Analyzing Arctic surface temperatures with Self Organizing-Maps: Influence of the maps size

D. Mewes, Ch. Jacobi

Universität Leipzig, Institute for Meteorology, Stephanstr. 3, 04103 Leipzig,
E-Mail: daniel.mewes@uni-leipzig.de

Summary: We use ERA-Interim reanalysis data of 2 meter temperature to perform a pattern analysis of the Arctic temperatures exploiting an artificial neural network called Self Organizing-Map (SOM). The SOM method is used as a cluster analysis tool where the number of clusters has to be specified by the user. The different sized SOMs are analyzed in terms of how the size changes the representation of specific features. The results confirm that the larger the SOM is chosen the larger will be the root mean square error (RMSE) for the given SOM, which is followed by the fact that a larger number of patterns can reproduce more specific features for the temperature.

Zusammenfassung: Wir benutzten das künstliche neuronale Netzwerk Self Organizing-Map (SOM), um eine Musteranalyse von ERA-Interim Reanalysedaten durchzuführen. Es wurden SOMs mit verschiedener Musteranzahl verglichen. Die Ergebnisse zeigen, dass SOMs mit einer größeren Musteranzahl deutlich spezifischere Muster produzieren im Vergleich zu SOMs mit geringen Musteranzahlen. Dies zeigt sich unter anderem in der Betrachtung der mittleren quadratischen Abweichung (RMSE) der Muster zu den zugeordneten ERA Daten.

1 Introduction

Finding specific patterns of meteorological variables is necessary to understand common features that governs the weather and climate of the Earth. One method that can been used to find patterns and to analyze time variation of those patterns is called Self Organizing-Maps. It was developed by Kohonen (1998) and has been used in multiple studies since (e.g. Hewitson and Crane (2002); Cassano et al. (2006); Lynch et al. (2016); McDonald et al. (2016); Ford and Schoof (2017)). The advantage of SOMs compared to empirical orthogonal Eigenfunction analysis is that SOMs are not limited to a linear assumption. The pattern recognition of SOMs seeks to map a user-defined number of patterns to a distribution of input data, while preserving the probability density function of the analyzed data. This means that the SOM method is reproducing patterns in a way that the patterns that are more frequent in the dataset occur more often in a SOM.
As a cluster algorithm, using the SOM method poses the problem to choose the number of patterns that shall be extracted. This number has to be defined *a priori* and may depend on the parameters to be investigated, the general variability of the dynamics, and the sciences questions. In general it can be said that a larger number of patterns result in

a more detailed decomposition of the data. In turn, a small number of patterns gives a broader picture of the meteorological states.
In this report we analyzed the clustering of high-latitude two meter temperature winter fields for differently sized SOMs. In Section 2 we explain the data that has been used and which sizes of SOMs were analyzed. Section 3 will show the results of the different SOMs. The last section 4 will shortly summarize and discuss the results.

2 Data and Method

For this study, synoptic (00, 06, 12, 18 UTC) ERA-Interim (ECMWF, 2017; Dee et al., 2011) two meter temperatures during the 1979-2016 winter seasons were used. The ERA-Interim data are available at a horizontal resolution of approximately 0.7 degrees. For the performed analysis the synoptic values were daily averaged, which lead to a total of 3340 daily temperature fields. The analysis was limited to regions north of 50°N.
To feed the temperature data into the SOM algorithm it is necessary to reshape from an three-dimensional data field (time, latitude, longitude) to a three-dimensional data field where the latitude and longitude dimensions are stacked onto each other. After applying the SOM algorithm, the shape of the latitudes and longitudes as horizontal coordinates is restored. Afterwards each pattern is subtracted by the ERA-Inetrim winter mean temperature in order to obtain the temperature anomalies with respect to the the whole analyzed time frame. Note, that for the structure of evolving patterns it does not matter whether anomalies are directly used in the SOM analysis, or whether the differences from the mean are calculated afterwards. The general process of the SOM method is summarized in Mewes and Jacobi (2017).
SOMs of the size of 3x1, 3x2, 4x3 and 5x4 were analyzed. For all of these test cases the set-up parameters for the SOM where the same: 1000 iterations, learning rate starting at 0.5 and linearly decreasing to 0.001 during the iterative process, and an decrease of the neighbourhood definition from the number of the columns to one during the iteration. To distinguish between patterns of the different SOMs the following notation is introduced: (r,c).(CxR), where r and c denotes the row and column of the pattern as described within one SOM. R and C denotes the maximum number of rows and columns of the specific SOM. For example, the leftmost pattern of the SOM with three patterns will be addressed as (0,0).(3x1). Speaking of a small (large) SOM means that the SOM consists of a small (large) number of patterns.
Further the root mean square error (RMSE) of each pattern of each SOM compared to the corresponding daily data fields was calculated. The RMSE was averaged over all patterns of one SOM. This averaged RMSE is used as a metric to characterize how well a size of a SOM corresponds to the representation of the daily data field. Hereby large values will indicate that the SOM only serves as an overview of the data while small values for the averages RMSE are corresponding to a SOM that can represent single features better.

