Geosci. Model Dev. Discuss., 8, C3794–C3799, 2016
www.geosci-model-dev-discuss.net/8/C3794/2016/

Geoscientific Model Development Discussions

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

D. Topping et al.

david.topping@manchester.ac.uk

Received and published: 25 January 2016

We would like to thank the reviewer for comments provided which we feel has improved the submitted manuscript. In the following text we respond to all comments made and highlight where changes have been made in the next version of the manuscript.

General comments:

Referee comment *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.
Response: We agree with this suggestion and have moved relevant details into a new section 3.1, incrementing further sections accordingly. The new section reads as follows

The UManSysProp website first provides a portal where users can enter or upload a SMILES string and predict the property of interest. Examples of supplying SMILES strings via the input are given in section 3.1 and 3.2. Whilst users have the option to display output on a new webpage via HTML as the default option, the following download options are also available. For more information on their use, please refer to the references given in parentheses
- HTML (view in web browser)
- Excel file
- Python pickle file https:// docs.python.org/ 2/ library/ pickle.html
-XML file https:// en.wikipedia.org/ wiki/ XML
-Zipped CSV file
-JSON file https:// en.wikipedia.org/ wiki/ JSON

If you want to access UManSysProp without using a web-browser we also provide a programmer friendly JSON API that enables you to call our suite of tools from your own code. This is described in detail on our ReadTheDocs.org webpage https:// umansysprop.readthedocs.org/ with an example provided in the Appendix where we briefly discuss future expansions.
In addition, this section is now referred to in the introduction during a breakdown of the manuscript and specific details moved to new section 3.1.We have also added the point regarding Windows and Mac OSX via the following sentence in section 6.2 *'[..] provided you have Python 2.7 or greater. This includes Microsoft Windows, Mac OS X*

and other operating systems.'

Referee comment *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.*
Response: This has now been changed and figure 2 removed from the document, including reference to it in section 2.1.

Referee comment *It would be nice to see the authors expound upon the statement (Section 3, p. 9764): C3383 "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-chemphys-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.*

Response: This is an important point and the referee is quite right in raising it. It can

be easy for developers to generate generic SMILES and introduce errors with regards to specificity. In the text that follows the brief introduction to this challenge, we discuss the specific approaches for ensuring certain predictive techniques captured the correct structural features for compounds of atmospheric interest from chemical mechanisms. This was actually carried out manually in the first instance. The facility does record atomic information and uses this to check there are no over- and under-counting of individual atoms by the construction of required functional groups. However we do not have a more detailed mechanism in place to check this for compounds outside of the chemical mechanisms mentioned. To make sure the reader is aware of this, we suggest the following addition to the end of section 2.1 *'All of the above checks of specificity were carried out by hand for atmospheric chemical mechanisms. Whilst the current facilities check for under- or over-counting of atoms for any given set of functional groups, a future development would need an automatic method of checking specificity for compounds falling outside of this subset following the discussions presented by Ruggeri and Takahama (2015).'*
Ruggeri, G. and Takahama, S.: Technical Note: Development of chemoinformatic tools to enumerate functional groups in molecules for organic aerosol characterization, Atmos. Chem. Phys. Discuss., 15, 33631-33674, doi:10.5194/acpd-15-33631-2015, 2015.
Referee comment *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.*
Response: We have now removed the lines *'As noted in that paper, whilst it is 'easy' to identify all primary alcohols (SMARTS 'a' in the table), the Nannoolal method requires primary alcohols to be split between NG 35 (carbon chain of 5 or more atoms with nomenclature defined in Nannoolal et al. (2008)) and NG 36 (primary alcohols on a C4 or smaller chain) although the exact criteria for this split is not clear in the literature. For our applications, the allocation of primary alcohols is achieved using a set of five SMARTS.'* and added the following text to illustrate this, taken from the Barley et al

(2011) paper: *'It is easy enough to identify all primary alcohols (SMARTS a in Table 3) but the Nannoolal method requires primary alcohols to be split between NG_35 (carbon chain of 5 or more atoms) and NG_36 (primary alcohols on a C4 or smaller chain) although the exact criteria for this split is not clear in the literature. In our work (Barley et al 2011) the allocation of primary alcohols is achieved using a set of five SMARTS. SMARTS b in Table 3 identifies whether the primary alcohol is on a "carbon" chain of 5 or more atoms. This chain has to be terminated by carbon atoms (which may bear functional groups that are not part of this count), but the intermediate atoms can be N or O as well as C. Hence (using SMILES notation) OCCCO and OCCCCO would both have two alcohol groups belonging to NG_36 while OCCCCC, OCCOCC and OCCN(C)CC would have primary alcohols belonging to NG_35. The other three SMARTS account for the possible branching of this heavy atom chain:- thus OCC(C)(C) and OCN(C)C would both be NG_36 alcohols while OCC(C)(C)C and OCN(C)CC would be NG_35 alcohols.'*
We have now also released all of the source code behind UManSysProp, despite including the web-portal and JSON API, so users can study SMARTS for all methods. By using the OpenBabel framework, we would encourage the user community to cycle through compounds of interest.

Minor comments:

Referee comment *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.*
Response: We now provide these tables as text files in the supporting information.

We agree on changing output options for future use and will respond to any similar requests during the initial uptake.
Referee comment *p. 9671, line 18: "It isn't clear" -> "It is not clear" C3384*
Response: This has now been corrected.
Referee comment *p. 9684, line 10: 'by appending ?' -> 'by appending "?"' or 'by appending a question mark'*
Response: This has now been corrected.
Referee comment *p. 9675, line 5: "Fig. 2" -> "Figure 2" p. 9676, line 10: "xx species"*
Response: This has now been corrected.
Referee comment *p. 9678, line 8: "Sub cooled liquid" -> "Sub-cooled liquid"*
Response: This has now been corrected.
Referee comment *p. 9676, line 27: "parenthese:" -> "parentheses:"*
Response: This has now been corrected.

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