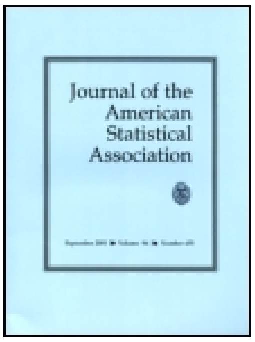

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://www.tandfonline.com/loi/uasa20

Modeling Waves of Extreme Temperature: The Changing Tails of Four Cities

Debbie J. Dupuis

To cite this article: Debbie J. Dupuis (2012) Modeling Waves of Extreme Temperature: The Changing Tails of Four Cities, Journal of the American Statistical Association, 107:497, 24-39, DOI: 10.1080/01621459.2011.643732

To link to this article: https://doi.org/10.1080/01621459.2011.643732

Accepted author version posted online: 06 Jan 2012.
Published online: 06 Jan 2012.

Modeling Waves of Extreme Temperature: The Changing Tails of Four Cities

Debbie J. DUPUIS

Heat waves are a serious threat to society, the environment, and the economy. Estimates of the recurrence probabilities of heat waves may be obtained following the successful modeling of daily maximum temperature, but working with the latter is difficult as we have to recognize, and allow for, not only a time trend but also seasonality in the mean and in the variability, as well as serial correlation. Furthermore, as the extreme values of daily maximum temperature have a different form of nonstationarity from the body, additional modeling is required to completely capture the realities. We present a time series model for the daily maximum temperature and use an exceedance over high thresholds approach to model the upper tail of the distribution of its scaled residuals. We show how a change-point analysis can be used to identify seasons of constant crossing rates and how a time-dependent shape parameter can then be introduced to capture a change in the distribution of the exceedances. Daily maximum temperature series for Des Moines, New York, Portland, and Tucson are analyzed. In-sample and out-of-sample goodness-of-fit measures show that the proposed model is an excellent fit to the data. The fitted model is then used to estimate the recurrence probabilities of runs over seasonally high temperatures, and we show that the probability of long and intense heat waves has increased considerably over 50 years. We also find that the increases vary by city and by time of year.

KEY WORDS: AR models; Change-point analysis; Exceedances over thresholds; Extreme value theory; Return periods; Seasonal volatility.

1. INTRODUCTION

Extreme weather and climate events negatively impact our society and environment (IPCC 2007). Heat waves, loosely defined as periods of unusually hot weather, are especially unforgiving for societies and infrastructures unable to cope or adapt. Stress from heat waves has been linked to excess human mortality and morbidity, violent behavior, drought, forest fires, tornadoes, decreased agricultural and livestock productivity, construction and transportation difficulties, and reduced electrical power supply (Smoyer-Tomic, Kuhn, and Hudson 2003). The epidemiological evidence of how climate variations and trends affect various health outcomes is well summarized in McMichael, Woodruff, and Hales (2006). Several authors have shown the impacts of heat waves on mortality, for example, Hajat et al. (2007) and Pauli and Rizzi (2008). The mortality excess is arresting: a 1995 heat wave led to at least 700 excess deaths in Chicago (Semenza et al. 1996), a 2003 heat wave accounted for more than 30,000 deaths throughout western Europe (Kosatsky 2005), and a preliminary estimate of the number of excess deaths during the 2010 Russian heat wave is 15,000 (Bloomberg 2010). Heat stress is already a leading cause of fatalities from natural phenomena (Kovats and Hajat 2008).

Three important trends are likely to exacerbate the problems: the aging population, urbanization, and climate change. The elderly appear to be disproportionately stressed when compared with other age groups (Kalkstein and Davis 1989). Worldwide, the number of persons aged 60 years or over is expected to almost triple, increasing from 672 million in 2005 to nearly 1.9 billion by 2050 (United Nations Population Division 2004). The *heat island effect* causes annual mean air temperature of a city with 1 million people or more to be 1.8°F–5.4°F warmer than that of the surrounding areas; and, in the evening, the difference can be as high as 22°F (U.S. Environmental Protection Agency 2010). An increasing proportion of the world's population lives in urban areas. Globally, this urbanization is projected to increase from 50.6% in 2010 to 70% in 2050 (UN-Habitat 2010). Finally, general warming of about 0.36°F per decade is projected for the next two decades for a range of special report on emissions scenarios (SRES) (IPCC 2007).

It is thus not surprising that the future outlook is bleak. Knowlton et al. (2007) projected a 70% increase in heat-related premature mortality in New York City by 2050. According to U.S. Global Change Research Program (2010), heat-related deaths in Chicago are expected to quadruple by 2050. A study of 44 large U.S. cities with metropolitan areas exceeding 1 million in population shows increases ranging from 70% for the most conservative 2050 Global Climate Models (GCM) to over 100% for the other GCM, even if the population acclimatizes to the increased warmth (Kalkstein and Greene 1997). The problem is global as the Australian Government estimates that heat-related deaths there could more than double to 2,500 a year by 2020 if there is no adaptation (Australian Government 2009). To combat the negative impacts of heat waves on our society and environment, governmental and nongovernmental organizations need to ensure adequate socioeconomic conditions for the vulnerable and dedicate the necessary resources to address problems. Policy decisions at local, national, and international levels, from those on issuing heat advisories to those on the proper dimensioning of engineering works, need to be based on good estimates of the recurrence probability of heat waves of different intensities and length. We need to provide these estimates and measures of their accuracy and uncertainty, along with projections of how they may change in the future. The former, along with estimating changes in recurrence probabilities of heat waves over 50 years, is the object of this article.

Debbie J. Dupuis is a Professor in the Department of Management Sciences, HEC Montréal; 3000, chemin de la Côte-Sainte-Catherine, Montréal (Québec), Canada H3T 2A7 (E-mail: *debbie.dupuis@hec.ca*). The author acknowledges the support of the Natural Sciences and Engineering Research Council of Canada. The author also thanks three anonymous referees for very constructive comments that helped improve a previous version of the article.

Journal of the American Statistical Association
March 2012, Vol. 107, No. 497, Applications and Case Studies
DOI: 10.1080/01621459.2011.643732

While there are many different technical definitions of heat waves (e.g., see Pauli and Rizzi 2008), most include a period of consecutive days, sometimes called a *run*, during which some measure of heat stress, for example, maximum daytime temperature or minimum nighttime temperature, is above a prespecified unusually high value. As early as Mearns, Katz, and Schneider (1984), there is interest in computing the probability of an extreme *run event*. A criterion is set for the run event and the relative frequencies of occurrence of the events over the observed data are estimates of these probabilities. Many authors have followed the original path of this article by adopting a heat wave definition and *counting* heat waves to show patterns in space or trends in time: Easterling et al. (2000), Frich et al. (2002), Meehl and Tebaldi (2004), Schär et al. (2004), Tebaldi et al. (2006), Della-Marta et al. (2007), Hunt (2007), Koffi and Koffi (2008), Ganguly et al. (2009), and Kuglitsch et al. (2010). The latter analyses are on past observations, GCM data, or on both. While these authors adopt various definitions of heat waves, they all show an observed increase in the occurrence of heat waves over time and/or project increases in the number of heat waves for the next century.

Some authors have obtained estimates for extreme run events not seen in the observed data by fitting an AR(1) model to daily maximum temperatures and generating long artificial series of daily maximum temperature (from which recurrence probabilities of heat waves are estimated). Mearns et al. (1984) fitted an AR(1) model to data for the month of July, assuming stationarity throughout the month and over the years. Macchiato et al. (1993) generalized this approach by including a deterministic part to capture the seasonal cycle in the mean and the variability, enabling a fit to data from the entire year. Kyselý (2010) further generalized by allowing a seasonal cycle and long-term trend in the first-order autocorrelation coefficient. One key assumption in these latter approaches cannot be ignored: the error term in the AR(1) process is normally distributed. Daily maximum temperatures, even after detrending and seasonal scaling, retain their *maximum*, that is, extreme, character, and have no reason to be normally distributed.

Extreme value theory (EVT) provides approximate probabilistic models allowing one to model and make inferences for extreme values (Coles 2001). These models for the tails of distributions are derived from asymptotic limit theorems. Using these models, certain regularity assumptions about the distributional tails, and a subset of large values from the data sample, one can make inferences about the extreme behavior of the underlying process that generated the data. The large values can be selected in various ways: block maxima (e.g., annual maxima), the r-largest values (e.g., the five largest events in the year), or threshold excesses (e.g., exceedances above a predefined threshold). Block maxima lead to generalized extreme value (GEV) models (see Coles 2001, chap. 3), while threshold excesses lead to generalized Pareto (GP) models (see Coles 2001, chap. 4).

While there have been many important developments in EVT in the last 50 years, the analysis of extreme temperature events remains very challenging. Zwiers and Kharin (1998) fitted stationary GEV models to annual maximum temperatures. Recognizing the changes in mean and variability over the years studied, Kharin and Zwiers (2005) then fitted nonstationary GEV models to annual maximum temperatures. A simple example of these types of analyses appears in Cooley (2009). Modeling annual maxima is insufficient however for the study of heat waves, and we have to work with daily maxima to eventually get the probability that successive daily maxima exceed a given level. Working with daily maxima increases the level of difficulty significantly as we now have to recognize, and allow for, not only time trends but also seasonality in the mean and in the variability, as well as serial correlation.

A first attempt was made by Nogaj et al. (2006) who removed the trend and seasonality in the mean using splines, hoped to limit the seasonality in the variability by considering only data for June–July–August (JJA), used a declustering procedure to ensure the independence of extremes, and then used the standard approach to modeling extremes (Davison and Smith 1990). This is an application of a now well-known approach to modeling the extremes of a nonstationary process: use a constant high threshold and introduce covariates into the threshold exceedance rate parameter and extreme value model parameters for the threshold exceedances to explain the nonstationarity. The approach was subsequently used by others to model extreme temperatures; for example, Abaurrea et al. (2007) who also considered JJA data but alternatively used first-order harmonic functions to describe the annual cycle in the mean and do not decluster.

An alternative approach to modeling the extremes of a nonstationary process is to consider time-varying thresholds. This approach was implemented for extreme temperature data by Coelho et al. (2008) and Kyselý, Picek, and Beranová (2010). Both articles consider only JJA data, but offer different ways of specifying the time-varying threshold.

While these two approaches to modeling the extremes of nonstationary processes allow one to get to extreme daily maxima over the months studied, they are deficient in two respects. First, they are unable to address heat waves as the daily serial correlation is not modeled. Second, these approaches have been applied only to JJA data and we do not know if they offer a good fit over other months. And, even if they did offer a good fit for certain months, we would have to find an acceptable way to chop the year into periods with similarly behaving daily maximum temperatures to which we could then fit the published models. It is important to be able to compute the probability of a heat wave for any particular time of year as heat waves early in the summer result in greater heat-related mortality and illness than heat waves later in the summer (Hajat et al. 2002), winter heat waves have negative impacts on resources and economies of mountainous areas (Beniston 2005), and autumn heat waves reduce the ability of northern ecosystems to sequester carbon (Piao et al. 2008). Ultimately, we wish to have a sufficiently general and flexible model that allows inferences about heat waves at any time of the year and can establish changes in the recurrence probabilities of such heat waves over the years.

Eastoe and Tawn (2009) suggested a third approach to modeling the extremes of a nonstationary process based on *preprocessing* the data. They showed that the preprocessing method gives a model that better incorporates the underlying mechanisms that generate the process and for which the form of the covariate selection in the parameters is invariant to the threshold choice, although the model was rather a poor fit to their ozone data.

