

# Review of the manuscript “Multi-model high-resolution analysis of Tropical-Like Cyclone

## Daniel with WRF and ICON: peculiarities and sensitivity to convection schemes”

by Serafini et al.

Recommendation: Major Revisions

### Summary

This manuscript presents a side-by-side, high-resolution (~2 km) intercomparison of WRF and ICON for Medicane Daniel (September 2023), with a focus on sensitivity to the convection scheme (explicit, shallow-only, grayzone, and fully parameterized). The authors take care to match the two models as closely as possible (domain, vertical discretization, initial/boundary conditions, resolution, timestep) and introduce two appealing diagnostics: a Python algorithm to produce comparable vertical level distributions across the two coordinate systems, and the "Temporal Annular Symmetric Mean" (TASM) for characterizing the time-mean, axisymmetric warm-core structure.

The topic is timely and relevant, Daniel is an important and well-documented case, and the matched-configuration philosophy is a genuine strength. The TASM and the level-matching algorithm are worthwhile methodological contributions. However, before the paper can be accepted, several issues regarding the verification reference, internal numerical consistency, the fairness of the "convection-scheme" attribution, and the framing of the resolution regime need to be addressed. My main concerns are detailed below.

### General comments

#### 1. Reconcile track errors with Figure 4(c)

Section 4.1 reports track errors that are substantially smaller than those shown in Figure 4(c). For example, the text states that WRF–EXP has a mean error of 33 km and an RMSE of 39 km, and that ICON–CU has a mean error of 55 km and an RMSE of 65 km. In contrast, Figure 4(c) appears to show much larger values for the same simulations, including WRF–EXP with  $\mu \approx 76$  km and  $\sigma \approx 48$  km, and ICON–CU with  $\mu \approx 113$  km. These values are not mutually consistent as currently presented. Since the ranking of the seven simulations by track skill is central to the paper’s conclusions, the authors should clarify whether the text and Figure 4(c) refer to different metrics, time windows, smoothing procedures, or samples. If not, the relevant numbers should be corrected, and a single internally consistent set of track-error statistics should be used throughout the manuscript.

## **2. Strengthen the observational reference for track and intensity**

The observed track and the observed CSLP/wind time series rely heavily on Hérincs (2023), which appears to be a non-peer-reviewed report rather than an operational best-track product or a peer-reviewed observational dataset. Since the model track RMSEs, CSLP evolution and maximum-wind conclusions are evaluated against this reference, the observational basis needs to be strengthened. I recommend that the authors: (a) cross-validate the observed track and intensity against independent sources, such as ERA5, ECMWF operational analyses, ASCAT center/wind estimates, available station/ship/buoy data, and the diagnostics discussed in Flaounas et al. (2025); and (b) explicitly justify the adequacy of the manually selected SEVIRI-HRV cyclone center as the primary reference track. The manuscript states that this track has an uncertainty of about 15 km at 3-hourly resolution. Given that some reported model differences and landfall-position errors are only a few tens of km, the observational uncertainty may overlap with the inter-model differences used to rank the simulations. The authors should therefore include an observational uncertainty estimate, or at least discuss how this uncertainty affects the robustness of the model ranking.

## **3. Assess statistical significance of inter-configuration differences**

The manuscript ranks the seven simulations partly on the basis of track-error statistics, but the reported spread appears large relative to the differences between configurations. For example, the text reports mean errors of roughly 33–55 km, while Figure 4(c) appears to show larger  $\mu$  values and substantial  $\sigma$  values. In either case, the separation between several configurations is small compared with the variability of the track error along the cyclone life cycle. It is therefore not clear whether the apparent differences between individual configurations, or between WRF and ICON as model families, are statistically or practically distinguishable.

Please add a robustness assessment of the track-skill ranking (statistical significance test?). At minimum, the authors should explicitly discuss whether the differences among configurations are robust enough to support statements such as “WRF performs slightly better than ICON” and “explicit and shallow cumulus schemes generally yield better results than fully parameterized schemes.”

