STORM TIDES AND EXTREME VALUE STATISTICS

BEE 6940 LECTURE 9

MARCH 20, 2023

TABLE OF CONTENTS
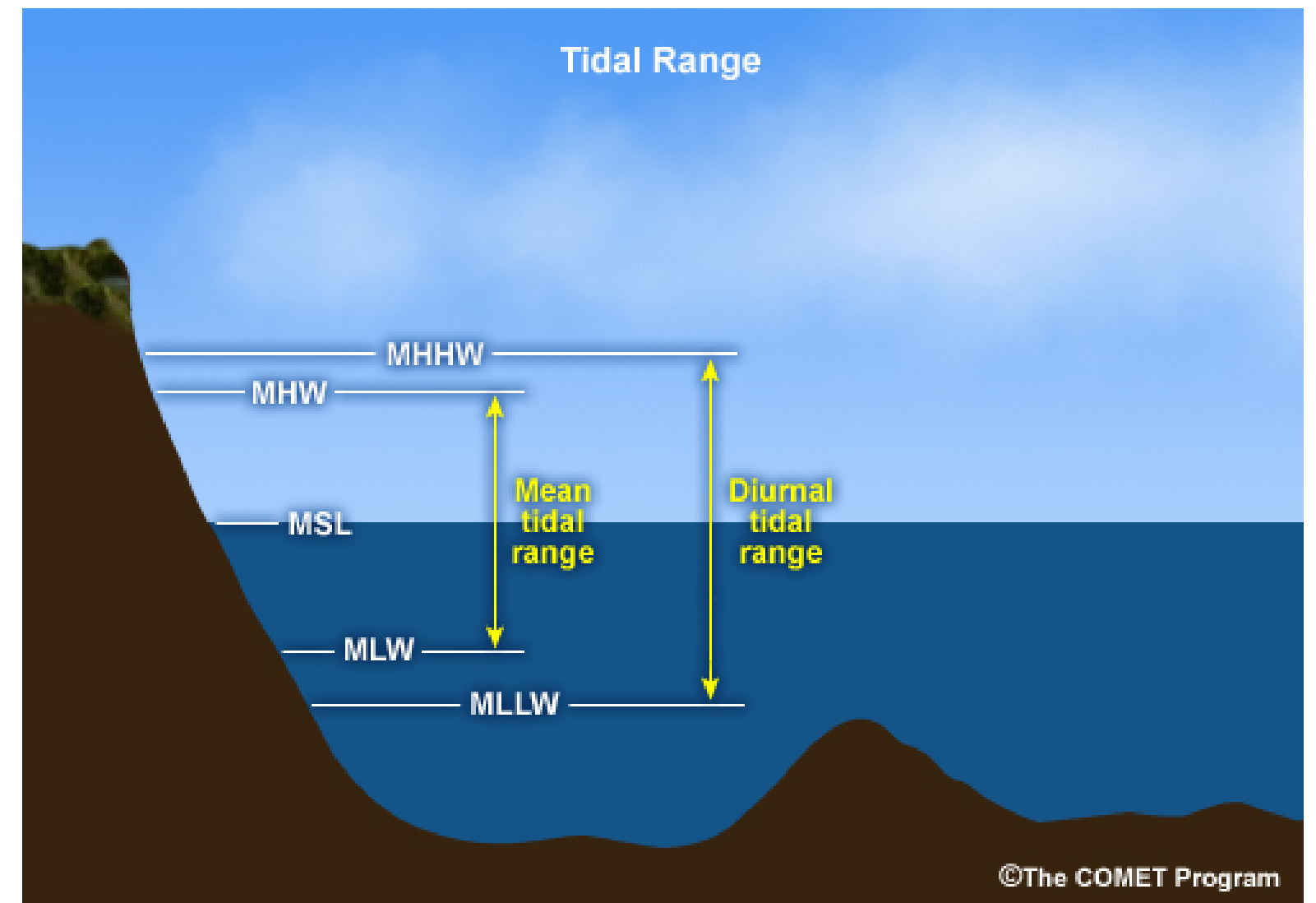
REVIEW OF COASTAL FLOOD HAZARDS

How Are Local High Water Levels Measured?

Tide gauge data comes in many "flavors", based on local tidal and diurnal cycles.
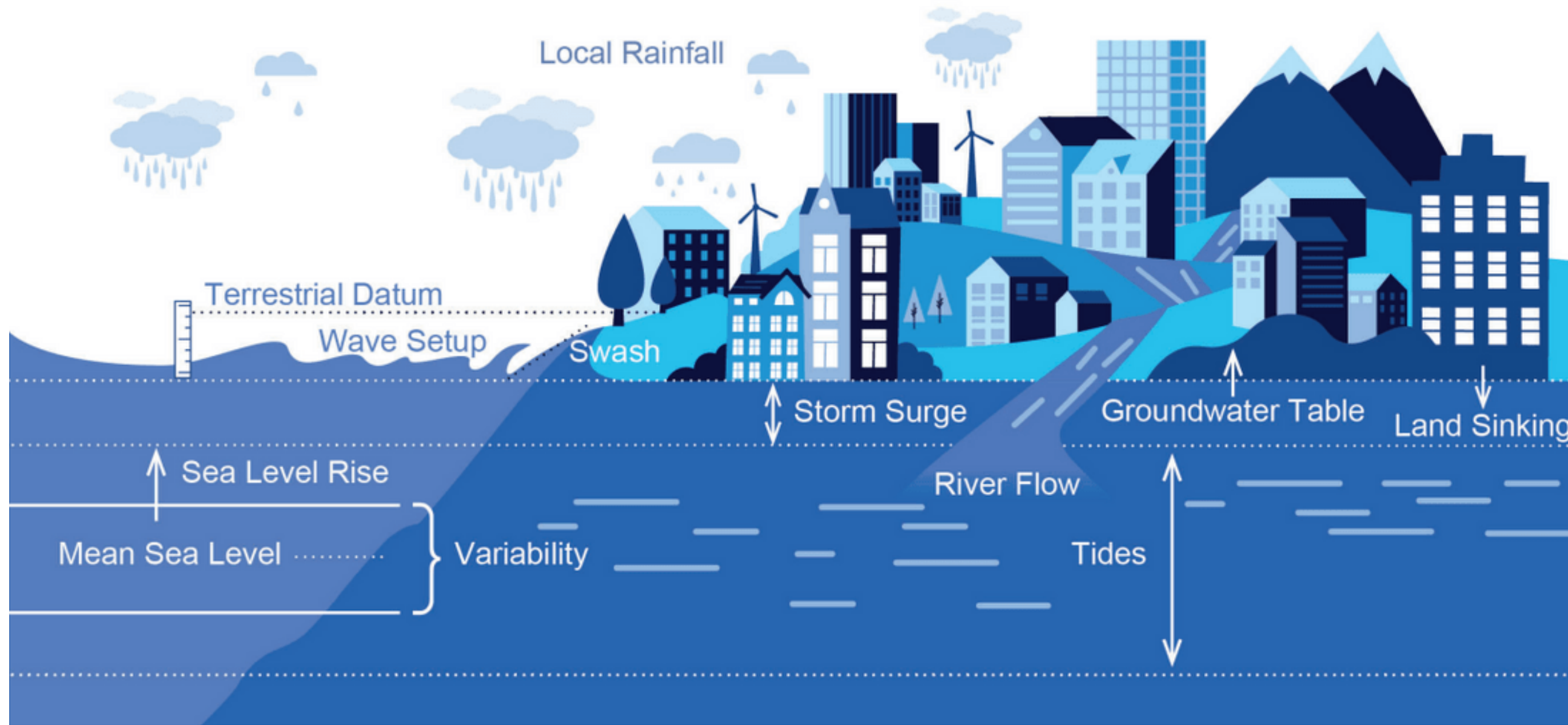
Mean Highest High Water (MHHW) is the typical "extreme" sea level datum.

Source: Inside the Eye Blog, National Hurricane Center, 01-29-2016

Contributors to Extreme Sea Levels

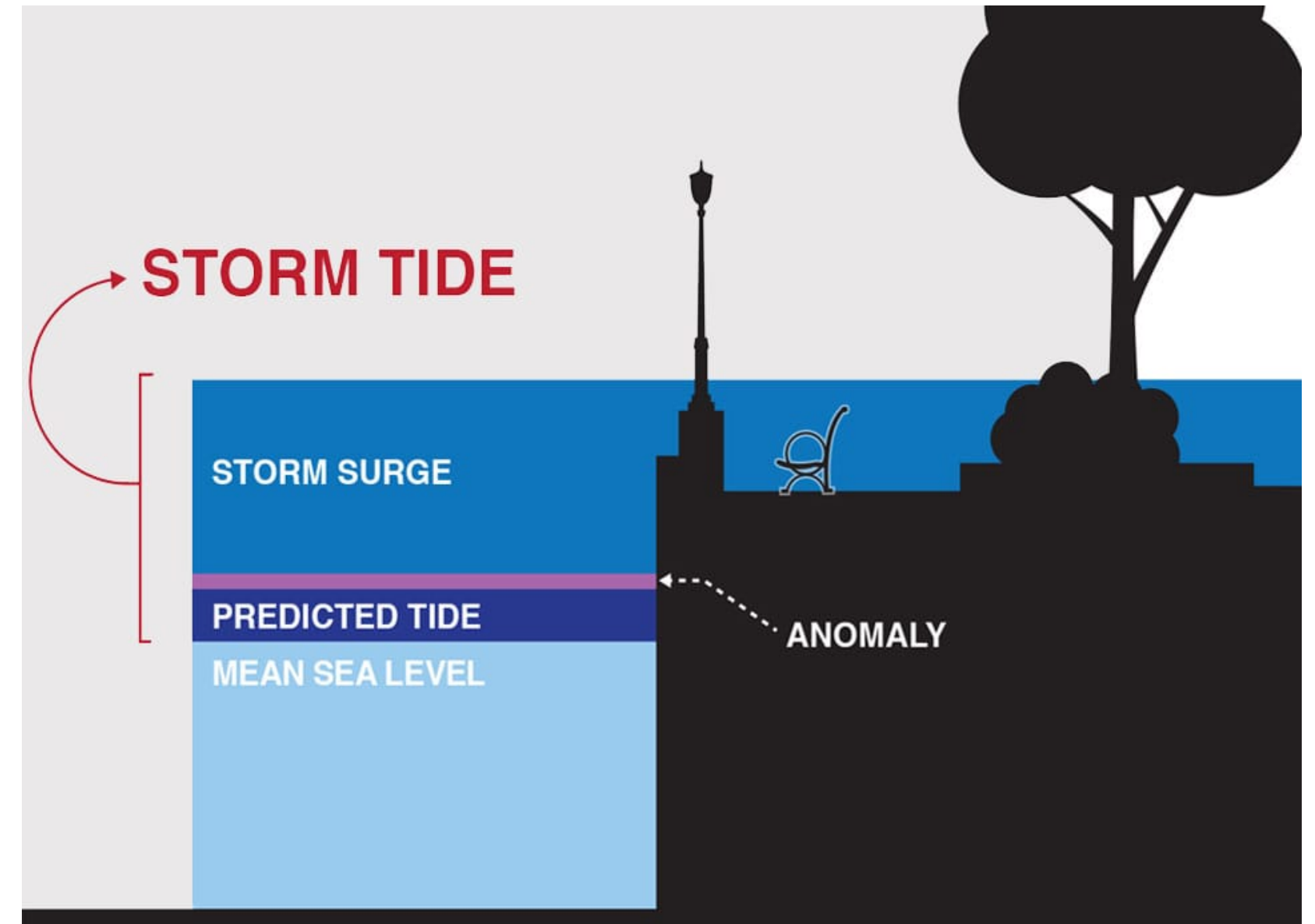Physical Factors Directly Contributing to Coastal Flood Exposure

