

استعمال طريقة الإمكان الأعظم مع وزن مقترح لتقدير تأثير بعض الملوثات في نهر دجلة - مدينة الكوت

خالد جمال كاظم**

أ. م. د. عمر عبد المحسن علي*

المستخلص

يهدف هذا البحث الى استعمال أسلوب الحصانة بالتشذيب حيث إن في كثير من الأحيان يفشل تحليل الإمكان الأعظم (ML) تماماً في حالة وجود قيم شاذة في الظاهرة المدروسة. حيث ستفقد MLE خصائصها لما تسببه الشواذ Outliers من تأثير سيئ على الظاهرة المدروسة ولعلاج هذه المشكلة تم تطوير أساليب إحصائية جديدة بحيث لا تتأثر بالقيم الشاذة بسهولة هذه الأساليب تمتاز بالحصانة أو (المقاومة)، ولذا كانت طريقة الإمكان الأعظم المشذب (MTL) كبديل جيد يحقق نتائج أكثر مقبولة وأمثلة، إلا أنه يمكن استخدام أوزان تؤدي الى زيادة كفاءة المقدرات الناتجة ولزيادة قوة التقدير باستخدام طريقة الإمكان الأعظم المشذب الموزونة (MWTL). ولتحقيق هذا البحث استدعت الحاجة الكشف والتقصي عن تأثير التلوث مياه نهر دجلة في مدينة الكوت بسبب مياه الصرف الصحي وبالذات التلوث بالمواد الصلبة غير الذائبة في الماء (TDS) كمتغير استجابة وتأثير ثلاث ملوثات هي (PO₄, CL, SO₄) كمتغيرات توضيحية. وتم تقديم ذلك ببحث احصائي وتقييمها بشكل دقيق ورفعها الى الجهات المختصة، حيث تم استعمال عينة بحجم (91) تم سحبها وفحصها في مختبر مديرية مجاري مدينة الكوت وتم إجراء التحليل باستعمال برنامج (MATLAB-2015). وكان معيار المفاضلة هو معيار الجذر التربيعي لمتوسط مربعات الخطأ (RMSE).

الكلمات المفتاحية: الإمكان المشذب، القيم الشاذة، الطرائق المشذبة، تلوث مياه الأنهار.

Using the greatest possible method with a suggested weight to estimate the effect of some pollutants on the Tigris River- City of Kut

Abstract

The aim of this research is to use the method of pruning immunity, as often the analysis of the greatest potential (ML) completely fails if there are abnormal values in the phenomenon studied. MLE will lose its properties because of the bad effect of the outliers on the studied phenomenon. Acceptability and optimization, however, weights can be used that increase the efficiency of the resulting capabilities and increase the power of estimation using the MWTL. In order to achieve this research, it was necessary to detect and investigate the impact of pollution of the Tigris River water in Kut city due to wastewater, especially pollution with water insoluble solids (TDS) as a response variable and the effect of three pollutants (PO₄, CL, SO₄) as illustrative variables. This was presented by statistical research and evaluated accurately and submitted to the competent authorities, where a sample size (91) was withdrawn and examined in the laboratory of the Directorate of sewage in the city of Kut and the analysis was conducted using the program MATLAB-2015)). The differentiation criterion was the RMSE.

Keywords: Trimmed Potential, Abnormal Values, Trimmed Methods, River Water Pollution.

1- مقدمة

تُعد مشكلة وجود قيم شاذة (Outliers) في البيانات من المواضيع التي حازت على اهتمام كبير في السنوات الأخيرة ، وذلك لإدراك الكثير من الباحثين أن استخدام الأساليب التقليدية في تقدير معلمات نموذج الانحدار في ظل وجود هذه المشكلة ستخرب بصورة كبيرة الافتراضات الموضوعية كشرط التحليل. لذا كان من الضروري البحث عن طرائق بديلة تكون حصينة لظهور الشواذ في البيانات. أذ فضل بعض الباحثين عدم تغيير قيم الشواذ بل يجب إيجاد طرائق بديلة للتقدير تكون غير حساسة لها عند وجود الشواذ في العينة لأنه كما هو معلوم عملية تقدير المعلمات للنموذج المشخص تعتبر مؤشرا للحصول على تمثيل جيد للمجتمع الذي أخذت منه العينة قيد الدراسة. على الرغم من وجود طرائق حصينة مختلفة للتقدير إلا أن أغلبها تشترك في نقطتين أساسيتين هما: إعطاء وزن أقل للقيمة الشاذة التي تتحرف لتقليل من تأثيرها والأخرى هي استخدام أسلوب التكرار. وغالبا ما تُظهر القيم الشاذة البيانات الحقيقية بصورة كبيرة جدا وأنها غالبا ما تظهر دون أن يلاحظها أحد وذلك بسبب أن البيانات تتم معالجتها بواسطة أجهزة الكمبيوتر من دون الكشف الدقيق. والقيم الشاذة لا تقتصر على قيم متغير الاستجابة y بل يمكن أن يكون في الجزء التفسيري x أيضاً. وكلا النوعين من الشواذ في كثير من الأحيان يفشلان في تحليل المربعات الصغرى (LS) تماماً وكذلك الامكان الأعظم (ML). ولعلاج هذه المشكلة تم تطوير أساليب إحصائية جديدة بحيث لا تتأثر بالقيم الشاذة بسهولة. هذه الطرائق تمتاز بالحصانة أو (المقاومة). حيث أن هذه القيم الشاذة أما أن تأتي من بيانات لتوزيعات غير متماثلة أي يكون فيها التواء عالٍ نحو اليمين أو نحو اليسار أو من نوعين من التوزيعات أحدهما التوزيع الأساسي والذي يولد مشاهدات جيدة بينما الآخر يسمى التوزيع الملوث والذي يولد قيما شاذة أو نتيجة لأسباب أخرى منها أخطاء يقع فيها الباحث عند تسجيل القياسات أو نتيجة وجود خلل في جهاز القياسات وخاصة في التجارب المختبرية، أو نتيجة أخطاء في الحسابات مما يؤدي إلى ظهور القيم الشاذة.

2- مشكلة البحث

تعاني البيانات من مشكلة وجود القيم الشاذة والتي لا يمكن أدراكها في ظاهرة من الظواهر والتي قد تؤدي الى ظهور مشاكل أخرى سيعاني منها نموذج الانحدار الذي سيمثلها رياضياً. وأن هذه القيم إذا تم حذفها مباشرة سوف يؤثر ذلك على النتائج المتوخاة وبالتالي على عملية التقدير.

