

UK QSAR & Chemoinformatics

Spring Meeting

14th April 2005

The Quality of QSAR Models: Myth and Reality

Evgueni Kolossov

QSAR Programme Manager

4th April 2005



- About 20 years ago T.Fujita published an article called “QSAR: Myth and Reality”.
- Today nobody is questioning the reality of QSAR.
- But what about the quality of QSAR models?
- Can we be sure we have the real criteria for this?
- What criteria is real and what is myth?

“All models are wrong but some are useful...”



Dr. Box is a Vilas Professor, the highest honour awarded to the faculty by the University of Wisconsin. Dr. Box is concerned with the planning and analysis of industrial experiments. The objective is to determine the important factors affecting product quality, then adjust them to their best levels. Recent work has concentrated on reduction of variance, as well as adjustment of mean levels. New Bayesian methods of analysis have been devised for highly fractionated designs. Studies of reduction of variance transmissions are in progress.

- A QSAR model should be associated with the following information:
 - A defined endpoint
 - An unambiguous algorithm
 - A defined domain of applicability
 - Appropriate measures of goodness-of-fit, robustness and predictivity
 - A mechanistic interpretation, if possible

[†] Formerly known as “Setubal principles“. OECD – Organisation for Economic Co-operation and Development

- A QSAR model should be associated with the following information:
 - A defined endpoint
 - An unambiguous algorithm
 - A defined domain of applicability
 - Appropriate measures of goodness-of-fit, robustness and predictivity
 - A mechanistic interpretation, if possible

[†] Formerly known as “Setubal principles”. OECD – Organisation for Economic Co-operation and Development

Goodness-of-fit, robustness and predictivity

$$D_C = D_D + D_M + D_A - C$$

Parameter	Myth	Reality
R^2	0 is worst, 1 is the best. Model is good if parameter over 0.6-0.7	Will improve as extra terms are added. Ignores systematic errors like constant shift. Not enough to make conclusions especially for PLS where the large number of components can lead to a very good correlation coefficient value
Q^2	1 is the best, lower values worse. Model is good if parameter over 0.6-0.7	Highly criticized during the last few years. Not actually applicable for the fragment-based models with a large number of fragments
$Q_{fp}^2 = r^2 - q^2$	0 is the best, higher values worse. Kind of instability measure.	If two previous parameters is not good enough, then the difference between them cannot be a good indicator as well
hi^2	0 is the best, higher values worse.	Depends on the number of points. Non-robust measure (can be affected by outliers).
SE (Standard error, standard deviation)	0 is the best, higher values worse.	Non-robust measure (can be affected by outliers).
MAR (Mean Absolute Residual)	0 is the best, higher values worse.	Non-robust measure (can be affected by outliers).
MSR (Mean Squared Residual)	0 is the best, higher values worse.	Non-robust measure (can be affected by outliers).
Intercept on Predicted vs Measured graph	0 is the best, absolute higher values worse.	Requires confidence limits evaluation as it may depend on the data range and data scatter.
Slope on Predicted vs Measured graph	1 is the best, lower/higher values worse.	Requires confidence limits evaluation as it may depend on the data range and data scatter.
F-Statistic significance	1 is best, 0 is worst	Depends on data range and data scatter.
t-test significance	1 is best, 0 is worst	Always 1 for linear models
Q-test significance	1 is best, 0 is worst	Ignores systematic errors like constant shift.

LOO:

- Advantage: it is reproducible
- Disadvantage:
 - ⋮ For large data sets, only a small part of the data set is omitted because it becomes computationally expensive
 - ⋮ The similar statistics obtained, with all the problems discussed above

LMO:

- Advantage: gives some confidence in the model
- Disadvantage:
 - ⋮ Needs to be performed many times
 - ⋮ No answer on by-chance correlation (can be tested by Y-randomization)
 - ⋮ No guarantee that the compounds taken out are inside the domain (the same with an external validation)

- Very simple yet effective tool
 - ... not only for model quality assessment
 - ... but also for assessment of over-fitting and under-fitting problems.
- More sensitive to model changes during the standard Leave-One-Out and Leave-Many-Out validation procedures than any other statistical coefficients
 - ... and can be used as a quantitative measure of model instability.
- The simplicity of the calculations suggests that these coefficients can be used for regulatory purposes during the QSAR model certification.

