

AL- Rafidain
University College

PISSN: (1681-6870); EISSN: (2790-2293)

مجلة كلية الرافدين الجامعة للعلوم

Available online at: <https://www.jrucs.iq>

JRUCS

Journal of AL-Rafidain
University College
for Sciences

تحديد المتغيرات المؤثرة على التلوث البيئي بغازي CO و SO₂ باستعمال أنموذج الانحدار Log-Logistic

أ.د انتصار عريبي فدعم	سماح صباح حسن
entsar_arebe@coadec.uobaghdad.edu.iq	Samahsabab1993@gmail.com
كلية الادارة والاقتصاد، جامعة بغداد، بغداد، العراق	وزارة الداخلية، بغداد، العراق

المستخلص

تعد مشكلة التلوث البيئي من المشاكل المهمة والخطيرة التي تواجه الانسان في الوقت الراهن لما لها من اثر مباشر على صحة الانسان والكائنات الحية الاخرى . فقد لوحظ في السنوات الاخيرة حدوث زيادة بمعدلات التلوث البيئي مما اثر بشكل كبير على صحة الانسان وادى الى ظهور العديد من الامراض مثل الامراض السرطانية والالتهابات الرئوية والتسمم وتشوه المواليد وغيرها ، كما ان تأثيره لم يقتصر على الانسان فحسب بل تعدى تأثيره على الحيوان والنبات ايضا .

ونظراً لأهمية الموضوع وخطورته وتأثيره المباشر على حياة الانسان ثم عمل هذا البحث لغرض معرفة نسبة مقدار التلوث بعاملين من اهم العوامل على تلوث الهواء وهو كل من غاز (CO و SO₂) بالاعتماد على المتغيرات التفسيرية هما متوسط درجات الحرارة ومتوسط نقطة الندى ومتوسط الرطوبة ومتوسط سرعة الرياح ومتوسط كمية النفط الخام المستعمل في عملية التصفية .

في هذا البحث تم تقدير دالة المخاطرة انموذج الانحدار Log-Logistic بأستعمال طرائق التقدير وهي طريقة تصغير مربع كاي وطريقة المربعات الصغرى الموزونة واخيراً طريقة الخوارزمية الجينية ، وتم تطبيق هذه الطرق على بيانات تلوث الهواء التي تم الحصول عليها من شركة مصافي الوسط في بغداد (مصفى الدورة) والتي تمثل القياسات اليومية لمركبات التلوث البيئي والتي تستند الى الوقت للمدة من 2018 الى 2022 .

معلومات البحث

تواريخ البحث:

تاريخ تقديم البحث: 25/2/2024
تاريخ قبول البحث: 12/4/2024
تاريخ رفع البحث على الموقع:
31/12/2024

الكلمات المفتاحية:

توزيع log-logistic، طريقة تصغير مربع كاي، طريقة المربعات الصغرى الموزونة، طريقة الخوارزمية الجينية، دالة المخاطرة، GLM

للمراسلة:

سماح صباح حسن

Samahsabab1993@gmail.comDOI: <https://doi.org/10.55562/jrucs.v56i1.22>

1. المقدمة

يلخص مفهوم التلوث pollution مختلف التهديدات للبيئة التي يتعرض لها الأفراد وأصبحوا في كثير من الأحيان أكثر ألفة بها حيث إن التلوث حالة من عدم النقاء أو عدم النظافة أو انه كل عملية تنتج هذه الحالة كما يعرف التلوث على انه تغير نوعي أو كمي في المكونات البيئية على إن يكون هذا التغير خارج مجال التذبذبات الطبيعية لأي من المكونات بحيث يؤدي إلى حدوث اختلال في اتزان الطبيعة مما يؤدي إلى تأثير مباشر أو غير مباشر على النظام البيئي . فالتلوث هو أي تغير في الخواص الطبيعية للبيئة يسبب اضرار بحياة الإنسان أو غيره من الكائنات كما يمكن تعريفه بأنه إضافة أو إدخال أي مادة غير مألوفة إلى البيئة مما يترتب عليه حدوث تغير في خواص هذه البيئة [11] .

تلوث الهواء الناجم عن استخدام نوعي الوقود المستخدم لتشغيل المولدات (البنزين والديزل) ونواتج احتراق هذا الوقود المتعددة .. مركبات الهيدروكربونات، اكاسيد الكبريت، اكاسيد النيتروجين، اكاسيد الكربون، الدقائق المادية والعناصر الثقيلة خصوصاً الرصاص وجميعها لها اضرار متنوعة على الصحة العامة وحياة الحيوان والنبات والممتلكات والبيئة ومن ابرز الاضرار على الصحة هي .. تخديش الجهاز التنفسي، تفاقم امراض القلب، الحساسية، بعض امراض العيون، التأثير على النمو الجسماني والذهني للأطفال أما التلوث الثاني فهو التلوث الضوضائي والذي يؤثر سلباً على صحة الانسان من النواحي البدنية والنفسية والعصبية ويتسبب في ضعف السمع وامراض القلب وتصلب الشرايين والاورام ونقص المناعة وسكر الدم وغيرها. اما

الأضرار غير المباشرة فتشمل تلوث المياه بالزيوت والمشتقات النفطية المتسربة جراء التشغيل وتلوث التربة والغطاء النباتي من جراء تسربات الوقود والزيوت ورمي فضلات ادامة المولد.

ان الهدف الاساسي هو تحليل العلاقة بين مجموعة من المتغيرات والحصول على افضل الصيغ لوصف الانموذج في حالة ان يكون متغير الاستجابة من النوع المتقطع لا يمكننا تطبيق الانموذج الخطي وانما نلجأ الى نماذج بديلة اخرى ، نذكر منها انموذج الانحدار Log-Logistic وهو من النماذج الاحصائية المهمة وهو حالة خاصة من حالات نماذج الانحدار الخطية المعممة ويسمى احياناً لوجت اللوغارتمي والذي لايتطلب افتراضات كثيرة ولايوجد تعدد خطي بين المتغيرات التوضيحية وان تكون حجم المشاهدات في كل مجموعة اكبر من خمس مرات عدد المعلمات [12] . ويكون على عدة انواع: الانحدار Log-Logistic الثنائي ويشمل هذا النوع عندما يأخذ متغير الاستجابة قيمتين مثلاً تأثير تلوث الهواء على جهاز التنفسي لكائنات الحية (وجود تلوث، عدم وجود تلوث) والانحدار Log-Logistic المتعدد يتم استعماله عندما يكون متغير الاستجابة متعدد القيم (اكثر من قيمتين) مثلاً ممارسة الرياضة (دائماً، احياناً، نادراً) والانحدار Log-Logistic الترتيبي وفيه يكون متغير الاستجابة رتبوي مثلاً مرحلة الدراسة الابتدائية (الاول، الثاني، ...، السادس) وفي هذا البحث سنركز على النوع الاول .

