

	<i>QMRF identifier (JRC Inventory):</i> Q17-22b-0055
	<i>QMRF Title:</i> QSARINS model for hydroxyl-mediated tropospheric degradation using DRAGON descriptors
	<i>Printing Date:</i> Dec 11, 2019

## 1. QSAR identifier

### 1.1. QSAR identifier (title):

QSARINS model for hydroxyl-mediated tropospheric degradation using DRAGON descriptors

### 1.2. Other related models:

#### 1.3. Software coding the model:

DRAGON

Software for the calculation of molecular descriptors, ver.5.5 for Windows, 2007 R. Todeschini, V. Consonni, A. Mauri, M. Pavan  
 info@talete.mi.it  
<http://www.talete.mi.it/>

MOBYDIGS

software for multilinear regression analysis and variable subset selection by Genetic Algorithm, ver. 1.0 beta for windows, 2004  
 Todeschini Roberto, Talete srl, Milan (Italy)  
<http://www.talete.mi.it/>

QSARINS

Software for the development, analysis and validation of QSAR MLR models, version 2.2, 2015  
 Paola Gramatica, email: [paola.gramatica@uninsubria.it](mailto:paola.gramatica@uninsubria.it)  
<http://www.qsar.it/>

## 2. General information

### 2.1. Date of QMRF:

21/06/2011

### 2.2. QMRF author(s) and contact details:

[1] Paola Gramatica University of Insubria, Varese +390332421573 [paola.gramatica@uninsubria.it](mailto:paola.gramatica@uninsubria.it)  
<http://www.qsar.it/>

[2] Stefano Cassani University of Insubria, Varese +390332421439 [stefano.cassani@uninsubria.it](mailto:stefano.cassani@uninsubria.it)  
<http://www.qsar.it/>

### 2.3. Date of QMRF update(s):

26/01/2015

### 2.4. QMRF update(s):

### 2.5. Model developer(s) and contact details:

Paola Gramatica University of Insubria, Varese +390332421573 [paola.gramatica@uninsubria.it](mailto:paola.gramatica@uninsubria.it)  
<http://www.qsar.it/>

### 2.6. Date of model development and/or publication:

**2.7.Reference(s) to main scientific papers and/or software package:**

[1]Roy PP, Kovarich S & Gramatica P (2011). QSAR Model Reproducibility and Applicability: A Case Study of Rate Constants of Hydroxyl Radical Reaction Models Applied to Polybrominated Diphenyl Ethers and (Benzo-)Triazoles. Journal of Computational Chemistry. 32, 2386–2396 Doi: 10.1002/jcc.21820

[2]DRAGON Software for the calculation of molecular descriptors, ver.5.5 for Windows, 2007 R. Todeschini, V.Consonni, A. Mauri, M.Pavan <http://www.talete.mi.it/>

[3]MOBYDIGS Software for multilinear regression analysis and variable subset selection by Genetic Algorithm, ver. 1.0 beta for windows, 2004 Todeschini Roberto, Talete srl, Milano (Italy). <http://www.talete.mi.it/>

[4]QSARINS 2.2, 2015. Software for the development, analysis and validation of QSAR MLR models <http://www.qsar.it/>

**2.8.Availability of information about the model:**

Non-proprietary. Defined and available algorithm. Training and prediction sets are available in the Supporting Information of the related paper [ref 2; sect 9.2], in the attached sdf files in this QMRF (see Section 9.3) and in the QSARINS database [ref 4,5; sect 9.2].

**2.9.Availability of another QMRF for exactly the same model:**

Non to date.

**3.Defining the endpoint - OECD Principle 1****3.1.Species:**

Not applicable

**3.2.Endpoint:**

2.Environmental fate parameters 2.2.b.Persistence: Abiotic degradation in air (Phototransformation). Indirect photolysis (OH-radical reaction, ozone-radical reaction, other)

**3.3.Comment on endpoint:**

Gas-phase reaction between photochemically produced hydroxyl radicals and organic chemicals at 25 °C and 1 atm for 460 heterogeneous organic chemicals [ref.3/ sect.2.7]. The units of the rate coefficient depend on the global order of reaction.

