UNIVERSITY OF CALIFORNIA, SAN DIEGO

High-Frequency Climate Variability Associated with Stochastic Weather-Climate Interaction

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Oceanography

by

Nicholas Robert Cavanaugh

Committee in charge:

Alexander Gershunov, Co-Chair
Arthur J. Miller, Co-Chair
Ery Arias-Castro
Sarah T. Gille
Joel R. Norris
Samuel S.P. Shen
Shang-Ping Xie

2014

The dissertation of Nicholas Robert Cavanaugh is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____
Co-Chair

_____
Co-Chair

University of California, San Diego

2014

DEDICATION

For my Dad.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to begin by thanking my **parents** for giving me the freedom to make my own choices from a very young age. Sorry for the worry that I've caused, I suppose on at least a couple of occasions, but it seems to have all worked out so far. **Dad**, thanks for giving me the support and encouragement to pursue my academic goals. **Mom**, thanks for giving me your "emotional IQ." That's what keeps it all together. Oh, and thanks for the countless free flights. And sunglasses.

Next, I would like to thank **Art Miller** for giving me the chance to study at Scripps. I still remember first speaking with you over the phone back in college and hearing that I'd been accepted. Celebration ensued. On that note, I would also like to thank you for giving me enough rope to quite possibly hang myself, but trusting that I would put it to good use. Thank you **Sasha Gershunov**, for giving Art the reassurance (which I'm assuming you did) that statistics isn't (complete) voodoo magic. Also, thank both of you for all the credibility I received, and will likely continue to receive, when I meet people away from Scripps and they thearn that I was advised by both of you. *Both of you.*

Thank you **Sam Shen** for moving in three doors down from me in Nierenberg Hall. That was a weird twist of fate and I never would have expected us to do so much good science and make so much progress so quickly. I look forward to a careers worth of collaboration. Thank you **Aneesh Subramanian**. Words can't describe how helpful and generous you are with your time. You'll make a great professor one day. I would also like to thank my committee members **Ery Arias-Castro**, **Sarah Gille**, **Joel Norris**, and **Shang-Ping Xie** for lending their time and guidance throughout my tenure at Scripps.

Yelena Nam, thank you continuing to be my loving girlfriend for all these years. You are the lighthouse that helps make sure I don't steer into the rocks. I mean that in the most endearing possible way.

And thanks to all the friends along the way! **Jeff Campana**, **Sam Wilson**, **Jess Millar/Masich**, and **Matt (Siggy) Siegfried** you guys rock. Thanks for making the landing so seamless. Thanks to the home team, team Oliver: **Mike DeFlorio**, **Nick Pizzo**, **Brett Lyall**, our benchwarmer **Timmy Myers**, and our

team mascot **Rayleigh**. The almost constant and relentless distractions helped keep me sane. Thanks to the away team: **Drew Archer** and **Jimmy Metzger**, I hope that you one day read this. And finally, a special thanks to the rest of **The Miller Group** and the **Students at SIO** who make Scripps such a great place to study.

Portions of this dissertation have been published in peer reviewed journals. Chapter 2 is published in its entirety in:

- **Cavanaugh, N.R.**, Allen, T., Subramanian, A., Mapes, B., Seo, H., and Miller, A.J. (2014). The skill of atmospheric linear inverse models in hindcasting the Madden-Julian Oscillation. *Climate Dynamics*, doi 10.1007/s00382-014-2181-x.

Chapter 3 has been published in its entirety in:

- **Cavanaugh, N.R.**, and Shen, S.S.P. (2014). Northern Hemisphere Climatology and Trends of Statistical Moments Documented from GHCN-Daily Surface Air Temperature Station Data from 1950 to 2010. *Journal of Climate*, 27, 5396-5410.

Chapter 4 has been submitted for publication in its entirety:

- **Cavanaugh, N.R.**, and Shen, S.S.P. (2014). Spatial Scaling Effects on the Statistical Moments and Their Trends for Global Daily Surface Air Temperature. *Journal of Climate*, Submitted, *sub judice.*

Chapter 5 has been submitted for publication in its entirety:

- **Cavanaugh, N.R.**, Gershunov, A., Panorska, A.K., and Kozubowski, T.J. (2014). On the Probability Distribution of Daily Precipitation Extremes. *Geophysical Research Letters*, Submitted, *sub judice.*

Chapter 6 has been submitted for publication in its entirety:

- **Cavanaugh, N.R.**, and Gershunov, A. (2014). Probabilistic Tail Dependence of Precipitation on Spatiotemporal Scale in Observations, Reanalyses, and GCMs. *Climate Dynamics*, Submitted, *sub judice.*

VITA

2010	B.A.S. in Biomedical Science, University of Pennsylvania
2012	M.Sc. in Oceanography, University of California, San Diego
2014	Ph.D. in Oceanography, University of California, San Diego

PUBLICATIONS

Cavanaugh, N.R., and Gershunov, A. (2014). Probabilistic Tail Dependence of Precipitation on Spatiotemporal Scale in Observations, Reanalyses, and GCMs. *Climate Dynamics*, Submitted, *sub judice.*

Cavanaugh, N.R., Gershunov, A., Panorska, A.K., and Kozubowski, T.J. (2014). On the Probability Distribution of Daily Precipitation Extremes. *Geophysical Research Letters*, Submitted, *sub judice.*

Cavanaugh, N.R., and Shen, S.S.P. (2014). Spatial Scaling Effects on the Statistical Moments and Their Trends for Global Daily Surface Air Temperature. *Journal of Climate*, Submitted, *sub judice.*

Klingaman, N.P., Woolnough, S.J., Jiang, J., Waliser, D., Xiavier, P.K., Petch, J., Caian, M., Hannay, C., Kim, D., Ma, H-Y., Merryfield, W., Miyakawa, T., Pritchard, M., Ridout, J., Roehrig, R., Shindo, E., Vitart, F., Wang, H., **Cavanaugh, N.R.,**, Mapes, B., Shelly, A., Zhang, G. (2014) Vertical structure and diabatic processed of the Madden-Julian oscillation: Linking hindcast fidelity to simulated diabatic heating and moistening. *Journal of Geophysical Research* Submitted, *sub judice.*

Seo, H.,Subramanaian, A.C., Miller, A.J., and **Cavanaugh, N.R.** (2014). Coupled impacts of the diurnal cycle of sea surface temperature on the Madden-Julian Oscillation. *Journal of Climate*, Accepted. doi 10.1175/JCLI-D-14-00141.1.

Cavanaugh, N.R., and Shen, S.S.P. (2014). Northern Hemisphere Climatology and Trends of Statistical Moments Documented from GHCN-Daily Surface Air Temperature Station Data from 1950 to 2010. *Journal of Climate*, 27, 5396-5410.

Cavanaugh, N.R., Allen, T., Subramanian, A., Mapes, B., Seo, H., and Miller, A.J. (2014). The skill of atmospheric linear inverse models in hindcasting the Madden-Julian Oscillation. *Climate Dynamics*, doi 10.1007/s00382-014-2181-x.

Chavas, D., Yonekura, E., Karamperidou, C., **Cavanaugh, N.R.**, and Serafin, K. (2012). US Hurricanes and Economic Damage: An Extreme Value Perspective. *Natural Hazards Review*, 14(4), 237-246.

ABSTRACT OF THE DISSERTATION

High-Frequency Climate Variability Associated with Stochastic Weather-Climate Interaction

by

Nicholas Robert Cavanaugh

Doctor of Philosophy in Oceanography

University of California, San Diego, 2014

Alexander Gershunov, Co-Chair
Arthur J. Miller, Co-Chair

The statistics of high-frequency climate variability in observations and reanalyses are markedly non-Gaussian and show coherence across spatiotemporal scales. Dynamically, this complexity comes about as a manifestation of nonlinear terms in the equations of state and motion which dictate the time evolution of geophysical fluids in the oceans and atmosphere. A different perspective is to consider the climate system as consisting of dynamically resolved low-frequency components augmented by unresolved high-frequency components parameterized as stochastic noise. A stochastic formulation such as this is naturally suited toward studying climate variability and uncertainty since all spatiotemporal scales are explicitly or

implicitly resolved in its dynamics. The purpose of this dissertation is to examine weather predictability, variability, and uncertainty in the atmosphere as a function of spatiotemporal scale. A particular emphasis is placed on the quantification of the non-Gaussianity observed in surface air temperature (SAT) and precipitation time series at daily resolution and how these distributions scale in space and time.

The linear stochastic predictability of the tropical atmosphere is first examined through the use of linear inverse modeling (LIM) techniques. LIM extended-range weather predictions are nearly as skillful as fully nonlinear numerical climate models, suggesting that the tropics at daily timescales behave as a primarily linear dynamical system. Next, the variability and trends of daily SAT are studied using the first four statistical moments. It is shown that daily SAT behaves as an approximately locally homogeneous quasi-Gaussian random field whose statistics are consistent with correlated additive and multiplicative stochastic noise. The probability distributions of SAT at scale are shown to be related to regionally varying correlation length scales in the atmosphere. It is also shown that SAT distributions have undergone significant systematic changes over time as a result of low-frequency variability or climate change. Finally, it is documented that daily precipitation extremes are heavy-tailed globally over many spatiotemporal scales, suggesting that precipitation rates are power-law distributed, which is again consistent with stochastic theory. Current generation numerical climate models and reanalyses are shown to, by and large, recreate the power-law distributions documented from observations at the equivalent spatial resolutions. Results from this dissertation shed light on high-frequency climate dynamics and have practical implications for the quantification of weather event probabilities across spatiotemporal scales.

Chapter 1

Introduction

The mathematical concept of weather prediction grew out of the studies of fluid dynamics and thermodynamics in the early 20th century and matured parallel with the studies of nonlinear dynamics and complexity. In the 1950s, the use of new computational technology to run large-scale weather simulations stimulated the development of what is now modern scientific computing. Gordon E. Moore first quantified the exponential growth in computer power, coined Moore's law, which gave rise to more and more sophisticated simulations and better predictions of the weather. As theory postulated, the fidelity of weather models would grow hand-in-hand with computing hardware ad infinitum; however in the 1990s, the skill of numerical weather forecasts began to saturate. The problem bifurcated into a study of scales; the line between *weather prediction* and *climate prediction* was drawn.

1.1 Background

1.1.1 The Prediction Problem

The coupled global climate system is a *high-dimensional*, *multi-scale*, *nonlinear*, *deterministically chaotic* system. The dimensionality of a system refers to the number of independent components or variables required to completely define the entire system at any given time; while the dimensionality of the real climate

system is unknown, the number of variables simulated by today's climate models is $10^6 - 10^8$, and it is unlikely that it possesses a definite set of low-dimensional attractors (*Lorenz*, 1991). While a differentiation between weather (i.e., observable atmospheric conditions at a point in space and time) and climate (i.e., long-term and large-scale averages of the weather) is often made, these terms in their essence define arguments of scale. A multi-scale nonlinear system is characterized by coherence across spatiotemporal scales, with energy transitions in between. This characteristic signifies that scale-separation, a common simplifying assumption in physics, is inappropriate. Nonlinearity implies the existence of feedback loops among system components, suggesting that no component is negligible with respect to the entire system. This characteristic of nonlinearity also implies the presence of non-Gaussian variability statistics (*Palmer*, 1993) which will be a reoccurring theme in this dissertation. Finally, a chaotic system is non-periodic and never returns to the same system state - a fundamental characteristic of dynamical unpredictability.

In the abstract of his 1963 paper "Deterministic Nonperiodic Flow," Edward Lorenz remarked about such systems:

> [For] finite systems of deterministic ordinary nonlinear differential equations designed to represent forced dissipative hydrodynamic flow with bounded solutions, it is found that nonperiodic solutions are ordinarily unstable with respect to small modifications, so that slightly differing initial states can evolve into considerably different states. ... A simple system representing [atmospheric] cellular convection is solved numerically. All of the solutions are found to be unstable, and almost all of them are nonperiodic. The feasibility of very-long-range weather prediction is examined in the light of these results (*Lorenz*, 1963).

Clearly, the inherent difficulty in the prediction problem has been known for quite some time.

Both numerical climate and weather prediction are accomplished through the discretization of mass and energy, either in physical space or in spectral space, followed by the integration of the system iteratively through time over very small discrete intervals. The dynamical cores of these models, which dictate the dynamics of energy transfer, are composed of the three so-called primitive equations:

1. the continuity equation, which dictates the conservation of mass within the system.

2. the conservation of momentum equation(s), consisting of the quadratically nonlinear Navier Stokes equation(s), which dictate hydrodynamical flows,

3. the thermal energy equation, which dictates the conservation of energy through the first law of thermodynamics and,

A thorough treatment of these equations and their numerical implementation can be found in *Warner* (2011)

1.1.2 The Forecasting Problem

Up until this point, I have been intentionally vague as to the definition of the term "prediction." Generally, we associate prediction with a seer-like premonition of the occurrence of a binary event some time in the future: event a will happen at time t. Predictions can be likely: *the sun will shine in San Diego tomorrow*. Predictions can be unlikely: *the sun will not shine in San Diego tomorrow*. Predictions are testable, assuming of course that San Diego and the sun do not cease to exist before tomorrow.

Numerical weather models naturally make predictions. Assuming that we can wholly define the entire state of the climate at an initial time, one can propagate the numerical model forward in time and define a new state at some time in the future, which is of course our prediction. More careful analysis of this methodology in the context of climate and weather prediction reveal many critical problems, such as,

- We do not have complete estimates of the climate state, do not know what the complete climate state constitutes, and what estimates we do have are uncertain,

- We have only uncertain knowledge about many of the complex physical systems critical to large-scale prediction, and

- We may be missing important physical systems or fundamental dynamical scales.

Each of these issues can wreak havoc on predictions due to the nonlinear and chaotic nature of the system in question.

Alternatively, we are also capable of making "forecasts." A forecast defines a range of possible outcomes, ideally with probabilities attached, which encapsulate all the possible future outcomes and their likelihoods; this concept is alternatively called probabilistic prediction. Unlike deterministic prediction, a forecast cannot be replicated by the real system since the actual outcome is only one possibility in a range of possible outcomes; rather, forecasts and forecasting methodologies can be validated over time to be skillful. While it may seem counter-intuitive in this setting since the atmosphere (or ocean) can only take one state at any one time, it is actually quite a natural construct for a viscously damped system such as the atmosphere (or ocean). In the numerical construct, a probabilistic prediction is usually accomplished by perturbing estimates of the initial conditions (which is done in weather forecasting) or the model physics (which is done in climate forecasting/projections), and integrating each perturbed sample separately to form an ensemble. The statistics of the ensemble constitute the basis for a probabilistic prediction, however such ensemble methods are extremely computationally demanding.

1.1.3 Stochastic Climate Models

An alternative point of view counter to the deterministic perspective on prediction offered in Sections 1.1.1 and 1.1.2 is to view the atmosphere (or ocean) as an idealized system of slowly evolving low-frequency deterministic components augmented by fast and unresolved high-frequency components approximated by stochastic noise. As alluded to previously, the dynamical cores of numerical models take the quadratically nonlinear form

$$\frac{du}{dt} = F + Au + B(u,u)\ ;\ \ u \cdot B(u,u) = 0 \tag{1.1}$$

where F is a constant forcing, A is a linear operator, B is a quadratically nonlinear operator, and $u \in \mathbf{R}^M$ is an M-dimensional representative state vector with $M >> 1$. Consider the separation of u into a resolved low-frequency component $x \in \mathbf{R}^N$, $N << M$, and an unresolved high-frequency component $y \in \mathbf{R}^{N'}$, $N' = M - N$ so that $u = (x_1, \ldots, x_N, y_1, \ldots, y_{N'})$. Then, following Majda et al. (2009), by assuming dx/dt has up to a quadratically nonlinear deterministic dependence on x plus the appropriately scaled diad and triad interactions between x and y, and dy/dt has linear dependence on y, the appropriately scaled triad and diad interactions between x and y, plus the appropriately scaled stochastic white noise, a stochastic differential equation (SDE) for the evolution of x can be shown to be

$$dx_i = A_{ij}x_j dt + \sum_p C_{ij,p} dW_{j,p} \tag{1.2a}$$

$$+ (constant)_i dt + F_i dt + B_i(x_i, x_j)dt \tag{1.2b}$$

$$- \sum_p \gamma_p^{-1} I_{ij,p}^2 x_j^3 dt + \sum_{j \neq i,p} \gamma_p^{-1} I_{ij,p} I_{ji,p} x_i x_j^2 dt \tag{1.2c}$$

$$+ \sum_p \gamma_p^{-1} \sigma_p (L_{i,p} - I_{ij,p} x_j) \circ dW_p \tag{1.2d}$$

where A is a square linear operator matrix, C is a square covariance matrix, p denotes the unresolved variables, L are linear damping coefficients, I arises from cross term interactions between x and y, σ and γ are scaling coefficients, and W is a Wiener process. In this case, the constant term can give rise to a non-physical climate drift and is likely zero, however drift from external forcing F_i can manifest itself as non-stationarity over time. A model in this form is called a stochastic climate model (*Majda et al.*, 2009).

Stochastic climate models rely on a somewhat subjective choice concerning the scope of the deterministic (or resolved) component of the model x. Usually, u is defined as the sum of the principal components (PCs) of empirical orthogonal functions (EOFs) taking the form

$$u = \sum_{k=1}^{N} u_k(t) e_k + \sum_{k=N+1}^{M} u_k(t) e_k \tag{1.3}$$

where each k constitutes an orthonormal energy component with an e eigenfunction, and thus x is defined as some number N of leading EOFs and y is implicitly

defined as the rest of the observed variability. In this formulation, eq. (1.2) would be called a reduced stochastic climate model, since the model only deterministically resolves some portion of the observed variability. Alternatively, if x is taken to be all the variables conserved in a traditional numerical climate model, then y would be analogous to a subgrid-scale parameterization, and thus the model of eq. (1.2) would have the same order of complexity as a full numerical model. Assuming a stationary system, if we take x to be the climatological anomalies, eq. (1.2) must be damped and in the limit as $t \to \infty$, the statistics of the forecasts are entirely stochastic; in Chapters 3 and 5, we investigate the statistics of daily surface air temperature (SAT) anomalies and precipitation, and in chapters 4 and 6, we investigate the statistical relationships between grid-scale and subgrid-scale variability.

Stochastic climate models are particularly well suited for producing forecasts since they

- include all spatiotemporal scales, explicitly through x and implicitly through y,
- are computationally efficient compared to full numerical models, particularly when in a reduced form and,
- provide an analytical construct for studying the distribution of x through the Fokker-Planck equation(s) as reduced, yet still physical meaningful, climate models.

1.2 Outline of the Dissertation

1.2.1 The Linear Inverse Model

By forcing the coefficients in eqs. (1.2b-d) to zero and setting $p = i$ and $dW_{i,p>1} = dW_{i,p=1}$, the linear inverse model (LIM) of *Penland and Magorian* (1993) is obtained. The LIM is one of the simplest of the possible stochastic climate models and has been used with relatively high degrees of success to model and

predict both oceanic and atmospheric conditions (see Chapter 2). In LIM, $x(t)$ can be solved analytically and used to make probabilistic predictions. The solution involves two components: the first term is a linear operator matrix which dictates the mean predicted state of the system at some time t in the future, analogous to an ensemble mean, and can also be dissected to yield insight about the dynamical interactions between the resolved components of the state vector. The second term is a stochastic noise component, which summarizes uncertainty in the probabilistic prediction as a multivariate Gaussian and can be used to make forecasts. It can also be decomposed to draw insight about the stochastic nature of the climate system (*Penland and Matrosova*, 1994).

In Chapter 2, we utilize reduced LIMs to make extended-range weather predictions in the in the tropics. We focus on hindcasting the Madden-Julian Oscillation, the major component of intraseasonal variability in the tropics, and demonstrate that simple linear stochastic representations of large-scale climate are nearly as skillful as full-form nonlinear numerical models at extended-range weather prediction. We additionally examine some dynamical features present in our models which provide physical intuition about the nature of prediction in the tropics and the propagation of energy across spatiotemporal scales.

1.2.2 Statistics of Daily Surface Air Temperature

By neglecting the nonlinear terms in eq. (1.2), the Fokker-Planck equation for the predicted steady-state probability distribution function (pdf) of $x(t)$, or $p(x)$, can be solved for analytically (*Sura*, 2011). As mentioned in Section 1.2.1, when neglecting eq. (1.2d) and solving only for eq. (1.2a), the resulting pdf is distributed as a multivariate Gaussian. Including eq. (1.2d) in the Fokker-Planck equation, however, yields non-Gaussian pdfs.

While atmospheric variables on monthly timescales are approximately normally distributed (*Stephenson et al.*, 2004), in Chapter 3 we demonstrate that the statistics of SAT at daily resolution estimated from weather stations are non-Gaussian and form coherent spatial structures in moments over the Northern Hemisphere. The observed weather statistics are consistent with those of the pdf for a

linear stochastic climate model forced with correlated additive and multiplicative (CAM) noise taking the form of eq. (1.2d) (*Sardeshmukh and Sura*, 2009). We additionally determine that these moments have statistically significant trends over the latter half of the 20th century, suggesting that low-frequency climate variability and/or climate change has had detectable impact on the statistics of daily weather probabilities.

1.2.3 Spatial Scaling of Surface Air Temperature

In Chapter 4, we investigate the relationship between SAT variability observed at point-scale with that estimated at grid-scale. We demonstrate that the statistics of daily SAT variability are closed at the fourth moment and constitute an approximately locally homogeneous quasi-Gaussian random field (*Millionshchikov*, 1941; *Monin and Yaglom*, 1971). As a consequence, we determine that the moments at grid-scale are mathematically related to correlation length scales which can be estimated from station data and vary geographically. This study provides a quantitative assessment about the impacts of spatial scaling on distribution by smoothing atmospheric variables.

1.2.4 Distribution of Daily Precipitation

In Chapter 5, we continue our investigation of daily weather variability observed at point-scale. We determine that in most places, globally, the fundamental distribution of daily precipitation is heavy-tailed. This conclusion is consistent with the stochastic theory defined in Section 1.1.3; as a consequence of the CAM noise discussed in Section 1.2.2, the tail of the pdf of x is expected to follow a power-law (*Penland and Sardeshmukh*, 2012). We demonstrate that the thickness of the high-valued probability tail, i.e. precipitation volatility, is physically related to storm type heterogeneity, which varies geographically.

1.2.5 Spatiotemporal Scaling of Extreme Precipitation and the Simulation of Precipitation Fields in Numerical Climate Models

In Chapter 6, we investigate the effects on distribution type resultant from spatiotemporal smoothing of precipitation fields. We determine that at high temporal resolutions, precipitation fields remain power-law distributed until over 200km and at high spatial resolutions, precipitation fields remain power-law to between 2- and 3-week accumulations. Following from our spatial scaling investigation, we find ourselves equipped with the data necessary to validate the distributions simulated by numerical climate models and reanalyses at the appropriate spatial scales; we go on to investigate the skill with which seven global reanalyses and forty global coupled climate models recreate the precipitation volatility observed over global land. Finally, we present a physically consistent global picture of daily precipitation distribution type estimated from the best performing global reanalysis dataset at 0.31° spatial resolution.

1.3 A Note to Readers

Each chapter contained within this dissertation has been designed to stand alone. It follows that in some instances, mathematical notation may differ between chapters, and redundancy in definitions, explanations, and/or references may occur.

Chapter 2

The Skill of Atmospheric Linear Inverse Models in Hindcasting the Madden-Julian Oscillation

A suite of statistical atmosphere-only linear inverse models (LIMs) of varying complexity are used to hindcast recent MJO events from the Year of Tropical Convection and the Cooperative Indian Ocean Experiment on Intraseasonal Variability / Dynamics of the Madden-Julian Oscillation mission periods, as well as over the 2000-2009 time period. Skill exists for over two weeks, competitive with the skill of some numerical models in both bivariate correlation and root-mean-squared-error scores during both observational mission periods. Skill is higher during mature Madden-Julian Oscillation conditions, as opposed to during growth phases, suggesting that growth dynamics may be more complex or nonlinear since they are not as well captured by a linear model. There is little prediction skill gained by including non-leading modes of variability.

2.1 Introduction

The Madden-Julian Oscillation (MJO) is a large-scale zonally propagating atmospheric signal in tropical rainfall and related fields (*Madden and Julian*, 1971) and is the dominant mode of intraseasonal variability in the tropics (*Wheeler and*

Hendon, 2004, hereafter WH04). The MJO modulates high-frequency weather, both in the tropics and extra-tropics through teleconnections (*Wallace and Gutzler*, 1981; *Ferranti et al.*, 1990; *Maloney*, 2000; *Matthews and Meredith*, 2004; *Cassou*, 2008; *Roundy and Gribble-Verhagen*, 2010; *Martin and Schumacher*, 2011), and has also been shown to affect longer timescale climate variability, for example the El Niño Southern Oscillation (*Lau and Chan*, 1985, 1988; *Kessler and Kleeman*, 2000; *Zhang*, 2001; *Subramanian et al.*, 2011). Many studies have suggested that the MJO may provide an avenue for predictability beyond the traditional 10-day limit (*Waliser et al.*, 2003; *Reichler and Roads*, 2005). Verification of predictions may be done in the space of the real-time multivariate MJO index (RMM, see WH04), comprising the leading two maximum covariance or empirical orthogonal function (EOF) modes of combined tropical 200mb and 850mb zonal wind (u200 and u850, respectively) and outgoing long-wave radiation (OLR) in the intraseasonal band. Bivariate correlations of forecasts and verifications of these two modes, and root-mean-squared error (RMSE) metrics proposed by *Lin et al.* (2008), make a well-accepted set of verification metrics (*Gottschalck et al.*, 2010) which we will use here. Specific details of the metrics and their interpretations can be found in those papers.

MJO hindcast skill studies utilizing high-dimensional numerical models have increased in recent years (*Zhang et al.*, 2013a). These hindcasts are usually produced for time periods coinciding with large, coordinated MJO research missions, such as the Year of Tropical Convection (YOTC, *Waliser and Moncrief*, 2007), and Cooperative Indian Ocean Experiment on Intraseasonal Variability / Dynamics of the Madden-Julian Oscillation (CINDY/DYNAMO, *Yoneyama et al.*, 2013). Several statistical forecast studies relevant to the MJO have been reported. Simple extrapolation by Fourier filtered zero-padded longitude-time sections (*Wheeler and Weickmann*, 2001) is one approach. Many kinds of covariance models have also been tried, including Principal Oscillation Patterns (POP, *von Storch et al.*, 1988); a singular value decomposition forecasting technique (*Waliser et al.*, 1999); and lagged regressions on EOFs of OLR and circulation indices, sometimes bandpass filtered (*Lo and Hendon*, 2000; *Jones et al.*, 2004). *Maharaj and*

Wheeler (2005) and *Jiang et al.* (2008) are regression models based on the RMM1 and RMM2 EOFs (convenient for verification). *Kang and Kim* (2010) summarize the predictability from a collection of statistical and dynamical models, but the variety of skill tests utilized across past studies complicates comparisons.

The Linear Inverse Model (LIM; *Penland and Magorian*, 1993) constitutes the least complex form of a reduced stochastic-dynamic climate model (*Majda et al.*, 2009) and has been used for diagnostics and prediction in several studies of the atmosphere (e.g.; *Winkler et al.*, 2001; *Newman et al.*, 2003; *Pegion and Sardeshmukh*, 2011) and coupled atmosphere-ocean system (*Newman et al.*, 2009). These LIMs have been shown to have comparable predictive capability to general circulation models, even though they have far fewer degrees of freedom. *Pegion and Sardeshmukh* (2011) compared a simple atmospheric LIM to coupled atmosphere-ocean numerical models to conclude that there is room for hindcast skill improvement in tropical climate prediction specifically.

In this paper, we explore the use of atmospheric LIMs in the established context of MJO forecast verification. We will see that simple stochastic-dynamic representations can provide hindcast skills comparable to other statistical approaches and to some GCMs, and offer useful informed hypotheses about MJO dynamics along the way. Section 2 introduces LIM theory briefly, and establishes the subset of models used in this analysis. Section 3 shows LIM hindcast skill for the DYNAMO period, two YoTC MJO events, and a ten-year continuous hindcast period. Finally, we conclude with a discussion and directions for future research.

2.2 Model Details and Methodology

Using the notation of *Newman et al.* (2009), the fundamental assumption underpinning LIM is that the governing dynamics of the system under consideration can be modeled as

$$\frac{d\mathbf{x}}{dt} = \mathbf{L}\mathbf{x} + \xi \tag{2.1}$$

where $\mathbf{x}$ represents an appropriate system state vector, $\mathbf{L}$ is a linear operator matrix, and ξ is a vector of stochastic temporally white but spatially structured

Gaussian noise. In a system where eq. (2.1) is stable, lag-covariance matrices decay exponentially, so $\mathbf{L}$ can be estimated from observational estimates of covariance matrices $\mathbf{C}(\tau)$ where $\mathbf{C}_{ij}(\tau) = \langle \mathbf{x}_i(t+\tau)\mathbf{x}_j(t) \rangle$ evaluated at any fixed lag, where subscripts i and j correspond to covarying observational time series, as well as the rows and columns of $\mathbf{C}$, respectively. For some chosen lead-time τ_0, $\mathbf{L}$ is estimated as $\mathbf{L} = \tau_0^{-1} \ln[\mathbf{C}(\tau_0)\mathbf{C}^{-1}(0)]$. Equation (2.1) can then be solved for analytically: $\mathbf{x}(t+\tau) = \mathbf{G}(\tau)\mathbf{x}(t)+\epsilon$ where $\mathbf{G}(\tau) = \exp(\mathbf{L}\tau)$ represents the decaying, predictable signals at forecast lead time τ, and ϵ is a random variable vector with covariance $\mathbf{E}(\tau) = \mathbf{C}(0) - \mathbf{G}(\tau)\mathbf{C}(0)\mathbf{G}^{-1}(\tau)$. The random vector ϵ is multivariate Gaussian and grows as a function of τ regardless of the initial condition at time t and can also be estimated from a suitable set of hindcast errors. More comprehensive overviews of LIM can be found in *Penland* (1989) and *Penland and Sardeshmukh* (1995) and its application in the tropics in *Newman et al.* (2009) and *Pegion and Sardeshmukh* (2011).

The data channels that constitute $\mathbf{x}$ may be time series of variables in physical space, or principal component (PC) time series of EOFs of the data comprising the desired forecast space. The diagonal components of $\mathbf{L}$ express the individual decay of each of the predicted variables, whereas the off-diagonal components of $\mathbf{L}$ represent modal interactions (if $\mathbf{x}$ is in a mode basis) or propagation (if $\mathbf{x}$ is in a spatial basis). Imaginary $\mathbf{L}$ components describe oscillations (e.g. in the POP model of *Penland*, 1989). After fitting $\mathbf{L}$, the residual noise can be obtained. Its EOFs describe leading spatially coherent patterns of stochastic forcing necessary to reproduce the observed data, and thus might be interpretable in terms of sources of high-frequency (spectrally white) turbulent or chaotic energy in nature, which are being parameterized in the LIM as noise (see *Penland and Matrosova*, 1994).

In this study, we chose a reduced climate state vector $\mathbf{x} = [\mathbf{W}\ \mathbf{O}]^T$, where $\mathbf{W}$ is the PCs associated with some number of leading EOFs of concatenated 850mb and 200mb u- and v- wind anomalies from the National Centers for Environmental Prediction Climate Prediction Center (NCEP/CPC) Reanalysis 2 (*Kanamitsu et al.*, 2002), and $\mathbf{O}$ is some number of leading PCs of OLR anomalies from the National Oceanic and Atmospheric Administration (NOAA) Interpolated OLR

dataset (*Liebmann and Smiith*, 1996). Both datasets are used from 1980-2012. To define these EOF-PC pairs, all variables were first truncated to T21 resolution, transformed to a Gaussian grid from 25°S - 25°N, and smoothed with a 7-day running mean filter, as in *Newman et al.* (2009). Unsmoothed data withheld from the training set are projected onto the leading EOFs from the smoothed data to produce out-of-sample hindcasts, which do not utilize any future data and are therefore valid as daily predictions. These smoothing techniques were selected to attenuate unpredictable (high-frequency) signals that can corrupt LIM dynamics through aliasing, but in this case MJO hindcast skills are largely insensitive to these filtering choices since EOF shapes are dominated by large-scale and low-frequency variability even in daily fine-grid data. LIMs constructed directly from unfiltered data EOFs provide similar predictions and corroborate the results of this study.

Four-week tropical hindcasts were initiated for each day Jan 1, 1999 - Dec 31, 2012, for each combination of $[4, 8, 12, \ldots, 40]$ EOFs of $\mathbf{W}$, $[4, 8, 12, \ldots, 40]$ EOFs of $\mathbf{O}$, and $\tau_0 = [2, 4, \ldots, 14]$, resulting in 700 models. LIM hindcasts are cross-validated by excluding the data during the year in which the hindcast is constructed when fitting the model. Model hindcasts were transformed back into physical space and then projected onto the WH04 RMM indices (which only involve zonal winds and OLR) for skill verification.

2.3 Results

The lowest complexity model is chosen as a baseline for comparison with a training lag time of eight days: four wind EOFs and four EOFs of OLR in the 25N-25S belt are included, summarized as 4-$\mathbf{W}$ 4-$\mathbf{O}$, and $\tau_0 = 8$. Figure 2.1 shows bivariate correlation (left column) and RMSE (right column) for the Jan 1, 2000 - Dec 31, 2009 interval. Overall, the all-season ten-year MJO hindcast skill is quite consistent across all our models. Decorrelation times, measured as the time when correlation falls below 0.5 (*Gottschalck et al.*, 2010), are about 14-16 days. Similarly, RMSE skill, as measured by the lead-time at which the RMSE crosses

2 (*Lin et al.*, 2008; *Vitart et al.*, 2010) across all models is also about 14 days.

Figures 2.1a,b show MJO skill sensitivity to OLR truncation. All conclusions are consistent in both the correlation and the RMSE skill metrics. Hindcast skill improves slightly by increasing the number of OLR predictors included in the model (Fig. 2.1a,b) until the skill level saturates at a high complexity. Similarly, including more **W** modes increases model skill essentially monotonically, particularly after about 7 days of hindcast (Fig. 2.1c,d), but with similar forecast skill prior to 7 days (when lower complexity therefore suffices), again before the skill level saturates at a high complexity. MJO hindcast skill is insensitive to τ_0 (Fig. 2.1e,f), indicating smoothly and exponentially decaying lag-covariances, as assumed in the LIM theory (*Penland and Sardeshmukh*, 1995). These results suggest that there is little linear MJO predictability (about 1-2 day skill extension) associated with non-leading EOFs, and that under this model construction, prediction skill is independent of the training lead-time. Sensitivities for models composed of unfiltered EOFs are slightly increased, however, the main conclusions remain consistent with those presented here.

Figures 2.2a-d show the structure of the first four EOFs of OLR used in the model. The first two EOFs show similar equatorially symmetric patterns roughly phase-shifted from each other. The first pattern shows a pattern of low OLR in the central Pacific region when the principal component is positive while the second reveals a pattern with lowest OLR in the Western Pacific region corresponding to a positive PC value. The third and fourth EOFs of OLR have both equatorial and off-equatorial structures in the West Pacific and reveal patterns with higher wavenumbers (wavenumbers 2-3) in the equatorial region.

Spatial patterns of the first four EOFs of the combined 850mb and 200mb winds are shown in Figure 2.3a-h. Shading indicates regions of convergence and divergence with blue indicating regions of divergence and red indicating regions of convergence corresponding to a positive PC value. The first EOF of the winds at 850mb and 200mb (Figures 2.3a-b) show regions of strong high-level convergence and low-level divergence over the Western Pacific Ocean region. This region also corresponds to the concentrated variability in OLR highlighted by the first pair of

EOFs of OLR. The second, third and fourth EOFs of winds reveal off-equatorial regions of increased variability in divergence especially in the region of the Inter-Tropical Convergence Zone.

Leading EOFs of both OLR and the combined winds have broad spectral peaks in the intraseasonal and interannual bands (not shown). Interactions of these modes, as quantified by the matrix of interaction coefficients $\mathbf{G}(\tau)$ (demonstrated in Figure 2.4 as $||\mathbf{G}(\tau = 7, 14, 21)||$), confirm that the magnitudes of the interactions between the leading EOFs (Figure 2.4, $\alpha - \delta$) are dominant in comparison to interactions between the non-leading EOFs, which tend to decay toward zero for lesser modes of variability particularly for large τ. Interestingly, interaction coefficients $\mathbf{G}_{i<j}(\tau)$ for the **OO** matrix block are generally of greater magnitude than coefficients $\mathbf{G}_{i>j}(\tau)$, indicating that the upscale interactions from modes of lower variability to those of higher variability play a more dominant role than those of downscale interactions in intraseasonal forecasts.

Figures 2.5a,b show bivariate correlation and RMSE for multiple hindcast intervals from the best-performing model (24-**W**, 32-**O**, $\tau_0 = 6$) based on bivariate correlation after 10 days over the whole 2000-2009 period. Results are comparable to the numerical model scores for the YoTC hindcast period of Oct. 10, 2009-Nov. 25, 2009 and Dec. 10, 2009-Jan. 25, 2010 (*Klingaman et al.*, 2014) and the DYNAMO hindcast period of Sept. 1, 2011-Mar. 31, 2012 (*Zhang et al.*, 2013a, Figure 2b). Useful bivariate correlation skill ranges during YoTC for numerical models are approximately 8-15+ days, and approximately 8-20+ days for DYNAMO. LIM hindcast skill for each comparable time period is on the low end of the dynamical model intercomparisons, but within the model spread, for both correlation and RMSE. This is particularly true during the DYNAMO period, where LIM skill is 18 days. MJO skill scores are broken out for periods where the hindcast initialization date has a combined RMM amplitude of < 1 (weak), $1 < ||\text{RMM}|| < 2$ (moderate), and $2 < ||\text{RMM}||$ (strong). LIM underperforms compared to overall hindcast skill during weak MJO conditions as indicated by rapidly decorrelating hindcasts, however performance is markedly improved for moderate and strong MJO hindcasts (Figure 2.5a) for each time period relative to

hindcasts initialized from weak MJOs. Figures 2.5c,d show bivariate correlation and RMSE separated by MJO phase at model initialization for the 2000-2009 period. LIM performance is relatively poor during phases 1, 2, and 6, with prediction skill to 14 days. Prediction during other phases is extended to between 15-19 days for bivariate correlation.

