Drawing inference without directly relevant data

Christian Genest

McGill University

CRM-Fields-PIMS Prize Lecture

April 20, 2023

First, credit where credit is due

This prize crowns 40 years of efforts in research.

Over 120 collaborators contributed to it, most notably

- ✓ Philippe Capéraà
- ✓ Johanna Nešlehová
- ✓ Bruno Rémillard
- ✓ Louis-Paul Rivest

who were major sources of inspiration in research. I am also very grateful for critical guidance provided at various stages of my career (and life) by

- ✓ Jim Zidek
- ✓ Mark Schervish
- ✓ Jack Kalbfleisch
- ✓ Jerry Lawless

- ✓ Luc Genest
- ✓ Lucie Lapointe

Finally, thanks to my family and children for their support and love.

Purpose and outline of this talk

To give an overview of some of the problems on which I worked, emphasizing their mathematical nature.

In hindsight, many modeling and inference issues I addressed where characterized by the absence of directly relevant data:

- ✓ expert use;
- ✓ dependence modeling;
- ✓ risk assessment.

This talk is meant to be accessible to undergraduate math students with minimal knowledge of statistics (apologies to the professionals).

A bit of background

Here are three basic concepts underlying much statistical thinking:

- ✓ stochastic model;
- ✓ parametric inference;
- ✓ Bayesian updating.

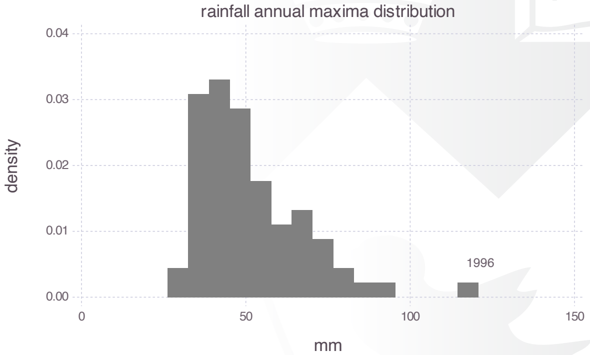
Picture on the right:
Saguenay flood
July 19-21, 1996

275 mm of rain in 48 hours

Here is a simple illustration having to do with extreme weather.

Side note: This "petite maison blanche," which survived the flood and was turned into a museum, belonged to Alyre Genest, my father's uncle.

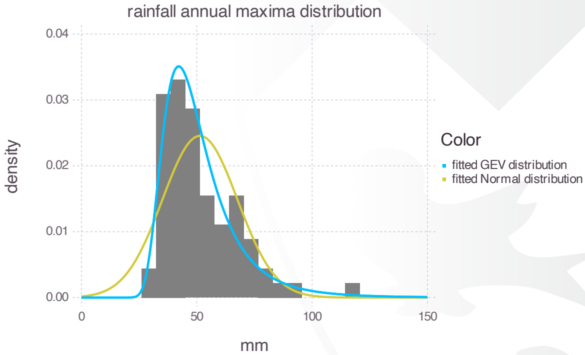
Source : http://montreal.weatherstats.ca/



Histogram of the rainiest day of the year at Trudeau International Airport between 1943 and 2017.

A model is a distribution for the data

A stochastic model is an "idealized histogram" that captures the key features of the data and makes it possible to evaluate risks and make predictions beyond what has been observed.



The yellow and blue curves are two different models for these data.

Mathematical formulation

Consider data $X_1, X_2, \ldots$ from distribution F and density $f = F'$.

The yellow (Gaussian) curve is given, for all $x \in \mathbb{R}$, by

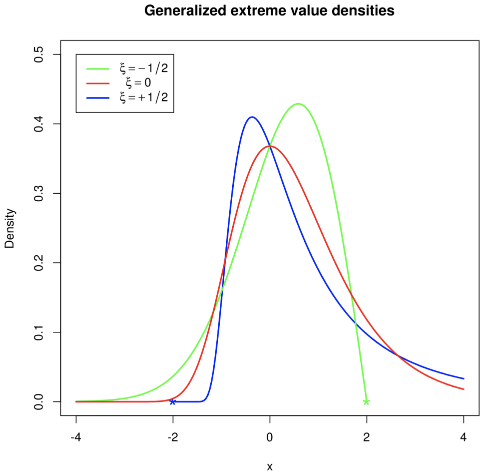
$$f(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}.$$

The blue (generalized extreme-value) curve is given, for all $x \in \mathbb{R}$, by

$$F(x \mid \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\},$$

where $a_+ = \max(0, a)$. These models involve parameters that must be selected, viz. $\mu \in \mathbb{R}$ (location), scale $\sigma > 0$ (scale), and $\xi \in \mathbb{R}^*$ (shape).

GEV model justified by the Fisher–Tippett theorem



All with $\mu = 0$, $\sigma = 1$. Asterisks mark support-endpoints

Inference

Parameters must be selected from the data in an optimal way, e.g., by matching moments or maximizing the likelihood function, viz.

$$\ell(\theta) = \ell_{x_1,\ldots,x_d}(\theta) = \prod_{i=1}^{n} f_\theta(x_i),$$

i.e., the joint distribution of the data viewed as a function of the parameter (vector) θ, e.g., $\theta = (\mu, \sigma)$ or $\theta = (\mu, \sigma, \xi)$.

In the Gaussian case, both approaches lead to

$$\hat{\mu} = \bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$$

but two (slightly) different formulas for $\hat{\sigma}^2$.

Note that estimators are random variables and are studied as such; their properties are an important subject of inquiry in Mathematical Statistics.

What can be done with this?

For the Trudeau Airport annual daily maximum, the MLEs are

Normal	GEV (Fréchet)
$\hat{\mu} = 51.36$	$\hat{\mu} = 43.56$
$\hat{\sigma} = 16.25$	$\hat{\sigma} = 10.58$
	$\hat{\xi} = 0.14$

Depending on the choice of model, one can then compute things such as:

- ✓ the probability of a 1996-like event, viz.

  $\Pr(X \geq 120.6) \approx 0.68\%$ versus 0.001%;

- ✓ the corresponding return period:

  once in 148 years versus 97,714 years!

