Predictive Analytics for Extreme Events in Big Data

Saahil Shenoy
Department of Physics
Stanford University, Stanford, CA
saahils@stanford.edu

Dimitry Gorinevsky
Department of Electrical Engineering
Stanford University, Stanford, CA
gorin@stanford.edu

***Abstract*—This paper presents an efficient computational methodology for longitudinal and cross-sectional analysis of extreme event statistics in large data sets. The analyzed data are available across multiple time periods and multiple individuals in a population. Some of the periods and individuals might have no extreme events and some might have much data. The extreme events are modeled with a Pareto or exponential tail distribution. The proposed approach to longitudinal and cross-sectional analysis of the tail models is based on non-parametric Bayesian formulation. The maximum a posteriori probability problem leads to two convex problems for the tail parameters. Solving one problem yields the trends for the tail decay rate across the population and time periods. Solving another gives the trends of the tail quantile level. The approach is illustrated by providing analysis of 10- and 100-year extreme event risks for extreme climate events and for peak power loads in electrical utility data.**

I. INTRODUCTION

The predictive analytics are used to model the data. They can be also applied to finding outliers, the anomalous data that defy prediction. The outliers that represent extreme (peak) events can be used to study risks to cost, safety, and other critical performance indicators. This paper presents methodology for analyzing the risks of these extreme events across the time (longitudinal study) and the populations (cross-sectional study). In Big Data problems, there can be thousands of one in a million peak events. This affords detailed modeing of the extreme statistics. Such modeling is described in this paper. Our methodology is scalable to petabyte size data sets. The methodology is illustrated by two examples: in power grid data and in climate data.

The branch of statistics describing the rare peak events corresponding to the tails of the probability distributions is known as extreme value theory (EVT). The Fisher-Tippett-Gnedenko theorem of EVT establishes generalized extreme value (GEV) distribution as a common limit distribution. The GEV tails are Pareto (power law), exponential, or finite. The long tails can be estimated using peaks over threshold (POT) method by fitting the model to the threshold exceedance data, see [1].

There seems to be little prior work on multi-period estimation of the tail models. One ad hoc approach assumes that tail distribution parameters depend on time in accordance with a given regression model. In [2], this is a linear trend model. In [3], a more complex regression is used. The non-convex problems in [2], [3] are computationally difficult. The multi-period model in [4] uses a Bayesian prior for the exceedance number, but not for the tail parameter; the non-smooth prior complicates the computation. The multi-period formulation in [5] is for extreme value distributions with finite tail.

Most of the earlier work on cross sectional estimation of extreme events trends is related to climate data. A non-parametric model of spatial dependence is developed in [6]; it is assumed that data for neighboring points in a square grid topology are related. A spatial parametric model is considered in [7]. Both of these paper solve specific problems and do not formulate a scalable broadly applicable approach.

The contributions of this paper are as follows. First, it develops scalable methodology for longitudinal and cross-sectional modeling of extreme events deviating from a predictive model in Big Data. The methodology allows for scalable parallelizable processing. As an example, a 1 PB dataset is reduced to 2 MB of longitudinal and cross-sectional tail model data.

Second, this paper presents a novel non-parametric Bayesian formulation for the tail distribution parameters as they vary across the time and the population. It includes both the longitudinal and cross sectional model of these parameters in a single optimal estimation setting. This is accomplished by using Bayesian priors to filter the raw data noise and extract the trends.

Third, we transform the tail parameter estimation problem into two separate unconstrained convex optimization problems for the distribution quantile and for the distribution tail. Once brought into this form, these problems can be solved efficiently. The solution is scalable. The non-parametric nature of the formulation allows for flexible modeling.

Fourth, we apply our methodology to electrical power grid data and climate data. The results yield non-trivial conclusions about the trends of extreme events. For the power grid data, these trends are analyzed across the service zones of the utility. For the climate data, the year-to-year trends of extreme temperature are analyzed across the 12 months of the year.

II. Problem Formulation

We consider a data set that includes multiple time periods and multiple individuals in a population

$$D = \{\{\{x_{tia}, y_{tia}\}_{a=1}^{M_{ti}}\}_{i=1}^{N}\}_{t=1}^{T} \tag{1}$$

where scalars y_{tia} are dependent variables and $x_{tia} \in \Re^n$ are independent variables (regressors). In (1), t is the time period, i is the individual index inside the population, a is the sample number for a given individual and the time period, M_{ti} is the number of samples for a given individual in a given time period, N is the number of the individuals in the population, and T is the total number of time periods. We consider the impact that individuals within the population might have on each other. Though we assume that we are dealing with Big Data, there is a possibility that $M_{ti} = 0$, there are no data for a given individual in a given time period. The developed analytics should be able to handle that.

We assume that data (1) for time period t and population i are i.i.d. samples of an underlying conditional probability distribution. This distribution is a mixture of a normal distribution (the body), with probability $1-q_{t,i}$, and a tail distribution, with probability $q_{t,i}$,

$$y_{tia} = x_{tia}^T\beta_{ti} + (1 - z_{ti})\, v_{ti}^{(n)} + z_{ti} v_{ti}^{(e)}, \tag{2}$$

$$z_{ti} \sim B(1, q_{t,i}), \tag{3}$$

$$v_{ti}^{(n)} \sim N(0, \sigma_{ti}^2), \tag{4}$$

$$v_{ti}^{(e)} \sim p(\theta_{ti}|x_{ti}), \tag{5}$$

where $B(1, q_{t,i})$ is the binomial distribution with $\{0, 1\}$ outcomes, $\beta_{ti} \in \Re^n$ is the regression parameter vector, σ_{ti} is the standard deviation of the normal distribution. Distribution (2)–(5) describes both the normal (distribution body) and the tail behavior of the data.

According to the EVT, the tail probability density $p(\theta_{ti}|x_{ti})$ can be modeled as either an exponential or Pareto distribution. In what follows, we assume that $p(\theta_{ti}|x_{ti})$ comes from exponential distribution with the rate parameter $\theta_{t,i}$. For the Pareto distribution, $\log v_{ti}^{(e)}$ is exponentially distributed and the same analysis can be applied with a slight modification, see [8].

In what follows, we assume that the tail intensity $q_{t,i}$ is a small parameter, $q_{t,i} \ll 1$. For the small residuals

$$v_{tia} = y_{tia} - x_{tia}^T\beta_{ti}, \tag{6}$$

we approximately have $v_{tia} \sim N$ (the distribution body in (4)). For large v_{tia}, the exponent dominates the Gaussian and we approximately have $v_{tia} \sim Exp$ (the distribution tail in (5)). The approximations described below separately consider the distribution body and its tail. They are the basis of the proposed approach to estimating the regression parameters and modeling the distribution tail. These approximations can be considered as a special case (for $q_{t,i} \ll 1$) of the more general probabilistic model for a mixture of asymmetric Laplace and Gaussian (MALG) distributions, see [9].

