3rd National Congress on Physical Sciences, 29 Sep. – 2 Oct. 2016, Sofia
Section: Physics of Earth, Atmosphere and Space

Стохастични модели на дневните суми на валежите за станция Златоград

Н. Нейкова, **П. Нейчев**

Национален институт по метеорология и хидрология,
Българска академия на науките,
бул. Цариградско шосе 66, 1784 София

Abstract. A stochastic daily precipitation model for Zlatograd station in south Bulgaria conditional on atmospheric predictors that characterize the atmospheric circulation over Balkans is developed. The model consists of two components describing the occurrence and intensity precipitation series. The intensity component is based on a hybrid between gamma and generalized Pareto distributions. The results of simulations designed to compare the models based on the hybrid distribution and those based on the standard gamma distribution are reported and some potential difficulties are outlined.

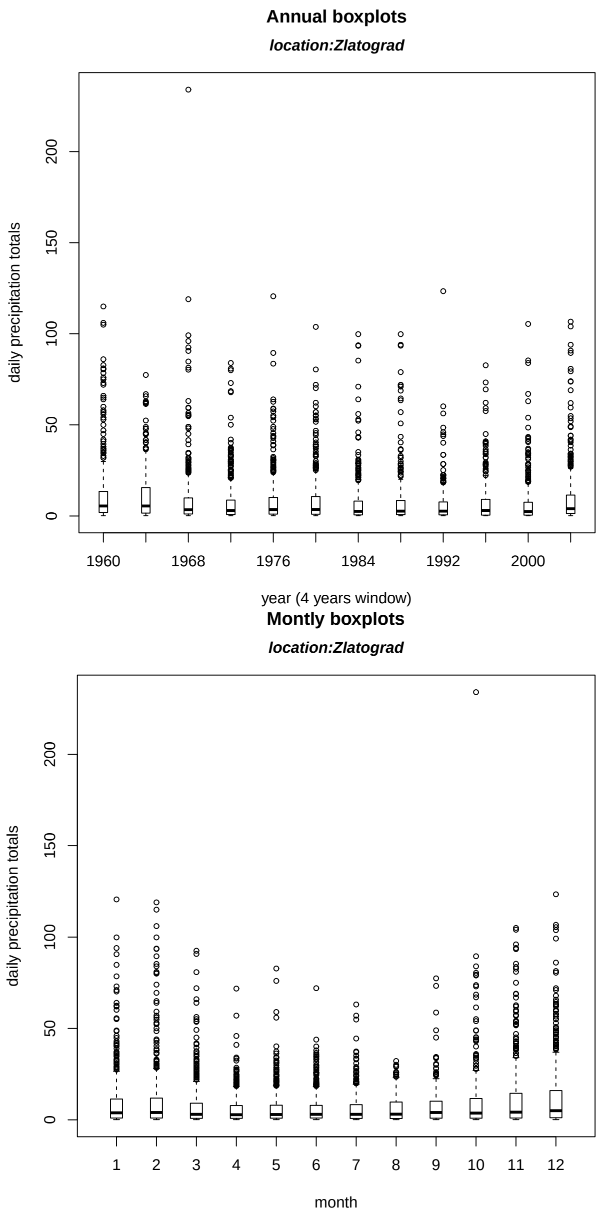
1 Въведение

Създаването на стохастични модели за валежите е актуална задача в социалната и стопанска дейност. Моделите на валежите се състоят от две компоненти, описващи появата на валеж и интензивността на валежа. За моделиране на появата се използва бинарната логистична регресия. За целта данните се превръщат в бинарни, след което се моделира вероятността за поява на валеж. За моделиране на интензивността се използват непрекъснати дясно скосени разпределения, например гама и Вейбул или логнормално разпределение. За извършване на пресмятанията може да бъде използван стандартен софтуер за обобщени линейни модели (GLMs). Основен недостатък на тези разпределения е, че чрез тях не могат да бъдат описани добре екстремалните стойности на валежите. За да избегнем този недостатък ние адаптираме подхода на [1], базиран на хибридни разпределения. По-точно ние използваме хибридно разпределение между Гама и обобщеното разпределение на Парето (GP) и хибридно разпределение между Вейбул и GP, за да направим стохастични дневни модели на валежите за станция Златоград. С помощта на хибридните разпределения става възможно симулирането (генерирането) на дневни суми на валежите, разпределението на които наподобява разпределението на реалните данни. Целта на симулирането на валежни данни е да увеличи обема на извадката от налични данни и върху тях да се проведе статистически анализ -– параметричен боотстрап (Б. Ефрон). Симулираните моделни редици от данни могат да бъдат използвани за оценка на риска от наводнения, засушавания както и за борба с ерозията на почвите и други. В тази статия е представено прилагането и адаптирането на подхода на [2], базиран на хибридни разпределения, и е предложен подобрен модел на дневните валежи с по-тежка опашка. Моделът описва редица от данни за валежа в град Златоград, България, като за предиктори включва атмосферни величини, характеризиращи поведението на атмосферната циркулация над Балканския полуостров.