Bayesian viewpoint

Encode the analyst's prior information concerning the unknown parameters in a distribution, e.g., a prior for the mean θ is

$$\pi(\theta) \equiv \mathcal{N}(\mu_0, \sigma_0^2).$$

One can update this information using Bayes' rule, viz.

$$\pi(\theta \mid X_1 = x_1, \ldots, X_n = x_n) \propto \pi(\theta)\ell(\theta).$$

For example if the data are Gaussian and σ is known, the posterior for θ is Gaussian with respective mean and precision $(= 1/\text{variance})$

$$\tau^2 \left(\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{x}_n}{\sigma^2} \right), \quad \frac{1}{\tau^2} = \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\,.$$

Easy Bayesian learning example

Consider the event A: a non-hemophiliac woman carryies the (F8) gene for this blood disorder.

Suppose $\Pr(A) = p$ a priori. How does p change with information through the hemophiliac status $X_1, \ldots, X_d$ of her sons?

$$\Pr(X_1 = \cdots = X_d = 0 \mid A) = 1/2^d,$$
$$\Pr(X_1 = \cdots = X_d = 0 \mid \bar{A}) = 1.$$

By Bayes' rule,

$$\Pr(A \mid X_1 = \cdots = X_d = 0) = \frac{p \times 2^{-d}}{p \times 2^{-d} + (1-p) \times 1},$$

which tends to 0 as $d \to \infty$ but jumps to 1 as soon as some $X_i = 1$.

VIGNETTE 1

Expert Use

UBC, Vancouver; CMU, Pittsburgh

Context

There are sometimes no data to inform risk management (e.g., chances of coastal flood in remote locations).

One way to carry out inference, risk management, and prediction, is to consult $d \geq 1$ experts, whose opinions may differ.

To make things simple, assume $\theta \in \mathbb{R}$ and the expert priors are densities

$$\pi_1, \ldots, \pi_d.$$

To benefit from all, one can combine them into a single distribution, viz.

$$(\pi_1, \ldots, \pi_d) \mapsto T(\pi_1, \ldots, \pi_d).$$

The map T is called a pooling operator.

Examples of pooling operators

The linear opinion pool (Stone, 1961):

$$T_{\mathrm{lin}}(\pi_1, \ldots, \pi_d) = w_1\pi_1 + \cdots + w_d\pi_d.$$

The logarithmic opinion pool (Madansky, 1964):

$$T_{\mathrm{log}}(\pi_1, \ldots, \pi_d) \propto \pi_1^{w_1} \times \cdots \times \pi_d^{w_d},$$

In both formulas, $w_1, \ldots, w_d \in [0, 1]$ are expert weights with

$$w_1 + \cdots + w_d = 1.$$

If $\pi_1 \sim \mathcal{N}(\theta_0, \sigma_1^2), \ldots, \pi_d \sim \mathcal{N}(\theta_0, \sigma_d^2)$, then T_{lin} is a mixture of Gaussians while T_{log} is a Gaussian distribution with precision

$$\frac{w_1}{\sigma_1^2} + \cdots + \frac{w_d}{\sigma_d^2}.$$

External Bayesianity

Now suppose data become available, leading to a likelihood $\ell(\theta)$.

Two options are then possible:

(a) Ask the experts to update their opinions and then pool, leading to

$$T\left(\frac{\pi_1}{\int \pi_1 d\nu}, \ldots, \frac{\pi_d}{\int \pi_d d\nu}\right).$$

(b) Pool first and then update yourself, leading to

$$T^*(\pi_1, \ldots, \pi_d) = \frac{\ell \times T(\pi_1, \ldots, \pi_d)}{\int \ell \times T(\pi_1, \ldots, \pi_d) d\nu}.$$

A pooling operator for which $T = T^*$ is called externally Bayesian (eB).

A characterization of eB pooling operators

Suppose that a pooling operator satisfies the local condition

$$T(\pi_1, \ldots, \pi_d)(\theta) = G\{\pi_1(\theta), \ldots, \pi_d(\theta)\} \quad \nu\text{-a.e.}$$

Then $T = T_{\log}$ under minimal conditions (Genest, 1984, *AoS*).

This result was later extended in various ways, e.g., Genest et al. (1986, *AoS*) show that if G can depend on θ, then

$$T(\pi_1, \ldots, \pi_d)(\theta) = \frac{g(\theta) \times \pi_1^{w_1}(\theta) \times \cdots \times \pi_d^{w_d}(\theta)}{\int g(\theta) \times \pi_1^{w_1}(\theta) \times \cdots \times \pi_d^{w_d}(\theta) d\nu(\theta)}$$

for some map g acting effectively as a (previously observed) likelihood.

A Bayesian view of linear opinion pools

Another line that Mark Schervish (Carnegie-Mellon) and I pursued is

"When does the linear opinion pool result from Bayes' rule?"

For simplicity, consider an event A (like hemophiliac status) and a vector $\mathbf{X} = (X_1, \ldots, X_d)$ of expert opinions, Bayes' rule yields

$$\Pr(A \mid \mathbf{X} = \mathbf{x}) = \Pr(A) \frac{\Pr(\mathbf{X} = \mathbf{x} \mid A)}{\Pr(\mathbf{X} = \mathbf{x})} .$$

Suppose that a decision maker (DM) does not wish (or cannot) specify the conditional distribution of $\mathbf{X}$ given A or its complement $\bar{A}$.

Result

Suppose that the DM merely specifies

$$\mathrm{E}(\mathbf{X}) = (\mu_1, \ldots, \mu_d)$$

and seeks a pooling operator T such that whatever the marginal law F of $\mathbf{X}$, there exists a compatible joint distribution for A and $\mathbf{X}$ such that

$$T(x_1, \ldots, x_d) = \Pr(A \mid \mathbf{X} = \mathbf{x}).$$

Genest & Schervish (1985, *AoS*) then show that

$$T(x_1, \ldots, x_d) = w_1(x_1 - \mu_1) + \cdots + w_d(x_d - \mu_d)$$

with $w_1, \ldots, w_d$ satisfying 2^{d+1} inequalities. A converse also holds.

A personal turning point

As it turns out, denoting $\mathbf{w} = (w_1, \ldots, w_d)^\top$, one finds

$$\mathrm{E}\{(\mathbf{1}_A - p)(\mathbf{X} - \mu)\} = \Sigma \mathbf{w} \quad \Rightarrow \quad \mathbf{w} = \Lambda^{-1} \sigma_d,$$

where $p = \Pr(A)$ and Σ is a $(d+1) \times (d+1)$ partitioned matrix, viz.

$$\mathrm{cov}(\mathbf{X}, \mathbf{1}_A) = \Sigma = \begin{pmatrix} \Lambda & \sigma_d \\ \sigma_d^\top & p(1-p) \end{pmatrix}.$$

Therefore, for each $i \in \{1, \ldots, d\}$, weight w_i should be a measure of additional information provided by expert i, over and above everyone else [regression analogy].

