

وزارة التعليم العالي والبحث العلمي

Ministry of Higher Education and Scientific Research

جامعة قاصدي مرباح ورقلة

University of Kasdi Merbah Ouargla

كلية الرياضيات وعلوم المادة

Faculty of Mathematics and Matter Sciences

قسم الكيمياء

Chemistry département



مذكرة لاستكمال نيل شهادة الماستر أكاديمي

التخصص: كيمياء المحيط

من إعداد الطالبتين :

- حفصة حميتي

- هيام نجاح غطاس

بعنوان:

**مزامنة المعامل الطوبولوجي وطرق المجموعات المساهمة لحساب
معاملات امتصاص التربة *koc* للمركبات العضوية**

نوقشت علنا بتاريخ: 2026/06/10

أمام أعضاء لجنة المناقشة المكونة من:

رئيساً	جامعة ورقلة	أستاذ محاضر (أ)	- عبد الفتاح علاوي
مناقشاً	جامعة ورقلة	أستاذ محاضر (أ)	- خولة شاوش
مشرفاً ومقرراً	جامعة ورقلة	أستاذ مساعد (أ)	- مصطفى سعيدات
مساعد المشرف	جامعة ورقلة	أستاذ	- عبد القادر بن منين

الموسم الجامعي: 2025/2026

ملخص:

تقترح هذه الدراسة منهجية لتقدير معامل الامتصاص على الكربون العضوي K_{oc} للمركبات العضوية اعتمادا على مساهمة المجموعات الوظيفية والخصائص البنوية للجزيئات. وترتكز هذه الطريقة على الدمج بين مساهمات المجموعات ومؤشرات الاتصال الطوبولوجية التي تعكس طبيعة وترتيب الذرات داخل البنية الجزيئية مما يسهم في رفع دقة التنبؤ بقيمة K_{oc} .

وقد طبقت هذه المنهجية على عدد كبير من المركبات العضوية ذات التركيب المختلفة بما في ذلك المركبات العطرية والحلقية وغير المتجانسة بهدف تقييم سلوكها البيئي ودراسة انتقالها في التربة والمياه. وتعتمد هذه التنبؤات اساسا على الخصائص البنوية للمركبات دون الحاجة الى بيانات تجريبية معقدة الامر الذي يجعل هذه الطريقة اداة فعالة وسريعة لتقدير قابلية ارتباط الملوثات بالمادة العضوية في الاوساط البيئية.

الكلمات المفتاحية: معامل امتصاص التربة، K_{oc} ، طرق المجموعات المساهمة، المؤشرات

الطوبولوجية، الواصفات الجزيئية.

Cette étude propose une méthodologie pour estimer le coefficient d'adsorption sur le carbone organique K_{oc} des composés organiques, en se basant sur la contribution des groupes fonctionnels et les caractéristiques structurales des molécules. Cette méthode repose sur l'intégration des contributions de groupes et des indices de connectivité topologique, qui reflètent la nature et l'arrangement des atomes au sein de la structure moléculaire, contribuant ainsi à améliorer la précision de la prédiction des valeurs de K_{oc} .

Cette méthodologie a été appliquée à un grand nombre de composés organiques de structures diverses, y compris des composés aromatiques, cycliques et hétérocycles, dans le but d'évaluer leur comportement environnemental et d'étudier leur mobilité dans le sol et l'eau. Ces prédictions s'appuient principalement sur les propriétés structurales des composés sans nécessiter de données expérimentales complexes, ce qui fait de cette méthode un outil efficace et rapide pour estimer la capacité de liaison des polluants à la matière organique dans les milieux environnementaux."

Mots clé : Coefficient de sorption de sol, K_{oc} , contribution de group, descripteurs moléculaire, indices topologique

شكر و عرفان

قبل كل شيء نحمد الله عز وجل الذي أنعم علينا بنعمة العلم ووفقنا إلى بلوغ هذه الدرجة ونقول

" اللهم لك الحمد حتى ترضى ولك الحمد اذا رضيت ولك الحمد بعد الرضا "

ونسأل الله عز وجل أن نكون قد ساهمنا ولو بشيء قليل في إثراء الأبحاث المستقبلية بالمعلومات.

كما نتقدم بالشكر الخاص إلى الاستاذ المشرف " مصطفى سعيدات " و الاستاذ مساعد المشرف

عبد القادر بن منين نشكرهم جزيل الشكر على التوجيهات القيمة والإرشادات الصائبة التي لم يخلنا

بها طيلة هذا الوقت.

كما نخص بالشكر لجنة المناقشة الاستاذة *عبد الفتاح علوي* والاستاذة *شاوش خولة* على قبولها

الإشراف على لجنة المناقشة وقبولهما مناقشة هذا العمل. وإلى كل من ساعدنا في هذا العمل.

وفي الأخير نتقدم بشكر إلى كل الأساتذة و عمال ادارة كلية علوم المادة بورقلة وإلى كل من بدل معنا

جهدا ووفر لنا وقتا ونصح لنا قولاً.

اهداء

اهدي هدا العمل الى مدرسة الحب والاخلاص الى منبع العطف والحنان الى التي وعانها
في الليل والنهار سر نجاحي *امي الغالية*.

الى الذي اشترى لي اول قلم فكان الشمعة التي احترقت لتنير درب حياتي الى من
حصد الاشواك عن دبي لي محمد لي طريق العلم والنجاح الى من كللت انامله لي تقدم لنا لحظة
سعادة والدي العالي .

الى اخواتي الاعزاء والى ابنا نعم البراعم والى جدي وجدتي والى اعمامي وعماتي واخوالي وخالاتي
والى اقابي ورفقاء دبي والى كل من تقاسم معهم لحظات الفرح والحزن صديقات في الاقامة
الجمعية.

الى كل استاذ سناونا في انجز هدا العمل من قريب او بعيد.

هيام نجاح غطاس

اهداء

الهي لا يطيب الليل الا بشكرك ولا يطيب النهار الا بطاعتك. ولا تطيب اللحظات الا بذكرك.

ولا تطيب الاخرة الا برحمتك وعفوك ولا تطيب الجنة الا برؤيتك

"الله جل جلاله"

الى من بلغ وادى الامانة.. ونصح الامة.. الى نبي الرحمة ونور العالمين

"سيدنا محمد صلى الله عليه وسلم"

الى من كلله بالهيبة والوقار.. الى من علمني العطاء بدون انتظار. الى من افعمني بصريح

العطف.. الى من اغدقني بفيض الاحسان.. الى من احمل اسمه بكل افتخار

"ابي"

املاكي في الحياة. الى التي غمرتني بخالص الحب. الى بسمه الحياة وسر الوجود.. الى من كان

دعائها سر نجاحي وحنانها بلسم جراحي

"امي"

الى الذين اثروني بالاخاء.. الى من ناصروني في كل ان.. الى من ارى التفاؤل بعيونهم والسعادة

في ضحكتهم. الى من تطلعوا لنجاحي بنظرات الامل

"اخواني واخوتي"

الى الذين صنعوا في النجاح بنور العلم. الى روائع البيان اولهن ثورية بن عمية ثم

"اساتذتي"

الى من تحلو بالاخاء وتميزوا بالوفاء والعطاء. الى من اخلصوني الوداد، واصدقوني الخلان

"خطيبي و اصدقائي"

اهداء الى من كنت حسنة من حسناتها"والدي" حفظهما الله!

حميتي حفصه

قائمة الجدول

الصفحة	الجدول	الرقم
11	مصفوفة وينر	1. I
28	مثال على تطبيق طريقة جوباك وريد	1. II
30	المخطط الانسيابي للمتغيرات	2. II
30	ارتباط χ^2 و K_{33} مع بعضهما	3. II
31	ارتباط χ^2 مع المجموعات العضوية (أقوى 5)	4. II
31	ارتباط K_{33} مع المجموعات العضوية (أقوى 5)	5. II
31	ارتباط χ^2 و K_{33} مع $\log K_{oc}$	6. II
31	مشاكل الارتباط العالي (Multicollinearity) بين المجموعات العضوية	7. II
32	تباين المكونات الأولى (Variance explained)	8. II
32	إسهامات المتغيرات في المكونات الرئيسية	9. II
33	الإحصاءات الوصفية لـ $\log K_{oc}$	10. II
33	اكتشاف القيم الشاذة (Outliers)	11. II
34	القيم التي تتجاوز 5.10 (شاذة مرتفعة)	12. II
34	قيم K_{oc} ، χ^2 و K_{33} لمتما كبات Xylenes	13. II
35	قيم K_{oc} ، χ^2 و K_{33} لمتما كبات Decalin	14. II
35	ملخص النتائج قبل الانحدار	15. II
36	قائمة المتغيرات النهائية التي سأستخدمها في الانحدار	16. II
42	معاملات التعديل (الخطية)	17. II
42	تحليل التباين (الخطي)	18. II
49	المجموعة الوظيفية وتأثيرها	1. III
50	نتائج الناتجة عن الانحدار	2. III
51	معاملات النموذج	3. III
52	تحليل التباين (ANOVA) للنموذج الخطي	4. III
53	مقارنة النتائج والتجريبية والتنبؤية لـ $\log K_{oc}$ تحليل مجموع المربعات	5. III

من النوع الاول		
54	الخصائص الإحصائية تحليل مجموع المربعات من نوع الثالث	6.iii
55	معلمات النموذج او بارامترات النموذج	7.iii
56	معلمات النموذج او بارامترات النموذج	8.iii
57	اختيار المتغيرات	9.iii
62	نتائج دقة التنبؤ للبنزين	10.iii
64	نتائج تحليل البواقي للبنزين	11.iii
65	نتائج دقة التنبؤ الكلية	12.iii
66	تحليل عدم اليقين في معاملات النموذج	13.iii

قائمة الاشكال

الصفحة		
15	مخطط بياني لتصنيف طرق التنبؤ بالخواص	1.أ
16	تمثيل لتاريخ طرق حساب الخواص الفيزيائية والثرموديناميكية	2.أ
21	خصائص الحاسوب	3.أ
28	مبدأ التجزئة	1.ب
40	واجهه البرنامج	2.ب
40	اجراء الانحدار الخطي المتعدد	3.ب
41	نافذة تحديد المتغير التابع والمتغيرات المستقلة في تحليل الانحدار الخطي المتعدد	4.ب
43	اعداد بيانات الانحدار غير الخطي في البرنامج XLSTAT	5.ب
44	XLSTAT نافذة الانحدار غير الخطي في برنامج المستخدمة لاختيار النموذج اللوغرتمي وتقدير المعاملات	6.ب
58	مخطط البواقي مقابل القيم المتوقعة	1.ج
58	مخطط البواقي المعيارية رسم بياني لتحليل التباين (Analysis of Variance - ANOVA Plot)	2.ج
59	مخطط تحليل المسافات (Distance Analysis Plot)	3.ج
59	مخطط الارتباط	4.ج
60	الرسم البياني الشريطي للبواقي المعيارية لمجموعة التحقق (Standardized Residuals Bar Chart for Validation Set)	5.ج
60	رسم بياني شريطي للبواقي المعيارية (Standardized Residuals Bar Chart) لمجموعة النشطة	6.ج
61	توزيع معاملات عدم اليقين (أعلى 12 معامل)	7.ج
63	نتائج محاكاة مونتني كارلو للبنزين	8.ج
63	نتائج محاكاة مونتني كارلو للبنزين	9.ج
65	توزيع الأخطاء المطلقة	10.ج
67	توزيع اخطا التنبؤ والمقارنة بين القيم التجريبية والقيم التنبؤية	11.ج

قائمة المعادلات

الرقم	عنوان المعادلة	الصفحة
الفصل الاول		
1.1	معادلة معامل توزيع التربة /ماء	6
2.1	معادلة معامل امتصاص الفئة للكربون العضوي	7
3.1	معادلة علاقة الارتباط الخطي	7
4.1	معادلة لحساب مؤشر وينر	8
5.1	معادلة لحساب مؤشر رانديتش	9
6.1	معادلة لحساب مؤشر كيروهول	9
7.1	معادلة QSAR للمركبات الكارهة للماء	9
8.1	المعادلة الرياضية لحساب مؤشرات الاتصال الجزيئي من درجة صفرية	9
9.1	المعادلة الرياضية لحساب مؤشرات الاتصال الجزيئي من درجة اولى	10
10.1	المعادلة الرياضية لحساب مؤشرات الاتصال الجزيئي من درجة الثانية	10
11.1	الاساليب القائمة على مبدا مساهمات المجموعات	14
12.1	الاساليب القائمة على مساهمة التفاعلات الجماعية	14
13.1	دالة الهدف	14
14.1	معادلة النموذج المقترح من قبل كونستانتينو وغانى	15
15.1	نموذج طريقة ماريو وغانى	19
16.1	معادلات جوباك وريد للسعة الحرارية	19
17.1	معادلات اثنالبي التبخر لجوباك وريد	19
18.1	معادلات لزوجة السائل لجوباك وريد	19
19.1	معادلات اثنالبي الانصهار لجوباك وريد	19
20.1	معادلات نماذج الانحدار المقترحة من قبل كونستانتينو وغانى	19
21.1	معادلات النموذج الجديد لدرجة الغليان ماريو-مارجون وبارديلو-فونديفيللا	19
الفصل الثاني		
1.11	معادلة نموذج الانحدار الخطي المتعدد	24
2.11	معادلة النموذج اللوغرتمي غير الخطي مع ثابت تصحيح	24

24	معادلة النموذج اللوغرتمي البسيط	3.ii
24	معادلة النموذج اللوغرتمي النسبي	4.ii
24	معادلة النموذج القوة	5.ii
32	معادلة النهائية المقترحة	6.ii
36	معادلة النموذج الانحدار الخطي متعدد	7.ii
39	معادلة نموذج اللوغرتمي غير الخطي	8.ii

التسميات والمختصرات

P	خاصية
I_c	مؤشر اتصال اما X_v^1 او X_v^0
b و a	معامل الارتباط
N_i	عدد المجموعات من النوع i
P_i	خاصية i مساهمة p
PCA	تحليل المكونات الرئيسية
MLR	الانحدار الخطي المتعدد
Y_i	القيمة المرصودة للمتغير التابع للملاحظة i
X_{ij}	القيمة التي يأخذها المتغير j
ϵ	الخطأ النموذج
X	المتغيرات المستقلة
$\log K_{oc}$	معامل امتصاص التربة للمركبات العضوية
R^2	معامل التحديد
R^2 Adjusted	معامل التحديد المعدل
MSE	متوسط مربعات الخطأ
(Degrees of freedom)DF	درجات الحرية
MLR	الانحدار الخطي المتعدد
(Durbin-Watson)DW	اختبار دوربين -واتسون
SSE	مجموع مربعات الخطأ
RMSE	الجذر التربيعي لمتوسط مربعات الخطأ
Kappa3	مؤشر كبا الشكلي من الرتبة الثالثة
Chi2v	مؤشر كاي من الرتبة الثانية المصحح بالتكافؤ
Analysis of variance (ANOVA)	تحليل تباين
Pr >F)P-Value	القيمة الاحتمالية المستخدمة لتحديد الدلالة الاحصائية
PC	المكون الرئيسي
Q	الربيع
G_i	عدد مرات ظهور كل مجموعة وظيفية
$beta_i$	المعاملات الوزن الذي ستحسبه XLSTAT لكل مجموعة

الفهرس

I	الملخص
II	شكر والعرفان
III	الاهداء
V	قائمة الجداول
VII	قائمة الاشكال
VIII	قائمة المعادلات
XI	قائمة التسميات
XII	الفهرس
1	مقدمة عامة

الفصل الاول: بحث مكتبي وطرح الاشكالية

5	المقدمة.....	1.I
5	المصطلحات.....	2. I
5	الخواص الفيزيائية والكيميائية.....	1.2. I
5	الخواص الفيزيائية الاساسية.....	2.2.I
6	خواص الذوبان والتوزع بين الاطوار.....	3.2.I
6	الخواص الترموديناميكية.....	4.2.I
6	الخصائص الفيزيائية الديناميكية الحرارية.....	5.2.I
7	تصنيف الخصائص.....	6.2.I
7	توزيع الملوثات وارتباط الخواص بالنمذجة.....	7.2.I
7	معاملات امتصاص التربة.....	3.I
8	تعريف K_{oc}	1.3.I
8	اهمية معاملات الامتصاص في الدراسة توزيع المركبات العضوية..	2.3.I
8	الحاجة الى نماذج تنبؤية لمعاملات K_{oc}	3.3.I
8	المتغيرات الجزيئية المستخدمة في نماذج QSAR.....	4.3.I
8	تعريف K_d	5.3.I
9	معامل امتصاص الكربون العضوي في التربة.....	6.3.I
9	علاقة الارتباط K_{oc} والخواص الفيزيائية والكيميائية.....	7.3.I
10	العلاقة بين K_{oc} و K_{ow}	8.3.I
10	العلاقة بين K_{oc} والخواص الاخرى ودور النماذج البنيوية.....	9.3.I
10	المؤشرات الطوبولوجية.....	4.I
11	استخدام المؤشرات الطوبولوجية.....	1.4.I
11	مؤشر وينر.....	2.4.I
11	مؤشر الترابط الجزيئي للتكافؤ.....	3.4.I
12	معادلة QSAR للمركبات الكارهة للماء.....	4.4.I
12	المعادلات الرياضية لحساب مؤشرات الاتصال الجزيئي.....	5.4.I
12	طرق الحساب.....	5.I
12	تصنيف الطرق الحساب.....	1.5.I
14	الطرق التجريبية.....	2.5.I
14	الطرق الشبه تجريبية.....	3.5.I
14	الطرق التقليدية للتقدير.....	4.5.I

14	اختيار طرق الحساب.....	5.5.I
16	طرق المجموعة المساهمة.....	6.I
16	تاريخ طرق تقدير الخصائص الفيزيائية والديناميكية الحرارية.....	1.6.I
17	تصنيف الاساليب والمجموعات المساهمة.....	2.6.I
17	الاساليب القائمة على مبد مساهمات المجموعات.....	3.6.I
17	الاساليب القائمة على مبد مساهمة التفاعلات الجماعية.....	4.6.I
17	طريقة الجمع البسيطة.....	5.6.I
18	طريقة الجمع المعقدة.....	6.6.I
18	العلاقة الكمية بين البنية والخاصية الجزيئية.....	7.I
19	التطبيقات في مجال الكيمياء العضوية.....	8.I
19	التنبؤ بمعامل امتصاص المركبات العضوية في التربة.....	1.8.I
19	تصميم المذيبات الصناعية الصديقة للبيئة.....	2.8.I
20	التنبؤ بالسمية الحادة للملوثات العضوية.....	3.8.I
20	حساب معاملات امتصاص التربة.....	4.8.I
20	التنبؤ بخواص المركبات النقية QSPR.....	5.8.I
20	ابتكار الجزيئات عضوية النقية بخصائص مستهدفة.....	6.8.I
20	دراسة العلاقة الكمية بين البنية والنشاط QSAR.....	7.8.I
20	برنامج XLSTAT.....	9.I
21	خصائص الحاسوب.....	1.9.I
21	محاكاة مونتني كارلوا.....	10.I
22	النموذج الرياضي.....	11.I
22	دور الخصائص CAPE.....	1.11.I
24	الخلاصه.....	12.I
		المراجع

الفصل الثاني: الطريقة المطورة

27	المقدمة.....	1. II
27	طريقة تطوير النموذج.....	2. II
29	جمع البيانات وإعداد قاعدة البيانات.....	1.2.II
29	تحليل البيانات الأولية باستخدام XLSTAT.....	2.2. II
30	الدراسة الإحصائية الكاملة (قبل الانحدار).....	3.2.II
36	تطوير النماذج باستخدام XLSTAT.....	4.2.II
36	الانحدار الخطي المتعدد (Multiple Linear Regression – MLR).....	1.4.2.II
36	تعريف المتغيرات (Variables).....	3.II
37	خطوات التطبيق في XLSTAT.....	1.3.II
38	الخطوات التفصيلية للانحدار على XLSTAT.....	2.3.II
43	الانحدار غير الخطي (Nonlinear Regression).....	4.II
45	خطوات التطبيق في مونتني كارلو.....	5.II
45	الخلاصه.....	6.II

المراجع

الفصل الثالث: تحليل النتائج ومناقشتها

49المقدمة	1.III
49نتائج الانحدار	2.III
49تفسير المعاملات	3.III
49تحليل العلامة والحجم	1.3.III
50تقييم أهمية مؤشر الاتصال	4.III
50مقارنة أداء النموذج	1.4.III
61تحليل نتائج محاكاة مونت كارلو لحساب $\log K_{OC}$	5.III
61منهجية التحليل	1.5.III
62تحليل النتائج الخاصة بمركب البنزين (كحالة دراسة)	2.5.III
65التحليل الشامل لجميع المركبات	3.5.III
67الاستنتاجات العلمية	4.III
68الخلاصه	5.III
69		الخلاصة العامة
72		التوصيات

مقدمة عامة

مقدمة عامة

تعد دراسة السلوك البيئي للمركبات العضوية من أكبر التحديات التي تواجه الباحثين في الهندسة البيئية والكيمياء، حيث ترتبط عمليات مثل الامتزاز، والانتشار، والتحلل بشكل وثيق بخصائص الجزيئات الفيزيائية والكيميائية [1]. ومن بين هذه الخصائص، يبرز معامل امتصاص الكربون العضوي في التربة (K_{oc}) كمعامل حاسم يحدد مدى انتقال الملوثات وتوزيعها بين الطور المائي والمادة العضوية [1]. ونظراً للحاجة الملحة لتقدير هذه القيم للمركبات الجديدة أو تلك التي يصعب قياسها مخبرياً، برزت الحاجة إلى تطوير نماذج تنبؤية تعتمد على بنية الجزيء، مثل درجة الغليان والانصهار التي تعكس قوة التماسك الجزيئي وتؤثر مباشرة على سلوك المركب البيئي [2][3].

تعتمد المنهجيات الحديثة في التنبؤ بالخصائص الثرموديناميكية على مبدأ "مساهمة المجموعات" (Group Contribution)، وهي طريقة تقسم الجزيء إلى مجموعات وظيفية بسيطة، مثل تلك التي اقترحها "جوباك وريد" والتي تشمل 41 مجموعة وظيفية [2]. ومع ذلك، فإن هذه الطريقة قد تواجه قصوراً في التمييز بين المتماكبات (Isomers) التي تمتلك نفس المجموعات ولكن بترتيب فراغي مختلف [4].

ومن هذا المنطلق كيف يمكن تطوير نموذج رياضي دقيق يعتمد على المؤشرات الطوبولوجية ومساهمة المجموعات لتحسين دقة تقدير قيم معامل امتصاص التربة للمركبات العضوية؟

ولتجاوز هذا القصور، تم إدخال "مؤشرات الترابط الجزيئي" (Molecular Connectivity Indices) كبصمات هيكلية فريدة ترفع من كفاءة النموذج في التفريق بين الجزيئات ذات الصيغ المتشابهة [1][5]. إن دمج هذه المؤشرات الطوبولوجية مع طرق مساهمة المجموعات يفتح آفاقاً جديدة لتطوير نماذج رياضية أكثر دقة وموثوقية في محاكاة السلوك البيئي للمركبات [6].

حيث توفر "مؤشرات تشي (Chi)" وصفاً دقيقاً للترابط الجزيئي وقدرة الجزيء على النفاذ والامتصاص [7]، بينما تساهم "مؤشرات كبا (Kappa)" في وصف شكل الجزيء ومرونته، مما يفسر الاختلافات السلوكية بين المتماكبات الفراغية (مثل cis و trans) [8] إن دمج هذه المعايير يتيح بناء نماذج قادرة على معالجة قواعد بيانات ضخمة تتجاوز 500 مركب بدقة عالية [9].

