Nonstationary seasonal model for daily mean temperature distribution bridging bulk and tails

Mitchell Krock, Julie Bessac, Michael L. Stein & Adam H. Monahan

2022

Faculty of Science

Faculty Publications

© 2022 Krock et al. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License
CC BY-NC-ND: https://creativecommons.org/licenses/by-nc-nd/4.0/

Original citation:
Krock, M., Bessac, J., Stein, M. L., & Monahan, A. H. (2022). Nonstationary seasonal model for daily mean temperature distribution bridging bulk and tails. *Weather and Climate Extremes*, *36*, 100438. https://doi.org/10.1016/j.wace.2022.100438

Downloaded from UVicSpace Research & Learning Repository

dspace.library.uvic.ca

University of Victoria | Libraries

Weather and Climate Extremes 36 (2022) 100438

Contents lists available at ScienceDirect

Weather and Climate Extremes

journal homepage: www.elsevier.com/locate/wace

Nonstationary seasonal model for daily mean temperature distribution bridging bulk and tails

Mitchell Krock [a,*], Julie Bessac [b], Michael L. Stein [a], Adam H. Monahan [c]

[a] Department of Statistics, Rutgers University, Piscataway, NJ, USA
[b] Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL, USA
[c] School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia, Canada

ARTICLE INFO

Keywords:
Temperature extremes
Bulk and tails
Nonstationary
Climate change

ABSTRACT

In traditional extreme value analysis, the bulk of the data is ignored, and only the tails of the distribution are used for inference. Extreme observations are specified as values that exceed a threshold or as maximum values over distinct blocks of time, and subsequent estimation procedures are motivated by asymptotic theory for extremes of random processes. For environmental data, nonstationary behavior in the bulk of the distribution, such as seasonality or climate change, will also be observed in the tails. To accurately model such nonstationarity, it seems natural to use the entire dataset rather than just the most extreme values. It is also common to observe different types of nonstationarity in each tail of a distribution. Most work on extremes only focuses on one tail of a distribution, but for temperature, both tails are of interest. This paper builds on a recently proposed parametric model for the entire probability distribution that has flexible behavior in both tails. We apply an extension of this model to historical records of daily mean temperature at several locations across the United States with different climates and local conditions. We highlight the ability of the method to quantify changes in the bulk and tails across the year over the past decades and under different geographic and climatic conditions. The proposed model shows good performance when compared to several benchmark models that are typically used in extreme value analysis of temperature.

1. Introduction

The probability distribution of surface air temperature (SAT) possesses nonstationary traits, such as seasonality and long-term trends, that can be difficult to capture with off-the-shelf models. This fact is particularly true for the tails of the distribution. In this paper, we study SAT using a model for the entire probability distribution that has versatile behavior in both tails. Some areas where such a model is useful include finance, insurance, and environmental science (Finkenstädt and Rootzén, 2003; Reiss and Thomas, 2007). The tails of the distribution may be more important than the bulk, as unlikely events in the stock market or environment can have more severe consequences than an ordinary event. With temperature data, both tails of the model distribution are important and of particular public interest due to growing impacts from climate change on human health, the environment, and the economy. Early in 2021, Texas experienced record low temperatures in February from winter storms, which led to massive power grid failures (Busby et al., 2021), while in the Pacific Northwest of North America, record high temperatures in June led to increased hospitalizations and fatalities (Schramm et al., 2021; Philip et al., 2021). On longer time scales (months to years), both tails of precipitation distributions matter for the same reasons as temperature. Even when the focus is on the extremes, an understanding of the full distribution is often still of interest. Temporal changes may be different in the upper and lower tails, and nonstationary patterns such as seasonality and long-term trends will also be observed in the bulk of the distribution. To our knowledge, this study is the first attempt to model nonstationary temperature data on the daily scale with a concentration on behavior in both tails.

Temporal changes in temperature distribution are influenced by various phenomena. Seasonal variability is largely driven by the seasonal cycle in solar radiation, which depends strongly on latitude. However, large-scale oceanic and atmospheric circulation patterns (McKinnon et al., 2013) and local geography (e.g., elevation, distance to the ocean) also have a large impact on the seasonal patterns at specific locations. For monthly means, temperature seasonality has been traditionally studied through circular harmonics, which have been shown to

* Corresponding author.
E-mail address: mk1867@stat.rutgers.edu (M. Krock).

https://doi.org/10.1016/j.wace.2022.100438
Received 19 October 2021; Received in revised form 20 March 2022; Accepted 21 March 2022
Available online 31 March 2022
2212-0947/© 2022 Argonne National Laboratory and The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

capture a large amount of the variability (up to 99%) (Legates and Willmott, 1990). However, at shorter time resolution, temperature exhibits more complex seasonal patterns.

Rising global mean temperature has drawn attention for many decades, and regional temperatures exhibit various rates and patterns of change, including in their extremes. A large literature has been generated on the topic; an extensive consolidation can be found in assessment reports of the Intergovernmental Panel on Climate Change (IPCC, 2021). In the following, we discuss literature focusing on statistical aspects of SAT extremes.

Most previous works have focused on statistics of a single tail of a temperature distribution but rarely on the entire distribution of temperature changing over time (Seneviratne et al., 2021). For instance, among the large quantity of work on global mean temperature, Rahmstorf and Coumou (2011) and Poppick et al. (2017) examined trends, and McKinnon et al. (2016) studied long-term changes of temperature quantiles. Meehl et al. (2009) showed that ratios of frequencies of daily hot maxima over frequencies of daily cold minima under nonstationary climate conditions in the United States exhibit strong asymmetry leading to more frequent hot extremes. Other studies have considered the evaluation and quantification of changing temperature in historical datasets (measurements and/or model output) (Legates and Willmott, 1990; Tarleton and Katz, 1995; Hansen et al., 2010; Rhines et al., 2017), while recent studies focus on projections under various greenhouse gas concentration pathways by leveraging the use of multi-model ensembles or large single-model ensembles (Huang et al., 2016; Haugen et al., 2018; Wehner, 2020). Several works have investigated the statistical aspects of extreme quantification under nonstationarity (Cheng et al., 2014; Gilleland et al., 2017; IPCC, 2021). Katz and Brown (1992) proposed a statistical hypothesis test to analyse the sensitivity of extreme events to changes in the location and shape of climate distributions, emphasizing the impact of the scale parameter on extreme occurrences. Robin and Ribes (2020) proposed a nonstationary framework for extremes including model data and observations. Grotjahn et al. (2016) reviewed statistical methodologies, dynamics, modeling efforts, and trends dedicated to temperature extremes.

Very few works have focused on jointly learning from the bulk and tails of SAT. However, pivotal insights about changing climate have been obtained by quantifying the link between changing extremes and other statistics of the bulk for the quantity of interest. For instance, Huybers et al. (2014) and Huang et al. (2016) respectively linked changing extremes to changing mean temperature using reanalysis data and climate model projections. Huybers et al. (2014) proposed a metric to link the mean to an extreme quantile. These works motivate the use of statistical models that can simultaneously assess changes in the tails and bulk of a distribution.

The classical approach to statistical extremes is to model a small fraction of the most extreme outcomes (large and/or small depending on context) of a sequence of independent, identically distributed (or stationary) random variables. Two common models, the generalized Pareto distribution (GPD) for threshold exceedances and the generalized extreme value distribution (GEV) for block-maxima, are motivated by asymptotic theory and can be readily fit by practitioners with pre-existing software. The restrictive assumption of stationarity can be addressed by treating parameters of the GPD or GEV distribution as functions of time-dependent covariates (Davison and Smith, 1990). See Coles (2001) for a more extensive introduction to extreme value theory. An essential limitation of GPD and GEV models is that they only fit a single tail of the distribution with only a small portion of the dataset. This practice of ignoring the majority of the available data is particularly concerning when trying to accurately infer any nonstationarities, as these patterns are strongly tied to the behavior in non-extreme observations. Nogaj et al. (2007), Eastoe and Tawn (2009), and Mentaschi et al. (2016) acknowledged this issue and proposed preprocessing of the data with location and scale functions to capture the nonstationarity. The processed data are assumed to be stationary, and standard GEV/GPD methods can then be used for inference. Analyses using this "transformed-stationary" approach demonstrate the value of modeling nonstationary extremes without ignoring the bulk of the data, but this methodology is limited to a single tail at a time, and is accompanied by standard issues with GEV/GPD fitting, such as selection of a block-size or threshold (Coles, 2001; Scarrott and MacDonald, 2012).

A recent model proposed by Stein (2021a) fits the entire distribution and has flexible behavior in both tails. Other works have also attempted to bridge the tails and bulk of a distribution; see Scarrott and MacDonald (2012) for a review. These proposals largely rely on mixture models (Frigessi et al., 2002; Carreau and Bengio, 2009; Bopp and Shaby, 2017; Yadav et al., 2021) or combining function composition with cumulative distribution functions (Naveau et al., 2016; Tencaliec et al., 2020; Stein, 2021b). Huang et al. (2019b) suggest a semiparametric approach incorporating log-histosplines. Vandeskog et al. (2022) recently used the five-parameter lambda distribution (Gilchrist, 2000) to model diurnal temperature range. A number of the limitations of these approaches, such as flexible behavior in only one tail, restrictions to positive or heavy-tailed variables, or the need to numerically compute a normalizing constant, are obviated by the approach of Stein (2021a), which provides a comprehensive approach to handle the bulk and both tails of a distribution.

In this paper, we extend the recent work of Stein (2021a) and demonstrate a flexible model for a random variable whose distribution changes over time. In particular, we extend the distribution from Stein (2021a) to a seasonal model accounting for seasonality, long-term trends, and the interaction between these two characteristics. We illustrate the flexibility and ability of the model to capture changing distributions of measurements of daily surface air temperature. In particular, the bulk, lower tail, and upper tail of temperature vary differently across seasons and over the long-scale study period. A number of stations representing the diversity of seasonality and long-term changes in temperature are selected from across the United States. We highlight the capability of the proposed model to quantify these changing temperature patterns and extremes.

We outline the paper as follows: in Section 2, we describe the daily temperature data used in this study. In Section 3, we describe the model from Stein (2021a) and its extension. Finally, in Section 4, we present results from applying our methodology to daily temperature data and compare it to alternatives in both the bulk and tails.

2. Surface air temperature data

Daily SAT used in this paper are provided by the National Climatic Data Center's Global Surface Summary of the Day (GSOD) (NCEI, 2020). The GSOD database contains meteorological measurements from weather stations across the world. The R package GSODR (Sparks et al., 2017) offers a helpful interface to work with this dataset. We selected several U.S. cities from varying locations with different geographies, climates, and local conditions: Bethel, Alaska; Colorado Springs, Colorado; Minneapolis, Minnesota; Boston, Massachusetts; Hilo, Hawaii; San Diego, California; Blythe, California; and Homestead, Florida. Under the Köppen climate classification (Peel et al., 2007), Bethel is a subarctic climate (Dfc), Colorado Springs is a dry semi arid climate (Bsk) at high elevation, Minneapolis is a humid continental region (Dfa) with large seasonal variation, Boston is a mixture between humid continental climate (Dfa) and a humid subtropical climate (Cfa), Hilo is a tropical rainforest which receives significant rainfall (Af), San Diego is a mixture between a Mediterranean climate (Csa) and a semi-arid climate (Bsh), Blythe is a hot desert climate (Bwh), and Homestead is a tropical savannah climate (Aw) that nears a tropical monsoon climate (Am). These eight locations are shown in Fig. 1. Time periods of the SAT observations range from the 1940's to near present (2020). The records present missing data that are not imputed in this study. Data are recorded on an hourly basis and subsequently averaged to create daily summary values. Table 3 in Appendix A provides additional information

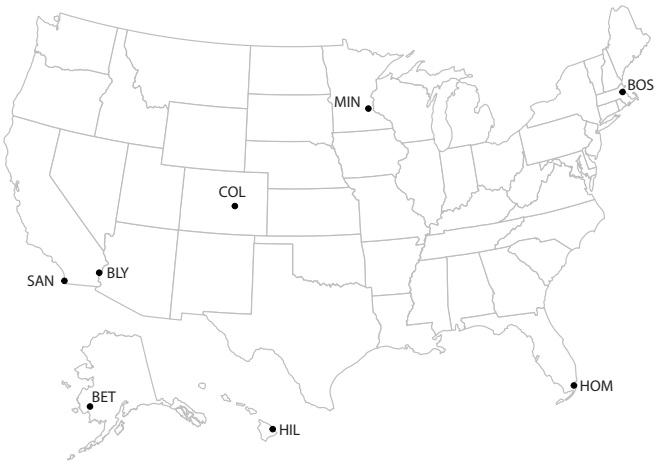

Fig. 1. The eight locations from the GSOD database that are studied.

about the SAT measurements at these eight stations.