The modeling of the tail for v_{tia} requires choosing a sufficiently large threshold Ω_{ti}. The tail risk model is defined as the probability of exceeding the threshold Ω_{ti} by the value u.

$$R_{tj}(u) = \mathbf{P}(v_{tj} - \Omega_{tj} \geq u|\theta_{t,j}, q_{t,j}) = q_{t,j} \cdot e^{-\theta_{t,j}u}. \tag{7}$$

This paper presents a method for analyzing the data (1) to estimate models for the tail risk (7). We consider combined longitudinal modeling (the dependence on the time period t) and cross sectional modeling (the dependence on the individual j in the population). The end goal is to model the risks of extreme events, such as 1-in-10 years events. This is a non-trivial problem since such events do not happen every time period to every individual in the population. The paper presents the method for solving this problem for very large datasets.

Figure 1 is a flowchart of the model estimation and risk evaluation steps that outlines the proposed methodology. Sections III and IV explain each of these steps in some detail. The approach is scalable to very large datasets since each step is highly scalable, parallelizable, and provides substantial data reduction. For illustrative purposes consider a data set of 1 PB size. This would roughly correspond to $T = 100$ time periods, $N = 1000$ individuals in the population, $M_{tj} = 10^7$ samples in each time period for each individual, and $n = 100$ regressors x_{tia}. We assume that all non-zero numbers are represented by a double precision type. Figure 1 shows how the data is reduced during the processing.

III. Model Formulation

A. Robust Regression

In this section we formulate how to estimate the model of the distribution body (2), (4). We define $\mathcal{C}_{ti}$

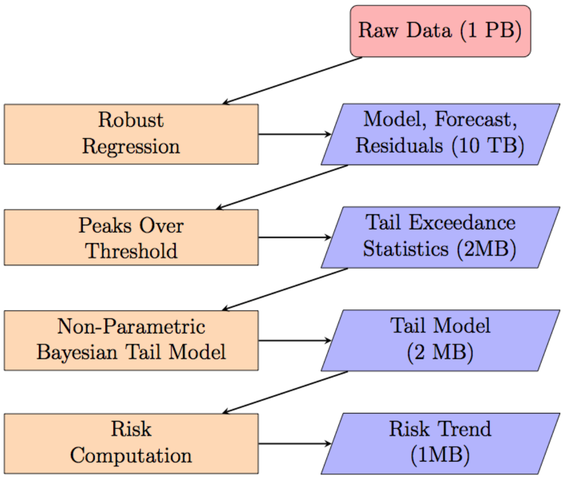

Fig. 1. Flow chart of model building logic and data reduction at each step.

as the set of indexes of the points that belong to the distribution body for time period t and population i. For the distribution body $z_{ti} = 0$, and (2), (4) yield the distribution $y_{ti} \sim \mathcal{N}(x_{ti}^T\beta_{ti}, \sigma_{ti}^2)$. The well-known maximum likelihood estimation (MLE) estimates of the model parameters $\hat{\beta}_{ti}$ and $\hat{\sigma}_{ti}$ are

$$\begin{aligned} \hat{\beta}_{ti} &= \arg \min_{\beta_{ti}} \|y_{\mathcal{C}_{ti}} - X_{\mathcal{C}_{ti}}\beta_{ti}\|_2^2 \\ &= \left(X_{\mathcal{C}_{ti}}^T X_{\mathcal{C}_{ti}}\right)^{-1} X_{\mathcal{C}_{ti}}^T y_{\mathcal{C}_{ti}}, \end{aligned} \tag{8}$$

$$\hat{\sigma}_{ti} = \text{card}(\mathcal{C}_{ti})^{-1/2}\|y_{\mathcal{C}_{ti}} - X_{\mathcal{C}_{ti}}\hat{\beta}_{ti}\|_2, \tag{9}$$

where $\text{card}(\cdot)$ is the cardinality (size) of the data set, $y_{\mathcal{C}_{ti}} = \{y_{tia} | a \in \mathcal{C}_{ti}\}$, and $X_{\mathcal{C}_{ti}}$ has rows corresponding to x_{tia}^T, $a \in \mathcal{C}_{tj}$. The estimation of β_{ti} (8) is known as a robust regression problem, see [10], [11].

The solution (8) is quite scalable when there are many data points, i.e., when $\text{card}(\mathcal{C}_{ti})$ is very large). To compute the scatter matrix $X_{\mathcal{C}_{ti}}^T X_{\mathcal{C}_{ti}}$ when the data matrix $X_{\mathcal{C}_{ti}}$ cannot fit into memory, one can use the following representation

$$X_{\mathcal{C}_{ti}}^T X_{\mathcal{C}_{ti}} = \sum_{a \in \mathcal{C}_{ti}} x_{tia} x_{tia}^T, \tag{10}$$

where each outer product $x_{tia}x_{tia}^T$ is a $n \times n$ matrix. For a moderately large number of regressors n, this matrix fits into the memory. For a large data set, the outer product terms can be added up using standard MapReduce techniques to yield $X_{\mathcal{C}_{ti}}^T X_{\mathcal{C}_{ti}}$. Since the scatter matrix $X_{\mathcal{C}_{ti}}^T X_{\mathcal{C}_{ti}}$ fits into the memory, the standard numerical techniques for matrix inversion can be used.

The product $X_{\mathcal{C}_{ti}}^T y_{\mathcal{C}_{ti}}$ in (8) is a vector in $\Re^n$. It can be represented as

$$X_{\mathcal{C}_{ti}}^T y_{\mathcal{C}_{ti}} = \sum_{a \in \mathcal{C}_{ti}} x_{tia} y_{tia}, \tag{11}$$

where each product $x_{tia}y_{tia}$ is a $n \times 1$ vector. For a large data set, these vectors can be added up using standard MapReduce techniques, see [12].

B. Model, Forecast, and Residuals

Once the scatter matrix (10) and vector (11) have been computed, one can solve for the regression model $\hat{\beta}_{ti}$ (8) in memory. The next step is to compute the predicted values $\hat{y}_{ti} \in \Re^{M_{ti}}$. The components of the vector $\hat{y}_{ti}$ are a result of scalable element-wise computation

$$\hat{y}_{tia} = x_{tia}^T \hat{\beta}_{ti}. \tag{12}$$

The above computations assume the knowledge of the set $\mathcal{C}_{ti}$. This is the set of the distribution body indexes, the set of indexes where $z_{ti} = 0$ in (2). Though the realization of z_{ti} is hidden, the set $\mathcal{C}_{ti}$ can be estimated from the data (1) and forecast (12).

$$\mathcal{C}_{ti} = \{a \mid A_{ti} > |v_{tia}|\}, \tag{13}$$

$$v_{tia} = y_{tia} - \hat{y}_{tia}, \tag{14}$$

where v_{tia} are the residuals (forecast errors) and A_{ti} is a threshold parameter, The selection of A_{ti} is discussed later on.

The standard deviation $\hat{\sigma}_{ti}$ (9) can be computed as $\hat{\sigma}_{ti}^2 = \text{card}(\mathcal{C}_{ti})^{-1}\|\hat{v}_{\mathcal{C}_{ti}}\|_2^2$. The scalable computation method sums up the squared components of $\hat{v}_{\mathcal{C}_{ti}} = \{v_{tia} | a \in \mathcal{C}_{ti}\}$.

