Statistical modeling of hot spells and heat waves

Eva M. Furrer[1,4,*], Richard W. Katz[1], Marcus D. Walter[2], Reinhard Furrer[3]

[1]Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research, Boulder, Colorado 80307, USA
[2]Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York 14850, USA
[3]Institute of Mathematics, University of Zürich, 8057 Zürich, Switzerland

[4]*Present address:* Institute for Social and Preventive Medicine, Biostatistics Unit, University of Zürich, 8001 Zürich, Switzerland

ABSTRACT: Although hot spells and heat waves are considered extreme meteorological phenomena, the statistical theory of extreme values has only rarely, if ever, been applied. To address this shortcoming, we extended the point process approach to extreme value analysis to model the frequency, duration, and intensity of hot spells. The annual frequency of hot spells was modeled by a Poisson distribution, and their length by a geometric distribution. To account for the temporal dependence of daily maximum temperatures within a hot spell, the excesses over a high threshold were modeled by a conditional generalized Pareto distribution, whose scale parameter depends on the excess on the previous day. Requiring only univariate extreme value theory, our proposed approach is simple enough to be readily generalized to incorporate trends in hot spell characteristics. Through a heat wave simulator, the statistical modeling of hot spells can be extended to apply to more full-fledged heat waves, which are difficult to model directly. Our statistical model for hot spells was fitted to time series of daily maximum temperature during the summer heat wave season in Phoenix, Arizona (USA), Fort Collins, Colorado (USA), and Paris, France. Trends in the frequency, duration, and intensity of hot spells were fitted as well. The heat wave simulator was used to convert any such trends into the corresponding changes in the characteristics of heat waves. By being based at least in part on extreme value theory, our proposed approach is both more realistic and more flexible than techniques heretofore applied to model hot spells and heat waves.

KEY WORDS: Climate change · Clustering of extremes · Generalized Pareto distribution · Point process approach · Heat wave simulator

1. INTRODUCTION

Heat waves are meteorological events that have received much attention in recent years, given the mortality associated with them (Gosling et al. 2009) and given the specter of increases in their frequency, duration, and intensity as part of global climate change (Meehl & Tebaldi 2004). In particular, the high mortality associated with the 2003 European heat wave generated much concern about whether climate change is playing a role (Schär et al. 2004). Other recent heat waves of note include the 1995 event in Chicago, Illinois, USA (Karl & Knight 1997). Because of their rarity and their severity, such events are naturally viewed as 'extreme', but statistical methods based on extreme value theory (e.g. Coles 2001) have only rarely, if ever,

been applied to this type of meteorological event in realistic climate applications. Even the statistical analysis of projections of future changes in heat wave characteristics, on the basis of climate change experiments using numerical models of the climate system, has generally avoided any use of the statistics of extremes (Tebaldi et al. 2006, Koffi & Koffi 2008).

Yet there is a long tradition of using statistical methods based on extreme value theory in the analysis of simple extreme meteorological events, most commonly in the form of the highest daily precipitation amount over a year or the highest temperature over the summer season (Gumbel 1958). While such analyses typically assume stationarity (i.e. an unchanging climate), they are starting to be extended to the case of temporal trends (e.g. Katz et al. 2002). The so-called point pro-

cess approach is a parsimonious way to model possibly non-stationary extremes, jointly modeling the occurrence of an event (e.g. an exceedance of a high threshold) and its severity (e.g. an excess over a high threshold) (Smith 1989, Coles 2001). This approach has recently been applied to detect trends in high temperature extremes (Brown et al. 2008). Other meteorological applications using the point process model are included in Furrer & Katz (2008) and Katz et al. (2002). In such analyses, it is common to 'decluster' the data and model only cluster maxima to account for temporal dependence. In the present application, these clusters constitute hot spells whose characteristics need to be modeled as well (especially hot spell length and temporal dependence of excesses within a hot spell) rather than discarded.

Hot spells and, to an even greater extent, heat waves have a complex temporal structure that makes the application of extreme value theory less than routine. Although some analyses have made at least limited use of the theory, attempts to date have tended to be rather ad hoc, among other things tied to somewhat arbitrary definitions of hot spells or heat waves (Katsoulis & Hatzianastassiou 2005, Khaliq et al. 2005, 2007, Abaurrea et al. 2007). Part of the problem relates to the difficulty in defining a heat wave, involving a choice of threshold, a minimal duration, and possibly other variables besides daily maximum temperature (Robinson 2001, Meze-Hausken 2008). As will be seen, an approach focused on hot spells, which are simply defined as consecutive days with maximum temperature over a certain threshold, with the statistical modeling based at least in part on extreme value theory, results in sufficient flexibility to be applicable to a wide variety of more complicated definitions of a heat wave.

In the statistical modeling of hot spells, it is essential that the temporal dependence of extreme high daily maximum temperature be realistically modeled (Mearns et al. 1984, Kysely 2002). In the statistics literature, models based on bivariate extreme value theory have been proposed to account for the persistence of temperature at high (or low) levels (Coles et al. 1994). In the present paper, we propose a simpler, but closely related approach that only makes use of the more familiar univariate extreme value theory and readily available software. All calculations in this work have been done with the free software environment for statistical computing and graphics, R, using the packages ismev and extRemes (see www.R-project.org and R Development Core Team 2009). One advantage of the proposed approach is being parsimonious enough to be readily extended to detect trends in the statistical characteristics of hot spells and related heat waves.

In Section 2, the statistical modeling of extreme temperature events with simple structure is provided as background, emphasizing the point process approach. Summer time series of daily maximum temperature at 3 locations, viz. Phoenix, Arizona (USA), Fort Collins, Colorado (USA), and Paris, France, were analyzed. This approach is extended to hot spells in Section 3, modeling hot spell length with a geometric distribution, and modeling the excess on a given day within a hot spell with a conditional generalized Pareto (GP) distribution whose scale parameter depends on the excess on the previous day. This statistical model for hot spells was further extended to allow for trends in the frequency, duration, and individual excesses of hot spells. A 'heat wave simulator' is introduced in Section 4 to demonstrate how characteristics of more full-fledged heat waves can be obtained from the underlying statistical model for hot spells. Finally, a brief discussion is provided in Section 5, emphasizing further extensions of the statistical modeling of hot spells to make the treatment of heat waves more realistic.

2. STATISTICAL MODEL FOR SIMPLE EXTREME TEMPERATURE EVENTS

The appropriate statistical tools to analyze simple extreme temperature events, such as excesses over high thresholds, are provided by the methods of extreme value theory. Well-known in the atmospheric science and hydrology literature are 2 approaches: (1) the modeling of block maxima (e.g. annual or seasonal maxima or, equivalently, minima) using the generalized extreme value (GEV) distribution; and (2) the peaks-over-threshold (POT) modeling of threshold excesses using the generalized GP distribution. Here, we advocate a third approach, closely related to the first 2, which models the occurrence of exceedances of a high threshold and the corresponding excesses jointly using a 2-dimensional Poisson process.

2.1. Point process approach

The core result of extreme value theory implies that the distribution of the (appropriately normalized) maximum $M_n = \max\{X_1, \ldots, X_n\}$ of an independent and identically distributed sample $X_1, \ldots, X_n$ from a distribution function F converges to the GEV distribution. Consistent with this result, the distribution of the excesses over a high threshold u is approximated by a GP distribution under mild conditions on F. In the context of this paper, the block maximum M_n corresponds to an annual or seasonal maximum temperature, whereas the excesses over u correspond to daily maximum temperatures exceeding the threshold u. The cumulative distribution function of the GEV is given by

$$G(x;\xi,\sigma,\mu) = \exp\left\{-\left[1+\xi\frac{x-\mu}{\sigma}\right]^{-1/\xi}\right\}, 1+\xi\frac{x-\mu}{\sigma} > 0 \quad (1)$$

and that of the GP by

$$H(x;\xi,\sigma_u,u) = 1 - \left[1+\xi\frac{x-u}{\sigma_u}\right]^{-1/\xi}, x > u, 1+\xi\frac{x-u}{\sigma_u} > 0 \quad (2)$$

Here ξ denotes the shape parameter, where positive ξ implies a heavy tail, negative ξ a bounded tail, and the limiting case of $\xi \rightarrow 0$ an exponential tail (i.e. the Gumbel distribution for block maxima and the exponential distribution for threshold excesses); σ, $\sigma_u > 0$ denote the scale parameters and $-\infty < \mu < \infty$ the location parameter. The scale parameters of the GEV and the GP distributions are related through $\sigma_u = \sigma + \xi(u - \mu)$ We anticipate obtaining negative shape parameters, i.e. a bounded tail, for temperature, as indicated e.g. by Brown & Katz (1995).

For most practical situations, for example if X_i represents the daily maximum temperature during the summer at a specific location, the independence assumption is obviously not realistic. One possible way to deal with this problem is to decluster excesses over the threshold u by identifying independent clusters using an empirical rule (e.g. a new cluster starts after r consecutive observations below u). Only 1 value per cluster is kept, e.g. the first excess of the cluster or the maximum excess of the cluster, reducing the sample size for further analysis. See Chapter 5 of Coles (2001) for a discussion of the need to decluster and the modeling of extremes of dependent series in general. A more general view on declustering schemes was provided by Ferro & Segers (2003).

