



المجلة العراقية للعلوم الإحصائية

www.stats.mosuljournals.com



التنبؤ لبيانات تلوث الهواء باستخدام الطريقة الهجينة RNN-Wavelet بالاعتماد على نموذج MLR

ختام وليد قادر^{ID} و اسامة بشير شكر^{ID}

قسم الاحصاء والمعلوماتية ، كلية علوم الحاسوب والرياضيات ، جامعة الموصل ، الموصل ، العراق

الخلاصة

ان دراسة الجسيمات المعلقة ((Particular matter (PM10) والتكهن بها ضروري للتقليل والسيطرة على الأضرار البيئية وصحة الانسان. هنالك العديد من مصادر التلوث او ما يسمى بالملوثات والتي ربما تؤثر على متغير PM10. كل هذه المتغيرات تصنف بياناتها كغير خطية. اخذ بيانات الدراسة من محطة مناخية في ماليزيا. وتم استخدام نماذج الانحدار الخطي المتعدد (MLR) كطريقة إحصائية خطية للتنبؤ بمتغير PM10 من خلال تأثره بمتغيرات الأرصاد الجوية المقابلة، لذلك فقد يعكس نتائج غير دقيقة عند استخدامه مع مجموعات البيانات غير الخطية. ولتحسين نتائج التنبؤ تم استخدام الشبكات العصبية المعادة (RNN) Recurrent neural network من خلال توقيفه مع نموذج MLR ضمن الطريقة الهجينة MLR-RNN. وكذلك تم استخدام تحليل المويجات Wavelet في ترشيح النتائج المستحصلة من طريقة MLR-RNN للحصول على تحسين اكثر للنتائج التنبؤ من خلال ما يسمى الطريقة الهجينة RNN-Wavelet واعتماداً على نموذج MLR. ان افضل نتائج التنبؤ بشكل عام كانت باستخدام الطريقة الهجينة RNN-Wavelet. ان افضل نتائج التنبؤ بشكل عام كانت باستخدام الطريقة المقترحة الهجينة RNN-Wavelet. كما عكست النتائج افضلية تنبؤات الطرق الهجينة مقارنة مع الطريقة التقليدية المتمثلة بنموذج MLR. وكاستنتاج في هذه الدراسة فمن الممكن استخدام التحليل المويجي بعد تهجينه مع الشبكات العصبية المعادة اعتماداً على نموذج الانحدار الخطي المتعدد كأسلوب فعال للحصول على افضل نتائج التنبؤ مع البيانات غير الخطية متعددة المتغيرات عندما يكون PM10 كمتغير معتمد.

معلومات النشر

تاريخ المقالة:
تم استلامه في 7 اذار 2021
تم القبول في 17 نيسان 2021
متاح على الإنترنت في 21 كانون الاول 2021

الكلمات الدالة:

الانحدار الخطي المتعدد (MLR) ،
الشبكة العصبية المتكررة (RNN) ،
الموجة ، مسألة معينة (PM10) ،
التنبؤ ،

MLR-RNN الهجين ،
RNN-Wavelet الهجين

المراسلة:

ختام وليد قادر

khetalzubaiddy86@gmail.com

DOI: <https://doi.org/10.33899/ijoss.2021.169969> , ©Authors, 2021, College of Computer and Mathematical Science, University of Mosul
This is an open access article under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

1. المقدمة Introduction

تم الاعتماد على نموذج الانحدار الخطي المتعدد MLR في التنبؤ بتلوث الهواء مع الأخذ في الاعتبار تأثيرات متغيرات الأرصاد الجوية المقابلة له. قام (Vlachogianni et al. (2011) بدراسة لتطوير نماذج التنبؤ باستخدام الانحدار الخطي المتعدد MLR ومقارنة التنبؤات مع طريقة الشبكة العصبية الاصطناعية ANN. ان نتائج تنبؤات الشبكات العصبية كانت أفضل من تلك المستحصلة من نموذج MLR. وقد استخدم (Janssen et al. (2011) نماذج الانحدار الخطي المتعدد لنمذجة بيانات PM10 لمدينة باريس للمستويات اليومية. كما وقدر (Ahmad et al. (2019) تركيز الجسيمات الدقيقة PM_{2.5} باستخدام اسلوب هجين للانحدار الخطي والشبكة العصبية الاصطناعية وأظهرت النتائج أن كلا من ANN و MLR متفان تماماً وقادران على تقدير تركيزات PM_{2.5} مع دقة اكبر للشبكات العصبية الصناعية ANN. اقترح الباحث (Okkan (2012) تحليل المويجات للتنبؤ بالتدفق الشهري لمخزون السد وتشير النتائج المحددة في