The selected stations across the U.S. exhibit different seasonal patterns of SAT, as shown in boxplots in Fig. 2 (arranged very roughly in geographic order). Most stations experience a slow increase of temperature at the end of winter, followed by a faster temperature decrease at the end of the summer, indicating complex seasonality patterns not captured by a single annual harmonic. Bethel in particular sees long winters and short summers. The most northern and continental locations (i.e., all cities in the top row of Fig. 2 and Blythe) exhibit the largest amplitude of seasonality. Coastal stations at Hilo, San Diego, and Homestead experience relatively warm temperatures year round; the climate at Hilo is especially uniform. San Diego has distinctive seasonality during the first part of the year (January to June), showing an almost linear increase in the temperature with an inflection around June–July. Fig. 2 also illustrates year-to-year variability of SAT on each day of the year. Seasonal variability is evident in the bulk of the distributions, represented by the box heights, as well as in the tails. Variability tends to be larger in winter and is particularly strong in the most northern locations such as Bethel. Lower and upper tails often exhibit different seasonal variability, which is evident in the various whisker[1] lengths and the number of observations beyond the whiskers. In Bethel, many more observations lie outside the boxplot whiskers in the summer than in the winter. In Colorado Springs, values beyond the whiskers are more abundant and spread out in the lower tail than in the upper tail. Minneapolis and Homestead also present a lack of values beyond the whiskers in the upper tail during winter months, while Boston and Hilo show this behavior to a lesser extent in the lower tails during summer. The upper tail of San Diego sees the greatest number of observations beyond the whiskers, as well as the most significant spread in extremes. It is noteworthy that tail heaviness does not imply greater variability, as the SAT range in San Diego is the second smallest of the locations considered (larger only than that of Hilo). The observed asymmetries in tail behavior indicate a departure from Gaussianity and the need for non-standard distribution models.

3. Seasonal model for bulk and tails of temperature distribution

In this section, we detail the statistical model used to fit the temperature data described in Section 2. The proposed model relies on a univariate model for bulk and tails (Section 3.1) and is extended to a seasonal model (Section 3.2) with parameters which are estimated by maximum likelihood (Section 3.3).

[1] See caption of Fig. 2 for a description of boxplot features.

3.1. Model for bulk and tails

Stein (2021a) proposed a flexible parametric model which we refer to as the bulk-and-tails (BATs) distribution. The "s" is included in the acronym to distinguish from "bulk-and-tail" models that tend to have limited flexibility in the lower tail. Consider the random variable X whose cumulative distribution function (cdf) $P(X \leq x)$ is given by $T_\nu(H_\theta(x))$, where T_ν is the cdf of a student-t random variable with ν degrees of freedom and H_θ is a monotone-increasing function with six parameters that control the upper and lower tails. To be precise, define $\Psi(x) = \log(1 + e^x)$ and

$$H_\theta(x) = \left\{1 + \kappa_1 \Psi\left(\frac{x - \varphi_1}{\tau_1}\right)\right\}^{1/\kappa_1} - \left\{1 + \kappa_0 \Psi\left(\frac{\varphi_0 - x}{\tau_0}\right)\right\}^{1/\kappa_0} \quad (1)$$

where $(\varphi_i, \tau_i, \kappa_i)$ are the location, scale, and shape parameters of the upper ($i = 1$) and lower ($i = 0$) tails. Like in the GPD and GEV distributions, the parameters κ_i control the shape of the tails, with positive values producing a heavy tailed distribution and negative values producing a thin tailed distribution with bounded support in that tail. Specifically, let (L, U) denote the interior of the support. If $\kappa_1 \geq 0$, the upper bound is $U = \infty$, while if $\kappa_1 < 0$, then $U = \varphi_1 + \tau_1 \Psi^{-1}(-1/\kappa_1)$. Similarly, $L = -\infty$ if $\kappa_0 \geq 0$ and $L = \varphi_0 - \tau_0 \Psi^{-1}(-1/\kappa_0)$ if $\kappa_0 < 0$. The cases $\kappa_i = 0$ are defined by continuity. For example, when both shape parameters equal zero, H_θ reads

$$H_\theta(x) = \exp\left(\frac{x - \varphi_1}{\tau_1}\right) - \exp\left(\frac{\varphi_0 - x}{\tau_0}\right).$$

Taking the derivative of the BATs cdf with respect to x, we obtain the probability density function (pdf) $t_\nu(H_\theta(x))\dot{H}_\theta(x)$, where t_ν is the pdf of the student-t distribution with ν degrees of freedom.

Stein (2021a) showed that the BATs distribution can behave like any three-parameter GPD in either tail. Derivatives of the BATs pdf with respect to model parameters can be calculated analytically, aiding maximum likelihood estimation. Moreover, there is no need to numerically compute any normalizing constant when writing the density, an obstacle often handled with Markov chain Monte Carlo methods (Gelman and Meng, 1998; Møller et al., 2006). Although the BATs distribution does not directly follow from any first-order limit theorems like in traditional extremes methodology, these properties produce a versatile and practical density for modeling purposes.

In this work, we attempt to model nonstationary data by allowing the scale parameters τ_i and location parameters φ_i to change with time, both seasonally and with long-term trends. In some scenarios, it may be appropriate to allow ν and κ_i to also vary with time, but estimation for these parameters is already difficult when they are held constant. With a relatively simple parameterization for τ_i and φ_i, we fit our model to daily SAT records at the eight U.S. cities discussed in Section 2. The code for fitting and working with the BATs model is available via the Julia package BulkAndTails.jl as well as an R interface to this package; see github.com/mlkrock/BulkAndTails.jl for more information. The package includes code for the standard seven-parameter BATs distribution and the nonstationary extension described in the next section.

3.2. Seasonal extension with long-term trend

Here we describe the seasonal model used to fit the daily SAT data to seasonal variations with a long-term trend. To capture nonstationary behaviors in SAT, namely seasonality, climate change trends, and their interaction, we introduce covariates which allow parameters of the H_θ function (1) to depend upon time. Reasonable choices for seasonal covariates include harmonics or a periodic spline basis; we choose the latter for more flexibility in the main seasonal term, as Fig. 2 shows evidence of complex seasonal patterns. For a climate change covariate, we use the logarithm of CO_2 equivalent (denoted log CO_2e) obtained from the PRIMAP emission time series (Gütschow et al., 2016) and freely

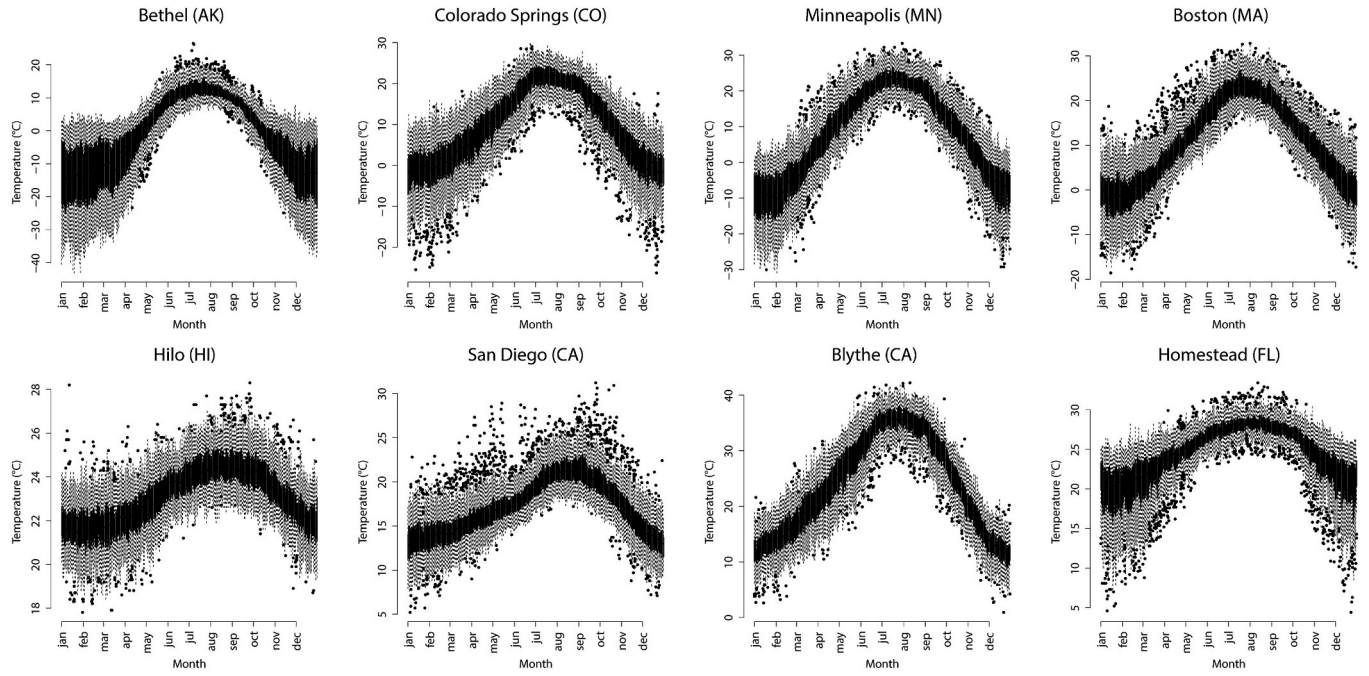

Fig. 2. Boxplots of daily mean SAT at the stations considered across every day of the year for the entire studied time-period. Boxes of the boxplot are delimited by the 0.25 and 0.75 quantiles $q_{0.25}$ and $q_{0.75}$, and whiskers are $q_{0.25} - 1.5 \times \text{IQR}$ and $q_{0.75} + 1.5 \times \text{IQR}$, with $\text{IQR} = q_{0.75} - q_{0.25}$ denoting the interquartile range.

available for download at https://www.pik-potsdam.de/paris-reality-check/primap-hist/. Values for CO_2e are derived from the six Kyoto greenhouse gases (CO_2, CH_4, N_2O, HFCs, PFCs, and SF_6) using the global warming potential from the Fourth Assessment Report. Version 2.2 of the PRIMAP time series is available on a yearly basis through 2018, and we regress these PRIMAP values on the historic Mauna Loa CO_2 dataset to predict the values at 2019 and 2020. In this study, log CO_2e serves as a proxy for anthropogenic climate change. Finally, as seasonal patterns can also be affected by the changing climate, an interaction term between seasonality and long-term trend is added, where the seasonality is modeled with annual harmonics.

