Probabilistic tsunami hazard maps for the western US coastline

Georgios Boumis[1,2]

[1]Department of Civil, Construction, and Environmental Engineering, University of Alabama, Tuscaloosa, AL, USA

[2]Center for Complex Hydrosystems Research, University of Alabama, Tuscaloosa, AL, USA

Introduction

Tsunami: a series of waves induced by vertical displacement of ocean water

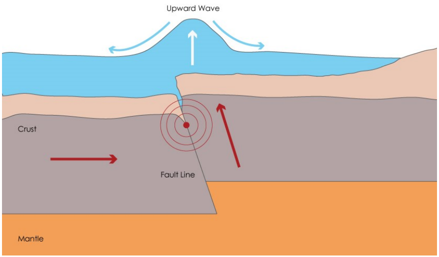


Figure 1: Schematic of tsunami generation due to a submarine earthquake.

Figure 2: Crescent City (California) harbor after the 2011 Tohoku tsunami.

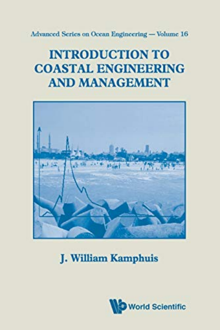
Introduction

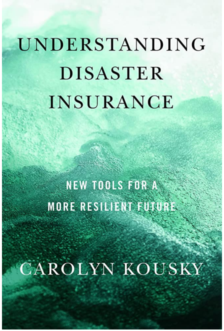
Quantifying tsunami hazard means providing the answer to the question:

"What is the probability that we observe a tsunami height of X meters close to the shore?"

Why is this important?

- Coastal defenses, e.g., seawalls, breakwaters, etc.
- Flood insurance policies





For a location of interest where empirical tsunami data exist:

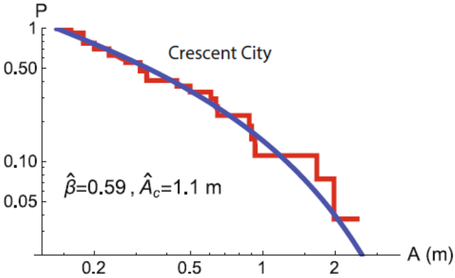


Figure 3: Tapered Pareto distribution for observed tsunami amplitude data at Crescent City (from **Geist and Parsons, 2016**).

Tsunamis are rare - for most cases only a few events have been recorded!

Introduction

An alternative: numerical physics-based simulation of earthquake, tsunami generation and propagation

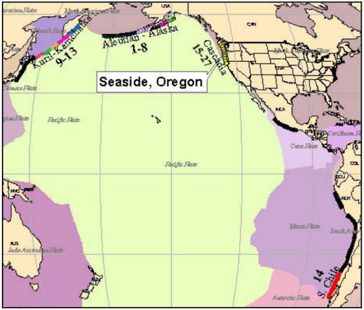
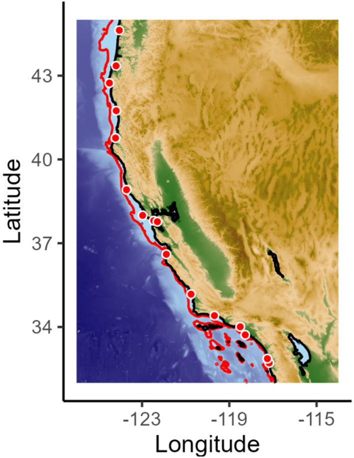


Figure 4: Numerical tsunami model domain for Seaside, Oregon (from **Wong et al., 2005**).

Tedious and time-consuming task...

This work

Instead... a spatial Bayesian hierarchical model to map tsunami hazard for the West Coast



Data & Pre-processing

Tsunami amplitude observations were retrieved from NOAA's (National Oceanic and Atmospheric Administration) Global Historical Tsunami Database

Figure 5: Tide gauge used for measuring sea levels.

- Events are between **1900-2021**
- Source validity is **definite tsunami**
- Cause of tsunami is **earthquake**
- Entry is **not doubtful**
- Tide gauges with **at least 10** such events

Methodology

A two-level hierarchical model:

$$\eta_{ij} \overset{iid}{\sim} TP(\beta_j, \theta_j), \tag{1}$$

for $i = 1, \ldots, n_j$ and $j = 1, \ldots, J = 16$ (number of stations). TP stands for Tapered Pareto with cumulative distribution function:

$$F(\eta; \beta, \theta) = 1 - \left(\frac{u}{\eta}\right)^{\beta} \exp\left(\frac{u - \eta}{\theta}\right) \tag{2}$$

where $u > 0$ is the minimum value of amplitude variable η, i.e., the threshold of record completeness, $\beta > 0$ is the shape parameter, and $\theta \geq u$ is the upper cutoff parameter.

$$\boldsymbol{\beta} \sim G_{\boldsymbol{\gamma}_\beta}, \quad \text{and} \quad \boldsymbol{\theta} \sim G_{\boldsymbol{\gamma}_\theta} \tag{3}$$

Parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_J)'$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_J)'$ are assumed to follow probability distributions $G_{\boldsymbol{\gamma}_\beta}(\cdot)$ and $G_{\boldsymbol{\gamma}_\theta}(\cdot)$ which can facilitate information sharing across stations.

