**Review for**

*Detection of upper-level troughs and ridges using deep learning – application in the Mediterranean*

**by**

*Ofir Ariel et al.*

---

**Summary:**

In their study, Ofir et al. introduce a new, machine-learning-based method to detect troughs and ridges. They explicitly compare their approach to existing algorithm, but also feed in their training data some expert-labelled trough and ridge axes. A special focus of the study is on the troughs and ridges of the Eastern Mediterranean, but the authors also show that their method can be applied on a global scale.

The strength of the study is on the detailed, systematic description of the machine learning approach. This, and the evaluation and (statistical, climatological) comparison against the existing identification methods, is the central part of the study, whereas the meteorological application is restricted to a minor part. This latter point is, in my opinion, also the major weakness of the study. In summary, for a generally WCD readership interested in trough and ridges as meteorological features, in machine learning to identify them, and in the climatologies and meteorological characteristics the balance between technical details and meteorology application might not be given. The method description is very detailed, and also very rich in ML terminology, whereas the meteorology/climatology could (and should) be expanded.

In summary, to make the work publishable in Weather and Climate Dynamics, I would (substantially) reduce the methodological, terminology-heavy part, and instead extend the meteorology part. It would also be highly appreciated if the authors more clearly show how the new method provides new, so far unknown insight into the lifecycles, characteristics of Eastern Mediterranean troughs and ridges. If the authors decide to keep the very detailed methodological part, while not extending the metorological part, I wonder whether WCD is the best journal to present the study.

Given all this, I only can recommend publication (with major revisions) if the major concerns and specific comments listed below are adequately be addressed.

**Major concerns**:

**1. Abundance of technical terminology:** The study introduces a new, machine-learning (ML) method to identify trough and ridge axes. The strength of the paper is certainly on applying state-of-the art ML approaches, but for a more general readership of Weather and Climate dynamics (WCD) it partly is too technical. This is reflected in a very detailed, ML-terminology-heavy description of the method in

section 3, followed by the results in section 4, which however also discuss rather methodological aspects (statistical measures, ablation study, global generalization, few-shot ridge transfer). If WCD were a methodology-focused scientific journal, that wouls all be fine; but, since WCD also addresses the weather and climate community, I think that a more meteorological and less technical presentation would be appreciated by the readership. I list here a few examples where I, as a more meteorology-focused reader, had to struggle:

- L169-170: “Soft curvature selection. The axis mask is the product of two sigmoids: one suppresses flat gradient regions, and the other selects curvature above a threshold for troughs or below its negative for ridges. The curvature and minimum gradient thresholds are learnable; a temperature parameter controls the sharpness of the selection” -> I kind of understand the idea, but do not really understand the details.

*- L181-189: This paragraph on the attention-based axis detector is very dense, full of ML-technical terminology, and might be to challenging for the general reader. It is nice that the steps are visualized in Figure 3, but possibly more explanations are needed.*

*- L207-210: Here, reference is made to ‘learning sigmoids’, which I did not immediately understand.*

*- L214: “A training procedure involving self-supervised pretraining with pseudo labels is evaluated as an ablation in Sect. 4.” -> difficult to understand if the reader is not very familiar with ML.*

*- L234: “using a marching-squares algorithm” -> I guess a reference would be sufficient here, but for a general reader it might not be clear how this algorithm works.*

*- L298: “All scores are computed out of fold” -> What does this mean?*

*- Section 4.4: “Ablation study” -> Again, the technical details could be a little overwhelming for a general readership; this is also reflected in Table 4 where reference is made to: Self-attention, Cross-attention Unified-channel attention Pretrain, Pretrain.... I have no doubt that these are all solid ML approaches, but they most likely go beyond what a general WCD reader interested in trough/ridge detection would understand.*

*- Section 4.7 (few-shot ridge transfer): Methodologically it is interesting whether the trough detector can be re-purposed into a ridge detector based on a small labelled set. But, again, I wonder whether this goes beyond what is of interest to the general readership of WCD? Further, I had some difficulties in readily understanding what exactly was done.*

*- I had difficulties in understanding the whole section 5.2 on the robustness of the architectural choices, and I assume that will also be the case for many reader for WCD that are not very familiar to ML methods.*

In summary, I invite the authors to carefully re-think what is really needed for the readership of WCD, and what could be omitted (or presented in a supplement) to make it more concise and clear.

