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 and beyond the suggestions of the reviewers. In particular, we produced new climate simulations to correct an issue in a physical parameterization in the PlaSim model, and we took a chance to add two more RCP scenarios, including one that is representative of the current climate targets according to Paris agreements. Soulivanh Thao has taken care of performing the simulations and he has been added to the authors'list.

We now base our analysis on an anomaly framework rather than on full atmospheric fields, as recommended by both reviewers. To improve readability, we moved the detailed description of the phenomena associated to each cold spell of interest to an appendix, and we changed all the figures, adopting a custom colorscale for better graphical clarity.

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

Comments on the revised version of Present and future synoptic circulation patterns associated with cold and snowy spells over Italy. By Miriam D'Errico1,*, Flavio Pons1,*, Pascal Yiou1, Cesare Nardini2, Frank Lunkeit3, and Davide Faranda

Remarks
The authors have taken many of the earlier comments by the reviewers seriously. The paper has improved considerably but also has seen quite a big change compared to the previous version. I am pleased with some of these, yet I have still difficulties understanding the results. Or perhaps I should say, new difficulties, different from before, because there is a lot of new material. New are a K-means clustering approach, and a focus on the frequency changes of the mslp analogs. Although there is potential in this new part, I have a number of concerns and remarks that may influence results and require new analysis. My main concern is that the most important new result (the "huge" increase in RCP85 cold-spell analog mslp conditions, at least for certain of the cases) is hard to digest without any suggestions as to the why of it. The paper does not offer any help with explanations. The same holds for the almost complete absence of effect on frequency, for the SST+4 runs where PLASIM global oceans are increased by 4

degrees.

Without a more proper interpretation of these results I cannot accept this paper. My recommendation based on this version, is revise with major revisions.

I list some of my main comments below.

1. Section 1 and 2 have not changed much. There still is a multipage long descriptive section 2 on the cases, making rather clear that they are quite different. To me providing the entire list with details on all cases is way too much given the amount of analysis that is undertaken subsequently. I leave the decision up to the editor, but I would be happy to see (some/most of it) put in an appendix.

We thank the reviewer and we agree with them, the detailed description of the cold spells has been moved to an Appendix.

2. To make some order in the chaos of all cases, the authors decided to conduct a K-means clustering analysis. This could be a useful thing sometimes indeed. They end up with two main clusters. However, there is no real argumentation for this. Figure 2, the scree plot, is poorly formatted with labels dropping off. It also doesn't tell anything as far as I can see, except that there is no favourable grouping. I would put this in supplementary material, but definitely tidy up the graphics!

We agree with the reviewer that the scree plot is poorly informative. As now pointed out in the text, this is probably due to the fact that gridded atmospheric fields are not an ideal type of dataset to satisfy technical requirements for k-means clustering, and in our case the cluster sizes also differ. However, k-means is also an established way to find weather regimes. In the new version, we remove the scree plot since it is not informative. The ratio of choosing 2 clusters comes from the fact that we tried with 3 and 4 groups, and in this case we obtain redundant clusters containing essentially the same configuration. We specify this in the text.

3. I think the domain chosen for the clusters is *way* too large. Although the authors warn the reader that they do this for a reason, the cluster domain now covers 120 degrees in the zonal direction, which is a 3rd of the earth. Have the authors experimented with using a domain that is more compact, to zoom in slightly more on the actual situation over Italy? Although the subsequent PLASim simulations are of course also rather coarse I think it would help make the analysis more relevant useful.

We decided to include this domain because the patterns influencing cold spell weather at the European level have a larger scale, and it would be more difficult to discriminate between them. From the maps in Fig. 3 it is evident that, while the position of the high and low pressures over Europe changes between clusters, the biggest difference is found between the Atlantic and North America, with a basically inverted position of positive and negative geopotential anomalies. Including this

portion of the domain helps better discriminate between the two weather regimes associated to Mediterranean cold spells.

4. Another basic question, has the clustering be performed on anomalies wrt to a climatology or to the full fields? Because of pre-existing large-scale pressure gradients, a full field framework is not recommended.

Thank you for the suggestion, it is now clarified in the text, we used standardized anomalies respect to the DJFM climatology of the historical period.

5. To augment my previous statement, it could help the authors to examine whether the differences in mslp between the cluster centroids are actually statistically significant over the prime region of interest: Italy. If not the authors have a problem with section 2.3.

We thank the reviewer for this suggestion; we have now given an extended description of the dynamical differences between the two clusters that is especially visible in the Z500 standardized anomalies and the SLP patterns.

More in general, we do not agree that a statistical test over the region is necessary nor sufficient to explain differences between the two regimes. Even in the case the SLP and Z500 anomalies over Italy were not significantly different, it is the position of the coupled high pressure that drives cold air over different paths (the Rhone Valley in the Atlantic ridge case, or from Russia with a NE-SW direction on the Scandinavian blocking case).