Source: NOAA 2022 Sea Level Rise Technical Report

STORM TIDES

Storm Tides vs. Storm Surge

Storm surge is the rise in water level due solely to the winds from a storm.

Storm tides are the total water level during a storm, including tides and the storm surge.

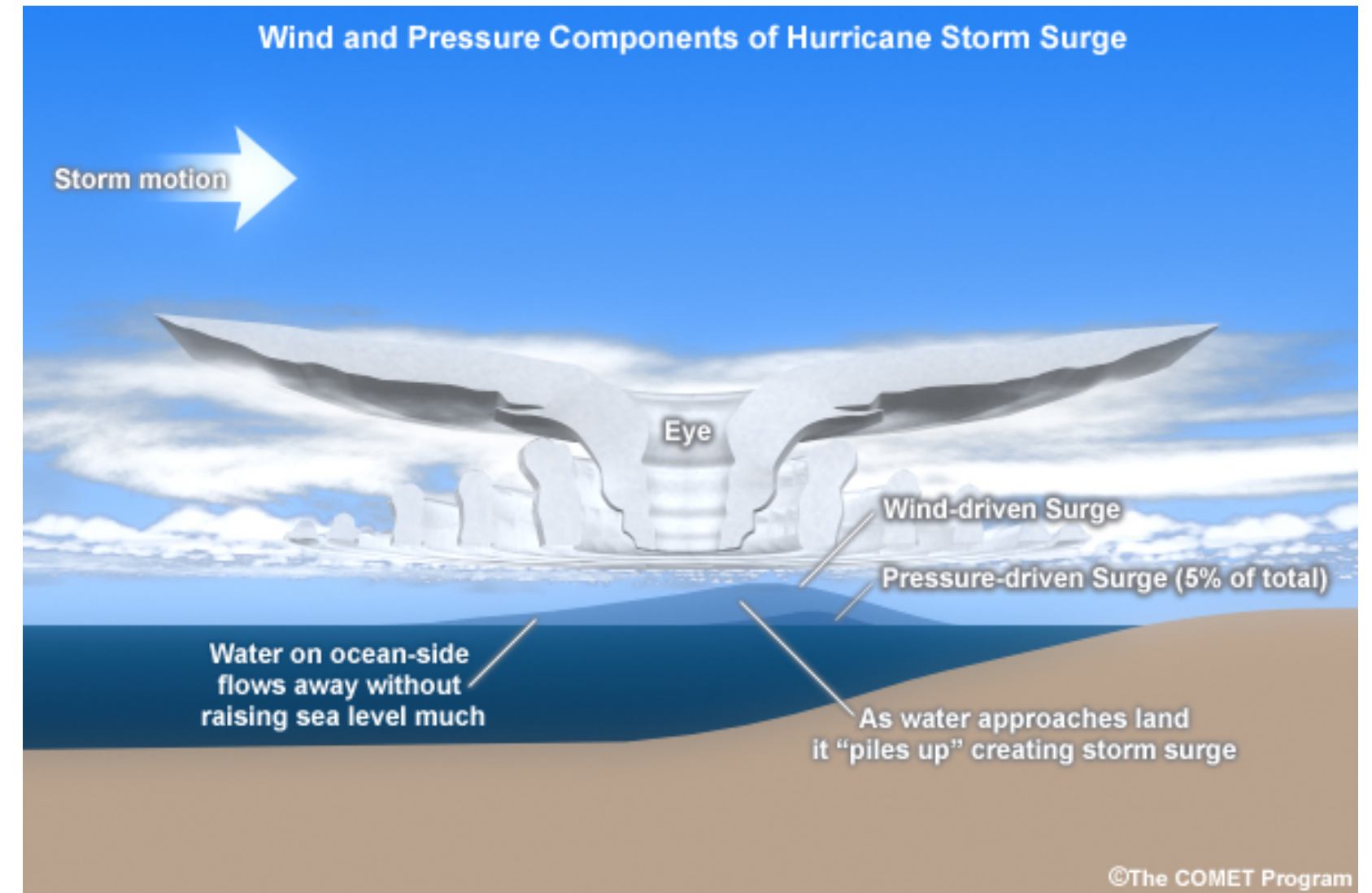*Source: NOAA Ocean Service*

CONTRIBUTORS TO STORM SURGE

The specifics of a storm surge event will depend on:

- Wind speed and angle (relative to shore)

- Coastal terrain

- Storm track and pressure

Source: NOAA National Hurricane Center

Risk Analysis and Storm Tides

This complexity means modeling storm surges often requires spatially-explicit physical models (or spatial emulators: more on this later).

But often (and particularly for risk analysis), we care about the distribution of storm tide levels and the potential to overwhelm flood mitigation infrastructure.

STATISTICAL MODELING OF STORM TIDES

As a result, we want to use statistical models to understand how probable flood events might be.

