

Dugoročna zavisnost

Danijel Grahovac*, Lucijana Grgić†

Sažetak

U analizi vremenskih nizova jedan od osnovnih ciljeva je pojave koje opažamo u vremenu opisati slučajnim procesima. Neke pojave pokazuju karakteristično jaku zavisnost tako da je čak i daleka prošlost značajno povezana s budućnosti. U ovom radu bit će opisani neki modeli dugoročne zavisnosti i njihova svojstva te metode uočavanja dugoročne zavisnosti u podacima. Primjena je ilustrirana na podacima o vodostaju rijeke Nil.

Ključne riječi: *dugoročna zavisnost, stacionarni slučajni procesi, autokorelacijska funkcija*

Long-range dependence

Abstract

In time series analysis one of the main goals is to model time dependent phenomena with stochastic processes. Some phenomena exhibit strong dependence so that even the distant past is significantly related to the future. In this paper we will describe some models of long-range dependence, their properties and methods for detecting long-range dependence in data. The application is illustrated on the yearly water levels of Nile river.

Keywords: *long-range dependence, stationary stochastic processes, autocorrelation function*

*Odjel za matematiku, Sveučilište u Osijeku, email: dgrahova@mathos.hr

†Odjel za matematiku, Sveučilište u Osijeku, email: lgrgic@mathos.hr

1 Uvod

Jedna od važnijih primjena statistike je modeliranje prirodnih pojava i procesa u raznim područjima poput hidrologije, geografije, ekonomije, klimatologije i agronomije. Kako se u tim područjima radi o procesima koji se ne mogu kontrolirati i koje pokreću komplicirani fizički mehanizmi i zakoni, za očekivati je da među promatranjima u različitim vremenskim trenucima postoji zavisnost. Odgovarajući stohastički modeli za takve pojave obično se biraju iz određene klase slučajnih procesa. Pri tome struktura zavisnosti modela treba što vjernije replicirati zavisnost opaženu na podacima.

Istraživanja u području primijenjene statistike pokazala su kako zavisnost među promatranjima, osim što se pojavljuje, može trajati duži period, odnosno prošlost u nekim procesima može imati dugoročan utjecaj na budućnost. Takve spoznaje izazivale su dosta prijepora u svoje vrijeme. Naime, do sredine 50-ih godina prošlog stoljeća prevladavali su modeli temeljeni na slučajnim procesima s tzv. Markovljevim svojstvom prema kojem je za ponašanje procesa u neposrednoj budućnosti dovoljno poznavati samo sadašnje stanje, a ne cijelu prošlost i sadašnjost.



Harold Edwin Hurst
(1880.—1978.)
britanski hidrolog

Začeci ideje o procesima sa svojstvom *dugoročnog pamćenja* pomalo neočekivano potekli su iz hidrologije i rada hidrologa Harolda Edwina Hursta [6]. Naime, zbog povećanih potreba za vodom, jedan od važnih problema u hidrologiji je izgradnja akumulacija pregrađivanjem riječnog toka branama kako bi se osigurala dostupnost vode. Pri tome visina brane diktira kapacitet akumulacije. Kapacitet akumulacije za vremensko razdoblje između trenutaka t i $t+k$ treba optimalno izabrati tako da izljev iz akumulacije bude jednolik, da u vremenskom trenutku $t+k$ akumulacija bude jednako puna kao što je bila u vremenskom trenutku t i da akumulacija ni u jednom trenutku ne bude prepunjena. Hurst je definirao tzv. *prilagođeni raspon* kao mjeru optimalne visine brane na sljedeći način. Ako je X_i priljev u vremenskom trenutku i i $Y_j = \sum_{i=1}^j X_i$ cjelokupan priljev do vremenskog trenutka j , tada je prilagođeni raspon

$$R(t, k) = \max_{0 \leq i \leq k} \left[Y_{t+i} - Y_t - \frac{i}{k} (Y_{t+k} - Y_t) \right] - \min_{0 \leq i \leq k} \left[Y_{t+i} - Y_t - \frac{i}{k} (Y_{t+k} - Y_t) \right].$$

U svrhu istraživanja svojstava koja su neovisna o skali, $R(t, k)$ se standardizira s

$$S(t, k) = \sqrt{\frac{1}{k} \sum_{i=t+1}^{t+k} (X_i - \bar{X}_{t,k})^2},$$

gdje je $\bar{X}_{t,k} = \frac{1}{k} \sum_{i=t+1}^{t+k} X_i$. Omjer

$$R/S(t,k) = \frac{R(t,k)}{S(t,k)} \quad (1)$$

se naziva *reskalirani prilagođeni raspon* ili R/S statistika (vidi [1, str. 33]) te predstavlja mjeru optimalne visine brane za period od k godina. Jasno je da će za određivanje optimalne visine biti ključne vrijednosti priljeva X_i , $i = t, \dots, t+k$. Kako se visina određuje za buduće priljeve (vodostaje) koji su nepoznati u trenutku izgradnje brane, prirodno je priljeve X_i promatrati kao slučajne varijable, odnosno niz (X_i) kao slučajan proces. Optimalna visina brane značajno će ovisiti o svojstvima procesa koji modelira vodostaje.

