

استخدام السلاسل الزمنية لمخرجات التحليل العنقودي مع التطبيق العملي

أ.م. وكاع علي هدية
كلية الإدارة والاقتصاد
جامعة كركوك

wakaa2017@uokirkuk.edu.iq

الباحثة: سمية علي حسين
كلية الإدارة والاقتصاد
جامعة كركوك

sumeya.ali@uokirkuk.edu.iq

المستخلص:

يعد التحليل العنقودي في بيانات السلاسل الزمنية من الموضوعات المهمة في تحليل البيانات وإيجاد اتجاهات متشابهة في السلاسل الزمنية الذي يمثل تحدياً كبيراً في مختلف المجالات. ازداد الاهتمام بدراسة عنقدة بيانات السلاسل الزمنية من قبل العلماء حيث أثبتت فعاليتها في توفير معلومات مهمة في مختلف المجالات استهدف البحث بصورة أساسية عملية العنقدة لبيانات السلاسل الزمنية لتسهيل التنبؤ للعناقيد المتكونة واستغلال الوقت والجهد، تم استخدام عنقدة بيانات السلاسل الزمنية في مجالات علمية متنوعة لاكتشاف الأنماط التي تمكن محلي البيانات من استخراج معلومات قيمة من مجموعة بيانات معقدة وضخمة ويتم تجميع (عنقدة) المجموعات المتجانسة معاً على أساس مقياس تشابه معين. واستخدمت البيانات الشهرية لإنتاجية الطاقة الكهربائية في كركوك لدراسة السلوكية الزمنية لها، تم استخدام طريقة العنقدة الهرمية Hierarchical Clustering Method والطريقة المعتمدة في طريقة الربط هي طريقة الربط الهرمية طريقة وورد (ward's method) بالاعتماد على مصفوفة التشابه واعتمدنا على مقياس مسافة القطاع مسافة مناهاتن City-block (manhattan) distance لإيجاد مصفوفة التشابه بين العناقيد وبغية الوصول إلى مجموعات (عناقيد) متجانسة التي تجمعها صفات مشتركة بالاعتماد على انتاجيتها، تستخدم العنقدة الهرمية التجميعية ورسم المخطط الشجري والتنبؤ بالقيم المستقبلية لإنتاجية العناقيد.

إن أهم النتائج التي توصل إليها البحث هي تكوين أربعة عناقيد وبناء نموذج سلاسل زمنية لكل عنقود، ومن خلال تحليل السلسلة تبين أنها غير مستقرة وغير عشوائية ولغرض تحقيق الاستقرار والعشوائية تم إجراء التحويلات اللازمة، واستخدام معايير المفاضلة (Schwartz Bayesian Criteria: SBIC) (Akaike, Information Criteria: AIC) (Root Mean Square Error: RMSE) (Hanna-Quinn Criterion: H-Q) لتشخيص النماذج المعنوية لاختيار النموذج الملائم والكفؤ، التنبؤ للعنقود الأول بالنموذج $ARIMA(0,1,1)$ ، النموذج الذي تم استخدامه للعنقود الثاني هو $SARIMA(2,0,0) \times (1,1,2)^{12}$ ، العنقود الثالث بسبب توقف الوحدات عن الإنتاج منذ سنة 2014 ولحد الآن الوحدات معطلة أي أنها لا تنتج الكهرباء أساساً لترك أعمال التأهيل فأن القيم المستقبلية المتنبئ بها هنا تكون صفر، العنقود الرابع تم استخدام النموذج $ARIMA(2,1,0)$ وكانت التنبؤات جيدة وقريبة من الواقع للفترة من تشرين الثاني 2020 لغاية تشرين الثاني 2022 لمدة سنتين.

الكلمات المفتاحية: عنقدة السلاسل الزمنية، التنبؤ، $ARIMA$ ، سلاسل زمنية، مسافة القطاع (مناهاتن).

Clustering of Time Series Data With Application

Researcher: Sumeya Ali Hussien
College of Administration and Economics
Kirkuk University

Assist. Prof. Wakae Ali Hadba
College of Administration and Economics
Kirkuk University

Abstract:

Cluster analysis in time series data is one of the important topics in analyzing data and finding similar trends in time series, which represents a major challenge in various fields. Scientists have increased interest in studying time series data clustering, as it has proven its effectiveness in providing important information in various fields. Clustering time-series data to facilitate prediction of formed clusters and exploit time and effort. Clustering time-series data has been used in various scientific fields to discover patterns that enable data analysts to extract valuable information from a large and complex data set. Homogeneous clusters are grouped together on the basis of a specific similarity measure. The monthly data of electrical energy productivity in Kirkuk was used to study its temporal behavior. The hierarchical clustering method was used, and the method adopted in the linkage method is the hierarchical linking method, the ward's method, based on the similarity matrix, and we adopted the city-block distance scale (manhattan) distance to find the similarity matrix between clusters and in order to reach homogeneous groups (clusters) that have common characteristics based on their productivity, hierarchical clustering, tree diagramming and prediction of future values of cluster productivity are used.

The most important findings of the research are the formation of four clusters and the construction of a time series model for each cluster, and through the analysis of the series, it was found that it is unstable and not random, and for the purpose of achieving stability and randomness, the necessary conversions were made, and the differentiation criteria (Akaik, Information Criteria: AIC) were used. (Schwartz Bayesian Criteria: SBIC), (Hanna-Quinn Criterion: H-Q), (Root Mean Square Error RMSE). To diagnose the significant models to choose the appropriate and efficient model, the prediction for the first cluster is in the ARIMA(0,1,1) model, the model that was used for the second cluster is SARIMA(2,0,0)x(1,1,2)¹², the third cluster due to stopping the units About production since 2014 and until now the units have been idle, meaning that they do not produce electricity as a basis for leaving the rehabilitation works, the future values predicted here are zero, the fourth cluster was used ARIMA (2,1,0) model and the predictions were good and close to reality for the period from November 2020 to November 2022 for a period of two years.

Keywords: time series clustering, Forecasting, ARIMA, time series, City-block (manhattan) distance.

المقدمة

الطاقة الكهربائية في العراق تعتبر قليلة جدا والإنتاجية لا تسد الحاجة الاستهلاكية للقطاعات المنزلية والصناعية وهذه مشكلة كبيرة جدا ويجب التطرق لها ومناقشتها ومحاولة الوصول الى حلول وفي هذه الدراسة نحاول مساعدة الجهات المعنية لوضع حلول لهذه المشكلة لما

لهذه الطاقة من أهمية في حياتنا المعاصرة وتعتمد عليها جميع النشاطات، يعد تحديد مجموعات (عناقيد) السلاسل الزمنية أمراً صعباً للغاية بسبب صعوبة تحديد التشابه للسلاسل الزمنية المختلفة والتي يمكن تحجيمها وتحديد شكل مختلف على كل من الأبعاد الزمنية والسلوكية لذلك يعد مفهوم التشابه مفهوماً مهماً جداً لعنقدة بيانات السلاسل الزمنية، تعتبر السلاسل الزمنية من الأدوات الإحصائية المهمة لوضع الخطط والبرامج المرتبطة بدراسة ظاهرة معينة على مدى فترة من الزمن، يحتوي عنقدة السلاسل الزمنية على العديد من التطبيقات في مجالات مختلفة منها الأسواق المالية والبيانات الطبية وتطبيقات علوم الأرض مثل درجات الحرارة أو الضغط.

في هذا البحث حاول الباحث أن يشخص الطريقة الأمثل لكيفية التنبؤ بإنتاجية الطاقة الكهربائية المجهزة وذلك في محافظة كركوك/محطة كهرباء الشهيد المهندس عبدالله حمزه الغازية للفترة من (2010-2020) وذلك باستخدام طريقة التحليل العنقودي للسلاسل الزمنية وبعدها نستخدم نماذج بوكس-جينكينز (Box & Jenkins) للتنبؤ لكل عنقود.

لقد اثبتت الدراسات ان التحليل العنقودي للسلاسل الزمنية له فعالية في توفير معلومات مفيدة في مختلف المجالات ويبدو ان هناك زيادة في الاهتمام بتجميع السلاسل الزمنية وكجزء من الجهد المبذول في ابحاث التنقيب عن البيانات الزمنية، يهدف موضوع عنقدة السلاسل الزمنية الى تكوين مجموعات ذات تشابه كبير داخل العنقود وتشابه منخفض بين العناقيد وعلى وجه التحديد يجب ان تظهر المتغيرات التي تنتمي الى نفس العنقود تشابه كبيراً مع بعضها البعض، بينما يجب ان تظهر المتغيرات التي تنتمي الى مجموعات (عناقيد) مختلفة تشابه اقل وبالتالي مسافة عالية مع بعضها البعض.

