



نموذج إحصائي مقترح للتنبؤ بمتوسط درجات الحرارة الشهرية ومدى تأثيرها بظاهرة الاحتباس الحراري بالتطبيق على محافظة القاهرة*

إيهاب حسن إبراهيم قنصوه

طالب دراسات عليا

د. عصام فوزي عزيز

أستاذ الإحصاء التطبيقي المساعد

كلية التجارة - جامعة عين شمس

جمهورية مصر العربية

د. مدحت محمد عبد العال

أستاذ الإحصاء التطبيقي

الملخص:

تناول هذا البحث التنبؤ بمتوسط درجات الحرارة الشهرية العظمى والصغرى، مع التطبيق على محافظة القاهرة، وذلك في الفترة من يناير 1961 إلى ديسمبر 2007. وقد تم استخدام أسلوبين هما أسلوب تحليل الانحدار في استجابات متعددة وأسلوب الشبكات العصبية باستخدام طريقة الارتداد الخلفي، حيث كانت المتغيرات التابعة هي المتوسط الشهري لدرجات الحرارة العظمى والمتوسط الشهري لدرجات الحرارة الصغرى، بينما كانت المتغيرات المستقلة هي الضغط الجوي، الرطوبة النسبية، فترات سطوع الشمس، سرعة الرياح، كمية البخر (متغير يعبر عن ظاهرة الاحتباس الحراري)، كمية الأمطار الشهرية. وقد توصل الباحثون إلى أن أسلوب الشبكات العصبية يعطي تنبؤات أكثر دقة من أسلوب الانحدار في استجابات متعددة، كما أن ظاهرة الاحتباس الحراري المتمثلة في متغير كمية البخر ليس لها تأثير على متوسط درجات الحرارة العظمى أو الصغرى.

المقدمة:

تعتبر ظاهرة الاحتباس الحراري من أهم المشكلات التي تهدد المناخ والتي قد يكون لها دور في حدوث ظاهرة ارتفاع درجة حرارة الأرض، حيث تعرف هذه الظاهرة بأنها زيادة نسب تركيز الغازات الدفيئة داخل الغلاف الجوي المحيط بالأرض، ومنها بخار الماء وثنائي أكسيد الكربون، مما يؤدي إلى احتباس الحرارة داخل الغلاف الجوي للأرض، ومن ثم فإن سطح الأرض يسخن ويقوم بدوره بإطلاق الحرارة في شكل أشعة تحت الحمراء. وقد انقسم علماء الأرصاد والمناخ إلى فريقين، الفريق الأول يرى أن هذه الظاهرة طبيعية وأن مناخ الأرض يشهد فترات ساخنة وفترات باردة، مستشهدين بالفترة الجليدية التي حدثت ما بين القرن السابع عشر والقرن الثامن عشر في أوروبا، أما الفريق الثاني فيرى أن ارتفاع درجات الحرارة يرجع إلى ظاهرة الاحتباس الحراري.

هدف البحث:

يهدف هذا البحث إلى دراسة تأثير مجموعة من المتغيرات المناخية، وكذلك تأثير متغير آخر يعبر عن ظاهرة الاحتباس الحراري على المتوسط الشهري لدرجات الحرارة العظمى والصغرى، بهدف الوصول إلى نموذج إحصائي مقترح للتنبؤ بالمتوسط الشهري لدرجات الحرارة العظمى والصغرى، وكذلك تحديد أهم المتغيرات التي تؤثر على درجات الحرارة حتى يمكن ترجيح كفة أحد الفريقين.

* تم تسلم البحث في سبتمبر 2010، وقيل للنشر في يناير 2011.

متغيرات البحث:

يوجد متغيران تابعان هما المتوسط الشهري لدرجات الحرارة العظمى Y_1 والمتوسط الشهري لدرجات الحرارة الصغرى Y_2 . أما المتغيرات المستقلة فيمكن تقسيمها إلى نوعين، النوع الأول.. متغيرات مناخية وهي الضغط الجوي X_1 ، الرطوبة النسبية X_2 ، فترات سطوع الشمس X_3 ، سرعة الرياح X_4 ، كمية الأمطار الشهرية X_6 . والنوع الثاني.. متغير يقيس الاحتباس الحراري يتم التعبير عنها بكمية البحر X_5 (wikipedia.org/wik). وقد تم التعامل مع سلسله زمنية شهرية تبدأ من يناير 1961 إلى ديسمبر 2007 لمحافظة القاهرة كحالة تطبيقية.

ويمكن تعريف متغيرات الدراسة كما يلي (Koutb, 2008: 130-141):

- متوسط درجات الحرارة العظمى الشهرية Y_1 تستخدم لقياس كمية الطاقة الحرارية التي يكتسبها الهواء من الإشعاع الشمسي أو الأرضي. أما بالنسبة للتغير اليومي لدرجة الحرارة فإنها ترتفع بعد شروق الشمس وتستمر في الزيادة حتى الساعة الثالثة بعد الظهر، ثم تأخذ في الهبوط وتستمر حتى ساعات الفجر، وتصل إلى نهايتها الصغرى الساعة الثالثة صباحاً، ويلاحظ أن المتوسط اليومي لدرجات الحرارة العظمى يعتمد على ثلاث قراءات، قراءة درجة الحرارة الساعة الثانية عشرة ظهراً، قراءة درجة الحرارة الساعة الثالثة عصرًا، قراءة درجة الحرارة الساعة السادسة مساءً.
- متوسط درجات الحرارة الصغرى الشهرية Y_2 تقاس بالدرجة المئوية، ويلاحظ أن المتوسط اليومي لدرجات الحرارة الصغرى يعتمد على ثلاث قراءات، قراءة درجة الحرارة الساعة الواحدة صباحاً، قراءة درجة الحرارة الساعة الثالثة صباحاً، قراءة درجة الحرارة الساعة السادسة صباحاً.
- الضغط الجوي X_1 ويُعرف الضغط الجوي عند أية نقطة بأنه القوة المبذولة على وحدة المساحات لسطح ما عند هذه النقطة نتيجة لتصادم جزيئات الهواء مع هذا السطح، ويقاس الضغط الجوي بالمللي بار.
- الرطوبة النسبية X_2 وتُعرف بأنها مقدار ما يحتويه الجو من بخار الماء وتقدر بالجرامات في المتر المكعب من الهواء في درجة حرارة معينة. وإذا زادت كمية بخار الماء حتى يصل الهواء إلى درجة التشبع، فإن النسبة بين كتلة بخار الماء الموجود فعلاً في وحدة الحجم من الهواء إلى الكمية اللازمة لتشبع وحدة الحجم هذه في درجة الحرارة نفسها تسمى الرطوبة النسبية، حيث تتأثر هذه الرطوبة حسب قرب أو بعد المكان عن البحر. وللرطوبة النسبية أهمية خاصة في التغيرات الجوية، ويسمى جهاز قياس الرطوبة الهكروميتر.
- فترة سطوع الشمس X_3 حيث تؤثر فترة سطوع الشمس تأثيراً عكسياً على كمية الأمطار الساقطة، فكلما قلت فترة سطوع الشمس، زاد احتمال سقوط أمطار، ومقياس فترة سطوع الشمس هو الساعة والدقيقة.
- سرعة الرياح X_4 حيث تؤثر سرعة الرياح في كمية الأمطار الساقطة تأثيراً طردياً، فكلما زادت سرعة الرياح، زادت كمية الأمطار.
- كمية البحر X_5 ويُعرف البحر بأنه تحول الماء إلى حالته الغازية في درجة الحرارة العادية، ويؤثر البحر على كمية الأمطار تأثيراً عكسياً.
- كمية الأمطار الشهرية X_6 وتقاس كمية الأمطار بالمليمتر.

