Spatio-temporal Graphical Models for Extreme Events

Hang Yu, Liaofan Zhang, and Justin Dauwels
School of Electrical and Electronics Engineering,
Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798

***Abstract*—We propose a novel statistical model to describe the spatio-temporal extreme events. The model can be used to estimate the extreme-value temporal pattern such as the seasonality and trend, and further to predict the distribution of extreme events in the future. The basic idea is to explore graphical models to capture the highly structured dependencies among extreme events measured in time and space. More explicitly, we first assume the single observation at each location and time instant follows a Generalized Extreme Value (GEV) distribution. The spatio-temporal dependencies are further encoded via graphical models imposed on the GEV parameters. We develop efficient learning and inference algorithms for the resulting non-Gaussian graphical model. Results of both synthetic and real data demonstrate the effectiveness of the proposed approach.**

I. Introduction

Analysis of multiple extreme-value time series has found applications and permeated the literature in a wide variety of domains, ranging from finance to climatology. Extreme precipitation, for example, can characterize climate change and cause flood or flash-flood related hazards. Therefore, assessing the temporal pattern of such events and making reliable predictions of future trends is crucial for risk management and disaster prevention.

One of the most common approaches for describing the extreme events of stationary data is the block maximum approach, which models the maxima of a set of contiguous blocks of observations using the Generalized Extreme Value (GEV) distribution [1]. It has been shown that extreme events that are sufficiently separated (e.g., annual maxima) are almost independent [1]. However, increasing the block size to well separate the block maxima often leads to a small sample size, and thus, the resulting estimates of GEV parameters are unreliable. The problem can be alleviated by considering the dependence between multiple GEV distributed variables (e.g., annual maximum precipitation at multiple locations) [2]. On the other hand, instead of using one single sample (i.e., the maximum) in a large block, we can shrink the size of the blocks and introduce temporal dependence to the extreme-value samples. The resulting analysis can utilize the data more effectively and often copes with spatio-temporal extreme values. A clear account of temporal dependence is presented in [1], which defines the temporal changes of GEV parameters through deterministic functions, such as linear, log-linear and quadratic functions. Unfortunately, the restrictive functional forms pose a serious limitation in practice. A more satisfactory approach is to replace the deterministic function with a linear combination of suitable basis functions, such as splines, and add a penalty to guard against overfitting [3]. As an alternative, hierarchical Bayesian models are also constructed, capturing spatial and temporal dependence in a flexible manner [4], [5]. The parameters of the model are inferred by Bayesian fitting with Monte Carlo Markov Chain method. However, such models can be quite computationally intensive and thus are prohibitive for large-scale problems. Recently, dynamic linear models (DLM) are employed in [6], [7], relating the present GEV parameters to the historical estimates while embedding the spatial dependence in the evolution matrix. Although DLMs are equipped with efficient inference algorithms, they use directed acyclic graphs to represent the dependence from past to present, and hence, the estimates of GEV parameters at time t only depend on extreme-value samples till t. Obviously, it is more tempting to take full advantage of all the observed samples, both before and after t, to yield more reliable estimates of GEV parameters. This indicates that GEV parameters at different time instants depend on each other, thus corresponding to an undirected graph.

In this paper, we propose to exploit undirected graphical models (i.e., Markov random fields) to capture the highly structured spatio-temporal dependencies among GEV parameters. We aim to estimate the temporal pattern of extreme events, such as the trend or seasonality of the data in time. Furthermore, we intend to predict the distribution of extreme events in the future based on the current trend. In the example of extreme rainfall, forecasting whether the size of extremes will increase in the future is the key for flood warning and strategic planning. To move forward to this goal, we first assume that the block maximum, particularly, monthly maximum, at each site and time position (i.e., each month) to follow a GEV distribution. We further assume that the shape and scale parameters of GEV distributions are constant in the spatio-temporal domain, therefore, the remaining location parameters characterize the variation and dependence in the data. The latter is well encoded by imposing a Gauss-Markov random field (GMRF) prior on the location parameters. We follow the empirical Bayes approach to estimate all the parameters. In particular, both the GEV parameters and the hyper parameters of the GMRF prior are learnt via expectation maximization (EM). To cope with the complicated functional form of GEV distributions, we recast the problem to a tractable one using a series of approximation algorithms. First of all, the GEV density is represented as a mixture of Gaussians for each specific shape parameter in a given discrete domain, as suggested

in [9]. We then select the shape parameter that maximizes the evidence of the data. In addition, we employ parallel expectation propagation (EP) [8] to compute the sufficient statistics in the E-step. Interestingly, the Gaussian mixture representation of GEV distributions aids in fast implementation of EP algorithm. Numerical results show that the proposed model can automatically recover the underlying pattern of location parameters across both space and time, although only one sample is observed at each location and time instant. Moreover, it also provides an effective tool to predict the future distribution of extreme events.