1. See Poster: QSAR Model Quality: A new approach to statistical analysis of model instability

Approaches:

- **Dwayne Moore¹: Mean rank for selected parameters for models in comparison – subjective selection of parameters**
- **John Dearden²: Percentage of compounds within the error ranges – required the same test set**
- **T.Schultz³: Confidence Index (CI) – nine factors used in equation in “normalized” form. Decision of what ‘normalised’ forms should be used for each factor was based on the expert opinions of the authors, in order to meet with their intuition of what should constitute high and low confidence/quality. Despite use of the term ‘confidence’, there is no attempt by the authors to place their normalisations on any statistical footing (to do so would not necessarily be appropriate, or even possible, given the nature of some of the measure of quality being used, in particular R_{conf} and S_{conf}).**

-
1. Moore, D.R.J. A Comparison of Model Performance for Six QSAR Packages that Predict Acute Toxicity to Fish, Environmental Canada, Ottawa, Ontario, Canada (2000).
 2. Dearden, J.C., Netzeva, T.I., Bibby, R. Comparison of a number of commercial software programs for the prediction of aqueous solubility, J.Pharm.Pharmacol., **2002**, 54 (Supplement), S-66.
 3. Schultz, T.W., Netzeva, T.I. and Cronin, M.T.D. (2004) “Quality and Confidence in QSARs”, SAR QSAR Environ. Res. **15**, 385-397.

$$D_C = D_D + D_M + D_A - C$$

- **PredictionBase: Model Quality Calculator (MQC) ->MQ²**
 - *Calculate the overall quality of a regression model by combining several existing measures.*
 - *Take into account weighted average.*
 - *Transformation applied automatically for known parameters like r^2 , q^2 , etc. to take into account the number of compounds in the model and the number of variables to allow a comparison between models with different number of compounds and variables.*
 - *For custom parameters, the weight, best and worst values are set by the user.*
 - *Each transformed measure must take non-negative values, with zero denoting perfection and higher values indicating worse quality.*
 - *All transformed measures must have the same units as one another. We adopt the convention that these measures are dimensionless, in other words that they are unchanged when multiplying the activities (or all the descriptors) by a constant.*
 - *The criterion of 'normalisation' can be satisfied if we express each measure relative to a prescribed confidence limit (for example, 95%). This ensures that a value of 1 in any transformed measure C has a common significance of 95%, say $P(C < 1) = 95\%$ under the null hypothesis that activity is independent of the descriptors (i.e., chance correlation).*
 - *Resulting index has values from 0 to 1.*

$$D_C = D_D + D_M + D_A - C$$

Calculator - Model Quality Calculator

Model Quality View Help

Quality Calculator

Calculator Graph

Model	# Structures	# Descriptors	r ²	q ² LOO	MIC LOO	q ² LMO	MIC LMO	MVIC LMO	test set q ²	Model Quality Index
LSF_F	1629	855	0.942586	0.942199	3.633	0.706453	0.808929	2.3628	-0.315723	0.910026
PLS_F	1657	43	0.977758	0.977541	80.4225	0.836097	14.991	14.1156	0.612013	0.81892

Model Parameters

- r²
- q² LOO
- MIC LOO
- q² LMO
- MIC LMO
- MVIC LMO
- test set q²

Model Parameter

Parameter Name:

Type

95% Conf. Value:

