

A Study to Evaluate the Effect of Some Water Quality Parameters on the Decay of Free Residual Chlorine in the Drinking Water Supply Network in Latakia City using Python Regression Models

Dr.Raed Jafar*
Dr.Hytham Jnad**
Dr.Adel Awad***
Mai Alfred Naser****

(Received 3 / 10 / 2022. Accepted 1 / 2 / 2023)

□ ABSTRACT □

This study aimed to evaluate the effect of some water quality parameters on the decay of free residual chlorine in the drinking water supply network in Latakia , Five sampling points were selected on the main network that feeds the Latakia city, starting from the treatment plant (Al-Sin) to Al-Montazah tank (Latakia) , These Samples were collected once a month , The study has lasted for seven months , from June (2021) to December (2021) , the measurement interval was determined and free residual chlorine concentration , temperature and pH were measured at the site while (turbidity - phosphates - nitrates - dissolved solids - electrical conductivity - sodium chloride) in the laboratory. The Python language was used to determine the correlation coefficient of the free residual chlorine with the studied parameters, which the correlation was inversely negative .Two models were built , Multiple linear regression and multivariate adaptive regression to determine the percentage of variance in free residual chlorine that is explained by the studied parameters and to evaluate the accuracy of the models.The variance was (86%) and this (86%) of the decrease in chlorine was related to the regression equation factors and that (14%) was due to other factors, the (R-squared) value for the first and second models, respectively, was ($R^2 = 0.81$, $R^2 = 0.86$) and (MSE = 0.008, MSE = 0.0046) thus based on the higher (R^2) value ($R^2 = 0.86$), the best and most suitable model for the data is multivariate adaptive regression .

Keywords: Chlorine decay - correlation coefficient - R - squared - Python language.

* Associate Professor, Department of Environmental Engineering, Faculty of Civil Engineering, Tishreen University, Lattakia, Syria.

** Assistant Professor, , Department of Environmental Engineering, Faculty of Civil Engineering, Tishreen University, Lattakia, Syria.

*** Professor, Department of Environmental Engineering, Faculty of Civil Engineering, Tishreen University, Lattakia, Syria.

**** Postgraduate Student (PhD), Department of Environmental Engineering, Faculty of Civil Engineering, Tishreen University, Lattakia, Syria.

دراسة لتقييم تأثير بعض بارامترات جودة المياه على تلاشي الكلور الحر المتبقي في شبكة الإمداد بمياه الشرب في مدينة اللاذقية باستخدام نماذج الانحدار بلغة (Python)

د. راند جعفر *

د. هيثم جناد **

د. عادل عوض ***

مي الفريد ناصر ****

(تاريخ الإيداع 3 / 10 / 2022. قُبِلَ للنشر في 1 / 2 / 2023)

□ ملخص □

هدفت هذه الدراسة إلى تقييم تأثير بعض بارامترات جودة المياه على تلاشي الكلور الحر المتبقي في شبكة الإمداد بمياه الشرب في مدينة اللاذقية ، تم اختيار خمس نقاط اعتيان على الشبكة الرئيسية التي تغذي مدينة اللاذقية بدءاً من محطة المعالجة (السن) حتى خزان المنتزه (اللاذقية) ، تم قطف العينات بمعدل مرة كل شهر ، استمرت الدراسة لمدة سبعة أشهر بدءاً من شهر حزيران (2021) حتى شهر كانون الأول (2021)، تم تحديد الفاصلة الزمنية للقياس وقياس تركيز الكلور الحر المتبقي ودرجة الحرارة والأس الهيدروجيني في الموقع وقياس كل من (العكارة -الفوسفات - النترات - المواد الصلبة المنحلة- الناقلية الكهربائية - كلوريد الصوديوم) في المخبر، تم استخدام لغة البايثون (Python) لتحديد معامل الارتباط للكلور الحر المتبقي مع البارامترات المدروسة حيث كانت علاقة الارتباط سلبية عكسية ، تم بناء نموذجي الانحدار الخطي المتعدد والانحدار التكيفي متعدد المتغيرات وذلك لتحديد نسبة التباين في الكلور الحر المتبقي الذي يتم تفسيره بواسطة البارامترات المدروسة وتقييم دقة النماذج ، بلغت نسبة التباين (86 %) وأن (86 %) من الانخفاض في الكلور مرتبط بعوامل معادلة الانحدار وأن (14 %) تعود إلى عوامل أخرى ، بلغت قيمة (R-squared) للنموذج الأول والثاني على التوالي ($R^2=0.86$, $R^2=0.81$) و ($MSE=0.008$) ($MSE=0.0046$) وبالتالي بناءً على قيمة (R^2) الأعلى ($R^2= 0.86$) فإن النموذج الأفضل والأكثر ملائمة للبيانات هو الانحدار التكيفي متعدد المتغيرات.

الكلمات المفتاحية : تلاشي الكلور -معامل الارتباط - R-squared - لغة البايثون (Python).