**3.4.Endpoint units:**

$\text{cm}^3\text{s}^{-1}\text{molecule}^{-1}$

**3.5.Dependent variable:**

$-\log(\text{OH})$

**3.6.Experimental protocol:**

Available at Atkinson, R. J Phys Ref Data 1989, Monograph 1,p1-246. [ref.3/ sect.9.2]

**3.7.Endpoint data quality and variability:**

Satisfactory models were also obtained in the past using the same dataset as well as AOPWIN package of EPI Suite have the same training set. The dataset is the famous and widely used Atkinson (1989, [ref.3/ sect.9.2]) set related to atmospheric reactivity.

## 4. Defining the algorithm - OECD Principle 2

### 4.1. Type of model:

QSAR

### 4.2. Explicit algorithm:

Multiple linear regression QSAR (OLS-Ordinary Least Square)

#### GA-OLS

The descriptors are: HOMO (Highest occupied molecular orbital energy), nX (number of halogen atoms), nCbH (number of unsubstituted sp<sup>2</sup>-carbon only in benzene-type rings) and IDE (Mean information content on the distance equality). See section 4.3 for a more detailed description of the four modeling descriptors.

#### Split model

Models were developed from three different training set of 191, 230, 230 compounds respectively based on structural similarity analysis (K-ANN, K-means) and random by sorting the response.

#### K-ANN

$$-\log(\text{OH}) = 3.70(\pm 0.64) - 0.76(\pm 0.06)\text{HOMO} + 0.17(\pm 0.03)\text{nCbH} + 0.36(\pm 0.06)\text{nX} - 0.35(\pm 0.10)\text{IDE}$$

#### Random

$$-\log(\text{OH}) = 4.06(\pm 0.70) - 0.73(\pm 0.06)\text{HOMO} + 0.39(\pm 0.06)\text{nX} + 0.15(\pm 0.03)\text{nCbH} - 0.34(\pm 0.11)\text{IDE}$$

#### K means

$$-\log(\text{OH}) = 3.60(\pm 0.68) - 0.78(\pm 0.06)\text{HOMO} + 0.18(\pm 0.03)\text{nCbH} + 0.35(\pm 0.06)\text{nX} - 0.37(\pm 0.09)\text{IDE}$$

#### Full model

Model developed on all available experimental data (training set of 460 compounds)

$$-\log(\text{OH}) = 4.07(\pm 0.48) - 0.72(\pm 0.04)\text{HOMO} + 0.37(\pm 0.04)\text{nX} + 0.16(\pm 0.02)\text{nCbH} - 0.34(\pm 0.07)\text{IDE}$$

### 4.3. Descriptors in the model:

[1]HOMO dimensionless Highest occupied molecular orbital energy. This descriptor characterizes the susceptibility of a molecule toward the attack by the electrophile OH radical, more reactive chemicals having higher HOMO energy

[2]nX dimensionless Number of halogen atoms. Molecules with more halogen atoms tend to have less reactivity

[3]nCbH dimensionless The number of unsubstituted sp<sup>2</sup>-carbon only in benzene-type rings. The descriptor nCbH, which is negatively correlated to the response in univariate models, is able to condense information on possible reactive sites in aromatic rings. The chemicals with higher number of hydrogen atoms can be more attacked by the hydroxyl radical and are, for this reason, more reactive

[4]IDE dimensionless Mean information content on the distance equality. This carries information regarding differences in atomic distribution and molecular dimension of chemicals.

### 4.4. Descriptor selection:

In this study we consider zero-, mono-, bi-dimensional descriptors in DRAGON 5.5 version and pruned those which are no longer in the updated version 6. Finally constant values and descriptors found to be correlated pair-wise were excluded in a pre-reduction step (one of any two descriptors with a K correlation greater than 0.95 was removed to reduce redundant and not useful information), thus obtaining a pruned set of 341 molecular descriptors.