This provides a principle for selecting the weights, but beyond that it was a turning point for me because, as I moved to Waterloo, I realized that

dependence between experts should really be modeled!

VIGNETTE 2

Dependence Modeling

UW, Waterloo; U. Laval, Québec

Fixed margins and copulas

Given distribution functions $F_1, \ldots, F_d$, how do you construct a multivariate distribution H with these margins?

Alternatively, you want to build a random vector $(X_1, \ldots, X_d)$ such that

$$\forall_{x_1,\ldots,x_d \in \mathbb{R}} \quad H(x_1, \ldots, x_d) = \Pr(X_1 \le x_1, \ldots, X_d \le x_d)$$

and, for all $i \in \{1, \ldots, d\}$,

$$\Pr(X_i \le x_i) = F_i(x_i).$$

In the case $d = 2$, an analog would be to fill a $K \times L$ frequency table whose row and column sums are fixed! [Sudoku analogy]

“My idea”

As pooling operators are (e.g., arithmetic or geometric) means, I thought of relying on the concept of “quasi-arithmetic” mean, viz.

$$M(x_1, \ldots, x_d) = \phi^{-1}\{w_1\phi(x_1) + \cdots + w_d\phi(x_d)\},$$

where ϕ is a monotonic function. This led me to look at maps of the form

$$C(u_1, \ldots, u_d) = \phi^{-1}\{\phi(u_1) + \cdots + \phi(u_d)\},$$

where $u_1, \ldots, u_d \in (0, 1)$, which is now universally known as an Archimedean copula.

My colleague János Aczél (1924–2020) alerted me to the work of Sklar and his (then recent) book with Berthold Schweizer, published in 1983.

Copulas

Under suitable conditions (see McNeil & Nešlehová, 2009, *AoS*),

$$C(u_1, \ldots, u_d) = \phi^{-1}\{\phi(u_1) + \cdots + \phi(u_d)\},$$

is a d-variate distribution function with uniform margins on $[0, 1]$.

As a result,

$$H(x_1, \ldots, x_d) = C\{F_1(x_1), \ldots, F_d(x_d)\}$$

is a d-variate distribution function with arbitrary margins $F_1, \ldots, F_d$.

Surprisingly, many of the multivariate models available back then were of this form (Genest & MacKay, 1986, *Canad. J. Statist.*).

Sklar's theorem

The following observation is key to dependence modeling, which has now become an industry.

If $(U_1, \ldots, U_d)$ *is a uniform random vector with copula* C*, then*

$$X_1 = F_1^{-1}(U_1), \ldots, X_d = F_d^{-1}(U_d)$$

has distribution $H = C(F_1, \ldots, F_d)$ *with margins* $F_1, \ldots, F_d$*. Moreover, any distribution* H *can be written in this form for some copula* C*, which is unique only when* $F_1, \ldots, F_d$ *are continuous.*

This is a 1959 *PISUP* result due to Abe Sklar (1925–2020).

Archimedean copulas is one class of copulas among many others.

The statistical issue (1–3)

Given data $X_1, \ldots, X_n$, how do you choose between different choices of ϕ, called Archimedean generator? Say,

$$\phi_\theta(t) = \frac{1}{\theta}\,(t^{-\theta} - 1), \quad \theta \in [-1, \infty)$$

$$\phi_\theta(t) = |\ln(t)|^\theta, \quad \theta \in [1, \infty)$$

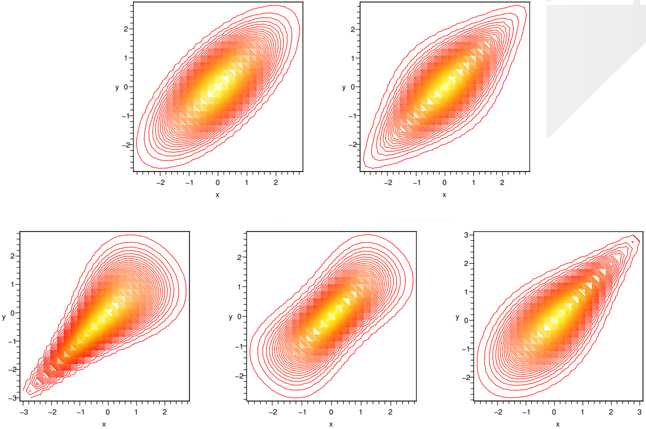
$$\phi_\theta(t) = -\ln\left(\frac{e^{-\theta t} - 1}{e^{-\theta} - 1}\right), \quad \theta \in \mathbb{R}$$

These lead to very different dependence structures (called the Clayton, Gumbel, and Frank model, respectively).

And given a family, how would you estimate θ from data?

Illustration with Gaussian margins

$$X \sim \mathcal{N}(0,1),\, Y \sim \mathcal{N}(0,1),\, C = ?$$



The statistical issue (2–3)

If you have data from the copula, i.e., mutually independent copies of

$$(U_1, \ldots, U_d) \sim C$$

things are relatively simple. For example, you might use

- ✓ the method of moments or maximum likelihood estimation;
- ✓ Kolmogorov–Smirnov statistics for goodness-of-fit;
- ✓ standard model selection techniques, cross-validation, etc.

A crucial point to understand, though, is that you don't get data from C!

The statistical issue (3–3)

Focusing on the bivariate case to simplify notation, what you get are data

$$(X_1, Y_1), \dots, (X_n, Y_n)$$

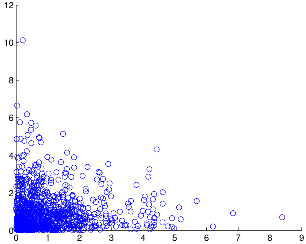
from $H = C(F, G)$. If margins F and G are known and continuous, then

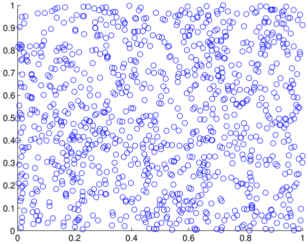
$$(U_1, V_1) = (F(X_1), G(Y_1)), \dots, (U_n, V_n) = (F(X_n), G(Y_n))$$

are indirect observations from C.

The (X, Y) sample and the (U, V) sample look quite different!

