Flood risk model

2021/2022 II. félév

Content

- Introduction: extreme value models
- Algorithm of the flood simulator
- Some simulation results
  - Effect of the most important settings
  - The most dangerous areas

Distribution of exceedances

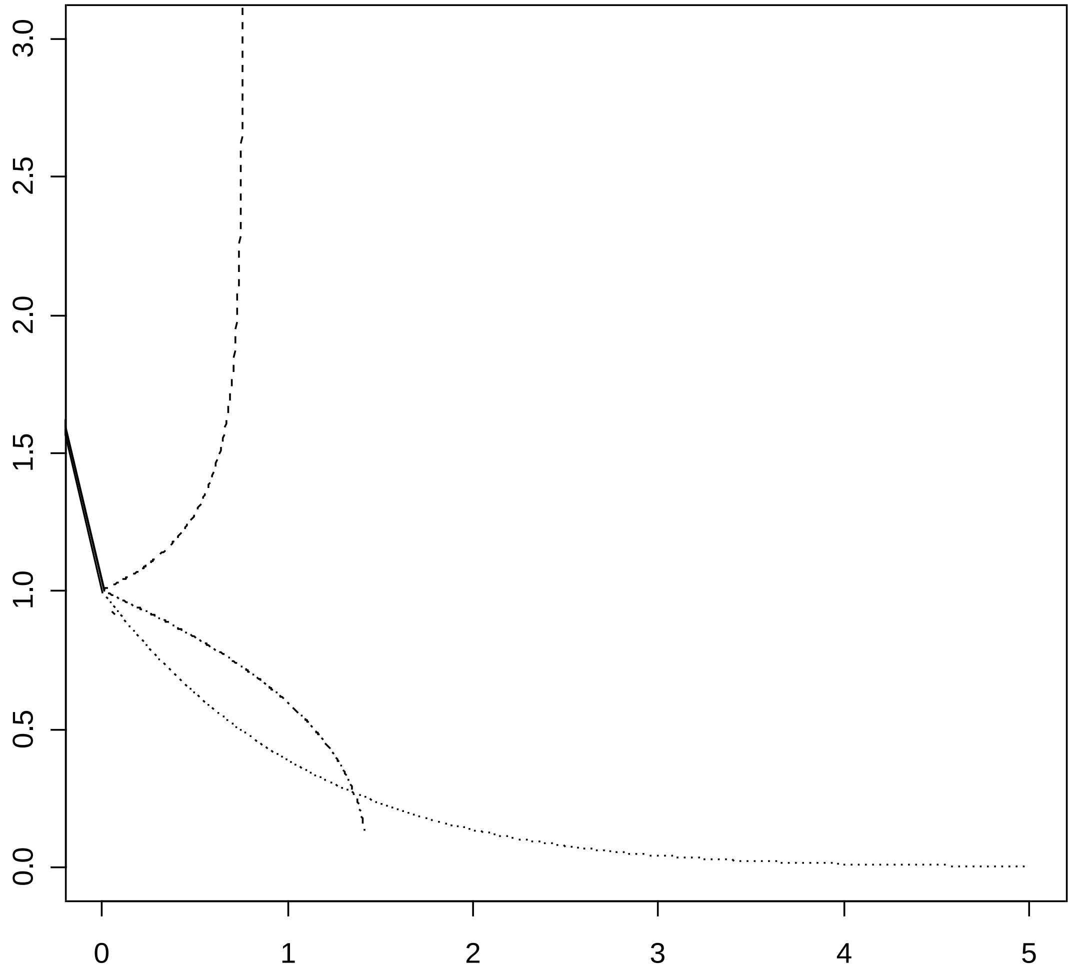
If X has a max-stable distribution, the conditional distribution function of X-u, under the condition that X>u, can be given as

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi},$$

if $\xi \neq 0$, $y > 0$ and $(1 + \xi y / \tilde{\sigma}) > 0$, where $\tilde{\sigma} = \sigma + \xi(u - \mu)$.
($H(y)$ is the so called generalized Pareto distribution, GPD).
If $\xi = 0$, $H(y) = 1 - e^{-y}$, $y > 0$ (the unit exponential distribution).

ξ is the same as the shape parameter of the corresponding GEV.

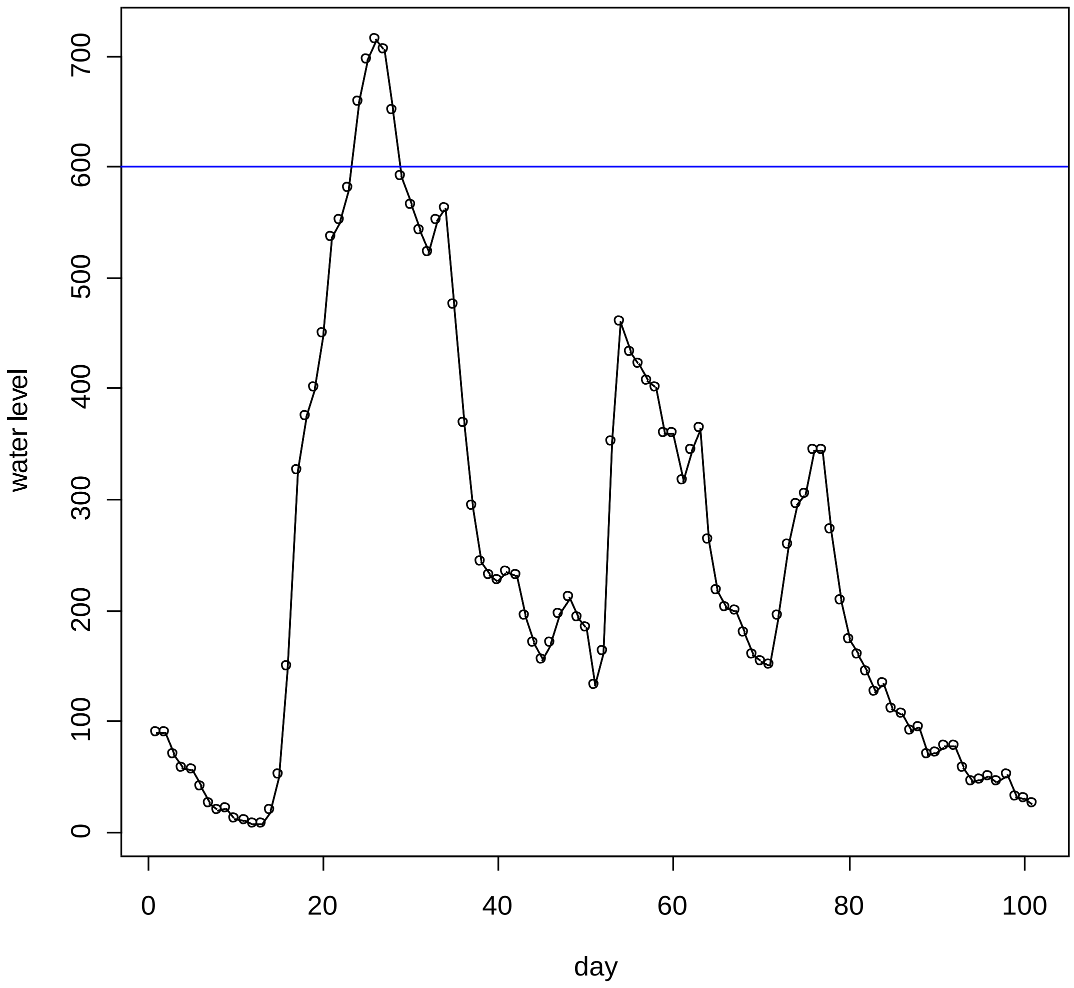
Densities of GPD with $\sigma=1$; solid: $\xi=0.5$, dotted: $\xi=-0.1$, dots-and-lines: $\xi=-0.7$, broken: $\xi=-1.3$



Threshold selection

- Mean excess plot:
  - For any *u* (threshold), plot the mean of *X-u* (for those observations for which *X*>*u*) against *u*.
  - If the Pareto model is true, this plot should be nearly linear.
  - The interpretation is made difficult by the great variability near the upper endpoint of the observations.
- Alternative:
  - Fit the distribution for different values of *u* and such a threshold should be chosen, where the parameter estimates are stable
- It is a trade-off between bias (likely to occur if the threshold is too low) and variance (occurs if the threshold is too high, and just a few observations are used for estimation).

