Vol. 62: 199–218, 2015
doi: 10.3354/cr01274

CLIMATE RESEARCH
Clim Res

Published online February 5

Self-organizing map analysis of widespread temperature extremes in Alaska and Canada

Elizabeth N. Cassano[1,*], Justin M. Glisan[2], John J. Cassano[3], William J. Gutowski Jr.[2], Mark W. Seefeldt[1]

[1]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309, USA
[2]Department of Geological and Atmospheric Sciences, Iowa State University, Ames, IA 50011, USA
[3]Cooperative Institute for Research in Environmental Sciences and Department of Atmospheric and Oceanic Sciences, University of Colorado, Boulder, CO 80309, USA

ABSTRACT: This paper demonstrates how self-organizing maps (SOMs) can be used to evaluate the large-scale environment, in particular the synoptic circulation associated with widespread temperature extremes. The paper provides details on how SOMs are created, how they can be used to understand extreme events, and lessons learned in applying this methodology for extremes analysis. Using a SOM can be helpful in understanding the underlying physical processes that control extreme events, and how the extremes and the processes that control them may change in time or differ across space. Examples of widespread daily temperature extremes in 4 regions: 2 each in Alaska and in northern Canada during winter (December, January, and February) for 1989–2007 are presented to illustrate the application of the methodology. For the regions studied, the size of the domain over which the synoptic circulation was defined—in particular using a smaller domain focused on particular regions—and a greater number of classes to represent the archetypical synoptic patterns for the regions, give the best relationship between synoptic circulation and extremes. The results are most robust for the Alaskan domains and less so for the Canadian domains, leading to the conclusion that further study is warranted to better understand extremes in the Canadian regions.

KEY WORDS: Self-organizing maps · Extreme events · Synoptic climatology

1. INTRODUCTION

Extreme weather events (e.g. floods, droughts, heat waves, and cold spells) can have detrimental and serious impacts on society, such as loss of life and property (IPCC 2007). One of the expected impacts of a changing climate is an increase in warm temperature extreme events (extreme events will hereafter be referred to as extremes), such as heat waves and warm nights and a decrease in cold extremes such as frost days (Tebaldi et al. 2006). The overarching goal of this research is to understand how large-scale circulation modulates widespread temperature extremes. This paper will focus on applying the methodology of the self-organizing map (SOM) technique (Kohonen 2001) to identify large-scale circulation features associated with extremes in 2 regions each in both Alaska and Canada. Using the SOM to identify relationships between extremes and large-scale atmospheric circulation allows analysis of extremes in data-sparse regions and under future climate-change scenarios, since global climate models and reanalysis products are fairly skillful in reproducing the large-scale atmospheric circulation (Randall et al. 2007, Flato et al. 2013).

Hewitson & Crane (2002) provide an overview of the application of the SOM technique to climate data. Many other studies have also used the SOM tech-

*Corresponding author: ecassano@cires.colorado.edu

nique for atmospheric analysis (e.g. Cavazos 1999, 2000, Gutowski et al. 2004, Reusch et al. 2005a,b, Cassano et al. 2006b, 2007, Lynch et al. 2006, Skific et al. 2009, Sheridan & Lee 2011 and references therein), including the study of extremes (Cassano et al. 2006a, Cavazos 1999, 2000).

The focus on relationships between large-scale circulation and extremes has been used as an approach for many other studies (e.g. Shabbar & Bonsal 2004, Kyselý 2008, Athar & Lupo 2010, Renom et al. 2011, Andrade et al. 2012), with specific attention on cluster analyses for evaluating these relationships (Maheras & Kutiel 1999, Rodríquez-Puebla et al. 2010, Heikkilä & Sorteberg 2012, Stefanon et al. 2012). Francis & Vavrus (2012) showed recent Arctic warming may be leading to slower moving and higher amplitude Rossby waves due to a reduction in the equator-to-pole temperature gradient. Slower propagation of waves can yield persistent patterns that may lead to daily extremes (Gutowski et al. 2008, Francis & Vavrus 2012, Kawazoe & Gutowski 2013a,b).

The main focus of this study is to describe the SOM methodology and its application to the evaluation of extremes. Technical details of creating a SOM and issues in applying the technique for this particular application are highlighted. Some initial results of the physical state associated with extremes are presented to illustrate the benefits and potential issues of using this methodology for this purpose.

2. DATA AND METHODOLOGY

2.1. Data

ERA-Interim (ERA-I) reanalysis (Dee et al. 2011) and regional climate model output from Version 3.1.1 of the Advanced Weather Research and Forecasting Model (WRF; Skamarock et al. 2008) for December, January, and February (DJF) 1989–2007 are used in this research. Daily mean sea-level pressure (SLP) and 2 m temperature (T2m) data are interpolated to a 50 km grid and used for all subsequent analyses. In this paper, extremes are defined using only ERA-I T2m data; extremes defined using WRF output will be used in future studies. The SOM representation of the synoptic climatology of the study area was created using daily SLP spatial anomalies from both datasets. SLP spatial anomalies were used as the basis for the synoptic climatology because the SLP gradients, rather than absolute values of SLP, are responsible for determining the near-surface circulation and therefore are of most interest in this analysis. Daily SLP anomalies were calculated for each day by subtracting each day's domain-averaged SLP from the grid point values of SLP for the same day. SLP values from locations with elevations >500 m were filtered out of the analysis due to errors associated with the reduction of surface pressure to SLP for high elevation locations (Wallace & Hobbs 1977, Mohr 2004).

2.2. Methodology

2.2.1. Extreme events

Extremes were defined as follows. The daily mean T2m at each grid point was sorted from warmest to coldest. For each grid point, days exceeding the threshold of the coldest and warmest 1% of all days were identified. Since the criteria for warm/cold extremes were applied to each grid point separately, the number of days that have grid points meeting the extreme criteria exceeds 1% of all days analyzed. For any day, if at least 25 grid points in the analysis region exceeded the warm or cold 1% threshold, this was deemed to be a widespread extreme day. The decision to sort and evaluate the data at each grid point individually allowed an analysis of what causes the coldest/warmest conditions at a particular location rather than the coldest/warmest conditions in the entire region, which would be biased towards areas that are climatologically colder or warmer (e.g. higher topography or northerly [cold]/southerly [warm] locations). Furthermore, the focus on widespread extremes was driven by the fact that a temperature extreme that occurs at a single point is more likely to be controlled by small-scale, very localized features (e.g. cold air pooling in a small valley) rather than tied to the large-scale circulation, which is the focus of the analysis presented here. There was no requirement that the grid points be adjacent, but, for most of the widespread extremes analyzed, the extreme grid points were indeed adjacent or in the same part of the region. However, there were a few days where the extreme grid points were spread out throughout the region, especially for the larger regions studied here, and these days in particular warrant further study.

The study area (the North American Arctic) for this project was divided into 4 regions (Fig. 1). Each of these regions was selected such that it is relatively homogenous in land type and will simultaneously be affected by a particular synoptic pattern (Glisan &

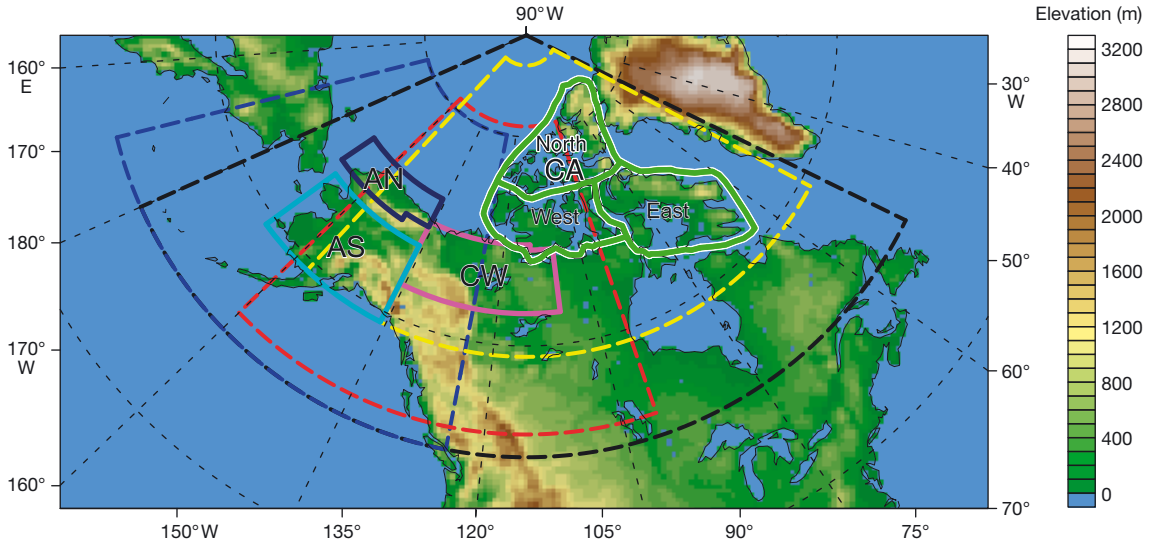

Fig. 1. Study regions—AN: Alaska north; AS: Alaska south; CW: Canada west; CA: Canada archipelago (and sub-regions in CA labeled north, west, and east as discussed in Section 3.3.3). The domains used to classify the synoptic circulation as discussed in Section 3 are outlined in dashed lines: full (black), Alaska (blue), Canada West (red), and Canada Archipelago (yellow)

Gutowski 2014). Furthermore, the size and distribution of regions allows a relatively large and contrasting number of case studies to explore, and also builds upon previous work (Cassano et al. 2006a, Glisan & Gutowski 2014). Only extremes over land grid points were considered. The first region is Alaska north (AN, 83 grid points), largely north of the Brooks Range. The second region is Alaska south (AS, 413 grid points), largely south of the Brooks Range, excluding the Alaskan panhandle and most of the Aleutian Island chain. The other 2 regions are in northern Canada. Canada west (CW, 430 grid points) is in northwestern Canada between 141 and 107°W and north of 63°N. Canada archipelago (CA, 602 grid points) encompasses most of the Canadian archipelago in addition to 2 peninsulas from mainland Canada that are in the same vicinity as the southern portion of the archipelago.

The synoptic climatology for each of the regions was defined using the SOM algorithm applied to SLP anomalies for a domain larger than the individual regions to fully encompass the broad synoptic circulation that affects each region. As will be outlined in detail in Section 3, there were 4 domains in total used to define synoptic climatology, the outlines of which are shown in Fig. 1. Throughout the paper, when 'region' is referenced, this refers to the specific areas over which extremes are defined (i.e. AN, AS, CW, CA). 'Domain' refers to the area over which the synoptic circulation is defined in the SOM classification, described in greater detail in the following section.

2.2.2. Self-organizing maps

The SOM algorithm (Kohonen 2001; software available at www.cis.hut.fi/research/som_lvq_pak.shtml) is used to define the synoptic climatology for the study area and to relate the extremes to the large-scale atmospheric circulation. The SOM algorithm employs a neural network method that uses unsupervised learning to determine generalized patterns in data. This technique reduces the dimension of large data sets by grouping similar data records together and organizing them into a 2-dimensional array that becomes a mapping of the pattern space occupied by the input data. Used in this way the SOM algorithm may be considered a clustering technique, but, unlike other clustering techniques, the SOM method does not need *a priori* decisions on data distribution and provides for smooth transitions from one pattern to another.

While principal component analysis (PCA) is common in climate research, Reusch et al. (2005a) found that a SOM-based analysis correctly identified known spatial patterns, while the PCA extraction, even with component rotation, was less capable of extracting known spatial patterns in the data, mixed patterns into single components, and tended to incorrectly partition variance among the components. As a result, the analysis presented here uses SOMs to identify the large-scale circulation patterns associated with extremes.