في هذا العمل، تمت دراسة تطوير نموذج رياضي هجين للتنبؤ بمعامل امتصاص الكربون العضوي في التربة (k_{oc}) للمركبات العضوية، وذلك من خلال المزامنة بين منهجية مساهمة المجموعات والمؤشرات الطوبولوجية، وقد تم تقسيم الدراسة وفق الهيكلية التالية:

- يتناول الفصل الأول (البحث المكتبي وطرح الإشكالية): استعراضاً شاملاً للمفاهيم الأساسية المرتبطة بالخواص الفيزيائية والثرموديناميكية ذات الصلة بالبيئة، مع التركيز على أهمية معامل ϵ_{oc} في تحديد سلوك الملوثات. كما يستعرض الفصل المناهج الحسابية التقليدية والأسس النظرية لطرق التنبؤ، وصولاً إلى صياغة إشكالية البحث وأهدافه.
- يتناول الفصل الثاني (منهجية بناء النموذج): شرحاً مفصلاً للمنهجية المقترحة، حيث يتم عرض كيفية تجزئة البنية الجزيئية وفق مجموعات "جوباك وريد"، ودمجها مع مؤشرات الاتصال الجزيئي للتكافؤ (مؤشرات تشي وكابا) التي تعمل كبصمة تمييزية بين المتماكبات. كما يوضح هذا الفصل قاعدة البيانات المستخدمة وإجراءات التحليل الإحصائي عبر الانحدار الخطي وغير الخطي.
- يتناول الفصل الثالث (تحليل النتائج ومناقشتها): عرضاً ومناقشة للنتائج التي تم الحصول عليها باستخدام برمجية XLSTAT ، حيث يتم تقييم دقة النموذج المطور ومقارنته بالطرق التقليدية. كما يتضمن هذا الفصل دراسة إحصائية معمقة للارتباط بين القيم التجريبية والمتوقعة، وتحليل فواصل الثقة والأخطاء لضمان موثوقية النموذج في التطبيقات البيئية.

الفصل الأول:

بحث مكتبي و طرح الاشكالية

بحث مكتبي و طرح الاشكالية

1.I المقدمة:

تعد معاملات امتصاص التربة ولا سيما معامل امتصاص التربة K_{oc} ، من أبرز الخصائص التي تؤثر بشكل كبير على سلوك المركبات العضوية في البيئة، فهي تلعب دورا رئيسيا في تحديد مدى توزيع هذه المركبات بين الطور المائي والمادة العضوية في التربة .

سنقدم في هذا الفصل تعريفات عامة لجميع المصطلحات المستخدمة في الفصول التالية.

2.I المصطلحات:

يهدف هذا الفصل إلى تقديم المفاهيم الأساسية المرتبطة بالخواص الفيزيائية والثرموديناميكية المتعلقة بالبيئة، إضافة إلى معاملات امتصاص التربة، كتمهيد لاستخدام المؤشرات الطوبولوجية في حساب K_{oc} للمركبات العضوية، كما يتطرق إلى استعراض المناهج الحسابية المختلفة المستخدمة في التنبؤ بهذه القيم.

1.2.I الخواص الفيزيائية والثرموديناميكية المتعلقة بالبيئة:

تلعب الخصائص الفيزيائية والثرموديناميكية دورا أساسيا في تحديد سلوك المركبات العضوية داخل الانظمة البيئية. حيث ترتبط عمليات مثل الامتزاز، الانتشار، التحلل، التبخر، والذوبان بشكل وثيق بخصائص الجزيئات بما في ذلك القطبية الذوبانية في الماء وثوابت الاتزان [1].

2.2.I الخواص الفيزيائية الاساسية:

هذه الخواص تستخدم في النماذج وفي اساليب مساهمة المجموعات.

• درجة الانصهار والغليان:

تعكس هاتين الخاصيتين قوة التماسك بين الجزيئات [4] وتستخدمان في تقدير التبخر وتصرف المركب عند تحولات الحرارة [10].

• الضغط البخاري:

يعبر ميل المادة المركبة لانتقال من الطور السائل/الصلب الى الطور الغازي [4] ويستخدم لتقسيم التطاير والانتقال الملوث من الماء او التربة الى الهواء [10].

• الكثافة:

تسهم في وصف تصرف السوائل والمخاليط العضوية وتدخل في نماذج تقدير السمات واختيار المذيبات والمكونات العضوية الملائمة للتطبيقات الصناعية والبيئية [5].

3.2.I خواص الذوبان والتوزع بين الاطوار:• **الذوبانية في الماء:**

تستخدم خاصية الذوبانية لتجديد نسبة المركب الذي يبقى في الطور المائي، مما يساعد في تقييم قدرته على الانتقال مع المياه السطحية او الجوفية [1] او على العكس من ذلك ميله الى الامتصاص على المادة العضوية [4]. تعد الذوبانية خاصية فيزيائية اساسية تستخدم في تصميم المذيبات، وكذلك في نماذج تحديد سلوك الملوثات [11].

• **معامل التوزيع اوكتانول/ماء (K_{ow}):**

يعد (K_{ow}) مؤشرا شائع الاستخدام لتحديد درجة كراهية المركبات للماء وميولها نحو الطور العضوي. وقد استخدم على نطاق واسع في العديد من الدراسات كأساس لتطوير $\log K_{oc}$ علاقات تجريبية تربط بين قيمة وبين معاملات الامتصاص او السمية. تشمل هذه الدراسات نماذج تعتمد على مؤشرات الترابط الجزيئي والموصفات الطوبولوجية [5] و بالإضافة الى ذلك فان ثابت هنري الذي يربط بين الضغط البخاري والذوبانية المائية يستخدم لوصف انتقال الملوثات بين الماء والهواء [1].

4.2.I الخواص الثرموديناميكية:

تعرف الخواص الثرموديناميكية بانها المتغيرات التي تحدد الحالة المايكروسكوبية للنظام وتصنف الى حالتين حوص مركزة وخواص شمولية [2] حيث تؤدي هذه الخواص دورا محوريا في تحديد اتجاه العمليات من خلال قياس التغير في الطاقة Gibbs الحرة والعشوائية [3].

تصنيف الخواص الثرموديناميكية

• الخواص المركزة: هي الخواص التي لا تعتمد على كمية المادة او حجم النظام مثل درجة الحرارة والضغط والكثافة.

• الخواص الشمولية: هي الخواص التي تتناسب قيمتها طرديا مع كمية المادة او الحجم النظام مثل الكتلة والحجم والطاقة الداخلية.

5.2.I الخصائص الفيزيائية – الديناميكية الحرارية:

خصائص المادة هي التي تستخدم لتحديدها او وصفها فعندما نقول ان الماء مبلل الفضة لامعة فإننا نصف المواد من حيث خصائصها، يمكن تقسيم الخصائص الى فئتين فيزيائية وكيميائية وهناك نوعان اساسيان من الخصائص الديناميكية الحرارية [12].

❖ **الخواص الكيميائية:**

هي الخصائص التي تغير الطبيعة الكيميائية للمادة، حيث لا يمكن ملاحظة هذه الخصائص الا اثناء التفاعل الكيميائي على سبيل المثال لا يمكنك تحديد مدى قابلية الكبريت للاشتعال الا عند محاكاة عملية احتراقه فعليا، وبالمثال، يعد التفاعل العنيف لعنصر الصوديوم مع الماء هو نموذج للخاصية الكيميائية.

❖ الخواص الفيزيائية:

تمثل الخصائص التي لا تغير من الطبيعة الكيميائية للمادة، بحيث يمكن ملاحظة الخصائص الفيزيائية لمادة نقية دون تغيير هويتها. وغلبا تتضمن هذه الملاحظات على نمطين من التوصيف: قياسات عددية كمية، او وجود وصف نوعي غير عددي للخاصية. وتجدر الإشارة الى ان المقرارات والمراجع العلمية المتخصصة تسرد نماذج متنوعة لهذه الخصائص تبعا لسياق الدراسة، من ابرز الامثلة الشائعة عليها : اللون، ونقطة التبريد، ونقطة الغليان [12].

6.2.I تصنيف الخصائص [12]:

يمكن تصنيف الخصائص الفيزيائية والديناميكية الحرارية الى فئتين:

❖ الخصائص الضمنية:

هي الخصائص التي لا يمكن حسابها الا تجريبيا.

❖ الخصائص الصريحة:

هي الخصائص التي يمكن تقدير قيمها بواسطة نماذج رياضية ويمكن ايضا قياسها تجريبيا.

7.2.I توزيع الملوثات وارتباط الخواص بالنمذجة:

نستخدم الخصائص بشكل فردي او مجتمعة، ضمن نماذج المحاكاة الخاصة بالتنبؤ بالمسار البيئي للملوثات وسلوكها الحركي، وكيفية توزيع المركبات العضوية بين الهواء والماء والتربة والكائنات الحية، اضافة الى تقدير مدة بقائها وامكانية تراكمها وانتقالها عبر السلاسل الغذائية [1]. وتعتبر هذه النماذج اساسية في تطوير علاقات خاصة-بنية QSRR/QSAR حيث يتم استخدام مواصفات بينية مثل المؤشرات الطبولوجية والوصف الطيفي Signature لربط بنية المركب بخصائصه البيئية [5].

I 3. معاملات امتصاص التربة:

تصنف معاملات امتصاص التربة من اهم الادوات المستخدمة في الكيمياء البيئية لدراسة سلوك المركبات العضوية داخل المكونات الارضية؛ فهي تلعب دورا رئيسيا في تحديد كيفية توزيع هذه المركبات بين الطور المائي والطور الصلب وخاصة في الطبقات الغنية بالمادة العضوية [1]. تساهم هذه المعاملات بشكل كبير في نماذج تتبع مصير الملوثات حيث تساعد في تقدير حركية المركبات العضوية، وامكانية تسربها الى المياه الجوفية بالإضافة الى قياس درجة تراكمها في الطبقات السطحية والرواسب لذا بعد فهم هذه

العمليات وروابطها بالخصائص الفيزيائية والكيميائية والبنية الجزيئية امرا اساسيا لتقييم المخاطر البيئية بدقة [5].

1.3.I تعريف K_{oc} :

يعرف معامل امتصاص التربة بأنه مقياس كمي يعبر عن قدرة التربة او الرواسب على امتصاص الملوثات العضوية من المحاليل المائية. يعتبر دليلا يساعد في تقييم حركة هذه المركبات داخل نظام التربة والماء [6].

2.3.I اهمية معاملات الامتصاص في دراسة توزيع المركبات العضوية:

- (1) تلعب عملية امتصاص المركبات الكيميائية في التربة و رواسب عاملا اساسيا يؤثر على انتقالها وحركتها داخل البيئة.
 - (2) يلعب الامتصاص دورا كبيرا في التأثير على التحولات الكيميائية والبيولوجية للمركبات في الوسط المائي.
- لذلك فان معرفة او تقدير معاملات الامتصاص مثل K_{oc} ضروري لتقييم توزيع الملوثات والمخاطر البيئية [6].

3.3.I الحاجة الى نماذج تنبويه لمعاملات K_{oc} :

بالرغم من اجراء تجارب مخبرية على مدى أكثر من 30 عاما الا ان البيانات المتوفرة حول معاملات الامتصاص لا تتجاوز 500 مركب كيميائي، اضافة الى ذلك تفنقر العديد من المركبات ذات الانتاج الكبير الى البيانات الموثوقة، بناء على ذلك و لتقدير معاملات QSAR تم تطوير نماذج الامتصاص استنادا الى الخصائص الجزيئية [6].

4.3.I المتغيرات الجزيئية المستخدمة في نماذج QSAR:

تم استخدامها كمؤشرات بنوية لأنها:

- متوفرة لألاف المركبات الكيميائية.
- يمكن حسابها بواسطة برامج عالية الدقة مثل معامل تقسيم الاوكتانول-الماء المحسوب حاسوبيا $ClogP$.
- يمثل مقياسا لمدى كراهية المركب للماء Hydrophobicity [6].

5.3.I تعريف k_a :

يعرف k_a كقياس لنسبة تركيز المركب الممتز على الطور الصلب في التربة الى تركيزه في الطور المائي عند الوصول الى حالة التوازن. يعكس هذا المعامل مدى ارتباط المركب بسطح التربة مقارنة ببقائه في الماء [1].

تشير القيم العالية لـ K_d الى امتزاز قوي للمركب وضعف حركته في الطور المائي، بينما تدل القيم المنخفضة على وجود نسبة أكبر من المركب في المحلول وزيادة قابليته للانتقال والترشح. ومع ذلك يتأثر K_d بشكل واضح بخصائص التربة مثل محتوى الكربون العضوي ونسبة الطين [1]. تعطى علاقة k_d بشكل التالي:

$$K_d = \frac{C_s}{C_w} \quad (1. I)$$

6.3.I معامل امتصاص التربة k_{oc} :

لتقليل تأثير الاختلاف في محتوى الكربون العضوي بين عينات التربة المختلفة، يتم اللجوء الى تطبيع (Normalization) معامل التوزيع العام (k_d) منسوبا الى الكسر الكتلي للكربون العضوي الكلي f_{oc} المتواجد في الطور الصلب، ويعبر عن معامل امتصاص الكربون العضوي (k_{oc}) وفقا للعلاقة الرياضية التالية [5]:

$$k_{oc} = \frac{k_d}{f_{oc}} \quad (2. I)$$

تشير قيم k_{oc} المرتفعة الى ميل المركب العضوي للامتصاص على المادة العضوية للتربة، مما يجعله اكثر خاصية تعكس السلوك الذاتي للمركب الكيميائي بغض النظر عن طبيعة التربة. يستخدم k_{oc} هذه القيم بشكل واسع في نماذج تتبع مصير الملوثات لتقدير حركة المركبات العضوية عبر اعمدة التربة وكذلك لتقييم احتمالات تسربها نحو المياه الجوفية او درجة تراكمها في الطبقات السطحية والرواسب الارضية؛ كما تم اعتماد هذه القيم كمتغير استجابة رئيسية في نماذج *Bahnick* و *Doucette*. تعتمد هذه النماذج على المؤشرات الطبولوجية لتقدير قدرة امتصاص المركبات العضوية على التربة [5].

7.3.I علاقة ارتباط K_{oc} والخواص الفيزيائية-الكيميائية:

نظرا لصعوبة المعايرة المخبرية المباشرة لمعامل K_{oc} ، فقد عكفت العديد من الدراسات على تطوير نماذج رياضية وعلاقات ارتباطية متعددة تربط بين $\log K_{oc}$ وبعض الخصائص الفيزيوكيميائية مثل معامل تقسيم اوكتانول-ماء k_{ow} والذوبانية المائية [1]. تعتمد هذه الروابط على نماذج انحدار خطية او غير خطية، يتم ضبط متغيراتها بناء على معطيات تجريبية لفئات محددة من الجزيئات العضوية، وتستعمل لاحقا لتقدير قيم K_{oc} بشكل موثوق عند توفير قيم $\log K_{ow}$ او S [5]

8.3.I العلاقة بين k_{oc} و k_{ow} :

تشير العديد من الدراسات والابحاث الى وجود ارتباط وثيق بين معامل k_{oc} و معامل k_{ow} لأن كلا المعاملين يعبران عن درجة الانجذاب المركب للطور العضوي مقارنة بالطور المائي [1,5]. عادة ما تتم صياغة هذه العلاقة باستخدام نماذج الانحدار الخطي تأخذ بشكل التالي:

$$\log K_{oc} = a \log K_{ow} + b \quad (3.1)$$

حيث تستخرج معاملات الانحدار a و b من بيانات تجريبية لفئات محددة من المركبات العضوية [1]، للإشارة الى العلاقة بين الخصائص الكيميائية. توضح هذه العلاقات ان المركبات ذات القيم العالية ل $\log K_{ow}$ تميل الى امتلاك قيم مرتفعة ل k_{oc} مما يجعل الملوثات اكثر قابلية قبل أن يتطور الامر الى استخدام النماذج الهيكلية التي تعتمد على المؤشرات الطوبولوجية والواصفات الجزيئية اكثر دقة للتمييز بين المتماكبات الكيميائية [1, 5].

9.3.I العلاقة بين K_{oc} والخواص الاخرى ودور النماذج البنوية:

أضافة الى الارتباط الوثيق بمعامل K_{ow} ، يمكن ربط معامل امتصاص التربة K_{oc} بصفات فيزيائية وكيميائية اخرى، مثل الذوبانية المائية S و بمحددات هيكلية تستخلص مباشرة من تركيبة الجزيئية (مثل المؤشرات الطوبولوجية والواصفات البنوية Signature descriptors ومساهمات المجموعات)، بهدف تحسين و رفع دقة التنبؤ بمعاملات الامتصاص لفئات متنوعة من الملوثات العضوية [1]. في هذا الاطار استخدمت مؤشرات الترابط الجزيئي والمؤشرات الطوبولوجية بنجاح في اعمال Bahnick و Doucette لربط بنية المركبات بمعاملات الامتصاص والسمية البيئية [5]، بينما طور Foulon و Doucette و Koch في ادراج واصفات البصمة الهيكلية Signature لاستعمالها في نماذج QSAR/QSPR [10].

وتوفر طرق مساهمات المجموعات التي وضعها Reid و Joback، ثم طورها [3]، بالإضافة الى اسهامات Friedenslund وزملائه في اطار تقدير الخواص الفيزيائية-الكيميائية للمركبات النقية، والتي يمكن ادخالها وادماجها كمتغيرات في نماذج تقدير ومعايرة معاملات الامتصاص البيئية الاخرى [4, 11, 13].

4.I المؤشرات الطوبولوجية:

تصف المؤشرات الطوبولوجية طبيعة الترابط الذري داخل الجزيء، كما تقدم هذه المؤشرات معلومات حول حجم الجزيء و شكله العام ومدى تفرعته. تعتمد هذه المؤشرات على نظرية المخططات البيانية المطبقة على جدول الترابط البنوي، والذي يقدم تمثيلا موجزا لألية ارتباط بين الذرات داخل الجزيء. ومن أكثر المؤشرات الطوبولوجية استخداما في هذه المجال: مؤشر وينر (Wiener) [7]، مؤشر رانديتش (Randic) [13] ومؤشر كير-هول (Kier-Hall) لترابط التكافؤ [8] ومؤشر بالابان (Balaban) [14]. يعبر عن مؤشر "وينر" باستخدام مصفوفة المسافة. هذه المصفوفة مربعة بأبعاد $(N_s \times N_s)$ وعناصرها d_{ij} تشير الى اقل عدد من روابط (اقصر مسار) التي تفصل بين درتين i و j [12].

1.4.I استخدام المؤشرات الطوبولوجية

اقترح الباحثون استخدام المؤشرات الطبولوجية للجزيئات لتقدير معاملات الامتصاص في التربة. وتتميز هذه المؤشرات بكونها تعتمد على البنية الجزيئية فقط، ويمكن حسابها رياضيا وتوقع قيمها بدقة دون الحاجة الى تجارب مخبرية [8].

2.4.I مؤشر وينر:

طور أول مؤشر طبولوجي بواسطة الكيميائي هاري وينر (H. Wiener) عام 1947. ويعرف مؤشر وينر رياضيا بأنه مجموع المسافات الطبولوجية بين جميع الذرات في المخطط البياني الجزيء. ويتم حساب هذا المؤشر انطلاقا من مصفوفة المسافة (D) [12]. وفقا للمعادلة التالية:

$$D = W(D) = \frac{1}{\sum_{i=1}^A} 2 \sum_{j=1}^A dij \quad (4.I)$$

الجدول (1.I): مصفوفة وينر

6	5	4	3	2	1	الذرة
2	4	3	2	1	0	1
1	3	2	1	0	1	2
2	2	1	0	1	2	3
3	1	0	1	2	3	4
4	0	1	2	3	4	5
0	4	3	2	1	2	6

$$W = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n d_{ij} = 32$$

3.4.I مؤشر الترابط الجزيئي للتكافؤ [12]:

يعد مؤشرا رانديتش وكيروهاال من أكثر المؤشرات الطبولوجية استخداما واعتمادا في الدراسات البنوية. وفيما يلي الصيغ العامة المعتمدة لحساب هذين المؤشرين

$$X^v = \sum_{i=1}^{N_s} \prod_{K=1}^{m+1} \left(\frac{1}{\delta_k^v} \right)^{\frac{1}{2}} \quad (5. I)$$

$$\delta_k^v = \frac{Z_k^v - H_k}{Z_k - Z_k^v - 1} \quad (6. I)$$

4.4.I معادلة QSAR للمركبات الكارهة للماء:

تشتق علاقة كمية بين البنية والنشاط (QSAR) للمركبات الكارهة للماء وفقا لعلاقة التالية:

$$\log K_{oc} = 0.70 + 0.52 \times x_1^V \quad (7. I)$$

5.4.I المعادلات الرياضية لحساب مؤشرات الاتصال الجزيئي (MCI) :

يمكن حساب مؤشرات الاتصال الجزيئي بالاعتماد على خوارزمية راندتش و التي تعتمد على مجموع مقاليب الجذور التربيعية لقيم الدلتا δ المرتبطة بالروابط والذرات في الجزيء. ويتم حساب المؤشرات لرتب مختلفة مثل (الرتبة الصفرية، الاولى، الثانية) وفقا للمعادلات التالية:

$$\chi^0 = \sum \frac{1}{\sqrt{\delta_i}} \quad (8. I)$$

$$\chi^1 = \sum \frac{1}{\sqrt{\delta_i \times \delta_j}} \quad (9. I)$$

$$\chi^2 = \sum \frac{1}{\sqrt{\delta_i \times \delta_j \times \delta_k}} \quad (10. I)$$

5.I طرق الحساب:

لتغلب على العقبات المصاحبة للقياسات التجريبية، يلجأ الباحثون الى اعتماد أساليب الحساب التنبؤية المرتكزة على مفهوم طرق مساهمة المجموعات الوظيفية (Group Contribution Methods) في التكوين الجزيئي للمادة. ومن ابرز هذه المنهجيات طريقة "يونفاك" (UNIFAC) التي طورها فريدنسلاند (Fredenslund) وآخرون لتخمين توازن الاطوار، بالإضافة الى المنهجيات المتقدمة المستعملة لتقدير الخصائص المواد النقية مثل طريقة ريد وآخرون [12].

رغم الأهمية البالغة لهذه الأساليب في التنبؤ الديناميكي الحراري، الا أن نطاق تطبيقها يظل محكوما بمجموعة من الافتراضات الأساسية التي قد تحد من دقتها التنبؤية، مما يجعل التنبؤ الدقيق ببعض الخصائص امرا معقدا. ورغم التحسينات المستمر والتي أدخلت على هذه النماذج كالتعديلات الهيكلية التي اجراها كونستانينو وغاني (Constantinon & Gani) الا انها تعد حولا مؤقتة وضمنية، كونها تشتت مسبقا توفير قاعدة بيانات تجريبية شاملة تغطي جميع الأنواع الكيميائية المدروسة تؤكد من موثوقيتها.

1.5.I تصنف طرق الحساب:

تصنف طرق الحساب الى فئتين رئيسيتين [12] :

اولا: الطرق المرجعية:

وهي الطرق التي يمكن التحقق من صحة نتائجها من خلال مقارنتها بطرق تقدير اخرى معتمدة او ابسط.

ثانيا: الطرق التقريبية:

وهي الطرق التي يتم الحصول عليها اعتمادا على تحليل الانحدار للبيانات التجريبية، ويمكن تقسيمها الى نوعين:

2.5.I الطرق التجريبية:

تعتمد على علاقات ارتباطية بين K_{oc} و K_{ow} و S وعلى ربط البيانات التجريبية [1] ، وتشمل نطاق الخصائص الأولية والثانوية. ويجد الاشارة الى أن هذه الطرق قابلة للتطبيق على تصميم العمليات البيوكيماوية والبيئية.

3.5.I الطرق الشبه تجريبية [12] :

ويشمل هذا النوع على الطرق التالية:

- (1) الطرق المعتمدة على مبدا الحالات المتقابلة.
 - (2) الطرق التي تعتمد على الطوبولوجية والهندسة الجزيئية.
 - (3) الطرق التي تقوم على اضافة مساهمات الذرات او الروابط او مجموعات الجزيئات. حيث تدمج ايضا بين المؤشرات الطوبولوجية والخواص الفيزيوكيميائية.
- وتعد الطرق المعتمدة على مبدأ الجمع، من أكثر الاساليب شيوعا في مجال تصميم المركبات وتطوير العمليات الصناعية. في حين تتميز الطرق المعتمدة القائمة على الطوبولوجيا والهندسة الجزيئية باتساع مجال تطبيقها [12] ، فان طرق المقارنة المبنية على الحالات المتقابلة تظل الاعلى دقة ،على الرغم من ان نطاق استخدامها يظل محدودا في بعض الاحيان [1.10].