But, unlike the statistical applications we've seen so far, we aren't as interested in the "typical" or average occurrence, we care about the **extremes**.
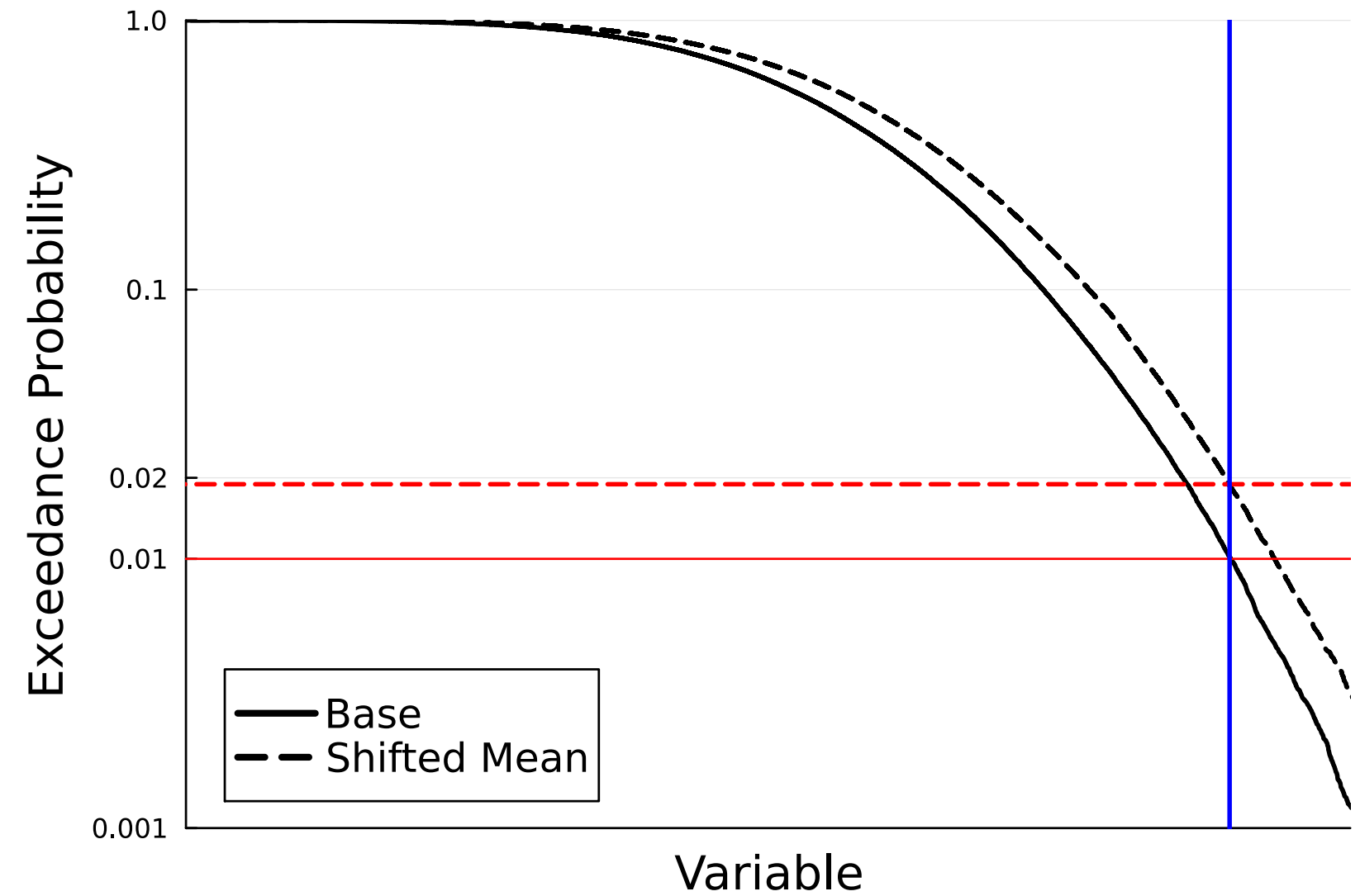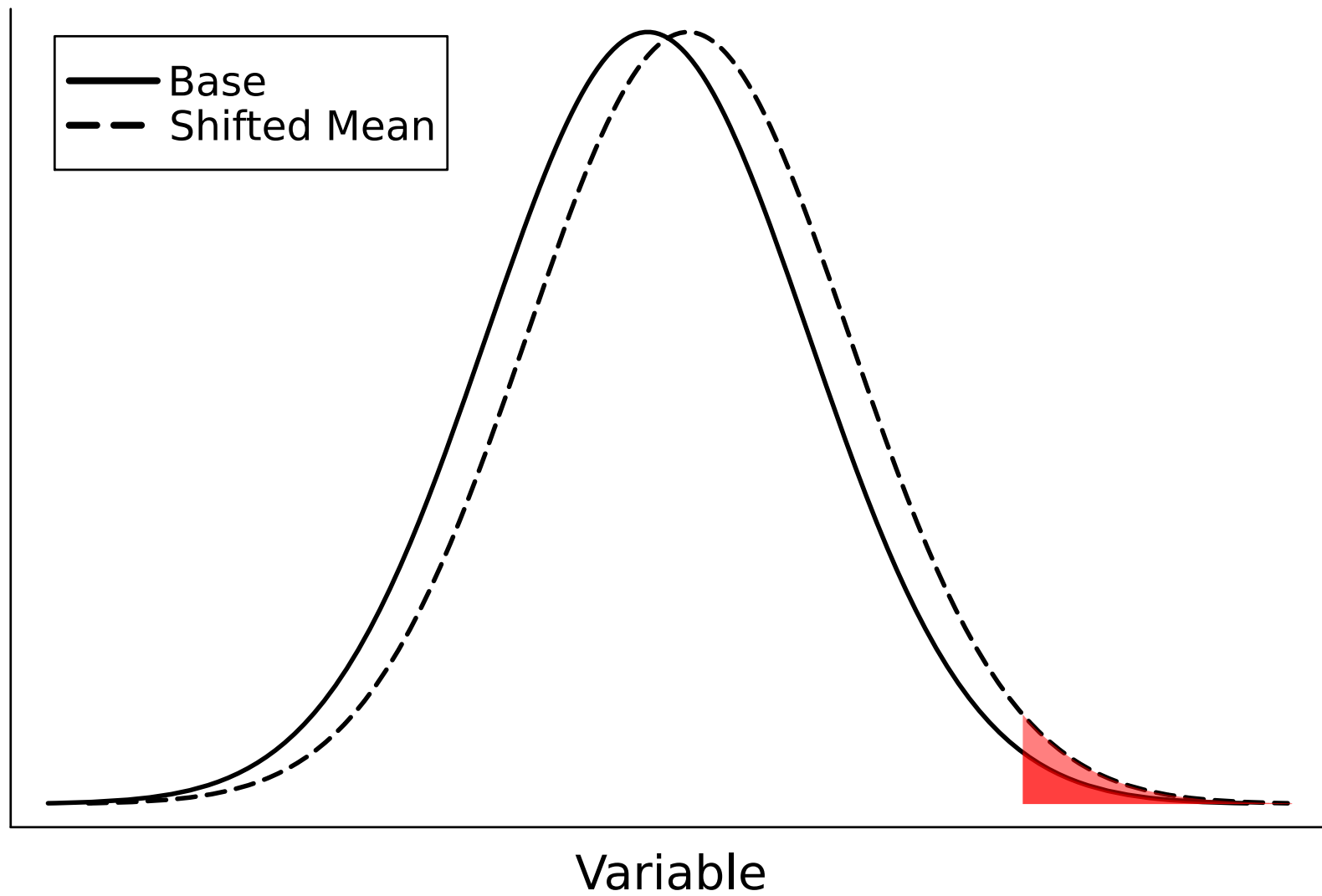
EXTREME VALUE STATISTICS

WHAT DO WE MEAN BY EXTREME VALUES?

Extreme values are those which occur far from the center of a probability distribution, and so are not well-represented by measures of central tendency.

Extreme values are of critical importance for risk analysis. Relatively small changes in underlying distributions can result in substantially-large changes in extremes.

EXAMPLE OF EXTREME VALUE CHANGES

Two Common Questions About Extremes

1. What is the distribution of "block" extremes, *e.g.* annual maxima?

Two Common Questions About Extremes

1. What is the distribution of "block" extremes, *e.g.* annual maxima?

2. What is the distribution of extremes which exceed a certain value?
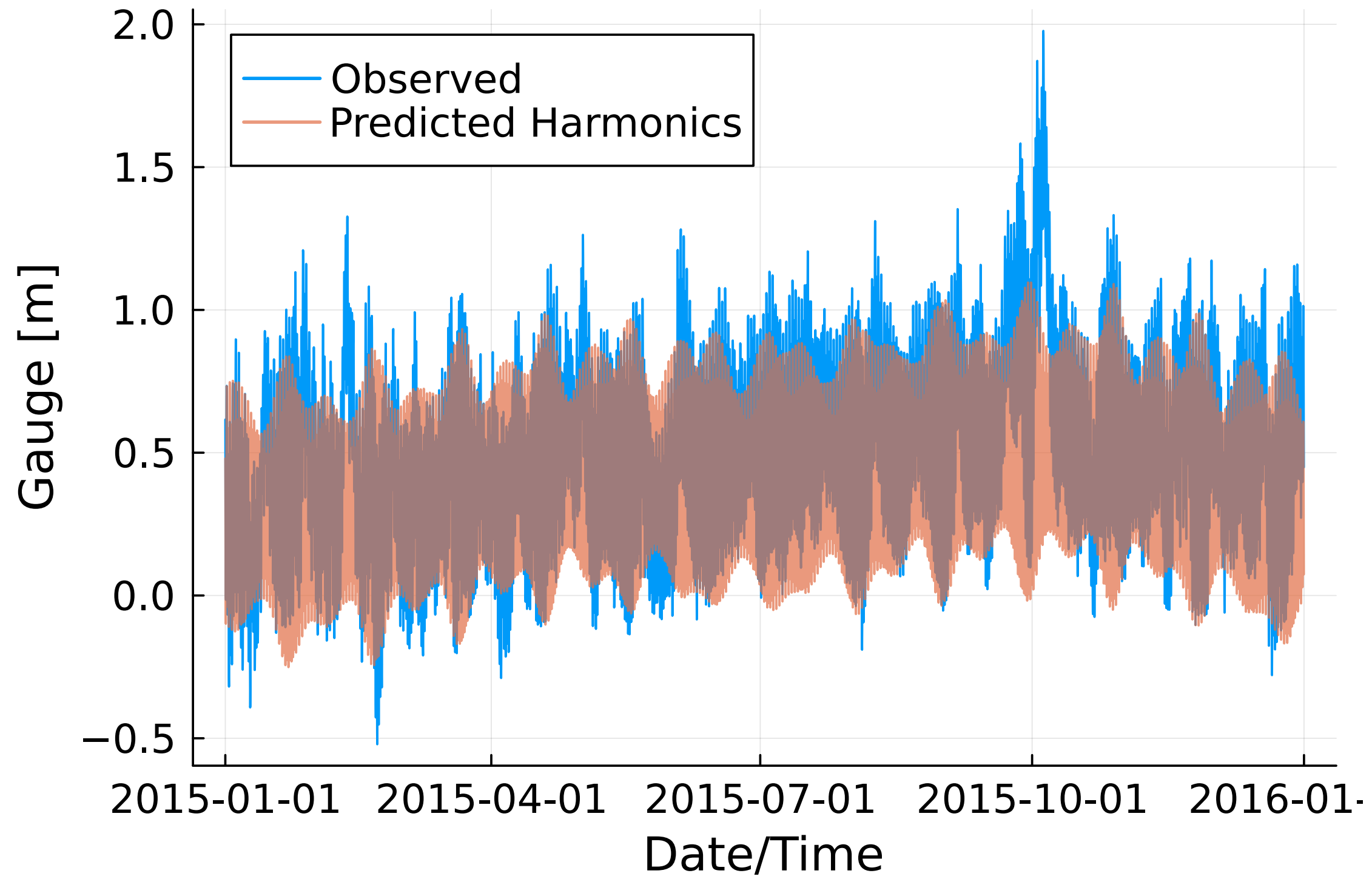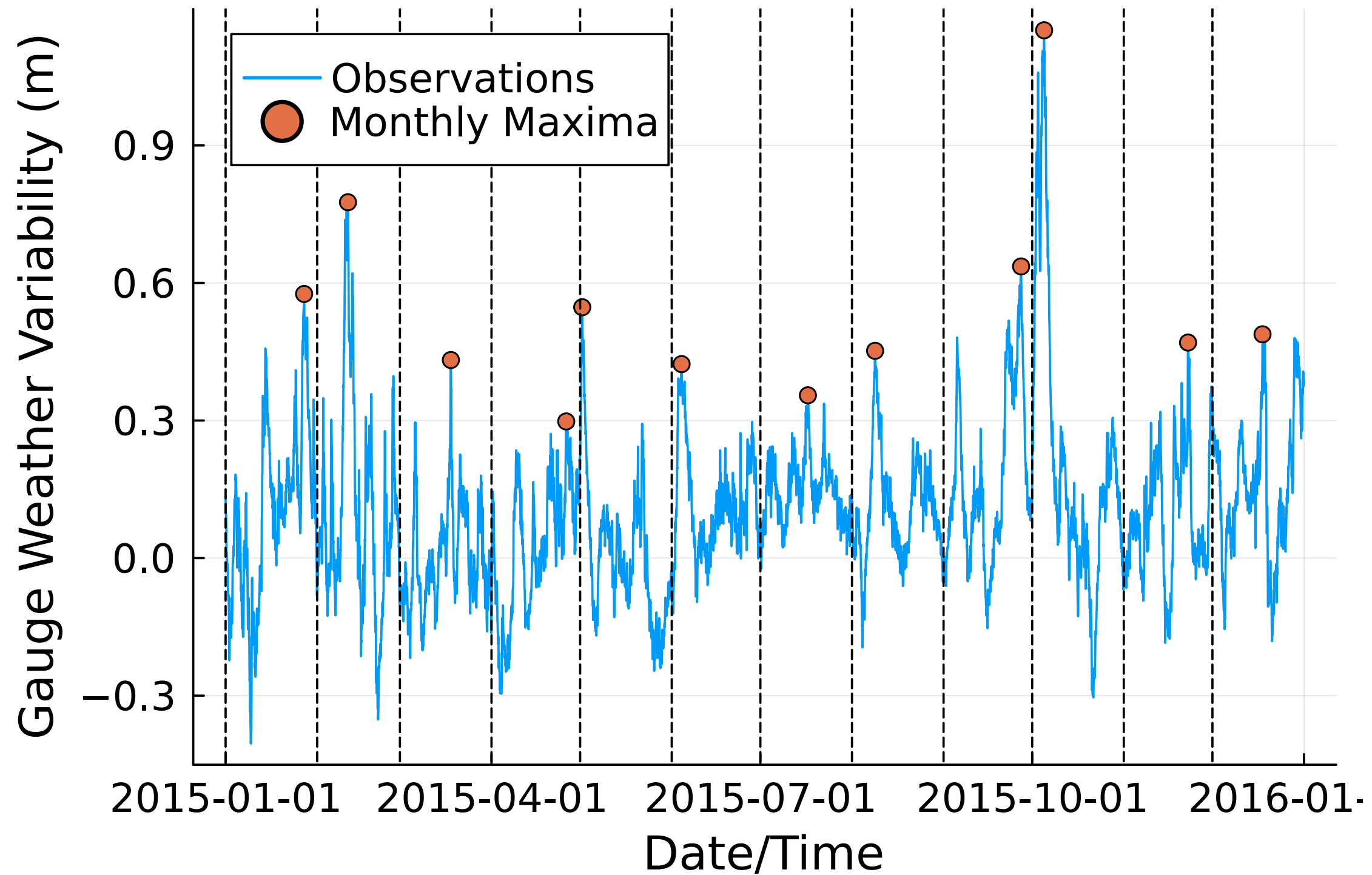
TWO COMMON QUESTIONS ABOUT EXTREMES