Hurst je u svom radu [6] promatrao kako se $R/S(t,k)$ statistika ponaša za 690 različitih vremenskih nizova vodostaja te došao do zaključka kako je proporcionalna k^H uz neku konstantu H . Iznenađujuće otkriće za Hursta je bilo što je u mnogim primjerima konstanta H značajno veća od 0.5. Naime, ako bi vodostaji bile nezavisne normalno distribuirane slučajne varijable, tada bi H morao biti približno 0.5. Činjenica da $R/S(t,k)$ statistika u mnogim primjerima raste brže od $k^{0.5}$ s povećanjem k postala je poznata kao *Hurstov fenomen*. Fenomen ima izravne posljedice na dizajniranje brana s obzirom da za duže vremenske periode veća vrijednost H povlači i veću optimalnu visinu brane.

Hurstovo istraživanje privuklo je dosta pozornosti. Tako je primjerice i istaknuti hrvatsko-američki matematičar William Feller u radu [4] pokazao da za svaki niz nezavisnih slučajnih varijabli konačne varijance, $R/S(t,k)$ statistika raste kao $k^{0.5}$. Ubrzo je postalo jasno da je za neuobičajeno brzi rast odgovorna jaka zavisnost među opaženim vrijednostima. Međutim, pokazalo se da se Hurstov fenomen ne može replicirati ni s jednim tada poznatim slučajnim procesom. Tako je Hurstov fenomen ostao neobjašnjen sve do ranih 60-ih godina prošlog stoljeća kada je slavni francuski matematičar Benoit Mandelbrot predložio moguće objašnjenje problema i potkrijepio ga odgovarajućim modelima. Više detalja o povijesti svojstva dugoročne zavisnosti može se u vidjeti u radu [5] i tamo navedenim referencama.

Od početaka razvoja teorije do danas objavljeno je nebrojeno mnogo znanstvenih radova na temu dugoročne zavisnosti i primjena. Razvijeni su brojni modeli te statističke metode detektiranja dugoročne zavisnosti u podacima. Pregled tema može se vidjeti primjerice u monografijama [1] i [8]. U nastavku rada u odjeljku 2 navodimo modernu definiciju dugoročne zavisnosti te u odjeljku 3 primjere slučajnih procesa sa svojstvom dugoročne zavisnosti. U odjeljku 4 prikazane su metode kojima se dugoročna zavisnost može prepoznati u podacima. Na kraju, u odjeljku 5 dan je primjer stvar-



William Feller
(1906.—1970.)
hrvatsko-američki
matematičar

nih podataka koji pokazuju svojstvo dugoročne zavisnosti.

2 Dugoročna zavisnost

Današnje poimanje pojma dugoročne zavisnosti uglavnom se temelji na specifičnom ponašanju autokorelacijske funkcije slučajnog procesa. Za početak, pretpostavimo da imamo vremenski niz podataka x_1, \dots, x_n zabilježen u uzastopnim vremenskim trenucima. U analizi vremenskih nizova, svrha nam je slučajnim procesom $\{X_t, t \in T\}$ modelirati stohastički mehanizam koji dovodi do realizacija koje čine vremenski niz. Pri tome je važno što vjernije replicirati strukturu zavisnosti vremenskog niza.

Strukturu zavisnosti slučajnog procesa možemo promatrati kroz *funkciju autokorelacija*:

$$\rho(t, s) = \text{Corr}(X_t, X_s) = \frac{\gamma(t, s)}{\sqrt{\gamma(t, t)\gamma(s, s)}}, \quad t, s \in T,$$

gdje je $\gamma(t, s) = \text{Cov}(X_t, X_s)$ *funkcija autokovarijanci* procesa. *Korelacija* između X_t i X_s , $\rho(t, s)$, uvijek je u intervalu $[-1, 1]$ i pokazuje koliko dobro se veza između X_t i X_s može opisati linearnom funkcijom. Tako vrijednost blizu 1 pokazuje da postoji jaka rastuća linearna veza (X_t raste povećanjem X_s), dok vrijednost blizu -1 pokazuje da postoji jaka padajuća linearna veza (X_t pada povećanjem X_s). Ako je $\rho(t, s) = 0$, X_t i X_s su nekorelirane.

U analizi vremenskih nizova posebno su važni stacionarni procesi kod kojih se struktura zavisnosti ne mijenja kroz vrijeme. Proces $\{X_t, t \in \mathbb{Z}\}$ je *stacionaran* ako ima konačnu varijancu, očekivanje se ne mijenja u vremenu $EX_t = EX_s$ za sve $t, s \in \mathbb{Z}$ i korelacija vrijednosti procesa u dva vremenska trenutka t i s ovisi samo o udaljenosti tih trenutaka $t - s$. Tako je za stacionarne procese funkcija autokorelacija funkcija jedne varijable

$$\rho(h) = \rho(0, h) = \frac{\gamma(h)}{\gamma(0)}, \quad h \in \mathbb{Z}.$$

Ovakva definicija stacionarnosti poznata je i kao slaba stacionarnost. Za detalje vidjeti primjerice [2], [3].

