The 24th International Electronic Conference on Synthetic Organic Chemistry

15 Nov – 15 Dec 2020

Evaluation of Organophosphate Pesticide Residues in Food using the Partial Least Squares Method

Gheorghe Ilia and Simona Funar-Timofei*

Coriolan Dragulescu Institute of Chemistry, Romanian Academy,

Bul. Mihai Viteazu 24, 300223 Timisoara, Romania

timofei@acad-icht.tm.edu.ro

The 24th International Electronic Conference on Synthetic Organic Chemistry

15 Nov – 15 Dec 2020

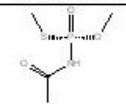
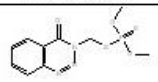
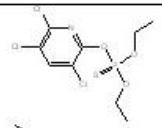
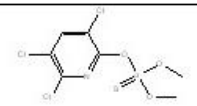
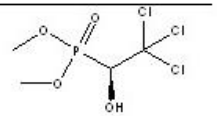
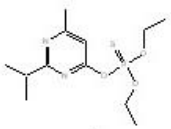
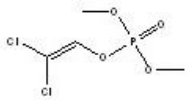
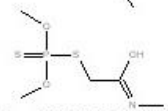
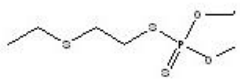
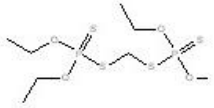
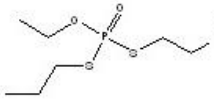
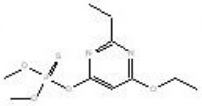
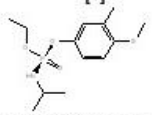
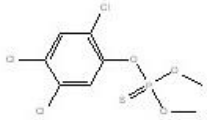
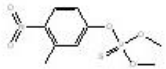
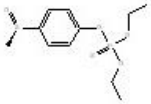
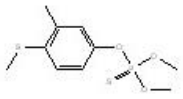
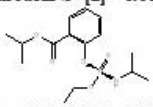
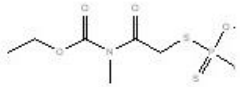
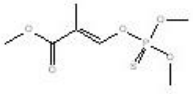
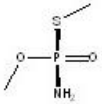
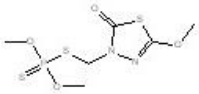
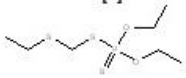
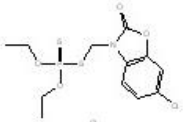
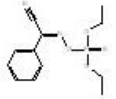
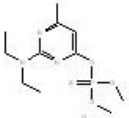
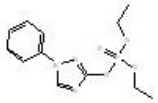
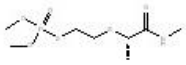
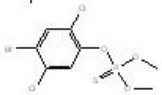
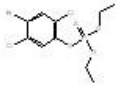
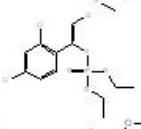
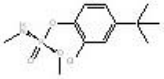
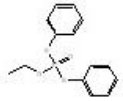
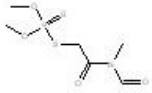
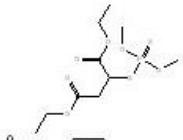
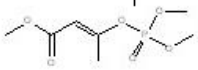
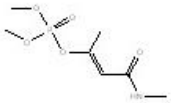
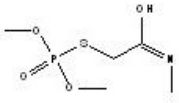
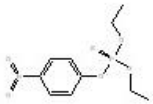
ABSTRACT

Organophosphourus (OP) chemicals were broadly used as insecticides and in the treatment of human diseases such as the malaria mosquitoes, parasitosis, myasthenia, and glaucoma. The OP toxicity is well known, they can cause environmental and health problems and the possibility to accumulate in the food chain. The acceptable daily intake (ADI) can be considered as a measure of the effect of pesticide residues in food on human health. In this paper the partial least squares (PLS) approach is used to evaluate the ADIs (expressed as pADIs) of a series of 46 structurally diverse OPs. OP structures were pre-optimized using the MMFF94s force field and structural descriptors were calculated for the minimum energy conformers. This dataset was divided into 26 training compounds, and 20 pesticides were included in the prediction set. Several criteria to check the model robustness, overfitting and the potential outliers in the X and Y space were employed. The PLS results indicate that new experimental toxicological data would be needed for five out of the 46 OPs, to improve their known ADI values, for qualitative and quantitative dietary long-term risk assessments.

