المجلة العراقية للعلوم الإحصائية(22) 2012
ص ص [121-135]

استخدام أسلوب انحدار الحرف في التنبؤ لتقدير نماذج الصندوق الأسود مع التطبيق

محاسن صالح الطالب*

الملخص

تناول البحث توفيق بعض نماذج الصندوق الأسود منها نماذج خطأ المخرجات التي تضم نموذج الانحدار الذاتي مع مدخلات إضافية ARX ونموذج الانحـدار الـذاتي والمتوسـطات المتحركة مع مدخلات إضافية ARMAX ،تم اختيار أفضل نموذج يمثل البيانات التي تتعلـق بدرجة الحرارة التي تتأثر بمجموعة من المتغيرات التوضيحية تمثلت بـ(الإشـعاع الشمـسي والسطوع الشمسي والتبخر).قدرت معلمات أفضل نموذج بطريقة انحدار الحرف عند وجـود معلومات مسبقة حول معلمات الأنموذج وعدمه وتمت مقارنة خطأ التكهن للنموذج الذي قـدر بطريقة المربعات الصغرى وطريقة انحدار الحرف عند عدم توفر المعلومات المـسبقة عـن المعلمات وعدمه.

Employing Ridge Regression Technique in Prediction of the Black box models with Application

Abstract

This paper is concerned with fitting some black box models. Some of them are, the outputs error model which contains the autoregressive and autoregressive moving average with additional inputs(ARX and ARMAX).The best model has been chosen which represents the data about the Temperature which is affected by some predictor variables which they were represented by (Brightness, unlike radiation, and evaporation).The parameters of the best model were estimated by the ridge regression method with and without the existence of prior information around the model parameters. The prediction errors of the model which has been estimated by least square and ridge regression when the prior information about the parameters is available were compared.

* مدرس / قسم الإحصاء والمعلوماتية / كلية علوم الحاسوب والرياضيات / جامعة الموصل.

تاريخ التسليم 2011/11/1 ________ تاريخ القبول 2012/2/26

1. المقدمة

يهتم علم القياس الاقتصادي بدراسة الظواهر الاقتصادية بطريقة كمية من خلال تحليل البيانات والتعرف على طبيعة العلاقة بين المتغيرات وقياس تلك العلاقة رقمياً، وتتطلب أية دراسة قياسية لظاهرة معينة ضرورة تحديد العوامل المؤثرة في تلك الظاهرة وصياغة العلاقة بين هذه العوامل في صورة نموذج قياسي يعبر عنها.

يعد تحليل الانحدار جزءاً مهما من علم الإحصاء وأسلوبا من أساليب الإحصاء التطبيقي عند دراسة كثير من الظواهر، ويعرف تحليل الانحدار Regression Analysis بشكل عام بأنه مقياس رياضي لمتوسط العلاقة بين متغيرين أو أكثر بدلالة وحدات قياس المتغيرات التوضيحية في العلاقة وغالباً ما تسمى العلاقات من هذا النوع بنماذج الانحدار وينقسم تحليل الانحدار إلى قسمين رئيسين هما الانحدار الخطي والانحدار اللاخطي (Linear Regression and Non-Linear Regression) ، الانحدار الخطي يكون بسيطاً إذا كان لدينا متغيران مرتبطان بعلاقة الأول يكون توضيحياً والثاني يمثل متغير استجابة ، والانحدار يكون متعددا (عاماً) إذا كان لدينا أكثر من متغيرا توضيحي.

إن نموذج الانحدار الخطي المتعدد بني على مجموعة من الفروض منها عدم توافر علاقة خطية تامة أو شبه تامة بين المتغيرات المستقلة أي غياب مشكلة تعدد العلاقة الخطية (Multicollinearity) (النعيمي،2005)، التي تتسبب في تشويه المقدرات التي تقدر بطريقة المربعات الصغرى الاعتيادية، ويشير المصطلح إلى وجود علاقة خطية تامة أو غير تامة بين اثنين أو أكثر من المتغيرات المستقلة في نموذج الانحدار مما يؤدي إلى خرق فرض من فروض التحليل، وتؤدي هذه المشكلة إلى الحصول على تقديرات غير متسقة ولها أخطاء معيارية كبيرة ولغرض معالجة مشكلة تعدد العلاقات الخطية فإنه يتطلب تحديد المتغير أو المتغيرات المسببة لهذه المشكلة(كاظم،2005) . ولغرض تشخيص مشكلة تعدد العلاقة الخطية في النموذج الذي يحوي على متغيرين مستقلين أو أكثر توجد عدة مقاييس منها **<u>عامل تضخم التباين</u>** **Variance Inflation Factor** حيث اقترح هذا المقياس Ferrar , Glaube (1967) وسماه (1970) Minquardt بـ (VIF)، يقيس هذا العامل التضخم في التباين للمعلمات المقدرة لكل المتغيرات المستقلة في النموذج ويعتمد هذا المقياس على فحص عناصر القطر الرئيس للمصفوفة $(X'X)^{-1}$ حيث أن VIF يساوي a_{jj} والتي تمثل $(1-R_j^2)^{-1}$ بحيث ان R_j^2 عبارة عن معامل التحديد لانحدار x_j على بقية المتغيرات المستقلة ، a_{jj} تمثل عناصر القيم القطرية. اقترح Gunst and mason (1980) بأنه إذا كان $a_{jj} \geq 4$ فإن ذلك دليل كافي على كون المتغير x_j يعاني من تضخم في تباين معامله وانه مرتبط ببقية المتغيرات

