North American Extreme Temperature Events and Related Large Scale Meteorological Patterns: Statistical Methods, Dynamics, Modeling, and Trends

Richard Grotjahn¹, Robert Black², Ruby Leung³, Michael F. Wehner⁴, Mathew Barlow⁵, Mike Bosilovich⁶, Alexander Gershunov⁷, William J. Gutowski Jr.⁸, John R. Gyakum⁹, Richard W. Katz¹⁰, Yun-Young Lee¹, Young-Kwon Lim¹¹, Prabhat⁴

Submitted to Climate Dynamics: 7 August 2014

¹Atmospheric Science Program, One Shields Ave., Dept. of L.A.W.R. University of California Davis, Davis CA USA 95616
²School of Earth and Atmospheric Sciences, 311 Ferst Drive, Georgia Institute of Technology, Atlanta, GA USA 30332-0340
³Pacific Northwest National Laboratory, Richland, WA USA 99352
⁴Lawrence Berkeley National Laboratory, Berkeley, CA USA 94720
⁵University of Massachusetts Lowell, Lowell, MA USA 01854
⁶NASA GSFC Global Modeling and Assimilation Office, Greenbelt MD 20771
⁷Climate, Atmospheric Science and Physical Oceanography (CASPO) Division, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA USA 92093
⁸Dept. of Geological and Atmospheric Sciences, Iowa State University, Ames, IA USA 50011
⁹Dept. of Atmospheric and Oceanic Sciences, McGill University, Montreal, Quebec, Canada
¹⁰Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research, Boulder CO USA 80307
¹¹NASA Goddard Space Flight Center, Global Modeling and Assimilation Office, and Goddard Earth Sciences Technology and Research / I.M. Systems Group, Greenbelt, MD USA 20771
Abstract

This paper reviews research approaches and open questions regarding data, statistical analyses, dynamics, modeling efforts, and trends in relation to temperature extremes. Our specific focus is upon extreme events of short duration (roughly less than 5 days) that affect parts of North America. These events are associated with large scale meteorological patterns (LSMPs).

Methods used to define extreme events statistics and to identify and connect LSMPs to extreme temperatures are presented. Recent advances in statistical techniques can connect LSMPs to extreme temperatures through appropriately defined covariates that supplement more straightforward analyses.

A wide array of LSMPs, ranging from synoptic to planetary scale phenomena, has been implicated as contributors to extreme temperature events. Current knowledge about the physical nature of these contributions and the dynamical mechanisms leading to the implicated LSMPs is incomplete. There is a pressing need for (a) systematic study of the physics of LSMPs life cycles and (b) comprehensive model assessment of LSMP-extreme temperature event linkages and LSMP behavior.

Generally, climate models capture the observed properties of heat waves and cold air outbreaks with some fidelity. However they overestimate warm wave frequency and underestimate cold air outbreak frequency, and underestimate the collective influence of low-frequency modes on temperature extremes. Climate models have been used to investigate past changes and project future trends in extreme temperatures. Overall, modeling studies have identified important mechanisms such as the effects of large-scale circulation anomalies and land-atmosphere interactions on changes in extreme temperatures. However, few studies have
examined changes in LSMPs more specifically to understand the role of LSMPs on past and future extreme temperature changes. Even though LSMPs are resolvable by global and regional climate models, they are not necessarily well simulated. Thus, additional research is needed to understand the limitations of climate models and improve model skill in simulating extreme temperatures and their associated LSMPs.

The paper concludes with unresolved issues and research questions.

**Keywords** large scale meteorological patterns for temperature extremes; heat waves; hot spells; cold air outbreaks; cold spells; statistics of temperature extremes; dynamics of heat waves; dynamics of cold air outbreaks; dynamical modeling of temperature extremes; statistical modeling of extremes
Temperature extremes have large societal and economic consequences. Heat waves in the U.S. cause a larger annual number of weather related deaths (170) than hurricanes (117) or flooding (74) in a 10-year average (1997-2006; Source: NOAA; http://www.nws.noaa.gov/om/hazstats.shtml). While many heat waves are short-lived (e.g. Chicago, 12-15 July 1995, more than 500 fatalities) longer events can have a large economic cost (e.g. $55B heat wave in the central part of the US during much of the summer of 1980).

Cold air outbreaks (CAOs) tend to be short-lived but carry large economic losses: $5.5B (20-30 December 1990) and $1.4B (13-17 January 2007) in California; while in Florida: $4.2B in 1983 and $2.3B in 1995 (Source: NOAA; http://www.ncdc.noaa.gov/billions/). Timing of the CAOs can be more important than the minimum temperatures of the freeze; during 4-10 April 2007 low temperatures across the South caused $2B in agricultural losses since many crops were in bloom or had frost sensitive buds or nascent fruit (Gu et al. 2008). The event also exemplifies how monthly means can be misleading: April 2007 average temperatures were near normal. In Florida during December 1989, 2 days of subfreezing weather wiped out half the citrus trees, though the monthly mean temperature was above normal. In short, both hot spells (HSs) and CAOs have great societal importance and they are short term events that do not necessarily appear in monthly mean data.

This report focuses on short-term (five-day or less) extreme temperature events in North America, though the extreme events discussed in this report generally occur over a portion of the continent. Multiple definitions of what constitutes a temperature extreme exist and may be based upon: frequency of occurrence, minimum length of event, whether subsequent temperatures exceed a threshold too. Temperature extremes considered in this paper include
both short-term hottest days (warm season) and CAOs (winter and spring) as these have the highest impacts. Such events, both observed and simulated, have received considerable attention (including research papers, e.g. Meehl and Tebaldi 2004; and active websites: http://www.esrl.noaa.gov/psd/ipcc/extremes/, http://www.ncdc.noaa.gov/extremes/cei/, and http://gmao.gsfc.nasa.gov/research/subseasonal/atlas/Extremes.html). However, this report focuses on the less well understood context for the extreme events. Our primary context is the large-scale meteorological patterns (LSMPs) that accompany these extreme events. From the LSMP context, this report surveys: 1) relevant extreme value analysis statistical tools, 2) synoptic and dynamical interactions between LSMPs and other scales from local to global, 3) model simulation issues, 4) trends in these temperature extremes, and finally 5) various open questions.

Temperature extreme events are usually linked to large displacements of air masses, hence a large amplitude wave pattern is produced, which is here called an LSMP. LSMPs have a spatial scale bigger than mesoscale systems but smaller than the near-global scale of some modes of climate variability. Because the pattern is superficially similar to a blocking ridge, a common forecast tool (a blocking index) can be used (Sillmann, et al. 2011). Extreme events have also been linked to circulation indices like the North Atlantic Oscillation (NAO) (Downton and Miller 1993; Cellitti et al. 2006; Brown et al. 2008; Kenyon and Hegerl 2008; Guirguis et al. 2011) the Madden-Julian Oscillation (MJO) (Moon et al. 2011) and El Nino / Southern Oscillation (ENSO) (Downton and Miller 1993; Higgins et al. 2002; Carrera et al. 2004; Meehl et al. 2007; Goto-Maeda et al. 2008; Kenyon and Hegerl 2008; Alexander et al. 2009; Lim and Schubert 2011). However, LSMPs are distinct from named climate modes such as the NAO, which are common modes of variability, whereas the LSMP is presumably as rare as the associated extreme event. Additionally, climate modes occur on a longer time scale than the LSMPs for the short term events focused upon in this article. Several studies identified
the LSMPs associated with specific extreme hot events (Grotjahn and Faure 2008; Loikith and Broccoli 2012) and CAO events (Konrad 1996; Carrera et al. 2004; Grotjahn and Faure 2008; Loikith and Broccoli 2012). Parts of these LSMPs tend to be uniquely associated with the corresponding extreme weather and those parts have some predictability (Grotjahn 2011).

Establishing a link between LSMPs and extreme temperature events that are rare, and occur on a smaller (regional) space provides a vehicle for interpreting the 'weather' created by a climate model. Being large in scale, the LSMPs are well resolved by current climate models. The smaller scale of the extreme may be more influenced by topography, surface conditions (e.g. vegetation type and landform) and sub-surface factors (including soil moisture for which there is hysteresis). However, the LSMPs for extreme events are not fully understood for different parts of North America. Also, interactions with local processes: topography, soil moisture, etc. play key roles but there is a knowledge gap in how well climate models simulate the LSMPs as well as these local processes and how the local and global modes interact with LSMPs. A critical pair of questions is: how well do climate models simulate the LSMPs, and is a given simulation of the distribution of extreme temperature 'good' for the right or wrong reasons? These questions are critical to address in order to understand the uncertainty of future projections and to drive model improvements.

Hence, now is an opportune time to examine the critical issues discussed above and identify key knowledge gaps in the understanding of climate extremes and their variability and trend because: 1) we often do not know if the current climate models used for future projections are producing extremes with the correct dynamical mechanisms, which reflects directly on the appropriate confidence in the projections, and 2) knowledge of the LSMPs may be used to improve downscaling (statistical or dynamical) by focusing attention on the large scale pattern that is fundamental to the occurrence of the extreme event. Global models which may not be able to reproduce the magnitude or duration of extreme temperature events accurately may still
be able to capture the correct LSMPs. Finally, there is now sufficient preliminary work to make a brief summary reasonable.

2 Extreme statistics and associated large scale meteorological patterns (LSMPs)

2.1 Definitions of extreme events

A series of indices developed by the Expert Team on Climate Change Detection and Indices (ETCCDI) under the auspices of the World Meteorological Organization’s CliVar program provide a useful but somewhat incomplete starting point to explore the relationship between extreme temperatures and LSMPs. Table 1 shows some of the ETCCDI temperature indices. The full set of ETCCDI indices from observations and climate models as well as software to calculate them are available at \url{http://etccdi.pacificclimate.org} and are described in Alexander et al. (2006). The ETCCDI indices summarized in Table 1 were designed for detecting and attributing the human effects on extreme weather and do not necessarily represent particularly rare events. However, application of extreme value statistical methodologies can be used to quantify the behavior of the tails of the distribution of certain ETCCDI indices and reveal insight into truly rare events (Brown et al. 2008; Peterson et al. 2013). Although a changing climate makes the usual interpretation of a return value (or level) as a measure of risk over a return period problematic, it still can be interpreted as a quantile that varies over time (e.g., the twenty-year return value may be interpreted as that value that has a 5% chance of occurrence in a particular year). Increasing trends in these measures of extreme temperature are detectible at the global scale (Brown et al. 2008) and have been attributed to human emissions of
greenhouse gases (Christidis et al. 2005). At local scales, patterns of increases and decreases are observed reflecting the significant amount of natural variability in extreme temperatures.

Figure 1 shows the observed trends over North America from 1950-2007 in the twenty-year return values of the hottest/coldest days and hottest/coldest nights. The temperature of very cold nights (figure 1d) exhibits pronounced warming over the entire continent as does the temperature of very cold days (figure 1b). The pattern of changes in the temperature of very hot days (figure 1a) and very hot nights (figure 1c) follows that of average temperature change with strong cooling in the southeastern United States. This “warming hole” has been connected to sea surface temperature patterns in the equatorial Pacific (Meehl et al. 2007) and to changes in anthropogenic aerosols in the eastern U.S. (Leibensperger et al. 2012). Changes are more pronounced at the higher latitudes, except in Quebec and Newfoundland. In this analysis, extended from Peterson et al. (2013), annual anomalies are used to define the extreme indices. Although illustrative of the statistical techniques and that certain trends are significant, analyses of annualized measures of extreme temperature can complicate the identification of associated LSMPs due to the strong seasonal dependence of the atmospheric circulation. Furthermore, events of high and low impact are better separated in seasonal analyses.

2.2 Application of extreme value statistical techniques

The observed changes in figure 1 are calculated using a time-dependent point process approach to fit “peaks over threshold” statistical models (Coles 2001). In this case, the extension of stationary extreme value methods to a time-dependent formalism used time as a “covariate” quantity to the ETCCDI indices (Kharin et al. 2013). One of the principal advantages of a fully time dependent formalism over invoking a quasi-stationary approximation (Wehner
is that the amount of data that can be used to calculate extreme value parameters is
substantially increased resulting in higher quality fitted distributions and hence more accurate
estimates of long period return values. For calculations involving climate model output,
additional statistical accuracy can be obtained by using multiple realizations from ensembles of
simulations, provided that they are independent and identically distributed.

The “block maxima” and “peaks over threshold” methods to fit the tails of the
distributions of random variables are asymptotic formalisms (Coles 2001). In this terminology,
‘block maxima’ refers to use of only the maximum value during each ‘block’ of time, usually a
single season or year. The resulting Generalized Extreme Value (GEV) or Poisson and
Generalized Pareto Distributions (GPD) are both three parameter functions and can be
transformed between each other. Hence provided that the data used to fit a distribution is in the
“asymptotic regime”, i.e. far out in the tail of the distribution, the two methods are equivalent.
However, uncertainty in the estimate of long period temperature return values resulting from
limited sample size can be appreciable and may be as large as that from unforced internal
variability (Wehner 2010). However, variations in these estimates from multi-model datasets
such as CMIP3/5 are generally significantly larger.

Confidence in the estimates of the statistical properties of the tail of the parent
distribution of a random variable can be ascertained by exploring the sensitivity to the sample
size used to fit the extreme value distribution. For block maxima methods, the length of the
block is a season (or effectively so for temperature if the block length is a year). Lengthening the
block makes the sample size smaller; shortening it makes the sample size larger since only one
extreme value is drawn from each block. Such sensitivities are somewhat more straightforward
to explore with peaks over threshold (POT) methods. Typical thresholds may be chosen
between 80% and 99% depending on the size of the parent distribution that the extreme values
are drawn from. However, as standard POT methods do not necessarily discriminate between
extreme values that may occur at successive instances, because the individual extremes values may not be truly independent. In these cases, declustering techniques (Coles 2001) should be applied to avoid biased (low) estimates of the uncertainty. The trends in extreme temperature shown in figure 1 were calculated using such a declustering technique and a POT formalism with time as a linear covariate. An alternative approach retains possibly dependent consecutive extremes, adjusting the estimates of uncertainty through either resampling or more advanced techniques for quantifying extremal dependence (e.g., Fawcett and Walshaw 2012).

Furthermore, LSMPs responsible for extreme events can be formed using only the dates of the onset of the event. (e.g. Grotjahn and Faure 2008; Bumbaco et al. 2013) reducing the risk of autocorrelated extremes. In the cited studies, a 5 day gap was typically required between events. Low frequency factors (or climate modes), such as ENSO, are best treated using the previously mentioned covariate techniques.

Some of the advantages and challenges of applying statistical methods based on extreme value theory to analyze non-stationary climate extremes have been pointed out previously (e.g., Katz 2010), but are still not necessarily well appreciated by the climate science community. Conventional approaches tend to be either: (i) less informative (e.g., analyzing only the frequency of exceeding a high threshold, not the excess over the threshold, a measure of the intensity of the event); or (ii) less realistic (e.g., based on assumption of distributions such as the normal that may fit the overall data well, but not necessarily the tails). When relating extremes to LSMPs, standard regression approaches would not quantify the uncertainty in the relationship as realistically as using extremal distributions with covariates. Challenges in extreme value methods include specifying the dependence on LSMPs of the parameters of the extremal distributions in a manner consistent with our dynamical understanding. Moreover, heat waves and CAOs are relatively complex forms of extreme events, some of whose characteristics can be challenging to incorporate into the framework of extreme value statistics
Finally, another advantage of the peaks over threshold approach over the block maxima approach is being able to incorporate daily indices of LSMPs as covariates (not just monthly or seasonally aggregated indices).

2.3 Identification of LSMPs related to extreme temperatures

There are several methods that have been used to identify the LSMPs that occur in association with temperature extreme events. These methods and their properties are summarized in Table 2 and the text which follows.

