



Supplement of

Classification of synoptic circulation patterns with a two-stage clustering algorithm using the modified structural similarity index metric (SSIM)

Kristina Winderlich et al.

Correspondence to: Kristina Winderlich (kristina.winderlich@dwd.de)

The copyright of individual parts of the supplement might differ from the article licence.

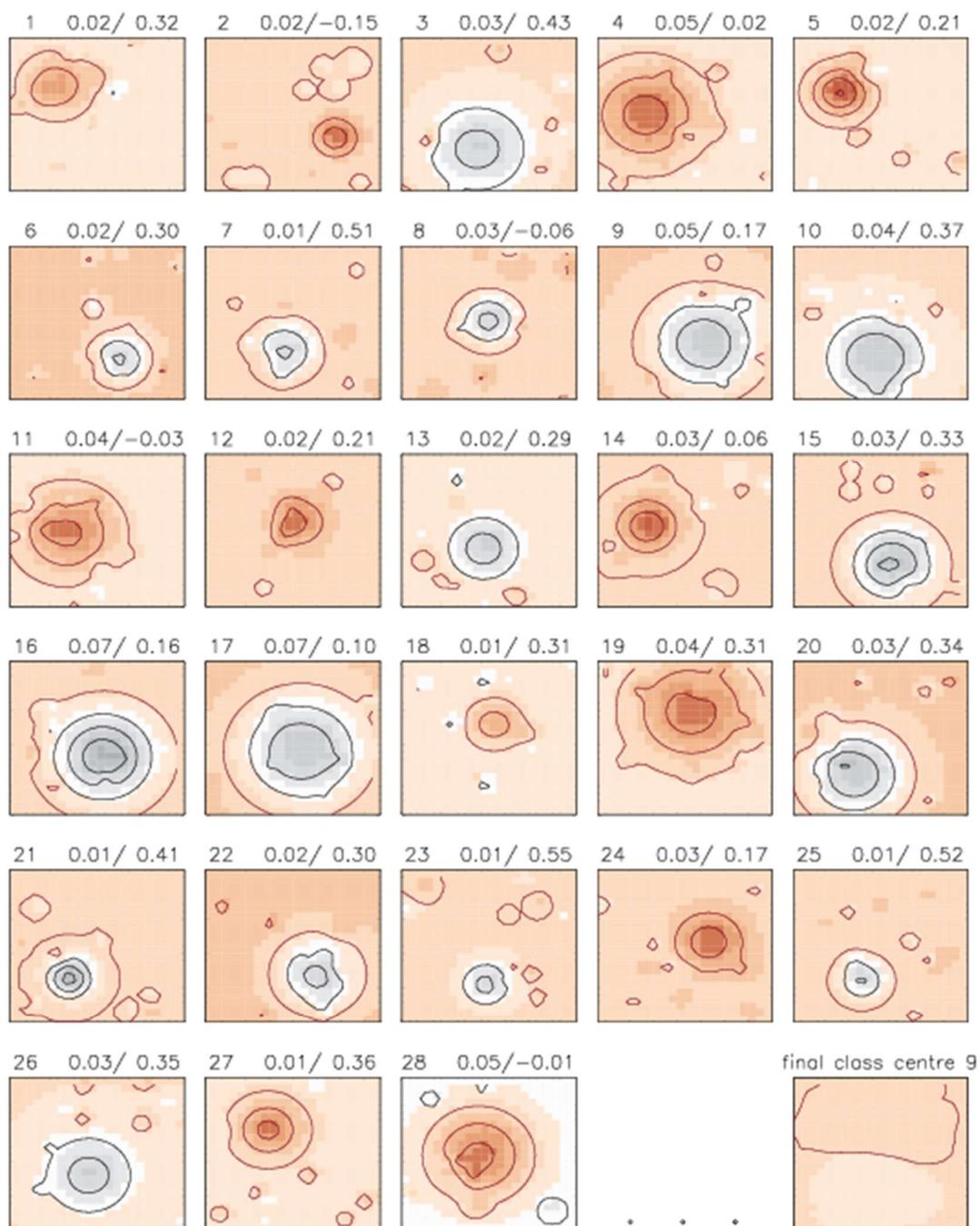


Figure S1: An example of a “snowball”-class resulted from the classification *k-mean-MSE*: class centre 9 (bottom right plot) has very weak structure, whereas its first 28 elements (only shown as examples out of 132) have strong anomalies of mixed sign and placed in various locations of the domain. Values of MSE and SSIM for an element to the class centre 9 are shown above plots for each element in form $\langle \text{MSE} \rangle / \langle \text{SSIM} \rangle$.