In this article, we implement the preprocessing approach and introduce new modeling tools to ensure a proper implementation on our extreme temperature data. Traditional preprocessing is used to account for trends and seasonality in the mean, seasonality in the variance, and serial correlation in daily maximum temperatures. The additional modeling, which includes the estimation of *seasons* over the year using change-point analysis and, then for each season, the estimation of *blocks* over time during which extreme daily maximum temperature exhibits similar tail behavior using maximum likelihood procedures for GP models with time-varying shape parameters, is necessary to properly model extreme daily maximum temperature over the 50-year period studied. In-sample and out-of-sample goodness-of-fit measures show that the proposed model is an excellent fit to the data. We then simulate a long series of daily maximum temperatures to assess the change in the recurrence probabilities of heat waves over the 50-year period studied. Results show that while heat waves have become significantly more intense and more frequent in the four cities studied, the 50-year evolution is not the same in the four cases. Furthermore, we show that changes are not uniform throughout the year; for example, in Tucson, we see that while intense 5-day heat waves in the early fall are more than five times more likely in 2005 than 50 years prior, the likelihood of very intense 2- to 5-day heat waves in the early spring in the city has less than doubled over the 50 years. To our knowledge, no other published model allows inferences and comparisons for such year-round extreme events.

The remainder of the article is organized as follows. In Section 2, we review some essential results in EVT, and in Section 3, we present our full preprocessing model. In Section 4, the data are presented and analyzed. Some discussion and conclusions appear in Section 5. Finally, some technical details appear in the Appendix.

2. EXTREME VALUE THEORY

2.1 Stationary Processes

Consider the problem of statistical inferences about the extreme values of a random process $\{Y_t\}$. Suppose that the process is stationary and has marginal distribution F with upper end point x^F. Pickands (1975) showed that if the distribution of excesses $Y_t - u$ of a high threshold u, $u < x^F$, scaled as a function of u, converges to a nondegenerate limiting distribution as $u \to x^F$, that distribution must be the generalized Pareto distribution (GPD). This derivation suggests that the GPD will be a practical family for statistical estimation when examining excesses over thresholds u, provided that the threshold is taken sufficiently high. The tail model thus includes the parameter $\phi_u = \Pr(Y > u)$ and the GPD scale parameter $\beta_u > 0$ and shape parameter ξ where

$$\Pr(Y > y + u | Y > u) = \left[1 + \frac{\xi y}{\beta_u}\right]_+^{-1/\xi}, \tag{1}$$

and $a_+ = \max(0, a)$. This threshold approach has been used numerous times to characterize the extreme behavior of various random processes in diverse fields of study since it was first presented in Davison and Smith (1990). The size and the rate of occurrence of the observations that exceed the threshold summarize the extreme behavior. In practice, the parameters must be estimated from the data. The maximum likelihood estimate (MLE) of the rate parameter ϕ_u is n_u/n, where n_u is the number of exceedances of the threshold u among the n observations of the process. The MLE of the GPD parameters is found by numerical optimization.

2.2 Nonstationary Processes

Environmental processes are often nonstationary since climate patterns cause systematic seasonal effects and long-term climate changes cause trends. The usual limit models are not applicable for nonstationary processes, but the standard EVT models can be used along with statistical modeling to provide useful inference. Here, we suppose that the nonstationarity is the result of a dependence on time t. Davison and Smith (1990) and Smith (1989) made the first proposal for extending the GPD to nonstationary processes. They continued to model the exceedances of a fixed high threshold u, but accounted for the nonstationarity in the exceedances by letting the parameters of the GPD be a function of t. That is, the rate of exceedance becomes $\phi_u(t)$ and the excesses follow a GPD($\beta_u(t)$, $\xi(t)$). Linear or quadratic terms in t are used to model long-term trends and harmonic functions can be used to capture seasonal effects.

An alternative approach to deal with the nonstationarity is to use a time-varying threshold to define the extremes on the original scale. Practitioners frequently use the artifice of splitting the data into seasons, allowing different thresholds in different seasons in the hope that stationarity will prevail within a season. We can, however, specify a time-varying threshold $u(t)$. This approach was implemented for extreme temperature data by Coelho et al. (2008) and Kyselý et al. (2010), where two different approaches to specifying $u(t)$ were outlined.

Finally, Eastoe and Tawn (2009) proposed a third approach for handling the nonstationarity. Proceeding as is commonly done in time series analysis, they preprocessed (or prewhitened as it is commonly referred to in finance) the data before fitting a model to the residuals. The preprocessing approach was outlined in the general context of nonstationarity due to dependent covariates $\{X_t\}$ by Eastoe and Tawn (2009), but we limit our presentation to a time dependence only as we use no other covariate data. We model

$$Y_t = \mu_t + \sigma_t Z_t, \tag{2}$$

where $\{Z_t\}$ is assumed to be approximately stationary. The nonextreme part of the distribution of $\{Z_t\}$, that is, the part below some threshold u, is modeled by its empirical distribution $\tilde{F}_{Z,u}$ and the extreme values of $\{Z_t\}$, that is, the values above the threshold u, by the first approach to nonstationary processes outlined in the beginning of this section. We thus have $\phi_{z,u}(t)$ as the rate of exceedance of u_z by Z_t, and $\beta_{z,u}(t)$ and $\xi_z(t)$ as the scale and shape parameters of the GPD, respectively. We thus have a two-step procedure. The first step is to estimate the location and scale parameters μ_t and σ_t. The second step is to estimate $\phi_{z,u}(t)$, $\beta_{z,u}(t)$, and $\xi_z(t)$.

Under the preprocessing approach, the conditional and marginal return levels of the nonstationary process are computed as follows. Consider a return level associated with a return period of $1/p$ observations. We can find the conditional return level $z_{p,t}$ for the transformed series $\{Z_t\}$. Then, since the conditional return probability is

$$p = \Pr(Y_t > y_{p,t}|\mathcal{F}_{t-1}) = \Pr(\mu_t + \sigma_t Z_t > y_{p,t}|\mathcal{F}_{t-1}), \quad (3)$$

where $\mathcal{F}_{t-1}$ denotes the information set available at time $t-1$, we easily back-transform to get $y_{p,t} = \mu_t + \sigma_t z_{p,t}$. Let $z_p(t)$ be the transformation under (2) of the marginal return level y_p. The marginal return level y_p is the solution to the equation

$$\begin{aligned} p &= \frac{1}{n}\Bigg[\sum_{t\in T}\Pr\{Z_t > z_p(t)|Z_t > u_z, \mathcal{F}_{t-1}\}\Pr(Z_t > u_z|\mathcal{F}_{t-1}) \\ &\quad + \sum_{t\notin T}\Pr\{Z_t > z_p(t)|\mathcal{F}_{t-1}\}\Bigg] \\ &= \frac{1}{n}\left(\sum_{t\in T}\left\{\phi_{z,u}(t)\left[1+\xi_z(t)\frac{z_p(t)-u_z}{\beta_{z,u}(t)}\right]_+^{-1/\xi_z(t)}\right\}\right. \\ &\quad \left. + \sum_{t\notin T}[1-\tilde{F}_{Z,u}\{z_p(t)\}]\right), \end{aligned} \quad (4)$$

where $T = \{t : z_p(t) > u_z\}$ and n is the total sample size. Note that both the empirical distribution $\tilde{F}_{Z,u}$ and the GPD ($\phi_{z,u}(t)$, $\beta_{z,u}(t)$, $\xi_z(t)$) intervene in (4) as the marginal return level y_p could be attained with a large value of z_t (i.e., larger than the threshold u_z, and thus the first sum) or with a value of z_t below u_z (and thus the second sum). The marginal return level integrates across the observed set of times and therefore corresponds to a "typical" return level within the period for which data are available. The marginal return level y_p therefore relates to a particular period of time, in a similar way that the conditional return level $y_{p,t}$ relates to a particular day t. Note finally that definitions (3) and (4) relate to the level which is exceeded with probability p on a particular day and that the often referred to "N-year return level" is obtained by taking $p = 1/(365.25 \times N)$.

3. A FULL PREPROCESSING MODEL

In this section, we give the modeling details of the two-step preprocessing approach as required for application to our nonstationary daily maximum temperature series.

3.1 Specification

We model daily maximum temperature as a regression between daily detrended and deseasonalized maximum temperatures. Let Y_t be the daily maximum temperature on day t. We consider the time series model

$$Y_t = \mu_t + \sigma_t Z_t,$$

where

$$\mu_t = S_t + \sum_{r=1}^{R}\alpha_r\,(Y_{t-r} - S_{t-r})\,, \quad (5)$$

$$\begin{aligned} S_t &= a + \sum_{k=1}^{K-1} h_k I\,[t\in B_{k+1}] + \sum_{i=1}^{H_1}\left\{a_i \sin\left(\frac{2i\pi}{365}d(t)\right)\right. \\ &\quad \left. + b_i\cos\left(\frac{2i\pi}{365}d(t)\right)\right\}, \\ \log\sigma_t^2 &= c + \sum_{i=1}^{H_2}\left\{c_i\sin\left(2i\pi\frac{d(t)}{365}\right) + d_i\cos\left(2i\pi\frac{d(t)}{365}\right)\right\}, \\ Z_t &\sim \text{stationary}, \end{aligned} \quad (6)$$

where I is the usual indicator function taking the value 1 if the time t is in time block B_{k+1} and 0 otherwise, and $d(t)$ is a repeating step function that cycles through $1,\dots,365$ (February 29 is dropped from all leap years). We consider K such that each B_k is a 5- or 10-year period. The values of H_1, H_2, and R are chosen following Bayesian information criterion (BIC) considerations (Schwarz 1978). Note that we have modeled the long-term trend in the mean as a step-like function over half or full decades. This allows us to capture the observed increases in mean daily maximum temperature without forcing the often adopted linear trend.

Unfortunately, data analysis in Section 4 shows that the derived series $\{Z_t\}$ is not stationary. The extreme values of $\{Y_t\}$ may have a different form of nonstationarity from the body of the $\{Y_t\}$, and the extreme values of $\{Z_t\}$ may not behave like extreme values of a stationary series. Our data analysis shows that exceedances of large residuals remain nonstationary with respect to the time of the year. This is likely due to the underlying physical conditions. For example, in winter, daily maximum temperature is driven by advection, while in the summer, clear skies and light winds can be the driving force. We wish to exploit the empirical evidence and scientific support for a multiseasonal structure and propose to adopt a change-point model in which the year is partitioned into *seasons*. Using change points in the extreme value context is not new, but our approach is innovative. Smith (2000) applied a Bayesian approach to the point process representation of extremes (Smith 1989) and developed a change-point model for stochastic volatility in financial time series. This is essentially equivalent to fitting the GPD to exceedances, assuming that the parameters change from one period to another according to a random change-point process. Coles and Pericchi (2003) used the same approach to fit a change-point model to the rainfall process in Venezuela where meteorological activity suggests a two-season pattern, but with unspecified seasonal timings. In both cases, maximum likelihood is not a viable option for fitting the model and Bayesian inference is required. Even the latter requires the imposition of some limited structure on the change-point parameters. To obtain estimates, Smith (2000) fixed the number of change points, while Coles and Pericchi (2003) put further constraints on their possible location.

We propose a new computationally simpler two-step approach: first, we identify the change points, and then we fit the GPD. To identify seasonal changes in the tail distribution of $\{Z_t\}$, we look at exceedances of $\{Z_t\}$ over a high threshold and create Bernoulli random variables $\{W_t\}$. We then fit a multiple change-point model to the observed Bernoulli variables to estimate *seasons* during the year with constant exceedance rates of the high threshold. More specifically, we define $W_t = I(Z_t > z_p)$, fix $\tau_0 = 1$, and assume a number m of change points $\tau_1,\dots,\tau_m$ such that the observations W_i with

$\tau_{j-1} < d(i) \le \tau_j$ follow a Bernoulli model with parameter p_j. The z_p is an upper quantile defined such that $\Pr(Z_t > z_p) = p$ and p is small.