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

**مدرس- قسم الهندسة البيئية - كلية الهندسة المدنية - جامعة تشرين - اللاذقية - سورية.

***أستاذ- قسم الهندسة البيئية - كلية الهندسة المدنية - جامعة تشرين - اللاذقية - سورية.

**** طالبة دراسات عليا (دكتوراه) قسم الهندسة البيئية- كلية الهندسة المدنية - جامعة تشرين - اللاذقية- سورية.

مقدمة:

أصبحت إدارة جودة المياه مهمة للغاية ، خاصة مع انتشار الأمراض المنقولة عن طريق المياه [1]. وإن الهدف الرئيسي لمعالجة مياه الشرب هو توفير مياه آمنة من الناحية الميكروبيولوجية، وأهم استخدام للمطهرات في معالجة المياه هو الحد من الأمراض التي تنقلها المياه وتعطيل الكائنات المسببة للأمراض في إمدادات المياه[2]. لا بد من أي نظام لتوزيع المياه أن يوفر المياه للمستهلك بكميات وضغط مناسبين وبجودة مقبولة من حيث الطعم والرائحة والمظهر[3]. يستخدم الكلور في جميع أنحاء العالم كمعقم لمواجهة التلوث الميكروبي وانتشاره في أنظمة إمدادات مياه الشرب[4]. ويتم اختياره على نطاق واسع في أنظمة مياه الشرب بسبب تكلفته المنخفضة وطويلة الأمد في شبكة توزيع المياه[5]. كما تعتبر مراقبة جرعات الكلور والتحكم فيها أمراً مهماً لضمان فعالية المعالجة بالكلور لتجنب الوصول إلى خصائص غير مرغوب فيها داخل نظام التوزيع ، فإذا كانت جرعة الكلور منخفضة للغاية فلن يكون هناك كلور متبقي كافٍ في نظام توزيع المياه للحفاظ على التعقيم حتى يصل الماء إلى المستهلكين وإذا كانت الجرعة عالية جداً يشككي المستهلكون من الذوق والرائحة، بالإضافة إلى ذلك ، من المعروف أن الكلور الزائد يساعد على تكوين منتجات ثانوية للتطهير (DBPs) مثل ثلاثي الهالوميثان (THMs) وأحماض الهالوأستيك المهلجنة (HAAS) التي يمكن أن تشكل خطراً صحياً على البشر [6].

تؤثر العديد من العوامل على جودة المياه في نظام التوزيع مثل مادة الأنابيب والقطر ووجود المنتجات الثانوية للتآكل والأغشية الحيوية ودرجة حرارة الماء وبقايا المطهرات والظروف الهيدروليكية [7]. كما يخضع الكلور إلى العديد من التفاعلات الكيميائية ، ويتلاشى على طول نظام التوزيع، يمكن أن يتأثر تلاشي الكلور بعدة عوامل الخصائص الفيزيائية للشبكة و مكونات النظام ومواد الأنابيب والعمر و معايير جودة المياه و درجة الحرارة ، تركيز الكلور الأولي ، المادة العضوية ، محتوى الحديد وعدد مرات إعادة الكلورة و تشغيل النظام وصيانته ، السعة التخزينية للخزانات وفقدان الماء الحقيقي والتشغيل المتقطع و الظروف الهيدروليكية[8].

مع التقدم التكنولوجي ازداد استخدام لغات البرمجة لبناء نماذج الانحدار لتحليل البيانات ومن أهم لغات البرمجة المستخدمة هي Python وهي لغة برمجة مستخدمة على نطاق واسع في المجالات العلمية بسبب احتوائها على عدد كبير من المكتبات التي توفر مجموعة كاملة من أدوات التحليل ومعالجة البيانات ، وبالمقارنة مع لغات البرمجة الأخرى المستخدمة عموماً لتحليل البيانات ، مثل R و Matlab ، لا توفر Python منصة لمعالجة البيانات فحسب، بل تحتوي أيضاً على بعض الميزات التي تجعلها فريدة مقارنة باللغات الأخرى والتطبيقات المتخصصة[9]. وتعتبر Python واحدة من أشهر لغات البرمجة وأكثرها طلباً ، ظهر استطلاع حديث لـ Stack Overflow أن Python قد استحوذت على لغات مثل Java و C و ++ C وشق طريقها إلى القمة[10].

تستخدم نمذجة الانحدار على نطاق واسع لإنشاء علاقة بين بعض المتغيرات المستقلة ومتغير الاستجابة، وهناك أشكال مختلفة من نماذج الانحدار يستخدم كل منها لأغراض مختلفة مثل وصف البيانات في مجالات الهندسة والعلوم، ومن بين نماذج الانحدار هذه ، النموذج المعروف والشائع الاستخدام هو نموذج الانحدار الخطي المتعدد[11]. ونموذج خطوط الانحدار التكيفية متعددة المتغيرات (MARS) هي تقنية انحدار يمكنها نمذجة العلاقة بين متغير الاستجابة (وتسمى أيضاً المتغير المستهدف) ومتغيرات التفسيرية (التنبؤية) المتعددة وتتمثل القوة الرئيسية لـ MARS في قدرتها على اكتشاف وتعزيز قابلية تفسير التفاعلات المعقدة بين متغير الاستجابة ومجموعة من متغيرات التنبؤ. يتم تمثيل نموذج MARS من خلال وظائف خطية بسيطة[12].