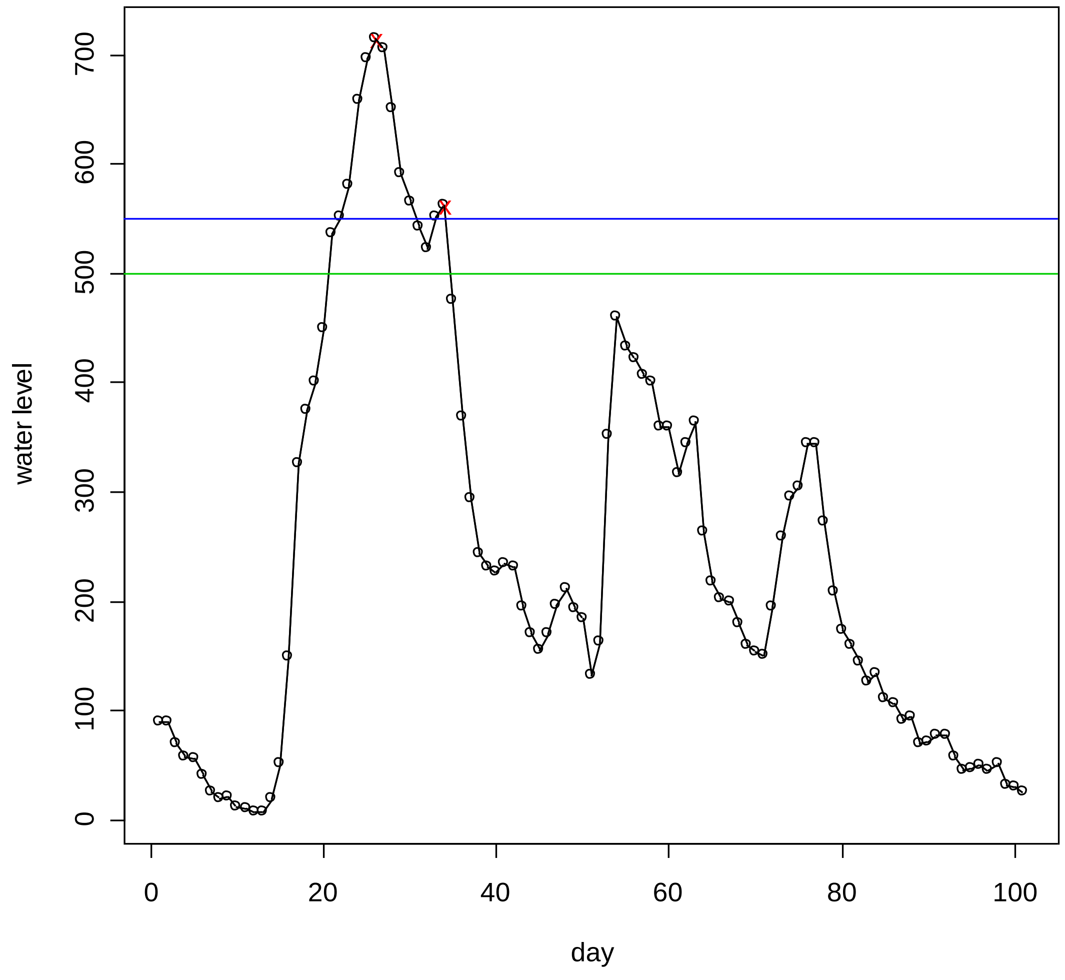
Dependence: a typical 100-day segment of the data



Declustering

One may take just the peak out of a cluster above the threshold.

The set of points does not increase monotonically as *u* decreases.



Practical solutions

- There are physical considerations resulting in a given time period, after which a new high value corresponds to a new flood, thus independence from the previous values is realistic.
- For water level data there is also a natural choice for the threshold u, as there is a level, when the water fills a wider area around the narrow riverbed.

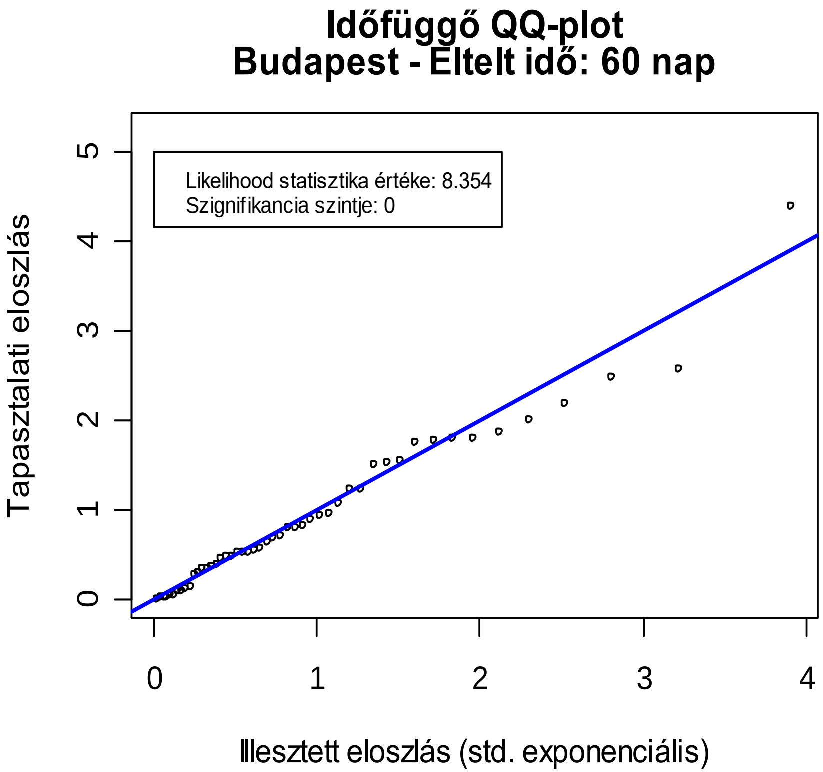
Modelling time-dependence

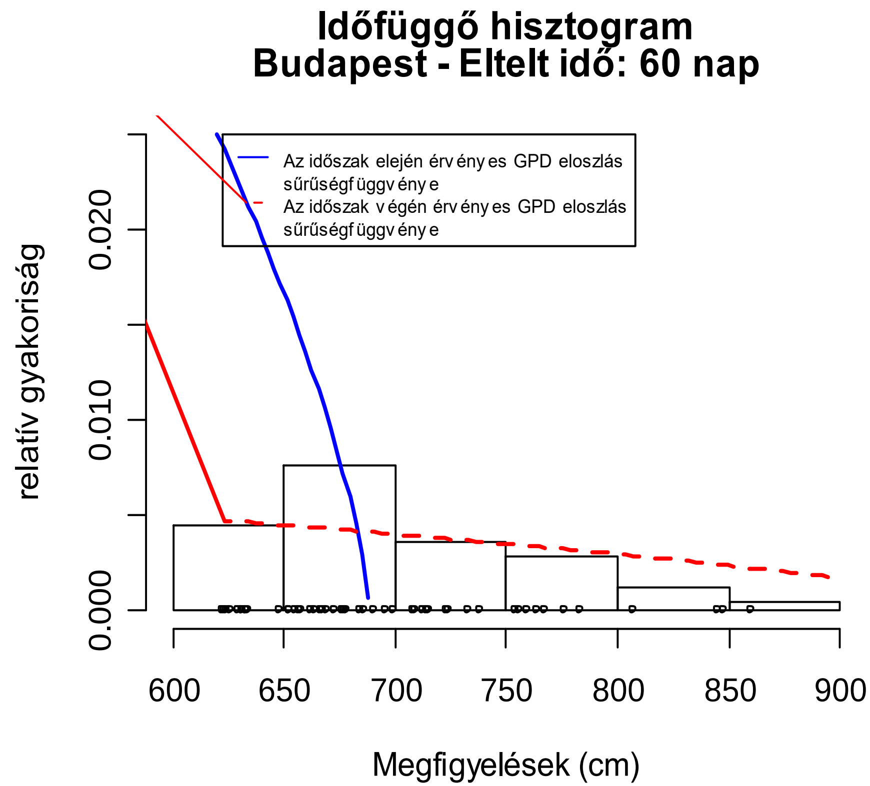
- Assumptions:
  - Shape does not change
  - Scale depends linearly on the time:

$$\sigma_t = \alpha \cdot t + \beta$$

- Positive coefficient (larger scale) here means larger observations as time advances
- Maximum likelihood estimator and the (modified) QQ plot for checking the fit are available

