

التنبؤ بمؤشر جودة المياه السطحية باستخدام التحليل التمييزي والانحدار الخطي المتعدد

د.م.رائد سلمان جعفر*

(الإيداع: 7 حزيران 2022، القبول: 10 آب 2022)

الملخص:

تضمنت هذه الدراسة تقييم وتفسير قيم القياسات المتعددة لبارامترات جودة المياه السطحية ومصادر التلوث في بحيرة سد الصوراني - طرطوس - سوريا، التي تسمح لنا بالحصول على معلومات متقدمة حول جودة المياه، التقييم والتصميم الأمثل لشبكة محطات المراقبة في منطقة الدراسة. لقد انتشرت تقنيات التنبؤ الإحصائية على نطاق واسع في الوقت الحاضر، وذلك من أجل حماية موارد المياه واستدامتها. تم حساب مؤشر الجودة الشامل (CPI) في محطات المراقبة السبعة المعتمدة لجميع القياسات في الفترة الزمنية بين كانون الثاني 2018 وكانون الأول 2020، إذ ينجم هذا التلوث الخفيف عن وصول مياه الصرف الصحي لتجمعات القرى وعن المنشآت الزراعية والصناعية والسياحية المجاورة للبحيرة. يشير هذا المؤشر بشكل عام إلى أن جودة المياه في البحيرة مقبولة إلى مقبولة بشكل أساسي. في هذه الدراسة، تم تطبيق تقنيات إحصائية، بما في ذلك التحليل التمييزي (DA) وتحليل الانحدار الخطي المتعدد (MLR) لتقييم جودة المياه في البحيرة. ركزت الدراسة على تحليل 21 بارامتر فيزيائية وكيميائية وميكروبيولوجية، في عينات المياه التي تم جمعها شهرياً على مدار ثلاثة أعوام (2018-2020) من 7 مواقع اعتيان مختلفة تقع حول وداخل البحيرة. تضمن التحليل الاستكشافي في البداية للبيانات المخبرية الدراسة الإحصائية الوصفية مع استبعاد القيم الشاذة والمعالجة التمهيدية للمعطيات. يتضح من النتائج أن قيم (CPI) المحسوبة من خلال نماذج الانحدار الخطي المتعدد المطبقة على البحيرة تحقق أعلى معامل ارتباط ($R = 0.98$). تساعد هذه التقنيات في تزويد سلطات إدارة المياه بالمعرفة المتعلقة باستخدام أو تعديل أو التنبؤ بمؤشر جودة المياه، والمساهمة في التخطيط المستقبلي لدراسات المياه لأغراض الشرب.

الكلمات المفتاحية: جودة المياه، التحليل التمييزي، تحليل الانحدار الخطي المتعدد، بحيرة سد الصوراني، مؤشر التلوث الشامل (CPI).

*أستاذ مساعد قسم الهندسة البيئية - كلية الهندسة المدنية - جامعة تشرين.

Predicting of surface water quality index using Discriminant analysis and multiple linear regression

Dr.Raed Selman Jafar*

(Received: 7 June 2022, Accepted: 10 August 2022)

Abstract:

This study focuses on evaluation and interpretation of multiple measurements data for surface water quality parameters and pollution sources in the Sourani dam lake – Tartous– Syria, which is permitting us to acquire innovative information about water quality, evaluation and optimal design of monitoring network. Currently, evaluation and forecasting techniques have commonly disseminate, in order to protect and sustain water resources.

The comprehensive pollution index (CPI) was calculated in the seven monitoring stations for all measurements during January 2018 and December 2020, as the source of this pollution outcomes from villages sewage water, agricultural and industrial and tourist facilities nearby to the lake. This index indicates generally that the water quality in the lake is qualified and basically qualified.

In this study, statistical techniques, including discriminant analysis (DA) and multiple linear regression (MLR) were applied to assess the water quality of the lake. The study concerns the analysis of 21 physical, chemical and microbiological parameters in water samples collected monthly over a period of three years (2018–2020) from 7 different sampling sites located around and within the lake. Exploratory analysis of laboratory data initially included statistical description, outliers elimination and pretreatment were done.

It is evidence from the results that the (CPI) calculated through multi–linear regression models were applied to the lake achieves the highest correlation coefficient of ($R = 0.98$). These techniques support and offer water management authorities with knowledge associated to the use, adjustment or prediction of a water quality index, and thus contribute to the future development studies of potable water.

Keywords: water quality, discriminant analysis (DA), multiple linear regression (MLR), Sourani–dam lake, comprehensive pollution index (CPI).

*Associate professor – Department of Environmental Engineering – Faculty of civil Engineering – Tishreen University.

1- المقدمة Introduction

يُعدّ تلوث المياه السطحية بالملوثات المتعددة الفيزيائية والكيميائية والبيولوجية مشكلة منتشرة في العديد من دول العالم. نظراً لأهمية الموارد المائية السطحية والتأثيرات التي يمكن أن يخلفها تلوث هذه الموارد على صحة الإنسان والنظام البيئي، تعتبر برامج إدارة جودة المياه حلاً ضرورياً وأساسياً للحماية والاستخدام الأمثل والمستدام لهذه الموارد.

بشكل عام، يمكن حصر المكونات الرئيسية لبرامج إدارة جودة الموارد المائية في ثلاثة محاور رئيسية هي الوقاية (Prevention) والرصد (Monitoring) والتحكم (Control). في مرحلة الوقاية، تُبذل الجهود لاتخاذ جميع الإجراءات والتدابير اللازمة لمنع تلوث موارد المياه السطحية وتدهور جودة مياها باستخدام أدوات مثل تقييم الأثر البيئي (EIA)، بينما تتضمن مرحلة الرصد برامج يتم تنفيذها لتقييم جودة المياه ومراقبتها، وأخيراً تشمل مرحلة التحكم إجراءات تنفيذية لمنع تلوث المياه الناتج عن مصادر التلوث المختلفة. تعتبر مرحلة الرصد في برامج إدارة جودة الموارد المائية هي المرحلة الوسطى التي تربط بين الوقاية والتحكم. الهدف من رصد جودة المياه هو الحصول على معلومات كمية حول التغيرات الزمانية والمكانية لبارامترات جودة المياه بحيث تعكس هذه البارامترات المقاسة في نقاط الاعتيان المختلفة قدر الإمكان التغيرات الإجمالية في جودة مياه المورد المائي المدروس.

نظراً لحاجة رصد جودة المياه إلى قياس بارامترات مختلفة، والذي بدوره يتطلب كلفة كبيرة ووقت طويل، فإن تصميم برنامج لرصد جودة المياه بأقل عدد ممكن من نقاط الاعتيان واختزال عدد البارامترات المقاسة وعدد مرات القياس، وبنفس الوقت يعطي تقييم بشكل صحيح لجودة المياه بأقل كلفة ممكنة يعتبر أمراً ضرورياً وهاماً جداً.

يلعب التحليل الإحصائي دوراً هاماً في تحليل وتفسير الظواهر الطبيعية والبيئية، ويعد أحد أهم طرق البحث العلمي الذي يستخدم عند دراسة المشاكل المختلفة البيئية والصحية والاقتصادية. إحدى الطرق التي تم استخدامها على نطاق واسع في السنوات الأخيرة لتصميم شبكات رصد جودة المياه هي الطرق الإحصائية متعددة المتغيرات (Multivariate Statistical Methods)، نذكر منها: التحليل العنقودي (Cluster Analysis)، تحليل المكونات الرئيسية (Principal Component Analysis)، التحليل التمييزي (Discriminant Analysis)، تحليل الانحدار الخطي متعدد المتغيرات (Multiple Linear Regression) (Hamil, Arab et al. 2018).

يعد أسلوب التحليل التمييزي من الأساليب الإحصائية للتحليل متعدد المتغيرات والذي يشاع استخدامه في المجالات البيئية، حيث يهتم التحليل التمييزي بكيفية التمييز بين مجموعتين أو أكثر من القياسات، وتصنيف القياسات الجديدة على المجموعات التي سبق تعريفها، يعتمد أسلوب التحليل التمييزي على الوصول إلى دالة تسمى دالة التمايز (Discriminate Function) والتي تعمل على زيادة الفروق بين متوسط المجموعات حيث أنه كلما كان هناك تباعد بين متوسط المجموعات كلما كان التمييز أكثر كفاءة، وبالتالي التخفيض من خطأ التصنيف (Liu, Zhang et al. 2021). يعتبر التحليل التمييزي استكشافياً بطبيعته حيث يستكشف أسباب الاختلاف المشاهدة عندما لا تستطيع فهم العلاقات السببية بدرجة كافية من الدقة (Chen, Zhang et al. 2018).

يعدّ مؤشر جودة المياه WQI الأداة الأكثر فعالية لتحويل بيانات بارامترات جودة المياه المعقدة الخام إلى معلومات بسيطة ومفهومة تعكس الحالة الكلية لتلوث المصدر المائي المدروس، بشكل عام تتراوح قيم بعض مؤشرات جودة المياه (WQIs) من 0 إلى 100. كلما زادت القيمة كلما كانت جودة المياه أفضل. تمثل القيمة 0 أسوأ جودة للمياه بينما تمثل القيمة 100 أفضل جودة للمياه، القيم بينهما تصنف في فئات جودة متدرجة ومختلفة.

تم تقديم هذه الطريقة من خلال المعادلات الرياضية لتحديد حالة جودة المياه في المصدر المائي بواسطة Horton (Uddin, Nash et al. 2021). بعد ذلك، تم تطوير عدة طرق لحساب مؤشرات جودة المياه لأغراض مختلفة من قبل العديد من

الهيئات والمنظمات في جميع أنحاء العالم، مثل مؤشر جودة المياه لمؤسسة الصرف الصحي الوطنية (Uddin, Nash et al. 2021) (NSFWQI)، مؤشر جودة المياه لمجلس الوزراء الكندي البيئي (CCMEWQI) (2001) (CCME)، مؤشر جودة المياه في كولومبيا البريطانية (BCWQI) (Zandbergen and Hall 1998)، مؤشر جودة المياه في ولاية أوريغون (OWQI) (Cude 2001)، مؤشر جودة المياه في ماليزيا (MWQI) (Shuhaimi-Othman, Lim et al. 2007)، مؤشر جودة المياه الشامل (CPI) (Vasanthi and Kumar 2019) (Jindal and Wats 2022).