3- هدف البحث

إن هذا البحث يهدف الى تسليط الضوء على تلوث مياه نهر دجلة في محافظة واسط بمياه الصرف الصحي وذلك لأهمية المياه في حياة جميع الكائنات الحية، وإن مياه الصرف الصحي تُعد من أخطر مصادر تلوث المياه لما تحتويه من مواد ضارة يصعب التخلص منها.

وقد تم إستعمال أسلوب التشذيب (trimming) ضد وجود القيم الشاذة، ودمجها بطرائق تقدير نموذج الانحدار المعروفة، مثل: الامكان الاعظم (MLE) لتصبح (MTLE)، حيث إن التشذيب في اللغة هو إبعاد الأجزاء التي لا يُرتجى منها فائدة وكذلك يهدف الى إستعمال أوزان حصينة يتم إدخالها على الطريقة المشدبة أعلاه لتصبح WMTL وكذلك التعرف في الوقت نفسه على القيم الشاذة وكشفها، ومقارنة افضلية هذا الاسلوب بإحدى معايير المقارنة مثل: (RMSE).

4- القيم الشاذة Outliers

هي المشاهدات في العينة التي تكون قيمها صغيرة جداً أو كبيرة جداً لمقارنة بقيّة المشاهدات في العينة حيث أن وجود هذه القيم يؤثر سلباً على كفاءة الطرائق الإحصائية ويجب الكشف عن هذه القيم للحدّ من تأثيرها [1]pp.252، أو هي المشاهدات التي تبدو غير منطقية عندما تتم مقارنتها بقيّة البيانات أو هي المشاهدات التي تقع بعيدة عن معادلة الانحدار ويكون لها خطأ كبير مقارنة بقيّة المشاهدات الطبيعية الأخرى. وقد وضح (Rousseeuw 1987) بياناً عن كيفية تأثير القيمة الشاذة في مقدرات طريقة المربعات الصغرى وكيف أن هذه القيمة تغير من اتجاه خط الانحدار وقد عرف العديد من الباحثين القيم الشاذة، إذ عرّف [2]pp.1 Hawkins المشاهدات الشاذة كالآتي: "هي المشاهدات التي تتحرف كثيراً عن المشاهدات الأخرى مما يثير الشك بأنه تم توليدها بألية مختلفة". أما [3]pp.2 Freeman فقد عرّف المشاهدات الشاذة كالآتي: "أي مشاهدة التي لم يتم توليدها بواسطة الآلية التي ولدت أغلبية المشاهدات في مجموعة البيانات". وبذا يمكن القول بأن القيمة الشاذة هي المشاهدات التي تقع بعيداً عن معظم مشاهدات العينة والتي لا تتسجم أو لا تتطابق مع بقيّة المشاهدات الموجودة في العينة تحت الدراسة. وأن هذه القيمة الشاذة يكون لها تأثير سلبياً على المعالم المقدرة وقد وضح [4]pp.1 Huber ذلك بقوله: "إن وجود مشاهد واحدة شاذة قد يهدم المزايا الجيدة لمقدرات المربعات الصغرى" أن تعاريف القيم الشاذة تقع في خمسة فئات [5] pp.1 وإن هذه الفئات تستند إلى:

1. التوزيع: أن المشاهدة الشاذة تعتبر شاذة إذا انحرفت عن التوزيع الأساسي وعلى سبيل المثال في حالة التوزيع الطبيعي فإن المشاهدة الشاذة تعتبر شاذة إذا كانت تبعد عن المتوسط بمقدار ثلاثة أمثال التباين.
2. المسافة: تعرف القيم الشاذة هي القيمة التي تقع على مسافة بعيدة من النسبة المئوية للبيانات.
3. التجمع أو التكتل: تعرف القيم الشاذة هي القيمة التي لا تلائم نمط التكتل.
4. الكثافة: يتم الكشف عن المشاهدة الشاذة من خلال المشاهدات.
5. العمق: المشاهدة الشاذة هي المشاهدة التي تقع في الطبقات الخارجية.

5- الكشف عن القيم الشاذة *Outlier Detection* pp.217 [9]:

يتم الكشف عن المشاهدات الشاذة في المتغيرات المستقلة في نموذج الانحدار الخطي المتعدد بواسطة المصفوفة H (*the hat matrix*) والتي تعرف بالصيغة الآتية

$$H = x(x'x)^{-1}x' \quad (1)...$$

فإذا كان $h_{ii} > \frac{2m}{n}$ or $h_{ii} > \frac{3m}{n}$ دل ذلك على أن المشاهدة (i) هي مشاهدة شاذة حيث أن (m) هي عدد المتغيرات المستقلة في النموذج و(n) هو حجم العينة، حيث أن h_{ii} هي عناصر القطر الرئيسي للمصفوفة (H).

6- طرائق التقدير

أ- طريقة الإمكان الأعظم *Maximum Likelihood Method*:

قدم العالم Fisher, R.A عام 1920 طريقة للتقدير أطلق عليها أسم (طريقة الإمكان الأعظم) حيث أن أساس هذه الطريقة يعتمد على إيجاد تقديرات للمعالم أو إيجاد مقدر مثل $\varphi(\hat{\theta})$ للدالة $\varphi(\theta)$ التي تجعل دالة الإمكان في نهايتها العظمى، وتعد طريقة الإمكان الأعظم من أهم طرائق التقدير في الإحصاء.

من المعروف أن دالة الإمكان (Likelihood Function) لـ (n) من قياسات العينة العشوائية $X = \{x_1, x_2, \dots, x_n\}$ مسحوبة من مجتمع ذو توزيع احتمالي $f(X; \theta)$ يعرف كالآتي

$$L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i, \theta) \quad (2)...$$

ويمكن إيجاد مقدر الإمكان الأعظم بحل المعادلة التالية بالنسبة لـ (θ)

$$\frac{\partial L(x_1, x_2, \dots, x_n; \theta)}{\partial \theta} = 0 \quad \dots (3)$$

بشرط أن

$$\frac{\partial^2 L(x_1, x_2, \dots, x_n; \theta)}{\partial \theta^2} < 0 \quad \dots (4)$$

فإذا كان لدينا نموذج انحدار متعدد

$$Y_i = B_0 + B_1x_{i1} + B_2x_{i2} + \dots + B_kx_{ik} + U_i \quad i = 1, \dots, n \quad (5)...$$

