Understanding and predicting temperature variability in the observational record

Understanding and predicting temperature variability in the observational record

a dissertation presented
by
Karen Aline McKinnon
to
The Department of Earth and Planetary Sciences

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Earth and Planetary Sciences

Harvard University
Cambridge, Massachusetts
April 2015

Thesis advisor: Professor Peter Huybers Karen Aline McKinnon

Understanding and predicting temperature variability in the observational record

Abstract

Temperature variability and change over land and ocean exhibit characteristic spatial and temporal structures. Understanding the physical mechanisms underlying these structures provides information about the movement and storage of heat in the climate system. In this thesis, I first analyze, and present an energy balance model for, seasonal temperature variability in the extratropics, which supports the idea that the advection of heat between land and ocean by the mean atmospheric circulation can explain the regional characteristics of seasonal variability over both land and ocean. The model is subsequently combined with a large, representative ensemble of Lagrangian atmospheric trajectories to provide a realistic model of the seasonal cycle in the Northern Hemisphere mid-latitudes. Second, based on the Lagrangian trajectories, a new spatially-resolved metric, termed Relative Land Influence, is developed. The metric quantifies the role of land as compared to ocean in influencing the temperature variability at a given location. In addition to explaining the majority of the spatial pattern of seasonal variability, Relative Land Influence is a significant predictor of the observed temperature change over both land and ocean independently since 1950, suggesting that similar physical processes influence temperature variability on seasonal and decadal timescales. Finally, I explore the tails of temperature distributions in the context of identifying the causes of anomalously hot days in the Eastern United States during peak summer. A coupled ocean-atmosphere mode in the central mid-latitude Pacific is identified, which evolves on a characteristic timescale and ultimately leads to the amplification of a mid-latitude wave train that includes a blocking high over the Eastern United States.

The early identification of the sea surface temperature precursors to this mode allows for skillful prediction of heat events at lead times greater than 40 days. The identification of physical processes underlying temperature variability on a range of timescales can inform predictions of how temperature variability may change in the future.

Contents

Acknowledgments

While there is a single name associated with a PhD thesis, the reality is that a great many minds combine to produce the work. My greatest thanks go to my adviser, Peter Huybers, who has been instrumental in guiding me through my entire scientific career thus far. I am thankful for both his mentorship, through which I have learned how to think critically yet creatively, and for the independence I was given to explore during my PhD. I would also like to thank the remainder of my committee — Zhiming Kuang, Jerry Mitrovica, and Dan Schrag — for their helpful feedback and advice on both my scientific work and career trajectory. I have also benefited from conversations with Brian Farrell and Carl Wunsch, which have been made possible by their willingness to keep their doors open to curious graduate students. Finally, thank you to Cecilia Bitz for hosting me at the University of Washington for a summer.

The administration in the Earth and Planetary Sciences department is unmatched in their support for graduate students, for which I am grateful. Special thanks to Sarah Colgan, Chenoweth Moffat, and Sabinna Cappo for helping me navigate the obstable course of graduate school. Thank you to Research Computing for supporting all my computing needs, and to both the NSF GRFP and NASA NESSF for providing funding.

I have had the priviledge of working with a wonderful research group during my time at Harvard. Thank you to Zan Stine, Martin Tingley, Andy Rhines, Ethan Butler, Cristi Proistosescu,

Marena Lin, and Nathan Mueller for all the conversations, laughs, and shared ideas. Thanks especially to Zan and Martin, who generously acted as both advisers and collaborators, and taught me about both science and how to be a scientist. Thanks also to Kathryn McKain, for being a great officemate, and putting up with all the Huybers group chatter.

The process of completing a PhD has been made infinitely more enjoyable due to a terrific group of friends and housemates. Thank you all for being there for both the adventures and the hard times.

Finally, thank you to my family for their support. Completing a PhD is a lot easier when you are following in the footsteps of both your parents and your brother! Thank you to Dad for stoking my scientific curiousity from a young age, to Mom for being an invaluable confidant, and to Daniel for always urging me to dig deeper.

Climate is what we expect, weather is what we get.

Robert Heinlein

Perhaps one should say that climate is what you expect, or better yet what you ought to expect, when you are not in a position to make a skillful weather forecast.

Edward Lorenz

1

Introduction

Surface temperature has traditionally been one of the most closely monitored aspects of climate, and serves as a convenient metric for climate variability and change. The earliest quantitative measurements of temperature were taken in the 17th century, after the thermometer was 'invented' by Santorio Santorio by adding a scale to a thermoscope, which provided only qualitative indications of relative temperature (Chang, 2004). The longest continuous temperature time series at a single location, the Central England Temperature dataset, was initiated more than a century later, in 1772 (Parker et al., 1992). Since then, weather stations measuring temperature and other meteorological variables have been established around the world with varying density, quality, and longevity. The United States in particular maintains a dense network of long-running stations, while less developed regions of the world contain fewer high-quality stations (Fig. 1.1).

Taking the temperature of the ocean is arguably more difficult than that of the land, but the first ship measurements of ocean temperature, including the subsurface, are thought to have been on Captain James Cook's second voyage during 1772–1775 (Abraham et al., 2013). Since

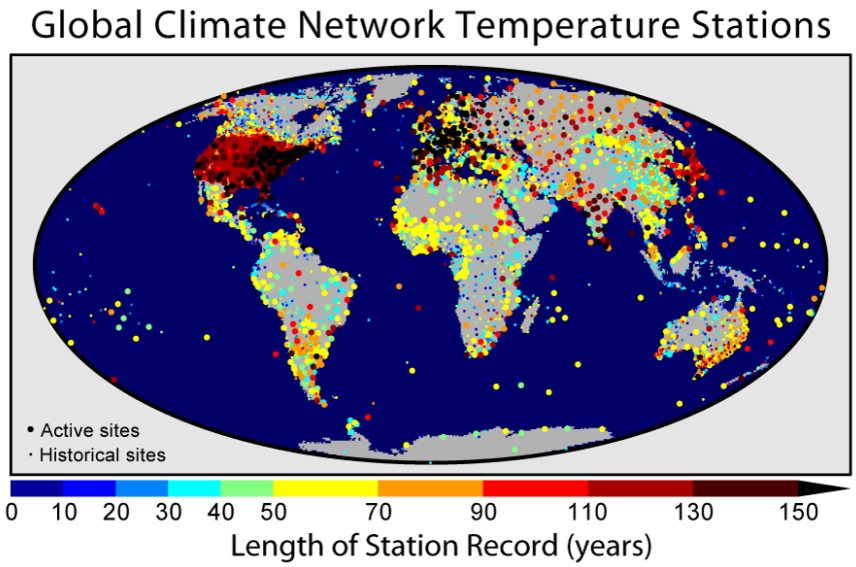

Figure 1.1: The spatial distribution of weather stations measuring temperature in the Global Historical Climatology Network Daily database. Color indicates the length of the station record. Image courtesy of Global Warming Art.

then, measurements of ocean temperature have been taken by merchant, navy, and scientific vessels, leading to a historic concentration of measurements along major shipping routes and in the Northern Hemisphere (Fig. 1.2). While the distribution of measurements of the world ocean has increased dramatically over the past century, only with the recent deployment of the Argo array have regions such as the high Southern latitudes become relatively well-sampled (Roemmich et al., 2009). In contrast to the subsurface ocean, sea surface temperatures (SSTs) have been continuously measured by satellites since the early 1980s (Reynolds et al., 2002).

The global – if uneven – spread of temperature measurements is a thus a relatively recent phenomena, but has been critical for our understanding of the earth system. Absent a phase change, such as water evaporating or ice melting, excess heating results in an increase in temperature, and temperature can thus serve as an imperfect indicator of the movement of heat throughout the climate system. Improved statistical and dynamical models for regional temperature vari-

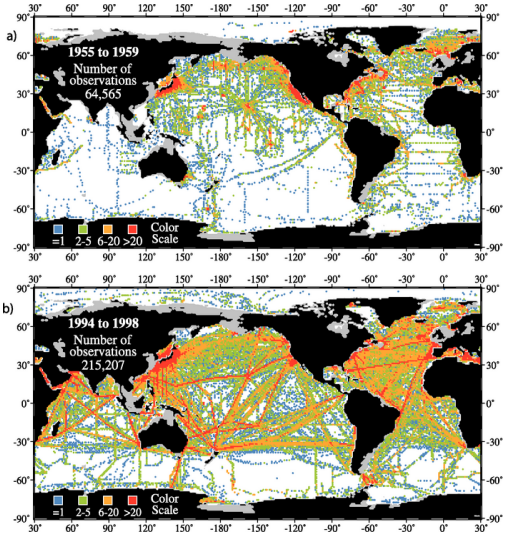

Figure 1.2: Counts of ocean observations at 250 meters depth for two different time periods. The observations are grouped into 1° grid boxes, and the color indicates count. Figure from the IPCC fourth assessment report (Solomon et al., 2007).

ability and change encapsulate information about transport, divergence, and storage of heat, and can provide insight into how temperature may be expected to change in response to the changing composition of the atmosphere. The regional responses, integrated globally, are related to climate sensitivity, defined generally as the temperature response to a perturbative radiative forcing.

The relationship between temperature, T, and heating, can be discussed, to first order, based on an energy balance model for regional temperature,

$$C\frac{dT}{dt} = R + \lambda T - \nabla \cdot \mathbf{F}, \tag{1.1}$$

where all variables can change in space and time, and all terms are anomalies from a specified base state. The left hand side of the equation represents the local heat storage, written under the assumption that the imbalance can be expressed as the time derivative of temperature scaled by

the heat capacity, C. The right hand side delineates the sources of the local heating (or cooling): incoming radiation, R, assumed to be independent of temperature; outgoing radiation, λT, assumed to be linearly dependent on temperature; and horizontal divergence, $\nabla \cdot \mathbf{F}$, which takes into account transport of heat from (or to) other regions. The scaling factor, λ, that links temperature and outgoing longwave radiation is termed the feedback factor, and encapsulates the equilibrium sensitivity of temperature to a radiative imbalance.

The model can be examined in order to make a few observations. First, the energy balance at any given location can be controlled by four factors, even in this simplified model, providing some indication of the challenges underlying an understanding of regional temperature variability. Second, the heat capacity of the system becomes irrelevant in equilibrium. Third, when globally averaged, the divergence term must disappear, and λ must be negative to allow for a stable system. The determination of λ, however, requires knowledge of the global heat storage and radiative balance.

This thesis takes a data-driven approach to address each of these three topics in an attempt to gain greater insight into the controls on regional surface temperature variability on daily, seasonal, and inter-decadal timescales.

1.1 Regional temperature variability and the seasonal cycle

Temperature varies at all frequencies, with variability increasing with decreasing frequency (Hasselmann, 1976). This behavior holds for global mean temperature on timescales of thousands of years, as well as for regional temperature on the daily to decadal timescales for which spatially-resolved data is available (Huybers and Curry, 2006). While the temporal variations of global mean temperature have been large over earth history, temperature variations across the globe are much larger, and these localized variations play a significant role in the global distribu-

tion of crops (Lobell and Field, 2007), diseases (Bhatt et al., 2013), and human populations (Small and Cohen, 2004). Despite their importance, the skill of general circulation models in predicting regional temperature variability and change tends to be limited (e.g. Sakaguchi et al., 2012).

The power spectra of regional temperature are surprisingly coherent (Fig. 1.3). In particular, the magnitude of variability on annual timescales ($f = \mathrm{yr}^{-1}$) at a given location is a good predictor of variability on both sub-annual and multi-decadal timescales (Huybers and Curry, 2006), providing a hint that the processes controlling the spatial structure of temperature variability may be similar from days to decades. Should this hold true, one timescale could be used as an analogue to study another.

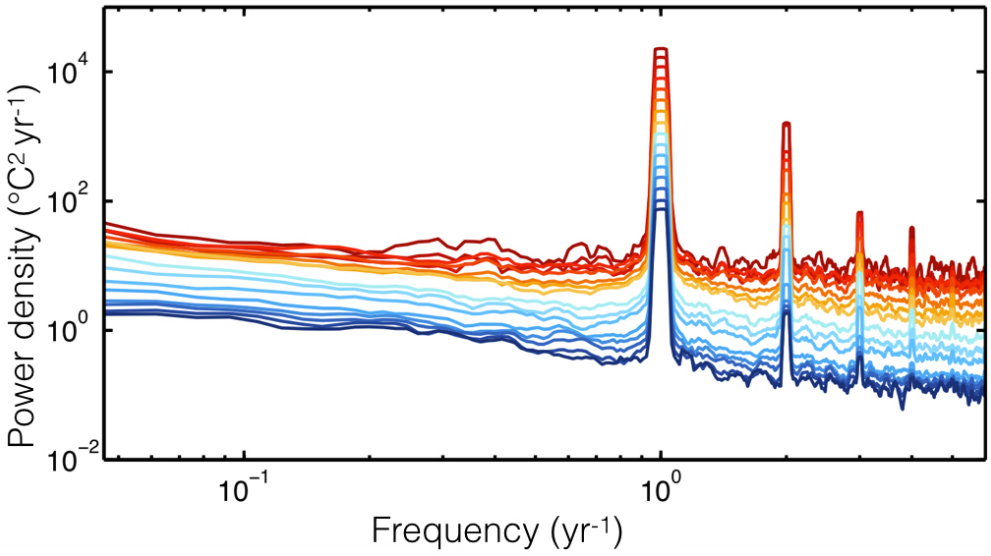

Figure 1.3: Power spectra for extratropical near-surface air temperature from NCEP-NCAR reanalysis (Kalnay et al., 1996), adapted from Huybers and Curry (2006). Grid boxes are grouped according to their annual period variability, which is also indicated by the line color. Power spectra are calculated using Thomson's multitaper power spectral density estimator (Thomson, 1982).

The seasonal cycle in particular is a promising timescale for exploration, because the magnitude of the solar forcing and temperature response are large, and it is observed every year. The signal-to-noise ratio for the seasonal cycle in temperature is perhaps the largest in the climate

system, leading to clear spatial patterns in the amplitude and phase of the seasonal cycle, relative to the solar forcing (Stine et al., 2009). In contrast, the observed temperature response pattern to increasing radiative forcing from anthropogenic influence is confounded by decadal variability (Stott, 2003), a decrease in the availability of data going back in time, and uncertainty about the magnitude and spatial pattern of the radiative forcing (Shindell and Faluvegi, 2009).

The spatial variability in the seasonal cycle has been used to understand the temperature response to orbital variations (Laepple and Lohmann, 2009) and to predict equilibrium temperature changes in different regions (Knutti et al., 2006). The skill of these approaches will be dependent on the extent to which the relevant physics holds on both timescales. Processes with timescales shorter than seasonal would be expected to be relevant on seasonal and longer timescales, as would processes that have timescales longer than both seasonal and the second, longer timescale. In contrast, if a process is slow with respect to seasonal timescales but fast with respect to the second, longer timescale, then the seasonal analogue may not be appropriate.

In Chapter 2, I present a description of the extratropical seasonal cycle, as measured by its amplitude and phase, including the identification of specific relationships between the amplitude and phase that have not been previously noted. I subsequently develop a set of energy balance models to explain these relationships. The models highlight and quantify the dominant influence of advective heat transports on the seasonal cycle.

In Chapter 3, I return to the advective model developed for the seasonal cycle, and present a metric, Relative Land Influence, that has predictive skill for not only the spatial structure of the seasonal cycle but also the pattern of temperature change observed since 1950. The model used to predict the pattern of temperature change also provides information about the magnitude of the climate feedback factor, discussed in the next section.

1.2 Estimating climate feedback factors and ocean heat content changes

A parameter of interest in the climate system is the climate sensitivity, typically defined as the global mean temperature response to a doubling of carbon dioxide concentrations from pre-industrial levels and proportional to λ^{-1} in Eq. (1.1). One of the challenges in estimating climate sensitivity from the observational record is a lack of knowledge about the magnitude of the heating imbalance, the vast majority of which (>90%) is heating the ocean (Loeb et al., 2012). In the case of a nonzero heating imbalance, the system has not yet equilibrated with the perturbative radiative forcing, so the observed temperature change is smaller than the theoretical equilibrium temperature change. Available observations have been leveraged to estimate climate sensitivity from the observational record, accounting for the uncertainty in radiative forcing and ocean heat uptake, leading to published ranges (90% confidence interval) of climate sensitivity including 1.4-7.7°C (Forest et al., 2002), 1.6°C with no upper bound (Gregory et al., 2002), and 0.9-5.0°C (Otto et al., 2013), respectively. By requirement, these observational estimates of climate sensitivity assume a constant, instantaneous value of the feedback factor. If these assumptions do not hold, a wide range of equilibrium climate sensitivities could emerge on long timescales (Proistosescu and Huybers, submitted).

Observational estimates can be complemented by general circulation model simulations. Due to the computational intensity of the models, it is not typically possible to run the simulations until equilibrium. Rather, analogous to the approach taken using the observational record, the climate feedback factor is assumed to be constant, and the modeled temperature response is extrapolated to a zero net heating imbalance. Using this approach, Andrews et al. (2012) found the range of equilibrium temperature responses across the recent Climate Model Intercomparison Project 5 (CMIP5) ensemble to be 2.1-4.7°C. Inherent in both the observational- and model-based estimates is the assumption that the currently observed feedback factor is representative

of that for longer timescales; thus, slow feedbacks emerging from changes to the land surface, cryosphere, or deep ocean are implicitly excluded.

In Chapter 3, I present a different method for estimating climate sensitivity based on the instrumental record, which takes into account both the magnitude of the observed warming and the spatial pattern. Based on insight from the seasonal cycle, the spatial pattern of warming is argued to primarily emerge from spatial variability in divergence (the final term in Eqn. (1.1)), and is used to jointly constrain the ocean heat uptake anomaly and the feedback factor under assumptions about the radiative forcing.

1.3 Describing and predicting extreme temperature

Superimposed on the slowly varying climate are synoptic-scale weather events. These events themselves are not, of course, distinct from the climate: rather, the average of the weather events defines the climate. It can be useful, however, to take advantage of the separation of timescales between climate and weather, and focus on the dynamical origins of specific weather events against the thermodynamic background state of mean climate.

Extreme weather events are of particular interest because of their substantial societal impacts; the focus here is on unusually hot days, or heat waves. In a thermodynamic sense, heat waves can be caused by advection of heat from warmer regions, by an increase in downward radiation due to clear skies, and/or by a decrease in the effective heat capacity of the land related to desiccation of the soil (Black et al., 2004; Miralles et al., 2014). These categories, however, are diagnostic rather than prognostic, and raise the question of what dynamical pathways lead to the anomalous heating and high temperatures sustained over a period of multiple days.

Heat waves are typically associated with a 'wavy' jet stream or blocking highs. A wavy jet stream can lead to advection of warm, remote air into a cooler region (Mazon et al., 2014), whereas

a blocking high initiates a set of positive feedbacks wherein the subsidence of warm, dry air and the clear skies associated with a blocking high lead to increased warming of the atmospheric column, thus reinforcing the high pressure system (Cerne et al., 2007). These types of events occur as part of random atmospheric variability (Hassanzadeh et al., 2014), but may be more probable during some climatic background states. Recent work has focused on the connection between high-latitude warming and sea ice loss, and the increased frequency of circulation anomalies associated with hot summer days (Tang et al., 2014; Coumou et al., 2014). A proven connection between the two would suggest a future increase in hot days not only due to mean warming, but also because of an increased frequency of circulation anomalies such as blocking highs. To solidify a connection between changes in the high-latitudes and mid-latitude weather circulation, however, it is necessary to identify a causal pathway and demonstrate prognostic skill, which remains to be done.

A related topic, addressed in Chapter 4, is whether certain climate states that occur within the envelope of unforced natural variability are more likely to precede heat waves in the United States. A correlation between North Pacific sea surface temperatures and hot summers in the United States was identified more than 30 years ago (Namias, 1982), but subsequent theoretical and modeling work has failed to identify a clear and consistent influence of mid-latitude sea surface temperatures on mid-latitude circulation (Kushnir et al., 2002). In Chapter 3, I use land and ocean temperature observations from the past three decades to identify an apparent coupled ocean-atmosphere mode that increases the probability of heat waves in the Eastern United States. Whether the frequency of the mode is changing is left for future work.

1.4 Road map

The three chapters of this thesis are published or submitted papers that are linked by a common theme of inferring the physical mechanisms underlying observed temperature variability. The chapters will be introduced with a brief note to contextualize their place in the thesis, but are otherwise reproduced with only slight modifications from their published forms. The thesis will conclude with a discussion of relevant future research questions inspired by the work done thus far.

2

The spatial structure of the annual cycle in surface temperature: amplitude, phase, and Lagrangian history

The annual cycle dominates the spectrum of temperature variability on timescales shorter than Milankovitch variability. In this chapter, I present a description of and physical model for the temperature response to annual-period solar forcing. The chapter is reproduced from its published form: McKinnon, K. A., A. R. Stine, and P. Huybers, 2013: The spatial structure of the annual cycle in surface temperature: amplitude, phase, and Lagrangian history. *Journal of Climate*, **26 (20)**, 7852–7862.

2.1 Abstract

The climatological annual cycle in surface air temperature, defined by its amplitude and phase lag with respect to solar insolation, is one of the most familiar aspects of our climate system. Here, we identify three first-order features of the spatial structure of amplitude and phase lag and explain them using simple physical models. Amplitude and phase lag (1) are broadly con-

sistent with a land and ocean end-member mixing model, but (2) exhibit overlap between land and ocean, and, despite this overlap, (3) show a systematically greater lag over ocean than land for a given amplitude. Based on previous work diagnosing relative ocean or land influence as an important control on the extratropical annual cycle, we use a Lagrangian trajectory model to quantify this influence as the weighted amount of time that an ensemble of air parcels has spent over ocean or land. This quantity explains 84% of the space-time variance in the extratropical annual cycle, as well as features (1) and (2). All three features can be explained using a simple energy balance model with land and ocean surfaces and an advecting atmosphere. This model explains 94% of the space-time variance of the annual cycle in an illustrative mid-latitude zonal band when incorporating the results of the trajectory model. The basic features of annual variability in surface air temperature thus appear to be explained by the coupling of land and ocean through mean atmospheric circulation.

2.2 Introduction

It has been long understood that the annual cycle in surface air temperature is largely controlled by the annual cycle in solar radiation, local surface conditions, and atmospheric circulation. Generally, oceanic climates have a small amplitude and large phase lag with respect to solar forcing, while continental climates have a large amplitude and small lag (Von Hann and Ward, 1903), with additional structure associated with the direction and strength of prevailing winds (Ward, 1906). This qualitative understanding of the systematic patterns in amplitude and lag of the annual cycle has also been supported by quantitative analysis, with a historical focus on obtaining a single measure of "continentality" that would reflect the relative influences of land and ocean.

Brooks (1917) used the land fraction in a series of concentric circles around a location as a predictor for the amplitude of the annual cycle, capturing the effects of nearby land or ocean, and

Brooks (1918) went on to account for the direction of the prevailing winds. Similar methods aiming to determine the amplitude of the annual cycle from a regression of geographic variables were presented by Spitaler (1922), Brunt (1924), and Hela (1953). Other work focused instead on the lag of temperature behind solar radiation as a measure of continentality (Prescott and Collins, 1951; van den Dool and Können, 1981), but amplitude and lag were not unified into a single framework for describing continentality.

More recently, Stine et al. (2009) focused on the relationship between amplitude and phase lag of the annual cycle, and showed that observations of the annual cycle could be approximately described as a linear mixture of two sinusoids, interpreted as theoretical ocean and land end-members. This conceptual framework provided a link between amplitude and phase, and also demonstrated that each provides a slightly different picture of the spatial structure of the annual cycle (see Fig. 1a,b in Stine et al., 2009). They further demonstrated the important role of the prevailing winds; however, their description of the annual cycle was fundamentally algebraic.

A number of simple models based on energy balance principles have also been applied to reproducing the annual cycle in surface temperature. Thompson and Schneider (1979) used a two-layer zonal model with diffusive heat transport and reproduced the zonally-averaged amplitude of the annual cycle, but the modeled temperature lagged the observations by one to two months. North et al. (1983) expanded on the one-layer diffusive energy balance model of North and Coakley (1979b) and North et al. (1981) to include realistic continental configurations, using heat capacity to distinguish between ocean, land, and coastal areas. Subsequent work used comparable models with application to the annual cycle, with comparable results (Hyde et al., 1989; Kim and North, 1992). These previous models parameterized atmospheric heat transport as a diffusive process and, while they do capture much of the first-order structure in the annual cycle, do not account for the role of of atmospheric advection (e.g. Brooks, 1918; van den Dool and Können, 1981). The influence of advection on seasonality is readily discerned in the east-to-west

structure across extratropical continents and oceans (Stine and Huybers, 2012).

Here, we analyze spatial variability in the climatological annual cycle in surface temperature and seek to explain, using simple physical models, three of its primary features: a first-order structure consistent with the extratropical annual cycle as a mixture of ocean and land end-member sinusoids (Stine et al., 2009), an overlap in amplitude and phase between ocean and land annual cycles, and a non-unique relationship between amplitude and phase, where the ocean systematically has a later phasing for the same amplitude. While the first feature can be well-predicted by westward distance from the coast (Stine et al., 2009), explanation of both the second and third features requires a more complete framework. It will be demonstrated that these features emerge from distinct ocean and land heat capacities in conjunction with mean circulation patterns.

2.3 Structure of the annual cycle

We use monthly climatological temperature data from the Hadley Centre Climate Research Unit (HadCRU, Morice et al., 2012). The climatology is based on the period 1961–1990 and is reported at 5° spatial resolution. Data are provided as monthly averages, which, given the Nyquist criterion that observing a sinusoid requires sampling at just less than half its period, is more than adequate to constrain the amplitude and phase of the annual cycle (Thompson, 1995). For example, our tests using the long ($>$ 30 years) extratropical records in the Global Historical Climatological Network (Menne et al., 2012) indicate an average correlation coefficient between estimates of phase lag and amplitude calculated using daily and monthly average temperatures of greater than 0.99, and a root mean square difference across the domain of 0.07 days for phase lag and 0.05 °C for amplitude. Only gridboxes where the HadCRUT4 product (Morice et al., 2012) has at least five of the thirty years of observational data for all months of the year are included in the

analysis. Each extratropical hemisphere is considered, but the tropics (23°N–23°S) are excluded because the annual cycle there contains substantial twice-per-year variability, and variations in local incoming solar radiation are a less dominant control on surface temperature relative to higher latitudes.

Over 99% of the space-time variance in the annual cycle in surface air temperature across the extratropics can be explained by the annual Fourier component, allowing for efficient representation of the monthly climatology as its amplitude and phase. In order to account for differences in solar insolation, we define gain as the amplitude normalized by the latitudinally-variable amplitude of the annual Fourier component in solar insolation (Berger and Loutre, 1991)[*], $G(x,y) = A_T(x,y)/A_{\text{sun}}(y)$, and lag as the difference in phase between the annual Fourier component of temperature and insolation, $\lambda(x,y) = \phi_T(x,y) - \phi_{\text{sun}}(y)$.

Gain and lag exhibit coherent spatial structure (Fig. 2.1). Gain is generally larger over Northern Hemisphere land masses, increases from west to east across continents, and increases more rapidly across North America than Eurasia. The smallest gains are found in the Southern Ocean and the North Atlantic, while the largest are in northeastern Eurasia. Lag exhibits a clearer land-ocean dichotomy, with an average lag of 28 days over land and 58 days over ocean. This structure can be efficiently represented in polar coordinates (Fig. 2.2), where gain is indicated by the distance from the origin, and lag by the angle measured counter-clockwise from the positive x-axis. The x-component is the gain that is in phase with the sun, and the y-component is the gain that is in quadrature with the sun. Polar coordinates offer the advantage that the mixing curve between two sinusoids is a straight line, and we use this representation to define three characteristics of the annual cycle:

(1) *Linearity*: The first-order structure in polar coordinates is linear with both the intercept and

[*]Code available at http://www.ncdc.noaa.gov/paleo/pubs/huybers2006b/huybers2006b.html

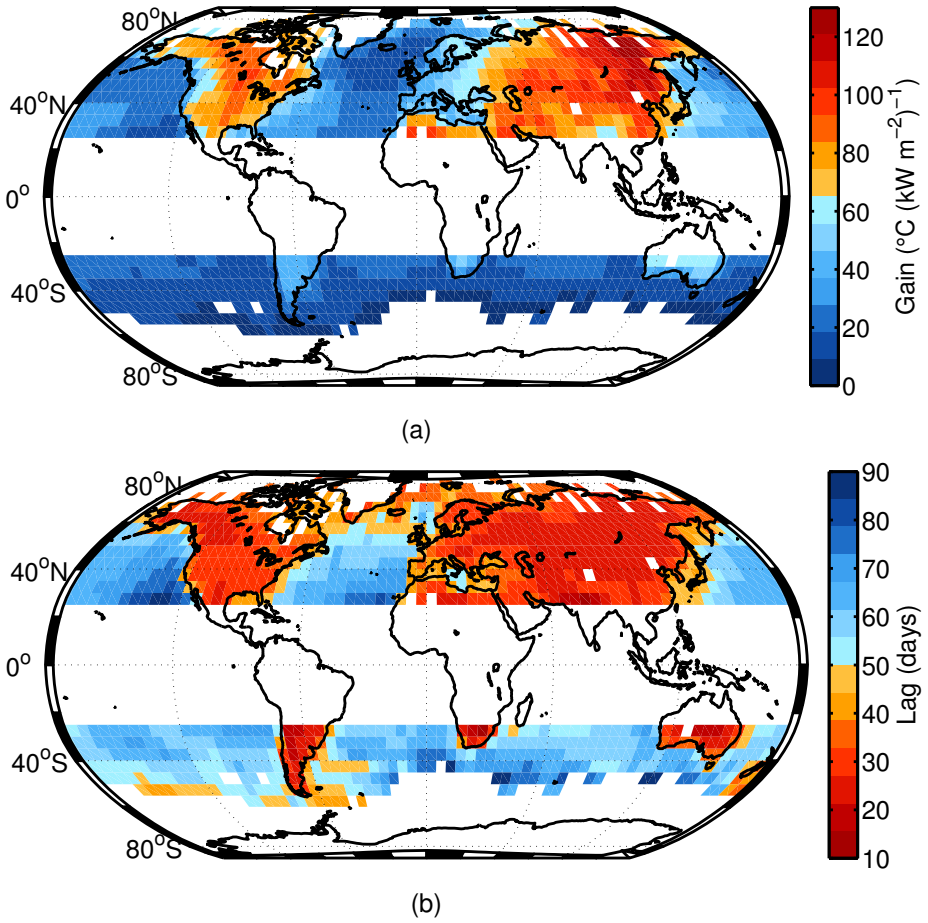

Figure 2.1: (a) Gain and (b) lag of the annual Fourier component of the climatological annual cycle.

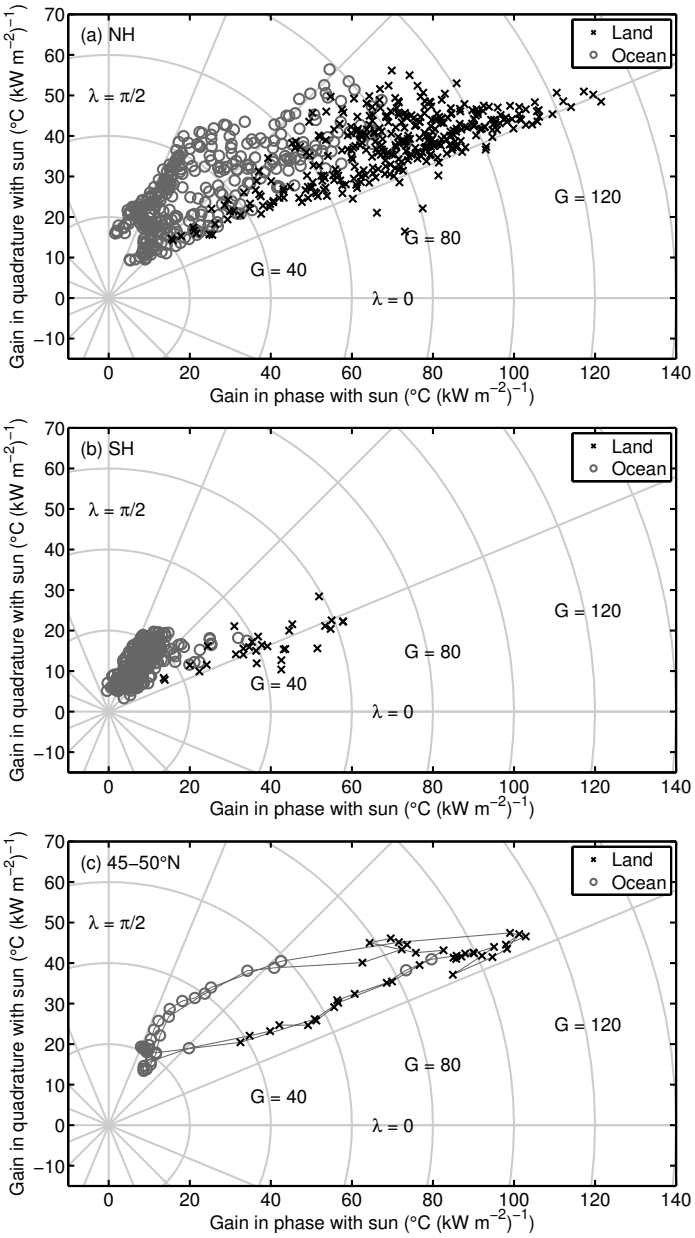

Figure 2.2: The gain and lag of the annual cycle in polar coordinates in the (a) Northern and (b) Southern Hemispheres, and (c) at 45-50°N. In (c), neighboring gridboxes are connected via a thin gray line to illustrate the connection between the pattern in map view and in polar coordinates. Black X's indicate land, and gray O's indicate ocean. Contours of equal gain (G) and lag (λ) are the labeled concentric circles and lines emanating from the origin, respectively.

slope positive. This relationship is consistent with a mixing model framework wherein a continental component that has high gain and small lag is linearly combined with an oceanic component that has low gain and large lag (Stine et al., 2009).

(2) *Overlap*: Ocean and land values overlap for both lag and gain. Land gridboxes are found with smaller gains and larger lags than ocean gridboxes, and the converse. This overlap is not a result of the coarseness of the grid that we employ because it exists even after excluding all gridboxes that contain mixed ocean and land, as indicated by a one-degree land mask. For instance, the gain of the annual cycle is 47% larger in the East China Sea (27.5°N, 127.5°E) than in central France (47.5°N, 2.5°E), and the lag in southern Nunavut, Canada (62.5°N, 97.5°W) is approximately five days longer than that in the middle of the Greenland Sea (77.5°N, 2.5°W). More generally, 29% of the extratropical gridboxes that we analyze have gains ranging between the minimum land gain and maximum ocean gain, and 10% have lags ranging between the maximum land lag and the minimum ocean lag.

(3) *Offset*: For ocean and land with the same gain, ocean gridboxes have a systematically higher lag. This indicates that the greater lag over ocean is not only due to a higher heat capacity. The differing land and ocean behavior can most easily be seen when focusing on a single latitude band (e.g. Fig. 2.2c). Rather than exhibiting a land-ocean symmetry, the amount of change in lag related to a specified change in gain is dependent on whether gain is increasing or decreasing in the direction of the prevailing Westerlies. Notably, this second-order structure indicates that it is not possible to define a single scalar value for continentality that completely describes the annual Fourier component of the climatological annual cycle.

We seek to explain these three features of extratropical seasonality using as simple a physical model as possible.

2.4 Influence of atmospheric circulation

We aim to quantify the predictive strength of the hypothesis that the large-scale spatial structure of the annual cycle is a consequence of land-ocean coupling through atmospheric circulation. An initial scale analysis supports the idea of a dominant role for atmospheric circulation: an approximate mid-tropospheric wind speed of 10 m s^{-1} and a radiative relaxation time of about a month (Goody and Yung, 1989) suggests that atmospheric temperature properties can be advected on scales of tens-of-thousands of kilometers. This simplistic analysis excludes latent heat processes and vertical adjustment, among other processes, but nonetheless indicates that oceanic influence can be present in continental interiors, and vice-versa. Below we quantify this influence through evaluating the Lagrangian history of air parcels over land and ocean and, subsequently, incorporating this information into an energy balance model in the spirit of North et al. (1981).

