Dear Editor,

Thank you for your feedback on our manuscript. We also thank the referees for their reviews during the open discussion.

We have revised the manuscript to ensure that all clarifications provided in our responses are now integrated into the main text.
Below you will find our point-by-point responses to both reviewers and the indications of the changes made in the marked-up manuscript.

Sincerely,

The Authors

Response to Referee #1

We are grateful to the reviewer for the assessment of our manuscript and for providing insightful suggestions. The comments provided are very constructive and have helped us significantly improve the quality of our work. Below, we provide a point-by-point response to each comment and detail the changes made to the revised text.
Please note that all line numbers correspond to the new revised manuscript, and all updated text is quoted in full below.

Major:

1) The discussion employing the classical Froude Number (Fr) to describe airflow dynamics over the Apennines is methodologically flawed. The standard criterion (Fr < 1 for flow splitting/blocking) is derived for isolated, three-dimensional topographic obstacles. For an extended, continuous 2D mountain barrier like the Central Apennines, lateral deflection is heavily constrained by the broad scale of the ridge. Consequently, stable air can frequently be forced over the barrier, rather than simple horizontal splitting. I recommend either removing the classical Froude number argument or revising the text to reflect the specific dynamics of elongated 2D ridges (referencing relevant literature on 2D mountain flow regimes).

R1) Following your recommendation, we have entirely removed the classical Froude number argument from the manuscript, considering that interpretation in terms of Fr is not straightforward. The corresponding text (lines 258-269 in the marked-up version) and related discussions have

been deleted to avoid misleading interpretations and to ensure the description of airflow dynamics remains physically accurate.

2) While the authors acknowledge the underestimation of the peak rainfall (180 mm simulated vs 419 mm observed at Cantiano), the manuscript would benefit from a deeper physical discussion on why the simulated storm decays prematurely after 17:00 UTC. Is this decay driven by excessive cold-pool propagation, insufficient surface latent heat fluxes, or microphysical parameterization limits? Expanding slightly on these potential physical mechanisms would provide greater insight into why NWP models systematically fail to reproduce the extreme tail (>400 mm) of such localized convective events.

R2) We completely agree that a deeper physical discussion regarding the premature decay of the simulated storm may add significant value to the manuscript.
Regarding the hypothesis of excessive cold-pool propagation, our analysis of the observational data did not reveal a distinct cold-pool signature during the event. Conversely, rather than a dynamic misrepresentation of a cold pool, we found that the premature decay of rainfall in the model is primarily driven by a marked near-surface thermodynamic bias. To investigate this, we expanded our analysis of the thermodynamic fields after 17:00 UTC. As now detailed in the revised Section 4.3 and shown in the new Figure 8, the simulation exhibits a significantly warmer near-surface environment, with temperatures remaining even 5-6°C higher than observations during the decay phase. This warm thermal structure directly affects the saturation state over time: considering that the specific humidity is well represented, the higher temperatures increase the saturation vapor pressure, causing the low troposphere to drop from a saturated state to well below 100%, shifting the modeled environment to a sub-saturated state that can no longer sustain condensation and heavy precipitation. This comprehensive physical discussion, together with the corresponding spatial maps and temporal profiles, has been integrated into Section 4.3 (lines 209-226) of the revised manuscript and now reads as follows:

“To investigate the physical mechanisms driving this premature convective decay, an analysis of the near-surface thermodynamic fields was conducted. For the 2 m temperature (T_2m) and 2 m relative humidity (RH_2m) analysis, data from 104 stations within domain d02 (which covers the entire Marche region) and 54 stations within domain d03 (covering only the portion of the region directly affected by the rainfall) were utilized. The validated station observations used for this comparison were retrieved from the official platform of the Civil Protection Department of the Marche Region (http://app.protezionecivile.marche.it/) and basic quality control procedures were applied to ensure data consistency and filter out anomalies. Figure 8 compares the observed and simulated T_2m and RH_2m spatial maps at 18:00 UTC (panels a and b), alongside their temporal evolution (panels c and d) for selected stations (Arcevia, Cantiano, and Colle).

