

التنبؤ بتلوث الهواء في مدينة بغداد باستخدام السلاسل الزمنية

نشأت جاسم محمد

احمد طلال جبار

الجامعة التقنية الوسطى / الكلية التقنية الادارية/ قسم تقنيات المعلوماتية

ahmedtq15@gmail.com

nashaat1974@gmail.com

الخلاصة

يعتبر الاهتمام بالبيئة احد الاولويات المهمة في البلدان المختلفة ، ويشكل التلوث البيئي اهم مصدر لتهديد البيئة حيث وصلت مستويات التلوث في جوانب البيئة المائية والجوية والارضية الى حدود خطيرة مما استدعى اهتمام الباحثين في مختلف العلوم للاهتمام بالابحاث التي من شأنها تقليل تلك المسببات ومراقبتها وتقليل نسبها ضمن الحدود المسموحة باستمرار. يعتبر تلوث الجو والهواء احد المخاطر الرئيسية في التلوث البيئي والتي لها تاثيرات مباشرة على حياة الانسان تمثلت بارتفاع درجة حرارة الارض وتآكل طبقة الاوزون بسبب انبعاث الغازات الخطير مباشرة الى الجو واهمهما غازي ثنائي أكسيد النتروجين (NO_2) .

بسبب التطور الكبير في تسجيل البيانات من خلال الاجهزة الرقمية الحديثة ، اذ تاخذ تلك البيانات شكل السلاسل الزمنية، وباستخدام تلك البيانات أُقترحت العديد من النماذج الرياضية لنمذجة السلوك الحركي للعديد من الملوثات لاستخدامها لغراض السيطرة والتنبؤ .

في هذا البحث سيتم توظيف طرائق واستخدام منهجية بوكس وجينكس متمثلة بأنموذج الانحدار الذاتي الاوساط المتحركة المتكاملة Auto Regressive Integrated Moving Average. حيث شمل البحث الملوث ثنائي أكسيد النتروجين (NO_2) في مدينة بغداد للمدة (2015-2017) وبمعدل اسبوعي 157 مشاهدة .

وقد بين البحث إن السلسلة الزمنية للمتغيرين مستقرة ، اذ كان النموذج الملائم لبيانات NO_2 هو $ARIMA(1,0,0)$.

الكلمات المفتاحية: التنبؤ ، تلوث الهواء ، السلاسل الزمنية ، منهجية بوكس وجينكس ، نماذج $ARIMA$.

Forecasting Air Pollution Using Time Series

Nasshat Jasim Mohammed

Ahmed Tallal Jabbar

Middle Technical University/ Technical College Of Management/ Baghdad

Abstract

Concern for the environment is an important priority in different countries; Environmental pollution is the most important source of environmental threat. Pollution levels in the water, air and land environment have reached serious limits, requiring researchers in various sciences to take care of researches that reduce, monitor and reduce their causes within the limits allowed. Air pollution is one of the main threats to environmental pollution, which has a direct impact on human life. It is due to the increase in the temperature of the earth and the depletion of the ozone layer due to the dangerous emissions of gases directly into the atmosphere, mainly NO_2 and SO_2 .

Because of the great evolution in recording data through modern digital devices, as these data take the form of time series, and using these data suggested several mathematical models to model the behavior of many of the pollutants for use for control and prediction .

In this research, the use of the Box & Jenkins method, Auto Regressive Integrated Moving Average (ARIMA). The polluted letter included NO_2 in Baghdad for the period 2015-2017 and a weekly average of 157 views.

This research showed that the time series of the variables is stable, Regressive Integrated Moving Average models. The appropriate model for the NO_2 data is $ARIMA(1.0.0)$.

Keywords: Forecasting, Air Pollution, Time Series, Box & Jenkins method, ARIMA

المقدمة

لقد اهتم الإنسان منذ أقدم العصور باستشراف المستقبل والتنبؤ بتغييراته والتنبؤ بقيمها. وعلى مستوى المؤسسات والدول فإن قراءة المستقبل تعد عملية بالغة الأهمية لارتباطها الصميمي الوثيق مع الخطط التخطيطية والتنمية المختلفة.

تقع هذه البحث في أربعة مباحث خصص المبحث الأول لكي يكون مقدمة عامة حول مشكلة البحث واهداف البحث ومنهجية وهيكلية البحث فضلاً عن الدراسات السابقة ذات الصلة بالبحث وخصص المبحث الثاني للجانب النظري لشرح مقدمة عامة للسلاسل الزمنية وأهم الطرائق والأساليب المستعملة في التنبؤ وخصص المبحث الثالث للجانب التطبيقي والعملية لإجراء التحليل الاحصائي والتنبؤ من خلال التطبيق على البيانات الرقمية لحالة ملوثات الهواء لمحافظة بغداد منطقة الوزيرية حيث كان مقاييس ملوثات الهواء وأستنتاج النتائج النهائية للتنبؤ بالنماذج المستخدمة في البحث وأوجز البحث في المبحث الرابع أهم الاستنتاجات والتوصيات التي خرج بها كما ألحق البحث بمجموعة من المصادر والملاحق بقصد إكمال الفائدة المرجوة .

يعد تحليل السلاسل الزمنية من أهم الطرق العلمية المستخدمة في التنبؤ ومن أبرز النماذج المستخدمة في تحليل السلاسل الزمنية (الموسمية وغير الموسمية) نماذج الإنحدار الذاتي المتكاملة مع المتوسطات المتحركة والتي تمت صياغتها من قبل الإحصائيين عام 1970م ولذلك تسمى نماذج بوكس وجنكنز وتعتمد هذه النماذج على الدمج بين نماذج الإنحدار الذاتي ونماذج المتوسطات المتحركة وتتلخص مشكلة البحث في الأسئلة التالية : ما هي نماذج بوكس - جنكنز ، وما هي منهجيتها ، وكيف يتم بناء النموذج ، وكيف تم إستخدامها في بيانات السلاسل الزمنية.

المبحث الاول : منهجية البحث

مشكلة البحث

مع نمو الكثافة السكانية وزيادة تعقيدها، يمكن تحديد العديد من العوامل المؤثرة في التلوث مثلًا (الغازات الملوثة والشوائب وعوادم السيارات والمصانع والرياح وما الى ذلك). ولذلك، أصبحت عملية التنبؤ على المدى البعيد أكثر تعقيداً وهناك حاجة إلى توقعات أكثر دقة. العلاقة بين التلوث بالغازات الملوثة والعوامل الخارجية هي معقدة ولا خطية، مما يجعل من الصعب جداً لنموذج من خلال التقنيات التقليدية التنبؤ بشكل دقيق . ومن هنا ظهرت مشكلة البحث التي تتمثل بأستخدام نماذج تقنيات تنقيب البيانات الخاصة بالتنبؤ والتحليل الظواهر البيئية للحصول على وصف دقيق وشامل وتنبؤات بعيدة المدى تخدم متخذي القرار.

اهمية البحث

تهدف كل الحكومات والمؤسسات إلى التوسع والنمو وتحقيق معدلات مرضية من الإستقرار والتطور على مستوى الدولة بأجهزتها المختلفة لتحقيق مستوى مقبول من الرفاهية الإقتصادية والإجتماعية للمجتمع والإستقرار المالي والنقدي للدولة ، وتعمل التنبؤات الدقيقة على رسم السياسات المستقبلية في كافة الاصعدة والمجالات ، إذ إنها تعمل على تقدير وتوقع أكثر الإحتمالات دقة وموضوعية لمسار الظاهرة محل الدراسة في المستقبل وبيان الإتجاهات لتطورها ومعدلات نموها والتغيرات المصاحبة لها قبل اتخاذ أي قرار بشأنها، لذلك التنبؤ يعتبر وسيلة لتحقيق عملية التخطيط للمستقبل على الوجه الأكمل وذلك بإستعمال التنبؤ والإستفادة منه لتوجيه الظروف المتاحة والحالية وفقاً للأهداف المراد بلوغها في ضوء الإمكانيات المتاحة لها.

