Bayesian approach to daily rainfall modelling to estimate monthly net infiltration using the Thornthwaite water budget and Curve Number methods.

E. ZIMMERMANN

E.Zimmermann (✉) Department of Hydraulics, National University of Rosario, Rio Bamba 245 bis - 2000 Rosario, Argentina.

E-mail: erikz@fceia.unr.edu.ar
Phone - Fax: +54-341-4808541

Abstract

The Thornthwaite and Mather water budget is a simple and frequently applicable tool to estimate surpluses of water, which are not stored in the soil profile. Combining it with the empiric CN-method of the US Soil Conservation Service (US-SCS), which is applied to daily rainfall records, it is possible to estimate the runoff, and this way, from the difference between surpluses and runoff, to estimate the net infiltration that would recharge a phreatic aquifer. In order to apply both methods during a sequence of years it is necessary to predict the number of rain events per month, and the rainfall depth for each event. In this work a methodology based on the theorem of Bayes to estimate the number of occurrences of rainy events in a considered month conditioning the forecast to the monthly rainfall is proposed. In addition to that, an exponential distribution to generate rainfall depth knowing the monthly rainfall was done. Both algorithms were applied in four stations of the southern region of Santa Fe province (Argentina). More than 7600 forecast of rain occurrences and rainfall depths were compared with the observed records. Moreover, values of runoff estimated by means of the US-SCS method, using the observed rainfall and using rainfalls predicted with the algorithms were also compared. In both cases, the obtained results were also very satisfactory. The proposed methodologies allow the correct application of the balance of Thornthwaite and Mather together with the US-SCS method and a good forecast of monthly runoff and net infiltration.

Keyboards: Groundwater recharge/water budget, Statistical modelling, Bayesian approach, Argentine, rainfall/runoff

Introduction

The amount of water that enters and leaves the groundwater environments needs to be quantified in order to make a proper sustainable management of groundwater resources.

Water budget methods can be used to estimate net infiltration in groundwater systems when direct methods, such as of water-table fluctuation, tracer techniques, etc. (Scanlon et al. 2002) can not be applied due to required data is not always available. Net infiltration is the amount of water that percolates downward in the unsaturated zone toward the water-table, after subtracting the portions of water that are subject to runoff, evapotranspiration, and soil storage. The total amount of net infiltration approximately corresponds to the direct recharge of groundwater.

The Thornthwaite and Mather (1957) water budget is a simple and frequently applicable tool to estimate surpluses of water, which are not stored in the soil profile. It is possible, by combining the Thornthwaite and Mather water budget with the Curve Number (CN) method (USDA-SCS, 1986), to estimate the runoff, and then, from the difference between surpluses and runoff, to estimate the net infiltration.

Scozzafava y Tallini (2001) modified the Thornthwaite method in order to determine the contribution of net infiltration alone, a result that cannot be achieved with the standard method. In the Thornthwaite method, the water surplus refers to all the excess water that flows out from the soil, without discriminating net infiltration from runoff. To distinguish these two contributions, they adopted the following process: (a) monthly runoff (RO) is determined with the SCS-CN method; (b) the RO value so obtained is subtracted from total rainfall (P); and (c) the water budget is computed on the basis of a fictitious rainfall value (PI) which is equal to P - RO. PI is the amount quantity of water available for net infiltration and for actual evapotranspiration, determined with the Thornthwaite formula.

The modified method can be applied to serial monthly data, and the net infiltration can be estimated in a long-term. Then, its average value, and its time of evolution can be estimated for water management and planner purposes.

Thornthwaite water budget only requires average monthly temperature data and monthly rainfall data. But considering that the CN method must be applied for

daily rainfall records, the following additional information will be required: (a) number of rainfall events for month, and (b) rainfall depths for each event.
Since this information is frequently unavailable, a model of daily rainfall is proposed in this paper. The model here presented is derived into two steps: the first one is the derivation of the probability distribution of the number of rainy events for a given month conditional upon the monthly rainfall; the second one is the derivation of the probability distribution function of the rainfall depth for each single event considering the number of rainy events.

Bayesian inference for a number of rainfall events

Given the monthly precipitation, P, and a random number of rainfall events N, in the considered month. Given *a priori* probability of the number of events for a given month f(N). A probability distribution function (pdf) for N, fitted for every month of the year, could be adopted, but the forecast would improve if additional information such as the monthly precipitation P is used.
Given the probability distribution function of the monthly rainfall P conditional upon the number of rain occurrences N, f(P|N).
Therefore, according to the Bayes's theorem, *a posteriori* probability of the number of rain occurrences N conditional upon the monthly rainfall P, f(N|P), can be determined as follows:

$$f(N \mid P) = \frac{f(P \mid N) f(N)}{f(P)} \qquad (1)$$

and,

$$f(P) = \sum_{j=1}^{Nmax} f(P \mid N_j) f(N_j) \qquad (2)$$

where Nmax is a maximum number of rainfall events for a month.
The problem is that usually neither a priori probability f(N) nor conditional probability f(P|N) are known. It is a common practice to assume that the occurrence of rainfall events is distributed according to a Poisson distribution (Todorovic, 1967, Appied Probability Course Notes, Fort Collins, Colorado, USA; Eagleson, 1972; Cox and Isham, 1994; Arnaud and Lavabre, 1999).

