An Approach to Determining Applicability Domains for QSAR Group Contribution Models: An Analysis of SRC KOWWIN

Nina Nikolova-Jeliazkova[1] and Joanna Jaworska[2]

[1]Institute of Parallel Processing, Bulgarian Academy of Sciences, Sofia, Bulgaria; [2]Procter and Gamble, Eurocor, Strombeek-Bever, Belgium

Summary — QSAR model predictions are most reliable if they come from the model's applicability domain. The Setubal Workshop report provides a conceptual guidance for defining a (Q)SAR applicability domain. However, an operational definition is necessary for applying this guidance in practice. It should also permit the design of an automatic (computerised) procedure for determining a model's applicability domain. This paper attempts to address this need for models that use a large number of descriptors (for example, group contribution-based models). The high dimensionality of these models imposes specific computational restrictions on estimating the interpolation region. The Syracuse Research Corporation KOWWIN model for prediction of the n-octanol/water partition coefficient is analysed as a case study. This is a linear regression model that uses 508 fragment counts and correction factors as descriptors, and is based on the group contribution approach. We conclude that the applicability domain estimation by descriptor ranges, combined with Principal Component rotation as a data pre-processing step, is an acceptable compromise between estimation accuracy and the amount of data in the training set.

Key words: *applicability domain, group contribution method, KOWWIN, QSAR.*

Address for correspondence: *Joanna Jaworska, Procter & Gamble, Eurocor, Central Product Safety, 100 Temselaan, 1853 Strombeek-Bever, Belgium.*
E-mail: *Jaworska.j@pg.com*

Introduction

QSAR model predictions are most reliable if they come from the model's applicability domain (AD). The OECD includes AD assessment as one of the QSAR acceptance criteria for regulatory purposes (1). The Setubal Workshop report (2) offered the following guidance for AD assessment: *The applicability domain of a (Q)SAR is the physico-chemical, structural, or biological space, knowledge or information on which the training set of the model has been developed, and for which it is applicable to make predictions for new compounds. The applicability domain of a (Q)SAR should be described in terms of the most relevant parameters, i.e. usually those that are descriptors of the model. Ideally the (Q)SAR should only be used to make predictions within that domain by interpolation not extrapolation*. This description is helpful for explaining the intuitive meaning of the "applicability domain" concept. However, the practical assessment of ADs needs guidance pertaining to methods and boundary criteria.

To date, two approaches to AD estimation have been proposed. The first, applied in this paper, estimates the interpolation region based on the training set in the model's descriptor space. This approach has recently been reviewed for regression and classification models (3). The second approach relies on similarity analysis according to the premise that *a QSAR prediction is reliable if the compound is "similar" to the compounds in the training set*. "Similarity" is a relative, not an absolute, concept (i.e. chemical A can't be similar to chemical B in absolute terms, but only with respect to some measurable key feature or features). Because different concepts of similarity are relevant to different endpoints, the similarity approach to AD assessment should be based on features relevant to the endpoint modelled (4, 5). An interesting application of similarity assessment to ADs is discussed in a paper by Sheridan *et al*. (6).

This paper attempts to develop practical guidance for the AD assessment of high-dimensional models such as group contribution-based models. The group contribution methods have two underlying premises. First, the additivity assumption implies that a particular property of a compound is a sum of the contributions associated with atoms or fragments. Second, the transferability assumption states that the contributions of identical atoms or fragments are identical to those in the original compounds (the training set) used to estimate these contributions. A group contribution method is a very robust approach for developing QSAR models for broad chemical classes. Group contribution models commonly use hundreds of fragments as model descriptors.

To examine which domain assessment methods are suitable for high-dimensional models, we analyse a very popular model for n-octanol/water partition coefficient (K_{ow}) estimation, namely, KOWWIN, a member of the EPIWIN (Estimation Programs Interface for Windows) suite of models. EPIWIN software is available free from Syracuse Research Corporation (SRC) at http://www.epa.gov/oppt/exposure/docs/episuitedl.htm. The KOWWIN model, an implementation of the Atom/Fragment Contribution (AFC) method (7) is a multivariate linear regression model for $\log(K_{ow})$, derived from 508 fragment counts and correction factors (both used as descriptors) and experimental K_{ow} values for 2434 compounds.

Methods

Interpolation in the multivariate space

Calculating an interpolation region in a multivariate space is equivalent to estimating a convex hull (3). In this paper, we compare the following convex hull approximations:

1. Ranges in descriptor space;
2. Euclidean distance;
3. City-block distance;
4. Mahalanobis distance; and
5. Leverage/Hotelling T^2.