The point process approach, mentioned in the Introduction, combines the modeling of the occurrence of exceedances of a high threshold and their corresponding excesses in 1 model. It uses the fact that the count of threshold exceedances within a certain time window can, under the same conditions as above for the GEV to arise, be approximated by a Poisson distribution with rate λ depending on the parameters μ, σ, ξ of the limiting GEV distribution of the corresponding block maximum. Chapter 7 of Coles (2001) introduces in an accessible way how the Poisson process approximation is obtained and summarizes mathematical and statistical details of this approach, especially the relation to the POT. Maximizing the likelihood of the Poisson process directly yields the GEV parameters μ, σ, ξ, and therefore indirectly the corresponding GP parameters σ_u, ξ. Furthermore, the Poisson rate of the number of clusters per season can be expressed as $\lambda = [1 + \xi(u - \mu)/\sigma]^{-1/\xi}$. The point process approach has several advantages over the block maxima and the POT approaches: (1) it uses considerably more data about extremes than a block maximum approach, resulting in more reliable results; (2) it can be formulated in terms of the GEV parameters, which are invariant to the choice of threshold, allowing non-stationarities such as trends to be easily and naturally introduced through covariate effects in the parameters; and (3) it includes the threshold excess rate in the inference, which is modeled separately in a POT approach. Note that parameter estimation via maximum likelihood requires specialized but straightforward numerical techniques in the non-stationary case.

In order to fit a point process model, it is necessary to select an appropriate threshold. A common approach is to fit the model using a set of candidate thresholds, and to consider only values of u for which the resulting parameter estimates are approximately stable. In the case of a point process model, it is also theoretically possible to vary the threshold in time, but this can lead to numerical instabilities in the maximization. In the case of heat waves, we concentrated on the summer season, so there was no need to consider time-varying thresholds.

2.2. Data

All considered models were tested using series of daily maximum temperature at 3 different stations: Sky Harbor International Airport in Phoenix, Arizona; Fort Collins, Colorado; and Parc Montsouris in Paris. The Phoenix data were obtained from the US National Climatic Data Center and span the period 1934–2007, where the years 1935–1937, 1939, and 1945–1947 are missing and were completely left out of the analysis. Note that Phoenix experienced a heat island effect over this period, with markedly increasing daily minimum temperature but a less pronounced increase in daily maximum temperature (see Balling et al. 1990). The Fort Collins data were obtained from the Colorado Climate Center at Colorado State University, and span the period 1900–1999 with no missing values. The Paris data were obtained from the European Climate Assessment and Dataset (see Klein Tank et al. 2002), and span the period 1900–2008.

For each station, we considered a summer period from 16 June to 15 September ($T = 92$ d) susceptible to the occurrence of hot spells and heat waves. Exploratory data analysis confirmed that daily maximum temperature did not have a marked cycle within this period at these locations, so we did not model seasonality of temperature. Nevertheless, the use of this specific period is a convenient oversimplification since, on one hand, the heat wave season is certainly longer in Phoenix than in Fort Collins or Paris, and, on the other hand, the length of the season itself may be subject to change. For a first application of the proposed method, the simplification seems adequate but may need to be relaxed in a more realistic model. In the

summer period considered, there were fewer than 10 additional missing values for Phoenix and only 2 for Paris. We set the value of the daily maximum temperature on these dates to the minimum observed value over the entire record period, so that they had no influence on the extremal analysis. The data from the USA were provided in heavily discretized form, rounded to the nearest degree Fahrenheit, and subsequently converted to degree Celsius. The data from France were provided rounded to the nearest one-tenth of a degree Celsius. Fig. 1 shows the time series of annual maximum temperature at the 3 sites.

Extreme temperature events are the focus of this paper, so data quality was of special importance. Moreover, since detecting possible trends in hot spells was a goal, we needed to assume homogeneity of the data series to justify fitting the proposed model. Klein Tank et al. (2002) mentioned that it is not untypical for climatic time series to be subject to certain inhomogeneities as, for example, changes in station location or instrumentation. Thus, one should be aware that any of the detected trends could be artifacts of these inhomogeneities, rather than reflecting real climate change.

2.3. Point process model fit

The discretization of the temperature data from the USA resulted in some numerical difficulties in the fitting of the point process model, being more than normally sensitive to the exact choice of the threshold. Cooley et al. (2007) ran simulations to show that using thresholds in the middle of the discretization interval provides numerically stable estimations, which is the approach we took here. Another possibility would be to artificially add noise to the observations to break the ties that cause the numerical difficulties (see Einmahl & Magnus 2008). The conversion from Fahrenheit to Celsius leads to seemingly arbitrary thresholds, which are simply explained as midpoints between distinct data values. Note that the data from Paris were much less discretized and were used here, at least in part, to ensure that the obtained results were not effects of the discretization.

For the traditional point process analysis, we used thresholds of 40.8°C (i.e. 105.5°F) for Phoenix, 30.8°C (i.e. 87.5°F) for Fort Collins, and 27°C for Paris, which were chosen following the approach described in Section 2.1. Clusters were separated by a single value below the threshold, i.e. $r = 1$, retaining the cluster maximum excesses to be treated as independent observations. Note that using the above thresholds and $r = 1$ only served statistical purposes in the modeling of clusters of high temperature; more societally meaningful thresholds and more meteorologically meaningful
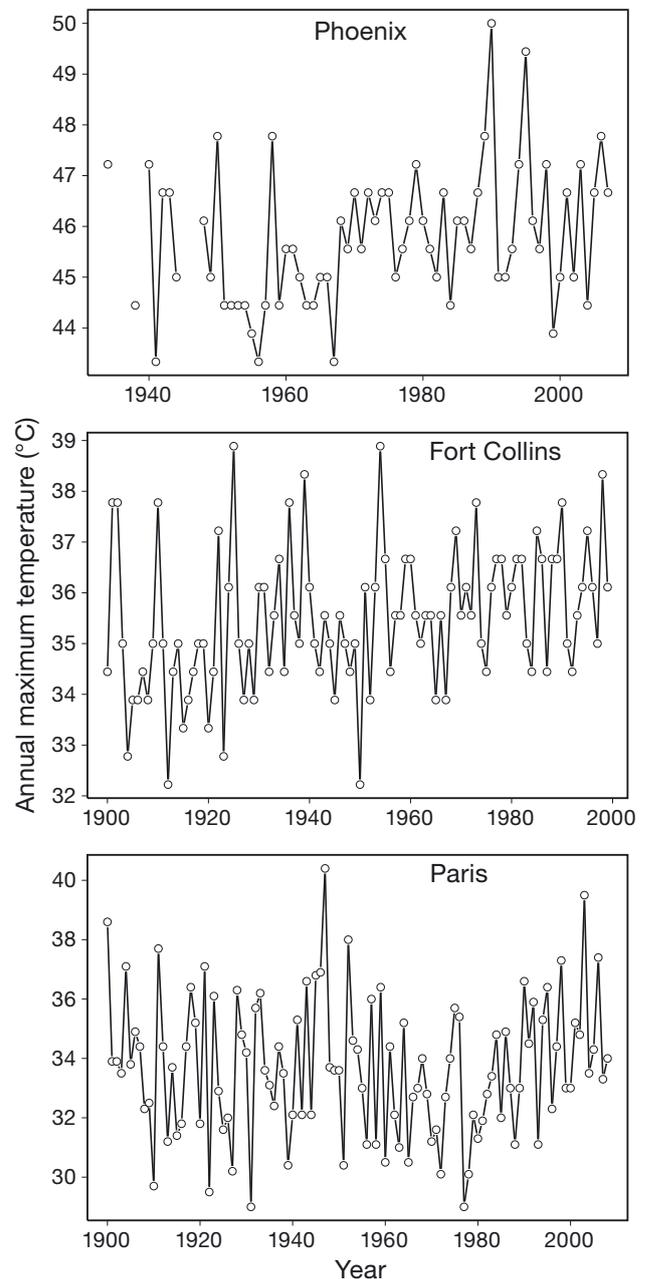

Fig. 1. Time series of annual maximum temperature at 3 sites

values for r were used when considering heat waves derived from hot spells (see Section 4.2).

Maximum likelihood estimates of the GEV parameters at all stations, as well as the above-mentioned thresholds, are given in Table 1. The estimates of the GP parameter σ_u and the Poisson parameter λ were derived from the corresponding values of the GEV parameters as indicated in Section 2.1. Table 1 includes standard errors for all parameter estimates. As anticipated, we obtained negative shape parameter estimates, i.e. a bounded tail, at all 3 locations. Recall that

Table 1. Parameter estimates of the generalized extreme value (GEV) parameters and the corresponding generalized Pareto (GP) and Poisson parameters (SE in parentheses), as well as p-values of the Poisson dispersion test, for 3 sites

Station	Threshold (°C)	Location μ	GEV Scale σ	Shape ξ	GP Scale σ_u	Poisson Mean λ	p
Phoenix	46.8	45.36 (0.141)	1.45 (0.044)	−0.27 (0.021)	2.66 (0.116)	9.69 (0.380)	0.99
Fort Collins	30.8	35.41 (0.010)	1.28 (0.032)	−0.30 (0.018)	2.67 (0.090)	11.24 (0.335)	0.83
Paris	27	33.33 (0.182)	2.31 (0.057)	−0.29 (0.020)	4.15 (0.161)	7.52 (0.263)	0.52

the shape parameter is identical in both parameterizations, the GEV, and the Poisson-GP, of the point process approach. We tested the Poisson hypothesis for the number of clusters per season with a Poisson dispersion test (Rice 1995), based on the approximate χ^2 distribution of the ratio of $n - 1$ times the variance divided by the mean; p-values of the Poisson dispersion test for all stations are also given in Table 1. The hypothesis was supported by the data for all 3 stations,

with Phoenix being the strongest case. The Poisson dispersion test is based on the assumption of stationarity; therefore, a rejection of the null hypothesis could be attributable to a Poisson distribution with a trend, rather than the lack of a Poisson distribution per se. Trends are addressed in Section 3.3.

