Open Access

Generation of Daily Synthetic Precipitation Series: Analyses and Application in La Plata River Basin

D. H. M. Detzel* and M. R. M. Mine

Department of Hydraulic and Sanitation, Federal University of Parana, Parana, Brazil

Abstract: Precipitation analysis is embedded in a range of important hydrological studies for hydraulic works construction and maintenance. However, flaws and limitations in records are obstacles often encountered by researchers. One feasible solution for overcoming these obstacles is to generate synthetic series. The main objective of this work is to structure and validate a model for generating synthetic rainfall series at a daily scale. A parametric model has been constructed, where the occurrences are determined by a stochastic Markov process and the cumulative rainfall quantities are computed using a mixed exponential probability distribution. Since no previous studies using the proposed probability distribution in La Plata Basin were found in the literature, several significance tests and relevant criteria were applied, in order to verify the model accuracy. The approach was studied in 11 rainfall stations inside Parana and Uruguay rivers basins, located in Brazilian South and Southeast regions, obtaining good results. Additional analyses of the model performance related to extreme events and droughts are also present.

Keywords: Markov chains, mixed exponential distributions, daily precipitation.

INTRODUCTION

Synthetic precipitation series generation has been used for some time by a few researchers to overcome limitations mainly involving the size or conditions of available historical records. Furthermore, the Monte Carlo Method growing use to estimate hydrological variables statistical properties increased the importance in generating consistent series. It should be remembered, however, that this solution still provokes much discussion, for precipitation is a highly complex process. It is useful to look at Zawadzki [1] and Waymire *et al.* [2]; these studies show the complexity involved in the dynamic rainfall process. It is clear that concerns about mathematical models limitations are appropriate, since none can fully reproduce natural phenomena. Obviously, synthetic series generation is not an exception, but even with restrictions it is considered a feasible solution for various problems.

Srikanthan and McMahon [3] developed one of the best references for researchers to generate synthetic rainfall series on annual, monthly and daily scales. Differently from these authors, however, the present study highlights only daily scales. Various existing approaches can be classified according to the processes used. Brissette *et al.* [4] organized them in three large groups: semiparametric or empirical, non-parametric and parametric models. In common, most of these models simulate precipitation occurrences, applying stochastic Markovian processes, or alternating renewal events processes. The difference between them lies in the precipitated amounts determination.

*Address correspondence to this author at the P.O. Box: 19.067-Lactec-Institute of Technology for Development; Brazil;
Tel: +55 41 3361 6178; Fax: +55 41 3361 3143;
E-mail: daniel_detzel@yahoo.com.br, daniel@lactec.org.br

Semiparametric models use histograms that are adjusted to the data, i.e., the process is formulated using statistical parameters and requires calibrations for the appropriate applications. Within these characteristics, the best known model was developed by Semenov *et al.* [5], and called LARS-WG. Boughton and Hill [6] also worked with a semiparametric approach.

The non-parametric models are known for not presenting a defined structure and, therefore, depend exclusively on data available to generate estimates. Researchers interested in working with this type of model concentrate their efforts mainly on trying to reproduce physical mechanism influences in the precipitation process. Bardossy and Plate [7] attempted to model daily precipitation taking into account atmospheric circulation patterns. Young [8], Lall *et al.* [9] and Harrold *et al.* [10, 11] are other examples of non-parametric models. In each study, needed parameters were determined by applying Kernel estimators [12]. Recently, Boulanger *et al.* [13] developed a generation model based on artificial neural networks, achieving good results.

Parametric models are the best known and most widely used by researchers. This approach does not have the same detail level as those presented previously, but offers greater flexibility and easy adjustment, attracting scholars' attention. Precipitated amounts are calculated by applying probabilistic distributions, mainly derived from the exponential family. Outstanding among them are the simple exponential [14-16], two-parameter exponential [17], three-parameter mixed exponential [18-20] and two-parameter gamma [21-25]. Goodness of fit provided by these distributions is variable, depending mainly on the region rainfall regime, where the model is to be applied. Anywise, some authors clearly prefer the two-parameter gamma distribution.

With the great number of existing univariate models, researchers attempted to generalize their formulations in order to generate series in multiple locations simultaneously. This is a very important issue in climate change studies, or spatial relationship evaluation between monitoring stations, for instance. However, it is not a direct process as new variables must be taken into account, especially those that consider correlations between the climatic data. In fact, the main obstacle to multivariate models good performance is precisely the correlation structure reproduction in the study region [4]. Wilks [18] was successful in estimating the synthetic correlation matrices and using them to generate different series by applying pairs of temporal independent, but spatially correlated random numbers. Other examples of multivariate models can be found in [4, 26-28], the latter in the non-parametric field.

The present study structures a model to generate synthetic daily scale precipitation series. Apart from monthly or annual scales, daily record series contain many "zeros", making the model to be developed much more complex. For this reason, researchers who choose to work with the daily scale generally structure the generation models in two distinct phases: occurrences determination and amounts calculation. In this study, occurrences were determined using first order, two states, Markov Chains and amounts were calculated with a three-parameter mixed exponential distribution.

There are also parallel analyses to assess the used considerations validity. Markovian model optimum order to be used or goodness of fit analyses for probability distribution involved, are examples. Assays of this kind, based on tests disseminated in specialized literature, are found throughout this study text.

MATERIALS AND METHODS

A typical parametric univariate model was structured; precipitation occurrences were determined by applying a Markovian stochastic process and accumulated quantities were calculated using a three-parameter mixed exponential distribution. The main references can be found in [16, 18 and 21], the latter in its univariate portion.

Precipitation Occurrences Determination

In many studies wet or dry days are determined using stochastic processes. In this work, first order, two states, Markov Chains were applied, i. e., the current and immediately previous days are considered, under assumption of dry or wet state. This choice was based on the good results obtained in several previous studies, like [15, 18 and 29], among others.

