Spatial effects are often separated into large scale (i.e., regional) effects and small scale (i.e., local) effects.

- In terms of a statistical model, regional spatial effects are often captured by characterizing how the marginal distribution varies over a study region.
- Local spatial effects are typically described by a dependence structure.

Given data, the distinction between the local and regional effects is likely not obvious. Can view the task of separating these two effects as analogous to the task of decomposing time series data into mean (trend and seasonal effects) and a stationary noise process described by a covariance structure.
Spatial data are necessarily multivariate as they are recorded at multiple locations. Throughout, we will assume that we analyze only one quantity (e.g., rainfall) at multiple sites.

Spatially analyzing multiple quantities is a possible extension of the work surveyed here.
The notion of max-stability forms the foundation of extreme value theory. Assume there exist normalizing sequences \( \{a_n\} \) and \( \{b_n\} \) such that

\[
\Pr \left( \frac{M_n - b_n}{a_n} \leq y \right) \rightarrow G(y), \quad \text{as } n \rightarrow \infty
\]

where \( G \) is non-degenerate. Then \( G \) belongs to the class of multivariate max-stable (equivalently, extreme value) distributions. Denote by \( Y = (Y_1, \ldots, Y_d)^T \) a max-stable random vector; that is, \( a_n^{-1}(M_n - b_n) \overset{d}{\rightarrow} Y \).
Unlike in the univariate case, no fully parametric representation exists for the multivariate max-stable distributions. The univariate marginal distributions must be univariate max-stable and therefore can be described by the generalized extreme value (GEV) distribution:

$$Pr(Y_j \leq y) = G_j(y) = \exp \left\{ - \left[ \left( 1 + \xi_j \frac{y - \mu_j}{\sigma_j} \right)^{-1/\xi_j} \right] \right\},$$

for $j = 1, \ldots, d$. Here, $\mu_j, \sigma_j$, and $\xi_j$ are the location, scale and shape parameters for the $j$-th component’s marginal. Representations for multivariate max-stable distributions assume the marginals have a common (convenient) max-stable distribution.
A fundamental construct for spatial extremes is the max-stable process, which is the infinite-dimensional analogue to a max-stable random vector. If for all $s \in S$ there exist renormalizing sequences $a_n(s)$ and $b_n(s)$ such that

$$a_n^{-1}(s) \left( \max_{i=1,\ldots,n} X_i(s) - b_n(s) \right) \xrightarrow{d} Y(s)$$

which has a non-degenerate distribution, then $Y(s)$ is a max-stable process. When the max-stable process has unit Fréchet margins, we will denote it by $Z(s)$. 
Fréchet Transformation

Assume $Z = (Z_1, \ldots, Z_d)^T$ has a multivariate max-stable distribution with unit-Fréchet marginals: $Pr(Z_j \leq z) = \exp(-z^{-1})$. Then

$$Pr(Z \leq z) = G^*(z) = \exp\{-V(z)\}, \text{ where}$$

$$V(z) = d \cdot \int_{\Delta_d} \bigvee_{j=1}^d \frac{w_j}{z_j} H(dw).$$

Here $\Delta_d = \{w \in \mathbb{R}_+^d \mid w_1 + \ldots + w_d = 1\}$ is the $(d - 1)$-dimensional simplex, and the angular (or spectral) measure $H$ is a probability measure on $\Delta_d$, which determines the dependence structure of the random vector. Due to the common marginals, $H$ has the moment conditions $\int_{\Delta_d} w_j H(dw) = 1/d$ for $j = 1, \ldots, d$. 

Cooley, Cisewski, Erhardt, Jeon, Mannshardt, Omolo, Sun

A Survey of Spatial Extremes
Fréchet Transformation

There is no loss of generality in assuming the multivariate max-stable distribution has unit-Fréchet margins as Resnick87 states that the domain-of-attraction condition is preserved under monotone transformations of the marginal distributions.

If \( a_n^{-1}(M_n - b_n) \xrightarrow{d} Y \) which does not have unit-Fréchet marginals, one can define marginal transformations

\[
T_j(x) = \left(1 + \xi_j \frac{x - \mu_j}{\sigma_j}\right)^{-\xi_j}, \quad j = 1, \ldots, d
\]

and define \( G^*(z_1, \ldots, z_d) = G(T_1^{-1}(z_1), \ldots, T_d^{-1}(z_d)) \), where \( T_j^{-1} \) is the inverse function of \( T_j \), for \( j = 1, \ldots, d \).