المستقلة.(سعيد، 1996)، (الصفاوي،يحيى،2009). وكذلك **معامل الارتباط البسيط** بين المتغيرات الذي يعد من ابسط مقاييس الكشف عن التداخلات الخطية. **ولمعالجة مشكلة تعدد العلاقة الخطية** يوجد هناك العديد من الأساليب التي تتم فيها معالجة التعدد الخطي بين المتغيرات منها:

1. التحويل المعياري للمتغيرات التوضيحية.
2. تضخيم البيانات.
3. حذف بعض المتغيرات.
4. استخدام طرائق التقدير التي تعتمد على المعلومات الكمية الأولية ومنها:

أ) طريقة المربعات الصغرى المقيدة.

ب) استخدام الطرائق المتحيزة في التقدير وأهمها:

1- طريقة المكونات الرئيسة Principle Component Analysis

2- طريقة انحدار الحرف Ridge Regression Method

أ) طريقة انحدار الحرف الاعتيادية.

ب) طريقة انحدار الحرف مع المعلومات المسبقة.

وسيتم الاعتماد في هذا البحث على هاتين الطريقتين (أ،ب) في تقدير معلمات النماذج .

2. هدف البحث

من المشاكل التي تواجه الباحثين هي مشكلة تعدد العلاقة الخطية التي تؤدي إلى أن تكون مقدرات معلمات النماذج باستخدام طريقة المربعات الصغرى غير دقيقة ولذلك يتم اللجوء إلى استخدام طرائق أخرى، في هذا البحث تم استخدام طريقة انحدار الحرف الاعتيادي وانحدار الحرف مع المعلومات المسبقة للتنبؤ بنماذج الصندوق الأسود ومقارنة النتائج .

3. انحدار الحرف Ridge Regression

أ) انحدار الحرف الاعتيادي :

تعد طريقة انحدار الحرف بأنها طريقة معدلة لطريقة المربعات الصغرى الاعتيادية في حالة وجود تداخل خطي بين المتغيرات التوضيحية، وتعتبر من الطرائق المتحيزة حيث تعتمد على إضافة المصفوفة (kIp) إلى المصفوفة $(X'X)$ قبل إيجاد المعكوس لها وكالآتي: (النعيمي،2005)

$$\hat{\beta} = (X'X + kIp)^{-1} X'Y \qquad ..(1)$$

حيث أن: (kIp) مصفوفة قطرية ذات بعد $(p \times p)$

1. معلمة الحرف: يمكن أن يعرف انحدار الحرف على انه دالة لمعلمة الحـرف k و k عدد موجب صغير يضاف إلى القطر الرئيس للمصفوفة $(X'X)$ بعد ضربه بالمصفوفة الأحادية (Ip)، $0 < k < 1$، إذ يعمل هذا العدد على تقليل تباين المعلمات المقدرة، ولكـن على حساب إدخال بعض التحيز على مقدرات تلك المعلمات ،إلا أن متوسط مربعـات الخطأ mean square error لمقدرات انحدار الحرف اقل من متوسط مربعات الخطـأ لطريقة المربعات الصغرى الاعتيادية.
2. تقدير معلمة الحرف:أن اختيار معلمة الحرف k يعود إلى خبرة الإحـصائي وعنـدما تكون فترة زيادة قيمة k زيادة معقولة وصحيحة وغير سريعة،إذ يكـون تحـسن فـي المقدرات وهذا التحسن غالباً ما يؤخذ من الواقع الذي تكون فيه المقدرات أكثر استقراراً أو ملاحظة التحسن في التنبؤ
3. كلما تكبر قيمة k ضمن الفترة $(0,1)$ فإن التباين يقل وتقدير المعلمـات يـصبح أكثـر استقراراً.

