General information about QSARINS

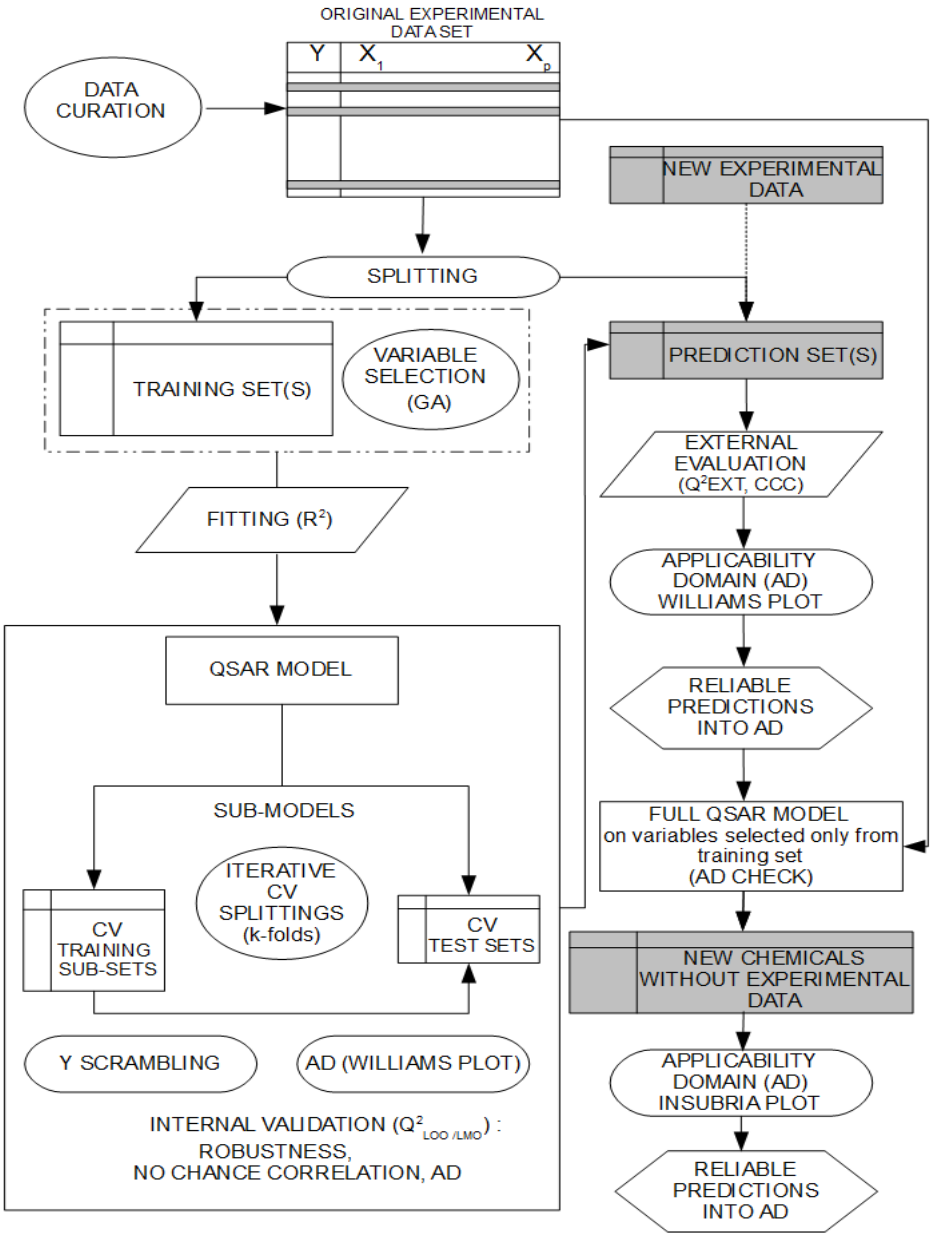
QSARINS (QSAR INSubria) is a software for the development and validation of Multiple Linear Regression (MLR) models by Ordinary Least Squares (OLS), developed by Nicola Chirico (2008-2014) and implemented according to the statistical approach for predictivity [Gramatica, 2007, 2009, 2012], which is applied by the QSAR group of the University of Insubria. The software is presented in the following paper:

Gramatica, P., Chirico, N., Papa, E., Kovarich, S., Cassani, S. QSARINS: A New Software for the Development, Analysis, and Validation of QSAR MLR Models. *Journal of Computational Chemistry*, *Software news and updates*, **2013**, 34, 2121-2132, DOI: 10.1002/jcc.23361.

The new version of QSARINS, including the new module "QSARINS-Chem", is presented in the following paper:

Gramatica, P., Cassani, S., Chirico, N. QSARINS-Chem: Insubria Datasets and New QSAR/QSPR Models for Environmental Pollutants in QSARINS. *Journal of Computational Chemistry*, *Software news and updates*, **2014**, Early View, DOI: 10.1002/jcc.23576.