2.3.1 Composites

Composite methods define the LSMP using a target ensemble average. The values at a grid point for a field on specified 'target' dates are averaged together. These dates might be when a temperature event begins (onset dates) defined as when some parameter(s) first meet some threshold and duration criteria. Table 3 illustrates various threshold and duration criteria that have been used to identify short-term extremely hot events. The different definitions yield somewhat different dates and results. A definition using a physiological hazard (Robinson 2001) is not satisfied at night in the coastal and some inland regions of the West, even though the daytime temperature threshold is well exceeded. Similarly, extremely hot temperature events can occur for fewer than 3 days and thereby included in some definitions but not others. Lower thresholds (90th percentile) generate more events thus increasing the sample size which can improve the statistical fit, though the behavior of the highest 1% may not be well fit. Similarly, the number of stations (or the size of the area over which those stations occur) impacts which
dates are identified, even for regions that would seem to be meteorologically consistent; for
example, which stations and how many are included from the Central Valley of California
strongly impacts which dates exceed a threshold. Also, different definitions target different
purposes: Grotjahn and Faure were interested in the LSMPs at the onset of the hottest events
while Meehl and Tebaldi were interested in finding the longest duration events of some
importance. One can view Table 3 as illustrative of why no single criterion is used and to
suggest various approaches to identifying events. These dates might be the dates for the event,
though other dates may be used, such as a specified number of days prior to the onset of each
event. The number of dates defines the number of ensemble members.

Compositing has several advantages. Meteorologically relevant full fields (or anomalies)
are found and can be tracked forward and back in time to gain synoptic and dynamical
understanding of what preceded the event and how it evolves over time. The method is non-
parametric in that it does not make any assumption about the pattern or the event statistics.
Unlike some other methods, criteria can be applied (typically a waiting period between events)
to ensure independence between the events. Significance can be assessed using a bootstrap
re-sampling procedure where the target ensemble value at a grid point is compared to the
distribution of values at that grid point found from a large number of ‘random ensembles’ (each
of which uses the same number but randomly-chosen dates). Values above (or below) a
threshold of the random ensemble distribution imply significantly high (low) values at that grid
point. For example, a target ensemble value as high or higher than the top 10 of 1000 random
ensemble values at that point is significant at the 99% level. Figure 2 shows the LSMPs for
California Central Valley cold air outbreaks and for heat waves obtained by this method,
including application of the bootstrap re-sampling significance testing.

Composting has some disadvantages compared to some other methods. The
identification of the extreme event dates must be done separately and before the compositing.
The composite produces one target ensemble average (for a specific field and level) for each set of target dates. If more than one LSMP can produce the extreme event, then that must be identified either with one or more additional criteria when choosing dates or identified by examining (perhaps qualitatively) the maps for each individual member of the target ensemble. A procedure like adding the number of positive and subtracting the number of negative anomaly values at a grid point in the target ensemble members (called the ‘sign count’ in Grotjahn 2011) can assist with identifying multiple patterns. (If the sign count equals the number of ensemble members, then all members have the same sign of the anomaly field at that location.)

2.3.2 Regression

Regression relates a known quantity (the predictor) to an unknown quantity (the predictand) using a specified function. The method is parametric in assuming a specific function applies to the relationship between the predictor and predictand. An example predictor might be daily minimum surface temperature and the predictand might be 700 hPa level meridional wind. At each grid point the value of the predictand can be estimated using a polynomial function of the predictor, where the coefficients of that polynomial are calculated to minimize a squared difference between actual predictand values and polynomial values at that grid point. In general, the coefficients differ from grid point to grid point. To find the LSMP in this example of extreme cold events, the polynomial can be used to construct the predictand at each grid point using a predictor value such as 2 standard deviations below normal.

The pattern obtained by regression can provide physical insights by directly linking the patterns in the predictand to extremes in the predictor. For example, lower tropospheric (700 hPa) winds prior to a California CAO flow from northern Alaska and northern western Canada to
reach California without crossing over the Pacific Ocean. Regression can be used to examine patterns leading up to (or after) event onset dates by offsetting in time the predictor values from the predictand values when calculating the regression line coefficients. The LSMP is again the resultant predictand estimates when the predictor is at some specified value (e.g. predictor equals 2 standard deviations below normal). Significance can be estimated by rejecting a null hypothesis (e.g. that the regression coefficient is zero at the 1% level using a student’s t test).

One disadvantage of regression is that assumption of a specific polynomial to represent how the predictand varies relative to the predictor. The fit of the regression line (polynomial) to extreme values may be notably altered by the order of the polynomial assumed. Regression, like composites, only finds one pattern. Regression does not incorporate event duration criteria (such as: the event must last at least 3 days). Regression treats all dates as independent, which they are not likely to be; however this can be somewhat mitigated by sub-sampling the data (e.g. only use every fifth day). Subsampling might be combined with low pass filtering to aggregate the mixture of onset and during-event dates. Another disadvantage is that a portion of the pattern may be highly significant but only have small correlation to the predictor.

2.3.3. Empirical Orthogonal Functions

Empirical Orthogonal Functions (EOFs) or Principal Components (PCs) can be used to identify LSMPs for extreme events. EOFs are the eigenfunctions of a matrix formed from the covariance between grid points on maps. EOFs from all maps in a time record will be ordered based on the eigenfunctions responsible for the largest amount of variance between time samples. Such eigenfunctions are the most common modes of variability and so not likely to be LSMPs of extreme events that are rare. However, EOFs can be formed only from maps selected in
reference to an extreme event, such as maps only on those dates of the onset of an event. (EOFs of common low frequency modes may influence short-term extreme events focused on in this paper, as discussed in section 3.1.) Weighting can be used for a variable grid spacing (such as occurs when using equal intervals of longitude across a range of latitudes).

One advantage of EOFs/PCs is the method finds multiple patterns, each of which is orthogonal (not a subset) of another pattern. The method can calculate the fraction of variability that is due to each EOF/PC. This approach is most often used with filtered data to find low frequency structures. This method is suitable for finding patterns leading up to and after the event by shifting the dates chosen by the event criteria (and only using those dates). In a study of California Central Valley hot spells, Grotjahn (2011) found the leading EOF based on dates satisfying the criteria in Table 3 to be very similar to the corresponding ensemble mean composite.

One disadvantage is that the patterns found may depend on the domain chosen, though EOF ‘rotation’ may help. Each EOF might explain only a small fraction of the variance and so no single EOF might be an LSMP thereby limiting physical insight. Different leading EOFs/PCs might have structures influenced by the required orthogonality and possibly not a pattern that occurs during an event. While the amount of variance associated with a given EOF is used to indicate the importance of the EOF, there is no inherent statistical significance test. Hence it can be unclear what portions of the EOF are significantly associated with the extreme event and what parts are not (and happen to reflect limited variation in the finite sample).

2.3.4. Clustering Analysis
Clustering analysis is the comprehensive terminology indicating a widely used partitioning procedure that identifies separate groups of objects having common structural elements. Clustering analysis has been used to classify distinct sets of LSMPs associated with extreme events (Park et al. 2011; Stephanon et al. 2012). When we have 100 historical hot spell events over a given region, not all extreme events may have the same LSMP on their onset phase. Some events may have a wave-like height field, others may have a dipole pattern, and still others may have a third pattern. By applying the cluster analysis to group similar onset patterns, one can isolate the dynamical origins of different extreme temperature events. Although detailed grouping procedures vary for every clustering technique, the basic concept is to minimize the overall distance between patterns among events in resultant groups. For example, the k-means clustering technique applies an iterative algorithm in which events are moved from one group to another until there is no additional improvement in minimizing the squared Euclidean point-to-centroid distance in a group (Spath 1985; Seber 2008), where each centroid is the mean of the patterns in its cluster.

Output of clustering analysis is just the average field of events in each cluster, which might be similar to the output from composite analysis. Unlike the composite analysis in which membership of clusters should be pre-identified, the essential point of clustering analysis is that it objectively classifies the events based on the spatial pattern similarity. Another advantage is that resultant clusters are based on physical maps without assumptions of orthogonality and symmetry such as in the mode separation by EOF/REOF. The robustness of a hot spell classification can be tested by Monte Carlo test as follows. The first step is to calculate a stability score which is the ratio of the number of verification period heat waves that are correctly attributed over the total number heat waves in the cluster. The second step estimates the PDF distribution of the null hypothesis that cluster assignment is purely random. Significance of each
cluster can be estimated by rejecting a null hypothesis (e.g. stability score is located within
highest 99% of PDF).

One disadvantage of clustering analysis is that one pre-specifies the number of clusters
(e.g. k in k-means clustering). Determining the number k is a little subjective if one does not
have sufficient prior knowledge of related physical patterns. There are several statistics to
check the optimal number of k such as ‘distance of dissimilarity’ (Stephanon et al. 2012).
Another disadvantage is the ambiguity of cluster assignment for certain events. The patterns of
such events cannot be assigned clearly to one group over another group. Part of pattern may
resemble Cluster #1, while the other part may be more similar to Cluster #2. To avoid the risk of
assigning some marginal events to specific clusters, probabilistic clustering methods (eg. Smyth
1998) are developed, which suggests the possibility (e.g. percentage) that event could be
assigned to each cluster rather than assigning it only to one cluster. If one increases the cluster
number, one can expect the decrease of such ambiguity in classification. However, the point of
clustering analysis is to give a physical insight with minimal groups but not to interpret every
single episode.

2.3.5. Self-Organizing Maps

Self-Organizing Maps (SOMs; Kohonen 1995) are two-dimensional arrays of maps that display
characteristic behavior patterns of a field (e.g., Cavazos 2000; Hewitson and Crane 2006;
Gutowski et al. 2004; Cassano et al. 2007). The SOM array is a discretization of the continuous
pattern space occupied by the field examined. Thus, in contrast to clustering analysis, SOMs do
not assume a clumping together of patterns, though such behavior can emerge if present in the
input data. Figure 3 gives an example of a SOM array of synoptic weather patterns in sea level
pressure over a region centered on Alaska. Individual maps in the array represent nodes in a projection of this continuous space onto a two-dimensional surface, with the size of the array determined by the degree of spatial discretization of the SOM space one “feels” is needed for the analysis at hand. The two dimensions show the two primary pattern transitions for the field examined. Although one could, in principle, use more than two dimensions, typical practice in climatological work has used only two.

The input maps themselves determine the degree and types of pattern transitions, hence the “self-organizing” nature of the resulting array. The SOM node array is trained on a sequence of input maps through an artificial neural net technique. The SOM array does not necessarily favor the largest scales in the input data, but rather the scales most relevant to the field for the domain and resolution examined. Consequently, SOMs can extract nonlinear pattern changes in fields, such as shifts in strong gradients. In addition, the pattern at each node is essentially a composite of input maps with similar spatial distribution for the field examined, so that patterns in the SOM array not only show archetypal patterns of the field examined, but they directly lend themselves to physical interpretation of actual climatic behavior. Typically, the SOM array displays behavior that has the highest temporal variance in the input data. From this perspective, the SOM array is roughly akin to a transformation of a rotated EOF from spectral space back to physical space.

An advantage of SOMs is that one can identify the nodes where extreme events occur frequently and thus the physical behavior yielding extremes. For example, extreme events may tend to cluster in a small portion of SOM space, thereby allowing identification of LSMPs yielding extreme events. A further advantage is that if more than one cluster emerges in the SOM-space frequency distribution, then the clustering provides a SOM-determined segregation of different types of extreme events. One can then focus analysis and composites of additional fields (e.g., precipitation, winds, temperature) on only events of the same type. For example,
Cassano et al. (2006) used a SOM of sea-level pressure patterns over Alaska to determine which synoptic weather patterns were responsible for extreme wind and temperature events at Barrow, Alaska. They then found robust links between these large-scale synoptic weather patterns and local weather features (precipitation, winds, and temperature).

Another advantage of SOMs is that one can construct estimates of the significance of differences in frequency distributions in SOM space through bootstrapping procedures to estimate the likelihood that frequency distributions are not simply the result of random, finite sampling of the pattern space. Thus one can compare frequency distributions between a present and projected climate to assess potential climate changes in LSMPs, or between an observational and model climates to assess similarity of observed and simulated LSMPs yielding extremes.

A disadvantage of SOMs is that the array size is pre-determined by the user, and there is no clear, objective guideline for selecting array size. There are, however, some factors that can affect the array-size choice. The issue of significance limits the degree of discretization (number of nodes) one applies to the SOM space. Fine discretization will allow apparent detection of small differences in how different data sets occupy pattern space, but fine discretization will also render very noisy frequency distribution functions of the fields in SOM space, thus undermining detection of any significant differences. Coarse discretization limits the ability of the SOM procedure to resolve features producing the extreme events, so a further disadvantage is that an insufficient array size may obscure clustering that may be present in the data. The training method also requires specification of parameters that govern the training process. A well-trained SOM is insensitive to these choices, but care is needed to ensure such a result. In addition, like some of the other methods described here, the extreme events are defined separately from the SOM analysis.
2.3.6. Machine learning and other advanced techniques

Looking to the future, we note that substantial progress has been made in the field of machine learning for extracting patterns from Big Data. Commercial organizations such as Google and Facebook rely on sophisticated, scalable analytics techniques for mining web-scale datasets. We believe that both supervised and unsupervised machine learning tools could play an important role in extracting spatio-temporal patterns from climate datasets. One technique in particular: Deep Belief Networks (Salakhutdinov and Hinton 2012) has been applied with tremendous success to the tasks of classifying objects in digital images (Krizhevsky et al. 2012) and speech recognition (Hinton et al. 2012). These methods have substantially outperformed existing techniques in the field with the same underlying learning algorithm. While these techniques have not yet been adapted for a multi-variate spatio-temporal dataset (such as in climate), research efforts are currently underway to evaluate the performance of such methods in extracting patterns as well as anomalies from datasets. It is too early to discern pros and cons fully for such methods.

2.4, including large scale patterns in extreme statistics

The application of covariates in extreme value methods is relatively new to the climate science community although it has been available to the larger statistics community for some time (Coles 2001). In fact, the book by Coles includes an example in which annual maximum sea level is related to a climate mode, the Southern Oscillation. In their study of changes in extreme
daily temperatures, Brown et al. (2008) used the NAO as a covariate in addition to a trend component.

Such techniques have proven useful in connecting extreme temperatures to LSMPs. Sillmann and Croci-Maspoli (2009) and Sillman et al. (2011) used a blocking index as a covariate for extremely cold European winter temperatures and found that extreme value distributions (based on block minima, equivalent to block maxima) were better fit and long period return values were somewhat colder. Furthermore, they concluded that projected future extremely cold events in Europe were less influenced by atmospheric blocking because of projected shifts in North Atlantic blocking patterns. Similarly, Photiadou et al. (2014) used a similar technique (but based on peaks over threshold approach rather than block maxima) to connect blocking and other indices to European high temperature events finding that while El Nino / Southern Oscillation (ENSO) does not exert much influence on extremely high temperature magnitudes or duration, the North Atlantic Oscillation (NAO) and atmospheric blocking do. However to date, such covariate techniques relating atmospheric blocking to extreme temperatures have not been applied to North America. Many physically based covariate quantities potentially offer insight into the mechanisms behind extreme temperature events and the response to changes in the average climate. Good North American candidates for covariates include indices measuring modes of natural variability such as those describing the ENSO, the Pacific Decadal Oscillation (PDO), the NAO, the North Atlantic Subtropical High (NASH), and various blocking indices.