Sensitivity of Jensen-Shannon distance metric

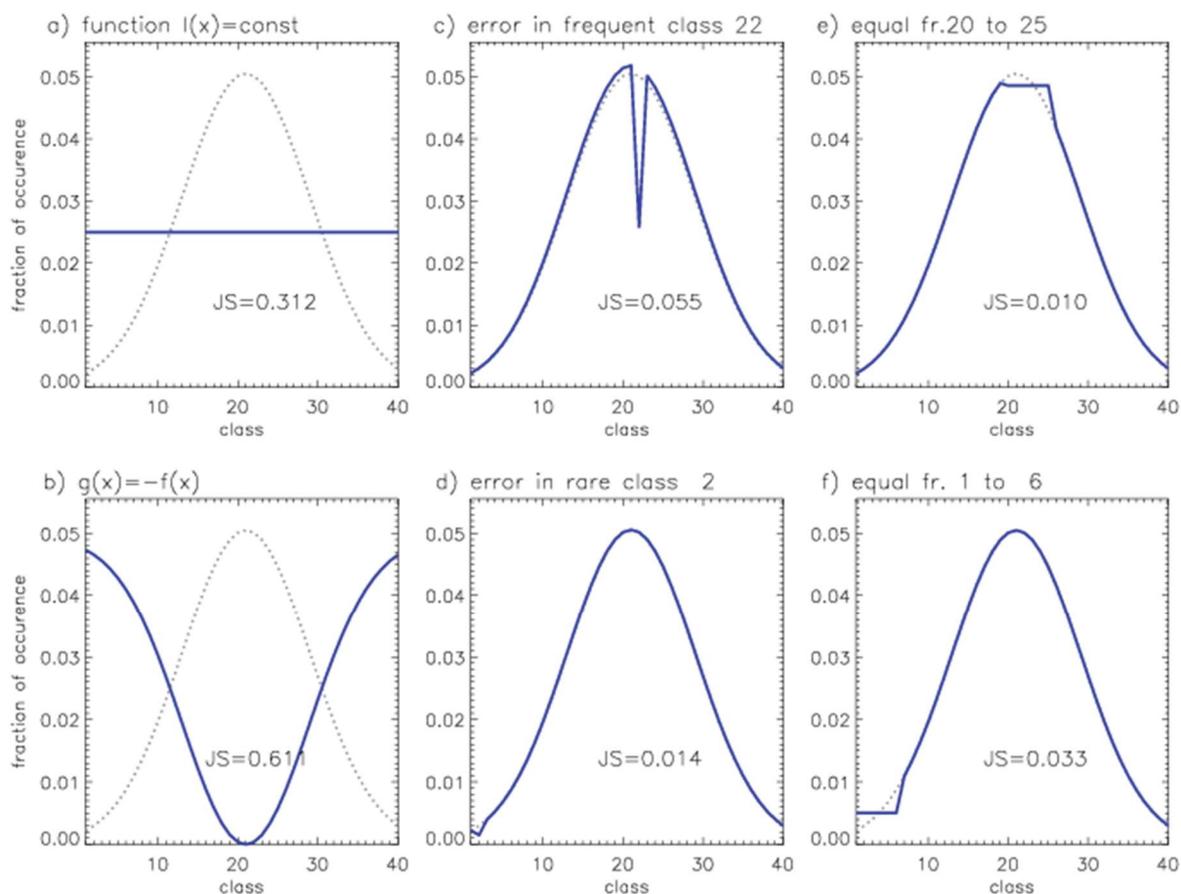


Figure S2: Jensen-Shannon distance between a reference Gauss-shape histogram $f(x)$ (dashed grey line) and six histograms of other shapes: a) equally frequent classes $l(x)=\text{const}$, b) “mirrored” $g(x)=a-f(x)$, c) with reduced fraction of one frequent element, d) with reduced fraction of one rare element, e) with six equally frequent classes, f) six equally rare classes.

Figure S2a shows that $JS=0.312$ between the Gauss-shape histogram and the uniformly distributed histogram should be considered a very large distance as the compared distributions are obviously very different. Figure S2b shows the largest [among the present examples] value of $JS=0.611$ as two “mirrored” distributions are compared, as expected. Please note, the Jensen–Shannon divergence is bounded by 1 for two histograms (using base 2 logarithm), and therefore, JS-distance is bounded by 1 as well.

It may be argued that rare classes have small impact on JS-distance, as demonstrated in the Figure S2d: an error in frequency of one rare class makes moderate contribution to the total JS-distance (as compared to the contribution of the similar error in one frequent class, Figure S2c). The plots “c” and “d” may lead the observer to a conclusion that errors in rare classes are negligible in computing JS-distance. This is a misleading conclusion. Having a look at Figure S2e and Figure S2f offers some clarification: JS-distance is higher in response to errors in multiple rare classes (plot “f”) as to the errors in frequent classes (plot “e”). The relative change in frequency of the rare classes by such error is quite high (the mean distribution used in computation of Jensen-Shannon Divergence undergoes large changes relatively to the original distributions).

