Structured Coupling of Probability Loss Distributions; Assessing the Flood Risk in Multiple River Basins

Anna Timonina, Ph.D.

International Institute for Applied Systems Analysis (IIASA); Risk, Policy and Vulnerability Program
Almost all European regions and, moreover, countries are very much subject to the flood risk;

The flood events have direct impact on the financial strength of affected countries, leading often to increasing taxes, additional public debts and budget diversions;

Though some loss distribution estimates are available for each individual country, a more global European flood loss estimate is not available yet;

Nevertheless, it is necessary to study the interdependency of flood losses between European countries, in order to address the whole complexity and to realistically assess the flood risk in Europe;

The study of flood interdependencies is also necessary for the purposes of the European Solidarity Fund.
River basins in Europe

Figure: River basins in Europe.

a. River basins, loss areas.

b. Tree structure of river basins.
How to study the interdependency between different pairs of basins/regions/countries?
Theory: Coupling Basins in Europe

Consider for simplicity just two river basins/regions with losses $L_1$ and $L_2$ correspondingly. The total loss in these river basins together is, clearly, $L = L_1 + L_2$. Suppose, that marginal densities for the basins are $f_1$ and $f_2$ respectively.

1. **Independent basins:**
   In case the basins/regions and, hence, losses, are independent, the density of the total loss $L_1 + L_2$ can be received by the convolution of marginals:
   \[ f(x) = f_1 \ast f_2 = \int f_1(x - y)f_2(y)dy. \]

2. **Dependent basins:**
   In case the basins/regions and, hence, losses, are dependent, the density of the total loss $L_1 + L_2$ can be received by the convolution over the copula $C(\cdot)$:
   \[ f(x) = f_1 \ast_C f_2 = \int c(F_1(x - y), F_2(y)) f_1(x - y)f_2(y)dy. \]
Up-scaling of losses over different levels

Figure: Up-scaling of losses over different levels.
**Choice of the Copula Type**

For the purposes of the flood risk analysis, we should find and use a copula type, that has the following properties:

1. The chosen copula $C_{\theta_{ij}}(F_i, F_j)$ should well describe the flood loss behavior: the interdependency between basins $i$ and $j$ is high for large flood events (i.e. $\theta_{i,j}$ is large) and is low for small events (i.e. $\theta_{ij}$ is small), i.e. if the flood event is rare and strong, it will most likely influence several basins and, hence, the interdependencies should be considered. At the same time, if the flood event is small, it is likely to influence only one basin or small number of basins;

2. It should be convenient for the fat tail estimation and for the study of tail interdependencies;

3. Coupling copulas of the same type should again produce the copula of the same type: this property does not hold for all types of copulas, but it is very important and useful for the study of dependencies between regions, for which the copulas are already obtained.
**Flipped Clayton Copula**

In our analysis we introduce the *Flipped* Clayton Copula, which density function for parameter $\theta > 0$ can be written as

$$C_\theta(u, v) = u + v - 1 + \left[(1 - u)^{-\theta} + (1 - v)^{-\theta} - 1\right]^{-\frac{1}{\theta}}$$  \hspace{1cm} (1)

and that satisfies all the necessary properties (1)-(3) and well describes the flood loss behavior.

*Figure:* Flipped Clayton CDF.
Estimation of the Copula Parameter (ML Estimation)

Suppose there is a sample $u_1, u_2, ..., u_N$ of $N$ independent and identically distributed observations. The likelihood function for the observations is

$$L(\theta) = \prod_{i=1}^{N} c_{\theta}(u_i, v_i),$$

and the log-likelihood is

$$\ln L(\theta) = \sum_{i=1}^{N} \ln c_{\theta}(u_i, v_i)$$

Maximizing the log-likelihood function, we get the parameter $\theta$, on which it depends, i.e.

$$\theta \in \arg \max_{\theta} \ln L(\theta).$$
Why geographical distance is a bad dependency measure? (Romania example)

a. Pair of basins along the Danube.

b. Copula for basins along the Donube.

Figure: The dependency is high and is equal to $\theta = 9.9999$, though the distance between the midpoints of the basins is large and equals $d = 242\, km$. 

Anna Timonina, Ph.D. (International Institute for Applied Systems Analysis (IIASA); Risk, Policy and Vulnerability Program)
Why geographical distance is a bad dependency measure? (Romania example)

a. Pair of basins in Romania not along the same river.

b. Copula for basins belonging to different rivers.

Figure: The dependency is low and is equal to $\theta = 1.5684$, though the distance between the midpoints of the basins is small and equals $d = 64km$. 

Anna Timonina, Ph.D. (International Institute for Applied Systems Analysis (IIASA); Risk, Policy and Vulnerability Program)
How to study the interdependency between multiple basins/regions/countries?
**Coupling multiple dependent loss distributions**

If we consider a set of basins \( I \in \{1, 2, ..., N\} \) and estimate the Flipped Clayton Copula between every pair of basins, we receive the following matrix

\[
\Theta = \begin{pmatrix}
\theta_{11} & \cdots & \theta_{1N} \\
\vdots & \ddots & \vdots \\
\theta_{N1} & \cdots & \theta_{NN}
\end{pmatrix} = \begin{pmatrix}
1 & \cdots & \theta_{1N} \\
\vdots & \ddots & \vdots \\
\theta_{N1} & \cdots & 1
\end{pmatrix}.
\]

*Figure:* Multivariate Flipped Clayton Copula for some basins in Romania.
Structured Coupling

*Figure:* River structure.

