

**استخدام طريقتي بوكس – جنكيز و التنقية المكيفة لتحديد افضل
نموذج للتنبؤ بتلوث مياه**

نهر دجلة باستخدام السلاسل الزمنية القوية

م. م عقيل حميد فرحان

جامعه ديالى – كلية الإدارة والاقتصاد – قسم الإحصاء

**Using the Box-Genghis and Adaptive Purification methods
to determine the best model for Redicting water pollution of
the Tigris River using class time series**

Asst. Ins . Aqeel Hmeed Farhan

University of Diyala

College of Administration & Economics

استخدام طريقي بوكس - جنكيز و التنقية المكيفة لتحديد افضل نموذج للتنبؤ بتلوث مياه نهر دجلة باستخدام السلاسل الزمنية الفئوية.....

م. م عقيل حميد فرحان

المخلص :

تعد السلاسل الزمنية الفئوية ذات أهمية بالغة لوجود العديد من التطبيقات المهمة لها في مختلف المجالات التطبيقية الواقعية إلا أنه يلاحظ وجود تغافل واضح لها في أدبيات موضوع السلاسل الزمنية وذلك بسبب افتقاد التقنيات لمعالجة تلك السلاسل الزمنية الفئوية.

كما أن التنبؤ باتجاهها في المستقبل يتيح لنا معرفة سلوك الظاهرة وعليه فان الهدف من هذه الدراسة هو التنبؤ للسلاسل الزمنية الفئوية لحالة التلوث الحاصلة لمياه نهر دجلة من خلال شبكات تصريف مياه الامطار بسبب تجاوز سكان المدن على هذه الشبكات الخاصة بتصريف مياه الامطار فقطني محافظة واسط باستعمال طريقي بوكس جنكيز وطريقة التنقية المكيفة وقد تم تقدير المعلمات باستخدام طريقة الإمكان الأعظم التي طبقت على بيانات اسبوعية لمدة سنة كاملة بواقع (51) اسبوعاً أخذت من مختبرات تلوث البيئة الخاصة بمحافظة واسط وأظهرت نتائج التطبيق أن النموذج الملائم والكفوء لتمثيل بيانات السلسلة الزمنية الفئوية هو النموذج من الدرجة ARMA(1, 2, 4) وأظهرت نتائج التنبؤ استنتاجاً أن طريقة التنقية المكيفة تظهر نتائج أدق من بوكس- جنكيز Box - Jenkins method بسبب ظهور تنبؤات متساقطة مع مثيلاتها في السلسلة الأصلية.

Abstract

The categorical time series is very important because there are many important applications in the various fields of applications. In fact, there is a clear disregard for them in the literature of the time series because of the lack of techniques to treat with these categorical time series .

Predicting its future direction also allows us to know the behavior of the phenomenon ,the objective of this study is to predict by the categorical time series of the state of pollution of the water of the Tigris River through rainwater drainage networks because the urban population exceeded these networks for the drainage of rain water only in Wasit Governorate, by using Box-Jenkins method and Adaptive Filtering method ,the parameters were estimated using the maximum likelihood method ,Which were applied to weekly data and for a full year of 51 weeks taken from environmental and private pollution laboratories in Wasit Governorate, The results of the application showed that the appropriate and efficient model for representing The categorical time series data is ARMA(4,2,1) ,The results of the prediction showed our conclusion that Adaptive Filtering method showed more accurate results than Box-Jenkins method because of predictions consistent with its Original Series

المقدمة

لقد أهتم الإنسان منذ أقدم العصور باستشراف المستقبل والتنبؤ بتغيراته والتنبؤ بقيمها على مستوى المؤسسات والدول فإن قراءة المستقبل تعد عملية بالغة الأهمية لارتباطها الصميمي الوثيق مع الخطط التخطيطية والتنموية المختلفة وأن من أهم العمليات التي تؤدي إلى تطور البلدان هي عملية التخطيط، والتخطيط يكون على أساس أحسن وجه إذا اعتمدت فيه الأساليب الإحصائية التي تؤدي إلى تخطيط قائم على أسس علمية متينة. ومن ابرز الأساليب الإحصائية هي السلاسل الزمنية ،التي تعد من الموضوعات المهمة والمستخدمه في تفسير سلوك الظواهر عن طريق دراسة تطورها التاريخي عبر مدة زمنية قد تكون يومية أو أسبوعية أو فصلية أو شهرية... الخ لأغراض التنبؤ بما سيحدث في مستقبل هذه الظواهر بأقل خطأ ممكن والغاية الأساسية من استعمال السلاسل الزمنية يتمثل بالحصول على أنموذج يمكن استخدامه لوصف المشكلة وبالنتيجة التنبؤ بمستقبل الظاهرة المدروسة وعلية سنقوم بدراسة نوع معين ومهم من السلاسل الزمنية هي السلاسل الزمنية الفئوية Categorical Time Series وهي سلاسل تمثل بشكل حروف أو رموز وليست أرقاماً أي تمثل أصنافاً معينة لحالة ما مثل [أسعار دور السكن تتغير تبعاً للمتغيرات التوضيحية (كمساحة حديقة الدار و موقع الدار) ومثل أسلوب التحرك عن طريق المركبات و القطارات والطائرات والسفن] ولها أهمية واضحة في التطبيقات الطبية والبيولوجية و الجيولوجية كذلك يتم التنبؤ باستعمال الشبكات العصبية التي توصف بأنها سلسلة من الحالات التي تمر بها ظاهرة معينة خلال انتقالها عبر المحور الزمني أو أي دليل آخر استناداً إلى قوانين احتمالية تدعى بالاحتمالات الانتقالية وقد تشعبت تطبيقات سلاسل ماركوف وشملت مساحه واسعة ابتداء من التطبيقات الإنسانية وحتى التطبيقات الطبية مروراً بالتطبيقات العلمية التقنية المختلفة.

الدراسات السابقة

1- في عام 2005 قامت الباحثة ابتهاج الكسو بتقديم أطروحة دكتوراه لتقدير رتبة سلسلة ماركوف على بيانات سلسلة زمنية فئوية باستخدام الشبكات العصبية طبقت على بيانات جبل بطمة في محافظة نينوى
2- وفي عام 2010 قدم الباحث Biswas بتقديم بحث استخدم فيه المعلومات المتبادلة لتحليل السلاسل الزمنية الفئوية إذ قام بتكوين نموذج انحدار ذاتي متقطع ثم لاءم بينه وبين العمليات الفئوية على وفق نموذج انحدار ذاتي والأوساط المتحركة ARMA إذ لاحظ عدم ثبات الارتباط في البيانات الفئوية لذا لجأ الباحث إلى المعلومات التبادلية (auto-mutual information).

مشكلة البحث

يعد استعمال السلاسل الزمنية الفئوية مشكلة بحد ذاتها وكيفية التعامل مع بيانات فئوية حقيقية ويمكن تعريف مشكلة البحث بانها ايجاد افضل طريقة للتنبؤ للسلاسل الزمنية الفئوية

هدف البحث

يهدف هذا البحث الى المقارنة بين طريقتي بوكس جنكيز والتنقية المكيفة في السلاسل الزمنية الفئوية للتنبؤ بحالة تلوث المياه في نهر دجلة للحصول على أفضل طريقة للتنبؤ من خلال تطبيقها على بيانات فئوية حقيقية تمثلت بحالة تلوث مياه نهر دجلة بسبب التجاوز من قبل السكان على شبكات الامطار الخاصة بتصريف مياه الامطار في محافظة واسط

الجانب النظري

سيتم استعمال السلاسل الزمنية الفئوية **Categorical Time Series** التي ستركز عليها بحثنا كما سنتناول مسألة التنبؤ بهذه السلسلة وذلك بالاعتماد على الطرائق الإحصائية التقليدية للسلاسل الزمنية إذ يمكن تعريف السلاسل الزمنية الفئوية **Categorical Time Series** بانها تلك السلاسل الزمنية التي يكون فضاء الحالة لها بشكل حروف أو رموز وليس أرقاماً. وعلى سبيل المثال قام الباحثون [Stoffer, et.al; 1988] بدراسة حالة النوم **Sleep-State** للطفل وذلك باستخدامهم جهاز **Electronic Phalographic (EEG)**. إن حالة النوم في تلك الدراسة قد صنفت (لكل دقيقة) إلى ست حالات هي:

أثر تناوب النوم الهادئ (**qt**) quiet sleep-trace alternant
الفولتية العالية للنوم الهادئ (**qh**) quiet sleep-high voltage
النوم الانتقالي (**tr**) transitional sleep
الفولتية المنخفضة للنوم النشط (**al**) active sleep- low voltage
الفولتية العالية للنوم النشط (**ah**) active sleep-high voltage
الاستيقاظ (**aw**) awake

إن المشاهدات المذكورة في تلك الدراسة هي كما يأتي (تقرأ من اليسار إلى اليمين):

{ah, ah, ah, ah, ah, ah, ah, ah, tr,}.