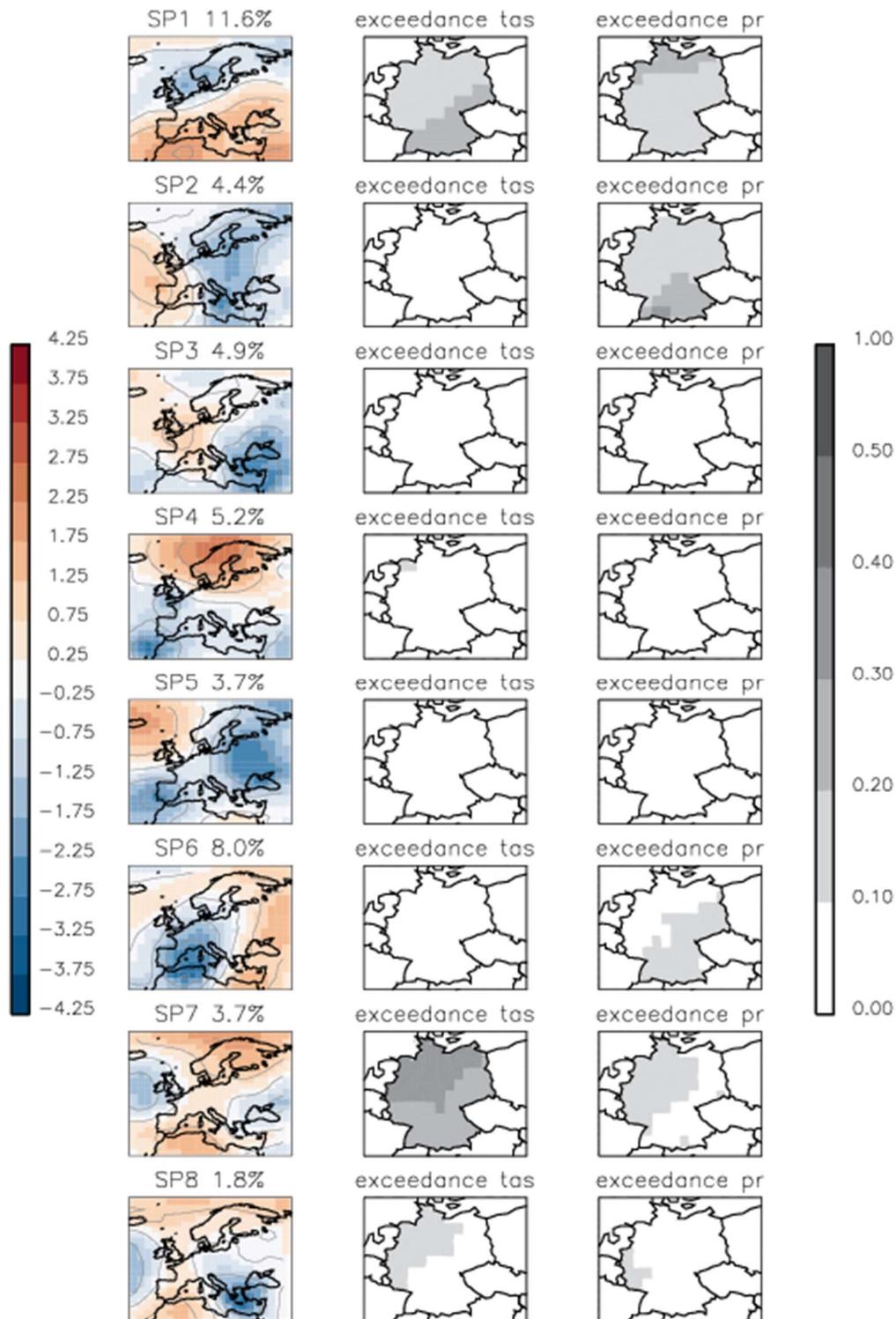


Figure S3: Synoptic patterns SP1-SP8 and corresponding maps of exceedance of 90th-percentile for near surface temperature (tas) and for precipitation (pr). Each map of exceedance shows the fraction of all days in the corresponding SP-class when temperature/precipitation exceeds the 90th-percentile of these variables computed over the period of 1979-2018.