قدم Pjotr Roelfsen دراسة بعنوان: (Time Series clustering, 2018)، تقوم هذه الدراسة بمقارنة بين الطرق المختلفة لعنقدة السلاسل الزمنية التي تم الحصول عليها من احدى الشركات للمبيعات لعميل معين، طبق خوارزميات التجميع الجزئي والتجميع الهرمي وتم المقارنة بينهم على اساس الوقت المستغرق لحسابها وايضا لمصفوفات المسافة المتكونة واستخدام نماذج (ARIMA) على العناقيد الثمانية المتكونة لمقارنة جودة التجميعات، قام باختبار اداء أفضل خوارزمية تجميع بالمقارنة بين ادائهم واستنتج ان أفضل طريقة للعنقدة هي:

Clustering Large Applications based on RANdomized Search: CLARANS clustering.

خوارزمية العنقدة وقانون المسافة الاقليدية (Euclidean distance) أعطت أفضل نتيجة وتنبؤ باستخدام نماذج ARIMA، والهدف الواضح هو استخدام العنقدة للحصول على العناقيد التي يمكن ان تستخدم للتنبؤ بدلا من عمل تنبؤات لجميع السلاسل الزمنية الفردية وعمل توقع واحد فقط لكل عنقود.

قدمت هناء طه عطالله الجوهرى اطروحة بعنوان (أسلوب احصائي مقترح لمعالجة بيانات السلاسل الزمنية العنقودية متعددة المتغيرات دراسة تطبيقية) في عام (2016)، اخذت سلاسل زمنية متعددة المتغيرات لمجموعة من مرضى الفشل الكلوي المزمن وتم معالجة القيم المفقودة باستخدام طريقة سلاسل ماركوف جين لمونتي كارلو (Markov chain Monte carlo)، وتم اقتراح أسلوب (Kernel multi-way principal component analysis) مع استخدام دالة كاوزين (Radial basis function) في عنقدة بيانات السلاسل الزمنية متعددة المتغيرات وقامت

بالمقارنة بين العناقيد قبل وبعد تطبيق الأسلوب المقترح وتوصلت الى انه تزداد جودة العناقيد بعد تطبيق الأسلوب المقترح.

مشكلة البحث: تظهر مشكلة البحث في ان الطاقة الكهربائية في العراق عموماً تعاني نقص كبير في الإنتاجية عبر تقدم الزمن مع التزايد المستمر على طلب الطاقة الكهربائية مما يستدعي الامر بالبحث والتقصي عن هذه المشكلة، سنحاول من خلال هذه الدراسة الإجابة على الإشكالية الآتية:

أ. كيف يمكن تطبيق طريقة التحليل العنقودي الهرمي على السلاسل الزمنية لوحداث الإنتاج وتكوين عناقيد (مجموعات) تضم وحدات الإنتاج المتشابهة من ناحية الإنتاج؟

ب. ما مدى دقة النماذج المقدرة في التنبؤ لكل عنقود مكون لإنتاجية الطاقة الكهربائية.

ج. هل النتائج التي حصلنا عليها وفق هذه الطريقة مجدية وهي التخلص من مشكلة وقت التنبؤ في حالة كون التنبؤ لكل وحدة انتاج الذي يستغرق وقت وجهد كبير لمحلل البيانات.

هدف البحث: يهدف البحث الى دراسة التحليل العنقودي للسلاسل الزمنية لإنتاجية الطاقة الكهربائية في محافظة كركوك/محطة كهرباء الشهيد المهندس عبدالله حمزة الغازية للفترة (2010-2020) وذلك بتطبيق التحليل العنقودي على السلاسل الزمنية لوحداث الإنتاج وعددها سبعة عشر (17) أي تصنيف وحدات انتاج الطاقة الى مجاميع (عناقيد) كل مجموعة تحتوي على عدد من الوحدات المتقاربة من ناحية الإنتاج، واستخدام منهجية Box-Jenkins على كل عنقود مكون اختيار نموذج يتنبأ بالإنتاجية للفترة الزمنية المستقبلية باستعمال بعض طرائق السلاسل الزمنية في التنبؤ اعتماداً على عينة البحث وإيجاد افضل نموذج عنقود مكون من عنقدة السلاسل الزمنية، ومقارنة النتائج في كل من العناقيد بالاعتماد على شكل الهيكل الشجري.

1. الجانب النظري:

1-1. عنقدة السلاسل الزمنية: ان عنقدة السلاسل الزمنية هي نوع خاص من التحليل العنقودي وتعرف سلسلة من العناصر المستمرة ذات القيمة الحقيقية باسم السلاسل الزمنية ويتم تصنيف السلاسل الزمنية بشكل رئيسي بأنها بيانات ديناميكية لان قيم ميزاتها تتغير كدالة مع الزمن وان السلاسل الزمنية هي بيانات بطبيعة الحال ذات حجم كبير جداً، وتتم عملية (تجميع) عنقدة بيانات السلاسل الزمنية اختصاراً للوقت ولاكتشاف انماط مثيرة لاهتمام في مجموعات بيانات السلاسل الزمنية وبذلك تتكون لدينا عناقيد تحتوي السلسلتين الزمنية المتشابهتين في عنقود واحد والتي تكون مختلفة عن باقي العناقيد وتتم العنقدة التجميعية بهذا الشكل الى ان يتكون لدينا عنقود واحد يحتوي جميع السلاسل الزمنية وهذه الطريقة العنقدة التجميعية عكس العنقدة الانقسامية (Aghabozorgi et al., 2015).

تأتي أهمية دراسة عنقدة السلاسل الزمنية لأسباب نذكر منها: (Liao, 2005)

- أ. بيانات السلاسل الزمنية تحتوي على معلومات مهمة يمكننا ان نحصل عليها من خلال اكتشاف الانماط ولذلك نستخدم العنقدة لأنها الحل الشائع الذي من خلاله نكشف الانماط في مجموعات السلاسل الزمنية.
- ب. أن بيانات السلاسل الزمنية كبيرة جداً ومن الصعب التعامل معها بشكل جيد من قبل الباحثين ولذلك يفضل الباحثين التعامل مع البيانات كمجموعات عناقيد منتظمة متشابهة الى حد كبير فيما بينها بدلاً من بيانات كبيرة جداً، وبذلك يتم تمثيل البيانات في مجموعات غير متداخلة او عن طريق التسلسل الهرمي مثلاً.

ج. يعتبر مفهوم عنقدة السلاسل الزمنية هو النهج الأكثر استخداماً باعتباره تقنية استكشافية لاستخراج البيانات المعقدة.

د. إن عملية العنقدة للسلاسل الزمنية مفيدة جداً في إعطاء تصور وفهم سريع لبنية البيانات والعناقيد والانتظامات الأخرى في البيانات.

فأذن نعرف الآن أن عملية فصل المجموعات تبعا لأوجه التشابه بينهم (بالعنقدة) وهناك مبدآن رئيسيان هنا: (Ergüner Özkoç, 2021)

❖ التشابه داخل العنقود هو الأعلى.

❖ التشابه بين العناقيد هو الأقل.

تنشأ مشكلة (تجميع) عنقدة السلاسل الزمنية عندما نلاحظ عينة من السلاسل الزمنية ونريد تجميعها في عناقيد أو مجموعات مختلفة هذه هي المشكلة المركزية في العديد من مجالات التطبيق وبالتالي فإن عملية عنقدة السلاسل الزمنية هو في الوقت الحاضر مجال بحث نشط في الاختصاصات المختلفة بما في ذلك التمويل والاقتصاد والطب والهندسة وعلم الزلازل والارصاد الجوية ومجالات أخرى كثيرة (Andrés M. Alonso, Francisco J. Nogales, 2020)، (Ergüner Özkoç, 2021)

2-1. قياس المسافة Distance Measures: عملية العنقدة هي العثور على انماط مخفية أو مجموعات متماثلة في البيانات وبذلك نعرف ونحدد فيما إذا كانت المشاهدات في أي من السلاسل الزمنية متشابهة من أجل عنقدة السلاسل الزمنية تكون طرق قياس التشابه مهمة وشديدة الأهمية إذ أنه يتم تحديد المسافة بين سلسلتين زمنيتين في جميع النقاط الزمنية (Li, 2017) (Roelofsen, 2015).

هناك ثلاث أنواع أساسية من طرق قياس التشابه:

❖ مقياس التشابه القائم على الشكل وهو أنه العنقود يحتوي على متسلسلة زمنية متماثلة في الوقت والشكل، مثل الاقليدية والتزييف الديناميكي للوقت (DTW) Dynamic Time Warping.

❖ ومقياس التشابه القائم على النموذج مناسب للسلسلة الزمنية القصيرة والطويلة منها معامل ارتباط بيرسون.

❖ مقياس التشابه القائم على الميزة وهذا يكون مناسب للسلسلة الزمنية الطويلة.

في بداية الأمر يجب أن نحدد مقياس التشابه لتحديد هذه المسافة سوف نبين في هذا القسم أكثر المقاييس شيوعاً واستخداماً توجد عدة طرق لقياس المسافة بين عناصر مصفوفة القرابة (الجوهري، 2016).

