**RESPONSE TO REVIEWERS**

**Quantitative structure-retention relationship model for predicting retention indices of constituents of essential oils of *Thymus vulgaris Lamiaceae***

YOUSSOUF DRIOUCHE1, DJELLOUL MESSADI1[[1]](#footnote-2)\*

*1* *Environmental and Food Safety Laboratory, Badji Mokhtar-Annaba University, BP.12, 23000 Annaba, Algeria*

**THE RESPONSE TO QUESTIONS**

**1. P2, Line 66: "algoRIthm" MEANS WHAT?**

It's algorithm.

**2.** **How do you justify 36 samples as sufficient sample size?**

The majority of QSRR models have been generally built on data sets of sizes ranging from 25 to 169 compounds, with most of these belonging to a single chemical series.

The aim of this study is to develop QSRR model for the RI prediction of 36 essential oil components, representing chemical variation of leaf essential oil, at different stage of (*Thymus vulgaris Lamiaceae*) of Iran plant growth [A. Nezhadali, M. Nabavi, M. Rajabian, M. Akbarpour, P. Pourali, F. Amini, *Beni-Seuf Univ. J. Appl. Sci*. **3** (2014) 87 (<https://doi.org/10.1016/j.bjbas.2014.05.001>)].

**3. Why KS method is used and not random method? KS usually give high R2
value.**

The simple random sampling is the most current method for fractionation of the data in development of the models, were the data are selected with a uniform probability. Sampling randomly is simple and easy to be realized and can effectively be carried out in only one passage on the data by using algorithms such as the algorithm of Knuth [D. E. Knuth, *The Art of Computer Programming, Volume 2 (3rd Ed.): Seminumerical Algorithms*, Addison-Wesley Longman Publishing Co., Inc, Boston, MA, USA, 1997].

However, the problem with this approach it is a chance that the scission of data suffers from the variance, or of partiality, in particular when the data are not distributed uniformly [G. D. [Tourassi](https://www.ncbi.nlm.nih.gov/pubmed/?term=Tourassi%20GD%5BAuthor%5D&cauthor=true&cauthor_uid=11797941), E. D. [Frederick](https://www.ncbi.nlm.nih.gov/pubmed/?term=Frederick%20ED%5BAuthor%5D&cauthor=true&cauthor_uid=11797941), M. K. [Markey](https://www.ncbi.nlm.nih.gov/pubmed/?term=Markey%20MK%5BAuthor%5D&cauthor=true&cauthor_uid=11797941), C.E. Jr. [Floyd](https://www.ncbi.nlm.nih.gov/pubmed/?term=Floyd%20CE%20Jr%5BAuthor%5D&cauthor=true&cauthor_uid=11797941).*Med. Phys.* **28** (2001) 2394 (<https://doi.org/10.1118/1.1418724>)].

**4. why training set is 75%, and test set 25%. Justify.**

The most effective method of validating a regression model with respect to its prediction performance is to collect fresh data and directly compare the model predictions against it. When this is not possible, a reasonable procedure is to split the available data into two parts: a training set from which the model is built and external set on which to evaluate its prediction power, the later should contain between 15 and 40% of the compounds in the full data set [E. Benfenati, J. R. Chrétien, G. Gini, N. Piclin, M. Pintore, A. Roncaglioni, *Validation of the models.* *In* Quantitative Structure-Activity Relationships (QSAR) for Pesticide Regulatory Purposes, Elsevier, (2007).  p. 185-199.  ([https://doi.org/10.1016/B978-044452710-3/50008-2)](https://doi.org/10.1016/B978-044452710-3/50008-2%29)]. The whole data was split arbitrarily into a 27 samples (75%) set and a 9 samples (25%) set.

**ADDITIONAL COMMENTS**

We had done some modifications in the text of the paper as following:

* In Introduction from line 50 to line 85 (Page 2, 3).
* In Experimental Data from line 89 to line 93 (page 3).
* In Training set and test set from line 105 to line 114 (page 4).
* In Chemometric Methods from line 121 to line 129 (Page 5).

And in Conclusion from line 244 to line 252 (Page 10).

1. \*Corresponding author. E-mail: d\_messadi@yahoo.fr [↑](#footnote-ref-2)