Consider the vector of residuals $v_{ti} \in \Re^{M_{ti}}$ with elements v_{tia}. Given the forecast $\hat{y}_t$ (12), this vector can be computed in accordance (14) element-wise. In our 1 PB data sets example, the vector v_{ti} takes 10 TB.

The Robust Regression method uses the residuals v_{tia} to set the threshold A_{ti} as a multiple of the standard deviation $\hat{\sigma}_{ti}$. Given A_{ti}, the set $\mathcal{C}_{ti}$ can be computed from (13). We then iteratively re-compute (8)–(12) with the new $\mathcal{C}_{ti}$. The iterations have converged when $\mathcal{C}_{ti}$ stops changing, see [13]. We initialize $\mathcal{C}_{ti}$ to include the indexes of all M_{ti} data points. This is equivalent to initializing $A_{ti} = \infty$.

C. Peaks Over Threshold

Peaks over threshold (POT) is a standard EVT method used to model the outliers, see [1]. Consider exceedances $e_{ti} \in \Re^{M_{ti}}$ for time period t and population i as

$$e_{ti} = \{v_{tia} - \Omega_{ti}\}_{a=1}^{M_{ti}}, \tag{15}$$

where Ω_{ti} is a threshold that is dependent on the empirical quantile.

Let us define set $\mathcal{T}_{ti}$ to be the set of indexes of points in the tail distribution (5). In the context of our problem, the POT exceedances are just the set of non-negative exceedances e_{tia}, which are in set $\mathcal{T}_{ti}$. We then define $\mathcal{T}_{ti}$ as

$$\mathcal{T}_{ti} = \{a |\ e_{tia} \geq 0\}. \tag{16}$$

The POT exceedances can be computed in a scalable, parallelizable way since $\mathcal{T}_{ti}$ is determined by independently considering each exceedance e_{tia}. If we assume that 1% of data points are POT exceedances, this reduces the data we are processing from 10 TB to 100 GB.

We are interested in estimating the risk R_{ti} (7) in the regime of large residuals v_{tia}, which correspond to large exceedances e_{tia}. We can rewrite the problem in terms of e_{ti}.

$$\begin{aligned} R_{ti}(u) &= \mathbf{P}(e_{ti} \geq u) \\ &= \mathbf{P}(e_{ti} \geq u | e_{ti} \geq 0) \cdot \mathbf{P}(e_{ti} \geq 0), \end{aligned} \tag{17}$$

where the conditional probability rule is used for the second line, see [14]. In accordance with (2),

$$q_{t,i} = \mathbf{P}(e_{ti} \geq 0). \tag{18}$$

In Subsection III-E, we estimate $q_{t,i}$ and the tail probability density function $p_{ti}(u)$, which is the derivative of the cumulative density function, $1 - \mathbf{P}(e_{ti} \geq u | e_{tj} \geq 0)$.

D. Tail Exceedance Statistics

Consider the POT data set $\mathcal{T}_{ti}$. In accordance with (7), the first multiplier in (17) follows the exponential distribution,

$$e_{ti} |\, (e_{ti} \geq 0, \theta_{t,i}) \sim Exp(\theta_{ti}). \tag{19}$$

This exponential distribution MLE parameter estimate is characterized by the zero and first order statistics of the exceedance data. More concretely, we just need to compute the number of non-negative exceedances n_{ti} and mean exceedances $\bar{e}_{ti}$ for each time period, which are given as follows

$$n_{ti} = \text{card}\{\mathcal{T}_{ti}\}, \tag{20}$$

$$\bar{e}_{ti} = \text{mean}_{a \in \mathcal{T}_{ti}}\{e_{tia}\}. \tag{21}$$

Finding the mean and the number of non-negative exceedances is a parallelizable process since each component of the exceedance vector can be processed separately. At this point, the problem reduces down to keeping $2 \cdot T \cdot N$ double precision numbers $\{\{n_{tj}, \bar{e}_{tj}\}_{j=1}^{N}\}_{t=1}^{T}$ for the tail exceedance statistics. In the example, the amount of data being processed is reduced from 100 GB to 2 MB.

The second multiplier in (17) is the quantile level parameter $q_{t,i}$ (18). We assume that each data point is an independent sample of the distribution. The point belongs to the tail with probability $q_{t,i}$ or not with the probability $1 - q_{t,i}$. Then, the number of the tail points n_{ti} (20) follows the Binomial distribution

$$n_{ti} | q_{t,i} \sim B(M_{ti}, q_{t,i}). \tag{22}$$

The tail exceedance statistics (20) and (21) are sufficient for characterizing the distribution (17).

E. Non-Parametric Bayesian Tail Model

To compute the risk R_{ti} (17), we must first estimate tail parameters $q_{t,i}$ (3) and $\theta_{t,i}$ (5). We do this using the distribution models (19) and (22).

1) Likelihood: In our longitudinal and cross-sectional analysis setting, the right tails for different time periods t and individuals i in the population are defined by $\{q_{t,j}, \theta_{t,j}\}_{j=1,t=1}^{N,T}$. We use maximum likelihood estimation (MLE) of these parameters, see [8]. Maximizing the log likelihoods for the two tail parameter distributions (19) and (22) yields

$$\hat{\theta}_{t,i} = \arg \max_{\theta_{t,i} \geq 0} \sum_{t=1}^{T} \sum_{j=1}^{N} n_{tj} \left(\log \theta_{t,j} - \theta_{t,j} \bar{e}_{tj} \right), \tag{23}$$

$$\begin{aligned} \hat{q}_{t,i} = \arg \max_{0 \leq q_{t,i} \leq 1} \sum_{t=1}^{T} \sum_{j=1}^{N} & [n_{tj} \log q_{t,j} \\ & + (M_{tj} - n_{tj}) \log(1 - q_{t,j})] . \end{aligned} \tag{24}$$

We make the following non-linear variable change

$$r_{t,i} = \psi(\theta_{t,i}) = \log \theta_{t,i}, \tag{25}$$

$$w_{t,i} = \phi(q_{t,i}) = \log(-\log q_{t,i}). \tag{26}$$

The MLE optimization problems can be compactly written as

$$\underset{r_{t,j}}{\text{maximize}} \sum_{t=1}^{T} \sum_{j=1}^{N} n_{tj} \Psi(r_{t,j}; \bar{e}_{tj}), \tag{27}$$

$$\Psi(x; \gamma) = x - \gamma e^{x}, \tag{28}$$

$$\underset{w_{t,j}}{\text{maximize}} \sum_{t=1}^{T} \sum_{j=1}^{N} n_{tj} \Phi(w_{t,j}; M_{tj}/n_{tj} - 1), \tag{29}$$

$$\Phi(x; \gamma) = -e^{x} + \gamma \log \left(1 - \exp\left(-e^{x}\right)\right). \tag{30}$$

The unconstrained convex optimization problems (27), (29) split into independent optimization problems for each individual i at one time period t. The optimal estimates can be obtained by differentiating each term in the sum (23) or (24) with respect to $\theta_{t,i}$ or $q_{t,i}\}$ and solving $\Psi'(r_{t,j}; \bar{e}_{tj}) = 0$, $\Phi'(w_{t,j}; M_{tj}/n_{tj} - 1) = 0$. This gives the MLE estimates

$$\hat{r}_{t,i} = -\log\left(\bar{e}_{ti}\right), \tag{31}$$

$$\hat{w}_{t,i} = \log\log(M_{ti}/n_{ti}). \tag{32}$$

F. Tail Model

Using (25), (26) to get back to the original coordinates, the estimated parameters (31), (32) are

$$\hat{\theta}_{t,i} = 1/\bar{e}_{ti}, \tag{33}$$

$$\hat{q}_{t,i} = n_{ti}/M_{ti}. \tag{34}$$

In our example, the tail exceedance statistic is 2 MB of the data. The tail model (33), (34) has the same size.