An example: Budapest





Parameters: time-dependent scale:1.79, scale:23.91, shape:-0.58

The distribution at the end of the time period is much more dangerous

The flood simulator

- Flood frequency estimator
  - By MCMC (result for the last period: „climate change”)
  - traditionally
- Random floods for the stations at upstream river sections
- Routing of upstream hydrographs by conceptual hydrodynamic models

Estimating the flood-frequency intensity

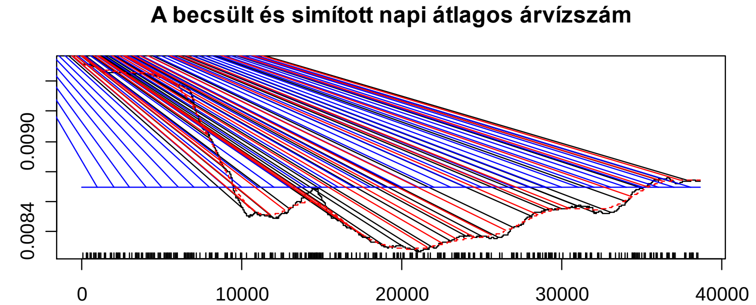
- It is assumed that the intensity is constant over a time period, but jumps are allowed (the data suggested such changes, corresponding to wet or dry periods in the climate)
- Reversible jump Markov Chain Monte Carlo method is used
- Apriori distribution
  - for the jumps: uniform
  - for the intensity: Gamma

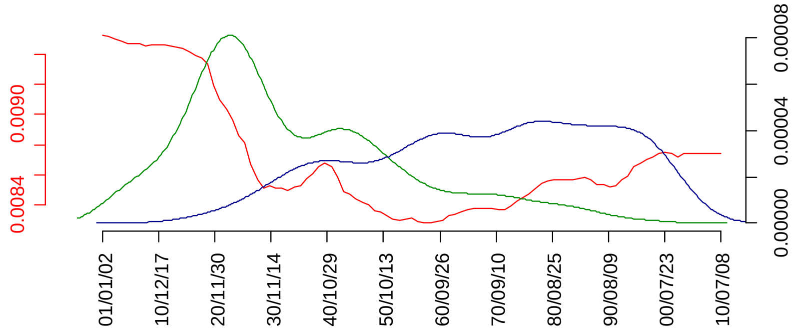
The MCMC algorithm

- Data: all flood peaks at the upstream stations (considered as different if the water level is low for at least 20 days)
- Three type of flood events
  - Flood of the Danube only
  - Flood of the Tisza and/or its tributaries only
  - Flood of both watersheds
- Steps of the algorithm
  - Change in intensity
  - Change the location of changepoints
  - Introducing a new changepoint
  - Removing an old changepoint

Illustration of the results (all floods)

- The simulated Markov chain can be in different states (different number of change points)
- The summary of the two change point-models show moderate fluctuation, but an increasing tendency for the last 50 years

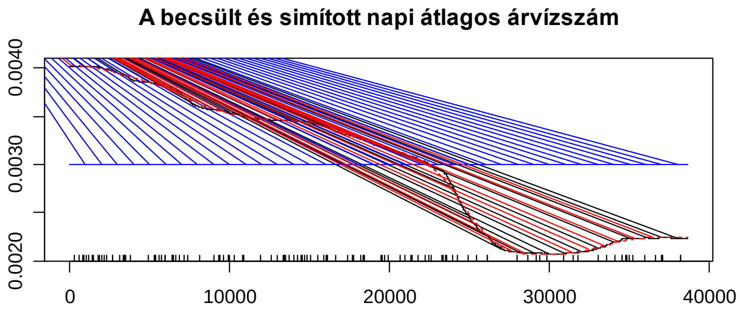




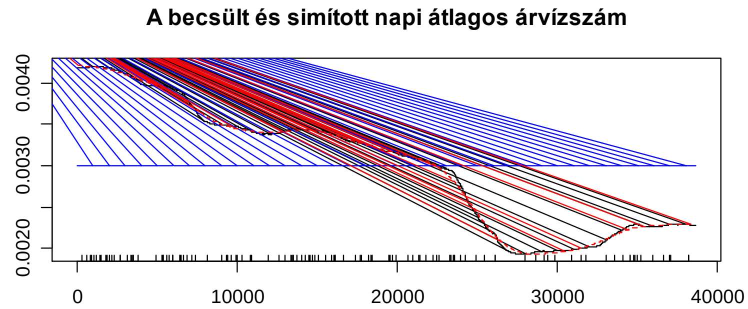
red: average of the daily intensities

green/blue: density of the first/second changepoint

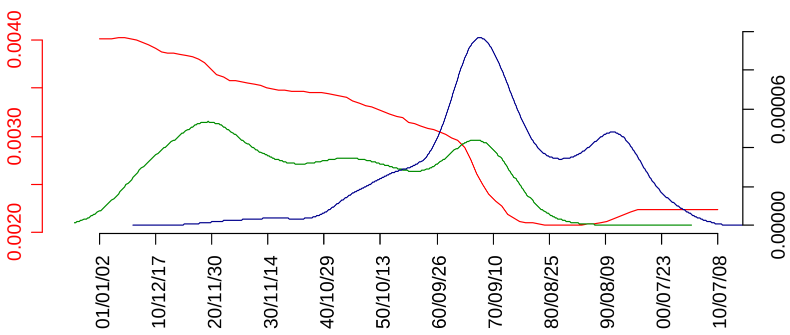
Results for floods of the Danube only

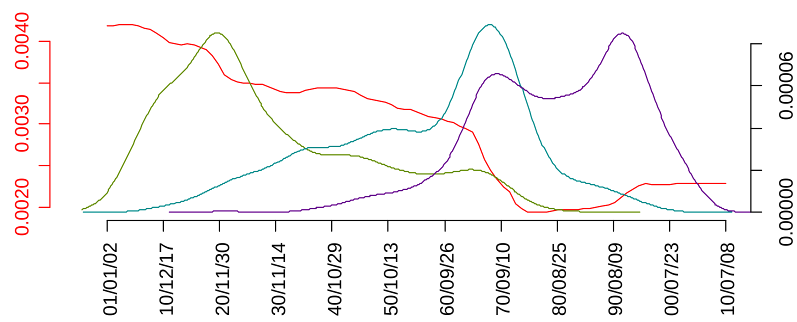


2-changepoints



3-changepoints

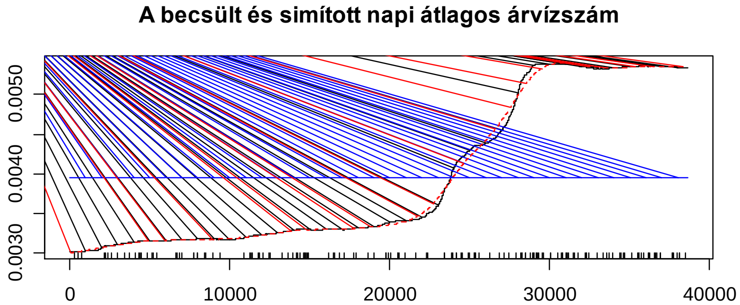




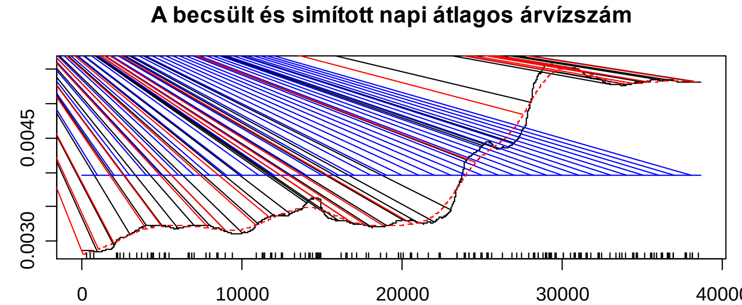
These models were the most frequent.

