

Transfer function modeling of the monthly accumulated rainfall series over the Iberian Peninsula

VIDAL L. MATEOS, JOSE A. GARCIA, ANTONIO SERRANO
and MARIA DE LA CRUZ GALLEGO

Dpto. de Física, Universidad de Extremadura, 06071 Badajoz, Spain
(Manuscript received June 12, 2001; accepted in final form April 3, 2002)

RESUMEN

Con el objeto de mejorar los resultados proporcionados por los modelos Autorregresivo - Media Móvil (ARMA) ajustados a las precipitaciones mensuales acumuladas registradas en 19 observatorios de la Península Ibérica se han usado modelos de función de transferencia (DLTFN) en los que se han empleado como variable independiente la presión local (LP), la presión a nivel del mar (SLP) o la temperatura de agua del mar (SST) en el Atlántico Norte.

En todos los casos analizados, los resultados obtenidos con los modelos DLTFN, medidos mediante la varianza explicada por el modelo, han sido mejores que los resultados proporcionados por los modelos ARMA. Los mejores resultados han sido dados por aquellos modelos que usan la presión local como variable de entrada, seguidos por los modelos que emplean la presión a nivel del mar, siendo los modelos que emplean la temperatura del agua del mar los que peores resultados proporcionan.

Con referencia a la situación geográfica, los modelos ajustados a los observatorios localizados al oeste de la Península dan mejores resultados que los ajustados a los observatorios situados al norte y este de la misma. Asimismo se ha encontrado que existe una región en el Atlántico Norte situada entre los paralelos 0° y 20° N donde la temperatura del agua parece tener alguna influencia sobre las precipitaciones peninsulares. Esta región se traslada hacia el norte cuando se emplean presiones a nivel del mar, a la vez que se intensifica su influencia.

ABSTRACT

In order to improve the results given by Autoregressive Moving-Average (ARMA) modeling for the monthly accumulated rainfall series taken at 19 observatories of the Iberian Peninsula, a Discrete Linear Transfer Function Noise (DLTFN) model was applied taking the local pressure series (LP), North Atlantic sea level pressure series (SLP) and North Atlantic sea surface temperature (SST) as input variables, and the rainfall series as the output series. In all cases, the performance of the DLTFN models, measured by the explained variance of the rainfall series, is better than the performance given by the ARMA modeling. The best performance is given by the models which take the local pressure as the input variable, followed by the sea level pressure models and the sea surface temperature models. Geographically speaking, the models fitted to those observatories located in the west of the Iberian Peninsula work better than those on the north and east of the Peninsula. Also, it was found that there is a region located between 0° N and 20° N, which shows the highest cross-correlation between SST and the peninsula rainfalls. This region moves to the west and northwest off the Peninsula when the SLP series are used.

Keywords: Autoregressive - Moving Average, Discrete Linear Transfer Function Noise, Rainfall Modeling.

1. Introduction

Autoregressive Moving Average (ARMA) models have frequently been used to parameterize accumulated monthly rainfall series (Delleur and Kavvas, 1978; Katz and Skaggs, 1981) and specifically to the Iberian Peninsula by Garrido and García (1993). In these models the rainfall series is regressed over itself lagged behind in time plus random noise. In an attempt to improve the results given by the ARMA models, Discrete Linear Transfer Function Noise models (DLTFN) were used (Box and Jenkins, 1976). There supposedly is a dynamic system where some variables (the predictors) act as inputs of this kind of model's system, and others (the predictands) as outputs, and that a change in the level of the input variables will produce a change in the output variables, though not necessarily at the same time. This implies a causal relation, in a statistical sense, between the input and output of the system. In the present paper the accumulated monthly rainfall series is taken as the output variable. For the input variables, one of the following is selected: monthly mean sea surface temperature (SST), which could be taken as a measure of the exchange of Sensible and Latent Heat between the ocean and the atmosphere, monthly mean sea level pressure (SLP), and monthly mean local station surface pressure (LP) (pressure measured at the same location as the accumulated monthly rainfall). The election of these variables was determined by the fact that according to Linés (1981) a good deal of the rainfall over the Iberian Peninsula is due to frontal systems coming from the North Atlantic Ocean, so that one might expect that a relation between the rainfall series measured over the Iberian Peninsula and the SST and SLP series over the North Atlantic Ocean could be established. The existence of such teleconnections between SST and SLP over the North Atlantic Ocean and large scale weather over Europe has been demonstrated by, amongst others, Ratcliffe and Murray (1970), Meehl and van Loon (1979), Barnett (1984). More specifically, the relation between SST and rainfall series has been studied, for example, by Nicholls (1989) in Australia, by Hastenrath *et al.* (1995) in South Africa, and in the Sahel by Folland *et al.* (1986), and Fontaine *et al.* (1995).

2. Data

The data to be analyzed in this study can be divided into two groups:

a. Continental data

Table 1. Observatory, geographic location in the Iberian peninsula, period spanned by each series, mean (μ) and standard deviation (σ_0) of the original accumulated monthly rainfall series, expressed in mm, mean (μ_1) and standard deviation (σ_1) of the root squared monthly rainfall series, standard deviation (σ_2) after removing the long-term trend and yearly cycle.

Observatory	Geographic Location	Period	μ_0	σ_0	μ_1	σ_1	σ_2
Badajoz	38°53' N, 6°58' W	1901-79	42.8	44.1	5.50	3.55	2.79
Barcelona	41°24' N, 2°9' E	1901-79	48.8	45.7	6.26	3.10	2.91
Coimbra	40°12' N, 8°25' W	1901-79	83.0	74.5	8.12	4.12	3.17
La Coruña	43°22' N, 8°25' W	1901-79	75.8	57.5	8.05	3.31	2.74
Evora	38°34' N, 7°54' W	1901-79	53.8	56.5	6.15	3.99	2.98
San Fernando	36°28' N, 6°12' W	1901-79	47.4	55.1	5.53	4.09	2.90
Huelva	37°16' N, 6°57' W	1903-79	40.1	48.6	4.99	3.90	2.90
Lisboa	38°43' N, 9°8' W	1901-79	60.8	62.2	6.60	4.15	2.94
Madrid	40°25' N, 3°41' W	1901-79	37.2	32.8	5.41	2.81	2.45
Málaga	36°43' N, 4°25' W	1906-79	42.0	54.4	5.02	4.09	3.14
Oporto	41°11' N, 8°36' W	1901-79	99.4	88.2	8.92	4.46	3.43
Salamanca	40°58' N, 5°41' W	1901-79	34.5	29.5	5.29	2.61	2.33
Santander	43°28' N, 3°49' W	1912-79	101.8	66.5	9.56	3.22	2.87
Sevilla	37°23' N, 5°59' W	1901-79	48.2	56.1	5.57	4.15	3.13
Soria	41°46' N, 2°28' W	1901-79	46.7	33.6	6.46	2.50	2.35
Valencia	39°28' N, 0°22' W	1901-79	36.2	46.5	5.01	3.33	3.05
Valladolid	41°39' N, 4°43' W	1901-79	35.2	29.5	5.34	2.77	2.33
Vitoria	42°51' N, 2°40' W	1919-79	72.5	46.3	8.05	2.76	2.49
Zaragoza	41°39' N, 0°53' W	1905-79	28.3	27.5	4.73	2.41	2.29

These data account for the accumulated monthly rainfall series measured at 19 observatories throughout the Iberian Peninsula, whose geographic situation is given in Table I and displayed in Figure 1. It also indicates the mean monthly pressure series measured at those observatories. This set of observatories coincides practically with that used by Lorente (1985) and Linés (1981), and could be considered representative of the Iberian Peninsula. Most of the temporal series extend from 1901 to 1979.

