Journal of Computational Science 5 (2014) 345–350

Contents lists available at ScienceDirect

Journal of Computational Science

journal homepage: www.elsevier.com/locate/jocs

Sensitivity, specificity, and accuracy of predictive models on phenols toxicity

Sorana D. Bolboacă [a], Lorentz Jäntschi [b,*]

[a] "Iuliu Haţieganu" University of Medicine and Pharmacy Cluj-Napoca, Department of Medical Informatics and Biostatistics, 6 Louis Pasteur, 400349 Cluj-Napoca, Cluj, Romania
[b] Technical University of Cluj-Napoca, Department of Physics and Chemistry, 103-105 Muncii Bdv., 400641 Cluj-Napoca, Romania

ARTICLE INFO

Article history:
Received 18 January 2013
Received in revised form 29 May 2013
Accepted 4 October 2013
Available online 16 October 2013

Keywords:
Quantitative structure–activity relationship (QSAR)
Residuals (r_i)
Hat-matrix leverage (h_i)
Cook's distance (D_i)
Model diagnostic

ABSTRACT

The investigation of how the influential affect the metrics and predictivity of multiple linear regressions on a set of phenolic compounds with toxicity on *Tetrahymena pyriformis* is presented. The investigation of influential was conducted using standardized residuals (r_i-model) and Cook's distance (D_i-model) approaches. The applied approaches let to improvement of model's metrics, robustness and accuracy on the investigated sample. Overall, the r_i-model proved higher accuracy and robustness in terms of sensitivity while D_i-model proved robustness in terms of specificity. Characterization of the withdrawn compounds is essential for advance in developing models for the toxicity of phenols.

1. Introduction

In silico methods are used alternatives to *in vivo* and *in vitro* testing of activities/properties of known or new compounds to minimize animal testing ("3Rs": replacement, refinement and reduction of animals in research [1]). *In silico* methods required a sample of known compounds with valid information on property/activity of interest [2]. Furthermore, these methods are also used to predict toxicity of chemicals under the umbrella of predictive toxicology [3,4], even if more still needed to be done in deeper understanding of the biological mechanisms for toxicity [5]. The results of *in silico* methods take the form of a mathematical model (called quantitative structure–activity/property relationship – QSAR/QSPR [6,7]) able to explain (estimate/predict) the toxicity using information extracted from structure of chemical compounds (theoretical descriptors [8–13]) or using experimental measurements [14]. The phenols had been investigated by many researchers using *in silico* approaches since it is known that they affect the environment [15] with toxicity including on humans [16,17]. Different QSAR models able to estimate and/or to predict the toxicity of phenols had been reported in the literature [18–22]. Classification (qualitative outcome variable, see for example [23]) and/or regression (linear whenever normal distributed quantitative outcome variable exists and/or logistic whenever not normal quantitative or qualitative outcome variable exists; see for example [24]) statistical methods were most frequent used to build QSAR/QSPR models [25,26].

Linear regression approaches are frequently used as statistical method in identification of QSAR models when continuous input and output data are available. The reported model must be statistically significant and robust and must have predictive power (the 4th OECD Validation Principle [27]). The robustness of model could be assessed by the stability of its parameters (coefficients of the regression model as relative importance of the descriptors) in training and tests sets while the predictivity is given by the accuracy of prediction (a higher proportion of accurate prediction indicates a more reliable model).

Sensitivity analysis in linear models is conducted to detect influential observation(s). A compound is considered as influential if its removal changes significantly the model. The approaches most frequent used in sensitivity analysis are standardized residuals, hat-matrix leverage and Cook's distance. The most common thresholds for standardized residuals (residuals normalized by the standard error) are 2.5 [28] or 3 [29]. The hat matrix approach maps the vector of the observed values to the vector of the fitted values, investigating the input data of the regression model. Any input data with hat-matrix leverage value $(h_i) > 2 \times (k+1)/n$ [30] or $>3 \times (k+1)/2$ [31] (where h_i = leverage for ith compound, k = number of independent variables in the regression model

* Corresponding author. Tel.: +40 2 64 401775; fax: +40 2 64 401 768.
E-mail addresses: sbolboaca@umfcluj.ro (S.D. Bolboacă), lorentz.jantschi@gmail.com, lori@academicdirect.org (L. Jäntschi).

1877-7503/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jocs.2013.10.003

Table 1
Metrics for characterization of MLR models.