3 Results

Figure 1 shows the SOM of the two meter temperature with 3 patterns. The pattern (0,0).(3x1) shows a colder than usual central Arctic, Greenland, Bering Strait and East Russia, and a warmer than usual North America and Eurasia. This composite is created from 1309 days. The lowest temperatures are observed northeast of Svalbard with -4 K and the warmest temperature in central Siberia with +4 K. On the opposite site (0,2).(3x1) of this 3x1 SOM every region is warmer than usual (up to +7 K), for parts of the North Atlantic, North America, and central Siberia (-3 K). Pattern (0,1).(3x1) produces cold Eurasia and North America (-4 K) anomalies and a warmer than usual sector from the North Pacific over the Northwest Passage to Greenland and the west coast of Greenland (+2 K). The last two patterns are each a composite from 1015 days.

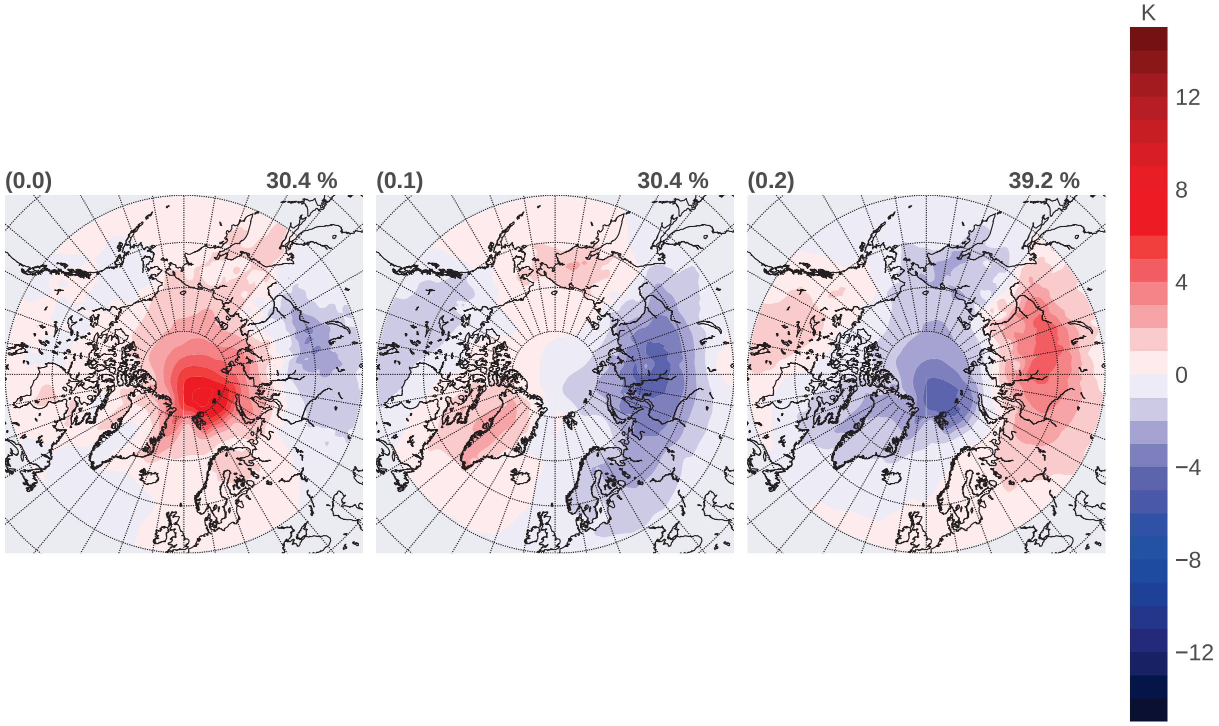

Fig. 1: 3x1 SOM of 2 meter air temperature, from ERA-Interim daily mean data for the winters from 1979/80 to 2015/16; numbers on the upper right show occurrence frequencies in percent. The numbers on the upper left show the number of the node.

For the next comparison the size of the SOM was doubled (see Figure 2). In general, going from three to six patterns creates about three new patterns and 3 patterns that can be related to the patterns from the 3x1 SOM. The range of represented days by each pattern varies from 484 to 701 for this SOM. Pattern (0,0).(3x2) is similar to pattern (0,0).(3x1) of the 3x1 SOM, but with bigger amplitudes of the anomalies (-5 K and +5 K). Similar results are found when comparing (1,1).(3x2) with (0,1).(3x1) and (1,2).(3x2) with (0,2).(3x1). New patterns identified in the 3x2 SOM are (0,1).(3x2), (1,0).(3x2), and (0,2).(3x2). Node (0,2).(3x2) shows an Arctic that is everywhere warmer than usual except for central Greenland. The new pattern (0,1).(3x2) shows a generally warmer situation while having colder regions reaching from East Russia to North Svalbard. Very cold anomalies over North Canada are shown in pattern (1,0).(3x2), which coincides with strong negative anomalies for the region north of the Barents sea and the Kara sea.
The general shape of the patterns in the corners of the following SOMs will be the same.