**2. Meteorological application:** The strength of the study is on its methodology, but still it aims to be published in WCD (with a focus on weather and climate *dynamics*), and also the title refers to an application of these upper-level troughs and ridges in the (Eastern) Mediterranean (EM ). Therefore, as a reader, I wonder at the end what I have learned new about the troughs and ridges specifically in this region beyond a new method to identify them and to build a climatology. Further, the climatological analysis does not go very deep, but is restricted to: a relatively short section 4.5 (seasonal climatology; link to geopotential field, vague discussion of link with precipitation) and section 5.7 (discussion of climatology; link to subsidence and precipitation, some further methodological aspects/limitations). What I am missing is some characteristics of the troughs/ridges: How 'intense' are they? What are there orientation, lifetimes, lengths? How are they linked to fields of vertical wind and/or precipitation, low-level winds and /possibly) moisture fluxes? I think that the new data set and the focus on the EM would allow for more in-depth analysis of the *meteorology* associated with the trough/ridge features. Would it be, for example, to categorize the trough axis into different classes and to study there impact on surface weather? Possibly, it would also be worthwhile to consider composites of meteorological fields (vertical velocity, precipitation,...) centered on the identified trough/ridge axes and to derive from this some weather implications for the EM?

In summary: I think that the climatologies presented are already very interesting, but they miss an opportunity to get further insights in trough/ridge characteristics in the EM and how they relate to impact weather. Given the WCD weather-and-climate focused readership, I think that the meteorological part of the study has to be extended (most likely on behalf of the detailed methodological aspects).

**3. Ambiguity of expert trough/ridge-axis identification:** I think it is an excellent idea to train a trough/ridge identification on expert-labelled features. But, it could be discussed more critically that also these labels come with some ambiguity, and hence that the objective ML training might be 'mislead'. This is briefly discussed in section 5.5, but I would appreciate a somewhat more detailed assessment on this. Note also that there exist similar issues in identifying fronts and cyclones. Two interesting references in this respect are (both including manual, subjective tracking):

> *Uccellini, L. W., Corfidi, S. F., Junker, N. W., Kocin, P. J., & Olson, D. A. (1992). Report on the surface analysis workshop held at the National Meteorological Center 25–28 March 1991. Bulletin of the American Meteorological Society, 459-472.*
>
> *Hewson, T. D., & Neu, U. (2015). Cyclones, windstorms and the IMILAST project. Tellus A: Dynamic Meteorology and Oceanography, 67(s1). https://doi.org/10.3402/tellusa.v67.27128@zela20.2015.67.issue-s1*

I don’t expect the authors to do a systematic analysis of this ‘expert bias’ and how it influences the ML-based trough/ridge identification. But they could make the reader aware that these biases also exist in other feature-extracting approaches (fronts, cyclones,...) and they might speculate a little how the outcome is affected by it. Personally, I think that this would be a very interesting aspect.

**4. Systematic comparison to Schemm et al. (2020) and other climatologies:** From the methodology section, it is clear that Schemm et al. (2020) was an important ‘motivation’ for the current study. Methodological steps are leaning towards Schemm et al. (2020), and/or it is argued how this study performs better. Therefore, I wonder whether the climatologies of the current study should be more systematically be compared to Schemm et al. (2020) and the other methodologies. Of course, I see that performance metrics are compared in Table 2 (for troughs) and Table 3 (for ridges), and for single cases in Figure 6 and 7. It would, however, also be interesting to see and discuss how the climatologies in Figure 8 and Figure 9 compare to the other methods. Are the trough/ridge hotspots the same? If yes,

Further, the authors mention in Section 5.7 that tracking the trough/ridge features, in addition to characterizing them with some quantitative metrics (intensity, length, orientation) would allow for addressing very relevant research questions. For example: Where do these features form? How fast do they propagate. Tracking the troughs and ridges was indeed done in Schemm et al. (2020), so in this sense the latter method goes beyond the current one. I would appreciate if the authors more systematically compare their approach with the outcome of other ones, thus also strengthening the meteorological application of their approach (see point 2 above).