اهداف البحث

يعدّ تلوث الهواء من المشكلات البيئية المهمة والخطيرة التي تواجه الانسان في الوقت الراهن لما له من اثر مباشر على حياة وصحة الانسان والكائنات الحية الاخرى(الحيوان والنبات)، ونظراً لأهمية هذا الموضوع وخطورته وتأثيره المباشر على صحة الانسان تم العمل في هذه البحث الى اختيار افضل أنموذج خطي للعوامل المؤثرة على تلوث الهواء في مدينة بغداد ولقد تم اختيار الملوث الرئيسي ثنائي أوكسيد النيتروجين NO_2 ولمدة ثلاث سنوات وبمعدل اسبوعي (157) اسبوع وذلك من خلال:-

- 1- تطبيق نماذج تنبؤ بالتلوث اعتماداً على النماذج التقليدية المتمثلة بالنماذج الاحصائية لتساعد الادارة في عمليات اتخاذ القرار .
- 2- برمجة كود في لغة برمجية متخصصة بالتحليل الاحصائي وكفوء لأدارة البيانات واستخدام نموذج التنبؤ المذكورة حيث تم استخدام برنامج R لتحليل هذه النماذج والحصول على تقارير تلبي حاجة القائمين في وزارة الصحة والبيئة العراقية.

البيانات المستخدمة في البحث

تم الحصول على البيانات الخاصة بالبحث والتي تتعلق بالتلوث البيئي في مدينة بغداد من سجلات وزارة الصحة والبيئة في منطقة الوزيرية ومن قسم ملوثات الهواء وتم اختيار الملوث الرئيسي ثنائي أوكسيد النيتروجين NO_2 ولمدة ثلاث سنوات وبمعدل اسبوعي (157) اسبوع وذلك من عام 2015 الى عام 2017 وهذه البيانات مقاسة بأجهزة حديثة ومتطورة ذات منشأ ياباني وهي موضحة في الملحق رقم (1).

حدود البحث

الحدود الموضوعية والمكانية : بناء نماذج باستخدام احد لغات البرمجة الكفوءة من الناحية الاحصائية للتنبؤ بنسب تلوث الهواء في مدينة بغداد باستخدام الطرق الاحصائية ليسانس موظفي شعبة الهواء

في وزارة الصحة والبيئة على التنبؤ بملوثات الهواء بدقة عالية.

الحدود الزمانية : بيانات تاريخية للعام 2015-2017 وبمعدل اسبوعي اي 157 اسبوع ولمتغير واحد (ملوث) استخدمت لتجربة وتقييم اداء البرامج والطرق المستخدمة.

الدراسات السابقة

إن موضوع البحث يحتاج الى دراسات معمقة بخصوص التنبؤ وفيما يلي وصف ملخص لبعض البحوث المنشورة ذات العلاقة بدراستنا :-

- في عام 2000 قام الباحثان Jehng & Shang بتوقع جودة الهواء في تايوان باستخدام السلاسل الزمنية التقليدية والشبكات العصبية. حيث جمع بيانات الملوثين ثنائي اوكسيد الكبريت والاوزون من محطتين رئيسيتين ومحطة صناعية، وكانت نتائج توقع الشبكة العصبية ل 24 خطوة مرضية وجيدة ومتفوقة على النموذج التقليدي ARIMA. [9]

- في عام 2008 قام الباحثون Ming, Yafeng & Min بلتنبؤ بتركيزات ملوثات الهواء كل ساعة بالقرب من قوانغتشو، الصين. وتُصنيف العوامل التي تؤثر على تركيزات الملوثات إلى أربع فئات: ذات الصلة حركة المرور ، وتركيز انبعاث المصانع ، والنشاط البركاني،الكثافة السكانية . تم قياس متوسطات كل ساعة من هذه العوامل المؤثرة وتركيزات أول أكسيد الكربون وثنائي أكسيد النيتروجين والمواد الخاصة والأوزون في ثلاثة مواقع مختارة إذ أظهرت دراسة مقارنة أن نماذج الشبكات العصبية تتفوق على نماذج الانحدار الخطي المتعددة. [10]

- في عام 2009 قام الباحثان Ujjwal & V. K. Jain باستخدام نماذج ARIMA و ARMA بالتنبؤ بتلوث الهواء (NO ، CO ، O_3)

Wavelet-GRNN، EMD-GRNN، GRNN
 GRNN و Wavelet-SVR. لذلك يمكن أن
 تكون النماذج الهجينة المقترحة تستخدم كأدوات
 فعالة وبسيطة للتنبؤ بالهواء والتنبؤ به وكذلك
 للإدارة.

المبحث الثاني: الإطار النظري للبحث

Time Series and مكوناتها ومكوناتها components

يمكن تعريف السلاسل الزمنية بأنها المشاهدات
 خلال الزمن، أي هي مجموعة من المشاهدات التي
 تتولد بالتوالي خلال الزمن وهذه المشاهدات هي بيانات
 مسجلة لظاهرة معينة خلال فترات زمنية سابقة لبعض
 الحالات الاجتماعية والاقتصادية والفيزيائية وغيرها من
 الظواهر التطبيقية، وتكون متسلسلة ومرتبطة حسب
 الزمن وغالباً تكون الفترات الزمنية بين مشاهدة وأخرى
 متساوية.

وعادة تكون المشاهدات المتتالية غير مستقرة أي تعتمد
 على بعضها البعض، ويمكن تمثيل السلاسل الزمنية
 على شكل رسم بياني يوضح أو يكشف وجود ذبذبات أو
 تحركات مميزة توجد بدرجات مختلفة كلها أو بعضها
 وتحليل هذه التحركات له أهمية كبيرة في كثير من
 الاستخدامات منها مشكلة التنبؤ. [1][6][7]

كذلك تعرف السلسلة الزمنية رياضياً بأنها متتابعة من
 المتغيرات العشوائية ضمن فضاء من الاحتمالية متعددة
 المتغيرات وترمز له بالرمز (t) والذي يعود الى مجموعة
 من الفترات (T) وعادة مايرمز للسلسلة الزمنية
 $\{X(t); t \in T\}$ فإذا كان الزمن (t) يأخذ قيماً مستمرة
 (Continuous) فإن السلسلة الزمنية تسمى مستمرة في
 الزمن ويرمز لها $\{X(t); -\infty \leq t \leq \infty\}$ أما اذا
 كان الزمن (t) يأخذ قيماً متقطعة (Discrete) اي
 $\{t=0, \pm 1, \pm 2, \dots\}$ فإن السلسلة الزمنية متقطعة في
 الزمن ويرمز لها $\{X(t); t = 0, \pm 1, \pm 2, \dots\}$.
 ومن الامثلة على السلاسل الزمنية :-

NO2) في مناطق تركيز حركة المرور في
 المناطق الحضرية نيودلهي، الهند. تم تطبيق
 التحول المناسب لتثبيت التباين على كل سلسلة
 زمنية من أجل جعلها ثابتة التباين بطريقة متنسقة.
 وتطبيق معايير المعلومات المختلفة فكانت نسبة
 الخطأ المطلق MAPE ل CO و NO2 و NO
 و O3، وهي 13.6 و 12.1 و 21.8 و
 24.1% على التوالي. [21]

• في عام 2013 قام الباحثون Snezhana, Atanas, Desislava & Doychin بتطبيق تحليل العوامل ومنهجية Box-Jenkins لدراسة تركيزات ملوثات الهواء الأولية مثل NO، NO2، SO2، PM10، NOx، ومستوى الاوزون O3 في بلدة Blagoevgrad، بلغاريا خلال فترة سنة واحدة، استناداً إلى القياسات بالساعة. باستخدام تحليل العوامل PCA و Promax rotation، تم اكتشاف علاقة خطية متعددة بين الملوثات الستة.

• وفي نفس العام قام الباحث وهيب بناء أنموذج ARIMA لمشكلة البطالة في مصر باستخدام بيانات متوفرة على الشبكة العنكبوتية لمدة من 1990 لغاية 2010 إذ تبين أن الانموذج الافضل لتمثيل السلسلة هو ARIMA(1,1,2) وذلك من خلال بعض معايير الخطأ المستخدمة حيث وجد بأن هناك زيادة مستمرة في البطالة من خلال ملاحظة القيم التنبؤية. [2][8]

• عام 2017 قام مجموعة من الباحثين Suling, Xiuyuan, Haixia, Jianming, Yuanyuan & Jinxing بتطبيق نماذج هجينة للتنبؤ بتلوث الهواء في الصين ومقارنة نتائج هذه النماذج مع نماذج ARIMA ونموذج SAR، تظهر نتائج التنبؤات أن النموذجين الهجين المقترحين يتفوقان على ARIMA & SVR والنماذج الهجينة هي

رياضي مناسب .ولتقدير أي سلسلة زمنية يتم التحقق من استقرارية هذه البيانات من حيث طبيعة التذبذبات الحاصلة في هذه الظاهرة لأي استقرارية البيانات لها دور كبير في عمليات التنبؤ .