An overview of formulae and ingoing assumptions for each method is provided in Table 1. Readers seeking more detail about the methods can consult a recent review by Jaworska *et al*. (3). Convex hull calculation in high-dimensional space (number of dimensions > 3) is computationally very intensive. Thus, in our analyses, we do not include the probability density estimation approach (3), because a) parametric probability methods based on the normal distribution yield the same results as 3, 4, 5 approaches, and b) the training set lacks sufficient data to permit the use of a non-parametric probability density method (8). However, we explore and compare results obtained for the raw data with the results obtained for data scaled, centred, and rotated by Principal Components Analysis (PCA), as advocated by Eriksson *et al*. (9) and Seber (10).

Because the AFC model uses only fragment counts as descriptors (correction factors are also counts of specially selected complex fragments), AD assessment in the descriptor space can also be interpreted as AD assessment by structural similarity. In this very special case, the descriptor space and structural space are the same. The selection of relevant features is addressed by the fact that the fragments used in the model itself are relevant to the property being investigated (i.e. K_{ow}). Calculated distances measure distances in both descriptor and structural space.

Applicability domain criteria

Compounds are labelled out of the domain, if:

1. At least one descriptor is out of range for the ranges approach.

2. The distance between the chemical and the centre of the training data set exceeds the threshold for distance approaches. The threshold for all kinds of distances and Hotelling T^2 is the largest distance among the training set data points to the centre of the training data set.

Although these criteria may appear different, they have the same outcome — the smallest space encompassing the whole training set is estimated.

Table 1: Formulae and assumptions for different interpolation methods

Method	Formula	Assumptions on data distribution
Ranges	$d(x,y) = x - y$	Uniform
Euclidean distance	$d_E(x,\mu) = \sqrt{(x-\mu)^T (x-\mu)}$ ()^T- transposed matrix	Normal, equal variances, uncorrelated variables
City block	$d_c(x,y) = \sum_{i=1}^{n} \lvert x_i - y_i \rvert$	Uniform
Mahalanobis distance (leverage and Hotelling T^2 are proportional to d_M, see [3])	$d_M(x,y) = (x_i - y_i)\Sigma^{-1} (x_i - y_i)$ where Σ^{-1} is the inverse of the covariance matrix	Normal, arbitrary variances, arbitrary correlation

We compare the approaches listed above by evaluating their ability to distinguish between good (low error) and bad (high error) predictions by the Root Mean Squared Error (RMSE), which is the square root of Mean Squared Error (MSE), calculated as a sum of squared errors divided by the number of points. By using RMSE, we do not question the analysed QSAR model's goodness of fit, but rather use RMSE as a relative measure of the accuracy of predictions in the domain and out of the domain. We calculate the RMSE for the training set compounds, the RMSE for validation set compounds found to be in the domain, and the RMSE for validation set compounds found to be out of the domain. In addition to these three RMSE values, we report the number of compounds in and out of the domain.

Results of the SRC KOWWIN Case Study

KOWWIN datasets

The KOWWIN training and validation datasets were kindly provided by Syracuse Research Corporation. These sets consist of CAS numbers, and observed and predicted log(K_{ow}) values for the 2343 compounds in the training set and the 10,910 compounds in the validation set. Recently, the datasets became available online at http://esc.syrres.com/interkow/KowwinData.htm.

KOWWIN descriptors determination

The description of the AFC method (7) provides only a partial list of fragments and correction factors. Fragments are described in textual form, and the explicit structures are not given in many cases. Several listed fragments have ambiguous descriptions that make it difficult to directly reuse the AFC method. Furthermore, the fragments list and the correction factors list differed slightly in successive versions of SRC KOWWIN software as more compounds and fragments were added. We therefore decided to use the most reliable source for the KOWWIN descriptor space — the full text output of SRC KOWWIN v1.66 software (Figure 1). This text file lists all fragments and factors, and their frequencies and weights, applicable to each compound in the training set.

We developed a software tool to parse the text output and produce a table summarising the relationships between fragments/corrections factors and compounds. In this table, the columns are all possible fragments/correction factors and the rows are compounds. Each cell in the table denotes how many times a fragment occurs in a compound. The KOWWIN model's descriptor space was obtained by running the training set's 2434 compounds through the software. This revealed 186 different fragments and 322 different correction factors, resulting in a 508-dimensional descriptor space (Table 2). The log(K_{ow}) values vary between –4.57 and 8.19. The validation set's 10,910 compounds were also run through the software. The validation set makes use of 172 of 186 fragments and 316 of 322 correction factors (Table 2). The log(K_{ow}) values in the validation set vary between –4.99 and 11.71. The quality of the very high log(K_{ow}) values (above about 8) may need to be reviewed, but it is beyond the scope of this paper. Because the accuracy of the KOWWIN model was originally reported by its authors when using the entire validation set, we decided not to remove any compounds from the validation set.