G. Risk Computation

Once we have our tail model, we compute our risk exceedance probability $R_{ti}(u_*)$ (7), with threshold u_*. We choose a threshold u_* such that the average exceedance probability is R_*. We start by aggregating all the data (1) into one time period and one individual in the population. We estimate q_* and θ_* of the POT exceedances (16). This allows to compute u_* for a given average exceedance probability R_* by solving

$$R_* = q_* \cdot e^{-\theta_* u_*}, \tag{35}$$

$$u_* = \theta_*^{-1} \log\left(q_*/R_*\right). \tag{36}$$

H. Risk Trend

The risk trend is the set of risk exceedance probabilities $\{\{R_{ti}(u_*)\}_{i=1}^{N}\}_{t=1}^{T}$ (7) for the threshold u_* (36). In our example, this data is 1 MB in size. The risk trend is related to R_* (35) as follows

$$\frac{1}{N \cdot T} \sum_{t=1}^{T} \sum_{i=1}^{N} R_{ti}(u_*) = R_*. \tag{37}$$

This is just the average risk probability over all time periods and individuals in the population.

IV. Optimal Bayesian Formulation

The MLE estimates described in Section III-E are very noisy and might change drastically between different time periods and individuals. Some of the time periods and individuals might not have any exceedance data at all. More reasonable smoothed estimates can be obtained by solving a Bayesian maximum a posteriori (MAP) estimation problem. We will formulate the MAP problem that includes non-parametric priors for the tail that enforces the smoothness.

A. Prior Distributions

We assume the priors for $\theta \equiv \{\theta_{t,i}\}$ and $q \equiv \{q_{t,i}\}$ that have the form

$$(\Delta\psi)_k(\theta) = \psi(\theta_{t_k,i_k}) - \psi(\theta_{t'_k,i'_k}) \sim \psi(\chi_k), \tag{38}$$

$$(\Delta\phi)_k(q) = \phi(q_{t_k,i_k}) - \phi(q_{t'_k,i'_k}) \sim \phi(\xi_k). \tag{39}$$

where $\psi(\cdot)$ is given by (25), $\phi(\cdot)$ is given by (26), The priors (38) and (39) relate the tail rate parameters θ and tail quantile parameters q at two indexes sets $\{t_k, i_k\}$ and $\{t'_k, i'_k\}$. They express the prior belief that the tail parameters for the individual i_k at time instance t_k are related, close, to the tail for the individual i'_k at time instance t'_k. The prior distributions in (38) and (39) are assumed to be given by $\chi_k \sim$ Gamma and $\xi_k \sim$ Beta, which are the conjugate prior distributions for (19) and (22) respectively, see [15].

We consider a set of N_p priors (38), (39),

$$\{(\Delta\psi)_k(\theta) \sim \phi(\chi_k),\ (\Delta\phi)_k(q) \sim \phi(\xi_k)\}_{k=1}^{N_p}. \tag{40}$$

Section V will specify the prior structure in more detail for particular applications. The technical aspects of the parameter estimate computations for $\theta_{t,i}$ and $q_{t,i}$ in this formulation are discussed in more detail in Subsections IV-B and IV-C below.

The parameters for the prior distributions in (40) are chosen as follows. For all distributions χ_k in (40) we assume $\chi_k \sim \text{Gamma}(\alpha_k + 1, \alpha_k)$, where α_k is the k-th prior distribution strength parameter. With the prior Gamma distribution parameters $(\alpha_k + 1, \alpha_k)$, the mode of $\psi(\chi_k)$ (22) is zero for $\alpha_k > 0$. The intuition is that the mode of $(\Delta\psi)_k(\theta)$ being zero means that in the absence of POT exceedance data in (22) the tail rate estimate for time period t_k and individual i_k is exactly equal to that for the time period t'_k and individual i'_k. For the similar reason, all distributions ξ_k in (40) are assumed to have $\xi_k \sim \text{Beta}(\eta_k, \eta_k(e-1)+1)$ in the conjugate Beta prior distribution for (39). The Beta distribution with the parameters $(\eta_k, \eta_k(e-1)+1)$ means $\phi(\xi_k)$ (39) has zero mode. In the absence of POT exceedance data in (19) the quantile parameter estimates for time period t_k and individual i_k is equal to that for the time period t'_k and individual i'_k.

The logarithm of the probability density function for $\text{Gamma}(\alpha+1, \alpha)$ has the form $\alpha\Psi(x)$, where Ψ is given by (28). The logarithm of the probability density function for $\text{Beta}(\eta, \eta(e-1)+1)$ has the form $\eta\Phi(x)$, where Φ is given by (30). Under these assumptions, the log priors for (40) can be be expressed as $\alpha_k\Psi\left((\Delta\psi)_k(\theta); 1\right)$ and $\eta_k\Phi\left((\Delta\phi)_k(q); e-1\right)$.

B. Posterior Maximization

Combining the formulated log likelihoods with the log priors for the parameters of the distribution yields the posterior to be maximized in the MAP formulation. For $\theta_{t,j}$ we get the following MAP problem

$$\begin{aligned} \underset{\theta_{t,j}\geq 0}{\text{maximize}} \sum_{t=1}^{T}\sum_{j=1}^{N} n_{tj}\Psi\left(\log\theta_{t,j};\bar{e}_{tj}\right) & \\ +\sum_{k=1}^{N_p}\alpha_k\Psi\left((\Delta\psi)_k(\theta);1\right). & \end{aligned} \tag{41}$$

The parameters α_k define the 'tightness' of the prior. For $\alpha_k = 0$, there is no prior and we get the MLE problem. For $\alpha_k = \infty$, the estimates are forced to be the same for (t, j) pairs in (38), (39). For $0 < \alpha_k < \infty$, the estimate is smoothed, with more smoothing for larger α_k.

For $q_{t,j}$ the MAP problem is

$$\begin{aligned} \underset{0\leq q_{t,j}\leq 1}{\text{maximize}} \sum_{t=1}^{T}\sum_{j=1}^{N} n_{tj}\Phi\left(\phi(q_{t,j});M_{tj}/n_{tj}-1\right) & \\ +\sum_{k=1}^{N_p}\eta_k\Phi\left((\Delta\phi)_k(q);e-1\right). & \end{aligned} \tag{42}$$

The prior parameters η_k play the same role as α_k in (41).