Let a time period of one month in which a sample of N storms has been recorded, and let λ_1 be the mean storm occurrence rate, i.e. the mean number of storms per month. If the number of occurrences N is a random variable distributed according to a Poisson distribution, the *a priori* probability f(N) can be written as follows:

$$f(N) = \frac{\lambda_1^N e^{-\lambda_1}}{N!} \qquad (3)$$

Todorovic (1967, quoted by Antigüedad et al 1995) proposed a cumulative distribution function of total precipitation P, produced by N storms, distributed according to a Gamma distribution, which can be written as:

$$F(P \mid N) = 1 - e^{-\lambda_2 P} \sum_{j=0}^{N-1} \frac{(\lambda_2 P)^j}{j!} \qquad (4)$$

The physical meaning of λ_2 is the inverse of the mean precipitation depth produced by a single storm. It can be estimated by:

$$\lambda_2 = \frac{\lambda_1}{P_{av}} \qquad (5)$$

where P_{av} is the average monthly precipitation.
The probability distribution function can be obtained by derivation of eq. (5), as follow:

$$f(P \mid N) = \frac{\lambda_2^N e^{-\lambda_2 P} P^{N-1}}{(N-1)!} \qquad (6)$$

This conditional distribution is also of Gamma type (Erlang's distribution) with N and λ_2 as parameters (Montgomery and Runger, 1996). Combining the equation (1), (2), (3), and (6) the *a posteriori* probability can be determined as follow:

$$f(N \mid P) = \frac{\frac{\lambda_2^N e^{-\lambda_2 P} P^{N-1}}{(N-1)!} \frac{\lambda_1^N e^{-\lambda_1}}{N!}}{\sum_{j=1}^{N\max} \frac{\lambda_2^{N_j} e^{-\lambda_2 P} P^{N_j-1}}{(N_j-1)!} \frac{\lambda_1^{N_j} e^{-\lambda_1}}{N_j!}} \qquad (7)$$

Now, it is possible to formulate an algorithm to define the most probable number of rainfall events for one month, when the monthly precipitation is known.

Algorithm to define numbers of monthly rainfall events conditioned to monthly precipitation values.

- Data available: A series of monthly precipitation, P, and monthly average values of number of rainfall events, N_{av}, and monthly rainfall, P_{av}.
- Assign parameters of the probabilistic distributions: λ_1=N_{av} and λ_2 estimated with (5).
- Calculate for each year and each month the values of the *a posteriori* probability f(N|P) using (7) where N ranged from 1 to Nmax.
- Select an optimal value of N, Nop, for each month and year such that f(Nop|P) is the greatest of f(N|P), N=1..Nmax.

Statistical approach to estimate rainfall depths

After to estimation of the number of monthly occurrences of the rainfall events is performed, it is necessary to determine the rainfall depth for each one. Many authors suggest exponential distributions to represent rainfall depths, spell between events and duration (Eagleson, 1972; Arnaud and Lavabre, 1999; Seoane and Valdés, 1994).

In the southern region of the Santa Fe province, the exponential distribution was the most appropriate to fit depths of daily rainfall (Zimmermann et al, 1996; Riccardi and Zimmermann, 2000).

Exponential distribution presents, also, the advantage of offering simplicity because it has one parameter, and an explicit expression for the *pdf.* If the rainfall depth P_n for a single event, in a given month, is distributed according to an exponential distribution, the cumulative density function is given by:

$$F(P_n) = 1 - e^{-\lambda_3 P_n} \qquad (8)$$

where n = 1..N, N is the number of occurrences of rainy events, and λ_3 is a parameter.

A value of rainfall depth can be generated when the corresponding cumulative density function is known. Solving eq. (8) for P_n:

$$P_n = -\frac{1}{\lambda_3} \ln\left[1 - F(P_n)\right] \qquad (9)$$

Additionally, it should be that:

$$\sum_{n=1}^{N} P_n = P \qquad (10)$$

where P is the monthly rainfall.

The cumulative distribution function $F(P_n)$ can be expressed empirically by the plotting position formulas. Given a set of ordered quantities, a list of the fractions of values at or below each quantity is known as a plotting position (Chow et al., 1994). A general expression for the empirical formulas can be given as:

$$F(P_n) = \frac{n - b}{N + 1 - 2b} \qquad (11)$$

where n is the position of P_n in a ordered set, and b is a parameter that ranges from 0 to 0,5. Two extreme empiric formulations have been analyzed: the Weibull's formulation and the Hazen's formulation.

Weibull's formulation

For the Weibull's formulation b = 0, and combining eq. (9), (10) and (11) the monthly rainfall could be calculated as the addition of N single rainfall depths, as follows:

$$P = -\frac{1}{\lambda_3} \ln\left[\frac{N}{N+1}\right] - \frac{1}{\lambda_3} \ln\left[\frac{N-1}{N+1}\right] - ..... - \frac{1}{\lambda_3} \ln\left[\frac{2}{N+1}\right] - \frac{1}{\lambda_3} \ln\left[\frac{1}{N+1}\right] \qquad (12)$$

Solving for the parameter λ_3:

$$\lambda_3 = -\frac{\sum_{n=1}^{N} \ln\left(\frac{N+1-n}{N+1}\right)}{P} \qquad (13)$$

If the distribution was strictly exponential, the parameter λ_3 should be defined as the inverse of the average depth of rainfall. But this definition is not compatible with (13), consequently, the distribution can be considered as pseudo-exponential.

