Anomalous Temperature Regimes during the Cool Season: Long-Term Trends, Low-Frequency Mode Modulation, and Representation in CMIP5 Simulations

REBECCA M. WESTBY, YUN-YOUNG LEE, AND ROBERT X. BLACK

Department of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, Georgia

(Manuscript received 13 December 2012, in final form 11 June 2013)

ABSTRACT

During boreal winter, anomalous temperature regimes (ATRs), including cold air outbreaks (CAOs) and warm waves (WWs), provide important societal influences upon the United States. The current study analyzes reanalysis and model data for the period from 1949 to 2011 to assess (i) long-term variability in ATRs, (ii) interannual modulation of ATRs by low-frequency modes, and (iii) the representation of ATR behavior in models from phase 5 of the Coupled Model Intercomparison Project (CMIP5).

No significant trends in either WWs or CAOs are identified for the continental United States. On interannual time scales, CAOs are modulated by the (i) North Atlantic Oscillation (NAO) over the U.S. Southeast and (ii) the Pacific–North American (PNA) pattern over the Northwest. WW frequency is modulated by (i) the NAO over the eastern United States and (ii) the combined influence of the PNA, Pacific decadal oscillation (PDO), and ENSO over the southern United States. In contrast to previous studies of seasonal-mean temperature, the influence of ENSO upon ATRs is found to be mainly limited to a modest modulation of WWs over the southern United States. Multiple linear regression analysis reveals that the regional collective influence of low-frequency modes accounts for as much as 50% of interannual ATR variability. Although similar behavior is observed in CMIP5 models, WW (CAO) frequency is typically overestimated (underestimated). All models considered are unable to replicate observed associations between ATRs and the PDO. Further, the collective influence of low-frequency modes upon ATRs is generally underestimated in CMIP5 models. The results indicate that predictions of future ATR behavior are limited by climate model ability to represent the evolving behavior of low-frequency modes of variability.

1. Introduction

During the boreal cool season, anomalous temperature regimes (ATRs), including cold air outbreaks (CAOs) and warm waves (WWs), affect regional economies and human safety over large portions of the United States via their significant impacts on energy consumption, local agriculture, and human health. The extremely cold temperatures associated with CAOs put humans at a higher risk of frostbite and hypothermia, resulting in approximately 30 deaths per year, while economic losses in the agriculture and transportation industries can reach billions of dollars owing to crop damage, livestock death, and transportation delays (Cellitti et al. 2006; Rogers and Rohli 1991). These impacts can be experienced in any region. However, they are especially pronounced along the East and Gulf Coasts, where the majority of the population is less accustomed to extreme cold temperatures (Cellitti et al. 2006). Historically, cool season WWs have received much less attention than CAOs, but they can have equally significant impacts. Abnormally warm temperatures associated with WWs can cause rapidly melting snow and ice, leading to the ice damming and flooding of rivers, as well as significant monetary losses in the winter sports industry. WWs also pose an increased threat of frost damage to agriculture by inducing premature plant development (Gu et al. 2008).

Previous research on CAO variability found no evidence of a trend toward fewer extreme cold events in the United States (Walsh et al. 2001; Portis et al. 2006). However, these studies do not include information encompassing the last decade. Therefore, extending these studies to examine recent trends in CAO variability is one of the goals of this study. Furthermore, this study also includes a new additional focus on the trends in WWs during the boreal cool season, which have rarely been considered in previous studies. Interestingly, a recent
study by Hankes and Walsh (2011) found warming trends within high-latitude CAO formation regions. However, considered alongside a lack of change in CAO frequency, this local warming trend presents an intriguing paradox. Two recent winters provide prime examples of this paradox, as both winters were exceptionally cold within particular regions (Cohen et al. 2010; Guirguis et al. 2011; Wang et al. 2010) despite a background consisting of anomalously warm hemispheric-average winter temperatures (Cohen et al. 2010). This suggests, then, that there must be some mechanism or process in the atmosphere with a stronger impact on the interannual variability and trends in CAOs other than a simple consideration of the mean background temperature. One such mechanism could be related to interannual variations in the large-scale circulation pattern.

Prior research has shown that ATRs—namely, CAOs—are enhanced by the large-scale low-frequency modes of variability, such as the positive phase of the Pacific–North American (PNA) pattern (Downton and Miller 1993; Vavrus et al. 2006; Cellitti et al. 2006; Rogers and Rohli 1991) and the negative phase of the North Atlantic Oscillation (NAO) (Walsh et al. 2001; Cellitti et al. 2006). In addition, a recent study by Lim and Schubert (2011) found that daily temperature extremes are strongly impacted by the Arctic Oscillation (AO), and this impact is comparable to or stronger than that of the El Niño–Southern Oscillation (ENSO). Unlike CAOs, little is currently known about the nature and low-frequency mode modulation of WWs during the boreal cool season, and this topic will therefore be another focus of this study. In addition, this study will quantify the respective roles of these modes of low-frequency variability in accounting for seasonal ATR behavior, which has also not been fully explored in previous studies.

A final goal of this study is to examine the capability of phase 5 of the Coupled Model Intercomparison Project (CMIP5) (Taylor et al. 2012) in representing observed regional ATR behavior, including long-term trends and low-frequency modulation. A new suite of model simulations is a useful test bed for analyzing the behavior of coupled atmosphere–ocean general circulation models (AOGCMs) in association with the CMIP5 models. Because human activities and the environment are greatly vulnerable to and affected by climate and weather extremes (Kharin et al. 2007), there is a strong interest for scientists and society alike to be able to forecast these events. However, the utility of AOGCMs in representing future changes in regional extreme weather, such as ATRs, is critically dependent upon their ability to replicate past observations. As a result, an assessment of simulated historical ATRs and their modulation by low-frequency modes is essential to estimate the likely success of predicting future ATR behavior.