The following Scheme summarizes the procedure:



QSARINS provides a user-friendly interface that includes tools for:

- Data normalization;

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

- Splitting of the data sets (by Random, Sorted response or Structure);

- 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_{LOO}, $Q^2_{LMO,}$ 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, Concordance Correlation Coefficient (CCC);

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

- Multi-Criteria Decision Making (MCDM);

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

- Check and validation of imported single models;

- PC Regression;

- PBT Index model;

- Application of several QSAR/QSPR models developed using PaDEL-Descriptor [Yap, 2011] software and stored in QSARINS;

- Ranking chemicals, based on PCA and MCDM;

- Database of compounds used to develop Insubria models; chemicals can be explored in different ways.

All outcomes of QSARINS are reported in the software with corresponding plots for quick graphical interpretation.

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 modelling method. Other chemometric tools (Principal Component Analysis (PCA), Multicriteria Decision Macking (MCDM)) for explorative analysis and ranking are also implemented, thus it is not limited to Quantitative Structure-Activity Relationships (QSAR) studies. The objects studied in QSAR modeling are chemicals, but they could be any kind of objects in other modeling studies.

QSARINS-Chem is the module where 2780 chemicals, studied by the Insubria group, are available with their 3D structure and experimental responses. In addition, 23 QSAR models of environmental end-points, based

on a free software for molecular descriptors (PaDEL-Descriptor 2.18) are available. These models, supported by their QMRF, can be applied for any new chemical.

It is important to note that any user can upload personal data sets and models and use QSARINS to manage them for storing, visualization, modeling, ranking etc.

Citation and acknowledgements

In any publication, derived from the use of QSARINS software, this must be cited as:

Gramatica, P., Chirico, N., Papa, E., Cassani, S., Kovarich, S., QSARINS: A New Software for the Development, Analysis, and Validation of QSAR MLR Models. Journal of Computational Chemistry, Software news and updates, **2013**, 34, 2121-2132, DOI: 10.1002/jcc.23361.

and

Gramatica, P., Cassani, S., Chirico, N. QSARINS-Chem: Insubria Datasets and New QSAR/QSPR Models for Environmental Pollutants in QSARINS. *Journal of Computational Chemistry*, *Software news and updates*, **2014**, Early View, DOI: 10.1002/jcc.23576.

and the provider should be acknowledged and the website www.qsar.it should be cited.

References:

Paola Gramatica, 2007. Principles of QSAR models validation: internal and external
QSAR Comb.Sci. 26(5), 694-701

Paola Gramatica, 2009. Chemometric Methods and Theoretical Molecular Descriptors in Predictive QSAR Modeling of the Environmental Behaviour of Organic Pollutants, Chapter 12 in *Recent Advances in QSAR Studies ,* Tomasz Puzyn - Jerzy Leszczynski - Mark T.D. Cronin Eds., (Challenges and Advances in Computational Chemistry and Physics), Springer-Verlag New York Inc, Nov. pp. 327-366.

Paola Gramatica, 2012. Modeling Chemicals in the Environment. Chap. 17 in *Drug Design Strategies-Quantitative Approaches,* D.J.Livingstone and A.M.Davies Eds., RSC Pub., pp. 458-478.

Yap Chun Wei, 2011. PaDEL-Descriptor: An open source software to calculate molecular descriptors and fingerprints. *J.Comput. Chem. 32 (7), 1466-1474*

QSARINS is currently used by the Insubria group and has been cited in the following peer-reviewed publications:

Nicola Chirico and Paola Gramatica (2011)
Real External Predictivity of QSAR Models: How To Evaluate It? Comparison of Different Validation Criteria and Proposal of Using the Concordance Correlation Coefficient.
J. Chem. Inf. Model. 51 (9), pp 2320–2335.

Nicola Chirico and Paola Gramatica (2012)
Real External Predictivity of QSAR Models. Part 2. New inter-comparable thresholds for different validation criteria and the need for scatter plot inspection.
J. Chem. Inf. Model 52 (8), pp 2044–2058

Paola Gramatica, Stefano Cassani, Partha Pratim Roy, Simona Kovarich, Yap Chun Wei, and Ester Papa (2012)
QSAR modeling is not "push a button and find a correlation": a case study of acute toxicity of (benzo-)triazoles on algae.
Mol.Inf. 31, pp 817-835.

Paola Gramatica (2013)
On the Development and Validation of QSAR Models.
Chap. 21 in Computational Toxicology: Volume II, Methods in Molecular Biology, vol. 930, Brad Reisfeld and Arthur N. Mayeno (eds.), 2012, DOI 10.1007/978-1-62703-059-5_21, Springer Science+Business Media, LLC, N.Y. (USA), 2013, pp 499-526.