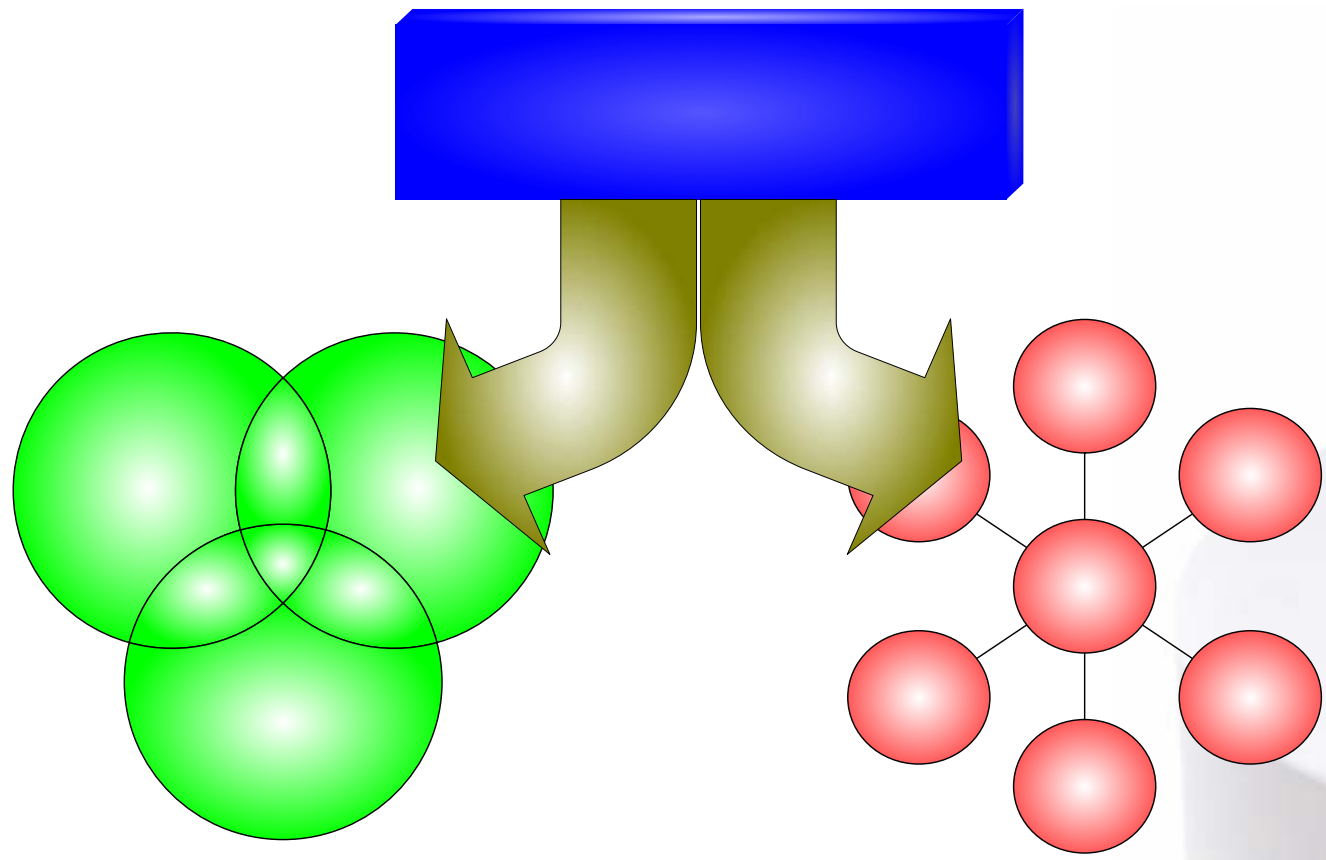
Best Value:

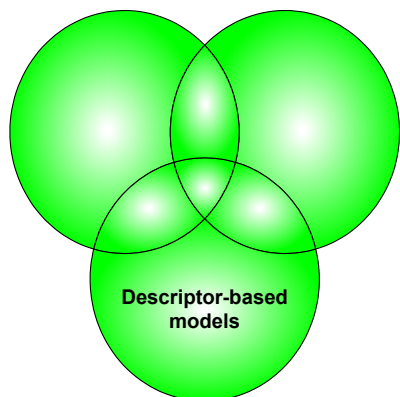
Weight:



Assessment of Domain of Applicability

$$D_C = D_D + D_M + D_A - C$$





➤ Many methods but none fully acceptable because of well-known problems:

- Shape of domain
- Poor handling of non-convex training sets
- Computational inefficiency *etc.* (Liverpool, 2004)

See presentation afternoon: Robert Stanforth, A QSAR Model's Domain of Applicability: Quantitative Measure of Distance from the Domain

- Traditional approach: a compound is outside the domain if any of the fragments generated from that compound are not present in the model's fragments set.
- This approach is unsatisfactory for the following reasons:
 - It does not take into account the total number of fragments generated – it is quite clear that the distance from the domain will depend critically on this factor.
 - It does not take into account whether the model contains similar fragments.
 - This definition provides no quantitative measure of the distance from the domain and/or an indication of errors in prediction.
- To find a satisfactory solution to these problems a new approach should be considered.

[†] UK Patent Application Number: GB 0506366.4. Filing date 30/03/05.

‘Error weight’ coefficient e_r from the diagonal of the inverse of the centered curvature matrix produced during the regression calculation:

$$(1) \quad e_r = E_{rr} \quad \text{where} \quad E = \left(X^T X - \frac{1}{N} \left(\sum_{k=1}^N \mathbf{x}_k \right) \left(\sum_{k=1}^N \mathbf{x}_k^T \right) \right)^{-1}$$

Where:

- N = number of structures in the model
- n = number of fragments in the model
- $X = [m_{kr}]$ - the ‘mapping count’: the number of times that fragment r occurs in structure k
- $\mathbf{x}_k^T = [m_{kr}]$ - the row of X containing the mappings for structure k

The coefficient e_r indicates the statistical variance in a_r relative to the model’s variance σ^2 . Thus, a mapping count of m_r for fragment r will contribute $m_r^2 e_r \sigma^2$ to the statistical variance of the predicted activity (this is a simplified formulation, ignoring correlation between the mappings of different fragments, aimed at highlighting the isolated effect of a single fragment). This statistical variance can be interpreted as ‘**uncertainty**’, or as the ‘**distance from the model**’, since a high value of e_r corresponds to little variation of the dataset with respect to m_r .