4.5.I الطرق التقليدية للتقدير:

من الطرق المستخدمة سابقا، هناك منهجيات تعتمد بشكل اساسي معامل التقسيم اوكتانول/ماء (K_{ow})، أو على قيمة الذوبانية في الماء، لكن هذه الطرق صالحة فقط لبعض اصناف المركبات حيث لا تتوفر فيها البيانات التجريبية الخاصة بقيم S او K_{ow} [8].

5.5.I اختيار طرق الحساب:

يعتمد اختيار طريقة حساب خصائص المركبات على العديد من العوامل الاساسية، ومن اهمها دقة النتائج المطلوبة، ومدى توفر المعطيات التجريبية، بالإضافة الى طبيعة الفئة الكيميائية للمركبات المدروسة والهدف من هذه الدراسة. حيث نميز حالتين من طرق الحساب:

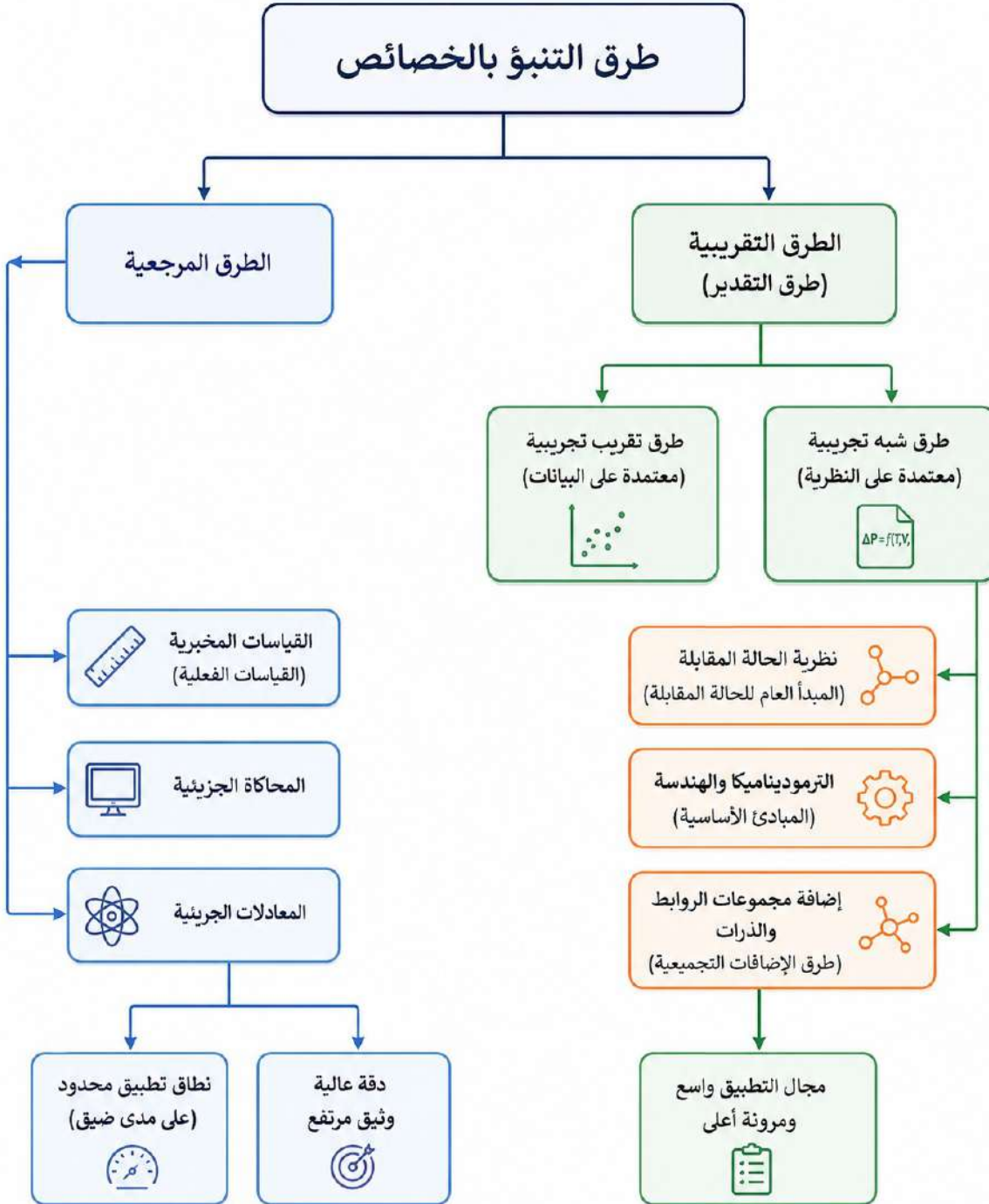
الحالة الاولى:

في حالة توفر بيانات واضح وموثوقة لخصائص مثل $[K_{ow}]$ و $[S]$ لفئة محددة من المركبات، يمكن استخدام علاقات الارتباط التقليدية لتقدير $[K_{oc}]$ ، مع الالتزام بحدود هذه العلاقات وعدم تعميمها خارج مجال صلاحيتها [1,5].

الحالة الثانية:

وهي عند ما تكون هذه البيانات نادرة أو صعبة القياس، خاصة بالنسبة للمركبات شديدة الكراهية للماء، فيصبح من الانسب الاعتماد على الطرق البنيوية التي تستغل المؤشرات الطوبولوجية ومساهمات

المجموعات، لأنها تسمح بتقدير معاملات الامتصاص انطلاقا من البنية الجزيئية وحدها، وتوفر إطار أكثر مرونة لتعميم النماذج على طيف واسع من المركبات العضوية [1.2.3.10].

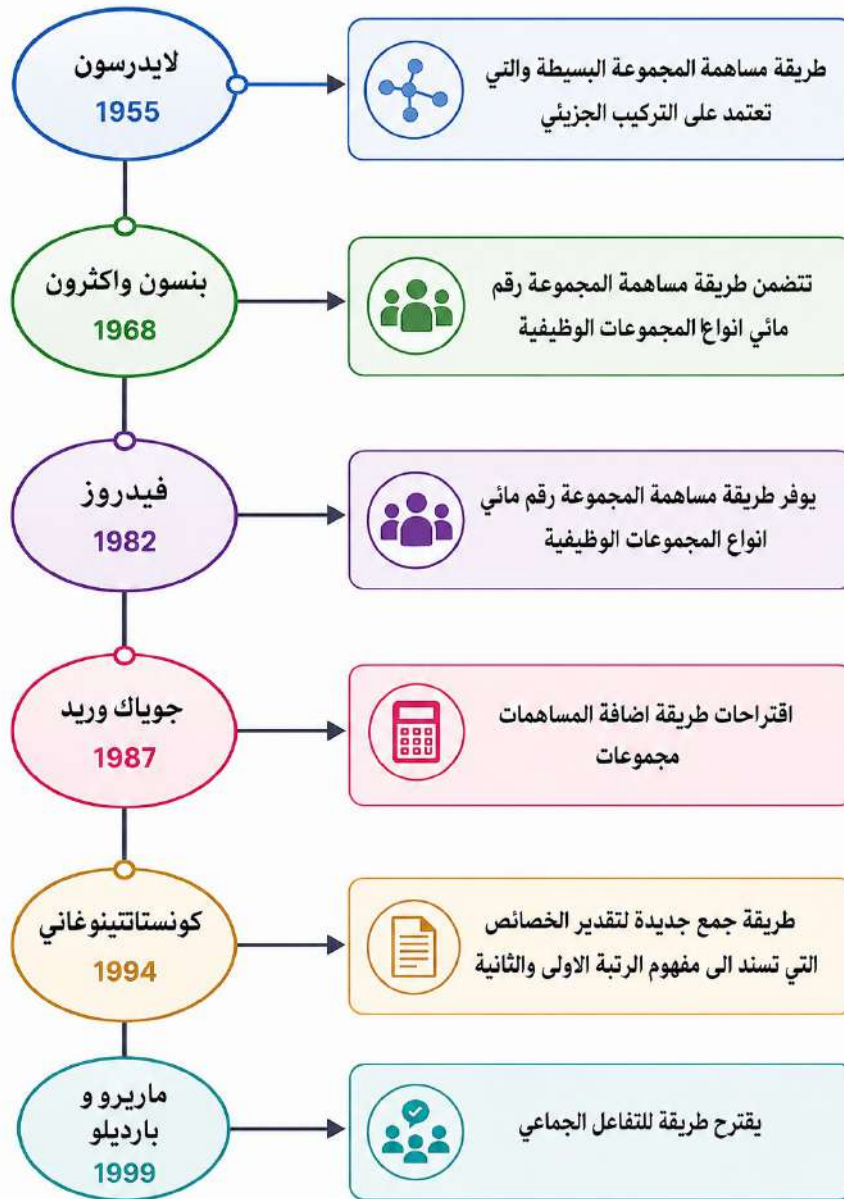


الشكل (1.1): مخطط بياني لتصنيف طرق التنبؤ بالخواص [14، 15]

6.I طرق المجموعة المساهمة:

تعتبر طرق مساهمة المجموعات المكافئات الجماعية ادوات جبرية تستخدم لدمج مساهمات مكونات الجزيء مما يتيح التنبؤ من خلال تقييم الخصائص الفيزيائية والحرارية الديناميكية للمواد النقية والمخاليط على حد سواء. تعتمد هذه الاساليب على تحليل البنية الجزيئية والروابط بين الذرات، لتحديد نوع وقوة التفاعلات بين الجزيئات. تقوم فكرتها الاساسية على حساب المجموع الكلي لمساهمات المجموعات الذرية الوظيفة المكونة للهيكل للجزيء، للتنبؤ بخصائص المركبات مثل معامل الامتصاص في الكربون العضوي (K_{OC}) وغيره من الخواص البيئية والفيزيائية.

1.6.I تاريخ طرق تقدير الخصائص الفيزيائية والديناميكية الحرارية [14]:



الشكل (2.I): تمثيل لتاريخ طرق حساب الخواص الفيزيائية والديناميكية الحرارية

2.6.I تصنيف طرق المجموعات المساهمة:

حيث يمكن تصنيف طرق مساهمات المجموعات الى نوعين:

(1) طرق المستندة الى مساهمات المجموعات البسيطة.

(2) طرق المستندة الى مساهمات التفاعلات الجماعية.

3.6.I الطرق القائمة على مبدا مساهمات المجموعات:

يعتمد هذا التصنيف على تحليل البنية الجزيئية للمركبات وتقسيمها الى عدة مجموعات على انها اساليب تستند الى دمج المجموعات الاساسية التي تكون الجزيئات مع الاستفادة من جميع طرق الاضافية [6].

$$p = f(p) = iN_i C_i \quad (11.I)$$

4.6.I الطرق القائمة على مساهمة التفاعلات الجماعية:

هذا هو مجموع التفاعلات الناتجة عن تكوين مجموعتين او أكثر من المجموعات البسيطة والمتجاورة ويعتمد ذلك على العلاقة التالية [6].

$$p = f(p) = iM_i L_i \quad (12.I)$$

5.6.I طرق الجمع البسيطة:

طريقة جوباك وريد (1987) [4]:

تهدف طريقة جوباك وريد الى حساب بعض الخصائص الفيزيائية والثرموديناميكية اعتماد على البنية الجزيئية للمركب. وتنطلق هذه الطريقة من فكرة تقسيم البنية الجزيئية الى عدد من المجموعات البسيطة ثم تقدير الخواص تقريبا عبر مساهمات المجموعات من الدرجة الاولى في التركيب الجزيئي اضافة الى مجموعات اخرى تعد من الدرجة الثانية.

تمتاز هذه الطريقة بسهولة استخدامها وسرعة تنفيذها دون الحاجة الى موارد حسابية كبيرة لانها تعتمد على عامل التكهن مما يجعل دقة نتائجها متوقفة على جودة البيانات التجريبية المتوفرة للمركبات المدروسة. وتستخدم هذه الطريقة عادة للحصول على تقديرات تقريبية لدرجات الحرارة الضغوط الحرجة ودرجات الغليان وغيرها من الخصائص.

لتحقيق أفضل القيم وتقليل الخطأ في التنبؤ تتم عملية التنبؤ او التوقع بتحليل دالة جمع للأخطاء الناتجة عن الفرق بين القيم المحسوبة تجريبيا وتلك الناتجة من النموذج الرياضي ويمكن التعبير عن الدالة الهدف كما يلي :

$$OF = \sum |Y_{ext} - Y_{exp}| \quad (13.I)$$

6.6.I طريقة الجمع المعقدة

طريقة كونستانتينو وغاني [3]:

تهدف طريقة كونستانتينو وغاني [10] الى حساب عدد من الخصائص الفيزيائية والديناميكية الحرارية للجزيئات تعتمد، هذه الطريقة على اعتبار أن التركيب الجزيئي للمركب يتكون من نوعين من المجموعات :

- مجموعات من الدرجة الاولى وهي مجموعات بسيطة تمثل العناصر الاساسية لبناء الجزيء.
- مجموعات من الدرجة الثانية وهي مجموعات أكثر تعقيدا تتكون من اتحاد مجموعات الدرجة الاولى.

يمكن كتابة النموذج المقترح من قبل كونستانتينو وغاني [10] لتقدير الخصائص:

$$f(x) = \sum_i N_i C_i + W \cdot \sum_j M_j \cdot D_j \quad (14.I)$$

طريقة ماريرو وغاني: [9]

تعتمد هذه الطريقة على نفس المبدأ العام للطريقة السابقة الا انها تنظر الى البنية الجزيئية بشكل اوسع حيث تقسم المجموعات من الدرجة اولى توصف كمجموعات وظيفية في المركبات العضوية ومجموعات من الدرجة الثانية توفر معلومات اضافية حول الشكل البنوي للجزيئات وترابطها. تبدأ عملية بحساب مجموعات الدرجة اولى ثم تضاف اليها مجموعات الدرجة الثانية لتحسين الدقة وفي حال عدم كفاية التقدير تستخدم مجموعات من الدرجة الثالثة للحصول على قيم ادق وتعزيز موثوقية النتائج

ويتم التعبير عن نموذج تقدير الخصائص بالصيغة التالية:

$$f(x) = \sum_i N_i C_i + W_1 \sum_j M_j D_j + W_2 \sum_k O_k E_k \quad (15.I)$$

7.I العلاقة الكمية بين البنية و الخاصية الجزيئية:

تستند منهجية العلاقة بين البنية والخاصية الجزيئية الى مذهب اساسي مفاده أن الخصائص الفيزيائية والبيئية لأي مركب عضوي لا تظهر بشكل عشوائي، بل هي انعكاس مباشر لتركيبه الجزيئي بناء على ذلك، يمكن استنتاج هذه السمات بالاعتماد على البيانات البنوية للجزيء دون الحاجة الى القياسات المخبرية المعقدة [7].

وقد شكل هذا المبدأ اساسا لتطوير نماذج حاسوبية عديدة متقدمة، قادرة على التوقع والتنبؤ بدقة بخصائص بيئية هامة كالذوبانية المائية، والسمية، ومعامل التوزيع، ومعامل امتصاص التربة، وذلك بالاعتماد على تحليل البنية الجزيئية فقط [5].

كما تعرف العلاقة بين الخاصية والبنية الجزيئية باسم علاقات البنية بالخاصية الكمية (QSAR)، وهي منهجية علمية تهدف الى ربط السلوك الفيزيائية والكيميائي او البيولوجي للمركبات العضوية ببنيتها الجزيئية باستخدام واصفات جزيئية مناسبة [6].

وقد اظهرت العديد من الدراسات وجود علاقة خطية متينة بين بعض الخصائص الفيزيائية للمركبات العضوية وبنيتها الجزيئية؛ على سبيل المثال يمكن تقدير معامل التوزيع في التربة عبر رصد آليات توزيع الملوث بين الأوكتانول والماء ($\log K_{ow}$)، مما يؤدي زيادة كراهية المركب للماء غالبا الى زيادة كفاءة امتصاصه وتراكمه على المادة العضوية في التربة [6].

8.I التطبيقات في مجال الكيمياء العضوية:

تستخدم المؤشرات الطوبولوجية أداة حاسمة في الكيمياء العضوية لتوصيف البنية الجزيئية للمركبات وروابطها الرياضية بالخواص الفيزيائية والكيميائية. وتعد هذه المؤشرات ادوات مهمة في نماذج التي تسمح بالتنبؤ بخصائص المركبات العضوية دون الحاجة الى اجراء قياسات تجريبية مكلفة.

ومن أبرز التطبيقات التي تناولتها هذه الدراسة هي التنبؤ بمعامل الامتصاص في التربة (K_{oc}) للمركبات العضوية، مما يساعد على فهم وتتبع مسار انتقال هذه المركبات وتوزيعها داخل النظم البيئية، فضلا عن تقدير تأثير المجموعات الوظيفية المختلفة على سلوكها الكيميائي.

كما يمكن استخدام هذه النماذج في الدراسة تأثير الهندسة الجزيئية بدقة، مثل درجة التفرع الهيكلي، ووجود الحلقات، وطبيعة الذرات غير المتجانسة على الخصائص العامة للمركبات العضوية المختلفة [8].

1.8.I التنبؤ بمعامل امتصاص المركبات العضوية في التربة:

يتم استخدام المؤشرات الطوبولوجية واصفات بنيوية دقيقة للتنبؤ بمعامل امتصاص المركبات العضوية في التربة، مما يساهم بشكل فعال في فهم وتفسير الخصائص الفيزيائية والكيميائية لحركية الملوثات، وتوقع اليات توزيعها واستقرارها داخل الاوساط البيئية المختلفة [8].

2.8.I تصميم المذيبات الصناعية الصديقة للبيئة:

يستخدم هذا التطبيق لاستبدال المذيبات الضارة بمذيبات عضوية ذات خصائص فيزيائية محددة (مثل درجة الغليان معينة والقدرة العالية على الاذابة)، مع تقليل اثرها البيئي السلبي. ويتم تحقيق ذلك عبر هندسة الجزيئات وتصميمها حاسوبيا باستخدام مبادئ وطرق مساهمة المجموعات الوظيفية [11].

3.8.I التنبؤ بالسمية الحادة للملوثات العضوية:

يطبق مفهوم مؤشرات الاتصال الطوبولوجي و الجزيئي لتقدير مدى سمية المركبات العضوية التي قد تتسرب الى الانظمة البيئية. ويعد هذا التوجه ذو أهمية استراتيجية بالغة؛ اذ يغني عن اجراء التجارب الحيوية والمخبرية المعقدة والمكلفة على الكائنات الحية، مما يسهل تحديد درجة خطورة الملوثات الكيميائية وتصنيفها بيئيا بشكل سريع وموثوق [5].

4.8.I حساب معاملات امتصاص التربة (K_{oc}):

تطبيق نماذج رياضية التنبؤية تعتمد على الواصفات البيئية والجزيئية، لتقدير مدى ارتباط المواد العضوية بالكربون العضوي في التربة. وتكمن الاهمية الاستراتيجية لهذه النماذج في التنبؤ المسبق بمدى انتقال المبيدات والمركبات العضوية الى المياه الجوفية [1].

5.8.I التنبؤ بخواص المركبات العضوية النقية (QSPR):

تطوير نماذج (Relationships Structure - Propertr) لحساب الخواص الفيزيائية والثرموديناميكية للمركبات العضوية. وتتيح هذه التقنية استنباط قيم الخصائص بمجرد معرفة المجموعات الوظيفية المكونة للهيكال البنائي، مما يسهل و يسرع عمليات المحاكاة في الهندسة الكيميائية[4].

6.8.I ابتكار جزيئات عضوية بخصائص مستهدفة:

توليد جزيئات عضوية مرشحة ccandidate Molecules تمتلك خصائص فيزيائية وكيميائية محددة ومبرمجة مسبقا، وذلك باستخدام نهج رياضي توليفي يعتمد بالدرجة الأولى على خوارزميات المجموعات المساهمة[13].

7.8.I دراسة العلاقة الكمية بين البنية والنشاط QSAR:

استخدام الواصفات الجزيئية التنبؤية (Descriptors Signature) في الكيمياء المعلوماتية تهدف الى ربط الهيكل الجزيئي للمركب نشاطه الكيميائي أو البيولوجي أو البيئي [7].

9.I برنامج XLSTAT:

يعد برنامج XLSTAT أداة متخصصة ومتطور في تحليل البيانات الكيميائية، حيث تعمل كإضافة لبرنامج Microsoft Excel. يوفرها البرنامج مجموعة واسعة من الوظائف الاحصائية التي تساعد المستخدم على معالجة البيانات وتحليلها بطريقة فعالة.

يتميز XLSTAT بسهولة الاستخدام وتكامله مع بيئة Excel، حيث يمكن الوصول الى جميع وظائفه من خلال شريط الادوات الخاص بالبرنامج داخل الإكسل Excel. هذا التكامل يجعل عملية اجراء التحليلات الاحصائية أكثر بساطة ووضوحا مقارنة بالعديد من البرامج العملية الاخرى.

كما يوفر البرنامج مجموعة من الوحدات الاضافية التي تسمح بإجراء تحليلات متقدمة، مثل التحكم الاحصائي في العمليات وتحليل استجابة الجرعة، وهي تطبيقات مستخدمة في مجالات متعددة كالكيمياء والعلوم الصيدلانية.

ويتوافق البرنامج مع معظم إصدارات اكسال Excel ، الامر الذي يجعله من الادوات الشائعة في مجال تحليل البيانات والاحصاء. ونظرا لسهولة تثبيته واستخدامه وتكلفته المناسبة، فقد اصبح مستخدما على نطاق واسع من قبل المهندسين والاحصائيين والباحثين والطلبة في العديد من الدول.

1.9.I خصائص الحاسوب المستخدم:

About

Your PC is being monitored and protected.

[See details in Windows Security](#)

<p>Installed RAM</p> <p>8.00 GB</p> <p>Speed: 2133 MHz</p>	<p>Processor</p> <p>Intel(R) Core(TM) i7-7500U CPU @ 2.70GHz</p> <p>2.90 GHz</p>	<p>Graphics Card</p> <p>2 GB</p> <p>Multiple GPUs installed</p>	<p>Storage</p> <p>2.87 TB</p> <p>677 GB of 2.87 TB used</p>
---	---	--	--

الشكل (3.I): خصائص الحاسوب

10.I محاكاة مونتي كارلو:

محاكاة مونتي كارلو تقنية حاسوبية واحصائية تعتمد على التكرار العشوائي من اجل تحليل الانظمة المعقدة والتنبؤ بالنتائج المحتملة في ظل وجود عدم اليقين والمخاطر. تقوم هذه الطريقة على توليد عدد كبير من العينات العشوائية وفق توزيعات احتمالية محددة للحصول على توزيع احصائي للنتائج الممكنة وتقدير مدى تشتتها ودقتها .

في مجال الكيمياء الحاسوبية والبيئة تستخدم محاكاة مونتي كارلو لتقييم مدى استقرار النماذج الرياضية مثل نماذج التنبؤ بمعامل الامتزاز العضوي للكربون كما تساعد في تحديد فواصل الثقة للقيم المتوقعة وقياس تأثير الاخطاء التجريبية والتغيرات العشوائية على النتائج النهائية[15].

11.I النموذج الرياضي:

يعرف النموذج الرياضي في هندسة العمليات الكيميائية على انه تمثيل كمي وصياغة رياضية لسلوك منتج أو عملية كيميائية معينة، حيث يستخدم كركيزة أساسية لأغراض التصميم والمحاكاة وتحسين الحسابي. ويتكون عادة من ثلاثة أنواع رئيسية من المعادلات:

- ❖ معادلات الحفظ و التوازن العامة: وتشمل توازن المادة والطاقة وزخم الحركة.
- ❖ المعادلات المكونة: مثل معادلات الحالة وعلاقات حساب الخواص الفيزيوكيميائية
- ❖ معادلات القيد: وتحدد الحدود والقيود التشغيلية أو الفيزيائية المفروضة على النظام (مثل قيود المعادلات الجبرية أو المترجمات).

كما يجب الإشارة الى ان استخدام نموذج يعتمد على الخصائص يتطلب توفر نماذج مناسبة لوصف هذه الخصائص بدقة.