Hazen formulation

For the Hazen's formulation b = 0.5, and combining eq. (9), (10) and (11) the monthly rainfall could be estimated by addition of N single rainfall depths, as follows:

$$P = -\frac{1}{\lambda_3}\ln\left[\frac{N-0.5}{N}\right] - \frac{1}{\lambda_3}\ln\left[\frac{N-1.5}{N}\right] - ..... - \frac{1}{\lambda_3}\ln\left[\frac{1.5}{N}\right] - \frac{1}{\lambda_3}\ln\left[\frac{0.5}{N}\right] \qquad (14)$$

Solving for the parameter λ_3:

$$\lambda_3 = -\frac{\sum_{n=1}^{N}\ln\left(\frac{N-n+0.5}{N}\right)}{P} \qquad (15)$$

Algorithm to define depth of rainfall events.

- Data available: A series of monthly precipitation, P, and monthly values of number of rainfall events, N.
- Assign parameters of the probabilistic distribution, λ_3, with (13) or (15).
- Generate N rainfall depth applying (9) and (11), where n varies in the interval [1 ..N] and b is defined according to the empiric formulation that is used.

Application of the algorithms to regional rainfall data.

The algorithms were applied in a plain region of Santa Fe province, Argentina; locations are shown in Fig. 1. The climate of the region is moderate and humid, the mean annual temperature is about 17 ºC, and the average annual rainfall is approximately 1,000 mm. The study area is in a flatland region with low slopes of about of 0.6 m/km and areas with natural surface storage and poor surface drainage. Groundwater is recharged mainly from precipitation that infiltrates through unconsolidated Quaternary loess.

Daily rainfall data are available in the study region, and it was possible to contrast the proposed methodology with observed registrations along a series of years and in different places. Four stations of a group of approximately twenty were selected since they present extensive periods of registration, good geographical covering of the region and good quality information. The selected stations and the amount of years of pluviometric registrations were the following: Bombal, 51 years; Chovet, 51 years; Santa Teresa, 52 years and Empalme, 17 years (Fig. 1).

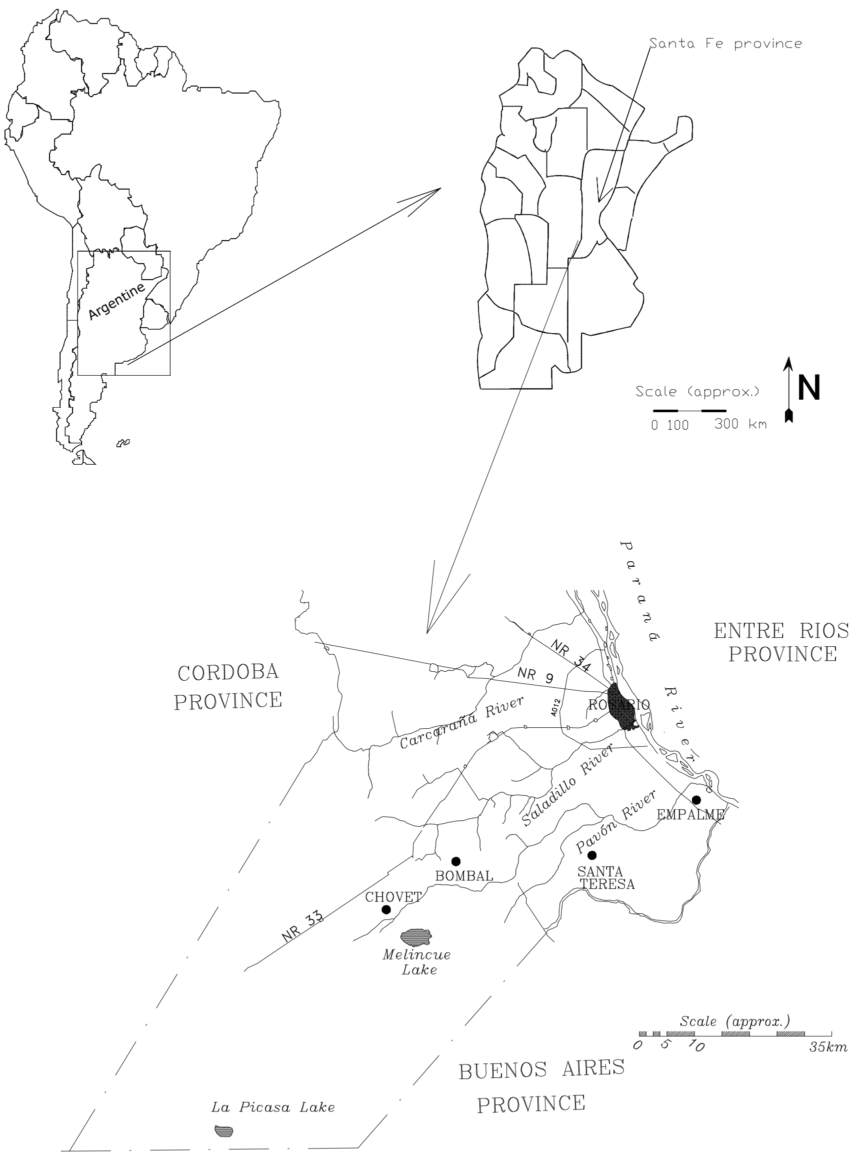

Figure 1. Location of study region and precipitation stations.