This approach of transforming to convenient marginals is similar to copula approaches, albeit with a marginal suggested by extreme value theory rather than Uniform \([0,1]\).
Max-Stable Approach

Asymptotic theory suggests the following general statistical methodology referred to as the block maxima approach.

Choose $n$ to be a fixed block size which is large enough such that the asymptotic theory holds approximately, and assume a sequence of i.i.d $X_i$, $i = 1, \ldots, nm$ are observed, where $m$ denotes the number of blocks.

Define $M_k = (\bigvee_{i=(k-1)n+1}^{kn} X_{i,1}, \ldots, \bigvee_{i=(k-1)n+1}^{kn} X_{i,d})^T$ for $k = 1, \ldots, m$ (note that the dependence on $n$ in the notation $M_k$ has been suppressed), and fit a multivariate max-stable distribution to the $M_k$.

It is important to note that $M_k$ will not appear in the observation record unless the occurrence times of each element’s block maximum coincide.
Max-Stable Approach

- Using representation (3) to fit a multivariate max-stable distribution requires that the marginals be unit Fréchet.
- Although transforming the marginals is a simple theoretical procedure, in practice, the marginal distributions must be estimated.
- Subsequently, utilizing (3) to perform a multivariate analysis of extremes involves two tasks: (1) estimating the marginals and, (2) characterizing the dependence via a model for $V(z)$ or $H(w)$.
- Tasks (1) and (2) seem sequential; however, we note that inference can be performed all-at-once either in the frequentist (Padoan 2010) or Bayesian (Ribatet 2011) settings.
For spatial extremes studies, the aforementioned regional and local spatial effects each can be associated with one of the above tasks.

- Most study regions are large enough that the marginal distribution of the studied quantity will vary over the region. Thus, in order to transform to a common marginal, one must first account for how the distribution’s tail varies by location.

- The local spatial effect is related to the spatial extent of individual extreme events and the resulting dependence in the data due to multiple sites being affected by the same event. In terms of (3), this dependence is captured by $V(z)$ or $H(w)$.

- We will refer to the dependence remaining after the marginal standardization as ‘residual’ dependence, as Sang and Gelfand (2010) termed the data after marginal transformation as ‘standardized residuals’.
Climate can be thought of as the distribution of weather (Guttorp, 2011), and climate varies with location. Performing the marginal transformation is akin to standardizing the climate across the study region.

Weather events have a spatial extent which is best captured by a (local) stochastic representation.
Fundamental differences between geostatistics and spatial extremes:

- Based on first and second moments, geostatistics focuses on central tendencies, not on the distribution’s tail.

- The Gaussian framework in traditional geostatistics analysis is incorrect for data that are maxima, as the Gaussian distribution is not max-stable.

- Dependence in extremes is described via the exponent measure function $V(z)$ or angular measure $H(w)$ which cannot be linked to covariance.

- Much of classical geostatistics is applied to situations where one has only one realization of the process $X(s)$, observed at multiple locations. To perform an EVA, it is necessary that multiple realizations $X_i(s)$ underlie the subset of extreme data.
Tail Dependence

To completely characterize the dependence among the components of a max-stable random vector requires one to specify the angular measure $H(w)$ or the exponent measure function $V(z)$, specification of which can be arduous, especially as the dimension $d$ grows.

It is useful to have summary measures of tail dependence, and several metrics have been developed which aim to summarize the amount of tail dependence in one number.

Tail dependence falls into two distinct categories: asymptotic dependence and asymptotic independence.
Let $Y$ be a $d$-dimensional max-stable random variable with common margins. The $d$-dimensional extremal coefficient $\theta_d$ can be implicitly defined as

$$Pr(Y_1 \leq y, \ldots, Y_d \leq y) = Pr^{\theta_d}(Y_1 \leq y),$$

The value $\theta_d$ can be thought of as the effective number of independent random variables in the $d$-dimensional random vector. The coefficient takes values between 1 and $d$, with a value of 1 corresponding to complete dependence among the locations, and a value of $d$ corresponding to complete independence. Although higher-order extremal coefficients are sometimes useful (Erhardt and Smith, 2011), $\theta_2$ is most widely used as it conveys the amount of dependence between a pair of components.
The madogram, a first-order semivariogram, (Matheron, 1987), requires the first-moment to be finite which is not always the case in extremes studies, Cooley, Naveau, and Poncet (2006) proposed the F-madogram, which first transforms the random variable by applying the cdf and is finite for any distribution. If $Y(s)$ is a stationary and isotropic max-stable process with marginal distribution $G$, then the F-madogram is:

$$\nu(h) = \frac{1}{2} \mathbb{E} \left| G(Y(s)) - G(Y(s + h)) \right|$$

(5)