AIM

- **Partial Least Squares (PLS) study of a series of organophosphourus (OP) pesticides. Structural pesticide parameters calculated for the minimum energy conformers are related to the accessible daily intake (expressed as pADI), as pesticide residue in food.**

Organophosphourus pesticide structures[1]

No	Structure	No	Structure	No	No	No	No
1*		2*		3	4* 	5* 	46* 
6 	7* 	8 	9*, **  PmodXPS+[2]= 0.039	10 			
11 	12 	13 	14*, **  PmodXPS+[2]= 0.016	15* 			
16 	17 	18 	19*,**  PmodXPS+[2]= 0.016	20 			
21 	22 	23* 	24 	25 			
26* 	27 	28 	29* 	30* 			
31 	32 	33* 	34 	35 			
36 	37 	38* 	39* 	40 			
41	42*,**	43*	44	45*,** PmodXPS+[2]= 0.006			

1. http://www.inchem.org/pages/pims.html

.

*** test compounds**

**** outliers in the X space**

METHODS

Dataset

- The OP derivatives were taken from literature.[1]
- The acceptable daily intake (ADI) values (mg/kg bodyweight), molar converted to pADIs, were used as the dependent variable.
- The OP structures were pre-optimized using the MMFF94 molecular mechanics force field of OMEGA v.2.5.1.4 (OpenEye) tool.[2,3] The default parameters were employed excepting the maximum number of conformers per compound set to 400 and an RMSD value of 0.5 Å.

Molecular descriptors

- The conformers of minimum energy were then used to calculate the structural parameters, using the DRAGON (Dragon Professional 5.5, 2007, Talete S.R.L., Milano, Italy) and InstanJChem (Instant JChem (2012) version 20.15.0, Chemaxon, http://www.chemaxon.com) software.

2. Hawkins, P.C.D.; Skillman, A.G.; Warren, G.L.; Ellingson, B.A.; Stahl, M.T. *J. Chem. Inf. Model.* 2010, 50, 572–584.
3. Hawkins, P.C,D.; Nicholls, A. *J. Chem. Inf. Model.* 2012, 52, 2919-2936.

METHODS

- **The Partial Least Squares (PLS) approach[4] was employed to relate the pADIs to the calculated structural descriptors, using the SIMCA (SIMCA P+12 12.0.0.0 2008, Umetrics, Sweeden, www.umetrics.com) program. The Variables Importance in the Projection (VIP) values and the sign of the variables' coefficients were used to explain the activity mechanism.**
- **The leave-7-out crossvalidation procedure was employed to select the most significant principal components and to check the internal model validation.**
- **Several criteria to check the potential outliers in the X and Y space were employed in the training and prediction sets: the score scatter plot, at the significance level of 0.05, the distance to the model in X space (for the selected dimension), for the observations used to fit the model (DmodX, with a significance level of 0.05), the probability of belonging to the model in the X space, for new observations in the prediction set combined with Hotelling's T^2 when the latter is outside the critical limit (PmodXPS+), The Hotelling's T^2 Range plot (which displays the distance from the origin in the score space for each selected observation, with a significance limit of 0.01).**

4. Wold, H., in: Kotz, S. and Johnson, N.L. (Eds.), Encyclopedia of Statistical Sciences, (Vol. 6), Wiley, New York, pp. 581 (1985).

