

Eötvös Loránd Tudományegyetem
Természettudományi Kar

Budapesti Corvinus Egyetem
Közgazdaságtudományi Kar



Max-stabilis folyamatok és alkalmazásuk az extrém biztosítási kockázatok modellezésében

diplomamunka

Szerző: Tóth Gábor

ELTE-BCE, Biztosítási és pénzügyi matematika MSc
aktuárius szakirány

Témavezető: Dr. Zempléni András

egyetemi docens
ELTE-TTK, Valószínűségelméleti és Statisztika Tanszék

2012

Tartalomjegyzék

0.1. Áttekintés	4
0.2. Köszönetnyilvánítás	5
1. A max-stabilis folyamatok elmélete	6
1.1. Az egydimenziós extrémérték-elmélet	6
1.2. Többdimenziós extrémérték-elmélet	8
1.3. Max-stabilis folyamatok nevezetes modelljei	10
1.3.1. Poisson-pontfolyamatok	10
1.3.2. A Smith-modell	11
1.3.3. A Schlather-modell	14
1.4. Illesztés tapasztalati adatokra	17
2. Alkalmazás valós földrajzi adatokra	20
3. Alkalmazás biztosítási káradatokra	26
3.1. Motivációk	26
3.2. Egy konkrét idősor elemzése	27

Ábrák jegyzéke

1.1. A Smith-modell szimulációi.	14
1.2. Whittle–Matérn és hatványexponenciális korrelációfüggvények különféle α -kra.	16
1.3. A Schalter-modell szimulációi.	17
2.1. Bemutató ábra a svájci csapadékadatokhoz.	21
2.2. Néhány mérési pontra való GEV-illesztés QQ-plotjai.	22
2.3. Extremális együtthatók szintvonalai, és a velük kapcsolatos QQ-plotok három, a csapadékadatokra illesztett max-stabilis modellnél.	23
2.4. A lineáris GEV-együttható becslés jóságát jellemző „QQ-plotok” hatványexponenciális modellnél.	24
2.5. „Tapasztalati” és modellezett extremális együtthatók a távolság függvényében, az eredeti és egy szimulált adatsorra.	25
2.6. Az éves csapadékmaximumok szimulációi hatványexponenciális Schalter-modellel. A zöld árnyalatok 20 mm körüli, a pirosak 60 mm körüli csapadékot jeleznek.	25
3.1. A kárkifizetések inflációval korrigált időszora, szezonaritása és adjusztált időszora.	28
3.2. QQ-plot a káradatokra való GEV-illesztésről, illetve néhány illesztett sűrűségfüggvény.	29
3.3. Az illesztett Schalter-modellek Bessel-féle korrelációfüggvényei.	31
3.4. A hétfők és csütörtökök együttes eloszlása a 168 kiigazított munkahétben, és annak szimulációiban.	33

Táblázatok jegyzéke

2.1. Becsült GEV-paraméterek a teljes adathalmazra és három mérési pontra.	22
3.1. Becsült GEV-paraméterek a teljes ötnapos munkahetek napjaira. . . .	30
3.2. Extremális együtthatók a munkanapokra.	32

0.1. Áttekintés

Ezen szakdolgozattal az a célom, hogy a térbeli extrémérték-elméletnek az elmúlt néhány évtizedben kifejlődött, és napjainkig leginkább a meteorológia terén használt módszereit bemutassam, továbbá hogy ezeknek a biztosítási kockázatok modellezésében való alkalmazhatóságát vizsgáljam.

A dolgozat első részében egy rövid elméleti áttekintés található. Ebben először is bevezetem a max-stabilis folyamat fogalmát, amely központi helyet foglal el a tárgyalt elméletben. Ezután bemutatom ennek a fogalomnak néhány fontos speciális esetét, kezdve a legklasszikusabbal, az egydimenziós extrémérték-elmélettel és az EVT-eloszlásokkal, folytatva néhány többdimenziós módszer említésével, végül pedig Smith és Schlather véletlen pontfolyamatokon alapuló modelljeinek tárgyalásával, különös tekintettel a gyakorlatban leginkább alkalmazott, normális alakfüggvényt illetve stacionárius Gauss-folyamatot felhasználó esetekre. A fejezet végén szólok néhány szót az illesztési, becslési módszerekről is.

A második fejezetben bemutatom az említett modellek és módszerek gyakorlati használatát az R programnak, illetve a hozzá letölthető `SpatialExtremes` csomagnak a segítségével. Ez utóbbi tartalmaz egy példa-adatbázist svájci csapadékadatokkal, amelyet szintén felhasználok, egyrészt a kiváló illusztrációs lehetőség miatt, másrészt abból a megfontolásból, hogy megfelelő földrajzi részletességgel összegyűjtött biztosítási káradatok is mutathatnak ehhez hasonló struktúrát.

Végül a harmadik fejezetben kitérek a fentiek tényleges biztosítási kárstatisztikákra való alkalmazásának lehetőségeire, elemzésnek vetve alá egy napi kárkifizetéseket tartalmazó idősort.

A programoknak a vizsgálatok végzése idején elérhető legfrissebb verzióit használtam: R 2.15.0 és `SpatialExtremes` 1.8-1. A dolgozatban közölt elemzések, eredmények és ábrák ezek használata mellett reprodukálhatók, a mellékelt CD-n lévő fájlok segítségével. A CD tartalma, ezen dolgozat elektronikus változatával együtt, elérhető a <http://www.cs.elte.hu/~tannin/szakdolgozat/> címen is.

0.2. Köszönetnyilvánítás

Szeretnék köszönetet mondani Zempléni Andrásnak a nagy türelemmel és odafigyeléssel végzett témavezetői munkájáért, hasznos tanácsaiért, és különösen az adatok beszerzésében való nélkülözhetetlen közbenjárásáért.

1. fejezet

A max-stabilis folyamatok elmélete

Az, hogy egy véletlen változó vagy folyamat max-stabilis, velősen összefoglalva annyit tesz, hogy maximumképzéssel már nem hozhatunk létre belőle lényegesen mást; a maximumképzés végtelen távoli végállomását adja, ebben a geometria ideális pontjaihoz hasonlít. Az extrémérték-elméletben ilyen ideákkal vetjük össze, és próbáljuk magyarázni a valóság extrém jelenségeit. Ebben a fejezetben a definíció kimondása után az egyszerűtől az összetett felé haladva ismertetjük a max-stabilis folyamatok nevezetes példáit.

1.0.1. Definíció (Max-stabilis folyamat). *Egy tetszőleges T indexhalmazon értelmezett $Y = \{Y_t, t \in T\}$ folyamatot max-stabilisnak nevezünk, ha minden $N \geq 1, t \in T$ esetén léteznek olyan $A_{N,t} > 0$ és $B_{N,t}$ konstansok, hogy ha Y^1, \dots, Y^N a folyamat N darab független példánya, és minden $t \in T$ -re*

$$Y_t^* = \frac{\max_{1 \leq n \leq N} Y_t^n - B_{N,t}}{A_{N,t}},$$

akkor az $Y^ = \{Y_t^*, t \in T\}$ folyamat ugyanolyan eloszlású, mint az Y folyamat.*

1.1. Az egydimenziós extrémérték-elmélet

Az extrémérték-elmélet (Extreme Value Theory, EVT) klasszikus problémája a következő: Adottak az $X_i, i = 1, 2, \dots$ független, azonos eloszlású valószínűségi változók, F eloszlásfüggvénnyel, amelyet praktikus okból mindig jobbról folytonos változatban fogunk használni. Ezek a legtöbb alkalmazásban valamilyen természeti jelenséggel kapcsolatosak (csapadék, vízállás), vagy valamilyen pénzügyi kárt, veszteséget jelenítenek

meg adott időpontokban, vagy helyeken. Az $M_n = \max(X_1, \dots, X_n)$ adott periódus alatti maximumérték viselkedéséről szeretnénk valamit mondani az $n \rightarrow \infty$ határátmenet vizsgálatával. A maximum eloszlása:

$$P(M_n \leq y) = P(X_1 \leq y, \dots, X_n \leq y) = F^n(y).$$

Ez természetesen elfajult eloszláshoz vezet, ha n -nel a végtelenhez tartunk, ezért érdemes valamilyen n -től függő konstansokkal centrálni, illetve normálni az (M_n) sorozatot, mint ahogyan a valószínűségi változók összegével tesszük a centrális határeloszlástételben. A lehetséges kimenetelről az alábbi nevezetes állítás szól ([4] p. 226):

1.1.1. Tétel (Fisher–Tippett, Gnedenko). *Ha léteznek olyan $A_n > 0$ és B_n konstansokból álló sorozatok, hogy*

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - B_n}{A_n} \leq x\right) = \lim_{n \rightarrow \infty} F^n(A_n x + B_n) = H(x),$$

teljesül valamilyen H nem elfajuló eloszlásfüggvényre, akkor H a GEV eloszláscsaládba tartozik.

1.1.2. Definíció (GEV eloszláscsalád). *Az extrémérték-eloszlásokat a*

$$H_\xi(x) = \begin{cases} \exp\left(- (1 + \xi x)^{-\frac{1}{\xi}}\right), & \xi \neq 0 \\ \exp(-e^{-x}), & \xi = 0 \end{cases}$$

eloszlásfüggvények határozzák meg, ahol $1 + \xi x > 0$. A $\xi \in \mathbb{R}$ értéket az eloszlás alakparaméterének nevezzük. Az általánosított extrémérték-eloszlások (Generalized Extreme Value, GEV) alakja:

$$H_{\xi, \mu, \sigma}(x) = H_\xi\left(\frac{x - \mu}{\sigma}\right),$$

ahol $\mu \in \mathbb{R}$, $\sigma > 0$ az eloszlás eltolás- és skálaparétérei.

Vegyük észre, hogy a $\xi = 0$ -hoz tartozó eloszlásfüggvény megkapható a másiktól a $\xi \rightarrow 0$ határátmenettel. Az eltolás és skálaparaméterek lényegesek a statisztikai elemzéseknél, mert Fisher–Tippett-tétel A_n és B_n paramétereit nem tudjuk becsülni.

Az alakparaméter alapján három csoportra szokás osztani a GEV eloszlásokat. Amikor $\xi > 0$, akkor **Fréchet-eloszlást** kapunk. Az $\alpha = \frac{1}{\xi}$, $m = \mu - \frac{\sigma}{\xi}$ és $s = \frac{\sigma}{\xi}$ helyettesítésekkel a Fréchet-eloszlás az alábbi alakban írható fel:

$$H_{\xi, \mu, \sigma}(x) = F_{\alpha, m, s}(x) = e^{-\left(\frac{x-m}{s}\right)^{-\alpha}}, \quad \text{ha } x > m.$$

Az $m = 0$, $\alpha = s = 1$ esetre a továbbiakban **sztenderd** vagy **egységnyi Fréchet-eloszlásként** fogunk hivatkozni. Fréchet-eloszlást vastag, hatványrendben lecsengő szélű eloszlások maximumainak határeloszlásaként kaphatunk meg, mint a Pareto-, vagy a Student-eloszlás. A tőzsdei hozamok és a felelősségbiztosítási kárkifizetések tipikusan vastag farkúak.

A $\xi = 0$ esete a **Gumbel-eloszlás**. Ez az előbb említetteknél gyorsabban – például exponenciálisan – csökkenő szélű eloszlások maximumainak határeloszlása lehet, mint az exponenciális, a normális vagy a lognormális eloszlás. Ha egy μ eltolás- és σ skálaparaméterű Gumbel-eloszlásból származó valószínűségi változót az exponenciális függvénybe helyettesítünk, akkor $F_{1/\sigma, 0, \exp(\mu)}$ eloszlásfüggvényű Fréchet-eloszlásút kapunk; sztenderd Gumbel esetén tehát a fenti értelemben vett sztenderd Fréchet-t.

