Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Hydrology 289 (2004) 286–302
Journal of Hydrology
www.elsevier.com/locate/jhydrol

Generation of daily amounts of precipitation from standard climatic data: a case study for Argentina

F. Castellví[a,*], I. Mormeneo[b], P.J. Perez[a]

[a]Dpto. de Medio Ambiente y Ciencias del Suelo, Universidad de Lleida, Av. Rovira Roure, 191, E-25198 Lleida, Cataluña, Spain
[b]Dpto. de Agronomia, Universidad Nacional del Sur, Bahía Blanca, Argentina

Received 17 May 2002; revised 22 October 2003; accepted 21 November 2003

Abstract

We propose a two-part model type for generating daily precipitation from standard climatic data. The objective was to cover the needs of Argentina, excluding its southernmost tip, although the model may also be used for other regions with similar available data. The input for the model was conditioned by the climatic data edited by the National Meteorological Service of Argentina (mean weather variables over 10 years). The model's performance was tested for three cases. In *Case 1*, the mean monthly amount and occurrence of precipitation were both available. In *Case 2*, the mean monthly amount of precipitation was available, but the mean monthly occurrence of precipitation was available for a nearby weather station. In *Case 3*, only the monthly amount of precipitation was available.

Calibration and validation of the model's algorithms was carried out using a wide range of climatic data from sites throughout the world and from all the available sites in Argentina with at least two decades of data covering the period 1950–1990. Use of the model was not recommended at sites near the Andes, beyond latitude 45°S and in Jujuy province. The excluded area represents less than 20% of Argentina's total surface area. Due to data availability, the full performance of the model was mainly evaluated in the province of Buenos Aires (the one with most engineering activity). At Bahía Blanca (in Buenos Aires province, 38°44′S, 62°10′W, Fig. 1) the model reliably reproduced the main features of precipitation required for agricultural, forestry and civil planning uses. In conclusion, the proposed model was very simple and fulfilled the objective of this work; furthermore, some of the results obtained could be extrapolated and applied for other regions.

Keywords: Weather generation; Precipitation; Markov chain; Single-parameter distribution function

1. Introduction

The generation of precipitation requires a range of models whose combination and configuration depend on the processes and temporal and spatial scales involved. Based on the physical processes involved, three general types of models can be classified (Cox and Isham, 1994):

(a) Empirical statistical models, based on stochastic models that are calibrated from actual data. These reproduce annual, monthly and daily precipitation data resembling actual data values.

* Corresponding author. Tel.: +34-973702620; fax: +34-973702613.
E-mail address: f-castellvi@macs.udl.es (F. Castellví).

0022-1694/$ - see front matter
doi:10.1016/j.jhydrol.2003.11.027

(b) Models of dynamic meteorology that incorporate complex non-linear partial differential equations representing different physical processes and that are used for weather forecasting.
(c) Intermediate stochastic models that incorporate a limited number of parameters determined from actual data collected at short time intervals (for example hourly data) and which are used to represent complicated physical phenomena associated with storm precipitation, such as rain cells, rain bands and cell clusters.

Empirical statistical models for generating daily precipitation data at a given site can broadly be classified into four groups: two-part models, transition probability matrix models, resampling models and ARMA time series models. For a complete review of the different models used for each of these last four groups, on different time and spatial scales, see Srikanthan and McMahon (2001).

Here we will concentrate on a type of two-part model for generating daily precipitation at a specific site. Two-part models for daily precipitation consist of two basic steps: first, a model for generating wet and dry events (rainy and non-rainy days); and, second, a model for assigning an amount of precipitation to a wet day. For a reliable simulation, some additional models and interpolation methods may need to be added to both steps in order to capture all the possible yearly, seasonal and daily variations in precipitation.

The first step (part I of the model) can be dealt with two methods as a basis: (a) Markov chains and (b) an alternating renewal process based on wet and dry spells distribution functions. Traditionally, most models have incorporated Markov chains of one or higher orders or hybrid orders (Chin, 1977; Eidsvik, 1980; Stern and Coe, 1984; Katz and Parlange, 1998; Wilks, 1999). The optimum order can be determined by either the corrected Akaike information criterion (Akaike, 1974; Hurvich and Tsai, 1989) or the Bayesian information criterion (Schwarz, 1978). Racsko et al. (1991) reported an alternating renewal process based on a distribution constructed as a mixture of two geometric distribution functions for generating consecutive spells of differing lengths. Using data for Hungary, they obtained a better agreement than using a first-order Markov chain (which follows a geometric distribution). However, Roldan and Woolhiser (1982), working with five US weather stations, found that the first-order Markov chain provided better results than an alternating renewal process using a truncated geometric distribution function for wet spells and a truncated negative binomial distribution function for dry spells. Some models combine Markovian and alternating renewal processes (Srikanthan and McMahon, 2001). Probably, the most commonly used techniques implemented to preserve time dependence in the different parameters involved in the models (such as transitional probabilities, variances and specific parameters) are the following: Fourier series and quadratic spline functions (see Richardson and Wright, 1984; Castellvi and Stockle, 2001; among others), mean-preserving segmented linear interpolation and disaggregated models (see Hershenhorn and Woolhiser, 1987; Mavromatis and Hansen, 2001; Hansen and Mavromatis, 2001; among others).

The second step (part II of the model) is based on the implementation of suitable specific precipitation distribution functions, such as two-parameter Gamma, mixed Exponential and skewed Normal distributions (among many others). Depending on the particular model, the parameters involved in the distribution function (and even the type of distribution function) may vary from year to year, month to month and even from day to day for a given wet day position in a wet spell (Srikanthan and McMahon, 2001).

The importance and utility of measurement precision and time period of precipitation data, depends on the research to be carried out. Series of precipitation data taken on a monthly basis are generally available but they may not be appropriate for certain purposes. Precipitation data recorded over shorter periods are therefore needed to improve planning decisions. For example, a whole month is too long a period for the growing season of some crops, since precipitation events only occur on a few days of the month in question. Data series for the amount of precipitation taken on a weekly or 10 daily basis may be more suitable for some general agricultural and hydrological purposes, but daily records are usually needed to obtain such series. Good quality and long series of daily precipitation are also required to fit or calibrate the set of models involved in rainfall-runoff and crop growth models.

In general, hydrological and agricultural models require complete weather generators which simulate a set of primary weather variables, such as temperature and solar radiation. Researchers have shown considerable interest in modelling and simulating primary weather variables as a tool for assessing planning decisions in agriculture and forestry (Nicks and Harp, 1980; Richardson and Wright, 1984; Geng et al., 1986; Jones and Kiniry, 1986; Pickering et al., 1988; Castellvi, 2001; Castellvi and Stockle, 2001; Castellvi et al., 2001, 2002; among many others). Since these primary weather variables are generally conditioned by rain events, simulating precipitation is a crucial part of the overall data generation process.

Unfortunately, at some sites the available daily series data for precipitation are too short, difficult to obtain due to financial and time constraints, and sometimes either incomplete or not readily available. There is therefore a need to generate daily precipitation data from available standard data for such sites and to thereby bypass these problems. The focus of the present work was to present a simple two-part model type for generating daily series of precipitation data, and to cover a specific need in Argentina. However, territories adjacent to the Antarctic continent remained outside the scope of this work. Monthly means for the amount and occurrence of precipitation were assumed to be available at various locations, but in order to consider different data sources the model was extended to cover sites where only monthly means precipitation data were available. Consequently, the model presented is rather empirical in nature. However, it may be possible to extrapolate some findings and apply them in other parts of the world with similar sources of available data.

2. The model

The two-part model proposed is constrained to the following three cases, or typical situations, which correspond to different quantities and qualities of common source data available for most of the sites in Argentina. All three cases assumed the availability of the monthly amounts of precipitation. In *Case 1*, *the monthly occurrence of precipitation is also available*. In *Case 2*, *the monthly occurrence of precipitation is available for a nearby station with a similar precipitation pattern*. In *Case 3*, *the monthly occurrence of precipitation is not available*. A rainy or wet day was defined as one with precipitation equal to, or greater than, 0.2 mm (rain gauge error). The construction of each part of the model is as follows.

