إستخدام نموذج (ARIMA) في استكمال السلسلة الزمنية لقيم التبخر في مدينة بغداد

ذر انتصار بكر

جامعة ديالى ,كلية العلوم ,قسم الفيزياء

الملخص

الهدف من هذا البحث هو دراسة إمكانية استخدام نموذج (ARIMA) الذي يعني نماذج الانحدار الذاتي المتكاملة مع المتوسطات المتحركة Auto Regressive Integrated Moving Average في معالجة القطوعات في السلاسل الزمنية لبيانات التبخر وذلك باستخدام السلسلة الزمنية للمعدلات الشهرية لقيم التبخر للفترة من (1971-2000). وقد تبين أن نموذج (ARIMA) يمكن أن يستخدم في معالجة القطوعات في السلاسل الزمنية لبيانات التبخر وخصوصا عند زيادة حجم البيانات التي تسبق القطع في السلسلة والتي تستخدم كمدخلات لنموذج (ARIMA). فعندما تكون مدخلات النموذج (348) شهرا تكون قيم ($MAE,RMSE,R^2$)هي (0.0978, 0.140574, 0.997544) على التوالي.وتتناقص دقة المعالجة كلما قل حجم البيانات التي تسبق القطع. أما إذا كانت مدخلات النموذج (120) شهرا كانت قيم كل من ($MAE,RMSE,R^2$) هي (0.3055, 1.079692, 0.889669) على التوالي .

الكلمات المفتاحية: السلاسل الزمنية,التبخر, ARIMA

Using(ARIMA) model in complete Time Series for evaporation value in Baghdad City

Dher I. Bakr

Diyala university , College of science , Physics department

Received 20 January 2014 ; Accepted 4 Febraury 2014

Abstract

The aim of the present work is to use (ARIMA) model (Auto Regressive Integrated Moving Average) in the treatment of the cuts in time series of evaporation data. Time series of monthly average evaporation data (1971-1999) has implemented in our study .It has been observed that the above mentioned model can be used and operated in the treatment of cuts in particular the increased of pre cuts data as when the input data (348) month the value of (MAE,RMSE, R^2) is (0.0978, 0.140574, 0.997544) respectively and hence decrease the accuracy with decrease of data. It is also noted that if the input (120) month (MAE,RMSE, R^2) is (0.3055, 1.079692, 0.889669) respectively.

Key words: Time series, Evaporation ,ARIMA

المقدمة

يعتبر التبخر من العناصر الجوية المهمة ويمكن ان يعرف على انه فقدان جزيئات الماء من السطح المائي إلى الغلاف الجوي[1]. تشكل عملية التبخر إحدى المركبات المهمة للدورة الهيدرولوجية(دورة المياه في الطبيعة) لذلك فقد اجتذبت هذه العملية اهتمام الكثير من الباحثين ولاسيما في المناطق التي تشح فيها الموارد المائية بما يتطلب الأمر الحفاظ على هذه الثروة الحيوية [2].ولأجل القيام بالدراسة حول هذا العنصر يجب توفر سلسلة زمنية لقيم هذا العنصر ولفترة كافية بحيث يمكن ان نحلل تلك السلسلة وعلى هذا الأساس يتم بناء التصورات واتخاذ القرارات اللازمة.

فيمكن أن تعرف السلسلة الزمنية بأنها مجموعة من المشاهدات حول ظاهرة ما أخذت بترتيب زمني معين عادة ما تكون فيه تساوي الفترات الزمنية مثل الساعات ,الأيام ,الأشهر أو السنوات [3]. وأيضا يمكن تعريفها بأنها مجموعة من القيم المتتالية منظمة خلال فترة زمنية معينة وهذه المشاهدات يتم تسجيلها خلال الفترة وحسب فترات متتالية وعادة ما تكون هذه الفترات الزمنية متساوية من حيث الطول [4].أما رياضياتيا فيمكن أن نعرف السلسة الزمنية على أنها متتابعة من المتغيرات العشوائية معرفة ضمن فضاء الاحتمالية متعددة المتغيرات ومؤشرة بالدليل (t) والذي يعود إلى مجموعة دليلية (T) ويرمز للسلسلة الزمنية عادة $\{X(t), t \in T\}$ او اختصارا $X(t)$ وتتكون من متغيرين احدهما توضيحي وهو متغير الزمن والأخر متغير الاستجابة وهو قيمة الظاهرة المدروسة ويمكن التعبير عنها رياضياتيا بـ ($X(t) = f(t)$) وأيضا يمكن تمثيل السلاسل الزمنية بيانيا[5].