The top row of Fig. 2 shows histograms of the number of clusters per summer along with the estimated Poisson probability function for all 3 sites. Although deviations from the estimated probability function are apparent, the general shape of the histograms does not strongly contradict the Poisson assumption. In the case of Phoenix, it seems that large numbers of clusters are observed more frequently than the Poisson model would predict. Quantile–quantile (Q–Q) plots for the cluster maximum excess under the GP hypothesis (as derived from the point process model fit) are shown in the bottom row of Fig. 2, and do not indicate any major departures from the assumed point process model.
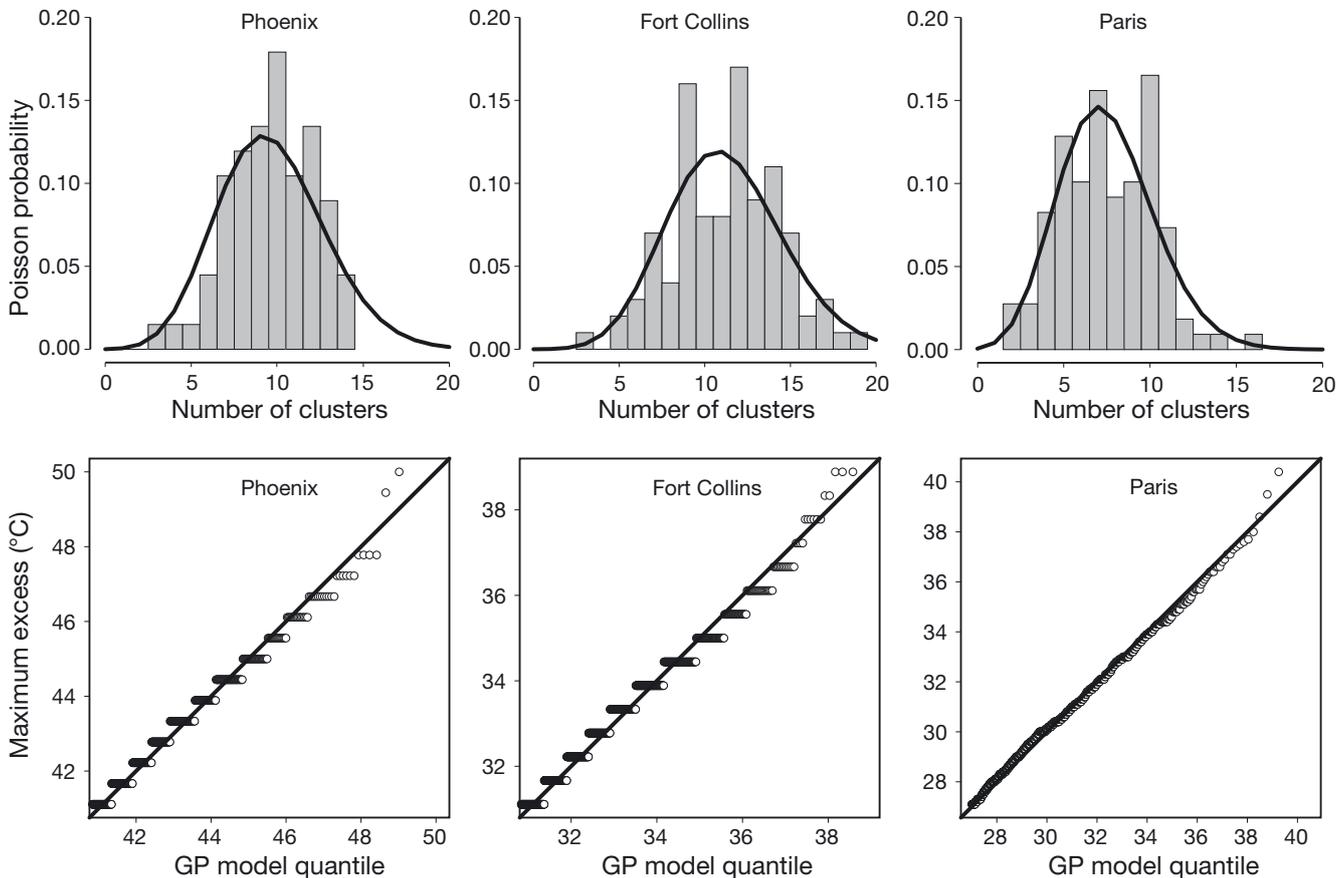

Fig. 2. Top row: histogram and estimated Poisson probability function for the number of clusters per summer at 3 sites. Bottom row: Q–Q plots for the cluster maximum excess under the generalized Pareto (GP) hypothesis at 3 sites

3. STATISTICAL MODEL FOR HOT SPELLS

In the previous section, we fitted a point process model to the cluster maximum temperatures, defining a high temperature cluster as consecutive days with maximum temperature above the threshold *u*, e.g. 40.8°C for Phoenix, where a new cluster of high temperatures starts if the temperature drops below *u* for at least 1 d (*r* = 1). From a more applied viewpoint, we call these clusters of high temperatures 'hot spells'; see Fig. 3 for an illustration of 9 hot spells in a season of 92 d. Again, the choice of threshold and of *r* = 1 is based on statistical considerations; societal and meteorological considerations will be important when applying the fitted model in the analysis of heat waves. Note that the point process fit of the previous section provides a Poisson model for the number of hot spells as well as a GP model for the hot spell maximum excess, but it does not provide a complete description of the entire hot spell. To be complete, we need to additionally provide a model for the spell length, as well as a model for the dependence of excesses within a spell.

3.1. Description of the hot spell model

Here, we propose to model the spell length through a geometric distribution, a simple enough model to allow the easy introduction of trends through a generalized linear model (GLM) framework. In addition, we propose to start from a simple GP model for the first excess of a spell and to assume conditional GP distributions for the remaining excesses, where the conditioning is on the excess of the previous day and the conditional relationship is assumed constant over the
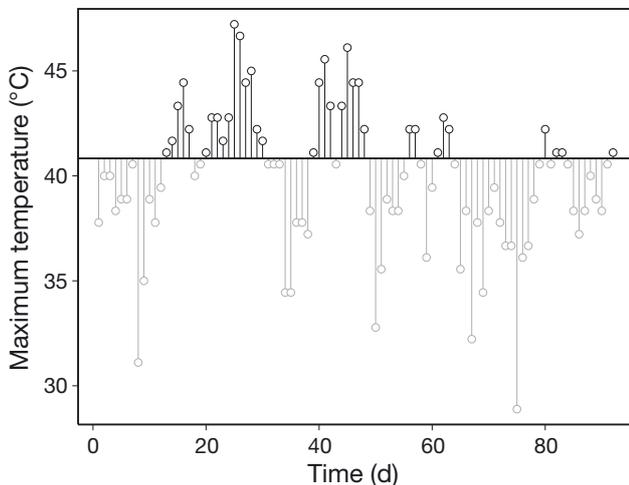

Fig. 3. Observed hot spells (black) during 1934 at Phoenix based on a threshold (horizontal line) of *u* = 40.8°C and *r* = 1 during 16 June (Day 1) to 15 September (Day 92)

length of the spell. Modeling the temporal dependence of the excesses within a hot spell through a Markov process, making use of bivariate extreme value theory, is an asymptotically correct approach (see Coles et al. 1994). Instead, we used only univariate extreme value theory, through conditioning in order to provide a simple approach that is easily applicable in practice while still making to a certain extent use of the theoretical advantages of extreme value models. For both types of GP model, simple and conditional, trends can be easily introduced through covariate effects in the parameters.

Given its memoryless property, the geometric distribution is the simplest plausible model for spell length. Smith et al. (1997) used it to model the cluster length of low minimum daily temperatures, although they found some evidence that a distribution with a heavier tail might be needed. The probability mass function of the geometric distribution is

$$P(k) = (1 - \theta)^{k-1}\theta, \qquad k = 1, 2,\dots \qquad (3)$$

with the reciprocal of the parameter θ being the mean. Parameter estimation is done using the method of moments (which in this case is equivalent to maximum likelihood). Under a wide range of conditions, the parameter θ corresponds to the so-called extremal index, which measures the tendency of the underlying process to cluster at extreme levels; see Chapter 5 of Coles (2001) for a brief discussion of this index.

Note that this model is specific to the threshold *u* in the sense that, if the fitted model is used at a higher threshold, the number of exceedances will no longer necessarily be geometric. Asymptotically correct extreme value models are not subject to this limitation. We circumvented this issue by simulating hot spells, i.e. using the original threshold, and obtaining results on heat waves.