Figure 1: KOWWIN output for a compound

```
SMILES : Oc(c(cc(c1)Cc(cc(c(O)c2C(C)(C)C)C(C)(C)C)c2)C(C)(C)C)c1C(C)(C)C
CHEM  : Phenol, 4,4'-methylenebis 2,6-bis(1,1-dimethylethyl)-
MOL FOR: C29 H44 O2
MOL WT : 424.67
--------+-----+-----------------------------------------+---------+--------
 TYPE  | NUM |      LOGKOW FRAGMENT DESCRIPTION        | COEFF   | VALUE
--------+-----+-----------------------------------------+---------+--------
 Frag  | 12  |  -CH3   [aliphatic carbon]              | 0.5473  |  6.5676
 Frag  |  1  |  -CH2-  [aliphatic carbon]              | 0.4911  |  0.4911
 Frag  | 12  |  Aromatic Carbon                        | 0.2940  |  3.5280
 Frag  |  2  |  -OH    [hydroxy, aromatic attach]      |-0.4802  | -0.9604
 Frag  |  4  |  -tert Carbon  [3 or more carbon attach]| 0.2676  |  1.0704
 Factor|  1  |  -CH2- (aliphatic), 2 phenyl attach correc|-0.2326 | -0.2326
 Factor|  2  |  Ring rx: -OH / di-ortho;sec- or t- carbon|-0.8500 | -1.7000
 Const |     |  Equation Constant                      |         |  0.2290
--------+-----+-----------------------------------------+---------+--------
                                  Log Kow  =  8.9931
```

Table 2: Fragment list for the KOWWIN training and validation sets[a]

	KOWWIN					
	Training set			Validation set		
Fragment	Frequency[b]	Min	Max	Frequency	Min	Max
Aromatic Carbon	1786 (73%)	2	24	8725 (80%)	1	30
CH_3[aliphatic carbon]	1388 (57%)	1	13	7353 (67%)	1	20
CH_2[aliphatic carbon]	1076 (44%)	1	18	7016 (64%)	1	28
CH[aliphatic carbon]	457 (18%)	1	16	3839 (35%)	1	23
C[aliphatic carbon-No H not tert]	229 (9%)	1	3	1343 (12%)	1	11
O[oxygen aliphatic attach]	108 (4%)	1	5	1231 (11%)	1	12
F[fluorine aliphatic attach]	103 (4%)	1	6	540 (5%)	1	23
Cl[chlorine aliphatic attach]	100 (4%)	1	6	354 (3%)	1	12
Si[silicon aromatic or oxygen attach]	15 (0.6%)	1	4	14 (0.1%)	1	9

[a]*Full list available from the authors*
[b]*Absolute (relative)*

The full list of fragments and correction factors used in the KOWWIN model, as well as the ranges for each fragment and correction factor, are not presented here, but are available from the authors of this paper.

The distributions of the descriptors were evaluated for a uniform distribution by the Kolmogorow–Smirnov test with MATLAB 6.5 R13, with a default rejection level of 5%: all failed. Next, the individual descriptors' distributions were evaluated for normality by the Jarque–Bera test with MATLAB 6.1 R13, with a default rejection level of 5%: all failed. This suggests that a more sophisticated technique, such as the non-parametric probability density estimation, is needed to determine the interpolation regions. However, we lack a sufficient amount of training data to permit the use of this method (3).

The KOWWIN model developers do not report performing scaling or PCA pre-processing steps. This may affect the model's quality and stability, as well as the feasibility for correctly estimating the AD. We scaled, centered and rotated the data by using PCA. This last step is especially important for the KOWWIN dataset, because the descriptors in the model are highly correlated. PCA corrects for these correlations, yielding orthogonal components (without linear correlation). The PCA results from the original data reveal that the first 16 principal components (PCs) explain 90% of the variance, and the first 36 PCs explain 95% of the variance. PCA on the scaled and centered data showed more balance: the first 197 PCs explain 90% of the variance; the first 282 PCs explain 95% of the variance.

Pre-treating the data during AD assessment, and not performing this step during model development, complicate the interpretation of the domain, because the model space and model domain space become different. Ideally, pre-treatment should be performed during model development, to allow the domain to be assessed in the model space.