وتتألف السلاسل الزمنية من عدة عناصر هي :-

1- الاتجاه العام والمقصود به الحركة المنتظمة للسلسلة عبر فترة زمنية طويلة نسبيا وعادة يعتبر من أهم العناصر المكونة للسلسلة الزمنية ودائما ما يكون هو العنصر الوحيد في بناء التوقعات المستقبلية [6].

2- العامل الموسمي وهو الذي يمثل التغيرات والتذبذبات الموسمية أو الفصلية الناتجة عن التغيرات في الفصول بسبب تأثير عوامل خارجية وهي تتم غالبا بطريقة منتظمة بشكل دورات لا يزيد طولها عن أسبوع أو شهر أو فصل أي أنها تمثل التغيرات المتشابهة التي تظهر في الأسابيع أو الأشهر أو الفصول المتناظرة خلال الفترات الزمنية المحتملة [7].

3- التغيرات الدورية وهي التغيرات التي تطرأ على السلسلة الزمنية والتي قد تحدث في فترات زمنية منتظمة أو غير منتظمة [8].

4- التغيرات العشوائية وهي التغيرات الناتجة عن عوامل عشوائية خارج نطاق السيطرة وهذه التغيرات لا يمكن التنبؤ بها لكن يمكن إظهار تأثيرها كتذبذبات صغيرة في السلسلة الزمنية [9].

نموذج ARIMA (p, d, q)

يقصد بنموذج (ARIMA) تلك المنهجية التي طبقها كل من George Box و Gwilyn Jenkins على السلاسل الزمنية سنة 1970 وتعتمد على ثلاثة أجزاء هي [10] :-

أ- نموذج الانحدار الذاتي Auto Regressive(AR) ويشير إلى إن القيم الحالية للمتغير Y تعتمد على قيم المتغير السابقة والذي يكتب بالشكل التالي [11]:-

$$Y_T = B_o + B_1 Y_{T-1} + B_2 Y_{T-2} + \cdots\cdots + B_p Y_{T-p} + e_T \qquad (1)$$

حيث ان :-

Y_T : قيم المتغير Y المتنبأ بها, $Y_{T-1}, Y_{T-2}, Y_{T-p}$ قيم المتغير Y المتأخرة زمنيا خلال الفترة T, B_o, B_1, B_2, B_p : معاملات معادلة الانحدار .

ب- نموذج المتوسط المتحرك Moving Average(MA) ويمكن ان يكتب بالصيغة التالية [12] :-

$$Y_T = W_o + e_T - W_1 e_{T-1} - W_2 e_{T-2} \cdots\cdots - W_q e_{T-q} \qquad (2)$$

حيث ان:

Y_T : قيم المتغير Y المتنبأ بها, $e_{T-1}, e_{T-2}, e_{T-q}$: تمثل المتأخرة للبواقي من تقدير المتغير , e_T : المتغير العشوائي , W_o, W_1, W_2, W_q : تمثل الأوزان.

جـ – نموذج الانحدار الذاتي والمتوسط المتحرك Auto Regressive Moving Average(ARMA) والذي يمثل الجمع بين النموذجين السابقين بنموذج واحد وبالشكل التالي:-

$$Y_T = B_o + B_1 Y_{T-1} + B_2 Y_{T-2} + \cdots\cdots + B_p Y_{T-p} + e_T + W_o + e_T - W_1 e_{T-1} - W_2 e_{T-2} \cdots\cdots - W_q e_{T-q} \qquad (3)$$

ويشار إلى هذا النموذج بـ ARMA من الرتبة (p, q) حيث يشير p إلى رتبة الانحدار الذاتي ويشير q إلى رتبة المتوسط المتحرك.