The paper is organized as follows. In Section II, we present the proposed spatio-temporal model for extreme events. We then develop the learning and inference algorithm in Section III. Numerical results are presented in Section 4. Section 5 closes the paper with concluding remarks.

II. Spatio-temporal Models

In this section, we present the proposed spatio-temporal graphical model for extreme events. Supposed that we have T monthly maxima $x_{p,t}$ at each of the P locations, where $p = 1, \cdots, P$ and $t = 1, \cdots, T$. The resulting dimension of the spatio-temporal model is $M = TP$. The set of observed spatio-temporal indices is denoted as $\mathcal{V}_o$.

A. Likelihood: Generalized Extreme Value Distribution

We assume that each observed $x_{p,t}$ follows a Generalized Extreme Value (GEV) distribution [1]:

$$F(x_{p,t}) = \exp\{-[1 + \frac{\xi_{p,t}}{\sigma_{p,t}}(x_{p,t} - u_{p,t})]_+^{-\frac{1}{\xi_{p,t}}}\}, \quad (1)$$

with shape parameter $\xi_{p,t}$, scale parameter $\sigma_{p,t}$ and location parameter $u_{p,t}$. In practice, the shape parameter for modeling sample extremes is usually in the domain of $[-1, 1]$. Let $f(x; \xi, \sigma, u)$ denote the GEV density function, thus general GEV distributions can be obtained from the GEV distribution with zero location and unit scale (i.e., $f_0(x; \xi) = f(x; \xi, 1, 0)$) through the following transition:

$$f(x; \xi, \sigma, u) = \frac{1}{\sigma} f_0(\frac{x - u}{\sigma}; \xi). \quad (2)$$

Direct learning of GEV parameters is usually problematic due to its complicated functional form. A reasonable remedy is to approximate $f_0(x; \xi)$ by a mixture of Gaussians, that is,

$$f_0(x; \xi) \approx \sum_{k=1}^{K} w_{k,\xi} \mathcal{N}(x; m_{k,\xi}, v_{k,\xi}), \quad (3)$$

where the weight $w_{k,\xi}$, mean $m_{k,\xi}$ and variance $v_{k,\xi}$ of each component k depend on the shape parameter ξ. Wand *et al.* [9] provide a look-up table for Gaussian mixture approximation with $K = 24$ components of GEV distributions with shape parameter $\xi \in \Xi = \{-1, -0.995, \cdots, 0.995, 1\}$. Taken (2) and (3) together, the auxiliary mixture approach can handle any GEV distributions with $\xi \in \Xi$.

B. Prior: Thin-plate Models

We now turn our attention the GMRF prior on the location parameters. Particularly, we construct the prior based on thin-plate models. As such, we give a short description of graphical models and the special case of thin-plate models.

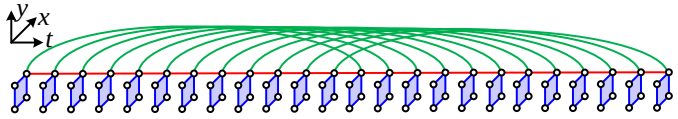

Fig. 1. Neighborhood structure of the spatio-temporal graphical model.

In an undirected graphical model (i.e., a MRF), the probability distribution is represented by an undirected graph $\mathcal{G}$ which consists of nodes $\mathcal{V}$ and edges $\mathcal{E}$. Each node i is associated with a random variable z_i. An edge (i, j) is absent if the corresponding two variables z_i and z_j are conditionally independent. In particular, for Gaussian distributed z, the graph $\mathcal{G}$ is characterized by the precision matrix (the inverse covariance) J, i.e., $J(i, j) \neq 0$ if and only if the edge $(i, j) \in \mathcal{E}$ [10]. The thin-plate model is a GMRF that is commonly used as smoothness prior as it penalizes the second-order difference. In other words, we model the second-order differences as a Gaussian distribution:

$$\Delta^2 z_i \sim \mathcal{N}(0, \alpha^{-1}). \quad (4)$$

As a result, the density function of a thin-plate model with a chain structure can be written as [10]:

$$p(\boldsymbol{z}) \propto \exp\Big(-\frac{\alpha}{2}\sum_{i=1}^{n-2}(z_i - 2z_{i+1} + z_{i+2})^2\Big) \quad (5)$$