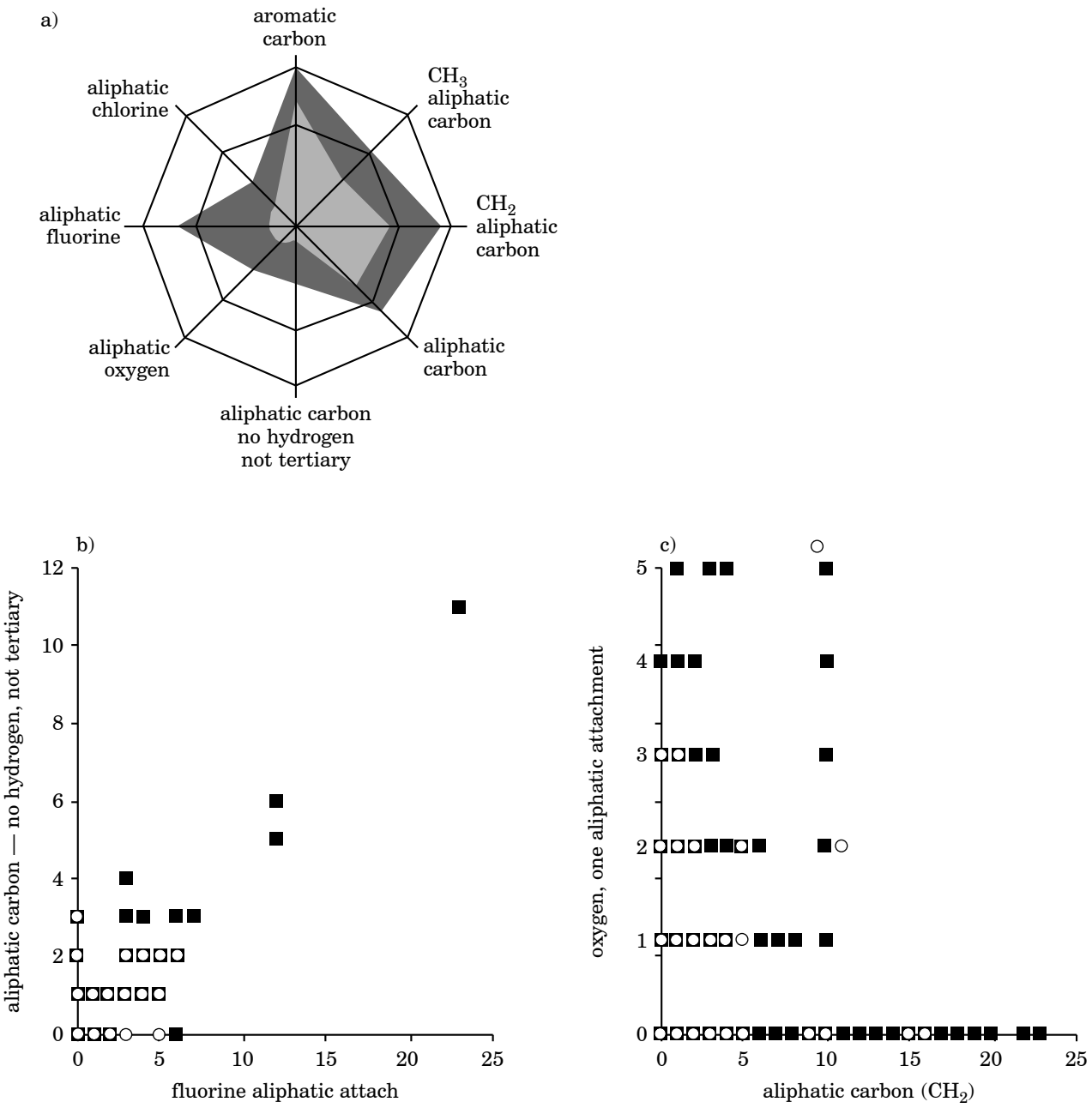
Comparison between KOWWIN training and validation set predictions

To assess the quality of AD assessment, we compared the observed results versus the predicted results for the validation set chemicals. The validation set is nearly ten times larger than the training set, but only partially overlaps the training set (Figure 2). The number of compounds found to be in the domain and out of the domain by different domain assessment methods are shown in Table 3. For selected compounds, the statistics for experimental and estimated $\log(K_{ow})$ values, and the absolute and relative prediction errors, are shown in Table 4. These compounds were selected to span the range of relative error for the set of compounds in and out of the domain, respectively. The entire list consists of more than 12,000 compounds, and is available from the authors.

Comparison of different methods to approximate training set interpolation regions

The numbers of validation compounds in the domain are similar for the various methods, except for the ranges approach after PC rotation of axes (Table 4). All the approaches result in lower RMSEs for validation compounds in the domain (0.43 to 0.6) than for compounds out of the domain (0.57 to 1.10). By comparison, the training set RMSE is 0.22. The ranges approach after PC rotation of axis

Figure 2: Projections of training set and validation set in descriptor space



a) web plot of 7 of the individual descriptors

= training set; = validation set.

b) fragment C and fragment F; c) fragment –O- and fragment CH_2

○ = training set; ■ = validation set

had a RMSE of 0.57, the lowest for the chemicals in the domain.

