Manuscript: egusphere-2025-1339

Before responding to the reviewers, we would like to clarify, as stated at the end of the abstract, that the so-called reinitialized method requires approximately 30 times less computational time than the continuous method in this study, while delivering fully comparable accuracy. This dramatic gain in efficiency strongly supports recommending it as the preferred approach, as its performance—considering both accuracy and computational efficiency—surpasses that of the continuous method.

At first sight, this finding may be surprising for researchers experienced with the continuous method for downscaling processes. Indeed, it was initially surprising for the authors as well. We conducted an extensive internal debate, exploring scenarios in which the continuous method could potentially provide better results (e.g., the appearance of discontinuities in the reinitialized method or atmospheric variables with a long memory). The results presented in this manuscript, which clearly demonstrate the superior performance of the reinitialized method, are the outcome of this thorough debate and detailed analysis of multiple metrics.

In our view, researchers have the right to know that the most widely used method is not necessarily the most optimal in terms of computational time (i.e., the real time required to obtain the final results), which can significantly accelerate their simulations. Finally, our findings are particularly relevant for mid-latitudes, where frontal systems and extratropical cyclones evolve rapidly, though they may not be directly applicable to other latitudes. Therefore, scientists modeling different regions should carefully assess the validity of our approach before applying it.

Reviewer #1

This manuscript investigates the performance of continuous versus daily reinitialized dynamical downscaling using the Weather Research and Forecasting (WRF) model. The simulations are driven by ERA5 and CMIP6 data over a large domain covering a wide area in the mid-latitudes of the Northern Hemisphere and are downscaled to 20-km resolution. The study compares both downscaling approaches in terms of their ability to reproduce key atmospheric variables such as wind speed, temperature, humidity, precipitation, surface pressure, and solar radiation, with ERA5 data used as the evaluation reference. The authors aim to assess whether the reinitialized approach—despite its lower computational cost—can produce results comparable to the more resource-intensive continuous simulations. Overall, while the manuscript addresses a relevant topic within regional climate modeling, the current version lacks sufficient novelty and depth of analysis to warrant publication. Key limitations in methodology and interpretation are not adequately addressed, and the conclusions are primarily descriptive without offering new insights into the physical processes or broader implications. Therefore, I do not recommend the manuscript for publication in its current form. To support this recommendation, I outline the main concerns as follows:

1. The methodology section lacks novelty and does not present a sufficiently innovative approach to advance the field. The study primarily compares continuous and daily reinitialized dynamical downscaling using the WRF model. The comparison between these two approaches is meaningful, but it does not offer any new insights in methodology. In addition, the experimental design is fairly conventional and lacks a thorough exploration of different experimental setups or configurations. For instance, the study does not investigate the impact of varying reinitialization intervals. Without a more creative or refined experimental design, the manuscript does not sufficiently push the boundaries of current understanding in regional climate modeling.

   - *We appreciate the reviewer's thoughtful comments. While we agree that our study does not introduce a new modeling technique or aim to push the methodological* *boundaries of regional climate modeling, we would like to clarify the originality and scope of our work. While previous studies have addressed similar comparisons between downscaling techniques (Pan et al., 1999; Qian et al., 2003; Lo et al., 2008; Carvalho et al., 2012; Lucas-Picher et al., 2013; Liu et al., 2018), to our knowledge* ***this is the first study*** *to simultaneously compare continuous and daily reinitialized WRF downscaling over such a large spatial domain, long temporal period, and with such a broad set of atmospheric variables and performance metrics. Additionally, our work is among the first to perform this evaluation using both a reanalysis dataset (ERA5) and a global climate model projection from CMIP6 as boundary conditions, providing a unique opportunity to assess the consistency of results across historical and future climate contexts. These aspects were added at the end of the introduction section in the revised version of the manuscript (lines 106-110).*