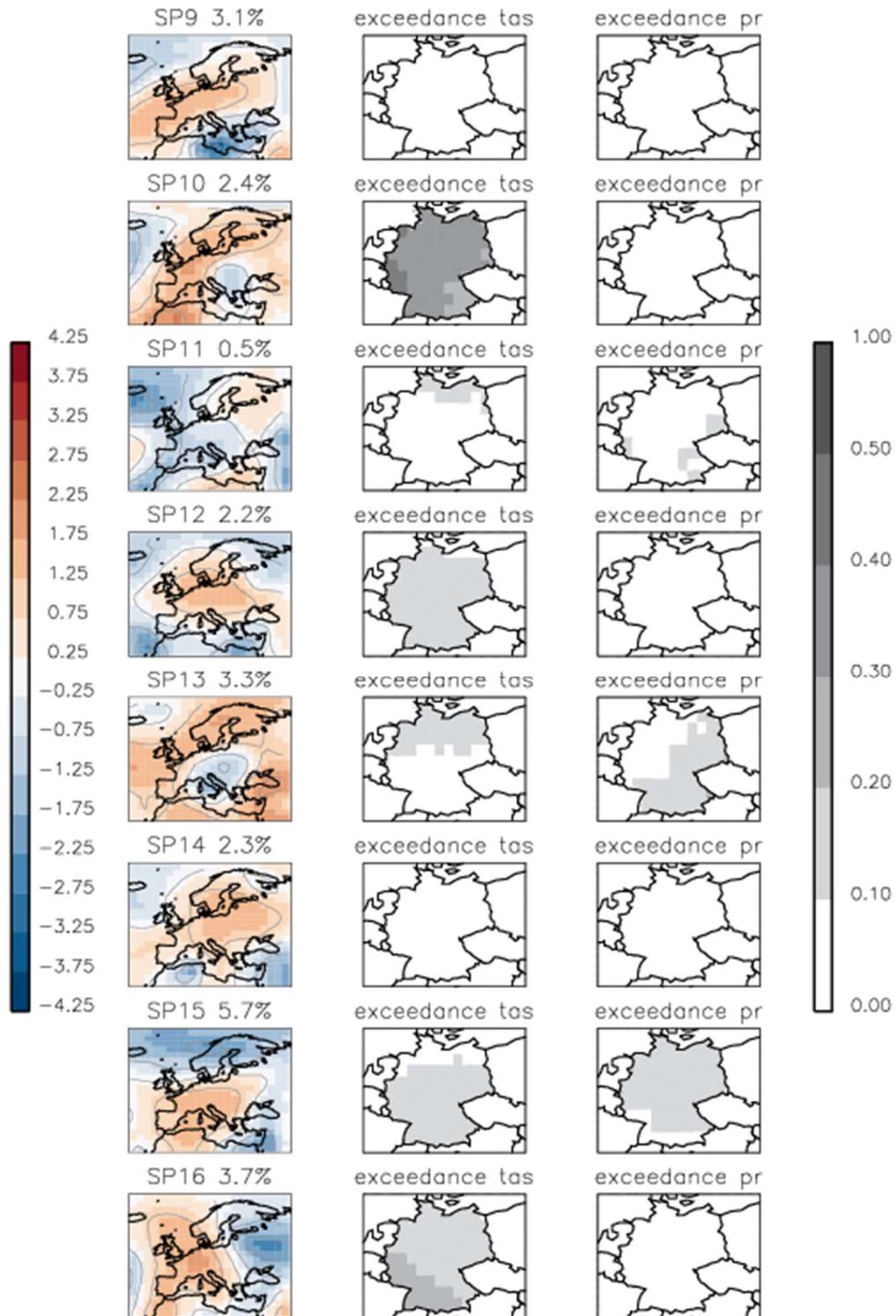


Figure S4: The same as Figure S 3 but for Synoptic patterns SP9-SP16

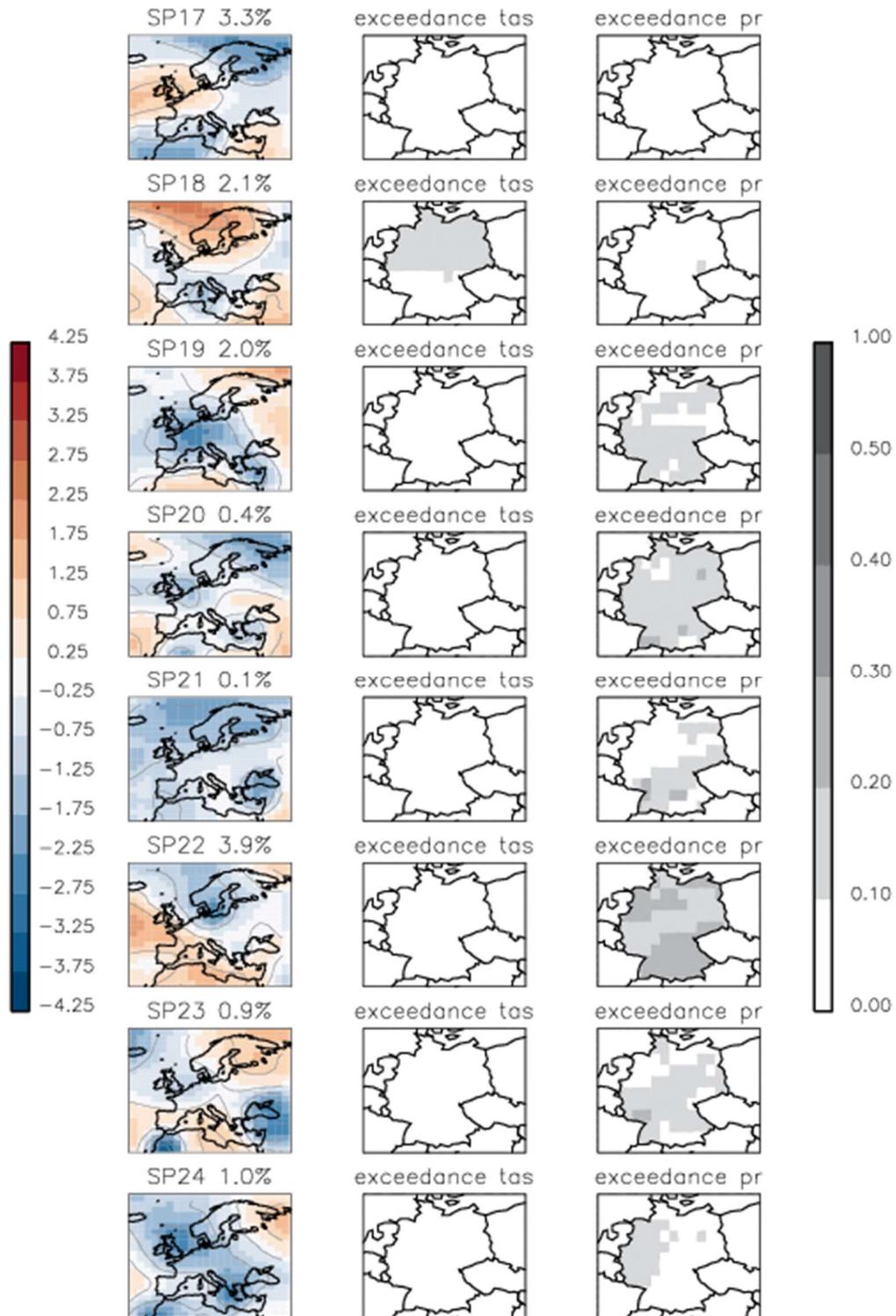


Figure S5: The same as Figure S 3 but for Synoptic patterns SP17-SP24

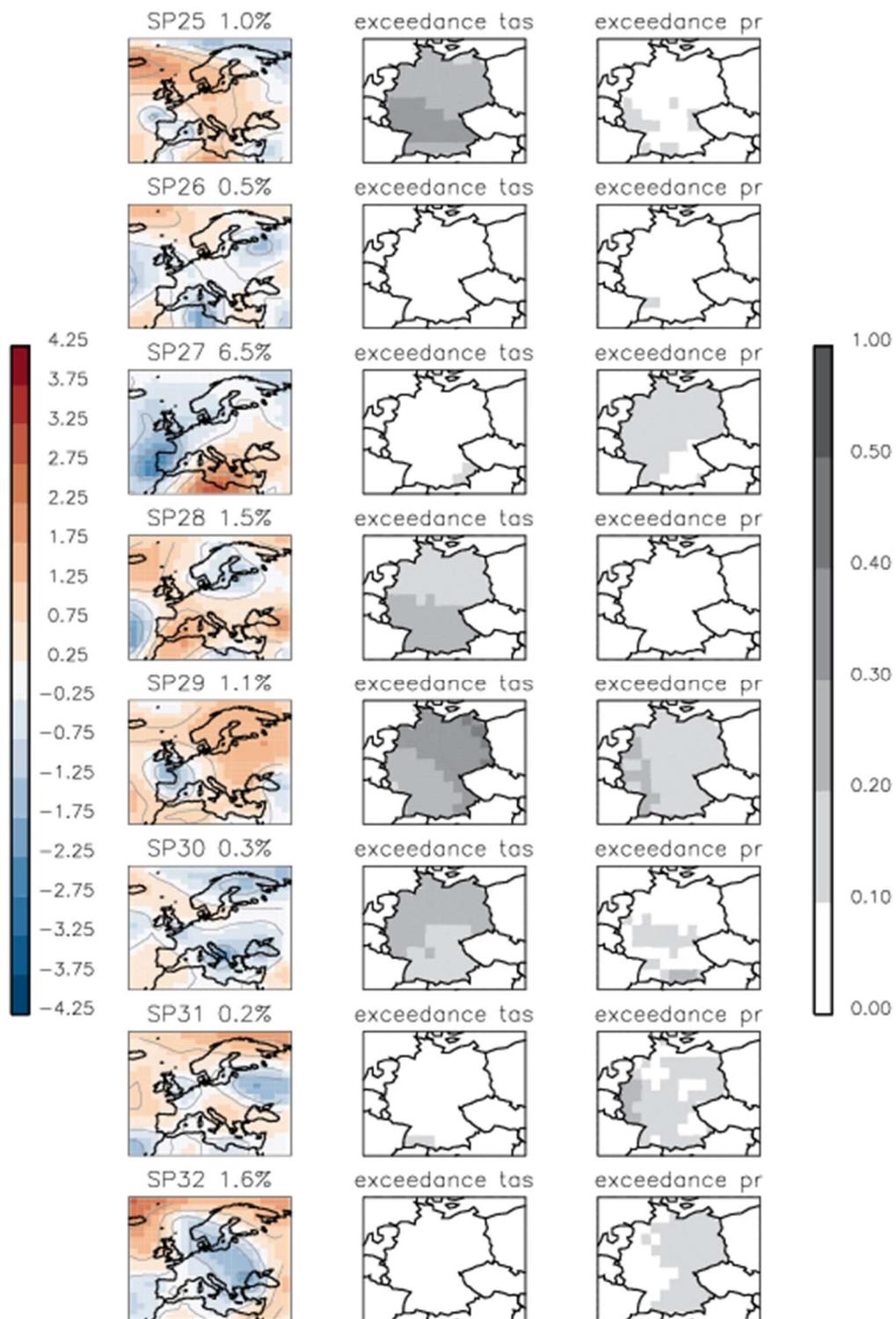


Figure S6: The same as Figure S 3 but for Synoptic patterns SP25-SP32

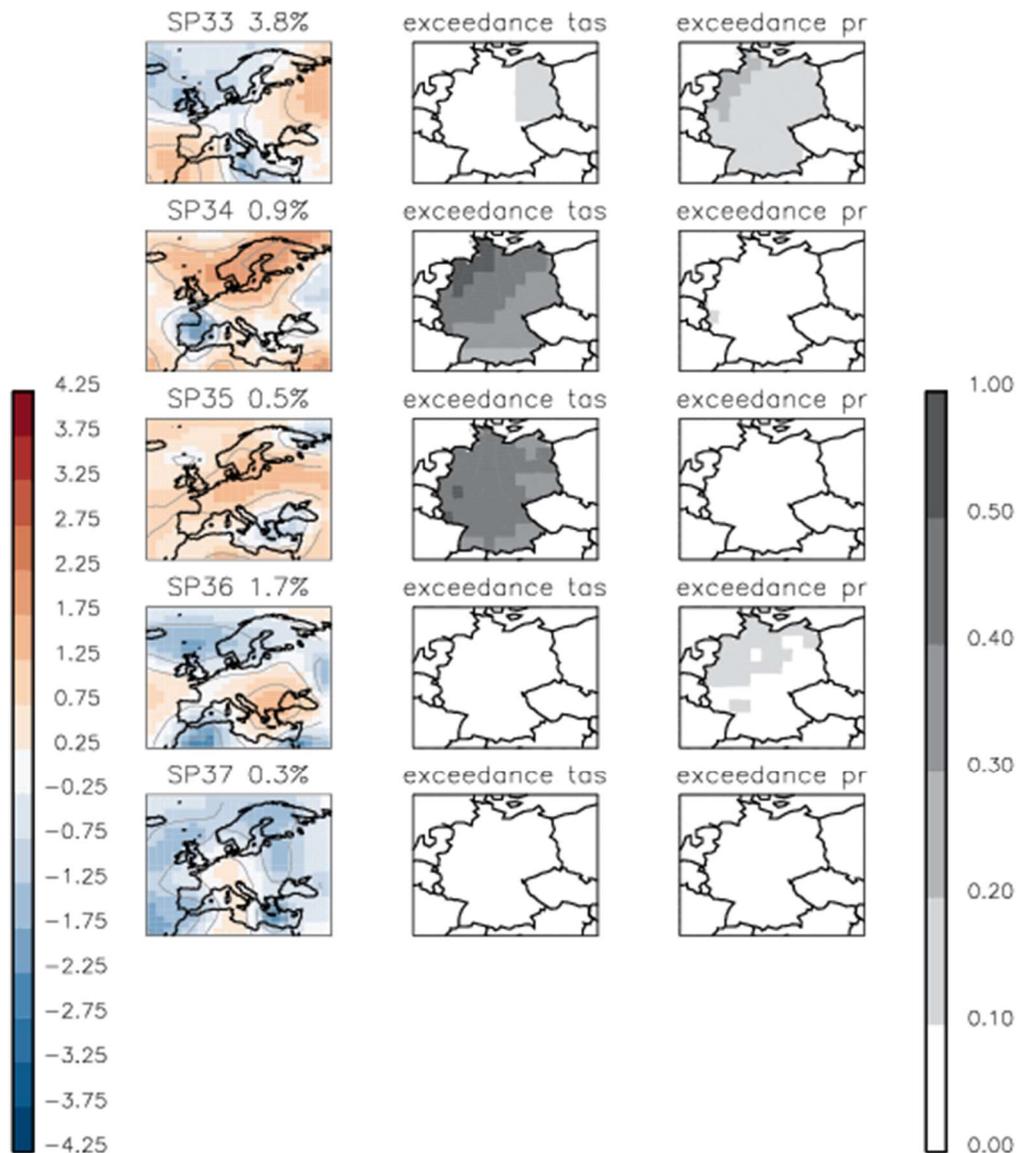


Figure S7: The same as Figure S 3 but for Synoptic patterns SP33-SP37

Analysis of the robustness of the estimates for the statistics *HIST*/*TRANSIT*/*PERSIST*.

It is conceptually difficult to assess the sampling uncertainty of the introduced statistics with only one realisation (which is furthermore not a simple Multinomial distribution). Therefore, we use resampled data (10-fold block cross validation, i.e. a sliding block of 10 years cut from the sample in a cyclic way) to build 40 different sets of 30 year periods for computing histograms *HIST*, *HIST_JFD*, *HIST_MAM*, *HIST_JJA*, *HIST_SON*, and *TRANSIT*, *PERSIST*-matrices. We decided to use the 30-years period as we observed that the classification algorithm produces stabilized number of classes using at least 20 years of data (see chapter “4.1. Synoptic classes, effect of the threshold TH_{merge} on the number of classes” in the main text).

From these 40 realizations, we estimate the mean and the standard deviations (*stdev*) for each element (frequency/persistence of a certain class, transition probability from one class to another) of these one- and two-dimensional statistics. As a very rough, zeroth-order check of robustness, we compare the estimated values in the frequency histograms with two times their resampled standard deviation. It appears that the uncertainty in the histograms is reasonably low: all values in the histogram are greater than their individual $2*stdev$, even for the rare classes (Figure S8).