The analysis reveals that the rapid decay of precipitation is related to near-surface thermodynamic biases. The simulation exhibits a significantly warmer environment, with temperatures remaining up to 5-6°C higher than observations during the decay phase.
This warm bias directly affects the near-surface relative humidity over time (Fig. 8d). Due to the higher simulated temperatures, the saturation vapor pressure increases, which in turn reduces the relative humidity (considering that the bias in specific humidity is small). Consequently, the simulated environment fails to maintain saturation. The drop in RH_2m represents a physical transition from a saturated state (necessary to sustain condensation and precipitation) to a sub-saturated environment that inhibits rainfall generation. This underestimation is further illustrated by comparing the 24 h maximum cumulative rainfall in the control simulation (REF) — evaluated both at the station point (REF_point) and using a 10 km neighborhood maximum (REF_10km) — with rain gauge observations (Table 3).”

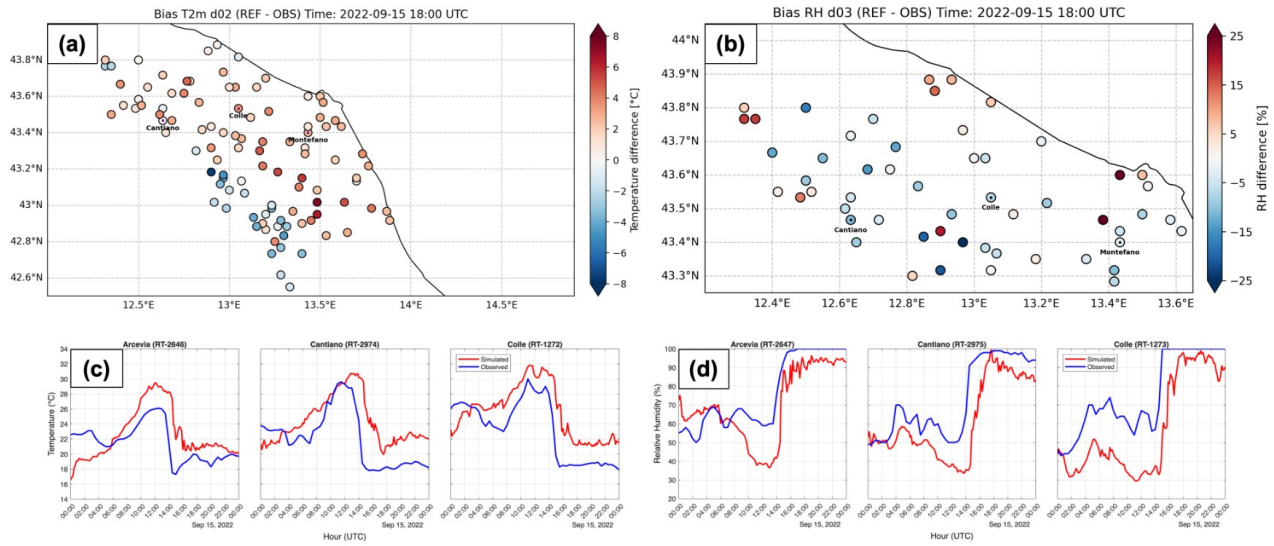


Figure 8. Comparison of observed and simulated (REF) near-surface (a) T_{2m} over a portion of the d02 domain and (b) RH_{2m} over the d03 domain at 18:00 UTC. Temporal evolution of (c) T_{2m} and (d) RH_{2m} comparing observations and model outputs (REF) for the selected stations of Arcevia, Cantiano, and Colle.

Minor:

3) To enhance readability and provide a clear overview for the reader, please include a summary table listing the maximum predicted 24-h precipitation across the different sensitivity experiments (NCEP GDAS, HRES ECMWF, NCEP GFS, ERA5) alongside the observed rain gauge maxima (e.g., Cantiano, Frontone).

R3) We have added a new table (Table 3 in the revised manuscript) comparing the 24 h maximum cumulative rainfall observed by the rain gauge network against the control simulation (REF), evaluated both at the exact station locations and within a 10 km neighborhood box. We have also

added the 24 h maximum cumulative rainfall predicted by each sensitivity experiment improving the existing Table 4.

Table 3. Comparison of 24 h maximum cumulative rainfall and corresponding percentage errors between the control simulation evaluated at the nearest grid point (REF_{point}), the maximum value within the 10 km neighborhood (REF_{10km}), and weather station observations (OBS) for selected locations.