PCA rotation affects the outcome of the ranges approach, but has little effect on distance-based approaches (Table 4). The distance-based approaches are not affected in the case of the KOWWIN data set, because a) the scale factors are nearly identical along all dimensions, and b) the distance approaches assume normally (i.e. symmetrical) distributed data. Because the principal components extracted from the original data are not the same as the principal components

Table 3: Selected KOWWIN validation set compounds in- and out-of the training set ranges, and corresponding experimental and calculated log K_{OW} values[a]

			SRC KOWWIN			
CAS No.	Name	In or out of domain	Estimated value	Experimental error	Absolute error	Relative error %
3298	Perfluoropmethyl cyclohexylpiperidine	In	4.79	7.10	–2.31	33
678262	Pentane, dodecafluoro-	In	5.05	4.40	0.65	15
355680	Perfluorocyclohexane	In	3.33	2.91	0.42	14
47071114	4,6-NH_2 2,2-DiMe1(4-CF3)Ph s-triazene	In	1.28	1.22	0.06	4.92
80616597	Butanamide, N-(5-amino-1H-1,2,4-triazol-3-yl)-2,2,3,3,4,4,4-heptafluoro-	In	1.53	1.54	–0.01	0.65
77963509	B30C10-Benzocrownether	Out	–0.15	0.03	–0.18	600
104946625	B33C11- Benzocrownether	Out	–0.43	–0.09	–0.34	378
63144763	B27C9- Benzocrownether	Out	0.12	0.23	–0.11	48
88116590	Iohexol derivative	Out	1.94	–2.80	4.74	169
2915	3Azaglutaramide analog A37	Out	4.00	3.60	0.40	11
93414552	Benzoic acid, 3,4,5-trimethoxy-, 2-[4-[[(2-ethoxy-2-oxoethoxy)imino] methyl]-2-methoxyphenoxy]ethyl ester	Out	3.26	2.94	0.32	11
121284206	[8,8]DB48C16-Dibenzocrownether	Out	0.56	0.52	0.04	8

[a]*Full list available from the authors*
CAS numbers and names are taken from KOWWIN output.

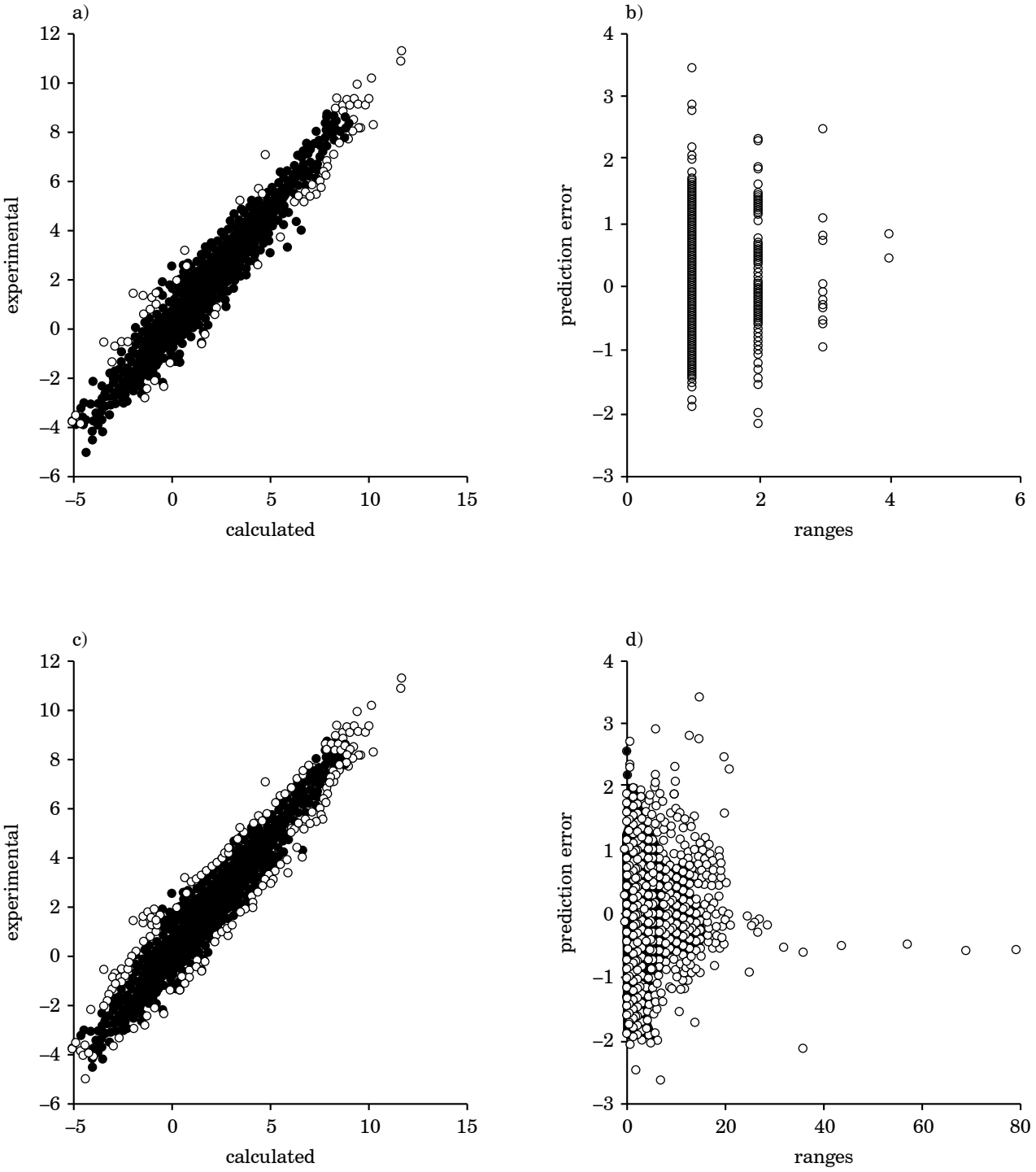
extracted from the scaled data (10), we carried out the scaling step before PCA rotation.

