Review of gmd-2020-303 by Hristo G. Chipilski

Title: TempestExtremes v2.1: A Community Framework for Feature Detection, Tracking and Analysis in Large Datasets
Authors: Paul A. Ullrich, Colin M. Zarzycki, Elizabeth E. McClenny, Marielle C. Pinheiro, Alyssa M. Stansfield, and Kevin A. Reed

Suggested decision: Accept with Minor Revisions

General comments
TempestExtremes (TE) is a framework for the identification and tracking of features in Earth system datasets. The underlying paradigm behind TE relies on the construction of abstract functions (kernels) that can be called directly from the command line and controlled via a highly configurable set of user parameters. In this work, the authors extend the original version of TE by carefully documenting all newly added kernels. Using several examples based on societally important meteorological features, they also demonstrate how one can configure TE for specific Earth system applications by sequentially combining relevant algorithm kernels. The robustness of the enhanced TE package is evident in its successful application to different geophysical features and the agreement of the obtained results with past studies. Because the upgraded version of TE generalizes previous tracking methods, the presented work constitutes an important contribution to the Earth system community as a whole. In view of this scientific merit and the high clarity of presentation, I strongly recommend the publication of the manuscript in GMD after the authors address my fairly minor comments below.

Thank you very much for your positive feedback and suggestions. We have extensively revised the manuscript in response to the two sets of reviewer comments and believe the resulting manuscript is much approved. Our response to individual comments can be found below.

Specific comments

L31: If the authors wish to expand their list of areal feature tracking algorithms, they could give reference examples pertaining to convectively-generated outflow boundaries, such is in my 2018 model-based work (Chipilski et al. 2018) or precursor observation-based techniques, such as those of Uyeda and Zrnić (1986), Smith et al. (1989) and Delanoy and Troxel (1993).
The complete references to these papers are as follows:

Chipilski, H. G., X. Wang, and D. B. Parsons, 2018: Object-based algorithm for the identification and tracking of convective outflow boundaries in numerical models. *Mon. Wea. Rev.*, **146**, 4179–4200, https://doi.org/10.1175/MWR-D-18-0116.1.

Delanoy, R. L., and S. W. Troxel, 1993: Machine Intelligent Gust Front Detection. *Lincoln Lab. J.*, **6**, 187–212.

Smith, S., A. Witt, M. D. Eilts, L. G. Hermes, D. Klingle-Wilson, S. Olson, and J. P. Stanford, 1989: Gust Front Detection Algorithm for the Terminal Doppler Weather Radar Part I: Current Status. *Proc. 3rd Intl. Conf. on the Aviation Weather System*, Anaheim, CA, 31.

Uyeda, H., and D. S. Zrnić, 1986: Automatic Detection of Gust Fronts. *J. Atmos. Ocean.*

Technol., **3**, 36–50, https://doi.org/10.1175/1520-0426(1986)003<0036:ADOGF>2.0.CO;2.

Thank you for pointing out these additional references. We are confident there are many types of feature detection schemes we've missed in our brief review, but are happy to add mention of convective outflow boundaries and gust fronts. Thus we've included references to Chipilski et al. (2018) and Delanoy and Troxel (1993) in the paper.

L44: Please define the abbreviation CF in CF-compliant.

Fixed.

L53: My advice is that you to not restrict to climate datasets only as TE can be applied with equal success to other types of Earth system datasets, e.g. outputs from Numerical Weather Prediction (NWP) models.

Rephrased to:

To the best of the authors' knowledge, no other comprehensive toolkit exists for general nodal and areal feature tracking in Earth system datasets.

L62: "*except from DetectNodes and StichNodes*" – remove this as these kernels are described in Section 2.1.

Fixed.

L75: Here the authors mention that "*filtering of existing quantities*" is one of the capabilities present in NodeFileEditor. However, it is not immediately clear how the filtering in NodeFileEditor differs from the filtering operations in NodeFileFilter, so please add a brief clarification on this point.

The output of NodeFileEditor is a nodefile (text file containing trajectories). The output of NodeFileFilter is a NetCDF file containing filtered quantities. To be more specific, the opening of the "NodeFileEditor" section has been rephrased as:

NodeFileEditor is a new addition to TE for editing nodefiles (i.e., output from StitchNodes). It includes options for (1) appending new details to trajectories, such as radial wind profiles or accumulated cyclone energy, (2) removing certain columns from nodefiles, or (3) filtering trajectories or points along a trajectory, e.g., when outside of a specific time interval.

The opening of the "NodeFileFilter" section has been rephrased as:

NodeFileFilter encapsulates algorithms for masking spatial data using nodefile information, i.e., effectively converting nodefiles into binary raster masks at each time slice and (optionally) applying them to available data.