غالباً ما يهدف الباحثون إلى دراسة ما إذا كان هناك ارتباط بين متغيرين وتقدير قوة هذه العلاقة، الارتباط بالمعنى الأوسع هو مقياس للارتباط بين المتغيرات. في البيانات المرتبطة، يرتبط التغيير في حجم متغير واحد بتغيير في حجم متغير آخر، إما في نفس الاتجاه (الارتباط الإيجابي) أو في الاتجاه المعاكس (الارتباط السلبي). يُنقَد أحياناً معامل الارتباط لأنه لا يحتوي على تفسير جوهري واضح، ويبلغ الباحثون أحياناً عن مربع معامل الارتباط. يسمى R^2 "معامل التحديد"، يمكن تفسيره على أنه نسبة التباين في متغير واحد يتم حسابه بواسطة الآخر [13].

أهمية البحث وأهدافه:

أهمية البحث

تتلخص أهمية البحث فيما يلي :

دراسة تلاشي الكلور في شبكة الإمداد بمياه الشرب في مدينة اللاذقية عن طريق قياس بعض بارامترات جودة المياه وتحديد نسبة التباين بينها وبين الكلور الحر المتبقي عن طريق بناء نماذج الانحدار والتوصل إلى النموذج الأفضل باستخدام لغة البايثون (Python).

أهداف البحث

يهدف البحث إلى :

- استخدام لغة البرمجة (Python) :

1- تحديد معامل الارتباط (R) (Correlation Coefficient) للكلور الحر المتبقي مع البارامترات المدروسة (الفاصلة الزمنية - درجة الحرارة - الأس الهيدروجيني - العكارة - الفوسفات - النترات - المواد الصلبة المنحلة - الناقلية الكهربائية - كلوريد الصوديوم).

2- بناء نماذج الانحدار (Building Regression Models) :

• نموذج الانحدار الخطي المتعدد.

(Multi Linear Regression Model)

• نموذج خطوط الانحدار التكيفية متعددة المتغيرات.

(Multivariate Adaptive Regression Splines Model)

3- تحديد جودة كل نموذج من النماذج المدروسة باستخدام معامل التحديد.

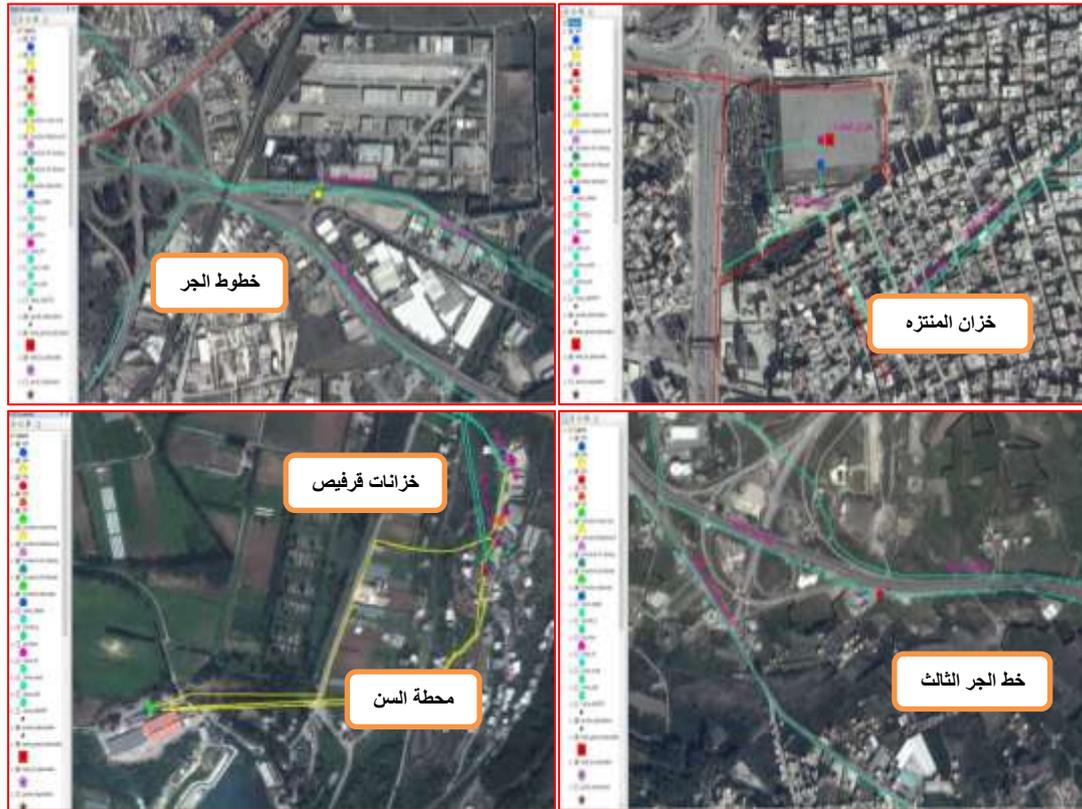
(r - Squared) (Coefficient Determination).

موقع منطقة البحث

وصف شبكة مياه الشرب

الحالة المدروسة : خط الجر الثالث الذي يغذي خزان المنتزه في مدينة اللاذقية.

يتم تغذية مدينة اللاذقية من محطة السن في بانياس عن طريق خزانات قرفيص الأربعة التي تقع على إرتفاع (92 m) عن سطح البحر بأربعة خطوط جر عن طريق الإسالة .



الشكل رقم (1) مخطط شبكة مياه الشرب الرئيسية التي تغذي مدينة اللاذقية.