The F-madogram’s values range from 0 to 1/6, which corresponds to complete dependence and independence respectively.
Extremal coefficient (left panel) and the F-Madogram (right panel) with unit Gumbel margins for the Schlather model with Whittle-Matérn correlation functions. The red lines are the theoretical extremal coefficient and F-madogram, gray points are pairwise estimates, and black crosses are binned estimates.
Asymptotic Independence

Two components $X_1$ and $X_2$ from a random vector $\mathbf{X}$ with common marginals are *asymptotically independent* if

$$\lim_{x \to x^*} \Pr(X_2 > x | X_1 > x) = 0,$$

where $x^*$ is the upper endpoint of the common marginal distribution.

- Asymptotic independence does not imply (complete) independence and, in fact, can mask relevant dependencies.
- Summarizing tail dependence with the extremal coefficient or madogram, or modeling tail dependence via $H(w)$ or $V(z)$, will ignore any dependence in an asymptotically independent couple.
- Models for and measures of tail dependence for spatial extremes have thus far been limited to the asymptotically dependent case. However, an understanding of the concept of asymptotic independence is essential to fully understanding the limitations of summary dependence metrics.
Ledford (1996) identified the problems arising with the existing modeling methodology in the case of asymptotic independence and proposed a new parameter $\eta$, the coefficient of tail dependence. If $(Z_1, Z_2)$ has unit Fréchet marginals, Ledford (1996) assume a joint survival function $\bar{F}$ satisfying

$$\bar{F}(z, z) = \Pr(Z_1 > z, Z_2 > z) \sim \mathcal{L}(z)z^{-1/\eta} \quad \text{as} \; z \to \infty,$$

where $\mathcal{L}$ is a slowly varying function ($\frac{\mathcal{L}(tz)}{\mathcal{L}(z)} \to 1$ as $z \to \infty$) and $\eta \in (0, 1]$. The coefficient of tail dependence $\eta$ is used to quantify the tail dependence in the asymptotic independent setting. $\eta = 1$ implies asymptotic dependence and $\eta < 1$ measures the degree of dependence.
Marginal Behavior

The representation of multivariate max-stable distributions as well as the spatial dependence models assume that the univariate marginal distributions are known and common at all locations in the study region. In most applications, the study regions are large enough that the assumption of a common marginal distribution is unrealistic and the marginal distribution is not known. Therefore, it becomes essential for one to model the marginal distribution at all locations within the study region.
A simplistic approach would be to individually estimate the marginal distributions at each location. This approach has been used in (non-spatial) multivariate applications (Heffernan, 2004; Cooley, 2010). Such an approach is less-than-ideal for spatial applications for a number of reasons:

▶ First, a goal of many spatial projects is to make inference at locations where there are no data, i.e., to perform spatial interpolation. Constructing unconnected models at each location does not allow one to readily interpolate.

▶ Second, there is a desire to borrow strength across locations when estimating marginal parameters.
Many spatial data have a temporal record of several decades. Such a data record is more-than-enough to pin down the central tendencies of the marginal distribution, but large uncertainties remain about tail quantities. It is well-documented that tail quantities, and in particular point estimates for the tail index $\xi$, can vary wildly over the spatial region when estimated individually (e.g., Cooley, 2010). Methods which borrow strength across locations ‘trade space for time’ and help to reduce uncertainties.
Regression approaches are frequently used to model the mean process $\alpha(s)$ of a geostatistical model. Similarly, regression approaches have been used to model the parameters of the extreme value distributions. When modeling annual maximum precipitation in the southeast United States, Padoan (2010) select a model in which the location parameter $\mu(s)$ and scale parameter $\sigma(s)$ are linear functions of latitude and elevation, and the shape parameter $\xi(s)$ is constant over the study region.
Generally, one wishes to employ covariates which are known at all locations $s \in S$ (both observed and unobserved), and often spatial coordinates are the only such covariates.