Furthermore three quantum-chemical descriptors (Highest Occupied Molecular Orbital (HOMO) and Lowest Unoccupied Molecular Orbital (LUMO) energies, HOMO-LUMO gap) were added to above pool of descriptors. Therefore input sets of 344 (DRAGON and MOPAC) descriptors underwent the subsequent selection for the best modeling variables. The Genetic Algorithm-Variable Subset Selection (GA-VSS), by Ordinary Least Squares regression (OLS), included in MOBYDIGS (and now reproduced in QSARINS [ref 4,5; sect 9.2]), was applied to select only the best combination of descriptors from the input pool: 4 descriptors selected from 344.

#### **4.5. Algorithm and descriptor generation:**

Multiple linear regression (MLR) and variable selection by GA-VSS were performed by Ordinary Least Squares regression (OLS) in order to develop the model. Descriptors were generated by DRAGON 5.5 from HYPERCHEM optimized structures (\*.hin files). Quantum chemical descriptors were calculated by the semi empirical molecular orbital program MOPAC (AM1 method for energy minimization) in the software HYPERCHEM version 7.03.

#### **4.6. Software name and version for descriptor generation:**

DRAGON Software for the calculation of molecular descriptors, ver.5.5 for Windows, 2007 R.

Todeschini, V. Consonni, A. Mauri, M. Pavan

Chemical structures, drawn in HyperChem 7.03 and used as input file (\*.hin) for DRAGON 5.5, energy minimised using AM1 procedure. These structures are available in QSARINS (QSARINS-Chem module) enabling an end user to regenerate the descriptors for a new compound  
info@talete.mi.it

<http://www.talete.mi.it/>

#### **4.7. Chemicals/Descriptors ratio:**

Split Model: 47.75 (191 chemicals / 4 descriptors)

57.50 (230 chemicals / 4 descriptors)

Full model: 115 (460 chemicals / 4 descriptors)

### **5. Defining the applicability domain - OECD Principle 3**

#### **5.1. Description of the applicability domain of the model:**

Quantitative measures of a model applicability domain (AD) are needed to evaluate the degree of extrapolation and for the identification of problematic compounds.

Response and descriptor space:

Range of experimental  $-\log(\text{OH})$  values: 9.44 - 15.7

Range of descriptors values:

HOMO: (-)7.3- (-)13.68

nX: 0-6

nCbH: 0-10

IDE: 0-3.358

The chemical space of the model includes alkanes, alkenes, alcohols, halogenated chemicals, amines, aromatics, and other functional groups.

### **5.2.Method used to assess the applicability domain:**

AD was verified by the leverage approach [6] (for the structural domain), and by the identification of response outliers (compounds with cross-validated standardized residuals greater than 2.5 standard deviation units).

Graphically, the plot of hat values (h) versus standardized residuals, i.e. the Williams plot, verified the presence of response outliers and training set chemicals that are structurally very influential in determining model parameters (compounds with leverage value (h) greater than  $3p/n$  ( $h^*$ ), where p is the number of the model variables plus one, and n is the number of the objects used to calculate the model). For our model  $h^*$  is equal to 0.033 (number of variables in the model are four and total number of compounds is 460)

For new compounds without experimental data, leverage can be used as a quantitative measure for evaluating the degree of extrapolation: for compounds with a high leverage value ( $h > h^*$ ), that are structural outliers, predictions should be considered less reliable. In QSARINS the Insubria graph allows chemicals to be identified for which the predictions are inter- or extrapolated by the model.

### **5.3.Software name and version for applicability domain assessment:**

QSARINS 1.0 (verified also on version 2.2)

Software for the development, analysis and validation of QSAR MLR models, ver. 2.2, 2015

Paola Gramatica, email: [paola.gramatica@uninsubria.it](mailto:paola.gramatica@uninsubria.it)

<http://www.qsar.it/>

### **5.4.Limits of applicability:**

a) Some common compounds have been found as outliers or influential in all the models:

Outliers for response (standardized residuals > 2.5 standard deviation units):

Overestimated: triethyl phosphate (61) and

2-(chloromethyl)-3-chloro-1-propene (403)

Underestimated: bromomethane (18), dimethylsulfide (37), diethyl sulfide (263), ethyl methyl sulfide (353), 3-methyl-1,2 butadiene (342).