Let y represent the year and d represent the Julian day since the first observation, modulo 365.25. Write $S_j(d)$ to denote the jth spline basis function at day d and $C(y)$ to denote the value of log CO_2e at year y. Suppressing the subscripts $i = 0$, 1 for τ_i and φ_i, we use the following parameterizations:

$$\begin{aligned}\varphi(d,y) = \alpha_0 \quad &+\textstyle\sum_{j=1}^{8} \alpha_j S_j(d) \\ &+C(y)\left(\beta_1 + \beta_2 \cos\left(\frac{2\pi d}{365.25}\right) + \beta_3 \sin\left(\frac{2\pi d}{365.25}\right)\right) \qquad (2)\\ \log(\tau(d)) = \gamma_0 \quad &+\textstyle\sum_{j=1}^{8} \gamma_j S_j(d).\end{aligned}$$

The long-term trend and its interaction with seasonality are only considered in the location parameters; we represent seasonality in the interaction term with circular harmonics for parsimony purposes. For simplicity, we use the same seasonal basis functions (periodic cubic splines from the R package pbs (Wang, 2013)) for the location and scale parameters. We fix the number of seasonal basis functions at eight. Using fewer basis functions at cities with more complicated seasonal temperature variations presented problems with the optimization convergence and also a worsened fit based on model diagnostics. We found that the choices made here provided a good overall compromise between capturing the nuances of the seasonal distributions at some sites and limiting the total number of parameters for reasons of computational stability and controlling model complexity.

3.3. Maximum likelihood estimation and its uncertainty

Suppose there are independent observations $x_1, \ldots, x_m$ of temperature at a location. The corresponding likelihood is

$$\prod_{i=1}^{m} t_\nu(H_\theta(x_i))H'_\theta(x_i) \qquad (3)$$

where H_θ is defined as in (1) and t_ν is the student-t pdf with ν degrees of freedom. Note that the likelihood (3) does not explicitly include the temporal dependence between daily observations, although the uncertainty quantification described in Section 3.3 does take this dependence into consideration. Ignoring the temporal dependence does not bias point estimates of marginal distributions of individual days (Varin et al., 2011). In total, there are 45 parameters which are estimated via maximum likelihood: 13 parameters each for the upper and lower locations, 8 parameters each for the upper and lower scales, two shape parameters, and the degrees of freedom ν. We enforced two constraints $\kappa_i/\nu > -0.5$ so that the likelihood is twice-differentiable at its endpoints when κ_i is negative. Optimization was performed in Julia with the IPOPT solver (Biegler and Zavala, 2009) and automatic differentiation from ForwardDiff.jl (Revels et al., 2016) to efficiently obtain the gradient and Hessian. In this high-dimensional parameter space, there is no guarantee that a unique maximum likelihood estimate exists, and different parameter starting values can lead to different local optima. In the rest of this paper, we call parameter estimates "locally optimal" if the optimization software converged and deemed the values a solution. Several different choices for our initial guess were tested. All coefficients related to climate change were set to zero. Seasonal coefficients for the location parameters (i.e., those corresponding to the periodic splines) were taken to be the linear regression coefficients from fitting the periodic splines to

the data. Coefficients for the log scale parameters were set to zero, except for the intercept term, which was profiled over with a small grid of values. We examined different initial values for ν and chose κ_i as either zero or slightly negative. These different initial guesses sometimes led to different local optima, and we explore the effect of these different estimates in more detail in Section 4.1. Unless stated otherwise, results are shown for the set of parameters providing the largest likelihood.

Obtaining uncertainties of estimated parameters is not straightforward. We perform uncertainty quantification with a bootstrapping procedure to account for temporal dependence. Using stratified block bootstrap resamples, it is possible to obtain confidence intervals for any desired function of the parameters (Efron and Tibshirani, 1993). Although classical bootstrapping operates under the assumption of independent and identically distributed data, the block bootstrap is commonly used in the setting of temporal dependence (Lahiri, 2003). Under the assumption that sub-annual temporal dependence is stronger than interannual dependence, selecting a block size of a year is natural choice, and we resample years based on a decadal stratification to preserve aspects of climate change seen over the observation period. That is, to create a bootstrapped dataset, we sample years (with replacement) from each decade, preserving the number of years in each decade from the original dataset. Maximum likelihood estimation is performed on each bootstrapped dataset, and the desired functions of parameters (e.g., quantiles, changes in quantiles over years, or parameters themselves) are computed for each fit. A pointwise 95% basic bootstrap confidence interval for the quantity of interest $\widehat{p}$ is given by $(2\widehat{p} - \widehat{p}_{97.5\%}, 2\widehat{p} - \widehat{p}_{2.5\%})$, where $\widehat{p}_{2.5\%}$ and $\widehat{p}_{97.5\%}$ are the 2.5 and 97.5 percentiles of the bootstrap estimates for $\widehat{p}$ (Paciorek et al., 2018). Bootstrap parameter estimates were obtained using the local optima from the fit on the historical dataset as the initial guess. Example confidence intervals from 200 bootstrap samples are shown later in Fig. 7 in Section 4.3, Fig. 4 in Section 4.1, and Table 4 in Appendix B. We emphasize that these uncertainty estimates are computed for the daily marginal distribution of SAT.

4. Results on changing daily mean temperature distributions

In this section, we present results of the proposed model fit to daily mean SAT. Section 4.1 provides a visual evaluation of the quality of the fit models. Section 4.2 shows a quantitative comparison of the proposed model to several benchmark models. In particular, we compare with skew-normal and generalized Pareto distributions to respectively assess the bulk and each tail of the fit distribution. Finally, in Section 4.3, we highlight how the proposed model is able to capture the changing seasonal patterns due to long-term trends.

To begin, we describe the benchmark distributions which are compared to the BATs distribution. First, with less of an emphasis on the tails of the distribution, we compare to a skew-normal distribution with time-varying location, scale, and skewness parameters. The skew-normal distribution was found to provide reasonable fits for temperatures in Stein (2021b). With φ and Φ respectively denoting the pdf and cdf of a standard normal random variable, the skew-normal pdf is

$$f(x) = \frac{2}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\alpha\left(\frac{x-\mu}{\sigma}\right)\right) \tag{4}$$

where α is a skewness parameter, μ is a location parameter, and $\sigma > 0$ is a scale parameter. We use the same spline-basis parameterization for the location and scale parameters as in (2) and also let the skewness vary in time using the same parameterization as the location parameter but without a climate-change covariate. Maximum likelihood estimation for the skew-normal model is performed in Julia.

To focus on the tails of the distribution, we compare with nonstationary GPD models. The GPD pdf is

$$f(x) = \frac{1}{\sigma}\left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-1-1/\xi} \tag{5}$$

where μ is a user-specified threshold, ξ is a shape parameter, and $\sigma > 0$ is a location parameter. If the shape parameter is negative, the tail is bounded; if the shape is nonnegative, the tail is unbounded. Specifically, the interior of the GPD support is (μ, U) where $U = \mu - \sigma/\xi$ if $\xi < 0$ and $U = \infty$ if $\xi \geq 0$. We construct the GPD threshold μ as a quantile regression[2] at the $p_\mu = 0.95$ quantile using the same seasonal and climate change covariates as the location parameter in (2). For GPD parameters, we consider a constant shape parameter and a temporal scale parameter whose logarithm has the same parametric form as the location parameter in (2). When modeling the lower tail with a GPD, we can multiply the data by -1 and work in the typical peaks-over-threshold setting. All GPD models were fit by maximum likelihood with the R package extRemes (Gilleland and Katz, 2016).

4.1. Visual evaluation of quantiles and densities

Given the model parameter estimates, it is possible to express the distribution of daily mean SAT at any day and year, provided that log CO_2e is available for that year. First, in Fig. 3, we examine estimated quantiles during the year 2020. Quantiles from the BATs model are shown for values 0.0001, 0.001, 0.01, 0.1 (blue), 0.25, 0.5, 0.75 (green), and 0.9, 0.99, 0.999, 0.9999 (red). The black lines in each panel are observed daily maxima, minima, and medians taken across each day of the year over all years. These maximum and minimum values serve as a proxy for upper and lower tail descriptions which would require a large amount of data to quantify precisely. We also emphasize that these maximum and minimum values are taken from the entire time series, unlike the BATs curves that are shown for the year 2020. Overall, the fitted BATs model provides a very flexible representation across the year for the bulk and both tails of the distributions, capturing: a) larger spread in both tails and sometimes in the interquartile range during winter (e.g., Colorado Springs and Homestead); b) different spreads in the lower and upper tails (e.g., San Diego); c) different seasonality patterns in each tail and in the interquartile range, such as in Bethel, where the upper and lower tails exhibit significantly different seasonal patterns; d) and asymmetric seasonal patterns across the year for the eight stations where the temperature increase at the end of the winter is slower than the decrease in early fall. Quantile regression can be used to study quantiles and tail behaviors in similar fashion to Fig. 3, but this approach typically involves separate analyses of the desired quantiles and comes with the possibility of producing crossing quantile curves. To avoid crossing quantiles in quantile regression, practitioners will need to turn to specialized methods (He, 1997; Chernozhukov et al., 2010; Cannon, 2018). The use of the BATs model for the entire distribution eliminates the risk of crossing quantiles.

A close examination of Fig. 3 reveals potential issues in some cities. At San Diego, there is a visible discrepancy between the BATs-estimated median of the data and the observed median. This is unsurprising given that the BATs median is shown for the year 2020 and San Diego has experienced a significant warming, which will be discussed in more detail in Section 4.3. More curious at San Diego is the behavior of the most extreme quantiles. Unlike at the seven other cities, the estimated 0.9999 quantile for each day at San Diego is much higher than the maximum historical observation for that day. Given a record of observations which is 10,000 days long, the expected number of exceedances of the 0.9999 quantile is one. Therefore, we would expect the 0.9999 quantile to be exceeded at least twice in the 76-year record of observations at San Diego, but there are zero exceedances of the 0.9999 quantile for this BATs model. Additionally, there is a drop in the 0.0001

[2] See Appendix C for a description of quantile regression.

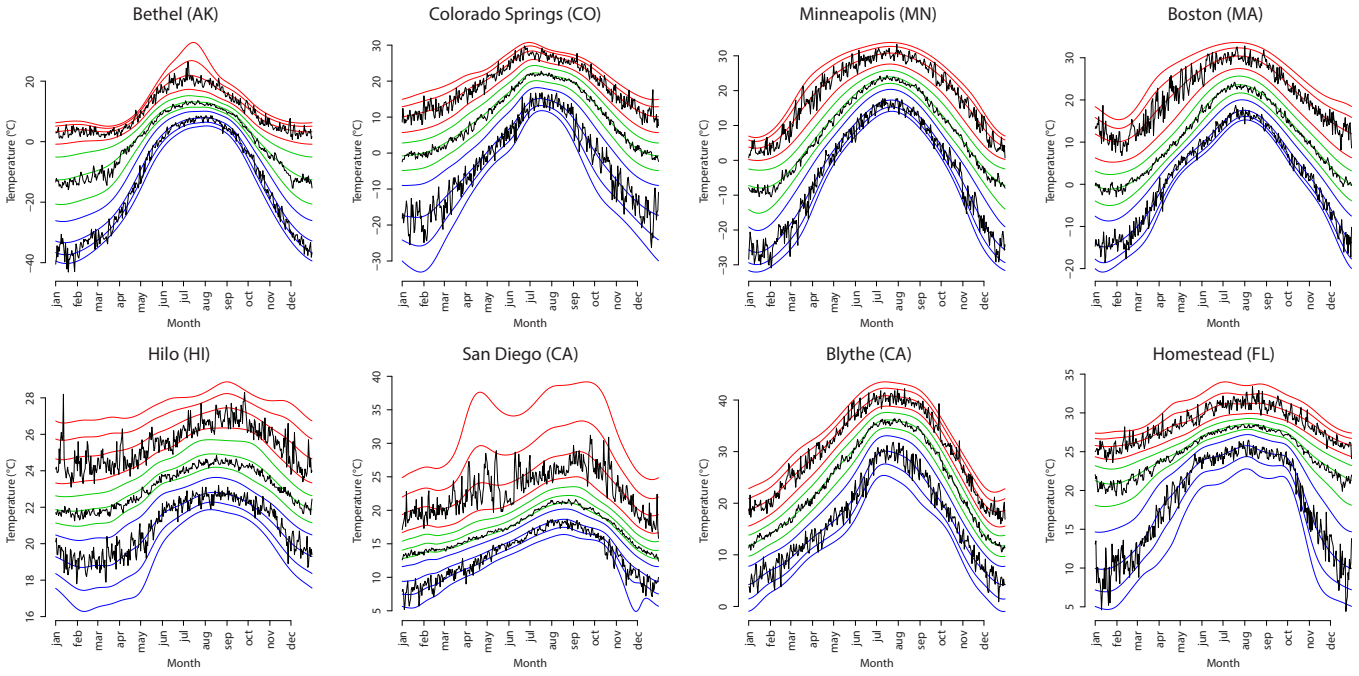

Fig. 3. BATs quantile estimates for the year 2020: 0.0001, 0.001, 0.01, 0.1 (blue), 0.25, 0.5, 0.75 (green), and 0.9, 0.99, 0.999, 0.9999 (red). Black lines show the minimum/median/maximum observation for that day of year, taken over all years. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