يُعدّ الانحدار الخطي المتعدد (MLR) من الأساليب الإحصائية المتقدمة والتي تضمن دقة الاستدلال من أجل تحسين نتائج البحث عن طريق الاستخدام الأمثل للبيانات في إيجاد علاقات سببية بين ظواهر موضوع البحث. والانحدار الخطي المتعدد هو عبارة عن إيجاد معادلة رياضية تعبر عن العلاقة بين متغيرين أو أكثر وتستعمل لتقييم قيم مقاسة وللتنبؤ بقيم مستقبلية، وهو عبارة أيضاً عن انحدار للمتغير التابع على العديد من المتغيرات المستقلة (Valentini, 2021). بالتالي يتم استخدام الانحدار الخطي المتعدد لشرح العلاقة بين متغير تابع مستمر ومتغيران مستقلان أو أكثر. يمكن أن تكون المتغيرات المستقلة مستمرة أو منقطعة. تعتمد فكرته على العلاقات الدلالية التي تستخدم ما يعرف بشكل التشتت أو الانتشار، فبإمكاننا التنبؤ باحتمالية تلوث المصدر المائي ودرجته مثلاً بالاعتماد على دراسة بارامترات التلوث المختلفة لهذا المصدر المائي (Arawo 2020).

يختلف أسلوب التمايز مع كل من تحليل التباين وتحليل الانحدار، حيث يكون المتغير التابع نوعي، بينما يكون في الأسلوبين الآخرين متغير كمي.

ونظراً لأن تلوث المياه الخطير يهدد بشكل مباشر صحة الإنسان، فإن برامج الرصد والتقييم المستمرة لنوعية المياه السطحية في بحيرة سد الصوراني (حالة الدراسة) أمر ضروري. من المشكلات الهامة في مراقبة جودة المياه هي التعقيد المرتبط بتحليل عدد كبير من المتغيرات، كل منها يحتوي على معلومات غنية حول خصائص موارد المياه، يمكن التعامل معها بشكل أكثر سهولة ودقة من خلال تطبيق الأساليب متعددة المتغيرات كالتحليل التمييزي وتحليل الانحدار الخطي المتعدد (Zhao, Xia et al. 2012).

2- أهمية البحث وأهدافه Research importance and objectives

تتجلى أهمية البحث في تقييم جودة مصادر المياه السطحية والحفاظ عليها من التلوث، من خلال دراسة بارامترات التلوث المختلفة الفيزيائية والكيميائية والبيولوجية باستخدام التقنيات الرياضية الإحصائية متعددة المتغيرات مثل التحليل العنقودي (Cluster Analysis (CA)، التحليل التمييزي (Discriminant Analysis (DA) وتحليل الانحدار الخطي المتعدد (Multiple linear Regression (MLR). تساهم هذه التقنيات في تفسير مجموعات البيانات، تقييم جودة المياه السطحية، تحديد مصادر التلوث المختلفة، كما تساعد في تقييم شبكة المراقبة وتخفيض عدد مواقع الاعتيان بالإضافة إلى خفض عدد التجارب المخبرية المختلفة وتواترها، بما يساهم في تخفيض الكلفة والوقت.

تهدف هذه الدراسة من خلال استخدام هذه التقنيات إلى تقييم نتائج التحاليل المخبرية لبارامترات تلوث بحيرة سد الصوراني في محطات الاعتيان المعتمدة، إيجاد أوجه التشابه والاختلاف الموجودة بين معايير جودة المياه المختلفة للمصدر المائي، وتحديد المتغيرات الخاصة بدراسة الاختلاف المكاني والزمني في البحيرة. كما تهدف أيضاً إلى التأكد من تأثير مصادر التلوث السكنية والزراعية والصناعية المحيطة بالبحيرة على جودة المياه، وذلك بحساب مؤشر تلوث المياه الشامل للبحيرة (Comprehensive Pollution Index (CPI) والتنبؤ به.

3- مواد وطرائق البحث Research materials and methods

3.1- منطقة الدراسة Study area

تقع منطقة الدراسة (بحيرة سد الصوراني) في منطقة الجبال الساحلية في سورية، التي تتميز بميول شديدة وانحدارات صعبة، على مسافة 50 Km شمال شرق مدينة طرطوس. يمتد الموقع بين خطي الطول $36^{\circ}07'30''E$ و $36^{\circ}06'54''E$ شرقاً ودائرتي العرض $35^{\circ}01'76''N$ و $35^{\circ}01'84''N$ شمالاً، تبلغ مساحة الحوض الساكب 76.65 Km^2 ، هذا ويوضح الشكل (1) موقع بحيرة سد الصوراني (حالة الدراسة). وهي مخصصة لأغراض الشرب وفق الدراسة الأولية والفنية للمشروع، حيث تجمع مياه الينابيع المنتشرة بوفرة في المنطقة.

يبلغ طول البحيرة 2.8 Km، والعرض الوسطي للبحيرة 119.9 m، والعمق الوسطي يصل حتى 13.4 m، أما مساحة سطح البحيرة فتبلغ 335720 m^2 ، وحجم التخزين الطبيعي 4.5 million m^3 . يجاور السد مجموعة من القرى والتجمعات السكنية الواقعة ضمن الحوض الصباب للنهر، ويقدر عدد سكانها بحوالي 12000 نسمة (وهي قرى: الصوراني، بغيدات، المجيدل، الوادي الأخضر وبرماننة المشايخ).

تصل الملوثات الخام لأغلبية التجمعات السكنية إلى بحيرة السد بدون أي معالجة. إن التركيب الجيولوجي لحوض نهر الصوراني هو عبارة عن مجموعة من التشكيلات الصخرية المختلفة، أهمها الصخور الكلسية الدولوميتية والصخور المارلية. يتشكل الوادي الرئيسي للسد من التقاء رافدي نهرين أساسيين، الأول: هو رافد السعنونية، الذي يتشكل بدوره من التقاء رافدي الصوراني وأبو الحمام، أما الرافد الثاني: فهو الوادي الأخضر.

يبلغ طول نهر الصوراني 8 Km، تصب مياه المفيض في نهر مرقية الذي بدوره يصب في البحر الأبيض المتوسط شمال مدينة طرطوس بمسافة 15 Km، لقد بلغ متوسط الجريان السنوي الوارد إلى سد الصوراني في عام 2017 ($25.72 \text{ million m}^3$).

ولتوصيف الموقع مناخياً تم الاعتماد على المعطيات المناخية لمحطة الصوراني. تتنوع أشكال الهطول المطري في منطقة الدراسة بين هطولات مطرية وثلجية وأحياناً تهطل زخات من البرد. وقد بلغ المتوسط السنوي للهطل المطري 1000 mm خلال الفترة (2008–2020).

يتأرجح متوسط درجات الحرارة الشهرية تبعاً لفصول وأشهر السنة، حيث يتراوح ما بين 12.5 درجة مئوية في كانون الثاني و 33.5 درجة مئوية في شهر تموز. في الوقت ذاته تتأثر مياه البحيرة بدرجة حرارة الهواء المحيط وقد سُجلت أعلى درجة حرارة (33 C°) صيفاً و (12.5 C°) شتاءً.



الشكل رقم (1): موقع بحيرة سد الصوراني - توزع نقاط مصادر التلوث ومواقع أخذ العينات.

3.2- جمع العينات وتحليلها Samples collection and analysis

تتعدد مصادر التلوث التي تؤثر على مياه بحيرة السد، أبرزها مياه الصرف الصحي القادمة من التجمعات السكنية لقرية برمانة المشايخ، ومخلفات الأنشطة الزراعية (الأسمدة - المبيدات الزراعية)، وكذلك مخلفات بعض المنشآت الصناعية (معاصر الزيتون - المقالع)، والمنشآت السياحية المحلية المجاورة لبحيرة السد. لذلك تأتي أهمية التحقق من مساهمة هذه المصادر في تلوث مياه البحيرة. تم اختيار شبكة اعتيان لقطف العينات، حيث حُددت مواقع أخذ العينات المائية من سطح بحيرة السد المدروسة في سبعة نقاط مختلفة على طول البحيرة (S1 قرب جسم السد، S2 على بعد 200 m من جسم السد، S3 رافد عين زعرور، S4 وسط البحيرة، S5 الوارد إلى البحيرة، S6 مدخل البحيرة، S7 وسط جنوب البحيرة) وفق الشكل (1)، وإحداثياتها مبينة في الجدول (1).

الجدول رقم (1): موقع وإحداثيات نقاط الاعتيان المعتمدة على البحيرة المدروسة

نقطة الاعتيان	Site الموقع	x (longitude)	y (latitude)
S1	قرب جسم السد	36.11377	35.02749
S2	على بعد 200 m من جسم السد	36.11544	35.02875
S3	رافد عين زعرور	36.11791	35.02861
S4	وسط البحيرة	36.12086	35.02775
S5	الوارد إلى البحيرة	36.13683	35.02497
S6	مدخل البحيرة	36.12727	35.02772
S7	رافد الوادي الأخضر	36.11875	35.02553

تم أخذ العينات من أجل إجراء التحاليل الفيزيائية والكيميائية والجرثومية بمعدل مرة واحدة في الشهر ولفترة امتدت لثلاث دورات هيدرولوجية ابتداءً من شهر كانون الثاني 2018 ولغاية شهر كانون الأول 2020، حيث تم قطف العينات بالتعاون مع الكادر الفني لمديرية الموارد المائية في محافظة طرطوس.

جُمعت العينات في عبوات بلاستيكية نظيفة ومغسولة، أما عينات التحليل الجرثومي، فتم جمعها في عبوات زجاجية نظيفة ومغسولة بالماء المقطر ومعقمة بالحرارة في فرن درجة حرارته (160 C°) لمدة ساعتين ونصف، وقد أخذت العينات على عمق (10-15 cm) تحت سطح الماء، تم اتباع كافة التعليمات اللازمة للقيام بقطف العينات. هناك بعض البارامترات التي تم قياسها في الحقل وهي: الأوكسجين المنحل (Dissolved Oxygen (DO)، وتغيرات درجة الحرارة Temperature Changes (TC).