We modeled the excess on the first day of a hot spell with a GP distribution with parameters σ_u and ξ, which we derived from a point process model fit to data retaining only the first excess per hot spell. The remaining excesses of the same spell were modeled conditionally on the excess of the preceding day. More precisely, the conditional excess E_l on Day *l* given the value of the excess E_{l-1} = ν on Day *l* – 1 follows a GP distribution with scale parameter $\sigma_{u,2} = \sigma_{u,2}$ (ν) depending on ν and constant shape parameter ξ_2.

Note that assuming a constant dependence structure throughout each hot spell reduces the number of parameters involved and increases the amount of data available to estimate each of them considerably, namely to all consecutive pairs of excesses. Obviously, it is possible to extend this simple and parsimonious approach by allowing the parameters of the conditional GP distribution to vary depending on which day within the spell is modeled. For the station data consid-

ered here, we encountered numerical problems while fitting such models, more precisely the estimated shape parameters being in some cases smaller than or very close to −0.5, a theoretical bound below which the maximum likelihood estimator is not valid (Coles 2001). The same type of numerical instability was observed for the data at the 3 sites when fitting bivariate extreme value models similar to those considered by Coles et al. (1994) and Smith et al. (1997).

The form of the scale parameter function $\sigma_{u,2}(\nu)$ remains to be chosen. Classical functional forms are the exponential $\sigma_{u,2}(\nu) = \exp(a + b \cdot \nu)$ and linear $\sigma_{u,2}(\nu) = a + b \cdot \nu$. The exponential function is more regularly used in statistics since it ensures positivity, and the classical bivariate extreme value models use a corresponding function. A linear functional form has the appeal of leading to a simpler model and later on to a simpler simulation procedure, and in practice, the positivity constraint is rarely violated within the range of the data.

The choice of the functional form has consequences for the theoretical properties of the conditional GP distribution. Since we assumed that the conditional distribution of the lth excess given the value of the $(l-1)$th excess is a GP distribution, the conditional mean (expectation) of the second excess E_2 given the first $E_1 = \nu$, for example, is given by

$$E[E_2|E_1 = \nu] = \frac{\sigma_{u,2}(\nu)}{1 - \xi_2} \tag{4}$$

i.e. it is again a linear function of the value of the first excess, if the linear form for $\sigma_{u,2}(\nu)$ is chosen. We close the description of the model by giving the formula for conditional quantiles, which will be used later on:

$$H^{-1}[p, \xi_2, \sigma_{u,2}(\nu)] = \frac{\sigma_{u,2}(\nu)}{\xi_2}[(1-p)^{-\xi_2} - 1] \tag{5}$$

where $0 < p < 1$. If $\sigma_{u,2}(\nu)$ is a linear function, then the conditional quantiles are obviously linear functions of ν

with steeper slopes for higher p resulting, for example, in a more rapidly increasing inter-quartile range than the median.

One of the drawbacks of the proposed approach is that the unconditional distribution of any given excess within a spell is not necessarily exactly a GP distribution, although it should be a close approximation. Even though the GP is the asymptotically correct model, it is in practice, i.e. for finite samples, only an approximation, and the conditional approach at worst only weakens this approximation somewhat further. Another possible limitation is that the stochastic process for daily intensities within a cluster is not time-reversible.

3.2. Hot spell model fit

Method of moments estimates of the parameter θ of the geometric distribution for the hot spell lengths as well as standard errors and the corresponding mean spell lengths are given in Table 2, and Fig. 4 shows histograms of hot spell lengths along with the fitted geometric distributions for all 3 sites. Especially for Phoenix, the tail of the geometric distribution seems to underestimate the observed frequency of longer spells, systematically indicating that it might not be heavy enough compared to the data. In spite of this possible drawback, we favor the geometric distribution over more heavy-tailed candidates such as the Zipf distribution (e.g. Section 11.20 of Johnson et al. 1992), since

Table 2. Estimates of the parameter of the geometric distribution θ ($\pm$SE) and mean spell lengths (d) for 3 sites

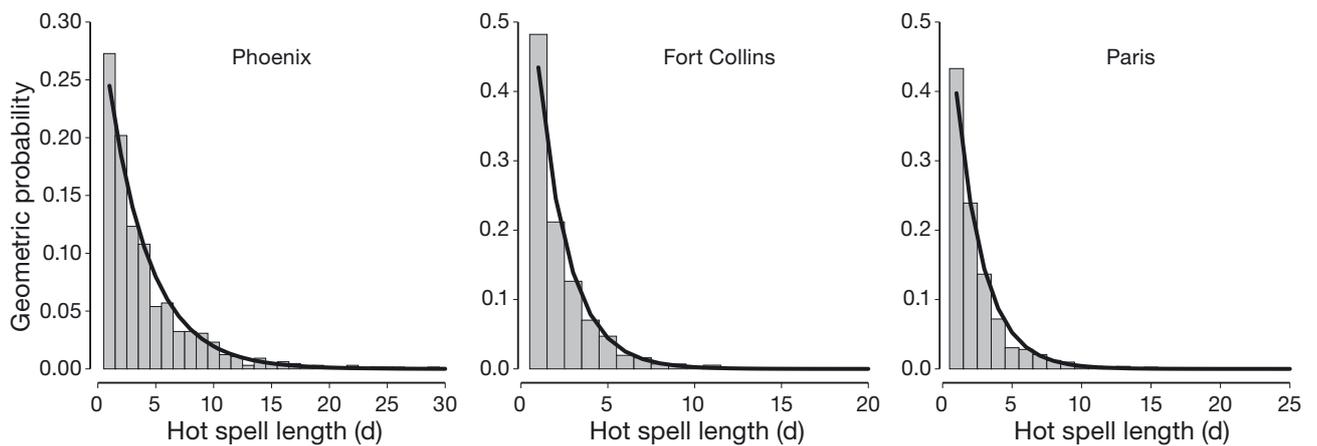| Station	θ	Mean spell length $1/\theta$
Phoenix	0.24 ± 0.008	4.08
Fort Collins	0.43 ± 0.010	2.30
Paris	0.40 ± 0.011	2.52

Fig. 4. Histogram and estimated geometric probability function for the hot spell lengths per summer at 3 sites

the effect due to the observed underestimation should be small and the possibility to easily introduce trends through a GLM approach should be more important.

Table 3 contains parameter estimates along with standard errors of the conditional GP distribution for both choices of the scale parameter function and for all

Table 3. Parameter estimates (±SE) of the conditional generalized Pareto (GP) distribution for both choices of the scale parameter function and for 3 sites

	Exponential function			Linear function		
	a	b	ξ_2	a	b	ξ_2
Phoenix						
	0.50	0.21	−0.36	1.46	0.60	−0.38
	±0.023	±0.013	±0.004	±0.042	±0.024	±0.003
Fort Collins						
	0.91	0.08	−0.38	2.47	0.25	−0.38
	±0.040	±0.010	±0.011	±0.091	±0.031	±0.010
Paris						
	1.31	0.09	−0.46	3.53	0.51	−0.47
	±0.035	±0.007	±0.013	±0.091	±0.032	±0.005

3 sites. Note that the estimates of the shape parameter are barely influenced by the choice of this function and lie in an acceptable range (i.e. negative as expected for temperature data but above −0.5). Fig. 5 shows the conditional relationship between all consecutive pairs of excesses (i.e. E_{l-1} and E_l) with respect to sample/observed (circles and vertical lines) and model/theoretical (thick and dashed lines) median and lower and upper quartiles: linear function (top panels) and an exponential function (bottom panels) for the scale parameter at all 3 sites. For Paris, we rounded the excesses to the nearest half degree Celsius in order to be able to calculate stable conditional sample quantiles. Circles in the plots without attached vertical lines correspond to a single pair of consecutive excesses for the given value of the first excess, i.e. no measure of spread can be calculated. In general, circles that are farther right in the plot are based on fewer values in the sample quantile calculation, i.e. they are less reliable. For Phoenix, the model using the exponential scale parameter function seems to provide a better fit to the last few points on the right. Farther to the left of
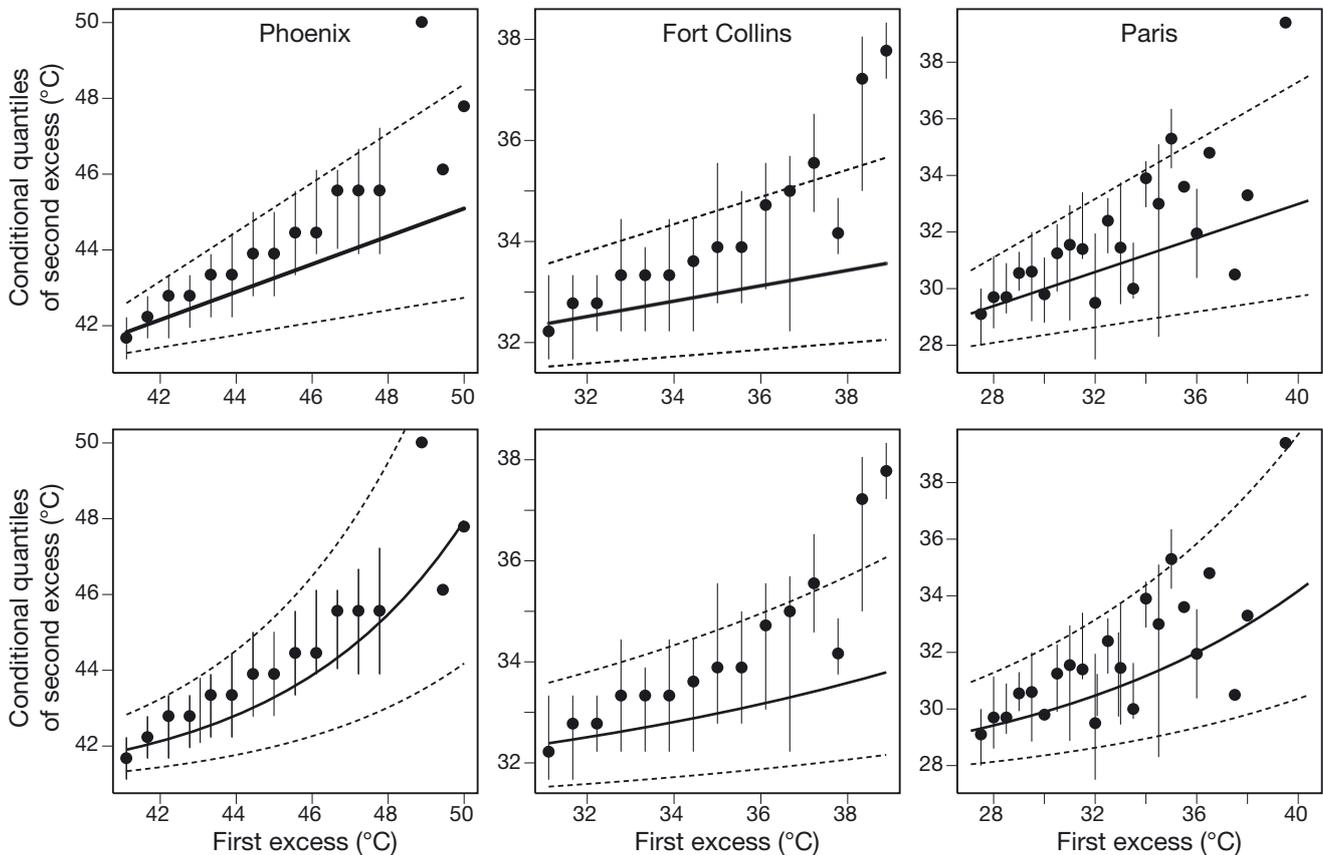