Pri modeliranju vremenskog niza stacionarnim procesom za pretpostaviti je da je korelacija između vrijednosti u dva vremenska trenutka manja što su ta dva trenutka udaljenija tako da $\rho(h) \rightarrow 0$ kad $h \rightarrow \infty$. Za uobičajene modele vremenskih nizova korelacije na udaljenim koracima postaju dovoljno male tako da je zbroj korelacija na svim vremenskim pomacima konačan:

$$\sum_{h=-\infty}^{\infty} |\rho(h)| < \infty. \quad (2)$$

Procesi za koje to ne bi vrijedilo nisu razmatrani sve do kraja 60-ih godina prošlog stoljeća kada ih je Benoit Mandelbrot predložio kao moguće objašnjenje Hurstovog fenomena. Danas se za takve procese obično koristi sljedeća definicija ([1, str. 42])).

Definicija 2.1. Neka je $\{X_t\}$ stacionaran proces za koji postoje realni brojevi $\alpha \in (0, 1)$ i $c_\rho > 0$ takvi da vrijedi

$$\lim_{h \rightarrow \infty} \frac{\rho(h)}{c_\rho h^{-\alpha}} = 1. \quad (3)$$

Tada $\{X_t\}$ zovemo stacionaran proces s *dugoročnim pamćenjem* ili *dugoročnom zavisnošću*¹.

Slučajni procesi s dugoročnom zavisnošću su stacionarni, no njihova zavisnost među trenutcima opada neuobičajeno sporo kako se ti trenutci udaljavaju i to toliko sporo da funkcija autokorelacija nije apsolutno sumabilna, odnosno vrijedi:

$$\sum_{h=-\infty}^{\infty} |\rho(h)| = \infty.$$

Primijetimo da ako je $\alpha > 1$ u (3), onda vrijedi (2). Takve procese nazivamo procesi s *kratkoročnim pamćenjem* ili *kratkoročnom zavisnošću*².

Parametar α u (3) obično predstavljamo u obliku $\alpha = 2 - 2H$ i H nazivamo *Hurstov parametar*. Pri tome dugoročna zavisnost nastupa kada je $1/2 < H < 1$.

3 Modeli

Hurstov fenomen opisan u uvodu privukao je pozornost brojnih matematičara. Veliki korak naprijed učinio je Benoit Mandelbrot koji je u nizu radova s koautorima ponudio objašnjenje fenomena. Mandelbrot i John van Ness su u radu [7] razmatrali *frakcionalno Brownovo gibanje* čiji prirasti daju prvi primjer procesa s dugoročnom zavisnošću. Iako se ovaj proces već pojavio u radu Andrey Nikolaeoviča Kolmogorova početkom 40-ih godina, tek je Mandelbrot prepoznao njegovu važnost, dao mu ime i istražio njegova svojstva ([9], [10]).



Benoit Mandelbrot
(1924.–2010.)
francusko-američki
matematičar

¹engl. *long memory, long-range dependence*

²engl. *short memory, short-range dependence*

Definicija 3.1. Za $0 < H < 1$, slučajan proces $B^H = \{B_t^H, t \geq 0\}$ koji je Gausovski (sve konačnodimenzionalne distribucije su Gaussove – normalne), ima očekivanje 0 i funkciju autokovarijanci

$$\text{Cov}(B_t^H, B_s^H) = \frac{1}{2}\sigma^2 \left[t^{2H} - (t-s)^{2H} + s^{2H} \right],$$

zovemo *frakcionalno Brownovo gibanje*.

Za $H = 1/2$, $\{B^{1/2}(t)\}$ je standardno Brownovo gibanje. Promotrimo za $i \in \mathbb{N}$ priraste frakcionalnog Brownovog gibanja

$$X_i = B^H(i) - B^H(i-1). \quad (4)$$

Može se pokazati ([9, str. 332]) da je $\{X_i, i \in \mathbb{N}\}$ stacionaran proces s funkcijom autokorelacija

$$\gamma(h) = \frac{1}{2}\sigma^2 \left(|h+1|^{2H} - 2|h|^{2H} + |h-1|^{2H} \right), \quad h \in \mathbb{Z}. \quad (5)$$

Ako je $H = 1/2$, onda je $\rho(h) = 0$ za $h \neq 0$ pa su prirasti nekorelirani. Međutim, za $1/2 < H < 1$ vrijedi ([1])

$$\lim_{h \rightarrow \infty} \frac{\rho(h)}{H(2H-1)h^{2H-2}} = 1,$$

pa je $\{X_i\}$ proces s dugoročnom zavisnošću s parametrom $\alpha = 2H - 2$. Proces $\{X_i, i \in \mathbb{N}\}$ često se naziva *frakcionalni Gausovski šum*. Zavisnost je jača što je parametar H bliže 1. Na slici 1 prikazane su simulacije trajektorija frakcionalnog Brownovog gibanja za neke vrijednosti H .