## **4. The "convection-scheme sensitivity" attribution is confounded by unmatched physics**

The manuscript aims to quantify sensitivity to convective treatment and to compare WRF and ICON under configurations that are “as comparable as possible”. This is a valuable objective. However, the cross-model interpretation should be more cautious. While convection treatment is varied within each model, other physical parameterizations are not matched between WRF and ICON. In particular, Table 1 shows different microphysics schemes — WRF WDM6 versus ICON Seifert–Beheng two-moment — and different boundary-layer/turbulence

treatments — WRF YSU + Smagorinsky versus ICON prognostic TKE. These differences are not secondary for a medicane-like system: microphysics–convection interactions, boundary-layer moisture transport, turbulent mixing and surface fluxes can directly affect intensity, precipitation distribution, latent heating, vertical warm-core structure and track evolution.

Therefore, systematic results such as WRF producing larger accumulated precipitation, deeper CSLP/stronger 10 m winds, or a deeper/more vertically coherent warm core than ICON cannot be attributed uniquely to convection-scheme treatment. They may partly reflect the combined model-physics package. I am not asking for a full additional matrix of microphysics/PBL experiments, but the limitation should be made explicit. The authors should separate conclusions about within-model convection sensitivity from conclusions about WRF–ICON differences, discuss the likely influence of microphysics and PBL/turbulence choices on the headline results, and temper statements that imply that the model-family differences are primarily controlled by convection parameterization alone. (I am not asking for a full additional experiment matrix, but the limitation must be made explicit and its likely influence on the results discussed. I have seen myself huge differences in precipitation fields and tracks in medicanes, among different MP physics.)

#### **5. Clarify and justify the "grayzone" framing for ~2 km**

The manuscript uses ~2 km grid spacing and includes an ICON–GZ experiment described as a "grayzone" configuration. However, the terminology needs to be clarified. In much of the current convection-permitting modelling literature, horizontal grid spacings around 2 km are usually treated as convection-permitting, while the deep-convection gray zone is more commonly associated with somewhat coarser grid spacings, where deep convective updrafts are only partly resolved. At the same time, 2 km is still not convection-resolving in a strict sense, especially for shallow convection, turbulent transport and individual convective updraft dynamics. Therefore, the manuscript should define explicitly what is meant by "gray zone" in this study.

This clarification is particularly important because the selected WRF cumulus scheme, KSAS, is itself a scale-aware mass-flux scheme developed for gray-zone applications. Thus, the contrast between "explicit", "parameterized" and "grayzone" configurations is not as clean as the current wording suggests. Please distinguish clearly between: (i) the grid-spacing regime of the simulations; (ii) the physical gray zone of partially resolved convection; and (iii) the specific ICON grayzone tuning used in ICON–GZ. The operational recommendations should then be revised so that they follow consistently from this definition, without implying that all 2 km simulations fall into the same gray-zone category or that the explicit/parameterized distinction is unambiguous at this resolution.

#### **6. Single-case scope vs. generalized conclusions**

The manuscript acknowledges that the analysis is based on a single event, but several conclusions are phrased more generally than the experimental design can support. Statements such as “the inclusion of a shallow convection parameterization proved to be important for both models” or that WRF may be “better tuned for the diabatic processes dominating TLC maintenance” should either be softened to apply specifically to Daniel under the present configuration, or more explicitly contextualized within the broader medicane sensitivity literature.

This is particularly important because the Introduction itself notes that there is no consensus on which model configuration performs best for TLCs, and that previous studies have shown strong sensitivity to microphysics, PBL schemes, initialization time, forcing dataset and cumulus parameterization. I therefore recommend that the authors revise the conclusions to distinguish clearly between: (i) findings that are robust within this Daniel case study; (ii) findings that are consistent with previous medicane studies; and (iii) hypotheses that remain to be tested across additional cases or initialization times. In particular, the authors should discuss where Daniel agrees with, or departs from, prior results in Ricchi et al. (2017, 2019), Pytharoulis et al. (2018), Miglietta et al. (2015), and Saraceni et al. (2023). Without this contextualization, operational recommendations about shallow convection or model-family performance should be phrased as case-specific rather than general guidance.

