

استخدام نموذج الانحدار المتعدد واسلوب التراصف الزمني للتنبؤ

بيانات PM10

Using the multiple linear regression model and the time-stratified approach to PM10

الباحثة / فنار عبد الرزاق محمد
كلية التربية للعلوم الصرفةأ.م.د. اسامة بشير شكر حنون
كلية الحاسوب والرياضيات

جامعة الموصل

المستخلص

ان نمذجة جودة الهواء اكتسبت أهمية كبيرة في تلوث الهواء الجوي بسبب الآثار السلبية على البيئة وصحة الإنسان مما دفع الكثير من الباحثين للنظر في ذلك الموضوع. وفي دراستنا هذه تم التطرق الى ملوثات الهواء التي كان لها تأثير مباشر على بيانات PM10 في الهواء وتم الوصول الى عدد كبير من المتغيرات وهي (O₃, CO, SO₂, NO, NOX) وكذلك سرعة الرياح ودرجة حرارة الهواء وغيرها). حيث تم استخدام تسعة متغيرات تفسيرية من بيانات الارصاد الجوية (Metrological) على مدى ثلاث سنوات في تطبيق نماذج الانحدار الخطي المتعدد. حيث كان تحليل البيانات لهذه الفترة صعب النكهن مما ادى الى وجوب النظر في البيانات على انها بيانات ذات أنماط اسبوعية و فصول موسمية و بذلك تم التوصل الى اسلوب التراصف الزمني (time-stratified approach (TS)) ليستخدم في هذه الدراسة لتتراصف البيانات في نمطين النمط الاسبوعي ونمط الفصول الموسمية. وتم الاخذ بالحسبان استخدام اسلوب التراصف الزمني من خلال ترتيب البيانات الى مجموعات اكثر تجانسا قد يحل صعوبة الوصول إلى نتائج أكثر دقة. ان نموذج الانحدار الخطي المتعدد (Multiple Linear Regression (MLR) وهو من الطرق الأكثر شيوعا لدراسة مثل هكذا عدد كبير من المتغيرات. ففي هذا البحث استخدمت بيانات الأرصاد الجوية الماليزية وعلى مدى ثلاث سنوات ولوحظ ان نتائج التكنهن بطريقة MLR من افضل الطرق المستخدمة لمثل هكذا بيانات، وان مرادفة البيانات باستخدام اسلوب التراصف الزمني تبعا للنمط الاسبوعي ونمط الفصول الموسمية كان افضل و اكثر دقتا في اظهار النتائج عند تطبيق MLR مقارنة مع البيانات الاصلية.

ABSTRACT

Air Quality Modeling gained great importance in atmospheric pollution because of its negative effects on the environment and human health. In our study, the rate of the PM10 has been effected by a large number of elements such as (O₃,CO₂,SO₃,NO,NOX). It is also effected by other explanatory variables such as wind speed, temperature and others. Nine independent variables have been used in the application of the multiple



regression model over three years of metrological datasets. The seasonal influences for weekly, seasonally periods lead up to difficult analyzing and forecasting. Classification the datasets into weekly and seasonally then using Time-stratified (TS) approach and taking only the compatible observations may solve the difficulty and lead to more accurate results. Reducing the number of variables may also lead to more accurate results. The multiple linear regression (MLR) model is the most common for studying like this number of variables.. Three years of Malaysian meteorological datasets are studied in this research. It was noted that the results of prediction in a method MLR one of the best methods to use such data. Appling the Time-stratified approach for PM10 data according to the weekly and seasonally pattern were better and more accurate for MLR comparing to the original data.

1. المقدمة:

يعرف الانحدار الخطي المتعدد MLR بأنه احد الاساليب الاحصائية المتقدمة للتحقق من العلاقة بين المتغيرات، ويستخدم لتشكيل معادلة بين المتغير المعتمد والعديد من المتغيرات التفسيرية و عادة يستخدم لشرح مكونات سجلات الأرصاد الجوية، اذ تكون بيانات الارصاد الجوية كبيرة الحجم ان كانت يومية أو أسبوعية أو أنماط موسمية اخرى، او بيانات متعددة المحطات مما يؤدي الى مواجهة صعوبات للتنبؤ بالبيانات المستقبلية عند استخدام عينة كبيرة الحجم. حيث استخدم بعض الباحثين عينة من البيانات لمدة ست سنوات أو أكثر. واستخدم آخرون بيانات من محطات متعددة أو مدن متعددة لإجراء مقارنة بين النتائج واختيار الأفضل (Güler & Güneri İşçi, 2016). فضلا عن استخدام الأنماط الموسمية أو الأسبوعية ومواجهة الآثار السلبية على البيانات ونتائج التحليل. حيث استخدم الاساليب التقليدية لتحليل السلاسل الزمنية والتكهن بالبيانات المستقبلية سوف لن يعطي الدقة الكافية المطلوبة، الآثار السلبية لأنماط البيانات الأسبوعية والموسمية في مجموعات البيانات ستعكس سلبا على النتائج. في هذا البحث هدفنا هو تحسين مستوى التكهن مقارنة مع الطرق التقليدية من خلال اقتراح MLR، و تجنب الآثار السلبية للنتائج الأسبوعية والموسمية وتحليل النتائج من خلال استخدام طريقة التراصف الزمني.

تم اقتراح اسلوب التراصف الزمني في العديد من الدراسات لاختيار فترات طويلة من الايام المسيطر عليها او الايام المثالية بسبب بعض الأحداث المتكررة في كل عام، و تم العمل على تراصفها في سلسلة أحادية المتغيرات لتجنب التحيز. لكي لا يكون هناك أي التباس او (خلط او تشويش) من خلال الخصائص المتباينة في مجموعة البيانات كما اقترحه العديد من الباحثين من قبل (Bateson & Schwartz, 2001; Levy et al., 2001).

يتم استخدام اسلوب التراصف الزمني للحصول على نمطين من البيانات اولها النمط المتراصف اسبوعيا وثانيها نمط الفصول الموسمية، اذ نظمت البيانات الاسبوعية على اساس اسلوب تراصف البيانات باختيار يوم العطلة من كل أسبوع وتم تحديد يوم السبت وهو اليوم الأكثر تأثيرا على بيانات PM10 وهو يوم العطلة المشترك لجميع الولايات الماليزية. اما انماط الفصول الموسمية فتم مراصفة البيانات للمواسم نفسها من كل عام وتكوين عينة طبقية لنفس المواسم من السنوات المقبلة حيث ان كل عام يحتوي على اربعة مواسم كل موسم يتكون من ثلاثة اشهر.

تحليل الانحدار الخطي المتعدد MLR يدرس العلاقة بين المتغير المعتمد و العديد من المتغيرات التي تؤثر على المتغير قيد الدراسة، فيجب تحديد المتغيرات ذات العلاقة الحقيقية مع ذلك المتغير واستبعاد المتغيرات الاخرى. قبل استخدام تحليل MLR يجب التأكد من عدم وجود علاقة خطية بين المتغيرات وفي حالة وجودها فيجب معالجة ذلك بالتخلص من بعض المتغيرات المترابطة. اذ عند وجود ارتباط تام بين متغيرين فيجب استبعاد احدهما والاقبل اهمية في التأثير على المتغير المعتمد كما اقترحها (Sousa et al., 2007).

وكذلك اقترح الباحث (Stadlober et al., 2008) ان نموذج الانحدار هو نموذج للمتوسطات اليومية لبيانات PM10 من خلال النموذج الخطي الذي سيكون الاساس لنماذج التكهن، و تعد نماذج الانحدار هي التحول الجذري للمتغير المعتمد الضروري لضمان تباين الخطأ المستمر وتجنب اخطاء النماذج.

