Response to the referee's comments; egusphere-2026-1539, Greciano-Zamorano et al.

Emilio Greciano-Zamorano[1,2,3], J. Fidel González-Rouco[2], Jorge Navarro-Montesinos[4], Cristina Vegas-Cañas[2], Félix García-Pereira[2], Elena García-Bustamante[4], Esteban Rodríguez-Guisado[3], Ernesto Rodríguez-Camino[5], Juan Carlos Sánchez-Perrino[3], Rita M. Cardoso[6], Sara Madera-Sánchez[1,4], Almudena García-García[7], Francisco José Cuesta-Valero[7], Hugo Beltrami[8,9], and Freddy Pinochet-Erazo[8,9]

[1]Faculty of Physics, Complutense University of Madrid, Madrid, Spain.
[2]Geosciences Institute, IGEO (CSIC-UCM), Madrid, Spain.
[3]Agencia Estatal de Meteorología (AEMET), Spain.
[4]Research Center for Energy, Environment and Technology (CIEMAT), Madrid, Spain
[5]Spanish Meteorological Society (AME), Madrid, Spain
[6]Institute Dom Luís (IDL), Faculty of Sciences, University of Lisbon, Lisbon, Portugal.
[7]Department of Remote Sensing, Helmholtz Centre for Environmental Research–UFZ, Leipzig, Germany.
[8]Climate & Atmospheric Sciences Institute, St. Francis Xavier University, Antigonish, Canada.
[9]Department of Earth and Atmospheric Sciences, University of Quebec at Montreal (UQAM), Montreal, Canada.

Correspondence: Emilio Greciano-Zamorano (emiliogr@ucm.es) and J. Fidel González-Rouco (fidelgr@ucm.es)

The authors would like to thank the reviewers for their constructive suggestions and the time they devoted to reading and evaluating the manuscript. We have tried to integrate all suggestions and think that the manuscript has improved with them. We do appreciate their contribution.

The next sections contain a detailed point-by-point response to the reviewers' comments. Comments are labeled by reviewers and in order of appearance, i.e. R2C3 is the third comment of reviewer 2. The original number by the reviewer is also preserved if it was given.

1 REFEREE 1

Referee's major comments:

R1C1: **I. Physics-Dynamics Coupling and Cumulus Parameterization. The authors use the New Tiedtke cumulus scheme across all domains, including the 1 km resolution (WRF3). As established in recent literature (e.g., Tomassini et al. (2023)), the range 1-5 km is the convective gray zone where convection is partially resolved. Unless I am misinterpreting the results presented in Figure 11, Simulation 3 (New Tiedtke-CPS, where parameterization is off at 1 km) appears to reduce the precipitation overestimation reported in the Reference run. In Panel (b), the WRF3 results (blue dots) for high-accumulation sites seem better aligned with the 1:1 observational diagonal than in Panel (a). The authors admit that "other parameterizations or CPS approaches may be more adequate" at 1 km. Could the authors clarify if I have overlooked a specific reason why the Reference configuration was preferred for the 30-year study despite their own sensitivity tests showing that scale-aware schemes or a CPS-off approach would have yielded superior results?**

Authors' response: **The authors consider this comment very reasonable. We agree that the CPS (Convection Permitting Scheme) approach in the 1 km simulation (WRF3) reduces the overestimation of precipitation, as shown by the 1:1 observational diagonal in Figure 11b. However, it must be noted that the sensitivity experiments presented in this manuscript were not designed as a preliminary tool to select the optimal setup before the 30-year run. In fact, the 30-year long simulation had been developed before this study was performed. We would need to make sure, but possibly the execution of this simulation preceded the publication cited by the reviewer. Instead, the sensitivity experiments were performed as a diagnostic analysis to explore the model's limitations, discuss the uncertainties within the convective gray zone, and better understand the physical mechanisms that cause the precipitation biases reported by the Reference run. Redoing such a long run was not an option for the preparation of this study. In spite of that, we think the present study offers insights about the precipitation in the Guadarrama Mountains and the realism of modeling in progressive resolution increments. Also, it provides further evidence supporting Tomassini et al. (2023). Of course, it will also influence the configuration of future experiments by the coauthors. Please, see also R2C3.**

We have added a statement in the text, as well as the reference Tomassini et al. (2023), as a result of this comment from the reviewer:

Line 143, in the tracked changes document: The 30-year-long simulation presented here included the New Tiedtke (Zhang and Wang, 2017; Tiedtke, 1993) for the cumulus parameterization. <u>The simulation had been completed before making it available for this study.</u> In order to assess the convenience of including convection-permitting schemes, several additional experiments were performed to simulate year 2009 data <u>during the development of this analysis,</u> using 6 different configurations: [...]