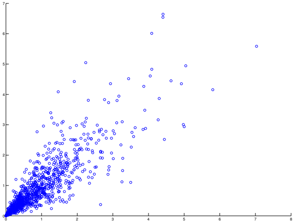
Illustration 1

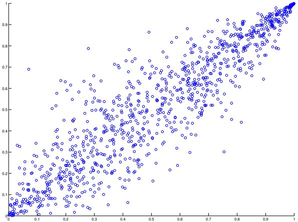




$$X, Y \sim F = G = \mathcal{E}(1) \quad \Rightarrow \quad (U, V) = (F(X), G(Y))$$

Illustration 2





$$X, Y \sim F = G = \mathcal{E}(1) \quad \Rightarrow \quad (U, V) = (F(X), G(Y))$$

Nonparametric inference (based on ranks)

What to do when you don't know the margins? Here is a simple but intriguing idea that goes a long way:

- ✓ Compute the ranks of $X_1, \ldots, X_n$, say $R_1, \ldots, R_n$.
- ✓ Compute the ranks of $Y_1, \ldots, Y_n$, say $S_1, \ldots, S_n$.

The pairs defined, for all $i \in \{1, \ldots, n\}$, by

$$(\hat{U}_1, \hat{V}_1) = (R_1/n, S_1/n), \ldots, (\hat{U}_n, \hat{V}_n) = (R_n/n, S_n/n)$$

are then pseudo-observations from C, but not a random sample.

Consequence: it is possible to carry out inference about a copula C (selection, estimation, validation) based on indirect observations.

Why does it work?

It works because if $X_1, \ldots, X_n$ form a random sample from F, one can define a map $\hat{F}_n$ by setting, for all $x \in \mathbb{R}$,

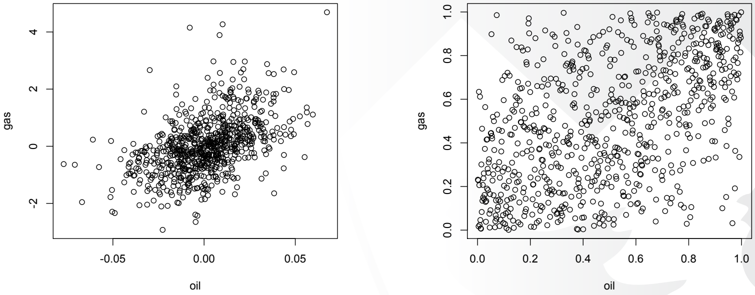
$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(X_i \leq x).$$

This map is called the empirical distribution function and, for all $i \in \{1, \ldots, n\}$,

$$\hat{F}_n(X_i) = R_i / n = \hat{U}_i.$$

The Glivenko–Cantelli lemma states that, as $n \to \infty$, $\hat{F}_n \to F$ point-wise. In other words, $\hat{F}_n$ is a consistent estimator of F!

Example (1–3)



Raw data and support of the empirical copula for the price of oil (Light Sweet Crude) and natural gas (mmBUT) data from 2004 to 2006
Source: Grégoire et al. (2008, *Energy Risk*)

Example (2–3)

If, say, the selected model is the Farlie–Gumbel–Morgenstern copula with density given, for all $u, v \in [0, 1]$, by

$$c_\theta(u, v) = 1 + \theta(1 - 2u)(1 - 2v),$$

one could estimate $\theta \in [-1, 1]$ by maximizing the pseudo log-likelihood

$$\ell(\theta) = \sum_{i=1}^{n} \ln[c_\theta\{\hat{F}_n(X_i), \hat{G}_n(Y_i)\}] = \sum_{i=1}^{n} \ln\{c_\theta(\hat{U}_i, \hat{V}_i)\},$$

or using a method of moments, e.g., Spearman's rho, exploiting the fact that in this model,

$$\text{corr}(U, V) = \theta/3.$$

Example (3–3)

Based on Spearman's rho, say, the estimator would be

$$\hat{\theta}_n = 3 - \frac{18}{n(n^2-1)} \sum_{i=1}^{n} (R_i - S_i)^2 .$$

This is a rank-based estimator. What are its properties, though?

- ✓ Is it unbiased? $\mathrm{E}(\hat{\theta}_n) = \theta$?
- ✓ Is it consistent? $\hat{\theta}_n \to \theta$ in probability as $n \to \infty$?
- ✓ Is it asymptotically Gaussian? $\sqrt{n}\,(\hat{\theta}_n - \theta) \rightsquigarrow \mathcal{N}(0, \sigma^2)$?
- ✓ Is it the most efficient estimator?
- ✓ Equivalently, is the asymptotic variance σ^2 as small as it could be?

Similar questions can be asked about the maximum pseudo-likelihood estimator, which is often not explicit, and others.

Central theoretical tool

Central to the theoretical developments is the empirical copula defined (in the bivariate case, for simplicity), for all $u, v \in [0,1]$, by

$$\hat{C}_n(u,v) = \frac{1}{n} \sum_{i=1}^{n} (\hat{U}_i \leq u, \hat{V}_i \leq v),$$

This is a random function or a stochastic process.

Its convergence was studied under various scenarios, e.g., by Rüschendorf (1976), Fermanian et al. (2004, *Bernoulli*), Segers (2012, *Bernoulli*), Genest et al. (2014, *Bernoulli*), and Genest et al. (2017, *JMVA*).

Philippe Capéraà and Louis-Paul Rivest (U. Laval), and most importantly Bruno Rémillard (UQTR, HEC Montréal) played a major part in helping me to unravel these issues.

Asymptotic result

If the copula C is "regular," then, as $n \to \infty$,

$$\hat{\mathbb{C}}_n = \sqrt{n}\,(\hat{C}_n - C) \rightsquigarrow \hat{\mathbb{C}},$$

where (in the bivariate case for notational simplicity)

$$\hat{\mathbb{C}}(u,v) = \mathbb{C}(u,v) - \frac{\partial C(u,v)}{\partial u}\,\mathbb{C}(u,1) - \frac{\partial C(u,v)}{\partial v}\,\mathbb{C}(1,v),$$

with $\mathbb{C}$ a Brownian sheet. In short, for large enough sample size n,

$$C_n \approx C.$$

The parts in red are "the price to pay" for not knowing the marginal distributions, although see Genest & Segers (2010), Genest et al. (2019).

As Tom Lehrer sang "Ah, ah, begins the fun..."