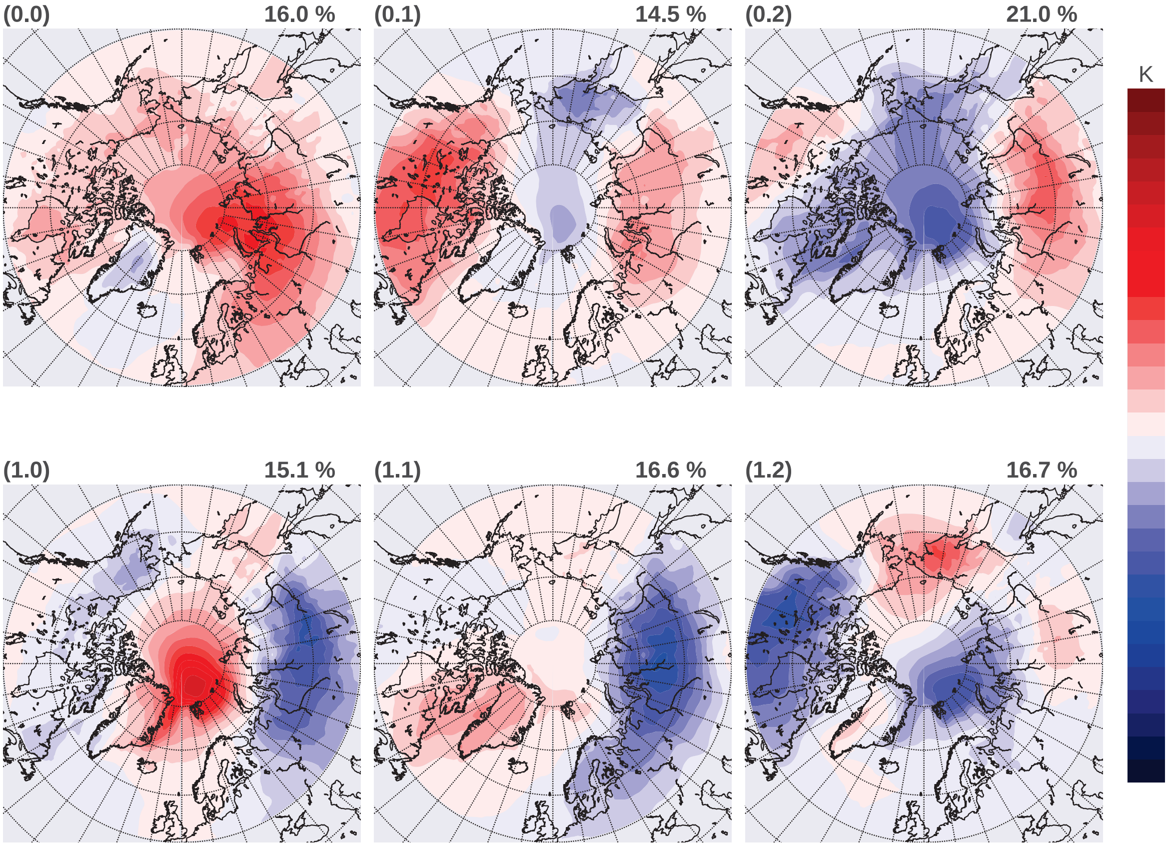

Fig. 2: As in Fig. 1 but for the 3x2 SOM.

For increasing SOM size the differences in the anomalies of the corner patterns are getting larger in space. Figure 3 shows the 2 meter temperature SOM with four by three patterns, which represents a doubling from six to twelve patterns. With this size of SOM each pattern represents about 193 to 337 days. Patterns (0,1).(4x3) and (0,2).(4x3) show that those two patterns might have developed from the pattern (0,1).(3x2). Similar can be seen for the patterns (2,1).(4x3) and (2,2).(4x3) compared to (1,1).(3x2). New patterns can be found in the new introduced row of patterns. Pattern (1,3).(4x3) shows to be very warm overall while having the center of maximal positive anomaly east of Svalbard, while pattern (1,2).(4x3) seems to be the opposite of the corner pattern in the bottom left ((2,0).(4x3) and (1,0).(3x2)). With Pattern (1,1).(4x3) a more neutral pattern has emerged while still having a colder than usual Siberia and central Arctic (-3 K) but slightly warmer North America (-3 K) and a warming of the Barents Sea (2 K). Pattern (1,0).(4x3) appears to combine features from (0,0).(3x2) and (1,0).(3x2).
Figure 4 shows an amount of twenty patterns, which is an increase of eight patterns compared to Figure 3. The amount of days represented by each pattern is in the range of 133 and 197 days. As it was stated above, the patterns in the corners are similar to the other two SOMs shown before. Most changes of patterns occur in the two middle rows compared to the 3x2 SOM. While in pattern (1,1).(4x3) data were clustered so that a warm Barents Sea was present, this pattern completely disappeared in the 5x4 SOM. One comparable new pattern would be (1,2).(5x4) but the warming is extending from the Barents Sea to the Kara Sea and northern Europe. Generally other patterns have emerged that were previously merged into other pattern of the smaller SOMs.