الأساليب الإحصائية المستخدمة في البحث:

استخدم البحث أسلوب تحليل الانحدار في استجابات متعددة Multi-responses Regression Analysis (MRRA) وأسلوب تحليل الشبكات العصبية Artificial Neural Networks (ANN) وذلك بهدف عمل دراسة مقارنة بين هذين الأسلوبين والتوصل إلى أفضل أسلوب إحصائي للتنبؤ بالمتوسط الشهري لدرجات الحرارة العظمى والصغرى.

أولاً- أسلوب الانحدار في استجابات متعددة (جونسون ووشرن، 1998):

يُعرف الانحدار الخطي في استجابات متعددة بأنه أسلوب إحصائي تتم فيه دراسة العلاقة بين مجموعة من المتغيرات المستقلة $X_1, X_2, X_3, \dots, X_r$ ومجموعة من المتغيرات التابعة $Y_1, Y_2, Y_3, \dots, Y_m$ والتي تعتبر استجابات تابعة للمتغيرات المستقلة.

ويمكن تمثيل نموذج الانحدار الخطي في استجابات متعددة بالمعادلة التالية:

$$Y_{n \times m} = X_{n \times r+1} \beta_{r+1 \times m} + \varepsilon_{n \times m} \quad (1)$$

ونجد أن توقع عنصر الخطأ يساوي "صفر"، وتباينه σ^2 كما أن تغاير عناصر الخطأ يمكن صياغتها كما يلي:

$$COV \varepsilon_i, \varepsilon_k = \sigma_{ik} I \quad i, k = 1, 2, \dots, m$$

حيث إن:

قيم المتغيرات المستقلة هي $X_{j0}, X_{j1}, X_{j2}, \dots, X_{jr} \quad j = 0, 1, 2, \dots, r$

قيم المتغيرات التابعة هي $Y_{i0}, Y_{i1}, Y_{i2}, \dots, Y_{im} \quad i = 0, 1, 2, \dots, m$

قيم الأخطاء المناظرة للمتغيرات التابعة هي $\varepsilon_{i0}, \varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{im} \quad i = 0, 1, 2, \dots, m$

ويكون نموذج الانحدار الخطي في استجابات متعددة من نموذج خطي خاص بكل استجابة (Y_i) أي أن:

$$\begin{aligned} Y_1 &= \beta_{01} + \beta_{11}X_1 + \beta_{21}X_2 + \dots + \beta_{r1}X_r + \varepsilon_1 \\ Y_2 &= \beta_{02} + \beta_{12}X_1 + \beta_{22}X_2 + \dots + \beta_{r2}X_r + \varepsilon_2 \\ &\vdots \\ &\vdots \\ Y_m &= \beta_{0m} + \beta_{1m}X_1 + \beta_{2m}X_2 + \dots + \beta_{rm}X_r + \varepsilon_m \end{aligned} \quad (2)$$

ونستطيع الحصول على المعالم المجهولة في نموذج الانحدار الخطي في استجابات متعددة بحساب تقديرات المربعات الصغرى $\hat{\beta}_i$ التي تتطلب استخدام المتغيرات المستقلة نفسها وذلك لحساب كل متغير تابع (Y_i)، ويتم توفيق النموذج ثم فحص البواقي لمعرفة مدى تبعية هذه البيانات للتوزيع الطبيعي واكتشاف القيم الشاذة، كما يتم تجميع تقديرات المربعات الصغرى بالنسبة لكل استجابة، أي لكل متغير تابع.

تقييم قدرة النموذج على التنبؤ:

يستخدم اختبار متباينة ثايل Theil's Inequality في قياس قدرة النموذج على التنبؤ ويمكن صياغة معامل متباينة ثايل كما يلي:

$$U = \sqrt{\frac{\sum d_f - d_a^2}{\sum d_a^2}} \quad (3)$$

حيث إن U معامل متباينة ثايل، d_f التغير في القيمة المتوقعة للمتغير التابع، d_a التغير في القيمة الفعلية للمتغير التابع. وتتراوح قيمة المتباينة U بين الصفر والواحد الصحيح وكلما صغرت قيمة U ، فإن ذلك يدل على ارتفاع قدرة النموذج على التنبؤ حتى إذا وصلت إلى الصفر، فإن النموذج يصل إلى حالة التنبؤ التام وعلى العكس عندما تساوي U الواحد الصحيح، فإن النموذج يعبر عن التنبؤ الرديء.