To develop the SOM array, the SOM patterns (which will be referred to as nodes for the remainder

of the paper) are initialized either randomly or using the 2 principal eigenvectors of the input data (the latter are used in this study). Tests were performed to determine if the final SOM differed using the 2 initializations, and the final results were fundamentally the same. Then each input data sample (i.e. gridded daily mean SLP anomaly) is presented to the SOM and the node to which it most closely matches (as measured by the Euclidean distance) becomes the 'winning' node. This node gets nudged towards the input sample, as do surrounding nodes to a lesser degree further away from the winning node. The resulting classification is thus organized such that similar nodes are located in the same portion of the array, with contrasting nodes on opposite sides of the array and in between a transition between the 2.

As with other clustering analyses, one of the critical choices to be made when using the SOM algorithm is how many nodes classify the input dataset. A balance between a tractable number of nodes and enough nodes to adequately represent the input data is, as with any synoptic classification, the goal of this choice. Choosing a larger number of nodes will provide greater detail and thus greater discrimination of circulation patterns, while a smaller number of patterns will represent the archetypical patterns of the area studied with few details of those patterns. The SOM technique differs from cluster analysis in that it identifies points in the data space that are representative of the surrounding data rather than grouping the data (Huth et al. 2008). The goal of the technique is not completely distinct patterns, but rather patterns that vary smoothly across the data space. The addition of more patterns allows a refinement of how the final patterns characterize the input data. Furthermore, the 2-dimensional architecture of the SOM array of patterns allows an ease of visualization, due to the organized nature of the final classification (i.e. each pattern in the SOM classification is similar to those around it) and therefore allows a greater number of patterns to be used to characterize the input data than is typically seen with other synoptic typing methods (Sheridan & Lee 2011).

For SOM training, there are 3 user-defined parameters that guide the training and resultant classification: number of iterations, learning rate, and neighborhood radius. The number of iterations is how many times the input samples are presented to the SOM algorithm. The learning rate determines the nudging strength of the winning node towards the input sample. This value gradually decreases to zero (no nudging) as training progresses. The neighborhood radius determines how many nodes surrounding the winning node get nudged towards the input sample. As training proceeds, this number gradually decreases to 1 (only the winning node is nudged towards the input sample at the end of the training).

One of the metrics by which to evaluate the final SOM classification is the quantization error (Q-error), a standard output from the software, which is proportional to the sum of the squared differences between the input data and the nodes to which they map. In creating one of the SOM classifications described in more detail in the next section, 180 tests were performed with differing numbers of neighborhood radii, learning rates, and number of iterations. Eighty percent of the tests had Q-errors within 2% of the SOM with the smallest Q-error. The tests that gave the largest Q-errors were those with a combination of a small neighborhood radius, large learning rate, and small number of iterations. SOM training using a larger radius (approximately equal to the smaller dimension of the SOM array), smaller learning rate, and larger number of iterations gave the smallest Q-errors. Using a small radius reduces the number of the surrounding nodes that are nudged towards the input sample. This then does not fully allow all of the surrounding nodes to self-organize, particularly if the number of iterations is small. A large learning rate indicates that for each time an input sample is presented to the SOM, the SOM node which it most closely matches is nudged to a large degree and is potentially forcing the node too strongly toward the most recent input sample. Again, combining this with a small number of iterations will not allow the SOM to fully represent the input data. Importantly, the resultant SOM classifications using different training parameters largely gave the same general types of patterns, indicating an overall lack of sensitivity to these parameters.

Alexander et al. (2010) performed tests using different numbers of SOM nodes, learning rates, and neighborhood radii. The synoptic classification was performed for Australia, which has a synoptic climatology significantly different than that evaluated in the present study, though the methodology can be applied in both regions. Root mean squared differences (RMSDs) were larger for small SOMs (6 and 12 nodes) than for SOMs with 20, 30, and 42 nodes; for the latter they found nearly identical RMSDs. Visually they determined that a 20-node SOM was optimal. Using fewer nodes did not encompass the full range of synoptic patterns that affected the area, while a SOM with a greater number of nodes included synoptic patterns that occurred infrequently for their area of study and made statistical analysis of

the results less robust. The impact of SOM array size and varying training parameters on SOM creation and the analysis of extreme events will be evaluated later in this paper.

Once the final SOM array has been created, the training data are 'mapped', that is, each daily sample is compared with each SOM node to find the node to which that input sample best matches. This results in a list of samples associated with each node, and this list allows one to determine the frequency of occurrence of each node, which nodes are responsible for extremes, and to associate other fields to the SOM for the samples that correspond to each node.

2.2.3. Evaluation methods

To evaluate how well the different SOM training trials represent the ERA-I SLP anomaly data, several quantitative measures were calculated. One of these metrics is the Q-error described in the previous section. Similar to the Q-error is the RMSD between each daily SLP anomaly field and the closest matching SOM pattern. For the remainder of the paper the RMSD rather than the Q-error will be highlighted, since this is a more commonly used measure of similarity between datasets. The RMSD can be calculated from the Q-error by dividing the Q-error by the number of grid points in the input data, and then taking the square root. Mean absolute difference (MAD) is calculated between each daily SLP anomaly field and the closest matching SOM pattern. The spatial correlation compares each point of gridded daily SLP anomaly with the corresponding point in the SOM node which it matched. SOMs with low RMSD and MAD, and high correlation indicate a good match between the SOM classification and the input data.

Another measure of the quality of the SOM training is how adjacent nodes relate to each other; this is an important consideration for subsequent analysis (Hewitson & Crane 2002). One way to visualize this relationship is Sammon mapping, which allows higher dimensional data to be visualized in a 2-dimensional array (Sammon 1969). The Sammon algorithm produces a grid that is an approximate projection onto 2-dimensional space showing the Euclidian distances between each SOM node and its neighbors. Fig. 2 shows the Sammon map of 2 of the SOMs created for testing here; the separation between points, which represent individual nodes in the SOM, shows the approximate relationship between the neighboring circulation maps.

The geometric shape of the Sammon map is important when determining how each node is related to every other node. A flat Sammon map indicates little distortion and is preferred for interpreting the physical behavior depicted on the SOM. A 'twistedness index' (TI) was developed as part of this work to provide a quantitative measure of the 'flatness' of the Sammon map. A flat Sammon map would produce a TI = 1 and a more distorted map would have an appreciably larger TI (Fig. 2). This index also provides a good counterweight to the RMSD, as SOM training that yields the smallest RMSD often has a larger TI (more details provided in the next section). Details of the TI calculation are provided in the Appendix.

Assessment of how well the SOM identifies unique circulation patterns for extreme events was quantified based on the percentage of nodes identified as having either cold or warm extremes and the percentage of nodes having both cold and warm extremes. Lower values of these metrics indicate that the SOM is able to distinguish unique circulation patterns associated with extreme events. In addition, a calculation was performed to determine if the distribution of SOM nodes associated with extremes could occur randomly in the SOM space. For this calculation, the total frequency of occurrence of days that mapped to nodes on which extreme events occurred was used to find the probability that a random distribution of days matching the number of days for which either warm or cold extremes occurred for a particular region would fall only on the nodes associated with extremes. For example, there were 20 cold extremes for the AN region; these are shown in the cells in bold print in Table 1. The percentages in this table are the frequency of occurrence for all days that mapped to the original SOM used for the analysis (see Fig. 3; described in more detail in the following section). The frequency of occurrence for only the nodes in which extremes occurred were summed,

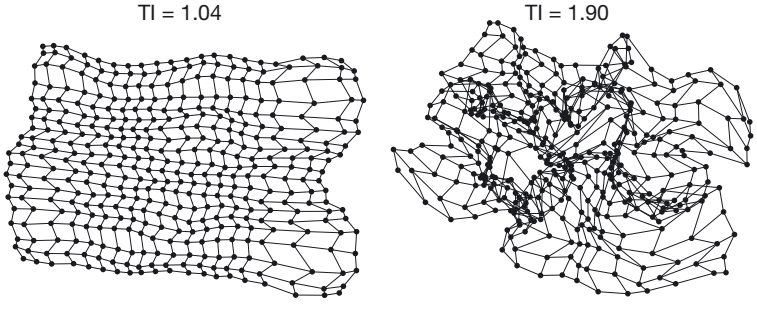

Fig. 2. The 20 × 20 Sammon maps showing the twistedness index (TI) for relatively flat (left) and twisted (right) geometries

Table 1. Frequency of occurrence (percentage) for all days that mapped to the original 5 × 4 SOM shown in Fig. 3. Each cell in this table corresponds directly with a node in Fig. 3. **Bold**: nodes on which cold extremes occurred for the Alaska north region

7.47	5.13	5.95	5.6	4.67
5.37	4.67	5.48	**4.49**	4.2
6.24	**3.56**	**3.85**	4.2	**4.14**
7.29	**4.55**	3.85	3.85	**5.43**

and this value was used to calculate the probability that a random sampling of days would only fall on these nodes. For this example, the total frequency of occurrence for the nodes on which extremes occur is 44.92%. The probability calculation in this example is then $44.92 / 100^{20} = 1.1 \times 10^{-7}$. The results from this calculation show the probability that a random sampling of days would only fall on nodes on which extremes occurred in general is very small (ranging from 1×10^{-4} to 1×10^{-18}). However, there are some cases in which the probability is not small, and those results will be discussed in the next section.

The overlap coefficient (OC; Inman & Bradley 1989) was calculated to assess the statistical significance of the difference between frequency distribution functions (FDFs) for the SOM climatology versus extremes. The OC was used to test the null hypothesis that the extremes' FDF is merely a random sampling of climatology. From the climatology FDF, we take N random draws (with replacement), where N is the number of extremes in a given region, and compute the OC. A Monte Carlo approach (Metropolis & Ulam 1949) was used to determine the likelihood that the extremes' OC is an outcome of a random sample of climatology, thus arriving at an estimate of statistical significance.

3. RESULTS

3.1. SOM array size

Initial analysis relating synoptic circulation to widespread temperature extremes used a SOM with 20 nodes (5 × 4 array) covering a domain that encompassed all 4 regions (referred to as the full SOM domain, Figs. 1 & 3). During DJF, every node had some aspect of either the Aleutian Low and/or the Icelandic Low, which are dominant modes of synoptic variability in these regions representing the preferred storm tracks in the North Pacific and North