In geostatistics, regression models on spatial coordinates are known as trend surface models (Schabenberger, 2005).

However simple regression models on available covariates are sometimes unable to fully capture complex spatial behavior.
Hierarchical (or multi-level) models have been extensively used in describing the relationship between observations and the complex processes that generate them.

- For many hierarchical models, the data collected is not well-suited for modeling within the usual Gaussian framework of geostatistics.
- To explain the spatial variation of the data, hierarchical models typically assume that the behavior of the data over the study region is driven by an unobserved or latent process.
Hierarchical models are often devised within a Bayesian framework, and typically have three levels:

(i) the data level
(ii) the process level
(iii) the prior level.
Hierarchical Modeling

Let the vector of parameters, $\psi$, be defined as $\psi = (\psi_1, \psi_2)$, where $\psi_1$ are the parameters in the data level, and $\psi_2$ are the parameters in the process level. Then, the posterior distribution of $\psi$ given the data $y$, $\pi(\psi|y)$, is given by

$$\pi(\psi|y) \propto \pi(y|\psi_1)\pi(\psi_1|\psi_2)\pi(\psi_2).$$

(7)

Here $\pi(y|\psi_1)$ defines the likelihood function, $\pi(\psi_1|\psi_2)$ describes the distribution of the process and $\pi(\psi_2)$ the (hyper) priors. When applied to extremes data, the likelihood is based on an EVD. Spatial modeling at the process level is designed to borrow strength across locations and flexibly capture spatial variation, showing how the marginal parameters of the EVD vary over the study region.
Hierarchical Modeling

In a spatial hierarchical model, the likelihood must account for the fact that the data are observed at multiple locations. A simple approach:

- Assume that the data at different locations are conditionally independent given the parameters \( \psi_1 \), which themselves are spatially dependent from the process level of the model.
- the likelihood becomes a product of the individual likelihoods at each location.
- This conditional independence assumption is widely made in hierarchical modeling, and is quite sensible in most epidemiological settings where disease counts at different locations can be assumed to be independent once the latent risk level is accounted for.
However, the conditional independence assumption is questionable when modeling weather data because individual events can affect more than one location.

- A hierarchical model which assumes conditional independence in the likelihood ignores any residual dependence which remains after accounting for marginal effects.
- Despite the aforementioned limitation of assuming conditional independence, there are several applications of BHMs in the spatial extremes literature whose primary aim is to characterize the marginal tail behavior and which make this assumption.
The definition of a max-stable process (Equation 2) as the infinite dimensional generalization of a multivariate max-stable distribution gives a well-defined class of models, but it does not suggest how to construct such processes. A conceptual construction of max-stable processes was first given with a spectral representation of extremal processes by de Haan (1984) and de Haan (2006).
Max-Stable Processes

Given a stochastic process \( \{X(s)\} \) and a Poisson process \( \Pi \) with intensity \( d\zeta/\zeta^2 \) on \((0, \infty)\). Let \( \{X_i(s)\}_{i \in \mathbb{N}} \) be independent realizations of a process \( X(s) \) with \( \mathbb{E}[X(s)] = 1 \), and let \( \zeta_i \in \Pi \) be points of the Poisson process. Then

\[
Z(s) = \max_{i \geq 1} \zeta_i X_i(s), \quad s \in S.
\]

is a max-stable process with unit-Fréchet margins and the distribution function is determined by

\[
\Pr(Z(s) \leq z(s), s \in S) = \exp \left( - \mathbb{E} \left[ \sup_{s \in S} \left\{ \frac{X(s)}{z(s)} \right\} \right] \right),
\]

where the negative argument of the exponential is the infinite-dimensional analogue to \( V \).
Max-Stable Processes - Models

Different choices of the process $X_i(s)$ lead to different classes of the max-stable process. Although Gaussian distributions and processes are not well-suited for modeling extremes, Gaussians can be used within the above max-stable construction to produce appropriate models.

