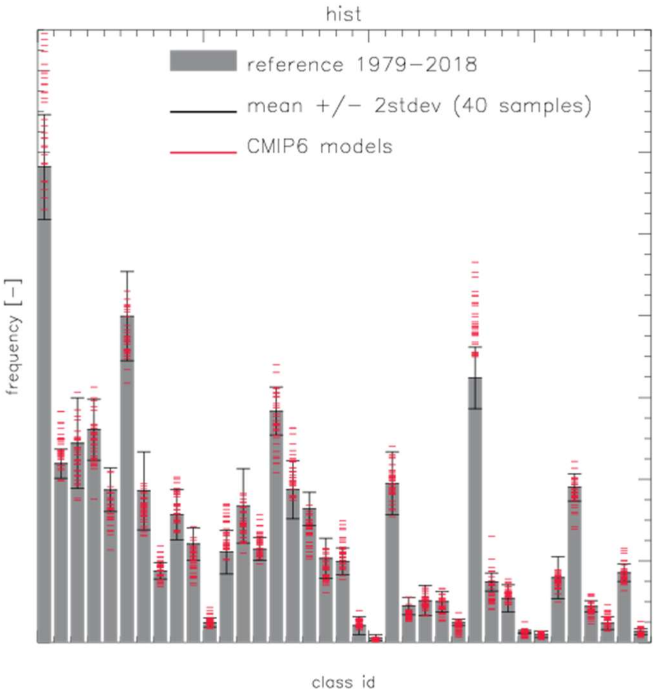
Answers to comments of all Reviewers.

Original comments of the Reviewers are listed in black, our answers - in blue colour.

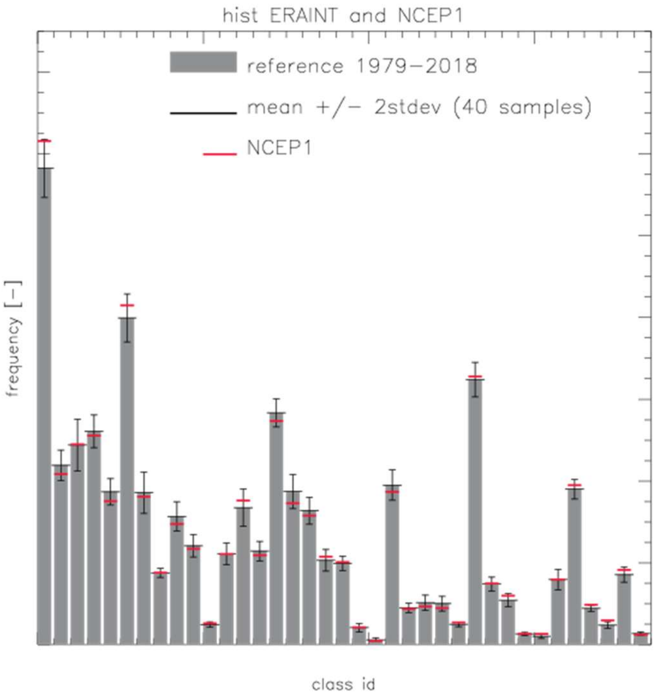
Answer to Reviewer 1.Firstly, we would like to thank reviewer 1 for going to such lengths and calculate his own simulations to illustrate his issue. Our comments on these simulations are the following: As already noted in our previous answer, the distribution of the synoptic patterns within the sample of 40 years of ERAINT is not a simple multinomial. Considering a single day a random variable, which can take one of 37 states (one state=one synoptic class), the days are neither independent nor identically distributed. Instead, the succession of weather patterns is highly auto-correlated (a great number of patterns is simply not possible to succeed each other) and the distribution depends heavily on the season and on low frequency climate variability. These preconditions narrow down the state space of the sequence considerably. Unfortunately, we are not able to quantify this restriction. But it entails the inconvenient consequence that the simple multinomial suggested by reviewer 1 is largely inappropriate as a null model for our data. A large portion of the state space spanned by the multinomial distribution, as far as physically possible, would generate a climate that differs from historic climate. What portion of the state space exactly would generate a different climate, how much different this climate would be and by which measures, we don't know. But the confidence intervals presented by reviewer 1 in his 2nd figure are definitely not applicable in our case.

As suggested by reviewer 1, we have repeated our resampling analysis with 20 instead of 30 years for each bootstrap sample. The results look very similar (to the Figure S9 in the supplement of our manuscript), albeit with wider confidence intervals as expected (see figure below):



This Figure (above) does not resemble the 3rd figure of reviewer 1. But as shown in Figure 6 of our manuscript, 20 years of daily data is the very lower limit of data necessary to capture all synoptic patterns that occur in the historical period. We therefore advised for longer data periods and do not consider the 20-year resampling a proper analysis of sampling error.

We believe that the sampling error is quite appropriately captured in the analysis of the alternative reanalysis NCEP1, which was generated by a completely different model in a completely independent attempt to reconstruct the historic climate, but of course based on the same observations ensuring that the state space is equal. The histogram of synoptic patterns captured in this alternative reanalysis NCEP1 is shown in the figure below (red colour):



The frequencies of each synoptic pattern in ERAINT and NCEP1 are close to each other. The NCEP1 frequencies lie within a 2*sdev interval of ERAINT for all synoptic patterns except one (SP class 35, 3rd from left in the above figure), which is nevertheless very close to its upper bound. In fact, one out of 37 classes to overshoot the *2*stdev* by a little margin is exactly what should be expected. The closeness of the NCEP1-genrated histogram to the reference histogram gives us an additional evidence that the frequencies of synoptic patterns are estimated quite well.

We add this latter Figure (with NCEP1-frequencies) and our discussion to the supplement of the manuscript. We hope this strengthen our argument about robustness of estimates for the SP-class frequencies.

Answer to Reviewer 2. Reviewer 2 recommended to accept the Manuscript without corrections.

Answer to Reviewer 3.

INE 869 – Should this line say, “whereas other models have *lower* values” not “higher values”?

Yes, we agree. This is a typo we overlooked to change as we switch from the Quality Index (best value is 1.0) to the Jensen-Shannon distance (best value is 0.0). In terms of HS-distance, the better performance of a Model is shown by a “lower” value as reviewer 3 noticed. Thank you very much!

We corrected the text to the following (Lines 868-869): *“… the climate simulation NorESM2-LM seems to underperform all other models (Mean JS=0.137) whereas other models show lower values (i.e. smaller distances to the reference statistics).”*