ومن هنا يمكن أن نقول ان بيانات السلسلة الزمنية تكون مستقرة إذا كانت في حالة موازنة أحصائية (Statistical Equilibrium) أي ان السلسلة الزمنية تمتلك (وسط حسابي وتباين والتباين الذاتي) كلها ثابتة مع الزمن ويمكن ان تكون السلسلة الزمنية مستقرة اعتماداً على الرسم البياني للملاحظات وعلى الرسم البياني لمصفوفة الارتباط الذاتي Correlgram اي في حالة السلسلة مستقرة فأن Correlgram منحدر بسرعة الى الاسفل والعكس في اذا كانت السلسلة غير مستقرة فإن Correlgram ينحدر ببطيء . التعبير للتوزيع الاحتمالي للملاحظات المستقرة التي لاتتأثر بتغير بالزمن اي ان:- [5][6]

$$F_{x_{t1}, x_{t2}, \dots, x_{tm}}(x_{t1}, x_{t2}, \dots, x_{tm}) = F_{x_{t1+k}, x_{t2+k}, \dots, x_{tm+k}}(x_{t1+k}, x_{t2+k}, \dots, x_{tm+k})$$

وأن (K) تمثل الازاحات (Lags) .

إذا فإن السلسلة الزمنية تكون مستقرة عند تحقيق الشروط الآتية :-

1. ثبوت الوسط الحسابي $E(X_t) = \mu$
2. ثبوت قيمة التباين $\text{var}(X_t) = \sigma_t^2$
3. إن التباين المشترك الذاتي (Auto Covariance Matrix) للسلسلتين (X_t, X_{t+k}) يعتمد على الازاحة (K) فقط أي

$$\gamma_k = E[X_t - \mu][X_{t+k} - \mu] ; k = 1, 2, 3, \dots, m$$

مستقرة لكونها تسير بصفة عامة في الاتجاه العام ، ولذلك لابد من تحويلها إلى سلاسل زمنية مستقرة ليسهل نمذجتها والتعامل معها ، ويمكن تحويل السلسلة الزمنية الى سلسلة مستقرة باستخدام الطريقة التالية :-

- مجموع الانتاج السنوي من النفط خلال عدة سنوات
- قيمة الصادرات السنوية خلال عدة سنوات
- قيمة المبيعات الشهرية لأحد المحلات التجارية خلال فترة معينة من الزمن
- اسعار الاسهم التجارية في سوق تبادل الأسهم

مكونات السلاسل الزمنية Time Series components

- توجد عدة عوامل تؤثر في تطور المتغير حسب الزمن تؤدي الى زيادة أو نقصان ومن اهم هذه العوامل ما يطلق عليها مركبات السلسلة الزمنية:- [14][11][1]
- الاتجاه العام Secular Trend
 - التغيرات الموسمية The Seasonal
 - التغيرات الدورية Cyclical Variation
 - التغيرات الغير المنتظمة (العشوائية) Irregular Fluctuations

استقرارية السلسلة الزمنية Stationary Time Series

إن استقرارية أو عدم استقرارية البيانات له أهمية في تحليل السلاسل الزمنية ، وكذلك ايجاد أفضل أنموذج

تحويل السلاسل الزمنية الغير مستقرة إلى سلسلة مستقرة

إن اغلب السلاسل الزمنية في الواقع العلمي التطبيقي تكون سلاسل غير مستقرة وقد ثبت ذلك من خلال الرسم البياني أو الاختبارات الاحصائية مثلاً نسب الملوثات في الهواء فهي تعتبر سلاسل زمنية غير

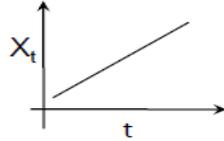
تعديل الفروق Differencing the series

استقرارية السلسلة ومعالجتها فإذا كانت X_t تتمثل بسلسلة زمنية غير مستقرة فإن التحويل لجعل هذه السلسلة مستقرة يكون بالشكل التالي:- [11][12]

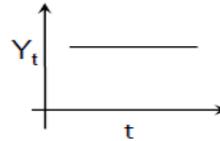
$$Y_t = X'_t = X_t - X_{t-1} \dots\dots\dots (1)$$

حيث X'_t تمثل الاختلاف الاول First order differencing

$$\{X_t\} = 2, 4, 6, 8, 10, \dots\dots\dots$$



$$\{Y_t\} = 2, 2, 2, \dots\dots\dots$$



شكل (1) يوضح اجراء تعديل الفروق في السلسلة

$$X''_t = X'_t - X'_{t-1} \dots\dots\dots (2)$$

حيث X''_t تمثل الاختلاف الثاني Second order differencing

$$\begin{aligned} &= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) \\ &= X_t - 2X_{t-1} + X_{t-2} \dots\dots\dots (3) \end{aligned}$$

تحويل الفرق وعند ادخال معامل الفرق d لنموذج $ARMA(p,q)$ تتم اضافة Integrated الى اسم الانموذج لتصبح انموذج انحدار ذاتي ووسط متحرك متكامل Autoregressive Integrated Moving Average ويرمز له $ARIMA(p,d,q)$ حيث ان p : رتبة الانحدار الذاتي

d : عدد مرات الفروق التي يتم اخذها للسلسلة الزمنية حتى تحقق الاستقرارية

q : رتبة المتوسطات المتحركة

والصيغة العامة لنماذج $ARIMA(p,d,q)$ كالآتي:-

[12][19]

$$\begin{aligned} \nabla Z_t &= \phi_1 \nabla Z_{t-1} + \phi_2 \nabla Z_{t-2} + \dots + \phi_p \nabla Z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \dots - \theta_q \varepsilon_{t-q} \\ \nabla Z_t &= \sum_{n=1}^p \phi_n \nabla Z_{t-n} + \varepsilon_t - \sum_{n=1}^q \theta_n \varepsilon_{t-n} + \varepsilon_t \dots\dots\dots (4) \end{aligned}$$

حيث ان ∇Z_t متغير عشوائي للسلسلة الزمنية بين زمنيين متغيرين

وغير موسمية) هي الطريقة التي اقترحها الباحثان (Box & Jenkins) عام 1970 وتتمثل هذه الطريقة بعملية بناء أنموذج $ARIMA(p,d,q)$ لتمثيل ظاهرة معينة والتنبؤ بها . وتتلخص عملية بناء الانموذج حسب

نماذج السلاسل الزمنية Time Series Models نماذج السلاسل الزمنية اللاموسمية)

الشائعة Non-Seasonal

نماذج الانحدار الذاتي المتكامل والمتوسطات المتحركة Autoregressive (ARIMA)(p,d,q)

Integrated Moving Average Models

تتصف معظم السلاسل الزمنية في الواقع العملي بخاصية عدم الاستقرارية Non Stationary مما يجعل تحليل هذه السلاسل الزمنية والتنبؤ بها امراً في غاية الصعوبة لذا يجب تحويلها الى سلاسل مستقرة باستعمال التحويلات اللازمة لها كما ذكرنا سابقاً . وبأستعمال

مراحل بناء الانموذج حسب منهجية بوكس جينكس (Box - Jenkins Method) :-

من الطرق الاحصائية الاحصائية المهمة لتحليل السلسلة الزمنية (المستقرة والغير مستقرة والموسمية

طريقة بوكس جينكس بالمرحل الاتية:- [6][5][3]

Stationary التحقق من استقرارية السلسلة الزمنية

Check

• تشخيص وتعريف الانموذج الذي يلائم البيانات

Identification & Diagnostic

• تقدير معاملات الانموذج المشخص

Parameters' Estimation

• أختبار ملائمة الانموذج Diagnostic

Checking

• التنبؤ Forecasting

المرحلة الاولى : التحقق من استقرارية السلسلة

الزمنية Stationary Check

وذلك برسم مخططٍ للسلسلة الزمنية ومخطط لكل من

دالة الارتباط الذاتي ACF ودالة الارتباط الجزئي

PACF لفحص استقرارية السلسلة الزمنية والاتجاه العام

للسلسلة والتغيرات الموسمية أو عن طريق الاختبار

الاحصائي جذر الوحدة لديكي وفولر (Augmented

Unit Root Tests) فإذا كانت

السلسلة غير مستقرة يطبق عليها تعديل الفروق

للحصول على الاستقرارية. [20]

