

This paper systematically analyzes the sensitivity of forecast skill for the 2025 extratropical cyclone Storm Éowyn to model grid resolution, latent heating, and forecast initialisation time, mostly using ICON simulations but with IFS ensemble forecasts used as a secondary point of comparison. The authors' chief findings are that (1) increasing the model resolution yields a more accurate representation of the cyclone's intensity and, to a lesser extent, its track; (2) reducing latent heating suppresses the forecast cyclone's intensity and shifts its track northward; and (3) forecasts at all model resolutions are generally less accurate at higher lead times, with a few significant outliers from this trend. These results are mostly consistent with earlier findings in the predictability literature, though the study's systematic methodology adds a fresh and valuable perspective to those findings. Result (2) is particularly interesting in the context of previous studies (e.g. Coronel et al. 2015 and Tamarin & Kaspi 2017, which the authors cite) showing that latent heating has the opposite effect on cyclone track displacement, albeit in a climatological sense using idealized models.

Overall, this is a readable and well-written paper, but it is quite long, and I would recommend some abridgements be made. As to the paper's scientific findings, I have some concerns about the paper's interpretation of the latent heating sensitivity experiments, specifically LH's effects on the cyclone track. Furthermore, the paper's discussion of initialisation-time sensitivity should undergo some substantial revisions; in its current state, it reads as overly descriptive and doesn't develop physical understanding to the extent that other sections do. Once these concerns are addressed, though, I believe this paper will be a good fit for publication in WCD.

1 Major concerns

1.1 Length/superfluous details

Three sections of the paper stood out to me as being unnecessarily verbose:

- Section 3 of the paper provides a thorough review of Storm Éowyn's meteorological history. This review is well-written and interesting in its own right, but some of the discussion is extraneous to the rest of the paper and may be excised. For instance, the paragraph discussing the storm's status as a diabatic Rossby wave (lines 180–92) feels parenthetical, and I'm not convinced it is entirely necessary. Similarly, Figures 3 and 5 are very information-dense and partially redundant with each other and with later figures in the text: e.g., both Figures 3 and 5 contain PV maps for 23 January at 00 UTC and illustrate similar meteorological features to those in figures 7 (vertical structure of the PV tower) and 11 (ridge building as the cyclone intensifies).

I would consider shortening this section to emphasize the meteorological features that are important for the later discussion of forecast sensitivity, and to highlight that importance explicitly. This might also allow you to make later sections more concise. I'd also consider eliminating Figure 3, since this feels the least significant to the later sections, and incorporating any especially important information therein into a revised Figure 5.

- Section 5.4 can also be cut and its contents redistributed to other sections. I do like the central figure (Figure 13), but I'm not convinced it deserves an entire section to itself; its

interpretation seems like a fairly straightforward consequence of the shifts in the cyclone track and intensity discussed previously. Moreover, the framing of this section strikes me as an empty hypothetical—if atmospheric latent heating suddenly turned off, a lot more would change than this specific storm! (This may be more a matter of taste though.)

- Much of the text of Section 6 reads as a “blow-by-blow” description of Figures 14 and 15, and this section would be strengthened if the writing focused more on *analyzing* the data in these figures (see later for more comments about this section).

There are a few other passages scattered throughout the study which feel superfluous, e.g.: “Although this is not the focus...indicating high confidence” (lines 231–3); “Interestingly, the stochastic perturbations...wind gusts” (lines 503–4). Additionally, the abstract should probably be shortened by a sentence or two—I have no specific requests for what to cut here, and defer to the authors’ judgment about which points they’d like to emphasize.

1.2 Track sensitivity to latent heating (section 5)

One of the central claims of this section of the paper is that reducing the diabatic forcings from latent heating yields a northwestward shift of the cyclone track (line 598), which contradicts idealized GCM studies on latent heating and cyclone propagation. However, I’m not sure exactly how robust these results are from the data that are shown. In Figure 10(b), for instance, it’s true that the no-LH cyclones lie poleward of the control-run cyclones at 23 Jan and 24 Jan, but it seems that by 25 Jan there’s very little systematic difference between the latitudes of the two sets of cyclones? Since the no-LH tracks are biased poleward at *initialisation* in this figure, this would imply that the control-run cyclones actually propagate northward (and eastward) faster to “catch up” to the no-LH ones, consistent with the earlier idealized studies (and that the bigger question is why the no-LH cyclones are more poleward at initialisation in the first place). These observations are just from inspection of the figure; if the enhanced northward propagation of the no-LH cyclones is actually (statistically) significant, it would be helpful to show this more systematically. There are similar ambiguities in Figure 11—for instance, in subfigure (e), it seems like the control and no-LH cyclones are at the same latitude on 23 Jan 12 UTC.

On a similar note, the Sutcliffe–Trenberth forcing analysis in section 5.3 offers additional support for the paper’s claim, and the northward shift in the low-level mean thermal wind in the no-LH experiment is clear in figure 12, but it would be interesting to see how this compares to the actual propagation velocity of the cyclone in *all* the experiments shown in Figure 10.

