Atmos. Chem. Phys. Discuss., 10, C1612–C1614, 2010
www.atmos-chem-phys-discuss.net/10/C1612/2010/


Atmospheric Chemistry and Physics Discussions



# *Interactive comment on* "Structure-activity relationships to estimate the effective Henry's law coefficients of organics of atmospheric interest" *by* T. Raventos-Duran et al.

**Anonymous Referee #1**

Received and published: 14 April 2010

General comments:

Using their newly conceived SAR, the authors of this article attempt to predict Henry's law constants for a large variety of organic compounds, many of which are of atmospheric interest. The article is well written and the subject is of relevance to atmospheric science, since Henry's law constants are of use to atmospheric models which address the partitioning of volatile organic compounds to the solution phase. For this reason, I would recommend publication of this article after the following points have been considered:

Specific comments:

As was mentioned in section 3.2, it is generally understood that multilinear regression models are prone to overfitting, hence there is a requirement for training sets in order to assess the predictive qualities of the model. There is also a necessity to reduce the number of variables used as model parameters for the same reason. Despite this, GROMHE includes unique descriptors for many different types of substitutions, such as peracid, which has only one entry in the entire database. My question is: have the authors considered a more generalized approach to the description of substitutions? The peracid group, for example, contains a hydroperoxide group and a carbonyl group, both of which are accounted for in the database. Another example would be ketones and aldehydes, where both contain a carbonyl oxygen atom and the only difference is an alkyl substitution. Therefore, would it be possible to describe many of these substitutions using smaller fragments, which are more general and better represented in the database?

If possible, the authors should provide more interpretation of their multilinear regressions. There is almost no explanation of the SAR regarding the mechanism by which Henry's law constants are affected by substitution. Exactly how does the presence of an electronegative substitution neighbouring another substitution affect the Henry's law constant and why? Questions such as this can become difficult to answer as the multilinear regression becomes more complex and the physical interplay between the descriptors becomes less obvious. So, is this approach just a black box, or can some physical interpretation be made about the results of this study?

In the final paragraph of this article a caveat is presented to the effect that the database that is used represents a subset of the available data. GROMHE was optimized for this database, but is compared with HWINb and SPARC, which are optimized for other databases. This is an inherently unfair comparison, and the general increase in performance associated with GROMHE is called into question. Since these models are complex and there may be a lot of work associated with training them, perhaps it is unreasonable to ask for these models to be re-optimized using the present database.



ACPD

10, C1612–C1614, 2010

However, is it possible to optimize GROMHE to the databases associated with HWINb and SPARC? This would result in a fair comparison and would add much more weight to any assertion that GROMHE possesses better predictive power.

Technical corrections:

Page 4618, line 25: Given that many of these compounds are likely to possess large, non-polar substitutions (e.g. alkyl groups), surely a large contingent of insoluble species is also expected.

Page 4619, line 27: SAR should be pluralized to SARs.

Page 4620, line 2: SAR should again be pluralized to SARs.

Page 4620, line 22: The employment of the acronym seamlessly into this sentence requires some revision. For example: "Furthermore, this database was used to develop a new SAR: the GROup contribution Method for Henry's law Estimate (hereafter named GROMHE).

Page 4623, line 4: constant should be plural.

Page 4623, line 22: carbon (atom).

Page 4632, line 3: log unit(s).

Page 4632, line 4: The sentence starting "This best agreement. . ." is phrased poorly and should be rewritten.

---

Interactive comment on Atmos. Chem. Phys. Discuss., 10, 4617, 2010.