The elements in Figure 3 illustrate the correspondence between AD and prediction error for the ranges approach. The (a) and (c) plots show predicted versus observed value scatter plots. The (b) and (d) plots show prediction errors, expressed as the residual between the observed and predicted log(K_{ow}) value, versus the number of dimensions where the point is out of the training set range (zero means in-range). Figure 4 illustrates the correspondence between AD and prediction error for Euclidean distance approach. The distance between the point and the centre of the training set is plotted on the x-axis. Table 4 and Figures 3 and 4 show that, on average, validation compounds outside the AD have much larger prediction errors than those inside the AD. For example, where the AD is determined by the ranges approach, the relative prediction error for validation compounds in the domain varies from 0.65% to 33%, while the relative prediction error for chemicals out of the domain varies from 8% to 600%.

Discussion

We assessed AD estimation by calculation of interpolation regions based on the training set in the model parameter space. Several methods approximating the interpolation region were compared. Based on experience with KOWWIN, we conclude that the PC-rotated ranges approach is the simplest acceptable solution for AD estimation for high dimension models. There is no benefit from the application of any of the distance methods we examined. KOWWIN's training data set distribution suggests the application of a non-parametric probability density method, but there is not enough

Figure 3: The correspondence between prediction error and the application domain boundary obtained by ranges approach in the original descriptor space (a, b) and PC rotated space (c, d)



● = *chemicals in the domain;* ○ = *chemicals out of the domain;*

(a, c) experimental vs. calculated values; (b, d) correspondence between prediction error and the number of dimensions where the point is out of training set range (i.e. zero means in-range).

Table 4: Summary of the statistics for different applicability domain estimation methods applied to the KOWWIN validation set

	Validation data set		Validation (in)		Validation (out)	
No.	Domain defined by:	PC space	No. compounds	RMSE	No. compounds	RMSE
1	Ranges		10247	0.46	597	0.74
3	Euclidean distance		10796	0.47	48	0.94
5	City block distance		10797	0.47	47	0.96
7	Hotelling T^2		10685	0.59	160	0.73
11	Ranges (scaled data)	Yes	7460	0.43	3384	0.57
13	Euclidean distance — (Mahalanobis) distance (scaled)	Yes	10187	0.47	27	1.10
15	Hotelling T^2/ leverage (scaled)	Yes	10749	0.60	96	0.67
17	City block distance (scaled data)	Yes	10708	0.46	136	0.97

*Training set (*n = *2434) RMSE is 0.22.*

RMSE = root mean squared error.

data in the training set to permit the application of this method. As a result, the KOWWIN domain determined by ranges includes considerable empty space. Nevertheless, the ranges approach represents a refinement currently implemented in KOWWIN AD assessment, where only the presence of a given fragment is verified. In addition, we found that pre-treating data by PCA was necessary, because the fragment counts are highly correlated in KOWWIN, a data characteristic that can lead to unstable models and erroneous AD assessment.