RESULTS AND DISCUSSION

- **A robust model with two significant principal components which explains 88% of the information content of the descriptor matrix (for 16 structural descriptors), with $R^2Y(CUM) = 0.81$ and $Q^2(CUM) = 0.77$ was obtained.**

- **The normal distribution pattern of descriptors[5] of the training and prediction sets was checked with a probability of 90% to find the X-outliers (for the training set) and the prediction compounds residing outside the AD, using the descriptor pool of the training and prediction set (included in the best PLS model). According to this criterion, compound 25 was found as potential outlier for the training set. This assumption was not confirmed by the PModXPS+ criterion, according to which compounds: 9, 14, 19, 42, and 45 do not belong to the prediction X space.**

5. Wold, H., in: Kotz, S. and Johnson, N.L. (Eds.), Encyclopedia of Statistical Sciences, (Vol. 6), Wiley, New York, pp. 581 (1985).

RESULTS AND DISCUSSION

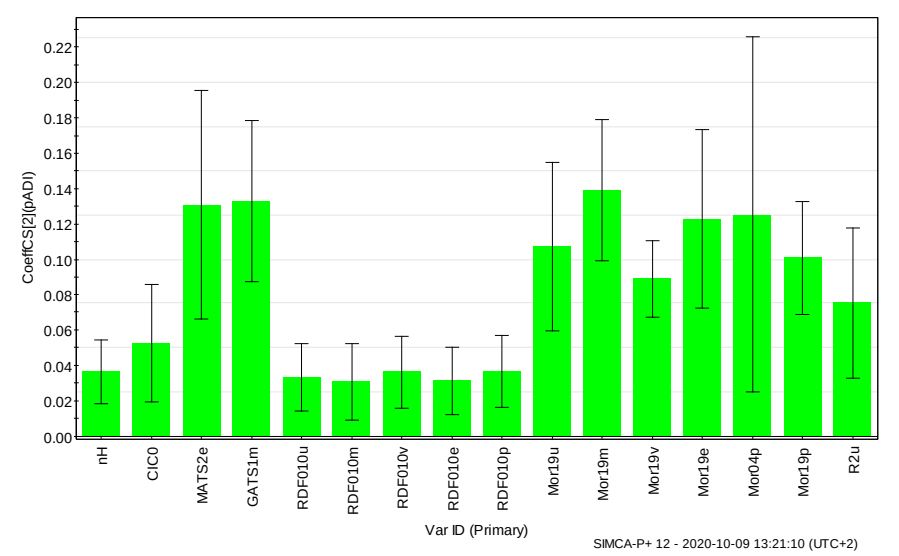


Figure 1. Coefficient plot of the final PLS model.

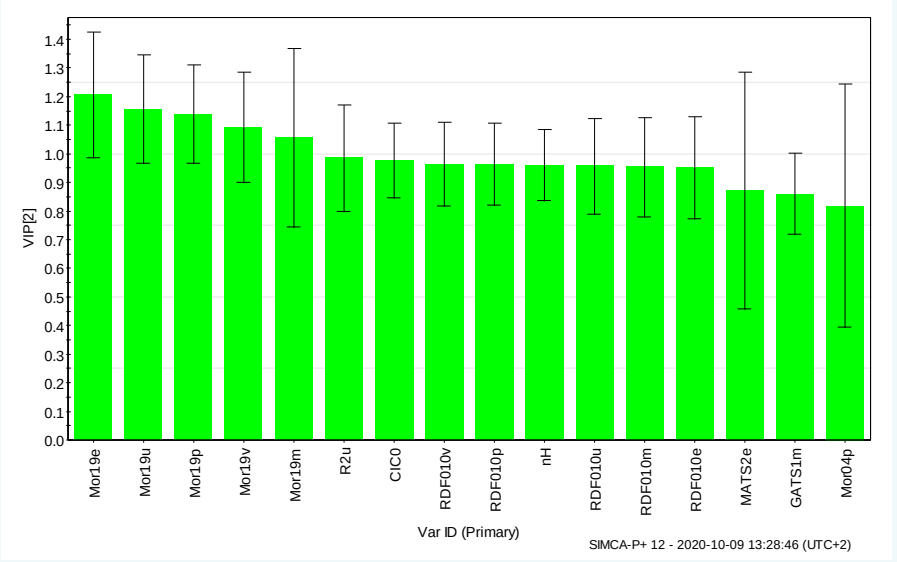


Figure 2. VIP plot for the final PLS model.

