihood estimation approach will be used for estimating the fitted model parameters (Jones, 1980; Harvey and Pierse, 1984; Ansley and Kohn, 1983; Wincek and Reinsel, 1986).

3 Seasonal vector ARIMA (sVARIMA) models

The multiplicative seasonal vector integrated autoregressive moving average model sVARIMA $(p, d, q) \times (P, D, Q)_s$ with mean vector zero for a k-dimensional time series of size n, $\boldsymbol{Z}_t = (Z_{1t}, \ldots, Z_{kt})'$, $t = 1, 2, \ldots, n$ (Reinsel, 1997; Lütkepohl, 2005; Tsay, 2014) can be written as

[2]Data includes very few observations in April, May and September.

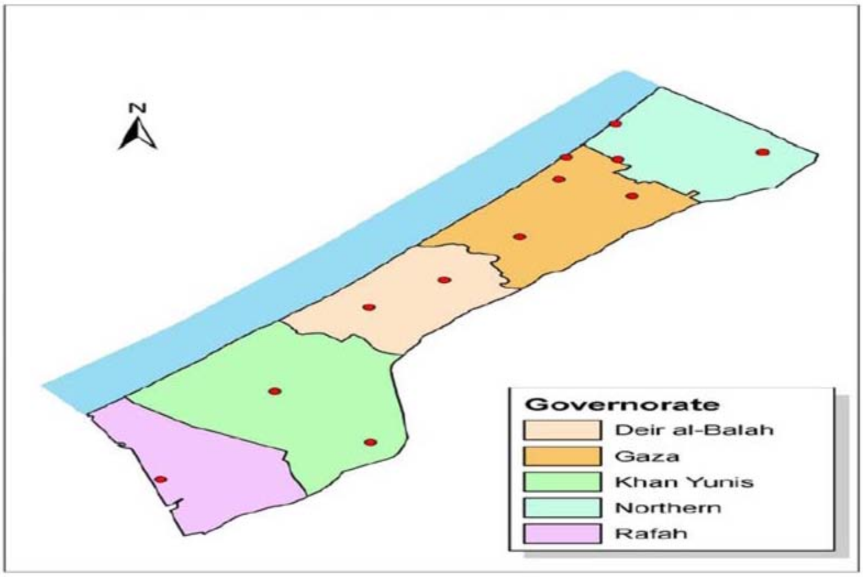

Figure 1: Location of rainfall stations in the Gaza strip.

$$\Phi_p(\boldsymbol{B})\boldsymbol{\Phi}_P(\boldsymbol{B}^s)\nabla^d\boldsymbol{\nabla}^{\boldsymbol{s}D}\boldsymbol{Z}_t = \Theta_q(\boldsymbol{B})\boldsymbol{\Theta}_{\boldsymbol{Q}}(\boldsymbol{B}^s)\boldsymbol{a}_t, \tag{1}$$

with $\Phi_p(\boldsymbol{B}) = \mathbb{I}_k - \varphi_1\boldsymbol{B}^1 - \cdots - \varphi_p\boldsymbol{B}^p$ and $\Theta_q(\boldsymbol{B}) = \mathbb{I}_k - \vartheta_1\boldsymbol{B}^1 - \cdots - \vartheta_q\boldsymbol{B}^q$ being non-seasonal polynomials in $\boldsymbol{B}$ of degrees p and q, respectively, and $\boldsymbol{\Phi}_P(\boldsymbol{B}^{\boldsymbol{s}}) = \mathbb{I}_k - \Phi_1\boldsymbol{B}^s - \cdots - \Phi_p\boldsymbol{B}^{sP}$ and $\boldsymbol{\Theta}_Q(\boldsymbol{B}^{\boldsymbol{s}}) = \mathbb{I}_k - \theta_1\boldsymbol{B}^s - \cdots - \theta_Q\boldsymbol{B}^{sQ}$ being seasonal polynomials in $\boldsymbol{B}^s$ of degrees P and Q, respectively. The integers $p \geq 0$ and $q \geq 0$ are the orders of the non-seasonal autoregressive (AR) model and moving average (MA) model respectively, whereas $P \geq 0$ and $Q \geq 0$ are the orders of the seasonal autoregressive (SAR) model and seasonal moving average (SMA) model respectively. $\boldsymbol{B}$ is the backshift operator on t and $s > 0$ is the length of the seasonal period; d is the order of the non-seasonal differences where $\nabla^d = (\mathbb{I}_k - \boldsymbol{B})^d$ is a $k \times k$ diagonal matrix of dimension k representing the non-seasonal differencing operator, whereas D is the order of seasonal differences and $\boldsymbol{\nabla}^{\boldsymbol{s}D} = (\mathbb{I}_k - \boldsymbol{B}^s)^D$ is a $k \times k$ diagonal matrix representing the seasonal differencing operator; both d and D may be selected to transform the non-stationarity in the sVARIMA models to stationary VARMA models. It is assumed that the VARMA model is stationary, invertible, and identifiable with no common roots between $\Phi_p(\boldsymbol{B})$ and $\Theta_q(\boldsymbol{B})$ and no common roots between $\boldsymbol{\Phi}_P(\boldsymbol{B}^s)$ and $\boldsymbol{\Theta}_Q(\boldsymbol{B}^s)$ (Reinsel, 1997; Lütkepohl, 2005; Tsay, 2014). The process $\{\boldsymbol{a}_t\}$ is a white noise process which is uncorrelated in time with a mean zero; that is, $\mathbb{E}(\boldsymbol{a}_t) = \boldsymbol{0}$ and $\mathbb{E}(\boldsymbol{a}_t\boldsymbol{a}'_{t-\ell}) = \Gamma_0\delta_\ell$ where Γ_0 is a $k \times k$ positive-definite variance covariance matrix and δ_ℓ is the usual Kronecker delta, which is equal to unity at $\ell = 0$ and zero elsewhere.