Weighted average estimator for the last 10 years: 0.0023

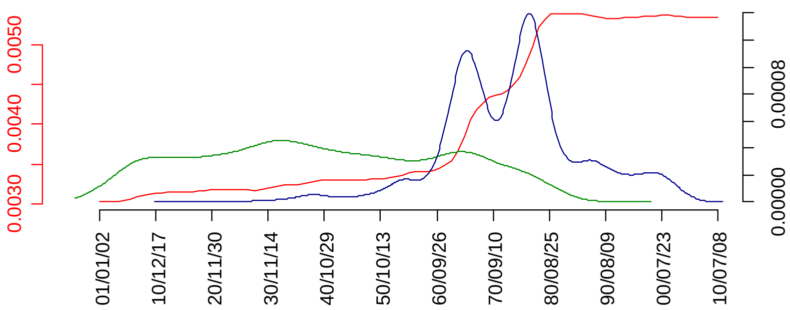
Floods of the Tisza and its tributary only

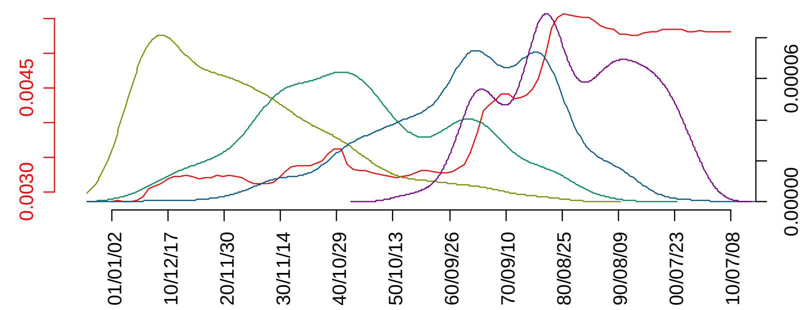


2-changepoints



4-changepoints

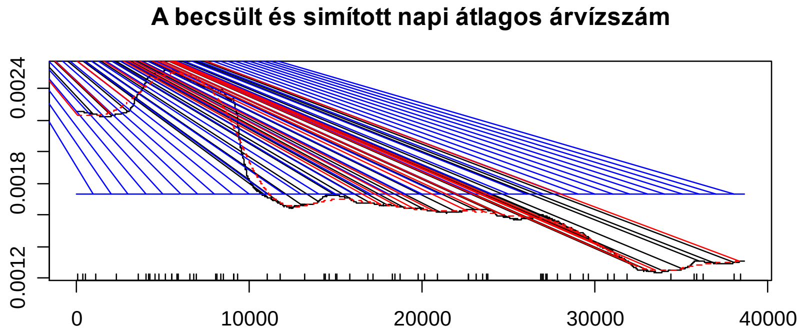


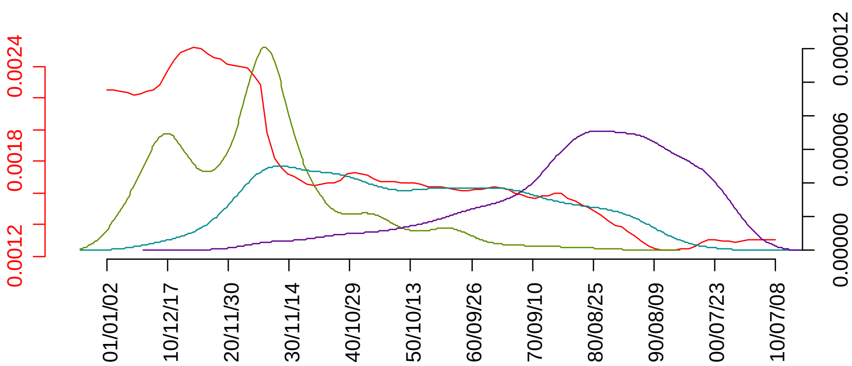


These models were the most frequent.

Weighted average estimator for the last 10 years: 0.0053

Floods affecting both rivers





Estimator: 0.0013

The fit of the nonhomogenous Poisson process was much better in all cases

The chosen upstream-stations

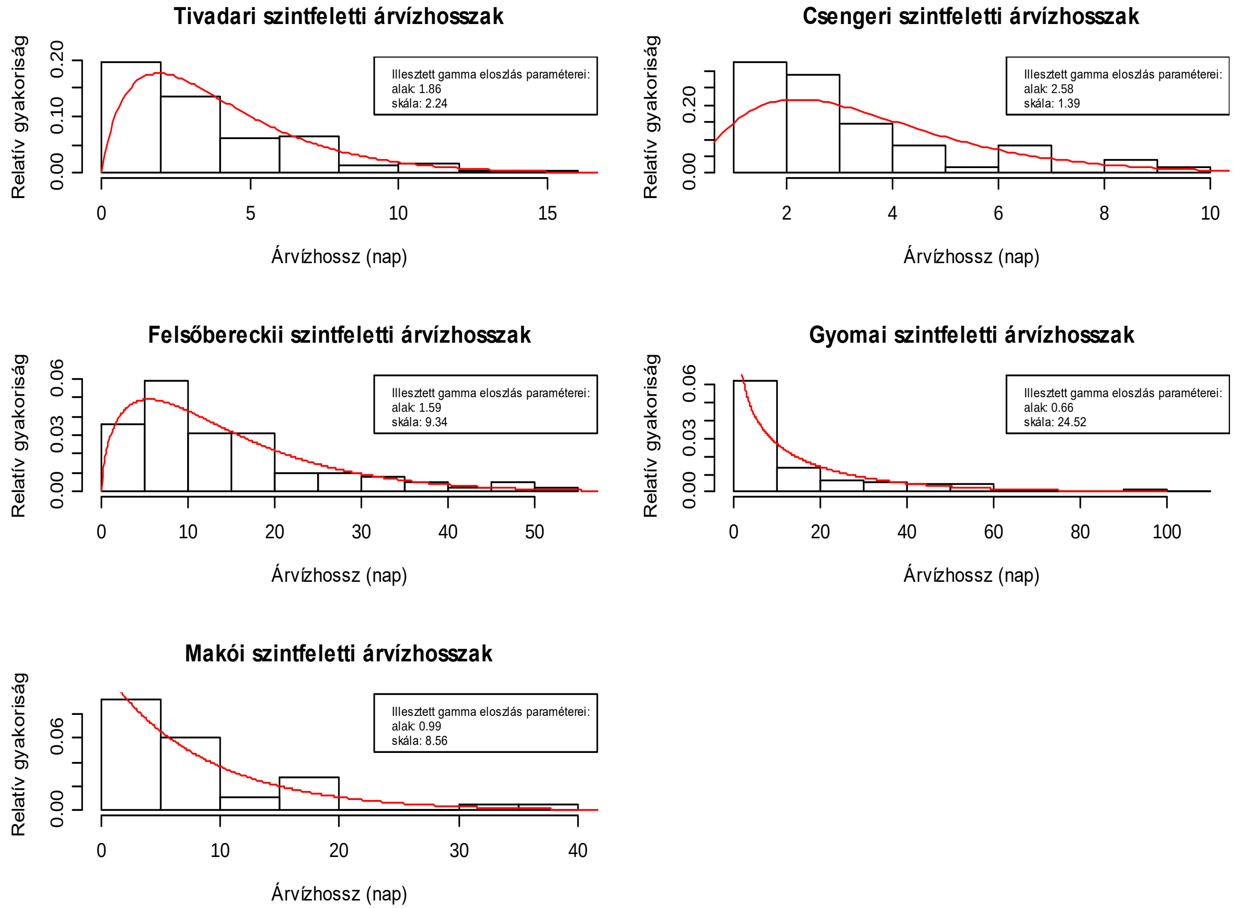
- Tisza and its tributaries:
  - Felsőberecki (Bodrog)
  - Tivadar (Tisza)
  - Csenger (Szamos)
  - Gyoma (Kőrös)
  - Makó (Maros)
- Danube and its tributaries :
  - Sárvár (Rába)
  - Komárom (Duna)
  - Budapest (Duna)