توجد عدة طرائق لاختيار معلمة الحرف منها الطريقة التكرارية التي تعد من الطرائق المهمة والناجحة لاختيار معلمة الحرف ،إذ أنها تتمتع بصغر متوسط مربعات الخطأ مقارنة بالطرائق الأخرى.(القصيمي،2000)

تعتمد الطريقة التكرارية في تقدير معلمة الحرف على قيمة الحرف k وتعويضها فـي معادلة انحدار الحرف لإيجاد عدة قيم لمقدرات الحرف، ومن ثم ملاحظة اسـتقرار هـذه المقدرات من خلال اخذ الفرق بين التقديرات اللاحقة والسابقة لهذه المعلمات بعد افتراض خطأ معين يمكن السماح به ويتم اختيار احد المقدرات ويعد هذا المقدر هو الأفضل وقيمة k المستخدمة هي القيمة المثلى. (القصيمي،2000) ،(القصيمي وآخرون،2007)

ب) انحدار الحرف مع المعلومات المسبقة(الأولية) (الحسو، 2009):

Ridge Regression with Prior Information

عمم الباحث (Swindle, 1976) نموذج Hoerl & Kennard وذلك بإضافة معلومـات مسبقة أولية أساسية إلى الطريقة المتحيزة لانحدار الحرف وبذلك تصبح معادلة انحدار الحرف بالصيغة الآتية:

$$\hat{\beta}(KI, J) = (X'X + KI_p)^{-1}(X'Y + KJ) \qquad (2)$$

إذ أن J هو متجه ثابت للمعلومات المسبقة لـ β والهدف من إضـافة الثابـت K ومتجـه المعلومات المسبقة J إلى المتجه $X'Y$ هو تصغير مقدار التحيز، وتم حساب قيمة المعلومـات المسبقة J من خلال المتوسط الحسابي لمعلمات طريقة المربعات الصغرى مـضروباً بمتجـه ثابت أحادي ، ويعبر عنه رياضياً بالصيغة الآتية :

$$J = \frac{\sum_{i=1}^{p} \hat{\beta}_{iLS}}{p} \begin{bmatrix} 1 \\ 1 \\ : \\ 1 \end{bmatrix} \qquad (3)$$

المعلومات المسبقة (الأولية) وعلاقتها ببيانات العينة:

Prior information related to sample data

في المجال التطبيقي قد لانستطيع الحصول على تقدير لمتجه المعلومات الأوليـة ، لـذلك يمكن استخدام المتجه الأولي المتعلق بالعينة ، وبذلك يمكن الحـصول علـى تقـدير لمتجـه المعلومات الأولية من خلال هذه الإمكانية إذ يمكن استخدام الوسط الحسابي لمعلمات طريقـة المربعات الصغرى (LS) أي $\hat{\beta}$ كعناصر للمتجه الأولي، وكالآتي :

$$J = \frac{1}{p} \begin{bmatrix} \sum \hat{\beta}_{iLS} \\ \sum \hat{\beta}_{iLS} \\ : \\ \sum \hat{\beta}_{iLS} \end{bmatrix} \qquad (4)$$

حيث أن :

$$\sum_{i=1}^{p} \hat{\beta}_{iLS} = \hat{\beta}_1 + \hat{\beta}_2 + \ldots + \hat{\beta}_p$$

ليكن $(X'X)^{-1}$ تمثل بالمصفوفة

$$(X'X)^{-1} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1p} \\ a_{21} & a_{22} & \ldots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \ldots & a_{pp} \end{bmatrix}$$

وليكن $X'Y$ يتمثل بالمتجه $\begin{bmatrix} S_1Y & S_2Y & \ldots & S_pY \end{bmatrix}$

$$\therefore \sum \hat{\beta}_{iLs} = \hat{\beta}_1 + \hat{\beta}_2 + \cdots + \hat{\beta}_p$$
$$= a_{11}S_1Y + a_{12}S_2Y + \cdots + a_{1p}S_pY + a_{21}S_1Y + a_{22}S_2Y + \cdots + a_{2p}S_pY$$
$$+ a_{31}S_1Y + a_{32}S_2Y + \cdots + a_{3p}S_pY + \cdots + a_{p1}S_1Y + a_{p2}S_2Y + \cdots + a_{pp}S_pY$$
$$= (a_{11} + a_{21} + \cdots + a_{p1})S_1Y + (a_{12} + a_{22} + \cdots + a_{p2})S_2Y + \cdots + (a_{1p} + a_{2p} + \cdots + a_{pp})S_pY$$

حيث أن:

$$\begin{aligned} b_1 &= a_{11} + a_{21} + \cdots + a_{p1} \\ b_2 &= a_{12} + a_{22} + \cdots + a_{p2} \\ &\vdots \\ b_p &= a_{1p} + a_{2p} + \cdots + a_{pp} \end{aligned}$$

$$\therefore J = BX'Y$$

$$B = \begin{bmatrix} b_1 & b_2 & \cdots & b_p \\ b_1 & b_2 & \cdots & b_p \\ \vdots & \vdots & \ddots & \vdots \\ b_1 & b_2 & \cdots & b_p \end{bmatrix} \qquad (5)$$