ويتطلب تقدير النموذج أن تكون السلسلة الزمنية مستقرة (Stationary) ويقصد بالاستقرارية من الناحية الإحصائية أن يكون التباين والوسط الحسابي ثابتين [13],أما إذا كانت السلسلة غير مستقرة فيجب تحويلها إلى مستقرة عن طريق اخذ الفروق ويعتبر عدد مرات الفروق المطلوبة لتحويلها الى مستقرة هي درجة تكامل السلسلة . وفي هذه الحالة يتحول النموذج من ARMA إلى ARIMA ويعني نموذج الانحدار الذاتي والتوسط المتحرك المتكامل والذي يتصف بثلاثة رتب هي :رتبة الانحدار(p) ,رتبة التكامل (d) ورتبة المتوسط المتحرك (q).وتكتب ARIMA(p ,d, q).

الجانب العملي

إن القيم المعتمدة هي المعدلات الشهرية لقيم التبخر بوحدة (ملم) وان هذه البيانات في هذا البحث أخذت للسلسلة الزمنية للفترة (1971-2000)لمدينة بغداد التي تقع على خط عرض (N^o 33.34) وعلى خط طول (E^o 44.45) وبارتفاع (34m)عن مستوى سطح البحر والتي تم الحصول عليها من الهيئة العامة للأنواء الجوية والرصد الزلزالي في بغداد.

إن الكثير من التطبيقات التي تعتمد على السلاسل الزمنية والتي تعتبرها أساسا في بناء القرارات او تعتبرها أساسا في التنبؤ عن أي ظاهرة او تطور في تلك الظاهرة وخصوصا العناصر الجوية تشترط ان تكون تلك السلسلة متجانسة وان تكون متصلة ولا تحوي على قطوعات في البيانات وان هذه القطوعات لا يمكن إهمالها وذلك بسبب تغير التسلسل الزمني للبيانات لذلك توجب إيجاد طريقة لمعالجة تلك القطوعات في السلسلة الزمنية وجعلها متكاملة البيانات وبأقل نسبة خطأ وذلك لإتمام الإفادة المثلى من السلسلة الزمنية .

ففي هذا البحث تم استخدام نموذج (ARIMA) في التنبؤ عن القيم المفقودة في السلسلة الزمنية للتبخر وذلك بالاعتماد على قيم التبخر للفترة التي تسبق القطع في السلسلة .فكان العمل على البيانات بطريقة افتراض قطع في السلسلة لمدة سنة واحدة (12) شهرا في أخر السلسلة الزمنية والمؤلفة من (360) شهرا فتكون البيانات التي تسبق القطع والتي تستخدم كمدخلات لنموذج (ARIMA)هي(348) شهرا ونقوم بتخمين القيم المقطوعة ومقارنتها مع القيم الحقيقية (التي تم قطعها من قبلنا) ومن ثم نجعل القطع في السلسة الزمنية للسنة (قبل الأخيرة) وبذلك تكون مدخلات (ARIMA)هي (336) شهرا وهكذا يستمر العمل بتحريك القطع المفترض وتخمينه ومن ثم مقارنته مع القيم الحقيقية,واحتساب قيم الارتباط و الإحصائيات التي تعتبر معايير للخطأ ونستمر بتحريك القطع للسنة (11) من السلسلة الزمنية أي أن تكون قيم المدخلات لمدة (120) شهرا. فعندها يتم التخمين للقيم المقطوعة وبنتائج جيدة فمن السهل والمنطقي استخدام النموذج في معالجة قطوعات لاحقة للفترة المخمنة ولنفس الفترة او اقل لان العناصر الجوية تتذبذب على مدار السنة بشكل متشابه تقريبا كذلك يتم تغيير قيم

كل من (p, d, q) للنموذج في كل اختبار لحين الوصول لأفضل قيم لـ (p, d, q) محققة بذلك أفضل قيم للإحصائيات. فالإحصائيات المستخدمة في هذه الدراسة هي :-