- ✓ Genest & Rivest (1995, *JASA*): Estimation for bivariate ACs
- ✓ Genest et al. (1995, *Biometrika*): Rank-based ML estimation
- ✓ Capéraà et al. (2000, *JMVA*): Archimax copulas
- ✓ Genest & Nešlehová (2007, *ASTIN*): Copulas for count data
- ✓ Genest & Rémillard (2004, *Test*): Rank-based tests of independence
- ✓ Genest et al. (2006, *Scand. J. Stat.*): Goodness-of-fit testing
- ✓ Genest & Segers (2009, *AoS*): Inference for extreme-value copulas
- ✓ Genest et al. (2011, *Test*): Estimation for multivariate ACs
- ✓ Genest et al. (2014, *Bernoulli*): Empirical multilinear copula process
- ✓ Côté et al. (2019, *Insur. Math. Econom.*): Copula regression models

Of course, many other people contributed to this topic — Remember that this is only a review of my work!

A glimpse of the technical challenges

Genest et al. (1995, *Biometrika*) show that, as $n \to \infty$, and under suitable regularity conditions (specified in the article),

$$\sqrt{n}\,(\hat{\theta}_n - \theta) \rightsquigarrow \mathcal{N}(0, \tau^2),$$

where τ^2 is larger than if the marginal distributions were known.

Such a result makes it possible to construct (asymptotic) confidence intervals and tests of hypotheses.

A similar result can be found when $\theta \in \mathbb{R}^p$, but things get more complicated when you estimate a generator ϕ, say.

Estimation of the Pickands dependence function

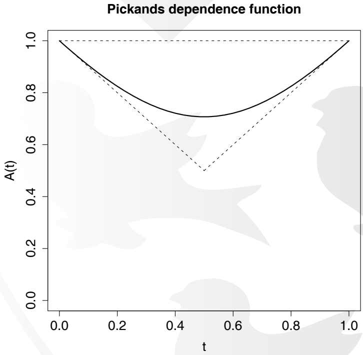
A generalized Fisher–Tippett theorem implies that for bivariate extremes,

$$C(u,v) = \exp\left[\ln(uv) A\left\{\frac{\ln(v)}{\ln(uv)}\right\}\right]$$

for all $u, v \in (0,1)$, where

- ✓ $A : [0,1] \to [1/2, 1]$ is convex;
- ✓ for all $t \in [0,1]$,

$$\max(t, 1-t) \leq A(t) \leq 1.$$



Genest–Segers estimators of A (1–2)

Recall that, for all $i \in \{1, \ldots, n\}$,

$$\hat{U}_i = R_i/n, \quad \hat{V}_i = S_i/n.$$

For any $t \in (0,1)$, set

$$\xi_i(t) = \frac{-\ln(\hat{U}_i)}{1-t} \wedge \frac{-\ln(\hat{V}_i)}{t} .$$

The estimators are, with γ denoting the Euler–Masceroni constant,

$$1/A_n^{\mathrm{P}}(t) = \frac{1}{n}\sum_{i=1}^{n} \xi_i(t) = \int_0^1 \hat{C}_n(u^{1-t}, u^t)\frac{du}{u},$$

$$\ln\{A_n^{\mathrm{CFG}}(t)\} = \int_0^1 \{\hat{C}_n(u^{1-t}, u^t) - \mathbf{1}(u > e^{-1})\}\frac{du}{u\ln(u)}.$$

Genest–Segers estimators of A (2–2)

If A is twice continuously differentiable, then as $n \to \infty$, one finds

$$\mathbb{A}_n^{\mathrm{P}} = \sqrt{n}\,(A_n^{\mathrm{P}} - A) \rightsquigarrow \mathbb{A}^{\mathrm{P}}, \quad \mathbb{A}_n^{\mathrm{CFG}} = \sqrt{n}\,(A_n^{\mathrm{CFG}} - A) \rightsquigarrow \mathbb{A}^{\mathrm{CFG}},$$

in the space $\mathcal{C}([0,1])$, where for all $t \in [0,1]$,

$$\mathbb{A}^{\mathrm{P}}(t) = -A^2(t) \int_0^1 \mathbb{C}(u^{1-t}, u^t)\,\frac{du}{u}\,,$$
$$\mathbb{A}^{\mathrm{CFG}}(t) = -A(t) \int_0^1 \mathbb{C}(u^{1-t}, u^t)\,\frac{du}{u\ln(u)}\,.$$

Alternatively, estimate A via splines (Cormier et al., 2014, *Extremes*) for which asymptotic theory now exists (Bücher et al., 2023, *Extremes*).

Goodness-of-fit testing

To test whether a copula family is adequate, one resorts to goodness-of-fit statistics, e.g., in the bivariate case

$$D_n = n \int_0^1 \int_0^1 \{C_{\hat\theta_n}(u,v) - \hat C_n(u,v)\}^2 \mathrm{d}\hat C_n(u,v)$$

$$= \sum_{i=1}^{n} \{C_{\theta_n}(\hat U_i, \hat V_i) - \hat C_n(\hat U_i, \hat V_i)\}^2,$$

called the Cramér–von Mises statistic.

As its limiting distribution depends on the unknown parameter value θ, one must approximate it through resampling methods such as the parametric bootstrap (Genest & Rémillard, 2008, *AIHP*).

Copula models with covariates

In regression, the dependence of a variable Y on covariates $X_1, \ldots, X_p$ is modeled through

$$\mathrm{E}(Y \mid x_1, \ldots, x_d) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p.$$

In contrast, copula regression centers on the distribution of $Y_1, \ldots, Y_d$ given $\mathbf{X} = (X_1, \ldots, X_p) = \mathbf{x} = (x_1, \ldots, x_p)$, so that, for all $y_1, \ldots, y_d \in \mathbb{R}$,

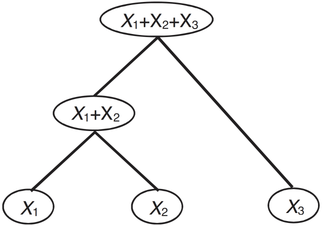
$$\begin{aligned} H(y_1, \ldots, y_d \mid x_1, \ldots, x_p) &= \Pr(Y_1 \le y_1, \ldots, Y_d \le y_d \mid \mathbf{X} = \mathbf{x}) \\ &= C\{F_{1\mathbf{x}}(y_1), \ldots, F_{d\mathbf{x}}(y_d)\}, \end{aligned}$$

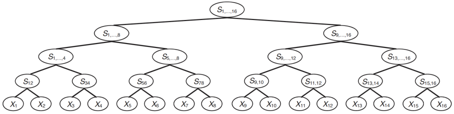
where $F_{i\mathbf{x}}(y_i) = \Pr(Y_i \le y_i \mid \mathbf{X} = \mathbf{x})$ for each $i \in \{1, \ldots, d\}$.