Outliers for structure (Hat cut off=0.033):

fluorinated chemicals: 1,1,2,2-tetrachloroethene (232),

1,1-dichloro-2,2,2-trifluoroethane (262), 1,1,1,2,2-pentafluoroethane

(265), hexafluorobenzene (267), 1-chloro-1,2,2,2-tetrafluoroethane (414)

and propylpentafluorobenzene (457)

b) Two classes of CADASTER chemicals, namely Polybrominated diphenylethers (PDBEs) and (benzo)triazoles (BTAs), were used to verify the applicability of our DRAGON model in the prediction of chemicals without experimental data.

All PBDEs are outside the AD. Chemicals with an increasing number of bromine atoms have the tendency to go far from the domain, and were extrapolated. For BTAZs 75% compounds are inside the domain and fluorinated compounds are structurally influential.

## **6.Internal validation - OECD Principle 4**

### **6.1.Availability of the training set:**

Yes

### **6.2.Available information for the training set:**

Yes

CAS RN: Yes

Chemical Name: Yes

Smiles: Yes

Formula: Yes

INChI: No

MOL file: No

### **6.3.Data for each descriptor variable for the training set:**

All

### **6.4.Data for the dependent variable for the training set:**

All

### **6.5.Other information about the training set:**

Type of chemicals: VOC (Volatile Organic Compounds)

Three different splitting procedures were adopted, two based on structural similarity analysis (K-ANN, K-means) and one random by sorting the response, in order to propose models that have a demonstrated high performance in predicting external chemicals of different typology, avoiding the bias derived from an unique split. The numbers of training set in three divisions (K-ANN, Random and K-means) are 191, 230 and 230 respectively.

### **6.6.Pre-processing of data before modelling:**

Transformation to logarithmic units and multiplied by -1 to obtain positive values

### **6.7.Statistics for goodness-of-fit:**

Here we have three different training sets for input set of descriptors.

Therefore we are reporting the statistical fittings of all the models.

$n_{\text{Training}}=191$ ,  $R^2=0.867$ ,  $R_a^2=0.864$ ,  
 $s=0.36$ ,  $F=302.25$

$n_{\text{Training}}=230$ ,  $R^2=0.826$ ,  $R_a^2=0.823$ ,  
 $s=0.42$ ,  $F=266.96$

$n_{\text{Training}}=230$ ,  $R^2=0.836$ ,  $R_a^2=0.833$ ,  
 $s=0.44$ ,  $F=286.83$

### **6.8.Robustness - Statistics obtained by leave-one-out cross-validation:**

$n_{\text{Training}}=191$ ,  $Q^2_{\text{LOO}}=0.856$ ;  $n_{\text{Training}}=230$ ,

$Q^2_{\text{LOO}}=0.817$ ;  $n_{\text{Training}}=230$ ,  $Q^2_{\text{LOO}}=0.827$ .

High value of  $Q^2_{\text{LOO}}$  (leave-one-out) means that the models, when verified for this technique of internal validation, are robust.

**6.9. Robustness - Statistics obtained by leave-many-out cross-validation:**

$Q^2_{LMO}$  was not calculated, since we calculated  $Q^2_{BOOT}$  (see 6.11).

**6.10. Robustness - Statistics obtained by Y-scrambling:**

$R^2_{ys} = 0.017-0.021$ ,  $Q^2_{ys} = 0.008-0.012$

(Values are in range for three splittings). The low values of Y-scrambled  $R^2$  and  $Q^2$  mean that the proposed models are not given by chance.

**6.11. Robustness - Statistics obtained by bootstrap:**

Split Models

$n_{Training} = 191$ ,  $Q^2_{BOOT} = 0.847$ ;  $n_{Training} = 230$ ,

$Q^2_{BOOT} = 0.809$ ;

$n_{Training} = 230$ ,  $Q^2_{BOOT} = 0.819$

Full Model

$Q^2_{BOOT} = 0.817$ .