وبشكل أكثر إختصار يمكن التعبير عن مجموعة المعادلات أعلاه بصيغة المصفوفات أي أن النموذج يصبح بالشكل التالي

$$Y = X\beta + U \quad (6)...$$

حيث أن

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \underline{U} = \begin{bmatrix} u_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \dots (7) \underline{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

وأردنا تقدير معالم النموذج فيتم كالآتي

بما أن $Y_i \sim N(X_i \underline{\beta}, \sigma^2)$ فإن دالة الإمكان للمتغير المعتمد (Y_i) تكون كالآتي

$$L(Y) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{U'U}{2\sigma^2}} \quad \dots (8)$$

حيث أن $U = Y - X\underline{\beta}$

حيث إن $Y = (y_1, y_2, \dots, y_n)$ ولتسهيل العمليات الرياضية يتم أخذ اللوغاريتم الطبيعي تصبح الدالة كالآتي

$$\ln L(Y) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{U'U}{2\sigma^2} \quad \dots (9)$$

إن الدالة الأخيرة تسمى دالة الإمكان، ولغرض الحصول على المعالم المقدرة يتم اشتقاق دالة الإمكان لكل معلمة على حدة ومساواة المشتقة إلى الصفر أي أن

$$\frac{\partial \ln L(Y)}{\partial B_i} = 0, i = 1, \dots, k$$

وحل المعادلات الناتجة وبذلك نحصل على المعالم المقدرة بطريقة الإمكان الأعظم وعليه فإن الصيغة النهائية لمقدرات الإمكان الأعظم هي كالآتي

$$\dots (10) \hat{\beta}_{MLE} = (x'x)^{-1}x'y$$

ب- طريقة الإمكان الأعظم المشذب (Maximum Trimmed Likelihood Method (MTL) [6]pp.141:

هوبر في عام (1964, 1972) وهامبل في عام (1974) اقترحوا طرق للتعامل بحصانة مع الشواذ في العمليات الإحصائية تستند على نسبة الإمكان الغير مقيدة، فكرة هوبر هي استعمال دوال مقيدة الدرجة (مقدرات M) لدوال معلميه مختلفة، في حين إن أسلوب هامبل يكون أقرب إلى المفاهيم التفاضلية (*differentiability concepts*) الذي يؤدي إلى دوال تأثير محدودة.

النظريتان تعطيان إمكانية إيجاد آلية حصينة مثلى، ولكن من الناحية العملية تصبح حساباتها معقدة جداً خصوصاً في حالة متعدد المتغيرات.

لنفترض دالة الكتلة المعملية $\{f_\theta\}$ ، حيث لابد من التعرف على مجال الدالة وكالآتي

$$S(F, \theta) = \int \log f_\theta(x) J[F\{y : \log f_\theta(y) \leq \log f_\theta(x)\}] dF(x) \quad \dots (11)$$

حيث إن

$$\dots (12) J(t) = \begin{cases} 0 & t \leq \alpha \\ 1 & \alpha < t \leq 1 \end{cases}$$

ومن البديهي يتم تقدير المعلمات من خلال

$$\dots (13) \hat{\theta}(F) = \arg \theta \{ \max_\theta S(F, \theta) \}$$

حيث $\hat{\theta}(F) = \theta$ ، إن القيم الصغيرة جداً من $\log f_{\theta}(x)$ لا يتم إعطاؤها أي وزن في المقدار $S(F, \theta)$ وهذا مكافئ إلى تشذيب المشاهدات التي لديها فرصة قليلة في الظهور، سوف نطلق على هذه الفكرة مصطلح "مبدأ الإمكان المشذب" (*Trimmed Likelihood Principle*)، هذا الأسلوب إلى حد ما قد أستعمل في متعدد المتغيرات، Campbell (1980) يوصي باستعمال مقدرات M للوسط الحسابي ومصفوفة التغاير مسبقاً لتحديد المشاهدات الشاذة من خلال رسم الاحتمالات لدالة مسافة مهلنوبس، يترتب على ذلك بتر القيم الشاذة وتطوير طرق تقدير المعلومات التي قد تؤدي إلى أساليب تقديرية يمكن أن تكون موازية إلى (13)

وفي سياق متصل Broffit و Clarke و Lachenbruch (1980) أخذ بنظر الاعتبار تشذيب نسبة من المشاهدات التي تكون مسافة مهلنوبس لها كبير قبل تعظيم الإمكان.

إن تشذيب نسبة من المشاهدات التي تكون بعيدة (مسافة مهلنوبس لها كبير) في الحقيقة هو نفس تطبيق مبدأ الإمكان المشذب للتوزيع الطبيعي في حالة بعد واحد (متغير واحد).

لأسباب الاشتقاق المحاذية يثبت أنه من المناسب النظر في التقدير الدالي باعتباره حلاً لمعادلة التقدير

$$L(F, \theta) = \int \phi(x, \theta) J[F\{y: \log f_{\theta}(y) \leq \log f_{\theta}(x)\}] dF(x) = 0 \quad (14)$$

حيث

$$\phi(x, \theta) = \frac{\partial}{\partial \theta} \log f_{\theta}(x)$$

بمعنى أن التقدير إلى $T(F)$ هو الحل إلى $L(F, \theta) = 0$ ، ويتضح أن مقدر الموقع في هذه الطريقة هو مماثل إلى مقدر المربعات الصغرى المشذبة، وهذه المقدرات لديها امتداد مقارب إلى النموذج الطبيعي التي تتزامن مع مقدرات M المعروفة باسم تخطي المتوسط (*skipped mean*) pp.158 [7]، على أي حال فإن إيجاد هذه المقدرات بهذه الطريقة ترتبط بالنماذج المعلمية (طبيعي، أسّي... الخ)، مقدر الإمكان المشذب يستطيع التعامل مع بيانات ملوثة غير متماثلة وكذلك مع بيانات متماثلة، وهذا لا يحدث في العديد من طرق التقدير الحصينة لمعلمة الازاحة.

وثمة هناك بديل آخر مثير للأهتمام هو عندما تكون الكثافة متناسبة إلى $e^{-|x|}$ ، فإن الاجراء في هذه الحالة يكون مقابل إلى تصغير المقدار الآتي

$$\sum_{i=1}^{[n(1-\alpha)]} |r|_{i:n} \quad \dots (15)$$

حيث أن $|r|_{i:n}$ هو أكبر خطأ مطلق من الرتبة i ، على الرغم من أن الكثافة غير قابلة للاشتقاق عند الصفر، نستطيع أن نجد خصائص الحل (المقدر) بطريقة مباشرة.

