QSARINS: SOFTWARE FOR THE DEVELOPMENT, ANALYSIS AND VALIDATION OF MLR MODELS AND QSARINS-CHEM: INSUBRIA DATASETS AND QSA(P)R MODELS FOR ENVIRONMENTAL POLLUTANTS

PAOLA GRAMATICA, NICOLA CHIRICO, ALESSANDRO SANGION,STEFANO CASSANI

QSAR Research Unit in Environmental Chemistry and Ecotoxicology, DiSTA, University of Insubria, Via J.H.Dunant 3, 21100 Varese, Italy, paola.gramatica@uninsubria.it. http://www.qsar.it.

If correctly developed and validated, QSA(P)R models are highly useful for screening and prioritizing chemicals without experimental data, even before their synthesis: this can be done in the ”benign by design” approach of Green Chemistry. The use of QSAR in regulation is also advised by REACH, especially to reduce costs and number of tests on animals. Particular attention has been devoted to the validation of QSAR models, and the “OECD principles for the validation of QSARs models for their application in regulation” (OECD 2004) have been established to increase the prediction reliability.

In this poster, the new software QSARINS (QSAR-INSubria) (GRAMATICA et al. 2013), for the development of Multiple Linear Regression (MLR) models, by Ordinary Least Squares (OLS) and Genetic AIgorithm (GA) for variable selection, is presented. This program is a user-friendly platform for QSAR modeling in agreement with the OECD Principles, mainly focused on validation (internal and external) by different statistical parameters. Additional features include tools for explorative analysis by Principal Component Analysis (PCA), splitting of the dataset, analysis of the Applicability Domain (e.g., detection of outliers and interpolated or extrapolated predictions), combined modeling, selection of the best model by Multi-Criteria Decision Making (MCDM) and various useful and informative plots.
QSARINS-Chem (GRAMATICA et al. 2014), a specific module of QSARINS, includes several datasets of environmental pollutants with the relative chemical structures and the corresponding end-points (physico-chemical properties and/or biological activities), modeled by Insubria group during the last two decades. This database can be accessed in different ways (by CAS RN, SMILES, names, etc.) and the 3D structure of stored chemicals can be visualized. Additionally, some QSA(P)R models, based on molecular descriptors calculated by the free open source software PaDEL-Descriptor (YAP 2011), are implemented in QSARINS-Chem, therefore they are available for predicitons of new chemicals without experimental data, checking the reliability with the Applicability Domain study. Among these models, there is the Insubria Persistent Bioaccumulative and Toxic (PBT) Index model for the prediction of the cumulative PBT behavior of molecules, recently applied to screen thousands of possible PBT compounds (GRAMATICA et al. 2015). The QSAR Model Reporting Format (QMRF) of all these QSA(P)Rs are available. QSARINS-Chem can be also used as a management tool of personal datasets and models and additional chemometric analyses can be done by PCA and MCDM in order to prioritize the most hazardous chemicals.