SocioBrains
ISSN 2367-5721, JOURNAL HOMEPAGE: WWW.SOCIOBRAINS.COM
INTERNATIONAL SCIENTIFIC REFEREED ONLINE JOURNAL WITH IMPACT FACTOR
ISSUE 54, FEBRUARY 2019

APPLICATION OF ARMA MODELS FOR RIVER RUNOFF FORECAST

Abstract: Almost all management decisions are based on predictions about future conditions. River runoff as a major element of water resource systems should be scientifically analyzed as best as possible to ensure that river forecasts are acceptable for practical application in water resource management.

Higher requirements for the adequacy of hydrological information in the conditions of high degree of water use in the context of global climate change necessitate updating and improving the methods and tools for assessing the multi-annual fluctuations of the runoff and its forecast.

Due to the random nature of the forecasts, finding a mathematical model for the replication of runoff time series aims not only to maximize information extraction from the limited available data but also to obtain a satisfactory generation of the historical process of runoff in the future. The ARMA models proved their performance in many applications related to forecasting procedures. Short and medium-term forecasts are used and they are continually updated upon new information. The mathematical formulation and algorithm for ARMA model, as well as the results of their application are given here.

Author information:

Anna Yordanova
Assoc. Prof., PhD
National Institute of Meteorology and Hydrology
✉ Anna.Yordanova@meteo.bg
Bulgaria

Irena Ilcheva
Assoc. Prof., PhD, Eng.
National Institute of Meteorology and Hydrology
✉ Irena.Ilcheva@meteo.bg
Bulgaria

Vesela Rainova
Assist. Prof. PhD, Eng.
National Institute of Meteorology and Hydrology
✉ Vesela.Rainova@meteo.bg
Bulgaria

Keywords:
river runoff, forecast, ARMA, water resource system

Въведение

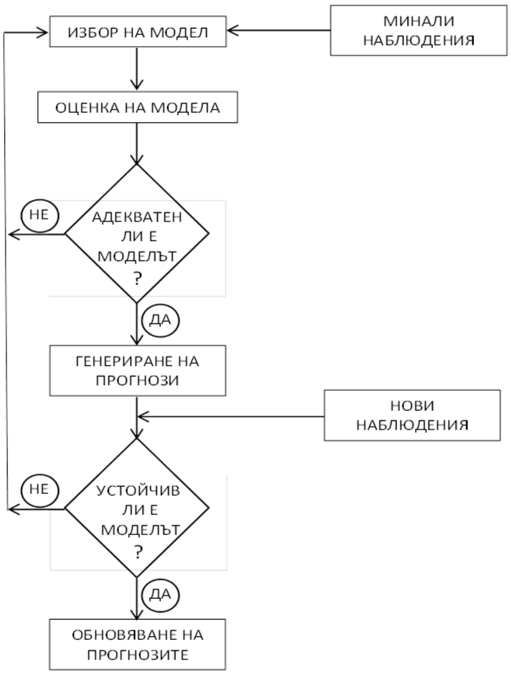
Осъвременяването и усъвършенстването на методите и средствата за оценка на многогодишните колебания на оттока и неговото прогнозиране се налага от процеса на нарастване на изискванията за адекватност на основната хидроложка информация, от високата степен на използване на водите, която се отразява неблагоприятно на природните условия, както и от предупреждението на учените за глобалната промяна на климата: затопляне, изтъняване на озоновия слой, нарастващи концентрации на парникови газове, киселинни валежи и др., които се отразяват неблагоприятно върху количеството и качеството на водите.

Възпроизвеждането на хидроложкия процес на речния отток е основна част от планирането и управлението на водните ресурси. Намирането на математически модел за възпроизвеждането на времевите редове за оттока цели не само максимално извличане на информация от ограничените налични данни, но и чрез получаване на задоволително моделиране на историческия процес на оттока да се извърши екстраполация в бъдещето, т.е. прогнозиране. Без представителна хидроложка информация е невъзможно провеждането на съвременни водностопански изследвания за рационално използване на водните ресурси и получаването на оптимален икономически, социален и екологически ефект.