فلو افترضنا أن $\{X_t; t=0,1,2,\dots\}$ تمثل السلسلة الزمنية لحالة النوم لذلك الطفل، فكما هو واضح فإن فضاء الحالة لهذه السلسلة هو المجموعة الفئوية $S=\{qt, qh, tr, al, ah, aw\}$. وكما هو معروف فإن التحليل الإحصائي لمثل هذه السلسلة الزمنية ليس بالأمر اليسير، لأن عناصر فضاء الحالة ليست قيماً عددية. لذا اقترح الباحثون [Stoffer, et.al, 1988] إجراء مقياس (Scaling) فأخذوا أولاً المقياس الآتي:

qt=1, qh=2, tr=3, al=4, ah=5, aw=6

ثم ناقشوا بعد ذلك، ولأسباب خاصة مقياساً آخر هو:

qt=1, qh=1, tr=2, al=3, ah=3, aw=4

لذا يلاحظ أن إجراء المقياس لهذه المشاهدات يكون عملية غير وحيدة وأن هذه الملاحظة تمثل نقطة ضعف واضحة في البيانات الفئوية مما يؤدي بالنتيجة إلى نتائج غير دقيقة ولا يمكن اعتمادها بثقة كبيرة وذلك لأن الرقم المقياس لا يعني شيئاً في هذه الحالة وعليه سيتم استخدام طرائق التنبؤ التالية :

طريقة بوكس - جنكيز [11,6,3] Box - Jenkins method

يقصد بنماذج بوكس - جنكيز (Box and Jenkins) هي تلك المنهجية التي طبقها كل من (George Box و Gwilyn Jenkins) على السلاسل الزمنية عام 1976 وهذه المنهجية تعتمد في صيغتها على ثلاثة أجزاء هي نموذج الانحدار الذاتي (Autoregressive) ونموذج المتوسطات المتحركة (Moving Average Model) والنموذج المختلط **Auto Regressive -Moving Average**.
وتعتمد هذه المنهجية على أربع مراحل لغرض التنبؤ .

1.مرحلة تحديد النموذج Identification.

2.مرحلة تقدير المعالم Estimation.

3.مرحلة اختبار دقة النموذج Diagnostic Checking.

4.مرحلة التنبؤ Forecasting.

إذ يتم التنبؤ بعد إكمال مرحلة اختبار دقة النموذج اذا كان النموذج غير ملائم فانه يتم تجاهله واختبار نموذج اخر وتعاد العملية من جديد أما اذا كان النموذج ملائماً حينها يتم اجراء عملية التنبؤ على أساسه أما أنواع نماذج بوكس - جنكيز Box - Jenkins فهي نموذج الانحدار الذاتي (AR) ونموذج المتوسطات المتحركة (MA) والنموذج المختلط (ARMA) والنموذج الموسمي (SARMA) وغيرها من نماذج بوكس - جنكيز وكالاتي.

نماذج الانحدار الذاتي Auto regressive models

ان صيغة نموذج الانحدار الذاتي من الدرجة (p) الذي يرمز له بالرمز AR(P) وهو كما يلي

$$Z_T - M + Q_1Z_{T-1} - Q_2Z_{T-2} - \dots - Q_aZ_{T-a}$$

حيث ان :

$$Z_{t-r} \text{ تمثل المشاهدات } t=1, 2, 3, \dots, p$$

$$(Q_i \text{ } i=1, 2, 3, \dots, p) \text{ تمثل مجموعة الاوزان للقيم السابقة الى } Z$$

M : عدد المعالم المقدره

نماذج الاوساط المتحركة Moving average models

ان الصيغة الاساسية لنماذج الاوساط المتحركة من الدرجة (q) ويرمز لها بالرمز MA(q) هي كالاتي

$$Z_T = a_t - \theta_1a_{t-1} - \theta_2a_{t-2} - \dots - \theta_a a_{t-a}$$

حيث Zt: هي قيمة المشاهدة في الفترة (t)

$$(\theta_t, \text{ } t=1, 2, 3, \dots, q) \text{ معالم ثابتة للنموذج } a_t \text{ تمثل الأخطاء العشوائية}$$

النماذج المختلطة Mixed auto regressive moving average models

ويرمز لهذه النماذج بالرمز ARMA(p, q) حيث (p) تمثل درجة الانحدار الذاتي ويمثل (q) درجة الاوساط

المتحركة وتكتب او توجد على وفق الصيغة التالية

$$Z_T - M - Q_1Z_{T-1} + Q_2Z_{T-2} + \dots + Q_pZ_{T-p} + a_t - \theta_1a_{t-1} - \theta_2a_{t-2} - \dots - \theta_q a_{t-q}$$

مرحلة تحديد النموذج Identification

تعتمد مرحلة تحديد النموذج على أيجاد معاملات الارتباط الذاتي (ACF) والارتباط الذاتي الجزئي (PACF)

واستخدام بعض المعايير التي تستخدم لتحديد درجة ونوع النموذج ومن أهمها معيار اكيائي (AIC) وكالاتي .

دالة الارتباط الذاتي ACF (Auto-Correlation Function): [14,3]

يعرف معامل الارتباط بأنه مقياس لدرجة العلاقة بين قيم المتغير نفسه عند فترات إزاحة موسمية مختلفة ويحسب

معامل الارتباط الذاتي في حالة السلاسل الزمنية الموسمية عند الإزاحة k حسب الصيغة الآتية

$$\rho_k = \frac{Cov(Z_t, Z_{t+k})}{\sqrt{Var(Z_t)} \sqrt{Var(Z_{t+k})}} \dots \dots \dots (2)$$

دالة الارتباط الذاتي الجزئي [14.3] Partial Auto-Correlation Function (PACF)

يعرف معامل الارتباط الذاتي الجزئي بأنه مقياس لدرجة العلاقة بين المشاهدتين Z_t, Z_{t+k} بثبوت بقية المشاهدات الأخرى $Z_{t+1} \dots \dots \dots Z_{t+k-1}$ ويمكن حسابها على وفق الصيغة الآتية .

$$\hat{\phi}_{k+1, k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_j} \dots \dots \dots (3)$$

وان دالة الارتباط الذاتي الجزئي (PACF) لا تقل أهمية عن دالة الارتباط الذاتي فهي أيضا أداة مهمة في تحليل السلاسل الزمنية وتستخدم أيضا في تشخيص النموذج من خلال عشوائية البواقي أخطاء البواقي الالية يتم تحديد الأنموذج الملائم للمشاهدات وهناك معايير أخرى لتحديد الأنموذج مثل .

هذه الطريقة اقترحها العالم Akaike في عام 1973 إذ اقترح معياراً اسماً اختصر بـ AIC اذ يعد هذا المعيار

اداة من أدوات التشخيص معيار معلومات اكيائي [9,6]: Akaike information criterion

فضلاً عن كونه معياراً لتحديد رتبة النموذج ويمكن حسابه على وفق الصيغة التالية .

$$AIC(M) = n \ln(\sigma_u^2) + 2M \dots \dots \dots (4)$$

إذ M تمثل العدد الكلي لمعاملات النموذج (p , q) .

n هي عدد مشاهدات السلسلة .

σ_a^2 : مقدار تباين الخطأ .

أذ أن الأنموذج الأفضل هو الأنموذج الذي يعطي اقل قيمة من AIC(M) .

مرحلة التقدير [14,8,6]: Estimation

إن عملية تقدير النماذج هي المرحلة الثانية من مراحل دراسة السلاسل الزمنية وتحليلها وتأتي بعد عملية تشخيص النموذج الملائم للسلسلة الزمنية اذ توجد عدة طرائق لتقدير معاملات النموذج واهم هذه الطرائق طريقة الإمكان الأعظم (Maximum Likelihood Method) لتقدير معاملات النموذج وذلك بجعل مجموع مربعات الخطأ اقل ما يمكن وصيغة دالة الإمكان الأعظم تكون على النحو التالي .

$$L(\phi, \sigma_\alpha^2 / Z) = (2\pi \sigma_\alpha^2)^{-\frac{T}{2}} \left| M_T^{-1} \right|^{\frac{1}{2}} \exp \left[\frac{-S(\phi)}{2\sigma_\alpha^2} \right] \dots \dots \dots (5)$$

وعند اخذ اللوغاريتم

سيكون :

استخدام طريقتي بوكس - جنكيز و التنقية المكيفة لتحديد افضل نموذج للتنبؤ بتلوث مياه نهر دجلة باستخدام السلاسل الزمنية الفتوية.....

$$\ln L(\phi, \sigma_\alpha / Z) = -\left(0 \frac{T}{2}\right) \ln(2\pi) - \left(\frac{T}{2}\right) \ln \sigma_\alpha^2 + \left(\frac{1}{2}\right) \ln |M_T^{(P)}| - \left[\frac{S(\phi)}{2 \sigma_\alpha^2} \right] \dots \dots \dots (6)$$

حيث ان :

$$S(\phi) = \sum_{t=p+1}^T (\alpha_t / Z, \phi)^2 = \sum_{t=p+1}^T (\alpha_t)^2 \dots \dots \dots (7)$$

وباشتقاق المعادلة (7) بالنسبة ل Φ ومساواتها للصفر وتبسيطها نحصل على تقدير Φ المطلوب

اختبار دقة النموذج [14,8,6]:Diagnostic Checking

بعد التعرف على الانموذج وتقدير معالمه تأتي مرحلة مهمة هي مرحلة اختبار مطابقة النموذج ومدى كفاءة النموذج وملاءمته لتمثيل بيانات السلسلة الزمنية في تمثيل الظاهرة المدروسة باستخدام اختبارات عدة منها .

أولاً: اختبار طبيعية البواقي أي مدى تقارب البواقي المقدره إلى التوزيع الطبيعي النظري المفترض أن تتوزع

$$\cdot a_t \approx N(0, \sigma^2) \text{ البواقي}$$

ثانياً: اختبار الاستقلالية يتم من خلال اختبار معاملات الارتباط الذاتي للخطأ وكالاتي .

$$H_0: P = 0 \quad H_1: P \neq 0$$

ونستخرج فيه الاحصاءة نفسها لاختبار فرضية المتوسط $u = \frac{e}{se(e)}$ وهذه الاحصاءة لها توزيع طبيعي

قياسي فإذا كانت $|u| < 1.96$ عند مستوى معنوية 0.05 عند ذلك تقبل فرضية العدم .