4 Exact maximum likelihood estimation

We first transform the sVARIMA model specified by Equation (1) into the Box-Jenkins VARMA $(\dot{P}, \dot{Q})$ representation with $\dot{P} = 1+p+P\times s$ and $\dot{Q} = 1+q+Q\times s$ as follows:

$$\dot{\boldsymbol{\Phi}}_{\dot{P}}(\boldsymbol{B})\ddot{\boldsymbol{Z}}_t = \dot{\boldsymbol{\Theta}}_{\dot{Q}}(\boldsymbol{B})\boldsymbol{a}_t, \tag{2}$$

with

$$\dot{\boldsymbol{\Phi}}_{\dot{P}}(\boldsymbol{B}) = \Phi_p(\boldsymbol{B})\boldsymbol{\Phi}_P(\boldsymbol{B}^s) \text{ and } \dot{\boldsymbol{\Theta}}_{\dot{Q}}(\boldsymbol{B}) = \Theta_q(\boldsymbol{B})\boldsymbol{\Theta}_Q(\boldsymbol{B}^s), \ddot{\boldsymbol{Z}}_t = \dot{\boldsymbol{\nabla}}\boldsymbol{Z}_t,$$

where $\dot{\boldsymbol{\nabla}} = diag[(\mathbb{I}_k - \boldsymbol{B})^d(\mathbb{I}_k - \boldsymbol{B}^s)^D]$.

It was shown that the state space form of the VARMA model associated with Kalman filtering procedures is a convenient representation for constructing the likelihood function in situations where we have values that are not observed at equally spaced times because of missing observations (Jones, 1980; Harvey and Pierse, 1984; Ansley and Kohn, 1983; Wincek and Reinsel, 1986).

Now suppose that the observations on at least some components of the vector series $\{\ddot{\boldsymbol{Z}}_t\}$ are available at the n integer times $t_1 < t_2 < \cdots < t_n$, not necessarily equally spaced, and $\ddot{\boldsymbol{Z}}_t$ follows the VARMA model as given in (2). Thus, at time t_i, we observe $\boldsymbol{Z}^\star_{t_i} = M_i\ddot{\boldsymbol{Z}}_{t_i}$, where M_i is a known incidence matrix of dimension $k_i \times k(k_i \le k)$, whose elements are equal to 1 or 0 to indicate the occurrence of an observation in each given component. In particular, $M_i = \mathbb{I}_k$ when all components of $\ddot{\boldsymbol{Z}}_{t_i}$are observed at time t_i so that $\ddot{\boldsymbol{Z}}_t$ has the state-space representation given by

$$\boldsymbol{Y_t} = \dot{\boldsymbol{\Phi}}\boldsymbol{Y_{t-1}} + \boldsymbol{\Psi}\boldsymbol{a_t}, \tag{3}$$

or equivalently

$$\boldsymbol{Y_t} = \begin{bmatrix} \boldsymbol{0} & \mathbb{I}_k & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \mathbb{I}_k & \dots & \boldsymbol{0} \\ \vdots & \boldsymbol{0} & \boldsymbol{0} & \dots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \dots & \mathbb{I}_k \\ \dot{\boldsymbol{\Phi}}_r & \dot{\boldsymbol{\Phi}}_{r-1} & \dots & \dots & \dot{\boldsymbol{\Phi}}_1 \end{bmatrix} \boldsymbol{Y_{t-1}} + \begin{bmatrix} \mathbb{I}_k \\ \Psi_1 \\ \vdots \\ \vdots \\ \Psi_{r-1} \end{bmatrix} \boldsymbol{a_t}, \tag{4}$$

with $\dot{\boldsymbol{\Phi}}_i = \boldsymbol{0}$ if $i > \dot{P}$, $\boldsymbol{\Psi}(\boldsymbol{B}) = \dot{\boldsymbol{\Phi}}_{\dot{P}}(\boldsymbol{B})^{-1}\dot{\boldsymbol{\Theta}}_{\dot{Q}}(\boldsymbol{B})$, $\ddot{\boldsymbol{Z}}_t = H\boldsymbol{Y}_t = [\mathbb{I}, \boldsymbol{0}, \dots, \boldsymbol{0}]\boldsymbol{Y}_t$, $\boldsymbol{Y}_t$ is the kr-dimensional state vector and $r = \max(\dot{P}, \dot{Q}+1)$ (Reinsel, 1997, Ch. 7).

It follows that the joint density of the vector of observations $\boldsymbol{z} = (\boldsymbol{Z}^{\star\prime}_{t_1}, \dots, \boldsymbol{Z}^{\star\prime}_{t_n})'$ can be expressed as

$$f(\boldsymbol{z}) = (\prod_{\boldsymbol{i=1}}^{\boldsymbol{n}} |\Sigma_{t_i|t_{i-1}}|^{-\frac{1}{2}}) \times \exp[\sum_{i=1}^{n} \boldsymbol{R}'\Sigma^{-1}_{t_i|t_{i-1}}\boldsymbol{R}] \tag{5}$$

with $\boldsymbol{R} = \boldsymbol{Z}^\star_{t_i} - \hat{\boldsymbol{Z}}^\star_{t_i|t_{i-1}}$ and $\Sigma_{t_i|t_{i-1}} = \mathbb{E}(\boldsymbol{R}\boldsymbol{R}')$.

The quantities $\hat{\boldsymbol{Z}}^\star_{t_i|t_{i-1}}$ and $\Sigma_{t_i|t_{i-1}}$ are directly determined from the recursive filtering calculations described in Equations 7.25-7.27, Chapter 7 of Reinsel (1997).