Criterion	Interpretation
MLR model	
R^2 = determination coefficient	The higher the better
R^2_{adj} = adjusted determination coefficient	The closeness to the R^2 the better
s = standard error of estimate	The lower the better
F-value (p-value)	The higher the better for F & lower the better for p-value
t-value (p-value)	The higher the better for t & lower the better for p-value
Leave-one-out model	
R^2_{loo} = determination coefficient	The closeness to the R^2 the better
s_{loo} = standard error of predicted	The lower the better
F_{loo} (p-value)	Similar to F-value (p-value)
t-Value (p-value)	Similar to t-value (p-value)

and n = sample size) is considered influential. Cook's distance (D_i = Cook's distance of ith compound), introduced in 1977 [32] and applied on the outcome of a model, combines residual and leverage in one indicator to identify influential in regression models. Different threshold found their practical applicability; a compound is considered influential if: $D_i > 0.85$ ($k = 2$, where k = number of regression parameters – simple linear models) [33], or if $D_i > 1$ [34] or if $D_i > 4/n$ (where n = sample size) [30]. Our study addresses the problem of sensitive analysis in QSAR models: we investigated the effect of influential compound(s) using standardized residuals, hat-matrix and Cook's distance on a sample of 250 phenolic compounds relating toxicity with structure.

2. Materials and methods

A sample of 250 phenolic compounds, with measured toxicity on *Tetrahymena pyriformis*, was investigated. The measured toxicity, expressed as $\log 1/IGC_{50}$ (the concentrations-mM producing a 50% growth inhibition on *T. pyriformis* expressed in logarithmical scale) were taken from Cronin et al. [35]. The calculated octanol/water partition coefficient ($\log P$) and LUMO (Energy of the Lowest Unoccupied Molecular Orbital) were taken from Zhao et al. [23]. Zhao et al. [23] classified the compounds in 17 groups based on the structure and functional groups of the compounds. The phenolic compounds, CAS number (a unique numerical identifier assigned by the Chemical Abstracts Service), the membership of the class according to Zhao et al. [23], measured toxicity and descriptors calculated for each compound ($\log P$ and LUMO) can be found in Table 1 of Supplementary Material. Zhao et al. [23] revealed that the determination coefficient for the whole sample of 250 compounds was slightly poor ($R^2 = 0.52$, R^2 = determination coefficient) but obtained good results for alkyl substituted phenols ($n = 35$, $R^2 = 0.90$, $s = 0.22$, $F = 287$, where n = sample size, s = standard error of the estimate, and F = Fisher's statistics).

The following steps were applied to identify influential compound(s) in our sample of 250 phenolic compounds:

- Test normality assumption of observed toxicity using Kolmogorov–Smirnov [36,37], Anderson–Darling [38] and Chi-Squared tests [39]. If data proved normal distributed construct and assess (using the metrics presented in Table 1) the MLR (multiple linear regression model).
- Identify the influential using the following thresholds:
  (a) Standardized residuals: $r_i > 3$
  (b) Hat-matrix leverage: $h_i > h_t$ (where $h_t = 2 \times (k+1)/n$)
  (c) Cook's distance: $D_i > 4/n$

Withdrawn the observation(s) that exceeded the threshold and construct the MLR again till no leverage exceed the threshold value or no improvement in the model is observed (used the metrics presented in Table 1).

- Construct and evaluate the final MLR models using the following validation metrics: MAE (mean absolute error), MAPE (mean absolute percentage error), SEP (standard error of prediction), REP% (relative error of prediction), RMSE (root-mean-square error), APV (average prediction variance [40]), TSE (total squared error [41]), APMSE (average prediction mean squared error [42]), %PredErr (percentage prediction error [43]). The model with the lowest values of the above statistical parameters was considering a better model.
- Diagnose the obtained models by calculation the accuracy ($AC = (TP + TN)/n \times 100$, where TP = number of true positive compounds, TN = number of true negative compounds, n = sample size, AC = accuracy), sensibility ($Se = (TP + FN)/n \times 100$, where FN = false negative) and specificity ($Sp = (TN + FP)/n \times 100$, where FP = false positive) accompanied by their associated 95% confidence intervals (95%CI) (more details about diagnosis statistics could be found in [44]). Sensitivity and specificity reflect the model robustness while the accuracy demonstrates the models predictivity.
- Compare the models in terms of correlation coefficient using Steiger's Z test [45] (use a significance level of 5%).
- Assess the predictivity of the models in training and test analysis. Use the simple randomization techniques [46] to split the full dataset in training and test sets (~2/3 compounds in training set).

3. Results

The measured/observed toxicity proved normal distributed (Kolmogorov–Smirnov statistics = 0.0228, p-value = 0.9992; Anderson–Darling statistics = 0.1470, Critical-value$_{\alpha=5\%}$ = 2.5018; Chi-Squared statistics = 1.8746, p-value = 0.9665).

The removal of compounds with h_i higher than threshold did not led to any improvement of the models characteristics and was not further investigated.