Line 141, in the tracked changes document: One issue that is discussed herein is the convenience of using convection-permitting schemes (Lind et al., 2020; Vergara-Temprado et al., 2020; Tomassini et al., 2023) [...]

R1C2: **II. Benchmarking against High-Resolution Reanalysis (CERRA). The study evaluates Added Value solely against ERA5 (31 km). Given that ERA5 fundamentally misses the sharp orographic peaks of the Sierra de Guadarrama, proving that WRF is "better" than a 31 km global model is a "low bar" that provides limited insight into the actual merit of the 1 km simulation. I would suggest expanding the benchmarking beyond ERA5. It would be very helpful for the reader to see how both the 3 km (WRF2) and 1 km (WRF3) climatologies compare to an existing regional reanalysis such as CERRA (5.5 km). Since the results show that WRF2 often matches or exceeds WRF3 in statistical skill, it would be valuable to demonstrate if these kilometer-scale simulations provide any physical value that a regional reanalysis does not already offer.**

Authors' response: **We consider the reviewer's suggestion highly reasonable. We have used ERA5 because we want to evaluate the potential improvement and deterioration of the simulations with respect to the reanalysis that has been used to provide boundary conditions for the WRF simulation. In this sense, the comparison is fair and addresses the question of whether improving resolution adds more realism and where this may or may not happen. Comparing with products like CERRA addresses a different (also interesting) question, namely, whether a 1 km dynamical downscaling with the WRF model performs relatively better or worse and where, in comparison to a dynamical downscaling of ERA5 that nearly reaches kilometer-scale resolution, like CERRA, and, importantly, it incorporates additional data assimilation.**

We are currently working on an extended domain that covers the entire Sistema Central. For the moment we are gathering the observational database, but we will truly consider your indication when we start to compare these observational data with simulations or reanalyses.

R1C3: **III. Microphysics Sensitivity. The manuscript identifies a persistent overestimation of wet-day frequency, reaching a factor of 2 in summer. While the study tests various cumulus schemes, the Thompson microphysics remains constant throughout all experiments. The overestimation of low-intensity events is often a symptom of grid-scale condensation and autoconversion processes, which are inherently microphysical. Considering that this overestimation is the primary driver for the overstated precipitation totals, would the authors expect substantial differences in these results if a different microphysics scheme were used? It would be scientifically valuable to discuss whether the current results—specifically the peaks of 1800 mm/year at the summits—are physically robust or if they might be sensitive to the choice of microphysics (e.g., using a different two-moment scheme or an alternative single-moment scheme like WSM6).**

Authors' response: **When we carried out the sensitivity tests for the year 2009 shown in the text for the cumulus parameterization, we also conducted an alternative simulation using the NSSL2 microphysics parameterization**

(Mansell et al., 2010). The results are shown below (Fig. R1C3-1). As the differences between the Thompson scheme and the fully two-moment NSSL2 scheme were not significant, we decided not to include these results in the original manuscript. The text has been modified by adding a new sentence in order to explain this:

Line 407, in the tracked changes document: Nevertheless, it allows for briefly inspecting with a few relatively short simulations of one year, the potential of changing how the simulation of convection is affected. <u>Additional simulations were performed changing the microphysics parameterization to the National Severe Storms Laboratory double-moment parameterization</u> (NSSL2; Mansell et al., 2010)<u>, but rendered non-significant results and are not reported here.</u>