واستخدم MLR لدراسة بيانات الارصاد الجوية وايجاد افضل النماذج التي تعبر عن العلاقة بين المتغير المعتمد وعدد من المتغيرات التفسيرية كما في (Vlachogianni *et al.*, 2011). وكذلك استخدمت تقنيات MLR لتطوير النماذج الخطية التي يكون فيها PM10 هو المتغير المعتمد في حين أن جميع المتغيرات اليومية التي تؤثر في نسبة PM10 من متغيرات الأرصاد الجوية تكون متغيرات مستقلة تفسيرية. نماذج الانحدار الخطية في هذه الحالة، هي طرق أساسية تقليدية للمقارنة مع الطرق الأخرى كما اشار لها الباحث في (Taşpınar & Bozkurt, 2014). واقترح الباحثين اسلوب التراصف الزمني لتحديد ايام السيطرة من الشهر او العام للحصول على بيانات اكثر تجانساً وترتيباً للوصول الى نتائج ادق كما اشار له الباحث في (Levy, 2001) وكذلك تم توظيفه من قبل عدد من الباحثين لاختيار الايام والاشهر التي تضم المواسم التي سيتم اعتمادها في الدراسة ومنهم (Bateson, 2001; Tobias, 2014).

2. اسلوب البحث

يستعرض هذا البحث بشكل مفصل كيفية الحصول على الانحدار الخطي المتعدد للنموذج الخاص بالنمط الاسبوعي (عطلة نهاية الاسبوع) و النماذج الخاصة بنمط الفصول الموسمية المتراصفة الاربعة، وسيتم عرض الانحدار الخطي المتعدد MLR و نماذجه للحصول على نتائج وتنبؤات دقيقة.

2.1. الانحدار الخطي المتعدد

يعرف الانحدار الخطي بانه أسلوب يمكن بواسطته تقدير قيمة أحد المتغيرين بمعلومية قيمة المتغير الآخر عن طريق معادلة الانحدار ويعبر عن العلاقة بمعادلة او نموذج يربط المتغير المعتمد بمتغير واحد او اكثر من المتغيرات التفسيرية وله أنواع ومنها الانحدار الخطي البسيط والانحدار الخطي المتعدد، الانحدار الخطي البسيط فكلمة " بسيط " تعني أن المتغير المعتمد Y يعتمد على متغير تفسيري مستقل واحد وهو X، وكلمة " خطي " تعني أن العلاقة بين المتغيرين (Y, X) علاقة خطية ويعد الانحدار الخطي المتعدد تعميم للانحدار الخطي البسيط، فكلمة بسيط تشير إلى وجود متغير مستقل واحد في النموذج، لذلك فإن كلمة متعدد تشير إلى وجود متغيرات تفسيرية عدة في النموذج والتي يعتقد أنها تؤثر في المتغير المعتمد (Nathans *et al.*, 2012).

يعرف الانحدار الخطي المتعدد MLR بانه أداة إحصائية تسمح بدراسة كيفية ارتباط المتغيرات التفسيرية المتعددة لمتغير معتمد واحد وبمجرد أن تحدد كيف ترتبط هذه المتغيرات المتعددة لمتغير المعتمد يمكن أن تأخذ معلومات عن جميع المتغيرات التفسيرية واستخدامها لجعل التكهانات أكثر قوة ودقة، الانحدار المتعدد يعبر عن العلاقة بين المتغيرات المستخدمة لتقدير القيم السابقة والتكهن بالقيم المستقبلية، ويحتوي على عدد كبير من المتغيرات التفسيرية التي قد تؤثر على متغير معتمد واحد قيد الدراسة، و يتم استخدامه للتنبؤ بالمتغير المعتمد الذي يؤثر عليه المتغيرات التفسيرية، و كثيرا ما تكون الحاجة للانحدار الخطي المتعدد لتحديد العوامل التي تؤثر في متغير ما، مثل تحديد العوامل التي تؤثر في بيانات PM10 أو التي تؤثر في ارتفاعه و انخفاضه عند حدوث تغيرات في الطقس التي تؤثر على بعض العوامل التي يتم رصدها من خلال الارصاد الجوية. وفي معظم الأحيان يكون لدينا عوامل عدة تؤثر في ذلك المتغير الذي يراد دراسته فلا يمكن الاعتماد على تحليل الانحدار الخطي البسيط، في حين ان الانحدار الخطي المتعدد يساعد على دراسة تأثير هذه العوامل على المتغير محل الدراسة مثل دراسة تأثير كل من سرعة الرياح و الرطوبة ودرجة حرارة الهواء وغيرها من العوامل التي تؤثر على قيم PM10 (Abderrahim *et al.*, 2016; Volume 40, 2006).

ويتطلب الانحدار المتعدد عددا كبيرا من الملاحظات أي يجب أن يتجاوز عدد البيانات الملاحظة بشكل كبير عدد المتغيرات التفسيرية التي تستخدم في الانحدار، كما يتطلب تحليل الانحدار الخطي المتعدد حجم عينة كبير. وهناك طرق عدة لتقييم تأثير المتغيرات التفسيرية في المتغير المعتمد ويمكن توضيح هذه الطرائق على النحو الآتي:

- الطريقة المتأنيبة ويحدد فيها الباحث مجموعة المتغيرات التفسيرية التي تشكل النموذج ويتم تقييم النموذج على هذا الأساس في التكهن بالمتغير المعتمد.
- الطريقة الهرمية وفيها يتم ادخال المتغيرات الى النموذج طبقا لترتيب محدد وهذا الترتيب يعكس بعض الاعتبارات النظرية او النتائج السابقة وإذا لم يكن هناك مثل هذه الاسس او النتائج التي يستند اليها الباحث في ترتيب دخول المتغيرات فلا يمكن الاعتماد على هذه الطريقة وبالطبع يتم تقييم مستوى تأثير كل متغير عند ادخاله الى النموذج وإذا لم يكن تأثيره له دلالة جوهرية في زيادة التكهن بالنموذج فيتم حذفه.
- الطرائق الاحصائية في هذه الطرائق فان ادخال المتغيرات او حذفها من النموذج يتحدد بناء على مقدار قوة الارتباط بينها وبين المتغير المعتمد وهناك اكثر من نوع في هذه الطريقة يطلق عليها الاختيار المتقدم، الاختيار الراجع، الاختيار متعدد الخطوات.

ولا يوجد ما يمنع من اتباع أسلوب وسط وهو أن يؤخذ في الحسبان بعض العوامل التي يكون لها تأثير قوي على المتغير المعتمد، ثم بعد ذلك نبدأ في إضافة المتغيرات الأخرى تباعا، عموما فان الاختيار بين هذه الطرائق لا يمثل مشكلة فكلها تؤدي في النهاية لنفس النتيجة. ويتم تحديد المتغير الذي يستبعد حيث يتم الاهتمام بقيمة P-Value لأنها تعني ما إذا كان هذا المتغير مؤثرا أم لا. ففي حالة الانحدار المتعدد يؤخذ في الحسبان قيمة P-Value فيتم حذف المتغير الذي له قيمة P-Value كبيرة وخاصة تلك التي تتجاوز 0.05