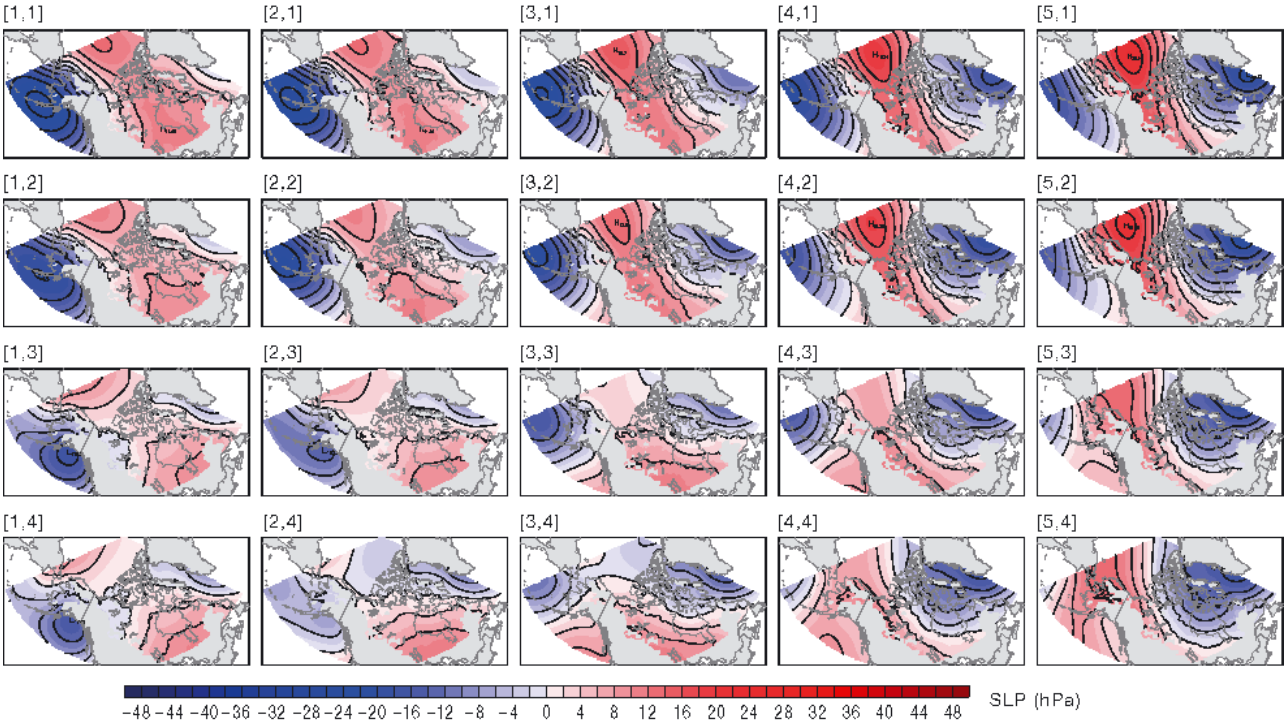

Fig. 3. The 5 × 4 sea-level pressure (hPa) anomaly full domain self-organizing map (SOM) representation of the synoptic circulation patterns impacting all 4 regions. Areas shaded in gray represent locations above 500 m that were filtered out of the analysis. Contours are color shaded every 2 hPa and black lines every 4 hPa

Atlantic Oceans, respectively (Trenberth & Paolino 1981, Serreze et al. 1997, Rodionov et al. 2005). In addition, many nodes also included high pressure in the Beaufort Sea, which is also an important feature of the synoptic circulation during winter for this area (McBean et al. 2005, Overland 2009).

Initial evaluation of nodes associated with extremes showed a large level of overlap between warm and cold extremes (i.e. warm and cold extreme days mapping to the same SOM node, particularly for the Canadian regions; Table 2, Fig. 4). Furthermore, the extremes for the Canadian regions were spread over the entire SOM space (resulting in a 100% probability of any randomly chosen day falling on a SOM node identified as an extreme node for the CA region, and for warm extremes for the CW region) though there was some clustering of both warm and cold extremes (e.g. the lower right corner of the SOM for cold extremes for CA, which are patterns with northerly flow on the backside of the Icelandic Low over much of the CA region). Nodes in which both warm and cold extremes occur do not necessarily indicate a failure to discriminate the patterns leading to these events; this aspect of these results is discussed in Section 3.1. & 3.4.

The substantial occurrence of warm and cold extremes on the same node, and the spreading of extremes over much of the SOM space suggested that this SOM was not adequately identifying differences in the SLP field responsible for different types of extremes and thus was not likely to be useful for

Table 2. Percent of SOM nodes that contain warm, cold, and both cold and warm extremes for each of the regions over which extremes are defined as described in Section 2.2.1 (AN: Alaska north; AS: Alaska south; CW: Canada west; CA: Canada archipelago). The first column of the table indicates SOM array size (5 × 4 or 7 × 5) and domain (full, Alaska, CW, or CA). For the CA region, node counts are also listed for the north, west, and east sub-regions as described in Section 3.3. **Bold**: lowest (best) percentages for each region

SOM	Nodes with warm extremes	Nodes with cold extremes	Nodes with both warm and cold extremes
AN region			
5 × 4, full	25.0	45.0	10.0
7 × 5, full	**20.0**	**25.7**	**2.9**
5 × 4, Alaska	25.0	35.0	10.0
7 × 5, Alaska	**20.0**	28.6	**2.9**
AS region			
5 × 4, full	75.0	65.0	40.0
7 × 5, full	62.9	51.4	31.4
5 × 4, Alaska	75.0	60.0	35.0
7 × 5, Alaska	**57.1**	**47.7**	**20.0**
CW region			
5 × 4, full	100.0	90.0	90.0
7 × 5, full	68.6	77.1	57.1
5 × 4, CW	85.0	80.0	65.0
7 × 5, CW	71.4	**62.9**	37.1
7 × 5, CW 850 hPa	**57.1**	74.3	**31.4**
CA region			
5 × 4, full	100.0	100.0	100.0
7 × 5, full	77.1	85.7	65.7
5 × 4, CA	95.0	100.0	95.0
7 × 5, CA	85.7	88.6	74.3
7 × 5, CA north	**28.6**	**14.3**	**2.9**
7 × 5, CA west	37.1	34.3	14.3
7 × 5, CA east	**28.6**	60.0	8.6

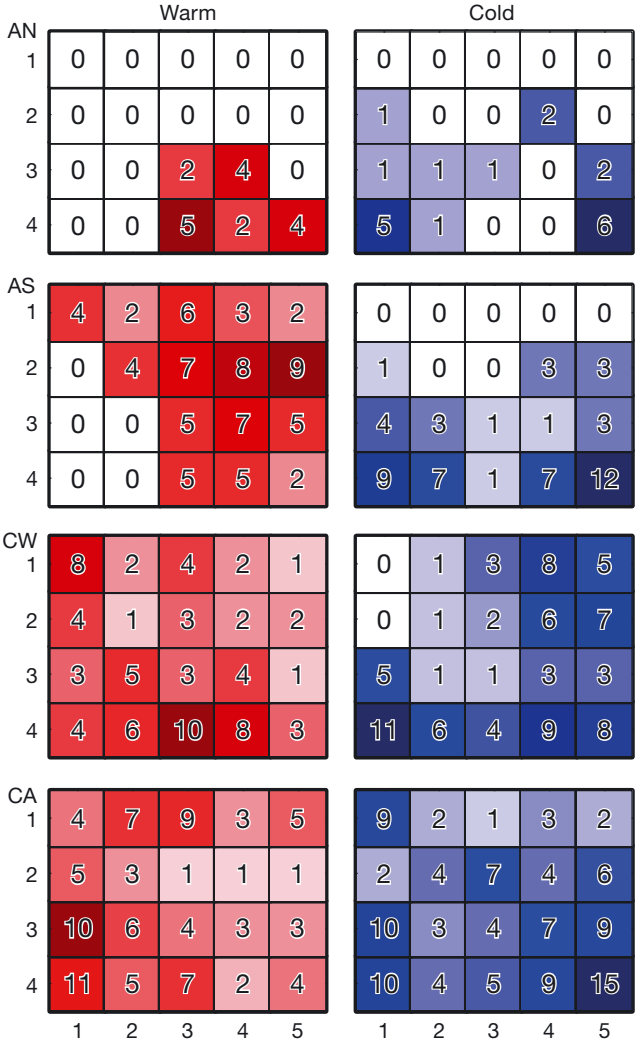

Fig. 4. Number of extreme days for each SOM node for the Alaska north (AN; upper row), Alaska south (AS; second row), Canada west (CW; third row), and Canada archipelago (CA; fourth row) regions for the 5 × 4 full domain SOM. The left (right) columns represent warm (cold) extremes. Darker (lighter) shading indicates more (fewer) extreme days on that node. The boxes in this figure correspond directly with the SOM nodes shown in Fig. 3

this research. The shortcomings of the initial SOM were thought to be related to the large domain and relatively small number of SOM nodes, both of which result in a SOM that is too generalized. The patterns lacked smaller scale features that may be relevant for extremes occurring on scales of ~250 × 250 km (equivalent to the twenty-five 50 km grid point criteria used to identify widespread extremes). Therefore, the level of generalization in the SOM patterns did not match the intended application of identifying differences in circulation associated with different types of extremes.

Ultimately the aim of atmospheric circulation classification is to find a balance between a tractable number of patterns that define the climatology of the study area and a sufficient number of patterns to identify variances in circulation types relevant to the analysis being performed. Additional SOM tests were undertaken to try to alleviate the issues described in the previous paragraph and to find a SOM more suitable for the goals of this research. SOMs containing between 12 and 400 nodes were created to test how well the SOM nodes matched the input data by the metrics of a low RMSD and TI.

Results show that, as the node count increases towards 100, the RMSD drops rapidly (Fig. 5). Above a node count of around 100, RMSD sensitivity levels off, while TI levels increase at a greater rate for node counts exceeding 50. Additionally, SOMs with node counts above 100 show little difference in their outer few rows and columns, with the interior nodes showing similar circulation patterns found in SOMs with a smaller number of nodes. Thus, if the array count is too high the training produces near-duplicate columns and rows of exterior nodes. This suggests that increasing the SOM size for this dataset and domain size beyond approximately 100 nodes provides little additional value. Thus, RMSD becoming insensitive to increasing SOM size appears to be a good indicator of maximum useful array size.

As described above, the selection of a SOM array most appropriate for a particular application involves not only choosing a sufficient number of nodes to represent circulation types relevant to the analysis but also a tractable number of nodes to facilitate visualizing the results. A 35 node SOM (7 × 5 array) was chosen as giving an appropriate balance between ease of visualization and reasonably low RMSD and TI (Fig. 5**).** This particular SOM array size has been used in previous climate studies (e.g. Hewitson & Crane 2002, Cassano et al. 2006b, 2011, Lynch et al. 2006, Schuenemann et al. 2009). Compared to the 5 × 4 SOM, this larger SOM (shown in Fig. A1 in the Appendix) has the same types of circulation patterns (i.e. Aleutian and Icelandic Lows, Beaufort High), but with additional fine-scale details distinguishing these main modes of variability. All of the metrics described in Section 2.2.3 were improved for the 7 × 5 SOM compared with the 5 × 4 SOM, with higher correlations, smaller RMSD and MAD, and a reduced percentage of nodes having overlap between both warm and cold extremes (Tables 2 & 3). These results indicate that the larger SOM was better able to capture the distinct circulation patterns associated with warm versus cold temperature extremes, although there was still considerable overlap, particularly for the Canadian regions. Thus, it was found that increasing the SOM array size to reasonable levels is not always sufficient to ensure that patterns identified by the SOM are adequate for meeting the research goals.