Using the non-linear variable change (25), the constrained non-convex optimization problem (41), is transformed into the problems

$$\begin{aligned} \underset{r_{t,j}}{\text{maximize}} \sum_{t=1}^{T}\sum_{j=1}^{N} n_{tj}\Psi\left(r_{t,j};\bar{e}_{tj}\right) & \\ +\sum_{k=1}^{N_p}\alpha_k\Psi\left((\Delta r)_k;1\right), & \end{aligned} \tag{43}$$

where in accordance with (38),

$$(\Delta r)_k = r_{t_k,i_k} - r_{t'_k,i'_k}. \tag{44}$$

Similarly, after the variable change (26), the constrained problem (42) becomes

$$\begin{aligned} \underset{w_{t,j}}{\text{maximize}} \sum_{t=1}^{T}\sum_{j=1}^{N} n_{tj}\Phi\left(w_{t,j};M_{tj}/n_{tj}-1\right) & \\ +\sum_{k=1}^{N_p}\eta_k\Phi\left((\Delta w)_k;e-1\right). & \end{aligned} \tag{45}$$

where in accordance with (39) and similar to (44), we have $(\Delta w)_k = w_{t_k,i_k} - w_{t'_k,i'_k}$. The unconstrained convex optimization problems (43) and (45) can be solved efficiently and reliably. The solution is discussed below.

C. Optimization Solution

The unconstrained convex problems (43), (45) can be solved using the Newton's iterations, see [16].

1) Quadratic Approximation: Both optimization problems (43) and (45) maximize the index of the form

$$L(\xi) = \sum_{t=1}^{T}\sum_{j=1}^{N} n_{tj} f(\xi_{t,j};\rho_{tj}) + \sum_{k=1}^{N_p}\lambda_k f((\Delta\xi)_k;\gamma), \tag{46}$$

where $\xi_{t,j}$ are the decision variables and $(\Delta\xi)_k$ is defined by (44). The parameters ρ_{tj}, λ_k, γ and the function $f(\cdot;\cdot)$ in (46) can be related back to problems (43) and (45). Expanding (46) into a quadratic Taylor series about $\xi^*_{t,j}$, with $\xi_{t,j} = \xi^*_{t,j} + \epsilon_{t,j}$ yields

$$\begin{aligned} L(\xi) \approx & \sum_{t=1}^{T}\sum_{j=1}^{N} n_{tj}\left(f'(\xi^*_{t,j};\rho_{tj})\epsilon_{t,j} + f''(\xi^*_{t,j};\rho_{tj})\frac{\epsilon^2_{t,j}}{2}\right) \\ & +\sum_{k=1}^{N_p}\lambda_k\left(f'((\Delta\xi^*)_k;\gamma)\cdot(\Delta\epsilon)_k\right. \\ & \left. + f''((\Delta\xi^*)_k;\gamma)\cdot\frac{(\Delta\epsilon)^2_k}{2}\right) + \text{const}, \end{aligned} \tag{47}$$

where f' and f'' are the first and the second derivatives of function f with respect to its first argument.

2) Newton Iteration: We can think of $\epsilon_{t,j}$ as the (t, j) element of a matrix. Let ϵ be the vectorization of this matrix. We can then write (47) as

$$\text{minimize} \sum_{m=0}^{1}(F_m\epsilon - d_m)^T B_m(F_m\epsilon - d_m), \tag{48}$$

$$\epsilon = \text{vec}\left\{\epsilon_{t,j}\right\}_{t=1,j=1}^{T,N}, \tag{49}$$

$$B_0 = \text{diag}\left(\text{vec}\left\{n_{tj}f''(\xi^*_{t,j};\rho_{tj})\right\}_{t=1,j=1}^{T,N}\right), \tag{50}$$

$$B_1 = \text{diag}\left(\text{vec}\left\{\lambda_k f''((\Delta\xi^*)_k;\gamma)\right\}_{k=1}^{N_p}\right), \tag{51}$$

$$d_0 = \text{vec}\left\{-\frac{f'(\xi^*_{t,j};\rho_{tj})}{f''(\xi^*_{t,j};\rho_{tj})}\right\}_{t=1,j=1}^{T,N}, \tag{52}$$

$$d_1 = \text{vec}\left\{-\frac{f'((\Delta\xi^*)_k;\gamma)}{f''((\Delta\xi^*)_k;\gamma)}\right\}_{k=1}^{N_p}, \tag{53}$$

where vec$\{\cdot\}$ denotes the vectorization operation. In (48), F_0 is the identity matrix. Matrix F_1 is the sparse 'difference' matrix that maps the vectorized matrix (49) into the vector $(\Delta\epsilon)_k$ in accordance with (44). Matrices B_0, B_1 and vectors d_0, d_1 depend on the expansion center $\xi^* \equiv \{\xi^*_{t,j}\}_{t=1,j=1}^{T,N}$.

Differentiating the optimization index (48) with respect to vector ϵ yields a system of linear equations.

Solving this system for ϵ gives

$$\epsilon = H^{-1} \sum_{m=0}^{1} F_m^T B_m d_m, \tag{54}$$

$$H = \sum_{m=0}^{1} F_m^T B_m F_m, \tag{55}$$

where H is the Hessian of the optimization index (48). The Hessian is extremely sparse matrix, which makes its inversion computationally feasible even for large problems. The quantity being multiplied by the inverse Hessian is the gradient of the quadratic objective. Since F_m and B_m are all sparse matrices, this multiplication can be done efficiently for large problems.

The Newton's method iterations go as follows. Let $\xi_{t,j}^{(i)}$ be the approximate solution at iteration i. We compute matrices (50)–(53) using $\xi_{t,j}^{*} = \xi_{t,j}^{(i)}$ as the approximation center. Then, the Newton's step $\epsilon_{t,j}^{(i)}$ is computed from (54). It is used to get the next iteration of the approximate solution as

$$\xi_{t,i}^{(i+1)} \leftarrow \xi_{t,i}^{(i)} + \epsilon_{t,j}^{(i)}. \tag{56}$$

The iterations continue until convergence is achieved. Since the problem is convex and smooth, the Newton's method iterations are guaranteed to converge. In the examples, they converge very fast.

A rough upper bound on the number of non-zero elements of the Hessian that needs to be inverted is $10 \cdot T(N+1)$. In our example, the tail exceedance statistics are 2 MB of data. As a sparse matrix, the Hessian takes approximately 10 MB with $T = 100$ and $N = 10^3$. This matrix can fit into memory and thus be inverted. The gradient will consist of an array of $T(N+1)$ double precision values, which corresponds to 1 MB.