1. **What is the distribution of "block" extremes, *e.g.* annual maxima?**

2. What is the distribution of extremes which exceed a certain value?

Example: Monthly Tide Gauge Maxima

EXAMPLE: MONTHLY TIDE GAUGE MAXIMA

BLOCK EXTREMES (MAXIMA)

A different way of framing this question:

Given independent and identically-distributed random variables $X_1, X_2, \ldots, X_{mk}$, what can we say about the distribution of maxima of "blocks" of size m:

$$\tilde{X}_i = \max_{(i-1)m < j \leq im} X_j,$$

for $i = 1, 2, \ldots, k$?

ANALOGY: CENTRAL LIMIT THEOREM

Recall that the **Central Limit Theorem** tells us:

If we have independent and identically-distributed variables $X_1, X_2, \ldots$ from some population with mean μ and standard deviation σ, the sample mean $\bar{X}$ has the approximate distribution

$$\bar{X} \sim \text{Normal}(\mu, \sigma/\sqrt{n}).$$

EXTREME VALUE THEOREM

The **Extreme Value Theorem** (*the stats one, not calculus*!) is the equivalent for block maxima.

If the limiting distribution exists, it can only by given as a **Generalized Extreme Value (GEV)** distribution:

$$H(y) = \exp\left\{-\left[1 + \xi\left(\frac{y - \mu}{\sigma}\right)\right]^{-1/\xi}\right\},$$
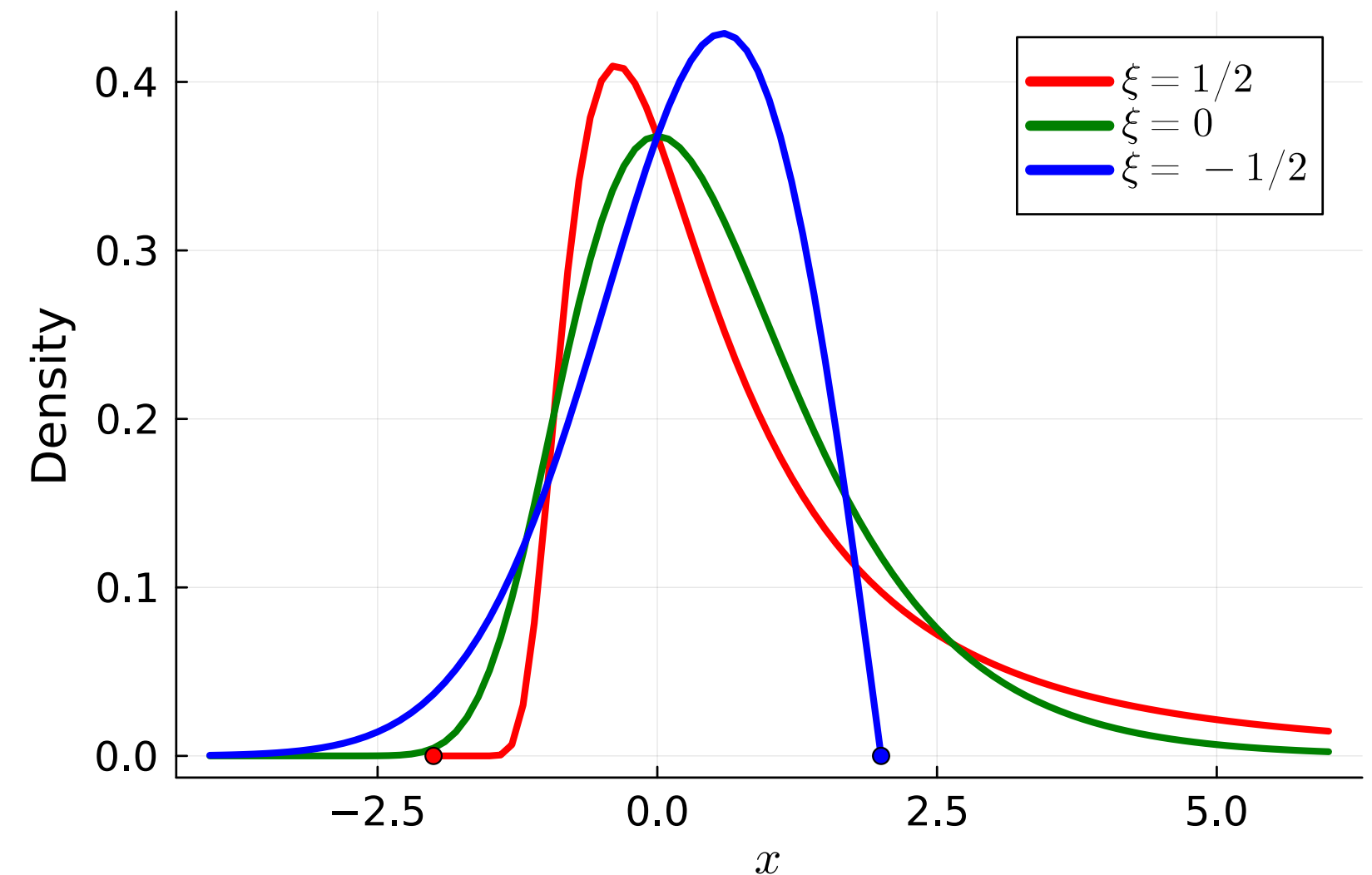
defined for y such that $1 + \xi(y - \mu)/\sigma > 0$.

Generalized Extreme Value Distributions

GEV distributions have three parameters:

- location μ;
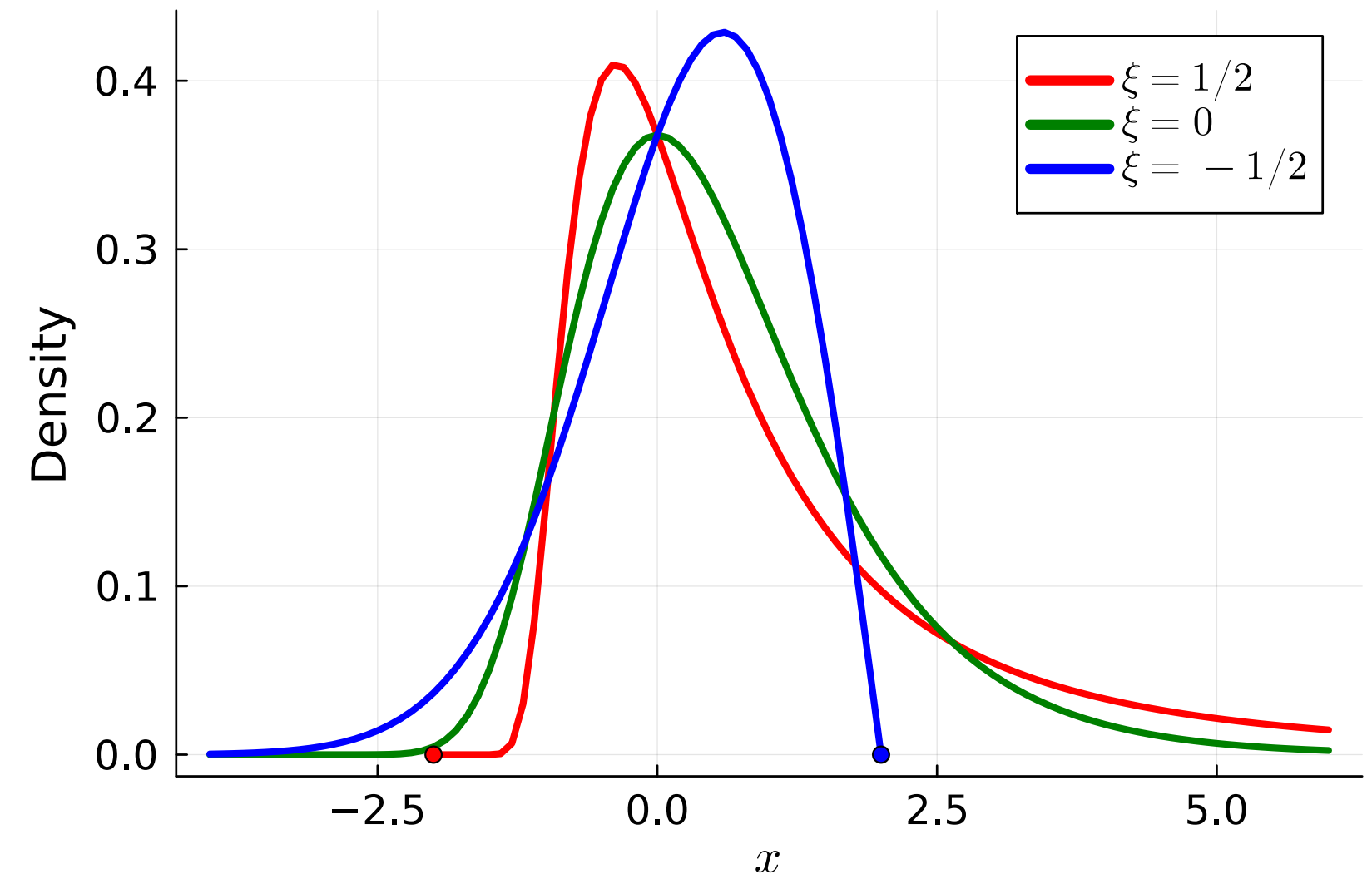- scale $\sigma > 0$;
- shape ξ.

Generalized Extreme Value Distributions

The shape parameter ξ is particularly influential, as the GEV distribution can take on three shapes depending on its sign.

GEV Types

- $\xi > 0$: Frechet (*heavy-tailed*)

- $\xi = 0$: Gumbel (*light-tailed*)

- $\xi < 0$: Weibull (*bounded*)

GEV TYPES
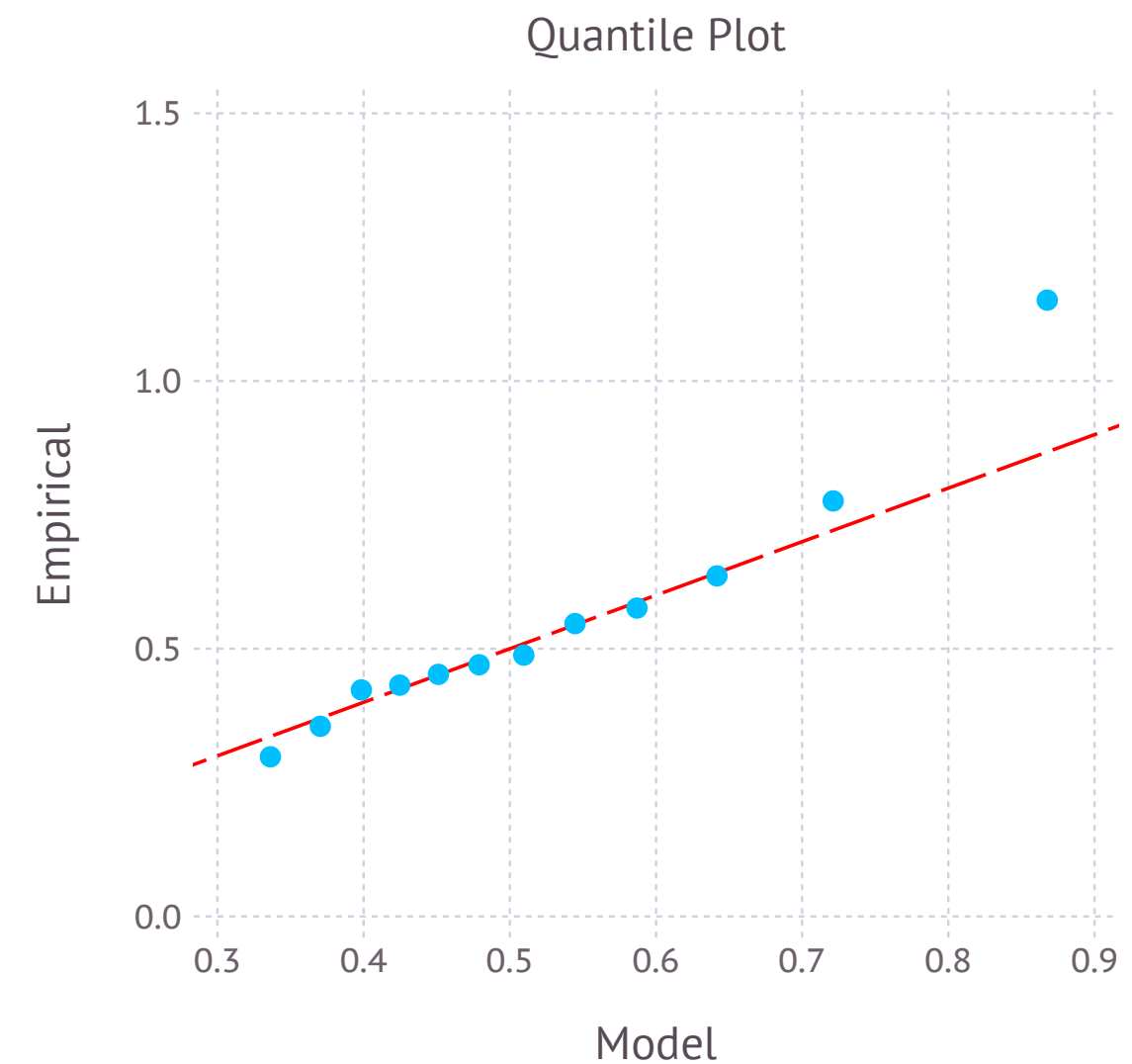
- $\xi < 0$ implies the extremes are *bounded* (the Weibull distribution comes up in the context of temperature and wind speed extremes).

- $\xi > 0$ implies that the tails are *heavy*, and there is no expectation. Common for streamflow, storm surge, precipitation.

- The Gumbel distribution ($\xi = 0$) is common for extremes from normal distributions, doesn't occur often in real-world data.

GEV Fit: Tide Gauge Data
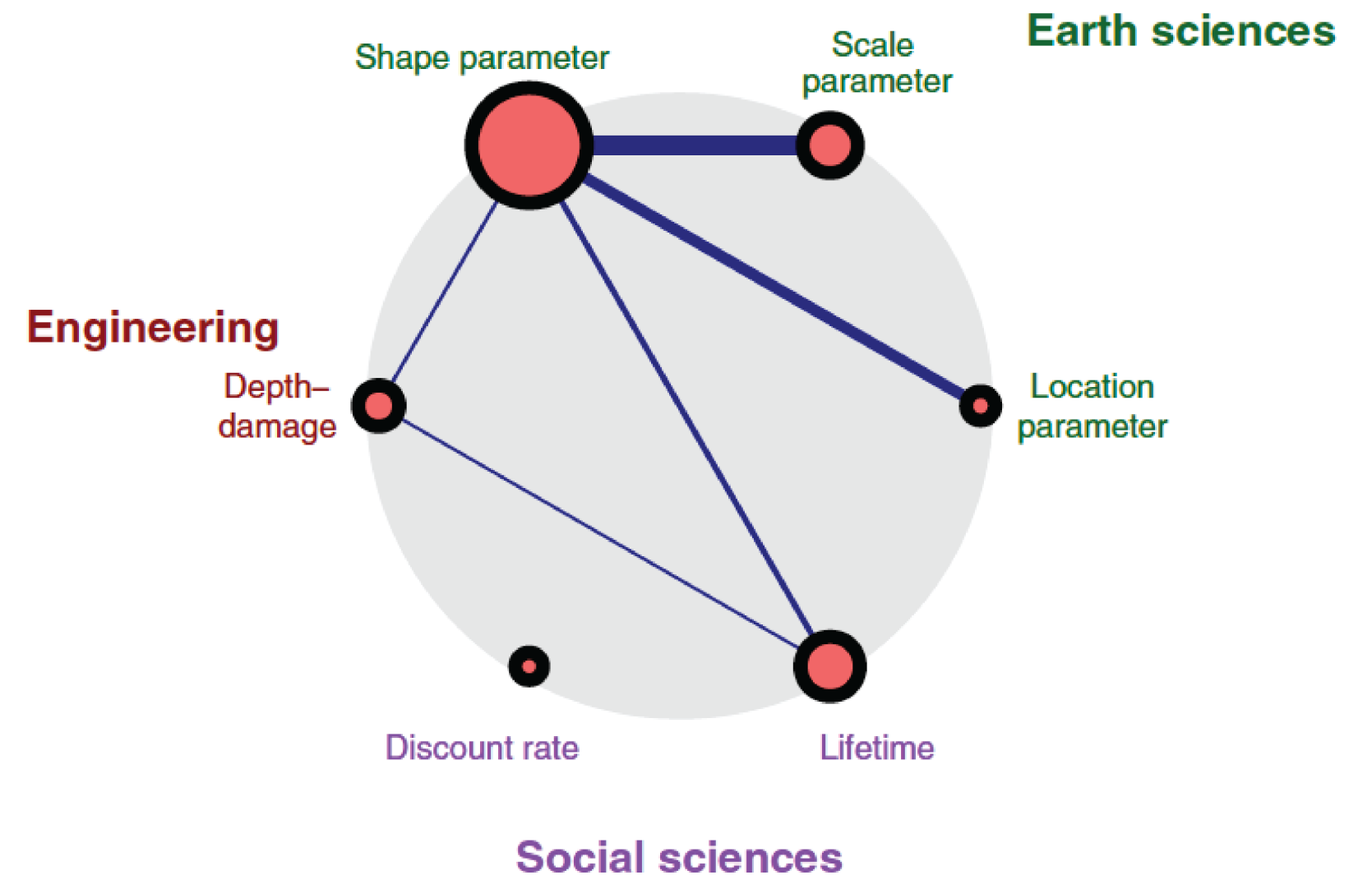
For example, for our tide gauge data, the maximum-likelihood estimate is:

- $\mu = 0.45$;

- $\sigma = 0.13$;

- $\xi = 0.21$.