1. Формулиране на процеса на прогнозиране

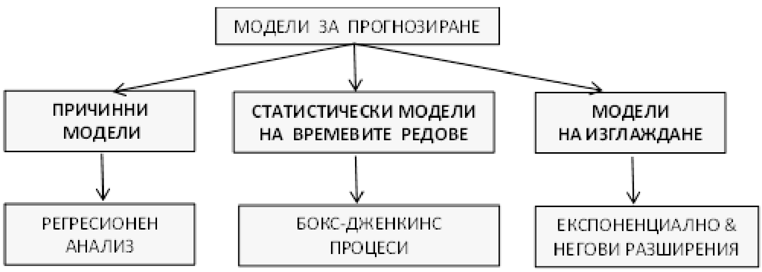
Основната идея зад прогнозирането на случаен процес, представен чрез времеви ред, е да се намери математическа формула, чрез която приблизително да се моделира историческия процес на времевия природен елемент, в случая речен отток.

Процесът на моделиране на прогнози за времеви редици е представен на схема 1[1]. Както се вижда от схемата, в процеса на моделиране на прогнозите има два етапа – етап на построяване на модела (т.е. избор на модела, неговата оценка и проверка на адекватността му) и етап на реализиране на прогнозирането (т.е. генериране на прогнозите, проверка на устойчивостта на прогнозирането при наличие на нови наблюдения и съответното обновяване на прогнозите).



Фиг.1 Схема на фазите при моделиране на прогнозите за времеви редове

Основните модели са класифицирани от проф. Клифорд Хурвич[1,2] в схема, представена на фигура 2:



Фиг.2 Класификация на Хурвич за широко използваните модели за прогнозиране на случаен процес

Причинните модели(детерминистични) се базират на установена зависимост между случайния процес, който ще се прогнозира и други вътрешни или външни фактори, които го определят. Броят на факторите определя броя на параметрите които ще участват в регресионното уравнение описващо процеса.

Статистическите модели за анализ и прогнозиране на времеви редове са доказали своята ефективност в много приложения при прогнозирането на речния отток. Най широко използваните за краткосрочни и средносрочни прогнози са ARMA(авторегресионни с пълзящо средно) модели, предложени от Бокс и Дженкинс.

Експоненциалното изглаждане е модел от типа на тежестното пълзящо средно(WMA), който позволява включване на тренд. Той дава по-голяма тежест на значенията на процеса, които са по-близки по време, отколкото на значенията за моменти, които се намират по-далеч във времето.

2. Описание на процедурата за прогнозиране с ARMA модел

Средномесечните значения $\mathbf{Z_t}$ на оттока могат да се разглеждат като динамични времеви редове, които имат сложна структура, съдържаща стационарни и нестационарни компоненти. Нестационарните се обуславят от климатичните промени, нарушаването на водния баланс при антропогенна дейност и други фактори, които не винаги могат да се отчетет количествено. При идентификацията на модела описващ такива времеви редове се извършва предварително елиминиране на тренда и периодичните компоненти и след това се прави анализ на стационарния остатък. Стационарният остатък възпроизвежда измененията с регулярен характер, които могат да се моделират с помощта на ARMA модели. Те обединяват два модела: на авторегресията (модел с „кратка памет“) и на пълзящото средно (модел с „дълга памет“).

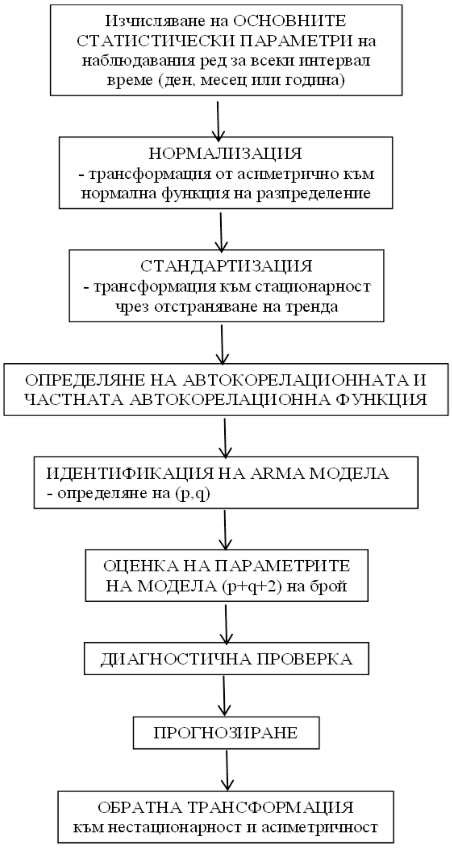
Математическият израз за ARMA модел с **p** на брой авторегресионни параметри **q** на брой параметри на пълзящото средно е следният:

$$z_t = \sum_{j=1}^{p} \phi_j z_{t-j} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