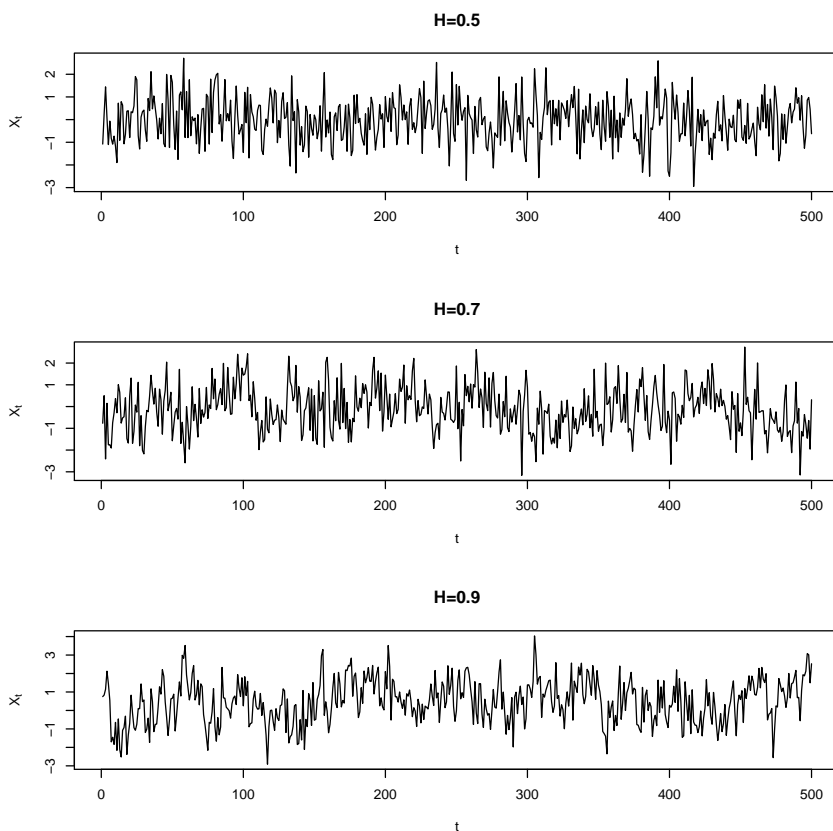
Osim frakcionalnog Gaussovskog šuma, danas postoje i brojni drugi modeli koji imaju svojstvo dugoročne zavisnosti (vidjeti [1], [2]). Među njima posebno su popularni frakcionalni ARIMA procesi. Oni su prirodno proširenje klasičnih ARIMA procesa i omogućavaju kombiniranje kratkoročne i dugoročne zavisnosti.

U analizi vremenskih nizova važno mjesto zauzimaju tzv. ARMA(p, q) procesi – *autoregresivni pomični prosjeci reda* (p, q), $p, q \geq 0$. Proces $\{X_t, t \in \mathbb{Z}\}$ je ARMA(p, q) ako je stacionaran i ako je

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \quad t \in \mathbb{Z}, \quad (6)$$

gdje su $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q \in \mathbb{R}$ koeficijenti i $\{Z_t\}$ bijeli šum, odnosno niz nekoreliranih slučajnih varijabli s očekivanjem nula. Osnovna ideja iza definicije (6) je prikazati proces kao linearnu kombinaciju prošlih vrijednosti procesa i sadašnje i prošlih vrijednosti nekog šuma.

DUGOROČNA ZAVISNOST



Slika 1: Simulirana trajektorija frakcionalnog Gaussovskog šuma za $H = 0.5$, $H = 0.7$ i $H = 0.9$.

Ako s B označimo operator pomaka unazad koji vraća vrijeme procesa jedan trenutak u prošlost, $BX_t = X_{t-1}$, onda (6) možemo zapisati i kao

$$\phi(B)X_t = \theta(B)Z_t,$$

gdje su ϕ i θ polinomi:

$$\begin{aligned}\phi(B) &= 1 - \phi_1 B - \dots - \phi_p B^p, \\ \theta(B) &= 1 + \theta_1 B + \dots + \theta_q B^q.\end{aligned}$$

ARMA procesi mogu imati samo kratkoročnu zavisnost. Poopćenje ARMA

procesa koje dopušta i dugoročnu zavisnost dano je frakcionalnim ARIMA procesima.

Definicija 3.2. Neka je $\{X_t, t \in \mathbb{Z}\}$ stacionaran proces takav da vrijedi

$$\phi(B)(1 - B)^d X_t = \theta(B)Z_t,$$

za $-\frac{1}{2} < d < \frac{1}{2}$, bijeli šum $\{Z_t\}$ i polinome ϕ, θ stupnjeva p, q redom. Tada $\{X_t\}$ zovemo *frakcionalni ARIMA proces reda* (p, d, q) ili FARIMA (p, d, q) .

Razlika u odnosu na ARMA procese dolazi od toga što jednadžba (6) vrijedi za $(1 - B)^d X_t$ umjesto za X_t . Izraz $(1 - B)^d$ predstavlja operator (frakcionalnog) diferenciranja i definiran je s

$$(1 - B)^d = \sum_{k=0}^{\infty} \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)} (-1)^k B^k,$$

gdje je $\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx$. Za $d = 1$ imamo $(1 - B)^d X_t = (1 - B)X_t = X_t - X_{t-1}$ što daje razlike procesa u uzastopnim trenucima. Frakcionalno diferenciranje može dovesti do dugoročne zavisnosti što pokazuje sljedeća propozicija ([1]).