Results of the GPD-modelling

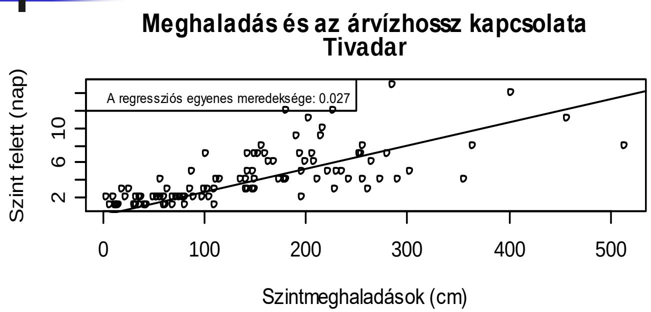
	Scale (1)	Scale (2)	Shape	Frequency	Flood/year
Sárvár	0.17	108.87	-0.35	91	0.95
Komárom	0.58	92.06	-0.39	107	1.02
Budapest	1.79	23.92	-0,59	50	0.52
Tivadar	1.22	143.68	-0.47	107	1.01
Csenger	0.24	127,23	-0.28	49	0.46
Felsőberecki	1.16	37.14	-0.59	78	1.03
Gyoma	0.53	134.57	-0.39	122	1.15
Makó	0.21	52.29	-0.05	37	0.39

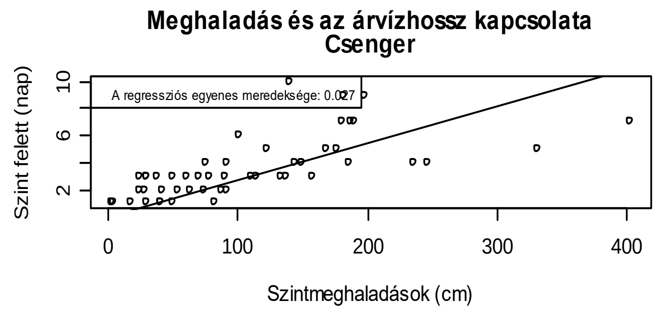
Duration of the floods

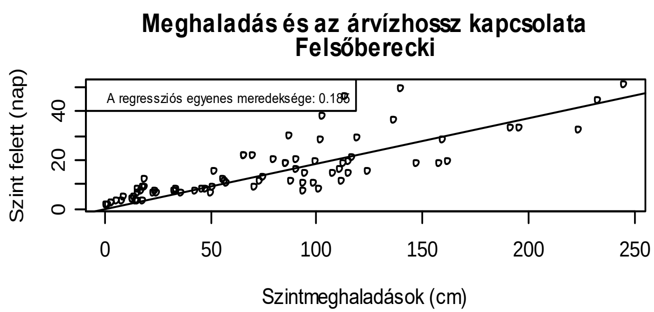
- Differences among the stations, basically two groups
- Gamma distribution was used for modelling the length

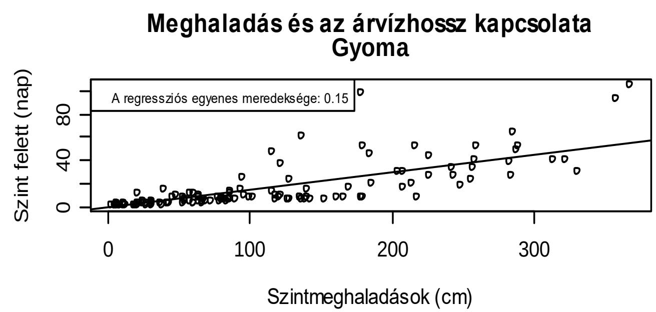


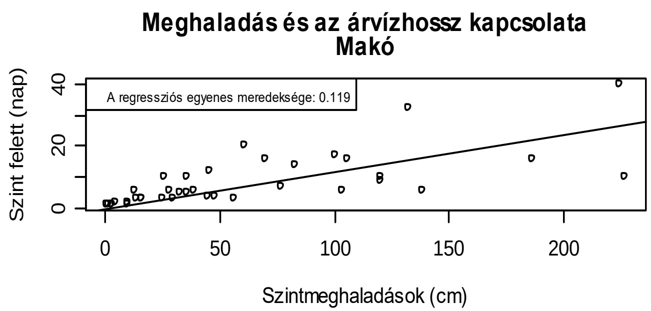
Dependence between the height and the duration of the floods











Linear regression was used, without an intercept and with a heteroscedastic, shifted Gamma distributed error term

The flood shape

- 4 different groups of stations was formed
- Having the height of the flood, 30 such hydrographs were chosen from the given group, where the height was nearest to the current simulated one
- Having the duration d, such a hydrograph was chosen from the above 30, for which the duration is at least $d/2$ and at most $2d$.

Dependence among the stations

- The most frequent scenarios out of the possible 31
- There were a couple of combinations, which have not occurred among the 164 floods, so a modification was needed

Berecki	Tivadar	Csenger	Gyoma	Makó	Frequency
0	0	0	1	0	32
1	0	0	0	0	21
0	1	0	0	0	17
1	0	0	1	0	15
1	1	0	0	0	14
0	0	0	0	1	10
1	1	1	1	1	8
0	1	0	1	0	8
1	1	0	1	0	7
1	1	1	0	0	5

Modified frequencies of joint flood occurrencies

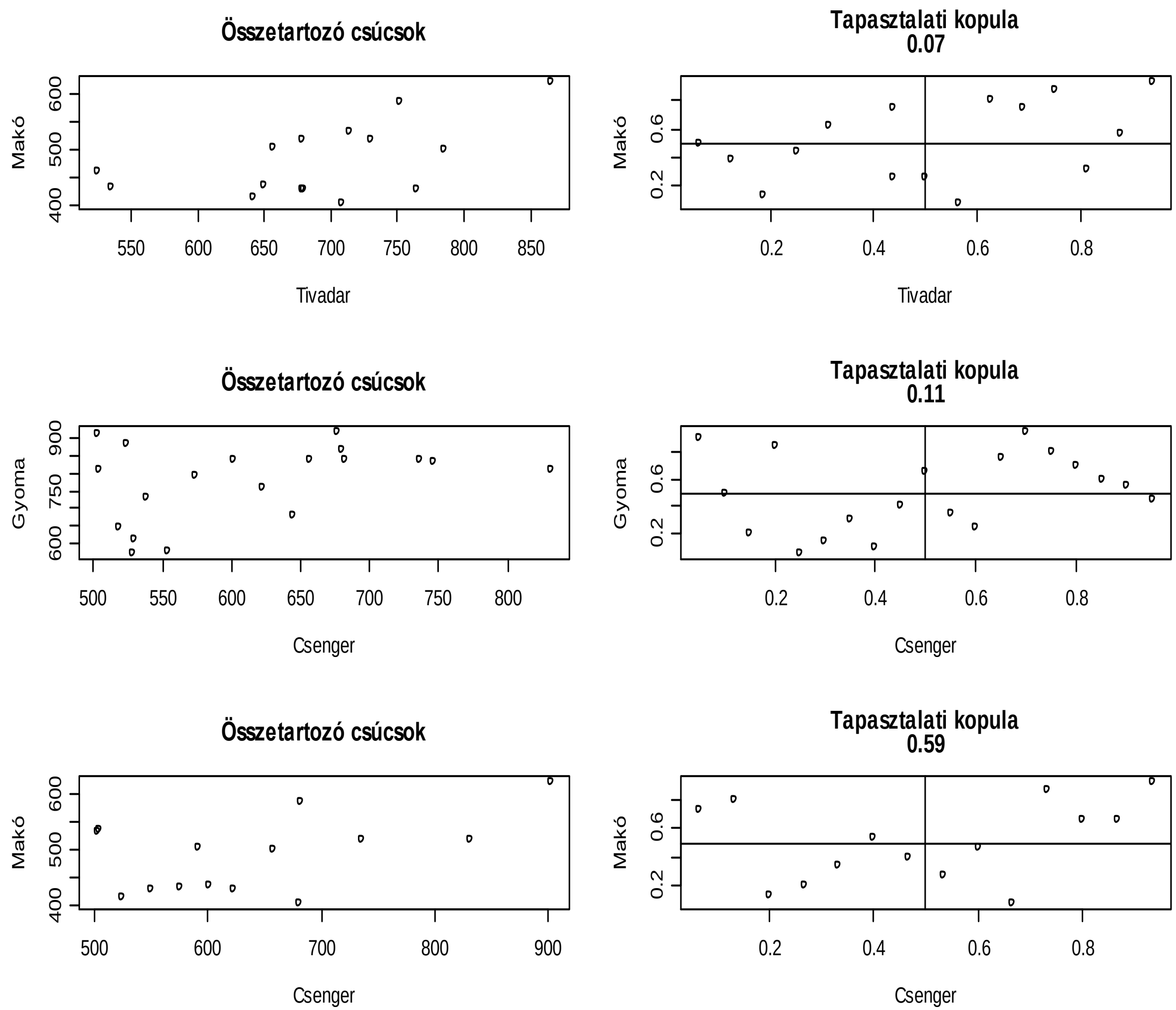
The near-floods were considered as floods with a probability (depending on the difference between the threshold and the observed peak).