-التنبؤ [10,3]:Forecasting

بعد تحديد النموذج الملائم من خلال مرحلة التشخيص وتقدير معالمه وفحص ملاءمة النموذج يتم استخدامه في التنبؤ بالقيم المستقبلية إلى ($L=1, 2, \dots$) فترة قادمة وذلك بأخذ التوقع الشرطي عند زمن الأصل (t) للنموذج وبعد كتابته عند الفترة ($t+L$) أي $E_t(Z_{t+L} / Z_t, Z_{t-1}, \dots)$ إذ باستخدام التوقع الشرطي نحصل على التنبؤات

$$\hat{Z}_t(L)$$

أذ يمكن حساب التنبؤات للنموذج المختلط الموسمي على وفق الصيغة الاتية:

حيث

$$Z_t = Z_t - \mu \quad \phi_{(p+d)}(\beta) = \phi(\beta)(1 - \beta)^{-1} = (1 - \phi_1\beta - \dots - \phi_{p+d}\beta^{p+d}) \quad \text{و}$$

ويسمى بمعامل الانحدار الذاتي غير المستقر من الدرجة ($p+d$).

ويمكن كتابة المعادلة على النحو الآتي:

$$Z_{t+L} = \hat{Z}_t(L) = \hat{\Phi}_S Z_{t+L-S} + \hat{\Phi}_{2S} Z_{t+L+2S} + \dots + \hat{\Phi}_{PS} Z_{t+L-PS} + a_{t+L} - \hat{\Theta}_S a_{t+L-S} - \dots - \hat{\Theta}_{QS} a_{t+L-QS} \dots \dots \dots (8)$$

حيث إن :

$$\begin{aligned}
 E_t[Z_{t-L}] &= Z_{t-L} & L = 0, 1, 2, \dots \\
 E_t[Z_{t+L}] &= Z_t(L) & L = 1, 2, \dots \\
 E_t[a_{t-L}] &= a_{t-L} = Z_{t-L} - Z_{t-L-s}(1) & L = 0, 1, 2, \dots \\
 E_t[a_{t+L}] &= 0 & L = 1, 2, \dots
 \end{aligned} \quad (9)$$

طريقة التنقية المكيفة [14,8] Adapting Filtering Method:

إن أول من بدأ العمل في هذا الحقل هو (Kolmogorov, 1941)، (Weiner, 1949) فضلاً عن كل من (Kalman, 1960) و (Kalman & Bucy, 1961) وقد فتحوا طريقاً واسعاً في استخدام هذا الأسلوب في مختلف التطبيقات الهندسية المتنوعة [14].

إن التنبؤ بطريقة التنقية المكيفة (Forecasting with Adaptive Filtering) والمعروضة من قبل (Makridakis & Wheelwright, 1973)، شملت خوارزميات (Algorithms) لتعديل (Revising) المعلومات للنماذج الخطية العشوائية وذلك من خلال إضافة حد لتصحيح المعلومات الأصلية والذي هو عبارة عن نسبة بين حاصل ضرب بواقى أحدث التنبؤ وقيم المشاهدات الأولية. على أن التنقية المعدلة في ذلك الوقت كانت تستخدم لنماذج الانحدار الذاتي (Autoregressive) فقط [14].

ولكن منذ عام (1977) توسعت في استخدامها لتشمل نماذج أخرى منها نماذج أخطاء الأوساط المتحركة (Moving Average Errors) لسلاسل مستقرة وغير مستقرة. وقد أشار (Nau & Oliver) في عام (1979) إلى أن التنبؤ بطريقة التنقية المكيفة المعروضة من قبل (Makridakis and Wheelwright) تعد أكثر دقة وكفاءة تقريباً من أسلوب التنقية المقدمة من قبل (Kalman) المطبقة بشكل فعال في نماذج الانحدار الذاتي. ويعرض الباحث فيما يلي التنبؤ بهذه الطريقة من فترة إلى أخرى وليست ثابتة.

التنقية المكيفة إلى الأنموذج الموسمي (الضربي)

[14,8, 5] Adaptive filtering to multiplicative model

أن نموذج ARIMA(p , d , q) يمكن كتابته على وفق الصيغة الآتية

$$Z_t = \phi_{1s}Z_{t-s} + \dots + \phi_{ps}Z_{t-ps} + a_t - \theta_{1s}a_{t-s} - \dots - \theta_{qs}a_{t-qs} \quad (10)$$

حيث أن : t = 1,2,3.....,n

s : فترة الموسم .

at : تمثل مركبة الخطأ العشوائي بوسط حسابي مساوٍ للصفر وتباين $\sigma^2 a_t$

$\phi_{1s}, \dots, \phi_{ps}$: تمثل معاملات AR غير المعلومة .

$\theta_{1s}, \dots, \theta_{qs}$: تمثل معاملات MA غير المعلومة .

استخدام طريقي بوكس - جنكيز و التنقية المكيفة لتحديد افضل نموذج للتنبؤ بتلوث مياه نهر دجلة باستخدام السلاسل الزمنية القوية.....

ويمكن تقدير المعلمات على وفق طريقة المربعات الصغرى غير الخطية non-linear least square method باستخدام ميول الانحدار السريع بمعنى استخدام الميول عند البحث عن الأصغر آذ أن الميول عند أي نقطة على سطح الاستجابة يمكن إيجادها عن طريق اشتقاق دالة متوسط مربعات الخطأ Mse المتمثلة a_t^2 وتكون على وفق الصيغة الآتية .

$$a_t = Z_t - \phi_{1s}Z_{t-s} - \dots - \phi_{ps}Z_{t-ps} + \theta_{1s}a_{t-s} + \dots + \theta_{qs}a_{t-qs} + \dots \quad (11)$$

وبتربيع طرفي المعادلة نحصل على .

$$Z_t - \phi_{1s}Z_{t-s} - \dots - \phi_{ps}Z_{t-ps} + \theta_{1s}a_{t-s} + \dots + \theta_{qs}a_{t-qs} \quad (12) = a_t^2$$

وباشتقاق الـ a_t^2 بالنسبة إلى ϕ_i والـ θ_j نحصل على :

$$\frac{\partial a_t^2}{\partial \phi_i} = -2a_t Z_{t-is} \quad i = 1, 2, \dots, p$$

$$\frac{\partial a_t^2}{\partial \theta_j} = 2a_t a_{t-js} \quad j = 1, 2, \dots, q$$

باستعمال الأسلوب التكراري يمكن إن نستنتج أن المعلمات المكيفة المعدلة هي [46]

$$\phi_{is}^* = \phi_{is} + 2k \nabla a^2 \quad \dots \dots \dots (13)$$

حيث أن :

ϕ_{is}^* : تمثل المعلمة المعدلة (المكيفة) الجديدة .

ϕ_{is} : تمثل المعلمة قبل التعديل

K : تمثل ثابتاً اختيارياً يقوم بالسيطرة على سرعه التقارب من خلال عدد مرات التكرار المستخدمة .

∇a^2 : متجه التدرج لـ a^2 ومنها نجد .

$$i = 1, 2, \dots, S \quad \phi_{it}^* = \phi_{is} + 2ka_t Z_{t-i}$$

$$\theta_{it}^* = \theta_{is} + 2ka_t a_{t-j} \quad j = 1, 2, \dots, S \quad \dots \dots \dots (14)$$

وعليه أن الصيغ(14) تطبق بشكل متكرر الى الحد او النقطة التي ينعدم فيها الاختزال او التخفيض لمتوسط مربعات الخطأ (Mse) عند هذه النقطة وعليه فإن عملية التعديل (التنقية) تنتهي وأن القيمة النهائية للمعلمة تستخدم للتنبؤ .

ومن الصيغة (65-2) تكون أخطاء البواقي المعدلة الجديدة .

$$a_t = Z_t - \phi_{1s}Z_{t-1} - \phi_{2s} - \dots - \phi_{ps}Z_{t-ps} + \theta_{ps}a_{t-1} + \dots + \theta_{qs}a_{t-qs} \dots \dots (15)$$

$$\nabla a_t = a_t^* - a_t$$

$$= -(\phi_{1s}^* - \phi_{1s})Z_{t-1} - \dots - (\phi_{ps}^* - \phi_{ps})Z_{t-ps} + (\theta_{1s}^* - \theta_{1s})a_{t-1} + \dots + (\theta_{qs}^* - \theta_{qs})a_{t-qs}$$

أو

$$a_t = -2ka_t [Z_{t-1}^2 \dots Z_{t-1}^2 + a_{t-1}^2 + \dots + a_{t-p}^2] \dots \dots \dots (16)$$

بقسمة المعادلة (16) على (at) ينتج .

$$\frac{|Va_t|}{a_t} = -2k[Z_{t-1}^2 \dots \dots Z_{t-1}^2 + a_{t-1}^2 + \dots \dots + a_{t-p}^2] \dots \dots \dots (16)$$

ومن العلاقة الأخيرة يمكن استنتاج

$$0 < k < 1[Z_{t-1}^2 \dots \dots Z_{t-1}^2 + a_{t-1}^2 + \dots \dots + a_{t-p}^2] \dots \dots \dots (17)$$

عامل التعديل [14, 6] Factor of Adjusting

اقترحها العالم (Makridakis) و(Wheel Wright) عام 1977 لغرض التخلص من التذبذب الحاصل في السلسلة الزمنية والوصول إلى القيم المثلى بصورة أسرع فإننا يمكن أن نجعل قيم السلسلة الزمنية قياسية وذلك بالقسمة على العامل القياسي (Standard Factor) واقترح الباحثان (Makridakis) و(Wheel Wright) بأن z_t ، a_t يجب أن تعدلا وقد قاما باقتراح عدة طرائق للتعديل من هذه الطرائق هي قسمة كل مشاهدة في السلسلة على اكبر قيمة فيها لكنهما عدلا باقتراحهما هذا بعد أن لاحظا قصوراً فيه من الناحية التطبيقية بالاقتراح

الثاني الذي يقول ان تعديل القيم يجب أن يكون بقسمة كل قيمة من القيم للسلسلة على $W_t = \sqrt{\sum_{i=1}^t z_i^2}$ ولكن الباحث (Chatf Fiold) وجد عام 1978 أن هذه الطريقة تعقد التحليل أحيانا ونتيجة لهذا الانتقاد قدم الباحث (Makridakis) و(Wheel Wright) اقتراحاً جديداً لمعامل التعديل الذي يتم تقسيم قيم السلسلة عليه وكما يلي:

$$W_t = \sum_{i=1}^T (Z_{t-i})^2 + \sqrt{W_{t-i}} \dots \dots \dots (2-74) \text{ For } .t = p+1, p+2, \dots \dots T$$

حيث أن :

W_t : عامل التعديل للفترة t ، T : طول السلسلة الزمنية

P : درجة السلسلة الزمنية

ويستعمل معامل التعديل لإزالة تأثير الاتجاه العام للسلسلة الزمنية مما يسهل المقارنة بين السلاسل المختلفة.