The characteristics of the models obtained after withdrawn of the identified influential using standardized residuals and Cook's distance are presented in Table 2. The Cook's distance approach reduces the sample size with 28% (95%CI [23–34%]) while the residual approach reduced the sample size with 12% (95%CI [8–17%]). Without any exception, all compounds identified as influential by standardized residuals approach were also identified by Cook's distance (4-Carboxylphenol; 3-Carboxylphenol; Salicylic acid; 4-Hydroxyphenylacetic acid; 2,3,5-Trichlorophenol; 4-Nitrophenol; 3-Methyl-4-nitrophenol; 2,6-Dichloro-4-nitrophenol; 2,4,6-Trinitrophenol; Catechol; Hydroquinone; Methylhydroquinone; 2,3-Dimethylhydroquinone; Trimethylhydroquinone; Tetramethylhydroquinone; Methoxyhydroquinone; Phenylhydroquinone; Chlorohydroquinone; Bromohydroquinone; Tetrafluorohydroquinone; Tetrabromocatechol; 1,2,4-Trihydroxybenzene; 1,2,3-Trihydroxybenzene; 2-Aminophenol; 4-Aminophenol; 4-Amino-2-cresol; 6-Amino-2,4-dimethylphenol; 4-Amino-2,3-dimethylphenol; 2-Amino-4-chlorophenol; 5-Amino-2-methoxyphenol; and 2,4-Diaminophenol). The proportion of chemical compounds withdrawn by Cook's distance approached proved significantly higher compared with proportion of compounds withdrawn by standardized residual approach (Z-statistics = 7.785, $p < 0.001$).

The following significant differences have been identified when the dependent and independent variables were compared for compound withdrawn by standardized residual approach (r_i-model): the mean of $\log P$ of withdrawn compound (1.41) proved

Table 2
Models characteristics: full model *vs.* residuals model (r_i) *vs.* Cook's distance model (D_i).

Model (n)	R^2	R^2_{adj}	s	F-value (p-value)	R^2_{loo}	s_{loo}	F_{loo}-value (p_{loo}-value)
Full ($n=250$)	0.5643	0.5608	0.5491	160 (2.77×10^{-45})	0.5513	0.5573	152 (8.48×10^{-44})
r_i ($n=219$)	0.8530	0.8517	0.3121	626 (1.15×10^{-90})	0.8482	0.3172	604 (1.33×10^{-89})
D_i ($n=179$)	0.9099	0.9089	0.2352	889 (9.94×10^{-93})	0.9068	0.2392	857 (4.56×10^{-92})

n = sample size; R^2 = determination coefficient; R^2_{adj} = adjusted determination coefficient; s = standard error of the estimate; F-value = Fisher's statistics; loo = leave-one-out analysis; s_{loo} = standard error of the predicted.

significantly lower compared with the mean of log P of remaining compounds (2.45) (t-statistics = 4.76, $p = 3.24 \times 10^{-6}$).

The withdrawn compounds were considered as external validation set and the following correlation coefficients were obtained: 0.1805 ($p = 0.1656$) for r_i-model and 0.5445 ($p = 4.60 \times 10^{-7}$) for D_i-model.

Graphical representations of the observed by estimated toxicity for the full model, the residual model (r_i-model) and the Cook's distance model (D_i-model) are presented in Fig. 1.

The validation of the QSAR models was conducted using nine statistics as presented in Section 2 and the results are presented in Table 3.

The r_i-model proved significant higher correlation coefficient compared with the full model ($Z = 2.564$, $p = 3.93 \times 10^{-12}$). D_i-model proved significant higher correlation coefficient compared with both full-model ($Z = 9.1$, $p < 0.0001$) and r_i-model ($Z = 2.6$, $p = 0.0052$).

The results obtained on diagnosis of the models, expressed as statistical parameter and associated 95% confidence interval, are presented in Table 4.

The data set of the full model was randomly split in training and test sets (34% of compounds in test set). 74 (34%, 95%CI [27–41%]) and respectively 61 (34%, 95%CI [27–41%]) compounds were in test sets for r_i-model and D_i-model. The correlation coefficients in training and test sets were as follows: 0.9159 ($p = 3.28 \times 10^{-63}$) in training and 0.9388 (8.31×10^{-37}) in test set for r_i-model; 0.9487 ($p = 2.99 \times 10^{-63}$) in training and 0.9646 (1.06×10^{-36}) in test set for D_i-model.

The performances of r_i-model and D_i-model in training and test sets are graphically depicted in Fig. 2.

The following results were obtained when the correlation coefficients obtained in training and test analyses by r_i-model and D_i-model were compared: training sets – Z-statistics = 2.038 ($p = 0.0208$) & test sets – Z-statistics = 1.584 ($p = 0.0566$).