الجدول (1): مقاييس التشابه في التحليل العنقودي

المقاييس Measure	صيغة المعادلة
Euclidean distance	$D(Y_i, Y_j) = [\sum_{j=1}^p (Y_i - Y_j)^2]$
City-block (manhattan) distance	$D(Y_i, Y_j) = [\sum_{j=1}^p Y_i - Y_j]$
Mikowski distance	$D(Y_i, Y_j) = [\sum_{j=1}^p Y_i - Y_j ^r]^{\frac{1}{r}}$
Dynamic time warping distance	$D = \min_k \frac{\sum_{k=1}^k W_k}{k}$

3-1. خوارزميات عنقدة السلاسل الزمنية Time-series Clustering Algorithms: في هذا الجزء ان الاعمال الحالية المتعلقة بتجميع السلاسل الزمنية تتركز على بيانات السلاسل الزمنية ومناقشتها ان البعض يستخدمون السلاسل الزمنية الخام والبعض الاخر يحاول استخدام طرق الاختزال قبل عملية تجميع السلاسل الزمنية، اما بالنسبة الى عملية عنقدة بيانات السلاسل الزمنية فتم تقسيمها الى ثلاث طرق وهي: (الجوهري، 2016)، (Li, 2017) (Kavitha & Punithavalli, 2010) (Aghabozorgi et al., 2015)

أ. اعتمادا على البيانات الخام (Row-data-based).

ب. اعتمادا على الصفات (Feature based).

ج. اعتمادا على النموذج (Model based).

في الطريقة الاولى نتعامل مع البيانات للسلاسل الزمنية دون اي تحويلات او دمج للبيانات وهذه الطريقة هي التي سوف نستخدمها في هذا البحث حيث تم اخذ السلاسل الزمنية لوحداث الانتاج بدون اي تغيير، المقارنة بين السلسلتين الزمنيتين في نفس الفترة الزمنية تكون اطوالها متساوية او قد تكون غير متساوية وتتخذ السلاسل الزمنية المتقاربة بشكل كاف لكل النقاط الزمنية وتعتبر في نفس المجموعة، نراجع التقنيات الأكثر استخداما لتجميع (عنقدة) السلاسل الزمنية المستندة الى الشكل الاسلوبي الأكثر شيوعا هما (k-means clustering) (hierarchical clustering) الطريقة الهرمية والطريقة الغير الهرمية، وفي الطريقة الثانية يستند على ملامح البيانات الخام وصفاتها، والطريقة الثالثة تعتمد العنقدة على معالم النموذج ويسمى احيانا بالأسلوب الاساسي للنماذج المعدلة (Bagnall & Janacek, 2005). هنالك اهداف رئيسة من عملية التعنقد للسلاسل الزمنية:

الهدف الاول: التشابه في الزمن similarity in time: هنا يكمن الهدف وهو تجميع (عنقدة) سلسلتين للتشابه في كل خطوه بالزمن، اي انه يكون التغير بين السلسلتين الزمنيتين يحدث بشكل متشابه جدا في كل خطوه زمنية (الجوهري، 2016).

الهدف الثاني: التشابه في الشكل Similarity in shape: تكون البيانات للسلاسل الزمنية لها خصائص متشابهة في الشكل اي انه هدفنا نجد سلاسل زمنية متشابهة في الخصائص وتتم العنقدة على هذا الأساس (الجوهري، 2016).

الهدف الثالث: التشابه في التغير Similarity in change: يكون الهدف من عملية التعنقد (التجميع) للسلاسل الزمنية هو انه يتم تجميع السلاسل المتشابهة في التغير من وقت لآخر (الجوهري، 2016).

ويمكن ان تكون العنقدة بشكل عام مصنفة الى اقسام:

أولاً. طريقة العنقدة الغير الهرمية Non-hierarchical clustering method: يوجد العديد من طرق التعنقد الغير الهرمية ومنها طريقة المتوسطات (k-means) وفكرته الأساسية تكمن في البحث حول k من المتوسطات تعتمد على فكرة التقسيم ولا يعتمد على بناء الشكل الهرمي وهي أكثر الطرق انتشارا، هنا تنقسم البيانات الى عناقيد كل منها لها متوسط خاص بها (أحمد، 2015).

ثانياً. طريقة العنقدة الهرمية Hierarchical clustering of time-series: تعمل هذه الطريقة التعنقد الهرمي عن طريق تجميع البيانات (السلاسل الزمنية) في شجرة من العناقيد والعنقدة الهرمية نوعية تكتلية تجميعية او انقسامية، اعتمادا ما اذا كان من الاسفل الى الاعلى أو من

الاعلى الى الاسفل حسب الاستراتيجية المتبعة، أن طريقة العنقدة الهرمية التكتلية أكثر استخداماً من الطريقة الانقسامية، حيث تكون طريقة التعنقد الهرمية التكتلية (Agglomerative hierarchical clustering) يبدأ بوضع كل عنقود في مجموعة ثم تدمج هذه المجموعات في مجموعات أكبر وأكبر حتى يتكون لدينا شجرة تحتوي على جميع العناقيد يتكون عنقود واحد، في كل خطوة من النهج الهرمي التكتلي أقرب مجموعتين يتم دمجهما في مجموعة جديدة واحدة وبالتالي فإن العملية لا رجعة فيها بالمعنى لا يمكن فصل أي عنصرين تم تجميعهما معا في مجموعة في وقت لاحق من الاجراء لا يمكن تصحيح أي أخطاء مبكرة (Bagnall & Janacek, 2005) (Liao, 2005) (Aghabozorgi et al., 2015).

بالمثل في الطريقة الهرمية الانقسامية (Divisive Hierarchical clustering) تكون من الأعلى الى الأسفل من خلال اعتبار مجموعة البيانات بأكملها كمجموعة (عنقود) واحد ويستمر بتقسيمها بشكل متكرر الى مجموعات أصغر الى ان تصل الى حالة ان يكون كل متغير بشكل منفرد، العنقدة الهرمية الانقسامية هي عملية عكسية للعنقدة الهرمية التجميعية، وتعتبر العنقدة الهرمية (hierarchical clustering) التي تتمتع بقدرات تصويرية أفضل من العنقدة الغير الهرمية (k-means clustering) حيث تشكل عملية العنقدة المخطط الشجري (نامق، 2010). إن العنقدة الهرمية لا تقتصر على السلاسل الزمنية العنقودية ذات الطول المتساوي وهو ينطبق على سلسلة من طول غير متكافئ إذا كان مقياس المسافة مرن مثل (DTW) تزييف الوقت الديناميكي.

هنا في هذه الطريقة على عكس معظم الخوارزميات في التسلسل الهرمي لا يتطلب تحديد عدد مجموعات (العناقيد) اي انها تعتبر نقطة قوة في مجموعات السلاسل الزمنية لأنه عادة يكون من الصعب تحديد عدد العناقيد (Rodrigues et al., 2008).

تنظيم هذه المجموعات من العناقيد المتداخلة كشجرة ويمكن ان نصورها بواسطة مخطط شجرة الاسنان ويكون المشاهدات في المحور السيني، والمحور الصادي يمثل المسافة بين المجموعات المختلفة (Rodrigues et al., 2008) (Roelofsen, 2015).

بعد عملية التعنقد لوحداث انتاج الطاقة الكهربائية البالغ عددها سبعة عشر وحدة (17) وتكون العناقيد الأربعة التي تضم وحدات الإنتاج المتشابهة بعدها نستخدم التنبؤ لكل عنقود متكون الذي يحوي السلاسل الزمنية المتشابهة وتكوين نماذج بوكس جينكيز (Box-Jenkins Model) لكل عنقود وسيتم توضيح مراحلها.

بمجرد تجميع السلاسل الزمنية والحصول على المجموعات (العناقيد) يمكننا القيام بالعديد من الإجراءات:

❖ وصف العناقيد (المجموعات) المتكونة.

❖ ننتبأ للعناقيد المتكونة.

4-1. نماذج السلاسل الزمنية (Time series Models): يعد الاسلوب الذي قدمه الباحثان بوكس-جينكيز في العام (1970) لتحليل السلاسل الزمنية من الادوات الاحصائية المهمة التي يمكن الاعتماد عليها في التحليل والتنبؤ للسلاسل الزمنية، تتطلب معظم طرق تحديد العلاقة الإحصائية بين المتغيرات ان تكون الأخطاء في التنبؤ غير مرتبطة ببعضها، ويجب أولاً تحديد هل السلسلة الزمنية مستقرة او لا نستخدم طريقة اختبارات الاستقرار (السكون) Time Series Stationary Tests هي دالة الارتباط

الذاتي (Autocorrelation Function: ACF) ودالة الارتباط الذاتي الجزئي (Partial Autocorrelation Function: PACF). (Newbold, 1983) (Nelson, 1998). (Metcalfe & Cowpertwait, 2009) (Karakaş, 2019) (N. M. Eze, 2020).