1- معامل الارتباط (r) Correlation Coefficient وهو مقياس لقوة الارتباط بين القيم الحقيقية والقيم المخمنة ويمكن أن يحسب من العلاقة الآتية:-

$$r = \sqrt[2]{R^2} \qquad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - Y_T)^2}{\sum_{i=1}^{n}(Y_T)^2} \qquad (5)$$

إذ إن :

R^2 : الارتباط (معامل التحديد)وهي قوة العلاقة بين متغيرين ويمثل التقارب بين القيمتين وليس له علاقة بصحة القيم المخمنة أو الحقيقية, وله قيم تتراوح بين(1-) و(1+)

Y_T: القيمة المخمنة, y_i : القيمة الحقيقية, y : المتوسط الحسابي للقيم, n : عدد القيم

2- الجذر المتوسط لمربع الخطأ Root Mean Square Error (RMSE) يمكن أن يعرف على انه قيمة الخطأ لمعدل الجذر التربيعي ويمكن أن يحسب من العلاقة الاتية:-

$$RMSE = \frac{1}{n}\left[\sum_{i=1}^{n}(y_i - Y_T)^2\right]^{\frac{1}{2}} \qquad (6)$$

3- متوسط الخطأ المطلق Mean Absolute Error (MAE) هو القيمة المطلقة للخطأ لمعدل الانحراف ويتم حسابه من المعادلة الآتية [14]:-

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - y_T| \qquad (7)$$

النتائج والمناقشة

بالاعتماد على طريقة العمل وعند افتراض القطع في السلسلة لبيانات التبخر في السنة الأخيرة (12) شهرا فتكون مدخلات النموذج (ARIMA)هي(348) شهرا نلاحظ أن ألـ (12) شهرا التي تم تخمينها وبالاعتماد على قيم الإحصائيات (الارتباط

ومعايير الخطأ) حيث أن قيم كل من($MAE,RMSE,R^2$)هي (0.0978,0.14057,0.9975) وهي أفضل من القيم (12) شهرا المخمنة عندما تكون قيم المدخلات للنموذج(ARIMA) هي(338) شهرا التي فيها قيم كل من ($MAE,RMSE,R^2$) هي (0.134,0.36537,0.9814) وهذه الحالة هي أفضل من ألـ (12) شهرا المخمنة عندما تكون مدخلات النموذج (ARIMA) هي(326) شهرا ويستمر التناقص في دقة المعالجة لحين الوصول الى قيم ($MAE,RMSE,R^2$) والتي هي(0.889669, 1.079692, 0.3055)عندما تكون البيانات الداخلة للنموذج هي(120) شهرا وهذا مبين في الجدول (1). مع ملاحظة ان الاختلاف او التغير في قيم الإحصائيات لا يكون بشكل منتظم مثلما يكون التناقص في بيانات الإدخال وذلك لان التغير والتذبذب في قيم مدخلات النموذج لا تكون بشكل منتظم وإنما التغير يخضع لعدة عوامل جوية و جغرافية تؤثر في قيم بيانات الإدخال هذا بالإضافة الى تغيرات مفاجئة تؤثر وتتأثر في ذلك العنصر زيادة على ذلك حجم البيانات الداخلة.

بصورة عامة إن أي نموذج يستخدم للتخمين والذي يعتمد على السلسلة الزمنية من البيانات السابقة للعنصر المراد تخمينه تكون نتائج التخمين أفضل كلما زادت قيم المدخلات للنموذج .