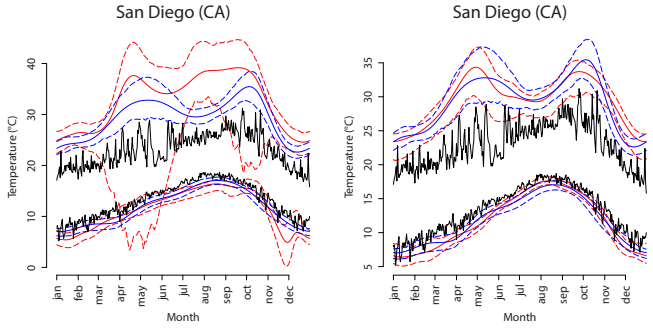

Fig. 4. 0.0001 and 0.9999 quantiles for BATs (red) and GPD (blue) at San Diego, along with a basic bootstrap 95% confidence interval. The left panel shows the BATs maximum likelihood estimate among all locally optimal sets of parameters that were found. This maximum likelihood estimate has positive shape parameters, as shown in Table 4. The right panel shows another local optima which has a slightly lower likelihood but negative shape parameters. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

quantile around the beginning of December which does not appear to be physically based. Fig. 4 shows the difference between using the primary BATs parameters versus a secondary set of locally optimal parameters whose log-likelihood is smaller by about 5 units. Not only are the 0.9999 confidence intervals narrower for the latter solution, but they also agree more closely with the equivalent quantile estimate from a GPD. Here, the GPD 0.998 quantile is used for comparison with the BATs 0.9999 quantile, as the threshold was taken from a 0.95 quantile regression, and $0.95 + 0.05 \times 0.998 = 0.9999$. Note that the 0.9999 quantile estimates from the GPD and second BATs fits are exceeded by historical observations once and twice, respectively, and the 0.0001 quantile estimates do not exhibit the drop in December 2020.

The key difference between the two sets of BATs parameters is the sign of their shape parameters. Typical shape parameter estimates for temperature distributions range from -0.3 to 0 or at times are slightly positive (Nogaj et al., 2006; Gilleland and Katz, 2006; Bommier, 2014). With the maximum likelihood estimate shown in Fig. 3, the shape parameters are both positive and κ_1 is fairly large, while in the second set of parameter estimates (in Fig. 4), both shape parameters are negative. These estimates possess markedly different tail behaviors, yet the log-likelihood values are fairly close. Indeed, there is a natural imbalance in estimating tail behavior with maximum likelihood: a tail estimate which is too large can barely affect the likelihood, yet a tail estimate which is too small can produce a likelihood of zero. Inducing bias in the tails is also a possibility when fitting extremes models that include the bulk of the distribution (Scarrott and MacDonald, 2012). The use of hierarchical models where the shape parameter has a prior that favors negative values could help address these issues. Weighted or censored likelihoods can be used to emphasize the greater importance of tail observations relative to bulk observations. Stein (2021a) and Stein (2021b) also discuss modifications of the seven-parameter BATs distribution that can provide more flexibility. We mention these options as means to correct misfitting in the far tails of the BATs distribution but do not explore them further in this paper. We suggest that the bootstrap uncertainties shown in Fig. 4 can act as a warning against erroneous behavior in the far-tail quantiles. These bootstrap confidence intervals for quantiles at other locations appeared reasonable, while at San Diego, the extremely large uncertainty in the 0.9999 quantile is a sign that BATs is likely misfitting the upper tail.

The difficulty in estimating the upper tail behavior at San Diego may be due to the complicated two-peaked seasonal pattern and high variability which emerge when looking far in the upper tail. As shown in Fig. 2, San Diego's hottest temperatures decreased in magnitude from spring to summer, and then increased again in early fall. Other cities may have problems in the most extreme quantile estimates, yet they do not seem to be problematic to the extent of those in San Diego. The lower quantiles in Colorado Springs, Hilo, and Homestead appear to be too low in some parts of the year, given that Fig. 3 is shown in the final observation year 2020, when these quantiles are at their largest. Another interesting point in Fig. 3 is the large spike in January in Hilo, which corresponds to heavy rainfall and flash flooding on January 12, 2020.

BATs estimates for Hilo exhibit a similar phenomenon to San Diego where a secondary set of parameters had slightly lower likelihood and negative shape parameters, while the primary set of parameters had positive shapes and a very large ν, indicating tail behavior similar to a Gaussian distribution. In terms of the primary BATs model, this January 12 event corresponds to the $1 - 1.5 \times 10^{-6}$ quantile, which is extremely far in the tail given the 50-year duration of the Hilo data. After refitting with this value removed, the January 12 value unsurprisingly lies even further in the tail at the $1 - 7 \times 10^{-7}$ quantile. With the secondary BATs estimates which had negative shape parameters, the event corresponds to the $1 - 8 \times 10^{-6}$ quantile, which is more reasonable, but after refitting with the outlier removed, the observation is the least likely overall at the $1 - 2 \times 10^{-7}$ quantile. This echoes the point from the previous paragraph that locally optimal estimates with very different tail behaviors can have similar likelihood values. It also illustrates how a single outlying observation can have a drastic impact on tail behavior. We note another important property of BATs in that the lower tail behavior κ_0/ν is unaffected by the January 12, 2020 observation and remains stable even after removal of this outlier.

To further explore the distributions across the year, we plot individual estimated monthly densities for the first day of each of January, April, July, and October from parametric models (BATs and skew-normal) along with empirical kernel density estimates (KDE) in Fig. 5. KDE curves come from a Gaussian kernel density estimate using observations from a fifteen-day window over all available years.[3] Seasonal BATs and skew-normal densities are calculated as a mixture of their yearly densities from the first observation year to 2020. The three density curves at each location largely agree with each other for all seasons and locations, but there are some notable differences. For example, in the winter and spring at Bethel, the KDE suggests a multimodal density which cannot be captured by the skew-normal estimate. In contrast, the BATs distribution has the ability to model bimodal distributions (as shown in Stein (2021a)). The presence of pronounced shoulders to the distributions which the BATs distribution can model but the skew-normal cannot is also found in the winters of Minneapolis and Homestead. These results suggest that the seasonal BATs model can capture daily mean temperature distributions more reliably than a skew-normal distribution and provide flexible densities matching observed ones. A more quantitative comparison of the skew-normal and BATs models is continued in the next section, along with a closer examination of tail properties.

4.2. Cross-validation comparison

We further compare our methodology to two relevant models through a quantitative analysis. To compare entire distributions, we calculate the cross-validated continuous ranked probability score (CRPS) (Gneiting and Raftery, 2007) of the BATs and skew-normal models over the entire year and for each of the four seasons individually. CRPS is typically used in the evaluation of probabilistic forecasts as a tool to assess calibration (statistical consistency between forecast and verification data) and sharpness (dispersion of the distribution to be evaluated). By definition, CRPS values are nonnegative, and smaller values indicate a better match. The CRPS between a single observation x and cdf F is defined as

$$CRPS(F, x) = \int_{-\infty}^{\infty} (F(y) - \mathbf{1}[y \geq x])^2 \, dy. \tag{6}$$

The CRPS is typically reported as the average of instantaneous CRPS as in (6) over a set of observations. Scores here are calculated in a cross-validation procedure where folds are obtained from blocks of four or five consecutive years, leading to between 10 and 20 folds depending upon the station. A scaled version of the average difference (BATs minus skew-normal) for CRPS values within a given fold is shown as a percentage in Table 1. In all cities, the CRPS difference value is negative, indicating that the BATs model fits better over the whole year, but there are some combinations of seasons and cities where the skew-normal performs better. The most noticeable improvement provided by BATs is seen in Bethel, particularly during the wintertime, consistent with the visual evidence of skew-normal misfitting the winter months shown in Fig. 5.

To compare the GPD and the non-thresholded models (BATs and skew-normal), we use a variant of the CRPS. A direct CRPS comparison is not sensible because the GPD only fits a single tail of a distribution. One option to deal with this is to change the cdf in the CRPS definition (6) to the cdf conditional on being larger than the threshold. Conditioning puts the cdf of the BATs model on the same range, [0, 1], as the GPD model when looking above the threshold. However, this approach does not take into account the model's ability to estimate the probability of being larger than the chosen threshold. Instead, we consider a new random variable

$$Z = \begin{cases} \mu & X \leq \mu \\ X & X > \mu, \end{cases}$$

which is censored to equal the threshold when an observation falls below the threshold. For the BATs and skew-normal models, which fit the entire temperature distribution, this censored cdf equals

$$F_Z(z) = \begin{cases} 0 & z < \mu \\ F_X(z) & z \geq \mu, \end{cases}$$

and for the GPD models, the censored cdf equals

$$F_Z(z) = \begin{cases} 0 & z < \mu \\ p_\mu + (1 - p_\mu) F_X(z) & z \geq \mu. \end{cases}$$

As indicated by the common notation, the censoring value μ and the GPD threshold μ are identical, but in theory they could be different. Here, $p_\mu = 0.95$ corresponds to the quantile regression level that creates the threshold μ.

To assess only tail behavior, we consider a weighted CRPS (Gneiting and Ranjan, 2011; Taillardat et al., 2019) where the weight function in the integral equals one if the observation lies in the tail and zero otherwise. To be precise, the indicator-weighted score expresses as

$$wCRPS(F, x; q) = \int_{-\infty}^{\infty} (F(y) - \mathbf{1}[y \geq x])^2 \mathbf{1}[y \geq q] \, dy$$

where F is a censored cdf described above. The role of q is to examine behavior of the distribution far in the tail by ignoring values smaller than q. Both the GPD threshold μ and the wCRPS threshold q are obtained from a cross-validated quantile regression with eight periodic splines as covariates, with μ taken as the $p_\mu = 0.95$ quantile and q as the p_q quantile, where $p_q \geq 0.95$. A similar procedure can be conducted to assess the lower tails of the distributions.

Results from this comparison are shown in Table 2. For the most part, the BATs model provides a slightly better fit than the GPD in this tail analysis, supporting the idea that studying nonstationary extremes with all possible data can improve the quality of fit in the tails. In Bethel, the skew-normal distribution fits the lower tail poorly, due primarily to its inability to reproduce behavior in DJF, as seen earlier in Fig. 5 and Table 1. Besides the lower tail in Bethel, the BATs model and skew-normal are relatively comparable with a modest advantage to BATs. While the GPD and BATs fits at Bethel are both much better than skew-normal, the GPD outperforms BATs in the smallest quantiles. The lower tail of Bethel was the only situation considered where we could not find any local BATs optima without a large positive shape parameter, which may be indicative of some overcompensation to better fit the bulk of the distribution. As discussed in Section 4.1, the possibility that the BATs fit worsens the farther we look into the tails is most apparent in the upper

[3] See Appendix D for a more precise definition of the kernel density estimate.