حيث أن B مصفوفة و b_j تمثل مجموع عناصر العمود j من أعمدة المصفوفة $(X'X)^{-1}$

$$\therefore \hat{\beta}_R = \left(X'X + KI\right)^{-1}\left(I + \frac{K}{P}B\right)X'Y \qquad \ldots(6)$$

4. نماذج الصندوق الأسود Black Box Models

يتكون تركيب النموذج الخطي العام General Linear Model Structure الذي تشتق منه كل النماذج الخطية Linear Models والذي عادةً ما لا يطبق عملياً لأنه يمثل الهيكلية الموحدة للأنظمة الخطية حيث يمكن حساب مخرجات النظام الخطي المحدد Y_t عند الزمن t من خلال ترشيح المدخلات u_t بمرشح خطي يرمز له $G(q)$ والذي يسمى بـالجزء المحدد،فضلا عن ترشيح التشويش الأبيض الذي يرمز له v_t بمرشح خطي آخر يرمز له $H(q)$ ويسمى بالجزء التصادفي وكالآتي :-

$$y_t = G(q)u_t + H(q)v_t \qquad \ldots \quad (7)$$

ويمكن إعادة صياغة هذا النموذج من خلال متعددات الحدود التي هي عبارة عن تراكيب خطية بالشكل الآتي:

$$y_t = \frac{B(q)}{F(q)A(q)}u_t + \frac{C(q)}{D(q)A(q)}v_t \qquad \dots (8)$$

حيث أن :

$$A(q) = 1 + a_1 q^{-1} + a_2 q^{-2} + \cdots + a_{na} q^{-na}$$
$$B(q) = b_1 q^{-1} + b_2 q^{-2} + \cdots + b_{nb} q^{-nb}$$
$$C(q) = 1 + c_1 q^{-1} + c_2 q^{-2} + \cdots + c_{nc} q^{-nc}$$
$$D(q) = 1 + d_1 q^{-1} + d_2 q^{-2} + \cdots + d_{nd} q^{-nd}$$
$$F(q) = 1 + f_1 q^{-1} + f_2 q^{-2} + \cdots + f_{nf} q^{-nf}$$

وان: nf, nd, nc, nb, na تمثل معلمات متعددات الحدود على التوالي وان : q^{-1} مشغل الإزاحة الخلفي Backward Shift

تقسم نماذج الصندوق الأسود إلى قسمين:-

1. نماذج خطأ المعادلة Equation Error Models

وتشمل نماذج ARX وARMAX وتتميز بوجود المرشح الخطي الشائع 1/A(q) في كل من نموذج العمليات المحددة ونموذج العمليات التصادفية وبعبارة أخرى تشترك جميع نماذج ARX وARMAX بمتعدد الحدود $A(q)$ كمقام حركي لدالة تحويل المدخلات ودالة تحويل التشويش وهذا ينسجم مع حقيقة أن التشويش لا يؤثر مباشرة في مخرجات النموذج ولكن بدلا من ذلك يدخل النموذج قبل المرشح 1/A(q) وبعبارة أخرى أن التشويش يدخل إلى العملية متأخرا ومن ثم فان ميزات التكرار تتشكل من خلال حركية العملية ويمكن تمثيل نموذج ARX بالشكل الآتي:

$$y_t = \frac{B(q)}{A(q)}u_t + \frac{1}{A(q)}v_t \qquad \dots (9)$$

أو يمكن كتابته بالشكل الآتي:-

$$A(q)y_t = B(q)u_t + v_t \qquad \dots (10)$$

نموذج ARMAX يمكن تمثيله بالشكل الآتي:

$$y_t = \frac{B(q)}{A(q)}u_t + \frac{C(q)}{A(q)}v_t \qquad \dots (11)$$

2. نماذج خطأ المخرجات Output Error Models

وتشمل كلاً من نموذج خطأ المخرجات Output Error ويرمز له بـ OE ونموذج بوكس-جنكينز Box-Jenkins ويرمز له بـ BJ .(Nells.2001)،(حياوي،2006)

5. معايير اختيار أفضل نموذج

1. دالة الخسارة Loss Function

تعرف في بعض الأحيان بدالة الكلفة حيث تعد المتطلب الأساس الأفضل في اختيار رتبة النموذج من خلال ملاحظة سلوك هذه الدالة مع زيادة رتبة النموذج، حيث تقل قيمتها بزيادة رتبة النموذج.(Nelles, 2001)

$$V = \frac{1}{2}\sum_{t=1}^{N} e_t^2 \qquad \dots (12)$$

إذ إن V : تمثل دالة الكلفة

N : تمثل عدد المشاهدات .