وتوجد عدة طرائق لتقدير دالة المخاطرة ومنها طريقة تصغير مربعات كاي (MCSE) من اجل الحصول على قيم المعلمة المثلى (افضل نتيجة تقدير) من خلال تقليل مجموع مربعات الخطأ العشوائي باستخراج المشتقة الاولى ومساواتها للصفر، وكذلك استعملنا طريقة المربعات الصغرى الموزونة (WLS) التي لاتحتاج الى قيم اولية، واخيراً طريقة الخوارزمية الجينية والتي تعد احد التقنيات التطويرية لبحث العشوائي التي تعالج مشكلة ما من اجل الوصول الى افضل النتائج الممكنة .

يشكل التلوث تحدياً رئيسياً في جميع انحاء العالم مع معظم المنظمات التي تسعى الى الحد من معدل التلوث في سبيل التخفيف من ظاهرة الاحتباس الحراري واحد من اهم المسببات التلوث المعروفة هي شركة انتاج ومعالجة وتكرير النفط التي تعد من بين الدوافع الرئيسية لتلوث الهواء والتربة والمياه والتلوث الاشعاعي على حد سواء بسبب العمليات المكثفة الطاقة المطلوبة في انتاج النفط .

في حالة البيانات الخاصة بالتلوث تم استعمال انموذج الانحدار Log-Logistic لمتغير الاستجابة الثنائي (1 وجود تلوث، 0 عدم وجود تلوث)، مع البيانات الطولية المصنفة الى طبقات ومن ثم الى مجموعات فرعية اكثر تجانساً يعد من النماذج التي ستعالج مثل هكذا بيانات وان تقدير مثل هكذا نماذج يكون من الصعوبة حلها لوجود ارتباطات داخل المشاهدات ضمن طبقات او العناقيد، لهذا يجب تحديد بنية الارتباط الصحيحة مما يساهم في ايجاد مقدرات موثوقة وغير متحيزة واكثر تقارباً

2. انموذج الانحدار log-logistic :

يعتبر التوزيع log-logistic بديلاً ممتازاً لتوزيع Weibull للنمذجة المعلمية لبيانات المخاطرة وحيث يكون معدل الخطر غير رتيب و يكون من المرغوب ايضاً تقارب دالة الخطر لعينات المنفصلة ويمكن تركيب النموذج على البيانات الخاضعة للرقابة من الجهة اليمنى واليسرى. ويمكن كتابة دالة التوزيع log-logistic لعينة واحدة على النحو التالي[20] :

$$F(y_i, \varphi, \theta) = \frac{1}{1 + y_i^{-\varphi} e^{-\theta}} \quad 0 \leq y_i \leq \infty ; \varphi > 0 \quad (1)$$

اذ ان :

y_i : يمثل المتغير الترتيبي (الوقت)

θ : يمثل معلمة القياس للموقع

φ : يمثل معلمة الشكل

بالنسبة لنموذج الانحدار الذي يشمل العديد من العينات سيكون φ هو نفسه لعينات بينما سيتم وضع θ بالشكل الاتي :

$$\theta_s = \beta_0 + \beta_1 y_{s1} + \dots + \beta_p y_{sp} = \beta' y_s \quad (2)$$

لجميع بيانات مخاطر التلوث في العينة (S=1,2,3,...,s) و y_s تمثل المتجه للمتغيرات المشتركة التي تحدد العينة β هو متجه لمعلومات غير معروفة، احتمالية الوفاة حسب الوقت y لفرد العينة S_{th} .

ودالة $F(y_i, \varphi, \theta)$ في معادلة رقم (1) يمكن كتابتها بدلالة $F_s(y)$

$$\log \left\{ \frac{F_s(y)}{1 - F_s(y)} \right\} = \varphi \log y + \beta' y_s \quad (3)$$

لتصبح نموذجاً خطياً لأحتمالات اللوغارتميات عند الوفاة بوجود المتغير الترتيبي y بالنسبة لعينتين مع متجهات من المتغيرات المشتركة y_1, y_2 والفرق بين دالة الاحتمالية يكون كالآتي :

$$\log \left\{ \frac{F_1(y)}{1 - F_1(y)} \right\} - \log \left\{ \frac{F_2(y)}{1 - F_2(y)} \right\} = \beta' (y_1 - y_2) \quad (4)$$

مستقل عن الوقت y_i وبالتالي فإن احدى دالة الاحتمالية هي مضاعف للأخرى بالنسبة لعينتين مع اختلاف وحدة في المتغير المشترك (جميع الاختلافات المتغيرة الاخرى تساوي صفر) فإن هذا الثابت يساوي e^{β_i} ويمكن التعرف على β_i على انه نسبة الاحتمالات اللوغارتميات هذه حالة خاصة لنموذج الاحتمالات النسبية في نموذج الاحتمالات النسبية العامة يمكن استبدال $\varphi \log y$ بأي دالة [14] y .

ويمكن اعطاء دالة المخاطرة للعينة بالشكل التالي :

$$h_s(y_i) = \frac{\varphi}{y_i(1 + y_i^{-\varphi} e^{-\theta_s})} \quad (5)$$

3. طرائق التقدير

3.1. طريقة تصغير مربع كاي

من الطرائق الشائعة الاستعمال في التقدير وتعتمد على تصغير احصاء مربع كاي لبيرسون المعروفة وحسب الصيغة الرياضية الآتية :

$$\chi^2 = R(\beta) = \sum_{i=1}^N \frac{(Y_i - \pi_i)^2}{\pi_i} \quad (6)$$

اذ ان :

Y_i : تمثل القيمة المشاهدة عند المستوى i

π_i : تمثل القيمة المتوقعة عند المستوى i

$R(\beta)$: تمثل احصاء مربع كاي لبيرسون

وفي حالة الانموذج log-logistic الثنائي الاستجابة فأن :

$$R(\beta) = \sum_{i=1}^N \frac{(Y_i - \pi_i)^2}{\pi_i} + \frac{[(1 - Y_i)(1 - \pi_i)]^2}{(1 - \pi_i)} \quad (7)$$