الدراسة إلى أن Wavelet هي أداة ناجحة لنمذجة سلسلة التدفق الشهري للسد ويمكن أن تعطي تنبؤاً جيداً مقارنة بالطرق الأخرى. كما وحلل (Mandrikova et al. (2017) الخصائص المميزة لبيانات الغلاف الأيوني باستخدام نظام هجين يعتمد على مجموعة من عمليات ترشيح الموجات والشبكات العصبية وكانت النتيجة تشير إلى كفاءة اختبار النظام الهجين. وتتأياً (Zhang et al. (2019) بمستوى المياه الجوفية في الأراضي المستصلحة على شاطئ البحر باستخدام النموذج الهجين للشبكة العصبية الاصطناعية وتحليل الموجات وظهرت النتيجة قدرة وفعالية هذه الطريقة في التنبؤ. وقد تنبأ أيضاً بالاعتماد على نموذج الإنحدار الخطي المتعدد وتحليل الموجات على تحسين نتائج تنبؤ PM_{10} . واستخدم (Zhou et al. (2020) نموذج هجين يجمع بين التحليل المويجي والشبكة العصبية الاصطناعية للتنبؤ بملوحة مصبات الأنهار وكانت النتائج جميعها تشير إلى تفوق نماذج Wavelet-ANN الهجينة في التنبؤ. تم في هذه الدراسة تقديم عدة طرائق للتنبؤ كنموذج الانحدار الخطي المتعدد (MLR) كطريقة تقليدية احصائية فقد تم استخدام احد الطرق الذكائية وهي الشبكات العصبية المعادة RNN لتحسين نتائج التنبؤ من خلال تهجينها مع الطريقة التقليدية. كما وتم استخدام تحليل الموجات Wavelet في ترشيح النتائج المستحصلة من طريقة MLR-RNN والحصول على تحسين اكثر لنتائج التنبؤ من خلال ما يسمى الطريقة الهجينة RNN-Wavelet اعتماداً على نموذج MLR. تم التطبيق على بيانات سلسلة معدلات تلوث الهواء المتمثلة بالجسيمات المعلقة PM_{10} بواقع (1034) مشاهدة حيث تتكون هذه السلسلة من المتغير المعتمد PM_{10} ومتغيرات تفسيرية مقابلة له. وتم استخدام برنامج (Minitab) وبرنامج أكسل للحصول على نموذج الانحدار الخطي المتعدد وتنبؤاته. ان البيانات التي تمت دراستها تم اخذها بفترتها الكلية وتبدأ من (1/1/2013) وتنتهي الى (31/10/2015). ان الطرق المستخدمة لتحسين التنبؤ بتلوث الهواء من خلال متغيرات PM_{10} وتم مقارنة الطرائق الثلاثة MLR و MLR-RNN و RNN-Wavelet للحصول على افضل تنبؤ بتلوث الهواء من خلال متغيرات PM_{10} بالاعتماد على معيار المقارنة MAPE.

2. الطرق المستخدمة للتنبؤ

في هذا الجانب تم التطرق إلى استخدام احد طرق التنبؤ شائعة الاستخدام للتنبؤ بالبيانات متعددة المتغيرات وهي نموذج الانحدار الخطي المتعدد MLR. ورغم ما يوفره نموذج الانحدار الخطي المتعدد من جودة في نمذجة البيانات الا انه قد يكون غير ملائم للتنبؤ بالبيانات غير الخطية على اعتبار ان نموذج MLR نموذج خطي مما يؤدي إلى ظهور بعض النتائج والتنبؤات بدقة قليلة أحياناً ولذلك فمن المفضل استخدام طرق غير خطية مثل الشبكات العصبية المعادة RNN مما يؤدي للحصول على نتائج أدق للتنبؤ (Jahandideh et al., 2009). كما ومن الممكن ترشيح نتائج التنبؤ للحصول على تحسين ونتائج اكثر دقة من خلال طريقة التحليل المويجي Wavelet Analysis.

2.1 نموذج الانحدار الخطي المتعدد (MLR)

ان نموذج الانحدار الخطي المتعدد يعتبر من الأساليب الإحصائية شائعة الاستخدام والمستخدمه كذلك في التنبؤ خصوصاً لبيانات السلاسل الزمنية عن طريق الاستخدام الأمثل للبيانات في إيجاد علاقات سببية بين بيانات الدراسة ويعرف الانحدار الخطي المتعدد بشكل عام بأنه أسلوب رياضي لتوضيح العلاقة بين المتغير المعتمد Dependent Variable ومتغير تفسيري واحد فيسمى بنموذج الانحدار الخطي البسيط او مجموعة من المتغيرات التفسيرية تسمى بنموذج الانحدار الخطي المتعدد MLR (Honarasa et al., 2015). ويمكن صياغة النموذج MLR لشرح العلاقة بين متغير معتمد ومتغيران تفسيريان أو أكثر على النحو الآتي: (Abrougui et al., 2019)

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m + e_i \quad (1)$$

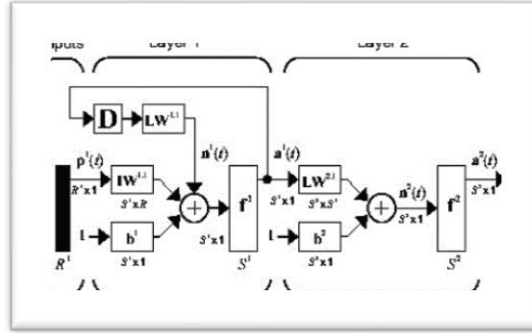
اذ أن y يرمز الى المتغير المعتمد و α ترمز الى القيمة الثابتة وترمز (x_1, x_2, \dots, x_m) للمتغيرات التفسيرية و ($\beta_1, \beta_2, \dots, \beta_m$) ترمز لمعاملات النموذج و e_i يرمز الى الخطأ العشوائي.