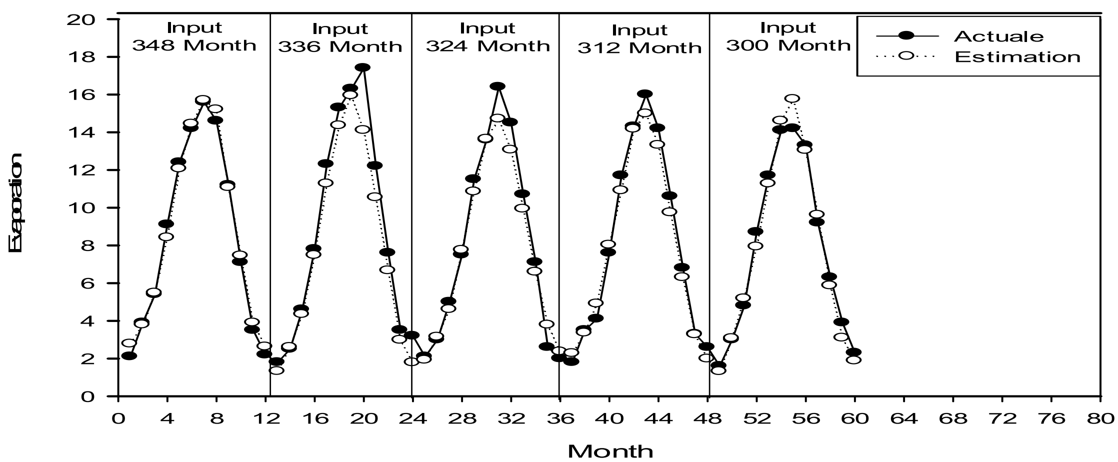
جدول(1) يبين تغير الارتباط وقيم الإحصائيات بتغير حجم المدخلات

Input Month	p,d,q	RMSE	MAE	R^2	Input Month	p,d,q	RMSE	MAE	R^2
348	5,0,2	0.140574	0.0978	0.997544	228	5,0,1	0.552368	0.2096	0.879916
336	5,0,2	0.365376	0.134	0.981469	216	5,0,1	0.540194	0.2428	0.95872
324	6,0,2	0.235425	0.1074	0.966324	204	6,0,1	0.365623	0.1567	0.980221
312	5,0,1	0.184304	0.0966	0.994989	192	5,0,2	0.473004	0.1582	0.973936
300	4,0,1	0.184904	0.0894	0.989198	180	6,0,1	0.284556	0.1041	0.991978
288	7,0,1	0.220274	0.128	0.987592	168	6,0,2	0.490739	0.1557	0.975103
276	5,0,1	0.484412	0.1525	0.973936	156	5,0,2	0.559196	0.1661	0.966604
264	7,0,1	0.375424	0.154	0.986708	144	6,0,1	0.292835	0.1251	0.987598
252	7,0,2	0.264797	0.2125	0.98668	132	5,0,2	0.354074	0.0984	0.985077
240	5,0,1	0.286108	0.1043	0.993399	120	7,0,2	1.079692	0.3055	0.889669

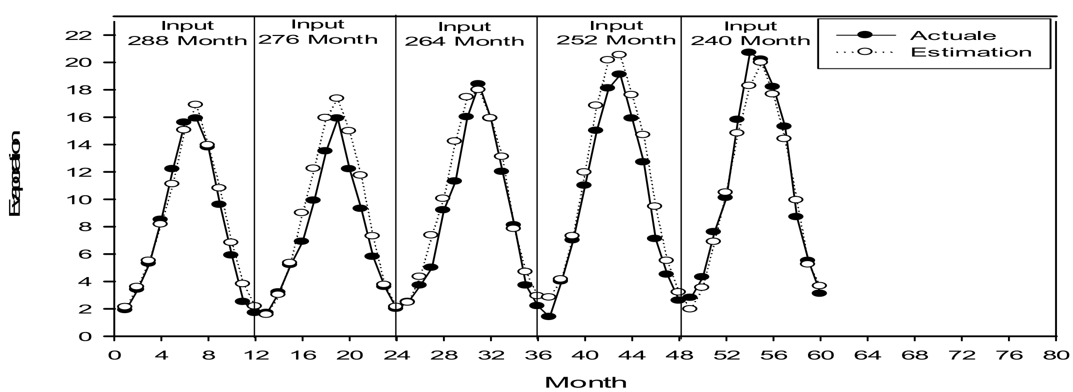
إن العناصر الجوية (والتي من ضمنها التبخر) بصورة عامة تكون متذبذبة بشكل منتظم وعلى مدار السنة .ولذلك فان التخمين (البيانات المخمنة) يأخذ نفس شكل السلسلة الزمنية الأولية وكما مبين في المخططات(1,2,3,4) . حيث يلاحظ ان التقارب يكون اكبر ما يمكن بين القيم الحقيقية والقيم المعالجة للفترة التي يسبقها حجم اكبر من البيانات من التي يسبقها حجم اقل .فيكون اكبر تقارب عندما يكون مدخلات النموذج (348) شهرا كما في المخطط (1) وتتناقص بتناقص حجم البيانات المدخلة الى نموذج (ARIMA) التي تسبق القطع كما في المخطط (2,3) حيث نلاحظ زيادة في الاختلاف بين القيم