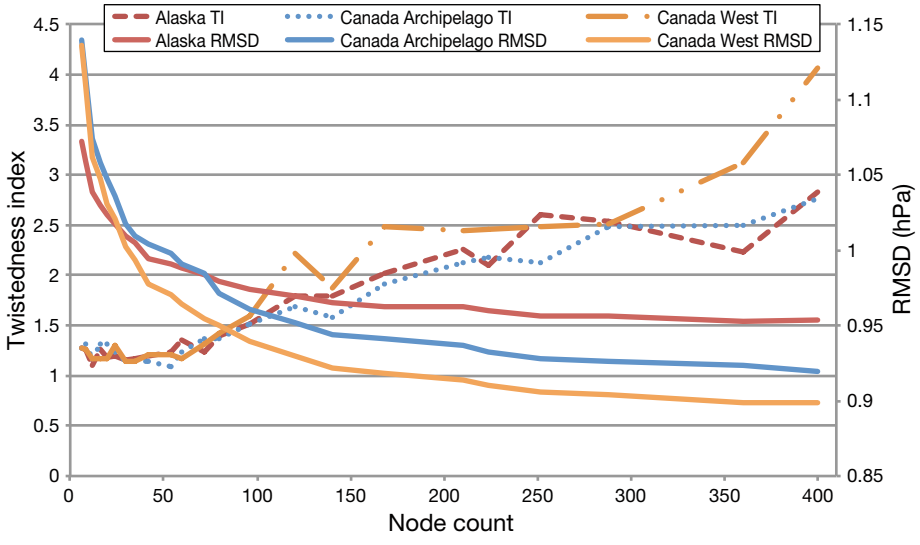

Fig. 5. Root mean squared difference (RMSD; solid lines) and twistedness index (TI; broken lines) versus number of SOM nodes for the Alaskan (red), Canada west (orange), and Canada archipelago (blue) domains

Table 3. Metrics of fit between the daily sea-level pressure anomaly and their respective SOM nodes for all days. RMSD: root mean squared difference; MAD: mean absolute difference. Units for RMSD and MAD are hPa. **Bold**: highest correlation and lowest RMSD and MAD for each region

SOM	Correlation	RMSD	MAD
5 × 4, full	0.76	8.34	6.52
7 × 5, full	0.78	7.94	6.21
5 × 4, Alaska	0.84	7.42	5.80
7 × 5, Alaska	**0.86**	**6.78**	**5.27**
5 × 4, Canada archipelago	0.80	6.16	4.73
7 × 5, Canada archipelago	**0.83**	**5.77**	**4.41**
5 × 4, Canada west	0.83	6.29	4.89
7 × 5, Canada west	**0.86**	**5.77**	**4.49**

3.2. Domain size

An important consideration in the creation of a synoptic circulation climatology is the domain over which to characterize the circulation. A domain should be chosen that highlights the pertinent circulation characteristics that impact the features to be studied. In particular, areas with high variability that are not relevant for the research questions being addressed should be excluded from the domain to avoid having those areas dominate the SOM training.

Given this and the results from the previous section, it was decided to create 6 new SOMs using both 5 × 4 (not shown) and 7 × 5 SOM arrays covering 3 smaller domains (Fig. 1): 1 focused on both Alaskan regions (Fig. 6a) and 2 separate SOMs for each of the Canadian regions (Fig. 6b [CW] and Fig. 6c [CA]). Visual comparison of the full-domain 5 × 4 SOM (Fig. 3) and the smaller domain 7 × 5 SOMs (Fig. 6) show that the latter better represent finer details of the archetypal patterns such as location and strength of cyclones that are likely important for extreme events. This conclusion is supported by the metrics defined in Section 2.2.3 which show that the 7 × 5 SOMs focusing on the individual domains have the highest correlations and lowest RMSD and MAD

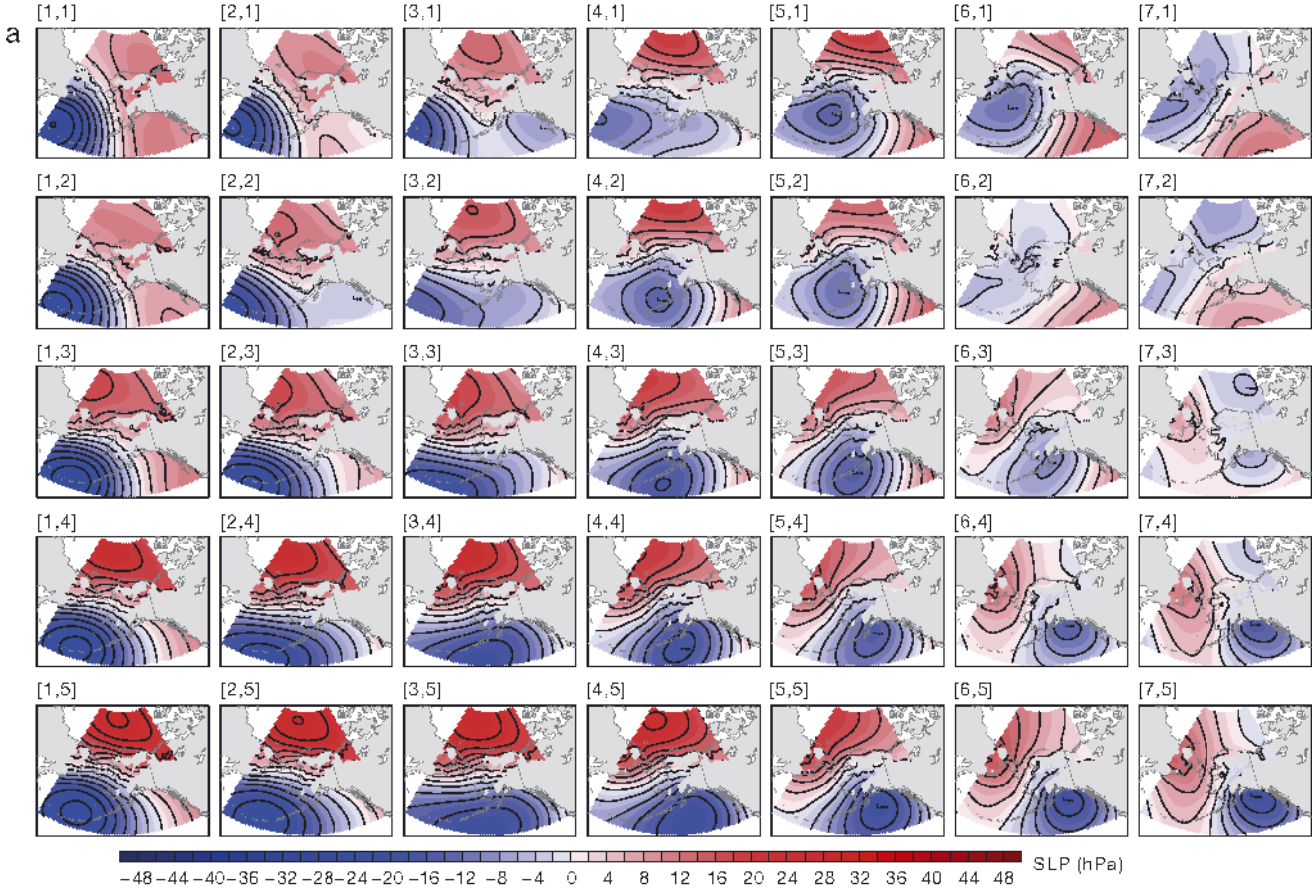

Fig. 6 (continued on next page). The 7 × 5 sea-level pressure (hPa) anomaly individual domain SOM representation of the synoptic circulation patterns for the (a) Alaska, (b) Canada west, and (c) Canada archipelago domains. The areas shaded in gray represent locations above 500 m that were filtered out of the analysis. Contours are color shaded every 2 hPa and black lines every 4 hPa

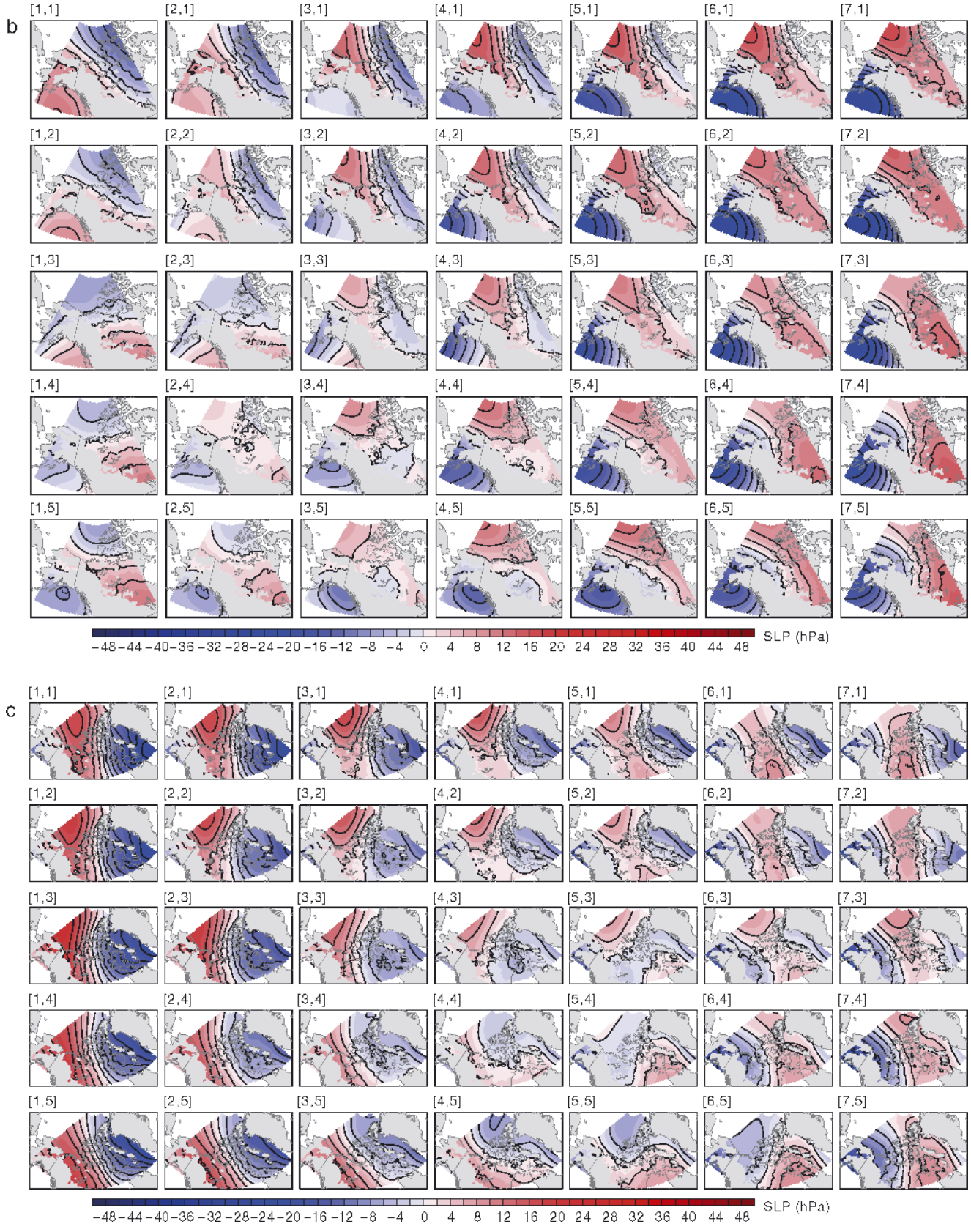

Fig. 6 (continued)

(Table 3). These calculations were also performed using only days that were defined as extreme for warm and cold extremes separately. Again, the correlations were highest and errors lowest (except for warm extremes in the AN region) for the 7 × 5 individual domain SOMs (Table 4).

The 7 × 5 individual domain SOMs usually have the lowest percentage of nodes associated with cold and warm extremes (Table 2, Figs. 4 & 7). For the AN and AS regions the percentage of nodes with either cold or warm extremes ranged from 25 to 75% on the 5 × 4 full SOM domain and decreased to 20–57.1% on the 7 × 5 Alaska SOM domain. For the CW and CA regions the percentage of nodes with either cold or warm extremes ranged from 90 to 100% on the 5 × 4 full SOM domain and decreased to 62.9–88.6% on the 7 × 5 individual SOM domains. The high percentage of nodes identified as being associated with warm or cold extremes in the 7 × 5 Canadian domain SOMs indicates that these SOMs may still not be ideal for the intended application.

The OC described in Section 2.2.3 was calculated to determine the amount of overlap between the extreme and non-extreme frequency distributions (Table 5). Smaller values of this calculation indicate fewer circulation patterns are responsible for extreme events. Higher values indicate multiple regions of SOM space (i.e. many more nodes) are accessed at a similar percentage. These results show, in agreement with the frequency results above, that the Canadian regions have a higher level of overlap between extreme and non-extreme frequencies than the Alaskan regions. A bootstrap method was used to test the hypothesis that the frequency distribution of extremes was simply the outcome of a random sampling of climatology. For all cases the extremes' OC was significantly different from climatology at the 99% level.