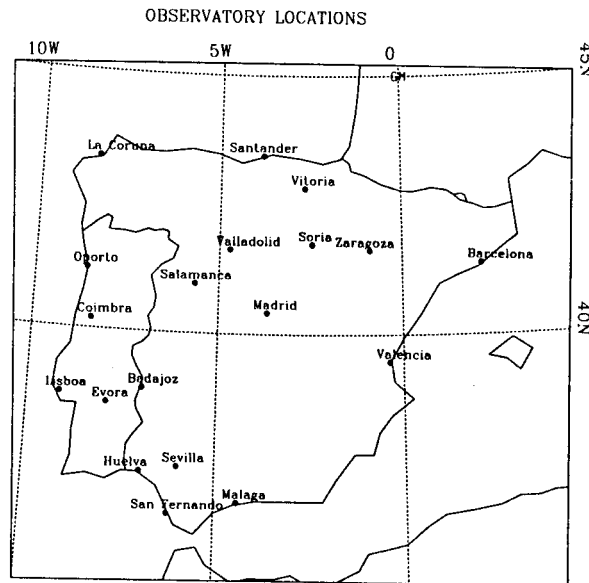


Fig. 1. Locations in the Iberian Peninsula of the rainfall observatories used in our study.

b. Ocean data

These data correspond to SST and SLP, and were obtained from groups 3 and 4 of the COADS (Comprehensive Ocean-Atmosphere Data Set) produced by the National Climatic Data Center of the USA (Woodruff *et al.*, 1987). Both groups average over 10 degree latitude-longitude boxes in the region 0° N to 70° N and 10° E to 70° W. Figure 2 shows the box labels as used by COADS. The period spanned by these series also run from 1901 to 1979. The SST and SLP series were selected for subsequent analysis for each observatory in the Peninsula. These series, once deseasonalized and linear detrended, have the highest cross-correlation coefficient with each rainfall series square root transformed, deseasonalized and linear detrended (see below).

Tables 2 and 3 present the highest correlation coefficient of each rainfall series with the SST and SLP series respectively, along with their standard errors, estimated as ± 1.96 times the lag-zero cross-correlation coefficient variance, which is given by Bennett (1979).

Table 2. Highest lag-zero cross-correlation coefficient and its standard error between each rainfall series and the SST series. Also listed is the label of the box with which the maximum value is obtained and the significance level of the cross-correlation coefficient.

Observatory	Cross-Correlation Coefficient	Standard Error	Label	Confidence Level
Budajoz	0.125	0.065	248	99.9
Barcelona	0.076	0.056	318	98.7
Coimbra	0.125	0.060	244	99.9
La Coruña	0.186	0.095	245	100.
Evora	0.141	0.059	248	100.
San Fernando	0.138	0.057	248	99.9
Huelva	0.175	0.061	248	100.
Lisboa	0.173	0.061	248	100.
Madrid	0.157	0.057	248	100.
Málaga	0.102	0.066	248	99.7
Oporto	0.152	0.061	244	99.9
Salamanca	0.133	0.064	215	99.9
Santander	-0.102	0.066	141	99.7
Sevilla	0.141	0.057	248	100.
Soria	0.097	0.061	248	99.8
Valencia	0.084	0.061	281	98.9
Valladolid	0.094	0.067	248	99.6
Vitoria	-0.135	0.074	141	99.9
Zaragoza	0.101	0.062	248	99.8

Table 3. Highest lag-zero cross-correlation coefficient and its standard error between each rainfall series and the SLP. Also listed is the label of the box with which the maximum value is obtained and the significance level of the cross-correlation coefficient.

Observatory	Cross-Correlation Coefficient	Standard error	Label	Confidence Level
Budajoz	-0.41	0.05	176	99.9
Barcelona	-0.21	0.06	214	99.9
Coimbra	-0.54	0.04	177	99.9
La Coruña	-0.50	0.04	177	99.9
Evora	-0.46	0.04	176	99.9
San Fernando	-0.43	0.04	212	99.9
Huelva	-0.47	0.04	212	99.9
Lisbon	-0.49	0.04	176	99.9
Madrid	-0.37	0.05	212	99.9
Málaga	-0.36	0.05	212	99.9
Oporto	-0.55	0.03	177	99.9
Salamanca	-0.38	0.05	177	99.9
Santander	0.23	0.04	174	99.9
Sevilla	-0.45	0.04	212	99.9
Soria	-0.39	0.05	177	99.9
Valencia	-0.15	0.06	212	99.9
Valladolid	-0.42	0.04	176	99.9
Vitoria	0.18	0.05	174	99.9
Zaragoza	-0.23	0.06	213	99.9

10 DEGREE BOXES

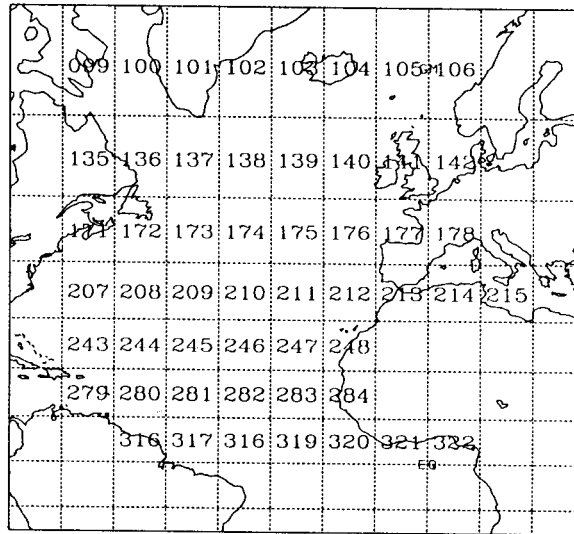


Fig. 2. 10 degree box labels as used by COADS

$$var[R(0)] = \frac{1}{T} \sum_{k=-\infty}^{\infty} [r_{xx}(k)r_{yy}(k) + r_{xy}(k)r_{yx}(k) + r_{yx}^2(0)\{r_{yx}^2(k) + \frac{1}{2}r_{xx}^2(k)\} - 2r_{yx}(0)\{r_{xx}(k)r_{yx}(k) + r_{yy}(-k)r_{yx}(k)\}],$$

Table 4. Model orders for local pressure/rainfall.

Observatory	r, s	p, q	Residual Variance
Budajoz	0, 0	1, 1	0.63
Barcelona	0, 0	3, 2	0.85
Coimbra	0, 0	0, 0	0.61
La Coruña	1, 1	0, 0	0.66
Evora	0, 0	1, 1	0.63
S. Fernando	1, 0	1, 1	0.63
Huelva	1, 0	1, 1	0.73
Lisboa	2, 1	1, 1	0.57
Madrid	0, 1	2, 2	0.72
Málaga	1, 1	0, 1	0.78
Oporto	0, 0	1, 1	0.65
Salamanca	0, 0	0, 0	0.74
Santander	1, 1	0, 1	0.89
Sevilla	0, 0	1, 1	0.63
Soria	1, 0	2, 2	0.72
Valencia	0, 0	0, 0	0.97
Valladolid	0, 0	1, 1	0.64
Vitoria	0, 2	2, 2	0.83
Zaragoza	1, 1	2, 1	0.86

where r_{xx} is the autocorrelation coefficient at lag k for the input series (SLP or SST), r_{yy} is the autocorrelation coefficient at lag k for the output series (rainfall series) and r_{xy} is the cross-correlation coefficient at lag k between the input and the output series. T is the period spanned by the series. The above sum was truncated at $T/4$.

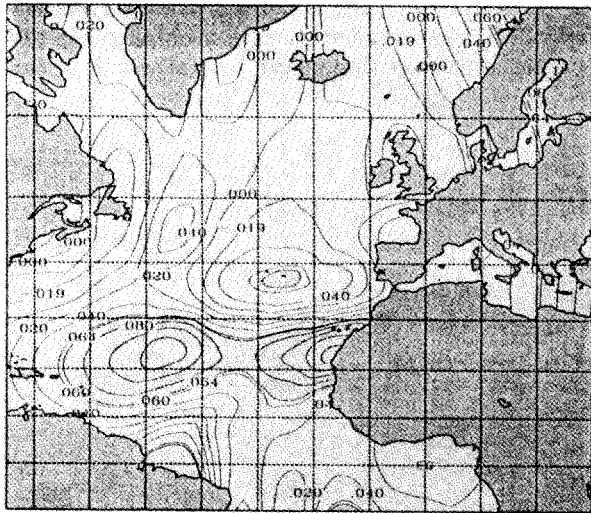
Table 5. Model orders for SLP/rainfall.