لتقييم نموذج الانحدار الخطي المتعدد في حالة الانحدار البسيط يعتمد على معامل التحديد R^2 ولكن في حالة الانحدار المتعدد فيستخدم معامل التحديد المعدل R^{-2} أي المعدلة. لأن قيمة معامل التحديد R^2 تزداد بشكل طبيعي كلما أضفنا متغيرا بمعنى أن قيمتها تدرس علاقة المتغير بمتغيرين ستكون أكبر منها عند استبعاد أحدهما. وهذا لا يساعد على معرفة ما إذا كان هذا المتغير الإضافي قد أفاد في التحليل أم لا. أما مع معامل التحديد المعدل R^2 فإن هذا لا يحدث لأن طريقة حسابه تأخذ في الحسبان عدد المتغيرات الداخلة في التحليل. لذلك لمعرفة إن كان إضافة متغير لها تأثير إيجابي على النموذج الرياضي (المعادلة التي تربط المتغير المعتمد بالمتغيرات التفسيرية) فيتم استخدام معامل التحديد المعدل R^{-2} . وهناك مقياس آخر أكثر دقة من معامل التحديد المعدل R^{-2} وهو (F Test)، وبدون الدخول في تفاصيل إحصائية فإن قيمة (F test) تزداد كلما تحسن النموذج وتقل كلما ساء النموذج أي أننا لو أضفنا متغيرا له علاقة إحصائية مؤثرة بالمتغير محل الدراسة فإن قيمة (F test) تزداد وهذا مشابه لما ذكرناه في R^{-2} غير أن قيمة (F test) لا تتراوح بين صفر و واحد بل تأخذ أي قيمة، والأمر المهم بمعرفة مدى دقة الدلالة لقيمة F test هو مشابهة لمعرفة تأثير أي متغير عن طريق P value. فإذا كانت الدلالة لقيمة F test أقل من أو تساوي 0.05 فإن هذا يعني أن النموذج مقبول إحصائيا وأما إذا زاد عن ذلك فإن النموذج يكون غير مقبول. حيث ان تحليل الانحدار الخطي المتعدد ليس طريقة واحدة فقط للتحليل بل هو عدد من الطرائق التي يمكن

استخدامها لمعرفة العلاقة بين المتغير المعتمد وعدد من المتغيرات التفسيرية (Stadlober *et al.*, 2008; Vlachogianni *et al.*, 2011).

وعادة ماتظهر ظاهرة الارتباط الذاتي في أغلب الدراسات التي تأخذ شكل السلاسل الزمنية وغالبا ما تكون المشاهدات المتتالية للسلسلة الزمنية غير التفسيرية، ولذلك يجب استخدام اساليب احصائية خاصة عند دراسة مثل هذه البيانات والتكهن بقيمها في المستقبل، و تظهر هذه المشكلة في البيانات المقطعية وخاصة التي تكون في شكل مجاميع واسباب هذه المشكلة عادة تكون ل

- حذف بعض المتغيرات التفسيرية من العلاقة المدروسة.
- التشخيص غير الدقيق للعلاقة بين المتغير المعتمد و المتغيرات التفسيرية.
- قد تكون هنالك عوامل عشوائية تؤثر على القيم المتتالية للبواري كما يحدث في حالات عدم الاستقرار والجفاف وغير ذلك من الحالات التي يمتد اثرها على مدى مشاهدات عديدة متعاقبة مما يحدث مشكلة ارتباط ذاتي بين أخطاء تلك المشاهدات

• إجراء تعديلات في البيانات او اللجوء الى تقدير قيم بعض المشاهدات اعتمادا على قيم مشاهدات اخرى.

في هذه الدراسة تم الكشف عن مدى وجود تعدد علاقة خطية بين المتغيرات التفسيرية من خلال حساب عامل التضخم للتباين (VIF) احد مؤشرات اختبار تعدد العلاقة الخطية، و يعبر تحليل عامل التضخم للتباين (VIF) عن مدى العلاقة بين المتغيرات التفسيرية في نماذج الانحدار الخطي المتعدد في تحليل الانحدار. حيث يعد تعدد العلاقة الخطية بين المتغيرات من المشاكل التي تواجه العديد من الباحثين لأنها يمكن أن تزيد من التباين في معاملات الانحدار، مما يجعلها غير مستقرة ويصعب تفسيرها. حيث يمثل ذلك العامل مقياس التباين لمعامل التضخم للانحدار مقارنة مع عدم وجود متغيرات مرتبطة بالتكهن. وهناك عدد من المعايير لقيم VIF لمعرفة مدى الارتباط بين المتغيرات فاذا كانت قيمة (VIF=1) فذلك يدل على عدم وجود ارتباط بين المتغيرات بشكل مطلق و اذا كانت قيمة (VIF> 1) >5) فذلك يعني انه هناك ارتباطا معتدلا اما عندما تكون قيمة (VIF>5) فذلك يعني ان هناك ارتباطا وثيقا غير مقبول بين المتغيرات التفسيرية. اما في حالة ان العلاقة المتبادلة المرتبطة لقيم VIF أكبر من 10 فتشير إلى ان تعدد العلاقة يؤثر بشكل غير ملائم على نتائج الانحدار. في هذه الحالة سوف يتم تقليل الارتباط عن طريق إزالة المسببات غير المهمة في النموذج (Dolatabadi *et al.*, 2010).

في هذا البحث تم إجراء تحليل الانحدار المتدرج لبيانات PM10 وتم التحقق من نتائج التعدد الخطي من خلال دراسة عامل التضخم للتباين (VIF) للمتغيرات التفسيرية (التكهنية) فتم الكشف عن وجود علاقة خطية بين متغيرات الدراسة حيث ان احد مسببات هذه العلاقة هو المتغير NO2 نتيجة لعدم ملائمة قيمه مع قيم بقية المتغيرات فكان الحل الامثل لهذه المشكلة بحذف المتغير من النماذج (Elbayoumi *et al.*, 2014)

واستخدم العديد من الباحثين الانحدار الخطي المتعدد في مجال تلوث الهواء ويمكن صياغة النموذج العام لمعادلة MLR على النحو الاتي:

$$Y = B_0 + B_1 X_1 + B_2 X_2 + \dots + B_i X_i + E \quad (1)$$

حيث ان Y هو المتغير المعتمد.

X_1, X_2, \dots, X_i هي المتغيرات التفسيرية.

هي معاملات الثابت او معاملات الانحدار B_0, B_1, \dots, B_i

E هو حد الخطأ لأقل تنبؤ رقم i حيث $i=1,2,\dots,n$ حيث ان عدد المشاهدات هو n يكون عدد المعادلات هو n (Elbayoumi et al., 2014) و يمكن صياغة المعادلة (1-2) في صورة مصفوفات كما يلي:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & X_{21} & X_{31} & \dots & X_{i1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{2n} & X_{3n} & \dots & X_{in} \end{bmatrix}, B = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_i \end{bmatrix}, E = \begin{bmatrix} E_1 \\ E_2 \\ \vdots \\ E_n \end{bmatrix} \quad (2)$$

حيث Y من درجة $(n \times 1)$ و المصفوفة X من درجة $(n \times (i+1))$ ودرجة B هي $(i \times 1)$ و درجة E هي $(n \times 1)$.

ويلاحظ أن العمود الاول في مصفوفة البيانات يحتوي على قيمة الواحد الصحيح عند كل المشاهدات من (1) إلى (n) وذلك لتقدير المعامل الثابت والعمود الثاني من المصفوفة يحتوي على قيم المتغير الاول (X_1)، وبذلك كل عمود يحتوي على قيم متغير تفسيري محدد، وباستخدام رموز المصفوفات يمكن اختصار كتابة نموذج الانحدار الخطي كما يلي:

$$Y = BX + E \quad (3)$$

ويمكن استخدام الانحدار الخطي المتعدد في حالة توافر الفروض الآتية :

1. أن تكون العلاقة خطية بين المتغيرات التفسيرية والمتغير المعتمد .
2. أن تكون البيانات موزعة توزيعاً طبيعياً للمتغيرات التفسيرية والمتغير المعتمد .
3. يجب أن تكون قيم المتغير المعتمد من المستوى الترتيبي على الأقل .
4. يوجد استقلال احصائي بين المتغيرات التفسيرية والخطأ العشوائي اي ان اعمدة المصفوفة مستقلة خطياً عن متجه الاخطاء العشوائية E .