Methodology (hierarchical Bayes' rule)

$$\underbrace{p(\boldsymbol{\beta},\boldsymbol{\theta},\gamma|\boldsymbol{\eta})}_{\text{posterior}} \propto \underbrace{p(\boldsymbol{\eta} \mid \boldsymbol{\beta},\boldsymbol{\theta})p(\boldsymbol{\beta} \mid \gamma_\beta)p(\boldsymbol{\theta} \mid \gamma_\theta)}_{\text{likelihood}}\underbrace{p(\gamma_\beta)p(\gamma_\theta)}_{\text{prior}},$$

where

$$\begin{aligned} p(\boldsymbol{\eta} \mid \boldsymbol{\beta},\boldsymbol{\theta}) &= \prod_{j=1}^{J}\prod_{i=1}^{n_j} f(\eta_{ij} \mid \beta_j, \theta_j), \\ p(\boldsymbol{\beta} \mid \gamma_\beta) &= f_N(\boldsymbol{\beta}; \boldsymbol{m}_\beta, \boldsymbol{K}_\beta), \\ p(\boldsymbol{\theta} \mid \gamma_\theta) &= f_N(\boldsymbol{\theta}; \boldsymbol{m}_\theta, \boldsymbol{K}_\theta), \end{aligned}$$

while $\boldsymbol{f_N}$ is the **Multivariate Normal** density. The covariance matrices are of the form:

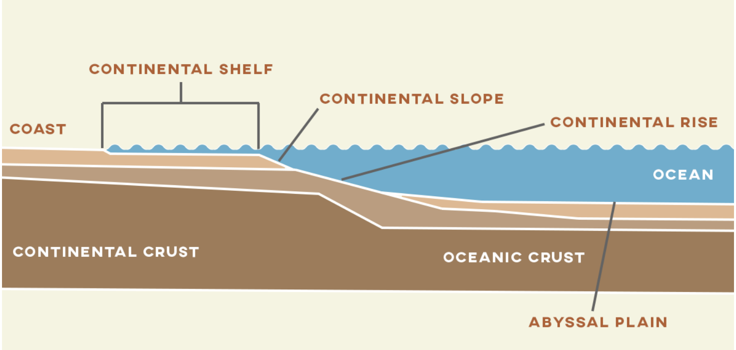
$$\boldsymbol{K}_{(i,j)} = \alpha^2 \exp\left(-\frac{|d_{ij}|}{\rho}\right) \tag{4}$$

whereas the mean vectors are expressed as:

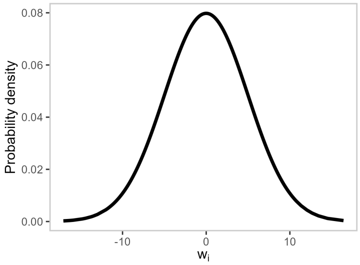
$$\boldsymbol{m} = w_0 + w_1\boldsymbol{y_1} + \cdots + w_q\boldsymbol{y_q} \tag{5}$$

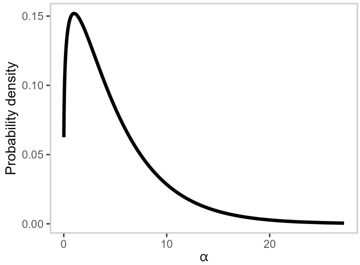
Methodology (covariates)

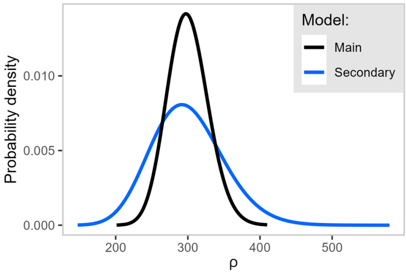
$q = 3$, i.e., z-scores of **Latitude**, **Longitude**, and width of the **Continental Shelf**



Methodology (priors)