تحديد المتغيرات المستقلة في النموذج:

لتطبيق نموذج الانحدار الخطي في استجابات متعددة يجب تحليل البيانات باستخدام مقياسين أو أكثر لكل توليفة (أحد المتغيرات التابعة ومجموعة من المتغيرات المستقلة تسمى توليفة)، هذه الاختبارات تستخدم في تحديد المتغيرات المستقلة التي يمكن إدراجها في نماذج الانحدار المختلفة لمجموعة المتغيرات التابعة، وهذه الاختبارات هي:

1- اختبار Wilk's Lambda:

يستخدم اختبار Wilk's Lambda لقياس المعنوية الكلية لتحليل التباين في اتجاهين، مع وجود تفاعل داخلي، وفيه يتم اختبار كل المتغيرات التابعة بمعنى اختبار وجود فروق بين المجموعات ومدى اختلافها بالنسبة للتوليفات الخطية للمتغيرات التابعة (French and Poulsen) باستخدام المعادلة:

$$S_{tot} = S_{IV1} + S_{IV2} + S_{effect} + S_{within-grouperror} \quad (4)$$

$$\Lambda = \left| S_{error} \right| \div \left| S_{effect} + S_{error} \right| \quad (5)$$

حيث إن S_{tot} تعبر عن التباين الكلي، S_{IV1} تعبر عن متوسط المربعات بين الأعمدة، S_{IV2} تعبر عن متوسط المربعات بين الصفوف، S_{effect} تعبر عن متوسط مربعات التفاعل بين الأعمدة والصفوف، $S_{within-grouperror}$ تعبر عن متوسط مربعات البواقي، كما إن Λ هي النسبة بين متوسط مربعات البواقي ومتوسط مربعات التفاعل بين الصفوف والأعمدة مضافاً إليها متوسط مربعات البواقي للحصول على توليفة خطية للمتغيرات التابعة باستخدام المعادلة: $\eta^2 = 1 - \Lambda$ وتتراوح قيمة هذا الاختبار بين (0، 1)، حيث إن القيمة صفر تشير إلى وجود علاقة قوية بين المتغير التابع ومجموعة المتغيرات المستقلة، والقيمة واحد تشير إلى عدم وجود علاقة جوهرية بين المتغير التابع ومجموعة المتغيرات المستقلة.

2- اختبار Lawley-hotelling Trace:

يقيس هذا الاختبار النسبة بين متوسط مربعات التفاعل بين الأعمدة والصفوف ومتوسط مربعات البواقي، ويمكن حسابه باستخدام المعادلة التالية:

$$\text{Hotelling's trace} = \left| S_{effect} \right| \div \left| S_{error} \right| \quad (6)$$

3- اختبار Pillai's Trace:

يستخدم هذا الاختبار للعينات الصغيرة والخلايا غير المتساوية ذات التغاير غير المتجانس وتمثل تباين التفاعل بين الأعمدة والصفوف.

$$\text{Pillai 's criterion} = S_{\text{effect}} \quad (7)$$

وقد تم تطبيق أسلوب الانحدار في استجابات متعددة على كل من المتغيرين التابعين ومجموعة المتغيرات المستقلة، وتم الحصول على النتائج التالية:

الجدول رقم (1) الآتي يعرض نتائج تحليل التباين للنموذجين، نموذج متوسط درجة الحرارة العظمى (Y_1) ونموذج متوسط درجة الحرارة الصغرى (Y_2).

جدول رقم (1)

جدول تحليل التباين لنموذج درجة الحرارة العظمى ونموذج درجة الحرارة الصغرى

P-Value		F-Ratio		متوسط المربعات		درجات الحرية	مجموع المربعات		مصدر التباين
Y_2	Y_1	Y_2	Y_1	Y_2	Y_1		Y_2	Y_1	
0.000	0.000	347.4	544.7	3725.4	2827.2	6	22352.1	16963.2	النموذج
				10.7	5.2	557	5972.3	2890.9	البواقي
						563	28324.3	19854.1	الكلية

ويتضح من الجدول رقم (1) أن قيمة P-Value لكلا النموذجين أصغر من 0.05 وهذا يدل على أن نموذج درجة الحرارة العظمى ونموذج درجة الحرارة الصغرى جوهريان.

ولتحديد المتغيرات المستقلة صاحبة أكبر تأثير على المتغيرين التابعين، تم تصميم تحليل التباين لكل متغير مستقل كما يتضح بالجدول رقم (2) التالي:

جدول رقم (2)

نتائج تحليل التباين لجميع المتغيرات المستقلة في كلا النموذجين

Y_2				Y_1			
P-Value	F-Ratio	مجموع المربعات	المتغيرات المستقلة	P-Value	F-Ratio	مجموع المربعات	المتغيرات المستقلة
0.000	542.93	5821.41	الرطوبة النسبية X_2	0.000	407.83	2116.65	الضغط الجوي X_1
0.000	337.15	3614.94	الضغط الجوي X_1	0.000	81.21	421.5	كمية الأمطار X_6
0.000	54.15	580.65	كمية الأمطار X_6	0.000	71.38	370.45	كمية البخر X_5
0.000	22.12	237.17	كمية البخر X_5	0.000	18.12	94.04	سرعة الرياح X_4
0.001	12.21	130.97	سرعة الرياح X_4	0.019	9.74	50.54	الرطوبة النسبية X_2
0.278	1.18	12.66	فترات سطوع الشمس X_3	0.278	1.18	6.12	فترات سطوع الشمس X_3
		28324.3				2890.85	البواقي

ويتضح من الجدول رقم (2) أنه قد تم ترتيب المتغيرات المستقلة طبقاً لقيمة F-ratio تنازلياً، وذلك لتحديد ترتيب الأهمية للمتغيرات المستقلة، هذا بالإضافة إلى أن القيمة الاحتمالية P-Value إذا كانت أصغر من 0.05 فهذا يدل على جوهرية التأثير على المتغيرين التابعين، ويتضح أيضاً أن متغير فترات سطوع الشمس X_3 ليس له تأثير جوهري في كلا النموذجين، ويحذف هذا المتغير وإعادة التحليل تم الحصول على النتائج التالية الموضحة في الجدول رقم (3)، حيث يعرض هذا الجدول نتائج تحليل التباين للمتغيرات المستقلة ذات التأثير الجوهري على المتغيرين التابعين.

جدول رقم (3)

نتائج تحليل التباين للمتغيرات المستقلة ذات التأثير الجوهري في كلا النموذجين

Y ₂				Y ₁			
P-Value	F-Ratio	مجموع المربعات	المتغيرات المستقلة	P-Value	F-Ratio	مجموع المربعات	المتغيرات المستقلة
0.000	541.89	5812.09	الرطوبة النسبية X ₂	0.000	443.33	2301.64	الضغط الجوي X ₁
0.000	367.43	3940.87	الضغط الجوي X ₁	0.000	114.93	596.70	كمية البحر X ₅
0.000	54.40	583.51	كمية الأمطار X ₆	0.000	81.52	423.21	كمية الأمطار X ₆
0.000	39.39	422.49	كمية البحر X ₅	0.000	19.64	101.95	سرعة الرياح X ₄
0.000	13.42	143.97	سرعة الرياح X ₄	0.001	10.27	53.32	الرطوبة النسبية X ₂
		5984.91				2896.97	البواقي

حيث يتضح من جدول رقم (3) أن أهم المتغيرات المستقلة التي تؤثر على متوسط درجة الحرارة العظمى هي الضغط الجوي، يليه في المرتبة الثانية متغير كمية البحر، ثم يأتي في المرتبة الثالثة في التأثير كمية الأمطار، أما المرتبة الرابعة فيأتي متغير سرعة الرياح، أما المرتبة الخامسة فمتغير الرطوبة النسبية. أما بالنسبة للمتغيرات المستقلة التي تؤثر على متوسط درجة الحرارة الصغرى فنجد ترتيب المتغيرات المستقلة طبقاً للأهمية هو متغير الرطوبة النسبية، الضغط الجوي، كمية الأمطار، كمية البحر، وأخيراً سرعة الرياح.

