# Return period maps for floods

24th May, 2023

Arfima
Financial Solutions
Calle José Bardasano Baos 9, 7
28016 Madrid, España
www.afs-services.com
+34 91 129 2874

ENERGYECOLAB
Universidad Carlos III de Madrid
Calle Madrid, 126,
28903 Madrid, España

# 1

# Data sources for hazards

## 1.1 Introduction

Dottori et al. (2022) provide a comprehensive description of the methodology regarding the creation of return period maps for flood events in Europe. Building upon this example, we aim to outline the procedure related to this problem.

## 1.2 Statistical basics of flood frequency analysis

The objective of flood frequency analysis is to estimate the frequency of floods at a given location. This procedure typically involves the following steps:

1. Collect the relevant data for the analysis.
2. Select a suitable probabilistic distribution that represents the frequency of flood events.
3. Choose a method to estimate the parameters of the selected distribution.
4. Calculate the reliability of the chosen parameters based on the available data.
5. Estimate the quantile for a given return period

To conduct the statistical analysis in the context of floods, we typically require a dataset of flood peak observations. These observations represent the highest recorded flood levels or flows at a specific location during a specific timeframe. This data can be obtained from various sources including hydrological and meteorological agencies, climate databases, and research studies.

Regarding the second step, the most common choices for frequency distributions representing flood peaks include:

- Generalized Extreme Value distribution (GEV)

$$f(x) = \frac{1}{a}\left[1 - k\left(\frac{\bar{z}-\mu}{a}\right)\right]^{\frac{1}{k}-1} \exp\left\{-1 - k\left(\frac{\bar{z}-\mu}{a}\right)^{\frac{1}{k}}\right\}$$

- Three-parameter log-normal distribution (LN3)

$$f(x) = \frac{1}{a\sqrt{2\pi}} \exp\left[-\log\left\{\frac{1-k(\bar{z}-\mu)}{a}\right\} - \frac{1}{2}\left[-\frac{1}{k}\log\left\{\frac{1-k(\bar{z}-\mu)}{a}\right\}\right]^2\right]$$

- Generalized Logistic distribution (GLO)

$$f(x) = \frac{1}{a}\left[1 - k\left(\frac{\bar{z}-\mu}{a}\right)\right]^{\frac{1}{k}-1}\left[1 + \left\{1 - k\left(\frac{\bar{z}-\mu}{a}\right)\right\}^{\frac{1}{k}}\right]^{-2}$$

- Gumbel distribution (GUM)

$$f(x) = \frac{1}{a}\exp\left[-\left(\frac{\bar{z}-\mu}{a}\right) - \exp\left(-\frac{\bar{z}-\mu}{a}\right)\right]$$

where $f(x)$ is a probability density function of each distribution.

In the context of parameter estimation, there are several commonly used methods that can be applied:

- Method of Moments (MOM)
- Probability Weighted Moments (PWM)
- Maximum Likelihood (ML) method
- LMoments (LM)

The performance of the model is assessed by comparing it with available data using suitable accuracy measures. Common accuracy measures for evaluating the fit include:

- Standard Error (SE) of parameter estimates
- Goodness-of-Fit (GOF) tests
- Accuracy measure methods

In the final step, our goal is to estimate the quantile, denoted as $z_T$, for a given return period $T$. This can be achieved using the following equation:

$$z_T = \phi\left(1 - \frac{1}{T}\right)$$

Here, $\phi$ represents the inverse function of the cumulative probability distribution $F$ that we previously estimated.

If the distribution cannot be expressed in inverse form as $z_T = \phi\left(1 - \frac{1}{T}\right)$, then we employ numerical methods to evaluate $z_T$ for a given value of $F$, relying on numerical relationships between $z_T$ and $F$.

It is important to acknowledge that there is inherent uncertainty in estimating flood magnitude. To quantify this uncertainty, one approach is to estimate the standard error (SE) of flood estimates and construct, for instance, 90% confidence intervals of flood quantiles for different return periods using the parametric bootstrap method.

## 1.3 Methodology overview

The approach used to generate the flood datasets can be summarized as follows. In our presentation, we will follow Dottori et al. (2022). The hydrological input data for flood simulations consists of daily river flow data spanning the years 1990-2016, obtained from the hydrological model LISFLOOD using interpolated daily meteorological observations.

The LISFLOOD simulations require a number of static input maps such as:

- land cover,
- a digital elevation model (DEM),
- a drainage network,
- soil parameters,
- the parameterization of reservoirs.

From the river flow data, three key outputs are derived:

- **Frequency distributions**
  Frequency distributions refer to the statistical distribution of flood events based on their occurrence or frequency. These distributions provide information about the probability or likelihood of different flood events happening within a given time period or at a specific location.

- **Peak discharges**
  Peak discharge refers to the maximum volume or rate of water flow in a river or stream. It represents the highest level of water discharge reached and is used to assess flood risk, design flood control measures, and develop flood management strategies.