Part I: occurrence of wet days. A first-order two-state Markov chain was used to stochastically generate dry and wet days. The reasoning behind this was as follows: the first-order model usually captures the distribution of wet spells as well as higher order models (Racsko et al., 1991; Wilks, 1999). Except for tropical sites, where Jones and Thornton (1993) suggest that, as a rule of thumb, higher order models are required, it has been shown that first-order model performs well for a wide range of different climates (Bruhn et al., 1980; Richardson and Wright, 1984; Lana and Burgueño, 1998; Wilks, 1999; Chineke et al., 1999; Castellvi and Stockle, 2001). The main deficiency associated with the use of first-order models is that long dry spells are not well reproduced (Racsko et al., 1991; Guttorp, 1995; Semenov and Porter, 1995). Simulations do not tend to generate long dry spells frequently.

The three sources, or cases, of available data seriously constrain the model: higher order and hybrid Markov models were not implemented because they require the respective determination of 2^k and $(k+1)$ parameters, where k denotes the model order (Wilks, 1999). It is also difficult to determine appropriate site-specific distribution functions for dry or wet spells and to implement corrections for low-frequency variability bias, as suggested by Hansen and Mavromatis (2001), to capture the large-scale atmospheric circulation pattern in Argentina, such as El Niño. The model therefore only accounts for high-frequency variability associated to a daily basis. Transitional probabilities were estimated on a monthly basis and assigned to the middle day of the month. A quadratic spline function was then used to assign daily transitional probabilities for specific days of the year.

Part II: amount of precipitation assigned to a wet day. The Gamma and Weibull precipitation distribution functions were selected because their site-specific shape can be estimated from the expected amount of wet day precipitation per month as, respectively, shown in Geng et al. (1986) and Selker and Haith (1990). The model's distribution function therefore varies from month to month. The expected amount of wet day precipitation can be determined in Case 1, but additional procedures need to be implemented in order to make estimations in Cases 2 and 3. To the best of the authors' knowledge, no previous research has compared the performance of the two one-parameter precipitation distribution functions: the Gamma and Weibull functions. Here we compared performances for a variety of climates with aim of establishing and recommending selection criteria.

2.1. Part I: generating wet and dry days

To implement a first-order Markov model, two transitional probabilities are needed. We used the probability of a wet day after another wet one, $p(w/w)$, and the probability of a wet day after a dry one, $p(w/d)$. Since precipitation either occurs or does not occur on a given day, the two complementary transition probabilities are $p(d/w) = 1 - p(w/w)$ and $p(d/d) = 1 - p(w/d)$. As transitional probabilities are conditional, the following expression holds:

$$f_{\mathrm{wet}} = p(w/d)(1 - f_{\mathrm{wet}}) + p(w/w)f_{\mathrm{wet}} \tag{1}$$

The two transitional probabilities needed were estimated for each available data source (the three cases mentioned above) as follows:

Case 1. Since *the monthly occurrence of precipitation is available*, the monthly frequency of wet days, f_{wet}, can be determined and the transitional probability of a wet day after a dry one for each month is estimated according to the following empirical expression:

$$p(w/d)_{\mathrm{est}} = \begin{cases} 0 & f_{\mathrm{wet}} = 0 \\ a_1 + a_2 f_{\mathrm{wet}} & f_{\mathrm{wet}} > 0 \end{cases} \tag{2}$$

where a_1 and a_2 are two site-specific coefficients, respectively. Combining Eqs. (1) and (2), we propose the following expression for estimating the transitional probability of a wet day after another wet one:

$$p(w/w)_{\mathrm{est}} = \begin{cases} \text{undefined} & p(w/d)_{\mathrm{est}} > \dfrac{f_{\mathrm{wet}}}{1 - f_{\mathrm{wet}}} \\ 1 - \dfrac{1 - f_{\mathrm{wet}}}{f_{\mathrm{wet}}} p(w/d)_{\mathrm{est}} & p(w/d)_{\mathrm{est}} \le \dfrac{f_{\mathrm{wet}}}{1 - f_{\mathrm{wet}}} \end{cases} \tag{3}$$

The boundaries in Eq. (3) are necessary because a probability is positive by definition and Eq. (2) derives from a regression analysis. The undefined value must be set to either zero or to a given threshold. Its implementation is required for very dry months when the frequency of wet days is either zero or close to zero.

Case 2. In this case *the monthly frequency of wet days is available for a nearby station (hereafter referred to as a primary station)*, *but not at the site of interest*. This case calls for a method for estimating the monthly frequency of wet days in order to apply Eqs. (2) and (3).

The proposed method is based on similarity. It assumes that both spatial variations in the monthly amount of precipitation and the processes of precipitation formation are similar for the whole region (orography relatively homogeneous). As precipitation patterns are similar, two hypotheses can be made:

Hypothesis 1. The expected amount of wet day precipitation in a given month is similar from place to place.

If the monthly amount of precipitation is also similar, it can therefore be assumed that the monthly frequency of wet days is constant throughout the region

$$f_{\mathrm{wet},i} = f_{\mathrm{wet},p} \tag{4}$$

Hereafter, the subindexes i and p will be used to denote the stations at the site of interest and the primary station, respectively. Eq. (4) assumes that, on a monthly basis, the probability of having a rainy day remains the same.

Hypothesis 2. The expected amount of wet day precipitation in month m (μ_m) cannot be considered similar at stations i and p.

Then, the following proportion is defined with w corresponding to month m

$$\mu_{m,i} = w\mu_{m,p} \tag{5}$$

where w is the ratio determined as the sum of the three monthly amounts of precipitation corresponding to months, $m-1$, m and $m+1$ at station i over the respective sum at the primary station. Thus, Eq. (5) can be rewritten in terms of the corresponding monthly frequency of wet days using the monthly amount of precipitation (P) available at each site as follows:

$$f_{\mathrm{wet},i} = f_{\mathrm{wet},p}\frac{P_i}{wP_p} \tag{6}$$

Eq. (6) assumes that, at sites with a similar monthly occurrence of wet days, daily rainfall totals are often heavier at one site than another.

Case 3. Since *only the mean monthly amount of precipitation is available*, this case therefore requires an empirical relationship linking monthly transitional probabilities and the amount of precipitation. We propose coupling Eqs. (2) and (3) through the following expression:

$$f_{\mathrm{wet}} = b_1(\ln P)^{b_2} \tag{7}$$

where b_1 and b_2 are two site-specific coefficients.

2.2. Part II: assigning amounts of precipitation to wet days

The Gamma precipitation distribution. The density function of the two-parameter (α and β) Gamma precipitation distribution is:

$$g(x,\alpha,\beta) = \frac{x^{\alpha-1}\mathrm{e}^{-x/\beta}}{\beta^{a}\Gamma(\alpha)} \tag{8}$$

where x is the daily amount of precipitation and Γ is the gamma function. Parameter α is dimensionless, usually less than one and mainly considers cases of small amounts of precipitation. Parameter β has units of precipitation. Both parameters are related to the expected daily precipitation, μ through the expression:

$$\mu = \alpha\beta \tag{9}$$

Parameter β is usually greater than one when the amount of precipitation is expressed in millimetres and takes into account heavy rainfall events, as it is also related to distribution variance (V) through the expression:

$$V = \mu\beta \tag{10}$$

Following Geng et al. (1986) parameter β can be estimated on a monthly basis through the expression:

$$\beta_m = c_1 + c_2\mu_m \tag{11}$$

where c_1 and c_2 are two site-specific coefficients. Combining Eqs. (9) and (11), the parameter α can be estimated from the expected amount of wet day precipitation. Therefore, Eq. (8) reduces to a monthly single parameter precipitation distribution function.

The Weibull precipitation distribution. Following Rodriguez (1977) the two-parameter Weibull precipitation distribution, may be converted to a single parameter distribution using the expression

$$W(x,\zeta) = 1 - \exp\left[-\left(\Gamma(1+1/\zeta)\frac{x}{\mu}\right)^{\zeta}\right] \tag{12}$$

where x is the daily amount of precipitation and ζ is a dimensionless parameter directly related with the coefficient of variation (CV) in the following way

$$1+\mathrm{CV}^2 = \frac{\Gamma(1+2/\zeta)}{\Gamma^2(1+1/\zeta)} \tag{13}$$

Assuming that Eq. (10) gives an accurate estimation of actual precipitation variability on a monthly basis, Eq. (13) can be rewritten as

$$1+\beta_m\mu_m^{-1} = \frac{\Gamma(1+2/\zeta_m)}{\Gamma^2(1+1/\zeta_m)} \tag{14}$$

Therefore, when the parameter β_m is estimated and after solving ζ_m in Eq. (14), the monthly single parameter Weibull distribution can be expressed in terms of monthly wet day precipitation only, implementing $\zeta = \zeta_m$ and $\mu = \mu_m$ in Eq. (12).