Table 5. Overlap coefficient (as described in Section 2.2.3) for each of the analysis regions over which extremes are defined (as described in Section 2.2.1) applied to the 7 × 5 individual SOMs

	Alaska north	Alaska south	Canada west	Canada archipelago
Cold	0.31	0.43	0.58	0.77
Warm	0.20	0.48	0.55	0.68

Table 4. Metrics of fit between the daily sea-level pressure anomaly and their respective SOM nodes for extreme days only for each of the regions over which extremes are defined as described in Section 2.2.1 (AN: Alaska north; AS: Alaska south; CW: Canada west; CA: Canada archipelago). The first column of the table indicates SOM array size (5 × 4 or 7 × 5) and domain (full, Alaska, CW, and CA). RMSD: root mean squared difference; MAD: mean absolute difference. Units for RMSD and MAD are hPa. **Bold**: highest correlation and lowest RMSD and MAD for each region

SOM	Warm extremes			Cold extremes		
	Correlation	RMSD	MAD	Correlation	RMSD	MAD
AN region						
5 × 4, full	0.70	8.81	6.80	0.65	8.01	6.32
7 × 5, full	0.76	**8.07**	**6.28**	0.67	7.85	6.21
5 × 4, Alaska	0.75	9.04	7.04	0.74	7.95	6.39
7 × 5, Alaska	**0.78**	8.14	6.37	**0.82**	**6.74**	**5.30**
AS region						
5 × 4, full	0.78	8.81	6.96	0.59	8.97	6.99
7 × 5, full	0.80	8.36	6.60	0.64	8.56	6.72
5 × 4, Alaska	0.84	8.02	6.29	0.73	7.93	6.17
7 × 5, Alaska	**0.86**	**7.52**	**5.84**	**0.78**	**7.24**	**5.65**
CW region						
5 × 4, full	0.71	9.03	7.14	0.70	8.75	6.82
7 × 5, full	0.74	8.62	6.82	0.73	8.34	6.53
5 × 4, CW	0.74	7.48	5.89	0.79	6.68	5.20
7 × 5, CW	**0.79**	**6.80**	**5.33**	**0.82**	**6.07**	**4.71**
CA regions						
5 × 4, full	0.69	9.31	7.29	0.73	8.46	6.58
7 × 5, full	0.72	8.88	6.95	0.77	7.99	6.20
5 × 4, CA	0.69	7.22	5.54	0.82	6.20	4.67
7 × 5, CA	**0.75**	**6.60**	**5.06**	**0.84**	**5.84**	**4.39**

The amount of overlap between the warm and cold extremes is reduced for the Alaska domain SOM, and to a much lesser degree for the Canadian domain SOMs (Table 2, Figs. 4 & 7). For the AN (AS) region the percentage of nodes with both warm and cold extremes decreased from 10% (40%) to 2.9% (20%) between the 5 × 4 full domain and 7 × 5 Alaska domain SOM. For the CW (CA) region the percentage of nodes with both warm and cold extremes decreased from 90% (100%) to 37.1% (74.3%) between the 5 × 4 full domain and 7 × 5 CW (CA) domain SOM. These results suggest that the 7 × 5 Alaska domain SOM is suitable for our intended application, but the CW and CA SOMs may not be ideal.

Thus, during SOM training, the selection of both SOM array size and domain is critical for creating a SOM that matches a project's research goals, but, in some cases, such as the Canadian domains, additional factors must be considered in creating appropriate SOMs. Given the results presented in this section all further analysis will focus on the 7 × 5 individual domain SOMs.

3.3. SOM training variable and examples of synoptic circulation patterns associated with temperature extremes

The selection of the variable with which to train the SOM is another key decision in applying this methodology. SLP or SLP anomaly was the most common variable used in previous SOM studies (e.g. Hewitson & Crane 2002, Cassano et al. 2007). For the current study SLP anomalies were selected to determine near-surface circulation patterns, as it was anticipated that these patterns play a large role in the resultant extremes. Results relating the SOM synoptic circulation patterns to widespread temperature extremes are shown in the following sub-sections to demonstrate the technique and to highlight particular issues of note in the use of SLP anomalies.

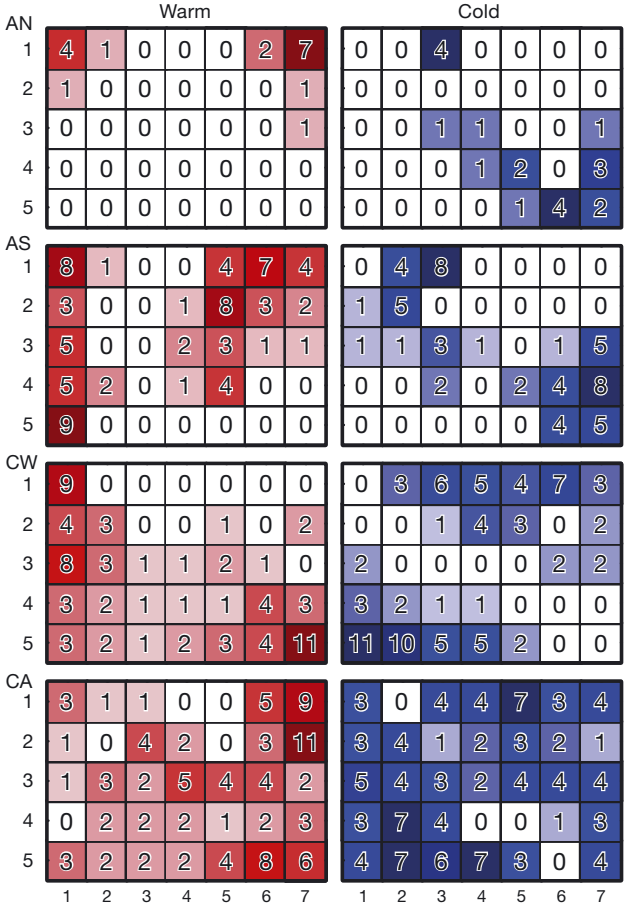

Fig. 7. Number of extreme days for each SOM pattern for the Alaska north (AN; upper row), Alaska south (AS; second row), Canada west (CW; third row), and Canada archipelago (CA; fourth row) regions for the 7 × 5 individual domain SOMs. The left (right) columns represent warm (cold) extremes. Darker (lighter) shading indicates more (fewer) extreme days on that node. The boxes in this figure correspond directly with the SOM nodes shown in Fig. 6

3.3.1. Alaska south

There are 74 d in this region with warm extremes over the time period of analysis. For the 7 × 5 Alaskan domain SOM these extreme days map to 20 (57%) of the nodes, and there are 2 primary clusters of nodes that are associated with warm extremes (Fig. 7). One is in the upper right corner of the SOM, which contains circulation patterns with low pressure just to the west of Alaska. The other cluster is along the left side of the SOM, which includes approaching strong Aleutian Lows. The primary circulation patterns for both of these node clusters result in strong southerly flow and warm air advection (WAA) over Alaska (Fig. 6a).

There are 55 d with cold extremes for the Alaska south region, and these days map to 16 (46%) of the nodes in the 7 × 5 Alaskan domain SOM. More than half of the cold extreme days fall on nodes in the lower right corner of the SOM (Fig. 7) which contain Gulf of Alaska low-pressure systems and high pressure in the Chukchi Sea, all of which have a northerly flow component into Alaska and associated cold air advection (CAA) (Fig. 6a). There is another cluster in the upper left corner of the SOM that has circulation patterns with Gulf of Alaska low pressure/Chukchi Sea high pressure, but an approaching Aleutian Low far to the west. Some of these circulation patterns also show a broad band of high pressure over Alaska, with a weak pressure gradient over the domain, and are likely dominated by the development of strong inversions, favored by the weak flow, and the strong radiative cooling that yields these types of cold extremes.

Further details on the different synoptic patterns that lead to cold extremes in AS are shown in Fig. 8. The upper panels in this figure show composites of the SLP anomaly patterns for all cold extreme days (left) and all non-extreme days (middle). In general, days with cold extremes have a broad high-pressure system over most of Alaska, with low pressure centered in both the Gulf of Alaska and west of the Aleutian Island chain. Non-extreme days show the overall typical synoptic pattern during DJF in this area, which contains Aleutian low- and high-pressure systems in the Beaufort and Chukchi Seas. The extreme minus non-extreme composite (upper right) shows higher pressure over most of Alaska, with lower pressure in the Alaskan panhandle and Canada just southeast of Alaska for days that are extreme. Days that map to 2 nodes from different parts of the SOM space were chosen to demonstrate the different types of synoptic patterns that can be associated with cold extremes for the AS region and the additional in-

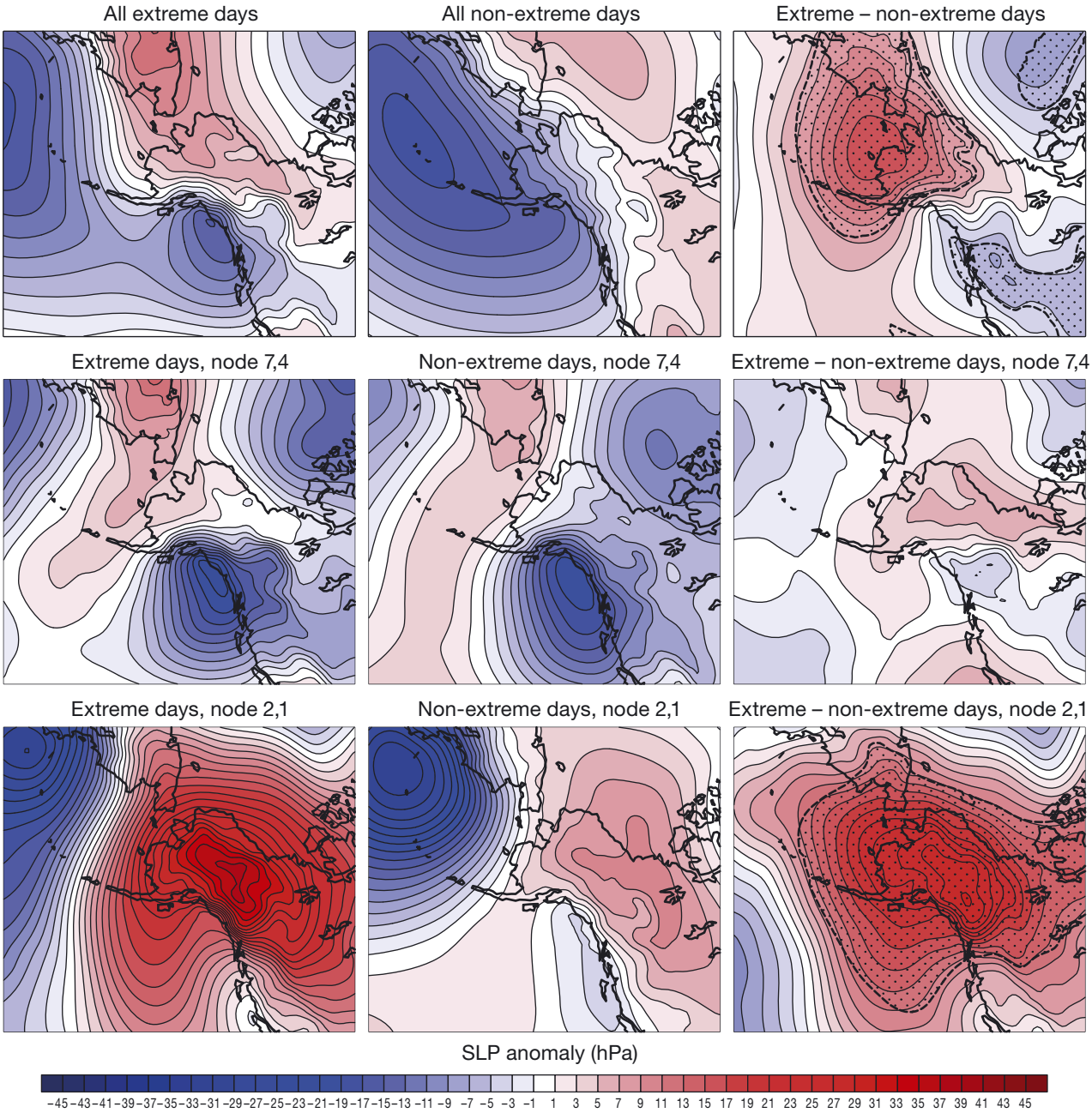

Fig. 8. Composites of sea-level pressure anomaly of cold extremes in the AS region for extreme days (left), non-extreme days (middle), and extreme minus non-extreme days (right). Top: all days; middle and bottom: days that map to Node 7,4 and Node 2,1 (see Fig. 6a), respectively. A Student's *t*-test was used to find statistically significant differences, and these are indicated by the stippling on the difference plots (right column)