Stan Application

```
functions{
    // Function to return Tapered Pareto log-likelihood for (log-)b and (log-)ac parameterization
    real tp_lpdf(vector y, real u, real b, real ac) {
        vector[rows(y)] lp;
        real loglike;
        int N;

        N = rows(y);

        for(i in 1:N){
            lp[i] = log(exp(b) / y[i] + 1 / exp(ac));
        }
        loglike = sum(lp) + exp(b) * N * log(u) - exp(b) * sum(log(y)) + (u * N) / exp(ac) - 1 / exp(ac) * sum(y);
        return(loglike);
    }
```

```
transformed parameters{
    // mean function for b
    vector[M] gfb = w0b * x[, 1] + w1b * x[, 2] + w2b * x[, 3] + w3b * x[, 4];

    // mean function for ac
    vector[M] gfac = w0ac * x[, 1] + w1ac * x[, 2] + w2ac * x[, 3] + w3ac * x[, 4];

    // covariance matrix for b and its Cholesky decomposition
    matrix[M, M] Kb = kernel(M, dist, ab, rhob);
    matrix[M, M] Lb = cholesky_decompose(Kb);

    // covariance matrix for ac and its Cholesky decomposition
    matrix[M, M] Kac = kernel(M, dist, aac, rhoac);
    matrix[M, M] Lac = cholesky_decompose(Kac);

    // non-centered parameterization
    vector[M] b = gfb + Lb * zb;
    vector[M] ac = gfac + Lac * zac;
}
```

Maps

Sampling for unmonitored sites:

$$\boldsymbol{\beta}_{v*} | \boldsymbol{\beta}_v \sim MVN_{s-16}, (\bar{\boldsymbol{\beta}}_{v*}, \bar{\boldsymbol{K}}_{\boldsymbol{\beta}_{v*}})$$

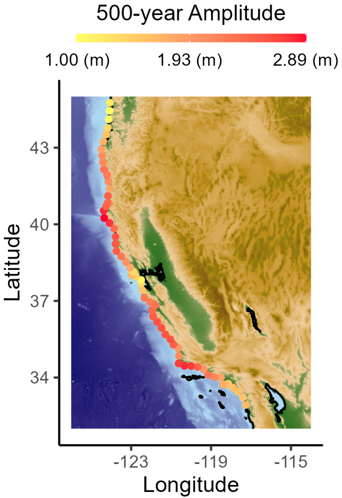
$$\bar{\boldsymbol{\beta}}_{v*} = \boldsymbol{m}_{\boldsymbol{\beta}_{v*}} + \boldsymbol{K}_{\boldsymbol{\beta}_{v,v*}} \times \boldsymbol{K}_{\boldsymbol{\beta}_v}^{-1} \times (\boldsymbol{\beta}_v - \boldsymbol{m}_{\boldsymbol{\beta}_v})$$

$$\bar{\boldsymbol{K}}_{\boldsymbol{\beta}_{v*}} = \boldsymbol{K}_{\boldsymbol{\beta}_{v*}} - \boldsymbol{K}_{\boldsymbol{\beta}_{v,v*}} \times \boldsymbol{K}_{\boldsymbol{\beta}_v}^{-1} \times \boldsymbol{K}_{\boldsymbol{\beta}_{v,v*}}$$

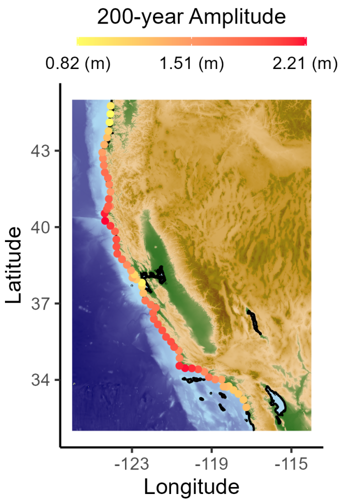
From probability to “return period (T)”:

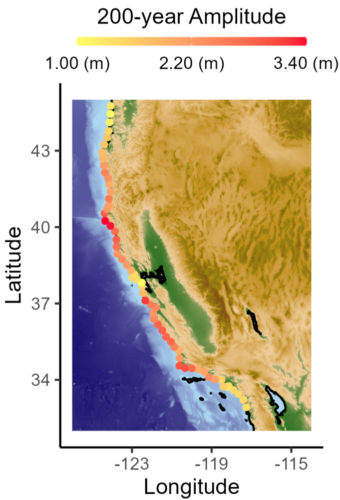
$$T = \frac{1}{\lambda p} = \frac{\tau}{p}$$

where τ is the mean interarrival time.



More maps...


200-year Amplitude
0.82 (m)
1.51 (m)
2.21 (m)
Latitude
43
40
37
34
-123
-119
-115
Longitude


200-year Amplitude
1.00 (m)
2.20 (m)
3.40 (m)
Latitude
43
40
37
34
-123
-119
-115
Longitude

Future research

So far:

$$\boldsymbol{K}_{(i,j)} = \alpha^2 \exp\left(-\frac{|d_{ij}|}{\rho}\right)$$

$$p(\boldsymbol{\eta} \mid \boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{j=1}^{J} \prod_{i=1}^{n_j} f(\eta_{ij} \mid \beta_j, \theta_j)$$

Preferably, we want:

- Anisotropic kernel
- Copulas