For each one of the precipitation stations, daily rainfall data was grouped as monthly rainfall data, P. Average values of monthly rainfall, P_{av}, were calculated. The number of rain occurrences per months, N, was considered for estimating an average number of event, N_{av}, in each month (Zimmermann, 2003). Average values and standard deviations for each station and analyzed variables are shown in Tables 1 and 2.

Table 1. Average (bold) and standard deviation (italic) for rainfall occurrences.

STATION NAME	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	YEAR
Bombal	**5.73**	**4.98**	**5.80**	**4.25**	**3.20**	**2.98**	**2.92**	**2.22**	**3.39**	**5.84**	**5.33**	**5.47**	**52.12**
	2.44	*2.43*	*3.14*	*2.65*	*2.61*	*2.45*	*2.64*	*2.09*	*2.36*	*2.87*	*2.67*	*2.91*	*10.34*
Chovet	**5.29**	**4.90**	**5.75**	**3.98**	**2.88**	**2.59**	**2.16**	**2.20**	**3.29**	**5.57**	**5.65**	**5.55**	**49.80**
	2.43	*2.33*	*2.61*	*2.22*	*2.08*	*1.92*	*1.51*	*2.14*	*1.98*	*2.74*	*2.36*	*2.74*	*9.00*
S.Teresa	**4.23**	**3.62**	**4.48**	**3.42**	**1.79**	**1.75**	**1.94**	**1.96**	**3.08**	**4.23**	**4.40**	**4.38**	**39.29**
	2.08	*2.13*	*2.46*	*1.96*	*1.58*	*1.47*	*1.61*	*1.56*	*2.01*	*2.16*	*1.94*	*2.18*	*6.65*
Empalme	**5.71**	**5.41**	**4.82**	**4.41**	**3.65**	**3.65**	**2.76**	**2.94**	**3.88**	**5.94**	**5.53**	**5.06**	**53.76**
	2.60	*2.62*	*2.32*	*3.32*	*2.98*	*1.50*	*1.89*	*2.99*	*2.20*	*3.40*	*2.53*	*2.70*	*13.90*

Table 2. Average (bold) and standard deviation (italic) for monthly rainfall.

STATION NAME	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	YEAR
Bombal	**109.80**	**100.18**	**143.02**	**77.90**	**41.86**	**31.84**	**40.57**	**28.86**	**55.69**	**98.05**	**84.20**	**110.71**	**922.67**
	63.78	*74.36*	*91.61*	*60.82*	*42.83*	*31.77*	*42.11*	*31.47*	*50.61*	*68.05*	*57.35*	*87.95*	*183.30*
Chovet	**109.72**	**104.31**	**139.29**	**75.94**	**36.47**	**35.10**	**31.90**	**24.14**	**49.73**	**97.98**	**93.78**	**99.39**	**898.75**
	76.00	*64.82*	*93.72*	*53.41*	*31.37*	*40.57*	*36.44*	*27.06*	*40.79*	*70.16*	*51.77*	*78.35*	*182.96*
S.Teresa	**105.35**	**106.79**	**137.15**	**78.29**	**43.38**	**35.71**	**37.71**	**30.87**	**61.81**	**94.81**	**92.23**	**103.31**	**927.40**
	69.78	*87.40*	*92.21*	*51.28*	*43.73*	*37.72*	*36.72*	*28.15*	*60.94*	*65.72*	*49.97*	*82.55*	*189.92*
Empalme	**134.44**	**160.47**	**111.59**	**73.94**	**48.82**	**54.47**	**43.62**	**43.71**	**85.76**	**117.97**	**99.44**	**122.41**	**1096.65**
	106.25	*121.63*	*74.39*	*58.91*	*51.87*	*40.51*	*41.71*	*46.26*	*64.62*	*79.00*	*60.95*	*104.59*	*298.11*

A similar behavior could be observed in the precipitation stations respect to the occurrences of rainy events. The variation coefficients ranged from 40% to 100%, with an important dispersion around the average values, especially for winter months. This descriptive aspect of the analyzed samples evidences the difficulty of predicting the number of rainfall occurrences if it was considered as an independent random variable. A forecast of N using random generation techniques (e.g. techniques based on Monte Carlo simulations), would be in a wide range of values for N. If the probability distribution function of N is conditional upon the monthly rainfall P degrees of freedom to the variable can be restricted allowing better forecast.

Concerning to the monthly precipitation, its variability is greater than the one registered for the rain occurrences but this variable represents inputs in a model.

Estimation of number of rainfall events

The proposed bayesian approach (eq. 7) was applied to the monthly series of precipitation of each station. The results obtained are presented in Table 3 and Figure 2.

Results show a general tendency to underestimate the number of events (Table 3). This can also be observed in Figure 2, since the fitted lines to the computations (dashed lines) have less slope than the coincidence lines (continuous lines).