2.4.1 Lagrangian trajectory analysis

To evaluate whether the basic features of the annual cycle can be explained as the result of atmospheric circulation coupling land and ocean, we employ the NOAA Air Resources Laboratory HYbrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model (Draxler and Hess, 1997, 1998; Draxler, 1997). We estimate 9,901,440 air parcel back trajectories, where 955 trajectories were initialized at each point on a 2.5° global grid. All back trajectories begin at 1000 meters above ground level, a rough approximation of the top of the boundary layer. Individual trajectories are equally spaced throughout 2006–2010 to obtain a representative annual average distribution of the source regions for each location (Fig. 2.3). Parcels are followed for the 28 days preceding their initialization, or until they go above 200 hPa, approximating the tropopause.

Fig. 2.3 shows three illustrative examples of the spatial distribution of the back trajectories, with parcels initialized in southeast Russia (57.5°N, 117.5°E), coastal California (37.5°N, 122.5°W),

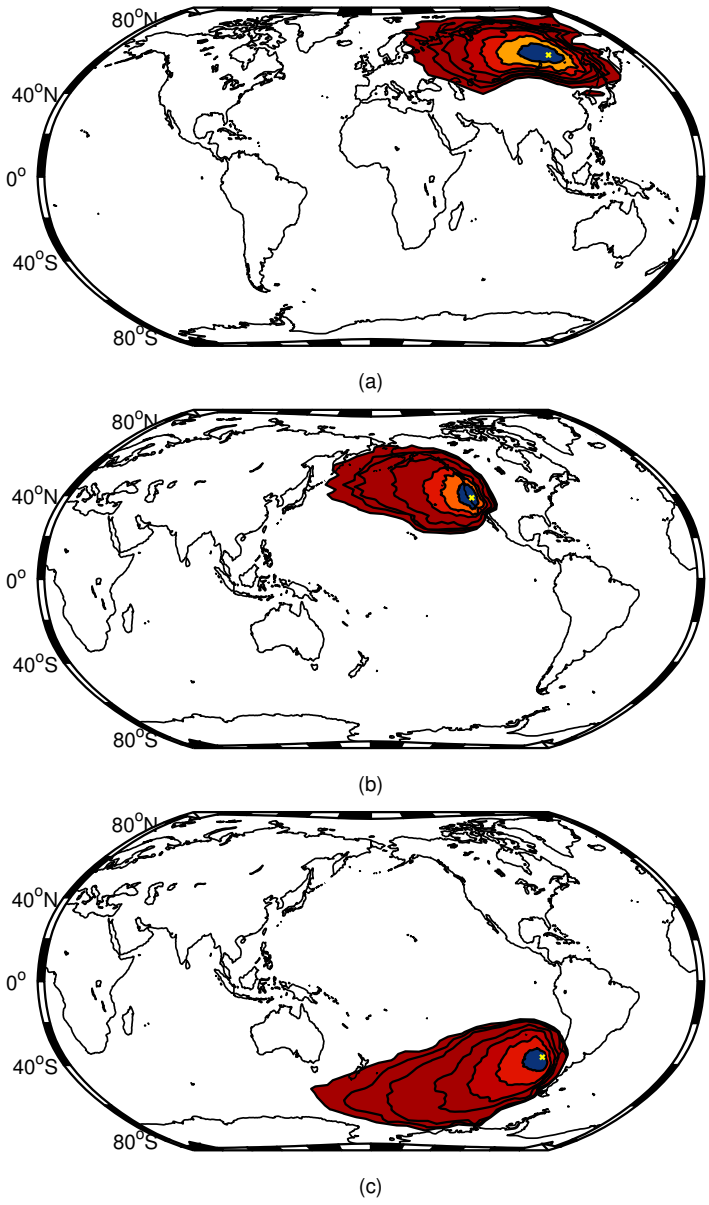

Figure 2.3: Source regions for parcels at the yellow 'x'. Each of the seven consecutively larger areas enclosed by the contour lines contains the closest 99% of parcel locations for each of seven days before initialization of the back trajectories.

and off the coast of Chile (37.5°S, 82.5°W). Each of the seven contours encloses the 99% of air parcels that are closest to the source location in each of the seven days before initialization, indicating the extent of the source region.

2.4.2 Relative Land Influence

Using the ensemble of HYSPLIT parcel trajectories, we define Relative Land Influence, RLI, as the weighted average of land and ocean that a set of air parcels in the atmospheric column have previously passed over,

$$RLI = \frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{t=0}^{28\text{ days}} Z(t)e^{-t/\tau}}{\sum_{t=0}^{28\text{ days}} e^{-t/\tau}}. \tag{2.1}$$

$Z(t)$ takes on values of zero or one depending on whether the parcel is over ocean or land, respectively, at time t, where a three hour resolution is used. τ is the globally averaged radiative relaxation time of the atmosphere, set at 25 days based on consideration of the lower 800 hPa of the atmosphere and using an emission temperature of 255 K (Goody and Yung, 1989). The spatial structure of RLI is largely insensitive to reasonable choices of τ; e.g., the pattern using a 15 day relaxation time is correlated at R = 0.99 to the one presented here.

To classify surface type, we use the NOAA GCOS one-degree land mask[†] and the ISLSCP II one-degree global sea ice concentration dataset[‡]. The land mask is constant in time but the sea ice mask is at monthly resolution. A gridbox is considered to be sea ice if its average concentration is greater than 50% for the month, and sea ice-covered gridboxes are treated as land.

A map of RLI (Fig. 2.4) shows the expected increase in land influence from west to east across continents and the complimentary decrease across ocean basins. The most continental region is in eastern Eurasia. No land gridbox is free from oceanic influence, and the converse. Note

[†]Available at http://data.nodc.noaa.gov/GCOS/software/
[‡]Available at http://daac.ornl.gov/

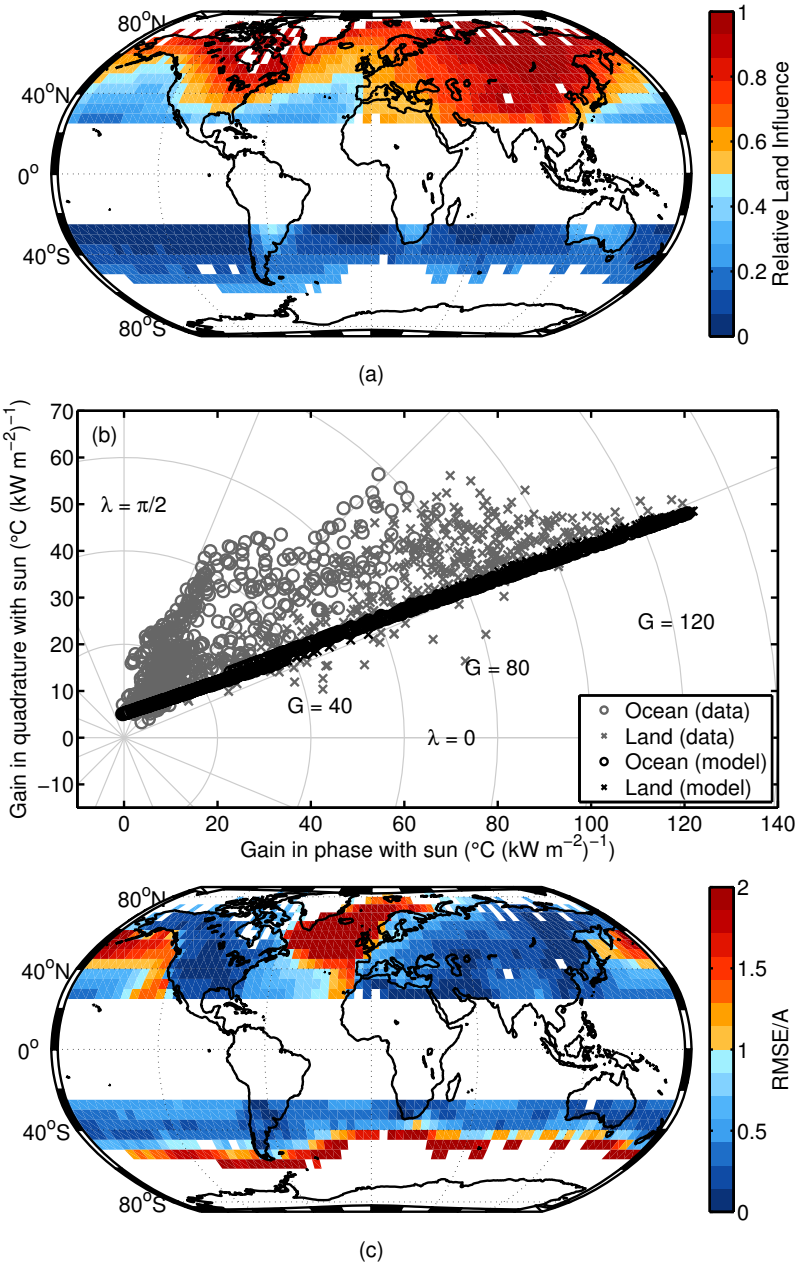

Figure 2.4: (a) Relative Land Influence (RLI) calculated from the HYSPLIT trajectory paths using Eq. (2.1). (b) Gain and lag of the annual cycle from the RLI-weighted end member mixing model (black) and, for comparison, the data (gray). (c) The root mean square error between the model and the data, normalized by the local amplitude of the annual Fourier component in temperature.

that values of RLI show a substantial overlap between land and ocean because ocean gridboxes off the eastern edge of continents have constituent air parcels mostly derived from over land, whereas land gridboxes on the western edge of continents are heavily influenced by oceanic conditions. This behavior is qualitatively consistent with the second feature of the annual cycle data identified in Sec. 2.3.

One means of converting RLI into units that can be directly compared with the gain and lag structure of the annual cycle is to define two time series for representative land and ocean annual cycles, for which we use the highest and lowest gain amongst the observed annual cycle climatologies, respectively. The land end member has a gain of 131 °C (kW m^{-2})$^{-1}$ and a lag of 22 days, whereas the ocean end member has a gain of 5.01 °C (kW m^{-2})$^{-1}$ and a lag of 96 days. It is then possible to convert RLI into a real time series of temperature at each gridbox, $T_{\mathrm{RLI}}(x, y, m) = RLI(x, y)T_{\mathrm{land}}(m) + (1 - RLI(x, y))T_{\mathrm{ocean}}(m)$, where m indicates month, and all temperatures are monthly averages. Albeit simplistic, this mixing model explains 84% of the space-time variance in climatological temperature, with better performance over land than ocean, where it explains 93% and 72% of the variance, respectively.

As can be seen when T_{RLI} is projected into components that are in phase and in quadrature with the solar annual cycle (Fig. 2.4b), the mixing model captures the first two features of the data: linearity and overlap. Linearity is reproduced by construction, whereas overlap results from the aforementioned properties of RLI. The non-unique relationship between gain and lag is, however, not captured by this framework (see Fig. 2.4b), and this issue is returned to in the next section using a more physical model of the annual cycle. Including when the parcel has been in the upper or lower troposphere during its trajectory as an additional predictor adds little to no explanatory power with respect to the annual cycle.

The misfit between our Lagrangian metric of the annual cycle, T_{RLI}, and the observations, T_{HadCRU}, can be measured by the root mean square error, RMSE(x,y) $= (\frac{1}{12}\sum_{\mathrm{m}=1}^{12}(T_{\mathrm{HadCRU,(x,y,m)}} -$

$T_{\mathrm{RLI},(x,y,m)})^2])^{1/2}$, normalized by the amplitude of the annual Fourier component of temperature (Fig. 2.4c). The RMSE relative to the annual amplitude is largest in regions of sea ice cover, indicating the importance of more fully accounting for this third surface type that is neither land nor ocean in future work. The importance of sea ice has recently been demonstrated by Dwyer et al. (2012), who showed that its loss can account for changes in the amplitude and phase of the annual cycle of temperature in general circulation model simulations. Globally, the majority of the misfit results from a mis-assignment of lag in the RLI model. Holding either lag or gain constant at the observed value for all gridboxes and changing the other to that predicted by RLI indicates that the error due to lag misfits is approximately twice that due to gain misfits.

2.5 Advection energy balance model

The Relative Land Influence metric accounts for the amount of time air parcels spend over land and ocean, but does not consider the physical interactions between surface and atmosphere along a trajectory. To explain the third feature of the annual cycle, offset, we turn to an energy balance model.

2.5.1 Model and idealized application

We present a simple energy balance model with east-west resolution of an advecting atmosphere atop a surface layer, permitting calculation of both atmospheric and surface temperature, T_a and T_s. The surface radiative balance is driven by absorbed shortwave radiation, $(1-\alpha)(1-f_{abs})S(t)$, where $S(t)$ is the annual cycle of top-of-the-atmosphere incoming solar radiation at a fixed latitude, α is surface albedo, and f_{abs} is atmospheric absorption. $S(t)$ is time-dependent whereas α and f_{abs} are constant. The surface is either land or ocean, distinguished by differing heat capacities, C_s. These surfaces are coupled to the atmosphere through a "leaky greenhouse" framework,

where the atmosphere absorbs and emits longwave radiation with an emissivity, ε, as well as through a linear diffusivity term, κ, that multiplies the local surface-atmosphere temperature difference and represents non-radiative heat exchange. The atmosphere flows west-to-east with velocity u, and has a heat capacity C_a, that together define a horizontal scale of influence. The advection energy balance model is expressed as,

$$C_s \frac{dT_s}{dt} = (1-\alpha)(1-f_{abs})S(t) + \kappa(T_a - T_s) + \sigma(\varepsilon T_a^4 - T_s^4), \tag{2.2}$$

$$C_a \frac{\partial T_a}{\partial t} = -uC_a \frac{\partial T_a}{\partial x} + f_{abs}S(t) + \kappa(T_s - T_a) + \varepsilon\sigma(T_s^4 - 2T_a^4). \tag{2.3}$$

Note that the model retains no vertical structure in either the atmosphere or ocean and that the temperatures in both can be viewed as representative of their respective mixed layers. The numerical values for all model parameters are in Table 2.1.

We apply the model to a simplified geometry, with constant, closed zonal flow around a latitude band consisting of one continent and one ocean basin of equal width. The modeled annual cycle along the band is shown in Fig. 2.5 and exhibits all three features discussed in Sec. 2.3: linearity, overlap, and offset. We are able to offer some perspective on the origins of the three features based on our simple model as follows:

(1) *Linearity*: The annual cycle in atmospheric temperature results from the relative influence of the land and ocean surfaces. The disparate heat capacities of land and ocean give differing basic annual cycles, in terms of gain and lag, that are then mixed through the advecting atmosphere.

(2) *Overlap*: The advective heat fluxes are of the same order of magnitude as the surface-atmosphere heat fluxes (Fig. 2.6), allowing for the penetration of remote oceanic conditions into continents,

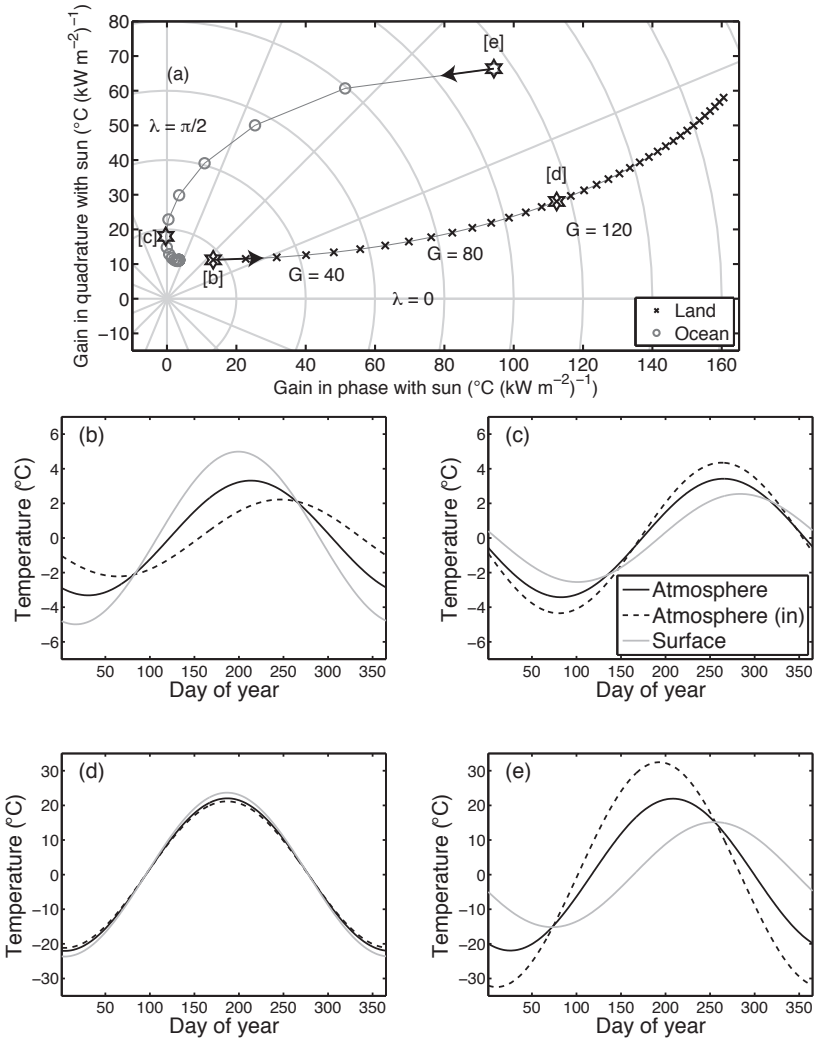

Figure 2.5: (a) Gain and lag of the annual cycle in polar coordinates calculated from the advection energy balance model (Eqs. (2.2) and (2.3)) using an idealized geometry. Black X's indicate land, gray O's indicate ocean. Neighboring gridboxes of the same surface type are connected by a thin gray line, and arrows indicate west-to-east movement across longitude. (b)-(e) Modeled annual cycles for the atmosphere atop land (b, d) and ocean (c, e) at the locations marked by black stars in (a). 'Atmosphere (in)' is the annual cycle of the atmosphere for the gridbox immediately to the west. The amplitude is the same in each row, but the ocean points exhibit a greater lag than the land points by 52 (b, c) and 21 (d, e) days.

and vice-versa.

(3) *Offset*: Absent atmospheric circulation in the model, land surfaces and the overlying atmosphere would have an annual cycle in temperature that lagged solar radiation by about a month, whereas ocean surfaces would lag the atmosphere by about three months and the overlying atmosphere by almost as much. Transport of heat by the atmosphere, represented using a constant zonal velocity, has the counter-intuitive effect of making the lag over land smaller and that over ocean greater. This broadening of land-ocean lags appears to be at the heart of the offset, because it almost entirely disappears if the surface annual cycle is prescribed to be equal to that attained in the no atmospheric circulation case.

The increased lag over ocean can be understood through considering the influence of atmosphere-surface heat fluxes relative to solar fluxes. The annual cycle in atmospheric temperature advected off the eastern edge of the continent into the western edge of the ocean basin leads to atmosphere-surface heat fluxes having a larger magnitude than direct radiative absorption. These fluxes of heat from the atmosphere generally lag behind the solar forcing, leading to a greater ocean lag than would be the case without atmospheric advection. As the continental influence decreases eastward, the solar heat fluxes become dominant and seasonality approaches that obtained with no atmospheric circulation.

The situation differs slightly over land. The annual cycle of the atmosphere advected onto the western edge of the continent primarily acts to damp the annual cycle over land, which decreases both its gain and lag. This is in analogy with a linear system driven by periodic forcing, e.g., $C\frac{dT}{dt} = A\exp[\imath\omega t] - \lambda T$, where the lag of T relative to the forcing also decreases with greater damping, $\phi = \frac{1}{\omega}\arctan[\frac{\omega C}{\lambda}]$. Thus, the oceanic influence acts to decrease the lag over land. These two effects combine to produce the lag offset in the model (e.g. Fig. 2.5b-e).

Examination of the modeled heat fluxes (Fig. 2.6) gives more insight into the controls on the

annual cycle. The advective and surface-atmosphere heat fluxes are large and in opposition. The shortwave forcing from atmospheric absorption is small compared to the other heat fluxes, only becoming comparable in the eastern interiors of continents and ocean basins, although it becomes a more dominant heating source when considering zonal mean quantities (Donohoe and Battisti, 2013). The tendency, which is proportional to but in quadrature phase with the annual cycle in temperature, is the residual of these large heat fluxes, and is generally an order of magnitude smaller than either the advective or surface-atmosphere heat fluxes. Furthermore, because the tendency is the sum of multiple heat fluxes in addition to solar forcing, the phase of air temperature is able to have a more-than-90-degree phase lag from the solar forcing.

We find the structure of the modeled annual cycle to be largely insensitive to reasonable variations in model design. For instance, although the majority of planetary albedo is due to atmospheric rather than surface reflection (Donohoe and Battisti, 2011), changing this partitioning does not affect the result in the sense that the resulting amplitude and phase structure are correlated with the base case at R>0.99. Including a seasonally-variable emissivity and varying albedo spatially or seasonally based on observations from the Earth Radiation Budget Experiment (Barkstrom, 1984) also has negligible influence on our results. Varying the coupling constant, κ, between 10 and 50 W m^{-2} K^{-1} has a slightly larger effect, particularly on the rate of change of lag across ocean basins, but the general structure is retained, and the gain and lag between the base case and these other choices for κ are correlated with R $\geq$ 0.97 and R $\geq$ 0.89, respectively.

The zonal flow model can also be applied to the land-ocean configuration of a specific latitude band, for which we choose 45-50°N, and compared to the observed annual cycle there (Fig. 2.7c, d). The model captures the first-order structures in gain and lag over ocean and land but does not reproduce differences such as those between North American and Eurasia. The misfit between the model and the observations can likely be ascribed to two obvious and related over-

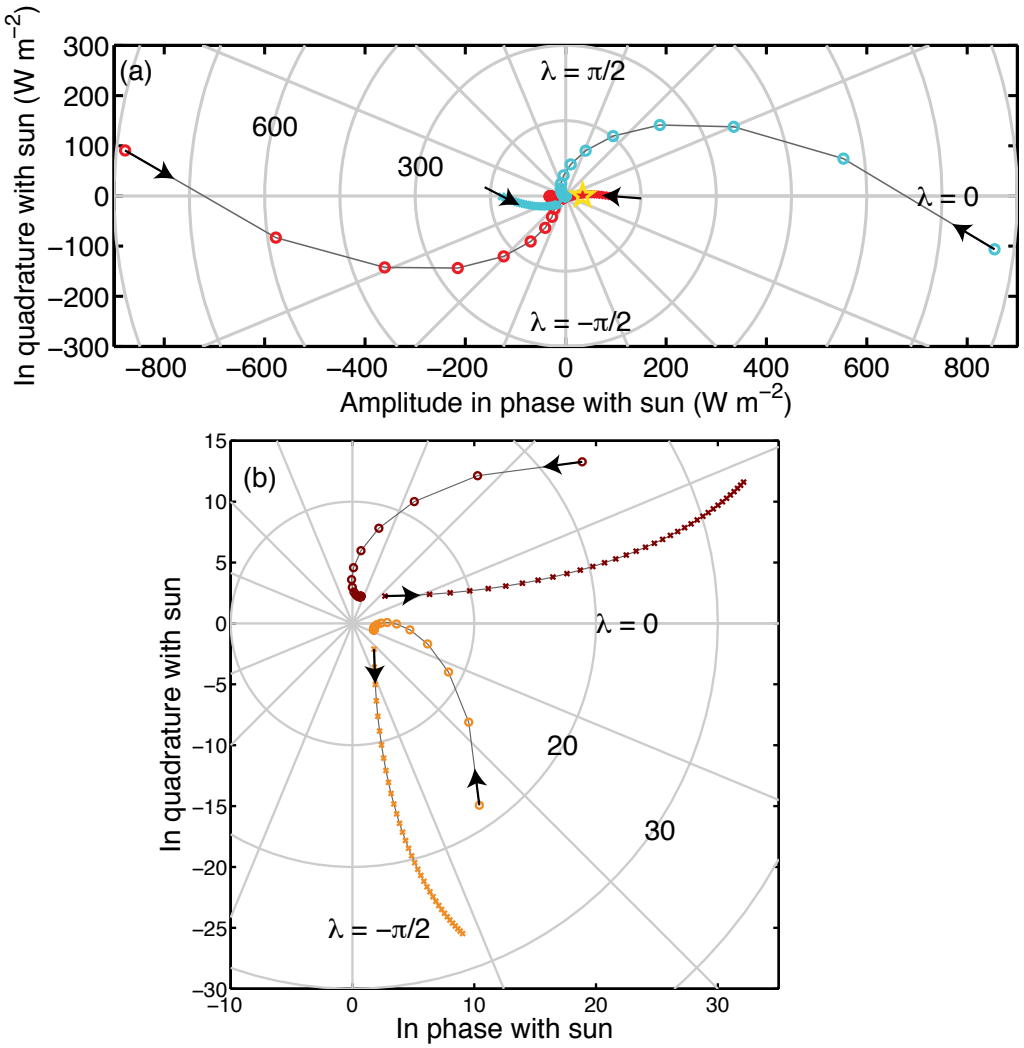

Figure 2.6: The amplitude and phase of the heat fluxes to the atmosphere in the advection energy balance model shown in polar coordinates. (a) Energy flux amplitude and phase broken down into components: advective ($-uC_a\frac{\partial T_a}{\partial x}$, light blue); sensible, latent, and longwave ($\kappa(T_s - T_a) + \varepsilon\sigma(T_s^4 - 2T_a^4)$, red); and solar ($f_{abs}S(t)$, yellow star). Mathematical representations correspond to Eqn. (2.3). X's indicate land, O's indicate ocean, neighboring gridboxes of the same surface type are connected by a thin gray line, and arrows indicate west-to-east movement across longitude. (b) Net heating (orange, Wm^{-2}) and the normalized temperature response (dark red, °C (kW m^{-2})$^{-1}$). The normalized temperature response is multiplied by a value of 0.2 to plot on the same axes as net heating.

simplifications in the zonal flow model: the prescription of constant velocity for the full latitude band and the use of the fixed land-ocean configuration at that latitude band, as opposed to taking into account differences in surface type that would result if the circulation was not purely zonal.

2.5.2 Incorporation of HYSPLIT results

To explore the idea that model mismatch arises because of deviations from pure zonal flow, we incorporate HYPSPLIT parcel trajectories. Consider a single trajectory that arrives at a single gridbox of the zonal model, for which we are interested in estimating the annual cycle. We construct a separate model to represent the 28 days preceding its arrival at the gridbox, using the land-ocean sequence that the trajectory follows as the surface type boundary condition in the model. Land and ocean boxes are determined from position using the GCOS one-degree land mask, and model velocity, u, is specified as the average zonal velocity of the trajectory. All other model parameters are in Table 2.1 and are identical to those used in the idealized simulation in the previous section. To capture the full seasonal cycle, the model is integrated to steady seasonality using solar forcing, $S(t)$ (as in Eqns. (2.2) and (2.3)). The initial condition for the air parcel comes from the closed-loop zonal flow model after also integrating it to a steady annual cycle under the same model parameters.

To sample a larger fraction of the distribution of parcel source regions (e.g. Fig. 2.3), we integrate the model using 642 different trajectory paths for each of the 72 five-degree gridboxes in the latitude band between 45-50°N. Trajectories are equally spaced across the time period 2006–2010 and are a subset of those used to calculate Relative Land Influence. The energy balance model outputs across each trajectory are then averaged to create a climatology for comparison to observations.

The model climatology captures 94% of the space-time variance of the monthly temperature

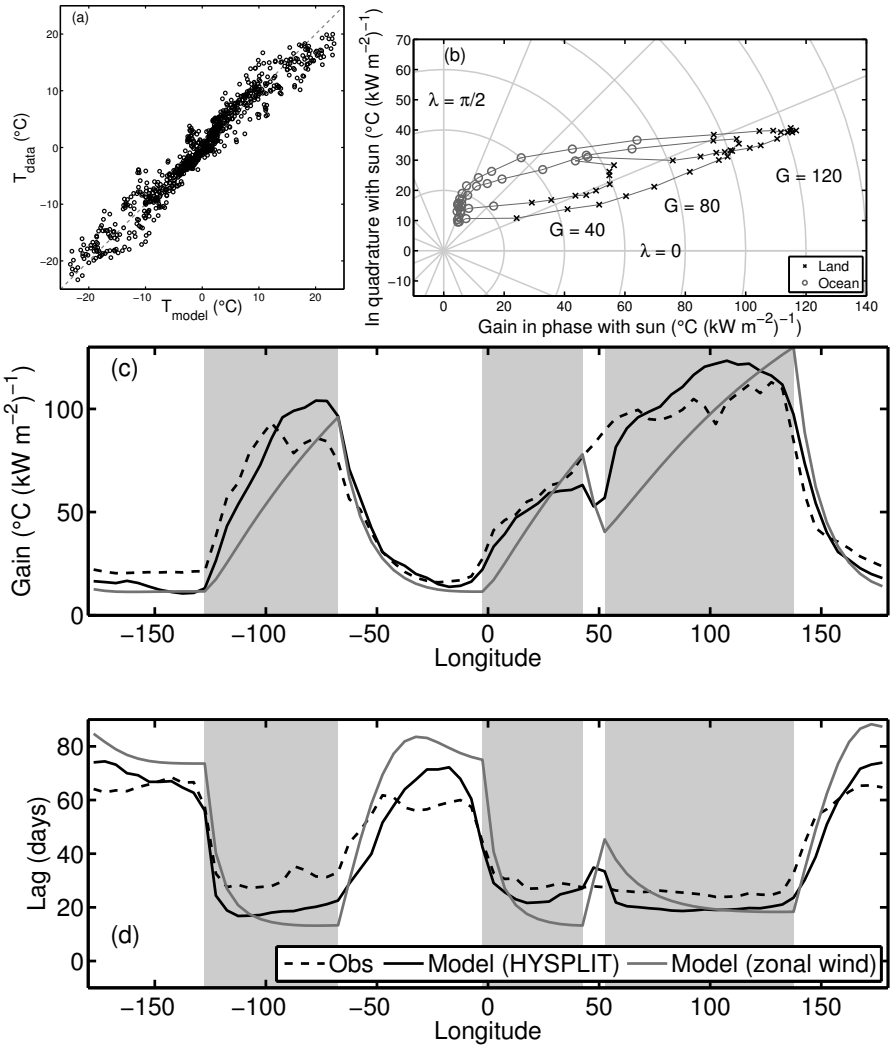

Figure 2.7: (a) Monthly temperature anomalies in the latitude band 45-50°N from the advection model driven by HYSPLIT trajectories versus observations. (b) The gain and lag of the modeled annual cycle in polar coordinates showing land (X's) and ocean (O's) boxes. Neighboring gridboxes are connected via a thin gray line. (c) The gain of the modeled annual cycle across longitude at 45-50°N using a zonal wind (gray) and with the inclusion of the HYSPLIT trajectory information (black), as compared to the observations (dashed). Land regions are indicated by shading. (d) Similar to (c) but for lag.

Table 2.1: Parameter values for the advection model. Zonal wind velocity is the average across the HYSPLIT trajectories at 45-50°N; see Sec. 2.52.5.2.

Name	Var. name	Value	Units
Zonal wind velocity	u	6.1	$\mathrm{m\ s^{-1}}$
Emissivity	ε	0.8	
Albedo	α	0.3	
Coupling coefficient	κ	50	$\mathrm{W\ m^{-2}\ K^{-1}}$
Shortwave absorption	f_{abs}	0.18	
Ocean heat capacity	C_o	75	m.w.e.
Land heat capacity	C_l	1	m.w.e.
Atmosphere heat capacity	C_a	1	m.w.e.
meter-water-equivalent	m.w.e.	4.18×10^6	$\mathrm{J\ m^{-2}\ K^{-1}}$
Latitude		47.5°N	

climatologies in the latitude band considered (Fig. 2.7a), including the more rapid changes in lag across ocean than land, the rapid decrease in gain at the western edge of the Pacific Ocean, and the more rapid increase in gain across North America than Eurasia. The energy balance model that incorporates the HYSPLIT information captures basin-scale features substantially better than is the case when a constant zonal flow is assumed (Fig. 2.7c,d).

The primary model-data misfit arises from the model producing too large a lag in the interiors of ocean basins, and we speculate this is due to lack of model mixed layer dynamics (e.g. Laepple and Lohmann, 2009). One might also have suspected that major features of the annual cycle would be controlled by processes such as temporal and spatial variability in cloudiness (Weaver and Ramanathan, 1997), intra-annual variability in atmospheric circulation patterns (Stine and Huybers, 2012), moisture availability over land (Stine et al., 2009), or higher order Fourier components of the annual cycle (Legates and Willmott, 1990). These additional processes presumably also contribute, but the present results indicate that the primary explanation for spatial variability in the the extratropical annual cycle structure is the land-ocean contrast, mediated by the circulation of the atmosphere.

2.6 Conclusion

The extratropical annual cycle is characterized by (1) a first-order linear structure consistent with an end-member mixing framework, (2) a land-ocean overlap in gain and lag, and (3) a systematically higher lag over ocean than land. Previous analysis identified feature (1) as due to relative land and ocean influences (Stine et al., 2009). We further find that accounting for the source region of air parcels through a Lagrangian approach captures feature (2), and that combining atmospheric advection with an energy balance model captures feature (3). Each component of our simple model appears necessary for describing these three features, suggesting that it is, in

a sense, a minimal representation. Atmospheric coupling between land and ocean gives rise to a linear structure between gain and lag when plotted in polar coordinates, the directional nature of atmospheric advection leads to overlap in gain and lag, and the time-dependence of atmospheric interaction with land and ocean produces a non-unique relationship between gain and lag.

These results support previous findings and have several further implications. Changes in continental configuration over geological timescales are expected to have far-reaching effects on the annual cycle, as the air parcel source region for a given location will become more continental or oceanic. Loss of sea ice (e.g. Stroeve et al., 2012) is also expected to moderate nearby annual cycles due to an increase in effective heat capacity (Dwyer et al., 2012). Further work on the parameterization of sea ice in this framework may be helpful in understanding the connection between sea ice loss and changes in the annual cycle. Finally, the coupling of land and ocean via mean circulation patterns is not specific to annual period variability, and this framework may also provide insight into the magnitude and timing of changes at synoptic and interannual timescales.

3

On using the seasonal cycle to interpret extratropical temperature changes since 1950

The investigation of the seasonal cycle in the previous chapter provided a framework with which to understand the observed spatial pattern of seasonal temperature variability in the extratropics. In this chapter, I test the hypothesis that the same mechanism – heat transport by the atmosphere between the low heat capacity land and high heat capacity ocean – can explain a large fraction of modeled and observed multi-decadal temperature changes. The chapter is reproduced from its published form: McKinnon, K. A. and P. Huybers, 2014: On using the seasonal cycle to interpret extratropical temperature changes since 1950. *Geophysical Research Letters*, **41 (13)**, 4676–4684.

3.1 Abstract

Extratropical near-surface air temperature variability is explored on three different timescales: the seasonal cycle, observed changes in temperature since 1950, and the equilibrium response

to increasing CO_2 in an atmospheric general circulation simulation with fixed sea surface temperatures. Exploration is undertaken using an energy balance model (EBM) that parameterizes advective land-ocean heat fluxes. The EBM is tuned only to the climatological seasonal cycle, yet captures 47% of the variability in observed multi-decadal temperature changes in the extratropics and 78% of the variability in the equilibrated model simulation. The subseasonal timescale of atmosphere-surface heat fluxes explains, at least in the context of this EBM, the ability to infer patterns of multi-decadal change using information primarily drawn from the seasonal cycle.

3.2 Introduction

The importance of ocean heat uptake in influencing patterns of temperature variability is well established. These patterns include the amplitude and phase of the seasonal cycle of surface temperature (North and Coakley, 1979a; North et al., 1981; McKinnon et al., 2013), internally generated surface temperature variability (Kim and North, 1991), and transient surface temperature changes in response to increased radiative forcing (Stouffer et al., 1989; Manabe et al., 1991; Kim et al., 1992). The common influence of ocean heat uptake amongst seasonal, interannual, and decadal timescales of temperature variation suggests that the observed response at one timescale may inform about the others. Indeed, the amplitude of the seasonal cycle in surface temperature has previously been used as a predictor for equilibrium climate sensitivity across an ensemble of general circulation models (Knutti et al., 2006), and the magnitude of the seasonal cycle is a good predictor of the magnitude of decadal variability in regional surface temperatures (Huybers and Curry, 2006).