Propozicija 3.1. Ako je $\{X_t\}$ FARIMA (p, d, q) proces i $0 < d < 1/2$, onda je

$$\lim_{h \rightarrow \infty} \frac{\rho(h)}{c_\rho h^{-(1-2d)}} = 1,$$

pa $\{X_t\}$ ima svojstvo dugoročne zavisnosti s parametrom $\alpha = 2d - 1$.

Hurstov parametar FARIMA procesa slijedi iz jednakosti $1 - 2d = 2 - 2H$ što povlači da je $H = d + 1/2$. Više detalja o FARIMA procesima može se vidjeti u [1], [8].

4 Metode detekcije i procjene

Sad kada je postavljen teorijski okvir dugoročne zavisnosti i pripadni modeli, postavlja se pitanje kako uopće na osnovu podataka znati kada su takvi modeli potrebni. Ovdje ćemo spomenuti nekoliko metoda kojima se može naslutiti dugoročna zavisnost u vremenskom nizu.

4.1 Korelogram

S obzirom da je dugoročna zavisnost definirana kao svojstvo funkcije autokorelacija, logičan pristup je prvo procijeniti funkciju autokorelacija na osnovu podataka.

Za niz podataka x_1, x_2, \dots, x_n funkciju autokovarijanci procjenjujemo s

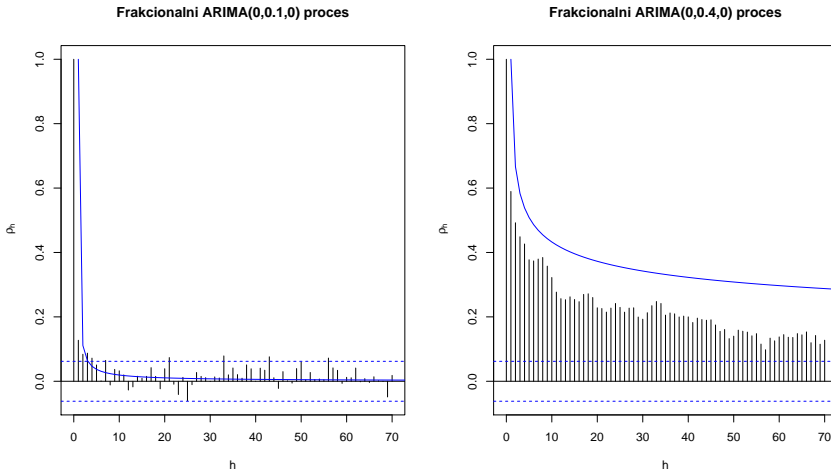
$$\hat{\gamma}(h) = \frac{1}{n} \sum_{j=1}^{n-h} (x_{j+h} - \bar{x}_n)(x_j - \bar{x}_n), \quad \text{za } 0 \leq h < n,$$

i $\hat{\gamma}(h) = \hat{\gamma}(-h)$, za $-n < h < 0$, gdje je $\bar{x}_n = \frac{1}{n} \sum_{j=1}^n x_j$. Funkciju autokorelacija procjenjujemo s

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)},$$

za $-n < h < n$. Graf parova $(h, \hat{\rho}(h))$ za $h \geq 0$ naziva se *korelogram*.

Dugoročnu zavisnost karakterizira neuobičajeno sporo opadanje autokorelacijske funkcije u nulu jer je ona proporcionalna s h^{2H-2} za neki $\frac{1}{2} < H < 1$. To može značiti da su korelacije male, ali je opadanje u nulu neuobičajeno sporo. Takvu asimptotsku karakteristiku teško je uočiti čak i kada imamo veliki uzorak.



Slika 2: Korelogram simuliranog frakcionalnog ARIMA(0, d , 0) procesa za $d = 0.1$ ($H = 0.6$) i $d = 0.4$ ($H = 0.9$) i njihova teorijska autokorelacijska funkcija (plavo).

Na slici 2 možemo primijetiti kako je kod frakcionalnog ARIMA(0, 0.1, 0) procesa većina vrijednosti procijenjene autokorelacijske funkcije vrlo mala. Za $H = d + \frac{1}{2}$ bliže 1, dugoročna zavisnost je uočljivija što pokazuje korelogram frakcionalnog ARIMA(0, 0.4, 0) procesa.

Prikladniji prikaz korelograma bi bio graf s logaritmiranim osima, tj. graf $\log |\hat{\rho}(h)|$ u odnosu na $\log h$. Ako je $\rho(h) \approx c_\rho h^{2H-2}$ za $h \rightarrow \infty$, onda će za velike h biti

$$\log |\rho(h)| \approx \log c_\rho + (2H - 2) \log h.$$

Na osnovu te relacije, parametar H može se procijeniti kao koeficijent smjera u linearnoj vezi $\log |\hat{\rho}(h)|$ u ovisnosti $\log h$. Takve procjene mogu poprilično odstupati od stvarnih vrijednosti. Primjerice, procjene za podatke sa slike 2 su $\hat{H} = 0.8364$ i $\hat{H} = 0.7002$ za stvarne vrijednosti $H = 0.6$ i $H = 0.9$, redom. Ipak, metoda može biti korisna u uočavanju dugoročne zavisnosti.