2. Data and methods

The dataset used for the identification of observed ATRs is the reanalysis data from the National Centers for Environmental Prediction–National Center for Atmospheric Research (hereafter NNR) (Kalnay et al. 1996). The NNR data is the longest reanalysis dataset currently available, and is therefore ideal for studying the long-term trends and variability of anomalous temperature regimes. The subset of data used for this analysis spans from January 1948 to February 2011. The variable used for ATR identification is the daily mean surface air temperature $T$ at the $\sigma = 0.995$ level, the closest available level to the surface. These data have a grid resolution of 2.5° latitude $\times$ 2.5° longitude. Statistical stationarity in the dataset is assessed in terms of long-term trends in the annual winter-mean [December–February (DJF)] temperature. Determining whether our dataset is statistically stationary is important because of how our events are defined (described in detail in a later paragraph). Our events are defined based on temperature anomalies since we are most interested in examining events that are deviations from normal, and this method implicitly assumes that the dataset is stationary. To determine the degree of statistical stationarity, further testing is required. Since statistically significant trends in the long-term annual winter-mean temperatures are identified, the dataset is therefore not statistically stationary, and the temperature records are first detrended by explicitly removing the long-term trend in annual winter-mean temperature for each grid point considered. The dataset detrending ensures that any observed trends in ATRs are not artificial due to changes in the background mean temperature. The detrended dataset is used in all subsequent calculations. The details of the aforementioned trends in the annual winter-mean temperatures will be discussed in the next section.

To assess the capability of CMIP5 models in representing regional ATR characteristics, including trends and low-frequency mode modulation, 16 historical AOGCM simulations from the CMIP5 data archive are used (listed in Table 1). Among the 16 model configurations, 7 are high-top models, which have their highest vertical level above the stratopause, while 8 others are low-top models and have their highest vertical level below the stratopause (Charlton-Perez et al. 2013). The single remaining model falls in between these two categories and is referred to as an intermediate model. In principle, in model analyses it is desirable to examine ensemble average results since model realizations can vary even with
Table 1. Linear trends in the yearly impact factor of cold days below $-1\sigma$ and warm days above $+1\sigma$ in the Southeast (SE), Northeast (NE), and Midwest (MW) regions, for DJF 1951–2005, unless otherwise specified. Boldface (italic) font represents significance at the 95% (90%) confidence level.

<table>
<thead>
<tr>
<th>Model/observation</th>
<th>Abbreviation</th>
<th>CAOs</th>
<th>WWs</th>
<th>CAOs</th>
<th>WWs</th>
<th>CAOs</th>
<th>WWs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long time period observations</td>
<td>NNR(1949–2011)</td>
<td>$-0.052$</td>
<td>$-0.040$</td>
<td>$-0.074$</td>
<td>$-0.070$</td>
<td>$-0.015$</td>
<td>$0.012$</td>
</tr>
<tr>
<td>Short time period observations</td>
<td>NNR(1951–2005)</td>
<td>$-0.144$</td>
<td>$0.078$</td>
<td>$-0.078$</td>
<td>$0.012$</td>
<td>$-0.029$</td>
<td>$0.065$</td>
</tr>
<tr>
<td>Hadley Centre Global Environmental Model, version 2 - Carbon Cycle</td>
<td>HadGEM2-CC</td>
<td>$0.034$</td>
<td>$-0.091$</td>
<td>$0.056$</td>
<td>$0.008$</td>
<td>$0.004$</td>
<td>$0.103$</td>
</tr>
<tr>
<td>L’Institut Pierre-Simon Laplace (IPSL) Coupled Model, version 5A, coupled with Nucleus for European Modelling of the Ocean (NEMO), low resolution</td>
<td>IPSL-CM5A-LR</td>
<td>$-0.226$</td>
<td>$0.168$</td>
<td>$-0.119$</td>
<td>$0.173$</td>
<td>$-0.171$</td>
<td>$0.216$</td>
</tr>
<tr>
<td>IPSL CM5A, coupled with NEMO, mid resolution</td>
<td>IPSL-CM5A-MR</td>
<td>$-0.092$</td>
<td>$0.148$</td>
<td>$-0.150$</td>
<td>$0.182$</td>
<td>$-0.133$</td>
<td>$0.156$</td>
</tr>
<tr>
<td>Model for Interdisciplinary Research on Climate, Earth System Model, Chemistry Coupled</td>
<td>MIROC-ESM-CHEM</td>
<td>$0.062$</td>
<td>$0.097$</td>
<td>$-0.122$</td>
<td>$0.172$</td>
<td>$0.094$</td>
<td>$0.039$</td>
</tr>
<tr>
<td>Model for Interdisciplinary Research on Climate, Earth System Model</td>
<td>MIROC-ESM</td>
<td>$-0.102$</td>
<td>$0.085$</td>
<td>$-0.272$</td>
<td>$0.212$</td>
<td>$-0.198$</td>
<td>$0.129$</td>
</tr>
<tr>
<td>Max Planck Institute Earth System Model, low resolution</td>
<td>MPI-ESM-LR</td>
<td>$-0.144$</td>
<td>$0.085$</td>
<td>$-0.312$</td>
<td>$0.089$</td>
<td>$-0.255$</td>
<td>$0.093$</td>
</tr>
<tr>
<td>Meteorological Research Institute Coupled Atmosphere–Ocean General Circulation Model, version 3</td>
<td>MRI-CGCM3</td>
<td>$-0.099$</td>
<td>$-0.074$</td>
<td>$-0.146$</td>
<td>$0.042$</td>
<td>$-0.177$</td>
<td>$0.094$</td>
</tr>
<tr>
<td>Community Climate System Model, version 4</td>
<td>CCSM4</td>
<td>$-0.100$</td>
<td>$0.059$</td>
<td>$-0.208$</td>
<td>$0.147$</td>
<td>$-0.239$</td>
<td>$0.243$</td>
</tr>
<tr>
<td>Geophysical Fluid Dynamics Laboratory Earth System Model with Generalized Ocean Layer Dynamics (GOLD) component (ESM2G)</td>
<td>GFDL-ESM2G</td>
<td>$-0.146$</td>
<td>$0.270$</td>
<td>$-0.074$</td>
<td>$0.092$</td>
<td>$-0.049$</td>
<td>$-0.039$</td>
</tr>
<tr>
<td>Centre National de Recherches Météorologiques Coupled Global Climate Model, version 5</td>
<td>CNRM-CM5</td>
<td>$-0.054$</td>
<td>$0.226$</td>
<td>$-0.076$</td>
<td>$0.129$</td>
<td>$0.029$</td>
<td>$0.133$</td>
</tr>
<tr>
<td>Commonwealth Scientific and Industrial Research Organisation, version 3.6.0</td>
<td>CSIRO Mk3.6.0</td>
<td>$-0.103$</td>
<td>$0.082$</td>
<td>$-0.064$</td>
<td>$0.064$</td>
<td>$-0.020$</td>
<td>$0.034$</td>
</tr>
<tr>
<td>Institute of Numerical Mathematics Coupled Model, version 4.0</td>
<td>INM-CM4.0</td>
<td>$0.024$</td>
<td>$-0.038$</td>
<td>$-0.030$</td>
<td>$0.013$</td>
<td>$-0.017$</td>
<td>$0.008$</td>
</tr>
<tr>
<td>Model for Interdisciplinary Research on Climate, version 5</td>
<td>MIROC5</td>
<td>$-0.094$</td>
<td>$0.105$</td>
<td>$-0.191$</td>
<td>$0.036$</td>
<td>$-0.060$</td>
<td>$-0.044$</td>
</tr>
<tr>
<td>Norwegian Earth System Model, version 1 (intermediate resolution)</td>
<td>NorESM1-M</td>
<td>$0.071$</td>
<td>$0.000$</td>
<td>$0.065$</td>
<td>$-0.053$</td>
<td>$0.043$</td>
<td>$-0.041$</td>
</tr>
<tr>
<td>Hadley Centre Coupled Model, version 3</td>
<td>HadCM3</td>
<td>$-0.039$</td>
<td>$0.085$</td>
<td>$-0.068$</td>
<td>$0.111$</td>
<td>$0.029$</td>
<td>$0.177$</td>
</tr>
<tr>
<td>Second Generation Canadian Earth System Model</td>
<td>CanESM2</td>
<td>$-0.058$</td>
<td>$0.162$</td>
<td>$-0.150$</td>
<td>$0.179$</td>
<td>$-0.130$</td>
<td>$0.094$</td>
</tr>
<tr>
<td>Model average</td>
<td></td>
<td>$-0.067$</td>
<td>$0.086$</td>
<td>$-0.116$</td>
<td>$0.100$</td>
<td>$-0.078$</td>
<td>$0.087$</td>
</tr>
</tbody>
</table>