والجدول رقم (4) التالي يعرض نتائج اختبار القدرة التفسيرية للنموذجين المقترحين باستخدام معامل التحديد، والخطأ المعياري للتقدير، متوسط الخطأ المطلق، مقياس ديربن واطسون.

جدول رقم (4)

نتائج اختبار القدرة التفسيرية للنموذجين

المؤشر	Y ₁	Y ₂
معامل التحديد	85.41%	78.87%
الخطأ المعياري للتقدير	2.28	3.28
متوسط الخطأ المطلق	1.77	2.50
مقياس ديربن واطسون	1.28	1.03
القيمة الاحتمالية	0.000	0.000

حيث يتضح من الجدول رقم (4) أن معامل التحديد لنموذج متوسط درجة الحرارة العظمى بلغ 85.41% ولنموذج متوسط درجة الحرارة الصغرى بلغ 78.87%، وهذا يفسر التغيرات التي تحدث في درجات الحرارة العظمى والصغرى والتي ترجع إلى تغيرات في المتغيرات المستقلة المذكورة في جدول رقم (3). كما أن متوسط الخطأ المطلق بلغ 1.77، 2.50 على الترتيب لكلا النموذجين، وهو يعبر عن الانحراف المعياري للبواقي، وهو يقيس حجم الخطأ في النموذج. وكلما اقتربت قيمة هذا المؤشر من الصفر، كان هذا دلالة على زيادة القدرة التفسيرية للنموذج. كما أن مؤشر ديربن واطسون الذي يقيس الارتباط الذاتي بين البواقي يشير إلى عدم وجود ارتباط ذاتي بين البواقي، حيث إن قيمة هذا الاختبار تتراوح بين $2 \pm$ كما إن القيمة الاحتمالية P-Value أصغر من 0.05 مما يؤكد عدم وجود ارتباط ذاتي بين البواقي لبيانات النموذجين.

ولتطبيق نموذج الانحدار الخطي في استجابات متعددة تم تحليل العلاقات المتداخلة بين المتغيرين التابعين ومجموعة المتغيرات المستقلة، وتحديد الملاءمة الكلية لنموذج تحليل التباين باستخدام مقياس Wilk's Lamda، و Hotelling's Trace، ومقياس Pillai's Trace كما هو موضح في الجدول رقم (5) التالي:

جدول رقم (5)

نتائج قياس معنوية اختبار تحليل التباين Multivariate test

المتغيرات/ المقياس الإحصائي	Walk's Lambda	Hotelling's Trace	Pillai Trace
الضغط الجوي X_1	0.557	0.795	0.443
الرطوبة النسبية X_2	0.132	6.594	0.868
سرعة الرياح X_4	0.965	0.036	0.035
كمية البحر X_5	0.760	0.316	0.240
كمية الأمطار X_6	0.868	0.152	0.132

حيث يتضح من جدول رقم (5) أن قيم مقياس Wilk's Lamda تتراوح ما بين صفر وواحد صحيح، مما يدل على وجود فروق بين متوسط مربعات الاختلاف بين المتغيرين التابعين والمتغيرات المستقلة X_1, X_2, X_4, X_5, X_6 مما يؤكد على وجود علاقات متداخلة بين المتغيرين التابعين Y_1, Y_2 والمتغيرات المستقلة، كما يتضح عدم وجود فروق بين متوسطات مربعات الاختلاف بين المتغيرين التابعين والمتغير المستقل X_3 والذي تم حذفه من النموذج.

وبحساب معامل متباينة ثايل لنموذج تقدير متوسط درجات الحرارة العظمى، اتضح أنه يساوي 0.013 ولنموذج تقدير متوسط درجات الحرارة الصغرى، اتضح أنه يساوي 0.027، مما يدل على ارتفاع قدرة النموذج على التنبؤ.

وقد تم استخدام المتغيرين التابعين Y_1, Y_2 والمتغيرات المستقلة X_1, X_2, X_4, X_5, X_6 للحصول على معادلة الانحدار لتقدير دالة متوسط درجات الحرارة العظمى ودالة متوسط درجات الحرارة الصغرى، فقد تم التوصل إلى النموذجين المقترحين الآتيين:

$$\text{MaxTemp} = 866.812 - 0.831X_1 + 0.024X_2 - 0.286X_4 + 0.538X_5 - 0.422X_6 \quad (8)$$

$$\text{MinTemp} = 1134.520 - 1.088X_1 - 0.256X_2 - 0.340X_4 + 0.453X_5 - 0.495X_6 \quad (9)$$

ثانياً- أسلوب الشبكات العصبية:

تعتبر الشبكات العصبية الصناعية ANN إحدى طرائق الذكاء الاصطناعي. ويرجع السبب وراء تفضيل الشبكات العصبية على طرق التحليل الإحصائي التقليدية إلى قدرة الشبكات العصبية على تحليل البيانات المشوشة التي تحتوي على قيم متطرفة، وإمكانية التحليل والتعلم من بيانات السلاسل الزمنية. ونجد أن هناك عيباً في استخدام الشبكات العصبية، وهو عدم القدرة على تطبيق الاختبارات الجوهرية الإحصائية للنماذج المقترحة وللمعاملات المقدرة (Refenes, et al., 1996)، وبالرغم من هذا العيب فإن نتائج العديد من الأبحاث أوضحت أن الشبكات العصبية يمكنها حل غالبية المشكلات إن لم يكن جميعها المشكلات بكفاءة أعلى من الطرق الإحصائية التقليدية.