### **Specific Comments**

1. Section 3.5.1 applies Hart’s original Cyclone Phase Space methodology, using a circular radius of  $R = 300$  km around the cyclone centre. Please justify why the original Hart CPS framework was selected for Daniel rather than a medicane-specific or medicane-adapted approach, such as that discussed in Miglietta et al. (2025), which is already cited in the manuscript. This justification is particularly important because Daniel is treated as a compact tropical-like cyclone, whereas the original CPS framework was developed for larger synoptic-scale systems. The authors should clarify whether the chosen CPS setup is appropriate for identifying the warm-core and symmetry characteristics of a compact medicane, and explain how the interpretation of the CPS diagrams is affected by using the original Hart framework rather than a medicane-specific adaptation.
2. TASM axisymmetry assumption (Sect. 3.5.2). The TASM is a useful idea, but the manuscript itself notes that TLC symmetry is poorly defined. Please (a) quantify the degree of asymmetry being averaged over (e.g., azimuthal variance), and (b) discuss how the 15 km center-position uncertainty propagates into the annular means.

3. FSS description and interpretation (Sect. 3.4 / 4.2). The verbal description of FSS behavior around lines 330–333 is imprecise and at one point conflates "more cells than the observations" with low FSS in a way that does not match the symmetric nature of the score. Please restate the FSS interpretation precisely and add a "useful scale" criterion (e.g., the  $FSS \approx 0.5 + f_0/2$  target of Roberts and Lean, 2008), which would make the skill claims more rigorous and easier to compare across thresholds.
4. Re-initialization experiment (lines 114–118). The three 72-h re-initialized runs and their degradation (attributed to hydrometeor imbalance and energy-budget) are interesting but presented without any figure or quantification. Please either show supporting evidence or present this as a brief methodological note without the strong claim.
5. Observed CSLP reliability vs. CSLP rankings (lines 294–300). The discussion of the 996 hPa observed minimum possibly being an overestimate is welcome and honest. However, the CSLP skill rankings are then made against this admittedly unreliable value. Please add a corresponding caveat to the CSLP comparisons, or use a reanalysis-based intensity estimate as a secondary reference.
6. Please specify which IMERG run (Early/Late/Final) and version were used. This affects the high-percentile (99th, 99.9th) FSS results that several key precipitation conclusions depend on.
7. Radius inconsistency for wind extraction. Section 3.4 states the maximum wind is extracted within 150 km of the center, whereas the Figure 5 caption states 100 km. Please reconcile.
8. Title of Paragraph 4.3 is "Physics," which does not accurately describe its content (an analysis of the cyclone's tropical characteristics and warm-core structure via CPS and TASM, rather than the model physics/parameterizations covered in Sect. 3.1). Consider retitling, e.g., "Tropical-like structure" or "Physical structure of the cyclone," for consistency with Sect. 3.5.
9. The manuscript describes a "dramatic lack of ground-based observations" (Sect. 3.2). While accurate for the offshore Ionian and Libyan phases, this overstates the situation over Greece during the cyclogenesis/extratropical phase, where the dense National Observatory of Athens automatic network (NOANN/METEO) (Lagouvardos et al., 2017) provides high-resolution surface and rainfall observations — data used by several Daniel studies the authors already cite. Please consider incorporating NOANN gauge data (at least for the Greek extratropical phase) to provide independent ground truth for the precipitation verification, which currently relies on satellite-derived

IMERG even though the cited literature documents IMERG biases over Thessaly for this event. At minimum, the "lack of observations" statement should be qualified by phase and region. In Line 74, the precipitation sums mentioned, come from the NOANN network. Please use either the Lagouvardos et al. 2017 citation or Flaounas et al., 2025 one.