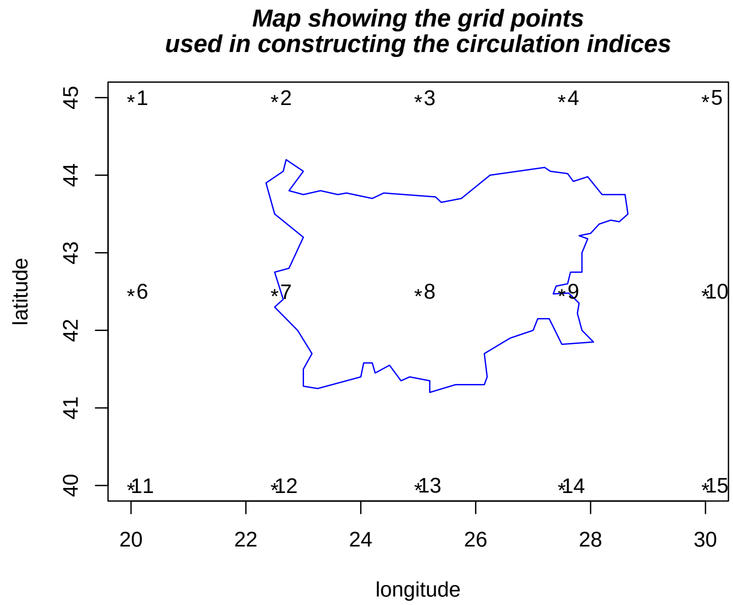
2 Описание на данните

Използвани са дневните суми на валежите за периода 1.01.1960 – 31.12.2007 год. за станция Златоград. Валежите се измерват веднъж в денонощието в 8:00 часа местно време, 6:00 GMT универсално време. Под валеж в най-общ смисъл се разбира падналата от облаците вода в течно или твърдо състояние. Валежите от тази станция представляват специален интерес, тъй като на 3.10.1970 год. е бил регистриран рекорден валеж от 234 литра за едно денонощие. В станция Златоград са регистрирани 5402 дни с валеж от общо 17474 дни. На Фиг. 1 е представено разпределението на дневните суми на валежите по години(вляво) и по месеци(вдясно). При създаването на модела на дневните суми на валежа са използвани различни атмосферни променливи (индекси), характеризиращи атмосферната циркулация над Балканския полуостров, формирани на различни геопотенциални височини във възлите дадени на Фиг. 2. Използвани са също така северно-атлантическата и арктическата осцилации NAO и AO, характеризиращи поведението на атмосферната циркулация над северното полукълбо. Тъй като дневната сума на валежа се измерва в 8:00 часа местно време, за моделиране на вероятността за появата и интензивността на валежа като предиктори са използвани стойностите на атмосферните индекси от предишния ден както и лагове от по-висок ред.

За NAO и AO бяха включени лагове до 5 дни. Включени са и следните индекси, характеризиращи атмосферната циркулация над Балканския полуостров: gama (gama850, gama700, gama300 и gama300) е влагоадиабатен градиент, стойността на който се променя в зависимост от налягането и температурата на съответното ниво 850 hPa и 700 hPa, 500 hPa или 300 hPa; adv.u.s е адвекция на влага в слоя между нива 850 hPa и 500 hPa по направление запад-изток по стойнос-



Фиг. 1: Разпределение на дневните суми на валежите по години(вляво) и по месеци(вдясно). Екстремалната стойност от 234 mm е измерена на 03.10.1970 г.



Фиг. 2: Разположение на възлите за атмосферната циркулация над територията на България.

тите на зоналната компонента на вятъра във възел 8 и стойностите на специфичната влажност във възли 7 и 9 на трите нива (850, 700 и 500 hPa); adv.v.s.850, adv.v.s.700, adv.v.s.500 – влажностната адвекция в направление север-юг изчислена по специфична влажност, характеризира нивото на облачността в средните слоеве на атмосферата, свързана е с обложни валежи; adv.u.t.300 и adv.v.t.300 са температурните адвекции по направление запад-изток и север-юг на ниво 300 hPa; adv.u.t.sig и adv.u.t.sig са зонални (север-юг и запад-изток) термични адвекции в близост до земята; prwtrv е валежната вода в атмосферен стълб над възела; prwtr е промяна на влагосъдържанието поради или движение на въздуха и/или изваляване; shum е абсолютната влажност; rhum е относителна влажност.