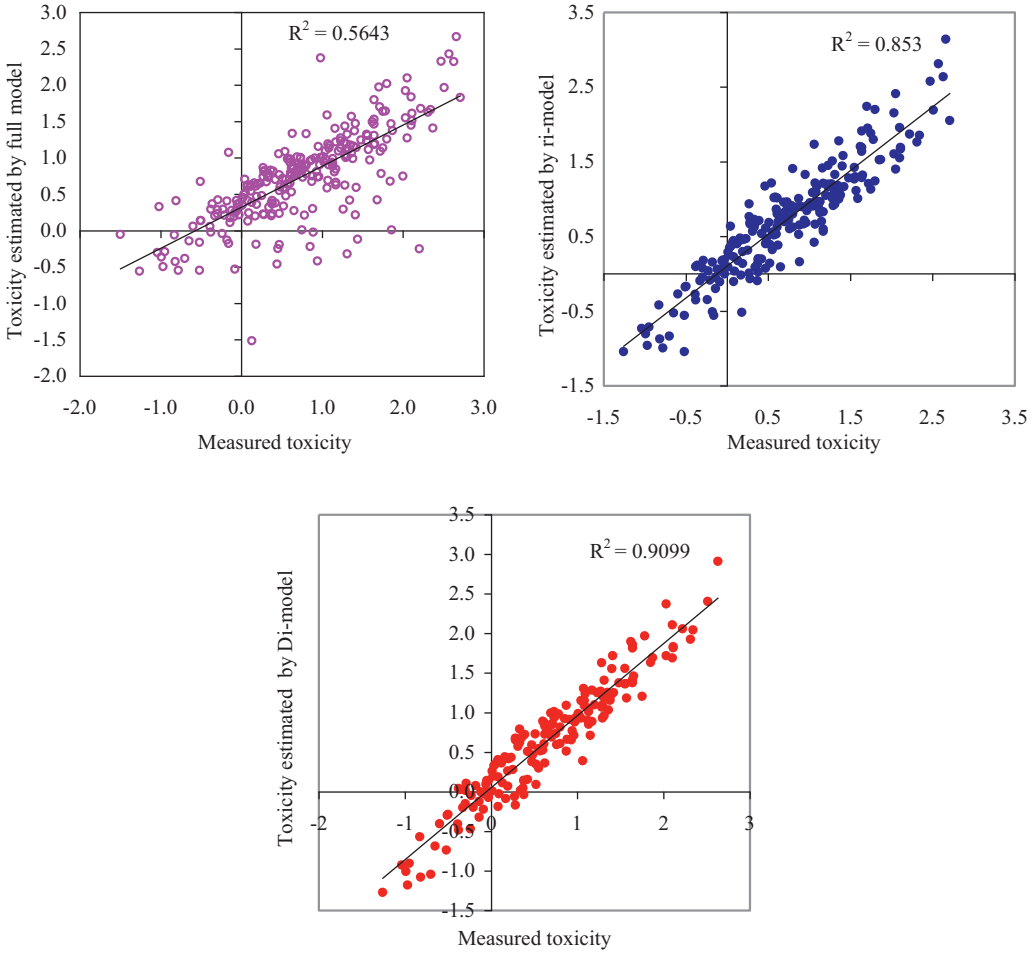

Fig. 1. The effects of influential on QSAR models: full model, r_i-model and Cook's distance model (D_i-model).

Table 3
Model performance metrics.

Validation parameter	Full (n = 250)	r_i > 3 (n = 219)	D_i > 4/n (n = 179)
MAE (mean absolute error)	0.4051	0.2569	0.1978
MAPE (mean absolute percentage error)	1.5639	0.9754	0.8280
SEP (standard error of prediction)	0.5468	0.3107	0.2338
REP(%) (relative error of prediction)	73.8294	42.8440	35.9884
RMSE (root-mean-square error)	0.3002	0.0970	0.0550
APV (average prediction variance)	0.3026	0.0978	0.0556
TSE (total squared error)	3	3	3
APMSE (average prediction mean squared error)	0.0012	0.0004	0.0003
%PredErr (percentage prediction error)	37.3708	20.7592	13.4657

r_i = standardized residuals model; D_i = Cook's distance model; n = sample size.

Table 4
Model diagnostic metrics.

Model	Sensitivity	Specificity	Accuracy
		Parameter [95% confidence interval]	
Full model (n = 250)	0.9463 [0.9065–0.9698]	0.4000 [0.2702–0.5455]	0.8480 [0.807–0.884]
r_i-model (n = 219)	0.9722 [0.9366–0.9881]	0.6923 [0.5358–0.8143]	0.9224 [0.879–0.949]
D_i-model (n = 179)	0.9583 [0.9121–0.9808]	0.7429 [0.5793–0.8584]	0.9162 [0.864–0.915]
		Z-statistics (p-value)	
Full *vs* r_i-model	−1.433 (0.1519)	−6.6489 (<0.0001)	−2.5633 (0.0104)
Full *vs* D_i-model	−0.5811 (0.5612)	−7.6162 (<0.0001)	−2.2191 (0.0265)
r_i-model *vs* D_i-model	0.7466 (0.4553)	−1.1204 (0.2626)	0.2255 (0.8216)

r_i-model = QSAR model obtained applying the standardized residual approach; D_i-model = QSAR model obtained applying the Cook's distance approach.

4. Discussion

This paper describes the effect of influential compounds identified using standardized residuals, hat-matrix leverage and Cook's distance to the QSAR models (as linear regression models) on some phenolic compounds with toxicity on *T. pyriformis*.

The hat-matrix leverage approach proved no effect on the model's metrics and was not further investigated. The two other investigated approaches led to removal of 12% (residual approach) and respectively 28% compounds (Cook's distance approach); with a significantly higher proportion of withdrawn compounds when the Cook's distance approach was applied ($p < 0.001$). The removal of high proportion of compounds using Cook's distance approach was also observed on other data sets (such as organohalogen compounds – 26% and aliphatic organic compounds – 27% [47], with variations between 3% and 21% on data sets reported on [48]).