Следователно за модела ARMA(p,q) трябва да се оценят (p+q+2) на брой параметри, а имено:

$$\overline{Z},\ \phi_1, \ldots, \phi_p,\ \varepsilon_1,\ \varepsilon_2, \ldots, \varepsilon_q,\ \sigma_\varepsilon^2$$

Стъпките на моделиращата процедура за времевия ред от значения на речния отток, която използва подхода на Бокс-Дженкинс (идентификация на ARMA модела, оценка на параметрите и диагностична проверка) и прогнозирането са представени на схемата[3,4]:

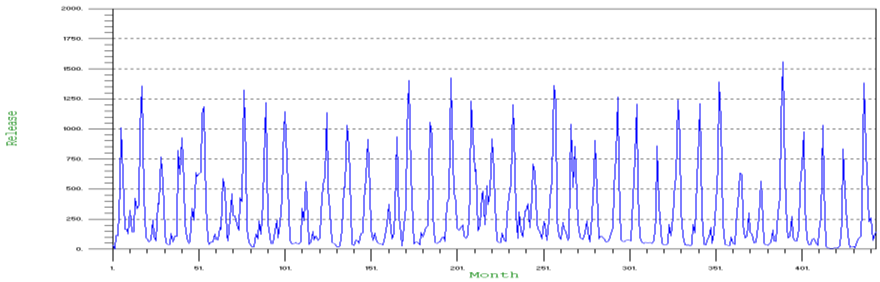


Фиг.3 Схема на алгоритъма при прогнозиране
на времеви редове

3. Пример

Разработена е програма на машинен език Fortran power Station, която реализира описаната процедура от 9 стъпки. Той е използван за месечните наблюдения X_t на ненарушеното речно количество Q(м^3/сек.) – приток към язовир Искър за период от 42 години.

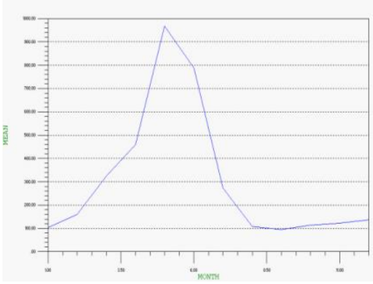
Месечният ходограф на речния отток X_t (t=1,…, 504 месеца) за периода 1959г. до 2000г., с всички нанесени корекции за постигане на хомогенност са илюстрирани на фиг.4:



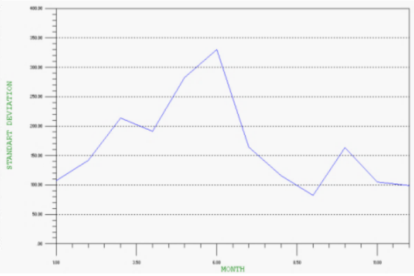
Фиг.4 Ходограф на месечният речен отток

<u>*СТЪПКА* 1.</u> *Изчисляване на основните статистически характеристики*

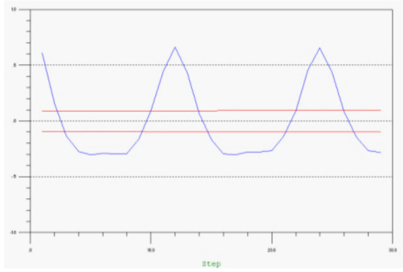
Основните статистически характеристики на оригиналния ред за всеки интервал от време са средната стойност μ_x, стандартното отклонение σ_x, коефициентите на вариация c_v и асиметрия c_s, автокорелационната функция r_k и асиметричното разпределение.



Фиг.5 Графика на средномесеч-ните значения на речния отток



Фиг.6 Графика на стандартното отклонение на месечния отток

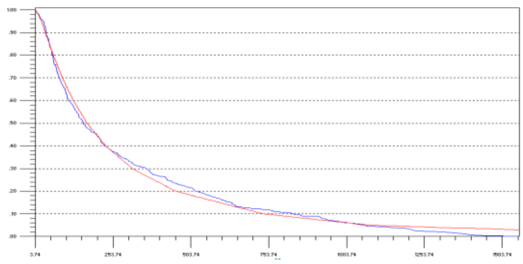


Фиг.7 Автокорелационната функция на наблюдавания месечен речен отток

Графичният анализ за поведението на основните статистически характеристики, определени в стъпка 1, показват, че оригиналния времеви ред на месечния речен отток е нестационарен и с периодичност от 12 месеца.