The low-frequency modes of interest are the NAO, PNA, the Pacific decadal oscillation (PDO), and ENSO (defined in terms of the Niño-3.4 sea surface temperatures), which have inherent time scales of a month or longer and are known or suspected to be associated with ATRs. The winter (DJF) mean low-frequency mode indices used in our investigation are from the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) (http://www.esrl.noaa.gov/psd/data/climateindices/). An active debate in the literature remains as to whether the AO or NAO is more appropriate for such analyses. However, the surface circulations for both modes are nearly identical over the region of interest, and our analysis of both modes produced quantitatively similar results. Thus, for brevity we choose to represent only the results for the NAO.
the CMIP5 model simulations, principal component (PC) analysis (Wilks 2011) is used to isolate simulated patterns of low-frequency variability and their associated time series. Typically, the NAO and PNA patterns are taken as the two leading modes of a rotated PC (RPC) analysis using the monthly mean 500-hPa geopotential height anomalies over the extratropical Northern Hemisphere (Barnston and Livezey 1987). In this study, the Kaiser varimax rotation is used in the RPC analysis. Kaiser varimax rotation is the method of applying varimax rotation with the Kaiser criterion for the unrotated EOFs. The Kaiser criterion assesses the appropriate number of factors that should be rotated. Specifically, all components with eigenvalues under 1.0 are removed from consideration by the Kaiser rule. The varimax method, the most common rotation option, maximizes the variance of the squared loadings of a factor (column) on all of the variables (rows) in a factor matrix (Child 2006; Fabrigrar et al. 1999; Thompson 2004). Accordingly, for each model the NAO-like (PNA like) mode is defined as the mode having the highest pattern correlation with respect to the observed NAO (PNA) mode. The PDO-like mode is defined to be the first loading vector of monthly mean sea surface temperature (SST) anomalies over the Pacific Ocean domain of 20°–60°N, 110°E–110°W (Mantua and Hare 2002; Mantua et al. 1997). Finally, the ENSO-like mode in the model simulations is assessed in the same manner as the observed Niño-3.4 index (CPC).

Clustering is a widely used partitioning procedure that identifies separate groups of objects having common structural elements. In atmospheric science, this method has been used to classify unique circulation types. Park et al. (2011) applied a hierarchical clustering analysis to further classify East Asia cold surges into two groups, the blocking type and wave type, while Fereday et al. (2008) investigated the existence of distinct weather regimes in North Atlantic–European circulations via the application of a k-means clustering analysis. Thus, clustering represents a useful analysis method for identifying distinct sets of patterns that fully characterize different circulation types (Christiansen 2007; Fereday et al. 2008).

To examine how the PDO-like modes in the models correspond to the observed PDO structure, a k-means clustering technique is applied to 14 of the 16 model loading vectors and the observations. The MIROC-ESM and MIROC-ESM-CHEM are omitted in the clustering analysis because SST is not available for these models. The k-means clustering technique is an iterative algorithm in which points are moved from one group to another until there is no additional improvement in minimizing the squared Euclidean point-to-centroid distance (Seber 2008; Spath 1985), where each centroid is the mean of the points in its cluster. The value of k indicates the number of clusters, which are specified here a priori to be four. The model structures that fall into the same cluster as the observations (NNR) are considered to provide a good representation of the observed PDO pattern.