2.2 الشبكات العصبية الاصطناعية (ANN) Artificial Neural Networks

تعد الشبكات العصبية الاصطناعية أحد أهم طرق الذكاء الاصطناعي والتي تستخدم في التنبؤ بالمتغير المعتمد للحصول على دقة اكبر، و تتمحور فكرتها بمجموعة من الخوارزميات يتم من خلالها محاكاة الدماغ (العقل) البشري في التعرف على الأنماط وتمييز الأشياء باستخدام الحاسب الآلي، والتي يتم فيها الاستفادة من الخبرات السابقة في سبيل الوصول إلى أفضل نتائج (Lin et al., 2020). والمميز في الشبكات العصبية الاصطناعية هو وجود طبقات عديدة تعمل على ما يسمى التعلم العميق، كل طبقة مختصة بعمل معين (Zhou et al., 2020). توجد عدة أنواع من الشبكات العصبية الشائعة الاستخدام واهم هذه الانواع والتي تؤدي غالباً الى نتائج اكثر دقة هي الشبكات

العصبية المعادة (RNN) Recurrent Neural Network (Torkashvand *et al.*, 2017). ان الشبكات العصبية المعادة RNN تستخدم مخرجات طبقة معينة وتعيدها مرة أخرى للشبكة لكن كمدخلات ونتيجة لذلك يمكن أن يساعد ذلك في التنبؤ بالعديد من النتائج المحتملة خلال أي طبقة من طبقات الشبكة فتحتفظ كل طبقة بذاكرة من الخطوة السابقة فيتذكر النظام التنبؤات الخاطئة ويتعلم منها لتحسين تنبؤاته التالية ولذلك تسمى باسم الشبكات المعادة لمعاودتها نفس الخطوات حتى الوصول إلى النتائج المطلوبة وبالتالي تستطيع شبكات RNN التعلم من كل خطوة للتنبؤ بالنتيجة في الخطوة التالية.

تحتوي RNN على طبقة واحدة او اكثر وهذا بدوره يعالج غير خطية البيانات ويحسن نتائج التنبؤ وكذلك تحوي على Daley Layer وهذا يحسن كثيراً التعامل مع مشكلة عدم تجانس البيانات وغير الخطية لانه يحتوي على ذاكرة اطول بقليل من خوارزمية الشبكة العصبية المغذية Feed-Forward Back Propagation والشكل التالي يمثل الشبكة وما تحتويه من ادخالات واخراجات وطبقات .



الشكل (1) : يوضح الشبكة العصبية المعادة RNN

في الشكل (1) فإن R هي الادخالات و $LW1$ هي وزن عشوائي للعصبون حيث يتم جمعها مع الجزء المتحيز $b1$ (التشويش الابيض) وناتجها سيكونان الدالة $f1$. حيث ان اخراج الدالة $f1$ سيعود كادخال ثالث في الطبقة الاولى وقبلها سوف يمر على دالة التأخير (Delay) لتكون وزنا عشوائيا اخر وفي الطبقة الثانية فإن اخراج الدالة f يكون الوزن العشوائي للخلية العصبية LW مجموع مع b وبالتالي تخرج لنا مصفوفة احادية. تحتوي RNN في هذه الدراسة على طبقتين الاولى طبقة الادخال، والثانية تكون طبقة الاخراج. حيث في طبقة الادخال سيكون هناك (R) من الادخالات وهذه الادخالات غالبا ماتكون توزن عشوائيا في كل طبقة مخفية وكذلك M من العصبونات. حيث يتم حساب العدد الامثل للعصبونات في الطبقة المخفية (Palit & Popovic, 2006; Sheela & Deepa, 2013) هي:

$$\text{عدد العصبونات} = R*2+1 \quad (2)$$

كل متغير ادخال Z موزون عشوائيا. وان اوزان N من الادخالات و M من العصبونات تجمع مع القيمة المتحيزة b بواسطة دالة التحويل (F) التي تمثل مجموع ادخالات المتغيرات وتصاغ رياضيا كما يأتي:

$$net_j(t) = \sum_{i=1}^N w_{i,j} Z_j(t) + b_j \quad (3)$$

ان الاوزان العشوائية $w_{i,j}$ للادخالات يمكن كتابتها كمصفوفة وعلى النحو التالي

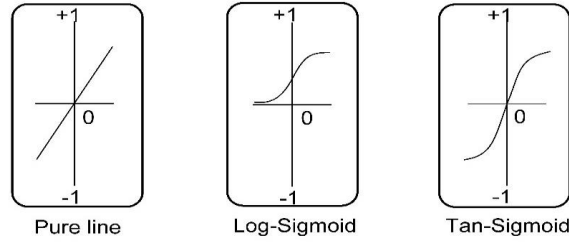
$$W = \begin{bmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,R} \\ w_{2,1} & w_{2,2} & \dots & w_{2,R} \\ \vdots & \vdots & \ddots & \vdots \\ w_{M,1} & w_{M,2} & \dots & w_{M,R} \end{bmatrix} \quad (4)$$

في حين تصاغ متغيرات الادخال كما يلي :

$$Z = [Z_{t1} \quad Z_{t2} \quad \dots \quad Z_{tR}]^T \quad (5)$$

وان اكثر دوال التحويل استخداما في الطبقة المخفية وطبقة الاخراج هي التحويل الزاوي (tan-sigmoid) ضمن الفترة (1,-1) والتحويل اللوغارتمي (log-sigmoid) ضمن الفترة (0,1) ودالة التحويل الخطي (linear) ضمن الفترة (1,-1). ان عملية اختيار الدالة للطبقتين امر

مهم جدا يؤدي الى تحسين دقة النتائج بالاعتماد على طبيعة البيانات والدالة المختارة ومدى التجانس بينهما، الشكل (2) يوضح الاختلافات بين دوال التحويل الممكن استعمالها للشبكة RNN.