إن طريقة الإمكان الأعظم المشذب في تقدير معالم نموذج الانحدار هي مشابهة إلى حد ما طريقة المربعات الصغرى المشذبة ولكن الفرق فيما بينهما هو أن في طريقة المربعات الصغرى المشذبة يتم اختيار المشاهدات التي يكون مجموع مربعات البواقي لها صغير، أما في طريقة الإمكان الأعظم المشذب فإنه يتم اختيار المشاهدات ذات الإمكان العالي أي تلك التي يكون الإمكان لها عال وتطبق طريقة الإمكان الأعظم الاعتيادية على هذه المشاهدات الجديدة.

إن الفرق بين طريقة المربعات الصغرى المشذبة وطريقة الإمكان الأعظم هو كالاتي، في طريقة المربعات الصغرى المشذبة يتم اختيار المشاهدات التي تكون البواقي لها صغيرة، أما في طريقة الإمكان الأعظم المشذب فيتم اختيار المشاهدات حسب مساهمتها في دالة الإمكان أي يتم اختيار المشاهدات ذات الإمكان العالي، وبعد أن تم تحديد المشاهدات المطلوبة يمكن أن يتم تقدير معالم نموذج الانحدار بطريقة الإمكان الأعظم المشذب والتي يمكن إيجادها على النحو التالي

$$\hat{\beta}_{MTLE} = (x'x)^{-1}x'y \quad (16)$$

ج- طريقة الإمكان الأعظم المشذب الموزونة pp.252 [10]:

إذا كان (X) متغير عشوائي يمتلك دالة كثافة احتمالية $p(x; \theta)$ التي تعتمد على المعلمة المجهولة (θ)، وافترض أن $(x_1, x_2, \dots, x_n)$ هي قيم ذلك المتغير العشوائي (X)، إن مسألة تقدير المعلمة المجهولة (θ) تتطلب طرق للتقدير، ومن الطرق الكلاسيكية تبرز طريقتي المربعات الصغرى وطريقة الإمكان الأعظم $Maximum\ likelihood(ML)$.

مقدر الإمكان الأعظم لـ (θ) هو قيمة (θ) التي تعظم دالة الإمكان (2) أو بشكل مماثل لو غاريتم دالة الإمكان

$$\sum_{i=1}^n \ell(\theta; x_i) \quad \dots (17)$$

حيث $\ell(\theta; x) = \ln p(x; \theta)$ هي دالة بدلالة (θ) التي تمثل مساهمة المشاهدة (i) الى لو غاريتم دالة الإمكان (17)، على الرغم من طرق التقدير الكلاسيكية هذه لديها خصائص مرغوب بها (خصوصاً طريقة ML) حيث أنها تنتج مقدرات تكون حساسة تجاه الشواذ، وفي السنوات الأخيرة تم التركيز بصورة واسعة على طرق التقدير الحصينة وهي الطرق التي تنتج مقدرات مقاومة تجاه وجود القيم الشاذة.

إن الطريقة المستخدمة لها علاقة مع الطرق الآتية:

- مقدرات M و GM .
- المربعات الصغرى المشدبة (LTS).
- مربعات الوسيط الصغرى (LMS).
- مقدر أصغر قطع ناقص (MVE).

الفكرة خلف هذه الطرق الحصينة هي استبدال معيار المربعات الصغرى بمعيار آخر يكون أقل حساسية تجاه وجود القيم الشاذة، إن الطرق المذكورة أعلاه مشتركة في جانب واحد، كل واحدة تستند نوعاً ما على تشذيب المشاهدات (لاحظ أن إعطاء الوزن هو نوع من التشذيب)، على سبيل المثال البيانات التي تقع بعيداً عن المركز يتم إعطاؤها وزن صغير أو تُحذف تماماً (يتم إعطاؤها وزن مساوي الى الصفر).

إن التشذيب يتطلب أن تكون البيانات مرتبة بترتيب معين (تصاعدي أو تنازلي)، الترتيب في حالة بيانات المتغير الواحد ($univariate$) يكون بصورة سهلة ولكن في حالة متعدد المتغيرات يكون ذلك صعباً.

في هذه الطريقة سوف نستبدل دالة الإمكان ببديل عنها أو بنسخة مشدبة منها أي أن المبدأ الذي تستند عليه هذه الطريقة هو تشذيب دالة الإمكان بدلاً من تشذيب المشاهدات مباشرة، لأنه إذا تم التشذيب من المشاهدات فإننا لا نعلم أي مشاهدة تحمل معلومات أكثر حول الظاهرة المدروسة لذلك يتم التشذيب من دالة الإمكان، ويتم تشذيب المشاهدات وفقاً لمساهمتها في دالة الإمكان بعد ترتيبها بنمط معين.

درجة التشذيب تعتمد على بعض المعالم التي تكون محددة من قبل محلل البيانات، مع الأخذ بنظر الاعتبار أنه كلما زاد تشذيب يعطي حصانة أكثر ولكن المقدرات الناتجة سوف تكون ذات كفاءة قليلة.

توجد بعض الملاحظات التوضيحية وهي لأي قيمة معلومة لـ (θ) فإن

$$\ell(\theta; x_1) \geq \ell(\theta; x_2) \geq \dots \geq \ell(\theta; x_n)$$

حيث أن $\ell(\theta; x_1)$ هي مساهمة المشاهدة (i) الى لو غاريتم دالة الإمكان.

من الواضح أن المؤشر الأصلي للبيانات أو العداد ربما قد لا يجعل قيم دالة الإمكان مرتبة بالترتيب الموضح أعلاه، إذا كان الترتيب لا يتحقق عند قيمة معلومة لـ (θ) يمكننا ببساطة تغيير العداد للمشاهدات حتى تكون مرتبة.

الأسلوب المتبع في هذه الطريقة هو استبدال دالة الإمكان لدالة لو غاريتم دالة الإمكان المشدب

$$\sum_{i=a}^b w_i \ell(\theta; x_i) \quad \dots (18)$$

حيث أن $a \leq b$ و $(a, b) \in [1, 2, \dots, n]$ والاوزان $w_i \geq 0$ وبتعظيم المقدار أعلاه نحصل على المقدرات لـ (θ) وهذه المقدرات يرمز لها بالرمز $\hat{\theta}(a, b, w)$ ومن الواضح أنها تعتمد على a, b, w حيث أن w هو متجه يحتوي على الاوزان، في حالة $w_i = 1, a \leq i \leq b$ في هذه الحالة التقديرات يرمز لها بالرمز $\hat{\theta}(a, b)$.