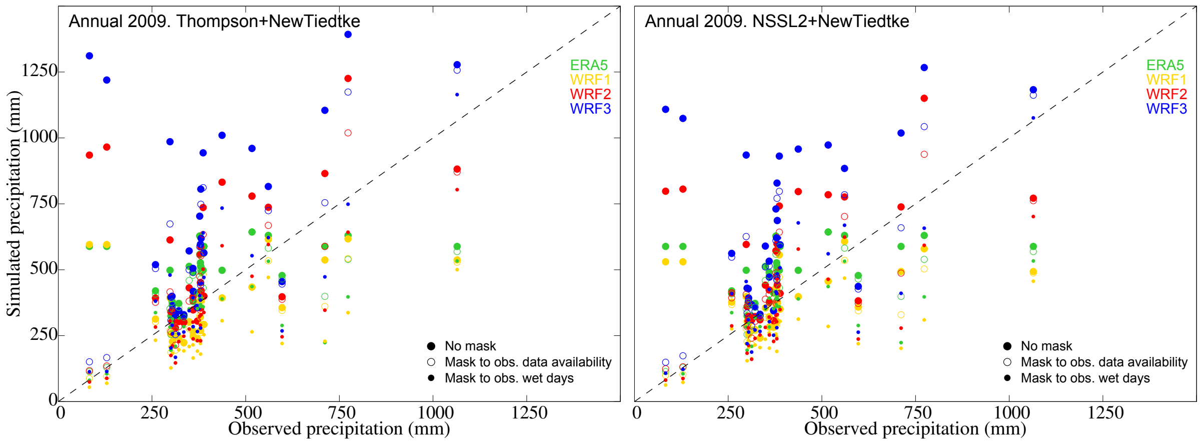

Figure R1C3-1. Comparison between observed and simulated annual accumulated precipitation by WRF and ERA5 within the D_3 domain for configurations of WRF involving different microphysics parameterizations: Thompson (reference simulation) and NSSL2. The symbol convention follows that of Figures 5 and 11, from the tracked changes document.

<u>Referee's minor comments:</u>

R1C4: **In Figures 5 and 11, the scatter plots have inconsistent ranges (e.g., in Fig. 5 (Top), Y-axis goes to 2000 mm while X-axis ends at 1450 mm) and non-square aspect ratios. For accurate visual assessment of the 1:1 line, the authors should use identical ranges for both axes and a perfectly square frame. This ensures that the reader can correctly perceive the magnitude of model biases.**

Authors' response: **Initially, these figures were designed with a square frame and equivalent axes; however, we later modified them to optimize space utilization within the page layout. Nevertheless, we understand that for an accurate visual assessment of the 1:1 line, it is better to return to our original design. Consequently, we have adjusted both Figures 5 and 11 according to the reviewer's indication. Please see the document with tracked changes.**

2 Referee 2

<u>Referee's major comments:</u>

R2C1: **1. The authors used a remarkable monitoring network installed and maintained in such a complex environment. However, Table 1 shows that some stations are only available for 3 or 4 years instead of the long coverage provided by most of the stations. For example, stations 3 (Hoyas), 8 (Cabeza Mediana), 18 (Alameda) and 26 (Herreria). Since the simulations are run for 30 years, I would not include these stations in the analyses, as they can produce fictitious biases when compared against the entire period as done in the first part of the analyses (Figs 2 and 3). In fact, the authors already mention this in lines 184-186. Thus, what is the reason to include these potentially problematic stations in the analysis?**

Authors' response: **We acknowledge that some stations, especially station 3 (Hoyas), cover a very short temporal period. We decided to include them in the study because they are located at the highest altitudes, which is of particular value for the case of this study. We did not want to lose such high-mountain data despite the limited number of years available. Consequently, temporal masking analyses have been carried out in our study, so that the simulation is strictly evaluated against each station only during the specific periods with available observations. Indeed, the results show that while some stations highlight the shortage of data in high-altitude zones, others provide consistent and useful information. This stems clearly from Figures 4, 5, and 7 as well as the maps in Figure 9 and thereafter. We think this provides valuable complementary information. In order to justify this decision, the text has been updated by including a new sentence:**

Line 90, in the tracked changes document: The GuMNet sites (circles in Fig. 1) are situated in the mountain environment between 920 and 2225 m a.s.l. Most of them are located on the southeast side of the Sierra, in the province of Madrid, except for one in the northwest side, in the province of Segovia (see stars indicating the location of nearby cities to aid geographical orientation in Fig. 1). The time coverage of these observations is limited from 2003 to 2020 in the best case and to two years of data in the case with less data (site 3). <u>Despite its limited time span, this short-record station was intentionally included due to its high-altitude placement</u>. The AEMET sites used in this work (squares in Fig. 1) were selected to allow for a meaningful comparison with model data and include all the available sites with good data coverage during the period of simulation, i.e. 1991 to 2020.