الشكل (2) : انواع دوال التحويل في RNN

تستعمل دوال التحويل في الطبقة المخفية لتعكس نوعية العلاقة بين الادخالات والاطراقات في حين تستعمل دوال التحويل في طبقة الاخراج لتعطي افضل وادق النتائج، والصيغ الرياضية للدوال الخطية واللوغارتمية والزاوية هي كما يلي وعلى التوالي:

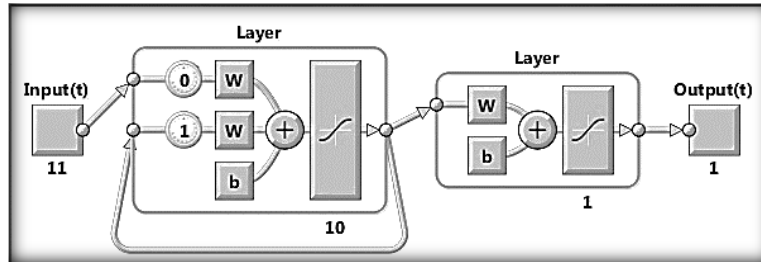
$$f(SUM) = SUM \quad (6)$$

$$f(SUM) = \frac{1}{1 + e^{-SUM}} \quad (7)$$

$$f(SUM) = \frac{2}{1 + e^{-2SUM}} - 1 \quad (8)$$

(Dawson & Wilby, 2001; Shrestha *et al.*, 2005; Yonaba *et al.*, 2010)

حيث ان الهيكل العام للشبكة RNN يحتوي على طبقة واحدة مخفية واخرى للاخراج، حيث الطبقة المخفية تحتوي على ثلاث مدخلات الوزن العشوي والتشويش الابيض بالاضافة الى الوزن الناتج من الخطوة السابقة اما الطبقة الخارجية فمدخلاتها ستكون الوزن الناتج من الطبقة المخفية مع التشويش الابيض ويمكن تمثله كما في الشكل (3) :



الشكل (3) : الهيكل العام للشبكة العصبية المعادة RNN

2.3 الطريقة الهجينة (MLR-RNN)

تتضمن هذه الطريقة تهجين الطريقتين التقليدية (MLR) والذكائية (RNN) من خلال الاستفادة من هيكلية المتغيرات التفسيرية في MLR لبناء الشبكة RNN وكما هو مدرج في الخطوات التالية:

أ. يتم ضرب كل متغير تفسيري في قيمة المعلمة المناظرة له واعتماد المتغيرات الناتجة كمدخلات للشبكة العصبية المعادة (RNN) وبناء طبقة الادخال .

ب. تحديد عدد العصبونات المستخدمة في الطبقة المخفية.

ج. بعد اعتماد هيكلية نموذج MLR لطبقة الادخال للشبكة العصبية حيث تتم عمليتي التدريب والاختبار للحصول على افضل التنبؤات وتسمى مخرجات شبكة RNN وهذه التنبؤات يتم استخدامها في الطريقة الهجينة للشبكات العصبية MLR-RNN، حيث كلما تكررت التدريبات وكانت اكثر كلما كانت النتيجة ادق.

د. يتم تدريب الشبكة على جميع دوال التحويل الخطية واللوغارتمية والزاوية في كلا الطبقتين المخفية والخراج وبأخذ جميع الاحتمالات الممكنة .

2.4 التحليل الموجي Wavelet Analysis

يعد تحليل الموجات من المواضيع الحديثة الاستخدام في كافة المجالات ويعتبر مكملاً للتحليل الطيفي الذي يعتمد على التحويل من مجال الزمن إلى مجال التكرار بالاعتماد على قواعد الدوال المثلثية أو الأسية المعقدة التي تؤدي إلى الكشف عن الدورية في حركة الموجة عن طريق الاعتماد على التحويل إلى مجال القياس والإزاحة باستخدام قواعد دوال متعامدة تؤدي إلى الكشف عن انتظام أو عدم الانتظام في حركة الموجة مما ساعد على استخدام التحليل الموجي في مجالات التنبؤ. ان مفهوم التحليل الموجي هو عبارة عن طريقة تحليلية رياضية تستخدم من أجل معالجة الإشارات للعديد من التطبيقات العملية، إذ تم تطوير النظرية الأساسية عدة مرات وظهرت الحاجة للتنبؤ باستخدام الموجات نتيجة لبعض العيوب التي تظهر أثناء القيام بعملية التنبؤ باستخدام نموذج الانحدار الخطي المتعدد. يستخدم التحليل الموجي في إعادة تشكيل الصور والإشارات باعتماد دالتين موجيتين إحداهما للجزئة والأخرى لإعادة التشكيل (Zhang et al., 2019). ونظراً للعيوب أو المعوقات التي قد تعيق أحياناً استخدام كلاً من طريقتي الانحدار الخطي المتعدد والشبكات العصبية الاصطناعية في التنبؤ فسيتم اعتماد نظام التحليل الموجي بعد تهجينه مع الشبكات العصبية المعادة RNN من خلال ترشيح نتائج طريقة MLR-RNN لتحسينها وجعلها أكثر دقة في التنبؤ إذ لا تستهلك الكثير من الوقت كالانحدار الخطي المتعدد والشبكات العصبية المعادة خاصة في حالة وجود العديد من المتغيرات التفسيرية التي تؤثر على المتغير المعتمد كما وتأخذ دوال التحليل الموجي في الاعتبار تلك المؤثرات الأخرى التي قد تؤثر بطريقة سلبية على دقة النتائج لتداخلها مع المتغيرات التفسيرية والتأثير عليها كما يحدث في حالة استخدام نموذج الانحدار الخطي المتعدد. تستطيع دوال التحليل الموجي كذلك فصل المتغيرات التي ليس لها تأثير جوهري على المتغير المعتمد بطريقة سلسة وبسيطة عن طريق تحليل موجات هذه المتغيرات أو المؤثرات بسرعة فائقة وإعطاء نتائج دقيقة لمدى تأثيرها على المتغير المعتمد ومن ثم استبعاد المتغير الذي ليس له تأثير والبقاء على تلك المتغيرات التي تؤثر بشدة على دقة التنبؤ بالمتغير المعتمد.