Taking into account all possible combinations of fragment counts and correction factors for the 508 descriptors and their ranges (as found in the KOWWIN training set), the training space consists of 5.44E+41 unique points. For the purposes of this study, a point is unique if it corresponds to a unique combination of descriptor values (i.e. fragment and correction factors counts). Of this enormous space, the training set uses only 2113 unique points (some of the 2434 points coincide). Therefore, only 3.88E–37% of the training space is covered by the training set points. Although it is obvious that not all possible combinations of fragment and correction factors reflect stable molecules, we note that nothing prevents users from applying the KOWWIN model to compounds with fragment combinations never seen in the training set.

The successful practical experience with the KOWWIN model means that fragment additivity and transferability are working reasonably well within the training set space. The AFC method has problems with fragment additivity for rigid aromatic molecules and for compounds where the same fragment occurs many times in a molecule (such as in a long aliphatic chain). The AFC method also fails for molecules with "uncommon" functional groups, because transferability of these fragments is difficult to establish due to large uncertainties in their estimated contributions. Fragment ranges provide only a rough estimation of additivity boundaries.

More-precise AD assessment for high-dimension models requires the development of approaches for which high dimensionality is not a limiting factor. The best approach would be to define where model assumptions are valid. Let us examine the possibility to verify the additivity and transferability assumptions of the group contribution method. Additivity (11, 12) implies that a compound's structural components each make a separate and additive contribution to the property of interest. Transferability assumes that these contributions are the same across a wide variety of compounds (11).

Additivity is a widely accepted hypothesis, with evidence provided from both empirical studies (12) and contemporary quantum theories (13). While quantum mechanics predicts the properties of open systems to be additive, this "additivity" could be observed experimentally, only when the contribu-

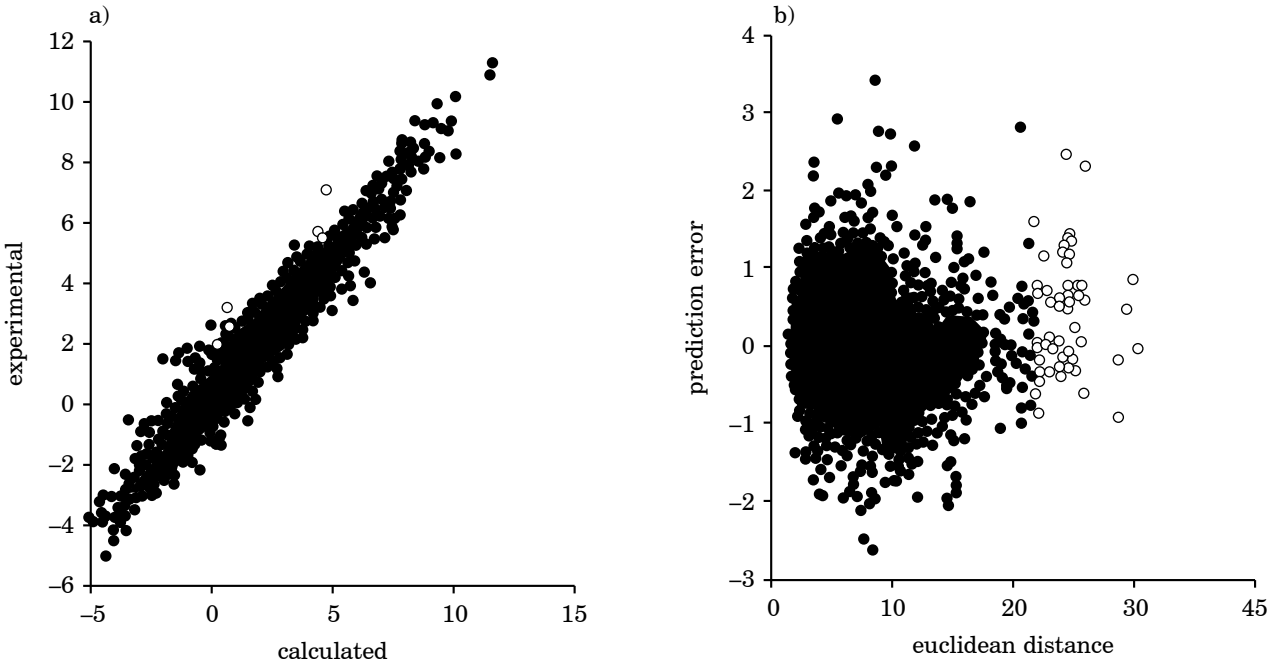
tion of the atom or fragment was also transferable without apparent change from one compound to another. Defining additivity and transferability boundaries has been thus far difficult to formalise, partly because, until recently, fragments had been determined empirically as was done for KOWWIN. Advances in understanding the additivity and, especially, the transferability of fragmental contributions may lead the way to redefining fragments based on theoretical considerations which are far easier to verify (12–14).