The high value of  $Q^2_{BOOT}$  means that the models are robust and stable.

**6.12. Robustness - Statistics obtained by other methods:**

No information available

**7. External validation - OECD Principle 4****7.1. Availability of the external validation set:**

Yes

**7.2. Available information for the external validation set:**

CAS RN: Yes

Chemical Name: Yes

Smiles: Yes

Formula: Yes

INChI: No

MOL file: No

**7.3. Data for each descriptor variable for the external validation set:**

All

**7.4. Data for the dependent variable for the external validation set:**

All

**7.5. Other information about the external validation set:**

We have distributed our dataset into training and prediction set using three different splitting procedure. One based on response and two are based on structural similarity analysis confirming well balance in the training and prediction set both in response and structure. The number of external validation set in three divisions (K-ANN, Random, K-means) are 269, 230 and 230 respectively.

**7.6. Experimental design of test set:**

The random by response splitting was obtained by ordering the chemicals according to their descending kinetic constant value, and then putting the most and the least reactive in the training set and one out every two chemicals in the prediction set (50% of the full dataset). This splitting guarantees that the prediction set spans the entire range of

the experimental measurements and is numerically representative of the dataset.

The splitting of the data set realized by Kohonen Artificial Neural Network (K-ANN) takes advantage of the clustering capabilities of K-ANN, allowing the selection of a structurally meaningful training set and a representative prediction set [sect.9.2/ ref.7].

Another approach for splitting into training and prediction sets is by using K-means clustering which ensures that the similarity principle can be employed for grouping chemicals and splitting them in balanced training and prediction sets [sect.9.2/ ref.8].

#### **7.7.Predictivity - Statistics obtained by external validation:**

n Prediction=269,  $Q^2F1$ [ref 9; sect 9.2]=0.797,  $Q^2F2$ [ref 10; sect 9.2]=0.794,  $Q^2F3$ [ref 11; sect 9.2]=0.766, RMSE=0.47, CCC [ref 12,13; sect 9.2,]=0.889 n Prediction=230,  $Q^2F1$ [ref 9; sect 9.2]=0.819,  $Q^2F2$ [ref 10; sect 9.2]=0.819,  $Q^2F3$ [ref 11; sect 9.2]=0.810, RMSE=0.44, CCC [ref 12,13; sect 9.2]=0.903 n Prediction=230,  $Q^2F1$ [ref 9; sect 9.2]=0.804,  $Q^2F2$  [ref 10; sect 9.2]=0.802,  $Q^2F3$ [ref 11; sect 9.2]=0.836, RMSE=0.43, CCC [ref 12,13; sect 9.2]=0.899

The high values of external  $Q^2$ , calculated in different ways (see references for more details), and CCC, show that the proposed models are predictive for new chemicals. In fact, the models show good results when applied to the chemicals not used during the model development (chemicals in the prediction sets).

#### **7.8.Predictivity - Assessment of the external validation set:**

The response range value of training sets for three splitting is [9.44-15.7] and the prediction set responses range are [9.5-14.77], [9.6-14.77], [9.6-14.6] respectively for K-ANN, Random and K-means clustering procedure. Therefore, the three splittings guarantee a balanced distribution of chemicals in training and prediction sets regarding the response.

HOMO:

Training set

K-ANN [(-)8.12- (-)13.31]; Random [(-)8.12- (-)13.31];K-means [(-)8.19- (-)13.31]

Prediction set

[(-)7.3- (-)13.68] for all three splits

nX:

Training sets: 0-6; Prediction sets: 0-5

IDE:

Training set

K-ANN [0- 3.236];Random [0- 3.358];K-means [0- 3.358]

Prediction set

K-ANN [0- 3.358]; Random [0- 3.104];K-means [0- 3.104]

nCbH:Training sets: [0-10]Prediction set:

K-ANN [0- 10]; Random [0- 9];K-means [0- 10].Therefore, the three splittings guarantee a balanced



distribution of chemicals in training and prediction sets regarding the structure. The prediction sets are all large and representative of the training sets, therefore the models can be reliably applied to the external sets.