Hierarchical copula structures (1–2)

In practice, e.g., actuarial applications, one often deals with tens, hundreds or thousands of variables.

In such cases, hierarchical structures are used, e.g.,

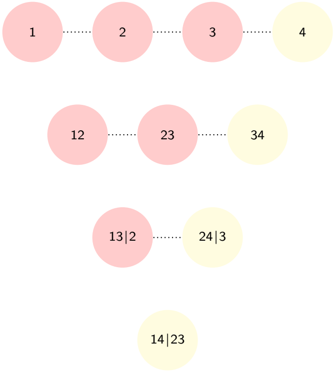




Source: Côté & Genest (2015, *Canad. J. Statist.*)

Hierarchical copula structures (2–2)

Identifying the “best structure” requires intensive computer searches!



$$f(x_1, x_2, x_3, x_4) =$$

$$f_1(x_1)\ f_2(x_2)\ f_3(x_3)\ f_4(x_4)$$

$$c_{12}\{F_1(x_1), F_2(x_2)\}$$
$$c_{23}\{F_2(x_2), F_3(x_3)\}$$
$$c_{34}\{F_3(x_3), F_4(x_4)\}$$

$$c_{13|2}\{F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2) \mid x_2\}$$
$$c_{24|3}\{F_{2|3}(x_2|x_3), F_{4|3}(x_4|x_3) \mid x_3\}$$

$$c_{14|23}\{F_{1|23}(x_1|x_2, x_3), F_{4|23}(x_4|x_2, x_3) \mid x_2, x_3\}$$

VIGNETTE 3

Risk Management

U. Laval, Québec; McGill, Montréal

Coastal flooding

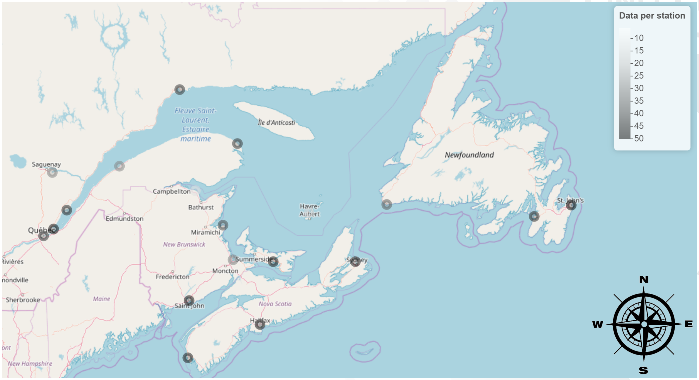
It is an important issue:

- ✓ Coastal floods accounted for 46% of natural disasters in 2018.
- ✓ Losses were estimated at $1 trillion between 1980 and 2013.
- ✓ There were more than 220,000 casualties.

Yet no Canadian insurance company covered this risk before 2015.

Sources: Winsemius et al. (*Nat. Climate Change*, 2013) and Munich Re

Your mission (if you choose to accept it)



Estimate the flooding risk everywhere without directly relevant data!

Data: Water level series recorded over 50 years in 21 locations (buoys)
Warning: Lots of missing values

Strategy for spatial extremes

Consider a spatial domain $\mathcal{D} \subseteq \mathbb{R}^2$ and for monitored sites $s_1, \ldots, s_d$, let

$$\boldsymbol{s} = (s_1, \ldots, s_d) \in \mathcal{D} \times \cdots \times \mathcal{D}.$$

At site s_j, assume that (daily, monthly, yearly) maxima M_{s_j} are such that

$$M_{s_j} \sim \text{GEV}(\mu_{s_j}, \sigma_{s_j}, \xi_{s_j}).$$

A latent spatial field is then constructed for the GEV parameters

$$\boldsymbol{\mu} = (m_{s_1}, \ldots, m_{s_d}), \quad \boldsymbol{\sigma} = (\sigma_{s_1}, \ldots, \sigma_{s_d}), \quad \boldsymbol{\xi} = (\xi_{s_1}, \ldots, \xi_{s_d}).$$

Simplifying assumption

Conditional independence is assumed between sites, viz.

$$M_{s_j} \perp M_{s_k} \mid (\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}).$$

This assumption is rarely met in practice but sufficient to get reliable results when the focus is on marginal extremal behavior.

Other option: Induce dependence through, say, a Student t copula with ν degrees of freedom and dispersion matrix

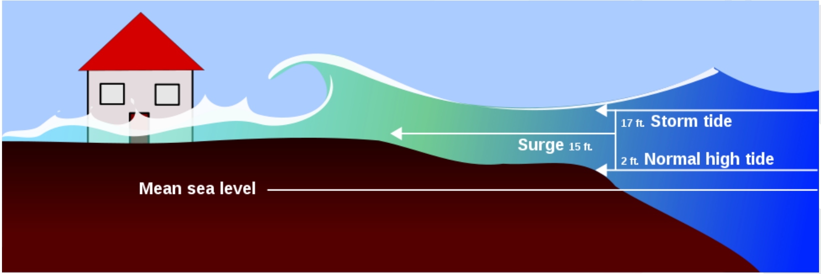
$$Q_{jk} = \exp(-d_{jk}/\rho_Q),$$

where d_{jk} is the distance between sites s_j and s_k (this is the exponential covariogram) and, say, $\rho_Q \sim \mathcal{U}[0, \max_{ij}(d_{ij}/3)]$.

Beck et al. (*Environmetrics*, 2020)

- ✓ Model Z_i, the annual maximum surge at location $s_i \in \mathcal{D}$, using an extreme Bayesian hierarchical model. In particular, each station will be modeled using a GEV.
- ✓ Incorporate spatiality, so as to allow information sharing across stations, improving fit, and interpolate to allow model to include un-monitored locations.
- ✓ Incorporate a copula, to further quantify dependence and ensure smoothness of the maximum surges across adjacent stations.
- ✓ Using realizations of extreme surges across the domain, generate potentially extreme water-levels and assess the risk of flooding from these events.