Many previous cold air outbreak studies apply a local temperature criterion to identify events, and for consistency, a similar approach is applied in this study. Anomalous temperature regimes are identified first for the continental United States at each grid point in terms of the daily mean surface air temperature anomalies $T'$. For each grid point, the daily mean $T$ climatology is first calculated and then the annual cycle of the daily climatology is smoothed (to remove residual high-frequency variability) by retaining the first six Fourier harmonics of the annual cycle. An analogous procedure is performed to obtain a smoothed annual climatological cycle of daily mean standard deviation $\sigma$. Daily temperature anomalies are then obtained by subtracting the smoothed daily mean $T$ climatology value from the daily mean $T$ at each grid point. Finally, the anomalies are normalized by the smoothed daily mean standard deviation.

To characterize the behavior of ATRs, an impact factor metric is devised that quantifies the cumulative effect of all warm or cold days per winter season: it is calculated by summing the absolute values of the normalized $T'$ values for all days during the winter season that exceed a threshold of $+n$ (a WW event) or $-n$ (a CAO event). For our study, $n$ is chosen to be 1. Thus, the impact factor represents an integrative measure of the seasonal impact of ATR events and combines information regarding both event frequency and intensity. The winter seasons are December–February and are labeled by the year in which January and February fall.

The long-term trends in ATRs are studied by correlating the time series of each ATR metric with time from 1949 to 2011. All trends are considered using a $p$ value for occurrence by chance, which is calculated using a two-tailed Student’s $t$ test. Trends are considered significant when $p$ is less than 0.05. The effect of autocorrelation in the time series is accounted for by computing the effective degrees of freedom using the procedure of Wilks (2011).

The modulation of ATRs by low-frequency modes is explored via correlation analysis between the seasonal-mean ATR metrics and the indices of the low-frequency modes of interest. This analysis is performed for 1950–2011, due to the restricted availability of the observed low-frequency mode indices. The statistical significance of these correlations is tested using the method described above. A multiple linear regression analysis is then performed using the indices of the significant low-frequency modes as predictor variables to quantify the
simultaneous influences of multiple low-frequency modes on ATR variability. In performing these regressions, there are two essential characteristics that need to be considered: (i) potential multicollinearities among predictors and (ii) autocorrelation in the residual term. Multicollinearities are situations in which two predictor variables in a multiple linear regression analysis are highly correlated with one another. Testing for multicollinearities is important because it is not always known a priori whether one predictor contains redundant information in relation to another predictor. Multicollinearities among predictors are detected using the Belsley–Kuh–Welsch (BKW) variance decomposition method (Belsley et al. 1980) and are eliminated by retaining only one of each set of implicated predictors in the multiple linear regression. Testing for autocorrelation in the residuals is equally important because this characteristic acts to degrade the regression and its associated statistics. Autocorrelation in the residuals of the regressions is tested via the Durbin–Watson (DW) statistic (Wilks 2011). In cases where a positive result is obtained, the multiple linear regressions were reperformed using the Cochrane–Orcutt (CO) method (Cochrane and Orcutt 1949). Finally, the net variance explained by the multiple linear regression is assessed, representing that part of the interannual ATR variability accounted for by the given set of predictor variables.

After the ATR metrics are assessed at each grid point in the domain considered, they are also areaically averaged over three regions in the eastern United States and then analyzed in the same manner as described above. We discuss how these regions were chosen in a later section. Since the CMIP5 model simulations considered only extend through 2005, the regional intercomparisons of observed and model simulated behavior are limited to the period 1951–2005, representing the full extent of the model simulations.

3. Observed characteristics of ATRs
   a. Long-term variability

Before examining the interannual variability and trends of anomalous temperature regimes, it is insightful to begin by addressing the question of statistical stationarity in the temperature dataset. Figure 1 shows the long-term trend in winter-mean temperature from 1949 to 2011. There are cooling trends across the Southeast, central United States and the West Coast, while there are warming trends over the upper Midwest and over the Rocky Mountains. Over much of the United States, the observed trends in winter mean temperature are small, and are statistically significant in only a few areas: along the Gulf Coast, over the upper Midwest, and over the Rocky Mountains. Thus, there is a regional structure of the trends in the DJF-mean winter temperatures. In addition, these statistically significant trends reach maximum values near 0.04 K yr\(^{-1}\). This finding is qualitatively similar to several previous studies (Lu et al. 2005; Cohen et al. 2012a,b; Wang et al. 2009). The regional minimum in temperature trends across the central United States and Southeast seems counterintuitive in the context of an overall warming global climate, but several studies have attributed this local cooling to feedbacks in the hydrological cycle (Wang et al. 2009; Pan et al. 2004; Portmann et al. 2009). To ensure that such trends in background winter-mean temperature do not influence our analysis of ATR trends, the background trends are explicitly removed from the temperature dataset prior to performing the remaining calculations.

![Figure 1. DJF seasonal-mean temperature trend values (shaded, K yr\(^{-1}\)), 1949–2011: black closed contours indicate statistical significance at the 95% confidence level.](image-url)
It is also useful to first examine the long-term seasonal average impact factor for cold air outbreaks and warm waves over the United States to provide a sense of where ATRs are most prevalent and to identify the spatial coherence of ATR behavior. Figure 2 shows the long-term seasonal average impact factor for CAOs and WWs. It is evident that CAOs tend to have the greatest impact between the Rocky Mountains and the Midwest and along the Gulf Coast, while WWs tend to strongly impact the upper Midwest and Southeast. WWs also appear to less severely impact the Pacific Northwest and Gulf Coast in comparison to CAOs.

The trend analysis reveals that the trends in ATRs from 1949 to 2011 are very small and insignificant across the continental United States (Fig. 3). The only statistically significant trend found is a decrease in WWs over a small portion of the Pacific Northwest. Thus, these results indicate that there have not been any significant changes in the seasonal-mean behavior of boreal cool season ATRs over most of the United States between 1949 and 2011. For comparison, this analysis was repeated using nondetrended data (not shown). In this case, the patterns in the trends of ATRs, especially WWs, are collocated with the patterns describing trends in the long-term winter-mean temperatures (Fig. 1). Therefore, we find that the trends in ATRs identified using the nondetrended data are an artifact of trends in the background seasonal-mean temperature. Thus, in order for our analysis to correctly describe variability in the behavior of temperature anomalies, it is deemed necessary to first detrend the temperature in our analysis.