4.2. الطريقة الهجينة (RNN-Wavelet)

1. سيتم استخدام الهيكل النهائي لافضل نموذج انحدار خطي متعدد من خلال (ضرب كل متغير في قيمة المعلمة المناظرة له) للحصول على متغيرات جديدة ومن ثم اعتمادها كمدخلات للشبكة العصبية المعادة (RNN).
2. بعد اعتماد هيكلية الشبكة العصبية المعادة اعتماداً على (MLR) فسيتم اجراء عمليتي التدريب والاختبار للحصول على افضل التنبؤات وتسمى هذه تنبؤات الطريقة الهجينة للشبكات العصبية RNN-Wavelet.
3. سيتم اجراء تحسينات على مخرجات الشبكة العصبية لتقليل اخطاء التنبؤ والحصول على تنبؤات ادق من خلال ترشيحها باستخدام طريقة الموجات .

ان المخرجات النهائية لطريقة الموجات من الممكن ان تسمى تنبؤات الطريقة الهجينة للموجات Wavelet والشبكات العصبية المعادة RNN (RNN-Wavelet) بالاعتماد على نموذج الانحدار الخطي المتعدد MLR لبيانات التدريب والاختبار .

3. مقاييس خطأ التنبؤ Forecasting Error Measurements

سيتم استخدام متوسط النسبة المئوية المطلقة للخطأ (MAPE) Mean Absolute Percentage Error للمقارنة بين الطرق المستخدمة. ويعرف الخطأ بأنه هو تقدير للفرق بين القيمة الحقيقية والقيمة المقدرة حيث كلما كان الخطأ قليل فستكون الدقة اكبر. وبحسب مقياس الخطأ MAPE وفق الصيغة الرياضية الآتية (Khair et al., 2017):

$$MAPE = \frac{1}{n} \left[\sum_{i=1}^n \left| \frac{e_i}{y_i} \right| \right] \times 100 \quad (9)$$

حيث e_i : تمثل خطأ التنبؤ، n : هي عدد المشاهدات، $i = 1, 2, 3, \dots, m$

y_i : هو السلسلة الحقيقية او الاصلية المستعملة كمتغير هدف. وبحسب خطأ التنبؤ كما يلي:

$$e_i = y_i - \hat{y}_i \quad (10)$$

4. النتائج والمناقشات

تم استخدام طريقة احصائية تقليدية شائعة الاستخدام متمثلة بنموذج MLR للتنبؤ بالجسيمات المعلقة بالهواء المتمثل بمقياس (PM₁₀). وكذلك تم استخدام طريقة اخرى لتحسين التنبؤ بتلوث الهواء لمتغير PM₁₀ بعد تهجينها مع الطريقة التقليدية. وقد تمت مقارنتها مع الانحدار الخطي المتعدد MLR في حالة البيانات الكلية وفي حالة البيانات المترافقة زمنياً ورغم ما يوفره الانحدار الخطي المتعدد من امكانية نمذجة البيانات لاستخدامه في التنبؤ بالمتغير المعتمد إلا أن بيانات التلوث الجوي والأرصاء الجوية تأخذ نمطاً غير خطياً مما يؤدي إلى ظهور بعض النتائج والتنبؤات بدقة قليلة أحياناً وقد حسنت الطريقة الهجينة المستخدمة من نتائج التنبؤ من خلال ما تضمنته من حلول وامكانات للتعامل مع البيانات غير الخطية. (Jahandideh et al., 2009). تم تقسيم البيانات الى قسمين قسم تدريب وقسم اخر للاختبار وكما مدرج ادناه:

1. بيانات التدريب : ستبدأ من الفترة (1/1/2013) الى (31/5/2015)

2. بيانات الاختبار: ستبدأ من الفترة (1/6/2015) الى (31/10/2015)