1. Multidimensional coupling;
2. Hierarchical coupling;
3. Ordered coupling.
Hierarchical coupling

For hierarchical coupling we need to estimate copula $\bar{C}$ suitable for coupling of copulas $C_{1,2}$ and $C_{3,4}$:

$$C(x_1, x_2, x_3, x_4) = \bar{C}(C_{1,2}(x_1, x_2), C_{3,4}(x_3, x_4)).$$

In this case, $C_{1,2}$, $C_{3,4}$ and $\bar{C}$ may be copulas of different types. But we focus on Flipped Clayton Copulas.

Hierarchical copulas follow a tree structure, but are difficult to estimate, especially if the topology of the tree has to be estimated as well.
Ordered coupling

Suppose, that $L_1$ influences $L_2$, that influences $L_3$, that influences $L_4$. Hence, the tree structure is the following:

In this case, one could estimate 2-dimensional copulas $c_{1,2}$, $c_{2,3}$ and $c_{3,4}$ and combine them in the following way to the 4-dimensional copula density:

$$c(x_1, x_2, x_3, x_4) = c_{1,2}(x_2|x_1) \cdot c_{2,3}(x_3|x_2) \cdot c_{3,4}(x_4|x_3).$$
Ordering of Copula Parameters

We use the minimax approach in order to choose $N$ pairs of basins from the matrix $\Theta$ and we show, that this method adequately represents the complexity of the tree structure and all the underlying interdependencies between basins.

Consider the following matrixes:

$$
\Theta = \begin{pmatrix}
\theta_{11} & \cdots & \theta_{1N} \\
\vdots & \ddots & \vdots \\
\theta_{N1} & \cdots & \theta_{NN}
\end{pmatrix}, \quad C = \begin{pmatrix}
c_{11} & \cdots & c_{1N} \\
\vdots & \ddots & \vdots \\
c_{N1} & \cdots & c_{NN}
\end{pmatrix}, \quad D = \begin{pmatrix}
d_{11} & \cdots & d_{1N} \\
\vdots & \ddots & \vdots \\
d_{N1} & \cdots & d_{NN}
\end{pmatrix}
$$

where $\Theta$ is the matrix of Flipped Clayton Copula parameters $\theta_{i,j}$ between basins $i$ and $j$; $C$ is the matrix of correlations with elements $c_{i,j}$ that represent correlations between basins $i$ and $j$; $D$ is the matrix of geographical distances between basins $i$ and $j$. 

Anna Timonina, Ph.D. (International Institute for Applied Systems Analysis (IIASA); Risk, Policy and Vulnerability Program)
Algorithm for ordering vector $\theta$ out of matrix $\Theta$

1. Initially we choose the maximal $\theta_{i,j}$ from the whole matrix $\Theta$ - this element refers to two basins $(i$ and $j)$ that are most dependent in the sense of flooding. So we couple them at first. At this moment our Order $= [i, j] = [j, i]$;

2. Then we should choose the next element (i.e. basin) $k$, that is suitable for both basins $i$ and $j$ (so that it is dependent not only on basin $i$, but also on basin $j$) (notice, that $k \neq i$, $k \neq j$). Therefore, we need to guarantee that the dependencies between pair $(i, j)$ and each of the basins $1, \ldots, N$ is greater than some values and then maximize over these values ($\textit{minimax}$);

3. Continue iterations until the vector of length $N$ is obtained.
Example: Minimax Ordering

Consider basins A, B, C, D, that are connected to each other by the following structure (the number on the edge refers to $\theta_{i,j}$):

According to the minimax approach, the order for coupling is $A, B, D, C$ here.
Comparison of Ordering Techniques (Romania)

Applying these orders on the matrix $\Theta$ we receive different possible vectors $\theta$, that should represent matrix $\Theta$, i.e. we get $\theta$, $\theta_c$ and $\theta_d$. Sampling multivariate Flipped Clayton Copulas with parameters $\theta$, $\theta_c$ and $\theta_d$ and fitting flood loss distributions to the received samples, we obtain three possible joint loss distributions.

**Figure:** Comparison of different ordering techniques for Romania.
Minimax v.s. Ordering (Part of the river structure)

The geographical distance strongly overestimates the flood losses. The question is now, if the minimax technique adequately represents the flood losses. To answer this question, we compare two loss curves: one, received based on the minimax approach, and the other, received, based on the geographical tree structure of rivers in Romania.

Figure: Two groups of basins in Romania and their loss curves.
Minimax v.s. Ordering (Independent coupling of trees)

Now, we apply independent convolution on the groups of basins (trees), that are not connected to each other through the river.

Figure: Comparison of different ordering techniques for Romania.
Thank you for your attention!