Geosci. Model Dev. Discuss., 8, C3382–C3385, 2015
www.geosci-model-dev-discuss.net/8/C3382/2015/
© Author(s) 2015. This work is distributed under
the Creative Commons Attribute 3.0 License.

Geoscientific Model Development Discussions

Interactive comment on "UManSysProp: an online facility for molecular property prediction and atmospheric aerosol calculations" *by* D. Topping et al.

S. Takahama (Referee)

satoshi.takahama@epfl.ch

Received and published: 17 December 2015

The authors introduce a new software tool that implements peer-reviewed methods for estimation of physical and chemical properties in pure compound phases, bulk mixtures, and individual particles. These properties include vapor pressures, solution densities, activity coefficients, and hygroscopic growth factors. Description of these methods with example applications are provided. The manuscript describes the method of programmatic mapping of compound structure to input parameters for various estimation methods using chemoinformatic techniques, and the construction of a software program that enables users to access this functionality through a central repository.

The interface is provided through both web and command line interface to accommodate various user types, but in a "software as a service" model like E-AIM (Wexler and Clegg, 2002; http://www.aim.env.uea.ac.uk/aim/aim.php) and AIOMFAC (Zuend et al., 2011; http://www.aiomfac.caltech.edu), the code and computation is hosted on the authors' servers. The software and technical description is of high quality and presents an important contribution to both the modeling and experimental community, and is recommended for publication in GMD after the comments below are considered.

Wexler, A. S. and Clegg, S. L. (2002): Atmospheric aerosol models for systems including the ions H+, NH4+, Na+, SO42−, NO3−, Cl−, Br− and H2O. J. Geophys. Res. 107, No. D14, art. no. 4207, doi:10.1029/2001JD000451.

Zuend, A., Marcolli, C., Luo, B. P., and Peter, T. (2008): A thermodynamic model of mixed organic-inorganic aerosols to predict activity coefficients, Atmos. Chem. Phys., 8, 4559–4593, doi:10.5194/acp-8-4559-2008.

General comments:

The description of the user interface and file formats for information exchange is mentioned in the main text in Sections 1 and 3, but it would be helpful for readers if a consolidated section is dedicated to its description and kept separate from the properties being predicted. In Appendix A1, it might be helpful to point out to the wide array of potential readers that "any machine" and "other platforms" specifically includes Microsoft Windows and Mac OS X, among other operating systems.

The reason for separation from Figure 2 from Figure 1 is unclear, as activity coefficients are bulk parameters (and therefore can be included in Figure 1) and SMARTS patterns are also used for property prediction (and therefore the requirement of SMARTS is not unique to the task depicted in Figure 2). It seems the two figures can be merged to give an overall picture of UManSysProp.

It would be nice to see the authors expound upon the statement (Section 3, p. 9764):

"For techniques used in UManSysProp, an extensive manual analysis of compounds used in the MCM (Jenkin et al., 2012), and a subset of GECKO mechanism (Aumont et al., 2005), were used to validate derived SMARTS libraries." As the authors note, the formulation of SMARTS patterns require special care to target specificity. Having realized these difficulties in our own work (http://www.atmos-chem-phys-discuss.net/15/33631/2015/), we have dedicated a technical note to present methods for validation and the structure of compounds and range (number of groups per compound) for which our formulated patterns were tested. While exhaustive description of the validation process is clearly not the focus of this work, many of the results rest on the correct enumeration of functional groups in molecules. Therefore, further discussion of the validation that the authors have already conducted can benefit users of the software and those seeking to adopt chemoinformatic approaches for structural queries. Additionally, some of the example property estimates presented in the current manuscript to some extent serve as even further validation.

Section 2, Parsing: introduction of examples containing contrasting primary -OH structures and how each are matched by one of the five groups defined might illustrate the elegance of this approach better to interested readers.

Minor comments:

The tables in appendices A1-A3 would be better served if provided (also) as text files in supporting information, as successful copying tabular data from PDFs can be depending on PDF viewer.

Regarding the web interface, for future versions it may be helpful to include the possibility of providing estimates of pure component properties (e.g., vapor pressures) from multiple methods in the same output file. It is likely that one of the uses for this technique will be to assess uncertainties in estimated properties and such a feature will allow users to more directly answer this question.

p. 9671, line 18: "It isn't clear" -> "It is not clear"

p. 9684, line 10: 'by appending ?' -> 'by appending "?"' or 'by appending a question mark'

p. 9675, line 5: "Fig. 2" -> "Figure 2"

p. 9676, line 10: "xx species"

p. 9678, line 8: "Sub cooled liquid" -> "Sub-cooled liquid"

p. 9676, line 27: "parenthese:" -> "parentheses:"

Interactive comment on Geosci. Model Dev. Discuss., 8, 9669, 2015.