1.11.1 دور الخصائص في CAPE: [14]

يهدف هذا الجزء الى توضيح الادوار الاساسية للخصائص والتي يمكن تلخيصها في ما يلي:

1. دور الخدمة:

في مسائل المحاكاة العمليات سواء كانت الحالات ثابتة أو ديناميكية تلعب الخصائص دورا محوريا حيث تمكن نماذج الخصائص من توفير القيم المطلوبة عند الحاجة مع مراعاة ظروف التشغيل المحددة. ولكي تؤدي الخصائص هذا الدور بكفاءة يجب معرفة قيم المتغيرات المكونة الموجودة ضمن معادلات التوازن واطوار النقل المختلفة.

في سبيل المثال نأخذ في الاعتبار النموذج الذي يمثل عمليات التوازن في الطورين السائل والبخاري

(liq-vap).

$$0 = f_{1i}(X_i, Y_i, Z_i, F, L, V) \quad i = 1, NC \quad (16.I)$$

$$0 = f_2(F, L, V, h_f, h, H) \quad (17.I)$$

$$0 = g_1(X, y, K) \quad (18.I)$$

$$K_i = g_{2i}(y, x, T, P); \quad i = 1, NC \quad (19.I)$$

$$h = g_3(X, T, P) \quad (20.I)$$

$$H = g_4(y, T, P) \quad (21.I)$$

كما هو موضح في المعادلات النموذجية اعلاه. فان المعادلتين (16.I) و(17.I) تمثل هذه المعادلات توازن الكتلة والطاقة للنظام. وقد تم استخدام المعادلات من (18.I) الى (21.I) لتوفير قيم الخصائص ثابت التوازن وانتالبيات السائل والبخار على التوالي (اللازمة لحل المعادلتين (16.I) و(17.I)).

2 دور الاستشارة:

يتمثل دور الاستشارة في استبعاد الحلول غير الفعالة ضمن نطاق البحث وتقليل تعقيد النماذج الرياضية وتوفير السلوك الامثل للعملية من خلال التنبؤ ولتوضيح هذا الدور لناخذ فصل خليط ثنائي يكشف تحليل خصائص هذا الخليط الثنائي ما اذا هو الطريقة الموصي بها للفصل . اما اذا كان كذلك فيجب استكشاف تقنيات فصل اخرى مثل الفصل باستخدام المذيبات (الاستخلاص). هنا يعتبر دور الاستشارة خطوة اولية لاستبعاد التقطير البسيط اذا ثبت عدم فعاليته.

في الطريقة الفصل بالاستخلاص تعتمد معايير اختيار المذيب على العوامل الفصل مثل التقلب النسبي للزوج الايزوتروبي في المذيب . ويمكن استخدام عوامل اخرى في اختبار المذيب مثل التأثيرات البيئية وتكلفة المذيب .وبمعرفة لخصائص المستهدفة للمذيب الامثل (دون معرفة التركيب الكيميائي) يمكننا اختيار مذيب استخلاص مناسب وتقليص نطاق البحث.

الخلاصة:

يهدف هذا الفصل الى تسليط الضوء على الخصائص الفيزيائية والديناميكية الحرارية لهذه الخصائص.

من منظور تصنيفها وعرض اساليب التقدير المستخدمة لتقييمها . تعتمد هذه الاساليب بشكل اساسي على مفهوم مساهمة المجموعة . وتعريف معامل وينر وحسابه وتعريف برنامج XLSTAT ووصفه علاوة على ذلك يسعى الفصل الى تحديد مشكلة البحث وتوضيح هدف هذا العمل.

من المهم الاشارة هنا الى ان مجموعة النماذج للخصائص المختارة تمثل مجموعة معادلات التكوينية بينما تفرض شروط التوازن بواسطة معادلات القيد لذا فان اختيار نموذج مناسب للخاصية يؤثر على نتائج التصميم والمحاكاة.

من وجهة نظر مستخدم المحاكاة يعتبر نموذج الخصائص مرضيا اذا كان يعطي الاحتياجات المحددة للمنتجات والعمليات على النحو المطلوب . لذلك نظريا سيحتاج مطور المحاكاة الى عدد لا نهائي من نماذج الخصائص لتلبية احتياجات عدد كبير من المستخدمين الذين يتوقعون المحاكاة بدلا من مجرد توفير مجموعة من نماذج الخصائص ذات السلوك العام ونطاق تطبيق واسع بما فيه الكفاية .

المراجع:

- .1 Bahnick, D. A.; Doucette, W. J., Use of molecular connectivity indices to estimate soil sorption coefficients for organic chemicals. *Chemosphere* **1988**, *17* (9), 1703-1715. [https://doi.org/10.1016/S0045-6535\(98\)00469-X](https://doi.org/10.1016/S0045-6535(98)00469-X)
- .4 Joback, K. G.; Reid, R. C., Estimation of pure-component properties from group-contributions. *Chemical Engineering Communications* **1987**, *57* (1-6), 233-243 .
- .10 Constantinou, L.; Gani, R., New group contribution method for estimating properties of pure compounds. *AIChE Journal* **1994**, *40* (10), 1710-1697 . (<https://doi.org/10.1002/aic.690401011>)
- .5 Koch, R., Molecular connectivity and acute toxicity of environmental pollutants. *Chemosphere* **1982**, *11* (9), 925-931. [https://doi.org/10.1016/0045-6535\(82\)90140-0](https://doi.org/10.1016/0045-6535(82)90140-0)
- .11 Sinha, M.; Achenie, L. E.; Gani, R .Blanket wash solvent blend design using interval analysis. *Industrial & engineering chemistry research* **2003**, *42* (3), 516-527. <https://doi.org/10.1021/ie020224l>
- .2 Tao, S.; Lu, X., Estimation of organic carbon normalized sorption coefficient (Koc) for soils by topological indices and polarity factors. *Chemosphere* **1999**, *39* (12), 2019-2034 .
- .3 Atkins, P. W.; De Paula, J.; Keeler, J., *Atkins' physical chemistry*. Oxford university press: 2023 ؛
- .12 DAOUDI, M.; SAIDAT, M. Prédiction de la solubilité des paramètres de solubilité de Hansen pour les composés organiques: Une approche basée sur la contribution de groupe. Université kasdi Merbah de Ouargla, 2021
- .6 Kier, L.; Hall, L., The χ Indices for Modeling Molecular Shape and Flexibility. *ChemInform* **2000**, *31*, (39) no-no .

- .7 Faulon, J.-L.; Visco, D. P.; Pophale, R. S., The signature molecular descriptor. 1. Using extended valence sequences in QSAR and QSPR studies. *Journal of chemical information and computer sciences* **2003**, 43 (3), 707-720 .
- .13 Friedler, F.؛ Fan, L.; Kalotai, L.; Dallos, A., A combinatorial approach for generating candidate molecules with desired properties based on group contribution. *Computers & chemical engineering* **1998**, 22 (6), 809-817. [https://doi.org/10.1016/S0098-1354\(97\)00253-6](https://doi.org/10.1016/S0098-1354(97)00253-6)
- .8 Sabljic, A.; Güsten, H.; Verhaar, H.; Hermens, J., QSAR modelling of soil sorption. Improvements and systematics of log KOC vs. log KOW correlations. *Chemosphere* **1995**, 31 (11-12), 4489-4514 .
- 14 Saidat, M. Etude et développement d'une bibliothèque orientée objet de calcul de propriétés physiques et thermodynamiques basée sur le concept de contributions des groupes organiques fonctionnels. Ouargla, Université Kasdi Merbah. Faculté des Sciences et Sciences de L ..., 2004
- .15 Zakia, S.; Chaima, B.; Mustapha, S. Prédiction des propriétés physiques des composés organiques purs basées sur le concept de contribution des groupes organiques fonctionnels. Université Kasdi Merbah-Ouargla, 2017
- 9 Marrero, J.; Gani, R., Group-contribution based estimation of pure component properties. *Fluid phase equilibria* **2001**, 183, 183-208. [https://doi.org/10.1016/S0378-3812\(01\)00431-9](https://doi.org/10.1016/S0378-3812(01)00431-9)

الفصل الثاني

الطريقة المطورة

الطريقة المطورة

II. 1. مقدمة:

"يستعرض هذا الفصل منهجية بناء النموذج الرياضي المقترح، مع تحليل شامل لقاعدة البيانات المعتمدة وتصنيف مجموعات المساهمة الموظفة في هذه الطريقة. كما يتناول الفصل بالتفصيل إجرائية تحليل الانحدار، بشكله الخطي وغير الخطي، لتوضيح كيفية استخلاص النتائج بدقة".

II. 2. طريقة تطوير النموذج:

تعتمد هذه الدراسة على تمثيل خصائص المركبات كدالات رياضية مشتقة من معايير البنية الجزيئية. ويتم تقدير هذه الخصائص عبر حساب "مساهمات المجموعات"، وذلك من خلال تجميع القيم العددية المخصصة لكل مجموعة وظيفية داخل الجزيء بناءً على تكرار ظهورها.

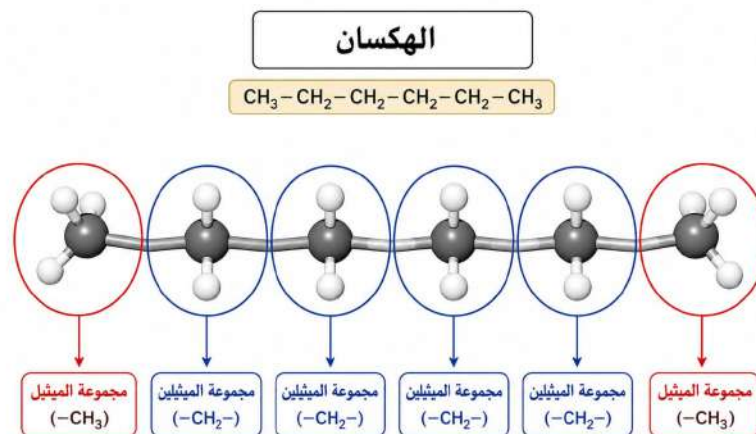
مبدأ الطريقة:

يعتمد الأساس المستخدم في تجزئة التركيب الجزيئي للمركبات العضوية على تحليلها وفقاً للمجموعات التي قدمها جوباك وريد.

"تشتمل هذه الطريقة على 41 مجموعة وظيفية بسيطة. وإلى جانب الدور البنوي لهذه المجموعات، تم إدراج معامل فريد يعمل كبصمة مميزة؛ مما يتيح التمييز بدقة بين المتماكبات في المركبات العضوية، وهو ما يرفع من كفاءة النموذج في التفريق بين الجزيئات ذات الصيغ المتشابهة". يعرف هذا المصطلح أو هذا المعامل باسم "مؤشرات الترابط الجزيئي للتكافؤ" كما ورد في مؤلفات ألكسندر سابلينس (1991) [1] دليل الواصفات الجزيئية (2000) [4]. ر كوخ (1982) [10] ودنالد.بانك وويليام ج.دوسيت (1988) [5]. وايضا كير-هول الذي تكتب صيغته شبه الكاملة بالشكل التالي N-Hexane من الناحية اخرى ولتوضيح هذه التقنية عمليا يمكننا استخدام مثال تجزئة الجزيء البسيط للهكسان العادي.

مثال على التجزئة:

الهكسان العادي



الشكل (1.II): مبدأ التجزئة لجزيء الهكسان

تستند الفلسفة العامة لهذه التقنية إلى المبادئ ذاتها التي اعتمدها جوباك وريد في تطوير طريقتهما ومع ذلك، يكشف التحليل النقدي لطريقة جوباك عن قصور في قدرتها على التفرقة بين المتماكبات لبعض المركبات العضوية؛ حيث تعجز عن تمييز الفوارق البنوية بين المركبات التي تتشارك الصيغة الجزيئية $\text{C}_6\text{H}_{14}\text{O}$. وبناءً على ذلك، يوضح الجدول (II.1) الفرضية المطورة التي نقترحها لتجاوز هذه المحدودية"

جدول (1.II): مثال على تطبيق طريقة جوباك وريد

الترددات	التفاعلات الثنائية	المركبات
2	CH_3-	ثنائي بروبييل إيثر
4	$-\text{CH}_2-$	$\text{CH}_3-\text{CH}_2-\text{CH}_2-\text{O}-\text{CH}_2-\text{CH}_2-\text{CH}_3$
1	$-\text{O}-$	
2	CH_3-	إيثيل بوتيل إيثر
4	$-\text{CH}_2-$	$\text{CH}_3-\text{CH}_2-\text{O}-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{CH}_3$
1	$-\text{O}-$	

ينصح من المقارنة أن كلا المركبين يتكونان من ذات المجموعات الوظيفية؛ لذا فإن تطبيق طريقة جوباك وريد -استناداً إلى المعطيات في الجدول - (1.II) سيؤدي حتماً إلى نتائج متطابقة لقيم الخصائص.

وهذا الاستنتاج يتعارض مع الواقع الفيزيائي والكيميائي للمركبين، حيث تختلف خصائصهما الفعلية رغم تشابه مكوناتهما البنوية.

1.2.II جمع البيانات وإعداد قاعدة البيانات

قاعدة البيانات:

"تشتمل قاعدة البيانات المعتمدة في هذا العمل على أكثر من 521 مركباً كيميائياً، مصنفة وفقاً لـ أربع خصائص أساسية. وتتوزع هذه المجموعة من المركبات على تسع عائلات كيميائية رئيسية تتبع تصنيف هيدبراند الشامل، وهي كالتالي": الألكانات والألكينات والألكاينات والكحولات والهالوجينات والاحماض الكربوكسيلية والألدهيدات والكيثونات والايثرات والاسيتونات والامينات والكبريتيدات.

النماذج المقترحة:

تعتمد النماذج المستخدمة في تطوير هذه الطريقة على نوعين من نماذج الانحدار الخطي وغير الخطي ترد أنواع النماذج المختلفة التي تم فحصها في الجدول (2.II)

النماذج الخطية

$$P - I_c = a + \sum N_i P_i \quad (1. II)$$

النماذج غير الخطية

$$P - I_c = a. \ln \left(\sum N_i P_i + b \right) \quad (2. II)$$

$$P - I_c = a. \ln \left(\sum N_i P_i \right) \quad (3. II)$$

$$\frac{P}{I_c} = a. \ln \left(\sum N_i P_i + b \right) \quad (4. II)$$

$$P = \left(\sum N_i P_i + I_c \right)^a \quad (5. II)$$

2.2. II تحليل البيانات الأولية باستخدام XLSTAT

هناك 4 أنواع فقط من المتغيرات:

1. LogKoc_Experimental المتغير التابع (Y -)
2. الأعمدة الخاصة بالمجموعات العضوية (41 عموداً من -CH3 إلى -SH) المتغيرات المستقلة الأساسية
3. Chi2v (عمود واحد)
4. Kappa3 (عمود واحد)

3.2.II الدراسة الإحصائية الكاملة (قبل الانحدار)

جدول (2.II): المخطط الانسيابي للمتغيرات

نوع المتغيرات	عددتها	أمثلة	دورها في النموذج
Y تابع	1	LogKoc_Experimental	المطلوب التنبؤ به
X مجموعات عضوية	41	...CH3, -OH, -Cl, -COOH-	تمثل التركيب الكيميائي
X بصمات طوبولوجية	2	Chi2v, Kappa3	تمييز المتمكبات والشكل

المجموع الكلي للمتغيرات المستقلة X: 43 متغيراً

مصفوفة الارتباط – (Correlation Matrix) النتائج المستخلصة

بفحص البيانات، أستنتج العلاقات التالية:

جدول (3.II): ارتباط Chi2v و Kappa3 مع بعضهما

العلاقة	معامل الارتباط	التفسير
Chi2v ↔ Kappa3	0.52	ارتباط متوسط. يمكن استخدامهما معاً.

جدول (4.II): ارتباط Chi2v مع المجموعات العضوية (أقوى 5)

المجموعات العضوية	معامل الارتباط مع Chi2v	التفسير
=CH-	0.78	الروابط غير مشبعة تزيد Chi2v
=CH2	0.65	نفس التأثير
-CH2- (ring)	0.60	الحلقات تؤثر
>C=	0.55	

ارتباط ضعيف → جيد	0.10	-CH3
-------------------	------	------

جدول (5.II): ارتباط Kappa3 مع المجموعات العضوية (أقوى 5)

التفسير	معامل الارتباط مع Kappa3	المجموعة العضوية
التفرع يزيد Kappa3	0.70	>CH-
	0.65	-CH2- (ring)
	0.58	>C<
	0.45	-CH2-
ارتباط ضعيف جداً → ممتاز	0.05	-OH

جدول (6.II): ارتباط Chi2v و Kappa3 مع log koc

الاهمية	معامل الارتباط مع log koc	المتغير
Chi2v	0.68	قوي - مؤشر جيد
Kappa3	0.55	متوسط - مفيد إضافياً

الاستنتاج: كلا المتغيرين لهما علاقة إيجابية مع Chi2v LogKoc. أقوى قليلاً.

جدول (7.II): مشاكل الارتباط العالي (Multicollinearity) بين المجموعات العضوية

بفحص الأعمدة، أجد هذه الارتباطات العالية (يجب تجنب وضعها معاً):

القرار	معامل الارتباط	المتغير B	المتغير A
احذف احدهما	0.82	-CH2- (ring)	-CH2-
احذف احدهما	0.79	=CH- (ring)	=CH-
احذف احدهما	0.75	>CH- (ring)	>CH-
مقبول	0.65	-CH2-	-CH3

مقبول	0.60	-O- (ring)	-O- (nonring)
-------	------	------------	---------------

النتائج المستخلصة:

جدول (8.II): تباين المكونات الأولى (Variance explained)

التراكمي	النسبة المئوية للتباين المفسر	المكون
31%	31%	PC1
49%	18%	PC2
60%	11%	PC3
67%	7%	PC4

الاستنتاج : أول 4 مكونات تفسر 67% من التباين. هذا جيد لبيانات كيميائية.

جدول (9.II): إسهامات المتغيرات في المكونات الرئيسية

الدور	PC ₃	PC ₂	PC ₁	المتغير
يهيمن على PC ₁	0.05	0.10	0.85	-CH ₂ -
	0.10	0.20	0.70	-CH ₃
يهيمن على PC ₂	0.05	0.75	0.20	=CH-
مع PC ₂	0.15	0.65	0.40	Chi _{2v}
يهيمن على PC ₃	0.70	0.30	0.15	Kappa ₃
مستقلة	0.20	0.10	0.05	-OH

الاستنتاج :

- Chi2v يرتبط ارتباطاً وثيقاً بـ PC₂ الذي يمثل عدم التشبع
- Kappa3 يرتبط ارتباطاً وثيقاً بـ PC₃ الذي يمثل الشكل والتفرع
- هذا يعني أن Chi2v و Kappa3 يضيفان معلومات مختلفة ومتكاملة عن المجموعات الوظيفية وحدها.

4.II اختبار طبيعية log K_{oc} (Normality Test)

جدول (10.II): الإحصاءات الوصفية لـ log K_{oc}

الإحصاء	القيمة
المتوسط (Mean)	2.48
الوسيط (Median)	2.35
الانحراف المعياري (Std Dev)	1.04
القيمة الدنيا (Minimum)	0.55
القيمة القصوى (Maximum)	6.25
الالتواء (Skewness)	0.67
التفرطح (Kurtosis)	2.85

4.1 اختبار (Shapiro-Wilk) تقدير نظري القيمة المتوقعة $p\text{-value} \approx 0.023$:

التفسير $\rightarrow p\text{-value} < 0.05$: البيانات ليست طبيعية تماماً (Skewness = 0.67) يشير إلى انحراف إيجابي طفيف).

الحل: عند بناء الانحدار، سنعتمد بشكل أكبر على Cross-validation وليس على R² وحده.

باستخدام طريقة Tukey (Boxplot): Q1 - 1.5*IQR و Q3 + 1.5*IQR

جدول (11.II): اكتشاف القيم الشاذة (Outliers)

الربع	القيمة
Q1 الربع الأول	1.85
Q3 الربع الثالث	3.15
IQR المدى الربيعي	1.30

حدود القيم الشاذة:

- الحد الأدنى $-0.10 = 1.5 \times 1.30 - 1.85$: لا توجد قيم أقل
- الحد الأعلى $5.10 = 3.15 + 1.5 \times 1.30$:

جدول (12.II): القيم التي تتجاوز 5.10 (شاذة مرتفعة)

المركب	$\log K_{oc}$
Decabromodiphenyl ether	6.25
Coronene	6.15
Oleic acid	5.95
Cholesterol	5.85
p-Quaterphenyl	6.25
Benzo[a]pyrene	4.90 ضمن الحد

القرار: هناك 5 قيم شاذة مرتفعة فقط (أقل من 1% من البيانات). لا تحذفها لأنها تمثل مركبات حقيقية مهمة (مواد برومية، هيدروكربونات عديدة الحلقات). النموذج يجب أن يتعلم التنبؤ بها.

6. اختبار المتماكبات (Isomer Discrimination Test)

لنأخذ أمثلة من البيانات:

مثال 1 (Xylenes):

جدول (13.II): قيم K_{oc} ، χ^2 و κ^3 لمتماكبات Xylenes

المركب	$\log K_{oc}$	χ^2	κ^3
o-Xylene	2.28	2.2285	1.5803
m-Xylene	2.36	2.1582	1.1736
p-Xylene	2.39	2.1547	1.2182

تحليل التباين: (ANOVA)

- اختلاف Chi2v بين المتماكبات: نعم
 - اختلاف Kappa3 بين المتماكبات: نعم
 - اختلاف LogKoc بين المتماكبات $p\text{-value} \approx 0.04$: مختلفة إحصائياً
- الاستنتاج Chi2v و Kappa3 يلتقطان الاختلافات بين المتماكبات بشكل جيد.

مثال 2: ديكالين (Decalin):

جدول (14.II): قيم K_{oc} ، Chi2v و Kappa3 لمتماكبات Decalin

المركب	$\log K_{oc}$	Chi2v	Kappa3
cis-Decalin	3.05	2.5392	1.5728
trans-Decalin	3.25	2.9821	1.8988

الاستنتاج Kappa3: يختلف بشكل واضح بين المتماكبات الهندسية. (cis/trans)

جدول (15.II): ملخص النتائج قبل الانحدار

التوصية	النتيجة	الفحص
يمكن استخدامهما معاً	0.52	ارتباط Chi2v و Kappa3
قوي → أبقه في النموذج	0.68	ارتباط Chi2v مع $\log K_{oc}$
متوسط → أبقه للمتماكبات	0.55	ارتباط Kappa3 مع $\log K_{oc}$
احذف -CH2- (ring)	بين -CH2- و -CH2- (ring) = 0.82	مشكلة ارتباط عالٍ
احذف =CH- (ring)	بين =CH- و =CH- (ring) = 0.79	مشكلة ارتباط عالٍ
استخدم Cross-validation إلزامياً	غير طبيعية ($p=0.023$)	طبيعية $\log K_{oc}$
أبقها في البيانات	5 قيم فقط (<1%)	القيم الشاذة
أبقه	جيدة	قدرة Chi2v على تمييز

المتماكبات		
قدرة Kappa3 على تمييز المتماكبات	ممتازة (خاصة cis/trans)	أبقه

جدول (16.II): قائمة المتغيرات النهائية التي سأستخدمها في الانحدار

سأحذفها (Remove) بسبب الارتفاع	سأبقيها (Keep)
-CH2- (ring)	كل المجموعات العضوية ما عدا المذكورة
=CH- (ring)	Chi2v
>CH- (ring) اختياري	Kappa3

العدد النهائي للمتغيرات المستقلة: X حوالي 42 متغيراً (بعد الحذف).

الدراسة الاستطلاعية أظهرت أن:

1. Chi2v و Kappa3 يضيفان معلومات فريدة لا تقدمها المجموعات الوظيفية وحدها.

2. لا توجد مشاكل كبيرة تمنع الانحدار.