ويتم اختصارها الى

$$R(\beta) = \sum_{i=1}^N \frac{(Y_i - \pi_i)^2}{\pi_i(1 - \pi_i)} \quad (8)$$

وبعد تعويض قيمة $(\pi_i, 1 - \pi_i)$ في المعادلة اعلاه نستنتج :

$$R(\beta) = \sum_{i=1}^n \left[\left(Y_i - \frac{e^{\hat{\chi}_i \beta}}{1 + e^{\hat{\chi}_i \beta}} \right)^2 \frac{(1 + e^{\hat{\chi}_i \beta})^2}{e^{\hat{\chi}_i \beta}} \right] \quad (9)$$

وبعد التبسيط تصبح المعادلة كالآتي :

$$R(\beta) = \sum_{i=1}^n [Y_i^2 e^{-\hat{\chi}_i \beta} + (1 - Y_i)^2 e^{-\hat{\chi}_i \beta} - 2Y_i(1 - Y_i)] \quad (10)$$

لايجاد β التي تعطي $R(\beta)$ يتطلب ايجاد المشتقة الاولى ومساوتها بالصفر كما يأتي :

$$\frac{\partial R(\beta)}{\partial \beta} = \sum_{i=1}^n \hat{X}_{ij} \left[(1 - Y_i)^2 \left(\frac{\pi_i}{1 - \pi_i} \right) - (Y_i)^2 \left(\frac{1 - \pi_i}{\pi_i} \right) \right] \quad (11)$$

وهي علاقة غير خطية وعند حلها يتطلب استعمال احدى الطرائق التكرارية مثلاً طريقة نيوتن - رافسون لايجاد تقديرات β حسب الصيغة الآتية :

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} + (\hat{X}_{v(m)} X)^{-1} \hat{X} (Y - P^{(m)}) \quad (12)$$

$$v^{(m)} = \text{diag} \left[(Y_i)^2 \left(\frac{\pi_i}{1 - \pi_i} \right) + (1 - Y_i)^2 \left(\frac{\pi_i}{1 - \pi_i} \right) \right] \quad (13)$$

3.2. طريقة المربعات الصغرى الموزونة

يعتبر انموذج الانحدار Log-Logistic حالة خاصة من النماذج الخطية العامة التي تكون امتداد للانموذج الخطي البسيط.

يمكن كتابة انموذج الانحدار الخطي العام كالآتي :

$$\theta_s = \beta_0 + \beta_1 y_{s1} + \dots + \beta_p y_{sp} + e_i \quad (14)$$

$$\theta_s = \beta_0 + \sum_{j=1}^p \beta_j y_{ij} + e_i, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, p \quad (15)$$

ويتم تقدير معاملات انموذج الانحدار Log-Logistic بأستعمال طريقة المربعات الصغرى الموزونة (WLS) للحصول

على افضل تقدير

$$W_i = \pi_i(1 - \pi_i) \quad (16)$$

اذ ان :

W_i : تمثل مصفوفة التباينات وهي الاوزان المختارة للمستوى i

فيكون تقدير المعلمات $(\beta_0, \beta_1, \dots, \beta_p)$ حسب (WLS) بايجاد القيم التي تجعل الفرق بين الاستجابة المشاهدة والاستجابة المقدرة اقل مايمكن اي تصغير مجموع مربعات الخطأ (SSE)

$$sse_i = \sum_{i=1}^n w_i (t_i - \hat{t}_i)^2 \quad (17)$$

$$t_i = \ln \frac{\pi_i}{1 - \pi_i} \quad (18)$$

$$sse_i = \sum_{i=1}^n w_i (Z_i - \hat{\beta}_1 y_{1i} - \dots - \hat{\beta}_p y_{pi})^2 \quad (19)$$

Z : يمثل متجه التحويل الخطي Log-Logistic للنموذج ذو رتبة (n*1)
W : يمثل مصفوفة مربعة عناصر قطرها الرئيسي يمثل التباينات ذو رتبة (m*n)
وبذلك نصل الى قيمة $\hat{\beta}$ المقدرة والتي تكون :

$$\hat{\beta}(XWX)^{-1} XWt \quad (20)$$

3.3. طريقة الخوارزمية الجينية

تعد الخوارزمية الجينية من طرائق البحث العشوائي تعالج مشكلة ما من اجل الوصول الى افضل النتائج الممكنة تتمحور حول تقنيات التطور من منطلق نظرية دار للتطور وتنص على البقاء للأفضل بتقليده عمل الطبيعة من خلال الاحتفاظ بالصفات الجيدة الموجودة في جيل الالباء ونقلها الى جيل الابناء هدفها الحصول على ذرية فورية تتمتع افضل صفات الالباء على اقل تقدير[6].

• مراحل الخوارزمية الجينية

تختلف الخوارزمية الجينية حسب اختلاف فروع التقنيات التطويرية الا انها مشتركة بهذا المراحل التالية [15]:

1. البداية: تتمثل في تعداد سكاني عشوائي من الكروموسومات (فضاء البحث) وبعبارة اخرى هي مجموعة حلول المسألة
2. التهيئة: هي عملية انشاء الجيل الابتدائي تتضمن توليد الكروموسومات عشوائية بقدر حجم المجتمع وحسب طبيعة المشكلة [9].
3. دالة الهدف ودالة التقييم (المفاضلة): من خلالها يتم تقييم الكروموسومات بحيث يعطي كل كروموسومات قيمة معينة تمثل مدى كفاءة الى مدى اقترابه من الحل وتكون دالة الهدف اما عظمى (Maximum) او صغرى (Minimum) وعليها تعتمد دالة التقييم .
4. الاستنتاج: عملية توليد جيل جديد من الافراد (الكروموسوم) التي يتم اختيارها من خلال عملية الاختيار وفق مبدأ البقاء للأصلح من ثم اجراء عليها عملية التهجين وعملية الطفرة لانتاج ابناء الجيل اللاحق .
5. مجتمع جديد: يتضمن توليد جيل جديد بتكرار المراحل التالية الى ان يكتمل الجيل ويتمثل الجيل بمايأتي[18] :
أ-الاختيار: هي عملية اختيار كروموسومات مناسبة من الجيل القديم لتكوين الجيل الجديد يتم انشاؤها وفقاً لقيم دالة التقييم لغرض وجود الكروموسومات مع اعلى قيمة لدالة التقييم في الاجيال الجديدة [17]
ب- التهجين: يتم اختيار الكروموسومات الجيدة من الجيل الاول يتم التزاوج بين كروموسومين لتكوين الجيل الجديد (الابناء) اعتماداً على الكروموسومات (الام) بأخذ الصفات الجيدة منها من طرائقها التهجين بنقطة وبنقطتين وغيرها [6].
- ت- الطفرة: بعد انشاء الجيل الجديد (الابناء) وباحتمال وجود الطفرة فتتم عملية الطفرة بأحداث بفترات عشوائية في صيغتها الكروموسومية وهذا يؤدي بنا الى الحفاظ على الصفات الجيدة بين الجينات في الكروموسوم الواحد والوصول الى شكل اسرع وفيها يحدث التبادل بين الكروموسوم وعندما لا تكون هنالك طفرة فتجري عملية استنساخ الكروموسومات (الاباء) مباشرة دون حدوث عملية التهجين [8].
6. الاختبار: نختبر الحل بتوفر شرط التوقف من عدمه فعند توفره تتوقف الخوارزمية الجينية ويفيد الحل الجديد من تكون اخر جيل [16].
7. معيار التوقف: يستمر تكوين الاجيال بشكل متعاقب من اجل تحسين امثلية الحل حتى يتحقق شرط التوقف الذي يعتمد على مقياس توقف الخوارزمية الجينية (الحل الامثل) ويختلف هذا المقياس حسب المشكلة المراد معالجتها .
8. الانتهاء : تنتهي الخوارزمية الجينية عند وجود احد العوامل التالية [12] :

✓ ايجاد الحل الامثل

✓ الوصول الى عدد الاجيال المطلوبة

✓ توفر قيمة معينة مثل حساب كلفة الانتاج

✓ الوقوع في قيمة الصغرى المحلية ولايمكن الخروج منها

✓ تطبيق مراحل الخوارزمية الجينية في أنموذج الانحدار log-logistic الثنائي :

نطبق مراحل الخوارزمية الجينية في معادلة دالة الهدف لكل طريقة لأيجاد تقديرات معلمات أنموذج الانحدار log-logistic الثنائي وفقاً لمايأتي [18]:

1. البداية: تكوين الكروموسوم من خلال قيم β_p التي تشكل جينات الكروموسوم وان ($P=0,1,2,...,p$) ضمن الاعداد الحقيقية .
2. التهيئة: انشاء الجيل الابتدائي عن طريق ايجاد قيمة اولية للجينات مع القيم العشوائية لمجموعة القيود الاخرى .
3. دالة الهدف: يتم تقييم الكروموسوم من حيث الكفاءة وصولاً الى الحل الامثل بتحديد قيمة β_p
4. اجراء عملية الاختبار للكروموسوم الذي يمتلك قيمة دالة الهدف صغيرة بأختيار الاحتمال الكبير لها وايجاد دالة التقييم له من خلال المعادلة التالية

$$Fitness Function = \frac{1}{1 + objective function} \quad (21)$$

من خلال صيغة دالة التقييم نستطيع ايجاد الاحتمالية هذه الدالة (افضل القيم) حسب الصيغة الرياضية الاتية [9] :

$$C_{(i)} = \frac{f(i)}{\sum_{i=1}^n f(i)} \quad (22)$$

اذ ان :

$C_{(i)}$: تمثل احتمالية الفرد i

$f(i)$: دالة التقييم الفرد i

n : يمثل حجم المشاهدات

1. في هذه الخطوة يتم تهجين الكروموسومات الجيدة في صفاتها عن طريق التزاوج بين كل كروموسومين ويطبق احد معاييرهم وهو التهجين المنظم بالاعتماد على احتمالية التهجين P_c وتقارن هذه القيمة مع قيمة الجينات للكروموسومين (الاباء) لتكوين الجيل الجديد الابناء ويحدث التبادل عندما تكون قيمة الجين الاكبر او تساوي القيمة الاحتمالية [13].
2. اخر خطوة يمكن ان تمر بها الكروموسومات هي عملية الطفرة وايضا تعتمد على مقدار احتمالي (P_m) للمعاملات استبدال جينات منقاة عشوائياً مع قيمة جديدة ايضاً حصلنا عليها بشكل عشوائي وفق الصيغة التالية :

مجموع الجينات = عدد الجينات في الكروموسوم * عدد السكان

4. الجانب العملي

4.1. التلوث البيئي

يعرف التلوث البيئي بأنه احدث تغير في النظام البيئي بشكل يؤدي الى عدم استطاعة النظام البيئي على اداء دوره الطبيعي في التخلص من الملوثات وخاصة العضوية منها بالعمليات الطبيعية الامر الذي يؤدي حدوث خلل في النظام البيئي [1]. وبمفهوم ادق يكون معنى التلوث كل ما يسهم في تغير اي عنصر من عناصر البيئة سواء كان هذا العنصر كائناً حياً كالانسان والحيوان والنبات او مكوناً طبيعياً غير حي كالهواء والماء والتربة والاشعاع وغيرها وفيما يأتي عرض موجز لأهم مصادر تلوث الهواء والاضرار الصحية والبيئة الناجمة عن هذه الملوثات [4].

4.2. مصادر ملوثات الهواء

بشكل عام تكون مصادر تلوث الهواء على نوعين مصادر صناعية وطبيعية , اما المصادر الصناعية فأنها تكون مصادر متحركة او ثابتة وتشمل كمحطات توليد الطاقة الكهربائية ومصانع الاسمنت والمعادن ومنشآت صناعة النفط والغاز الطبيعي التي تساهم هذه المصادر في زيادة خطورة التلوث بدرجات مختلفة فعلى سبيل المثال صناعة النفط تؤدي الى تلوث الهواء بأكاسيد الكبريت والنيتروجين والامونيا واوكسيد الكربون وكبريتيد الهيدروجين وغيرها من الغازات كما تنطلق بعض المواد المركبات العضوية الطيارة في الهواء والمحيط بمستويات البترول ومصانع الاصباغ... الخ [4]. لذلك تسعى في هذا البحث الى اثبات صحة هذه الفكرة الشائعة بفحص مستويات التلوث بالنسبة الى حجم انتاج النفط على وجه التحديد من خلال بيانات المقدمة من شركة مصافي الوسط.