The change points $\tau_1, \ldots, \tau_m$ define our seasons, that is, periods of constant threshold exceedance rates. We can assume a constant threshold crossing rate within a season and the GPD can be fitted to data from each season. However, we note that for a given season, the GPD parameters β and ξ could also be time-varying because of changes in weather patterns over the years due to climate change, industrialization, etc. We fit a GPD with various time covariate models: (i) allowing for one change point in the value of ξ, (ii) allowing for one change point in the value of β, (iii) allowing for one change point in the values of both ξ and β, and (iv) log-linear trend for β and one change point in the value of ξ. More precisely, for $\{Z_t\}$ with $\tau_{j-1} < d(t) \le \tau_j$, we model $(Z_t - u_{z,j})|Z_t > u_{z,j} \sim \mathrm{GPD}(\beta_j(t), \xi_j(t))$ where model (i) $\xi_j(t) = \xi_{j0} + \xi_{j1} * I(t > \eta_j)$ and $\beta_j(t) = \beta_j$, model (ii) $\beta_j(t) = \beta_{j0} + \beta_{j1} * I(t > \eta_j)$ and $\xi_j(t) = \xi_j$, model (iii) $\beta_j(t) = \beta_{j0} + \beta_{j1} * I(t > \eta_j)$ and $\xi_j(t) = \xi_{j0} + \xi_{j1} * I(t > \eta_j)$, model (iv) $\beta_j(t) = \exp(\gamma_j t)$ and $\xi_j(t) = \xi_{j0} + \xi_{j1} * I(t > \eta_j)$, and u_z is a high threshold. We will refer to the change point η_j as the change year. We select the model based on likelihood considerations, carrying out likelihood ratio tests where appropriate.

3.2 Estimation

We assume that detrended and deseasonalized temperatures follow an AR model with exogenous explanatory variables in the conditional variance equation; see (5) and (6). We obtain the detrended and deseasonalized data in a first step, and then fit our AR model with exogenous explanatory variables in the conditional variance equation to these data. More precisely, we first use a linear regression analysis to find estimates of a; h_k, $k = 1, \ldots, (K-1)$; a_i, b_i, $i = 1, \ldots, H_1$. We consider $K = 5, 10$; $H_1 = 1, \ldots, 10$ and choose the values of K and H_1 using BIC, retaining the corresponding estimates of a; h_k, $k = 1, \ldots, (K-1)$; $a_i, b_i, i = 1, \ldots, H_1$. Then, we regress the detrended and deseasonalized daily maximum temperature on day t against those of previous days to find α_r, $r = 1, \ldots, R$. The AR model with seasonal disturbances can be estimated using maximum likelihood with a normal conditional distribution. We estimate model parameters α_r, $r = 1, \ldots, R$; c, c_i, d_i, $i = 1, \ldots, H_2$, using the `garch` function in the `S+FinMetrics` module of S-Plus, which allows ARMA terms as well as exogenous explanatory variables in the conditional mean equation and in the conditional variance equation (see Zivot and Wang 2002). Again, we choose the values of R and H_2 using BIC. Note that we also estimated the model using other thicker-tailed conditional distributions, for example, a Student-t distribution with four degrees of freedom, and found that for our data, the BIC-selected models under these alternative distributions were always subsets of those selected under normality. The model selection thus seems robust to our conditional distributional assumptions and we are confident that the important covariates have been retained.

Using $p = 0.02$, we then create the series of Bernoulli random variables $\{W_t\}$ and use the dynamic programming (DP) algorithm of Hawkins (2001) to find the change points $\tau_1, \ldots, \tau_m$. Details are given in the Appendix. Finally, for $\{Z_t\}$ with $\tau_{j-1} < d(t) \le \tau_j$, we estimate $\beta_j(t)$ and $\xi_j(t)$ by maximizing the log-likelihood of the GPD model for the data $(Z_t - u_{z,j})|Z_t > u_{z,j}$ where the high threshold $u_{z,j}$ is such that $\Pr(Z_t > u_{z,j}) = 0.02$ for $\{Z_t\}$ with $\tau_{j-1} < d(t) \le \tau_j$. We perform the maximization over the entire parameter space of β's and ξ's, but limit possible values of η_j to integer-valued years yielding sufficiently large samples before and after η_j to assure the GPD fit. Again here, we found only mild sensitivity to the threshold selection as long as it was selected sufficiently large. Using $\Pr(Z_t > u_{z,j}) = 0.02$ yielded good results, that is, good performance under the goodness-of-fit measures discussed in the following section, while allowing the use of the largest amount of data.

3.3 Goodness of Fit

We carry out a simulation to compute marginal return levels for daily maximum temperatures based on our full preprocessing model approach. Simulation is necessary since the AR component means that it does not suffice to look at the GPD estimates on day t to get estimates for the object of interest: maximum temperatures for day t, regardless of the maximum temperatures on days leading up to day t. We produce a QQ-plot for marginal return levels to assess the fit. Assuming that each series $\{Z_t\}$ before (after) η_j with $\tau_{j-1} < d(t) \le \tau_j$ is stationary, we can obtain bootstrap confidence intervals. We obtain nonparametric bootstrap samples of residuals before (after) the change year η_j within each estimated season j, and *glue* the samples together to form a 50-year bootstrapped series. Then, we back-transform the bootstrapped series to the original scale by using the parameters fitted to the original data. Our proposed model is fitted to the resulting bootstrapped sample and the results from repeatedly applying these steps are used to obtain a sampling distribution for the return levels. Note that our analysis shows no evidence for autocorrelation within the residuals $\{Z_t\}$ so a block bootstrap approach is unnecessary. Note that the bootstrap procedure does not account for uncertainty in the timing of the change points τ_j or the change year η_j.

To assess the goodness of fit of the conditional distribution, we take an out-of-sample approach and look at the forecasting distribution for day $t+1$ given data up to day t. We assess the adequacy of the fit of the conditional distributions of our full preprocessing model in the following way. Let F_t be the forecast distribution and y_t be the outcome. In accordance with the prequential principle in Dawid (1984), a forecast distribution must be assessed on the basis of the forecast-observation pairs (F_t, y_t) only, regardless of its origin. The outcome y_t is a random number with distribution G_t chosen by nature. If $F_t = G_t$ for all t, the forecaster is said to be ideal. Following Dawid (1984) and Diebold, Gunther, and Tay (1998), we use the value of the probability integral transform (PIT) $p_t = F_t(y_t)$ to assess the adequacy of the fit of the conditional distribution. If the forecaster is ideal and F_t is continuous, then p_t has a uniform distribution. That is, the uniformity of the PIT is a necessary condition for the forecaster to be ideal. Histograms of the PIT values help identify departures from uniformity when the sample

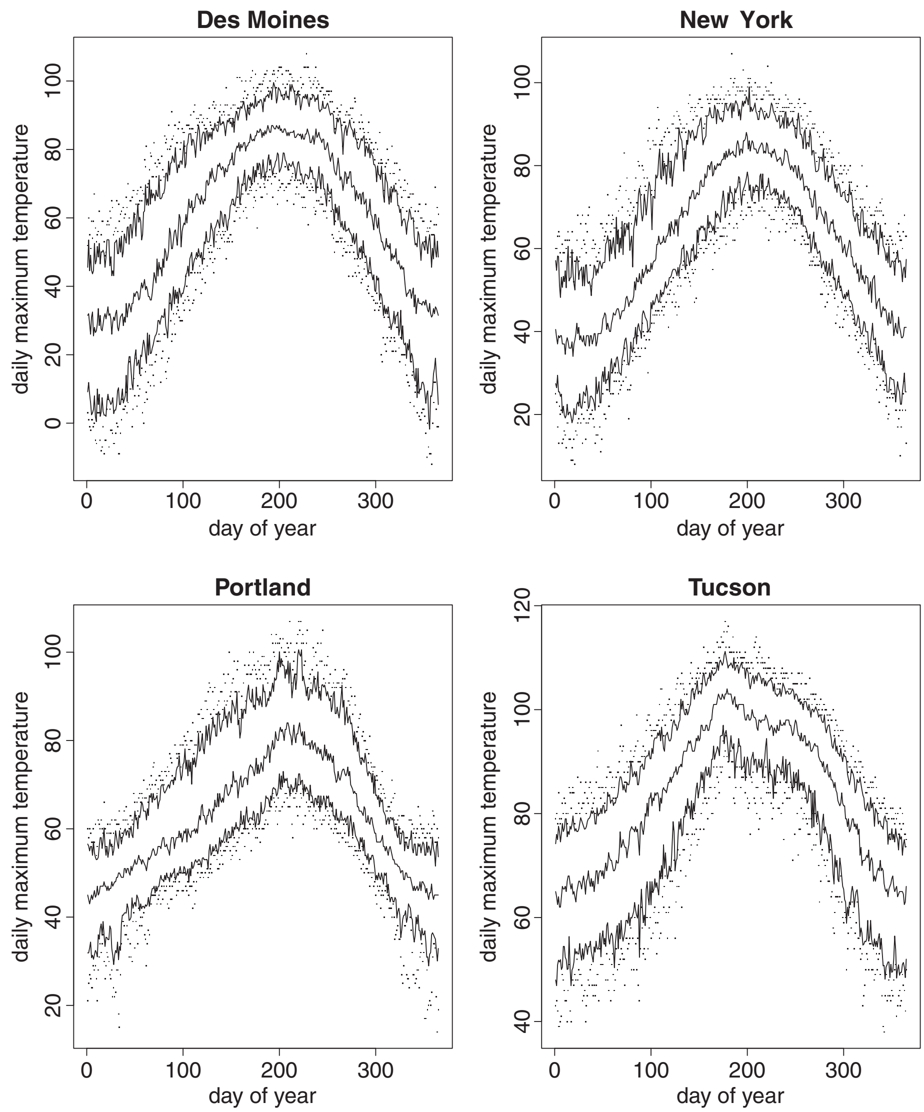

Figure 1. Minimum, 5% quantile, median, 95% quantile, and maximum observed daily maximum temperatures (in °F) for Des Moines, New York, Portland, and Tucson, 1956–2005.

size is large. We can aggregate the PIT by month, year, etc., to check whether the conditional distribution is adequate.

4. DATA ANALYSIS

We consider the historical temperatures for four cities: Des Moines, New York, Portland, and Tucson. The Weather Bureau Army Navy (WBAN) station numbers are 14933, 14732, 24229, and 23160 for the four sites, respectively. We purchased the weather data from these stations from the National Climatic Data Center (NCDC). Daily maximum temperatures are recorded to the nearest degree Fahrenheit. All the stations have been functioning for over five decades and we have used 50 years of data, from 1956 to 2005, inclusively. There are the occasional missing observations, but there are no more than 50 missing points over 18,250 daily observations (365 days per year for 50 years). Given the small proportion of missing values, we simply filled in missing points by linear interpolation.

The 2009 population estimates for the metropolitan area of these cities are 562,906, 19,069,796, 2,241,841, and 1,020,200, respectively (U.S. Census Bureau 2009). These cities offer different levels of the heat island effect, are subject to different seasonal variability and different weather patterns, and form an eclectic sample of U.S. cities. Quantile curves for the observed daily maximum temperatures are plotted in Figure 1. The data show the now well-known behaviors: seasonality in both mean and volatility, both of which are addressed in our full preprocessing model.

4.1 Estimated Models

Our proposed model as outlined in Section 3.1 is fitted to the data from the four cities following the estimation procedures detailed in Section 3.2. Estimated time trend parameters for the mean are shown in Table 1. It is interesting to note that mean levels for the last two decades are greater than the respective 1956–1965 mean levels for all cities. Tucson has both the largest absolute and relative increases with the mean daily maximum level in 1986–1995 rising 3.1°F above the 1956–1965 level of 81.7°F. Our model for the seasonality trend function $S(t)$ is a good fit to the data and is capturing 74%–80% of the variability

Table 1. Estimated time trend parameters for mean: a is mean level for 1956–1965, h_k, $k = 1, \ldots, 4$, are changes (in degrees Fahrenheit) with respect to a for 1966–1975, 1976–1985, 1986–1995, and 1996–2005, respectively

	Des Moines	New York	Portland	Tucson
a	59.08 (0.17)	61.35 (0.13)	61.68 (0.12)	81.68 (0.11)
h_1	0.12 (0.25)	−0.45 (0.19)	1.14 (0.17)	−0.69 (0.16)
h_2	0.66 (0.25)	−0.26 (0.19)	0.68 (0.17)	0.14 (0.16)
h_3	1.47 (0.25)	1.55 (0.19)	2.22 (0.17)	3.13 (0.16)
h_4	1.25 (0.25)	1.28 (0.19)	1.33 (0.17)	2.00 (0.16)
H_1	3	3	6	5
R^2	78.0%	80.1%	73.8%	77.3%

Second to last row shows value of H_1 as selected by BIC. Last row shows R^2 of linear regression model for $S(t)$.

in the daily maximum temperatures. The order H_1 of the harmonic series selected following BIC also appears in Table 1. The value of H_1 is not the same for all four cities.