وبهذا سيكون هناك (881) مشاهدة للتدريب و(153) مشاهدة للاختبار وسيتم استخدام طريقة الانحدار MLR حيث تم حذف المتغيرات غير المعنوية اعتماداً على قيمة P-value كما سنشاهده في النتائج ادناه وقد تم اختيار افضل معادلة نموذج انحدر خطي متعدد اعتماداً على افضل تفسير للعلاقة والتأثير ومعنوية المعلمات حيث ان افضل نموذج انحدر خطي متعدد في حالة البيانات الكلية كما مدرج ادناه:

$$y = 79.01x_1 + 3501x_2 - 2029x_3 + 197.6x_4 \quad (11)$$

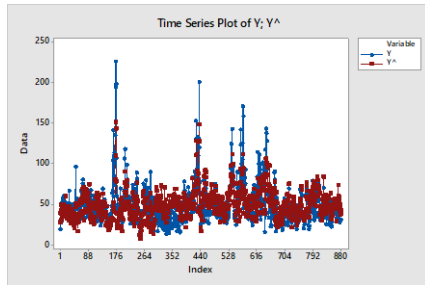
حيث يمكن تعريف متغيرات ومعلمات نموذج الانحدار الخطي المتعدد في (11) كما هو ادناه:

y : هو متغير الاستجابة PM₁₀ و x₁ : يمثل متغير احادي اكسيد الكربون CO و x₂ : يمثل متغير ثنائي اكسيد الكبريت SO₂ و x₃ : يمثل متغير احادي اكسيد النيتروجين NO و x₄ : يمثل متغير الاوزون O₃. نلاحظ من خلال المعادلة (11) التي تمثل نموذج الانحدار الخطي المتعدد ان معاملات المتغيرات التفسيرية β₁, β₂, β₃, β₄ تساوي (79.01)، (3501)، (-2029)، (197.60) على التوالي والتي تمثل قيمتها قوة تأثير المتغير التفسيري على المتغير المعتمد بينما توضح الاشارة اتجاه ذلك التأثير.

الجدول(1): يوضح معاملات نموذج الانحدار ومعنوياتها

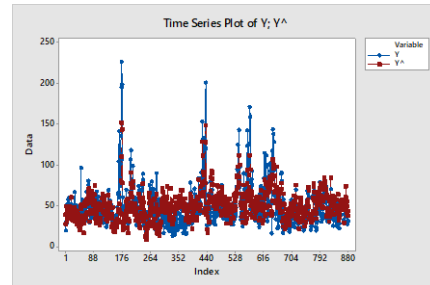
TERM	COEFF	T-value	P-value
SO ₂	3501.00	7.27	0.000
NO	-2029.00	-15.61	0.000
O ₃	197.60	2.97	0.003
CO	79.01	25.82	0.000

حيث ان جميع المعلمات المقدره معنوية لان قيم p-value هي اقل من مستوى المعنوية(0.05) مما يدل على معنوية المعلمات وكفاءة النموذج ولذلك فيعتبر النموذج الامثل.



الشكل (4): التطابق بين البيانات الاصلية مع البيانات التقديرية خلال الفترة الكلية لبيانات التدريب بطريقة PM₁₀(سلسلة التنبؤ)

MLR.



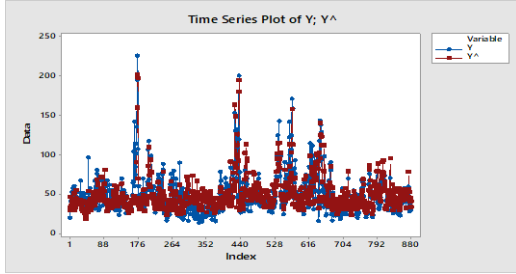
الشكل (5): التطابق بين البيانات الاصلية مع البيانات التقديرية خلال الفترة الكلية لبيانات الاختبار بطريقة PM₁₀ (سلسلة التنبؤ)

MLR.

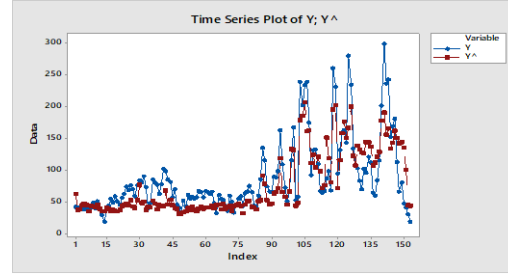
في هذه الدراسة تم استخدام طريقة احصائية تقليدية متمثلة بنموذج MLR للتنبؤ بالجسيمات المعلقة بالهواء المتمثل بمقياس (PM_{10}) . وكذلك تم اقتراح طريق اخرى لتحسين التنبؤ بتلوث الهواء لمتغير PM_{10} . وقد تمت مقارنة الطرق المستخدمة وهما الطريقتان الهجيتان MLR-RNN و RNN-Wavelet مع نموذج الانحدار الخطي المتعدد MLR.

الجدول(2): قيم(MAPE) في حالة البيانات الكلية للتنبؤ باستخدام RNN

بيانات التدريب	بيانات الاختبار
22.0665	29.6825



الشكل(6): التطابق بين البيانات الاصلية مع البيانات التقديرية (سلسلة التنبؤ) (PM_{10}) خلال الفترة الكلية لبيانات التدريب باستخدام MLR-RNN .



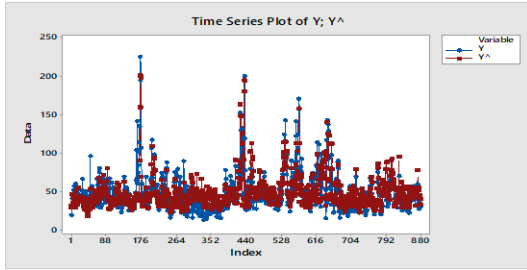
الشكل(7): التطابق بين البيانات الاصلية مع البيانات التقديرية (سلسلة التنبؤ) للمتغير المعتمد (PM_{10}) خلال الفترة الكلية لبيانات الاختبار باستخدام MLR-RNN .

