Supplementary Material for "Multi-model high-resolution analysis of Tropical-Like Cyclone Daniel with WRF and ICON: peculiarities and sensitivity to convection schemes."

Piero Serafini[1,2], Antonio Ricchi[1,2], Chiara Marsigli[3], Cristiano D'Amico[1,2], Matteo Nastasi[1,2], Renata Pelosini[4], and Rossella Ferretti[1,2]

[1]UNIVAQ (DSFC) - University of L'Aquila (Department of Physical and Chemical Sciences), L'Aquila, Italy
[2]CETEMPS - Center of Excellence in Telesensing of Environment and Model Prediction of Severe Events, L'Aquila, Italy
[3]ARPAE - Regional Agency for Environmental Protection in Emilia-Romagna, Bologna, Italy
[4]ARPAP - Regional Agency for Environmental Protection in Piemonte, Torino, Italy

Correspondence: Piero Serafini (piero.serafini@graduate.univaq.it)

S1 Robustness Assessment of tracks error

To assess the true robustness of the track-skill ranking, we have added a dedicated serial-correlation-aware significance assessment.

The method deserves a short description, as a naive test assuming independent hourly pairs would severely overstate significance. We compare configurations through paired differences of the hourly displacement errors at common valid times. Consecutive hourly errors are strongly serially correlated because position errors persist over many hours; both the standard paired t-test and the Wilcoxon signed-rank test assume independent pairs, so raw p-values computed on the full hourly sample would be spuriously small.

We account for this serial correlation in two complementary ways:

- (i) Paired t-test with effective sample size: The test is applied using an effective sample size $n_{\text{eff}} = n(1 - r_1)/(1 + r_1)$, where r_1 is the lag-1 autocorrelation of the difference series (**??**).
- (ii) Subsampled Wilcoxon signed-rank test: As a distribution-free check, the Wilcoxon test is applied to differences subsampled at 12-hourly spacing to ensure temporal independence relative to the cyclone's evolution.

A difference is reported as significant only when both tests reject the null hypothesis at the 5% level ($\alpha = 0.05$).
The outcome of this serial-correlation-aware analysis is as follows:

- Sample Size & Autocorrelation: With $n = 60$ common hourly samples and lag-1 autocorrelations (r_1) of the difference series ranging between 0.466 and 0.872, the effective sample sizes range between $n_{\text{eff}} = 4.1$ and $n_{\text{eff}} = 21.8$.
- Pooled WRF vs. ICON Families: The difference between the pooled WRF-family and pooled ICON-family series ($-23.34\,\text{km}$) is not significant ($t = -1.224$, $p = 0.2847$ for the n_{eff}-corrected t-test; $p = 1.0000$ for the 12-hourly Wilcoxon test).
- Pairwise Comparisons: Among the pairwise comparisons, no pair was found to be statistically significant once serial correlation was accounted for; the remaining pairs are not distinguishable from baseline variability.
- Observational Uncertainty Benchmark: As a secondary, phase-resolved result, the mean displacement error of every configuration exceeds the $15\,\text{km}$ observational uncertainty outside the cyclogenesis phase (one-sample t-test with the same n_{eff} correction), which quantifies where in the life cycle the models lose the track.

S2 Precipitation insight

S2.1 Precipitation per phases

An analysis of the rainfall divided by cyclone phase is presented. During the extratropical phase, all configurations reproduce the event fairly accurately in terms of both location and intensity (Fig.S1). The first differences appear in the transition phase:

WRF–SH shows a completely de-localized maxima and ICON–CU has no cells with values above the 99th percentile (Fig.S2). Finally, in the tropical phase, all configurations maintain a fairly good localization of the precipitation, except for ICON–CU (Fig.S3).