أختبار ديكي - فولر الموسع Augmented

Dickey-Fuller test(ADF)

يعدُّ اختبار جذر الوحدة لديكي - فولر

الموسع (ADF) (Dickey & Fuller, 1989) من

بين الاختبارات القوية للكشف عن استقرار السلسلة

الزمنية ، في نماذج بوكس وجينكس في حالة تكون

جذور الوحدة كثيرة الحدود خارج دائرة الوحدة إذا

السلسلة تكون مستقرة ، لذا فإن اختبار ديكي وفولر

لتحديد جذر الوحدة يعمل بفرضية العدم لجذر الوحدة

باستخدام الانموذج الاتي: [20]

$$\nabla y_t = \alpha + \beta_1 t + \beta_2 t^2 + \gamma y_{t-1} + \varphi_1 \nabla y_{t-1} + \dots + \varphi_{p-1} \nabla y_{t-p+1} + \varepsilon_t \dots(5)$$

حيث ان :

$\nabla y_t = y_t - y_{t-1}$ عامل الفرق حيث

α ثابت الاتجاه العام للأنموذج trend

β_1 معامل اتجاه الوقت time trend

p عدد القيم المزاحة للأنحدار الذاتي AR

ε_t سلسلة عشوائية بوسط ثابت

ان فرضية العدم لجذر الوحدة

$$H_1 : \gamma < 0 , H_0 : \gamma = 0$$

حيث ان :

H_0 : تمثل فرضية العدم (y_t لديه جذر الوحدة) /

السلسلة الزمنية غير مستقرة

H_1 : تمثل الفرضية البديلة (y_t لا يمتلك جذر الوحدة)

/ السلسلة الزمنية مستقرة

يطبق الاختبار الاحصائي على قيمة γ المقدره وتحسب

قيمة (τ) كالاتي :

$$\tau = \frac{\hat{\gamma}}{\sigma_{\hat{\gamma}}} \dots\dots(6)$$

وان $\hat{\gamma}$ هو تقدير المعلمة (γ) ، وان $\sigma_{\hat{\gamma}}$ يمثل

الانحراف المعياري.

وبمقارنة القيمة المحسوبة لإحصائية (τ) مع القيم

الجدولية الخاصة باختبار ديكي- فولر

(Dickey-Fuller) فإذا كانت القيمة المطلقة

لإحصائية (τ) المحسوبة اكبر من قيمة (τ) الدرجة

فإننا نرفض الفرضية الصفرية وتكون السلسلة الزمنية

مستقرة (Stationary) ، أما اذا كانت قيمة (τ)

المحسوبة اقل من قيمة (τ) الدرجة فإننا لا نرفض

الفرضية الصفرية وفي هذه الحالة تكون السلسلة الزمنية

غير مستقرة (Non Stationary) ، لتحديد الدرجة

التي تصل السلسلة عندها الى الاستقرار فإننا نعيد

الاختبار بعد اخذ الفرق الاول فاذا استقرت السلسلة فهذا

يعني ان البيانات متكاملة من الدرجة الاولى وهكذا الى

(d) من الاختبارات.

المرحلة الثانية : تشخيص وتعريف الانموذج الذي**يلائم البيانات Identification & Diagnostic**

إن مرحلة تشخيص نماذج السلاسل الزمنية هي مرحلة مهمة من مراحل بناء الانموذج باستعمال طريقة بوكس - جينكس لكونها تمكننا من معرفة سلوك السلسلة الزمنية والتعرف على خصائصها وذلك بأستعمال البيانات التاريخية للسلسلة الزمنية والكشف عن الكيفية

جدول (1) سلوك الدالتين

الانموذج	دالة الارتباط الذاتي ACF	دالة الارتباط الذاتي الجزئي PACF
AR(p)	يتناقص اسياً	ينقطع بعد الازاحة p
SAR(P)	يتناقص اسياً	ينقطع بعد الازاحة P
MA(q)	ينقطع بعد الازاحة q	يتناقص اسياً
SMA(Q)	ينقطع بعد الازاحة Q	يتناقص اسياً
ARMA(p,q)	يتناقص اسياً	يتناقص اسياً
SARMA(P,Q)	يتناقص اسياً	يتناقص اسياً

الذاتي وللدقة اكثر فقط طور الباحث اكاكي عام (1978-1979) توسيع بيز لطريقة AIC ويسمى بمعيار بيز Bayesian information criterion ويرمز له BIC علماً ان الباحث Akaike توصل الى ان معيار BIC له اقل احتمالية في تقدير رتبة الانحدار الذاتي . وتحسب قيمة BIC كالآتي:- [18]

$$BIC = -2 \cdot \text{Ln} \hat{L} + k \cdot \text{Ln}(n) \dots\dots\dots(8)$$

حيث ان \hat{L} تمثل اعلى قيمة لدالة الاحتمالية Likelihood للانموذج M وتحسب وفق المعادلة :-

$$\hat{L} = p(x|\hat{\theta}, M)$$

$\hat{\theta}$: قيمة المعلمة التي تحتسب دالة Likelihood

x : قيمة المشاهدات

n : حجم العينة اي حجم المشاهدات الكلي

k : عدد معلمات الانموذج المقدر .

المرحلة الثالثة تقدير معلمات الانموذج المشخص**Parameters' Estimation**

بعد مرحلة تشخيص الانموذج الملائم للسلسلة الزمنية نقوم بتقدير المعلمات لهذا الانموذج المشخص والتي تتمثل $(\varphi_i, i = 1, 2, \dots, p)$ في انموذج AR(p)

ولغرض الحصول على أفضل انموذج مقدر يمثل البيانات يستعمل عدة معايير المقارنة وان الانموذج الافضل يظهر عند اقل المعايير :-

معيار معلومات اكاكي Akaike s' information criterion (AIC)

اقترحه العالم الياباني هيروتوغو اكاكي سنة 1973 ويستخدم كأداة في تشخيص واختيار الانموذج الامثل ويمكن حسابه وفق الصيغة الآتية:- [17]

$$AIC = 2K - 2\text{Ln}(L) \dots\dots\dots(7)$$

حيث ان k : عدد المعالم في parameters الانموذج L : تمثل اعلى قيمة لدالة الاحتمال Likelihood للانموذج

وبعد إجراء التجارب على عدد من النماذج يكون الانموذج الافضل الذي له اقل قيمة AIC والذي يمكن استخدامه.

معيار معلومات بيز Bayesian information criterion (BIC)

أظهر الباحث shibata عام 1976 ان معيار AIC يميل الى اخذ رتبة اعلى من الرتبة المحددة للانحدار

للبيانات ، وبشكل عام فإن اختبار Ljung-Box يكون وفق الفرضيات الآتية :-

H_0 : الانموذج يلائم البيانات

H_1 : الانموذج لا يلائم البيانات

يعرف الاختبار الاحصائي كالاتي :

$$Q = n(n + 2) \sum_{k=1}^m \frac{\hat{f}^2}{n-k} \dots \dots \dots (9)$$

حيث \hat{f}^2 تمثل قيم الارتباط الذاتي المقدر للسلسلة

الزمنية عند الازاحة k و m تمثل عدد القيم المزاحة

الاختبار يرفض فرضية العدم اذا كان

$Q > \alpha$ ، حيث ان α هي قيم جدولية

بتوزيع مربع كاي Chi-square بدرجة حرية ومستوى

معنوية α

ولأن هذا الاختبار مطبق على سلسلة البواقي فإن درجة

الحرية يجب حسابها على عدد معاملات الانموذج المقدر

اي : $h = m - p - q$

حيث ان p,q تمثل درجة نماذج الانحدار الذاتي

والمتوسط المتحرك .