Station	**REF_{point} [mm]**	**REF_{10km} [mm]**	**OBS [mm]**	**Error (REF_{point})** $\frac{OBS-REF_{point}}{OBS} \cdot 100$ **[%]**	**Error (REF_{10km})** $\frac{OBS-REF_{10km}}{OBS} \cdot 100$ **[%]**
Arcevia	35.2	109.7	129.2	72.8	15.1
Cantiano	164.2	180.1	419.0	60.8	57.0
Colle	72.4	89.9	204.0	64.5	55.9
Jesi	34.4	78.8	126.0	72.7	37.5
Frontone–Monte Acuto	172.4	180.6	384.4	55.2	53.0
Montefano	99.8	122.5	150.8	33.8	18.8

Table 4. Summary of the WRF sensitivity experiments. All simulations share the same domains and grid spacing (10 km, 2 km, 0.66 km). Δt_{BC} indicates the update frequency of the lateral boundary conditions. All simulations share the same physical parameterizations, except for the "Modified physics" experiment (panel d). This specific run employs the Thompson microphysics scheme (Thompson et al., 2008), the Kain-Fritsch cumulus parameterization (Kain, 2004), the Dudhia shortwave and RRTM longwave radiation schemes (Dudhia, 1989; Mlawer et al., 1997), and the Unified Noah Land Surface Model (Tewari et al., 2004).

Panel	**Sensitivity**	**Forcing Dataset**	**Start Time (UTC)**	**Δt_{BC} [h]**	**24 h Max. Rain [mm]**	**Remarks**
a	REF	NCEP GDAS/FNL (0.25°)	15/09/2022 00:00	3	180.6	Control simulation
b	I.C.	NCEP GDAS/FNL (0.25°)	15/09/2022 06:00	3	98.9	Delayed initialization
c	B.C.	NCEP GDAS/FNL (0.25°)	15/09/2022 00:00	6	197.1	Lower boundary update frequency
d	Physics	NCEP GDAS/FNL (0.25°)	15/09/2022 00:00	3	74.2	Modified physics (see caption)
e	I.C. / B.C.	NCEP GFS (0.25°)	15/09/2022 00:00	3	129.5	Historical forecast archive
f	B.C.	I.C. GDAS / B.C. GFS	15/09/2022 00:00	3	156.1	Mixed forcing datasets
g	I.C. / B.C.	HRES ECMWF (0.1°)	15/09/2022 00:00	3	49.3	High-resolution operational analysis
h	I.C. / B.C.	ERA5 (0.25°)	15/09/2022 00:00	1	108.5	Hourly reanalysis

Response to Referee #2

We would like to thank the reviewer for the time spent evaluating our work. The comments helped us improve the manuscript. Please find below our point-by-point response to the feedback, along with details of the changes made.
Please note that all line numbers correspond to the new revised manuscript, and all updated text is quoted in full below.

General comments:

1) The rainfall evaluation documents important limitations of the reference simulation. It produces about 160 to 180 mm where the observed maximum exceeds 400 mm, misses the later weaker rainfall phase, and delays the main phase by about 1 to 2 h. The closest agreement is obtained with the 10 by 10 km neighbourhood method, which extracts a maximum from the model field rather than comparing the simulated value at the gauge. Please distinguish this useful displacement tolerant diagnostic from a direct station evaluation, report simple aggregate errors for both approaches, and restructure the Abstract and Conclusions accordingly. In particular, a retrospectively selected GDAS based reconstruction cannot establish operational predictive skill or the early identification of hydrogeological risk. It can support a more limited statement about the simulated precursors and mechanisms of this case.

R1) We have introduced a new table (Table 3) presenting a quantitative comparison of the 24 h cumulative rainfall in the observations and in the control simulation using both the station points and the 10 km neighborhood method, including the relative percentage errors to quantify the model underestimation. We added text at lines 199-201 to clearly explain the conceptual differences between the two evaluation approaches and why the neighborhood method accounts for minor spatial shifts.

Finally, we changed the Abstract (lines 11-17), Introduction (lines 61-64), and Conclusions (lines 407-410) to discuss the different roles of both evaluation methods and to moderate the statements regarding the potential use for early warning systems or operational predictive skill. The manuscript now frames the GDAS-driven (REF) simulation as a diagnostic investigation focused on evaluating the physical precursors and dynamical mechanisms of the convective event.

