Atmos. Chem. Phys. Discuss., 10, C3344–C3350, 2010
www.atmos-chem-phys-discuss.net/10/C3344/2010/

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 #2

Received and published: 29 May 2010

General Comments

The authors compile effective Henry's law coefficients (H) in air-water systems at 298 K from literature, for a large group (488) of organic compounds bearing functional groups of atmospheric relevance. From this data set, GROMHE, a new Structure Activity Relationship (SAR) based on a group contribution approach, for estimating Henry's law constants for organic compounds of atmospheric interest in air-water systems is developed. The reliability of GROMHE, together with those of the existing HENRYWIN bond contribution method and SPARC approach, for the determination of H are then assessed.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

Overall, the manuscript is well written. The language and structure of the presentation are clear with very few ambiguities. The presentation of methods, arguments and results is easy to follow for someone who is familiar with the core of the suject matter, but has not worked with the specific area of model development, or with the statistical and structure contribution methods involved. The presentation of results in the figures is intuitive and the tables provided in the supporting material are convenient and valuable compilations, for reading reference as well as for other applications.

Describing the partitioning of organic compounds between the gas and condensed phases in the atmosphere is a key step in determining the amount, properties, fate, and effects of atmospheric organic aerosol. As the level of detail in atmosperic modelling increases in terms of both number and complexity of the compounds and processes considered, there is a corresponding need to describe the properties and interactions of all the species involved. Since experimental data for the relevant parameters cannot be obtained for the very wide composition and concentration ranges found in the atmosphere, reliable estimation methods become essential. The authors present a tool for prediction of the Henry's law coefficient, which is one such key property for the description and model inclusion of a number of central atmospheric processes, as for example related to vapor-aerosol-cloud interactions. I therefore believe the results presented by the authors are relevant and useful for application to several areas of the atmospheric sciences, and are thus well-suited for publication in Atmospheric Chemistry and Physics.

Specific Comments

- "Most secondary organics are expected to be water soluble" (P4618 L 25): Although it is a fringe-note, this statement is ambiguous and can be misleading.

Almost anything can be dissolved to some extent in water, depending on the relative amounts of water and organic present, as well as other specific (intensive) properties of the system. I suggest to give explicit (for example, order of magnitude of some property) gauges for the term "water soluble", or reformulate the statement.

- I suggest chemical and mathematical equations defining central concepts are emphasized, and specifically not stated as regular text. In particular, there are different forms of Henry's law given in the literature. The version used in the present work for defining the Henry's law constant (H) is given in {P4619 L 11}, but should be easier to locate. This would from the beginning avoid any ambiguity concerning the definition of this central concept and immidiately clarify exactly what quantity the developed model predicts. It is especially important since the authors use the inverse form of Henry's law compared to several of the references provided (e.g. Hilal et al. (2008))
- The authors refer to both the Henry's law "constant" and the Henry's law "coefficient" (e.g. P4624). The view that may emerge is a distinction between the "*intrinsic* constant" and the "*effective* coefficient". If such a distiction is intended, it may be too subtle, and I suggest the different quantities are explicitly defined. If the different terminology is used for the same parameter, I suggest this ambiguity between a constant and a coefficient (that can vary as a function of various system parameters) is avoided.
- As the authors note, the version of Henry's law used for defining the Henry's law constant is a limiting law (P4619 L 13), without activity coefficient correction and thus pertaining to the infinite dilution state of the solute. An underlying assumption of the SPARC method (Hilal et al., 2008) is that it also applies in the solubility limit of a saturated solution (P4620 L 10). As mentioned in several of the references provided, the variability or uncertainty related to Henry's law con-

stants measured by composite methods is in some cases related to confusion of states and the extrapolation of Henry's limiting law to the solute solubility limit. A clarifying comment about the uniformity of the reference state assumptions involved at the different stages in the model development and the significance of the predictions with GROMHE to atmospheric applications (where aqueous aerosol solutions are rarely dilute) might be useful. In fact, the discussion might benefit from adding a short theory section clarifying these ambiguities.