ونجد أن الشبكات العصبية تتكون من طبقتين أو أكثر أو تتكون من مجموعات من عناصر التشغيل التي تسمى نيورون Neuron، وهذا المصطلح يشير إلى الوحدة الأساسية في الشبكات العصبية التي يمكن استخدامها لتشغيل البيانات. إن النيورونات تتصل بالشبكة العصبية بحيث تعتبر مخرجات أي نيورون هي المدخلات لنيورون آخر. وهذا يعني وجود اتصال بين النيورونات، وبالتالي يوجد اتجاه لهذه الاتصالات. وطبقاً لاتجاه الاتصال فإن طريقة الاتصال قد تكون اتصالاً أحاديًا One-directional أو اتصالاً ثنائيًا Bi-directional، وطبقاً لكثافة الاتصال فإن الاتصال قد يكون اتصالاً نشطاً أو اتصالاً غير نشط. ويتم تقسيم النيورونات في مجموعات تمثل الطبقات، ونجد أن هناك ثلاثة أنواع من الطبقات هي طبقة المدخلات Input Layer والطبقة الخفية Hidden layer وطبقة المخرجات Output Layer.

ويلاحظ أن مهمة طبقة المدخلات هي إستقبال البيانات من مصدر خارجي ثم إرسالها إلى الطبقة الخفية، وفي داخل الطبقة الخفية يتم تشغيل البيانات ثم إرسال نتائج هذا التشغيل إلى طبقة المخرجات، حيث تتم مقارنة مخرجات الشبكة بالمخرجات الحقيقية، وذلك لحساب مقدار الخطأ في التقديرات، ثم يلي ذلك إرجاع الخطأ الذي تم حسابه مرة أخرى إلى الشبكة ويتم تعديل أوزان الاتصال Connection Weights بين النيورونات، ثم يتم تشغيل البيانات مرة أخرى وإعادة الخطوات السابقة عدة مرات حتى يتم الحصول على مخرجات الشبكة قريبة جداً للمخرجات الحقيقية، وبهذا يتم تصغير الخطأ الذي يمثل الفرق بين مخرجات الشبكة والمخرجات الحقيقية.

وجدير بالذكر أن أوزان الاتصال تمثل قوة الاتصال بين النيورونات، فإذا كان النيورون j يتصل بالنيورون i فإن W_{ji} تمثل وزن الاتصال بين النيورون j والنيورون i ، وكذلك فإن W_{ij} هي وزن الاتصال بين النيورون i والنيورون j ، وإذا كان النيورون i متصل بالنيورونات 1، 2، 3،، n فإن أوزان الاتصال يتم حفظها في المتغيرات الآتية $W_{1i}, W_{2i}, W_{3i}, \dots, W_{ni}$. ونجد أن النيورون يستقبل العديد من المدخلات طبقاً لعدد النيورونات المتصلة به، ثم يقوم هذا النيورون باستخدام هذه المدخلات لإنتاج مخرجات واحدة ترسل إلى نيورون آخر، وذلك طبقاً لدالة التحويل المستخدمة.

ويلاحظ أن عمليات تصميم الشبكة العصبية تتكون من أربع خطوات هي (Zahedi, 1993):

- 1- تحديد النيورونات في مختلف الطبقات.
- 2- تحديد طبيعة الاتصال بين النيورونات، ونجد أن الاتصال قد يكون اتصالاً بين الطبقات Inter-layer أو اتصالاً داخل الطبقة الواحدة Intera-layer.
- 3- تحديد اتجاه استقبال المدخلات واتجاه إنتاج المخرجات.
- 4- تحديد قاعدة التعلم لتعديل أوزان الاتصال.

عند استخدام الشبكة العصبية فإن البيانات يتم تقسيمها إلى ثلاث فئات هي فئة التدريب Training Set وفئة الصلاحية Validation set وفئة الاختبار Test Set، حيث نجد أن هناك تشابهاً بين فئة الصلاحية وفئة الاختبار، ويمكن تعريف هذه الفئات كما يلي:

فئة التدريب: هي مجموعة من المشاهدات تستخدم لتقدير معاملات النموذج المقترح. وفئة الصلاحية: هي مجموعة من المشاهدات تستخدم لتعديل تقديرات المعلمات التي تم حسابها باستخدام فئة التدريب، ويمكن استخدامها أيضاً لتحديد

عدد الطبقات الخفية في الشبكة. أما فئة الاختيار: فهي مجموعته من المشاهدات تستخدم لتعميم النتائج أو لاختبار التقديرات التي تم حسابها باستخدام فئة التدريب وتم تعديلها باستخدام فئة الصلاحية (Ripley, 1996).

ويلاحظ أن الهدف من استخدام الشبكات العصبية هو الحصول على أفضل تمثيل للبيانات، وأن أبسط أسلوب للمقارنة بين الشبكات العصبية المختلفة التي يمكن بناؤها هو تقدير دالة الخطأ باستخدام مجموعة بيانات تكون مستقلة عن البيانات التي يتم استخدامها لتدريب الشبكة. وهناك العديد من الشبكات العصبية يمكن تدريبها من خلال تصغير أو تدنيه دالة الخطأ، حيث نجد أن المفاضلة بين الشبكات العصبية تتم من خلال مقارنة دالة الخطأ، وذلك باستخدام فئة الصلاحية وأن الشبكة صاحبة أقل خطأ يتم اختيارها، ونجد أن هذا الإجراء يمكن أن يؤدي إلى ما يسمى بمشكلة Over Fitting لفئة الصلاحية، وفي هذه الحالة يتم استخدام الفئة الثالثة من البيانات وهي فئة الاختبار للمفاضلة بين الشبكات العصبية (Bishop, 1995: 372).

إن أسلوب الارتداد الخلفي أو التغذية الخلفية هو إحدى الطرائق التي يمكن استخدامها في تطبيقات كثيرة مختلفة، حيث إن الهدف من استخدام هذا الأسلوب هو حساب الخطأ في طبقة المخرجات ثم إعادته مرة أخرى لكل طبقة خفية لتعديل أوزان الاتصال حتى يتم الوصول إلى طبقة المدخلات (Dhar and Stein, 1996). ويهدف هذا الأسلوب إلى تدنيه عنصر الخطأ الذي يتمثل في الفرق بين مخرجات الشبكة والمخرجات الحقيقية، ثم بعد حساب الخطأ بين مخرجات الشبكة والمخرجات الحقيقية يتم تغذية هذا الخطأ في اتجاه عكسي (عكس اتجاه المدخلات) للشبكة لتعديل الأوزان، وبالتالي يتم تخفيض الخطأ (Fahlmann, 1988: 88-162)، ويتم إعادة هذه الإجراءات حتى يصل الخطأ إلى أقل قيمة ممكنة. ويمكن استخدام المعادلة رقم (10) لتعديل الأوزان W_{ij} من العقدة (النيورون) N_i إلى العقدة أو النيورون: N_j