The chance of multiple flooding was increased.

Berecki	Tivadar	Csenger	Gyoma	Makó	Freq.	Modified freq.
0	0	0	1	0	32	25.49
1	0	0	0	0	21	20.26
0	1	0	0	0	17	14.262
1	0	0	1	0	15	21.874
1	1	0	0	0	14	13.07
0	0	0	0	1	10	4.054
1	1	1	1	1	8	15.1553

Possible conditional dependence between the flood peaks

- If we know that there were floods at two stations, is there any relation between the height of these floods?
- The conditional independence can be accepted.



Simulations for the stations, where there was no flood

- Dependence is strong for neighbouring stations (two groups were formed on the Tisza)
  - For the neighbouring stations a height from the conditional distribution of the (non-flood) peaks was chosen, and then a real part of the hydrograph, which had this height as peak and the whole distribution was near enough to the observed one
  - For other stations, only the peak was set as before

Summary of the algorithm

1. Number of floods for the given year (from the Poisson distribution)
2. Simulation of the affected upstream stations
3. Simulation of the flood peaks from the time-dependent GPD (conditionally independent)
4. Duration of the flood (by regression model)
5. The shape of the flood (random choice from the observed sequences, by transformation)

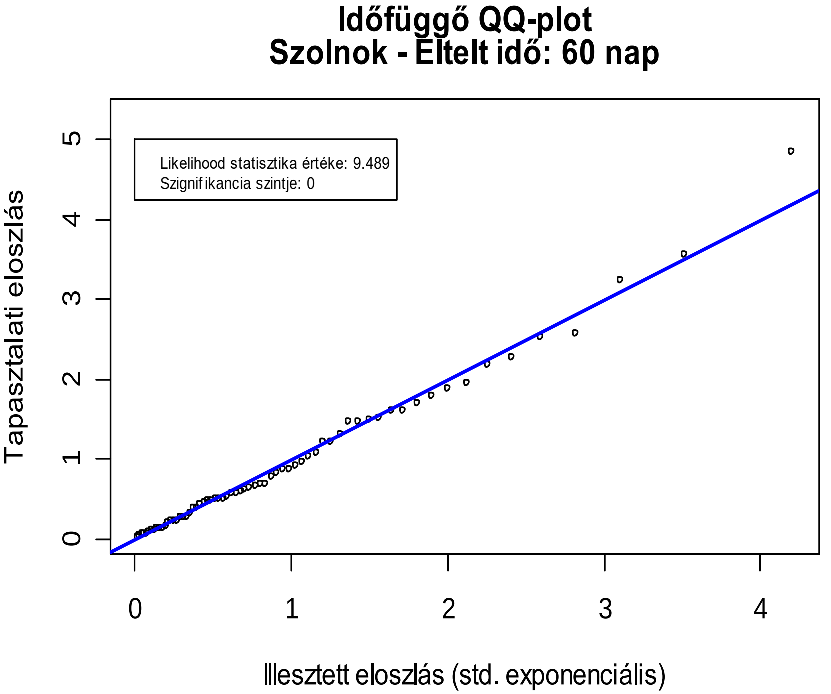
Algorithm/2

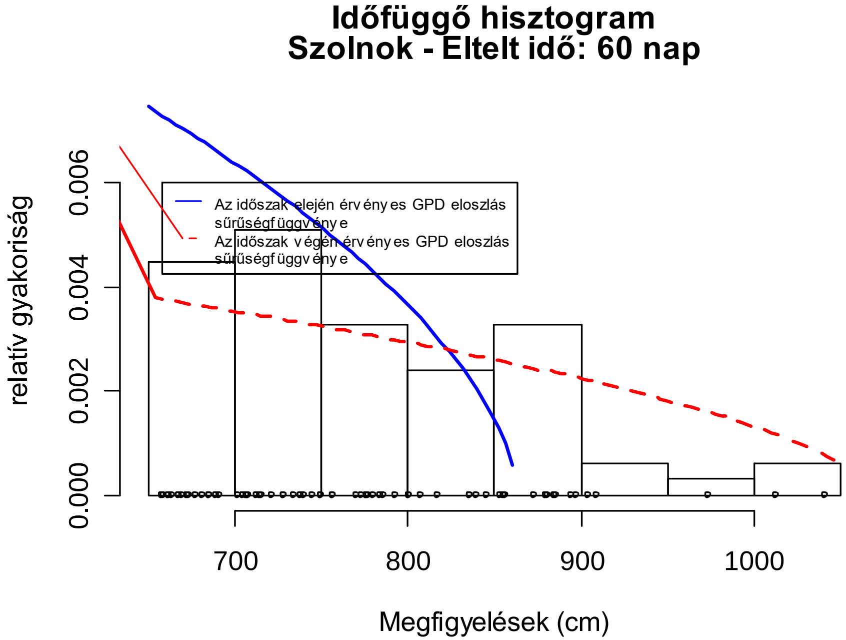
- The choice of medium and low flow data (from the conditional distribution, as a real, non-transformed sequence)
- Special case: the data for Budapest is got by Komárom (approx. 100 km upstream from Budapest) (by regression for two groups: were there floods on the catchment between)

Hydrological routing

- By the discrete linear cascade model, for the most important representative floodplain sections:
- For the Tisza: Vásárosnamény, Záhony, Tokaj, Polgár, Szolnok, Szeged
- For the Danube: Baja