Figure 2.6 shows MJO hindcast propagation for the YoTC E case, Oct 10, 2009-Nov 25, 2009 for the best-performing model. The YoTC E case is selected since both the behavior of observations and its hindcasts most clearly illustrate the essential performance features of LIM, which are common across many MJO events, as well as the inability of LIM to capture (perhaps nonlinear) deviations from a smoothly propagating MJO evident in this particular event. Figure 2.6a shows multiple 3-week hindcasts initialized every 7 days over the same MJO event. Model initialization RMM values do not perfectly coincide with observation RMMs due to our EOF truncation and filtering choices. LIM forecasts are poor during the weak and moderate stages of MJO initiation, which correspond to times preceding amplitude errors that grow rapidly. Figure 2.6a shows rapidly intensifying amplitude errors during MJO initiation, which are consistent with information from Figures 2.5c and d that highlight the relatively poor performance of MJO hindcasts during phases 1 and 2. This result indicates that rapid MJO initiation depends on either non-linear phenomena (e.g. *Straub*, 2013; *Kemball-Cook and Weare*, 2001) and/or processes, such as air-sea interaction in the Indian Ocean (*Yoneyama et al.*, 2013) perhaps, with no linear projection on the set of current model variables in the historical training data. Once the MJO has reached a mature amplitude, however, hindcasts performance improves both in propagation speed and amplitude.

Figure 2.6b shows a series of one-week lead-time hindcasts. The nature of this plot is that each hindcast day is initialized one-week prior and the hindcast plot responds as it "learns" of MJO initiation information. In this sense, the LIM hindcast tends to mimic MJO behavior when it incorporates the observed MJO state at the constant lead-time lag. For example, the LIM produces a stalled MJO around October 30 (index date 20 on Fig. 2.6b) since during the initiation day, October 23 (index 13 on Fig. 2.6b), the MJO was stalled. The failure of LIM

to capture the stalling MJO behavior indicates that this YoTC E specific MJO initiation feature is inconsistent with linearized historical observations. Model behavior in these hindcast plots is typical and is similar for other MJO events observed during the YoTC and DYNAMO time periods (not shown).

2.4 Concluding Remarks

This paper presents hindcast skill specific to MJO for a suite of LIMs composed of OLR and 200mb and 850mb winds over multiple intervals from 2000 to 2012. Results show that LIM skill is on the low end of current full-physics numerical models, but is within the model spread for bivariate correlation and RMSE during two comparison periods. This study highlights that these extremely simple empirical models perform competitively with GCMs at the low-end of the skill spread in MJO hindcasts. The LIMs presented here also perform competitively with other MJO statistical forecasting techniques, but are designed to reproduce full spatial fields of tropical variability, making them more comparable to numerical model output. We have chosen here to focus on skill related to the RMM indices to maximize the relevancy in the MJO community. Very little additional prediction skill is gained through the inclusion of more modes of variability, which in this case suggests that the simplest LIMs are likely most appropriate for MJO prediction, however the more complex LIMs may be useful toward achieving other goals, such as forecasting higher wavenumber features. This study also provides a forecast performance baseline for comparison with other forecast methodologies, particularly if only a short hindcast or observation period is available. Lastly, the analytical noise solution ϵ provides the statistics of an infinite-member ensemble forecast that may be used for uncertainty quantification at a much lower computational cost than producing large GCM ensembles to sample the uncertainty space.

The LIM performs particularly well during mature stages of the MJO. This suggests either that the LIM formulations presented here are missing key components of MJO initiation and amplification that are not linearly related to $\mathbf{x}$ or

that MJO amplification marks a period of dominantly non-linear deviation from an otherwise more linear system which may be captured by better performing numerical models. Alternately, smoothly propagating mature MJO events, whose amplitudes and phase speeds are well captured by LIM, suggest that at maturity, MJO propagation behaves as a more linear system of traveling waves. Prediction skill during the YoTC events is lower than that during the DYNAMO and 2000-2009 validation periods, however this behavior may be a product of the short YoTC validation period.

There are many possible extensions to the LIMs presented here. *Kondrashov et al.* (2013) investigated the use of a quadratic inverse model of RMMs for MJO forecasting, which produce similar hindcast skill to our models. This is perhaps because the dominant quadratic terms are two orders of magnitude smaller than the linear coefficients (*Kondrashov et al.*, 2013, Supplementary material) and that there is at least some predictability beyond the RMMs which is captured by our models. That same study, however, achieved large hindcast skill improvements by training the model noise component only from time periods when the historically observed dynamics were similar to those observed at the hindcast initiation with so called past-noise forecasting. Another possible extension is to add a seasonally cyclostationary time-dependence, originally proposed by *OrtizBevia* (1997), by conditioning $\mathbf{L}$ and ξ on the time of the year either as piecewise linear or subject to a continuous empirical model. Previous studies have constructed seasonal LIMs (e.g. *Winkler et al.*, 2001; *Newman et al.*, 2003; *Pegion and Sardeshmukh*, 2011) and have achieved marginal skill increases over an all-season LIM. Ocean-atmosphere coupled LIMs may also add realism, but are unlikely to yield additional MJO hindcast skill due to the nearly uncoupled nature of the ocean and atmosphere at subseasonal timescales (*Newman et al.*, 2009). Since our all-season LIM does not perform equally well at all amplitudes and phases of the MJO, it is possible that an MJO behavior-, amplitude- or phase-dependent LIM, resulting in a piecewise-stationary linear model, could yield large increases in MJO forecast skill, provided any hindcast stitching procedures do not produce amplifying errors and that the observed model skill deficiencies are not merely manifestations of the

intrinsic predictability of the tropical atmosphere itself.

Lastly, it is well known that linear stochastic models with Gaussian white noise produce Gaussian forecasts, whereas high-frequency climate statistics are markedly non-Gaussian (*Perron and Sura*, 2013). An alternative mode reduction strategy, strategic choice of additional independent observable inputs containing MJO precursor and dynamics information, and/or augmentation of the system with correlated additive and multiplicative noise, suggested by *Sura et al.* (2005), may improve forecasts by accounting for more aspects of variability, while maintaining the simplicity of a linear stochastic framework.

2.5 Acknowledgements

This material is based partly upon work supported by the National Science Foundation under Grant No. 0731520. We also gratefully acknowledge funding from ONR (Grants N00014-10-1-0541, N00014-13-1-0139, and N00014-13-1-0704) and NSF (OCE-0960770). We would like to thank Prashant Sardeshmukh, Cecile Penland, and Mathew Newman for conversations regarding LIM. Allen would like to thank Kathy Pegion of NOAA for her time and discussions towards understanding the LIM framework and its applications. We greatly appreciate the insightful comments of the four reviewers, whose input greatly enhanced the quality of this manuscript. In addition, Allen is grateful for the Wallace Fellowship administered by The University of Miami that helped to foster collaborative efforts among the authors. The chapter has been published in: **Cavanaugh, N.R.**, Allen, T., Subramanian, A., Mapes, B., Seo, H., and Miller, A.J. (2014). The skill of atmospheric linear inverse models in hindcasting the Madden-Julian Oscillation. *Climate Dynamics*, doi 10.1007/s00382-014-2181-x.

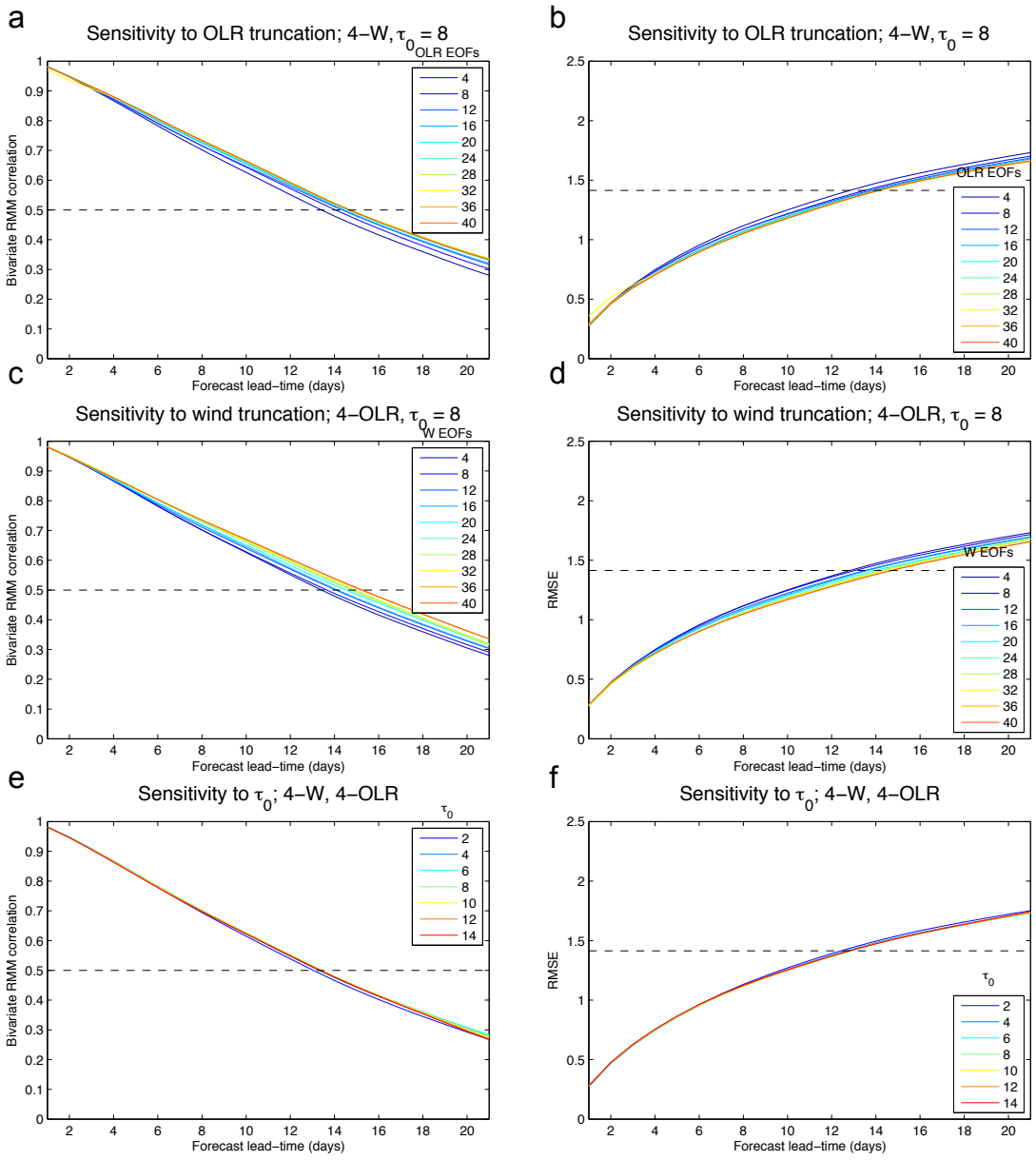

Figure 2.1: Bivariate correlation (left column) and RMSE (right column) for LIM suite hindcasts over the 2000-2009 validation period. Subplots a-b show MJO skill sensitivity to OLR truncation. Subplots c-d and e-f are the same as a-b except for winds and τ_0 sensitivity, respectively.

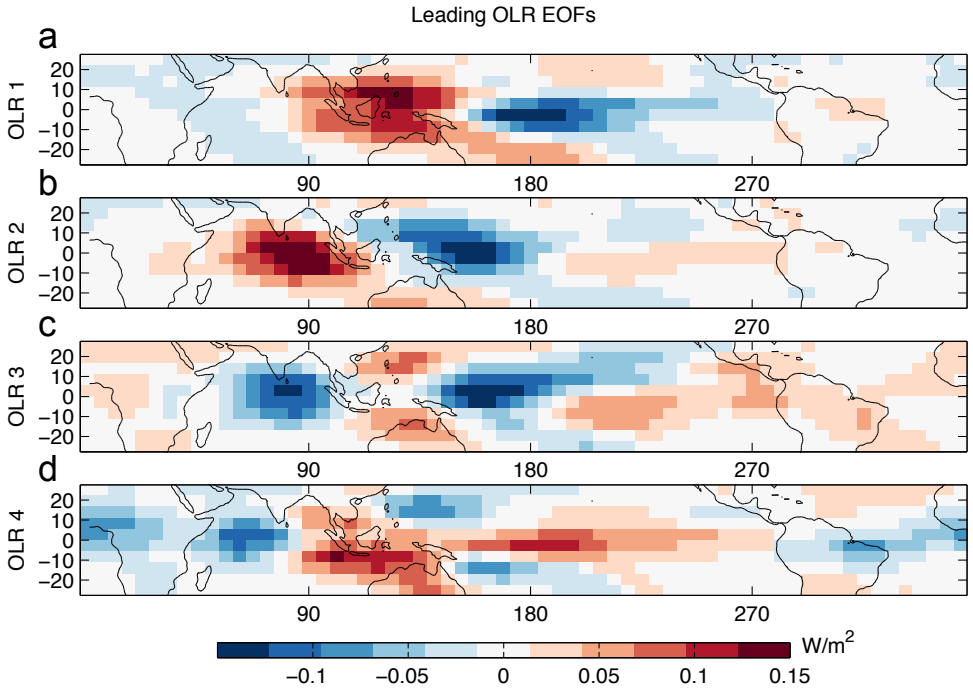

Figure 2.2: First four leading EOFs of tropical OLR. (a) represents 8.3% of the smoothed variance field. (b) 5.5%. (c) 3.6%. (d) 3.0 %.

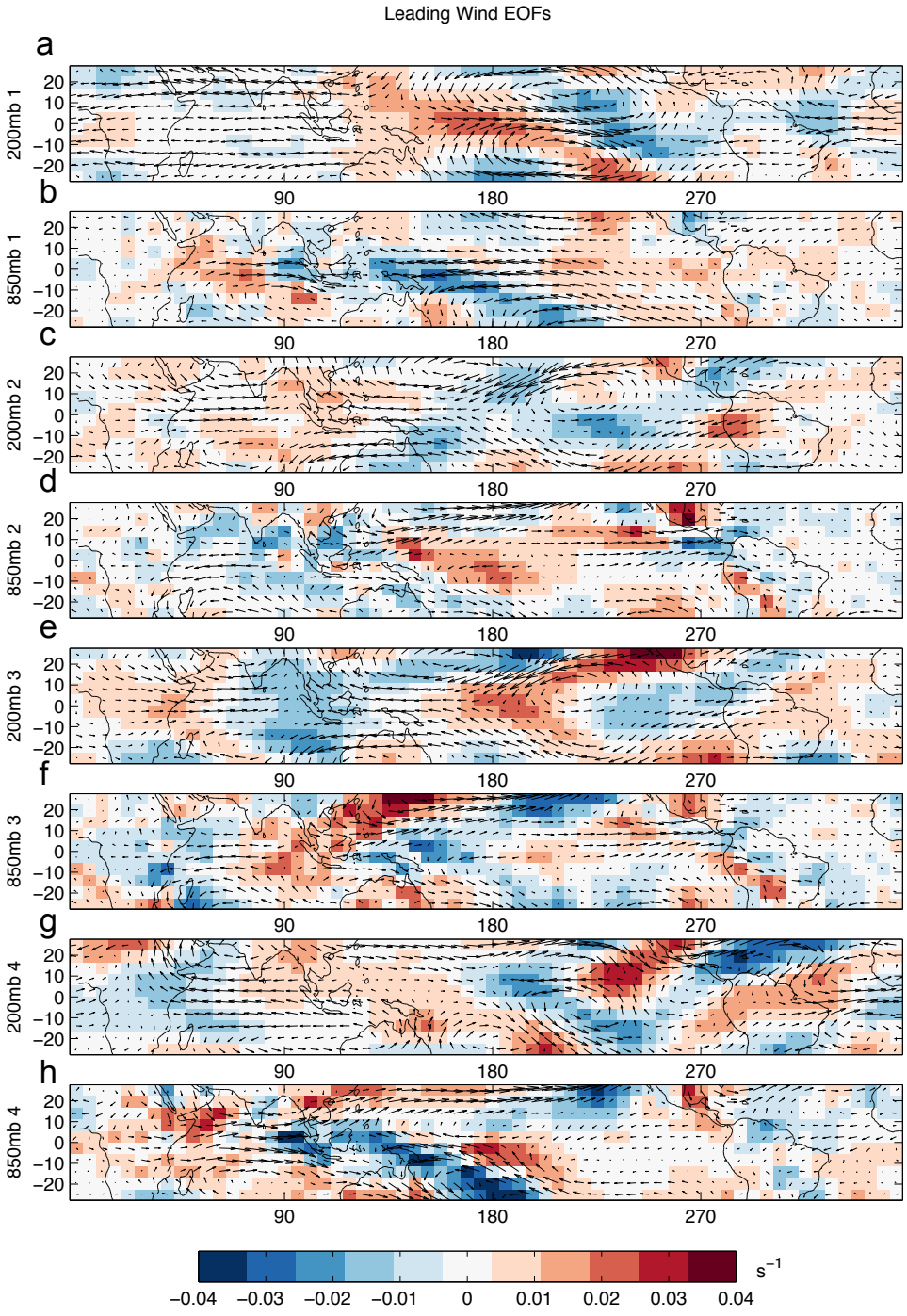

Figure 2.3: Same as Figure 2.2 for combined u200, u850, v200, and v850. Winds are plotted as vector fields, and magnitudes are shaded as 2-dimensional divergence fields. (a-b) represents 6.9% of the smoothed combined wind fields. (c-d) 4.2%. (e-f) 3.0%. (g-h) 2.4%.

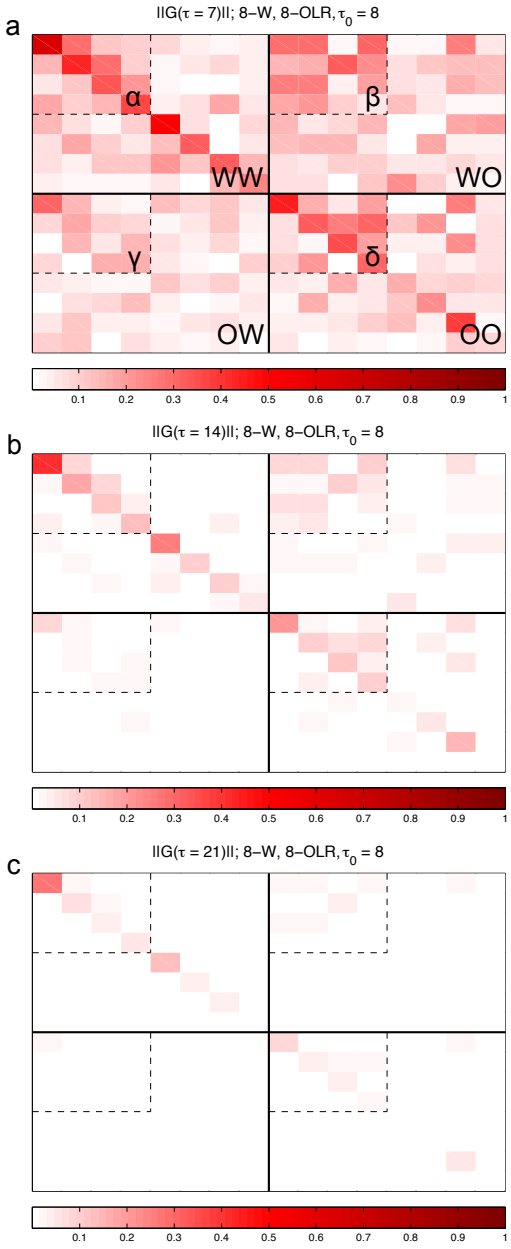

Figure 2.4: Magnitude of the modal interaction forecasting coefficients for the base model at (a) $\tau = 7$ days, (b) $\tau = 14$ days, and (c) $\tau = 21$ days. The **WW** submatrix summarizes the magnitude of wind pattern interactions, **OO** summarizes OLR interactions, with cross-interactions **OW** and **WO** in the bottom left and top right, respectively, delineated by solid lines. Submatrices $\alpha - \delta$, delineated by dashed lines, summarize interactions amongst the first four EOFs of each (combined) field and are only labeled in (a) for simplicity. Rows and columns are oriented as increasing from the upper left.

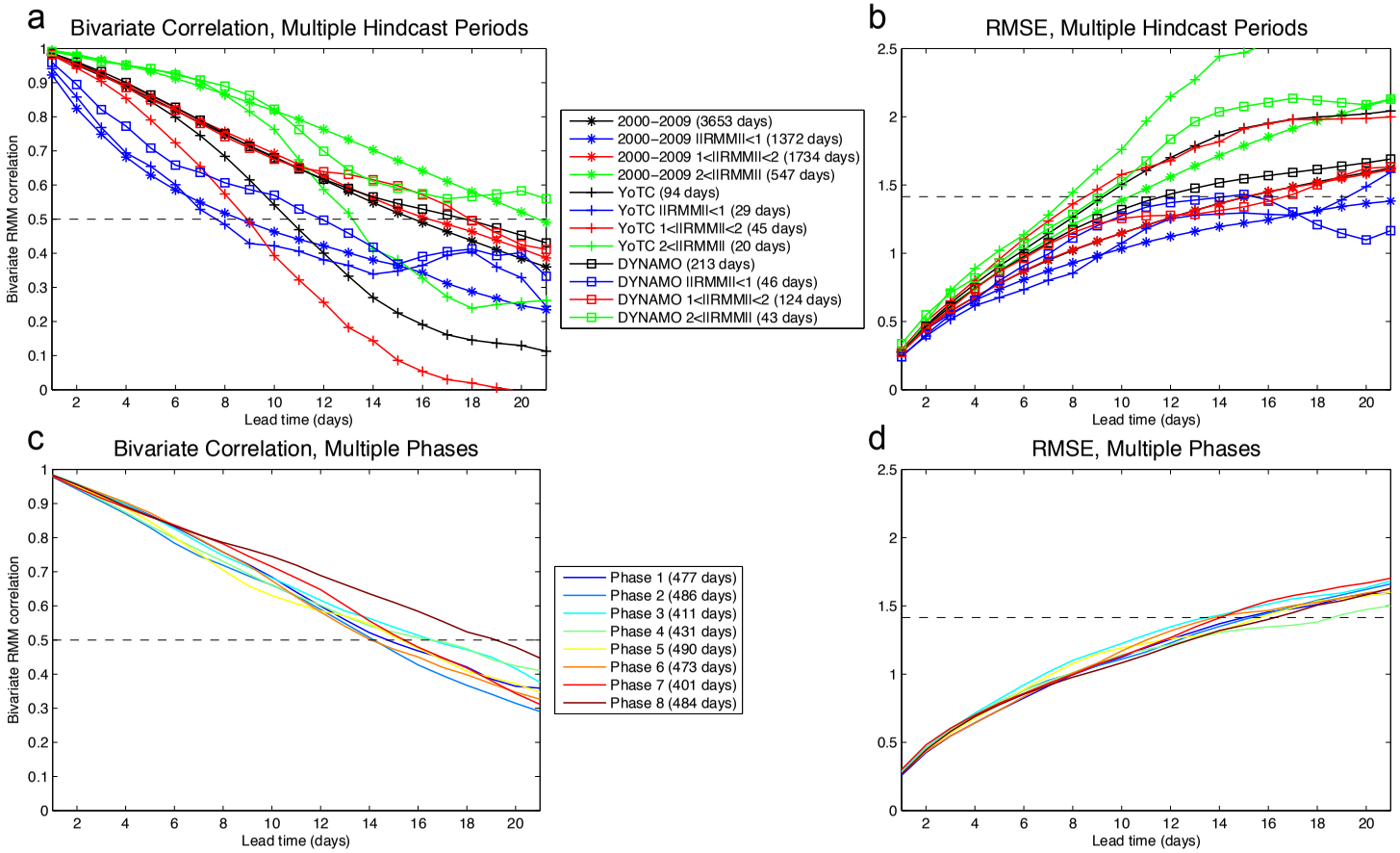

Figure 2.5: Bivariate correlation (left column) and RMSE (right column) for 24-**W** 32-**O** $\tau_0 = 6$ LIM hindcasts. Subplots a,b show MJO skill over multiple time periods, DYNAMO, YoTC, and 2000-2009, separated additionally by their RMM amplitude performance. Subplots c,d show hindcast skill over the 2000-2009 time period separated by MJO phase. The DYNAMO performance in (a) is directly comparable to *Zhang et al.* (2013a) Figure 2b. YoTC performance is directly comparison to forthcoming figures in *Klingaman et al.* (2014). The number of sample days is summarized in the figure legends.

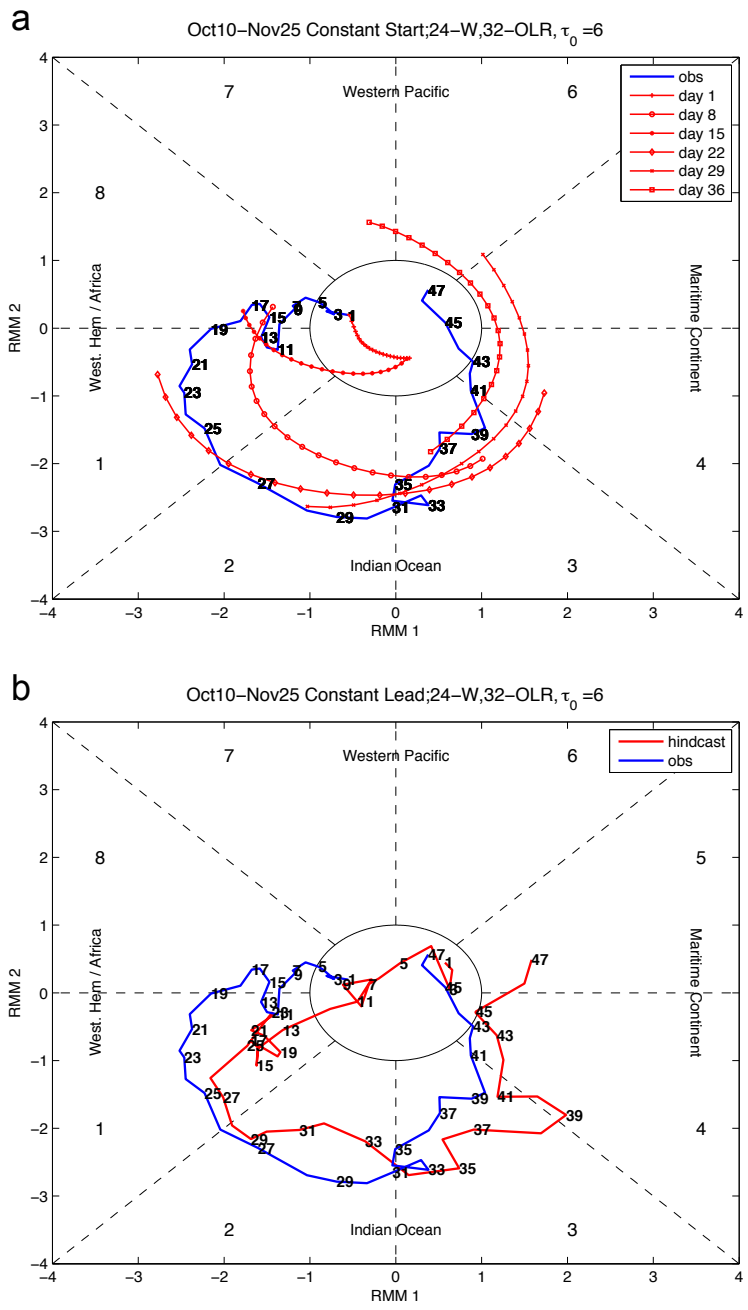

Figure 2.6: (a) Multiple 3-week hindcasts for the YoTC E MJO initialized at multiple days over the MJO event. (b) Constant one-week lead-time forecast of the YoTC E MJO. Models are initialized with the EOF truncated observation data at hindcast-day 0 and trained with EOF truncated observation data which does not include the year in which the hindcast was made. Numbers located near plotted lines indicate days into the forecast period.

Chapter 3

Northern Hemisphere Climatology and Trends of Statistical Moments Documented from GHCN-Daily Surface Air Temperature Station Data from 1950-2010

The first four statistical moments and their trends are calculated for the average daily surface air temperature (SAT) from 1950-2010 using the Global Historical Climatology Network Daily station data for each season relative to the 1961-1990 climatology over the northern hemisphere. Temporal variation of daily SATs probability distributions are represented as generalized linear regression coefficients on the mean, standard deviation, skewness, and kurtosis calculated for each 10-year moving time window from 1950-1959 to 2001-2010. The climatology and trends of these statistical moments suggest that daily SAT probability distributions are non-Gaussian and are changing in time. The climatology of the first four statistical moments has distinct spatial patterns with large coherent structure

for mean and standard deviation and relatively smaller and more regionalized patterns for skewness and kurtosis. The linear temporal trends from 1950-2010 of the first four moments also have coherent spatial patterns. The linear temporal trends in the characterizing statistical moments are statistically significant at most locations and have differing spatial patterns for different moments. The regionalized variations specific to higher moments may be related to the climate dynamics that contribute to extremes. The non-zero skewness and kurtosis make our detailed documentation on the higher statistical moments useful for quantifying climate changes and assessing climate model uncertainties.

3.1 Introduction

Weather related damage often occurs on the timescales of days or shorter. As such, understanding and quantifying climate risk requires consideration of daily variability as well as variability on longer timescales. For temperature, damages at (sub-) daily timescales often concern human health (e.g., heat stress; *Sherwood and Huber*, 2010) and agriculture (e.g., frost days; *Meehl et al.*, 2004). Historically, most climate variability studies focus on either timescales of months or longer (e.g., *Parker et al.*, 1994; *Michaels et al.*, 1998; *IPCC*, 2007; *Hansen et al.*, 2012), or on spatial averages over regions, such as individual countries or over an entire hemisphere (e.g., *Karl et al.*, 1995; *Gong and Ho*, 2004; *Shen et al.*, 2011). While these analyses deduce useful results that can aid in the understanding of global climate changes, finer space-time scale details for not only the mean but also higher statistical moments relevant to high-frequency geophysical dynamics and regional climate risk still require careful documentation as climate data and models improve. The purpose of this paper is to document the higher statistical moments, their seasonal climatology, their regional structure of temporal trends, and their associated non-Gaussian probability density functions (PDFs).

It is known that statistical moments can determine a PDF via a Taylor series expansion and Fourier transform (*Wackerly et al.*, 2008). *Perron and Sura* (2013) used the third and fourth statistical moments to capture the non-Gaussian nature

of daily atmospheric variables. They compiled a climatology of non-Gaussian atmospheric statistics from the NCEP / NCAR Reanalysis I dataset, which includes 925 mb temperature. *Petoukhov et al.* (2007) examined the ECMWF ERA40 reanalysis dataset, focusing on skewness and mixed third-order moments for a variety of atmospheric variables, also including temperature. *Shen et al.* (2011) examined changes in the higher-order moments of surface air temperature (SAT) of the contiguous United States from 1901-2000 and found decreasing second and third order moments and increasing fourth order moments. *Donat and Alexander* (2012) calculated the differences between the moment statistics (up to skewness) of two 30-year climatologies, 1951-1980 and 1981-2010, for the daily maximum and minimum temperatures in HadGHCN-D (Hadley Global Historical Climatology Network - Daily), a gridded dataset interpolated from the station data used here.

Quite often, approximate Gaussianity is assumed for simplicity in studies of geophysical systems (e.g., in independent EOF analysis for physical interpretations and the independent data and model errors in Kalman filter data assimilation). However, non-Gaussian variability in observations and its prediction from ensembles of nonlinear dynamical systems perturbed with Gaussian white noise (*Palmer*, 1993, and references therein) has been known for quite some time. Principally, non-Gaussian ensemble statistics necessarily arise from nonlinear deterministic dynamics or a non-Gaussian noise term when considering a stochastic forcing. Recent theory in empirical stochastic-dynamic modeling has adapted to capture non-Gaussianity through the form of multiplicative noise (*Berner*, 2005; *Sura et al.*, 2005) and has been shown to spatially approximate statistics from idealized atmospheric models (*Sardeshmukh and Sura*, 2009). Non-zero skewness measures the asymmetry of a PDF and is sufficient to demonstrate non-Gaussianity, and the non-Gaussian structure of climate data has been documented for several climate datasets as aforementioned.

Two major pitfalls are detrimental to the study of high-frequency probabilistic climate variability. First, temporal averages asymptotically approach Gaussian anomaly distributions due to the central limit theorem. For many variables and locations, monthly or longer time averages are approximately Gaussian

(*Stephenson et al.*, 2004). Second, spatial correlation scales vary globally, necessitating that close attention be paid to the aggregation of data over homogeneous regions. Averaging or aggregating data over large scales can sully the underlying statistics.

The current paper studies not only the first two statistical moments of SAT variability, but also the higher order moments, whose changes may imply deviations in the structure of the governing dynamical system. These changes of higher statistical moments, which impart PDF changes, can help diagnose dynamic variability and quantify variations of many applications-oriented indices including frost days, heating and cooling degree days, corn-heat-units, and heat stress index. Our work provides a comprehensive documentation of non-Gaussian structure for Northern Hemisphere SAT using the ground station observations of the Global Historical Climatology Network Daily (GHCN-D) from January 1, 1950 - December 31, 2010. Information on non-Gaussianity can help explain why certain EOF modes are not independent and/or have little physical meaning (*Monahan et al.*, 2009). The climatology may practically be used as benchmark distribution characteristics for validating model output mean, variance and extremes. They can also be used for a method of moments estimation for chosen distributions and to derive inferences about the probabilities of specific climatic extremes (e.g., *Rahmstorf and Coumou*, 2011). While this dataset suffers from incomplete spatial and temporal coverage, no a priori spatial or temporal averaging has been performed, which reduces the chance of systematic statistical bias and increases the chance of the results obtaining fidelity to truth. No attempt at assessing field significance about the spatial average of a large region is made since any demarcation would be inherently arbitrary without a more complete assessment of climatic regions based on daily PDFs or other high-frequency regime measures.

This paper is organized as follows. Section 3.2 discusses the observational dataset and its preprocessing steps, followed by a thorough explanation of further processing methodology. Section 3.3 describes data quality measures, statistical characterizations, regression methods and PDF approximations. Section 3.4 includes topical regional analysis of results derived directly from GHCN-D data for

both the climatology and trends. A discussion regarding the spatial patterns of climate change is supplemented with two examples to facilitate understanding and to demonstrate the use of these metrics for regional risk analysis. Section 3.5 concludes with a summary of findings and suggestions for future work.

3.2 GHCN-D Data and Preprocessing

GHCN-D contains in situ observational data from over 80,000 stations in 180 countries and territories worldwide (*Menne et al.*, 2012). Recorded variables commonly include daily maximum (T_{max}) and minimum temperature (T_{min}), total daily precipitation, snowfall, and snow-depth. In this study, T_{max} and T_{min} records from approximately 4,000 stations distributed across the Northern Hemisphere are analyzed. Each of these station records must pass both the GHCN-D consistency and quality assurance (QA) measures, outlined below, as well as record completeness measures from 1950-2010, outlined in Section 3.

The GHCN-D source data have been compiled with the goal of maximizing spatial coverage. Records from numerous agencies have been collected and quality controlled, but not homogenized. Before new data are added to the dataset, they are first cross-checked with included data to confirm record uniqueness, spatial consistency, and temporal consistency, as outlined by *Menne et al.* (2012), which reduces the chances of redundancy. Each record is then subject to 19 QA measures and flagged with an appropriate tag so that data can be utilized or discarded as necessary by a user (*Durre et al.*, 2010). Toward the goal of capturing trends in the moments of daily anomalies aggregated over many seasons, each flagged temperature datum is viewed as questionable and removed from the analysis. It is assumed that each seasonal daily distribution is not under-sampled over a 10-year period, obviating the need to retain questionable data for the sake of robust statistical moment estimation.

The daily mean SAT anomaly, denoted by T_{avg}, for this study is defined as the average of T_{max} and T_{min}. First, a running average of five days is performed on T_{max} and T_{min}, and seasonal cycles are computed from the 1961-1990

smoothed values, producing the 30-year daily T_{max} and T_{min} climatologies, which are then removed from the GHCN-D T_{max} and T_{min} data to produce T_{max} and T_{min} anomalies. Only stations that include at least 20 years of data (7,300 days) over the 1961-1990 period are included in this study. This anomaly calculation procedure follows *Alexander et al.* (2006) and *Caesar et al.* (2006) to facilitate intercomparison with other works. The station density per 3.75° longitude by 2.5° latitude grid cell is shown in Figure 3.1 for stations that are at least 75% complete (i.e., 15 years of data in each 20-year period).

3.3 Methodology

3.3.1 Record Completeness Measures

Before each station's T_{avg} anomaly record is analyzed, it is subject to record completeness checks to ensure robust statistical moment estimation for each 10-year period as well as for completeness over the 1950-2010 regression period. As a preliminary measure, each station record is required to span the entire regression period; in general, for a given station, data quality and completeness (inferred by number of missing days) improves nearer to present, so each record is required to include data for Jan 1, 1950 or before, but not explicitly required to include data for Dec 31, 2010 or after, since many stations lack the most recent values (likely due to unreporting) but are otherwise of high quality for the most recent decade. To ensure an adequate sampling of the regression period, each station record that is less than 75% complete is removed.

It is assumed here that SAT anomalies are approximately inter-seasonally piecewise stationary, but maintain inter-annual non-stationarity (following *Shen et al.*, 2011). The data from each station T_{avg} anomaly record is thus divided into four seasonal records, Dec-Jan-Feb (DJF), Mar-Apr-May (MAM), Jun-Jul-Aug (JJA), and Sep-Oct-Nov (SON). Each seasonal record is then subject to a running 10-year data-aggregating moving time window so that the first bin contains the 1950-1959 days, the second bin the 1951-1960 days, and so forth. Thus, we have used 52 bins, beginning with 1950-1959 and ending with 2001-2010. A 10-year

bin is selected to smooth the fluctuations of yearly anomaly statistics arising from the El Niño Southern Oscillation (ENSO; see *Stefanova et al.*, 2013), to yield an adequate sample size for robust estimation of summary statistics, to enable a robust trend detection over the relatively short regression period, and to help satisfy the statistical assumptions associated with the regression method.