Végül $\xi < 0$ értékekre kapjuk a **Weibull-eloszlást**, amely véges jobboldali végpontú, ezért felülről korlátos valószínűségi változók vizsgálatánál használható, mint például a hitelkockázat, illetve a biztosítási kockázatvállalások nagy része.

A számegyenesen értelmezett eloszlások között a GEV családot *karakterizálja a max-stabilis tulajdonság* ([8] p. 2); a szükséges konstansok:

$$A_N = N^\xi, \quad B_N = \begin{cases} (1 - N^\xi) \left(\mu - \frac{\sigma}{\xi} \right) & \xi \neq 0 \\ \sigma \ln N & \xi = 0 \end{cases}$$

Sztenderd Fréchet esetén például $A_N = N$ és $B_N = 0$ a stabilizáló konstansok. Későbbi fontosságára való tekintettel ezt külön le is vezetjük. Ha az X_i -k sztenderd Fréchet-k, akkor

$$P\left(\frac{M_N}{N} \leq y\right) = P(M_N \leq N \cdot y) = \left[e^{-\frac{1}{N \cdot y}}\right]^N = e^{-\frac{1}{y}} = P(X_1 \leq y) \quad (1.1)$$

Megjegyzés: Az itt tárgyalt elmélet természetesen alkalmazható minimumok vizsgálatára is, hiszen $\min(X_1, \dots, X_n) = -\max(-X_1, \dots, -X_n)$.

1.2. Többdimenziós extrémérték-elmélet

Legyenek most az \mathbf{X}_i , $i = 1, 2, \dots$ független, közös F eloszlásfüggvényű, d dimenziós valószínűségi változók, és jelöljük az i -ediknek a j -edik koordinátáját X_i^j -vel. Ekkor az n elemű minta maximumát koordinátánkénti maximumként értelmezzük, azaz:

$$\mathbf{M}_n = \left(\max_{1 \leq i \leq n} X_i^1, \dots, \max_{1 \leq i \leq n} X_i^d \right).$$

Vegyük észre, hogy \mathbf{M}_n nem feltétlenül eleme a mintának. A következő definícióban a vektorok közötti műveletek és összehasonlítás koordinátáinként értendő ([4], p. 311).

1.2.1. Definíció (MEV eloszláscsalád). *Ha létezik olyan F d -dimenziós eloszlásfüggvény, és léteznek olyan $\mathbf{A}_n > \underline{0}_d$ és \mathbf{B}_n determinisztikus vektorokból álló sorozatok, hogy*

$$\lim_{n \rightarrow \infty} P \left(\frac{\mathbf{M}_n - \mathbf{B}_n}{\mathbf{A}_n} \leq \mathbf{x} \right) = \lim_{n \rightarrow \infty} F^n(\mathbf{A}_n \mathbf{x} + \mathbf{B}_n) = H(\mathbf{x}),$$

teljesül valamilyen H d -dimenziós eloszlásfüggvényre, akkor azt mondjuk, hogy H a többdimenziós extrémérték-eloszlások (Multivariate Extreme Value, MEV) családjába tartozik.

Ha H egyetlen marginálisában sem elfajuló, akkor 1.1.1 alapján tudjuk, hogy a peremeloszlásai GEV-ek. Már csak a közöttük lévő összefüggőségi struktúra kérdéses. Ennek vizsgálatakor feltehető, hogy a peremeket egységnyi Fréchet-re transzformáltuk. Egy klasszikus megközelítési mód, amit érdemes megemlíteni, a H -hoz tartozó, a Sklar-tétel szerint egyértelműen létező kopula vizsgálata, amelyhez egy szép eredmény kapcsolódik. Eszerint a kopulára is egyfajta stabilitás teljesül: ([4] p. 311-312)

1.2.2. Tétel. *Ha H a MEV eloszláscsaládba tartozik és egyetlen marginálisában sem elfajuló, akkor a hozzá egyértelműen létező C kopulára igaz, hogy $C(\mathbf{u}^t) = C^t(\mathbf{u})$, $\forall t > 0$.*

Egy újabb megközelítés, ami számunkra fontosabb, az extrémális együtthető (lásd pl. [5] p. 19-től, [3] p. 7). Legyenek az \mathbf{X} véletlen vektor X^1, X^2, \dots, X^d koordinátái egységnyi Fréchet-eloszlásúak. Írjuk fel az együttes eloszlást a következő alakban:

$$P(X^1 \leq x_1, \dots, X^d \leq x_d) = \exp(-V(x_1, \dots, x_d))$$

A V függvény jellemzi az összefüggőségi struktúrát. Tökéletes függetlenség akkor áll fenn, amikor $V(x_1, \dots, x_d) = \sum_{j=1}^d \frac{1}{x_j}$, tökéletes összefüggőség pedig abban az esetben, ha $V(x_1, \dots, x_d) = \max_{1 \leq j \leq d} \frac{1}{x_j}$. Az (1.1) miatt:

$$\begin{aligned} \exp(-nV(nx_1, \dots, nx_d)) &= P(X^1 \leq nx_1, \dots, X^d \leq nx_d)^n \\ &= P\left(\frac{\mathbf{M}_n^1}{n} \leq x_1, \dots, \frac{\mathbf{M}_n^d}{n} \leq x_d\right) = P(X^1 \leq x_1, \dots, X^d \leq x_d) \\ &= \exp(-V(x_1, \dots, x_d)). \end{aligned}$$

A V függvény tehát -1 rendben homogén, és ezért

$$P(X^1 \leq x \dots, X^d \leq x) = \exp\left(-\frac{\theta_d}{x}\right),$$

ahol $\theta_d = V(\underline{1}_d)$, amit *extremális együtthathónak* nevezünk. A θ_d értéke az $[1; d]$ intervallumba esik, 1 jelenti a tökéletes összefüggőséget, d pedig a tökéletes függetlenséget.

Az 1.3.2 és 1.3.3 szakaszokban bemutatandó folytonos modellek is lényegében az extrémális együtthathó segítségével írják le két a két pont közötti összefüggőséget, úgy, hogy az értéke a pontok közötti távolságvektortól függjön.

1.3. Max-stabilis folyamatok nevezetes modelljei

1.3.1. Poisson-pontfolyamatok

Mielőtt a nevezetes modelleket ismertetnénk, [10] alapján bevezetjük a Poisson-pontfolyamat fogalmát, amely tetszőleges σ -véges mértéktérre való általánosítása a jól ismert, számegegyenesen definiált Poisson-folyamat szakadási helyeiből kapott pontfolyamatnak.

1.3.1. Definíció (Pontfolyamat és Poisson-pontfolyamat). *Legyen $(\Omega, \mathcal{B}, \mathcal{P})$ valószínűségi mértéktér, (S, \mathcal{A}, μ) pedig σ -véges mértéktér. Legyen Z az összes (s_1, s_2, \dots) , $s_i \in S$, $i = 1, 2, \dots$ (esetleg véges) pontrendszer halmaza az S téren. Jelölje $z(\mathbf{A})$ a z pontrendszer \mathbf{A} halmazba eső pontjainak számát. Legyen továbbá*

$$\mathcal{F} = \sigma\left(\left\{z \mid z \in Z, z(\mathbf{A}_1) = k_1, \dots, z(\mathbf{A}_j) = k_j\right\}, j = 1, 2, \dots;\right. \\ \left.\forall 1 \leq l \leq j : k_l \in \mathbb{N}, \mathbf{A}_l \in \mathcal{A}, \mu(\mathbf{A}_l) < \infty\right)$$

Ekkor egy $\xi : (\Omega, \mathcal{B}) \rightarrow (Z, \mathcal{F})$ mérhető leképezést az (S, \mathcal{A}) téren értelmezett pontfolyamatnak nevezünk.

Azt mondjuk, hogy ez a ξ pontfolyamat Poisson-pontfolyamat az (S, \mathcal{A}) téren μ számlálómértékkel, ha tetszőleges $j \in \mathbb{Z}^+$ esetén, tetszőleges $\mathbf{A}_l \in \mathcal{A}$, $\mu(\mathbf{A}_l) < \infty$, $1 \leq l \leq j$ diszjunkt halmazrendszert véve a $z(\mathbf{A}_l)$, $1 \leq l \leq j$ valószínűségi változók függetlenek; továbbá minden $\mathbf{A} \in \mathcal{A}$, $\mu(\mathbf{A}) < \infty$ esetén $z(\mathbf{A})$ Poisson-eloszlású $\mu(\mathbf{A})$ paraméterrel.

1.3.2. Állítás. *Tetszőleges (S, \mathcal{A}, μ) σ -véges mértéktéren létezik Poisson-pontfolyamat μ számlálómértékkel.*

Bizonyítás. Lásd a [10]-ban felvázolt konstrukciót. □

1.3.2. A Smith-modell

A most következő módszert max-stabilis folyamatok generálására [7] és [5] alapján tárgyaljuk. Legyen S tetszőleges mérhető tér σ -véges ν mértékkel, $\{(\xi_i, s_i) \mid i = 1, 2, \dots\}$ pedig a $(0, +\infty) \times S$ téren értelmezett Poisson-pontfolyamat realizációja, $(\xi^{-2} d\xi) \times \nu$ számlálómérték mellett. Legyen továbbá $f : S \times S \rightarrow \mathbb{R}_0^+$ függvény olyan, hogy

$$\int_S f(s, t) \nu(ds) = 1, \quad \forall t \in S. \quad (1.2)$$

Ekkor a Smith-féle konstrukció a következő:

$$Y_t = \sup_i \{\xi_i f(s_i, t)\}, \quad \forall t \in S. \quad (1.3)$$

Hogy megértsük az emögött rejlő szemléletet, képzeljük el, hogy különböző záporokból származó csapadékmennyiséget mérünk S pontjaiban – legyen például $S = \mathbb{R}^2$. Ekkor s_i az i -edik zápor centruma, ξ_i az intenzitása, $t \mapsto f(s_i, t)$ pedig az s_i közép-pontú zápor alakját meghatározó függvény, és így $\xi_i f(s_i, t)$ az i -edik záporból származó csapadékmennyiség a t helyen. Ezen csapadékadatok közül vesszük pontonként a legnagyobbat.