RESULTS AND DISCUSSION

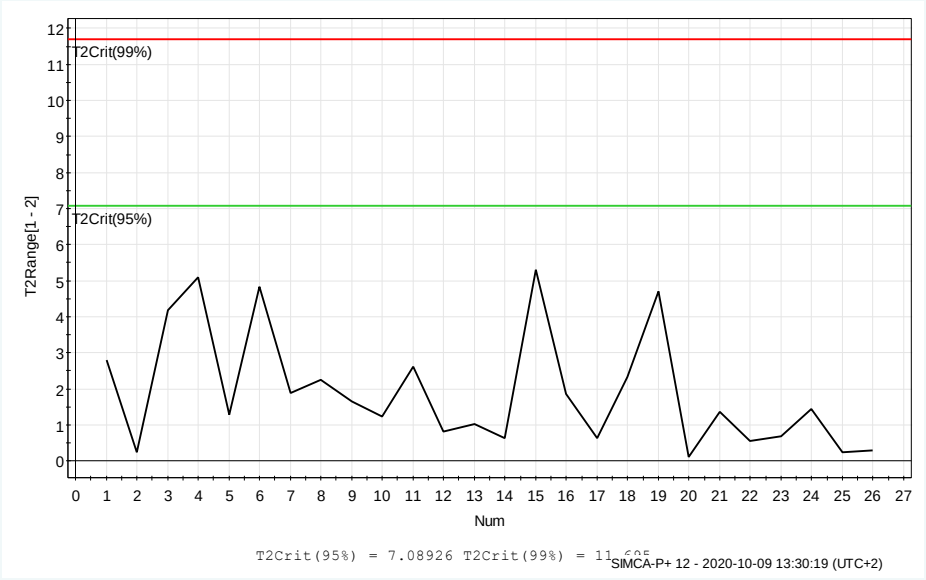


Figure 3. The Hotelling's T^2 range plot of the PLS model.

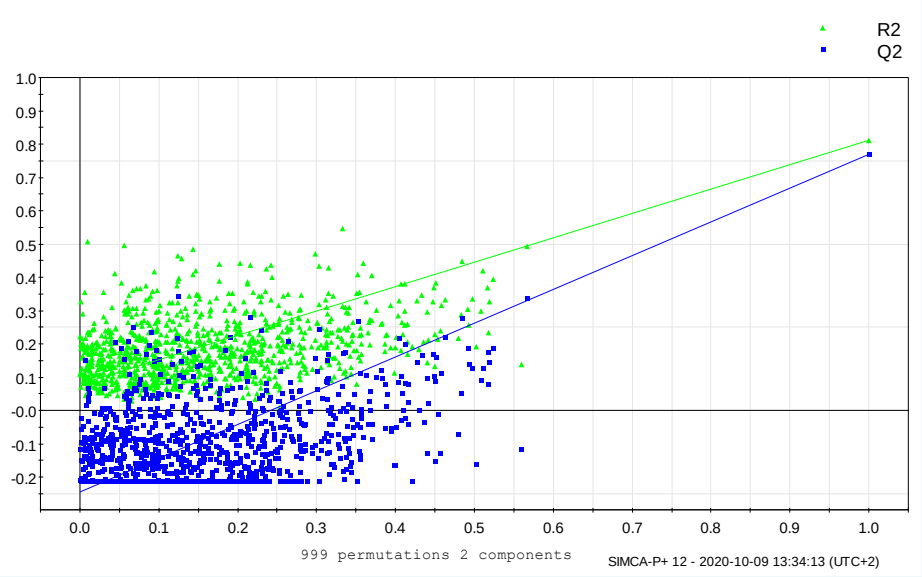


Figure 4. Y-scramble plots for the PLS model.

