Distribution of QSARs correlation coefficients

Lorentz JÄNTSCHI[a], Sorana D. BOLBOACĂ[b], Mircea V. DIUDEA[c], Radu E. SESTRAŞ[d]

[a] *Tehnical University of Cluj-Napoca, Cluj, Romania*

[b] *"Iuliu Haţieganu" Medicine and Pharmacy University of Cluj-Napoca, Cluj, Romania*

[c] *"Babeş-Bolyai" University Cluj-Napoca, Cluj, Romania*

[d] *Agricultural Sciences and Veterinary Medicine Cluj-Napoca, Cluj, Romania*

Abstract

For set of ordnance compounds, with size ranging from 5 (one case) to 8 (11 cases) were previously studied the effect in the aquatic environment (on the marine life). For a number of 24 biological activities at which a computed property were added (OPLS-SAE) were created the pool of simple linear regressions relating the activities with the structure by using the MDFV methodology (implies creating the pool of descriptors and applying of adaptation criteria to the descriptors). From the pool of simple linear regressions were selected the ones being statistically significant (probability to reject the linear model less or equal to 0.05). A study regarding the distribution of the correlation coefficients for the selected simple linear regressions was conducted. Three distribution laws were proved to be relevant for the population of correlation coefficients (Beta, Generalized Pareto and Pert). For these three distributions a further study were conducted regarding the classification of the activities into one or another distribution. The study shown that the Chi-Square statistic was the best classifier of the activities: only one disagreement with the true positive classification simultaneous independent events of all three. The outcome of the study is that 48% of the correlation coefficients populations of significant simple linear regressions are Beta distributed, 40% are Generalized Pareto, and the rest - 12% - including the OPLS-SAE, are Pert distributed.

Keywords:

Structure-Activity Relationships; Distribution laws; Correlation coefficients

Acknowledgments:

This paper is dedicated to Professor Ante Graovac on the occasion of his 65th birthday.

References:

Jäntschi L, 2000. Prediction of Physical, Chemical and Biological Properties using Mathematical Descriptors. PhD Thesis (Chemistry) - supervisor Prof. Diudea MV. "Babeş-Bolyai" University Cluj-Napoca, Cluj-Napoca, January 2000. http://getcited.org/pub/103434624

Jäntschi L, 2009. Distribution Fitting 1. Parameters Estimation under Assumption of Agreement between Observation and Model, Bulletin of University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca. Horticulture, 66(2):684-690. http://arxiv.org/abs/0907.2829

Jäntschi L, Bolboacă SD, 2009. Distribution Fitting 2. Pearson-Fisher, Kolmogorov-Smirnov, Anderson-Darling, Wilks-Shapiro, Kramer-von-Misses and Jarque-Bera statistics, Bulletin of University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca. Horticulture, 66(2):691-697. http://arxiv.org/abs/0907.2832

Bolboacă SD, Jäntschi L, 2009. Comparison of QSAR Performances on Carboquinone Derivatives, TheScientificWorldJOURNAL, 9(10):1148-1166. http://dx.doi.org/10.1100/tsw.2009.131

Jäntschi L, 2010. Genetic Algorithms and their Applications. PhD Thesis (Horticulture) - supervisor Prof. Sestraş RE. Agricultural Sciences and Veterinary Medicine Cluj-Napoca, Cluj-Napoca, July 2010. http://getcited.org/pub/103483641