ان منهجية هذا الاسلوب تعتمد في صياغتها على النماذج الخطية الثلاثة الشائعة الاتية:

أ. **نموذج الانحدار الذاتي من الرتبة (p) Autoregressive Model (AR)**: في هذا النموذج قيمة المتغير في الفترة الحالية (Y_t) تعتمد على قيمته في الفترات السابقة (Y_{t-1}) نموذج الانحدار الذاتي هو عبارة عن انحدار خطي لقيم السلسلة الزمنية (متغير تابع) مع واحد او اكثر من القيم السابقة للسلسلة الزمنية كمتغيرات غير معتمدة (متغيرات مستقلة) ويرمز له بالرمز $AR(p)$ (Karakaş, 2019) ويمكن كتابة النموذج بالصيغة الاتية:

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t \quad \dots (1)$$

(μ): الحد الثابت ($-\infty < \mu < \infty$) (p): تمثل رتبة النموذج، (Y_t) يمثل مشاهدات السلسلة

الزمنية في الزمن t ، (ϕ): معاملات النموذج، ($-1 < \phi_i < 1$)، (a_t): الخطأ العشوائي (التشويش

الابيض) (white Noise) ويتوزع توزيعاً طبيعياً ($a_t \sim WN(0, \sigma^2)$)

ب. **نموذج الاوساط المتحركة من الرتبة (q) (Moving Average Model)(MA)**: يمكن تمثيل السلسلة الزمنية بواسطة نموذج المتوسطات المتحركة ويعبر عن القيمة الحالية للسلسلة (Y_t) بدلالة المجموع الموزون للقيم السابقة للأخطاء اي انه يعتمد على الاخطاء السابقة لتمثيل السلسلة الزمنية (Karakaş, 2019) ويرمز له بالرمز $MA(q)$ ويكتب بالصيغة الاتية:

$$Y_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad \dots (2)$$

θ_i : معالم النموذج ($-1 < \theta < 1$)، a_t : حد الخطأ العشوائي، q : رتبة الأوساط المتحركة.

ج. **النموذج المختلط للانحدار الذاتي والوساط المتحركة ARMA(p,q)**: نجد احيانا عند بناء نموذج عشوائي بأن كثير من السلاسل الزمنية لا يمكن تمثيلها باستخدام أنموذج الانحدار الذاتي AR فقط او أنموذج الوساط المتحركة MA لأنه ممكن ان تمتلك السلسلة خواص للنموذجين معا وبذلك نحصل على نموذج مناسب اكثر عند دمجها معا ويسمى النموذج المختلط من رتبة (p, q)، تعد نماذج (ARMA) اكثر نماذج السلاسل الزمنية استخداما حيث يمكن ان نشق جميع النماذج منها سواء الانحدار الذاتي أو الأوساط المتحركة أو المختلط (Chakravarti et al., 1973) (إسماعيل، 2014)، ويكتب بالصيغة الاتية:

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t - \theta_1 Y_{t-1} - \theta_2 Y_{t-2} - \dots - \theta_q Y_{t-q} \quad \dots (3)$$

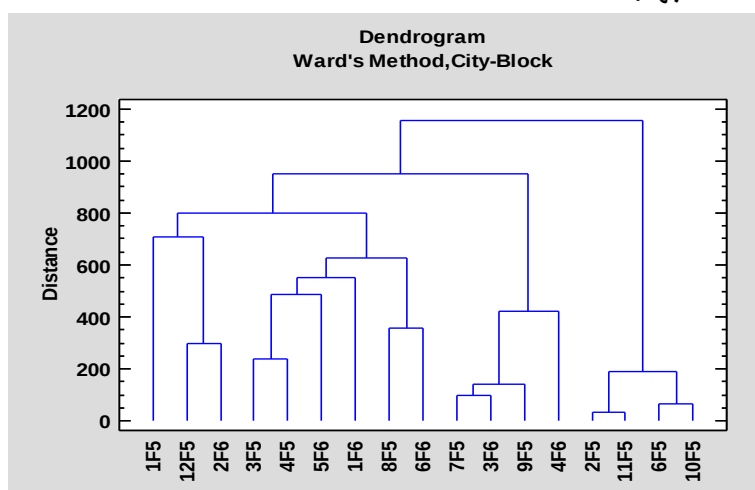
د. **النموذج المتكامل المختلط Integrated Mixed Model (ARIMA)**: نماذج ARIMA للسلسلة الزمنية الخطية الغير المستقرة، في حالة كون السلسلة الزمنية تكون غير مستقرة في المتوسط وبذلك نقوم بأجراء تحويلات وذلك بأخذ الفروق وتحويلها الى سلسلة مستقرة والفروق تأخذ بأعداد صحيحة موجبة (d) وتعد نماذج (p, d, q) (ARIMA) اكثر نماذج السلاسل الزمنية استخداما في عملية التنبؤ بالقيم المستقبلية، في نموذج أريما القيمة المستقبلية للمتغير هي مزيج خطي من القيم السابقة والأخطاء الماضية (Zhang, 2003) (Chakravarti et al., 1973) (Mumbare et al., 2014) (الكلابي، 2018) (علوان، 2020)، وصيغته كالتالي:

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B)a_t \quad \dots (4)$$

2. الجانب التطبيقي:

2-1. وصف البيانات: في هذا الجانب من الدراسة سنعرض النتائج التي حصلنا عليها عند اتباع خطوات تطبيق التحليل العنقودي على بيانات السلاسل الزمنية وخطوات تطبيق نماذج بوكس-جينكيز (Box and Jenkins) المتمثلة في مراحل (التشخيص، التقدير، فحص ملائمة النموذج، التنبؤ) لتقدير النموذج الملائم والكفؤ لتمثيل بيانات السلسلة الزمنية لكل عنقود مكون. ولإنجاز خطوات العمل تم استخدام برنامج (STATGRAPHICS-18).

2-2. التطبيق باستخدام التحليل العنقودي Application using cluster Analysis: لغرض تحليل المتغيرات والتي هي وحدات الانتاج تم اعتماد طريقة التحليل العنقودي الهرمي للسلاسل الزمنية (Hierarchical clustering) حيث لا يتطلب المعرفة المسبقة بعدد العناقيد التي سيتم تصنيف الحالات على أساسها وتكوين مخطط الشجرة لتوضيح مقدار التشابه بين المتغيرات، ولغرض تطبيق هذا النوع من التحليل للمتغيرات يجب اولا استخراج مصفوفة مسافة مناهاتن Manhattan distance (City-Block) او تسمى بمسافة القطاع وبالاتماد على تلك المصفوفة يمكن استخراج مستويات الالتحام بين وحدات الانتاج وتقرر التشابه بين الحالات وطريقة الربط الهرمية طريقة وورد (ward's method)، وبينت النتائج انه تكونت أربعة عناقيد cluster تضمنت الوحدات المتشابهة.



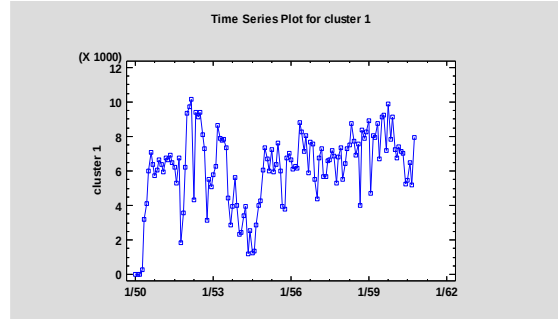
الشكل (1): الشكل الشجري للعناقيد

2-3. التطبيق باستخدام نماذج ARIMA على العناقيد:

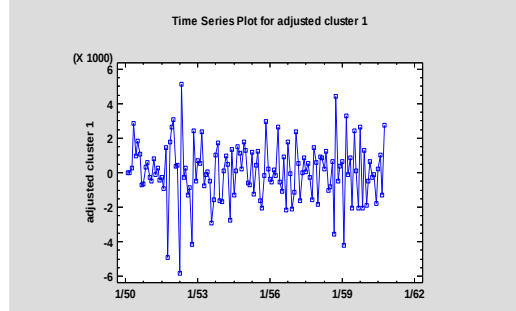
(Application using ARIMA Models for cluster):

2-3-1. التطبيق باستخدام نماذج ARIMA على العنقود الأول Cluster:

أولاً. التشخيص: نفحص السلسلة الزمنية هل هي مستقرة بالوسط والتباين ويمكننا من خلال الرسم الاطلاع على طبيعة التذبذب فيها، ان العنقود الاول يحتوي على (1F5, 12F5, 2F6) وحدات الانتاج هذه متشابهة من ناحية الانتاج لمدة 11 سنة تم ترتيب ربط المجموعات (العناقيد) وفقا للمسافة بينهم وبذلك المسافة هنا بين الوحدات صغيرة وتعقدت معا، حيث تم رسم بيانات الكمية المنتجة شهريا من الكهرباء ولرسم هذا السلاسل يجب ان نأخذ المعدل لهذه السلاسل لان كل عنقود يحتوي عدة سلاسل زمنية وبهذا تتكون لدينا سلسلة جديدة ونرسمها، يتبين من الشكل (2) ان السلسلة الزمنية مما يدل ان السلسلة غير مستقرة في المتوسط ولجعلها مستقرة تم أخذ الفروق الأولى وتحقيق الاستقرارية.