1.3.3. Állítás. *A fent definiált Y max-stabilis folyamat az S téren, továbbá Y_t sztenderd Fréchet-eloszlású minden t -re.*

Bizonyítás. Rögzítsük az $y_t > 0$ értékeket minden $t \in S$ -re. Vezessük be a következő halmazt:

$$B = \{(\xi, s) \mid \exists t : \xi f(s, t) > y_t\}$$

Az $\{Y_t \leq y_t \text{ minden } t\text{-re}\}$ esemény azt jelenti, hogy a pontfolyamatnak nem esik pontja a B halmazba. A B halmazba eső pontok száma Poisson-eloszlású, paramétere a halmaz mértéke, ezért:

$$\begin{aligned} P(Y_t \leq y_t \text{ minden } t\text{-re}) &= \exp\left(-\int_S \int_{\inf_t \frac{y_t}{f(s,t)}}^{\infty} \xi^{-2} d\xi \nu(ds)\right) \\ &= \exp\left(-\int_S \sup_t \frac{f(s,t)}{y_t} \nu(ds)\right). \end{aligned} \quad (1.4)$$

Az (1.4) egyenlőségből levezethető állításunk mindkét fele. A max-stabilitás megmutatásához vegyünk Y^1, \dots, Y^N független, Y -nal megegyező eloszlású folyamatokat. Ekkor $\frac{\max_{1 \leq i \leq N} Y^i}{N}$ eloszlása is megegyezik Y eloszlásával, hiszen egyrészt abból,

hogy független Poisson-eloszlások összege Poisson-eloszlás és a paraméterek összedódnak, másrészt (1.3)-ból következik, hogy $\max_{1 \leq i \leq N} Y^i$ maga is Smith-féle konstrukció $N \cdot [(\xi^{-2} d\xi) \times \nu]$ számlálómértékkel, és ezért:

$$\begin{aligned} P\left(\frac{\max_{1 \leq i \leq N} Y_t^i}{N} \leq y_t \text{ minden } t\text{-re}\right) &= \exp\left(-\int_S \sup_t \frac{f(s,t)}{Ny_t} (N \cdot \nu)(ds)\right) \\ &= P(Y_t \leq y_t \text{ minden } t\text{-re}) \end{aligned}$$

Végül pedig a marginálisok sztenderd Fréchet-k, mivel tetszőleges $t_0 \in T$ -re:

$$\begin{aligned} P(Y_{t_0} \leq y_{t_0}) &= \lim_{t \neq t_0: y_t \rightarrow \infty} P(Y_t \leq y_t \text{ minden } t\text{-re}) \\ &= \exp\left(-\int_S \frac{f(s, t_0)}{y_{t_0}} \nu(ds)\right) \stackrel{(1.2)}{=} \exp\left(-\frac{1}{y_{t_0}}\right) \end{aligned}$$

□

Tekintsük most azt a speciális esetet, amikor $S = \mathbb{R}^d$, ν a Lebesgue-mérték, továbbá $f(s, t) = f_0(s - t)$, ahol f_0 az $N(\underline{0}, \Sigma)$ d -dimenziós normális eloszlás sűrűségfüggvénye. A záporsemények alakja tehát elliptikus, melyet lényegében a Σ kovarianciamátrix határoz meg, amelyről azt is feltesszük, hogy pozitív definit. Smith megmutatta, hogy ekkor a folyamat kétdimenziós határeloszlásai a következőképpen írhatók fel ([7]):

$$P(Y_{t_1} \leq y_1, Y_{t_2} \leq y_2) = \exp\left[-\frac{1}{y_1} \Phi\left(\frac{a}{2} + \frac{1}{a} \log \frac{y_2}{y_1}\right) - \frac{1}{y_2} \Phi\left(\frac{a}{2} + \frac{1}{a} \log \frac{y_1}{y_2}\right)\right], \quad (1.5)$$

ahol Φ a sztenderd normális eloszlásfüggvény, valamint

$$a^2 = (t_1 - t_2)^\top \Sigma^{-1} (t_1 - t_2). \quad (1.6)$$

Bizonyítás. (1.4)-ből következik, hogy:

$$\begin{aligned} -\log P(Y_{t_1} \leq y_1, Y_{t_2} \leq y_2) &= \int_{\mathbb{R}^d} \int_{\min\left(\frac{y_1}{f_0(s-t_1)}, \frac{y_2}{f_0(s-t_2)}\right)}^{\infty} \xi^{-2} d\xi ds \\ &= \int_{\mathbb{R}^d} \max\left(\frac{f_0(s-t_1)}{y_1}, \frac{f_0(s-t_2)}{y_2}\right) ds \\ &= \int_{\mathbb{R}^d} \frac{f_0(s-t_1)}{y_1} I\left\{\frac{f_0(s-t_1)}{y_1} > \frac{f_0(s-t_2)}{y_2}\right\} \\ &\quad + \frac{f_0(s-t_2)}{y_2} I\left\{\frac{f_0(s-t_2)}{y_2} > \frac{f_0(s-t_1)}{y_1}\right\} ds \end{aligned}$$

$$= E \left[\frac{1}{y_1} I \left\{ \frac{f_0(X)}{y_1} > \frac{f_0(X - t_2 + t_1)}{y_2} \right\} + \frac{1}{y_2} I \left\{ \frac{f_0(X)}{y_2} > \frac{f_0(X - t_1 + t_2)}{y_2} \right\} \right],$$

ahol $I\{A\}$ az A esemény indikátora, X pedig $N(0, \Sigma)$ eloszlású valószínűségi változó, melynek sűrűségfüggvénye:

$$f_0(x) = (2\pi)^{-\frac{d}{2}} |\Sigma|^{-1} \exp \left(-\frac{1}{2} x^\top \Sigma^{-1} x \right).$$

Ebbe X -et behelyettesítve levezethető:

$$\begin{aligned} & I \left\{ \frac{f_0(X)}{y_1} > \frac{f_0(X - t_2 + t_1)}{y_2} \right\} \\ &= I \left\{ X^\top \Sigma^{-1} (t_1 - t_2) > \log \frac{y_1}{y_2} - \frac{1}{2} (t_1 - t_2)^\top \Sigma^{-1} (t_1 - t_2) \right\}. \end{aligned}$$

Vegyük észre, hogy $X^\top \Sigma^{-1} (t_1 - t_2)$ egy 0 várható értékű és $(t_1 - t_2)^\top \Sigma^{-1} (t_1 - t_2) = a^2$ varianciájú normális eloszlású valószínűségi változó, ezért a fenti esemény valószínűsége $\Phi \left(\frac{a}{2} + \frac{1}{a} \log \frac{y_2}{y_1} \right)$, és így

$$E \left[\frac{1}{y_1} I \left\{ \frac{f_0(X)}{y_1} > \frac{f_0(X - t_2 + t_1)}{y_2} \right\} \right] = \frac{1}{y_1} \Phi \left(\frac{a}{2} + \frac{1}{a} \log \frac{y_2}{y_1} \right),$$

ami természetesen az 1, 2 indexeket felcserélve is igaz; ebből következik az állítás. \square

Az extrémális együttható értéke tehát $\theta(t_1 - t_2) = 2\Phi \left(\frac{a}{2} \right)$. Az a érték a két pont súlyozott euklideszi távolsága, amely **Mahalanobis-távolság** néven ismert. Ez is interpretálható a pontokbeli határeloszlások összefüggőségének mértékeként, ugyanis:

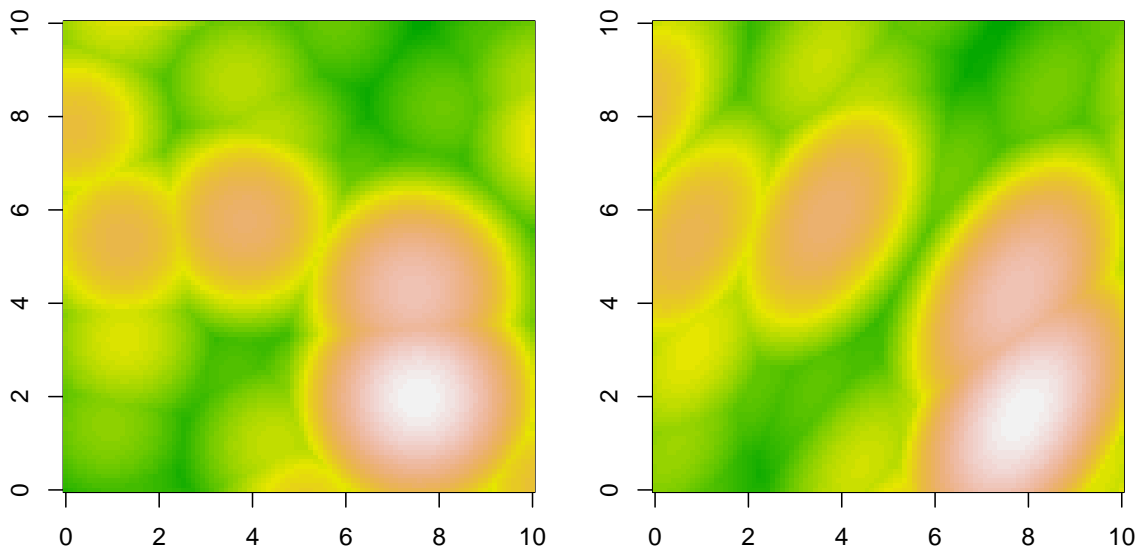
$$\begin{aligned} \lim_{a \rightarrow +\infty} P(Y_{t_1} \leq y_1, Y_{t_2} \leq y_2) &= P(Y_{t_1} \leq y_1) \cdot P(Y_{t_2} \leq y_2), \text{ továbbá} \\ \lim_{a \rightarrow +0} P(Y_{t_1} \leq y_1, Y_{t_2} \leq y_2) &= P(Y_{t_1} \leq \min(y_1, y_2)), \end{aligned}$$

az előbbi a tökéletes függetlenséget, az utóbbi a tökéletes összefüggőséget jelenti.

A Σ kovariancimátrixunk szimmetrikus és pozitív definit, ezért felírható $\Sigma = U\Lambda U^\top$ alakban, ahol Λ a pozitív sajátértékeket tartalmazó diagonális mátrix, U pedig ortogonális mátrix, soraiban a megfelelő sajátvektorokkal. Például:

$$\Sigma = \begin{pmatrix} 1 & \frac{\sqrt{3}}{3} \\ \frac{\sqrt{3}}{3} & \frac{5}{3} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & \frac{2}{3} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}.$$

Tehát a fenti kovariancimátrix hatását a sztenderd esettel összehasonlítva úgy lehet jellemezni, hogy a záporok alakja kétszeresére nyúlik $+60^\circ$ irányába, és $\frac{2}{3}$ -szorosára szűkül össze -30° irányába, mint ahogyan ez az 1.1 ábrán is megfigyelhető.



1.1. ábra. A Smith-modell két szimulációja, a bal oldalon egységnyi kovarianciamátrixszal, a jobb oldalon pedig $\sigma_{11} = 1$, $\sigma_{12} = \frac{\sqrt{3}}{3}$ és $\sigma_{22} = \frac{5}{3}$ paraméterekkel. A szemléletesebb megjelenítés érdekében Gumbellé transzformáltuk a marginálisokat.

1.3.3. A Schlather-modell

Smith módszerét Schlather általánosította 2002-ben ([6]). A fő eltérés a kettő között, hogy míg az előző szakaszban ismertetett modellben az egyes események térbeli eloszlása – a záporok alakja – előre rögzített, a most következő konstrukcióban ez is véletlenszerű, mint az intenzitásuk.

Legyenek a $\{\xi_i, i = 1, 2, \dots\}$ értékek, mint az előbb, egy Poisson-pontfolyamat realizációja a $(0, +\infty)$ téren $\xi^{-2} d\xi$ számlálómértékkel. Legyen továbbá X stacionárius folyamat az \mathbb{R}^d téren, melyre igaz, hogy

$$E[\max(0, X_t)] = 1, \quad \forall t \in \mathbb{R}^d \quad (1.7)$$

Ekkor a Schlather-féle konstrukció:

$$Y_t = \sup_i \{\xi_i \max(0, X_t^i)\}, \quad \forall t \in \mathbb{R}^d \quad (1.8)$$

Ahol az $X^i, i = 1, 2, \dots$ folyamatok az X független példányai.