The ETCCDI indices were designed for climate change detection and attribution purposes rather than for exploring the mechanisms causing extreme events. While not ideal for connecting extreme temperature events to LSMPs and they do not describe particularly rare events, ETCCDI indices are designed to be robust over the observational record. They have been calculated and described for the CMIP5 models by Sillmann et al. (2013a) and are freely
available (see Table 1). Furthermore, they are intended to be applied globally and be meaningful in areas of sparse observations. The relatively dense network of North American observations since the beginning of the 20th century permits the construction of more specialized LSMP extreme indices linked to specific extreme events. Grotjahn (2011) defines an index which is simply an unnormalized projection of key parts of a target ensemble LSMP onto a daily map of the corresponding variable. (He combined such projections onto 850 hPa temperature and 700 hPa meridional wind to form his ‘circulation index’.) His target ensemble members are from dates satisfying criteria listed in Table 3. The key parts of the fields used are those where all the extreme events in the training period were consistent, at least in having the same anomaly sign. Grotjahn (2011) found that extreme values of such an index (based on upper air data) occurred on many of the same dates as extreme values of surface stations in the California Central Valley (CCV). Statistics of extremes techniques also indicate statistically significant relationships between the circulation index and both the rate of CCV daily maximum temperature exceeding a high threshold and the distribution of the excess over the threshold (Katz and Grotjahn 2014). Grotjahn (2013) used such an index to show that a particular climate model was notably under-predicting the occurrence frequency (by half) of CCV hot spells. Grotjahn (2014) used such an index to show how that same model compared with a 55 year historical record and what the model implied for CCV hot spells during the last half of the twenty-first century under two representative concentration pathways (RCPs) of greenhouse gases.

3. Large Scale Meteorological Patterns related to Extreme Temperature Events
Intraseasonal extreme temperature events (ETEs) are almost always associated with regional air mass excursions induced by circulation anomalies part of large-scale meteorological patterns (LSMPs). The nature of LSMPs responsible for ETEs ranges from mobile synoptic scale disturbances (e.g., midlatitude cyclones; Konrad 1996) up to quasi-stationary teleconnection patterns spanning the entire North American continent (e.g., the Pacific-North American teleconnection pattern; Cellitti et al. 2006). In some cases, it is the juxtaposition of distinct LSMPs that leads to ETE events (Lim and Kim 2013). Heuristically, the role of LSMPs in producing ETEs could be considered the result of either (1) a direct contribution to the large-scale circulation that facilitates the air mass excursion or (2) the indirect modulation of sub-scale variability, such as regional modulation of storm track behavior by blocking patterns. Besides such dynamically-driven impacts, there exist possible local impacts related to the interaction of the LSMP with local topography or coastline features, leading to possible local symmetries in the response pattern (e.g., Loikith and Broccoli 2012). Here we overview current knowledge of the remote forcing, dynamics and local forcing of LSMPs associated with ETEs.

3.1 Remote forcing of LSMPs and ETEs

3.1.1 Connection to low frequency modes of climate variability.

Numerous observational studies have ascertained that ETE behavior is modulated by recurring large scale teleconnection patterns, particularly during winter. On intraseasonal time scales there is a substantial modulation of North American ETEs during winter by the Pacific-North American (PNA) pattern, North Atlantic (or Arctic) Oscillation (NAO or AO) and blocking patterns (Walsh et al. 2001, Celitti et al. 2006, Guirguis et al. 2011). The modulation of winter ETEs by
the PNA and AO/NAO carries over to interannual time scales (Lim and Schubert 2011, Westby et al. 2013). On interannual and longer time scales additional climate modes such as El Nino-Southern Oscillation (ENSO) and the Pacific Decadal Oscillation (PDO) are also implicated (Westby et al. 2013). Interestingly, the same studies suggest that the influence of climate modes upon North American winter warm waves (WWs) is actually greater than the influence on cold air outbreaks (CAOs). General relationships that have emerged from these statistical analyses are illustrated in Fig. 4: The positive (negative) phase of the NAO favors the occurrence of warm (cold) events over the eastern (southeastern) United States. The positive (negative) phase of the PNA tends to favor cold events over the southeastern (northwestern) US.

These connections to climate modes are neither unique nor independent. For example, the regional influence of the PNA pattern on ETEs largely mirrors that of both the PDO and ENSO (Fig. 4) since the midlatitude atmospheric signatures of both ENSO and the PDO project on the PNA pattern. Also, the prevalence of atmospheric blocking patterns is intrinsically linked to particular phases of primary climate modes (Renwick and Wallace 1996). As an example, atmospheric blocking patterns tend to be more frequent during the negative phase of the AO pattern (Thompson and Wallace 2001). In this case blocking may serve as an LSMP conduit that links ETEs to climate modes.

There have been pronounced episodes of climate modes influencing ETEs during recent winters. Cold extremes over Europe and the southeastern United States during recent winters (2009-10 and 2010-11) were primarily accounted for by the anomalous blocking associated with persistent episodes of large amplitude negative phase of the NAO (Guirguis et al. 2011). Although the influence of climate modes upon major temperature extremes is expected to be less prominent during the warm season (when such climate modes are less active), there is nonetheless evidence of an important role for stationary Rossby wave patterns in contributing to North American temperature extremes during summer (Schubert et al 2011; Wu et al. 2012).
These wave patterns appear to arise from internal forcing associated with intraseasonal transient eddies (Schubert et al. 2011).

3.1.2 Connection to remote anomalous surface forcing

Observational studies show that the Arctic sea ice extent varies considerably during any given year with a maximum in February–March and minimum in August–September (Comiso 2006), along with dramatic long-term reduction in sea ice extent (sea ice area and concentration as well) during the last three decades (Jahn et al. 2012). Overland et al. (2008) argues that the significant loss of Arctic sea ice in recent decades is attributed to a combination of factors including natural variability (e.g., Arctic Oscillation (Overland and Wang 2010; Liu et al. 2012)), human impacts such as increased greenhouse gases, and ice–ocean feedbacks. One might think that warming would be observed over large parts of Northern Hemisphere (NH) in conjunction with the recent sea ice reduction. However, the observed warming of the atmosphere attributable to sea ice reduction is confined to high-latitudes northward of ~60°N while the cooling is observed over the mid-latitudes (Kumar et al. 2010; Lim et al. 2012; Liu et al. 2012). Liu et al. (2012) argues that the circulation change due to the decline of Arctic sea ice leads to more frequent events of atmospheric blocking that cause severe cold surges over large parts of northern continents.

As discussed above, two more commonly-recognized remote influences upon North American ETEs are associated with ENSO and the PDO (e.g., Westby et al. 2013), both involving local sea surface temperature anomalies and atmosphere-ocean coupling. These generally operate in conjunction with PNA-like teleconnection patterns that extend from the coupling region downstream into North America. Similar to the effect of climate modes, the impact of remote forcing upon warm season ETEs is limited by the relative inactivity and spatial extent of the primary climate modes, which serve as horizontal pathways for Rossby wave energy between the remote forcing region and the local surface response (Schubert et al. 2011). Nonetheless, Alfaro et al. (2006), studying seasonal controls on summer temperature
extremes, find both remote and local sources of predictability. In this case the PDO was
determined to have a significant impact particularly on summertime minimum temperatures,
while extremes in maximum daily temperatures were more closely related to local soil moisture
conditions. In addition, some recent studies link summertime heat waves to Arctic sea ice loss
(Tang et al. 2013).

3.1.2 Connection to sea ice and snow cover

The process of the atmospheric response to the Arctic sea ice reduction is found to be Arctic
warming and destabilization of the lower troposphere, increased cloudiness, and weakening of
the poleward thickness gradient and polar jet stream (Francis et al. 2009; Outten and Esau
2012). Outten and Esau (2012) revealed that cooling over the NH mid-latitudes is associated
with extreme warming around the Kara Sea attributable to sea ice reduction. As the Arctic
warms faster than lower latitudes, the meridional temperature gradient in the Arctic latitudes
weakens leading to a weaker polar jet. It in turn indicates that the role of the westerly jet in
helping maintain the mild European climate by transporting heat from the Atlantic is interrupted.
Francis and Vavrus (2012) found that a weaker zonal flow (i.e., polar jet) from weakened
meridional temperature gradient slows the eastward Rossby wave progression and tends to
create larger meridional excursions of height contours and associated temperature
displacements resulting in both slower moving circulation systems and higher probability of
extreme weather. Francis et al. (2009), Overland and Wang (2010), Jaiser et al. (2012) and Lim
et al. (2012) found that there is a delayed atmospheric response to the Arctic sea ice.

Specifically, the Arctic sea ice extent in summer to fall influences the atmospheric circulation in
the following winter over the northern mid- to high-latitudes, affecting the seasonal winter
temperature and subseasonal warm/cold spells.

As to the 21st century climate change associated with the long-term trend of the Arctic
sea ice extent, Vavrus et al. (2012) simulated the Arctic climate using the Community Climate
System Model, version 4 (CCSM4). The study found that a much warmer, wetter, cloudier, and
stormier Arctic climate with substantial reduction in sea ice and salinity is predicted under a
strong radiative forcing scenario. Zhang and Walsh (2006) simulates decline of the Arctic ice
coverage, on average by \(-3.82 \times 10^5\) km\(^2\) per decade, or by \(-56.8\%\) by the end of the 21st
century. Gordon and O’Farrell (1997) and Zhang and Walsh (2006) predict faster decline of
summer sea ice extent than winter ice areas in the 21st century, implying that Arctic warming
and mid-latitude cooling could be enhanced through the Arctic sea ice loss since the winter
temperature tends to have a delayed response to the summer to fall sea ice extent.

An analogous forcing of wintertime ETE events over North America is noted to arise
from autumnal variability in Eurasian snow cover (Cohen and Jones 2011). In this case,
autumnal snow cover anomalies induce a subsequent weakening of the stratospheric polar
vortex during winter which, in turn, leads to a persistent negative phase episode of the
tropospheric AO favoring North American cold events (Cohen et al. 2007). Interestingly, there is
new evidence suggesting that the changing Arctic sea ice extent may play a direct role in
bringing about changes in the autumnal advance in Eurasian snow cover (Cohen et al. 2013).
Thus, these two distinct surface forcing mechanisms may be intrinsically linked to one another
with the latter serving as a possible communication pathway for the former in terms of exciting
the AO pattern. We note, however, that there is some uncertainty about the statistical
stationarity of the Eurasian snow cover-AO link (Peings et al. 2013).

27
3.1.3 Large-scale climate "markers" for climate model assessment

Representation of fundamental climate modes. One obvious essential minimum requirement for climate models to properly represent the modulation of ETEs by climate modes is the extent to which the models are able to represent the primary climate modes, themselves. Thus, fundamental markers for model assessment are metrics that measure the representation of key extratropical climate modes including those internally forced on intraseasonal time scales (PNA, AO/NAO and atmospheric blocking) and those externally forced on longer time scales (the extratropical response to ENSO and the atmospheric part of the PDO). Atmospheric models have had historical difficulty in representing some types of intraseasonal low frequency variability (Black and Evans 1998). A particular problem is an under-representation of atmospheric blocking activity (Scaife et al. 2010). In some cases even though the power spectrum of low frequency variability is properly simulated, the underlying structure and dynamical origin may be incorrect (Robinson and Black 2005). This can be linked to parallel errors in representing the three-dimensional structure of the climatological-mean flow (Lee and Black 2013). For example, one energy source for the PNA pattern is the barotropic extraction of energy from the mean flow (Simmons et al 1983). This mechanism relies on the existence of an extended region of diffluence in the exit region of the East Asian jet. In a similar vein, the representation of externally forced extratropical modes connected to ENSO and PDO depends on how well the coupled climate models simulate the associated oceanic phenomenological behavior.

CMIP provides an ideal resource for assessing the ability of modern coupled climate models to represent the behavior of climate modes. Analyses of past versions of CMIP suggest a mixed performance in representing climate modes (Stoner et al. 2009, Casado and Pastor
A recent analysis of CMIP5 models indicates that while most models studied perform well in representing the basic aspects of the PNA pattern, a small subset of models have difficulty qualitatively replicating the NAO pattern (Lee and Black 2013; Table 4). Otherwise, differences among model patterns consist of horizontal shifts or amplitude variations in the circulation anomaly pattern features. It is also determined that models with a well-resolved stratosphere (“high top” models) do no better (in some cases worse) than those without in representing the NAO and PNA. (See Table 4.) CMIP5 models generally underestimate the regional frequency of winter blocking events while summertime blocking events occurring over the high latitude oceanic basins are typically overestimated (Masato et al. 2013). Polade et al. (2013) studied the representation of ENSO and PDO teleconnections to North American wintertime extreme precipitation events. They contrasted the performance of CMIP3 and CMIP5 models and found a modest improvement in the representation of teleconnections emerging from the two main Pacific modes of climate variability to North America. Conversely, Westby et al (2013) found serious deficiencies in the representation of the PDO by CMIP5 models with direct impacts upon the modulation of anomalous temperature regimes.

*Regional flow parameters impacting remote dynamical communication.* The pathway between anomalous remote forcing and the regional circulation response involves several distinct factors. For example, in the case of the extratropical response to tropical heating anomalies, tropical divergent outflow interacts with subtropical vorticity gradients to produce a forced Rossby wave train that extends into the extratropics (Sardeshmukh and Hoskins 1988). The forced Rossby wave train can then dynamically interact with the background extratropical mean flow (such as barotropic deformation in the jet exit region) and the midlatitude storm tracks leading to a “net” large scale circulation pattern extending from the tropics into the midlatitude region of interest (e.g., Franzke et al 2011). One could ascribe an analogous pathway between stratosphere polar vortex variability and the modulation of regional extreme events (via AO/NAO
variability; e.g., Thompson and Wallace 2001). The ability of coupled climate models to accurately represent such pathways is dependent upon a concomitant representation of several regional atmospheric phenomena and circulation structures including:

1) The tropical large-scale circulation response to tropical diabatic heating
2) Upper tropospheric meridional vorticity gradients in the subtropics
3) Barotropic deformation structures in the jet exit regions
4) The structure and intraseasonal variability of the Atlantic and Pacific storm tracks

As discussed earlier, teleconnection pathways are most active during the cool season when the midlatitude jets (and embedded dynamical structures) are strongest, displaced southward and exhibit the strongest variation. We suggest that model validation activities concentrate upon the above features in order to uncover likely sources for existing model deficiencies in leading teleconnection patterns. For both model validation activities and assessment of future changes, it is also important to consider the impact of potential non-stationarity in both the climatological-mean flow (Bond et al 2003) and teleconnection patterns (Johnson et al. 2008).

3.2: Dynamics of LSMPs

3.2.1 Diagnostic tools to study dynamics of LSMP onset/decay

The onset and decay of LSMP structures linked to ETEs generally occur on relatively short intraseasonal time scales. As such, the range of dynamical forcing mechanisms that can directly account for LSMP time evolution is limited to internal atmospheric processes (given the relatively long time scales associated with boundary forcing). The atmospheric processes
affecting LSMP evolution can be local or remote in nature. For example, local synoptic-scale cyclogenesis can usher in a regional Arctic air outbreak while a transient episode of the PNA teleconnection pattern can provide a remote downstream influence on North American extremes. In some cases it is the optimal juxtaposition of both local and remote influences that is required to produce an extreme event.

Ultimately most ETE events are associated with lateral air mass excursions, which are induced by large-scale circulation anomalies in the lower troposphere embedded within the implicated LSMP (e.g., Loikith and Broccoli 2012). The anomalous circulation serves as a “dynamical trigger” for ETE events. We can consider the delineation of LSMP dynamics as a two-stage process: First it is of interest to assess whether the main energy source(s) are local or remote in nature. Once the energy source location is determined, the second stage is to assess the specific physical mechanism providing the proximate energy source in this location. An effective means for assessing the source location regions for large-scale atmospheric waves is the application of wave activity flux analyses (Plumb 1985, Takaya and Nakamura 2001). The wave activity flux is parallel to the Rossby wave group velocity and thereby traces out a three-dimensional pathway between a wave source region (where flux divergence occurs) and a wave sink region (where flux convergence occurs). Past studies have been very successful in applying this diagnostic to make inferences about the wave source location for various large-scale circulation patterns (e.g., Black 1997, Yu et al. 2007).