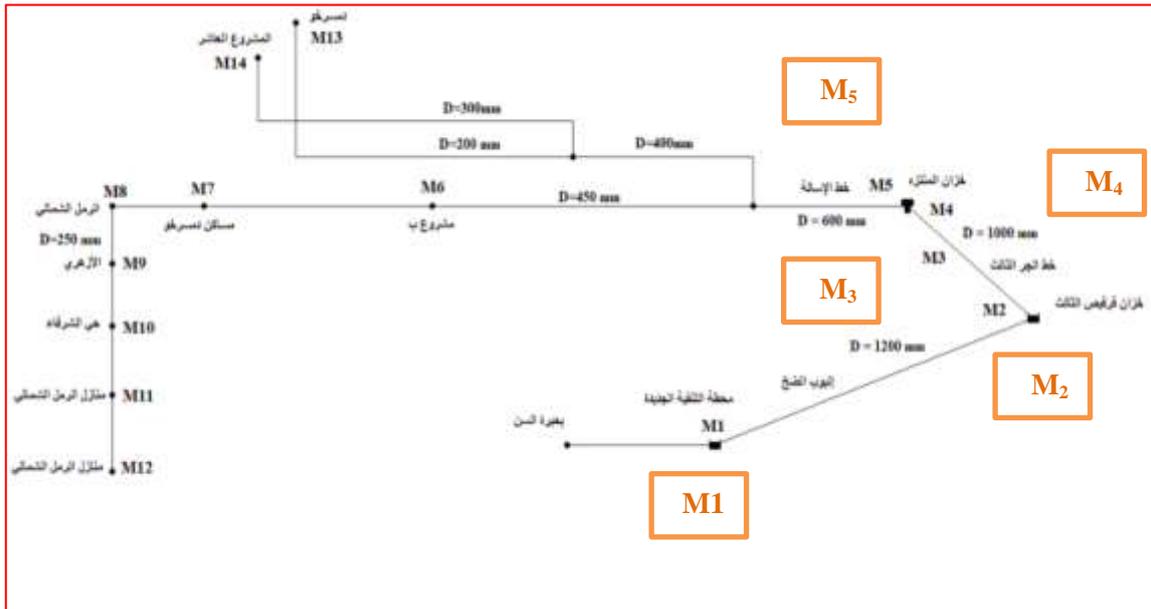
في الحالة المدروسة تم اختيار خط الجر الثالث الذي يغذي خزان المنتزه (50000 m^3) من خزان قرفيص الثالث (5000 m^3) بخط الجر الثالث بطول ($L=42 \text{ Km}$) وقطر ($D = 1000 \text{ mm}$) مصنوع من مادة الفونت.

طرائق البحث ومواده:

تناولت الدراسة الحقلية والمخبرية اختيار شبكة رصد مناسبة (محطة المعالجة - خزانات رئيسية - خطوط التغذية الرئيسية) لدراسة تأثير بعض بارامترات جودة المياه على تلاشي الكلور الحر المتبقي في مياه الشرب مؤلفة من خمس نقاط اعتيان موزعة على الشكل التالي ومبينة بالشكل رقم (2) :

■ الشبكة الرئيسية من محطة المعالجة (السن) حتى خزان المنتزه في مدينة اللاذقية

نقطة الاعتيان	موقع نقطة الاعتيان
M ₁	نقطة من خزان التجميع عند إضافة جرعة الكلور الأولية في محطة السن.
M ₂	نقطة دخول المياه إلى خزان قرفيص الثالث (T ₃) (نهاية إنبوب الضخ من محطة التنقية إلى خزان قرفيص الثالث (T ₃)).
M ₃	نقطة عند مدخل مدينة اللاذقية (الهنادي) (نقطة على خط الجر الثالث على أوتوستراد اللاذقية - دمشق تبعد حوالي (6 km) عن خزان المنتزه)
M ₄	نقطة قبل دخول المياه إلى خزان المنتزه (من خط الجر الثالث مباشرة).
M ₅	نقطة خروج المياه من خزان المنتزه (بداية خط الإسالة (600 mm)).



الشكل رقم (2) : مواقع نقاط الاعتيان المحددة.

طريقة جمع العينات

تم أخذ العينات المائية من نقاط الاعتيان المحددة بمعدل مرة واحدة في الشهر لإجراء التحاليل اللازمة لقياس بارامترات جودة المياه المحددة اعتباراً من شهر حزيران (2021) حتى شهر كانون الأول (2021) ، تم جمع العينات في عبوات زجاجية سعة (1000 ml) وذلك بعد غسل العبوة بماء العينة عدة مرات ثم إغلاقها بإحكام مع تثبيت شريط ورقي لاصق على كل عبوة مدون عليها (نوع العينة - رقم العينة - تاريخ قطف العينة) ثم نقلت العينات بواسطة حاوية مبردة إلى المخبر ليتم إجراء التحاليل اللازمة.

أما بالنسبة لقياس تركيز الكلور الحر المتبقي ودرجة الحرارة والأس الهيدروجيني تم إجراء القياسات في الموقع مع تحديد زمن القياس.

تم استخدام الأجهزة المتوفرة في مخبر كلية الهندسة المدنية والمعهد العالي لبحوث البيئة في جامعة تشرين.