تضمن العمل المخبري إجراء عدد من التحاليل على العينات المأخوذة، وشملت تحديد قيم العديد من مؤشرات نوعية المياه والتي تمثلت بما يلي: درجة الحرارة (Temperature)، العكارة (Turbidity)، الرقم الهيدروجيني (pH-Value)، الناقلية الكهربائية (EC Electrical Conductivity)، الأملاح المنحلة (TDS Total Dissolved Solid)، الأوكسجين المنحل (Dissolved Oxygen DO)، تركيز المواد الصلبة المعلقة الكلية (TSS Total Suspended Solid)، الطلب الحيوي للأوكسجين (BOD₅ Biochemical Oxygen Demand)، الطلب الكيميائي للأوكسجين (Chemical Oxygen

(Demand COD)، القساوة الكلية (TH Total Hardness)، القلوية الكلية (TA Total Alkalinity)، شاردة النترات (NO_3^-)، شاردة النتريت (NO_2^-)، شاردة الأمونيا (NH_4^+)، شاردة الفوسفات (PO_4^{3-})، شاردة الكبريتات (SO_4^{2-})، شاردة الكالسيوم (Ca^{+2})، شاردة المغنيزيوم (Mg^{+2})، شاردة الصوديوم (Na^+)، شاردة البوتاسيوم (K^+)، التحاليل الجرثومية (تعداد العصيات الجرثومية الغائطية) (FC Fecal Coliform). يوضح الجدول (2) بعض المؤشرات الإحصائية الوصفية للبيانات المقاسة.

الجدول رقم (2): المؤشرات الإحصائية الوصفية

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Temperature (C°)	252	12,30	35,20	22,14	6,42
Turbidity (NTU)	252	1,04	15,50	4,30	2,44
pH	252	7,10	9,22	8,30	0,40
TDS (mg/l)	252	150,00	375,50	254,75	42,20
DO (mg/l)	252	6,77	12,90	9,43	1,50
NO_3^- (mg/l)	252	0,00	20,00	3,68	3,90
PO_4^{3-} (mg/l)	252	0,00	0,45	0,04	0,05
BOD_5 (mg/l)	252	0,50	28,30	6,36	4,10
FC (CFU/100 mL)	252	4,00	320,00	68,16	53,66
TSS (mg/l)	252	0,20	25,30	5,87	3,54
COD (mg/l)	252	0,70	51,20	9,16	5,84
Ca (mg/l)	252	12,00	100,00	44,18	15,15
Mg (mg/l)	252	2,40	78,00	25,69	11,92
NO_2^- (mg/l)	252	0,00	0,35	0,04	0,05
SO_4^{2-} (mg/l)	252	7,90	329,50	31,50	19,86
K (mg/l)	252	0,40	8,80	2,61	1,64
Na (mg/l)	252	2,20	40,80	11,07	5,59
NH_4^+ (mg/l)	252	0,00	0,30	0,04	0,03
EC ($\mu\text{s}/\text{cm}$)	252	240,00	592,00	407,17	65,89
Total_Hardness (mg/l)	252	120,00	368,00	210,14	45,32
General_Alkaline (mg/l)	252	100,00	320,00	182,92	39,34

4- النتائج والمناقشة Results and discussion

4.1- مؤشر التلوث الشامل (CPI) Comprehensive Pollution Index

يساعد مفهوم مؤشر جودة المياه في تقييم خصائص المصدر المائي الفيزيائية والكيميائية والجرثومية، حيث يعرف المؤشر على أنه طريقة رياضية لحساب قيمة واحدة من نتائج اختبارات متعددة إذ تمثل نتيجة هذا المؤشر مستوى جودة المياه في المصدر المدروس، يعد هذا المؤشر هاماً لمراقبي جودة المياه على فترات من الزمن من أجل رصد التغيرات في النظام المائي والبيئي. يمكن إجراء اختبار جودة المياه والتحقق من ملاءمتها لأغراض الشرب من خلال حساب مؤشر الجودة الخاص بها. يعطي مؤشر الجودة دلائل عن الحالة الصحية للمياه في نقاط رصد مختلفة، ويمكن استخدامه لتحليل وتعقب التغيرات التي تطرأ على جودة المصدر المائي مع مرور الزمن.

بناءً على تقييم مؤشر العامل الفردي وبالنظر إلى التأثير المشترك لجميع العوامل من خلال عملية تقييمها، تم حساب مؤشر التلوث الشامل (CPI) Comprehensive Pollution Index باستخدام نماذج رياضية مختلفة وتحديد درجات التلوث بالطريقة الملائمة (المعادلات 1 و 2) (Vasanthi and Kumar 2019)، (Jindal and Wats 2022)، (Pramanik, Majumdar et al. 2020).

$$(Eq.1) \quad CPI = \frac{1}{n} \sum_{i=1}^n PI$$

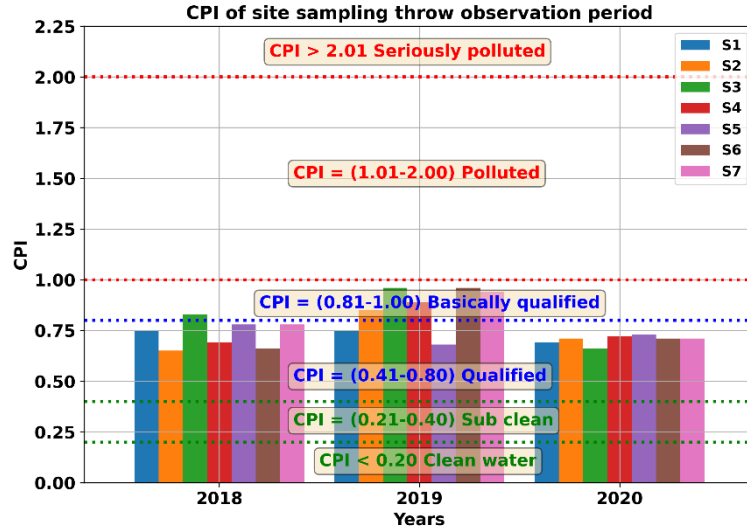
$$(Eq.2) \quad PI = \frac{\text{measured value of individual parameter}}{\text{standard permissible value of parameter}}$$

قمنا بحساب مؤشر التلوث الشامل لكل عينة، وتم تصنيفها باستخدام الحدود الموضحة في الجدول (3).

الجدول رقم (3): حدود مؤشر التلوث الشامل لنوعية المياه

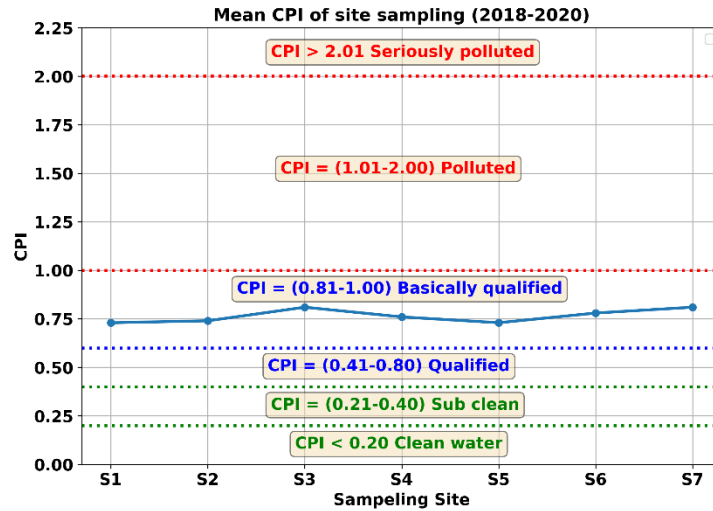
مؤشر التلوث الشامل (CPI)	التصنيف (الحالة)	جودة المياه واستخداماتها
0.00 – 0.20	نظيفة (Clean)	جيدة جداً وتستخدم للشرب والري والأغراض الصناعية.
0.21 – 0.40	نظيفة جزئياً (Sub clean)	جيدة واستخداماتها للأغراض المنزلية والري والصناعية.
0.41 – 0.80	مقبولة (Qualified)	تم الكشف عن بعض الملوثات ولكن تراكيزها تتوافق مع المعايير، أي الجودة مقبولة وتستخدم لأغراض الري والصناعة.
0.81 – 1.00	مقبولة بشكل أساسي (Basically qualified)	تجاوزت تراكيز بعض الملوثات المعايير، أي الجودة رديئة واستخداماتها فقط لأغراض للري.
1.01 – 2.00	ملوثة (Polluted)	تتجاوز تراكيز جزء كبير من الملوثات المعايير، أي نوعية رديئة للغاية (ملوثة) والاستخدام المقيد للري.
> 2.01	ملوثة بشكل خطير (Seriously polluted)	تتجاوز تراكيز جزء كبير من الملوثات المعايير عدة مرات، أي جودة المياه ملوثة للغاية ومعالجة مناسبة مطلوبة قبل الاستخدام.

يبين الشكل (2) نتائج حساب مؤشر التلوث الشامل CPI على كافة مواقع أخذ العينات للفترة الزمنية (2018-2020) بشكل سنوي، حيث تدل هذه النتائج على أن جودة مياه البحيرة مقبولة في العامين (2018,2020) إلى مقبولة بشكل أساسي في العام (2019).



الشكل رقم (2): مؤشر التلوث الشامل CPI في كافة مواقع الاعتيان .

يوضح الشكل (3) نتائج حساب مؤشر التلوث الشامل CPI في كافة مواقع أخذ العينات بشكل وسطي للفترة الزمنية المدروسة الكلية (2018-2020)، تدل هذه النتائج على أن نوعية مياه البحيرة مقبولة إلى مقبولة بشكل أساسي، حيث نجد قيم مؤشر التلوث الأعلى في الموقعين (S3, S7).



الشكل رقم (3): متوسط مؤشر التلوث الشامل CPI في البحيرة خلال فترة الدراسة (2018-2020)

4.2- التقنيات الإحصائية Statistical techniques

لقد استخدمت التقنيات الإحصائية على نطاق واسع لعدة عقود نظراً لدقة نتائجها، وهي أدوات شائعة الاستخدام تساعد في التقليل من التحيز في اختيار البارامترات، تعدّ هذه الأساليب الأدوات الأكثر فعالية لفهم التباين بين العديد من المتغيرات وتحويلها إلى مجموعات أصغر من المتغيرات المستقلة من خلال التعرف على أنماطها (Ghesquière, Walter et al. 2020, Busico, Kazakis et al. 2015). ترتبط هذه المنهجية بالصفات الإحصائية، مثل القيم المتوسطة وتراكيز البارامترات المختلفة المحسوبة على مدى فترة طويلة. تم تطبيق هذه التقنيات أيضاً بشكل فعال لتقييم جودة المياه من قبل العديد من مطوري مؤشر جودة المياه. لذلك، تُستخدم الطرق السائدة في التقنيات متعددة المتغيرات لاختيار وتجميع بارامترات التلوث المختلفة (Hamil, Arab et al. 2018)، مثل التحليل العنقودي Cluster Analysis، التحليل التمييزي

Chen, Zhang et) Multiple Linear Regression المتعدد التحليل الانحدار الخطي المتعدد Discriminant Analysis، تحليل الانحدار الخطي المتعدد Discriminant Analysis (al. 2018).