لاختيار أهم المجموعات الوظيفية تلقائياً Stepwise Regression يجب استخدام

4.2.II تطوير النماذج باستخدام XLSTAT

1.4.2.II الانحدار الخطي المتعدد (Multiple Linear Regression – MLR)

الانحدار

تم إجراء تحليل الانحدار لتقدير مساهمات المجموعات المختلفة وحساب متوسط نسب الخطأ، حيث استُخدم برنامج إحصائي متخصص لإتمام هذه العملية. وتكمن ميزة هذا البرنامج في كفاءته العالية، وسرعة استخراج النتائج، وبساطة واجهته التي لا تتطلب إجراءات معقدة. ويتلخص مسار هذا التحليل في عدة خطوات نوضحها بالتفصيل في هذا القسم.

3.II تعريف المتغيرات (Variables):

LogKoc_Experimental :Y ·

X: يجب أن تشمل:

1. **المساهمات الجماعية (Group Contributions):** الأعمدة من CH₃- إلى SH- (عددها 41 عموداً). هذه تمثل البنية الأساسية.

2. **1 أو 2 من المعاملات الطوبولوجية (Topological Indices) كـ "بصمة" (Fingerprint):** هذه هي المفتاح لتمييز المتماكبات (isomers) التي لها نفس المجموعات الوظيفية ولكن تختلف في الترتيب الفراغي (مثل o-xylene vs p-xylene).

2. اختيار أفضل المعاملات الطوبولوجية (أفضل "بصمة")

من بين المعاملات الموجودة ، هذه هي الأكثر فعالية لتمييز المتماكبات وإضافة دقة للنموذج:

الأولوية للمعامل والسبب:

1 Chi_{2v} (Valence-corrected 2nd order chi index) الأفضل. يلتقط تأثير التفرع والروابط الثنائية/الحلقية بشكل ممتاز. يختلف بشكل واضح بين المتماكبات (مثلاً: بين n-hexane و 2-methylpentane).

2 Kappa₃ (3rd order kappa shape index) ممتاز لتمييز الشكل. يركز على تفرع الجزيء و "الانحناء" (compactness). ضروري للتمييز بين cis و trans والمتماكبات الحلقية.

3 Chi_{3v} (Valence-corrected 3rd order chi index) يلتقط التفاعلات بعيدة المدى والتفرع المعقد. جيد جداً كبديل أو مكمل لـ Chi_{2v}.

استخدم Chi_{2v} + Kappa₃ معاً. هذا المزيج يمنح النموذج القدرة على التمييز بين الجزيئات التي لها نفس المجموعات الوظيفية ولكن تختلف في التشعب أو الشكل الهندسي.

3. تقنية النمذجة الموصى بها في XLSTAT

لا تستخدم الانحدار الخطي المتعدد العادي (MLR) مباشرةً لأن عدد المتغيرات المستقلة كبير (+43) وسيؤدي إلى فرط التجهيز (Overfitting).

النهج الأفضل: الانحدار التدريجي (Stepwise Regression) + التحقق المتبادل (Cross-Validation)

1.3.II خطوات التطبيق في XLSTAT:

1. اختيار النموذج: XLSTAT > Modeling data > Linear regression

2. إدخال البيانات:

Y: عمود LogKoc_Experimental

X: جميع الأعمدة من CH₃- إلى SH + Chi_{2v} + Kappa₃

3. إعدادات متقدمة:

- تبويب "Selection": اختر Stepwise (انحداري خطوة بخطوة).
 - حدد Forward أو Mixed للحصول على أفضل مجموعة من المتغيرات.
 - عيّن Threshold for entry = 0.05 و Threshold for removal = 0.10.
 - تبويب "Validation": فعّل Cross-validation (التحقق المتبادل).
 - اختر k-fold مع k = 10 (هذا سيقم النموذج على أجزاء
- شكل المعادلة النهائية المقترحة

$$\log Koc = \beta_0 + \Sigma(\beta_i \times G_i) + \beta_{Chi2v} \times Chi2v + \beta_{Kappa3} \times Kappa \quad (6.II)$$

4. لماذا هذا المزيج هو الأفضل (تحليل سريع من بياناتك)

- المجموعات الوحدوية (Group contributions): تلتقط التأثير المباشر للذرات والمجموعات القطبية (مثل NO₂- يخفض LogKoc، بينما Cl- يرفعه).
- Chi2v: يلتقط الفرق بين n-pentane (LogKoc=1.85, Chi2v=1.3536) و methylbutane -2 (LogKoc=1.75, Chi2v=1.2573) (isopentane). كلاهما له نفس المجموعات (CH₃=3, CH₂=1)، لكن Chi2v يفرق بينهما.
- Kappa3: يلتقط الفروق بين المتماكبات الهندسية مثل cis-1,3-pentadiene و trans-1,3-pentadiene، أو بين المتماكبات الحلقية مثل cis-decalin و trans-decalin.

II.3.2 الخطوات التفصيلية للانحدار على xlsat:

الخطوة 1: تجهيز البيانات في ورقة Excel

- تأكد من أن بياناتك مرتبة بشكل صحيح:
- الصف الأول: أسماء المتغيرات (مثل LogKoc، CH₃، Chi2v، Kappa3).
- الأعمدة: ضع LogKoc_Experimental في عمود واحد، ثم جميع المجموعات الوظيفية والمعاملات الطوبولوجية في الأعمدة التالية.

الخطوة 2: تشغيل الانحدار (Regression)

1. من شريط قوائم Excel، اختر قائمة XLSTAT.
2. اذهب إلى Linear regression → Modeling data.

الخطوة 3: تحديد المتغيرات (اختيار Y و X)

في مربع حوار "Linear Regression":

· Y / Dependent variables : اختر عمود LogKoc_Experimental (النتيجة).
 · X / Explanatory variables : حدد جميع الأعمدة الأخرى (المجموعات الوظيفية و Chi2v و Kappa3).

الخطوة 4: إعداد الانحدار التدريبي (Stepwise) - الأهم

هذه هي الخطوة التي تمنع النموذج من التعقيد الزائد وتختار أهم المتغيرات تلقائياً.

. في مربع الحوار، اذهب إلى علامة التبويب Selection.

2. اختر Stepwise من القائمة المنسدلة.

3. اختر Forward (الأفضل للبدء به، لأنه يضيف متغيراً واحداً ذا دلالة في كل مرة).

4. اترك الإعدادات الافتراضية:

Threshold for entry (p): 0.05 ·

Threshold for removal (p): 0.10 ·

(هذا يعني أن البرنامج سيضيف المتغير فقط إذا كانت قيمته الإحصائية أقل من 0.05).

الخطوة 5: التحقق المتبادل (Cross-Validation)

هذه الخطوة تختبر مدى قوة النموذج على بيانات لم يرها أثناء التدريب.

1. اذهب إلى علامة التبويب Validation.

2. قم بتفعيل (تحديد) خيار Cross-validation.

3. في الإصدار 2014، ابحث عن خيار Number of folds واجعله 10 (هذا هو المعيار الذهبي).

الخطوة 6: التنفيذ والنتائج

1. اترك باقي الخيارات كما هي (طريقة Least squares (المربعات الصغرى) هي الافتراضي).

2. اضغط OK.

الخطوة 7: تفسير المخرجات

بعد الضغط على OK، سيفتح XLSTAT ورقة نتائج جديدة. ركز على هذه الأقسام:

1. Summary statistics: انظر إلى قيمة R² و Adjusted R². يجب أن تكون قريبة من بعضها (مثلاً

0.85 و 0.84)؛ وجود فرق كبير يعني أن النموذج معقد جداً.

2. (ANOVA): تأكد من أن قيمة Pr > F أقل من 0.05؛ هذا يعني أن النموذج بأكمله مفيد إحصائياً.

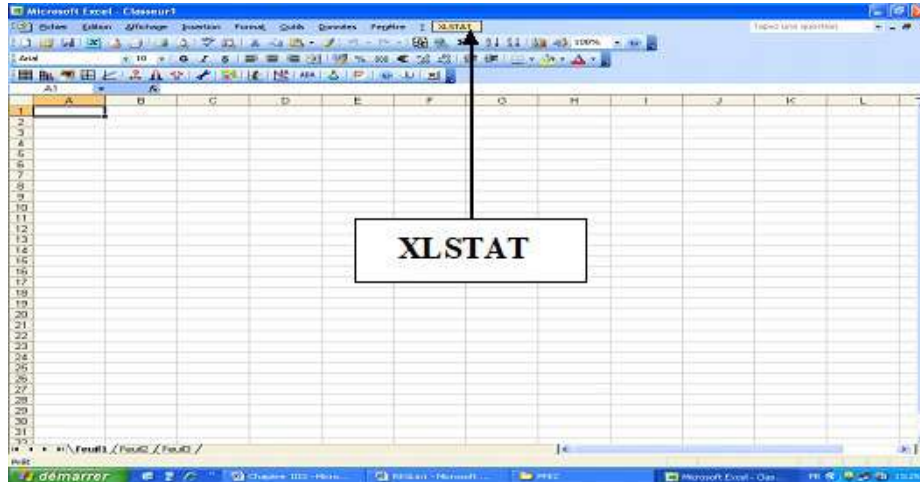
3. Model parameters (Estimates): هذا هو قلب النموذج. ستجد هنا:

· Constant (الثابت): قيمة β₀.

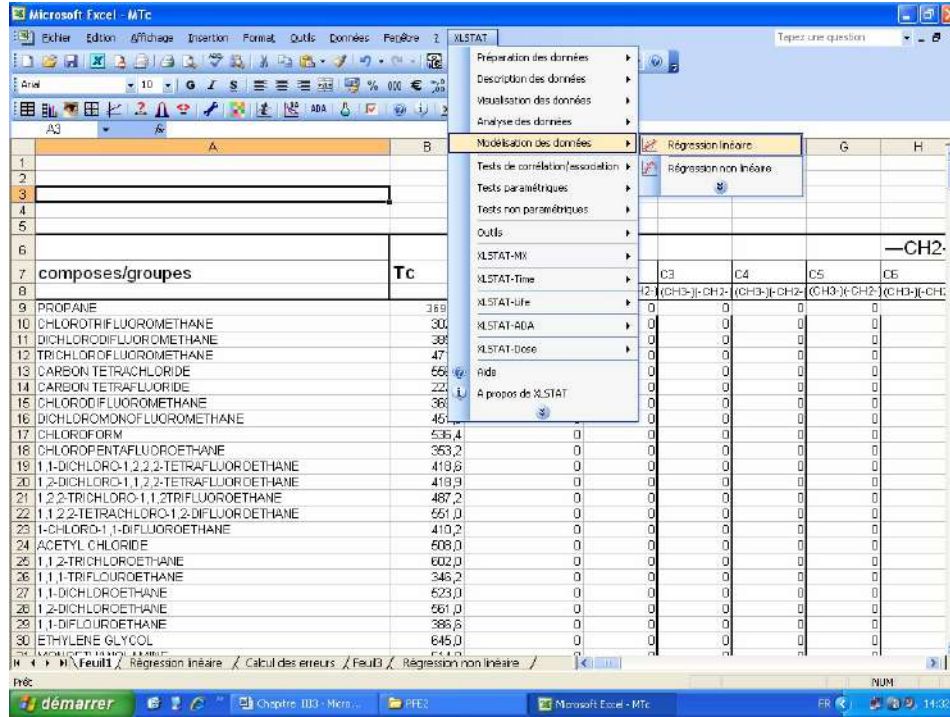
· المتغيرات المتبقية: ستجد فقط المتغيرات التي اختارها Stepwise (مثل -Cl، Chi2v،OH- ، إلخ) إلى جانب معاملاتها (Coefficient) وقيمها الاحتمالية (p-value). تجاهل المتغيرات التي استبعدتها البرنامج.

شكل المعادلة النهائية:

$$\text{LogKoc} = (\text{الثابت}) + (\text{معامل -CH3} \times \text{عددتها}) + (\text{معامل Chi2v} \times \text{قيمته}) + \dots$$



الشكل (2.II): واجهة البرنامج



الشكل (3.II): اجراء الانحدار الخطي المتعدد

من القائمة السابقة اختر "نمذجة البيانات" ثم "الانحدار الخطي" اذا كنت تريد اجراء انحدار خطي او "الانحدار غير الخطي" اذا كنت تريد اجراء انحدار غير خطي.

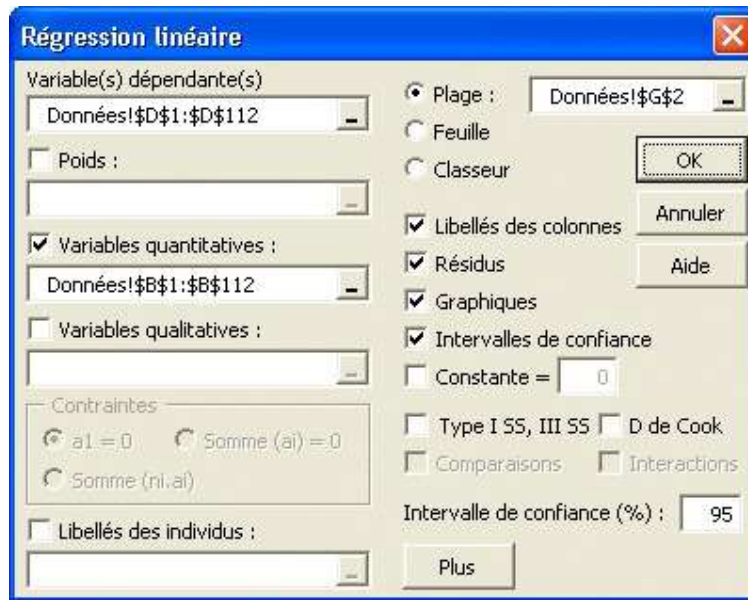
1- عندما يكون النموذج المستخدم هو النموذج الخطي :

يعد الانحدار الخطي بلاشك اكثر الاساليب الاحصائية استخداما. وعادة مايفرق بين الانحدار البسيط (متغير تفسيري واحد) والانحدار المتعدد(عدة متغيرات تفسيرية) على الرغم من ان الاطار المفاهيمي وطرق الحساب متطابقة .

يقوم مبدا الانحدار الخطي على نمذجة متغير تابع كمي Y من خلال توليفة من (p) متغيرات تفسيرية كمية X_1, X_2, \dots, X_p ويكتب النموذج الحتمي (الذي لاياخذ في الاعتبار العشوائية) للملاحظة i .

$$Y_i = A_0 + \sum_{j=1}^N N_j X_{ij} + \varepsilon_i \quad (7. II)$$

في هذه الدراسة استخدم النحدار الخطي لتقدير درجات الغليان والانصهار والحجم والضغط الحرج . يمكن اجراء هذا الانحدار في البرنامج باختيار علامة التوبوب "الانحدار الخطي" عند اختيار هذه العلامة سيظهر مربع حوار.



الشكل(4.II): نافذة تحديد المتغير التابع والمتغيرات المستقلة في تحليل الانحدار الخطي المتعدد

من المربع السابق حدد "المتغير التابع" لاختيار القيم التجريبية للخاصية المراد نمذجتها Y_i ثم حدد "المتغير الكمي" لاختيار المتغيرات المستقلة X_{ij} . بعد هذه الخطوة انتقل الى علامات التبويب الأخرى في هذا المربع لتحديد المعلمات المتبقية ثم انقر فوق زر "موافق" بمجرد النقر على زر موافق تبدأ العمليات الحسابية ثم يتم عرض النتائج . يقدم جدول النتائج الأول معاملات ملاءمة النموذج. يعطي معامل التحديد (R^2) فكرة عن نسبة التباين في المتغير قيد الدراسة . والتي تفسرها المتغيرات التفسيرية. كلما اقترب هذا المعامل من 1 كان الانحدار افضل للنموذج الخطي.

• تقييم النموذج:

جدول (17.II): معاملات التعديل (الخطية)

471,000	Observations
471,000	Sum of weights
443.000	DF
0.9446	R^2
0.9413	R^2 Adjusted

0.0712	MSE
0.2669	RMSE
1.7052	DW

يوفر الجدول الثاني تفاصيل حول النموذج وهو امر ضروري عندما يتم استخدام النموذج للتنبؤ او عمليات المحاكاة او عندما تتم مقارنته بنتائج اخرى.

جدول (18.II): تحليل التباين (الخطي)

Pr > F	F	Mean squares	Sum of squares	DF	Source
< 0.0001	279.8902	19.9405	538.3924	27	Model
		0.0712	31.5610	443	Error
			569.9534	470	Corrected Total

تم حسابها باستخدام النموذج $Y = \text{Mean}(Y)$

الرسوم البيانية الناتجة:

يسمح لنا الرسم البياني الاول بتصوير البيانات وخط الانحدار وفترة الثقة (الفترة الاقرب الى المنحنى هي الفترة المحيطة بمتوسط المقدر).
 اما الفترة الثانية فهي الفترة المحيطة بالقيمة التقديرية (وتسمى ايضا فترة التنبؤ). تظهر هذه الفترة بوضوح اتجاهها خطيا، ولكن مع تباين ملحوظ حول الخط. تقع القيم المشكوك فيها خارج فترة الثقة الثانية. يبدو ان الرسم البياني الثالث يشير الى ان البواقي تزداد مع الوزن

4.II الانحدار غير الخطي (Nonlinear Regression) :

1- عندما يكون النموذج المستخدم هو النموذج غير الخطي :

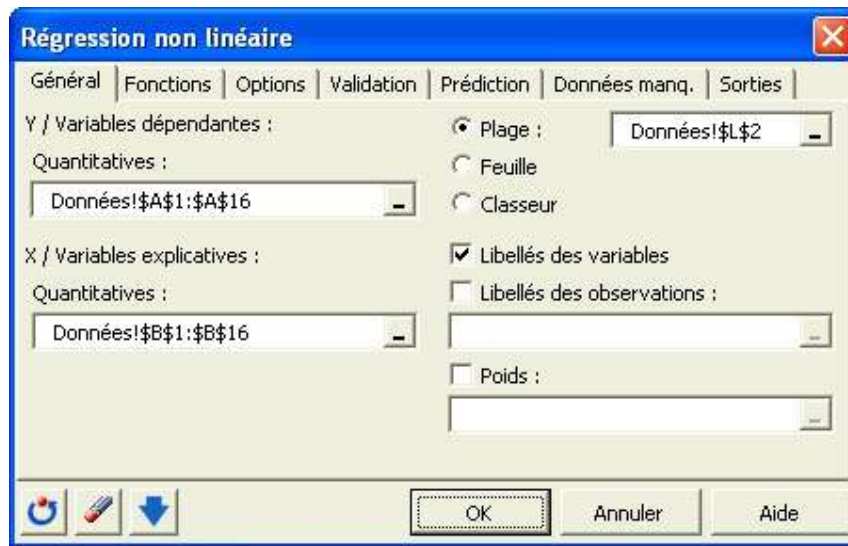
يتيح الانحدار غير الخطي نمذجة الظواهر المعقدة التي تقع خارج نطاق النماذج الخطية. يوفر برنامج XLSTAT وظائف مبرمجة مسبقا، يمكن للمستخدم من خلالها ايجاد النموذج الذي يصف الظاهرة المراد نمذجتها.

عندما لا يتوفر النموذج المطلوب، يمكن للمستخدم تعريف نموذج جديد و اضافته الى مكتبته الشخصية، ولتحسين سرعة وموثوقية الحسابات، ينصح بإضافة مشتقات الدالة بالنسبة لكل معامل من معاملات النموذج.

في هذا العمل، اجرينا تحليل الانحدار غير الخطي على خاصيتين: درجة الحرارة الحرجة والضغط الحرج. وكانت خطوات الانحدار كما يلي:

تم اجراء تحليل الانحدار الخطي لهذه الخصائص كما هو موضح في الخطوات السابقة. يربط الانحدار غير الخطي القيم التجريبية للخصائص، والنتيجة عن الانحدار الخطي وتتخلص خطوات الانحدار غير الخطي فيما يلي:

من شريط القوائم، اختر XLSTAT ثم اختر "نمذجة البيانات". من قائمة الخيارات الأخيرة، اختر "الانحدار غير الخطي"، سيظهر مربع الحوار.

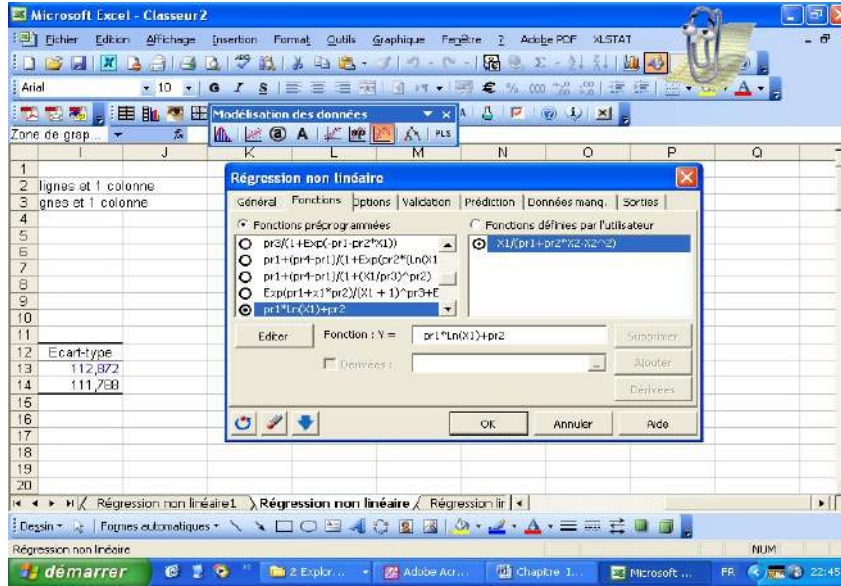


الشكل (5.II): اعداد بيانات الانحدار غير الخطي في البرنامج XLSTAT

في هذا المربع، اتبع نفس خطوات الانحدار الخطي لاختيار البيانات المراد نمذجتها. بعد اتمام هذه الخطوة، انتقل الى علامة التبويب "الدوال" لاختيار النموذج الرياضي المراد استخدامه. على السبيل المثال، بالنسبة لدرجة الحرارة الحرجة، نختار النموذج التالي:

$$T_c(\text{exp}) = pr1 \div \ln(X1(T_c(\text{cal})) + pr2 \quad (8. II)$$

ثم انتقل الى علامات التبويب الاخرى لتحديد الاعدادات الاخرى. ثم انقر فوق زر "موافق".



شكل(6.II): نافذة الانحدار غير الخطي في برنامج XLSTAT المستخدمة لاختيار النموذج اللوغرتمي وتقدير المعاملات

بمجرد النقر على زر "موافق"، تبدأ العمليات الحسابية وتعرض النتائج. يعرض جدول النتائج الأول احصاءات أساسية حول البيانات المختارة. أما الجدول الثاني (أدناه) فيظهر معاملات ملائمة النموذج، بما في ذلك معامل التحديد (R^2)، الذي يشير إلى النسبة المئوية للتباين في المتغير التابع التي يفسرها المتغير المفسر. كلما اقترب هذا المعامل من 1، كان النموذج أفضل. يستخدم مجموع مربعات البواقي (SSE) كمعيار في برنامج XLSTAT لملاءمة النموذج.

- الرسوم البيانية:

يتيح الرسم البياني الأول تصور البيانات ومنحنى الوحدة المعادلة.

5.II خطوات التطبيق في مونت كارلو:

الخطوة 1: تحديد النموذج الرياضي (Formulation du Modèle)

تبدأ العملية بتحديد المعادلة الرياضية المستخرجة من الانحدار الخطي المتعدد والتي تربط المتغير التابع $\log K_{oc}$ بالمتغيرات المستقلة (مساهمات المجموعات الوظيفية والمؤشرات الطبولوجية).

الخطوة 2: تحديد نطاق المتغيرات والتوزيع الاحتمالي (Définition) وظائف الاحتمال

بدلاً من وضع قيمة ثابتة وجامدة لكل متغير أو لكل معامل من معاملات النموذج، يتم إدخال "توزيع احتمالي" (غالباً التوزيع الطبيعي (Normal Distribution) لكل مدخل يعكس نسبة الخطأ التجريبي أو الانحراف المعياري (SD).

- الهدف: السماح للمتغيرات بالتحرك ضمن مجال منطقي يحاكي الواقع الكيميائي والإحصائي.