النتائج والمناقشة:

حساب معامل الارتباط (R) (Correlation Coefficient)

تم استخدام لغة البرمجة (Python) لحساب معامل الارتباط (Correlation Coefficient) بين الكلور الحر المتبقي والبارامترات المدروسة (الفاصلة الزمنية - درجة الحرارة - الأس الهيدروجيني - العكارة - الفوسفات - النترات - المواد الصلبة المنحلة - الناقلية الكهربائية - كلوريد الصوديوم).

يوضح الجدول رقم (1) الكود البرمجي لمعامل الارتباط الخطي (R) بلغة البايثون (Python) لكل الأشهر المدروسة .

(CORRELATION COEFFICIENT(R) MODEL)

1- استدعاء المكتبات

```

import pandas as pd
import numpy as np
import statistics as st
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score
import seaborn as sns

DEFINE ( X AND Y )

Time_Interval=[0,58,105,131,141,0,42,80,94,124,0,28,78,98,108,0,21,65,78,88,0,20,73,84,100,0,26,80,100,110,0,23,45,65,83].
Water_Temperature=[21.5,23.3,23.5,25.8,23.7,18.7,21.1,22.3,23.3,22.3,18.7,21.2,23.3,23.7,23.5,18.7,21.3,22.3,23.5,21.7,17.5,18.7,21.5,23.5,22.5,16.5,16.7,17.7,18.7,18.3,10.3,10.5,11.5,13.7,12.7].
pH=[7.78,7.63,7.74,7.71,7.82,7.82,8.07,7.52,7.94,7.88,7.87,7.83,7.88,7.77,7.88,7.82,7.74,7.68,7.71,7.82,8.18,8.15,8.17,8.13,8.16,7.78,7.77,7.76,7.78,7.75,7.75,7.76,7.77,7.78,7.75].
Turbidity=[1.17,1.38,1.48,1.58,1.73,1.23,1.37,1.53,1.75,1.87,1.35,1.45,1.55,1.68,1.78,2.53,2.63,2.73,3.15,3.57,2.78,2.88,3.13,3.25,3.37,3.18,3.35,3.58,3.68,3.88,3.15,3.33,3.75,4.15,4.53].
Phosphate=[0.137,0.281,0.318,0.435,0.453,0.128,0.277,0.313,0.436,0.487,0.127,0.215,0.317,0.437,0.457,0.138,0.232,0.318,0.438,0.455,0.145,0.215,0.332,0.434,0.413,0.157,0.261,0.355,0.456,0.415,0.137,0.265,0.356,0.478,0.417].
Nitrate=[2.511,2.512,3.515,3.518,3.555,2.515,2.517,3.518,3.525,3.556,2.525,2.527,3.528,3.535,3.558,2.527,2.528,3.527,3.545,3.575,2.535,2.537,3.537,3.547,3.577,2.537,2.538,3.543,3.557,3.578,2.555,2.557,3.556,3.558,3.573].
Total_Dissolved_Solids=[221,231,235,237,253,225,235,237,241,255,227,237,243,245,256,235,238,247,248,257,237,243,245,247,258,245,247,253,257,273,251,255,257,263,278].
Conductivity=[441,442,443,445,453,443,444,445,453,455,445,446,447,455,457,447,448,453,457,458,434,445,447,456,455,433,444,445,455,453,435,445,447,453,455].
Nacl=[0.9,0.9,0.9,0.9,0.9,0.9,0.8,0.7,0.8,0.9,0.9,0.8,0.8,0.8,0.8,0.9,0.9,0.9,0.9,0.9,0.8,0.9,0.9,0.9,0.9,0.8,0.9,0.9,0.9,0.9,0.8,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9].
Free_Residual_Chlorine=[0.44,0.32,0.28,0.14,0.12,0.54,0.36,0.32,0.14,0.12,0.35,0.21,0.17,0.15,0.14,0.80,0.60,0.48,0.41,0.31,0.63,0.53,0.41,0.36,0.24,0.57,0.47,0.35,0.27,0.23,0.78,0.65,0.43,0.33,0.25].
dictt={'Time_Interval':Time_Interval,'Water_Temperature':Water_Temperature,'pH':pH,'Turbidity':Turbidity,'Phosphate':Phosphate,'Nitrate':Nitrate,'Total_Dissolved_Solids':Total_Dissolved_Solids,'Conductivity':Conductivity,'Nacl':Nacl,'Free_Residual_Chlorine':Free_Residual_Chlorine}
df=pd.DataFrame(dictt)
df
x=df[['Time_Interval','Water_Temperature','pH','Turbidity','Phosphate','Nitrate','Total_Dissolved_Solids','Conductivity','Nacl']]
y=df['Free_Residual_Chlorine']
df.corr()

```

يوضح الجداول رقم (2) نتائج قيم معامل الارتباط (R) (Correlation Coefficient) للكلور مع البارامترات المدروسة .

تحليل نتائج قيم معامل الارتباط (R) للأشهر المدروسة :

إن قيمة معامل الارتباط (R) تدل على أن علاقة الارتباط بين الكلور الحر المتبقي وبعض البارامترات المدروسة علاقة ارتباط سلبية (Negative Correlation).