Estimated AR parameters for model (5) with seasonal disturbances (6) where R and H_2 are selected following BIC are shown in Table 2. It is interesting to note that at least the first three lag parameters are significant in each city. Tucson shows the strongest lag-1 dependence, estimates for Des Moines and Portland are approximately the same (and weaker than Tucson), while New York shows the weakest lag-1 dependence. This is expected as the Atlantic jet stream leads to faster moving weather systems in New York. The significance of at least the first three lags also suggests that fitting AR(1) models to daily maximum temperatures to estimate recurrence probabilities of heat waves by simulation (Mearns et al. 1984; Macchiato et al. 1993; Kyselý 2010) is ill-advised for these cities. The estimated constant c of the seasonal variance term varies considerably across the four cities, as does the volatility of detrended and deseasonalized daily maximum temperature (see Figure 2). The BIC-selected order H_2 of the harmonic series, which is listed in Table 2, is effectively capturing the seasonality in each city.

Scaled residuals Z_t are then calculated and inspected for serial correlation. Estimated lag coefficients are very small (not shown) and there are no significant lags. The nonextreme part of the distribution of the Z_t is modeled by its empirical distribution $\tilde{F}_{Z,u}$, but we need to check for, and model if necessary, any nonstationarity in the extremes of the $\{Z_t\}$ process. We use the following approach to check the tail behavior. Under the null hypothesis that the Z_t are iid, the expected number of the exceedances of any n of them over a high threshold u_p such that $\Pr(Z_t > u_p) = p$ is np, and the total number of exceedances, $\sum_{t=1}^{n} I(Z_t > u_p)$, follows a binomial$(n, p)$ distribution. We slice the 50-year span into different periods and perform a two-sided binomial test of the correct (expected) number of exceedances against the alternative of too few or too many exceedances over a given period. Results for tests over 10-year blocks are listed in Table 3 and show no evidence of nonstationarity.

It is thought, however, that the extremes of Z_t are nonstationary due to the fundamental difference in heat-generating mechanisms over seasons. We thus repeat the analysis of Table 3, but now focus on residuals by season over the 50-year period. Results for the tests are listed in Table 4. This analysis highlights the fact that the Z_t is not identically distributed over the year.

While we show the nonstationarity with respect to the four calendar-based seasons, it is inadequate to adopt a model for Z_t simply based on these divisions. It is preferable to estimate *seasons* as described in Section 3.2. Estimated seasonal breaks are shown in Figure 3. We see that the optimal seasons are city dependent and do not correspond to the calendar-based divisions either in location or in number. While the exceedance rate of the Z_t process over high thresholds is the same within these estimated seasons, we wish to allow for possible changes in the distribution of these exceedances over the 50-year period. We fit the time covariate GPD models listed in Section 3.1. For Des Moines and Tucson, the likelihood values for model (i) are always greater than those for model (ii). For New York, the likelihood value for model (i) is much greater than that for model (ii) for season 2, while the two models are comparable with only a slight preference for model (ii) for the other two seasons. For Portland, the likelihood values for models (i) and (ii) are comparable with only a slight preference for model (ii). The likelihood ratio tests comparing model (i) and models (iii) and (iv) never support model (iii) or (iv) for any of the seasons in any of the cities. Given this, and in the interest of being able to better compare changes across the cities, we opted to retain the fit of model (i) in all cases.

Table 2. Estimated AR parameters for BIC-selected model with seasonal disturbances

	Des Moines	New York	Portland	Tucson
α_1	0.690 (0.007)**	0.626 (0.007)**	0.707 (0.007)**	0.834 (0.006)**
α_2	−0.097 (0.008)**	−0.111 (0.008)**	−0.088 (0.008)**	−0.181 (0.008)**
α_3	0.045 (0.008)**	0.050 (0.007)**	0.021 (0.009)*	0.037 (0.007)**
α_4	0.013 (0.008)	–	0.034 (0.007)**	–
α_5	−0.004 (0.008)	–	–	–
α_6	0.014 (0.008)	–	–	–
α_7	−0.008 (0.007)	–	–	–
α_8	0.040 (0.007)**	–	–	–
c	4.06 (0.01)**	3.74 (0.01)**	3.36 (0.01)**	3.07 (0.007)**
H_2	2	3	4	4

Second to last row shows estimate of constant c in model (6). Last row shows value of H_2 as selected by BIC. *Significant at 0.05; **significant at 0.001.

Table 3. Number of exceedances of residuals Z_t over high thresholds $u_{0.10}$ [1], $u_{0.05}$ [2], $u_{0.01}$ [3], and $u_{0.001}$ [4]. Expected number of exceedances over the four thresholds are 365, 182.5, 36.5, and 3.65, respectively

	Des Moines				New York				Portland				Tucson			
	[1]	[2]	[3]	[4]	[1]	[2]	[3]	[4]	[1]	[2]	[3]	[4]	[1]	[2]	[3]	[4]
1956–1965	390	199	27	4	388	207*	43	4	369	189	44	7	354	173	27	3
1966–1975	374	191	49*	1	359	171	31	3	378	180	40	4	374	186	38	7
1976–1985	352	169	44	4	333	173	34	5	366	189	40	4	339	184	35	1
1986–1995	346	178	36	6	379	189	46	6	363	176	29	3	382	183	43	4
1996–2005	363	176	27	4	366	173	29	1	349	179	30	2	376	187	40	4

*Significant at 0.05, i.e., too few or too many exceedances over the threshold during the period listed in Column 1.

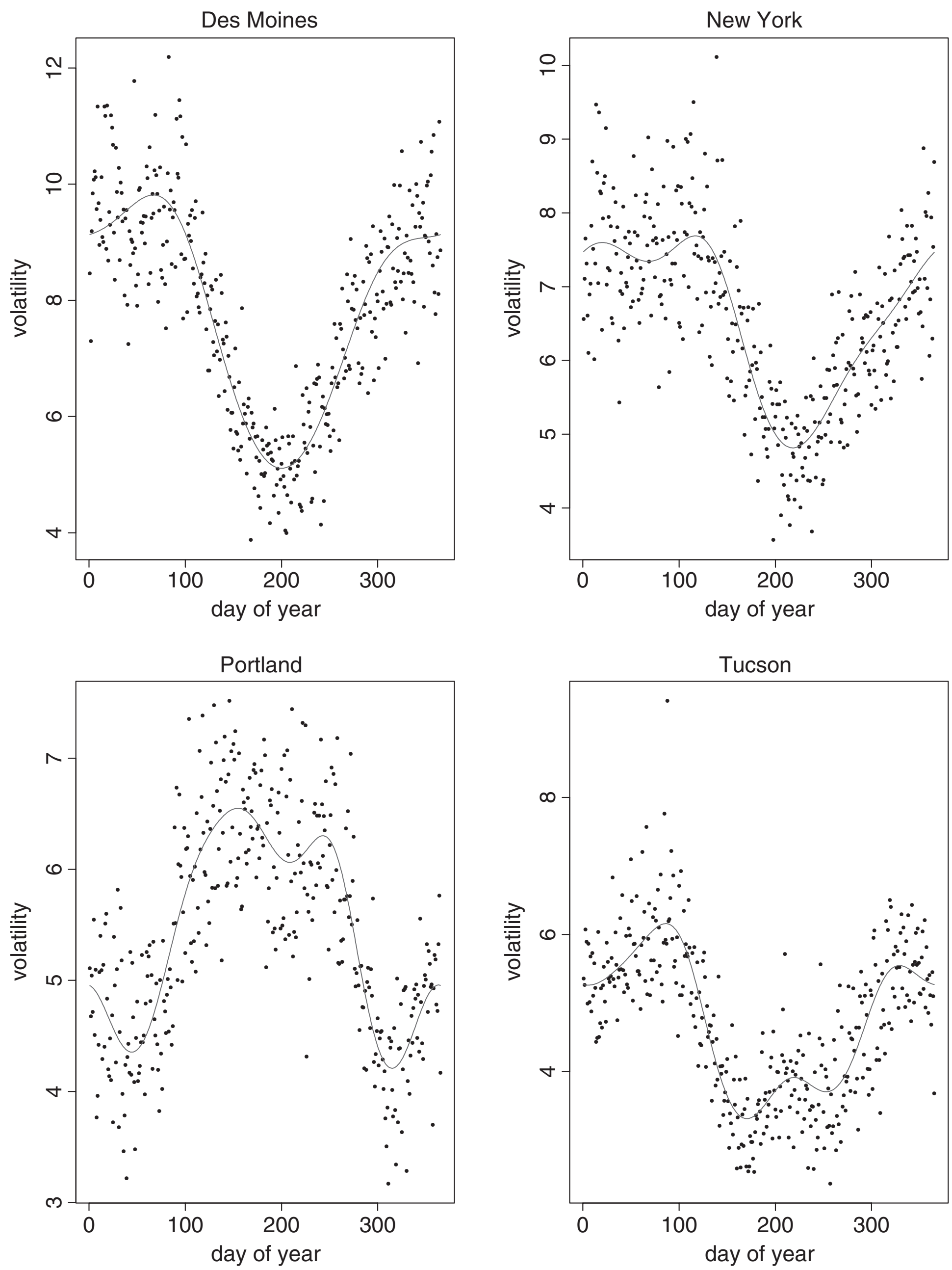

Figure 2. Observed and estimated volatility of detrended and deseasonalized daily maximum temperature for Des Moines, New York, Portland, and Tucson, 1956–2005.

Table 4. Number of exceedances of residuals Z_t over high thresholds $u_{0.10}$ [1], $u_{0.05}$ [2], $u_{0.01}$ [3], and $u_{0.001}$ [4]. Expected number of exceedances over the four thresholds are approximately 455, 227.5, 45.5, and 4.55, respectively. Based on data for 1956–2005

	Des Moines				New York				Portland				Tucson			
	[1]	[2]	[3]	[4]	[1]	[2]	[3]	[4]	[1]	[2]	[3]	[4]	[1]	[2]	[3]	[4]
Winter	510**	290**	62*	5	482	250	61*	5	450	246	70**	9	513**	268**	37	2
Spring	448	206*	53	8	467	241	61*	8	504*	244	24**	2	446	221	47	5
Summer	395**	184**	32*	1	422	193*	37	5	405**	190**	27**	0	435	218	55	7
Fall	472	233	36	5	454	229	24**	1	466	233	62*	8	431	206	44	5

*Significant at 0.05; **significant at 0.01, i.e., too few or too many exceedances over the threshold during the period listed in Column 1.

The estimated change year η for each estimated season, as obtained following the optimization procedure laid out in Section 3.2, also appears in Figure 3. There is no consensus on one particular change year across the cities, but five of the nine estimated years are in the 1970s. Table 5 shows the remaining estimated GPD parameters for model (i). Note that $\xi = 0$ corresponds to exponential tails, whereas $\xi < 0$ and $\xi > 0$ correspond to finite upper bound and heavy tails, respectively. Thus, a change from a very large in absolute value negative estimate for ξ, for example, $\hat{\xi}_1 = -0.49(0.14)$ in pre-1975 season 2 of New York, to a zero or positive ξ, for example, $\hat{\xi}_2 = 0.05(0.17)$ in post-1975 season 2 of New York, indicates that large exceedances above normal mean daily maximum levels (which we have already allowed to increase over time) are more likely post-1975 than they were pre-1975 during this season. This change in time toward a greater possibility of daily maximum temperatures far above seasonal maxima is most pronounced for seasons 1 and 3 of Des Moines and season 2 of New York.