Quantile Plot

Be Careful About the Shape Parameter!

If $\xi > 0$, risk projections are **extremely** sensitive to its value, as larger values will yield disproportionately larger tail samples.

Source: Zarekarizi et al (2020)

RETURN PERIODS

The **return period** of an extreme value is the inverse of the exceedance probability, *e.g.* a value with an annual exceedance probability of 1% (0.99 quantile) has a 100-year return period ("100-year storm").

RETURN PERIODS

The **return period** of an extreme value is the inverse of the exceedance probability, *e.g.* a value with an annual exceedance probability of 1% (0.99 quantile) has a 100-year return period ("100-year storm").

A major challenge with return periods is that we often don't have enough data to constrain these values, but we can use fitted GEV distributions to estimate the m-year return period by computing the $1 - 1/m$ quantile.

PEAKS OVER THRESHOLDS
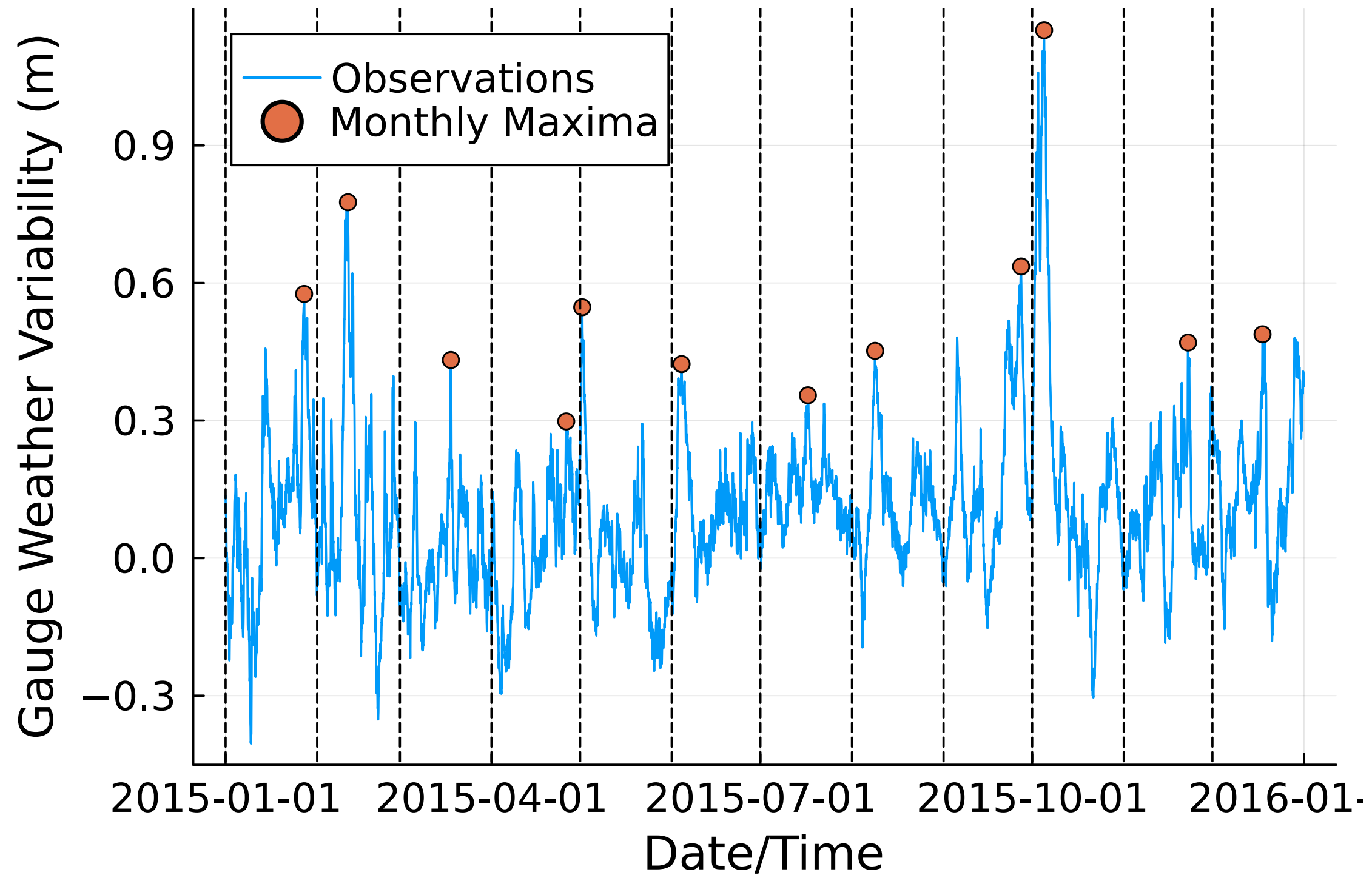
TWO COMMON QUESTIONS ABOUT EXTREMES

1. What is the distribution of "block" extremes, *e.g.* annual maxima?

2. **What is the distribution of extremes which exceed a certain value?**

DRAWBACKS OF BLOCK MAXIMA

The block-maxima approach has two potential drawbacks:

1. Uses a limited amount of data;

2. Doesn't capture the potential for multiple exceedances within a block.
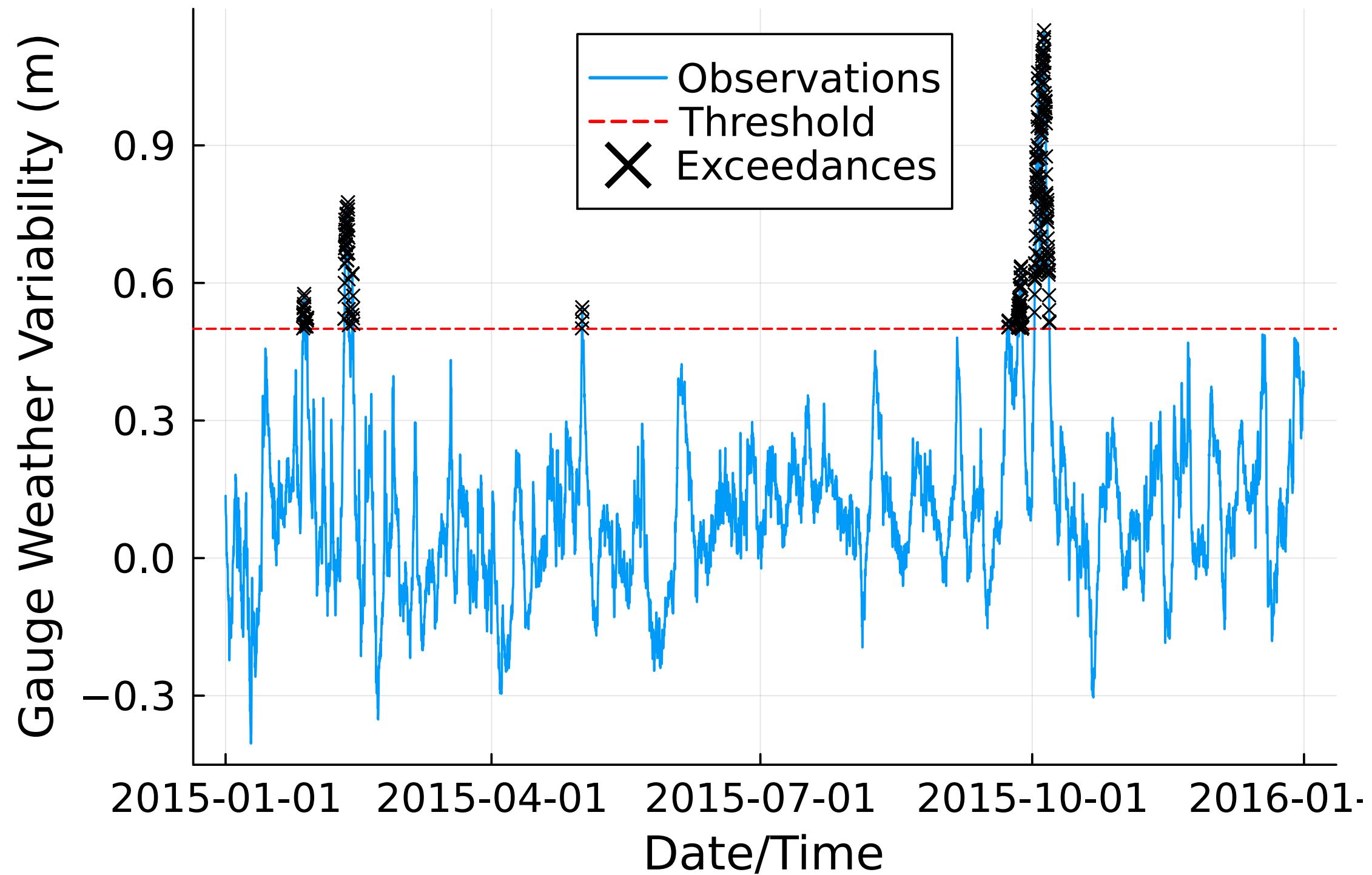
EXAMPLE: MONTHLY TIDE GAUGE MAXIMA

PEAKS OVER THRESHOLDS

An alternative approach is to model the distribution of events where a random variable $X_1, X_2, \ldots$ exceeds a sufficiently high threshold u.