3 Моделиране на дневните суми на валежите

3.1 Стохастичен модел на валежа

Нека Y_t е сумата на валежа в деня t, $\mathbf{Z}_t = (Z_{1t}, \ldots, Z_{kt})'$ е вектор от атмосферни променливи, свързан с Y_t за $t = 1, \ldots, T$. Денят t е сух (без валеж) ако $Y_t < c$, където c е предварително зададена константа, чиято стандартна стойност е $c = 0.1$ mm, и мокър (с валеж) ако

$Y_t \geq c$. Редицата от сухи и мокри дни е представена с помощта на индикаторната функция $I_t = I_{[y_t \geq c]}$, която приема стойност 0, ако денят t е сух, и 1, ако денят t е мокър. Наблюдаваните стойности на Y_t и $\mathbf{Z}_t$ и I_t ще означаваме с малки букви. Да означим с $p_t(\mathbf{x}_t)$ вероятността денят t да е с валеж при наблюдавани стойности на вектора $\mathbf{x}_t = (i_{t-1}, \ldots, i_{t-m}, y_{t-1}, \ldots, y_{t-m}, z_{1t}, \ldots, z_{kt})'$, в който са включени лагове от ред m на индикаторната променлива и дневната сума на валежа. Интензивност на валежа за деня t се дефинира като $R_t = Y_t$, ако $Y_t \geq c$, и $R_t =$ missing в противен случай. Означаваме с $q(r_t|\mathbf{x}_t)$ условната плътност на разпределение, което е от тип дясно скосено разпределение, понеже преобладават валежите с малка интензивност.

Редицата от дневни суми на валежите се моделира като смес от две разпределения. Едното разпределение е дискретно с маса в нулата, отчитащо дните без валеж, докато второто разпределение е непрекъснато дясно скосено, отчитащо интензивността на дните с валеж, [3]. Понеже двете състояния са взаимно изключващи се (несъвместими), то съответната преходна вероятност е:

$$\begin{aligned} f_t(y_t|\mathbf{x}_t) &= (1 - p_t(\mathbf{x}_t))\, I_{[y_t<c]} + p_t(\mathbf{x}_t) q_t(r_t|\mathbf{x}_t) I_{[y_t \geq c]} \\ &= (1 - p_t(\mathbf{x}_t)) \left(1 - I_{[y_t \geq c]}\right) + p_t(\mathbf{x}_t) q_t(r_t|\mathbf{x}_t) I_{[y_t \geq c]} \\ &= (1 - p_t(\mathbf{x}_t))^{\left(1 - I_{[y_t \geq c]}\right)} \left(p_t(\mathbf{x}_t) q_t(r_t|\mathbf{x}_t)\right)^{I_{[y_t \geq c]}}. \end{aligned}$$

Най-често използваните разпределения в практиката за $q_t(r_t|\mathbf{x}_t)$ са гама, Вейбул, лог-нормално. Ако се интересуваме от модели на екстремалните дневни валежни суми, тогава се използват обобщеното разпределение на екстремалните стойности (GEV) или обобщеното разпределение на Парето (GP).

При условие, че $p_t(\mathbf{x}_t)$ и $q_t(r_t|\mathbf{x}_t)$ нямат общи параметри, функцията на правдоподобие за $(y_{t-m-1}, \ldots, y_T)$ се дефинира, както следва:

$$\begin{aligned} L &= \prod_{t=m+1}^{T} f_t(y_t|\mathbf{x}_t) = \prod_{t=m+1}^{T} (1 - p_t(\mathbf{x}_t))^{\left(1 - I_{[y_t \geq c]}\right)} \left(p_t(\mathbf{x}_t) q_t(r_t|\mathbf{x}_t)\right)^{I_{[y_t \geq c]}} \\ &= \prod_{t=m+1}^{T} (1 - p_t(\mathbf{x}_t))^{\left(1 - I_{[y_t \geq c]}\right)} \left(p_t(\mathbf{x}_t)\right)^{I_{[y_t \geq c]}} \prod_{t=m+1, y_t > c} q_t(r_t|\mathbf{x}_t). \end{aligned}$$

От тази факторизация на функцията на правдоподобие се вижда, че за оценяването на неизвестните параметри могат да бъдат използвани стандартни програмни процедури за обобщени линейни модели като *glm*, *vglm* от *VGAM* и други от програмната среда R. Това е така, понеже първата компонента представлява функцията на правдоподобие на бинарен времеви ред, докато втората компонента е функцията на правдоподобие на интензивността на валежа с някое от споменатите по-горе разпределения. За моделиране на вероятността за поява на валеж $p_t(\mathbf{x}_t)$ се използва логистичната регресия както в [1] и [3]. В тази статия вниманието ще бъде насочено към моделиране на интензивността на валежа.