1.3.4. Állítás. *Az így definiált Y folyamat stacionárius, max-stabilis folyamat, továbbá Y_t sztenderd Fréchet-eloszlású minden t -re.*

Bizonyítás. A stacionaritás közvetlen következménye X stacionaritásának. A max-stabilitáshoz és a határeloszláshoz értelemszerűen, pontról pontra másolhatjuk az **1.3.3** állítás bizonyítását, (1.2) és (1.3) helyett az (1.7) és (1.8) tulajdonságokra hivatkozva. \square

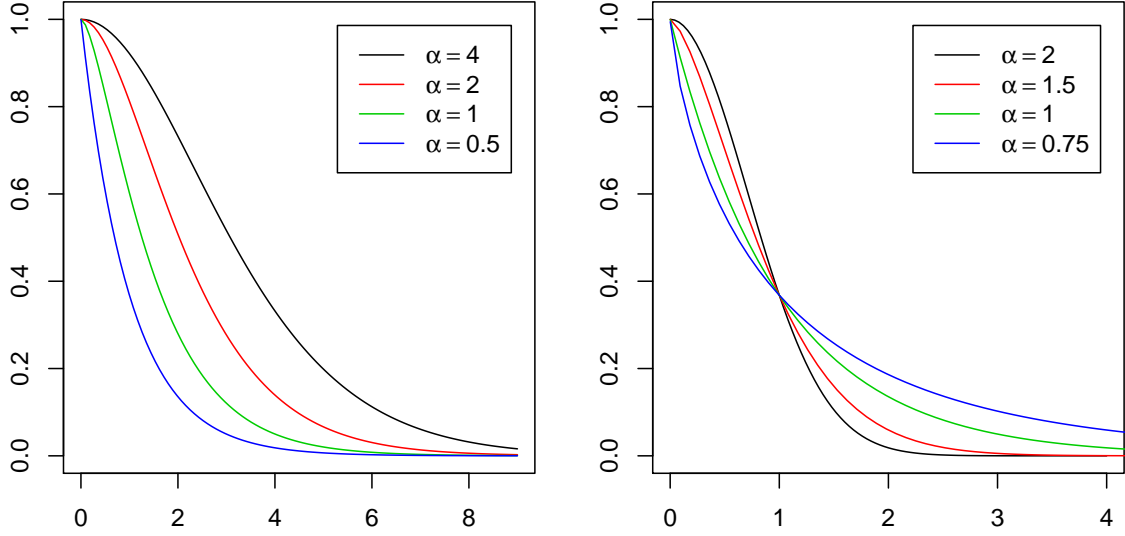
A Schlather-modell valóban magában foglalja az 1.3.2-beli modellt, azon megkötések mellett, hogy $S = \mathbb{R}^d$, ν a Lebesgue-mérték konstansszorozosa – és így az $\{s_i\}$ Poisson-pontfolyamat homogén –, továbbá $f(s, t) = f_0(t - s)$ valamilyen f_0 sűrűségfüggvény mellett. Ekkor az $X_t^i = f_0(t - s_i)$ választással azonos eloszlású stacionárius folyamatokat kapunk ([5] p. 7).

Természetesen a gyakorlatban való alkalmazhatósághoz további feltevések szükségesek az X folyamatról. Schlather megmutatta ([6] Appendix) hogy ha X olyan, (1.7)-t teljesítő stacionárius Gauss-folyamat, melynek pontonkénti határeloszlásai sztenderd normálisok, továbbá tetszőleges (t, s) pontpárra $\text{corr}(X_t, X_s) = \rho(t - s)$ valamilyen $\rho : \mathbb{R}^d \rightarrow [-1, 1]$ korrelációfüggvény mellett, akkor a kétdimenziós eloszlások így fejezhetőek ki:

$$P(Y_{t_1} \leq y_1, Y_{t_2} \leq y_2) = \exp \left[-\frac{1}{2} \left(\frac{1}{y_1} + \frac{1}{y_2} \right) \left(1 + \sqrt{1 - 2(\rho(t_1 - t_2) + 1) \frac{y_1 y_2}{(y_1 + y_2)^2}} \right) \right], \quad (1.9)$$

Ha $\rho(t_1 - t_2) = -1$, vagyis X_{t_1} és X_{t_2} mozgása tökéletesen ellentétes, akkor Y_{t_1} és Y_{t_2} függetlenek. Ennek szemléletes magyarázata, hogy ilyenkor az i indexeket az X^i -k előjele szerint szétválogatva független, identikus Schlather-modellekből származtathatjuk a folyamat értékeit a két pontban. Amikor $\rho(t_1 - t_2) > -1$, akkor Y_{t_1} és Y_{t_2} pozitív korrelációjú; a tökéletes összefüggőség pedig természetesen $\rho(t_1 - t_2) = 1$ esetén érhető el ([6] p. 39). Az extrémális együttható értéke $\theta(t_1 - t_2) = 1 + \sqrt{\frac{1 - \rho(t_1 - t_2)}{2}}$.

Egy korrelációfüggvényt *izotróp*nak mondunk, amikor értéke az iránytól nem, csak a pontok közötti euklideszi távolságtól függ. Ekkor kezelhetjük egyváltozós, $\rho : [0, +\infty) \rightarrow [-1, 1]$ függvényként, és ezzel az változattal $\text{corr}(X_t, X_s) = \rho(|t - s|)$ írható. Mivel $\rho(0) = 1$ -nek triviálisan teljesülnie kell, a további tárgyalás során csak pozitív argumentummal foglalkozunk. Alább mutatjuk be az izotróp korrelációfüggvények néhány, gyakorlatban használatos paraméteres családját, melyeket a **SpatialExtremes** csomag is támogat ([5] p. 7):

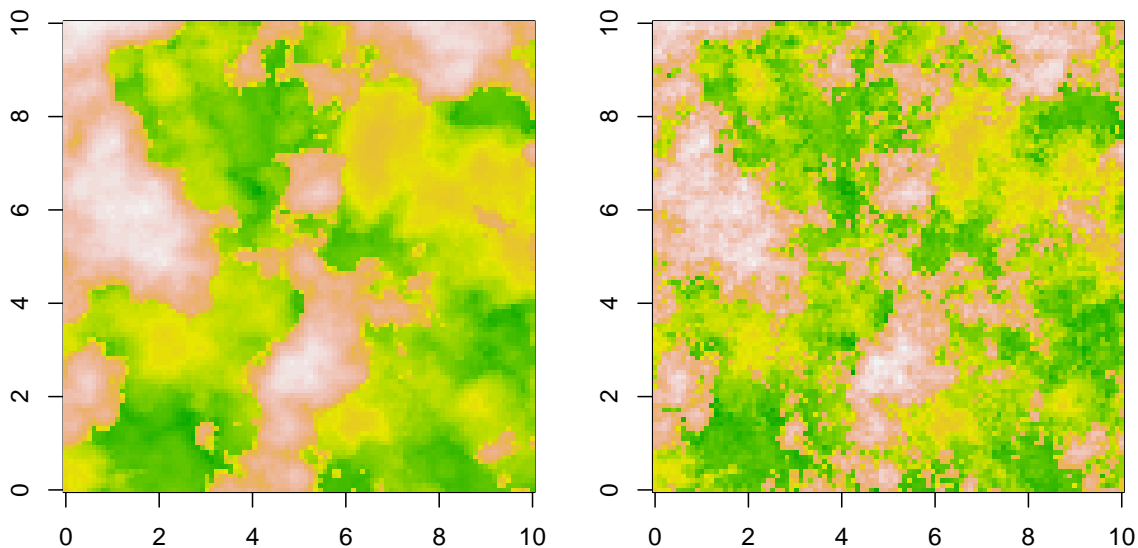


1.2. ábra. A Whittle–Matérn és hatványexponenciális korrelációfüggvények grafikonjai $c = 1$ mellett, α értékét futtatva, *nugget* hatás nélkül.

Whittle–Matérn	$\rho(h) = \frac{2^{1-\alpha}}{\Gamma(\alpha)} \left(\frac{h}{c}\right)^\alpha K_\alpha\left(\frac{h}{c}\right),$	$c > 0, \quad \alpha > 0$
Cauchy	$\rho(h) = \left[1 + \left(\frac{h}{c}\right)^2\right]^{-\alpha},$	$c > 0, \quad \alpha > 0$
Általánosított Cauchy	$\rho(h) = \left[1 + \left(\frac{h}{c}\right)^{\alpha_2}\right]^{-\frac{\alpha}{\alpha_2}},$	$c > 0, \quad \alpha, \alpha_2 > 0$
Hatványexponenciális	$\rho(h) = \exp\left[-\left(\frac{h}{c}\right)^\alpha\right],$	$c > 0, \quad 0 < \alpha \leq 2$
Bessel	$\rho(h) = \Gamma(\alpha + 1) \left(\frac{2c}{h}\right)^\alpha J_\alpha\left(\frac{h}{c}\right),$	$c > 0, \quad \alpha \geq \frac{d-2}{2}$

Az α és α_2 alakparaméterek határozzák meg a korreláció lecsengésének sebességét, és így a kapott max-stabilis folyamat simaságát (*smooth*; vessük össze az 1.2 és 1.3 ábrákat), c pedig skálaparaméter (*range*). Γ a Gamma-függvény, J_α és K_α pedig α -rendű Bessel-függvények. Az előbbi elsőfajú Bessel-függvény, amely a 0 körül oszcillál, és az abszolút értéke $h^{-\frac{1}{2}}$ rendben csökken; az utóbbi pedig másodfajú módosított Bessel-függvény, amely exponenciális sebességgel csökkenve konvergál a 0-hoz (lásd [11]). Ezek alapján belátható, hogy a fenti függvények a 0-ban 1-hez, a végtelenben 0-hoz tartanak, és a negatív értéket is felvelő Bessel kivételével monoton csökkenőek.

Tapasztalati adatokra való illesztés során indokolt lehet egy olyan 1-nél nem nagyobb konstans becslése is, amely szorzótényezőként áll a fenti alakok valamelyikét felvevő korrelációfüggvény előtt (*sill* paraméter); annak ellenére, hogy ez sérti a folytonosságot a 0 környezetében. Ez az ún. *nugget* hatás. Az 1.8-as verziótól kezdődően,



1.3. ábra. A Schlather-modell szimulációi Whittle–Matérn és hatványexponenciális korrelációfüggvények mellett, $\alpha = c = 1$, paraméterekkel, *nugget* hatás nélkül. A jobb megjelenítés érdekében most is Gumbel-eloszlásra transzformáltuk a peremeket.

a `SpatialExtremes` csomag `fitmaxstab()` függvénye a *sill* helyett az ugrás mértékét kifejező *nugget* paramétert becsüli, amely a *sill*-hez hozzáadva éppen 1-et ad.

Izotróp korrelációfüggvényekből könnyen gyárthatunk ún. elliptikusokat, amelyek szintfelületei gömbök helyett ellipszoidok:

$$\rho_e(t - s) = \rho \left(\sqrt{(t - s)^\top A (t - s)} \right),$$

ahol A tetszőleges pozitív szemidefinit mátrix.

1.4. Illesztés tapasztalati adatokra

A marginálisok együttes eloszlására az (1.5) és (1.9) eredményekkel találkoztunk eddig. Sajnos kettőnél több marginális esetén nem ismert ilyen egzakt formula ([5], p. 29), ezért maximum likelihood becslés nem végezhető a szokásos módon. Helyette használhatjuk az ún. *kompozit likelihood* módszerek egyik fajtáját, a *páronkénti likelihood* módszerét.

