Review – Ambrogio Volonté

Objective identification of high-wind features within extratropical cyclones using a probabilistic random forest (RAMEFI). Part I: Method and illustrative case studies

This manuscript contains a very interesting study, illustrating an objective approach at the identification of strong wind features in extratropical cyclones. The manuscript is well-written and insightful and the results shown are indeed promising. Novel methods are used, sometimes beyond the typical expertise of WCD readers (and reviewers!). On this, I would recommend addressing the main comments made by the first reviewer. My concerns echo those three key points and therefore I'm not repeating them here. I add below some other comments, all generally minor. I would be happy to see this manuscript accepted for publication once all these points are addressed.

Comments:

Abstract:

- Line 9, "of spatial dependencies and gradients": this is quite vague and not really clear.

Introduction:

Line 25, "belong to": I would write "can produce some of" or something similar.

Line 34: I would write "or, in short, the warm jet (WJ). Same point applies to the CJ.

Characteristics of high-wind features:

- Line 92, "little or no precipitation": can you provide a reference for this statement? Warm conveyor belts are, as you write, airstreams ascending in the warm sector of the cyclone, and are associated with vigorous moist processes (condensation, …). Therefore, the reader would be surprised to hear that little or no precipitation is associated with them.

Line 109, "ground 850 hPa": missing word here, perhaps "below 850 hPa"?

Section 2.5: I find feature naming not totally consistent here. If dry intrusion and post-cold-frontal convection are both subsets of CS, why does Figure 1 display CS + pCFC in the cold sector? Shouldn't it be DI + pCFC (= CS) ?

Data:

Line 153: "wind speed at 10m" (same applies to wind direction)

Line 153: I'm being pedantic here, but could you replace "RR" with "R", given that it's precipitation amount and not rain rate? (or if it's actually rain rate, please state it)

Method:

Line 218, “Section 4”: this line is already in Section 4, so I guess this should be referring to a different section.

Line 219: I would argue that wind speeds are not “enhanced” by a strong pressure gradient, as this is what causes winds to be strong (at meso-synoptic scale) in the first place! In my view, factors enhancing wind speeds are those not accounted for by the (gradient-wind-) balanced flow, such as convective downdraughts for CFC and symmetric instability for SJ.

Table 2: could you include the height of max wind location? (and if it’s > 800 m, as I think Zugspitze is, add the max winds below 800m?)

Figure 2: could you make the colour progression more intuitive? (e.g., using green and dark green in consecutive years, instead of having red in between)

Lines 225-226: could you mention the features (after NF) according to the number of points, in decreasing order? I think it would make for an easier reading of the sentence.

Line 228: I’m not sure I understand the rationale of merging CJ and SJ points. I would understand doing this if you weren’t able to separate them, but you have just listed them separately, so I don’t get why you then decide to put them back together. Is this because you think RAMEFI wouldn’t be able to separate them? If so, state it explicitly.

Line 249-250: Yes, I think this is really important

Line 267: “Appendix A1”.

Line 286: “are provided in Appendix A2 (practical implementation) and B1 (mathematical formulation)”. (Do you reference to A3 anywhere? I couldn’t find where)

Illustrative case study:

Figure 3: I’m not sure I’m reading it correctly, as it seems to me that also locations where $\underline{v} < 0.8$ (i.e., outside that dotted contours) are included in panel d. Also, I can’t see any solid contours. Could you clarify?

Line 349: Where would a “hook-shaped structure of the winds” be considered in your algorithm?

Statistical evaluation:

Line 374: “in Appendix B2”.

Lines 366-367, “Further, […] or 1”. Could you rephrase this sentence? It is not very clear to me.

Discussion:

Figure 7 and all-pairs approach: Does the order of features matter? In other words, is (0 = CJ, 1 = CS) equal to (0 = CS, 1 = CJ). If that's not the case, the missing panels should be included and discussed.

Lines 476-478: Wouldn't high winds related to a strong synoptic-scale pressure gradient (which to me just means a deeper, more intense cyclone) still fall in the same categories (features) but with higher wind speed values than a shallower cyclone? Or are you implying that deeper cyclones have on average a different structure? Please clarify this.

Section 7.2: Still on the same point, but I think this is crucial for the understanding. I don't think the expression "the highest wind speeds are enhanced by a strong background pressure gradient" is physically correct (see my "Line 219" comment). I agree that with a deeper cyclone, more locations can record strong winds, but I would still assume that even if they're not directly related to CJ or WJ, they would still be located either in the warm or cold sector (and thus fall in the CS category in the latter case). Could it be that you need a WS category?

Lines 498-499: please explain the difference between front and convergence line here, to improve clarity.

Lines 539-540: One could argue that this shows that the set of surface parameters used to train the RF is broad enough to allow it to include both SJ and CJ in the "CJ" category. Are you ruling out that separating SJ from CJ in the observation would lead to RF correctly identifying them individually? If so, on what basis?

Sections 7.2 and 7.4: My understanding is that using normalised values (of pressure, theta, etc…) you can identify features even in cases departing from "your climatology" (i.e., the mean values in the 12 cases selected). However, in these sections, you show that "anomalous" cases can be a challenge for RAMEFI. Do you expect this issue to be partially or totally solved when RAMEFI is trained on more data? Is this explored in the Part II paper?

Conclusions:

Lines 574-575: What surface values do you think could best distinguish CJ from CS (apart from p)? Maybe spatial temperature/theta gradients could be useful? Otherwise, the most obvious to me would be cloud cover, but I'm not sure how many stations would have radiation/sunshine measurements.

Lines 588-589: This is indeed an encouraging result, but this sentence is probably too bold, given you've just looked at 12 cases in observations vs 1 reanalysis dataset. Maybe you could replace "should" with "could"?