4.2 R/S statistika

Najstarija metoda detekcije dugoročne zavisnosti potječe još od Hursta i temelji se na promatranju reskaliranog prilagođenog raspona ili R/S statistike $R/S(t, k)$ definirane u (1) kada $k \rightarrow \infty$. U okviru hidrologije radi se o optimalnoj visini brane za dugo razdoblje u budućnost.

Korištenje R/S statistike temelji se na sljedećoj heuristici koja se u velikoj mjeri može opravdati uz dodatne pretpostavke. Za kratkoročno zavisne nizove u pravilu vrijedi centralni granični teorem te se može pokazati da u tom slučaju $R/S(t, k)$ raste kao $k^{1/2}$ s porastom k . S druge strane, za nizove sa svojstvima dugoročne zavisnosti $R/S(t, k)$ raste kao k^H za $1/2 < H < 1$. Za više detalja vidjeti [8].

Metoda uočavanja dugoročne zavisnosti preko R/S statistike može se sažeti na korake:

1. Izračunaj $R/S(t, k)$ za sve moguće vrijednosti od k i t .
2. Prikaži graf vrijednosti $\log R/S(t, k)$ u odnosu na $\log k$.
3. Procijeni pravac $y = a + b \log k$ metodom najmanjih kvadrata. Tada \hat{b} predstavlja procjenu parametra H , tj. \hat{H} .

Ono što predstavlja problem u ovoj metodi je to što distribucija R/S statistike ovisi o distribuciji podataka pa je upitno je li metoda najmanjih kvadrata primjerena u ovom slučaju. Nadalje, raspoložemo konačnim i vrlo često manjim skupom podataka i problematično je odrediti od kojeg k počinje ono „granično ponašanje”. Unatoč ovim nedostacima, R/S graf je koristan i može nam dati prvu procjenu o dugoročnoj zavisnosti.

4.3 Druge metode

Osim spomenutih metoda postoje i druge metode koje se temelje na različitim svojstvima procesa s dugoročnom zavisnošću. Metoda varijance primjerice temelji se na činjenici da varijanca parcijalnih suma raste različitom brzinom za kratkoročno i dugoročno zavisne procese. Druga klasa metoda temelji se na analizi u spektralnoj domeni i rastu spektralne gustoće u okolini nule.

Procjenitelji metodom maksimalne vjerodostojnosti ili ML procjenitelji su efikasni i asimptotski normalni, stoga se ova metoda često koristi za procjenu Hurstovog parametra H . Metoda maksimalne vjerodostojnosti se oslanja na distribuciju procesa zbog čega se pretpostavlja da je proces Gaussovski. Kako je izračun funkcije vjerodostojnosti za frakcionalne ARIMA procese spor i zahtjevan, koristi se Whittleova aproksimacija funkcije vjerodostojnosti. U tom slučaju moguće je i oslabiti pretpostavku da je proces Gaussovski. Dobiveni procjenitelj poznat je kao Whittleov aproksimativni ML procjenitelj.

Više o ovim metodama može se pronaći u [1, Poglavlje 5, Poglavlje 6] i [8].

5 Primjer – vodostaj rijeke Nil

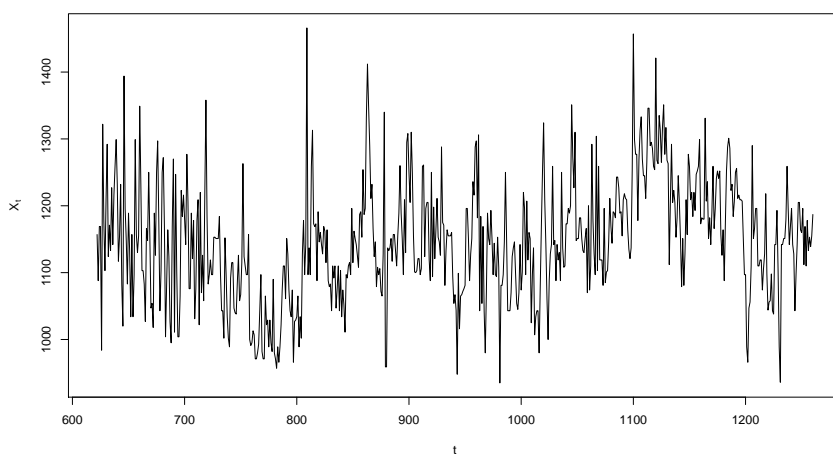
U ovom odjeljku promotrit ćemo vremenski niz podataka o vodostaju rijeke Nil koji je doveo do razvoja dugoročne zavisnosti kako je opisano u uvodu. Vremenski niz $\{x_t\}$ sastoji se od 663 godišnja podatka o minimalnom vodostaju rijeke Nil u razdoblju od 622. do 1281. godine koji su mjereni na mjerilu na otoku Roda pokraj Kaira. Originalni podaci su dani u monografiji [1].