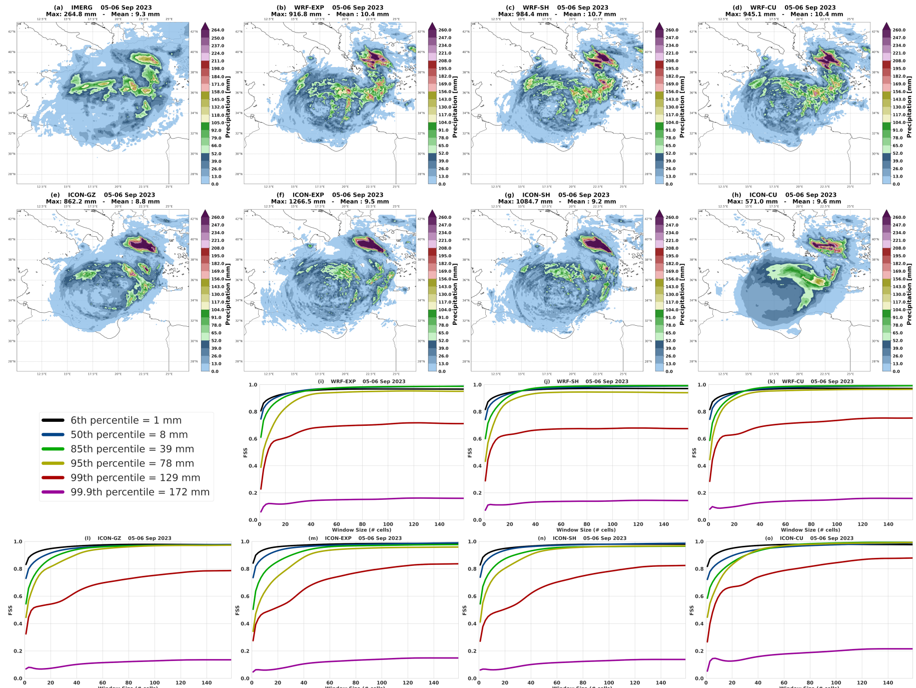

Figure S1. Accumulated total precipitations during extratropical phase (from 2023-09-05 to 2023-09-06). (a) Observations from IMERG, (b) WRF–EXP, (c) WRF–SH, (d) WRF–CU, (e) ICON–GZ, (f) ICON–EXP, (g) ICON–SH, (h) ICON–CU. Color shades as for FSS percentile thresholds. FSS scores for each simulation with respect to IMERG. On the x-axis the window size used for the FSS computation.

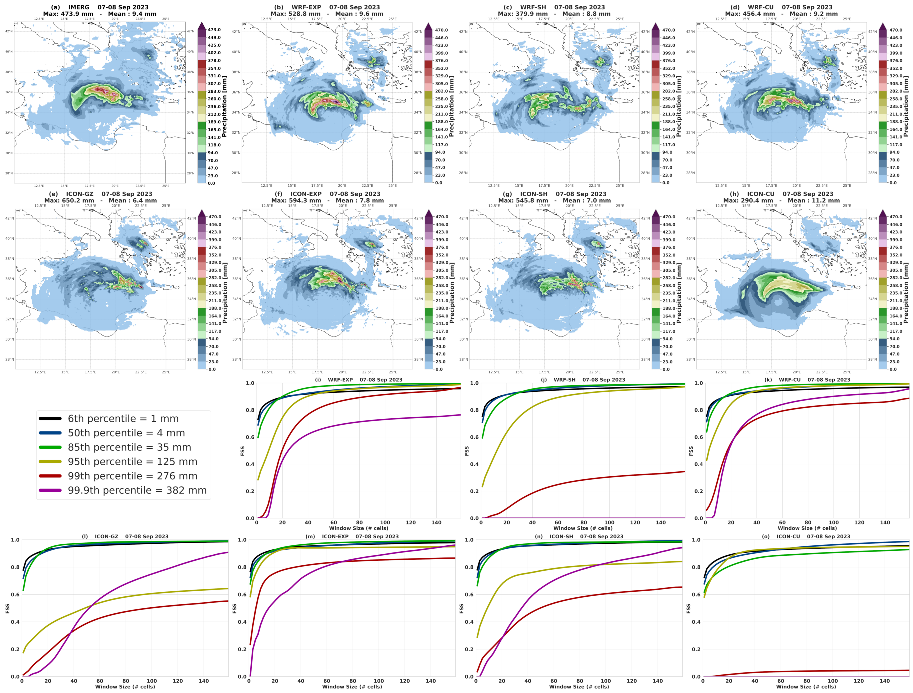

Figure S2. Accumulated total precipitations during transition phase (from 2023-09-07 to 2023-09-08). (a) Observations from IMERG, (b) WRF–EXP, (c) WRF–SH, (d) WRF–CU, (e) ICON–GZ, (f) ICON–EXP, (g) ICON–SH, (h) ICON–CU. Color shades as for FSS percentile thresholds. FSS scores for each simulation with respect to IMERG. On the x-axis the window size used for the FSS computation.