The Cook's distance approach identified as influential all compounds identified by standardized residuals approach leading to a model (D_i-model) which excluded the compounds identified as influential by both applied approaches. The analysis of models metrics led to the followings:

- Withdrawing of influential led to a significantly improvement of models goodness-of-fit ($p < 0.0001$; Table 2).
- All models proved internal valid, sustained by the metrics obtained in leave-one-out analysis (Table 2).
- The highest determination coefficient and F-statistics as well as the lowest standard error of the estimate and predicted were obtained by D_i-model (Table 2). Furthermore, this model also obtained the lowest distance between R^2 and R^2_{loo}.
- D_i-model better fit to a straight line (Fig. 1) compared to r_i-model.
- Validation metrics identified that D_i-model is the best model (8 criteria out of 9 – TSE proved identical value to all 3 models – compared to both full-model and r_i-model; Table 3). The lowest values of mean absolute error and mean absolute percentage error, as well as lowest relative error of prediction, average prediction mean squared error and percentage prediction error sustain this conclusion. Furthermore, r_i-model proved also better

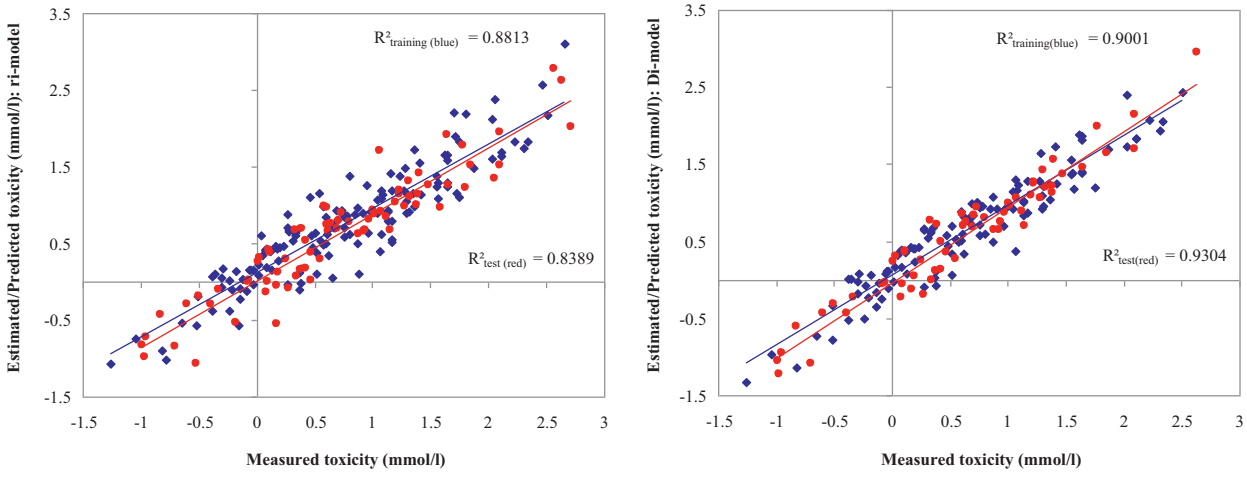

Fig. 2. Models performances in training and test sets: r_i-model (left-hand plot) and D_i-model (right-hand plot).

than full-model (with the same observation as above regarding TSE; Table 3).
- The predictivity of the r_i-model assessed through accuracy seems slightly better compared with predictivity of D_i-model but not statistically significant (Table 4). However, the predictivity of D_i-model according to the value of the accuracy metric is higher than the predictivity of the full-model.
- The r_i-model proved more robust in terms of lowest false negative classification of compounds while the D_i-model proved more robust in terms of lowest false positive classification of compounds (Table 4).

The compounds identified as influential by standardized residuals proved significantly different by the remained compounds. But in which aspects these compounds are different? If the withdrawn compounds were seen as external validation sets, low goodness-of-fit were identified (0.1805 for r_i-model and 0.5445 for D_i-model). However, correlation coefficient obtained on the compounds withdrawn by Cook's distance proved statistically significant ($p < 0.0001$) but did not accomplished the criterion of being higher than 0.6 [49]. The analysis of the withdrawn compounds identified that one of independent variable had statistically different values compared with the remaining compounds ($\log P$) in r_i-model. Differences were not identified neither for independent variables nor for dependent variable when the D_i-model was analyzed.

The prediction performances of r_i- and D_i-models were furthermore investigated in leave-many-out analysis. The random split of compounds was performed using the whole sample of 250 compounds resulting ~34% compounds in both test sets. The metrics of the models in training and test analyses proved predictive power of both investigated models (r_i-model and D_i-model; Fig. 2) with significant higher correlation coefficient in D_i-training set compared with r_i-training set ($p = 0.0208$) and no significant difference in test sets ($p = 0.0566$).