The revised sentences are reported below:

- Lines 199-201: “Direct point evaluation misses the intensity of the peak rainfall even if the location of the simulated storm is correctly reproduced, being shifted by just a few kilometers. In contrast, the neighborhood approach removes errors due to small spatial offsets and provides a more reliable estimate of the actual storm intensity.”

- Abstract (lines 11-17): "Furthermore, the simulations suggest that gravity waves induced by the mountains may have helped to sustain the vertical motion of the impinging air, and highlight the role of upper-level dynamics (passage of a jet streak). The comparison with surface stations shows that the simulation underestimates peak accumulations (160-180 mm simulated versus over 400 mm observed), delays the main phase by 1 to 2 h, and misses the final phase. However, overall the model captures the system's spatial position, orientation, and quasi-stationary evolution associated with orographic anchoring.
  Rather than evaluating predictive skill in an operational-like perspective, this study investigates the physical precursors and dynamical mechanisms of the event. "
- Introduction (lines 61-64): "In this context, the use of high-resolution numerical models becomes essential to improve our understanding of the atmospheric drivers associated with severe hydrogeological impacts in the complex terrain of the Apennine."
- Conclusions (lines 407-410): "Nevertheless, the model's ability to correctly identify the structural precursors — such as the capping inversion, the convergence line, and the moisture transport — highlights its utility in representing the essential physical mechanisms driving this extreme convective event."

Table 3. Comparison of 24 h maximum cumulative rainfall and corresponding percentage errors between the control simulation evaluated at the nearest grid point (REF_{point}), the maximum value within the 10 km neighborhood (REF_{10km}), and weather station observations (OBS) for selected locations.

Station	**REF_{point} [mm]**	**REF_{10km} [mm]**	**OBS [mm]**	**Error (REF_{point})** $\frac{OBS-REF_{point}}{OBS} \cdot 100$ **[%]**	**Error (REF_{10km})** $\frac{OBS-REF_{10km}}{OBS} \cdot 100$ **[%]**
Arcevia	35.2	109.7	129.2	72.8	15.1
Cantiano	164.2	180.1	419.0	60.8	57.0
Colle	72.4	89.9	204.0	64.5	55.9
Jesi	34.4	78.8	126.0	72.7	37.5
Frontone–Monte Acuto	172.4	180.6	384.4	55.2	53.0
Montefano	99.8	122.5	150.8	33.8	18.8

2) The experiments convincingly show that the simulation is sensitive to the forcing dataset, but the GDAS and ECMWF configurations differ in more than the vertical resolution of the input fields. A comparison of 500 hPa geopotential height and theta-e alone does not prove that the large scale forcing is irrelevant to predictability, nor does it isolate the lower level pressure gradient as the cause of the rainfall displacement. Please present these results as evidence of sensitivity among the forcing datasets and soften causal statements. A concise quantitative comparison of the relevant temperature, moisture, wind, and pressure differences at the initial and boundary times would make the interpretation more transparent.

R2) We agree with the reviewer and have updated lines 376-389 accordingly (marked-up version). We removed the statements regarding the lack of influence of the synoptic scale forcing and reframed the analysis around model sensitivity to the input data. Following the reviewer's

suggestion, we added a quantitative comparison of 850 hPa geopotential height, temperature, relative humidity, and SST at the initial and boundary times, supported by Fig. S3 in the Supplementary Information. The text now describes how low-level differences in moisture and temperature timing are linked to the simulated rainfall shifts.

The revised text at lines 376-389 now reads as follows:

“While mid-tropospheric fields (500 hPa geopotential height and 500 hPa θe) are broadly similar in the driving datasets, clear differences appear in the lower troposphere (850 hPa). The initial and boundary fields (Fig. S3) reveal a timing mismatch in lower-tropospheric moisture and temperature throughout the event. From initialization up to 06:00 UTC, the run driven by ECMWF forcing simulates warmer temperatures (from +1 to +3 °C) and higher relative humidity (from +10 to +40 %) over central Italy than in REF, alongside lower 850 hPa geopotential height (~ -15 to -20 gpm), suggesting an earlier transit of the trough axis compared to REF. Therefore, during the time period when convective initiation starts (12:00-18:00 UTC) the pattern reverses: REF becomes warmer (from +1 to +3 °C) and wetter (from +20 to +40 % higher relative humidity) directly over the target area. Meanwhile, Sea Surface Temperature differences remain small (~ 0.5-1 °C) and play only a secondary role; although REF exhibits slightly higher temperatures in the western Mediterranean and Alboran Sea, these features appear too far from our target domain to fully explain the different simulation outcomes. These results highlight the model sensitivity to the low-level forcing dataset, showing that differences in timing and location of moisture transport relative to the Apennine topography may contribute to significant changes in the simulated rainfall.”

