Dear editor,

We are glad to resubmit the article titled "Present and future synoptic circulation patterns associated with cold and snowy spells over Italy" to be considered for publication on Earth System Dynamics.

We took several actions to improve the paper according to the suggestions of the reviewers. We are confident that the changes we made satisfy the requirements of the two reviewers, and that they have actually improved both the scientific quality of the results and the cogency of the article.

Best regards
Flavio Pons & Davide Faranda on behalf of all the authors

Reviewer 1

Review comments on esd-2020-61 "Present and future synoptic circulation patterns associated with cold and snowy spells over Italy", version 30 Nov 2021, by M. D' Errico and co-authors

GENERAL COMMENTS
This is my third review of this manuscript. Again, there have been extensive changes that solve some of the previous problems but also raise new questions and concerns. I particularly appreciate the condensation of the previous case-by-case results to the cluster-wise analysis in Table 1. The way in which the results are represented in this table is useful, because it shows that the large relative increases in the frequency of the circulation analogues in fact reflect relatively modest increases in the absolute numbers.
As the authors now apply their analogy search to anomalies in Z500, which I presume to have been calculated against the time means in the RCP simulations, the increase in the frequency of good analogies apparently cannot be explained by biases or changes in the time mean state. Yet, the manuscript still leaves it unclear to me why this increase takes place. I realize that it may be difficult to give a fully satisfying answer to this question. However, one way could be to pick up a fixed number (e.g. top 2%) of the best analogies for both the CTRL and the RCP simulations, and calculate the local root-mean-square difference between these and the cluster centroids for each grid point in the analysis domain (22.5º-70ºN, 10ºW-70ºE). Doing this and comparing the resulting maps, it would at least be possible to deduce whether the analogies in the RCP simulations are better in a specific part of the domain or whether the improvement is more or less evenly distributed in the area.

We thank the reviewer for the suggestion. This analysis is now included in the new version of the manuscript.

In this version of the manuscript, bias corrections have been introduced that adjust both the time mean bias and the standard deviation. The correction for the mean bias seems redundant as far as the analogies are based on anomaly values only. There is also a risk that the correction of the standard deviation makes more harm than good. If the standard deviation biases in PlaSim are different in different parts of the domain, then the correction of the standard deviation may result in anomaly patterns that differ

(e.g., in the locations of their maxima and minima) from the originally simulated patterns. In such a situation, the dynamical interpretation of the results becomes difficult.
What made me pay attention to such potential bias correction artefacts is Fig. 4g-h. These maps show, for the RCP8.5 scenario, Z500 height fields with a strong (> 6000 m?) maximum in the northernmost part of the domain. This is completely different from the patterns in the CTRL (Figs. 4a-b). Moreover, as there is a deep minimum in SLP in the same area in Figs. 6g-h, such a maximum in Z500 would only be consistent with the hydrostatic balance if the temperatures in the northern part of the domain were much higher than those elsewhere. However, Fig. 7g-h shows that, even for RCP8.5, T850 is lower in the northern than the southern part of the domain. Thus, Z500, SLP and T850 and SLP are physically inconsistent with each other. Such physical inconsistences could result if the correction of standard deviation is also affecting the climate change signal and not only the variability around the mean state. Therefore, my advice is to repeat the analysis without the standard deviation correction, or at least modify the implementation of this correction so that it has no impact on the time mean state in the RCP simulations.

We thank the reviewer for these insightful comments. We agree about the - previously overlooked by us - physical inconsistency of the results obtained with BC data. Moreover, bias correction is not usually performed on variables such as SLP and Z500. We do not perform any BC on the two dynamical observables in the current version of the article.
However, we still perform linear transform BC on near-surface and 850 hPa temperatures, and a quantile mapping BC on precipitation, as these variables must be unbiased to correctly evaluate the likelihood of cold and snowy conditions under different forcing scenarios.