Fig. 5. Theoretical (thick and dashed lines; linear function in top panels, exponential function in bottom panels) and sample (dots and vertical lines) medians and lower and upper quartiles for the conditional generalized Pareto (GP) model for all consecutive pairs of excesses (i.e. the pair E_{l-1} and E_l, with the threshold added to show original temperature values rather than excesses) at 3 sites

the plots, both functional forms result in a similar fit of the model to the data for all 3 stations.

Note that all model characteristics shown in these plots are derived from the fitted model, with the displayed sample characteristics not being directly fitted explaining at least some of the apparent less than ideal performance of the models. In view of the fact that we only allowed 1 parameter to vary (the scale parameter of the conditional GP distribution) and that the right-hand side of the plots is naturally based on extremely few observations, the fit of the conditional GP models seems adequate for all 3 stations. We therefore chose to use the simpler linear function for the scale parameter in the following. A major advantage over more conventional approaches like a conditional normal model is that the conditional GP model is able to capture the effect of increasing variability with increasing median or mean.

3.3. Trends in hot spells

We intentionally constructed our hot spell model such that the introduction of trends in duration, frequency, and intensity of hot spells, and later on indirectly for heat waves, is easily possible. Technically these 3 characteristics correspond to the components of the hot spell model: the geometric model for spell length, the Poisson model for number of spells per season, and the (conditional) GP model for the sizes of the temperature excesses within a spell. While duration and frequency are direct consequences of the definition of a hot spell, intensity can be measured in different ways, e.g. by a mean, maximum, or total excess of a spell. In our case, we concentrated on the first excess as an indicator since it can be assessed easily in our modeling framework and since trends in the first excess will induce changes in these other measures as well.

For all 3 model components, we considered parameters fixed over the heat wave season within a given year but allowed shifts from one year to another, i.e. for each year y of the record period 1, …, P we considered $\theta = \theta(y)$ for the geometric parameter, $\lambda = \lambda(y)$ for the Poisson parameter, and $\sigma_u = \sigma_u(y)$ for the GP scale parameter for the first excess within a spell. The GP shape parameter was kept fixed, since changes in shape are rarely observed and are difficult to model. For the geometric model, trends were introduced through a GLM framework. For the point process/Poisson-GP model, there are 2 possibilities: either (1) introducing trends indirectly through covariate effects in the GEV parameters and then transforming to the Poisson-GP parameterization; or (2) introducing trends directly but separately through a GLM framework in the Poisson model and through covariate effects in the GP scale parameter. We preferred the second possibility because of the advantage that statistical significance can be evaluated separately for the number of spells and excesses. Covariate

Table 4. Parameter estimates (±SE) and p-values (**bold** values indicate significance at the 0.05 level) of the likelihood ratio test for the 3 components of the hot spell model, number of spells, first excesses and spell length, with and without trend, for all 3 sites, where $y = 1, …, P$ denotes the year in the record period

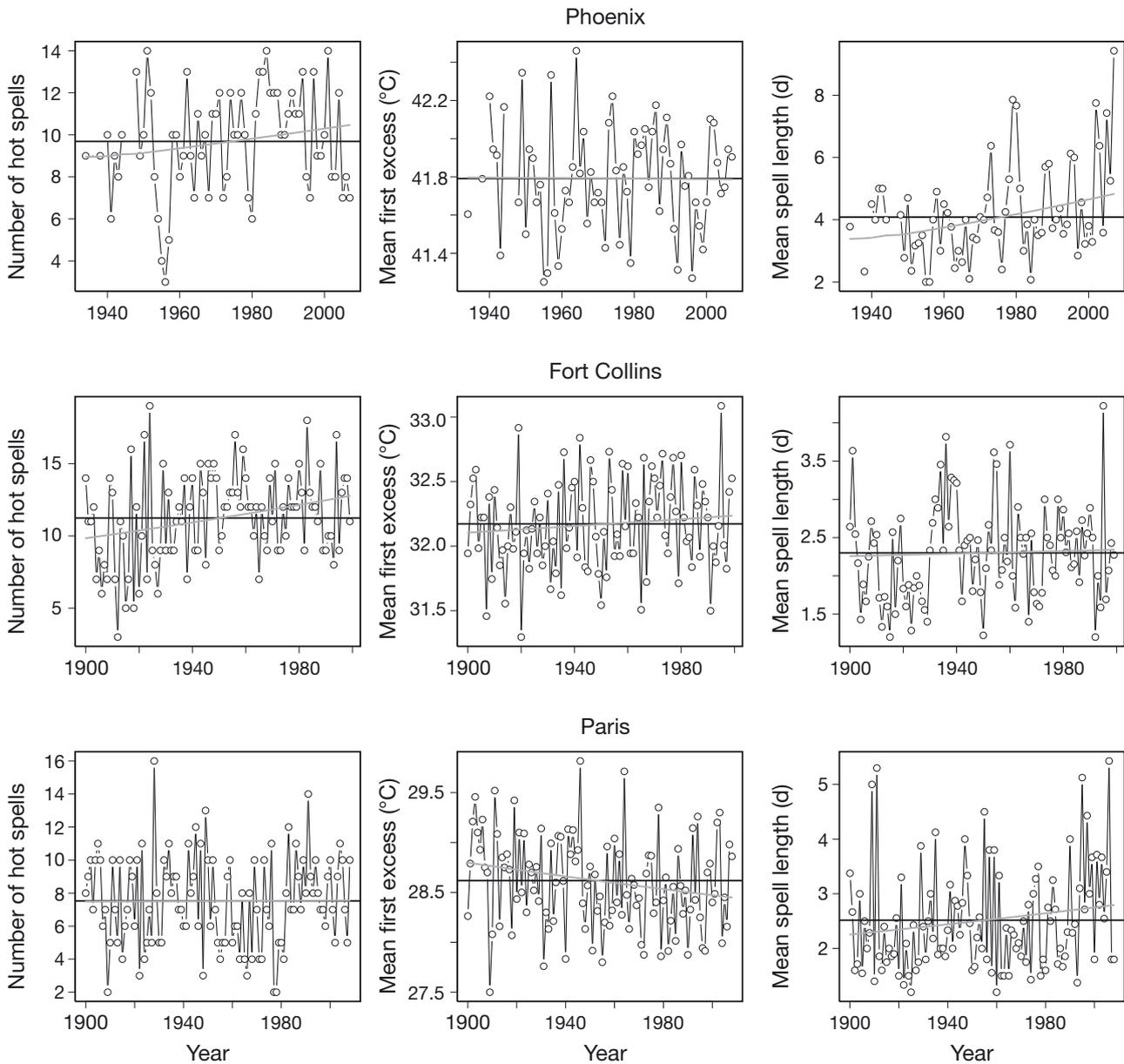|	Number $\lambda(y)$	First excess $\sigma_u(y)$	First excess ξ	Length $1/\theta(y)$
Phoenix				
No trend	9.69	1.13	−0.18	1/0.24
	±0.380	±0.050	±0.022	±0.008
Trend	$\exp(2.2 + 0.002y)$	$\exp(0.1 - 1.6 \times 10^{-4} y)$	−0.18	$\exp(1.2 + 0.005y)$
	±0.081 ±0.002	±0.074 ±0.002	±0.022	±0.080 ±0.002
p	0.24	0.92		**0.01**
Fort Collins				
No trend	11.24	1.61	−0.20	1/0.43
	±0.335	±0.055	±0.018	±0.010
Trend	$\exp(2.3 + 0.003y)$	$\exp(0.4 - 0.001y)$	−0.20	$\exp(0.8 + 3.6 \times 10^{-4} y)$
	±0.062 ±0.001	±0.056 ±0.001	±0.017	±0.054 ±0.001
p	**0.01**	0.24		0.69
Paris				
No trend	7.52	1.88	−0.16	1/0.40
	±0.263	±0.073	±0.018	±0.012
Trend	$\exp(2.0 + 8.0 \times 10^{-5} y)$	$\exp(0.7 - 0.002y)$	−0.16	$\exp(0.8 + 0.002y)$
	±0.070 ±0.001	±0.061 ±0.001	±0.019	±0.065 ±0.001
p	0.94	**0.03**		**0.05**

Fig. 6. Observed number of hot spells, mean first excess, and mean length of hot spells per summer (circles connected by black lines) compared to mean values of the generalized Pareto (GP), Poisson, and geometric distributions, respectively, with (gray lines) and without (black horizontal lines) trend at 3 sites

effects in the parameters are obviously also possible for the conditional GP model, but in view of the difficulties of fitting even our basic model for the dependence of excesses we refrained from that possibility altogether.