The expected monthly amount of wet day precipitation in month m, can be determined in *Case 1*, and estimated in *Cases 2 and 3* (using Eqs. (4) and (6), and Eq. (7), respectively) through the expression:

$$\mu_m = \frac{P}{f_{\mathrm{wet}}k} \tag{15}$$

where k is the number of days in month m.

3. The database, background and previous results: the proposed model

3.1. The database. General climatic features

Complete daily data was only available for one Argentinian weather station: Bahía Blanca ($-38°44'$ latitude, $62°10'$ longitude West, in Buenos Aires province, Fig. 1), which had daily precipitation data for a 40 year period. Averaged data relating to monthly amounts and occurrences of precipitation for 2, 3 and four 10-year periods were available at several stations. These data were obtained from 10-year summaries published by Argentina's National Meteorological Service. Data were available for a total of 140 weather stations spread throughout Argentina (excluding its Antarctic territories). Fifty-five sites had four decades of available data corresponding to the period 1951–1990. Forty-one sites had three decades of data: 20 corresponding to the period 1951–1980 and 21 corresponding to 1961–1990. Forty-four sites had two decades of available data corresponding to different decades within the period 1951–1990: 14 sites for 1951–1970, 11 sites for 1961–1980 and 19 sites for 1971–1990. Fig. 1 shows a map of Argentina with the locations of the different weather stations.

Buenos Aires province lies between latitudes -33 and $-41°$. Its climate is temperate due to the maritime influence, though there are also signs of a continental influence in the west. The mean annual temperature

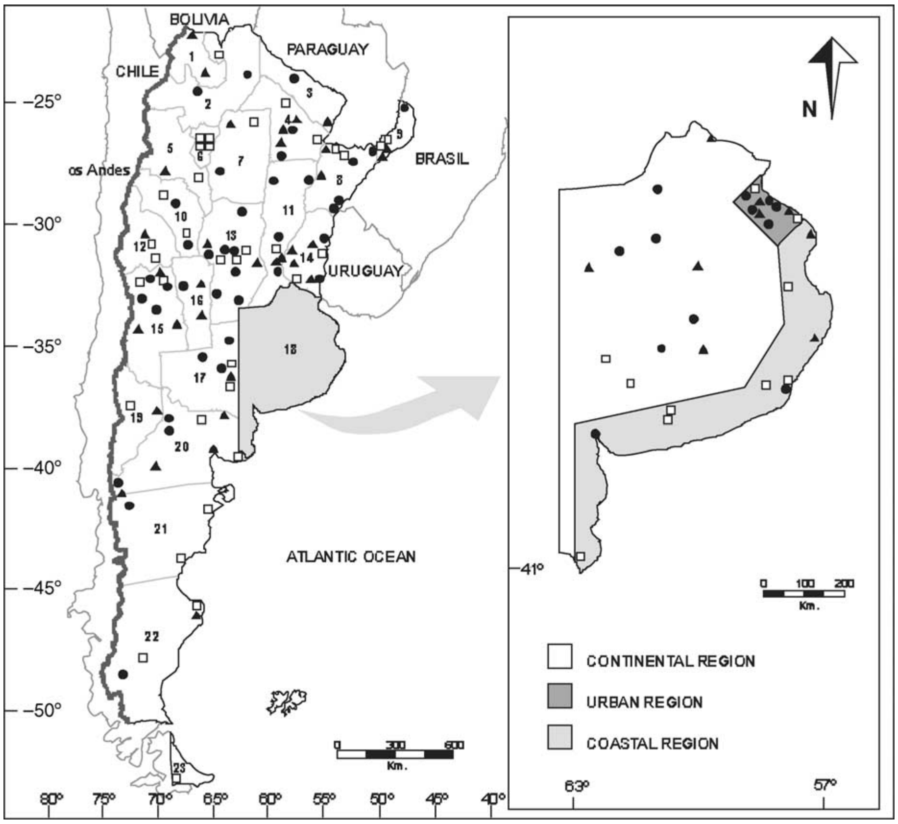


Fig. 1. Main geographical features and locations of available weather stations in Argentina. The symbols ●, ▲ and □ indicate sites with data records available for periods of four, three and two decades, respectively. Numbers denote the different provinces: Jujuy (1), Salta (2), Formosa (3), Chaco (4), Catamarca (5), Tucumán (6), Santiago del Estero (7), Corrientes (8), Misiones (9), La Rioja (10), Santa Fe (11), San Juan (12), Cordoba (13), Entre Ríos (14), Mendoza (15), San Luis (16), La Pampa (17), Buenos Aires (18), Neuquén (19), Río Negro (20), Chubut (21), Santa Cruz (22), Tierra del Fuego, Antártida and Islas del Atlántico Sur (23).

and precipitation at different points in the province are as follows: values of 12 °C and 530 mm in the south near the coast; 16 °C and 950 mm in the continental interior; and 16 °C and 1250 mm at Gran Buenos Aires area (in the vicinity of Buenos Aires city, in the north-east near the coast). The climate at Bahía Blanca is temperate with annual average temperatures oscillating between 14 and 20 °C. Mean annual precipitation is 645.7 mm and the seasons are well differentiated (Capelli and Campo De Ferreras, 1994).

Generally speaking, in Argentina it is possible to recognise three main different regions, excluding the Antarctic territories. The north and east have humid climates; Patagonia, in the south, is dry and characterised by its steep temperature gradient, with very cold zones beyond latitude $-45°$; while the western part of the country exhibits a range of climatic features many of which, like the Föhn effect, are caused by the natural barrier of the Andes.

Another database containing a long series of daily amount of precipitation data was used to complement that available for Argentina. This database, with a total of 17 available weather stations (Table 1), characterised a wide range of different climates around the world. It represented cold and temperate climates in the USA and Europe, and tropical climates in the Philippines, Australia and Africa. The humidity regime was different at each site. Fig. 2 shows the expected amount of wet day precipitation versus the frequency of wet days for each month and weather station. Note the different precipitation patterns covered by the database. A set of determined α and β parameters for Eq. (8), which corresponded to 139 stations spread throughout the USA and obtained from Richardson and Wright (1984), were also added.

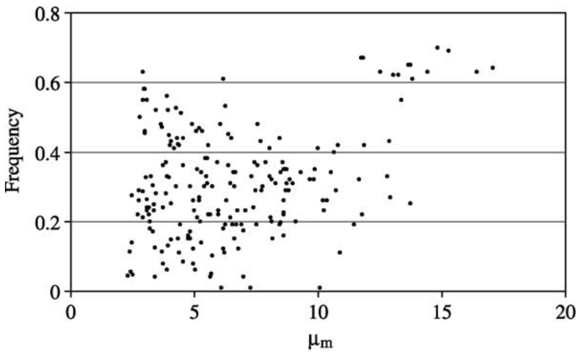


Fig. 2. Monthly frequency of wet days versus expected monthly amount of wet day precipitation μ_m (mm) for climatic zones located in different parts of the world (Table 1).

Table 1
Sites and locations with long data series for daily amounts of precipitation

Site	Country	Latitude	Number of years
Akron	CO, USA	40.09°N	33
Central	Andorra	43.63°N	25
Engolasters	Andorra	43.65°N	40
IRRI (*Dry*)	Philippines	14.22°N	14
IRRI (*Wet*)	Philippines	14.18°N	14
Katerine	Australia	14.28°S	32
Kimberly	ID, USA	42.40°N	30
Lleida	Spain	41.62°N	50
Manhattan	KS, USA	39.20°N	32
Montpellier	France	43.60°N	35
Myal Vale	Australia	30.10°S	31
Pullman	WA, USA	46.77°N	30
Ransol	Andorra	43.64°N	30
Rodeplaat	Africa	25.58°S	30
Glen	Africa	28.95°S	30
Versailles	France	48.9°N	35
Wageningen	The Netherlands	51.95°N	30

IRRI (*Dry*) Longitude 121°15′E Latitude: 14°13′N. *IRRI* (*Wet*) Longitude 121°15′E Latitude: 14°11′N.

3.2. Background and previous results

Calibration and validation of the model requires a database with long period records of daily precipitation at weather stations spread throughout the country because Argentina has a variety of climatic patterns. Unfortunately, as mentioned in the previous section complete information about daily precipitation was not available. Some previously obtained results were compared and combined with findings reported by other authors in order to build a representative model for as wide a range of climates as possible.