R2C2: **2. The authors employ the New Tiedtke scheme as cumulus parameterization even in the 1 km resolution domain. However, at the end of section 2, they mention some sensitivity experiments done for year 2009 to evaluate the performance of the model when the parameterization is switch-off in that domain, or when using alternative parameterization options. However, why was 2009 considered to run those experiments? Did these alternative runs also include a 1 year spin-up as the main simulation?**

Authors' response: **The year 2009 was selected due to the good density of observational data available from AEMET in the study region. Regarding the spin-up procedure, we did not perform independent one-year spin-ups for each sensitivity run. Instead, all experiments were initialized directly from the same restart file of the main simulation to ensure that all sensitivity runs share identical initial states. Also, the spin-up process is mostly oriented to initialize the slow evolving processes, basically related to the soil. This approach was carried out in order to guarantee that the differences in the outputs are strictly due to the physical parameterizations under test and not to to internal model drift or differences in spin-up trajectories. The text has been modified to address this issue. We appreciate the reviewer's clarification.**

Line 145, in the tracked changes document: In order to assess the convenience of including convection-permitting schemes, several additional experiments were performed to simulate year 2009 data using 6 different configurations: the configuration used in the 1991–2020 simulation (Table 3) has been compared to 5 additional simulations in which cumulus parameterizations have been changed or switched off to allow for the direct simulation of convection (simulations 1-5; Table 3). All these test simulations start from the same restart file resulting from the last time step of 31 December 2008, thus sharing the same initial conditions.

R2C3: **3. Also related to the sensitivity experiments. The authors stated in the discussion than they obtain be[er results with these simulations than with the configuration of the main simulation (406-408). Thus, why was that configuration used in the first place and not the one from the sensitivity experiments?**

Authors' response: **The authors agree with the reviewer's point. This comment addresses the same concern raised by R1C1. Please refer to our detailed response to Reviewer 1 above, where we explain the nature of the sensitivity experiments and the computational constraints of the project. For clarity, we briefly state here that the sensitivity experiments were performed as a retrospective diagnostic tool to explore the model's limitations and biases in the convective gray zone, rather than as a preliminary setup selection, since the simulation had already been completed before the beginning of this analysis. Furthermore, re-running the continuous 30-year climate simulation down to a 1 km resolution with the CPS configuration was computationally prohibitive due to high-performance computing resource and time limitations. Please see our response to R1C1 and the related changes in the manuscript.**

R2C4: **4. The authors investigate the precipitation also during the days in which it rains in reality. The authors justify this stating that "the assessment focuses on the realism of the simulations in representing precipitation when it rains in the reality" (lines 156-157). However, using such metric, I think that the authors are not evaluating the full performance of the model and that they are biasing their results. For example, such metric is only applicable in this particular case in which some stations are available, but it could not be valuable for alternative regions where the stations are limited. Thus, I would consider removing this metric from the final version and using only**

the one based on the temporal masking, which includes also the days in which the model simulates rain, but it is not observed in reality.

Authors' response: **We understand that evaluating precipitation exclusively on observed wet days could omit model false alarms. However, we would like to clarify that the 'wet masking' analysis was never intended to replace the full performance evaluation, but rather to serve as an additional diagnostic tool. Its purpose is to assess the physical realism of the model strictly when it rains, given the fact that the model produces more wet days than reality and tends to agree with observations when rain occurs. How much it agrees when it rains is what is assessed here. To guarantee an unbiased evaluation, our study additionally presents the results for both the full unmasked analysis and the temporal masking (masked to data existence) analysis. These explicitly account for all days, including those where the model simulates false alarms or misses events. Therefore, we consider that the 'wet masking' analysis does not bias our final conclusions, but rather provides additional information about the physical performance of the model.**

R2C5: **5. Figures 2 and 3. I don't see the point of including ERA5 information as points, as if each grid-point was a station. Why not including an additional panel showing the map from ERA5 data. Additionally, I think the figures would benefit from including some contours at different elevations, such as 1500 m and 2000 m. I suggest those values since they are used in lines 220-226.**