المرحلة الخامسة التنبؤ Forecasting

المرحلة الاخيرة من مراحل تحليل السلاسل الزمنية هي

مرحلة التنبؤ ولا يمكن الانتقال الى هذه المرحلة الا بعد

اتمام جميع الفحوصات الاحصائية الضرورية لتشخيص

الانموذج الذي يتم اختياره كما مر ذكره في مراحل

التحليل حيث ان هذه المرحلة هي مرحلة التطبيق

العملي للانموذج المقترح من خلال الحصول على القيم

التنبؤية المتوقعة للسلسلة وبعد تحديد درجات الانموذج

التنبؤية المتوقعة للسلسلة وبعد تحديد درجات الانموذج

(p,d,q) وتقدير الانموذج ومن ثم استخدامه للتنبؤ وذلك

بأحلال القيم الحالية والماضية للمتغير التابع Z_t

والبواقي α_t كقيمة تقديرية لحد الخطأ للحصول على

القيمة المستقبلية الاولى Z_{t+1} وهو ما يسمى بالتنبؤ

لفترة مستقبلية واحدة ، ويمكن الحصول على القيمة

المستقبلية الثانية Z_{t+2} بأحلال القيمة المستقبلية الاولى

Z_{t+1} التي تم النوصل اليها في الخطوة الاولى للتنبؤ

و $(\theta_i, i = 1, 2, \dots, q)$ في انموذج MA(q) و φ_i

و θ_i في انموذج ARMA(p,q) . أذ توجد عدة طرق

لتقدير هذه المعلمات وأهم هذه الطرق طريقة الاماكن

الاعظم Maximum Likelihood method

Approximate التي تم الاعتماد عليها في هذا البحث

وذلك بجعل مجموع مربعات الخطأ أقل ما يمكن.

[14][4]

المرحلة الرابعة اختبار ملائمة الانموذج Diagnostic

Checking

بعد ان تم التعرف على أنموذج مبدئي وأيجاد تقديرات

معلمات الانموذج تأتي مرحلة مهمة هي مرحلة الاختبار

اي اختبار مطابقة الانموذج ومدى كفاءته وملاءمة

الانموذج ومدى صلاحيته لتمثيل بيانات السلسلة الزمنية

وذلك بأستعمال اختبارات عدة منها اختبار طبيعة

البواقي اي مدى تقارب البواقي المقدر الى التوزيع

الطبيعي ، كذلك حساب دالة الارتباط الذاتي لسلسلة

البواقي (residuals) فأذا كانت معاملات دالة الارتباط

الذاتي للبواقي تقع ضمن مستوى ثقة معين فهذا يعني

ان سلسلة البواقي عشوائية وبالتالي يمكننا القول ان

النموذج المشخص ملائم ، كذلك يمكن استخدام اختبار

بوكس وليونك Ljung-Box لمعرفة ملائمة الانموذج.

[15][12]

في حال تطبيق هذه الاختبارات وكان الانموذج غير

ملائم وسلسلة البواقي مترابطة فيجب الرجوع الى مرحلة

التشخيص والانموذج الذي يلائم البيانات وبناء أنموذج

آخر .

أختبار بوكس وليونك لسلسلة البواقي - Test Ljung-

Box For Residuals

طور الباحثان Box&Ljung في عام 1978 هذا

الاختبار ويستخدم هذا الاختبار بعد مرحلة التشخيص

والتقدير الانموذج المناسب للبيانات ، يحدد هذا

الاختبار m من الارتباطات الذاتية للبواقي ، أذا كان

الارتباط الذاتي قليل جدا نستنتج بأن الانموذج مناسب

التنبؤ بأنموذج الانحدار الذاتي المتكامل والاسواط

المتحركة ARIMA

ومن خلال تصدير البيانات الاسبوعية الى المنصة التنفيذية لبرنامج R ووفقاً للكود البرمجي في ملحق رقم (2) .

بيانات ثاني اوكسيد النيتروجين NO₂

اظهرت نتائج التحليل الوصفي ان جميع القيم الاسبوعية كانت موجبة وان اقل قيمة كانت 0.008 واكبر قيمة كانت 0.089 وان المتوسط تقريبا يساوي الوسيط اي ان البيانات تتوزع تقريبا طبيعياً ، وكما موضح في الجدول (2).

في الطرف الايمن من المعادلة ، وهكذا حتى نصل الى الدالة المطلوبة مع افتراض ان حد الخطأ خارج العينة للدالة يساوي صفر أي: [13]

معادلة التنبؤ للفترة الاولى هي :-

$$Z_{t+1} = \phi Z_t + \phi_1 Z_{t-1} + \dots \dots \dots (10)$$

$$a_t - \theta_1 a_{t-1}$$

معادلة التنبؤ للفترة الثانية هي :-

$$Z_{t+2} = \phi_1 Z_{t+1} + a_t - \dots \dots \dots (11)$$

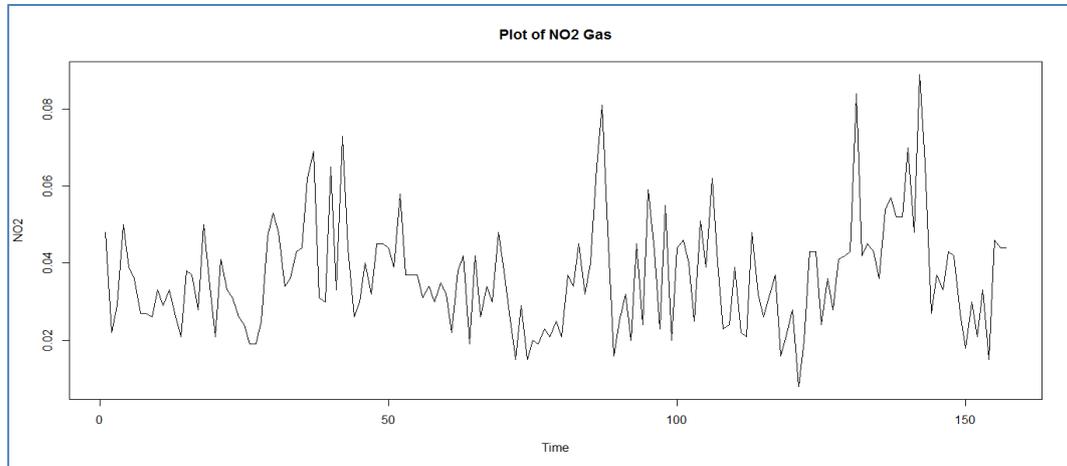
$$\theta_2 Z_t$$

المبحث الثالث: الاطار العملي للبحث

جدول (2) نتائج التحليل الوصفي

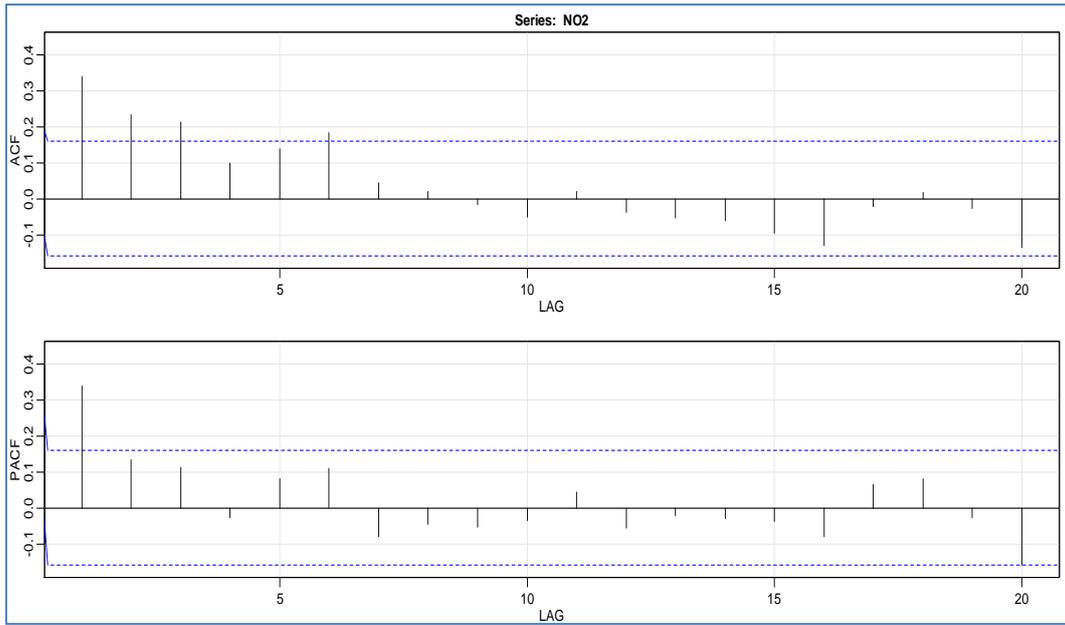
NO ₂	Min	1st Qu	Median	Mean	3rd Qu	Max
	0.00800	0.02600	0.03500	0.03668	0.04400	0.08900

من خلال رسم بيانات غاز ثاني اوكسيد النيتروجين مقابل الزمن نلاحظ تذبذب البيانات حول متوسط ثابت تقريبا ، كما ان مقدار التذبذب يزداد خلال النصف الثاني من السلسلة إذ تؤشر وجود أربعة قيم شاذة تقريبا ، وكما موضح في الشكل رقم (2).