Observatory	r, s	p, q	Residual Variance
Badajoz	0, 0	1, 1	0.82
Barcelona	0, 0	0, 0	0.94
Coimbra	0, 2	0, 0	0.69
La Coruña	0, 0	1, 1	0.71
Evora	1, 1	0, 0	0.77
S. Fernando	0, 0	0, 0	0.80
Huelva	0, 0	0, 0	0.77
Lisboa	0, 0	0, 0	0.75
Madrid	0, 0	0, 0	0.85
Málaga	0, 0	0, 0	0.86
Oporto	0, 0	0, 0	0.68
Salamanca	0, 0	1, 1	0.83
Santander	0, 0	0, 0	0.80
Sevilla	1, 1	1, 1	0.78
Soria	0, 1	0, 0	0.83
Valencia	0, 0	3, 3	0.91
Valladolid	0, 0	2, 2	0.80
Vitoria	0, 0	3, 3	0.86
Zaragoza	0, 1	0, 0	0.93

The labels of the boxes with which the highest cross-correlation value is obtained and the significance level, the probability that the cross-correlation coefficient is significantly different from zero, determined using a bootstrap procedure (Efron, 1993; Kahl *et al.*, 1993) are also indicated. As observed in the SST table, except for Barcelona, Santander, Vitoria and Salamanca, the most repeated box is # 248 (see Figure 2 for the location of the boxes). To further analyze the relationship between the SST series and the rainfall series, a contour plot of the cross-correlation between each rainfall series and all the SST series was made. Figure 3 shows four cases (Evora, Madrid, Santander and Valencia) out of the nineteen contour plots obtained. As suggested by the Evora and Madrid plots, there is a belt from 20°N to 30°N and 10°W to 70°W that presents, however, significant positive cross-correlations between the SST and the the rainfall series. The results obtained for these two observatories are typical examples of those located in the west and southwest of the Peninsula. This is more clearly seen in Figure 2, which is a contour-plot of the cross-correlations between box # 248 and all of the rainfall series. Also, it was found that for some observatories (Coimbra, Evora, S. Fernando, Lisboa, Malaga, Oporto, Sevilla and Soria) there is a small area north of the positive cross-correlation region, with significant negative cross-correlation, leading to a kind of dipole with the positive cross-correlation region. An analysis of the cross-correlations between the rainfall series and the SST series formed as the difference between the positive and the negative region do not seem to improve the results obtained with the SST themselves. Hence, the SST series of the box with the maximum

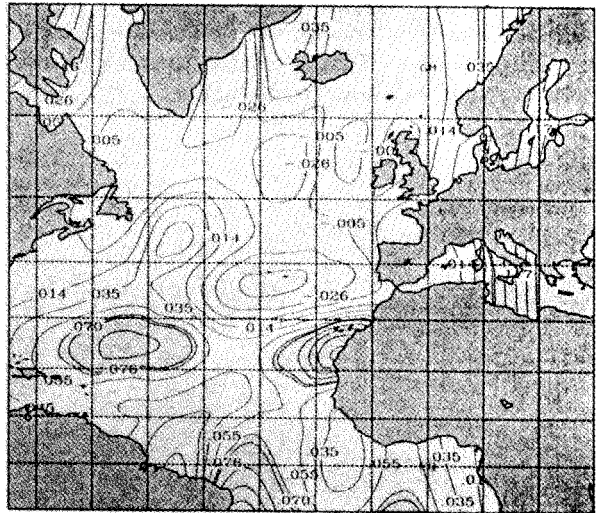
cross-correlations with the rainfall series will be used in subsequent analysis. As shown by Rowtree (1976), using a hemispheric model, a tropical ocean temperature anomaly causes a fall in surface pressure which extends well north of 60°N, which in turn induces an increase in rainfall at middle latitudes. These theoretical results match with our finding of a relationship between box # 248 and the rainfall over the Iberian Peninsula.

CORRELATIONS SST/EVORA



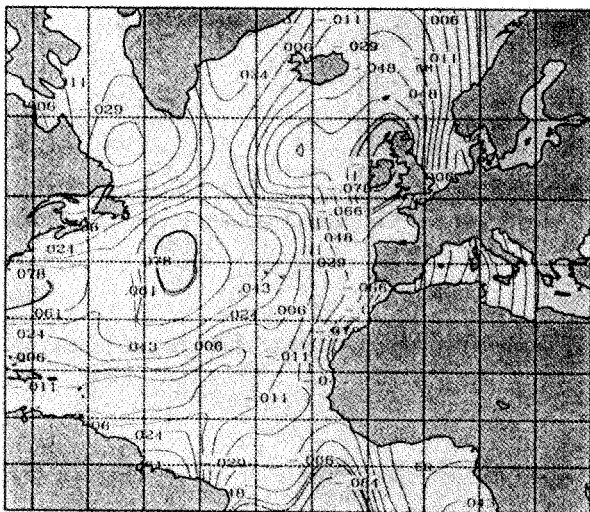
CONTOUR FROM - 004 TO 141 BY 020

CORRELATIONS SST/MADRID



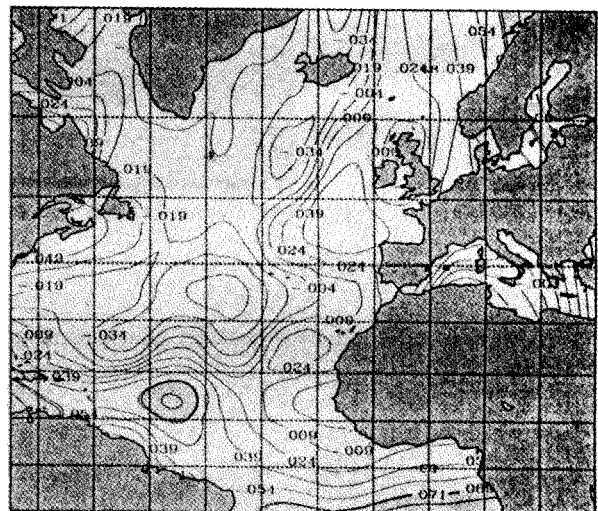
CONTOUR FROM - 070 TO 156 BY 020

CORRELATIONS SST/SANTANDER



CONTOUR FROM - 103 TO 079 BY 010

CORRELATIONS SST/VALENCIA



CONTOUR FROM - 071 TO 084 BY 014

Fig. 3. Maps of the cross-correlation between the SST and four observatories in the peninsula. Shaded regions over the ocean are those with significant cross-correlation coefficients.

With reference to the SLP table, except for Barcelona, Santander, Vitoria and Zaragoza, boxes with greater cross-correlations are adjacent to the north, west and southwest of the Iberian Peninsula (boxes # 176, 177 and 212). In this case, the cross-correlation coefficients are notably higher than those obtained for the SST series, and they are negative, which means that the highest precipitations are obtained with low

pressure centers situated in or near these boxes. As before, we made contour plots of the cross-correlation between each rainfall series and the SLP series, Figure 5 shows four examples of the contour plots obtained. The results for Evora and Madrid are representative of those obtained at observatories located in western and southwestern Spain. As the figure shows, the region in boxes # 176 and 212 presents the highest (negative) cross-correlation values. The relation between the rainfall over the Peninsula and the SLP over the above mentioned boxes is more clearly seen in Figure 6, which shows a contour plot of the cross-correlations between the average of the SLP boxes # 176, 177 and 212 with the rainfall series. It should be noted that, for the Santander observatory (and for Vitoria, not shown), the maximum cross-correlation region is a positive cross-correlation one, instead of the negative cross-correlation region obtained at the other observatories. This is related to most favorable synoptic situations that cause rainfall north of the Peninsula, as it will later be discussed