3.2 Разпределения на интензивността

За моделиране на разпределението на интензивността на валежа са използвани гама и GP разпределенията. Поради различните параметризации на тези разпределения, в тази част са дадени техните плътности, които са използвани в библиотеката VGAM, за да бъдат избегнати недоразумения с други алтернативни представяния в литературата. Плътност на гама разпределението:

$$f(x,a,b) = \begin{cases} \frac{b^a \ r^{a-1} \exp(-br)}{\Gamma(a)} & r > 0 \\ 0 & r = 0, \end{cases}$$

където $\Gamma(a)$ е гама функцията, $a > 0$ е параметър на формата (shape), $b > 0$ е параметърът, свързан с мащаба (скалата, rate). При тази параметризация очакването и дисперсията са съответно равни на $\mu = a/b$ и $\sigma^2 = /b^2$.

Плътност на обобщеното разпределение на Парето (GP) с праг u:

$$g(r,u,\sigma,\xi) = \frac{1}{\sigma}\left[1 + \frac{\xi(r-u)}{\sigma}\right]_+^{-\frac{1}{\xi}-1}$$

където $r > u$, u, $\sigma > 0$ и ξ са параметрите на положението, мащаба и формата, $[A]_+ = \max(A,0)$. Параметърът ξ характеризира типа на GP разпределението, както следва: с тежка опашка, ако $\xi > 0$; с крайна опашка ,ако $\xi < 0$; и експоненциално (изместено с u) разпределение, ако $\xi = 0$.

При моделиране на данни с обобщени линейни модели (GLMs) и разпределения на екстремалните стойности (EVD) се използват следните свързващи функции, чрез които се задава връзката между предикторните (регресионните) променливи и параметрите на гама и GP разпределенията

$$\log(a) = u(\mathbf{x}_t) = \theta_0 + \sum_{l=1}^{p} g_l(y_{t-l}) + \sum_{l=1}^{k} g_{p+l}(\mathbf{z}_{lt}) + g_{p+k+1}(t).$$

В линейния предиктор $u(\mathbf{x}_t)$ се включват лагови предиктори на интензивността y_{t-l} на валежа, краен ред на Фурие, за да бъде отчетена сезонността на валежите, както и атмосферни предиктори, където g_l за $l = 1, \ldots, p + k + 1$ са непрекъснати функции. Пример за подобен модел:

$$\log(a) = \theta_0 + \theta_1 \mathrm{y}_{t-1} + \theta_2 \mathrm{C}_t + \theta_3 \mathrm{S}_t + \theta_4 \mathrm{NAO}_{t-1}$$

където $\mathrm{C}_t = \cos(2\pi t/365.25)$ и $S_t = \sin(2\pi t/365.25)$, а NAO е северно-атлантическата осцилация. Предикторите в този модел са:
$\mathbf{x}_t = (1, \mathrm{y}_{t-1}, \mathrm{C}_t, \mathrm{S}_t, \mathrm{NAO}_{t-1})$, а θ_i са неизвестни параметри.

$$\log(a) = \theta_1^T \mathbf{x}_{1t},\ \log(b) = \theta_2^T \mathbf{x}_{2t},\ \log(\sigma) = \theta_3^T \mathbf{x}_{3t},\ \xi = \theta_4^T \mathbf{x}_{4t}.$$

Във функциите θ_i са неизвестните векторни параметри, $\mathbf{x}_{it}$ са образувани от вектора на предикторите $\mathbf{x}_t$ за $i = 1, \ldots, 4$. Използването на логаритъма като свързаваща (log-link) функция осигурява положителност на параметрите a, b и σ. Определянето на оценките на неизвестните параметри θ_i се получава чрез максимизиране на функцията на правдоподобие.

4 Хибридни разпределения

Хибридното разпределение (гама-GP) на основата на гама и GP разпределенията е дефинирано в [2], както следва:

$$h(x) = \begin{cases} f(x) & x \le u \\ (1 - F(u))g(x) & x > u \end{cases}$$

където $F(x)$ е функцията на гама разпределението, $f(x)$ и $g(x)$ са плътностите на гама и GP разпределенията, а $(1 - F(u))$ е нормиращ множител.