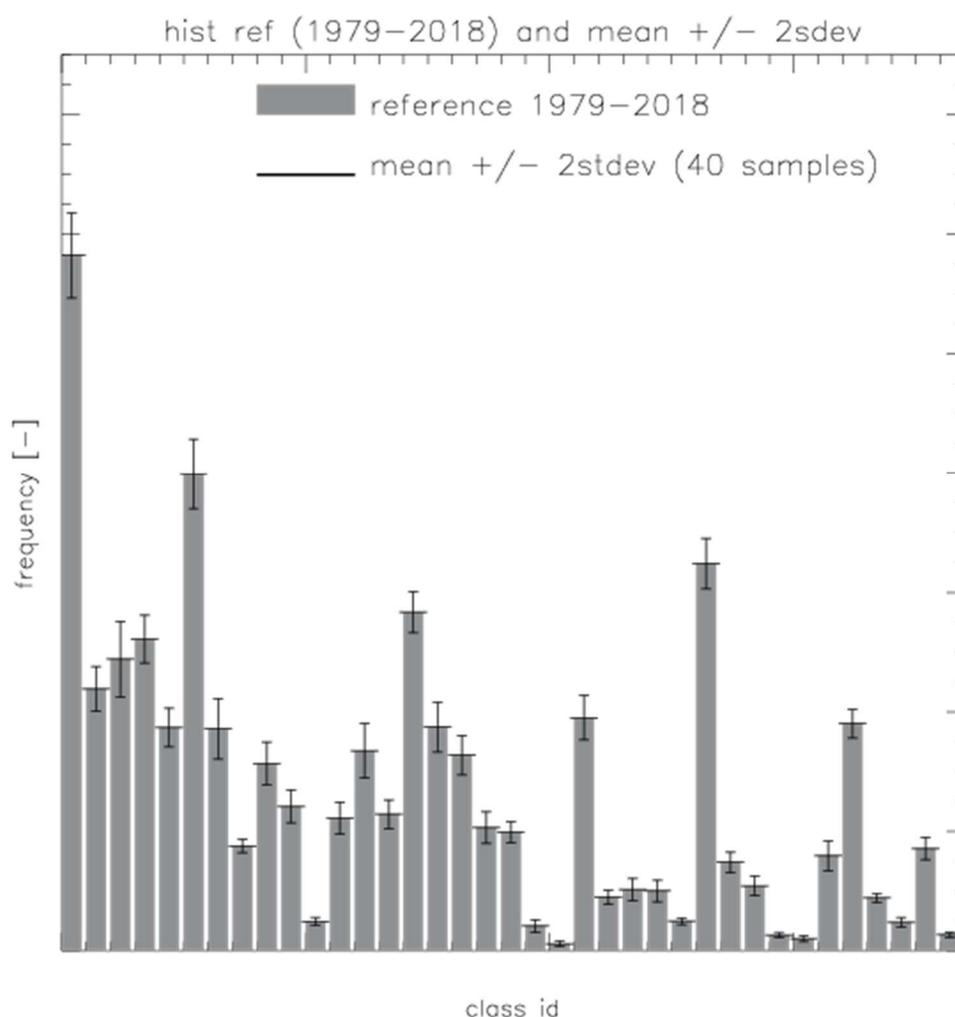


Figure S8: Histogram *HIST* for class frequency of 37 classes obtained using the full set of 40 years of ERA-Interim reanalysis data. The black horizontal lines show the *mean* frequencies and $2*stdev$ interval of 40 samples of these classes using 30-year data volumes sampled on the full set of the ERA-Interim reanalysis.

Furthermore, the values produced by the CMIP6 projections, though some do fall inside our “confidence intervals”, show far higher departures from the reference than the resampled data, both in frequent and rare classes. Interestingly, many models seem to shift weight from a number of rare classes to the frequent ones, which would indicate a reduced diversity of circulation types (Figure S9). This histogram shows the usefulness of our approach comparing the frequency histograms of the projections to the reference.