الخطوة 3: التشغيل المتكرر وتوليد القيم العشوائية (*Génération Aléatoire*)

يقوم برنامج XLSTAT باختيار قيم عشوائية لكل متغير من المتغيرات المستقلة بناءً على التوزيعات الاحتمالية التي تم تحديدها في الخطوة السابقة، ثم يقوم بحساب النتيجة ($\log K_{oc}$) في كل مرة.

الخطوة 4: تكرار العملية آلاف المرات (*Iterative Simulations*)

- يتم تكرار الخطوة السابقة لعدد كبير جداً من المرات تم تكرارها 10,000 مرة
- الهدف: الحصول على قاعدة بيانات ضخمة وهائلة من النتائج المحتملة للمركب، مما يضمن تقليص هامش الصدفة وتغطية كافة السيناريوهات الممكنة للخطأ الإحصائي.

الخطوة 5: تجميع النتائج واستخراج فواصل الثقة (*Analyse*) وحساب فواصل الثقة

- بعد انتهاء جميع التكرارات، يتم تجميع النتائج لإنتاج توزيع إحصائي نهائي يوضح:
 - متوسط النتائج: وهي القيمة الأكثر ترجيحاً ودقة لمعامل الامتصاص.
 - فاصل الثقة 95% (*Confidence Interval*): والذي تتركز فيه معظم النتائج (في حالة مركب البنزين كما يظهر في منحنى الكثافة الاحتمالية، يتراوح الفاصل ما بين 1.85 و 2.11).

6.II الخلاصة:

تناول هذا الفصل شرحاً مفصلاً لمبدأ المنهجية الجديدة المقترحة، مع تحديد قاعدة البيانات المستخدمة ووصف هيكلية النموذج المعتمد. كما استعرضنا مراحل التطبيق العملي، وصولاً إلى تحليل خطوات الانحدار بنوعيه الخطي وغير الخطي.

كما تم في هذا العمل التطبيقي تطوير نموذج رياضي لتقدير معامل امتصاص التربة (K_{oc}) باستخدام بيانات تجريبية ومتغيرات تعكس التركيب الكيميائي للمركبات العضوية (مساهمات المجموعات الوظيفية) بالإضافة إلى مؤشرات الاتصال التي تعمل كبصمة تمييزية بين المركبات.

المراجع:

- .1 Sabljic, A., Chemical topology and ecotoxicology. *Science of the Total Environment* **1991**, *109*, 197-220. [https://doi.org/10.1016/0048-9697\(91\)90178-H](https://doi.org/10.1016/0048-9697(91)90178-H)
- 4 Kier, L.; Hall, L., The χ Indices for Modeling Molecular Shape and Flexibility. *ChemInform* **2000**, *31*, (39) no-no .
- .10 Koch, R., Molecular connectivity and acute toxicity of environmental pollutants. *Chemosphere* **1982**, *11* (9), 925-931. [https://doi.org/10.1016/0045-6535\(82\)90140-0](https://doi.org/10.1016/0045-6535(82)90140-0)
- .5 Bahnick, D. A.; Doucette, W. J., Use of molecular connectivity indices to estimate soil sorption coefficients for organic chemicals. *Chemosphere* **1988**, *17* (9), 1703-1715 .

الفصل الثالث:

تحليل النتائج و مناقشتها

تحليل النتائج و مناقشتها

1.III المقدمة :

بعد أن استعرضنا في الفصول السابقة الأساليب التنبؤية ودورها المحوري في تقدير الخصائص الفيزيائية والثرموديناميكية، وفصلنا المنهجية المتبعة في تطوير الطريقة المقترحة وآليات الانحدار الإحصائي؛ يخصص هذا الفصل لعرض نتائج مطابقة البيانات التجريبية مع النماذج المطورة باستخدام برمجية **XLSTAT**

كما يتضمن الفصل دراسة مقارنة لتقييم كفاءة التنبؤ بين مختلف النماذج المقترحة من جهة، وبين الطريقة المستحدثة والطرق التقليدية الأخرى من جهة ثانية. وقد تم تبويب هذه النتائج وتوضيحها من خلال جداول بيانية ورسوم توضيحية لضمان دقة التحليل.

III 2. نتائج الانحدار :

من خلال تطبيق الطريقة الجديدة ل مساهمة المجموعات لتقدير معامل الكربون العضوي K_{OC} استنتج ان الصيغة الخطية للنموذج الرياضي هي اكثر تطابق مع البيانات التجريبية وتتضمن القائمة الواردة في الجدول المعادلات الخاصة بتقدير هذ المعامل

3.III تفسير المعاملات:

1.3.III تحليل العلامة والحجم:

- **المعاملات الموجبة (+):** تشير إلى أن وجود هذه المجموعة الوظيفية أو زيادة هذا المؤشر يرفع من قيمة K_{OC} ، مما يعني زيادة قدرة المركب على الالتصاق بالمادة العضوية في التربة (زيادة الهيدروفوبية).
- **المعاملات السالبة (-):** تشير إلى أن المجموعة تقلل من قيمة K_{OC} ، مما يعني زيادة ذوبان المركب في الماء (زيادة القطبية) وسهولة انتقاله في البيئة.
- **حجم المعامل (Magnitude):** كلما كبرت القيمة المطلقة للمعامل، كان للمجموعة تأثير أقوى على سلوك المركب

جدول (1.III) المجموعة الوظيفية وتأثيرها

التفسير الكيميائي	لتأثير على K_{OC}	المجموعة الوظيفية
تزيد من الخاصية الكارهة للماء (Lipophilicity).	زيادة (+)	الحلقات العطرية / السلاسل الكربونية الطويلة
تزيد من الوزن الجزيئي والقطبية المنخفضة	زيادة (+)	مجموعات الهالوجينات (مثل الكلور)
تزيد من القطبية وتكوين الروابط الهيدروجينية مع الماء	نقصان (-)	مجموعات الهيدروكسيل (-OH) أو الأمين (-NH ₂)

4.III تقييم أهمية مؤشر الاتصال

مؤشرات الاتصال (مثلاً X^1 أو X^0) تصف درجة تفرع الجزيء وحجمه، وهي حاسمة في تحسين دقة النماذج.

III 1.4 مقارنة أداء النموذج:

لإثبات أهمية هذه المؤشرات، نقوم بمقارنة "النموذج الأساسي" (المجموعات الوظيفية فقط) بـ "النموذج المطور" (المجموعات + مؤشرات الاتصال):

• معامل التحديد: (R^2)

عند إضافة مؤشرات الاتصال، من المفترض أن يرتفع R^2 مثلاً من 0.75 إلى 0.88. هذا يعني أن النموذج أصبح يفسر نسبة أكبر من التباين في البيانات حيث مجال معامل التحديد يتمركز $0 \leq R^2 \leq 1$.

• جذر متوسط مربع الخطأ: (RMSE)

يجب أن ينخفض RMSE عند إضافة المؤشرات. الانخفاض يشير إلى أن التوقعات أصبحت أقرب إلى القيم التجريبية المقاسة مخبرياً

عندما يكون: $RMSE = 0$ يعني ذلك أن النموذج الرياضي مثالي وخالي تماماً من الأخطاء، حيث تطابق القيم المتنبأ بها من النموذج تماماً القيم التجريبية الحقيقية المقاسة في المختبر لكل المركبات العضوية.

- كلما اقتربت القيمة من الصفر: دل ذلك على ارتفاع دقة النموذج وقوة تطابقه مع الواقع التجريبي.
- كلما كبرت القيمة (اتجهت نحو المالانهاية): دل ذلك على زيادة حجم الأخطاء وتباعد القيم المتنبأ بها عن القيم الحقيقية، مما يعني ضعف كفاءة النموذج.

جدول (2.III) نتائج الانحدار

R^2	النموذج	الخاصية
0.9446	$\log K_{oc} = a + \sum N_i C_i + \beta K_{app} + \gamma chi2v$	معامل امتصاص التربة K_{oc}

جدول (3.III): معاملات النموذج (Coefficients of the Model)

Variable	bservation/ith miss	without miss	Minimum	Maximum	Mean	d. deviation	
LogKoc_Exp	471	0	471	0.4500	6.3500	2.2432	1.1012
-CH3	471	0	471	0.0000	6.0000	1.0170	1.1203
-CH2-	471	0	471	0.0000	16.0000	1.7176	2.8500
≡CH	471	0	471	0.0000	2.0000	0.0552	0.2550
=CH2	471	0	471	0.0000	2.0000	0.0828	0.3188
=CH-	471	0	471	0.0000	6.0000	0.1529	0.5743
-OH (alcohol)	471	0	471	0.0000	3.0000	0.1423	0.4363
O=CH- (aldehyde)	471	0	471	0.0000	1.0000	0.0085	0.0919
-COOH (acid)	471	0	471	0.0000	2.0000	0.1104	0.3698
>C=O (nonring)	471	0	471	0.0000	2.0000	0.2229	0.4603
≡C-	471	0	471	0.0000	4.0000	0.0998	0.3922
-CN	471	0	471	0.0000	1.0000	0.0106	0.1026
>CH-	471	0	471	0.0000	3.0000	0.1168	0.3983
-COO- (ester)	471	0	471	0.0000	3.0000	0.1720	0.4728
-CH2- (ring)	471	0	471	0.0000	12.0000	0.5817	1.7532
-NH2	471	0	471	0.0000	2.0000	0.0955	0.3534
>CH- (ring)	471	0	471	0.0000	14.0000	1.8514	2.9172
>C< (ring)	471	0	471	0.0000	12.0000	0.7834	1.5395
-OH (phenol)	471	0	471	0.0000	4.0000	0.0510	0.2797
-Cl	471	0	471	0.0000	6.0000	0.1507	0.6908
-NO2	471	0	471	0.0000	2.0000	0.0191	0.1518
=CH- (ring)	471	0	471	0.0000	4.0000	0.0510	0.3881
=CH< (ring)	471	0	471	0.0000	3.0000	0.0191	0.1892
-O- (nonring)	471	0	471	0.0000	5.0000	0.0934	0.4225
>N- (nonring)	471	0	471	0.0000	2.0000	0.0637	0.2693
-O- (ring)	471	0	471	0.0000	6.0000	0.0616	0.4405
-S- (ring)	471	0	471	0.0000	2.0000	0.0234	0.1647
-S- (nonring)	471	0	471	0.0000	1.0000	0.0106	0.1026
>C=O (ring)	471	0	471	0.0000	2.0000	0.0722	0.3181
>C<	471	0	471	0.0000	8.0000	0.0955	0.6303
=C<	471	0	471	0.0000	2.0000	0.0425	0.2403
-Br	471	0	471	0.0000	10.0000	0.0488	0.5722
-I	471	0	471	0.0000	1.0000	0.0021	0.0461
-N= (nonring)	471	0	471	0.0000	3.0000	0.0425	0.3234
-N= (ring)	471	0	471	0.0000	2.0000	0.0085	0.1302
-F	471	0	471	0.0000	19.0000	0.1529	1.4897
>NH (nonring)	471	0	471	0.0000	2.0000	0.0297	0.1821
>NH (ring)	471	0	471	0.0000	1.0000	0.0170	0.1294
-SH	471	0	471	0.0000	1.0000	0.0021	0.0461
Kappa3	471	0	471	-27.0400	9507.9600	51.2621	624.7052
Chi2v	471	0	471	0.0000	12.7677	2.2487	1.4732

يمثل الجدول (3.III) (جدول المعاملات) القيم العددية لمعاملات نموذج الانحدار المطور. من خلال عمود 'القيمة (Value)', تم استنباط المعادلة الرياضية التي تربط معامل التوزيع K_{oc} بالخصائص الجزيئية والمؤشرات التوبولوجية المختارة. وتُظهر قيم الاحتمالية (P-value) المرتبطة بكل متغير أن العناصر المدرجة في النموذج تمتلك معنوية إحصائية عالية حيث أن $(P < 0.05)$ ، مما يؤكد مساهمتها الفعلية في دقة التنبؤ بسلوك المركبات العضوية في التربة".

جدول(4.III): تحليل التباين (ANOVA) للنموذج الخطي

Variable	bservator/ith miss	in:hout miss	Minimum	Maximum	Mean	d. deviatio	
LogKoc_Ex	50	0	50	0.7500	6.4500	2.3180	1.3564
-CH3	50	0	50	0.0000	4.0000	0.9200	1.0069
-CH2-	50	0	50	0.0000	14.0000	1.7000	2.9294
≡CH	50	0	50	0.0000	1.0000	0.0400	0.1979
=CH2	50	0	50	0.0000	1.0000	0.0800	0.2740
=CH-	50	0	50	0.0000	2.0000	0.2400	0.5911
-OH (alcoh	50	0	50	0.0000	2.0000	0.1800	0.4375
O=CH- (alc	50	0	50	0.0000	1.0000	0.0200	0.1414
-COOH (ac	50	0	50	0.0000	3.0000	0.2400	0.5911
>C=O (nor	50	0	50	0.0000	1.0000	0.1800	0.3881
≡C-	50	0	50	0.0000	1.0000	0.0600	0.2399
-CN	50	0	50	0.0000	1.0000	0.0200	0.1414
>CH-	50	0	50	0.0000	1.0000	0.0800	0.2740
-COO- (es	50	0	50	0.0000	2.0000	0.1600	0.4219
-CH2- (ring	50	0	50	0.0000	6.0000	0.1800	0.9409
-NH2	50	0	50	0.0000	1.0000	0.0600	0.2399
>CH- (ring	50	0	50	0.0000	18.0000	2.7400	4.3697
>C< (ring)	50	0	50	0.0000	12.0000	1.0800	2.2392
-OH (pher	50	0	50	0.0000	0.0000	0.0000	0.0000
-Cl	50	0	50	0.0000	2.0000	0.1000	0.4165
-NO2	50	0	50	0.0000	3.0000	0.0600	0.4243
=CH- (ring	50	0	50	0.0000	2.0000	0.0400	0.2828
=CH< (ring	50	0	50	0.0000	0.0000	0.0000	0.0000
-O- (nonri	50	0	50	0.0000	2.0000	0.1200	0.4352
>N- (nonri	50	0	50	0.0000	1.0000	0.0600	0.2399
-O- (ring)	50	0	50	0.0000	0.0000	0.0000	0.0000
-S- (ring)	50	0	50	0.0000	1.0000	0.0200	0.1414
-S- (nonri	50	0	50	0.0000	0.0000	0.0000	0.0000
>C=O (ring	50	0	50	0.0000	2.0000	0.0400	0.2828
>C<	50	0	50	0.0000	1.0000	0.0600	0.2399
=C<	50	0	50	0.0000	0.0000	0.0000	0.0000
-Br	50	0	50	0.0000	0.0000	0.0000	0.0000
-I	50	0	50	0.0000	0.0000	0.0000	0.0000
-N= (nonri	50	0	50	0.0000	0.0000	0.0000	0.0000
-N= (ring)	50	0	50	0.0000	0.0000	0.0000	0.0000
-F	50	0	50	0.0000	0.0000	0.0000	0.0000
>NH (nonr	50	0	50	0.0000	0.0000	0.0000	0.0000
>NH (ring)	50	0	50	0.0000	0.0000	0.0000	0.0000
-SH	50	0	50	0.0000	0.0000	0.0000	0.0000
Kappa3	50	0	50	0.0000	13.4700	3.0911	2.8749
Chi2v	50	0	50	0.3149	6.2007	2.2003	1.3961

يُظهر جدول تحليل التباين (ANOVA) معنوية النموذج الإحصائي المستخدم؛ حيث بلغت قيمة الاحتمالية (P-value) أقل من 0.0001، وهي قيمة أصغر بكثير من مستوى الدلالة المعتمد (0.05). هذا يشير إلى أن المتغيرات المستقلة المدرجة في النموذج (مثل المؤشرات التوبولوجية وبارامترات هانسين)

تمتلك قدرة تفسيرية حقيقية وذات دلالة إحصائية عالية للتنبؤ بمعامل K_{oc} ، مما ينفي فرضية العدم ويؤكد جودة الارتباط الخطي.

جدول (III. 5): مقارنة النتائج والتجريبية والتنبؤية لـ $\log K_{oc}$ تحليل مجموع المربعات من النوع الاول

Source	DF	m of square	mean square	F	Pr > F
-CH3	1	0.1741	0.1741	2.4444	0.1187
-CH2-	1	72.4250	72.4250	1016.5789	< 0.0001
≡CH	0	0.0000			
=CH2	0	0.0000			
=CH-	1	9.8265	9.8265	137.9283	< 0.0001
-OH (alcohol)	1	15.0383	15.0383	211.0826	< 0.0001
O=CH- (aldehyde)	0	0.0000			
-COOH (acid)	0	0.0000			
>C=O (nonring)	1	2.7311	2.7311	38.3352	< 0.0001
≡C-	1	2.3586	2.3586	33.1059	< 0.0001
-CN	0	0.0000			
>CH-	1	10.1522	10.1522	142.4994	< 0.0001
-COO- (ester)	0	0.0000			
-CH2- (ring)	1	8.7272	8.7272	122.4975	< 0.0001
-NH2	1	23.2132	23.2132	325.8269	< 0.0001
>CH- (ring)	1	231.4394	231.4394	3248.5524	< 0.0001
>C< (ring)	1	87.4679	87.4679	1227.7258	< 0.0001
-OH (phenol)	0	0.0000			
-Cl	1	15.2732	15.2732	214.3794	< 0.0001
-NO2	0	0.0000			
=CH- (ring)	1	5.1545	5.1545	72.3500	< 0.0001
=CH< (ring)	1	2.3152	2.3152	32.4963	< 0.0001
-O- (nonring)	1	0.8123	0.8123	11.4019	0.0008
>N- (nonring)	1	0.8476	0.8476	11.8978	0.0006
-O- (ring)	1	2.2534	2.2534	31.6298	< 0.0001
-S- (ring)	1	1.9013	1.9013	26.6866	< 0.0001
-S- (nonring)	0	0.0000			
>C=O (ring)	1	2.7340	2.7340	38.3753	< 0.0001
>C<	1	21.5478	21.5478	302.4517	< 0.0001
=C<	1	7.6957	7.6957	108.0188	< 0.0001
-Br	1	9.2780	9.2780	130.2284	< 0.0001
-I	1	0.4691	0.4691	6.5843	0.0106
-N= (nonring)	1	0.5734	0.5734	8.0491	0.0048
-N= (ring)	0	0.0000			
-F	0	0.0000			
>NH (nonring)	1	1.6073	1.6073	22.5600	< 0.0001
>NH (ring)	1	1.0534	1.0534	14.7862	0.0001
-SH	0	0.0000			
Kappa3	0	0.0000			
Chi2v	1	1.3225	1.3225	18.5636	< 0.0001

يؤكد الجدول فعالية النموذج المقترح في التنبؤ لـ $\log K_{oc}$ حيث تتقارب القيم المتوقعة بشكل وثيق مع القيم المرجعية ان انخفاض نسبة الخطأ في عينة الاختبار يعكس قدرة النموذج على التعميم خارج نطاق البيانات التدريب وهو ما يعد مؤشرا قويا على نجاح المزامنة بين البارامترات الهيكلية والمساهمة المجموعه.

جدول (6.III): الخصائص الإحصائية لتحليل مجموع المربعات من نوع الثالث

Source	DF	m of square	mean square	F	Pr > F
-CH3	1	5.6784	5.6784	79.7040	< 0.0001
-CH2-	1	202.3394	202.3394	2840.0964	< 0.0001
≡CH	0	0.0000			
=CH2	0	0.0000			
=CH-	1	20.5955	20.5955	289.0840	< 0.0001
-OH (alcohol)	1	1.3999	1.3999	19.6489	< 0.0001
O=CH- (aldehyde)	0	0.0000			
-COOH (acid)	0	0.0000			
>C=O (nonring)	1	1.3436	1.3436	18.8586	< 0.0001
≡C-	1	3.1699	3.1699	44.4942	< 0.0001
-CN	0	0.0000			
>CH-	1	4.8120	4.8120	67.5433	< 0.0001
-COO- (ester)	0	0.0000			
-CH2- (ring)	1	55.3121	55.3121	776.3778	< 0.0001
-NH2	1	4.1304	4.1304	57.9756	< 0.0001
>CH- (ring)	1	67.2480	67.2480	943.9124	< 0.0001
>C< (ring)	1	33.7756	33.7756	474.0841	< 0.0001
-OH (phenol)	0	0.0000			
-Cl	1	15.1256	15.1256	212.3075	< 0.0001
-NO2	0	0.0000			
=CH- (ring)	1	1.6268	1.6268	22.8339	< 0.0001
=CH< (ring)	1	2.1429	2.1429	30.0790	< 0.0001
-O- (nonring)	1	1.2793	1.2793	17.9567	< 0.0001
>N- (nonring)	1	1.1216	1.1216	15.7427	< 0.0001
-O- (ring)	1	1.9493	1.9493	27.3612	< 0.0001
-S- (ring)	1	1.9528	1.9528	27.4095	< 0.0001
-S- (nonring)	0	0.0000			
>C=O (ring)	1	2.8254	2.8254	39.6578	< 0.0001
>C<	1	20.5797	20.5797	288.8631	< 0.0001
=C<	1	7.3679	7.3679	103.4180	< 0.0001
-Br	1	7.2149	7.2149	101.2703	< 0.0001
-I	1	0.5058	0.5058	7.1000	0.0080
-N= (nonring)	1	0.4311	0.4311	6.0505	0.0143
-N= (ring)	0	0.0000			
-F	0	0.0000			
>NH (nonring)	1	1.5873	1.5873	22.2799	< 0.0001
>NH (ring)	1	1.0241	1.0241	14.3745	0.0002
-SH	0	0.0000			
Kappa3	0	0.0000			
Chi2v	1	1.3225	1.3225	18.5636	< 0.0001

جدول (III.7): معاملات النموذج او بارامترات النموذج

Source	Value	Standard error	t	Pr > t	Lower bound (95%)	Upper bound (95%)
Intercept	0.5417	0.0366	14.8190	< 0.0001	0.4699	0.6136
-CH3	0.1465	0.0164	8.9277	< 0.0001	0.1142	0.1787
-CH2-	0.2779	0.0052	53.2926	< 0.0001	0.2677	0.2882
≡CH	0.0000	0.0000				
=CH2	0.0000	0.0000				
=CH-	0.3918	0.0230	17.0025	< 0.0001	0.3465	0.4370
-OH (alcohol)	-0.1387	0.0313	-4.4327	< 0.0001	-0.2002	-0.0772
O=CH- (aldehyde)	0.0000	0.0000				
-COOH (acid)	0.0000	0.0000				
>C=O (non-ring)	0.1354	0.0312	4.3427	< 0.0001	0.0741	0.1967
≡C-	0.2204	0.0330	6.6704	< 0.0001	0.1555	0.2854
-CN	0.0000	0.0000				
>CH-	0.3138	0.0382	8.2185	< 0.0001	0.2388	0.3889
-COO- (ester)	0.0000	0.0000				
-CH2- (ring)	0.2485	0.0089	27.8636	< 0.0001	0.2309	0.2660
-NH2	-0.2916	0.0383	-7.6142	< 0.0001	-0.3668	-0.2163
>CH- (ring)	0.2147	0.0070	30.7232	< 0.0001	0.2010	0.2285
>C< (ring)	0.2844	0.0131	21.7735	< 0.0001	0.2587	0.3100
-OH (phenol)	0.0000	0.0000				
-Cl	0.2980	0.0205	14.5708	< 0.0001	0.2578	0.3382
-NO2	0.0000	0.0000				
=CH- (ring)	0.1685	0.0353	4.7785	< 0.0001	0.0992	0.2378
=CH< (ring)	0.3863	0.0704	5.4844	< 0.0001	0.2479	0.5248
-O- (nonring)	-0.1302	0.0307	-4.2375	< 0.0001	-0.1906	-0.0698
>N- (nonring)	-0.1962	0.0495	-3.9677	< 0.0001	-0.2934	-0.0990
-O- (ring)	-0.1713	0.0327	-5.2308	< 0.0001	-0.2357	-0.1069
-S- (ring)	0.4176	0.0798	5.2354	< 0.0001	0.2609	0.5744
-S- (nonring)	0.0000	0.0000				
>C=O (ring)	0.2659	0.0422	6.2974	< 0.0001	0.1829	0.3489
>C<	0.3498	0.0206	16.9960	< 0.0001	0.3093	0.3902
=C<	0.5617	0.0552	10.1695	< 0.0001	0.4531	0.6702
-Br	0.2503	0.0249	10.0633	< 0.0001	0.2014	0.2992
-I	0.7157	0.2686	2.6646	0.0080	0.1878	1.2435
-N= (nonring)	0.0960	0.0390	2.4598	0.0143	0.0193	0.1727
-N= (ring)	0.0000	0.0000				
-F	0.0000	0.0000				
>NH (nonring)	-0.3378	0.0716	-4.7202	< 0.0001	-0.4785	-0.1972
>NH (ring)	-0.3703	0.0977	-3.7914	0.0002	-0.5623	-0.1784
-SH	0.0000	0.0000				
Kappa3	0.0000	0.0000				
Chi2v	0.0426	0.0099	4.3086	< 0.0001	0.0231	0.0620