where the variables z_i's are evenly located on the chain and the smoothness parameter α harnesses the penalty on the curvature. It is easy to tell from (5) that the thin-plate model is invariant to the addition of a linear function along the Markov chain. As such, this prior captures the increasing or decreasing trend in the data in a straightforward manner. Furthermore, the unspecified boundary condition of (5) aids in predicting (or extrapolating) the future values [10], i.e.,

$$E(z_{i+1}|z_1, \cdots, z_i; \alpha_z) = z_i + (z_i - z_{i-1}). \quad (6)$$

Therefore, the conditional mean of z_{i+1} is simply the linear extrapolation based on the last two observations z_{i-1} and z_i. In addition, constant boundary condition is applied more often in practice, that is, $z_{i+1} = z_i$. Consequently, we can define the second-order differences at the boundary variable z_i as $\Delta^2 z_i = z_{i-1} - 2z_i + z_{i+1} = z_{i-1} - z_i$, and the resulting thin-plate model with constant boundary condition is [11]:

$$p(\boldsymbol{z}) \propto \exp\Big(-\frac{\alpha}{2}\sum_{i=1}^{n}(|N(i)|z_i - \sum_{j\in N(i)} z_j)^2\Big), \quad (7)$$

where $N(i)$ denotes the neighboring nodes of z_i and $|N(i)|$ is the number of neighbors z_i have. In (7), each node is modeled to be close to the average of its neighbors. This model can easily be extended to the case of multiple dimensions.

In the proposed model, the location parameters $u_{p,t}$'s in the spatio-temporal domain have a neighborhood structure as shown in Fig. 1. In the spatial domain, each node typically has four neighbors (two vertical and two horizontal) as represented by the blue lattices. In the temporal domain, we first connect consecutive nodes $u_{p,t}$ and $u_{p,t+1}$ together as denoted by the red edges. Moreover, in order to establish the periodic annual pattern, we connect $u_{p,t}$ and $u_{p,t+\tau}$ via green edges, where $\tau = 12$. For simplicity, we only show four nodes in each spatial lattice and only depict the temporal dependence

structure across time at one location. Since the smoothness across the subgraphs defined by the edges with different colors is usually different, we introduce a thin-plate model for each subgraph. Concretely, we employ the thin-plate model with constant boundary condition (7) for the blue and red subgraphs. In addition, we utilize the thin-plate model without boundary conditions (5) to capture the annual dependence, for the sake of future forecast. We then integrate the three models together and obtain the spatio-temporal prior:

$$p(\boldsymbol{u}) \propto |J_{\text{prior}}|_+^{0.5} \exp\{-\frac{1}{2}\boldsymbol{u}^T J_{\text{prior}} \boldsymbol{u}\}, \tag{8}$$

where $J_{\text{prior}} = \alpha_s J_s + \alpha_m J_m + \alpha_y J_y$, and $|J_{\text{prior}}|_+$ denotes the product of the non-zero eigenvalues of J_{prior}. Note that $\alpha_s J_s$ and $\alpha_m J_m$ are the precision matrices resulting from Expression (7). They capture the spatial and monthly dependence respectively. On the other hand, $\alpha_y J_y$ corresponds to the precision matrix of Expression (5) and characterizes the yearly dependence. The resulting prior precision matrix J_{prior} is rank deficient, thus, we use the improper density (8).

C. The Spatio-Temporal Extreme-Value Graphical Model

In the proposed model, we assume the shape parameter ξ and scale parameter σ are constant for all the locations and time instants, as suggested in [6], [7]. It therefore follows from the last two subsections that

$$p(\boldsymbol{x}, \boldsymbol{u}|\boldsymbol{\theta}) = p(\boldsymbol{u}|\alpha_s, \alpha_m, \alpha_y) \prod_{\{p,t\}\in\mathcal{V}_o} f(x_{p,t}; \xi, \sigma, u_{p,t}), \tag{9}$$

where $\boldsymbol{\theta}$ denotes the set of parameters $\{\xi, \sigma, \alpha_s, \alpha_m, \alpha_y\}$. Note that expression (9) can be regarded as the unnormalized joint density of a MRF with non-Gaussian unary potential functions and Gaussian pairwise potential functions. Our objective is to estimate both the GEV parameters and the smoothness parameters given observed extreme values $\boldsymbol{x}$, which is elaborated on in the next section.