Authors' response: **Concerning the presentation of ERA5 data, our motivation for including its grid points directly onto the WRF fields was to facilitate an immediate comparison within a single and unified map. We think the manuscript already has a considerable number of figures and tried to avoid adding further ones. For instance, Figure 2 helps to show (diamonds) that ERA5 produces more precipitation in the area of the Sierra and allows for local comparisons as altitudes increase with resolution. As resolution increases, the large differences in precipitation are visible. We therefore think that the figure aids in its current form to conduct the analysis. However, if the reviewer maintains that this is essential, we can add an extra column with the ERA5 information separated from the rest. Regarding the topography, we have modified both Figure 2 and Figure 3 to include elevation contours as suggested by the reviewer. Please see the tracked changes document.**

R2C6: **6. Some parts of the text mention that it is difficult to say with the available stations if some precipitation amounts are realistic or not (e.g., line 212). To investigate more on that, why some information from gridded precipitation datasets based on satellite data have not been included in the comparison? I think the manuscript would benefit from a complementary source of information about precipitation.**

Authors' response: **We thank the reviewer for this suggestion regarding gridded satellite-based precipitation datasets. The study domain is approximately 95 km per side, and we utilize a network of 37 observational stations. This configuration yields an average spatial coverage of approximately one station for every 16 km x 16 km**

area, if they were uniformly distributed. Notably, in higher-altitude areas, the station density is even greater than this average (Figure 1 of the manuscript), indicating that the study region is reasonably well-sampled from an observational perspective. However, as we note throughout the manuscript (lines 2, 206, and 212), highly complex orographic zones generate complex, non-homogeneous precipitation patterns that would theoretically require an even higher number of stations for a perfect characterization. In some areas where the simulations show interesting spatial variability, like on the mountain tips, having more observational sites would be beneficial, but those are not available. Although satellite data can be effective for filling spatial gaps in other contexts, their application here is limited by scale and complex topography, given that these products range from 0.05° to 0.25° in resolution and carry high uncertainties regarding localized orographic rainfall. Additionally, they have their own biases that must be corrected in comparison with observations. We think that would require a separate study elsewhere.

R2C7: **7. Figures 4 and 6: If I am not mistaken, these figures show the mean of the annual accumulated precipitation (mean of 30 values since 30 years are available). However, I think that the analysis would benefit from including some kind of information about the variability during the entire period. This information is included in Figure 7, but I think it should be included too in these figures.**

Authors' response: **Figure 4 presents the average percentage of rainy days per year, calculated by counting the total number of days with appreciable precipitation (> 0.2 mm) across the whole temporal period and dividing it by the total number of available data points for each station. Since Fig. 4 is specifically focused on the frequency of wet days, it addresses the probability of rain occurrence in all datasets and illustrates the model bias in this regard. Each station already displays multiple bars divided into solid, hatched, and hollow boxes to evaluate temporal masking, so that adding more layers would probably compromise visual readability. It is noteworthy that the different levels of masking produce consistent results across sites, revealing increments in the percentage of wet days with altitude, comparable model-data percentages if only observational wet days are considered, and overestimated percentages by the WRF model if available observational days or total numbers are considered; all of this being robust across sites. Figure 6 illustrates the daily precipitation absolute and relative frequencies distributed by accumulation thresholds and already shows information about variability through horizontal whisker plots displaying the first, second (median), and third quartiles (Q1, Q2, and Q3). This analysis is extended to all sites in Figure 7, which definitely shows data variability through boxplots with quartiles and extreme percentiles. Taking this into account, we hope the reviewer agrees that the current structure balances comprehensive statistical detail with visual clarity. However, if the reviewer thinks it is very important for the manuscript to represent interannual variability for the Figure 4 analysis, of course we would rethink a different way of representing what was requested without compromising the readability of the current figures.**

Referee's minor comments:

R2C8: **Line 7: the WRF 4.2.2 model -> the WRF model (version 4.2.2)**

Authors' response: **The text has been changed according to the reviewer's indication:**

Line 8, in the tracked changes document: A regional simulation with the WRF ~~4.2.2~~ model (version 4.2.2), driven with boundary conditions provided by the ERA5 reanalysis, was performed.