Note also that Schemm et al. (2020) considered other meteorological fields and/or characteristics of the trough and ridge axes, many of which would also be available in the current study. Hence, the authors could take Schemm et al. (2020) (or other methods) as ‘guidance’ how to deepen the meteorological analysis of the study.

A first step in comparing the different methods is presented in section 5.1, which is highly appreciated. Still, this comparison focus more on the methodological strengths and weaknesses of the methods. In addition, a meteorology-based comparison of the methods would be very interesting. Basically, I wonder what do we learn from the new trough/identification of this study compared to what was already known before – and this regarding climatology, characteristics, trends (e.g. decadal trends)?

**Specific/minor comments:**

- L10: Note every reader might be familiar what the F1 is referring to. I see that it refers to the F1 score which is often used in evaluating classification models, still, I wonder whether it should be introduced in one sentence?

- L33-34: Here, different approaches to identify trough axis are introduced. Whereas single-centre and vorticity-based methods are discussed in the following paragraphs,

this is not (in an obvious way) the case for geometry-based and continuous-axis methods.

- L62-63: “Notably, Bueh and Xie (2015) is the only study reviewed here that reports some kind of objective evaluation of their method against an independent test dataset” -> It is true that systematic and objective evaluations are missing. But, some steps in this direction where, e.g., taken by Schemm et al. (2020) by considering independent (not included in trough identification) meteorological fields that are expected to differ to the west and east of the trough axis. In their case, it was quasi-geostrophic vertical-wind forcing. Maybe, the authors could discuss in somewhat greater detail what such an objective evaluation should consider, e.g., possible meteorological fields?

- L85-86, & Figure 1: In which sense is the trough structure leading to the surface observations? To me it is not obvious how the weak trough in Figure 1a leads to the precipitation in Figure 1b? The precipitation mostly seems to be near coasts, i.e., where topographic and land/sea-circulations might be of importance. Further, the trough seems very weak to me, and the trough axis far to the west of the precipitation. Further, it would be interesting to know whether the trough’s influence is mainly due to directing the low-level flow, or whether it also associated with substantial vertical lifting. Does the trough axis separate regions with subsidence (to the west) from regions with ascent (to the east)? Finally, in Figure 1a it might be worthwhile to zoom out a little to see the full scale of the trough (also its extension to the west of the axis).

-L103: “explicitly grounding the learning process with atmospheric physic” -> I am not sure whether I really understand this statement! Explain further. Also, I wonder whether it should already at this point made clear in which way the new method goes beyond Schemm et al. (2020), as this method is often referred to.

- L119: Move the URL to the section on data availability.

- Figure 3: What does ‘QKV’ in panel (c) stand for?

- L160: Is the spherical geometry of the Earth (convergence of meridians towards the pole) taken into account in the smoothing algorithm, and more generally in the further algorithmic steps? As long as the analysis is applied in the EM, I assume that the spherical geometry does not affect the outcome, but this might change if troughs/ridges are identified in polar regions? More general: To which degree does the methodology rely on a regular latitude/longitude grid, or would it also work if the input data are represented on a different geographical projection (e.g., polar stereographic projection)?