Possible primary physical mechanisms providing wave activity sources in this context include large-scale barotropic growth, baroclinic growth/instability and nonlinear forcing by synoptic-scale eddies (Evans and Black 2003). These mechanisms may be augmented via secondary feedbacks related to internal diabatic processes or interactions of the LSMP with the local topography or land surface. Past studies have introduced comprehensive dynamical frameworks for studying the dynamical mechanisms leading to the growth and decay of large-
scale circulation anomalies (e.g., Feldstein 2002, 2003). These are based upon a local analysis of tendencies in either the geopotential (Evans and Black 2003) or streamfunction (Feldstein 2002) fields. In these studies local tendencies are decomposed into separate forcing terms that can be related to distinct physical processes. Using such an approach, Feldstein (2002) inferred that the intraseasonal life cycle of the PNA pattern is due to linear dynamical processes characterized by barotropic growth over the North Pacific with energy dispersion downstream to North America. On the other hand, in his 2003 paper Feldstein found that nonlinear dynamical processes (synoptic eddy feedback) play a fundamental role in the intraseasonal life cycles of the NAO pattern. We suggest that the two-stage process outlined above is a generally useful means for the dynamical diagnosis of LSMP life cycles in both observations and climate model simulations.

3.2.2 North American Arctic air mass formation

Both dynamics and thermodynamics have roles in the formation of extreme cold-air masses. Wexler (1936, 1937) postulated that the cold air was formed at the surface by radiative cooling from a snow-covered ground under clear, windless conditions, creating an intense temperature inversion restricted to a very shallow layer at the surface. Actual soundings, however, indicated a deeper layer of cooling than could be explained by this model because the subsidence and high stability accompanying the surface cooling restricts the possibility for upward mixing of the cold air. Finding that the Wexler model could not adequately explain the depth of the cold layer observed in soundings, Gotaas and Benson (1965) showed that the existence of suspended ice crystals were crucial to upper-level cooling. Curry (1983) later modeled the effect with the introduction of condensate, particularly ice-crystals, in the layer. The study used a one-
dimensional model following the trajectory of air over a homogeneous snow surface. In an experiment without moisture, the inversion that formed after 2 weeks of radiative cooling, was less than 1000 meters deep. The effect of various parameters including turbulent mixing and subsidence were examined. Only with the addition of ice crystals did the inversion rise above 900 hPa. More recently, Emanuel (2009) also used a single-column radiative-convective model, beginning with a tropical sounding, to simulate the formation of arctic air. The rates and depth of cooling in the model were sensitive to the amount of water vapor and clouds. When clouds were allowed to form, they cooled the layers at their top through radiation while warming the lower layers, both radiatively and with latent heat release.

Turner and Gyakum (2011), in their composite study of 93 Arctic air mass formations, found that the cold air mass lifecycle consisted of a multi-stage set of processes. Each event is defined as three or more consecutive days of minimum temperature anomalies below one monthly standard deviation for at least five of the ten first-order meteorological stations in northwestern Canada. During the first stage, snow falls into a layer of unsaturated air in the lee of the Rocky Mountains, causing moisture increases in the sub-cloud layer. Simultaneously, the mid-troposphere is cooled by cloud-top radiation emissions. On the second day, snowfall abates, the air column dries and clear-sky surface radiational cooling predominates, augmented by the high emissivity of fresh snow cover. The surface temperature falls very rapidly, up to a maximum of 18°C day⁻¹ in one event. On the third day, after near-surface temperatures fall below their frost point, ice crystals and, nearer the surface, ice fog form. At the end of formation, there is cold-air damming (Forbes et al. 1987; Fritsch et al. 1992), with a cold pool and anticyclone in the lee of the Rockies, lower pressure in the Gulf of Alaska and an intense baroclinic zone oriented northwest to southeast along the mountains.

Figure 5, illustrating the LSMPs of a subset (53) of these cases, shows surface ridging building southeastward along the eastern slopes of the Rockies. The weather associated with
the in-situ cooling of the air mass is primarily snow, rather than clear skies. The authors argue that diabatic cooling is needed to explain the cooling observed. This cooling may consist of a combination of sublimational cooling from precipitation falling into a dry layer, and/or radiational cooling from suspended ice crystals.

Turner et al. (2013), in a case study of an extremely cold arctic air mass formation and maintenance during 27 January through 13 February 1979 in northwestern Canada, found structures and mechanisms similar to those found in the air mass composite study. What was particularly distinguishing about this case was its extreme longevity (17 days, compared with an average length of five days). Additionally, its dissipation period was characterized by the cold air migrating southeastward, with the associated strong jet stream being responsible for a clustering of explosive cyclogenesis from 10-16 February 1979 (Sanders and Gyakum 1980). Subsequent to this cyclone clustering period was the cold-air mass' association with the often-studied Presidents' Day Cyclone of 18-19 February 1979 (e. g., Bosart 1981; Bosart and Lin 1984; Spaete et al. 1994; Uccellini et al 1984, 1985).

3.2.3 CONUS wintertime cold air outbreaks

Cold air outbreaks over North America consist of a two-stage process: The first stage is the formation of an Arctic air mass surface anticyclone at higher latitudes over Canada. This is followed by the rapid horizontal transport of the air mass to lower latitudes. The latter stage is enabled by the lower-tropospheric circulation anomaly embedded within LSMPs. North American cold air outbreaks are typically associated with a 500 hPa geopotential height anomaly pattern consisting of a broad region of negative anomalies over the CAO region, itself, along with a positive anomaly feature located to the northwest (figure 6, panels Jan Tx5 and Jan
This 500 hPa dipole structure is linked to a positive sea level pressure anomaly feature that extends between the upper level features along with a considerably weaker negative SLP anomaly located to the southeast. Thus, the CAO events are linked to a near-surface northerly flow embedded within a westward tilting anticyclonic circulation anomaly. Loikith and Broccoli observe that although these circulation anomaly patterns resemble large-scale teleconnection patterns, the inherent spatial length scale is closer to synoptic scale. This illustrates the importance of local synoptic features in the life cycles of ETE events. On the other hand, recent papers by Westby et al (2013) and Loikith and Broccoli (2014) clearly demonstrate the existence of a statistically significant modulation of ETEs by larger-scale teleconnection patterns. It is likely that, in some cases, both synoptic-scale and larger-scale elements are required in order for the ETE event to occur.

West coast events (e.g., CA Central Valley events). Compared with cold air outbreaks affecting the eastern and central parts of the US, less has been written about outbreaks affecting the western US. Extreme cold over California is related to a large scale pattern that brings cold air from the Arctic and northern Canada without crossing the Pacific, and hence over the Rocky Mountains (Grotjahn and Faure 2008). As discussed in Grotjahn and Faure, the large scale pattern is similar to outbreaks affecting the areas east of the Rockies however with a primary difference being a small but statistically significant ridge over the southeastern US. (See figure 2a,b.) This ridge precedes the CAO by several days, including the large ridge that develops subsequently over Alaska. A cold air mass between these ridges is directed further westward (a more northerly or northeasterly flow of the cold air) than for eastern CAOs. The distortion of the flow by that southeastern ridge enables the cold air mass to cross the Rockies. Loikith and Broccoli (2012) capture a portion of this pattern in geopotential heights at 500 hPa using their ‘grand composite’ technique. Regional properties are compared to that grand composite and that grand composite does include a weak small ridge near the southeastern US.
when the grand composite is shifted to locations in the far western US. Loikith and Broccoli also show a sea level pressure (SLP) composite for cold in the southwest deserts having higher pressure to the northwest and lower to the southeast with consequent forcing of northerly winds. More local to the area of the extreme, colder temperatures are associated with an adjacent surface high (e.g. the conceptual model of Colle and Mass 1995). Favre and Gershunov (2006) pursue this link to anticyclones affecting western North America; and develop indices based on frequency and central pressure of transient cyclones and anticyclones in the eastern North Pacific. Their 'CA' index is the difference, cyclone minus anticyclone, in strength and frequency. Negative CA values are correlated with the coldest 10% of winter minimum temperatures. Cassano et al. (2011) calculate self organizing maps of SLP and identify those associated with colder temperatures at 5 Alaskan stations; in broad terms, the colder temperatures are associated with higher SLP and flow implied by the SLP pattern that advects air from colder regions.

Eastern US events (i.e., east of Rockies). Several statistical and synoptic studies either directly or indirectly relate to the topic of US CAOs occurring east of the Rockies (e.g., Konrad and Colucci 1989, Colle and Mass 1995, Konrad 1996, Walsh et al 2001, Celliti et al 2006, Portis et al 2006). Common precursors to eastern CAO events include anomalously high surface air pressure over western Canada (linked to a polar air mass), occurrences of the negative (positive) phase of the NAO (PNA) teleconnection pattern and/or an anomalously weak stratospheric polar vortex. CAO onset is characterized by the south/southeastward advection of the polar air mass in association with one or more of the following synoptic LSMP features: southward extension or propagation of the surface high pressure from Canada, surface cyclogenesis over the eastern US, anomalously low 500 hPa geopotential heights over the Great Lake region and the southeastward movement of an upper level shortwave from Canada.
Distinguishing between east coast (EC) and Midwest (MW; just east of the Rockies) CAO events, EC events result from the geostrophic cold advection associated with a southeastward propagation of the surface high pressure system (often coincident with surface cyclogenesis further east) leading to the detachment of the polar air mass from its source (Konrad and Colucci 1989, Walsh et al. 2001). MW events are linked to the southeastward meridional extension of surface high pressure from Canada leading to a more continuous cold air transport (Walsh et al. 2001). For MW events occurring just east of the Rockies, the cold advection is related to low-level northerly ageostrophic flow within a region of cold-air damming against the topography (Colle and Mass 1995). In all cases the salient LSMP features (cyclogenesis, anticyclone extension/propagation, upper level wave propagation and horizontal circulations) are synoptic-scale in nature and, thus, governed by generally well-understood dynamical physical principles considered to be well-represented by weather and climate models. The CAO process, itself, is largely due to radiative cooling (during polar air mass formation and early stages of CAO onset) and southward air mass transport (by the horizontal wind) partly offset by adiabatic warming induced by large-scale subsidence (Konrad and Colucci 1989, Walsh et al. 2001, Portis et al. 2006). Although the proximate physics associated with the CAO process itself is reasonably well characterized, our knowledge of the physics of implicated LSMP patterns is implicit and depends on subjective associations (between advective circulation features and synoptic LSMPs) that have been made in earlier studies. As such, there is an existing need to (a) more objectively link critical advective CAO circulation features to dynamical entities (e.g., in terms of potential vorticity attribution) and (b) subsequently assess the physical origin of these dynamical entities. The local analyses presented in the composite study of Loikith and Broccoli (2012) represents one step toward achieving (a).
3.2.4 Summer heat wave events

Hot spells over North America have an intense upper level ridge as one expects from the hypsometric equation. This ridge is seen in figures 2 and 6. There can also be a shallow layer of relatively lower pressure, sometimes called a ‘thermal low’. Hot spells occur from a combination of factors: a hot air mass displaced from its normal location (displaced from the southwestern desert), strong subsidence (causing adiabatic warming), and solar heating. Solar heating is more effective when surface latent heat fluxes are constrained (by drought, for example). The first two factors are enhanced and help define the LSMP for a hot spell; there may be an upstream ridge (seen in figures 2 and 6) which enhances southwesterly advection, presumably of warmer air. Other factors influence how the LSMP develops. What follows are details for different regions of North America.

West coast events. Heat waves affecting the west coast are linked to an upper air LSMP that has a local ridge. For California Central Valley heat waves, that ridge is typically aligned with the west coast and the LSMP also has upstream features: a trough south of the Gulf of Alaska and a ridge further west, south of the Aleutians (Grotjahn and Faure 2008). See figure 2. The local ridge is easily understood as resulting from high thickness due to the anomalous high temperatures through the depth of the troposphere generally centered along the west coast of the US. (Plots can be found here: http://grotjahn.ucdavis.edu/EWEs/heat_wave/heat_wave.htm) Grotjahn and Faure also show that a statistically significant ridge in the northwest Pacific develops prior to the significant intensification of that west coast ridge. (Grotjahn and Faure also find a significant ridge the southeastern US prior to onset that is not present in a longer record of more events; Grotjahn 2014.) These features are essentially equivalent-barotropic through the depth of the troposphere. While Grotjahn and Faure found variation amongst the events, the ensemble average consists of a significant temperature anomaly that in the lower troposphere
(850 and 700 hPa) is located just off the Oregon and northern California coast and elongated meridionally. The narrow zonal with longer meridional scale is consistent with Bachmann (2008). Bachmann showed that extreme surface temperature dates in Sacramento were more frequently matched by corresponding extreme dates at places far to the north (e.g. Seattle) than with places across the Sierra Nevada or Cascade ranges to the east (e.g. much closer Reno). Grotjahn (2011) shows that the temperature anomaly leads to a thermal low being at the coast which sets up a low level pressure gradient that opposes penetration by a cooling sea breeze. The lower and mid-tropospheric flow has anomalous significant easterlies which are also downslope over some regions, most notably the Sierra Nevada mountains. Finally, there is notable sinking that lowers the climatological subsidence inversion and hence sunlight rapidly heats up the shallow layer beneath. (These factors are seen in figure 7).

Gershunov et al. (2009, hereafter GCI) make the novel distinction of considering separately heat waves that have high daytime maximum and events with high nighttime minimums. In so doing they uncover an LSMP with unusually high values of precipitable water (PW) over their region of interest occurring during their ‘nighttime’ heat waves (but not their ‘daytime’ events). GCI emphasize a trend of increasing occurrence of ‘nighttime’ events. Gershunov and Guirguis (2012) also find that trend in all their sub-regions of California and also show another trend: increasing longitudinal extent of Central Valley heat waves. GCI show maps of anomalous geopotential heights which also show the west coast ridge (height anomaly centered over Washington State) with upstream (and downstream) troughs, consistent with Grotjahn and Faure. GCI show the LSMP for SLP (high over the Great Plains and low off the west coast) and remark that both daytime and nighttime events have a corresponding general southeasterly flow of air out of the desert southwest. This flow occurs throughout the lower troposphere with associated high frequency (<7 day period) heat fluxes that are prominent just off the coast (Grotjahn 2014); and consistent with setting up the offshore pressure gradient (to
oppose a cooling sea breeze). GCI emphasize the record-breaking event during July 2006 (Blier 2007) since it is extreme both as a daytime and nighttime event. GCI consider heat waves affecting the combined region of California and Nevada. In the past 35 years that 2006 event has the third highest temperature value and anomaly in Sacramento, exceeded recently only by the great September 1988 heat wave (which affected the west coast, northwest US and southwest Canada more than Nevada) and a short extreme event in early July 1991. Gershunov and Guirguis (2012) discuss how heat waves differ between 6 zones of California. Gershunov and Guirguis state that a sea breeze can keep temperatures low in the narrow strip along the coast even as inland temperatures soar. In contrast to other zones, extreme heat along the coastal zones occurs for off shore flows in late summer (September) especially for ‘daytime’ events on the southern California coast due to the development of Santa Ana conditions.

For heat waves affecting the western areas of Washington and Oregon, Bumbaco et al. (2013, hereafter BDB) show an upper level (500 hPa) ridge over the west coast and trough upstream over the Gulf of Alaska. That trough and ridge pattern is generally similar to the pattern for California heat waves, including the tropospheric height anomaly centered at the coast (Grotjahn 2011). Using a regression approach Lau and Nath (2012) find the upstream trough and a ridge even further upstream (over the western North Pacific) which is stronger several days prior to the event onset both the location and timing are consistent with Grotjahn and Faure. BDB find a significant trough downstream (over the central plains of southern Canada and northern US) but such a significant downstream trough is not seen by Grotjahn and Faure for California heat waves. Lau and Nath find a downstream trough (centered over the eastern North America) develops after the event. Similar to Grotjahn (2011, 2013, 2014), BDB find the lower tropospheric temperature anomaly (850 hPa) to be centered at the west coast of North America. Hence, BDB find negative values of sea level pressure anomaly centered offshore which sets up an offshore low level pressure gradient and offshore (northeasterly) flow.
(a similar pattern for California heat waves is discussed in Grotjahn 2011, 2014). Consistent with the heat fluxes shown in Grotjahn (2014), the near-coast thermal trough appears to migrate from the southwestern deserts across the length of California and finally reaching western Oregon and Washington (Brewer et al. 2012). Similar to GCI, BDB separately consider nighttime heat waves. While the peak values of the height anomaly ridge are less for nighttime than daytime events, the difference is not quite significant. BDB find significantly low PW content during daytime events and significantly high PW at the end of nighttime events affecting their study region, but BDB did not find the difference to be significant at the onset.