4.2.1- التحليل التمييزي Discriminant Analysis

يستخدم التحليل التمييزي من أجل تصنيف الحالات في مجموعات وذلك بناء على أوزان أو نسب أو درجات يحصلون عليها في توليفة من المتغيرات التي تنتبأ بتحديد عضويتهم في مجموعتين كأن يتم تصنيفهم إلى متغيرات يتوقع أن تكون فيزيائية أو كيميائية على سبيل المثال، وقد يتم تصنيف الحالات إلى أكثر من مجموعتين كأن يتم تصنيفها مثلاً إلى ثلاث مجموعات: فيزيائية وكيميائية وبيولوجية.

بشكل عام يمكن القول بأن التحليل التمييزي يعمل على إيجاد دالة للتمييز Discriminant Function وذلك من خلال حساب قيم لمتغيرات كمية تستخدم في التنبؤ، حيث تقوم دالة التمييز بالتنبؤ برقم المجموعة التي ينتمي إليها كل متغير (Jindal and Wats 2022). فالدالة التمييزية هي عبارة عن توليفة من المتغيرات المستقلة التي يمكن استخدامها في عملية التنبؤ بانتماء الحالات إلى إحدى مجموعتين أو أكثر. وبالتالي تتم عملية التصنيف على أساس دالة تمييزية واحدة عندما يكون هناك مجموعتين. أما في حالة وجود ثلاث مجموعات أو أكثر، فإن عدد الدوال التمييزية يبني على القاعدة التالية: (عدد المجموعات - 1) أو عدد المتغيرات الكلية أيهما أقل (Pramanik, Majumdar et al. 2020). يوجد ثلاث أنواع من التحليل التمييزي تبعاً لطريقة إدخال المتغيرات تتمثل في التحليل التمييزي المباشر (Direct Discriminate Analysis)، التحليل التمييزي الهرمي (Hierarchical Discriminate Analysis)، التحليل التمييزي المتدرج (Stepwise Discriminate Analysis). يتم الحكم على جودة النموذج التمييزي من خلال قيمة معامل الارتباط القانوني (Canonical Correlation) حيث تدل القيم المرتفعة لمعامل الارتباط القانوني على مؤشر لجودة التوفيق العالي للنموذج التمييزي، وبترتيب قيمة معامل الارتباط القانوني يتم الحصول على قيمة معامل التحديد (R^2) الذي يحدد نسبة مساهمة المتغيرات المستقلة في التمييز والتصنيف.

تم تطبيق التحليل التمييزي بعد التحقق من شروط استخدامه على مجموعة البيانات لإنشاء الدوال التمييزية ضمن بيئة عمل (SPSS)، وذلك لتقييم التغيرات الزمانية والمكانية في جودة مياه البحيرة، حيث تم اعتبار فصول العام ومحطات الاعتيان (المراصد) كمتغيرات تابعة، وكافة المتغيرات المقاسة كمتغيرات مستقلة، كما يلي:

- دراسة تأثير التغيرات الزمنية على بارامترات جودة المياه:

قمنا بتقسيم مجموعة البيانات إلى أربع مجموعات تبعاً للفصول الأربعة (الشتاء، الربيع، الصيف والخريف) وذلك من أجل تقييم التغيرات الزمنية في بارامترات جودة المياه، حيث تم اعتبار البارامترات كمتغيرات مستقلة واعتبار الفصول كمتغيرات تابعة.

من أجل دراسة التأثير الفصلي تم تطبيق طريقة التمايز المتدرج (Use Stepwise Method) والتي يتم فيها تجميع البيانات بدءاً من البيانات الأكثر تأثيراً وحتى الأقل أهمية في التمايز بين المجموعات الفصلية.

يوضح الجدول (4) التمايز الزماني والتأثير الفصلي من خلال قيم المعنوية Sig وقيم ويلكس لامبدا Wilks' Lambda وقيم (F) درجات الحرية كما يلي:

الجدول رقم (4): قيم المعنوية Sig لتأثير التغير الفصلي على المتغيرات

المتغير	Wilks' Lambda	F	Sig.	المتغير	Wilks' Lambda	F	Sig.
Temperature	0,165	418,876	0,000	Ca	0,704	34,745	0,000
Turbidity	0,910	8,137	0,000	Mg	0,980	1,726	0,162
pH	0,950	4,341	0,005	NO ₂	0,921	7,102	0,000
TDS	0,775	23,980	0,000	SO ₄	0,969	2,632	0,051
DO	0,568	62,989	0,000	K	0,830	16,981	0,000
NO ₃	0,727	31,044	0,000	Na	0,947	4,609	0,004
PO ₄	0,900	9,171	0,000	NH ₄ ⁺	0,957	3,714	0,012
BOD ₅	0,865	12,917	0,000	EC	0,809	19,542	0,000
FC	0,801	20,510	0,000	Total_Hardness	0,902	8,993	0,000
TSS	0,875	11,778	0,000	General_Alkaline	0,810	19,449	0,000
COD	0,856	13,875	0,000				

يبين الجدول السابق من خلال قيم المعنوية Sig أن كافة المتغيرات لها دلالة معنوية وتتأثر تراكيزها بالتغير الفصلي (P-value < 0.05) ما عدا متغيري (المغنيزيوم والكبريتات) التي ليس لديها ارتباط معنوي وهذا يعني أنه لا يوجد تأثير للتغير الفصلي على تراكيزها وبالتالي فهي مرتبطة بمياه الصرف الصناعي والري.

يعطي الجدول (5) ثلاثة دوال تمييزية تفسر مجموعات الارتباط. تعكس قيمة الارتباط العالية قوة الارتباط بين المتغيرات الداخلة في التحليل. يشير اختبار Wilks Lambda إلى مدى أهمية الدالة التمييزية في التمييز بين المجموعات، ولتقييم جودة التمييز للدالة نلاحظ أنه بلغت قيمة Wilks Lambda 0.054 للدالة الأولى و0.342 للدالة الثانية و0.779 للدالة الثالثة وبمستوى دلالة 0.00 للدوال التمييزية الثلاثة، بالتالي يمكن الاستنتاج أن هناك فروق ذات دلالة إحصائية بين المجموعات تعود إلى المتغيرات المستقلة، وهذا يشير إلى القدرة الجيدة للدالة على التمييز بين المجموعات وذلك يدل على أن الاختلاف في التصنيف بين المجموعات هو اختلاف جوهري ولا يعود للصدفة (Kappa).

الجدول رقم (5): الدوال التمييزية للمتغيرات المستقلة بدلالة الفصول السنوية

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation	Wilks' Lambda	Chi-square	Sig.
الدالة التمييزية	العامل الكامن (الذاتي)	التباين المفسر %	التباين التراكمي %	معامل الارتباط التجميعي			
1	5,380	77,5	77,50	0,918	0,054	715,57	0,00
2	1,280	18,4	95,90	0,749	0,342	262,46	0,00
3	0,283	4,1	100,00	0,470	0,779	60,96	0,00

يبين الجدول (5) أعلاه القيم الذاتية للدوال التمييزية حيث أن العامل الذاتي للدالة الأولى 5.38 ونسبة تفسيرها 77.5%، والعامل الذاتي للدالة الثانية 1.28 ونسبة تفسيرها 18.4%، والعامل الذاتي للدالة الثالثة 0.283 ونسبة تفسيرها 4.1%، وجميعها ذات دلالة معنوية Sig=0.00، وبلغت نسبة تفسير التباين الكلي للدوال التمييزية الثلاثة 100% للعلاقة بين الفصول السنوية والمتغيرات. أما فيما يتعلق بمعامل الارتباط فقد بلغ 91,8% بالنسبة للدالة التمييزية الأولى، 74,9% للدالة التمييزية الثانية و47% للدالة التمييزية الثالثة، مما يدل على جودة توافق الدوال. يبين الجدول (6) المتغيرات الأساسية التمييزية التي لها أثر واضح من خلال التغيرات الفصلية.

الجدول رقم (6): المعاملات المعيارية للدالة التمييزية

المتغير	الفصل Season			
	Winter	Spring	Summer	Autumn
Temperature	3,774	4,724	6,210	4,829
DO	9,404	10,716	9,129	8,609
NO ₃	1,021	1,657	1,223	0,968
PO ₄	28,351	15,189	34,309	19,121
FC	0,089	0,060	0,076	0,073
TSS	0,023	-0,016	0,204	-0,067
COD	0,289	0,504	0,492	0,419
K	2,486	1,184	1,741	2,182
General_Alkaline	0,246	0,211	0,228	0,221

يوضح الجدول (6) أعلاه معاملات كل دالة تمييزية معيارية بالنسبة للفصول والتي تحدد الأهمية النسبية لكل متغير من المتغيرات المستقلة، حيث نلاحظ أن درجة الحرارة والفسفور (Temperature, PO₄) هي أكثر المتغيرات أهمية في فصل الصيف، بينما كان البوتاسيوم (K) المتغير الأهم بالنسبة لفصلي الشتاء والخريف، في حين يمثل الأوكسجين المنحل (DO) المتغير الأكثر أهمية في فصل الربيع من بقية المتغيرات. تظهر الدوال التمييزية وجود 9 متغيرات تمييزية أساسية لها أثر واضح في التغيرات الفصلية، كما هي مبينة في الجدول السابق. بالنتيجة تعطي المتغيرات المستقلة الموضحة نسبة تفسير تباين عالية بالنسبة لعدد المتغيرات الكلية وذلك من خلال تطبيق التحليل التمييزي، حيث تم تصنيف (91.3%) من الحالات المجموعة الأصلية بشكل صحيح.

