Review of "Assessing the skill of high-impact weather forecast in southern South America: a study on Cut-off Lows" by Belén Choquehuanca et al. (2024)

This article analyses the predictability of cut-off lows during Autumn that frequent the south-west coast of South America. The authors analyse the forecast lead-time of cut-off low formation identified objectively in GEFS forecast data. Understanding the predictability of cut-off lows is important to understand, particularly in this part of the world, where they have high impact, as the authors suggest. There are however some gaps in how various analyses and processes are discussed in this work that need to be improved before publication.

Some general, specific and technical comments are below.

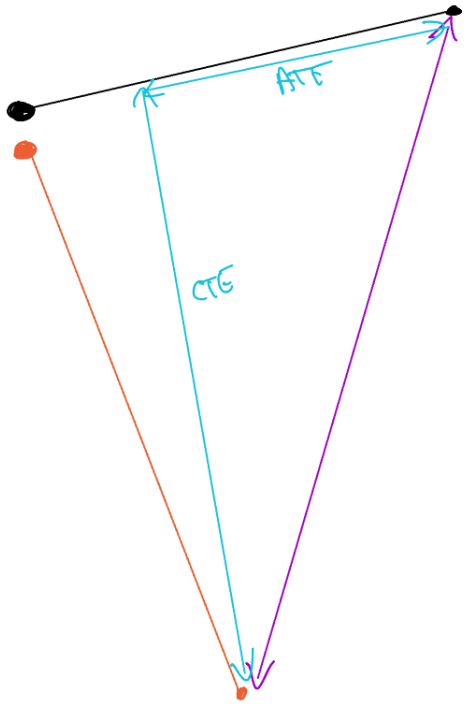
General comments

1. Methods non-specific (Section 2.3)
   The methodologies used within this study need to be clarified. I can understand that they are based on methodologies used and referenced in previous work, however at least some minor detail needs to be added to contextualise the results for the reader. This is especially true in Section 2.3. For example, "certain restrictions" - what are they?; "winds on the polar side of the labelled COL" - how far poleward?. These details are important for the reader to understand in order to understand the result presented.
2. Verification metrics used
   I have concerns that some of the verification metrics used may not be reflective of the processes they are trying to verify. This is particularly true of the CTE and ATE metrics. I am not sure that the ATE will consistently show an error in the speed of the track and CTE an error in the positional error of the track. An example situation of this is described in the specific comments below. Would showing speed and bearing not be more appropriate?
3. Sample size
   The sample of COLs used in this work seems incredibly small given the number of COLs that this region receives - this represents less than 1 COL per year. I wonder therefore if the region or criteria is too strong in this study. Relaxing it and getting more events would help significantly in making sure the results are more robust. I was also unclear exactly how COLs in the study region are selected - is it only COLs that begin (as Stage 2) in the region that are selected? Or are COLs that move through the region as Stage 2 also counted?
4. Some conclusions reached seem to lack evidence
   The authors make several conclusions related to the precipitation patterns, depth or vertical coupling and thermodynamics of the COLs related to errors in the forecast of these systems. Many of these are expressed as important results in both the conclusion and the abstract, however, in my view, these largely have very little evidence to back them up. These are largely based on two individual case studies of COL simulations by analysis of the ensemble mean. Can holistic conclusions be drawn for just these two cases? Statements such as "predictive skill of COLs had a significant impact on the precipitation forecasts downstream" or "underestimating the strength of the cold core of COLs can significantly alter the vertical coupling of COLs" need to be robustly backed up by the results. One way to make this more robust

would be analyse each of the members and leverage the ensemble spread to understand the processes associated with small and large positional errors.

Specific comments

- L70: The precipitation patterns studied seem very local and not “downstream”. In my view therefore, this posed question is not answered as stated.
- L88: Previous authors (e.g. Reboita et al. (2010)) have found large differences in frequencies of COLs for different reanalyses, with ERA reanalyses producing more COLs than NCEP reanalyses. Since the authors use GEFS, would the use of NCEP reanalysis not be a fairer comparison?
- L90-93: Does the orography affect your choice of using 850hPa in your thickness criteria? Presumably are large proportion of the Andes is generally above 850hPa.
- L94: I assume the “accumulated precipitation” is ERA5 data. General I have found that these totals are generally error prone. Is ERA5 precipitation reliable?
- L108: What are these restrictions?
- L109: I assume you look at a minimum in thickness? Provide details on this procedure
- L111: How far poleward?
- L113: What are these specific criteria?
- L117-120: Not sure I understand this point. Consider clarifying.
- L128-132: How robust is the use of ATE to represent speed and CTE to represent spatial error? I can envisage situations in which the along track error vastly under-estimates the speed of propagation (see my poor recreation of your Fig 2. below). In this case, the speed error is presumably under-estimated if one looks at the ATE, however in reality the speed is actually faster than in reality. Would using speed/distance travelled (to determine speed errors) and bearing (to represent bias in left/right not be better?