Our detrended results extend the findings of Walsh et al. (2001) and Portis et al. (2006), who found that there has been no decrease in boreal cool season CAO events over a shorter data record. In contrast to our study, these prior studies did not detrend their data. However, it is likely that the background temperature trends were less of an issue in these earlier studies owing to the earlier time period analyzed. Nonetheless, our results are consistent with these prior studies. In addition, a recent study by Hankes and Walsh (2011) found that the number of extreme cold air mass events in the subarctic of North America, a source region for the cold air masses responsible for CAOs in the continental United States, has dramatically decreased over the past several decades. It seems plausible, then, that these changes should also be reflected to some extent in the regions to the south that are affected by CAOs originating from these higher latitudes. Paradoxically, however, our current results suggest that changes in extreme cold air masses in the subarctic have not translated into a concomitant change in CAO frequency or amplitude over the continental United States (also see Walsh et al. 2001; Portis et al. 2006).

![Figure 2: DJF seasonal-mean impact factor of days the detrended $T'$ was (a) below $-1 \sigma$ (CAOs) and (b) above $+1 \sigma$ (WWs), 1949–2011.](image)

![Figure 3: Trends (shaded) in the yearly impact factor for (a) CAOs and (b) WWs using detrended data, 1949–2011: black closed contours indicate statistical significance at the 95% confidence level.](image)
b. Modulation of ATRs by low-frequency modes

We next present the results of our assessment of the relationship between seasonal-mean ATR behavior and various low-frequency modes. Figure 4 presents a local correlation analysis that reveals several spatially coherent and statistically significant associations between ATRs and prominent low-frequency modes, indicating that part of the interannual variability in ATRs may be linked to these low-frequency modes. More specifically, ATRs within specific regions of the United States are preferentially modulated by subsets of the low-frequency modes considered. For instance, in the southeastern United States, CAOs are more likely to occur when the NAO is in its negative phase (Fig. 4, top left panel), which is consistent with the findings of prior studies (e.g., Walsh et al. 2001; Cellitti et al. 2006), while CAOs are more likely to occur over portions of the western United States during the negative phases of the PNA, PDO, and ENSO (Fig. 4, left). On the other hand, more WWs over the eastern United States are favored when the NAO (PNA, PDO, and ENSO) is in its positive (negative) phase (Fig. 4, right). However, we note that the statistically significant relationship...
between WWs and the PNA, PDO, and ENSO is restricted in its northward extent compared to that for the NAO. The correlation analysis also reveals that a greater number of low-frequency modes are implicated in modulating WWs (4) compared to CAOs (1) in the eastern United States, while the reverse is true over the western United States (0 and 3, respectively). This is a previously unrecognized result. Based on these findings, for the eastern United States it appears easier to attribute interannual variability in WWs to low-frequency mode modulation than is possible for CAOs, while the opposite appears to be true for ATRs over the western United States. Thus, overall, the correlation analysis indicates that (i) ATRs in specific regions of the United States are modulated by certain low-frequency modes and (ii) there is an apparent asymmetry between the regional low-frequency modulation of CAOs and WWs. This asymmetry may be related to the higher-order statistical moments, such as skewness and kurtosis.

c. Quantifying the low-frequency mode modulation of ATRs

The multiple linear regression (MLR) analysis reveals that a linear combination of two predictors explains the maximum amount of variance in most cases, even though the optimal combination of predictors was not limited to any specific number. Overall, where statistically significant correlations between ATRs and low-frequency modes exist in the United States (i.e., Fig. 4), various combinations of these low-frequency modes explain between 10% and 50% of the variance in the ATR metrics. To provide a specific example, we consider the behavior of boreal cool season WWs and CAOs over the Southeast United States, which is impacted by all four low-frequency modes considered. The results of our MLR indicate that one predictor, the NAO, is the optimal subset of predictors for CAOs over the U.S. Southeast, explaining approximately 30% of the interannual variability (Fig. 5, left panel). This finding agrees with the correlation maps provided in Fig. 4, which show that of the four low-frequency modes of interest, only the NAO provides a statistically significant modulation of CAOs in the Southeast. In contrast, the combination of the NAO and PNA indices is the optimal set of predictors for WWs in this region, explaining over 50% of the interannual variability (Fig. 5, right panel). In fact, this combination of predictors explains between 40% and 50% of the variability in the impact factor of warm days over a very large area, covering most of Texas and extending eastward through the Southeast and then northeastward into Kentucky and Virginia. These results are also consistent with the correlation patterns plotted in Fig. 4. Conversely, very little of the WW variability over the western United States is accounted for by the same two predictors. This is because neither the PNA nor the NAO have a statistically significant relationship with WWs in this region (Fig. 4, right). Therefore, in principle, in regions where low-frequency modes are significantly correlated with ATR metrics, predictions of the implicated low-frequency modes could be used to forecast at least a portion of the seasonal impact of ATRs.

4. Representation of ATRs in CMIP5 simulations

For ease of presentation and comparison among models, the observed and model-simulated ATR metrics are next analyzed over three distinct regions in the eastern United States: the Southeast (SE), Midwest (MW), and the Northeast (NE) (Fig. 6). These three regions were chosen to include several highly populated metropolitan cities, while also minimizing the moderating effect on surface temperature over the oceans. Our focus on the eastern half of the United States is largely based on the high societal impact of ATRs in this region (Fig. 2), as well as the strong correlations between the ATR metrics and low-frequency modes observed in these regions (Fig. 4), especially for WWs. Further, there are ongoing research efforts focusing on temperature extremes over California and the western United States (e.g., Grotjahn

![Figure 5](https://example.com/figure5.png)

**Fig. 5.** Variance explained (%) by multiple linear regressions for the yearly impact factor of warm (cold) days with the NAO and PNA (NAO) as predictors, detrended, 1950–2011.
Thus, we decided to focus our attention primarily on the eastern United States. In addition, due to the availability of the model data, these analyses are conducted over the time period 1951–2005.

a. Surface air temperature climatologies and trends in ATRs in the CMIP5 simulations