sights into circulation patterns that lead to extremes that can be gained by using the SOM methodology. In the middle row, the average SLP anomaly patterns for days that map to node 7,4 (8 cold extreme days; located in the lower right corner of the SOM shown in Fig. 6a) are quite similar to the SLP anomaly composite of all cold extreme days (comparing the upper left panel [all extremes] to the middle left panel [extreme days that map to node 7,4]). For this node, the differences between extreme and non-extreme days are not significant, with higher pressure over Alaska and lower pressure to the south and west of the Aleutian Islands for extreme days (middle right panel of Fig. 8). However, the synoptic pattern for extremes that map to node 2,1 (4 cold extreme days; located in the upper left corner of the SOM shown

in Fig. 6a) is substantially different from the SLP anomaly patterns associated with cold extreme days for node 7,4 (comparing the bottom to the middle left panels of Fig. 8) and the composite of all cold extreme days (comparing the bottom to upper left panels of Fig. 8). The cold extremes for days that map to node 2,1 are associated with strong high pressure over all of Alaska, the details of which are lost in the composite of all cold extreme days. This shows that the SOM can discriminate between different types of synoptic patterns that can lead to extremes, and also retains information about the circulations associated with extremes that can get lost in a composite analysis that includes all extremes.

In the previous sections, one of the criteria by which a SOM was evaluated for the application in this study was the ability to discriminate between warm and cold extremes. An investigation into a particular node for the AS region was undertaken to demonstrate this occurrence and describe the situation in which both cold and warm extremes can occur for the same synoptic pattern. Node 5,4 in Fig. 6a has a Gulf of Alaska low-pressure system, with high pressure over eastern Siberia. This synoptic pattern had 4 warm and 2 cold extremes (Fig. 7; AS). An example of a warm extreme associated with this node was a 3 d event from 26 to 28 February 1992. The grid points that were extreme for this event were in the eastern portion of the domain in an area ahead of a low-pressure system in the Gulf of Alaska. As the event progressed over the 3 d, the low-pressure system moved northward into Alaska, likely bringing cloudy conditions which would inhibit radiative cooling. An example of a cold extreme that mapped to this node was one in which the extreme grid points were in the western part of the domain. The SOM node shows high pressure in the northwestern part of the domain, and the actual SLP pattern for this day had a strong high-pressure system over eastern Siberia, placing western Alaska in the northerly flow and associated CAA on the front side of the high-pressure system. Therefore, analysis regions that are large can have both warm and cold extremes for a single node if that node represents a strong pattern with varied flow across the region.

3.3.2. Canada west

There are 76 d with warm extremes and 84 d with cold extremes over the time period of interest for the CW region. On the 7 × 5 CW domain SOM, warm extremes occur on 25 (71%) of the nodes and cold extremes occur on 22 (63%) of the nodes. There is less isolation of the extremes across the SOM than what is seen with the AS region (Table 2). In addition, 13 (37%) of the 35 nodes have at least 1 occurrence of both a warm and a cold extreme. These results suggest that a SOM trained with SLP may not be appropriate for understanding the circulation that drives extremes in this region. One possible reason for this lack of discrimination of extremes is that many of the SLP anomaly values are removed (Fig. 6b) since they are at elevations above 500 m. This means that some of the more local circulation features of potential importance to extremes are not gleaned in the SLP anomaly SOM.

Based on a review of additional meteorological fields for individual extremes, the circulation at 850 hPa appears to have a stronger direct correlation to the presence of extremes in the CW region. A 7 × 5 SOM of 850 hPa geopotential heights (Fig. 9) is used to further analyze circulation features associated with extremes in this region. Instead of height anomalies, 850 hPa geopotential heights were selected for this analysis because these provide information on both circulation and temperature (e.g. lower heights indicate a colder column of air from the surface to 850 hPa).

The 850 hPa geopotential height SOM offers improvement in isolating the warm extremes across fewer nodes (57%) compared to the SLP anomaly SOM (71.4%; Table 2). The warm extremes occur predominantly along the farthest left column of the SOM, which are patterns with an elongated 850 hPa ridge across the far northwestern region of Canada and the Gulf of Alaska. The warm extremes also occur in nodes located on the right side of the SOM, which are patterns with low pressure to the west of the CW region bringing WAA across the CW region. These results are not surprising given that WAA or downslope warming drives warm extremes during the winter months.

Mapping cold extremes to the 850 hPa geopotential height SOM results in less discrimination of nodes compared to the SLP anomaly SOM. The 84 cold extremes occur in 26 (74%) of the 850 hPa SOM nodes compared to 22 (62.9%) of the SLP anomaly SOM nodes (Table 2). Analysis of meteorological fields prior to and during cold extremes reveals that events prior to the occurrence of the extreme are important. Prior to the extreme, very cold air is advected into the CW region. Next, a weak pressure gradient with light winds and enhanced radiative cooling occurs in the region, producing extremely cold conditions. This sequence of events leading to

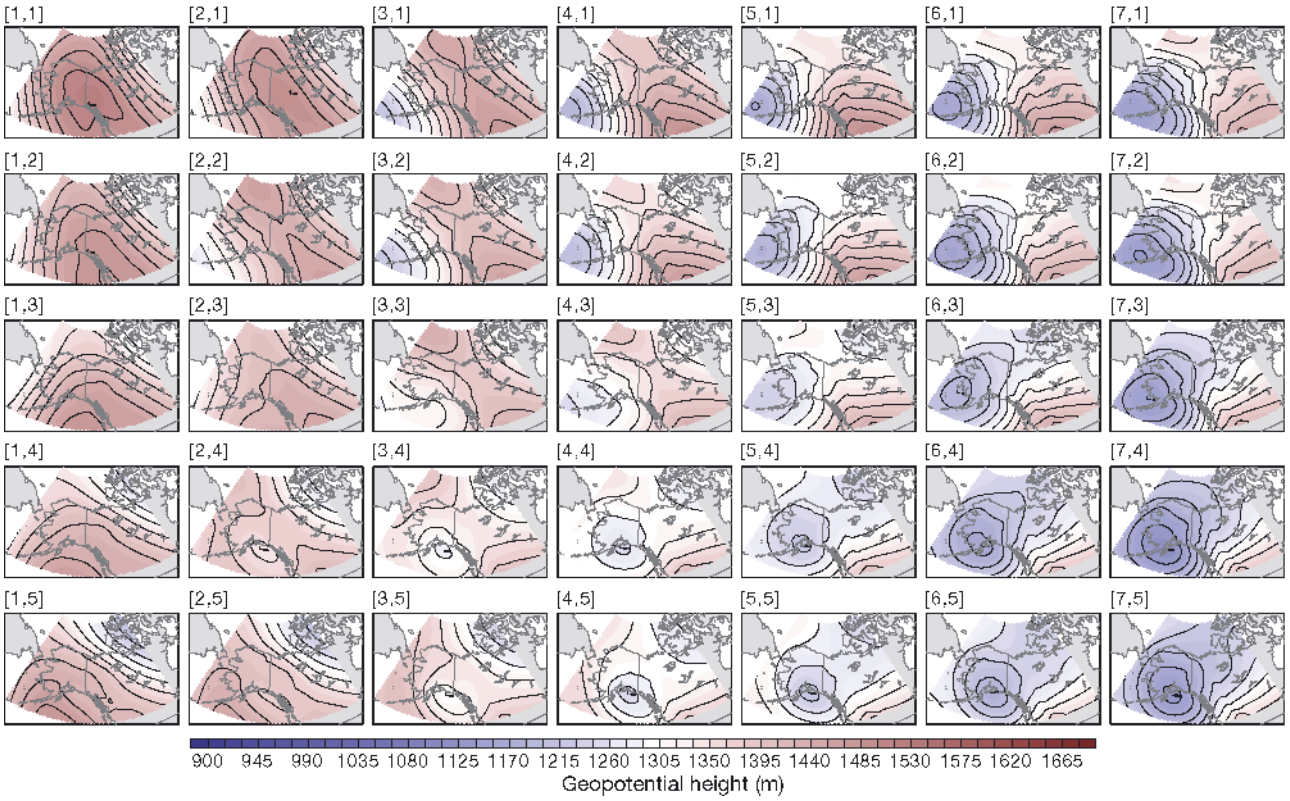

Fig. 9. The 7 × 5 SOM of 850 hPa geopotential heights representing the synoptic circulation patterns for the Canada west domain. Contents are color shaded every 15 m and black lines every 30 m

the cold extreme makes it less beneficial to evaluate cold extremes based solely on matching the circulation on the day of the extreme to the corresponding SOM node. Such results indicate that SOMs are an effective tool in analyzing extremes, but that an analysis technique that works in some cases may not work as well in other cases.

3.3.3. Canadian archipelago

The CA region had the most extreme days based on the definition used in this study: 116 (cold) and 100 (warm). The warm and cold extremes were spread over >85% of the nodes of the SLP anomaly 7 × 5 CA domain SOM, and 74% of the nodes were identified as having both warm and cold extremes (Table 2). These statistics indicate that this SOM is not adequate for identifying unique circulations associated with extremes in this region. This may be due to the fact that the CA region has nearly 50% more grid points than the AS and CW regions, and an even larger difference in spatial area given the percentage of non-land area that is a part of this region. To determine if the large size of the CA region was responsible for the lack of extreme event discrimination in this SOM, the CA region was divided into 3 subregions of roughly the same number of grid points: north (north of roughly 74°N), west (west of roughly 90°W and south of 74°N), and east (east of roughly 90°W and south of 74°N) (Fig. 1). The reasoning behind this direction of analysis was to determine if, for such a large region, the SOM was still able to naturally separate in SOM space the circulation that was associated with extremes in particular portions of the region.

A day was determined to be extreme in these subregions if 25 grid points or more fell in a particular sub-region and fewer than 5 grid points fell outside of that sub-region. For the warm extremes, there is a definite improvement on the separation in SOM space for extremes that fall in the different regions (Fig. 10). The extremes in the west are focused in the lower rows of the SOM, while in the east these are concentrated in the upper right corner of the SOM. The extremes in the northern part of the domain are more spread out in SOM space though somewhat focused in the middle of the SOM. However, for the cold extremes, there is less separation in SOM space between the different regions.