Table 3. Results achieved for forecasting of number of rainfall occurrences

STATION NAME	N OBSERVED	N CALCULATED	CORRELATION COEFFICIENT
Bombal	2658	2410	0.769
Santa Teresa	2043	1874	0.757
Chovet	2540	2343	0.812
Empalme	910	854	0.674

Dispersion around the coincidence line indicates that the monthly precipitation is not the only variable that controls the number of rainfall events. There are complex climatological factors that command the studied variable and that, necessarily, they are outside of the proposed algorithm. Nevertheless, keeping in mind the simplicity of the proposal, and due to the frequently available information the results could be considered as satisfactory.

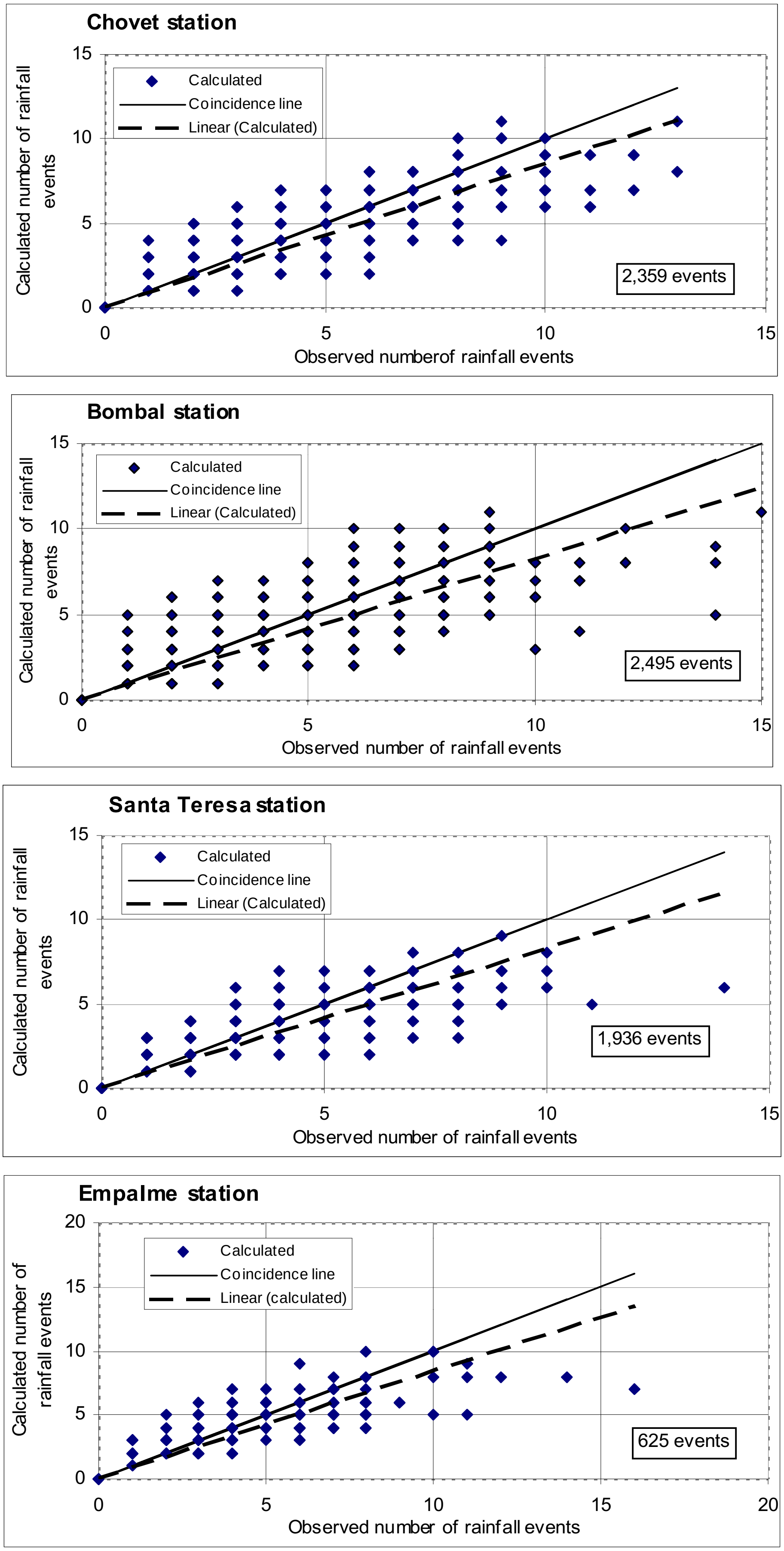

Figure 2. Forecasting of number of rainfall events for study stations of Santa Fe province.

Estimation of depths of rainfall events

Depths of rainfall depths for each station and monthly series of precipitation were estimated applying the algorithm proposed, and considering the number of events observed N. The procedure was applied for the two formulations of empirical frequencies. Series of calculated rainfall depths were compared with the observed ones, and correlation coefficients are shown in Table 4.
The obtained results indicated that the formulation of Hazen presented a better performance to forecast rainfall depths.

Table 4. Correlation coefficients achieved for forecasting of rainfall depths

STATION NAME	WEIBULL'S FORMULATION	HAZEN'S FORMULATION
Bombal	0.892	0.910
Chovet	0.886	0.906
Santa Teresa	0.907	0.907
Empalme	0.914	0.904

Results using Hazen's formulation are shown in Figure 3. As can be observed, the fitted lines to the computations (dashed lines) approach very good to the coincidence lines (continuous lines). In this way, the average tendency of the forecasts is very satisfactory.