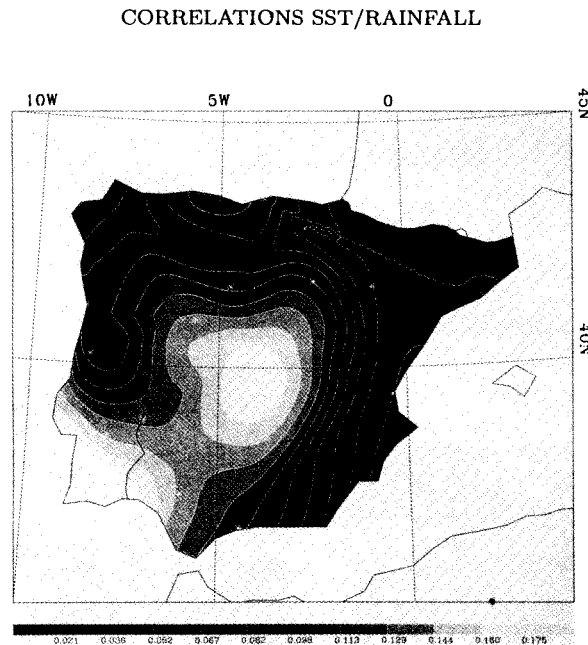
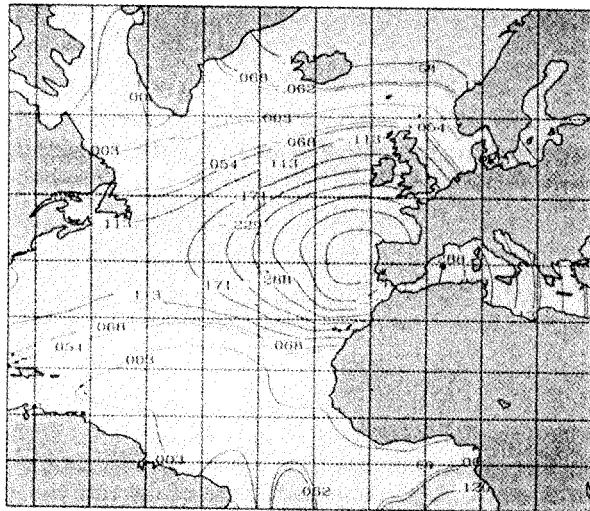


Fig. 4. Contour-shaded plot of the cross-correlation between SST box # 244 and the rainfall at each observatory in the peninsula.

Prior to their use in DLTFN models, all the series were: *i*) Deseasonalized, by fitting them to the following function:

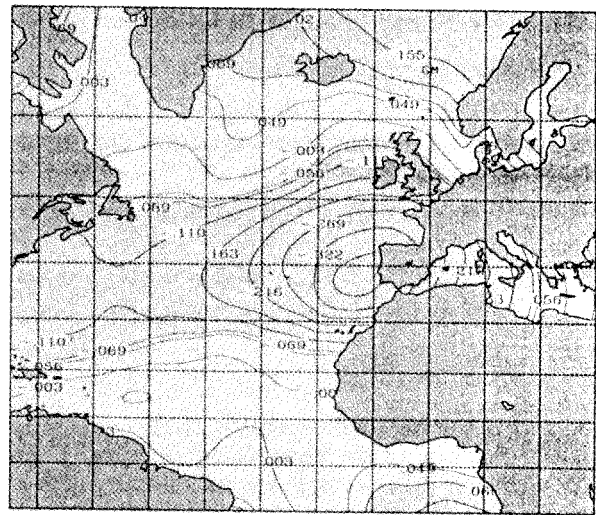
$$y(t) = A_0 + \sum_{j=1}^4 A_j \sin(2\pi jt/365) + B_j \cos(2\pi jt/365) + y_r(t), \quad (1)$$

CORRELATIONS SLP/EVORA



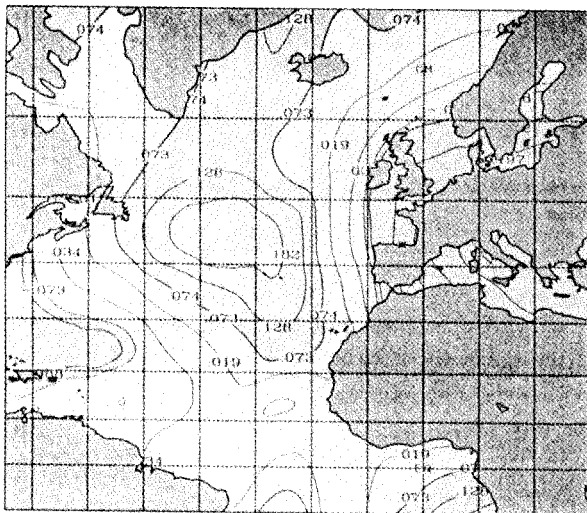
CONTOUR FROM - .463 TO .12 BY .058

CORRELATIONS SLP/MADRID



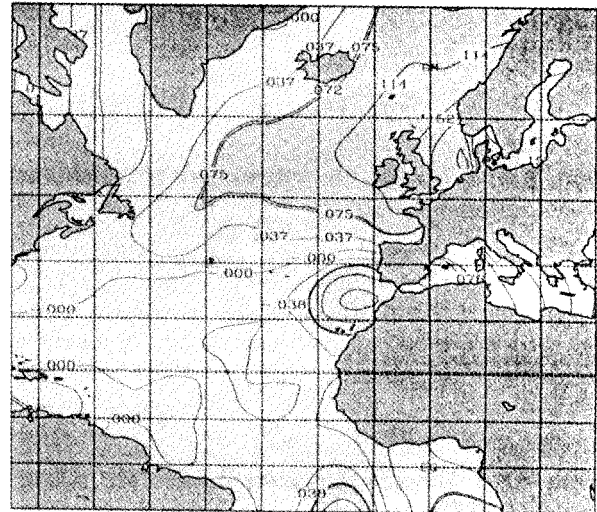
CONTOUR FROM - .375 TO .155 BY .053

CORRELATIONS SLP/SANTANDER



CONTOUR FROM - .305 TO .236 BY .054

CORRELATIONS SLP/VALENCIA



CONTOUR FROM - .153 TO .228 BY .038

Fig. 5. Maps of the cross-correlation between the SLP and four observatories in the peninsula. Shaded regions over the ocean are those with significant cross-correlation coefficients.

where only the yearly cycle and its first three harmonics have been included. No other period, apart from the annual, was found to be significant (Garrido and García, 1992). In the case of accumulated rainfall series, they were previously square root transformed in order to get an approximately normally distributed series (see Box and Cox, 1964; Delleur and Kavvas, 1978). Also, some peninsular pressure series were corrected for changes in location. *ii*) Detrended, by submitting each series to a homogeneity analysis (Mateos, 1993). In general, a linear trend was found to be enough to make the series stationary. *iii*) Finally, all the series were standardized, dividing them by their standard deviation. Table 1 shows the mean (μ) and the standard deviation (σ_0) of the original rainfall series, the mean (μ_1) and the standard deviation (σ_1) of the square

root-transformed rainfall series and the standard deviation (σ_2) of the deseasonalized series.