3) Quadratic Approximation Coefficients: For the optimization problem (43), the matrices (50)–(53) an be obtained by setting the function f and parameters ρ_{tj}, λ_k, γ as follows

$$\begin{aligned} f(x;\nu) = -\Psi(x;\nu) = \nu e^x - x, \\ \rho_{tj} = \bar{e}_{tj}, \ \ \gamma = 1, \ \ \lambda_k = \alpha_k. \end{aligned} \tag{57}$$

The minus sign comes from going from a maximization problem to a minimization problem in the Newton's iterations. The first and the second derivative of $\Psi(x,\nu)$ with respect to x are given by

$$\Psi'(x;\nu) = 1 - \nu e^x, \tag{58}$$

$$\Psi''(x;\nu) = -\nu e^x. \tag{59}$$

For problem (45)

$$\begin{aligned} f(x;\nu) = -\Phi(x;\nu) = e^x - \nu \log\left(1 - \exp\left(-e^x\right)\right), \\ \rho_{tj} = M_{tj}/n_{tj} - 1, \ \ \gamma = e - 1, \ \ \lambda_k = \eta_k. \end{aligned} \tag{60}$$

The first and the second derivative of $\Phi(x,\nu)$ with respect to x are given by

$$\Phi'(x;\nu) = e^x \left(\frac{\nu}{\exp\left(e^x\right) - 1} - 1 \right), \tag{61}$$

$$\Phi''(x;\nu) = e^x \left(\nu \frac{\exp(e^x)\left(1 - e^x\right) - 1}{\left(\exp(e^x) - 1\right)^2} - 1 \right). \tag{62}$$

V. Applications

We apply the formulated methodology to electric utility data and to a global temperature dataset.

A. Power Load Peak Demand

The data set from [17] includes hourly electrical power loads and ambient temperature data for a US utility. The loads are in the range of 1 kW to 500 MW. This data set is 16 MB in size. The data covers a time range of 48 consecutive months with the sampling interval of one hour, 33,740 samples in all, and 20 different service zones of the utility. We considered each month as a time period and each zone as an individual in the population, $T = 48$ periods and $N = 20$ zones in all. The computations for the proposed analytics, performed on single processor, take under 1 sec.

The power load has strong seasonal component. Therefore, the chosen regressors x_{tia} (1) are the binary vectors that indicate the calendar month for the current time period. More specifically, we take

$$x_{tia}^T = \left[\begin{array}{cccc} m_0(t) & m_1(t) & \cdots & m_{11}(t) \end{array} \right], \tag{63}$$

$$m_j(t) = \mathbb{1}\left(\mathrm{mod}(t,12) = j\right), \tag{64}$$

where $\mathbb{1}(\cdot)$ is the indicator function. More detailed regression models of the electric power load are considered in [13], [18].

Let L_{tia} be the a-th sample of electric power load (in GW) for month t and zone i. We choose the dependent variable $y_{tia} = L_{tia}$. We choose $A_{ti} = 3 \cdot \hat{\sigma}_{ti}$, where $\hat{\sigma}_{ti}$ is estimated from (9). Figure 2 shows the load profile and forecast for the entire month of August 2004 in zone 8.

We choose the POT threshold Ω_{ti} in (15) based on the empirical quantile of the data for each individual aggregated across each time period. The threshold Ω_{ti} is picked such that only 12 data points for each individual exceed the threshold. This approximately corresponds to the quantile level of 99.98%.

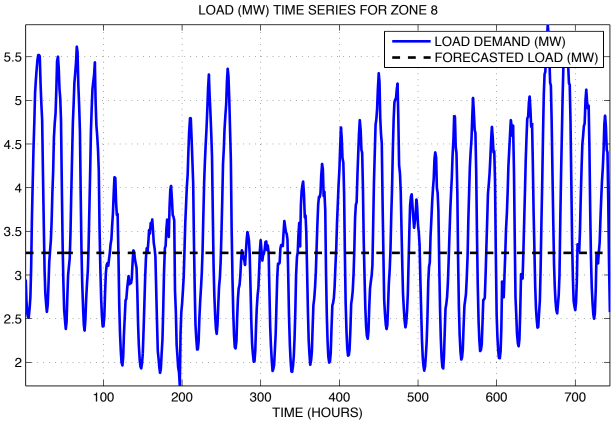

Fig. 2. Load and forecast (in MW) for a August 2004 in zone 8.

We choose the two following prior structures (38)–(39). For the tail rate $\theta_{t,j}$

$$\text{vec}\left\{(\Delta\psi)_k(\theta)\right\} = \text{col}(v', v''), \tag{65}$$

$$v' = \text{vec}\left\{\psi(\theta_{t,j}) - \psi(\theta_{t-1,j})\right\}_{t=2,j=1}^{T,N}, \tag{66}$$

$$v'' = \text{vec}\left\{\psi(\theta_{t,j}) - \psi(\theta_{t,0})\right\}_{t=1,j=1}^{T,N}. \tag{67}$$

For the tail quantile parameter $q_{t,j}$

$$\text{vec}\left\{(\Delta\phi)_k(q)\right\} = \text{col}(w', w''), \tag{68}$$

$$w' = \text{vec}\left\{\phi(q_{t,j}) - \phi(q_{t-1,j})\right\}_{t=2,j=1}^{T,N}, \tag{69}$$

$$w'' = \text{vec}\left\{\phi(q_{t,j}) - \phi(q_{t,0})\right\}_{t=1,j=1}^{T,N}. \tag{70}$$

The priors (66), (69) relate the parameters at two sequential time periods for the same zone and express that they are close. We assume that the priors (66) have strength α' and the priors (67) have strength α''. The priors (67), (70) relate the parameters for all zones at a given time period and express that that they are close to a common (average) value for all the zones. The priors (69) have strength η' and the priors (70) have strength η''. Table I summarizes the parameters $\alpha', \alpha'', \eta', \eta''$ used in the numerical examples.

This prior structure is captured through matrix F_1 in (48). This matrix is given by

$$F_1 = \left[\begin{array}{c|c} \mathbf{0}_{(T-1)N\times T} & I_N \otimes D_T \\ \hline -\mathbf{1}_N \otimes I_T & I_{TN} \end{array}\right], \tag{71}$$

where $\otimes$ is the Kronecker product, $\mathbf{0}_{K\times L} \in \Re^{K\times L}$ is a matrix of zeros, $D_T \in \Re^{(T-1)\times T}$ is the two-diagonal first difference matrix with -1 on the main diagonal (except the last, zero, entry) and 1 above the main diagonal, $I_K \in \Re^{K\times K}$ is the identity matrix, and $\mathbf{1}_L \in \Re^L$ is a vector of ones.

After solving the convex MAP problems (43) and (45), we get the estimates for $\theta_{t,i}$ and $q_{t,i}$ for each month t and zone i. These allow to estimate the risk $P_{ti}(u_*)$ at at threshold u_* (36) computed as

TABLE I

Prior Strength Parameters in Two Examples

	α'	α''	η'	η''
Power Load	50	5	5000	500
Extreme Weather	500	50	5000	50

$$P_{ti}(u) = 1 - \left(1 - R_{ti}(u)\right)^{720}. \tag{72}$$

The data is sampled at an hourly rate and $R_{ti}(u_*)$ is the hourly probability of exceeding threshold u_*. With 720 hours in a month, $P_{ti}(u_*)$ is the monthly probability of u_* exceeded in at least one hour. We pick u_* that yields an 1-in-10 years event, with $R_* = 0.1/8760$ (35), for the pooled data (all months and zones).