عندما تكون $a=1, b=n$ فأئنا نحصل على مقدرات الإمكان الأعظم لـ (θ) ، و، أما إذا كانت $a > 1$ أما $a < n$ أو $b < n$ حيث في هذه الحالة يتم تشذيب لوغاريتم دالة الإمكان ويشار الى هذه الطريقة بطريقة الإمكان الأعظم المشذب *Maximum Trimmed Likelihood (MTL)* وللمقدرات الناتجة بهذه الطريقة بمقدرات الإمكان الأعظم المشذب *Maximum Trimmed Likelihood Estimator (MTLE)*.

هناك بعض القيم المهمة لـ (a, b)

1. عندما تكون $\hat{\theta}(1, n)$ فأئنا نحصل على مقدرات الإمكان الأعظم.
2. عندما تكون $\hat{\theta}(1, n)$ والبيانات تتوزع توزيع طبيعي فأئنا نحصل على مقدرات المربعات الصغرى.
3. عندما تكون $a=1$ و $b < n$ أي $\hat{\theta}(1, b)$ هي مقدرات المربعات الصغرى لـ *Neykov* و *Vandev*، وإذا أضفنا معلمة موقع (θ) والبيانات لها توزيع طبيعي فإن $\hat{\theta}(1, b)$ تصبح مقدرات *(LTS)*
4. عندما تكون n عدد فردي وأن $a = b = \frac{1}{2}(n + 1)$ وأن هذا المقدار يكون مساوي الى m أي أن $m = \frac{1}{2}(n + 1)$ فإن دالة الهدف (19) ، تختزل الى $\ell(\theta; x_i)$ ، وبالتالي $\hat{\theta}(m, m)$ هي قيمة θ التي تعظم الوسيط (بدلاً من الوسط الحسابي) للمقدار $\ell(\theta; x_i), i = 1, \dots, n$ ويجب ملاحظه أن وسيط الإمكان $\ell(\hat{\theta}(m, m); x_m)$ ، ليس بالضرورة أن يكون مساوي الى المقدار $\ell(\hat{\theta}(m, m); x_{(m)})$ حيث أن $x_{(m)}$ هي الاحصاء المرتبة ذات الرتبة (m) (الوسيط في هذه الحالة)، عندما تكون θ معلمة موقع وتوزيع البيانات يكون توزيع طبيعي، فإن $\hat{\theta}(m, m)$ هي مقدرات *LMS* ($\hat{\theta}_{LMS}$) كما سيتم توضيح ذلك لاحقاً.
5. عندما تكون (n) عدد زوجي، سيكون لدينا وسيطين أماكن (وسيطين إثنين) هما $\ell(\theta; x_m)$ و $\ell(\theta; x_{m+1})$ حيث أن $m = \frac{1}{2}n$ ، لذلك توجد ثلاث حالات لتقدير θ باستخدام (19) إن هذه الحالات تنتج عندما تمتلك (a, b) قيم معينة وكالاتي

الحالة الأولى: عندما تكون $a = m$ و $b = m + 1$

وبالتالي فأئنا نختار θ لتعظيم التركيبة الخطية (الوسط الحسابي إذا كان $w=0.5$) لوسيطي الإمكان، حيث سنرمز لهذه المقدرات بالرمز $\hat{\theta}_1(m, m + 1)$.

الحالة الثانية: عندما تكون $a = b = m$ أي أن

$$\text{maximize}_{\theta} \ell(\theta; x_m)$$

في هذه الحالة نختار θ التي تعظم أعظم وسيطي الإمكان، حيث يجب أن تكون $\ell(\theta; x_m) \geq \ell(\theta; x_1)$ و $\ell(\theta; x_2) \geq \dots \geq \ell(\theta; x_n)$ ، حيث سنرمز لهذه المقدرات بالرمز $\hat{\theta}_2(m, m)$.

الحالة الثالثة: عندما تكون $a = b = m + 1$ أي أن

$$\text{maximize}_{\theta} \ell(\theta; x_{m+1})$$

عندما تكون θ معلمة موقع وتوزيع البيانات توزيع طبيعي، فإن المقدرات الناتجة هي مقدرات *LMS* لـ θ ، حيث سنرمز لهذه المقدرات بالرمز $\hat{\theta}_3(m + 1, m + 1)$.

في حالة (5و4) سوف نشير الى هذه الطريقة بطريقة أعظم وسيط إمكان *Maximum Median Likelihood (MML)* وللمقدرات الناتجة من هذه الطريقة بـ *Maximum Median Likelihood Estimators (MMLE)*

خوارزمية لإيجاد مقدرات *MML*:

سيتم الأخذ بنظر الاعتبار خوارزميات لحساب مقدرات *MTL* في حالة كون المعلمة θ قيمه عديدة (قيمة واحدة) السبب وراء ذلك هو

1. عندما تكون θ متجه من القيم، فإن إيجاد *MTLE* حسابيا مسألة أكثر صعوبة

2. الحالة الأكثر مواجهه في التطبيق عندما تكون θ متجه من القيم هي التوزيع الطبيعي المتعدد والنموذج الخطي العام.

بسبب أن طريقة MML تنتج خوارزميات بسيطة الى حد ما، نأخذ بنظر الاعتبار الاتي خوارزميات لحساب $MMLE$ ، وهذه الخوارزميات فيما بعد تُعمم الى حالة MTL ، أما الان سننتقل الى خوارزمية $MMLE$.

أفترض أن لدينا عينة $\{x_1, x_2, \dots, x_n\}$ مستقلة ومتماثلة التوزيع (i.i.d) هي مشاهدات متغير عشوائي يمتلك دالة الكثافة الاحتمالية $p(x; \theta)$ التي تعتمد على المعلمة المجهولة θ .

أفترض $\{y_1, y_2, \dots, y_d\}$ هي المجموعة التي تحتوي على قيم متميزة من $\{x_1, x_2, \dots, x_n\}$ وأفترض أيضا $\{f_1, f_2, \dots, f_d\}$ هي التكرارات المقابلة لها.

أفترض أن θ_{rr} هي قيمة θ التي تجعل $\ell(\theta, y_r)$ يبلغ أعظم قيمة لذلك θ_{rr} نحصل عليها من خلال

$$\frac{\partial}{\partial \theta} \ell(\theta, y_r) = 0, r = 1, \dots, d \quad \dots (19)$$

وبافتراض أن الشروط الاعتيادية متحققة لـ $p(x; \theta)$ بحيث أن المعادلة تنتج قيمة θ التي تعظم $\ell(\theta, y_r)$ ضمن مجال القيم الممكنة لـ θ .