5 Criteria for model specification

To choose the most appropriate low order mixed seasonal VARIMA model based on the observations of a dataset, we usually appeal to model selection criteria such the Akaike Information Criterion (AIC) and/or the Bayesian information criterion (BIC). The AIC model selection criteria (Akaike, 1974, 1976) is given by

$$\begin{aligned} \text{AIC} &= n^{-1}(-2 \times \log(\textit{maximized likelihood}) + 2r) \\ &= \log(|\Sigma_r|) + 2rn^{-1} + \textit{constant}, \end{aligned} \quad (6)$$

where r denotes the number of parameters estimated by maximum likelihood (ML) in the sVARIMA model and Σ_r is the corresponding ML residual covariance matrix estimate of $\Sigma = \mathbb{C}ov(\boldsymbol{a}_t)$.

The BIC criteria (Schwarz, 1978) imposes more penalty for the number of estimated model parameters than does AIC and can be calculated by the formula

$$\text{BIC} = \log(|\Sigma_r|) + r\log(n)n^{-1}. \quad (7)$$

These model selection criteria are used to compare various models fitted by ML to the series. The fitted model that yields a minimum value for a given criteria is chosen (Reinsel, 1997).

6 Multivariate time series diagnostic checking

Under the null hypothesis that the model in (1) has been correctly identified, the residuals $\widehat{\boldsymbol{a}}_{tk} = \boldsymbol{Z}_{tk} - \widehat{\boldsymbol{Z}}_{tk},\ t = 1, \ldots, n$ of the fitted model may be estimated and used to test the null hypothesis that

$$\mathcal{H}_0 : \ \Gamma_\ell = \mathbf{0}, \ell = 1, 2, \ldots, m,$$

where $\Gamma_\ell = \mathbb{C}ov(\boldsymbol{a}_t, \boldsymbol{a}_{t-\ell})$ and m is the maximum lag that covers all lags, ℓ, of interest. The portmanteau test statistics are usually used to check the adequacy of the fitted VARIMA models.

Chitturi (1974); Hosking (1980) and Li and McLeod (1981) introduced multivariate portmanteau test statistics that can be seen as multivariate analogs of the well-known univariate portmanteau statistic proposed by Ljung and Box (1978). These test statistics may be expressed in terms of the residual autocovariances $\widehat{\Gamma}_\ell = n^{-1}\sum_{t=\ell+1}^{n} \widehat{\boldsymbol{a}}_t \widehat{\boldsymbol{a}}'_{t-\ell},\ \ell \geq 0$ as follows

$$\mathcal{Q}_m = n\sum_{\ell=1}^{m} tr(\widehat{\Gamma}'_\ell \widehat{\Gamma}_0^{-1} \widehat{\Gamma}_\ell \widehat{\Gamma}_0^{-1}), \quad (8)$$

where $tr(\cdot)$ denotes the trace of a matrix and $\widehat{\Gamma}_{-\ell} = \widehat{\Gamma}'_\ell$.

Recently, Mahdi and McLeod (2012) introduced a non-seasonal VARMA test statistic that extends the univariate portmanteau test proposed by Peña and Rodríguez (2006). The multivariate portmanteau test statistic of Mahdi and McLeod (2012) is

$$\mathcal{D}_m = -3n(2m+1)^{-1}\log|\widehat{\mathfrak{R}}_m|, \quad (9)$$

with

$$\widehat{\mathfrak{R}}_m = \begin{pmatrix} \mathbb{I}_k & \hat{R}_1 & \dots & \hat{R}_m \\ \hat{R}_{-1} & \mathbb{I}_k & \dots & \hat{R}_{m-1} \\ \vdots & \dots & \dots & \vdots \\ \hat{R}_{-m} & \hat{R}_{-m+1} & \dots & \mathbb{I}_k \end{pmatrix}, \tag{10}$$

$\hat{R}_\ell,\ \ell = 1, \dots, m$, being the residual autocorrelation matrix defined by Hosking (1980) as $\hat{R}_\ell = \hat{L}'\widehat{\Gamma}_\ell \hat{L}$, where $\hat{L}$ is the lower triangular Cholesky decomposition of $\widehat{\Gamma}_0^{-1}$.

Mahdi and McLeod (2012) derived the asymptotic distribution of their test statistic as a chi-square distribution with $3k^2m(m+1)(4m+2)^{-1} - p - q$ degrees of freedom, where k represents the dimension of the time series. In addition, Mahdi and McLeod (2012) proposed a Monte-Carlo version of $\mathcal{D}_m$ and provided a simulation comparison study to demonstrate that both methods (the asymptotic distribution of $\mathcal{D}_m$ and its Monte-Carlo version) are more powerful than Chitturi (1974); Hosking (1980) and Li and McLeod (1981) portmanteau test statistics with the correct size level.

Replacing the residual autocorrelations included in the Toeplitz matrix specified in (10), $\hat{R}_\ell$, by $\hat{R}_{\ell s}$ will extend $\mathcal{D}_m$ to test the seasonality (Mahdi, 2016). The multivariate seasonal portmanteau statistic to test whether the seasonal autocorrelations at multiple lags s of time series are different from zero as proposed by Mahdi (2016) is

$$\mathcal{D}_{ms} = -3n(2m+1)^{-1} \log |\widehat{\mathfrak{R}}_m(s)| \sim \chi^2_{3m(m+1)(4m+2)^{-1} - P - Q} \tag{11}$$

where $\widehat{\mathfrak{R}}_m(s)$ is the seasonal residual autocorrelations of the Toeplitz matrix appearing in (10) with $\hat{R}_\ell$ replaced by $\hat{R}_{\ell s}$.