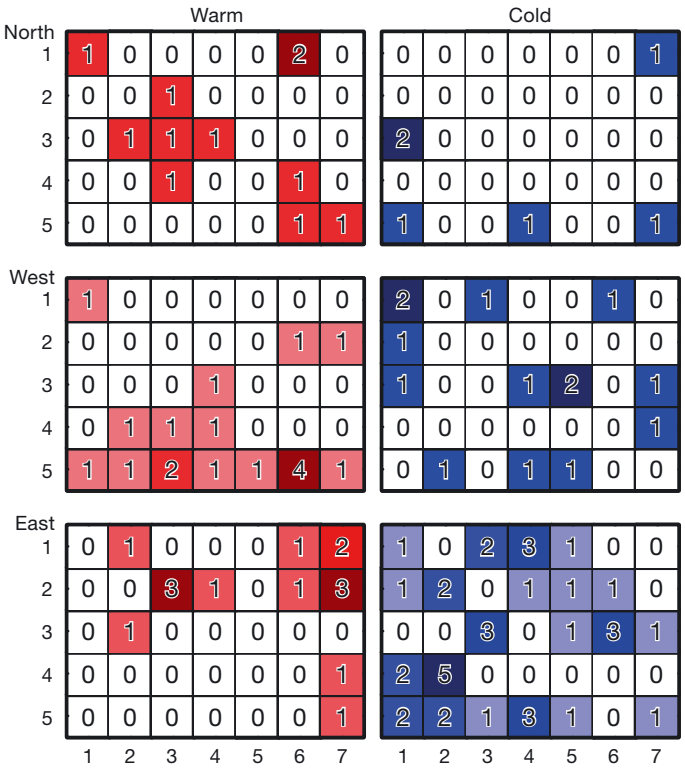

Fig. 10. Number of extreme days for each SOM pattern mapping to the 7 × 5 Canada archipelago (CA) domain SOM for the north (top), west (middle), and east (bottom) sub-regions in CA. The left (right) columns represent warm (cold) extremes. Darker (lighter) shading indicates more (fewer) extreme days on that node. The boxes in this figure correspond directly with the SOM nodes shown in Fig. 6c

The patterns in the bottom row of the SOM (Fig. 6c) represent an example of cold and warm extremes that fall on the same nodes (Fig. 10), where the extreme grid points are located in different parts of the region. Synoptic patterns with low pressure in the northeastern part of the domain (over western Greenland) and higher pressure over the southwestern portion of the domain are found on the left side of the bottom row of the SOM (Fig. 6c). Moving to the right in the bottom row of the SOM, the circulation patterns are characterized by low pressure over Alaska, with ridging over the eastern part of the domain. In the eastern sub-region of CA (Fig. 10), cold extremes tended to occur in the left portion of the bottom row, while, in the west, warm extremes occurred in every node in the bottom row and 4 extreme days occurred in Node 6, 5, which places the western portion of the CA region on the back side of the ridge in an area of southerly flow. Though these details were uncovered in this analysis, as was found for the CW region, the SLP-based SOM does not fully differentiate the physical processes responsible for the cold extremes in this region, and other considerations need to be taken into account to evaluate and understand extremes in this region.

4. DISCUSSION

Based on the results presented in this study the SLP-based SOM was better able to classify patterns responsible for temperature extremes in Alaskan regions than in Canadian regions. There are several likely reasons. The first is region size; the Alaskan regions are smaller, particularly AN, and there are simply fewer extreme days to evaluate. Also the areas that are extreme are smaller and more likely to be caused by a similar synoptic pattern. Another possible reason for this is that, in the winter, the weather in the Alaskan regions is quite strongly affected by the North Pacific storm track. These types of strong circulation patterns are not as prevalent in the Canadian regions, with the exception of the influence of the North Atlantic storm track in the southeastern portion of the CA region. In addition, changing sea ice coverage may have an impact on extremes, particularly in the CA region, which has a considerable marine influence. This region is largely ice covered in the winter. However, even a small amount of open water can have a large impact on air temperature due to the large temperature difference between open water and the atmosphere at this time of year. Therefore, the extremes in the Canadian regions may less likely be tied to specific synoptic patterns.

A goal of this work is to understand the underlying physical processes associated with extremes by using the SOM to differentiate the synoptic situations that lead to both warm and cold extremes. As the domain size decreased and the number of SOM nodes increased, there was a better separation in SOM space of warm versus cold extremes. However, for patterns with strongly varying flow across the region, e.g. highly meridional flow, warm and cold extremes can occur with the same synoptic patterns. Although there were no examples of this occurring for the regions studied or the definition of extremes used in this study, warm and cold extremes can occur on the same day in a particular region (e.g. a strong cyclone with WAA ahead of the cyclone and CAA behind it).

The use of an 850 hPa geopotential height SOM was shown to better identify circulation features associated with extremes in the CW region than the SLP-based SOM. Creation of node-averaged fields based on days associated with each SOM node allows analysis of multiple variables related to ex-

tremes as well as the events leading up to an extreme. These additional analyses, along with evaluation of the large-scale circulation information available with the SLP-based SOM, can give a more complete picture of the physical processes that lead to an extreme. Future work will expand on this application of SOMs for analyzing the forcing of extremes.

The method of matching extremes to specific nodes is effective in analyzing extremes, but it frequently comes with limitations, including some that have been discussed above. The true power and effectiveness in using SOMs comes with innovative analyses leveraging the advantages conferred by objectively determined pattern clustering provided by SOMs. There are many ways of combining a SOM with additional SOM-based analyses to provide answers to pertinent questions.

5. SUMMARY AND CONCLUSIONS

This study describes a method that uses the self-organizing map (SOM) algorithm to analyze the large-scale circulation associated with widespread temperature extremes. In addition, several items of importance were highlighted in creating and selecting a SOM classification and using it for this type of analysis. The SOM algorithm was used to both characterize large-scale circulation patterns and to relate these to widespread temperature extremes for 4 regions in the North American Arctic. An important part of the process is the identification of the climatology of synoptic patterns that represent the area of interest and how these patterns relate to the extremes. Many SOMs were created testing SOM array size, the domain representing the synoptic climatology, and the parameters used to create the SOM. The large-scale synoptic circulation for the extreme events was evaluated in the context of the SOM classification to determine the relationships between this circulation and widespread temperature extremes, and how these relationships differ from non-extreme days.

Some of the lessons learned from this analysis are given below.

Methods:

1. Tests of SOM creation show that using a larger neighborhood radius (approximately equal to the smaller dimension of the SOM array), a smaller learning rate, and a larger number of iterations result in the smallest final Q-errors and RMSDs.

2. For the domain and time period used for analysis in this study, SOMs with a node count of >100 did not add any additional diagnostic power, and this large number of SOM nodes was found to be too cumbersome to allow easy visualization of all of the synoptic patterns. The number of SOM nodes, or classes, to choose for the analysis needs to be large enough to clearly identify the features pertinent to what is being studied, but not too large if no additional benefit is gained and the analysis and presentation of results becomes cumbersome.

3. For the application in this paper, a high level of detail was needed to identify patterns associated with extremes. Going from the SOM with the fewest number of nodes covering a larger domain to SOMs with a larger array of nodes covering a smaller, more focused domain resulted in better separation in SOM space for the extremes, particularly for the Alaskan domains, and a better fit of the input data to the SOM nodes to which they mapped.

4. A final SOM size was chosen that was a balance between a tractable number of nodes and enough nodes to obtain the necessary circulation details. The 7 × 5 SOM focused on the individual domains was able to provide useful discrimination of circulation features responsible for temperature extremes. A SOM of 20 nodes (5 × 4) covering the full North American Arctic domain was found to miss some of the fine-scale features that are significant in identifying extremes based on a review using subjective and objective measures.

5. For some regions, sea level pressure is not the optimal variable for defining the synoptic circulation patterns and assessing the physical processes that lead to extreme temperature events. The targeted field should be one that has the strongest influence on the extreme being studied. In the Canada west (CW) region, a SOM created with 850 hPa geopotential heights was better able to identify important circulation features, particularly for warm extremes.

Extremes:

1. For the Alaska regions, warm extremes for both regions (north and south) were associated with an approaching low-pressure system and warm air advection (WAA) ahead of this system. The circulation patterns identified by the SOM indicated a more northerly location of low pressure for warm extremes in Alaska north (AN) versus Alaska south (AS). Cold extremes were associated with either cold air advection (CAA) or a broad band of high pressure likely associated with strong radiative cooling.

2. For CW, warm extremes were either due to WAA events or downslope warming. For cold extremes, the days leading up to the extreme were important in setting the cold air in place.

3. For Canada archipelago (CA), dividing the region into smaller sub-regions led to better separation in SOM space for the extremes, particularly for the warm extremes. However, the results were less clear for the region overall, and much more study is required to fully understand the extremes in this region.

The methodology presented in this study will be used in future work to understand in greater detail the physical processes that lead to widespread temperature extremes and the relationships between circulation and extremes for each of the 4 regions. This analysis will also be extended to summer (June, July, and August) temperature extremes, as well as to the evaluation of precipitation extremes. This will also be applied to future projections of climate change to determine if and how these extremes may change in the future.

Acknowledgements. This work was supported by NSF grant ARC-1023369 and US Department of Energy grants DESC0006643 and DEFG0207ER64463. Matt Higgins performed some of the original data analysis, and Cody Phillips contributed to analysis discussions. The authors thank 3 anonymous reviewers for their helpful comments and suggestions, which improved the manuscript.

LITERATURE CITED

Alexander LV, Uotila P, Nicholls N, Lynch A (2010) A new daily pressure dataset for Australia and its application to the assessment of changes in synoptic patterns during the last century. J Clim 23:1111–1126

Andrade C, Leite SM, Santos JA (2012) Temperature extremes in Europe: overview of their driving atmospheric patterns. Nat Hazards Earth Syst Sci 12:1671–1691

Athar H, Lupo AR (2010) Scale analysis of blocking events from 2002 to 2004: a case study of an unusually persistent blocking event leading to a heat wave in the Gulf of Alaska during August 2004. Adv Meterol 610263, doi: 10.1155/2010/610263

Cassano EN, Lynch AH, Cassano JJ, Koslow MR (2006a) Classification of synoptic patterns in the western Arctic associated with extreme events at Barrow, Alaska, USA. Clim Res 30:83–97

Cassano JJ, Uotila P, Lynch A (2006b) Changes in synoptic weather patterns in the polar regions in the twentieth and twenty-first centuries. 1. Arctic. Int J Climatol 26: 1027–1049

Cassano JJ, Uotila P, Lynch AH, Cassano EN (2007) Predicted changes in synoptic forcing of net precipitation in large Arctic River basins during the 21st century. J Geophys Res 112:G04S49, doi:10.1029/2006JG000332

Cassano EN, Cassano JJ, Nolan M (2011) Synoptic weather pattern controls on temperature in Alaska. J Geophys Res 116:D11108, doi:10.1029/2010JD015341

Cavazos T (1999) Large-scale circulation anomalies conducive to extreme precipitation events and derivation of daily rainfall in northeastern Mexico and southeastern Texas. J Clim 12:1506–1523

Cavazos T (2000) Using self-organizing maps to investigate extreme climate events: an application to wintertime precipitation in the Balkans. J Clim 13:1718–1732

Dee DP, Uppala SM, Simmons AJ, Berrisford P and others (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. QJR Meteorol Soc 137:553–597

Flato G, Marotzke J, Abiodun B, Braconnot P and others (2013) Evaluation of climate models. In: Stocker TF, Qin D, Plattner GK, Tignor M and others (eds) Climate change 2013: the physical science basis. Contribution of Working Group I to the 5th assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge

Francis JA, Vavrus SJ (2012) Evidence linking Arctic amplification to extreme weather in mid-latitudes. Geophys Res Lett L06801, doi:10.1029/2012GL051000

Glisan JM, Gutowski WJ Jr (2014) WRF summer extreme daily precipitation over the CORDEX Arctic. J Geophys Res 119:1720–1732

Gutowski WJ, Otieno F, Arritt RW, Takle ES, Pan Z (2004) Diagnosis and attribution of a seasonal precipitation deficit in a US regional climate simulation. J Hydrometeorol 5:230–242