3.2.1. Part I: generating wet and dry days

3.2.1.1. The transitional probabilities. Results previously found by other authors showed robust values for the site specific coefficients in Eq. (2). Analysing seven different climatic zones from around the world (Los Baños, Philippines; Wageningen, The Netherlands; Phoenix (AR), Miami (FL), Boise

(ID) and Boston (MA) in the USA), Geng et al. (1986) obtained $p(w/d)_{\text{est}} = 0.75f_{\text{wet}}$, with a global determination coefficient $R^2 = 0.965$ and an intercept a_1 in Eq. (2) that was not statistically different from zero. As a consequence of Eq. (3), for estimating the transitional probability of having a wet day after another wet day Geng et al. (1986) proposed the expression

$$p(w/w)_{\text{est}} = (1 - a_2) + p(w/d)_{\text{est}} \tag{16}$$

regardless of the value for the frequency of wet days. This expression indicates that when plotting the two transitional probabilities versus month, they should tend to exhibit generally parallel trends. The global determination coefficient obtained by estimating $p(w/w)$ was not reported in Geng et al. (1986).

In this work, Eq. (2) was determined using the world database given in Table 1. A good linear relationship between $p(w/d)$ and f_{wet} was also found. The general best fit line obtained was

$$p(w/d)_{\text{est}} = 0.7f_{\text{wet}} + 0.007 \tag{17}$$

$$p(w/w)_{\text{est2}} = \begin{cases} 0.05 & p(w/d)_{\text{est}} > \dfrac{f_{\text{wet}}}{1 - f_{\text{wet}}} \\[2ex] 1 - 0.7(1 - f_{\text{wet}}) - 0.007\dfrac{(1 - f_{\text{wet}})}{f_{\text{wet}}} & p(w/d)_{\text{est}} \le \dfrac{f_{\text{wet}}}{1 - f_{\text{wet}}} \end{cases} \tag{19}$$

with a determination coefficient $R^2 = 0.955$ and with an intercept that, in global, was not statistically different from zero (Fig. 3). However, when using Eq. (17) in Eq. (3) for months with a low frequency of wet days, the departures obtained when estimating

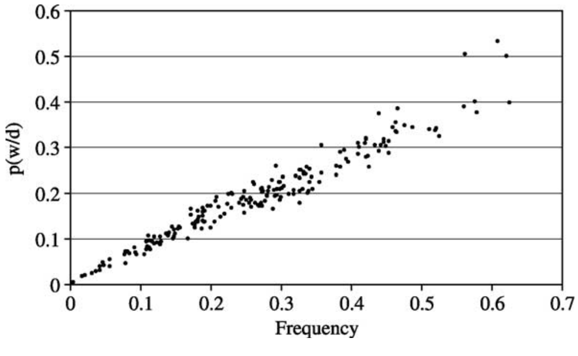


Fig. 3. Relationship between the monthly probability of a wet day after a dry one and the frequency of wet days for the climatic zones presented in Table 1.

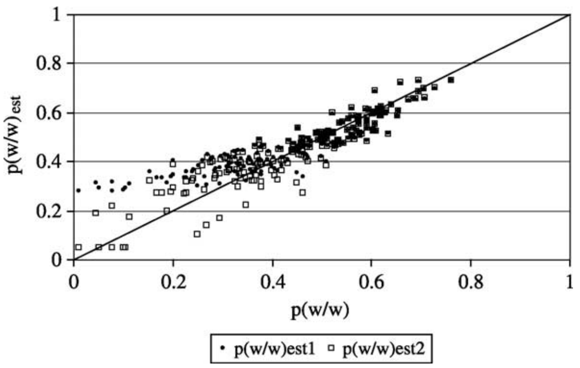


Fig. 4. Comparison between the actual probability of wet day after another $p(w/w)$ and the estimating equations $p(w/w)_{est1}$ (Eq. (18)) and $p(w/w)_{est2}$ (Eq. (19)). The 1:1 line is introduced for comparison.

$p(w/w)$ suggested to us that the intercept could not be set to zero. This can be seen in Fig. 4, which compares the estimated and actual probability of having a wet day after another wet day, using the following two expressions for Eq. (3):

$$p(w/w)_{\text{est1}} = 1 - 0.7(1 - f_{\text{wet}}) \tag{18}$$

In Eq. (19) a threshold of 0.05 was set as a rule of thumb, as it is unusual to find a site without a single rain event. The variances captured using the expressions $p(w/w)_{\text{est1}}$ and $p(w/w)_{\text{est2}}$ were 46 and 83.5%, respectively. The empirical linear relationship represented by Eq. (17) with no intercept, is able to capture more than 95% of the temporal and spatial variability for a wide range of climatic patterns (see Fig. 2). However, it was generally found that significant errors arise when estimating $p(w/w)$ at arid sites. It is therefore necessary to maintain the intercept in Eq. (17) in order to improve estimations of $p(w/w)$ and also to incorporate a wider range of climates. In general, a parallel trend between the two transitional probabilities cannot be assumed as Eq. (3) is not linear.

3.2.1.2. The frequency of wet days. In *Case 1*, the observed frequency of wet days is known. In *Case 2*, in order to avoid empirical equations, both procedures for estimating the frequency of wet days were exempt from calibration. In *Case 3*, Eq. (7) was determined using data from all of Argentina's weather stations with three or four decades of data; a total of 96 well distributed sites. Fig. 5a shows the frequency of wet days versus precipitation for all stations except La Quiaca (−22°06′, 65°36′W) located in Jujuy province. Fig. 5a also shows how trends at four sites, all located in the Andes, differed from the general pattern. These sites were: Bariloche (−41°09′, 71°10′W) and El Bolson (−41°56′, 71°33′W) in Rio Negro province, Rivadavia (−24°10′, 62°54′W) in Salta province and Jujuy (−24°11′, 65°18′W) in Jujuy province. After using regression analysis to adjust Eq. (7), the following relationship was obtained

$$f_{\text{wet}} = 0.03(\ln P)^{1.43} \tag{20}$$

which was able to capture 85.5% of the variance at all of the sites shown in Fig. 5a except those located in and close to the Andes. For the four sites located in the Andes, the best relationship for estimating the frequency of wet days was

$$f_{\text{wet}} = 0.024P^{0.6} \tag{21}$$

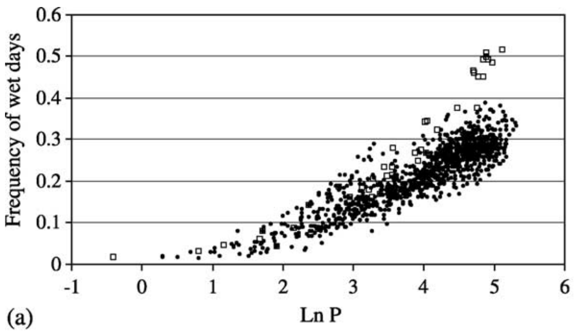
which captured 94.5% of the variance.

Fig. 5b shows that La Quiaca did not follow the two patterns shown in Fig. 5a. La Quiaca is located in the driest part of Jujuy province and suffers a severe Föhn effect deriving from humid air masses proceeding from the Pacific Ocean, crossing Chile and entering Argentina via the highest part of the Andes (3459 m). The Föhn effect at other sites in Jujuy and surrounding provinces is less intensive because they also receive the influences of other air masses proceeding from the east and north-east. For La Quiaca, the empirically obtained best fit was the following linear relationship

$$f_{\text{wet}} = 0.0065P \tag{22}$$

which captured more than 99% of the variance (Fig. 5b).

Overall, without taking into account the Antarctic continent, Eq. (20) can be considered representative for a large region covering almost the whole of Argentina. It only excludes the Andes and adjacent areas such as Jujuy province. Since Eq. (21) was obtained using only four stations and Eq. (22) only one, larger databases covering both the Andes and Jujuy province would be needed to consider these last two relationships as statistically representative. As a result, Eqs. (21) and (22) cannot be recommended with confidence for spatial extrapolation in those specific zones.



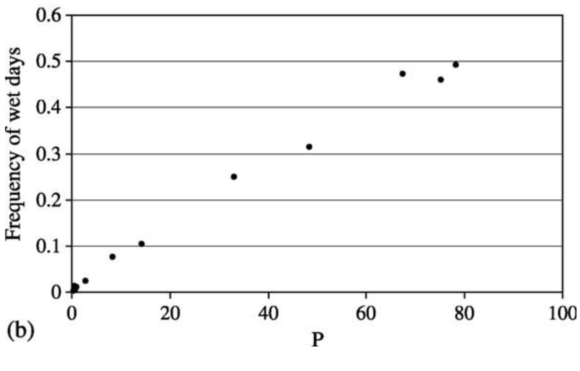


Fig. 5. (a) Relationship between the monthly frequency of wet days and the amount of precipitation P (mm). The general trend is described by Eq. (7). (b) Particular linear relationship corresponding to the driest region of Argentina (La Quiaca, Jujuy province).

