

MODELING SPATIAL EXTREMES VIA ENSEMBLE-OF-TREES OF PAIRWISE COPULAS

*Hang Yu**, *Wayne Isaac T. Uy**, and *Justin Dauwels*

School of Electrical and Electronics Engineering, School of Physical and Mathematical Sciences
Nanyang Technological University, 50 Nanyang Avenue, Singapore, 639798

ABSTRACT

Assessing the risk of extreme events in a spatial domain, such as hurricanes, floods and droughts, presents unique significance in practice. Unfortunately, the existing extreme-value statistical models are typically not feasible for practical large-scale problems. Graphical models are capable of handling enormous number of variables, yet have not been explored in the realm of extreme-value analysis. To bridge the gap, an extreme-value graphical model is introduced in this paper, i.e., ensemble-of-trees of pairwise copulas (ETPC). In the proposed graphical model, extreme-value marginal distributions are stitched together by means of pairwise copulas, which in turn are the building blocks of the ensemble of trees. By exploiting this particular structure, novel efficient inference algorithms are derived that are applicable to large-scale statistical problems involving extreme values. It is proven that, under mild conditions, the ETPC model exhibits the favorable property of tail-dependence between an arbitrary pair of sites (variables), and therefore is reliable to capture the dependence between extremes at different sites. Real data results further demonstrate the advantages of the ETPC model.

Index Terms— extreme events, pairwise copulas, graphical models, ensemble of trees, tail dependence

1. INTRODUCTION

Extreme events such as hurricanes, earthquakes, and floods often have a major impact on our society. To assess the likelihood of such events, statisticians have developed extreme-value theory [1]. However, the existing extreme-value models are often limited to tens of variables while many practical problems, for instance in Earth Sciences, involve hundreds or thousands of sites (variables). Although graphical models can harness such large number of variables [2, 3], they have not yet been applied to extreme-event analysis. In this paper, we address this gap by introducing an extreme-value graphical model, i.e., an ensemble-of-trees of pairwise copulas.

Spatial extremes are often modeled in two stages [4]. First the marginals of the extreme values at each site are learned. The marginal parameters are typically coupled in space to address the issue of inaccurate marginal estimation due to the lack of extreme-value samples [5, 6, 7]. In the second

stage, the marginal models are “glued” together via copulas or max-stable processes. For example, in the approach of [8], a Gaussian copula connects the marginal distributions to form a joint distribution. Unfortunately, the resulting dense covariance matrix in the Gaussian latent layer is computationally prohibitive for high-dimensional data. In [9], the problem is settled instead by using a Gaussian copula graphical model with a sparse inverse covariance matrix. A major limitation, however, is that Gaussian copulas are asymptotically independent, rendering them incapable of capturing the tail dependence between extremes. This shortcoming has sparked interest in asymptotically dependent models, including max-stable processes and extreme-value (max-stable) copulas [10, 11]. However, due to their complexity, only the density functions of bivariate max-stable processes or copulas are tractable. A promising strategy is to use bivariate extreme-value copulas as building blocks to construct large multivariate extreme-value models. A fruitful step in that direction has been taken in [10, 11], where composite likelihoods [12] are constructed from pairwise copulas. There is no principled approach to impose structure on composite likelihoods, however, posing a serious drawback. On the other hand, vine copula models [13] decompose the joint density into a product of pairwise conditional densities and approximate them by pairwise copulas. Despite its popularity, the complexity of the vine structure increases quadratically with respect to the number of variables, making it intractable in large scale cases.

In this paper, we introduce an ensemble-of-trees model [14] of pairwise copulas (ETPC) to deal with high-dimensional spatial extremes. As a starting point, the sites in the spatial domain are arranged on a lattice. The probability density function (PDF) of the ETPC model is a weighted sum of the PDF of all possible spanning trees on that lattice. The PDF of these trees in turn are constructed from pairwise copulas [15]. In this setting, the extremes can be modeled as asymptotically tail dependent or independent [16, 17] by choosing the pairwise copulas appropriately. It can be proven that tail dependence in the ETPC model is preserved if all the pairwise copulas in the model are tail dependent. We then propose efficient learning algorithms, and also derive scalable inference algorithms to impute extreme values at unobserved sites and to perform conditional simulation. Numerical results for extreme precipitation data in Japan suggest that the proposed

*Both authors contributed equally to the work.

ETPC model is suitable for spatial extreme-event analysis.