L123-124 (discussion relevant to Figure 1): Could you please clarify in the text whether the ID

of a merged object is equal to the smallest ID from the set of merging objects? In your example from Figure 1, objects 1 and 2 from time level 2 merge to a new object with an ID=1 at time level 3, i.e. 1=min{1,2}. Is this always the case?

Merged objects are relabeled so that the global ids are sequential and without gaps. So the resulting global ID could be completely unrelated to the original tag. To clarify, the following sentence has been added to the text:

Note that global ids start at 1 and are consecutive thereafter; they are assigned only after connected components of the graph are identified, and as such are unrelated to the blob id on each time slice.

L226-L227: Please elaborate on the meaning of "*This further provides an example of the ability of TE to evaluate functional relationships at run-time*" as it is not clear from the context what these functional relationships are and how you have defined them.

This sentence has been rephrased to:

It is also an example of TE's ability to evaluate functions of meteorological fields at run-time.

Figure 2: On L262-263, you state there is a high correlation between the algorithm-derived TC climatology and the observed TC tracks provided by IBTrACS. Is it possible to add the IBTrACS data as a subpanel in Figure 2 so that your readers confirm this conclusion? Ideally, I would like to see an algorithm-observation comparison similar to that shown in Figure 3.

Added an additional panel to Fig. 2 that shows IBTrACS pointwise trajectories (in addition to top panel, which shows the TempestExtremes tracked cyclones in ERA5).

L295: Briefly explain your choice of "*159 bins of width 0.125 degrees*" by either using a reference from the existing literature or a physically-based reasoning.

Thank you for pointing this out. The following explanation has been added to the text:

The number of bins and bin width were chosen based on the horizontal grid spacing of the ERA5 wind data, which is approximately 31 km. The bin width of 0.125° was chosen to adequately sample points at this grid spacing to create the radial wind profiles. The number of bins was chosen to ensure the radial averaging extended out far enough from the TC center points to capture the storms' complete wind circulations.

L393-L395: "*strong advection of warm, moist, equatorward air*" – avoid quantifying the strength of advection unless you decide to overlay the near-surface winds in Figure 5. Similarly, it is not possible to conclude "*that the heaviest ETC precipitation is associated with the warm conveyer belt*" in the absence of wind information.

To address this, we have done two things. One, we have improved the composite panels, which now include wind barbs (to highlight advection) and vertical velocity contours (to highlight the

WCB). We also plot IVT instead of the static low-level moisture field as this is more representative in showing poleward/upward advection of moisture on the eastern side (in the Northern Hemisphere) of the composite ETC.

In addition to the updated figure, the text as been modified to read:

"Figure 5 shows the composited precipitation rate field (PRECT), along with analogously calculated composites of 850 hPatemperature (T850) and integrated vapor transport (IVT). Total precipitation is largest near the storm center. Further, advection of warm, moist air wrapping cyclonically around the eastern side of the storm center is seen in the 850 hPa temperature field (composite wind vectors shown in black). Lastly, the collocation of high values of IVT and rising motion in the mid-troposphere(600 hPa omega contours shown in white) shows strong upward and poleward moisture advection associated with the warm conveyor belt, as previously shown in hand-compositing studies (e.g., Browning, 1986; Field and Wood, 2007)."

L425: Here you could reword your subsection as "*Step 2: Create AR mask with NodeFileFilter*" in order to establish a better connection with the following "*Step 3: Apply AR mark to VIWVN*".

Agreed. Something like this was also suggested by Reviewer 2 to make the paper read less like a technical document. As such all sections have been relabeled to emphasize the purpose rather than the operation.

Figure 7: Replace "*northward*" with "*poleward*" to reflect that IVT refers to either the Northern or Southern Hemispheres.

We believe in the top figures it is correct to use "northward", as values in the southern hemisphere are negative (i.e., poleward). In the bottom figures the term "northward" or "poleward" are interchangeable. The text is unchanged.

Technical corrections

L143: Remove "*a*" in "*followed by a several examples*".

Fixed.

L144: "*subsequent employ*" should be changed to "*subsequent utilization*".

Fixed.

L146-L148: Remove the sentence starting with "*In each of these composite algorithms …*" as you already mention this information earlier in your paragraph.

Fixed.

L209: Remove "*is*" in "*… our ERA5 data is comes …*".

Fixed.

L229: Avoid repeating "*output*"; e.g., you could replace the second "*output*" with "*written*".

Fixed.

L265: Please add the publication year to your Zarzycki et al. reference.

Fixed.

L545: It might be better to use "*integrating*" in lieu of "*developing*".

Fixed.