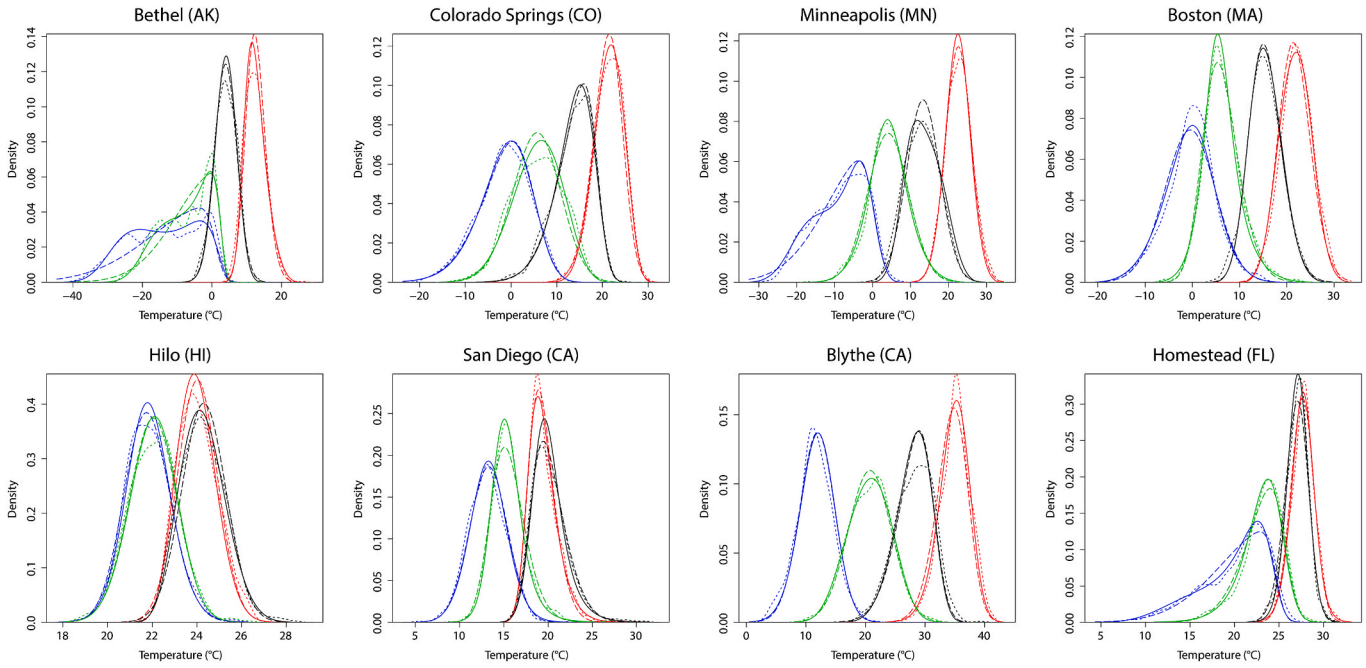

Fig. 5. Density estimates for SAT distributions from the seasonal BATs model (solid), skew-normal model (long dash), and a kernel density estimate (short dash). Blue, green, red, and black curves correspond to the first day of January, April, July, and October, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Cross-validated CRPS comparison between the entire distribution for BATs and skew-normal. Table values equal $100\sum_{k=1}^{K}\texttt{mean}(\texttt{BATs}(k) - \texttt{Skew}(k)) \,/\, \sum_{k=1}^{K}\texttt{mean}(\texttt{BATs}(k))$ where k is a cross-validation fold. Negative values indicate better performance in the BATs fit. Results are tabulated for the entire year ('Year' row) and by season (DJF: December-January-February, MAM: March-April-May, JJA: June-July-August, SON: September-October-November).

	BET (AK)	BLY (CA)	BOS (MA)	COL (CO)	HIL (HI)	HOM (FL)	MIN (MN)	SAN (CA)
Year	−0.95	−0.14	−0.10	−0.10	−0.27	−0.18	−0.01	−0.44
DJF	−3.04	−0.27	−0.05	−0.07	−0.22	−0.12	−0.18	−0.22
MAM	−0.57	0.09	−0.10	−0.11	−0.07	−0.05	−0.08	−0.70
JJA	0.03	−0.48	−0.15	−0.21	−0.18	−0.12	0.08	−0.42
SON	−0.26	0.09	−0.10	0.01	−0.60	−0.39	0.12	−0.40

tail of San Diego. This hypothesis is indeed reflected in these wCRPS values at San Diego; however, the scores still suggest that the quality of the BATs fit is competitive with that of the GPD, as there is not a large difference like with the skew-normal in the lower tail at Bethel. We note that the wCRPS may not be the best way to compare distributions when one is mainly interested in extreme events (Taillardat et al., 2019). The BATs scores also degrade below GPD scores as we look farther into the upper tail of Boston and the lower tail of Colorado Springs, although the

Table 2
Cross-validated wCRPS comparison of the tail behavior beyond quantile p_q. Estimates of μ (GPD threshold where observations are censored; corresponds to $p_\mu = 0.95$ quantile (upper tail) or $p_\mu = 0.05$ quantile (lower tail)) and q (wCRPS threshold; corresponds to p_q quantile in left column) are obtained from cross-validated quantile regressions with eight periodic splines as covariates. Table values equal $100\sum_{k=1}^{K}\texttt{mean}(\texttt{BATs}(k) - \texttt{GPD}(k))/\sum_{k=1}^{K}\texttt{mean}(\texttt{GPD}(k))$ and $100\sum_{k=1}^{K}\texttt{mean}(\texttt{Skew}(k) - \texttt{GPD}(k))/\sum_{k=1}^{K}\texttt{mean}(\texttt{GPD}(k))$ where k is a cross-validation fold. Negative values indicate better performance than the GPD fit.

p_q	BET (AK)		BLY (CA)		BOS (MA)		COL (CO)		HIL (HI)		HOM (FL)		MIN (MN)		SAN (CA)	
	BATs	Skew	BATs	Skew	BATs	Skew	BATs	Skew	BATs	Skew	BATs	Skew	BATs	Skew	BATs	Skew
0.95	−0.50	−0.42	−0.51	−0.87	−0.13	−0.03	−0.46	−0.30	−4.32	−2.89	−2.41	−2.38	−0.28	−0.25	−1.20	−1.12
0.99	−0.11	−0.19	−0.71	−1.05	−0.10	0.03	−0.35	0.10	−2.15	−1.12	−2.26	−2.17	−0.09	0.00	−0.04	−0.04
0.995	−0.07	−0.16	−0.84	−1.19	−0.06	0.06	−0.36	0.18	−1.16	−0.63	−2.06	−1.89	−0.09	−0.01	0.18	0.00
0.999	−0.14	−0.11	−1.89	−2.17	0.03	0.18	−0.26	0.33	−0.57	−0.33	−1.61	−1.41	−0.05	0.02	0.51	0.05
0.9995	−0.26	−0.12	−2.11	−2.34	0.02	0.26	−0.14	0.69	−0.48	−0.32	−1.76	−1.56	−0.01	0.13	0.95	−0.03
0.9999	−0.35	−0.09	−3.58	−3.70	0.04	0.35	−0.10	0.87	−0.40	−0.29	−1.70	−1.48	−0.05	0.10	1.08	0.01
0.05	−0.44	3.34	0.27	−0.27	−0.13	−0.06	−0.23	−0.21	−0.74	−0.46	−0.24	−0.30	−0.37	0.15	−1.50	−0.10
0.01	−0.07	13.38	0.91	−0.26	−0.05	0.02	−0.09	−0.09	−0.50	0.03	0.18	−0.04	−0.30	0.90	−0.82	0.15
0.005	−0.09	20.44	0.94	−0.37	−0.06	0.00	−0.06	−0.06	−0.41	0.07	0.30	0.06	−0.25	1.18	−0.45	0.13
0.001	0.19	58.81	1.86	−0.22	−0.05	0.05	−0.01	−0.04	−0.37	0.23	1.04	0.39	−0.14	1.88	−0.44	−0.02
0.0005	0.35	69.44	1.30	−0.19	−0.06	0.17	0.05	−0.05	−0.27	0.36	1.16	0.56	−0.17	1.62	−0.38	−0.04
0.0001	0.33	148.6	1.64	−0.16	−0.05	0.20	0.04	−0.03	−0.23	0.58	0.42	0.14	−0.15	3.02	−0.55	−0.13

drop in quality is minor compared to that in San Diego and Bethel. As an illustration of the separate behaviors of BATs tails, recall that boxplots for Colorado Springs in Fig. 2 showed asymmetric behavior in the upper and lower tails, with the lack of observations beyond the upper boxplot whiskers indicating a non-Gaussian upper tail. Accordingly, the BATs and skew-normal wCRPS values in the lower tail are quite similar, with BATs slightly worse in the far lower tail, yet BATs provides a better fit in the upper tail. In other cases where the performance of BATs is particularly poor—namely, the lower tails at Blythe and Homestead—the trouble comes from misfit in the first cross-validation fold, where fitting was done on the hold-out data which did not include initial observation years due to the configuration of the cross-validation folds. The temperature records at these two locations had significant gaps after their earliest years, with Homestead missing over a decade, and Blythe missing over 30 years (see Appendix A). Table 2 provides more evidence that extrapolation into the far tails of the BATs distribution can be problematic. We recommend examining bootstrap uncertainties to scrutinize the fit of far tail quantiles before extrapolating into the far tails.

4.3. Changing distributions over the years

We conclude our analysis by illustrating the ability of the proposed nonstationary BATs model to capture changes in SAT distributions and their seasonal patterns between years in a nonstationary climate. Our model explicitly accounts for a proxy of climate change through log CO_2e incorporated in the model parameters and through the interaction between seasonality and this long-term trend. First, in Fig. 6, we show how estimated seasonal quantiles from the seasonal BATs model fits evolve over the period of the observational record. Specifically, the graphs use gradations of shading from grey to black to show time evolution of the differences of the estimated 0.0001, 0.1, 0.5, 0.9, and 0.9999 yearly quantiles from the 0.5 quantile at the starting year. Most combinations of quantiles and stations show a warming trend (the curves become darker with increasing time). However, the details of this trend vary substantially across locations, quantile levels, and time of the year. Bethel displays substantial warming, particularly in its lower quantiles at winter months. Spring and summer months in San Diego also experience a prominent increase in the 0.9999 quantile. The long-term trend and its interaction with seasonal patterns can capture evolving seasonal patterns. Changes in seasonality over the years exhibit very different patterns depending upon station location and the quantile level considered. For instance, Hilo has the largest increases in median and hot temperatures over late summer and early fall, whereas cold quantiles at Bethel warm the most during winter. A notable departure from warming is seen at Hilo, Blythe, and Colorado Springs, where the 0.0001 quantile shows a cooling trend in August through December. Many cold quantiles exhibit a less pronounced seasonality in the recent years compared with past years. In particular, the seasonality of the winter cold quantiles at warmer cities flattens over time; this phenomenon is less prominent in the warm quantiles.