III. Learning and Inference

As mentioned in Section I, we estimate all the parameters through an empirical Bayes approach. Specifically, we first estimate the common shape and scale parameter ξ and σ as well as the smoothness parameters by solving the following problem:

$$(\hat{\xi}, \hat{\sigma}, \hat{\alpha}_s, \hat{\alpha}_m, \hat{\alpha}_y) = \operatorname{argmax} p(\boldsymbol{x}|\xi, \sigma, \alpha_s, \alpha_m, \alpha_y). \tag{10}$$

(10) can be solved using expectation-maximization (EM). Due to the intractable functional form of GEV distributions, we follow [9] to approximate it by a Gaussian mixture for each ξ defined in the discrete domain Ξ (cf. Section II-A). We then select ξ such that:

$$\hat{\xi} = \operatorname*{argmax}_{\xi} \max_{\sigma,\alpha_s,\alpha_y,\alpha_y} p(x|\xi, \sigma, \alpha_s, \alpha_m, \alpha_y). \tag{11}$$

The interpretation is that we will perform the EM algorithm once for each candidate $\xi \in \Xi$ and then choose the ξ with the largest evidence. On the other hand, due to the non-Gaussian nature of the graphical model, we exploit expectation propagation (EP to approximate posterior density $p(\boldsymbol{u}|\boldsymbol{x}, \xi, \sigma, \alpha_s, \alpha_m, \alpha_y)$ by a Gaussian in the E-step. Below we describe the proposed algorithms at length.

A. The EM-EP algorithm

1) The E-step: Here, EP is employed to estimate the joint density $p(\boldsymbol{u}|\boldsymbol{x}, \boldsymbol{\theta})$. Equivalently, we approximate the incomplete likelihood $p(\boldsymbol{x}, \boldsymbol{u}|\boldsymbol{\theta})$ by a unnormalized Gaussian $q(\boldsymbol{x}, \boldsymbol{u}|\boldsymbol{\theta}) = Z\mathcal{N}(\boldsymbol{u}; \boldsymbol{\mu}_{\text{post}}, J_{\text{post}}^{-1})$ [12]. It is apparent that $p(\boldsymbol{u}|\boldsymbol{x}, \boldsymbol{\theta})$ is approximated by $q(\boldsymbol{u}|\boldsymbol{x}, \boldsymbol{\theta}) = N(\boldsymbol{u}; \boldsymbol{\mu}_{\text{post}}, J_{\text{post}}^{-1})$. Moreover, integrating out $\boldsymbol{u}$ yields the approximation of the evidence in (11), that is,

$$Z = \int q(\boldsymbol{x}, \boldsymbol{u}|\boldsymbol{\theta}) d\boldsymbol{u} \approx p(\boldsymbol{x}|\boldsymbol{\theta}). \tag{12}$$

Following the framework of EP, we factorize $q(\boldsymbol{x}, \boldsymbol{u}|\boldsymbol{\theta})$ as:

$$q(\boldsymbol{x}, \boldsymbol{u}|\boldsymbol{\theta}) = p(\mu|\alpha_s, \alpha_m, \alpha_y) \prod_{(p,t)\in\mathcal{V}_o} \tilde{t}(u_{p,t}), \tag{13}$$

where $\tilde{t}(u_{p,t}) = \tilde{Z}_{p,t}\mathcal{N}(u_{p,t}|\tilde{\mu}_{p,t}, \tilde{\nu}_{p,t})$ is the approximation to the GEV density $f(x_{p,t}; \xi, \sigma, u_{p,t})$. The parallel EP [8] then iterates between the following two steps until convergence:

1) Update the approximated posterior $\mathcal{N}(\boldsymbol{u}; \boldsymbol{\mu}_{\text{post}}, J_{\text{post}}^{-1})$ as:
$$J_{\text{post}} = J_{\text{prior}} + C^T \tilde{V}^{-1} C, \tag{14}$$
$$\boldsymbol{\mu}_{\text{post}} = J_{\text{post}}^{-1} C^T \tilde{V}^{-1} \tilde{\boldsymbol{\mu}}. \tag{15}$$
where $\tilde{V}$ is a diagonal matrix with $\tilde{\nu}_{p,t}$ on the diagonal, $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_{p,t})$, and C is the selection matrix that selects the location parameter $u_{p,t}$ where the extreme-value observations $x_{p,t}$ are available (i.e., $(p,t) \in \mathcal{V}_o$). Compute the marginals $q(u_{p,t}) = \mathcal{N}(u_{p,t}; \mu_{p,t}, \nu_{p,t})$ from the approximated joint distribution, where $v_{p,t}$'s are the diagonal elements of J_{post}^{-1}.
2) For each $(p,t) \in \mathcal{V}_o$, we update $\tilde{t}(u_{p,t})$ as follows:
   a) Remove the approximated unary potential $\tilde{t}(u_{p,t})$ from the marginal $q(u_{p,t})$ to obtain the cavity distribution $q_{\setminus p,t}(u_{p,t}) = q(u_{p,t})/\tilde{t}(u_{p,t})$. The corresponding mean and variance of $q_{\setminus p,t}(u_{p,t})$ can be computed as:
   $$\nu_{\setminus p,t} = (\nu_i^{-1} - \tilde{\nu}_i^{-1})^{-1}, \tag{16}$$
   $$\mu_{\setminus p,t} = \nu_{\setminus p,t}(\nu_i^{-1}\mu_i - \tilde{\nu}_i^{-1}\tilde{\mu}_i) \tag{17}$$
   b) Multiply $q_{\setminus p,t}(u_{p,t})$ and the true unary potential $f(x_{p,t}; \xi, \sigma, u_{p,t})$ to get $p(u_{p,t})$. Fit an unnormalized Gaussian $\hat{q}(u_{p,t}) = \hat{Z}_{p,t}\mathcal{N}(u_{p,t}; \hat{\mu}_{p,t}, \hat{\nu}_{p,t})$ to $p(u_{p,t})$ by minimizing the KL-divergence $\text{KL}(p(x_{p,t})\|\hat{q}(x_{p,t}))$. Recall that $f(x_{p,t}; \xi, \sigma, u_{p,t})$ is approximated by a mixture of Gaussians. Therefore, $\hat{\mu}_{p,t}$, $\hat{\nu}_{p,t}$, and $\hat{Z}_{p,t}$ can be updated efficiently as:
   $$\hat{Z}_{p,t} = \sum_{k=1}^{K} w_k c_k, \tag{18}$$
   $$\hat{\mu}_{p,t} = \hat{Z}_{p,t}^{-1} \sum_{k=1}^{K} w_k c_k \hat{\mu}_{p,t,k}, \tag{19}$$
   $$\hat{\nu}_{p,t} = \hat{Z}_{p,t}^{-1} \sum_{k=1}^{K} w_k c_k (\hat{\mu}_{p,t,k}^2 + \hat{\nu}_{p,t,k}) - \hat{\mu}_{p,t}^2, \tag{20}$$
   where
   $$c_k = \big(2\pi(\nu_{\setminus p,t} + \sigma^2\nu_k)\big)^{-0.5} \times \exp\Big(-\frac{(x_{p,t} - \sigma m_k - \mu_{\setminus p,t})^2}{2(\sigma^2\nu_k + \nu_{\setminus p,t})}\Big), \tag{21}$$
   $$\hat{\nu}_{p,t,k} = (\nu_{\setminus p,t}^{-1} + (\sigma^2\nu_k)^{-1})^{-1}, \tag{22}$$

$$\hat{\mu}_{p,t,k} = \hat{\nu}_{p,t,k}\Big(\frac{\mu_{\setminus p,t}}{\nu_{\setminus p,t}} + \frac{x_{p,t} - \sigma m_k}{\sigma^2 \nu_k}\Big). \quad (23)$$

c) Update the approximated unary potential as $\tilde{t}(u_{p,t}) = \hat{q}(x_{p,t})/q_{\setminus p,t}(u_{p,t})$, i.e.,

$$\tilde{\nu}_{p,t} = (\hat{\nu}_{p,t}^{-1} - \nu_{\setminus p,t}^{-1})^{-1}, \quad (24)$$

$$\tilde{\mu}_{p,t} = \tilde{\nu}_{p,t}(\hat{\nu}_{p,t}^{-1}\hat{\mu}_{p,t} - \nu_{\setminus p,t}^{-1}\mu_{\setminus p,t}) \quad (25)$$

$$\tilde{Z}_{p,t} = \hat{Z}_{p,t}\big(2\pi(\nu_{\setminus p,t} - \tilde{\nu}_{p,t})\big)^{0.5} \times \exp\Big(\frac{(\mu_{\setminus p,t} + \tilde{\mu}_{p,t})^2}{2(\nu_{\setminus p,t} + \tilde{\nu}_{p,t})}\Big). \quad (26)$$

Different from the original EP algorithm, parallel EP recomputes marginal means and variances only after updating all the $\tilde{t}_{p,t}$'s, thus reducing the number of computing the posterior mean and variance in (14) and (15) and speeding up the EP process.