A Markov chain is constructed mainly based on the states definition and on transition probabilities among them. Let X be the current state with indexes "0" for dry day and "1" for wet day. The transition probabilities are, therefore, expressed by:

$$\begin{aligned}\Pr\left\{X_t(k)=1 \mid X_{t-1}(k)=0\right\} &= p_{10}(k);\\ \Pr\left\{X_t(k)=1 \mid X_{t-1}(k)=1\right\} &= p_{11}(k)\end{aligned} \tag{1}$$

This representation is interpreted as $p_{10}(k)$ indicating a dry day preceded by a wet day and $p_{11}(k)$ indicating a wet day preceded by another wet day. Further, t is the index of time and k represents the locality involved. Transition probabilities matrix is built with the complementary probabilities $p_{01}(k)$ and $p_{00}(k)$, however, only the probabilities in (1) are sufficient to implement the model. Calculations were performed by direct counting the respective events found in the regions of interest historical records.

Occurrences synthetic series are obtained initially by defining a critical probability p. This critical probability is redefined every new generated day and assumes the values $p_{10}(k)$ or $p_{11}(k)$ according to the process evolution. Random numbers, uniformly distributed on the interval (0, 1] are used to execute the comparisons that will define both initial state (corresponding to the first day) and the other states.

During their studies, some authors realized that higher order Markovian models had better results when applied under different conditions. Chin [30] confirmed this when noticed that Markovian processes optimum orders can vary according to seasons, geographical location, or even sample size available. Choosing a higher dependence degree on the model implies the use of information from a longer sequence of preceding days to define the current state. Deni *et al.* [31], Jimoh and Webster [32] and Azevedo and Leitão [33] are examples of studies that used various orders Markovian chains to determine precipitation events.

As a parallel analysis, two criteria were used to investigate whether first order Markovian chains are really appropriate events generation in the present study interest region: Akaike's Information Criterion (AIC) [34] and Bayesian Information Criterion (BIC) [35]. Both criteria are based on the principle of parsimony; the model optimum order is obtained from an equation that allows compliance with the adjustment (through the likelihood functions L_m) and a penalty that increases proportionally with the number of parameters (or orders) to be used. The equations are defined by:

$$\mathrm{AIC}(m) = -2L_m + 2s^m(s-1) \tag{2}$$

$$\mathrm{BIC}(m) = -2L_m + s^m(\ln n) \tag{3}$$

Where m represents the order of the Markov chain to be tested, s the number of states, and n the sample size.

Precipitated Amounts Determination

The statistical distribution employed to calculate the precipitated amounts during the days considered wet was the three-parameter mixed exponential. According to Wilks [36] some physical mechanisms have more than one generating process, so they are incompletely represented by a simple statistical distribution. Furthermore, mixed distributions give the models a good flexibility degree, directly reflecting on the results to be obtained.

The probability density function (PDF) is given by:

$$\begin{aligned}f_R\left[r(k)\right] &= \frac{\alpha(k)}{\beta_1(k)}\exp\left[\frac{-r(k)}{\beta_1(k)}\right]+\\ &\frac{1-\alpha(k)}{\beta_2(k)}\exp\left[\frac{-r(k)}{\beta_2(k)}\right]\\ &\beta_1(k)\geq\beta_2(k)>0, \qquad 0<\alpha(k)\leq 1\end{aligned} \tag{4}$$

Where R represents the random variable, r represents the value assumed by the random variable (rainfall amount, proper) and α, β_1 and β_2 represent the parameters. The unknowns entire set is attached to a specific site k. This mixed distribution can also be seen as the sum of two simple exponential functions (one parameter each) intermediated by a probability factor.

The method used to estimate the parameters was Maximum Likelihood, acknowledged for providing good quality estimators, especially in samples with asymptotic tendencies. However, on applying the method classical form, one can note that the mixed exponential distribution parameters are implicit in the formulation, as also attested in [37]:

$$L_m(x;\alpha;\beta_1;\beta_2) = \sum_{i=1}^{n} \ln \left\{ \begin{array}{l} \dfrac{\alpha(k)}{\beta_1(k)} \exp\left[\dfrac{-r(k)}{\beta_1(k)}\right] + \\ \dfrac{1-\alpha(k)}{\beta_2(k)} \exp\left[\dfrac{-r(k)}{\beta_2(k)}\right] \end{array} \right\} \tag{5}$$

So, in order to overcome this undesirable situation, the EM (Expectation–Maximization) algorithm technique was applied. Initially developed by Dempster *et al.* [38], it was mentioned by Wilks [36] as an excellent solution for problems involving mixed statistical distributions. More applications can be seen in [27 and 39].

Once the distribution parameters for a given locality have been estimated, rainfall amounts can be generated. The purpose becomes to determine the random variable R corresponding to the probability function involved. One can use the simple exponential inverted PDF, conditioned to the mean β:

$$r_i(k) = r_{min} - \beta(k).\ln(v_i) \tag{6}$$

where r_{min} represents the minimum quantintity of precipitation for a day to be considered wet and v_i is a random number, uniformly distributed in the interval (0,1]. The mean β assumes the value β_1 or β_2.The choice is made by generating another uniform random number u_i; if $u_i \leq \alpha$, the mean β_1 is chosen; if $u_i \geq \alpha$ the mean chosen is β_2.

The generation model is completed by equation (7):

$$Y_i(k) = X_i(k).r_i(k) \tag{7}$$

Where X_i assumes the values 0 or 1, depending on the occurrences results, as explained in the previous section.