شكل (2) بيانات غاز ثاني اوكسيد النيتروجين مقابل الزمن

كما نلاحظ ان قيم دالة الارتباط الذاتي تتناقص على نحو سريع نسبياً مع انخفاض الحدود الدنيا والعليا لفترة الثقة الخاصة بها ودالة الارتباط الذاتي الجزئي غير معنوية بعد الفرق الاول، وكما موضح في الشكل رقم (3) .



شكل (3) قيم دالة الارتباط الذاتي ودالة الارتباط الذاتي الجزئي

نقبل الفرضية البديلة التي تنص بان السلسلة الزمنية مستقرة ولا داعي لاخذ الفروق لجعلها مستقرة بالمتوسط وهذا قد يكون بسبب اخذ المتوسط الاسبوعي لها، وكما موضح في الجدول (3)

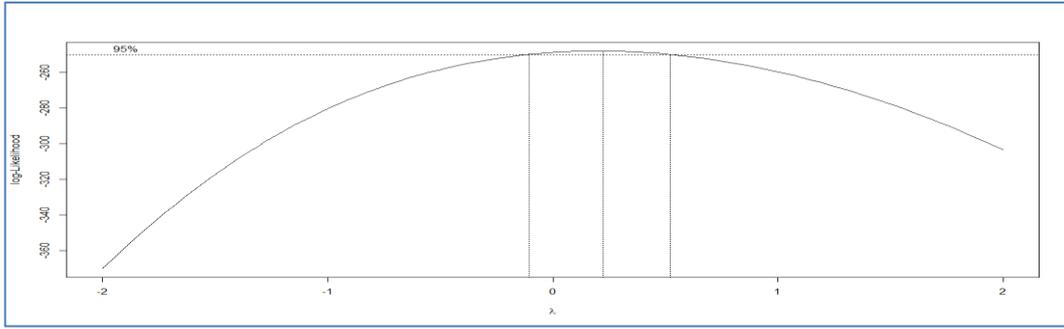
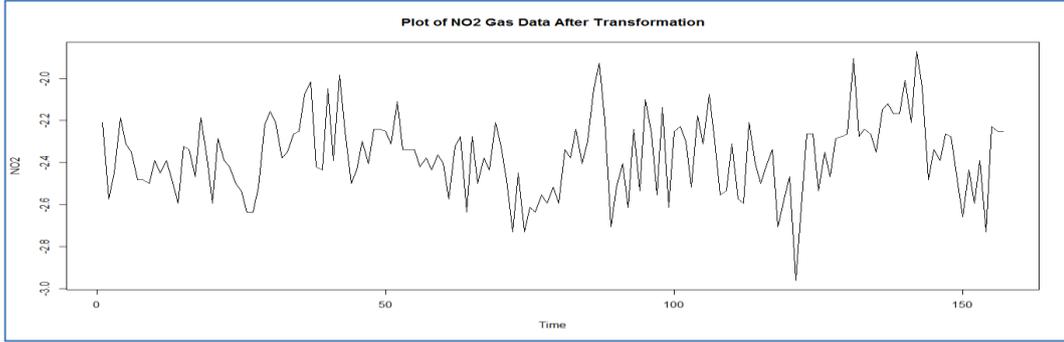
وبالاعتماد على تلك المؤشرات الصورية السابقة ومن خلال تطبيق اختبار ديكي فولر الموسع Augmented Dickey-Fuller Test المستخدم لاكتشاف عدم الاستقرار للسلسلة الزمنية ، اذ كانت قيمة احصاء الاختبار معنوية اي اننا

جدول (3) تطبيق اختبار ديكي فولر الموسع Augmented Dickey-Fuller Test

Augmented Dickey-Fuller Test
data: NO2
Dickey-Fuller = -8.7981, Lag order = 0, p-value = 0.01
alternative hypothesis: stationary

البيانات وجعل تباينها مستقر. ومن خلال تطبيق الخوارزمية الخاصة بالاختبار تبين ان قيمة λ المقدر تساوي 0.222 ، وكما موضح في الشكل رقم(4) حيث تمثل القيمة التي تعظم لوغاريتم دالة الامكان الاعظم. وبعد اجراء التحويل المناسب ستكون البيانات كما في الشكل (5).

ولاختبار استقرارية التباين تم تطبيق اختبار بوكس كوكس Box-Cox ، والذي يعتمد على تقدير المعلمة λ التي تمثل معلمة التحويل ، فاذا كانت تساوي صفراً فيتم اخذ التحويل اللوغاريتمي واذا كانت تساوي واحد فلاداعي للتحويل وغير ذلك يتم اعتماد المعادلة لاجراء التحويل المناسب على

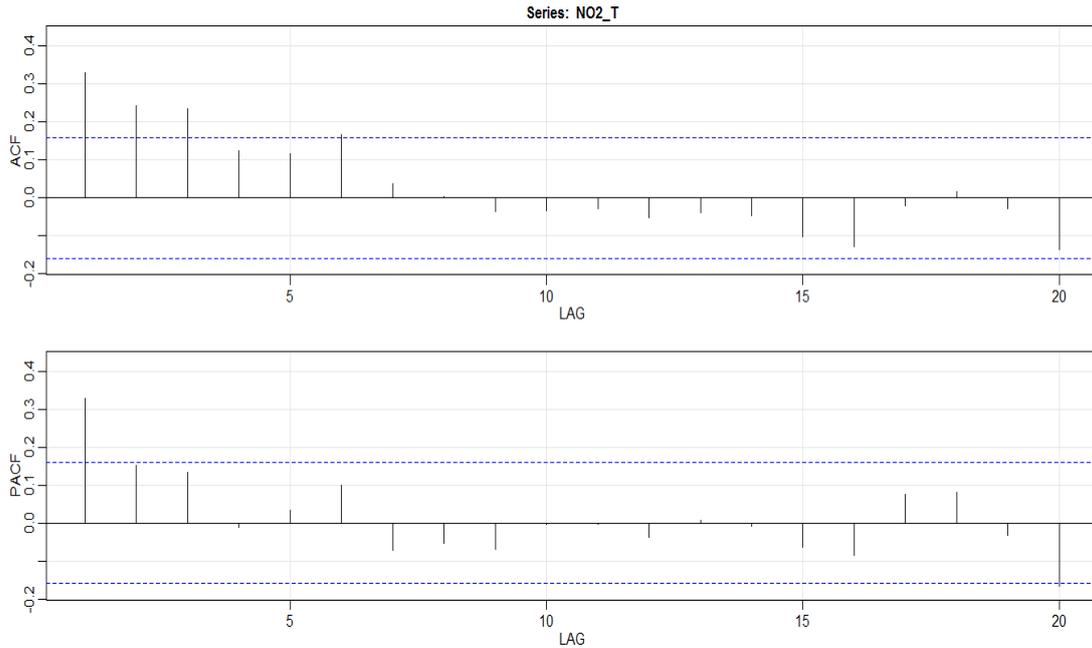
شكل (4) قيمة λ المقدرة ومن خلال تطبيق الخوارزمية

شكل (5) أختبار استقرارية التباين بتطبيق اختبار بوكس كوكس Box-Cox

الى حد كبير سلوك ذاتي الارتباط الذاتي والارتباط
الذاتي الجزئي لانموذج الانحدار الذاتي من الرتبة
الاولى . ولذلك فان الانموذج المرشح لتقدير بيانات
غاز ثنائي اوكسيد النيتروجين المحولة ، سيكون انموذج
الانحدار الذاتي من الرتبة الاولى

من خلال مقارنة السلوك النظري لذاتي
الارتباط الذاتي والارتباط الذاتي الجزئي مع نظيراتها
المقدرة نلاحظ ان دالة الارتباط الذاتي تتلاشى تدريجيا
تقريبا بعد فرق معين ، وان دالة الارتباط الذاتي الجزئي
تكون غير معنوية بعد الفرق الاول ، وهذا السلوك يشبه

ARIMA($p = 1, d = 0, q = 0$) ، وكما موضح في الشكل رقم (6)