Consider the **conditional excess distribution function**

$$F_u(y) = \mathbb{P}(X > u + y | X > u),$$

which is the cumulative distribution of values by which X exceeds u (given that the exceedance has occurred).

PEAKS OVER THRESHOLDS EXAMPLE

PEAKS OVER THRESHOLDS

It turns out that, for a large number of underlying distributions of X (including most of the typical ones, such as normal and log-normal), $F_u(y)$ is well-approximated by a **Generalized Pareto Distribution (GPD)**:

$$F_u(y) \to G(y) = 1 - \left[1 + \xi \left(\frac{y - \mu}{\sigma} \right)^{-1/\xi} \right],$$

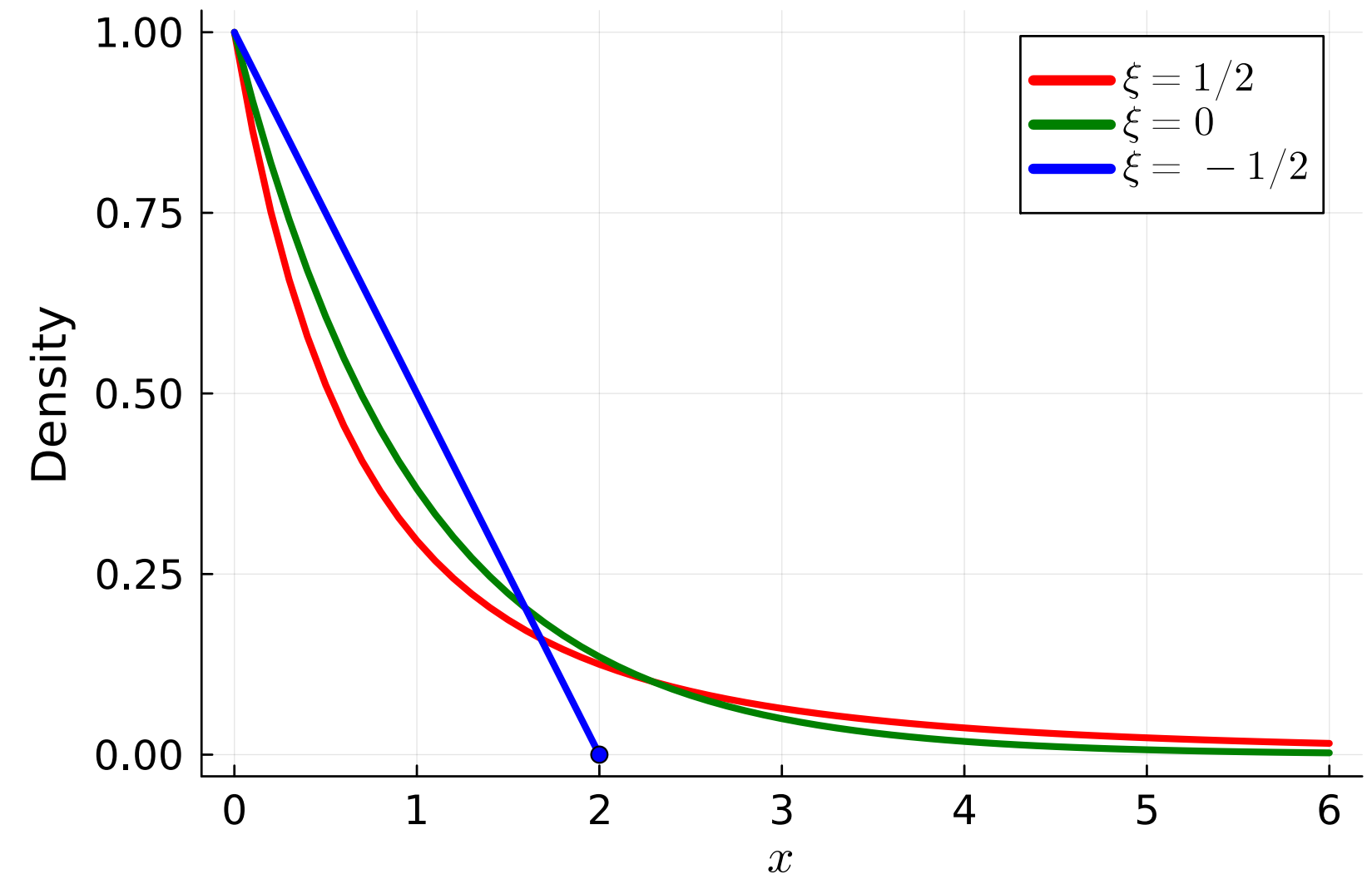defined for y such that $1 + \xi(y - \mu)/\sigma > 0$.

GENERALIZED PARETO DISTRIBUTIONS

Similarly to the GEV distribution, the GPD distribution has three parameters:

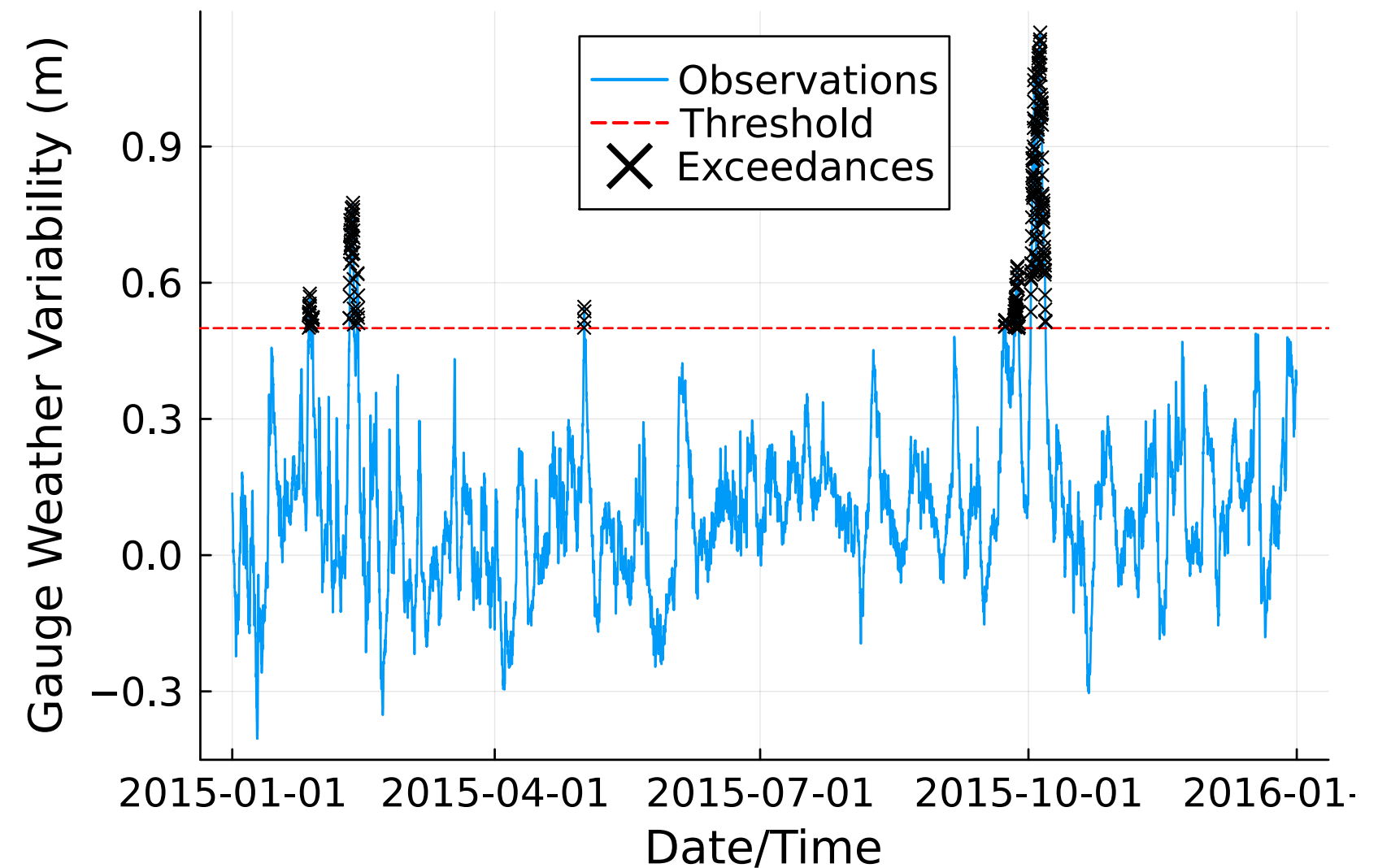- location μ;

- scale $\sigma > 0$;

- shape ξ.