المخمنة والقيم الحقيقية بسبب التناقص في بيانات الداخلة للنموذج .إلى أن يصل إلاختلاف الى اكبر ما يمكن عندما تكون مدخلات النموذج (120) شهراكما في المخطط(1).

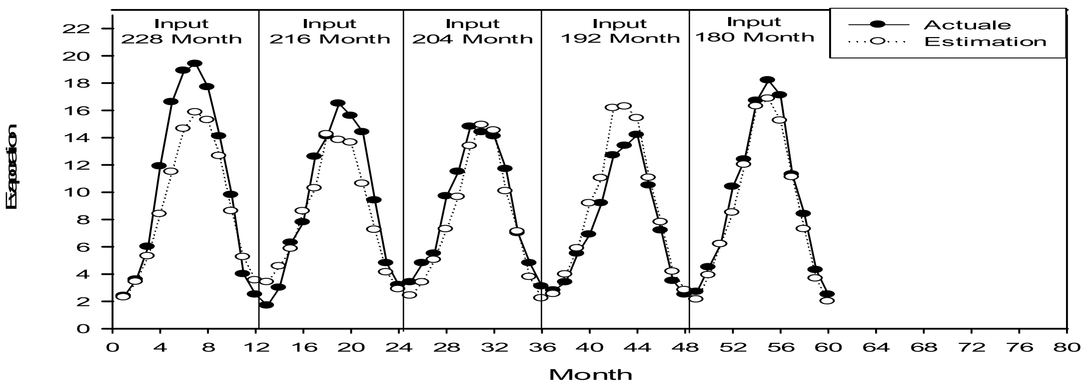
إن هذه المعالجات للبيانات المفقودة او المقطوعة للسلسلة الزمنية تمت على افتراض ان مدة القطع هي (12) شهرا أما إذا كانت فترة القطع في السلسلة اقل من تلك المدة فيمكن معالجتها وبنفس الطريقة .



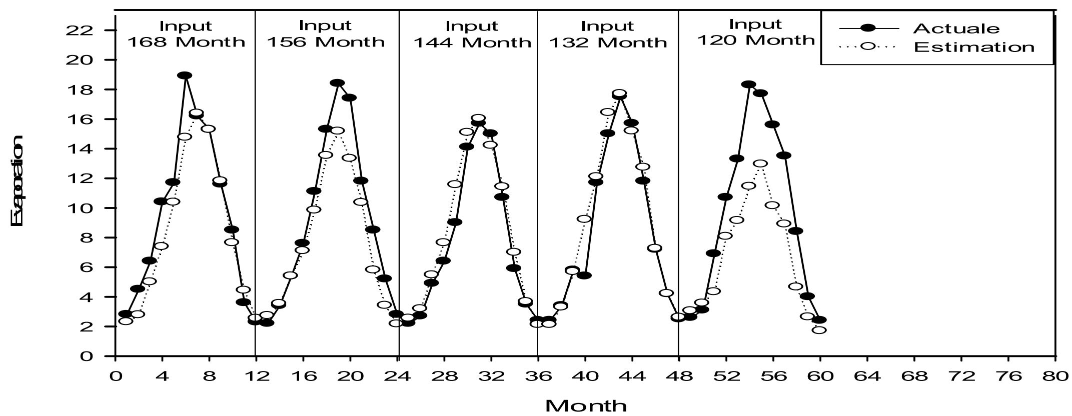
المخطط (1) يوضح العلاقة بين القيم الحقيقة المخمنة للقطع بتناقص المدخلات من (300-348) شهرا