2.2. اسلوب التراصف الزمني وبيانات البحث:

اسلوب التراصف الزمني هو أداة تحليلية لتقدير آثار محفزات النتائج من التعرض البيئي و يضمن تقديرات الانحدار اللوجستي المشروط غير المنحازة ويتجنب التحيز بسبب الاتجاه الزمني في سلسلة التعرض، ويمكن تكييف الطبقة المحددة للسيطرة على المربعات الزمنية المتغيرة حسب التصميم المستخدم. ويستخدم التراصف الزمني لتصميم المتراصفات المتقاطعة التطبيقية على نطاق واسع في علم الأوبئة البيئية لتحليل الآثار القصيرة الأجل لعوامل خطر البيئة، مثل تلوث الهواء أو درجة الحرارة، واثرها على صحة الإنسان. و اليوم الذي يحدث فيه الحدث الصحي (يوم الحالة) يدعى بيوم السيطرة المختار من الشهر نفسه من كل عام. و يمكن تطبيقه عموماً في سياقات أخرى ويعمل على الوصول الى بيانات اكثر تجانساً من

البيانات الكلية للوصول الى نتائج ادق (Malig et al., 2016; Tobias et al., 2014) ويمكن ايجاز الخطوات التي يتم بها الترافف الزمني بالنقاط الاتية:

1. الخطوة الاولى رسم البيانات بسلاسل اعتمادا على الاشهر المتكررة من كل عام ضمن الفترة الزمنية المحددة للبيانات.
2. الخطوة الثانية تحديد الفترات الثابتة للأنماط الاسبوعية والفصول الموسمية اعتمادا على النمط المحدد للبيانات وتحديد يوم متميز من كل اسبوع او فصل من كل عام.
3. الخطوة الثالثة بأخذ هذه الفترات من السلسلة و مرافقتها زمنيا لذلك اليوم المميز من كل اسبوع والوصول الى النمط الاسبوعي المترافف واخذ الفترات المحددة للفصول الموسمية من كل عام والوصول الى اربع فصول موسمية متراففة.
4. بمراففة البيانات بالنمط الاسبوعي او نمط الفصول الموسمية يتم التوصل الى سلسلة بيانات اكثر تجانسا من البيانات الاصلية.

في هذا البحث سيتم دراسة البيانات اليومية لعدد من ملوثات الهواء (CO, O3, SO2, NO, NOX) ، NO2، درجة حرارة الهواء، ... الخ) وتأثيرها على البيانات جسيمات PM10 في احدى الولايات الماليزية. اذ تم الكشف عن وجود مشكلة تعدد علاقة خطية بين المتغيرات وذلك بإيجاد عامل التضخم للتباين (VIF) والذي هو احد مؤشرات تعدد العلاقة الخطية وتم التوصل الى وجود علاقة ارتباط بين المتغيرات التفسيرية وتم الكشف عن المتغير المسبب لذلك الارتباط وهو المتغير NO2 وكان الحل الامثل لإنهاء هذه المشكلة التي كانت تؤثر في نتائج نماذج الانحدار الخطي المتعدد بحذف ذلك المتغير من جميع نماذج الانحدار (Sattari & Gharagheizi, 2008). في هذا البحث، يتم التكهّن بالبيانات اليومية الخاصة بقيم PM10 من محطة الأرصاد الجوية باستخدام نمطين من البيانات (أسبوعية و الفصول الموسمية)، حيث غطت هذه الدراسة السنوات الثلاث من 1 كانون الثاني 2013 إلى 31 تشرين الاول 2015 واستخدام جزء من البيانات للتدريب والباقي للاختبار و باستخدام أسلوب الترافف الزمني لأنماط البيانات المتراففة النمط الأسبوعي و نمط الفصول الموسمية، اذ تم التكهّن بعطلة نهاية الأسبوع و كان تركيز الاهتمام على ثلاثة أيام من الأسبوع هي الجمعة والسبت والأحد ايام العطل الاكثر شيوعا في الولايات الماليزية، ولكن يومي الجمعة والأحد لم تكن العطلة المشتركة في جميع الولايات، فكان من الافضل اعتبار عطلة نهاية الاسبوع هو يوم السبت لأنه يوم العطلة المشترك في جميع الولايات وبذلك تم تحديد نمط البيانات الاسبوعية، اما في نمط الفصول الموسمية فقد تم التوصل له بترافف انماط الفصول الموسمية على مدى ثلاث سنوات حيث يتكون كل عام من أربعة مواسم حيث كل موسم من ثلاثة أشهر. و أصبحت الفصول المتراففة زمنيا بالاعتماد على نمط الفصول الموسمية تتألف من تسعة اشهر (ثلاثة فصول موسمية) استخدمت ستة اشهر لكل مجموعة (فصلين موسمين) منها للتدريب وثلاثة اشهر (فصل موسمي واحد) للاختبار. باستثناء الفصل الموسمي الأول اذ تم اضافة شهري كانون الثاني 2013 و شباط 2013 كي لا يتم فقدان جزء من البيانات أو اهمالها فتألف الفصل الموسمي المترافف الاول من خمسة اشهر تدريب (كانون الثاني 2013 و شباط 2013 و كانون الاول 2013، كانون الثاني 2014 و شباط 2014) وثلاثة اشهر من البيانات للاختبار (كانون الاول 2014، كانون الثاني 2015 و شباط 2015). وتكون الفصل الموسمي المترافف الثاني من ستة أشهر (اذار 2013، نيسان 2013، ايار 2013 و اذار 2014، نيسان 2014، ايار 2014) تم استخدامهم للتدريب

وثلاثة أشهر (أذار 2015 و نيسان 2015 و أيار 2015) استخدمت للاختبار. وضم الفصل الموسمي المتراصف الثالث ستة اشهر (حزيران 2013، تموز 2013، اب 2013 و حزيران 2014، تموز 2014، اب 2014) تم استخدامهم للتدريب و ثلاثة أشهر (حزيران 2015 و تموز 2015 و اب 2015) استخدمت للاختبار. و اخرهم الفصل الموسمي المتراصف الرابع ضم ستة أشهر (أيلول 2013، تشرين الاول 2013، تشرين الثاني 2013 و أيلول 2014، تشرين الاول 2014، تشرين الثاني 2014) تم استخدامهم للتدريب و ثلاثة أشهر (أيلول 2015، تشرين الاول 2015، تشرين الثاني 2015) استخدمت للاختبار.

2.3: مقاييس الخطأ:

الخطأ يعني تقدير الفرق بين القيمة الحقيقية والقيمة المقدرة، وكلما قل الخطأ زادت الدقة في القياس، حيث \hat{Y} هي القيمة المقدرة ل Y التي هي القيمة الحقيقية المتمثلة بقيمة PM10 ولإيجاد قيمة الخطأ و قياس دقة النماذج سيتم استخدام بعض المقاييس لذلك (Wang *et al.*, 2015)، ويمكن إيجاد الخطأ بالمعادلة الآتية:

(4)

$$e = Y - \hat{Y}$$

حيث Y هي PM10 و \hat{Y} هي المقدرة.

وتم إيجاد مقاييس الخطأ (المقاييس الاحصائية RMSE, MSE, MAE) والتي بدورها تحدد نوعية النماذج الأقل خطأ. وسيجري استعراض القانون لكل مقياس:

MAE : هو متوسط الخطأ المطلق، الذي يقيس أداء متوسط الأخطاء المطلقة.

$$MAE = \frac{1}{N} \sum_{i=1}^N |e_i| \quad (5)$$

MSE : يقيس أداء مربع الخطأ ويسمى (متوسط مربع الخطأ).

$$MSE = \frac{1}{N} \sum_{i=1}^N e_i^2 \quad (6)$$

RMSE : هو مقياس لجذر متوسط مربع الخطأ الذي يقيس الاختلافات بين القيم التي يتنبأ بها النموذج .