RESULTS AND DISCUSSION

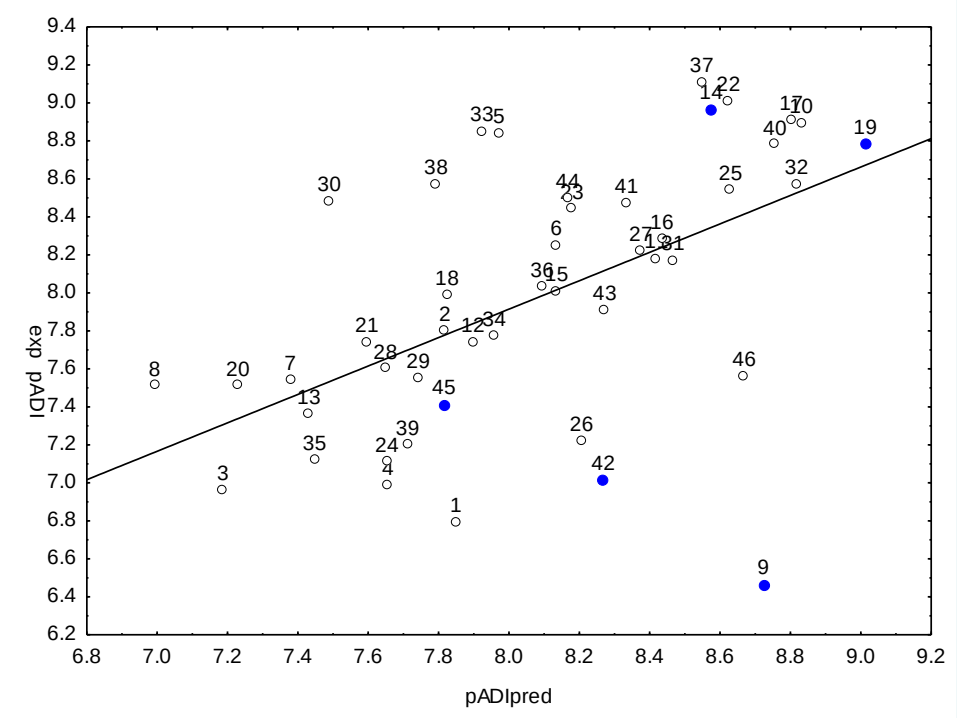


Figure 5. Experimental versus calculated pLC_{50} values (blue circles-outliers in the X space).

