QSARINS (QSAR-INSubria)

Software for QSAR MLR model development and validation

From ideas of Prof. Paola Gramatica, applied in “QSAR Research Unit in Environmental Chemistry and Ecotoxicology” at University of Insubria, Italy

How to cite:

Each user of QSARINS (whether for QSAR models’ development/validation or for application of QSARINS-Chem models or for use of the stored Insubria datasets or for analysis and validation of personal models) must cite the web: www.qsar.it and the following papers, where the software is presented:


The provider Prof. Paola Gramatica should also be acknowledged for the free license.
General information about QSARINS

**QSARINS (QSAR-INSubria)** is a software for the development and validation of Multiple Linear Regression (MLR) models by Ordinary Least Squares (OLS) and Genetic Algorithm (GA) for variable selection, based on QSAR experience of Prof. Paola Gramatica since 1995 and developed by Nicola Chirico (2008-2012). It is implemented according to the statistical/chemometric approach for QSAR models’ predictivity (Tropsha et al. 2003; Gramatica 2007, 2009, 2012, 2013, 2014, 2016; Chirico and Gramatica 2011, 2012; Gramatica et al. 2012; Gramatica and Sangion 2016), which is applied by the QSAR group of the University of Insubria in all the modeling researches.

In the module **QSARINS-Chem** the 3D-chemical structures of many Insubria datasets are stored, in addition to several QSAR models, based on PaDEL Descriptors (Yap 2011), developed by the Insubria QSAR group and applicable also to new chemicals, directly in QSARINS, where the PaDEL Descriptors can be calculated.

The following scheme (from Gramatica et al. 2012) summarizes the procedure for a validated QSAR modeling, applicable in QSARINS:
**QSARINS** provides a user-friendly interface that includes tools for:

- Data normalization;

- Dataset analysis (by Principal Component Analysis (PCA), etc.);

- Splitting of the data sets (by Random, Ordered response or Structure by PCA of molecular descriptors);

- OLS Model development by All Subset Selection;

- Genetic Algorithm (GA) for Variable Selection (with tracing of the models while being developed);

- Analysis of all individual variables, as well as those more frequently selected in the models;

- Internal validation, including the following criteria: $Q^2_{\text{LOO}}$, $Q^2_{\text{LMO}}$, Concordance Correlation Coefficient (CCC), RMSE, MAE, Y-Scrambling, etc...;

- External validation, including the following criteria: $Q^2_{F1}$, $Q^2_{F2}$, $Q^2_{F3}$, Golbraikh and Tropsha parameters, $r^2_m$ metrics, $\text{CCC}_{\text{EXT}}$, RMSE, MAE;

- Structural applicability domain by the leverage from the diagonal values of the Hat matrix (including Williams plot for chemicals with experimental data, y axis: standardized residuals and Insubria graph for chemicals without data, y axis: predicted values);

- Principal Components (PCs) Regression;

- Multi-Criteria Decision Making (MCDM) for selection of the best models;

- Consensus modeling (weighted and not, different tools for selection of models to be average);

- Check and validation of imported single models;

- Calculation of molecular descriptors and fingerprints with the open source software PaDEL-Descriptor (Yap 2011) (version currently included: 2.21).

- Application of several QSAR/QSPR models for environmental pollutants, developed by PaDEL-Descriptor software and stored in the QSARINS-Chem module, with the corresponding QMRF (QSAR Model Reporting Format). In particular: the PBT Index model (Papa and Gramatica 2010), re-developed (Gramatica et al. 2013, 2015) using PaDEL-Descriptor (Yap 2011);

- Database of compounds used to develop Insubria QSAR models: chemicals can be explored in different ways (CAS, SMILES, names) and visualized in 3D;

- Ranking of chemicals, based on PCA and MCDM
All outcomes of QSARINS are reported in the software with corresponding plots for quick graphical interpretation.

Example of some available plots:
SCREENSHOTS

DATA SETUP - Selection of variables and response. Three different splitting (Random, by ordered response, by structure based on PCA of molecular descriptors). PCA of the selected descriptors and generation of PC scores as variables for Principal Components Regression.
SINGLE MODEL – Model’s parameters related to internal and external validation, predicted values, HAT values, and standardized residuals are calculated. In addition, it is possible to view the PCA and the correlation matrix of the modeling descriptors. If available, the QMRF can be exported. It is present for the large majority of the models available in QSARINS-Chem.

IMPORTANT INFORMATION: Any user can upload personal MLR models and use QSARINS to manage them for storing, visualization, validation, ranking etc..
QSARINS-Chem Module

QSARINS-Chem is the module where 3012 chemicals, collected from the literature, curated and modeled by the Insubria group, are available with their 3D structure and experimental responses. In addition, 45 QSAR models of environmental end-points for organic pollutants, based on free software for molecular descriptors (Yap 2011), are available. These models, supported by their QMRF, can be applied for any new chemical without experimental data or also not yet synthesized, verifying the structural applicability domain by the Insubria graph.