7 The fitting of the seasonal VARIMA model

We now fit the seasonal vector ARIMA model for spatio-temporal rainfall data. We utilize the statistical software R (Mahdi et al., 2014; R Development Core Team, 2015; Tsay, 2016) to analyze the daily rainfall data obtained from the 12 Palestinian meteorological stations located across the five governorates of the Gaza Strip.

As the rainfall amount may vary from one station to another based on the weighted region area, we use the Thiessen polygons technique to compute the amount of the daily rainfall measured in each station based on its weighted area (Thiessen, 1911). The amount of the rainfall measured in Million Cubic Meter (MCM) over an area A [3] based on the Thiessen polygons method is given by:

$$R_v = R_a \times A \times 10^{-3}, \tag{12}$$

where R_v stands for the rainfall volume from region of area A in Million Cubic Meter (MCM) and R_a is the rainfall amount in Millimeters (Mm).

[3]See Figure 2 and Table 1 for the station network areas.

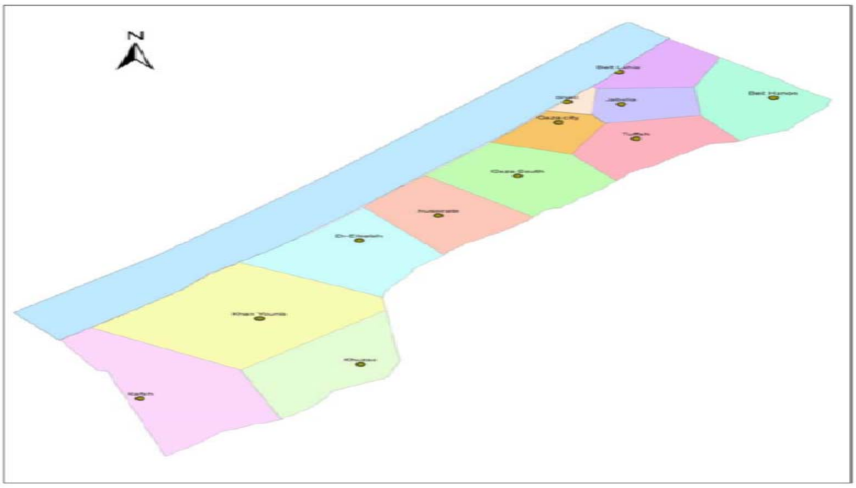

Figure 2: Rainfall Thiessen network of the Gaza strip.

Motivated by the fact that the geographical area of the Gaza Strip is small, so that the spatial-dependence within such a small geographic space is almost perfect, it is reasonable to restrict our analysis to the temporal-dependence. We compile the district information of the 12 meteorological stations into 5 groups (governorates) and do the analysis based on the amount of precipitation that has accumulated within the 5 governorates of the Gaza Strip. In this respect, the rainfall dataset from the Gaza Strip can be seen as a multivariate time series of dimension 5 so that the spatio-temporal dependence structure in this data can be explained by multivariate time series models. Then, we split the data into accumulated monthly data to obtain a multivariate time series of size 486×5.

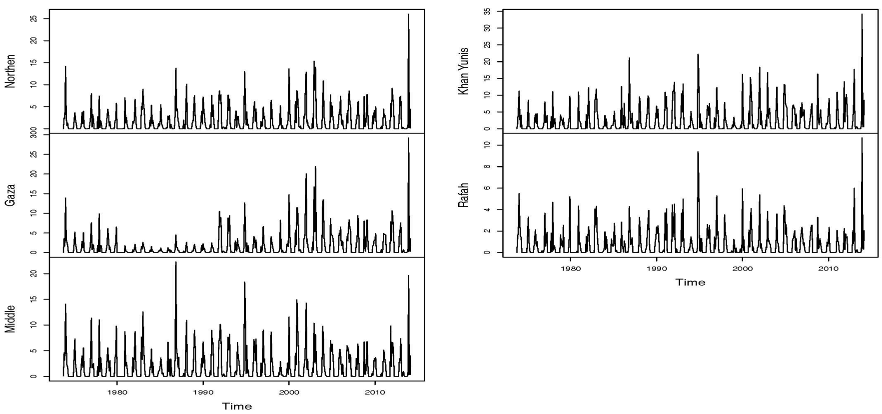

Figure 3: Monthly cumulative rainfall (MCM) in 5 governates of the Gaza strip.

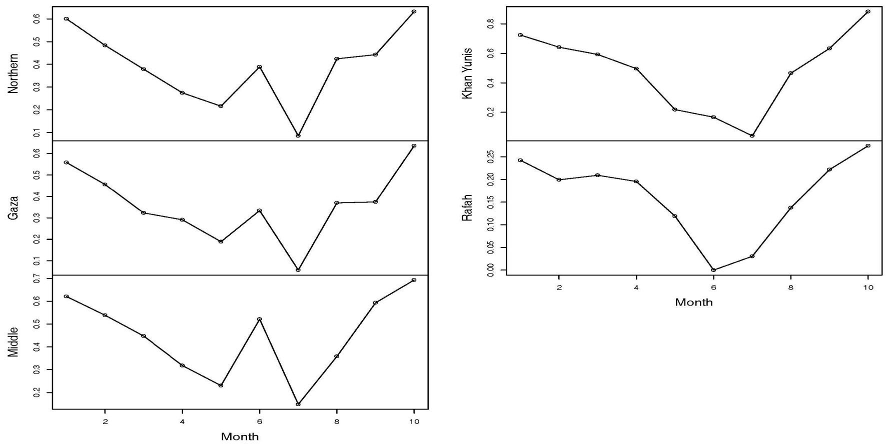

Figure 4: Means of rainfall (MCM) by month for the years 1973-2014 in 5 governates of the Gaza Strip.

Figure 3 shows the plots of the monthly cumulative rainfall amounts for the 5 governates in MCM . The patterns of peaks and troughs in Figure 3 suggest that the monthly seasonal rain effect is presents within the 42 years.