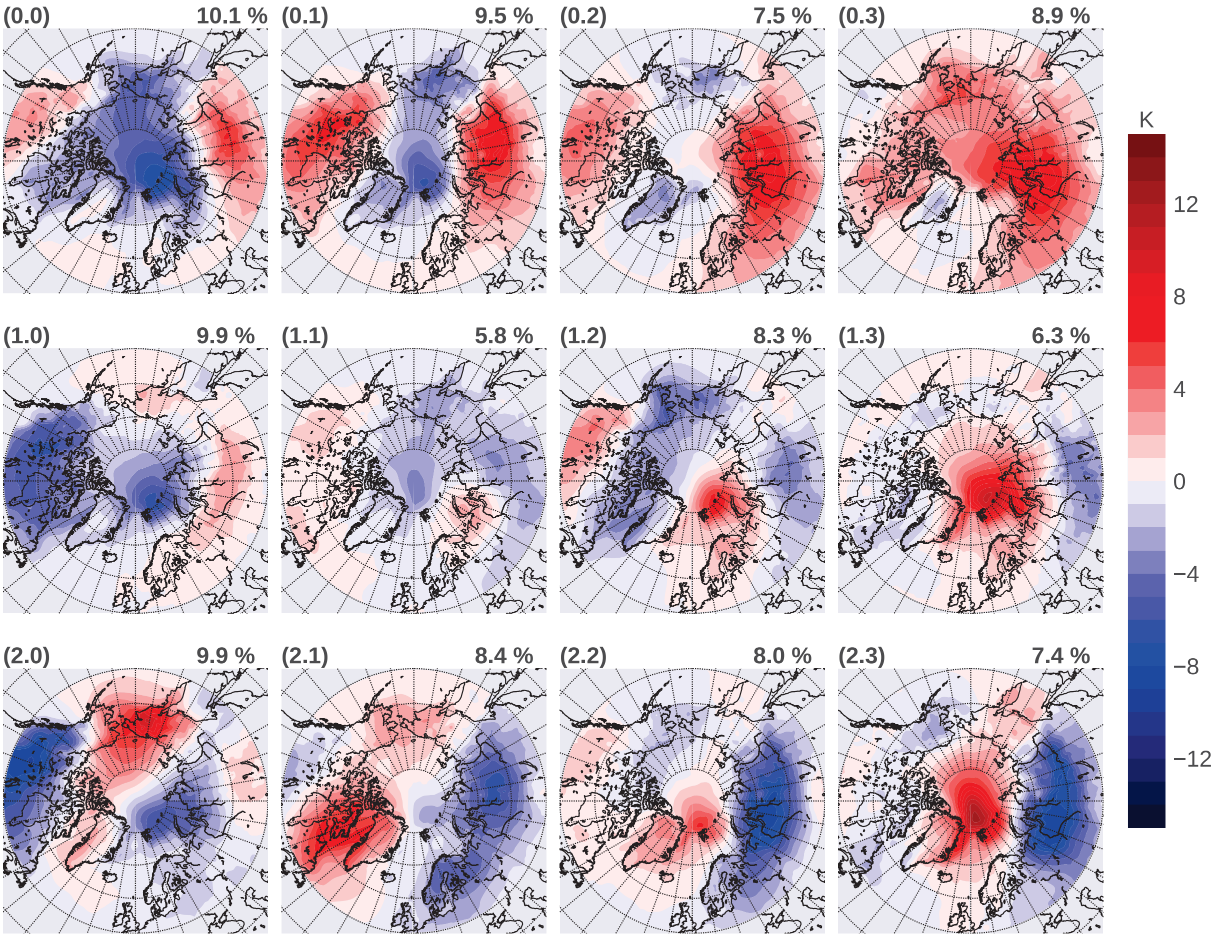

Fig. 3: As in Fig. 1 but for the 4x3 SOM.

Table 1 shows the RMSE in dependence of the number of Patterns (N). In addition to the SOMs shown here two other SOMs were created and analyzed with respect to their RMSE, one with 50 and one with 100 patterns. Generally the results show that with an increasing number of the patterns the mean RMSE is decreasing. Moreover, a linear

Table 1: RMSE depending on the number of patterns (N)