$$W_{ij,(t+1)} = W_{ij,t} + \eta \delta_j \bar{N}_i \quad (10)$$

حيث إن: W_{ij} تمثل وزن الاتصال بين النيورون N_i والنيورون N_j ، t تمثل عدد مرات تعديل الأوزان في الشبكة، η تمثل معدل التعلم (أحد معاملات التعلم في الشبكة)، δ_j تمثل حساسية النيورون N_j لتغير الوزن W_{ij} . (معدل التعلم هو حدود تعديل أوزان الاتصال، فإذا كان معدل التعلم كبيراً، فهذا يؤدي إلى عدم ثبات الشبكة. وإذا زاد معدل التعلم بقدر كبير، فإن قدرة الشبكة على التنبؤ تصبح منخفضة، أما إذا كان معدل التعلم صغيراً للغاية، فهذا يعتبر ضياعاً للوقت، حيث تستغرق الشبكة وقتاً أكثر لإجراء عملية التعلم (Garson, 1991: 398-434). ونجد أن زيادة معدل التعلم تعتبر إجراء مفيداً لزيادة قدرة الشبكة على التعلم، وبالتالي يصبح من الممكن تعديل الأوزان). وبلي ذلك حساب إجمالي المدخلات للنيورون N_j كما يلي:

$$S_j = \sum_i N_i W_{ij} \quad (11)$$

حيث إن: S_j تمثل مجموع جميع المدخلات لنيورون N_j ، N_i تمثل مخرجات النيورون السابق، W_{ij} تمثل وزن الاتصال بين النيورون i ، والنيورون j في الطبقة السابقة. يلي خطوة حساب إجمالي المدخلات للنيورون i تحويل المخرجات باستخدام دالة التنشيط أو التحويل اللوجستي والتي يمكن صياغتها كما يلي:

$$N_j = \frac{1}{1 + e^{-S_j}} \quad (12)$$

حيث إن: N_j تمثل إجمالي المخرجات من النيورون j ، ونجد أن الخطأ الكلي للبيانات التي مرت خلال الشبكة يمكن حسابه كما يلي:

$$E = \frac{1}{2} \sum_{Output} N_j - D_j^2 \quad (13)$$

حيث إن: D_j تمثل المخرجات الحقيقية للنيورون j ، هذا الخطأ الذي تم حسابه يتم إرجاعه إلى الشبكة في اتجاه عكسي حتى يتم تخفيض الخطأ، وهذه العملية يمكن تنفيذها من خلال أربع خطوات هي:

(1) يتم حساب الخطأ لمخرجات النيورون والتي يرمز لها بالرمز O_j ونجد أن هذا الخطأ يمكن حسابه باستخدام المعادلة التالية:

$$\varepsilon O_j = N_j - D_j \quad (14)$$

(2) إن مقدار الخطأ يتغير كلما تغيرت مدخلات النيورون لمخرجات نيورون آخر، وهذا المقدار يجب حسابه، ويمكن تحديد هذا المقدار بمقدار التغير في المعادلة رقم (13) عندما يتغير إجمالي المدخلات في المعادلة رقم (11) والتي يمكن التعبير عنها رياضياً كما يلي:

$$\varepsilon S_j = \varepsilon O_j N_j - N_j \quad (15)$$

(3) يتم حساب مقدار تعديل الأوزان W_{ij} من النيورون N_i في الطبقة التي تقع أسفل الطبقة الحالية إلى النيورون N_j الموجود في الطبقة الحالية كما يلي:

$$\varepsilon W_{ij} = \varepsilon S_j N_j \quad (16)$$

(4) يتم إعادة هذه العملية من النيورون في الطبقات الدنيا، وذلك بالسماح للنيورونات في الطبقة الخفية لعرض مخرجات الطبقة، ثم يتم حساب إجمالي الخطأ من جميع المدخلات إلى الطبقة الخفية. هذا بالإضافة إلى أنه يتم حساب مقدار الخطأ في النيورون الخفي عن طريق حساب التغير في النيورون في الطبقة التي تقع أعلى الطبقة الخفية. ويمكن حساب مقدار تعديل الأوزان باستخدام المعادلة التالية:

$$\varepsilon H_i = \sum_j \varepsilon S_j W_{ij} \quad (17)$$

حيث إن: H_i هو الفرق بين وزن الاتصال في خطوة ووزن الاتصال في الخطوة التالية والذي يمكن حسابه باستخدام المعادلة التالية:

$$H_i = W_{ij(1)} - W_{ij(2)} \quad (18)$$

ونجد أن الخطأ في الشبكة يتم إرجاعه عدة مرات حتى يتم تخفيض الخطأ الكلي لمخرجات الشبكة (Dhar and Stein, 1996).

وقد تم بناء شبكة عصبية باستخدام أسلوب الارتداد الخلفي Back Propagation، وتم التوصل إلى النتائج التالية، حيث يعرض الجدول رقم (6) التالي بعض المقاييس التي يمكن استخدامها كمؤشر لاختلاف الوسط الحسابي للقيم المُشاهدة والقيم المتوقعة.

جدول رقم (6)

نتائج تطبيق الشبكات العصبية لنموذج متوسط درجات الحرارة العظمى والصغرى

متوسط درجات الحرارة الصغرى		متوسط درجات الحرارة العظمى		المقياس
فئة الاختبار	فئة التدريب	فئة الاختبار	فئة التدريب	
0.11	0.52	6.37	6.75	معامل الاختلاف
7.65	162.73	248.33	269.37	متوسط مربعات الخطأ
2.08	11.44	157.49	158.76	متوسط الخطأ المُطلق
10.49	42.85	681.97	733.05	نسبة متوسط الخطأ المُطلق
%00	%84.89	%00	%88.45	نسبة التباين
%89.98		%88.87		معامل التحديد

ومن المعروف أن جذر متوسط مربعات الخطأ (RMSE) Root Mean Square Error يعتبر مقياساً لمتوسط الخطأ، والخطأ هو الفرق بين القيم المُشاهدة والقيم المتوقعة، ويتم تربيع هذه الفروق ثم حساب متوسطها، ثم يتم حساب الجذر التربيعي لها. وحيث إن الخطأ قد تم تربيعه قبل حساب المتوسط فإن RMSE يعطى وزناً نسبياً أكبر للأخطاء أو الفروق الكبيرة، ولهذا يُفضل استخدام هذا المقياس في حالة الفروق أو الأخطاء الكبيرة. ويلاحظ أيضاً أن جذر مربعات الخطأ بعد تحويله إلى توزيع طبيعي (NMSE) Normalized Mean Square Error هو RMSE مقسوماً على مدى البيانات المُشاهدة، وكلما انخفضت قيمة هذا المقياس NMSE دل ذلك على انخفاض تباين البواقي.