To detect seasonality, we plot the average of the accumulated rainfall of the observed months during the 42 years and the Autocorrelation Function (ACF) as seen in Figures 4 and 5, respectively, which clearly show that there exist monthly differences. Accordingly, we deseasonalized the series data using the first seasonal difference of lag 12 as seen in Figure 6. The graphs in Figures 6 and 7 suggest that the data has mean zero and no further differencing is needed.

Table 2: Information criteria values of suggested sVARIMA models

sVARIMA model	AIC	BIC
sVARIMA $(1, 0, 0) \times (1, 1, 0)_{12}$	3.3249	3.7638
sVARIMA $(0, 0, 1) \times (1, 1, 0)_{12}$	3.3130	3.7520
sVARIMA $(2, 0, 0) \times (0, 1, 0)_{12}$	4.4941	4.9330
sVARIMA $(0, 0, 2) \times (1, 1, 0)_{12}$	3.3377	3.9962

Then, we estimate the parameters of the sVARIMA model with the maximum likelihood method and select the best model among several fitted models (Table 2) using the proposed information criteria. The AIC and BIC suggest that the preferred model is the sVARIMA $(0, 0, 1) \times (1, 1, 0)_{12}$ model.

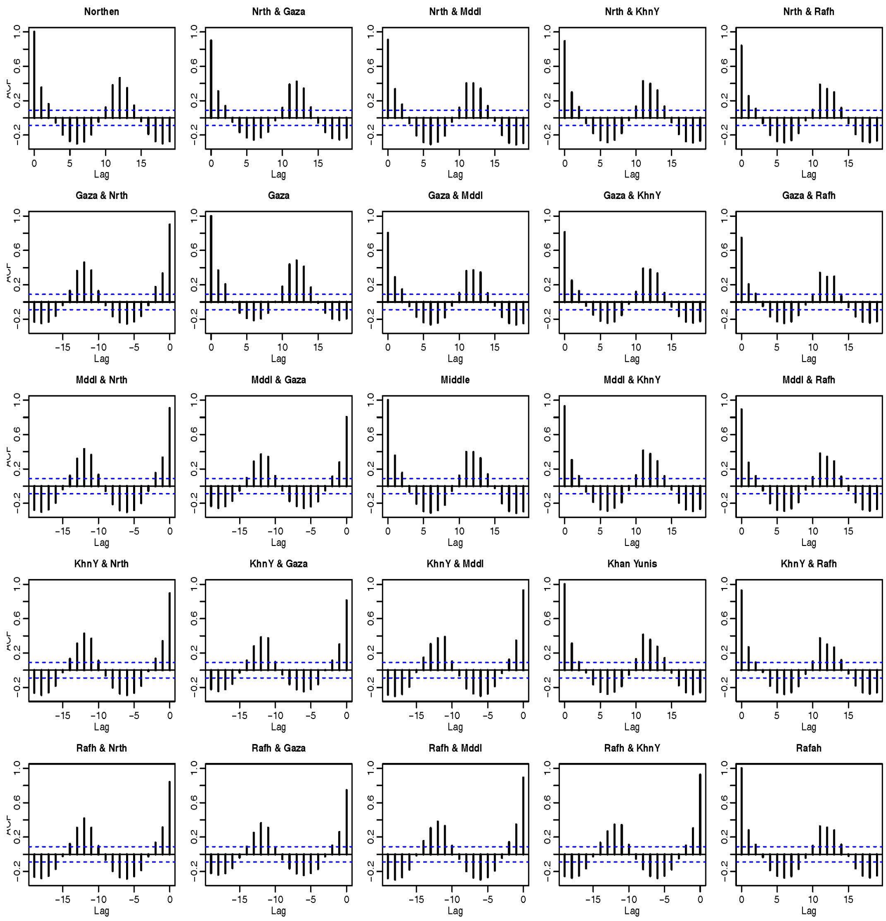

Figure 5: ACF of monthly cumulative rainfall (MCM) in 5 governates of the Gaza strip.

The ML estimates of the selected model parameters (with estimated standard errors in parenthesis) are

$$\widehat{\Phi}_1 = \begin{bmatrix} \underset{(0.101)}{-0.328} & \underset{(0.077)}{0.066} & \underset{(0.088)}{-0.189} & \underset{(0.077)}{0.171} & \underset{(0.176)}{-0.392} \\ \underset{(0.106)}{0.256} & \underset{(0.081)}{-0.422} & \underset{(0.093)}{-0.200} & \underset{(0.081)}{0.195} & \underset{(0.185)}{-0.594} \\ \underset{(0.122)}{0.112} & \underset{(0.091)}{0.009} & \underset{(0.106)}{-0.553} & \underset{(0.092)}{0.134} & \underset{(0.210)}{-0.340} \\ \underset{(0.157)}{0.142} & \underset{(0.117)}{0.086} & \underset{(0.136)}{-0.050} & \underset{(0.118)}{-0.420} & \underset{(0.269)}{-0.425} \\ \underset{(0.052)}{0.014} & \underset{(0.039)}{0.052} & \underset{(0.046)}{0.016} & \underset{(0.040)}{0.004} & \underset{(0.091)}{-0.601} \end{bmatrix}$$