3. Discrete linear transfer function models

a. Definition

Let $y(t)$ and $x(t)$ be a pair of series of data available at equally spaced intervals of time Δt , taken usually as 1, that could be considered as output and input of some dynamic system. $y(t)$ and $x(t)$ are supposedly related through a linear filter:

$$\begin{aligned} y(t) &= v_0x(t) + v_1x(t-1) + v_2x(t-2) + \dots = \\ &= v_0x(t) + v_1Bx(t) + v_2B^2x(t) + \dots = \\ &= V(B)x(t) \end{aligned} \quad (2)$$

where B is the backshift operator $Bx(t) = x(t-1)$. The polynomial $V(B) = v_0 + v_1B + v_2B^2 + \dots$ is the transfer function of the filter. Weights v_0, v_1, \dots are called the input response of the system. The problem with equation (2) is that it has an infinite number of parameters. A parsimonious class of discrete linear transfer function models is given by (Box and Jenkins, 1976):

$$(1 - \delta_1B - \delta_2B^2 - \dots - \delta_rB^r)y(t) = (\omega_0 - \omega_1B - \omega_2B^2 - \dots - \omega_sB^s)x(t), \quad (3)$$

or

$$\delta(B)y(t) = \omega(B)x(t), \quad (4)$$

where $\delta(B)$ and $\omega(B)$ are polynomials of order r and s respectively. Equation (4) can be written as

$$y(t) = \delta(B)^{-1}\omega(B)x(t). \quad (5)$$

Equating (2) and (5), produces a relation between δ_j , ω_j and v_j . Noise has to be added to the input because the system is usually perturbed by unknown factors, not taken into account by the input $x(t)$:

$$y(t) = \delta(B)^{-1}\omega(B)x(t) + n(t) \quad (6)$$

where $n(t)$ represents the external noise which is usually shaped as an ARMA model,

$$\Phi(B)n(t) = \Psi(B)a(t) \quad (7)$$

$\Phi(B)$ and $\Psi(B)$ being two polynomials of orders p and q in the backshift operator B , and $a(t)$ white noise. The final form of the DLTFN is

$$y(t) = \frac{\omega(B)}{\delta(B)}x(t) + \frac{\Psi(B)}{\Phi(B)}a(t). \quad (8)$$

b. Parameter estimation

The parameter estimation is carried out in four steps. See Box and Jenkins (1976) for a full description of the fitting process.

Identification of the model: From the cross-correlation function between the input series x_t and the output series y_t , weights \hat{v}_j are estimated, from which the coefficients of the polynomials $\hat{\delta}(B)$ and $\hat{\omega}(B)$ are obtained for some selected values of the orders r and s . In this case, all possible combinations of $1 \leq r \leq 4$ and $1 \leq s \leq 4$, were tried keeping the one that gave minimum noise variance for subsequent analysis. Also, the roots of the $\hat{\delta}(B)$ polynomial were evaluated in order to assure the stability of the model. In all cases, the minimum noise variance model was found to be stable.

Identification of the noise: An estimate $\hat{n}(t)$ of the noise of the system is provided by subtracting the estimate \hat{y} from the observed output of the system $y(t)$, given by

$$\hat{y} = [\hat{\delta}(B)]^{-1} \hat{\omega}(B) x(t). \quad (9)$$

Once the noise series has been obtained, it is ARMA modeled

$$[\hat{\Phi}(B)]^{-1} \hat{\Psi}(B) \hat{n}(t) = a(t), \quad (10)$$

where $a(t)$ is white noise.

Parameter estimation: The final estimate of the parameters was obtained by fitting the series $\{x(t), y(t)\}$ to Equation (8) by a non-linear least squares method using the Levenberg-Marquardt algorithm. The order of the model, i.e. r, s, p, q and the initial parameters for the non-linear least squares algorithm are those obtained in previous steps. A $100(1-\alpha)\%$ confidence interval for the parameters was calculated according to Box and Jenkins (1976).

$$\delta\beta_i = \pm t_{\alpha/2} \sqrt{\sigma_a^2} \sqrt{C_{ii}} \quad (11)$$

where $\delta\beta_i$ is the confidence interval for each parameter of the model, σ_a^2 is the noise variance and C_{ii} are the main diagonal elements of the covariance matrix $(\mathbf{X}'\mathbf{X})^{-1}$, where \mathbf{X} is the Jacobian $x_{i,t} = -\partial a_t / \partial \beta_i$ at the minimum, $t_{\alpha/2}$ is a two-tailed $100\alpha\%$ critical value of Student's t , and α is taken to be 0.05.

Checking the model:

The adequacy of the fit was tested by comparing the statistics (see Box and Jenkins (1976) for a thorough discussion)

$$Q = M \sum_{k=1}^K r_{aa}^2(k), \quad (12)$$

against a χ^2 with $K - m - n$ degrees of freedom, where m, n are the orders of the ARMA model fitted to

the noise $n(t)$ and $M = N - p$ is the number of values of $a(t)$ that are actually available for computation, and statistics purposes

$$S = M \sum_{k=0}^K r_{\alpha\alpha}^2(k), \quad (13)$$

against a χ^2 with $K + 1 - (r + s + 1)$ degrees of freedom, where $(r + s + 1)$ is the number of parameters fitted in the DLTFN model. The Q test checks the adequacy of the noise model, and the S test the adequacy of the DLTFN model.

4. Forecasting

One of the main goals of time stochastic models is to forecast future values of the observed series $y(t)$. The method proposed by Box and Jenkins (1976) was followed. Using square brackets to denote conditional expectation at time t , and $\tilde{y}_t(l)$ and $\tilde{x}_t(l)$ to denote lead- l forecasts of $y(t)$ and $x(t)$:

$$\begin{aligned} \tilde{y}_t(l) = [y_{t+l}] = & \delta_1^*[y_{t+l-1}] + \dots + \delta_{p+r}^*[y_{t+l-p-r}] + \omega_0^*[x_{t+l}] - \dots - \omega_{p+s}^*[x_{t+l-p-s}] + \\ & + [a_{t+l}] - \phi_1^*[a_{t+l-1}] - \dots - \phi_{q+r}^*[a_{t+l-q-r}], \end{aligned} \quad (14)$$

where

$$[y_{t+j}] = \begin{cases} y(t+j) & j \leq 0 \\ \tilde{y}_t(j) & j > 0 \end{cases}$$

$$[x_{t+j}] = \begin{cases} x(t+j) & j \leq 0 \\ \tilde{x}_t(j) & j > 0 \end{cases}$$

$$[a_{t+j}] = \begin{cases} a(t+j) & j \leq 0 \\ \tilde{a}_t(j) & j > 0. \end{cases}$$

The $a(t)$ series is calculated as $y(t) - \tilde{y}_{t-1}(1)$, and the $\tilde{x}(t)$ series is obtained from the ARMA model fitted to the input series that was used for pre whitening. The variance of the lead-1 forecast is given by

$$V(1) = \omega_0^2 \sigma_\alpha^2 + \sigma_a^2. \quad (15)$$

5. Results and discussion

For each accumulated rainfall series analyzed (one per observatory), three DLTFN models were obtained, one for each input series: LP, SST and SLP. An ARMA model was also fitted for comparison.

Tables 4 to 6 give the orders of the DLTFN models obtained, and Table 7 the orders of the ARMA models. The residual variance is also listed. Tables 8 to 10 give the results of the S and Q tests. The tables present the values of Q and S obtained using Equations (12) and (13) with $K = 100$, the degree of freedom ν and the χ^2 critical value at a 5% level of significance. The model is not rejected when $Q, S \leq \chi^2$. According to these tables, the models are rejected in only a few cases [San Fernando, Huelva (Q test, LP/rainfall), Barcelona (Q test, SST/Rainfall)].

Table 6. Model orders for SST/rainfall.

Observatory	r, s	p, q	Residual Variance
Badajoz	1, 1	2, 3	0.97
Barcelona	1, 1	0, 1	0.98
Coimbra	0, 3	0, 0	0.96
La Coruña	1, 0	0, 0	0.97
Evora	2, 0	3, 3	0.97
S. Fernando	1, 0	1, 1	0.97
Huelva	0, 0	2, 2	0.95
Lisbon	0, 0	1, 2	0.92
Madrid	2, 2	0, 2	0.96
Málaga	0, 0	1, 1	0.96
Oporto	2, 0	0, 0	0.96
Salamanca	0, 0	2, 2	0.96
Santander	3, 1	3, 3	0.92
Sevilla	0, 0	0, 1	0.97
Soria	0, 0	3, 3	0.91
Valencia	0, 0	0, 0	0.98
Valladolid	1, 1	0, 1	0.96
Vitoria	2, 2	0, 2	0.94
Zaragoza	1, 1	2, 2	0.99

Table 7. ARMA model orders for rainfall.