The final evidence can be approximated as:

$$p(\boldsymbol{x}|\boldsymbol{\theta}) \approx \prod_{(i,p)\in\mathcal{V}_o} \tilde{Z}_{p,t} \frac{\mathcal{N}(\boldsymbol{u}; \tilde{\boldsymbol{\mu}}, \tilde{V}) p(\boldsymbol{u}|\alpha_s, \alpha_m, \alpha_y)}{N(\boldsymbol{u}; \boldsymbol{\mu}_{\text{post}}, J_{\text{post}}^{-1})}. \quad (27)$$

Taking logarithm on both sides yields:

$$\log p(\boldsymbol{x}|\boldsymbol{\theta}) \approx \sum_{(i,p)\in\mathcal{V}_o} \log(\tilde{Z}_{p,t}) - \frac{1}{2}\log\det\tilde{V} - \frac{1}{2}\tilde{\boldsymbol{\mu}}^T\tilde{V}^{-1}\tilde{\boldsymbol{\mu}} + \frac{1}{2}\log|J_{\text{prior}}|_+ - \frac{1}{2}\log\det J_{\text{post}} + \frac{1}{2}\boldsymbol{\mu}_{\text{post}}^T C^T\tilde{V}^{-1}\tilde{\boldsymbol{\mu}}. \quad (28)$$

2) M-step: We now aim to maximize $w.r.t.$ $(\sigma, \alpha_s, \alpha_m, \alpha_y)$ the following lower bound:

$$\int q(\boldsymbol{u}|\boldsymbol{x}, \boldsymbol{\theta}) \log p(\boldsymbol{x}, \boldsymbol{u}|\theta) d\boldsymbol{u}. \quad (29)$$

For the smoothness parameters, (29) boils down to

$$\log|J_{\text{prior}}|_+ - c_s\alpha_c - c_m\alpha_m - c_y\alpha_y, \quad (30)$$

where c_s, c_m and c_y are constants. By equating partial derivatives $w.r.t$ $(\alpha_s, \alpha_m, \alpha_y)$ to zero and solving the system of equations, (30) can be further simplified as:

$$(\alpha_s, \alpha_m, \alpha_y) = \operatorname{argmax}\log|J_{\text{prior}}|_+ \quad (31)$$

$$\text{s.t. } c_s\alpha_s + c_m\alpha_m + c_y\alpha_y = M - 1. \quad (32)$$

The above determinant maximization problem can be solved using interior point method. On the other hand, when estimating the common scale parameter σ, (29) becomes:

$$\int q(\boldsymbol{u}|\boldsymbol{x}, \boldsymbol{\theta}) \sum_{(p,t)\in\mathcal{V}_o} \log f(x_{p,t}|\xi, \sigma, u_{p,t}) d\boldsymbol{u}. \quad (33)$$

Since the integration is intractable, we approximate $f(x_{p,t}|\xi, \sigma, u_{p,t})$ by a Gaussian distribution. The mean and variance of the latter (as a function of σ) can be obtained from the Gaussian mixture approximation of $f(x_{p,t}|\xi, \sigma, u_{p,t})$. The function to be maximized therefore has a nice form of $c_1/\sigma^2 + c_2/\sigma + c_3\log\sigma$, where c_1, c_2 and c_3 are constants, and can be solved analytically.

B. Prediction of Future GEV Parameters

A primary goal of the spatio-temporal model is to predict the future GEV distributions. Since we assume the shape and location parameters are time invariant, we only need to estimate the location parameters. Recall that the thin-plate model without boundary conditions serves as a natural tool to predict future using current trend (cf. Section II-B), we thus incorporate the future locations parameters at the right end of the neighborhood structure of the spatio-temporal model (see Fig. 1). As a result, the future location parameters $\boldsymbol{u}_f$ and the estimated observed location parameter u_o together form a GMRF. According to Schur complement, we can obtain the MAP estimates:

$$\boldsymbol{u}_f = -[J_{\text{prior}}]_{f,f}^{-1}[J_{\text{prior}}]_{f,o}\boldsymbol{u}_o. \quad (34)$$

Note that $J_{\text{prior}} = \alpha_s J_s + \alpha_m J_m + \alpha_y J_y$. Moreover, as α_m (i.e., the smoothness parameter dictating monthly dependence) is usually small, only $\alpha_s J_s$ and $\alpha_y J_y$ are required for prediction. Consequently, (34) boils down to

$$\boldsymbol{u}_{(1:P),t} = \alpha_y(\alpha_s[J_s]_{P\times P} + \alpha_y I_{P\times P})^{-1} \times (2\boldsymbol{u}_{(1:P),t-\tau} - \boldsymbol{u}_{(1:P),t-2\tau}), \quad (35)$$

where $u_{(1:P),t}$ represents the location parameters of all the P sites at time t, $I_{P\times P}$ is the $P \times P$ identity matrix, and τ denotes the period that equals 12 in our case.