За да осигурят непрекъснатост в праговата стойност u тези автори налагат условието

$$f(u) = [1 - F(u)]g(u) = [1 - F(u)]/\sigma.$$

В резултат на това за мащабния параметър на GP разпределението получаваме

$$\sigma = (1 - F(u))/f(u).$$

Това означава, че параметърът на мащаба на GP разпределението се дефинира изцяло в термините на гама разпределението, с което успешно могат да бъдат моделирани наблюденията с по-малки стойности до прага u. Авторите предлагат следната процедура за оценяване на параметрите на хибридното разпределение: (i) моделиране на интензивността на валежа със стандартна процедура за обобщен линеен модел с гама разпределение (свързваща функция – log link function) на неизвестните параметри с предикторните променливи по пълната извадка от данни; (ii) моделиране на интензивностите на валежа, превишаващи прага u, със стандартна процедура за анализ на екстремалните стойности, основана на GP разпределението, като за целта бъдат използвани подходящи свързаващи функции; (iii) заместване на параметъра на скалата на GP разпределението с обратно пропорционалната стойност на хазартната функция на гама разпределнието в стойността на прага u. Предложената процедура за дефиниране на хибридни разпределения е универсална, тъй като не зависи от вида на разпределението $F(x)$. В [4] се разглеждат модели на дневните суми на валежите за станция Ихтиман с гама, Вейбул и хибридни разпределения между гама и GP, и Вейбул и GP, в които модели NAO е използвана като предиктор.

5 Резултати

5.1 Модел за интензивността на валежите

За моделиране на интензивността на дневните валежи сме използвали гама разпределението с линеен предиктор, в който са включени северно-атлантическата и арктически осцилации NAO и AO и PNA. При създаването на модел за интензивността на дневния валеж са използвани стойностите на всички атмосферни индекси от предходния ден, както и някои взаимодействия с лага на интензивността от предишния ден, докато за NAO, AO и PNA са използвани лагове до 5 дни. За модела на интензивността на валежа са използвани 5406 дни, в които е регистриран валеж в станция Златоград. Голям брой от включените атмосферни индекси се оказаха статистически незначими предиктори, според критерия на Стюдент при стандартните нива на съгласие. Това се потвърждава и от резултатите при проверката на хипотези за значимост на оценките на параметрите пред съответните атмосферни индекси с теста, основан на отношението на правдоподобие. Тези резултати са дадени в Табл. 1 на отклоненията (Deviance table) от изхода на glm процедурата. Тази таблица се редуцира до ANOVA таблица за класическия модел на множествена линейна регресия с Гаусово разпределение на грешката. Моделът е с изразен сезонен характер, а оценките на неизвестните параметри пред някои от атмосферните индекси, за които е известно че формират интензивността на валежа, са статистически значими. Проведен бе допълнителен анализ чрез стъпковата процедура stepAIC от библиотеката MASS за избор на статистически значими предиктори с помощта на информационния критерий на Бейс (BIC). В резул-

Табл. 1: ANOVA таблица на модел на интензивността на дневните валежи – редуциран модел след използването на stepAIC процедурата

Terms	Df	Dev.Resid.	Df	Resid.Dev	Pr(¿Chi)	Signif.
NULL			5380	11555.9		
sin	1	56.48	5379	11499.5	2.285e-08	***
cos	1	353.82	5378	11145.7	¡ 2.2e-16	***
sin2	1	1.98	5377	11143.7	0.2957719	
cos2	1	5.51	5376	11138.2	0.0809456	.
sin3	1	3.06	5375	11135.1	0.1934717	
cos3	1	16.03	5374	11119.1	0.0029043	**
y_{t-1}	1	172.67	5373	10946.4	¡ 2.2e-16	***
AO_{t-1}	1	44.50	5372	10901.9	7.010e-07	***
NAO_{t-2}	1	17.53	5371	10884.4	0.0018493	**
prwtrv_{t-1}	1	40.19	5370	10844.2	2.421e-06	***
prwtru_{t-1}	1	83.30	5369	10760.9	1.139e-11	***
gama700_{t-1}	1	184.88	5368	10576.0	¡ 2.2e-16	***
gama300_{t-1}	1	373.55	5367	10202.4	¡ 2.2e-16	***
shumx850_{t-1}	1	366.41	5366	9836.0	¡ 2.2e-16	***
adv.u.s.700_{t-1}	1	32.71	5365	9803.3	2.105e-05	***
adv.u.t.sig_{t-1}	1	197.87	5364	9605.5	¡ 2.2e-16	***
adv.v.s.850_{t-1}	1	186.72	5363	9418.7	¡ 2.2e-16	***
adv.v.s.700_{t-1}	1	5.39	5362	9413.3	0.0841121	.
adv.v.t.300_{t-1}	1	216.33	5361	9197.0	¡ 2.2e-16	***
adv.v.s.10m_{t-1}	1	88.49	5360	9108.5	2.637e-12	***
adv.v.t.sig_{t-1}	1	43.46	5359	9065.1	9.442e-07	***
slpx.2m_{t-1}	1	107.18	5358	8957.9	1.370e-14	***
rhumv_{t-1}	1	32.40	5357	8925.5	2.304e-05	***
pratev_{t-1}	1	21.50	5356	8904.0	0.0005642	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