2. معيار أكاكي لخطأ التنبؤ النهائي Akaik's Final Prediction Error Criteria

يعد من المعايير المهمة في تحديد الرتبة المناسبة للنموذج، حيثُ عرفه العالم Akaike عام 1969 ويرمز له اختصاراً بـ FPE ويمثل مقياس خطأ التنبؤ النهائي ويعرف على انه تباين خطأ التنبؤ لفترة قادمة ويحسب بالشكل الآتي :(Ljung, 1999)

$$FPE = \frac{1+\frac{m}{N}}{1-\frac{m}{N}}V \qquad \dots (13)$$

إذ إن

FPE : تُعبر عن خطأ التنبؤ النهائي

m : تُعبر عن عدد المعلمات في النموذج .

3. معيار اكاكي للمعلومات Akaike's Information Criteria

عرف هذا المعيار أيضا من قبل العالم Akaike عام (1973-1974) ويرمز له اختصارا بـ AIC وقدم من خلاله معلومات لاختيار الرتبة الملاءمة لنموذج ARIMA من بين عدة نماذج بحيث تقابل الرتبة المناسبة اقل قيمة لمعيار AIC وتمثل الرتبة الأكثر

ملاءمة،وهذا المعيار يساوي ضعف عدد المعلمات مطروحاً منه ضعف دالة الإمكان الأعظم للمشاهدات، ويعبر عنه رياضياً بالشكل الآتي:- (Nelles, 2001,)

$$AIC = 2m - 2\ln L \quad \dots (14)$$

إذ إن:-

AIC : تُعبر عن معيار اكاكي للمعلومات

L : تُعبرعن دالة الإمكان الأعظم.

علما أن معظم البرمجيات الجاهزة مثل Matlab تستخدم دالة الكلفة في حـساب قيمـة هذا المعيار ويمكن التعبير عنه بالشكل الآتي:-(Ljung, 2004)

$$AIC = \log\left(V\left(1+2\frac{m}{N}\right)\right) \quad \dots (15)$$

6. اختبار دقة النتائج التنبؤية

في بعض الأحيان يطلق يثعبر عن الدقة بكلمة حسن المطابقة التي تشير إلى الأسـلوب الذي يستخدم لجعل نموذج التنبؤ قادراً على توليد بيانات ذوات كفاية، هناك عدة معايير يـتم من خلالها ضبط دقة النموذج في التكهن منها : (Makridakis, 1998)

1. معدل القيم المطلقة للأخطاء Mean Absolute Error

ويمكن حسابه بالصيغة الآتية:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|e_t|$$
$$e_t = (Y_t - F_t) \quad \dots (16)$$

حيث أن : e_t :تمثل البواقي

Y_t : تمثل القيم الحقيقية للسلسلة، F_t :تمثل القيم المتنبأ بها للسلسلة. n : تمثل فترة التكهن

2. متوسط الخطأ النسبي Mean Percentage Error

ويحسب بالصيغة الآتية :

$$MPE = \frac{1}{n-m}\sum_{t=1}^{n} PE_t$$
$$PE = \left(\frac{Y_t - F_t}{Y_t}\right) * 100 \quad \dots (17)$$

m : عدد معلمات النموذج.

3. متوسط الخطأ النسبي المطلق Mean Absolute Percentage Error

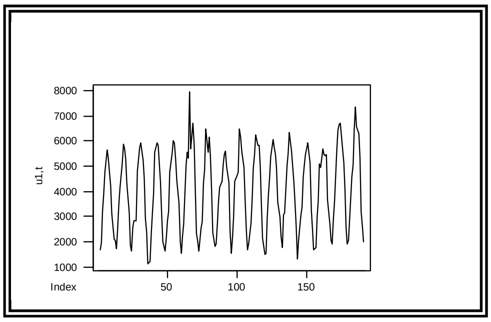
ويمكن حسابه بالصيغة الآتية:

$$MAPE = \frac{1}{n-m}\sum_{t=1}^{n}|PE_t| \quad \dots (18)$$

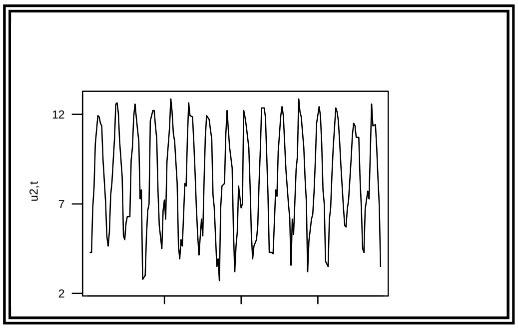
7. الجانب العملي

يستند البحث إلى بيانات حقيقية واقعية مأخوذة من دائرة الأنواء الجوية في محافظة نينوى تمثل المعدلات الشهرية لدرجات الحرارة والعوامل المؤثرة فيها للفترة (1985-2000) وهي عبارة عن المدخلات المتمثلة بالإشعاع الشمسي $x_{1,t}$ الذي يقاس بـ (واط.يوم/م^2)، وسطوع الشمس $x_{2,t}$ الذي يقاس بـ (ساعة) ، والتبخر $x_{3,t}$ الذي يقاس بـ (ملمتر) فضلاً عن المخرجات وهي عبارة عن درجة الحرارة y_t وتتضمن (192) مشاهدة لكل من المدخلات والمخرجات .