Mixed exponential distribution performance proved superior in some important studies. Roldán and Woolhiser [37] compared this to the simple exponential distribution and the gamma distribution. After careful analysis, the author's conclusion determined that the mixed exponential distribution was much superior to the others, as it provided better results. This was also observed by Wilks [18] and Foufoula-Georgiou and Lettenmaier [20]. However, this distribution behavior in the present study region was unknown, since no similar work applied to this area was found in the literature. For this reason, some analyses about the mixed exponential distribution goodness of fit were performed by constructing probability curves and applying classical statistical inferences: Chi square, Kolmogorov-Smirnov and Filliben's Probability Plot Correlation Coefficient test (PPCC) [40]. The distribution challenged was the two parameters gamma, widely used by researchers. Its parameters were also estimated using the Maximum Likelihood method, with formulation found in Botelho and Morais [41] and Wilks [36].

Model Validation

When one works with synthetic series to be applied in various studies, a large number of them is always generated, intending to reduce the sample error. The generated series set must be able to reproduce the same original series statistical characteristics. It is also useful if they are indistinguishable, yet equiprobable. In agreement with these facts, a set of 1000 synthetic series was generated for each raingauge station chosen. Markov Chains parameters were noted (transition probabilities), as well as the results of the AIC and BIC tests. Then dry and wet days number in the historical series was compared to the generated ones.

For the precipitation amounts, several calculations were performed. Like in the occurrence procedure, mixed exponential distribution parameters were also noted. Then the long term means and standard deviations, total precipitation and maximum daily precipitation were compared. It should be emphasized that for all calculations referring to precipitated amounts, rainless days were ignored. The last parameter calculated for validation was the cross correlation coefficient between the historical and the generated series. As mentioned at the beginning of this section, it is useful for the series to be indistinguishable from each other. Therefore, the result expected from this cross correlation coefficient would be as close to zero as possible.

Worth mentioning that the reproduction of basic statistical characteristics reveals only that the model was correctly implemented. So, with the model duly refined, the validation process begins, focusing analyses in extreme or low frequency events. It should be remembered that these events are not limited to high intensity rains; they also extend to longer than normal dry spells. Thus, the maximum dry or wet consecutive days sequence per period of 1 to 10 days and the empirical probabilities distribution for the dry day sequences were accounted.

Study Area Description

Raingauge stations located in the Parana and Uruguay rivers basins (Brazilian portion of the La Plata Basin) were chosen to apply the model. Both basins play a major role in the hydroelectric power scenario in Brazil. Together, they contain approximately 63.6% of the installed Brazilian hydroelectric potential, besides the potential remaining for future facilities [43]. All this potential is a great motivational factor in choosing these areas.

According to the Köppen-Geider climatic classification most recent form [44], both basins have a predominantly Cfa type climate. Types Cwa and Cwb are also found in the north of the Parana River basin; the northeast of the same basin is classified as Aw type. Finally, isolated points in the north of the Uruguay River basin present a Cfb type climate. More detailed climatic characteristics of each region can be found in [45].

Table 1. Selected Raingauge Stations

#	Station	Code ANA*	Latitude (West)	Longitude (South)	Altitude (m)
MAM	Monte Alegre de Minas	01848000	18°52'20"	48°52'10"	730.00
UCC	Usina Couro do Cervo	02145007	21°20'37"	45°10'13"	813.00
MM	Monte Mor	02247058	22°57'39"	47°17'45"	560.00
Ca	Caiuá	02151035	21°50'00"	51°59'00"	350.00
To	Tomazina	02349033	23°46'00"	49°57'00"	483.00
UV	União da Vitória - 396	02651000	26°13'41"	51°04'49"	736.00
Ta	Taiamã	01655003	16°43'39"	55°31'17"	163.00
Co	Caracol	02257000	22°01'51"	57°01'45"	247.00
PM	Passo Marombas	02750009	27°19'51"	50°45'03"	829.00
LC	Linha Cescon	02753004	27°48'42"	53°01'40"	350.00
Cq	Cacequi	02954001	29°52'40"	54°49'25"	100.00

*ANA is the acronym for Brazilian Water Agency.

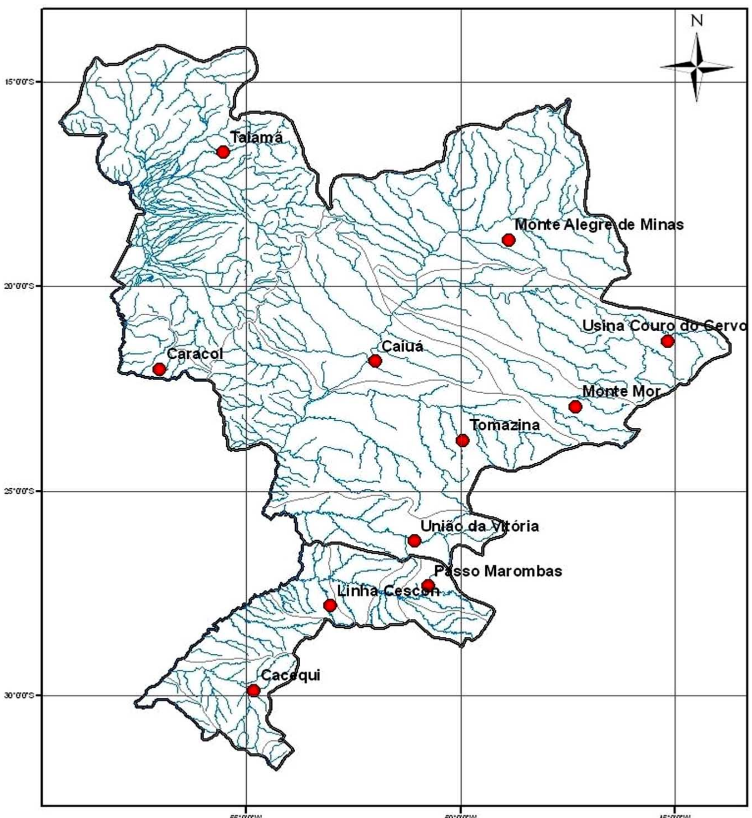


Fig. (1). Study area with selected raingauge.