For at least some events, the continental anticyclonic flow at mid-levels is part of a larger pattern of anomalies with remote centers over the Pacific and Atlantic (Namias 1982, Chang and Wallace 1987, Livezey and Tinker 1996, Lau and Nath 2012, Lokith and Broccoli 2012, Teng et al. 2013), with some suggestion that the anomalies are generated in the Pacific (Namias 1982, Lyon and Dole 1995, Livezey and Tinker 1996, Teng et al. 2013), with possible predictability (Teng et al. 2013). Teng et al. (2013) have noted the similarity of the circulation pattern to a Rossby wave number 5 wavetrain in the jetstream waveguide. As previous studies have used a range of definitions and domains, it is somewhat difficult to assess the commonality of the wave train and the details of its structure and evolution.
Studies of the surface flow have been less common, but negative SLP anomalies have been noted (Chang and Wallace 1987, Lau and Nath 2012, Loikith and Broccoli 2012), and Lau and Nath (2012) have shown anomalous southerly (or southwesterly) flow at the surface, consistent with strong horizontal warm advection. The regional-scale flow also interacts with local mechanisms, particularly the Urban Heat Island (UHI) effect (Kunkel et al. 1996, Palecki et al. 2001).

Another common element identified in several studies is the presence of drought (Chang and Wallace 1987, Karl and Quayle 1981, Namias 1982, Lyon and Dole 1995) as well as simultaneous precipitation deficits (Lau and Nath 2012). Despite this link with dry conditions, the presence of high values of low-level moisture has played an important role in high dewpoints and related societal impacts in several severe heat waves (Kunkel et al. 1996, Livezey and Tinker 1996, Palecki et al. 2001).

*Eastern events.* While the large-scale meteorological patterns associated with heat waves in the eastern US have not yet seen much study, they have been examined for the Northeast and Central US Gulf Coast as part of the study of Lau and Nath (2012) and for piedmont North Carolina by Chen and Konrad (2006). For both the Northeast and Gulf Coast, Lau and Nath found a mid-level ridge centered over the region as part of a wavetrain and precipitation deficits, broadly similar to the results for the Midwestern regions, although the wavetrain was not as well-defined for the Gulf Coast events. For piedmont North Carolina, Chen and Konrad showed that, at upper levels, a strong ridge over or just upstream of the region was a common feature, and, at lower levels, adiabatic warming was a common feature, associated with an arm of the Bermuda High extending northwestward from the Gulf of Mexico and adiabatically warming as it descended the Blue Ridge mountains into the piedmont region.
4. Modeling of temperature extremes and associated circulations

Temperature extremes have some predictability at weather and climate time scales that can be exploited by models to improve short term predictions and long term projections of heat waves and cold air outbreaks. At weather time scales the predictability of North American extreme temperature events (ETEs) is largely dependent upon the nature of the LSMPs that help organize their occurrence. The greatest predictability is expected to occur during the boreal cool season during which time ETEs are, at least in part, influenced by low frequency modes (PNA, NAO and blocking events) with intrinsic time scales of several days to weeks (Feldstein 2000). Both warm waves and cold air outbreaks over North American are affected by the PNA and NAO teleconnection patterns (Westby et al. 2013). However, this relationship is a statistical modulation of ETE behavior and does not guarantee their occurrence (i.e., necessary but not sufficient). The observational study of Loikith and Broccoli (2012) illustrates that the local LSMPs linked to ETEs generally exhibit synoptic spatial scales rather than planetary scales. Implicated synoptic-scale disturbances include east coast cyclones and southward moving polar anticyclones over the Midwest (Konrad and Colucci 1989, Walsh et al. 2001). Given the essential role of synoptic-scale phenomena in ETE, pointwise predictability of ETEs is ultimately limited by our ability to forecast the details of synoptic-scale phenomena several days in advance (Hohenegger and Schär 2007).

On the other hand, since the statistical likelihood of ETEs is modulated on longer time scales by low frequency modes, there may also be useful probabilistic predictability on time scales ranging from a week (in the case of PNA/NAO or blocking) to a season (for ENSO or the PDO). Such probabilistic predictability is dependent upon (a) the strength of the regional ETE modulation by a low frequency mode and (b) the intrinsic time scale of the low frequency mode.
involved. In some cases, the constructive interaction of multiple low frequency modes may serve to increase this predictability (e.g., Westby et al. 2013). Hence a significant challenge for models to predict temperature extremes is their ability to predict or simulate the synoptic scale phenomena and low frequency mode that provide the large-scale meteorological context for the extreme events. However, the surface temperature response to the large scale circulation anomaly is strongly dependent on both atmospheric processes such as turbulence mixing, cloud and radiation, and land surface processes and properties that influence the surface heat fluxes. Both dynamical and statistical models have been developed and used to simulate and project changes in temperature extremes. This section briefly summarizes the methods and skills of the models, and analysis of observed trends and projections of future trends are summarized in Section 5.

4.1 Global and regional climate model skill in simulating temperature extremes

Both global and regional climate models have been used to study temperature extremes to elucidate contributing processes and evaluate model skills. The ability of global climate models to reproduce the observed temperature extreme statistics has been assessed by Sillmann et al. (2013a,b), Westby et al. (2013) and Wuebbles et al. (2014) using the Coupled Model Intercomparison Project (CMIP5) multi-model ensembles. Figure 8 shows a performance portrait of the normalized root mean square errors over the North American land area from the 28 available CMIP5 “historical” models for eight temperature based ETCCDI indices over the period 1979-2005 as compared to the ERA Interim reanalysis (Dee et al. 2011). As discussed in Gleckler et al. (2008), in order to plot errors from multiple variables on the same scale, they are normalized by the median error of the CMIP5 models using the formula
Here, $E_{\text{median}}$ is the median root mean square error (RMSE) of the CMIP5 models, $E_j$ is the RMSE of the $j^{th}$ model and $E_j^R$ is that model’s “relative RMSE” and is plotted for seasonal means of the indices. In this analysis, the median RMSE in equation 1 is calculated for each variable over all seasons and then applied to normalize each season in order to assess the relative seasonal performance. Blue colors represent errors lower than the median error, while red colors represent errors larger than the median error. Seasons are denoted by triangles within each square. The different models are arranged in order of increasing average relative error with the models with the lowest average relative error on the left. The average relative error is positive for all 8 indices indicating that figure 8 has more deep red than deep blue colors even though the number of positive and negative relative errors is equal. Spring has significantly lower relative error (3.8%) averaged across all models and variables, compared to winter (5.0%), summer (5.2%) and fall (5.8%).

Sillmann et al. (2013a) found that the spread amongst CMIP5 temperature extreme indices tends to be smaller than that of the CMIP3 models (earlier version of the CMIP5), indicating reduced uncertainty. They identified an increase in the monthly maximum daily maximum temperature and a decrease in the monthly minimum daily minimum temperature by the CMIP5 models over the high northern latitudes, compared to CMIP3. Globally, upward trends were found over the period from the latter half of the 20th century through the early 21st century (i.e., 2005) for some temperature extreme indices such as the hottest day of a year, the coldest night of the year, the number of summer days (daily maximum temperature exceeding 25°C) and tropical nights (daily minimum temperature greater than 20°C). In contrast, the numbers of frost days (daily minimum lower than 0°C) and ice days (daily maximum lower than 45
0°C) exhibited downward trends. The study found that these trends are also reflected in the percentile indices (Alexander et al. 2006; Morak et al. 2011; Sillmann et al. 2013a). Analysis of twenty year return values in the CMIP3 models by Zwiers et al. (2011) found that models tend to warm the minimum temperature less than observed and that they warm the maximum temperature more than observed over the latter half of the 20th century.

Westby et al. (2013) analyzed reanalysis data and CMIP5 historical simulations to study the statistics and low frequency mode modulation of wintertime cold air outbreaks (CAOs) and warm waves (WWs) occurring over the continental United States during 1949-2011. The observational results indicate (a) a lack of significant long-term trends in ETE frequency and (b) a seasonal modulation of CAOs (WWs) by the NAO and PNA (NAO, PNA, ENSO and PDO) patterns (Fig. 4). Similar behavior is found in the CMIP5 models, showing CAOs mainly modulated by the NAO over the southeastern US and by the PNA over the northwestern US, while WW frequency was modulated by the NAO over the eastern US and by a combination of the PNA, PDO and ENSO over the southern US. The study also found modest influence of ENSO on WWs over the southern US, in good agreement with Lim and Schubert (2011).

Comparison of WWs and CAOs between the CMIP5 models and observation indicated that the models tend to overestimate WW frequency, but underestimate CAOs frequency.

Overall, CMIP5 models properly represent many of the significant associations between ETEs and low frequency modes, particularly the modulation by the NAO and PNA patterns. Similar to Lee and Black (2013) who noted model deficiencies in representing low-frequency variability, Westby et al. (2013) found that the CMIP5 models underestimated the collective influence of low-frequency modes on temperature extremes (e.g., Fig 9). One notable model failure is the virtual absence of a seasonal modulation of ETEs by the PDO. This is largely attributed to the inadequacy of the CMIP5 models in representing the nature and physics of the PDO. On the other hand, CMIP5 models do a considerably better job in representing the
general behavior of the NAO and PNA patterns (Lee and Black 2013). Westby et al. conclude that predictions of future ETE behavior are ultimately limited by the ability of state-of-the-art coupled climate models to properly represent the (evolving) behavior of prominent low frequency climate modes. Nonetheless, consistent with observations, little evidence of significant trends (in either WW or CAO frequency) is found in the model simulations over the continental U.S. from 1949 - 2011. The presence of volcanic forcing signal in the CMIP5 model data was examined by Sillmann et al. (2013a), who found a decrease in the number of warm days and nights and an increase in the number of cold days and nights in years following major volcano eruptions.

Heat waves are associated with anomalous large-scale circulation patterns and cover large spatial extent and therefore should be well resolved by GCMs. Grotjahn (2013, 2014) describes how the LSMPs for California hot spells simulated by a CMIP5 climate model compare with the observed LSMPs; while the model generates a similar large scale pattern and included the needed large scale sinking, the model is unable to simulate sea breezes so the effect of LSMP blocking sea breezes during extreme heat episodes is likely not captured by the global model. In general, heat wave intensity and duration can vary at local to regional scales because surface temperature is influenced by other processes such as land surface fluxes, turbulence and winds, and clouds and radiation. For example, land-atmosphere interactions were found to play an important role in European summer heat waves of the past decades (Fischer et al. 2007). Using a regional model, Jaeger and Seneviratne (2011) and Lorenz et al. (2010) highlighted the key role of soil moisture memory on heat wave intensity and persistence, respectively. Gershunov and Douville (2008) considered the spatial scale of heat waves and diagnosed heat wave activity over Europe and North America from observations and a CMIP3 climate model. They showed a strong negative relationship between preceding winter and spring precipitation and summertime heat in the central, Midwestern, and eastern US. In
regions with limited soil moisture, surface temperature can respond more strongly to the large-scale circulation anomaly that leads to clear sky and strong surface solar heating because surface cooling from evapotranspiration is limited. As soil moisture is influenced by antecedent precipitation and surface properties such as land cover, topography, and soil characteristics that vary at local to regional scales, heat wave intensity and duration can vary similarly, within the context of a large-scale influence of circulation anomaly. Thus the ability of climate models to simulate and predict heat waves is also dependent on physics parameterizations and grid resolution.

Lau and Nath (2012) analyzed GCM simulations performed at 200km and 50km grid resolutions and compared the simulated heat waves and associated meteorological conditions with the North American Regional Reanalysis (NARR). They found considerable resemblance between the temperature anomaly patterns from the coarse and fine resolution simulations, except for some local details generated by the higher resolution model. Both models were able to capture realistic heat wave intensity, duration, and frequency in various key regions of North America and the synoptic features accompanying the warm episodes as revealed in NARR. The impacts of model resolution on simulations of heat waves were more systematically investigated by Kunkel et al. (2010), who compared regional climate simulations performed at 30km resolution over North America driven by two GCM historical simulations with the respective GCM simulations and observations. Although the two GCM simulations have opposite temperature biases, downscaling by the RCM provided improved simulations of heat waves in both cases, with the overall averaged biases reduced by a factor of two compared to the GCMs. Gao et al. (2012) compared a regional simulation at 4km grid resolution over the eastern U.S. driven by a GCM with observations and the GCM simulation. They found statistically significant improvements in heat wave intensity and duration in the regional simulation compared to the global simulation in 14 out of 16 states in the eastern U.S. They attributed the improvements to
In summary, global climate models have demonstrated useful skill in simulating heat waves and cold air outbreaks over North America and their linkages to low frequency variability such as NAO, PNA, and ENSO. However, models generally underestimate the combined influence of low frequency mode on temperature extremes, with PDO and its modulations on temperature extremes being most notably deficient. Although heat waves and their LSMP are well resolved by global models, some studies have demonstrated improved skill in simulating heat waves using regional models that may better capture regional processes such as land-atmosphere interactions influenced by surface heterogeneity.

4.2 Statistical modeling of temperature extremes

Most statistical methods for downscaling of temperature can be thought of as extensions of the model output statistics and related approaches to statistical weather forecasting (e.g., Wilks 2006). The most popular method is multiple regression analysis (e.g., Easterling 1999). Yet regression methods cannot be expected to represent extremes well, as only the conditional mean of temperature given the values of the predictor variables (such as indices of LSMPs) is statistically modeled. In particular, the error term in the regression equation is usually assumed, either explicitly or tacitly, to be normally distributed (i.e., not necessarily a realistic distribution for temperature extremes).

Recently, bias correction techniques involving quantile mapping (Thrasher et al. 2012), sometimes called quantile matching (Ho et al. 2012), have become popular to adjust the entire probability distribution of temperature. Because of the mismatch in spatial (and sometimes
temporal) scales involved in downscaling, the downscaled temperatures typically do not possess high enough variance. This issue is crucial for extremes, with it being preferable to increase the variance through randomization rather than direct “inflation” (Maraun 2013; von Storch 1999). Ideally, quantile mapping for the extreme tails of the temperature distribution should use statistical methods based on extreme value theory (e.g., Coles 2001). In fact, Kallache et al. (2011) proposed such a method, in which the quantile mapping involves fitting the Generalized Pareto (GP) distribution to the upper (or lower) tail of the distribution. Although their application is to precipitation extremes, the technique should apply equally well to temperature extremes.

The statistical methods for relating extreme temperatures to LSMPs and climate modes, described earlier in Section 2, could also be applied in the context of statistical downscaling. In this approach, the extremal distribution (GEV or GPD) is conditioned on indices of LSMPs or climate modes, as well as possibly on other covariates in the context of statistical downscaling. Brown et al. (2008) relate daily maximum and minimum temperatures to the NAO index; Sillmann et al. (2011) relate European CAOs to a blocking index; Katz and Grotjahn (2014) relate California hot spells to an associated LSMP index. The chief limitation is that, given the effective reduction in data when only considering extremes, not as many predictors necessarily could be included as in more conventional statistical downscaling.

5. Observed and projected trends of temperature extremes

5.1 Observed trends and variability
Traditional methods in climate research for detecting trends in temperature extremes tend to focus on the relative frequency of exceeding a high threshold (e.g., the 95th or 0.99th percentile of daily maximum temperature) or of falling below a low threshold (e.g., the 1st or 5th percentile of daily minimum temperature), as well as indices derived from such rates (Karl et al. 1996, Gleason et al. 2008). Such approaches generally fail to give a complete picture of trends in temperature extremes. In particular, examining only the rate of occurrence of an extreme event (e.g., the event of exceeding a high threshold) neglects the intensity of the event (e.g., in terms of the “excess” over a high threshold, or by how much the threshold is exceeded). It should be noted that other forms of indices are also commonly used to monitor changes in temperature extremes (Zhang et al. 2011).