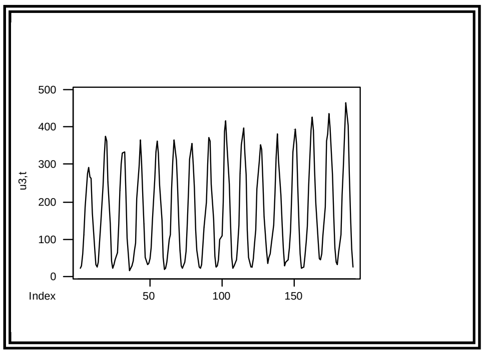
لمعرفة ثبوت المدخلات والمخرجات واستقراريتها تم تمثيل الرسم الزمني للمدخلات والمخرجات على التوالي:



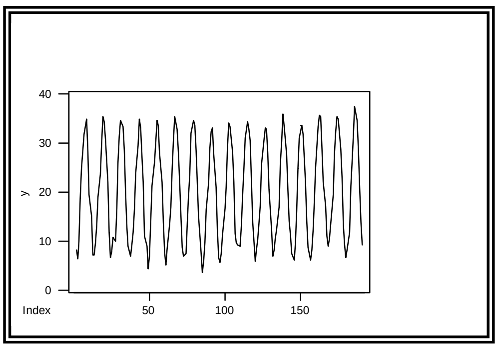
الشكل (1) : الرسم الزمني للاشعاع الحراري



الشكل (2) : الرسم الزمني للسطوع



الشكل (3) : الرسم الزمني للتبخر



الشكل (4) : الرسم الزمني للحرارة

نلاحظ من الأشكـال (1) , (2), (3), (4) أعلاه أن البيانات غيـر مستقرة لذلك تـم إجراء أحد التحويلات عليها وهو الجذر التربيعي وبعدها تم اخذ الفروق الموسمية من الرتبـة الأولى للإزاحة الموسمية(12) وثم اخذ الفرق اللاموسمي الأول وعليه تم الحصول على 179 مشاهدة ، استخدمت أول 30 مشاهدة كمعلومات أولية (مسبقة) .

وباستخدام التوافيق لإيجاد نماذج خطأ المعادلة والمتمثـلة بـ ARMAX ,ARX فقد اخذ 140 مشاهدة من البيانات المتبقية واستخدم الباقي للتنبؤ , تم توفيق العديد مـن النمـاذج بمعلمات متعددة ومختلفة لكل نموذج من نماذج خطأ المعادلة واختيار أفضل نمـوذج حـسب المعايير الإحصائية AIC, FPE , Loss function الموضحة في الجدول (1).

الجدول (1): يبين النموذج الأفضل لنماذج خطأ المخرجات حسب المعايير

المعايير / النماذج	AIC	Loss fun.	FPE
ARMAX	-2.575	0.04053	0.05059
ARX	-2.9775	0.042294	0.0501432

من الجدول(1) تم تحديد أفضل نموذج من نماذج الصندوق الأسود الذي يمتلك اقل قيمة بالنسبة للمعايير الإحصائية المذكورة وهو النموذج ARX وصيغته بالشكل الآتي:

$$y_t = -0.4321^* y_{t-1} - 0.5084^* y_{t-2} - 0.4141^* y_{t-3} - 0.157^* y_{t-4} - 0.1118^* y_{t-5} - 0.00688^* y_{t-6}$$
$$- 0.0317^* y_{t-7} + 0.005736^* X_{1t} + 0.01404^* X_{1,t-1} + 0.3531^* X_{2t} - 0.1205^* X_{3t}$$
$$- 0.06515^* X_{3,t-1} - 0.05574^* X_{3,t-2} + e_t \qquad (19)$$

بعـد ذلك تم إيجاد أفضل معادلـة انحدار للمتغيـر المعتمد مـع المتغيـرات التوضـيحية $x_{1,t}$, $x_{2,t}$, $x_{3,t}$, y_t على التوالي , حيث اختبر هـل هنـاك علاقـة بيـن هـذه المتغيرات من خلال استخدام مصفوفة الارتباط أولا , و كالآتي :

$$R = \begin{bmatrix} 0.886 & 0.940 & 0.966 \\ 0.940 & 0.925 & 0.912 \\ 0.966 & 0.912 & 0.935 \end{bmatrix}$$

نلاحظ من مصفوفة الارتباط أن الارتباط بين $x_{1,t}$, y_t هو 0.886 والـذي يؤكـد على وجـود علاقـة قويـة وكـذلك بالنـسبة $x_{2,t}$, y_t قيمـة الارتبـاط كانـت0.940

وبيـن $x_{1,t}$, $x_{2,t}$ كـان الارتبـاط 0.925 ونفـس الحالـة بـين $x_{3,t}$ مـع كـل مـن $x_{1,t}$, $x_{2,t}$, y_t التي تمتلك قيـم ارتباط عالية جدا .