Fig. 7 provides uncertainty estimates for the changes in the 0.0001 and 0.9999 quantiles between the first and last years of the studied period. These results are obtained from the stratified block bootstrap procedure accounting for temporal dependence discussed in Section 3.3. Observe that, in general, the uncertainty bounds are not symmetric around the estimated quantile change. In particular, San Diego exhibits a very large uncertainty associated with the change in the 0.9999 quantile from April to June and September to November, consistent with the high variability in the upper tail observed in Figs. 2 and 3 at those times of the year. Table 4 shows that the BATs model in San Diego favors a heavy upper tail, and the confidence interval for the shape parameter of the upper GPD tail also includes positive values, indicating that a heavy tail is plausible. As shown in Fig. 4, the 95% bootstrap confidence interval for the 0.9999 quantile at San Diego is very wide. Fig. 4 also shows the same results for a secondary set of BATs parameters with negative shape values, which displayed more realistic behavior in the far tails and narrower bootstrap confidence intervals. Note that the shape parameters for bootstrapped data typically displayed the same sign as the shape estimates for the historical data, as the historical parameter estimates were used as the initial guess in the bootstrap parameter estimation. Indeed, this large upper tail uncertainty in San Diego must be interpreted with caution, but overall, the block bootstrap helps us quantify uncertainty without neglecting temporal dependence or

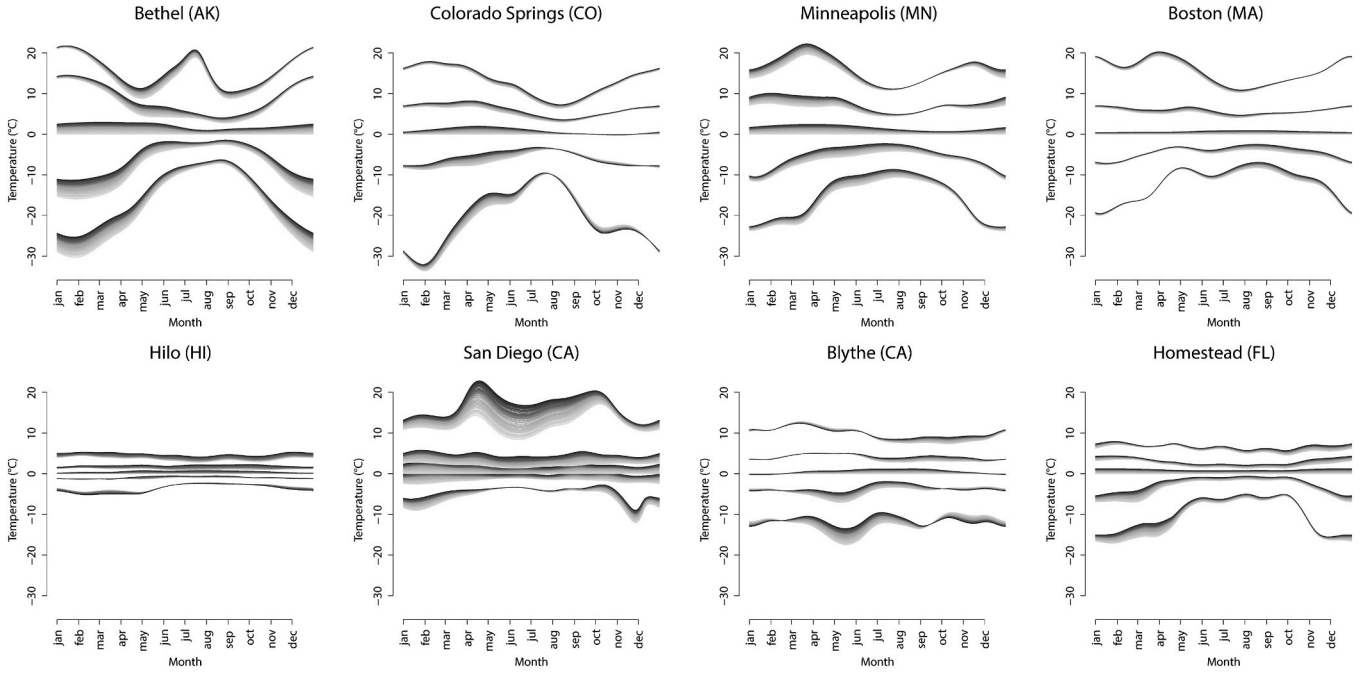

Fig. 6. Changes of SAT quantiles from the seasonal BATs model from the beginning of the observation period (light) to the end (dark). Curves are the 0.0001, 0.1, 0.5, 0.9, and 0.9999 quantiles minus the 0.5 quantile at the earliest year. To emphasize differences in magnitude between stations, all subplots have the same vertical axis.

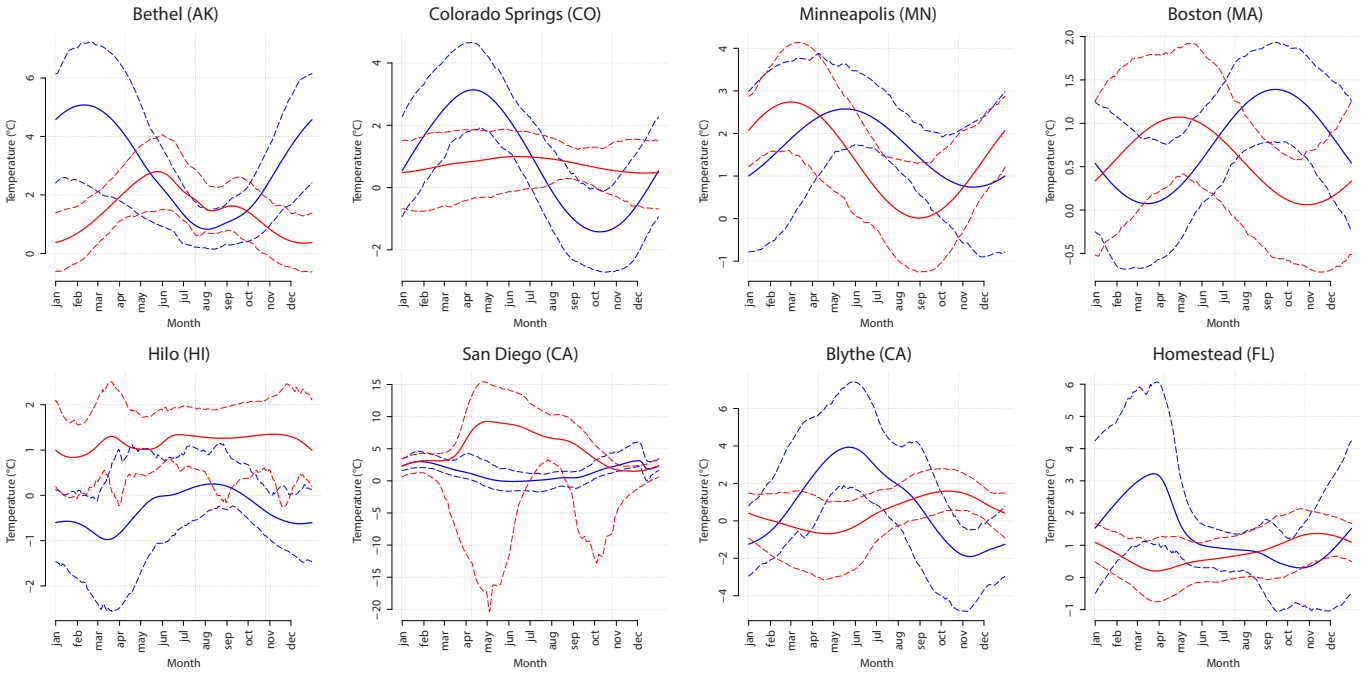

Fig. 7. Basic bootstrap uncertainty estimates for the change in 0.0001 (blue) and 0.9999 (red) SAT quantiles from the starting observation year to 2020. Solid line represents the estimated quantile change, and dashed lines represent pointwise 95% bootstrap confidence intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

climate change.

Finally, we consider the yearly evolution of quantile differences across the study period (Fig. 8), which enables us to quantify changes in the spread of bulk and tails of the distribution. Specifically, we examine the interquartile range and quantile differences in the tails to assess how the bulk spread and tail spreads change over time. For comparison with the BATs model fits, quantile estimates are also obtained from a quantile regression whose covariates are eight periodic splines to quantify seasonal patterns, a long-term trend to quantify climate change, and an interaction term for long-term change in seasonal patterns (i.e., same covariates as the location parameter in (2)). We observe a strong spatial variability in the behavior of quantile spreads. In general, the width of

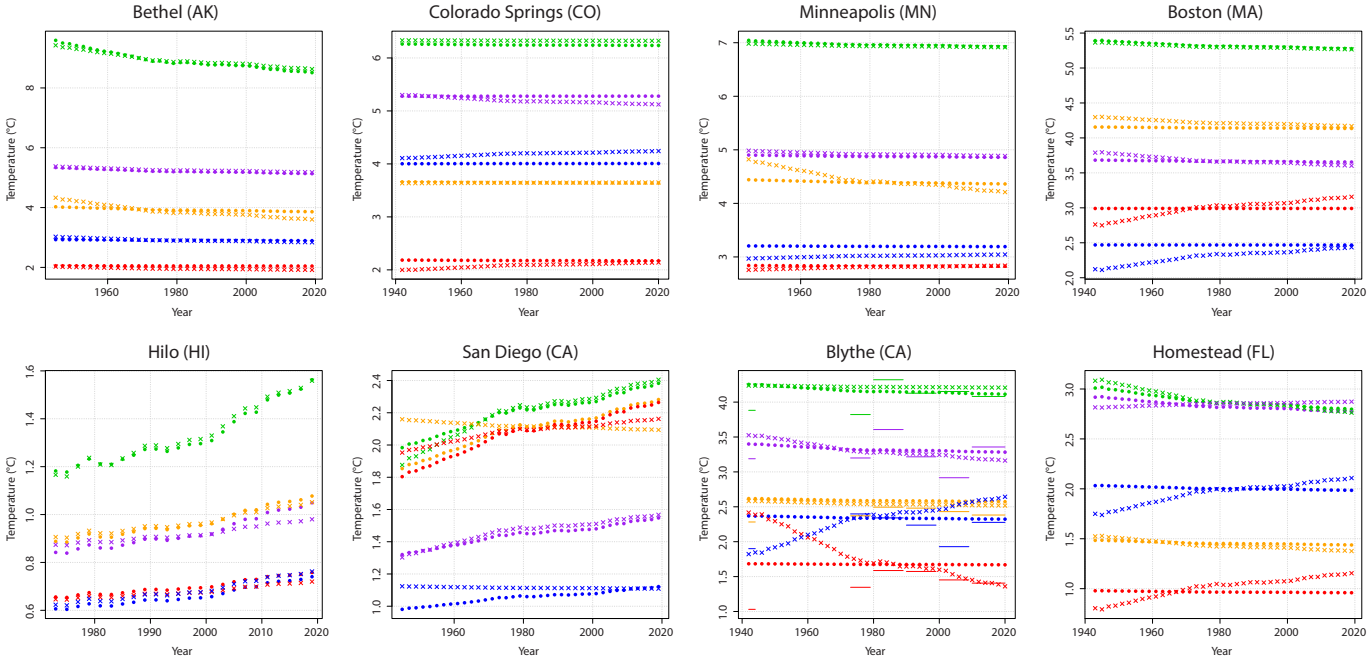

Fig. 8. Average differences between quantiles for each year based on the BATs model (o) and quantile regression (x). Only every other year is shown. Red is $q_{0.99} - q_{0.95}$, orange is $q_{0.95} - q_{0.75}$, green is $q_{0.75} - q_{0.25}$, purple is $q_{0.25} - q_{0.05}$, and blue is $q_{0.05} - q_{0.01}$. Due to missing data at Blythe, we add lines which show stationary quantile regressions for each decade of data. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

the bulk of the distribution decreases over time, with the IQR decreasing over time in most places (e.g., Bethel) but increasing elsewhere (e.g., Hilo). Quantile differences in both the near (orange and purple) and far (red and blue) tails experience different rates of changes for cold and warm temperature. Huang et al. (2016) and Rhines et al. (2017), respectively in climate projections and historical reanalysis, reported a decrease in variability of winter temperature in most North American regions. The far warm tails (red) often show a stretching trend, whereas the near cold tails (purple) tend to contract. For most locations considered, the far tail spreads (red and blue) are smaller than the near tail spreads (orange and purple), with a few exceptions. In Colorado Springs and Homestead, both lower tail spreads are larger than the upper tail spreads. In San Diego, both upper tail spreads are larger than the lower tail spreads, and the far upper tail spread has changed at a pace that rivals the interquartile range. At Hilo, all measures of spread appear to be increasing over time. For the most part, results from BATs and quantile regression agree with one another in capturing these aspects of SAT variability. The largest discrepancy is in Blythe, where the BATs model lacks the increase in far lower tail spread and decrease in far upper tail spread found in results for the quantile regression. This difference seems to result from the many years of missing data at Blythe, as a stationary quantile regression for each decade (shown with horizontal lines) does not exhibit this crossing of far upper and lower tail spreads. Other stations also display a general characteristic of "flatness" in the BATs results when compared to the corresponding estimates from quantile regression. Allowing other parameters in the BATs model to depend upon log CO_2e may help capture more subtle year-to-year changes in distribution.