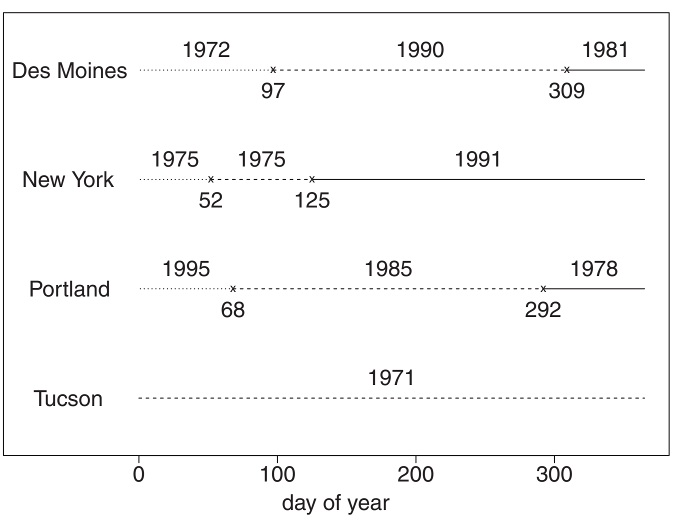

Figure 3. Estimated change point day τ (under segment) marking the change of seasons during the year and estimated change year η (over segment) at which there was a change in the shape parameter of the GPD of scaled residuals for each season for Des Moines, New York, Portland, and Tucson, 1956–2005. For example, we estimate that there are three seasons in Des Moines: January 1 to April 7 (Day 97), April 8 to November 7 (Day 309), and November 8 to December 31. And, the GPD has shape parameter ξ_1 for January 1 to April 8 data up to 1972, and shape parameter ξ_2 for January 1 to April 8 data after 1972. See Table 5.

Furthermore, while the scaled residuals show no evidence of serial correlation, it is important to verify the level of dependence at extreme levels as it might be quite different from that which occurs at mean levels. We look for any evidence of clustering within the extreme values of Z_t by estimating the extremal index θ (Leadbetter 1983). Estimated values of θ following the intervals estimator of Ferro and Segers (2003) are listed in Table 5. In general, there is very weak dependence at the extreme levels (dependence at this extreme level is negligible when $\theta = 1$) with the strongest dependence appearing over the colder months. We will nonetheless account for this dependence in our simulation-based approach for the estimation of return levels for heat waves in Section 4.3 as dependence in the extremes is central to the idea of a heat wave.

Finally, note that the presence of seasonal change points to delimit periods of constant exceedance rates is quite intuitive and is consistent with the underlying meteorological mechanisms. The functional form found for the GPD shape and scale parameters is also quite simple and agrees with our intuition regarding changes in extreme temperature over the observed period. It is thus reasonable to think that most, if not all, of the nonstationarity of $\{Y_t\}$ has been removed, or at least simplified, and that the use of standard methods of analyses for the extremes of $\{Z_t\}$ is justified.

4.2 Goodness of Fit

QQ-plots of estimated marginal return levels against the observed data, along with 95% bootstrapped confidence intervals, are shown in Figure 4. The fits are excellent for all cities. We further assess the goodness of fit by looking at the fit of the conditional model as described in Section 3.3. We get out-of-sample next-day daily maximum temperature forecasts over a period of 30 years, 1976–2005, using a moving window of only 20 years of historical data. We produce out-of-sample results for 30 years so that we can aggregate over different criteria and investigate possible lack of fit over many facets. Figure 5 shows the histograms of PIT by month over all years in 1976–2005 for the city of New York. All forecasts are good, and this is true no matter how we aggregate and for all cities (other plots not shown). We are confident that our full preprocessing model is capturing the behavior of daily maximum temperature of these four cities over the 50-year period studied.

Table 5. Estimated GPD parameters. ξ_1 (ξ_2) is the estimated shape parameter for the period before (after) the change year η shown in Figure 3

	Season	u_z	β	ξ_1	ξ_2	θ	$n_1 : n_2$
Des Moines	1	2.11	0.41 (0.06)	−0.34 (0.12)	−0.01 (0.12)	0.91	28 : 69
	2	1.84	0.41 (0.04)	−0.09 (0.06)	−0.44 (0.08)	0.95	162 : 50
	3	2.00	0.51 (0.10)	−0.52 (0.17)	−0.03 (0.19)	0.82	33 : 23
New York	1	2.03	0.60 (0.10)	−0.38 (0.17)	−0.65 (0.13)	0.80	15 : 37
	2	2.35	0.49 (0.08)	−0.49 (0.14)	0.05 (0.17)	0.92	34 : 39
	3	1.99	0.37 (0.03)	−0.11 (0.07)	−0.04 (0.10)	0.98	173 : 67
Portland	1	2.25	0.47 (0.09)	−0.06 (0.15)	0.18 (0.36)	0.83	59 : 9
	2	1.94	0.35 (0.03)	0.03 (0.08)	−0.20 (0.06)	0.99	137 : 87
	3	2.30	0.51 (0.09)	−0.05 (0.18)	−0.43 (0.13)	0.96	37 : 36
Tucson	–	1.65	0.31 (0.02)	0.19 (0.09)	−0.10 (0.05)	0.93	100 : 265

Estimated extremal index θ at threshold u_z is in the second to last column. Last column gives the number of exceedances before (n_1) and after (n_2) the change year η.

4.3 Heat Waves

Having established the good fit of our models, we now use them to obtain estimates for the recurrence probabilities of heat waves by simulation. Return periods (i.e., one over annual recurrence probabilities) are shown for heat waves of different intensities and different lengths in Figure 6. Abscissae are chosen for each city to show events of similar return periods in all cases. Estimates are based on 500,000 simulations and thus the approximate Monte Carlo sampling error $(1/p^2)\sqrt{p(1-p)/500{,}000}$ for the $1/p$-year return event is no greater than 1.4, 45, and 1,414 years for 100-, 1,000-, and 10,000-year return events, respectively. Estimates are shown for 1957 and 2005. We compute estimates for 1957 since estimates for 1956 can only be obtained as for the $(R+1)$th day of January since the use of an AR(R) means that we need R days to

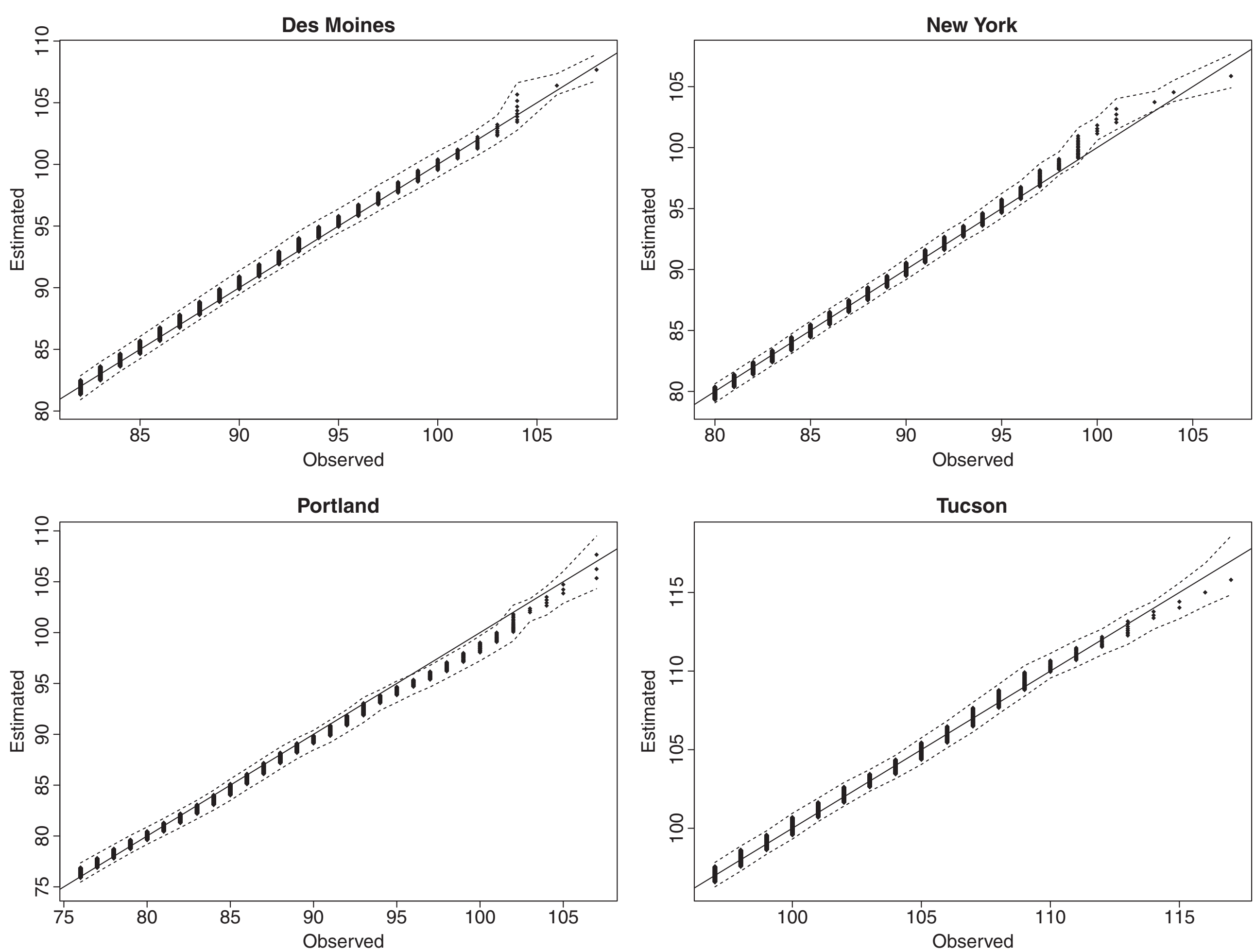

Figure 4. QQ-plots of estimated marginal return levels against observed data using the full preprocessing model for Des Moines, New York, Portland, and Tucson. Top 22% of the data are shown. Dotted lines indicate 95% bootstrapped confidence intervals; dashed line is 45° line and indicates a perfect fit.