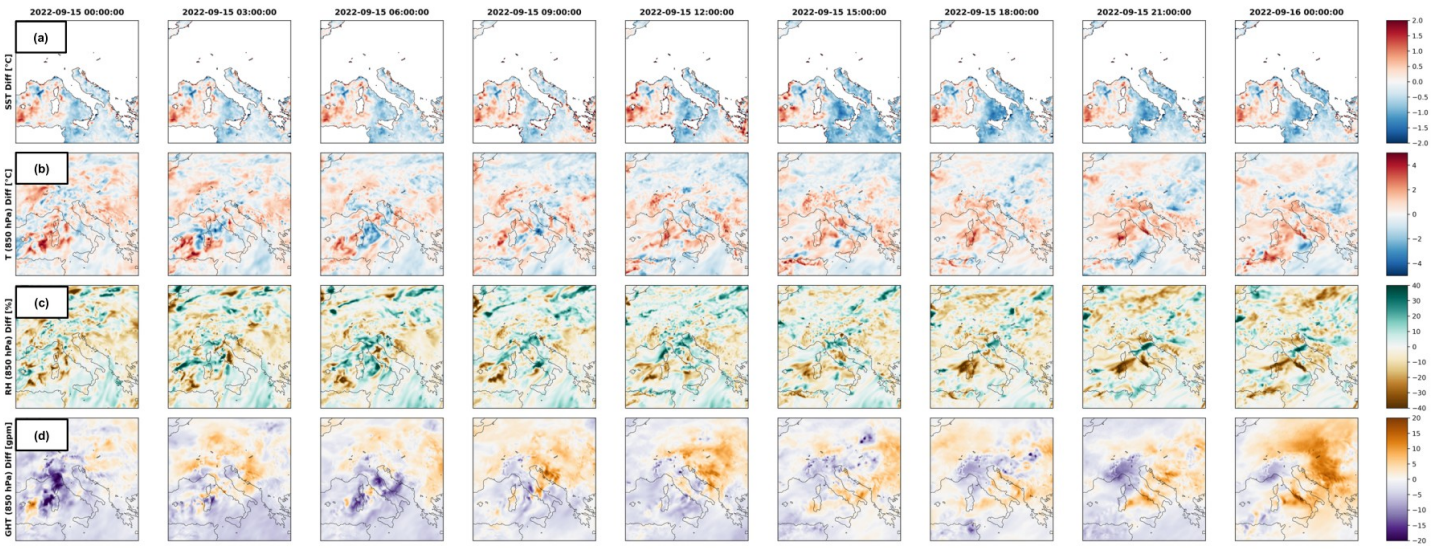


Figure S3. Spatiotemporal evolution of lower-troposheric and surface differences between the driving forcing dataset (REF minus ECMWF) throughout the event. Rows from top to bottom show differences in: (a) Sea Surface Temperature (°C), (b) 850 hPa temperature (°C), (c) 850 hPa relative humidity (%), and (d) 850 hPa geopotential height (gpm). Columns show the evolution at 3 h steps throughout the simulation.

3) The terrain reduction experiments support a contribution from the Apennines to the location and amplification of the simulated rainfall. However, the present evidence is less decisive for the stronger conclusion that gravity waves are generated and are essential to anchor the system. Alternating vertical motion in one cross section is consistent with a wave response, but it is not a complete diagnosis. The Froude number also uses a fixed 1000 m mountain height across a complex barrier. Please either add a direct wave diagnostic and assess the sensitivity of the Froude estimate to the chosen barrier height, or use more qualified language throughout the manuscript. A figure showing the spatial and temporal evolution of the Froude number would also help support this interpretation. The method used to modify the terrain and any associated changes to static fields should be described clearly.

