QSPR Study of the Water Solubility of a Diverse Set of Agrochemicals: Hybrid (GA/ MLR) Approach

Etude QSPR de la Solubilité Aqueuse d'un Ensemble de Divers Produits Agrochimiques: Approche Hybride (AG/RLM)

Amel Bouakkadia, Hamza Haddag, Nabil Bouarra, Djelloul Messadi*

Environmental and Food Safety Laboratory, Badji Mokhtar University, Annaba 23000, Algeria.

Soumis le : 05/03/2015 *Révisé le : 21/03/2016* *Accepté le : 12/04/2016*

ملخص

أجريت علاقة بين كمية الهيكل و الخاصية للتنبؤ بانحلالية المبيدات منتمية إلى أربعة أقسام كيميائية: أحماض، اليوريا، تريازين، وكاربامات. المجموعة المكونة من 77 مبيد قسمت إلى مجموعة بناء من 58 مبيد ومجموعة اختبار من 19 مبيد بتقنية. النموذج بستة متغيرات بمعامل ارتباط (R^2) يساوي 0.8895 و خطا معيار التقدير (s) يساوي 0.52 وحدة تم تطويره بتطبيق التراجع المتعدد الخطي باستخدام المربعات الصغرى واختيار مجموعة المتغيرات تم باستعمال الخوارزمية الجينية. قوة النموذج المقترح تأكدت باستخدام عدة تقنيات للتقييم leave- one- out ، bootstrap ، الاختبارات العشوائية، والتحقق من خلال مجموعة الاختبار.

الكلمات الدالة: *المبيدات ، الانحلالية، QSPR ، الموصفات الجزيئية، التراجع المتعدد الخطي.*

Abstract

A quantitative structure- property relationship (QSPR) was performed for the prediction of the aqueous solubility of pesticides belonging to four chemical classes: acid, urea, triazine, and carbamate. The entire set of 77 pesticides was divided into a training set of 58 pesticides and a test set of 19 pesticides according to the Snee technique. A six descriptor model, with squared correlation coefficient (R^2) of 0.8895 and standard error of estimation (s) of 0.52 log unit, was developed by applying multiple linear regression analysis using the ordinary least square regression method and genetic algorithm- variable subset selection. The reliability of the proposed model was further illustrated using various evaluation techniques: leave- one- out cross- validation, bootstrap, randomization tests, and validation through the test set.

Key Words: *pesticides- aqueous solubility- QSPR- molecular descriptors- multiple linear regression.*

Résumé

Une relation quantitative structure-propriété (QSPR) a été réalisée pour la prédiction de la solubilité aqueuse des pesticides appartenant aux quatre classes chimiques: acide, urée, triazine, et carbamate. L'ensemble des 77 pesticides a été divisé en un ensemble de calibrage de 58 pesticides et un ensemble de test de 19 pesticides selon la technique de Snee. Un modèle de six descripteurs, avec un coefficient de corrélation (R^2) de 0,8895 et une erreur standard d'estimation (s) de 0,52, a été développé en appliquant une analyse de régression linéaire multiple en utilisant la méthode de régression des moindres carrés ordinaires et les algorithme-génétiques pour la sélection des sous-ensembles de variables. La fiabilité du modèle proposé a été en outre illustrée en utilisant diverses techniques d'évaluation: validation croisée par leave- one- out, bootstrap, tests de randomisation, et la validation par l'ensemble de test.

Mots clés: *pesticides- solubilité aqueuse- QSPR- descripteurs moléculaires- régression linéaire multiple.*

* *Corresponding author : d_messadi@yahoo.fr*

©UBMA - 2016