That there are common controls on surface temperature variability for the seasonal cycle and longer timescales has previously been demonstrated using an energy balance model (EBM), wherein atmospheric heat transport was parameterized as a diffusive process (Kim and North,

1991; Kim et al., 1992). Here, we revisit this topic from the standpoint of a longer instrumental record and an EBM that permits for analytical solutions. Furthermore, as opposed to the diffusive and isotropic representation of horizontal heat fluxes used in the earlier EBM studies, heat transport is parameterized based on mean atmospheric circulation, allowing for anisotropy due to the climatological wind patterns.

The EBM is written in a minimalist manner so as to isolate the influence of advection of heat between land and ocean on surface temperature patterns. Excess heat not realized as a temperature increase is stored in the ocean. While this approach ignores many additional important processes, such as changes in meridional heat transport, spatial structure of perturbative radiative forcing, and spatial and temporal variability in mixed layer depth, it allows us to isolate the effect of ocean heat storage on temperature variability. The model is presented and minimally tuned to the seasonal cycle (Sec. 2), shown to give accurate predictions at longer timescales using equilibrium results from a general circulation model with fixed sea surface temperatures (Sec. 3), and then compared to observed temperature variations over recent decades (Sec. 4). The skill – or lack thereof – of the EBM is used a metric to assess the importance of advective heat fluxes and ocean heat uptake in influencing surface temperature as compared to other climate processes that are excluded from the model.

3.3 Amplitude of the seasonal cycle in surface temperature

3.3.1 EBM formulation

The near-surface air temperature anomaly (hereafter simply 'temperature'), $T_{i,j}$, is modeled as a mixture between land and ocean temperatures, weighted by a mixing parameter, $m_{i,j}$:

$$T_{i,j} = m_{i,j} T_{\text{land}} + (1 - m_{i,j}) T_{\text{ocean}}. \tag{3.1}$$

'Land' and 'ocean' refer to the end-member cases where the heat flux divergence is zero, so the end-members are in local radiative equilibrium. The indices i, j indicate spatial dependence. These end-members are modeled using two linearized energy balance models,

$$C_{\text{land}} \frac{dT_{\text{land}}}{dt} = R + \lambda T_{\text{land}}, \tag{3.2}$$

$$C_{\text{ocean}} \frac{dT_{\text{ocean}}}{dt} = R + (\lambda - \kappa) T_{\text{ocean}}, \tag{3.3}$$

where C is heat capacity, with $C_{\text{land}} < C_{\text{ocean}}$; R is a perturbative radiative forcing; λ is a feedback parameter; and κ is a linear diffusivity that controls the rate of heat uptake by the deep ocean. Parameters lacking subscripts are constant in space. Ocean temperature, T_{ocean}, represents the mixed layer, and the deep ocean is assumed to have infinite heat capacity for the scenarios considered here.

Energy balance equations similar to Eqns. (3.2) and (3.3) are more typically applied with respect to the global mean (e.g. Wigley and Schlesinger, 1985; Raper et al., 2002; Gregory and Forster, 2008); here, the end-members are reflective neither of the global mean nor of the average land and ocean temperatures. Instead, each model equation captures the temperature response for land or ocean if each was controlled only by its heat capacity, the feedback parameter, and—in the case of the ocean—heat uptake by the interior.

Local temperatures, $T_{i,j}$, are then expressed as,

$$C_{\text{land}} \frac{dT_{i,j}}{dt} = R + \lambda T_{i,j} - H_{i,j}, \tag{3.4}$$

$$H_{i,j} = (1 - m_{i,j}) \left[(C_{\text{ocean}} - C_{\text{land}}) \frac{dT_{\text{ocean}}}{dt} + \kappa T_{\text{ocean}} \right]. \tag{3.5}$$

In this formulation, temperature is modeled as a fast, land-like response to forcing that is damped by the horizontal, dominantly zonal heat flux divergence, $H_{i,j}$, due to atmospheric advection.

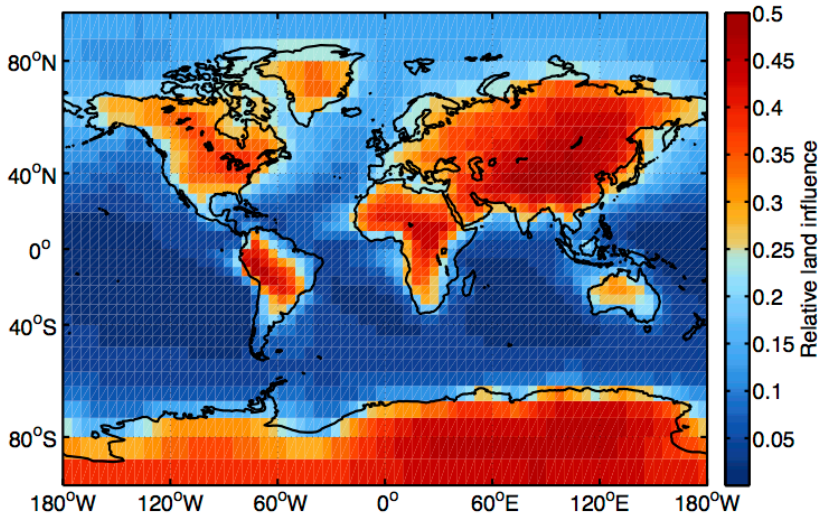

Figure 3.1: A map of the Relative Land Influence field

The magnitude of damping is a function of Relative Land Influence, $m_{i,j}$. Relative Land Influence is a scalar measure of the relative magnitude of atmospheric heat fluxes from land as compared to ocean at a given location, and is calculated from an ensemble of Lagrangian air parcel back trajectories. The parcel trajectories are simulated using HYSPLIT (Draxler and Hess, 1997, 1998; Draxler, 1997) and the meteorological fields from NCEP-DOE reanalysis (Kanamitsu et al., 2002). Values from zero to one grade from purely marine to purely continental influence (Fig. 3.1). Further details regarding Relative Land Influence can be found in McKinnon et al. (2013) and the auxiliary material. The model is written such that the net diverged heat is implicitly stored in the mixed layer (first term on the right-hand-side of Eqn. (3.5)) and deep ocean (second term on the right-hand-side of Eqn. (3.5)).

3.3.2 Application to the seasonal cycle

To model the extratropical seasonal cycle, we prescribe a sinusoidal forcing function, R, with annual frequency, ω. The seasonal cycle is negligibly influenced by the deep ocean in the EBM, so the value of κ in Eqn. (3.3) is set to zero for simplicity. Consistent with this assumption, the seasonal timescale results are insensitive to the choice of κ. The amplitude and phase of R vary in latitude, so we examine the annual component of temperature normalized by that of the forcing, i.e., the gain (Stine et al., 2009). Annual components of temperature and forcing are calculated as the magnitude of the once per year Fourier component (Thomson, 1995). Eqns. (3.4) and (3.5) yield a gain, $G_{i,j}$, of,

$$G_{i,j} = \left[\frac{\lambda^2 + \omega^2 (C_{\text{land}}(1 - m_{i,j}) + C_{\text{ocean}} m_{i,j})^2}{(\lambda^2 + C_{\text{land}}^2 \omega^2)(\lambda^2 + C_{\text{ocean}}^2 \omega^2)} \right]^{1/2}. \tag{3.6}$$

A related expression is obtained for the phase offset, or lag, between the forcing and the temperature response. Heat capacities, C_{land} and C_{ocean}, are respectively set to 2 and 76 meters of water equivalent, consistent with previously used values (McKinnon et al., 2013). All parameter values are given in Table 3.1.

To estimate gain from observations (Fig. 3.2a), we use the Hadley Centre monthly temperature climatology at 5°x5° resolution (Morice et al., 2012), scaled by the seasonal forcing. The forcing is the sum of two annual period terms: (1) the top-of-the-atmosphere (TOA) solar forcing, scaled by a co-albedo, and (2) the seasonal cycle in meridional heat flux convergence. We explicitly include meridional heat flux convergence in the forcing because it is not otherwise parameterized in the model, but acts to damp the seasonal cycle in heating (Donohoe and Battisti, 2013), and, therefore, should also be accounted for when estimating gain. Note that the availability of observationally-based estimates of seasonal heat flux convergence and our approach of using the

Table 3.1: Model parameters for the EBM in the three different scenarios. Parameters inferred from fitting the model to the data, via minimizing the mean squared deviations across gridboxes, are shown with their 95% confidence intervals.

Name	Symbol	Value	Units
Land heat capacity	C_{land}	2	m.w.e.
Ocean heat capacity	C_{ocean}	76	m.w.e.
Angular frequency	ω	2π (year)$^{-1}$	year^{-1}
Radiative forcing (2x CO_2)	R_{2x}	3.48	W m^{-2}
Radiative forcing rate (before 1982)	β_1	0.06	W m^{-2} (decade)$^{-1}$
Radiative forcing rate (after 1982)	β_2	0.53	W m^{-2} (decade)$^{-1}$
Meters-water-equivalent	m.w.e.	4.18×10^6	J m^{-2} K^{-1}
Scaling parameter for Relative Land Influence	k	0.52 ± 0.01	(unitless)
Seasonal cycle feedback parameter	$\lambda_{\text{seasonal}}$	-2.48 ± 0.15	W m^{-2} K^{-1}
Fixed SST equilibrium feedback parameter	$\lambda_{\text{AM2.1}}$	-2.09 ± 0.09	W m^{-2} K^{-1}
Multi-decadal change feedback parameter	$\lambda_{\text{HadCRUT4}}$	-0.57 ± 0.04	W m^{-2} K^{-1}
Linearized ocean diffusivity	κ	7.69 ± 2.02	W m^{-2} K^{-1}

seasonal cycle to tune the EBM is distinct from later applications of the EBM, when it is used to predict anomalies in temperature at longer timescales and anomalies in heat flux convergence are not included. Both the TOA solar forcing and the co-albedo are from the CERES-EBAF climatology product (Loeb et al., 2009). Meridional heat flux convergence is from the analysis of Fasullo and Trenberth (2008). The amplitude and phase of the forcing varies as a function of latitude but is constant in longitude, such that all zonal structure in the EBM is due to structure in Relative Land Influence.

Analysis is restricted to the extratropics (10-60° in both hemispheres) due to weak once-per-year solar influence relative to other processes in the deep tropics (Sobel et al., 2001) and the combination of fewer data and strong local feedbacks (Hwang et al., 2011) in the high latitudes. Locations where less than 50% of the variance in the monthly climatology is explained by an annual period sinusoid are also excluded from all analyses. The remaining domain covers 64% of the global area. Estimation of the annual period components of temperature and forcing using daily rather than monthly values gives essentially the same amplitude (McKinnon et al., 2013).

Two free scalar parameters are jointly adjusted to fit the EBM to the seasonal cycle. The first, k, scales the range of Relative Land Influence. Specifically, while the spatial pattern of Relative Land Influence is fixed from the analysis of the Lagrangian back trajectory ensemble (see auxiliary material), it spans an arbitrary range between zero and one. The scalar parameter, k, translates the unscaled values of Relative Land Influence, $m^*_{i,j}$, from Eqn. (A1) to scaled ones, $m_{i,j}$, as $m_{i,j} = km^*_{i,j}$. The scaled values allow for a proper allocation of local heat fluxes between outgoing longwave radiation, $\lambda T_{i,j}$, and heat flux divergence, $H_{i,j}$ (see Eqns. (3.4) and (3.5)). The second free parameter is the feedback parameter, λ. The two parameters are jointly constrained (see Fig. 3.4b) through minimizing the sum of squared deviations across gridboxes between the EBM and observations. Operationally, we use the MATLAB non-linear fitting toolbox, which utilizes the Levenberg-Marquardt algorithm for minimization. The value of k is fixed to the value found

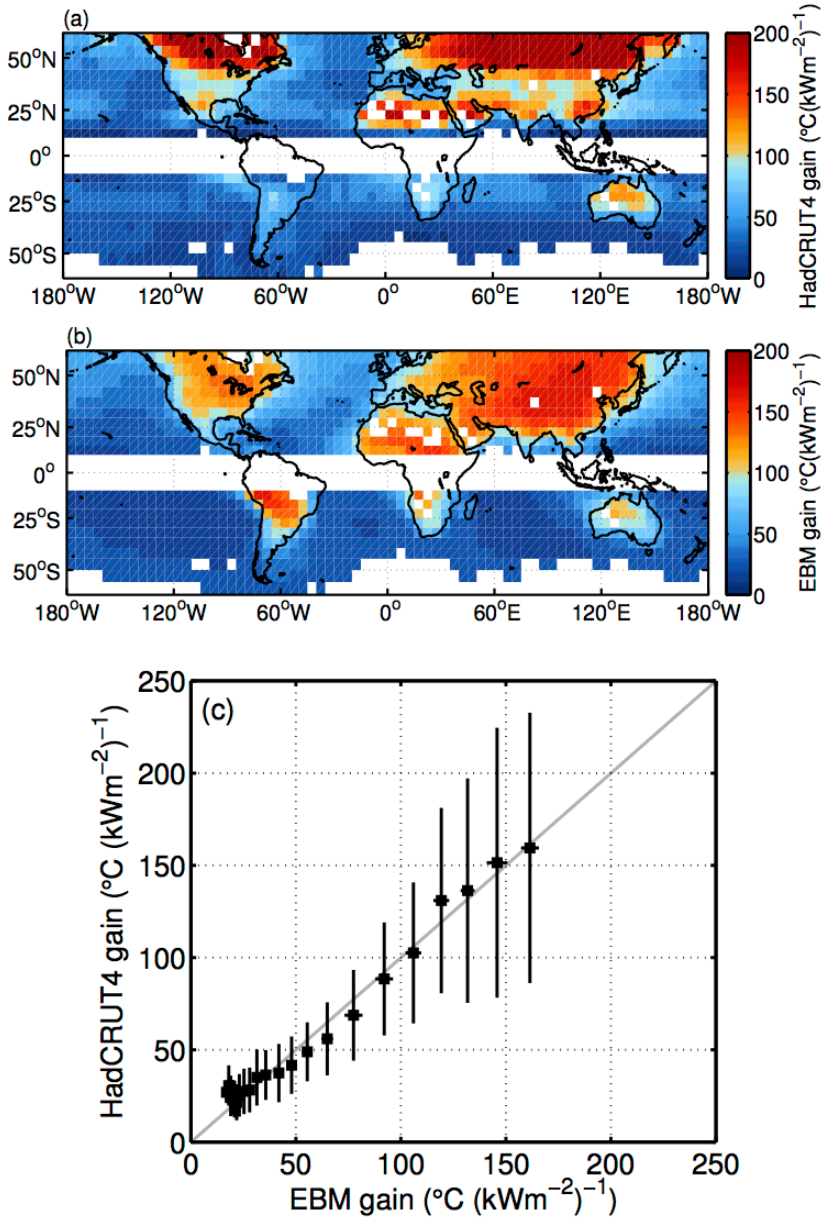

Figure 3.2: EBM fit to the seasonal cycle. (a) Gain of the observed seasonal cycle from the HadCRU climatology. (b) Gain predicted by the EBM (c) Observed gain (vertical axis) compared to modeled gain (horizontal axis) after binning. The squares show the mean of the bin, whereas the whiskers show one standard deviation across the gridboxes within the bin. Number of bins is assigned as a function of number of gridboxes, n, according to the Rice Rule as $\lceil 2n^{1/3} \rceil$, and there are approximately the same number of gridboxes in each bin.

here in all subsequent analyses.

The EBM fit to the seasonal cycle indicates a feedback parameter value of -2.48 ± 0.15 $\mathrm{Wm^{-2}K^{-1}}$, where a more negative λ is associated with a smaller equilibrium temperature response to a given forcing. Uncertainties are 95% confidence intervals based on model fit to the data and do not include the effects of error in model choice or in the observational data. The gains predicted from the EBM and observed in the Hadley Centre data have a correlation of 0.81 across space (Fig. 3.2). This correlation indicates that the model is largely capable of reproducing the seasonal climatology in the selected domain, but because we approach the seasonal cycle as a tool for model tuning, we do not make any claims as to the significance of the correlation.

3.4 An equilibrium experiment

To explore whether the seasonal results are also informative for longer timescale responses, we examine the annual mean temperature changes (Fig. 3.3a) from a set of GFDL AM2.1 general circulation model simulations (The GFDL Global Atmospheric Model Development Team, 2004). The simulations begin in equilibrium with preindustrial CO_2 concentrations. Sea surface temperatures (SSTs) are fixed to their original equilibrated values, then CO_2 concentrations are doubled, and the model is integrated until it attains a new equilibrium, which occurs almost entirely within a model year. The fixed SST experiment allows us to examine temperature changes due to increasing CO_2 in the end-member case where all temperature change is confined to the land and atmosphere, because the ocean adjustment timescale is effectively infinite. This maximizes the signal of differential land and ocean warming in response to increased radiative forcing.

Temperature change is calculated as the difference in annual average temperature between the control and the 2x CO_2 simulations, where 60 years of model output from three different simulations are used in the control average, and 40 years from two different simulations are used in the

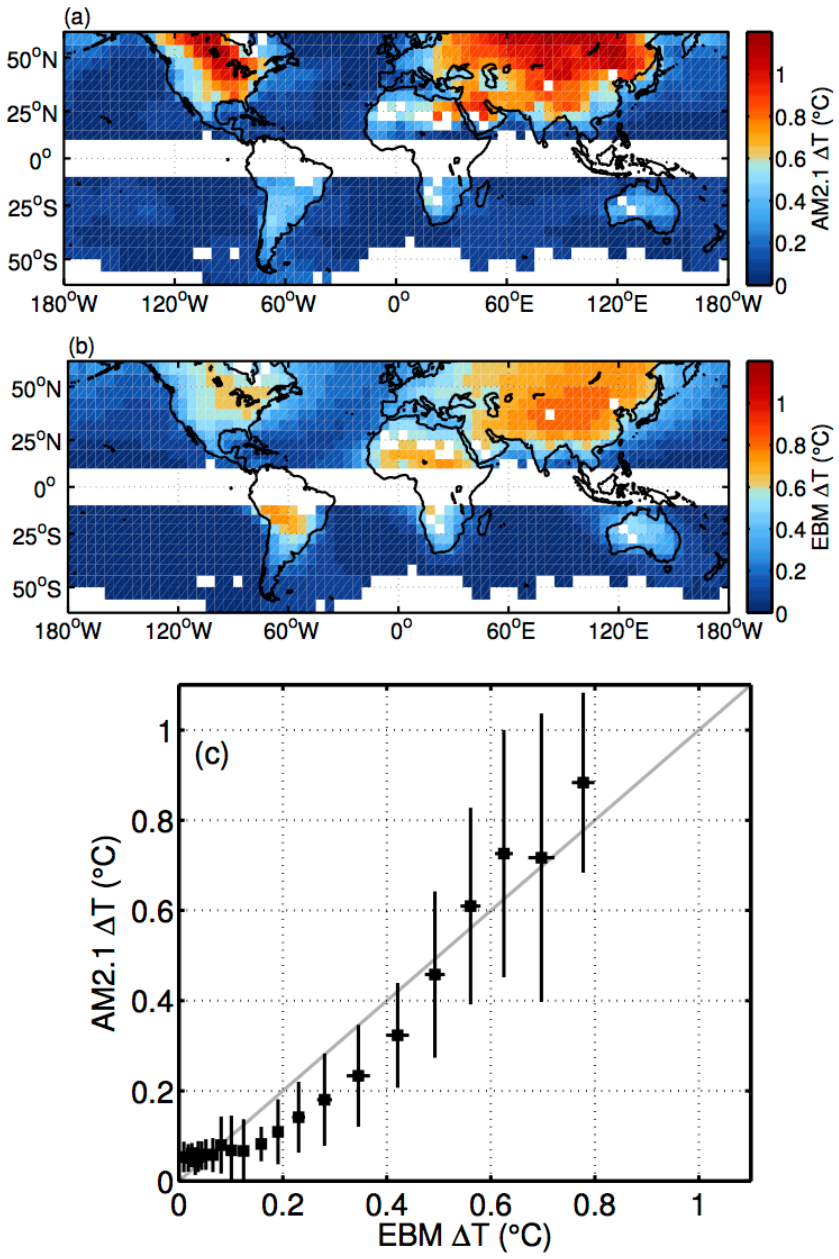

Figure 3.3: Results from fitting the EBM to the output from the GFDL AM2.1 model experiment, with fixed SSTs. The figure includes (a) the map of the AM2.1 temperature change; (b) the EBM best-fit to the field; and (c) the relationship between the two fields, shown in a box and whisker plot.

2x CO_2 average. We re-grid the AM2.1 model output to the 5° resolution of the Hadley temperature data.

Representation of the equilibrium near-surface air temperature change, $\Delta T_{i,j}$, due to land-ocean heat fluxes parameterized in the EBM, is particularly simple for a fixed SST model simulation. In this case, all changes in temperature are related to warming over land; warming over the ocean is due to advection of heat from the land, as represented by Relative Land Influence. Using Eqns. (3.4) and (3.5), this temperature change can be written as,

$$\Delta T_{i,j} = -\frac{R_{2x}}{\lambda} m_{i,j}. \tag{3.7}$$

R_{2x} is the radiative forcing due to CO_2 doubling, set at 3.48 Wm^{-2} (Held et al., 2010). Although R_{2x} is expected to have latitudinal structure, we use a single global-mean value for the purpose of isolating the contribution of advective land-ocean heat fluxes to the pattern of warming. Unlike the seasonal cycle analysis, meridional heat flux convergence anomalies are not included or parameterized. The values of Relative Land Influence, $m_{i,j}$, are fixed from the seasonal cycle analysis, and the only adjustable parameter in the model is λ. As can be seen in Eqn. (3.7), temperature change (Fig. 3.3b) is expected to be linear in Relative Land Influence and have zero warming for zero Relative Land Influence. The feedback parameter can be calculated directly from the best-fit line through the origin between $m_{i,j}$ and $\Delta T_{i,j}$, giving $\lambda = -2.09 \pm 0.09$ $\mathrm{Wm}^{-2}\mathrm{K}^{-1}$.

The EBM predicts the majority of the equilibrium temperature structure in the AM2.1 model simulation (r = 0.87, p-value < 0.01, Fig. 3.3). Model skill is not simply due to the land/ocean contrast in warming as model predictions are independently significant over land (r = 0.66, p-value < 0.01) and ocean (r = 0.71, p-value < 0.01). Significance of these correlations is estimated using a Monte Carlo method whereby surrogate fields of Relative Land Influence are obtained using the Iterative Amplitude Adapted Fourier Transform algorithm (Venema et al., 2006), which

involves phase randomizing the original field using a two-dimensional Fourier transform. This algorithm preserves the spatial autocorrelation structure (Theiler et al., 1992) as well as the distribution of the original field. We then compute the correlation between the EBM estimate of $\Delta T_{i,j}$ based on the surrogate field, and the $\Delta T_{i,j}$ produced in the AM2.1 simulation. This process is repeated 10,000 times to provide a null distribution against which the correlation obtained from using the actual Relative Land Influence field is compared. We present one-sided p-values because a positive relationship between the observed and modeled temperature change is expected.

The primary residual between the EBM and AM2.1 is that the EBM shows less warming as a function of latitude, overestimating warming at lower latitude and underestimating it at higher latitudes. The cross-over point occurs between 40-45° in both hemispheres. On average, the EBM overestimates the warming by 0.06°C (0.02°C) at latitudes below this cross-over point, and underestimates it by 0.07°C (0.02°C) above this point in the Northern (Southern) Hemisphere. Additionally, the ratio of land to ocean warming is greater in the AM2.1 simulation (6.7) than in the EBM (4.8), suggesting that the EBM may overestimate the magnitude of heat fluxes from land to ocean. Finally, the warming over central South America is overestimated. Misfit in this region is also apparent for the seasonal cycle case and the multi-decadal predictions considered in the next section. The origin of this misfit may stem from misspecification of Relative Land Influence because of biases in NCEP related to poor representation of the effects of the Andes (e.g. Rusticucci and Kousky, 2002).

3.5 Interpretation of recent decadal trends

Having tuned the EBM to the seasonal cycle and demonstrated skill in reproducing simulated equilibrium conditions, we now turn to comparing the EBM results to observed temperature

changes since 1950, interpreted as a response to a trend in radiative forcing. Other influences on temperature such as land use changes, spatial variability in forcing or feedbacks, and variations in circulation are not accounted for.

Based on Eqns. (3.4) and (3.5), assigning zero temperature change at t_0, and assuming that radiative forcing increases at a constant rate β, the temperature change at t_1 is,

$$\Delta T_{i,j} = m_{i,j}\left[\frac{\beta C_{\text{land}}}{\lambda^2}\left[\exp[\frac{\lambda t_1}{C_{\text{land}}}] - 1\right]\right] + (1 - m_{i,j})\left[\frac{\beta C_{\text{ocean}}}{(\lambda - \kappa)^2}\left[\exp[\frac{(\lambda - \kappa)t_1}{C_{\text{ocean}}}] - 1\right]\right] - \frac{\beta t_1}{\lambda}. \quad (3.8)$$

The final term, $-\frac{\beta t_1}{\lambda}$, is equivalent to the equilibrated case in Eqn. (3.7). The first and second terms are the transient damping induced by land and ocean heat uptake, though the small heat capacity of the land makes the first term negligible. The only free model parameters are λ and κ, which are chosen based on minimizing the sum of squared deviations across gridboxes between the EBM and the observations. The heat capacities and Relative Land Influence field are fixed to the values used in the seasonal cycle analysis. Like the AM2.1 scenario, meridional heat flux anomalies are not included.

We set t_0 to be 1850, when the perturbative temperature change was likely to be small. For this long period, a piecewise rather than single linear forcing function is employed, but the resulting estimate of $\Delta T_{i,j}$ is similar to that in Eqn. (3.8); specifically, the expected relationship between $m_{i,j}$ and $\Delta T_{i,j}$ remains linear. The piecewise linear forcing has a single breakpoint in 1982, where the breakpoint and trends are determined from maximizing the fit to the ensemble median CMIP5 global average forcing as estimated by Forster et al. (2013) (Fig. 3.4a).

To compare the EBM to the observations, we calculate temperature change in the HadCRUT4 dataset (Morice et al., 2012) as the difference between the first and last 20-year intervals of the period 1950-2012 (Fig. 3.5a). Gridboxes are only included in the analysis if they have data available for all months of at least 80% of the years in each interval.

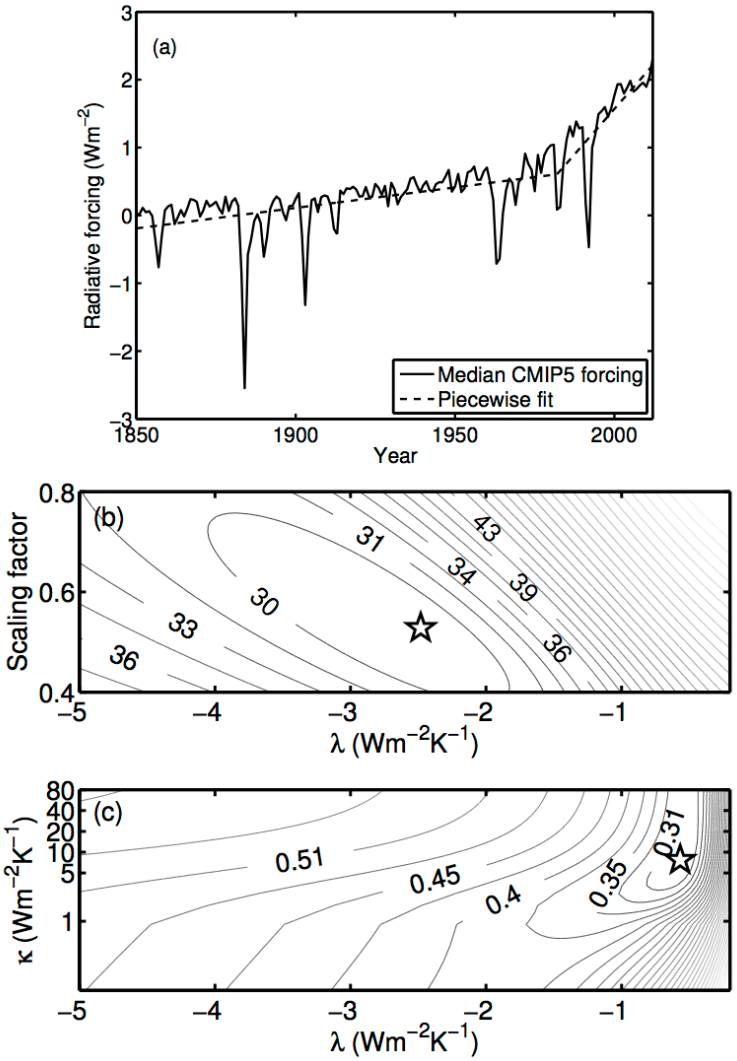

Figure 3.4: Radiative forcing and timescale-dependent sensitivity. (a) Median CMIP5 radiative forcing (natural and anthropogenic) from Forster et al. (2013), and the piecewise linear fit used to force the EBM for multi-decadal temperature change. (b) The root mean square deviation (rmsd, in °C(kWm^{-2})$^{-1}$) between the EBM and the observed gain as a function of λ and the scaling factor, k, for Relative Land Influence (see Sec. 2.2). The seasonal cycle model is insensitive to values of κ, which is set to zero. (c) The rmsd (°C) for 1950-2012 temperature change as a function of λ and κ. In both (b) and (c), the black star indicates the best-fit value for the EBM.

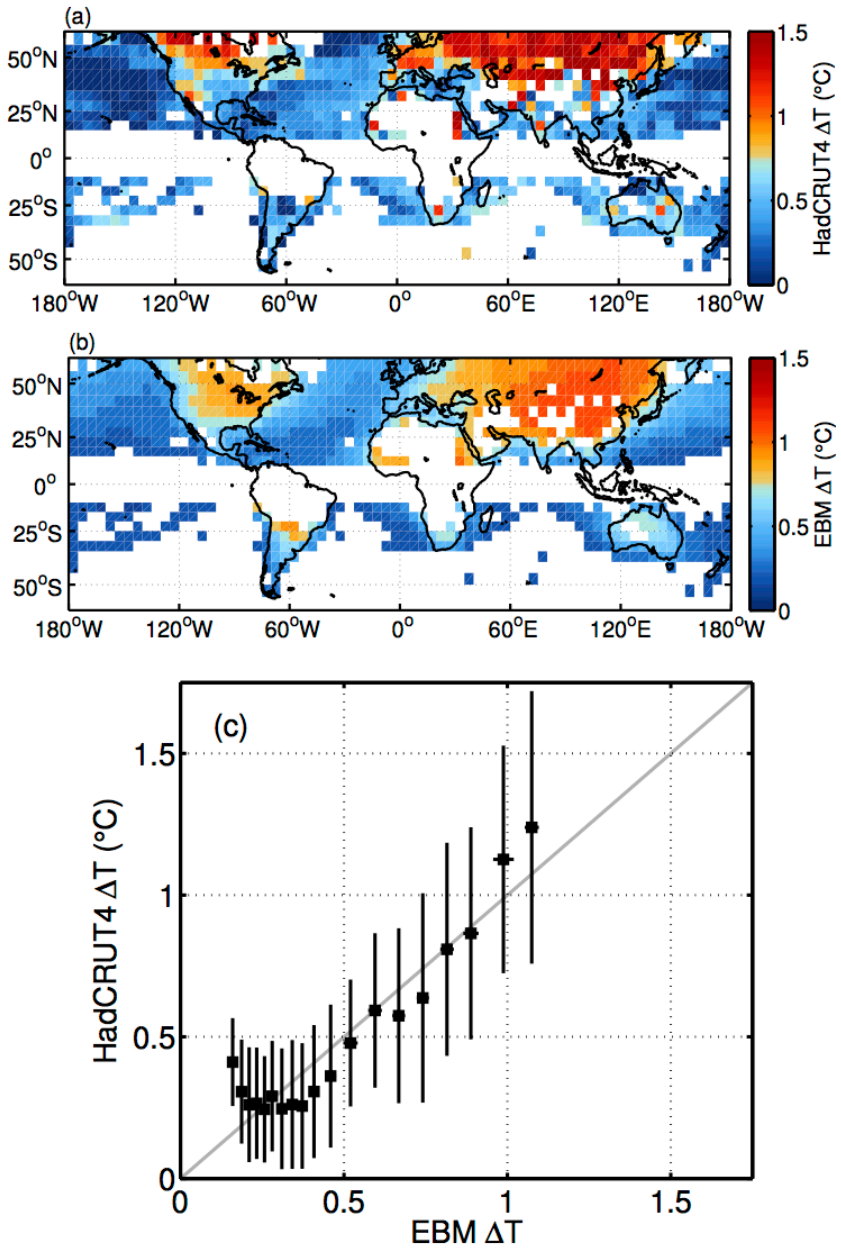

Figure 3.5: As in Fig. 3.2, but for the observed change in temperature in the HadCRUT4 dataset between 1950 and 2012.

The pattern of warming expected due to ocean heat uptake (Fig. 3.5b) is significantly correlated with the observed warming (r = 0.68, p-value < 0.01). Correlations between the EBM predictions of temperature change and the observations range between 0.45-0.70 for any pair of intervals ranging in length from 10 to 30 years, where the second interval must end in 2012, but the first interval can begin as early as 1900 (Fig. 3.6). The correlation declines for earlier intervals when fewer gridboxes meet the inclusion criteria, as well as for shorter intervals spanning the recent past, presumably because the ratio of interannual variability in temperature to the temperature response radiative forcing is larger. The correlation between the EBM estimate of $\Delta T_{i,j}$ and the observations remains significant when individually considering ocean (r = 0.28, p-value < 0.03) or land (r = 0.52, p-value < 0.04), indicating skill beyond the first order land/ocean contrast. The correlation over the ocean is lower than over land, likely related to the fact that the EBM lacks representation of ocean heat transport and changes in mixed layer depth.

Consistent with previous work using observations to determine climate parameters (e.g., Forest et al., 2002), the linear diffusivity, κ, which parameterizes deep ocean heat uptake, is only weakly constrained (see Table 2). Note that our definition of κ (Eqn. (3.3)) scales the change in temperature of the ocean end-member to calculate ocean heat uptake, rather than the global mean temperature change, as is done in Gregory and Forster (2008).