The second main weakness in the revised manuscript concerns the new Section 4 on the CMIP5 simulations. This seems like a last-minute addition since it is mentioned neither in the abstract nor the concluding section in the manuscript. More importantly, the description of what was done is so brief that it is difficult to understand what the results mean.
On L283-285, it is stated that the (observed?) daily Z500 anomalies for the 32 cold spell events were embedded into the historical simulations for 1951-2000 and the RCP4.5 and RCP8.5 simulations for 2051-2100. Does this mean that you only used the time mean Z500 fields in the CMIP5 simulations, and then generated Z500 fields corresponding to the cold-spell cases by adding the observed anomalies to these time mean fields? This sounds rather different from what you did with PlaSim, for which the analysis focused on a comparison of the simulated and observed anomalies. Or did you in fact use the daily Z500 data in the CMIP5 models and repeat the analysis done for PlaSim (which would require a major effort)?
The follow-up on the two types of bias correction on L291-296 is equally difficult to grasp. First, “bias correction” is a somewhat misleading word for removing the spatially averaged Z500 trend (which I assume to mean the area and time mean difference in Z500 between 2051-2100 and 1951-2000). Second, what does the bias correction do if the spatial mean trend is preserved – do you also correct for the biases in the local 1951-2000 mean values? In either case, it is intractable to me how you constructed the anomalies that were compared with the observed (cluster 1 or cluster 2 mean?) anomalies in the end.
Please add detail to the description of the CMIP5 analysis. Otherwise, the results in Section 4 are impossible to interpret. Also, you should mention this analysis both in the abstract and in Section 5, as it appears to be an important part of your work.
Because of these reasons, the manuscript still requires major revisions. Further

comments on the details are given below.

We agree with the reviewer that the analysis of the CMIP5 models would deserve more information. After discussing with the coauthors and in light of the referees' comments, we decided to remove this analysis from the present manuscript and to postpone its publication in another work that will be dedicated to forced CMIP5 and CMIP6 analyses. The main reason is that we cannot really perform on forced simulations the same procedure applied for PlaSim simulations with stationary forcing.

SPECIFIC COMMENTS

1. L6-8. The CMIP5 analysis should also be mentioned in the abstract.

The CMIP5 analysis has been removed from the manuscript (see below)

2. L27-40. Consider also citing O'Gorman (2014): https://www.nature.com/articles/nature13625

We thank the reviewer for the suggestion of this paper that we had somehow overlooked, we added a citation of their main result.

3. The legend in 1 indicates that the size is proportional to the duration. Does this indicate that the diameter of the circles if proportional to the duration, or that their area is proportional to the duration?

The duration is proportional to the diameter, as now indicated in the caption of the figure.

4. Figures 3a-d. Please include the temperature anomalies in the maps. You can add them as contour plots and simultaneously retain the shading for the absolute values.

Thank you for the suggestion, we now present absolute values in color shading and standardized anomalies as isolines.

5. L175. A better general reference to the RCP scenarios: van Vuuren, D.P., Edmonds, J., Kainuma, M. et al. The representative concentration pathways: an overview. Climatic Change 109, 5 (2011). https://doi.org/10.1007/s10584-011-0148-z

Thank you, we replaced the previous citation with your suggestion.

6. L177-178. The actual CO 2 concentrations in the RCP scenarios are lower. The quoted numbers are equivalent CO 2 concentrations, which include the net effect of all anthropogenic greenhouse gas forcing.

We thank the reviewer for the observation. Indeed, PlaSim simulations are based on equivalent CO2 concentrations including the net effect of all GHG. We now specify this in the indicated lines, and we use the expression "equivalent CO2 concentration" throughout the article.

7. Table 1, cluster 2, RCP85. +71.3% (in parentheses) appears too small. Should rather be ~121%?

Thank you, indeed there was a typo. However, all values changed due to the differences in our data analysis strategy.

8. Caption of Table 1. Please specify the meaning for the numbers that are given for the RCP scenarios (i) before the parentheses (e.g. 0.978) and (ii) in the parentheses (e.g. 0.977).

Thank you, we proceeded to specify this detail in the caption.