- دراسة تأثير التغيرات المكانية على بارامترات جودة المياه:

قمنا أيضاً بتطبيق التحليل التمييزي وفق نتائج التحليل العنقودي الذي اعطى عنقودين، يضم العنقود الأول (Cluster 1) مجموعتين مكونه من خمس محطات رصد من الجهة الشمالية للبحيرة وهي المجموعة الأولى (S1) قرب جسم السد، S2 على بعد 200 m من جسم السد، S3 رافد عين زعرور) والمجموعة الثانية (S4 وسط البحيرة، S6 مدخل البحيرة) وهذا العنقود مقبول بشكل أساسي وفق مؤشر التلوث الشامل، حيث ينجم تلوثه عن وصول مياه الصرف الصحي لتجمعات القرى وعن المنشآت الصناعية والسياحية المجاورة للبحيرة. بينما شمل العنقود الثاني (Cluster 2) على مجموعة واحدة مكونة من محطتي الرصد (S5 الوارد إلى البحيرة، S7 رافد الوادي الأخضر) ويمكن تصنيف هذا العنقود بأنه مقبول وأقل تلوثاً من العنقود الأول، حيث ينجم تلوثه عن وصول مياه الصرف الصحي عن بعض القرى المجاورة للبحيرة. تم اعتبار المتغيرات المدروسة كمتغيرات مستقلة والعناقيد كمتغيرات تابعة، ومن أجل دراسة التأثير المكاني قمنا بتطبيق طريقة التمايز المتدرج حيث تم تجميع البيانات بدءاً من البيانات الأكثر تأثيراً وحتى الأقل أهمية. يوضح الجدول (7) التمايز المكاني والتأثير لتغير الموقع من خلال قيم المعنوية Sig وقيم ويلكس لامبدا (Wilks lambda) وقيم (F) درجات الحرية كما يلي:

الجدول رقم (7): قيم المعنوية Sig لتأثير تغير الموقع على المتغيرات

المتغير	Wilks' Lambda	F	Sig.	المتغير	Wilks' Lambda	F	Sig.
Temperature	0,998	0,378	0,539	Ca	0,968	8,340	0,004
Turbidity	0,998	0,470	0,494	Mg	0,851	43,828	0,000
pH	0,996	0,983	0,322	NO ₂	0,997	0,733	0,393
TDS	0,798	63,419	0,000	SO ₄	1,000	0,024	0,877
DO	0,972	7,256	0,008	K	1,000	0,025	0,875
NO ₃	0,999	0,212	0,646	Na	0,998	0,472	0,493
PO ₄	0,958	10,862	0,001	NH ₄ ⁺	0,993	1,727	0,190
BOD ₅	0,994	1,439	0,231	EC	0,807	59,886	0,000
FC	0,975	6,529	0,011	Total_Hardness	0,846	45,459	0,000
TSS	0,988	3,128	0,078	General_Alkaline	0,882	33,552	0,000
COD	0,999	0,181	0,671				

يبين الجدول (7) من خلال قيم المعنوية Sig أنه يوجد عدد من المتغيرات تملك دلالة معنوية وهي تتأثر بالتغير المكاني وتعطي قيمة ذات ارتباط معنوي $P\text{-value} < 0.05$ وهي: (TDS, DO, PO₄, FC, TSS, Ca, Mg, EC, Total Hardness, General Alkaline)، أما بقية المتغيرات لها قيمة $P\text{-value} > 0.05$ مما يدل على عدم وجود ارتباط معنوي بالتالي لا يوجد تأثير للتغير المكاني على تراكيزها. نجد من خلال الجدول (8) أنه لدينا دالة تمييزية واحدة والتي تفسر مجموعات الارتباط كما يلي:

الجدول رقم (8): دالة التمييز للمتغيرات المستقلة بدلالة تغير موقع الاعتيان

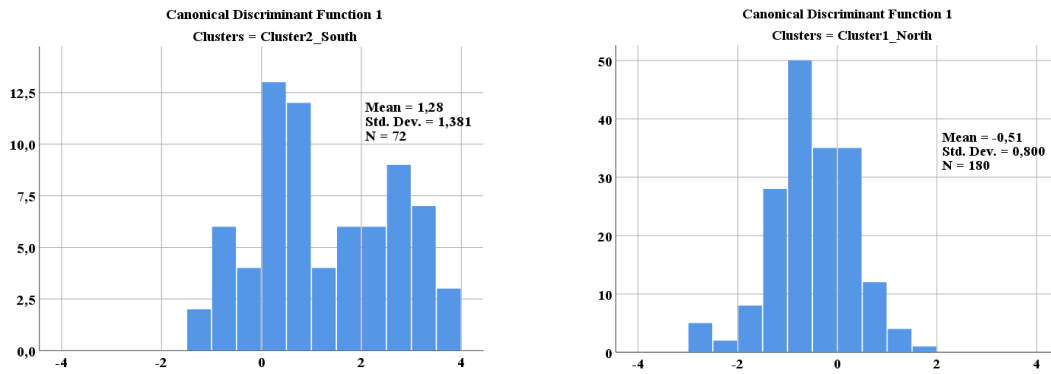
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation	Wilks' Lambda	Chi-square	Sig.
الدالة التمييزية	العامل الكامن (الذاتي)	التباين المفسر %	التباين التراكمي %	معامل الارتباط التجمعي			
1	0,660	100,0	100,0	0,630	0,603	124,889	0,00

يظهر الجدول (8) القيمة الذاتية للدالة التمييزية 0,66، ولها دلالة معنوية احصائياً $Sig=0.00$ ، ونسبة تفسيرها للتباين 100% في العلاقة بين المواقع والمتغيرات وهي نسبة ممتازة لتفسير بارامترات المياه. كما بلغ معامل الارتباط للدالة التمييزية 63% الذي يدل على جودة توافق الدالة. يبين الجدول (9) المتغيرات الأساسية التمييزية التي لها أثر واضح من خلال تغير مواقع الاعتيان.

الجدول رقم (9): المعاملات المعيارية للدالة التمييزية

Classification Function Coefficients		
تصنيف المعاملات وفق الموقع		
المتغيرات المستقلة	العناقيد Clusters	
	Cluster1_North العنقود الأول شمال البحيرة	Cluster2_South العنقود الثاني جنوب البحيرة
Temperature	2,745	2,951
TDS	0,295	0,333
DO	6,222	5,935
COD	0,437	0,552
Ca	0,394	0,426
Mg	0,141	0,196
General_Alkaline	0,117	0,135

بيّن تطبيق التحليل التمييزي على مجموعة البيانات وجود 7 متغيرات تنبؤية كما هو موضح في الجدول (9) وهي (Temperature, TDS, DO, COD, Ca, Mg, General Alkaline) باعتبارها أهم معايير التمييز بين مجموعتي العناقيد، ويوضح الجدول معاملات الدالة التمييزية المعيارية التي تحدد الأهمية النسبية لكل متغير من المتغيرات المستقلة، يشكل الاوكسجين المنحل أهم المتغيرات المستقلة في دالة التحليل في العنقود الأول (شمال البحيرة)، بينما تشكل درجة الحرارة المتغير الأكثر أهمية في دالة التحليل التمييزي بالنسبة للعنقود الثاني (جنوب البحيرة). بالنتيجة تعطي المتغيرات المستقلة الموضحة في الجدول السابق نسبة تفسير تباين كاملة بالنسبة لعدد المتغيرات الكلية وذلك من خلال تطبيق التحليل التمييزي للمتغيرات من أجل تقييم التغيرات المكانية لمواقع الاعتيان. حيث تم تصنيف (83.7 %) من الحالات المجمعّة الأصلية بشكل صحيح. يوضح الشكل (4) مخطط الدالة التمييزية وفق العنقودين المدروسين.



الشكل رقم (4): مخطط الدالة التمييزية وفق العنقودين الأول والثاني

– دراسة تأثير تغيرات بارامترات التلوث على درجة تصنيف جودة المياه:

وفقاً لنتائج قيم مؤشر التلوث الشامل (CPI) قمنا بتقسيم قاعدة البيانات إلى ست مجموعات تبعاً لتصنيف درجات التلوث (نظيفة، نظيفة جزئياً، مقبولة، مقبولة بشكل أساسي، ملوثة وملوثة بشكل خطير)، وذلك من أجل تقييم تغيرات بارامترات التلوث على جودة المياه باستخدام التحليل التمييزي، حيث تم اعتبار البارامترات كمتغيرات مستقلة واعتبار درجات مؤشر

التلوث الشامل كمتغيرات تابعة. مع ملاحظة أنه نتج لدينا وفق قاعدة بيانات التحليل المعتمدة في الدراسة للسنوات (2018-2020) ونتائج حساب مؤشر التلوث الشامل فقط ثلاث درجات تلوث وهي (مقبولة، مقبولة بشكل أساسي وملوثة). من أجل هذه الدراسة تم تطبيق طريقة التمايز المتدرج (Use Stepwise Method) والتي يتم فيها تجميع البيانات بدءاً من البيانات الأكثر تأثيراً وحتى الأقل أهمية في التمايز بين تصنيفات درجات التلوث. يوضح الجدول (10) تمايز قيم التلوث وتأثير بارامترات التلوث من خلال قيم المعنوية Sig وقيم ويلكس لامبدا Wilks' Lambda (λ) وقيم (F) درجات الحرية كما يلي:

الجدول رقم (10): قيم المعنوية Sig لتأثير تغير المتغيرات المستقلة في درجات التلوث

المتغير	Wilks' Lambda	F	Sig.	المتغير	Wilks' Lambda	F	Sig.
Temperature	0,990	1,289	0,277	Ca	0,974	3,281	0,039
Turbidity	0,839	23,877	0,000	Mg	0,982	2,307	0,102
pH	0,969	3,928	0,021	NO ₂	0,989	1,367	0,257
TDS	0,962	4,851	0,009	SO ₄	0,999	0,104	0,902
DO	0,885	16,182	0,000	K	0,980	2,569	0,079
NO ₃	0,997	0,402	0,669	Na	0,987	1,598	0,204
PO ₄	0,994	0,762	0,468	NH ₄ ⁺	0,994	0,719	0,488
BOD	0,414	176,544	0,000	EC	0,951	6,460	0,002
FC	0,993	0,835	0,435	Total_Hardness	0,973	3,434	0,034
TSS	0,749	41,825	0,000	General_Alkaline	0,964	4,585	0,011
COD	0,392	193,064	0,000				

يبين الجدول السابق من خلال قيم المعنوية Sig أنه لدينا إحدى عشر متغير لها دلالة معنوية وتؤثر تراكيزها بدرجات جودة المياه ($P\text{-value} < 0.05$) وهي (Turbidity, pH, TDS, DO, BOD, TSS, COD, Ca, EC, Total Hardness and General Alkaline) بينما بقية المتغيرات ليس لديها ارتباط معنوي.