The first four statistical moments for each 10-year data bin that contains at least a 75% complete record, for each season and each station are calculated. Out of these seasonal records, only records that contain summary statistics for 75% of the regression period (i.e., at least 39 out of the 52 binned 10-year periods) are included in the trend analysis. This regression period completeness measure is in addition to the completeness measure applied to the daily data. The 75% completeness measures were selected as an intermediate level of strictness and for simplicity. The results are insensitive to this choice with values from 50-90% tested.

3.3.2 Statistical Moments

Data in the binned 10-year periods are used to calculate the first four moments, as follows. The expectation of each 10-year seasonal distribution is characterized by the biased arithmetic sample mean

$$\hat{\mu}_j = \bar{X}_j = \frac{1}{\tau}\sum_{t=1}^{\tau} X_{j,t} \tag{3.1}$$

for the j^{th} station for time index t with non-missing $X_{j,t}$ contained within the 10-year distribution. The biased estimator, a simple average, is used since the sample size τ is over 600, large enough so that the bias is effectively reduced. In comparison, the sample size for monthly or yearly data may be small and the unbiased estimator may be more appropriate.

To assess changes in the second moment, the sample standard deviation

$$\hat{\sigma}_j = \left[\frac{1}{\tau}\sum_{t=1}^{\tau}(X_{j,t} - \hat{\mu}_j)^2\right]^{1/2} \tag{3.2}$$

for the j^{th} station for time index t contained within the 10-year distribution is used. The standard deviation is chosen so that the units of the first two moments are comparable.

The third and fourth moments are characterized by the sample skewness and sample (excess) kurtosis, respectively:

$$Skew(X) = \frac{1}{\tau}\sum_{t=1}^{\tau}\frac{(X_{j,t} - \hat{\mu}_j)^3}{\hat{\sigma}_j^3} \tag{3.3}$$

$$Kurt(X) = \frac{1}{\tau}\sum_{t=1}^{\tau}\frac{(X_{j,t} - \hat{\mu}_j)^4}{\hat{\sigma}_j^4} - 3 \tag{3.4}$$

For a Gaussian distribution, both the skewness and excess kurtosis are zero. It is customary to leave out excess in discussions of kurtosis (as will be done in this paper).

Skewness and kurtosis are considered descriptors of extreme probabilistic behavior in that they help define the tail behavior of a PDF. Skewness is a measure of the degree of asymmetry in a distribution. Positive (negative) skewness indicates the presence of a long right (left) tail; for a unimodal distribution this manifests itself as a more positive (negative) mean than mode, and generally indicates that equally probable high-valued (low-valued) events are more extreme than low-valued (high-valued) events. Conversely, this implies that equally extreme high-valued (low-valued) events are more (less) probable than their low-valued (high-valued) counterparts.

Kurtosis is a measure of the peakedness of a distribution relative to a Gaussian distribution. Simultaneously, kurtosis is also a measure of the tail thickness relative to a Gaussian distribution, since the area under a PDF must integrate to unity. Geometrically, distributions with positive kurtosis (and zero skewness in this case) will show greater density around the mean and tails than a Gaussian and are commonly considered fat-tailed, indicating higher probabilities for very large anomalies.

3.3.3 Generalized Linear Regression Model

For dynamic systems, non-zero autocorrelation structure leads to a reduced effective number of observations. In this case, mid-latitude atmospheric degrees of freedom can be considered largely controlled by the propagation of atmospheric Rossby waves (*Perron and Sura*, 2013). Oceanic variables and climate oscillations

are more complex with decorrelation times on the order of years to decades. The methodology described above introduces another form of autocorrelation in that each binned moment observation is a function of each seasonal-daily observation of the previous 10 years. For the mean, this autocorrelation is analogous to that induced by a low-pass running mean filter. Within the context of the simple linear regression model, this leads to over-confident estimates of the slope, β (*Wilks*, 2011). Common methods to estimate confidence intervals in autocorrelated time-series trend analysis involve the estimation of a decorrelation time and subsequent adjustment of confidence intervals, bootstrap sampling methods, non-parametrics, and/or assumptions of a particular autocorrelation structure such as an autoregressive (AR) model. In this analysis, moments are assumed to have an AR(1) structure of the form

$$M_{j,t+1} = \alpha_j + \beta_j t + u_{j,t} \tag{3.5a}$$

$$u_{j,t} = \rho_j u_{j,t-1} + \epsilon_{j,t} \tag{3.5b}$$

$$\epsilon_{j,t} \sim N(0, \sigma_j^2 \tag{3.5c}$$

for the seasonal moment M at station j for year t, with M-intercept α, slope β, error term u, AR(1) coefficient ρ, and random normal noise ϵ of zero mean and σ standard deviation. Further details are discussed by *Box et al.* (1994).

3.4 Results

3.4.1 Seasonal Climatology of Non-Gaussian T_{avg} Statistics

The left panels of Figures 3.2-3.5 show each station's 1961-1990 moment climatology for MAM, JJA, SON, and DJF, respectively. The mean is presented as the seasonal mean over the climatology period, and the higher moments are calculated from the distribution of daily anomalies. Each moment subplot from each season can be analyzed individually to assess the effect of regional geography on that particular distribution characteristic. Seasonal dependence of regional distribution characteristics can be assessed by comparing the appropriate subplots from each season, and the four moment subplots from each individual season can

be taken in conjunction to draw conclusions about the behavior of T_{avg} variability relative to regional geography. Grid cells corresponding to 3.75° longitude × 2.5° latitude bins on skewness and kurtosis plots are hatched if at least 50% of the stations contained within that particular grid cell are significantly different from 0 at a 5% significance level, based on the standard errors of skewness (eq. 3.6a) and kurtosis (eq. 3.6b) (*Brooks and Carruthers*, 1953),

$$SE(skew) \approx \sqrt{6/N_{eff}} \tag{3.6a}$$

$$SE(kurt) \approx \sqrt{24/N_{eff}} \tag{3.6b}$$

where N_{eff} is the effective number of sampled days. *Perron and Sura* (2013) show that these standard errors are approximately valid for the weakly non-Gaussian distributions observed in the atmosphere. We assume here that $N_{eff} = N/7$ based on the reasoning put forth by *Perron and Sura* (2013), that mid-latitude Rossby waves traverse the globe in approximately one week. For all stations, we take N equal to 20 years worth of seasonal data, the minimum to be considered in the climatology.

The most profound (and possibly most subtle) observation is that for each moment that characterizes an aspect of probability distributions in each individual season, the spatial structure is smooth at the synoptic scale, but the emergent patterns over the Northern Hemisphere in each of the moments are different. The implications of this are twofold. First, for any particular region the contained stations' anomaly PDFs are the same or similar regardless of local elevation or other geography. This is another justification for using anomalies to assess climate changes. Second, each climatological statistical moment has its own coherent spatial pattern and the direction and magnitude of the trend of each moment also has a coherent spatial pattern. These coherent spatial patterns provide the potential for skill in both spatial and temporal interpolation (and perhaps temporal extrapolation). Moment interpolation, as opposed to time series interpolation, has the benefit of preserving probability distribution characteristics and hence extremes. Comparing the patterns in moments contained within each season, the negative linear relationship between mean and standard deviation noted earlier by *Robeson*

(2002) is evident (see Fig. 3.6). Cloud cover, cloud type, humidity, and wind speed are key controllers of SAT, and their variability in time is impacted by the variability of synoptic conditions. *Robeson* (2002) suggests generally that regions with higher average temperatures are less likely to experience highly variable synoptic scale conditions, for example the frequent passage of cold fronts, which lead to increased standard deviations in SAT. While a parabolic relationship between skewness and kurtosis is expected for a diffusive system, a quite precise inequality for a random variable governed by a Fokker-Planck equation forced by multiplicative noise is $kurt \geq (3/2) \times skew^2 + constant$ (*Sura and Sardeshmukh*, 2008), which is also valid for our data as shown in Figure 3.7. This suggests that T_{avg} might also be modeled by such a stochastic system. The offset by a small negative constant is discussed in detail by *Sura and Perron* (2010). The dynamical generator of these higher order moments in the climate system is thought to be local adiabatic turbulence (*Sardeshmukh and Sura*, 2009). These relationships do not hold over coastal regions where near-shore stations have a strong marine influence.

Unsurprisingly, climatological means of the raw field (Figs. 3.2-3.5a), as opposed to the anomaly, exhibit a strong meridional gradient where higher (lower) latitudes are colder (warmer) which can be explained simply by top-of-atmosphere (TOA) insolation; deviations from the TOA insolation pattern are dynamically induced, and mirror to a large extent the climatological jet stream path. Winter months produce colder means nearer to the pole, while the tropics remain relatively constant year-round. Regions with a more maritime influence (e.g., Western Europe), whose temperature fluctuations are strongly damped, exhibit much lower standard deviations relative to other areas within the same zone. MAM, SON, and DJF (Figs. 3.2c, 3.4-3.5c) show a similar standard deviation pattern with slightly elevated values above 45°N during DJF. Seasonal cycles of standard deviations are evident over the entire NH with maxima in DJF (Fig. 3.5c) and minima in JJA months (Fig. 3.3c), which is consistent with our understanding of the mean and standard deviation relationship.

Although the large spatial patterns of the mean and standard deviation in panels a and c in Figures 3.2-3.5 are not surprising, the smaller spatial scales of

skewness and kurtosis seem striking (panels f and h). Skewness is phase shifted from the locations of semi-permanent low-level high and low pressure systems, as originally proposed by *Loikith and Broccoli* (2012), suggesting that large-scale circulation patterns may play a role in modulating the likelihood or magnitude of asymmetrical temperature extremes. For example, during DJF (Fig. 3.5e), the areas east (west) of the Canadian and Siberian Highs show positive (negative) skewness, and areas east (west) of the Aleutian and Icelandic Lows show negative (positive) skewness. These results are largely consistent with those of *Perron and Sura* (2013). In other areas and at different times, more local effects appear to be at play. These effects include modulation of extreme distribution characteristics by large bodies of water, and perhaps in some places, atmospheric conditions at elevation. Coastal areas, particularly near the west and north sides of continents, exhibit positive skewness which becomes both more prominent and extends further poleward during the warmer months. Skewness is most prominently negative along the Continental Divide in the Americas during cold seasons (Fig. 3.5e). Similarly, kurtosis values are most prominently positive in these areas at the same times. It is unknown exactly how the synoptic-scale wave-like patterns of skewness are dynamically linked to nonlinear atmospheric dynamics (i.e., Chapters 6-7, *Holton*, 2004). Northern Hemisphere kurtosis reaches a maximum in MAM and SON months (Figs. 3.2g and 3.4g) and is predominantly positive below 30°N year-round. Soil moisture levels regulate surface air temperature extremes (*Durre et al.*, 2000, and references therein), and seasonal precipitation (see the GPCP monthly climatology: http://jisao.washington.edu/data/gpcp/#climatology) qualitatively appears to play an important role in the extremes and contributes to skewness and kurtosis values (particularly in the American North West and Southeast Asia), but the exact relationship is unclear and left to future research.

3.4.2 1950-2010 Temporal Trends of Statistical Moments

The right panels of Figures 3.2-3.5 show trends of the first four moments for each season in a 61-year period from 1950-2010. Each station's trend is plotted individually. Grid cells corresponding to 3.75° longitude × 2.5° latitude bins are

hatched if at least 50% of the stations contained within that particular grid cell are significant at a 5% significance level. At this p value, hatched regions will be termed "robust."

Positive trends in the mean for the 61-year regression period are widespread over the northern hemisphere, with many regions exhibiting coherent robust trends, particularly over Eurasia during MAM (Fig. 3.2b). One notable exception is the so called warming hole located over the central southwest United States (*Portmann et al.*, 2009) which shows robust weak cooling and is observable predominantly during JJA and DJF (Figs. 3.3b, 3.5b). Warming trends are strongest over Eurasia during MAM and DJF at about 0.4°C/decade (Figs. 3.2b, 3.5b). During MAM (Fig. 3.2b), there is also a weak cooling trend over northern Mexico.

Trends in standard deviation are seasonally variable, although broadly, most of the robust and regionally coherent trends tend to be negative over the northern hemisphere. During MAM (Fig. 3.2d), the dominant robust features include positive trends over the Iberian Peninsula and negative trends over the southeast United States, the Pacific Northwest regions of the United States and Canada, Europe, western Russia, Central Asia, and non-coastal Siberia. During JJA (Fig. 3.3d), robust trends include positive trends over Western Europe and negative trends over the United States (except in the NW), western Canada, Alaska, and inland Siberia. During SON (Fig. 3.4d), there are no coherent, robust, regional, positive trends in standard deviation, but negative robust trends are observed over the contiguous United States (except the Rocky Mountains region), Canada, Alaska, Great Britain, Eastern Europe, and Central Asia. During DJF (Fig. 3.5d), robust positive trends are over the Ural Mountains and robust negative trends are over most of the contiguous United States (except the mid-west region), Canada, Europe, Siberia, and Central Asia.

During MAM (Fig. 3.2f), positive robust trends of skewness are observed over the Ural Mountains and Central Asia, with negative robust trends over the Iberian Peninsula, Eastern Europe, much of Siberia, and Mexico. During JJA (Fig. 3.3f), robust positive trends exist over parts of Eastern Europe, with widespread robust negative trends over Alaska, central Canada, the Pacific Northwest, Ap-

palachia, Western Europe, the Ural Mountains, Central Asia, and most of Siberia. During SON (Fig. 3.4f), robust trends include localized positive trends over much of Russia and negative trends over the west coast of the United States and Canada and in eastern Siberia. During DJF (Fig. 3.5f), positive robust trends are observed over Western Europe, eastern Siberia, and the Ural Mountains, with negative robust trends over Central Asia, central Siberia, Northern Canada, Alaska, California, and Mexico. It remains to be investigated whether these trend patterns can be linked to changing climate dynamics.

Robust trends for kurtosis over each season also tend to be predominantly negative. During MAM (Fig. 3.2h), negative robust trends are observed over the western United States and Canada, Alaska, the Iberian Peninsula, Eastern Europe, Russia and Siberia, and Central Asia. During JJA (Fig. 3.3h) negative trend regions include central Canada, the United States (except for the west coast), Western Europe, the Balkans, Central Asia, the Ural Mountains, and eastern Siberia. During SON (Fig. 3.4h), negative trends are observed over the United States (except the Midwest), Eastern Europe, Russia and Siberia, and Central Asia. There are no robust positive trend regions over MAM, JJA, or SON. During DJF (Fig. 3.5h), robust positive trend regions are observed over central Canada and Siberia, with negative trends over the United States (except the Midwest), mainland Europe, the Ural Mountains, and western Russia.

3.4.3 Discussion of the Results

As shown previously in Sub-section 3.4.1, the climatological moments, and thus the PDFs that they describe, are spatially coherent and smoothly transition at synoptic scales (left columns of Figs. 3.2-3.5). This suggests that after removing a seasonal cycle, the underlying anomaly PDFs of individual GHCN-D stations over large areas are remarkably similar. The spatial patterns of each moment for each season, however, are different. There is a markedly negative linear relationship between the mean and standard deviation and a quadratic (parabolic) relationship between skewness and kurtosis, as shown in Figs. 3.6 and 3.7. This observed skewness-kurtosis relationship is predicted through moment closure in stochastic-

dynamic theory (*Sura*, 2011) and suggests that surface temperature may be appropriately modeled with a linear stochastic-dynamic system with correlated additive and multiplicative noise.

The climatological mean and standard deviation exhibit a strong meridional gradient, which has been widely discussed in existing literature. Contour lines of 250mb geopotential height (Z250) are shown in Figs. 3.2-3.5a and c, and appear to be conspicuously aligned with zonal waves of mean and standard deviation. These zonal waves are particularly prominent over DJF (Figs. 3.5a,c) and are coincident with the locations of the greatest zonal circulation gradients. Northern hemisphere extratropical circulation gradients are smaller and exhibit weaker wave formation during JJA (Figs. 3.3a,c), resulting in reduced zonal fluctuations in mean and standard deviation during these months.

Recent studies have investigated the linkage between Z500 and surface temperature extremes and suggest that persistent Z500 anomalies project strongly (*Meehl et al.*, 2004) and, particularly in DJF, linearly (*Loikith and Broccoli*, 2012) on SAT extremes. *Loikith and Broccoli* (2012) also suggests that regional variations in skewness may point to regions where temperature extremes are dominated by local or regional circulation dynamics. During DJF, when semi-permanent high and low pressure systems are strongest, geographical skew patterns are coherent on synoptic scales suggesting that DJF extremes are dominated by large-scale dynamic regimes, consistent with *Loikith and Broccoli* (2012). During JJA, geographical skew patterns are much more regional and resemble climatological relative humidity (*Dai*, 2006) suggesting that JJA temperature extremes are dominated by local conditions such as soil moisture content (e.g., *Hirschi et al.*, 2010), also consistent with *Loikith and Broccoli* (2012). During transition months, it is clear that there is a mixture of both synoptic and regional impacts on T_{avg}. Patterns in kurtosis may result by the linear mechanism suggested by *Loikith and Broccoli* (2012) from distributions of Z500 whose DJF kurtosis climatologies (*Perron and Sura*, 2013) resemble the kurtosis climatologies here or may also result from local nonlinear interactions, exemplified by the strong positive kurtosis region in the NW United States during DJF.

We additionally investigate the trends of these moments over a 1950-2010 regression period against an AR(1) model. Trends in the mean show familiar global warming trend patterns for each season that are compatible with estimates of recent climate change (*IPCC*, 2007). Warming trends during MAM are the most suggestive, particular over Eurasia. Robust trends in the standard deviation are predominantly negative globally, which is also consistent with recent research (*Karl et al.*, 1995; *Michaels et al.*, 1998; *Shen et al.*, 2011). Although standard deviation trends during MAM and JJA are more positive overall relative to the other seasons and show spatially coherent positive trends over large areas, the positive standard deviation trends tend not to be robust at grid-scale, while the negative trends are. Trends in skewness and kurtosis show robustness at grid-scale and exhibit spatial coherence as well. Over the United States, trends in skewness are predominately positive while kurtosis trends are predominantly negative, which is in agreement with the literature (*Shen et al.*, 2011). Here we include regional details, which make the results more useful as benchmark data for climate change assessment via other means, such as remote sensing and mathematical modeling. There is no other comparable literature for other regions.

Karl and Katz (2012) briefly discussed how sampling Gaussian distributed variability over non-stationary time periods when means and standard deviations have changed can in fact lead to non-zero, biased estimates of skewness. If such non-stationary time series have nonlinear trends in the first two moments, it could also lead to biases in the estimates of higher moment trends as well. We investigated this possibility briefly through Monte Carlo simulation (not shown). Biases in skewness and kurtosis estimates arising from linear trends in the mean and standard deviation of the order of magnitude of those observed in the atmosphere are at least an order of magnitude smaller than observed T_{avg} skewness and kurtosis. Additionally, nonlinear trends in the mean and standard deviation lead to trends in skewness and kurtosis which are also at least an order of magnitude smaller than those observed for all reasonable mean and standard deviation temporal trends.

No attempt to quantitatively assess trend significance at larger than grid scale is made here, since it is unclear which stations can be appropriately considered

homogeneous representations of a climate (trend) region. For example, a previous study by *Shen et al.* (2011) estimated higher moment trends over the contiguous United States; however it is apparent from this station-scale analysis that both the climatological PDFs and moment trends for stations within the United States are not a homogeneous sample. It is also clear, though, that taken individually there is strong regionalization for trends in each moment (whose regions may or may not coincide with the trend regions for other moments), suggesting that the underlying dynamic mechanisms responsible for these PDF characteristics also have changes on regional scales, globally. One exceptional note is that, in general, during winter months, skewness trends tend to oppose their climatological values which is perhaps an indication of weakening or shifting climatological high- and low-pressure systems which manifest themselves in skewness statistics.

Information regarding entire distributions, both the bulk of observations and the extremes, is critical for understanding and adapting to regional climate change, and high-frequency variability (e.g., daily statistics) driven by low-frequency climate modes (e.g., ENSO phase) is increasingly in the public interest. Recent climate extremes, such as the Russian heat wave/European blocking event of 2010, has piqued interest in regional climate extremes and global teleconnections (e.g., *Trenberth and Fasullo*, 2012; *Rahmstorf and Coumou*, 2011), as well as the role of natural variability in the context of high frequency extreme events (e.g., *Dole et al.*, 2011; *Matsueda*, 2011). In Figure 3.8, we choose weather stations located near Moscow, Russia during JJA to illustrate how changes in moment statistics are manifested through shifts in entire PDFs. Figure 3.8 shows that the PDFs of T_{avg} from each station located within the grid cell for the climatological period are nearly identical, as suggested by the smoothness of climatological moments in space. During JJA, the climatological anomaly PDFs are nearly Gaussian, as indicated by skewness and kurtosis near zero (Figs. 3.3e, g). Over the 1950-2010 regression period, this region shows positive trends in mean and skewness (Figs. 3.3b, f), which are indicated in Fig. 3.8 as a positive shift of the bulk of the distribution as well as a lengthening and thickening of the positive anomaly tail, which is particularly noticeable at the extremes. The cumulative probability den-

sity of large positive anomalies over JJA have noticeably increased each decade for the last 4 decades, indicating that the probability of large positive anomalies has increased, particularly during the 2001-2010 period.

Similarly, recent reduction in water levels in the Colorado River basin have prompted concern regarding recorded and projected climate change over the western United States. Studies have shown that while business-as-usual climate projections show slightly decreased precipitation levels (*Christensen et al.*, 2004), temperature changes and the early onset of spring conditions, which result in an increased rain-snow ratio and contribute to early season snow melt, have the largest impact on the basin (*Cayan et al.*, 2001; *Stewart et al.*, 2004). Examining the climatological PDFs of stations in the region (Fig. 3.9) again reveals similar PDFs for each of the stations contained in the same grid cell. The Colorado stations have a slightly elongated negative anomaly tail, which is manifested in the climatologial statistics as negative skewness over the region (Fig. 3.3e). Over the 1950-2010 regression period, the Colorado region exhibits a non-robust warming trend, with a robust negative trend in kurtosis. Neither trend is particularly distinguishable in decadal PDFs (Fig. 3.9) as shifts in kurtosis are subtle and difficult to discern; however, the most recent decadal PDF does display a distinct linear, positive shift, and PDFs for the last two decades have become less peaked, which is consistent with a negative shift in kurtosis.

3.5 Conclusion

We present non-Gaussian 1961-1990 climatology statistics for GHCN-D station T_{avg} data in the Northern Hemisphere for each season, as well as trends in these statistics over a 1950-2010 regression period. The relevant climatology results for gridded data mentioned in the introduction resemble those derived from station data in this paper. Non-Gaussian probability functions of SAT reflect strongly on the probabilities of both high and low temperature extremes (*Ruff and Neelin*, 2012), and proper assessment of climate distribution tail behavior is critical for proper model and reanalysis verification and climate risk assessment under cli-

mate change. Climatological moments show coherence on regional to synoptic scale and are seasonally dependent, suggesting that anomaly PDFs are quite similar at regional scales but exhibit a strong seasonal dependence. Mean and standard deviation show synoptic coherence and a strong meridional gradient and are modulated zonally by circulation patterns, while skewness and kurtosis fluctuate on smaller scales, implicating regional dynamics in the modulation of temperature extremes, particularly for summer months. Trends in moments show large regionalization with robust changes at grid-scale for all moments and all seasons. For mean and standard deviation, trends are coherent on synoptic scales and show robust warming in most areas during each season and reduced variability in most areas during most seasons. Trends in skewness and kurtosis vary on smaller regional scales and suggest changes to the regional dynamics that modulate higher order statistics at those scales. We further present as case studies the climatological and decadal PDFs for the regions around Moscow, Russia and Colorado, USA to illustrate the homogeneity of regional station PDFs and how shifts in moments are represented in entire PDFs over the regression period.

Further works on the mechanisms of these shifts are needed to gain a physical understanding of the (changing) processes that manifest themselves in the (changing) statistical moments. If explicit dynamical links can be made to climatological moments, either through stochastic dynamics (e.g., multiplicative noise discussed previously) or through numerical models (e.g., the specific contribution of extreme weather events to variability statistics), trends in these moments might provide further insight regarding local, regional, and global climate change. Additionally, a greater understanding of the spatial correlation structure of station time series is needed so that biases in the moments arising from averaging can be anticipated and potentially corrected or avoided. Spatial averaging and/or interpolation results in reduced standard deviations, as well as skewness and kurtosis values closer to zero, as is expected by the central limit theorem. Finally, further work on assigning regions of climate change with respect to daily variability (e.g., through cluster analysis; *Loikith et al.*, 2013) is needed so that spatial correlation in moment trends can be appropriately accounted for to provide regional estimates

of climate change statistical significance.

Our moment method can be used to assess climate changes, diagnose climate dynamics, and validate models in terms of probability distributions. Toward the specific ends of examining extremes in observational data, *Donat et al.* (2013a) have compiled the GHCN-DEX, a global database of extremes indices compiled from GHCN-D. As for daily-observed data, we are also aware of the existence of the gridded version of GHCN-D data: HadGHCN-D. A detailed assessment of the differences between the station GHCN-D and gridded GHCN-D will be made in another study, which will particularly facilitate climate model validations when regarding climate model output at a grid point as a grid box average.

3.6 Acknowledgements

This study was supported in part by the US National Science Foundation (Award AGS-1015926). Shen was also supported by the US National Oceanographic and Atmospheric Administration (Award EL133E09SE4048), US National Science Foundation (Award AGS-1015957), and US Department of Energy (Award DE-SC002763). We greatly appreciate the efforts of two anonymous reviewers whose comments greatly improved the clarity of this paper. Cavanaugh would also like to acknowledge funding received from the Scripps Institution of Oceanography and to thank Alexander Gershunov for his thoughtful contributions to the paper. This chapter has been published in: **Cavanaugh, N.R.**, and Shen, S.S.P. (2014). Northern Hemisphere Climatology and Trends of Statistical Moments Documented from GHCN-Daily Surface Air Temperature Station Data from 1950 to 2010. *Journal of Climate*, 27, 5396-5410.

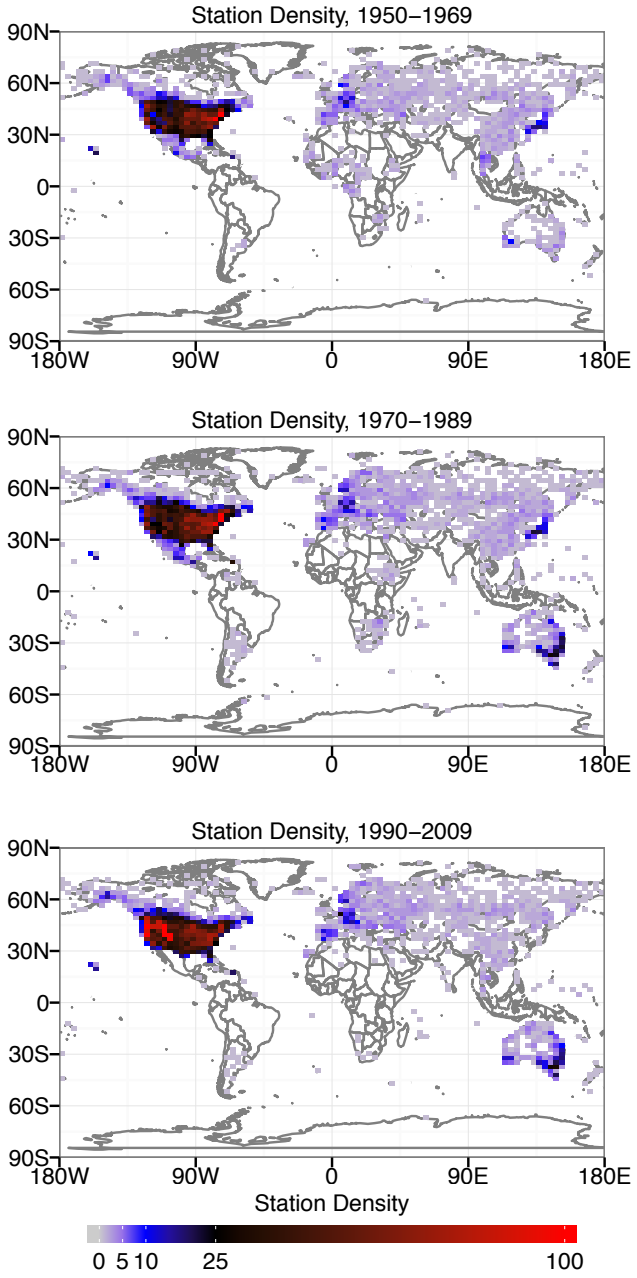

Figure 3.1: GHCN-D station density (number of stations/grid cell) over the regression period, separated into 20-year intervals. Stations accounted for are those that are at least 75% complete over the 20-year interval.

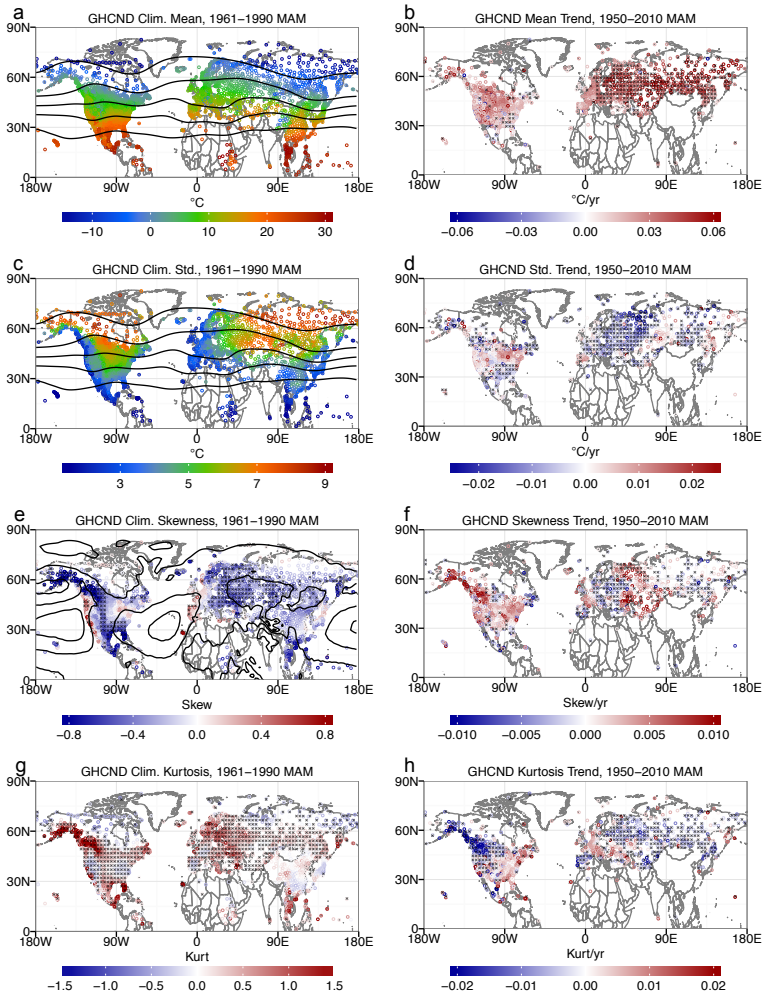

Figure 3.2: Climatological moment statistics (left column) and trends over the 1950-2010 regression period (right column) for MAM. The mean (a) is calculated from the seasonal climatology, whereby the variability statistics (c,e,f) are calculated from the anomalies. Individual subplots for each season are plotted on the same color scale to facilitate seasonal inter-comparison. Skewness and kurtosis grid cells are hatched if at least 50% of the contained stations are significantly different from 0 at a 5% significance level using standard error formulas (eqs. 3.6a,b). Trend grid cells are hatched if at least 50% of the contained stations show a significant trend at 5% one-tailed significance level, taken here to be robust. Contours on a and c show climatological 250mb geopotential height with 200 meter spacing; contours on e show climatological sea level pressure with 5mb spacing, both from the NCEP-NCAR reanalysis I (*Kalnay et al.*, 1996). Extremes and outliers are truncated slightly to enhance interpretability.

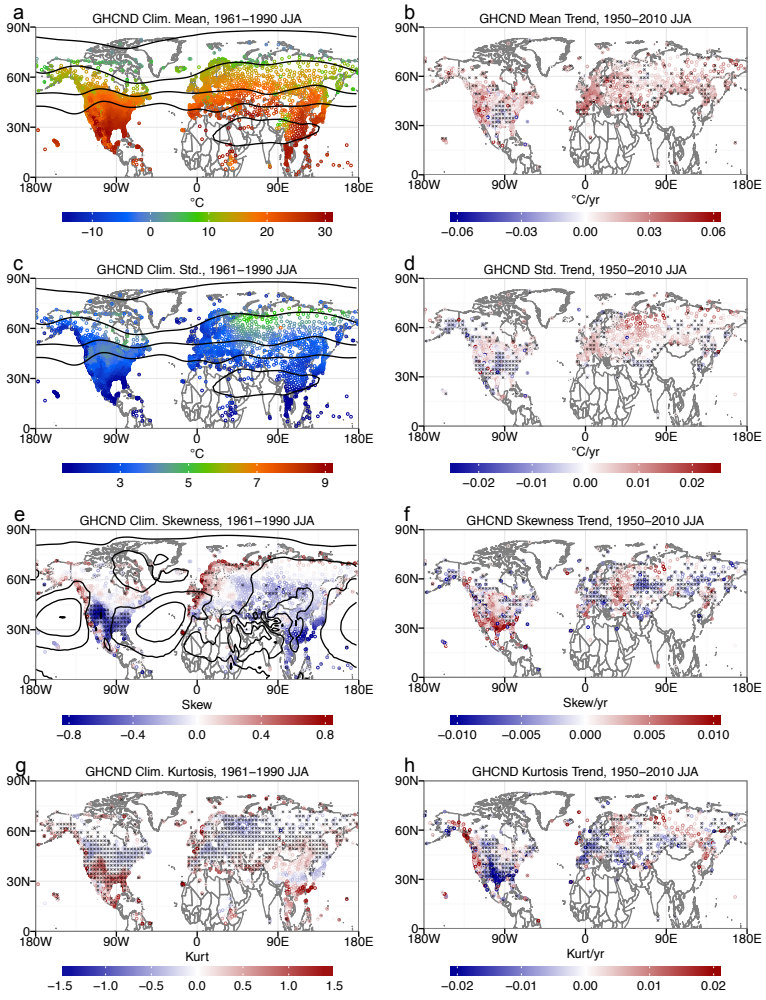

Figure 3.3: Same as Figure 3.2 for JJA.

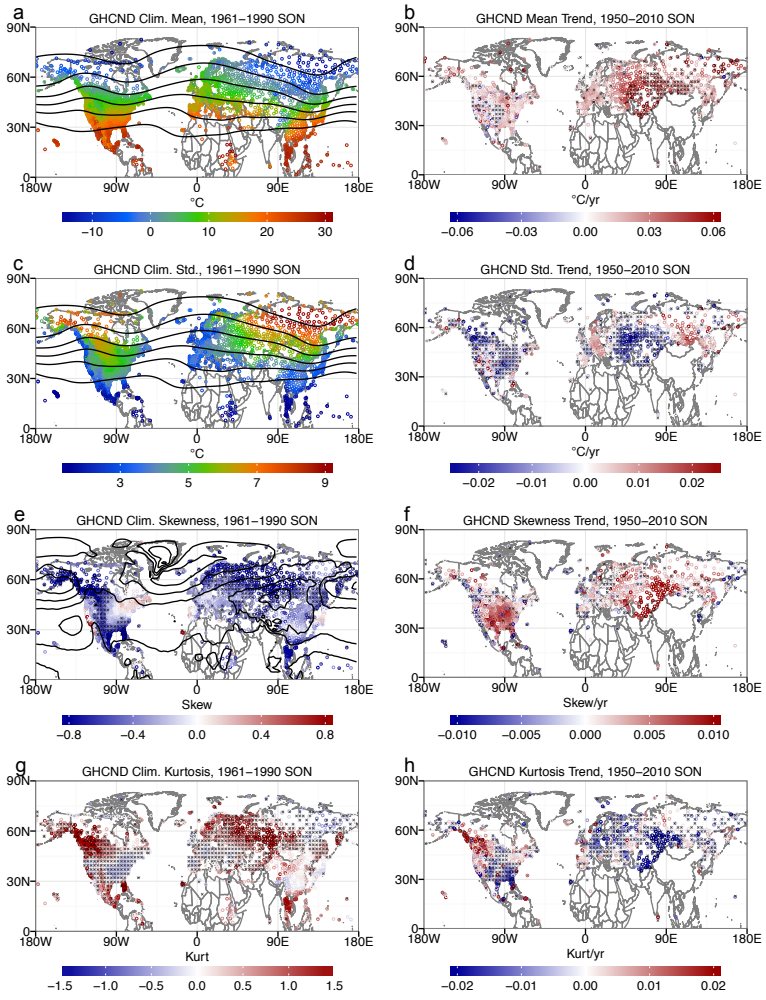

Figure 3.4: Same as Figure 3.2 for SON.