The mechanism that generates precipitations in the studied basins is strongly influenced by the Polar Atlantic Front. In the Parana River basin Center-North portion, many different types of weather are found during the year. The tendency is towards warm, moist weather in the winter, with an average precipitation of 2000 mm/year. Both Parana River basin South portion and Uruguay River basin, on the other hand, have a discrepancy between the temperature and the rainfall regimes during the year. Despite great thermal variability, the precipitations are regularly distributed and the averages are between 1250 mm and 2000 mm. In all the regions the snowfall records are limited to a few points in the mountain areas, so that the precipitation regime can be characterized by rainfall alone.

In order to test the model under the different climatic conditions found within the study area, it was decided to choose raingauge stations that are well distributed in each basin. Thus, the selected stations number was established as 11, showed in Table **1** and Fig. (**1**). Series used in the model have the same length, covering the period of 01/01/1969 to 12/31/2003, summing 35 years or 12760 recorded days for each raingauge.

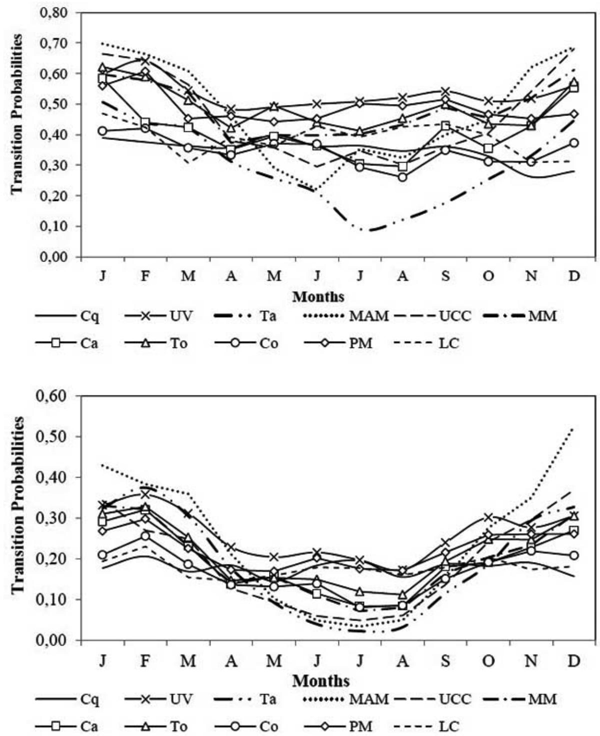


Fig. (2). Transition probabilities behavior (upper - p_{11}; lower - p_{10})

APPLICATION AND RESULTS

A few initial considerations were adopted to obtain the results proposed in this study objectives. First, just as in [14, 18 and 24], it was assumed that there is a monthly seasonality with stationary historical series. For processes dealing with time series, stationarity appears to be a necessary prerequisite for a good reproduction of natural dynamic phenomena, through a finite data interval. Among all the authors researched to develop the present work, few tried to model precipitations considering a non-stationary system [42]; even so, the model had a non-parametric connotation. In parametric models case the stationarity condition is unanimous among researchers.

It is necessary to define the limit of rainfall for a day to be considered wet. Deni *et al*. [31] stressed its importance when noticed that different limit values can change the Markovian Chain optimum order to be used. The limit value for the present study is the same used in [4, 18 and 21]: 0.3 mm.

First results obtained refer to the Markov Chains transition probabilities. It is interesting to note that these probabilities intrinsically provide information about each raingauge dry or wetspells, as well as their magnitude. To illustrate, two comparative graphs were prepared in which rain after a dry day (p_{10}) and of rain after a wet day (p_{11}) probabilities are plotted for all raingauge stations (Fig. **2**). On analyzing these graphs, the different climatic behaviors become clear.

Table **2** provides the general accountancy, in absolute and percentage terms, for AIC and BIC criteria verdict in all 11 considered raingauge (12 seasons each).

As to the mixed exponential distribution, it was observed that the estimated parameters vary considerably, even within a same raingauge station. As previously described, this statistical distribution has the advantage of offering flexibility in fitting the data. It can be measured calculating the ratio between the mean parameters (β_1/β_2) [18]. One can interpret as a differentiation degree indicator that the mixed exponential distribution can perform regarding rainfall intensity. In other words, on being represented by two means, the probability distribution is careful in considering events with different intensities. In this work, the average value found was 3.4. As to the probability parameter (α), the average value found was 0.66, while its standard deviation between all raingauge stations, in a pre-fixed month, presented an average value of 0.17.

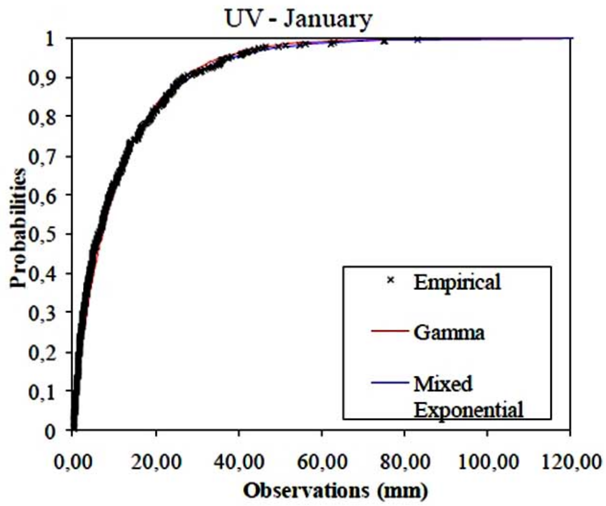
Goodness of fit analysis is shown next. Initially the two probability models are compared to the empirical distribution. Next the Probability-Probability graphs are plotted. They are expressed, respectively, By Fig. (**3**) and Fig. (**4**), elaborated only for large or small samples cases (many or few rainy days, respectively), due to the numerous results (11 raingauge and 12 seasons, summing 132 cases).