R2C9: **Lines 70, 142, 413: use the n-dash for the periods stated in the manuscript. Example: 1991–2020**

Authors' response: **All the time periods in the text have been rewritten according to the reviewer's indication.**

R2C10: **Lines 85-87: I would reference only once the Table 1 in the sentence.**

Authors' response: **The text has been changed according to the reviewer's indication.**

Line 86, in the tracked changes document: Observational data (Table 1; Fig. 1) were obtained from two different sources in order to cover a wider altitudinal range. Eight sites belong to the Guadarrama Monitoring Network (GuMNet; Vegas-Cañas et al., 2020; García-Pereira et al., 2024; Fernández-Castillo et al., 2025~~, Table 1 ,Fig. 1~~) and 29 to the Spanish Meteorological Agency (AEMET~~; Table 1 ,Fig. 1~~)

R2C11: **Lines 130-131: I think there is a typo in the labels of the domains. WRF1 should be D1 and the 3 km resolution domain D2.**

Authors' response: **We would like to point out that WRF1, WRF2, and WRF3 represent the three resolutions of the WRF model analyzed in this study over the D3 domain, whereas D1, D2, and D3 refer to the physical simulation domains. The correspondence between both nomenclatures is that WRF1, WRF2, and WRF3 are the three distinct resolution outputs simulated strictly inside the D3 domain, which is the specific region where this study is conducted. For this reason, we consider that the labels referring to the domains should remain as D1, D2, and D3, while those referring to the specific outputs within D3 should be kept as WRF1, WRF2, and WRF3. However, we have modified the text with the purpose of clarifying this point.**

Line 132, in the tracked changes document: The three separate WRF resolutions within the D_3 domain are going to be referred to as WRF1, WRF2, and WRF3 hereafter. These three correspond to the 9 km, 3 km, and 1 km resolution outputs, respectively. ~~The 9 km resolution WRF output inside the D_3 domain is going to be referred to as WRF1 hereafter, the 3 km resolution WRF output inside the D3 domain is going to be identified as WRF2 and similarly WRF3 for the 1 km resolution WRF output.~~

R2C12: **Line 221: appreciable (orange boxes; Fig. 4) precipitation -> appreciable precipitation (orange boxes; Fig. 4)**

Authors' response: **The text has been changed according to the reviewer's indication.**

Line 228, in the tracked changes document: [...]the observed percentage of days with appreciable ~~(orange boxes; Fig. 4)~~ precipitation (orange boxes; Fig. 4) is around 20 %.

R2C13: **Figure 8. The labels of the bottom right panel are too small.**

Authors' response: **Figure 8 has been changed according to the reviewer's indication. Please see the tracked changes document.**

References

Lind, P., Belušić, D., Christensen, O. B., Dobler, A., Kjellström, E., Landgren, O., Lindstedt, D., Matte, D., Pedersen, R. A., Toivonen, E., et al.: Benefits and added value of convection-permitting climate modeling over Fenno-Scandinavia, Clim. Dyn., 55, 1893–1912, https://doi.org/10.1007/s00382-020-05359-3, 2020.

Mansell, E. R., Ziegler, C. L., and Bruning, E. C.: Simulated electrification of a small thunderstorm with two-moment bulk microphysics, J. Atmos. Sci., 67, 171–194, https://doi.org/10.1175/2009JAS2965.1, 2010.

Tiedtke, M.: Representation of Clouds in Large-Scale Models, Mon. Weather Rev., 121, 3040–3061, https://doi.org/10.1175/1520-0493(1993)121<3040:ROCILS>2.0.CO;2, 1993.

Tomassini, L., Willett, M., Sellar, A., Lock, A., Walters, D., Whitall, M., Sanchez, C., Heming, J., Earnshaw, P., Rodriguez, J. M., et al.: Confronting the convective gray zone in the global configuration of the Met Office Unified Model, J. Adv. Model. Earth Syst., 15, https://doi.org/10.1029/2022MS003418, 2023.

Vergara-Temprado, J., Ban, N., Panosetti, D., Schlemmer, L., and Schär, C.: Climate models permit convection at much coarser resolutions than previously considered, J. Clim., 33, 1915–1933, https://doi.org/10.1175/JCLI-D-19-0286.1, 2020.

Zhang, C. and Wang, Y.: Projected Future Changes of Tropical Cyclone Activity over the Western North and South Pacific in a 20-km-Mesh Regional Climate Model, J. Clim., 30, 5923–5941, https://doi.org/10.1175/JCLI-D-16-0597.1, 2017.