Observatory	(p, q)	Residual Variance
Badajoz	1, 2	0.989
Barcelona	0, 0	0.99
Coimbra	0, 2	0.987
La Coruña	1, 1	0.972
Evora	0, 1	0.996
San Fernando	0, 1	0.992
Huelva	0, 1	0.998
Lisbon	1, 2	0.989
Madrid	0, 1	0.993
Málaga	0, 1	0.993
Oporto	3, 3	0.983
Salamanca	0, 1	0.985
Santander	1, 2	0.984
Sevilla	0, 1	0.996
Soria	0, 2	0.994
Valencia	0, 0	1.0
Valladolid	1, 1	0.987
Vitoria	1, 2	0.978
Zaragoza	1, 1	0.992

The best fit, measured by the variance reduction, is obtained when LP is the input variable, followed by SLP. The worst is obtained for the SST input series. In all cases, the variance reduction obtained with the DLTFN models is greater than the reduction in variance obtained with the ARMA models.

Table 8. Results of the Q and S tests for the LP/rainfall models: χ^2 is the chi-square cumulative probability at the 5% level of significance, ν is the degree of freedom, Q is the value of the Q statistic with $K = 100$, and S is the value of the S statistic with $K = 100$.

Observatory	Q Test			S Test		
	Q	χ^2	ν	S	χ^2	ν
Bajaioz	96.12	121.0	97	72.9	119.9	96
Barcelona	96.3	117.6	94	78.6	119.9	96
Coimbra	84.0	120.9	97	83.2	121.0	97
La Coruña	73.0	124.3	100	91.4	122.1	98
Evora	95.6	122.1	98	81.3	119.9	96
San Fernando	142.5	119.9	96	86.2	121.0	97
Huelva	270.5	119.9	96	109.2	121.0	97
Lisbon	90.3	117.6	94	81.9	121.0	97
Madrid	97.2	119.9	96	94.4	121.0	97
Málaga	116.6	123.2	99	116.5	121.0	97
Oporto	76.5	122.1	98	84.7	121.0	97
Salamanca	108.5	122.1	98	77.9	122.1	98
Santander	77.9	122.1	98	90.5	119.9	96
Sevilla	107.2	119.9	96	94.4	121.0	97
Soria	108.3	119.9	96	91.2	119.9	96
Valencia	106.2	123.2	99	113.6	122.1	98
Valladolid	86.02	121.0	97	92.7	119.9	96
Vitoria	98.3	118.7	95	75.8	121.0	97
Zaragoza	105.1	123.2	99	105.1	120.1	96

Table 9. Results of the Q and S tests for the SLP/rainfall models: χ^2 is the chi-square cumulative probability at the 5% level of significance, ν is the degree of freedom, Q is the value of the Q statistic with $K = 100$, and S is the value of the S statistic with $K = 100$.

Observatory	Q Test			S Test		
	Q	χ^2	ν	S	χ^2	ν
Bajaioz	75.6	122.1	98	78.5	119.9	96
Barcelona	93.8	124.3	100	74.0	118.7	95
Coimbra	67.9	121.0	97	100.65	124.3	100
La Coruña	68.1	122.1	98	85.1	118.7	95
Evora	75.5	121.0	97	70.6	119.9	96
San Fernando	73.7	124.3	100	101.9	119.9	96
Huelva	71.7	124.3	100	105.3	119.9	96
Lisbon	75.7	124.3	100	74.8	118.7	95
Madrid	73.7	124.3	100	101.7	119.9	96
Málaga	72.7	122.1	98	111.2	123.2	99
Oporto	68.0	122.1	98	87.4	119.9	96
Salamanca	68.0	122.1	98	91.3	119.9	96
Santander	103.6	105.2	83	71.0	119.9	96
Sevilla	73.6	119.9	96	113.4	118.7	95
Soria	68.1	123.2	99	85.5	118.7	95
Valencia	77.7	108.6	86	77.7	124.3	100
Valladolid	75.4	117.6	94	70.7	121.0	97
Vitoria	99.5	117.6	94	93.6	121.0	97
Zaragoza	62.8	124.3	100	82.2	118.7.1	95

CORRELATIONS SST/RAINFALL

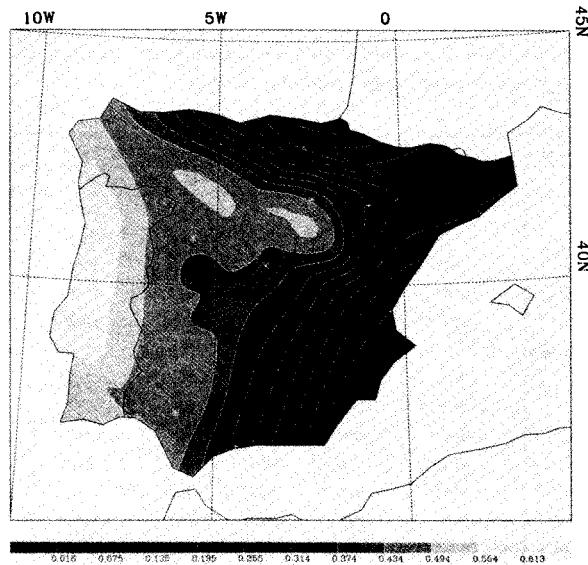


Fig. 6. Contour-shaded plot of the cross-correlation between the average of SLP boxes # 176, 177 and 244 and the rainfall at each observatory in the peninsula.

For the LP/rainfall models the explained variance is greater in observatories near the Atlantic coast. In moving away from the Atlantic, the explained variance decreases. The worst case (in terms of the explained variance) is obtained for Barcelona and Valencia on the Mediterranean coast, and for Santander and Vitoria near the Bay of Biscay. This is clearly seen in the contour plot of the explained variance displayed in Figure 7.

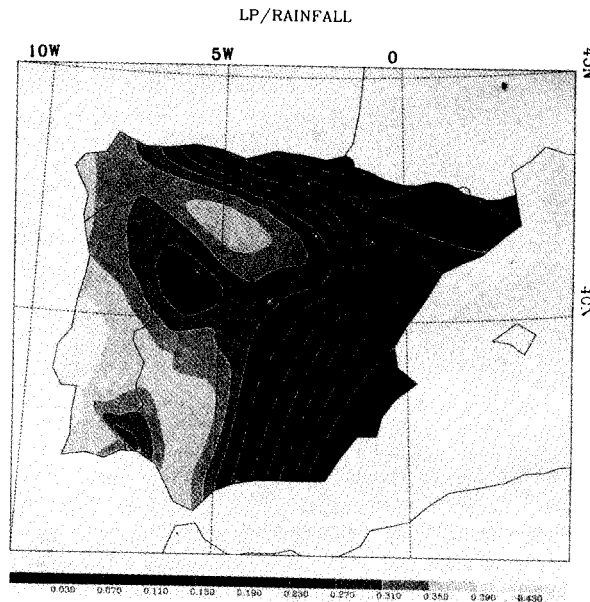


Fig. 7. Contour-shaded plot of the variance explained by the LP models.

With reference to the SLP/rainfall models, Figure 8 shows a contour plot of the explained variance.

As before, the models work better for the observatories situated in the northwestern part of the Iberian Peninsula, decreasing in quality of performance as approaching the Mediterranean coast.

Table 10. Results of the Q and S tests for the SST/rainfall models: χ^2 is the chi-square cumulative probability at the 5% level of significance, ν is the degree of freedom, Q is the value of the Q statistic with $K = 100$, and S is the value of the S statistic with $K = 100$.