The obtained models appear to have good correlation performances (with R^2 values higher than 0.8) [49] in leave-one-out and leave-many-out analyses which is useful in explanatory models. The D_i-model proved good retrospective fit (fit to the original data after removal of influential compounds) and fair prospective fit (fit to new data represented by compounds withdrawn as influential). In terms of diagnostic metrics, could be observed that the refined models showed no significant improvements of model sensitivity but had significant improvements of both specificity and accuracy (see Table 4).

The withdrawn of influential compounds can led to loss of pertinent information. Therefore, it is not an easy task to judge if a data point is truly influential. Furthermore, it is essential to understand how the withdrawn compounds are different from compounds keep in the sample in order to translate the mathematical model in practical knowledge. For the class of compounds investigated in this study, Schultz and co-authors explained the "mechanism" of outliers (compounds that do not fit QSAR models) as a result of electo(nucleo)philic interactions [50]. Previous studies on phenols toxicity for *T. pyriformis* identified that frequently the compounds with pre-electrophile mechanism of action are statistically identified as outliers [51,52]. As consequence, Enoch and co-authors suggested that a better approach is to apply the modeling approach for strictly defined mechanism of action to avoid the poor prediction of certain compounds into a global model [19]. Aptula and co-authors showed that polyhydroxybenzene derivatives with hydroxy groups oriented meta had toxicity correlated with hydrophobicity while the compounds with hydroxy groups oriented ortho or para had the toxicity related to the electrophilic chemistry of their oxidation products [53] The characterization of the compounds identified as influential was beyond the aim of our study, even if it is an essential step when a model with practical applicability is of interest. We focused in our study only on the problem of influential from statistical point of view which is could not bring a real advance in developing models for toxicity of phenols where a practical interpretation of the model is essential. From a statistical or modeling print of view, in our analysis, when the compounds were removed by standardized residuals or Cook's distance approach, the R^2 increased from 0.5643 to 0.8530 and respectively 0.9099 (see Table 2). The increasing of R^2 may suggest that not important information was discarded. Furthermore, the advantage of standardized residual approach and of Cook's distance approach could be a matter of thresholds selection. The analysis of how the thresholds adjustments influence the metrics of the models and how the influential compound are different by the compounds kept in the sample according to applicability domain is now conducted in our laboratory. Moreover, we prepare an analysis to identify if the removed compounds are the same when the thresholds are adjusted and if an algorithm of removing influential compounds using more than one criterion led to better results. Another way of approaching the problem is to adjust the thresholds that both methods to remove the same number of compounds and to analyze how it is reflect in the models metrics and performances.

5. Conclusion

The use of standardized residuals and Cook's distance approaches let to improvement of QSAR's metrics, robustness and accuracy. The standardized residuals approach identified those compounds with lower values of measured $\log P$.

Overall, the r_i-model proved higher accuracy and robustness in terms of sensitivity compared to D_i-model while D_i-model proved robustness in terms of specificity.

Identification of influential compounds is recommended whenever a linear regression model is preferred. Further researches are needed to analyze the nature of the differences between compounds identified as influential and remaining compounds.

Conflict of interest

The authors declare that they have no conflicts of interest in the research.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jocs.2013.10.003.

References

[1] W.M.S. Russell, R.L. Burch, The Principles of Humane Experimental Technique, Methuen, USA, 1959.
[2] A. Sacan, S. Ekins, S. Kortagere, Applications and limitations of in silico models in drug discovery, Methods Mol. Biol. 910 (2012) 87–124.
[3] C. Helma, Predictive Toxicology, Taylor & Francis Group, New York, USA, 2005.
[4] S. Kar, K. Roy, Predictive toxicology using QSAR: a perspective, J. Indian Chem. Soc. 87 (12) (2010) 1455–1515.
[5] M.P. Gleeson, S. Modi, A. Bender, R.L. Marchese Robinson, J. Kirchmair, M. Promkatkaew, S. Hannongbua, R.C. Glen, The challenges involved in modeling toxicity data in silico: a review, Curr. Pharm. Des. 18 (9) (2012) 1266–1291.
[6] L.P. Hammett, Some relations between reaction rates and equilibrium constants, Chem. Rev. 17 (1935) 125–136.
[7] C. Hansch, A.R. Steward, The use of substituent constants in the analysis of the structure–activity relationship in penicillin derivatives, J. Med. Chem. 7 (1964) 691–694.
[8] R. Gozalbes, J.P. Doucet, F. Derouin, Application of topological descriptors in QSAR and drug design: history and new trends, Curr. Drug Targets Infect. Disord. 2 (1) (2002) 93–102.
[9] A.Z. Dudek, T. Arodz, J. Gálvez, Computational methods in developing quantitative structure–activity relationships (QSAR): a review, Comb. Chem. High Throughput Screen. 9 (3) (2006) 213–228.