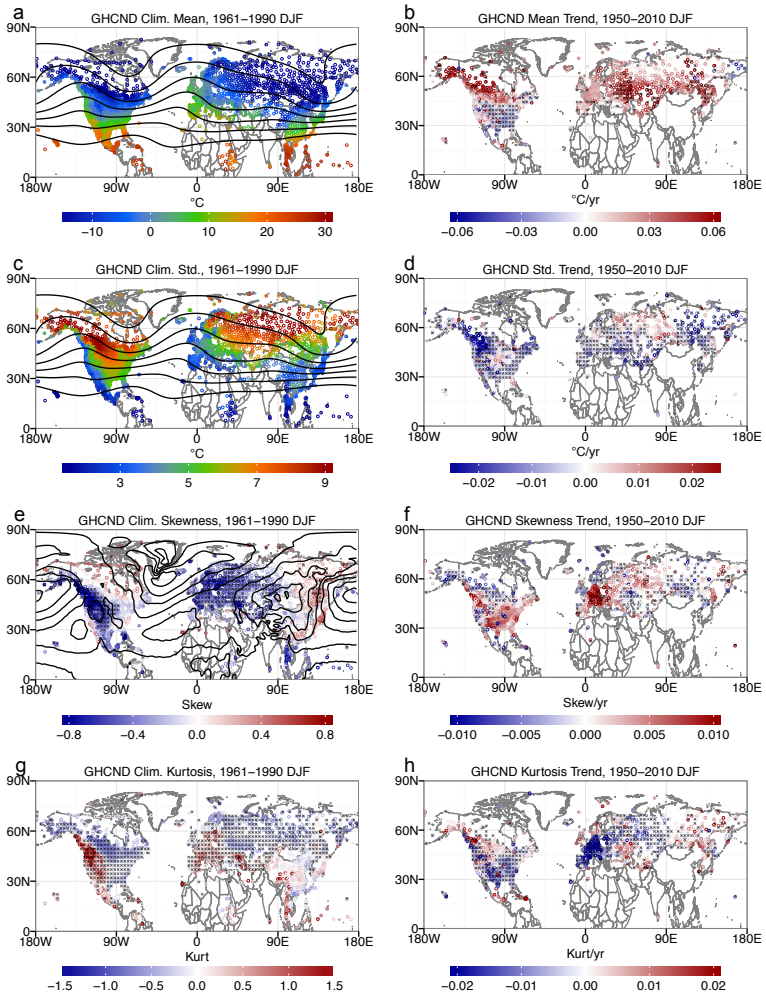

Figure 3.5: Same as Figure 3.2 for DJF.

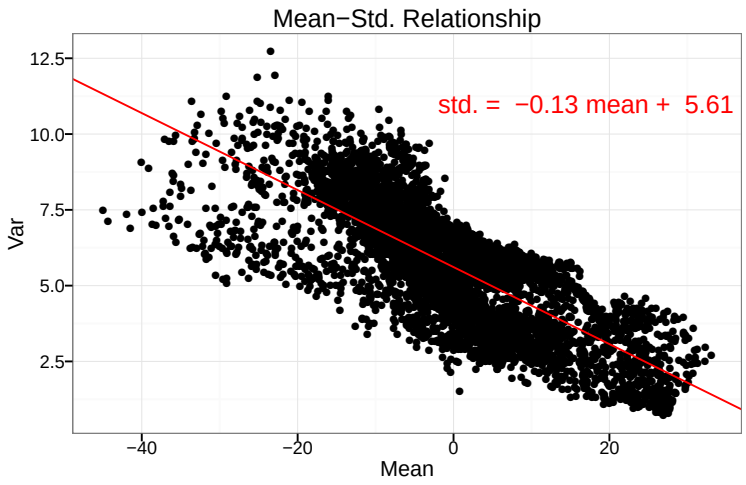

Figure 3.6: 1961-1990 climatological mean and standard deviation for DJF for GHCN-D stations included in the climatology. Other seasons are qualitatively similar. An empirical linear fit is plotted in red.

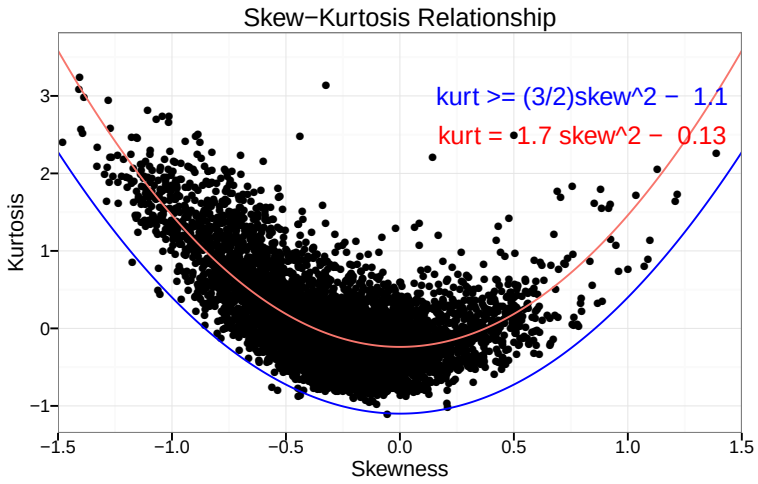

Figure 3.7: 1961-1990 climatological skewness and kurtosis for DJF for GHCN-D stations included in the climatology. Other seasons are qualitatively similar, except with varying levels of vertical shift. The kurtosis-skewness inequality is plotted in blue. An empirical parabolic fit is plotted in red.

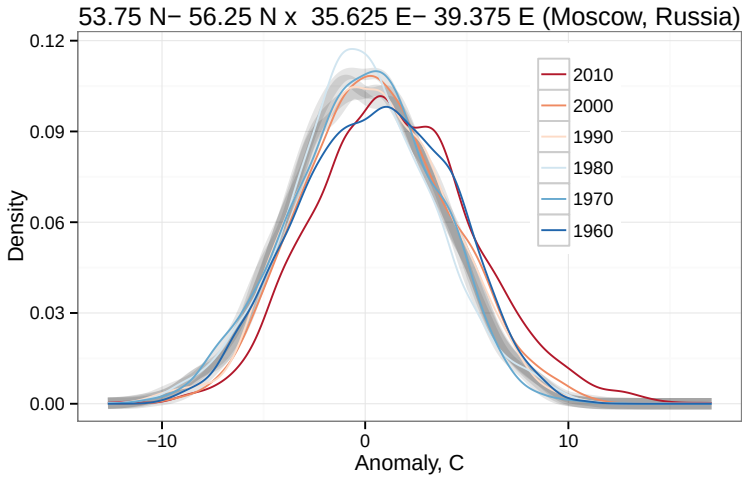

Figure 3.8: 1961-1990 climatological PDFs for each station contained in the grid cell covering Moscow, Russia (grey) overlaid with decadal PDFs determined by aggregating data from each station in the contained grid cell for each interval (colored). Decades are marked as the last year in the decadal period (e.g., 2010 represents the 2001-2010 interval).

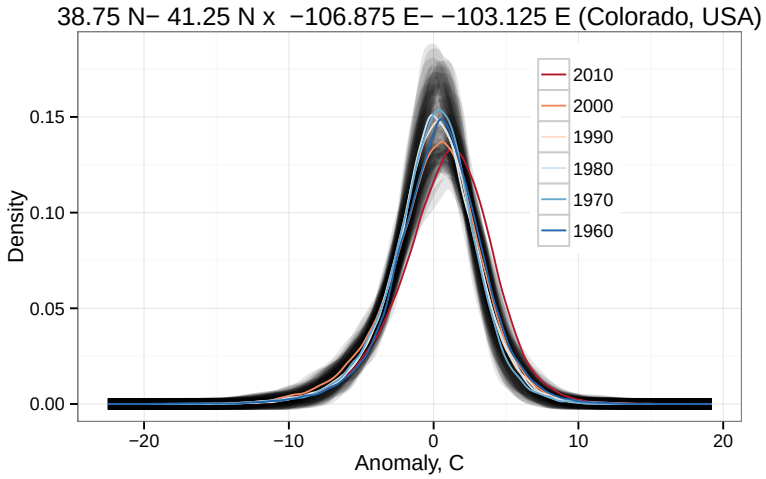

Figure 3.9: Same as Figure 3.8 for the region over central Colorado, USA.

Chapter 4

Spatial Scaling Effects on the Statistical Moments and Their Trends of Global Daily Surface Air Temperature

This paper explores the effects of weather station data aggregation on the first four statistical moments of daily minimum and maximum surface air temperature (SAT) anomalies over the entire globe. The Global Historical Climatology Network-Daily (GHCN-D) and UK Met Office Hadley Center's GHCN-D (HadGHCN-D) datasets from 1950-2010 are examined. The GHCN-D station data exhibit large spatial patterns for each moment and statistically significant moment trends from 1950-2010, indicating that SAT probability density functions are non-Gaussian and have undergone characteristic changes in shape due to decadal variability and/or climate change. Comparisons with gridded data show that grid-scale data underestimates climate variability and has altered moment trends that are in some cases opposite in sign over large geographic areas. A statistical mechanics approach is taken to explore SAT's higher-order moments and correlations and several critically important and intrinsic relationships are identified among the statistical moments affecting variability at grid-scale. The higher-order and nonlin-

ear spatial correlations up to fourth-order demonstrate that higher-order moments at grid-scale can be determined by functions of station pair correlations that depend on regional decorrelation length scales. The applicable scaling functions are identified and discussed. These results can aid in the development of constraints to reduce uncertainties in climate models and have implications for studies of atmospheric variability, extremes, and climate change using geographically smooth variables across spatial scales.

4.1 Introduction

Understanding climate variability from a probabilistic distribution point of view requires detailed information about various orders of statistical moments. Owning to spatial heterogeneity, various orders of spatial correlations are also critical information. A variety of gridded data and spatiotemporal resolutions are needed in the climate sciences, and numerous gridded observation data products focused on the first and second moments, i.e., the mean and variance, have been produced to facilitate the objective validation of climate models and studies of global climate change, e.g., NOAA OI SST (*Reynolds et al.*, 2002), GPCP (*Huffman et al.*, 2009), and NASA GISS (*Hansen et al.*, 2010). These gridded datasets are convenient for many applications, such as model forecast validation and studies of large-scale climate variability. It is, however, still unclear exactly how grid-scale variability is quantitatively related to weather variability observed at one point in space. To quantify station-level variability, *Cavanaugh and Shen* (2014) documented the Northern Hemisphere climatology and trends of higher-order statistical moments from the Global Historical Climatology Network-Daily (GHCN-D) surface air temperature (SAT) station dataset. A comparable study of grid-scale variability is also needed to facilitate climate model validation and climate risk analysis when higher moments are important, such as when studying high-frequency variability and climate extremes.

The importance of research on climate extremes and higher-order moments cannot be overstated. Changes to climate extremes, as a manifestation of non-

stationary low-frequency variability, are an important aspect of climate change and impact many facets of peoples lives. The reliable quantification of high-frequency variability and a precise description of the relevant physical mechanisms are critical in order to improve numerical weather and climate model predictions. However, these tasks are known to be both challenging and elusive in the climate community. While monthly and longer averages are approximately Gaussian (*Stephenson et al.*, 2004), the non-Gaussianity of daily data adds another facet of complexity. Non-Gaussian high-frequency climate processes require that more care be taken towards understanding the relevant probability distributions, higher-order moments, tail characteristics and the probabilities of certain extremes, as well as how they scale in space. The main purpose of this paper is to examine these properties for daily SAT anomalies over the entire globe.

Numerous gridding algorithms have been developed to translate observed point data through weighted averaging to grid-scale approximations; canonical examples include kriging-based optimal interpolation (*Gandin*, 1963) and angular distance weighting interpolation (*Shepard*, 1968). Considering that each observation can be thought of as a realization of a random variable (RV) at a given time and location, it is possible that the distribution of the resultant interpolated RV might be altered compared to those of the contributing RVs defined at point-scale. However, when using interpolated geophysical data, it is commonly assumed that the gridding effects on variability distributions are of second-order compared to the variability of the underlying mean signal and thus that the resulting gridded data can be routinely used for studies of climate variability. This paper will demonstrate that this assumption can be violated in some cases and that cautions need to be taken when dealing with higher-order moments and non-Gaussian variability.

Daily SAT is a smooth spatial anomaly field with large decorrelation distances. *Cavanaugh and Shen* (2014) examined the climatological moments of daily SAT anomalies (T_{avg}) up to fourth order and demonstrated that in most locations and seasons T_{avg} distributions are non-Gaussian and that their spatial patterns in moments are highly coherent on regional scales, implying that the SAT variability distributions of nearby weather stations are (nearly) identically distributed.

Caesar et al. (2006) investigated the decorrelation distances for daily maximum and minimum SAT (T_{min} and T_{max}, respectively) and showed that e-folding correlation distances are approximately 1000km globally, with variations of up to a couple hundred km according to the variable, location, and season. *Caesar et al.* (2006) also discussed the compilation of the UK Met Office Hadley Centers GHCN-D (HadGHCN-D) dataset, the only currently available daily gridded observation SAT dataset derived from the Global Historical Climatology Network Daily dataset (GHCN-D; *Menne et al.*, 2012). This study further examines the moments of HadGHCN-D SAT up to fourth order through the lens of statistical mechanics to demonstrate the effects on variability distributions resulting from data aggregation and interpolation. We aim to reveal the relationships between variability observed at point-scale by weather stations and that estimated or simulated at larger scales, i.e., gridded data. We also aim to highlight parallels between methods common to statistical mechanics and those common to climatology in order to provide a new perspective on climate variability. To accomplish these tasks, we first develop the basic mathematical theory necessary to understand non-Gaussian variability distributions across spatial scales. We then compare the moment statistics and trends from the gridded dataset, HadGHCN-D, with those taken directly from the GHCN-D stations themselves to illustrate our results. As a byproduct of these goals, we present both station and gridded SAT datasets of mean, variance, skewness and kurtosis for the entire global land (where observations are available) for T_{avg}, T_{min}, and T_{max}.

In Section 4.2, we derive the mathematical theory by moments for weighted sums of dependent RVs and interpret its meaning from the perspective of statistical mechanics. Section 4.3 discusses the GHCN-D weather station dataset and the general form of gridding algorithms commonly used in practice. We also briefly describe our methodology for calculating climatological moments as well as trends over 1950-2010. Section 4.4 evaluates the (higher-order) correlation terms originally identified in Section 4.2. Section 4.5 describes climatological moments for the gridded dataset and their comparisons to station data. We present trends in moments over 1950-2010 and show how gridded variability trends can deviate or

oppose those estimated directly from station data. Section 4.6 explains our findings in the broader context of multi-scale climate variability, and Section 4.7 contains conclusions and discussion.

4.2 Higher-Order Moments and Their Relations to Higher-Order Covariances

4.2.1 A Central Limit Theorem by Moments

SAT at a given location or grid cell can be treated as a RV with realizations in time indexed by t. For any discrete RV X, its first moment, μ_1, is the expected value

$$\mu_1 = E[X] = \sum_t p_t x_t = \overline{X} \tag{4.1}$$

where x_t is a random outcome with uniform probability p_t in the X's sample space (see any introductory statistics textbook, e.g., *Ross*, 2010). The overbar represents an arithmetic ensemble mean or expected value and is equivalent to $E[\cdot]$. The n^{th}-order moments of X, μ_n for $n \geq 2$, are defined as

$$\mu_n = E[(X - \overline{X})^n] \tag{4.2}$$

The 2^{nd}-order moment is commonly referred to as the variance and has the same units as X^2 (and its square root is the standard deviation with the same units as X). The 3^{rd}- and 4^{th}-order moments are directly related to the skewness and kurtosis, which are suitably non-dimensionalized by the standard deviation (see eqs. (4.10) and (4.11) below, also *Cavanaugh and Shen* (2014), *Wilks* (2011), *Ross* (2010)). The skewness measures the asymmetry of a distribution and is usually sufficient to demonstrate non-Gaussianity since Gaussian RVs necessarily have zero skewness. The kurtosis measures the tail thickness and peakedness of the distribution, with higher values indicating increased probabilities of large deviations from the mean in symmetric distributions. Gaussian distributions have kurtosis equal to 3.

The transformation of station data into data on grid cells using weighted averages leads to a grid cell based RV, $\dot{X}$, which is the weighted sum of any number of station RVs from the set $i = [1, 2, \ldots, s]$ (which are now defined as the zero-centered anomalies without loss of generality) with weights w_i and a total number of stations s. We wish to explore how the moments of $\dot{X}$, denoted as $\dot{\mu}$, are related to higher-order covariance terms.

For the mean, the expectation of the grid box RV remains zero, unchanged from those of the station RVs, i.e.,

$$\dot{\mu}_1 = \overline{\sum_i w_i X_i} = \sum_i w_i \overline{X_i} = 0 \tag{4.3}$$

The n^{th}-order moments, $\dot{\mu}_n$, when $n \geq 2$, are much more complex and are functions of the weights as well as various higher-order polynomial covariance terms. The mathematical expressions for the second through fourth moments are displayed below:

$$\dot{\mu}_n = \overline{(\sum_i w_i X_i)^n}$$

$$\dot{\mu_2} = \sum_i w_i^2 \overline{X_i^2} + \sum_{i \neq j} w_i w_j \overline{X_i X_j} \tag{4.4a}$$

$$\dot{\mu_3} = \sum_i w_i^3 \overline{X_i^3} + 3 \sum_{i \neq j} w_i^2 w_j \overline{X_i^2 X_j} + \sum_{i \neq j \neq k} w_i w_j w_k \overline{X_i X_j X_k} \tag{4.4b}$$

$$\begin{aligned} \dot{\mu_4} = & \sum_i w_i^4 \overline{X_i^4} + 4 \sum_{i \neq j} w_i^3 w_j \overline{X_i^3 X_j} + 3 \sum_{i \neq j} w_i^2 w_j^2 \overline{X_i^2 x_j^2} + \\ & + 6 \sum_{i \neq j \neq k} w_i^2 w_j w_k \overline{X_i^2 X_j X_k} + \sum_{i \neq j \neq k \neq l} w_i w_j w_k w_l \overline{X_i X_j X_k X_l} \end{aligned} \tag{4.4c}$$

Equation (4.4a) defines the variance for the weighted sum of RVs and is equal to the weighted sum of all of the terms in the 2-dimensional covariance matrix of dimension $s \times s$. Similarly, eq. (4.4b) defines the third-order moment for the weighted sum and is equal to the weighed sum of all the terms in the 3-dimensional coskewness matrix of dimension $s \times s \times s$. Lastly, eq. (4.4c) defines the fourth-order moment for the weighted sum and is equal to the weighted sum of all the terms in the 4-dimensional cokurtosis matrix of dimension $s \times s \times s \times s$. If $\dot{\mu}_3$ and $\dot{\mu}_4$ are normalized by $\dot{\mu}_2$ to be dimensionless, they define the skewness and kurtosis of

the weighted sum, respectively. The last terms in eqs. (4.4a-c) are worth special attention as they are related to the collinearity among the measurements. The non-zero values of these terms make it apparent that weighted aggregation is needed to form gridded data from station data. A traditional simple average using equal weights is the optimal average only for the case of a spatially white noise field since the high-order covariances are all zero. Climate researchers have noted collinearity in climate data for a long time. Empirical orthogonal functions (EOFs) are a way to remove the collinearity in the EOF spectral space (*North*, 1984; *Wilks*, 2011) and to formulate optimal averaging and optimal gridding algorithms for globally inhomogeneous and correlated random fields, such as SAT and precipitation (*Shen et al.*, 1994, 1998, 2004, 2014). While second-order covariances are familiar to the climate research community in the usual definition, the higher-order expectation matrices, termed co-skewness and co-kurtosis, are uncommonly used in climate research and were coined in the field of financial mathematics in the 1970s for the purpose of valuing risk portfolios. Additionally, their contained higher-order covariance terms are routinely used to examine the properties of random fields in the study of statistical fluid mechanics. For convenience, we will employ the simple notation $B_{o(i),o(j),o(k),o(l)}$ to denote correlations terms of their respective order (e.g. $B_{2,1} = \overline{X_i^2 X_j}$, $B_{2,1,1} = \overline{X_i^2 X_j X_k}$, $B_{1,1,1,1} = \overline{X_i X_j X_k X_l}$, etc).

Since we will see that the moments of $\mathrm{T_{min}}$ and $\mathrm{T_{max}}$ are nearly equal over large spatial areas, we can consider the RVs X_i in eqs. (4.3) and (4.4) to be identically distributed. In the case of independent RVs, B_2 is the only contributing term to $\dot{\mu}_2$, B_3 is the only contributing term to $\dot{\mu}_3$, and B_4 and $B_{2,2}$ are the only contributing terms to $\dot{\mu}_4$, with the dominance of $B_{2,2}$ increasing as the number of weighted RVs increases. In the special case when all X_i are perfectly dependent, then $\dot{\mu}_n = \mu_n$ and the probabilistic distribution of the weighted sum is perfectly preserved. With regard to SAT, this would imply perfectly homogeneous weather stations with identical time series. However, in practice, station SAT correlations usually decrease with distance (*Caesar et al.*, 2006) when the teleconnection range is not yet reached. As such, the Cauchy-Schwarz inequality implies that $\dot{\mu}_2 \leq \mu_2$, or more simply that the variance of the interpolated data is less than that of the

contributing stations (*Shen et al.*, 2012). Although not rigorously proven, it might be reasonable to also claim that $|\dot{\mu}_3/\dot{\mu}_2^{3/2}| \leq |\mu_3/\mu_2^{3/2}|$ and $|\dot{\mu}_4/\dot{\mu}_2^2 - 3| \leq |\mu_4/\mu_2^2 - 3|$ will hold for finite s.

4.2.2 Quasi-Gaussian Random Fields

Time series of SAT anomaly measurements are assumed piecewise-stationary for each season. According to moment generation function theory (*Wackerly et al.*, 2008), the probability density function (pdf) of an RV can be determined by its moments. Some pdfs can be determined by finitely many moments. For example, a Gaussian RV's distribution can be determined by the moments up to the second order. However, for some RVs a non-zero third-order moment (as is the case for SAT) is important, and even higher-order correlation terms are non-negligible and can require more simultaneous means, such as triplets and quadruplets (see eq. (4.4)). We will truncate the terms at quadruplets in this paper.

Millionshchikov (1941) proposed that quasi-normal or quasi-Gaussian fields are characterized by vanishing fourth-order cumulants, leading to closure of the fourth-order moments of the following form,

$$\overline{X_i X_j X_k X_l} = \overline{X_i X_l} \cdot \overline{X_j X_k} + \overline{X_i X_k} \cdot \overline{X_j X_l} + \overline{X_i X_j} \cdot \overline{X_k X_l} \tag{4.5}$$

implying that fourth-order correlations can be decomposed into functions of pair correlations. A thorough treatment of this theory is given by *Monin and Yaglom* (1971).

By combining the Cauchy-Schwarz inequality with the quasi-Gaussian approximation (eq. (4.5)), *André et al.* (1976) demonstrated analytically that bounds on the third-order correlations that prohibit the nonphysical development of negative energy spectra take the form

$$|\overline{X_i X_j X_k}| \leq \min \left\{ \begin{array}{l} [\overline{X_i^2}(\overline{X_j^2 X_k^2} + \overline{X_j X_k}^2)]^{1/2} \\ [\overline{X_j^2}(\overline{X_i^2 X_k^2} + \overline{X_i X_k}^2)]^{1/2} \\ [\overline{X_k^2}(\overline{X_i^2 X_j^2} + \overline{X_i X_j}^2)]^{1/2} \end{array} \right\} \tag{4.6}$$

and provide the criteria for the clipping approximation of numerical simulations of turbulent inhomogeneous flows (*Andre et al.*, 1976). Eq. (4.6) implies that the

triple correlations are bounded by functions of the smallest of the squared pair correlations for standardized variables. The relationships in eqs. (4.5) and (4.6) will be utilized in Section 4.4 to structure the parallels between statistical fluid mechanics theory and climatological observations, and Sections 4.5 and 4.6 will further discuss their relevance to the study of multi-scale climate variability.

4.3 Data and Methodology

4.3.1 Station GHCN-D Data

GHCN-D contains *in situ* observational data from over 90,000 weather stations in 180 countries and territories worldwide (*Menne et al.*, 2012). The GHCN-D source data have been compiled with the goal of maximizing spatial coverage. Records from numerous agencies have been collected and quality controlled, but not homogenized. Before new data are added to the dataset, they are first cross-checked with included data to confirm record uniqueness, spatial consistency, and temporal consistency (*Menne et al.*, 2012), which reduces the chances of redundancy. Each record is then subject to 19 quality assurance (QA) measures and flagged with appropriate tags so that data can be utilized or discarded as necessary by a user (*Durre et al.*, 2000).

Our study leaves out any data that have been QA flagged. In order to define our anomaly time series, a running average of five days is first performed on the T_{max} and T_{min} records in order to fill in the holes of a few days in the climatology period 1961-1990, and seasonal cycles are computed from the 1961-1990 smoothed values, producing daily T_{max} and T_{min} climatologies for each station. Each station's T_{max} and T_{min} climatology is then subtracted from the raw T_{max} and T_{min} data to produce T_{max} and T_{min} anomalies. Only stations that include at least 20 years of data (7,300 days) over the 1961-1990 period are included in this study.

4.3.2 Gridded GHCN-D Dataset

HadGHCN-D is a recent attempt at producing a gridded surface observation-based climate database at daily resolution focused on providing information relevant to temperature and precipitation extremes and climate model evaluation (*Caesar et al.*, 2006). It is comprised of GHCN-D stations which have been interpolated to a grid based on an angular distance-weighting algorithm (*Alexander et al.*, 2006; *Caesar et al.*, 2006). The goal of performing weighted average interpolation is to leverage shared information summarized by the dependencies between nearby stations so that weighted combinations of data sources contribute to a more spatiotemporally complete record. Often in station records incompleteness and inconsistencies arise from equipment malfunctions, human error, and systematic changes, which manifest themselves in the record as unbelievable or missing data, instrumental drift, and discontinuous jumps, among other imperfections. Weighted interpolation seeks to aggregate stations contained within some distance scale D, in the case of HadGHCN-D an e-folding correlation distance, which are thought to contain at least some component of the information that should represent a grid cell centered at a certain longitude-latitude location, ϕ' and θ', and weight those stations as a function usually of their distance from the grid-cell, d_i (favoring closer stations), and their geographic location, ϕ_i and θ_i.

$$X'(\phi', \theta') = \sum_i w_i(d_i, \phi_i, \theta_i) X_i \; ; \;\; d_i < D \; , \; \sum_i w_i = 1 \tag{4.7}$$

Since varying station density near a grid-cell can bias this interpolation procedure, it is common to require a minimum number of stations (3 for HadGHCN-D) and also to limit the maximum number of contributing stations (10 for HadGHCN-D). Fig. 4.1 shows a general diagrammatic representation of the HadGHCN-D gridding procedure. Other optimal or variogram-based interpolation methods do not explicitly impose a D scale, but rather utilize the covariance structure between stations to implicitly define this limit by giving lower weights to uncorrelated stations (*Gandin*, 1963). The relationship of eq. (4.7) to the theory outlined in Section 4.2 is immediately evident.

4.3.3 Methods for Calculating Higher Moments and Their Trends

Following *Cavanaugh and Shen*'s (2014) approach, we calculate higher statistical moments over the climatological period 1961-1990 and over the moving time window of each decade during 1950-2010 for both station and gridded data. We also calculate the moment trends over the 1950-2010 period. We include only the T_{max} and T_{min} data streams with at least 20 years of data over the climatological period 1961-1990. The data streams are separated into four seasons, Dec-Jan-Feb (DJF), Mar-Apr-May (MAM), Jun-Jul-Aug (JJA), and Sept-Oct-Nov (SON). These seasonal records are then decomposed into their first four moments. The arithmetic mean measures the first moment

$$\overline{X} = \frac{1}{\tau}\sum_{i=1}^{\tau} x_t \tag{4.8}$$

with time index t for non-missing datum x_t contained in the 30-year distribution for each season. We use the standard deviation to represent the second moment

$$\hat{\sigma} = \left[\frac{1}{\tau}\sum_{t=1}^{\tau}(x_t - \overline{X})^2\right]^{1/2}, \tag{4.9}$$

so that the units of the first two moments are equal. The third and fourth standardized moments are characterized by the sample skewness and sample kurtosis, respectively,

$$Skew(X) = \frac{1}{\tau}\sum_{t=1}^{\tau}\frac{(x_t - \overline{X})^3}{\hat{\sigma}^3}, \tag{4.10}$$

$$Kurt(X) = \frac{1}{\tau}\sum_{t=1}^{\tau}\frac{(x_t - \overline{X})^4}{\hat{\sigma}^4}, \tag{4.11}$$

which are scaled by the standard deviation to be dimensionless. Note that $\overline{X} = 0$ in our case, since anomaly time-series have zero mean for the entire climatology period.

To examine trends in the moments over a 1950-2010 period, the same moment decomposition methodology described above is applied to decade-length moving bins, the first being 1950-1959, then 1951-1960, and so on, for a total of 52

bins. Each decade is required to be at least 75% complete for its moments to be computed. The resultant 52-year time-series for each moment are then fit with a generalized linear regression model with a first order autoregressive (AR) process,

$$M_{j,t+1} = \alpha_j + \beta_j t + u_{j,t} \tag{4.12a}$$

$$u_{j,t} = \rho_j u_{j,t-1} + \epsilon_{j,t} \tag{4.12b}$$

$$\epsilon_{j,t} \sim N(0, \sigma_j^2) \tag{4.12c}$$

for the seasonal moment μ at for year t, with y-intercept α, slope β, error term u, AR(1) coefficient ρ, and random Gaussian noise ϵ of zero mean and σ standard deviation. Thus, for each moment and for each season, the slope parameter β summarizes the change rate per year for that moment.

4.4 Relationships Between Higher-Order SAT Correlations

Each station i that has at least 20 years of climatological data is used in this study, both at the station level and in the gridding procedure. A maximum search radius of 4000km is used to identify which other stations are within the proximity of the station i in question. From the subset of stations identified for station i, all possible combinations of stations are used to estimate the expectation terms that arise in eqs. (4.4a-c) using the definition of the expectation in eq. (4.1), including the higher-order terms that contribute to the coskewness and cokurtosis matrices. This process is repeated for each contributing station (about 8200 in total) for both $\mathrm{T_{min}}$ and $\mathrm{T_{max}}$. The third- and fourth-order expectations are sub-sampled in order to make the problem computationally tractable, resulting in about 36-38 million estimates for each covariance term. Station records are first normalized by their standard deviations to ease interpretation.

4.4.1 Pair Correlation

Figure 4.2 displays elements contributing to the second term in eq. (4.4a), i.e., the station pair correlations $B_{1,1}$, as a function of the distance between stations

normalized by an interpolated e-folding correlation distance for station i, λ_i (see Fig. 4.3). Following *Hansen and Lebedeff* (1987), we separate correlation plots into 30° latitude bands. Due to the very large number of data points, station correlations are plotted here as a probability density estimated from station pairs that fall within each 0.1-width bin spanning $d/\lambda = [0, 4]$. Solid and dashed lines represent the mean and uncertainty bounds estimated for each bin and thus can be interpreted as a continuous box-and-whisker plot.

In agreement with *Hansen and Lebedeff* (1987), *Caesar et al.* (2006) and others, our station pair correlations for T_{min} and T_{max} decay smoothly to zero as a function of the distance separating the stations. Also, when normalizing by a locally-determined e-folding correlation distance, the decay rates globally collapse to the same functional form for both T_{min} during DJF and T_{max} during JJA, as well as for other seasons. Uncertainty bounds increase as station density decreases and the magnitude and shape of the uncertainty envelope is similar across each latitude bin and for both T_{min} and T_{max}.

Figure 4.3 shows the e-folding correlation distances for each station estimated by spline interpolation from the subset of stations within 4000km plotted geographically. The e-folding correlation distances vary globally largely in response to regional geography, particularly for T_{min}. SAT decorrelation distances tend to be larger during each station's respective synoptically active seasons (fall, winter, spring), which is indicative of more pronounced large-scale meteorological conditions affecting or determining SAT at those times.

4.4.2 Coskewness

Figure 4.4 displays eq. (4.4b)'s elements that contribute to the coskewness matrix, scaled by the third moment of station i to facilitate understanding; station pairs/triplets whose scaled higher-order correlations are between -1 and 1 indicate that the third moment of the weighted sum of those stations is closer to zero than the third moment of station i; this might also imply that the skewness of the weighted sum is closer to zero since the B_3 term in eq. (4.4b) decays faster than the B_2 term in eq. (4.4a). Only stations whose standardized third moment

is significantly different from zero, based on the standard error of skewness (see *Cavanaugh and Shen*, 2014) are plotted. Figs. 4.4a,b show $B_{2,1}$ correlations plotted against their respective $B_{1,1}$ correlations. $B_{2,1}$ scales approximately linearly with $B_{1,1}$, where higher $B_{1,1}$ indicates that $B_{2,1}$ is nearer to observed third moment for station i.

Figs. 4.4c,d display the $B_{1,1,1}$ correlations plotted against the minimum pair correlation between each of the three stations. Plotting against the minimum pair correlation is inspired by the inequality in eq. (4.6), which is satisfied by our observations. $B_{1,1,1}$ scales approximately linearly with the minimum pair correlation, where higher average pair correlations indicate that triple correlations are nearer to the observed third moment for station i.

In all cases, normalized third-order expectation distributions are unimodal and appear to be approximately symmetric about the mean (solid line). Uncertainty bounds for third-order terms are large relative to those of the other moments and are proportional to increasing pair correlations. Some portion of this is due to the method of scaling the observed third-order correlations by the third moment of station i. Third moments close to zero may inflate the ratio between the third moment and the higher-order correlation in question, thereby amplifying the uncertainty bounds.

4.4.3 Cokurtosis

Figure 4.5 shows the elements of the cokurtosis matrix defined in eq. (4.4c) plotted against functions of the pair correlations predicted for a quasi-Gaussian random field (see eq. (4.5)). Cokurtosis terms are scaled by the fourth moment of station i; a scaled fourth-order correlation less than 1 indicates that the fourth moment of the weighted sum of those stations is closer to 0 than the fourth moment of station i; this might also signify that the kurtosis for the weighted sum is closer to that of a Gaussian distribution since the B_4 term in eq. (4.4c) decays faster than the B_2 term in eq. (4.4a). In all cases, scaled $B_{3,1}$ (Figs. 4.5a,b), $B_{2,2}$ (Figs. 4.5c,d), $B_{3,1,1}$ (Figs. 4.5e,f) and $B_{1,1,1,1}$ (Figs. 4.5g,h) have a one-to-one relationship with their respective pair correlation function for a quasi-Gaussian system and have

very small error bounds relative to the equality signal. This result signifies that SAT constitutes a nearly quasi-Gaussian system with fourth-order moment closure (see eq. (4.5)).

4.5 Observed Multi-Scale SAT Variability

In light of the observed dependence measures documented in Section 4.4, we examine the climatological SAT moments estimated from GHCN-D seasonal data during 1961-1990 and compare these distribution characteristics to those computed using HadGHCN-D. The trends of these moments over 1950-2010 computed from GHCN-D and HadGHCN-D datasets are compared. Significant disagreement between trends and/or trend patterns exist between the two datasets. This paper only shows the results for T_{min} during DJF and T_{max} during JJA to maximize interest among the climate community.

4.5.1 Climatological T_{min} and T_{max}

Figures 4.6 and 4.7 illustrate the spatial distribution climatological moments for T_{min} during DJF and T_{max} during JJA, respectively. The left columns of Figs. 4.6 and 4.7 are the moments computed from GHCN-D station data, and the right columns are the moments computed from the gridded HadGHCN-D data minus those from the GHCN-D data. The station anomaly pdfs for both T_{min} and T_{max} are regionally approximately identical, as evidenced by the smooth and coherent spatial structure of moments at regional scales (Figs. 4.6-4.7a,c,e; discussed in detail by *Cavanaugh and Shen*, 2014). Additionally, distributions of these SAT variables are non-Gaussian and have somewhat complex spatial structures that are influenced by regional geography and land/atmosphere characteristics.

Immediately evident from Figs. 4.6-4.7b is that the standard deviations estimated from gridded observations are greatly reduced from those estimated directly from station data. Globally, reductions in standard deviations are largest in regions where station density is low, and lowest where station density is high. The clearest example of this can be observed at the border of the United States and

Mexico where station density goes from high to low when crossing the border going south. This effect is due to the contribution of the $B_{1,1}$ terms in eq. (4.4a); in areas with high station density, only stations that are geographically close to the center of the grid cell are used in interpolation, and thus their pair correlations are close to one (see Figs. 4.2a,b). Large reductions in standard deviation arise in large-scale regions where the standard deviations are highest in the climatological record, e.g., for T_{min} (Fig. 4.6a), in Northern Europe and Russia and for T_{max} (Fig. 4.7a), in the North West United States, Northern Europe, and Russia. These reductions occur resultant from the $B_{1,1}$ terms in eq. (4.4a), where the magnitude of the reduction should arise proportionally to the climatological standard deviations of the contributing stations. There are also more localized regions where there are large reductions in standard deviation, e.g., for T_{min}, along the west coast of North America, and for T_{max}, on the Iberian Peninsula and in Mexico. These reductions result from reduced $B_{1,1}$ terms in eq. (4.4a) arising from the short decorrelation distances in those areas (Figs. 4.3a,b).

In some relatively rare cases, the climatological standard deviation from the gridded data is slightly increased relative to the station record. These increases might arise do to local-scale station inhomogeneity (for example those arising from variable station exposures or wind-chill effects, see *Fall et al.* (2011); and/or landscapes and land cover variability, see *Montandon et al.* (2011)), weighting effects of the chosen interpolation algorithm (which might occur, for instance, in the case where poorly sited stations are isolated in space), and/or from non-static observation records or recording techniques, all of which are likely the major contributors in the United States and Western Europe. In other areas that are less station-dense, station inhomogeneity within a large decorrelation length scale is the more likely contributor, as is probably the case, for example, near Afghanistan (see Fig. 4.6b).