Monthly runoff forecasting

The utility of the proposed methodology could be evaluated when monthly runoff calculated using the observed rainfall is compared with those monthly runoff calculated with the predicted rainfall using both algorithms.
To perform this evaluation, the US-SCS method was applied to the observed records of daily rainfall and to the calculated rainfall events, which were synthetically generated. Runoffs calculated for each event were cumulated to monthly values. Finally, monthly runoffs estimated by both procedures were compared (Table 5).

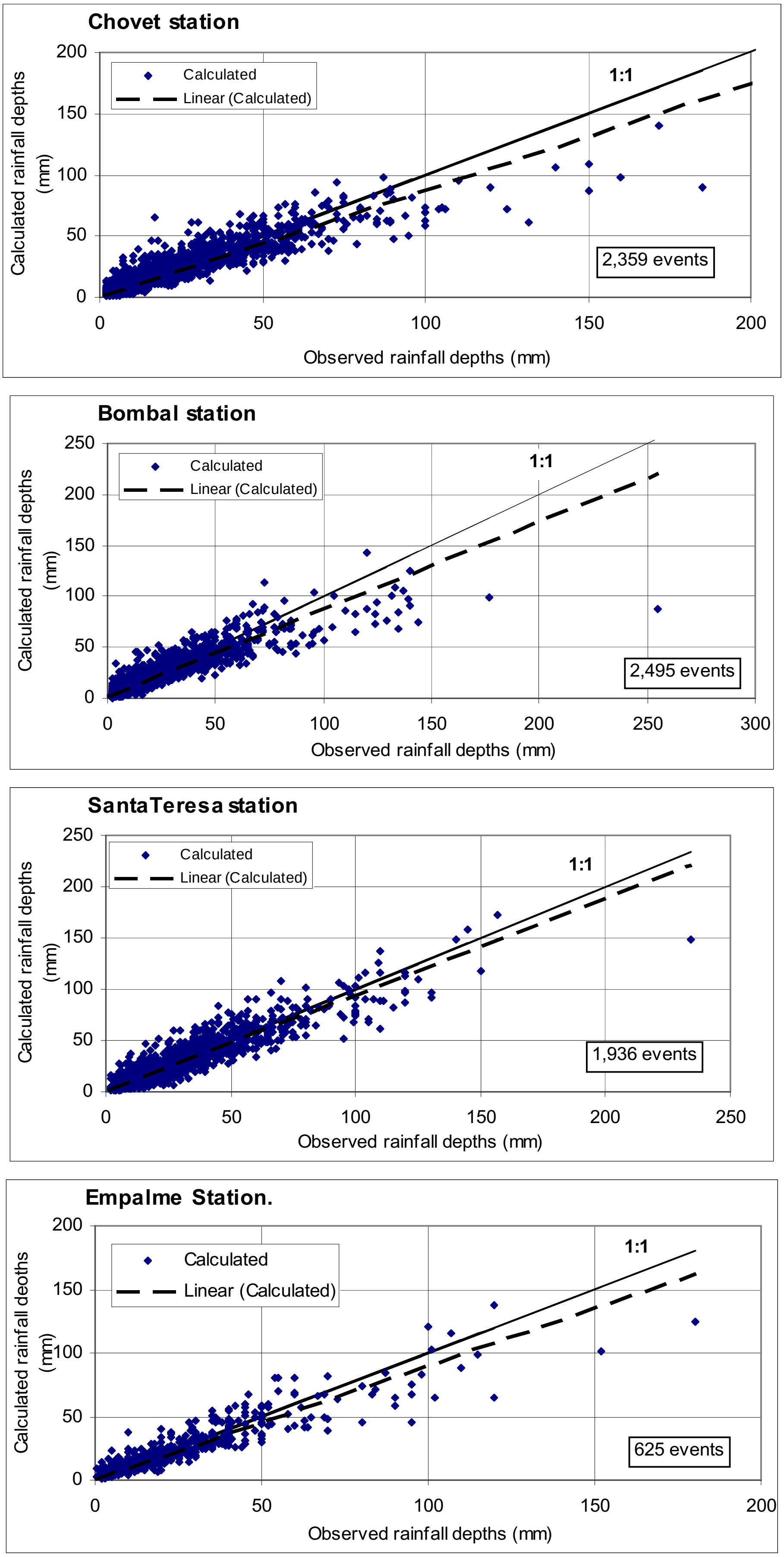

Figure 3. Forecasting of rainfall depths for study stations of Santa Fe province.