[10] S.D. Bolboacă, L. Jäntschi, Comparison of QSAR performances on carboquinone derivatives, Sci. World J. 9 (2009) 1148–1166.
[11] R. Todeschini, V. Consonni, Molecular Descriptors for Chemoinformatics, Wiley-VCH Verlag, Weinheim, 2010.
[12] R.E. Sestraş, L. Jäntschi, S.D. Bolboacă, Quantum mechanics study on a series of steroids relating separation with structure, J. Planar Chromatogr.-Mod. TLC 25 (6) (2012) 528–533.
[13] S.D. Bolboacă, L. Jäntschi, From molecular structure to molecular design through the Molecular Descriptors Family Methodology, in: E.A. Castro (Ed.), QSPR-QSAR Studies on Desired Properties for Drug Design, Research Signpost, Transworld Research Network, 2010, pp. 117–166.
[14] I. Lessigiarska, A.P. Worth, T.I. Netzeva, J.C. Dearden, M.T. Cronin, Quantitative structure–activity–activity and quantitative structure–activity investigations of human and rodent toxicity, Chemosphere 65 (10) (2006) 1878–1887.
[15] S. Wasi, S. Tabrez, M. Ahmad, Toxicological effects of major environmental pollutants: an overview, Environ. Monit. Assess. (2012) 1–9.
[16] D.J. Brancato, Recognizing potential toxicity of phenol, Vet. Hum. Toxicol. 24 (1) (1982) 29–30.
[17] J. Michałowicz, W. Duda, Phenols – sources and toxicity, Pol. J. Environ. Stud. 16 (3) (2007) 347–362.
[18] D.X. Jiang, Y. Li, J. Li, G.X. Wang, Prediction of the aquatic toxicity of phenols to Tetrahymena pyriformis from molecular descriptors, Int. J. Environ. Res. 5 (4) (2011) 923–938.
[19] S.J. Enoch, M.T. Cronin, T.W. Schultz, J.C. Madden, An evaluation of global QSAR models for the prediction of the toxicity of phenols to Tetrahymena pyriformis, Chemosphere 71 (7) (2008) 1225–1232.
[20] P.R. Duchowicz, A.G. Mercader, F.M. Fernandez, E.A. Castro, Prediction of aqueous toxicity for heterogeneous phenol derivatives by QSAR, Chemom. Intell. Lab. Syst. 90 (2008) 97–107.
[21] I.V. Tetko, I. Sushko, A.K. Pandey, H. Zhu, A. Tropsha, E. Papa, T. Oberg, R. Todeschini, D. Fourches, A. Varnek, Critical assessment of QSAR models of environmental toxicity against Tetrahymena pyriformis: focusing on applicability domain and overfitting by variable selection, J. Chem. Inf. Model. 48 (9) (2008) 1733–1746.
[22] M. Doğa Ertürk, M. Türker Saçan, M. Novic, N. Minovski, Quantitative structure–activity relationships (QSARs) using the novel marine algal toxicity data of phenols, J. Mol. Graph. Model. 38 (2012) 90–100.
[23] Y.H. Zhao, X. Yuan, L.M. Su, W.C. Qin, M.H. Abraham, Classification of toxicity of phenols to Tetrahymena pyriformis and subsequent derivation of QSARs from hydrophobic, ionization and electronic parameters, Chemosphere 75 (7) (2009) 866–871.
[24] Y.H. Zhao, X.J. Zhang, Y. Wen, F.T. Sun, Z. Guo, W.C. Qin, H.W. Qin, J.L. Xu, L.X. Sheng, M.H. Abraham, Toxicity of organic chemicals to Tetrahymena pyriformis: effect of polarity and ionization on toxicity, Chemosphere 79 (1) (2010) 72–77.
[25] P. Gramatica, On the development and validation of QSAR models, Methods Mol. Biol. 930 (2013) 499–526.
[26] P. Liu, W. Long, Current mathematical methods used in QSAR/QSPR studies, Int. J. Mol. Sci. 10 (5) (2009) 1978–1998.
[27] OECD, Guidance Document on the Validation of (Quantitative) Structure–Activity Relationships [(Q)SAR] Models. Available from: http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=env/jm/mono(2007)2&doclanguage=en
[28] A. Tropsha, P. Gramatica, V.K. Gombar, The Importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models, QSAR Comb. Sci. 22 (1) (2003) 69–77.
[29] P. Gramatica, Principles of QSAR models validation: internal and external, QSAR Comb. Sci. 26 (5) (2007) 694–701.
[30] K.A. Bollen, R. Jackman, Regression diagnostics: an expository treatment of outliers and influential cases, in: J. Fox, J. Scott Long (Eds.), Modern Methods of Data Analysis, Sage, Newbury Park, 1990, pp. 257–291.
[31] L. Eriksson, J. Jaworska, A.P. Worth, M.T.D. Cronin, R.M. McDowell, P. Gramatica, Methods for reliability and uncertainty assessment and for applicability evaluations of classification and regression-based QSARs, Environ. Health Perspect. 111 (2003) 1361–1375.
[32] R.D. Cook, Detection of influential observations in linear regression, Technometrics 19 (1977) 15–18.
[33] B. Mc Donald, A teaching note on Cook's distance – a guideline, Res. Lett. Inf. Math. Sci. 3 (2002) 127–128.