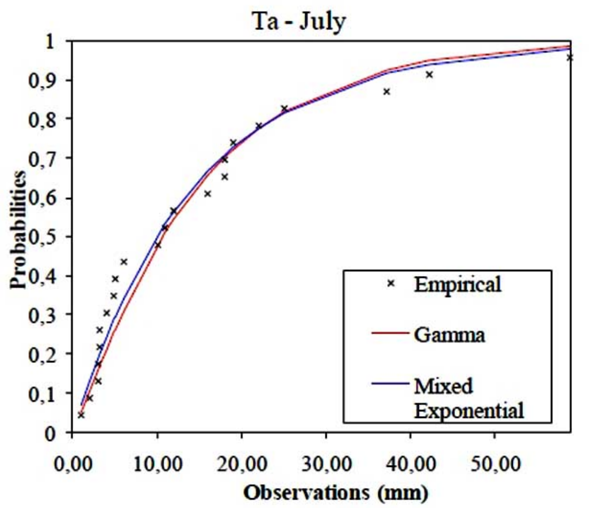
As to the statistical tests applied, the mixed exponential distribution proved slightly superior to the gamma. In the case of the Chi-square test, the statistics for the calculated mixed exponential resulted in a value that was further away from the tabled limit than the gamma distribution. For the Kolmogorov–Smirnov test the gamma distribution was rejected in some cases, but the value calculated was very close to critical, therefore, the rejection can be considered irrelevant. The mixed exponential distribution presented null hypothesis acceptance in all cases.

The PPCC test [40], quoted by many authors as a simple, but powerful test [36] showed somewhat more rigorous than the others aforementioned. Among all raingauge seasons and

Table 2. AIC and BIC Results

	Absolute		Percentage	
	AIC	BIC	AIC	BIC
Optimum order 0	0	8	0%	6%
Optimum order 1	82	119	62%	90%
Optimum order 2	50	5	38%	4%
Totals	132	132	100%	100%





Fig, (3). Comparisons among probabilistic and empirical distributions.

months combinations, four rejections were appointed to the gamma distribution and one for the mixed exponential distribution. For all the tests performed in this study, the confidence level adopted was 95%.

Once the verifications had been performed and the parameters calculated, the series were generated. Compared with the original series, the results were positive, as one can see in Table **3**.

In turn, Table **4** shows results for the accumulated rain amounts, from 1 to 10 days spell. Results are expressed in errors form, calculated in absolute terms (means of maximum values obtained in each series) and in average terms (mean of the values obtained in each series). It should be pointed out that the values in each period were obtained through moving averages which included all elements of the series.

The last item to be evaluated in this study is the empirical probabilities distribution for the low flow periods. As in previous plot showed, only two cases were selected due to the numerous results. Fig. (**5**) exhibits these cases referred to the best and worse fits, respectively.

Table 3. Average and Maximum Errors in the Model Results

	Average Error	Maximum Error
Means	0.3 mm	0.5 mm
Standard Deviations	0.5 mm	1.2 mm
Total Precipitated	82.7 mm	179.7 mm
# Dry/Wet days	1 day	3 days
Maximum daily amount	13.7 mm	52.8 mm
Cross Correlation	Null for all cases	

Table 4. Errors for the Total Amount (mm) Per Period (Days)

Period (days)	Absolute (mm)	Average (mm)
1	13.7	0,1
2	20.3	0.2
3	25.6	0.2
4	27.6	0.3
5	29.6	0.4
6	32.5	0.4
7	34.6	0.5
8	34.5	0.6
9	35.1	0.7
10	36.8	0.7

DISCUSSIONS

Obtained results can be considered quite interesting, from the Markov Chains order determination to the mixed exponential performance. On analyzing AIC and BIC criteria (Table **2**), both indicate a clear preference for the first order. It is perceived, however, that the number of indications for each order differs. These results agree with Katz [46]; in his study, the author concluded that AIC criterion tends to overestimate the chain order to be used. BIC estimator presents more consistent results, and it is recommended for the model optimum order determination. Further, Wilks [18] emphasizes that, in large samples (with more than 1000 elements), the most recommended criterion is BIC.

Still considering Wilks' work, a mixed exponential distribution was fit to rainfall data in some North American raingauge. To evaluate the flexibility degree offered by both β_1 and β_2 means, the author introduced the ratio between them, obtaining 4.8. As showed before, this procedure was also applied in the present study, obtaining 3.4. This lower value can be associated to the large number of raingauge stations in regions where frontal type precipitation predominates.