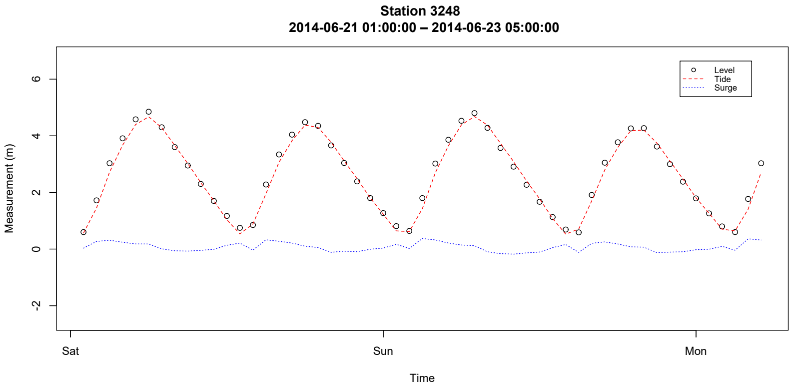
Storm tides and surges



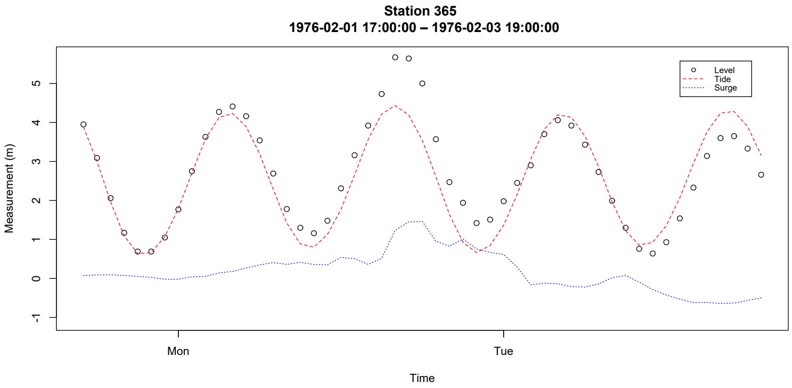
The surge is the water level rise beyond its expected level at high tide.

The tide is a covariate whose cycles must be taken into account.

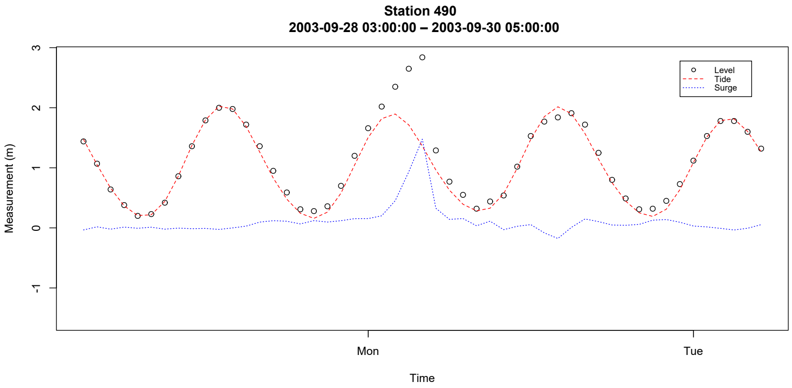
Surges observed at Québec, QC



Surges observed at Yarmouth, NS



Surges observed at Halifax, NS



Hierarchical model

At each of the 21 stations, the maximum surge is modeled with a GEV,

$$M_i \sim \text{GEV}(\xi, \mu_i, \sigma_i),$$

with $\xi > 0$ (Fréchet).

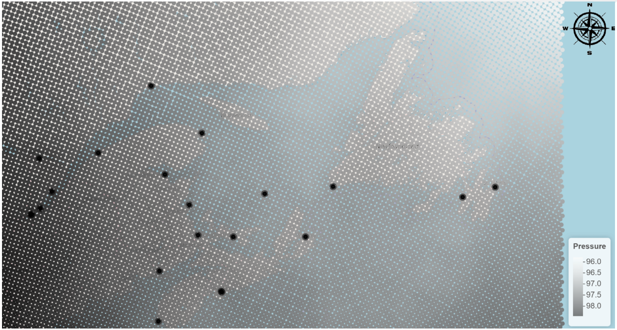
To link the components of $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{21})$ and $\boldsymbol{\Phi} = (\phi_1, \ldots, \phi_{21})$, covariates must be used, e.g., proximity or sea level water pressure, viz.

$$\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{X}\beta_{\boldsymbol{\mu}}, \tau^2_{\boldsymbol{\mu}}\Sigma_{\boldsymbol{\mu}}), \quad \boldsymbol{\Phi} \sim \mathcal{N}(\boldsymbol{X}\beta_{\boldsymbol{\Phi}}, \tau^2_{\boldsymbol{\Phi}}\Sigma_{\boldsymbol{\Phi}}).$$

However, the shape parameter ξ is assumed to be constant.

Sea level pressure

Important covariate obtained through a reconstruction of the past climate using the Canadian Regional Climate Model 5 developed by Ouranos.



Benefit of the covariate

Return Period	Individual Model	Spatial Model	%RC
10	1.81	2.26	25
25	1.95	2.63	35
50	2.06	2.93	42
100	2.17	3.25	50
250	2.31	3.71	61
500	2.41	4.09	69
1000	2.52	4.49	78

Station 3248: Vieux Québec

Return Period	Individual Model	Spatial Model	%RC
10	2.30	2.30	0
25	2.57	2.68	4
50	2.78	2.99	8
100	2.98	3.32	12
250	3.24	3.80	17
500	3.43	4.19	22
1000	3.62	4.61	27

Station 3250: Lauzon

%RC = percent relative change

The spatial model gives more realistic and spatially consistent results.

Interpolation (1–2)

Once the model has been adjusted at the 21 stations, one can interpolate anywhere on the grid.