[34] R.D. Cook, S. Weisber, Residuals and Influence in Regression, Chapman & Hall, New York, 1982.
[35] M.T.D. Cronin, A.O. Aptula, J.C. Duffy, T.I. Netzeva, P.H. Rowe, I.V. Valkova, T.W. Schultz, Comparative assessment of methods to develop QSARs for the prediction of the toxicity of phenols to Tetrahymena pyriformis, Chemosphere 49 (2002) 1201–1221.
[36] A. Kolmogorov, Confidence limits for an unknown distribution function, Ann. Math. Stat. 12 (4) (1941) 461–463.
[37] N.V. Smirnov, Tables for estimating the goodness of fit of empirical distributions, Ann. Math. Stat. 19 (1948) 279.
[38] T.W. Anderson, D.A. Darling, Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes, Ann. Math. Stat. 23 (2) (1952) 193–212.
[39] K. Pearson, On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, Philos. Mag. 50 (1900) 157–175.
[40] C.L. Mallow, Some comments on Cp, Technometrics 15 (1973) 661–675.
[41] J.W. Gorman, R.J. Toman, Selection of variables for fitting equations to data, Technometrics 8 (1966) 27–51.
[42] J.W. Tukey, Discussion, J. R. Stat. Soc. 29 (1967) 47–48.
[43] A.-L. Milac, S. Avram, A.-J. Petrescu, Evaluation of a neural networks QSAR method based on ligand representation using substituent descriptors: application to HIV-1 protease inhibitors, J. Mol. Graph. Model. 25 (1) (2006) 37–45.
[44] S.D. Bolboacă, L. Jäntschi, Predictivity approach for quantitative structure-property models. Application for blood–brain barrier permeation of diverse drug-like compounds, Int. J. Mol. Sci. 12 (7) (2011) 4348–4364.
[45] J.H. Steiger, Tests for comparing elements of a correlation matrix, Psychol. Bull. 87 (1980) 245–251.
[46] S.D. Bolboacă, Assessment of random assignment in training and test sets using generalized cluster analysis technique, Appl. Med. Inform. 28 (2) (2010) 9–14.
[47] S.D. Bolboacă, Is the Cook's distance useful in refining the quantitative structure-property linear models? The case of partition coefficient, in: L. Stoicu-Tivadar, G.I. Mihalaş, O.S. Lupşe, M.M. Vida, C. Vernic (Eds.), Innovation and Cooperation – Proceeding Conference of the Romanian Society of Medical Informatics, ROMEDINF2012, Editura Victor Babeş, 2012, pp. 87–97.
[48] S.D. Bolboacă, L. Jäntschi, The effect of leverage and influential on quantitative structure–activity relationships, Comb. Chem. High Throughput Screen. 16 (4) (2013) 288–297.
[49] A. Golbraikh, A. Tropsha, Beware of q2!, J. Mol. Graph. Model. 20 (4) (2002) 269–276.
[50] T.W. Schultz, M.T.D. Cronin, J.D. Walker, A.O. Aptula, Quantitative structure–activity relationships (QSARs) in toxicology: a historical perspective, J. Mol. Struct. (THEOCHEM) 622 (1) (2003) 1–22.
[51] M.T.D. Cronin, T.W. Schultz, Structure–toxicity relationships for phenols to *Tetrahymena pyriformis*, Chemosphere 32 (1996) 1453–1468.
[52] J. Devillers, Linear versus nonlinear QSAR modeling of the toxicity of phenol derivatives to *Tetrahymena pyriformis*, SAR QSAR Environ. Res. 15 (2004) 237–249.
[53] A.O. Aptula, D.W. Roberts, M.T.D. Cronin, T.W. Schultz, Chemistry–toxicity relationships for the effects of di- and trihydroxybenzenes to *Tetrahymena pyriformis*, Chem. Res. Toxicol. 18 (2005) 844–854.

Sorana D. Bolboacă received a PhD in Medical Sciences from "Iuliu Haţieganu" University of Medicine and Pharmacy Cluj-Napoca (2006). Since October 2001 she has been working for the "Iuliu Haţieganu" University of Medicine and Pharmacy Cluj-Napoca, and is now assistant professor at this university (2006-present). Her research topics are QSPR/QSAR approaches and applied statistics for medical and/or life-science research. She published over 100 papers and is a author of a book chapter on QSAR Studies on Desired Properties for Drug Design (Research Signpost, 2010).

Lorentz Jäntschi received a PhD in Organic and Computational Chemistry from the Babeş-Bolyai University Cluj-Napoca (2000) and a PhD in Genetics and Plant Breeding from the University of Agricultural Science and Veterinary Medicine Cluj-Napoca (2010). He is now associate professor at the Technical University of Cluj-Napoca. His main research interests concern the modeling and simulations of different systems, algorithms and client-server applications. He published over 150 papers and is a co-author of a book on Molecular Topology (Nova Science, 2001).