Generalized Pareto Distributions Types

- $\xi > 0$: *heavy-tailed*

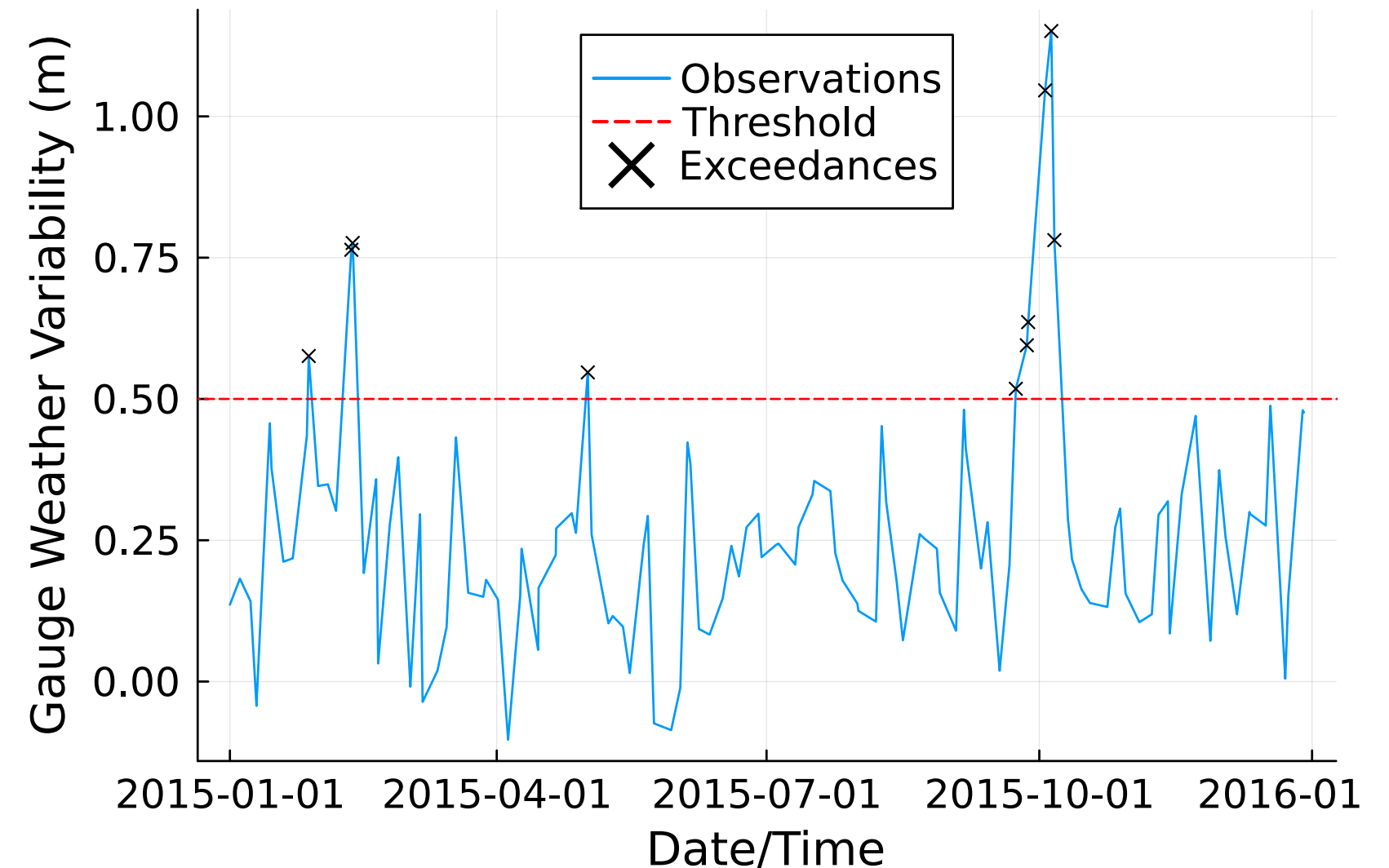- $\xi = 0$: *light-tailed*

- $\xi < 0$: *bounded*

PEAKS OVER THRESHOLDS EXAMPLE

Note that exceedances can occur in clusters due to the same meteorological forcing: this violates the assumption of independence.

PEAKS OVER THRESHOLDS EXAMPLE
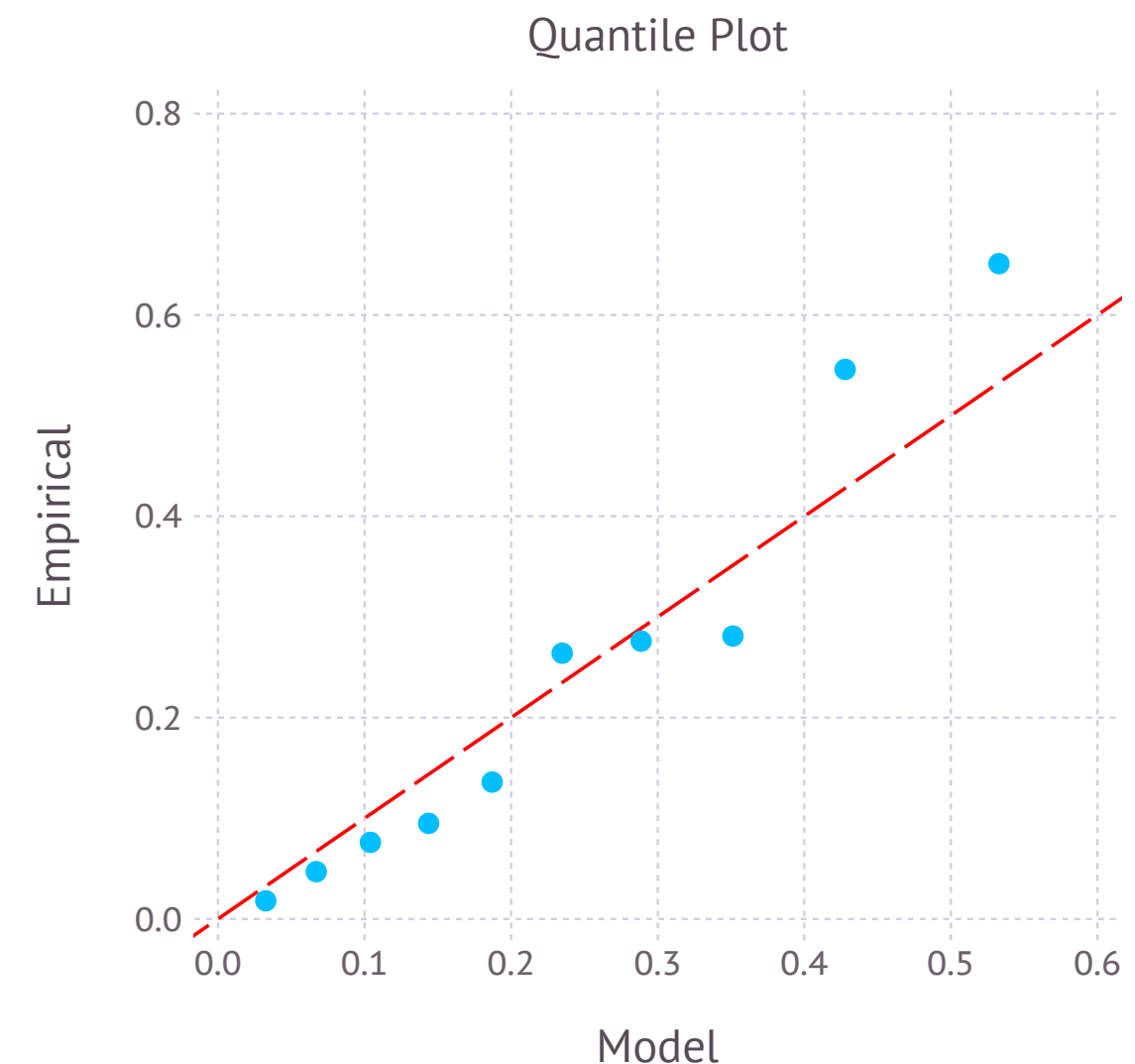
However, as Arns et al (2015) note, there is no clear declustering time period to use: need to rely on physical understanding of events and "typical" durations.

GP FIT: TIDE GAUGE DATA

For a threshold of 0.5m for the weather contribution of our tide gauge data, the maximum-likelihood estimate is:

- $\mu = 0$ (assumed);

- $\sigma = 0.35$;

- $\xi = -0.41$.

Quantile Plot

GP Fit Depends on the Threshold!

Note that the GP fit is for the *amount by which the distribution exceeds the threshold* (usually σ changes).

So a new threshold means a new fit is required.

Selecting a Threshold

Selecting a threshold requires some careful thought (and hopefully is decision-relevant!).

Too high: Not many exceedances, estimator won't be great.

Too low: Too many exceedances, distribution will be poorly approximated by a GP.

Poisson-GP Processes

Peaks over Thresholds models are often paired with a **Poisson process**, which models the number of times an event occurs using a Poisson distribution,

$$n \sim \text{Poisson}(\lambda_u).$$

Then, for each $i = 1, \ldots, n$, sample

$$X_i \sim \text{GeneralizedPareto}(u, \sigma, \xi).$$

RETURN LEVELS FOR PP-GP PROCESSES

Return levels and periods for these processes need to account for both the *rate* of exceedance λ_u and the distribution of peaks over the threshold:

$$\mathbb{P}(X \leq x) = 1 - \lambda_u \left[1 + \xi \left(\frac{x - u}{\sigma} \right) \right]^{-1/\xi},$$

so a return level corresponding to return period s is obtained by setting this equal to $1 - 1/s$ and solving for x.

KEY TAKEAWAYS

KEY TAKEAWAYS

- Risk analysis is often concerned with extremes (*e.g.* occurrence of storm tides).

- Extreme values can be modeled as block maxima or peaks-over-thresholds.

- Block maxima: Generalized Extreme Value distributions.

- Peaks-Over-Thresholds: Generalized Pareto distributions (plus maybe Poisson processes).

- Statistical models are highly sensitive to details: shape parameters ξ, thresholds u, etc.

- **Models assume independent variables.**

WHAT WE HAVEN'T DISCUSSED

- What happens if these extremes change? This is typically how climate change impacts systems. This is the world of **nonstationarity**, which we will discuss later.

WHAT WE HAVEN'T DISCUSSED

- What happens if these extremes change? This is typically how climate change impacts systems. This is the world of **nonstationarity**, which we will discuss later.

- Multivariate extremes are difficult: what does this even mean?

  - Often require **copulas** to "glue" distributions together. We might discuss this if we have time.

Upcoming Schedule

Upcoming Schedule

Wednesday: Lab on extreme value distributions.

Next Monday: Nonstationarity and hypothesis testing.