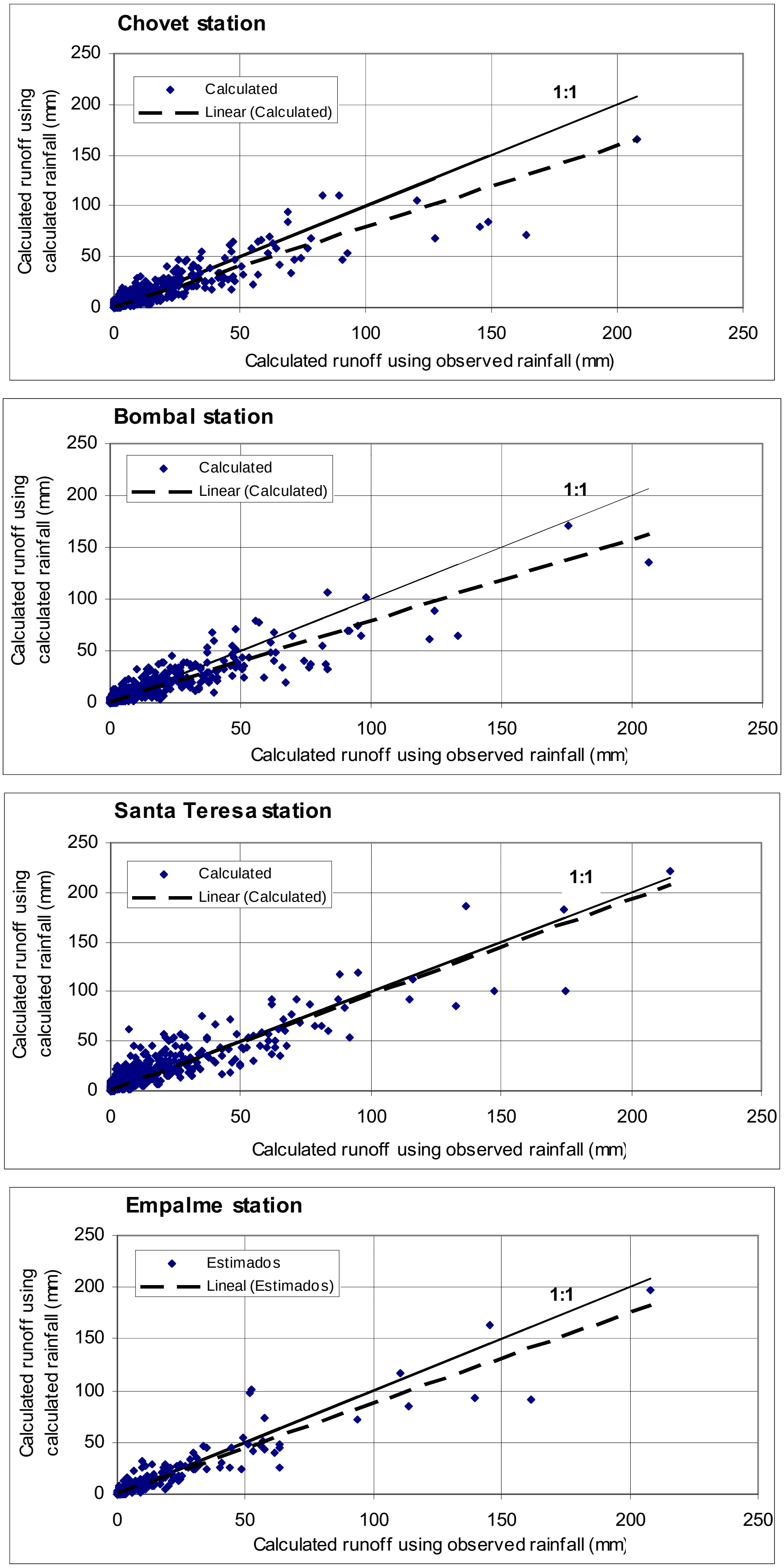

Figure 4. Forecasting of monthly runoff for study stations of Santa Fe province.

Table 5. Correlation coefficients achieved for forecasting of monthly runoff

STATION NAME	HAZEN'S FORMULATION CN = 78
Bombal	0.893
Chovet	0.892
Santa Teresa	0.898
Empalme	0.928
Annual runoff for all stations	*0.807*

Figure 4 show runoffs generated from observed rains and from calculated rains in each precipitation station. The value of curve number (CN) used was 78, which is quite representative of soil conditions, vegetation covering and average moisture in the region.

Figure 5 shows a forecast of annual runoff for all stations. It can be observed that the fitted line to the computations (dashed lines) approaches very good to the coincidence line (continuous line).

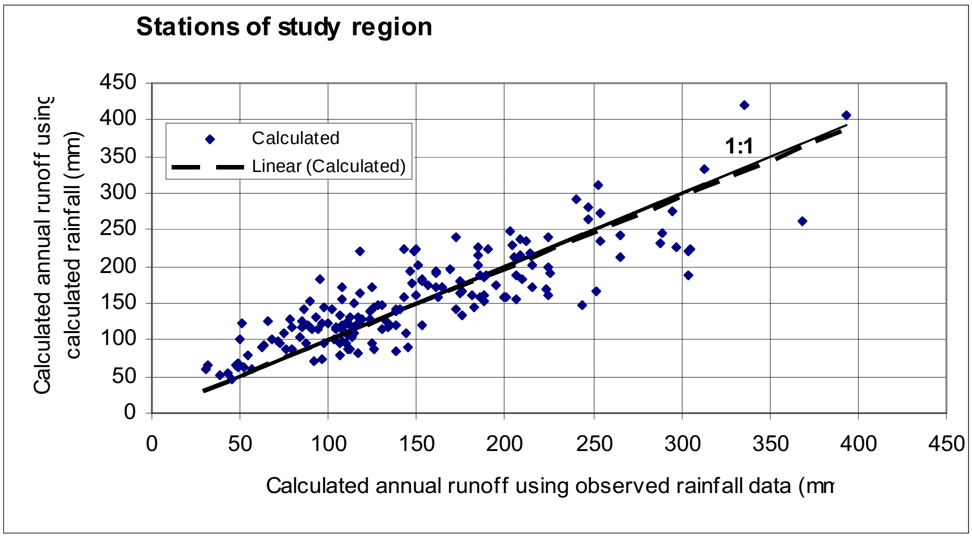

Figure 5. Forecasting of annual runoff for all stations.