Although we do not expect the global ocean heat uptake anomalies calculated by the EBM to be accurate because the model is constrained using only extratropical data and does not account for ocean dynamics, amongst other important processes, we do compare our estimates to those from the literature as a consistency check. We find that our estimates are smaller than those obtained from combining the estimates of Lyman et al. (2010) and Purkey and Johnson (2010), as well as smaller than those from Otto et al. (2013) for the most recent decade. They are, however, consistent with estimates from Otto et al. (2013) for the 1970s, 1980s, and 1990s, as well as those from Wunsch and Heimbach (2014) for 1992-2011, and from Loeb et al. (2012) for 2000-2008. Note

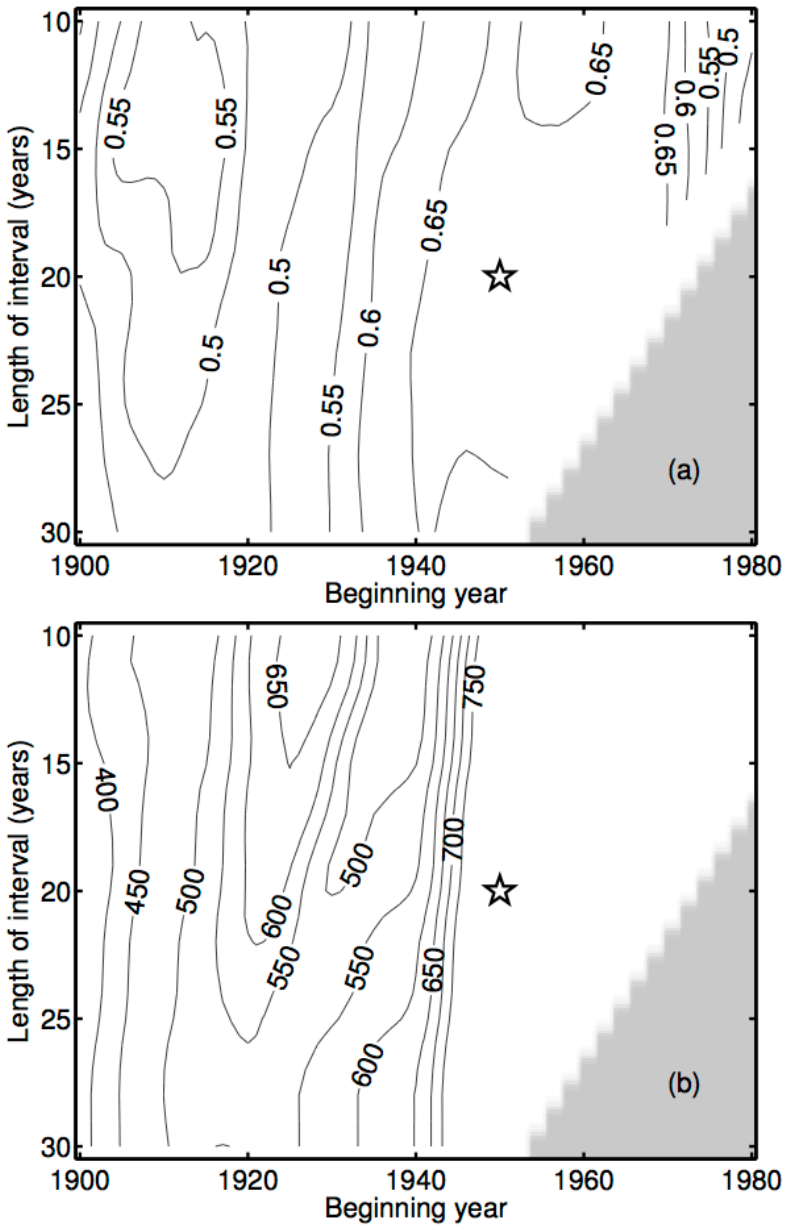

Figure 3.6: The effect of time interval on model results. (a) Correlation between the EBM patterns of temperature change and those observed as a function of the length and timing of the averaging intervals used to calculate the change in temperature in the observations. (b) the number of datapoints included in the model fit as a function of the length and timing of the averaging intervals used to calculate the change in temperature in the observations.

that Loeb et al. (2012) and Otto et al. (2013) consider the planetary, rather than ocean only, heat uptake anomalies; however, the numbers are roughly comparable because the majority (approximately 90%) of the planetary energy imbalance is stored in the ocean.

The residuals between the HadCRUT4 data and the EBM are surprisingly similar to those from the AM2.1 analysis, despite the fact that the latter was a fixed-SST model experiment. Specifically, the EBM underestimates the meridional gradient in temperature change. Like the AM2.1 analysis, the cross-over between over- and underestimating warming occurs between 40-45°, with an average overestimation in the lower latitudes of 0.12°C, and an average overestimation in the higher latitudes of 0.11°C in the Northern Hemisphere. Whether meridional patterns in Southern Hemisphere residuals exist is unclear because the data is sparse.

3.6 Discussion and conclusions

The significance of the relationship between the patterns of warming predicted by the EBM and those from an atmospheric GCM equilibrium simulation and observed over the past half-century suggest that there may be value in interpreting the inferred values of the feedback parameter, λ. The value of λ, however, is not strictly comparable to estimates of effective sensitivity in climate models because the domain we consider is not global in scope; therefore, λ incorporates the influence of any heat fluxes into or out of the domain. Instead, this feedback parameter is a metric of the temperature sensitivity to radiative forcing in the extratropics. The value of λ, -0.57 $\pm$ 0.04, found from fitting our model to the instrumental record exceeds the upper value of the CMIP3 estimates (Soden and Held, 2006), which range from -1.64 to -0.88 $\mathrm{Wm^{-2}K^{-1}}$. The difference could either reflect heat flux convergence into the extratropics or a more sensitive extratropics as compared to the globe. The AM2.1 feedback parameter, -2.09 ± 0.09 $\mathrm{Wm^{-2}K^{-1}}$, is more negative than that found for the coupled version of the model, CM2.1, of -1.37 $\mathrm{Wm^{-2}K^{-1}}$ (Soden and

Held, 2006), possibly reflecting the fact that fixing ocean surface temperature in AM2.1 limits the strength of positive feedbacks, like the water vapor feedback.

The analysis also suggests that the magnitude of the feedback parameter depends on the timescale under consideration. The feedback parameter for the seasonal cycle is significantly more negative than that for the 1950-2012 temperature change. This difference in sensitivity is not simply due to the different magnitude of ocean heat uptake on seasonal versus longer timescales because the EBM's feedback parameter and ocean diffusivity are jointly constrained by the observations (Fig. 3.4b,c). Such timescale-dependent differences in feedback strength have been found for the water vapor and lapse rate feedbacks in general circulation models (Colman and Hanson, 2013), as well as the total feedback parameter over the oceans based on an analysis of observations and reanalysis (Bony et al., 1995). A separate application of radiative kernels (Shell et al., 2008) to NCAR CCSM4 also suggests that the net feedback for the seasonal cycle is more negative than that for the response to ramped radiative forcing.

Although the intentionally simple nature of the EBM allows for ease of interpretability, there are substantial limitations. A diversity of important climate processes were explicitly excluded including possible changes in meridional heat transport, spatial variability in radiative forcing, spatial or temporal structure in the ocean's mixed layer, and land cover distributions and their changes. We cannot conclusively attribute model residuals to any of these processes, or ensure that the strength of the model fit is not in part due to coincidental cancelation of processes not modeled. Other consideration include that we cannot apply this model in the deep tropics because tuning relies on the presence of a strong once-per-year season cycle, and that our our metric of Relative Land Influence would be compromised by changes in atmospheric circulation.

It appears, however, that this minimalist model at least partially captures the advective land-ocean heat fluxes that lead to similar patterns of variability at different timescales. This result may seem surprising given the substantial spatial structures in forcing and feedbacks (Zelinka

and Hartmann, 2012; Armour et al., 2013), but may result from horizontal heat transport diverging heat away from regions experiencing larger forcing and/or less negative feedbacks. Such an effect has been shown using aquaplanet simulations in which temperature changes due to feedbacks are strongly anti-correlated with those due to transport (Feldl and Roe, 2013). The meridional structure of the residuals between the EBM predictions and the AM2.1 simulation, as well as the HadCRUT4 observations, are consistent with simulations showing increased poleward heat transport in response to increased radiative forcing, with the switch from divergence to convergence occurring around 40-45°N in model simulations (Held and Soden, 2006).

The interpretability of the equilibrium and multi-decadal patterns of temperature change predicted by the EBM suggests that the repeatedly-observed seasonal cycle contains valuable information for constraining and understanding longer-term temperature responses to radiative forcing. In future work, it would be useful to further explore similarities between seasonal and longer-term temperature variability in the context of more complete models.

4

Long-lead prediction of US heat events using the Pacific Extreme Pattern

The previous chapters have explored the role of heat exchanges between the land, oceans, and atmosphere in influencing the spatial pattern of temperature variability and change on seasonal and longer timescales. These mean structures present an average over many synoptic events, which themselves transfer heat, moisture, and momentum through the climate system. This chapter moves to the shorter timescales of individual weather events, with a focus on anomalously hot days in the Eastern United States. Rather than demonstrating the importance of direct, thermodynamic exchanges of heat between the land, oceans, and atmosphere, it is demonstrated that specific sea surface temperature patterns in the Pacific Ocean can initialize a dynamic response in the atmosphere that ultimately leads to a convergence of heat over the United States. The identification of this process allows for long-lead predictability of these detrimental events. This chapter is a modified form of a submitted manuscript, McKinnon, K.A., Rhines, A., Tingley, M.P. and P. Huybers, Long-lead prediction of United States heat events using the Pacific Extreme Pattern, wherein the supplementary information has been incorporated into the main text.

4.1 Abstract

Predictability of mid-latitude weather is generally limited to no more than 10 days, beyond which growth of uncertainty precludes skill. Identifying atmospheric interactions with more slowly-varying components of the climate system can extend this time horizon. Clustering and regression on historical land and sea surface temperature data reveals an ocean-atmosphere mode whereby enhanced meridional temperature gradients in the mid-latitude Pacific ultimately promote the development of summer blocking patterns over the Eastern United States. Predictions based on this mode are skillful at lead times greater than 40 days and involve three-fold changes in the odds that an extreme heat event will occur. Advanced warning at these long lead times has not previously been possible, and should aid in preparing for heat events.

4.2 Introduction

Weather models provide forecasts for mid-latitude weather events on timescales of a week, and specific events may be predicted up to 15 days in advance based on precursor atmospheric circulation regimes (Teng et al., 2013). Predictions at longer lead times rely on more slowly varying components of the climate system, and a substantial body of work has identified connections between tropical sea surface temperature (SST) variations and mid-latitude weather during the winter season (Trenberth et al., 1998), leading to valuable seasonal predictions (Higgins et al., 2004). In contrast, connections between summer mid-latitude weather and SSTs in the tropics appear limited (Chen and Yen, 1993; Newman and Sardeshmukh, 1998). There is some evidence that mid-latitude Pacific SST anomalies are connected with summertime high temperature anomalies in the United States (Namias, 1982; Lyon and Dole, 1995), although, to our knowledge, no skillful predictive relationship has previously been demonstrated between SST patterns and summer heat events.

4.3 Data and methods

Heat events can be characterized by their magnitude or spatial extent, and are often considered in a binary sense. In this study, heat events are quantified using a continuous variable, the severity index, that accounts for both the magnitude and extent of the event. Magnitude is assessed at each station as anomalies in daily maximum temperature (hereafter simply temperature) relative to an 11-year running average of the station's 95th percentile of summer temperatures. Extent is then accounted for by computing the temperature anomaly exceeded by the hottest 5% of the spatial domain on a given day, yielding the severity index. An index of zero means that 5% of stations have a temperature anomaly in the top 5% of their distribution. The severity index is defined for each summer day and is approximately normally distributed (Fig. 4.1a).

Temperature data is from the Global Historical Climatology Network-Daily (GHCND (Menne et al., 2012)), which is provided at the station level. Weather stations provide a more direct measurement of near-surface air temperature than atmospheric reanalyses (Hofstra et al., 2010), which do not directly ingest station data (Simmons et al., 2004). To be included in the analysis, stations are required to have at least 80% coverage during June, July, and August for at least 80% of the years considered in the analysis. There are 1813 stations in the Eastern United States that fulfill these requirements for the 1982-2013 period, and 1222 for the longer 1950-2013 period. The running average is used to focus on interannual variability, and results are similar if 21- or 31-year windows are used instead. To minimize the influence of the seasonal cycle, we focus on predicting the severity index between June 24 – August 22, the 60 hottest days of summer based on the average climatology across United States weather stations (FIg. 4.1b).

The region of study is defined by clustering United States weather stations together on the basis of synchronously having temperatures that exceed the 95th percentile threshold. Specifically, we use hierarchical agglomerative clustering that minimizes the average intracluster Jaccard dis-

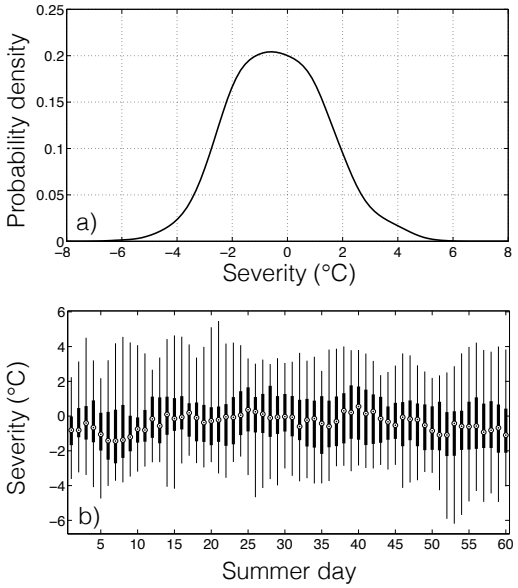

Figure 4.1: The severity index. (a) The distribution of the severity index. (b) The severity index as a function of summer day (June 24 – August 22). The center of each box is the median, the box spans the 25th-75th percentiles of the data, and the whiskers show the span of the remaining data.

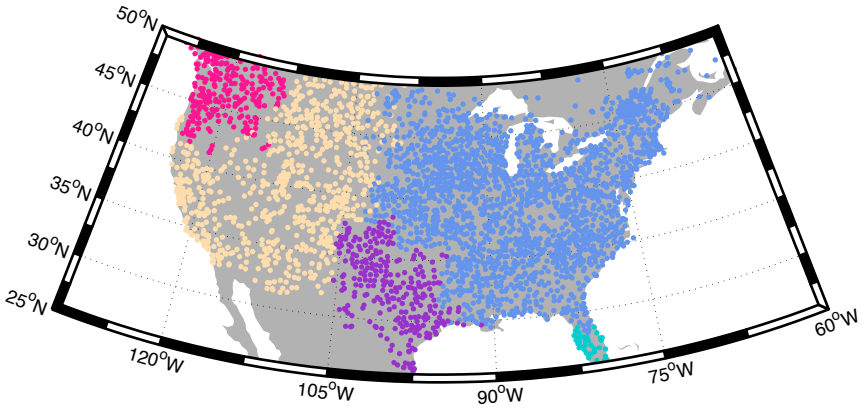

Figure 4.2: The five largest clusters of the GHCND station data in the United States (and Southeast Canada), based on if stations experience hot days synchronously.

tance (Jaccard, 1912) between stations, where the Jaccard distance is a function of a binary variable that indicates whether the temperature at a station was above the 95th percentile threshold or not. The algorithm incorporates no geographic information but produces spatially-distinct groups of weather stations (Fig. 4.2). Using four large clusters divides the United States into interpretable climate zones: the arid West, the semiarid and subtropical Southwest, the marine Northwest, and the humid and subtropical Eastern United States. We focus on the latter cluster because it contains major population centers and primary regions of agriculture, both of which tend to be disproportionally affected by heat events (Smoyer et al., 2000; Battisti and Naylor, 2009).

4.4 Identifying the Pacific Extreme Pattern

In order to identify precursors to heat events, we calculate averages of daily SST anomalies weighted by the severity index. SST data is at daily resolution from the NOAA OI SST2 dataset (Reynolds et al., 2002), and is available from September 1981 to the present. SST anomalies are calculated as deviations from the climatology, estimated as the average SST for that day of year during the full period of analysis (1982-2013, or 1950-2013). The SST anomaly pattern synchronous with heat events shows coherent spatial structures in the mid-latitude oceans (Fig. 4.3d), with a tripole pattern in the mid-latitude Pacific composed of anomalies as large as $\pm 0.4^{\circ}$C. The tripole SST pattern tends to evolve over the 50 days leading up to a heat event in the domain between 20-50°N and 145-230°E (Fig. 4.4), and we term this evolving structure the Pacific Extreme Pattern, or PEP. Similar structures are evident for individual events (Figs. 4.5, 4.6).

The temporal evolution of PEP can be described in three stages. In the first stage, spanning the 50 to 10 days before the heat event (Fig. 4.4), an anomalous meridional dipole in the mid-latitude Pacific enhances the background climatological gradient in SST. During this period,

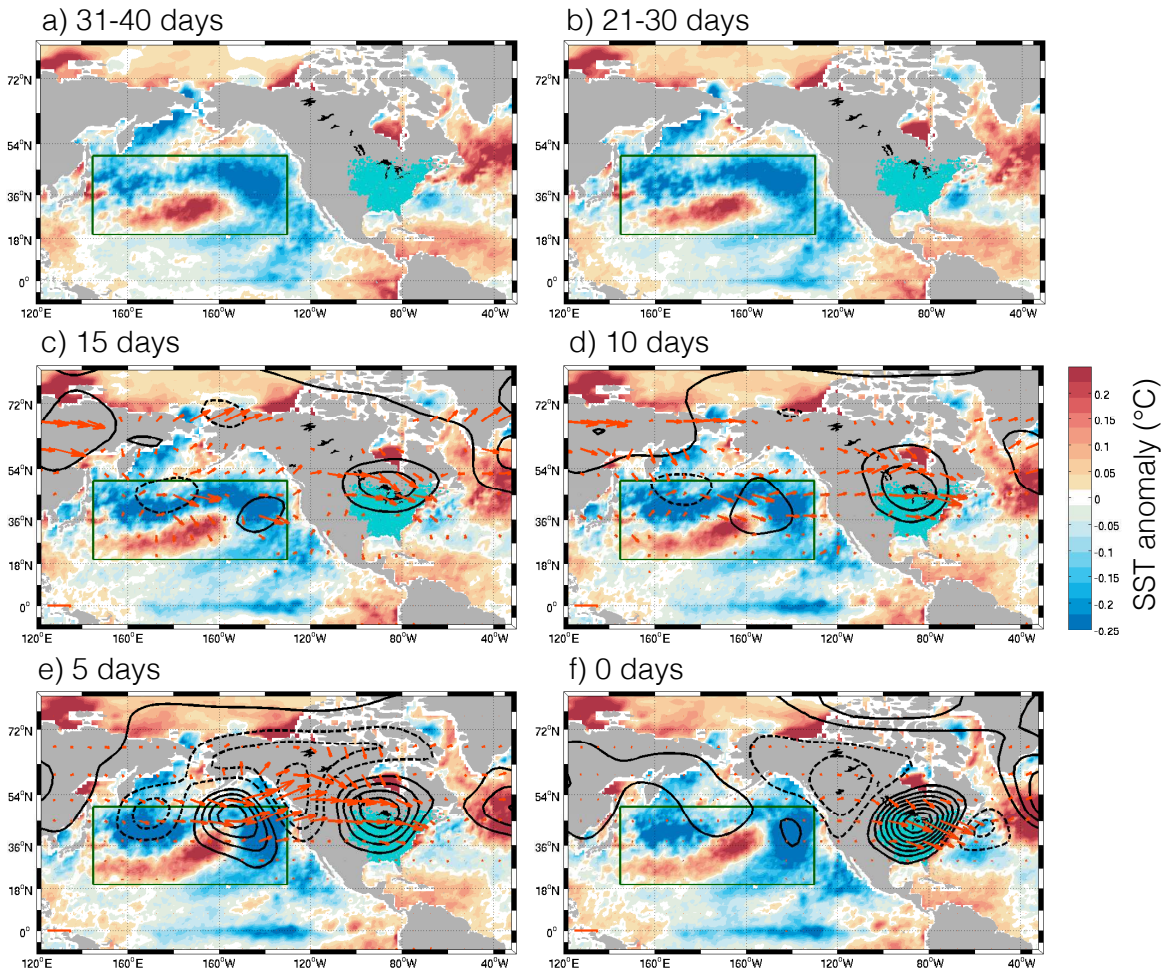

Figure 4.3: Patterns associated with Eastern United States heat events. (a-b) SST anomalies averaged over 10-day periods for lead times of (a) 31-40 and (b) 21-30 days. (c-f) Anomalies in SST (colors), 300mb height (contours), and wave activity flux (Takaya and Nakamura, 1997) (red arrows, $5m^2s^{-2}$ scale bar in lower left of all plots) for lead times of (c) 15, (d) 10, (e) 5, and (f) 0 days. The Pacific Extreme Pattern is defined inside the green box. Weather stations in the Eastern cluster are in turquoise. Contours are in 10m increments, with negative contours dashed and the zero contour suppressed. The occurrence of the warm SST anomaly in the North Atlantic does not have predictive skill for heat events.

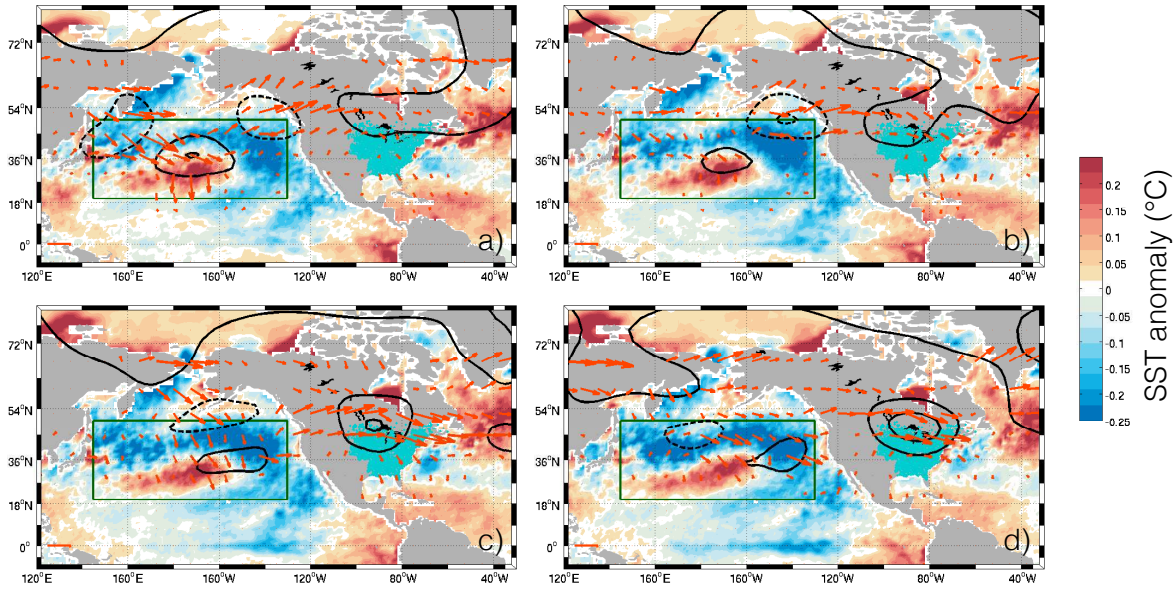

Figure 4.4: As in Fig. 4.3, but for 10-day averages at longer lead times of (a) 41-50 days, (b) 31-40 days, (c) 21-30 days, and (d) 11-20 days.

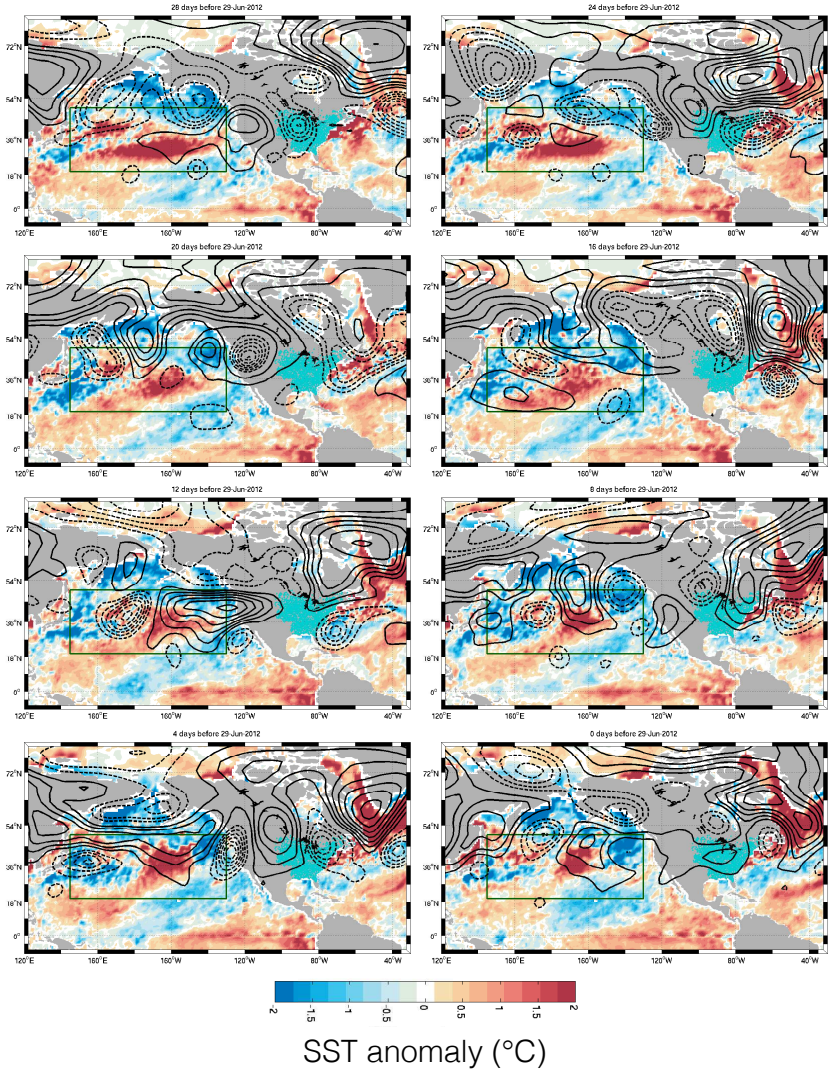

Figure 4.5: Case study in predicting an individual heat event. The SST anomalies (colors) and 300mbar geopotential height anomalies (contours, 50m interval, negative contours dashed) leading up to June 29th, 2012. The severity index on June 29th was 4.19°C, with 42% of the Eastern United States having temperatures above the decadal-scale 95th percentile threshold. The PEP index was positive for the 97 days preceding the event, and surpassed 1 for the 43 days before the event.

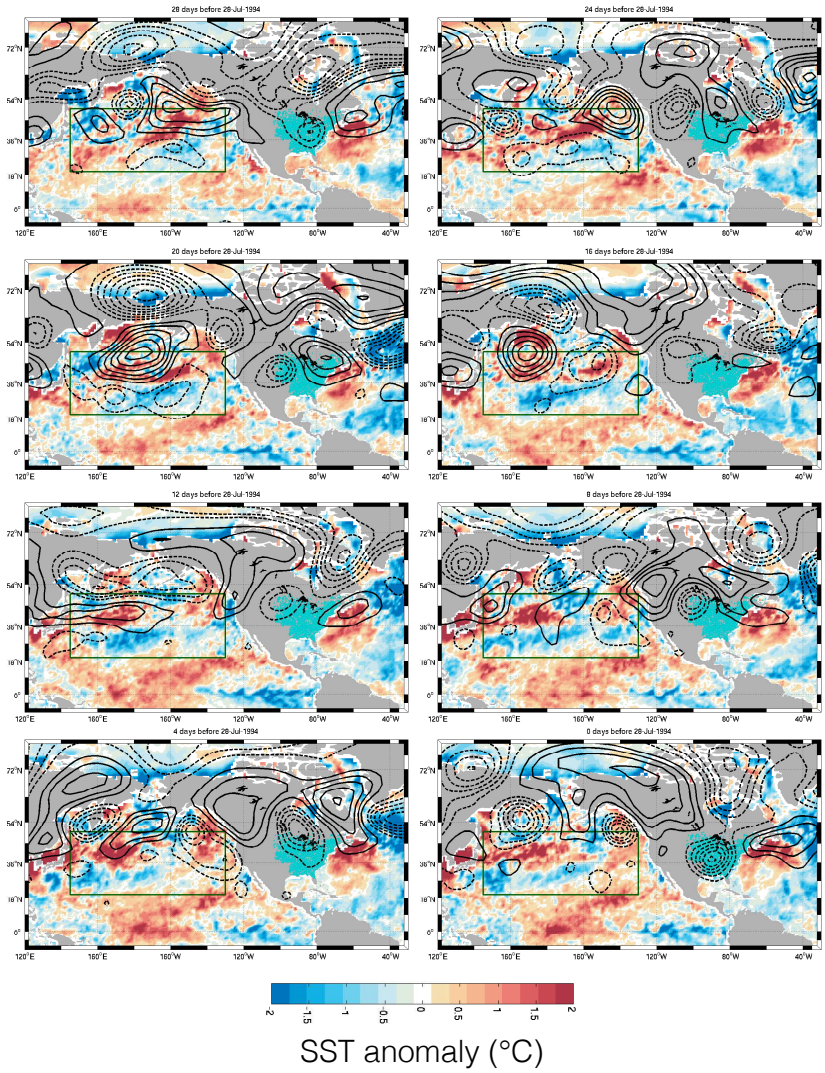

Figure 4.6: As in Fig. 4.5, but for predicting the suppression of a heat event. The severity index on July 28th, 1994 was -3.46°C, with only 3 stations in the Eastern United States having temperatures above the 95th percentile threshold. The PEP index was negative for the 26 days before the event, and was less than -1 for the 20 days before the event.

anomalous atmosphere-ocean heat fluxes further intensify the SST gradient (Fig. 4.7), perhaps resulting from the integrative response of SST to stochastic atmospheric forcing (Frankignoul and Hasselmann, 1977). The enhanced meridional SST gradient from the first stage is expected

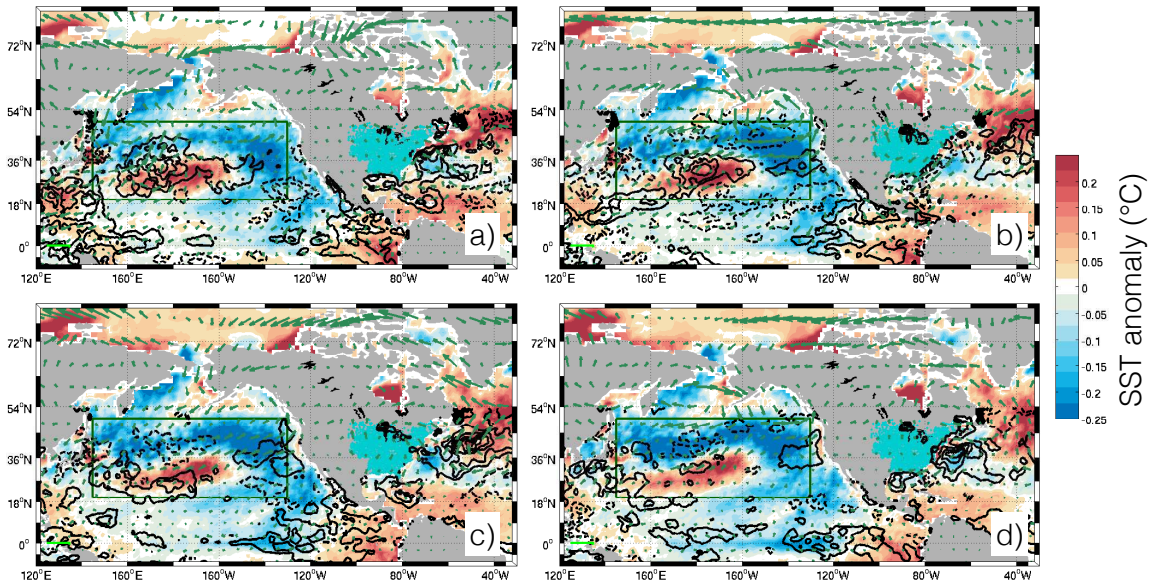

Figure 4.7: SST, ocean-atmosphere flux, and 10m wind anomaly patterns associated with Eastern United States heat events at long lead times. Each panel shows a 10-day average. Anomalies in SST (colors), ocean-atmosphere fluxes (contours at 3Wm^{-2} intervals, positive is into-ocean), and 10m winds (green arrows, 5ms^{-1} scale in the lower left corner of all plots) for lead times of (a) 41-50, (b) 31-40, (c) 21-30, and (d) 11-20 days.

to increase low-level baroclinicity, and in the second stage ocean-atmosphere interactions appear to form and amplify a mid-latitude wave train (Fig. 4.3). The possibility of a similar finite-amplitude growth of ocean-atmosphere anomalies was derived analytically in an earlier study (Pedlosky, 1975). Approximately 10 days before the heat event, a cyclone-anticyclone pair establishes over the Pacific that transports heat into the center and removes heat from the edges of the Northern cold anomaly (Fig. 4.8), giving rise to a zonal cold-warm-cold tripole. The com-

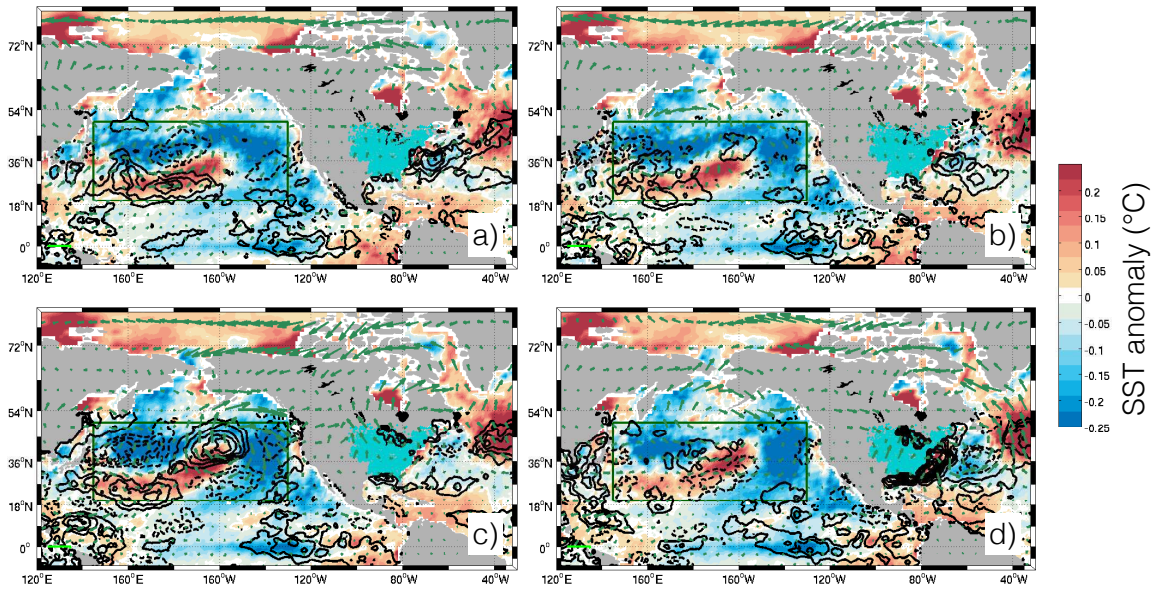

Figure 4.8: As in Fig. 4.7, but showing the weighted averages for individual days at shorter lead times of (a) 15, (b) 10, (c) 5, and (d) 0 days.

posite atmospheric structure is equivalent barotropic (Fig. 4.9), as is typical in the mid-latitudes (Hoskins and Karoly, 1981). Atmospheric pressure anomalies are in quadrature with SST anoma-

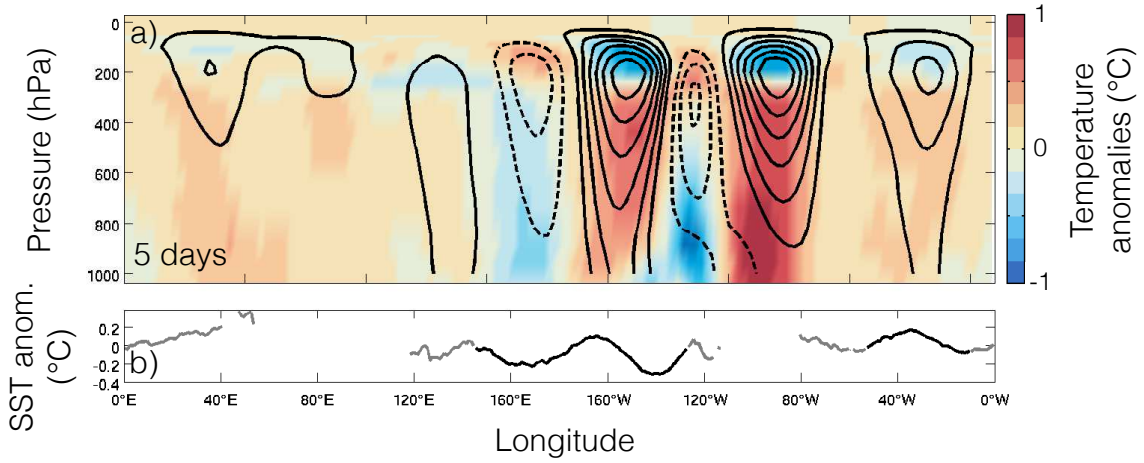

Figure 4.9: Air temperature, geopotential height, and SST anomalies associated with heat events averaged between 30-50°N at a lead time of five days. (a) Vertical structure of air temperature (colors) and geopotential height anomalies (contours). Contours are in 5m increments, with negative contours dashed and the zero contour suppressed. (b) SST anomalies for longitudes containing ocean. Partial ocean coverage is indicated by grey lines.

lies and extend northward of the Rocky Mountains, suggesting that orography may additionally contribute to controlling the phase of the disturbance.