This paper is structured as follows. In Section 2 we describe the extreme-value marginal distributions. In Section 3, we explain our proposed ETPC model, which couples the extreme-value marginal distributions through pairwise copulas, arranged in an ensemble of trees. Inference methods for the ETPC model are outlined in Section 4. A theoretical guarantee on tail dependence of the ETPC model is presented in Section 5. Numerical results for precipitation data from Japan are provided in Section 6. Finally, we offer concluding remarks in Section 7.

2. EXTREME-VALUE MARGINAL DISTRIBUTIONS

Here we briefly review the theory of marginal distributions of extreme values. In this study, we consider maxima over a particular time period, e.g., monthly or annual maxima. Suppose that we have N samples $x_i^{(n)}$ at each of P sites, where $i = 1, \dots, P$ and $n = 1, \dots, N$. The extreme value theory states that the block maxima of i.i.d. univariate samples x_i at each location converge to the three-parameter Generalized Extreme Value (GEV) distribution with cumulative distribution function (CDF) [1]:

$$F(x_i) = \begin{cases} \exp\{-[1 + \frac{\xi_i}{\sigma_i}(x_i - \mu_i)]^{-\frac{1}{\xi_i}}\}, & \xi_i \neq 0 \\ \exp\{-\exp[-\frac{1}{\sigma_i}(x_i - \mu_i)]\}, & \xi_i = 0, \end{cases} \quad (1)$$

where $\mu_i \in \mathbb{R}$, $\sigma_i > 0$, and $\xi_i \in \mathbb{R}$ denote the location, scale and shape parameter, respectively. To improve the accuracy of the estimated GEV parameters, we couple those parameters in space by means of a thin-membrane model as in [5, 9].

3. ENSEMBLE-OF-TREES OF PAIRWISE COPULAS

In this section, we proceed to tie the GEV marginal distributions together by means of statistical copulas, i.e., an ensemble-of-trees of pairwise copulas (ETPC). To this end, we first introduce copulas, and then present the ETPC model.

3.1. Copulas

According to Sklar's Theorem [18], any joint distribution can be expressed as:

$$F(x_1, \dots, x_p) = C(u_1, \dots, u_p), \quad (2)$$

where the function C is defined to be the copula, F_i is the marginal CDF of x_i , and $u_i = F_i(x_i)$ follows unit uniform distributions. Assuming that the partial derivatives exist, the probability density function can be written as:

$$f(x_1, \dots, x_p) = c(F_1(x_1), \dots, F_p(x_p)) \prod_{i=1}^p f_i(x_i), \quad (3)$$

where c is the copula density function [19].

We are primarily concerned with upper tail dependence in this study which is relevant in the analysis of extremes. This can be mathematically expressed as:

$$\lambda_U = \lim_{u \rightarrow 1^-} P(U_1 > u | U_2 > u) = \lim_{u \rightarrow 1^-} \frac{1 - 2u + C(u, u)}{1 - u},$$

where λ_U is the upper tail dependence coefficient, U_1 and U_2 are uniformly distributed and C is the copula defined on

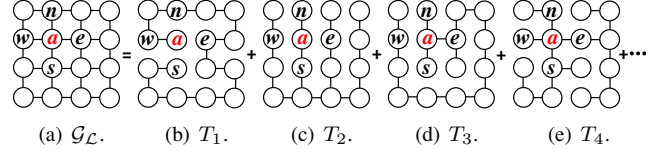


Fig. 1. ET model: the lattice (a) and spanning trees (b) - (e).

U_1 and U_2 . We consider here eight commonly used copulas: Gaussian, t , Clayton, Frank, Gumbel, Galambos, t -EV, and Hüsler-Reiss copula [19, 20, 21]. Among them, only Gaussian, Clayton and Frank copulas are not upper-tail dependent.