وللتأكـد مـن معنويـة هـذه الارتباطـات تـم اختبارهـا حسب الـصيغة $t = r(n-2)\big/\sqrt{1-r^2}$ حيث أن r تمثل معامل الارتباط البسيط بين x_{jt}, y_t و $x_{jt}, x_{j't}$ وبعد مقارنتها بالقيمة الجدولية المناسبة $t_{(\alpha/2,n-2)}$ تبين أن جميع الارتباطات معنويـة وكمـا موضحة في الجدول الآتي :

الجدول (2) معاملات الارتباط وقيم t المحسوبة

Correlation	t-calculate
$r_{x_1,y} = 0.886$	363.043
$r_{x_2,y} = 0.940$	523.486
$r_{x_3,y} = 0.966$	709.904
$r_{x_1 x_2} = 0.925$	462.540
$r_{x_1,x_3} = 0.912$	422.438
$r_{x_2,x_3} = 0.935$	500.920

وحسب عامل تضخم التباين Variance Inflation Factor (VIF) وكانت النتائج موضحة في الجدول الآتي:

الجدول (3): يوضح قيم VIF لمتغيرات الاستجابة

Predictor	VIF
x_1	7.8
x_2	10.6
x_3	9.1

نلاحظ أن قيم VIF التي تمثل عامل التضخم في تباينات المعاملات المبينة في الجدول (3) كونها اكبر مـن 4, فمن خلال نتائج هذين الاختبارين لوحظ وجـود مـشكلـة تعـدد خطـي ولمعالجة مشكلـة التعدد الخطي استخدمت إحدى طرائق معالجـة التعـدد الخطـي وهـي طريقة انحدار الحرف وبأسلوبين الأول انحدار الحرف الاعتيادي حيث تـم اختيـار

عـدة قيم للثابت k التي تقع قيمها بين (1 ,0)، إن أفضل قيمة للثابت هي k=0.95 وتحـت خطأ مسموح به اقل أو يساوي 0.001 التي تعطي أفضل معادلة انحدار وصـيغتها بالـشكل الآتي:

$$y = -0.000815 - 0.00262 * x_1 - 0.319261 * x_2 + 0.129519 * x_3 \quad (20)$$

والثاني انحدار الحرف مع المعلومات المسبقة ، تبين أن أفضل قيمة للثابت k هـي0.1 وبنفس الخطأ في الأسلوب الأول والتي تعطي أفضل معادلة انحدار وهي

$$y = -0.001152 - 0.001106 * x_1 - 0.363211 * x_2 + 0.134437 * x_3 \quad (21)$$

واختيار أفضل نموذج حسب المعـايير الإحـصائية AIC, FPE , Loss function الموضحة في الجدول الآتي .

الجدول (4):النموذج الأفضل حسب المعايير

المعايير / النماذج	AIC	Loss fun.	FPE
ARX	**-2.9775**	**0.0422894**	**0.0501432**
انحدار الحرْف	**0.686092**	**4.59154**	**4.86163**
انحدار الحرْف مع المعلومات المسبقة	**0.685432**	**4.58457**	**4.85425**

من الجدول(4) تم تحديد أفضل نموذج من النماذج الثلاثة الذي يمتلك اقل قيمة بالنسبة للمعايير الإحصائية المذكورة وهو النموذج ARX والمتمثل بالمعادلة (19).

وبعد ذلك تم إيجاد القيم التنبؤية لنموذج ARX ونماذج الانحدار وكما موضحة في الجدول الآتي :

الجدول (5) :القيم الحقيقية والتنبؤية لنموذج ARX ونماذج انحدار الحرف وانحدار الحرف مع المعلومات المسبقة

t	Y	$\hat{y}_{ARX}$	$\hat{y}_r$	$\hat{y}_{pr}$
1.	0.429862	0.310990	0.302934	0.310240
2.	-0.302375	-0.287039	-0.468635	-0.486753
3.	0.098722	0.069887	0.119377	0.120019
4.	0.192260	-0.057750	0.165328	0.167917
5.	-0.191246	-0.105770	-0.108309	-0.113723
6.	0.016141	-0.087240	0.052993	0.062266
7.	-0.204892	0.118822	-0.043578	-0.062772