and

$$\widehat{\vartheta}_1 = \begin{bmatrix} \underset{(0.114)}{-0.069} & \underset{(0.089)}{0.084} & \underset{(0.102)}{-0.066} & \underset{(0.082)}{0.107} & \underset{(0.133)}{0.041} \\ \underset{(0.127)}{-0.097} & \underset{(0.095)}{-0.034} & \underset{(0.107)}{-0.086} & \underset{(0.081)}{0.246} & \underset{(0.092)}{-0.001} \\ \underset{(0.143)}{0.182} & \underset{(0.105)}{-0.033} & \underset{(0.120)}{-0.095} & \underset{(0.101)}{-0.022} & \underset{(0.188)}{0.172} \\ \underset{(0.183)}{0.301} & \underset{(0.133)}{-0.126} & \underset{(0.156)}{-0.122} & \underset{(0.133)}{-0.070} & \underset{(0.245)}{0.308} \\ \underset{(0.061)}{0.138} & \underset{(0.045)}{-0.013} & \underset{(0.052)}{-0.068} & \underset{(0.045)}{-0.033} & \underset{(0.086)}{0.102} \end{bmatrix},$$

with estimated ML residual variance-covariance matrix

$$\Sigma_r = \begin{bmatrix} 6.093 & 5.513 & 6.173 & 7.745 & 2.349 \\ 5.513 & 6.763 & 5.702 & 7.243 & 2.187 \\ 6.173 & 5.702 & 8.716 & 10.135 & 3.078 \\ 7.745 & 7.243 & 10.135 & 14.644 & 4.378 \\ 2.349 & 2.187 & 3.078 & 4.378 & 1.625 \end{bmatrix},$$

whose determinant is 22.24.

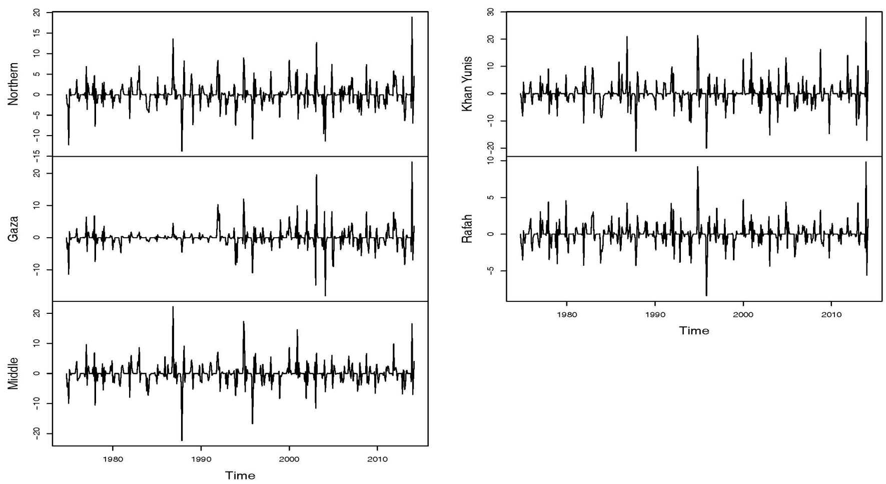

Figure 6: Deseasonalized monthly cumulative rainfall in 5 governates of the Gaza Strip.

As some of the coefficient estimates of $\widehat{\Phi}_1$ and $\widehat{\vartheta}_1$ are not significant, we can refine the fitted sVARIMA model by setting all the insignificant estimates to zero. The AIC and BIC applied to the refined model are 3.3036 and 3.5758, respectively, which suggest suggests this model is preferable to the full model.

The final selected model is

$$(\mathbb{I}_5 - \widehat{\Phi}_1 \boldsymbol{B}^{12})\dot{\boldsymbol{Z}}_{\boldsymbol{t}} = (\mathbb{I}_5 - \widehat{\vartheta}_1 \boldsymbol{B})\widehat{\boldsymbol{a}}_t, \tag{13}$$

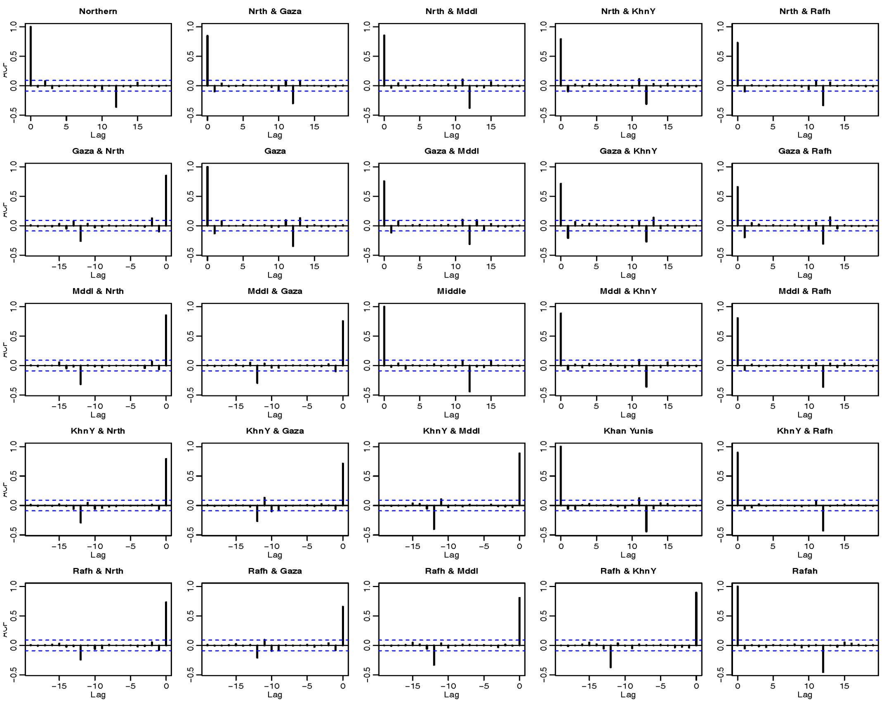

Figure 7: ACF of deseasonalized monthly cumulative rainfall in 5 governates of the Gaza strip.