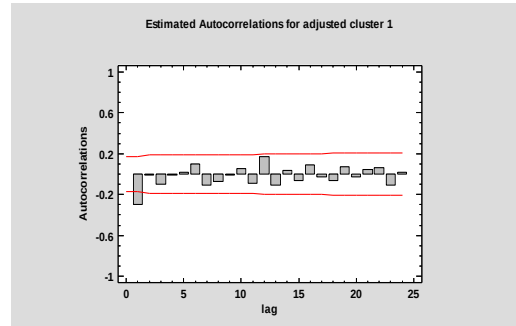


الشكل (2): المنحنى التكراري للسلسلة الزمنية الاصلية

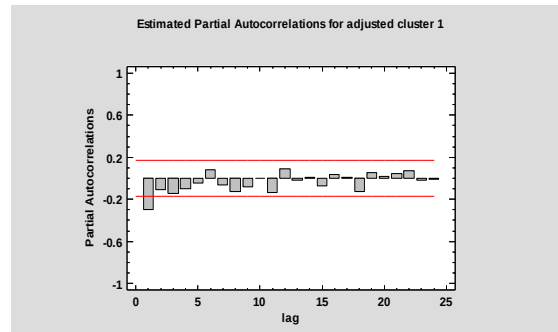


الشكل (3): المنحنى التكراري للسلسلة الزمنية بعد أخذ الفرق الأول

❖ **تحقيق الاستقرارية:** للوصول الى حالة الاستقرارية للسلسلة الزمنية نأخذ الفرق الأول وتحديد رتبة عنصر التكامل ($d=1$) ويجب ان نحدد رتبة كل من الانحدار الذاتي (AR) والمتوسط المتحرك (MA) من خلال رسم دالتي الارتباط الذاتي ودالة الارتباط الذاتي الجزئي.



الشكل (4): دالة الارتباط الذاتي ACF للسلسلة المعدلة



الشكل (5): دالة الارتباط الذاتي الجزئي PACF للسلسلة المعدلة

❖ **اختبار عشوائية السلسلة الزمنية بعد اخذ الفروق الأولى:** بعد اخذ الفرق الأول للسلسلة الزمنية وإعادة اختبار العشوائية للسلسلة الزمنية من خلال اختبار العشوائية (Ljung-Box test) يعتبر

اختبار لسلسلة الضوضاء البيضاء حيث إذا كانت قيمة (p-value) أكبر (0.05) تعتبر السلسلة مستقلة وعشوائية (independent and identically distribution: i.i.d) لا تحتوي على ارتباطات، وإذا كانت قيمة (p-value) أقل من (0.05) دل على وجود ارتباطات عالية بين البيانات.

نلاحظ ان (p-value > 0.05) لا نرفض فرضية العدم التي تنص ان السلسلة الزمنية عشوائية كما في الجدول (2):

الجدول (2): اختبار (ljung & box test) لاختبار العشوائية للسلسلة الزمنية المعدلة بعد اخذ الفرق الاول

فرصيات الاختبار	قيمة إحصاءة Q-	p-value
H ₀ : السلسلة عشوائية	30.3287	0.174095
H ₁ : السلسلة غير عشوائية		

تم الاعتماد على الاقتراح الذاتي من قبل البرنامج الاحصائي (STATGRAPHICS-18) واقتراح (11) نموذجاً واخذ اربعة نماذج (4) من اجل اختيار النموذج الملائم للبيانات الذي يمتلك اقل قيمة لاختبارات عديدة تم تطبيقها عليها تم الاعتماد على ثلاثة معايير احتساب معيار اكاكي (AIC)، ومعيار معلومات شوارتز (BIC) ومعيار حنان-كوين (HQC) كما في الجدول:

الجدول (3): النماذج المقترحة للسلسلة الزمنية للعنقود الاول Cluster1

Model	AIC	HQC	SBIC
ARIMA(0,1,1)	14.7872	14.7962	14.8093
ARIMA(1,1,0)	14.8124	14.8213	14.8344
ARIMA(1,1,1)	14.7929	14.8109	14.8370
ARIMA(2,1,1)	14.7877	14.8146	14.8539

ان النموذج الملائم هو النموذج الاول ARIMA(0,1,1) الذي تم اختياره بالاعتماد على معايير يمتلك فيها اقل قيمة لمعيار اكاكي (AIC) وكذلك اقل قيمة لمعيار حنان كوين (HQC) ومعيار شوارتز (SBIC).

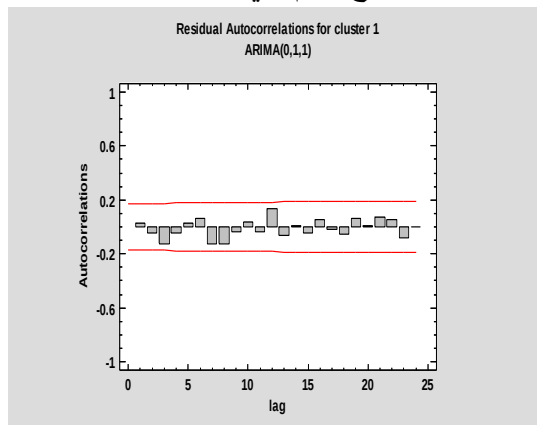
ثانياً. التقدير Estimation: بعد تحديد النموذج الملائم تم تقدير معاملات الانموذج ARIMA (0,1,1) وذلك لمعرفة مدى معنوية المعلمات المقدرة الجدول يوضح المعلمات المقدرة وفحصها إذا كانت معنوية أم لا، وذلك بالاعتماد على قيمة (p-value < 0.05) اي انه نرفض فرضية العدم ونقبل الفرضية البديلة القائلة بأن المعلمة المقدرة للأنموذج هي معنوية وتختلف عن الصفر وبالتالي تبقى في الأنموذج وتكون صيغة النموذج:

$$Y_t = 54.2056 + Y_{t-1} + a_t - 0.393941a_{t-1}$$

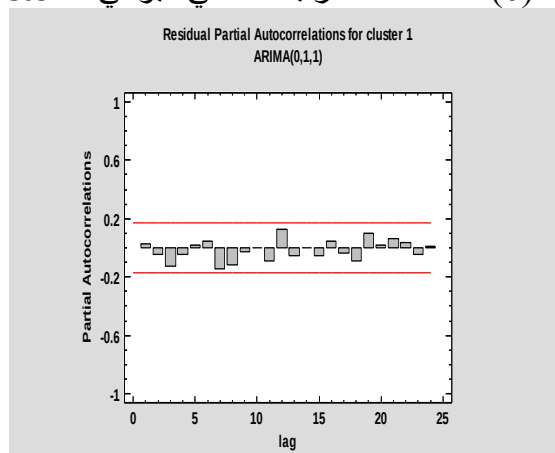
الجدول (4): تقدير قيم معاملات النموذج ARIMA (0,1,1)

Parameter	Estimate	Std. Error	T	P-value
MA(1)	0.393941	0.0826353	4.76723	0.000005
Mean	54.2056	88.6511	0.611449	0.541995
Constant	54.2056			

ثالثاً. فحص مدى ملائمة النموذج **Diagnostic Checking**: بعد ايجاد وتحديد معلمات النموذج الملائم تأتي مرحلة فحص مدى ملائمة الأنموذج وذلك بحساب البواقي Residuals (وهي الفرق بين القيم الفعلية والقيم التقديرية) ولأبد من اختبار صحة هذا النموذج من خلال اختبار معاملات الارتباط الذاتي للبواقي ومعاملات الارتباط الذاتي الجزئي للبواقي ويجب ان تكون جميع القيم ضمن حدود الثقة وهنا تحقق الشرط وجميع القيم هي ضمن حدود الثقة.