4.3. مراحل اجراء الجانب العملي

✓ المرحلة الاولى : اسلوب جمع البيانات

تضمن البحث الحصول على بيانات مناخية من موقع (weather history for KQTZ) على شبكة الانترنت بأعتبره من افضل المواقع التي يوفر سجلات الطقس التاريخية بسهولة لجميع البلدان حسب الرمز البريدي الخاص بكل منطقة (مدينة). اما بيانات ملوثات الهواء فقد تم الحصول عليها من شركة مصافي الوسط التي تمثل القياسات اليومية لمركبات التلوث البيئي التي تستند للمدة ايلول 2018 الى كانون الاول 2022 وقد تم استحصاها موافقة شركة مصافي الوسط (مصفى الدورة) واذا هذه مجموعة البيانات تتعقب العناصر المسببة للتلوث الجوي من خلال القياسات المتكررة للمركبات مثل مركبات الكربون (CO_x) ومركبات الكبريت (SO_x) ومركبات النيتروجين (NO_x) فضلاً عن المتغيرات التفسيرية بما في ذلك سرعة الرياح , درجات الحرارة , كميات النفط المنتجة مع الاشارة الى انه تم اخذ قياسات البيانات للعوامل المناخية المسجلة في نفس الساعة التي تم فيها قياس كل ملوثات في المصفي .

✓ المرحلة الثانية : وصف البيانات

بالنظر لخلفية البيانات فقد تألفت مجموعة البيانات على ذلك (9) سمة مميزة اذ كانت اثنتين فئوية (التاريخ , الوقت) وكانت متغيرين الاستجابة هما:

- غاز اول اوكسيد الكربون (CO): يعد هذا الغاز انه عديم اللون والرائحة وينشأ نتيجة لأنبعاثات من عملية تكرير وانتاج النفط وهذا الغاز له القدرة على التداخل في الدم واعاقه قدرته على امتصاص وحمل غاز الاوكسجين ويؤدي احيانا الى الوفاة.
 - غاز ثاني اوكسيد الكبريت (SO₂): عبارة عن غاز لونه بني مصفر ينتج من مصافي النفط يسبب التعرض لهذا الغاز الى امراض خطيرة في القلب وامراض الجهاز التنفسي وخصوصاً الرئتين كما انه يعتبر سام للنباتات ويعمل هذا الغاز على تآكل المعادن عند زيادة نسبته في الهواء .
- وخمس متغيرات تفسيرية هما متوسط درجة الحرارة لكل ساعة بالـ (C⁰) ومتوسط نقطة الندى ومتوسط الرطوبة لكل ساعة بالدرجة المئوية ومتوسط سرعة الرياح بالـ (كم / ساعة) ومتوسط كمية النفط الخام المستعمل في عملية التنصيف (م³ / ساعة).

✓ المرحلة الثالثة : ترميز البيانات

بما ان تم قياس مستوى بيانات التلوث لمدة اكثر من سنتين (28) فإن عدد القياسات يكون صغير نسبياً لتحديد مسارات معقدة في متغيرات الاستجابة وبناءً على ذلك تم افتراض أنموذج logit للعينة الذي بموجبه تم تحويل البيانات الى استجابات ثنائية للتنبؤ بنتائج التلوث، وتم معالجة البيانات استناداً الى الحدود القصوى المسموح بها لمكونات الهواء المنبعثة من مصادر الاحتراق. اذ ان تعد القيم الدنيا والمساوية الى الحدود القصوى غير ملوثة [0] وتعد القيم الاعلى من الحدود القصوى ملوثة [1] . وان الحدود القصوى المسموح بها لتركيز كل ملوث من ملوثات التي يسمح بطرحها بموجب المعايير الوطنية ان غاز CO قيمته 500 ug/m³ و غاز SO₂ قيمته 10 ug/m³

✓ المرحلة الرابعة : اختبار فرضية للأنموذج العام

غالباً ما يتم اختبار معنوية المعلمات المناظرة لقبول او رفض فرضية العدم اقل 0.05 والتي تنص على ان
 H_0 : الزيادة في معدل التلوث لاتعتمد على زيادة في الظروف المناخية وانتاج النفط
 H_1 : الزيادة في معدل التلوث تعتمد على زيادة في الظروف المناخية وانتاج النفط

✓ المرحلة الخامسة : تحليل البيانات

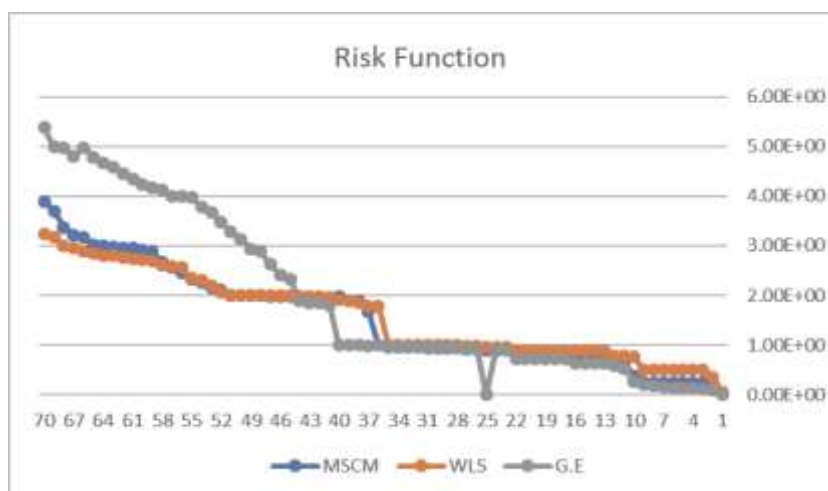
سوف يتم عرض نتائج الجانب العملي وتحليلها للوصول الى مدى ملائمة البيانات الحقيقية مع أنموذج الانحدار log-logistic المقدر من خلال اجراء الاختبارات الخاصة بمقدرات دالة المخاطرة لأنموذج log-logistic وفيما يلي سوف يتم عرض النتائج في الجداول التي سيتم تحليلها وحسب الجداول وكالاتي :

جدول (1): يمثل مقدرات دالة المخاطرة لأنموذج الانحدار log-logistic الثنائي بطريقة تصغير مربعات كاي وطريقة المربعات الصغرى الموزونة و طريقة الخوارزمية الجينية

parameter Estimated	$\hat{h}(MCSE)$	$\hat{h}(WLS)$	$\hat{h}(G.E)$
$\hat{\beta}_0$	0.540	0.058	4.422
$\hat{\beta}_1$	0.089	0.0741	0.005
$\hat{\beta}_2$	1.547	0.1937	0.003
$\hat{\beta}_3$	0.654	0.092	0.137
$\hat{\beta}_4$	0.255	2.032	0.383
$\hat{\beta}_5$	1.453	0.425	0.020