i.e.,

$$\boldsymbol{Z}_t - \boldsymbol{Z}_{t-12} - \widehat{\Phi}_1 \boldsymbol{Z}_{t-12} - \widehat{\Phi}_1 \boldsymbol{Z}_{t-24} = \widehat{\boldsymbol{a}}_t - \widehat{\vartheta}_1 \widehat{\boldsymbol{a}}_{t-1}, \tag{14}$$

where $\dot{\boldsymbol{Z}}_{\boldsymbol{t}} = \boldsymbol{Z}_t - \boldsymbol{Z}_{t-12}$ represents the monthly cumulative rainfall measured in MCM in the 5 governates of the Gaza Strip at time t, $\widehat{\Phi}_1$ and $\widehat{\vartheta}_1$ are the respective ML estimates of the seasonal autoregressive and non-seasonal moving average parameter matrices (with estimated standard errors in parenthesis) and $\widehat{\boldsymbol{a}}_t$ denotes the residuals of the fitted model at time t with estimated ML residual variance-covariance matrix Σ_r, the ML estimates being

$$\widehat{\Phi}_1 = \begin{bmatrix} \underset{(0.064)}{-0.356} & \underset{(0.039)}{0.049} & \underset{(0.051)}{-0.170} & \underset{(0.051)}{0.143} & \underset{(0.137)}{-0.296} \\ \underset{(0.077)}{0.220} & \underset{(0.052)}{-0.430} & \underset{(0.065)}{-0.181} & \underset{(0.061)}{0.166} & \underset{(0.153)}{-0.498} \\ \underset{(0.058)}{0.055} & \underset{(0.000)}{0.000} & \underset{(0.047)}{-0.535} & \underset{(0.051)}{0.100} & \underset{(0.152)}{-0.205} \\ \underset{(0.058)}{0.104} & \underset{(0.000)}{0.000} & \underset{(0.000)}{0.000} & \underset{(0.047)}{-0.467} & \underset{(0.175)}{-0.246} \\ \underset{(0.000)}{0.000} & \underset{(0.016)}{0.043} & \underset{(0.000)}{0.000} & \underset{(0.000)}{0.000} & \underset{(0.046)}{-0.535} \end{bmatrix},$$

$$\widehat{\vartheta}_1 = \begin{bmatrix} \underset{(0.000)}{0.000} & \underset{(0.029)}{0.091} & \underset{(0.000)}{0.000} & \underset{(0.028)}{0.069} & \underset{(0.000)}{0.000} \\ \underset{(0.000)}{0.000} & \underset{(0.000)}{0.000} & \underset{(0.050)}{-0.054} & \underset{(0.044)}{0.185} & \underset{(0.000)}{0.000} \\ \underset{(0.051)}{0.181} & \underset{(0.000)}{0.000} & \underset{(0.000)}{0.000} & \underset{(0.000)}{0.000} & \underset{(0.103)}{0.006} \\ \underset{(0.078)}{0.283} & \underset{(0.045)}{-0.094} & \underset{(0.000)}{0.000} & \underset{(0.000)}{0.000} & \underset{(0.140)}{0.036} \\ \underset{(0.029)}{0.125} & \underset{(0.000)}{0.000} & \underset{(0.020)}{-0.043} & \underset{(0.000)}{0.000} & \underset{(0.053)}{0.014} \end{bmatrix},$$

and

$$\Sigma_r = \begin{bmatrix} 6.262 & 5.676 & 6.355 & 7.947 & 2.410 \\ 5.676 & 6.931 & 5.861 & 7.406 & 2.239 \\ 6.355 & 5.861 & 8.947 & 10.412 & 3.162 \\ 7.947 & 7.406 & 10.412 & 14.993 & 4.482 \\ 2.410 & 2.239 & 3.162 & 4.482 & 1.659 \end{bmatrix},$$

whose $\det(\Sigma_r)$ is 23.88.

Finally, we applied the Monte-Carlo seasonal and nonseasonal Mahdi and McLeod test statistic to check the adequacy of the proposed model (Mahdi, 2016; Mahdi and McLeod, 2012). The results presented in Table 3 suggest that this model is adequate.

Table 3: Monte-Carlo seasonal and nonseasonal portmanteau test statistics

Lags	seasonal tests		nonseasonal tests	
	statistic	p-value	statistic	p-value
5	77.18547	0.9810190	112.5859	0.26773227
10	123.20723	1.0000000	160.4557	0.98801199
15	293.18354	0.6953047	333.2310	0.14685315
20	430.40587	0.2437562	471.8811	0.09497502

8 Conclusion

In this paper, we implemented a multivariate time analysis to model monthly cumulative rainfall data obtained from gauge stations located in the Gaza Strip. Our analysis suggests that the sVARIMA $(0,0,1)\times(1,1,0)_{12}$ model is the most appropriate for forecasting rainfall amounts a month-ahead in the governorates of the Gaza Strip. It can be concluded that the sVARIMA model applied to the current and previous rainfall data provides a useful method to forecast rainfall data. It could prove very helpful to environmental scientists, especially those based in Middle Eastern countries, to utilize our model to forecast the amount of rainfall as a preliminary guideline toward short and long-term sustainable water resources management.

Acknowledgements

The authors would like to extend their gratitude and appreciation to Dr. Yasser El-Nahhal, Associate Professor of Environmental Chemistry and Eco-toxicology, The Islamic University-Gaza, Palestine Environmental Science for providing the historical daily rainfall data. The authors thank the Editor, the Co-Editor, and the three anonymous referees for helpful comments.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19(6):716–723.

Akaike, H. (1976). *Canonical Correlation Analysis of Time Series and the Use of an Information Criterion*, chapter In System Identification, pages 27–96. Academic Press, New York. Advances and Case Studies, eds. R. K. Mehra and D. G. Lainiotis.