3.2. From Copulas to Trees and Ensemble-of-Trees

The PDF of a tree graphical model $T_i = (\mathcal{V}, \mathcal{E}_i)$, as illustrated in Fig. 1(b) - (e), can then be written as:

$$f(\mathbf{x}|T_i) = \prod_{j \in \mathcal{V}} f_j(x_j) \prod_{(j,k) \in \mathcal{E}_i} \frac{f_{jk}(x_j, x_k)}{f_j(x_j)f_k(x_k)} \quad (4)$$

$$= \prod_{j \in \mathcal{V}} f_j(x_j) \prod_{(j,k) \in \mathcal{E}_i} c_{jk}(F_j(x_j), F_k(x_k)). \quad (5)$$

The last expression follows from (3) [15]. Since trees have limited flexibility in modeling dependence, we construct the Ensemble-of-Trees of Pairwise Copulas (ETPC) model by computing the weighted sum over all possible spanning trees of Fig. 1(a) as follows:

$$f(\mathbf{x}) = \sum_{T_i} P(T_i) f(\mathbf{x}|T_i), \quad (6)$$

where $P(T_i) = 1/Z \prod_{(j,k) \in \mathcal{E}_i} \beta_{jk}$ is a decomposable prior proposed by [22]. According to the matrix-tree theorem [22], the normalizing constant $Z = \sum_{T_i=(\mathcal{V}, \mathcal{E}_i)} \prod_{(j,k) \in \mathcal{E}_i} \beta_{jk} = \det[Q(\beta)]$, where β is the edge weight matrix that is symmetric with the diagonal entries being zero and with the (j, k) entry being β_{jk} . In addition, $Q(\beta)$ denotes the first $P-1$ rows and columns of the $P \times P$ Laplacian matrix corresponding to a graph with edge weight matrix defined by β .

By substituting (5) into (6), the ET model can be succinctly formulated as [14, 15, 22]:

$$f(\mathbf{x}) = \sum_{T_i} \frac{1}{Z} \prod_{j \in \mathcal{V}} f_j(x_j) \prod_{(j,k) \in \mathcal{E}_i} \beta_{jk} c_{jk}(F_j(x_j), F_k(x_k)) \\ = \prod_{j \in \mathcal{V}} f_j(x_j) \frac{\det[Q(\beta \odot \mathbf{c})]}{\det[Q(\beta)]}, \quad (7)$$

where \odot denotes componentwise multiplication and \mathbf{c} is the copula density matrix whose (j, k) entry equals c_{jk} .

Since the parameters of the marginal density and of the copula can be estimated in advance, we only need to infer β . The resulting optimization problem can be formulated as:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} N \log \det[Q(\beta)] - \sum_{n=1}^N \log \det[Q(\beta \odot \mathbf{c}^{(n)})],$$

$$\text{s. t. } \beta_{jk} = 0 \quad \forall (j, k) \notin \mathcal{E} \quad , \quad \beta_{jk} \geq 0 \quad \forall (j, k) \in \mathcal{E}$$

$$\text{and } \|U(\beta)\|_2 = 1.$$

The expression $\|U(\beta)\|_2$ is the Euclidean norm for the upper triangular part of the matrix β . The above problem can be solved via the projected gradient method [14].

4. INFERENCE OF MISSING DATA

We now direct our attention to inference. To the best of our knowledge, no inference methods have been proposed for the ensemble-of-trees (ET) models [14, 15, 22]. Here we consider the inference problem of imputing missing data. More explicitly, we aim to infer missing values \mathbf{x}_M at a set of sites given observed data \mathbf{x}_O at other sites. A reasonable approach is the maximum a posteriori (MAP) estimation:

$$\hat{\mathbf{x}}_M = \operatorname{argmax}_{\mathbf{x}_M} f(\mathbf{x}_M | \mathbf{x}_O) \quad (8)$$

$$= \operatorname{argmax}_{\mathbf{x}_M} \prod_{\mathbf{x}_i \in \mathbf{x}_M} f_i(x_i) \cdot \det[Q(\boldsymbol{\beta} \odot \mathbf{c}(\mathbf{x}_M, \mathbf{x}_O))], \quad (9)$$

where the notation $\mathbf{c}(\mathbf{x}_M, \mathbf{x}_O)$ indicates that the copula density matrix \mathbf{c} is a function of both \mathbf{x}_M and \mathbf{x}_O . In most standard graphical models (e.g., trees), the problem can be simplified as estimating \mathbf{x}_M given the values of the respective adjacent nodes [2]. However, for the ET model, such simplification is not possible as indicated by the following theorem.

Theorem 1. *Every pair of variables (x_j, x_k) in the ensemble of trees (ET) model corresponding to a connected graph \mathcal{G} are conditionally dependent given other variables in the model.*

As such, a plausible alternative is to solve (9). Due to the determinant in the right hand side (RHS) of (9), the maximization in (9) is computationally infeasible for high-dimensional problems. By exploiting the matrix-tree theorem, we can decompose that determinant into a summation of tractable terms, reducing the computational complexity significantly.