10. Figure 4: panels (a) and (b) rely on an eight-way color code that is defined only in the caption, with no in-figure legend. Several of the assigned colors are difficult to distinguish, making the overlaid tracks (a) and stacked-bar timeseries (b) hard to interpret. Please add an explicit legend to the figure and, if possible, adopt a more clearly separable color palette. Ensure color assignments are consistent across panels (a)–(c) and the corresponding text.
11. Line 248: "The cyclone structure is typically analysed using vertical cross-sections of  $\theta_e$ ." The word "typically" asserts a methodological convention without support. Since this claim motivates the introduction of the TASM, please provide citations establishing that vertical  $\theta_e$  cross-sections are the conventional diagnostic for (tropical-like) cyclone warm-core structure, or rephrase to avoid the unsupported generalization.
12. The manuscript repeatedly invokes operational use to justify model and resolution choices "widespread use in... operational forecasting" (line 45) and "operationally this is approximately the resolution used by both models" (line 108) but never specifies which centers run WRF and ICON operationally, at what resolution, domain, and convection configuration. Please name the relevant operational forecast suites (e.g., the DWD ICON convection-permitting systems and the specific WRF deployments intended) with citations, and confirm that 2 km is genuinely representative of operational practice for both models rather than one. This is important because the 2 km choice and the operational recommendations in Section 5 rest on this assertion. The grayzone scheme's description as "NWP-specific tuning" (line 152) should likewise be attributed.
13. Line 28: Missing a reference for socio-economic impacts. The sentence on the importance of correctly forecasting these phenomena "for early warning, prevention, and adaptation" would be well supported by the recent Reviews of Geophysics synthesis on the socio-economic impacts of Mediterranean cyclones. Please consider to add the reference Khodayar et al., 2025.

## Minor comments

- The Abstract characterizes Daniel as featuring a "rapid transition from a baroclinic disturbance to a compact tropical-like vortex." This appears to contradict the manuscript's own Case Study (Sect. 2), which describes a five-phase evolution over ~6 days (4–10 September) with a gradual transition from 7–9 September, and especially the Discussion (lines ~433–437), which explicitly states that "unlike systems that rapidly attain a mature structure, Daniel evolved through a prolonged sequence of physically distinct stages." Please reconcile these: either rephrase the Abstract (e.g., "multi-stage" transition, or restrict "rapid" to the final barotropic alignment on 9 September if that is what is meant) or justify the "rapid" characterization against the stated timeline.
- Line 84: "70–80 km/[t]" — units appear to be km/h; correct the symbol.
- Line 90: "western Egyptian desert" -> "Western Desert of Egypt"
- Line 61: "future recherche" → "future research."
- Line 263: "Aegean Sea" → "Ionian Sea"
- Line 359: "keeping a cold core and chaotic structure" — "chaotic" carries a specific dynamical-systems meaning that is not what is intended here. Replace with a more precise term such as "disorganized" or "incoherent."
- Line 378: "a horizontal extension of a few tents km" → "a few tens of km" (also line 390 region: check similar instances).
- Line 459: SST is mentioned, but never defined. Define it in Line 77 when it is first mentioned.
- Table 1: unresolved markers such as "+ [t]." appear to be stray LaTeX artifacts; please clean up.
- Figure 5 vs. Section 3.4 radius mismatch (see Specific Comment 7).
- Figures 8, 9, and 10 are first introduced and discussed in Section 4.3 (Results) but are typeset within Section 5 (Discussion and Conclusion), separated by one to several pages from the text that interprets them. Please relocate them closer to their first citation to aid readability.

## References:

- Lagouvardos, K., Kotroni, V., Bezes, A., Koletsis, I., Kopania, T., Lykoudis, S., Mazarakis, N., Papagiannaki, K., and Vougioukas, S.: The automatic weather stations NOANN network of the National Observatory of Athens: operation and database, *Geoscience Data Journal*, 4, 4–16, <https://doi.org/10.1002/gdj3.44>, 2017
- Khodayar, S., Kushta, J., Catto, J. L., Dafis, S., Davolio, S., Ferrarin, C., et al.: Mediterranean cyclones in a changing climate: A review on their socio-economic impacts, *Reviews of Geophysics*, 63, e2024RG000853, <https://doi.org/10.1029/2024RG000853>, 2025