1.3 Initialisation time analysis (section 6)

The sensitivity of the forecast skill of this storm to initialisation time is a rich and complex subject for inquiry, but as currently written, Section 6 mainly seems to *describe* the results and emphasize conclusions which are already common knowledge. For instance: “A correlation exists between the SLP*_{min} deviation (ICON vs. ERA5) and the initialisation time, demonstrating that later initialisations yield a more accurate representation of cyclone intensity across various horizontal

resolutions’’ (lines 613–5). This is just the general nature of forecast error growth, as formally demonstrated by, e.g., Lorenz (*Tellus*, 1982), and for the Lagrangian statistics of extratropical cyclones by Froude et al. (*Mon. Wea. Rev.*, 2007). Similarly, that increased horizontal resolution improves the skill of short-range forecasts is well-attested in the literature (Buizza, *Quart. J. Roy. Meteor. Soc.*, 2010; Davis et al., *Wea. Forecasting* 2010; Lei & Whitaker, *JAMES* 2017).

The discussion in section 6 highlights the non-monotonicity of this correlation, with anomalously poor forecasts occurring on 18 January and 20 January (18 UTC) for instance. On the origins of this monotonicity, the paper says that ‘‘[t]he generally poor performance at these initialisation times is suggested to be attributed to data assimilation errors that yield incorrect initial conditions’’ (lines 564–5). Can you demonstrate this more concretely, for example by considering the nature of the error growth in these particular forecasts? Being more specific on this point would help strengthen this section considerably, I think. (Though, given the complexity of the subject, this may turn out to be an analysis for another paper!)

There are a number of similar passages in this section that have a similarly speculative tone that I would rather see supported by more rigorous data analysis or reasoning: ‘‘The stronger wind gusts...model resolution’’ (lines 501–2); ‘‘This suggests that at coarser model resolutions...prior to cyclogenesis’’ (542–4), etc.

On a similar note, there are some interesting properties of the data which I feel are under-discussed in the text. For instance, why did so many IFS ensemble members fail to form a cyclone on 18 January? Why do the ICON simulations systematically underestimate the cyclone pressure anomaly relative to ERA5/IFS from 19 Jan on (Figure 14, left panel)—is this a resolution effect? Also, as shown in Figure D1, the position biases seem to be of similar direction across model resolutions and shift somewhat coherently with lead time (e.g. from a mostly southern bias on 17 Jan to a north-northwestern bias on 20 Jan). Any thoughts as to why this is the case? Can we learn anything from this about biases in the model or the data assimilation? I bring these points up merely as additional examples of observations where I think this paper could develop some interesting physical insights.

2 Minor/technical comments

Some of these are small technical suggestions, others points where I feel like the text can be clarified somewhat.

- All maps in figures, e.g. Figure 5(a)–(c) and Figure 11, should have some kind of latitude-longitude labels on them—the continental outlines are not easy to see amid the other details shown.
- Section 2.4: the phrase ‘‘start date’’ (line 149) is ambiguous—I assume you mean the start date for the algorithm (i.e. the time of maximum cyclone intensity), not the cyclogenesis date, so I would say that explicitly.
- Figure 4: How is the ‘‘maximum 24-hour deepening rate’’ defined? Why is it always positive, even when the cyclone is weakening?

- Section 4.2: Regarding jumps in the storm track early in Éowyn’s life cycle, the paper remarks that “The 50 IFS ensemble forecasts help explain this behavior, as they split into two distinct track branches during this stage” (lines 252–3). This isn’t really an explanation, though, merely the IFS and ICON tracks showing similar behavior?
- Figure 6: In Section 4.2, the paper invokes this figure to support the claim that “[a]s the cyclone evolves further, higher horizontal resolution becomes crucial for capturing the landfall location” (lines 273–4). This isn’t obvious from the figure, though, since all 5 forecast tracks shown in Figures 6(b) and 6(c) fall within the spread of the IFS ensemble tracks, and their accuracy doesn’t vary monotonically with resolution. Wouldn’t Figure D1 provide stronger support for this claim, since it shows that the higher resolution forecasts are systematically more accurate across a range of lead times?
- In Figure 8 and the discussion of it in section 4.3, why perform the correlation calculation using all the data shown? It seems clear from the figure that the linear correlation between low-level PVU and high-wind coverage does not hold early in the cyclone’s life cycle. I’d also wonder whether the trend that “[o]verall, higher resolution yields stronger correlations” (line 304) if the early-life-cycle points are removed.
- When you write, in section 5.1, “This corresponds to a relative contribution of LH to the total cyclone deepening in the control of 27%” (lines 319-20), are you using the DDC metric defined in section 5.2?

I also noticed the occasional small grammatical error in the paper, e.g. “The interested reader might has spotted...” (line 699, should be *might have*). These errors didn’t substantially affect readability but should be double-checked for in review.