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N e_i^2} \quad (7)$$

حيث e هو خطأ القيمة في الوقت i و N هو العدد الاجمالي للملاحظات (Cortina-Januchs *et al.*, 2015).

3. نتائج البحث

ان اولى الخطوات الرئيسية في عمل بعض البحوث هي الحصول على البيانات التي تتناسب مع الجانب العملي من الدراسة من جانب ودقتها من جانب اخر وهي الركيزة الأساسية للوصول إلى النتائج التي يمكن الاعتماد عليها في أي عمل بحثي، ولذلك كانت دراسة بيانات الأرصاد الجوية في ماليزيا التي غطت ثلاث سنوات من 2013-2015. وتم اختيار البيانات الأكثر تأثرا على بيانات PM10 ثم تحديد أهم الفترات في هذه السنوات الثلاث لتمثل بداية انفراجه في هذه الدراسة لتصبح مجموعات أكثر تجانسا وملائمة للدراسة. وتم اعتماد نوعين من الانماط في البيانات وهي الأسبوعية والموسمية واستخدامها في طريقة التراصف الزمني، وتمت مقارنة نتائج البيانات الكلية والبيانات المتراصفة للنمط الاسبوعي او نمط الفصول الموسمية وكذلك المقارنة بين الانماط الاسبوعية و الفصول الموسمية. بالنسبة للنمط الاسبوعي فقد تم تحديد اليوم الذي يظهر فيه انخفاض في قيم PM10 لأنه يوم العطلة المشتركة في جميع الولايات الماليزية وتم تحديد يوم السبت لمراصفته زمنيا مع جميع ايام السبت في السلسلة الزمنية. اما في نمط الفصول الموسمية المتراصفة فقد حددت أربع مجاميع متراصفة ايضا بالاعتماد على الفصول الموسمية في كل سنة ومراصفتها زمنيا مع جميع الفصول المشابهة في السلسلة الزمنية ، وفي هذا الفصل سيتم عرض نتائج التكهينات باستخدام مقاييس خطأ التكهين MAE و RMSE لطريقة MLR المستخدمة في هذا البحث وكذلك مناقشتها، وتمثيل الاتساق بمخططات توضيحية بين سلسلة الدالة الاصلية PM10 وسلسلة التكهين Y باستخدام الطريقة المذكورة، حيث تم تنفيذ الجانب التطبيقي من الدراسة باستخدام البرامج Minitab, MATLAB, Excel لتنفيذ الاوامر التي تخدم هذه الدراسة.

اذ صممت هذه الدراسة للتنبؤ بقيم PM10 في الغلاف الجوي وبعض ملوثات الهواء التي تؤثر في بيانات PM10 وهي (CO, O3, SO2, NO, NOX) ودرجة حرارة الهواء والرطوبة وسرعة الرياح بمقياسين 10m و xxm) حيث تغطي البيانات اليومية لمدة ثلاث سنوات من 1 كانون الثاني 2013 الى 31 تشرين الاول 2015 في احدى الولايات الماليزية، باستخدام طريقة التراصف الزمني. وسيتم حساب نماذج MLR باستخدام احدى البرامج (SPSS او Minitab) اما في هذه الدراسة تم استخدام برنامج Minitab لإيجاد نماذج الانحدار و تم الحصول على النموذج التالي للبيانات الكلية.

$$PM10 = - B0 + B1 CO - B2 O3 + B3 SO2 + B4 Nox - B5 NO + B6 AT - B7 H + B8 WS10m - B9 WSxxm$$

(8)

اذ ان (سرعة الرياح: WS10m, WSxxm, درجة الحرارة: AT, الرطوبة: H) .
 و Y هو PM10 و (X₁, X₂, X₃, X₄, X₅, X₆, X₇, X₈, X₉) هي CO, O3, SO2, NOX, NO, درجة حرارة الهواء و الرطوبة وسرعة الرياح بالمقياسين (10m, xxm) على التوالي، و كذلك B0, B1,.....,B9 معاملات الثوابت للمتغيرات التفسيرية، و سيتم استخدام برنامج Minitab لإيجاد نماذج الانحدار الخطي المتعدد، والذي يتضمن عددا كبيرا من البيانات والمتغيرات التي يصعب حلها دون استخدام الحاسوب. وتم الحصول على النموذج الأول الذي يمثل الانحدار للبيانات الكلية في النموذج أدناه، والذي تم إيجاد المقاييس الإحصائية له

$$PM10 = -206 + 122 CO - 21.8 O_3 + 368 SO_2 + 569 NO_x - 2255 NO - 5.63 AT - 0.0185 H + 4.68 WS_{10m} - 0.0010 WS_{xxm}$$

(9)

وتم استخدام جزء من البيانات الكلية لجعلها بيانات تدريب وإيجاد نماذج الانحدار لهذه البيانات ليتم إيجاد أفضل النماذج والجزء الباقي للاختبار للتأكد من قوة التكهّن بالنماذج. في هذه الدراسة تم أخذ بيانات التدريب بدء من 1 كانون الثاني 2013 إلى 30 نيسان 2015 والجزء المتبقي من البيانات المستخدمة للاختبار بدء من 1 ايار 2015 إلى 31 تشرين الأول ثم وجد نموذج الانحدار الخطي المتعدد للبيانات الكلية الخاصة بالتدريب كما في النموذج التالي (Dolatabadi *et al.*, 2010).

$$PM10 = -152 + 110 CO + 55.8 O_3 + 2336 SO_2 + 397 NO_x - 1866 NO + 4.08 AT - 0.108 H + 4.25 WS_{10m} + 0.0048 WS_{xxm}$$

(10)

و بعد التكهّن بالبيانات الخاصة بالمتغير المعتمد PM10 للبيانات الكلية الخاصة بالتدريب كان من الأفضل ترتيب البيانات بطرق احصائية أكثر تجانساً فتم توجيه الاهتمام الى نمطين من البيانات المترافقة وهي (النمط الاسبوعي و نمط الفصول الموسمية). و بإيجاد نموذج MLR للبيانات الكلية سوف يكون الهدف التالي هو البيانات المترافقة للنمط الأسبوعي وإثبات أن البيانات المترافقة أفضل من البيانات الكلية. و سيتم مناقشة الطريقة الأولى (النمط الأسبوعي) ومقارنتها مع البيانات الكلية.

الفصل الموسمي المترافق الأول من البيانات المترافقة بالنمط الاسبوعي حيث تم فحص البيانات لأيام العطل و كان توجيه النظر نحو الايام التي يتكرر فيها انخفاض قيم PM10، حيث كان افضل يوم اعتباره عطلة نهاية الاسبوع هو يوم السبت لأنه يوم عطلة مشتركة في جميع الولايات الماليزية، وبذلك تم اتباع احد الاساليب المستخدمة من قبل بعض الباحثين باختيار اجزاء من العينة من خلال رصد ايام السيطرة من الشهر او العام نفسه واستخدام اسلوب الترافف الزمني للحصول على احد انماط البيانات المترافقة ومنها النمط الاسبوعي ودراستها وتحليلها كما ذكرت في البحث (Deeb *et al.*, 2007) وكذلك تمت الإشارة له في البحث (Gu *et al.*, 2010).