9. L283-285. Did I understand this correctly? You simply take the 50-year time mean of Z500 in the CMIP5 simulations and add the observed anomalies in the 32 cold spell cases?
10. L284-285. What was the time resolution of the CMIP5 data? Daily or monthly? Overall, Section 4 is difficult to follow because the methodology is not explained in sufficient detail.
11. L297. Euclidean distance relative to the mean of the cluster (1 or 2) in which the event belongs to?
12. L299. Why does the inclusion of the spatially averaged trend lead to closer analogies? One would expect that, with a general warming of climate, the Z500 anomalies relative to the 1951-2000 mean will become systematically positive. If anything, this should make the analogues worse.
13. Section 5. The CMIP5 analysis should also be mentioned in this section.

After discussing with the coauthors and in light of the referees' comments, we decided to remove this analysis from the present manuscript and to postpone its publication in another work that will be dedicated to forced CMIP5 and CMIP6 analyses. The main reason is that we cannot really perform on forced simulations the same procedure applied for PlaSim simulations with stationary forcing.

TECHNICAL COMMENTS
1. L203. to the control and RCP simulations?
2. L214. persistence?
3. L216. cold spells
4. L254. these results
5. L292. that trend / those trends
6. L297. more slightly or more strongly?
7. L298. these decreases
8. Caption of Figure 10, L1-2. each dot represents
9. Caption of Figure 10, L2. given
10. L324. requirements of the Paris agreement (Arias et al. 2021).

Reviewer 2

Comments on the re-revised version of Present and future synoptic circulation patterns associated with cold and snowy spells over Italy, by M. d'Errico et al.

January 27, 2022

Remarks
The authors have taken many of the earlier comments by the reviewers seriously. Again the paper has seen big changes. The SST-run has disappeared, and new scenario runs have appeared. Some figures are gone, many new have appeared. Section 4 is new material, showing some analysis with CMIP5 data. I appreciate the extra effort but have difficulties to understand them. I think most of the changes work out well and the paper has improved. However, I still have several points/remarks. My suggestion to the editor is that the paper still requires some revisions prior to acceptance.
Points (mostly remarks to the authors' response document)

1. My previous point 3 ("too large size of the clustering domain") has not been used. Instead, arguments are given why the authors think their large domain (120degree wide in longitude) is necessary. I can follow their arguments, but still hesitate as to its justification, especially if the final target area is that small. Large-scale embedding is essential of course, but in my view the large domain now leads to flow fields not tied very strongly to the region of interest. This implies that the results are perhaps less clear than would otherwise have been possible, in the sense that for the target region, the response fields likely regress to the mean. But in the end, it is the responsibility of the authors to squeeze the best possible results out of the data. My suggestion would be to include a brief discussion on how experimenting with domain-size would impact the results. This would also help to address issues raised below under 4 of my minor points.

We thank the reviewer for the observation. Initially, we used a large domain to find the clusters, to get an idea of the large-scale configurations leading to cold spells over Italy, but we considered the smaller domain sufficient to detect analogues producing similar phenomena over the region. As suggested, we repeated our analysis using the same domain for clustering and analogue search, while implementing requests from the other reviewer, along with the correction of issues we detected while revising our work. We find that, indeed, using the same domain is more appropriate, but results are clearer using the larger domain for both analyses, rather than the smaller one. In particular, we find that Z500 analogues found with the smaller domain produce SLP and temperature composites that are less compatible with the cluster centroids compared to the composites of analogues found using the larger domain.

2. I can agree to the authors' reply to my Point 5 ("significant differences"). If one can argue that the flow is different outside the target area, this should also be sufficient. as it points to different types of air masses involved.

We are happy that the reviewer agrees with our view.

3. It is appreciated that the figures have been improved. Yet I have some further suggestions (see below).

We have taken the further suggestions on figures into account

4. On my point 8 ("alarming results of PlaSim"): I am pleased that an error was discovered and corrected for and am happy with the authors' response and putting these results into a better context of existing results/debate. Still the huge increase in frequency is difficult to grasp.