тат на това се достигна до редуциран модел на интензивността, включващ най-значимите предиктори като сезонен тренд от три хармоники $\sum_{k=1}^{3}[\sin(2\pi tk/365.25)+\cos(2\pi tk/365.25)]$, интензивността на валежа от предишния ден y_{t-1} и атмосферните индекси $AO_{t-1}, NAO_{t-2}, prwtrv_{t-1}, prwtru_{t-1}, gama700_{t-1}, gama300_{t-1}$, $shumx850_{t-1}$, $adv.u.s.700_{t-1}, adv.u.t.sig_{t-1}, adv.v.s.850_{t-1}, pratev_{t-1}$, $adv.v.s.700_{t-1}, adv.v.t.300_{t-1}, adv.v.s.10m_{t-1}, adv.v.t.sig_{t-1}, slpx.2m_{t-1}$, $rhumv_{t-1}$. Ще отбележим, че стойността на функцията на отклоненията без предикторни променливи, т.е., само модел със свободен параметър a_0, е 11555.9 при 5380 степени на свобода. Тази стойност се редуцира до 8904.0 при 5356 степени на свобода след използването на сезонните компоненти, интензивността на валежа от предишния ден, 17те атмосферни индекса и свободният член като предиктори в модела. Тези атмосферни индекси са статистически значими предиктори, понеже съответните им параметри са статистически значими. Така например, влажността в атмосферата $shumx850_{t-1}$ и интензивността на валежа от предишния ден y_{t-1} редуцират стойността на функцията на отклоненията, съответно с 366.71 и 172.67. Линейната комбинция от sin и cos, характеризираща сезонното поведение на появата на валеж, е статистически значима, тъй като поне една от нейните компоненти е значима. Резултатът от процедурата stepAIC е еквивалентен на определянето на оптимален модел сред множество от алтернативни модели по обучаваща и валидираща извадки от данните. В заключение можем кажем, че разгледаният модел за интензивността, характеризира много добре историческите данни.

5.2 Поверка на хипотезата за тежка опашка на разпределението на екстремалните валежа

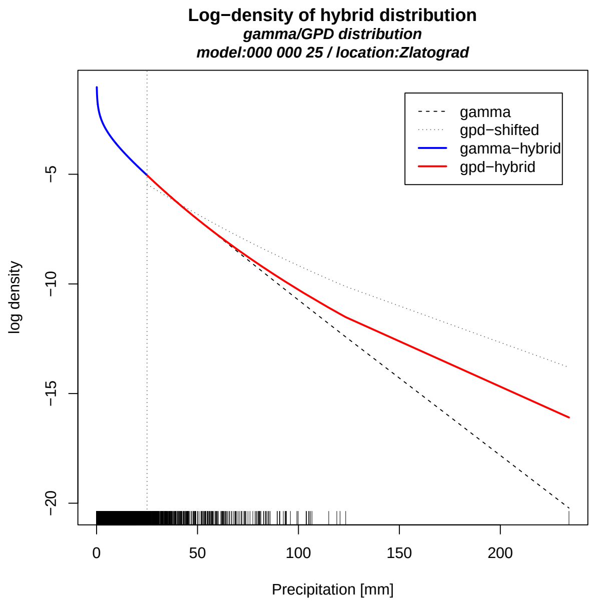
Създаването на хибридно разпределение е целесъобразно при условие, че данните са генерирани от разпределение с тежка опашка. Проверява се хипотезата $\mathrm{H}_0 : \xi = 0$ за експоненциален закон на разпределението срещу алтернативната $\mathrm{H}_1 : \xi > 0$, т.е. разпределението на интензивността да е с тежка опашка. Оценяваме параметрите на GP и експоненциалното разпределение по извадката от данни, които превишават предварително зададен емпиричен квантил на интензивността на дневните валежи. Стойността на прага се избира измежду 80% - 95% емпиричен квантил на интензивността на валежа, за който качеството на апроксимацията на теоритичните GP квантили спрямо емпиричните е удовлетворителна. Понеже интензивността на дневните валежи е времеви ред и данните са зависими, за да постигнем известна независимост данните се деклъстерират (определяне на групи от данни във времето, които превишават зададен емпиричен квантил на интензивността на валежа). Необходимите пресмятания са извършени в средата на $vglm$ процедурата, следвайки стандартната методология за оценяване на есктремалните стойности: (1) деклъстериране на данните, след което извличане на максималната стойност за всяка група от интензивности, превишаваща предварително зададен праг; (2) по метода на максималното правдоподобие и извлечените максимални стойности от стъпка (1) оценяваме параметрите на експоненциалното и GP разпределението; (3) от обектните файлове, които означаваме с $fit.exp$ и $fit.gpd$, съответно извличаме стойности на максимумите на функциите на правдоподобие $logLik(fit.exp)$ и $logLik(fit.gpd)$; (4) образуваме отношението на правдоподобие; (5) с програмата $pchisq$