- *We recognize that only two downscaling approaches were tested. However, these two methods represent the key paradigms used in dynamical downscaling: the continuous method, which is conventional and widely adopted, and the daily reinitialized method, which, while less commonly applied, is significantly faster and increasingly considered for operational and resource-constrained applications. Exploring additional reinitialization intervals (e.g., weekly or monthly) can be of interest, but such configurations have already been addressed in studies like Lo et al. (2008), and are not the focus here, given the already substantial length of the manuscript. Our goal was not to develop a novel technique or experimental setup, but to deliver a robust and large-scale comparison of two realistic and widely relevant downscaling strategies. We believe that this fills an important gap in the literature as it proves that the reinitialized dynamical downscaling technique drastically reduces simulation time (by 30 times in this study compared to the continuous method) while showing no losses in data quality. Therefore, we believe that while the methodology of this paper is not new, its goals and conclusions are.*

2. The manuscript provides a general description of the results but fails to explore the underlying physical mechanisms behind the observed differences between the two downscaling techniques. While the study acknowledges that neither method reliably captures wind speed and surface pressure in complex terrain, particularly in mountainous regions, it does not sufficiently address the physical processes that contribute to these limitations. Without addressing these physical mechanisms involved, the study does not offer a thorough understanding of the factors influencing the performance of the downscaling techniques.

    - *Indeed, both the continuous and reinitialized WRF simulations show limited skill in reproducing wind speed and surface pressure over complex terrain, with no significant difference between the two approaches. We agree with the reviewer that analysing the physical mechanisms involved would enhance the quality of the study, therefore such analysis has been added at the end of sections 3.1.1.1 (lines 345-352) and 3.1.1.5 (lines 568-574). Though, the main goal of this manuscript is to demonstrate to researchers involved in downscaling processes the potential of the less frequently used reinitialized dynamical downscaling method, which offers substantial computational efficiency without losing accuracy.*

3. The manuscript presents the analysis of various atmospheric variables, but these variables such as wind speed, temperature, humidity, and precipitation are analyzed independently, without considering their interdependencies. In regional climate modeling, it is crucial to explore the interactions between these variables, as they often exhibit complex relationships that can impact model accuracy. The absence of such an integrated analysis limits the depth of the study's findings and fails to offer a better understanding of the underlying climate processes.

- *The primary aim of this study is not to gain a deeper understanding of the physical processes occurring during dynamical downscaling, but rather to evaluate the model's performance. Specifically, the objective is to provide climate modelers with an independent assessment of the quality of the downscaled atmospheric data, and we actually demonstrated that the reinitialized dynamical downscaling technique cuts simulation time dramatically—by a factor of 30 in this study—without compromising data quality. Nevertheless, we acknowledge the reviewer's valuable comment regarding variable interdependency. In response, we have conducted an additional analysis (which methodology and results can be found in new sections 2.4 and 3.1.2, respectively) to explore these interactions. We focused on extreme precipitation events characterized by a 24-hour period of light rainfall immediately followed by a 24-hour period of very heavy rainfall. Locations and instances of such events were identified simultaneously in the ERA5 dataset, as well as in the outputs of both continuous and reinitialized WRF simulations forced with the same reanalysis data. We then examined the behavior of other atmospheric variables between the two 24-hour periods, analyzing their relative differences. The results show that both WRF downscaling approaches exhibit patterns and magnitudes comparable to those found in ERA5. Specifically, during these periods of increased precipitation, wind speed, humidity, and IVT tend to increase, while temperature, pressure, and solar radiation decrease.*

Besides, there are several minor issues that need to be addressed.

- Lines 17-18: "115ºW-40ºE in longitude and 20ºN-60ºN latitude" should be "115ºW-40ºE in longitude and 20ºN-60ºN in latitude".

  This has been modified in the revised version of the manuscript.

- Lines 99 and 114: "from -115ºW to 40ºE" should be "from 115ºW to 40ºE" or "from -115ºE to 40ºE".

  This has been modified in the revised version of the manuscript.

- In Figure 2, the latitude and longitude labels are too small and difficult to read. I recommend increasing the font size to improve readability and overall presentation quality.

  In Figures 2, 4, 6, 8, 10, 12 and 14, it is true that longitude and latitude labels are difficult to read, but we would rather not increase their size since it would reduce the portion of the figures containing actual data. Nevertheless, it has been increased in Figure 1, representing the domain of study, for better clarity.