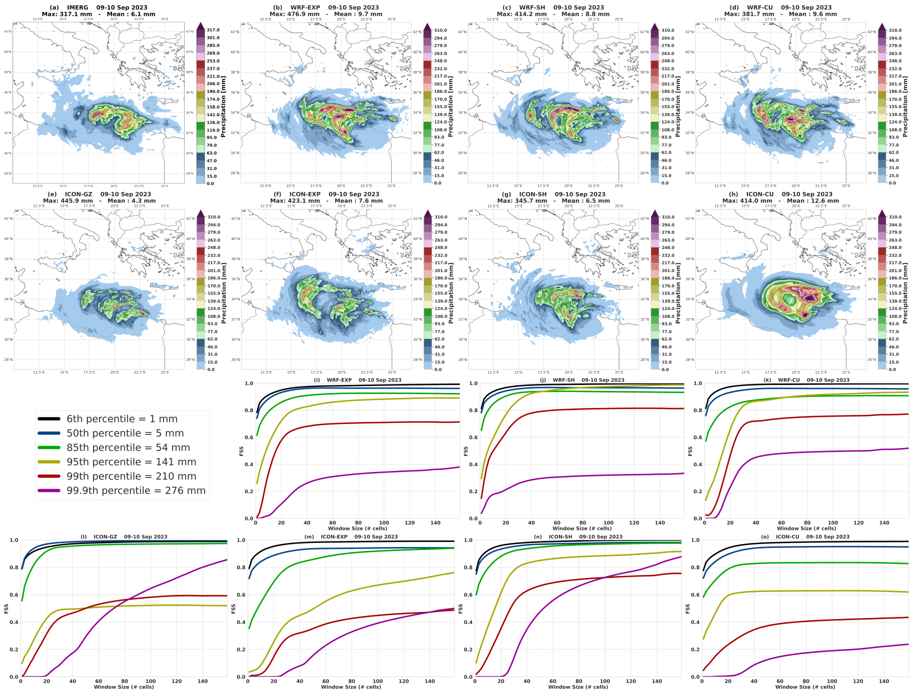

Figure S3. Accumulated total precipitations during tropical phase (from 2023-09-09 to 2023-09-10). (a) Observations from IMERG, (b) WRF–EXP, (c) WRF–SH, (d) WRF–CU, (e) ICON–GZ, (f) ICON–EXP, (g) ICON–SH, (h) ICON–CU. Color shades as for FSS percentile thresholds. FSS scores for each simulation with respect to IMERG. On the x-axis the window size used for the FSS computation.

S2.2 Various comparisons between observations and precipitation

Model-minus-IMERG difference maps of the event-total accumulated precipitation for all seven configurations. Both model fields and IMERG were first conservatively regridded to the common $0.1°$ verification grid, the same used for the FSS, so that the differences are not contaminated by resolution mismatch. The maps show that the dominant error structure along the cyclone track is a dipole pattern centred on the observed rainfall maxima, indicating that displacement, rather than a systematic intensity bias, drives most of the local disagreement during the marine phases. Over Greece, the WRF configurations overestimate accumulation along the windward slopes of the Pindus range, while the ICON configurations underestimate precipitation on the southern flank of the cyclone during the transition phase. The dipole amplitudes are smallest in the SH configurations of both models, consistent with sub-grid transport distributing convective precipitation more evenly and limiting grid-scale moisture convergence.