List of QSAR models in QSARINS-Chem:

2. BFR MP (Papa et al. 2009; Gramatica et al. 2014) 
3. BFR VP (Papa et al. 2009; Gramatica et al. 2014) 
5. BTAZ Log Kow (Bhhatarai and Gramatica 2011a; Gramatica et al. 2014) 
6. BTAZ MP (Bhhatarai and Gramatica 2011a; Gramatica et al. 2014) 
7. BTAZ O. mykiss tox (Cassani et al. 2013; Gramatica et al. 2014) 
8. BTAZ P. subcapitata tox (Cassani et al. 2013; Gramatica et al. 2014) 
9. BTAZ Sw (Bhhatarai and Gramatica 2011a; Gramatica et al. 2014) 
10. BTAZ VP (Bhhatarai and Gramatica 2011a; Gramatica et al. 2014) 
11. Endocrine Disruptor Chemicals (EDC) Estrogen Receptor Binding (Li and Gramatica 2010; Gramatica et al. 2014) 
12. Esters D. magna tox (Papa et al. 2005a; Gramatica et al. 2014) 
13. Esters P. promelas tox (Papa et al. 2005a; Gramatica et al. 2014) 
14. Fish Biotrans. logHLn M1_day (Papa et al. 2014) 
15. Fish Biotrans. logHLn M2_day (Papa et al. 2014) 
16. Fish Biotrans. logHLn M3_day (Papa et al. 2014) 
17. Global Half-Life Index (GHLI) (Gramatica and Papa 2007; Gramatica et al. 2014) 
18. Human Biotrans. logHLB1_h (Papa et al. 2018) 
19. Human Biotrans. logHLB2_h (Papa et al. 2018) 
20. Human Biotrans. logHLB3_h (Papa et al. 2018) 
22. Human Biotrans. logHLT_h (Papa et al. 2018) 
23. Log Koc of Pesticides (Gramatica et al. 2007a, 2014) 
24. Nitro PAH Mutagenicity (Gramatica et al. 2007b, 2014) 
25. Pimephales promelas tox (Papa et al. 2005b; Gramatica et al. 2014) 
26. Persistence Bioaccumulation Toxicity (PBT) Index (Papa and Gramatica 2010; Gramatica et al. 2013) 
27. Personal Care Products (PCP) Aquatic Toxicity Index-ATI (Gramatica et al. 2016) 
28. PCP D.magna acute tox (Gramatica et al. 2016) 
29. PCP P.promelas acute tox LogP-based (Gramatica et al. 2016) 
30. PCP P.promelas acute tox (Gramatica et al. 2016)
31. PCP *P.subcapitata* acute tox (Gramatica et al. 2016)*
32. PerFluorinated Chemicals (PFC) Mouse Inhalation tox (Bhhatarai and Gramatica 2010; Gramatica et al. 2014)§
33. PFC Rat Inhalation tox (Bhhatarai and Gramatica 2010; Gramatica et al. 2014)§
34. PFC Rat Oral tox (Bhhatarai and Gramatica 2011b; Gramatica et al. 2014)§
35. PFC Sw (Bhhatarai and Gramatica 2011c; Gramatica et al. 2014)§
36. PFC VP (Bhhatarai and Gramatica 2011c; Gramatica et al. 2014)§
37. Pharm. Aquatic Toxicity Index-ATI (Sangion and Gramatica 2016a)*
38. Pharm. *D.magna* acute tox (Sangion and Gramatica 2016a)*
39. Pharm. *O.mykiss* acute tox (Sangion and Gramatica 2016a)*
40. Pharm. *P.promelas* acute tox (Sangion and Gramatica 2016a)*
41. Pharm. *P.subcapitata* acute tox (Sangion and Gramatica 2016a)*
42. PPCP intertox *D.magna-O.mykiss* (Sangion and Gramatica 2016b)
43. PPCP intertox *D.magna-P.promelas* (Sangion and Gramatica 2016b)
44. PPCP intertox *O.mykiss-P.promelas* (Sangion and Gramatica 2016b)
45. PPCP intertox *P.promelas-O.mykiss* (Sangion and Gramatica 2016b)

§Models originally developed using proprietary software (DRAGON v 5.5 or former versions) and updated using PaDEL Descriptor software (models are described in Gramatica et al. 2014).

*Models reported in QSARINS-Chem Standalone version

**Additional information**

QSARINS can be used for every modeling work involving Multiple Linear Regression (MLR) calculations, based on Genetic Algorithm for variable selection and Ordinary Least Squares (OLS) as modeling method. However, any personal MLR model, even if developed by other software, can be analyzed by the plots available in QSARINS as well as validated by several statistical validation parameters (both for internal and external validation: see above list).

Other chemometric tools (Principal Component Analysis (PCA), Multicriteria Decision Making (MCDM)) for explorative analysis and ranking are also implemented, therefore QSARINS is not limited to QSAR studies.

It is also important to note that not only chemicals, but any kind of objects can be analyzed or modeled in a multivariate way.
Bibliography


QSARINS is currently used by the Insubria group and several international QSAR groups: about 500 free licenses active into 2018.

QSARINS citations

QSARINS has been cited in the following peer-reviewed publications (December 2018, source ISI Web of Knowledge and Google Scholar):


Chayawan, Vikas, 2016. Quantum-mechanical parameters for the risk assessment of multi-walled carbon-nanotubes: A study using adsorption of probe compounds and its application to...


Judycka, U., Jagiello, K., Gromelski, M., Bober, L., Błażejowski, J., Puzynt, T., 2018. Chemometric approach to correlations between retention parameters of non-polar HPLC columns and


Önlü, Serli, Saçan, M.T., 2016. An in silico approach to cytotoxicity of pharmaceuticals and personal care products on the rainbow trout liver cell line RTL-W1. Environmental Toxicology and Chemistry. https://doi.org/10.1002/etc.3663


Wang, Y., Bai, F., Cao, H., Li, J., Liu, H., Gramatica, P., 2015. A Combined Quantitative Structure-