حيث يتأثر تركيز الكلور الحر المتبقي بشكل عكسي مع بعض البارامترات المدروسة وإن زيادة تركيز أحد البارامترات المدروسة بمقدار وحدة واحدة يؤدي إلى انخفاض تركيز الكلور بمقدار معين يحدد لاحقاً وهو يختلف من بارامتر إلى آخر. نلاحظ أن الكلور الحر المتبقي يتأثر بشكل عكسي بالبارامترات التالية (الفاصلة الزمنية والفسفات والنترات والناقلية الكهربائية ودرجة الحرارة) في الأشهر المدروسة.

بلغت قيمة معامل الارتباط للكلور الحر المتبقي مع الفاصلة الزمنية ($R = -0.820907$) في الأشهر المدروسة.

بلغت قيمة معامل الارتباط للكلور الحر المتبقي مع الفوسفات ($R = -0.735094$) في الأشهر المدروسة.

بلغت قيمة معامل الارتباط للكلور الحر المتبقي مع النترات ($R = -0.66439$) في الأشهر المدروسة.

بلغت قيمة معامل الارتباط للكلور الحر المتبقي مع الناقلية الكهربائية ($R = -0.577086$) في الأشهر المدروسة.

بلغت قيمة معامل الارتباط للكلور الحر المتبقي مع درجة الحرارة ($R = -0.565092$) في الأشهر المدروسة.

بناء نماذج الانحدار (Regression Models)

• نموذج الانحدار الخطي المتعدد (Multi Linear Regression Model)

يوضح الجدول رقم (3) الكود البرمجي لنموذج الانحدار الخطي المتعدد بلغة البايثون (Python) لكل الأشهر المدروسة.