† UK Patent Application Number: GB 0506366.4. Filing date 30/03/05. .

For details see poster: **Assessment of Domain of Applicability for the Fragment-based QSAR Models**

When applying this fragment model to a new structure, some of the structure's fragments may not be present in the model. In these cases we can substitute the most similar fragment in the model, with similarity defined based on the occurrence of certain atom/bond configurations and measured by the Tanimoto or any other similarity coefficient, T . The dissimilarity $1-T$ gives rise to another source of **uncertainty**, which can also be interpreted as a '**distance from the model**'.

We require a method of combining both these uncertainties, over all fragments in a new structure, to seek a single measure d_{fm} of distance of the structure from the fragment model's domain. One such method is to incorporate the product of the '**error weight**' **uncertainty** and the '**dissimilarity**' **uncertainty** (for each fragment), which suggests the following measure:

$$(2) \quad d_{fm}^2 = \frac{1}{m} \sum_{r=1}^m (1 - T_r) m_r^2 e_r$$

where the sum is over all m fragments occurring in the structure and T_r is the similarity of fragment r to its closest match in the model. It is sensible to take the expression in (2) as defining a **squared** distance measure as it involves squared factors (m_r^2, E) in each term; this is analogous to standard Euclidean distance and is consistent with existing standard measures for a regression model such as r^2 and q^2 .

This method of combining the uncertainties means that only fragments that are not in the model contribute to d_{fm}^2 , the contribution of each of these fragments is proportional to its dissimilarity to the closest fragment in the model. It follows that if all fragments of a certain structure are in the model then $d_{fm} = 0$. This ensures that training structures have a distance of zero from the model.

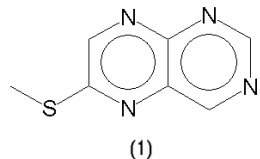
† UK Patent Application Number: GB 0506366.4. Filing date 30/03/05. .

For details see poster: **Assessment of Domain of Applicability for the Fragment-based QSAR Models**

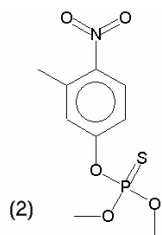
Fragment-based models: Example†

$$D_C = D_D + D_M + D_A - C$$

During external test set validation of a fragment-based model, for two following molecules (7-Methylthiopteridine (1) and Fenitrothion (2)) some generated fragments were not found in the model. Calculated values for both compounds are based on the sum of contributions from all fragments generated from the molecules:



Fragment	Contribution	Map Value	Sum	
1337	0.372844	4	1.49138	
1353	0.919518	2	1.83904	
10435	0.99403	1	0.99403	
10450	-1.96769	1	-1.96769	
10497	-0.517476	1	-0.517476	
10570	6.48473	1	6.48473	
10581	-4.44095	1	-4.44095	
10588	-2.05982	1	-2.05982	
11882	-1.94157	1	-1.94157	
new	-0.517476	1	-0.517476	Replaced by 10497
new	-2.05982	1	-2.05982	Replaced by 10588
new	-2.05982	1	-2.05982	Replaced by 10588
CONSTANT	-1.53666		-1.53666	
.....
RESULT			-6.29212	



Fragment	Contribution	Map Value	Sum	
426	-1.13723	1	-1.13723	
523	-0.658474	1	-0.658474	
8577	-1.2151	1	-1.2151	
10496	0.788824	1	0.788824	
10593	-3.67947	1	-3.67947	
10594	2.46572	1	2.46572	
10938	0.266484	1	0.266484	
11667	-0.169088	1	-0.169088	
11754	1.04783	1	1.04783	
11797	1.95595e-002	1	1.95595e-002	
11804	-0.478921	1	-0.478921	
new	0.582525	1	0.582525	Replaced by 430
new	-0.478921	1	-0.478921	Replaced by 11804
CONSTANT	-1.53666		-1.53666	
.....
RESULT			-4.18293	

Both structures contain approximately the same number of fragments (12 and 13). For 7-Methylthiopteridine, there are three fragments not found in the model. For Fenitrothion there are two missing fragments. In these cases, contributions of the most similar fragments from the model were used.

† UK Patent Application Number: GB 0506366.4. Filing date 30/03/05. .

For details see poster: **Assessment of Domain of Applicability for the Fragment-based QSAR Models**

Fragment-based models: Example†

$$D_C = D_D + D_M + D_A - C$$

Which predicted values can be trusted and which cannot?