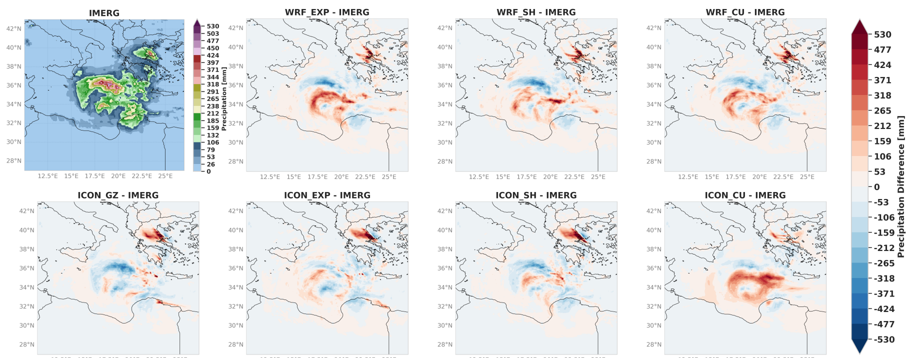

Figure S4. Precipitation field difference Model–IMERG

We computed the frequency distributions of event-total precipitation for all simulations and for IMERG. To avoid areal-averaging artefacts, all distributions are evaluated on the common $0.1°$ verification grid. The probability density function (PDF) is computed for the bins $[1, 10, 50, 100, 200, 300, ..., 1800]$. All configurations reproduce the observed distribution up to approximately the 50th percentile, with a consistent behaviour up to 85th percentile. Above it, the model tails are heavier than the IMERG tail; the excess is largest for ICON–EXP and smallest for the ICON–CU configuration, in agreement with the FSS results at the 99th and 99.9th percentile thresholds. Part of the tail disagreement lies in the reference itself: IMERG Final applies a monthly climatological gauge adjustment and underestimated the observed maxima over Thessaly during this event (Flaounas et al., 2025), so the model tail excess over land should be read as an upper bound. Overall these results give no more informations than what already presented in the previous analysis.

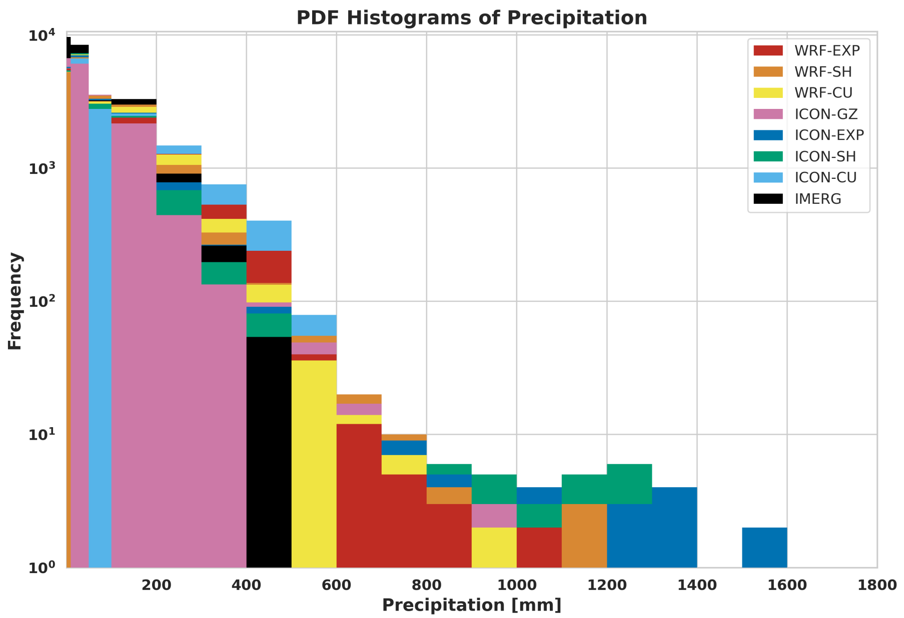

Figure S5. Precipitation PDF. WRF-EXP in red, WRF-SH in orange, WRF-CU in yellow; ICON-EXP in blue, ICON-SH in green, ICON-GZ in magenta, ICON-CU in light-blue, IMERG observations in black.