Whether these SST anomalies cause remote change depends on the mean state of the atmosphere. Model simulations indicate that tropically-forced Rossby waves do not effectively propagate into the mid-latitudes during the summer (Yun et al., 2011); rather, North America appears particularly sensitive to anomalies in the mid-latitude Pacific with length-scales similar to PEP (Newman and Sardeshmukh, 1998). At lead times of 10 days, synchronous with the growth of the cold-warm-cold tripole, there is increased wave activity flux (Takaya and Nakamura, 1997)

over the mid-latitude Pacific. Wave activity flux then propagates eastward, consistent with a ducting of eddy energy into the wave guide, and, in the third stage, begins converging over the Eastern United States. An anomalous high grows rapidly in the region (Fig. 4.3c,d), and stationary dry and clear conditions permit for increased radiative heating of the surface. Surface drying may also lead to amplified temperature anomalies(Miralles et al., 2014). The atmospheric circulation anomalies identified in the third and final stage are similar to those identified by (Teng et al., 2013) in a long model simulation.

4.5 Quantifying predictive skill

In order to evaluate the predictive skill associated with PEP, we construct a linear statistical model that uses the covariance of PEP with observed SSTs between 1982-2013 to predict the severity index. Covariance is divided by its standard deviation across summer days for interpretability. The normalized covariance is termed the PEP index, and a PEP index of 1 is equal to 0.034°C. The linear relationship between the PEP index and the severity index is significant at all lead times up to 47 days (p-value < 0.05), where significance is assessed using a block bootstrap with a block size of one summer (Fig. 4.10c). All analyses are performed using *leave-one-year-out* cross validation, such that data from a given year is excluded when calibrating the prediction model for that year. Excluding only the predicted year from the analysis is sufficient because mid-latitude summer SST anomalies rarely persist for more than six months (Deser et al., 2003).

Predictions are compared to baselines of persistence and climatology (Fig. 4.10a,b). Persistence predicts the severity index on a given day as the same severity index as was observed at the specified lead time. Climatology is defined as the average severity on a given day of year across all years except that being predicted. Unsurprisingly, persistence predictions perform best in the two days immediately preceding a heat event. More relevant for long-lead predictions is

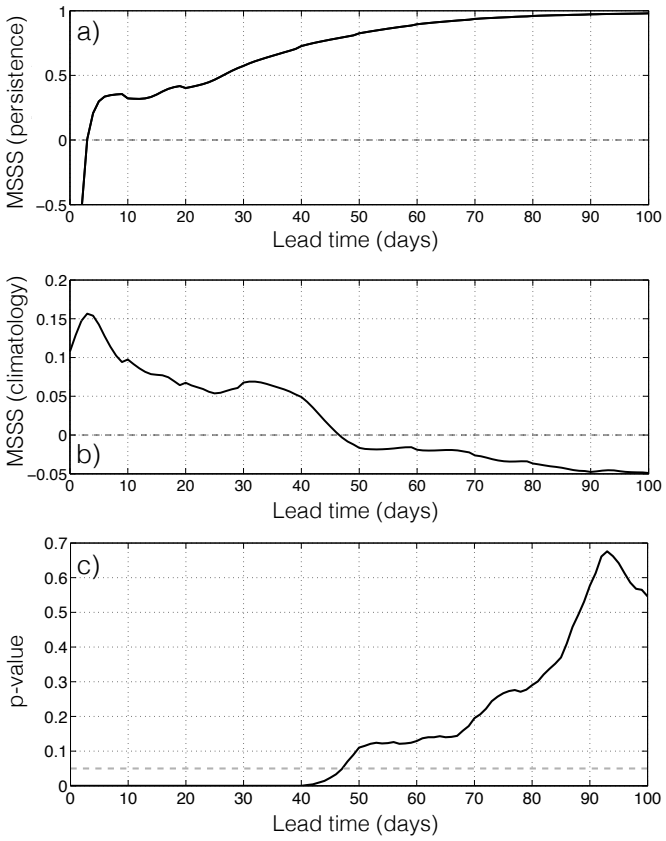

Figure 4.10: Model skill for predicting the severity index with the PEP index using a linear model. (a-b) The mean square skill score (MSSS) for baseline models of (a) persistence and (b) climatology. (c) The p-value for the linear model. The p-value is calculated using a block bootstrap, with a block size of one summer. The dashed gray line is at a p-value of 0.05.

that our cross-validated statistical model outperforms climatology at lead times of 46 days or less as measured by the mean square skill score (Murphy and Epstein, 1989).

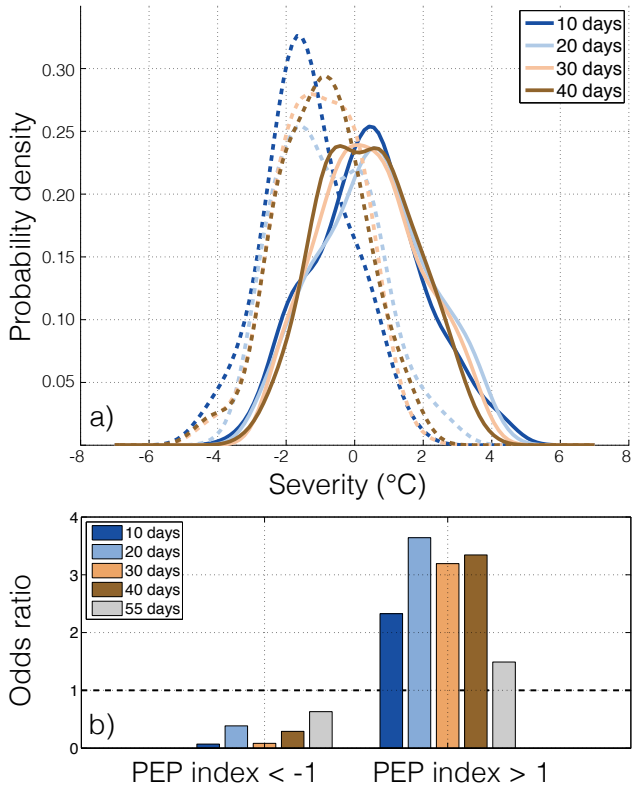

Figure 4.11: Shifts in the severity index, conditional on the PEP index. (a) The distribution of the severity index across days with a PEP index greater than 1 (solid lines) and less than -1 (dashed lines). Distributions are smoothed using a kernel density smoother with a bandwidth of 0.5. (b) The odds ratio for predicting a severity index in the top 15% of days for lead times of 10, 20, 30, 40, and 55 days, given a PEP index less than -1 (left) and greater than 1 (right).

Because the ocean-atmosphere system does not deterministically evolve, it is useful to define a prediction interval that averages over $\pm 10\%$ of the lead time, where interval length is rounded downward (e.g. a 4-day half-window for lead times between 40-49 days). Results are similar if no averaging is used, albeit with reduced skill at longer lead times. Differences in the distributions of the severity index when conditioning on PEP indices either greater than 1 or less than

-1 are large (Fig. 4.11a). The medians of the two severity distributions differ by 1.88°C, 1.43°C, 1.45°C, and 1.33°C for lead times of 10, 20, 30, and 40 days, respectively. Predictions can also be made in terms of threshold-crossing events. Consistent with (Barnston and Mason, 2011), we define an extreme event as one in the top 15% of severity, and find that the odds of extreme events increase by 2 to 4 times when a PEP index greater than 1 is observed in the 10 to 40 days prior (Fig. 4.11b).

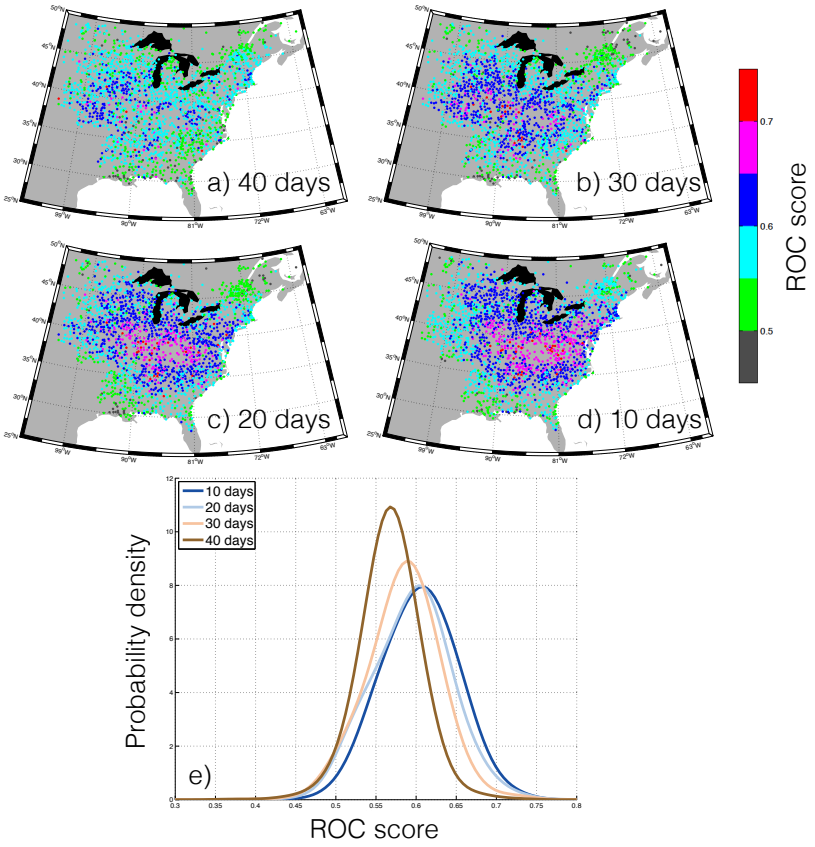

Figure 4.12: ROC scores at the station level. The spatial distribution of ROC scores for lead times of (a) 40, (b) 30, (c) 20, and (d) 10 days. (e) The distribution of ROC scores. The distributions are created using a kernel density smoother with a bandwidth of 0.05. A ROC score of 0.5 indicates no model skill.

Predictions can also be resolved at the station level (Fig. 4.12) using Relative Operating Char-

acteristic (ROC) curves Swets (1973). ROC curves illustrate the trade-off between false positives and true positives in a prediction. The area under the curve can be integrated to calculate the ROC score, and scores greater than 0.5 indicate model skill at discriminating between an extreme event occurring, or not. Extreme events at stations in the interior of the domain are better predicted than those on the edges, likely due to a combination of coastal effects and the most likely location of the blocking high. While skill at individual stations is lower than that for the region as a whole, the model retains skill at 96% to 99% of stations for lead times of 10 to 40 days.

Finally, the model is used to predict extreme heat events between 1950 – 1981, a period independent of all foregoing analyses, using an SST dataset that extends further back in time (Kennedy et al., 2011). The distribution of the severity index again exhibits shifts on the order of 1°C when conditioning on PEP indices greater than 1 or less than -1, and observing a PEP index greater than 1 leads to a 1.5 – 3 times increase in the odds ratio of an extreme event (Fig. 4.13).

Although we see a clear sequence of development of PEP in the mid-latitude Pacific, it is also relevant to consider if more established modes of variability permit for predictability of Eastern United States heat events. Averaged over lead times of 10 to 40 days before heat events, PEP explains 42% of the interannual variance in the summer-average severity index (p-value < 0.01, Fig. 4.14; significance is assessed using leave-one-year-out cross validation). In contrast, we find that the June-August summer average of the indices associated with the Pacific North America pattern, the Arctic Oscillation, the Pacific Decadal Oscillation, the Northern Annular Mode, and the El Niño-Southern Oscillation are not significantly correlated with summer average severity, and explain no more than 5% of the interannual variability. The lack of skillful predictions based on tropical indices is consistent with the aforementioned minimal tropical influence on the mid-latitudes during summer (Yun et al., 2011).

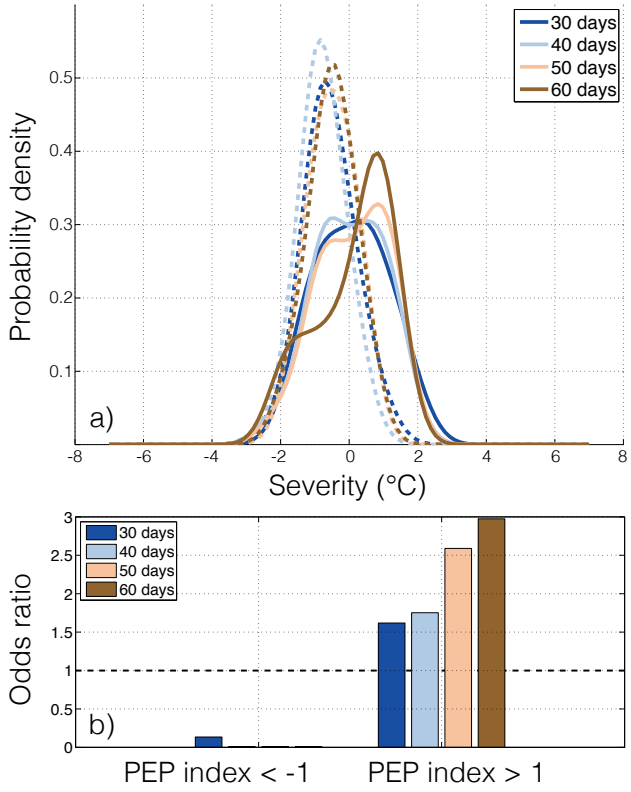

Figure 4.13: As in Fig. 4.11, but using the HadSST3 data.

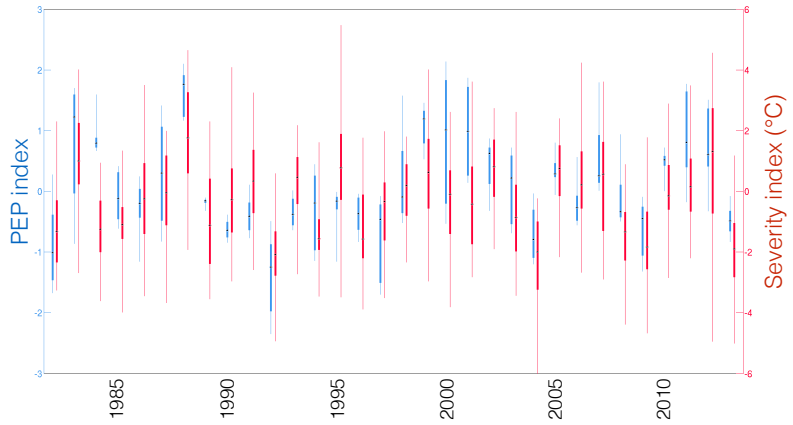

Figure 4.14: Interannual variability in the severity index (red, right axis), and the PEP index (blue, left axis), averaging over lead times of 10-40 days. The center of each box is the median, the box spans the 25th-75th percentiles of the data, and the whiskers show the span of the remaining data.

4.6 Conclusion

The 21st century has already featured a number of ill-predicted (Dole et al., 2011) and costly (Smith and Katz, 2013) heat events that have had negative impacts on crop production (Battisti and Naylor, 2009) and human health (Kovats and Hajat, 2008). Trends over the last several decades towards increased frequency and magnitude of summer heat events in the United States (DeGaetano and Allen, 2002) are expected to continue (Seneviratne et al., 2012), making forecast skill increasingly important. We demonstrate that PEP permits for skillful prediction of heat events at lead times greater than 40 days. These long-lead predictions can be combined with early warning systems to reduce the impacts of extreme heat events (Ebi and Schmier, 2005). Dynamical prediction of sea surface temperature evolution may further extend the time horizon of predictability.

5

Conclusion and further work

Temperature exhibits coherent variability across space and time. Understanding the origin of these patterns of variability, and how they will change due to anthropogenic influences on the climate, has been the focus of this thesis. Initial investigations focused on the seasonal cycle, due to its strong signal-to-noise ratio. The amplitude and phase of the seasonal cycle exhibited both expected and surprising behavior, the latter of which included the fact that western seas can have a more variable seasonal cycle than western edges of continents due to the strong influence of upstream seasonal cycles. This, and other subtle characteristics, appear to emerge from advective heat transport by many individual air parcels, averaged over time. Because the Lagrangian timescale of the atmosphere is short compared to the seasonal cycle, and the difference in heat capacity between the land and ocean is timescale invariant, one is encouraged to hypothesize that a similar process may be important on other timescales. Consistent with this hypothesis, a simple model tuned using the seasonal cycle is shown to provide skillful predictions for the pattern of equilibrium temperature change in a model with a fixed ocean, as well as for the observed pattern of transient temperature change.

Implicit in the seasonal and decadal estimates of temperature variability and change is an average over thousands of individual weather events; the distribution of the weather events 'define' the climate on long enough timescales*. Heat waves are a subset of weather events that sample from the high end of the distribution of temperature at a given location. The frequency of days that exceed a pre-specified temperature threshold will increase in the simple case of an increase in the mean of the temperature distribution, but this prediction is statistical rather than deterministic. To predict individual events, it is necessary to delve into their dynamical origins. The final chapter of this thesis identifies conditions in the Pacific Ocean that tend to precede Eastern US heat waves, thus allowing for long-lead, skillful predictions. The empirically-identified pathway from Pacific SSTs to the Eastern US suggests the importance of coupled ocean-atmosphere interactions that lead to transient growth in an anomalous mid-latitude wave train.

The collection of methods and results presented here suggest future research directions.

The seasonal cycle in deep time

The study of present-day climate benefits from the existence of observations that cover much of the earth and are highly temporally resolved, whether they are point observations at a weather station or fully gridded satellite data products. In contrast, studies of paleoclimate tend to depend on a relatively small set of point observations (ice and sediment cores, pollen data, glacial moraines) that may preferentially sample one portion of the seasonal cycle more than another (Schneider et al., 2010). To move from these proxy data to estimates of global or regional temperature, assuming that the relationship between the proxy and temperature is known (itself a major challenge), one must assume a spatio-temporal covariance structure (Tingley et al., 2012). The analysis of the seasonal cycle presented here can serve as a guide for the interpretation of

*Ideally, it would be possible to define the climate from observing the weather over infinite time, assuming that the climate system is stationary. Unfortunately, we do not have infinite observations. Additionally, due to climate change, the climate system is not stationary. Thus, more ad hoc definitions of climate are necessary, such as the commonly-used average over 30 years.

some paleodata, under the assumption that the mean atmospheric circulation has not undergone dramatic shifts. While the atmospheric circulation data used in the seasonal cycle model are based on today's climate, results would be similar if one were to assume that the extratropical circulation is entirely zonal, with some modification by existing orography. As such, the model could be applied to paleo continental configurations to infer the spatial structure of seasonal variability. An an example, a puzzle from the Eocene is the existence of palms (Wing and Greenwood, 1993; Greenwood and Wing, 1995) and crown group crocodilian fossils (Markwick, 1998) in modern-day Wyoming, a region typically associated with the cold winters of continental interiors. These fossils, however, have been interpreted to mean that winter temperatures in the Eocene could not have dropped below 5.5°C. Climate models have generally struggled to simulate these warm winter temperatures without increasing greenhouse gases to very unlikely levels (Huber and Caballero, 2011). Applying the seasonal model to the Eocene, including a modified continental configuration, could provide constraints on the range of parameter space (optical thickness of the atmosphere, zonal wind speed, atmosphere-surface coupling, climate sensitivity) that is consistent with the fossil record.

Seasonality of circulation

The seasonal cycle model of Chapter 2 is based upon the annual average behavior of ensembles of Lagrangian parcels. While this metric leads to a good model for the seasonal cycle, it ignores the substantial seasonal structure in atmospheric circulation patterns. For instance, the seasonal cycle in the strength of the jet would be expected to make mid-latitude land locations more maritime during the winter and more continental during the summer. A more complete model could account for the seasonality of the circulation, and be compared to the model presented here to quantify the importance of the seasonality of circulation in influencing the seasonal cycle. Relatedly, there is evidence that seasonal shifts in atmospheric circulation patterns have altered the

amplitude and phase of the seasonal cycle since 1950 (Stine and Huybers, 2012). A changing seasonal cycle has major implications for the biosphere (Walther et al., 2002), crop production (Tao et al., 2006), and water resource management (Regonda et al., 2005). The relationship between changing circulation and the seasonal cycle can be examined by parsing trajectory data by year, and determining the extent to which the model reproduces the interannual variability and trends that have been observed for the seasonal cycle.

Improving estimates of ocean heat content anomalies

There is likely a positive energy imbalance at the top of the atmosphere (Loeb et al., 2012), the majority of which is accommodated by an increase in oceanic heat content. Quantifying the temporal change in the heat content of the ocean is critical for understanding the likely climatic response to increasing radiative forcing: larger values of oceanic heating imply a more sensitive climate in equilibrium, holding everything else equal. The planetary energy imbalance, however, is difficult to quantify. Satellites measure top of the atmosphere radiation with great precision (Loeb et al., 2007), but with a bias that is larger than the signal of interest (Loeb et al., 2009). Meanwhile, even with the impressive Argo array, we are lacking measurements of the deep and high-latitude oceans, as well as the marginal seas (Roemmich et al., 2009). The current best estimates of the planetary energy imbalance use ocean heat content data to 'pin' the satellite data to the correct mean; unfortunately, this approach does not leave any remaining independent data with which to test the estimates of the imbalance. An alternative approach is to first focus on the seasonal cycle, which is resolved by the highly-precise satellite data and should align in amplitude and phase with planetary storage data. Storage data can be gleaned from Argo, as well as independent estimates of atmospheric and cryospheric storage. Misfits on seasonal timescales can be clearly identified, and may have implications for the decadal trend.

Quantifying and predicting changes in extremes

Many first-order questions about extreme weather events remain unanswered. A fundamental one is simply whether the frequency of extreme events has changed over a specified time period. Preliminary work using quantile regression (Koenker and Bassett Jr, 1978), a non-parametric method which generalizes estimates of trends in the median to other percentiles, suggests that seasonal temperature extremes in the United States are changing in spatially-coherent manner. For example, while the 95th percentile of daily summer maximum temperature is increasing across most of the country, there is a band of cooling that spans the major cropping regions of the country (Rhines et al., submitted). These types of basic data analyses can be extended to temperature and precipitation in other regions to provide a first-order understanding of how the full distributions of temperature and precipitation are shifting. A second step is to assess why these changes are occurring. A clear extension from the heat wave prediction work of Chapter 4 is to determine whether the frequency of the Pacific Extreme Pattern is changing, or is projected to change in the future in climate models.

Mid-latitude ocean-atmosphere interactions

Great gains have been made in understanding the role of tropical SSTs in influencing mid-latitude circulation (e.g., Trenberth et al., 1998). In contrast, it remains unclear whether, or to what extent, mid-latitude SST anomalies have a dynamical influence on the mid-latitude atmosphere (Kushnir et al., 2002). The empirical analysis presented in Chapter 4 is consistent with coupled ocean-atmosphere interactions that lead to the transient growth of both atmospheric and oceanic anomalies; however, data analysis alone cannot conclusively demonstrate this dynamical hypothesis. If the system is truly coupled, fixed SST model experiments would not be expected to reproduce the empirical results. Rather, further work using a hierarchy of physical models that allow for some level of bidirectional coupling could be used to determine if the interactions do

require coupling, and to explore whether similar coupled interactions happen in other regions or during other seasons.

The interplay between extreme temperature and precipitation

The focus of this thesis has been on temperature, but there are clear, important extensions to be made by addressing precipitation variability. The hydrological cycle is an integral part of the energy cycle, and the redistribution of heat in a changing climate has implications for both temperature and precipitation changes. Furthermore, the occurrence and impact of extreme temperatures is modulated by precipitation: low soil moisture can lead to an amplification of hot temperatures (Miralles et al., 2014), since a greater fraction of heat is allocated to warming (sensible heat) rather than evaporation (latent heat). While drought, a major natural disaster, does not have a single definition, it generally occurs during conditions that are both anomalously warm and anomalously dry (Heim Jr, 2002). Important open questions include quantifying the (spatially-variable) timescale over which precipitation influences temperature variability, determining the covariance structure between temperature and precipitation for both the mean and the extremes, and analyzing the significance of any trends in these quantities.

Collectively, these topics will expand our understanding of both 'natural' and forced climate variability, with a focus on observables – temperature and precipitation – that have a major influence on human society and the biosphere. Natural climate variability is large, and its study can provide fundamental information about important climate processes, many of which will be relevant for understanding and predicting climate changes. On subseasonal timescales, improved predictions of extreme events like heat waves and droughts could reduce their negative impact via early preparation. On longer timescales, placing better constraints on ocean heat content changes will provide important dataset against which to test our climate models, and will

help reduce the uncertainty in estimates of climate sensitivity. Interactions between climate scientists and farmers, business owners, and policy makers are necessary to ensure that improved knowledge of our climate spreads beyond the ivory tower.

A

Relative Land Influence calculation

Relative Land Influence, m, is calculated from an ensemble of N Lagrangian atmospheric back trajectories from HYSPLIT (Draxler and Hess, 1997, 1998; Draxler, 1997) based on the meteorological fields from the NCEP-DOE reanalysis (Kanamitsu et al., 2002) as,

$$m = km^* = k\frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{t=0}^{28\text{ days}} Z(t)e^{-t/\tau}}{\sum_{t=0}^{28\text{ days}} e^{-t/\tau}}. \tag{A.1}$$

$Z(t)$ is defined to be equal to one over land and zero over ocean. The value of k is found from tuning the model to the seasonal cycle, as discussed in the main text. Parcels are initialized at 1000 meters above ground level, approximating the top of the boundary layer. For a given parcel trajectory, the influence of the previous land or ocean the parcel traveled over is weighted as a function of the time, t, between when it passed over that surface type and the present, and extending back to 28 days before the present. The ensemble of parcels are simulated from 2006-2010, and sample evenly across time of day and day of year; the values of Relative Land Influence used here are averaged across all trajectories, so retain no seasonality. See McKinnon et al.

(2013) for more detail.

The weighting and interpretation of the relaxation timescale, τ, can be understood through a simplified model. Consider a parcel of air moving atop a surface that can either be land or ocean. The atmosphere exchanges heat with the surface via a linear diffusivity, κ_z. For simplicity, surface temperature, T_s, is held fixed.

The energy balance of the parcel of air can be written as,

$$C\frac{dT}{dt} = \kappa_z(T_s - T), \tag{A.2}$$

where C is the parcel's heat capacity, and T is its temperature anomaly. In order to apply this model to the Lagrangian back trajectories, for which parcel location is recorded every three hours, the system is discretized in time as,

$$T_{i+1} = T_i + \frac{\kappa_z dt}{C}(T_{s,i+1} - T_i) = T_i + \frac{\kappa_z}{\kappa_x}(T_{s,i+1} - T_i), \tag{A.3}$$

where $\kappa_x = \frac{C}{dt}$ and reflects the strength of advective transport.

Assuming that the parcel begins in equilibrium with a surface condition, $T_{s,0}$, its temperature at time step n is,

$$T_n = (1 - \frac{\kappa_z}{\kappa_x})^n T_{s,0} + \frac{\kappa_z}{\kappa_x}\sum_{i=1}^{n}[(1 - \frac{\kappa_z}{\kappa_x})^{n-i} T_{s,i}]. \tag{A.4}$$

A given surface condition, $T_{s,i}$, is weighted in the final sum by $(1 - \frac{\kappa_z}{\kappa_x})^{n-i}$, where the index $n - i$ is the number of time steps since the parcel was over surface box i, and can be expressed as $\frac{-t}{dt}$, where dt is less than zero, consistent with the definition of t in Eqn. (A.1). The contribution of the initial condition, $T_{s,0}$, is small and neglected. Because this system is discretized, a stable solution requires that $0 < \frac{\kappa_z}{\kappa_x} < 1$. Note, however, that both κ_z and κ_x are a function of the spatial discretization scale, so the constraint is mathematical rather than physical.

The relative influence of a given land or ocean surface on the final parcel temperature is weighted by $w = (1 - \frac{\kappa_z}{\kappa_x})^{\frac{-t}{dt}}$, which can also be written as an exponential weighting, $w = \exp[\frac{-t}{\tau}]$, consistent with the exponential weighting using in Eqn. (A.1). The relaxation time, τ, is $dt/log(1 - \frac{\kappa_z}{\kappa_x})$. This representation indicates that τ is a measure of the relative strengths of horizontal and vertical coupling. In particular, the five day relaxation time used in this analysis reflects the relative influence of advective and surface heat fluxes, as opposed to the longer timescales that would be expected from radiative relaxation of the atmosphere (e.g. Goody and Yung, 1989; Donohoe, 2011).

B

Amplified warming in the tails of summertime maximum temperatures

Surface temperature is changing in response to anthropogenic forcing and natural variability. The magnitude of this change is often discussed in terms of the annual and global mean, but changes on smaller spatial and temporal scales can be more important for humans and the biosphere. In this paper, we explore the changes in peak summer daily temperature observed since 1979 over the United States. Since daily temperature is typically not normally distributed, we use quantile regression to calculate the trends in the daily data for different percentiles. The spatial patterns of change are spatially coherent, and generally show curtailment in the variability of temperature due a large warming of the lowest daily maximum summer temperatures. Furthermore, changes in the $5^{th}$and 95^{th}percentiles are significantly correlated to changes in the median across stations, with an amplification of both the $5^{th}$and 95^{th}percentiles relative to the median change, suggesting that changes in the tails of the temperature distributions may be able to be predicted from changes in the median. The subsequent text was primarily authored by Martin Tingley, with contributions from myself, as well as Andrew Rhines and Peter Huybers. My spe-

cific contributions entailed calculation and analysis of the quantile regression slopes, as well as the development of the Bayesian line fitting algorithm.

B.1 Abstract

The contribution of average warming to increases in high temperature extremes and decreases in low temperature extremes is well established, but the role of changes in other aspects of the temperature distribution is unclear. Using quantile regression, we explore trends in the median and tails of summertime daily maximum temperature observations from North American weather stations, and find spatially coherent linear relationships between them. Trends in the tails are significantly amplified and offset relative to changes in the median, demonstrating contraction in the lower half of the temperature distribution and variable changes in the width of the upper half. Particularly high correlation between trends in the upper tail and center ($r^2 = 0.90$) indicates the potential for improved projections of future high temperature extremes.

B.2 Introduction

The distribution of extreme climate events is sensitive to both changes in mean and variability (Katz and Brown, 1992; IPCC, 2007). Whereas trends in mean temperature are unequivocally contributing to the current increase in warm temperature extremes and decrease in cold temperature extremes (Meehl and Tebaldi, 2004; Meehl et al., 2009; Rahmstorf and Coumou, 2011; Hansen et al., 2012; IPCC, 2007), changes in other aspects of the temperature distribution are less well understood (Robeson, 2002; Schär et al., 2004; Rahmstorf and Coumou, 2011; Simolo et al., 2011; Rhines and Huybers, 2013; Huntingford et al., 2013; Alexander and Perkins, 2013).

Most recent studies have analyzed gridded or binned temperature estimates (Simolo et al., 2011, 2012; Donat and Alexander, 2012; Hansen et al., 2012; Rhines and Huybers, 2013; Tingley

and Huybers, 2013; Caesar et al., 2006; Brown et al., 2008; Mannshardt et al., 2013). Station density has not remained constant, however, and has even decreased over recent decades in certain regions, with the expected result being increased noise variance in the associated gridded temperature estimates (Kennedy et al., 2011; Rhines and Huybers, 2013; Huntingford et al., 2013). Gridding also tends to decrease the variability of the temperature signal relative to station-level estimates because of averaging across sub-gridscale features (Haylock et al., 2008; Zhang et al., 2011). The common practice of using a subset of the analyzed years for purposes of defining anomalies (Simolo et al., 2011, 2012; Donat and Alexander, 2012; Hansen et al., 2012; Brown et al., 2008; Caesar et al., 2006) can introduce further biases in estimates of trends in extremes (Rhines and Huybers, 2013; Tingley, 2012).

B.3 Temperature data and quantile regression

Here we assess linear trends in the distribution of daily summertime maximum temperatures (T_X) from more than 3800 stations in the Global Historical Climatology Network (GHCN-D, Menne et al., 2012) from the heavily sampled North American region. We analyze stations that are nearly complete over the 1979–2012 interval, which represents a compromise between interval length relative to data availability and the appropriateness of a linear model. This interval also permits for direct comparison of results with the National Center for Environmental Prediction's Reanalysis II product (NCEP II, Kanamitsu et al., 2002). Analyses are restricted to summertime conditions, defined here as the warmest 60-day period in the climatology averaged across all stations. The choice of 60 days represents a compromise between an interval short enough to reflect peak summer conditions and long enough to capture peak summer throughout the domain. Consistent results are obtained when using 50 or 70 day summer intervals (see Sec. B.8.7). All observations with negative quality control flags are removed, leaving a total of 7.5 million

temperature observations. While the GHCN-D data is not homogenized, the estimated uncertainties obtained from a spatially smoothed version of our results indicate that inhomogeneities do not influence our conclusions (see Sec. B.8.5).

To differentiate between trends in the center of the temperature distribution and those in the upper and lower tails, we apply quantile regression (Koenker and Bassett Jr, 1978) to separately estimate the 5^{th}, 50^{th}(median), and 95^{th} percentiles of the distribution as linear functions of the year, treating the T_X observations as exchangeable within each year. The present analysis entails two orders-of-magnitude more stations than previous quantile regression analyses of station data from Korea (Lee et al., 2013) and Central Europe (Barbosa et al., 2011), and unlike approaches employed elsewhere (Brown et al., 2008; Simolo et al., 2012; Mannshardt et al., 2013), does not require assumptions regarding the parametric form of the underlying probability distribution. Our focus is on the median, as opposed to the mean, because it is more amenable to disentangling changes in the center and tails of the distribution — for example, an increase confined to the top 10% of the distribution affects the mean but not the median.

Quantile regression results can be strongly biased when applied to rounded data, and the GHCN-D observations are rounded to units varying between 0.1-1°C (Fig. B.1). We circumvent this source of bias through a recommended procedure (Koenker and Bassett Jr, 1978; Machado and Silva, 2005) of adding uniform random noise realizations to approximately restore the data distribution to its original, continuous form. To determine the requisite range of the uniform distribution, a Hidden Markov Model is used to detect the degree of rounding in the GHCN-D observations (see Sec. B.8.4).

Significance of both the trends and the differences between them is assessed at each station using block bootstraps on years. As a second determinant of significance, we smooth the station results using a thin plate spline, and then assess significance of the trends and their differences using the standard errors of the estimated surfaces (see Sec. B.8.5). Discussion below focuses on

spatially cohesive structures that significantly differ from a null of no trend at the 95% significance level according to both approaches to uncertainty quantification.

B.4 Spatial patterns of temperature change

By way of a representative example of results (Fig. B.1), we consider the trends in the 5^{th}, 50^{th}, and 95^{th} percentiles at a weather station in Bluff, on the Colorado plateau of Southern Utah. Trends in all three percentiles are positive, with the 5^{th} the largest and the 50^{th} the smallest, consistent with a distribution with decreasing variance and increasing skew. The relationships between percentiles for Bluff are generally representative of stations that feature strong warming in the median. A more anomalous example, from Ukiah, 200 km north of San Francisco, features warming in the 5^{th} percentile, cooling in the 50^{th} percentile, and stronger cooling in the 95^{th} percentile (Fig. B.1).