مقاييس دقة التنبؤ:

نتيجة لاستعمال عدة طرائق للتنبؤ ومنها طريقة بوكس - جنكيز وطريقة الشبكات العصبية للسلاسل الزمنية لابد من توافر مقاييس او معايير لاختيار الطريقة الأفضل للتنبؤ .

متوسط الخطأ النسبي المطلق (MAPE) Mean Absolute Percentage Error

وتحسب على وفق الصيغة الآتية .

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| * 100 \dots \dots \dots (18)$$

متوسط مربعات الخطأ Mean Square Error

يعد متوسط مربعات الخطأ من أكثر وأدق المقاييس الإحصائية استعمالاً وذلك لأنه يعد مقياساً لمدى قرب المعلمة المقدرة من المعلمة الحقيقية وأيضاً يعد مقياساً جيداً بين مختلف طرائق التقدير ويحسب على وفق الصيغة الآتية .

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2 \dots \dots \dots (19)$$

الجانب التطبيقي :

تم الاعتماد على بيانات حقيقية أخذت من مختبرات دائرة البيئة في محافظة واسط والتي استخدمت في هذا البحث تؤلف سلسلة زمنية فئوية Categorical Time Series بواقع (51) اسبوعاً تمثل حالات التلوث لمياه نهر دجلة بسبب التجاوز الحاصل على شبكات تصريف مياه الامطار من قبل السكان في المدينة سميت فئوية لكونها بيانات ليست عددية بل عبارة عن أسماء تمثلت بأسماء (ملوثة - تلوث مقبول - غير ملوثة) والتي مثلت بالحروف أي كونت فضاء الحالة S=[A , B ,C] وهي بيانات فئوية (وصفية) إذ لا يمكن التنبؤ باستخدام الطرائق الإحصائية الاعتيادية للملاحظات الفئوية ولمعالجة مسألة التنبؤ لهذه السلسلة نحتاج أولاً إلى ترميز المشاهدات لحالات الطقس الثلاث لتحويلها إلى بيانات عددية بحيث يمكن التنبؤ لها علماً انه تم ترميز المشاهدات إلى ست حالات مرمره والتي سميت بـ (C1 , C2 , C3 , C4 , C5 , C6) إذ تم ترميزها وبشكل عشوائي وكما يلي :

1 الحالة الأولى (C1) A=1 , B=C=2 ,=3

2 الحالة الثانية (C2) A=B=2 , C=3 ,=1

1 الحالة الثالثة (C3) A=3 , B=C=1 ,=2

2 الحالة الرابعة (C4) A=B=2 , C=1 ,=3

1 الحالة الخامسة (C5) A=1 , B=C=3 ,=2

2 الحالة السادسة (C6) A=B=3 , C=2 ,=1

والتي تستخدم لإظهار نتائج التنبؤ بطريقة (Box - Jenkins) ومقارنتها مع طريقة التنقية المكيفة وذلك لمعرفة أي الطرائق أفضل للتنبؤ باستعمال السلاسل الزمنية الفئوية وتم الاعتماد على البرامج الحاسوبية (Minitab 15 , MATLAB ,) لإظهار النتائج التنبؤية للظاهرة المدروسة.

٢- التنبؤ باستخدام طريقة بوكس جنكيز

لمعرف سلوك المشاهدات يجب علينا رسم شكل الانتشار للسلسلة الزمنية الفئوية للملاحظات الأصلية ومن خلال ملاحظة الشكل رقم (1) نلاحظ أن السلسلة مستقرة في الوسط كونها حصرت بين قيم الترميز التي اختيرت من قبل الباحث .

- مرحلة تحديد النموذج Identification

لمعرفة وتحديد النموذج يتم من خلال رسم دالة الارتباط الذاتي (ACF) ودالة الارتباط الذاتي الجزئي (PACF) للملاحظات الأصلية المرمره للسلسلة الزمنية الفئوية و يتضح من شكل رقم (2) و (3) التي تمتلك دالتي الارتباط الذاتي والجزئي أنها تتناقص تدريجياً مع ازدياد فترات الإزاحة (k) سالكة سلوك دالة الجيب المتضائلة تدريجياً ومن خلال هذا المؤشر نستنتج أن النموذج هو نموذج (ARMA) الاعتيادية وباستخدام معيار المفاضلة (AIC) لتحديد الأنموذج الأفضل وكما مبين في الجدول رقم (1)

(المصدر اعداد الباحث اعتمادا على بيانات الدراسة باستعمال برنامج [Minitab 19])

جدول (1)

يمثل قيمه معيار (AIC) لنماذج ARIMA لتلوث مياه نهر دجلة

Case	ARIMA	M	MS	AIC
C1	(2, 2, 2)	4	1.057	10.8272
	(3, 1, 1)	4	0.98	6.9697
	(3, 1, 2)	5	0.9728	8.5936
	(4, 2, 1)	5	0.952	5.4913
C2	(2, 2, 2)	4	0.952	5.4913
	(3, 1, 1)	4	0.89	2.0568
	(3, 1, 2)	5	1.342	25.002
	(4, 2, 1)	5	0.822	0.00324
C3	(2, 2, 2)	4	0.578	-19.957
	(3, 1, 1)	4	0.6008	-17.981
	(3, 1, 2)	5	0.628	-13.726
	(4, 2, 1)	5	0.5529	-20.2215
C4	(2, 2, 2)	4	1.054	10.6822
	(3, 1, 1)	4	0.9804	6.9905
	(3, 1, 2)	5	0.9521	7.4967
	(4, 2, 1)	5	0.822	0.00324
C5	(2, 2, 2)	4	0.578	-19.957
	(3, 1, 1)	4	0.6008	-17.981
	(3, 1, 2)	5	0.628	-13.726
	(4, 2, 1)	5	0.5529	-20.2215
C6	(2, 2, 2)	4	1.057	10.827
	(3, 1, 1)	4	0.9804	6.9905
	(3, 1, 2)	5	0.9728	8.5936
	(4, 2, 1)	5	0.9521	7.4967

حيث أن M : تمثل عدد المعالم المقدره في النموذج

مما تقدم نستنتج أن اقل قيمة لمعايير المفاضلة المبينة في الجدول رقم (1) يحملها النموذج ARIMA (4, 2, 1) وهو النموذج الملائم للبيانات .

تقدير معالم النموذج : Estimation :

قدرت معالم النموذج على وفق طريقة الإمكان الأعظم (Maximum Likelihood Method) وكما موضح في الجدول رقم (2) والتي استخرجت باستخدام البرنامج الجاهز (Minitab 15) وكالاتي.

جدول (2)

النتائج الاحصائية لتلوث مياه نهر دجلة لنموذج ARIMA

مقدرات الحالة الأولى C1				مقدرات الحالة الثانية C2			
Type	Coef	StDev	T	Type	Coef	StDev	T
AR 1	-0.4040	0.1370	-2.95	AR 1	-1.9250	0.5170	-3.72
AR 2	-0.1997	0.1475	-1.35	AR 2	-1.3760	0.5867	-2.35
AR 3	-0.2055	0.1470	-1.40	AR 3	-0.3452	0.3939	-0.88
AR 4	-0.4221	0.1353	-3.12	AR 4	0.0806	0.1535	0.52
MA 1	0.9499	0.0700	13.56	MA 1	-0.9344	0.5007	-1.87
Number of observations: Original series 51, after differencing 49				Differencing: 2 regular differences Number of observations: Original series 51, after differencing 49			
Residuals: SS = 41.8938 (backforecasts excluded)				Residuals: SS = 36.1676 (backforecasts excluded)			
MS = 0.9521 DF = 44				MS = 0.8220 DF = 44			
مقدرات الحالة الثالثة C3				مقدرات الحالة الرابعة C4			
Type	Coef	StDev	T	Type	Coef	StDev	T
AR 1	-0.5478	0.1437	-3.81	AR 1	-1.9250	0.5170	-3.72
AR 2	-0.4700	0.1505	-3.12	AR 2	-1.3760	0.5867	-2.35
AR 3	-0.4551	0.1498	-3.04	AR 3	-0.3452	0.3939	-0.88
AR 4	-0.2637	0.1417	-1.86	AR 4	0.0806	0.1535	0.52
MA 1	0.9585	0.0596	16.07	MA 1	-0.9344	0.5007	-1.87
Differencing: 2 regular differences Number of observations: Original series 51, after differencing 49				Differencing: 2 regular differences Number of observations: Original series 51, after differencing 49			
Residuals: SS = 24.3275 (backforecasts excluded)				Residuals: SS = 36.1676 (backforecasts excluded)			
MS = 0.5529 DF = 44				MS = 0.8220 DF = 44			
مقدرات الحالة الخامسة C5				مقدرات الحالة السادسة C6			
Type	Coef	StDev	T	Type	Coef	StDev	T
AR 1	-0.5478	0.1437	-3.81	AR 1	-0.4040	0.1370	-2.95
AR 2	-0.4700	0.1505	-3.12	AR 2	-0.1997	0.1475	-1.35
AR 3	-0.4551	0.1498	-3.04	AR 3	-0.2055	0.1470	-1.40
AR 4	-0.2637	0.1417	-1.86	AR 4	-0.4221	0.1353	-3.12
MA 1	0.9585	0.0596	16.07	MA 1	0.9499	0.0700	13.56
Differencing: 2 regular differences Number of observations: Original series 51, after differencing 49				Differencing: 2 regular differences Number of observations: Original series 51, after differencing 49			
Residuals: SS = 24.3275 (backforecasts excluded)				Residuals: SS = 41.8938 (backforecasts excluded)			
MS = 0.5529 DF = 44				MS = 0.9521 DF = 44			

تم تقدير معالم النموذج لكل حالة وتعد تقديرات جيدة وكفوءة والتي تنبئ بتوليد تنبؤات جيدة لمستقبل الظاهرة

- اختبار دقة النموذج Diagnostic Checking :

بعد أن تم تحديد النموذج وتقدير معالمه لا بد من إجراء بعض التشخيصات على البواقي لبيان مدى دقة النموذج للسلسلة المشاهدة وحسب الاختبارات التالية .