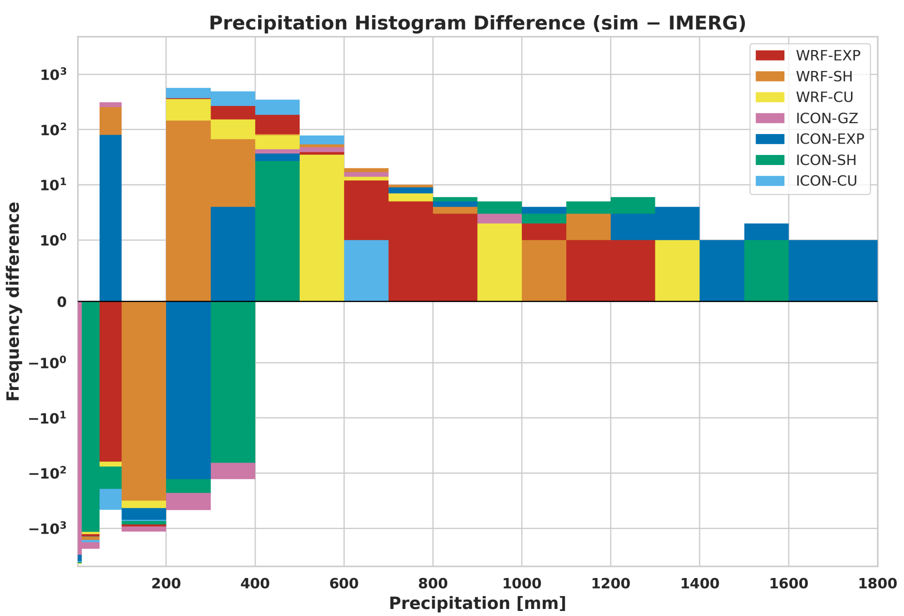

Figure S6. Precipitation PDF Model–IMERG. WRF-EXP in red, WRF-SH in orange, WRF-CU in yellow; ICON-EXP in blue, ICON-SH in green, ICON-GZ in magenta, ICON-CU in light-blue.

S3 TASM axisymmetry and displacement

To assess the degree of asymmetry averaged over the annular domains, we calculated the azimuthal standard deviation of the equivalent potential temperature (θ_e) for each radius and vertical level.

The target control (CTRL) simulation is the WRF–SH one, as it provides the best overall result. The azimuthal mean is defined as:

$$\overline{\theta_e}(r,z) = \frac{1}{N}\sum_{i=1}^{N}\theta_e(r,\phi_i,z)$$

Consequently, the azimuthal dispersion is computed as:

$$\sigma_\phi(r,z) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\theta_e(r,\phi_i,z) - \overline{\theta_e}(r,z)\right]^2} \tag{S1}$$

where r is the radial distance from the cyclone center, z is the vertical level, ϕ_i represents the i-th azimuthal point within the annulus, and N is the total number of grid points within the annular ring. The parameter $\sigma_\phi(r,z)$ provides a quantitative, localized estimate of the field's asymmetry.

The axisymmetry assumption holds remarkably well in the inner core of the cyclone. In this central region, σ_ϕ generally remains below 2 K Fig. S7. As expected, the asymmetry naturally increases moving radially outward. Specifically, we identify a localized maximum of asymmetry at approximately 180 km from the center at the 500 hPa level. Beyond a radius of 250 km, the σ_ϕ values exceed 3, indicating a rapid degradation of the symmetric structure. Therefore, we can confidently state that the axisymmetry assumption underlying the TASM is physically robust and well-justified within the inner core of the medicane (up to a radius of 200–250 km).