СТЪПКА 2. Нормализация.

Средномесечният речен приток за яз.Искър е с теоретична функция на разпределение на Джонсън.



Фиг.8 Апроксимация на емпиричното разпределение на месечния наблюдаван отток с теоретичното Джонсъново разпределение

Извършва се трансформация от асиметрично към нормално разпределение на изходния времеви ред за всеки интервал от време като се използва обратната функция на така определеното асиметрично разпределение.

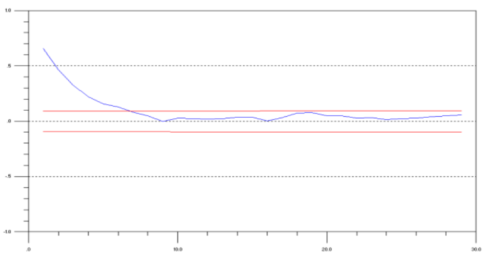
СТЪПКА 3. Стационарност.

За отстраняване на нестационарността се извършва стандартизация на оригиналния времеви ред по формулата :

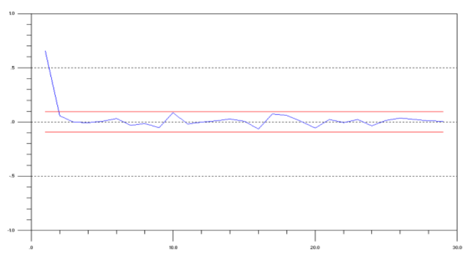
$$z_{i,\tau} = \frac{x_{i,\tau} - \overline{x}_\tau}{s_\tau} \quad , \quad \tau = \overline{1,\ 12} \quad , \quad i = \overline{1,\ 42}$$

СТЪПКА 4. Изчисляване и изчертаване на АКФ и ЧАКФ на трансформираните данни

Функциите автокорелационна функция (АКФ) и частна автокорелационна функция (ЧАКФ) на трансформираните данни са изчислени и изобразени на фиг.9 и фиг.10:



Фиг.9 Автокорелационна функция на трансформираните данни



Фиг.10 Частна автокорелационна функция на трансформираните данни

Тези функции не показват периодично изменение, защото периодичните средно и стандартно отклонение бяха предварително отстранени от реалните данни.

СТЪПКА 5. *Идентификация.*

Функциите АКФ и ЧАКФ се използват за графично идентифициране на ARMA- модела за представяне на трансформирания ред. От графиките може да се направи предположение за наличие на ARMA (2,1) процес.

СТЪПКА 6. *Оценяване на параметрите на модела*

За моделиране на вариациите на средномесечния речен отток са предложени 5 конкуриращи се ARMA модели: ARMA(1,0), ARMA(2,0), ARMA(1,1), ARMA(2,1), ARMA(1,0).

Изчисляват се началните оценки за авторегресионните параметри ϕ_1, ϕ_2, …, ϕ_p и за параметрите на пълзящото средно θ_1, θ_2, …, θ_q по метода на моментите, както и съответната дисперсия на остатъците. Тези оценки участват като начални стойности при изчисляване на по-точните оценки за авторегресионните и за пълзящото средно параметри по метода на максималното правдоподобие. Оценките на параметрите са дадени в таблица 1.

Табл.1 Оценки на параметрите на модела

	модел	ϕ_1	ϕ_2	θ_1	$s^2(e)$
1	ARMA(1,0)	0,663	-	-	0,53
2	ARMA(2,0)	0,621	0,056	-	0,53
3	ARMA(1,1)	0,705	-	0,082	0,53
4	ARMA(2,1)	0,610	0,063	-0,011	0,53
5	ARMA(0,1)	-	-	-0,512	0,51

СТЪПКА 7. *Тест за добро приближение*

- Тест за независимост на остатъчните членове

Приложеният тест този на Portmanteau. Статистиките на този тест за разглежданите модели са дадени в таблица 2.