Table 4 contains parameter estimates along with standard errors and p-values of the likelihood ratio test, which indicate significant trends if smaller than a certain level, usually taken as 5%, for all 3 components of the hot spell model, with and without trends, and all 3 sites. Fig. 6 contains the observed evolution over time of the same 3 components compared to mean values of the respective model distribution with and without

trends at all 3 sites. As indicated in the table, there was a significant trend in spell length for Phoenix, in the number of spells for Fort Collins, and in the first excess and spell length for Paris, which were more or less all confirmed by the visual impression of Fig. 6. The downward trend in mean of first excesses in Paris is surprising, but might simply be due to the fact that hot spell intensity is not adequately represented by the mean first excess. The visual impression of the observed series confirms a decreasing variance over the years, which matches the sign of the parameter estimates in Table 4. Note that since we considered a

season of fixed length, it will be difficult to allow for a trend in both spell length and number of spells. As mentioned before, this at least somewhat unrealistic assumption will probably need to be relaxed in an expanded version of the model.

4. HEAT WAVE SIMULATOR

Here we introduce a 'heat wave simulator', i.e. a stochastic simulation algorithm that generates temperature series from the discussed hot spell model in order to demonstrate how characteristics of more full-fledged heat waves can be obtained.

4.1. Algorithm

The algorithm to simulate a time series of hot spells starts by generating the number of hot spells for each year in the desired simulation period from the Poisson distribution (which is derived from the point process model of the first excess of a spell). Then, for each of these hot spells, a spell length is generated from the

geometric distribution. From the theory of Poisson processes, it follows that the distribution of the hot spells within the season is uniform, and we use this fact to simulate the alternation between hot spells and intervals between spells over the entire simulation period. The next step is to generate an excess over the threshold for the first day of each hot spell from the GP distribution (which is derived from the point process model), and finally to generate excesses for the remaining days of each hot spell recursively using the conditional GP model. The technical details for the implementation of the simulation algorithm are given in Appendix 1.

As a demonstration of the simulator applied to Phoenix, Fig. 7 shows boxplots of different characteristics of the observed temperature series of 67 yr (74 yr between 1934 and 2007 with 7 yr of missing data), along with minimum/maximum, lower/upper quartile, and median of 100 simulated temperature series of length 74 yr (with no trends in any parameters of the hot spell model). Here, the mean excess per summer is shown as an example of an indicator of the intensity of a hot spell. This characteristic is calculated from all excesses in all spells of a season, i.e. the simulated values are drawn from the GP model of the first excess and the
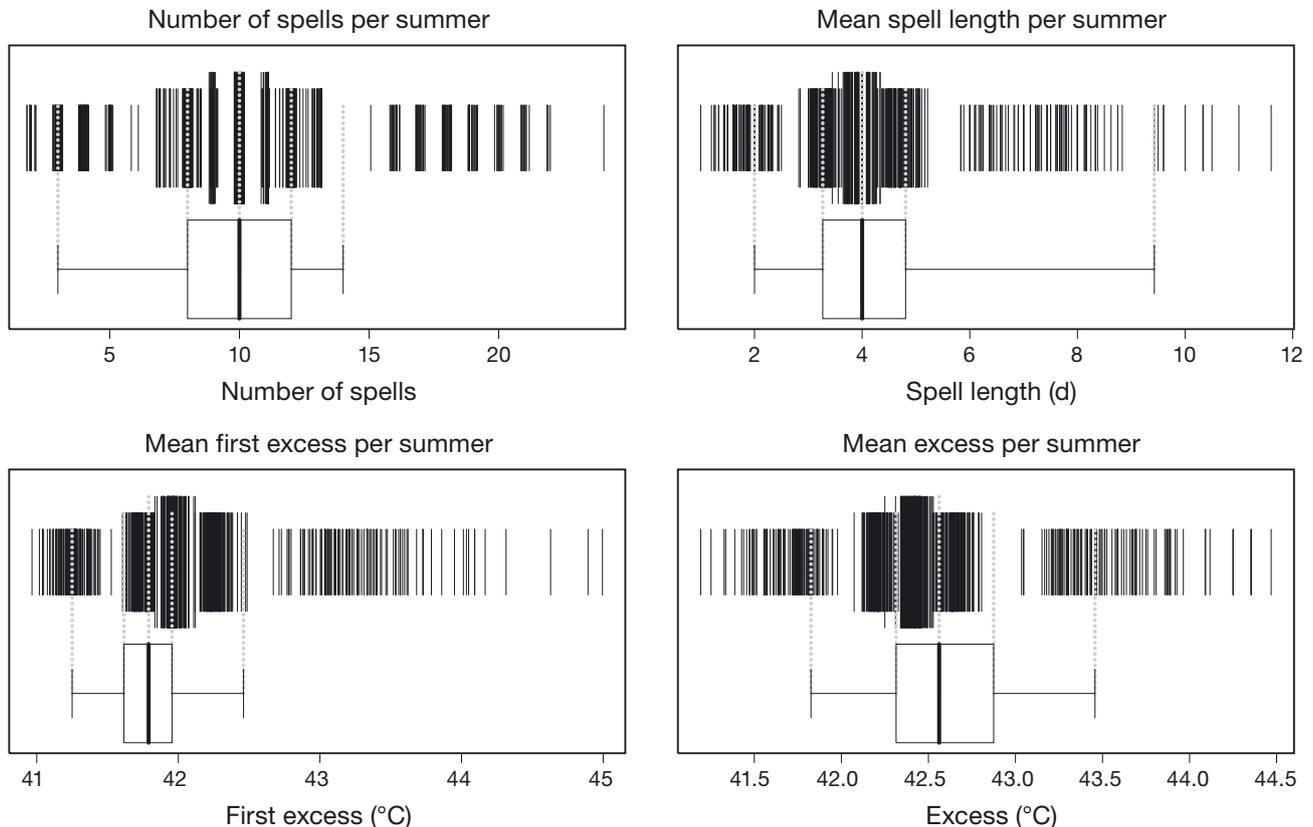

Fig. 7. Boxplots of several hot spell characteristics of observed (67 yr) temperature series over the threshold of $u = 40.8°C$ for Phoenix. Black vertical lines above boxplots correspond with increasing length to minimum/maximum, lower/upper quartile, and median of 100 simulated temperature series of length 74 yr (with no trends in any parameters of the hot spell model). Gray dotted lines extend quantiles of boxplots to facilitate comparison with the simulation
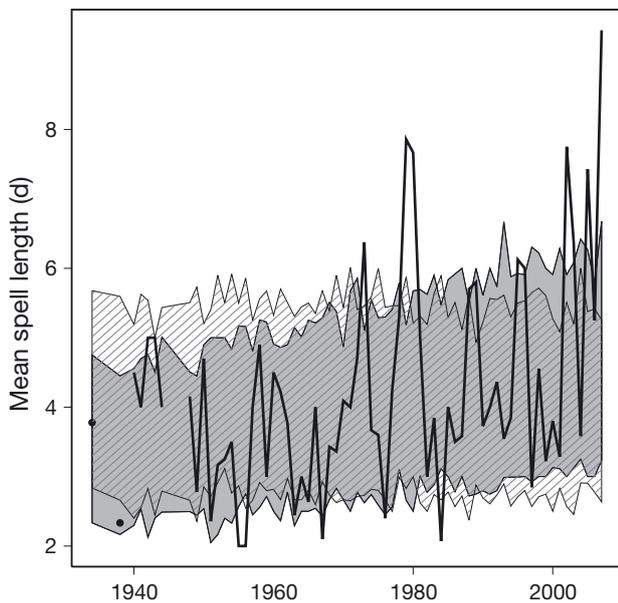
Fig. 8. Mean spell lengths: observed temperature series for Phoenix (thick black line), pointwise 10 and 90% quantiles of 100 simulated temperature series of length 74 yr (hatched: without trend; gray: with trend)

conditional GP model for the remaining excesses of a spell. The deviation in the central part of the boxplot is a concern here, reflecting again that the conditional GP model is not a perfect approximation of the true underlying process (see also Fig. 5).