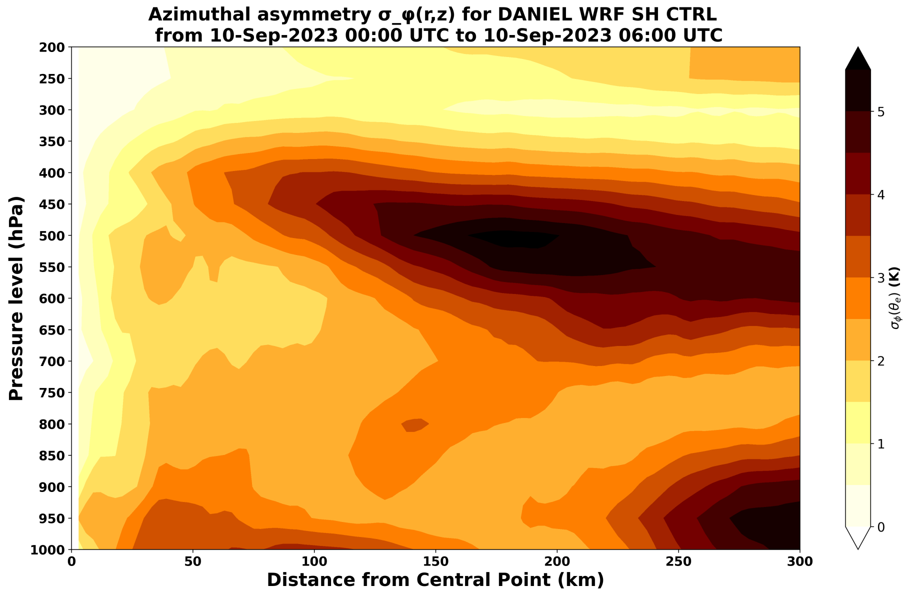

Figure S7. Azimuthal asymmetry (Eq. S1)in the CTRL configuration (WRF–SH). Warm color shades from white to dark brown proportionally to the asymmetry.

To address the propagation of the 15 km positional uncertainty into the annular means, we performed a dedicated spatial sensitivity test on the CTRL simulation dataset. The tracked center of the cyclone was artificially displaced by $0.14°$ (approximately 15 km) in the four cardinal directions (North, South, East, and West). For each of these displaced configurations, we recalculated the TASM and compared it against the original, centered CTRL output Fig.S8.

The statistical comparison reveals that the 2D spatial correlation between the displaced TASM fields and the original reference field is exceptionally high, ranging strictly between 0.9990 and 0.9992.

Therefore, we can state that a center-position uncertainty of 15 km does not significantly propagate into the annular means. The azimuthal averaging effectively dampens the localized spatial noise caused by minor misplacements. Consequently, the TASM calculation proves to be highly resilient to typical tracking algorithm uncertainties, guaranteeing that the resulting thermal anomaly profiles are physically reliable and not artifacts of tracking precision.

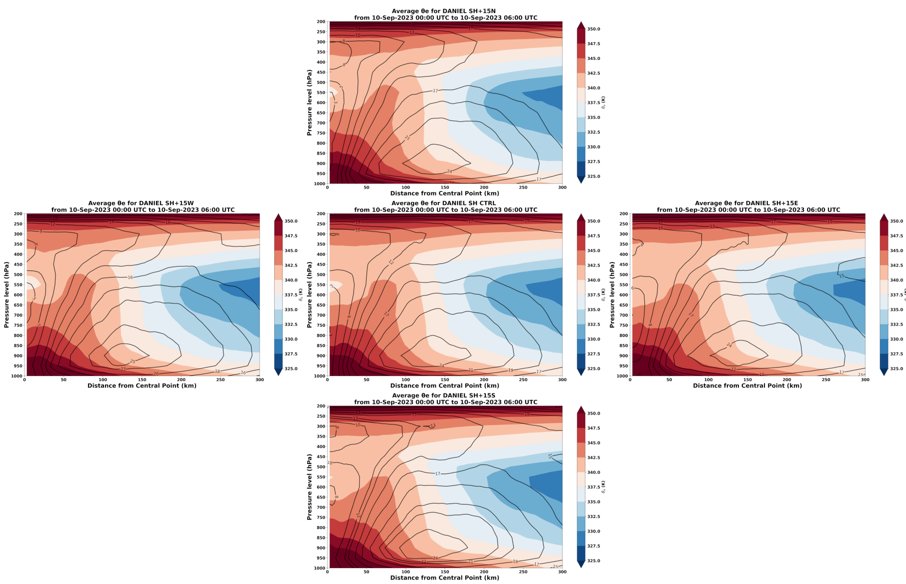

Figure S8. Displacement test in the computation of TASM. CTRL run in the center and cardinal displacement respectively Northward (up), Southward (below), Eastward (right) and Westward (left). Wind speed (m/s, black contours) and equivalent potential temperature (K, filled contours)