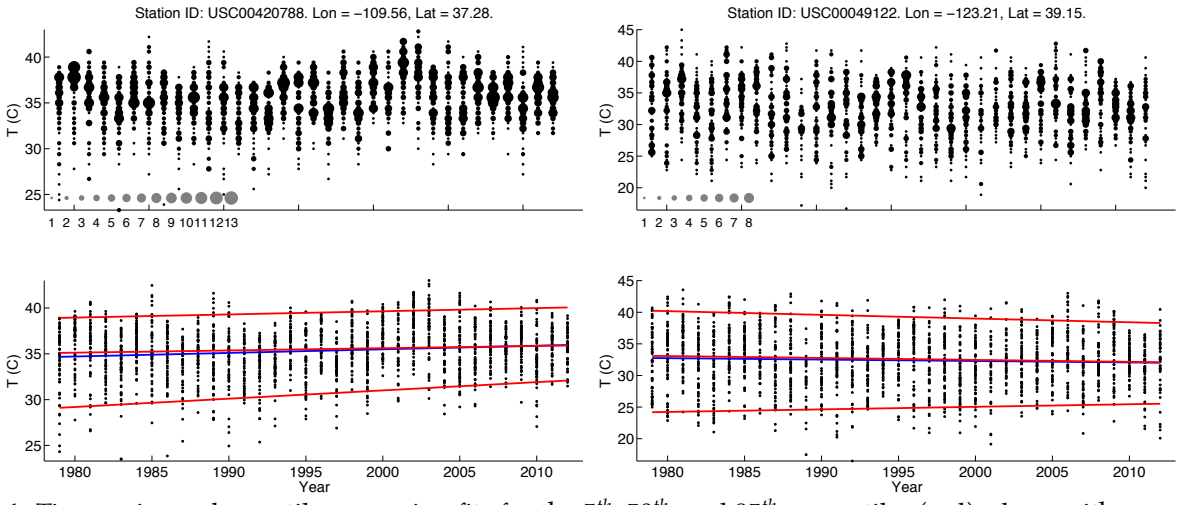

Figure B.1: Time series and quantile regression fits for the 5^{th}, 50^{th}, and 95^{th} percentiles (red), along with ordinary least squares fits to the mean (blue), for stations located in Bluff, UT (left) and Ukiah, CA (right). The top panels show the original GHCN-D data, with the size of the dots giving the number of superposed observations (numbered gray dots along the x-axis give the scale) and the vertical spacing between them the precision. The bottom panels display the data after adding random jitter to the observations, with the width of the uniform distributions determined separately for each observation.

Trends are generally positive for the 5^{th} percentile, with increases in excess of 1°C decade^{-1} in the mountainous West and along the Eastern Seaboard (Figs. B.2, B.10). Florida, Texas, and a swath roughly centered on the Mississippi River are exceptions, showing decreases in the 5^{th} percentile. The patterns of 50^{th} and 95^{th} percentile trends likewise feature generally positive values but with a swath of negative trends west of the Great Lakes and in the southern Atlantic states. Cooling in T_X in regions to the west of the Great Lakes has been reported elsewhere and attributed to changes in land use and land cover (Kalnay and Cai, 2003; Lobell et al., 2008; Diffenbaugh, 2009) as well as to regional feedbacks between circulation and soil moisture (Pan et al., 2004). Stations located near the margin of the West coast (e.g., the Ukiah station in Fig. B.1) also show negative trends for some or all of the three percentiles. Smoothing the trend maps accentuates and confirms the significance of large-scale patterns, though is not amenable to confirming the significance of smaller-scale patterns, such as the cooling seen along the margin of the West coast (Fig. B.10).

As with the trends themselves, there are large-scale spatial patterns in the differences between them, evident in both station-specific and smoothed maps (Figs. B.2, B.10). The largest magnitude differences occur between slopes in the 95^{th} and the 5^{th} percentiles, with differences in the trends indicating an increase in the spread of summer temperature in excess of 1°C decade^{-1} for stations in Texas, Oklahoma, and Kansas, and trends toward decreased spread of comparable magnitude in the West and along portions of the Eastern Seaboard. Such changes in spread map onto changes in variance, but the fact that the difference trends for the 95^{th}-50^{th} percentiles are small relative to those for the 5^{th}-50^{th} percentiles point to changes in skewness or other higher-order moments.

All three percentiles of the temperature distribution feature coherent regions of significant cooling interspersed within the generally warming trends (Figs. B.2, B.10, Table B.1). As the standard deviations of the trends across space are larger than the corresponding error standard devi-

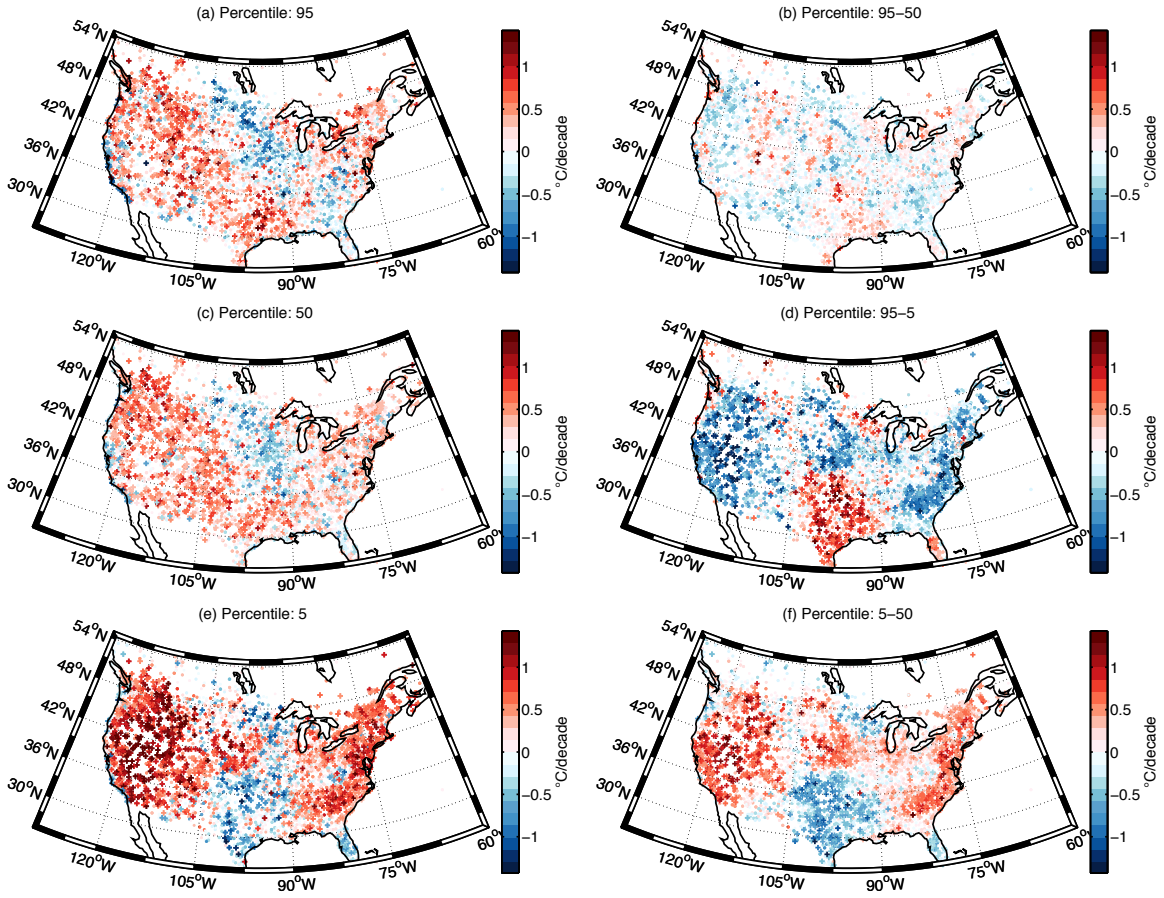

Figure B.2: Maps of station-level quantile regression trends and their differences, in °C decade^{-1}, for daily summertime T_X from GHCN-D. (a), (c), (e). Trends in the 95^{th}, 50^{th}, and 5^{th} percentiles. (b), (d), (f). The difference in trends between the 95^{th} and 50^{th} percentiles, the 95^{th} and 5^{th} percentiles, and the 5^{th} and 50^{th} percentiles. In all panels, large plus-signs indicate values that are significantly different from zero at the 95% level according to an annual block bootstrap, while small dots indicate stations that pass data-quality screenings but where the slope is not significantly different from zero.

ations in the estimation of these trends (Table B.1), the variability of the estimated trends about the spatial mean is not consistent with estimation uncertainty alone. Analyses based on binning or aggregating data in space (Caesar et al., 2006; Brown et al., 2008; Simolo et al., 2011, 2012; Donat and Alexander, 2012; Hansen et al., 2012; Rhines and Huybers, 2013) are, therefore, likely to average over disparate phenomena and obtain diluted results.

Table B.1: Means and standard deviations of the station-wise GHCN-D quantile regression slopes (in $^{\circ}C/\text{decade}^{-1}$) and the differences between the slopes, along with a pooled estimate of the residual bootstrap standard deviation, calculated as the square root of the mean of the bootstrap variances. Also shown are means and standard deviations for the spatially smoothed maps of the GCHN and NCEP II analysis, along with the mean of the standard errors in the estimated smooth surface. The final two rows show correlations and root mean square errors (RMSE) between the smoothed GHCN-D and NCEP II fields. All calculations for the smoothed fields are area weighted, while station-wise results are not area weighted.

		95^{th}	50^{th}	5^{th}	$95^{th} - 50^{th}$	$95^{th} - 5^{th}$	$5^{th} - 50^{th}$
GHCN-D	Mean	0.11	0.14	0.28	−0.03	−0.17	0.14
(stations)	Std.	0.41	0.35	0.53	0.24	0.50	0.40
	σ_{BSr}	0.34	0.25	0.36	0.26	0.41	0.29
GHCN-D	Mean	0.14	0.17	0.28	−0.03	−0.14	0.10
(smoothed)	Std.	0.22	0.17	0.39	0.14	0.37	0.33
	SE	0.10	0.08	0.13	0.08	0.13	0.10
NCEP	Mean	0.03	0.18	0.37	−0.14	−0.34	0.20
(smoothed)	Std.	0.46	0.44	0.74	0.30	0.59	0.51
	SE	0.08	0.08	0.17	0.08	0.16	0.14
Correlation		0.51	0.33	0.53	0.42	0.48	0.55
RMSE		0.39	0.39	0.55	0.25	0.48	0.37

B.5 Relationships between quantiles

Promising, in terms of ascribing a simple interpretation to the overall changes in the distribution of summertime maximum temperatures, are strong linear relationships between the spatial maps of the three trends (Fig. B.3). For instance, the estimated squared-cross-correlation between trends in the 95^{th} and 50^{th} percentiles is 0.90. To estimate linear relationships between

percentiles, we use a Bayesian line fitting procedure that accounts for correlated errors in the estimated slopes (see Sec. B.8.6). This approach avoids regression dilution (Frost and Thompson, 2000) and yields more accurate estimates of the linear relationships between trends in the median and those in the 5^{th} or 95^{th} percentile.

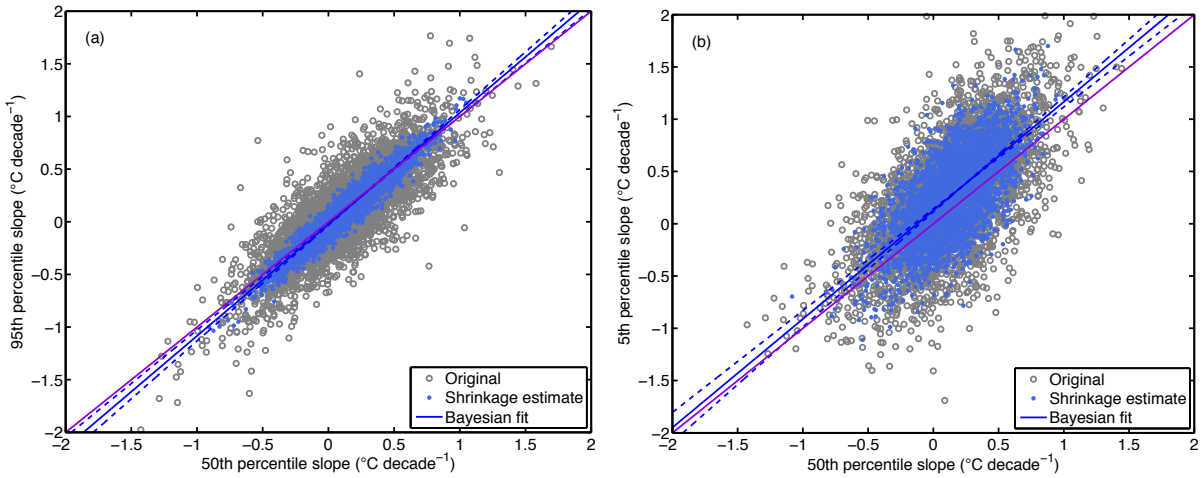

Figure B.3: Regression of trends in (a) the 95^{th} percentile and (b) the 5^{th} percentile onto trends in the 50^{th} percentile. Gray circles depict the estimated trends in the percentiles for each station (compare with Fig. B.2), while the blue dots are a draw of the shrinkage estimates of the corresponding latent slopes, based on a partitioning of variance between observational errors and the variability of the true –as opposed to estimated– trends (see supplementary methods). Bayesian regression fits to the latent slope processes, along with 95% uncertainties are shown in blue, and as a reference, the one-to-one line is shown in purple. Numerical values of the intercepts and slopes are given in Table B.2.

Regressing trends in the 95^{th} percentile onto those in the 50^{th} percentile yields a slope of 1.06 ± 0.04 °C/°C and an intercept of -0.024 ± 0.010 °C decade $^{-1}$ (95% intervals; see Fig. B.3, Table B.2). Opposite from previous findings (Caesar et al., 2006), the strength of the linear relationship ($r^2 = .90$) points towards the practicality of inferring changes in warm extremes by scaling changes in the center of the distribution (Santer et al., 1990). The larger than unity slope demonstrates an amplification of trends in warm extremes with respect to trends in the median, suggesting a mechanism whereby increases in maximum daily temperature lead to a disproportionately larger influence on the hottest summer days.

Table B.2: Posterior medians of the slopes (in °C/°C) and intercepts (in °C decade $^{-1}$) for regressions of trends in the 95^{th} and 5^{th} percentiles onto trends in the 50^{th} percentile (Fig. B.3), using the GHCN-D and NCEP II data sets. Also shown are estimates of the r^2 value for each regression. 95% central posterior intervals are included as a measure of uncertainty for all parameters. For the GHCN-D analysis, results are presented using all stations (*all*), only those stations with constant precision throughout the time interval (*precision*), and for the 1000 stations with constant precision that feature the most complete data availability (*data*). Results are qualitatively similar using 500 or 1500 stations.

		95^{th} on 50^{th}	5^{th} on 50^{th}
GHCN-D *(all)*	Intercept (95% interval)	-0.024 (-0.034, -0.013)	0.127 (0.113, 0.143)
	Slope (95% interval)	1.06 (1.02, 1.10)	1.04 (0.97, 1.10)
	r^2 (95% interval)	0.90 (0.87, 0.93)	0.47 (0.42, 0.50)
GHCN-D *(precision)*	Intercept (95% interval)	-0.028 (-0.037, -0.018)	0.125 (0.110, 0.141)
	Slope (95% interval)	1.06 (1.01, 1.10)	1.04 (0.97, 1.10)
	r^2 (95% interval)	0.91 (0.88, 0.93)	0.46 (0.41, 0.50)
GHCN-D *(data)*	Intercept (95% interval)	-0.045 (-0.067, -0.024)	0.160 (0.127, 0.193)
	Slope (95% interval)	1.10 (1.01, 1.20)	0.98 (0.83, 1.13)
	r^2 (95% interval)	0.94 (0.85, 0.97)	0.37 (0.28, 0.46)
NCEP II	Intercept (95% interval)	0.054 (0.036, 0.073)	-0.011 (-0.038, 0.017)
	Slope (95% interval)	0.74 (0.54, 0.90)	1.98 (1.79, 2.17)
	r^2 (95% interval)	0.56 (0.32, 0.71)	0.90 (0.83, 0.94)

One possibility is that higher temperatures are associated with lower soil moisture and a disproportionate loss of evapotranspirative cooling on the hottest summer days (Seneviratne et al., 2010). Also relevant is that the intercept term in the regression significantly differs from zero, indicating the presence of an overall cooling influence in the 95^{th} percentile irrespective of the slope in the median. Such an offset is indicative of a temporally intermittent mechanism; one

possibility is that circulation patterns associated with summer heat waves have altered (Loikith and Broccoli, 2012, 2014).

The competing effects of amplification and offset lead to separate regimes, featuring, respectively, increasing and decreasing trends in the 95^{th} percentile relative to the median. Only those stations having median warming trends exceeding 0.42 °C decade^{-1}, or 19% of all stations, are expected to feature excess warming in the 95^{th} percentile relative to the 50^{th} percentile (Table B.1). These competing effects are also apparent in the maps of the trends and their differences (Figs. B.2, B.10), which show amplification of the 95^{th} percentile trend in regions where the trend in the 50^{th} percentile is large in magnitude, and a general shift towards lower values everywhere else. Note that this amplified and offset relationship between trends in the percentiles would be obscured through regional averaging.

Regressing trends in the 5^{th} percentile onto those in the 50^{th} percentile yields a best estimate slope that is slightly greater than unity (1.04 ± 0.06 °C/°C) and a significant, positive intercept of 0.127 ± 0.015 °C decade $^{-1}$ (Fig. B.3, Table B.2). The marginally greater than unity slope in combination with a strong positive intercept leads to a general expectation of contraction of the lower tail, though there are regional departures. The Gulf Coast region, for instance, features larger slopes in the 50^{th} percentile than the 5^{th} percentile (Figs. B.2, B.10), possibly because it features peak summer conditions earlier than other locations. These regional departures are consistent with there being only moderate correlation between 50^{th} and 5^{th} percentile trends ($r^2 = 0.47$). The strength of the intercept points to a temporally intermittent mechanism associated with relatively cold summer days. A possible mechanism is amplified polar warming (Screen and Simmonds, 2010; Bekryaev et al., 2010) leading to preconditioning of air masses associated with cold anomalies in the mid-latitudes toward warmer temperatures.

Results of the two regressions together indicate that the lower half of the distribution contracts for all trends in the median, and the upper half either contracts or, for 19% of stations, expands

at a slower rate than the contraction in the lower half. Such changes in the distribution are consistent with median warming leading to an overall decrease in variance and an increase in skew.

Analysis has focused on the median, as opposed to the mean, because it permits for isolating changes in the center of the distribution from those in the tails. For completeness, however, we note that the slope is similar when regressing trends in the 95^{th} percentile onto trends in the mean, while the intercept is smaller (-0.051 ± 0.008 °C decade $^{-1}$; Table B.3, Fig. B.4). The smaller intercept is associated with strong trends in the lower tail uniformly increasing the trends in the mean; such influences from the lower tail point to the utility of using the median as a measure of center. For similar reasons, regressing trends in the 5^{th} percentile onto trends in the mean yields a smaller intercept (0.073 ± 0.016 °C decade $^{-1}$). There is also increased correlation between mean and 5^{th} percentile trends ($r^2 = 0.77$; Table B.3), resulting in a steeper estimated slope (1.22 ± 0.06 °C/°C). Indeed, the regressions on the mean feature higher r^2 values in both cases, consistent with trends confined to a single tail affecting mean but not median trends.

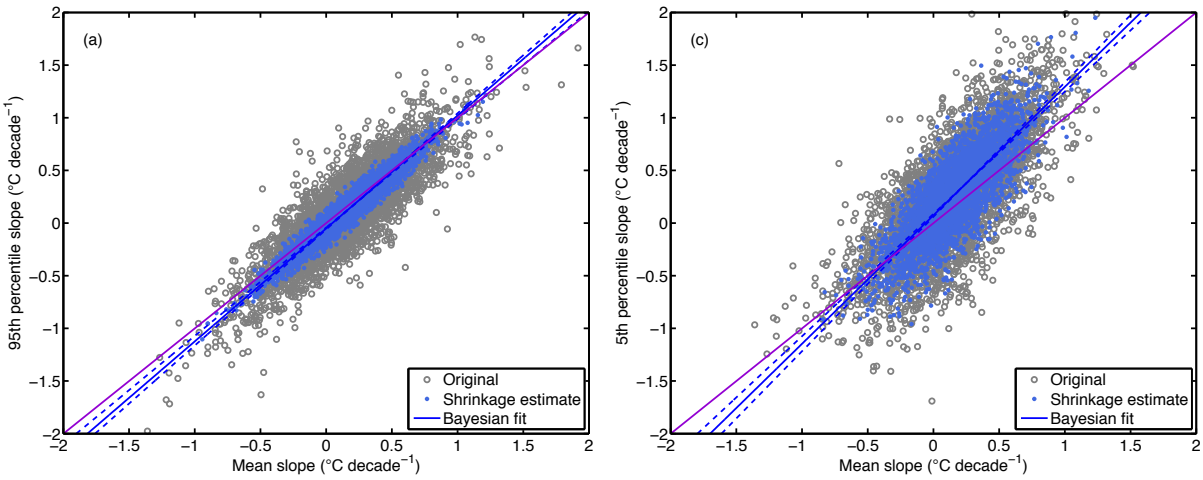

Figure B.4: As in Fig. B.3, but using the OLS trends in the mean to predict trends in the (a) 95^{th} and (b) 5^{th} percentiles. Numerical values of the intercepts and slopes are given in Table B.3.

Table B.3: Posterior medians of the slopes (in °C/°C) and intercepts (in °C decade^{-1}) for regressions of the trends in each of the three percentiles onto trends in the mean (Fig. B.4), using all stations in the GHCN-D data set, along with estimates of the r^2 value for each regression. 95% central posterior intervals are included as a measure of uncertainty for all parameters.

		95^{th} on mean	50^{th} on mean	5^{th} on mean
GHCN-D	Intercept	-0.051	-0.019	0.073
(all)	95% interval	$(-0.059, -0.043)$	$(-0.023, -0.015)$	$(0.056, 0.087)$
	Slope	1.07	0.97	1.22
	95% interval	$(1.03, 1.10)$	$(0.96, 0.99)$	$(1.16, 1.28)$
	r^2	0.966	0.990	0.774
	95% interval	$(0.953, 0.977)$	$(0.987, 0.992)$	$(0.747, 0.801)$

B.6 Comparison to trends in NCEP II reanalysis

The spatially smoothed results from the GHCN-D analysis have only a weak correspondence with results obtained from applying the same quantile regression analysis to the NCEP II reanalysis (Kanamitsu et al., 2002). Spatial cross-correlations between the smoothed GHCN-D and smoothed NCEP II estimates of the six trend or difference fields range from only 0.33 to 0.55 (Table B.1). The root mean squared differences between the smoothed station data and reanalysis are at least three times larger than the estimated standard errors in the smoothed station fields (Fig. B.5, Table B.1), leading us to conclude that the reanalysis contains substantial biases. The apparent biases in the NCEP II reanalysis may be introduced through changes in observational coverage (Bengtsson et al., 2004) or unaccounted for changes in land use (Kalnay and Cai, 2003). Similarly, spurious spatial structure may be introduced though approximations for surface boundary conditions, such as those involving topography (Rusticucci and Kousky, 2002).

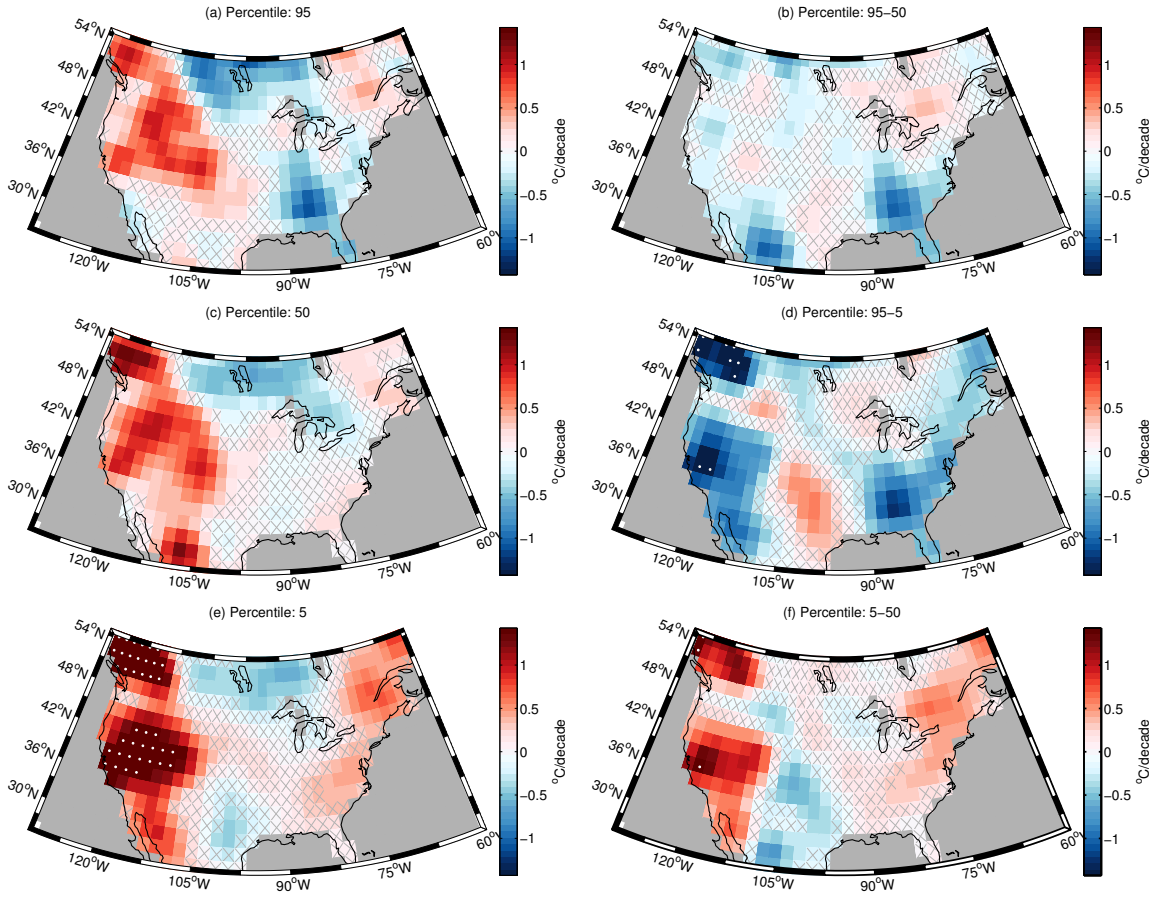

Figure B.5: As in Fig. B.5, but for an analysis of the daily maximum of the four times daily 2m air temperature from the NCEP II reanalysis. Results are smoothed in the same manner as the GHCN-D results to facilitate comparisons. White dots demarcate locations with values outside the bounds of the color scale.

These results also suggest that analyses of extremes using other reanalysis products may contain similar biases relative to station data (Huntingford et al., 2013). See Sec. B.8.8 for further discussion.

B.7 Conclusion

Through applying quantile regression to station-level temperature data, we identify significant trends in the 5^{th}, 50^{th}, and 95^{th} percentiles of summertime T_X. Furthermore, and in contrast to previous work (Caesar et al., 2006; Brown et al., 2008; Huntingford et al., 2013), including our own (Rhines and Huybers, 2013; Tingley and Huybers, 2013), we identify large scale patterns of significant differences between trends in the tails and center of the summertime T_X distribution. A particularly strong linear relationship exists between trends in the median and upper tail of the distribution, indicating a general amplification of trends in the hottest summer temperatures relative to the median, tempered by an overall offset toward cooling. With the caveat that trust-worthy predictions also require an understanding of the offset in trends, the general amplification points to the practicality of using model projections that capture changes in the center of the temperature distribution to predict changes in warm extremes.

B.8 Supplementary methods and discussion

B.8.1 Data

We use daily T_X observations from the Global Historical Climatology Network (GHCN-D) (Menne et al., 2012) database in the densely sampled North American region, from 130°W to 60°W and 25°N to 55°N. Previous studies (Barbosa et al., 2011; Caesar et al., 2006; Brown et al., 2008) have sought to remove the seasonal cycle prior to examining trends in extremes, but because the variance and higher-order moments of temperature anomalies about the seasonal cycle themselves depend on the season (Huybers and Curry, 2006), and the phase and amplitude of the seasonal cycle are time- and space-dependent (Thomson, 1995; Stine and Huybers, 2012; McKinnon et al., 2013), establishing an appropriate algorithm for removing seasonality is non-trivial, particu-

larly if one seeks to differentiate between changes in the center as compared to the tails of the temperature distribution. Here we take the simpler approach of only considering summer conditions, taken as the 60 day interval from calendar day 175 to 234, inclusive (June 24 to August 22 for a non-leap year), which is the warmest 60-day interval in the climatology defined by the area-weighted average across all stations and across all years between 1979-2012. Section B.8.7 discusses results under 50 and 70 day summer intervals.

Any observation featuring a negative quality control flag in the GHCN-D metadata is discarded from the analysis, and stations are further screened for data availability using the criteria that included records must contain at least 20% of summer observations per decade and near-complete (80%) summers in at least two years per decade. A total of 3894 stations in the region meet these data quality and availability criteria, constituting approximately 7.5 million individual observations.

B.8.2 Quantile regression

Formally, quantile regression (Koenker and Bassett Jr, 1978; Koenker and Hallock, 2001; Cade and Noon, 2003) proceeds as a minimization problem under a tilted absolute value loss function, with the angle determining the quantile, τ (Fig. B.6). For values of $\tau > 1/2$, the tilt in the absolute value loss function results in positive residuals contributing more to the loss function than negative residuals, and vice versa for values of $\tau < 1/2$. In contrast, ordinary least squares regression is based on minimizing a symmetric, quadratic loss function. The quantile regression minimization problem is generally solved via a linear programming algorithm (Koenker and Bassett Jr, 1978), and we make use of the Matlab and R implementations of the quantile regression algorithms provided by Roger Koenker (`www.econ.uiuc.edu/~roger/research/rq/rq.html`).

Quantile regression shares a first-order equivalence with least squares regression on empirical quantiles calculated from data subsets defined by the predictor variable (Bassett Jr et al., 2002).

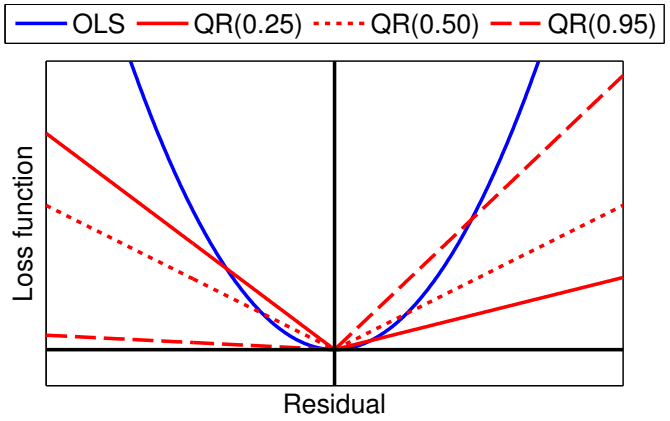

Figure B.6: Loss functions for ordinary least squares (OLS) regression (blue), and for quantile regression (QR) for the .25, .50, and .95 quantiles (red). Axes are intentionally blank as it is the shape of the loss functions that is important.

Indeed, past studies have taken the latter approach (Caesar et al., 2006), but pooled data across years to increase the number of observations available for calculating empirical quantiles. As our interest is in trends in the tails at the annual time scale, we view quantile regression as a preferred approach.

B.8.3 Estimating conditional quantiles from rounded observations

Presentations of quantile regression generally assume that the conditional distribution function is continuous (Koenker and Bassett Jr, 1978). The GHCN-D observations are rounded, with the level of precision varying both between records and over time within a given record (Fig. B.1). Although the level of rounding is not independently recorded, exploratory analysis suggests that the most precise observations within the subset of the database that we analyze are rounded to the nearest 0.1 °C and the least precise to 1°C. Stations from the United States are rounded to 1°F and then converted to °C.

Estimates of medians from discrete or rounded observations have different properties than estimates of means, and these properties are inherited by estimates of quantile regression pa-

rameters. As illustration, consider the distribution of the sample mean ($\hat{\mu}$) and median ($\hat{M}$) of $2N + 1$ independent, random draws from $\mathcal{U}(0.5, 2.5)$, where $\mathcal{U}$ denotes the uniform distribution and the sample size is forced to be odd for convenience. $\hat{M}$ is then the $N + 1$ order statistic and, therefore, follows a shifted-and-scaled beta distribution. The probability density function for the sample median, like that for the sample mean, is continuous with support on the entire $(0.5, 2.5)$ interval. If the observations are rounded to the nearest integer, the resulting sample median, $\hat{M}_r$, can take on only the values one or two, each with probability 0.5—regardless of the sample size. Rounding adds coarseness to estimates of the median that, in the large sample limit, is not present for estimates of the mean.

As with estimates of the median, rounding results in slopes estimated from quantile regression preferentially falling on particular values, at the expense of near-by regions of the real line (Figs. B.7, B.8ab). These biases can be corrected by imposing a degree of smoothness on the rounded or discrete observations by "jittering" them; that is, by adding independent random noise (Koenker and Bassett Jr, 1978; Machado and Silva, 2005). The jitter should, ideally, exactly match the distribution of values associated with the process of rounding, as the aim is to probabilistically undo the rounding through realizing a continuous distribution that preserves ordering of the data up to the precision of the rounding. In the median estimation problem discussed above, for example, adding a draw from a $\mathcal{U}(-0.5, 0.5)$ to each rounded sample restores the original distributional properties of the sample median.

Adding zero-centered uniform jittering, with width given by the rounding interval, generally suffices to counter the distributional effects of rounding, assuming that the observations have been rounded according to convention and that the rounding interval is small relative to the variability of the temperature data distribution. For the daily GHCN-D T_X observations, the half-width of the jittering noise is at most 0.5°C, while the standard deviation of daily T_X at any location is generally an order of magnitude larger, justifying the use of a uniform jittering distri-

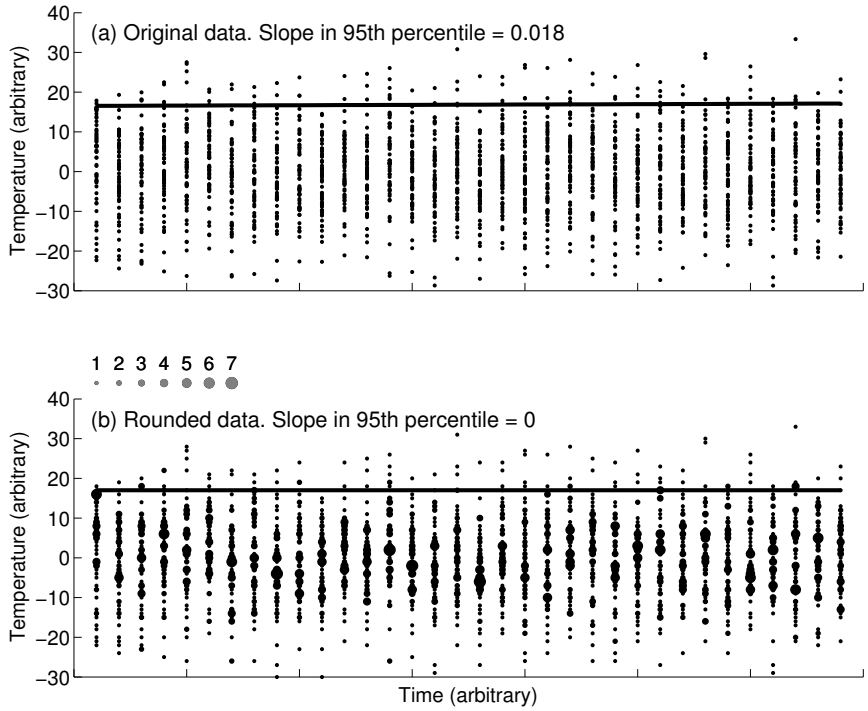

Figure B.7: The impact of rounding on quantile regression, for normal errors on a weak linear trend. (a) Original data, formed as IID $\mathcal{N}(0, 10^2)$ samples added to a trend of 0.01 'temperature' units per 'time' unit. To match the GHCN-D data used in the analysis, there are 60 'temperature' observations at each of 34 time intervals. The estimated slope in the 95^{th} percentile for this sample is 0.018. (b) The data from panel (a), rounded to the nearest integer, with the size of the dots giving the number of superposed observations. The estimated slope in the 95^{th} percentile for the rounded data is 0 to machine precision.

bution. In addition, a uniform jittering distribution avoids the necessity of modeling the distribution of the underlying, continuous data-generating process (Reich and Smith, 2013).