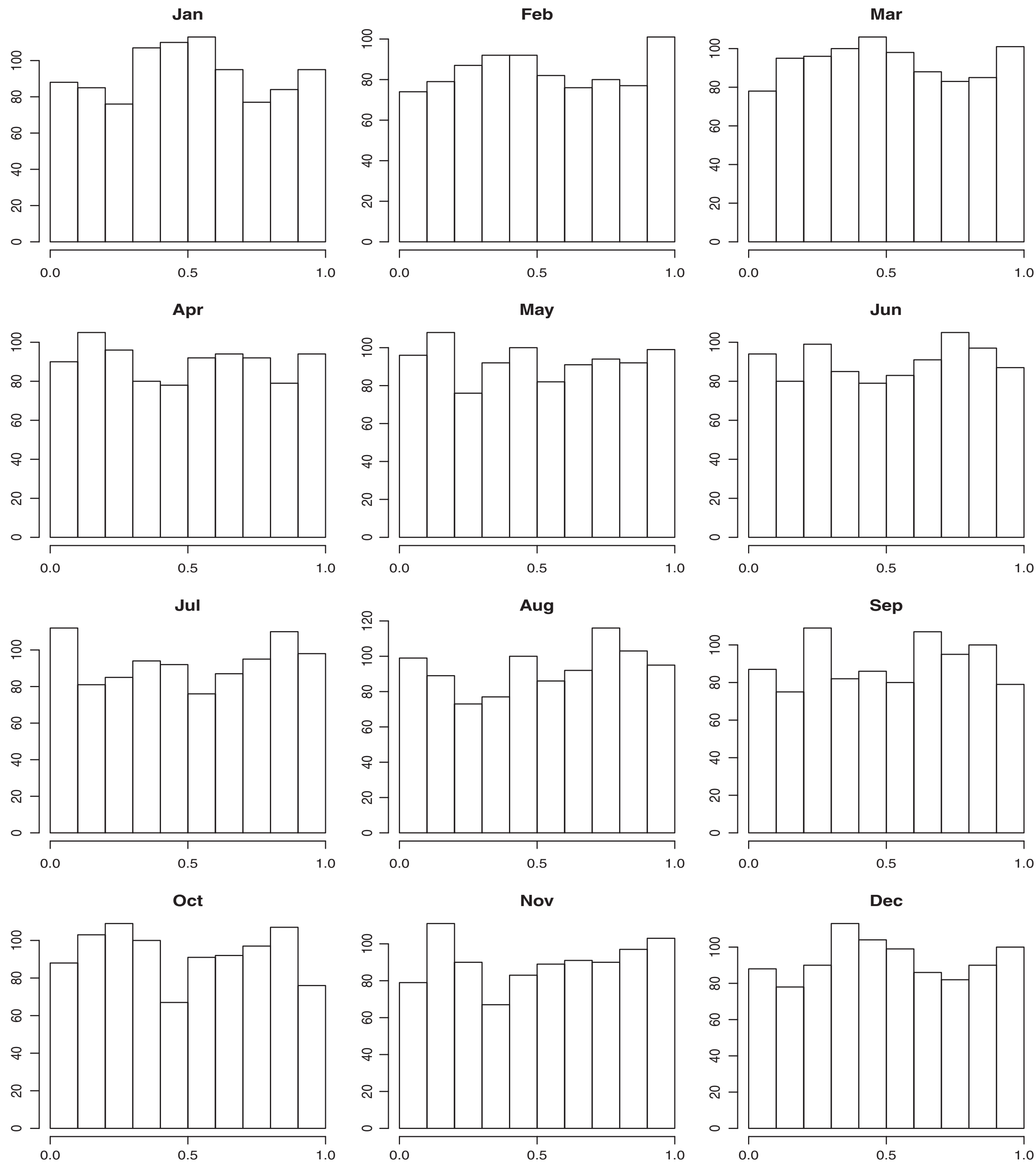

Figure 5. Histograms of PIT by month for full preprocessing model forecasts for all days over 1976–2005, for the city of New York. Forecasts based on 20-year moving window.

start the simulated series. First, we note that while heat waves have become significantly more intense and more frequent in all four cities, the 50-year evolution is not the same in the four cases. In Des Moines and Portland, 2005 return periods for 5-day heat waves above the temperatures shown are about one half of their 1957 values. In 1957, a 5-day run above 99°F was an $\approx$ 250-year event in Des Moines; in 2005, it had become an $\approx$ 115-year event. In Portland, the changes are similar: a 5-day run above 98°F was an $\approx$ 300-year event in 1957 and an $\approx$ 160-year event in 2005. Changes in 2-day runs for the two cities are a bit smaller as 2005 return periods are ≈ 0.70 of their 1957 values. Increases in heat wave intensity and frequency are greater for New York. In 1957, a 5-day run above 96°F was a 306-year event in New York; in 2005, it had become a 105-year event. Return periods in 2005 of 2-day runs for the temperatures shown are 0.4–0.6 of their 1957 values. Finally, an already very hot Tucson is getting hotter. Heat waves have become more intense and more frequent over the 50-year period, and changes

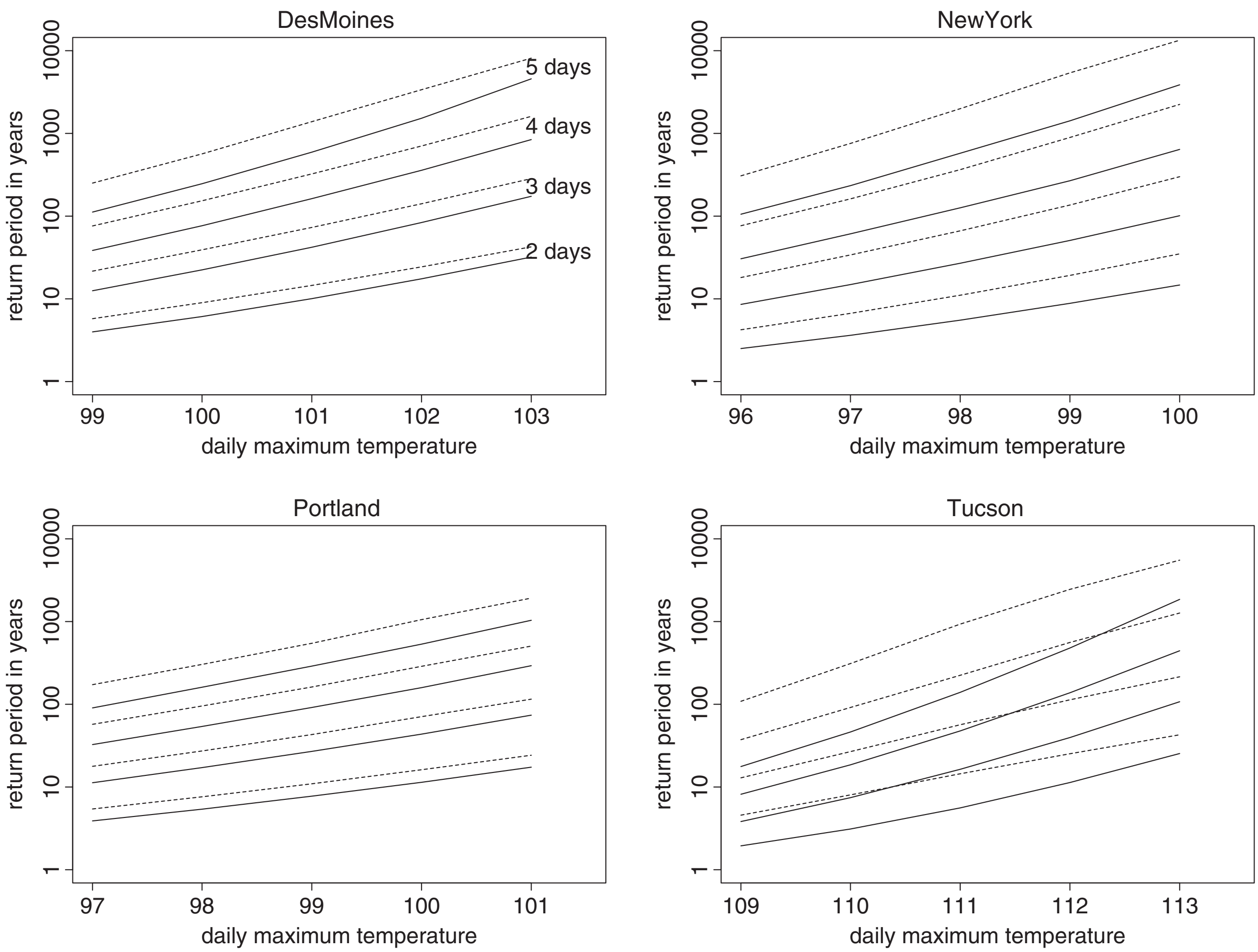

Figure 6. Estimated return periods of heat waves for any time of the year in Des Moines, New York, Portland, and Tucson. Dotted lines are for 1957, solid lines are for 2005. Lines are for 2-day, 3-day, 4-day, and 5-day heat waves (bottom to top) where daily maximum temperature (in °F) was at least the value on the abscissa.

are greater here than in the other three cities. In 1957, a 5-day run above 110°F was a 310-year event in Tucson; in 2005, it had become a 46-year event. Fortunately, the increase in frequency lessens with increased intensity: in 1957, a 5-day run above 113°F was a 5,500-year event; in 2005, it had become a 1,850-year event. Similarly, year 2005 values of return periods of 2-day runs above 109°F and 110°F are about 0.40 of their 1957 values, whereas a 2-day run above 113°F was a 42-year event in 1957, and a 25-year event in 2005.

Recall that one of the strengths of our full preprocessing approach is that it allows us to compute return periods for any time of the year, not just over the entire year. Figure 7 shows the estimated return periods for March–April heat waves. For all four cities, heat waves in these months have incurred smaller reductions in return periods over the 50 years than those seen on an annual level. City comparisons are similar to those on an annual level: changes over 50 years are relatively the same in Des Moines and Portland, a bit greater in New York, and much greater for certain temperatures in Tucson.

Figure 8 shows the estimated return periods for September–October heat waves. While heat waves in these months have incurred roughly the same reductions in return periods over the 50 years than those seen on an annual level for Des Moines, New York, and Portland, the reductions are even greater over these months than they were at the annual level in Tucson. A 5-day run above 103°F in September–October was a 85-year event in 1957; in 2005, it had become a 18-year event. And, the increase in frequency is maintained for increased intensity: a 5-day run above 105°F in September–October was a 600-year event in 1957; in 2005, it had become a 100-year event.

5. DISCUSSION AND CONCLUSIONS

This article tackles inferences for a complex extreme event: heat waves. A complete model that enables the estimation of the recurrence probability of a heat wave of any intensity and any length over any period of the year is presented. We define heat waves as a period of consecutive days where daily maximum temperature is above a prespecified high value and choose a preprocessing approach to model our nonstationary daily maximum temperature series. We could model daily minimum temperatures in an analogous fashion. This would allow us to address heat waves that are defined as a run of days where both daily minimum and maximum temperatures were above prespecified temperatures.

Ballester, Giorgi, and Rodó (2010) argued that shifts in European temperature extremes can be more reliably derived

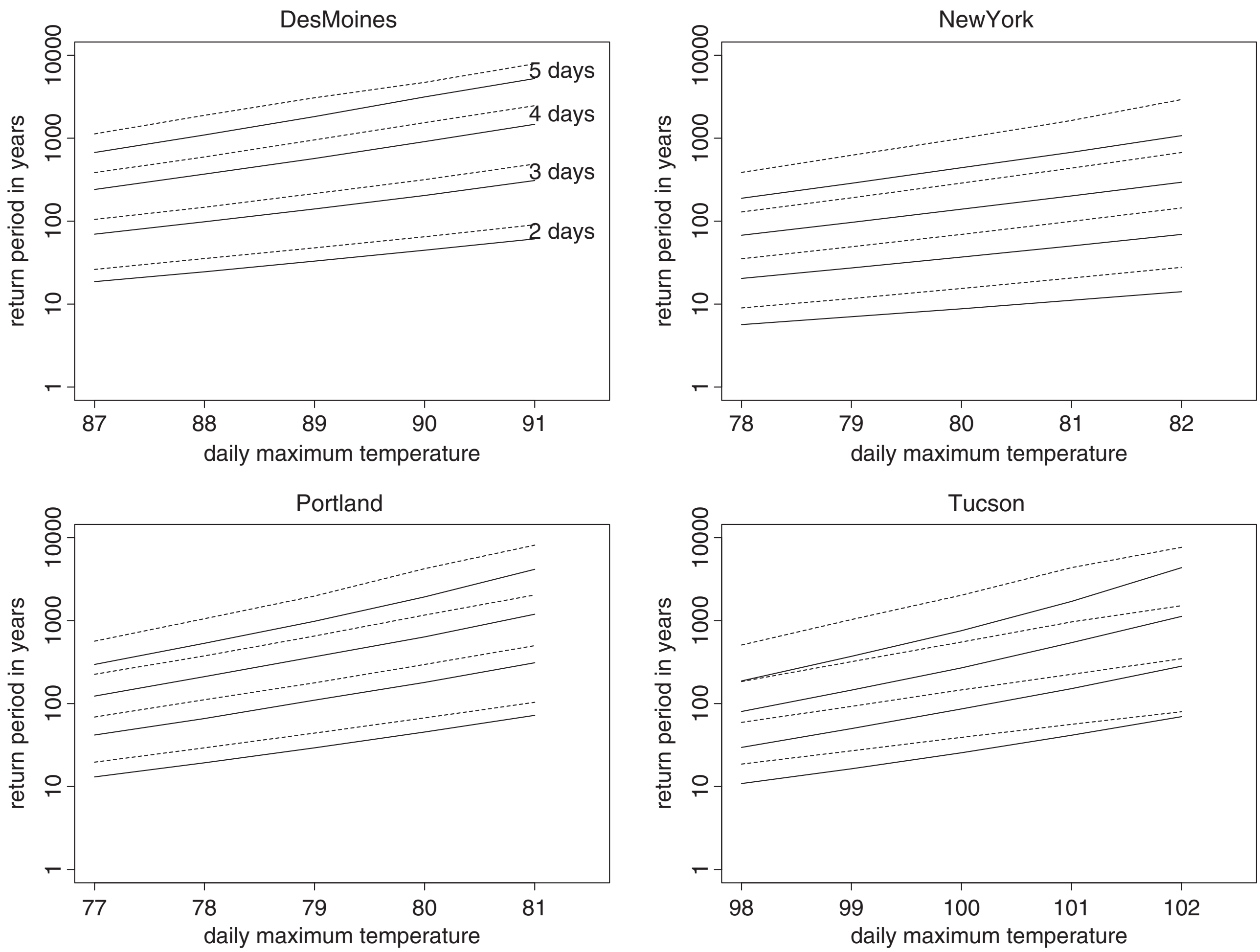

Figure 7. Estimated return periods of heat waves in March–April for Des Moines, New York, Portland, and Tucson. Dotted lines are for March–April 1957, solid lines are for March–April 2005. Lines are for 2-day, 3-day, 4-day, and 5-day heat waves (bottom to top) where daily maximum temperature (in °F) was at least the value on the abscissa.