At large-scale, skewness values in the gridded dataset for both T_{min} and T_{max} (Figs. 4.6-4.7d) tend to be closer to zero than in the station observations (i.e., the deviations are negatively correlated with the station observations of skewness illustrated in Figs. 4.6-4.7c), indicating that gridded variability is more sym-

metric than in the stations themselves. Spatial correlations between Figs. 4.6c,d and 4.7c,d are -0.435 and -0.502, respectively. Similarly, kurtosis deviations (Figs. 4.6-4.7f) also tend to be negatively correlated with observations (Figs. 4.6-4.7e), indicating that the peakedness of gridded variability is closer to that of a Gaussian distribution than the stations themselves, as one would expect from the CLT. Spatial correlation between Figs. 4.6e,f and 4.7e,f are -0.532 and -0.559, respectively. Both of these effects can be anticipated from Figs. 4.4 and 4.5. As can be deduced by comparing the broad distributions of third-order moments (Fig. 4.4) to the narrow distributions of the fourth-order moments (Fig. 4.5), the magnitude of the anti-correlations between station skewness and gridded skewness are smaller than those of kurtosis, indicating that convergence to distribution symmetry is less-strict than convergence to symmetric exponential tail-decay. Gridded skewness and kurtosis biases tend to be largest both where the magnitudes of the station moments are the largest and where the decorrelation distances are the shortest. It is difficult to ascribe which contribution is dominant, since these regions tend to overlap when comparing the station climatologies (Figs. 4.6-4.7c,e) to the e-folding distances (Figs. 4.3a,b). These results can be expected from Figs. 4.4 and 4.5. Third- and fourth-order correlations scale as functions of the pair correlations of the contributing stations and are thus related to the distances separating the stations used for interpolation. Since we have chosen to normalize the higher-order correlations by the third and fourth moments of station i, in each instance where the normalized third- or fourth-order correlation is less than one (which is where the bulk of the observations lie), we can expect from eqs. (4.4) and (4.5) that the higher-order moments of the gridded observations should be more Gaussian. Similar to the explanation of results regarding the second moment, gridded data are most Gaussian both where station density and/or local decorrelation distances (Figs. 4.3a,b) are low.

Area averaging over homogeneous stations minimizes bias with respect to the mean at the cost of changes to the distribution around it. While it is commonplace to use gridded data in studies of extremes, reductions and modulations to the observed variability resultant from area averaging have the largest impacts on

the tails of pdfs that dictate extremal probabilities. In the least biased scenario of approximately Gaussian station distributions with very high station density (e.g., over parts of the United States), the probability of exceeding two standard deviations at the station level is generally about twice that estimated from gridded data located over the same region (not shown). By contrast, in greatly skewed regions with low station density (e.g., over parts of Russia), these probabilities can exceed five times those estimated from gridded data located over the same region (not shown). It is thus quite important to stress that caution must be taken when utilizing gridded data to study climate extremes.

4.5.2 Moment Trends 1950-2010

The observed climatological distributions of SAT are non-Gaussian; additionally, the differences in moments between station and gridded SAT data are of the same order of magnitude as the moments themselves (Figs. 4.6-4.7). The first result necessitates the inclusion of higher-order moment effects in studies of non-stationary high-frequency climate variability. The second result suggests that trends in the moments of gridded data might also differ from those of the station data, potentially polluting inferences concerning the effects of climate change on weather variability. This section describes linear trends in the moments of T_{min} during DJF and T_{max} during JJA from 1950-2010 for both station data and gridded data using the methodology recapitulated in Section 4.3.3 (for details see *Cavanaugh and Shen* (2014)). We would like to explore what effect a gridding algorithm may have on trends compared with station data.

Figures 4.8 and 4.9 illustrate trends in moment statistics in the left column and biases in these trends caused by interpolation to a grid in the right column. Both T_{min} and T_{max} show large-scale spatially-coherent trends in moments. For T_{min}, the standard deviation trends are largely negative over the Northern Hemisphere, indicating less variability in winter nighttime temperatures. Surprisingly, trends in standard deviations for gridded data are mostly more positive than the trends from the station data; in many cases (particularly over Russia, Western Canada, and Alaska), the directions of the trends oppose each other. Standard

deviation trends from T_{max} are more mixed globally, although again, in many cases (particularly over the Northern United States and Western Russia), gridded trends oppose those taken from the stations directly.

SAT trends in the skewness and kurtosis from gridded data are mixed globally, but are largely spatially coherent, and are in many regions significantly nonzero. For T_{min}, the largest-magnitude trends in skewness tend to be positive (over the Northwest and Southeast of the United States and in western Europe), and also tend to coincide with the areas of dominant negative trends in kurtosis. For T_{max}, higher-order moment trends are variable at the regional level, and in some regions, areas with large trends in skewness and kurtosis tend to coincide. Under these circumstances, this may indicate a particularly large change in the structure of SAT extremes at these locations over decadal timescales. Similar to standard deviations, biases in the gridded data higher-order moment trends are mostly regionally coherent, and in some cases contribute to trends that oppose those observed in the station records themselves. These differences in moment trends are obviously of practical significance when considering the likelihood of events in a non-stationary climate.

While it is possible that intrinsic variability at grid-scale could have different trends than variability at point-scale, this possibility is unlikely. As such, the opposite signs of the trends observed in station data and gridded data in some regions make it necessary to identify the mechanisms responsible for those differences, although pinpointing specific mechanisms is a very challenging task because many factors can contribute simultaneously. A few are listed below:

1. *The number of stations contributing to the area average increases (decreases) over the trend period.* Trends in gridded standard deviations are more negative (positive) than in the stations themselves. Trends in higher moments are biased toward becoming more (less) Gaussian than in the stations themselves. Both changes are resultant from an increased (decreased) contribution of off-diagonal terms in eqs. (4.4a-c).

2. *The distances from the center of the grid to the stations contributing to the area average increases (decreases) over the trend period*, resulting in greater

station heterogeneity (homogeneity). Trends in gridded standard deviations are more negative (positive) than in the stations themselves. Trends in higher moments are biased toward becoming more (less) Gaussian than in the stations themselves. Both changes are resultant from a decreased (increased) magnitude of off-diagonal terms in eqs. (4.4a-c), assuming stations are uniformly distributed.

3. *The distances between stations contributing to the area average increases (decreases) over the trend period.* Trends in gridded standard deviations are more negative (positive) than in the stations themselves. Trends in higher moments are biased toward becoming more (less) Gaussian than in the stations themselves. Both changes are resultant from a decreased (increased) magnitude of off-diagonal terms in eqs. (4.4a-c).

4. *The weighting scheme between stations contributing to the area average changes over time as a result of changing station locations and data availability.* Gridded moment trends can be biased in either direction.

5. *(Higher-order) correlations between stations contributing to the area average are non-stationary.* Gridded moment trends can be biased in either direction.

Forcing the interpolation scheme to be constant in time can alleviate these artifacts; however, it may come at the cost of reducing accuracy with respect to the probabilistic mean and sacrificing some potentially useful information.

4.6 General Implications for Multi-Scale SAT Variability

While weather station observations from the GHCN-D dataset are taken at one point in space and time, the moment statistics presented in Section 4.4 can shed additional light on multi-scale SAT variability across both space and time. Specifically, the observed moment statistics, which describe a quasi-Gaussian random field, suggest that the first four moments are sufficient in order to characterize

all observable or estimable SAT distributions at any scale via eq. (4.4). The descriptive moment climatologies presented in Figs. 4.6 and 4.7 quantify the upper bounds on the degree non-Gaussianity for any SAT distribution since averaging results in distributions that are increasingly Gaussian. As explicitly discussed in Section 4.4, spatial averages over homogeneous areas, which by the law of large numbers should approximate the means in probability over those areas, are merely weighted sums (or integrals with a weighting function) of quasi-Gaussian RVs. The resulting distributions will in fact be closer to Gaussian than the distributions at point locations in space. Analogously, atmospheric autocorrelations decay in time, and moments for temporal averages follow the same functional form as in eq. (4.4), implying that the distributions for temporal averages will also be more Gaussian than those of the contributing instantaneous observations.

Additionally, observed SAT distributions are closed at the fourth moment, indicating that moments for gridded data can be estimated from $B_{1,1}$ pair correlation terms (see Figs. 4.4 and 4.5) and the moments of the stations that contribute to the grid cell approximation. The implication is that by keeping track of which stations contribute to averages in gridding algorithms, their respective assigned weights, and their higher-order correlations, a mapping between the averaged variability and variability at a point in space can be deduced (this, again, analogously applies in time as well). For spatial averages, this of course can practically be accomplished (approximately) by merely keeping track of the distances between contributing stations. This result is critical for understanding statistical downscaling and helps explain its success when applied to SAT since it is now clear how spatial variability is related to variability at a point.

The extreme characteristics of area-averaged SAT variability, which are deducible by the skewness and kurtosis statistics, are dependent only upon characteristic length scales related to spatial pair-correlations, and take either a linear (for skewness) or quadratic (for kurtosis) functional form. Taken to a more abstract level, this may indicate that explicit mappings could be made between the gridded variability of numerical simulations and the unobserved and intrinsic subgrid-scale variability smoothed over by the model simulation, particularly if characteristic

length scales for model grid cells can be deduced either from models or observations. Additionally, the correlation terms presented in Section 4.4 could be used as evaluation criteria or constraints in climate model development. Future research on this topic would be of great importance toward understanding the local-scale effects of decadal variability and climate change, particularly pertaining to extremes.

4.7 Conclusions and Discussion

We have taken a bottom-up approach toward understanding the multi-scale variability of SAT in weather station observations through an examination of the first four statistical moments. We focus on observations of T_{min} during DJF and T_{max} during JJA as examples of random fields observed at points in space and instantaneously in time. We first derive the moments for weighted sums of dependent RVs, which contain both linear and nonlinear correlations up to fourth order. We then examine these (higher-order) correlations through the lens of statistical fluid mechanics and identify SAT observations as approximately locally homogeneous quasi-Gaussian random fields whose third- and fourth-order correlations are functionally related to the pair correlations of contributing stations. We also report on station pair correlation decay with distance, and illustrate the spatial distribution of pair correlation e-folding distance, globally. These results are particularly important as they provide a bridge into the relatively well-developed field of statistical mechanics and may help to elucidate useful applications to climate-related disciplines.

Changes to moment statistics when moving from point-scale to grid-scale are expected from observed (higher-order) correlations. We compare a gridded observation dataset with station data as a point of reference to examine these effects. The second moments of gridded observation data are greatly reduced globally compared to the second moments of contributing stations. Higher-order moments become closer to those of a Gaussian distribution than the contributing stations themselves. It is shown that these changes can be expected from mathematical theory, and that they are physically related to the distances between the

contributing stations.

We also examine trends in moments over a 1950-2010 regression period and show that gridding algorithms that change over time can contribute to trends in moments, which can deviate from (or even oppose) those estimated from station data directly. These results in particular have large impacts on assessments of weather risk and the study of extremes within the context climate change. As an example, wintertime trends in SAT standard deviation are in fact significantly negative over much of the world and appear to be driven by reduced variability in nighttime temperatures (which also has significantly negative trends in standard deviation). This result is consistent with climate change theory due to increased downwelling long-wave radiation for a moister atmosphere and potentially reduced baroclinicity in the tropics, however, observed trends at station-level are underestimated by gridded data. It is uncertain at this time whether these differences arise from biases due to gridding algorithms, discussed in Section 4.5.2, or if distribution trends should be fundamentally different across spatiotemporal scales.

Lastly, we discuss the general implications of this research on multi-scale SAT variability. Through a careful examination of moment statistics up to fourth-order, we are able to shed light on the SAT variability potentially observable or estimable at any spatial or temporal scale, which may have even further implications for the study of variability and extremes using numerical simulations. Provided a method for deducing characteristic spatial length scales for grid cell based model output, quantitative inference concerning the effects of climate change on temperature extremes could be greatly improved.

This article aims both to shed light on SAT variability and to demonstrate some potential pitfalls in the analysis of high-frequency climate variability in general, particularly when studying extremes. Further research is needed in identifying robust and unbiased methods for studying high-frequency climate variability and extremes, particularly in the context of climate change. For example, a recently proposed distribution-preserving gridding algorithm, the hybrid interpolation technique (*Shen et al.*, 2001), is well supported by these analyses, particularly if it is extended to include the higher moments. This study highlights that it is perhaps

particularly important that gridding algorithms be consistent both spatially and temporally in order to study the statistics of daily variability, since algorithmic choices and data aggregation have first-order effects on observed variability. An overview of current gridding methods for daily data is available in *Shen et al.* (2001); however it may be important to modify current or innovate new methods in order to optimize the study of high-frequency variability. Index-based gridded datasets for extremes, such as the recently released GHCNDEX (*Donat et al.*, 2013a), also provide a promising avenue for distributionally consistent examinations of high-frequency climate variability.

4.8 Acknowledgements

This research was supported in part by the National Science Foundation (NSF) under Grant No. AGS-1015957, AGS-1015926, AGS-1419526, OCE0960770 and OCE1419306. This chapter has been published in: **Cavanaugh, N.R.**, and Shen, S.S.P. (2014). Spatial Scaling Effects on the Statistical Moments and Their Trends of Global Daily Surface Air Temperature. *Journal of Climate*, Submitted, *sub judice.*

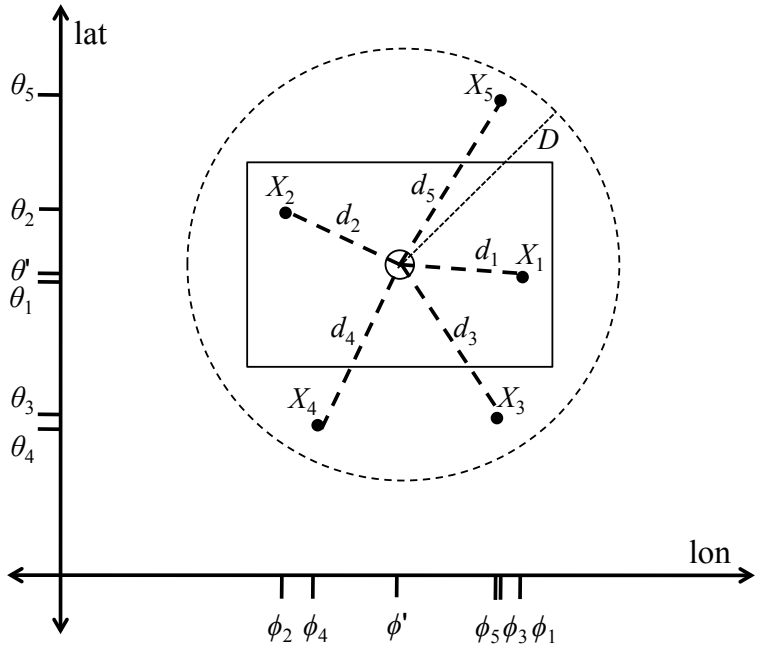

Figure 4.1: Diagrammatic depiction of the HadGHCN-D gridding algorithm. The solid rectangle indicates the grid cell being interpolated to, with the small solid circle indicating the grid cell center whose longitude and latitude are ϕ' and θ'. The large dashed circle indicates the furthest extent of the search radius D, inside of which stations may be used for interpolation. X_i indicates the locations of weather stations being aggregated for interpolation whose distance to the grid center is $d_i \leq D$ and whose longitude and latitude are ϕ_i and θ_i.

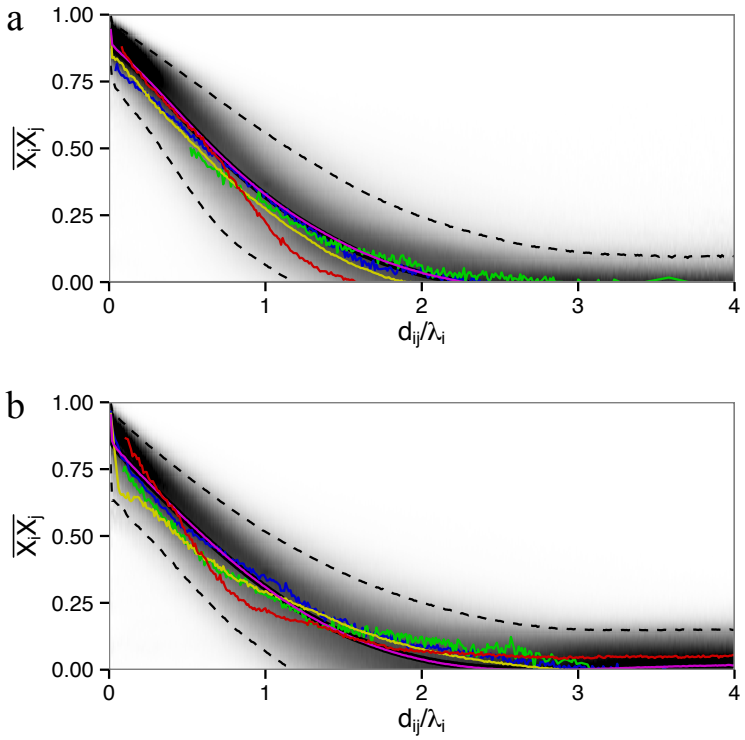

Figure 4.2: Pair correlations $\overline{X_iX_j}$ between weather station data time series X_i and X_j for (a) T_{min} during DJF and (b) T_{max} during JJA as a function of station separation distance d_i normalized by a local e-folding correlation distance λ_i given in Figure 4.3. Shading corresponds the spread of observations and can be interpreted as the uncertainty. The solid black line indicates the mean correlation. Dashed lines represent the 5% and 95% percentiles estimated from observations and can be interpreted as uncertainty bounds. Means for latitude-bands are further broken out by colors: (red) 60N - 90N, (magenta) 30N - 60N, (yellow) 0 - 30N, (green) 30S - 0, and (blue) 60S - 30S.

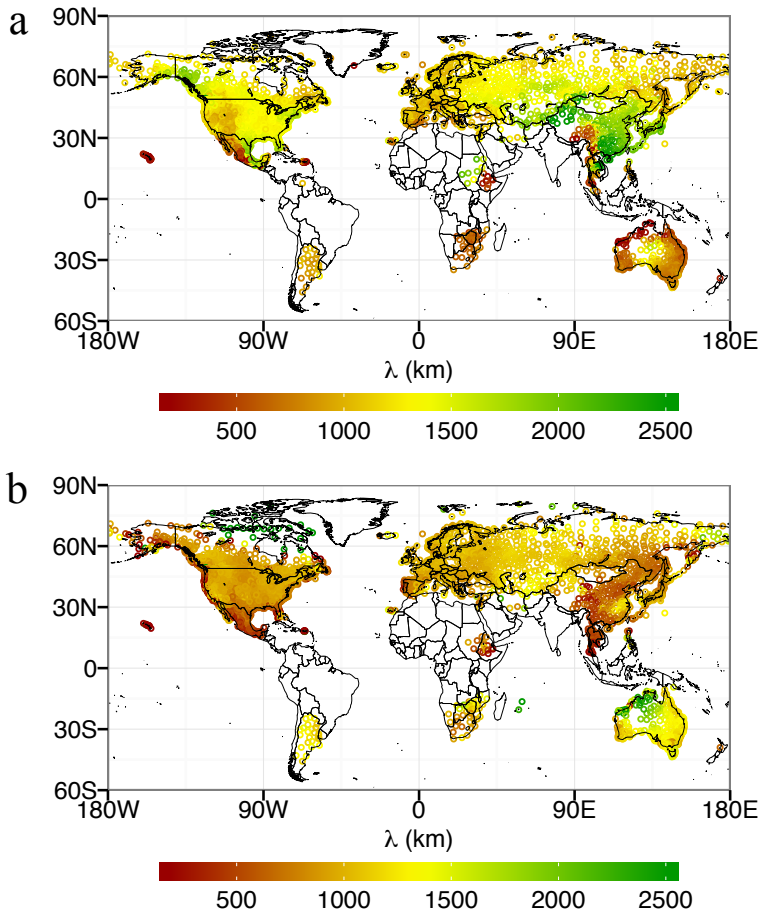

Figure 4.3: Spline interpolated local e-folding correlation distance λ for each GHCN-D station according to daily data for (a) T_{min} during DJF and (b) T_{max} during JJA. The color scale is limited to 200km and 2500km.

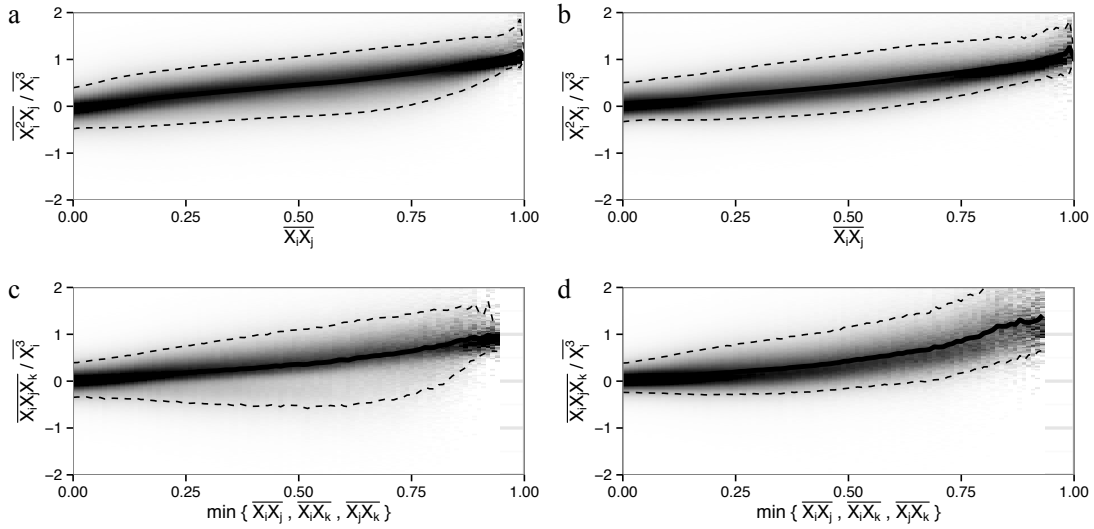

Figure 4.4: Higher-order correlation terms in coskewness matrices for T_{min} during DJF (left column) and T_{max} during JJA (right column) normalized by the third moment $\overline{X^3}$of station i plotted as functions of pair correlations. Figure shading and lines are as in Figure 4.2. Figures 4a,b document $B_{2,1}$. Figures 4c,d document $B_{1,1,1}$.

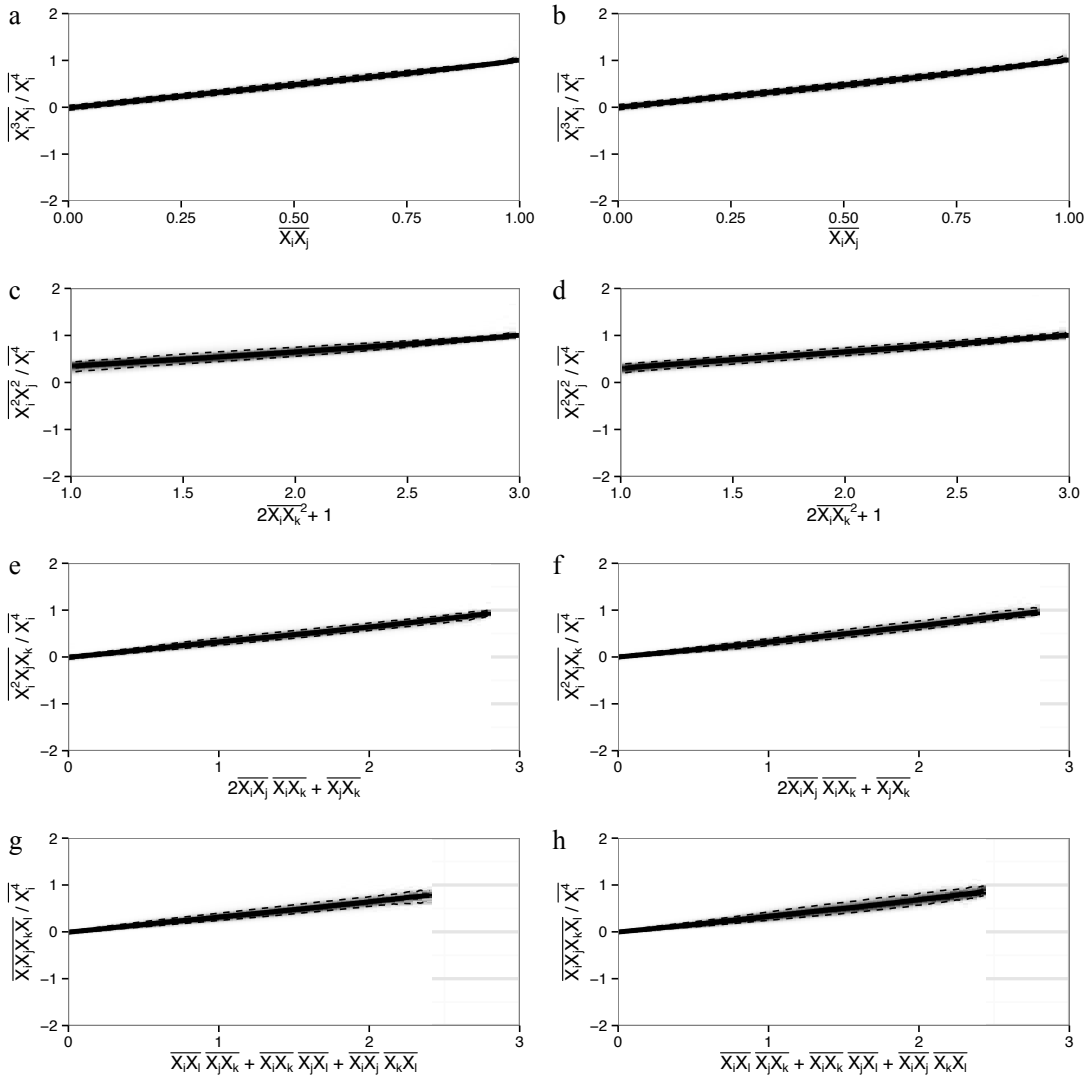

Figure 4.5: Higher-order correlation terms in cokurtosis matrices for T_{min} during DJF (left column) and T_{max} during JJA (right column) normalized by the fourth moment $\overline{X^4}$ of station i plotted as functions of pair correlations for a quasi-Gaussian field. Figure shading and lines are as in Figure 4.2. Figures 4.5a,b document $B_{3,1}$. Figures 4.5c,d document $B_{2,2}$. Figure 4.5e,f document $B_{2,1,1}$. Figures 4.5g,h document $B_{1,1,1,1}$.

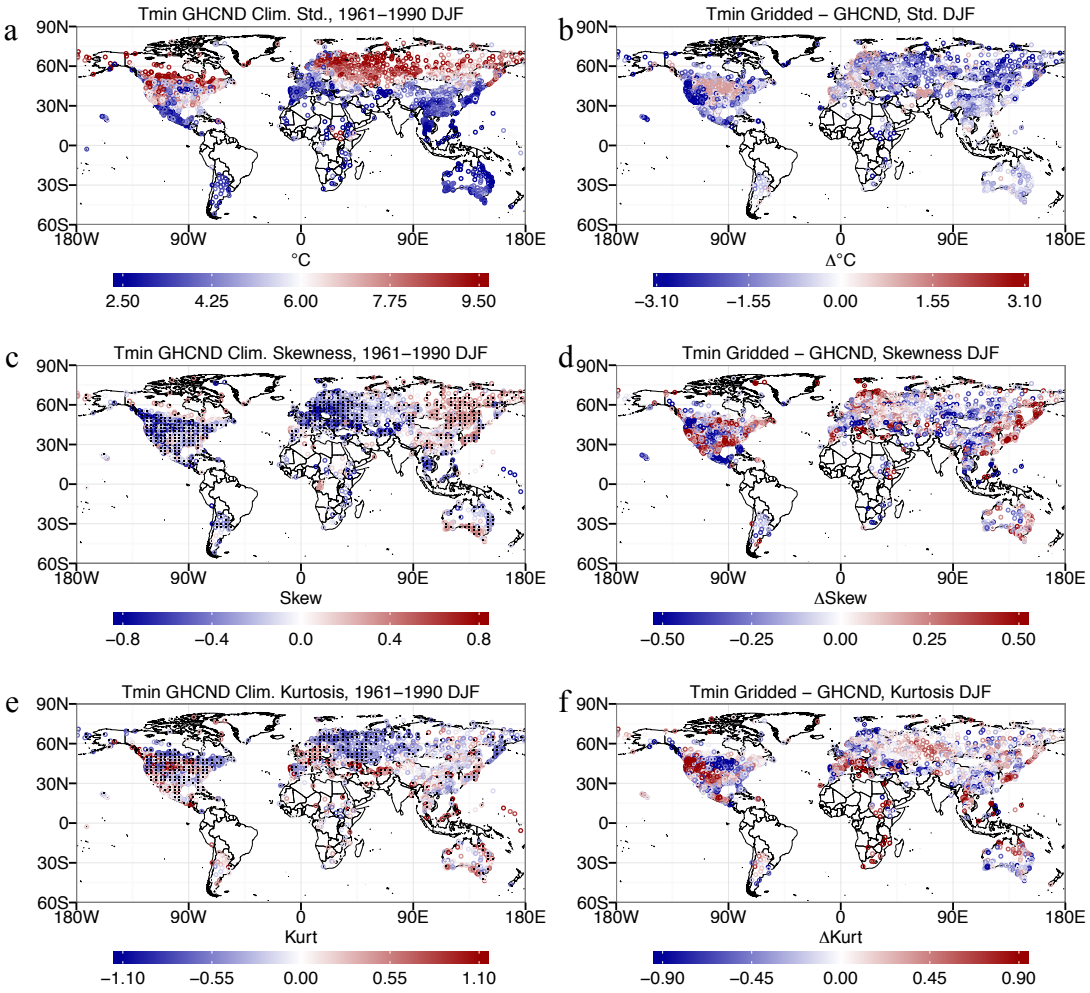

Figure 4.6: Station-level climatological moments (left column) and the bias of gridded data (right column) for T_{min} during DJF. The biases at each station are represented as the moments of the HadGHCN-D grid cell minus the moments from station data inside the same grid cell. Grid cells in c and e are marked if at least 50% of the contained stations have moments which significantly differ from 0 based on the standard error of skewness and kurtosis at a 95% confidence level (a,b) standard deviation. (c,d) skewness. (e,f) kurtosis.

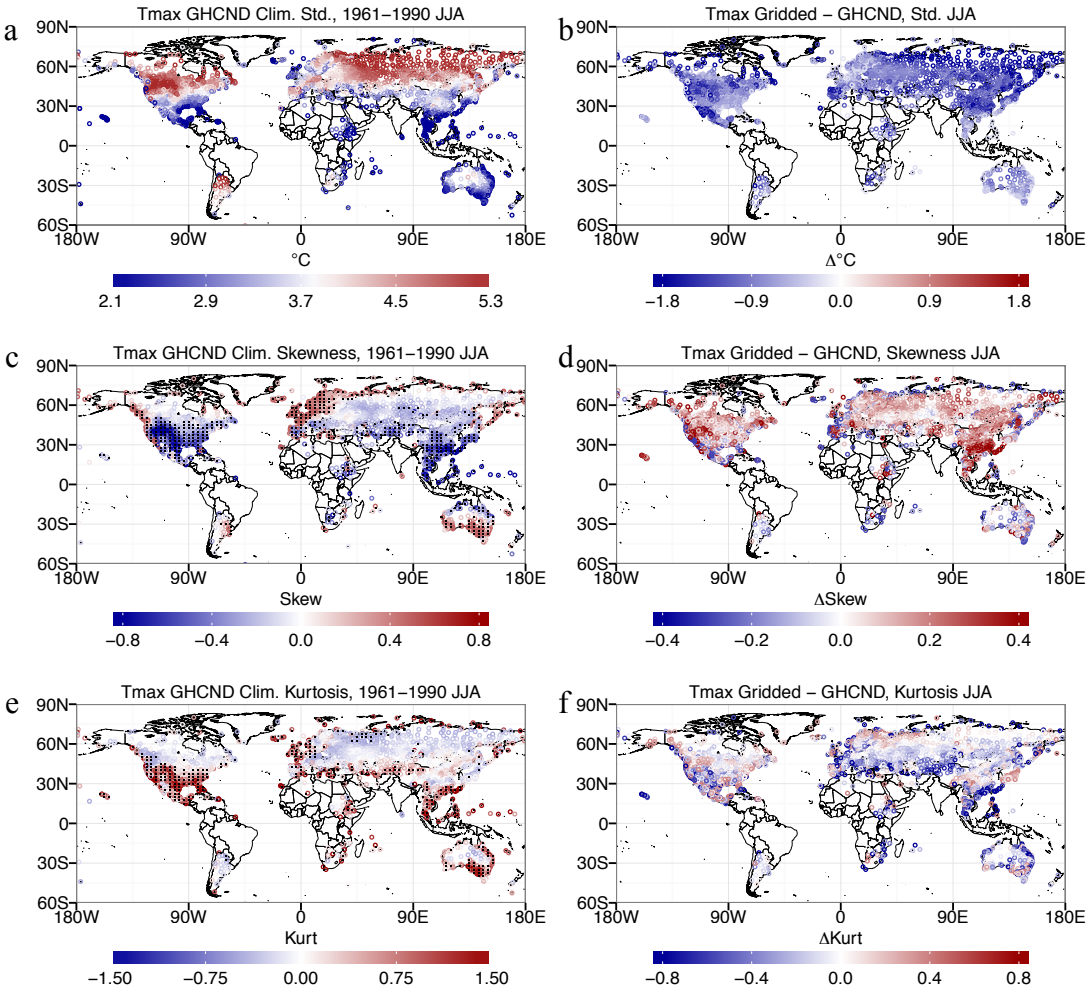

Figure 4.7: Same as Figure 4.6 for T_{max} during JJA.

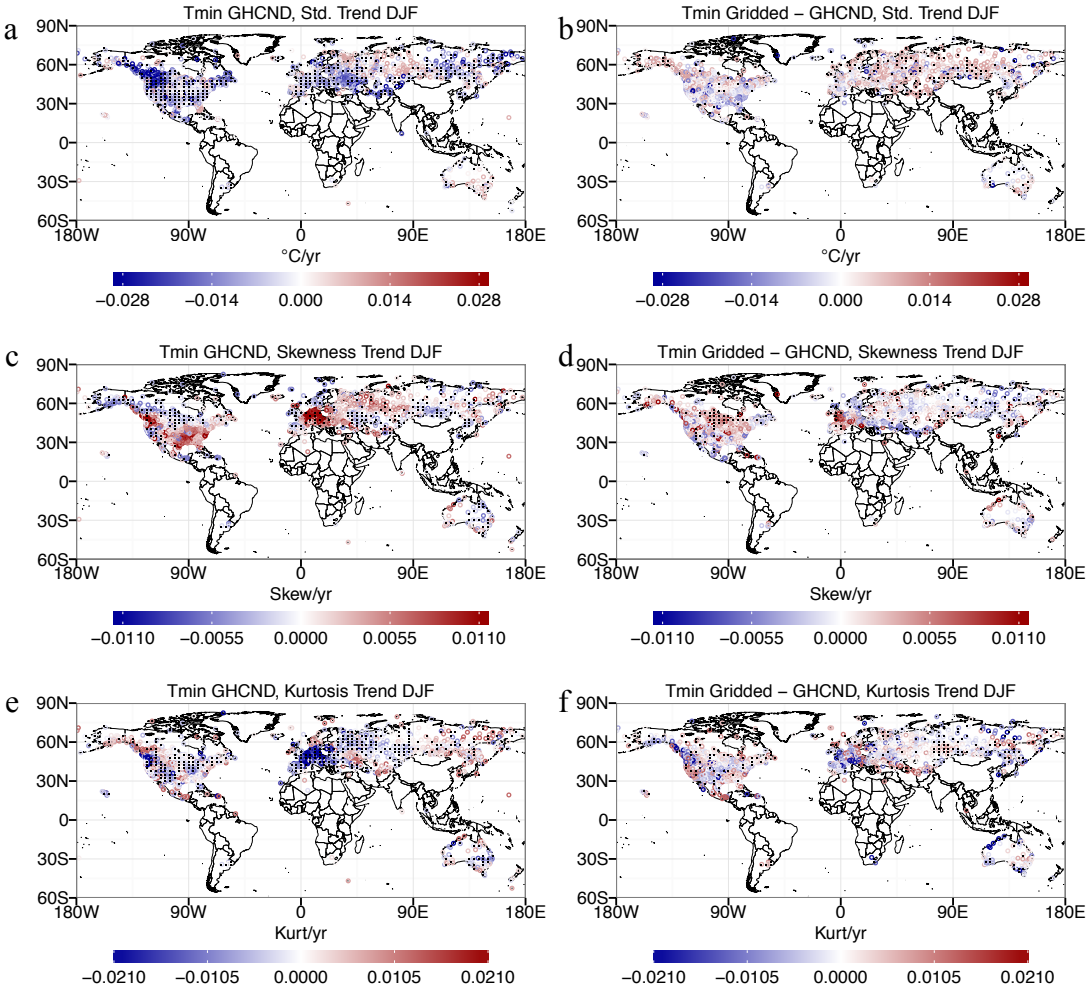

Figure 4.8: Trends in moments over the 1950-2010 time period for T_{min} during DJF computed from GHCN-D data (left column). Trends in moments for stations differenced from the trends in moments estimated from the HadGHCN-D dataset (right column). The trend biases at each station are represented as the moment trends of the HadGHCN-D grid cell minus the moment trends from station data inside the same grid cell. Grid cells for a,c,e are marked if at least 50% of the contained stations have statistically significant trends estimated against an AR1 at 95% confidence. Grid cells for b,d,f are marked if the direction of the trend for the gridded data opposes the direction of the average trend of each station within the same grid cell.