علاوة على ذلك أفترض θ_{rj} هي مجموعه قيم θ التي تحقق الاتي

$$\ell(\theta, y_r) = \ell(\theta, y_j); r = 1, \dots, d-1; j = r+1, \dots, d \quad \dots (20)$$

قيم θ_{rr} و θ_{rj} التي نحصل عليها من المعادلتين التاليتين (20) و (21) يمكن ترتيبها في مصفوفة مثلث علوي على النحو التالي

$$\begin{pmatrix} \theta_{11} & \theta_{12} & \dots & \theta_{1d} \\ & \theta_{22} & \dots & \theta_{2d} \\ & & \ddots & \vdots \\ & & & \theta_{dd} \end{pmatrix}$$

$\theta_{rr}(y_1, y_2, \dots, y_d)$ هي مجموعة (الضرورة) من القيم المميزة لـ θ_{rr} التي يتم الحصول عليها من المعادلة (19)، $\theta_{rj}(y_1, y_2, \dots, y_d)$ هي مجموعة القيم المميزة لـ θ_{rj} التي يتم الحصول عليها من (20) حيث أن

$$\theta(y_1, y_2, \dots, y_d) = \theta_{rr}(y_1, y_2, \dots, y_d) \cup \theta_{rj}(y_1, y_2, \dots, y_d)$$

افترض أن M هو عدد العناصر في $\theta(y_1, y_2, \dots, y_d)$ ومن الواضح أن

$$M \leq \frac{1}{2}n(n+1)$$

الخوارزميات لـ $MMLE$ تستند على المبرهنة التالية، حيث نفترض أن هناك على الأكثر حل واحد θ_{rj} الى (20) لكل زوج المؤشرات (r, j) ، هذه الافتراضات تسري على معظم التوزيعات التي نواجهها في التطبيق.

مبرهنة:

افترض $\theta(y_1, y_2, \dots, y_d)$ ، $\theta_{rr}(y_1, y_2, \dots, y_d)$ و $\theta_{rj}(y_1, y_2, \dots, y_d)$ ، θ تقدر بواسطة $\max_{\theta} \ell(\theta; x_m), m \in \{1, 2, \dots, n\}$ فإن

1. إذا كانت $\theta_{rj}(y_1, y_2, \dots, y_d)$ مجموعة خالية، فإن مقدر لـ θ هو عنصر يوجد في $\theta_{rr}(y_1, y_2, \dots, y_d)$ ، هذا المقدر يتم الحصول عليه بحل (19) عند $y_r = x_m$.

إذا كانت $\theta_{rj}(y_1, y_2, \dots, y_d)$ مجموعة ليست خالية، فإن مقدر لـ θ هو عنصر يوجد في $\theta(y_1, y_2, \dots, y_d)$ الذي يعظم $\ell(\theta; x_m)$. حيث إن برهان هذه المبرهنة موجود في [10].

إن هذه المبرهنة لها آثار مهمة وهي

1. عند وضع $m = \frac{1}{2}(n + 1)$ ، $MMLE$ يتم الحصول عليها كحالة خاصة.
2. طالما المقدرات تنتمي الى مجموعة التي عناصرها تكون مقيدة من الأعلى التي لا تتجاوز M ، إن المبرهنة تثبت وجود $MMLE$.
3. في الحالة الأولى يمكن أن نحصل على نتيجة مماثلة عند حل (19) عندما يكون $y_r = x_m$ ، ومما يجب ملاحظته أن $p(x_i; \theta)$ يجب أن تكون رتيبة في x في هذه الحالة.
4. للحالة الثانية، يمكن تطوير خوارزمية البحث لتحديد أي عنصر من المجموعة $\theta(y_1, y_2, \dots, y_d)$ يكون هو $MMLE$ ، حيث يتم توضيح ذلك في الخوارزمية التالية وهي كالآتي

1. باستخدام (19) نحسب قيمة θ_{rj} التي تجعل $\ell(\theta, y_r)$ ؛ $r = 1, \dots, d$ تبلغ أقصى حد، بالإضافة الى ذلك يتم حساب قيمة θ_{rj} باستخدام (20) ووفقاً للمبرهنة (4,1) فإن $MMLE$ يجب أن تكون أحد نقاط

2. أحسب قيمة θ_{rj} ؛ $r = 1, \dots, d$ ؛ $j = r, \dots, d$

بعد ذلك يتم ترتيب الأزواج (θ_{rj}, ℓ_{rj}) ترتيب تنازلي وفقاً الى قيم ℓ_{rj} .

3. يتم اختيار أول قيمة θ_{rj} في الخطوة الثانية حيث أن هذه القيمة هي التي ستكون $MMLE$ ، بحيث أن

$$\ell_{rj} = \text{median}\{\ell(\theta_{rj}; y_1), \dots, \ell(\theta_{rj}; y_d)\}$$

حيث يتم الأخذ بنظر الاعتبار التكرار المطلق لـ y_r .

هذه الخوارزمية تتطلب عدد من العمليات الحسابية والتي تكون مساوية الى $\frac{1}{4}d(d + 1)$ من التخمينات لـ θ_{rj} و ℓ_{rj} بالإضافة الى عدد عشوائي من التخمينات لـ $\ell(\theta, y_r)$ في حدود $\frac{1}{4}d^3$ ، هذه الخوارزمية على الرغم من الناحية النظرية بسيطة قد لا تكون كفؤة حسابياً.

خوارزميات أكثر كفاءة تستند على تتبع منحنى وسيط الإمكان تم تطويرها وتعميمها للحصول على مقدرات MTL و حتى الآن لا يعرف الكثير عن مقدرات (TL) لتوزيعات مختلفة عن التوزيع الطبيعي [11] pp.504 اما خوارزمية إيجاد MTL يمكن الرجوع اليها في (أنظر 10)

7- الجانب التطبيقي

7-1 وصف البيانات والعينة:

تم الحصول على عينة بحجم 91 من مختبر دائرة مجاري واسط وكانت هذه العينة تحتوي على متغير معتمد (TDS) وثلاث ملوثات مستقلة وهي (PO_4, CL, SO_4) ، حيث تم سحب بيانات هذه العينة من قبل المختصين وتم سحبها من مواقع مختلفة لمدينة الكويت لعام 2015، والبيانات موضحة في الجدول التالي.