بعد ترتيب البيانات بأسلوب الترافف الزمني وقبل إيجاد MLR للبيانات سيتم استخدام الجزء الأكبر من البيانات لمجموعة التدريب والجزء المتبقي للاختبار من خلال مجموعة التدريب يتم إيجاد أفضل نموذج ثم يتم فحص قوة التكهّن بمجموعة الاختبار، وقد غطت البيانات ثلاث سنوات لنمط البيانات المترافقة بالنمط الاسبوعي للفترة من (1 كانون الثاني 2013 إلى 14 اب 2015)، والتي استخدمت للتدريب وتم تخصيص ثلاثة أشهر من البيانات (15 اب 2015 إلى 31 تشرين الأول 2015) للاختبار وتم تحديد يوم السبت (عطلة نهاية الأسبوع). وبتطبيق معادلة الانحدار تم الحصول على نموذج النمط الاسبوعي للبيانات المترافقة عطلة نهاية الأسبوع وسيتم عرض النتائج التي تم التوصل إليها في هذا الصدد، ويمكن عرض نموذج النمط الاسبوعي بالمعادلة (11) الآتية:

$$PM10 = -198 + 82.3 CO + 62.0 O3 + 2656 SO2 + 1454 NOX - 2325 NO + 4.56 AT + 0.158 H + 5.16 WS10m + 0.0304 WSxxm \quad (11)$$

الفصل الموسمي المتراسف الثاني من نمط الفصول الموسمية المتراسفة وتعد واحدة من أهم الأساليب الإحصائية المستخدمة من قبل العديد من الباحثين وقد تم التوصل الى اربعة مواسم من كل عام وباستخدام اسلوب التراسف الزمني الذي اتبعه العديد من الباحثين بأخذ اجزاء من العينة بتراسف البيانات موسميا للحصول على نتائج اكثر دقة كما اشار لها العديد من الباحثين في البحث التالي (Perez, 2012) و كذلك اقترحه الباحث في (Huebnerova & Michalek, 2014)

حيث تم التوصل الى اربعة انماط موسمية من خلال بيانات PM10 من حيث الارتفاع والانخفاض المنكرر خلال السنوات الثلاث بنفس الفترة الزمنية و حسب التأثيرات الجوية من امطار او جفاف او اعتدال، و يضم الفصل الموسمي المتراسف الاول الاشهر كانون الاول و كانون الثاني و شباط، و الفصل الموسمي المتراسف الثاني اذار و نيسان و ايار، و الفصل الموسمي المتراسف الثالث حزيران و تموز و اب، و الفصل الموسمي المتراسف الرابع ايلول و تشرين الاول و تشرين الثاني وعلى مدى ثلاث سنوات.

في هذه الدراسة تم استخدام التراسف الزمني لتشكيل انماط فصول موسمية متراسفة موسميا، و قبل ايجاد MLR يتم ايجاد مجموعتين جزئيتين فرعيتين جديدتين من البيانات الجزء الأكبر يستخدم للتدريب والجزء الآخر للاختبار تستخدم مجموعة التدريب للحصول على أفضل النماذج و تستخدم مجموعة الاختبار للتحقق من دقة النماذج وسيتم عرض النماذج التي تم التوصل اليها لأنماط الفصول الموسمية المتراسفة الاربعة بالمعادلات الاتية وعلى التوالي:

$$PM10 = -31.4 + 8.19 CO + 597 O3 + 4241 SO2 + 2253 NOX - 1916 NO + 1.79 AT - 0.396 H + 1.05 WS10m - 0.0498 WSxxm \quad (12)$$

$$PM10 = -136 + 161 CO + 103 O3 + 1964 SO2 - 1534 NOX - 1016 NO + 4.58 AT - 0.597 H + 4.34 WS10m + 0.133 WSxxm \quad (13)$$

$$PM10 = -129 + 128 CO + 514 O3 - 2713 SO2 + 42 NOX - 924 NO + 3.24 AT - 0.151 H + 6.16 WS10m - 0.0080 WSxxm \quad (14)$$

$$PM10 = -99 + 79 CO + 113 O3 + 5075 SO2 + 794 NOX - 2385 NO + 1.48 AT + 0.221 H + 5.82 WS10m - 0.0293 WSxxm \quad (15)$$

حيث سيتم عرض نتائج MAE و RMSE لطريقة MLR في الحالتين التدريب والاختبار للبيانات الكلية، و نمط البيانات المتراففة اسبوعيا و نمط الفصول الموسمية المتراففة، فضلا عن نتائج MAE و RMSE لطريقة PC في حالتي التدريب والاختبار للبيانات وكذلك لباقي الطرق MLR-GA و PC-GA الهجينة والمقارنة بين النتائج التي تم الحصول عليها للبيانات الكلية والبيانات المتراففة فالنمط الاسبوعي يمثل Saturday اما S1, S2, S3, S4 هي انماط الفصول الموسمية المتراففة الاربعة. سيعرض الجدول 1 النتائج التي تم التوصل اليها لطريقة MLR للمقياسين MAE و RMSE لبيانات التدريب والاختبار.

الجدول:1 : قياس مفسرات الخطأ MAE, RMSE لطريقة MLR للتدريب والاختبار

Method	Training		Testing	
	MAE	RMSE	MAE	RMSE
Full data	11.3	15.3	16.3	23.3
Saturday	10.2	13.5	34.9	54.1
S1	6.7	8.4	14.3	17.5
S2	12.9	16.6	9.8	12.1
S3	9.8	13.5	9.3	12.0
S4	10.9	13.9	34.9	48.9

الجدول 1 يعرض MAE و RMSE لطريقة MLR للتدريب و الاختبار للبيانات الكلية حيث كانت MAE هي 11.3 و 16.3 و RMSE هي 15.3 و 23.3 على التوالي، في حين في الاسبوعية كانت MAE هي 10.2 و 34.9 و RMSE هي 13.5 و 54.1 على التوالي، وكانت نتائج انماط الفصول الموسمية المتراففة الاربعة اذ في (S1) كانت MAE هي 6.7 و 14.3 و RMSE هي 8.4 و 17.5 على التوالي، اما في (S2) كانت MAE هي 12.9 و 9.8 و RMSE هي 16.6 و 12.1 على التوالي، اما في (S3) كانت MAE هي 9.8 و 9.3 و RMSE هي 13.5 و 12.0 على التوالي، اما في (S4) كانت MAE هي 10.9 و 34.9 و RMSE هي 13.9، 48.9 على التوالي. ونلاحظ بأن هناك تحسنا في نتائج النمط الاسبوعي و نمط الفصول الموسمية باستثناء الفصل الموسمي المترافف الثاني في بيانات التدريب وكذلك في الاختبار. وعموما تم ملاحظة التحسن في كثير من الأحيان، وبذلك يمكن القول ان البيانات التي خضعت لأسلوب الترافف الزمني في النمطين الاسبوعي و الفصول الموسمية كانت افضل من البيانات الكلية ولذلك تم استخدام الترافف الزمني للبيانات للحصول على نتائج افضل.

اذ عند اجراء مقارنة بين البيانات الكلية و النمط الاسبوعي لطريقة MLR كانت قيم MAE و RMSE للتدريب و الاختبار للبيانات الكلية حيث كانت MAE هي 11.3 و 16.3 و RMSE هي 15.3 و 23.3 على التوالي، و في الاسبوعية كانت MAE هي 10.2 و 34.9 و RMSE هي 13.5 و 54.1 على التوالي التي لوحظت لطريقة MLR للنمط الاسبوعي أفضل من البيانات الكلية في التدريب أما في الاختبار لم تكن أفضل ولكنها كانت مقاربة، ويمكن الإشارة الى وجود تحسن في التدريب في حين