Na slici 3 je prikazan minimalni vodostaj rijeke Nil kroz vrijeme u razdoblju od 622. do 1260. godine. Neke numeričke karakteristike vremenskog niza $\{x_t\}$ prikazane su u Tablici 1.

Minimum	Donji kvartil	Medijan	Prosjek	Gornji kvartil	Maksimum
935	1089	1146	1147	1200	1466

Tablica 1: Numeričke karakteristike niza podataka o godišnjem minimalnom vodostaju rijeke Nil od 622. do 1260. godine.

Na slici 3 možemo uočiti kako se izmjenjuju dugi periodi visokih i niskih vrijednosti što je karakteristika nizova s dugoročnom zavisnošću. Benoit Mandelbrot je, osvrćući se na Bibliju, ovo ponašanje prozvao *Josipov efekt*.



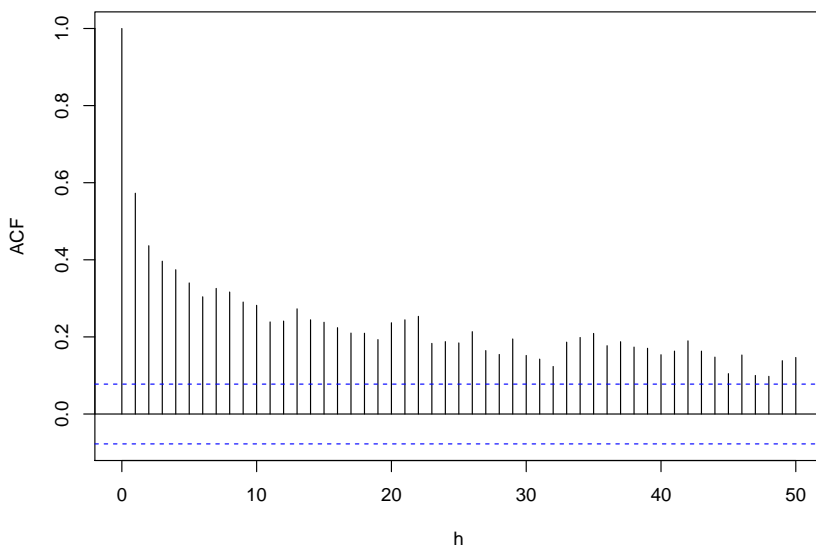
Slika 3: Godišnji minimalni vodostaj rijeke Nil u razdoblju od 622. do 1260. godine.

Naime, u Knjizi Postanka 41, 29–30, Josip je interpretirao faraonov san o sedam lijepih i sedam ružnih krava: „Dolazi, evo, sedam godina velikog obilja svoj zemlji egipatskoj. A poslije njih nastat će sedam gladnih godina, kada će se zaboraviti sve obilje u zemlji egipatskoj.” Također, na slici 3 uočavamo puno ciklusa različitih frekvencija za što ne možemo reći da je posljedica nekakve sezonalnosti.

Na slici 4 prikazana je procijenjena autokorelacijska funkcija promatranog niza. Sporo opadanje uzoračke autokorelacijske funkcije u nulu koje je za ovaj niz podataka očito, sugerira postojanje dugoročne zavisnosti. U klasičnoj analizi vremenskih nizova ovakvo ponašanje se tipično pripisuje nestacionarnosti te se različitim transformacijama pokušava doći do niza koji ima obilježja stacionarnosti. Međutim, modeli s dugoročnom zavisnošću pokazuju da je takvo sporo opadanje procijenjenih autokorelacija moguće opisati u okviru stacionarnih procesa.

Kako bismo provjerili sumnju da je niz $\{x_t\}$ realizacija stacionarnog procesa s dugoročnom zavisnosti, možemo provesti metode za statističku procjenu dugoročne zavisnosti koje su opisane u odjeljku 4.

Na slici 5 prikazani su grafovi tih metoda. Korelogram vremenskog niza u logaritmiranim osima može se vidjeti na slici 5(a). Procjena koeficijenta smjera regresijskog pravca na osnovu podataka $\log |\hat{\rho}(h)|$ u odnosu na $\log h$



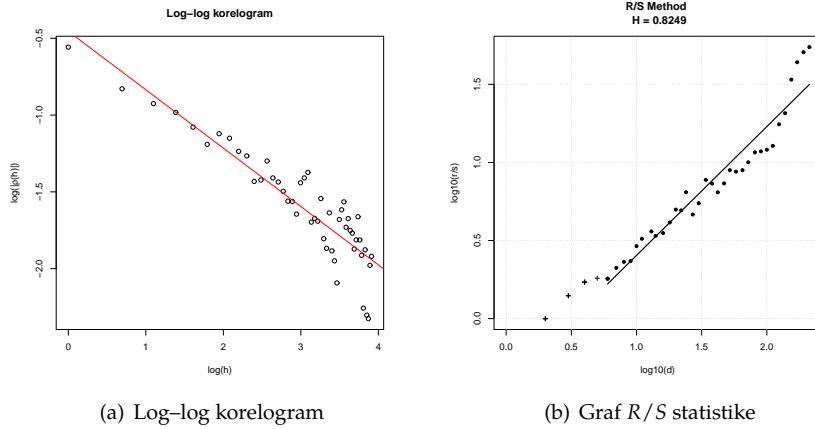
Slika 4: Uzoračka autokorelacijska funkcija vremenskog niza vodostaja rijeke Nil

je $-\hat{\alpha} = -0.3801$, što znači da je po ovoj metodi $\hat{H} = 0.80995$.