There are now few new studies that confirm important modifications in the atmospheric circulation patterns with climate change. We have added the following discussion: "This increased frequency of Atlantic Ridge and Scandinavian blocking patterns in PlaSim simulations and warming scenarios could be associated to a wealth of phenomena driven by the mean anthropogenic climate change but still debated in the current scientific literature such as the Arctic Amplification or the increased land-sea temperature contrast (Cohen et al., 2020; Hamouda et al., 2021). Contrary arguments show that there is an increase in flow zonality over the North Atlantic but mostly for the Autumn (de Vries et al., 2013) and the summer seasons (Fabiano et al., 2021)."

5. On my point 9 ("alternative suggestion"). I am pleased to see that the authors have followed up my suggestion.

Thank you

6. On my point 11 ("4K ocean"). I think it is good that the authors removed that part in the re-revised version. The additional simulations with different scenarios are a good replacement.

Other comments (mostly minor, except 4).

1. In Section 2.1 (line 53) the Delta-SST-run is mentioned. I think it is not used anymore (see my above point 6). This sentence can be removed.

Thank you, we removed the erroneously retained mention of the 4SST run.

2. Cluster 2. The authors refer to it (sec 2.2, line 124) as a "Scandinavian" blocking. However, when I look at the MSLP pattern, it doesn't look quite like a Scandinavian blocking, as the high pressure is much too far south. Maybe the authors could rephrase it into "resembles/is more like a Scandinavian blocking pattern". It could help to show MSLP also as an anomaly. Generally, I think also Figure 3 would benefit from taking an anomaly perspective, with fewer colors. The color-scale for precipitation is not well chosen (I'd use the conventional BrBG option in an (relative)-anomaly framework, again with fewer colors).

We thank the reviewer for the suggestion. We no longer refer to the configuration as "Scandinavian blocking". We changed the colorscale for precipitation and, also following the suggestions of the other reviewer, we now show full fields as color shading and standardized anomalies as isolines.

3. Table 1: the results for Cluster 2 for RCP8.5 (bottom-right numbers) seem inconsistent. The number between brackets should be 53.4% higher than the number outside the bracket (or, likely, the number outside the bracket should be 53.4% lower than the one inside). Can I assume this is a typo?

We thank the reviewer for the observation. Indeed, this was a typo. Since we now use a larger domain for all analyses and other adjustments have been made, while the general tendency to an increase in frequency is preserved, absolute numbers are overall different from the previous version.

4. Figure 4: It is not instructive to start with this figure, as you mainly see the

increase of the Z500 level with increased warming levels. Figure 5 is the key figure. It is strikingly similar under all 4 scenarios (as they should I think, since you based your analogs on these patterns). I first wondered why these patterns do not agree to the ones obtained from observations... Similarly for figure 6: These MSLP fields don't look at all like the patterns obtained earlier based on the observations (Figure 2e-f)? Based on these figures I would call neither of them Scandinavian blocking. The most pronounced features in Z500 (fig 5) are the low's rather than the highs (The highs are far away over the Atlantic). The reason that the patterns are different of course is that you now search for analogues in a much smaller domain, more connected to the flow in the region of the target region. I do appreciate this shift, but it somehow conflicts with your earlier arguments (of choosing a wide domain for the clustering). Why did you first construct cluster centroids for a larger domain (making them not very different in the target area) but subsequently use a smaller domain for the analogues? I would rather use the same domains for both exercises. The current choice makes the story much more complex than needed.

Thank you for the suggestions. After exploring both possibilities, we decided that it is appropriate to use the same domain, but we chose to maintain the larger one.

5. Figure 7-9: Again, using an anomaly framework would be much more instructive. Now, upon visual inspection one can see hardly any difference between Cluster 1 and 2.

6. Section 4: I cannot really follow what is being done here based on the information presented. It is stated that the "We then embed these observed events into historical simulations...". What does this mean? Do you project all days in CMIP5 for each winter and model on the observed cold spells, and compute their Euclidean distance? If so, why would a average reduction of the distance between the observed (cold-spell) and modelled (CMIP5) circulation be an indication that the cold-spells become more frequent? Or do you perhaps select cold days in some way from CMIP5 first? In its current formulation I cannot see how the information in Figure 10 helps us.

This section has been removed, as we decided to keep the analysis of full complexity models, possibly CMIP6, for a more detailed future work.