أما متوسط الخطأ المُطلق Mean Absolute Error (MAE) فهو مقياس لمتوسط الخطأ أو الفرق بين القيم المُشاهدة والقيم المتوقعة مع إهمال اتجاه الخطأ، وهذا المقياس خطي، أي أن جميع الفروق لجميع المُشاهدات لها نفس الوزن عند حساب الوسط الحسابي للفروق أو الأخطاء. ونجد أن مقياس نسبة التباين Proportion of Variance (Var%) يشرح التباين الذي يمكن تفسيره بواسطة النموذج ويحدد نسبة تطابق القيم المتوقعة مع القيم المُشاهدة، فإذا كان التطابق تاماً، فهذا يعني أن هذه النسبة تساوي 0%.

كما يتضح أن البيانات قد تم تقسيمها إلى ثلاث فئات هي فئة التدريب وفئة الصلاحية وفئة الاختبار، وتم تسجيل نتائج فئة التدريب وفنتي الصلاحية والاختبار معاً لكل نموذج على حدة، كما يتضح أن قيم هذه المقاييس قد انخفضت في فنتي الصلاحية والاختبار عنها في فئة التدريب، كما أن نسبة التباين قد ارتفعت وهذا يفسر على أن الشبكة قد أدت إلى الحصول على توقعات أكثر دقة.

ولتحديد أهم المتغيرات التي تؤثر على المتغيرين التابعين، تم استخدام مقياس أهمية المتغيرات Variable Importance كما يتضح بالجدول رقم (7) التالي:

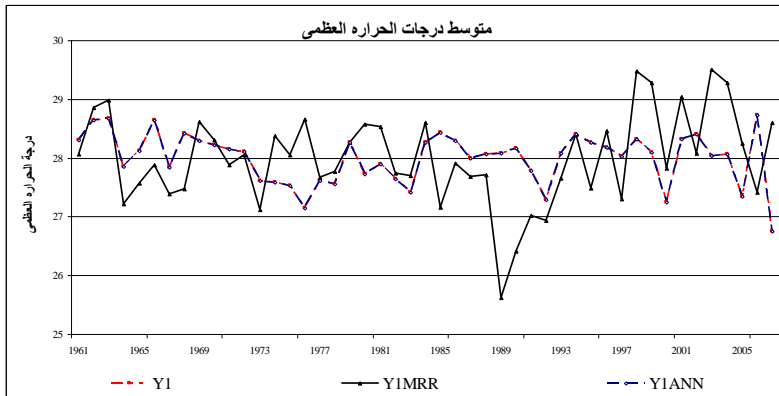
جدول رقم (7)
مقياس أهمية المتغيرات لكل نموذج

متوسط درجات الحرارة الصغرى		متوسط درجات الحرارة العظمى	
الأهمية النسبية	المتغيرات	الأهمية النسبية	المتغيرات
100	الضغط الجوي	100	الضغط الجوي
22.37	البخر	47.67	الرطوبة النسبية
17.23	الأمطار	11.41	الأمطار
2.37	الرطوبة النسبية	8.95	البخر
2.02	سرعة الرياح	3.64	سرعة الرياح
0.89	فترات سطوع الشمس	2.58	فترات سطوع الشمس

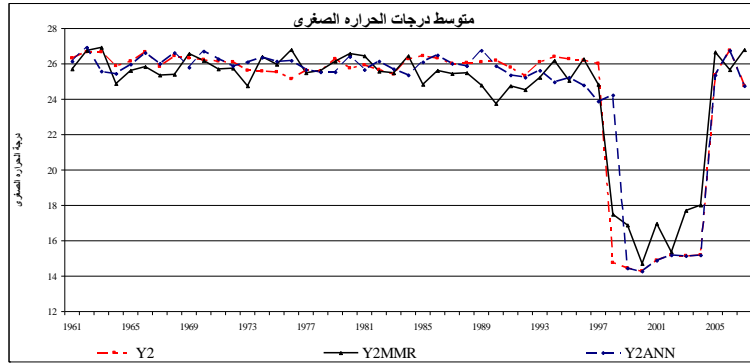
حيث يتضح من الجدول رقم (7) أن أهم المتغيرات المستقلة صاحبة أكبر تأثير على المتغير التابع هي التي تأخذ أهمية نسبية 100%، كما يتم تجاهل المتغيرات المستقلة التي تقل أهميتها النسبية عن 40% (Thomas, et al., 2007: 61-91)، ومن ثم نجد أن أهم المتغيرات التي تؤثر على متوسط درجات الحرارة العظمى هي متغير الضغط الجوي ومتغير الرطوبة النسبية، كما أن متوسط درجات الحرارة الصغرى تتأثر فقط بالضغط الجوي.

النتائج والتوصيات:

مما تقدم نجد أنه تم استخدام أسلوب الانحدار في استجابات متعددة وأسلوب الشبكات العصبية بطريقة الارتداد الخلفي، وذلك للتوصل إلى أفضل نموذج يمكن استخدامه للتنبؤ بمتوسط درجات الحرارة العظمى ومتوسط درجات الحرارة الصغرى في محافظة القاهرة، فقد اتضح أن معامل التحديد لنموذج متوسط درجات الحرارة العظمى باستخدام أسلوب الانحدار في استجابات متعددة قد بلغ 85.41% في حين أن معامل التحديد لنموذج الشبكات العصبية بلغ 88.87%، وبالنسبة لنموذج متوسط درجات الحرارة الصغرى فقد بلغ معامل التحديد 78.87% في نموذج الانحدار في استجابات متعددة وبلغ 89.98% في نموذج الشبكات العصبية، وهذا يعطى مؤشراً قوياً على أن استخدام أسلوب الشبكات العصبية سوف يعطى تنبؤات أكثر دقة. كما تمت المقارنة بيانياً بين القيم المُشاهدة والقيم المتوقعة باستخدام كل أسلوب من أساليب التحليل كما يتضح من الشكلين التاليين:



شكل رقم (1): المقارنة بين القيم المُشاهدة والتوقعات باستخدام أسلوب الانحدار وأسلوب الشبكات لمتوسط درجات الحرارة العظمى



شكل رقم (2): المقارنة بين القيم المُشاهدة والتوقعات باستخدام أسلوب الانحدار وأسلوب الشبكات لمتوسط درجات الحرارة الصغرى

ومن الشكلين السابقين يتضح أن أسلوب الشبكات العصبية يعطي توقعات أكثر دقة من أسلوب الانحدار الخطي في استجابات متعددة. أما بالنسبة إلى تأثير ظاهرة الاحتباس الحراري على متوسط درجات الحرارة والتي تم التعبير عنها باستخدام متغير كمية البحر، فقد توصل الباحثون إلى أن هذا المتغير ليس له تأثير على كل من متوسط درجات الحرارة العظمى ومتوسط درجات الحرارة الصغرى، ومن ثم فإنه يمكن إرجاع ظاهرة ارتفاع درجات الحرارة إلى أن الأرض تمر بفترات ساخنة وفترات باردة وليس بسبب الاحتباس الحراري.