Comparing the theoretical distributions to the empirical one (Fig. **3**), it is noted that both distributions adjustment becomes worse as wet days number in the sample diminishes. Even so, comparatively, both gamma and mixed ex-

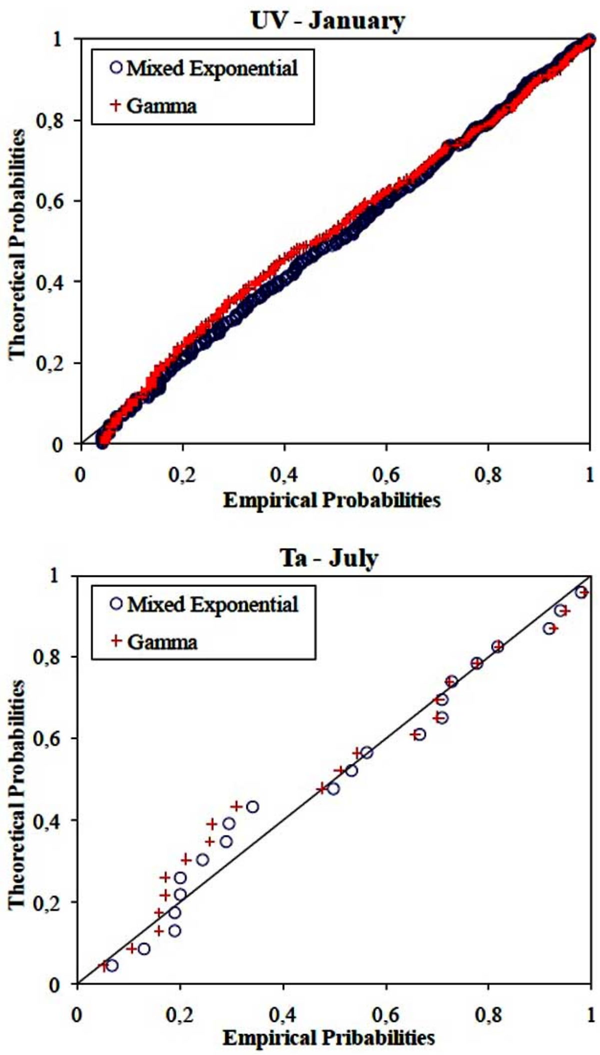


Fig. (4). Probability-Probability plots.

ponential distributions are adequate to represent the precipitated amounts in these regions. Likewise, the plots shown in Fig. (**4**) demonstrate a similar adjustment between the confronted distributions. As in previous graphs, goodness of fit worsens as the number of elements in the sample diminishes, i.e., the drier the period used, more difficult it is for probability models to reproduce the observed values.

Generated series provided the results in Table **3** and Table **4**. It is perceived that the model fulfilled its objectives very well. In agreement with theoretical goodness of fit analyses, the model precision was slightly inferior in dry seasons for all considered raingauge. This is perfectly justifiable, because with less rainy days, the model has fewer information to determine the parameters, which has a direct effect on its precisions.

As to the total precipitations, if it is assumed that the mean error was uniformly distributed among the 35 years of the series, it would result in approximately 2.4 mm/month and 28.3 mm/year. Based on the presupposition that the study area raingauge average precipitation is above 1500.0 mm/year, the error can be considered practically negligible.

Model reverse appears in daily maximum precipitation determination. Errors magnitudes are elevated, which indicates the model deficiency in reproducing high intensity events. Unfortunately, these failures commonly occur in many models [8, 14 and 24], which motivated specific studies as in [23]. Generally, heavy rainfalls are caused by extraordinary atmospheric arrangements and, since most models structures are not able to consider it, extreme phenomena reproduction is impaired.

As to the wet and dry days number, applying the Markov chains produced an excellent result. Errors showed in Table **3** are very small compared to the sample size. In many cases, the model managed to reproduce precisely the same dry and wet days number as the original historical series.

On the statistical long term analyses, it was noted that, in average terms, the model relatively kept up the good performance presented previously. However, it is understood that obtaining excellent results for accumulated totals in absolute terms may have been impaired for two reasons: first, theoretical, is related to the sample variability present in generated series. Since no bias analysis was performed in the estimators, there is no concise notion of its reliability. Thus, even presenting good values in average terms, it may be difficult for the model to reproduce values located at the sample interval extremities. The second reason is related to precipitations physical mechanism; analyzing the historical series, high intensity rainfall events are found, which directly influence the obtainment of totals precipitated per period. Even knowing that some generated series did managed to reproduce these higher intensity events, results were compared with the series averages, for the simple reason that there is an excessively large number of series for individual treatment. Therefore, when one considers the average values among all series, there is an undesirable loss of information.

In maximum events sequences in a same state evaluation, one might say that the model performed well, especially for wet days. For the low flow periods, the model was less precise, resulting in a mean error of six days. Finally, Fig. (**5**) analysis makes it clear that the model performed well in reproducing the empirical probability distribution for drought periods. However, in both cases shown, the sample was relatively large (high number of wet days), since the plot for drier seasons was jeopardized due to the extreme rainless days in a row. In order to overcome this situation, it would be necessary to have longer historical series than the ones used here.

Topics About Multisite Generalization

Once the model was tested and validated, next step would be a generalization for simultaneous application in multisite raingauges. Wilks [18] was the first successful author in achieving consistent results in this matter. As also attested by other researchers [4 and 48] the complexity lies on reproduce the synthetic spatial correlation structure for generate correlated random numbers, which will drive the model. The solution introduced by Wilks was to identify the existing monotonic relationship between generated standard Normal variables and sample correlation and compute, by trial and error, a similar synthetic correlation matrix.

Nevertheless, Brissette *et al.* [4] stressed that sample correlation matrix among raingauges may be non-positive defi-