6. The quality of figures 3-5 is poor and does hardly provide insight in the way they are presented now. They should be improved. My advice is to combine mslp and T850 in the same plot (or maybe even T2M as well), using shading and contours. And use an anomaly framework! So make these plot wrt a DJF or whatever climatology 1981-2010 or so. And use much tighter colour bands. It is almost not possible to make out differences in temperature at all this way and even for mslp it is hard to see how the flow is organised.

We thank the reviewers for the suggestions, the figures have been completely re-made in the current version of the paper.

7. Then onwards from section 3, we turn to the model world of PlaSIM. I appreciate that the authors have brought some of my earlier suggestions into practice by focussing less on the thermodynamic aspects. However, the results that are produced by the analogon approach are quite surprising/disturbing/alarming, at least the one for the RCP85 scenario. In there we see spectacular increases in the frequency of cases. Although the world also warms, this might not yield extremely cold/snowy situations in the end (which the authors warn for already), but still.

8. To me, the huge increase seen in RCP85 raises an alarm bell. Why/how does this occur? Many existing climate model ensembles exists (e.g. CMIP5, CMIP6), but as

far as I know, none does produce such extreme changes in the tails of the distribution. So the authors at least have to come up with a convincing story here.

We thank the reviewer for these comments. Indeed, we found that the previous PlaSim run contained an error due to a problem in the parameterization of ocean fluxes. We performed new simulations correcting the problem and including more emission scenarios. We still observe an increase in the frequency of these analogues with increasing CO2 concentration, but results are less dramatic.
As we point out in the article, the increased frequency of these configurations may be due to a more wavy jet-stream under climate change due to reduced meridional temperature gradient. We are aware and we state in the article that there is no consensus about this, and that the result's validity is limited to the framework of this specific model.

9. An 11-step scheme is presented to obtain structures that are similar to each of the 32 events. However, by now focusing on each of them (in the table), the reader may wonder how strange/anomalous each of them was. The readers have no idea about the mslp fields underlying each case, and therefore have no feeling about what the numbers in the tables indicate. Why not simply use the two cluster centroids decided on in section 2.3 and use these to find analogs for?! One could even use these two cluster centroids and search for distribution changes in the way done e.g. in the snow paper by de Vries et al. Clim Dyn. DOI 10.1007/s00382-012-1583-x. (eg their figure 6).

We thank the reviewer for this suggestion, we agree that this makes more sense and we decided to follow and implement it. In the new version, we consider analogues of the Z500 anomaly fields averaged over the two clusters.

10. The same question I had on the domain size applies here as well. First: is the analoging done on anomalies or full field, and have the authors experimented with the domain size? If the final interpretation is to hold for Italy specifically, it should (I believe) be demonstrated that a domain is chosen that at least for that region provides meaningful results.

Thank you for the question. We use the extended domain only to find the regimes associated to the two clusters. For analogues search, we use the smaller domain indicated in Section 3.3. Figures 5 and 6 show that these analogues catch well the difference between the two configurations over the region of interest, with the high pressure centred over the UK for analogues of cluster 1, and over Scandinavia for analogues of cluster 2.

11. The rationale for using a +4SST is that the MedSea warms faster than the rest, but in PLASIM the oceanwater is globally raised by 4 degrees. I am then surprised to see that this leads to no adjustment whatsoever.

We thank the reviewer for the comment. Indeed, we realised that the PlaSim framwork did not allow to tackle properly the role of Mediterranean SSTs for this and other

reasons, and we decided to remove it from our study, and focus on more emission scenarios instead.

12. Figure 7 is unclear what we see. Is it climatological mean snowcover? Units seem to be kg/m2, which is probably the same as cm snow. But then showing the snowcover up to natural logarithm values of -15 is rather small/meaningless..

13. Figure 8-9 same story as for figures 3-5. (see above comments)

14. Finally, in those figures RCP85 mslp analogs are combined with the T850 conditions. This reduces the number of cases accordingly.

Thank you for these comments. Figures have been re-made completely, and we do not consider combined analogues anymore.

____
_
____
_

Review of “Present and future synoptic circulation patterns associated with cold and snowy spells over Italy” by M. D’ Errico et al.

GENERAL COMMENTS

The manuscript has been extensively revised from its first version. The catalogue of combined heavy snowfall and low temperature events in Italy between the years 1954 and 2018 is retained, but large changes have been made to the rest of the paper. First, a cluster analysis has been used to divide the observed cold spells to two clusters based on the associated sea level pressure (SLP) fields. Second, the analysis of the intermediate complexity PlaSim climate model simulations has been refocused on the search of situations that are analogous to the 32 identified cold spells in terms of the SLP and the 850 hPa temperature (T850) distributions.