جدول رقم (1) البيانات الحقيقية والتي تمثل متغيرات الدراسة

	Po4	cl	So4	Tds		Po4	cl	So4	Tds
1	0	3.65	7.81	16	25	0.08	1.84	4.93	10
2	0	3.5	5.76	14	26	0.01	2.65	6.17	16
3	0	4.48	9.87	22	27	0.2	3.45	3.29	12
4	0	10.64	17.28	34	28	0.2	2.48	4.52	16
5	0	3.82	7.81	20	29	0.2	2.58	6.17	18
6	0.082	1.99	4.93	12	30	0.08	1.41	3.7	12
7	0.2	2.42	3.29	10	31	0.2	2.49	3.29	14
8	0.158	3.4	6.17	18	32	0.052	3.41	4.52	14
9	0.2	5.01	10.28	26	33	0.13	1.29	2.88	10

10	0.2	2.97	4.11	12	34	0.2	4.58	2.88	18
11	0.075	1.87	4.93	12	35	0.02	3.31	9.87	22
12	0.12	3.82	6.17	22	36	0.116	2.03	2.05	10
13	0.2	2.12	4.93	12	37	0.17	8.28	5.76	38
14	0.2	3.49	5.34	18	38	0.195	1.56	2.88	14
15	0.2	4.89	8.64	24	39	0.156	1.36	2.05	10
16	0.145	1.95	3.29	10	40	0.2	1.53	1.23	8
17	0.2	2.44	2.46	12	41	0.16	2.83	2.88	16
18	0.2	82.95	51.02	242	42	0.2	1.82	3.29	14
19	0.013	3.22	3.29	10	43	0.2	2	2.88	10
20	0.041	2.49	5.76	12	44	0.15	1.77	2.88	10
21	0.19	2.59	2.05	12	45	0.112	2.49	4.11	12
22	0.2	4.4	4.11	14	46	0.132	2.69	3.7	14
23	0.15	3.4	5.76	16	47	0.19	2.59	3.7	14
24	0.182	4.37	8.64	22	48	0.11	4.08	5.76	18
49	0.068	2.02	2.47	10	71	0.06	3.17	6.17	14
50	0.098	2.85	2.88	16	72	0.2	2.42	3.7	16
51	0.2	1.9	5.76	12	73	0.2	2.81	4.52	14
52	0.021	2.49	5.34	16	74	0.2	1.95	2.46	8
53	0.11	3.63	6.99	24	75	0.2	1.17	3.7	10
54	0.08	3.02	14.4	34	76	0.1	4.68	7.4	20
55	0.2	41.09	16.04	58	77	0.2	3.51	6.58	22
56	0.19	2.84	3.7	12	78	0.156	2.92	3.29	16
57	0.2	4.8	5.76	20	79	0.152	1.36	3.7	10
58	0.2	1.98	2.05	10	80	0.18	1.75	2.88	10
59	0.2	2.97	6.58	18	81	0.2	1.98	2.88	12
60	0.2	2.38	2.05	10	82	0.2	2.44	3.29	10
61	0.2	1.78	3.7	10	83	0.028	3.36	9.05	22
62	0.14	2.76	4.93	16	84	0.02	2.12	4.93	14
63	0.136	3.12	7.4	14	85	0.11	1.77	3.29	10
64	0.04	2.65	3.7	12	86	0.2	2.65	28.39	82
65	0.09	3.54	15.22	40	87	0.082	2.42	0.82	8
66	0.2	2.97	6.99	18	88	0.14	1.41	4.11	12
67	0.04	1.81	4.52	14	89	0.09	2.41	6.58	16
68	0.088	1.41	2.05	10	90	0.108	2.12	3.7	12
69	0.14	2.46	0.82	12	91	0.143	1.84	3.29	12
70	0.13	34.46	16.46	66					

1-المواد الذائبة الكلية (TDS) [12]:

المواد الصلبة الذائبة هي كافة المواد التي تمر من خلال أوراق الترشيح المختبري مشتملة على المواد الغروية وتشكل المواد الصلبة الذائبة في المياه حوالي 90% والباقي على شكل غروي. تشكل المواد العضوية حوالي 40% من أجمالي المواد الصلبة الذائبة والباقي غير عضوي والجزء الغروي يكون في الغالب عضوي التركيب.

تعتبر المواد الصلبة الذائبة من أهم معايير صلاحية المياه والري وأن زيادة تركيز هذه المواد يؤثر على نمو النباتات حيث تحد من كميات المياه التي تمتصها الجذور لتقوم بعملية النتح وتقلل بالتالي كمية المياه المنتقلة الى الأوراق والمياه المالحة تؤدي الى تراكم الاملاح في التربة مما يستوجب غسل التربة بعد كل موسم.

2- الكلوريدات (CL) Chloride [12]:

تعتبر الكلوريدات من الايونات السالبة المهمة الموجودة في المياه الطبيعية وعند تواجدها بتركيز عالية يكتسب الماء تأثير تآكلي قد تظهر على الانابيب و المنشأة المعدنية ويتم فحص الكلوريدات عند استخدام الماء للري في المياه المعالجة وذلك لتأثيره على أوراق المحاصيل الحساسة ويكون على شكل املاح الصوديوم والبوتاسيوم و الكالسيوم و يعطي الطعم المالح إذا ارتبط مع أيون الصوديوم وشكل ملح كلوريد الصوديوم (ملح الطعام) NaCl, ولا يعطي الطعم المالح إذا ارتبط مع الايونات الأخرى مثل الكالسيوم و أيون المغنيسيوم, وتحتوي فضلات الانسان مثل العرق والادرار على (6) غرام من الكلوريدات للشخص الواحد في اليوم بسبب ملح الصوديوم الذي لا يتغير عبر مروره في الجهاز الهضمي لذلك ينطرح مع الادرار, وإن مصدر الكلوريدات تأتي من الترسبات و الصخور (أي من تحلل الصخور الموجودة في الأنهر وتحلل الرمال و من ادرار الانسان) ولا يمكن تقدير نتيجة الكلوريدات من اللون إلا من الحجم.

3- الكبريتات SO4 Sulphate [12]:

أن وجود الكبريتات في المياه نتيجة المخلفات الصناعية منها السيارات، الأسمدة، الكيماويات الغير عضوية، البلاستيك، تكرير البترول، الصلب، الصناعات الورقية. إن تأثير الكبريتات في المياه يسبب العسرة الدائمة وتكون على شكل كبريتات الكالسيوم والمغنيسيوم. وتحديد صلاحية المياه للري والبناء (ضروري لنمو النبات، ويؤثر على نوعية الكونكريت ويعمل على تفتته) ويعطي رائحة غير مقبولة للماء ويسبب تآكل الانابيب.