شكل (6) قيم دالة الارتباط الذاتي ودالة الارتباط الذاتي الجزئي بعد الفرق الاول

المجموعة المناظرة لاقبل قيمة من قيم المعايير السابقة على الترتيب تعتبر مقدرات لرتب انموذج الانحدار الذاتي الاوساط المتحركة. ومن جدول (4) تبين ان افضل رتب بالاعتماد على المعايير الثلاثة وبالاجماع هو الانموذج المختلط من الرتبة الاولى للانحدار الذاتي والرتبة الاولى للاوساط المتحركة.

$$ARIMA(p = 1, d = 0, q = 1)$$

وبسبب كون تقديرات دالة الارتباط الذاتي والارتباط الذاتي الجزئي تعتمد على الملاحظة البصرية في تقدير رتب انموذج الارتباط الذاتي والايوساط المتحركة وهذا بجميع الاحوال تعتبر غير دقيقة ويمكن ان تمثل تقدير اولي لتلك الرتب، ولذلك سوف يتم تطبيق معايير اخرى مثل معيار معلومات اكيائي AIC ومعيار معلومات اكيائي المطور AICc ومعيار بيز المعلوماتي BIC لعدد من التراتيب الممكنة والقريبة من التقدرات السابقة و

جدول (4) تطبيق معايير المعلوماتية لاختبار رتب الانموذج

Model	p	q	AIC	AICc	BIC
1	0	1	-2.461236	-2.447499	-3.422303
2	1	0	-2.4888	-2.475062	-3.449867
3	1	1	-2.513493	-2.499078	-3.455094
4	2	0	-2.499571	-2.485156	-3.441171
5	0	2	-2.469814	-2.455399	-3.411414
6	1	2	-2.500769	-2.4855	-3.422903
7	2	1	-2.500767	-2.485497	-3.422901

وان تقدير حد الخطأ هو $\hat{\sigma}^2 = 0.02868$ ، بدرجة

حرية = 154

AIC=-2.513493 ، AICc=-2.499078

,BIC= -3.455094

ومن خلال قيم البواقي لسلسلة NO₂ الموضحة في ملحق رقم (3) رسم البواقي المعيارية والتي يجب ان تسلك سلوك المشاهدات المسحوبة من التوزيعات مستقلة ومتماثلة لكي تبين صحة الانموذج المقدر في وصف البيانات، نلاحظ التالي:

1- إن قيم البواقي المعيارية تتذبذب تقريباً حول المتوسط المساوي للصفر.

2- يظهر مخطط قيم دالة الارتباط الذاتي ان جميع القيم تقع ضمن الحد الاعلى والادنى للمشاهدات المستقلة التي تخضع للتوزيع الطبيعي.

وبالاعتماد على طريقة الامكان الاعظم لتقدير معلمات الانموذج السابق والبالغ عددها اربع معلمات (واحدة لانموذج الانحدار الذاتي وواحدة لانموذج الاوساط المتحركة والمتوسط العام وتباين حد الخطأ) وبالاعتماد على البيانات التي تم تحويلها بالخطوة السابقة، اذ تعتمد هذه الطريقة على طرائق الحل العددي للحصول على قيم المقدرات، وبعد ادراج فقط المقدرات ذات التأثير المعنوي كانت النتائج كما يلي:

$$x_t = -0.5458 + 0.7694x_{t-1} + \hat{w}_t - 0.5131\hat{w}_{t-1}$$

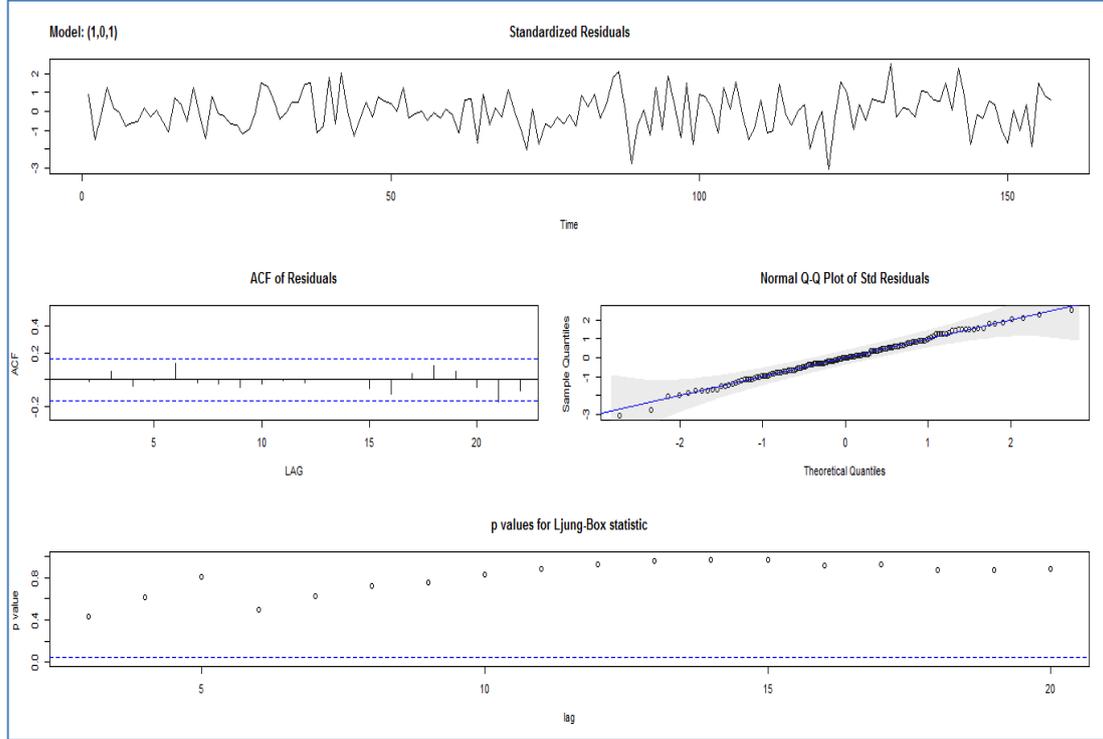
(0.0001) (0.0433) (0.00006)

إن القيم بين الاقواس تمثل قيم الاحتمالية المرافقة للاختبار المعنوية للمعلمات (P_value) ومنه نلاحظ ان جميعها معنوية .

3-معظم البواقي المعيارية كانت ضمن الخط المائل لمخطط Q-Q-plot ، وهذا يؤشر خضوعها للتوزيع الطبيعي . ما عدا المشاهدات في الاطراف والتي سببها وجود بعض القيم الشاذة في البيانات.

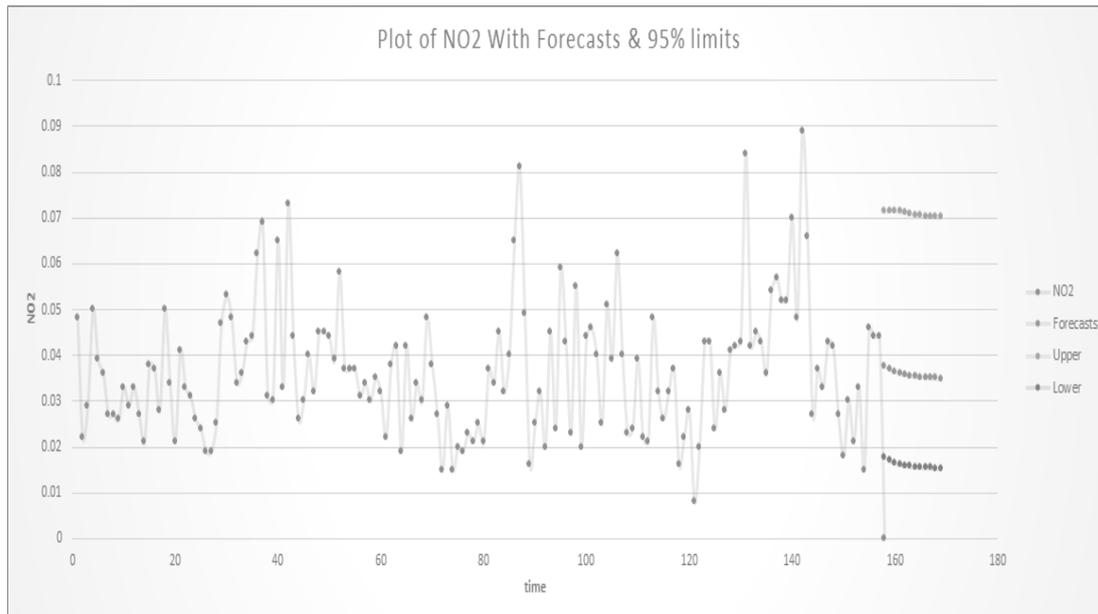
4-جميع قيم المعنوية لاختبار Ljung-Box كانت غير معنوية (اكبر من 0.05) ولجميع الفروق المستخدمة ، اي اننا نقبل فرضية العدم لذلك الاختبار التي تنص بكفاية الانموذج احصائياً في وصف البيانات. وكما في الشكل (7).