Let us first consider the scenario where data is missing at only *one* node x_a , i.e. $\mathbf{x}_M = x_a$. Without loss of generality, we consider a lattice $\mathcal{G}_{\mathcal{L}}$ where each site is connected to its four nearest neighbors. Consequently, an unobserved node a has four observed neighbors n, s, e, w as shown in Fig. 1(a). We now consider the 5-point stencil of a , which is the subgraph that is made up of a and its four neighbors. We denote by \mathcal{S} the set of all non-empty subgraphs S_i of that 5-point stencil. Each spanning tree of $\mathcal{G}_{\mathcal{L}}$ contains one or more subgraphs S_i . The spanning trees can be clustered according to the largest subgraph S_i embedded in them; this leads to $|\mathcal{S}|$ clusters τ_{S_i} of spanning trees, each associated with a subgraph S_i , where $|\mathcal{S}|$ is the cardinality of \mathcal{S} .

As an example, the spanning trees in Fig. 1(b)-(e) are classified into four different clusters with corresponding subgraphs $\{(a, e)\}$, $\{(a, s), (a, n)\}$, $\{(a, w), (a, s), (a, n)\}$, and $\{(a, e), (a, w), (a, s), (a, n)\}$ respectively. The matrix-tree theorem allows us to decompose the determinant in (9) according to the $|\mathcal{S}|$ subgraphs S_i . Following the theorem, we can regard $\det[Q(\boldsymbol{\beta} \odot \mathbf{c}(x_a, \mathbf{x}_O))]$ as the sum of the weights of all the spanning trees in a graph with edge weight matrix $W(x_a, \mathbf{x}_O) = \boldsymbol{\beta} \odot \mathbf{c}(x_a, \mathbf{x}_O)$. As a result, if we define the weight w_{S_i} of each cluster τ_{S_i} as the sum of the weights of all spanning trees in that cluster, the determinant can be decomposed as the summation of all the cluster weights w_{S_i} . Consequently, we can write the posterior marginal $f(x_a | \mathbf{x}_O)$

as:

$$f(x_a | \mathbf{x}_O) \propto f_a(x_a) \det[Q(\boldsymbol{\beta} \odot \mathbf{c}(x_a, \mathbf{x}_O))] \quad (10)$$

$$= f_a(x_a) \sum_{i=1}^{|\mathcal{S}|} w_{S_i}(x_a, \mathbf{x}_O). \quad (11)$$

Since all the spanning trees in each cluster τ_{S_i} share the subgraph S_i , its counterpart in w_{S_i} can be factored out, i.e., the product of pairwise copulas $\prod_{(a,j) \in S_i} c_{aj}(F_a(x_a), F_j(x_j))$ corresponding to the edges (a, j) in S_i . Only this common factor varies with x_a , and the remaining part is a constant γ_{S_i} independent of x_a . In summary, the weight w_{S_i} of cluster τ_{S_i} can be factorized as:

$$w_{S_i}(x_a, \mathbf{x}_O) = \gamma_{S_i} \prod_{(a,j) \in S_i} c_{aj}(F_a(x_a), F_j(x_j)). \quad (12)$$

To determine the constant γ_{S_i} , we compute $w_{S_i}(x_a, \mathbf{x}_O)$ for an arbitrary value \hat{x}_a of x_a , leading to:

$$\gamma_{S_i} = \frac{w_{S_i}(\hat{x}_a, \mathbf{x}_O)}{\prod_{(a,j) \in S_i} c_{aj}(F_a(\hat{x}_a), F_j(x_j))}. \quad (13)$$

Once the constants γ_{S_i} have been computed, the posterior marginal $f(x_a | \mathbf{x}_O)$ can be evaluated as:

$$f(x_a | \mathbf{x}_O) \propto f_a(x_a) \sum_{i=1}^{|\mathcal{S}|} \left[\gamma_{S_i} \prod_{(a,j) \in S_i} c_{aj}(F_a(x_a), F_j(x_j)) \right], \quad (14)$$

which follows from substituting (12) in (11). We will next elaborate on the calculation of the weights $w_{S_i}(x_a, \mathbf{x}_O)$, required to compute the constants γ_{S_i} (13). Specifically, four separate cases will be considered, depending on the number of edges in subgraph S_i .