Extreme value statistics, methods based on the statistical theory of extreme values (Coles 2001), formally account for both the rate of exceeding a high threshold and in the excess over a high threshold. In particular, they can be applied to detect trends in each of these two components. Yet their application in the context of non-stationary climate extremes has been quite limited so far (Katz 2010).

A fundamental result in extreme value theory is that block maxima (e.g., summer highest daily maximum temperature), suitably normalized, have an approximate generalized extreme value (GEV) distribution. In practice, it is more efficient and can be more informative to fit the GEV indirectly using the peaks over threshold (POT) approach. This approach involves two components: (i) a Poisson distribution to approximate the rate of exceeding a high threshold; and (ii) a generalized Pareto (GP) distribution to approximate the excess over the high threshold. Possible trends in extremes can be introduced into these extreme value distributions through using time as a covariate. That is, the location, scale, and shape parameters of the GEV (or, equivalently, the rate parameter of the Poisson distribution and the scale and shape parameters of the GP distribution) can possibly shift with time (Coles 2001). See Brown et al.
for examples of the application of the POT approach to trend analysis of temperature extremes.

Recently, a few papers have made use of extreme value statistics in attempts to detect trends in observed temperature extremes, either for the U.S. alone or globally. Most noteworthy, Peterson et al. (2013) include such a trend analysis for the U.S. by applying the POT approach with a threshold corresponding to the 99th percentile for the upper tail (or 1st percentile for the lower tail) specific to each grid point. To avoid temporal dependence of extremes, only the single highest (or lowest) daily temperature within a run of consecutive days exceeding the threshold was retained (termed “runs declustering” in the extreme value statistics literature; Coles 2001). Using the equivalent parameterization in terms of the GEV distribution, a linear trend in the location parameter of the GEV was fitted by the technique of maximum likelihood (Coles 2001), assuming that the scale and shape parameters were constant over time. To adjust for seasonality, the analysis was performed in terms of daily anomalies.

This approach was applied to daily maximum and minimum temperature from the HadEX Global Climate Extremes Indices database (Alexander et al. 2006) for grid points in North America over the time period 1950-2012. Figure 1 shows the estimated change in 20-yr return levels between 1950 and 2012, along with an indication of statistical significance. The broadest region of warming occurs for the cold tail of minimum temperature (panel d). Cooling occurs in the upper tail of both daily maximum (panel a) and minimum (panel c) temperature in the southeastern US. Although these results are not explicitly in terms of heat waves or cold air outbreaks, they do reflect changes in extremes often part of such multi-day events (Furrer et al. 2010).

Atmospheric reanalyses have provided a significant source of data for weather and climate research because of the assimilation of the multitudes of available observing systems
and the ability to provide continuous fields through regions or variables where direct observation
is either sparse or non-existent. Beyond the identification of extreme events, reanalyses allow in
depth evaluation of the general circulation and regional weather and processes at the onset,
development and decay of extreme occurrences. For example, drought conditions often develop
and are maintained by both large-scale circulation and local heating feedback (Trenberth and
Guillemot 1996). Nonetheless, reanalyses, like purely observed data sets, have associated
uncertainties that should be considered, especially when evaluating extreme temperature
conditions.

Reanalyses provide the large-scale environment associated with extremes over
decades, and have been used to identify the mechanisms behind extreme weather conditions.
For example, Francis and Vavrus (2012) used reanalyses thickness fields to examine the
connections between Arctic warming and mid-latitude storm tracks and persistent blocking
events that lead to drought, flood and heat waves in a climatological sense. Lau et al. (2012)
examine the large scale weather conditions associated with a coupling of a drought and flood,
utilizing not only the reanalyses atmospheric circulation data, but also atmospheric energy
budget data.

Utilizing reanalyses in the calculation of extreme indices should be approached with a
strong validation strategy (as discussed in Zhang et al. 2011). In the case of temperature,
processes near the surface can influence the value of temperature and its biases enough to
influence the distribution of temperature. However, Simmons et al. (2010) show that ERA
Interim reanalysis surface temperature reproduces that of HadCRUTv very closely, owing to the
correction of soil water by the analysis of near-surface atmospheric water and temperature.
Further, Bosilovich (2013) shows that even without surface analysis, reanalyses can provide
robust interannual variability of seasonal surface temperature at regional scales. Even so,
reanalysis temperature, augmented with observations, can provide a substantially improved
representation of the diurnal cycle of temperature (maximum and minimum) consistently over the global land mass (Wang and Zeng 2013).

The continuous nature of reanalysis data permits calculation of diagnostics everywhere with every time, which may have some advantage in detecting occurrences of extreme events. For example, two reanalyses, the Modern-Era Retrospective-analysis for Research and Applications (MERRA) (Rienecker et al. 2011) and the Climate Forecast System Reanalysis (CFSR) (Saha et al. 2010) have stored data at 1 hourly intervals which will allow evaluation of the diurnal cycle, especially maximum and minimum temperatures that are important to extreme temperature events. Figure 10 shows the U.S. summertime 2011 Warm Spell Duration Index (WSDI, e.g. Alexander et al. 2006) computed from the occurrences of daily mean temperature exceeding the 90th percentile for 6 days or longer from MERRA. This diagnostic relies on continuous data to derive the seasonal index, but also identify the spatial extent. Figure 11 shows the US trends in WSDI from MERRA. Positive trends are apparent in many areas of the US with some regions having statistical significance (above 90% confidence). Of course, the diagnostic is intended to examine the wings of the distribution of temperature, and it may have different characteristics than large-scale mean temperature changes.

Walsh et al. (2001) used global reanalysis data to study cold air outbreaks and identified Midwest, Gulf Coast, and East Coast as regions influenced by CAOs. They did not identify any apparent trend in CAOs in North America between 1948 and 1999, although analysis from a single station suggests that extreme outbreaks may have been more frequent in the late 1800s and early 1900s. Extending the study to include the recent decade from the global reanalysis data may provide further confidence in potential trends in CAOs in North America.

5.2 Projected trends
In addition to analysis on the present climate, the CMIP3 and CMIP5 multimodel ensembles have also been used to investigate projected changes of temperature extremes for the next couple of decades (e.g., mid 21st century) (e.g., Tebaldi et al. 2006; Orlowsky and Seneviratne 2012; Sillmann et al. 2013b; Wuebbles et al. 2014). Climate change simulations using different emission scenarios in the CMIP5 multimodel ensembles showed greater changes in ETCCDI based on daily minimum temperatures than in ETCCDI based on daily maximum temperatures. Also, the strong emission scenario such as Representative Concentration Pathways (RCPs) 8.5 (Moss et al. 2010; van Vuuren et al. 2011) produced more definitive changes in temperature extreme statistics than weaker emission scenarios (e.g., RCP 2.6 and 4.5). Furthermore, the changes in ETCCDI under RCP8.5 tend to be greater than changes under any of the scenarios in the Special Report on Emission Scenarios (SRES) (Nakicenovic et al. 2000). The spatial distributions of Sillmann et al. (2013b) showed that the northern high latitude regions have the strongest increase in the minimum daily minimum temperature, while changes in the maximum daily maximum temperatures tend to be evenly distributed globally. The study also noted that the percentile indices such as warm and cold nights, exhibit the highest increases in the tropical regions. This is a result of differences in interannual temperature variability between the two regions, which is relatively large in mid- to high latitudes but small in the low-latitudes.

Figure 12 shows the projected ensemble mean increase in two of the ETCCDI indices assessed in figure 8 at the end of this century (2080-2100) from the reference period of 1985-2005 under the RCP8.5 forcing scenario of the 5th Assessment Report of the Intergovernmental Panel on Climate Change (IPCC AR5). Twenty-six models are included in this multi-model average. Models marked with an “*” in figure 8 are not included in this projection due to a lack of future simulations. The two indices, cold nights (Tnn) and hot days (Txx), are chosen to illustrate the seasonal differences of changes at the extremes of both ends of the distribution of daily
surface air temperatures. The upper left panel shows that the wintertime increase in cold nights has large changes northward of the middle of the U.S. largely corresponding to present day winter snow covered areas. Kodra et al. (2011) studied cold extremes in the 21st century using nine GCMs and found that despite future warming, extreme cold events, both in terms of intensity and duration, can persist in the future. Cool nights in the summer, shown in the upper right panel, do not exhibit such a monotonic poleward behavior. Rather, the continental interior warms more than the edges for this measure of extreme temperature. Wintertime increases in warm days, shown in the lower left panel, are projected to be significantly less than wintertime cold nights, although some of the poleward gradient property is present. Summertime increases in hot days are projected slightly warmer than but generally similar to summertime increases in cold nights. One noticeable difference is that future extreme temperatures in the Mexican interior warm more in the summer than in the winter.

The complexity of the seasonality in the projected changes of these two indices is a reflection of the complex mechanisms of change affecting extreme as well as mean temperatures in North America. Changes in snow cover certainly affect the future winter mean and extreme temperatures. Likewise, decreases in soil moisture affect the future summer mean and extreme temperatures. The additional possibility of changes in LSMPs affecting extreme temperatures must be considered in light of these and other mechanisms. For example, van Oldenborgh et al. (2009) argued that warming over western Europe in the last decades was underestimated by climate models involved in the Ensemble SimulationS of Extreme weather events under Nonlinear Climate changE (ESSENCE) project, which covers a 17-member ensemble of climate runs using the ECHAM5/MPI-OM climate model (Jungclaus et al. 2006). Changes in the large-scale circulation, including a shift towards a more westerly circulation and the North Atlantic current need to be better simulated especially in winter and spring for improving the underestimated temperature mean and extremes (van Oldenborgh et al. 2009).
Sillmann and Croci-Maspoli (2009) investigated the role of atmospheric blocking in influencing extreme climate using ECHAM5/MPI-OM data. The model suggested that European blocking events influence particularly the winter cold extremes in Europe. For future projection, decrease in blocking frequency with increase in maximum blocking duration was simulated under the A1B scenario. Also, the blocking pattern shifts northeastward, affecting a larger part of Europe by giving rise to anomalously cold winter months.

To partly address the question of whether changes in heat waves in the future are associated with a shift in the daily maximum temperature or changes in LSMP, Lau and Nath (2012) compared the PDFs of daily maximum surface temperature simulated by a GCM at 200km and 50km over North America. They noted that the PDFs for the current and future climate over most regions that experience heat waves in the U.S. have similar shape but were shifted by the mean warming in the daily maximum temperature, and the magnitude of skewness of the PDFs exhibited only minor changes from the current to future climate. Further statistical manipulations of the model outputs suggest that the increase in heat wave intensity and frequency in the future is primarily associated with a shift in the daily maximum surface temperature, which suggests that changes in the characteristics of LSMPs associated with heat waves may be minor. Future changes in heat waves may also be modulated by changes in land-atmosphere interactions. For example, global warming may shift the climate regime northward and establish new transitional zones with strong land-atmosphere coupling strength (Seneviratne et al. 2006) and drying over the subtropics and mid-latitude during summer in the future may increase land-atmosphere feedbacks in general (Dirmeyer et al. 2012) and result in amplified heat wave response to global warming.

Regional climate models have also been used to investigate temperature extreme changes in the future. Kunkel et al. (2010) analyzed downscaled climate change scenarios for North America at 30km resolution driven by current and future climate simulated by two GCMs.
for 1980 – 1989 and 2090 – 2099. They showed that the regional model has superior skill in simulating heat waves compared to the global models, but it largely inherited the climate sensitivity from the GCMs. They found considerable regional variations in future heat wave characteristics depending on the emission scenarios and GCMs used. Using a suite of regional climate simulations driven by 5 ensemble members of a GCM, Diffenbaugh and Ashfaq (2010) investigated changes in the hot extremes defined by the exceedance of the 95th percentile daily maximum temperature for the present climate and heat wave duration in the U.S. They found that hot extremes are intensified over much of the U.S. associated with a summer anticyclonic circulation anomaly in the future, which reduces precipitation and soil moisture, and amplifies severe hot and dry conditions. The regional simulations showed summer anticyclonic anomaly that is more widespread than the GCM.

Illustrative details appear in studies of trends in California heat waves. Grotjahn (2014) shows trends in extreme values and duration above a threshold for a hot spells LSMP index. The model considered by Grotjahn (2014) had an excessive increasing trend during the historical period (1951-2005) compared with reanalysis data and an increasing trend the last half of this century. Duration above one standard deviation (of the LSMP index) was dramatically increased, and 4-5 day periods above the threshold became the norm for the RCP8.5 simulation. Gershunov and Guirguis (2012) find three out of four GCM’s considered did not adequately capture the synoptic causes of California heat waves. When sub-regional heat waves are defined relative to the changing local climate (a ‘non-stationary’ perspective), coastal heat waves have greater impact than those inland even though the summertime average warming is stronger inland. Gershunov et al. (2013) discuss future hot and cold extremes over the Southwest.

Overall, some robust changes in temperature extremes have emerged from analysis of the CMIP3 and CMIP5 multimodel ensembles as well as targeted global and regional climate
modeling efforts. Model projections generally indicate increased frequency, intensity, and duration of heat waves, and despite global warming, cold extremes may persist in some regions due to changes in blocking events. However, more detailed analyses using multimodel ensembles may provide further insights on the mechanisms of heat wave changes associated with land-atmosphere coupling and potential changes in LSMP affecting both heat waves and cold air outbreaks in the future. Additionally, well-designed climate modeling experiments combined with observational analyses that target specific mechanisms and hypotheses may help isolate the roles of different factors and provide constraints for more robust projections of temperature extreme changes in the future.

6 Questions and gaps in knowledge

Below is a list of open questions and gaps in knowledge that became apparent during the preparation of this article.

Are the observational (and model) data records adequate? How long do observational records need to be? How many realizations of model runs are necessary? Long observational records and reanalyses are needed to understand historical trends and attribute causes. The length needed may be different for different variables or metrics.

Are uncertainties in the observed datasets adequately quantified? As new observing technologies are established how will those changes affect historical trend analysis, especially if new systems replace old systems? What are the consequences of using reanalysis output instead of the actual observational data? What are the effects of gridding station data on extreme temperature?
How shall multivariate extreme statistical tools be further developed? Sometimes neither variable in a combination is extreme, but that particular combination of values is rare and important. Meteorological examples include: hot-humid-no wind versus hot-dry-windy situation. Other examples may need combinations of factors like snow cover and a path that minimizes airmass transformation by avoiding large water bodies (for CAOs) or enhanced downslope motion and suppressed sea breeze (for hot spells).

Are we considering appropriate extreme values needed by various application sectors? For example, longer duration of extreme cold or hot temperatures can have more severe impact than a brief but more extreme value. Also, combinations of variables are important like high heat and high humidity; for example, livestock and humans are both sensitive to apparent temperature.

How effectively can LSMPs be identified for extreme hot or cold events (ETEs) in different locations? Similarities have been found (Loikith and Broccoli 2012) since such events involve large shifts of air masses and hence of associated fields: thickness, geopotential, gradient winds. Is there more than one type of LSMP pattern responsible for each type of ETE?

How sensitive are the results of how ETEs are defined? In this report we argue for multiple approaches to identify LSMPs since no single technique seems uniformly most advantageous.

What are the synoptic-dynamic properties of these LSMPs? What physical mechanisms are responsible for the formation and maintenance of these rare LSMP events? What other factors besides the LSMP, itself, serve to reinforce the ETE? How does the synoptic-dynamics vary seasonally? What is the relative importance of the different mechanisms identified as important (solar heating, adiabatic warming due to subsidence, horizontal temperature advection) and how do these factors relate to the circulation patterns?
Extremes result from a combination of phenomenological factors on synoptic to interannual time scales. How do low frequency phenomena such as blocking, MJO and ENSO contribute to the ETE? What are the roles of local factors such as topography, land surface conditions, and land use changes in the ETE severity? What diagnostic methodologies can delineate among local processes, large scale dynamics, remote forcing, and long term variability? Can a well-defined, broadly-applicable ‘recipe’ of procedures to perform this separation be created?