جدول (8.III): معاملات او بارامترات النموذج

Source	Value	Standard error	t	Pr > t	Lower bound (95% bound)	Upper bound (95% bound)
-CH3	0.1490	0.0167	8.9277	< 0.0001	0.1162	0.1818
-CH2-	0.7193	0.0135	53.2926	< 0.0001	0.6928	0.7458
≡CH	0.0000	0.0000				
=CH2	0.0000	0.0000				
=CH-	0.2043	0.0120	17.0025	< 0.0001	0.1807	0.2279
-OH (alcohol)	-0.0550	0.0124	-4.4327	< 0.0001	-0.0793	-0.0306
O=CH- (aldehyde)	0.0000	0.0000				
-COOH (acid)	0.0000	0.0000				
>C=O (nonring)	0.0566	0.0130	4.3427	< 0.0001	0.0310	0.0822
≡C-	0.0785	0.0118	6.6704	< 0.0001	0.0554	0.1017
-CN	0.0000	0.0000				
>CH-	0.1135	0.0138	8.2185	< 0.0001	0.0864	0.1407
-COO- (ester)	0.0000	0.0000				
-CH2- (ring)	0.3956	0.0142	27.8636	< 0.0001	0.3677	0.4235
-NH2	-0.0936	0.0123	-7.6142	< 0.0001	-0.1177	-0.0694
>CH- (ring)	0.5688	0.0185	30.7232	< 0.0001	0.5324	0.6052
>C< (ring)	0.3976	0.0183	21.7735	< 0.0001	0.3617	0.4334
-OH (phenol)	0.0000	0.0000				
-Cl	0.1870	0.0128	14.5708	< 0.0001	0.1617	0.2122
-NO2	0.0000	0.0000				
=CH- (ring)	0.0594	0.0124	4.7785	< 0.0001	0.0350	0.0838
=CH< (ring)	0.0664	0.0121	5.4844	< 0.0001	0.0426	0.0902
-O- (nonring)	-0.0500	0.0118	-4.2375	< 0.0001	-0.0731	-0.0268
>N- (nonring)	-0.0480	0.0121	-3.9677	< 0.0001	-0.0718	-0.0242
-O- (ring)	-0.0685	0.0131	-5.2308	< 0.0001	-0.0943	-0.0428
-S- (ring)	0.0624	0.0119	5.2354	< 0.0001	0.0390	0.0859
-S- (nonring)	0.0000	0.0000				
>C=O (ring)	0.0768	0.0122	6.2974	< 0.0001	0.0528	0.1008
>C<	0.2002	0.0118	16.9960	< 0.0001	0.1770	0.2233
=C<	0.1226	0.0121	10.1695	< 0.0001	0.0989	0.1463
-Br	0.1300	0.0129	10.0633	< 0.0001	0.1046	0.1554
-I	0.0299	0.0112	2.6646	0.0080	0.0079	0.0520
-N= (nonring)	0.0282	0.0115	2.4598	0.0143	0.0057	0.0507
-N= (ring)	0.0000	0.0000				
-F	0.0000	0.0000				
>NH (nonring)	-0.0559	0.0118	-4.7202	< 0.0001	-0.0791	-0.0326
>NH (ring)	-0.0435	0.0115	-3.7914	0.0002	-0.0660	-0.0210
-SH	0.0000	0.0000				
Kappa3	0.0000	0.0000				
Chi2v	0.0569	0.0132	4.3086	< 0.0001	0.0310	0.0829

جدول (9.III):اختبار المتغيرات

Variable	Variable IN/C	Status	MSE	R ²	Adjusted R ²	F	allows' C	Akaike's AIC	Schwarz's SB	Nemeniya's F
1	>C< (ring)	>C< (ring)	IN	0.8538	0.2974	0.2959	5113.8248	-72.4453	-64.1356	0.7056
2	-CH2- / >C	-CH2-	IN	0.5389	0.5575	0.5556	3049.7750	-288.2174	-275.7528	0.4462
3	-CH2- / -C(CH2- (ring		IN	0.4572	0.6254	0.6230	2512.5939	-364.6539	-348.0345	0.3794
4	-CH2- / -C>CH- (ring)		IN	0.3825	0.6872	0.6845	2023.4136	-447.6256	-426.8513	0.3181
5	-CH2- / '=C	=CH-	IN	0.3318	0.7293	0.7264	1691.1159	-513.6921	-488.7630	0.2765
6	-CH2- / '=C	>C<	IN	0.2804	0.7717	0.7688	1356.3903	-591.9145	-562.8305	0.2342
7	-CH3 / -CH	-CH3	IN	0.2371	0.8074	0.8045	1074.7503	-670.0287	-636.7898	0.1984
8	-CH3 / -CH	-Cl	IN	0.1943	0.8425	0.8398	798.2213	-762.6923	-725.2986	0.1629
9	-CH3 / -CH	-Br	IN	0.1743	0.8590	0.8563	668.7443	-812.9826	-771.4340	0.1464
10	-CH3 / -CH=CH- (ring)		IN	0.1591	0.8716	0.8688	570.9784	-854.9358	-809.2324	0.1340
11	-CH3 / -CH	-NH2	IN	0.1457	0.8827	0.8799	484.8681	-895.4885	-845.6302	0.1229
12	-CH3 / -CH	=C<	IN	0.1330	0.8931	0.8903	404.0136	-937.3462	-883.3330	0.1125
13	-CH3 / -CH	>CH-	IN	0.1240	0.9006	0.8977	346.8426	-969.3720	-911.2039	0.1051
14	-CH3 / -CH>C=O (ring)		IN	0.1165	0.9068	0.9039	299.6319	-997.6713	-935.3485	0.0989
15	-CH3 / -CH	≡C-	IN	0.1088	0.9131	0.9103	251.1252	-1028.9384	-962.4607	0.0926
16	-CH3 / -CH=O (nonring)		IN	0.1014	0.9193	0.9164	204.3196	-1061.4839	-990.8514	0.0864
17	-CH3 / -CH=CH< (ring)		IN	0.0960	0.9237	0.9208	171.2329	-1085.9842	-1011.1967	0.0820
18	-CH3 / -CHNH (nonring)		IN	0.0921	0.9269	0.9240	147.2863	-1104.5870	-1025.6446	0.0788
19	-CH3 / -CH -S- (ring)		IN	0.0886	0.9299	0.9269	125.8177	-1122.0316	-1038.9344	0.0760
20	-CH3 / -CH -O- (ring)		IN	0.0860	0.9321	0.9291	110.5308	-1134.8859	-1047.6339	0.0739
21	-CH3 / -CHH (alcohol)		IN	0.0831	0.9345	0.9314	93.2284	-1150.0454	-1058.6385	0.0716
22	-CH3 / -CH>NH (ring)		IN	0.0805	0.9367	0.9336	77.6852	-1164.2025	-1068.6408	0.0695
23	-CH3 / -CH	Chi2v	IN	0.0780	0.9388	0.9357	62.8062	-1178.2892	-1078.5726	0.0674
24	-CH3 / -CHD- (nonring)		IN	0.0755	0.9409	0.9378	48.1892	-1192.6818	-1088.8103	0.0654
25	-CH3 / -CHN- (nonring)		IN	0.0730	0.9430	0.9398	33.7096	-1207.5224	-1099.4961	0.0634
26	-CH3 / -CH	-I	IN	0.0721	0.9439	0.9406	28.8697	-1212.6930	-1100.5119	0.0627
27	-CH3 / -CH= (nonring)		IN	0.0712	0.9446	0.9413	24.8620	-1217.0824	-1100.7464	0.0621

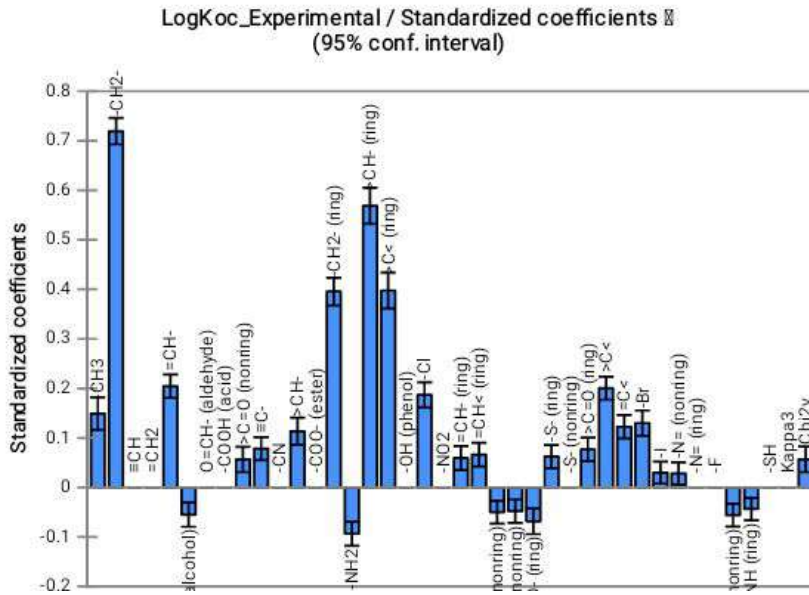
الرسوم البيانية:

تعد الرسوم البيانية وسيلة جوهرية لتمثيل البيانات وتحليلها؛ حيث ركزنا في هذا القسم على ثلاثة أنواع رئيسية نتيج لنا تصور الفجوة بين القيم المتوقعة (Predictions) والقيم التجريبية الملاحظة (Observations).

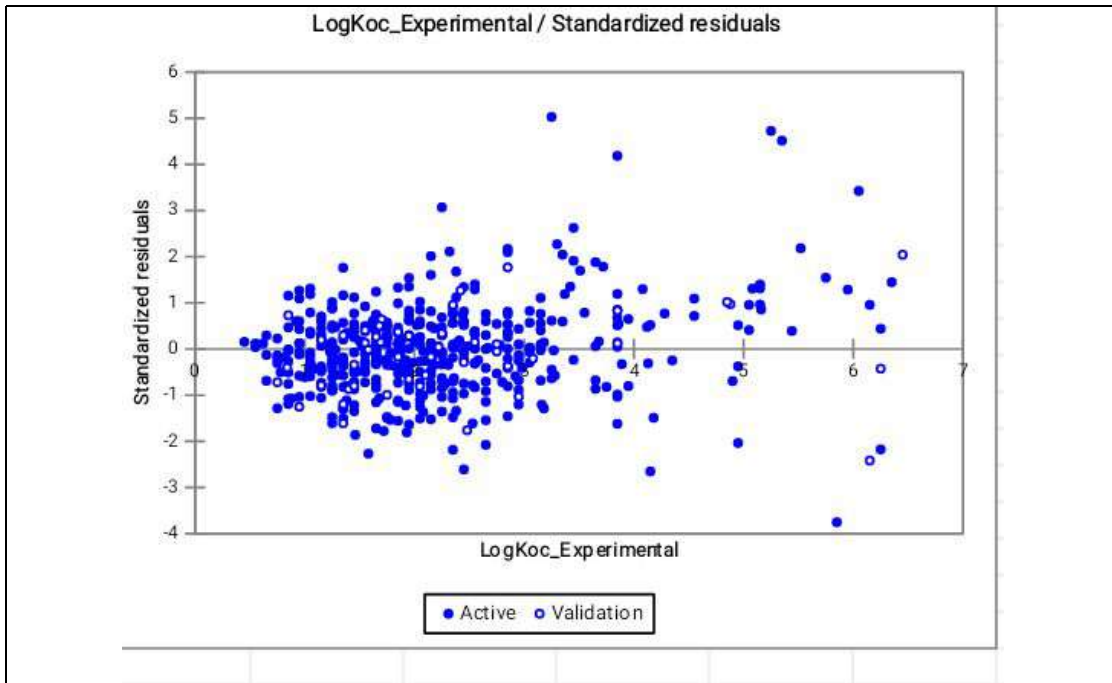
1. **مخطط الارتباط (Correlation Plot):** في النموذج المثالي، يُفترض أن تقع جميع النقاط على خط المنصف خط $(y = x)$ ، مما يشير إلى تطابق تام بين التوقعات والنتائج الفعلية.

2. **الرسوم البيانية الشريطية للبواقي المعيارية (Standardized Residuals):** خصصنا رسمين بيانيين لهذا الغرض؛ الأول يتعلق بـ **المجموعة النشطة (Training Set)**، والثاني بـ **مجموعة التحقق (Validation Set)** التي يتم اختيارها عشوائياً لتقييم كفاءة النموذج.

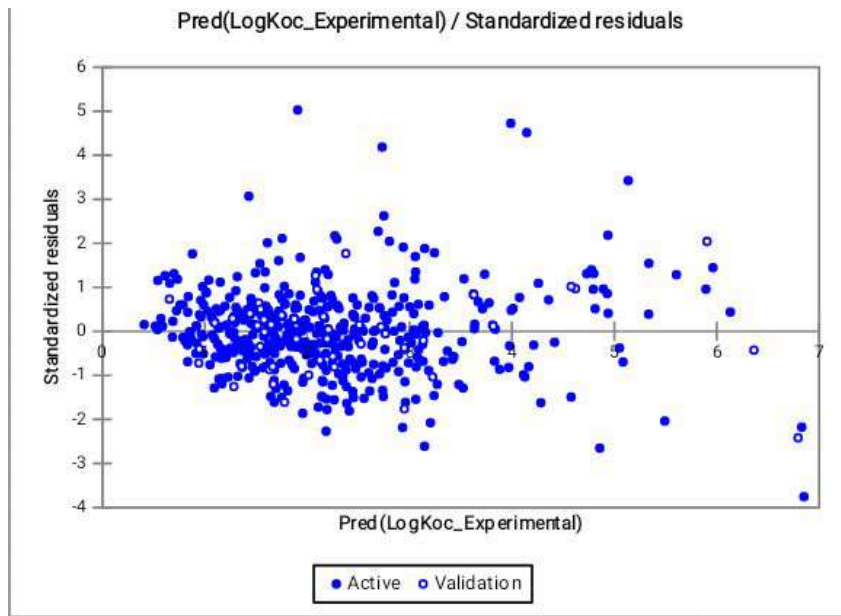
تكم أهمية هذين الرسمين الأخيرين في الكشف عن "القيم الشاذة"، ومعرفة ما إذا كان هناك عدد غير طبيعي من البيانات يقع خارج نطاق الثقة المعتاد. فبافتراض التوزيع الطبيعي، يجب أن تقع حوالي **95%** من البيانات ضمن الفترة المحصورة بين **[-2, +2]**، مما يؤكد دقة النموذج وصلاحيته للتنبؤ.



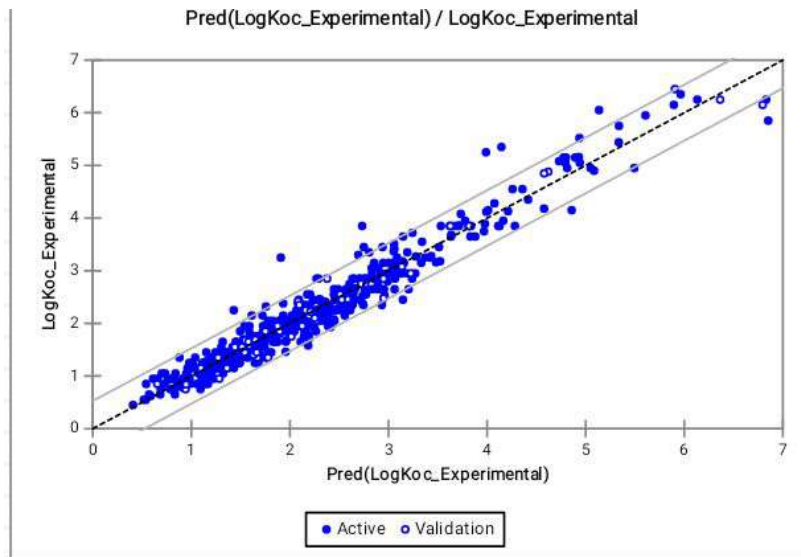
الشكل (1.III): مخطط البواقي مقابل القيم المتوقعة (Residuals vs Predictions Plot)



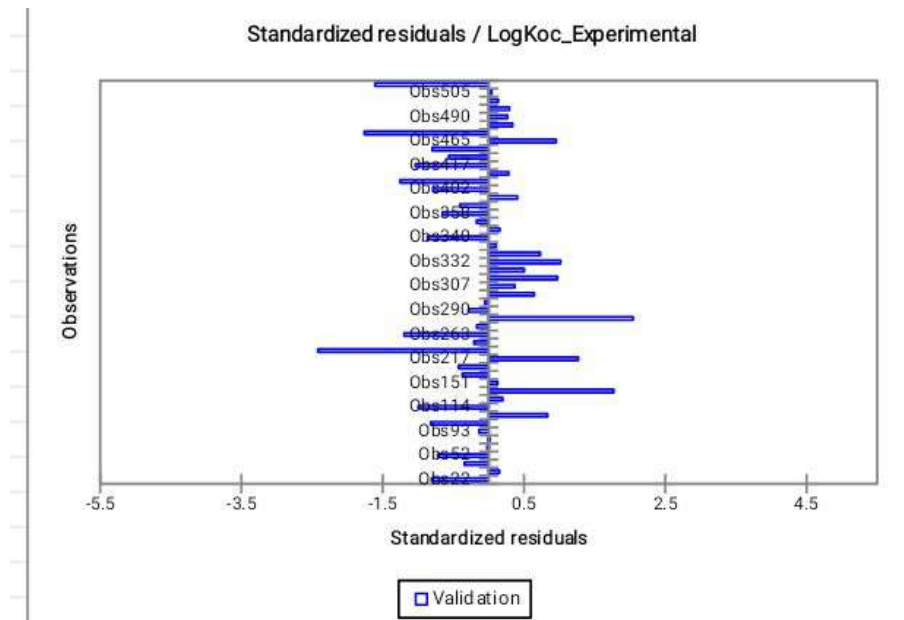
الشكل (2.III): مخطط البواقي المعيارية رسم بياني لتحليل التباين (Analysis of Variance - ANOVA Plot)



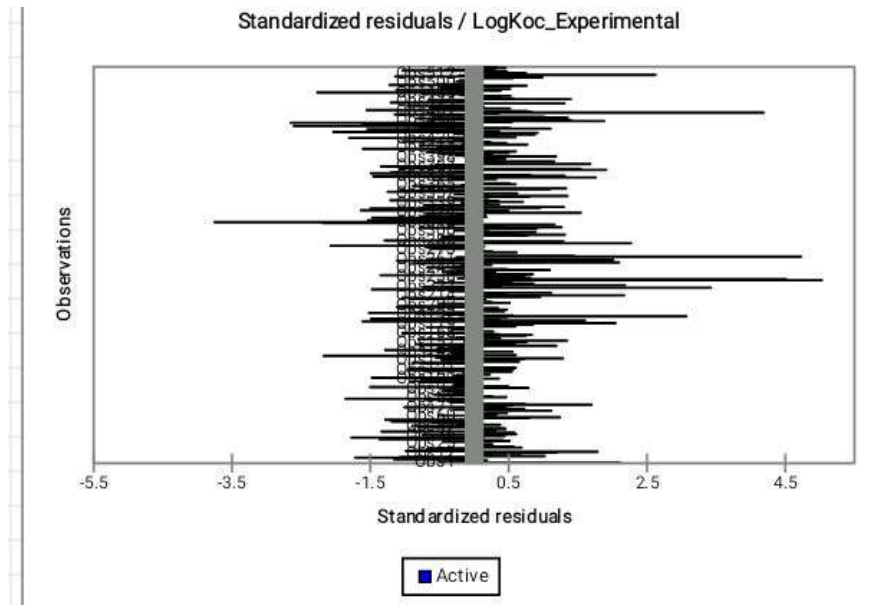
الشكل (3.III): مخطط تحليل المسافات (Distance Analysis Plot)



الشكل (4.III): مخطط الارتباط



الشكل (5.III): الرسم البياني الشريطي للبقاقي المعيارية لمجموعة التحقق (Standardized Residuals Bar Chart for Validation Set)



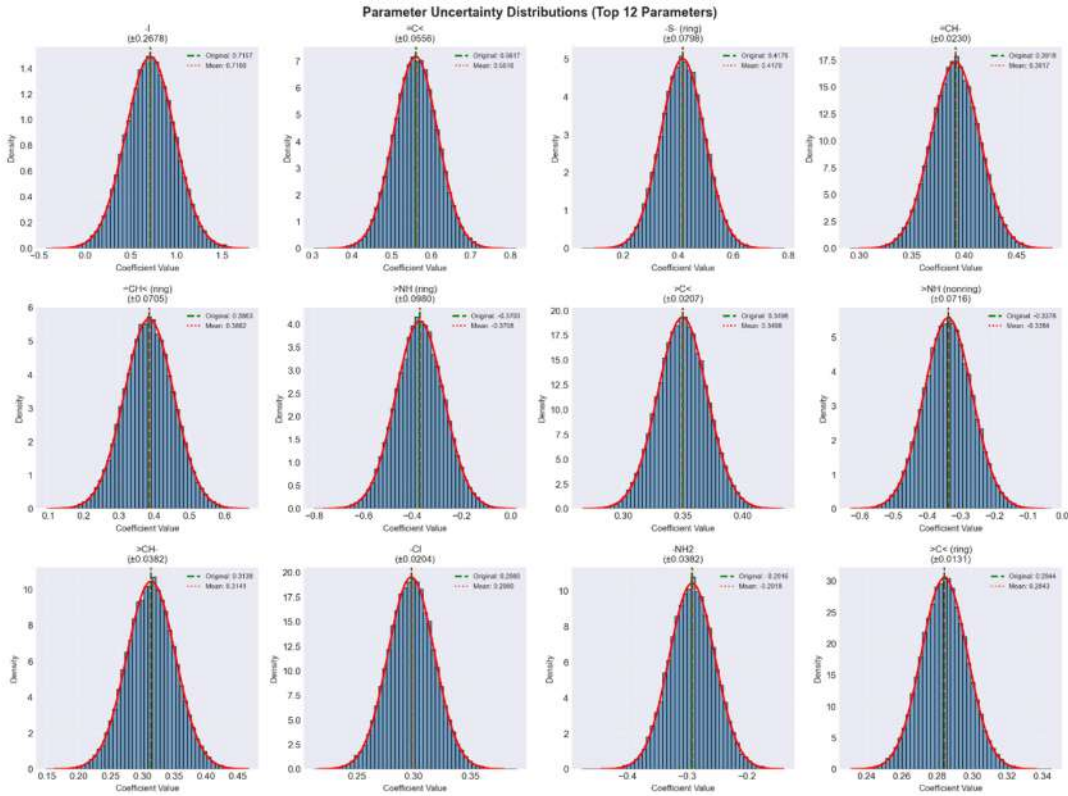
الشكل (6.III): رسم بياني شريطي للبقاقي المعيارية (Standardized Residuals Bar Chart) لمجموعة النشطة

5.III تحليل نتائج محاكاة مونت كارلو لحساب $\log K_{OC}$: دراسة حالة لمركب البنزين والتقييم الشامل للمركبات العضوية

أُجريت محاكاة مونت كارلو لتقييم عدم اليقين في حساب قيم $\log K_{OC}$ باستخدام نموذج يعتمد على معاملات مساهمة ذرية وجزيئية (C, S, CH, NH) وغيرها. تم تطبيق المحاكاة على مجموعة واسعة من المركبات العضوية تضم 521 مركباً، مع تحليل تفصيلي لمركب البنزين كمثال. يقدم هذا التقرير تحليلاً إحصائياً للقيم المنتبأ بها، وتوزيع الأخطاء، وتقييم دقة النموذج، مع تحديد حدود تطبيقه.