Általánosan ([9] p. 2)): Adott az (S, \mathcal{F}) téren az Y valós értékű, $y \mapsto f(\theta, y)$ sűrűségfüggvényű véletlen folyamat. Az ismeretlen θ , p -dimenziós paramétervektort sze-

retnénk becsülni az Y adott $y \in \mathbb{R}^S$ realizációja mellett. Legyen $\{\mathcal{A}_1, \dots, \mathcal{A}_K\} \subset 2^{\mathcal{F}}$ marginálisokkal kapcsolatos, esetleg feltételes σ -algebrák halmaza, a hozzájuk rendelt súlyok w_k , a hozzájuk tartozó likelihoodok pedig $\theta \mapsto f_k(\theta, y)$, $1 \leq k \leq K$. Ekkor a kompozit log-likelihood:

$$\ell_c(\theta, y) = \log \text{LC}(\theta, y) = \sum_{k=1}^K w_k \cdot \log f_k(\theta, y).$$

Ha az Y -ra n darab független megfigyelésünk van: $\underline{y} = (y_1, y_2, \dots, y_n)$, akkor

$$\ell_c(\theta, \underline{y}) = \sum_{l=1}^n \ell_c(\theta, y_l)$$

Speciálisan a mi esetünkben ([5] p. 30): Legyen N darab mérési pontunk az \mathbb{R}^d téren, továbbá jelölje minden $1 \leq i < j \leq N$ párra $n_{i,j}$ azon független megfigyelések számát, melyekben mind az i -edik, mind a j -edik helyről van adatunk, valamint $\underline{y}_{i,j}^k$ az ilyen közös megfigyelésekből a k -adikat ábrázoló kétdimenziós vektort. Jelölje ezeken felül $f_{i,j}(\theta; \cdot, \cdot)$ az illeszteni kívánt, egységnyi Fréchet peremű max-stabilis folyamat kétdimenziós peremeloszlásának sűrűségfüggvényét, paraméterek adott θ vektora mellett. Ekkor a páronkénti log-likelihood:

$$\ell_p(\theta, \underline{y}) = \log \text{LP}(\theta, \underline{y}) = \sum_{i < j} \sum_{k=1}^{n_{i,j}} \log f_{i,j}(\theta; \underline{y}_{i,j}^k).$$

A paraméter becsült értékét, $\hat{\theta}_c$ -t az $u(\theta, \underline{y}) = \nabla_{\theta} \ell_c(\theta, \underline{y})$ függvény gyökeként kaphatjuk meg, amely az egyes peremeken vett log-likelihoodok deriváltjainak lineáris kombinációja. A kompozit likelihood természetesen nem a valódi likelihood, hiszen a különféle peremeloszlások nem függetlenek; ezért a becslés hibájának megállapításhoz nem használhatjuk a Fisher-féle információs mátrixot. Helyette bevezetjük a következő, *Godambe-féle információs mátrixot* ([9] p. 4):

$$G(\theta) = H(\theta)J(\theta)^{-1}H(\theta), \quad \text{ahol}$$

$$H(\theta) = E_{\theta}[-\nabla_{\theta} u(\theta, Y)], \quad J(\theta) = \text{var}_{\theta}(u(\theta, Y))$$

Ha ℓ_c a klasszikus log-likelihood, akkor G , H és J is megegyeznek a Fisher-féle információs mátrixszal. A kompozit likelihood aszimptotikus viselkedéséről a következő eredmény ismert ([9] p. 5): ha n tart a végtelenhez, akkor

$$\sqrt{n}(\hat{\theta}_c - \theta) \xrightarrow{d} N_p(0, G^{-1}(\theta))$$

Becslésünk tehát aszimptotikusan torzítatlan, az aszimptotikus hatásosságát pedig a $G(\theta)$ és az $I(\theta)$, a Fisher-információ összehasonlításával lehet jellemezni – a mi max-stabilis folyamatainkra persze ez csak szimulációval lehetséges. Bizonyos speciális esetben a páronkénti likelihood hatásos, sőt, ugyanarra az eredményre vezet, mint a maximum likelihood; ilyen például a többdimenziós normális eloszlás paramétereinek becslése ([9], p. 19).

A $\hat{\theta}_c$ megtalálása után tapasztalati becslés adható a H és J mátrixokra, amely, a sztenderd hiba kiszámítása mellett, felhasználható modellszelekcióra is. A maximum likelihood becslésnél használt Akaike-féle információs kritériumot (AIC) a kompozit likelihood módszereknél a *Takeuchi-féle információs kritérium* (TIC) helyettesíti ([5] p. 36):

$$\text{TIC} = -2\ell(\hat{\theta}_c) - 2\text{tr}\left[\hat{J}(\hat{\theta}_c)\hat{H}^{-1}(\hat{\theta}_c)\right]$$

Azt a modellt tekintjük jobbnak, amelyikre a fenti érték kisebb.

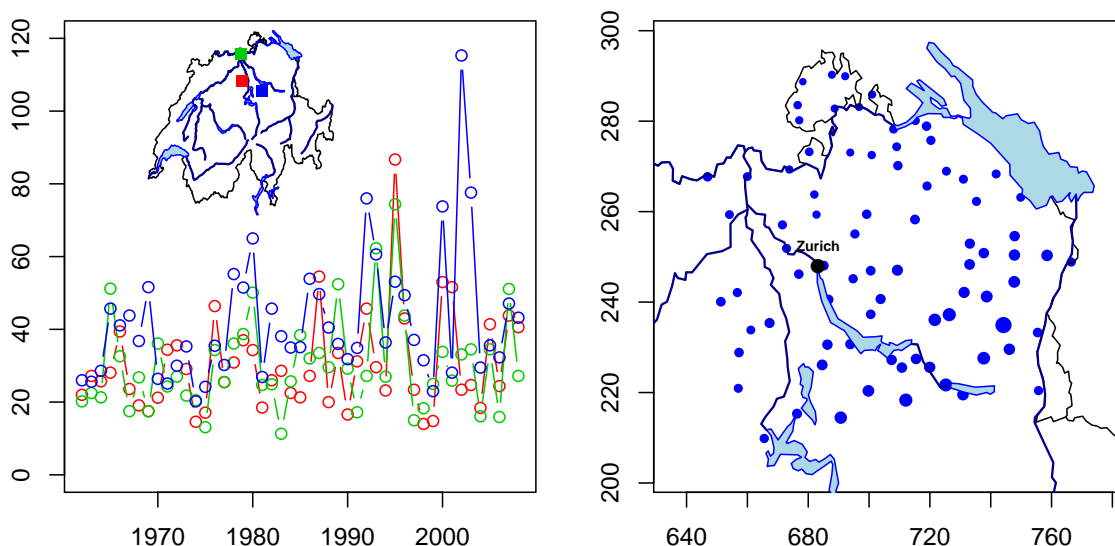
2. fejezet

Alkalmazás valós földrajzi adatokra

Már a klasszikus extrémérték-elméletet születését is természeti folyamatok (pl. vízállás változása) leírása motiválta, a pénzügyi kockázatok terén később kezdték el alkalmazni. Az extrém jelenségek térbeli struktúrájának elemzése is különösen a meteorológiában terjedt el a közelmúltban, (pl. szélerősség, csapadékmennyiség [1], [3], hóvastagság [2]). Én is ezen a nyomon elindulva mutatom be az előző fejezetben ismertetett max-stabilis modellek alkalmazását a gyakorlatban.

Ehhez a `SpatialExtremes` csomaghoz mellékeltem `rainfall` adatbázist használtam fel, amely két mátrixot tartalmaz. A `rain` 79 északkelet-svájci helyszínen mm-ben mért napi csapadékmennyiségek éves maximumait tartalmazza 1962 és 2008 között, a `coord` pedig ezen mérési pontok szélességi és hosszúsági koordinátáit, valamint a méterben mért tengerszint feletti magasságot. Az első két dimenzió km-ben van megadva úgy, hogy a Svájc térképét ábrázoló `swiss()` függvénnyel kompatibilis legyen. A most következő vizsgáltokhoz és szemléltetésekhez csak e két dimenziót fogom használni.

Marginálisok vizsgálata. Nelder–Mead-féle numerikus optimalizációval becsültem GEV-paramétereket, egyrészt mind a 49×79 adatra, másrészt külön-külön az egyes mérési pontokra. Ezután QQ-plotokkal ellenőriztem az illeszkedést, ami kiváló lett az egész adatsort tekintve, és a marginálisok esetén sem tapasztalhatók olyan nagy kilengések a teoretikus eloszlástól, ami miatt le kellene mondanunk az extrémális modellezésről (2.2 ábra). A becsült alakparaméterek többnyire 0,1 nagyságrendű pozitív számok, a csapadékmaximumok tehát enyhén Fréchet-jellegűek.



2.1. ábra. *Bal oldal:* A mm-ben mért napi csapadékértékek éves maximumainak alakulása három mérési pontban ($V1$, $V10$ és $V20$ jelűek az adatsorban) 1962 és 2008 között. *Jobb oldal:* A 79 mérési pont. A körök átmérője a csapadékmaximumok 49 éves átlagával arányos. A tengelybeosztás km-ben értendő.

Kétdimenziós illesztések. Az így becült GEV-paraméterekkel egységnyi Fréchet-eloszlásra transzformáltam a peremeket, és kétdimenziós max-stabilis modellek illesztésével próbálkoztam, a földrajzi szélesség és hosszúság figyelembe vételével. A Smith-modellből származó kovarianciamátrix elemei $\sigma_{11} = 362,7$, $\sigma_{12} = 55,4$ és $\sigma_{22} = 209,9$, ami nyugat-keleti irányban kissé megnyúlt záporokat jelent (2.3 ábrán felül az első). A paraméterek sztenderd hibái rendre 3,4, 2,6 és 5,4.

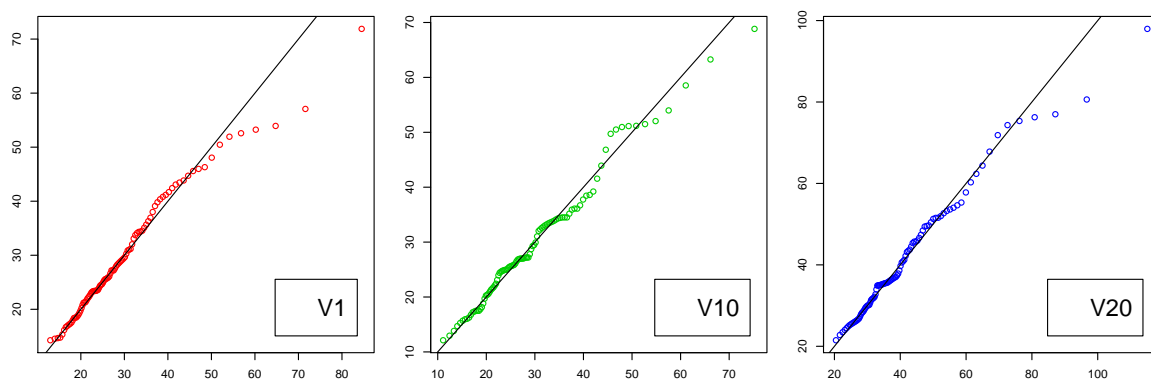
Am a páronkénti devianciák és a TIC-értékek alapján a Schlather-modellek jobban teljesítettek a Smith-modellnél, különösen a Bessel-függvénnyel illesztett. Ez utóbbinak a jósága akkor látszik igazán, amikor összehasonlítjuk az adatainkból Schlather–Tawn-módszerrel ([5] p. 20-21, [1]) becült páronkénti extrémális együtthatókat a modellből származókkal (2.3 ábrán alul a harmadik).

A Besselre becült paraméterek: $nugget = 0,32$, $range = 2,19$, $smooth = 150,22$. Sztenderd hibákat és TIC-et itt nem tudott számolni a program, de a páronkénti deviancia sokkal kisebb, mint a többi sűrűségfüggvénnyel.

Észrevettem, hogy a marginálisokra illesztett GEV-eloszlások eltolás- és skálapa-

	μ	σ	ξ
Teljes	26,84	10,55	0,15
V1	23,91	8,24	0,19
V10	24,17	9,10	0,08
V20	33,51	10,10	0,23

2.1. táblázat. Becsült GEV-paraméterek 1.1.2 jelöléseivel, a teljes adathalmazra és három mérési pontra.