indirectly from changes in the overall probability distribution of a climate variable (e.g., shifts in the mean and standard deviation) than through direct statistical modeling of extremes. Our model includes a long-term trend in the mean. We checked for the presence of a long-term trend in the variance in several ways. We tried including a time trend term in (6) in several ways: linear in time, linear by year, and step-like functions over half or full decades. None of these models were favored by the BIC criterion. Furthermore, there is no evidence of an increase of variances when we examine calendar-day variances of estimated scaled residuals over time. As our subsequent modeling of scaled residuals showed a significant change over time of their tail distribution, we show that the Ballester et al. (2010) arguments do not hold for our four U.S. cities.

We have only included a time covariate in our model for daily maximum temperature; however, other covariates could be useful when modeling heat waves. Heat waves are generally associated with high dew point temperatures, high urban island effects, high solar radiation levels, and light winds (Kunkel et al. 1996; Palecki, Changnon, and Kunkel 2001). It thus seems plausible that accounting for rainfall in previous days, differences in temperature (daily maxima or nightly minima) between urban and rural stations, solar radiation levels, and average daily wind speed could improve heat wave models. This is a topic of future research.

Katz (2010) highlighted the lack of use of the statistics of extremes in climate change research, more particularly in the modeling of heat wave events. As our model provides an excellent fit to observed data, it represents a large step toward filling the void in climate change research. The model can be applied to GCM data and if good fits are obtained, we could then properly quantify projected changes in heat wave intensity and frequency under different SRES emissions scenarios. This is a topic of future research.

Finally, it is interesting to look at estimated return periods in 1957 and 2005 of recent heat waves. In New York, the daily maximum temperature was at least 95°F from August 29, 2010, to September 2, 2010. We can use our fitted models to estimate the return period of a 5-day run of at least 95°F from, for example, August 15 to September 15. We find that the latter was a 6,460-year event in 1957 and a 1,450-year event in 2005. Clearly, we have not seen the end of increases in heat wave intensity and frequency and our fitted models need to be continuously updated with the most recent data for estimated return periods to accurately reflect current heat stress loads.

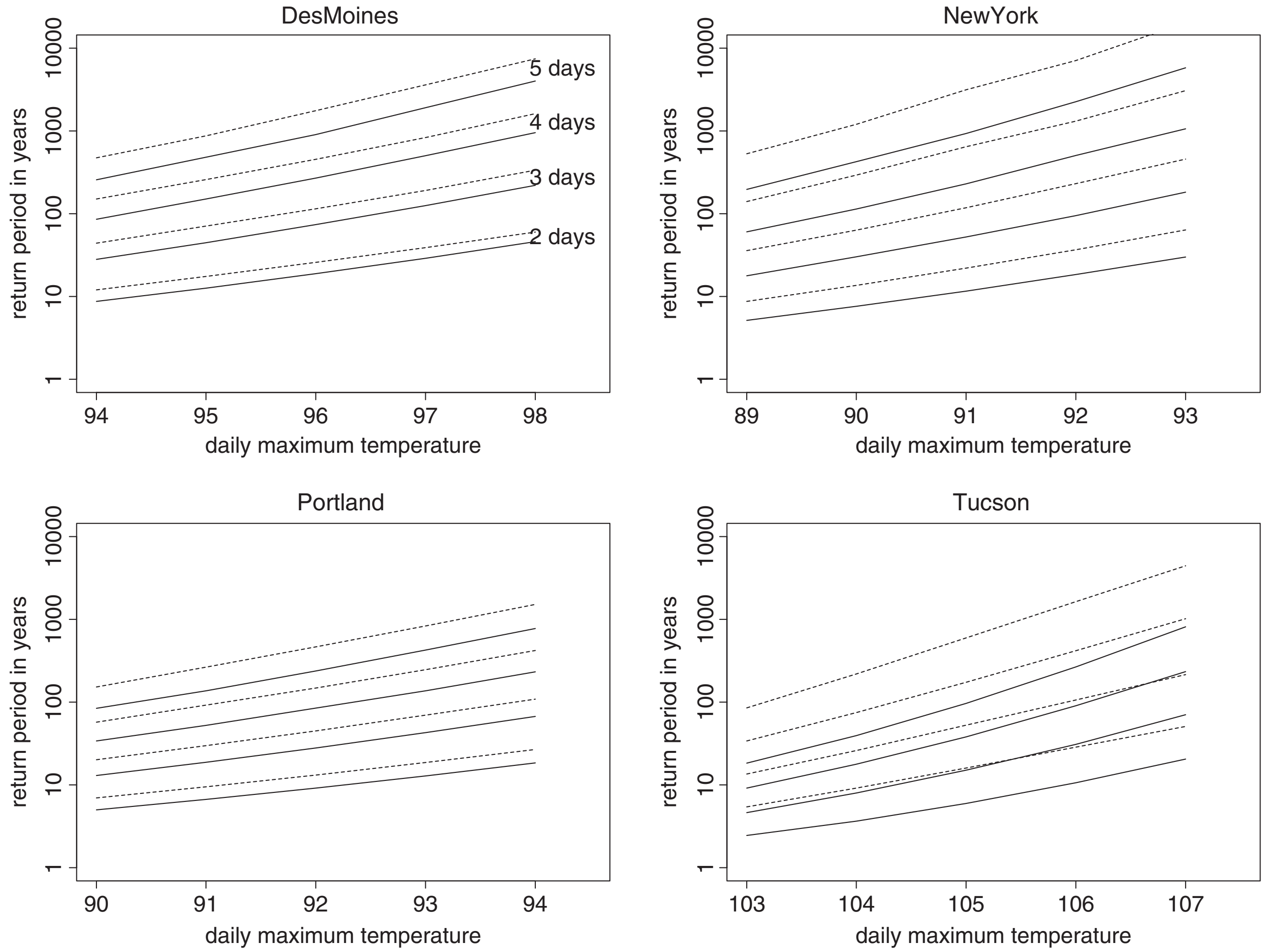

Figure 8. Estimated return periods of heat waves in September–October for Des Moines, New York, Portland, and Tucson. Dotted lines are for September–October 1957, solid lines are for September–October 2005. Lines are for 2-day, 3-day, 4-day, and 5-day heat waves (bottom to top) where daily maximum temperature (in °F) was at least the value on the abscissa.

APPENDIX: ESTIMATING CHANGE-POINTS

Hawkins (2001) developed an exact approach for finding MLE of change points and within-segment parameters when the within-segment distribution is from the general exponential family. He showed that the likelihood is separable and that optimal splits can be found with a DP algorithm.

The GPD is not in the general exponential family, so the approach does not apply directly. However, focusing first on the exceedance (or not) of a high threshold, we create Bernoulli-distributed variates and use the approach to find optimal splits of the year into seasons with constant threshold crossing rates. We create the series of Bernoulli variates $\{W_t\}$ where $\{W_t\} = I(Z_t > z_p)$, where z_p is the upper quantile defined such that $\Pr(Z_t > z_p) = p$ and p is small. We choose $p = 0.02$.

Consider a number of change points, $\tau_1, \tau_2, \ldots, \tau_k$ such that the observations $\{W_t\}$ with $\tau_{j-1} < d(t) \le \tau_j$ follow a Bernoulli distribution with parameter p_j. If the sequence $W_{h+1}, \ldots, W_m$ follows a Bernoulli distribution with parameter p, the MLE of p is $\overline{W}_{h,m} = \sum_{i=h+1}^{m} W_i/(m-h)$. Write $Q(h,m)$ for -2 times the maximized log-likelihood obtained by substituting $\overline{W}_{h,m}$ for p in the log-likelihood of this subsequence of the data. We have

$$Q(h,m) = -2(m-h)\Big[\overline{W}_{h,m} \log \overline{W}_{h,m} + (1-\overline{W}_{h,m}) \times \log(1-\overline{W}_{h,m})\Big]. \qquad (7)$$

We apply the Hawkins (2001) DP algorithm:

For each $m = 1, \ldots, n$, calculate $F(1,m) = Q(0,m)$.
For each $r = 2, \ldots, k$, calculate $F(r,m)$, $m = 1, \ldots, n$, where

$$F(r,m) = \min_{0<h<m} [F(r-1,h) + Q(h,m)].$$

For each $F(r,m)$, we keep a record of $H(r,m)$, the h value yielding the minimum.

The maximized log-likelihood of the k segment model fitted to the full data is $-\frac{1}{2}F(k,n)$. The estimates $\widehat{\tau}_j$ are given by the DP back-tracing operation $\widehat{\tau}_k = n$, and for $r = (k-1), \ldots, 1$, $\widehat{\tau}_r = H(r+1, \widehat{\tau}_{r+1})$.

The Hawkins (2001) DP algorithm does not find the optimal number of change points, but rather the optimal k change points for a fixed k. We do not fix k a priori and initially allow for an arbitrary number of change points, but we stop adding change points when one of the *seasons* that they define contains too few days. Since the GPD will be fitted to exceedances over high thresholds of data within a season, a season cannot be too short, otherwise there will be an insufficient number of exceedances to enable the fitting of the GPD. When using 50 years of data, we found that we could allow a season to consist of as few as 30 days and had no problem obtaining our GPD fits to the resulting 30 $(= 50 \times 30 \times 0.02)$ exceedances. While choosing too small a p is not useful since then most Bernoulli variables are equal to zero and no change points are found, we found that $0.02 \leq p \leq 0.10$ gives approximately the same good performance.

[Received October 2010. Revised July 2011.]