4- الفوسفات (PO4) Phosphate [12]:

أن المصدر الرئيسي للفوسفات في المياه العادمة هو الأسمدة ومساحيق الغسيل وللكشف عنها يتم الاعتماد على اللون حيث أن شدة اللون الأصفر تتناسب مع تركيز الفوسفات الموجود في المياه العادمة ويستخدم جهاز سبكتروفوتوميتر لقياس الامتصاصية أو النفاذية ومن خلالها نستطيع أن نعرف تركيز الفوسفات باستخدام الرسوم البيانية المعدة مسبقاً. حيث أن الفوسفات يفيد في تقييم الطاقة البيولوجية للمياه السطحية وتشغيل محطات التنقية ودراسة تلوث الجداول.

ان الهدف من هذا الاجراء هو معرفة تأثير المواد الصلبة الذائبة (الاملاح) في مياه الصرف الصحي والتي تشمل الكبريتات والكلوريدات والفوسفات والتي تصبها محطات الصرف الصحي في مياه النهر مما يؤثر على الانسان والنبات والتربة.

2-7 النتائج

باستخدام المصفوفة (1) تم الكشف عن القيم الشاذة في المتغيرات المستقلة حيث تم الكشف عن كل متغير بصورة منفردة لمعرفة القيم الشاذة في كل متغير وأيضا معرفه قيمتها، وتم وضع هذه القيم الشاذة في الجدول التالي.

جدول رقم (2)

القيم الشاذة في كل متغير بواسطة المصفوفة h

Cl	So4
8295	5102
4109	2839
3446	

إن القيم الشاذة توجد فقط في كل من Cl و So4, وبعد أن تم التعرف على القيم الشاذة بواسطة المصفوفة h تم تطبيق كل من الطرق الاتية (MLE, MTLE, MWTLE) باستخدام برنامج (MATLAB 2015) على هذه الملوثات المذكورة أعلاه وكانت نتائج هذه الطرق على النحو التالي:

جدول رقم (3)
 معالم النموذج حسب كل طريقة

Pram. Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
MLE	-4.4179	1.0845	27.6734	2.5566
	(1.5405) *	(0.1025) *	(8.8035) *	(0.1604) *
MTLE	2.4713	0.7596	6.8637	1.7596
	(1.0003) *	(0.0610) *	(4.4868) *	(0.1262) *
MWTLE	2.4882	0.7166	5.4762	1.8256
	(0.5871) *	(0.0782) *	(2.5903) *	(0.0898) *

*: تمثل الانحراف المعياري لكل معلمة عند كل طريقة

جدول رقم (4)
 تباين الخطأ وتباين النموذج وكفاءه النموذج ومعيار RMSE

Pram. Method	$v(\hat{y})$	$v(e)$	RMSE	eff**
MLE	651.5331	30.1015	5.4865	100
MTLE	110.4754	3.9237	1.9808	16.9562
MWTLE	109.3834	3.9884	1.9972	16.7886

** : تمثل كفاءة النموذج عند القيم التقديرية لـ $(\hat{y})$

8- الاستنتاجات

1. ان طريقة الإمكان الأعظم المشذب أفضل من طريقة الإمكان الأعظم الاعتيادية حيث أدت الى تحسين التقدير وأدت الى التقليل من تأثير وجود الشواذ وتلاحظ ذلك من خلال كفاءة معيار (RMSE) الذي كانت قيمة مساوية الى 1.9808 في جدول رقم (4).
2. طريقة الإمكان الأعظم المشذب كانت افضل من طريقة الإمكان الأعظم الموزون المشذب مما يدل على ان طريقة الإمكان الأعظم المشذب هي الأفضل

9- التوصيات

1. زيادة الوعي البيئي ويتم ذلك من قبل مختصين في مجال البيئة والصحة حيث يتم ذلك عن طريق عمل دورات تثقيفية في مجال المحافظة على البيئة أو عن طريق عمل بعض المنشورات أو الإعلانات الدعائية والتي تهدف الى زيادة الوعي البيئي لدى كافة الناس.
2. زيادة كفاءة محطات معالجة مياه الصرف الصحي في معالجة التلوث والاهتمام بها من ناحية الأجهزة المخبرية والمعدات اللازمة والمتطورة لعمل هذه المحطات على أتم وجه، ومحاولة التقليل من وجود هذه الملوثات وعدم رميها في مياه الأنهر بصورة مباشرة الا بعد التخلص من هذه الملوثات الضارة.

المصادر

- [1] Tiku, M. L. (2004). Robust estimation and hypothesis testing. New Age International.
- [2] Hawkins, D. M. (1980). Identification of outliers (Vol. 11). London: Chapman and Hall.
- [3] Freeman, P. R. (1980). On the number of outliers in data from a linear model. *Trabajos de estadística y de investigación operativa*, 31(1), 349-365.
- [4] Huber, P. J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *The Annals of Statistics*, 799-821.
- [5] Jin, W., Tung, A. K., & Han, J. (2001, August). Mining top n local outliers in large databases. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 293-298). ACM.
- [6] Bednarski, T., & Clarke, B. R. (1993). Trimmed likelihood estimation of location and scale of the normal distribution. *Australian & New Zealand Journal of Statistics*, 35(2), 141-153.
- [7] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (2011). Robust statistics: the approach based on influence functions (Vol. 196). John Wiley & Sons.
- [8] Vandev, D. (1993). A note on the breakdown point of the least median of squares and least trimmed squares estimators. *Statistics & probability letters*, 16(2), 117-119.
- [9] Leroy, A. M., & Rousseeuw, P. J. (1987). Robust regression and outlier detection. *Wiley Series in Probability and Mathematical Statistics*, New York: Wiley, 1987.
- [10] Hadi, A. S., & Luceño, A. (1997). Maximum trimmed likelihood estimators: a unified approach, examples, and algorithms. *Computational Statistics & Data Analysis*, 25(3), 251-272.
- [11] Müller, C. H., & Neykov, N. (2003). Breakdown points of trimmed likelihood estimators and related estimators in generalized linear models. *Journal of Statistical Planning and inference*, 116(2), 503-519.
- [12] American Public Health Association, & American Water Works Association. (1989). *Standard methods for the examination of water and wastewater*. American public health association.