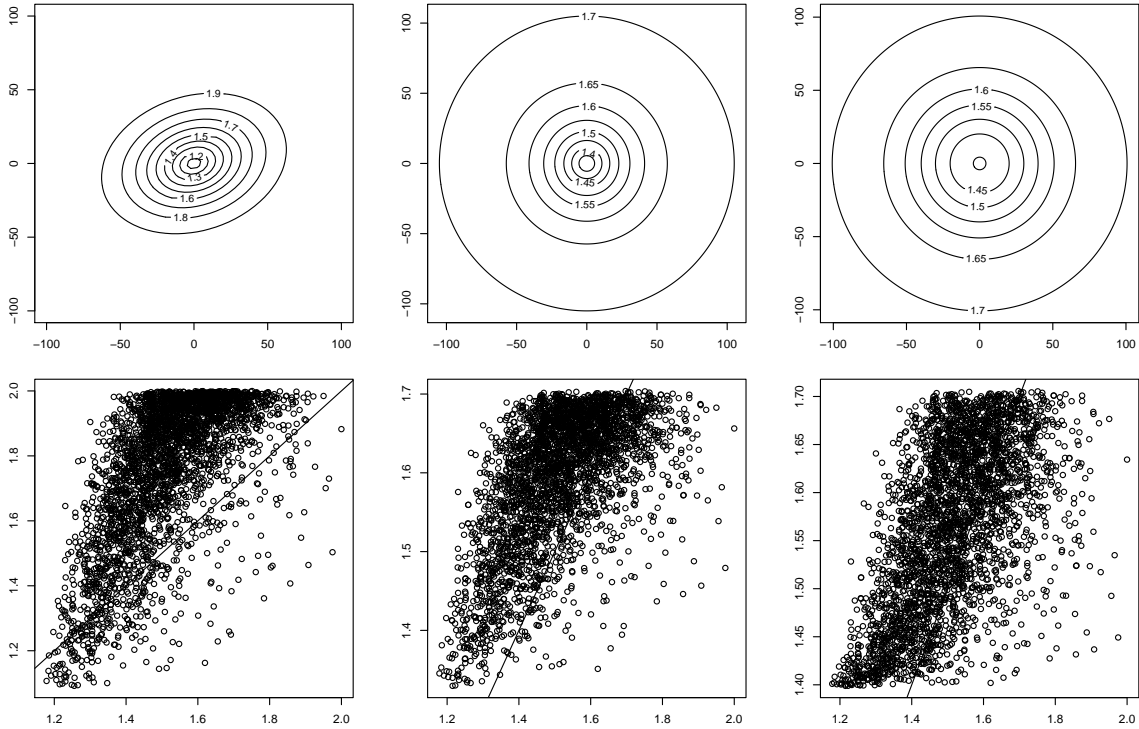
2.2. ábra. A GEV-illesztések jóságát szemléltető QQ-plotok a 2.1 ábrán is vizsgált három helyszínen. A vízszintes tengelyről az illesztett eloszláshoz tartozó, a függőlegesről a tapasztalati percentilisek olvashatók le (csapadék mm-ben).

paraméterei trendszerűen növekednek a vizsgált földrajzi terület délkeleti sarka felé haladva (ez tükröződik 2.1 ábrán is). Ezért olyan illesztéssel is megpróbálkoztam, amely – a Bessel-függvény együtthatói mellett – e két GEV-paramétert a földrajzi szélesség és hosszúság lineáris függvényeként becsüli; az alakparamétert pedig konstansnak vettem. Az így keletkezett modellben a korrelációs függvény paraméterei: $nugget=0,36$, $range=2,26$, $smooth=150,26$; a marginálisok struktúrájára pedig a következő jött ki:

$$\mu = 21,87 + 0,041 \cdot lon - 0,108 \cdot lat$$

$$\sigma = 9,10 + 0,006 \cdot lon - 0,023 \cdot lat$$

$$\xi \equiv 0,135$$



2.3. ábra. *Balról jobbra:* Smith-, hatványexponenciális és Bessel-modell. *Felül:* Az elméleti extrémális együttthatók szintvonalai a síkon. *Alul:* A Schlather–Tawn-módszerrel páronként becsült extrémális együttthatókat (vízszintes) a modell alapján kalkuláltakkal (függőleges) összehasonlító QQ-plotok.

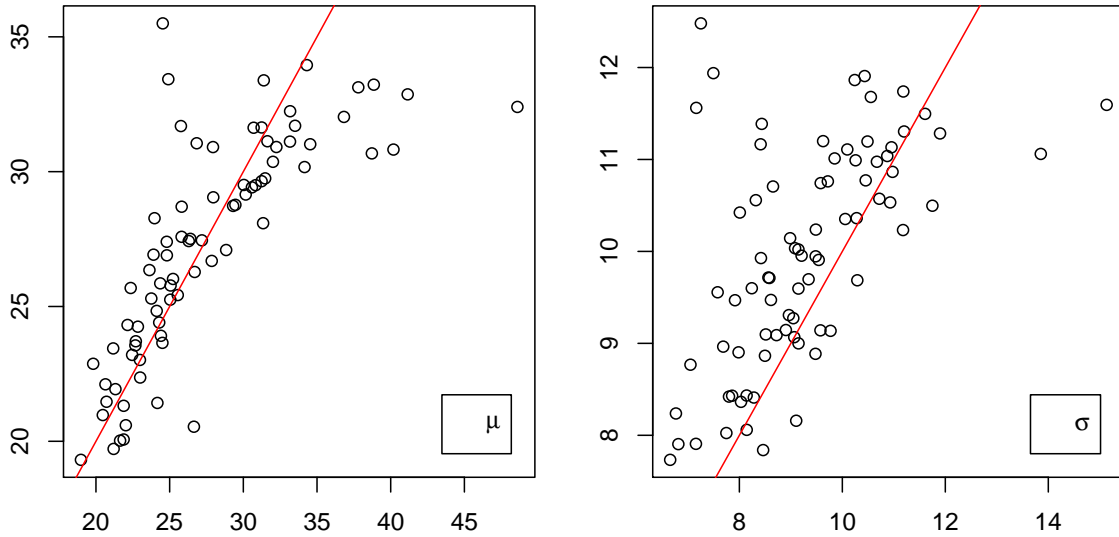
Itt „lon” jelenti a hosszúságot és „lat” a szélességet. Az eredeti GEV-becslésünkkel való összehasonlítás után úgy találtam, hogy a lineáris modell jó megközelítés volt, csupán a konstans meghatározásában mutatkozott bizonytalanság, modellünk ugyanis szisztematikusan alulbecslülte a Nelder–Mead-optimalizálás eredményét.

Az R-nek azonban nehézséget okozott a fenti modellel való szimuláció, ezért lefuttattam egy ugyanilyet hatványexponenciális függvénnyel is, amely a többiek közül a legjobb TIC-et adta. A becsült paraméterek: $nugget=0,21$, $range=36,95$, $smooth=1,44$, sztenderd hibáik rendre 0,09, 8,61 és 0,51; a lineáris struktúra pedig:

$$\mu = 19,29 + 0,068 \cdot lon - 0,160 \cdot lat$$

$$\sigma = 3,10 + 0,025 \cdot lon - 0,042 \cdot lat$$

$$\xi \equiv 0,188$$

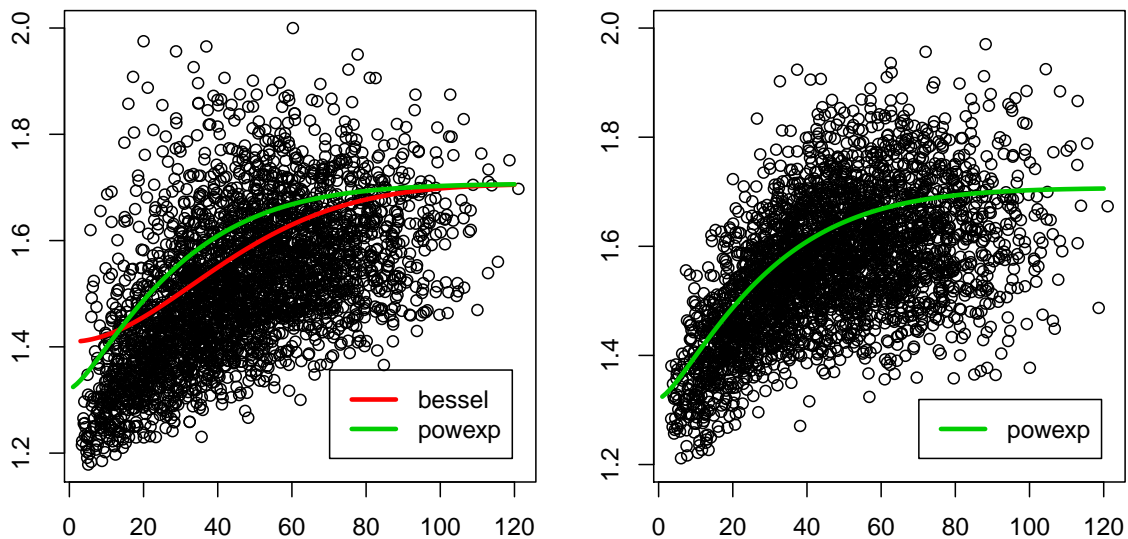


2.4. ábra. A mintánkra eredetileg becsült (vízszintes) és a lineáris modellből származó (függőleges) GEV-együtthatókat összehasonlító „QQ-plotok”, hatványexponenciális Schlather-modell esetén.

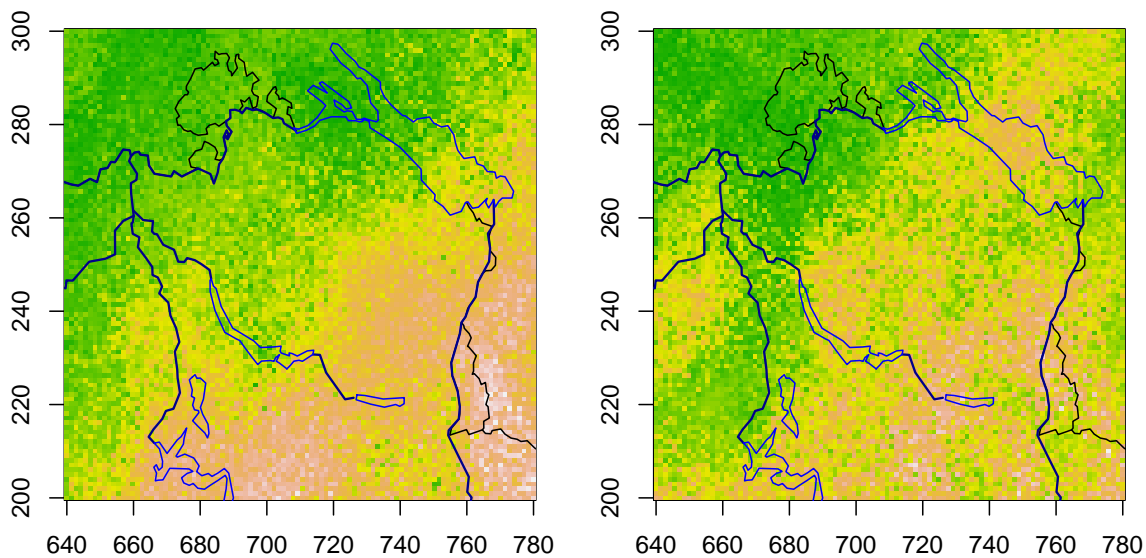
Ez már jobban illeszkedett a Nelder–Mead-optimalizálás eredményéhez, a bizonytalan konstans miatti eltolódás enyhült (2.4 ábra). A most kapott Bessel- és hatványexponenciális modellek összevethetők az alapján is is, mennyire írja le jól a modellből származó, a távolságtól függő extrémális együttható a valódi összefüggőségi struktúrát. Ezt a szempontot már figyelembe vettük a 2.3 ábra QQ-plotjain, de egyben is láthatjuk az összehasonlítást a 2.5 ábra bal felén. Ez árnyalja a képet, mivel látszik, hogy a Bessel-modell kis távolságokra nem illeszkedik jól. Az ábra jobb fele a hatványexponenciális modell 49 független szimulációjából számolt extrémális együtthatókat mutatja.

