Reviewer 3

Reviewer's comment No. 1 — The authors utilize machine learning to predict saturation vapor pressure and two equilibrium-partitioning coefficients for gas-particle partitioning. For training and validating the machine learning model they use a dataset obtained by COSMOtherm calculations of theses observables for atmospheric oxidation product molecules.

The paper is well written, the topic timely and of great interest for the readers of ACP and I recommend publishing but ask the authors to take the following comments and suggestions into account.

I have one very general concern, which does not relate to the machine learning approach presented here, but to the underlying COSMOtherm data set. The authors write (e.g. line 49 page 2) that the COSMOtherm predictions have an order of magnitude accuracy. However, for a number of compounds at low saturation vapor pressures there have been studies comparing experimental saturation vapor pressures with COSMOtherm predictions and finding much larger deviations (e.g. Bannan et al., 2017, Krieger et al. 2018). It should be pointed out that the COSMOtherm model has been "calibrated" with a parametrization dataset of known compounds, which are potentially biased to high saturation vapor pressures (Klamt et al. 1998). Therefore, the accuracy of the underlying reference data may be only several orders of magnitude for low saturation vapor pressure components.

Authors' reply: We completely agree. By "order of magnitude" we meant "at best order of magnitude", to contrast with the factor of 3.7 quoted in the COSMOtherm documentation. Fortunately, proper consideration and selection of conformers, as well as improvements to the H-bonding treatment in newer versions of COSMOtherm, are slowly decreasing the disagreement between the saturation vapor pressure predictions and the limited number of experimental data points for atmospherically relevant low-volatility polyfunctionals. As noted in our reply to reviewer 2, our current best estimate, based on direct comparisons to the very limited number of available experiments on relevant compounds (see e.g. Kurtén et al 2018, Krieger et al 2018), is that the error margin of the computed saturation vapor pressures are probably around an order of magnitude for moderately complex (2-3 functional groups) molecules, possibly increasing by as much as a factor of 5 per intra-molecular hydrogen bond. This has now been noted in the manuscript as discussed above.

Reviewer's comment No. 2 — For gas-particle partitioning, the saturation vapor pressure range from about 10-11 kPa to about 10-3 kPa is relevant (e.g. Valorso et al. 2011, or the discussion starting in the last paragraph of page 2). However, Fig. 3c shows that there are hardly any molecules in the dataset below 10-8 kPa. Actually about half of the dataset contains compounds, which will be entirely in the gas phase under atmospheric conditions. Does this pose a problem?

Authors' reply: Yes, this poses a serious problem for predicting volatilities of large and complex molecules, and because of it, this study should be considered a proof-of-concept pilot for finding

appropriate combinations of descriptors and machine learning algorithms. We are in the process of performing additional COSMOtherm calculations and the corresponding machine learning on a substantially larger and much more complex set of compounds generated by the GECKO algorithm. We hope to be able to report preliminary result on this work relatively soon. Plans for future directions have been added to the conclusions - section of the manuscript.

Reviewer's comment No. 3 — Related: the last paragraph on page 6 states that Wang's dataset is rather small for machine learning but internally consistent. I intuitively understand that this helps the machine-learning model to succeed in predicting well. However, the authors write that Sanders's dataset for 17350 Henry's law constant are not internally consistent (as Wang's dataset). But what if the Sander's data are the correct ones? What if the real world is more complex than what is predicted by COSMOtherm? Would the machine learning approaches fail because it there are no easy "rules" the machine-learning algorithm can pick out of the dataset? Would the output of a model trained with these data just produce random partitioning coefficients within the range of the data set? These questions are probably impossible to answer without doing the experiment. It would have been very interesting to see how the machine-learning model perform on the dataset of Sander, but this is clearly beyond the work presented here.

Authors' reply: By Sanders's dataset "not being internally consistent" we mean primarily the fact that this (impressively large) set often contains multiple entries for the same compound (corresponding to e.g. different experimental studies, often with different methods), and the actual values can vary widely. For example for many polyols, Henry's law constants in the dataset vary by 6 orders of magnitude or more. This result can obviously not be correct, as a particular compound must have precisely one Henry's law constant at one temperature. This has been clarified in the manuscript. The other type of "internal inconsistency" (or complexity) presumably referred to by the reviewer would be e.g. strong non-additivity of the effects of various functional groups, and/or cases where very small differences in structures lead to very large differences in properties. We agree that the real world contains examples of this type of inconsistency or complexity, though typically the most extreme cases tend to be for chemical reactivity rather than physical molecular properties. Certainly such complexity also makes it more challenging to define rules for predicting properties based on structures (i.e. structure-activity or structure-property relationships). COSMOtherm is able to account for some, but probably not all, of these cases, as evidenced from the discussion on the effects of intra-molecular H-bonds also in the references cited by the reviewer. We agree that experimental methods capable of probing volatilities of very complex molecules will be needed to definitively answer the question.

On a final note, uncertainties in the data, e.g. experimental noise, can easily be taken into account in probabilistic machine learning models. We are working on such probabilistic models and will report their results in a future publication. It has to be emphasized, however, that even a noisy dataset has to be internally consistent. If Henry's law constants differ by 6 orders of magnitude for a certain compound, the dataset needs to be refined.

"For example, the Sander dataset contains several molecules with multiple entries for the same property, sometimes spanning many orders of magnitude."

Reviewer's comment No. 4 — I find section 2.2.4 rather brief. For me – being not familiar

with the topic – it is not possible to follow despite Fig. 4d. May be extent a bit?

Authors' reply: We improved the description as follows:

"TopFP first extracts all topological paths of a certain lengths. The paths start from one atom in a molecule and travel along bonds until k bond lengths have been traversed as illustrated in Fig. 4d. The path depicted in the figure would be OCCO. The list of patterns produced is exhaustive: Every pattern in the molecule, up to the pathlength limit, is generated. Each pattern then serves as a seed to a pseudo-random number generator (it is "hashed"), the output of which is a set of bits (typically 4 or 5 bits per pattern). The set of bits is added (with a logical OR) to the fingerprint. The length of the bitvector, maximum and minimum possible path lengths k_{max} and k_{min} and the length of one hash can be optimized. "

Reviewer's comment No. 5 — Discussion on page 16: Related to my comments above, without experimental vapor pressures for the C10 compounds being available, this discussion is interesting, but there may be surprises if experimental vapor pressures become available. I feel the authors should clearly state that the COMOtherm predictions are not validated in this pressure regime at all.

Authors' reply: We agree, and this has now been stated explicitly.

"However, we caution that COSMOtherm predictions have not yet been properly validated against experiments for this pressure regime. As discussed above, we can hope for order-of-magnitude accuracy at best."

Reviewer's comment No. 6 — Technical comment: Page 12, line 292: Figure 5 should be Fig. 3, correct?

Authors' reply: Many thanks to the reviewer for catching this issue, we have now corrected it on page 12.