1.5.III منهجية التحليل:

• المدخلات: معاملات النموذج مع عدم يقين موضحة في الشكل التالي:



شكل (7.III): منحني توزيع معاملات عدم اليقين (أعلى 12 معامل)

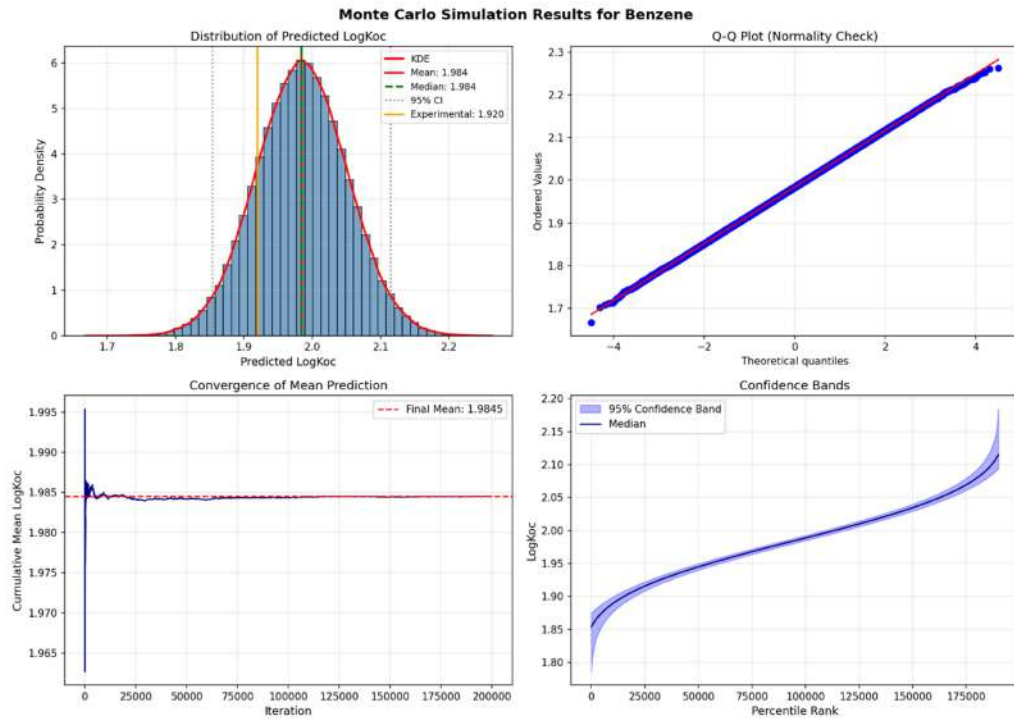
- عدد التكرارات: آلاف التكرارات لكل مركب (حتى تقارب المتوسط).
- المخرجات: متوسط متوقع، وسيط، فترات ثقة (68% و 95%)، توزيع الأخطاء.
- المركبات: مجموعة شاملة تتضمن مركبات بسيطة ومعقدة (أكثر من 50 مركباً موثقاً).

2.5.III تحليل النتائج الخاصة بمركب البنزين (كحالة دراسة):
2.1 دقة التنبؤ

جدول (10.III): نتائج دقة التنبؤ للبنزين

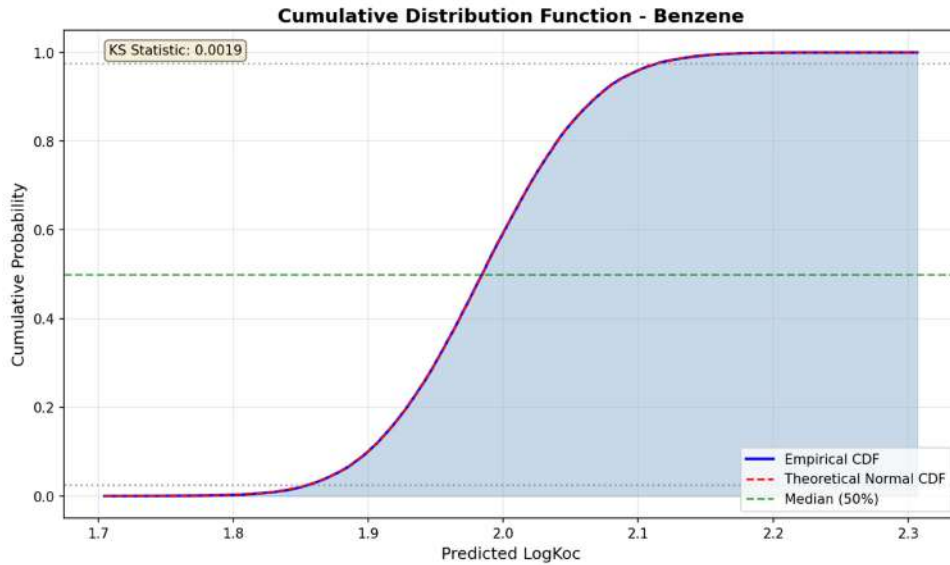
المقياس	القيمة
القيمة التجريبية	1.920
المتوسط المتنبأ به	1.984
الوسيط	1.984
الخطأ المطلق	0.064
الانحراف المعياري	0.066

- يشير التقارب بين المتوسط والوسيط إلى توزيع متماثل للقيم المتنبأ بها، مما يؤكد ملاءمة استخدام المتوسط كمقياس نزعة مركزي. الخطأ المطلق (0.064) يقع ضمن نطاق الدقة المقبول للمركبات البسيطة.
- 2.2 مجالات الثقة وعدم اليقين:**
- مجال الثقة 68%: تتراوح بين قرابة 1.92 و 2.05.
 - مجال الثقة 95%: تتراوح قيمه بين قرابة 1.85 و 2.12 (كما يبينه الشكل (III.2)).



شكل (8.III): منحنيات نتائج محاكاة مونت كارلو للبنزين

- المدى الكلي: يتراوح مداه بين 1.69 و 2.29 ، مما يشير إلى وجود قيم متطرفة نادرة.
- 2.3 التوزيع الإحصائي:
- اختبار كولموغوروف-سميرنوف: (KS) :
يشير الشكل الموالي الى تحليل مؤشرات التوزيع الطبيعي او ما يسمى اختبار كولموغوروف – سميرنوف



شكل (9.III): منحنى نتائج محاكاة مونت كارلو للبنزين

- تقيس إحصائية KS أقصى مسافة عمودية بين منحنى التوزيع التراكمي التجريبي (Empirical CDF) والنظري (Normal CDF)، حيث كلما كانت القيمة أصغر، كان التوزيعان أكثر تطابقاً. تشير القيم أقل من 0.05 عادةً إلى عدم وجود فرق ذي دلالة إحصائية بين التوزيعين. أي أن التوزيع التجريبي للقيم المنتبأ بها لا يختلف بشكل ذي دلالة إحصائية عن التوزيع الطبيعي النظري.
 - (الالتواء) (Skewness): -0.011 قريب من الصفر
 - (التفرطح) (Kurtosis): 0.02 قريب من التوزيع الطبيعي
- بناءً على هذه المؤشرات، يُستنتج أن توزيع القيم المنتبأ بها للبنزين لا تختلف بشكل ذي دلالة إحصائية عن التوزيع الطبيعي. هذه الخاصية تبرر استخدام الإحصاءات البارامترية (المتوسط، الانحراف المعياري) في توصيف عدم اليقين.

2.4 تحليل البواقي – (Residuals) للبنزين:

جدول (11.III): نتائج تحليل البواقي للبنزين

المؤشر	القيمة
متوسط البواقي	0.064
الانحراف المعياري للبواقي	0.066
نسبة البواقي ضمن $\pm 1\sigma$	48.7%
نسبة البواقي ضمن $\pm 2\sigma$	84.5%
نسبة البواقي ضمن $\pm 3\sigma$	97.9%

تشير النسب المئوية إلى انحراف طفيف عن التوزيع الطبيعي المثالي) حيث يُتوقع 68%، 95%، 99.7% على التوالي). هذا الانحراف لا يعتبر مشكلة منهجية كبيرة، بل يعكس طبيعة البيانات والتكرارات المحدودة في ذيول التوزيع. لم يُظهر مخطط Residuals vs Predicted نمطاً محددًا (مثل الشكل المخروطي أو الاتجاه)، مما يؤكد عدم وجود علاقة بين حجم الخطأ وقيمة التنبؤ.

2.5 استقرار محاكاة مونت كارلو (Convergence)

أظهر مخطط monte_carlo_Benzene.png قسم Convergence of Mean Prediction أن المتوسط التراكمي استقر عند القيمة 1.9845 بعد عدد محدود من التكرارات، مع عدم وجود اتجاه صعودي أو هبوطي بعد ذلك. هذا يدل على أن عدد التكرارات كان كافيًا لتحقيق التقارب الإحصائي.

3.5.III التحليل الشامل لجميع المركبات 3.1 مؤشرات الدقة الكلية:

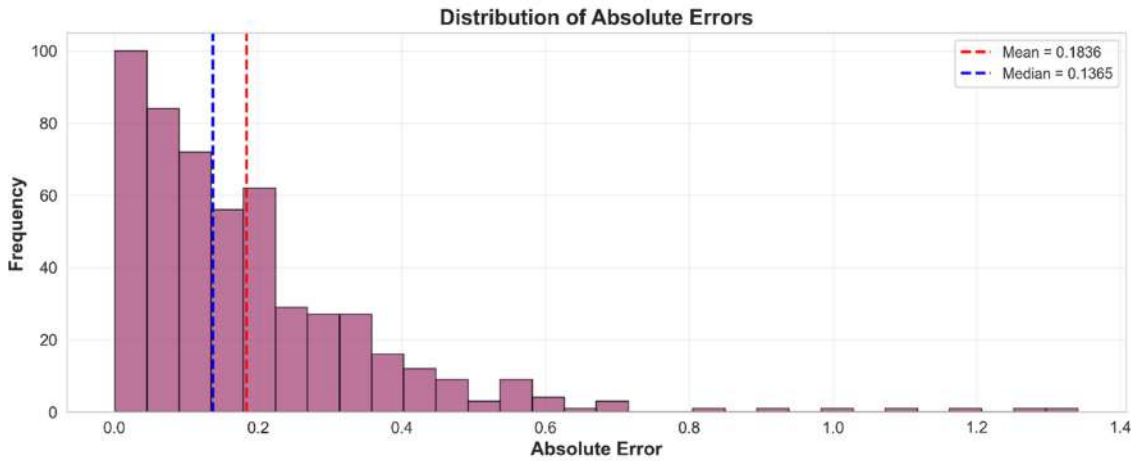
جدول (12.III): نتائج دقة التنبؤ الكلية

المقياس	القيمة
متوسط الخطأ المطلق (MAE)	0.1836
جذر متوسط مربع الخطأ (RMSE)	0.2561

تعتبر هذه القيم جيدة إلى ممتازة مقارنة بال نماذج المنشورة في بحوث سابقة لحساب LogKoc ، والتي غالبًا ما تتراوح RMSE فيها بين 0.3 و 0.5.

3.2 توزيع الأخطاء المطلقة (Absolute Errors) :

استنادًا إلى المنحنى المبين في الشكل (III.3) :



شكل (10.III): مخطط توزيع الأخطاء المطلقة

- الذروة الرئيسية تقع في الفئة 0.0 – 0.2.
- يتناقص التكرار تدريجيًا حتى 1.0.
- يوجد ذيل طويل يمتد حتى 1.35 .

التفسير الكمي:

- حوالي 80% من المركبات لديها خطأ مطلق < 0.3
- حوالي 10-15% من المركبات هي المسؤولة عن الأخطاء الكبيرة (> 0.6)
- هذا النمط ("ذيل طويل") شائع في النماذج التي تعمل بشكل جيد على الأغلبية، ولكنها تفشل في فئة فرعية معينة من البيانات.

3.3 تحليل أفضل وأسوأ المركبات:

أفضل 30 مركبًا: (Bottom 30 Absolute Errors)

- نطاق الخطأ: 0.0009 – 0.0135

- أمثلة تمثيلية (0.0118) Fluorene ، (0.0051) m-Xylene ، (0.0019) Vinyl acetate
- الخصائص المشتركة: مركبات عطرية بسيطة، هيدروكربونات صغيرة، كيتونات وإسترات خطية قصيرة السلسلة، أحماض كربوكسيلية بسيطة.

أسوأ 30 مركبًا: (Top 30 Absolute Errors)

- نطاق الخطأ: 1.35 – 0.42
- أمثلة تمثيلية (0.98) Cholesterol ، (1.25) Farnesol ، Hexabromocyclododecane (1.20) ، (1.35) Poly(ethyleneglycol)
- الخصائص المشتركة:
 - وزن جزيئي مرتفع جدًا (> 300 Da) غالبًا
 - وجود هالوجينات (Br, Cl) متعددة
 - هياكل ستيرويدية أو تربينية معقدة
 - مجموعات وظيفية متعددة ومتباعدة مكانيًا

3.4 تحليل عدم اليقين في معاملات النموذج:

من parameter_uncertainty.png

جدول (13.III) تحليل عدم اليقين في معاملات النموذج

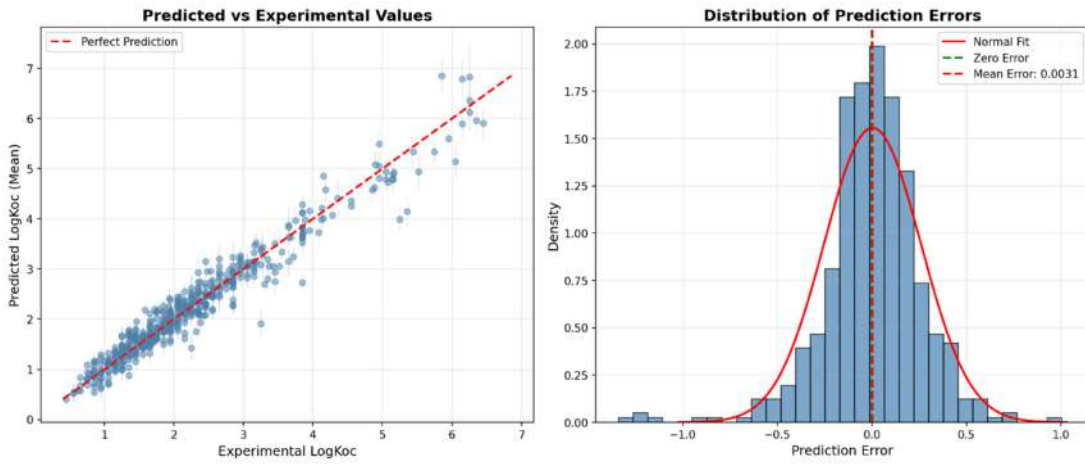
الملاحظة	المدى (Min - Max)	القيمة المركزية	المعامل
عدم يقين كبير جدًا	0.0000 – 0.7175	0.2678	C (كربون)
عدم يقين كبير	0.0000 – 0.5571	0.0556	S (كبريت)
عدم يقين متوسط	0.0000 – 0.3918	0.0230	CH
عدم يقين متوسط	-0.0370 – 0.3730	0.0980	NH

عدم اليقين في المعاملات C و S هو المصدر الرئيسي لانتشار القيم المتنبأ بها في محاكاة مونت كارلو. أي تحسين في حساب هذه المعاملات (عبر بيانات تجريبية إضافية) سينعكس إيجابًا على دقة النموذج لجميع المركبات.

3.5 العلاقة بين الخطأ وحجم $\log K_{oc}$:

أظهر منحنى الشكل (III.4) وجود علاقة طردية بين قيمة $\log K_{oc}$ المتوقعة وحجم الخطأ:

- عند $\log K_{oc} \sim 2.0$ الخطأ ~ 0.12
- عند $\log K_{oc} \sim 4.0$ الخطأ ~ 0.24
- عند $\log K_{oc} \sim 6.0$ الخطأ ~ 0.36



شكل (11.11): مخطط توزيع اخطا التنبؤ والمقارنة بين القيم التجريبية والقيم التنبؤية

التفسير: تميل المركبات شديدة الكارهة للماء قيم $\log K_{oc}$ العالية إلى أن تكون جزيئات كبيرة ومعقدة، مما يقع خارج نطاق صلاحية النموذج الخطي القائم على المساهمة الذرية البسيطة

4.111. الاستنتاجات العلمية

أ. النموذج مناسب للاستخدام لمجموعة واسعة من المركبات العضوية البسيطة والمتوسطة التعقيد،

محققاً $MAE = 0.184$ و $RMSE = 0.256$

ب. البنزين يمثل حالة مثالية للنموذج، حيث أظهر المحاكاة أن حساب $\log K_{oc}$ (1.984) يقع ضمن فاصل ثقة 95% ضيق ويبعد 0.064 فقط عن القيمة التجريبية (1.920). التوزيع طبيعي والمتوسط مقارب.

ج. **الفشل النظامي للنموذج:** يظهر بوضوح في فئتين من المركبات:

- المركبات الضخمة (وزن جزيئي $(Mw) < 300$)
- المركبات متعددة الهالوجين.

د. **المصدر الرئيسي لعدم اليقين** في المحاكاة هو المعاملات المرتبطة بالكربون (C) والكبريت (S). تقييد هذه المعاملات ببيانات تجريبية إضافية من شأنه تحسين الدقة الكلية.

هـ. **الموثوقية:** يمكن الوثوق بالتنبؤات للمركبات المشابهة للقائمة السفلية (bottom 30) يجب عدم استخدام النموذج للمركبات المشابهة للقائمة العلوية (top 30) دون معايرة إضافية أو نموذج بديل.

5. التوصيات:

- إذا كان الهدف هو حساب $\log K_{oc}$ لمجموعة جديدة من المركبات، يُنصح أولاً بمقارنة خصائصها (الوزن الجزيئي، وجود الهالوجينات، التعقيد البنوي) بالمركبات الواردة في قائمتي best و worst.
- يمكن استخدام محاكاة مونت كارلو الحالية لتوليد مجالات ثقة للتنبؤات، بدلاً من الاعتماد على قيمة نقطية وحيدة.
- يُقترح تحسين النموذج بإضافة حدود قصوى لمساهمة المعاملات C و S بناءً على بيانات تجريبية للمركبات كبيرة الحجم.

الخلاصة:

في هذا الفصل يمكن القول ان النتائج المستخلصة من عملية النمذجة الاحصائية وتحليل البيانات قد اكدت كفاءة الطريقة المطورة التي تعتمد على مساهمة المجموعات والمؤشرات الطوبولوجية في تقدير معامل الامتصاص الكربوني(Koc) بدقة عالية. حيث يمكن تلخيص اهم نتائج التي توصلنا اليها في هذه الدراسة فيما يلي:

- ❖ **جودة النموذج الاحصائي(الرياضي):** وفر النموذج المطور دقة عالية بمعامل تحديد(R^2) وصل الى 0.9446 مع معنوية احصائية قوية ($P\text{-value} < 0.0001$) لجميع المتغيرات المدرجة.
 - ❖ **دور مؤشرات الاتصال:** اثبت الحليل المقارنة ان اضافة مؤشرات الاتصال ساهمت بشكل كبير وجوهري في رفع كفاءة النموذج وتقليل جدر متوسط مربع الخطأ (RMSE)، مما اده الى تفسير افضل للتباين في لبيانات.
 - ❖ **التفسير الكيميائي:** تشير النتائج ان المجموعات الكارهة للماء(كالحلقات العطرية والسلاسل الكربونية الطويلة)ترفع من قيمة Koc، حيث تعمل المجموعات القطبية (كالهيدروكسيل والامين)على خفضها، مما يزيد من قابلية انحلال المركب في الماء.
 - ❖ **الموثوقية والتوزيع الطبيعي:** اثبات اختبارات التوزيع الطبيعي و"محاكاة مونت كارلو" استقرار التنبؤات وتوطن الاخطاء حول الصفر. كما تقع حوالي 95% من البيانات ضمن نطاق الثقة المعتمد.
 - ❖ **القدرة التنبؤية:** اظهر هذا النموذج قدرة فائقة على التعميم، مما يجعله اداة موثوقة لتوقع السلوك البيئي للمركبات العضوية، وذلك اعتمادا على تقارب النتائج بين المجموعة النشطة والمجموعة التحقق.
- نستنتج ان النموذج المطور اداة تنبؤية دقيقة وفعالة تجمع بين البارامترات الهيكلية والمساهمة المجموعاتي لتقليص الانحرافات في تقدير معامل Koc.

خلاصة عامة

خارصه عامه

خلصت هذه الدراسة إلى تطوير نموذج رياضي متقدم يعتمد على منهجية "مساهمة المجموعات" (Group Contribution Method) للتنبؤ بمعامل الامتصاص الكربوني ($\log oc$). أثبتت النتائج أن هذا النموذج يمثل أداة فعالة لتقدير السلوك الديناميكي للملوثات العضوية في الأوساط البيئية، حيث يتيح الحساب الدقيق لقيم الامتصاص استناداً فقط إلى المعطيات الهيكلية والمجموعات الوظيفية للجزيء، مما يلغي الحاجة تماماً للقياسات المخبرية المكلفة والمستهلكة للوقت.

تم تطوير نموذج تنبؤي جديد لتقدير معامل الامتصاص على الكبون العضوي للمركبات العضوية للمركبات العضوية بالاعتماد على الدمج بين المساهمات المجموعات الوظيفية والمؤشرات الطوبولوجية .

من خلال المعالجة الإحصائية المتقدمة باستخدام برمجية **XLSTAT**، والاعتماد على قاعدة بيانات تضم 521 مركباً، أفضت الدراسة إلى الاستنتاجات العلمية التالية:

- **الأداء الإحصائي الفائق:** حقق النموذج المطور دقة تنبؤية استثنائية، حيث بلغت قيمة معامل التحديد (R^2) 0.9446، وهو ما يعكس قدرة النموذج على تفسير الغالبية العظمى من التباينات الفيزيوكيميائية للمركبات المدروسة، مدعوماً بانخفاض ملحوظ في معدلات الخطأ ($RMSE=0.2669$).
- **التأثير البنوي والمجموعي:** كشف التحليل التدريجي (Stepwise) أن الهيكل الكربوني الأساسي والحلقي يمثل المحدد الجوهرى لخاصية الامتصاص، بينما تلعب المجموعات الوظيفية (خاصة الهالوجينية منها كالكلور والبروم) دوراً حاسماً في تعزيز الصفة الكارهة للماء، مما يرفع من ثباتية المركبات في المادة العضوية للتربة.
- **الدقة في التمييز الجزيئي:** أظهرت المنهجية المتبعة مرونة عالية في التمييز بين المتماكبات (Isomers)، وهو عامل حيوي في الكيمياء البيئية نظراً لاختلاف السلوك الفيزيائي والحيوي للمركبات حتى وإن تشابهت صيغتها الإجمالية.
- **الجدوى البيئية والتطبيقية:** يساهم هذا النموذج بشكل مباشر في دعم استراتيجيات تقييم المخاطر البيئية، حيث يوفر وسيلة سريعة لتصنيف الملوثات حسب قدرتها على الانتقال والوصول إلى المياه الجوفية، مما يسهل اتخاذ قرارات وقائية لحماية الأنظمة البيئية.

- **الموثوقية والشمولية:** إن التنوع الكبير في قاعدة البيانات المستخدمة يعزز من موثوقية النتائج ويجعل النموذج قابلاً للتطبيق على مدى واسعة من العائلات الكيميائية، مع إمكانية تحديثه وتطويره باستمرار ليشمل مركبات كيميائية أكثر تعقيداً في المستقبل.

التوصيات

التوصيات

ختاماً وفيما يخص العمل المستقبلي نقترح التوصيات التالية

- ❖ توسيع قاعدة البيانات،
- ❖ اختبار واصفات جزيئية جديدة،
- ❖ تطبيق النموذج على ملوثات عضوية اخرى،
- ❖ مقارنة النموذج مع تقنيات الذكاء الاصطناعي.