Case 1. The subgraph S_i only contains one edge, i.e., $S_i = \{(a, j)\}$ for $j = n, s, e, w$ (see Fig. 1(b)). The cluster weight equals $w_{S_i}(x_a, \mathbf{x}_O) = \det[Q(W_{S_i}(x_a, \mathbf{x}_O))]$ where the weight matrix $W_{S_i}(x_a, \mathbf{x}_O)$ is obtained from $W(x_a, \mathbf{x}_O)$ by setting all elements (a, k) and (k, a) with $k \neq j$ to zero.

Case 2. The subgraph S_i contains two edges (see Fig. 1(c)), i.e., $S_i = \{(a, j_1), (a, j_2)\}$, $j_1, j_2 \in \{n, s, e, w\}$, $j_1 \neq j_2$. Let us now consider the weight matrix $W_{S_i}(x_a, \mathbf{x}_O)$ obtained from $W(x_a, \mathbf{x}_O)$ by setting all elements (a, k) and (k, a) with $k \neq j_1, j_2$ equal to zero. The weight matrix $W_{S_i}(x_a, \mathbf{x}_O)$ corresponds to all spanning trees that do not contain any of the edges (a, k) with $k \neq j_1, j_2$. Those spanning trees may contain either (a, j_1) , (a, j_2) , or both. As we are interested in the latter case, we can therefore compute the desired cluster weight w_{S_i} as follows:

$$w_{S_i} = \det[Q(W_{S_i}(x_a, \mathbf{x}_O))] - w_{\{(a, j_1)\}} - w_{\{(a, j_2)\}}, \quad (15)$$

where $w_{\{(a, j_1)\}}$ and $w_{\{(a, j_2)\}}$ are the weights for $S_i = \{(a, j_1)\}$ and $S_i = \{(a, j_2)\}$ respectively (cf. Case 1).

Case 3. The case where S_i comprises three edges (as shown in Fig. 1(d)) is similar to Case 2, and the problem can be dealt with in an identical manner.

Case 4. When all edges are present in S_i (Fig. 1(e)), w_{S_i} can be computed by noting that the sum of all the weights w_{S_i} equals $\det[Q(W(x_a, \mathbf{x}_O))]$.

After decomposing the determinant in (9), the expression (14) depends on x_a in a more transparent manner, specifically, through products of pairwise copulas. To maximize (14) w.r.t. x_a , we employ the interior point method with four starting points, corresponding to the values of x_a that maximize the likelihood of the four pairwise copulas $c_{aj}(x_a, x_j)$ with $j = n, s, e, \text{ or } w$. Our experiments indicate that this optimization scheme typically results in the global maximum of $f(x_a|\mathbf{x}_O)$.

In the scenario when there are multiple missing sites, we employ the Iterative Conditional Mode (ICM) algorithm [24] in order to decompose the multi-site case into the single-site case. Typically, the initial estimates for missing sites are set to the average of the extreme values at the neighboring observed sites. The experimental results show that the ICM algorithm converges to a unique maximum.

5. THEORETICAL RESULTS

In this section, we outline our theoretical results on tail dependence. We defer the detailed proof to the journal version of this work. We begin with a proposition about the tail dependence between two arbitrary nodes in a tree.

Proposition 1. *Given a tree graphical model T whose joint PDF can be written as in (5), the upper tail dependence between all components of \mathbf{X} exists if all the pairwise copulas c_{ij} corresponding to the edges \mathcal{E} in the tree are upper tail dependent and X_i and X_j (variables conditionally dependent in the graph) are stochastically increasing with each other. Specifically, for any two variables X_i and X_j in the graph, the lower bound of their tail dependence coefficient is the product of the tail dependence coefficients of all the pairwise copulas corresponding to the edges in the path connecting X_i and X_j .*

We next generalize the result to the ETPC model by providing a closed form expression for the lower bound of the upper tail dependence.

Proposition 2. *The upper tail dependence between any two nodes in the Ensemble-of-Trees model is bounded below by $\det[Q(\beta \odot \lambda)] / \det[Q(\beta)]$ where $Q(X)$ takes the first $P - 1$ rows and columns of the Laplacian matrix corresponding to the graph defined by the $P \times P$ edge weight matrix X , λ is the pairwise upper tail dependence coefficient matrix, and \odot denotes componentwise multiplication.*

Therefore, by properly selecting pairwise copulas, the ETPC model can flexibly describe the dependencies between extreme values, both in a tail dependent and independent manner. Moreover, the ETPC model allows us to simulate extremal values while preserving tail properties; this is of high practical value in spatial extremes modeling (e.g. [25, 26]).