Al-Khatib, M. and Al-Najar, H. (2011). Hydro-geochemical characteristics of groundwater beneath the gaza strip. *Journal of Water Resources and Protection*, 3(5):341–348.

Al-Najar, H. (2011). The integration of fao-cropwat model and gis techniques for estimating irrigation water requirement and its application in the gaza strip. *Natural Resources*, 2:146–154.

Ansley, C. F. and Kohn, R. (1983). Exact likelihood of vector autoregressive-moving average process with missing or aggregated data. *Biometrika*, 70:275–278.

Ansley, C. F. and Kohn, R. (1995). A scale model of multivariate rainfall time series. *Journal of Hydrology*, 168(1-4):1–15.

Banu, S. S., Manjula, S., Priya, S. S., Devi, V. Y., and Thangamani, M. (2016). Predictive analysis of rainfall data to help the farmers. *International Journal of Advanced Research in Computer Science and Software Engineering*, 6(3):93–115.

Bras, R. L. and Rodríguez-Iturbe (1985). *Random functions and hydrology.* Addison-Wesley, Reading, MA, 4th edition.

Chitturi, R. V. (1974). Distribution of residual autocorrelations in multiple autoregressive schemes. *Journal of the American Statistical Association*, 69(348):928–34.

Cox, D. R. (1981). Statistical analysis of time series: some recent developments (with discussion). *Scandinavian Journal of Statistics*, 8:93–115.

Cristiano, E., claire ten Veldhius, M., and van de Giesen, N. (2016). Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas – a review. *Hydrology and Earth System Sciences.*

El-Nahhal, Y., Abadsa, M., and Affifi, S. (2013a). Adsorption of diuron and linuron in gaza soils. *American Journal of Analytical Chemistry*, 4(7):94–99.

El-Nahhal, Y., Awad, Y., and Safi, J. (2013b). Bioremediation of acetochlor in soil and water systems by cyanobacterial mat. *International Journal of Geosciences*, 4(6):880–890.

El-Nahhal, Y., Tubail, K., Safi, M., and Safi, J. (2013c). Effect of treated wastewater irrigation on plant growth and soil properties in Gaza Strip Palestine. *American Journal of Plant Science*, 4(9):1736–1743.

Harvey, A. C. and Pierse, R. G. (1984). Estimating missing observations in economic time series. *Journal of the American Statistical Association*, 79:125–131.

Hosking, J. R. M. (1980). The multivariate portmanteau statistic. *Journal of American Statistical Association*, 75(371):602–607.

Jones, R. H. (1980). Maximum likelihood fitting of arma models to time series with missing observations. *Technometrics*, 22:389–395.

Li, W. K. and McLeod, A. I. (1981). Distribution of the residual autocorrelation in multivariate arma time series models. *Journal of the Royal Statistical Society, Series B*, 43(2):231–239.

Ljung, G. and Box, G. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65:297–303.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, New York.

Mahdi, E. (2016). Portmanteau test statistics for seasonal serial correlation in time series models. *SpringerPlus*, 5(1485).

Mahdi, E. and McLeod, A. I. (2012). Improved multivariate portmanteau test. *Journal of Time Series Analysis*, 33(2):211–222.

Mahdi, E., Xiao, K. J., and McLeod, A. I. (2014). portes: Portmanteau tests for univariate and multivariate time series models. R package version 2.1-3.

Melloul, A. J. and Collin, M. (2000). Sustainable groundwater management of the stressed coastal aquifer in the gaza region. *Hydrological Sciences-Journal-des Sciences Hydrologiques*, 45(1):147–159.

Palestinian Water Authority (2013). Status report of water resources in the occupied state of palestine-2012. Technical report, Annual Water Resources Status Report 2012.

Peña, D. and Rodríguez, J. (2006). The log of the determinant of the autocorrelation matrix for testing goodness of fit in time series. *Journal of Statistical Planning and Inference*, 8(136):2706–2718.

R Development Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reinsel, G. C. (1997). *Elements of Multivariate Time Series Analysis*. Springer-Verlag, New York, 2nd edition.

Safi, J., Awad, Y., and El-Nahhal, Y. (2014). Bioremediation of diuron in soil and by cyanobacterial mat. *American Journal of Plant Sciences*, 5(8):1081–1089.

Salas, J., Delleur, J., Yevjevich, V., and Lane, W. (1980). *Applied Modeling of Hydrologic Time Series*. Water Resources Publication, Littleton, Colorado, 1st edition.

Schwarz, G. (1978). Estimation the dimension of a model. *Annals of Statistics*, 6:461–464.

Shomer, B., Mueller, G., and A., Y. (2004). Potential use of treated wastewater and sludge in agricultural sector of the gaza strip. *Clean Technologies and Environmental Policy*, 6(2):128–137.

Soltani, S., Modarresa, R., and Eslamianb, S. S. (2007). The use of time series modeling for the determination of rainfall climates of Iran. *International Journal of Climatology*, 27:819–829.

Stern, R. D. and Coe, R. (1984). A model fitting analysis of daily rainfall data. *Journal of the Royal Statistical Society, Series A*, 147(1):1–34.

Thiessen, A. (1911). Precipitation averages for large areas. *Monthly Weather Review*, 39(7):1082–1084.

Tsay, R. S. (2014). *Multivariate Time Series Analysis: With R and Financial Applications*. John Wiley and Sons, Inc., New York.

Tsay, R. S. (2016). Mts: All-purpose toolkit for analyzing multivariate time series (mts) and estimating multivariate volatility models. R package version 0.33.

Wincek, M. A. and Reinsel, G. C. (1986). An exact maximum likelihood estimation procedure for regression-rma time series models with possibly nonconsecutive data. *Journal of the Royal Statistical Society, Series B*, 48:303–313.