за χ^2 пресмятаме вероятността на опашката на LRT. Максимално правдоподобните оценки на параметъра на формата на GP разпределението $\xi = shape$ за прагове 10 и 15 са 0.2214157 и 0.2016251, съответно. Отхвърляме хипотеза за експоненциален закон на разпределение на интензивността на валежите за станция Златоград, тъй като стойностите на вероятността на опашката на χ_1^2 са твърде малки - 3.104358e-13 и 3.363121e-08, съответно. Т.е. разпределението на екстремалните стойности е с тежка опашка.

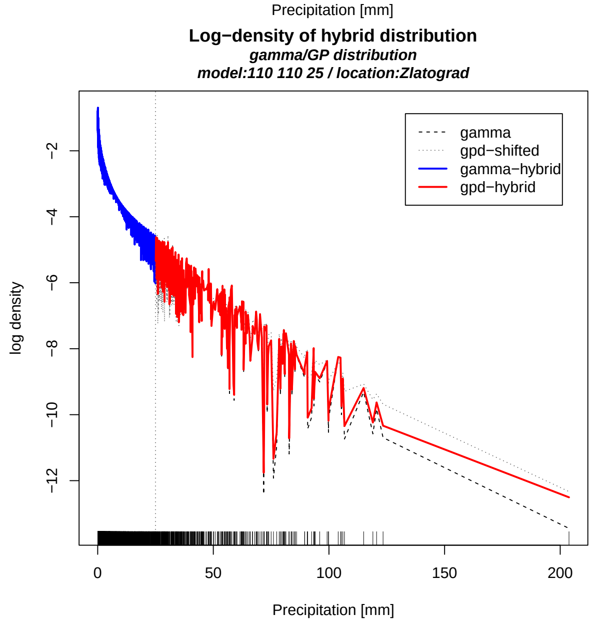
5.3 Хибридни разпределения. Сравнителен анализ

Изборът на праг е от изключително значение за конструирането на хибридните разпределения. Разгледани са модели с прагове 5 mm, 10 mm и 25 mm. Така например: (1) ако бъде избран праг с голяма стойност, оценяването на параметрите на GP разпределението ще бъде по извадка с твърде малък обем, което ще се отрази на надеждността на резултатите, поради големите стойности на оценката на стандартните грешки; (2) ако бъде избран праг с малка стойност, тогава ще бъдат нарушени предположенията за валидност на GP разпределението, докато формалното му използване ще доведе до голяма тежест на опашката, вследствие на което бихме получавали твърде големи предсказани квантилни стойности за дневни валежни суми. На Фиг. 3.а) са показани логаритмите от максимално правдоподобните оценки на плътности с хомогенни параметри (без използване на предиктори): на гама разпределението са представени с непрекъсната синя линия до праг с 25 mm и пунктирна линия след съответния праг; на GP разпределенето са представени с линия от многоточие след прага; на хибридно гама-GP – с непрекъсната линия (синя и червена).

Вертикалната права на плота представлява прага, а интензивността на дневните валежи е дадена с малките вертикални черти по хоризонталната ос. Интерпретацията на резултатите от плота на Фиг. 3.б) е като на плот (а), но за МП оценка на плътността на гама разпределението са използвани атмосферни предикторни променливи. Във връзка с надеждността и приложимостта на разглежданите модели на дневните суми на валежа бе проведен сравнителен анализ между симулираните по модела интензивности на валежа и наблюдаваните интензивности. На Фиг. 4 са показани 95% и 99% емпирични квантили (точките) и моделните квантили с гама (непрекъсната линия) и хибридно гама-GP (пунктирна линия) разпределения на дневните интензивности на валежа с праг 5 mm за модел без атмосферни предиктори. Екстремната стойност от 234 mm, измерена на 5.08.2005 г., е маркирана с плътен триъгълник. Ефектът от хибриди-