6. EXPERIMENTAL RESULTS

To assess our model, we consider the extreme precipitation in four 10×10 regular grids in South Japan, where heavy rainfall is often the cause of floods [27]. For each of the four regions, we compare two possible configurations under the

Table 1. Comparison of goodness of fit of two models.

Region No.	ETPC-mixture			ETPC-Gaussian
	tail independent	tail dependent	BIC	BIC
Region 1	2	178	-5.22×10^4	-4.76×10^4
Region 2	0	180	-8.52×10^4	-6.89×10^4
Region 3	0	180	-9.82×10^4	-8.76×10^4
Region 4	1	179	-5.98×10^4	-5.44×10^4

Table 2. The averaged MAE of imputed precipitation values with respect to the length of the square region of missing values for 5 quantiles.

Quantile	Copula Configuration	MAE (mm)					
		1×1	2×2	3×3	4×4	5×5	6×6
99 th	mixture	4.77	7.36	9.00	14.38	16.40	17.36
	Gaussian	10.76	12.53	16.47	31.92	39.29	45.89
90 th	mixture	1.69	9.93	10.50	11.47	14.61	17.96
	Gaussian	4.49	11.51	14.50	16.50	22.04	31.26
80 th	mixture	2.26	8.05	8.65	7.83	7.34	9.08
	Gaussian	3.04	6.81	7.32	8.55	11.04	12.84
70 th	mixture	2.53	3.86	5.26	7.64	6.88	8.82
	Gaussian	1.74	3.54	5.15	12.97	15.34	16.11
60 th	mixture	0.81	1.30	2.62	3.54	4.23	6.03
	Gaussian	0.52	1.54	2.33	4.67	9.20	11.47

Ensemble-of-Trees model: 1) the case when all pairwise copulas are Gaussian (ETPC model with Gaussian copula, denoted as “ETPC-Gaussian”) and 2) the case when pairwise copula selection is performed through the minimization of the Bayesian Information Criteria (BIC) (ETPC model with a mixture of copulas, denoted as “ETPC-mixture”). We compute the BIC score of the overall model, and also the mean absolute error (MAE) between the true and imputed value over all unobserved sites. For imputation purposes, we retain five samples corresponding to the 99th, 90th, 80th, 70th, and 60th quantiles of all the total rainfall amounts per event for the region as the testing data, while the remaining are treated as training data. For each testing sample, we consider 6 instances in which the data is missing in a 1×1 , 2×2 , 3×3 , 4×4 , 5×5 , and 6×6 area at the center of the 10×10 grid. We then average the imputation results over the four regions. The numerical results are listed in Table 1 and Table 2.

Apparently, for the ETPC-mixture model, almost all the chosen pairwise copulas have tail dependence. The smaller BIC score further suggests the use of tail dependent copulas for modeling spatial extremes. Additionally, the ETPC-mixture model greatly outperforms the ETPC-Gaussian model when imputing the missing data of extreme rainfall events with high quantile (90th quantile and above), probably because Gaussian copulas fail to capture the upper tail dependence in the data. Moreover, the ETPC-mixture yields a low MAE even in the case of the largest missing region (6×6), indicating that the proposed model can reliably simulate the extreme events in a spatial domain given the boundary conditions.

7. CONCLUSION

We have proposed the ETPC model for multivariate analysis of spatial extremes. Such model is equipped with tractable and efficient learning and imputation algorithms, while enjoying the attractive property of preserving the upper tail dependence under certain mild conditions.