وبناءً على ما تقدم، يوصي الباحثون بالاعتماد على نماذج الشبكات العصبية للتنبؤ بالمتوسط الشهري لدرجات الحرارة العظمى والصغرى، وذلك باستخدام المتغيرات المستقلة الأكثر تأثيراً على متوسط درجات الحرارة وهذه المتغيرات هي:

1- الضغط الجوي والرطوبة النسبية بالنسبة للتنبؤ بمتوسط درجات الحرارة العظمى.

2- الضغط الجوي فقط بالنسبة للتنبؤ بمتوسط درجات الحرارة الصغرى.

وبالطبع يمكن استخدام هذه النماذج للتنبؤ بدرجات الحرارة اليومية أو بكمية الأمطار أو أي من الظواهر الطبيعية.

المراجع

أولاً- مراجع باللغة العربية:

- جونسون، ريتشارد؛ دين وشرن. (1998). "التحليلات الإحصائية للمتغيرات المتعددة من وجهة التطبيقية". ترجمة د/ عبد المرضي حامد عزام. الرياض: دار المريخ للنشر.

ثانياً- مراجع باللغة الأجنبية:

- Ahmadian, M. H. (1995). "Artificial Neural Networks: A New Tool in Technology", *Tech Directions*, 54: 9, April. pp 33-36.
- Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press, p 372.
- Cheng, B. and D. M. Titterington. (1994). "Neural Networks: A Review from a Statistical Perspective", *Statistical Science*, 9, pp 2-54.
- Dhar, V. and R. Stein. (1996). *Intelligent Decision Support Methods: The Science of Knowledge Work*. New Jersey: Prentice Hall Business Publishing, Upper Saddle River.
- Dillon, W. R. and M. Goldstein. (2004). *Multivariate Analysis: Methods and Applications*. New York: John Wiley.
- Fahlmann, S. E. (1988). *An Empirical Study of Learning Speed in Back-propagation Networks*. CMU Technical Report, CMU-CS- June, pp. 88-162.
- French, Aron and John Poulsen. *Multivariate Analysis of Variance*, Online.sfsu.edu/~efc/classes/biol710/manova/manova
- Garson, G. D. (1991). "A Comparison of Neural Network and Expert Systems Algorithms with Common Multivariate Procedures for Analysis of Social Science Data", *Social Science Computer Review*, fall, 9 (3), pp. 398-434.
- Kaastra, I. and M. S. Boyd. (1995). "Forecasting Futures Trading Volume Using Neural Networks", *Journal of Futures Markets*, 15 (8), December, pp. 953-970.
- Koutb, Aly. *Journal of Meteorology*, Vol. 14, July 2008, Cairo, Egypt, pp 130-141.
- Kuan, C. M. and H. White. (2002). "Artificial Neural Networks: An Econometric Perspective", *Econometric Reviews*, 13, pp. 81-91.
- Ormerod, P.; J. C. Taylor and T. Walker. (1991). "Neural Networks in Economics", In: M. P. Taylor, *Money and Financial Markets*, Cambridge, Mass and Oxford: Blackwell, pp. 341-53.
- Refenes, A. N.; Y. Abu-Mostafa; J. Moody and A. Weigend. (1996). "Neural Networks in Financial Engineering", Proceedings of: *The Third International Conference on Neural Networks in Capital Markets*, World Scientific.
- Richard, A. Johnson, and Dean W. Wichern. (2002). *Applied Multivariate Statistical Analysis*, 5th Ed., New Jersey: Prentice Hall.
- Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press.
- Shtub, Avraham and Ronen Versano. (1999). "Estimating the Cost of Steel Pipe Bending, A Comparison between Neural Networks and Regression Analysis", *International Journal of Production Economics*, Vol. 62 (3), September, pp 201-07

- Thomas, D. R.; P. Zhu and Y. J. Decady. (2007). “Point Estimates and Confidence Intervals for Variable Importance in Multiple Linear Regression”, *Journal of Educational and Behavioral Statistics*, Vol. 32; Num B 1, pp. 61-91.
- Tibshirani, R. (1996). “A Comparison of Some Error Estimates for Neural Network Models”, *Neural Computation*, 8, pp. 152-163.
- Verkooijen, W. (2003). “A Neural Network Approach to Long-run Exchange Rate Prediction”, *Computational Economics*, 9, pp 51–65.
- wikipedia.org/wik
- Wong, F. S. (1999). “Time Series Forecasting Using Back Propagation Neural Networks”, *Neurocomputing*, 2, pp. 147–159.
- Zahedi, F. (1993). *Intelligent Systems for Business, Expert Systems with Neural Networks*. Wodsworth Publishing Inc., 1993.

**Proposed a Statistical Model to Predict
Average Monthly Temperatures and the Extent Affected by the
Phenomenon of Global Warming in Cairo Governorate**

Dr. Medhat M. Abdel Aal
Professor
of Applied Statistics

Dr. Essam F. Aziz
Assistant Professor
of Applied Statistics

Ihab H.I. Qansuh
Graduate student

Faculty of Commerce - Ain Shams University
Arab republic of Egypt

ABSTRACT

This research addressed the problem of forecasting an average monthly temperature, maximum and minimum with the application on the governorate of Cairo, in the period from January 1961 to December 2007.

Two statistical methods are used; Multi-responses Regression Analysis (MRRRA) and Artificial Neural Networks (ANN) method of Back Propagation. Where the dependent variables are the average monthly temperatures, maximum and minimum, while the independent variables are the air pressure, relative humidity, periods of sunshine, wind speed, the amount of evaporation (variable reflects the phenomenon of global warming), the amount of monthly rainfall.

The research found that the neural networks method gives more accurate forecasting comparing with Multi-Responses Regression Analysis, and the variable amount of evaporation does not substantially affect the average temperature maximum or minimum.