5. Discussion and conclusions

In this paper, we have applied the seasonal bulk-and-tails (BATs) model with a long-term trend to daily mean surface air temperature data from eight cities in the United States. The model has demonstrated its ability to capture different seasonality and climate change patterns as well as different behaviors in the upper and lower tails. A cross-validation comparison shows superior fit of the BATs distribution to the skew-normal distribution on the entire distribution at all cities and to the generalized Pareto distribution in both tails at many cities.

For practitioners, this model offers a great deal of flexibility with a closed-form density that is relatively easy to work with. Only a few statistical models proposed in extremes literature possess a density describing the bulk and both tails (Naveau et al., 2016; Tencaliec et al., 2020; Stein, 2021a,b). Extreme value analysis is still traditionally performed with block-maxima or peaks-over-threshold techniques, which limits analysis to a tail at a time and avoids an all-encompassing model that describes the entire distribution (Coles, 2001). The proposed framework also allows parameters in the BATs distribution to change over time, both seasonally and long-term. A stratified block bootstrap procedure was conducted for uncertainty quantification.

Of the seven parameters in the standard BATs model, we have taken the shape parameters κ_0 and κ_1 and the smoothness ν to be constant over time. This choice was made due to the difficulty of estimating the shape and smoothness even when these parameters are constant. To capture long-term variation of the nonstationary parameters, we included an annual log CO_2e covariate in the location parameters. Other parameters could depend upon this covariate, particularly the scale parameters, which would reflect a long-term change in the temperature variability, or the shape parameters, which would reflect long-term changes in tail behavior. Fig. 8 indicates that it may be worth allowing other parameters to depend on the long-term trend and its seasonal interaction. Although log CO_2e serves as a proxy for climate change induced by greenhouse gases, it is important to note that not all long-term changes in this paper are due to global-scale anthropogenic climate change. Some sites may be affected by localized changes, particularly the larger cities and those with measurements taken at airports.

The proposed model provides a flexible statistical tool to examine the bulk and tails of distributions changing over time. We have quantified how different parts of the SAT distribution (i.e., different quantiles) exhibit different seasonal and long-term changes at cities representative of different geographies and local climates. Hot and cold quantiles tend to experience larger long-term changes than median quantiles and also present different seasonal patterns. In particular, studied locations with warmer climates experience the most warming for hot quantiles in summer, and locations with colder climates experience the most warming for cold quantiles in winter. Quantifying the robustness of these changes can be further studied with large climate model ensembles (Huang et al., 2021).

An area where BATs can potentially struggle is in extrapolation far into the tails of the distribution. Fig. 4 highlights the issue at San Diego, where the primary set of BATs parameters indicated a heavy upper tail and a large amount of uncertainty in the 0.9999 quantile. A secondary set of parameters with lower likelihood but negative shape parameters produced more realistic quantiles which were closer to the historical observations and showed greater confidence in the quantile estimates. Climate model large ensembles could be used to perform detailed tests of the accuracy of BATs tail estimates and their uncertainties in an idealized setting (Haugen et al., 2018, 2019). Assessing bootstrap confidence intervals for extreme quantiles can help diagnose whether BATs is misfitting the far tails of the distribution. It is possible to add new parameters to the BATs distribution to provide even greater flexibility in the tails. Allowing the shape parameter to change over time could also help here, but using a constant shape parameter in this work was shown to be relatively effective and already difficult to estimate. Weighting or censoring the likelihood function to downplay the influence of the bulk of the distribution could also improve the fit in the tails. Bayesian modeling is another approach which could regulate tail behavior in two ways: first, by placing appropriate priors on shape parameters (e.g., favoring slightly negative values for daily temperature data), and second, by allowing shape parameters to vary smoothly in space according to a Bayesian hierarchical model. Indeed, it can be difficult to estimate a shape parameter using observations from only a single site (Stein, 2020).

Finally, the proposed model is fit to observations while neglecting spatial and temporal dependence. Modeling multidimensional extremes (as would be needed to explicitly account for spatial and/or temporal dependence) is an on-going challenge for the community (Huser and Wadsworth, 2020), where various techniques have been proposed, relying on hierarchical models (Gaetan and Grigoletto, 2007; Cooley et al., 2007), copulas (Lee and Joe, 2018; Krupskii and Genton, 2021), and conditional modeling (Wadsworth and Tawn, 2019; Simpson and Wadsworth, 2021; Huang et al., 2021). Most of these methods ignore the bulk of the data or only focus on a single tail. Developing a bulk-and-tails model for multiple variables is a formidable task because of the wide range of behaviors that can occur in distributions for multivariate extremes (Huser and Genton, 2016; Huser et al., 2017; Huser and Wadsworth, 2019; Huang et al., 2019a; Davison et al., 2013).

Author statement

1. Mitchell Krock: Conceptualization, Software, Formal Analysis, Writing - Original Draft
2. Julie Bessac: Conceptualization, Software, Formal Analysis, Writing - Original Draft
3. Michael Stein: Methodology, Supervision, Writing - Review & Editing, Project Administration
4. Adam Monahan: Supervision, Writing - Review & Editing, Project Administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

Acknowledgements

The effort of Mitchell Krock, Julie Bessac and Michael Stein is based in part on work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR) under Contract DE-AC02-06CH11347. Adam Monahan acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) (funding reference RGPIN-2019-204986).

Appendix A. Complementary Data Information

Table 3
Summary description of the data. In particular, the percentage of daily mean SAT measurements taken with an average of over 20 h of data, the starting year of measurements, and any years which are entirely missing data. End year is 2020.

	%	Start	Missing
BET (AK)	89	1945	
BLY (CA)	86	1942	1945–1972
BOS (MA)	95	1943	
COL (CO)	100	1942	1965–1972
HIL (HI)	99	1973	
HOM (FL)	94	1943	1946–1955, 1971–1972, 2000–2004
MIN (MN)	100	1945	1965–1972
SAN (CA)	98	1945	

Appendix B. Parameter Estimates

Table 4
Parameter estimates with 95% confidence interval. Two left columns are tail behaviors of the BATs model, and two right columns are tail behaviors of the GPD models. Confidence intervals were obtained from the basic bootstrap method with 200 block bootstrap samples.

	κ_0/ν		κ_1/ν		ξ_0		ξ_1	
BET (AK)	0.005	(0.000, 0.019)	−0.015	(−0.030, 0.019)	−0.219	(−0.254, −0.109)	−0.163	(−0.196, −0.091)
BLY (CA)	−0.034	(−0.080, 0.001)	−0.005	(−0.025, 0.019)	−0.218	(−0.271, −0.109)	−0.099	(−0.149, −0.011)
BOS (MA)	−0.245	(−0.276, −0.164)	−0.190	(−0.231, −0.122)	−0.204	(−0.232, −0.124)	−0.219	(−0.261, −0.151)
COL (CO)	−0.088	(−0.152, −0.041)	−0.156	(−0.241, −0.044)	−0.186	(−0.232, −0.120)	−0.133	(−0.171, −0.059)
HIL (HI)	0.000	(−0.014, 0.009)	0.000	(−0.024, 0.000)	−0.212	(−0.247, −0.093)	−0.082	(−0.107, 0.005)
HOM (FL)	−0.002	(−0.037, 0.047)	−0.009	(−0.036, 0.029)	−0.208	(−0.249, −0.131)	−0.101	(−0.142, 0.042)
MIN (MN)	−0.220	(−0.282, −0.166)	−0.251	(−0.293, −0.202)	−0.164	(−0.186, −0.087)	−0.237	(−0.284, −0.172)
SAN (CA)	0.003	(−0.028, 0.147)	0.119	(−0.001, 0.216)	−0.144	(−0.172, −0.071)	−0.021	(−0.055, 0.043)

Appendix C. Quantile Regression

The τ^{th} quantile $(0 < \tau < 1)$ of a random variable Y with cdf F is defined as $Q_Y(\tau) = \inf\{y\colon \tau \le F(y)\}$. Quantile regression is used to estimate quantiles as a linear function of covariates X; that is, $X\beta \in \mathbb{R}^n$, where X is a $n \times p$ matrix of covariates. Estimates of the coefficient vector are given by

$$\widehat{\beta} = \underset{\beta \in \mathbb{R}^p}{\arg\min} \sum_{i=1}^{n} \rho_\tau(y_i - x_i\beta)$$

where x_i is the ith row of X and $\rho_\tau(y) = y(\tau - \mathbf{1}[y < 0])$. All quantile regressions were fit with the R package quantreg (Koenker, 2021).

Appendix D. Kernel Density Estimate

Suppose we are interested in the KDE at day of year d_0. Let $A_{1943} = \{d_{-7}, \ldots, d_7\}$ be the window of days centered at d_0 for the first observation year, say 1943. With $A = A_{1943} \cup \cdots \cup A_{2020}$, the KDE formula reads

$$K_h(x) = \frac{1}{|A|} \sum_{x_i \in A} K\left(\frac{x - x_i}{h}\right)$$

where K is a Gaussian kernel function and h is a bandwidth. Estimates were obtained from the R function "density" with bandwidth selected according to Sheather and Jones (1991).

References

Biegler, L., Zavala, V., 2009. Large-scale nonlinear programming using ipopt: an integrating framework for enterprise-wide dynamic optimization. Comput. Chem. Eng. 33, 575–582.

Bommier, E., 2014. Peaks-Over-Threshold Modelling Of Environmental Data. PhD thesis. Uppsala University.

Bopp, G.P., Shaby, B.A., 2017. An exponential–gamma mixture model for extreme Santa Ana winds. Environmetrics 28 (8), e2476.