Before assessing ATR behavior in the CMIP5 models, we first study the representation of the climatological annual cycle of surface air temperature in the models. The observed seasonal cycle of surface air temperature is encompassed within the range of the seasonal cycles exhibited by the 16 models, noting that the range is not very large (Figs. 7a–c). In addition, both high-top and low-top model averages show climatological seasonal cycles similar to observations. When directly comparing the high-top and low-top model averages, there are no prominent differences between them in representing the

![Image of the three regions in the United States for which anomalous temperature regimes (ATRs) were identified and analyzed using CMIP5 model data.](image)

![Graphs showing the climatological seasonal cycle of the (top) mean and (bottom) standard deviation (std) of surface air temperature (at the 0.995 sigma level) in the (left) SE, (center) NE, and (right) MW regions, based on the time period from 1 Jan 1950 to 31 Dec 2005. The average seasonal cycle of high-top (red solid line) and low-top (blue solid line) models and the observed (NNR) seasonal cycle (black dashed line) are shown, as well as the range for the 16 models (gray solid lines).](graphs)
seasonal cycle of surface air temperature. Therefore, we demonstrate that the climatological seasonal cycle of the daily mean surface air temperature appears well represented in most models. Similar conclusions can generally be made about the climatological cycle of the daily mean surface air temperature standard deviation, albeit with a few minor differences. For the NE and MW regions, the annual peak in the standard deviation of the daily mean surface air temperature is slightly delayed in both high-top and low-top models compared to the observed annual cycle (Figs. 7e,f), while in the SE region, the model-simulated variability during boreal winter is underestimated compared to the observations (Fig. 7d). However, it is expected that the differences in the high-top and low-top model averages in the seasonal cycle of the standard deviation will be slightly larger than those in the seasonal cycle of the long-term mean. Nevertheless, the 16 models mimic the expected cyclical pattern of stronger variability during boreal winter compared to boreal summer.

Next, the regional ATR impact factor measures are calculated for each model and the observations. The long-term mean and interannual standard deviation of the CAO and WW impact factor metrics for each of the three regions are first analyzed (Fig. 8). In the SE and NE regions, the average impact factor for the 16 models is less than (exceeds) observed values for CAOs (WWs) (Figs. 8a,c). In the MW region, the mean impact factor of the 16 models slightly exceeds (falls below) the observed mean impact factor of CAOs (WWs). In contrast, the interannual variability (or standard deviation) of the impact factor metric for both CAOs and WWs in all three regions is consistently overestimated by the models (Figs. 8b,d). Focusing next on the comparison between the high-top and low-top models, the 55-season means of the high-top models are more similar to the observations (black) than those of the low-top models (blue) in most cases (Figs. 8a,c). With regard to the interannual variability of ATRs, the high-top model performance is also generally closer to the observed value than the low-top models (Figs. 8b,d).

Upon comparing the results presented in Figs. 7 and 8, it is evident that ATR representation in the models is not very sensitive to the model ability to replicate the observed seasonal cycle. In addition, the models that replicate the seasonal cycle of the mean or standard deviation of surface air temperature better do not necessarily represent ATR behavior better. For example, the low-top models show slightly better results in representing the seasonal cycle of the standard deviation of daily mean surface air temperature for the SE region (Fig. 7d) but, in terms of representing the mean and standard deviation of the SE impact factor metric (Figs. 8a,c), they perform more poorly than the high-top models.
The linear trends in the winter-mean ATR impact factor for the 16 CMIP5 simulations are assessed next and are compared to the observed values (Table 1). In the observations, there are some slight differences in the trend values for WWs between the longer (1949–2011) and shorter (1951–2005) time periods for the SE and NE regions. More specifically, the observed trend for WWs in these two regions for 1949–2011 is slightly negative, while that for 1951–2005 is weakly positive. This distinction is not considered to be of great concern, however, as neither is statistically significant. Similarly, the majority of the trends in the models are also not statistically significant (only 6 of the 96 model trend values are deemed significant at the 95% confidence level). Thus, the models are consistent with observations in representing a general lack of significant ATR trends over the latter portion of the twentieth century.

b. Low-frequency mode modulation of ATRs in CMIP5

To further investigate the ability of the model simulations to represent the interannual and longer-term variability in ATRs, we study the modulation of ATRs by low-frequency modes in the models via correlation, clustering, and multiple linear regression analyses. First, a correlation analysis is performed to examine how well the CMIP5 model simulations represent the general relationship between the ATR impact factors and the primary low-frequency modes. To facilitate direct comparison between observations and model data, the correlation coefficients between the observed ATR impact factor metrics and the observed low-frequency mode indices were recalculated for the shortened time period (1951–2005).

In the SE region, observed CAOs (WWs) have a statistically significant association with the NAO (NAO, PNA, PDO, and Niño3.4), while the NAO and PDO modes exhibit statistically significant associations only with observed WWs in the NE region (Fig. 9). In the MW region, none of the low-frequency modes has statistically significant associations with either CAOs or WWs. Some of the models are able to simulate the significant observed relationships between the impact factor metrics and the low-frequency modes quite well. In fact, the average correlation values of the 16 models closely resemble those of the observations in most cases. It is also interesting to note that the majority of the statistically significant correlations in the model simulations tend to be related to two modes: the NAO-like and PNA-like modes. In contrast, correlations with the oceanic (PDO-like and ENSO-like) modes are relatively small and not statistically significant for most models, with the average correlation coefficient for the models near zero.

In apparent contradiction to conventional understanding, there is little to no relationship between ATRs
and ENSO over the regions of interest except for WWs over the SE region (Figs. 4, 9). However, much of the conventional wisdom is based upon an analysis of seasonal-mean temperature (Mo 2010; Ropelewski and Halpert 1986; Yu et al. 2012), in contrast to our current focus on shorter-term intraseasonal events. In addition, the multiple linear regression analysis indicates that the ENSO impact on temperature extremes across the contiguous United States is weaker than the other atmospheric modes. Our analyses suggest that, in fact, the ENSO impact upon intraseasonal temperature variability differs from the impact upon seasonal-mean temperature.