يلاحظ من الجدولين (1 و 2) والأشكال (6 و 7) ان هنالك تقارب وانسجام بين القيم الحقيقية والقيم المقدرة التنبؤية باستخدام MLR-RNN وبصورة اكثر انسجاما مما كانت عليه في نموذج MLR وهذا يدل على ان الطريقة الهجينة RNN-MLR افضل من الطريقة التقليدية وهي نموذج MLR. وبعد ان تم ايجاد قيم (MAPE) للشبكات العصبية المعادة فسيتم اختيار الشبكة التي تحتوي على اقل قيم للخطأ ومن ثم استخدام التحليل المويجي لاجراء تحسينات على مخرجات الشبكة العصبية المعادة والتي تمت تسميتها بتنبؤات الطريقة الهجينة MLR-RNN لتقليل اخطاء التنبؤ والحصول على تنبؤات افضل مما كانت عليه من خلال استخدام مرشحات تحليل المويجات وباستخدام برنامج Matlab والادوات الجاهزة Toolbox الخاصة بالمويجات. ان المخرجات النهائية لتحليل المويجات من الممكن ان تسمى تنبؤات الطريقة الهجينة للمويجات Wavelet والشبكات العصبية المعادة RNN (RNN-Wavelet) بالاعتماد على نموذج الانحدار الخطي المتعدد وليبيانات التدريب والاختبار. الجدول(3) ادناه يبين قيم مقياس الخطأ MAPE للطريقة الهجينة RNN-Wavelet بالاعتماد على نموذج MLR.

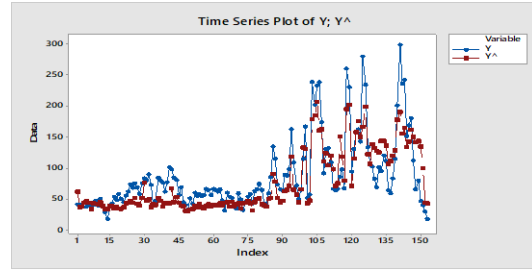
الجدول(3): قيم (MAPE) للطريقة الهجينة RNN-Wavelet بالاعتماد على MLR

	MLR	MLR-RNN	RNN-Wavelet
تدريب	27.2236	22.0665	20.3933
اختبار	28.0112	29.6825	28.9185

يلاحظ من خلال الجدول (3) تفوق الطريقة الهجينة RNN-Wavelet على الطريقة الهجينة MLR-RNN حيث انها اعطت اقل قيم للخطأ مقارنة مع الطرق ال اخرى. الشكل (8) والشكل (9) يوضحان مدى التطابق بين البيانات الاصلية مع البيانات التقديرية (سلسلة التنبؤ) للمتغير المعتمد (PM_{10}) لبيانات التدريب من (2013/1/1) الى (2015/5/31) وليبيانات الاختبار من (2015/6/1) الى (31/10/2015) على التوالي باستخدام الطريقة الهجينة RNN-Wavelet.



الشكل (8): التطابق بين البيانات الاصلية مع البيانات التقديرية (سلسلة التنبؤ) للمتغير المعتمد (PM_{10}) لبيانات التدريب باستخدام RNN-Wavelet الهجينة.



الشكل (9): التطابق بين البيانات الاصلية مع البيانات التقديرية (سلسلة التنبؤ) للمتغير المعتمد (PM_{10}) لبيانات الاختبار باستخدام RNN-Wavelet الهجينة.

يلاحظ من الشكلين (8 و9) ومقارنات الجدول (3) ان هناك تقارب وانسجام اكبر بين القيم الحقيقية والقيم المقدرة اكثر مما كان عليه في الطرق الاخرى وكذلك نتائج ادق للطريقة الهجينة RNN-Wavelet مما يدل على جودتها مقارنة بطريقة MLR-RNN الهجينة ونموذج .MLR

5. الاستنتاجات

من خلال ما تم عرضه من نتائج ومناقشات للطرائق المستخدمة وتطبيقها على بيانات الدراسة فان استخدام التحليل الموجي Wavelet بعد تهجينه مع الشبكات العصبية المعادة RNN اعتمادا على نموذج الانحدار الخطي المتعدد MLR يعد كأفضل اسلوب فعال للحصول لترشيح النتائج والحصول على افضل دقة في نتائج التنبؤ مع البيانات غير الخطية متعددة المتغيرات عندما يكون PM_{10} كمتغير معتمد. وكذلك فإنه من الممكن استخدام الطريقة الهجينة MLR-RNN لتحسين نتائج التنبؤ مقارنة بالطرق التقليدية.