The efficacy of our jittering approach can be examined with respect to example data sets with characteristics specified to mimic the actual GHCN-D stations. An example realization is produced as mean-zero Gaussian noise with standard deviation of 10 superimposed on a linear trend of 0.01 per unit time. The length of the interval is 34, and there are 60 observations for each time unit, approximating the degrees of freedom for each GHCN-D station. Quantile regression slopes are then estimated for the original data; after rounding all observations to the nearest inte-

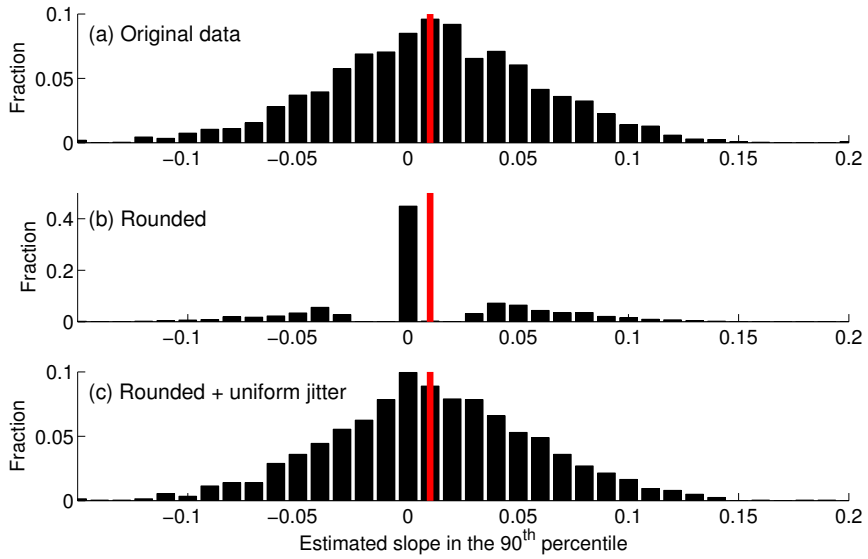

Figure B.8: The effects of rounding and subsequent jittering on estimates of quantile regression slopes. (a) Histogram of estimated slopes in the 95^{th} percentile for 2000 data samples produced as in Fig. B.7a. The true slope value of 0.01 used to create the data is marked in red in all panels. (b) As in (a), but each data point is rounded to the nearest integer prior to estimating the slope. (c) As in (b), but $\mathcal{U}(-1/2, 1/2)$ random jitter is added to each rounded observation prior to estimating the slope.

ger; and after then adding independent $\mathcal{U}(-0.5, 0.5)$ draws to each rounded observation. Under repeated sampling, a two-sample Kolmgorov-Smirnov test comparing the distribution of slopes from the original and rounded data clearly rejects the null hypothesis that the samples follow the same distribution, with a p-value less than 0.001. The jittering, however, produces a distribution of slopes that is indistinguishable from that of the original observations (Fig. B.8), with the Kolmgorov-Smirnov test p-value of 0.69 giving no indication that the distributions are different.

Standard analytical treatments of the uncertainties associated with quantile regression also assume continuity of the conditional density function. Continuity facilities asymptotic approaches to estimating functional forms for standard errors and normal approximations to the quantile regression estimator (Koenker and Bassett Jr, 1978; Simpson et al., 1987; Knight, 1998). The rounded nature of the GHCN-D data and the expectation that daily temperature observations display au-

tocorrelation renders standard approaches difficult to implement, and we instead quantify uncertainty using a block bootstrap procedure, discussed below.

B.8.4 Determining the precision of the GHCN-D data

It is necessary to infer the precision, or degree of rounding, in the GHCN-D data set directly from the observations. Initial examination of the time series of first differences reveals that observations are rounded to the nearest 0.1°C, 0.5°C, 1°C, or to the nearest 1°F or 2°F prior to the GHCN-D aggregation to the nearest 0.1°C (Durre et al., 2010). There are some rare exceptions to these differences that are likely attributable to errors either in the reading, recording, or transferring of data. Other potential issues with station data, including the preferential occurrence of certain values of the last digit of the observations (Nese, 1994), are not addressed here.

We model the level of precision for each observation using a Hidden Markov Model (Wikle and Berliner, 2006), defining the state vector in units of °C as $S = (0.1,\ 0.5,\ 1.0,\ 5/9,\ 10/9)$. A prior transition matrix is generated by assuming that each state will persist from day to day with probability α, and will transition with equal probability to any of the other states. Depending on which state is active, it is possible to observe different subsets of temperature levels, and for each observation we encode this information as a consistency vector. The emissions matrix then maps the states in S to their likelihood of emitting an observation with each possible consistency vector under 0.1°C rounding. The Baum-Welch algorithm (Rabiner, 1989), based on expectation-maximization, is first used to estimate the true transition probabilities from the data and the transition prior, and the Viterbi algorithm (Forney Jr, 1973) is then used to recursively solve the hidden Markov model for the optimal state path of the data, given the estimated transition probabilities and the emission matrix.

As the linear trends for the GHCN-D stations tend to be small relative to the variability, the discretization of the observations results in the expected spurious structures in histograms of

the quantile regression slopes across stations (Fig. B.9). In particular, there is substantial zero-inflation at the expense of small, non-zero values. The addition of uniform random noise to the observations, with parameters determined according to the inferred level of rounding, ameliorates the zero inflation and restores the distribution of slopes to approximate normality (Fig. B.9).

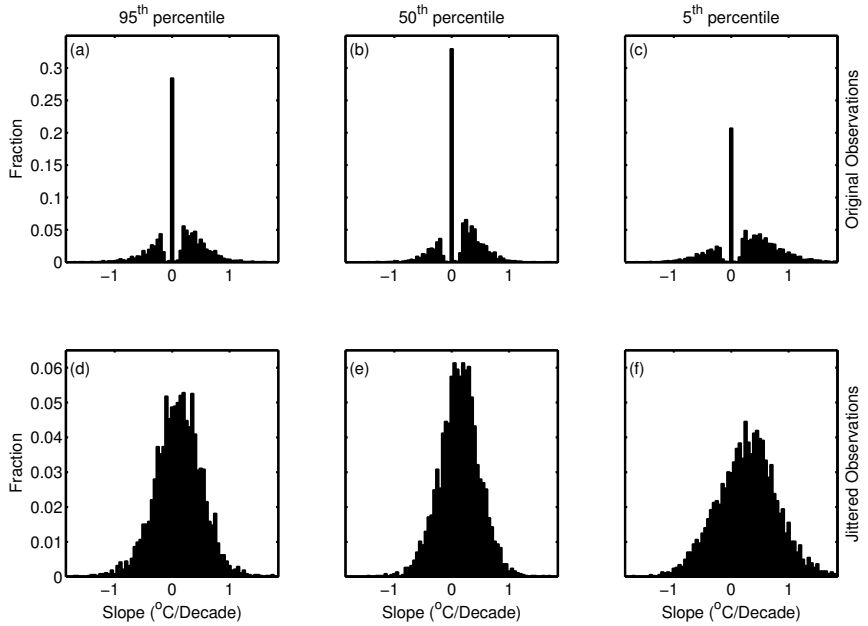

Figure B.9: Slopes, in °C decade^{-1}, for quantile regression fits to the daily summer T_X data from GHCN-D. (a)–(c) Slopes in the 5^{th}, 50^{th}, and 95^{th} percentiles of the original data. (d)–(f) Slopes in the 5^{th}, 50^{th}, and 95^{th} percentiles after adding random uniform jitter to the original data. Note the change in vertical scaling between the top and bottom row.

B.8.5 Uncertainty quantification: bootstrapping and spatial mapping

The distributions of station-level slopes consistent with the data are assessed using annual block bootstraps (Barbosa et al., 2011; Lee et al., 2013), and we additionally disentangle the uncertainty due to the jittering outlined above. To assess the uncertainty due to the precision of the observations, we calculate the variances of the slope estimates under 1000 independent realizations of

the jittering procedure. We assess uncertainty due to the availability and variability of the observations by forming bootstrap samples for each station by resampling, with replacement, entire years of observations, and then calculating slopes and differences for the three percentiles, with the jittering realization held constant. In effect, the bootstrap randomizes the years, with replacement, while preserving any temporal correlation structure within each year, and the bootstrap variance is an order of magnitude larger than that from the jittering. The overall uncertainty is formed from the sum of the two variances, and is strongly dominated by the bootstrap uncertainty. Slopes and differences that lie outside of the corresponding 95% central uncertainty intervals are considered significant, in the sense that the value is larger in magnitude than would be expected under a random resampling of the years (Fig. B.2). Significance in this context is not a statement of anthropogenic attribution, but rather that there is low-frequency structure inconsistent with annual samples that are independent of one another.

A second block bootstrap is performed on the residuals about the initial fits, and the estimated error standard deviations and correlations associated with the estimated slopes used in the regression analysis between trends in different percentiles (Fig. B.3 and see subsequent). Whereas the first bootstrapping procedure aims to assess the significance of the linear trends with respect to a null assumption of independence between years, the second quantifies the uncertainty in the model parameter estimates, and generally produces smaller uncertainties.

As an alternate assessment of spatial patterns and uncertainty, we smooth the station-level results and spatially map them onto the approximately 2° latitude by 2° longitude grid associated with the NCEP II Reanalysis product (Kanamitsu et al., 2002). The spatial smoothing is performed using `fastTPS` within the R (Ihaka and Gentleman, 1996) package `fields` (Furrer et al., 2013). `fastTPS` is a spatial mapping scheme based on the compactly supported, isotropic and stationary Wenland covariance function, and we set the spatial range to $\theta = 750$ miles (great circle distance), so that the estimated, smoothed value at each location is affected by all values

within 750 miles. Results are visually similar for larger (1000) and smaller (500) values of θ, and we choose a value of 750 miles to accentuate the large-scale features evident in the station-level results (Fig. B.2), at the expense of smaller-scale features, such as those seen on the West Coast. `fastTPS` uses a cross validation procedure to determine the partitioning of variance between the smooth interpolating surface and the spread of the observations around the smooth surface. The uncertainty in the estimate of the smoothed field at each location is determined by the data availability in the surrounding area, as well as by the overall partitioning of variance. The standard errors associated with the estimated smoothed surface provides an independent means of quantifying uncertainty, and we consider as significant those grid boxes with values that are more than 1.96 standard errors from zero, to be consistent with the 95% uncertainty range used for the station data (Fig. B.10).

These two disparate approaches to uncertainty quantification both display robustness to the presence of inhomogeneities (Menne et al., 2012) in the original station level data series. Correcting step-function inhomogeneities, such as those arising from changes in instrument location, type, or observation protocol, will generally decrease the variance of the original series (Figs. 4, 5 of Menne and Williams Jr, 2009). The presence of inhomogeneities will therefore, modulo pathological instances where the inhomogeneities lead to a better linear fit, increase variance about any climatic trends, and as such increase the bootstrap standard deviations used to characterize uncertainty in the station-level results. The contributions of such stations are then down weighted visually (Fig. B.2) and in the assessment of the relationships between trends in the three quantiles (see subsequent).

Inhomogeneities will also, in general, make the trends at near-by stations more distinct from one another – indeed, the assumed similarly of near-by records is the main diagnostic for identifying and correcting inhomogeneities (Menne and Williams Jr, 2009). Provided the presence and magnitude of spurious slopes arising from inhomogeneities are uncorrelated with the large-

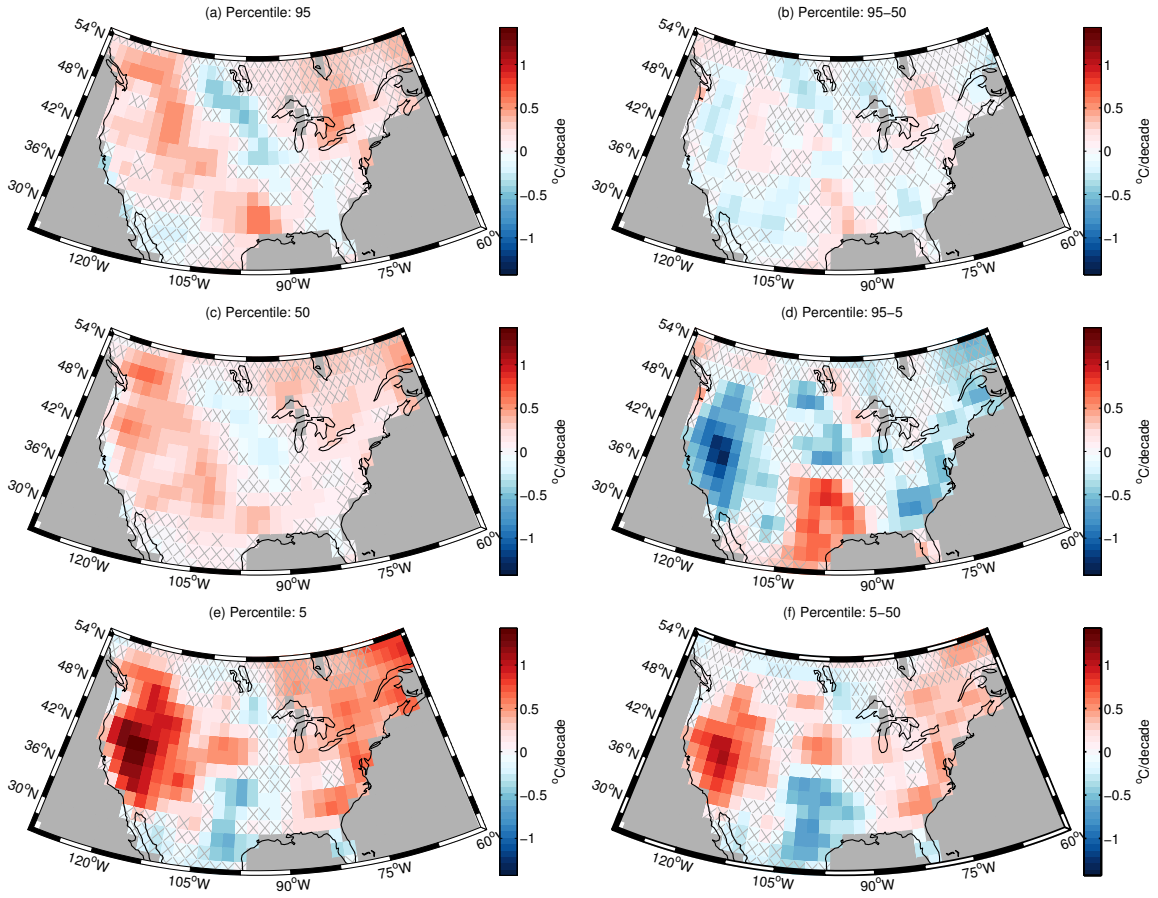

Figure B.10: Spatially smoothed maps, interpolated onto the NCEP II grid, of quantile regression trends and their differences, in °C decade^{-1}, for daily summertime T_X from GHCN-D. (a), (c), (e). Trends in the 95^{th}, 50^{th}, and 5^{th} percentiles. (b), (d), (f). The difference in trends between the 95^{th} and 50^{th} percentiles, the 95^{th} and 5^{th} percentiles, and the 50^{th} and 5^{th} percentiles. In all panels, gray hashing indicates locations where the smoothed field is within 1.96 standard errors of zero, so that an absence of hashing indicates locations that are significantly different from zero at the 95% level.

scale spatial variability that is accentuated by the spatial smoothing (Fig. B.10), inhomogeneities will tilt the variance partitioning inherent to the smoothing more towards variability about the smooth surface, and will likewise increase the standard errors in the estimated smooth surface. For all three percentiles, the standard deviation of the smooth surface is at least twice as large as the average standard error in the estimation of the smooth surface (Table B.1), indicating that the

variability in the smooth field is substantially larger than the uncertainty in its estimation.

B.8.6 Fitting straight lines to scatter plots of noisy observations

In assessing the relationships between linear trends in the 50^{th} and 95^{th} percentiles, and the 5^{th} and 95^{th} percentiles, we adopt a Bayesian regression formalism that takes into account the error covariances in the estimated slopes and models the values of the independent variable as draws from a probability distribution (as opposed to being set in advance as part of the experimental design). The explicit modeling of observational uncertainty in the independent variable avoids regression dilution – underestimation of the slope arising from noise in the independent variable – while modeling of the observational uncertainty in the dependent variable reduces uncertainty in the estimation of the regression parameters. We defer general discussions of regression dilution (Frost and Thompson, 2000; Tingley and Huybers, 2010), and the difficulties of straight-line fitting in the presence of errors in both variables (Brown, 1993; Christiansen, 2013), in favor of presenting the specific model used to fit the regressions.

Let $x_o(\mathbf{s}_i)$ denote the estimated slope at location $\mathbf{s}_i$ for the percentile treated as the independent variable, and $x_t(\mathbf{s}_i)$ the corresponding 'true', or latent, value of the slope at this location – that is, we acknowledge the estimated slope is an imperfect estimate, due to the limited sample size. Likewise, let $y_o(\mathbf{s}_i)$ and $y_t(\mathbf{s}_i)$ denote the observed and true values of the slope at location $\mathbf{s}_i$ for the percentile treated as the response. We seek to fit a model of the form,

$$\begin{aligned} y_t(\mathbf{s}_i) &= \alpha + \beta \cdot x_t(\mathbf{s}_i) + \varepsilon_y \\ \varepsilon_y &\sim \mathcal{N}(0, \tau^2), \end{aligned} \tag{B.1}$$

where the primary complication is that only the $x_o(\mathbf{s}_i)$ and $y_o(\mathbf{s}_i)$ are available, while the regression is in terms of $\mathbf{y}_t$ and $\mathbf{x}_t$.

The aim of the analysis is thus to find the conditional distribution, $\pi(\alpha, \beta, \tau^2 | \mathbf{y}_o, \mathbf{x}_o)$, of the regression parameters, α, β, τ^2, given the observations, $\mathbf{y}_o, \mathbf{x}_o$, where π denotes a probability distribution. We do so via a Bayesian approach, specifying an improper, uninformative prior for the three regression parameters, $\pi(\alpha, \beta, \tau^2) \propto 1/\tau^2$, corresponding to flat priors on α and β, and the Jeffery's prior on τ^2 (Gelman et al., 2003). Bayes' rule then gives,

$$\pi(\alpha, \beta, \tau^2 | \mathbf{y}_o, \mathbf{x}_o) \propto f(\mathbf{y}_o, \mathbf{x}_o | \alpha, \beta, \tau^2) \cdot \pi(\alpha, \beta, \tau^2) \tag{B.2}$$

where f denotes the likelihood. Exploiting the identify $\pi(A) = \int_B \pi(A|B)\pi(B)$ by conditioning on and integrating over $\pi(\mathbf{x}_t, \mathbf{y}_t)$ gives,

$$\pi(\alpha, \beta, \tau^2 | \mathbf{y}_o, \mathbf{x}_o) \propto \int_{\mathbf{x}_t, \mathbf{y}_t} f(\mathbf{y}_o, \mathbf{x}_o | \mathbf{y}_t, \mathbf{x}_t, \alpha, \beta, \tau^2) \pi(\mathbf{y}_t, \mathbf{x}_t | \alpha, \beta, \tau^2) \pi(\alpha, \beta, \tau^2) \tag{B.3}$$

where we use f exclusively to indicate that the distribution is in terms of the observations. Simplification can be achieved by noting that the prior on (α, β, τ^2) is independent of $\mathbf{x}_t$ and $\mathbf{y}_t$, so can be removed from the integral. Further simplification follows by modeling the observations, $\mathbf{x}_o$ and $\mathbf{y}_o$, as conditionally independent of the regression parameters, given $\mathbf{x}_t$ and $\mathbf{y}_t$. We do, however, explicitly model the correlation between the errors in the observations:

$$\begin{pmatrix} x_o(\mathbf{s}_i) \\ y_o(\mathbf{s}_i) \end{pmatrix} \Bigg| \begin{pmatrix} x_t(\mathbf{s}_i) \\ y_t(\mathbf{s}_i) \end{pmatrix} \sim \mathcal{N} \left[\begin{pmatrix} x_t(\mathbf{s}_i) \\ y_t(\mathbf{s}_i) \end{pmatrix}, \begin{pmatrix} \sigma_x^2(\mathbf{s}_i) & \rho(\mathbf{s}_i)\sigma_x(\mathbf{s}_i)\sigma_x(\mathbf{s}_i) \\ \rho(\mathbf{s}_i)\sigma_y(\mathbf{s}_i)\sigma_x(\mathbf{s}_i) & \sigma_y^2(\mathbf{s}_i) \end{pmatrix} \right]. \tag{B.4}$$

As a result, and suppressing explicit dependence of all parameters on $\mathbf{s}_i$, we have,

$$\begin{aligned} x_o | y_o, x_t, y_t &\sim \mathcal{N}(x_t + \rho \frac{\sigma_x}{\sigma_y} (y_o - y_t), \sigma_x^2 \left(1 - \rho^2\right)) \\ y_o | x_o, x_t, y_t &\sim \mathcal{N}(y_t + \rho \frac{\sigma_y}{\sigma_x} (x_o - x_t), \sigma_y^2 \left(1 - \rho^2\right)) \end{aligned} \tag{B.5}$$

Each of these forms will be useful in deriving the full conditional distributions of $x_t(\mathbf{s}_i)$ and $y_t(\mathbf{s}_i)$.

All that remains is to specify a form for $\pi(\mathbf{y}_t, \mathbf{x}_t|\alpha, \beta, \tau^2)$, and here we exploit the identity $\pi(A, B) = \pi(A|B)\pi(B)$; the regression relationship in Eqn. (B.1); and the fact that the values of the dependent variable are best modeled as draws form a common normal distribution with unknown mean and variance:

$$\begin{aligned}
\pi(\mathbf{y}_t, \mathbf{x}_t|\alpha, \beta, \tau^2) &= \pi(\mathbf{y}_t|\mathbf{x}_t, \alpha, \beta, \tau^2)\pi(\mathbf{x}_t|\alpha, \beta, \tau^2) = \pi(\mathbf{y}_t|\mathbf{x}_t, \alpha, \beta, \tau^2)\pi(\mathbf{x}_t) \\
y_t(\mathbf{s}_i)|x_t(\mathbf{s}_i), \alpha, \beta, \tau^2 &\sim \mathcal{N}(\alpha + \beta \cdot x_t(\mathbf{s}_i), \tau^2) \\
x_t(\mathbf{s}_i)|\mu_x, \delta_x^2 &\sim \mathcal{N}(\mu_x, \delta_x^2) \\
\pi(\mu_x, \delta_x^2) &\propto 1/\delta_x^2 \qquad \text{(B.6)}
\end{aligned}$$

As with the regression parameters, we use a flat, improper prior for μ_x and the Jeffery's prior for δ_x^2, so that $\pi(\mu_x, \delta_x^2) \propto 1/\delta_x^2$ (Gelman et al., 2003). Eqn. (B.3) then simplifies to,

$$\pi(\alpha, \beta, \tau^2|\mathbf{y}_o, \mathbf{x}_o) \propto \pi(\alpha, \beta, \tau^2) \int \pi(\mathbf{y}_t|\mathbf{x}_t, \alpha, \beta, \tau^2)\pi(\mathbf{x}_t|\mu_x, \delta_x^2)\pi(\mu_x, \delta_x^2) f(\mathbf{x}_o, \mathbf{y}_o|\mathbf{y}_t, \mathbf{x}_t)$$

where the integration is over $\mathbf{x}_t, \mathbf{y}_t$, and the parameters μ_x and δ_x^2. The last term can equivalently be written as $f(\mathbf{x}_o|\mathbf{y}_o\mathbf{x}_t, \mathbf{y}_t)f(\mathbf{y}_o|\mathbf{y}_t)$ or $f(\mathbf{y}_o|\mathbf{x}_o\mathbf{x}_t, \mathbf{y}_t)f(\mathbf{x}_o|\mathbf{x}_t)$; in the former case, only the first term is dependent on $\mathbf{x}_t$, and in the latter case, only the first term is dependent on $\mathbf{y}_t$.

We adopt an empirical Bayes (Gelman et al., 2003) approach, setting $\sigma_{x,y}^2(\mathbf{s}_i)$ to the station specific residual bootstrap variances, and $\rho(\mathbf{s}_i)$ to the station specific correlations between bootstrap samples (Table B.4). Note that these error correlations are non-negligible, with median values across stations of 0.656 between the 95^{th} and 50^{th} percentiles, and 0.634 between the 5^{th} and 50^{th} percentiles.

The regression framework partitions the observed variance in the independent variable be-

Table B.4: Estimates of the error standard deviations (in $^\circ$C/decade^{-1}) and error correlations (unitless) for trends in the 95^{th}, 50^{th}, and 5^{th} percentiles for the full GHCN-D data set. Values are derived from the residual block bootstrap on years, and subsequently used in the regressions between trends. Separate values are estimated for each station, and both the median and a central 95% interval are shown.

		95^{th}	50^{th}	5^{th}
Standard Deviations	Median	0.304	0.245	0.329
	95% interval	$(0.157, 0.555)$	$(0.131, 0.377)$	$(0.165, 0.586)$
		95^{th} and 50^{th}		5^{th} and 50^{th}
Correlations	Median	0.656		0.636
	95% interval	$(0.397, 0.776)$		$(0.363, 0.776)$

tween correlated observational or measurement errors $[\sigma_x^2(\mathbf{s}_i), \rho(\mathbf{s}_i)]$ (assumed known), and the estimated variance across the stations in the true values of the independent variable $[\delta_x^2]$. For the dependent variable, the variance in the observations is partitioned between the specified correlated measurement error variance $[\sigma_y^2(\mathbf{s}_i), \rho(\mathbf{s}_i)]$, the variability due to the dependence on $\mathbf{x}_t$ $[\beta^2\delta_x^2]$, and equation error in the regression model $[\tau^2]$. Posterior draws of $\mathbf{x}_t$ and $\mathbf{y}_t$ are effectively shrunk towards the mean value and the regression line, respectively, according to the variance partitioning.

Note that there are five unknown scalar parameters – $(\alpha, \beta, \mu_x, \tau^2, \delta_x^2)$ – and that the modeling assumptions are equivalent to assuming that $(x_t, y_t)^T$ is bivariate normal with unknown mean (two parameters) and covariance matrix (three parameters). In effect, we fit the parameters of the joint distribution via the identity $P(A, B) = P(A|B)P(B)$, which, under normality, results in a linear expression for the conditional mean. An estimate of the squared correlation for the regression follows from equating the five parameters of a bivariate normal to the five parameters

estimated here (Table B.2):

$$r^2 = \frac{\beta^2 \sigma_x^2}{\tau^2 + \beta^2 \sigma_x^2}.$$

The sampling of (α, β, τ^2) and integration over $(\mathbf{x}_t, \mathbf{y}_t, \mu_x, \delta_x^2)$ then proceeds numerically, via Markov Chain–Monte Carlo sampling. Using a Gibbs sampler (Gelman et al., 2003), we draw from the full conditional distribution of each unknown in turn, given the most recent draws of all other parameters. The result is an estimate of the joint distribution of $(\alpha, \beta, \tau^2, \mu_x, \delta_x^2, \mathbf{x}_t, \mathbf{y}_t)$, so that the samples of the parameters of interest – (α, β, τ^2) – integrate over the distributions of the nuisance parameters.

Given the priors discussed above, the full conditional posteriors have the following forms, where the notation $z|\cdot$ indicate the conditional distribution of z given all other variables.

The full conditional posteriors for α, β, and μ_x are normal:

$$\begin{aligned}
\alpha|\cdot &\sim \mathcal{N}(1^T \cdot (\mathbf{y}_t - \beta \mathbf{x}_t)/N, \tau^2/N) \\
\beta|\cdot &\sim \mathcal{N}(\mathbf{x}_t^T(\mathbf{y}_t - \alpha)/(\mathbf{x}_t^T \mathbf{x}_t), \tau^2/(\mathbf{x}_t^T \mathbf{x}_t)) \\
\mu_x|\cdot &\sim \mathcal{N}(1^T \cdot \mathbf{x}_t/N, \delta_x^2/N),
\end{aligned} \tag{B.7}$$

where N gives the total number of stations, 1 is a vector of ones, and a superscript ^{T} denotes a transpose.

The full conditional posterior for τ^2 and δ_x^2 are Inverse-Gamma:

$$\begin{aligned}
\tau^2|\cdot &\sim \mathcal{IG}(N/2, (\mathbf{y}_t - \beta \mathbf{x}_t - \alpha)^T(\mathbf{y}_t - \beta \mathbf{x}_t - \alpha)/2) \\
\delta_x^2|\cdot &\sim \mathcal{IG}(N/2, (\mathbf{x}_t - \mu_x)^T(\mathbf{x}_t - \mu_x)/2),
\end{aligned} \tag{B.8}$$

where we use the notation of Gelman et al. (2003).

The full conditional posteriors for the $x_t(\mathbf{s}_i)$ and $y_t(\mathbf{s}_i)$, where the equations hold separately at

each location $\mathbf{s}_i$ and $(x_t, y_t, x_o, y_o, \sigma_x^2, \sigma_y^2, \rho)$ are all implicitly dependent on $\mathbf{s}_i$, are normal:

$$\begin{aligned} x_t(\mathbf{s}_i)|\cdot &\sim \mathcal{N}(V_x\Psi_x, \Psi_x) \\ V_x &= \frac{\beta(y_t-\alpha)}{\tau^2} + \frac{x_o - \rho\frac{\sigma_x}{\sigma_y}(y_o - y_t)}{\sigma_x^2(1-\rho^2)} + \frac{\mu_x}{\delta_x^2} \\ \Psi_x &= \left(\frac{\beta^2}{\tau^2} + \frac{1}{\sigma_x^2(1-\rho^2)} + \frac{1}{\delta_x^2}\right)^{-1} \end{aligned} \tag{B.9}$$

$$\begin{aligned} y_t(\mathbf{s}_i)|\cdot &\sim \mathcal{N}(V_y\Psi_y, \Psi_y) \\ V_y &= \frac{\beta x_t + \alpha}{\tau^2} + \frac{y_o - \rho\frac{\sigma_y}{\sigma_x}(x_o - x_t)}{\sigma_y^2(1-\rho^2)} \\ \Psi_y &= \left(\frac{1}{\tau^2} + \frac{1}{\sigma_y^2(1-\rho^2)}\right)^{-1}. \end{aligned} \tag{B.10}$$

B.8.7 Robustness of results to the length of the summer interval

The phase of peak seasonality varies from roughly calendar day 190 in the Southwest to day 210 along the Pacific coast, with a mean value of 204.5 used to define thermal summer for the region as a whole as days 175 to 234. Given the spatial variability in the phase of peak summer, we view a 60 day interval as a reasonable compromise between an interval short enough to reflect peak summer conditions, and long enough to capture peak summer throughout the domain.

Using a 50 or 70 day definition of peak summer leads to spatial maps of trends that are qualitatively and quantitatively similar to those resulting from a 60 day summer interval, with correlations in all cases greater than 0.94. Correlations under changing summer intervals are smallest for trends in the 5^{th} percentile, and largest –in excess of 0.98– for trends in the 95^{th} percentile. As expected, there is a greater sensitivity to seasonality for trends in the lower tail.

The regression relationships between the trends in the three percentiles likewise display some sensitivity to the length of the interval used to define summer (Table B.5). The amplification of the trends in the 95^{th} percentile as compared with trends in the 50^{th} percentile is more pronounced under a shorter definition of summer, and less so using a 70-day interval, likely due to seasonality affecting estimates of trends in the 50^{th} percentile under a longer definition of summer. The offset between trends in the 5^{th} percentile as compared with the 50^{th} percentile is evident under all three summer intervals, and is largest for the 50 day interval. In all cases, the slope between trends in the 5^{th} percentile and the 50^{th} percentile slope is not different from unity.

Table B.5: As in Table B.2, but showing results for the GHCN-D data set using 50 and 70 day summer intervals.

		95^{th} on 50^{th}	5^{th} on 50^{th}
GHCN-D	Intercept	-0.013	0.17
(50 days)	95% interval	$(-0.024, -0.004)$	$(0.161, 0.189)$
	Slope	1.11	0.98
	95% interval	$(1.07, 1.16)$	$(0.91, 1.04)$
	r^2	0.90	0.47
	95% interval	$(0.88, 0.93)$	$(0.42, 0.52)$
GHCN-D	Intercept	-0.042	0.110
(70 days)	95% interval	$(-0.051, -0.033)$	$(0.097, 0.123)$
	Slope	1.02	1.03
	95% interval	$(0.98, 1.06)$	$(0.97, 1.08)$
	r^2	0.87	0.56
	95% interval	$(0.84, 0.90)$	$(0.52, 0.60)$

B.8.8 Comparisons with results from the NCEP II reanalysis

As the GHCN-D data cover the same time interval as the National Center for Environmental Prediction's Reanalysis II product (Kanamitsu et al., 2002), and results are additionally smoothed and interpolated onto the NCEP II grid, it is fruitful to explore the agreement between the station-level results and those for the maxima of the four times daily 2m air temperature from NCEP II. To facilitate comparison, we first apply to the NCEP II results the same thin plate spline smoothing as was applied to the station level results, as the spatial smoothing length scale of 750 miles is much greater than the 2° NCEP II grid boxes.

There is generally weak agreement between the smoothed trend maps from the NCEP II and GHCN-D analyses (Figs. B.10, B.5). Both the correlation and root mean squared error (RMSE) between corresponding smoothed trend fields are largest for the 5^{th} percentile ($r = 0.53$, RMSE= 0.55) and smallest for the 50^{th} percentile ($r = 0.33$, RMSE=0.39), indicating that the pattern of agreement is worst for the 50^{th}, while the errors are smallest due to the generally smaller magnitudes (Table B.1).

The NCEP II 95^{th} percentile trends feature more significant warming in the West as compared with GHCN-D; a different pattern of cooling in the centre of the domain; and similar warming in eastern Canada, where GHCN-D is data sparse (Figs. B.10, B.5). The 50^{th} percentile trends from NCEP II feature clusters of strong warming throughout the western portion of the domain; with the exception of the southernmost warming cluster, where there are no stations, these features are amplifications of those seen in the GHCN-D results. Cooling regions in the the center of the domain for both the 50^{th} and 95^{th} percentiles are generally polewards of the corresponding features seen in GHCN-D. For trends in the 5^{th} percentile, NCEP II captures the strong warming in the West, and some of the warming along the Eastern Seaboard that is evident in the smoothed GHCN-D results, including in regions where GHCN-D is data-sparse. The similarities and in-

congruities between NCEP II and GHCN-D are likewise reflected in maps of the differences in trends between quantiles for each analysis. NCEP captures aspects of the increasing variability seen in GHCN-D for the Texas region, the moderate reduction in variability in the eastern part of the domain, and a stronger reduction in variability in the West.

The most striking difference between the two analyses are the much larger values, of both signs, seen in the NCEP II analysis. The largest magnitude trends for the smoothed NCEP analysis are about 2.5°C decade^{-1}, substantially larger than the maximum of about 1.4°C decade^{-1} for the smoothed GHCN-D results. We note that the spatially smoothed trends from the GHCN-D analysis are not immediately comparable to those found for the NCEP II gridded product, as the latter measures trends in the quantiles of spatially smoothed temperatures, and in the former we produce spatially smoothed maps from station-level trend estimates. In general, calculating extremal indices on gridded climate data should reduce the severity of the extremes, due to the averaging or smoothing inherent to the gridding (Haylock et al., 2008; Zhang et al., 2011).

It is therefore somewhat surprising that the smoothed trends estimated from the NCEP II analysis feature larger values and larger variability than the smoothed trends from the GHCN-D data (Table B.1 and Figs. B.10, B.5). A number of the NCEP grid boxes, even after smoothing to the same level as the GHCN-D results, feature slopes, particularly for the 5^{th} percentile, that are nearly twice as large in magnitude as those for the smoothed GHCN-D results. The slopes are most amplified for grid boxes that feature strongly varying topography, where there is often a clear multi-modality in the NCEP temperature values, persistent throughout much of the year and lasting from about 1979–1999 (Fig. B.11), perhaps indicating that the boundary layer model struggles to accommodate sharp elevation gradients prior to about 1999 (Rusticucci and Kousky, 2002). As the multi modality is time dependent, slopes, particularly in the 5^{th} percentile, are much larger than any observed in the GHCN-D data set.