الشكل (6): معاملات الارتباط الذاتي للبواقي cluster 1



الشكل (7): معاملات الارتباط الذاتي الجزئي للبواقي cluster 1

يمكن معرفة مدى ملائمة النموذج المقترح $ARIMA(0,1,1)$ لبواقي السلسلة الزمنية لاختبار العشوائية للسلسلة وذلك باختبار الفرضية الموضحة في الجدول:

الجدول (5): اختبار العشوائية للنموذج $ARIMA(0,1,1)$

الفرضية	p-value
H_0 : النموذج ملائم	0.864421
H_1 : النموذج غير ملائم	

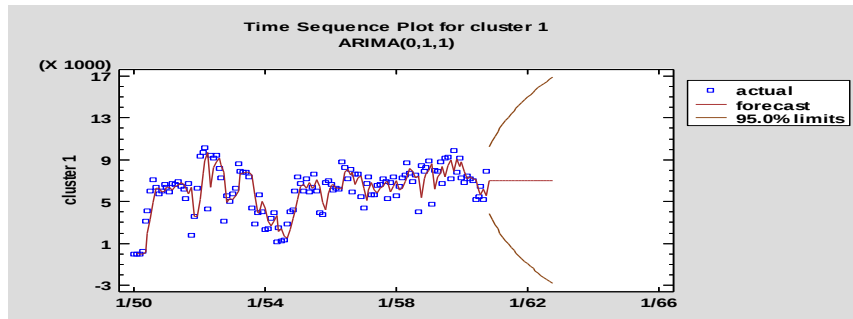
نرى ان قيمة p-value هي أكبر 0.05 هذا يدل بأننا لا نرفض فرضية العدم وان البواقي عشوائية وتتوزع بشكل مستقل وبذلك أثبتنا ان النموذج المقترح $ARIMA(0,1,1)$ هو نموذج جيد.

رابعاً. **التنبؤ Forecasting**: بعد ان تم تحديد النموذج الملائم للسلسلة الزمنية وتم تقدير معاملاته نصل الان الى اخر مرحلة وهي التنبؤ بكميات الإنتاجية للعنقود الاول الذي يضم (1F5, 12F5, 2F6) وحدات الانتاج في محطة كهرباء المهندس الشهيد عبدالله حمزة الغازية والتنبؤ يكون بشكل شهري لمدة سنتين.

الجدول (6): قيم التنبؤ لكمية الانتاج الكهربائي لسنتين للعنقود الاول cluster1

		Lower 95%	Upper 95%
Period	Forecast	Limit	Limit
11/2020	7013.1	3821.26	10204.9
12/2020	7013.1	3274.43	10751.8
1/2021	7013.1	2797.95	11228.2
2/2021	7013.1	2370.12	11656.1
3/2021	7013.1	1978.51	12047.7
4/2021	7013.1	1615.24	12411.0
5/2021	7013.1	1274.93	12751.3
6/2021	7013.1	953.69	13072.5
7/2021	7013.1	648.648	13377.5
8/2021	7013.1	357.573	13668.6
9/2021	7013.1	78.7044	13947.5
10/2021	7013.1	-189.375	14215.6
11/2021	7013.1	-447.827	14474.0
12/2021	7013.1	-697.622	14723.8
1/2022	7013.1	-939.574	14965.8
2/2022	7013.1	-1174.38	15200.6
3/2022	7013.1	-1402.64	15428.8
4/2022	7013.1	-1624.86	15651.1
5/2022	7013.1	-1841.52	15867.7
6/2022	7013.1	-2052.99	16079.2
7/2022	7013.1	-2259.65	16285.8
8/2022	7013.1	-2461.79	16488.0
9/2022	7013.1	-2659.72	16685.9
10/2022	7013.1	-2853.67	16879.9

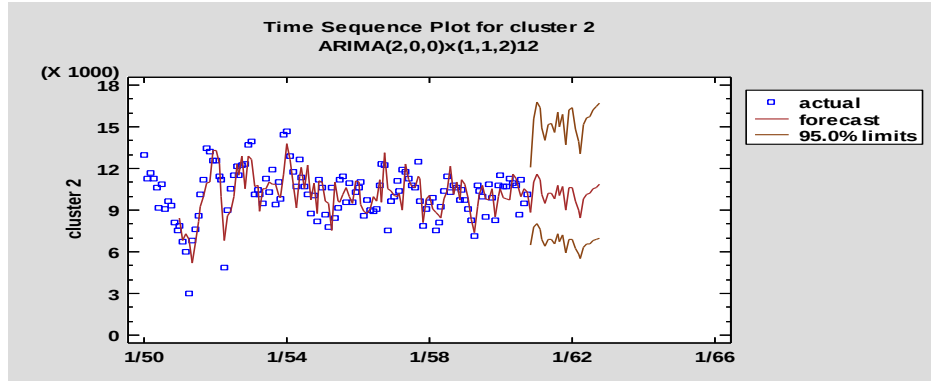
يوضح الجدول القيم المستقبلية ان جميع القيم المتنبئ تكون محصورة بين الحدين الاعلى والادنى وبنسبة ثقة 95%.



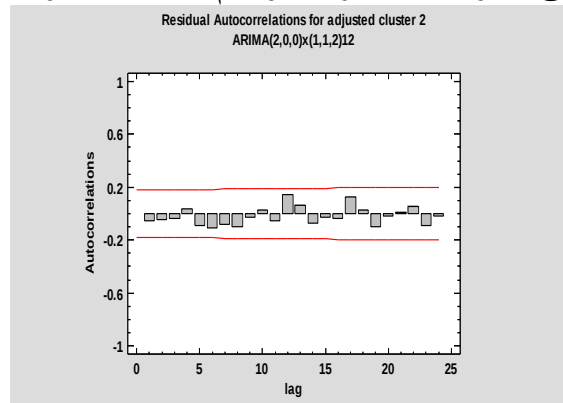
الشكل (8): منحني التنبؤ للسلسلة الزمنية للعنقود الاول cluster1

تطبيق نماذج ARIMA على العناقيد الثاني والثالث والرابع: تم تطبيق نفس الخطوات الموضحة في الجزء السابق في تحليل العنقود الاول cluster1 وتوصلنه الى انه في العنقود الثاني

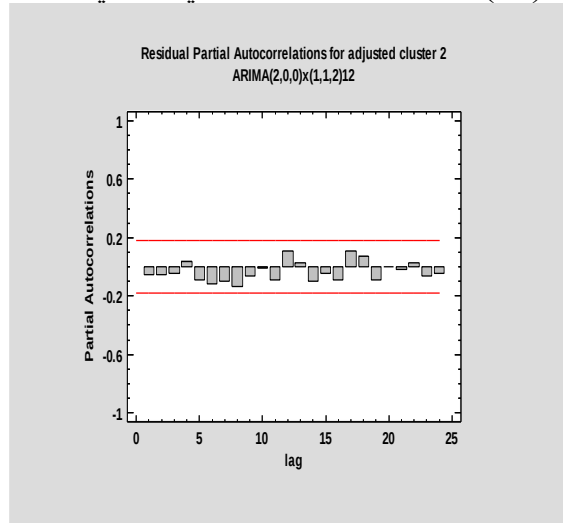
Cluster 2 وحدات الإنتاج في هذه المحطة هي (3F5, 4F5, 5F6, 1F6, 8F5, 6F6) ان السلسلة الزمنية تحتوي على المركبة الموسمية وهي غير مستقرة ولإزالة هذه التأثيرات نأخذ اللوغاريتم الطبيعي مع اخذ الفرق الاول الموسمي ويكون نموذج موسمي مضاعف $SARIMA(2,0,0) \times (1,1,2)^{12}$ نموذج كفو وجيد للتنبؤ بالقيم المستقبلية ومعاملات الارتباط الذاتي للبواقي والارتباط الذاتي الجزئي للبواقي جميع قيمها تقع ضمن حدود الثقة واثبتنا ان البواقي عشوائية حسب اختبار (Ljung & box test) ان قيمة $p\text{-value} > 0.05$ يدل ان النموذج ملائم.



الشكل (9): منحى التنبؤ للسلسلة الزمنية والقيم الاصلية للعنقود الثاني cluster2



الشكل (10): معاملات الارتباط الذاتي للبواقي cluster2



الشكل (11): معاملات الارتباط الذاتي الجزئي للبواقي cluster2

❖ اختبار **ljung& box test** للبواقي (الأخطاء): لابد من اختبار صحة ملائمة النموذج $SARIMA (2,0,0) \times (1,1,2)^{12}$ وذلك من خلال اختبار (ljung-box test) وبما ان قيمة $(p\text{-value} > 0.05)$ لا يمكن رفض فرضية العدم اي ان السلسلة الزمنية عشوائية ونستنتج ان الارتباطات الذاتية غير معنوية ويشير الى ان البواقي عشوائية وتتوزع بشكل مستقل واعتبار النموذج المشخص هو ملائم كما في الجدول التالي:

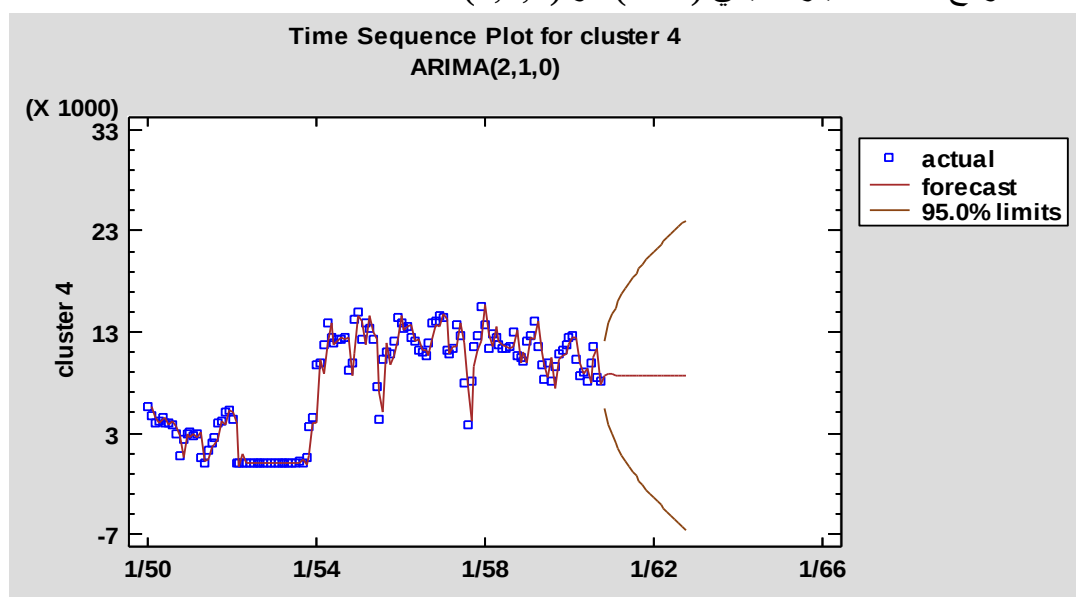
الجدول (7): اختبار ljung-box للعشوائية لسلسلة البواقي في نموذج

$$SARIMA (2,0,0) \times (1,1,2)^{12}$$

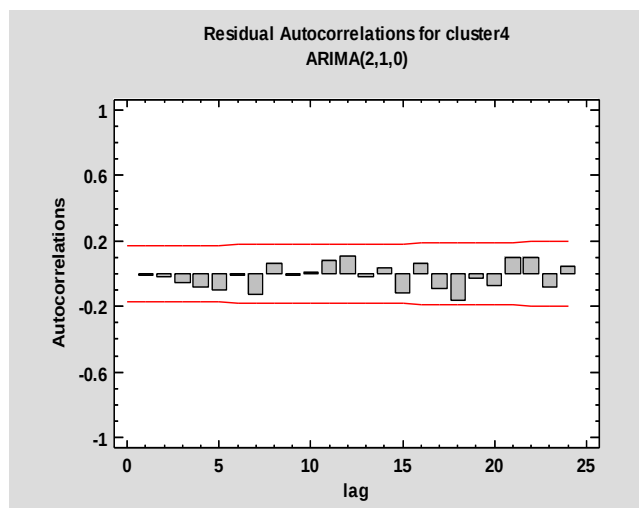
الفرضية	p-value
H_0 : السلسلة عشوائية	0.698563
H_1 : السلسلة غير عشوائية	

اما بالنسبة للعنقود الثالث Cluster 3 يحتوي على أربعة سلاسل زمنية لوحداث الإنتاج (7F5, 3F6, 9F5, 4F6) والتي هي متوقفة عن العمل منذ سنة 2014 و 2015 لتترك اعمال التأهيل من قبل احدى الشركات المختصة وفسخ العقد معها وصدور الامر القضائي بتغريم الشركة لتتركها موقع العمل وعدم تكملة اعمال التأهيل وبذلك تكون هذه المحطات هي خارج نطاق العمل ولا تعتبر قيم مفقودة لأنها متوقفة عن العمل أي ان الإنتاجية (صفر)، وبذلك يكون التنبؤ بالقيم المستقبلية لإنتاجية الطاقة الكهربائية في هذه الوحدات هي صفر.

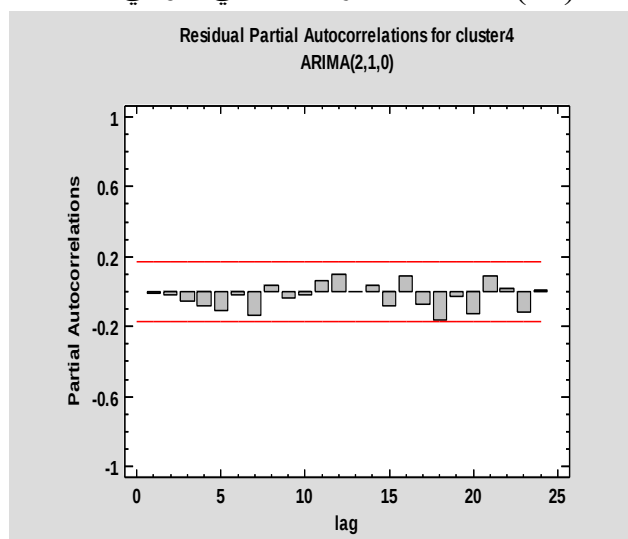
والعنقود الرابع Cluster 4 تحتوي اربعة سلاسل زمنية لأربعة وحدات لإنتاجية الطاقة الكهربائية (2F5, 11F5, 6F5, 10F5) السلسلة الزمنية المتكونة غير مستقرة بعد رسمها وإيجاد معاملات الارتباط الذاتي ACF ومعاملات الارتباط الذاتي الجزئي PACF واختبار (Ljung-box test) لاختبار العشوائية وكانت قيمة $p\text{-value} < 0.05$ يدل على عدم استقرار السلسلة الزمنية وتم اخذ الفروق الأولى للوصول للاستقرارية وبذلك توصلنا الى حالة الاستقرارية وأفضل نموذج حسب معيار اكاكي (AIC) هو $ARIMA (2,1,0)$



الشكل (12): يبين القيم المتنبى بها وقيم المشاهدات للسلسلة الزمنية للعنقود الرابع Cluster 4



الشكل (13): معاملات الارتباط الذاتي للبواقي cluster4



الشكل (14): معاملات الارتباط الذاتي الجزئي للبواقي cluster4

لمعرفة مدى ملائمة النموذج المقترح ARIMA (2,1,0) للسلسلة الزمنية من خلال اختبار (Ljung-box test) من خلال الفرضيات الموضحة في الجدول:

الجدول (8): اختبار العشوائية لنموذج ARIMA (2,1,0)

الفرضيات	قيمة الإحصاءة Q	p-value
H ₀ : النموذج ملائم	21.6769	0.479311
H ₁ : النموذج غير ملائم		

نلاحظ ان قيمة $P\text{-value} > 0.05$ فهذا يعني باننا لا نرفض فرضية العدم اي انه البواقي عشوائية وتتنوع توزيع مستقل وتتبع نمط السلسلة الضوضاء البيضاء (White Noise) مما يدل على استقلالية البواقي مما يؤكد ان النموذج جيد وملائم للتنبؤ بالقيم المستقبلية.

4-2. مقارنة بين نماذج العناقيد المتكونة: استخدم معايير ضبط دقة التنبؤ الجذر التربيعي لمتوسط مربع الأخطاء (Root Mean Square Error) (RMSE) لاختيار أفضل نموذج تنبؤ بين النماذج المستخدمة للتنبؤ لكل عنقود متكون كما موضح في الجدول (9):

الجدول (9): مقارنة بين نماذج العناقيد المستخدمة لتمثيل البيانات

ت	العنقود	النموذج	RMSE
1	Cluster 1	ARIMA (0,1,1)	1613.1
2	Cluster 2	SARIMA(2,0,0)x(1,1,2) ¹²	1250.37
3	Cluster 4	ARIMA(2,1,0)	1681.85

يوضح الجدول ان الأنموذج الثاني SARIMA (2,0,0)x(1,1,2)¹² للعنقود الثاني cluster 2 هو الأفضل لتسجيل اقل قيمة للمعيار RMSE أي انه الأفضل في تمثيل البيانات وان العنقود الثاني هو الأفضل بين العناقيد الأربعة المتكونة الذي يضم ستة وحدات لإنتاجية الطاقة الكهربائية التي تكون انتاجيتها في فصل الصيف اقل نسبيا بسبب تأثرها بدرجات الحرارة العالية على العكس في الأشهر الأخرى تكون انتاجيتها اعلى نسبيا.