Figure 3 shows risk $P_{ti}(u_*)$ computated using the baseline MLE estimates of $\theta_{t,i}$ and $q_{t,i}$, i.e., without the priors. Figure 4 shows risk $P_{ti}(u_*)$ computated for the MAP (smoothed) estimates of $\theta_{t,i}$ and $q_{t,i}$. These smoothed estimates are much more informative than the baseline MLE estimates. In this example, the computational time of all processing steps in Figure 1, performed on single processor is just under 1 sec.

The risk exceedance probability is approximately 0.8% on a monthly average, which roughly translates to 10% on a yearly average. Figure 4 shows that that there is an upward trend in the probability of a 1-in-10 years event across all zones. This upward trend can be attributed by the increasing variability of the grid load because of the introduction of distributed energy resources, such as solar and increasing use of the wind power. Another possible explanation is related to increasing trend of extreme weather events, which is discussed in the next subsection.

B. Extreme Weather in Changing Climate

The temperature data set from [19] contains monthly global temperature readings. Temperature measurements range from -73.78°C to 41.75°C. We use this data for a period from 1968 to 2012. The spatial resolution of this data set is 2.5° latitude by 2.5° longitude. We use this data for longitude between 195° to 300°E and latitude from 22.5° to 55°N, which roughly corresponds to the continental United States excluding Alaska. The data set is 2.6 MB in size. We considered each year as a time period and each month as an individual in the population. There are $T = 45$ time periods in all, $N = 12$ calendar months, and 602 different spatial

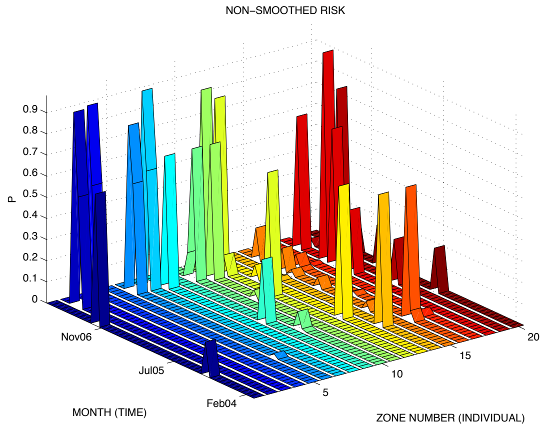

Fig. 3. Baseline MLE estimate for risk $P_{ti}(u_*)$ of 10-year power load exceedance in each service zone depending on time (months).

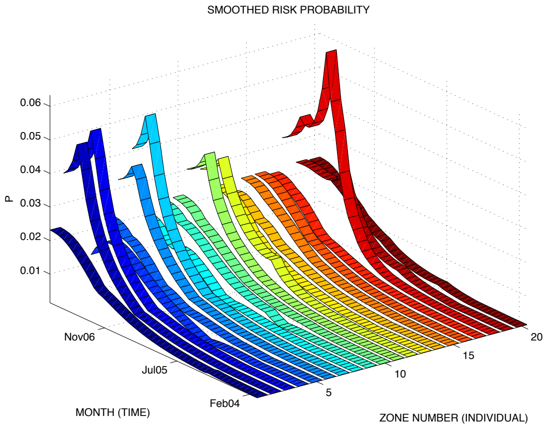

Fig. 4. MAP estimate for risk $P_{ti}(u_*)$ of 10-year power load exceedance in each service zone depending on time (months).

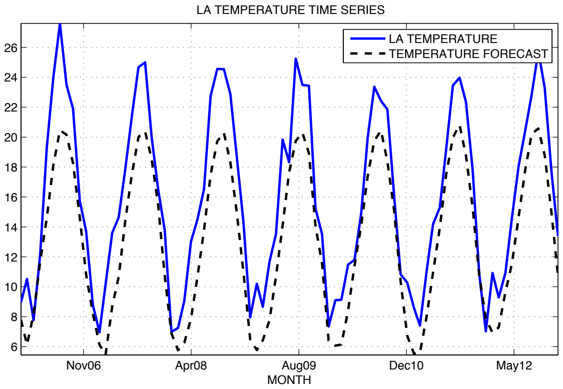

Fig. 5. Mean monthly temperature in °C at Los Angeles, CA location compared to monthly mean forecast from 2006 to 2012.

locations. In this example, the computational time of all processing steps in Figure 1, performed on single processor is under 1 sec.

As regressors x_{tia} (1) we used the binary vectors that indicate the calendar month for the current time period. More specifically,

$$x_{tia}^T = \left[\begin{array}{cccc} m_0(i) & m_1(i) & \cdots & m_{11}(i) \end{array}\right], \tag{73}$$

where $m_k(i)$ is defined in (64).

Let T_{tia} be the temperature in °C at location with index a for year t and month i. We choose the dependent variable $y_{tia} = T_{tia}$. We select $A_{ti} = 1.66 \cdot \hat{\sigma}_{ti}$. Figure 5 is a sample monthly temperature time series at the grid node location near Los Angles, CA compared to the monthly forecast.

We choose Ω_{ti} from (15) based on quantile level of 99.91% for each individual. This means that 99.91% of data points for each individual is below this threshold.

Using the same notation as Section V-A, we choose the two following prior structures (38)–(39). For $\theta_{t,j}$,

$$\text{vec}\left\{(\Delta\psi)_k(\theta)\right\} = \text{col}(v', v'', v'''), \tag{74}$$

$$v' = \text{vec}\left\{\psi(\theta_{t,j}) - \psi(\theta_{t-1,j})\right\}_{t=2,j=1}^{T,N}, \tag{75}$$

$$v'' = \text{vec}\left\{\psi(\theta_{t,j}) - \psi(\theta_{t,j-1})\right\}_{t=1,j=2}^{T,N}, \tag{76}$$

$$v''' = \text{vec}\left\{\psi(\theta_{t,1}) - \psi(\theta_{t-1,N})\right\}_{t=2}^{T}. \tag{77}$$

The structure of the priors for $q_{t,j}$ is

$$\text{vec}\left\{(\Delta\phi)_k(q)\right\} = \text{col}(w', w'', w''), \tag{78}$$

$$w' = \text{vec}\left\{\phi(q_{t,j}) - \phi(q_{t-1,j})\right\}_{t=2,j=1}^{T,N}, \tag{79}$$

$$w'' = \text{vec}\left\{\phi(q_{t,j}) - \phi(q_{t,j-1})\right\}_{t=1,j=2}^{T,N}, \tag{80}$$

$$w''' = \text{vec}\left\{\phi(q_{t,1}) - \phi(q_{t-1,N})\right\}_{t=2}^{T}. \tag{81}$$

The priors (75) and (76), (77) have strengths α' and α'', α'' respectively. The priors (79) and (80), (81) have strengths η' and η'', η''. Table I summarizes the parameters $\alpha', \alpha'', \eta', \eta''$ used in the numerical example.