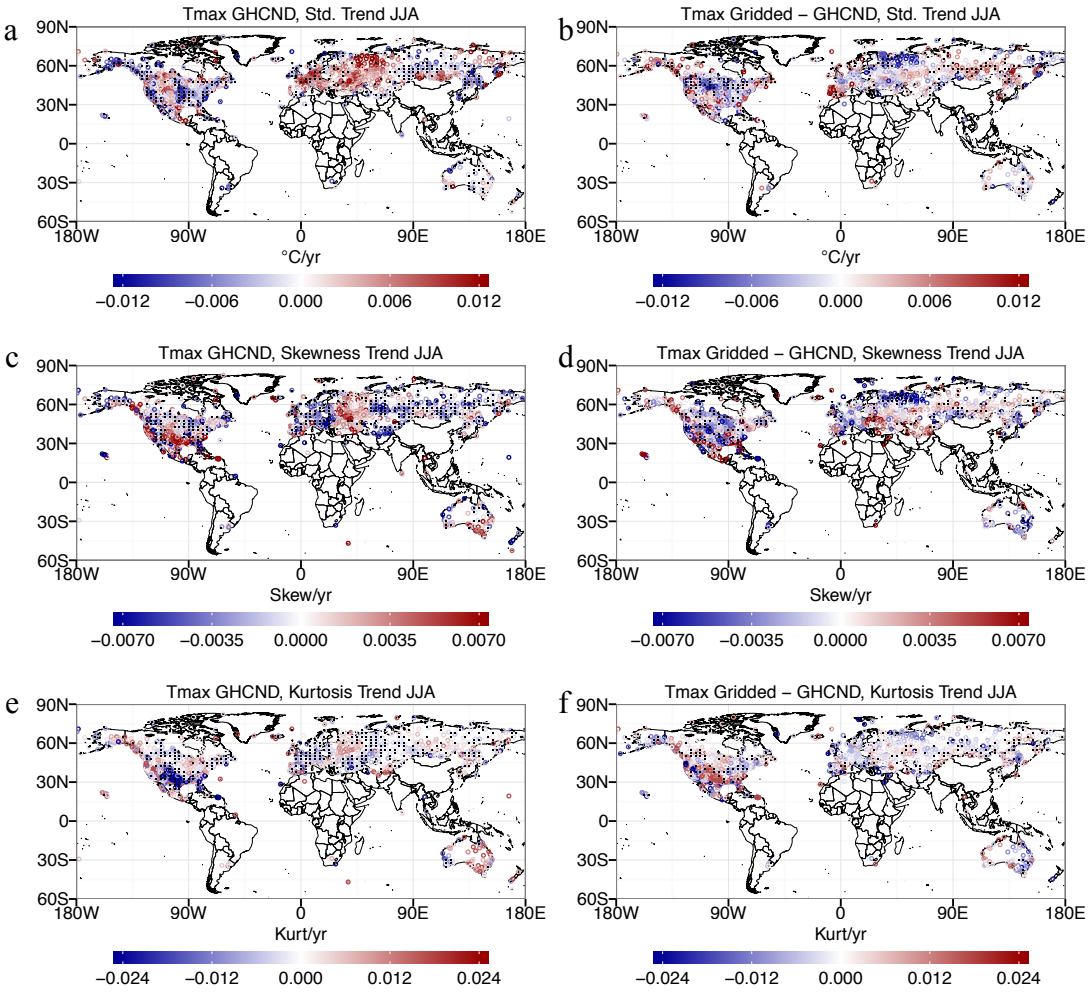

Figure 4.9: Same as Figure 4.8 for T_{max} during JJA.

Chapter 5

On the Probability Distribution of Daily Precipitation Extremes

Standard statistical approaches to modeling daily precipitation cannot account for observed volatility and typically underestimate the probability of extreme events. This can result in dangerously incorrect risk assessment for practical applications as well as obstruct climatic studies of extreme meteorological events. Stochastic theory and a unique test are used to quantify volatility of daily precipitation amounts at over 22,000 weather stations globally and to identify mathematically and physically consistent models for the probability of daily precipitation extremes. Results indicate that, at most stations, precipitation volatility can be explained by heavy-tailed probability models, but not typically by the customary distributions with exponential tails. Moreover, by examining geographical and seasonal structures in extreme precipitation behavior, the authors illustrate that the degree of volatility, as well as the fundamental probability tail structure observed at specific locations, is determined by the diversity in precipitation-producing mechanisms. These results represent progress in stochastic modeling of high-frequency precipitation extremes and a better understanding of the climatic reasons for their observed geographic behavior. They also have important practical implications for deducing the probabilities of high-magnitude precipitation events.

5.1 Introduction

A correct parametric approach to modeling the probability of high-frequency hydrologic extremes is required to appreciate weather in the context of climate variability and change as well as for accurate estimation of extreme event probabilities necessary for important practical applications. The urgency of this problem is amplified by the prospect that global hydrologic change may disproportionately manifest itself in increased frequency and intensity of extreme precipitation. This view is supported by theoretical reasoning (*Allen and Ingram*, 2002; *Trenberth et al.*, 2003; *Karl and Trenberth*, 2003) and a host of climate model projections (e.g., *Kharin et al.*, 2013; *Toreti et al.*, 2013; *Zhang et al.*, 2013b; *Chen et al.*, 2014; *Polade et al.*, 2014; *Wuebbles et al.*, 2014) as well as empirical evidence for increasing trends in the frequency of extreme daily precipitation worldwide (e.g., *Tsonis*, 1996; *Easterling and Evans*, 2000; *Groisman et al.*, 1999, 2004, 2005; *Alexander et al.*, 2006; *Donat et al.*, 2013b). It is difficult, however, to account for high-frequency precipitation extremes without a reasonable probability distribution model. Distinguishing likely distributions through extreme value analyses is becoming increasingly popular; however, large-scale attempts at distribution classification still rely on the subjective selection of parametric distributions (e.g., *Papalexiou et al.*, 2013).

A plethora of theoretical probability distribution functions (PDFs) is available for modeling any type of environmental data, including rainfall. The correct choice of PDF is not trivial. At best, it can be determined from a combination of physical reasoning and a careful examination of the connection between data and mathematics. At worst, it is dictated by convention and involves no thorough investigation of the reasons why a particular PDF should fit the data. Needless to say, the two approaches can lead to significantly different results. The latter paradigm seems to be the typical approach to modeling daily rainfall with exponentially-tailed PDFs. This approach, however, often leads to discarding extremely large events as outliers. If the extreme events are of primary importance, a more flexible approach is needed to encompass them into the body of the statistical model.

An alternative is provided by heavy-tailed distributions, such as the Pareto

(Johnson et al. 1994) or the stable laws (*Samorodnitsky and Taqqu*, 1994). The stable laws arise naturally as approximations to sums of random quantities such as total precipitation accumulated over a time period. The Pareto distribution, via the Peaks Over Threshold theory, is also a natural model for excesses over high thresholds. Alternatively, we can say that heavy-tailed distributions have power-law or fat tails. The most important, essentially qualitative, difference between the heavy- and exponentially-tailed random quantities is in the size of their large percentiles relative to the majority of the data. The heavy-tailed distributions have much larger high percentiles than the exponentially-tailed laws, which reflects their higher volatility.

Heavy-tailed behavior has been documented in precipitation and streamflow data measured at specific locations or regions for time series of extreme values, e.g. annual peak record streamflow (*Smith*, 1989; *Katz et al.*, 2002) and precipitation (*Katz et al.*, 2002), as well as threshold exceedances for streamflow (*Anderson and Meerschaert*, 1998). Heavy tails have also been suggested as reasonable models for daily precipitation amount at a large number of U.S. stations (*Smith*, 2001; *Panorska et al.*, 2007) and globally (*Papalexiou and Koutsoyiannis*, 2013). However, in climate and hydrology research, exponentially-tailed PDFs are still most-often used to model daily rainfall, including extremes (e.g., *Tsonis*, 1996; *Groisman et al.*, 1999; *Zolina et al.*, 2004; *Wilson and Toumi*, 2005; *Katz*, 2010; *Chen and Brissette*, 2014).

Below, we describe an innovative statistical test derived from probability theory developed to distinguish between heavy and exponential precipitation tails (*Panorska et al.*, 2007; *Kozubowski et al.*, 2009). Further, we apply this test to thousands of weather stations distributed globally and show with high statistical confidence that daily precipitation tails are predominantly heavy. In other words, daily precipitation extremes at a large majority of the stations are not exponential, but more closely follow a power law. This means that statistical distributions traditionally used to model daily rainfall (e.g. exponential, Weibull, Gamma, log-normal) generally underestimate the probabilities of extremes. The degree of this distortion, i.e. volatility, depends on regional and seasonal climatic peculiarities

over the globe, in agreement with North American results of *Panorska et al.* (2007). It is most severe in regions that experience precipitation from a variety of meteorological systems producing wildly different precipitation rates. However, even at many stations that experience precipitation almost exclusively from consistent synoptic sources, heavy-tailed models may be superior to exponential ones and can lead to more realistic and safer estimates of extreme event probabilities, return periods, n-year events, and design limits (*Panorska et al.*, 2007). The correct choice of the distribution function is therefore essential to safe engineering design, hazard assessment and other applications. It will also facilitate further investigations of extreme weather events and climate.

Our methodological focus is on distribution classification, rather than probability estimation. We classify the entire distribution of precipitation excesses as one of two (Pareto or exponential) possible models. Additionally, we note the theoretical connection between the distribution of the excesses and the statistical distribution of all precipitation at a given location. This broadens the impact of our results and sheds light on how the volatility of precipitation varies geographically and how extreme daily event probability structure is related to local climatology.

5.2 Methodology

We now review the methodology that was developed by *Panorska et al.* (2007) and described in theoretical detail by *Kozubowski et al.* (2009). Aiming to do a probability tail classification, we are immediately confronted by the fact that many choices exist for the specific form of the PDF with exponential and heavy tails. We consider a random variable X to be heavy tailed if $P(X > x) \sim cx - \alpha$ for large values of x, where c and α are positive constants. Note, that according to this definition, none of the distributions traditionally used to model precipitation (e.g. Gamma, lognormal, Weibull, double, stretched or extended exponential, etc.) are heavy-tailed. Examples of heavy-tailed distributions include Pareto, Cauchy, Frchet, and stable laws. To resolve the problem of distributional multiplicity, i.e. to avoid having to choose specific PDFs from the exponential and

heavy-tailed families, we apply the peak-over-threshold (POT) theory. Since our primary interest is in estimating probabilities of large (extreme) events, the POT method helps focus the search for the reasonable models and provides a rigorous mathematical foundation for the results.

The POT method involves examination of the data falling above a threshold. For any data value X and threshold u, $X[u]$ is the exceedance, that is the (conditional) value of X given that X exceeds u. We consider excesses over the threshold u, i.e. $X[u] - u$. The Balkema-de Haan-Pickands theorem (*Balkema and Haan*, 1974; *Pickands III*, 1975) provides the limiting distribution of excesses. The theorem states that when the threshold (u) increases, the distribution of the excesses ($X[u] - u$) converges to a Generalized Pareto (GP) distribution. Any GP distribution has to be one of the following three kinds: exponential, Pareto or beta. The importance of this result for practical applications is: no matter what the specific original distribution of X is, as long as the limiting distribution of excesses over threshold u exists, the excess $X[u] - u$ has (approximately) one of only three distributions. The three distributions of the excesses correspond to the tails of the original distribution of X. If X has an exponential tail, then $X[u] - u$ will have an approximately exponential distribution. If X has a heavy tail, then $X[u] - u$ will have an approximately Pareto distribution. The Beta distribution has finite support and thus it is not considered. The problem of finding the correct PDF for describing daily extreme event probabilities reduces, therefore, to classifying excesses at a particular station. We seek a decision rule, ideally a formal statistical test, for classifying the precipitation excesses over threshold into either exponential or Pareto models.

Panorska et al. (2007) approached this problem using ideas from the theory of likelihood ratio tests (*Lehmann*, 1997). Formally, the procedure seeks to test data against the null hypothesis H_0: data come from an exponential distribution with the survival function

$$S(x) = P(X > x) = \exp(-x/\sigma) \tag{5.1}$$

versus the alternative H_1: data come from a Pareto distribution with the survival

function

$$S(x) = \left(1 + \frac{x}{\sigma\alpha}\right)^{-\alpha} \tag{5.2}$$

The approach is to consider the ratio of the maxima of the likelihoods of the observed sample under the Pareto or exponential (in the numerator) and exponential (in the denominator) models. The logarithm of the likelihood ratio, the L statistic, is:

$$L = \log \left[\frac{\max\left(\sup_{\alpha>0,\sigma>0} L_{Pareto}(\vec{x}|\alpha,\sigma), \sup_{\sigma>0} L_{exp}(\vec{x}|\sigma) \right)}{\sup_{\sigma>0} L_{exp}(\vec{x}|\sigma)} \right] \tag{5.3}$$

where $\vec{x}$ is the observed sample of excesses and L_{Pareto} and L_{exp} are the likelihood functions of the sample under Pareto and exponential models, respectively. In the Pareto case, the α parameter determines the thickness of its tail and is of primary importance. The scale parameter σ is of secondary importance. As α decreases to zero, the tail of the Pareto distribution becomes heavier (larger volatility) causing the probabilities of extremes to increase. In the exponential case, σ is the scale parameter. Statistic L is scale invariant. Note that as the tail parameter α increases, the Pareto survival function converges to the exponential one, and thus that the Pareto is able to seamlessly model both power-law and exponentially-distributed tails. The distribution and appropriate critical values of L can be obtained through Monte Carlo simulation for finite sample sizes, while the limiting distribution of L consists of an atom at zero and a continuous distribution over $L > 0$. More specifics of the test and numerical procedure can be found in *Kozubowski et al.* (2009).

5.3 Data

We have computed L for a subset of over approximately 90,000 available station records of daily rainfall distributed across the globe taken from the Global Historical Climatology Network Daily (GHCN-D; *Menne et al.*, 2012) dataset. Over 22,000 best-quality stations were selected that pass both our quality control and temporal completeness standards. Each datum that has been flagged as

having failed one of the GHCN-D quality control tests (*Durre et al.*, 2010) was first removed from the dataset (although this strictness has little qualitative nor quantitative effect on the results of this paper). At least 30-years worth of data were required to be present at each station for the common observational time period: January 1, 1950 to December 31, 2013. Here, we present results based on the 90th percentile threshold, although reasonable deviations in threshold choice do not substantially affect the results in most regions. The problem of choosing the optimal threshold is recognized as a difficult and still open one (e.g., *Smith*, 1987; *Davison and Smith*, 1990; *Smith*, 1994; *Gross et al.*, 1994).

5.4 Results

Figure 5.1 shows the log likelihood ratio statistic (L, eq. (5.3)) at each station over the entire year and is color-coded according to the heaviness and type of the local precipitation tail. This result illustrates a spatially coherent dependence of tail type on geographical location with respect to climate. First, we see that the vast majority of local tails are, strictly speaking, non-exponential at the 95% confidence level (64.9% of stations in the Northern Hemisphere and 60.2% of stations in the Southern Hemisphere). We see additionally that an even larger portion of these stations (91.1% and 86.6%) are more likely power-law than exponential. The degree of departure from the exponential model depends on the peculiarities of regional climatic regimes. Precipitation at stations with a large variety of meteorological influences (i.e., frontal, thunderstorm, tropical cyclone, intense and organized convection) tends to display highly non-exponential tails (near to black in color). This occurs where extreme weather interacts with noteworthy topography (e.g., the Mexican Gulf coast, sections of Coastal Brazil, Southeast Coastal Australia, Japan), regions of hurricane and tropical cyclone landfalls (e.g., Gulf and East Coasts of the U.S., East Asian Coasts), and in the Taiga where volatility may be enhanced by a blend of precipitation types across seasons. By contrast, predominantly exponential behavior occurs where precipitation is preferably of one type (i.e., generated by similar types of systems), for example, the exclusively convective

summer precipitation along the central Mexican high plateau, the Brazilian Highlands, and the East African Highlands west of the Afromontane. Regions of almost exclusively frontal-system-generated precipitation (e.g. California) tend to display more exponential tails, although topography appears to complicate this picture. So, over the North American mountainous West, Scandinavia, the Iberian Peninsula, as well as in regions of mixed frontal and convective precipitation (France, the U.S. Southwest) tails are mixed.

A simple although rough way to categorize precipitation by frontal and convective types is to consider seasonal data separately. Figure 5.2 presents L computed for each season independently, i.e. for each season's daily excesses over the 90^{th} percentile determined from the Dec-Jan-Feb (DJF, Fig. 5.2a), Mar-Apr-May (MAM, Fig. 5.2b), Jun-Jul-Aug (JJA, Fig. 5.2c) and Sep-Oct-Nov (SON, Fig. 5.2d) seasonal data. Tail thickness is generally heavier over each hemisphere's respective summer, and ranges from 20.3% (during DJF; Fig. 5.2a) to 30.3% (during JJA; Fig. 5.2c) heavy-tailed in the Northern Hemisphere and 25.4% (during JJA; Fig. 5.2c) to 31.7% (during DJF; Fig. 5.2a) heavy-tailed in the Southern Hemisphere. Increasing (decreasing) the exceedance threshold tends to decrease (increase) the percentage of significantly heavy-tailed stations, consistent with the scaling power of the test relative to the sample size (see, *Kozubowski et al.*, 2009); compounding this effect, increasing (decreasing) the exceedance threshold also has the tendency of decreasing (increasing) the diversity of sampled precipitation types, which corroborates one of our main findings: that precipitation diversity leads to volatility. The spatial patterns for L and the fraction of Pareto-leaning stations, however, are less affected by threshold choice.

During DJF, power-law tails are observed over the Cascade Mountains, coastal Alaska, Northern Europe, and Japan, and are likely associated with variable precipitation types from orographically enhanced mid-latitude storms dominant in winter mixed with less variable stratiform precipitation (Fig. 5.2a-i). Light-to-moderate snowfall mixed with occasionally warmer wintertime precipitation modulated by descending fronts from the north contribute to heavy tails observed in the northern plains and Midwest US (Fig. 5.2a-ii), while moisture drawn in from

the Pacific and the Gulf of Mexico by the subtropical jet contributes to volatility in the southern States (Fig. 5.2b-iii). Occasionally, post-frontal high-pressure systems establish over the southern States resulting in a cold-air damming effect and a redirection of the moist subtropical jet to further south. This condition promotes strong winds and heavy precipitation along the eastern Sierra Madre Mountains in Mexico and contributes to power-law tails along the Mexican Gulf Coast during the cold season (i.e., the Nortes, *Hurd*, 1929, see Fig. 5.2b-iv). Along the South African, and Australian coasts, power-law tails most likely result from orographically enhanced precipitation driven by moist sub-tropical air contributing to spurious coastal convection (Fig. 5.2b-v).

By MAM, midlatitude storm frequency has diminished, reducing the magnitude and variability of precipitating events along western midlatitude coasts and Japan, which largely persists until SON. The beginning of tornado season is signaled by increased heavy-tails throughout the central plains in the US that continues through the warm season before diminishing in SON (Fig. 5.2b-vi). The boreal forests of Russia, the Taiga, move into their wet season characterized by a variety of rainfall producing events derived from Arctic, Atlantic, and Mediteranean fronts and local convection, yielding mostly power-law tails through to the following cold season (Fig. 5.2b-vii). Australia begins a trend toward less heavy tails resultant from decreased rainfall variability and magnitude in the cold season.

JJA brings convective precipitation to the US Northeast (Fig. 5.2c-viii) and marks the beginning of hurricane season (Fig. 5.2c-ix), resulting in increasing power-law tails for the Gulf and Atlantic Coasts which continue till the cold season, peaking in SON. Increased regional convective variability in Europe produces a popcorn-like effect in power-law tail pattern (Fig. 5.2c-x). The onset of seasonal summer monsoons in Southeast Asia is marked by extremely heavy power-law tails over much of the region (Fig. 5.2c-xi). SON brings the gradual return of DJF-like conditions. Most tropical island nations experience power-law tails year-round (marked in Fig. 5.2d-xii).

The most general comparison of Figures 5.1 and 5.2, however, indicates that daily precipitation tails are heavier when all-season data are considered compared

to individual seasons, clearly illustrating the idea that the diversity of precipitation-producing systems is essentially responsible for precipitation volatility. Moreover, all-season heavy tails tend to occur preferentially in regions where heavy tails are found in at least one of the contributing seasons. In the warm season, when rainfall is produced by a larger variety of storm types, precipitation tends to exhibit heavier tails than in the cold season when mostly frontal systems are active.

Notably, strong spatial gradients in volatility are particularly visible along or near significant topographic gradients. As discussed above, this is likely due to the greater variety of precipitation producing mechanisms along preferred slopes of major topography where interactions between orography and atmospheric circulation leads to precipitation of different intensities being produced by various circulation/storm types on one side of the mountain range, but not on the other.

5.5 Summary and Conclusions

Here, we aim to investigate the heaviness of the daily precipitation PDF tails for thousands of globally distributed stations and to examine the spatial structure of the results. We introduce and utilize the approach of *Panorska et al.* (2007) and *Kozubowski et al.* (2009) for differentiating between exponential- and heavy-tailed distributions of precipitation. The elegant part of the statistical methodology adapted here is that it allows us to simply evaluate whether the tails of an empirical distribution are exponential or heavy, regardless of the specific form of the PDF from the exponential or heavy-tailed families of distributions. So, it is not necessary to test whether the data is or is not exponential, Gamma, Gumbel, Weibull, Double (or sometimes stretched or extended) exponential or any other exponentially-tailed PDF commonly used to model daily rainfall. These are accepted or rejected under the broad umbrella of the exponentially-tailed family of distributions.

The test requires numerical optimization to compute only one parameter (scale) and we have a theoretical bound on the interval in which to perform this optimization. Thus, our search for the maximum likelihood estimates is mathematically rigorous and the numerical routines are tightly managed, which results

in robust conclusions with accompanying significance levels. Results reported here confirm that most of the precipitation exceedance probabilities follow the Pareto law and therefore, most precipitation PDF tails are power-law distributed, not exponential. Threshold exceedances at stations that may have approximately exponential tails can also be thought of as limiting, special cases of the Pareto model, i.e. the Pareto with a very large tail parameter, α. As α increases, the Pareto distribution gets close to the exponential and the L statistics may not be sensitive enough for high values of α. Such cases appear to be atypical and, in practice, the heavy tail is superior at most locations.

A visual examination of the geographic and seasonal distributions of this measure leads to inferences about regional climatic processes that give rise to spatial patterns of precipitation volatility and heavy vs. exponential tails in general. The divergence from exponential is a matter of degrees. The most closely exponential tails are found at stations where similar systems and consistent processes produce virtually all of the precipitation. This also accounts for seasonal differences in the probability tail. During DJF, when most Northern Hemisphere extratropical precipitation is produced by midlatitude cyclones, exponential tails are just about as common as heavy tails across North America and Eurasia. However, even stations that receive precipitation from almost exclusively similar systems can exhibit heavy-tailed behavior. This may be due to variations in dynamics (e.g. position, intensity, moisture content) of the circulation systems themselves and will be investigated further in a more focused geographic and specific topographic context. How far from exponential the tails are depends strongly on location. Places that receive precipitation from a variety of systems tend to exhibit severely non-exponential precipitation tails; diversity leads to volatility. Topographic diversity also plays a role. Major topography, via its interaction with atmospheric circulation, appears capable of dominating precipitation tail shape on preferred slopes of mountain ranges.

The question as to whether heavy tails can result from mixing exponentially-tailed distributions is an obvious and important one. In theory, this is not possible; at least one of the distributions must be heavy-tailed. Our consideration of the four

seasons separately is only a rough general attempt to limit storm type diversity. From this exercise, it appears that the all-season tail (Figure 5.1) tends to be heavy only where heavy tails are found in at least one of the seasons (Figure 5.2). This observation supports theory. In order to provide a more rigorous practical answer, we would need to perform the classification on precipitation amounts classified by storm type. That exercise is beyond the scope of the present work. A practical significance of this work is that when all precipitation is considered, heavy-tail models are appropriate to model the entire precipitation distribution at a great majority of stations. Applying the customary models with exponentially decaying tails will result in gross underestimation of extreme event probabilities (e.g., *Katz et al.*, 2002; *Panorska et al.*, 2007).

The present work is a rigorous attempt to simply and efficiently relate hydrologic weather extremes to climate. It is based on a physically and mathematically consistent attempt to choose a correct model for describing the entire probability distribution and extreme precipitation events in particular directly from the stochastic properties of the data. Our empirical results strongly suggest the superiority of heavy tails to the traditional exponential on the global scale. We believe that the spatial structure and coherence of our results and their solid grounding in climate and weather patterns demonstrate the benefits of our simple approach based on letting the data show us the nature of its tails. Limiting the distributional possibilities based on the stochastic properties of the data is a positive step forward because the scientific potential and practical implications of classifying precipitation probability tails into heavy or exponentially-tailed families are considerable.

The real prospects of climatic change impacting precipitation regimes (e.g., *Polade et al.*, 2014) emphasize the need for mathematical models of extremes consistent with reality and easy to apply in practice. The choice between heavy-tailed and exponentially-tailed models is of a qualitative nature. The heavy-tailed distributions have much larger high percentiles relative to the rest of the data values than the exponentially-tailed ones. This implies that in places where heavy-tailed models are appropriate, and especially if observed and modeled trends continue,

future large events may be much larger than those observed to date. We need to be prepared for such possibilities. The exponentially-tailed models of precipitation will not be able to predict very large (relative to the observed data) events, because their mathematical properties do not allow such extremes. Addressing precipitation under climate change would have to involve analyzing precipitation produced by climate models; however, it is not clear at this point that climate models realistically simulate high frequency precipitation probability tails. Before comparing observed and modeled tails, the question of whether and how spatial scale is related to tail structure must be answered. This will be our next focus of investigation.

5.6 Acknowledgements

This work was initially supported through NSF grant ATM-0236898 Modeling, variability and predictability of North American hydrologic extremes. Cavanaugh was supported in part by NSF grants OCE0960770 and OCE1419306. We would like to thank J. Fein for his personal encouragement of this work. Funding was also provided by the California Climate Change Center, sponsored by the California Energy Commission's Public Interest Energy Research Program, by the NOAA Office of Global Programs, under the California and Nevada Applications Program, and by the Department of the Interior via the Southwest Climate Science Center. This chapter has been published in: **Cavanaugh, N.R.**, Gershunov, A., Panorska, A.K., and Kozubowski, T.J. (2014). On the Probability Distribution of Daily Precipitation Extremes. *Geophysical Research Letters*, Submitted, *sub judice.*

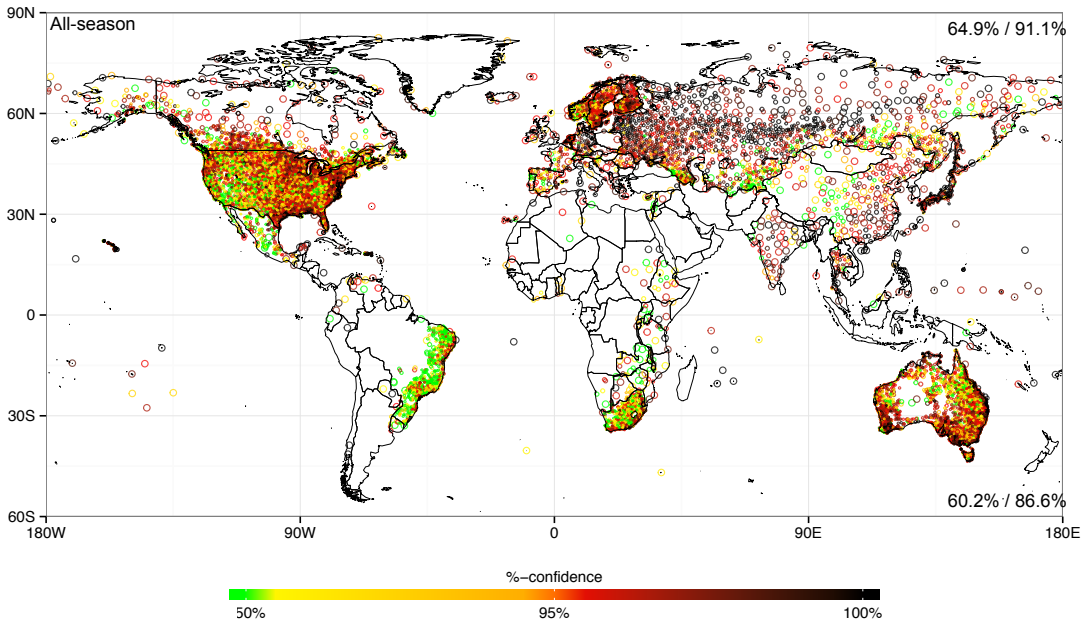

Figure 5.1: Log likelihood parameter (L) plotted geographically. Marker size is adjusted based on station density to enhance interpretability; areas of higher station density are marked by smaller points. Stations marked in green are indistinguishable from an exponential distribution. Stations marked in yellow-orange can be considered Pareto-leaning with confidence levels 50-95%. Stations marked in red-black are significantly power-law at (over) 95% confidence. Percentages in the upper- and lower-right corners mark the %-Pareto / % Pareto-leaning stations in each hemisphere, respectively.

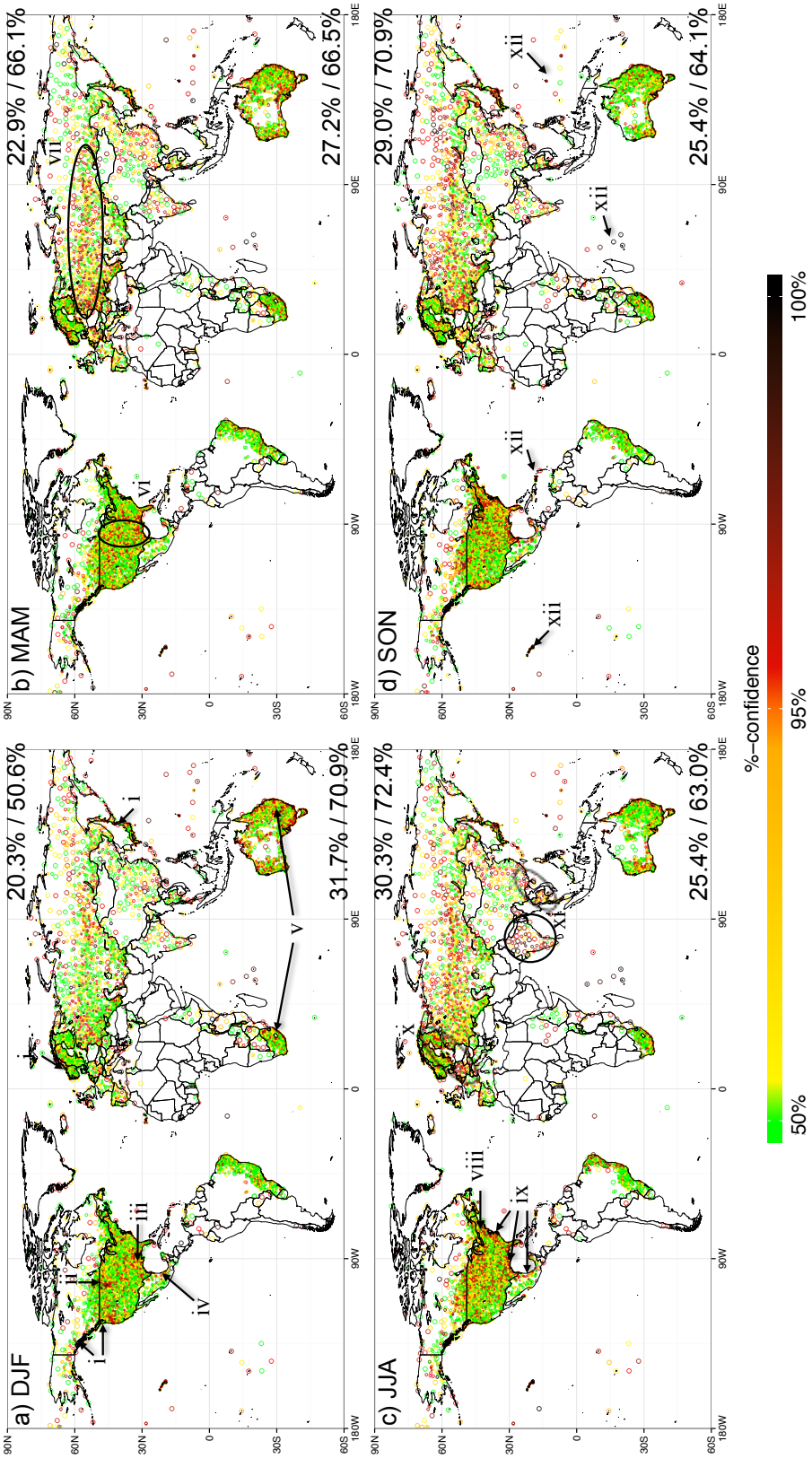

Figure 5.2: Same as Figure 5.1 however separated by season. Regional power-law features are marked by Roman numerals, discussed in text.

Chapter 6

Probabilistic Tail Dependence of Extreme Precipitation on Spatiotemporal Scale in Observations, Reanalyses, and GCMs

Daily precipitation variability as observed from weather stations is heavy-tailed at most locations around the world. It is thought that diversity in precipitation-causing weather events is fundamental in producing heavy-tailed distributions, and it arises from theory that at least one of the precipitation types contributing to a heavy-tailed climatological record must also be heavy-tailed. Precipitation is a multi-scale phenomenon with a rich spatial structure and short decorrelation length and timescales; the spatiotemporal scale at which precipitation is observed is thus an important factor when considering its statistics and extremes. In this study, we examine the spatiotemporal scaling behavior of precipitation extremes from point-scale to large grid cells and from one day to four weeks over the entire globe. We go on to validate the current generation of historically-forced climate models and reanalyses against observational data at consistent spatial scales. Our

results demonstrate that the prevalence and magnitude of heavy tails in observations decrease when moving to lower spatiotemporal resolutions, as is consistent with stochastic theory. Reanalyses and climate models generally reproduce large, synoptic scale distribution classifications, but struggle to reproduce the statistics in regions that are strongly affected by mesoscale phenomena. We discuss these results in relation to physically consistent atmospheric conditions. We conclude with a global view of precipitation distribution type at daily resolution as calculated from the best-performing renalysis, the Climate Forecast System Reanalysis.

6.1 Introduction

Weather estimated or simulated at grid-scale exhibits less variability than weather observed at point-scale, which can lead to inaccurate estimates of exceedance probabilities and n-year events. To account for this, many statistical downscaling methods have been developed with the objective to transform the data so that the variability observed at larger spatial scales matches the statistics observed by weather stations at point-scale (see *Wilks*, 2010, 2012). This can be accomplished either through the direct estimation of point-scale time series (e.g., regression-based approaches) or through the estimation of variability distributions through either stochastic (e.g., Monte Carlo simulation) or empirical (e.g., parametric downscaling, quantile-mapping) statistical methods.

Although it is generally accepted that grid-based datasets derived from observations, reanalyses, and numerical weather or climate models are meant to be representative of an areal average of the weather in that region, it is less well understood what the variability at grid-scale *should be* since it is impossible to observe weather everywhere simultaneously in an unbiased manner. If we consider observed weather variability at the surface to be a continuous random field spanning space and time, then grid-scale data should approximate the double integral of the random field over the desired spatial and temporal bounds. Since we are unable to observe the entire random field, it often suffices in practice to treat station-based observations as discrete random variables and utilize some form of weighted aver-

aging to estimate the mean (in probability) over the area in question; canonical examples of weighted average interpolation algorithms include the kriging-based optimal interpolation (*Gandin*, 1963) and the angular-distance weighting method (*Shepard*, 1968). These algorithms all take the common form

$$x'(\phi', \theta') = \sum_{i=1}^{n} w_i(\phi_i, \theta_i) x_i \; ; \quad \sum_i w_i = 1 \tag{6.1}$$

where the x represent the $i = [1, \ldots, n)$ station-based observations taken at longitudes ϕ and latitudes θ with weights w, and x', ϕ' and θ' are their grid-cell based counterparts. From eq. (6.1), it is evident that the variability at grid-scale, X' estimated by the sample space x' defined over time, is distributed as a weighted sum of the point-scale distributions X_i, estimated by the sample space of each x_i defined over time. For independent random variables, the asymptotic distribution of X' is Gaussian and is determined by the central limit theorem. For correlated random variables, the properties of X' can be determined by the higher-order cross-covariance terms of the X_i (*Cavanaugh and Shen*, 2014).

In this paper, we examine the statistics of daily precipitation. Daily precipitation has a relatively short e-folding correlation distance (about 200km, depending on location and season; *Wilby et al.*, 2003; *Hofstra and New*, 2009). As such, while the results from the central limit theorem do not hold exactly, the distribution of X' should converge to a Gaussian once the degree of spatial smoothing exceeds the decorrelation threshold. A particularly important characteristic of the Gaussian distribution is that it falls into the exponential family of distributions, that is, it is characterized by tail probabilities that decay exponentially.

As discussed in depth in previous papers (*Panorska et al.*, 2007; *Cavanaugh et al.*, 2014), the Balkema-de Haan-Pickands theorem (*Balkema and Haan*, 1974; *Pickands III*, 1975) states that sufficiently large samples from any distribution (i.e., excesses over a high threshold), can be categorized as approximately one of only three distributions: the bounded Beta distribution, the thin-tailed exponential distribution, and heavy- or fat-tailed Pareto distribution which decays as a power-law. This theorem leads to the particularly elegant result that classifying entire-sample precipitation distribution families relies only on classification of the

extreme values. *Cavanaugh et al.* (2014) utilized a novel statistical test developed by *Panorska et al.* (2007) and *Kozubowski et al.* (2009) to determine that, globally, most daily precipitation records observed from weather stations are heavy-tailed and that their tail thickness is related to regional climate factors. However, when we consider grid-scale precipitation, it is expected that at some spatial scale, the distribution of precipitation extremes should become indistinguishable from the exponential due to the effects of area averaging.

Accurately classifying the distribution of multi-day precipitation accumulations is also of practical importance in the hydrology community since stakeholders can be affected by a variety of temporal scales. In addition to spatial correlation length scales, daily precipitation is also characterized by decaying autocorrelation functions (*Wilks*, 2011) and is usually modeled using a low-order autoregressive model. When aggregating precipitation values over time, it is then expected that multi-day precipitation extremes should also become exponentially decaying after some temporal threshold has been reached; determining the spatiotemporal scaling behavior and the thresholds for the exponentiality of precipitation rates is the first goal of this study. Only beyond this threshold can the exponential-type distributions traditionally used to model precipitation be considered appropriate.