Can model data be made more useful? For short term extremes, the synoptic-dynamics analyses require ‘instantaneous’ data multiple times per day at many levels for multiple variables. The storage, access to, and visualization of these data remain ongoing challenges. What model limitations are important? For example, parameterization issues such as too deep planetary boundary layer and too strong surface winds influence how well models simulate the strength of land-atmosphere coupling. Other example processes are listed near the end of section 3.1.3.

How well do global and regional climate models simulate the relevant LSMPs? Are these simulated LSMPs similar to those observed (or present in reanalyses) and if not, what dynamical or numerical issues contribute to the differences?

How well must the relevant LSMP be simulated by global climate models to have sufficiently accurate lateral boundary conditions for regional climate models to achieve a specified level of local accuracy? Large scale flow of global climate model important as BC to regional climate model. Example: complex western topography is generally too simple and complex shoreline inadequately resolved in the global models to simulate surface values, but RCM can compensate to some unknown(?) extent. What would the level of accuracy be, or
more specifically, how would a useful level of accuracy be identified? A starting place might be a comparison of statistical distributions of key variables in observations and models.

How to address uncertainty in model projections? Beyond ensemble modeling, coordinated experiments are needed to systematically evaluate models and attribute the sources of model differences; perturbed physics experiments could be used – but how long a record is long enough? How does one adjust the interpretation of model extremes given known model biases?

Does extreme temperature have resolution-dependence? Are model physics tuned to fit the middle of a statistical distribution at some expense to simulating the tails? Are there examples one might postulate for temperature, such as the response of simulated surface vegetation to extreme drought?

7 Summary

This review paper summarizes our current knowledge of, and context for developing new understandings about, extreme hot and cold temperature events affecting regions of North America. The topic of extremes encompasses many scientific issues and a breadth of time and space scales. Understanding extreme events ranges from how events are defined and measured, to how extreme events are studied statistically, theoretically, and with models. To reduce the breadth of scales the paper focuses upon the large scale meteorological patterns (LSMPs) that accompany extreme temperature events of short duration. However, even with this narrower focus, one easily sees that much further progress is needed to understand the properties of extreme temperature events and their supporting LSMP processes. Thus, the primary goal of this report is to provide guidance to researchers interested in studying extreme
temperature events, especially in relation to their associated LSMPs. As such, a variety of techniques are described for identifying the LSMPs, including the relative merits of each approach. A variety of analysis tools are identified which highlight the linkage of LSMPs to both remote low frequency phenomena and local factors. Also highlighted is the information gleaned from multiple climate models including simulation issues and projected trends.

Section 2 presents some methods of defining extreme temperature events statistics using both simple indices and extreme value statistical techniques. A survey of methods used to identify and connect LSMPs to extreme temperatures is also presented. Recent advances in statistical techniques offer an opportunity to connect LSMPs to extreme temperatures through appropriately defined covariates that supplement more straightforward analyses.

Section 3 provides an overview of our current knowledge of LSMPs related to ETEs. Although phenomena ranging from synoptic-scale waves to planetary-scale climate modes are implicated as contributors to ETEs, existing information on (a) the physical nature of these contributions and (b) the dynamical mechanisms responsible for the implicated LSMPs is uneven and incomplete. A diagnostic formalism is put forth for systematically isolating the underlying physics of LSMP life cycles with an ancillary goal of identifying essential large-scale circulation “markers” for climate model validation purposes.

Section 4 summarizes the approaches used to model extreme temperatures, including dynamical methods using global and regional climate models and statistical models. Although climate models generally capture heat waves and cold air outbreaks with some levels of fidelity compared with observations, they overestimate warm wave frequency and underestimate CAO frequency. Furthermore, while CMIP5 models properly represent many of the significant associations between ETEs and low frequency modes, they underestimate the collective influence of low-frequency modes on temperature extremes, particularly related to the Pacific.
Decadal Oscillation. Statistical methods used to relate LSMPs with extremes are limited mainly by small sample sizes of extreme events, so that many predictors could be included to develop more robust statistical relationships.

Section 5 surveys observed and projected trends in extreme temperatures. The studies that form the basis of the IPCC AR5 reports and the 3rd US National Climate Assessment do not consider the role of LSMPs in the magnitude and trends of extreme temperatures. However, a survey of the limited literature exploring these roles in North America and Europe suggests that future assessments, particularly at the regional scale, must include the connection between LSMPs and extreme temperature events.

In the process of preparing this report it quickly became apparent that much work remains to understand better the past and future occurrences of extreme events and their underlying physical causes. A list of open questions was tallied and presented which span the data, statistical, dynamical, and simulation topics. While many gaps exist, investigators now have a large variety of tools and a useful LSMP framework by which to pursue a better understanding of both historical and future extreme temperature events.

Acknowledgments

The authors thank Dr. Christopher J. Paciorek for his assistance in preparing this article. The authors also thank Dr. Steven Vavrus for his comments. Most of the authors are members of the US CLIVAR Extremes working group who greatly appreciate the support provided by the US CLIVAR office. This report was enhanced by discussions held at the 2013 workshop on Analyses, Dynamics, and Modeling of Large Scale Meteorological Patterns Associated with Extreme Temperature and Precipitation Events held at the Lawrence Berkeley National Laboratory and also funded by US CLIVAR. Research by Grotjahn was funded in part by NSF.
grant 1236681. Leung and Wehner were supported by the Regional and Global Climate Modeling Program of the Office of Biological and Environmental Research in the Department of Energy Office of Science. Pacific Northwest National Laboratory is operated by Battelle Memorial Institute for the DOE under contract DE-AC05-76RL01830 and Lawrence Berkeley National Laboratory is under contract DE-AC02-05CH11231. Black was supported by the U.S. Department of Energy, Office of Biological and Environmental Research, award DE-SC0004942 and the National Science Foundation grant ARC-1107384. Gutowski was supported by National Science Foundation grant ARC1023369 and Department of Energy grant DESC0006643. Gyakum was supported by a Natural Sciences and Engineering Research Council of Canada Discovery Grant, and by an International Polar Year Grant. The National Center for Atmospheric Research is sponsored by the National Science Foundation.
References


doi: http://dx.doi.org/10.1175/JAMC-D-12-094.1


72
Gosling SN, McGregor GR, Paldy A (2007) Climate change and heat-related mortality in six
cities Part 1: Model construction and validation. Int. J. Biometeorol. 51: 525-540

Gotaas, Y, Benson CS (1965) The Effect of Suspended Ice Crystals on Radiative Cooling. J.
Appl. Meteor., 4, 446-453

Goto-Maeda Y, Shin D, O'Brien J (2008), Freeze probability of Florida in a regional climate

Grotjahn R (2011) Identifying extreme hottest days from large scale upper air data: a pilot
scheme to find California central valley summertime maximum surface temperatures.

Grotjahn R (2013) Ability of CCSM4 to simulate California extreme heat conditions from
evaluating simulations of the associated large scale upper air pattern. Clim Dyn 41:
1187-1197. DOI 10.1007/s00382-013-1668-1

Grotjahn R (2014) Western North American extreme heat, associated large scale synoptic-
dynamics, and performance by a climate model. In: Dynamics and Predictability of
Large-Scale, High-Impact Weather and Climate Events. Eds: J Li, R Swinbank, H
Volkert, R Grotjahn, Cambridge Univ Press

Grotjahn R, Faure G (2008) Composite predictor maps of extraordinary weather events in the
doi:10.1175/2007WAF2006055.1

eastern U.S. spring freezes: Increased cold damage in a warming world? BioScience
58:253–262


Loikith PC, Broccoli AJ (2014) The Influence of Recurrent Modes of Climate Variability on the
Occurrence of Winter and Summer Extreme Temperatures over North America. J Clim
27: 1600–1618.


doi: http://dx.doi.org/10.1175/2009JCLI3101.1

Wave-Droughts. J Clim, 8: 1658–1675

Maraun, D (2013) Bias correction, quantile mapping, and downscaling: Revisiting the inflation
issue. J Clim, 26: 2137-2143

in CMIP5 Models. J. Clim, 26: 7044–7059

Meehl G, Tebaldi C (2004) More intense, more frequent, and longer lasting heat waves in the


1149

Morak S, Hegerl GC, Kenyon J (2011), Detectable regional changes in the number of warm
Moss RH, Edmonds JA, Hibbard KA, Manning MR, Rose SK, van Vuuren DP, Carter TR, Emori
S, Kainuma M, Kram T, Meehl GA, Mitchell JFB, Nakicenovic N, Riahi K, Smith SJ,
February 2010) | doi:10.1038/nature08823

Nakicenovic N, Alcamo J, Davis G, de Vries HJM, Fenhann J, Gaffin S, Gregory K, Grubler A,
Jung TY, Kram T, La Rovere EL, Michaelis L, Mori S, Morita T, Papper W, Pitcher H,


Orlowsky B, Seneviratne SI (2012) Global changes in extreme events: Regional and seasonal

Outten SD, Esau I (2012) A link between Arctic sea ice and recent cooling trends over Eurasia.

Overland JE, Wang M (2010) Large-scale atmospheric circulation changes associated with the

10.1111/j.1600-0870.2008.00327.x


Rienecker M, and Coauthors (2007) The GEOS-5 data assimilation system—Documentation of
versions 5.0.1 and 5.1.0. NASA GSFC Tech. Rep. Series on Global Modeling and Data


Meteor. Soc., 91: 1015–1057


Sanders F, Gyakum JR (1980) Synoptic-dynamic climatology of the "bomb." Mon Wea Rev,
108: 1589-1606


biases in climate models, J. Clim, 23(23): 6143–6152

extremes in the Northern Hemisphere: The role of stationary Rossby waves. J Clim,
24: 4773–4792

471-69121-6


European mean and extreme climate. Geophys Res Lett , 36: L10702

82


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, Climatic Change, 79: 185-211


86


Fig. 1 Change over 1950-2007 in estimated 20-year annual return values (°C) for a) hot tail of daily maximum temperature (TXx), b) cold tail of daily maximum temperature, (TXn) c) hot tail of daily minimum temperature, (TNx) and d) cold tail of daily minimum temperature (TNn). Results are based on fitting extreme value statistical models with a linear trend in the location parameter to exceedances of a location-specific threshold (greater than the 99th percentile for upper tail and less than the 1th percentile for lower tail). As this analysis was based on anomalies with respect to average values for that time of year, hot minimum temperature values, for example, are just as likely to occur in winter as in summer. The circles indicate the z-score for the estimated change (estimate divided by its standard error), with absolute z-scores exceeding 1, 2, and 3 indicated by open circles of increasing size. Higher z-score indicates greater statistical significance.

Fig. 2 Example large scale meteorological patterns (LSMPs) obtained as target ensemble mean composites for two types of California Central valley extreme events. Cold air outbreaks in winter (DJF) at (a) 72 hours prior and (b) at onset of the events are shown in the 500 hPa geopotential height field. Heat waves during summer (JJAS) (c) 36 hours prior and (d) at the onset are shown in the 700 hPa geopotential height field. Shading indicates significance at the highest or lowest 5% level, with the innermost shading significant at the 1.5% level. Further discussion is in Grotjahn and Faure (2008).

Fig. 3 Self-organizing map of synoptic weather patterns in a region focused on Alaska. The SOM array maps give the departure (in hPa) of sea level pressure (SLP) from the domain
averaged sea level pressure. The SOM used daily December-January-February (DJF) SLP for 1997-2007 from ERA-Interim reanalyses and output from a regional climate model. Locations with elevation exceeding 500 m are not included in the maps to avoid using SLP in regions strongly influenced by methods used to extrapolate SLP from surface pressure.

Fig. 4 Correlation between the local seasonal impact of cold days (left column) and warm days (right column) and the seasonal mean NAO (first row), PNA (second row), PDO (third row) and Niño 3.4 Indices (fourth row) during winter, 1950-2011. The black contours encompass regions having correlations statistical significant at the 95% confidence level (Figure from Westby et al. 2013).

Fig. 5 (a) Daily (0600–0600 UTC) composites of 53 arctic air mass formations’ 1000–500-hPa thickness (dam) (shaded) and sea level pressure (white contours, every 4 hPa) for days 1–3 of the genesis period. Contours of 12-hPa standard deviation of sea level pressure (dashed) and 14-dam standard deviation of 1000–500-hPa thickness (black contours) are also shown. (b) 1000–500 thickness (dam) anomalies (shaded) and sea level pressure anomalies (contours, every 4 hPa) for days 1–3 of the genesis period. Anomalies are calculated relative to the 1948–2008 climatological daily mean. (Reproduced from Turner and Gyakum 2011)

Fig. 6 Grand composites of anomalies associated with temperature regimes over North America: SLP anomalies (hPa) are shaded and \( Z_{500} \) height anomalies are contoured every 20 m; red (blue) contours are positive (negative) \( Z_{500} \) anomalies. Grand composites are shown for (top) January and (bottom) July extreme (from left to right) cold maximum (Tx5), warm
maximum (Tx95), cold minimum (Tn5), and warm minimum (Tn95) temperatures. From Loikith and Broccoli (2012; see text for further details).

Fig. 7 Composite synoptic weather patterns at the onset of the 14 Sacramento California summer (JJAS) heat waves studied by Grotjahn and Faure (2008). a) Temperature at 850 hPa with a 2 K interval. b) 700 hPa level pressure velocity with 2 Pa/s interval and where positive values mean sinking motion. c) Sea level pressure with 2 hPa interval, d) Surface wind vectors with shading applicable to the zonal component. Areas with yellow (lighter inside dark)) shading are positive (above normal) anomalies that are large enough to occur only 1.5% of the time by chance in a same-sized composite; areas that are blue (darker inside light) shading are negative anomalies occurring only 1.5% of the time.

Fig. 8 Performance portrait of the CMIP5 models’ ability to represent the temperature based ETCCDI indices over North American land. The colors represent normalized root mean square errors of seasonal indices compared to the ERA Interim reanalysis. Blue colors represent errors lower than the median error, while red colors represent errors larger than the median error. Seasons are denoted by triangles within each square. Models marked with “**” are not included in the RCP8.5 projections.

Fig. 9 The variance explained in an annual metric of the impact of wintertime Warm Waves on the southeast United States from a multiple linear regression using the NAO and PNA indices as predictors. Results are displayed for individual CMIP5 model simulations (blue bars) while the light and dark gray lines denote variance values for observations and the model mean, respectively (From Westby et al. 2013).
Fig. 10 Warm Spell Duration Index from MERRA for 2011 summer over the United States. WSDI is computed from the 90th percentile of daily mean temperatures for the summer season.

Fig. 11 Trend in WSDI as determined from MERRA for US Summertime temperatures. Dashed and solid black contours indicate statistical significance (at 90% and 95% confidence respectively).