REFERENCES

Abaurrea, J., Asín, J., Cebrián, A. C., and Centelles, A. (2007), "Modeling and Forecasting Extreme Hot Events in the Central Ebro Valley, a Continental-Mediterranean Area," *Global and Planetary Change*, 57, 43–58. [25]

Australian Government (2009), *Climate-Change Health Risk Assessments and Burdens for Australia* [on-line], Canberra: National Health and Medical Research Council. Available at *http://www.nhmrc.gov.au/your_health/climate_change* (retrieved 2010-09-01). [24]

Ballester, J., Giorgi, F., and Rodó, J. (2010), "Changes in European Temperature Extremes Can Be Predicted From Changes in PDF Central Statistics: A Letter," *Climatic Change*, 98, 277–284. [35]

Beniston, M. (2005), "Warm Winter Spells in the Swiss Alps: Strong Heat Waves in a Cold Season?," *Geophysical Research Letters*, 32, L01812. [25]

Bloomberg. (2010), "Russia Heat Wave May Kill 15,000, Shave $15 Billion of GDP." Available at *http://www.bloomberg.com/news/2010-08-10/russia-may-lose-15-000-lives-15-billion-of-economic-output-in-heat-wave.html* (retrieved 2010-09-03). [24]

Coelho, C. A. S., Ferro, C. A. T., Stephenson, D. B., and Steinskog, D. J. (2008), "Methods for Exploring Spatial and Temporal Variability of Extreme Events in Climate Data," *Journal of Climate*, 21, 2072–2092. [25,26]

Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, London: Springer-Verlag. [25]

Coles, S., and Pericchi, L. (2003), "Anticipating Catastrophes Through Extreme Value Modelling," *Journal of the Royal Statistical Society,* Series C, 52, 405–416. [27]

Cooley, D. (2009), "Extreme Value Analysis and the Study of Climate Change," *Climatic Change*, 97, 77–83. [25]

Davison, A. C., and Smith, R. L. (1990), "Models for Exceedances Over High Thresholds," *Journal of the Royal Statistical Society,* Series B, 52(3), 393–442. [25,26]

Dawid, A. P. (1984), "Statistical Theory: The Prequential Approach" (with discussion), *Journal of the Royal Statistical Society,* Series A, 147, 278–292. [28]

Della-Marta, P. M., Haylock, M. R., Luterbacher, J., and Wanner, H. (2007), "Doubled Length of Western European Summer Heat Waves Since 1880," *Journal of Geophysical Research-Atmospheres*, 112, ARTN D15103. [24]

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998), "Evaluating Density Forecasts With Applications to Financial Risk Management," *International Economics Review*, 39, 863–883. [28]

Easterling, D. R., Evans, J. L., Groisman, P. Ya, Karl, T. R., Kunkel, K. E., and Ambenje, P. (2000), "Observed Variability and Trends in Extreme Climate Events: A Brief Review," *Bulletin of the American Meteorological Society*, 81, 417–425. [24]

Eastoe, E. F., and Tawn, J. (2009), "Modelling Non-Stationary Extremes With Application to Surface Level Ozone," *Applied Statistics*, 58, 1, 25–45. [25,26]

Ferro, C. A. T., and Segers, J. (2003), "Inference for Clusters of Extreme Values," *Journal of the Royal Statistical Society,* Series B, 65, 545–556. [32]

Frich, P., Alexander, L. V., Della-Marta, P., Gleason, B., Haylock, M., Klein Tank, A. M. G., and Peterson, T. (2002), "Observed Coherent Changes in Climatic Extremes During the Second Half of the Twentieth Century," *Climate Research*, 19, 193–212. [24]

Ganguly, A. R., Steinhaeuser, K., Erickson, III, D. J., Branstetter, M., Parish, E. S., Singh, N., Drake, J. B., and Buja, L. (2009), "Higher Trends but Larger Uncertainty and Geographic Variability in 21st Century Temperature and Heat Waves," *Proceedings of the National Academy of Sciences of the United States*, 106, 15555–15559. [24]

Hajat, S., Armstrong, B., Baccini, M., Biggeri, A., Bisanti, L., Russo, A., Paldy, A., Menne, B., and Kosatsky, T. (2007), "Impact of High Temperatures on Mortality: Is There an Added Heat Wave Effect?" *Epidemiology*, 17, 632–638. [24]

Hajat, S., Kovats, R. S., Atkinson, R. W., and Haines, A. (2002), "Impact of Hot Temperatures on Death in London: A Time Series Approach," *Journal of Epidemiology and Community Health*, 56, 367–372. [25]

Hawkins, D. M. (2001), "Fitting Multiple Change-Point Models to Data," *Computational Statistics & Data Analysis*, 37, 323–341. [28,37]

Hunt, B. G. (2007), "A Climatology of Heat Waves From a Multimillennial Simulation," *Journal of Climate*, 20, 3802–3821. [24]

IPCC (2007), *Climate Change 2007: Synthesis Report*, Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change Core Writing Team, eds. R. K. Pachauri and A. Reisinger, GenevaSwitzerland: IPCC, 104 pp. [24]

Kalkstein, L. S., and Davis, R. E. (1989), "An Evaluation of Demographic and Interregional Responses in the United States," *Annals of the Association of American Geographers*, 79, 44–64. [24]

Kalkstein, L. S., and Greene, J. S. (1997), "An Evaluation of Climate/Mortality Relationships in Large U.S. Cities and the Possible Impacts of a Climate Change," *Environmental Health Perspectives*, 105, 84–93. [24]

Katz, R. W. (2010), "Statistics of Extremes in Climate Change," *Climatic Change*, 100, 71–76. [36]

Kharin, V. V., and Zwiers, F. W. (2005), "Estimating Extremes in Transient Climate Change Simulations," *Journal of Climate*, 18, 1156–1173. [25]

Knowlton, K., Lynn, B., Goldberg, R. A., Rosenzweig, C., Hogrefe, C., Rosenthal, J. K., and Kinney, P. L. (2007), "Projecting Heat-Related Mortality Impacts Under a Changing Climate in the New York City Region," *American Journal of Public Health*, 97, 2028–2034. [24]

Koffi, B., and Koffi, E. (2008), "Heat Waves Across Europe by the End of the 21st Century: Multiregional Climate Simulations," *Climate Research*, 36, 153–168. [24]

Kosatsky, T. (2005), "The 2003 European Heat Waves," *Euro Surveillance*, 10, 148–149. [24]

Kovats, R. S., and Hajat, S. (2008), "Heat Stress and Public Health: A Critical Review," *Annual Review of Public Health*, 29, 41–55. [24]

Kuglitsch, F. G., Toreti, A., Xoplaki, E., Della-Marta, P., Zerefos, C. S., Türkes, M., and Luterbacher, J. (2010), "Heat Wave Changes in the Eastern Mediterranean Since 1960," *Geophysical Research Letters*, 37, L04802. [24]

Kunkel, K. E., Changnon, S. A., Reinke, B. C., and Arritt, R. W. (1996), "The July 1995 Heat Wave in the Midwest: A Climatic Prespective and Critical Weather Factors," *Bulletin of the American Meteorological Society*, 77, 1507–1518. [36]

Kyselý, J. (2010), "Recent Severe Heat Waves in Central Europe: How to View Them in a Long-term Prospect?" *International Journal of Climatology*, 30, 89–109. [25,30]

Kyselý, J., Picek, J., and Beranová, R. (2010), "Estimating Extremes in Climate Change Simulations Using the Peaks-Over-Threshold Method With a Non-Stationary Threshold," *Global and Planetary Change*, 72, 55–68. [25,26]

Leadbetter, M. R. (1983), "Extreme and Local Dependence in Stationary Sequences," *Z. Wahrsch. Verw. Gebiete*, 65, 291–306. [32]

Macchiato, M., Serio, C., Lapenna, V., and La Rotonda, L. (1993), "Parametric Time Series Analysis of Cold and Hot Spells in Daily Temperature: An Application in Southern Italy," *Journal of Applied Meteorology*, 32, 1270–1281. [25,30]

McMichael, A. J., Woodruff, R. E., and Hales, S. (2006), "Climate Change and Human Health: Present and Future Risks," *Lancet*, 367, 859–869. [24]

Mearns, L. O., Katz, R. W., and Schneider, S. H. (1984), "Extreme High-Temperature Events: Changes in Their Probabilities With Changes in Mean Temperature," *Journal of Climate and Applied Meteorology*, 23, 1601–1613. [24,25,30]

Meehl, G. A., and Tebaldi, C. (2004), "More Intense, More Frequent, and Longer Lasting Heat Waves in the 21st Century," *Science*, 305, 994–997. [24]

Nogaj, M., Yiou, P., Parey, S., Malek, F., and Naveau, P. (2006), "Amplitude and Frequency of Temperature Extremes Over the North Atlantic Region," *Geophysical Research Letters*, 33, L10801. [25]

Palecki, M. A., Changnon, S. A., and Kunkel, K. E. (2001), "The Nature and Impacts of the July 1999 Heat Wave in the Midwestern United States: Learning From the Lessons of 1995," *Bulletin of the American Meteorological Society*, 82, 1353–1367. [36]

Pauli, F., and Rizzi, L. (2008), "Analysis of Heat Wave Effects on Health by Using Generalized Additive Model and Bootstrap-Based Model Selection," *Applied Statistics*, 57, 473–485. [24]

Piao, S. L., Ciais, P., Friedlingstein, P., Peylin, P., Reichstein, M., Luyssaert, S., Margolis, H., Fang, J. Y., Barr, A., Chen, A. P., Grelle, A., Hollinger, D. Y., Laurila, T., Lindroth, A., Richardson, A. D., and Vesala, T. (2008), "Net Carbon Dioxide Losses of Northern Ecosystems in Response to Autumn Warming," *Nature*, 451, 49–53. [25]

Pickands, J. (1975), "Statistical Inference Using Extreme Order Statistics," *The Annals of Statistics*, 3, 119–131. [26]

Schär, C., Vidale, P. L., Lüthi, D., Frei, C., Häberli, C., Liniger, M. A., and Appenzeller, C. (2004), "The Role of Increasing Temperature Variability in European Summer Heatwaves," *Nature*, 427, 332–336. [24]

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464. [27]

Semenza, J. C., Rubin, C. H., Falter, K. H., Selanikio, J. D., Flanders, W. D., Howe, H. L., and Wilhelm, J. L. (1996), "Heat-Related Deaths During the

July 1995 Heat Wave in Chicago," *The New England Journal of Medicine*, 335(2), 84–90. [24]

Smith, R. L. (1989), "Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground-Level Ozone," *Statistical Science*, 4, 367–377. [26,27]

——— (2000), "Measuring Risk With Extreme Value Theory," in *Extremes and Integrated Risk Management* (Chap. 2), ed. P. Embrechts, London: Risk Books, pp. 19–35. [27]

Smoyer-Tomic, K. E., Kuhn, R., and Hudson, A. (2003), "Heat Wave Hazards: An Overview of Heat Wave Impacts in Canada," *Natural Hazards*, 28, 463–485. [24]

Tebaldi, C., Hayhoe, K., Arblaster, J., and Meehl, G. (2006), "Going to the Extremes," *Climatic Change*, 79, 185–211. [24]

UN-Habitat (2010), *Urban Trends: Urbanization and Economic Growth* [on-line], Nairobi, Kenya: United Nations Human Settlements Programme. Available at *http://www.unhabitat.org/documents/SOWC10/R7.pdf* (retrieved 2010-09-01). [24]

United Nations Population Division (2004), *World Population 2004* [on-line], New York: Department of Economic and Social Affairs, United Nations Population Division. Available at *http://www.un.org/esa/population/publications/WPP2004/World_Population_2004_chart.pdf* (retrieved 2010-09-01). [24]

U.S. Census Bureau (2009), "Annual Estimates of the Population of Metropolitan and Micropolitan Statistical Areas: April 1, 2000 to July 1, 2009." Available at *http://www.census.gov/popest/data/metro/totals/2009/tables/CBSA-EST2009-01.csv* (retrieved 2012-03-21). [29]

U.S. Environmental Protection Agency (2010), *Heat Island Effect* [on-line], Washington, DC: US EPA. Available at *http://www.epa.gov/heatisld/* (retrieved 2010-09-01). [24]

U.S. Global Change Research Program (2010), "Human Health. Washington D.C.: U.S. Global Change Research Program." Available at *http://www.globalchange.gov/publications/reports/scientific-assessments/us-impacts/climate-change-impacts-by-sector/human-health* (retrieved 2010-09-01). [24]

Zivot, E., and Wang, J. (2002), *Modeling Financial Time Series With S-PLUS*, New York: Springer. [28]

Zwiers, F., and Kharin, V. (1998), "Changes in the Extremes of the Climate Simulated by CCC GCM2 Under CO_2 Doubling," *Journal of Climate*, 11, 2200–2222. [25]