Gutowski WJ, Willis SS, Patton JC, Schwedler BRJ, Arritt RW, Takle ES (2008) Changes in extreme, cold-season synoptic precipitation events under global warming. Geophys Res Lett 35:L20710; doi:10.1029/2008GL035516

Heikkilä U, Sorteberg A (2012) Characteristics of autumn–winter extreme precipitation on the Norwegian west coast identified by cluster analysis. Clim Dyn 39:929–939

Hewitson BC, Crane RG (2002) Self-organizing maps: applications to synoptic climatology. Clim Res 22:13–26

Huth R, Beck C, Philipp A, Demuzere M and others (2008) Classifications of atmospheric circulation patterns. Ann NY Acad Sci 1146:105–152

Inman HF, Bradley EL (1989) The overlap coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. Comm Stat Theory Methods 18:3851–3874

IPCC (Intergovernmental Panel on Climate Change) (2007) Climate change 2007: the physical science basis. Contribution of Working Group I to the 4th assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, New York, NY

Kawazoe S, Gutowski WJ (2013a) Regional, extreme daily precipitation in NARCCAP simulations. J Hydrometerol 14:1212–1227

Kawazoe S, Gutowski WJ (2013b) Regional, extreme daily precipitation in CMIP5 simulations. J Hydrometerol 14: 1228–1242

Kohonen T (2001) Self-organizing maps. Springer, New York, NY

Kyselý J (2008) Influence of the persistence of circulation patterns on warm and cold temperature anomalies in Europe: analysis over the 20th century. Global Planet Change 62:147–163

Lynch A, Uotila P, Cassano JJ (2006) Changes in synoptic weather patterns in the polar regions in the twentieth and twenty-first centuries. 2. Antarctic. Int J Climatol 26:1181–1199

Maheras P, Kutiel H (1999) Spatial and temporal variations in the temperature regime in the Mediterranean and their relationship with circulation during the last century. Int J Climatol 19:745–764

McBean G, Alekseev GV, Chen D, Førland E and others (2005) Arctic climate: past and present. In: Arctic Climate Impact Assessment scientific report. Cambridge University Press, Cambridge, p 22–60

Metropolis N, Ulam S (1949) The Monte Carlo method. J Am Stat Assoc 44:335–341

Mohr M (2004) Problems with the mean sea level pressure field over the western United States. Mon Weather Rev 132:1952–1965

Overland JE (2009) Meteorology of the Beaufort Sea. J Geophys Res 114:C00A07, doi:10.1029/2008JC004861

Randall DA, Wood RA, Bony S, Colman R and others (2007) Climate models and their evaluation. In: Solomon S, Qin D, Manning M, Chen Z and others (eds) Climate change 2007: the physical science basis. Contribution of Working Group I to the 4th Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge

Renom M, Rusticucci M, Barreiro M (2011) Multidecadal changes in the relationship between extreme temperature events in Uruguay and the general atmospheric circulation. Clim Dyn 37:2471–2480

Reusch DB, Alley RB, Hewitson BC (2005a) Relative performance of self-organizing maps and principal component analysis in pattern extraction from synthetic climatological data. Polar Geogr 29:188–212

Reusch DB, Hewitson BC, Alley RB (2005b) Towards ice core-based synoptic reconstructions of West Antarctic climate with artificial neural networks. Int J Climatol 25:581–610

Rodionov SN, Overland JE, Bond NA (2005) The Aleutian Low and winter climatic conditions in the Bering Sea. 1. Classification. J Clim 18:160–177

Rodríguez-Puebla C, Encinas AH, García-Casado LA, Nieto S (2010) Trends in warm days and cold nights over the Iberian Peninsula: relationships to large-scale variables. Clim Change 100:667–684

Sammon J (1969) A nonlinear mapping for data structure analysis. IEEE Trans Comput C-18:401–409

Schuenemann KC, Cassano JJ, Finnis J (2009) Synoptic forcing of precipitation over Greenland: climatology for 1961-1999. J Hydrometerol 10:60–78

Serreze MC, Carse F, Barry RG, Rogers JC (1997) Icelandic low cyclone activity: climatological features, linkages with the NAO, and relationships with recent changes in the Northern Hemisphere circulation. J Clim 10:453–464

Shabbar A, Bonsal B (2004) Associations between low frequency variability modes and winter temperature extremes in Canada. Atmos Ocean 42:127–140

Sheridan SC, Lee CC (2011) The self-organizing map in synoptic climatological research. Prog Phys Geogr 35: 109–119

Skamarock WC, Klemp JB, Dudhia J, Gill DO and others (2008) A description of the advanced research WRF Version 3. NCAR/TN 475+STR, NCAR, Boulder, CO

Skific N, Francis JA, Cassano JJ (2009) Attribution of projected changes in atmospheric moisture transport in the Arctic: a self-organizing map perspective. J Clim 22: 4135–4153

Stefanon M, D'Andrea F, Drobinski P (2012) Heatwave classification over Europe and the Mediterranean region. Environ Res Lett 7:014023, doi:10.1088/1748-9326/7/1/014023

Tebaldi C, Hayhoe K, Arblaster JM, Meehl GA (2006) Going to the extremes: an intercomparison of model-simulated historical and future changes in extreme events. Clim Change 79:185–211

Trenberth KE, Paolino DA Jr (1981) Characteristic patterns of variability of sea level pressure in the Northern Hemisphere. Mon Weather Rev 109:1169–1189

Wallace JM, Hobbs PV (1977) Atmospheric science: an introductory survey. Academic Press, San Diego, CA

Appendix. Twistedness index calculation and 7 × 5 SLP anomaly full domain SOM representation (Fig. A1)

Here, the equations used to calculate the 'twistedness index' (TI) for a Sammon map of N × M dimensions shown in Fig. 2 are presented. The calculation's only input data are the node indices in the SOM array and their Sammon map node coordinates in the Sammon map's 2-dimensional space. The latter are outputted from the SOM training in an ASCII file.

The TI finds each node's nearest neighbor in Sammon map space and computes their separation in terms of their node indices in the SOM array. The TI is the average of all such separations computed for each node of the SOM array. As a reference, a perfectly flat Sammon map will yield a TI value of 1; each node's nearest neighbor is just 1 index away in the array of SOM nodes. Values >1 will be expected as a function of the Sammon map's degree of twistedness. Calculation steps are as follows.

1. Select a SOM node as an anchor point. Find the distance between the anchor point and its nearest neighboring node in Sammon map space. Then compute their separation in the SOM array's index space. Repeat this step for each point.

$$\text{Sammon}_{Distance} = \left((x_i - x_j)^2 + (y_i - y_j)^2\right)^{1/2} \quad \text{(A1)}$$

$$\text{Index}_{Distance} = \left((X_1 - X_2)^2 + (Y_1 - Y_2)^2\right)^{1/2} \quad \text{(A2)}$$

where:

$x_1 = i$ MOD Dim_x

$y_1 = i$ / Dim_x

x_2 = min_index MOD Dim_x

y_2 = minindex / Dim_x

2. Compute the twistedness index.

$$Twistedness\ Index = \frac{\sum \text{Index}_{Distance}}{Dim\_Sammon} \quad \text{(A3)}$$

Table A1. Summary of the variables used in calculation of the twistedness index

Variable name	Variable description
Dim_x	x dimension range of the Sammon Map node coordinates
Dim_y	y dimension range of the Sammon Map node coordinates
Dim_Sammon	Total dimension size of Sammon Map (e.g. Dim_x × Dim_y)
Dim_array	Dim_Sammon − 1; allows for automated computation

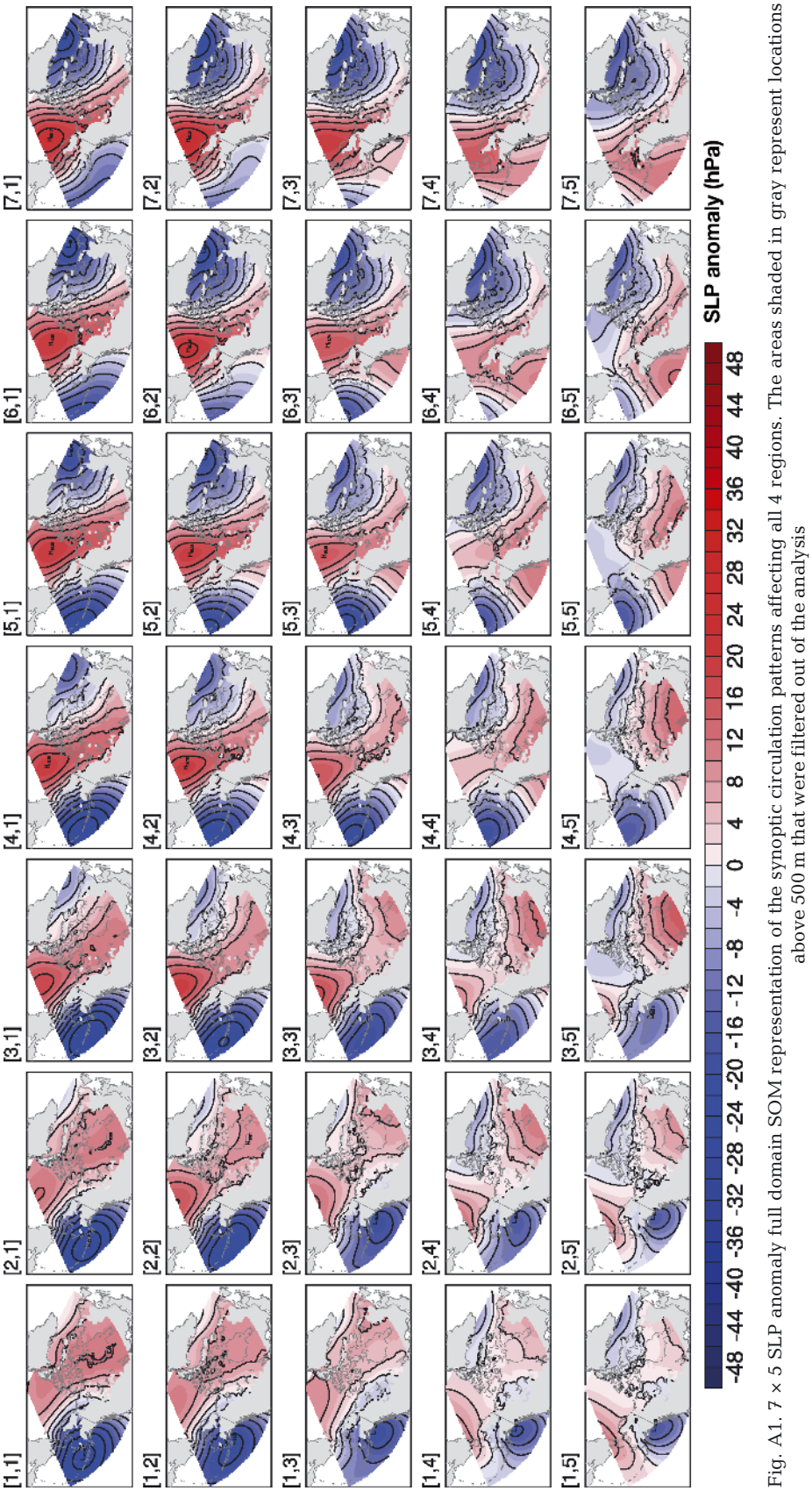

Fig. A1. 7 × 5 SLP anomaly full domain SOM representation of the synoptic circulation patterns affecting all 4 regions. The areas shaded in gray represent locations above 500 m that were filtered out of the analysis

Editorial responsibility: Oliver Frauenfeld, College Station, Texas, USA

Submitted: March 27, 2014; Accepted: November 7, 2014
Proofs received from author(s): January 31, 2015