Our approach in this paper is to incrementally upscale precipitation rates observed at weather stations, first to a high-resolution $0.25^\circ \times 0.25^\circ$ global grid using area-weighted interpolation, and then to lower-resolution grids through conventional area averaging, examining the extreme value statistics at every step. Second, we incrementally test our set of station (see *Cavanaugh et al.*, 2014) and gridded data at increasing degrees of temporal smoothing, probing the extreme value behavior as a function of temporal scale as well. We present our results both spatially to increase regional interpretability and to maximize their usefulness within the climate, weather, and hydrology communities, as well as through summary statistics to enhance the clarity and interpretability of the fundamental conclusions. Our results indicate a clear spatiotemporal threshold where the majority of observations become approximately exponential; at high-resolutions, observations become 75% exponential at two-to-three week accumulations, while

at lower-resolutions, observations become 75% exponential at less than one week accumulations. Our results also suggest that at high temporal resolutions, the modal maxima in the fraction of Pareto-tailed observations varies as a function of spatial resolution; high spatial resolution observations exhibit a modal maxima with no temporal smoothing; however at lower spatial resolutions, observations show a peak at about three days.

Equipped with the data and methodology necessary to examine the probability structure of high-frequency precipitation extremes at multiple scales, we then focus on the validation of available gridded reanalyses and climate models which have complete spatial coverage over the entire globe. Seven publicly available reanalyses and 40 historically-forced climate models from the climate model intercomparison project (CMIP5; *Taylor et al.*, 2012) are examined. We validate each dataset against one of our observationally-derived datasets at a comparable spatial resolution in the regions where station data are available. Our goal is to determine if the available gridded datasets both recreate the broad spatial features of precipitation distribution type, as well as observed regional volatility at the appropriate scale. Reanalyses and climate models show a broad range in skill, with reanalyses generally outperforming climate models in both large-scale spatial structure and simulated volatility. Finally, we present a global picture of precipitation volatility derived from the best-performing global reanalysis, the Climate Forecast System Reanalysis (CFSR; *Saha et al.*, 2010).

Section 6.2 contains documentation about the station, model, and reanalysis data used in this study, as well as our methodology for interpolating station data to a high-resolution global grid. Section 6.3 briefly describes the statistical test used for distribution classification. Section 6.4 presents the observational results and dataset intercomparison. In section 6.5, we summarize and conclude.

6.2 Data and Gridding Methodology

6.2.1 Point-Scale

Weather station data for this study comes from the Global Historical Climatology Network Daily dataset (GHCN-D; *Menne et al.*, 2012), which includes approximately 90,000 quality controlled precipitation records from weather stations distributed globally. As in *Cavanaugh et al.* (2014), all data that have been flagged as having failed one of the GHCN-D quality control tests (*Durre et al.*, 2010) have been first removed from the station records; however this choice does not impact the results presented here. Over 22,000 records that include at least 30-years worth of daily data over the common observational time period of January 1, 1950 to December 31, 2013 are included in the study of point-scale precipitation variability. These same station records are also used in the creation of the gridded station datasets described in the next section.

6.2.2 Grid-Scale

We follow closely the methodology of *Hofstra and New* (2009), scaled globally using the 22,000 station records described in Section 6.2.1. This methodology uses the angular-distance weighting methodology of *Alexander et al.* (2006) and *Caesar et al.* (2006), which was adapted from the original manuscript of *Shepard* (1968). Angular-distance weighting is a popular scheme for interpolating irregularly spaced data to regularly spaced points or grid-cells and has been demonstrated to yield accurate results for daily precipitation (*Hofstra et al.*, 2008). One major benefit to using angular-distance weighting, a deterministic-type weighting scheme, for this task is that the computational overhead required is much smaller than it is for variogram-based methods (such as Kriging), which proved to be computationally intractable with available resources.

Angular-distance weighting follows the functional form of eq. (6.1). The

weights w_i take the form

$$w_i = \delta_i \left[1 + \frac{\sum_k \delta_k \left(1 - \cos(\theta_k - \theta_i)\right)}{\sum_k \delta_k} \right] \ , \quad i \neq k \tag{6.2}$$

where δ is a distance-weighting function for station i

$$\delta_k = (e^{-d_i/D})^m \tag{6.3}$$

In eq. (6.3), d_i is the distance of station i from the center of the grid-cell, D is a correlation length scale, and m is a positive constant taken here to be 4. The minimum and maximum number of contributing stations to any one grid cell at any time is 3 and 10, respectively, chosen as closest stations to the center of the grid-cell with recorded observations for that day. The length scale D is usually taken to be an e-folding correlation distance, as is done in *Hofstra and New* (2009), however here we will take D to be approximately half an e-folding correlation distance (100km) so that the minimum correlation between any two contributing stations is approximately e. We modify the methodology of *Hofstra and New* (2009) in this way since our goal is to accurately estimate and study grid-scale extremes rather than maximize the spatial coverage of the gridded dataset.

We use the angular-distance weighting methodology (eq. (6.2)) to produce a global 0.25° × 0.25° gridded dataset of daily precipitation variability, which generally includes observations where station density is moderate. From this highest-resolution gridded dataset, we upscale using simple area averaging with equal weights to grids of 0.5°, 0.75°, 1°, 1.25°, 1.5°, 2°, 2.5°, and 3°. We require all of the 0.25° subgrid-scale data to be present in order to estimate an average for a lower resolution grid-cell so that no additional interpolation is performed at each upscaling step; as a byproduct of this strict requirement, spatial coverage decreases with decreasing resolution.

6.2.3 Model-Derived

We compare and validate data taken from seven publicly available reanalyses and 40 global climate model runs archived in the CMIP5 model database; we

analyze one ensemble member for each CMIP5 model that includes precipitation rate at daily resolution as a model output. Models are forced with historically observed greenhouse gas concentrations, and we utilize the data spanning January 1, 1950 December 31, 2013. The specifics of models can be found in *Taylor et al.* (2012), as well as from model-specific references included therein. The reanalysis datasets aim to be a recreation of historical variability; for each reanalysis, we utilize its entire span of data included within the 1950-2013 observation window; the specific years included in testing differ by reanalysis. Summaries of the reanalyses and climate models used in the study are presented in Table 6.1 and Table 6.2, respectively.

6.3 Statistical Testing Methodology

In this study, we follow the methodology and testing procedure described by *Cavanaugh et al.* (2014). We present results based on exceedances of the 90th percentile of rain days in both station data and gridded data. We define rain days as those in which rainfall exceeds 0.5mm. We utilize the likelihood ratio based test of *Panorska et al.* (2007) and *Kozubowski et al.* (2009) for testing the null hypothesis H_0: data come from an exponential distribution (and are thin-tailed), versus the alternative hypothesis H_1: data comes from a Pareto distribution (and are heavy-tailed). We present results based on the L statistic

$$L = \log \left[\frac{\max \left(\sup_{\alpha>0,\sigma>0} L_{Pareto}(\vec{x}|\alpha,\sigma), \sup_{\sigma>0} L_{exp}(\vec{x}|\sigma) \right)}{\sup_{\sigma>0} L_{exp}(\vec{x}|\sigma)} \right], \tag{6.4}$$

which is simply the log of the likelihood ratio between a Pareto distribution with parameters α and σ and an exponential distribution with the single parameter σ, or zero (whichever is greater). Additional formal details and critical values can be found in *Kozubowski et al.* (2009).

In order to examine extremal dependency on temporal scale, we use a running sum over the observational time period with a window size equal to the desired temporal scale. This technique has the effect of smoothing variability and reducing

the degrees of freedom in the data. When testing smoothed data with temporal scales greater than one day, the likelihood ratio test is corrected so that the critical values based on an effective number of samples are used, calculated by dividing the number of observations by the temporal scale. This adjustment, however, has little effect on the outcome of the test, since L is scale invariant and we retain a large (effective) number of samples at all temporal scales.

6.4 Results

6.4.1 Spatiotemporal Scaling of Observations

As a baseline for the comparison of gridded daily precipitation extremes derived from station data to the extremes from stations at point-scale, we present the L statistic estimated from the $0.25° \times 0.25°$ observation-based dataset in Figure 6.1 plotted spatially. At this high-resolution, the spatial patterns in volatility match those, to a large extent, observed in station data at point-scale (*Cavanaugh et al.*, 2014, Figure 5.1). Mesoscale features become better defined, making the topographical effects on precipitation extremes even more pronounced. Many regions with local-scale heterogeneity in station classifications are smoothed out by the interpolation procedure. In addition to the conclusions about volatility derived from station data in *Cavanaugh et al.* (2014), we note that the largest geographic regions of exponentiality occur both on the windward foothills of major mountain regions (e.g., west of the Cascades, Sierra Nevada, and Rockies in North America, along coastal Norway and Spain, and in Georgia and Central Asia) where orographically enhanced frontal precipitation is the dominant regime and where mesoscale convective complexes account for the majority of precipitation (e.g., east of the Rockies, *Jiang et al.* (2006); the Brazilian Highlands, *Rasmussen and Houze* (2011); South Africa, *Blamey and Reason* (2012)). These results further confirm previous findings that exponentiality is observed only in regions that experience only one extremal precipitation type.

Next, we investigate how classification behavior responds as a function of spatiotemporal resolution. In Figure 6.2, we present 20 plots of L estimated over

North America arranged from highest spatiotemporal resolution (point-scale, daily accumulations) in the bottom left to lowest spatiotemporal resolution (2°, two week accumulations) in the top right. We choose to illustrate only North America since spatial coverage, data quality, and station density are all considered to be relatively high. Our analysis of features and trends, however, hold globally, and this choice does not effect our qualitative or quantitative results. In Figure 6.3, we summarize Figure 6.2 and map the percentage of heavy-tailed observations as a function of spatiotemporal scale.

The general trend is toward the increasing prevalence of exponential tails when moving toward lower spatiotemporal resolutions, which is consistent with the theory outlined in Section 6.1. At high temporal resolutions (i.e., daily accumulations), mesoscale patterns in L become more defined until spatial smoothing nears the decorrelation distance, which is at about a 2° resolution. At 2° spatial resolution and daily temporal resolution, the majority of grid-cells are approximately exponential (Figure 6.3). Interestingly, at this scale, many regions that can be considered exponential at higher-resolutions become heavy-tailed (e.g., in the U.S. South West over Southern California, Arizona, and New Mexico), pointing to the influence of diverse larger-scale precipitation regimes on extremes in these locations. Specific to the southwest coastal United States, this result may be attributable to the effect of atmospheric rivers that produce peak precipitation rates over relatively large areas at roughly three-day accumulations (*Ralph and Dettinger*, 2012). At high spatial resolution (i.e., point-scale or 0.25°), the degree of exponentiality monotonically increases when moving toward lower temporal resolutions (Figure 6.3). Areas that retain their heavy-tailed behavior at low temporal resolutions tend to be regions that experience extremely heavy and large-scale precipitation events, for example, tropical cyclones along the Gulf and Atlantic coasts and heavy orographic frontal precipitation in the mountains, such as from atmospheric rivers along the windward slopes of west-coast mountain ranges (*Ralph and Dettinger*, 2011). These same regions retain their heavy-tailed behavior at the lowest spatiotemporal resolutions.

A deeper inspection of Figure 6.3 indicates that at low spatial resolutions

(i.e., 1° and 2°), a modal maximum in the percentage of heavy-tailed observations exists at about the 3-day accumulation rate. With this fact in mind, an examination of Figure 6.2 reveals very heavy tails along the continental divide and across the southern and eastern states, which may be indicative of the influence of persistent synoptic conditions on high-frequency precipitation rates.

6.4.2 Model-Derived Precipitation

Armed with a suite of observation-based datasets at multiple resolutions and with spatial coverage over a broad range of locations and climatic regions globally, we seek to investigate whether patterns of precipitation volatility in gridded reanalyses and climate models match those observed in the climate record at the appropriate spatial scale. In Figure 6.4, we present summary statistics for 40 CMIP5 climate models and 7 reanalyses, again validated over North America. We match each model-derived dataset with its comparable observation-based dataset by selecting the closest resolution from among the subset of observation-based datasets with higher-resolution than the model-derived data. We restrict our comparable datasets to only those that have higher-resolution so that any bias among the model-derived data validations is consistent across datasets; in this case, we expect that all model-derived data should be slightly biased toward thin tails since we are validating against a less-smooth (higher-resolution) dataset, however this bias is likely small. We choose to compare each model-derived grid-cell with the closest observation-based grid-cell, rather than interpolating to a common grid, to minimize the effects of interpolation on precipitation variability and L statistics.

In general, most climate models and reanalyses recreate the synoptic scale features of precipitation volatility as that estimated from the observational record. Reanalyses tend to outperform CMIP5 models at recreating both the observed large-scale patterns in distribution type as well as domain averaged volatility. CFSR performs best among the reanalyses and scores a false classification rate of 33% and an average L bias of about 2.5, which is to say that CFSR produces heavier tails on average than those in the 0.25° observation-based dataset. CFSR also has the highest resolution among our reanalysis comparison, 0.31°x0.31°.

Among CMIP5 models with less than 2° resolution, CCSM4 (#8), two of the CESM1 models (#9, #11), and BCC-CSM1.1(m)(#4) all yield false classification rates of less than 40%, but are all biased toward heavy tails as well. At between 2°-3° resolution, the NASA GISS models (#23, #24) and MIROC-ESM models (#34, #35) all yield false classification rates of 35% or less, but are all biased toward light tails. HadCM3 (#25) performs best at the lowest resolution rung, although L statistic heterogeneity at 3.75° resolution is minimal since most regions are exponential. We note that volatility in model-derived data that succeeds in recreating the large-scale distribution patterns is generally biased in the same direction (either all too light or all too heavy, if there is any bias) globally.

Both climate models and reanalyses struggle to emulate the statistics, and thus the precipitation rates, observed over areas affected by mesoscale convective complexes (not shown) which act on subgrid scales and tend to produce exponential tails in observational data (Figure 6.1). The difficulty in simulating these multiscale convective systems is a noted deficiency in climate models (*Pritchard et al.*, 2011). These regions have the largest negative impact on model misclassification rates, since in general models and reanalyses otherwise tend to produce similar large-scale features to observations.

In Figure 6.5, we present a global picture of precipitation distribution classification at daily resolution computed from the best performing reanalysis, CFSR. Areas marked in white have insufficient data to reasonably estimate the extremal dependence structure, and are thus left out of the analysis. We caution explicitly that *while CFSR does tend to recreate large scale features most consistent with observations, there are noted differences and this documentation should not be taken as absolute truth.* Immediately evident from Figure 6.5 is the striking difference between the tropics and extratropics in distribution classification; the tropics in general are predominantly heavy-tailed except in instances of significant land-ocean-atmosphere interaction. To the contrary, the extratropics are predominantly exponentially-tailed, particularly over the Northern and Southern Hemisphere storm tracks. Topography clearly impacts precipitation extremes and has the effect, in most cases, of yielding exponential tails along the windward sides

of major mountain ranges; this is most visible along tropical mountain ranges in South America and Indonesia, but is also visible elsewhere. Polar regions, particularly those covered by land, appear to be almost exclusively exponential-type as well.

6.5 Summary and Conclusions

In this paper, we build on work of *Cavanaugh et al.* (2014) and *Panorska et al.* (2007) to examine how spatiotemporal scaling effects the probability tail thickness of high-frequency precipitation, and thus the classification of precipitation distributions as either exponential or power-law. We use the statistical test of *Panorska et al.* (2007) and *Kozubowski et al.* (2009) which utilizes a likelihood ratio based approach to determine whether exceedances of a high-threshold (i.e., the extremes), taken here to be 90%, can be considered as heavy-tailed Pareto-type or thin-tailed exponential-type. Results indicate that at high-spatiotemporal resolution, the bulk of observations are power-law distributed, except in regions influenced by orographically enhanced frontal precipitation or mesoscale convective complexes, which tend to be dominated by that one homogeneous precipitation type. Results also indicate that when transitioning to lower spatiotemporal resolution, the bulk of observation shift toward exponentiality, as is expected from theory. At high temporal resolutions, dependence between distribution type and spatial scale is indicated by the maximum amount of observed power-law tails shifting from 1-day accumulations at high spatial resolution to 3-day accumulations are low spatial resolutions.

We use these results to determine whether current generation climate models and reanalyses reproduce the same types of extreme statistics as are observed in comparable resolution observation-based datasets. In general, both climate models and reanalyses are skillful at recreating large-scale spatial features in distribution classification, but struggle to correctly classify regions influenced by mesoscale convective precipitation. On the other hand, climate models and reanalyses succeed in classifying precipitation regimes as heavy-tailed over most of the observation

domain, and generally recreate the patterns observed at synoptic scales, including areas of topographic interaction. Lastly, we present a global picture of precipitation distribution type estimated from the best performing reanalyses dataset, CFSR. This global summary indicates a higher prevalence of extremely heavy-tails over the tropical oceans, exponentiality along the mid-latitude storm tracks and over polar land, and mixed distribution types over landmasses, which is largely consistent with observations.

The implications of these results for the climate and hydrology communities are broad reaching. The L statistic is both a measure of tail-thickness (or volatility) of precipitation extremes, as well as a means for classifying the distribution type or family of the entire precipitation distribution, including the higher-probability events. The first perspective on L signifies that it can be used as a summary statistic across all high-level quantiles, since they are all distributionally related to each other. As such, this analysis expands on the more common method of determining certain event probabilities since it examines all extremal probabilities simultaneously. From the second perspective, the classification of distribution type can be helpful in drawing conclusions about the underlying physical mechanisms which wholly contribute to precipitation variability. In theory, at least one of the contributing precipitation regimes for a heavy-tailed observational record must also be heavy-tailed, and thus the failure of a climate model or reanalyses to correctly classify the distribution type points to a deficiency in the simulation of the contributing precipitation events. This study constitutes an initial step toward a more thorough investigation of precipitation-influencing parameterizations in climate models and their intercomparison, however, the full investigation is beyond the scope of this paper.

It is certainly a promising result that the more recent reanalyses and climate models tend to reproduce precipitation volatility more faithfully. With this fact in mind, the logical next question to ask is whether high-value precipitation volatility is stationary in time or if fundamental changes in the distributions of precipitation can be detected conditional on low-frequency climate variability or in response to changing concentrations of greenhouse gasses in the atmosphere. The answers to

these questions could profoundly impact our understanding of hydrologic extremes and could help in producing informed decisions about variability across forecasting timescales.

6.6 Acknowledgements

We thank Mary Tyree for archiving CMIP5 data sets and making them available locally. This work contributes to research supported by DOI via the Southwest Climate Science Center, by NOAA via the RISA program through the California and Nevada Applications Center, and by the California Energy Commission PIER Program. We acknowledge the World Climate Research Programmes Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups who produced the models listed in supplementary paragraph S1 for making available their model output. For CMIP the U.S. Department of Energys Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. Cavanaugh was supported in part by NSF grants OCE0960770 and OCE1419306. This chapter has been published in: **Cavanaugh, N.R.**, and Gershunov, A. (2014). Probabilistic Tail Dependence of Precipitation on Spatiotemporal Scale in Observations, Reanalyses, and GCMs. *Climate Dynamics*, Submitted, *sub judice.*

Table 6.1: Reanalysis datasets used in precipitation inter-comparison.

#	**Name**	**Lon. Res.**	**Lat. Res.**	**t. Span**	**Reference**
1	NCEP CFSR	0.31	0.31	1979-2010	*Saha et al.* (2010)
2	ECMWF ERA-i	0.70	0.70	1979-2013	*Dee et al.* (2011)
3	ECMWF ERA-40	1.13	1.12	1958-2001	*Uppala et al.* (2005)
4	JRA-55	0.56	0.56	1958-2013	*Ebita et al.* (2011)
5	NCEP/NCAR R1	1.88	1.91	1950-2013	*Kalnay et al.* (1996)
6	NCEP2/DOE R2	1.88	1.91	1979-2013	*Kanamitsu et al.* (2002)
7	NASA MERRA	0.66	0.50	1979-2013	*Rienecker et al.* (2011)

Table 6.2: Historically-forced CMIP5 models used in precipitation intercomparison.

#	Name	Center	Lon. Res.	Lat. Res.
1	ACCESS1.0	CSIRO-BOM	1.88	1.25
2	ACCESS1.3	CSIRO-BOM	1.88	1.25
3	BCC-CSM1.1	BCC	2.81	2.77
4	BCC-CSM1.1(m)	BCC	1.13	1.11
5	BNU-ESM	GCESS	2.81	2.77
6	CanCM4	CCCMA	2.81	2.77
7	CanESM2	CCCMA	2.81	2.77
8	CCSM4	NCAR	1.25	0.94
9	CESM1(BGC)	NSF-DOE-NCAR	1.25	0.94
10	CESM1(CAM5)	NSF-DOE-NCAR	1.25	0.94
11	CESM1(FASTCHEM)	NSF-DOE-NCAR	1.25	0.94
12	CMCC-CESM	CMCC	3.75	3.68
13	CMCC-CM	CMCC	0.75	0.74
14	CMCC-CMS	CMCC	1.88	1.85
15	CNRM-CM5	CNRM-CERFACS	1.41	1.39
16	CSIRO-Mk3.6.0	CSIRO-QCCCE	1.88	1.85
17	EC-EARTH	EC-EARTH	1.13	1.11
18	FGOALS-g2	LASG-CESS	2.81	2.79
19	FGOALS-s2	LASG-CESS	2.81	1.65
20	GFDL-CM3	NOAA GFDL	2.50	2.00
21	GFDL-ESM2G	NOAA GFDL	2.50	1.52
22	GFDL-ESM2M	NOAA GFDL	2.50	1.52
23	GISS-E2-H	NASA GISS	2.50	2.00
24	GISS-E2-R	NASA GISS	2.50	2.00
25	HadCM3	MOHC	3.75	2.50
26	HadGEM2-AO	NIMR/KMA	1.88	1.25
27	HadGEM2-CC	MOHC	1.88	1.25
28	HadGEM2-ES	MOHC	1.88	1.25
29	INM-CM4	INM	2.00	1.50
30	IPSL-CM5A-LR	IPSL	3.75	1.89
31	IPSL-CM5A-MR	IPSL	2.50	1.27
32	IPSL-CM5B-LR	IPSL	3.75	1.89
33	MIROC5	MIROC	1.41	1.39
34	MIROC-ESM	MIROC	2.81	2.77
35	MIROC-ESM-CHEM	MIROC	2.81	2.77
36	MPI-ESM-LR	MPI-M	1.88	1.85
37	MPI-ESM-MR	MPI-M	1.88	1.85
38	MPI-ESM-P	MPI-M	1.88	1.85
39	MRI-CGCM3	MRI	1.13	1.11
40	NorESM1-M	NCC	2.50	1.89

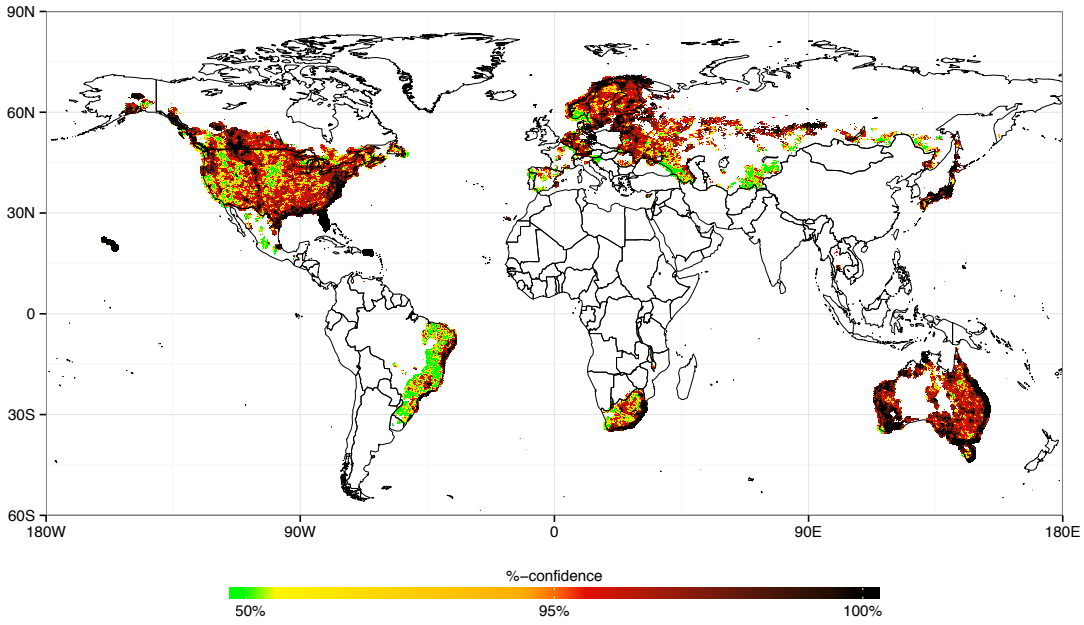

Figure 6.1: Log likelihood parameter (L) plotted geographically for the 0.25°x0.25° dataset. Grid-cells marked in green are indistinguishable from an exponential distribution. Grid-cells marked in yellow-orange can be considered Pareto-leaning with confidence levels 50-95%. Grid-cells marked in red-black are significantly power-law at (over) 95% confidence.

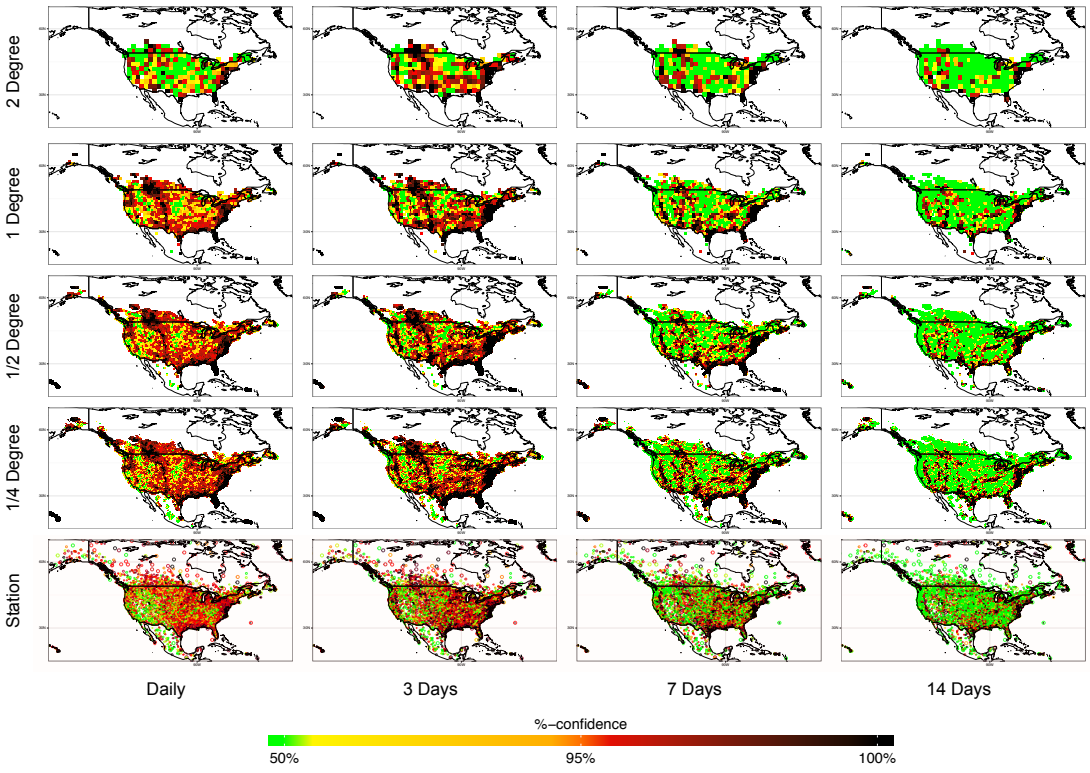

Figure 6.2: Log likelihood parameter (L) plotted geographically for multiple spatial resolutions and temporal scales over North America. Other geographical regions exhibit similar scaling behavior. Colors as in Figure 6.1.

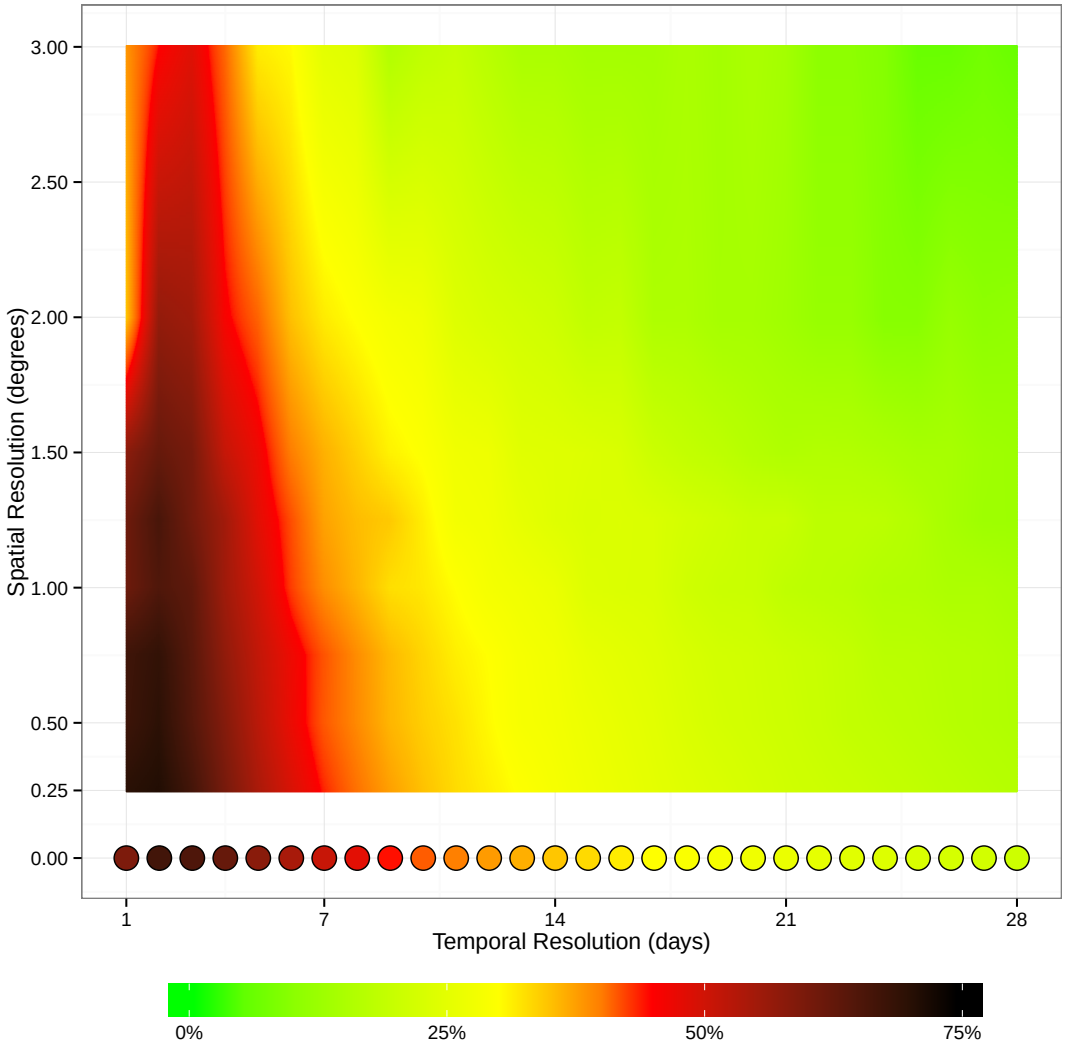

Figure 6.3: Percentage of observations over North America (as shown in Figure 6.2) that reject exponentiality at over 95% confidence plotted as a function of spatial and temporal resolution. The percentage of stations are plotted along the x-axis and marked as 0° spatial resolution.

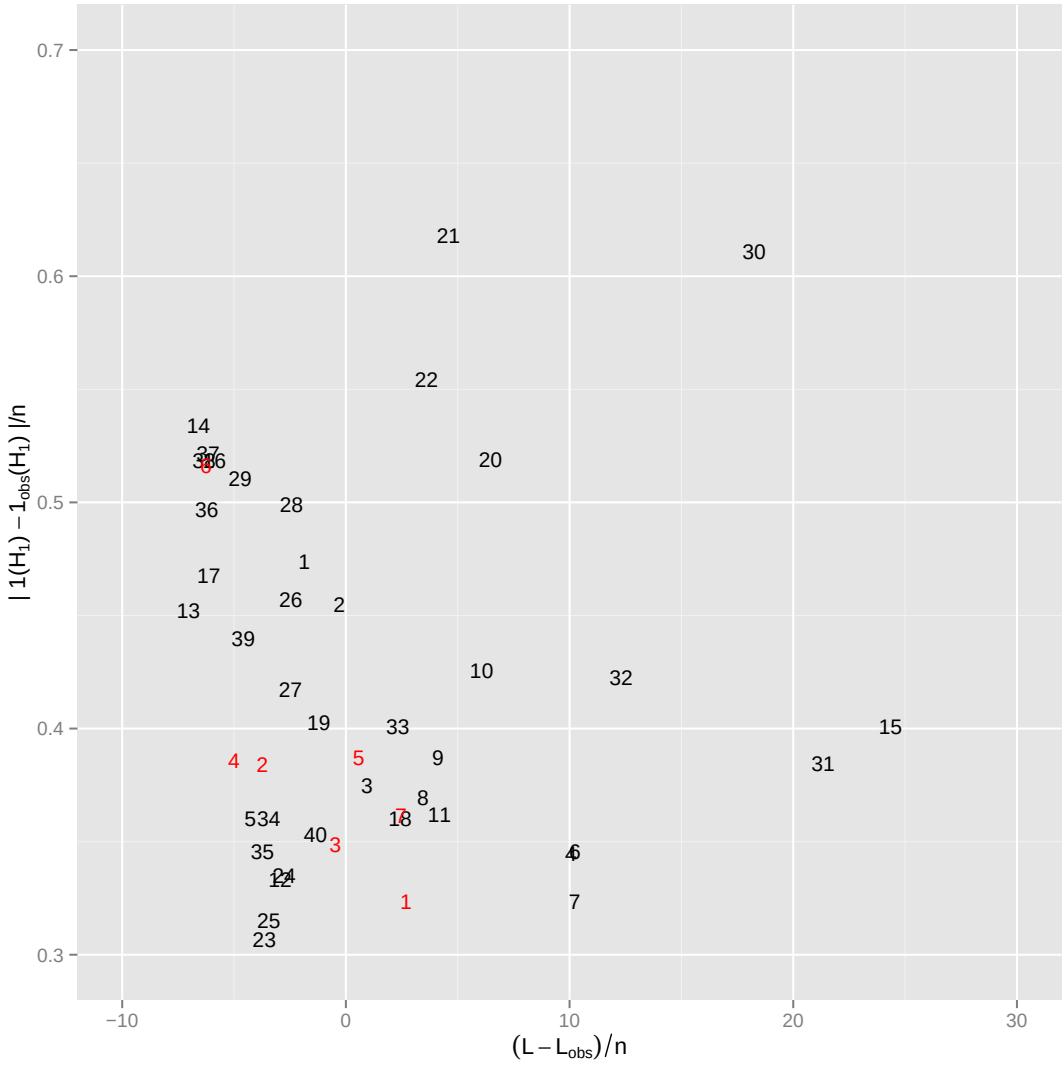

Figure 6.4: Performance of each reanalysis and CMIP5 climate model relative to the closest comparable gridded observation dataset. CMIP5 models are plotted in black. Reanalyses are plotted in red. Numbers correspond to datasets designated in Tables 6.1 (reanalyses) and 6.2 (models). The x-axis is the average grid-cell deviation in L from the model or reanalysis compared to the equivalent grid-cells in the comparable observation-based dataset and can be considered as a relative performance measure of degree of tail thickness (volatility). The y-axis is an averaged binary measure where $\mathbf{1}(H_1)$ is an indicator function, which is true if exponentiality is rejected at 95% confidence and false otherwise; the metric takes the value 1 if the datasets agree and 0 if not. This measure is equivalent to a false classification rate and can be considered as a relative performance measure of the large-scale spatial structure of observed power-law tails.

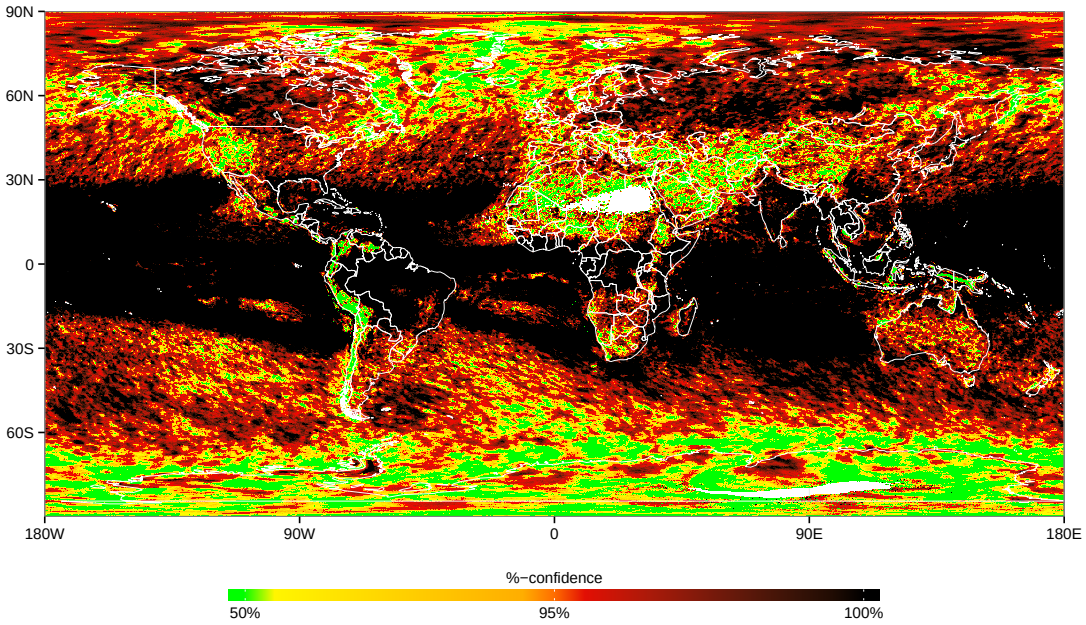

Figure 6.5: Log likelihood parameter (L) plotted geographically for CFSR (reanalysis #1). Colors as in Figure 6.1.