One substantial problem area for all the models considered is in the PDO modulation of ATR variability, which, in observations, is significant for WWs in both SE and NE regions. None of the models are capable of replicating a similar significant association between the PDO-like mode and the regional ATR metrics, which warrants additional investigation in the current study. This discrepancy can be understood in terms of a theoretically straightforward mechanism. The PDO is a prominent mode of coupled variability between the lower atmosphere and upper ocean over the North Pacific characterized by a zonal asymmetric SST anomaly pattern (Mantua and Hare 2002; Mantua et al. 1997; Schneider and Cornuelle 2005). Because the midlatitude baroclinicity over the North Pacific Ocean is determined by the horizontal temperature gradient strength in the lower atmosphere over the upper surface of the ocean, the horizontal structure of the PDO strongly influences both the intensity and dominant path of the atmospheric synoptic-scale waves traveling through this region. In other words, the PDO strongly influences the North Pacific storm track, which in turn impacts regional climate extremes, such as ATRs, over the continental United States (Lee et al. 2011). Therefore, if the spatial structure of the PDO-like mode in the model is dissimilar to observations, it is also likely that the model will be unable to replicate the observed downstream influences upon regional climate variability, such as ATR behavior.

To examine this issue further in relation to the CMIP5 models used here, the performance of the 14 model simulations in representing the spatial pattern of the PDO-like mode is assessed via a cluster analysis (Fig. 10). Interestingly but perhaps not surprisingly, none of the model structures ends up being classified into the same cluster category as the observations (Cluster 1). The robust zonal asymmetry exhibited in the observed PDO structure is weakened in the three other clusters, and the composite loading pattern for Cluster 3 locates the strong region of cold anomalies over the western North Pacific instead of the central North Pacific. In addition, Cluster 4 also exhibits latitudinal asymmetry in the form of a north–south dipole pattern, with enhanced warm anomalies at the high latitudes of the eastern North Pacific. Therefore, that many of the model simulations have a poorly resolved PDO-like mode likely explains why the models also lack any substantial PDO modulation of downstream ATR behavior. A similar argument may also apply to some of the other low-frequency modes with unrealistic relationships with ATRs. In addition, such discrepancies (particularly in representing the PDO) may help in accounting for some of the differences between the observed and simulated ATR metrics in Fig. 8.
Several other studies have also ascertained the poor simulation of PDO structures in coupled general circulation models. Overland and Wang (2007) examined the PDO-like pattern in twentieth-century simulations for the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4). This study found that only 10 coupled models among 18 have spatial patterns with a correlation of 0.7 or greater with observed PDO spatial patterns. Newman (2007) found that the eigenmodes in the IPCC models do not strongly correspond to canonical ENSO and PDO patterns. In Furtado et al. (2011), the spatial correlations of the leading EOFs of North Pacific wintertime sea level pressure (SLP) between the twentieth-century climate simulation (20C3M) scenario runs of the models and observations are much larger than those of the leading EOFs of North Pacific wintertime SST. Thus, this research is consistent with the indication from our current study that the best representations of the low-frequency modulation of ATRs by the models are associated with the atmospheric modes (i.e., NAO and PNA) and not the oceanic modes (i.e., PDO).

Besides the PDO, an evaluation of the other three modes is also relevant. The simulated behavior of the NAO-like, PNA-like, and ENSO-like mode spatial structures in the CMIP5 simulations is also extensively analyzed via a clustering analysis (Lee and Black 2013). Unlike the PDO mode, the authors find that either three or four of the model spatial patterns for these three modes are included in Cluster 1, which indicates that the respective low-frequency modes in those models have a horizontal structure and intensity similar to the observed low-frequency mode structure. Therefore, the models clearly do a much better job of representing the NAO, PNA, and ENSO patterns than the PDO pattern. This is consistent with our current results that the relationship between ATRs and the NAO, PNA, and ENSO modes is generally well represented by the CMIP5 models (Fig. 9). This is in stark contrast to the model performance in representing the PDO, as well as its connection to ATRs, motivating our current focus on validating the PDO pattern.

Given the imperfect ability of individual models to simulate the associations between the ATR metrics and some of the low-frequency modes, one important question remains: Do these model shortcomings preclude their use as predictors for the current and future behavior of ATRs? This question is investigated indirectly via parallel multiple linear regression analyses. As discussed in section 3c and shown in Fig. 5, the combination of the NAO and PNA indices best replicates the observed interannual variability of the WW impact factor over the U.S. Southeast. To examine how these two predictors perform in the model simulations, the time series of the yearly impact factor in the SE region is reconstructed using a multiple linear regression based on the NAO-like and PNA-like modes as resolved in the individual models. The $R^2$-squared values for each of the models are plotted in Fig. 11, and are contrasted with the $R^2$-squared values for the observations and the model average. In the observations, these two modes account for slightly over 50% of the interannual variance in the WW impact factor for the SE region. In the models, the amount of variance explained by these two modes is, on average, just under 40% ($\sim$12% less than for observations).

One remaining question associated with the collective influence of the NAO and PNA modes upon ATRs regards the difference in the performance of high-top and low-top models. Considering the variance explained for the individual models, four of the models are associated with values below 20%, while five models have values that
exceed observations. Interestingly, of the latter five models (MPI-ESM-LR, CCSM4, GFDL-ESM2G, CNRM-CM5, and NorESM1-M), four are low-top models (CCSM4, GFDL-ESM2G, CNRM-CM5, and NorESM1-M). In contrast, of the four models that explain less than 20% of the variance (HadGEM2-CC, IPSL-CM5A-LR, IPSL-CM5A-MR, and INM-CM4.0) three are high-top models (HadGEM2-CC, IPSL-CM5A-LR, IPSL-CM5A-MR). On average, the amount of variance explained by high-top models is 29.7%, while the average amount of variance explained by low-top models is 45.8%. Therefore, the current results suggest that the combined use of the NAO-like and PNA-like modes in predicting WWs over the SE region is better replicated in low-top models than in the high-top models. This result is quite interesting and is related to the results presented in a separate paper (Lee and Black 2013). In that paper, the authors find that the overall behavior of the NAO and PNA patterns is better represented in the low-top models than in the high-top models. Thus, one possible reason for the poorly represented relationship between ATRs and low-frequency modes found in the current study for the high-top models could be the poorly represented spatial structure of the low-frequency modes in these models. Another notable point is that even though the individual model configurations often cannot precisely model all aspects of ATR metrics, important information regarding the variability in ATRs can still be gleaned from several of the models as well as the model average. In addition, the results presented in Fig. 11 likely provide a useful discriminating tool that can be used in selecting an optimal subset of models to study in an assessment of future changes in ATR–low-frequency mode linkages.