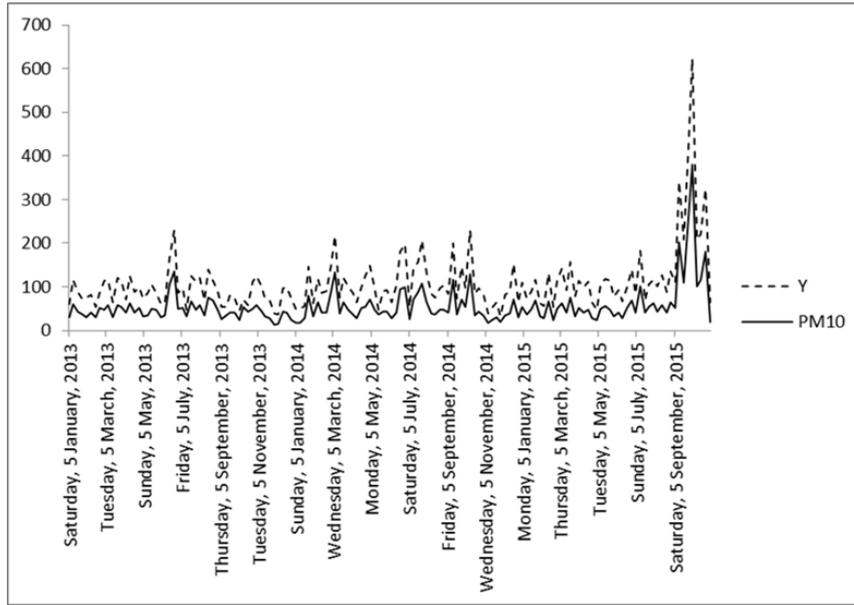
أن مؤشر الاختبار لم يشير الى تحسن، وعموماً يمكن ان يشير ذلك التحسن الى ان استخدام أسلوب الترافف الزمني للبيانات افضل من استخدام البيانات الكلية.

وبمقارنة نتائج MAE و RMSE للبيانات الكلية ونمط الفصول الموسمية المتراففة التي تم الحصول عليها بطريقة MLR للبيانات الكلية حيث كانت MAE هي 11.3 و 16.3 و RMSE هي 15.3 و 23.3 على التوالي. و نتائج نمط الفصول الموسمية المتراففة حيث في (S1) كانت MAE هي 6.7 و 14.3 و RMSE هي 8.4 و 17.5 على التوالي، اما في (S2) كانت MAE هي 12.9 و 9.8 و RMSE هي 16.6 و 12.1 على التوالي، اما في (S3) كانت MAE هي 9.8 و 9.3 و RMSE هي 13.5 و 12.0 على التوالي، اما في (S4) كانت MAE هي 10.9 و 34.9 و RMSE هي 48.9 على التوالي. في الجدول (1) حيث أظهرت النتائج في التدريب ان هناك تحسناً في الفصول الموسمية الثلاثة (الأول والثالث والرابع) باستثناء الفصل الموسمي المترافف الثاني اذ لم يكن هناك تحسناً وكانت النتائج متقاربة. وعند مقارنة النتائج لبيانات الاختبار بين البيانات الكلية و نمط الفصول الموسمية المتراففة، تم ملاحظة أن هناك تحسناً في الفصول الموسمية الثلاثة (الأول والثاني والثالث) باستثناء (الرابع) لم يكن هناك تحسن. وعموماً، تم ملاحظة التحسن في نتائج بيانات نمط الفصول الموسمية المتراففة أفضل من البيانات الكلية في حالي التدريب والاختبار وعموماً تم استنتاج ان البيانات المتراففة افضل من البيانات الكلية لطريقة MLR.

وعند اجراء مقارنة بين نمطي البيانات المتراففة أسبوعياً و الفصول الموسمية المتراففة مع MLR، التي ادرجت في الجدول (1) كانت نتائج MAE و RMSE في النمط الاسبوعي MAE هي 10.2 و 34.9 و RMSE هي 13.5 و 54.1 على التوالي، وكانت نتائج نمط الفصول الموسمية المتراففة في (S1) MAE هي 6.7 و 14.3 و RMSE هي 8.4 و 17.5 على التوالي، اما في (S2) فكانت MAE 12.9 و 9.8 و RMSE 16.6 و 12.1 على التوالي، اما في (S3) فقد كانت MAE 9.8 و 9.3 و RMSE 13.5 و 12.0 على التوالي، اما في (S4) فكانت MAE 10.9 و 34.9 و RMSE 13.9، 48.9 على التوالي. حيث أظهرت النتائج أنه في التدريب كان هناك تحسناً في الفصلين الموسمين المتراففين (الأول و الثالث)، ونلاحظ أن الفصلين الموسمين المتراففين (الثاني و الرابع) لم تشير الى تحسن كبير ولكنها كانت متقاربة جداً، اما في الاختبار تم ملاحظة أن هناك تحسناً في الفصول الموسمية المتراففة الاربعة، اي يمكن القول عموماً ان نمط الفصول الموسمية المتراففة افضل من النمط الاسبوعي.

وسوف يتم توضيح الاتساق بين السلسلتين الاولى الاصلية PM10 و الثانية لسلسلة التكهين Y لبيانات التدريب والاختبار للبيانات الكلية و المتراففة للنمط الاسبوعي و نمط الفصول الموسمية لطريقة MLR وطريقة PCR وكذلك طريقتي MLR-GA و PC-GA الهجينة و رسمها باستخدام برنامج Excel وادراجها بالأشكال الخاصة لكل طريقة.

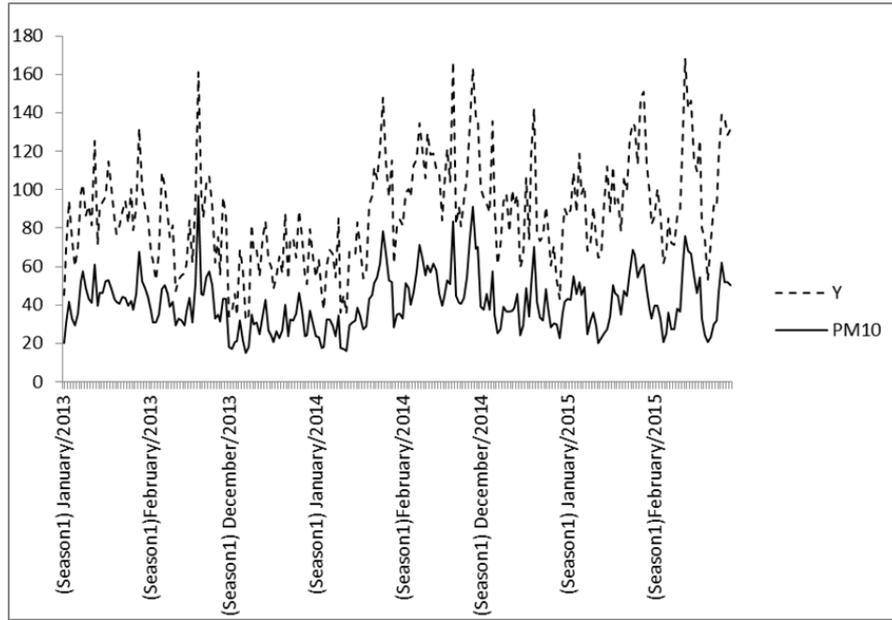
حيث سيوضح الشكل 1 الاتساق بين السلسلتين الاصلية و التكهين لنمط البيانات الاسبوعية المتراففة لطريقة MLR ويضم الجزء التدريبي الذي يبدأ من 5 كانون الثاني 2013 إلى 8 اب 2015، وجزء الاختبار بدءاً من 15 اب 2015 إلى 31 تشرين الاول 2015.



الشكل 1: الاتساق بين السلسلة الاصلية PM10 للنمط الاسبوعي للتدريب والاختبار وسلسلة التكهين Y باستخدام طريقة MLR.

الشكل (1) يعرض اتساقا جيدا بين السلسلتين الاصلية PM10 وسلسلة التكهين Y ليوم السبت للتدريب، والاختبار باستخدام طريقة MLR حيث PM10 تم تمثيلها بالسلسلة المتصلة و Y مثلت بسلسلة متقطعة لتميزهم عن بعض.