A lineáris GEV-moddal bővített illesztésből az egész téren is szimulálhatunk. A hatványexponenciális változatból készítettem szimulációkat egy 100×100 -as, Északkelet-Svájc térképére illeszkedő rácsra. Először egységnyi Fréchet-peremű Schlather-modellt szimuláltam az illesztett hatványexponenciális korrelációfüggvénnyel, majd az egyes rácspontokat a lineáris modellből adódó GEV-eloszlásúvá transzformáltam. Két ilyen szimuláció eredménye a 2.6 ábrán látható.

A fentiek alapján láthatjuk, hogy a max-stabilis modellek illesztése valóban jó és rugalmas módszer lehet az extrém meteorológiai folyamatok megértéséhez.



2.5. ábra. Schlather–Tawn-módszerrel becsült páronkénti extrémális együtthatók a km-ben mért távolság függvényében, bal oldalon az eredeti, jobb oldalon a hatványexponenciális modellből szimulált csapadékmaximumokra. Velük együtt ábráztuk az eredeti adatainkra illesztett Schlather-modellekből származó elméleti együtthatófüggvényeket.



2.6. ábra. Az éves csapadékmaximumok szimulációi hatványexponenciális Schlather-modellel. A zöld árnyalatok 20 mm körüli, a pirosak 60 mm körüli csapadékot jeleznek.

3. fejezet

Alkalmazás biztosítási káradatokra

3.1. Motivációk

Először is vegyük sorra, hogy a vagyombiztosítási károk időbeli és térbeli eloszlásának mely tulajdonságai motiválhatják a max-stabilis folyamatokkal való megközelítést, illetve milyen nehézségekkel kell megküzdenünk, ha ilyen jellegű elemzésbe fogunk.

Az információk gazdaságos felhasználása szempontjából komoly előny lehet, ha adott időegység alatt, adott földrajzi helyszínen képződött kötelezettségek értékét lényegében néhány extrém nagy kár határozza meg, és a kis károk összértéke hozzájuk képest vagy elenyésző, vagy legalábbis kis szórású, konstansként kezelhető a diverzifikációnak köszönhetően. Ekkor ugyanis nem feltétlenül szükséges maximumokat venni, maga a nyers adatsor is mutathat max-stabilis jelleget, GEV-szerű marginálisokat. Különösen a felelősségbiztosításban tipikus a károk ilyen alakulása.

Másik fontos jellemzője a vagyombiztosítási károknak, hogy egy-egy természeti katasztrófa, vagy más rendkívüli esemény hatására adott időintervallumban és földrajzi régióban mind a gyakoriságukban, mint nagyságukban megnövekedhetnek. Ezt a hatást neveztük az 1.3.2 és 1.3.3 szakaszokban szemléletesen záporseménynek.

Tekinthetjük csak az idő dimenzióját is, ekkor a megfigyeléseink, amelyekre max-stabilis folyamatot illesztünk, lehetnek egymást követő hetek, hónapok, évek stb. A szezonalitást vizsgálhatjuk ezen elmélet keretei között is, például az egyes napokra, mint marginálisokra történő GEV-beccsléssel. A szezonális jelleg miatt elképzelhető, hogy egy-egy megfigyelt időintervallum eleje és vége között szorosabb az együttmoz-

gás, mint ami a tényleges időbeli távolságból következne. Ekkor jó ötlet lehet például, ha a marginálisokat egyenes szakasz helyett körvonalon helyezzük el, és így illesztünk rájuk kétdimenziós Schlather-modellt.

A trendet, inflációt könnyen kiszűrhetjük az idősorból, gondot jelenthet viszont, ha a volatilitás is változik benne. Ezt a stacionárius modellejeinkkel aligha tudjuk hűen visszadni; jóval nehezebben vizsgálható, heteroszkedasztikus modellek kellenének, de ezek illesztése nem célunk ebben a dolgozatban.

Problémát jelent az is, ha nem tudjuk a kötelezettségeket a károk bekövetkezésének ideje szerint rendszerezni, hanem csak a károk bejelentésének, vagy kifizetésének időpontja érhető el. Ekkor egyrészt sérül az összefüggőségi struktúra, másrészt a katasztrófák fentebb említett hatása is csak időben elnyújtva, vagy egyáltalán nem érzékelhető. Ekkor segíthet, ha mozgó átlagolással simítjuk az adatsort, vagy pedig, ha elég sok adatunk van, elemezhetünk klasszikus módon maximumokat is.

3.2. Egy konkrét idősor elemzése

Valós és kellő részletességű biztosítói kárkimutatásokhoz nem könnyű hozzájutni, az ilyen információ legtöbbször üzleti titok.

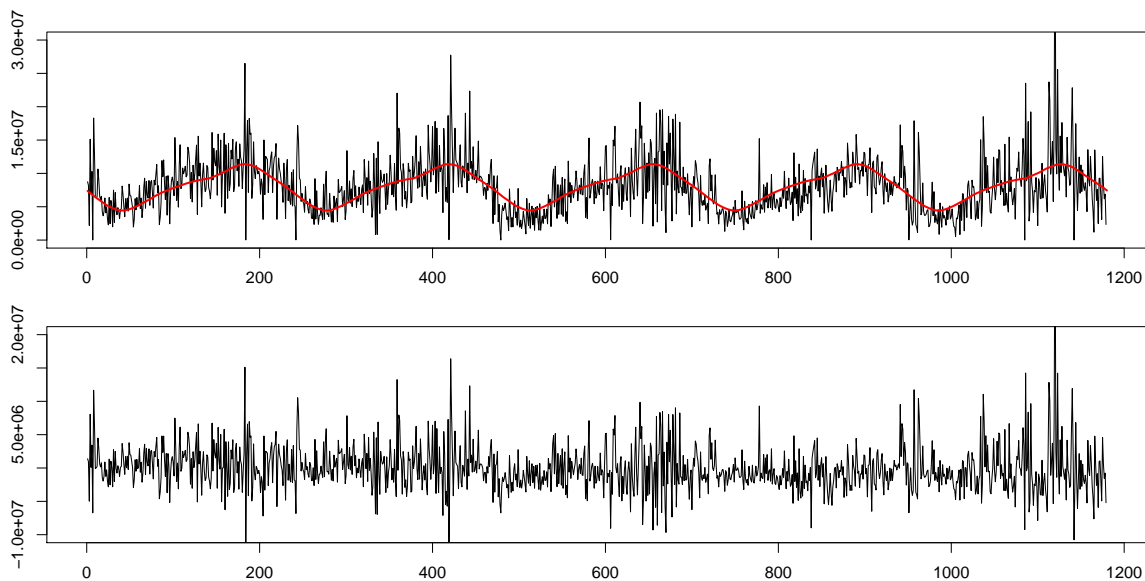
Az idősort, amelyet módomban állt elemezni, egy magyar biztosító valamely vagyonszámításos üzletágának napi kárkifizetési adataiból állították elő, természetesen némileg transzformálva a tényleges pénzmennyiségeket. Az idősor 1826 egymást követő nappól áll, melyek közül csupán 1179-en történt kárkifizetés. A héttel osztható sorszámú napokon soha nincs pozitív adat, és az őket megelőző napok túlnyomó többségén sem; nyilvánvalóan ezek a hétvégék. A tipikus hét öt kárkifizetési nappól áll, de sok csonka hét is van, néhol egész hetek hiányzanak.

Ezen megfigyelések alapján adataink alkalmasak lehetnek az ötnapos munkahét vizsgálatára, nagyobb időegységekre viszont nem. Év nagyságrendűre azért nem, mert nagyon kevés lenne a megfigyelés, hónap nagyságrendűre pedig azért nem, mert nagyon nehéz az adatsort egybevágó időintervallumokra felosztani, annak lukacsos szerkezete miatt. Heti maximumokat is képezhetünk, és elemi módon vizsgálhatjuk őket, de max-stabilis folyamatok illesztéséhez ezek sem adnak elégséges méretű információhalmazt. Fő célként marad tehát a teljesen ismert munkahetek elemzése, amelyekből összesen

168 darab van.

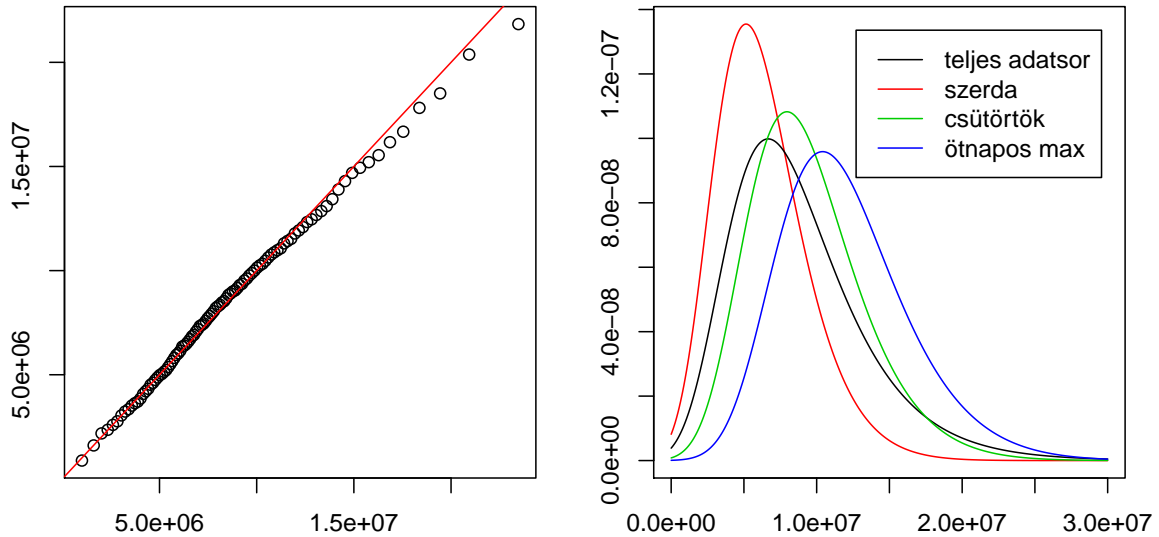
A számos megközelítés és modellillesztés közül – amelyekkel az elemzés során többkevesebb sikerrel kísérleteztem, és próbáltam a lehető legtöbbet kihozni az adatokból – itt egyetlen irányvonalat szeretnék bemutatni, amely, ha erre a szerény méretű és torzított adatsorra nem is ad látványos eredményt, de a biztosítók számára megítélesem szerint hasznos lehet.

Inflációs és szezonális kiigazítás. A hetek vizsgálata során mind a trend, mind a szezonális zavaró tényező, amit ki kell szűrni. Az idősorban szignifikáns lineáris trend jelentkezik, de a növekedés értelmezhető inflációként is. Én ez utóbbi értelmezést választottam abból a megfontolásból, hogy pénzügyi adatokra ez egészen biztosan hatással van. Az exponenciális trendként becsült öt éves infláció mintegy 18,9%-nak adódott. Ez átlagosan 3,5%-os inflációt jelent évente, ami valószínű is az elmúlt évtized magyar gazdasági viszonyai között. Az ezzel való korrekció után nem maradt szignifikáns lineáris trend az idősorban. Ezután a szezonalitást becsültem lineárisan simított napi átlagolással. Ezt kivonva is szignifikáns maradt az ötnapos autokorreláció, megerősítve számunkra, hogy érdemes a munkahetet tanulmányozni.



3.1. ábra. Fent a hiányzó adatoktól tisztított, inflációval korrigált idősor és a rá illesztett szezonális zavaró tényező. Alatta a kettő különbsége, a szezonálisan kiigazított idősor.

GEV-illesztések, és a heti maximumok. A 3.1-beli hipotézisünk teljesült: kitűnően illeszthető GEV-eloszlás a tisztított adatainkra, speciálisan Gumbel is – persze lehet, hogy mesterségesen transzformálták ilyenné az idősort. A Gumbel-jelleg, ahogyan azt várnánk is, átöröklődik mind a teljes ötnapos hetek, mind a hétnapos hetek maximumaira, bár az utóbbira való becslés bizonytalanabb az adatsor inhomogén szerkezete miatt.