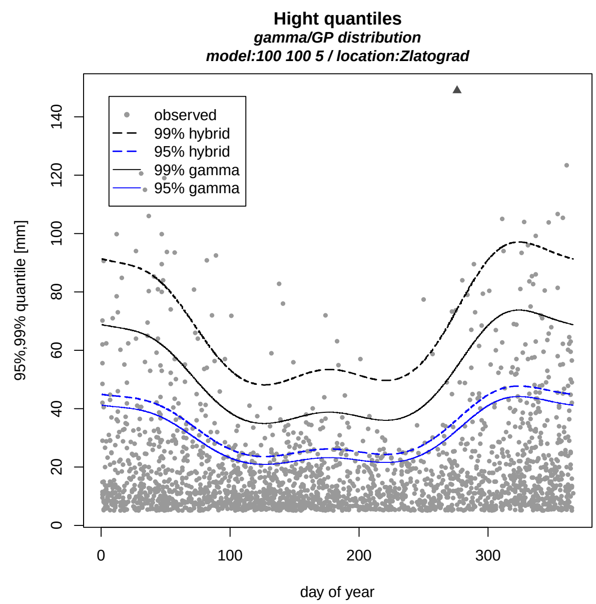
(а)



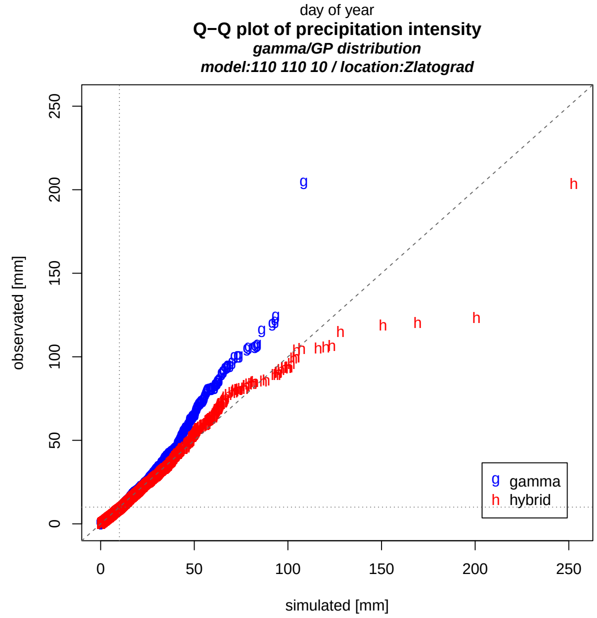
(б)

Фиг. 3: Логаритми от максимално правдоподобните оценки на плътностите на гама, GP и хибридно гама- GP разпределенията. (а) без използване на предиктори в модела; (б) с използване на атмосферни предикторни променливи.

(а)



(б)



Фиг. 4: (а) Емпирични и моделни квантили (95% и 99%) с гама и хибридно гама-GP разпределенията; (б) Q-Q плот на наблюдавани и симулирани чрез гама (g) и хибридно гама-GP (h).

зацията на разпределенията се откроява ясно при по-високите квантили и по-малка стойност от 5 mm за праг. На Фиг. 4.б) е показан Q-Q плот на наблюдавани и симулирани чрез гама (g) и хибридно гама-GP (h) квантили на дневните суми на валежите с праг 10 mm. Ефектът от хибридизацията на разпределенията се откроява ясно. Вижда се, че моделите на интензивността, използващи стандартно дясно скосено разпределение като гама, генерират сходни с наблюдаваните дневни интензитети с изключение на екстремалните стойности на дневните валежни суми. Ефектът от хибридизацията е виден: чрез хибридните модели можем да генерираме дневни валежни суми в целия спектър на наблюдаваните валежи, дори по-големи от наблюдаваните, макар и с малка вероятност.

6 Заключение

В това изследване бяха разгледани няколко модела за интензивността на дневните валежи за станция Златоград, включващи атмосферни предиктори, характеризиращи поведението на атмосферната циркулация над Балканския полуостров. За моделиране на разпределението на интензивността на валежа бяха използвани гама разпределението и хибридното разпределение между него и опашката на обобщеното разпределение на Парето. В резултат на изследването беше установено, че разпределението на дневните суми на валежите за станция Златоград е с тежка опашка. Необходимите изчисления бяха проведени със стандартни статистически процедури като glm и vglm от R. Определянето на прага за свързване на основното разпределение с GP разпределението е напълно субективен, което се оказа сериозен проблем дори при сравнително малък брой атмосферни предикторни променливи. Придобитият опит ще бъде от полза при създаването на модели на часовите суми на валежите за тази и за други станции, което е актуална задача, тъй като те се използват за оценка на риска от наводнения, при проектирането на канализационни мрежи за управление на водите и водните ресурси, за борба с ерозията на почвите и други.

Литература

[1] E. M. Furrer and R. W. Katz (2007) *Clim. Res.* **34** 129
[2] E. M. Furrer and R. W. Katz (2008) *Water Resour. Res.* **44** W12439.
[3] G. K. Grunwald and R. J. Jones (2000) *Environmetrics* **11** 327
[4] N. M. Neykov, P. N. Neytchev and W. Zucchini (2014) *Nat. Hazards Earth Syst. Sci.* **14** 2321.