3.2.2. Part II: the precipitation distribution functions

3.2.2.1. The single parameter Gamma precipitation distribution. From Eq. (9) and after calibration of Eq. (11), the two-parameter Gamma precipitation distribution becomes a single parameter Gamma distribution. For the site-specific coefficients of Eq. (11), Geng et al. (1986) recommended the values $c_1 = 2.16$ and $c_2 = 1.83$. With these coefficients, Eq. (11) explained up to 96% of the total variance for the seven climatic regions used in their work (mentioned in above section). However, in our case when Eq. (11) was calibrated using a total of 156 sites (including all the sites used in Geng et al. (1986)) we obtained $c_1 = -1.31$ and $c_2 = 1.61$. With these new coefficients Eq. (11) was capable of capturing 94% of the variance.

Since the β parameter is defined positive such linear relationship found from our database is not valid for very dry months. In order to capture as wide a range of climate features as possible, we propose the following relationship

$$\beta = \mu_m^{1.17} \tag{23}$$

Results show that Eq. (23) was capable of capturing the 96.5% of the variance. The good performance of Eq. (23) is shown in Fig. 6a.

3.2.2.2. The single parameter Weibull precipitation distribution. Monthly values of the ζ parameter in the Weibull precipitation distribution (Eq. (13)) were determined for 156 sites, ranging from 0.65 to 0.99.

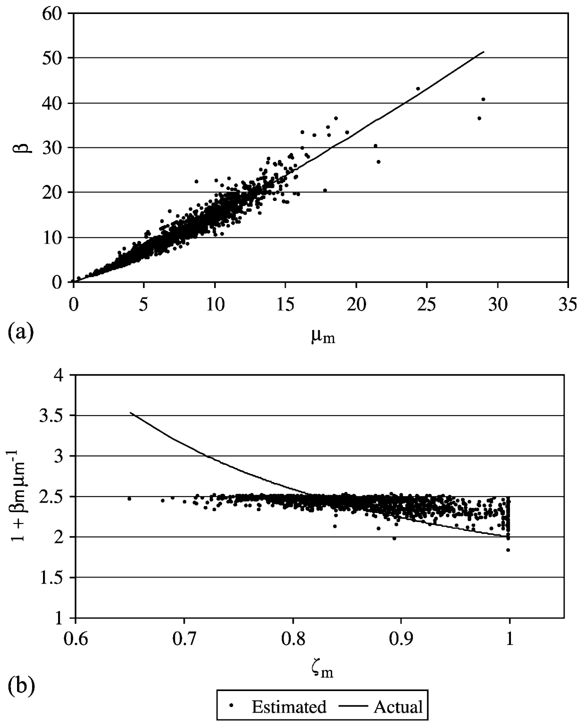


Fig. 6. (a) Empirical relationship between the monthly β parameter in the Gamma precipitation distribution and the expected amount of wet day precipitation μ_m (mm) (Eq. (23)). (b) Actual and estimated monthly values of the function $1 + \beta_m \mu_m^{-1}$ (Eq. (14)) versus the actual ζ_m parameter for the monthly Weibull precipitation distribution (see text).

Then, Eq. (14) was analysed. When the actual and estimated $(1 + \beta_m \mu_m^{-1})$ values were represented versus the ζ_m parameter (Fig. 6b), results showed that, despite the good performance of Eq. (23) (Fig. 6a), it was not possible to estimate ζ_m from the expected monthly amount of wet day precipitation. This was due to the narrow range of the ζ_m parameter. Low departures in β_m introduce large errors in ζ_m.

The ζ parameter in Eq. (13) was obtained by Selker and Haith (1990) using a total of 33 weather stations spread throughout the USA. They proposed a global optimum ζ_m parameter equal to 0.75, regardless of month and climate. Therefore, the recommended single parameter Weibull distribution function (Eq. (12) for month m) is as follows

$$W_m(x) = 1 - \exp\left[-\left(1.191 \frac{x}{\mu_m} \right)^{0.75} \right] \tag{24}$$

This precipitation distribution performed better than the beta-P (Pickering et al., 1988) and exponential (Eq. (12) for $\zeta_m = 1$) distributions at 11 sites representing a wide range of different precipitation patterns in the USA. In this work we therefore adopted the single parameter Weibull distribution proposed by Selker and Haith (1990).

3.2.2.3. Comparing the single parameter Gamma and Weibull precipitation distributions. We used the database in Table 1 for this comparison applying the non-parametric Kolmogorov–Smirnov test. The critical levels of significance obtained using the single parameter Gamma precipitation distribution were higher than using the Weibull distribution in 91.7% of the total number of months. This was mainly due to the fact that the single parameter Weibull precipitation distribution was unable to capture events involving low amounts of precipitation. The single parameter Gamma precipitation distribution passed 80.2% of tests at the 5% significance level.

We sought a filter, or warning relationship, capable of anticipating possibility that performance for a given climate may be limited, and use this to raise user awareness before application of the single parameter Gamma precipitation distribution. This task proved very difficult from the data sources whatever is the case. No relationship was found between the monthly frequency of wet days and the expected monthly amount of wet day precipitation at sites and in months in which the observed precipitation distribution was different (at the 5% significance level) from the single parameter Gamma distribution. The range of the frequency of wet days and the expected monthly amount of wet day precipitation for all the climates in Table 1, were 0.01–0.7 and 2.33–17.6 mm, respectively (Fig. 2). This indicates that Eq. (23) did not apparently lack performance for determined climatic features. Some sites may therefore require another type of precipitation distribution function.

3.3. The proposed model

Without taking into account the Antarctic continent, from the results obtained in the previous sections the proposed two-part model for Argentina corresponding to *Case 1* is constituted by the following set of equations. Eqs. (17) and (19) constitute the part I of the model for simulating wet and dry days. These monthly transitional probabilities were assigned to the middle day of each month and quadratic splines functions were used to assign the first-order Markov chain to each day of year. Eqs. (8), (9) and (23) constitute the part II of the model for simulating amounts of precipitation. Therefore, the single parameter Gamma precipitation distribution for month m, $G(x, \mu_m)$, is expressed as:

$$G(x,\mu_m)=\frac{1}{\beta^{a}\Gamma(\alpha)}\int_0^x x^{\alpha-1}e^{-x/\beta}dx \quad \text{with} \quad \begin{cases}\alpha=\mu_m^{-0.17}\\ \beta=\mu_m^{1.17}\end{cases} \tag{25}$$

For *Cases 2 and 3*, the model also includes the corresponding equations for estimating the monthly frequency of wet days and Eq. (15) for estimating the expected monthly amount of wet day precipitation. At sites holding true to the hypotheses made in *Case 2*, the model implements Eqs. (4) and (6), which are exempt of calibration. If not (*Case 3*), then Eq. (20) is implemented for the whole of Argentina (except for sites located in, or influenced by, the Andes) and Eqs. (21) and (22) should only be implemented at (or close to) sites used for calibration.

4. Procedure for testing the proposed model in Argentina

The database available in Argentina conditioned the validation study of the two parts of the model. Daily data was only available at one station (Bahía Blanca in Buenos Aires province). It was therefore not possible to analyse the performance of the equations for estimating the monthly transitional probabilities and precipitation distribution function using traditional statistical tests at different sites: this was only possible at Bahía Blanca. To test the performance of Eqs. (4) and (6) (*Case 2*), an homogeneous region with a relatively dense weather station network is required because a primary weather station is needed. This situation only existed in the case of Buenos Aires province (Fig. 1). In Buenos Aires province the spatial distribution of precipitation throughout the year depends both on distance from the ocean and on latitude. Therefore, the weather stations were clustered in the following three regions: the *urban region* (the northern part of the province close to the sea), the *coastal region* (the southern part of the province close to the sea), and the *continental region* (Fig. 1). Each weather station played the role of a primary weather station. The urban, coastal and continental region had eleven, four and eight weather stations, respectively (Fig. 1).

In *Case 3*, to analyse the performance of Eq. (7), since calibration was carried out using stations with three and four decades of data, the test was carried out using those stations where two single decades of data were available (90 stations). It should be noted that those stations were well spread throughout Argentina (Fig. 1) and represented different two decades in the period 1951–1990. For the above mentioned reasons, the procedures carried out for testing the performances were as follows.