3-معظم البواقي المعيارية كانت ضمن الخط المائل لمخطط Q-Q-plot ، وهذا يؤشر خضوعها للتوزيع الطبيعي . ما عدا المشاهدات في الاطراف والتي سببها وجود بعض القيم الشاذة في البيانات.



شكل (7) جميع قيم المعنوية لاختبار Ljung-Box

وبذلك تكون القيم التنبؤية لمدة 12 اسبوع مع حدود الثقة لها بثقة 95% موضحة كما في الشكل (8) والجدول (5)



شكل (8) بيانات القيم التنبؤية لمدة 12 اسبوع مع حدود الثقة لها بثقة 95%

الجدول (5) القيم التنبؤية لمدة 12 اسبوع مع حدود الثقة لها بثقة 95%

Forecasts	Standard Error	Upper	Lower
0.037602	0.1693386	0.0713306	0.017819744
0.03693	0.1748096	0.071594	0.016989391
0.03642	0.1779689	0.0715323	0.016455467
0.036031	0.1798128	0.0713373	0.016099945
0.035735	0.1808955	0.0711031	0.015856681
0.035508	0.1815333	0.0708746	0.015686618
0.035334	0.1819098	0.0706707	0.015565505
0.0352	0.1821323	0.0704973	0.015477956
0.035098	0.1822639	0.0703543	0.015413886
0.035019	0.1823417	0.0702387	0.015366551
0.034959	0.1823878	0.0701464	0.01533124
0.034912	0.182415	0.0700733	0.01530471

نتائج التنبؤ Forecast Results

أستخدمت مقاييس خطأ الأتية لتقييم أداء التنبؤ المستخدم في هذا البحث :

MODEL	MSR	RMSR
ARIMA	0.02868	0.169352

المبحث الرابع: الاستنتاجات والتوصيات

سيتم مناقشة أهم الاستنتاجات والتوصيات التي تم التوصل إليها من خلال اجراء التجارب العملية.

الاستنتاجات

1. إن السلسلة الزمنية للملوث NO2 لديها سلوك مستقراً نوعاً ما وياتجاه عام ثابت حول وسطها الحسابي أي إنها مستقرة في المتوسط .

2. بيانات السلسلة الزمنية للملوث NO2 كانت تحتوي على قيم مفقودة عدم تسجيل القراءات في ايام العطل الرسمية مما استدعى معالجتها قبل التحليل .

3. أظهرت السلسلة عدم الاستقرار في التباين مما استدعى الى تحويلها الى سلسلة مستقرة في التباين باستخدام تحويل Box-cox .

التوصيات

1. توصي الدراسة بضرورة بناء قاعدة بيانات مركزية ومتكاملة (تسجيل ايام العطل) لتسجيل بيانات الملوثات في وزارة الصحة والبيئة بغية الحصول على تقارير بيانات وجداول تنسجم وحاجة الباحثين.

2. ضرورة توفير بيانات الملوثات في وزارة الصحة والبيئة بشكل مجاني لغرض تشجيع الباحثين لدراسة تلك السلاسل وبناء النماذج الرياضية لها .

3. تطبيق تقنيات هجينة بين نماذج تنقيب البيانات والنماذج الاحصائية مثل تقنيات هجينة تدمج ما بين الشبكات العصبية وأنموذج ARMA مثال (NARIMA) والمقارنة بين النتائج المستخدمه بالبحث.

4. تطبيق النماذج المستخدمة في هذا البحث على بيانات الملوثات في محافظات اخرى من بلدنا.

المصادر

1. شومان، عبد اللطيف حسن ، " مقدمة في الاحصاء التطبيقي " ، الطبعة الاولى - دائرة المكتبة الوطنية، المملكة الاردنية الهاشمية ، رقم الايداع 2008/1/7 ، أخرج وتدقيق دار الجنان .

2. وهيب، مؤيد سلطان (2013) ، " بناء نموذج ARIMA للتنبؤ بحجم البطالة في مصر " ، مجلة جامعة كركوك للعلوم الادارية والاقتصادية /المجلد 3 / العدد 1 .

13. Box, George EP, et al. "Time series analysis: forecasting and control". John Wiley & Sons, 2015 .
14. Box, George EP, and Gwilym M. Jenkins. "Time series analysis: forecasting and control" , revised ed. Holden-Day, 1976 .
15. Shumway, Robert H., and David S. Stoffer. "Time series analysis and its applications: with R examples" . Springer Science & Business Media, 2010 .
16. Suling Zhu , Xiuyuan Lian , Haixia Liu , Jianming Hu , Yuanyuan Wang , Jinxing Che ." Daily air quality index forecasting with hybrid models: A case in China" . Elsevier , Environmental Pollution , Volume 231, Part 2, December 2017, Pages 1232-1244
17. Akaike, Hirotugu. "Akaike's Information Criterion." International Encyclopedia of Statistical Science. Springer Berlin Heidelberg, 2011. 25-25.
18. Volinsky, Chris T., and Adrian E. Raftery. "Bayesian information criterion for censored survival models." Biometrics 56.1 (2000): 256-262.
19. Dickey, David A., David P. Hasza, and Wayne A. Fuller. "Testing for unit roots in seasonal time series." Journal of the American Statistical Association 79.386 (1984): 355-367.
20. Edlund, Per-Olov. "Identification of the multi-input box-Jenkins transfer function model." Journal of Forecasting 3.3 (1984): 297-308.
21. Ujjwal Kumar & V. K. Jain ." ARIMA forecasting of ambient air pollutants (O3, NO, NO2 and CO)" springer , Stochastic Environmental Research and Risk Assessment , July 2010, Volume 24, Issue 5, pp 751–760 .
3. R.TSAY. (2005) ,"Analysis of Financial Time Series Financial Econometrics" .John Wiley&Sons, Inc, USA .
4. Anderson ,O.D.(1976) ,"Time series analysis and Forecasting " Butter Werlths London and Boston .
5. Wei,W.S.(2006),"Time series Analysis Univariate and Multivariate Methodes ", Addison- Wesly publishing company INC,USA .
6. Box, G. E. P, Jenkins, G. M., and Reinsel, G. C. (2008), "Time Series Analysis: Forecasting and Control" ,4rd Ed., Holden-Day, San Francisco .
7. Brockwell, P. J and Davis, R. A. (1996) , "Introduction to Time Series and Forecasting" , Springer-Verlag , New York .
8. Snezhana Georgieva & Atanas Valev& Desislava& Doychin Todorov." Time series analysis and forecasting for air pollution in small urban area: an SARIMA and factor analysis approach" springer, Stochastic Environmental Research and Risk Assessment , May 2014, Volume 28, Issue 4, pp 1045–1060
9. Jehng-Jung Kao & Shang-Shuang Huang (2000) "Forecasts Using Neural Network versus Box-Jenkins Methodology for Ambient Air Quality Monitoring Data", Elsevier, Journal of the Air & Waste Management Association, 50:2, 219-226
10. MingCai& Yafeng Yin& Min Xie. (2008)" Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach" Elsevier, Transportation Research Part D: Transport and Environment , Volume 14, Issue 1, January 2009, Pages 32-41 .
11. Montgomery,Douglas C., Lynwood A. Johnson, and John S. Gardiner."Forecasting and time series analysis". McGraw-Hill Companies,1990.
12. Montgomery,Douglas C., Cheryl L. Jennings, and Murat Kulahci. "Introduction to time series analysis and forecasting". John Wiley& Sons, 2015 .

الملاحق

لاطلاع على جداول البيانات والكود المستخدم بلغة R

المستخدمة في هذا البحث يرجى زيارة الرابط التالي :

<https://drive.google.com/file/d/1wMGnaFHOBCNAOuHIkUd2Pv55uMioF-oy/view?usp=sharing>