Chapter 7

Concluding Remarks

This dissertation constitutes an extensive examination of weather and climate statistics through the lens of stochastic dynamical systems. The results of this work can be applied in a theoretical setting toward understanding how high-frequency climate variability manifests itself within the scope of the entire climate system, but also in a practical setting for model validation and by those seeking physically consistent methods of determining the probabilities of weather events at some time in the future. This section summarizes the results of the dissertation and their implications for theoretical and practical science moving forward.

In Chapter 2, we considered the forecasting problem of Section 1.1.2. We utilized the simplest version of a reduced stochastic dynamic climate model (eq. 1.2), the Linear Inverse Model (LIM; eq. 2.1), to examine the structure of dynamic and stochastic variability and make forecasts of the Madden Julian Oscillation. Our results suggest that at grid-scale and daily resolution, the dynamics contributing to predictability in the atmosphere are predominantly linear, since our hindcasts proved nearly as skillful as fully nonlinear numerical models in prediction. This result has since been validated by other works for the tropics specific to the prediction of the Madden Julian Oscillation (*Kondrashov et al.*, 2013) which concluded that the quadratic prediction terms (the last term of eq. 1.2b) are negligible compared to the linear terms (eq. 1.2a). Linearity may also hold for the extratropics based on previous studies (*Winkler et al.*, 2001; *Newman et al.*, 2003; *Pegion and Sardeshmukh*, 2011).

In Chapter 3, we provide an in depth study of the moments of daily surface air temperature (SAT) as recorded by weather stations at point-scale. The statistics of SAT variability are significantly non-Gaussian and follow the skewness-kurtosis inequality predicted through moment closure for a linear stochastic system with correlated additive and multiplicative (CAM) noise of the form in eq. (1.2d). This result suggests that the coefficients of eq. (1.2d), and simultaneously the cubic terms of eq. (1.2c), are non-zero. The cubic coefficients manifest themselves as additional damping on the system (*Majda et al.*, 2009); however, they are likely small compared to the coefficients in eq. (1.2a) for daily resolutions, since the statistics of the LIM in Chapter 2 are (usually) already stable and yield skillful predictions and recreations of climatological variability.

SAT statistics show spatially coherent climatological moment patterns consistent with our physical understanding of the atmosphere. In particular, anomaly distributions are nearly identically distributed over synoptic-scale regions, implicating the dominant role of large-scale dynamics in the modulation of weather statistics on smaller spatial scales. SAT moment statistics also show significant and spatially coherent trends over the last 60-years which imply that low-frequency variability and/or climate change has had a measurable and significant impact on the probability distributions of daily weather events as well as the probabilities extremes. These results might be interpretable as external forcing (F_i, eq. 1.2b). Practically speaking, these results mean that deducing the probability of weather events from the climatological record requires the use of non-Gaussian distributions. The role that climate change plays in the modulation of high-frequency weather statistics has numerous open problems. Our results from this study suggest at the very least, that they are measurable and significant for temperature and precipitation.

In Chapter 4, we investigate the spatial scaling of SAT variability. We determine that SAT constitutes an approximately locally homogeneous quasi-Gaussian field with fourth-order moment closure. We also determine that the moments of SAT anomalies at scale can be determined by higher-order correlation terms, which in turn are related to pair correlation length scales that vary geographically.

These results form the quantitative backbone of statistical downscaling and have practical implications for determining the likelihoods of events at scale.

In Chapters 5 and 6, we examine the tail structure of multi-scale precipitation extremes. We determine that high-frequency precipitation rates more closely follow a power-law than an exponential-law, and thus that the underlying physically consistent distributions necessary to model precipitation variability come from the Pareto family of distributions. We also investigate the fidelity with which the current generation of numerical climate models and reanalyses recreate observed precipitation volatility patterns globally. The current generation of available gridded datasets do correctly classify precipitation distribution types in most cases, except where mesoscale convection is the dominant contributor to precipitation extremes. These results suggest that the fundamental distributions of precipitation are heavy-tailed, and thus that heavy-tailed distributions must be used to model precipitation extremes in order to capture and correctly predict the probabilities of high-valued events. It is an encouraging result that the current generation of specific climate models correctly simulate precipitation extremes at the appropriate spatial scales; the next logical step is to examine the stationarity of precipitation volatility within the context of low-frequency climate variability and climate change. The outcome of that study could have very important implications for water resources management such as flood risk management and hydrological engineering.

This body of research constitutes only the first steps toward the consistent treatment of high-frequency variability in the climate system. The upper bounds on climate predictability remain uncertain, and it is still unclear whether the fidelity of numerical predictions using conventional climate models will continue to increase. The study of stochasticity in the climate system is a relatively new concept and may provide effective alternate avenues for the study of multi-scale variability and weather-climate interaction.

Chapter 8

Bibliography

Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, a. M. G., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., Tagipour, A., Rupa Kumar, K., Revadekar, J., Griffiths, G., Vincent, L., Stephenson, D. B., Burn, J., Aguilar, E., Brunet, M., Taylor, M., New, M., Zhai, P., Rusticucci, M., and Vazquez-Aguirre, J. L. (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research*, 111(D05109):1–22.

Allen, M. R. and Ingram, W. J. (2002). Constraints on future changes in climate and the hydrologic cycle. *Nature*, 419:224–32.

Anderson, P. and Meerschaert, M. (1998). Modeling river flows with heavy tails. *Water Resources Research*, 34(9):2271–2280.

André, J., Moor, G. D., Lacarrere, P., and Du Vachat, R. (1976a). Turbulence Approximation for Inhomogeneous Flows: Part I. The Clipping Approximation. *Journal of the Atmospheric Sciences*, 33:476–481.

André, J., Moor, G. D., Lacerrere, P., and Du Vachat, R. (1976b). Turbulence Approximation for Inhomogeneous Flows: Part II. The Numerical Simulation of a Penetrative Convection Experiment. *Journal of the Atmospheric Sciences*, 33:482–491.

Balkema, A. and Haan, L. D. (1974). Residual Life Time at Great Age. *The Annals of Probability*, 2(5):792–804.

Berner, J. (2005). Linking Nonlinearity and Non-Gaussianity of Planetary Wave Behavior by the Fokker–Planck Equation. *Journal of the Atmospheric Sciences*, 62(7):2098–2117.

Blamey, R. C. and Reason, C. J. C. (2012). Mesoscale Convective Complexes over Southern Africa. *Journal of Climate*, 25:753–766.

Box, G., Jenkins, G., and Reinsel, G. (1994). *Time Series Analysis Forecasting and Control*. Prentice-Hall International, Inc., 3rd edition.

Brooks, C. and Carruthers, N. (1953). *Handbook of Statistical Methods in Meteorology*. Her Majesty's Stationery Office.

Caesar, J., Alexander, L., and Vose, R. (2006). Large-scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded data set. *Journal of Geophysical Research*, 111(D05101):1–10.

Cassou, C. (2008). Intraseasonal interaction between the Madden-Julian Oscillation and the North Atlantic Oscillation. *Nature*, 455:523–527.

Cavanaugh, N. R., Gershunov, A., Panorska, A. K., and Kozubowski, T. J. (2014). On the Probability Distribution of Daily Precipitation Extremes. *Geophysical Research Letters, sub judice*.

Cavanaugh, N. R. and Shen, S. S. P. (2014). Northern Hemisphere Climatology and Trends of Statistical Moments Documented from GHCN-Daily Surface Air Temperature Station Data from 1950 to 2010. *Journal of Climate*, 27:5396–5410.

Cayan, D., Kammeridiener, S., Dettinger, M., Caprio, J., and Peterson, D. (2001). Changes in the Onset of Spring in the Western United States. *Bulletin of the American Meteorological Society*, 82(3):399–415.

Chen, H., Sun, J., and Chen, X. (2014). Projection and uncertainty analysis of global precipitation-related extremes using CMIP5 models. *International Journal of Climatology*, 34:2730–2748.

Chen, J. and Brissette, F. (2014). Stochastic generation of daily precipitation amounts: review and evaluation of different models. *Climate Research*, 59:189–206.

Christensen, N., Wood, A., Voisin, N., Lettenmaier, D., and Palmer, R. (2004). The Effects of Climate Change on the Hydrology and Water Resources of the Colorado River basin. *Climatic Change*, 62:337–363.

Dai, A. (2006). Recent Climatology, Variability, and Trends in Global Surface Humidity. *Journal of Climate*, 19:3589–3606.

Davison, A. and Smith, R. (1990). Models for Exceedances over High Thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(3):393–442.

Dee, D. P., Uppala, S. M., Simmons, a. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. a., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, a. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes,

M., Geer, a. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kå llberg, P., Köhler, M., Matricardi, M., McNally, a. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137:553–597.

Dole, R., Hoerling, M., Perlwitz, J., Eischeid, J., Pegion, P., Zhang, T., Quan, X.-W., Xu, T., and Murray, D. (2011). Was there a basis for anticipating the 2010 Russian heat wave? *Geophysical Research Letters*, 38(L06702):1–5.

Donat, M., Alexander, L., Yang, H., Durre, I., Vose, R., and Caesar, J. (2013a). Global Land-Based Datasets for Monitoring Climatic Extremes. *Bulletin of the American Meteorological Society*, 94:997–1006.

Donat, M. G. and Alexander, L. V. (2012). The shifting probability distribution of global daytime and night-time temperatures. *Geophysical Research Letters*, 39(L14707):1–5.

Donat, M. G., Alexander, L. V., Yang, H., Durre, I., Vose, R., Dunn, R. J. H., Willett, K. M., Aguilar, E., Brunet, M., Caesar, J., Hewitson, B., Jack, C., Klein Tank, a. M. G., Kruger, a. C., Marengo, J., Peterson, T. C., Renom, M., Oria Rojas, C., Rusticucci, M., Salinger, J., Elrayah, a. S., Sekele, S. S., Srivastava, a. K., Trewin, B., Villarroel, C., Vincent, L. a., Zhai, P., Zhang, X., and Kitching, S. (2013b). Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *Journal of Geophysical Research: Atmospheres*, 118:2098–2118.

Durre, I., Menne, M. J., Gleason, B. E., Houston, T. G., and Vose, R. S. (2010). Comprehensive Automated Quality Assurance of Daily Surface Observations. *Journal of Applied Meteorology and Climatology*, 49:1615–1633.

Durre, I., Wallace, J., and Lettenmaier, D. (2000). Dependence of Extreme Daily Maximum Temperatures on Antecedent Soil Moisture in the Contiguous United States during Summer. *Journal of climate*, 13:2641–2651.

Easterling, D. and Evans, J. (2000). Observed Variability and Trends in Extreme Climate Events: A Brief Review*. *Bulletin of the American Meteorological Society*, 81:417–425.

Ebita, A., Kobayashi, S., Ota, Y., Moriya, M., Kumabe, R., Onogi, K., Harada, Y., Yasui, S., Miyaoka, K., Takahashi, K., Kamahori, H., Kobayashi, C., Endo, H., Soma, M., Oikawa, Y., and Ishimizu, T. (2011). The Japanese 55-year Reanalysis "JRA-55": An Interim Report. *SOLA*, 7:149–152.

Fall, S., Watts, A., Nielsen-Gammon, J., Jones, E., Niyogi, D., Christy, J. R., and Pielke, R. a. (2011). Analysis of the impacts of station exposure on the U.S. Historical Climatology Network temperatures and temperature trends. *Journal of Geophysical Research*, 116(D14120):1–15.

Ferranti, L., Palmer, T., Molteni, F., and Klinker, E. (1990). Tropical-Extratropical Interaction Associated with the 30-60 Day Oscillation and Its Impact on Medium and Extended Range Prediction. *Journal of the Atmospheric Sciences*, 47(18):2177–2199.

Gandin, L. S. (1963). *Objective Analysis of Geophysical Fields*. Israeli Program for Scientific Translations.

Gong, D.-Y. and Ho, C.-H. (2004). Intra-Seasonal Variability of Wintertime Temperature over East Asia. *International Journal of Climatology*, 24:131–144.

Gottschalck, J., Wheeler, M., Weickmann, K., Vitart, F., Savage, N., Lin, H., Hendon, H., Waliser, D., Sperber, K., Nakagawa, M., Prestrelo, C., Flatau, M., and Higgins, W. (2010). A Framework for Assessing Operational Madden–Julian Oscillation Forecasts: A CLIVAR MJO Working Group Project. *Bulletin of the American Meteorological Society*, 91:1247–1258.

Groisman, P., Karl, T., Easterling, D., Knight, R. W., Jamason, P. F., Hennessy, K. J., Suppiah, R., Page, C. M., Wibig, J., Fortuniak, K., Razuvaev, V. N., Douglas, A., Forland, E., and Zhai, P.-M. (1999). Changes in the Probability of Heavy Precipitation: Important Indicators of Climatic Change. *Climatic Change*, 42:243–289.

Groisman, P., Knight, R., Easterling, D. R., Karl, T. R., Hegerl, G. C., and Razuvaev, V. N. (2005). Trends in Intense Precipitation in the Climate Record. *Journal of climate*, 18:1326–1350.

Groisman, P., Knight, R., Karl, T. R., Easterling, D. R., Sun, B., and Lawrimore, J. H. (2004). Contemporary Changes of the Hydrological Cycle Over the Contiguous United States: Trends Derived From in Situ Observations. *Journal of Hydrometeorology*, 5:64–85.

Gross, J., Heckert, A., Lechner, J., and Simui, E. (1994). Novel extreme value estimation procedures: Application to extreme wind data. In *Extreme Value Theory and Applications, Proceedings. Volume 1.*, pages 225–284. Kluwer Academic Publishers.

Hansen, J. and Lebedeff, S. (1987). Global Trends of Measured Surface Air Temperature. *Journal of Geophysical Research*, 92(D11):13,345–13,372.

Hansen, J., Ruedy, R., Sato, M., and Lo, K. (2010). Global Surface Temperature Change. *Reviews of Geophysics*, 48(RG4004):1–29.

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

Hirschi, M., Seneviratne, S. I., Alexandrov, V., Boberg, F., Boroneant, C., Christensen, O. B., Formayer, H., Orlowsky, B., and Stepanek, P. (2010). Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. *Nature Geoscience*, 4:17–21.

Hofstra, N., Haylock, M., New, M., Jones, P., and Frei, C. (2008). Comparison of six methods for the interpolation of daily, European climate data. *Journal of Geophysical Research*, 113(D21110):1–19.

Hofstra, N. and New, M. (2009). Spatial variability in correlation decay distance and influence on angular-distance weighting interpolation of daily precipitation over Europe. *International Journal of Climatology*, 29:1872–1880.

Holton, J. (2004). *An Introduction to Dynamic Meteorology*, volume 1. Academic press.

Huffman, G. J., Adler, R. F., Bolvin, D. T., and Gu, G. (2009). Improving the global precipitation record: GPCP Version 2.1. *Geophysical Research Letters*, 36(L17808):1–5.

Hurd, W. E. (1929). Northers of the Gulf of Tehuantepec. *Monthly Weather Review*, pages 192–194.

IPCC (2007). Climate change 2007: the physical science basis. Technical report.

Jiang, X., Lau, N.-C., and Klein, S. a. (2006). Role of eastward propagating convection systems in the diurnal cycle and seasonal mean of summertime rainfall over the U.S. Great Plains. *Geophysical Research Letters*, 33(L19809):1–6.

Jiang, X., Waliser, D. E., Wheeler, M. C., Jones, C., Lee, M.-I., and Schubert, S. D. (2008). Assessing the Skill of an All-Season Statistical Forecast Model for the Madden–Julian Oscillation. *Monthly Weather Review*, 136:1940–1956.

Jones, C., Carvalho, L., Higgins, R. W., Waliser, D. E., and Schemm, J.-K. E. (2004). A statistical forecast model of tropical intraseasonal convective anomalies. *Journal of Climate*, 17:2078–2095.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, B., Janowiak, J., Mo, K., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D. (1996). The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*, 77(3):437–471.

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

Kang, I.-S. and Kim, H.-M. (2010). Assessment of MJO Predictability for Boreal Winter with Various Statistical and Dynamical Models. *Journal of Climate*, 23:2368–2378.

Karl, T., Knight, R., and Plummer, N. (1995). Trends in high-frequency climate variability in the twentieth century. *Nature*, 377:217–220.

Karl, T. R. and Katz, R. W. (2012). A new face for climate dice. *Proceedings of the National Academy of Sciences of the United States of America*, 109(37):14720–14721.

Karl, T. R. and Trenberth, K. E. (2003). Modern Global Climate Change. *Science (New York, N.Y.)*, 302:1719–1723.

Katz, R., Parlange, M., and Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in water resources*, 25:1287–1304.

Katz, R. W. (2010). Statistics of extremes in climate change. *Climatic Change*, 100:71–76.

Kemball-Cook, S. and Weare, B. (2001). The onset of convection in the Madden-Julian oscillation. *Journal of Climate*, 14:780–793.

Kessler, W. and Kleeman, R. (2000). Rectification of the Madden-Julian Oscillation into the ENSO cycle. *Journal of Climate*, 13:3560–3575.

Kharin, V. V., Zwiers, F. W., Zhang, X., and Wehner, M. (2013). Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, 119:345–357.

Klingaman, N. P., Woolnough, S. J., Jiang, X., Xavier, P. K., Petch, J., Caian, M., Hannay, C., Kim, D., Ma, H.-y., Merryfield, W. J., Miyakawa, T., Pritchard, M., Vitart, F., Ridout, J. A., and Wang, H. (2014). Vertical structure and diabatic processes of the Madden – Julian oscillation : Linking hindcast fidelity to simulated diabatic heating and moistening. *Journal of Geophysical Research, sub judice*.

Kondrashov, D., Chekroun, M. D., Robertson, a. W., and Ghil, M. (2013). Low-order stochastic model and "past-noise forecasting" of the Madden-Julian Oscillation. *Geophysical Research Letters*, 40:5305–5310.

Kozubowski, T. J., Panorska, A. K., Qeadan, F., Gershunov, A., and Rominger, D. (2009). Testing Exponentiality Versus Pareto Distribution via Likelihood Ratio. *Communications in Statistics - Simulation and Computation*, 38:118–139.

Lau, K. and Chan, P. (1985). Aspects of the 40-50 Day Oscillation during the Northern Winter as Inferred from Outgoing Longwave Radiation. *Monthly Weather Review*, 113:1889–1909.

Lau, K. and Chan, P. (1988). Intraseasonal and Interannual Variation of the Tropical Convection: A Possible Link between the 40-50 Day Oscillatio and ENSO? *Journal of the Atmospheric Sciences*, 45(3):506–521.

Lehmann, E. (1997). *Testing Statistics Hypotheses, 2nd edition*. Springer.

Liebmann, B. and Smiith, C. (1996). Description of a Complete (Interpolated) Outgoing Longwave Radiation Dataset. *Bulletin of the American Meteorological Society*, 77(6):1275–1277.

Lin, H., Brunet, G., and Derome, J. (2008). Forecast Skill of the Madden–Julian Oscillation in Two Canadian Atmospheric Models. *Monthly Weather Review*, 136:4130–4149.

Lo, F. and Hendon, H. (2000). Empirical Extended-Range Prediction of the Madden-Julian Oscillation. *Monthly Weather Review*, 128:2528–2543.

Loikith, P., Lintner, B., Kim, J., Lee, H., Neelin, J. D., and Waliser, D. E. (2013). Classifying reanalysis surface temperature probability density functions (PDFs) over North America with cluster analysis. *Geophysical Research Letters*, 40:3710–3714.

Loikith, P. C. and Broccoli, A. J. (2012). Characteristics of Observed Atmospheric Circulation Patterns Associated with Temperature Extremes over North America. *Journal of Climate*, 25:7266–7281.

Lorenz, E. (1963). Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, 20:130–141.

Lorenz, E. N. (1991). Dimension of weather and climate attractors. *Nature*, 353(6341):241–244.

Madden, R. and Julian, P. (1971). Detection of a 40-50 Day Oscillation int he Zonal Wind in the Tropical Pacific. *Journal of the Atmospheric Sciences*, 28:702–708.

Maharaj, E. a. and Wheeler, M. C. (2005). Forecasting an Index of the Madden-Oscillation. *International Journal of Climatology*, 25:1611–1618.

Majda, A. J., Franzke, C., and Crommelin, D. (2009). Normal forms for reduced stochastic climate models. *Proceedings of the National Academy of Sciences of the United States of America*, 106(10):3649–3653.

Maloney, E. D. (2000). Modulation of Hurricane Activity in the Gulf of Mexico by the Madden-Julian Oscillation. *Science*, 287:2002–2004.

Martin, E. R. and Schumacher, C. (2011). Modulation of Caribbean Precipitation by the Madden–Julian Oscillation. *Journal of Climate*, 24:813–824.

Matsueda, M. (2011). Predictability of Euro-Russian blocking in summer of 2010. *Geophysical Research Letters*, 38(L06801):1–6.

Matthews, A. J. and Meredith, M. P. (2004). Variability of Antarctic circumpolar transport and the Southern Annular Mode associated with the Madden-Julian Oscillation. *Geophysical Research Letters*, 31(L24312):1–5.

Meehl, G. a., Tebaldi, C., and Nychka, D. (2004). Changes in frost days in simulations of twentyfirst century climate. *Climate Dynamics*, 23:495–511.

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

Michaels, P., Balling, R., Vose, R., and Knappenberger, P. (1998). Analysis of trends in the variability of daily and monthly historical temperature measurements. *Climate Research*, 10:27–33.

Millionshchikov (1941). On the theory of homogeneous isotropic turbulence. *Doklady Acad. Nauk SSSR*, 32:611–614.

Monahan, A. H., Fyfe, J. C., Ambaum, M. H. P., Stephenson, D. B., and North, G. R. (2009). Empirical Orthogonal Functions: The Medium is the Message. *Journal of Climate*, 22:6501–6514.

Monin, A. and Yaglom, A. (1971). *Statistical Fluid Mechanics: Mechanics of Turbulence*. The MIT Press.

Montandon, L. M., Fall, S., Pielke, R. a., and Niyogi, D. (2011). Distribution of Landscape Types in the Global Historical Climatology Network. *Earth Interactions*, 15(6):1–24.

Newman, M., Sardeshmukh, P., Winkler, C., and Whitaker, J. S. (2003). A Study of Subseasonal Predictability. *Monthly Weather Review*, 131:1715–1732.

Newman, M., Sardeshmukh, P. D., and Penland, C. (2009). How Important Is Air–Sea Coupling in ENSO and MJO Evolution? *Journal of Climate*, 22:2958–2977.

North, G. (1984). Empirical Orthogonal Functions and Normal Modes. *Journal of the Atmospheric Sciences*, 41(5):879–887.

OrtizBevia, M. (1997). Estimation of the cyclostationary dependence in geophysical data fields. *Journal of Geophysical Research*, 102(D12):13,473–13,486.

Palmer, T. (1993). Extended-Range Atmospheric Prediction and the Lorenz Model. *Bulletin of the American Meteorological Society*, 74(1):49–65.

Panorska, A. K., Gershunov, A., and Kozubowski, T. J. (2007). From Diversity to Volatility : Probability of Daily Precipitation Extremes. In Tsonis, A. A. and Elsner, J. B., editors, *Nonlinear Dynamics in Geophysics*, pages 465–484. Springer.

Papalexiou, S. M. and Koutsoyiannis, D. (2013). Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resources Research*, 49:187–201.

Papalexiou, S. M., Koutsoyiannis, D., and Makropoulos, C. (2013). How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrology and Earth System Sciences*, 17:851–862.

Parker, D., Jones, P., Folland, C., and Bevan, A. (1994). Interdecadal changes of surface temperature since the late nineteenth century. *Journal of Geophysical Research*, 99(D7):14,373–14,399.

Pegion, K. and Sardeshmukh, P. D. (2011). Prospects for Improving Subseasonal Predictions. *Monthly Weather Review*, 139:3648–3666.

Penland, C. (1989). Random Forcing and Forecasting Using Principal Oscillation Pattern Analysis. *Monthly Weather Review*, 117:2165–2185.

Penland, C. and Magorian, T. (1993). Prediction of Nino 3 Sea Surface Temperatures Using Linear Inverse Modeling. *Journal of Climate*, 6:1067–1076.

Penland, C. and Matrosova, L. (1994). A Balance Condition for Stochastic Numerical Models with Application to the El Nino-Southern Oscillation. *Journal of Climate*, 7:1352–1372.

Penland, C. and Sardeshmukh, P. (1995). The Optimal Growth of Tropical Sea Surface Temperature Anomalies. *Journal of Climate*, 8:1999–2024.

Penland, C. and Sardeshmukh, P. D. (2012). Alternative interpretations of power-law distributions found in nature. *Chaos (Woodbury, N.Y.)*, 22(023119):1–5.

Perron, M. and Sura, P. (2013). Climatology of Non-Gaussian Atmospheric Statistics. *Journal of Climate*, 26:1063–1083.

Petoukhov, V., Eliseev, A. V., Klein, R., and Oesterle, H. (2007). On statistics of the free-troposphere synoptic component: an evaluation of skewnesses and mixed third-order moments contribution to the synoptic-scale dynamics and fluxes of heat and humidity. *Tellus*, 60A:11–31.

Pickands III, J. (1975). Statistical Inference Using Extreme Order Statistics. *the Annals of Statistics*, 3(1):119–131.

Polade, S. D., Pierce, D. W., Cayan, D. R., Gershunov, A., and Dettinger, M. D. (2014). The key role of dry days in changing regional climate and precipitation regimes. *Scientific reports*, 4:4364.

Portmann, R. W., Solomon, S., and Hegerl, G. C. (2009). Spatial and seasonal patterns in climate change, temperatures, and precipitation across the United States. *Proceedings of the National Academy of Sciences of the United States of America*, 106(18):7324–7329.

Pritchard, M. S., Moncrieff, M. W., and Somerville, R. C. J. (2011). Orogenic Propagating Precipitation Systems over the United States in a Global Climate Model with Embedded Explicit Convection. *Journal of the Atmospheric Sciences*, 68:1821–1840.

Rahmstorf, S. and Coumou, D. (2011). Increase of extreme events in a warming world. *Proceedings of the National Academy of Sciences*, 109:1–5.

Ralph, F. and Dettinger, M. (2011). Storms, floods, and the science of atmospheric rivers. *Eos, Transactions American Geophysical Union*, 92(32):265–272.

Ralph, F. M. and Dettinger, M. D. (2012). Historical and National Perspectives on Extreme West Coast Precipitation Associated with Atmospheric Rivers during December 2010. *Bulletin of the American Meteorological Society*, 93:783–790.

Rasmussen, K. L. and Houze, R. a. (2011). Orogenic Convection in Subtropical South America as Seen by the TRMM Satellite. *Monthly Weather Review*, 139:2399–2420.

Reichler, T. and Roads, J. (2005). Long-Range Predictability in the Tropics. Part II: 30-60-Day Variability. *Journal of Climate*, 18:634–650.

Reynolds, R., Rayner, N., Smith, T. M., Stokes, D. C., and Wang, W. (2002). An Improved In Situ and Satellite SST Analysis for Climate. *Journal of Climate*, 15:1609–1625.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J. (2011). MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *Journal of Climate*, 24:3624–3648.

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

Ross, S. (2010). *A First Course in Probability, 8th ed.* Prentice Hall.

Roundy, P. E. and Gribble-Verhagen, L. M. (2010). Variations in the Flow of the Global Atmosphere Associated with a Composite Convectively Coupled Oceanic Kelvin Wave. *Journal of Climate*, 23:4192–4201.

Ruff, T. W. and Neelin, J. D. (2012). Long tails in regional surface temperature probability distributions with implications for extremes under global warming. *Geophysical Research Letters*, 39(L04704):1–6.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-Y., Juang, H.-M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., Van Den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M. (2010). The NCEP Climate Forecast System Reanalysis. *Bulletin of the American Meteorological Society*, 91:1015–1057.

Samorodnitsky, G. and Taqqu, M. (1994). *Stable non-Gaussian Random Processes*. Chapman & Hall/CRC.

Sardeshmukh, P. D. and Sura, P. (2009). Reconciling Non-Gaussian Climate Statistics with Linear Dynamics. *Journal of Climate*, 22:1193–1207.

Shen, S., Gurung, A., Oh, H., Shu, T., and Easterling, D. R. (2011). The twentieth century contiguous US temperature changes indicated by daily data and higher statistical moments. *Climatic change*, 109:287–317.

Shen, S. S., Basist, A. N., Li, G., Williams, C., and Karl, T. R. (2004). Prediction of sea surface temperature from the global historical climatology network data. *Environmetrics*, 15(3):233–249.

Shen, S. S., Dzikowski, P., Li, G., and Griffith, D. (2001). Interpolation of 1961-97 Daily Temperature and Precipitation Data onto Alberta Polygons of Ecodistrict and Soil Landscapes of Canada. *Journal of Applied Meteorology*, 40:2162–2177.

Shen, S. S., Lee, C. K., and Lawrimore, J. (2012). Uncertainties, Trends, and Hottest and Coldest Years of U.S. Surface Air Temperature since 1895: An Update Based on the USHCN V2 TOB Data. *Journal of Climate*, 25:4185–4203.

Shen, S. S., North, G., and Kim, K. (1994). Spectral Approach to Optimal Estimation of the Global Average Temperature. *Journal of Climate*, 7:1999–2007.

Shen, S. S., Smith, T., Ropelewski, C., and Livezey, R. E. (1998). An Optimal Regional Averaging Method with Error Estimates and a Test Using Tropical Pacific SST data. *Journal of Climate*, 11:2340–2350.

Shen, S. S., Tafolla, N., Smith, T. M., and Arkin, P. A. (2014). Multivariate Regression Reconstruction and Its Sampling Error for the Quasi-Global Annual Precipitation from 1900 to 2011. *Journal of the Atmospheric Sciences*, 71:3250–3268.

Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the 1968 23rd ACM National Conference*, pages 517–524.

Sherwood, S. and Huber, M. (2010). An adaptability limit to climate change due to heat stress. *Proceedings of the National Academy of Sciences*, 107(21):9552–9555.

Smith, J. (1989). Regional Flood Frequency Analysis Using Extreme Order Statistics of the Annual Peak Record. *Water Resources Research*, 25(2):311–317.

Smith, R. (1987). Estimating Tails of Probability Distributions. *The Annals of Statistics*, 15(3):1174–1207.

Smith, R. (2001). Extreme Value Statistics in Meteorology and the Environment. In *Environmental statistics*, pages 701–761.

Smith, R. L. (1994). Multivariate threshold methods. In *Extreme Value Theory and Applications, Proceedings. Volume 1.*, pages 225–284. Kluwer Academic Publishers.

Stefanova, L., Sura, P., and Griffin, M. (2013). Quantifying the Non-Gaussianity of Wintertime Daily Maximum and Minimum Temperatures in the Southeast. *Journal of Climate*, 26:838–850.

Stephenson, D. B., Hannachi, A., and O'Neill, A. (2004). On the existence of multiple climate regimes. *Quarterly Journal of the Royal Meteorological Society*, 130:583–605.

Stewart, I., Cayan, D., and Dettinger, M. (2004). Changes in Snowmelt Runoff Timing in Western North America under a 'Business as Usual' Climate Change Scenario. *Climatic Change*, pages 217–232.

Straub, K. H. (2013). MJO Initiation in the Real-Time Multivariate MJO Index. *Journal of Climate*, 26:1130–1151.

Subramanian, A. C., Jochum, M., Miller, A. J., Murtugudde, R., Neale, R. B., and Waliser, D. E. (2011). The Madden–Julian Oscillation in CCSM4. *Journal of Climate*, 24:6261–6282.

Sura, P. (2011). A general perspective of extreme events in weather and climate. *Atmospheric Research*, 101:1–21.

Sura, P., Newman, M., Penland, C., and Sardeshmukh, P. (2005). Multiplicative Noise and Non-Gaussianity: A Paradigm for Atmospheric Regimes? *Journal of the Atmospheric Sciences*, 62:1391–1409.

Sura, P. and Perron, M. (2010). Extreme Events and the General Circulation: Observations and Stochastic Model Dynamics. *Journal of the Atmospheric Sciences*, 67:2785–2804.

Sura, P. and Sardeshmukh, P. D. (2008). A Global View of Non-Gaussian SST Variability. *Journal of Physical Oceanography*, 38:639–647.

Taylor, K. E., Stouffer, R. J., and Meehl, G. a. (2012). An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, 93(4):485–498.

Toreti, A., Naveau, P., Zampieri, M., Schindler, A., Scoccimarro, E., Xoplaki, E., Dijkstra, H. a., Gualdi, S., and Luterbacher, J. (2013). Projections of global changes in precipitation extremes from Coupled Model Intercomparison Project Phase 5 models. *Geophysical Research Letters*, 40:4887–4892.

Trenberth, K. E., Dai, A., Rasmussen, R. M., and Parsons, D. B. (2003). The Changing Character of Precipitation. *Bulletin of the American Meteorological Society*, 84:1205–1217.

Trenberth, K. E. and Fasullo, J. T. (2012). Climate extremes and climate change: The Russian heat wave and other climate extremes of 2010. *Journal of Geophysical Research: Atmospheres*, 117(D17103):1–12.

Tsonis, A. (1996). Widespread increases in low-frequency variability of precipitation over the past century. *Nature*, 382:700–702.

Uppala, S. M., KÅllberg, P. W., Simmons, a. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, a., Kelly, G. a., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. a., Beljaars, a. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, a., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. a. E. M., Jenne, R., Mcnally, a. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. a., Saunders, R. W., Simon, P., Sterl, a., Trenberth, K. E., Untch, a., Vasiljevic, D., Viterbo, P., and Woollen, J.

(2005). The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, 131(612):2961–3012.

Vitart, F., Leroy, A., and Wheeler, M. C. (2010). A Comparison of Dynamical and Statistical Predictions of Weekly Tropical Cyclone Activity in the Southern Hemisphere. *Monthly Weather Review*, 138:3671–3682.

von Storch, H., Bruns, T., Fischer-Bruns, I., and Hasselmann, K. (1988). Principal Oscillation Pattern Analysis of the 30to 60day Oscillation in General Circulation Model Equatorial Troposphere. *Journal of Geophysical Research*, 93(D9):11022–11036.

Wackerly, D., Mendenhall, W., and Scheaffer, R. (2008). *Mathematical Statistics with Applications*. Cengage Learning.

Waliser, D., Jones, C., Schemm, J.-K., and Graham, N. R. (1999). A Statistical Extended-Range Tropical Forecast Model Based on the Slow Evolution of the Madden-Julian Oscillation. *Journal of Climate*, 12:1918–1939.

Waliser, D. and Moncrief, M. (2007). The Year of Tropical Convection (YOTC) Science Plan: A joint WCRP - W WRP/THORPEX International Initiative. In *WMO/TD No. 1452, WCRP - 130, WWRP/THORPEX - No 9.*, number March, pages 1–24, WMO, Geneva, Switzerland.

Waliser, D. E., Stern, W., Schubert, S., and Lau, K. M. (2003). Dynamic predictability of intraseasonal variability associated with the Asian summer monsoon. *Quarterly Journal of the Royal Meteorological Society*, 129:2897–2925.

Wallace, J. and Gutzler, D. (1981). Teleconnections in the Geopotential Height Field during the Northern Hemisphere Winter. *Monthly Weather Review*, 109:784–812.

Warner, T. (2011). *Numerical Weather and Climate Prediction*. Cambridge University Press.

Wheeler, M. and Weickmann, K. (2001). Real-Time Monitoring and Prediction of Modes of Coherent Synoptic to Intraseasonal Tropical Variability. *Monthly Weather Review*, 129:2677–2694.

Wheeler, M. C. and Hendon, H. H. (2004). An All-Season Real-Time Multivariate MJO Index: Development of an Index for Monitoring and Prediction. *Monthly Weather Review*, 132:1917–1932.

Wilby, R., Tomlinson, O., and Dawson, C. (2003). Multi-site simulation of precipitation by conditional resampling. *Climate Research*, 23:183–194.

Wilks, D. (2011). *Statistical Methods in the Atmospheric Sciences*, volume 100. Academic press, 2nd edition.

Wilks, D. S. (2010). Use of stochastic weather generators for precipitation downscaling. *Climate Change*, 1(Nov/Dec):898–907.

Wilks, D. S. (2012). Stochastic weather generators for climate-change downscaling , part II : multivariable and spatially coherent multisite downscaling. *Climate Change*, 3(May/June):267–278.

Wilson, P. S. and Toumi, R. (2005). A fundamental probability distribution for heavy rainfall. *Geophysical Research Letters*, 32(L14812):1–4.

Winkler, C., Newman, M., and Sardeshmukh, P. D. (2001). A Linear Model of Wintertime Low-Frequency Variability. Part I: Formulation and Forecast Skill. *Journal of Climate*, 14:4474–4494.

Wuebbles, D., Meehl, G., Hayhoe, K., Karl, T. R., Kunkel, K., Santer, B., Wehner, M., Colle, B., Fischer, E. M., Fu, R., Goodman, A., Janssen, E., Kharin, V., Lee, H., Li, W., Long, L. N., Olsen, S. C., Pan, Z., Seth, A., Sheffield, J., and Sun, L. (2014). CMIP5 Climate Model Analyses: Climate Extremes in the United States. *Bulletin of the American Meteorological Society*, 95:571–583.

Yoneyama, K., Zhang, C., and Long, C. N. (2013). Tracking Pulses of the Madden–Julian Oscillation. *Bulletin of the American Meteorological Society*, 94:1871–1891.

Zhang, C. (2001). Intraseasonal Perturbations in Sea Surface Temperatures of the Equatorial Eastern Pacific and Their Association with the Madden-Julian Oscillation. *Journal of Climate*, 14:1309–1322.

Zhang, C., Gottschalck, J., Maloney, E. D., Moncrieff, M. W., Vitart, F., Waliser, D. E., Wang, B., and Wheeler, M. C. (2013a). Cracking the MJO nut. *Geophysical Research Letters*, 40:1223–1230.

Zhang, H., Fraedrich, K., Blender, R., and Zhu, X. (2013b). Precipitation Extremes in CMIP5 Simulations on Different Time Scales. *Journal of Hydrometeorology*, 14:923–928.

Zolina, O., Kapala, A., Simmer, C., and Gulev, S. K. (2004). Analysis of extreme precipitation over Europe from different reanalyses: a comparative assessment. *Global and Planetary Change*, 44:129–161.