4.1. Performance of the model at Bahía Blanca

The χ^2-test and the non-parametric Kolmogorov–Smirnov test were, respectively, applied to compare the actual and estimated monthly transitional probabilities and precipitation distribution function (Eq. (25)). In order to illustrate an application of the proposed model, it was used to generate daily precipitation data from the corresponding input data available in each case. The generated series for the different cases were statistically compared with observed data, to test the capacity of the model to reproduce precipitation features that are useful for general engineering assessment. As observed daily data are needed for such task, this operation could only be carried out for Bahía Blanca. The statistics compared were: means and standard deviations for monthly and 10-day periods; some return periods; and the distribution functions for the amount of precipitation, maximum daily precipitation, and sequences of dry and wet days. The Gumbel distribution function performed well in the annual maximum daily precipitation distribution. The t-test, F-test and non-parametric Kolmogorov–Smirnov test were applied for the means, standard deviations and distribution functions, respectively. The null hypothesis was stated as follows: *the generated daily amount of precipitation for a sample of 40 years is a possible manifestation of the observed climate*. The reliability of the global results obtained was analysed from an agricultural and hydrological point of view.

4.2. The performance of the model at the rest of the sites

Since actual daily data was not available, the tests were applied to determine if the estimated frequencies introduced statistically significant departures in the monthly transitional probability and precipitation distribution function. Therefore, the χ^2-test was applied to the transitional probabilities assuming that the one obtained using the actual monthly frequency of wet days (*Case 1*) corresponded to the observed climate. The non-parametric Kolmogorov–Smirnov test was applied assuming that Eq. (25) determined from the actual expected monthly wet day precipitation was the *true or observed precipitation distribution function* at the site. Eq. (25) determined using the estimated expected monthly wet day precipitation, will be referred to as the *estimated precipitation distribution*.

Therefore, the maximum departure between the *true* and the *estimated* distributions was evaluated. For a given amount of precipitation, the departure Δ in the actual monthly precipitation distribution $G(x, \mu_m)$ due to the estimated expected amount of wet day precipitation in month (μ_{m_est}) is

$$\Delta = |G(x, \mu_m) - G(x, \mu_{m\_\mathrm{est}})| \tag{26}$$

In terms of the non-parametric Kolmogorov–Smirnov test, the maximum departure permitted (D_{ls}) for Δ to accomplish the null hypothesis is (Essenwanger, 1986):

$$D_{\mathrm{ls}} = K_{\mathrm{ls}} \left[\frac{n_\mathrm{o} + n_\mathrm{g}}{n_\mathrm{o} n_\mathrm{g}} \right]^{1/2} \tag{27}$$

where n_o and n_g are the number of observed and generated wet days, respectively. The coefficient K_{ls} is a tabulated value listed in Massey's tables that depends on the level of significance (ls) for acceptance of the null hypothesis. The lower the value of K_{ls}, the greater the level of significance. When the number of data pairs is higher than or equal 30, the values of K_{ls} for the levels of significance corresponding to $\mathrm{ls} = 10\%$ and $\mathrm{ls} = 5\%$ are $K_{0.1} = 1.22$ and $K_{0.05} = 1.36$, respectively.

The monthly transitional probabilities used as input in the first part of the model are reproduced in long series of generated data, therefore, so do the frequencies of wet days used in Eqs. (17) and (19). If the number of years generated coincides with the available at each weather station, expressing the number of wet days of the respective actual and generated series in terms of the frequency of wet days and using Eq. (15), the null hypothesis will therefore be accepted when the following expression is accomplished,

$$K_{\mathrm{ls}} \left[\frac{(\mu_m + \mu_{m\_\mathrm{est}})\mu_m \mu_{m\_\mathrm{est}}}{NP} \right]^{1/2} \geq \Delta \tag{28}$$

where N is the number of years and Δ is evaluated using Eq. (25).

5. Results

5.1. Comparing the algorithms of the proposed model

Bahía Blanca station. The performance obtained from Eqs. (17), (19) and (25) after applying the mentioned tests were as follows: whatever the case, all tests were comfortably accepted as values for the critical level of significance were as high as 99.5

and 70% for the χ^2-test and the non-parametric Kolmogorov–Smirnov tests, respectively.

The rest of stations. The results obtained from the test corresponding to *Cases 2 and 3* in Argentina were as follows: whatever the case and site, no transitional probabilities were rejected at the 10% critical level of significance. Some different performance was obtained when comparing the precipitation distribution function for Buenos Aires province. At the 10% level, *Hypothesis 2* made in *Case 2* performed slightly better than *Hypothesis 1* for the urban and continental regions, but not for the coastal region. At the 5% level, both hypotheses performed similarly in each region. Despite the Case 2 did good performance in all three regions (up to 87% of months, the tests accepted the null hypotheses at 5% level), *Case 3* showed its best performance in Buenos Aires province, where none of the tests was rejected at the 10% level of significance.

Fig. 7 shows the values of K_{ls} obtained at sites spread over Argentina from Eq. (28) corresponding to *Case 3*. The line corresponding to $K_{0.1}$ is also represented. It is shown that 94.7% of the months passed the test at the minimum significance level of 10 and 97.3% at 5%. The tests rejected were mainly obtained in San Julian ($-49°19'$, $67°45'$W, Santa Cruz province), and in Usuahia ($-54°48'$, $68°19'$W, Tierra del Fuego province), both in the south of Argentina (Fig. 1). Overall, the simplest and most affordable method (*Case 3*) performed excellently, covered a large part of the country and can be recommended

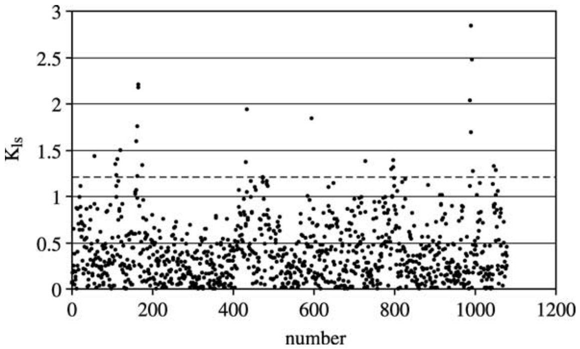


Fig. 7. Values of coefficient K_{ls} (Eq. (28)) to test the monthly single parameter Gamma precipitation distribution, corresponding to *Case 3* at sites where two decades of actual data were available. Dashed line correspond to $K_{ls} = 1.22$, value for the 10% level of significance (see text).

regardless of the available data in Case 2. The total area corresponding to the Andes and the provinces of Jujuy, Santa Cruz and Tierra del Fuego, represents less than 20% of Argentina's total surface area.

5.2. Comparing actual and generated daily precipitation at Bahía Blanca

Three runs of 40 years, one per case, were carried out at Bahía Blanca. Eqs. (4) and (6), corresponding to *Case 2*, showed similar performances at this site so, in order to simplify the results, data was generated estimating the frequency of wet days using Eq. (6) and Mar del Plata was used as a primary station. Based on significance levels obtained in the last subsection for this particular site, Mar del Plata corresponded to the worst selection. It is the furthest station (465 km to the north-east). This was done in order to obtain as wide variability as possible in the model's output.

Reproducing monthly periods. Table 2 shows the results, in terms of significance level, obtained applying the non-parametric Kolmogorov–Smirnov tests to distribution functions for each month. The *Cases 1 and 3* performed well at the 5 and 10% significance levels; as the test would be rejected when $ls < 0.05$ and $ls < 0.10$, respectively, all the tests were passed. At the 5% significance level, the tests for February, June and July failed in *Case 2*. The means and standard deviations determined from the actual and generated series are also shown in Table 2. All the *t*-tests accepted the null hypothesis in *Cases 1 and 3* at the level of 10%. For *Case 2*, at the 5% significance level, the tests for February, June and July did not pass the null hypothesis. For *Cases 2 and 3*, none of the *F*-tests rejected the null hypothesis. However, using *Case 1* the *F*-tests for February, June and July were rejected.