Observatory	Q Test			S Test		
	Q	χ^2	ν	S	χ^2	ν
Badajoz	93.1	118.7	95	100.1	122.1	98
Barcelona	510.6	122.1	98	98.4	119.9	96
Coimbra	106.5	124.3	100	76.1	117.6	94
La Coruña	95.0	124.3	100	62.9	121.0	97
Evora	92.8	115.3	92	110.4	121.0	97
San Fernando	93.2	122.1	98	77.7	122.1	98
Huelva	93.0	113.1	90	84.3	121.0	97
Lisboa	92.8	115.3	92	125.8	122.1	98
Madrid	93.3	122.1	98	69.6	119.9	96
Málaga	99.4	117.6	94	77.1	119.9	96
Oporto	106.5	124.3	100	68.5	118.7	95
Salamanca	109.3	118.7	95	96.8	118.7	95
Santander	100.6	112.0	89	104.1	119.9	96
Sevilla	93.2	122.1	98	97.5	119.9	96
Soria	92.8	115.4	92	87.8	119.9	96
Valencia	43.2	122.1	98	24.4	119.9	96
Valladolid	93.3	122.1	98	85.8	121.0	97
Vitoria	92.5	121.0	97	108.0	119.9	96
Zaragoza	94.6	118.7	95	71.7	121.0	97

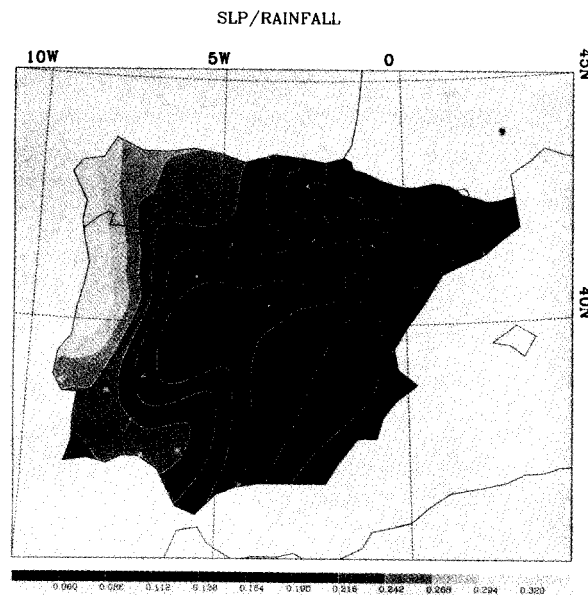


Fig. 8. Contour-shaded plot of the variance explained by the SLP models.

For the SST/rainfall models, Figure 9 presents the results of the explained variance. The explained variance is so low that there seems to be no definite region where the models work better than in others (except perhaps for Soria, which is believed to have no climatological significance).

SST/RAINFALL

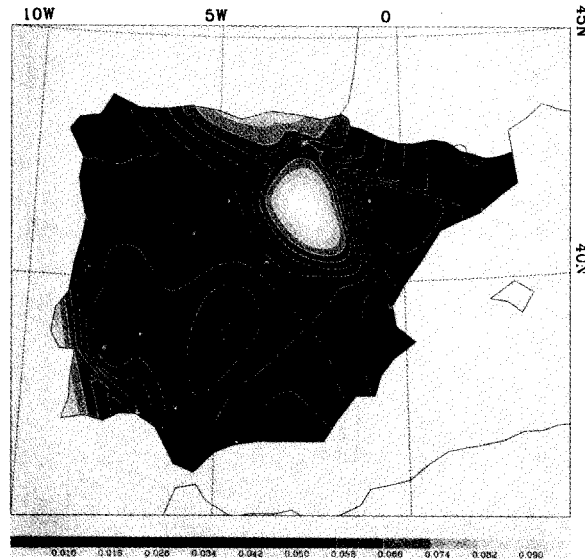


Fig. 9. Contour-shaded plot of the variance explained by the SST models.

As Table 4 shows, most of the models present either a zeroth-order polynomial $\omega(B)$ or an $\omega(B)$ with its first-order coefficient ω_1 much smaller than the zeroth-order coefficient ω_0 . This means that the models are really diagnostic rather than prognostic, and their forecasting ability will really be that of the ARMA models fitted to the input series x_t . Because in practically all the cases analyzed (LP, SLP and SST) these ARMA models behave as white noise, the forecasting models only for the most favorable cases should be developed, that is, when the LP are the input series, and using the models as diagnostic rather than prognostic models. Hence, we shall limit the research to the observational period, to using a lead-1 forecast, and to using the observed pressure value instead of the predicted value when applying Equation (14). In order to compare the forecast value with the observed one, the path followed to the normalization has to be reversed, i.e., Equation (1) must be replaced by

$$y(t) = \left(y_r(t)\sigma_2 + A_0 + \sum_{j=1}^4 A_j \sin(2\pi jt/365) + B_j \cos(2\pi jt/365) \right)^2, \quad (16)$$

where $y_r(t)$ is the forecast value given by the DLTFN model using (14) and σ_2 is the standard deviation of the deseasonalized rainfall series. The square has been added because the rainfall series were previously square root transformed. A $100(1-\alpha)\%$ confidence interval for the forecast value was calculated as

$$\delta_y = \pm t_{\alpha/2} 2 \left(y_r(t)\sigma_2 + A_0 + \sum_{j=1}^4 A_j \sin(2\pi jt/365) + B_j \cos(2\pi jt/365) \right) \sigma_2 \sqrt{V(1)}, \quad (17)$$

where $t_{\alpha/2}$ is a two-tailed $100\alpha\%$ critical value of Student's t , and α is taken to be 0.05. $V(1)$ is the variance of the lead-1 forecast. Figure 10 shows the results obtained at some observatories. Filled circles show the observed values while open circles indicate the forecast ones. The error bars are the confidence intervals calculated using (17). As the figure shows, the forecast values are not much too different from the observed ones, but the error bars are quite large due to the high variance left unexplained. Obviously, observatories

with smaller lead-1 forecast variance show smaller error bars. This is easily seen by comparing the results of Valladolid or Madrid with those of La Coruña or Lisboa.

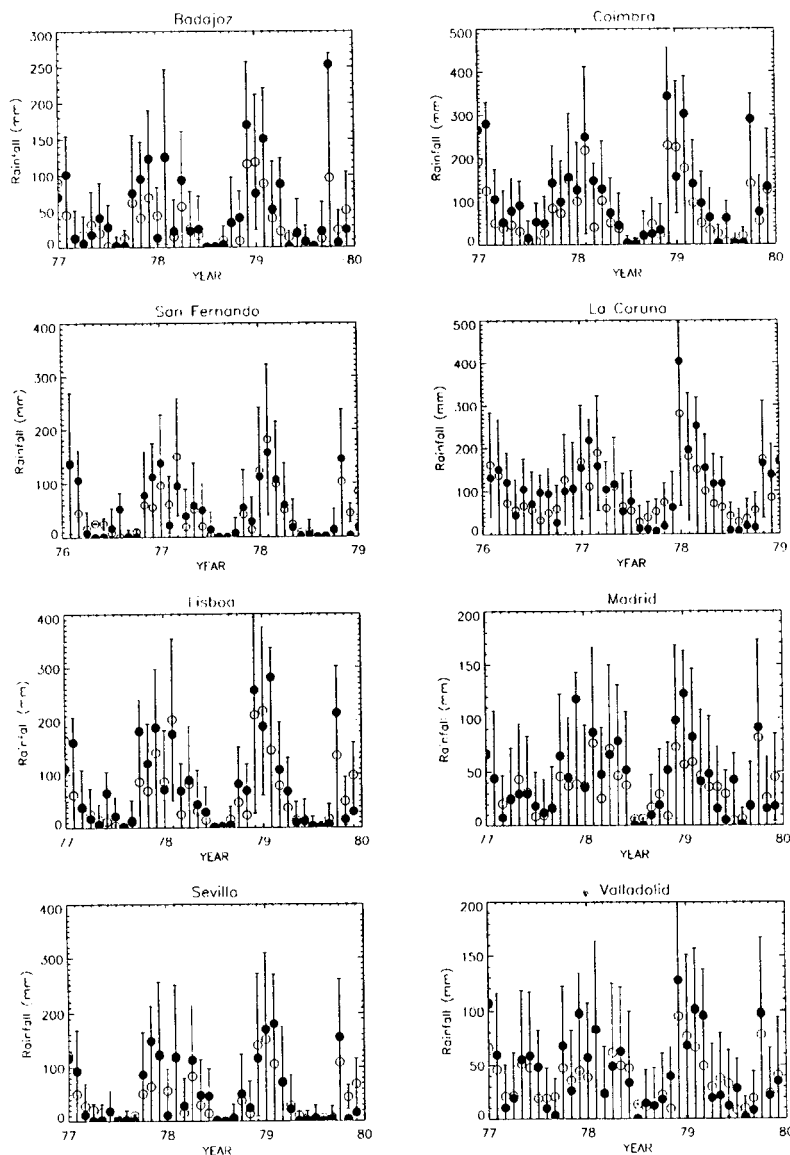


Fig. 10. Time series plot for several observatories of observed (solid circles) and 'forecast' (open circles) rainfalls for the last three years of the recording period. The LP/rainfall models were applied with lead-1 forecast and the observed pressures were used.