- Figures 3, 5, 7, 11, 13 and 15: Please check the figure legend — it seems that "renitialized" is a typo, and it should be corrected to "reinitialized".

  Indeed, this was an error that has been corrected in the revised version of the manuscript.

Reviewer #2

This study compares two dynamical downscaling techniques—continuous and daily reinitialized simulations—using the Weather Research and Forecasting (WRF) model. The simulations were driven by ERA5 (upscaled data 1 every 4) and CMIP6 datasets and covered a broad domain. The downscaled data were evaluated against high-resolution ERA5 (all points) data to assess accuracy in key climate variables. The authors present that both methods showed good to excellent agreement with the reference data overall but for wind speed or surface pressure in mountainous areas. From these results the authors state that given that the reinitialized method requires significantly less computational time, approximately 30 times less in this study, it is recommended as the preferred approach since its performance is comparable to or better than that of continuous downscaling.

While the manuscript covers an important subject in regional climate modeling, it does not present enough innovation or thorough analysis in its current form. Critical issues related to the methodology and interpretation of results are not adequately explored, which weakens its overall scientific value. Therefore, I cannot recommend this paper for publication and it should be rejected. Here I expose some reasons.

From the computational point of view, the authors repeatedly fail to clearly distinguish between computational cost and wall-clock time. While the reinitialization approach (i.e., time-parallel integration) can reduce wall time if sufficient computational resources are available, it is inherently more expensive in terms of total computational cost. This is due to the increased number of model spin-ups and the lack of continuity in the simulations. The manuscript should clarify this distinction, as it is critical for properly assessing the efficiency and scalability of the downscaling methods. Note that if, for example, NN CPUs are available and the model configuration allows for their efficient use in parallel (i.e., with minimal loss in scalability), then the computational cost of a continuous simulation would be significantly lower. In such cases, running a single continuous simulation can be more efficient than multiple reinitialized runs, both in terms of total CPU time and resource utilization. This highlights the importance of considering scalability and resource availability when evaluating the computational efficiency of each approach. The message conveyed by the authors may be misleading or even problematic for less experienced users. Without a clear distinction between computational cost and wall-clock time, and without properly addressing the implications of model scalability and resource allocation, there is a risk that users may draw incorrect conclusions about the efficiency of the reinitialized approach. This could lead to inefficient use of computational resources or misinformed methodological choices in future studies.

- *We agree with the reviewer that the manuscript did not clearly distinguish between simulation time (wall-clock time) and computational cost. Indeed, while the reinitialized approach can reduce wall-clock time by using multiple cores to run several independent simulations in parallel, it inherently increases the total computational cost because each run requires its own spin-up, whereas only one is needed in a continuous simulation. These clarifications have been incorporated into the revised manuscript, particularly in the Introduction (lines 64-67), Section 2.2 (lines 171-174), and the Conclusion (lines 768-769).*

- *We also acknowledge that, in an ideal scenario such as the one described by the reviewer—where the model configuration enables efficient scaling across many cores and sufficient computational resources are available—a continuous simulation would indeed be both faster and computationally cheaper, since it requires only one spin-up and avoids overlap. However, in practice—and this is precisely the situation our study aims to assess—such conditions are rarely achievable. Users often lack access to the number of cores required to perform long continuous simulations in less time than multiple shorter, reinitialized simulations distributed across available* *resources. It is generally not feasible to perform long, continuous simulations within a reasonable timeframe since the use of all available CPUs is generally restricted, and the queue management system imposes considerable waiting time for long simulations.*

On the other hand, the authors seem to overlook the fundamental purpose of dynamical downscaling. If the goal is merely to enhance spatial resolution, the time-slice integration approach can be useful, as it acts like a form of pseudo-nudging—allowing the incorporation of higher spatial detail with physical meaning, particularly in regions with complex topography. However, this method can introduce discontinuities in the time series, and its application should be approached with caution, depending on the physical processes being studied. For instance, when used for regional climate change projections, insufficient spin-up time may lead to physical inconsistencies, undermining the reliability of long-term simulations.