Busby, J.W., Baker, K., Bazilian, M.D., Gilbert, A.Q., Grubert, E., Rai, V., Rhodes, J.D., Shidore, S., Smith, C.A., Webber, M.E., 2021. Cascading risks: understanding the 2021 winter blackout in Texas. Energy Res. Social Sci. 77, 102106.
Cannon, A., 2018. Non-crossing nonlinear regression quantiles by monotone composite quantile regression neural network, with application to rainfall extremes. Stoch. Environ. Res. Risk Assess. 32.
Carreau, J., Bengio, Y., 2009. A hybrid Pareto model for asymmetric fat-tailed data: the univariate case. Extremes 12 (1), 53–76.
Cheng, L., AghaKouchak, A., Gilleland, E., Katz, R.W., 2014. Non-stationary extreme value analysis in a changing climate. Climatic Change 127 (2), 353–369.
Chernozhukov, V., Fernández-Val, I., Galichon, A., 2010. Quantile and probability curves without crossing. Econometrica 78 (3), 1093–1125.
Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer Series in Statistics. Springer-Verlag London, Ltd., London.
Cooley, D., Nychka, D., Naveau, P., 2007. Bayesian spatial modeling of extreme precipitation return levels. J. Am. Stat. Assoc. 102 (479), 824–840.
Davison, A.C., Huser, R., Thibaud, E., 2013. Geostatistics of dependent and asymptotically independent extremes. Math. Geosci. 45 (5), 511–529.
Davison, A.C., Smith, R.L., 1990. Models for exceedances over high thresholds. J. Roy. Stat. Soc. B 52 (3), 393–442 (With discussion and a reply by the authors).
Eastoe, E.F., Tawn, J.A., 2009. Modelling non-stationary extremes with application to surface level ozone. J. Roy. Stat. Soc.: Series C (Applied Statistics) 58 (1), 25–45.
Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. In: Monographs on Statistics and Applied Probability, vol. 57. Chapman and Hall, New York.
Finkenstädt, B., Rootzén, H., 2003. Extreme Values in Finance, Telecommunications, and the Environment. Chapman and Hall/CRC.
Frigessi, A., Haug, O., Rue, H., 2002. A dynamic mixture model for unsupervised tail estimation without threshold selection, 2003 Extremes 5 (3), 219–235.
Gaetan, C., Grigoletto, M., 2007. A hierarchical model for the analysis of spatial rainfall extremes. J. Agric. Biol. Environ. Stat. 12 (4), 434–449.
Gelman, A., Meng, X.-L., 1998. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. Stat. Sci. 13 (2), 163–185.
Gilchrist, W., 2000. Statistical Modelling with Quantile Functions. Chapman and Hall/CRC, Boca Raton.
Gilleland, E., Katz, R., 2006, 18th Conference on Climate Variability and Change, 86th American Meteorological Society (AMS) Annual Meeting. Analyzing Seasonal to Interannual Extreme Weather and Climate Variability with the Extremes Toolkit, vol. 29.
Gilleland, E., Katz, R., 2016. extRemes 2.0: an extreme value analysis package in R. J. Stat. Software 72 (1), 1–39.
Gilleland, E., Katz, R.W., Naveau, P., 2017. Quantifying the risk of extreme events under climate change. Chance 30 (4), 30–36.
Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. J. Am. Stat. Assoc. 102 (477), 359–378.
Gneiting, T., Ranjan, R., 2011. Comparing density forecasts using threshold- and quantile-weighted scoring rules. J. Bus. Econ. Stat. 29 (3), 411–422.
Grotjahn, R., Black, R., Leung, R., Wehner, M.F., Barlow, M., Bosilovich, M., Gershunov, A., Gutowski, W.J., Gyakum, J.R., Katz, R.W., et al., 2016. North American extreme temperature events and related large scale meteorological patterns: a review of statistical methods, dynamics, modeling, and trends. Clim. Dynam. 46 (3), 1151–1184.
Gütschow, J., Jeffery, M., Gieseke, R., Gebel, R., Stevens, D., Krapp, M., Rocha, M., 2016. The PRIMAP-hist national historical emissions time series. Earth Syst. Sci. Data 8 (2), 571–603.
Hansen, J., Ruedy, R., Sato, M., Lo, K., 2010. Global surface temperature change. Rev. Geophys. 48 (4).
Haugen, M.A., Stein, M.L., Moyer, E.J., Sriver, R.L., 2018. Estimating changes in temperature distributions in a large ensemble of climate simulations using quantile regression. J. Clim. 31 (20), 8573–8588.
Haugen, M.A., Stein, M.L., Sriver, R.L., Moyer, E.J., 2019. Future climate emulations using quantile regressions on large ensembles. Adv. Stat.. Climatol.Meteorol. Oceanogr. 5 (1), 37–55.
He, X., 1997. Quantile curves without crossing. Am. Statistician 51 (2), 186–192.
Huang, W.K., Cooley, D.S., Ebert-Uphoff, I., Chen, C., Chatterjee, S., 2019a. New exploratory tools for extremal dependence: χ networks and annual extremal networks. J. Agric. Biol. Environ. Stat. 24 (3), 484–501.
Huang, W.K., Monahan, A.H., Zwiers, F., 2021. Estimating concurrent climate extremes: a conditional approach. Weather Clim.Extremes 33, 100332.
Huang, W.K., Nychka, D.W., Zhang, H., 2019b. Estimating precipitation extremes using the log-histospline. Environmetrics 30 (4), e2543.
Huang, W.K., Stein, M.L., McInerney, D.J., Sun, S., Moyer, E.J., 2016. Estimating changes in temperature extremes from millennial-scale climate simulations using generalized extreme value (GEV) distributions. Adv. Stat.. Climatol.Meteorol. Oceanogr. 2 (1), 79–103.
Huser, R., Genton, M.G., 2016. Non-stationary dependence structures for spatial extremes. J. Agric. Biol. Environ. Stat. 21 (3), 470–491.
Huser, R., Opitz, T., Thibaud, E., 2017. Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures. part A Spat. Stat. 21, 166–186.
Huser, R., Wadsworth, J.L., 2019. Modeling spatial processes with unknown extremal dependence class. J. Am. Stat. Assoc. 114 (525), 434–444.
Huser, R., Wadsworth, J.L., 2020. Advances in Statistical Modeling of Spatial Extremes. Wiley Interdisciplinary Reviews: Computational Statistics page e1537.
Huybers, P., McKinnon, K.A., Rhines, A., Tingley, M., 2014. US daily temperatures: the meaning of extremes in the context of nonnormality. J. Clim. 27 (19), 7368–7384.
IPCC, 2021. Climate Change 2021: the Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press (in press).
Katz, R.W., Brown, B.G., 1992. Extreme events in a changing climate: variability is more important than averages. Climatic Change 21 (3), 289–302.
Koenker, R., 2021. Quantreg: Quantile Regression. R package version 5.85.
Krupskii, P., Genton, M.G., 2021. Conditional normal extreme-value copulas. Extremes 1–29.
Lahiri, S.N., 2003. Resampling Methods for Dependent Data. Springer Series in Statistics. Springer-Verlag, New York.
Lee, D., Joe, H., 2018. Multivariate extreme value copulas with factor and tree dependence structures. Extremes 21 (1), 147–176.
Legates, D.R., Willmott, C.J., 1990. Mean seasonal and spatial variability in global surface air temperature. Theor. Appl. Climatol. 41 (1), 11–21.
McKinnon, K.A., Rhines, A., Tingley, M.P., Huybers, P., 2016. The changing shape of northern hemisphere summer temperature distributions. J. Geophys. Res. Atmos. 121 (15), 8849–8868.
McKinnon, K.A., Stine, A.R., Huybers, P., 2013. The spatial structure of the annual cycle in surface temperature: amplitude, phase, and Lagrangian history. J. Clim. 26 (20), 7852–7862.
Meehl, G.A., Tebaldi, C., Walton, G., Easterling, D., McDaniel, L., 2009. Relative increase of record high maximum temperatures compared to record low minimum temperatures in the US. Geophys. Res. Lett. 36 (23).
Mentaschi, L., Vousdoukas, M., Voukouvalas, E., Sartini, L., Feyen, L., Besio, G., Alfieri, L., 2016. The transformed-stationary approach: a generic and simplified methodology for non-stationary extreme value analysis. Hydrol. Earth Syst. Sci. 20 (9), 3527–3547.
Møller, J., Pettitt, A.N., Reeves, R., Berthelsen, K.K., 2006. An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. Biometrika 93 (2), 451–458.
Naveau, P., Huser, R., Ribereau, P., Hannart, A., 2016. Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. Water Resour. Res. 52 (4), 2753–2769.
NCEI, 2020. Global Surface Summary of Day (GSOD), National Centers for Environmental Information, National Oceanic and Atmospheric Administration. U.S. Department of Commerce. https://www7.ncdc.noaa.gov/CDO/GSOD_DESC.txt.
Nogaj, M., Parey, S., Dacunha-Castelle, D., 2007. Non-stationary extreme models and a climatic application. Nonlinear Process Geophys. 14 (3), 305–316.
Nogaj, M., Yiou, P., Parey, S., Malek, F., Naveau, P., 2006. Amplitude and frequency of temperature extremes over the North Atlantic region. Geophys. Res. Lett. 33 (10), L10801.
Paciorek, C.J., Stone, D.A., Wehner, M.F., 2018. Quantifying statistical uncertainty in the attribution of human influence on severe weather. Weather Clim.Extremes 20, 69–80.
Peel, M., Finlayson, B., Mcmahon, T., 2007. Updated world map of the Koppen-Geiger climate classification. Hydrol. Earth Syst. Sci. Discuss. 4.
Philip, S.Y., Kew, S.F., van Oldenborgh, G.J., Yang, W., Vecchi, G.A., Anslow, F.S., Li, S., Seneviratne, S.I., Luu, L.N., Arrighi, J., Singh, R., van Aalst, M., Hauser, M., Schumacher, D.L., Marghidan, C.P., Ebi, K.L., Bonnet, R., Vautard, R., Tradowsky, J., Coumou, D., Lehner, F., Wehner, M., Rodell, C., Stull, R., Howard, R., Gillett, N., Otto, F.E.L., 2021. Rapid Attribution Analysis of the Extraordinary Heatwave on the Pacific Coast of the US and Canada June 2021. https://www.worldweatherattribution.org/western-north-american-extreme-heat-virtually-impossible-without-human-caused-climate-change/.
Poppick, A., Moyer, E.J., Stein, M.L., 2017. Estimating trends in the global mean temperature record. Adv. Stat.. Climatol.Meteorol. Oceanogr. 3 (1), 33–53.
Rahmstorf, S., Coumou, D., 2011. Increase of extreme events in a warming world. Proc. Natl. Acad. Sci. Unit. States Am. 108 (44), 17905–17909.
Reiss, R.-D., Thomas, M., 2007. Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields, third ed. Birkhäuser Verlag, Basel. With 1 CD-ROM (Windows).
Revels, J., Lubin, M., Papamarkou, T., 2016. Forward-mode Automatic Differentiation in Julia. CoRR, abs/1607, 07892.
Rhines, A., McKinnon, K.A., Tingley, M.P., Huybers, P., 2017. Seasonally resolved distributional trends of North American temperatures show contraction of winter variability. J. Clim. 30 (3), 1139–1157.
Robin, Y., Ribes, A., 2020. Nonstationary extreme value analysis for event attribution combining climate models and observations. Adv. Stat.. Climatol.Meteorol. Oceanogr. 6 (2), 205–221.
Scarrott, C., MacDonald, A., 2012. A review of extreme value threshold estimation and uncertainty quantification. REVSTAT 10 (1), 33–60.
Schramm, P.J., Vaidyanathan, A., Radhakrishnan, L., Gates, A., Hartnett, K., Breysse, P., 2021. Heat-Related Emergency Department Visits during the Northwestern Heat Wave.
Seneviratne, S.I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Luca, A.D., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S.M., Wehner, M., Zhou, B., 2021. Weather and climate extreme events in a changing climate. In: Climate Change 2021: the Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press (in press).
Sheather, S.J., Jones, M.C., 1991. A reliable data-based bandwidth selection method for kernel density estimation. J. Roy. Stat. Soc. B 53 (3), 683–690.
Simpson, E.S., Wadsworth, J.L., 2021. Conditional modelling of spatio-temporal extremes for red sea surface temperatures. Spatial Statistics 41, 100482.
Sparks, A., Hengl, T., Nelson, A., 2017. GSODR: global summary daily weather data in R. J. Open Source Software 2.

Stein, M.L., 2020. Some statistical issues in climate science. Stat. Sci. 35 (1), 31–41.
Stein, M.L., 2021a. A parametric model for distributions with flexible behavior in both tails. Environmetrics 32 (2), 24. Paper No. e2658.
Stein, M.L., 2021b. Parametric models for distributions when interest is in extremes with an application to daily temperature. Extremes 24 (2), 293–323.
Taillardat, M., Fougères, A.-L., Naveau, P., de Fondeville, R., 2019. Extreme Events Evaluation Using CRPS Distributions.
Tarleton, L.F., Katz, R.W., 1995. Statistical explanation for trends in extreme summer temperatures at Phoenix, Arizona. J. Clim. 8 (6), 1704–1708.
Tencaliec, P., Favre, A.-C., Naveau, P., Prieur, C., Nicolet, G., 2020. Flexible semiparametric Generalized Pareto modeling of the entire range of rainfall amount. Environmetrics 31 (2), e2582.
Vandeskog, S.M., Thorarinsdottir, T.L., Steinsland, I., Lindgren, F., 2022. Quantile based modeling of diurnal temperature range with the five-parameter lambda distribution. Environmetrics, e2719.
Varin, C., Reid, N., Firth, D., 2011. An overview of composite likelihood methods. Stat. Sin. 21 (1), 5–42.
Wadsworth, J.L., Tawn, J., 2019. Higher-dimensional Spatial Extremes via Single-Site Conditioning. *arXiv preprint arXiv:1912.06560*.
Wang, S., 2013. Pbs: Periodic B Splines. R package version 1.1.
Wehner, M.F., 2020. Characterization of long period return values of extreme daily temperature and precipitation in the CMIP6 models: Part 2, projections of future change. Weather Clim.Extremes 30, 100284.
Yadav, R., Huser, R., Opitz, T., 2021. Spatial hierarchical modeling of threshold exceedances using rate mixtures. Environmetrics 32 (3), e2662.