R3) Following the suggestions of both reviewers, we have removed the Froude number analysis (lines 258-269) and moderated our conclusions regarding gravity waves. In lines 285-288 and 400-403, we changed the text, identifying in the terrain primarily a mechanical barrier that drives low-level convergence and uplift, rather than a trigger for anchoring gravity waves.

Also, we updated the text to clarify the terrain modification method (lines 273-274), i.e., the surface elevation was reduced (to 50% and 0%) within the static geographic pre-processing files (geo_em) across all nested domains in the region (42.0 – 45.0°N, 10.5 – 15.0°E). Running the initialization pipeline after these modifications ensured that all model fields adapted to the modified terrain before starting the simulation.

The revised sentences read as follows:

- Lines 273-274: “The modifications were applied to the static geography fields across all nested domains prior to model initialization.”
- Lines 285-288: “This result reaffirms that the mesoscale flow dynamics determine the storm’s triggering and positioning, whereas the orography is essential for anchoring the system and providing the mechanical lift. While vertical velocity fields suggest mountain wave activity, the terrain mainly acts as a barrier forcing low-level convergence and uplift.”
- Lines 400-403: “In addition, vertical cross-sections show alternating vertical motion patterns consistent with orographic gravity wave activity, which may have locally modulated the updrafts leeward of the mountain barrier.”

Specific Comments:

1) The final namelists, preprocessing configuration, and postprocessing scripts should be archived in a public repository rather than being available only on request.

R1) As requested, we have archived the final namelists, preprocessing configuration, and postprocessing scripts in a public repository. They are now freely available on Zenodo at the following DOI: https://doi.org/10.5281/zenodo.22808447. The manuscript has been updated accordingly (see the "Data Availability" section).

2) The 12:00 UTC Pratica di Mare comparison shows simulated MUCAPE of 1731 J kg−1 against 3092 J kg−1 in the sounding. This is useful information, but it should be described as a substantial underestimation rather than qualitative agreement. A compact comparison of the humidity and wind profiles would help assess whether the model represents the ingredients relevant to the later convergence and moisture transport.

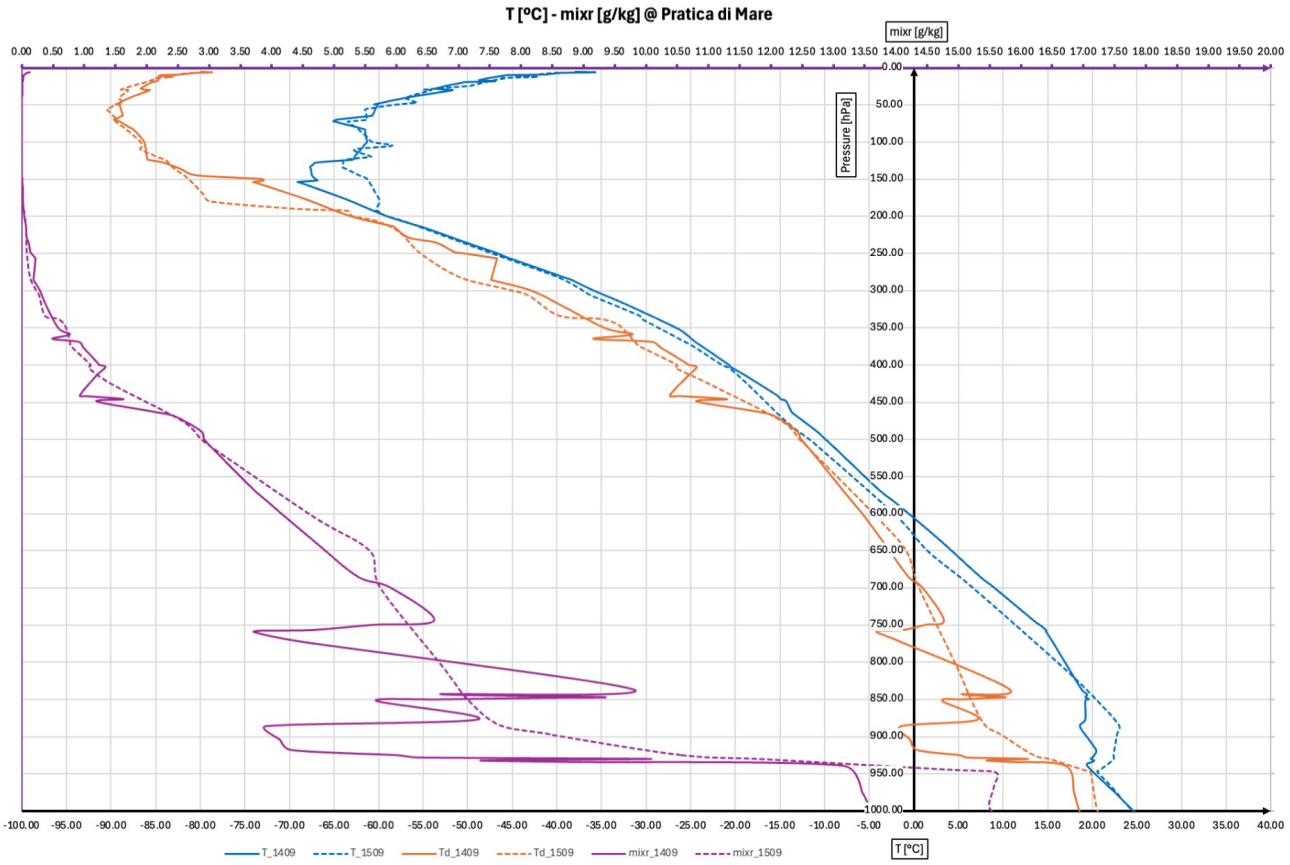
R2) In lines 174-176 of the marked-up manuscript, we updated the text to explicitly state that the model substantially underestimates the available CAPE, despite it captures the presence of an unstable atmosphere.
Regarding the vertical profiles, we show hereafter a comparison between the observed Pratica di Mare soundings at 12:00 UTC on 14 September and 15 September (see the figure attached below; please note that we preferred not to include it in the Supplementary Material). The comparison shows the significant increase in low-level moisture on the day of the event. However, because Pratica di Mare is situated far upstream from the impacted area, these two vertical profiles alone do not reveal the local mechanisms driving the rainfall underestimation further inland.