Conclusions

The water budget of Thornthwaite and Mather, applied jointly with the CN-method of the US-SCS, allows estimating the net infiltration, which can be assimilated to the recharge of phreatic aquifer. To apply the CN-method together with the water budget in a long-term, it is necessary to know the number of monthly rainfall occurrences, N, and the depth of rainfall for each event, information that frequently is not available.

A daily rainfall model was proposed here, which is derived into two steps: (a) derivation of the probability distribution of the number of rainy events for a given month conditional upon the monthly rainfall, and (b) derivation of the probability distribution function of the rainfall depth for each single event. A combination of an Erlang's pdf with a Poisson pdf was proposed for the first step. A pseudo-exponential pdf was proposed for the second step.

Both algorithms were applied in four rainfall stations of the southern region of Santa Fe (Argentina). More than 7600 forecasts of rainfall occurrences and depths were compared with the observed registrations.

It could be observed that the monthly precipitation is not the only variable that controls the number of rainfall events. There are complex climatological factors that command the studied variable and that, necessarily, they are outside of the proposed algorithm. Nevertheless, keeping in mind the simplicity of the model, and the available information the results were adequate.

To evaluate the performance of the proposed methodologies, monthly runoff calculated using CN-method and observed rainfall records were compared with monthly runoff calculated using predicted rainfalls. The results for all the pluviometric stations were satisfactory.

The methodologies proposed allow (a) the correct application of the Thornthwaite's water budget together with the CN-method, and (b) a good prediction of monthly runoff.

In order to validate the water budget model, calculations of net infiltration should be contrasted with other methodologies to estimate direct recharges to the phreatic aquifers.

References

Antigüedad I, García Muñiz J, Llamas J. (1995) A mathematical model for precipitation in the Basque Country, Spain. Hydrological Sciences Journal. 40 (3): 291-301. IAHS Press. Wallingford. UK.

Arnaud P, Lavabre J. (1999) Using a stochastic model for generating hourly hyetographs to study extreme rainfalls. Hydrological Sciences Journal. 44 (3): 433-445. IAHS Press. Wallingford. UK.

Cox DR, Isham V. (1994). Stochastic models of precipitation. In: Barnett V and Turkman K (eds) Statistics for the environment, Water related issues. 3-18. Wiley, Chichester.
Chow VT, Maidment D, Mays L. (1994) Hidrología Aplicada [Applied Hydrology]. Mc Graw-Hill Interamericana SA. Colombia.
Eagleson PS (1972) Dynamics of floods frequency. Water Resources Research. 8 (4): 878-898.
Montgomery D, Runger G (1996). Probabilidad y Estadística Aplicadas a la Ingeniería [Applied Statistics and Probability for Engineers]. Mc Graw-Hill. México.
Riccardi G, Zimmermann E. (2000) Modelo Probabilístico de Tormentas Puntuales Aplicado a Rosario Aero (Santa Fe, Argentina) [Punctual Probabilistic Storm Model Applied to Rosario Aero Station (Santa Fe province, Argentina)] In: XVIII Congreso Nacional del Agua.Termas de Río Hondo, Argentina.
Scanlon B, Healy R, Cook P. (2002) Choosing appropriate techniques for quantifying groundwater recharge. Hydrogeology Journal.10 (1): 18-39. Springer-Verlag. Germany.
Scozzafava M, Tallini M. (2001) Net Infiltration in the Gran Sasso Massif of central Italy using the Thornthwaite water budget and curve-number method. Hydrogeology Journal. 9 (5): 461-475. Springer-Verlag. Germany.
Seoane R, Valdés J. (1994) Desarrollo de una Función de Distribución Derivada para la Estimación de Caudales Extremos en Cuencas con Datos Escasos. [Development of a Derived Distribution Function for Estimation of Extreme Flows in Catchments with Scarce Data] XVI Congreso Latinoamericano de Hidráulica. (2): 193-204. IARH. Santiago de Chile. Chile.
Thornthwaite CW, Mather JR (1957) Instructions and tables for computing potential evapotranspiration and the water balance, 5th printing. 10 (3), CW Thornthwaite Associates, Laboratory of Climatology , Elmer, NJ, USA.
USDA-SCS (1986) National Engineering Handbook, Sect 4, Hydrology. US Department of Agriculture-Soil Conservation Service, Washington, DC, USA.
Zimmermann E, Navarro R, Silber M. (1996) Un Modelo Probabilístico para la Generación de Tormentas Sintéticas [A Probabilistic Model for the Generation of Synthetic Storms] XVII Congreso Latinoamericano de Hidráulica. IAHR. Guayaquil, Ecuador.