Regressions between the trends in percentiles from the NCEP II analysis do not reproduce the

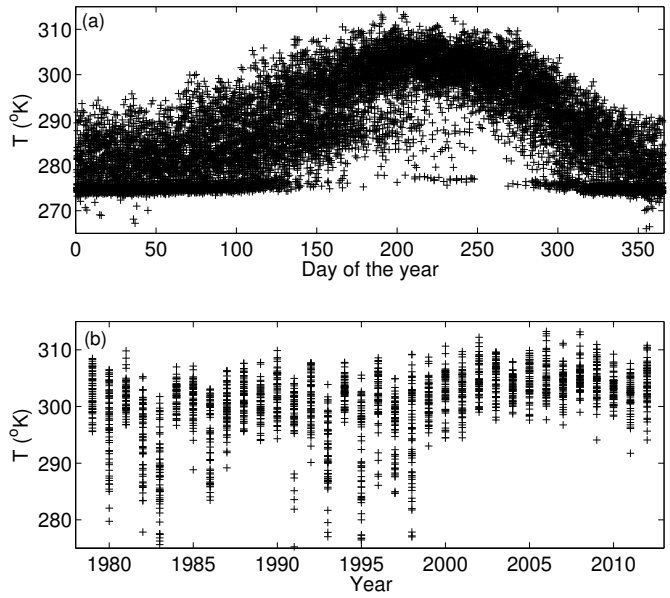

Figure B.11: Maximum of four times daily 2m air temperature from the NCEP II reanalysis, for a grid box centered between Fresno and Yosemite, in the central valley of California. (a) T_X plotted as a function of day of the year, for 1979–2012. (b) T_X for days 175–234, plotted as a function of year.

features seen in the GHCN-D analysis (Table B.2). The regression of trends in the 5^{th} percentile onto trends in the 50^{th} percentile from NCEP II features strong amplification in the slope and an intercept that is not significantly different from zero, whereas the GHCN-D results feature a large intercept and a slope not different from unity. For the NCEP II analysis, trends in the 95^{th} percentile are damped with respect to those in the 50^{th} percentile, and the intercept is positive; in contrast, GHCN-D results feature amplification in the slope and a negative intercept. These discrepancies suggest that NCEP II is not capturing aspects of the robust distributional changes seen in the station data. In interpreting the NCEP-II results we stress that NCEP II and GHCN-D results are not immediately comparable, as the former is an inherently smoothed product. In addition, the time dependent multi modality evident in a number of the mountainous NCEP-II grid boxes naturally inflates trends in the 5^{th} percentile as compared with the 50^{th} or 95^{th} percentile, further complicating the comparison with the GHCN-D results.

References

Abraham, J., et al., 2013: A review of global ocean temperature observations: Implications for ocean heat content estimates and climate change. *Reviews of Geophysics*, **51 (3)**, 450–483.

Alexander, L. and S. Perkins, 2013: Debate heating up over changes in climate variability. *Environmental Research Letters*, **8 (4)**, 041 001.

Andrews, T., J. M. Gregory, M. J. Webb, and K. E. Taylor, 2012: Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophysical Research Letters*, **39 (9)**, L09 712.

Armour, K. C., C. M. Bitz, and G. H. Roe, 2013: Time-varying climate sensitivity from regional feedbacks. *Journal of Climate*, **26 (13)**, 4518–4534.

Barbosa, S., M. Scotto, and A. Alonso, 2011: Summarising changes in air temperature over Central Europe by quantile regression and clustering. *Natural Hazards and Earth System Science*, **11 (12)**, 3227–3233.

Barkstrom, B. R., 1984: The earth radiation budget experiment (ERBE). *Bulletin of the American Meteorological Society*, **65 (11)**, 1170–1185.

Barnston, A. G. and S. J. Mason, 2011: Evaluation of IRI's seasonal climate forecasts for the extreme 15% tails. *Weather and Forecasting*, **26 (4)**, 545–554.

Bassett Jr, G. W., M.-Y. S. Tam, and K. Knight, 2002: Quantile models and estimators for data analysis. *Metrika*, **55 (1-2)**, 17–26.

Battisti, D. S. and R. L. Naylor, 2009: Historical warnings of future food insecurity with unprecedented seasonal heat. *Science*, **323 (5911)**, 240–244.

Bekryaev, R. V., I. V. Polyakov, and V. A. Alexeev, 2010: Role of polar amplification in long-term surface air temperature variations and modern Arctic warming. *Journal of Climate*, **23 (14)**, 3888–3906.

Bengtsson, L., S. Hagemann, and K. I. Hodges, 2004: Can climate trends be calculated from reanalysis data? *Journal of Geophysical Research: Atmospheres (1984–2012)*, **109 (D11)**.

Berger, A. and M.-F. Loutre, 1991: Insolation values for the climate of the last 10 million years. *Quaternary Science Reviews*, **10 (4)**, 297–317.

Bhatt, S., et al., 2013: The global distribution and burden of dengue. *Nature*, **496 (7446)**, 504–507.

Black, E., M. Blackburn, G. Harrison, B. Hoskins, and J. Methven, 2004: Factors contributing to the summer 2003 European heatwave. *Weather*, **59 (8)**, 217–223.

Bony, S., J.-P. Duvel, and H. Le Trent, 1995: Observed dependence of the water vapor and clear-sky greenhouse effect on sea surface temperature: comparison with climate warming experiments. *Climate Dynamics*, **11 (5)**, 307–320.

Brooks, C., 1917: Continentality and temperature. *Quarterly Journal of the Royal Meteorological Society*, **43 (182)**, 159–174.

Brooks, C., 1918: Continentality and temperature — second paper: The effect of latitude on the influence of continentality on temperature. *Quarterly Journal of the Royal Meteorological Society*, **44 (188)**, 253–270.

Brown, P. J., 1993: *Measurement, regression, and calibration*. Clarendon Press, Oxford.

Brown, S., J. Caesar, and C. Ferro, 2008: Global changes in extreme daily temperature since 1950. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **113 (D5)**.

Brunt, D., 1924: Climatic continentality and oceanity. *The Geographical Journal*, **64 (1)**, 43–49.

Cade, B. S. and B. R. Noon, 2003: A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, **1 (8)**, 412–420.

Caesar, J., L. Alexander, and R. Vose, 2006: Large-scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded data set. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **111 (D5)**.

Cerne, S. B., C. S. Vera, and B. Liebmann, 2007: The nature of a heat wave in eastern Argentina occurring during SALLJEX. *Monthly Weather Review*, **135 (3)**, 1165–1174.

Chang, H., 2004: *Inventing temperature: Measurement and scientific progress*. Oxford University Press.

Chen, T.-C. and M.-C. Yen, 1993: Interannual variation of summertime stationary eddies. *Journal of Climate*, **6 (12)**, 2263–2277.

Christiansen, B., 2013: Straight line fitting and predictions: On a marginal likelihood approach to linear regression and errors-in-variables models. *Journal of Climate*, **27 (5)**, 2014–2031.

Colman, R. and L. Hanson, 2013: On atmospheric radiative feedbacks associated with climate variability and change. *Climate Dynamics*, **40 (1-2)**, 475–492.

Coumou, D., V. Petoukhov, S. Rahmstorf, S. Petri, and H. J. Schellnhuber, 2014: Quasi-resonant circulation regimes and hemispheric synchronization of extreme weather in boreal summer. *Proceedings of the National Academy of Sciences*, **111 (34)**, 12 331–12 336.

DeGaetano, A. T. and R. J. Allen, 2002: Trends in twentieth-century temperature extremes across the United States. *Journal of Climate*, **15 (22)**, 3188–3205.

Deser, C., M. A. Alexander, and M. S. Timlin, 2003: Understanding the persistence of sea surface temperature anomalies in midlatitudes. *Journal of Climate*, **16 (1)**, 57–72.

Diffenbaugh, N. S., 2009: Influence of modern land cover on the climate of the United States. *Climate dynamics*, **33 (7-8)**, 945–958.

Dole, R., et al., 2011: Was there a basis for anticipating the 2010 Russian heat wave? *Geophysical Research Letters*, **38 (6)**, L06 702.

Donat, M. G. and L. V. Alexander, 2012: The shifting probability distribution of global daytime and night-time temperatures. *Geophysical Research Letters*, **39 (14)**, L14 707.

Donohoe, A., 2011: Radiative and dynamic controls of global scale energy fluxes. Ph.D. thesis, University of Washington.

Donohoe, A. and D. S. Battisti, 2011: Atmospheric and surface contributions to planetary albedo. *Journal of Climate*, **24 (16)**, 4402–4418.

Donohoe, A. and D. S. Battisti, 2013: The seasonal cycle of atmospheric heating and temperature. *Journal of Climate*, **26 (14)**, 4962–4980.

Draxler, R., 1997: HYSPLIT_4 user's guide. NOAA Tech. Memo ERL ARL-230, NOAA Air Resources Laboratory, Silver Spring, MD.

Draxler, R. and G. Hess, 1997: Description of the HYSPLIT_4 modeling system. NOAA Tech. Memo ERL ARL-224, NOAA Air Resources Laboratory, 24 pp., Silver Spring, MD.

Draxler, R. and G. Hess, 1998: An overview of the HYSPLIT_4 modelling system for trajectories, dispersion, and deposition. *Australian Meteorological Magazine*, **47 (4)**, 295–308.

Durre, I., M. J. Menne, B. E. Gleason, T. G. Houston, and R. S. Vose, 2010: Comprehensive automated quality assurance of daily surface observations. *Journal of Applied Meteorology and Climatology*, **49 (8)**, 1615–1633.

Dwyer, J. G., M. Biasutti, and A. H. Sobel, 2012: Projected changes in the seasonal cycle of surface temperature. *Journal of Climate*, **25 (18)**, 6359–6374.

Ebi, K. L. and J. K. Schmier, 2005: A stitch in time: improving public health early warning systems for extreme weather events. *Epidemiologic Reviews*, **27 (1)**, 115–121.

Fasullo, J. T. and K. E. Trenberth, 2008: The annual cycle of the energy budget. Part II: Meridional structures and poleward transports. *Journal of Climate*, **21 (10)**, 2313–2325.

Feldl, N. and G. H. Roe, 2013: The nonlinear and nonlocal nature of climate feedbacks. *Journal of Climate*, **21 (21)**, 8289–8304.

Forest, C. E., P. H. Stone, A. P. Sokolov, M. R. Allen, and M. D. Webster, 2002: Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science*, **295 (5552)**, 113–117.

Forney Jr, G. D., 1973: The viterbi algorithm. *Proceedings of the IEEE*, **61 (3)**, 268–278.

Forster, P. M., T. Andrews, P. Good, J. M. Gregory, L. S. Jackson, and M. Zelinka, 2013: Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. *Journal of Geophysical Research: Atmospheres*, **118 (3)**, 1139–1150.

Frankignoul, C. and K. Hasselmann, 1977: Stochastic climate models, Part II: application to sea-surface temperature anomalies and thermocline variability. *Tellus*, **29 (4)**, 289–305.

Frost, C. and S. Thompson, 2000: Correcting for regression dilution bias: comparison of methods for a single predictor variable. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **163 (2)**, 173–189.

Furrer, R., D. Nychka, and S. Sain, 2013: *fields: Tools for spatial data*. URL `http://www.image.ucar.edu/Software/Fields`, R package version 6.8.

Gelman, A., J. Carlin, H. Stern, and D. Rubin, 2003: *Bayesian Data Analysis*. 2d ed., Chapman & Hall/CRC, Boca Raton.

Goody, R. and Y. Yung, 1989: *Atmospheric Radiation: Theoretical Basis*. 2d ed., Oxford University Press, New York.

Greenwood, D. R. and S. L. Wing, 1995: Eocene continental climates and latitudinal temperature gradients. *Geology*, **23 (11)**, 1044–1048.

Gregory, J. and P. Forster, 2008: Transient climate response estimated from radiative forcing and observed temperature change. *Journal of Geophysical Research: Atmospheres*, **113 (D23)**, D23 105.

Gregory, J., R. Stouffer, S. Raper, P. Stott, and N. Rayner, 2002: An observationally based estimate of the climate sensitivity. *Journal of Climate*, **15 (22)**, 3117–3121.

Hansen, J., M. Sato, and R. Ruedy, 2012: Perception of climate change. *Proceedings of the National Academy of Sciences*, 1–9.

Hassanzadeh, P., Z. Kuang, and B. F. Farrell, 2014: Responses of midlatitude blocks and wave amplitude to changes in the meridional temperature gradient in an idealized dry GCM. *Geophysical Research Letters*, **41 (14)**, 5223–5232.

Hasselmann, K., 1976: Stochastic climate models part I. Theory. *Tellus*, **28 (6)**, 473–485.

Haylock, M., N. Hofstra, A. Klein Tank, E. Klok, P. Jones, and M. New, 2008: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research: Atmospheres*, **113 (D20)**.

Heim Jr, R. R., 2002: A review of twentieth-century drought indices used in the United States. *Bulletin of the American Meteorological Society*, **83 (8)**, 1149–1165.

Hela, I., 1953: Regional distribution of the continentality in the climate of the oceans. *Geophysica*, **4 (2)**, 41–47.

Held, I. M. and B. J. Soden, 2006: Robust responses of the hydrological cycle to global warming. *Journal of Climate*, **19 (21)**, 5686–5699.

Held, I. M., M. Winton, K. Takahashi, T. Delworth, F. Zeng, and G. K. Vallis, 2010: Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *Journal of Climate*, **23 (9)**, 2418–2427.

Higgins, R., H. Kim, and D. Unger, 2004: Long-lead seasonal temperature and precipitation prediction using tropical Pacific SST consolidation forecasts. *Journal of Climate*, **17 (17)**, 3398–3414.

Hofstra, N., M. New, and C. McSweeney, 2010: The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. *Climate Dynamics*, **35 (5)**, 841–858.

Hoskins, B. J. and D. J. Karoly, 1981: The steady linear response of a spherical atmosphere to thermal and orographic forcing. *Journal of the Atmospheric Sciences*, **38 (6)**, 1179–1196.

Huber, M. and R. Caballero, 2011: The early Eocene equable climate problem revisited. *Climate of the Past*, **7 (2)**, 603–633.

Huntingford, C., P. D. Jones, V. N. Livina, T. M. Lenton, and P. M. Cox, 2013: No increase in global temperature variability despite changing regional patterns. *Nature*, **500**, 327–330.

Huybers, P. and W. Curry, 2006: Links between annual, Milankovitch and continuum temperature variability. *Nature*, **441 (7091)**, 329–332.

Hwang, Y.-T., D. M. Frierson, and J. E. Kay, 2011: Coupling between Arctic feedbacks and changes in poleward energy transport. *Geophysical Research Letters*, **38 (17)**, L17 704.

Hyde, W. T., T. J. Crowley, K.-Y. Kim, and G. R. North, 1989: Comparison of GCM and Energy Balance Model Simulations of Seasonal Temperature Changes over the Past 18000 Years. *Journal of Climate*, **2 (8)**, 864–887.

Ihaka, R. and R. Gentleman, 1996: R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, **5**, 299–314.

IPCC, 2007: Summary for policymakers. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change.*, C. Field, V. Barros, T. Stocker, D. Qin, D. Dokken, K. Ebi, M. Mastrandrea, K. Mach, G.-K. Plattner, S. Allen, M. Tignor, and P. Midgley, Eds., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jaccard, P., 1912: The distribution of the flora in the alpine zone. *New Phytologist*, **11 (2)**, 37–50.

Kalnay, E. and M. Cai, 2003: Impact of urbanization and land-use change on climate. *Nature*, **423 (6939)**, 528–531.

Kalnay, E., et al., 1996: The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77 (3)**, 437?471.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. Hnilo, M. Fiorino, and G. Potter, 2002: NCEP–DOE AMIP-II Reanalysis (R-2). *Bulletin of the American Meteorological Society*, **83 (11)**, 1631–1643.

Katz, R. W. and B. G. Brown, 1992: Extreme events in a changing climate: variability is more important than averages. *Climatic change*, **21 (3)**, 289–302.

Kennedy, J., N. Rayner, R. Smith, D. Parker, and M. Saunby, 2011: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties. *Journal of Geophysical Research: Atmospheres*, **116 (D14)**, D14 104.

Kim, K.-Y. and G. R. North, 1991: Surface temperature fluctuations in a stochastic climate model. *Journal of Geophysical Research: Atmospheres*, **96 (D10)**, 18 573–18 580.

Kim, K.-Y. and G. R. North, 1992: Seasonal cycle and second-moment statistics of a simple coupled climate model. *Journal of Geophysical Research*, **97 (D18)**, 20 437–20 448.

Kim, K.-Y., G. R. North, and J. Huang, 1992: On the transient response of a simple coupled climate system. *Journal of Geophysical Research: Atmospheres*, **97 (D9)**, 10 069–10 081.

Knight, K., 1998: Limiting distributions for L1 regression estimators under general conditions. *Annals of Statistics*, **26 (2)**, 755–770.

Knutti, R., G. A. Meehl, M. R. Allen, and D. A. Stainforth, 2006: Constraining climate sensitivity from the seasonal cycle in surface temperature. *Journal of Climate*, **19 (17)**, 4224–4233.

Koenker, R. and G. Bassett Jr, 1978: Regression quantiles. *Econometrica: journal of the Econometric Society*, **46 (1)**, 33–50.

Koenker, R. and K. Hallock, 2001: Quantile regression. *Journal of Economic Perspectives*, **15 (4)**, 143–156.

Kovats, R. S. and S. Hajat, 2008: Heat stress and public health: a critical review. *Annu. Rev. Public Health*, **29**, 41–55.

Kushnir, Y., W. Robinson, I. Bladé, N. Hall, S. Peng, and R. Sutton, 2002: Atmospheric GCM response to extratropical SST anomalies: synthesis and evaluation. *Journal of Climate*, **15 (16)**, 2233–2256.

Laepple, T. and G. Lohmann, 2009: Seasonal cycle as template for climate variability on astronomical timescales. *Paleoceanography*, **24 (4)**, PA4201.

Lee, K., H.-J. Baek, and C. Cho, 2013: Analysis of changes in extreme temperatures using quantile regression. *Asia-Pacific Journal of Atmospheric Sciences*, **49 (3)**, 313–323.

Legates, D. and C. Willmott, 1990: Mean seasonal and spatial variability in global surface air temperature. *Theoretical and Applied Climatology*, **41 (1)**, 11–21.

Lobell, D. B., C. J. Bonfils, L. M. Kueppers, and M. A. Snyder, 2008: Irrigation cooling effect on temperature and heat index extremes. *Geophysical Research Letters*, **35 (9)**, L09 705.

Lobell, D. B. and C. B. Field, 2007: Global scale climate-crop yield relationships and the impacts of recent warming. *Environmental Research Letters*, **2 (1)**, 014 002.

Loeb, N. G., J. M. Lyman, G. C. Johnson, R. P. Allan, D. R. Doelling, T. Wong, B. J. Soden, and G. L. Stephens, 2012: Observed changes in top-of-the-atmosphere radiation and upper-ocean heating consistent within uncertainty. *Nature Geoscience*, **5**, 110–113.

Loeb, N. G., B. A. Wielicki, D. R. Doelling, G. L. Smith, D. F. Keyes, S. Kato, N. Manalo-Smith, and T. Wong, 2009: Toward optimal closure of the Earth's top-of-atmosphere radiation budget. *Journal of Climate*, **22 (3)**, 748–766.

Loeb, N. G., et al., 2007: Multi-instrument comparison of top-of-atmosphere reflected solar radiation. *Journal of Climate*, **20 (3)**, 575–591.

Loikith, P. C. and A. J. Broccoli, 2012: Characteristics of observed atmospheric circulation patterns associated with temperature extremes over North America. *Journal of Climate*, **25 (20)**, 7266–72 681.

Loikith, P. C. and A. J. Broccoli, 2014: The influence of recurrent modes of climate variability on the occurrence of winter and summer extreme temperatures over North America. *Journal of Climate*, **27 (4)**, 1600–1618.

Lyman, J. M., S. A. Good, V. V. Gouretski, M. Ishii, G. C. Johnson, M. D. Palmer, D. M. Smith, and J. K. Willis, 2010: Robust warming of the global upper ocean. *Nature*, **465 (7296)**, 334–337.

Lyon, B. and R. M. Dole, 1995: A diagnostic comparison of the 1980 and 1988 US summer heat wave-droughts. *Journal of Climate*, **8 (6)**, 1658–1675.

Machado, J. A. F. and J. S. Silva, 2005: Quantiles for counts. *Journal of the American Statistical Association*, **100 (472)**, 1226–1237.

Manabe, S., R. Stouffer, M. Spelman, and K. Bryan, 1991: Transient responses of a coupled ocean-atmosphere model to gradual changes of atmospheric CO_2. Part I. Annual mean response. *Journal of Climate*, **4 (8)**, 785–818.

Mannshardt, E., P. Craigmile, and M. Tingley, 2013: Statistical modeling of extreme value behavior in North American tree-ring density series. *Climatic Change*, **117**, 843–858.

Markwick, P. J., 1998: Crocodilian diversity in space and time: the role of climate in paleoecology and its implication for understanding K/T extinctions. *Paleobiology*, **24 (4)**, 470–497.

Mazon, J., D. Pino, and M. Barriendos, 2014: Rapid and sudden advection of warm and dry air in the Mediterranean Basin. *Natural Hazards and Earth System Sciences*, **14**, 235–246.

McKinnon, K. A., A. R. Stine, and P. Huybers, 2013: The spatial structure of the annual cycle in surface temperature: amplitude, phase, and Lagrangian history. *Journal of Climate*, **26 (20)**, 7852–7862.

Meehl, G. and C. Tebaldi, 2004: More intense, more frequent, and longer lasting heat waves in the 21st century. *Science*, **305 (5686)**, 994–997.

Meehl, G., C. Tebaldi, G. Walton, D. Easterling, and L. McDaniel, 2009: Relative increase of record high maximum temperatures compared to record low minimum temperatures in the us. *Geophysical Research Letters*, **36 (23)**, L23 701.

Menne, M. J., I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, 2012: An overview of the Global Historical Climatology Network-Daily database. *Journal of Atmospheric & Oceanic Technology*, **29 (7)**, 897–910.

Menne, M. J. and C. N. Williams Jr, 2009: Homogenization of temperature series via pairwise comparisons. *Journal of Climate*, **22 (7)**, 1700–1717.

Miralles, D. G., A. J. Teuling, C. C. van Heerwaarden, and J. V.-G. de Arellano, 2014: Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. *Nature Geoscience*, **7**, 345–349.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research*, **117 (D8)**, D08 101.

Murphy, A. H. and E. S. Epstein, 1989: Skill scores and correlation coefficients in model verification. *Monthly Weather Review*, **117 (3)**, 572–582.

Namias, J., 1982: Anatomy of Great Plains protracted heat waves (especially the 1980 US summer drought). *Monthly Weather Review*, **110 (7)**, 824–838.

Nese, J. M., 1994: Systematic biases in manual observations of daily maximum and minimum temperature. *Journal of climate*, **7 (5)**, 834–842.

Newman, M. and P. D. Sardeshmukh, 1998: The impact of the annual cycle on the North Pacific/North American response to remote low-frequency forcing. *Journal of the Atmospheric Sciences*, **55 (8)**, 1336–1353.

North, G. and J. Coakley, 1979a: Differences between seasonal and mean annual energy balance model calculations of climate and climate sensitivity. *Journal of the Atmospheric Sciences*, **36 (7)**, 1189–1204.

North, G. R., R. F. Cahalan, and J. A. Coakley, 1981: Energy balance climate models. *Reviews of Geophysics*, **19 (1)**, 91–121.

North, G. R. and J. A. Coakley, 1979b: Differences between seasonal and mean annual energy balance model calculations of climate and climate sensitivity. *Journal of the Atmospheric Sciences*, **36**, 1189–1204.

North, G. R., J. G. Mengel, and D. A. Short, 1983: Simple Energy Balance Model Resolving the Seasons and the Continents: Application to the Astronomical Theory of the Ice Ages. *Journal of Geophysical Research*, **88 (C11)**, 6576–6586.

Otto, A., et al., 2013: Energy budget constraints on climate response. *Nature Geoscience*, **6**, 415–416.

Pan, Z., R. W. Arritt, E. S. Takle, W. J. Gutowski, C. J. Anderson, and M. Segal, 2004: Altered hydrologic feedback in a warming climate introduces a 'warming hole'. *Geophysical Research Letters*, **31 (17)**, L17 109.

Parker, D. E., T. Legg, and C. K. Folland, 1992: A new daily central England temperature series, 1772–1991. *International Journal of Climatology*, **12 (4)**, 317–342.

Pedlosky, J., 1975: The development of thermal anomalies in a coupled ocean-atmosphere model. *Journal of the Atmospheric Sciences*, **32 (8)**, 1501–1514.

Prescott, J. and J. Collins, 1951: The lag of temperature behind solar radiation. *Quarterly Journal of the Royal Meteorological Society*, **77 (331)**, 121–126.

Proistosescu, C. and P. Huybers, submitted: Divergence of warming scenarios at long timescales due to slow feedbacks.

Purkey, S. G. and G. C. Johnson, 2010: Warming of global abyssal and deep southern ocean waters between the 1990s and 2000s: Contributions to global heat and sea level rise budgets. *Journal of Climate*, **23 (23)**, 6336–6351.

Rabiner, L. R., 1989: A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, **77 (2)**, 257–286.

Rahmstorf, S. and D. Coumou, 2011: Increase of extreme events in a warming world. *Proceedings of the National Academy of Sciences*, **108 (44)**, 17 905–17 909.

Raper, S. C., J. M. Gregory, and R. J. Stouffer, 2002: The role of climate sensitivity and ocean heat uptake on AOGCM transient temperature response. *Journal of Climate*, **15 (1)**, 124–130.

Regonda, S. K., B. Rajagopalan, M. Clark, and J. Pitlick, 2005: Seasonal cycle shifts in hydroclimatology over the western United States. *Journal of Climate*, **18 (2)**, 372–384.

Reich, B. J. and L. B. Smith, 2013: Bayesian quantile regression for censored data. *Biometrics*, **69 (3)**, 651–660.

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *Journal of Climate*, **15 (13)**, 1609–1625.

Rhines, A. and P. Huybers, 2013: Frequent summer temperature extremes reflect changes in the mean, not the variance. *Proceedings of the National Academy of Sciences*, **110 (7)**, E546.

Rhines, A., M. P. Tingley, K. A. McKinnon, and P. Huybers, submitted: Robust decrease in North American winter temperature extremes.

Robeson, S. M., 2002: Relationships between mean and standard deviation of air temperature: Implications for global warming. *Climate Research*, **22 (3)**, 205–213.

Roemmich, D., et al., 2009: The Argo Program: Observing the global ocean with profiling floats. *Oceanography*, **22 (2)**, 34–43.

Rusticucci, M. M. and V. E. Kousky, 2002: A comparative study of maximum and minimum temperatures over Argentina: NCEP–NCAR reanalysis versus station data. *Journal of Climate*, **15 (15)**.

Sakaguchi, K., X. Zeng, and M. Brunke, 2012: The hindcast skill of the CMIP ensembles for the surface air temperature trend. *Journal of Geophysical Research*, **117 (D16)**, D16 113.

Santer, B., T. Wigley, M. Schlesinger, and J. Mitchell, 1990: *Developing climate scenarios from equilibrium GCM results*. Report No. 47, Max-Planck-Institut für Meteorologie, Hamburg.

Schär, C., P. Vidale, D. Lüthi, C. Frei, C. Häberli, M. Liniger, and C. Appenzeller, 2004: The role of increasing temperature variability in European summer heatwaves. *Nature*, **427 (6972)**, 332–336.

Schneider, B., G. Leduc, and W. Park, 2010: Disentangling seasonal signals in Holocene climate trends by satellite-model-proxy integration. *Paleoceanography*, **25 (4)**, PA4217.

Screen, J. A. and I. Simmonds, 2010: The central role of diminishing sea ice in recent arctic temperature amplification. *Nature*, **464 (7293)**, 1334–1337.

Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling, 2010: Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, **99 (3)**, 125–161.

Seneviratne, S. I., et al., 2012: Changes in climate extremes and their impacts on the natural physical environment. *Managing the risks of extreme events and disasters to advance climate change adaptation*, 109–230.

Shell, K. M., J. T. Kiehl, and C. A. Shields, 2008: Using the radiative kernel technique to calculate climate feedbacks in NCAR's Community Atmospheric Model. *Journal of Climate*, **21 (10)**, 2269–2282.

Shindell, D. and G. Faluvegi, 2009: Climate response to regional radiative forcing during the twentieth century. *Nature Geoscience*, **2**, 294–300.

Simmons, A., et al., 2004: Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature. *Journal of Geophysical Research: Atmospheres*, **109 (D24)**, D24 115.

Simolo, C., M. Brunetti, M. Maugeri, and T. Nanni, 2011: Evolution of extreme temperatures in a warming climate. *Geophysical Research Letters*, **38 (16)**, L16 701.

Simolo, C., M. Brunetti, M. Maugeri, and T. Nanni, 2012: Extreme summer temperatures in Western Europe. *Advances in Science and Research*, **8 (1)**, 5–9.

Simpson, D. G., R. J. Carroll, and D. Ruppert, 1987: M-estimation for discrete data: asymptotic distribution theory and implications. *The Annals of Statistics*, **15 (2)**, 657–669.

Small, C. and J. E. Cohen, 2004: Continental physiography, climate, and the global distribution of human population. *Current Anthropology*, **45 (2)**, 269–277.

Smith, A. B. and R. W. Katz, 2013: US billion-dollar weather and climate disasters: data sources, trends, accuracy and biases. *Natural Hazards*, **67 (2)**, 387–410.

Smoyer, K. E., D. G. Rainham, and J. N. Hewko, 2000: Heat-stress-related mortality in five cities in Southern Ontario: 1980–1996. *International Journal of Biometeorology*, **44 (4)**, 190–197.

Sobel, A. H., J. Nilsson, and L. M. Polvani, 2001: The weak temperature gradient approximation and balanced tropical moisture waves. *Journal of the Atmospheric Sciences*, **58 (23)**, 3650–3665.

Soden, B. J. and I. M. Held, 2006: An assessment of climate feedbacks in coupled ocean-atmosphere models. *Journal of Climate*, **19 (14)**, 3354–3360.

Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller, (Eds.) , 2007: *Climate change 2007: The physical science basis*. Cambridge University Press, 996 pp.

Spitaler, R., 1922: Klimatische kontinentalität und ozeanität. *Petermanns Geographische*.

Stine, A., P. Huybers, and I. Fung, 2009: Changes in the phase of the annual cycle of surface temperature. *Nature*, **457 (7228)**, 435–440.

Stine, A. R. and P. Huybers, 2012: Changes in the seasonal cycle of temperature and atmospheric circulation. *Journal of Climate*, **25 (21)**, 7362–7380.

Stott, P. A., 2003: Attribution of regional-scale temperature changes to anthropogenic and natural causes. *Geophysical Research Letters*, **30 (14)**, 1728.

Stouffer, R., S. Manabe, and K. Bryan, 1989: Interhemispheric asymmetry in climate response to a gradual increase of atmospheric CO_2. *Nature*, **342 (6250)**, 660–662.

Stroeve, J., M. Serreze, M. Holland, J. Kay, J. Malanik, and A. Barrett, 2012: The Arctic's rapidly shrinking sea ice cover: a research synthesis. *Climatic Change*, **110 (3–4)**, 1005–1027.

Swets, J. A., 1973: The relative operating characteristic in psychology. *Science*, **182 (4116)**, 990–1000.

Takaya, K. and H. Nakamura, 1997: A formulation of a wave-activity flux for stationary Rossby waves on a zonally varying basic flow. *Geophysical Research Letters*, **24 (23)**, 2985–2988.

Tang, Q., X. Zhang, and J. A. Francis, 2014: Extreme summer weather in northern mid-latitudes linked to a vanishing cryosphere. *Nature Climate Change*, **4**, 45–50.

Tao, F., M. Yokozawa, Y. Xu, Y. Hayashi, and Z. Zhang, 2006: Climate changes and trends in phenology and yields of field crops in China, 1981–2000. *Agricultural and Forest Meteorology*, **138 (1)**, 82–92.

Teng, H., G. Branstator, H. Wang, G. A. Meehl, and W. M. Washington, 2013: Probability of US heat waves affected by a subseasonal planetary wave pattern. *Nature Geoscience*, **6**, 1056–1061.

The GFDL Global Atmospheric Model Development Team, 2004: The new GFDL global atmosphere and land model AM2-LM2: Evaluation with prescribed SST simulations. *Journal of Climate*, **17 (24)**, 4641–4673.

Theiler, J., S. Eubank, A. Longtin, B. Galdrikian, and J. Doyne Farmer, 1992: Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, **58 (1)**, 77–94.

Thompson, R., 1995: Complex demodulation and the estimation of the changing continentality of Europe's climate. *International Journal of Climatology*, **15 (2)**, 175–185.

Thompson, S. L. and S. H. Schneider, 1979: A seasonal zonal energy balance climate model with an interactive lower layer. *Journal of Geophysical Research*, **84 (C5)**, 2401–2414.

Thomson, D. J., 1982: Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, **70 (9)**, 1055–1096.

Thomson, D. J., 1995: The seasons, global temperature, and precession. *Science*, **268 (5207)**, 59–68.

Tingley, M., 2012: A Bayesian ANOVA Scheme for Calculating Climate Anomalies, with Applications to the Instrumental Temperature Record. *Journal of Climate*, **25**, 777–791.

Tingley, M. and P. Huybers, 2010: A Bayesian Algorithm for Reconstructing Climate Anomalies in Space and Time. Part 2: Comparison with the Regularized Expectation-Maximization Algorithm. *Journal of Climate*, **23 (10)**, 2782–2800.

Tingley, M. P., P. F. Craigmile, M. Haran, B. Li, E. Mannshardt, and B. Rajaratnam, 2012: Piecing together the past: statistical insights into paleoclimatic reconstructions. *Quaternary Science Reviews*, **35**, 1–22.

Tingley, M. P. and P. Huybers, 2013: Recent temperature extremes at high northern latitudes unprecedented in the past 600 years. *Nature*, **496 (7444)**, 201–205.

Trenberth, K. E., G. W. Branstator, D. Karoly, A. Kumar, N.-C. Lau, and C. Ropelewski, 1998: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *Journal of Geophysical Research: Oceans*, **103 (C7)**, 14 291–14 324.

van den Dool, H. and G. Können, 1981: Strong variations in the delay of the annual cycle in the air temperature near the coast. *First International Conference on Meteorology and Air/Sea Interaction of the Coastal Zone*, Boston, MA, American Meteorological Society, 325–327.

Venema, V., et al., 2006: Surrogate cloud fields generated with the iterative amplitude adapted Fourier transform algorithm. *Tellus A*, **58 (1)**, 104–120.

Von Hann, J. and R. Ward, 1903: *Handbook of climatology*. The MacMillan Company, New York.

Walther, G.-R., et al., 2002: Ecological responses to recent climate change. *Nature*, **416 (6879)**, 389–395.

Ward, R., 1906: The classification of climates: I. *Bulletin of the American Geographical Society*, **38 (7)**, 401–412.

Weaver, C. and V. Ramanathan, 1997: Relationships between large-scale vertical velocity, static stability, and cloud radiative forcing over Northern Hemisphere extratropical oceans. *Journal of Climate*, **10 (11)**, 2871–2887.

Wigley, T. and M. Schlesinger, 1985: Analytical solution for the effect of increasing CO_2 on global mean temperature. *Nature*, **315**, 649–652.

Wikle, C. K. and L. M. Berliner, 2006: A Bayesian tutorial for data assimilation. *Physica D*, **230**, 1–16.

Wing, S. L. and D. R. Greenwood, 1993: Fossils and fossil climate: the case for equable continental interiors in the Eocene. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **341 (1297)**, 243–252.

Wunsch, C. and P. Heimbach, 2014: Bidecadal thermal changes in the abyssal ocean. *Journal of Physical Oceanography*, **44 (8)**, 2013–2030.

Yun, K.-S., S.-Y. Kim, K.-J. Ha, and M. Watanabe, 2011: Effects of subseasonal basic state changes on rossby wave propagation during northern summer. *Journal of Geophysical Research: Atmospheres*, **116 (D24)**, D24 102.

Zelinka, M. D. and D. L. Hartmann, 2012: Climate feedbacks and their implications for poleward energy flux changes in a warming climate. *Journal of Climate*, **25 (2)**, 608–624.

Zhang, X., L. Alexander, G. C. Hegerl, P. Jones, A. K. Tank, T. C. Peterson, B. Trewin, and F. W. Zwiers, 2011: Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdisciplinary Reviews: Climate Change*, **2 (6)**, 851–870.