IV. Experimental Results

In this section, we apply our model to synthetic and real data; we benchmark the proposed model against a spatial model (without considering the temporal dependence) and a temporal model (without considering the spatial dependence). We compare the three models on two criteria: (1) mean squared error (MSE) between the estimated GEV parameters and the ground truth; (2) Akaike information criterion (AIC).

A. Synthetic Data

We generate synthetic data by first specifying the GEV parameters in the spatio-temporal domain and then drawing one single sample at each location and time instant. Specifically, we select 25 sites arranged on a 5×5 grid with 240 monthly maximum observations for each site. The shape and scale parameters are chosen to be constant, and equal to 0.15 and 3 respectively. The location parameter is a quadratic function across the space and a trigonometric function with period $\tau = 12$ across time. We further add a quadratic temporal trend to the location parameters.

We first depict the approximated evidence (28) as a function of shape parameter ξ in Fig. 2; the evidence attains the maximum when $\xi = 0.15$, which is the consistent with the ground truth. This indicates that the approximated evidence offers a reliable criterion for selecting proper shape parameter.

Next, we summarize the results of estimating location parameters in Fig. 3 and Fig. 4. We can see that the proposed spatio-temporal model outperforms the other two models; the estimates given by this model closely follow the ground truth across both space and time. On the other hand, although the spatial model can recover the spatial pattern of the location parameters, it overestimates those parameters, and more unfortunately, it ignores the temporal variation of the locations parameters. Similarly, the temporal model overestimates the location parameters changing with time and it fails to characterize the spatial variation of the locations parameters. We further list the MSE and AIC score resulting from the three models in Table I. It is worth noticing that even though there is the only one sample with noise available at each location and

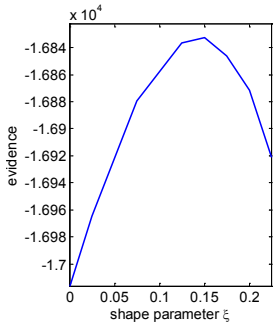

Fig. 2. Evidence as a function of shape parameter ξ.

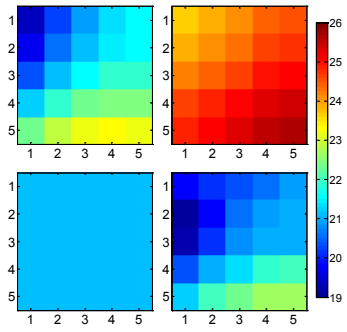

Fig. 3. Location parameters across space given sequentially by (upper left to lower right) the ground truth; the temporal model; the spatial model; and the spatio-temporal model.

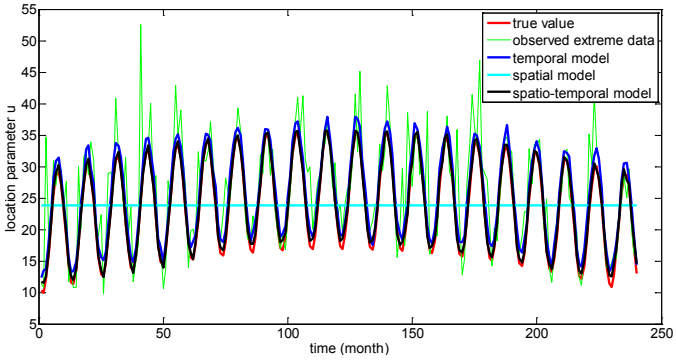

Fig. 4. Estimation of location parameters across time at one location.

time instant, the proposed model still yields reliable estimates of GEV parameters. Furthermore, the AIC scores suggest that the spatio-temporal model fits the data best.

Finally, we predict future GEV distributions in the next year using the method developed in Section III-B. The resulting MSE between the predicted location parameters and the ground truth is 63.3624, 3.5600, and 1.7114 respectively for the spatial model, the temporal model and the spatio-temporal model. Obviously, the proposed model achieves the best performance.

TABLE I
QUANTITATIVE COMPARISON OF DIFFERENT MODELS.