الجدول رقم (11): الدوال التمييزية للمتغيرات المستقلة بدلالة تغير درجات تصنيف التلوث

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation	Wilks' Lambda	Chi-square	Sig.
الدالة التمييزية	العامل الكامن (الذاتي)	التباين المفسر %	التباين التراكمي %	معامل الارتباط التجمعي			
1	2,431	92,8	92,8	0,842	0,245	345,939	0,00
2	0,190	7,2	100,0	0,399	0,841	42,690	0,00

يظهر الجدول (11) القيمة الذاتية للدالتين التمييزيتين، الأولى 2.431 ونسبة تفسيرها للتباين 92.8% في العلاقة بين درجات تصنيف التلوث والمتغيرات المستقلة وهي نسبة عالية لتفسير بارامترات جودة المياه، والثانية 0.19 ونسبة تفسيرها للتباين 7.2%، ولهما دلالة معنوية احصائية Sig=0.00. كما بلغ معامل الارتباط للدالة التمييزية الأولى 84.2% والذي يدل على جودة توافق الدالة، بينما بلغ معامل الارتباط للدالة التمييزية الثانية 40%.

الجدول رقم (12): المعاملات المعيارية للدالة التمييزية

Classification Function Coefficients			
تصنيف المعاملات وفق درجة التلوث			
المتغيرات المستقلة	CPI_Class درجات التصنيف		
	مقبول Qualified	مقبول بشكل أساسي Basically qualified	ملوث Polluted
Turbidity	0,561	0,933	0,999
NO ₃	-0,586	-0,628	-0,229
BOD	0,919	1,407	2,078
TSS	0,431	0,669	0,743
COD	0,222	0,262	1,015
NO ₂	-10,813	-7,261	17,719
EC	0,124	0,134	0,113
(Constant)	-30,098	-41,214	-57,163

بين أيضاً تطبيق التحليل التمييزي على مجموعة البيانات وجود سبع متغيرات تنبؤية كما هو موضح في الجدول (12) وهي (Turbidity, NO₃, BOD, TSS, COD, NO₂, EC) باعتبارها أهم معايير التمييز بين مجموعات التلوث التصنيفية الثلاث، ويوضح الجدول معاملات الدالة التمييزية المعيارية التي تحدد الأهمية النسبية لكل متغير من المتغيرات المستقلة، تشكل العكارة، الطلب الحيوي للأوكسجين، المواد الكلية العالقة، الطلب الكيميائي للأوكسجين وثاني أكسيد النتروجين أهم المتغيرات المستقلة في دالة التحليل بالنسبة لفئة التلوث (Polluted)، بينما كان المتغير المستقل النترات أهم المتغيرات بالنسبة لفئة (Basically qualified). بالنتيجة تعطي المتغيرات المستقلة الموضحة في الجدول السابق نسبة تفسير تباين عالية بالنسبة لعدد المتغيرات الكلية وذلك من خلال تطبيق التحليل التمييزي، حيث كانت (94%) من الحالات المجمعة الأصلية مصنفة بشكل صحيح الجدول (13).

الجدول رقم (13): نتائج تصنيف المجموعات

نتائج التصنيف ^a Classification Results						
درجات التصنيف CPI_Class			Predicted Group Membership			Total
			مقبول Qualified	مقبول بشكل أساسي Basically qualified	ملوث Polluted	
Original	Count	Qualified	174	14	0	188
		Basically qualified	1	41	0	42
		Polluted	0	0	22	22
	%	Qualified	92,6	7,4	0,0	100,0
		Basically qualified	2,4	97,6	0,0	100,0
		Polluted	0,0	0,0	100,0	100,0

a. 94,0% of original grouped cases correctly classified.

4.2.2- الانحدار الخطي المتعدد Multiple Linear Regression

يعتبر الانحدار الخطي المتعدد من الأساليب الإحصائية المتقدمة والتي تضمن دقة الاستدلال من أجل تحسين نتائج البحث عن طريق الاستخدام الأمثل للبيانات التي تم جمعها في إيجاد علاقات سببية بين الظواهر موضوع البحث (تلوث مياه البحيرة).

الانحدار الخطي المتعدد هو عبارة عن إيجاد معادلة رياضية تعبر عن العلاقة بين متغيرين وتستعمل لتقدير قيم سابقة وللتنبؤ بقيم مستقبلية، وهو عبارة أيضاً عن انحدار للمتغير التابع (Y) (مؤشر تلوث البحيرة) على العديد من المتغيرات المستقلة المستمرة (بارامترات التلوث) X_1, X_2, \dots, X_n بالتالي فهو يستخدم في التنبؤ بتغيرات المتغير التابع الذي يؤثر فيه عدة متغيرات مستقلة أي تركز فكرته على العلاقات الدلالية التي تستخدم ما يعرف بشكل التشتت أو الانتشار، لذا يمكننا التنبؤ بالمستوى الرقمي لمؤشر تلوث البحيرة بالاعتماد على عدد من بارامترات التلوث (الفيزيائية، الكيميائية، الجرثومية والعضوية) (Abba, Hadi et al. 2017).

تقدم المعادلة (3) الشكل الرياضي لطريقة تحليل الانحدار الخطي المتعدد:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_iX_i \quad (\text{Eq.3})$$

حيث أن:

$Y =$ المتغير التابع.

$b_0 =$ قيمة ثابتة *Constant* أو *Intercept* ، $b_1 =$ ميل الانحدار y على المتغير المستقل الأول، $b_2 =$ ميل الانحدار

y على المتغير المستقل الثاني، $b_i =$ ميل الانحدار y على المتغير المستقل رقم i .

$X_1 =$ المتغير المستقل الأول، $X_2 =$ المتغير المستقل الثاني، $X_i =$ المتغير المستقل رقم i .

يمكن استخدام الانحدار الخطي المتعدد في حالة وجود علاقة خطية بين المتغيرات المستقلة والمتغير التابع، والبيانات موزعة توزيعاً طبيعياً، وقيم المتغير التابع من المستوى الترتيبي على الأقل.

قمنا أيضاً قبل الشروع بتطبيق الانحدار الخطي المتعدد ضمن بيئة عمل بايثون (Python) بالتأكد من عدم وجود مشكلة العلاقة الخطية المتعددة multicollinearity بين المتغيرات وهنا يجب أن نحسب قيم معاملات تضخم الفروق Variance Inflation Factors (VIF) معادلة (4)، حيث تعد العلاقة الخطية المتعددة مفهوماً إحصائياً ترتبط فيه العديد من المتغيرات المستقلة في النموذج. يتم اعتبار متغيرين مستقلين على علاقة خطية متداخلة تماماً إذا كان معامل الارتباط بينهما كبيراً $+/-$ 1.0 (أي أنهما مرتبطان ببعضهما البعض ارتباطاً وثيقاً في نموذج الانحدار). إن العلاقة الخطية المتعددة بين المتغيرات المستقلة ستفضي إلى استدلالات إحصائية أقل موثوقية.

$$VIF = (1/(1-R^2)) \quad (\text{Eq.4})$$

قمنا بعد حساب قيم VIF الجدول (14) بحل مشكلة العلاقة الخطية المتعددة من خلال إبعاد البارامترات التي تملك قيم عالية لمعامل تضخم الفروق بالتالي ($VIF > 3$)، مع الأخذ بعين الاعتبار مستوى المعنوية (Significant) لهذه البارامترات في نموذج الانحدار الخطي المتعدد.

الجدول رقم (14): نتائج قيم معاملات تضخم الفروق VIF

No.	variables	VIF	No.	Variables	VIF	No.	variables	VIF
1	Temperature	30.09	8	BOD ₅	13.42	15	SO ₄	3.86
2	Turbidity	7.37	9	FC	3.63	16	K	7.82
3	pH	244.73	10	TSS	7.08	17	Na	6.77
4	TDS	273.42	11	COD	13.30	18	Ammonia	2.92
5	DO	70.01	12	Ca	22.43	19	EC	317.83
6	NO ₃	4.99	13	Mg	15.47	20	Total_Hardness	84.20
7	PO ₄	2.39	14	NO ₂	2.22	21	Alkaline_General	42.80

بعد التحقق من توفر شروط الاختبار المذكورة أعلاه، قمنا بإجراء تحليل الانحدار الخطي المتعدد بطريقة Enter (يقوم البرنامج بإدخال جميع المتغيرات المستقلة) بعدة سيناريوهات للعوامل الداخلة في النماذج، واختبار جودة نموذج الانحدار باستخدام F-Test ثم حساب النسبة التي يفسرها كل متغير مستقل في التغيير الحاصل في جودة الأداء، ويتم اختيار النموذج الذي يحقق معامل الارتباط الأعلى. نوضح النتائج التي تم الحصول عليها في الجداول (15, 16, 17) للنماذج (A, B, C) على التوالي.

الجدول رقم (15): نتائج اختبار تحليل الانحدار الخطي المتعدد - نموذج (A).

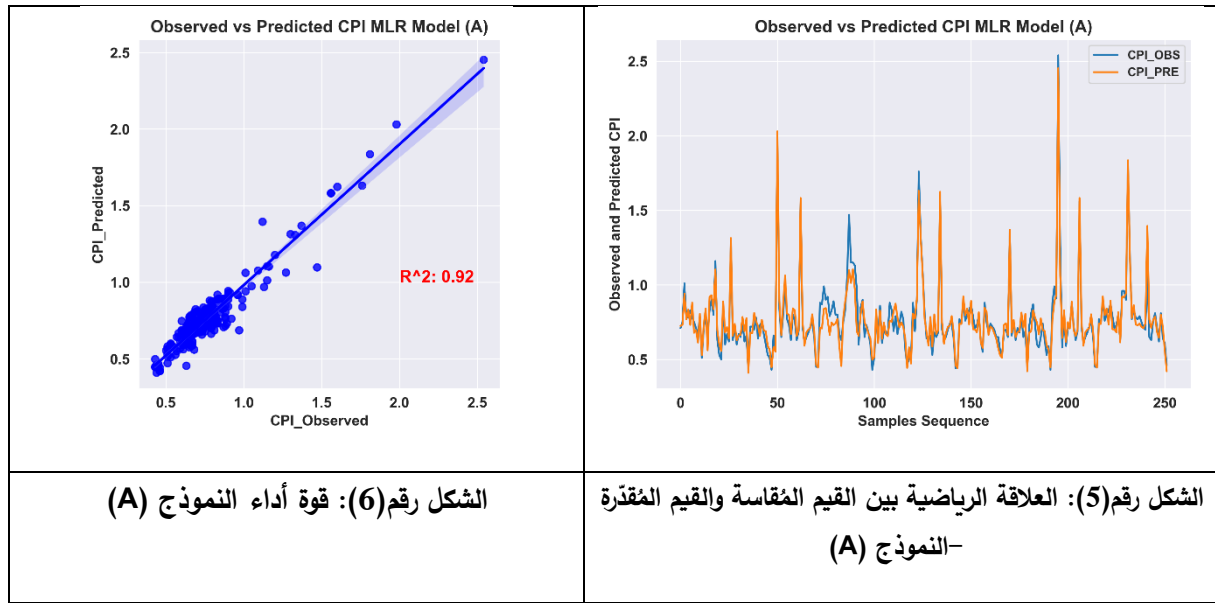
جودة النموذج		معامل تضخم التباين VIF	معامل التحديد R ²	معامل الارتباط R	قيمة Sig	قيمة t	قيمة B	النموذج (A)
قيمة Sig	قيمة F				0,000	6,699	0,201	الثابت (Constant)
0,00	717,456	1,019	0,921	0,96	0,001	3,304	0,004	النترات NO ₃
		1,119			0,000	48,669	0,039	الطلب الكيميائي للأوكسجين COD
		1,004			0,012	2,540	0,001	الكبريتات SO ₄
		1,120			0,000	5,814	0,018	الأوكسجين المنحل DO