الاستنتاجات والتوصيات

اولاً. الاستنتاجات: يمكن توضيح أهم الاستنتاجات بما يأتي:

1. تكمن أهمية التنبؤ بالإنتاجية الشهرية من الطاقة الكهربائية في دوره لتحسين سياسات المؤسسة حيث ان التنبؤ الجيد يؤدي الى تحسين أداء الإنتاجية من الطاقة التي بالتالي تؤدي الى تأمين احتياجات المجتمع من الطاقة الكهربائية الاستهلاكية اللازمة والضرورية للحياة المعاصرة.
2. تشير نتائج بحثنا العلمي الى أن خوارزمية السلاسل الزمنية العنقودية الهرمية Hierarchical clustering of time-series طريقة وررد Ward linkage وباستخدام قانون مسافة القطاع City Block Distance وتسمى ايضا مسافة منهاتن Manhattan اعطت نتائج هي الأفضل لهدف التجميع (العنفدة) لدينا والحصول على مجاميع متجانسة تكونت لدينا اربعة مجاميع (عناقيد) كل عنقود يحتوي على وحدات انتاج للطاقة الكهربائية متشابهة من حيث الإنتاج الى حد كبير، حيث العنقود الأول يحتوي على 3 وحدات انتاج (1F5, 12F5, 2F6)، والعنقود الثاني يحتوي على ستة وحدات انتاج (3F5, 4F5, 5F6, 1F6, 8F5, 6F6)، العنقود الثالث يحتوي 4 وحدات انتاج وهي (7F5, 3F6, 9F5, 4F6)، والعنقود الرابع يحتوي اربع وحدات انتاج (2F5, 11F5, 6F5, 10F5).
3. وباستخدام طريقة بوكس-جينكينز على كل عنقود cluster متكون لدينا وجد ان السلسلة الزمنية للفترة الزمنية (2020-2010) لوحدات الإنتاج غير مستقرة (ليس عشوائيا) وبذلك تم اجراء التحويلات اللازمة بهدف تحقيق الاستقرار.
4. تم اختيار أفضل نموذج للتنبؤ بالقيم المستقبلية لكل عنقود متكون بالاعتماد على معايير المفاضلة (أقل قيمة لتباين النموذج معيار اكاكي (AIC) ومعيار شوارتز (SBIC) ومعيار حنان كوين (HQC))
5. وجد ان النموذج الكفو والملائم للعنقود الأول هو ARIMA(0,1,1) والنموذج المناسب للعنقود الثاني هو SARIMA (2,0,0)x(1,1,2)¹²، وبالنسبة للعنقود الثالث يحتوي على وحدات انتاج للطاقة المتوقعة عن العمل من سنة 2014 تقريبا ولحد الان هيه متوقفة عن العمل وبذلك التنبؤ بالإنتاجية يكون صفر للسنتين القادمتين، والنموذج الذي تم التوصل اليه للعنقود الرابع ويعتبر كفو هو ARIMA (2,1,0).

6. وفقا لهذه النماذج تم التنبؤ بكميات الإنتاج بالطاقة الكهربائية في محافظة كركوك/العراق محطة كهرباء الشهيد المهندس عبدالله حمزة الغازية لفترة (24) شهرا من شهر تشرين الثاني 2020 لغاية تشرين الثاني 2022 حيث أظهرت هذه القيم تناسقا مع مثيلاتها في السلسلة الأصلية، وقدمت لنا صورة مستقبلية لواقع انتاج الطاقة الكهربائية لنستنتج ان طريقة التحليل العنقودي الهرمي للسلاسل الزمنية افضل بكثير من عملية التنبؤ بشكل منفرد لكل وحدة انتاج لأنها تختصر الوقت والجهد حيث اختزلت الوحدات من 17 الى اربع مجموعات (عناقيد) متجانسة فيما بينها، وتعطي معلومات ادق وبشكل مفصل لكل وحدة انتاج موجودة في العناقيد (المجموعات).

ثانياً. التوصيات:

1. نوصي باعتماد طريقة التحليل العنقودي للسلاسل الزمنية والتوسع للبحث وأخذ المحطات في كركوك كلها وبالإمكان الأخذ بنتائج البحث وتعميمه على بقية المحافظات التي تعاني انخفاض كبير في انتاج الطاقة الكهربائية في كل العراق والتنبؤ لإنتاجيتها والتوسع بتحليل سلاسل زمنية للغاز او النفط وغيرها.
2. بما ان العراق في الوقت الحالي بحاجة الى عملية اعمار واسعة نوصي بزيادة الطاقة الإنتاجية للطاقة الكهربائية لأنها لا تسد الحاجة الاستهلاكية للمجتمع.
3. نوصي بالاستفادة من النتائج لوضع الخطط والسياسات المستقبلية ونوصي بأخذ الاستهلاك بكل جوانبه الاستهلاك المنزلي والصناعي للطاقة الكهربائية.
4. استخدام الطرق الأخرى للعنقدة للسلاسل الزمنية مثل الطرق الغير الهرمية للعنقدة. (K-means Clustering).

المصادر

اولاً. المصادر العربية:

1. أحمد، طالب، (2015)، تصنيف المحافظات السورية حسب الانفاق الاستهلاكي للأسرة باستخدام التحليل العنقودي، مجلة جامعة تشرين للبحوث والدراسات العلمية، مجلد (37)، العدد (2).
2. إسماعيل، أبوذر اسماعيل مفرح، (2014)، المقارنة بين نموذج السلاسل الزمنية والانحدار البسيط في التنبؤ بحجم المبيعات، رسالة ماجستير، جامعة السودان للعلوم والتكنولوجيا، كلية الدراسات العليا، قسم الاقتصاد التطبيقي.
3. الجوهري، هناء طه عطاالله، (2016)، أسلوب احصائي مقترح لمعالجة بيانات السلاسل الزمنية العنقودية متعددة المتغيرات (دراسة تطبيقية)، أطروحة دكتوراه، جامعة المنصورة، كلية التجارة، قسم الإحصاء التطبيقي والتأمين، 2016.
4. علوان، اقبال محمود وحمزة، لمى كريم، (2020)، استعمال السلاسل الزمنية في التنبؤ بكمية النفقات الصلبة لمحافظة بغداد للمدة 2008-2018، مجلة كلية الرافدين الجامعة للعلوم، وقائع المؤتمر العلمي الخامس عشر والدولي الثاني للتطبيقات الإحصائية-الجمعية العراقية للعلوم الإحصائية، العدد 46، ص270-284.
5. الكلابي، صفاء مجيد مطشر، (2018)، استعمال بعض طرائق التنبؤ المختلفة لتحليل اعداد المصابين بالأورام الخبيثة رسالة ماجستير، جامعة كربلاء، كلية الإدارة والاقتصاد، قسم الاحصاء.
6. نامق، فيصل ناجي، أسلوب التحليل العنقودي لتصنيف الانفاق على السلع والخدمات الأساسية وفقا للمستوى البيئي (حضر وريف) للسنوات 1971-2007، الكلية التقنية الإدارية-بغداد، مجلة كلية بغداد للعلوم الاقتصادية الجامعة، العدد 25، 2010، ص 331-352.

ثانياً. المصادر الأجنبية:

1. Aghabozorgi, S., Seyed Shirkhorshidi, A., & Ying Wah, T., (2015), Time-series clustering, A decade review. In Information Systems (Vol. 53, pp. 16–38). <https://doi.org/10.1016/j.is.2015.04.007>
2. Andrés M. Alonso, Francisco J. Nogales, and C. R., (2020), Hierarchical Clustering for Smart Meter Electricity Loads Based on Quantile Autocovariances Andrés.
3. Bagnall, A., & Janacek, G., (2005), Clustering Time Series with Clipped Data. Machine Learning, 58 (2–3), 151–178. <https://doi.org/10.1007/s10994-005-5825-6>
4. Chakravarti, I. M., Box, G. E. P., & Jenkins, G. M., (1973), Time Series Analysis Forecasting and Control. In Journal of the American Statistical Association (Vol. 68, Issue 342, p. 493). <https://doi.org/10.2307/2284112>
5. Ergüner Özkoç, E., (2021), Clustering of Time-Series Data. In Data Mining - Methods, Applications and Systems. IntechOpen. <https://doi.org/10.5772/intechopen.84490>
6. Karakaş, E., (2019), Forecasting Automotive Export Revenue of Turkey using ARIMA Model. Journal of Yasar University, 318–328.
7. Kavitha, V., & Punithavalli, M., (2010), Clustering Time Series Data Stream-A Literature Survey. <http://arxiv.org/abs/1005.4270>
8. Li, J., (2017), Clustering and forecasting for rain attenuation time series data. In Journal of Atmospheric and Solar-Terrestrial Physics. <https://doi.org/10.1016/j.jastp.2017.12.001>
9. Liao, T. W., (2005), Clustering of time series data - A survey. In Pattern Recognition (Vol. 38, Issue 11, pp. (1857–1874). <https://doi.org/10.1016/j.patcog.2005.01.025>
10. Mumbare, S., Gosavi, S., Almale, B., Patil, A., Dhakane, S., & Kadu, A., (2014), Trends in average living children at the time of terminal contraception: A time series analysis over 27 years using ARIMA (p, d, q) nonseasonal model. Indian Journal of Community Medicine, 39(4), 223. <https://doi.org/10.4103/0970-0218.143024>
11. Rodrigues, P. P., Gama, J., & Pedroso, J. P., (2008), Hierarchical Clustering of Time-Series Data Streams. IEEE Transactions on Knowledge and Data Engineering, 20(5), 615–627. <https://doi.org/10.1109/TKDE.2007.190727>
12. Roelofsen, P., (2015), Time-series clustering. In Master thesis Business Analytics (pp. 241–264). <https://doi.org/10.1201/b19706>
13. Zhang, P. G., (2003), Time series forecasting using a hybrid ARIMA and neural network model. In Neurocomputing (Vol. 50, pp. 159-175). [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)