

Geosci. Model Dev. Discuss., referee comment RC1
<https://doi.org/10.5194/gmd-2022-44-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2022-44

Anonymous Referee #1

Referee comment on "A machine learning methodology for the generation of a parameterization of the hydroxyl radical" by Daniel C. Anderson et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2022-44-RC1>, 2022

This paper describes the application of a gradient boosted regression tree machine learning approach to derive a parameterization for tropospheric OH based on CCM simulations. The approach is shown to reproduce simulated OH well under current conditions even for cases it has not been trained on, and it behaves acceptably, albeit with increasing errors, when applied to future conditions outside the standard training set. There is substantial novelty in the approach taken, and the results offer a degree of interpretability that is very interesting. The paper is generally well structured, clearly written, and appropriately illustrated. The authors have been thorough in evaluating their approach, and it is particularly good to see robust testing of input variable choice and hyperparameter value selection.

The main weakness of the paper is that the potential applications of the approach are not clearly identified. What does the parameterization add beyond the simulated climatology that was used to generate it? The chemical inputs used for the parameterization are dependent on OH, and hence a full model simulation is required to capture the feedbacks. In its present form, the parameterization is not sufficient to replace CCM chemistry, and can only reproduce OH from existing simulations. If a full CCM simulation is required to generate the inputs to the parameterization, then what does the parameterization add? If the approach is to be used as a simplified chemistry, like ECCOH, how will the oxidised inputs such as MHP and H₂O₂ be adjusted? A clear description of the application or purpose of the parameterization is needed, ideally with an example. Could the approach be applied with aircraft or satellite measurements to estimate OH concentrations? What is its value beyond simply reproducing existing simulations?

Overall, the methodological approaches developed here appear sound, and the parameterization has substantial potential, but practical application of the approach is not described. The paper is not suitable for publication in GMD until the authors have made the purpose and application clear to the reader.

Specific Comments:

Title: The paper describes a machine learning methodology, but does not convincingly demonstrate a tool to improve computational efficiency, as independent application of the approach is not described. The second half of the title should be dropped.

The approach is trained on an existing near present-day CCM simulation. Why was the approach not trained over a much wider range of conditions from preindustrial to possible future? A more thorough and complete set of training data would permit generation of a much more robust parameterization that was applicable to a wider range of conditions. Weaknesses in reproducing simulated future OH with the same CCM and chemistry demonstrate the frailty of the approach.

Table 1 needs to be presented more tidily to conform to journal standards. Note also that isoprene is included twice; the first occurrence should be ethane. Does UV albedo refer to UV surface albedo? Does cloud fraction apply to the column fraction above a given level or to the local fraction at that level? Optical depth above and below are separate inputs, while C4/C5 alkanes are combined; this is not clear from the Table, but is evident in Fig 5.

Line 232: What is the likely origin of the biases poleward of 30 degrees? Could this be related to averaging of cloud cover over the diurnal cycle, a factor lost when using daily-mean input variables?

The methane lifetime is matched well, but the small bias high is systematic throughout the year and indicates a consistent underestimate of mass- or methane-weighted OH (as also seen in Fig S5). Can the authors suggest why this arises?

Typos and Minor Issues

Lines 74-76: Sentence grammar needs revision (or remove "Though")

Line 174: "balance" would be clearer as "remainder" (or a similar word)

Lines 35, 219 and 436: "comports" is somewhat archaic; "accords with" would be clearer

Fig 3: Red percentage labels are difficult to read over color backgrounds. Please adjust the font color so that they are legible.

Fig 5: What do the colors represent?

Line 457: The citation for Shi et al. is missing the publication date (2018)

There is a Section 4.1 (with subsections) but no Section 4.2, so renumbering of sections is needed here.