نلاحظ من الجدول (15) وجود علاقة ارتباط بين المتغيرات المستقلة (بارامترات التلوث) مع المتغير التابع (مؤشر التلوث الشامل CPI) بنسبة 96% بدلالة (0.00) أصغر من مستوى المعنوية (0.05) وهذا يدل إحصائياً على وجود علاقة طردية تبين أنه كلما زاد المتغير المستقل درجة ازدياد درجة مؤشر التلوث الشامل والعكس صحيح. وكانت قيمة الاختبار F

(717.456) دالة احصائياً عند مستوى معنوية (0.00) أصغر من (0.05) وهي تدل على جودة نموذج العلاقة وصحة الاعتماد على نتائج النموذج بدون أخطاء. تشير قيمة R^2 (0.921) إلى أن عوامل التلوث المستقلة التفسيرية الداخلة في النموذج تفسر التغيرات الحاصلة في مؤشر التلوث الشامل للبحيرة بنسبة 92% وتبقى نسبة 8% تفسرها العوامل الأخرى بالإضافة إلى الأخطاء العشوائية. ولتحديد أي من المتغيرات المستقلة كانت سبباً لإضافة التفسير الجوهرى للتغير الحاصل في المتغير التابع مؤشر جودة التلوث الشامل (CPI)، يتضح لنا من معاملات الانحدار في النموذج أن المتغيرات الدالة معنوياً هي COD حيث جاءت قيمة t (48.669) بقيمة احتمالية (0.00) أصغر من (0.05)، و DO حيث جاءت قيمة t (5.814) بقيمة احتمالية (0.00) أصغر من (0.05)، و NO_3 حيث كانت قيمة t (3.304) بقيمة احتمالية (0.001) أصغر من (0.05)، وأخيراً SO_4 حيث جاءت قيمة t (2.54) بقيمة احتمالية (0.012) أصغر من (0.05)، كما يوضح الجدول السابق نتائج اختبار التعددية الخطية Multicollinearity من خلال ملاحظة معامل تضخم التباين للنموذج الذي جاءت قيمه أصغر من (3)، بالتالي تتكون معادلة الانحدار من قيم بيتا (B) لكل من المتغيرات المستقلة السابقة وقيمة الثابت Constant على النحو التالي:

$$\text{مؤشر التلوث الشامل CPI} = 0.201 + 0.004 (NO_3) + 0.039 (COD) + 0.001 (SO_4) + 0.018 (DO)$$

يبين الشكل (5) العلاقة الرياضية بين القيم المقاسة والقيم المقدرة من خلال تطبيق النموذج (A) على مصفوفة المعطيات. كما يبين الشكل (6) قوة أداء النموذج من خلال قيمة R معامل الارتباط.



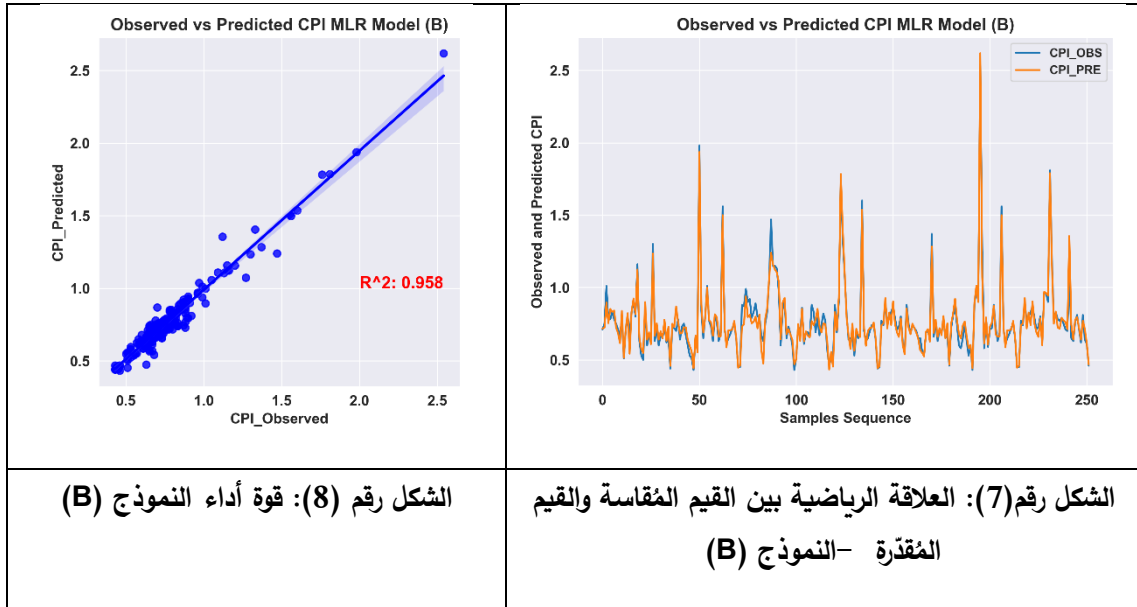
الجدول رقم (16): نتائج اختبار تحليل الانحدار الخطي المتعدد - نموذج (B).

جودة النموذج		معامل تضخم التباين VIF	معامل التحديد R^2	معامل الارتباط R	قيمة Sig	قيمة t	قيمة B	النموذج (B)
قيمة Sig	قيمة F							
0,00	1413,186	1,178	0,958	0,979	0,000	12,150	0,254	الثابت (Constant)
		1,211			0,000	60,820	0,037	الطلب الكيميائي للأوكسجين COD
		1,607			0,001	3,236	0,008	الأوكسجين المنحل DO
		1,458			0,000	12,190	0,014	المواد الصلبة العالقة الكلية TSS
					0,004	2,939	0,005	العكارة Turbidity

نلاحظ من الجدول (16) وجود علاقة ارتباط بين بارامترات التلوث مع مؤشر التلوث الشامل CPI بنسبة 98% بدلالة (0.00) أصغر من مستوى المعنوية (0.05) وهذا ذو دلالة احصائية. كانت قيمة الاختبار F (1413.186) دالة احصائياً عند مستوى معنوية (0.00) أصغر من (0.05) وهي تدل على جودة نموذج العلاقة وصحة الاعتماد على نتائج النموذج بدون أخطاء. تشير قيمة R^2 (0.958) إلى أن عوامل التلوث المستقلة التفسيرية الداخلة في النموذج تفسر التغيرات الحاصلة في مؤشر التلوث الشامل للبحيرة بنسبة 96% وتبقى نسبة 4% تفسرها العوامل الأخرى بالإضافة إلى الأخطاء العشوائية. تتكون معادلة الانحدار من قيم بيتا (B) لكل من المتغيرات المستقلة السابقة وقيمة الثابت Constant على النحو التالي:

$$\text{مؤشر التلوث الشامل CPI} = 0.254 + 0.037 (\text{COD}) + 0.008 (\text{DO}) + 0.014 (\text{TSS}) + 0.005 (\text{Turbidity})$$

يبين الشكل (7) العلاقة الرياضية بين القيم المقاسة والقيم المقدرة من خلال تطبيق النموذج (B) على مصفوفة المعطيات. كما يبين الشكل (8) قوة أداء النموذج من خلال قيمة R معامل الارتباط.



قمنا بتقليص عدد المتغيرات الداخلة في النموذج السابق إلى ثلاث متغيرات فقط وذلك بحذف المواد الصلبة العالقة الكلية TSS، الجدول التالي يلخص نتائج التحليل الانحدار الخطي المتعدد.

الجدول رقم (17): نتائج اختبار تحليل الانحدار الخطي المتعدد - نموذج (C).