MULTI LINEAR REGRESSION MODEL

IMPORT LIBRARIES

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error
```

DEFINE (X AND Y)

<pre>x=[[0,21.5,7.78,1.17,0.137,2.511,221,441,0.9],[58,23.3,7.63,1.38,0.281,2.512,231,442,0.9], ,[105,23.5,7.74,1.48,0.318,3.515,235,443,0.9],[131,25.8,7.71,1.58,0.435,3.518,237,445,0.9], ,[141,23.7,7.82,1.73,0.453,3.555,253,453,0.9],[0,18.7,7.82,1.23,0.128,2.515,225,443,0.9], ,[42,21.1,8.07,1.37,0.277,2.517,235,444,0.8],[80,22.3,7.52,1.53,0.313,3.518,237,445,0.7],[9 4,23.3,7.94,1.75,0.436,3.525,241,453,0.8],[124,22.3,7.88,1.87,0.487,3.556,255,455,0.9],[0, 18.7,7.87,1.35,0.127,2.525,227,445,0.9],[28,21.2,7.83,1.45,0.215,2.527,237,446,0.8],[78,2 3.3,7.88,1.55,0.317,3.528,243,447,0.8],[98,23.7,7.77,1.68,0.437,3.535,245,455,0.8],[108,2 3.5,7.88,1.78,0.457,3.558,256,457,0.8],[0,18.7,7.82,2.53,0.138,2.527,235,447,0.9],[21,21.3 ,7.74,2.63,0.232,2.528,238,448,0.9],[65,22.3,7.68,2.73,0.318,3.527,247,453,0.9],[78,23.5,7 .71,3.15,0.438,3.545,248,457,0.9],[88,21.7,7.82,3.57,0.455,3.575,257,458,0.9],[0,17.5,8.18 ,2.78,0.145,2.535,237,434,0.8],[20,18.7,8.15,2.88,0.215,2.537,243,445,0.9],[73,21.5,8.17,3 .13,0.332,3.537,245,447,0.9],[84,23.5,8.13,3.25,0.434,3.547,247,456,0.9],[100,22.5,8.16,3. 37,0.413,3.577,258,455,0.9],[0,16.5,7.78,3.18,0.157,2.537,245,433,0.8],[26,16.7,7.77,3.35, 0.261,2.538,247,444,0.9],[80,17.7,7.76,3.58,0.355,3.543,253,445,0.9],[100,18.7,7.78,3.68, 0.456,3.557,257,455,0.9],[110,18.3,7.75,3.88,0.415,3.578,273,453,0.9],[0,10.3,7.75,3.15,0. 137,2.555,251,435,0.8],[23,10.5,7.76,3.33,0.265,2.557,255,445,0.9],[45,11.5,7.77,3.75,0.3 56,3.556,257,447,0.9],[65,13.7,7.78,4.15,0.478,3.558,263,453,0.9],[83,12.7,7.75,4.53,0.41 7,3.573,278,455,0.9]</pre>
SPLIT THE DATASET IN TRAINING SET AND TEST SET
<pre>x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.3, random_state = 42).</pre>
TRAIN THE MODEL IN THE TRAINING SET
<pre>LR = LinearRegression() LR.fit(x_train,y_train)</pre>
PREDICT THE TEST RESULTS
COMPUTE THE COEFFICIENTS
<pre>print(LR.coef_)</pre>
COMPUTE THE INTERCEPT
<pre>print(LR.intercept_)</pre>
EVALUTE THE MODEL
<pre>score=r2_score(y_test,y_prediction)</pre>
COMPUTE THE MEAN SQUARED ERROR
<pre>mean_sqrd_error=mean_squared_error(y_test,y_prediction)</pre>

• نموذج الانحدار التكيفي متعدد المتغيرات

(Multivariate Adaptive Regression Splines Model)

يوضح الجدول رقم (4) الكود البرمجي لنموذج الانحدار التكيفي متعدد المتغيرات بلغة البايثون (Python) لكل الأشهر المدروسة.

MULTIVARIATE ADAPTIVE REGRESSION SPLINES

IMPORT LIBRARIES

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from pyearth import Earth
from sklearn.metrics import r2_score
```

```

from sklearn.metrics import mean_squared_error

x=[[0,21.5,7.78,1.17,0.137,2.511,221,441,0.9],[58,23.3,7.63,1.38,0.281,2.512,231,442,0.9]
,[105,23.5,7.74,1.48,0.318,3.515,235,443,0.9],[131,25.8,7.71,1.58,0.435,3.518,237,445,0.9]
],[141,23.7,7.82,1.73,0.453,3.555,253,453,0.9],[0,18.7,7.82,1.23,0.128,2.515,225,443,0.9],
[42,21.1,8.07,1.37,0.277,2.517,235,444,0.8],[80,22.3,7.52,1.53,0.313,3.518,237,445,0.7],[9
4,23.3,7.94,1.75,0.436,3.525,241,453,0.8],[124,22.3,7.88,1.87,0.487,3.556,255,455,0.9],[0,
18.7,7.87,1.35,0.127,2.525,227,445,0.9],[28,21.2,7.83,1.45,0.215,2.527,237,446,0.8],[78,2
3.3,7.88,1.55,0.317,3.528,243,447,0.8],[98,23.7,7.77,1.68,0.437,3.535,245,455,0.8],[108,2
3.5,7.88,1.78,0.457,3.558,256,457,0.8],[0,18.7,7.82,2.53,0.138,2.527,235,447,0.9],[21,21.3
,7.74,2.63,0.232,2.528,238,448,0.9],[65,22.3,7.68,2.73,0.318,3.527,247,453,0.9],[78,23.5,7
.71,3.15,0.438,3.545,248,457,0.9],[88,21.7,7.82,3.57,0.455,3.575,257,458,0.9],[0,17.5,8.18
,2.78,0.145,2.535,237,434,0.8],[20,18.7,8.15,2.88,0.215,2.537,243,445,0.9],[73,21.5,8.17,3
.13,0.332,3.537,245,447,0.9],[84,23.5,8.13,3.25,0.434,3.547,247,456,0.9],[100,22.5,8.16,3.
37,0.413,3.577,258,455,0.9],[0,16.5,7.78,3.18,0.157,2.537,245,433,0.8],[26,16.7,7.77,3.35,
0.261,2.538,247,444,0.9],[80,17.7,7.76,3.58,0.355,3.543,253,445,0.9],[100,18.7,7.78,3.68,
0.456,3.557,257,455,0.9],[110,18.3,7.75,3.88,0.415,3.578,273,453,0.9],[0,10.3,7.75,3.15,0.
137,2.555,251,435,0.8],[23,10.5,7.76,3.33,0.265,2.557,255,445,0.9],[45,11.5,7.77,3.75,0.3
56,3.556,257,447,0.9],[65,13.7,7.78,4.15,0.478,3.558,263,453,0.9],[83,12.7,7.75,4.53,0.41
7,3.573,278,455,0.9]]
y=[0.44,0.32,0.28,0.14,0.12,0.54,0.36,0.32,0.14,0.12,0.35,0.21,0.17,0.15,0.14,0.80,0.60,0.
48,0.41,0.31,0.63,0.53,0.41,0.36,0.24,0.57,0.47,0.35,0.27,0.23,0.78,0.65,0.43,0.33,0.25].

FIT THE MODEL
model=Earth()
COMPUTE ( Y ) PREDICT
y_pred=model.predict(x)
EVALUTE THE MODEL
r2_score(y,y_pred)
COMPUTE THE MEAN SQUARED ERROR
mean_squared_error(y,y_pred)
SUMMARY OF RESULTS
print(model.summary())
    
```

يوضح الجدول رقم (5) مقارنة النتائج لنموذج الانحدار الخطي المتعدد و نموذج الانحدار التكيفي متعدد المتغيرات.

Compare the results between the two models		
	MULTI LINEAR REGRESSION	MULTIVARIATE ADAPTIVE REGRESSION SPLINES
EVALUTE THE MODEL (R ²)	0.8172305779858619	0.8625
COMPUTE THE MEAN	0.008449989261417289	0.0046

SQUARED ERROR (MSE)		
--------------------------------------	--	--

من الجدول السابق نلاحظ أن قيمة (R^2) في نموذج (MULTI LINEAR REGRESSION) هي ($R^2 = 0.81, MSE=0.008$) وأن قيمة (R^2) في نموذج (MULTIVARIATE ADAPTIVE REGRESSION SPLINES) هي ($R^2 = 0.86, MSE=0.0046$). وبالتالي بناءً على قيمة (R^2) الأعلى ($R^2 = 0.86$) فإن النموذج الأفضل والأكثر ملائمة للبيانات هو (MULTIVARIATE ADAPTIVE REGRESSION SPLINES) ونلاحظ أن (86%) من التباين في المتغير التابع يتم تفسيره بواسطة المتغيرات المستقلة، وأن (86%) من الانخفاض في الكلور مرتبط بعوامل معادلة الإنحدار وأن (14%) تعود إلى عوامل أخرى، نلاحظ أن كلما ارتفعت قيمة (R^2) يتم تفسير التباين من خلال متغيرات الإدخال ويكون النموذج أفضل.