Stefano Cassani, Simona Kovarich, Ester Papa, Partha Pratim Roy, Leon van der Wal, and Paola Gramatica (2013)
Daphnia and fish toxicity of (benzo)triazoles: validated QSAR models, and interspecies quantitative activity-activity modelling.
J.Hazard Mat. 258-259, pp 50-60.

Stefano Cassani, Simona Kovarich, Ester Papa, Partha Pratim Roy, Magnus Rahmber2, Sara Nilsson, Ullrika Sahlin, Nina Jeliazkova, Nikolay Kochev, Ognyan Pukalov, Igor V. Tetko, Stefan Brandmaier, Mojca Kos Durjava, Boris Kolar, Willie Peijnenburg and Paola Gramatica (2013)
Evaluation of CADASTER QSAR models for aquatic toxicity of (benzo-)triazoles and prioritization by consensus.
ATLA, 41, pp 49-64.

Vikas, Reenu, Chayawan(2013)
Does electron-correlation has any role in the quantitative structure–activity relationships?

J.Mol.Graph. Model. 42, pp 7-16.

Ester Papa, Simona Kovarich, and Paola Gramatica (2013)
QSAR prediction of the competitive interaction of emerging halogenated pollutants with human transthyretin
SAR QSAR Environ. Res., 24 (4), pp 599-615.

Jiazhong Li, Huanxiang Liu, Xing Huo, Paola Gramatica (2013)
Structure-Activity Relationship Analysis of the Thermal Stabilities of Nitroaromatic Compounds Following Different Decomposition Mechanisms.
Mol.Inf. 32, pp 193-202.

Vijay H. Masand et al. (2013)
Does tautomerism affect the outcome of QSAR modeling?
Med. Chem. Res. Just accepted paper

Gulcin Tugcu, Birkan Yilmaz, Melek Türker Saçan (2013)
Comparative performance of descriptors in a Multiple Linear and Kriging Models: A case study on the acute toxicity of organic chemicals to algae
Poster presented at ICCE 2013, Barcelona, Spain. 25-28 June 2013.

Jiazhong Li, Shuyan Li, Chongliang Bai, Huanxiang Liu, Paola Gramatica (2013)
Structural requirements of 3-carboxyl-4(1H)-quinolones as potential antimalarials from 2D and 3D QSAR analysis
J. Mol. Graphics Modell. 44, pp 266-277.

Simona Funar-Timofei, Sorin Avram, Ana Borota (2013)
Structure-Toxicity Study of Some Pyrethroidal Esters Insecticides.
Proceedings of the 17th Int. Electron. Conf. Synth. Org. Chem., 1-30 November 2013; Sciforum Electronic Conference Series, Vol. 17, **2013**, e012; doi:10.3390/ecsoc-17-e012, http://www.sciforum.net/conference/ecsoc-17/paper/2250

Simona Funar-Timofei, Liliana Halip, Ana Borota, Alina Bora (2013)
Quantitative Structure-Activity Study of Some Cyclin-Dependent Kinase 5/P25 Inhibitors
Timisoara's Academic Days, XIIIth Edition. New Trends and Strategies in the Chemistry of Advanced materials with relevance in Biological Systems, technique and Environmental Protection, 13-14 June 2013, Timisoara, Romania, ISSN: 2065-0760, p. 46, poster.

Vijay H. Masand, Devidas T. Mahajan, Taibi Ben Hadda, Rahul D. Jawarkar, Hemant Chavan, B. P. Bandgar, Harsh Chauhan (2014)
Molecular Docking and Quantitative Structure Activity Relationship (QSAR) Analyses Of Indolylarylsulfones As HIV-1 Non-Nucleoside Reverse Transcriptase Inhibitors

*Med. Chem. Re*s. 23, pp 417-425. DOI: 10.1007/s00044-013-0647-8

Yonghua Shao, Jining Liu, Meixia Wang, Lili Shi, Xiaojun Yao, Paola Gramatica (2014)
Integrated QSPR models to predict the soil sorption coefficient for a large diverse set of compounds by using different modeling methods
Atmos. Environ. Accepted, in press. doi: 10.1016/j.atmosenv.2013.12.018.

Ester Papa, Leon van der Wal, John. A. Arnot, and Paola Gramatica (2014)
Metabolic biotransformation half-lives in fish: QSAR modeling and consensus analysis
Sci. Total Environ. 470-471, pp 1040-1046.

Vikas, Chayawan (2014)
Externally predictive quantitative modeling of supercooled liquid vapor pressure of polychlorinated-naphthalenes through electron-correlation based quantum-mechanical descriptors
Chemosphere 95, pp 448-454.