To address the causes of the rainfall underestimation and the premature storm decay after 17:00 UTC, we expanded our analysis in Section 4.3 (lines 209-226) and added Figure 8. Observational surface data showed no distinct cold-pool signature during the event. Instead, the simulation develops a near-surface warm bias of 5-6°C compared to observations during the decay phase. This higher temperature increases the saturation vapor pressure, causing the relative humidity to drop below 100% and preventing condensation, which leads to the premature decay of precipitation (see also the response to Reviewer 1, comment R2).

The revised sentences read as follows:

- Lines 174-176: “A comparison with the observed sounding (MUCAPE = 3092 J kg−1) reveals that the model captures the presence of conditions favorable to instability, although it significantly underestimates the value of MUCAPE.”
- Lines 209-226: “To investigate the physical mechanisms driving this premature convective decay, an analysis of the near-surface thermodynamic fields was conducted. For the 2 m temperature (T_2m) and 2 m relative humidity (RH_2m) analysis, data from 104 stations within domain d02 (which covers the entire Marche region) and 54 stations within domain d03 (covering only the portion of the region directly affected by the rainfall) were utilized. The validated station observations used for this comparison were retrieved from the official platform of the Civil Protection Department of the Marche Region

(http://app.protezionecivile.marche.it/) and basic quality control procedures were applied to ensure data consistency and filter out anomalies. Figure 8 compares the observed and simulated T_2m and RH_2m spatial maps at 18:00 UTC (panels a and b), alongside their temporal evolution (panels c and d) for selected stations (Arcevia, Cantiano, and Colle). The analysis reveals that the rapid decay of precipitation is related to near-surface thermodynamic biases. The simulation exhibits a significantly warmer environment, with temperatures remaining up to 5-6°C higher than observations during the decay phase. This warm bias directly affects the near-surface relative humidity over time (Fig. 8d). Due to the higher simulated temperatures, the saturation vapor pressure increases, which in turn reduces the relative humidity (considering that the bias in specific humidity is small). Consequently, the simulated environment fails to maintain saturation. The drop in RH_2m represents a physical transition from a saturated state (necessary to sustain condensation and precipitation) to a sub-saturated environment that inhibits rainfall generation. This underestimation is further illustrated by comparing the 24 h maximum cumulative rainfall in the control simulation (REF) — evaluated both at the station point (REF_point) and using a 10 km neighborhood maximum (REF_10km) — with rain gauge observations (Table 3).”