6. Reference

1. Abrougui, K., Gabsi, K., Mercatoris, B., Khemis, C., Amami, R., & Chehaibi, S. (2019). Prediction of organic potato yield using tillage systems and soil properties by artificial neural network (ANN) and multiple linear regressions (MLR). *Soil and Tillage Research*, 190, 202-208 .
2. Ahmad, M., Alam, K., Tariq, S., Anwar, S., Nasir, J., & Mansha, M. (2019). Estimating fine particulate concentration using a combined approach of linear regression and artificial neural network. *Atmospheric Environment*, 117050, 219 .
3. Dawson, C., & Wilby, R. (2001). Hydrological modelling using artificial neural networks. *Progress in physical Geography*, 25(1), 80-108 .
4. Honarasa, F., Yousefinejad, S., Nasr, S., & Nekoeinia, M. (2015). Structure–electrochemistry relationship in non-aqueous solutions: predicting the reduction potential of anthraquinones derivatives in some organic solvents. *Journal of Molecular Liquids*, 212, 52-57 .
5. Jahandideh, S., Jahandideh, S., Asadabadi, E. B., Askarian, M., Movahedi, M. M., Hosseini, S., & Jahandideh, M. (2009). The use of artificial neural networks and multiple linear regression to predict rate of medical waste generation. *Waste management*, 29(11), 2874-2879 .
6. Janssen, N. A., Hoek, G., Simic-Lawson, M., Fischer, P., Van Bree, L., Ten Brink, H., . . . Brunekreef, B. (2011). Black carbon as an additional indicator of the adverse health effects of airborne particles compared with PM10 and PM2.5. *Environmental health perspectives*, 119(12), 1691-1699 .
7. Khair, U., Fahmi, H., Al Hakim, S., & Rahim, R. (2017). *Forecasting error calculation with mean absolute deviation and mean absolute percentage error*. Paper presented at the Journal of Physics: Conference Series.
8. Lin, L., Dekkers, I. A., Tao, Q., & Lamb, H. J. (2020). Novel artificial neural network and linear regression based equation for estimating visceral adipose tissue volume. *Clinical Nutrition* .
9. Mandrikova, O., Polozov, Y., & Geppener, V. (2017). Method of ionospheric data analysis based on a combination of wavelet transform and neural networks. *Procedia engineering*, 201, 756-766 .
10. Okkan, U. (2012). Wavelet neural network model for reservoir inflow prediction. *Scientia Iranica*, 19(6), 1445-1455 .
11. Palit, A. K., & Popovic, D. (2006). *Computational intelligence in time series forecasting: theory and engineering applications*: Springer Science & Business Media.

12. Sheela, K. G., & Deepa, S. N. (2013). Review on methods to fix number of hidden neurons in neural networks. *Mathematical Problems in Engineering*, 2013 .
13. Shrestha, R. R., Theobald, S., & Nestmann, F. (2005). Simulation of flood flow in a river system using artificial neural networks. *Hydrology and Earth System Sciences Discussions*, 9(4), 313-321 .
14. Torkashvand, A. M., Ahmadi, A., & Nikraves, N. L. (2017). Prediction of kiwifruit firmness using fruit mineral nutrient concentration by artificial neural network (ANN) and multiple linear regressions (MLR). *Journal of integrative agriculture*, 16(7), 1634-1644 .
15. Vlachogianni, A., Kassomenos, P., Karppinen, A., Karakitsios, S., & Kukkonen, J. (2011). Evaluation of a multiple regression model for the forecasting of the concentrations of NO_x and PM₁₀ in Athens and Helsinki. *Science of the total environment* .1571-1559 ,(8)409 .
16. Yonaba, H., Anctil, F., & Fortin, V. (2010). Comparing sigmoid transfer functions for neural network multistep ahead streamflow forecasting. *Journal of Hydrologic Engineering*, 15(4), 275-283 .
17. Zhang, J., Zhang, X., Niu, J ,Hu, B. X., Soltanian, M. R., Qiu, H., & Yang, L. (2019). Prediction of groundwater level in seashore reclaimed land using wavelet and artificial neural network-based hybrid model. *Journal of Hydrology*, 577, 123948 .
18. Zhou, F., Liu, B., & Duan, K .(2020) .Coupling wavelet transform and artificial neural network for forecasting estuarine salinity. *Journal of Hydrology*, 125127 .

Forecasting of air pollution data using the RNN-Wavelet hybrid method based on the MLR model

Khitam Walid Qader Osama Bashir Shukr
College of Computer and Mathematical Science, University of Mosul

Abstract

Particular matter (PM₁₀) studying and forecasting is necessary to control and reduce the damage of environment and human health. There are many pollutants as sources of air pollution may effect on PM₁₀ variable. This type of dataset can be classified as nonlinear. Studied datasets have been taken from climate station in Malaysia. Multiple linear regression (MLR) is used as linear statistical method for PM₁₀ forecasting through its influencing by corresponding climate variables, therefore it may reflect inaccurate results when used with nonlinear datasets. To improve the results of forecasting, recurrent neural network (RNN) has been suggested to be used after combining with MLR in hybrid in this study. Wavelet analysis is proposed filtering the result of MLR-RNN method for more improving of forecasting results through RNN-Wavelet hybrid method based on MLR model. In general, the best results of forecasting were for using RNN-Wavelet method. . In addition, the results of hybrid methods were outperformed comparing to MLR model as traditional method. As conclusion in this study, Wavelet analysis can be used after hybridizing with RNN based on MLR as active approach to obtain better forecasting results with nonlinear datasets in which PM₁₀ is to dependent variable.

Keywords: Multiple linear regression (MLR), Recurrent neural network (RNN), Wavelet, Particular Matter (PM₁₀), Forecasting, Hybrid MLR-RNN, Hybrid RNN-Wavelet.