Fig. 12 Projected seasonal changes in North American extreme temperatures from the CMIP5 multi-model at the end of this century under the RCP8.5 forcing scenario. The reference period is 1985-2005 while the future period is 2080-2100. Winter changes are shown on the left while summer changes are shown on the right. The top figures represent changes in cold nights (Tnn) while the lower figures represent changes in hot days (Txx). Units: Kelvins.
Table 1  Some temperature related ETCCDI indices (Sillman et al. 2013a). For a complete list and formal definitions, see http://etccdi.pacificclimate.org/list_27_indices.shtml

<table>
<thead>
<tr>
<th>ETCCDI index name</th>
<th>Semi-formal definition</th>
<th>Plain English</th>
</tr>
</thead>
<tbody>
<tr>
<td>TX90p</td>
<td>The percentage of days when the high temperature is greater than 90% of those in reference period</td>
<td>Hot days</td>
</tr>
<tr>
<td>TX10p</td>
<td>The percentage of days when the high temperature is less than 10% of those in reference period</td>
<td>Cold days</td>
</tr>
<tr>
<td>TN90p</td>
<td>The percentage of days when the low temperature is greater than 90% of those in reference period</td>
<td>Hot nights</td>
</tr>
<tr>
<td>TN10p</td>
<td>The percentage of days when the low temperature is less than 10% of those in reference period</td>
<td>Cold nights</td>
</tr>
<tr>
<td>TXx</td>
<td>monthly or seasonal maximum of daily maximum temperature</td>
<td>Hottest day</td>
</tr>
<tr>
<td>TXn</td>
<td>monthly or seasonal minimum of daily maximum temperature</td>
<td>Coldest day</td>
</tr>
<tr>
<td>TNx</td>
<td>monthly or seasonal maximum of daily minimum temperature</td>
<td>Hottest night</td>
</tr>
<tr>
<td>TNn</td>
<td>monthly or seasonal minimum of daily minimum temperature</td>
<td>Coldest night</td>
</tr>
<tr>
<td>HWDI</td>
<td>Heat Wave Duration Index</td>
<td>Length of a heat wave</td>
</tr>
<tr>
<td>CWDI</td>
<td>Cold Wave Duration Index</td>
<td>Length of a cold spell</td>
</tr>
<tr>
<td>FD</td>
<td>Days below freezing</td>
<td>Frost days</td>
</tr>
<tr>
<td>Method (example reference)</td>
<td>Approach</td>
<td>Attributes</td>
</tr>
<tr>
<td>---------------------------</td>
<td>----------</td>
<td>------------</td>
</tr>
<tr>
<td>Composites (e.g. Grotjahn and Faure 2008)</td>
<td>Average together an ‘ensemble’ of maps on pre-identified dates As with some other procedures described here, it is important to assess how independent the dates are, e.g. consulting the autocorrelation</td>
<td>Ensemble average provides physical insights. Can be used to examine patterns leading up to and subsequent to the event onset by shifting dates used. Non-parametric in making no assumption about the pattern. The dates can be chosen to be independent.</td>
</tr>
<tr>
<td>Regression (e.g. Lau and Nath 2012)</td>
<td>Fit a regression line (polynomial in the predictand) to a time series of a predictor at each grid point.</td>
<td>Pattern provides physical insights. Parametric in assuming a specific regression line (polynomial). Can be used to examine patterns leading up to dates chosen event if event lasts for one time sample.</td>
</tr>
<tr>
<td><strong>EOFs or PCs</strong> (Wu et al. 2012)</td>
<td><strong>Calculate the eigenvectors of a space-weighted covariance (between values at different grid points) matrix. If only inputting extreme dates, then it is important to assess how independent the dates are, e.g. consulting the autocorrelation</strong></td>
<td><strong>Finds multiple patterns, each of which is orthogonal (not a subset) of another pattern. Can identify fraction of variability that is due to that EOF/PC. Most often used with filtered data to find leading low frequency structures.</strong></td>
</tr>
<tr>
<td>SOMs (e.g. Hewitson and Crane 2002)</td>
<td>Obtain the distribution of patterns covered by a set of input maps, using neural-net training</td>
<td>Resultant maps span the pattern space of input fields, represent nodes of a continuous pattern distribution, and can be related to a transformation of rotated EOFs back to physical space.</td>
</tr>
</tbody>
</table>
| Clustering analysis (e.g. Stephanon et al. 2012) | Assigning events to K clusters so that it minimizes total sum of the generalized Euclidian distance between patterns in a cluster (Spath 1985; Seber 2008). If only inputting extreme dates, then it is | Objectively classify events and find relating LSMPs that provide physical insight. No assumptions about the patterns such as orthogonality and symmetry. | The number of LSMPs can be somewhat arbitrary since number of cluster is pre-specified, although larger separation of clusters is acquired a little from the ‘dissimilarity index’. The cluster assignment can be vague for | Significance of classification stability can be obtained from Monte Carlo test by rejecting the null hypothesis that the verification period temperature extreme cannot be classified in the clusters obtained from remaining periods. In theory,
| **Machine Learning** (Salakhutdinov and Hinton 2012) | Train a multi-layer neural network on a dataset. Training is done one layer at a time, with spatio-temporal patches at the bottom layer, and class labels being assigned at the top layer. | DBNs have proved to be powerful in capturing a range of patterns. It is likely that they will extract invariant patterns as well as anomalies. | These techniques have produced state-of-the-art results in computer vision and speech recognition tasks. They have not yet been applied to climate datasets. | Has not been conducted in the context of climate/LSMP applications. Classification performance of method has been conducted extensively for images by testing on held-out data. |
Table 3 Sampling of Various Criteria Used in Heat Wave and Hottest Day Definitions

<table>
<thead>
<tr>
<th>Source</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robinson (2001)</td>
<td>A period of at least 48 hours during which neither the overnight low nor the daytime heat index $Hi$ falls below the NWS heat stress thresholds (80°F and 105°F). At stations where more than 1% of both the high and low $Hi$ observations exceed these thresholds, the 1% values are used as the heat wave thresholds</td>
</tr>
<tr>
<td>Hajat et al. (2002)</td>
<td>3-day moving average temperatures $&gt; \text{the } 99^{\text{th}} \text{ percentile of the whole record of temperature}$</td>
</tr>
</tbody>
</table>
| Meehl and Tebaldi (2004)    | The longest period of consecutive days satisfying the following 3 conditions:  
2. Average daily maximum temperature $> T_1$ for entire period  
3. Daily maximum temperature $> T_2$ for every day of entire period,  
where  
$T_1$ (threshold 1) = 97.5$^{\text{th}}$ percentile of distribution of maximum temperatures in the observations and in simulated present day climate  
$T_2$ = 81$^{\text{st}}$ percentile |
| Beniston (2004)             | Maximum $T$ exceeding the 90$^{\text{th}}$ quantile of summer temperature (30C) at a station (Basil, Switzerland).                                                                                       |
| Lipton et al. (2005)        | Daily maximum high temperature remains 2 standard deviations above normal for at least 2 consecutive days                                                                                                     |
| Gosling et al. (2007)       | For 3 or more days the maximum $T$ must be $\geq 95^{\text{th}}$ percentile of the maximum $T$ in the summer climatology                                                                                     |
| Grotjahn and Faure (2008)   | At least 3 consecutive days during which the daily maximum temperatures are above 100°F (38°C), and with at least one above 105°F (40.5°C)                                                               |
| Bachmann (2008)             | Two combinations of criteria, I and II, were tested:  
I. Must satisfy both conditions (where anomaly is relative to long term daily temperature)                                                  |
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>mean):</strong></td>
<td>1). At least 3 consecutive days with daily anomaly maximum temperature ≥ 10°C</td>
</tr>
<tr>
<td></td>
<td>2). At least 1 day must have maximum temperature anomaly ≥ 15°C</td>
</tr>
<tr>
<td><strong>II. Must satisfy the 2 conditions above plus this additional condition:</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.) The average maximum temperature for the event ≥ 100 F (38°C)</td>
</tr>
<tr>
<td><strong>Gershunov et al. (2009)</strong></td>
<td>Individual stations exceeding the 99th percentile for 1, 2, or 3 dates in a row are aggregated, with the highest aggregation of values over the region including all of California and Nevada determining a ranking for an event. Daytime maximum and nighttime (highest) minimum treated separately.</td>
</tr>
<tr>
<td><strong>Lyon (2009)</strong></td>
<td>Daily maximum temperature must exceed the 90th percentile for at least 3 consecutive days, where the percentile is based daily values from the 3-month summer season. Also tested, same temperature criterion over 5 consecutive days.</td>
</tr>
<tr>
<td><strong>Grotjahn (2011)</strong></td>
<td>Daily maximum temperature anomaly (relative to long term daily mean) normalized by daily long term mean standard deviation at all three CV stations (KRBL, KFAT, KBFL) must all exceed 1.6 Note: this defines hottest days, not heat waves.</td>
</tr>
<tr>
<td><strong>Bumbaco et al. (2013)</strong></td>
<td>Daily maximum temperature anomalies for stations in a region are averaged together. Heat wave when regional average daily anomaly exceeds 99th percentile for 3 or more consecutive days</td>
</tr>
</tbody>
</table>
Table 4 Correlation coefficients between the loading pattern of observation and that of each CMIP5 model for NAO-like and PNA-like modes. Right-most column includes the averaged coefficients of two modes. Dark (light) grey shadings denote high-top (low-top) models (Reproduced from Lee and Black 2013).

<table>
<thead>
<tr>
<th>Model</th>
<th>NAO</th>
<th>PNA</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;GFDL-ESM2G&quot;</td>
<td>0.94</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>&quot;MPI-ESM-LR&quot;</td>
<td>0.83</td>
<td>0.92</td>
<td>0.88</td>
</tr>
<tr>
<td>&quot;HadCM3&quot;</td>
<td>0.90</td>
<td>0.82</td>
<td>0.86</td>
</tr>
<tr>
<td>&quot;CSIRO-Mk3-6-0&quot;</td>
<td>0.85</td>
<td>0.81</td>
<td>0.83</td>
</tr>
<tr>
<td>&quot;CCSM4&quot;</td>
<td>0.71</td>
<td>0.91</td>
<td>0.81</td>
</tr>
<tr>
<td>&quot;CanESM2&quot;</td>
<td>0.73</td>
<td>0.83</td>
<td>0.78</td>
</tr>
<tr>
<td>&quot;CNRM-CM5&quot;</td>
<td>0.78</td>
<td>0.77</td>
<td>0.78</td>
</tr>
<tr>
<td>&quot;MIROC-ESM-CHEM&quot;</td>
<td>0.73</td>
<td>0.78</td>
<td>0.75</td>
</tr>
<tr>
<td>&quot;inmcm4&quot;</td>
<td>0.72</td>
<td>0.76</td>
<td>0.74</td>
</tr>
<tr>
<td>&quot;NorESM1-M&quot;</td>
<td>0.56</td>
<td>0.91</td>
<td>0.73</td>
</tr>
<tr>
<td>&quot;IPSL-CM5A-MR&quot;</td>
<td>0.61</td>
<td>0.84</td>
<td>0.73</td>
</tr>
<tr>
<td>&quot;MIROC5&quot;</td>
<td>0.68</td>
<td>0.77</td>
<td>0.73</td>
</tr>
<tr>
<td>&quot;HadGEM2-CC&quot;</td>
<td>0.74</td>
<td>0.70</td>
<td>0.72</td>
</tr>
<tr>
<td>&quot;MIROC-ESM&quot;</td>
<td>0.72</td>
<td>0.70</td>
<td>0.71</td>
</tr>
<tr>
<td>&quot;IPSL-CM5A-LR&quot;</td>
<td>0.48</td>
<td>0.88</td>
<td>0.68</td>
</tr>
<tr>
<td>&quot;MRI-CGCM3&quot;</td>
<td>0.49</td>
<td>0.87</td>
<td>0.68</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>0.72</td>
<td>0.83</td>
<td>0.77</td>
</tr>
</tbody>
</table>
Fig. 1 Change over 1950-2007 in estimated 20-year annual return values (°C) for a) hot tail of daily maximum temperature (TXx), b) cold tail of daily maximum temperature, (TXn) c) hot tail of daily minimum temperature, (TNx) and d) cold tail of daily minimum temperature (TNn). Results are based on fitting extreme value statistical models with a linear trend in the location parameter to exceedances of a location-specific threshold (greater than the 99th percentile for upper tail and less than the 1th percentile for lower tail). As this analysis was based on anomalies with respect to average values for that time of year, hot minimum temperature values, for example, are just as likely to occur in winter as in summer. The circles indicate the z-score for the estimated change (estimate divided by its standard error), with absolute z-scores exceeding 1, 2, and 3 indicated by open circles of increasing size. Higher z-score indicates greater statistical significance.
Fig. 2 Example large scale meteorological patterns (LSMPs) obtained as target ensemble mean composites for two types of California Central valley extreme events. Cold air outbreaks in winter (DJF) at (a) 72 hours prior and (b) at onset of the events are shown in the 500 hPa geopotential height field. Heat waves during summer (JJAS) (c) 36 hours prior and (d) at the onset are shown in the 700 hPa geopotential height field. Shading indicates significance at the highest or lowest 5% level, with the innermost shading significant at the 1.5% level. Further discussion is in Grotjahn and Faure (2008).
Fig. 3 Self-organizing map of synoptic weather patterns in a region focused on Alaska. The SOM array maps give the departure (in hPa) of sea level pressure (SLP) from the domain averaged sea level pressure. The SOM used daily December-January-February (DJF) SLP for 1997-2007 from ERA-Interim reanalyses and output from a regional climate model. Locations with elevation exceeding 500 m are not included in the maps to avoid using SLP in regions strongly influenced by methods used to extrapolate SLP from surface pressure.
**Fig. 4** Correlation between the local seasonal impact of cold days (left column) and warm days (right column) and the seasonal mean NAO (first row), PNA (second row), PDO (third row) and Niño 3.4 Indices (fourth row) during winter, 1950-2011. The black contours encompass regions having correlations statistical significant at the 95% confidence level (Figure from Westby et al. 2013).
Fig. 5 (a) Daily (0600–0600 UTC) composites of 53 arctic air mass formations’ 1000–500-hPa thickness (dam) (shaded) and sea level pressure (white contours, every 4 hPa) for days 1–3 of the genesis period. Contours of 12-hPa standard deviation of sea level pressure (dashed) and 14-dam standard deviation of 1000–500-hPa thickness (black contours) are also shown. (b) 1000–500 thickness (dam) anomalies (shaded) and sea level pressure anomalies (contours, every 4 hPa) for days 1–3 of the genesis period. Anomalies are calculated relative to the 1948–2008 climatological daily mean. (Reproduced from Turner and Gyakum 2011)
Fig. 6 Grand composites of anomalies associated with temperature regimes over North America: SLP anomalies (hPa) are shaded and $Z_{500}$ height anomalies are contoured every 20 m; red (blue) contours are positive (negative) $Z_{500}$ anomalies. Grand composites are shown for (top) January and (bottom) July extreme (from left to right) cold maximum (Tx5), warm maximum (Tx95), cold minimum (Tn5), and warm minimum (Tn95) temperatures. From Loikith and Broccoli (2012; see text for further details).
Fig. 7 Composite synoptic weather patterns at the onset of the 14 Sacramento California summer (JJAS) heat waves studied by Grotjahn and Faure (2008). a) Temperature at 850 hPa with a 2 K interval. b) 700 hPa level pressure velocity with 2 Pa/s interval and where positive values mean sinking motion. c) Sea level pressure with 2 hPa interval, d) Surface wind vectors with shading applicable to the zonal component. Areas with yellow (lighter inside dark) shading are positive (above normal) anomalies that are large enough to occur only 1.5% of the time by chance in a same-sized composite; areas that are blue (darker inside light) shading are negative anomalies occurring only 1.5% of the time. Figure reproduced from Grotjahn (2011; DOI 10.1007/s00382-011-0999-z)
Fig. 8 Performance portrait of the CMIP5 models’ ability to represent the temperature based ETCCDI indices over North American land. The colors represent normalized root mean square errors of seasonal indices compared to the ERA Interim reanalysis. Blue colors represent errors lower than the median error, while red colors represent errors larger than the median error. Seasons are denoted by triangles within each square. Models marked with “**” are not included in the RCP8.5 projections.
Fig. 9 The variance explained in an annual metric of the impact of wintertime Warm Waves on the southeast United States from a multiple linear regression using the NAO and PNA indices as predictors. Results are displayed for individual CMIP5 model simulations (blue bars) while the light and dark gray lines denote variance values for observations and the model mean, respectively (From Westby et al. 2013).
Fig. 10 Warm Spell Duration Index from MERRA for 2011 summer over the United States. WSDI is computed from the 90th percentile of daily mean temperatures for the summer season.
Fig. 11 Trend in WSDI as determined from MERRA for US Summertime temperatures. Dashed and solid black contours indicate statistical significance (at 90% and 95% confidence respectively).
Fig. 12 Projected seasonal changes in North American extreme temperatures from the CMIP5 multi-model at the end of this century under the RCP8.5 forcing scenario. The reference period is 1985-2005 while the future period is 2080-2100. Winter changes are shown on the left while summer changes are shown on the right. The top figures represent changes in cold nights (Tnn) while the lower figures represent changes in hot days (Txx). Units: Kelvins.