8. REFERENCES

- [1] S. G. Coles, *An Introduction to Statistical Modeling of Extreme Values*. London: Springer, 2001.
- [2] A. T. Ihler, S. Krishner, M. Ghil, A. W. Robertson, and P. Smyth, "Graphical Models for Statistical Inference and Data Assimilation," *Physica D* **230**, pp. 72-87, 2007.
- [3] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, P. Li, and F. Kschischang, "The factor graph approach to model-based signal processing," *Proceedings of the IEEE* **95**(6), pp. 1295-1322, 2007.
- [4] D. Cooley, J. Cisewski, R. J. Erhardt, S. Jeon, E. Mannshardt, B. O. Omolo, and Y. Sun, "A survey of spatial extremes: measuring spatial dependence and modeling spatial effects," *REVSTAT* **10**, pp. 135-165, 2012.
- [5] 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.
- [6] H. Sang, and A. E. Gelfand, "Hierarchical modeling for extreme values observed over space and time," *Environmental and Ecological Statistics* **16**(3), pp. 407-426, 2009.
- [7] P. J. Northrop, and P. Jonathan, "Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights," *Environmetrics* **22**(7), pp. 799-809, 2011.
- [8] H. Sang, and A. E. Gelfand, "Continuous Spatial Process Models for Spatial Extreme Values," *Journal of Agricultural, Biological and Environmental Statistics* **15**, pp. 49-65, 2010.
- [9] H. Yu, Z. Choo, W. I. T. Uy, J. Dauwels, and P. Jonathan, "Modeling Extreme Events in Spatial Domain by Copula Graphical Models," *Proc. FUSION*, pp. 1761- 1768, 2012.
- [10] A. C. Davison, and M. M. Gholamrezaee, "Geostatistics of extremes," *Proc. R. Soc. Lond. Ser. A* **468**, pp. 581-08, 2012.
- [11] A. C. Davison, S. Padoan, and M. Ribatet, "Statistical modeling of spatial extremes," *Statist. Sci.* **27** (2):161-186, 2012.
- [12] C. Varin, N. Reid, and D. Firth, "An overview of composite likelihood methods," *Statistica Sinica* **21**(1):5-42, 2011.
- [13] K. Aas, C. Czado, A. Frigessi, and H. Bakken, "Pair-copula constructions of multiple dependence," *Insurance: Mathematics and Economics* **44**, pp. 182-198, 2009.
- [14] Y. Lin, S. Zhu, D. D. Lee, and B. Taskar, "Learning Sparse Markov Network Structure via Ensemble-of-Trees Models," *12th International Conference on Artificial Intelligence and Statistics.*, 2009.
- [15] S. Kirshner, "Learning with tree-averaged densities and distributions," *Advances in Neural Information Processing Systems (NIPS)*, 2008.
- [16] A. W. Ledford, and J. A. Tawn, "Statistics for near independence in multivariate extreme values," *Biometrika* **83**(1), pp. 169-187, 1996.
- [17] G. Draisma, H. Drees, A. Ferreira, and L. de Haan, "Bivariate tail estimation: dependence in asymptotic independence," *Bernoulli* **10**, pp. 251-280, 2004.
- [18] A. Sklar, "Fonctions de répartition à n dimensions et leurs marges," *Publications de l'Institut de Statistique de L'Université de Paris* **8**, pp. 229-231, 1959.
- [19] P. K. Trivedi, and D. M. Zimmer, *Copula Modeling: An Introduction for Practitioners*. Hanover, MA, USA: now Publishers Inc, 2007.
- [20] S. Demarta, and A. J. McNeil, "The t copula and related copulas," *International Statistical Review* **73**(1), pp. 111-129, 2005.
- [21] G. Gudendorf, and J. Segers, "Extreme-value copulas," In W. H. P. Jaworski, F. Durante and T. Rychlik (Eds.), *Proceedings of the Workshop on Copula Theory and its Applications*. Springer, pp. 127-146, 2010.
- [22] M. Meilä, and T. Jaakkola, "Tractable Bayesian learning of tree belief networks," *Statistics and Computing* **16**(1): 77-92, 2006.
- [23] D. P. Bertsekas, *Nonlinear programming*, Athena Scientific., 2003.
- [24] J. Besag, "On the Statistical Analysis of Dirty Pictures," *Journal of the Royal Statistics Society B* **48**(3) , pp. 259-302, 1986.
- [25] J. E. Heffernan, and J. A. Tawn, "A conditional approach for multivariate extreme value," *J.R. Statist. Soc. B* **66**, pp. 1-34, 2004.
- [26] C. Keef, J. Tawn, and C. Svensson, "Spatial risk assessment for extreme river flows," *J.R. Statist. Soc. C* **58**(5), pp. 601-618, 2009.
- [27] K. Kamiguchi, O. Arakawa, A. Kitoh, A. Yatagai, A. Hamada, and N. Yasutomi, "Development of APHRO.JP, the first Japanese high-resolution daily precipitation product for more than 100 years," *Hydrological Research Letters* **4**, pp. 60-64, 2010.