نلاحظ في جدول رقم (1) تنفيذ مقدرات دالة المخاطرة لأنموذج الانحدار log-logistic من خلال تحديد متجه

المعلمات $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5)$ وتعويضها مع المتغيرات التوضيحية نحصل على مقدرات أنموذج الانحدار log-logistic الثنائي



شكل (1): يمثل مقدرات دالة المخاطرة لأنموذج الانحدار log-logistic الثنائي بطريقة تصغير مربع كاي ومربعات الصغرى الموزونة والخوارزمية الجينية

نلاحظ من شكل رقم (1) ان سلوك مقدرات دالة المخاطرة لطرائق التقدير ظهرت بشكل متزايد لأنموذج الانحدار log-logistic الثنائي .

جدول (2): يمثل مقدرات دالة المخاطرة والخطأ المعياري لكافة المتغيرات التوضيحية لأنموذج الانحدار log-logistic الثنائي بطريقة تصغير مربعات كاي

Estimated parameter	$\hat{h}(MCSE)$	Standard error $SE(\hat{\beta}_i)$	$\frac{\hat{\beta}}{SE(\hat{\beta}_i)}$	significations
$\hat{\beta}_0$	0.540	0.031	17.451	Sig
$\hat{\beta}_1$	0.089	0.039	2.283	Sig
$\hat{\beta}_2$	1.547	0.327	4.731	Sig
$\hat{\beta}_3$	0.654	1.491	0.441	N0N-Sig
$\hat{\beta}_4$	0.255	0.012	21.273	Sig
$\hat{\beta}_5$	1.453	0.437	3.327	Sig

نلاحظ في جدول رقم (2) تقديرات دالة المخاطرة لأنموذج وقيم الخطأ المعياري لكل معلمة مقدرة والنتائج التي تم الحصول عليها وهي تتبع التوزيع الطبيعي القياسي بمستوى دلالة ($\alpha=0.05$) وتقران مع قيمة Z الجدولية $Z_{\frac{1}{2}(1-0.05)} = \pm 1.96$ والعمود الاخير يمثل معنوية المتغيرات التوضيحية (اقل من 0.05) قيد الاختبار يكون معنوياً .

وعليه تم ازالة ($\hat{\beta}_3$) لانها لا تؤثر بشكل كبير في الانموذج المقدر اي ان المتغير متوسط الرطوبة ليس لها اي تأثير، بينما ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_4, \hat{\beta}_5$) متوسط درجة الحرارة ومتوسط نقطة الندى ومتوسط سرعة الرياح ومتوسط كمية النفط المنتج له تأثير معنوي على غاز (CO) و (SO_2) في انتشاره بكميات كبيرة في الهواء. وفيما اذا ذلك ان قوة العلاقة بين المتغيرات التوضيحية ومتغير الاستجابة من خلال احتساب احصاءة R^2 ظهرت قيمتها تساوي 0.52 وهذا يعني ان متوسط درجة الحرارة ومتوسط نقطة الندى ومتوسط سرعة الرياح ومتوسط كمية النفط المنتج تفسر الاختلافات في الاستجابات المشاهدة لحالات (CO) و (SO_2) بنسبة 52 % .

جدول (3): يمثل مقدرات دالة المخاطرة والخطأ المعياري لكافة المتغيرات التوضيحية لأنموذج الانحدار log-logistic الثنائي بطريقة المربعات الصغرى الموزونة

Estimated parameter	$\hat{h}(WLS)$	Standard error $SE(\hat{\beta}_i)$	$\frac{\hat{\beta}}{SE(\hat{\beta}_i)}$	significations
$\hat{\beta}_0$	0.058	0.043	1.361	N0N-Sig
$\hat{\beta}_1$	0.0741	0.1674	0.443	N0N-Sig
$\hat{\beta}_2$	0.1937	0.1521	1.274	N0N-Sig
$\hat{\beta}_3$	0.092	0.093	0.998	N0N-Sig
$\hat{\beta}_4$	2.032	0.741	2.743	Sig
$\hat{\beta}_5$	0.425	0.214	1.988	Sig

نلاحظ في جدول رقم (3) تقديرات دالة المخاطرة لأنموذج وقيم الخطأ المعياري لكل معلمة مقدرة والنتائج التي تم الحصول عليها وهي تتبع التوزيع الطبيعي القياسي بمستوى دلالة ($\alpha=0.05$) وتقران مع قيمة Z الجدولية $Z_{\frac{1}{2}(1-0.05)} = \pm 1.96$ والعمود الاخير يمثل معنوية المتغيرات التوضيحية (اقل من 0.05) قيد الاختبار يكون معنوياً .

وعليه تم ازالة ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$) لانها لا تؤثر بشكل كبير في الانموذج المقدر اي ان المتغيرات (متوسط درجة الحرارة ومتوسط نقطة الندى ومتوسط الرطوبة) ليس لها اي تأثير، بينما ($\hat{\beta}_4, \hat{\beta}_5$) (متوسط سرعة الرياح ومتوسط كمية النفط المنتج) له تأثير معنوي على غاز (CO) و (SO_2) في انتشاره بكميات كبيرة في الهواء. وفيما اذا ذلك ان قوة العلاقة بين المتغيرات التوضيحية ومتغير الاستجابة من خلال احتساب احصاءة R^2 ظهرت قيمتها تساوي 0.47 وهذا يعني ان متوسط سرعة الرياح ومتوسط كمية النفط المنتج تفسر الاختلافات في الاستجابات المشاهدة لحالات (CO) و (SO_2) بنسبة 47 % .