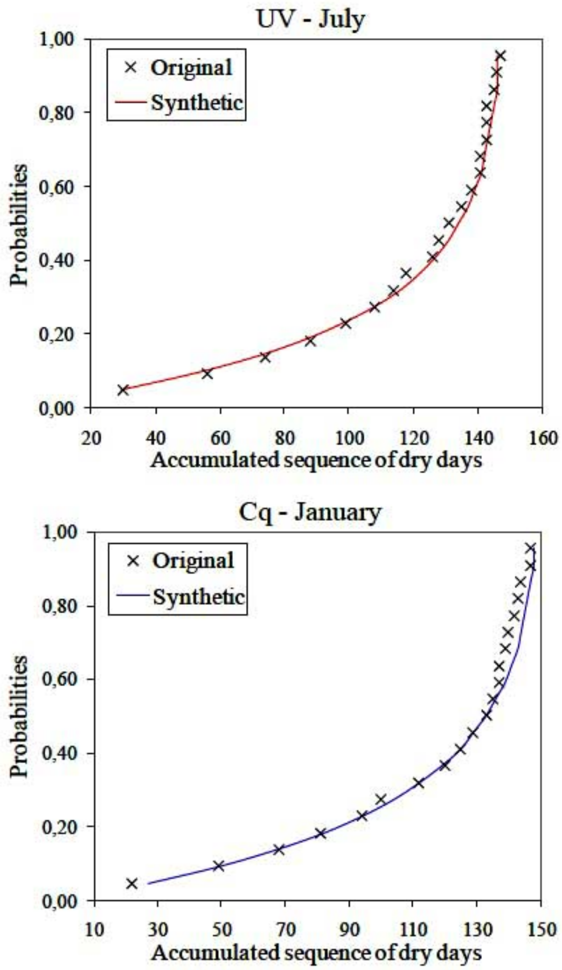


Fig. (5). Empirical probabilities for the drought periods

nite, making it impossible to apply numerical technics, such as Cholesky decomposition, in order to create a standard Normal field. There are many reasons for this, but one can be highlighted: distance between raingauges. With long distances, sample correlation matrices become unstable, mainly due to climatic variability. In his study, Wilks [18] applied his model in rainfall stations from 10 to 500 km apart; Mehrotra and Sharma [47], on the other hand, used raingauges spaced from 20 to 340 km.

In the present work, the multisite generalization is rather critical, since rainfall stations are very distant from each other: the shortest (MM to UCC) is 280 km away and the longest (Ta to Cq) is 1.450 km away. Besides, as attested with this work results, climatic behavior in these regions are distinct, leading to weak sample correlation matrix which would, almost certainly, result in non-positive definite correlation matrices. The recommendation is, then, to limit the study area in smaller regions easing the multisite generalization process.

CONCLUSIONS AND RECOMMENDATIONS

This paper studied a two-part model for daily precipitation synthetic series generation, applying first order, two states, Markov Chains for occurrences determination and a three-parameter mixed exponential distribution for amounts calculation. Its predominantly parametric structure is heavily based on statistical concepts, using information from the historical series. As a main reference, Wilks [18], in its univariate part, was used. Initially developed and executed for the New York State, United States, the model presented very good results, which led the author to credit its performance to the mixed exponential probabilistic distribution, little used in previous studies so far.

Since no previous studies were found in the literature, it was assumed no changes in the structure of Wilks' model, precisely in order to analyze its performance in a climatologically distinct region. Analyzing obtained results in Parana and Uruguay rivers basins, it became very clear that the model repeated its good performance mainly in humid regions. Some worse results were detected in rather arid regions.

Doubts about the mixed exponential distribution performance in the study region were solved after applying diverse analyses. Compared to the gamma distribution (which had, until then, been used more often for hydrological models to generate daily precipitation series), mixed exponential distribution proved superior in all cases. This, together with the AIC and BIC criteria applied to determine the Markov Chain optimum, supplied extra safety in applying the model. Even so, the less precise results obtained, related mainly to dry periods or seasons, lead to endorse future applications preferably in humid climate regions.

Concerning the generalization for multisite raingauges, the authors strongly recommend to limit the study area in order to avoid unstable correlation matrices and, consequently, non-positive definite matrices.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement Nº 212492 (CLARIS LPB. A Europe-South America Network for Climate Change Assessment and Impact Studies in La Plata Basin). First author would also like to thank Coordination of Improvement of Higher Education Personnel (*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* – CAPES) for financial scholarship support. Second author thanks the National Council for Scientific and Technological Development (*ConselhoNacional de DesenvolvimentoTecnológico e Científico* – CNPq) for fellowship granted research productivity.

REFERENCES

[1] Zawadzki II. Statistical properties of precipitation patterns. J Appl Meteorol 1973; 12: 459-72.

[2] Waymire E, Gupta VK. Rodriguez-Iturbe, I. A spectral theory of rainfall intensity at the meso-β scale. Water Resour Res 1984; 20 (10): 1453-65.

[3] Srikanthan R, McMahon TA. Stochastic generation of annual, monthly and daily climate data: a review. Hydrol Earth Syst Sci 2001; 5(4): 653-70.

[4] Brissette FP, Khalili M, Leconte R. Efficient stochastic generation of multi-site synthetic precipitation data. J Hydrol 2007; 345: 121-33.

[5] Semenov MA, Brooks RJ, Barrow EM, Richardson CW.Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates. Climate Res 1998; 10: 95-107.

[6] Boughton WC, Hill P. A design flood estimation procedure using data generation and daily water balance model. Cooperative Research Center for Catchment Hydrology.1997; Paper 97/8.

[7] Bardossy A, Plate EJ. Space-time model for daily rainfall using atmospheric circulation patterns. Water Resour Res 1992; 28(5): 1247-59.

[8] Young KC. A multivariate chain model for simulating climatic parameters from daily data. J Appl Meteorol 1994; 33: 661-71.

[9] Lall U, Rajagopalan B, Tarboton DG. A nonparametric wet/dry spell model for resampling daily precipitation. Water Resour Res 1996; 32 (9): 2803-23.

[10] Harrold TI, Sharma A, Sheather SJ. A nonparametric model for stochastic generation of daily rainfall occurrence. Water Resour Res 2003; 39(10): 1300.

[11] Harrold TI, Sharma A, Sheather SJ. A nonparametric model for stochastic generation of daily rainfall amounts. Water Resour Res 2003; 39(12): 1343.

[12] Parzen E. On estimation of a probability density function and model. Ann Math Stat 1962; 33(3): 1065-76.