As a second demonstration of the simulator, this time including a trend in the parameter of the geometric distribution for spell length, Fig. 8 shows mean spell lengths per season from the observed temperature series from Phoenix along with pointwise 10 and 90% quantiles of 100 simulated temperature series of length 74 yr. This display emphasizes the positive effect of the introduction of a trend on the simulation, but again indicating that the geometric model is probably not heavy tailed enough (see also Fig. 2).

4.2. Heat waves

The extreme value methodology we apply requires, on the one hand, that the threshold be high enough for the asymptotic theory to be valid but, on the other hand, low enough such that enough data are available for the analysis to be stable. Another requirement is that clusters/hot spells need to be approximately independent, which is usually achieved by a declustering scheme of which we used the simplest one: 2 spells are separated by at least 1 d of lower temperatures. To study heat waves, we used the fact that at least for some definitions, they can be indirectly derived from hot spells, for example by using a higher threshold (Fig. 9, middle), using only longer hot spells, merging spells, i.e. using $r > 1$ (Fig. 9, right) or other functionals of the spell (e.g. mean or total excess). The principal idea is to model hot spells using extreme value theory, and then derive conclusions on heat waves, which themselves cannot be analyzed as easily for various reasons: (1) there are too few data for direct models of heat waves to be as reliable; (2) different definitions of heat waves would require repeated model fitting, if not different modeling approaches; and (3) heat waves are, depending on the definition, considerably more complex to model directly.

As an example, we used the fitted hot spell model (i.e. $u = 40.8°C$ and $r = 1$) for Phoenix and accordingly simulated hot spell series with and without the detected trend in spell length. From these simulated hot spells, we indirectly obtained simulations of heat waves, defined as temperatures exceeding the higher threshold of 43.6°C (i.e. 110.5°F), where 2 heat waves are separated by at least 1 d of lower temperatures (i.e. still $r = 1$). Fig. 10 shows an observed series of the number of heat waves (top), mean length (middle), and mean excess (bottom, i.e. 1 measure of heat wave intensity) along with corresponding pointwise 5% and 95% quantiles of 100 simulated temperature series of the same length (except for years of missing data) with and without trend.
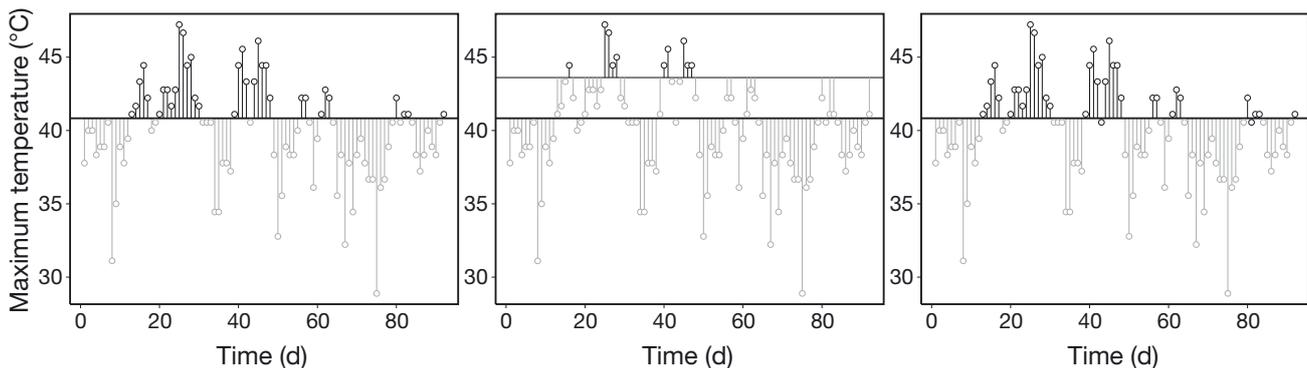

Fig. 9. Observed hot spells (black) during 1934 at Phoenix based on a threshold of $u = 40.8°C$ and $r = 1$ (left), $u = 43.6°C$ and $r = 1$ (middle), and $u = 40.8°C$ and $r = 2$ (right)
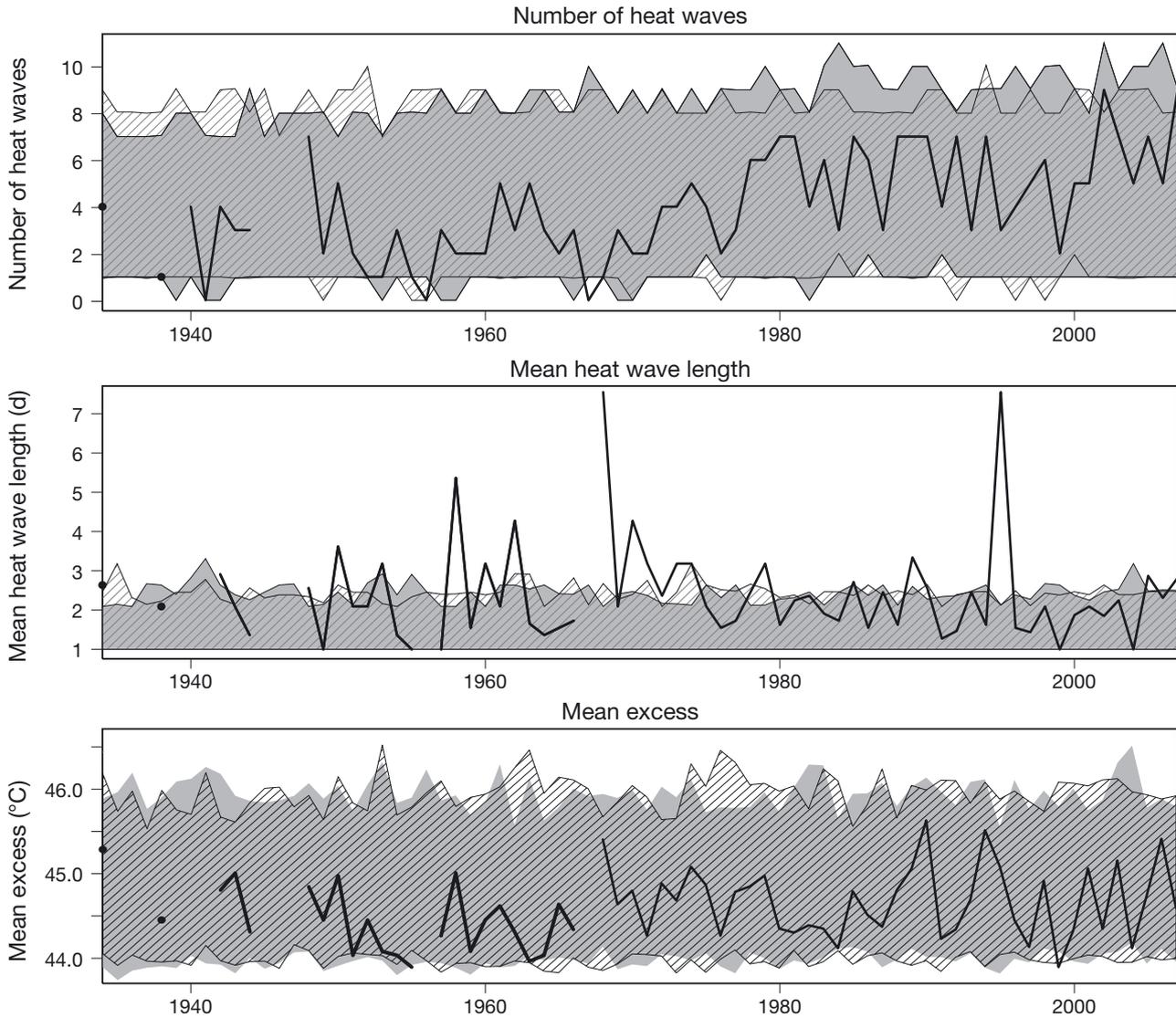
Fig. 10. Number of heat waves (u = 43.6°C, r = 1), mean heat wave length (d, middle), and mean excess (bottom); thick lines correspond to the observed series at Phoenix of length 67 yr, areas between pointwise 5 and 95% quantiles of 100 simulated temperature series of length 74 yr are hatched for the hot spell model without trend and gray for the model including a trend. The extra gaps (in addition to the years of missing data) in the time series (middle and bottom panel) are due to years in which no heat waves occurred (see top panel)

In Fig. 10, the observed number of heat waves in Phoenix seems to increase more systematically than the observed number of hot spells (recall Fig. 6), whereas the length and mean excess during heat waves do not seem to show systematic changes over time. Possibly the observed (and statistically significant, see Table 4) increasing trend in hot spell length results in heat waves occurring more frequently. Most of the line corresponding to the number of heat waves is contained in the shaded area, as it is rather wide. The dark gray area, corresponding to the hot spell model with a trend in the geometric spell length distribution, seems to reflect the potential trend in the number of heat waves, more convinc-

ingly so if bands with lower confidence (e.g. 10 and 90%) are used (not shown). For mean heat wave length, it is clear that the simulations are not able to reproduce the observed sudden spikes, and for the mean excess during heat waves, the confidence bands seem again rather wide compared to the observed values.