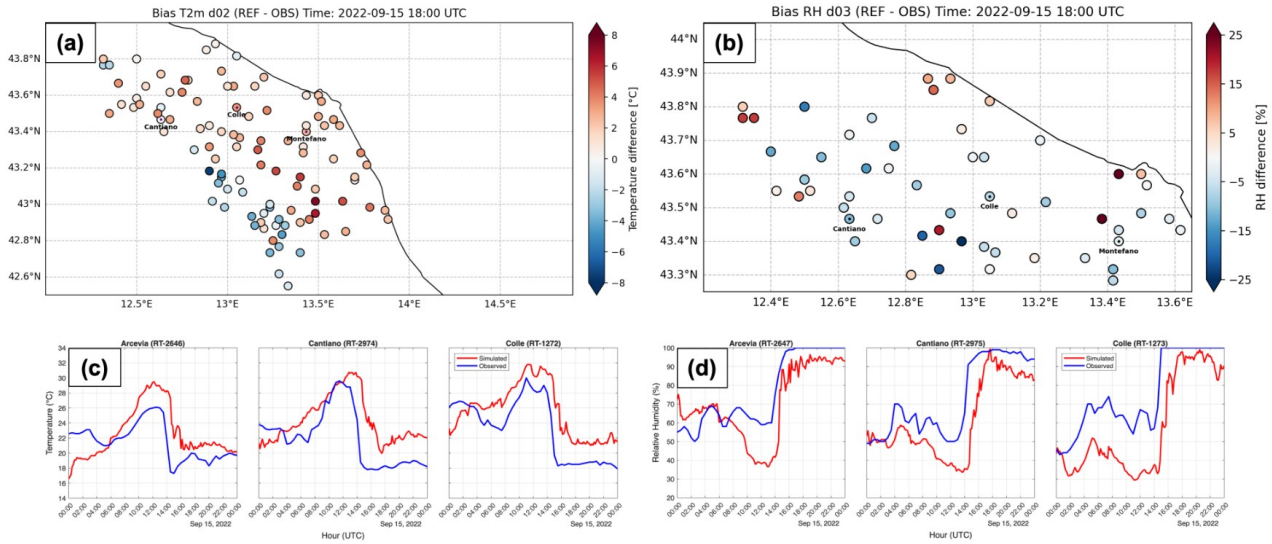


Figure 8. Comparison of observed and simulated (REF) near-surface (a) T_{2m} over a portion of the d02 domain and (b) RH_{2m} over the d03 domain at 18:00 UTC. Temporal evolution of (c) T_{2m} and (d) RH_{2m} comparing observations and model outputs (REF) for the selected stations of Arcevia, Cantiano, and Colle.

3) The modified physics experiment changes the microphysics, convection, radiation, and land surface schemes simultaneously. It therefore shows sensitivity to this physics package rather than the contribution of any individual scheme. Please state this explicitly when discussing Fig. 14.

R3) We agree with the reviewer and have updated lines 328-332 to state this point clearly. The manuscript now clarifies that this experiment demonstrates sensitivity to the overall physics package as a whole, rather than isolating the contribution of any individual scheme.

The updated text in lines 328-332 now reads as follows:

"Furthermore, the choice of the parameterization schemes plays a crucial role in the spatial distribution and intensity of the precipitation; as shown in Fig. 14d, changing the model physics package — by simultaneously modifying the cumulus convection, microphysics, radiation, and land surface schemes — significantly alters the precipitation: this sensitivity run shifts the rainfall pattern northward and decreases the rainfall amount compared to the control simulation."

4) For the Froude number calculation, please identify the precise upstream averaging area, the temporal sampling, and the treatment of terrain height. The statement that the flow was definitely able to surmount the barrier should be softened because this bulk index is only one approximation for such an irregular, nonuniform terrain profile.

R4) We thank the reviewer for this remark. Following the recommendation of Reviewer 1, we have completely removed the Froude number analysis from the manuscript. The text now focuses

directly on how the Apennine terrain acts as a mechanical barrier driving low-level convergence and moisture transport.