Models	Mean Square Error (MSE)			AIC
	ξ	σ	u	
Spatial model	0.1348	0.8281	49.8018	1.9951×10^{6}
Temporal model	0.0140	3.3546	1.4293	8.1243×10^{4}
Spatio-temporal model	0	5.8510×10^{-4}	0.3846	4.4600×10^{4}

B. Real Data

We now consider the extreme precipitation in Thailand, which are the major causes of severe floods in recent years. The daily rainfall data from 1951 to 2007 is compiled and interpolated onto a grid with resolution 0.25° [13]. We select a 5×5 grid with spacing 0.75°, and then fit the proposed spatio-temporal model using 120 monthly maxima (i.e., 10 years). We then predict the location parameters in the 11^{th} year. We also compute the 95% confidence interval of the extreme rainfall amount given the predicted GEV parameters. The result of one arbitrary location is shown in Fig. 5. We can find that the observed extremes lie in the predicted the 95% confidence interval, implying the reliability of prediction.

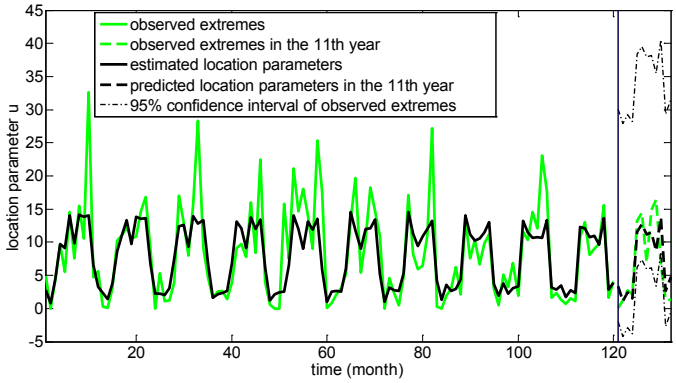

Fig. 5. Estimation and prediction of location parameters across time at one location for extreme precipitation in Thailand.

We then compare the proposed model with spatial model and temporal model in terms of AIC score; the AIC scores for the three models are respectively 2.5315×10^4, 6.9735×10^4 and 2.2909×10^5. Obviously, the proposed model is superior to other models when expressing the behavior of monthly extreme rainfall amount.

V. CONCLUSION

A novel spatio-temporal graphical model is proposed to model extreme events. It can estimate the GEV parameters reliably given one single extreme-value sample at each location and time instant.

REFERENCES

[1] S. G. Coles, *An Introduction to Statistical Modeling of Extreme Values*, Springer, London, 2001.

[2] H. Yu, Z. Choo, J. Dauwels, P. Jonathan, and Q. Zhou, "Modeling Spatial Extreme Events using Markov Random Field Priors," *Proc. ISIT*, pp. 1453–1457, 2012.

[3] V. Chavez-Demoulin and A. C. Davison, "Modelling Times Series Extremes," *Statistical Journal*, vol. 10, pp. 109-133, 2012.

[4] H. Sang, and A. E. Gelfand, "Hierarchical modeling for extreme values observed over space and time," *Environmental and Ecological Statistics* vol. 16, pp. 407–426, 2009.

[5] S. Ghosh and B. K. Mallick, "A hierarchical Bayesian spatio-temporal model for extreme precipitation events,", *Environmetrics*, vol. 22, pp. 192-204, 2011.

[6] G. Huerta and B. Sansó, "Time-Varying Models for Extreme Values," *Environmental and Ecological Statistics*, vol. 14, pp. 285-299, 2007.

[7] Y. Liu, M. T. Bahadori, and H. Li, "Sparse-GEV: Sparse Latent Space Model for Multivariate Extreme Value Time Series Modeling," *Proc. ICML*, 2012.

[8] B. Cseke, and T. Heskes, "Approximate Marginals in Latent Gaussian Models," *Journal of Machine Learning Research*, vol. 12, pp. 417-454, 2011.

[9] M. P. Wand, J. T. Ormerod, S. A. Padoan, and R. Frühwirth, "Mean Field Variational Bayes for Elaborate Distributions," *Bayesian Analysis*, vol. 7, no. 2, pp. 847-900, 2012.

[10] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications*, Chapman & Hall, 2005.

[11] Y. Yue and P. L. Speckman, "Nonstationary Spatial Gaussian Markov Random Fields," *Journal of Computational and Graphical Statistics*, vol. 19, no. 1, pp. 96-116, 2010.

[12] C. Skar, "The Expectation Propagation Algorithm for use in Approximate Bayesian Analysis of Latent Gaussian Models," S. M. Thesis, Norwegian University of Science and Technology, 2010.

[13] A. Yatagai, K. Kamiguchi, O. Arakawa, A. Hamada, N. Yasutomi and A. Kitoh, "APHRODITE: Constructing a Long-term Daily Gridded Precipitation Dataset for Asia based on a Dense Network of Rain Gauges," *Bulletin of American Meteorological Society* , doi:10.1175/BAMS-D-11-00122.1, 2012