Табл.2 Статистиките Q и χ^2 за теста на Portmanteau

Модел	Q	χ^2 (95%)
ARMA(1,0)	20,822	40,113
ARMA(2,0)	19,869	38,885
ARMA(1,1)	19,878	38,885
ARMA(2,1)	19,874	37,652
ARMA(0,1)	125,712	40,113

От таблицата се вижда, че нулевата хипотеза за независимост на остатъчните членове ε_t не се отхвърля при първите четири модела, при които изчислената статистика $Q < \chi^2$ (95%), и се отхвърля за последния ARMA(0,1)-модел, поради което той не участва в следващите изследвания.

- Информационен критерий на Акайк (AIC)

Получените стойности по критерия на Акайке за четирите модела са дадени в Таблица 3:

Табл.3 Икономичност на моделите по критерия AIC

Модел	AIC
ARMA(1,0)	0,534
ARMA(2,0)	0,531
ARMA(1,1)	0,530
ARMA(2,1)	0,535

Минималната стойност, получена за модела ARMA(1,1) показва, че той “най-добре” от предложените модели описва средномесечните изменения на речния отток.

ARMA(1,1) моделът за генериране на времеви ред Z_t от месечни значения на речния отток има вида :

$$z_t = 0{,}705.z_{t-1} + \varepsilon_t + 0{,}082.\varepsilon_{t-1}$$

СТЪПКА 8. Прогнозиране

За целта оригиналният ред е разделен на две части, като първата (първите 39 години) се използва за оценяване на параметрите на ARMA (1,1) модела, а втората (последните 5 години), за тестване на прогнозирането. Изборът на разделянето по този начин е да не се намали съществено и без това късия период от години, върху който се определят параметрите на модела.

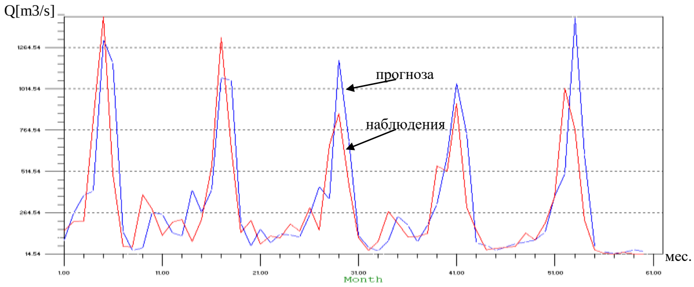
Прогнозирането с L стъпки напред се извършва чрез рекурсивната връзка:

$$\begin{aligned} z_t(1) &= 0{,}705.z_{t-1} + 0{,}082.\varepsilon_{t-1} \\ z_t(2) &= 0{,}705.z_t(1) \\ &\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. \\ z_t(L) &= 0{,}705.z_t(L-1) \end{aligned}$$

В случая е извършено прогнозиране със L=5 стъпки напред.

СТЪПКА 9. Извършва се обратната трансформация на генерирания стационарен ред Z_t към асиметричен ред X_t.

Сравнението графично на наблюдаваните последни 5 години с прогнозираните такива е показано на фигура 11.



Фиг.11 Сравнение на наблюдавания и прогнозирания редове от 60 месечен речен приток към яз.Искър

4. Точността на прогнозата

Метод на Крицкий и Менкел

В руската хидроложка литература и практика за оценка точността на прогнозата се използва метод на Крицкий и Менкел [5], основан на съпоставянето на вероятност, която характеризира повтаряемостта на даденото явление при отсъствие на прогноза с вероятността, получена при използване на прогнозтични зависимости. А именно: изчислява се корелационното отношение:

$$\eta = \sqrt{1 - \left(\sigma_z \big/ \sigma\right)^2},$$

което характеризира точността на връзката между предсказаното явление и факторите на прогнозата. При линейна зависимост, това отношение съвпада с коефициента на корелация.

Тук:

σ - средно квадратично оотклонение на прогнозираната величина;

σ_z - средно квадратично отклонение на емпиричните точки от установената зависимост, т.е. средноквадратичната грешка на проверяваните прогнози, която се изчислява с:

$$\sigma_z = \frac{\sqrt{\sum (z_p - z_o)^2}}{2},$$

където z_p и z_o са съответно прогнозираната и наблюдаваната величини; n – брой на проверяваните прогнози. По настоящем е приета следната скала за оценяване точността на хидроложките прогнози:

Табл.4 Точност на прогнозите по критерия на Крицкий и Менкел

	η	$\sigma_z \big/ \sigma$
добра	$\geq 0{,}87$	$\leq 0{,}5$
задоволителна	0,60 - 0,86	0,51 – 0,80

Тест на Фишер

Като строг критерий за добро приближение на поргнозирани значения спрямо наблюдаваните такива за речния отток се препоръчва тестът на Фишер[5].