- The authors might consider spending a few lines describing the difference between the GROMHE, HWINb and SPARC models, since the performance of GROMHE in comparison to these existing tools is a main result presented. What the significant new contribution of the GROMHE, compared to the HWINb and SPARC? Besides the argued improvement of statistics, any new concepts should be more explicitly emphasized to clarify this further.
- P4621: The authors point out that the data set is very limited for some compound classes and this affects the reliablility and/or predictive ability of the model for other such compounds. I understand that the purpose of this work is to develop the GROMHE tool from already existing data. Nevertheless, it would be most useful for targeting future experimental work if the authors could provide recommendations and/or guidelines for measurements that would specifically facilitate further optimization of the model developed.
- P4622 L7: "The data were exclusively taken from experimental values either from direct or indirect measurements." As opposed to what - which other forms of experimental values could have been used? I have an idea what the authors mean, but perhaps this could be clarified.
- P4622 L21: What is the sensitivity of the model to the estimated variation in the desolvation enthalpy (ΔH_{solv})? Although I understand that the experimental uncertainty for H is generally large, it would be good if the authors could provide a

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

remark about whether such sensitivity is comparable to or minor in comparison with these uncertainties, and for how many compounds the temperature extrapolation is relevant.

- The model predicts H at 298 K, but how sensitive are model predictions with GROMHE to atmospherically relevant temperatures? What do the authors believe is the valid temperature range of the model? What is considered to be the relevant temperature range for atmospheric applications? For example, how would the model perform for, say, 278 K?

- P4624: The authors address hydration of carbonyls and the effect on observed effective Henry's law coefficients, as well as the means to account for this effect in the developed model. What about the effects of other solute-solvent equilibria, e.g. hydrolysis reactions of acid functionalities and protonation of bases?

- P4624 L18: This paragraph starts very abruptly and it is difficult to follow. The purpose of the paragraph is not immediately clear, if you are not already familiar with the concepts, as it introduces and discusses new quantities without much explanation. For example, what is a "descriptor", and what is the purpose of introducing it? I was also confused about the sudden discussion of aromatics, aliphatics, then aldehydes and ketones: Maybe "Taft and Hammett σ" values (P4624 L22) could be briefly defined and put into context?

- The greatest difficulty for me in assessing the methods and results presented by the authors concerns the statistical methods and parameters for which values are provided. Although it may be argued that this should ideally be basic knowledge to everyone working in science, at least I lack the proper knowledge foundation to evaluate the quality and performance of the developed model. Specifically,

  1. What is precisely meant by "reliability", "quality", "prediction ability", "contribution", and "weight in the regression"? What are the definition and units of

the "standard errors" of the descriptors?

2. What is the significance/interpretation of the different types of errors discussed, i.e. the RMSE, MAE, MBE (P4625, P4628)? How do these quantities assess the model and method? What does a certain value mean? What is "good" or "bad"? Also, the "n" in the defining equations is unspecified?
3. It seems like a "Catch 22" that the GROMHE is evaluated in comparison with other models against the data set on the basis of which it was developed? I see this is noted by the authors themselves in the conclusions section (P4632 L5), but again, some remark elaborating a bit further on the significance of this could be very informative.

I believe it would be a great benefit if the central statistical parameters and their interpretation was explained shortly, since it is the foundation of the development and evaluation of the model described. It might furthermore enable the reader to better appreciate the significance and quality of the results. Naturally, the authors cannot write a new statistics text book, but a little more basic information might greatly improve readability to people outside the specific field, notably the end users of the model developed. This would enable the users to evaluate the strengths and weaknesses of the model for the various application purposes. Such a statistics summary could be included as supporting material or an appendix.

Technical Comments

- P4620 L 18: Pluralize "Henry's law constant".
- Tables in supporting material (1, 2): Regarding the typography of the molecular formulae, the authors could

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

1. use the "back-slash{chem}" command supported in the Copernicus LaTeX package.
2. capitalize the appropriate elemental designation letters.
3. clarify what is meant by "SMILES" (table T1 and T2, leftmost columns' headings): at least I am not familiar with this term.

- Tables in the main text (1, 2, 3): When the tables use terms and abbreviations that must be well understood in order to understand the meaning of the table, and reference is given to the body text for explanation, maybe it would be good to give the specific section where the concept is defined.

REFERENCES

Hilal, S. H., Ayyampalayam, S. N., and Carreira, L. A. (2008). Air-Liquid Partition Coefficient for a Diverse Set of Organic Compounds: Henry's Law Constant in Water and Hexadecane. *Environ. Sci. Technol.*, 42:9231–9236.

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