8.	0.268467	0.088120	-0.034392	-0.023778
9.	-0.087026	0.167572	-0.038606	-0.031089

وباستخدام معايير التنبؤ لاختيار أفضل نموذج تم الحصول على النتائج المبينة في الجدول الآتي:

الجدول (6): يوضح معايير التنبؤ للمقارنة بين نموذجي الانحدار و ARX

المعايير / النموذج	MAE	MPE	MAPE
ARX	0.151174	-348.718	-348.718
انحدار الحرف	0.108129	3.31844	70.9222
انحدار الحرف مع المعلومات المسبقة	0.107066	-4.97653	76.8703

يتبين من الجدول (6) أن النموذج الذي يمتلك اقل قيمة لكل معيار هو النموذج الذي يعطي أفضل قيم تقديرية لمعلمات النموذج وهنا نلاحظ أن نموذج الانحدار مع المعلومات المسبقة هو الأفضل .

8. الاستنتاجات

تبين من خلال النتائج التي تم الحصول عليها ما يأتي:

1. وجود علاقة خطية بين المتغيرات التوضيحية الثلاثة وهي (السطوع والإشعاع الشمسي والتبخر) مع المتغير المعتمد المتمثل بالحرارة وتم الكشف عن وجود تعدد في العلاقة الخطية من خلال أساليب الكشف عن وجود التعدد الخطي بين المتغيرات.
2. نجاح أسلوب انحدار الحرف مع المعلومات المسبقة في الحصول على نموذج يعطي قيماً تنبؤية أفضل من نماذج الصندوق الأسود وتبين ذلك في الجانب العملي من خلال قيم معايير اختبار دقة النتائج التنبؤية.
3.

9. المصادر

1. الحسو، طلال عبدالرزاق،(2009)، "المعلومات المسبقة غي انحدار الحرف"، رسالة ماجستير (غير منشورة)، كلية علوم الحاسبات والرياضيات، جامعة الموصل،العراق.

2. حياوي، هيام عبد المجيد، (2006):"تشخيص النظم الحركية الخطية التصادفية عند تغيرها مع الزمن"، أطروحة دكتوراه (غير منشورة)، كلية علوم الحاسبات والرياضيات، جامعة الموصل،العراق.

3. سعيد، هيفاء عبد الجواد،(1996):"طرق التعرف على تعدد العلاقة الخطية وكيفية معالجتهـا بطرائق التقدير المتحيزة"، رسالة ماجستير، كلية الإدارة والاقتصاد،جامعة الموصل، العراق.

4. الصفاوي،صفاء يونس،يحيى، مزاحم محمد،(2009):"استخدام انحدار الحـرف والانتروبـي العظمى العامة في تحليل التلوث لمعمل اسمنت كركوك"، المؤتمر العلمي الثاني للرياضيات- الإحصاء والمعلوماتية، المنعقد للفترة 6-2009/12/7 كلية علوم الحاسـبات والرياضـيات، جامعة الموصل،العراق.

5. القصيمي ، عزة مصطفى عبد القادر،الطالب، محاسن صـالح، الـشبخون، عبيـر أسـامة، (2007)،:"مقارنة طريقتي .H.S.L والطريقة التكرارية لتقدير معلمة الحرف"، مجلـة تنميـة الرافدين، المجلد 29، العدد 86، ص.ص171-179.

6. القصيمي،عزة مصطفى عبد القادر ،(2000):"استخدام أسلوب المحاكاة في مقارنـة مقـدرات انحدار الحرف"، رسالة ماجستير .(غير منشورة)، كلية علوم الحاسبات والرياضيات، جامعـة الموصل، العراق.

7 . كاظم،أموري هادي ،(2005):"مقدمة في القياس الاقتصادي"، الطبعة الأولى، مطبعة جامعـة الموصل، العراق.

8. النعيمي ، أسوان محمد طيب، (2005):"اختيار المتغيرات فـي انحـدار الحـرف"، رسـالة ماجستير .(غير منشورة)، كلية علوم الحاسبات والرياضيات، جامعة الموصل، العراق.

9. Ljung, L., (1999):"**System Identification Theory for user**", 2nd ed. Prentice Hall Upper Saddle River N.J. London, UK.

10. Ljung, L., (2004):"**System Identification Toolbox for use with MATLAB**", 6.0 Mathworks Inc.

11. Makridakis, S., Wheelwright, S., & Mcgee, E., (1998):"**Fore-casting :Methods and Application**s ", 2nd ed. John Wiley & Sons New York U.S.A.

12. Nelles, O., (2001):"**Nonlinear System Identification from Classical Approach to Neural Network and Fuzzy Models**", Springer Verlag Belin Heidelberg, Germany.