RESULTS AND DISCUSSION

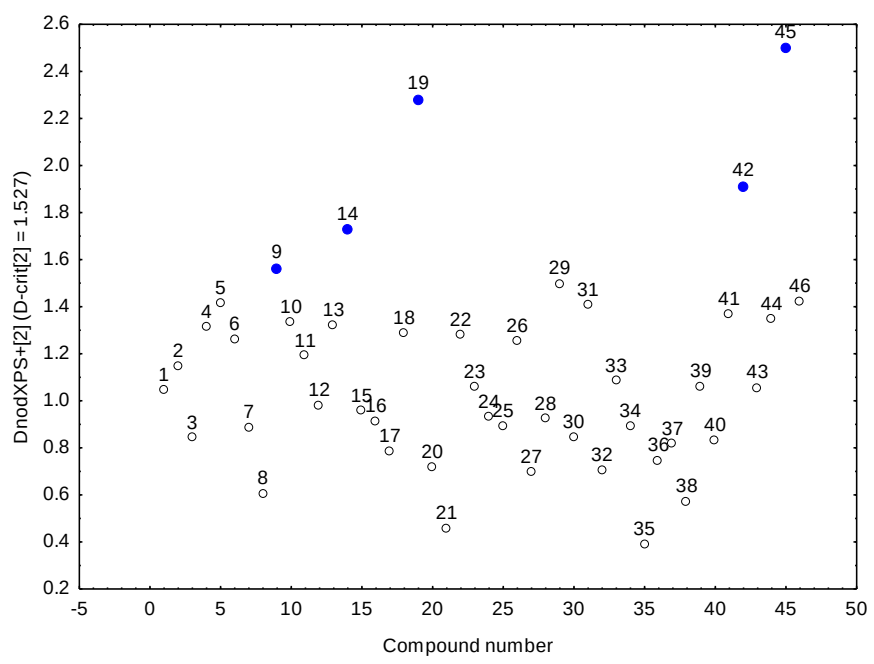


Figure 6. DmodX plot of the final PLS model (D-crit[2] = 1.527). Outliers in the X space are marked as blue circles.

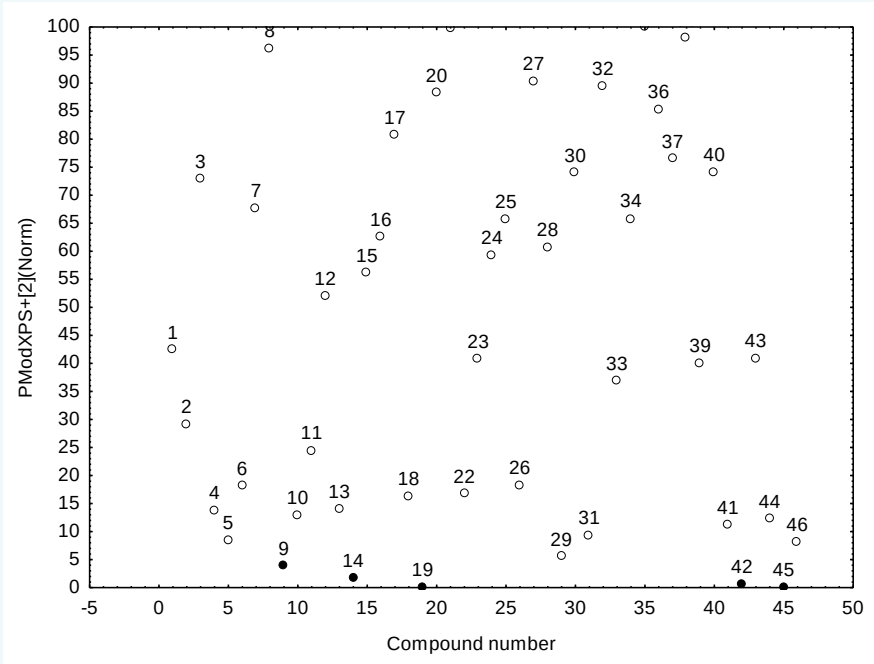


Figure 7. PmodXPS+ plot of the final PLS model. Outliers in the X space are marked as black circles.

CONCLUSIONS

- **The partial least squares (PLS) approach was used to study the acceptable daily intake (ADI), considered to be a measure of qualitative and quantitative dietary long-term risk assessments of a series of 46 organophosphourus pesticides.**

- **The OP structures were pre-optimized using the MMFF94 molecular mechanics force field.**

- **Structural descriptors derived from the minimum energy conformers were related to the pADI values.**

- **Several criteria to verify the model stability and the potential outliers in the X and Y space were applied to establish if new experimental toxicological data would be needed for this dataset.**

- **Five OPs were found as potential outliers in the X and Y space and new ADIs would be needed to be established for these compounds.**

Acknowledgments

This project was financially supported by Project 1.1 of the Coriolan Dragulescu Institute of Chemistry of the Romanian Academy. Access to the OpenEye Ltd., and Chemaxon Ltd. software are greatly acknowledged by the authors.