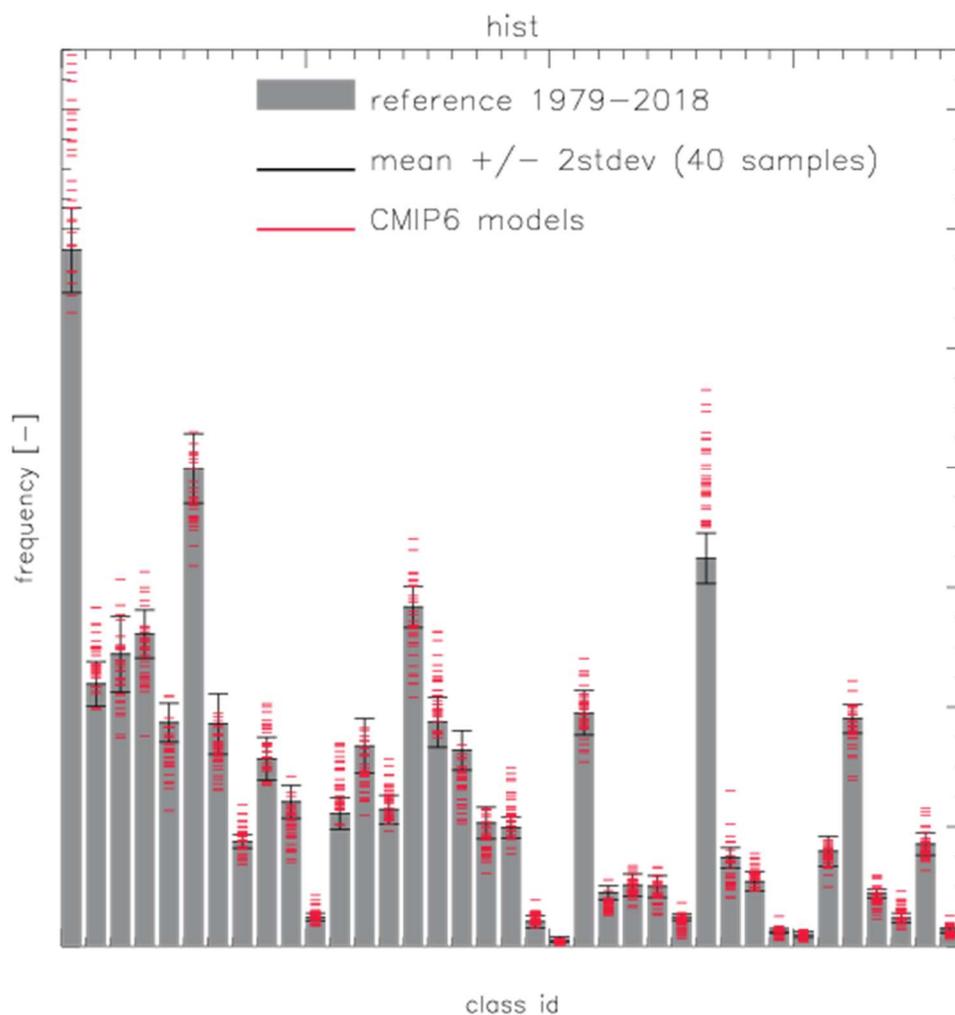


Figure S9: Histogram *HIST* for class frequency of 37 classes obtained using the full set of 40 years of ERA-Interim reanalysis data. The black horizontal lines show the mean frequencies and $2*stdev$ interval of 40 samples of these classes using 30-year data volumes sampled on the full set of the ERA-Interim reanalysis. Red horizontal lines show values of histograms computed for 32 CMIP6 models.

Another way to support the robustness of the estimated frequencies of synoptic classes is to look for these classes in the alternative reanalysis, NCEP1, and to compute their frequencies. The NCEP1 reanalysis was generated by a completely different model in a completely independent attempt to reconstruct the historic climate, but based on the same observations providing an alternative view of the “true” state of the climate. Therefore, if the class frequencies are robustly estimated in the reference reanalysis ERAINT, the frequencies of these classes in the alternative reanalysis NCEP1 should be close (fall within $2*stdev$ interval) to the frequencies in ERAINT. This is exactly the case (Figure S10): the frequencies of each synoptic pattern in ERAINT and NCEP1 are close to each other; the NCEP1 frequencies lie within $2*stdev$ interval of ERAINT for all synoptic patterns except one (*SP-class* 35, 3rd from the left), 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-generated histogram to the reference histogram gives us an additional evidence that the frequencies of synoptic patterns are estimated quite well.

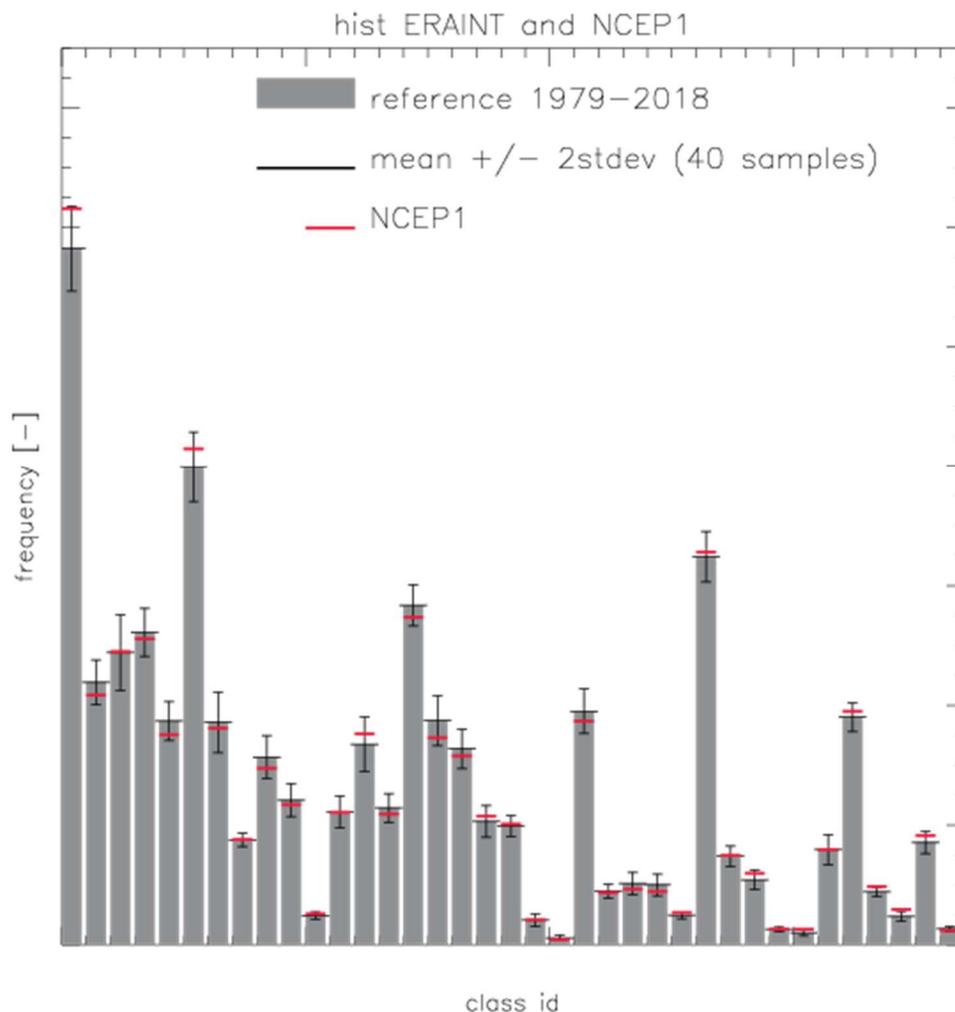


Figure S10: Histogram *HIST* for class frequency of 37 classes obtained using the full set of 40 years of ERA-Interim reanalysis data. The black horizontal lines show the mean frequencies and $2*stdev$ interval of 40 samples of these classes using 30-year data volumes sampled on the full set of the ERA-Interim reanalysis. Red horizontal lines show values of histograms computed for or NCEP1-reanalysis (data of this alternative reanalysis were assigned to the reference set of 37 SP-classes using SSIM).