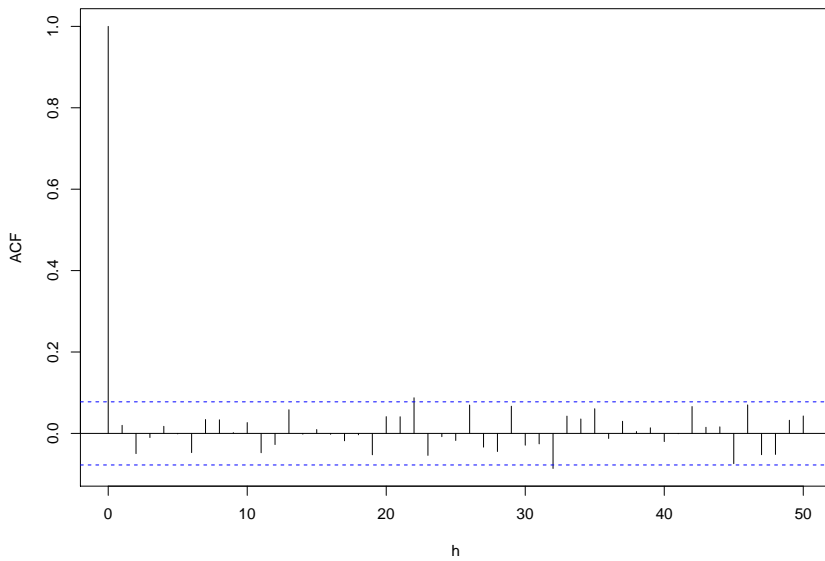
Graf R/S statistike prikazan na slici 5(b) pokazuje kako se vrijednosti R/S statistike gomilaju oko pravca s koeficijentom smjera većim od $\frac{1}{2}$. Procjena koeficijenta smjera pravca koji aproksimira podatke R/S statistike je 0.8248789 što je ujedno i procjena za parametar dugoročne zavisnosti H po ovoj metodi.

Procjena parametra H dobivena pomoću Whittleovog aproksimativnog ML procjenitelja je $\hat{H} = 0.89656833$. Ukoliko bi niz vodostaja modelirali FARIMA modelom, za parametar d u FARIMA(p, d, q) modelu možemo uzeti $d = \hat{H} - \frac{1}{2} = 0.397$. Ako je takav FARIMA model zaista prikladan, onda bi niz koji dobijemo primjenom operatora $(1 - B)^{0.397}$ na vodostaje trebao zadovoljavati jednadžbu (6). Procijenjena autokorelacijska funkcija tog niza prikazana je na slici 6.

Primjećujemo kako su vrijednosti korelacija vrlo male tako da se dobiveni niz može opisati ARMA procesom, pa čak i bijelim šumom koji odgovara ARMA(0, 0) procesu.



Slika 5: Statistički procjenitelji za parametar dugoročne zavisnosti vodostaja rijeke Nil



Slika 6: Uzoračka autokorelacijska funkcija frakcionalno diferenciranog vremenskog niza operatorom $(1 - B)^{0.397}$

Literatura

- [1] J. Beran, *Statistics for Long-Memory Processes*, Chapman & Hall, New York, 1994.
- [2] P.J. BROCKWELL, R. A. DAVIS, *Introduction to Time Series and Forecasting*, Springer-Verlag, New York, 2002.
- [3] P.J. BROCKWELL, R. A. DAVIS, *Time Series: Theory and Methods*, Springer-Verlag, New York, 1991.
- [4] W. Feller, *The asymptotic distribution of the range of sums of independent random variables* The Annals of Mathematical Statistics **22**(3) (1951), 427-432
- [5] T. Graves, R. Gramacy, N. Watkins, C. Franzke, *A brief history of long memory: Hurst, Mandelbrot and the road to ARFIMA, 1951–1980.*, Entropy **19**(9) (2017), 437
- [6] H. E. Hurst, *Long-term storage capacity of reservoirs*, Trans. Amer. Soc. Civil Eng. **116** (1951), 770-799
- [7] B. B. Mandelbrot, J. .W. Van Ness, *Fractional Brownian motions, fractional noises and applications*, SIAM review, **10**(4) (1968), 422-437
- [8] V. Pipiras, V., M. S. Taqqu, *Long-range dependence and self-similarity*, Cambridge University Press, Cambridge, 2017.
- [9] G. Samorodnitsky, M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, CRC Press, New York, 1994.
- [10] M. S. Taqqu, *Benoît Mandelbrot and fractional Brownian motion*, Statistical Science, **28**(1) (2013), 131-134