In locations $\mathcal{S}^* = \{\boldsymbol{s}_{22}, \ldots, \boldsymbol{s}_d\}$, the mean vector $\boldsymbol{\mu}_{\mathcal{S}\cup\mathcal{S}^*}$ is

$$\boldsymbol{\mu}_{\mathcal{S}\cup\mathcal{S}^*} \sim \mathcal{N}_d(\boldsymbol{X}_{\boldsymbol{\mu}_{\mathcal{S}\cup\mathcal{S}^*}} \beta_{\boldsymbol{\mu}_{\mathcal{S}\cup\mathcal{S}^*}}, \tau^2_{\boldsymbol{\mu}_{\mathcal{S}\cup\mathcal{S}^*}} \Sigma_{\boldsymbol{\mu}_{\mathcal{S}\cup\mathcal{S}^*}}),$$

where $\boldsymbol{X}_{\boldsymbol{\mu}_{\mathcal{S}\cup\mathcal{S}^*}} = (\boldsymbol{X}^\top_{\boldsymbol{\mu}_\mathcal{S}}, \boldsymbol{X}^\top_{\boldsymbol{\mu}_{\mathcal{S}^*}})^\top$ and

$$\Sigma_{\boldsymbol{\mu}_{\mathcal{S}\cup\mathcal{S}^*}} = \begin{bmatrix} \Sigma_{\boldsymbol{\mu}_\mathcal{S}} & \Sigma_{\boldsymbol{\mu}_{\mathcal{S},\mathcal{S}^*}} \\ \Sigma_{\boldsymbol{\mu}_{\mathcal{S},\mathcal{S}^*}} & \Sigma_{\boldsymbol{\mu}_{\mathcal{S}^*}} \end{bmatrix}.$$

Given $\boldsymbol{\mu}_{\mathcal{S}} = \boldsymbol{m}$ in observed locations, the parameters in non-observed locations are uncertain but this uncertainty can be quantified, viz.

$$\boldsymbol{\mu}_{\boldsymbol{\mathcal{S}}^*} \mid \boldsymbol{\mu}_{\mathcal{S}} = \boldsymbol{m} \sim \mathcal{N}_{d-21}(\bar{\boldsymbol{\mu}}_{\mathcal{S}^*}, \tau^2_{\boldsymbol{\mu}_{\mathcal{S}^*}} \bar{\Sigma}_{\boldsymbol{\mu}_{\mathcal{S}^*}}),$$

where

$$\bar{\boldsymbol{\mu}}_{\mathcal{S}^*} = \boldsymbol{X}_{\boldsymbol{\mu}_{\boldsymbol{\mathcal{S}}^*}} \beta_{\boldsymbol{\mu}_{\boldsymbol{\mathcal{S}}^*}} + \Sigma_{\boldsymbol{\mu}_{\mathcal{S},\mathcal{S}^*}} \Sigma^{-1}_{\boldsymbol{\mu}_{\mathcal{S},\mathcal{S}}} (\boldsymbol{m} - \boldsymbol{X}_{\boldsymbol{\mu}_{\boldsymbol{s}}} \beta_{\boldsymbol{\mu}_{\boldsymbol{s}}})$$

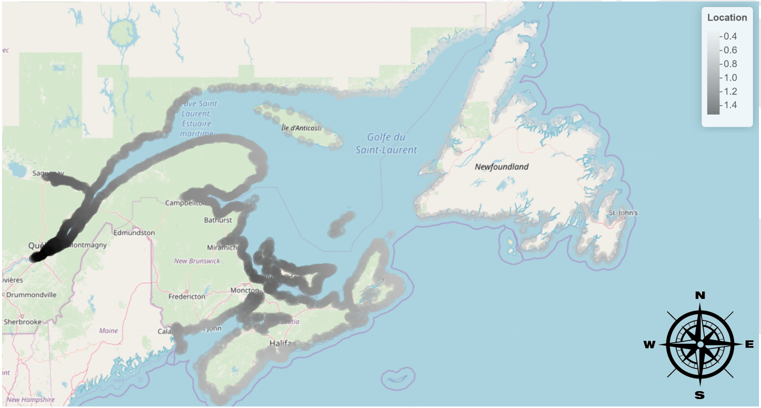
and

$$\bar{\Sigma}_{\boldsymbol{\mu}_{\mathcal{S}^*}} = \Sigma_{\boldsymbol{\mu}_{\mathcal{S}^*}} - \Sigma_{\boldsymbol{\mu}_{\mathcal{S},\mathcal{S}^*}} \Sigma^{-1}_{\boldsymbol{\mu}_{\mathcal{S}}} \Sigma_{\boldsymbol{\mu}_{\mathcal{S},\mathcal{S}^*}}.$$

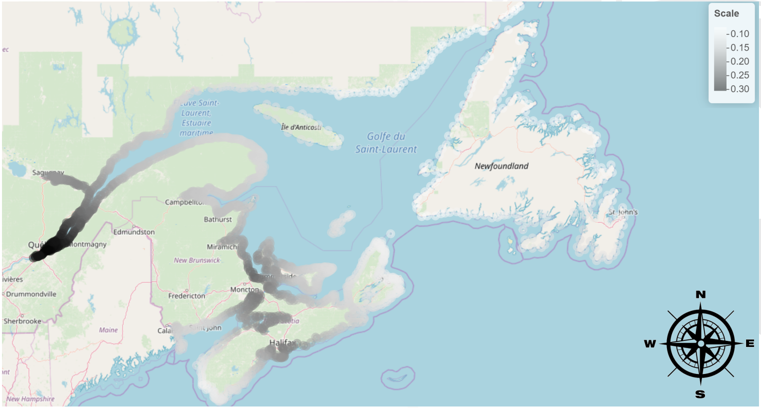
It is then possible to simulate surges everywhere on the coast and to compute the probability and magnitude of future catastrophes.

To do so, of course, one must factor in the systematic tidal effect.

Mean interpolation



Scale interpolation



Other examples of application (1–2)

At McGill, my team has combined Bayesian modeling, expert use and copula dependence to investigate many issues, e.g.,

Jalbert et al. (2019, *JRSS-C*)

Estimate the return period of the 2011 Richelieu Valley flood
[Related to flood hazard map updating and boundary water management]

Li et al. (2021, *Environmetrics*)

Predict the frequency of drought periods on the Rivière des Mille Îles
[Relevant for city waste water management and fresh water supply]

Other examples of application (2–2)

Côté et al. (2022, *Bayesian Anal.*)

Model multivariate multilevel insurance claims exploiting partial information from open, unsettled claims
[Used by a large Canadian insurance company]

Jalbert et al. (2022, *JABES*):

Production of intensity-duration-frequency (IDF) curves
[Implemented by Hydro-Québec and Québec Department of Environment]

Acknowledgments

UNIVERSITÉ LAVAL

Fonds de recherche sur la nature et les technologies Québec

CANADA RESEARCH CHAIRS

University of Waterloo

NSERC CRSNG

IFM²

INSTITUT DE FINANCE MATHEMATIQUE DE MONTREAL

McGill

TROTTIER ISPP
INSTITUTE FOR SCIENCE AND PUBLIC POLICY
INSTITUT POUR LA SCIENCE ET LES POLITIQUES PUBLIQUES