Reproducing 10-day periods. The *t*-tests and *F*-tests for means and variances that were rejected at the 5% significance level are also shown in Table 2. In *Cases 1 and 3*, the means were fully reproduced since all the *t*-tests were accepted, and in *Case 2* the null hypothesis was rejected for one 10-day period in 3 months, February, June and July. The monthly variability for 10-day periods was not fully reproduced in any of the cases. *Case 1* passed the *F*-test in 6 months, January, March, April, May, August and September; but two 10-day periods

Table 2
Level of significance (ls) obtained applying the Kolmogorov–Smirnov test, and observed and generated monthly means and standard deviations (SD). Rejected tests are shown in italics. Every 10-day period within a month rejected by the corresponding test is denoted by (+)

Statistic	Case	J	F	M	A	M	J	J	A	S	O	N	D
ls	1	0.913	0.573	0.759	0.573	0.400	0.263	0.263	0.164	0.573	0.263	0.313	0.760
	2	0.263	*0.003*	0.097	0.100	0.173	*0.000*	*0.015*	0.759	0.573	0.400	0.759	0.400
	3	0.766	0.766	0.990	0.479	0.165	0.266	0.990	0.405	0.165	0.990	0.266	0.279
Mean	Actual	65.2	62.4	85.2	59.7	39.9	35.6	28.2	28.3	45.6	64.5	60.0	71.1
	1	70.2	50.3	90.5	64.2	45.9	23.8	30.3	31.2	62.2	57.2	53.4	72.8
	2	73.5	*91.8*$^{+}$	106.4	82.8	46.1	*56.6*$^{+}$	*40.0*$^{+}$	33.8	53.7	69.2	56.1	78.6
	3	71.8	54.5	85.1	56.1	31.6	38.6	28.3	28.6	41.9	62.5	47.4	55.2
SD	Actual	42.3	52.1	58.4	48.0	29.8	40.7	26.1	25.2	29.7	42.0	42.5	45.7
	1	47.8	*33.2*$^{++}$	59.5	50.1	30.8	*20.9*$^{+}$	*20.4*$^{+}$	23.2	37.9	33.0^{++}	36.0^{+}	43.9^{+}
	2	35.1	53.7^{++}	54.2^{+}	40.9	26.8	35.5^{+}	23.2^{+}	25.1	38.3	39.4	40.5	41.7^{+}
	3	46.4	42.4	49.5^{++}	35.8^{+}	28.0^{+}	29.8^{+}	25.1^{+}	22.5^{+}	31.8^{+}	39.0	33.7^{+}	38.3^{+}

rejected the null hypothesis in February and October, and in the rest of the months one period was rejected. *Case 2* passed the F-test in 7 months, January, April, May, August, September, October and November; and in the rest of the months one of three was rejected, except in February when two periods were not reproduced. For *Case 3*, the monthly variability was reproduced in 3 months, January, February and October. March was the most difficult month to reproduce, since the null hypothesis was rejected for two periods. For the rest of the months, one of three did not pass the F-test.

Reproducing annual maximum daily precipitation: return periods. The maximum absolute differences between actual and generated Gumbel distribution functions were 0.120, 0.215 and 0.247, using *Cases 1, 2 and 3*, respectively. Since Eq. (27) provided a value $D_{0.05} = 0.3$, the test accepted the null hypothesis at the 5% significance level, for all cases. Table 3 shows the agreement between actual and generated values for extreme events of daily precipitation, for different return periods. *Case 2* did the best performance in reproducing this statistic.

Reproducing sequences of wet and dry days. The maximum number of consecutive dry days observed in the real data was 49, whereas *Cases 1, 2 and 3* generated 53, 44 and 58, respectively. All *Cases* were not able to reproduce the actual dry spells distribution. They performed poorly for sequences of from 1 to 3 days. The maximum number of consecutive wet days observed in the real data was 10, whereas *Cases 1, 2 and 3* generated 10, 12 and 10, respectively. The wet spells distribution were only well reproduced in *Case 1*. *Cases 2 and 3* performed poorly for sequences of 1 and 2 days.

Reliability of the model. Wheat is the most important crop in the coastal region, and the soils are mainly Petrocalcic Paleustoll: fine-loamy, mixed and termic (Soil Survey Staff, 1999). On average, the percentage of total agricultural land dedicated to wheat production in Bahía Blanca province was between 76 (Gargano et al., 1990) and 95.4% (Saldungaray et al., 1996). Wheat sowing and collection take place in June or July and in late December or early January. The harvest period for most crops and fruit trees is before the onset of winter.

Good performances were obtained in all cases when reproducing means on a monthly and 10-day period basis, except in February, June and July using the *Case 2*. In general, monthly variability was also well replicated, but intravariability for periods of 10 days in a given month was difficult to reproduce for

Table 3
Actual and generated annual maximum daily precipitation for various return periods (T)

T	Actual	Case 1	Case 2	Case 3
2	62.6	57.8	59.4	58.8
5	86.8	78.6	87.4	78.4
10	102.8	92.3	106.0	91.4
20	118.1	105.5	123.8	103.9
30	127.0	113.0	134.1	101.1
50	138.0	122.5	146.9	110.0
100	152.9	135.3	164.2	142.1

about 50% of the months. In such cases, crop varieties growing on sandy soils with short root systems could be affected by the reduction in soil moisture. This is not, however, generally the case in the province.

The variability in the amount of precipitation during the months of February, June and July was difficult to reproduce using *Cases 1 and 2*, and also for *Cases 1 and 3* in October and March, respectively. For general agricultural purposes, except for *Case 2* this question was not relevant for February, March, June and July. As previously mentioned, the last wheat harvest takes place during January. June and July are the driest months, and water percolation can be negligible, so the most crucial parameter is the total amount of precipitation rather than how it is distributed over 10-day periods during this month. Furthermore, in most areas sowing starts in mid June, while October is a relevant month as crops are growing. However, percolation is not important because the evapotranspiration rate is not very significant until the period from mid November to mid February, and its influence on the water balance and on crops may be negligible since means over 10-day periods were well reproduced. In general, for Bahía Blanca the *Cases 1 and 3* did the best performance being the *Case 3* slightly better than *Case 1*.

The distribution of annual maximum daily precipitation and return periods were well reproduced. The model may therefore be useful for long-term hydrological planning assessment. The estimated transitional probabilities were not able to reproduce dry and wet sequences over short periods. As precipitation conditions all other primary weather variables, the utility of the model for assessing problems deriving from short-term weather events seems questionable. However, this needs to be tested implementing the proposed model in other models.

Generation processes for other weather variables, which are conditioned by the dry or wet status of a particular day, has been widely used in several weather generators such as WGEN (Richardson and Wright, 1984), WXGEN (Wallis and Griffiths, 1995), CLIGEN and USCLIMATO (Johnson et al., 1996) or CLIMGEN (Castellvi and Stockle, 2001; Stockle et al., 2001), among others. The precipitation model proposed in this work may therefore be incorporated into more complete weather generators. This will make possible to assess the consequences deriving from dry and wet spells, which include water stress, drying processes, fire risk and erosion and to evaluate the reliability of the fully reproduced climate.

6. Summary and concluding remarks

We aimed to develop a model to generate daily amounts of precipitation for Argentina (excluding its Antarctic region). The three most frequently available sources of data were analysed. *Case 1* assumed the availability of data of the monthly amount and occurrence of precipitation. *Case 2* assumed the availability of the monthly amount of precipitation data and its occurrence in the surrounding area. And *Case 3* simply assumed the availability of data on the monthly amount of precipitation. To serve our purpose, we needed a model exempt from site-specific algorithms, because local calibration was not possible. Most weather precipitation generators require daily data as input and for this reason we based our model on empirical algorithms, determined using a wide range of different types of climate around the world, and capable of including a variety of climate patterns.

The empirical model was not able to include all the climate patterns present in Argentina. As a consequence, its application was not recommendable either at sites located beyond latitude 45°S, in Jujuy province (the coldest and driest parts of Argentina, respectively), or in the Andes (Fig. 1). More data would be needed to calibrate the proposed algorithms for these areas, but they occupy less than 20% of Argentina's total surface area.

The model recommended was based on a first-order Markov chain with two states and the single parameter Gamma precipitation distribution function. The two transitional probabilities were estimated using and depending on the data available at each site. A simple relationship between the transitional probabilities and the frequency of wet days was derived that was applicable for a wide range of climates.

It was only possible to test *Case 2* in Buenos Aires province because it requires a relatively dense weather station network in a region with homogeneous orography. The results obtained using *Case 3* were slightly better than those using *Case 2*. In the estimation of the frequency of wet days from the amount of precipitation, the good performance of Eq. (7) allowed us to develop a simple, straightforward model that served our overall purpose. Since Eq. (7) captured 85.5% of the spatial and time variability for Argentina, it was therefore unnecessary to cluster similar precipitation patterns in order to apply *Case 2*. It was also unnecessary to find specific regional calibrations of Eq. (7) in order to apply *Case 3*. However, the form of Eq. (7) cannot be extrapolated to other countries. For example, the climates listed in Table 1 do not fulfil this relationship (not shown).