#### **7.9. Comments on the external validation of the model:**

Models selected by GA from three different splitting procedures (K-ANN, K-means, random) demonstrated high performance in predicting external chemicals of different typology avoiding the bias derived from unique split.

### **8. Providing a mechanistic interpretation - OECD Principle 5**

#### **8.1. Mechanistic basis of the model:**

The model was developed by a statistical approach. No mechanistic basis was defined a priori.

#### **8.2. A priori or a posteriori mechanistic interpretation:**

A posteriori mechanistic interpretation.

The most relevant combination of descriptors from a slightly different set of descriptors developed from updated versions of DRAGON is (HOMO, nX, nCbH, IDE).

Highest occupied molecular orbital (HOMO) energy. This descriptor characterizes the susceptibility of a molecule toward the attack by the electrophile OH radical, more reactive chemicals having higher HOMO energy. nX is the number of halogen atoms. Molecules with more halogen atoms tend to have less reactivity. nCbH is the number of unsubstituted sp<sup>2</sup>-carbon only in benzene-type rings. The descriptor (nCbH), which is negatively correlated to the response in univariate models, is able to condense information on possible reactive sites in aromatic rings. The chemicals with higher number of hydrogen atoms can be more attacked by the hydroxyl radical and are, for this reason, more reactive. Less important is the topological descriptor IDE, that carry information regarding differences in atomic distribution and molecular dimension of chemicals.

#### **8.3. Other information about the mechanistic interpretation:**

No information available

### **9. Miscellaneous information**

#### **9.1. Comments:**

The model is transparent in its reproducibility by DRAGON 5.5 software also by updated version. In order to predict chemicals without experimental activity it is suggested to use the full model developed from all available (n=460) chemicals with wider domain of applicability.

The statistical quality of full model

n=460, R<sup>2</sup>=0.824, Q<sup>2</sup><sub>LOO</sub>=0.819, Q<sup>2</sup><sub>BOOT</sub>=0.817,

RMSE<sub>tr</sub>=0.43, RMSE<sub>CV</sub>=0.43

The applicability of our models was verified on two classes of CADASTER chemicals, namely Polybrominated diphenylethers (PDBEs) and (benzo)

triazoles (BTAZs), comparing the predictions also with EPISuite (AOPWIN). It was verified that all the PBDEs are outside the applicability domain of our model, therefore the predicted data, which however are similar to those obtained by AOPWIN, are extrapolated. It was verified that for BTAZs almost 75% of the chemicals are within its applicability domain, therefore interpolated. In this regard special emphasis is given to applicability domain regarding interpolation and extrapolation.