Future improvements in AD estimation will further reduce uncertainty in the prediction in the domain, but will never completely eliminate it. A particular representation of a chemical compound by its properties may not be unique (i.e. two different compounds may have the same representation by the subset of selected properties). Non-unique representation poses a potential risk of obtaining the correct result for one compound and an incorrect result for another. The lack of uniqueness could only be avoided if the set of descriptors used contained all the information about a chemical compound, but this is practically impossible. Thus, models that use few descriptors are especially prone to the non-unique representation of molecules, while models using many parameters (for example, AFC), are less prone to this, because the chances of missing a relevant parameter that explains chemical activity are smaller. Moreover, predictions from outside the AD could be correct, so such predictions could be marked as outside the domain, but should not be automatically rejected.

Conclusions

The key component to evaluating the quality of QSAR predictions is determining whether a compound belongs to the model's AD. The interpolation region based on the training data set provides a basis for estimating the model's AD. For high-dimension models, choosing the estimation method is not trivial, as the number of data in the training set and their distribution severely limits the choices. We recommend the simplest approach of ranges (combined with a PCA pre-processing step) as a practical compromise for group contribution

Figure 4: The correspondence between prediction error and the application domain boundary obtained by the Euclidean distance in the original descriptor space



Results in PC rotated space are very similar and not shown; ● = chemicals in the domain; ○ = chemicals out of the domain.

a) experimental vs. calculated values; b) correspondence between prediction error and the distance between a point and the mean for the test set.

models. At the same time, we recognise the need for research to develop methods for which dimensionality is not a limiting factor. This could be achieved by improving our theoretical understanding of two key assumptions in the group contribution method, additivity and transferability.

Acknowledgements

The training and validation sets of KOWWIN models were kindly provided by Syracuse Research Corporation (P. Howard). Nina Nikolova's work was funded by a Procter & Gamble postdoctoral fellowship. We also acknowledge partial funding by EU JCR ECVAM (project CCR.496575-Z).

Received 7.2.05; received in final form 7.6.05; accepted for publication 13.6.05.

References

1. OECD (2004). OECD Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure-Activity Relationship Models. Paris, France: OECD. Website http://www.oecd.org/document/23/0,2340,en_2649_201185_33957015_1_1_1_1,00.html (Accessed 12.9.05).
2. Jaworska, J., Comber, M., Van Leeuwen, C. & Auer, C. (2003). Summary of the workshop on regulatory acceptance of QSARs. *Environmental Health Perspectives* **111**, 1358–1360.
3. Jaworska, J., Nikolova-Jeliazkova, N. & Aldenberg, T. (2005). Review of methods for QSAR applicability domain estimation by the training set. *ATLA* **33**, 445–459.
4. Nikolova, N. & Jaworska, J. (2003). Approaches to measure chemical similarity — a review. *QSAR & Combinatorial Science* **22**, 1006–1026.
5. Bender, A. & Glen, R.C. (2004). Molecular similarity: a key technique in molecular informatics. *Journal of Organic and Biomolecular Chemistry* **2**, 3204–3218.
6. Sheridan, R., Feuston, R.P., Maiorov, V.N. & Kearsley, S. (2004). Similarity to molecules in the training set is a good discriminator for prediction accuracy in QSAR. *Journal of Chemical Information and Computer Science* **44**, 1912–1928.
7. Meylan, W.M. & Howard, P.H. (1995). Atom/fragment contribution method for estimating octanol–water partition coefficients. *Journal of Pharmacological Sciences* **84**, 83–92.
8. Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability. 170pp. London, UK: Chapman and Hall.
9. Eriksson, L., Jaworska, J., Worth, A., Cronin, M.T.D., McDowell, R.M. & Gramatica, P. (2003). Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environmental Health Perspectives* **111**, 1351–1375.
10. Seber, G.A.F. (1984). *Multivariate Observations*. 671pp. New York, NY, USA: Wiley & Sons.
11. McNaught, A.D. & Wilkinson, A., eds. (1997). *Compendium of Chemical Terminology*. 103pp. London, UK: Blackwell Science.
12. Benson, S.W., Cruickshank, F.R., Golden, D.M., Haugen, G.R., O'Neal, H.E., Rodgers, A.S., Shaw, R. & Walsh, R. (1969). Additivity rules for the estimation of thermo chemical properties. *Chemical Reviews* **69**, 279–324.
13. Bader, R. & Bayles, D. (2000). Properties of atoms in molecules: Group additivity. *Journal of Physical Chemistry A* **104**, 5579–5589.
14. Curutchet, C., Salichs, A., Barril, X., Orozco, M. & Luque, F.J. (2003). Transferability of fragmental contributions to the octanol/water partition coefficient: an NDDO based MST study. *Journal of Computational Chemistry* **24**, 32–4.