-اختبار استقلالية البواقي :

بعد ان تم اختيار الانموذج تتم عملية اختبار الانموذج وكما مبين في الشكل (4) و (5) والتي اتضح منها انها تتبع أنماط الضجة البيضاء أي أنها غير مترابطة وبذلك فهي إن كانت طبيعية فهي مستقلة .

-الاختبار الطبيعي لتوزيع البواقي .

بعد ملاحظة أن توزيع البواقي يقترب كثيراً من التوزيع الطبيعي بمتوسط صفر وتباين ثابت وذلك من خلال قبول فرضية طبيعة التوزيع فان الرسم يبين مقدار التقارب مع التوزيع الطبيعي .

-التنبؤ Forecasting:

يعد التنبؤ آخر مرحلة من مراحل تحليل السلاسل الزمنية وبعد ان تم تحديد الانموذج الملائم وإيجاد معاملات الانموذج المقدر لكل حالة مرمزه وعليه سنقوم بالتنبؤ الداخلي (Prediction) للسلسلة الزمنية لكل حالة مرمزه، الذي يتم من خلاله المقارنة واختبار دقة التنبؤ إذ تم قطع اخر ستة اسابيع من السلسلة وذلك لاستخدامها في التنبؤ الداخلي وبعد أن وجدت قيم التنبؤ الداخلي وباستعمال برنامج Minitab تم إيجاد قيم مقاييس أخطاء التنبؤ ودونت في الجدول رقم (3) كالاتي

جدول (3)

يوضح نتائج مقاييس أخطاء التنبؤ لطريقة بوكس -جنكز

الحالات	MSE	MAPE
C1	0.762484	11.1531
C2	0.384937	30.07053
C3	0.200222	9.163263
C4	0.348937	8.691941
C5	0.200222	9.88452
C6	0.762484	51.03153

وتبين من نتائج مقاييس أخطاء التنبؤ في جدول رقم(3) نتائج جيدة وذلك بسبب قلة الأخطاء وبعد التنبؤ المستقبلي (Forecasting) للملاحظات ظهرت نتائج التنبؤ المستقبلي ودونت في الجدول رقم (4) وكالاتي إذ يتم إيجاد القيم المستقبلية للسلسلة المدروسة باستخدام أفضل أنموذج ملائم للسلسلة وكالتالي .

الجدول (4)

يمثل قيم التنبؤ المستقبلي للحالات المرمزة C1, C2, C3, C4, C5, C6 بطريقة بوكس -جنكز

ت	قيم التنبؤ C1	تقريب قيم التنبؤ C1	تحويل الأرقام إلى رموز C1	ت	قيم التنبؤ C2	تقريب قيم التنبؤ C2	تحويل الأرقام إلى رموز C2
1	0.87920	1	ملوثة	1	2.00148	2	ملوثة
2	0.80721	1	ملوثة	2	2.00016	2	ملوثة
3	0.739625	1	ملوثة	3	2.00204	2	ملوثة
4	0.685335	1	ملوثة	4	2.00118	2	ملوثة
5	0.665747	1	ملوثة	5	2.00225	2	ملوثة
6	0.607985	1	ملوثة	6	2.00207	2	ملوثة
7	0.554120	1	ملوثة	7	2.00280	2	ملوثة
8	0.493561	1	ملوثة	8	2.00267	2	ملوثة
9	0.428126	1	ملوثة	9	22.00347	2	ملوثة
10	0.381309	1	ملوثة	10	2.00330	2	ملوثة

ت	قيم التنبؤ C3	تقريب قيم التنبؤ C3	تحويل الأرقام إلى رموز C3	ت	قيم التنبؤ C4	تقريب قيم التنبؤ C4	تحويل الأرقام إلى رموز C4
1	3.08587	3	ملوثة	1	1.99852	2	ملوثة
2	3.12314	3	ملوثة	2	1.99984	2	ملوثة
3	3.14867	3	ملوثة	3	1.99796	2	ملوثة
4	3.16351	3	ملوثة	4	1.99882	2	ملوثة
5	3.19129	3	ملوثة	5	1.99775	2	ملوثة
6	3.23384	3	ملوثة	6	1.99793	2	ملوثة
7	3.26990	3	ملوثة	7	1.99720	2	ملوثة
8	3.29951	3	ملوثة	8	1.99733	2	ملوثة
9	3.32580	3	ملوثة	9	1.99653	2	ملوثة
10	3.35631	3	ملوثة	10	1.99670	2	ملوثة

ت	قيم التنبؤ C5	تقريب قيم التنبؤ C5	تحويل الأرقام إلى رموز C5	ت	قيم التنبؤ C6	تقريب قيم التنبؤ C6	تحويل الأرقام إلى رموز C6
1	0.914132	1	ملوثة	1	2.96982	3	ملوثة
2	0.876861	1	ملوثة	2	2.99136	3	ملوثة
3	0.851328	1	ملوثة	3	2.95411	3	ملوثة
4	0.836492	1	ملوثة	4	2.97528	3	ملوثة
5	0.808713	1	ملوثة	5	2.94054	3	ملوثة
6	0.766159	1	ملوثة	6	2.95899	3	ملوثة
7	0.730104	1	ملوثة	7	2.92681	3	ملوثة
8	0.700487	1	ملوثة	8	2.94285	3	ملوثة
9	0.674204	1	ملوثة	9	2.91297	3	ملوثة
10	0.6436688	1	ملوثة	10	2.92680	3	ملوثة

-التنبؤ باستخدام طريقة التنقية المكيفة Adaptive Filtering Method :

بعد الحصول على النموذج الملائم للسلسلة الزمنية وتقدير معاملات الأنموذج للسلسلة الزمنية تم حساب قيم التنبؤ الداخلي أي ضمن الفترة الزمنية للسلسلة وتم إيجاد قيم مقاييس أخطاء التنبؤ ودونت في الجدول رقم (5) كالآتي

جدول (5)

يوضح نتائج مقاييس أخطاء التنبؤ لطريقة التنقية المكيفة

الحالات	MSE	MAPE
C1	0.640625	9.6047
C2	0.19063	29.1094
C3	0.0222	8.495
C4	0.19063	3.3854
C5	0.0222	3.495
C6	0.640625	28.3854

استخدام طريقي بوكس – جنكيز و التنقية المكيفة لتحديد افضل نموذج للتنبؤ بتلوث مياه نهر دجلة باستخدام السلاسل الزمنية الفئوية.....

وحسبت قيم التنبؤ المستقبلي لطريقة التنقية المكيفة ودونت في الجدول رقم (6). إذ يتم أيجاد القيم المستقبلية للسلسلة المدروسة باستخدام أفضل أنموذج ملائم للسلسلة وكالاتي .

الجدول (6)

يمثل قيم التنبؤ المستقبلي للحالات المرمزة C1, C2, C3, C4, C5, C6 بطريقة التنقية المكيفية

ت	قيم التنبؤ C1	تحويل الأرقام C1 إلى رموز	ت	قيم التنبؤ C2	تحويل الأرقام C2 إلى رموز	ت	قيم التنبؤ C3	تحويل الأرقام إلى رموز C3
1	3	غير ملوثة	1	2	ملوثة	1	1	تلوث مقبول
2	2	تلوث مقبول	2	3	تلوث مقبول	2	3	ملوثة
3	2	تلوث مقبول	3	3	تلوث مقبول	3	3	ملوثة
4	2	تلوث مقبول	4	3	تلوث مقبول	4	3	ملوثة
5	3	غير ملوثة	5	2	ملوثة	5	1	تلوث مقبول
6	3	غير ملوثة	6	2	ملوثة	6	1	تلوث مقبول
7	3	غير ملوثة	7	2	ملوثة	7	1	تلوث مقبول
8	3	غير ملوثة	8	2	ملوثة	8	1	تلوث مقبول
9	1	ملوثة	9	1	غير ملوثة	9	2	غير ملوثة
10	3	غير ملوثة	10	2	ملوثة	10	1	تلوث مقبول

ت	قيم التنبؤ C4	تحويل الأرقام إلى رموز C4	ت	قيم التنبؤ C5	تحويل الأرقام إلى رموز C5	ت	قيم التنبؤ C6	تحويل الأرقام إلى رموز C6
1	2	ملوثة	1	1	ملوثة	1	3	ملوثة
2	1	تلوث مقبول	2	2	غير ملوثة	2	1	غير ملوثة
3	1	تلوث مقبول	3	2	غير ملوثة	3	1	غير ملوثة
4	1	تلوث مقبول	4	2	غير ملوثة	4	1	غير ملوثة
5	2	ملوثة	5	1	ملوثة	5	3	ملوثة
6	2	ملوثة	6	1	ملوثة	6	3	ملوثة
7	2	ملوثة	7	1	ملوثة	7	3	ملوثة
8	2	ملوثة	8	1	ملوثة	8	3	ملوثة
9	3	غير ملوثة	9	3	تلوث مقبول	9	2	تلوث مقبول
10	2	ملوثة	10	1	ملوثة	10	3	ملوثة

الاستنتاجات والتوصيات :

-الاستنتاجات: مما تقدم يمكن تلخيص النتائج التالية .