جدول (4): يمثل مقدرات دالة المخاطرة والخطأ المعياري لكافة المتغيرات التوضيحية لأنموذج الانحدار -logistic الثنائي بطريقة الخوارزمية الجينية

Estimated parameter	$\hat{h}(G. E)$	Standard error $SE(\hat{\beta}_i)$	$\frac{\hat{\beta}}{SE(\hat{\beta}_i)}$	significations
$\hat{\beta}_0$	4.422	3.471	1.274	N0N-Sig
$\hat{\beta}_1$	0.005	0.013	0.391	N0N-Sig
$\hat{\beta}_2$	0.003	0.002	1.510	N0N-Sig
$\hat{\beta}_3$	0.137	0.146	0.943	N0N-Sig
$\hat{\beta}_4$	0.383	0.192	1.999	Sig
$\hat{\beta}_5$	0.020	0.006	3.451	Sig

نلاحظ في جدول رقم (4) تقديرات دالة المخاطرة وقيم الخطأ المعياري لكل معلمة مقدرة والنتائج التي تم الحصول عليها وهي تتبع التوزيع الطبيعي القياس بمستوى دلالة ($\alpha=0.05$) وتقارن مع قيمة Z الجدولية $Z_{\frac{1}{2}(1-0.05)} = \mp 1.96$ والعمود الأخير يمثل معنوية المتغيرات التوضيحية (أقل من 0.05) قيد الاختبار يكون معنوياً .

وعليه تم ازالة ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$) لأنها لا تؤثر بشكل كبير في الانموذج المقدّر أي ان المتغيرات (متوسط درجة الحرارة ومتوسط نقطة الندى ومتوسط الرطوبة) ليس لها أي تأثير، بينما ($\hat{\beta}_4, \hat{\beta}_5$) (متوسط سرعة الرياح ومتوسط كمية النفط المنتج) له تأثير معنوي غاز (CO) و (SO_2) على انتشاره بكميات كبيرة في الهواء. وفيما إذا ذلك ان قوة العلاقة بين المتغيرات التوضيحية ومتغير الاستجابة من خلال احتساب احصاءة R^2 ظهرت قيمتها تساوي 0.62 وهذا يعني ان متوسط سرعة الرياح ومتوسط كمية النفط المنتج تفسر الاختلافات في الاستجابات المشاهدة لحالات (CO) و (SO_2) بنسبة 62 % .

5. الاستنتاجات

- بالنسبة لأنموذج الانحدار log-logistic الثنائي بطريقة تصغير مربع كاي تبين من خلال اختبار قيم (Z) ان العوامل التي كان لها تأثير معنوي وأقل من 0.05 على متغير الاستجابة لغاز (CO) و (SO_2) هي متوسط درجات الحرارة ومتوسط نقطة الندى ومتوسط سرعة الرياح وكمية النفط المنتج .
- كان عامل متوسط سرعة الرياح ومتوسط كمية النفط المنتج هو الأكثر تأثيراً وبمعنوية أقل من 0.05 على متغير الاستجابة لغاز (CO) و (SO_2) باستعمال اختبار (Z) لأنموذج الانحدار log-logistic الثنائي بطريقة مربعات الصغرى الموزونة وطريقة الخوارزمية الجينية .
- اثبتت جميع طرائق التقدير ان المتغيرين متوسط سرعة الرياح ومتوسط كمية النفط المنتج يعتبران من العناصر الرئيسية على انتشار غاز (CO) و (SO_2) وهذا ما تم تأكيده من قبل المهندسين الكيميائيين بهذا المجال .
- اظهرت مقدرات دالة المخاطرة بشكل متزايدة بأزيد انتشار نسبة غاز (CO) و (SO_2) في الهواء

6. التوصيات والدراسات المستقبلية

- استعمال الطرائق التي تم عرضها لتقدير دالة المخاطرة لأنموذج الانحدار log-logistic الثنائي والذي يمكن استعماله لنمذجة البيانات في مجالات التلوث البيئي مثل تلوث الماء والتربة والتلوث الإشعاعي .
- اجراء دراسات مستقبلية تتناول الاعتماد على طرائق التقدير الاعتيادية والمحسنة بالخوارزمية الجينية المذكورة في البحث والمقارنة بينهما في حالة وجود قيم شاذة في البيانات .
- استعمال متغيرات اخرى غير التي ذكرت في البحث مثل متغيرات الاستجابة التي تشمل غاز اكاسيد نيتروجين والامونيا واوكسيد الكربون ثنائي والاوزون .
- اضافة بيانات من اجل تسهيل تطوير أنموذج تنبؤ مثالي واقتراح ان البيانات اليومية للمدة من 10 الى 20 سنة تسفر عن نموذج أكثر دقة .
- ينبغي ان تنظر شركة مصافي الوسط ايضاً في تحسين عمليات جمع البيانات بمواصلة جمع البيانات عن الملوثات المحتملة في مختلف الظروف المناخية مثل النفايات السائلة والصلبة التي يتم الحصول عليها من خلال عمليات التكرير وذلك لتسهيل تقييم المخاطر على مسار الهواء في الدراسات المستقبلية المماثلة .
- واحدة من الدراسات الرئيسية هي لفحص المستويات المستمرة لانبعاثات الكربون الناجمة عن عمليات التكرير في مصافي النفط بالنسبة لكميات النفط المنتجة .
- ان تناول هذا الموضوع من شأنه تسهيل التعرف على إمكانية تلوث الكربون عن طريق مصافي النفط باعتبار ان المتغيرات البيئية ايضاً تديم التلوث وتحفز التفاعلات بين الملوثات