- **Flood hydrographs**
  Flood hydrographs refer to graphical representations of the water flow characteristics in a river or stream over a specific period of time. They illustrate the relationship

between time and the magnitude of the flood, showing how the water level or discharge changes over time.

The flood hydrographs are then utilized to simulate the flooding processes at a local scale using the LISFLOOD-FP hydrodynamic model. Subsequently, a validation exercise is conducted, and different approaches and input datasets are compared.

## 1.4 Input to flood simulations - LISFLOOD model

The LISFLOOD model is a hydrological-physical rainfall-runoff model available at: https://ec-jrc.github.io/lisflood/. We use it to perform a long-term hydrological simulation from 1990 to 2016 at a 5 km grid spacing and daily resolution. This simulation provides the hydrological input data for the flood simulations. The long-term run of LISFLOOD relies on gridded meteorological maps generated by interpolating meteorological observations from stations and precipitation datasets. In addition to meteorological data, LISFLOOD simulations require static input maps such as land cover, a digital elevation model (DEM), a drainage network, soil parameters, and reservoir parameterization.

The streamflow dataset, obtained from the long-term run of LISFLOOD, is utilized in the following manner:

- We extract annual maxima for each pixel of the river network from the streamflow dataset covering the period 1990-2016.
- The Gumbel distribution is selected as the frequency distribution for flood events.
- The parameters of the Gumbel distribution are estimated using the L-moments approach.
- This process is repeated for various return periods, including 10, 20, 50, 100, 200, and 500 years.

Additionally, we calculate a flow duration curve (FDC) and the synthetic flood hydrographs from the streamflow dataset. In the context of floods, a flow duration curve is a graphical representation of the cumulative distribution of flow rates or discharge over a specific period. It shows the percentage of time that a certain flow rate or discharge is equaled or exceeded. The curve is typically plotted with flow rate or discharge on the y-axis and the corresponding percentage of time on the x-axis.

## 1.5 Flood simulation

In the subsequent step, the output of the LISFLOOD model is utilized in the flood simulations. These simulations involve running local flood scenarios along the entire river network, following a methodology similar to Alfieri et al. (2014).

The flood simulation employs the two-dimensional hydraulic model LISFLOOD-FP, with each 100m section of the river being considered. The simulation incorporates the CCM DEM (Digital Elevation Model) as elevation data, synthetic hydrographs as the hydrological input, and a combination of CORINE Land Cover for 2016 (from the Coordination of Information on the Environment; Copernicus Land Monitoring Service, 2017) and Copernicus GLOBCOVER for 2009 (Bontemps et al., 2011) to estimate the friction coefficient based on land use.

Subsequently, flood maps corresponding to the same return period are merged to generate continental-scale flood hazard maps. Additionally, the dataset includes a separate map of the 100m river network, which delineates the water courses considered in creating the flood hazard maps.

## 1.6 Validation

In the last step we want to preform the validation of the generated dataset with official datasets.

### 1.6.1 validation areas and maps

To validate large-scale flood hazard maps, it is necessary to employ benchmarks comprising one or more datasets that possess similar extent and accuracy as the modeled maps.

In Europe, both EU member states and the UK have created national datasets containing flood hazard maps for various flood probabilities, typically indicated by the flood return period, in compliance with the EU Floods Directive (EC, 2007) guidelines. These maps are typically generated using multiple hydrodynamic models of different complexities (AdB Po, 2012) based on high-resolution topographic and hydrological datasets. While it is acknowledged that official maps may have errors or be incomplete (Wing et al., 2017), they are expected to offer greater accuracy compared to the modelled maps presented in this study. As a result, the official maps have been chosen as the reference maps for validation purposes.

### 1.6.2 Performance metrics and validation procedure

We assess the performance of the simulated flood maps by comparing them to reference maps using several indices proposed in the literature (Bates and De Roo, 2000; Alfieri et al., 2014; Dottori et al., 2016a; Wing et al., 2017).

Let $F_0$ represent the total observed flooded area, and $F_m$ denote the area flooded as predicted by the model. The validation metrics employed include:

- The hit rate (HR)

$$HR = \frac{F_m \cap F_0}{F_0} \cdot 100$$

where $F_m \cap F_0$ stand for area correctly predicted by the model. The HR indicator

- false-alarm rate (FAR)

$$FAR = \frac{F_m \backslash F_0}{F_m} \cdot 100$$

where $F_m \backslash F_0$ stand for area wrongly predicted by the model.

- critical-success index (CSI)

$$CSI = \frac{F_m \cap F_0}{F_m \cup F_0} \cdot 100$$

where $F_m \cap F_0$ is the union of observed and simulated flooded areas.

These metrics provide insights into the accuracy and reliability of the model's flood predictions when compared to the reference maps.