The most dangerous area





Parameters: time-dependent scale:1.22, scale:133, shape:-0.63

The simulated floods were even more dangerous

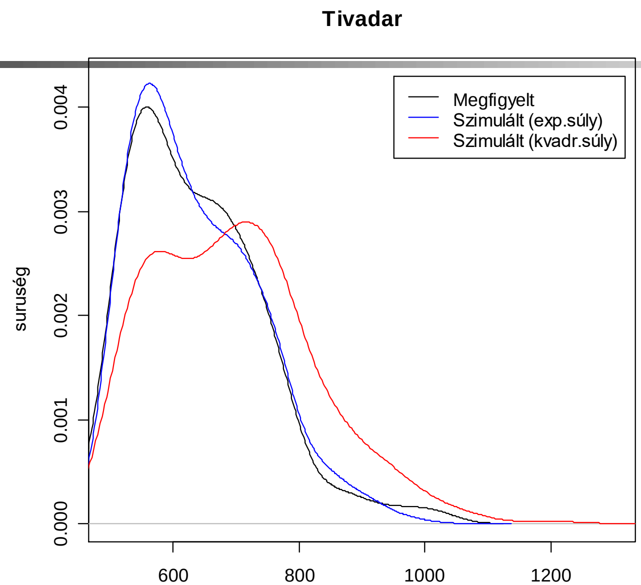
Data base for floods

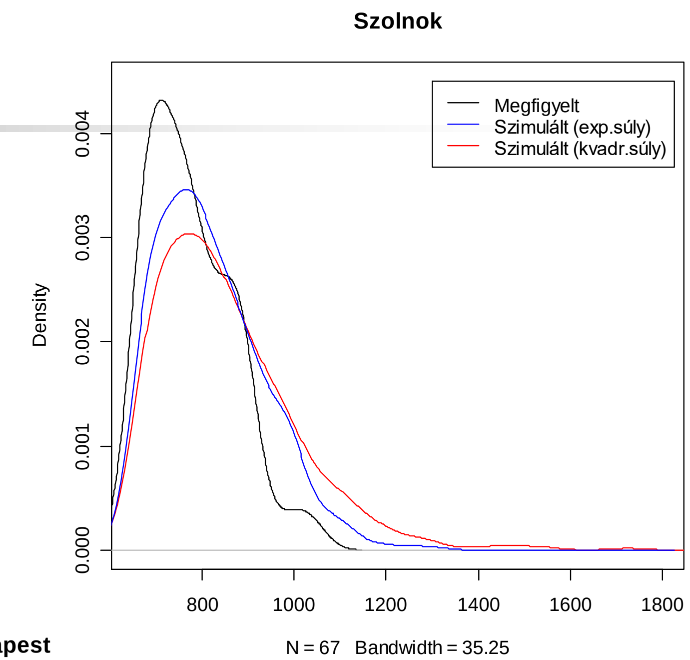
- As the simulation of floods for the upstream stations and the hydrological routing are computer intensive, the simulation uses a data base of floods generated as described above.
- 20000 floods were generated
  - 10000 from the described model
  - another 10000 from a bootstrap version, where the error of the GPD fit was also taken into account
- 1000 floods were chosen so that the fit to the observed data should be acceptable, but the sample should also include exceptionally dangerous ones (900 from the original sample, 100 from the bootstrap).

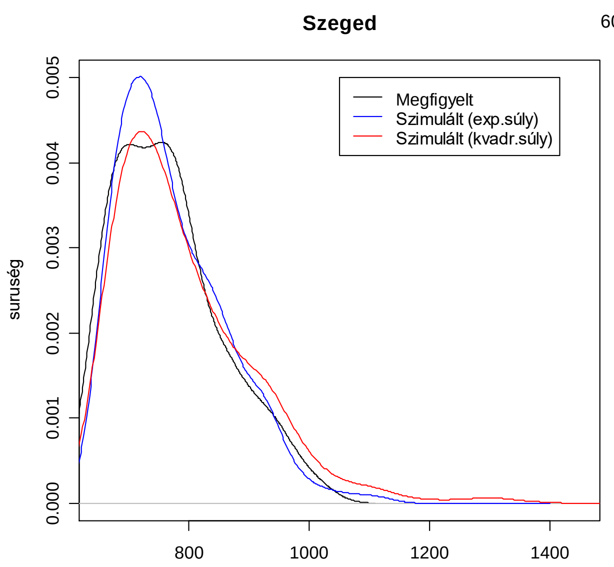
Weights for the simulator

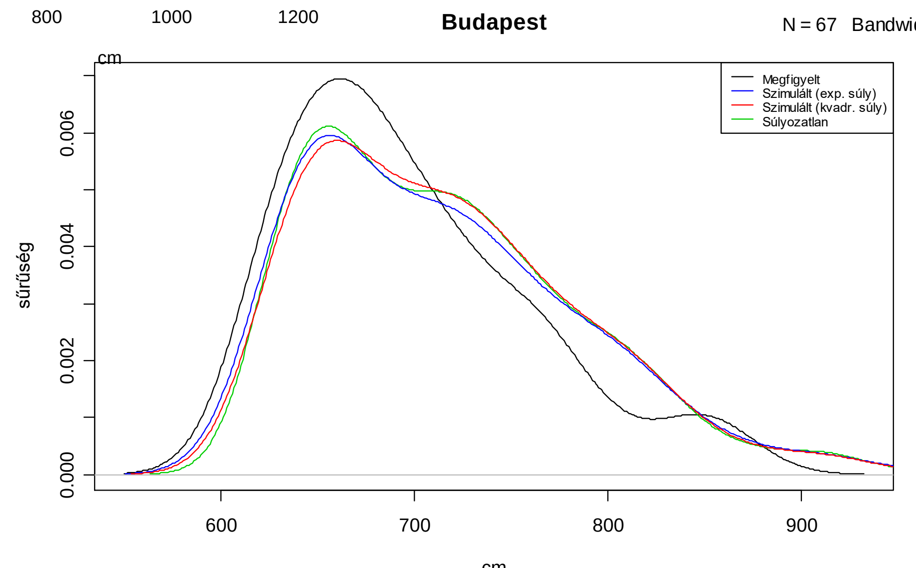
- We had dangerous floods in the data base, but the simulator should comply with the reality. We set the chosen risk by weights. The chosen characteristic was the duration (time length) of the flood.
- Exponential weights: by the likelihood-ratio of the Gamma distribution, fitted to the observed and simulated flood length at the given station.
- Quadratic weights: the extreme long floods were downweighted by a quadratic function (minimising the discrepancy between the observed and the simulated regression model for the exceedance/duration pair).

Examples


Tivadar
Megfigyelt
Szimulált (exp.súly)
Szimulált (kvadr.súly)
suruség
0.000
0.001
0.002
0.003
0.004
600
800
1000
1200


Szolnok
Megfigyelt
Szimulált (exp.súly)
Szimulált (kvadr.súly)
Density
0.000
0.001
0.002
0.003
0.004
800
1000
1200
1400
1600
1800
N = 67 Bandwidth = 35.25


Szeged
Megfigyelt
Szimulált (exp.súly)
Szimulált (kvadr.súly)
suruség
0.000
0.001
0.002
0.003
0.004
0.005
800
1000
1200
1400


Budapest
cm
Megfigyelt
Szimulált (exp. súly)
Szimulált (kvadr. súly)
Súlyozatlan
sűrűség
0.000
0.002
0.004
0.006
600
700
800
900
cm