- L176-180: Here, it is discussed that Z500 gradients are used as input, as well as the horizontal wind components U/V. At this point I wondered whether there could really be additional information in the wind field, given that geostrophic wind balance is nearly fulfilled at mid-tropospheric levels. I see that this point is discussed further below in the manuscript, but I still wonder whether it would be better to use only the *ageostrophic* wind components instead of the full ones. This might be a ‘sharper’ input field, thus creating clearer signals on how the wind field contributes to the

outcome of the classification. But I can also imagine that the ML algorithm somehow' extracts exactly these components? In any case: I think it is a very interesting point and I would appreciate an in-depth discussion on this aspect. How does the ageostrophic wind contributes to the trough/ridge identification? Does the algorithm realizes that ridges/troughs are super- and sub-geostrophic? Or is it because of curvature effects on the ageostrophic wind?

- L231 and 242: and others: Is 'blob' the best word to use? It's rather 'informal'. Maybe, use 'feature', 'structure', 'region',... instead of 'blob'.

- L294-296: I appreciate this plain-word description/summary why all the metrics before are needed. Possibly, it can even be the introduction to section 3.5, so that the reader knows from beginning what the metrics describe, before 'working through' their formal definition.

- L314: Please complete/add missing references.

- Figure 6: What are the colour shadings showing? This information is missing in the figure caption.

- Section 4.6: I see that the authors would like to make the point that the method also works on a global scale. However, the key topic of the study is troughs/ridges in the Mediterranean, and thus this very short subsection 4.6 distracts the reader from the main story. How about moving this part to an appendix, or a supplement?

- L403: At the beginning of this subsection it is stated: "The ablations help define a practical final configuration". This is fine, however, then as a reader I would expect at the end of the subsection to find the answer to this 'question'. So, what is the final, best configuration? Maybe, the answer is hidden in the text and I missed it?!

- Section 5.6: In this section the generalisability of the method is discussed. I wonder whether section 4.6. (application to global scale) would better fit here? In this way, section 4.6 does not distract from the main focus, which according to title and introduction is the Eastern Mediterranean.

- L459-460: "Others reflect something more fundamental: annotating synoptic features is inherently subjective and open to meteorological interpretation" -> As suggested in major concern 3, add (and discuss) references to other features where human annotation can lead to biases (fronts, cyclones,...).

- L482: What 'another axis type' could this be? Possibly, be more specific.

- L485-499: Is this part discussion, or would it fit better into an outlook section at the end of section 6?

- Section 5.7: Please insert paragraphs to guide the reader; at the moment the whole section 5.7 is one single paragraph.

- L533-535: "The natural continuation of the present study, and the focus of our future work, is therefore to track trough axes through time and characterize their properties over the EM, including their propagation speed, tilt, amplitude, and meridional extent." -> I fully agree that this would be a worthwhile extension of the

study. Please make the reader aware that a first step in this direction is already included in Schemm et al. (2020).

- L522-524: “For the same reason, polar lows that occasionally reach the northern boundary of the study domain, as well as the persistent high-pressure systems a lower latitudes, are generally not identified. These systems tend to be too broad, rounded, and zonally oriented to be classified as troughs or ridges.” > I agree that troughs and ridges become very broad at northern latitudes. But I wonder whether this is essentially an artefact due to the choice of geographical projection (equidistant latitude/longitude), and whether this problem would be solved if, e.g., a polar stereographic projection is used instead.

- L507-510: “ Consequently, although such systems occur frequently, they less often develop the type of trough identified by our detection method. By contrast, systems over the Mediterranean more frequently take the form of meridionally elongated troughs, which may explain the higher frequency of detected troughs in this region.” -> I am not sure whether I understand this point. In which sense do the EM troughs differ from the western European troughs? I was expecting the new method to detect also troughs that are less meridionally elongated. Further, it would be interesting to better understand why the European troughs behave differently in their rain-producing characteristic. Would it be, for example, be helpful to create trough-centered composites of meteorological fields (precipitation, vertical velocity, low-level flow) for European and EM troughs, and see if and how they differ. In short: I think it would be very interesting to investigate this different meteorological behaviour more systematically.

- Appendices: The manuscript now comes with nine appendices, which I think is too many. I am not sure whether all of them are needed, or whether some could be removed or possibly shifted to the supplementary material.