- 1.بينت الاختبارات الإحصائية أن السلسلة الزمنية القوية مستقرة في الوسط والتباين وذلك بسبب طريقة الترميز التي اختيرت من قبل الباحث فضلاً عن كونها غير مستقرة هذا ما تبين من خلال دالة الارتباط الذاتي والارتباط الذاتي الجزائي
- 2.تم اختيار أفضل نموذج من بين النماذج الممكنة باستخدام معايير المفاضلة (اقل قيمة لتباين النموذج ومعيار AIC) وتم فحص ملاءمة النموذج المقترح إحصائياً من تحليل دالة الارتباط الذاتي للبواقي والتوزيع الطبيعي للبواقي .
- 3.وجد أن النموذج الملائم والكفوء لتمثيل بيانات السلسلة الزمنية القوية هو النموذج الموسمي المضاعف من الدرجة $ARIMA(4,2,1)$.
- 4.يستنتج الباحث أن 'طريقة التنقية المكيفة' تظهر نتائج أدق من نماذج بوكس - جنكيز Box - Jenkins method وذلك بسبب ظهور تنبؤات متناسقة مع مثيلتها في السلسلة الأصلية وقدمت لنا تنبؤات مستقبلية لواقع حالة التلوث الحاصلة في مياه نهر دجلة مستقبلاً .
- 5.كلما زاد تقسيم البيانات إلى مجاميع أكثر كانت الدقة اكبر .
- 6.يتضح أن طريقة التقدير هي الأفضل لكونها هي طريقة غير متحيزة .

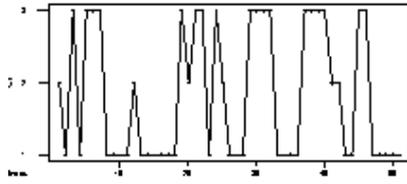
-التوصيات : من خلال النتائج التي تم التوصل إليها نوصي بما يلي:

- 1.الأخذ بنتائج هذا البحث والصيغة المعتمدة للتنبؤ من قبل الجهات ذات العلاقة لاعتماد الأسلوب العلمي الملائم في التنبؤ .
- 2.تعميم هذا البحث الى دراسات مناظرة وإجراء مقارنة بينها .
- 3.يوصي الباحث بأخذ بيانات أكثر للحالات المرمره بتوسيع عدد المجاميع المرمره للبيانات
- 4.يوصي بإجراء عدة طرائق لتقدير المعلمات و المقارنة بينها باستخدام المحاكاة .
- 5-يوصي بالاهتمام بشبكة الامطار ومراقبتها بسبب التلوث الذي يحمل مواد سامه وقتل الاحياء في نهر دجلة

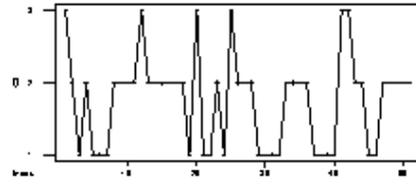
المصادر :

- 1.الازدي، أيمن سليمان محمد. (2002). "تقدير رتبة سلاسل ماركوف مع تطبيق على سلاسل الـ DNA"، رسالة ماجستير، كلية علوم الحاسبات والرياضيات، جامعة الموصل، العراق.
2. تاج , لطفي & سرحان , عمار (2002) "مقدمة في العمليات العشوائية " كتاب دار الطباعة والنشر
- 3 . الخضيرى , محمد قدوري عبد , (1996)"دراسة مقارنة لطرائق التقدير والتنبؤ لبعض نماذج بوكس – جنكيز الموسمية "رسالة ماجستير إحصاء ,كلية الإدارة والاقتصاد ,جامعة بغداد , العراق .
- 4.أسعدي ,اسماء , (2002)"تقدير الاحتمالات الانتقالية لسلاسل ماركوف المنتهية غير المستقرة "رسالة ماجستير , كلية الإدارة والاقتصاد ,جامعة بغداد , العراق.
- 5.السعدي , خليل برهان الدين , (2007)"تقلبات سعر الدينار العراقي للمدة من عام 1996 ولغاية منتصف عام 2005 والتنبؤ بسعره في المستقبل باستخدام سلاسل ماركوف " رسالة ماجستير ,كلية الإدارة والاقتصاد , جامعة بغداد , العراق .
- 6.الصراف , نزار مصطفى وشومان ,عبد اللطيف حسن , (2013) "السلاسل الزمنية والأرقام القياسية " دار الدكتور للعلوم الإدارية والاقتصادية ,شارع المتنبى ,بغداد , العراق .
- 7.الگسو، ابتهاج عبد الحميد.(2005):"استخدام الشبكات العصبية في تقدير رتب سلاسل ماركوف مع التطبيق على سلسلة جبل بطمة في محافظة نينوى"، أطروحة دكتوراه ،كلية علوم الحاسبات والرياضيات ،جامعة الموصل، العراق .
- 8.كنيهر , عباس لفته , (2010) "بناء نموذج سلاسل زمنية للتنبؤ بإعداد المرضى المراجعين لمستشفى الزهراء في محافظة واسط بحث منشور في مجلة كلية الادارة والاقتصاد ,جامعة واسط ,العراق.
- 9-Akaike, "A new look at the statical model Identification", IEEE, transaction on automatic control, vol. (AC-19), No. (6), p. (716), 1974.
- 10-Anderson ,O.D.(1976)"Time series analysis and Forecasting "Butter Werlhs London and Boston.
- 11- Box, G.P. and Jenkins, G.M. (1976). "Time Series Analysis Forecasting and Control", Revised Edition Holden-Day Inc. San Francisco .
- 12- Lee.T .C & Judge .G.G & Zellner .A.(1970) "Estimating the Parmeters of Markov Propability Model From augregate tiam series data " (2and edition) North Holland Porbishim compauy.Nothevland Amsterdam.
- 13-Schnatter .S, Pamminger. C,(2009)"Bayesian Clustering of Categorical Time Series Using Finite Mixtures of Markov Chain Models " Department of Applied Statisyics Johannes Kepler University Linz Altenbergerstrabe 69,A-4040 Austria.
- 14-Wei, W. W. S. (1990): "Time Series Analysis – Univariate and Multivariate Methods", Addison – Wesley Publishing Company, Inc., The Advanced Book Program, California, U.S.A .

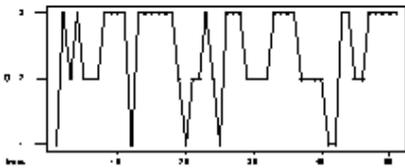
ملحق الاشكال



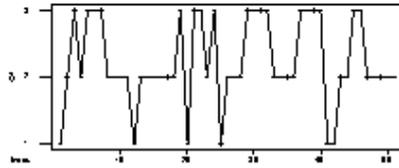
(C1)



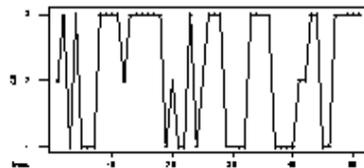
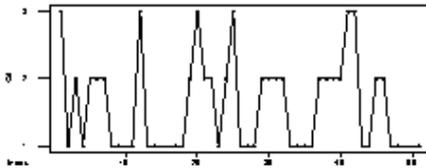
(C2)



(C3)