Similarly to the analysis for *HIST*, we analyse the elements of the *TRANSIT*-matrix and see them likely to be robustly estimated if they exceed $2*sd$. Figure S11 shows the *TRANSIT*-matrix and its re-sampled standard deviation. As the transition-elements with larger absolute value (higher frequencies of occurrence) are outlined with black contour (this indicates the element is $\geq 2*sd$) they are likely to be robustly estimated. The transition-elements with smaller values (rare transitions) are less likely to be robustly estimated, as expected. We suggest, that a larger data set would help to estimate all elements of the *TRANSIT*-matrices more robustly.

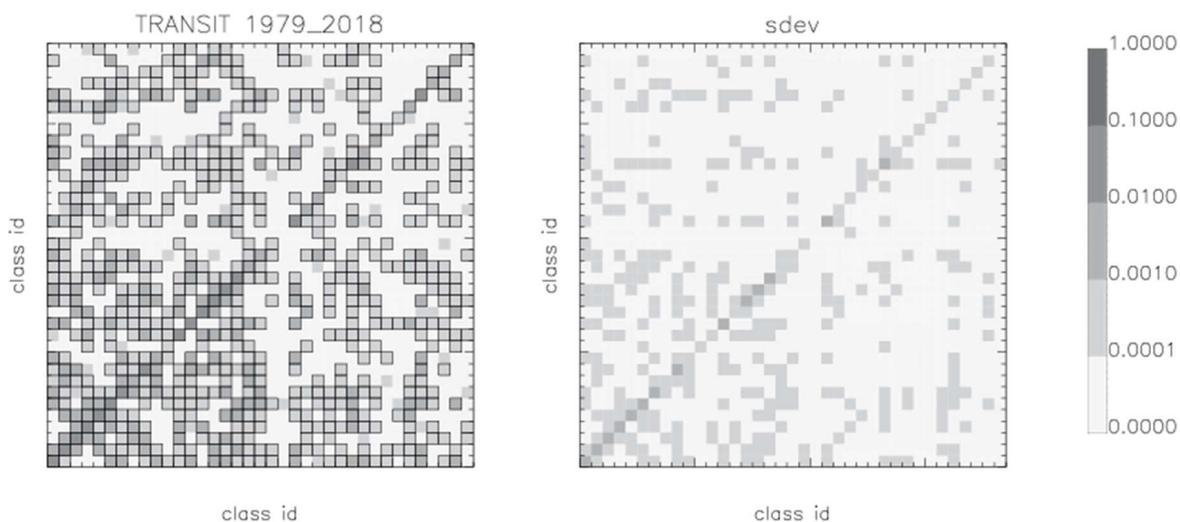


Figure S11: *TRANSIT*-matrix computed on the full set of data 1979-2018 (left) and its re-sampled standard deviation (right) computed over 40 samples of the *TRANSIT*-matrix. Matrix elements highlighted with black contour are $\geq 2*sd$. Note: shades of grey show the probability of class-to-class transition, scale is logarithmic.

In contrast to the *TRANSIT*-matrix, the *PERSIST*-matrix seems to be sufficiently robust, i.e. all elements are greater than their $2*sd$. This might result from the smaller number of elements contained in the matrix.

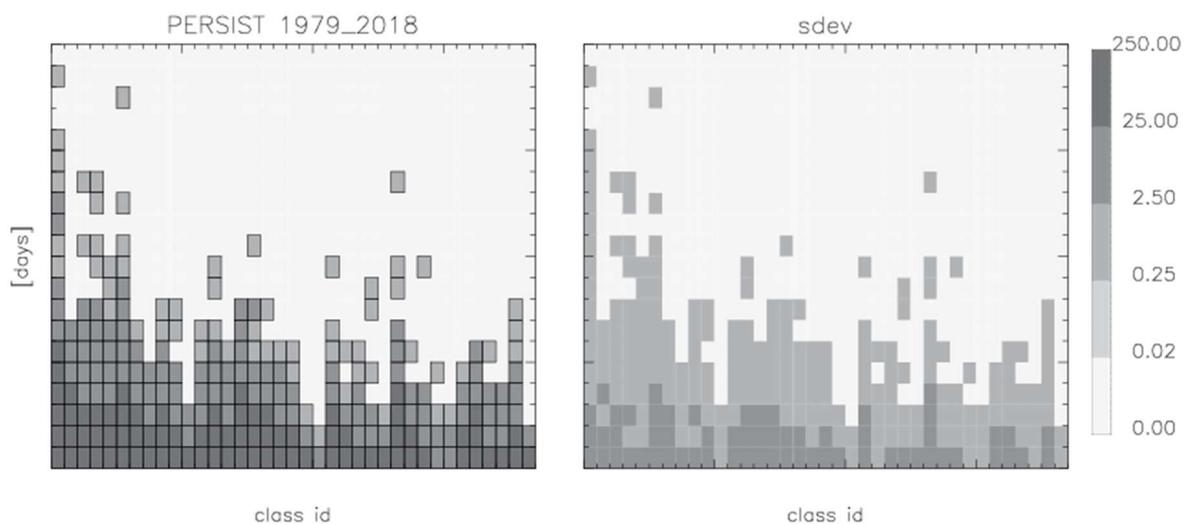


Figure S12: *PERSIST*-matrix computed on the full set of data 1979-2018 (left) and its re-sampled standard deviation (right) computed over 40 samples of the *PERSIST*-matrix. Matrix elements highlighted with black contour are $\geq 2*sd$. Note: shades of grey show the number of occurrences for each class of the duration 1-to-8 days, scale is logarithmic.

Despite the large number of elements contained in the histograms and the persistence matrix, the resampling of these statistics shows quite similar values with small standard deviations of the samples. Therefore, we suggest these statistics could be used to compute differences between a model and a reference (i.e. reanalysis). In the case of the transition matrix, a larger data amount could provide more robust estimates for rare elements.