- *We appreciate the reviewer's insightful comment. We agree that the time-slice integration approach may introduce discontinuities in the time series if insufficient spin-up time is provided. To address this concern, we evaluated whether a 12-hour spin-up is sufficient for ensuring consistency in daily reinitialized simulations. Specifically, we compared the values of key variables at 00h on day D with those at 18h on day D–1, which originate from different runs in the reinitialized approach. Our analysis shows that the differences between reinitialized and continuous simulations are comparable to those observed in ERA5, indicating that the chosen spin-up period is adequate to avoid inconsistencies and discontinuities. The description of the methodology has been included in new section 2.5 of the revised manuscript, and the results of this evaluation in new section 3.1.3.*

If I have understood correctly, the authors compare the output of GCM-driven simulations with reanalysis data just for one year. This is a serious methodological flaw. Reanalysis data represent an approximation to observed atmospheric states, while GCM-driven simulations for a given year has nothing to do the reality. It is like compare one year to another. Comparing them directly, without proper bias correction and in a climatological context, is not scientifically valid and can lead to misleading conclusions.

- *We strongly believe our methodology is scientifically valid. Indeed, what the reviewer describes is only the case of section 3.2, where the outputs of CMIP6-driven simulations are compared to reanalysis data (ERA5). We agree with the reviewer that both datasets could not be compared directly, that is why no instantaneous metric of comparison (RMSE, CC, STD) was used for this analysis, but only the OP which is a climatologic one. Indeed, it only compares the distribution of the variables' values without taking into account their temporal consistency. Nevertheless, section 3.1 compares the output of ERA5-driven simulation to the same ERA5 data. In this case, instantaneous metrics of comparison were also used since both datasets are reanalysis. The end of the Introduction (lines 98-103) has been rewritten in order to avoid all possible confusion.*

- *Lastly, only the year 2014 was modelled but it is considered as a typical year, as the major regional variability patterns, such as the North Atlantic Oscillation and El Niño–Southern Oscillation, did not show pronounced phases (Fernández-Alvarez et al., 2023a).*

The paper is unnecessary long, specially the results section. Probably all results could be reduced to a couple of tables and panels.

- *We agree that the paper is long, but we believe all results shown are of importance. Indeed, they strengthen the conclusion that reinitialized and continuous downscaling* *show similar results, both in instantaneous values and climatologically, and both when using reanalysis or GCM data to perform the dynamical downscaling. While tables and Taylor diagrams show results averaged according to the soil type (over sea or inland), maps also permit to give local information about where one method performs better than the other, as well as for which variables. Nevertheless, we decided to remove figures 16 to 19 from the manuscript since they showed information similar to Table 9.*

The physical interpretation is largely absent, as the authors do not provide any explanation of the underlying physical processes responsible for the differences observed between the downscaling approaches or physical consistence of the atmosphere and surface (for example) fields.

- *While this is not the main point of this study, we agree with the reviewer that no major physical interpretation was present in the manuscript. Therefore, in accordance with another reviewer's comment, we have added explanations about the poor modeling of surface wind speed (lines 345–352) and pressure (lines 568–574) in the revised version of the manuscript. Additionally, an analysis of the physical interdependency between atmospheric variables was performed, with methodology described in new section 2.4 and results in new section 3.1.2. Further detailed analysis would be of interest, nevertheless this is not the aim of the present study and the manuscript already shows a consequent length.*

The experimental design is basic and I would say wrong. For example, the authors use a very big domain. Later they use just a part (data above 20°N was selected) because they think the results are worse in tropical-equatorial regions. Also, the skill scores used are not the best for checking one year of data.

- *Actually, the objective has always been to study the area above 20°N and we knew that the chosen parameterization performed optimally in these subtropical and extratropical regions (Insua-Costa & Miguez-Macho, 2018; Insua-Costa et al., 2019). Therefore, we deliberately positioned the southern boundary of the WRF domain far from the area of interest. While it could have been placed closer to the Equator, we opted for a more conservative approach to ensure reliability. This has been clarified in the section 2.1 (lines 126-127) of the revised manuscript.*

- *Regarding the skill scores, we employed a comprehensive set of widely used metrics, including instantaneous measures (NRMSE, CC, STD) as well as a climatological one (OP). We believe that, even with one year of data, these metrics are appropriate and provide a robust assessment of model performance.*