In a multiple linear regression analysis, it is also sometimes informative to partition the variance explained among the various predictors. The partial variance explained that is uniquely attributable to individual predictors is computed by comparing two regression models: a complete model and a reduced model (Kerr et al. 2002; Tabachnick and Fidell 2001). The complete model is the multiple linear regression with all of the predictors in inclusion. The partial variance explained that is uniquely attributable to the NAO is the variance explained for the complete model minus the variance explained for the reduced model in which the variable of interest (the NAO) is omitted. The remaining variance explained that is not attributable to either solely the NAO or PNA is regarded as the confounded variance between the modes and reflects the correlation between the two predictors. In the observations, the PNA explains more than three times the amount of variance of WWs in the SE region compared to the variance explained by the NAO (Table 2). In the model average, the amount of variance explained by the NAO is similar to observations, while the amount of variance explained by the PNA is far underestimated (less than half of the observed value). In addition, the confounded variance for the CMIP5 models is more than double that for observations, and the model average total explained variance is less than the observed value. This reduced total explained variance in WWs over the SE region may be attributed to differences in the detailed structures of the PNA pattern over the continental United States and the resultant misrepresentation of ATR modulation.

5. Conclusions

During the boreal cool season, anomalous temperature regimes (ATRs), including cold air outbreaks (CAOs) and warm waves (WWs), affect regional economies and human safety over large portions of the United States via their significant impacts on energy consumption, local agriculture, and human health. Through this research, we provide a baseline quantification of long-term ATR statistics. The present study applies a range of statistical analyses to (i) examine long-term variability in the regional impact of ATRs, (ii) identify and quantify interannual modulation of ATR behavior by prominent modes of low-frequency variability, and (iii) examine the representation of ATR behavior in the CMIP5 model simulations.

Analyses of long-term ATR variability reveal that there have not been any statistically significant trends in either WWs or CAOs over most of the continental United States from 1949 to 2011. Nonetheless, strong interannual variability in ATRs is still evident and is linked to several prominent low-frequency modes. More specifically, a correlation analysis shows that regional ATR behavior is modulated by certain low-frequency modes with important asymmetries between the low-frequency mode modulation of CAOs and WWs. In particular, we discover that the low-frequency modulation of WWs over the eastern United States is considerably more robust than the parallel modulation of CAOs.

<table>
<thead>
<tr>
<th>Source</th>
<th>NAO (unique)</th>
<th>PNA (unique)</th>
<th>Confounded</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>NNR</td>
<td>11.29</td>
<td>37.29</td>
<td>4.44</td>
<td>53.02</td>
</tr>
<tr>
<td>Model average</td>
<td>10.29</td>
<td>19.21</td>
<td>9.48</td>
<td>38.98</td>
</tr>
</tbody>
</table>

Table 2. The partial variance explained by a multiple linear regression for the yearly impact factor of warm days above +1σ over the SE region.
Further, the influence of ENSO upon ATRs is found to be mainly limited to a modest modulation of WWs over the southern United States. This result is contrary to prior studies of seasonal-mean temperature. Additionally, multiple linear regression analyses demonstrate that the optimal collective influence of low-frequency modes can account for as much as 50% of the interannual variability in ATR behavior.

Parallel statistical analyses of 16 AOGCM models from the CMIP5 project are also performed. Although most of the models are able to replicate the observed seasonal cycle of daily mean surface air temperature, WWs tend to be overestimated by the models, while CAOs tend to be underestimated. This is particularly true for the SE and NE regions. In addition, the average interannual variability of ATRs is larger in the model simulations than found in the observations. Similar to observations, trend analyses reveal few statistically significant trends in either WWs or CAOs over the United States. Also, the model correlation analysis reveals that the models represent several significant associations between the ATR metrics and low-frequency modes of variability, particularly the NAO- and PNA-like modes. However, the significant modulation of ATRs by the PDO-like mode is virtually nonexistent in most models. This is most likely due to model inadequacy in representing the underlying physics of the PDO. The simulated PDO-like structures show marked distinctions from the typical zonally asymmetric SST anomaly pattern found over the North Pacific Ocean in association with the observed PDO. Finally, the multiple linear regression analysis indicates that the collective influence of the NAO- and PNA-like modes accounts for almost 40% of the interannual variability in the impact factor of WWs in the SE region, with several individual models performing very close to observations. Interestingly, we find that there is no apparent benefit of high-top models in replicating the observed characteristics of ATR behavior and its low-frequency modulation. The current results provide fundamental information on simulated ATR variability that provides a useful benchmark for assessing the representation of regional climate variability in global coupled climate models and determining likely future variability in ATR behavior.

Anomalous temperature regimes will remain of interest to weather forecasters because of their socioeconomic impacts. Therefore, accurate prediction and forecasting of these events is desirable. However, our analysis has thus far been primarily statistical in nature, and therefore does not establish the physical processes or mechanisms connecting ATRs to low-frequency modes. In addition, little research currently exists on ATR triggers and how they operate. As a result, our future work will seek to identify and quantify the dynamical mechanisms responsible for ATRs. This research will be pursued via detailed diagnostic analyses of three-dimensional synoptic and dynamic circulation structures so as to isolate the physical features of ATRs and assess the roles of low-frequency modes in modulating ATRs. The manifestation of these dynamical mechanisms will also be explored in the CMIP5 models.

Acknowledgments. This research was conducted under support by the U.S. Department of Energy, Office of Biological and Environmental Research, Award DE-FOA-0000242.

REFERENCES