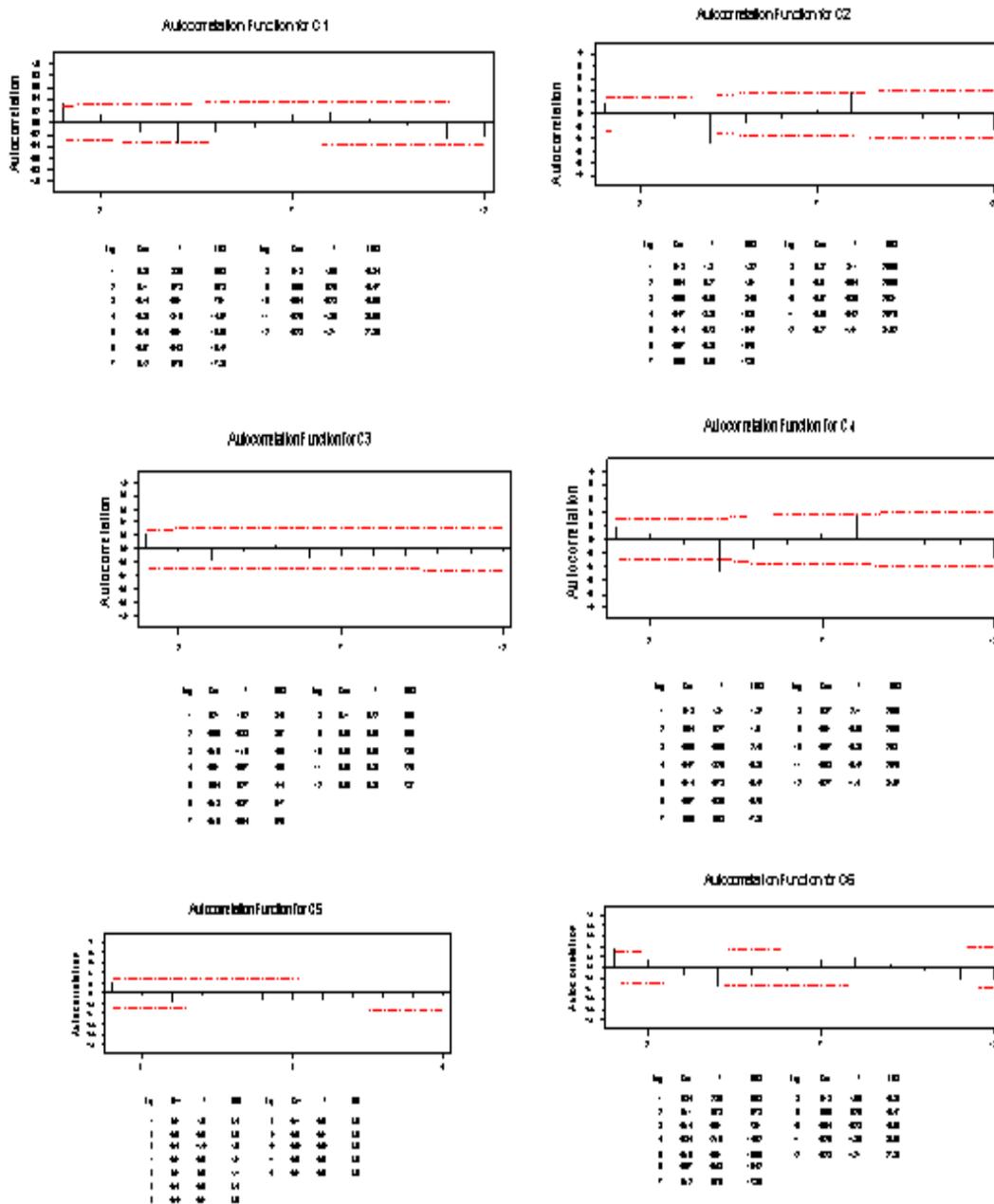


(C4)



شكل (1)

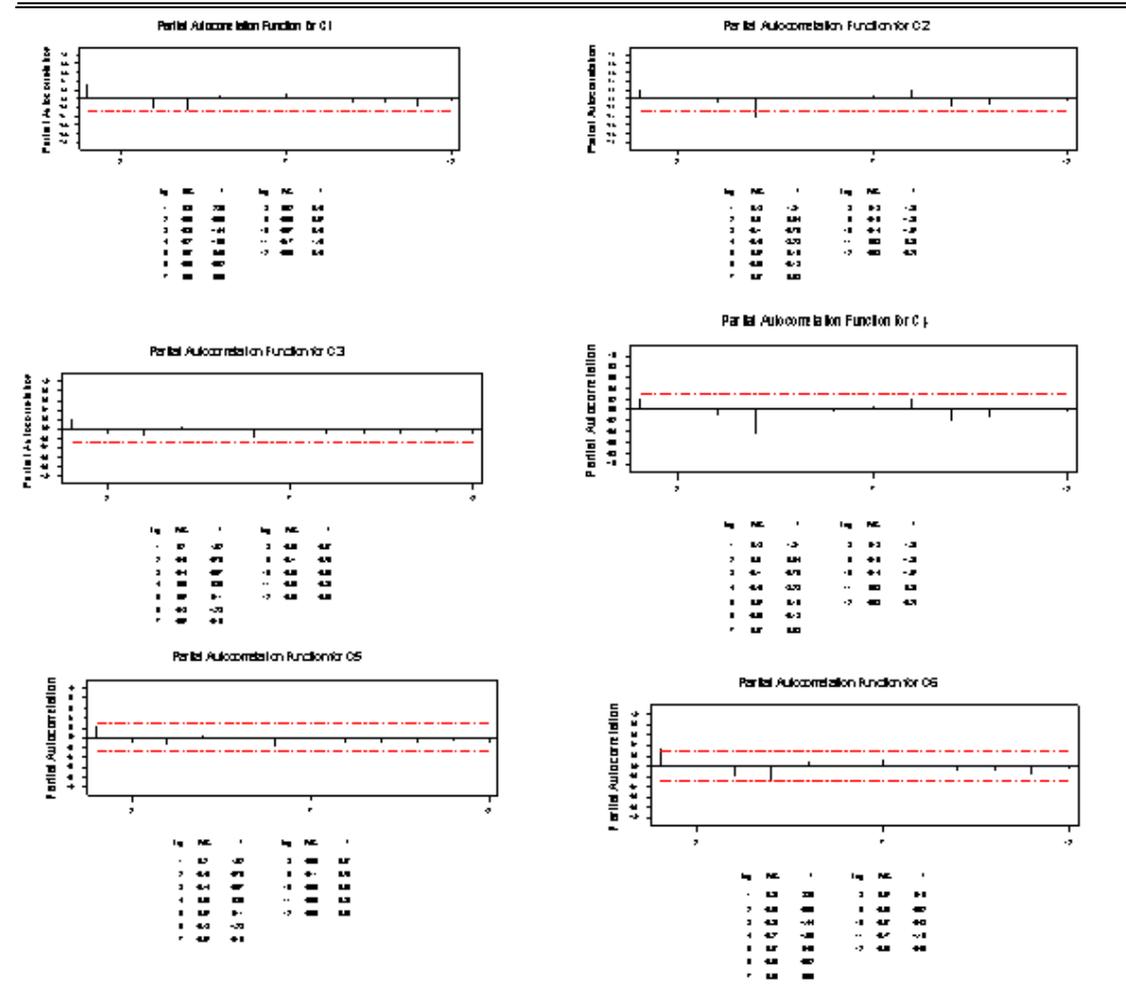
يمثل رسم السلسلة للحالات الست المرمزة [Minitab 19]



شكل (2)

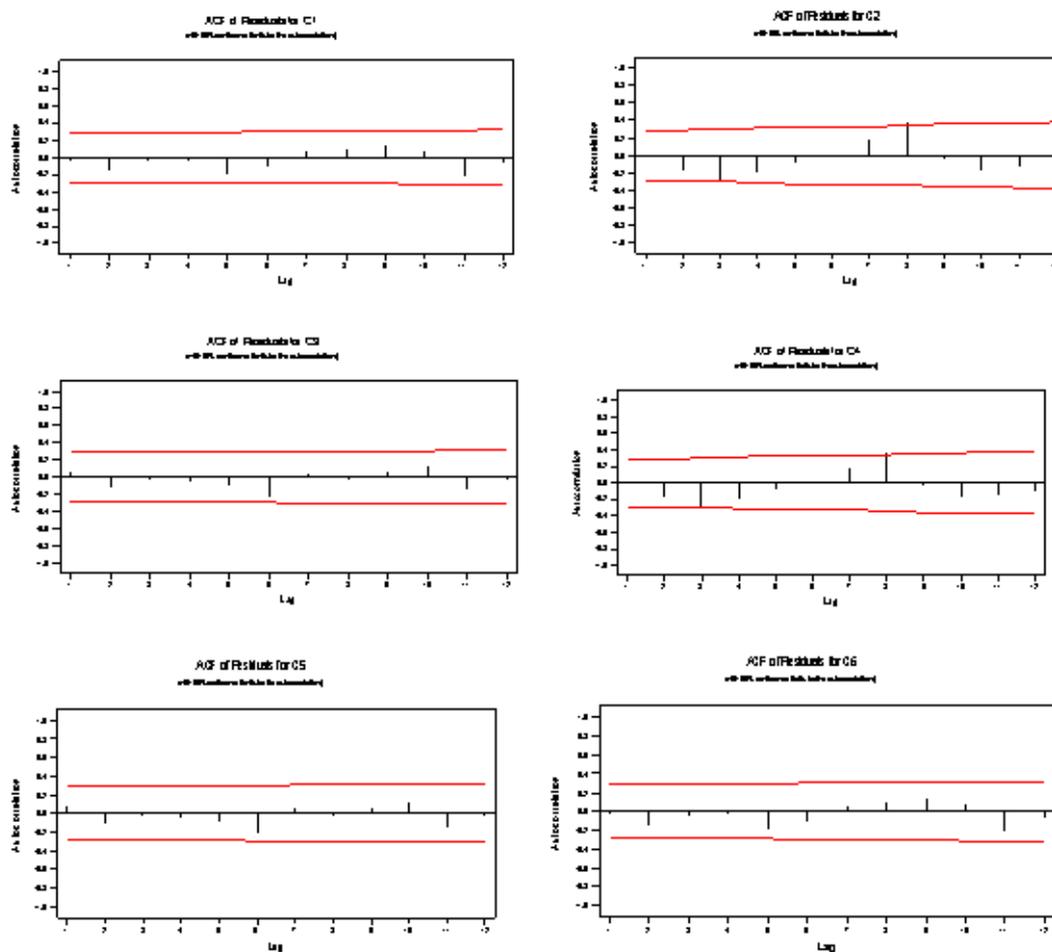
يمثل دالة الارتباط الذاتي (ACF) للسلسلة [Minitab 19]

استخدام طريقتي بوكس - جنكيز و التنقية المكيفة لتحديد افضل نموذج للتنبؤ بتلوث مياه نهر دجلة باستخدام السلاسل الزمنية القوية.....