Unfortunately, it was only possible to test the proposed model at Bahía Blanca, where it performed very reasonably. This good performance was obtained by replicating some climatic features that are useful for general engineering purposes (means and variances for monthly and 10-day periods, precipitation distribution, maximum daily rainfall, return periods and dry and wet spells). Soil properties and agricultural activity can be extrapolated for the majority of the coastal area. If there was nothing to suggest that a good performance obtained at one particular site could be considered a purely local question, similar results would be expected in the coastal region. It should be noted that Buenos Aires is the most heavily populated province with 40% of the country's total population. It consequently concentrates the majority of Argentina's engineering activity.

In conclusion, the proposed model coupled with a map showing the monthly amount of precipitation may constitute a useful tool for assessing engineering projects across most of Argentina. This model may be combined with others to generate other primary weather variables and thereby complete a useful weather generator for use in Argentina. Some of the results obtained in this work, the estimations of the transitional probabilities and the reduction of the Gamma distribution as a function of a single parameter, could be considered useful for applying to sites in other countries with similar needs to those highlighted in this work.

Acknowledgements

The authors would like to thank Raul Diaz for providing daily precipitation data for Bahía Blanca, Buenos Aires. The authors thank C.O. Stockle for providing data and encouragement, and also thank the valuable reviewers' comments that improved the manuscript. This work was funded by the Spanish Ministry of Science and Technology under project REN2001-1630/CLI, the AECI program and the Universities and Research Department of the Generalitat de Catalunya under project 2001SGR-00306.

References

Akaike, H., 1974. A new look at statistical model identification. IEEE Trans. Autom. Control, AC 19, 716–722.

Bruhn, J.A., Fry, W.E., Fick, G.W., 1980. Simulation of daily weather data using theoretical probability distributions. J. Appl. Meteorol. 19, 1029–1036.

Capelli, A., Campo De Ferreras, A., 1994. La transición climática en el sudoeste Bonaerense. Sigeo. Departamento de Geografía, Universidad Nacional del Sur, Bahía Blanca, Argentina, 77 pp.

Castellvi, F., 2001. A new simple method for estimating monthly and daily solar radiation. Performance and comparison with other methods at Lleida (NE of Spain); a semiarid climate. Theor. Appl. Climatol. 69, 231–238.

Castellvi, F., Stockle, C.O., 2001. Comparing the performance of Climgen and Wgen in the generation of temperature and solar radiation. Trans. ASAE 44 (5), 1683–1687.

Castellví, F., Stockle, C.O., Ibañez, M., 2001. Comparing a locally-calibrated versus a generalized temperature weather generation. Trans. ASAE 44 (5), 1143–1148.

Castellví, F., Stockle, C.O., Mormeneo, I., Villar, J.M., 2002. Testing the performance of different processes to generate temperature and solar radiation: a case study at Lleida (northeast Spain). Trans. ASAE 45 (3), 571–580.

Chin, E.H., 1977. Modelling daily precipitation occurrence process with Markov chain. Water Resour. Res. 13, 949–956.

Chineke, T.C., Jagtap, S.S., Aina, J.I., 1999. Applicability of a weather simulation model based on observed meteorological data in humid tropical climate. Theor. Appl. Climatol. 64, 15–25.

Cox, D.R., Isham, V., 1994I. In: Barnett, V., Turkman, K.F. (Eds.), Stochastic Models of Precipitation. Statistics for the Environment 2. Water Issues. Wiley, New York, pp. 3–18.

Eidsvik, K.J., 1980. Identification of models for some time series of atmospheric origin with Akaike's information criterion. J. Appl. Meteorol. 19, 357–369.

Essenwanger, O.M., 1986. General Climatology, 1B. Elements of Statistical Analysis. Elsevier, Amsterdam, 423 pp..

Gargano, A.O., Aduriz, M.A., Saldungaray, M.C., 1990. Sistemas agropecuarios de Bahía Blanca: 1-Clasificacion y descripcion mediante indices. Revista Argentina de Produccion Animal 10 (5), 361–371.

Geng, S., Frits, W.T., de Vries, P., Supit, I., 1986. A simple method for generating daily rainfall data. Agric. For. Meteorol. 36, 363–376.

Guttorp, P., 1995. Stochastic Modelling of Scientific Data. Chapman & Hall, London, Chapter 2.
Hansen, J.W., Mavromatis, T., 2001. Correcting low-frequency variability bias in stochastic weather generators. Agric. For. Meteorol. 109, 297–310.
Hershenborn, J., Woolhiser, D.A., 1987. Disagregation of daily rainfall. J. Hidrol. 95, 299–322.
Hurvich, C.M., Tsai, C.L., 1989. Regression and time series model selection in small samples. Biometrika 76, 297–307.
Johnson, G.L., Hanson, C.L., Hardegree, S.P., Ballard, E.B., 1996. Stochastic weather simulation: overview and analysis of two commonly used models. J. Appl. Meteorol. 53, 366–372.
Jones, C.A., Kiniry, J.R. (Eds.), 1986. CERES-Maize: A Simulation Model of Maize Growth and Development. Texas A&M University Press, College Station, TX, p. 194.
Jones, P.G., Thornton, P.K., 1993. A rainfall generator for agricultural applications in the tropics. Agric. For. Meteorol. 63, 1–19.
Katz, R.W., Parlange, M.B., 1998. Overdispersion phenomenon in stochastic modelling of precipitation. J. Climate 11, 591–601.
Lana, X., Burgueño, A., 1998. Daily dry-wet behaviour in Catalonia (NE Spain) from the viewpoint of Markov chains. Int. J. Climatol. 18 (7), 793–816.
Mavromatis, T., Hansen, J.W., 2001. Interannual variability characteristics and simulated crop response of four stochastic weather generators. Agric. For. Meteorol. 109, 283–296.
Nicks, A.D., Harp, J.F., 1980. Stochastic generation of temperature and solar radiation data. J. Hidrol. 48, 1–7.
Pickering, N.B., Stedinger, J.R., Haith, D.A., 1988. Weather input for nonpoint source pollution models. J. Irrig. Drain. Eng. 114 (4), 674–690.
Racsko, P., Szeidl, L., Semenov, M., 1991. A serial approach to local stochastic weather models. Ecol. Model. 57, 27–41.
Richardson, C.W., Wright, D.A., 1984. WGEN: A Model for Generating Daily Weather Variables. US Department of Agriculture, Agricultural Research Service, ARS-8, 83 pp.
Rodriguez, R.N., 1977. A guide to the Burr type XII distributions. Biometrika 64, 129–134.
Roldan, J., Woolhiser, D.A., 1982. Stochastic daily precipitation models. 1. A comparison of occurrence processes. Water Resour. Res. 18, 1451–1459.
Saldungaray, M.C., Gargano, A.O., Aduriz, M.A., 1994. Evaluacion fisico-economica de los sistemas agropecuarios de Bahia Blanca en 1994 comparados con los de 1988, Revista Argentina de Economia Agraria. XXVII Reunion anual, AAEA, Santa Fé, Argentina, 11 pp..
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6, 461–464.
Selker, J.S., Haith, D.A., 1990. Development and testing of simple parameter precipitation distributions. Water Resour. Res. 26 (11), 2733–2740.
Semenov, M.A., Porter, J.R., 1995. Climatic variability and the modelling of crop yields. Agric. For. Meteorol. 73, 265–283.
Soil Survey Staff, 1999. Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Surveys, Agriculture Handbook 436, second ed., United States Department of Agriculture, Natural Resources Conservation Service, USA, 863 pp.
Srikanthan, R., McMahon, T.A., 2001. Stochastic generation of annual, monthly and daily climate data: a review. Hydrol. Earth Syst. Sci. 5 (4), 653–670.
Stern, R.D., Coe, R., 1984. A model firing analysis of daily rainfall data. J. Roy. Stat. Soc. A 147 (Part 1), 1–34.
Stockle, C.O., Nelson, R., Donatelli, M., Castellvi, F., 2001. Climgen: a flexible weather generation program. Proceedings of the European Society for Agronomy Congress: Agroclimatology and Modeling, July 2001, Italy, pp. 229–230.
Wallis, T.W.R., Griffiths, J.F., 1995. An assessment of the weather generator (WXGEN) used in the erosion/productivity impact calculator (EPIC). Agric. For. Meteorol. 73, 115–133.
Wilks, D.S., 1999. Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. Agric. For. Meteorol. 93 (3), 153–169.