Realization of Gaussian extreme value process (top left panel), extremal Gaussian process (top right panel), and Brown-Resnick process (bottom panel) from SpatialExtremes R package.
A barrier to fitting max-stable processes to data is that closed-form expressions for the joint likelihood can only be written out in low dimensional settings. This means that if the data are observed at \( d > 2 \) locations in space, the joint likelihood cannot be written in closed form. Padoan (2010) proceeded with a likelihood-based approach to fitting max-stable processes by substituting a composite likelihood for the unavailable joint likelihood. We first introduce composite likelihoods, then show the connection to max-stable processes.
If $f(z; \psi)$ is a statistical model and we have a set of marginal or conditional events $\{A_j \subseteq \mathcal{F}, j = 1, \ldots, J\}$ where $\mathcal{F}$ is a sigma algebra on $\mathcal{Z}$ and each event $A_j$ has associated likelihood $\mathcal{L}_j(\psi; z) \propto f(z \in A_j; \psi)$, then a composite log-likelihood is a weighted sum of log-likelihoods for each event.

$$\ell_C(\psi; z) = \sum_j w_j \cdot \log f(z \in A_j; \psi),$$

where $w_j$ is a weight function on $j$th event. If the weights are all equal they may be ignored, though non-equal weights may be used to improve the statistical performance in certain cases.
One example of a composite log-likelihood is the pairwise log-likelihood, defined (in a spatial application) as

$$
\ell_C(\psi; z) = \sum_{k=1}^{m} \sum_{j=1}^{d-1} \sum_{j'=j+1}^{d} \log f(z_k(s_j), z_k(s_{j'}); \psi),
$$

where each term $f(z_k(s_j), z_k(s_{j'}); \psi)$ is a bivariate marginal density function based on observations at locations $j$ and $j'$, and $\psi$ is a spatial dependence parameter.

Padoan et al (2010) used the composite likelihood to model the joint spatial dependence of extremes and accounted for regional effects with a regression model on the GEV parameters.
Copulas provide another framework for representing the dependence structure of a multivariate distribution with known marginals. Copulas are multivariate distributions with standard uniform marginal distributions, and they characterize the dependence structure of a multivariate distribution from univariate marginal distributions by defining a joining mechanism (Nelsen, 2006, Joe, 1997).
Copula Approaches

Given a $d$-dimensional random vector $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_d)^T$ with corresponding marginal cdfs $F_j$ for $j = 1, \ldots, d$ and joint distribution function $F$, a copula is a function, $C : [0, 1]^d \rightarrow [0, 1]$ such that

$$F(\mathbf{Y}) = C(F_1(Y_1), \ldots, F_d(Y_d)).$$

(8)

If the marginal cdfs of $\mathbf{Y}$ are all continuous, then the copula function $C$ is uniquely defined by (8). Conversely, for a copula $C$ and continuous margins $F_1, \ldots, F_d$, the copula $C$ corresponds to the distribution of $F_1(Y_1), \ldots, F_d(Y_d)$, i.e.,

$$C(u_1, \ldots, u_d) = F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)).$$

(9)
The copula framework has appeal for modeling multivariate extremes. If working with multivariate block maximum data, extreme value theory suggests that each marginal should be approximately GEV distributed. Equation (9) suggests one can combine knowledge of the marginal distributions with a copula model to obtain a valid cdf. Further, (8) says that one can obtain a copula model from any multivariate distribution.
While this approach allows one great flexibility to create multivariate distributions with GEV marginals, these distributions will not correspond to a MEVD as characterized in above unless one uses an extremal copula model (Joe, 1997), which essentially correspond to the existing parametric MEVD subfamilies. Use of (nonextremal) copula models to describe extremes has been controversial.
Copula Approaches

For spatial data which are typically observed at many locations, one would need a copula which can handle high dimensions, and further, whose pairwise dependence can be linked to distance. Two obvious choices are to use the multivariate Gaussian or multivariate Student $t$ distributions to generate a copula.
The study of spatial extremes is young, as evidenced by the fact that the majority of the work referenced in this survey article has been done in the last 10 years, and nearly all of it in the last 20 years. The study has a well-developed foundation in probability theory, and the field has made great strides in developing realistic models based on this theory.

However, there are limitations in practice - e.g. one cannot "throw climate model output into max-stable methods". Dan Cooley?
Future Work

Further development of methods and models for spatial extremes is anticipated, and it is imagined that this work will follow similar themes to recent work in geostatistics. It remains a challenge to fit max-stable process models to data recorded at many locations, for example climate model output which has thousands of locations.

Available software for analyzing spatial extremes data:

- R package **SpatialExtremes** - can be used both to estimate spatial dependence and to fit max-stable process models.
- **RandomFields** package - is useful for simulating max-stable processes.