[13] Boulanger J–P, Martinez F, Penalba O, Segura EC. Neural network based daily precipitation generator (NNGEN-P). Climate Dyn 2007; 28: 307-24.

[14] Haan CT, Allen DM, Street JO. A markov chain model of daily rainfall. Water Resour Res 1976; 12 (3): 443-9.

[15] Richardson CW. Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resour Res 1981; 17(1): 182-90.

[16] Todorovic P, Woolhiser DA. A stochastic model of n-day precipitation. J Appl Meteorol 1975; 14: 17-24.

[17] Nascimento CES, Kelman J. A stochastic model for generate daily precipitation (in Portuguese) Braz Eng Mag – Water Resour Pap 1995; 13(2): 19-36.

[18] Wilks DS. Multisite generalization of a daily stochastic precipitation generation model. J Hydrol 1998; 210: 178-91.

[19] Smith RE, Schreiber HA. Point process of seasonal thunderstorm rainfall – 2. Rainfall depth probabilities. Water Resour Res 1974; 10 (3): 418-23.

[20] Foufoula-Georgiou E, Lettenmaier DP. A markov renewal model for rainfall occurrences. Water Resour Res 1987; 23(5): 875-84.

[21] Katz WR. Precipitation as a chain-dependent process. J Appl Meteorol 1997; 16(7): 671-76.

[22] Krüger CM, Kaviski E, Müller II. Daily precipitation stochastic generation in the Parana State (in Portuguese). Brazilian Water Resour Mag 1998; 3(2): 87-94.

[23] Grondona MO, Podestá GP, Bidegain M, Marino M, Hordij H.A stochastic precipitation generator conditioned on ENSO phase: a case study in southeastern south-america. J Clim 2000; 13(16): 2973-86.

[24] Liao Y, Zhang Q, Chen D. Stochastic modeling of daily precipitation in China. J Geogr Sci 2004; 14 (4): 417-26.

[25] Gomes AC, Parizi ARC, Peiter MX, Robaina AD, Calgaro M. Modeling daily precipitation in Santiago region, (in Portuguese). [Online] Proceedings of the 14th Brazilian Meteorology Congress; 2006 Nov 27 - Dec 1 [cited 2009 Aug 12]; Florianópolis, Brazil.

[26] Quian B, Gameda S, Hayhoe H, De Jong R,Bootsma A. Comparison of LARS-WG and AAFC-WG stochastic weather generators for diverse Canadian climates. Climate Res 2002; 26: 175-91.

[27] Thompson CS, Thompson PJ, Zheng X, Fitting a multisite daily rainfall model to New Zeland data. J Hydrol 2007; 340: 25-39.

[28] Buishand TA, BrandsmaT. Multisite simulation of daily precipitation and temperature in the Rhine basin by nearest-neighbor resampling. Water Resour Res 2001; 37 (11): 2761-76.

[29] Mehrotra R, Sharma A. Preserving low-frequency variability in generated daily rainfall series. J Hydrol 2007; 345: 102-20.

[30] Chin EH. Modeling daily precipitation occurrence process with markov chain. Water Resour Res 1977; 13(6): 949-56.

[31] Deni SM, Jemain AA, Ibrahim K. Fitting optimum order of Markov chain models for daily rainfall occurrences in Peninsular Malaysia. Theor Appl Climatol 2008; 97: 109-21.

[32] Jimoh OD, Webster P. The optimum order of a markov chain model for daily rainfall in Nigeria. J Hydrol 1996; 185: 45-69.

[33] Azevedo PV, Leitão MMVBR. Application of markov chains processes to daily precipitation in the Paraiba State (in Portuguese). Braz Meteorol Mag 1990; 5(1): 389-402.

[34] Akaike H. A new look at the statistical model identification. IEEE Trans on Automatic Control 1974; 19(6): 716-23.

[35] Schwartz G. Estimating the dimension of a model. Ann Math Stat 1978; 6(2): 461-4.

[36] Wilks DS. Statistical methods in the atmospheric sciences.2nd ed. San Diego: Academic Press 2006.

[37] Roldán J, Woolhiser D. A. stochastic daily precipitation models - 2. A comparison of distribution amounts. Water Resour Res 1982; 18 (5): 1461-68.

[38] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data *via* EM algorithm. J R Stat Soc B Met 1977; 39: 1-38.

[39] Katz WR, Zheng X. Mixture model for overdispersion of precipitation. J Climate 1999; 12: 2528-37.

[40] Filliben JJ. The probability plot correlation coefficient test for normality. Technometrics 1975;17(1):111-7.

[41] Botelho VA, Morais AR. Gamma parameter estimates for pluviometric data of Lavras Town (in Portuguese). Agrotech Sci 1999; 23(3): 697-706.

[42] Rajagopalan B, Lall U, Tarboton DG. Nonhomogeneous markov model for daily precipitation. J Hydrol Eng 1996; 1(1): 33-40.

[43] ANEEL. Brazilian Eletric Power Atlas (in Portuguese). 4th ed. Brasília: ANEEL, 2008.

[44] Peel MC, Finlayson BL, McMahon TA, Updated world map of the Köppen-Geider climate classification. Hydrol Earth Syst Sci 2007; 11: 1633-44.

[45] Mendonça F, Danni-Oliveira IM. Climatology: basic notions and Brazil climates. São Paulo: Oficina de Textos, 2007.

[46] Katz WR. On some criteria for estimating the order of a Markov Chain. Technometrics 1981; 23(3): 243-9.

[47] Mehrotra R, Sharma A. A semi-parametric model for stochastic generation of multi-site daily rainfall exhibiting low-frequency variability. J Hydrol 2006; 335: 180-93.

Received: July 26, 2010 Revised: April 13, 2011 Accepted: April 27, 2011