We describe the prior structures (75)–(81) with matrix F_1. Using the same notation as in (66)–(70), we get

$$F_1 = \left[\begin{array}{ccc} \multicolumn{3}{c}{I_N \otimes D_T} \\ \hline \multicolumn{3}{c}{D_N \otimes I_T} \\ \hline V_1 & \mathbf{0}_{(T-1)\times T(N-2)} & V_2 \end{array}\right], \tag{82}$$

where $D_N \in \Re^{(N-1)\times N}$ is the first difference matrix, $V_1 \in \Re^{(T-1)\times T}$ is a sparse matrix with 1 above the main diagonal, and $V_2 \in \Re^{(T-1)\times T}$ is a sparse matrix with -1 on the main diagonal.

Just as in Section V-A, we compare risk $R_{ti}(u_*)$ (7) for the baseline MLE estimates with the risk computed for the MAP estimates of the parameters $\theta_{t,i}$ and $q_{t,i}$. The average probability of exceeding u_* is a 1-in-

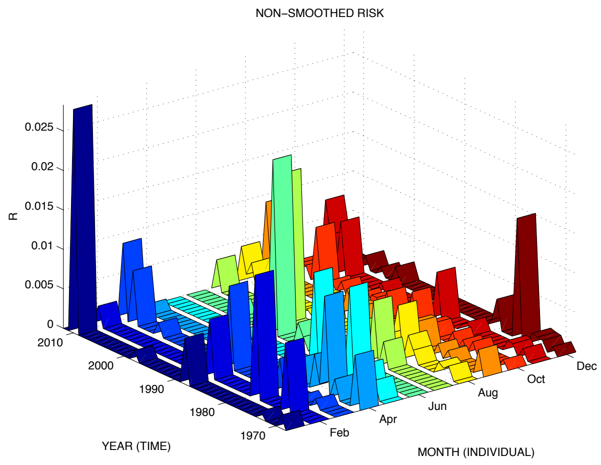

Fig. 6. Baseline MLE estimate of 100-year extreme temperature risk $R_{ti}(u_*)$ for each calendar month (individual) depending on year.

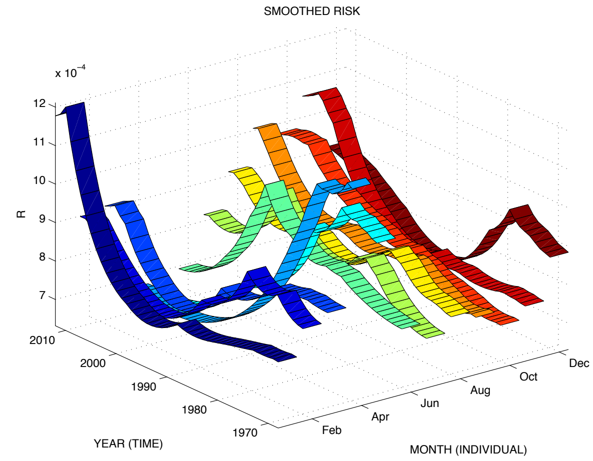

Fig. 7. MAP estimate of 100-year extreme temperature risk $R_{ti}(u_*)$ for each calendar month (individual) depending on year.

100 years event, with $R_* = 0.01/12$. Figure 6 shows the computation of risk R_{ti} with the baseline MLE parameter estimates. Figure 7 shows the computation of risk R_{ti} with the MAP estimates. As before, the smoothed MAP estimates are clearly superior to and more informative then the baseline MLE estimates.

The risk exceedance probability is approximately 0.08% on a monthly average, which roughly translates to 1% on a yearly average. Figure 7 shows that there is a noticeable increase in risk probability of an extreme high temperature event in January, September, October, and November. There is a slight or negligible change in risk for the other months. The increasing trend in high temperature risk could be attributed to the global warming trend. The formulated combined longitudinal and cross-sectional analysis provides the additional insight about risk trends for different calendar months.

References

[1] L. de Haan and A. Ferreira, *Extreme Value Theory: An Introduction*. New York: Springer, 2006.

[2] R. L. Smith, "Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone," *Statistical Science*, pp. 367–377, 1989.

[3] R. L. Smith and T. S. Shively, "Point process approach to modeling trends in tropospheric ozone based on exceedances of a high threshold," *Atmospheric Environment*, vol. 29, no. 23, pp. 3489–3499, 1995.

[4] V. Chavez-Demoulin, P. Embrechts, and S. Sardy, "Extreme-quantile tracking for financial time series," *Journal of Econometrics*, vol. 181, no. 1, pp. 44–52, 2014.

[5] D. A. Clifton, L. Clifton, S. Hugueny, D. Wong, and L. Tarassenko, "An extreme function theory for novelty detection," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 1, pp. 28–37, 2013.

[6] E. Vanem, "Long-term time-dependent stochastic modelling of extreme waves," *Stochastic Environmental Research and Risk Assessment*, vol. 25, no. 2, pp. 185–209, 2011.

[7] P. J. Northrop and P. Jonathan, "Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights," *Environmetrics*, vol. 22, no. 7, pp. 799–809, 2011.

[8] S. Shenoy and D. Gorinevsky, "Estimating long tail models for risk trends," *IEEE Signal Processing Letters*, vol. 22, pp. 968–972, 2015.

[9] S. Shenoy and D. Gorinevsky, "Gaussian-Laplacian mixture model for electricity market," in *IEEE Conference on Decision and Control*, (Los Angeles, CA), IEEE CDC, December 2014.

[10] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts," *Signal Processing Magazine, IEEE*, vol. 29, no. 4, pp. 61–80, 2012.

[11] P. J. Huber and E. M. Ronchetti, *Robust Statistics*. New Jersey: Wiley, 2nd ed., 2009.

[12] A. Rajaraman and J. Ullman, *Mining of Massive Datasets*. New York: Cambridge University Press, 2012.

[13] S. Shenoy and D. Gorinevsky, "Risk adjusted forecasting of electric power load," in *American Control Conference (ACC), 2014*, pp. 914–919, IEEE, 2014.

[14] A. Shiryaev and S. Wilson, *Probability*. Graduate Texts in Mathematics, Springer New York, 1995.

[15] H. Raiffa, *Applied Statistical Decision Theory*. Div. of Research, Graduate School of Business Administration, Harvard Univ., 1974.

[16] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge University Press, 2004.

[17] Kaggle.com, "Global energy forecasting competition 2012 load forecasting." Available: https://www.kaggle.com/c/global-energy-forecasting-competition-2012-load-forecasting.

[18] S. Shenoy, D. Gorinevsky, and S. Boyd, "Non-parametric regression modeling for stochastic optimization of power market forecast," in *American Control Conference*, (Chicago, IL), July 2015.

[19] Earth Systems Research Laboratory, "NCEP/NCAR reanalysis 1." Available: http://www.esrl.noaa.gov/psd/gridded/data.ncep.reanalysis.surface.html.