الاستنتاجات والتوصيات:

الاستنتاجات

تم التوصل إلى مايلي :

- 1- علاقة الارتباط بين الكلور الحر المتبقي والبارامترات المدروسة علاقة ارتباط سلبية .
- 2- بلغت قيمة معامل الارتباط للكلور الحر المتبقي مع الفاصلة الزمنية ($R = -0.820907$) .
- 3- بلغت قيمة معامل الارتباط للكلور الحر المتبقي مع الفوسفات ($R = -0.735094$) .
- 4- بلغت قيمة معامل الارتباط للكلور الحر المتبقي مع النترات ($R = -0.66439$) .
- 5- بلغت قيمة معامل الارتباط للكلور الحر المتبقي مع الناقلية الكهربائية ($R = -0.577086$) .
- 6- بلغت قيمة معامل الارتباط للكلور الحر المتبقي مع درجة الحرارة ($R = -0.565092$) .
- 7- قيمة (R^2) في نموذج (MULTI LINEAR REGRESSION) هي ($R^2 = 0.81$) .
- 8- قيمة (R^2) في نموذج (MULTIVARIATE ADAPTIVE REGRESSION SPLINES) هي ($R^2 = 0.86$) .
- 9- قيمة (R^2) الأعلى ($R^2 = 0.86$) فإن النموذج الأفضل والأكثر ملائمة للبيانات هو : (MULTIVARIATE ADAPTIVE REGRESSION SPLINES)
- 10- إن (86%) من التباين في المتغير التابع يتم تفسيره بواسطة المتغيرات المستقلة.
- 11- إن (86%) من الانخفاض في الكلور مرتبط بعوامل معادلة الإنحدار وأن (14%) تعود إلى عوامل أخرى.
- 12- كلما ارتفعت قيمة (R^2) يتم تفسير التباين من خلال متغيرات الإدخال ويكون النموذج أفضل.

التوصيات:

- 1- ضرورة قياس الحديد والمنغنيز وإضافتها للنماذج لتحديد تأثيرها على تلاشي الكلور الحر المتبقي.
- 2- إجراء دراسة لتأثير نوعية الأنابيب على تلاشي الكلور.

References:

- 1- Hamdy, D. Moustafa, M.A.E. Elbakri, W. Free Residual Chlorine Calibration by WaterCAD at El-Nozha Water Network in Alexandria Governorate, Egypt, Journal of Environmental Protection, 2014, 5, 845-861.
- 2- Shamrukh,M. Hassan,Y. Chlorination and Optimal Chlorine Dosage For Nile Water, Firdt AIN Ahams University International Conference on Environmental Engineering, Egypt,2005,644-654.
- 3- Castro,P. Neves,M.Chlorine Decay in Water Distribution Systems Case Study Lousada Network, Electronic Journal of Environmental Agricultural and Food Chemistry,2003,2,261-266.
- 4- Monteiro,L. Figueiredo,D. Dias,S. Freitas,R. Covas,D. Menaia,J. Coelho,S.T. Modeling of chlorine decay in drinking water supply systems using EPANET MSX, 12th International Conference on Computing and Control for the Water Industry,2014, 1192 – 1200.
- 5- Sonia,AL. Effects of water quality character on chlorine decay in water distribution networks, Department of Sanitary and Environmental Engineering, Budapest Hungary,2017.
- 6- Tiruneh,A.T.Debessai,T.A.Bwembya,G.C.Nkambule,S.J.Zwane1,L. Variable Chlorine Decay Rate Modeling of the Matsapha Town Water Network Using EPANET Program, Journal of Water Resource and Protection, 2019, 11, 37-52.
- 7- Kahil,M.A. Application of First Order Kinetics for Modeling Chlorine Decay in Water Networks, International Journal of Scientific & Engineering Research,2016,7,331-336.
- 8- Ramos, H. M. Loureiro, Loureiro, D. Lopes ,A. Fernandes, C. Covas ,D. Reis , L. F. Cunha, M. C. Evaluation of Chlorine Decay in Drinking Water Systems for Different Flow Conditions: From Theory to Practice, Water Resour Manage,2010,24 ,815–834.
- 9- N. Fabio. Python Data Analytics, Springer Science+Business Media New York,2015, 348-4505.
- 10- S. A.L.Sayeth. T.Vinothraj, MMM.Fareez. Popular python libraries and their application domains, International Journal of Advance Engineering and Research Development,2020,7, 2348-4470.
- 11- W. Gerhard-Wilhelm,B. _Inci,K. Gulser,T. Pakize,O. Fatma Yerlikaya. CMARS: A New Contribution to Nonparametric Regression with Multivariate Adaptive Regression Splines Supported by Continuous Optimization, Turkish Scientific and Technological Research Institute (TUBITAK), 105-138.
- 12- E.Yvette, S.Justin, An introduction to Multivariate Adaptive Regression Splines for the cane industry, School of Engineering and Physical Sciences, James Cook University,2011,33,1-15.
- 13- S.Patrics, B.Christa, Correlation Coefficients: Appropriate Use and Interpretation, Wolters Kluwer Health, Inc.on behalf of the International Anesthesia Research Society.,2018, 126,1763-1768.