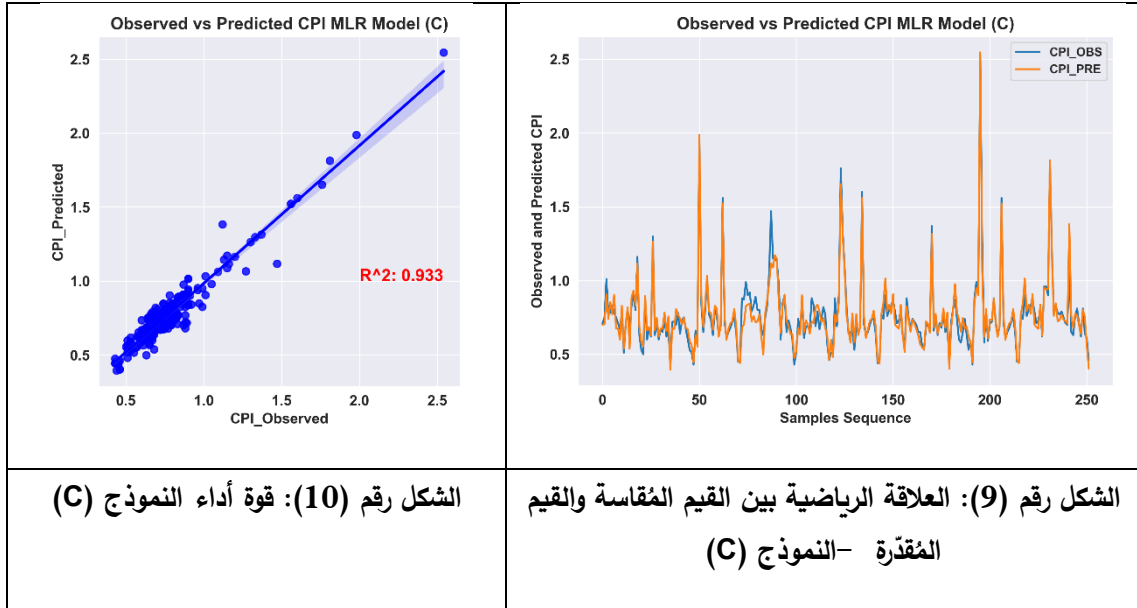
جودة النموذج		معامل تضخم VIF	معامل التحديد R^2	معامل الارتباط R	قيمة Sig	قيمة t	قيمة B	النموذج (C)
قيمة Sig	قيمة F				0,000	8,558	0,224	الثابت (Constant)
0,00	1150,157	1,137	0,933	0,966	0,000	50,856	0,038	الطلب الكيميائي للأوكسجين COD
		1,155			0,000	4,750	0,014	الأوكسجين المنحل DO
		1,109			0,000	8,083	0,014	العكارة Turbidity

نلاحظ من الجدول (17) وجود علاقة ارتباط بين بارامترات التلوث مع مؤشر التلوث الشامل CPI بنسبة 97% بدلالة (0.00) أصغر من مستوى المعنوية (0.05) وهذا ذو دلالة احصائية. كانت قيمة الاختبار F (1150.157) دالة احصائياً

عند مستوى معنوية (0.00) أصغر من (0.05) وهي تدل على جودة نموذج العلاقة وصحة الاعتماد على نتائج النموذج بدون أخطاء. تفسر العوامل التلوث المستقلة التفسيرية الثلاثة الداخلة في النموذج التغيرات الحاصلة في مؤشر التلوث الشامل للبحيرة بنسبة 93% وتبقى نسبة 7% تعزى إلى عوامل الأخرى بالإضافة إلى الأخطاء العشوائية. تتكون معادلة الانحدار من قيم بيتا (B) لكل من المتغيرات المستقلة السابقة وقيمة الثابت Constant على النحو التالي:

$$\text{مؤشر التلوث الشامل CPI} = 0.224 + 0.038 (\text{COD}) + 0.014 (\text{DO}) + 0.014 (\text{Turbidity})$$

يبين الشكل (9) العلاقة الرياضية بين القيم المقاسة والقيم المقدرة من خلال تطبيق النموذج (C) على مصفوفة المعطيات. كما يبين الشكل (10) قوة أداء النموذج من خلال قيمة R معامل الارتباط.



5- الاستنتاجات Conclusions

يتضح من خلال حساب مؤشر التلوث الشامل CPI في كافة مواقع الاعتيان بشكل وسطي لفترة الدراسة (2018-2020) وتقييم التغيرات (الزمانية والمكانية وبارامترات التلوث) على جودة مياه البحيرة بتطبيق التحليل التمييزي على مجموعة البيانات من أجل إنشاء الدوال التمييزية أن:

1- نوعية مياه البحيرة مقبولة إلى مقبولة بشكل أساسي، حيث كانت أعلى القيم لمؤشر التلوث في الموقعين المجاورين للتجمعات السكانية (S3, S7).

2- بالنسبة لتقييم التغيرات الزمانية، كان لكافة المتغيرات دلالة معنوية وتتأثر تراكمها بالتغير الفصلي ماعدا متغيرين (المغنيزيوم والكبريتات) مما يدل على ارتباطهما بمياه الصرف الصناعي والري. كان الأوكسجين المنحل أكثر المتغيرات أهمية في فصل الربيع.

3- فيما يتعلق بتقييم التغيرات المكانية، كانت المتغيرات المتأثرة وذات ارتباط معنوي هي: (TDS, DO, PO4, FC, TSS, Ca, Mg, EC, Total Hardness, General Alkaline).

4- بين التحليل التمييزي وجود 7 متغيرات تنبؤية وهي (Temperature, TDS, DO, COD, Ca, Mg, General Alkaline) باعتبارها أهم معايير التمييز بين مجموعتي عنقيد مواقع الرصد، شكّل الأوكسجين المنحل أهم المتغيرات المستقلة في دالة التحليل في العنقود الأول وفي فصل الربيع.

5- بالنسبة لتقييم بارامترات التلوث المؤثرة في تصنيف درجات جودة البحيرة وفق مؤشر التلوث الشامل (CPI) فقد تبين وجود سبع متغيرات تنبؤية هي (Temperature, DO, BOD, FC, EC, General Alkaline) باعتبارها أهم معايير التمييز بين مجموعات التلوث التصنيفية الثلاث، تشكل العكارة، الطلب الحيوي للأوكسجين، المواد الكلية العالقة، الطلب الكيميائي للأوكسجين وثنائي أوكسيد النتروجين أهم المتغيرات المستقلة في دالة التحليل بالنسبة لفئة التلوث (Polluted)، بينما كان المتغير المستقل النترات أهم المتغيرات بالنسبة لفئة (مقبول بشكل أساسي).

6- تم تطبيق ثلاثة نماذج انحدار خطي متعدد لبارامترات الجودة مع مؤشر التلوث الشامل (CPI). كان النموذج الثاني (B) أفضلها، حيث حقق علاقة ارتباط قوية، وفسرت عوامل التلوث المستقلة الداخلة في النموذج التغيرات الحاصلة في مؤشر التلوث الشامل للبحيرة بنسبة 98%.

6- التوصيات Recommendations

- 1- نوصي بأهمية حماية مصادر المياه السطحية والحفاظ عليها من التلوث، وذلك بمعالجة مياه الصرف الصحي الناتجة عن القرى المجاورة قبل تصريفها في البحيرة.
- 2- توعية المواطنين في ترشيد استهلاك المياه وعدم الإفراط في استخدام المواد والاسمدة الكيميائية في الزراعة.
- 3- دراسة تأثير العناصر الثقيلة على نوعية مياه البحيرة.
- 4- ضرورة متابعة برامج المراقبة والتحليل الدوري للعينات، وحفظ نتائج التحاليل المتعددة ضمن قواعد بيانات خاصة يمكن الاستفادة منها لاحقاً في الأبحاث المستقبلية، بحيث تساعد المعنيين وأصحاب القرار في تقييم جودة مصادر المياه ووضع الخطط المستقبلية لإدارة واستثمار هذه الموارد بالشكل الأمثل.

6- References

- 1-(CCME) (2001). Canadian Water Quality Index 1.0 Technical Report and User's Manual; Canadian Environmental Quality Guidelines Water Quality Index Technical Subcommittee. Gatineau, QC, Canada.
- 2-Abba, S., S. J. Hadi and J. Abdullahi (2017). "River water modelling prediction using multi-linear regression, artificial neural network, and adaptive neuro-fuzzy inference system techniques." Procedia computer science **120**: 75-82.
- 3-Arawo, C. (2020). "Multiple Linear Regression (MLR) Model: A Tool for Water Quality Interpretation." Momona Ethiopian Journal of Science **12(1)**: 123-134.
- 4-Busico, G., N. Kazakis, E. Cuoco, N. Colombani, D. Tedesco, K. Voudouris and M. Mastrocicco (2020). "A novel hybrid method of specific vulnerability to anthropogenic pollution using multivariate statistical and regression analyses." Water Research **171**: 115386.
- 5-Chen, T., H. Zhang, C. Sun, H. Li and Y. Gao (2018). "Multivariate statistical approaches to identify the major factors governing groundwater quality." Applied Water Science **8(7)**: 1-6.

- 6–Cude, C. G. (2001). "Oregon water quality index a tool for evaluating water quality management effectiveness 1." JAWRA Journal of the American Water Resources Association **37**(1): 125–137.
- 7–Ghesquière, O., J. Walter, R. Chesnaux and A. Rouleau (2015). "Scenarios of groundwater chemical evolution in a region of the Canadian Shield based on multivariate statistical analysis." Journal of Hydrology: Regional Studies **4**: 246–266.
- 8–Hamil, S., S. Arab, A. Chaffai, M. Baha and A. Arab (2018). "Assessment of surface water quality using multivariate statistical analysis techniques: a case study from Ghrib dam, Algeria." Arabian Journal of Geosciences **11**(23): 1–14.
- 9–Jindal, R. and M. Wats (2022). "Evaluation of surface water quality using water quality indices (WQIs) in Lake Sukhna, Chandigarh, India." Applied Water Science **12**(1): 1–14.
- 10–Liu, J., D. Zhang, Q. Tang, H. Xu, S. Huang, D. Shang and R. Liu (2021). "Water quality assessment and source identification of the Shuangji River (China) using multivariate statistical methods." PloS one **16**(1): e0245525.
- 11–Pramanik, A. K., D. Majumdar and A. Chatterjee (2020). "Factors affecting lean, wet–season water quality of Tilaiya reservoir in Koderma District, India during 2013–2017." Water Science **34**(1): 85–97.
- 12–Shuhaimi–Othman, M., E. C. Lim and I. Mushrifah (2007). "Water quality changes in Chini lake, Pahang, west Malaysia." Environmental monitoring and assessment **131**(1): 279–292.
- 13–Uddin, M. G., S. Nash and A. I. Olbert (2021). "A review of water quality index models and their use for assessing surface water quality." Ecological Indicators **122**: 107218.
- 14–Valentini, M., G. B. dos Santos and B. Muller Vieira (2021). "Multiple linear regression analysis (MLR) applied for modeling a new WQI equation for monitoring the water quality of Mirim Lagoon, in the state of Rio Grande do Sul—Brazil." SN Applied Sciences **3**(1): 1–11.
- 15–Vasanthi, S. S. and A. S. Kumar (2019). "Application of artificial neural network techniques for predicting the water quality index in the Parakai Lake, Tamil Nadu, India." Applied ecology and environmental research **17**(2): 1947–1958.
- 16–Zandbergen, P. A. and K. J. Hall (1998). "Analysis of the British Columbia water quality index for watershed managers: a case study of two small watersheds." Water Quality Research Journal **33**(4): 519–550.
- 17–Zhao, Y., X. Xia, Z. Yang and F. Wang (2012). "Assessment of water quality in Baiyangdian Lake using multivariate statistical techniques." Procedia Environmental Sciences **13**: 1213–1226.