شكل (3)

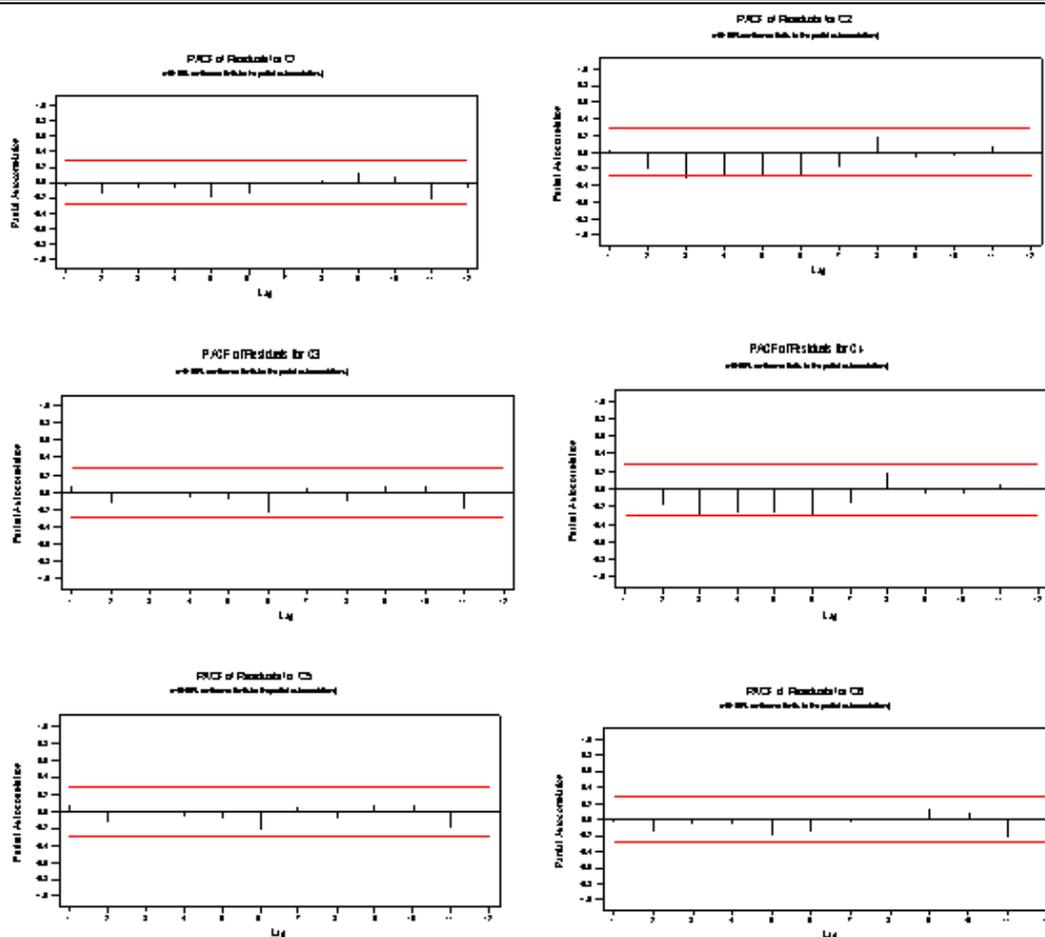
يمثل دالة الارتباط الذاتي الجزئي (PACF) للسلسلة [Minitab 19]



شكل (4)

يمثل دالة الارتباط الذاتي لاختبار الأنموذج [Minitab 19]

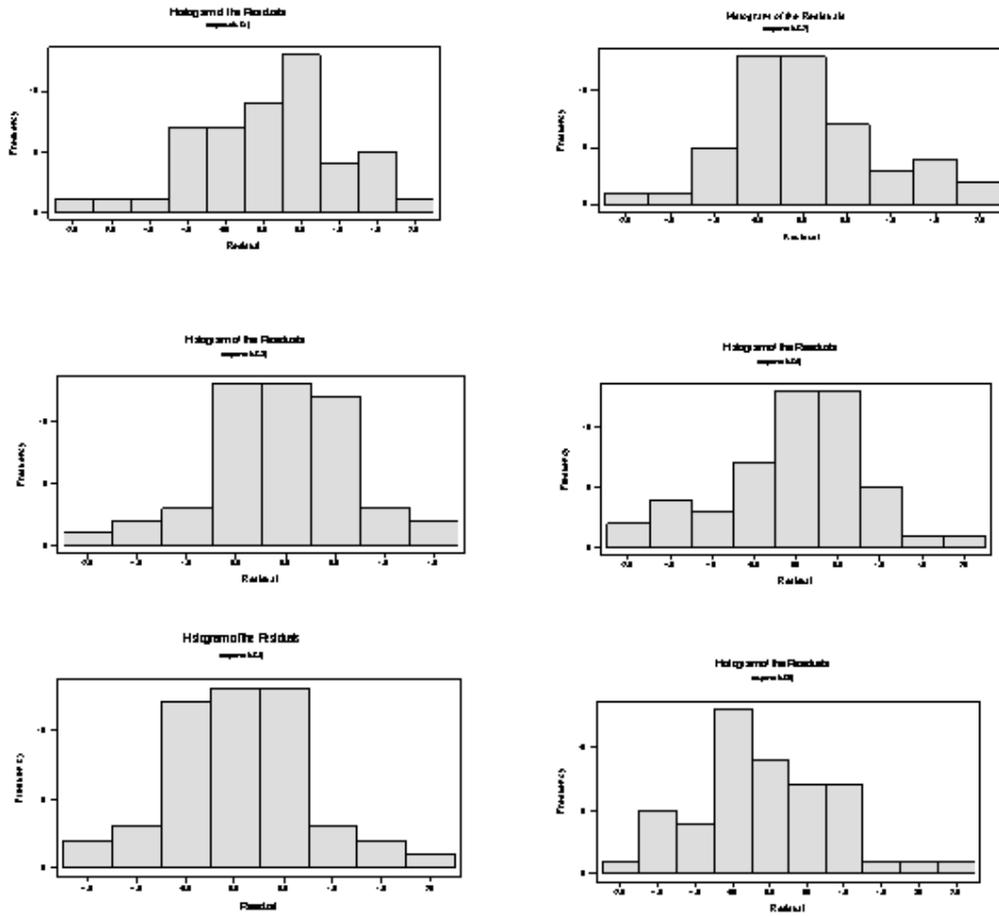
استخدام طريقتي بوكس - جنكيز و التنقية المكيفة لتحديد افضل نموذج للتنبؤ بتلوث مياه نهر دجلة باستخدام السلاسل الزمنية القوية.....



١

شكل رقم (5)

يمثل دالة الارتباط الذاتي الجزئي لاختبار الأنموذج

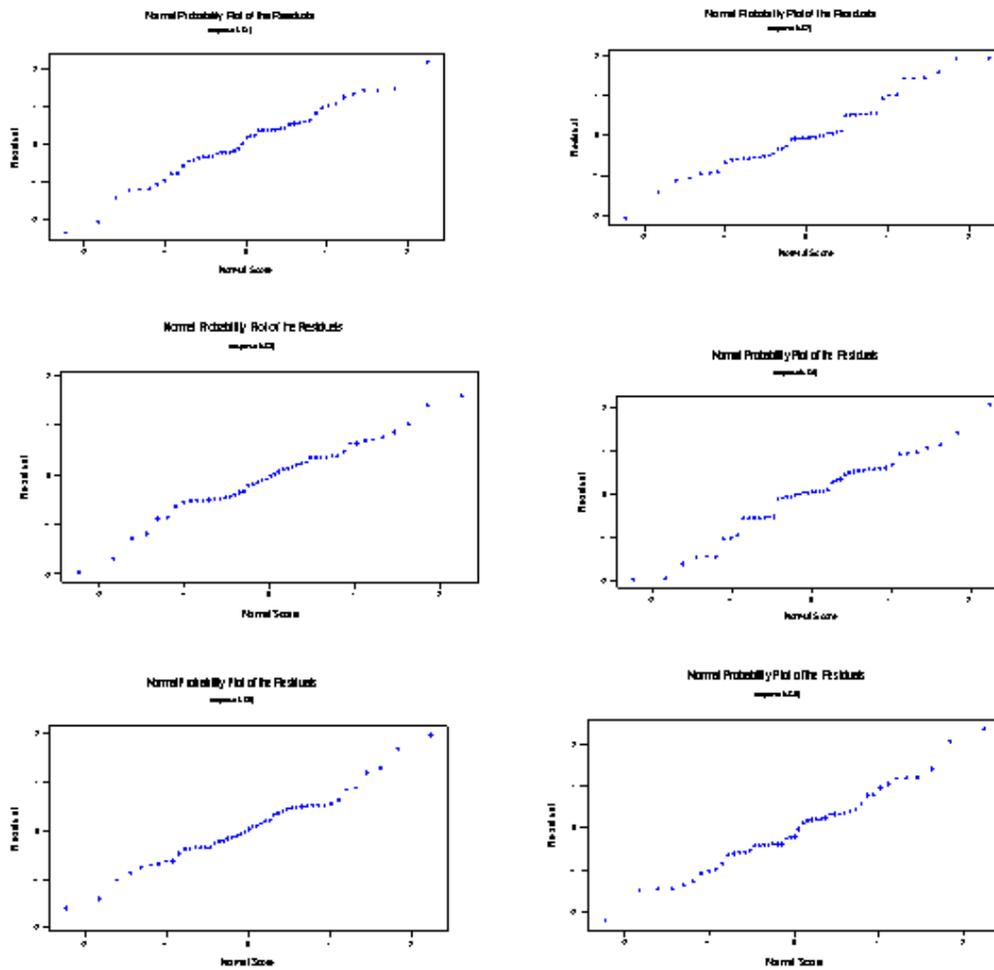


شكل (6)

يمثل توزيع البواقي لاختبار النموذج [Minitab 19]

استخدام طريقتي بوكس - جنكيز و التنقية المكيفة لتحديد افضل نموذج للتنبؤ بتلوث مياه نهر دجلة باستخدام السلاسل الزمنية القوية.....

+



شكل (7) يمثل احتمال انتشار البواقي لاختبار الأنموذج [Minitab 19]