За прилагането на този тест се изисква :

1. наблюденията в двете извадки да са случайни и независими;
2. разпределенията на двете генерални съвкупности да са приблизително нормални.

Нека:

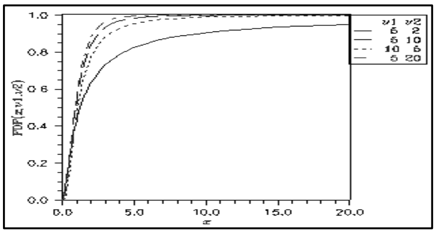
- σ_1^2 и σ_2^2 са двете дисперсии на неизвестни генерални съвкупности;
- n_1 и n_2 са броят на съответно първата и втората извадки;
- S_1^2 и S_2^2 са оценките на дисперсиите , получени от двете извадки.

Хипотезите се дефинират:

H_o: $\sigma_1^2 = \sigma_2^2$ Нулевата хипотеза гласи, че между дисперсиите на двете съвкупности не съществува статически значима разлика.

H_1: $\sigma_1^2 \neq \sigma_2^2$ В общия случай алтернативната хипотеза гласи, че между дисперсиите на двете съвкупности съществува статистически значима разлика.

За проверка на хипотезите в този случай се използва т.нар. F-разпределение (познато още като разпределение на Фишер):



Фиг.12 Разпределение на Фишер

Разпределението има два параметъра: v_1 – степен на свобода на числителя и v_2 – степен на свобода на знаменателя в емпирична характеристика, наричана F-отношение:

$$F = \begin{cases} S_1^2 / S_2^2 & , при\ S_1^2 > S_2^2 \\ S_2^2 / S_1^2 & , при\ S_1^2 < S_2^2 \end{cases}$$

Нулевата хипотеза се отхвърля, ако: емпиричната характеристика F е по-голяма от теоретичната такава, т.е.:

$$F > F_{(\alpha/2,\ v_1,\ v_2)}$$

5. Резултати и изводи

По критерия на Крицкий и Менкел са получени следните резултати:

Табл.5 Резултати за точност на прогнозите по критерия на Крицкий и Менкел

	$\eta = 0{,}74$	$\sigma_Z/\sigma = 0{,}56$
добра	$\geq 0{,}87$	$\leq 0{,}5$
задоволителна	0,60 - 0,86	0,51 – 0,80

Резултатите показват, че получените прогнози са задоволителни.

При теста на Фишер са получени значенията: $F = 2{,}415$ $F_{(\frac{\alpha}{2},\gamma_1,\gamma_2)} = 2{,}978$ т.е.:

$F < F_{(\frac{\alpha}{2},\gamma_1,\gamma_2)}$ и следователно нулевата хипотеза НЕ се отхвърля (между дисперсиите на прогнозираните и наблюдаваните съвкупности не съществува статистически значима разлика).

Въз основа на направените изследвания и анализа на получените резултати могат да се направят следните изводи:

- ARMA моделите са лесно приложими за хидроложки редове от годишни, месечни или дневни значения;

- Различни варианти на тези модели съществуват в литературата, като различията се изразяват само в начина, по който нестационарността и периодичността се изолират от наблюдаваните редове;

- Тези модели са несложни, удобни за формулиране, неикономични на параметри, но е установено, че адекватно представят характеристиките на екстремалните случаи – пълноводие и суша.

References:

1. Box,G., Jenkins, G.: Time series analysis-forecasting and control. Revised edition, Holden-Day, 1976.
2. Chang, Dayly Precipitation Modelling by Diiscrete Autoregressive Moving Avarege Processes, v.20, 1984.,No.5.
3. Joseph, E.S. (1972) Some Applications of Statistic Hydrological Models. Proc. Int. Symp. on Modeling Techniques in Water Resources Systems, Canada.
4. Salas, J.D.,D.C. Boes, ARMA Modeling of Seasonal Hydrologic Series. J. Hyd. Div., ASCE, 1980.
5. Shtorm, R., P.Myuler, P.Noyman, Tablitsы po matematicheskoy statistike, Moskva, „Finansы i statistika“, 1982.