- L139: How is intensity defined in this study?
- L139-141: “7 days before” - Do you refer to forecast lead-times here? Clarify this statement.
- L149: 34 COLs - seems a small number (~1 COL per season). Would increasing your domain size give you more COLs and increase the robustness of your results? This number is too small to get a robust result in my view
- Figs 6-9: I strongly suggest that these figures are combined into a single plot to help the reader compare these different metrics. Either stack each metric on an init axis,or stack inits on a metric axis.
- L178-179: Do all your tracks last at least 3 days in order to study the track evolution after 3 days? Most climatologies (some of which are cited here) shows the majority of COLs last for only 1-2 days. Please clarify.
- Section 3.2: Forecast COLs are generally weaker than observed. How does this impact their predictability? Do they generally terminate more quickly than observed as they are weaker?
- L192: “except for init 2” - init 0 also fairly symmetric
- L217-220 and 240-243: This synoptic description is unnecessary. This circulation is there by definition of your COL identification technique.
- L226-228: It is unclear from your analysis how the vertical coupling was affected by the COLs incorrect forecast position. Be specific and add detail to this analysis
- Figures 10-11: Both Sections 3.3.1 and 3.3.2 make reference to low-level circulation patterns, however these are not shown.
- L231-233: I disagree from Figure 10, that one could say that GEFS had “difficulties” in predicting the split jet and “failed” to capture the cold-core. Both these features are present in Figure 10j,k,l
- L235: Do you feel that this is a robust statement? In general GEFS seems to have done a decent job. Your longest lead-time produced a weaker COL in the mean which affected the forecast of higher precipitation. However I feel that the statement that GEFS (which is relatively course resolution here) may not perform well for COL precipitation forecasts is very harsh given the data provided and the severely limited cases looked at.
- L244-246: Is this frontal zone not linked at all to the COL? Ie. Is this front not connected to the surface low (if there is one) associated with the COL or is it linked to a completely separate mid-latitude cyclone? Be specific.
- L251: “the circulation at low and mid-levels” - how do we know this? This implies that the low and mid-level circulation are completely independent of the upper-levels, which they are are not of course. Clarify and add detail to explain your analysis.
- L257-259: Similarly to case 1, I think the rainfall patterns presented look pretty good, especially considering that the model is 1x1 degree. What is the effect of the resolution on your analysis?
- L262: The final row of Fig 11 appears not to be cut-off, but called a COL here. Would this have been detected as a COL in your analysis?

Technical comments

- L32: “suppose” -> “pose”?
- L82-83: is data from the second week used here? It seems that only lead-times up to 7 days are investigated. If not, remove this detail.
- L137: “position is biased fast (slow)” - consider rewording

- L168: “Conversely,...” - this sentence is not converse to the previous statements. Both indicate large spread, the one is simply larger than the other.
- L172: “direction” -> “bias”?
- Figure 10-11: Maybe “Obs” is not a representative label, but rather “Reanalysis”.
- L176 and elsewhere: You use the terminology “segregation”, “onset” and “start” throughout the article. It remains unclear whether these are different or mean the same thing. If there are the same thing, stick to one terminology. If not, define them explicitly and clarify.
- L182: “ERA” -> “ERA5”?
- L187-202: “increases” - do you mean “decreases”? The value gets less. The magnitude of the error increases. Be specific.
- L221: “record-breaking rainfall”: 25-50mm seems low for a rainfall record. If this is the record, then add a reference to backup this statement.
- L243: “northeast of the country” - which country?
- L238: “the previous COL” -> “Case 1”
- L243-244: “Figure 11c” - this is thickness and geopotential height at 300hPa and not low-level circulation as specified.
- L282: “have a prominent” -> “tend to have a” - the bias is not that prominent is it?