5. DISCUSSION

A new technique has been proposed for the statistical modeling of hot spells. Unlike most previous research on this topic, our method is based as much as is

feasible on the statistical theory of extreme values. Given that hot spells are an extreme meteorological phenomenon, this reliance on extreme value theory naturally produces an approach that treats the basic characteristics (i.e. frequency, duration, and intensity) of such events in a more realistic manner statistically than techniques heretofore applied. Perhaps less obvious, the point process technique for extreme value analysis results in a more powerful approach for systematically studying the statistical features of extreme high temperatures. We have demonstrated how the statistical characteristics of more full-fledged heat waves can be derived from our statistical model for hot spells. In particular, attention need no longer be restricted to a rigid definition of a heat wave, about which there is not necessarily any consensus.

The proposed technique has been intentionally kept simple enough for trends in its various components to be incorporated. Thus, there remain a number of respects in which the technique could be extended, both to make its treatment of hot spells more realistic statistically and of heat waves meteorologically. As already mentioned, it would seem more reasonable to allow a trend in the length of the heat wave season, along with any trends in other characteristics. A more appealing, but less parsimonious, approach would consist of introducing seasonality into the parameters of the statistical model, rather than holding them fixed over an entire season. A longer summer season could be modeled, with hot spells being less likely at the beginning and end of the season. Concerning meteorological realism, it would be straightforward to apply the technique to daily time series of apparent temperature instead of maximum temperature, thus taking into account humidity (e.g. Karl & Knight 1997). Indices of atmospheric circulation patterns, such as blocking, could be used as covariates instead of, or in addition to, a trend component (e.g. Sillmann & Croci-Maspoli 2009). Much more challenging would be the simultaneous treatment of both daily maximum and minimum temperature, thus taking into account night-time weather conditions as well.

This research has implications for the statistical modeling of temperature variables more generally. Specifically, climate change scenarios are frequently produced by stochastic weather generators (e.g. Semenov 2008). Conventional weather generators are based on autoregressive-type models for time series of daily minimum and maximum temperature. As such, they cannot be expected to adequately represent the statistical characteristics of extreme high temperatures, especially the temporal dependence of excesses within a hot spell. How to modify such weather generators to improve their performance in terms of simulating heat waves remains an open question.

LITERATURE CITED

Abaurrea J, Asín J, Cebrin AC, Centelles A (2007) Modeling and forecasting extreme hot events in the central Ebro valley, a continental-Mediterranean area. Global Planet Change 57:43–58

Balling RC Jr, Skindlov JA, Phillips DH (1990) The impact of increasing summer mean temperatures on extreme maximum and minimum temperatures in Phoenix, Arizona. J Clim 3:1491–1494

Brown BG, Katz RW (1995) Regional analysis of temperature extremes: spatial analog for climate change? J Clim 8:108–119

Brown SJ, Caesar J, Ferro CAT (2008) Global changes in extreme daily temperature since 1950. Global Planet Change 113:D05115 doi:10.1029/2006JD008091

Coles S (2001) An introduction to statistical modeling of extreme values. Springer, London

Coles SG, Tawn JA, Smith RL (1994) A seasonal Markov model for extremely low temperatures. Environmetrics 5:221–239

Cooley D, Nychka D, Naveau P (2007) Bayesian spatial modeling of extreme precipitation return levels. J Am Stat Assoc 102:824–840

Einmahl JHJ, Magnus JR (2008) Records in athletics through extreme-value theory. J Am Stat Assoc 103:1382–1391

Ferro CAT, Segers J (2003) Inference for clusters of extreme values. J R Stat Soc B Stat Methodol 65:545–556

Furrer EM, Katz RW (2008) Improving the simulation of extreme precipitation events by stochastic weather generators. Water Resour Res 44:W12439 doi:10.1029/2008WR007316

Gosling SN, Lowe JA, McGregor GR, Pelling M, Malamud BD (2009) Associations between elevated atmospheric temperature and human mortality: a critical review of the literature. Clim Change 92:299–341

Gumbel EJ (1958) Statistics of extremes. Columbia University Press, New York, NY

Johnson NL, Kotz S, Kemp AW (1992) Univariate discrete distributions, 2nd edn. John Wiley & Sons, New York, NY

Karl TR, Knight RW (1997) The 1995 Chicago heat wave: How likely is a recurrence? Bull Am Meteorol Soc 78:1107–1119

Katsoulis BD, Hatzianastassiou N (2005) Analysis of hot spell characteristics in the Greek region. Clim Res 28:229–241

Katz RW, Parlange MB, Naveau P (2002) Statistics of extremes in hydrology. Adv Water Resour 25:1287–1304

Khaliq MN, St-Hilaire A, Ouarda TBMJ, Bobée B (2005) Frequency analysis and temporal pattern of occurrences of southern Quebec heatwaves. Int J Climatol 25:485–504

Khaliq MN, Ouarda TBMJ, St-Hilaire A, Gachon P (2007) Bayesian change-point analysis of heat spell occurrences in Montreal, Canada. Int J Climatol 27:805–818

Klein Tank AMG, Wijngaard JB, Können GP, Böhm R and

others (2002) Daily dataset of 20th-century surface air temperature and precipitation series for the European climate assessment. Int J Climatol 22:1441–1453

Koffi B, Koffi E (2008) Heat waves across Europe by the end of the 21st century: multiregional climate simulations. Clim Res 36:153–168

Kysely J (2002) Probability estimates of extreme temperature events: stochastic modeling approach vs. extreme value distributions. Stud Geophys Geodaet 46:93–112

Mearns LO, Katz RW, Schneider SH (1984) Extreme high-temperature events: changes in their probabilities with changes in mean temperature. J Clim Appl Meteorol 23:1601–1613

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

Meze-Hausken E (2008) On the (im-)possibilities of defining human climate thresholds. Clim Change 89:299–324

R Development Core Team (2009) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Rice JA (1995) Mathematical statistics and data analysis.

Duxbury Press, Belmont, CA

Robinson P (2001) On the definition of a heat wave. J Appl Meteorol 40:762–775

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

Semenov MA (2008) Simulation of extreme weather events by a stochastic weather generator. Clim Res 35:203–212

Sillmann J, Croci-Maspoli M (2009) Euro-Atlantic blocking and extreme events in present and future climate simulations. Geophys Res Lett 36:L10702 doi:10.1029/2009GL038259

Smith RL (1989) Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone (with discussion). Stat Sci 4:367–393

Smith RL, Tawn JA, Coles SG (1997) Markov chain models for threshold exceedances. Biometrika 84:249–268

Tebaldi C, Hayhoe K, Arblaster JM, Meehl GA (2006) Going to the extremes: an intercomparison of model-simulated historical and future changes in extreme events. Clim Change 79:185–211

Appendix 1. Implementation of the heat wave simulator

Recall that the discussed hot spell model is based on several parameters: λ for the Poisson parameter of the number of hot spells, θ for the geometric parameter of hot spell length, σ_u and ξ for the generalized Pareto (GP) parameters of the first excess of each spell, and a, b, ξ_2 for the parameters of the conditional GP distribution with scale parameter modeled with a linear function for the excesses within a spell. Given a season length T, a number of seasons S to be simulated, and values (estimated or hypothetical) for the model parameters, the algorithm to generate series of hot spells is given by the following pseudo-code:

for y in $1,\ldots S$ repeat [1] to [8]
[1] generate $N(y) \sim \text{POIS}(\lambda)$
[2] if $N(y) > [T/2]$ goto 1
[3] for $i \in 1,\ldots,N(y)$ generate $L_i(y) + 1 \sim \text{GEO}(\theta)$
[4] if $\Sigma_i L_i(y) + 2N(y) - 1 > T$ goto 3
[5] draw $u_1,\ldots,u_{N(y)}$ without replacement from $\{0,\ldots,T_{\text{cold}} = T - \Sigma_i L_i(y) - N(y)\}$,
[6] divide $\{1,2,\ldots,T_{\text{cold}}\}$ using $u_{[1]},\ldots,u_{[N(y)]}$
[7] for $i \in 1,\ldots,N(y)$ generate $E_{i,1}(y) \sim \text{GP}(\sigma_u, \xi)$
[8] for $i \in 1,\ldots,N(y)$, $j \in 1,\ldots,L_i(y)$ generate $E_{i,j+1}(y) \sim \text{GP}(a + bE_{i,j}(y), \xi_2)$

Here $N(y)$ denotes the number of hot spells, $L_i(y) + 1$ is the length of the ith hot spell (with spells being at least 1 d long), there are $T_{\text{cold}} = T - \Sigma_i L_i(y) - N(y)$ days not in a hot spell, $u_{[1]},\ldots,u_{[N(y)]}$ is the ordered sample, and $E_{i,j}(y)$ is the excess on day j of the ith hot spell. In step [1], the notation $N(y) \sim \text{POIS}(\lambda)$ refers to a random number $N(y)$ distributed according to a Poisson distribution with parameter λ, similarly for the geometric distribution in step [3] and the GP distribution in step [7] and [8]. Note that steps [2] and [4] ensure that there are enough days in the season to fit the generated number of hot spells with corresponding length and at least 1 d of colder temperature between spells (i.e. at least $N(y) - 1$ 'cold' days).

The simulation algorithm is virtually the same if the model contains a trend in 1 or several of the parameters, the only difference being that a different parameter value is used in the generation for each season. For example, if the observed trend in the spell length for Phoenix is included in the model, then the third step uses $L_i(y) + 1 \sim \text{GEO}(\theta(y))$.