3.2. ábra. *Bal oldal:* A teljes idősorra való GEV-becslés jóságát szemléltető QQ-plot, vízszintes tengelyen az elméleti, függőlegesen a tapasztalati percentilisekkel. *Jobb oldal:* Az ebben, és még három, a teljesen ismert ötnapos munkahetekre való GEV-becslésben kapott elméleti sűrűségfüggvények.

Az egyes munkanapokon való kárkifizetésekre is szépen illeszthető GEV-eloszlás, de itt a becsült alakparaméterek $-0,1$ nagyságrendűek, enyhén Weibull-szerűek. Ennek az lehet az oka, hogy a csonka hetek napjaira nagyobb kifizetések koncentrálnak, amelyek most kiesnek a látókörünkéből. Az eltolás- és skálaparaméterek értékei is közel azonosak az egyes napokon, csak a szerda lóg ki közülük, amikor átlagosan 30 – 40%-kal kevesebb a kártérítés mértéke, mint a többi napon.

Mindezek a megállapítások igazak akkor is, ha a szezonálisan igazított adatsorból indulunk ki, bár az illeszkedés jósága kissé romlik.

	μ	σ	ξ
hétfő	6973459	3523604	-0,054
kedd	6987113	3354810	-0,063
szerda	4905739	2725561	-0,085
csütörtök	7647442	3411493	-0,087
péntek	7223460	3711054	-0,062

3.1. táblázat. Becsült GEV-paraméterek a teljes ötnapos munkahetek napjaira.

Schlather-modell illesztése a munkahétre. A marginális feltételek teljesüléséről meggyőződünk, már csak az összefüggőségi struktúra a kérdéses. A hét napjai közötti Pearson-féle korrelációk részben szignifikánsak maradtak a szezonális levonása után is, de nem alkotnak szabályos struktúrát, ami egy korrelációs függvény illesztéséhez kellene. A GEV-bebecslések felhasználásával egységnyi Fréchet-re transzformáltam a peremeket, de ez szinte teljesen eltüntette ezeket a szignifikáns korrelációkat is.

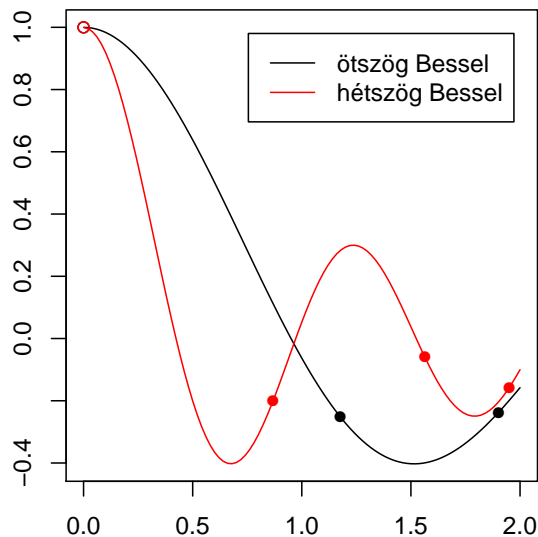
Ennek ellenére megpróbálkoztam az illesztéssel. A napokhoz háromféleképpen rendeltem koordinátákat:

1. Egy dimenzióban, számok 1-től 5-ig.
2. Két dimenzióban, egységkörbe írt szabályos ötszög csúcsai.
3. Két dimenzióban, egységkörbe írt szabályos hétszög szomszédos csúcsai.

Az 1.3.3-ban ismertetett izotróp korrelációfüggvények közül az általánosított Cauchy kivételével – amelyet illeszteni ugyan lehet a `SpatialExtremes`-ben, de szimulálni vele nem – mindet kipróbáltam.

A legtöbb becslésben a program nem tudott sem TIC értékeket számolni, sem sztenderd hibákat – vagy ha igen, akkor nagyon nagyok voltak. Ezért szimulációval, a szimulált adatok korrelációs struktúrájának ellenőrzésével és újrabecsléssel vizsgáltam, mennyire fogadhatóak el. A legjobbnak a Besselrel való illesztések bizonyultak. Ez nem meglepő, hiszen a megismert függvénycsaládokból csak a Bessel mehet negatív érték-tartományba, és így vele tudunk kis pozitív korrelációkat modellezni. Az egydimenziós Bessel-modell sajnos nem futott le; a kétdimenziósokból kapott korrelációfüggvények

a 3.3 ábrán láthatók. A többi függvénytípusra elfajult becslések születtek, amelyeket ábrázolni nem lehet.



3.3. ábra. Az egyégyeni Fréchet-pereműre transzformált munkahetekre illesztett korrelációfüggvények a Bessel-családból. A pontok a sokszögek oldal- és átlóhosszainál felvett értékeket jelzik.

A Bessel-modellből kalkulált extrémális együtthatókat sem érdemes ábrázolni, hiszen csak 10 van belőlük, és ötszögre kétféle, hétszögre háromféle értéket vehetnek fel; helyette a 3.2 táblázatban gyűjtöttem össze őket, összehasonlítva a modelltől függetlenül becsült értékekkel.

Mindez azonban igen kevés információ ahhoz, hogy az előző fejezetben látottakhoz hasonlóan értékeljük modellünk jóságát. Ennek alaposabb vizsgálatához most is szimulációra van szükség. Példaként a 3.4 ábrán összehasonlítottam az adjusztált teljes munkahetek hétfői és csütörtöki napjainak együttes eloszlását két különböző szimulációban tapasztalattal. Azért ezt a két napot választottam, mert közöttük a legerősebb a tapasztalati Pearson-korreláció, 0,357, és ezért fontos ennek a viszonylatnak a helyes visszaadása. Ugyanez a tapasztalati korreláció 0,304-nek adódott az ötszögre illesztett modellben és 0,294-nek a hétszögre illesztettben. Az ábra alapján szemlátomást jól sikerült visszaadni a kapcsolatot, és az illeszthető lineáris trend is nagyon hasonló lett.

Persze ez a hasonlóság jórészt onnan származik, hogy az egységnyi Fréchet-peremű

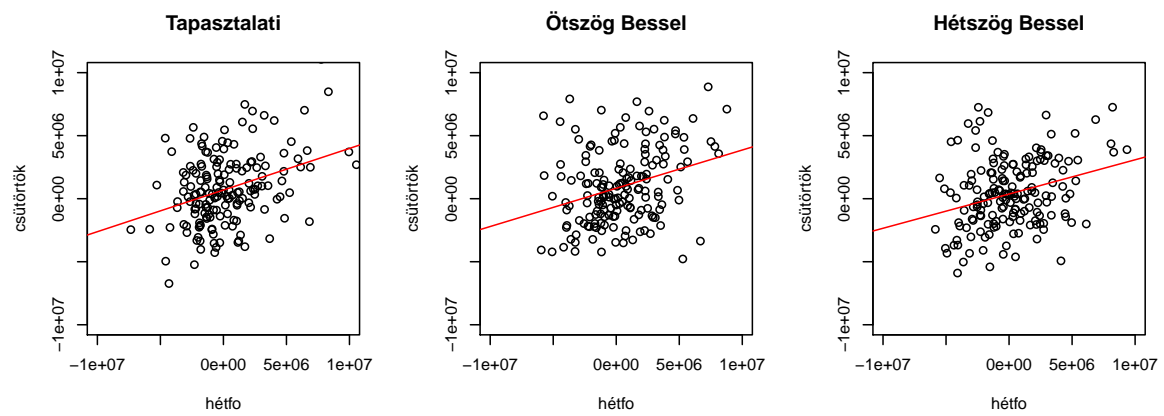
	ST	B5	B7		ST	B5	B7
H–K	1,750	1,791	1,775	K–Cs	1,763	1,786	1,728
H–Sz	1,862	1,786	1,728	K–P	1,740	1,786	1,761
H–Cs	1,693	1,786	1,761	Sz–Cs	1,961	1,791	1,775
H–P	1,977	1,791	1,761	Sz–P	1,882	1,786	1,728
K–Sz	1,764	1,791	1,775	Cs–P	2,000	1,791	1,775

3.2. táblázat. A páronkénti extrémális együttható értékei Schlather–Tawn-módszerrel, valamint az ötszögre, illetve hétszögre illesztett Bessel-modellekből becslülve.

szimulációkat ugyanazokkal a GEV-paraméterekkel transzformáltam vissza, mint amelyekkel egységnyi Fréchet-re hoztam a megfigyeléseket a becsléshez. Megjegyzendő az is, hogy modelljeink olyan nappárokhoz is szoros együttmozgást rendelnek, amelyek között szignifikáns korrelációt nem tudunk kimutatni.

Összefoglalás. A mintánk, mint már említettem, maximumok képzése nélkül is alkalmasnak bizonyult az extrémérték-modellezésre, időbeli összefüggőségen alapuló max-stabilis modellek illesztéséhez viszont egyrészt túl kicsi, másrészt nem mutatható ki benne elégséges mélységű, egybefüggő korrelációs struktúra. Ennek egyik oka lehet, hogy nem a károk bekövetkezésének, hanem a kifizetésének napját ismerhetjük meg belőle (lásd a 3.1-ben felsorolt problémákat). Mozgóátlagolás segíthetne, de a minta inhomogén szerkezete miatt ezt nem tudjuk kellő szofisztikáltsággal megtenni.

Mindazonáltal láttunk érveket amellet is, hogy biztosítási kárkimutatások alkalmasak lehetnek térbeli extrémérték-modellezésre, de természetesen ehhez elég nagy és részletes adatbázis szükséges.



3.4. ábra. A hétfők és csütörtökök együttes eloszlása a 168 kiigazított munkahétben, és annak szimulációiban, a lineáris trend feltüntetése mellett.

Irodalomjegyzék

- [1] L. **Bel**, J.-N. **Bacro**, Ch. **Lantuéjoul**: „Estimation of the extremal coefficient function of a stationary random field. Application to rainfalls maxima.”, 2006, http://www.unavarra.es/metma3/Papers/PDFS_ORAL/Bel.pdf
- [2] J.**Blanchet**, A. C. **Davison**: „Spatial Modeling of Extreme Snow Depths”, *The Annals of Applied Statistics*, **5**(3): 1699-1725, 2011.
- [3] A. C. **Davison** et al.: „Models for Spatial Extremes, with some Applications”, 2009, <http://extremes.epfl.ch/files/content/sites/extremes/files/users/111184/public/WorkshopNov09/Davison.pdf>
- [4] A. J. **McNeil**, R. **Frey**, P. **Embrechts**: *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, 2005.
- [5] M. **Ribatet**: „A User’s Guide to the SpatialExtremes Package”, melléklet a SpatialExtremes R-csomaghoz, 2009.
- [6] M. **Schlather**: „Models for Sstationary Max-stable Random Fields”, *Extremes*, **5**(1): 33-44, 2002.
- [7] R. L. **Smith**: „Max-stable Processes and Spatial Extremes”, kiadatlan kézirat, 1990, <http://www.stat.unc.edu/postscript/rs/spatex.pdf>
- [8] R. L. **Smith**: „Spatial Extremes”, 2009, <http://www.unc.edu/~rls/samsi/SpatialExtremes.pdf>
- [9] C. **Varin**, N. **Reid**, D. **Firth**: „An Overview of Composite Likelihood Methods”, *Statistica Sinica*, **21**: 5-42, 2011.

[10] „Poisson folyamatok, exponenciális eloszlások”, <http://www.math-inst.hu/~major/valszam/poisson.pdf>

[11] http://en.wikipedia.org/wiki/Bessel_function