The refocusing of the PlaSim analysis from snowfall to the atmospheric dynamical features has improved the manuscript, since the model is clearly more suitable for the simulation of the latter than the former. Furthermore, the revised manuscript is honest in acknowledging that cold spells in warmer future climate will produce less snow.

However, I still have some questions and concerns related to the interpretation of the main results. The analysis of the PlaSim simulations suggests that the frequency of SLP circulation states that resemble those in the observed cold spells is increasing strongly in the RCP8.5 scenario, whereas the corresponding change resulting for an artificial 4 K warming of the sea surface temperatures is negligible. To me this result is counterintuitive, and its physical significance is difficult to assess without further information on what the increase in RCP8.5 results from. Specifically, the increase in the frequency of circulation analogies could be associated with

1. A change in the winter mean SLP field that makes the average SLP distribution more similar to that observed during the cold spells, or
2. A change in the variability of SLP around its mean state, resulting in a larger frequency of SLP anomaly fields that resemble the SLP anomaly fields during the cold spells.

These two possibilities could be distinguished by repeating the cluster analysis for SLP anomalies relative to the observed or simulated (present-day and RCP8.5 separately) DJFM mean SLP field. If the increase in frequency is still seen when the mean state change has been eliminated by focusing on the anomalies, if must originate from changes in the simulated variability. If it disappears, then the change in the mean SLP field is the key.

Furthermore, if the change in the mean SLP field turns out to explain the increase in the frequency of the circulation analogies, a follow-up question is how and where the mean state changes. In which parts of the (rather large) analysis domain (22.5-70°N, 80°W-50°E) does the new mean state approach the SLP fields during the observed cold spells?

We thank the reviewer for this comment. Indeed, we moved to an anomaly framework, and we also decided to use Z500 instead of SLP for clustering.

Finally, biases in the simulated present-day winter mean SLP field might affect the change in the frequency of the circulation anomalies in a non-intuitive way. To check for this possibility, it would be prudent to repeat the analysis after also applying the simple linear scaling bias correction to SLP, not only to T850 as was apparently done.

We thank the reviewer for the suggestion, we proceeded to apply the scaling bias correction also to SLP and Z500 fields to eliminate the effect of biases.

I also have some concern about the ability of the clustering algorithm to identify good circulation anomalies in the model simulations. There appear to be quite large differences between the observed (Fig. 3) and the simulated (Figs. 8-9) SLP fields for both two clusters of the cold spell cases. Furthermore, the difference between the two clusters appears much smaller for the simulations. I wonder if this might be improved by selecting a somewhat smaller domain in the search of the circulation analogies.

We thank the reviewer for the comment. Besides these issues, we also realised there was a problem with the parameterization of fluxes in PlaSim, which we corrected before running new simulations. In the new version of the paper, we directly look for analogues of the two clusters; results obtained from the simulations (Fig 5 and 6) are in good agreement with the fields associated to the two clusters.

Aside of this issue, the figures need improvement. All the maps (Figs. 3-5 and 7-9) use a very fine-grained colour scale, resulting in weak contrasts between the individual shades. Therefore, it is difficult to estimate any quantitative values from the maps. At least for SLP and T850, traditional isoline plots with labelled contours (with or without colours superimposed) would most likely be more informative.

We thank the reviewer for the comment; after trying isoline plots and also considering the comments from the other reviewer, we produce new figures with an enhanced colorscale, but we decided to keep a map format. We believe that the new maps are nonetheless clear and easily interpretable by the reader.

More detailed comments follow below.

SPECIFIC COMMENTS

We thank for the detailed comments, they have been addressed in the new version of the paper, or they refer to sentences that have been removed or deeply changed based on the new results.

1. Figure 1. Please provide an absolute scale for the duration of the events
2. L277. Where was -23°C observed, if it was even colder in Marcesina?
3. L325. The lowest temperature in Finland in February 2012 was -42.7C (https://www.ilmatieteenlaitos.fi/lampimin-ja-kylmin-paikka-vuosittain)
4. L434-435. respectively 12 and 20 for k = 1 and k = 2?
5. L436. the 20 events in cluster 2
6. Figure 2. y axis labels are only partly visible
7. L503. (expected to lead to ~+3.5 K SST) This kind of numbers are meaningless without mentioning the emission scenario. From reading Adloff et al. (2015), this number most likely represents the high SRES A2 scenario.
8. L585-587. Are "decreased", "increasing" and "unchanged" the right words, when comparing PLASIM with the real world (NCEP) frequencies? Rather "smaller", "larger" and "the same"?
9. Caption of Table 1: cold spell SLP analogues?
10. Caption of Figure 7. Is this really a natural logarithmic scale from 0 to -15? Exp(-15) would mean about 3*10-7 kg m-2 of snow, which seems incredibly little even in Southern Europe (equivalent to having one day with 1 kg m-2 of snow