المصادر

- [1] A. Agresti, "Categorical Data Analysis", 2nd edition, Jhon Wiley & Sons Inc , Hoboken, New Jersey, 2002.
- [2] A. Alrahamneh, and O. Hawamdeh, "The Factors Affecting Eye Patients (Cataract) In Jordan by Using the Logistic Regression Model", Modern Applied Science, ISSN, pp. 38-42 , 2017.
- [3] A. Misra , "Portfolio optimization of commercial Banks- An Application of Genetic Algorithm", EJBM, pp. 120-129, 2013.
- [4] A.A Hassan, and Sh." Comparison of some robust estimators to estimate the parameters of the two-response logistic regression model using simulation", Journal of Economics and Administrative Sciences, Pages 423- 440 , 2018 .
- [5] B. Altshuler, "Theory for the measurement of competing risks in animal experiments", Math. Biosci., 6, 1-11 , 1970.
- [6] D. Garson, "Logistic Regression :Binary and Multinomial" Statistical Associates Publishing. All rights reserved worldwide in all media ,ISBN , Retrieved from , 2014 , [http : www.statisticalassociates.com](http://www.statisticalassociates.com).
- [7] D. R. Cox, "A note on the graphical analysis of survival data". Biometrika, 66, 188-190 , 1979.
- [8] D. Raghupathikumar ,and K.Raja, "Genetic Algorithm Based Scheduling of an Input Queued Switch", IJCA, pp.37-42 , 2012.
- [9] E. Demir , Ö. Akkus , "An Introductory Study on How the Genetic Algorithm Works in the Parameter Estimation of Binary Logit Model", IJS:BAR, pp.162-180 , 2015.
- [10] H. Liu, C.Ong , "Variable selection in clustering for marketing segmentation using genetic algorithms", ELSEVIER, pp. 502-510 ,2008.
- [11] I. Yousef , "Some methods for estimating the conditional logistic regression model in the case of longitudinal data and their application in environmental pollution " A master's thesis in statistics submitted to the College of Administration and Economics at the University of Baghdad , 2017.
- [12] J. H. Roger, and S. D. Peacock, "Fitting the scale as a GLIM parameter for Weibull, extreme value, logistic and log-logistic regression models with censored data". GLIMN ewslatter , 6, 30-37, (1982).
- [13] J. Pasia , A. Hermosilla, and et al. , "A useful tool for statistical estimation genetic algorithm", JSCS, pp. 237 – 251 , 2005.
- [14] J. S. Cramer, "Logit Models From Economics and other fields", Cambridge University Press cape Town, New York, ISBN, pp.33-45, 2003.
- [15] J. Whitehead, "Fitting logistic and log-logistic regression models to censored data using GLM " . GLIM Newsletter, 4, 12-19. Correction, 5, 3, 1981.
- [16] K.H. Lee, K.W. Kim, "Performance comparisons of particle swarm optimization and genetic algorithm for inverse surface radiation problem", ELSEVIER, PP. 330-337 ,2015.
- [17] M. Aitkin , D. Clayton , "The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using GLIM", Appl. Statist., 29, 156-163, 1980.
- [18] M. Marlene , "Generalized Linear Models", Fraunhofer Institute for Industrial Mathematics (ITWM),(Germany) , [www. Marlenmuller . ed / publication / hand book CS. Pdf](http://www.Marlenmuller.de/publication/handbookCS.Pdf) , 2004 .
- [19] M.S. Ahmed , "Forecasting of Environmental Pollution Using Box-Genghis Models for Al-Waziriya Station", Diploma Thesis in Statistics, College of Administration and Economics at the University of Baghdad , 2015.
- [20] N. Pettitt, "Fitting generalized logistic and log-logistic models to censored data using GLIM" . GLIM Newsletter, 7, in press , 1983.
- [21] Ö. Akkus , E. Demir , " Comparison Som Classical And Meta-Heuristic Optimazation Techniques in The Estimation Of The Logit Model Parameters", IJAR, pp.1026-1042 , 2016.
- [22] P. McCullagh, "Regression models for ordinal data". J. R. Statist. Soc. B, 42, 109-142, 1980.
- [23] P. Mccullagh, and J. Nelder, "Generalized Linear Models", London, Chapman and Hall ,1983.

- [24] S. Bennett , "Log-logistic Regression Models for Survival Data" , University of Reading, UK , Appl. Statist, 32,No.2,pp. 165-171, 1983.
- [25] S. Bennett, "Analysis of survival data by the proportional odds model". Statistics in Medicine, 2, in press,1983.
- [26] S. Hadji, and et al. , "Theoretical and experimental analysis of genetic algorithms based mppt for PV systems", ELSEVIER, pp. 772-787 , 2015.
- [27] S. Menard, "Applied Logistic Regression Analysis", 2nd Edition Thousand Oaks Edition Thousand Oaks , CA : Sage Publications , Series Quantitative Applications in the Social Sciences , PP. 1 – 111, 2002.
- [28] S. Zager, R. Karim , "Generalized Linear Models with Random Effects : A Gibbs Sampling Approach", JASA , pp. 79-86 , 1991 .
- [29] S.A Al-Rudaini, "The use of the genetic algorithm in estimating the parameters of the binary logistic regression model with practical application" A master's thesis in statistics submitted to the College of Administration and Economics at the University of Baghdad ,2019
- [30] S.A Rabab, M.A Sara , " The use of the Biz method and classical methods in estimating the parameters of the binary logistic regression model" , Journal of Economics and Administrative Sciences , Vol.. 25 , No..113 , Pages 543- 556 , 2019 .
- [31] S.Sefian , M. Benbouziane , "Portfolio Selection Using Genetic Algorithm" , MPRA , pp. 1-12 , 2012 .



AL- Rafidain
University College

PISSN: (1681-6870); EISSN: (2790-2293)

**Journal of AL-Rafidain
University College for Sciences**

Available online at: <https://www.jrucs.iq>

JRUCS

Journal of AL-Rafidain
University College
for Sciences

Determine the Variables Affecting Environmental Pollution with CO and SO₂ Gases Using the Log-Logistic Regression Model

Prof. Dr. Entsar A. Fadam AL.Doori	Samah S. Hassan
entsar_arebe@coadec.uobaghdad.edu.iq	Samahsabah1993@gmail.com
Department of Statistics, College of Administration and Economics, University of Baghdad, Baghdad, Iraq	Ministry of Interior, Baghdad, Iraq

Article Information

Article History:

Received: February, 25, 2024

Accepted: April, 12, 2024

Available Online: December,
31, 2024

Keywords:

Log-logistic distribution, chi-square minimization method, weighted least squares method, genetic algorithm method, risk function, GLM

Abstract

The problem of environmental pollution is one of the most important and dangerous problems facing humans at present because of its direct impact on human health and other living organisms. In recent years, it has been observed that there has been an increase in environmental pollution rates, which has greatly affected human health and led to the emergence of many diseases such as cancer, pneumonia, poisoning, birth defects, and others. Its impact has not only been limited to humans but has extended beyond its impact on animals and plants as well.

Given the importance of the topic, its seriousness, and its direct impact on human life, this research was conducted to determine the proportion of the amount of pollution based on two of the most important factors on air pollution, which are both gases (CO and SO₂), based on the explanatory variables, which are the average temperature, the average dew point, and the average humidity: the average wind speed and the average amount of crude oil used in the filtering process.

In this research, the risk function was estimated using the Log-Logistic regression model using estimation methods: the Chi-square minimization method, the weighted least squares method, and the genetic algorithm method. These methods were applied to air pollution data from the Central Refineries Company in Baghdad (Al-Dora Refinery). which represents daily measurements of environmental pollution compounds, which are based on time for the period from 2018 to 2022.

Correspondence:

Samah S. Hassan

Samahsabah1993@gmail.com

DOI: <https://doi.org/10.55562/jrucs.v56i1.22>