N	Mean RMSE in K
3	4.7164
6	4.5194
12	4.3435
20	4.2188
50	4.0038
100	3.8512

relationship between the RMSE and the logarithm of N has been found. After a linear fit the function for the stated relationship is as follows:

$$RMSE = (-0.566 \pm 0.012)\ \log(N) + 4.967 \pm 0.016$$

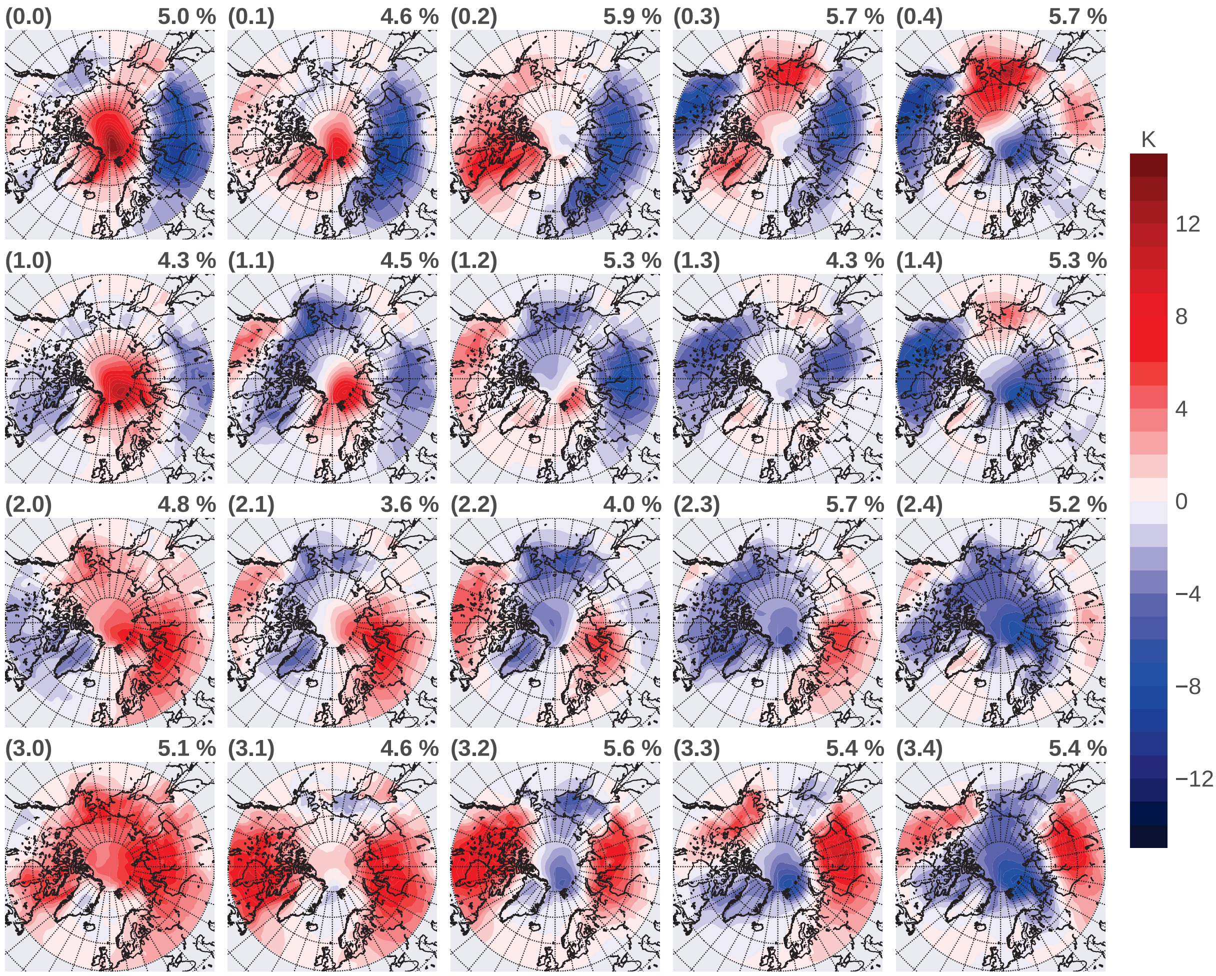

Fig. 4: As in Fig. 1 but for the 5x4 SOM.

This model between RMSE and the number of patterns results into an r^2 of 0.998 and a standard error of 0.0123. This means that an increase by 10 patterns leads to an improvement of the RMSE of 11.38% compared to the starting RMSE.

