Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-105-RC2, 2019

Geoscientific Model Development Discussions

Interactive comment on "tobac v1.0: towards a flexible framework for tracking and analysis of clouds in diverse datasets" *by* Max Heikenfeld et al.

Anonymous Referee #2

Received and published: 4 July 2019

1 Reviewer's summary of the manuscript

In "tobac v1.0: towards a flexible framework for tracking and analysis of clouds in diverse datasets," Heikenfeld et al. present a new, open source software framework for detecting and tracking clouds in datasets ranging from satellite observations to model output. They describe the details and justification for their Python-based implmentation, which extensively leverages existing libraries that offer a number of benefits (e.g., lazy data loading from dask), and they describe in detail the various algorithmic steps of tobac. They further provide two illustrating examples of how tobac might be used: tracking of deep convective clouds in 3D CRM output and tracking/comparison of clouds in

2D satellite data and 2D model output. The authors also discuss some limitations of tobac–especially restrictions on the appropriate temporal resolution of the input data, and they describe how the modular design of tobac will facilitate the authors (and/or future users) to easily improve/upgrade various algorithmic components.

2 Summary of Review

Overall, Heikenfeld et al. present a well-written and thorough description of a new, open-source framework that other geoscientists will likely find useful. The paper is well within the scope of GMD. The author's algorithmic choices are both well-explained and justified, and the methods by which the authors evaluate tobac are sound: particularly with respect to evaluation of the impact of temporal resolution.

I ordinarily have many paragraphs of feedback when reviewing papers, but in this case, the authors have put forth a solid manuscript; I do not have much feedback to offer, beyond a suggestions for clarifying in a few places. Based on this assessment, I recommend that the manuscript be accepted for publication in GMD pending some minor revisions. The authors should be commended on assembling an excellent manuscript and describing a useful open-source code.

3 General Comments

3.1 Some additional citations and discussion

In section 1, the authors give a reasonably thorough overview of other examples of cloud tracking in the literature. There are a few additional references that I would suggest the authors discuss:

- Wilcox, Eric M. “Spatial and Temporal Scales of Precipitating Tropical Cloud Systems in Satellite Imagery and the NCAR CCM3.” Journal of Climate 16, no. 22 (November 1, 2003): 3545–59. https://doi.org/10.1175/1520-0442(2003)016<3545:SATSOP>2.0.CO;2.

Note that the following papers focus on identifying clouds in time-slices, and not necessarily tracking them in time. But the concept is close enough–the papers are essentially steps 1–3 in the tobac workflow (Figure 1)–that these papers warrant discussion.

- Wilcox, Eric M, and V Ramanathan. “Scale Dependence of the Thermodynamic Forcing of Tropical Monsoon Clouds: Results from TRMM Observations.” Journal of Climate 14, no. 7 (April 1, 2001): 1511–24. https://doi.org/10.1175/1520-0442(2001)014<1511:SDOTTF>2.0.CO;2.
- Wood, Robert, and Paul R. Field. “The Distribution of Cloud Horizontal Sizes.” Journal of Climate 24, no. 18 (2011): 4800–4816. https://doi.org/10.1175/2011JCLI4056.1.
- O’Brien, Travis A, Fuyu Li, William D Collins, Sara A Rauscher, Todd D Ringler, Mark Taylor, Samson M Hagos, and L Ruby Leung. “Observed Scaling in Clouds and Precipitation and Scale Incognizance in Regional to Global Atmospheric Models.” Journal of Climate 26, no. 23 (December 2013): 9313–33. https://doi.org/10.1175/JCLI-D-13-00005.1.
- Igel, Matthew R., Aryeh J. Drager, and Susan C. van den Heever. “A CloudSat Cloud Object Partitioning Technique and Assessment and Integration of Deep Convective Anvil Sensitivities to Sea Surface Temperature.” Journal of Geophysical Research: Atmospheres 119, no. 17 (2014): 10515–35. https://doi.org/10.1002/2014JD021717.

- Guillaume, A., B. H. Kahn, Q. Yue, E. J. Fetzer, S. Wong, G. J. Manipon, H. Hua, and B. D. Wilson. “Horizontal and Vertical Scaling of Cloud Geometry Inferred from CloudSat Data.” Journal of the Atmospheric Sciences 75, no. 7 (July 2018): 2187–97. https://doi.org/10.1175/JAS-D-17-0111.1.

As a side note, I am a bit surprised that one of these references wasn’t already included, as it was written by one of the co-author’s former Ph.D. students.

In addition to these additional references, it might be useful to add a paragraph or two that describes what–in terms of science–was has been learned by using cloud tracking. For readers unfamiliar with cloud tracking, this might help justify the scientific motivation for a flexible and open-source software package for cloud tracking.

3.2 Feature ID vs Segmentation

Perhaps I’m being a bit dense about this, but I re-read both sections 2.2 and 2.3 several times and could not determine the functional difference between feature identification and segmentation.

Based on what I’m reading, it sounds like the only output of the feature identification step is the set of feature positions (weighted mean centers), from which the segmentation starts. Is this understanding correct? If so, this point should be emphasized, and if not, the text would benefit from a revision to make the distinction between feature identification and segmentation more clear.

In later parts of the manuscript, it seems that one functional difference might be that feature identification and segmentation might use different variables: e.g., max vertical velocity vs cloud condensate mixing ratio. Would the authors get the same result if they segmented the condensate field, based on any values above the segmentation threshold, and then filtered out objects with max vertical velocity below a certain threshold?

I'm not suggesting that this should be done, but rather I am illustrating that it might be useful to discuss the feature-ID/segemntation approach versus other plausible approaches; such a comparison might help readers grasp what appears (to me) to be a subtle distinction.

3.3 Style of package/software names

The manuscript contains a lot of references to software libraries, e.g.,: tobac, pandas, xarray, scipy, etc.. It is good, and useful, that the authors do this, but it might be useful to use a different text style for these software package names in order to visually distingush package names from English words. This is important to make the manuscript accessible to readers who might not be familiar with the Python ecosystem of packages, and it is especially important for software packages that could easily be confused for English words. For example, on pg 10, line 8, the sentence "The trajectories are recorded in a pandas DataFrames" might trip-up a non-Python-initiated reader (or at least amuse them). If the authors used LaTeX to compose the manuscript, and if GMD style rules allow it, I would strongly suggest that the authors use the `\verb:package-name:` macro.

3.4 Nice code

This is a compliment rather than constructive feedback: I appreciate the pervasive use of comments in the code and the clear and consistent documentation of functions. It was easy to skim through the code and get a general understanding of how the code is functioning.

4 Minor issues

p 6, line 25-26: The use of the term 'erosion' here is a bit unclear: the use of this word, in this sense, is not common in the geophysical sciences. The term 'erosion' should be explained here.

p 7, lines 6-16: this multithreshold approach is interesting. I don't think I've seen this in the literature before - do the authors have a reference for this, or is this an innovation of this study? It should be stated either way.
Also, it might be useful to state whether the framework is flexible enough to permit a single threshold value, which might be appropriate in cases where a person wants to track any contiguous feature with non-zero cloud condensate.

p10 lines 2-5: "The trajectories.." <– it isn't clear to me what this sentence means. It might need to be rephrased.

p 10, lines 12-13: "Instead, the algorithm. . . " it might be useful to produce an illustration of this. I'm finding that I'm unable to visualize how cloud trajectories from tobac would appear when there are cloud splits/mergers

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-105, 2019.