بعد ايجاد نماذج MLR لنمط الفصول الموسمية المتراصفة وعرض نتائج MAE و RMSE ومناقشتها، سيتم تمثيل الاتساق بين السلسلة الاصلية وسلسلة التكهين للفصول الموسمية المتراصفة الاربعة وسيتم عرض الشكل 2 للفصل الموسمي الأول كعينة من بيانات نمط الفصول الموسمية المتراصفة، وذلك بسبب العدد الكبير من الرسومات في المجاميع، الذي يوضح جزئي التدريب والاختبار حيث جزء التدريب الذي يبدأ من 1 كانون الثاني 2013 إلى 28 شباط 2014 وجزء الاختبار بدءا من 1 كانون الاول 2014 إلى 28 شباط 2015.



الشكل 2: الاتساق بين السلسلة الاصلية PM10 للنمط S1 وسلسلة التكهن Y للتدريب والاختبار باستخدام طريقة MLR.

الشكل (2) يعرض اتساقا جيدا بين السلسلة الاصلية PM10 وسلسلة التكهن Y باستخدام طريقة MLR للفصل الموسمي الاول من نمط الفصول الموسمية المتراصفة لبيانات التدريب، والاختبار حيث PM10 تم تمثيلها بالسلسلة المتصلة و Y مثلت بسلسلة متقطعة لتميزهم عن بعض.

4. الاستنتاجات Conclusions

سيتم عرض الاستنتاجات للتنبؤ بالطريقة المقترحة للدراسة وهي MLR والحصول على الاستنتاجات للبيانات الكلية والبيانات التي خضعت لأسلوب التراصف الزمني (البيانات المتراصفة بالنمط الاسبوعي والمتراصفة بنمط الفصول الموسمية)، اذ يتم استعراض الاستنتاجات التي تستند إلى البيانات المتراصفة وهي:

1. ان مرارفة البيانات تبعا للنمط الاسبوعي و نمط الفصول الموسمية كان افضل او له دور كبير في إظهار نتائج ادق عند تطبيق الطريقة المقترحة مقارنة مع البيانات الكلية.
2. ان مرارفة البيانات اعتمادا على نمط الفصول الموسمية المتراصفة اظهرت افضل النتائج وتغلبت على الاسبوعية ايضا وبذلك تكون المرارفة للفصول الموسمية في اعلى مرتبة لإظهار ادق النتائج من البيانات الكلية و نمط المرارفة الاسبوعية بعد تطبيق جميع الطرق المقترحة عليها.

5. المصادر REFERENCES

- .1 Abderrahim, H., Chellali, M. R., & Hamou, A. (2016). Forecasting PM10 in Algiers: efficacy of multilayer perceptron networks. *Environmental Science and Pollution Research*, 23(2), 1634-1641 .
- .2 Bateson, T. F., & Schwartz, J. (2001). Selection bias and confounding in case-crossover analyses of environmental time-series data. *Epidemiology*, 12(6), 654-661 .
- .3 Cortina-Januchs, M. G., Quintanilla-Dominguez, J., Vega-Corona, A., & Andina, D. (2015). Development of a model for forecasting of PM10 concentrations in Salamanca, Mexico. *Atmospheric Pollution Research*, 6(4), 626-634.
- .4 Deeb, O., Hemmateenejad, B., Jaber, A., Garduno-Juarez, R., & Miri, R. (2007). Effect of the electronic and physicochemical parameters on the carcinogenesis activity of some sulfa drugs using QSAR analysis based on genetic-MLR and genetic-PLS. *Chemosphere*, 67(11), 2122-2130 .
- .5 Dolatabadi, M., Nekoei, M., & Banaei, A. (2010). Prediction of antibacterial activity of pleuromutilin derivatives by genetic algorithm-multiple linear regression (GA-MLR). *Monatshefte für Chemie-Chemical Monthly*, 141(5), 577-588 .
- .6 Elbayoumi, M., Ramli, N. A., Md Yusof, N. F. F., Yahaya, A. S. B., Al Madhoun, W., & Ul-Saufie, A. Z. (2014). Multivariate methods for indoor PM10 and PM2.5 modelling in naturally ventilated schools buildings. *Atmospheric Environment*, 94, 11-21.
- .7 Gu, J., Bai, Z., Liu, A., Wu, L., Xie, Y., Li, W . . . , Zhang, X. (2010). Characterization of atmospheric organic carbon and element carbon of PM2. 5 and PM10 at Tianjin, China. *Aerosol Air Qual. Res*, 10, 167-176 .
- .8 Güler, N., & Güneri İşçi, Ö. (2016). The regional prediction model of PM10 concentrations for Turkey. *Atmospheric Research*, 180, 64-77.
- .9 Huebnerova, Z., & Michalek, J. (2014). Analysis of daily average PM10 predictions by generalized linear models in Brno, Czech Republic. *Atmospheric Pollution Research*, 5(3), 471-476.
- .10 Kumar, A., & Goyal, P. (2011). Forecasting of air quality in Delhi using principal component regression technique. *Atmospheric Pollution Research*, 2(4), 436-444.
- .11 Levy, D., Lumley, T., Sheppard, L., Kaufman, J., & Checkoway, H. (2001). Referent selection in case-crossover analyses of acute health effects of air pollution. *Epidemiology*, 12(2), 186-192 .
- .12 Malig, B. J., Pearson, D. L., Chang, Y. B., Broadwin, R., Basu, R., Green, R. S., & Ostro, B. (2016). (A time-stratified case-crossover study of ambient ozone exposure and emergency department visits for specific respiratory diagnoses in California (2005-2008). *Environmental health perspectives*, 124(6), 745 .
- .13 Nathans, L. L., Oswald, F. L., & Nimon, K. (2012). Interpreting multiple linear regression: A guidebook of variable importance. *Practical Assessment, Research & Evaluation*, 17(9) .(
- .14 Sattari, M., & Gharagheizi, F. (2008). Prediction of molecular diffusivity of pure components into air: a QSPR approach. *Chemosphere*, 72.1302-1298 ,(9)
- .15 Sousa, S., Martins, F., Alvim-Ferraz, M., & Pereira, M. C. (2007). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling & Software*, 22(1), 97-103 .

- .16 Stadlober, E., Hörmann, S., & Pfeiler, B. (2008). Quality and performance of a PM10 daily forecasting model. *Atmospheric Environment*, 42(6), 1098-1109.
- .17 Taşpınar, F., & Bozkurt, Z. (2014). Application of Artificial Neural Networks and Regression Models in the Prediction of Daily Maximum PM10 Concentration in Düzce, Turkey .
- .18 Tobias, A., Armstrong, B., & Gasparri, A. (2014). *Analysis of time-stratified case-crossover studies in environmental epidemiology using Stata*. Paper presented at the United Kingdom Stata Users 'Group Meetings 2014.
- .19 Ul-Saufie, A. Z., Yahaya, A. S., Ramli, N. A., Rosaida, N., & Hamid, H. A. (2013). Future daily PM10 concentrations prediction by combining regression models and feedforward backpropagation models with principle component analysis (PCA). *Atmospheric Environment*, 77, 621-630.
- .20 Vlachogianni, A., Kassomenos, P., Karppinen, A., Karakitsios, S., & Kukkonen, J. (2011). Evaluation of a multiple regression model for the forecasting of the concentrations of NOx and PM10 in Athens and Helsinki. *Science of The Total Environment*, 409(8), 1559-1571.
- .21 Volume 40, I, July 2006, Pages 4126–4136. (2006). PM10 modeling of Beijing in the winter .
- .22 Wang, P., Liu, Y., Qin, Z., & Zhang, G. (2015). A novel hybrid forecasting model for PM10 and SO2 daily concentrations. *Science of The Total Environment*, 505, 1202-1212.