4 Summary and Discussion

We analyzed the influence of the size of a SOM on the decomposition and clustering of ERA-Interim daily mean 2 meter temperature fields north of 50°N.
It could be shown that increasing the SOM size from 3x2 on, the corner patterns remained structurally the same. This was expected due to the fact that those pattern where mathematically the most different from each other and thus had to been placed far apart from each other according to how the method works. Most new pattern were developing along an added row comparing different sized SOMs. With increasing number of patterns it could be seen that in some cases multiple patterns were evolving at the expense of a single one and these more specific cases of the temperature anomalies could be identified. This also could be seen with respect to the RMSE. With increasing number of patterns, each pattern has to represent a fewer amount of daily data and following this the general error from the pattern to the data connected to it has to decrease. This leads to the conclusion that patterns of larger SOMs can represent specific patterns much better than smaller SOMs. Through this shift of data to different/new patterns they could be partly recognized

from more general patterns of smaller SOMs.
In summary, it can be seen that the choice of the size of the SOM can change the representation of specific patterns dramatically. The general advise for using Self-Organizing Maps for distinguishing meteorological situations would be to create a fairly large SOM (with many patterns) and then re-group distinct patterns with similar meteorological features manually. This will help to understand and control better, which pattern might fit together. Leaving this grouping to the method by just simply using a smaller map size might result into composites of days that might fit mathematically well to each other, but not under a meteorological point of view. This problem occurs due to the fact that the used package to create the SOM works by simply calculating the Euclidean distances between daily fields and the patterns to assign the specific daily fields to a specific pattern.

Acknowledgements

ERA-Interim reanalyses data are provided http://apps.ecmwf.int/datasets/data/interim-full-moda. We gratefully acknowledge the support from the German Research Foundation (Deutsche Forschungsgemeinschaft) within the Transregional Collaborative Research Center (TR 172) “Arctic Amplification: Climate Relevant Atmospheric and Surface Processes, and Feedback Mechanisms $(AC)^3$.”

References

Cassano, E. N., Lynch, A. H., Cassano, J. J., and Koslow, M. R., 2006: Classification of synoptic patterns in the western Arctic associated with extreme events at Barrow, Alaska, USA, Climate Res., 30, 83–97, doi:10.3354/cr030083.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F., 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quart. J. Roy. Meteor. Soc., 137, 553–597, doi:10.1002/qj.828, URL `http://dx.doi.org/10.1002/qj.828`.

ECMWF: ERA Interim, Daily, URL `http://apps.ecmwf.int/dataset/data/interim-full-daily`, 2017.

Ford, T. W. and Schoof, J. T., 2017: Characterizing extreme and oppressive heat waves in Illinois, J. Geophys. Res. Atmos., 122, 682–698, doi:10.1002/2016JD025721.

Hewitson, B. and Crane, R., 2002: Self-organizing maps: applications to synoptic climatology, Climate Res., 22, 13–26, doi:10.3352/cr022013.

Kohonen, T., 1998: The self-organizing map, Neurocomputing, 21, 1–6, doi:10.1016/S0925-2312(98)00030-7.

Lynch, A. H., Serreze, M. C., Cassano, E. N., Crawford, A. D., and Stroeve, J., 2016: Linkages between Arctic summer circulation regimes and regional sea ice anomalies, J. Geophys. Res. Atmos., 121, 7868–7880, doi:10.1002/2016JS025164.

McDonald, A. J., Cassano, J. J., Jolly, B., Parsons, S., and Schuddeboom, A., 2016: An automated satellite cloud classification scheme using self-organizing maps: Alternative ISCCP weather states, J. Geophys. Res. Atmos., 121, doi:10.1002/2016JD025199.

Mewes, D. and Jacobi, C., 2017: Variability of horizontal temperature fluxes over the Arctic, Wiss. Mitteil. aus dem Instit. f. Meteorol. der Univ. Leipzig, 55, 51–60, URL `http://nbn-resolving.de/urn:nbn:de:bsz:15-qucosa2-167700`.