المخطط (2) يوضح العلاقة بين القيم الحقيقة المخمنة للقطع بتناقص المدخلات من (240-288) شهرا



المخطط (3) يوضح العلاقة بين القيم الحقيقة المخمنة للقطع بتناقص المدخلات من (180-228) شهرا



المخطط (4) يوضح العلاقة بين القيم الحقيقة المخمنة للقطع بتناقص المدخلات من (120-168) شهرا

الاستنتاجات

1. إن نتائج معالجة القطوعات باستخدام نموذج (ARIMA) تكون أدق واقرب للواقع كلما زاد حجم البيانات التي تسبق القطع في السلسلة الزمنية والتي تكون مدخلات للنموذج.
2. إن نموذج (ARIMA) يعتمد على البيانات السابقة للعنصر المراد معالجة القطع في بياناته فقط ولا يعتمد على بيانات العناصر المرتبطة به كما الحال في العديد من النماذج الرياضياتية.
3. التغير في قيم الارتباط ومعايير الخطأ لا يكون بشكل منتظم كالتناقص المنتظم لبيانات الإدخال لنموذج (ARIMA).
4. بناءا على قيم الارتباط ومعايير الخطأ يمكن استخدام نموذج (ARIMA) كإحدى طرق معالجة القطوعات في بيانات السجلات المناخية لعنصر التبخر لإتمام الاستفادة من هذه السجلات بالطريقة المثلى من قبل العاملين في هذا الحقل .

المصادر

1. Surinder Deswal, and Mahesh Pal (2008)"Artificial Neural Network based Modeling of Evaporation Losses in Reservoirs" World Academy of Science ,Engineering and Technology 39, p 279.
2. العاني ,افتخار عبد الجواد عبد الحميد (2007) "أنموذج شبكة عصبية اصطناعية لتخمين التبخر – نتح المرجعي اليومي لمنطقة الموصل"
3. عوض منصور وعزام صبري"مبادئ الإحصاء" دار الصفاء للنشر والتوزيع,عمان,الطبعة الأولى,2000 ص 239.
4. نصيب رجم"الإحصاء التطبيقي"دار العلوم للنشر والتوزيع,الجزائر,ص 37 .
5. ألعبيدي عبد الغفور جاسم سالم (1989)"تحليل و نمذجة السلسلة الزمنية لدرجات الحرارة في مدينة الموصل"
6. Jean Pierre Vedrines "technique quantitative de gestation" librairie Velbert ,Paris,1985,p17.
7. البلداوي ,عبد المجيد عبد الحميد "الإحصائيات للعلوم الإدارية والتطبيقية"دار الشروق,عمان,الطبعة الأولى,ص563.
8. الطائي,فاضل عباس (2009)"التنبؤ والتمهيد للسلاسل الزمنية باستخدام التحويلات مع التطبيق.
9. المشهداني , محمود حسن والدليمي,محمد مناجد عيفان (1985)"طرق لإحصاء الأرقام القياسية والسلاسل الزمنية"جامعة بغداد ,كلية الإدارة والاقتصاد.
10. العاني,احمد حسين بتال"استخدام نماذج ARIMA في التنبؤ الاقتصادي, كلية الإدارة والاقتصاد جامعة الانبار.
11. John Hanke and Arthur Reitsch (1991)"Understanding Business Statistics" Richard D. Irwn Inc, Boston, p 718.
12. بخيت حسين علي,سحر فتح الله"مقدمة في الاقتصاد القياسي"الدار الجامعية للطباعة والنشر ,بغداد,2002 ص 213.
13. الغنام حمد بن عبد الله(2003) "تحليل السلسلة الزمنية لمؤشر أسعار الأسهم في المملكة العربية السعودية باستخدام منهجية بوكس جينكينز"مجلة جامعة الملك عبد العزيز,م 17,ع 2,ص26-3.
14. Willmott,C.J., S.G. Ackleson, R.E.Davis, J.J. Feddema and K.M.Klink ,1985,Statistics for the Evaluation and Comparison of Models.Journal of Geophysical Research , Ottawa, 90(5):8995-9005.