In order to better understand the results, we must analyze the origin of precipitation over the Iberian peninsula. The classical works on this topic are those of Linés (1981), who analyzed the perturbations and the associated rainfall that affect the Iberian Peninsula, Font (1983), who made a classification of the different weather types that affect the Peninsula, and more recently Marroquin *et al.* (1995), who analyzed the synoptic situations that seem responsible for the daily rainfall over Badajoz (one of the observatories studied, and a typical example of these results) and Ribalaygua and Borén (1996), who analyzed the relationship between the daily synoptic situations and the daily rainfall patterns over the Peninsula.

According to the results obtained by the aforementioned authors, a good deal of the rainfall over most of the Iberian Peninsula is related to the passages of fronts, moving eastward, with the associated low pressure center located in or near boxes #176, 177 and 212. This kind of synoptic situation is particularly important for the western two-thirds of the Iberian Peninsula, whereas the Mediterranean coast rainfall has a major convective and orographic origin with westerly flow and a low located in the Gulf of Genova (Catalonian rainfall) or the Gulf of Cadiz (Valencia rainfall). Also a good deal of the rainfall over the Gulf of Biscay coast is associated with a northern flow with high pressures west of the British Isles, and a low in the Gulf of Genova. Our results agree fairly well with the above in that:

- The LP/rainfall models work well for the western part of the country (rainfall due to frontal passage).
- The SLP/rainfall models work well for the western part of the country (lows located to the west/northwest of the Peninsula).

6. Summary and conclusions

In order to try to improve the results given by ARMA modeling of monthly accumulated rainfall series, a Discrete Linear Transfer Function Noise model was used, thus introducing new information given by an external variable — in this case, North Atlantic SST, SLP and Local Pressure, LP.

The preliminary analysis shows that there is a small but significant relation between SST measured along a belt from 10 to 20°N and the rainfall measured at observatories on the west and southwest of the Iberian Peninsula. This relation is much greater when the SLP data are taken as input variables. In this case, the highest correlation North Atlantic zone stretches from 30 to 50°N and lies just to the west of the Iberian Peninsula.

In all cases analyzed, the performance of the DLTFN models, measured by the explained variance of the rainfall series, is better than that of ARMA models. The model that 'works' best is the one where the input variables are LP, followed by SLP, and SST.

In most of the models, the first coefficient ω_1 of the polynomial $\omega(B)$ is either zero or much smaller than the zeroth-order coefficient ω_0 , which means that the models are diagnostic rather than prognostic. We therefore used them as a prognostic tool, keeping our 'forecasting' within the observational period.

Acknowledgements

Thanks are due to the Instituto Nacional de Meteorología of Spain for providing us with rainfall and local pressure data, and to the Spanish CICYT for their financial support, project number CLI96-1871-C04-03. Thanks are also due to the anonymous reviewers for their useful comments.

REFERENCES

- Barnett, T. P., 1984. Statistical prediction of seasonal air temperature over Eurasia. *Tellus*, **36A**, 132–146.
- Bennett, R.J., 1979. *Spatial Time Series*. Pion Limited. London.
- Box, G. E. P. and D. R. Cox, 1964. An analysis of transformations. *J. Roy. Statist. Soc., Ser. B*, **26**, 211–252.
- Box, G. E. P. and G. M. Jenkins, 1976. *Time Series Analysis: Forecasting and Control* (rev.). Holden-Day. San Francisco.
- Delleur, J. M. and M. L. Kavvas, 1978. Stochastic models for monthly rainfall forecasting and synthetic generation. *J. Appl. Meteor.*, **17**, 1528–1536.

- Efrom, B. and R. J. Tibshirani, 1993. *An Introduction to the Bootstrap*. Chapman and Hall. New York.
- Folland, C. K., T. N. Palmer and D. E. Parker, 1986. Sahel rainfall and the worldwide sea temperatures 1901-85. *Nature*, **320**, 602-607.
- Font, I., 1983. *Climatología de España y Portugal*, Instituto Nacional de Meteorología, Madrid, Spain.
- Fontaine, B., S. Janicot and V. Moron, 1995. Rainfall anomaly patterns and wind field signals over West Africa in August (1958-1989). *J. Climate*, **8**, 1503 - 1510.
- Garrido, J. and J. A. García, 1992. Periodic signals in Spanish monthly precipitation data. *Theor. Appl. Climatol.*, **45**, 97-106.
- Garrido, J. and J. A. García, 1993. Aplicación de los procesos de autoregresivos-media móvil para modelizar series temporales de precipitación en la España peninsular. *Anales de Física*, **89**, 50-56.
- Hastenrath, S., L. Greischar and J. van Heerden, 1995. Prediction of summer rainfall over South Africa. *J. Climate*, **8**, 1511-1518.
- Kahl, J. D., D. J. Charlevoix, N. A. Zaitseva, R. C. Schnell and M. C. Serreze, 1993. Absence of evidence for greenhouse warming over the Arctic Ocean in the past 40 years. *Nature*, **361**, 335-337.
- Katz, R. W. and R. H. Skaggs, 1981. On the use of autoregressive-moving average processes to model meteorological times series. *Mon. Wea. Rev.*, **109**, 479-484.
- Linés, A., 1981. *Perturbaciones típicas que afectan a la Península Ibérica y precipitaciones asociadas*. Instituto Nacional de Meteorología de España, Madrid.
- Lorente, J. M., 1985. *La variabilidad de las precipitaciones atmosféricas sobre la España peninsular*. *Revista de Geofísica XIV*, Madrid.
- Marroquin A., J. A. García, J. Garrido and V. L. Mateos, 1995. Neymann-Scott cluster model for daily rainfall processes in Lower Extremadura (Spain). Rainfall generating mechanisms. *Theor. Appl. Climatol.*, **52**, 183-195.
- Mateos, V. L., 1993. *Modelos de Funcion de Transferencia para series de precipitacion en la peninsula Iberica*, Ph. D. Thesis, Universidad de Extremadura, España.
- Meehl G. A. and H. van Loon, 1979. The seesaw in winter temperatures between Greenland and northern Europe. Part III: Teleconnections with Lower Latitudes. *Mon. Wea. Rev.*, **107**, 1095-1106.
- Nicholls, N., 1989. Sea surface temperature and Australian winter rainfall. *J. of Climate*, **2**, 965-973.
- Ratcliffe, R. A. S. and R. Murray, 1970. New lag associations between North Atlantic sea temperature and European pressure applied to long-range weather forecasting. *Quart. J. Roy. Meteor. Soc.*, **96**, 226-246.
- Ribalaygua, J. and R. Borén, 1996. *Clasificación de Patrones de Precipitación Diaria Sobre la España Peninsular y Baleárica*, Informe N°3, Servicio de Analisis e Investigación del Clima, Instituto Nacional de Meteorología, Spain.
- Rowntree, P. R., 1976. Response of the atmosphere to a tropical Atlantic Ocean temperature anomaly. *Quart. J. R. Met. Soc.*, **102**, 607-625.
- Woodruff, S. D., R. J. Slutz, R. J. Jenne and P. M. Steurer, 1987. A comprehensive ocean-atmosphere data set. *Bull. Amer. Meteor. Soc.*, **68**, 1239-1250.