## 9.2. Bibliography:

- [1] Gramatica P, Pilutti P & Papa E (2004). Validated QSAR Prediction of OH Tropospheric Degradation of VOCs: Splitting into Training Test Sets and Consensus Modeling. *Journal of Chemical Information and Computer Sciences* 44, 1794-1802. DOI: 10.1021/ci049923u
- [2] Roy PP, Kovarich S & Gramatica P (2011). QSAR Model Reproducibility and Applicability: A Case Study of Rate Constants of Hydroxyl Radical Reaction Models Applied to Polybrominated Diphenyl Ethers and (Benzo-)Triazoles. *Journal of Computational Chemistry* 32, 2386–2396. DOI: 10.1002/jcc.21820
- [3] Atkinson R (1989). Kinetics and Mechanisms of the Gas-Phase Reactions of the Hydroxyl Radical with Organic compounds. *Journal of Physical and Chemical Reference Data, Monograph* 1, 1-246. DOI: 10.1021/cr00071a004
- [4] Gramatica P et al (2013). QSARINS: A new software for the development, analysis and validation of QSAR MLR models. *Journal of Computational Chemistry (Software News and Updates)* 34 (24), 2121-2132. DOI: 10.1002/jcc.23361
- [5] Gramatica P et al (2014). QSARINS-chem: Insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS. *Journal of Computational Chemistry (Software News and Updates)* 35 (13), 1036-1044. DOI: 10.1002/jcc.23576
- [6] Gramatica P (2007). Principles of QSAR models validation: internal and external. *QSAR & Combinatorial Science* 26, 694-701. DOI: 10.1002/qsar.200610151
- [7] Gasteiger J & Zupan J (1993). Neural Networks in Chemistry. *Angewandte Chemie International Edition* 32, 503-527. [http://web.uni-plovdiv.bg/plamenpenchev/mag/files/ang\\_chem2.pdf](http://web.uni-plovdiv.bg/plamenpenchev/mag/files/ang_chem2.pdf)
- [8] Leonard JT & Roy K (2006). On selection of training and test sets for the development of predictive QSAR models. *QSAR & Combinatorial Science* 25, 235-251. DOI: 10.1002/qsar.200510161
- [9] Shi LM, Fang H, Tong W, Wu J, Perkins R, Blair RM, Branham WS, Dial SL, Moland CL & Sheehan DM (2001). QSAR Models Using a Large Diverse Set of Estrogens. *Journal of Chemical Information and Computer Sciences* 41, 186-195. DOI: 10.1021/ci000066d
- [10] Schüürmann G, Ebert RU, Chen J, Wang B & Kühne R (2008). External Validation and Prediction Employing the Predictive Squared Correlation Coefficient Test Set Activity Mean vs Training Set Activity Mean. *Journal of Chemical Information and Modeling* 48, 2140-2145. DOI: 10.1021/ci800253u
- [11] Consonni V, Ballabio D & Todeschini R (2009). Comments on the Definition of the Q2 Parameter for QSAR Validation. *Journal of Chemical Information and Modeling* 49, 1669-1678. DOI: 10.1021/ci900115y
- [12] Chirico N & Gramatica P (2011). Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *Journal of Chemical Information and Modeling* 51, 2320-2335. DOI: 10.1021/ci200211n
- [13] Chirico N & Gramatica P (2012). Real External Predictivity of QSAR Models. Part 2. New

### 9.3. Supporting information:

qmr321_DRAGON_OH_Training set _Full model	<a href="http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf478_qmrf321_DRAGON_OH_Training set _Full model.sdf">http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf478_qmrf321_DRAGON_OH_Training set _Full model.sdf</a>
qmr321_DRAGON_OH_k means_Trainingset	<a href="http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf478_qmrf321_DRAGON_OH_k means_Trainingset .sdf">http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf478_qmrf321_DRAGON_OH_k means_Trainingset .sdf</a>
qmr321_DRAGON_OH_Random_Trainingset	<a href="http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf478_qmrf321_DRAGON_OH_Random_Trainingset.sdf">http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf478_qmrf321_DRAGON_OH_Random_Trainingset.sdf</a>
qmr321_DRAGON_OH_KANN_Trainingset	<a href="http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf478_qmrf321_DRAGON_OH_KANN_Trainingset .sdf">http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf478_qmrf321_DRAGON_OH_KANN_Trainingset .sdf</a>
qmr321_DRAGON_OH_k means_Predictionset	<a href="http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf478_qmrf321_DRAGON_OH_k means_Predictionset.sdf">http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf478_qmrf321_DRAGON_OH_k means_Predictionset.sdf</a>
qmr321_DRAGON_OH_Random_Predictionset	<a href="http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf478_qmrf321_DRAGON_OH_Random_Predictionset.sdf">http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf478_qmrf321_DRAGON_OH_Random_Predictionset.sdf</a>
qmr321_DRAGON_OH_KANN_Predictionset	<a href="http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf478_qmrf321_DRAGON_OH_KANN_Predictionset.sdf">http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf478_qmrf321_DRAGON_OH_KANN_Predictionset.sdf</a>

## 10. Summary (JRC QSAR Model Database)

### 10.1. QMRF number:

Q17-22b-0055

### 10.2. Publication date:

2017-09-27

### 10.3. Keywords:

DRAGON;hydroxyl;tropospheric degradation;QSARINS;INSUBRIA;

### 10.4. Comments:

old# Q47-19-49-478