Choosing from the weights

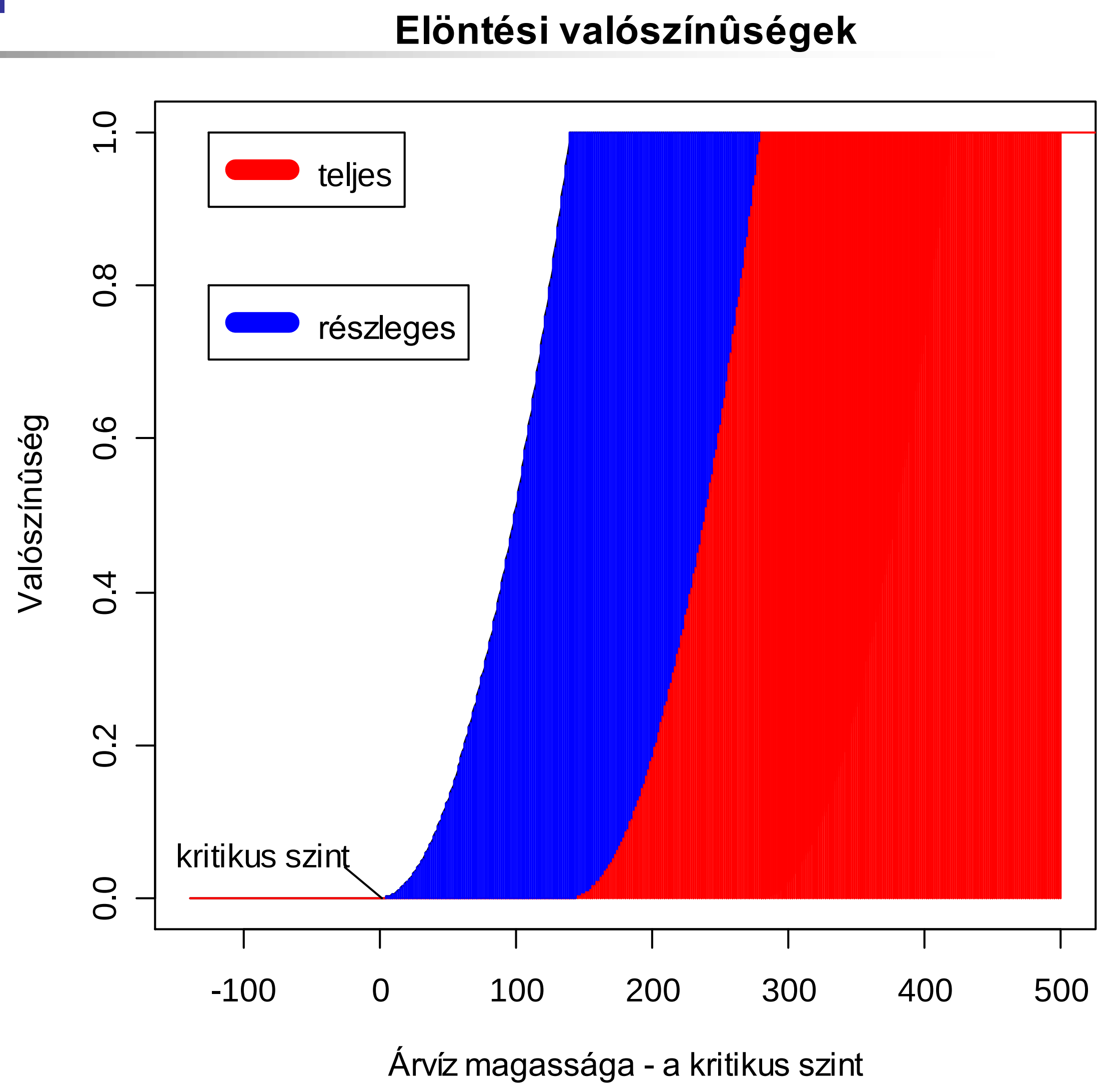
- The quadratic weights put more weight on the longer floods (“wet years”),
- The exponential weights provide floods more similar to the observed ones (“dry years”).
- Both cases are more risky than the observed data.

The floodplain sections

- The are more than 100 such sections, bordered by natural or artificial dykes.
- The dykes are not everywhere in good shape, the risks are determined by the weakest part in the floodplain sections.
- Thus a just safe (critical) level was determined for all floodplain sections, and it was supposed that a 140 cm higher flood automatically causes a dyke breachment.

Risk model

- Probability of breaching the dikes was given by hydrologists, a distinction between partial and full inundation of the floodplain section was done.



Localisation

- Even for larger dike failures it is usually possible to localise the flooding
- There are official contingency plans to confine inundation for the two largest floodplain sections
  - Budapest-Baja
  - Fegyvernek-Mesterszállás

  where secondary protection lines are marked by the plans
- For the others: a decrease of the probability for complete flooding in accordance with the area of the floodplain section was included into the model.
- The cities of Szolnok, Szeged, Győr are expected to get exceptional care, so their flooding is less likely (it is supposed to happen only if the peak level is higher by more than 140 cm than the level of the surrounding flood embankments/levees)

Effect of dike breach

- The water level is decreased by a couple of centimetres in the neighbouring downstream stations. This effect depends on the size and topology of the inundated floodplain section.
- We used a simple model: the effect is $10cm/100km^2$, which is halved every 50 km.

Experiences

- In the last decades there were just a few breaching of dykes
  - Bereg, 2001 being the most recent (where we had claim data from)
- These corresponded to extreme high floods (with estimated return period over 100 years)
- There were problems with the dykes around Szolnok during the recent floods. This validates our observation of this area being the most dangerous.

Claim ratio

- It is possible to choose from different distributions (beta, or normal/t for the logit transform).
- The expected value and variance of the claim can be parametrised. The default value is the estimator we got by the available data (it depends on the wall type, floor etc.)
- The dependency between the building and content-losses is also included into the model. Here we used the copulas for finding suitable structure to a given correlation.

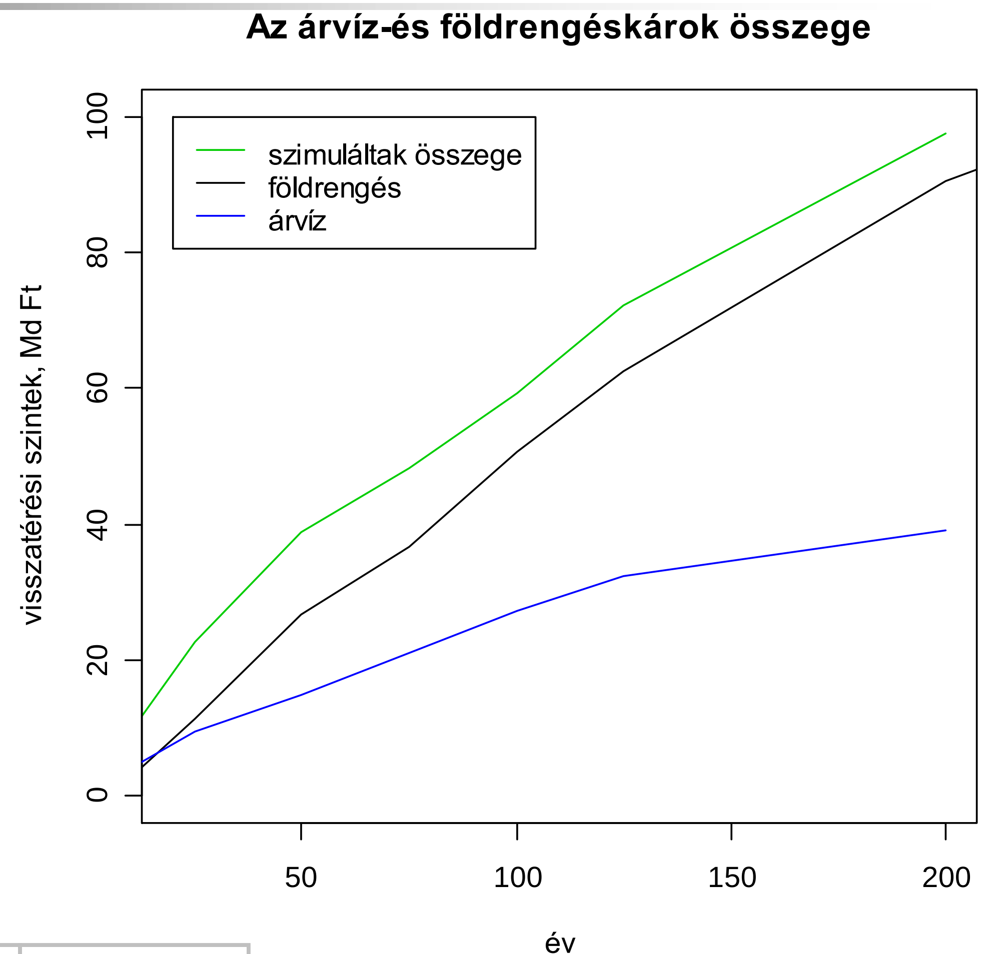
Some parameters of the losses

	partial	flooding	total	flooding
	building	content	building	content
stone	0.0261	0.1173	0.1236	0.1173
adobe	0.0318	0.1173	0.2635	0.1959
other	0.0339	0.1173	0.2934	0.1959

- These losses are supposed to be halved for 2nd floor flats and further halved for higher levels.

Convolution of the flood and the quake losses

- The simplest, and also the most accurate method was the simulation. The flood results are calculated for the wet years, including the effect of the climate changes.



10	25	50	100	200	year
9.38	22.55	38.40	57.66	93.84	return level (thousands of M HUF)