To answer this question we need to calculate the distance of the compound from the domain of applicability using formula (2), based on the statistics:

Compound	Num. of	Replacement Fragments ID	Similarity to the original fragment	Mapping value	Error Weight	Contribution	dfm ²
	Fragments						
7-Methylthiopteridine	12	10497	0.606061	1	1.351	0.532211	
		10588	0.34466	1	167.9	110.0315	
		10588	0.36	1	167.9	107.456	
							218.0197
Fenitrothion	13	430	0.744186	1	0.768	0.196465	
		11804	0.779221	1	2.122	0.468493	
							0.664958

$$d_{fm}^2(\text{7-Methylthiopteridine}) = (1/12) * [(1-0.606061) * 1^2 * 1.351 + (1-0.34466) * 1^2 * 167.9 + (1-0.36) * 1^2 * 167.9] = (1/12) * [0.5322 + 110.0 + 107.5] = 218.0 / 12 = \mathbf{18.17}$$

$$d_{fm}^2(\text{Fenitrothion}) = (1/13) * [(1-0.744186) * 1^2 * 0.7680 + (1-0.779221) * 1^2 * 2.122] = (1/13) * [0.1965 + 0.4685] = 0.6650 / 13 = \mathbf{0.05115}$$

† UK Patent Application Number: GB 0506366.4. Filing date 30/03/05. .

For details see poster: **Assessment of Domain of Applicability for the Fragment-based QSAR Models**

$$D_C = D_D + D_M + D_A - C$$

Examination of d_{fm}^2 values for **7-Methylthiopteridine (18.17)** and **Fenitrothion (0.05115)** shows that Fenitrothion is very close to the domain of the model whereas 7-Methylthiopteridine is very far from this domain. The predicted value for 7-Methylthiopteridine should therefore be treated with extreme caution. Examination of the experimental values and residuals for these structures is further evidence of the validity of the method:

7-Methylthiopteridine: experimental value is **-1.55**; predicted value is **-6.292**; absolute residual value is **4.742 (305.9%)**.

Fenitrothion: experimental value is **-4.04**; predicted value is **-4.183**; absolute residual value is **0.1429 (3.54%)**.

Clearly the predicted value for Fenitrothion is much more accurate compared to that for 7-Methylthiopteridine which is outside the domain of the model.

† UK Patent Application Number: GB 0506366.4. Filing date 30/03/05. .

$$D_C = D_D + D_M + D_A - C$$

- We have introduced a simple and quantitative novel method for assessment of the domain of applicability for fragment-based QSAR models.
- This method takes into account the total number of fragments generated from a structure, the availability of replacement fragments in the model and their contribution weight values.
- Implementation of this method allows examination of the statistical significance of predictions and solves the problem of missing fragments, which is crucial for fragment-based QSAR models.
- A method of assessing if predicted structures are within the applicability domain of fragment-based models is an important milestone towards the use of QSAR models in regulatory assessment, providing a new metric to gauge the confidence of prediction.
- It is our hope that this will be a crucial step in the adoption of QSAR methods in regulatory usage and will lead to a reduction in the requirement for animal testing.

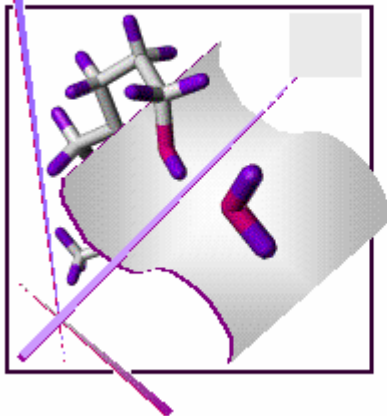
† UK Patent Application Number: GB 0506366.4. Filing date 30/03/05. .

For details see poster: **Assessment of Domain of Applicability for the Fragment-based QSAR Models**

- OECD principles need revising with a clear definition of the statistic parameters for quality assessment
- Model instability coefficients is recommended as one of the most important parameters for quality assessment
- MQ^2 is recommended as a unique measure of QSAR model quality
- The distance from the domain for the fragment-based models is recommended for the assessment of domain of applicability for this type of model.

➤ The author would like to thank:

- Dr. Andrew Lemon (IDBS)
- Dr. Andrew English (IDBS)
- Beth Thomas (IDBS)
- IDBS QSAR Team:
 - ⋮ Roy Saggars
 - ⋮ Virendra Mishra
 - ⋮ Andrei Groznyi
 - ⋮ Robert Stanforth
 - ⋮ Oliver Clancy



**UK QSAR
& Chemoinformatics**

Spring Meeting

14th April 2005