

Forecasting local daily precipitation patterns in a climate change scenario

Jesús Abaurrea*, Jesús Asín

Dpto. Métodos Estadísticos, Universidad de Zaragoza; Pedro Cerbuna 12, Zaragoza 50009, Spain

ABSTRACT: The present study introduces a statistical procedure for obtaining long-term local daily precipitation forecasts in a climate change scenario. It is based on a regression model that uses climate variables properly reproduced by a General Circulation Model (GCM) as predictors. The daily rainfall model used consists of a logistic regression as the occurrence model and a generalized linear model (GLM) with Gamma error distribution as the quantity model. The ability of the model to generate plausible long-term projections is analysed by studying and comparing its behaviour using observed and GCM simulated data as input. The method is applied to forecast the rainfall pattern in the area of Zaragoza (Spain) for the period 2090–2100, in an IS92a scenario. We use the data corresponding to an experiment with the CGCM1 model, the first version of the coupled GCM of the Canadian Centre for Climate Modelling and Analysis (CCCma). The results obtained show that no significant change in global rainfall frequency or in the annual accumulated amount are to be expected; however, an important modification of the seasonal cycle, with a high decrease in rainfall frequency and in the amount collected in spring, is forecasted.

KEY WORDS: Statistical downscaling · Daily rainfall · Climate change

Resale or republication not permitted without written consent of the publisher

1. INTRODUCTION

The development of complex models capable of simulating the ocean–land–atmosphere system has allowed climate scientists to obtain projections for long-term climate evolution in different scenarios. However, these models are not able to provide output at the required spatial resolution for representing the processes of interest in hydrological applications, such as the impact produced by a rainfall pattern change in the runoff regime or the erosion process in a river basin, which require daily rainfall data on a local scale. This has motivated the development of precipitation downscaling methods, which can be roughly classified into 2 groups: those using a numerical model (a regional climate model nested in a GCM) and those based on a statistical model which relates the local rainfall to the GCM output.

There are several approaches for the statistical modeling of daily precipitation. Hughes et al. (1999), Good-

ess (2000), Stehlik & Bardossy (2002) and Sumner et al. (2003) link precipitation with the atmospheric state which governs the model parameters. Palutikof et al. (2002) and Beersma & Buishand (2003) use the daily atmospheric state for selecting a subset of the historical dataset, from which the rainfall observation is obtained through a sampling schema. Others, such as Wilby et al. (1999) or Beckmann & Buishand (2001, 2002), obtain, through regression analysis, the function linking the rainfall model parameters with the variables representing the atmospheric situation. Although the atmospheric circulation explains the rainfall behaviour to a large extent, several authors (e.g. Buishand & Klein Tank 1996, Wilby et al. 1999, Cavazos 2000, Beckmann & Buishand 2002) show that rainfall is also tied to other variables not directly related with pressure, such as humidity or temperature in different atmosphere levels.

The statistical method proposed in this paper is based on a regression model and can characterise,

*Email: abaurrea@unizar.es

quite satisfactorily, the daily rainfall behaviour (Abau-rea & Asín 2003, Asín 2003). It is applied to the series measured at the Zaragoza Airport observatory and to those obtained from CGCM1 data corresponding to the grid-cell where this observatory is located. The CGCM1 model is the first version of the coupled general circulation model (GCM) developed by the Canadian Centre for Climate Modelling and Analysis (CCCma).

Zaragoza is situated in the middle of the Ebro Valley, at 240 m above sea level. The Ebro river flows from northwest to southeast into the Mediterranean Sea, and its basin (85 362 km²) borders on the Pyrenees and the Cantabrian mountains to the north, the Iberian chain to the southwest and the Coastal-Catalonian chain to the east. This orographical situation favours a NW wind predominance and a dry climate in the central part of the valley, which is affected by climate extremes and convective storms. The observatory is free from an urban effect and is the only one in the region with radiosonde data.

The time periods considered are 1990–2000 and 2090–2100; the data length is limited by the upper air measurements available in Zaragoza since August 1990. All models have been fitted using 1990–1997 data and we have set aside the 1998–2000 period for validation. Thus, the fitting is based on 2153 days for the occurrence model and on 530 days for the quantity model.

The procedure implementation consists of 4 steps: (1) A comparative analysis of GCM output and the observed climate variables is carried out to select the potential predictors. (2) The statistical model built from the selected predictors is estimated and checked. (3) The resulting model is evaluated, as a forecasting tool, by analysing its achievement in reproducing the rainfall pattern using observed and GCM simulated data as input. In this way, the possible bias introduced by the model structure can be detected and the model rebuilt. (4) Using the resulting model, the precipitation pattern in the climate change scenario is forecasted and the proposed changes are analysed.

The statistical model we used is composed of a logistic regression model for fitting the daily rainfall probability and a generalised linear model with Gamma error distribution for fitting the rainfall amount measured on wet days. This approach is close to that of Stern & Coe (1984), the main difference being that our model fits a different rainfall probability and amount distribution every day.

In Section 2 we introduce the data used; in Section 3, the statistical rainfall model is shown and the degree of achievement in fitting the different aspects of daily rainfall is discussed; how the model can actually reproduce the observed rainfall pattern when GCM gener-

ated data are used as input is also analysed. In Section 4 we discuss the properties of the forecasted rainfall pattern. Section 5 summarises the results and states some conclusions.

2. DATA

2.1. Data employed

The observed data were provided by the Spanish Meteorological Institute (INM); the CCCma provided the data simulated by the CGCM1 model. The GCM data correspond to a GHG+A experiment with an IPCC IS92a forcing scenario, in which the evolution of greenhouse gases (GHG) corresponds to that observed for the period 1900 to 1990; thereafter, a 1% increase rate per year is hypothesised up to 2100; the effect of sulphate aerosols is also included. Details about the model and simulation experiments are given in Boer et al. (2000a,b).

The daily data for the GHG+A1 experiment for the period 1961–2100 are available from the CCCma website; the data are provided by a 97×48 Gaussian grid (approximately 3.75 latitude \times 3.75 longitude). We downloaded the information relative to the periods 1990–2000 and 2090–2100 for the cell centered at 42.68° N, 0.00° E, where the Zaragoza Airport observatory (41.66° N, 1.01° W) is located.

At the Zaragoza observatory, the surface meteorological situation is measured every day at 0, 7, 12 and 18 h and the upper air information is collected by 2 radiosondes launched at 0 and 12 h. We point out that the number of missing values increases with height because the radiosonde is launched only once if it does not reach the 500 hPa level at the first attempt.

The variables in the model are those offered by the CCCma for the GCM experiment; as response variables: the rainfall amount recorded each day (0 to 24 h) (R_t) and a derived variable (Y_t), which equals zero if R_t is less than 0.1 mm m^{-2} and 1 otherwise; as surface predictors: the daily maximum and minimum temperatures (T_{MAX} and T_{MIN}), the daily temperature average (T), the average of the atmospheric pressure at 0 and 12 h (P), the specific humidity daily average (Q) and the daily average of U and V wind components. The information on the upper atmosphere exists for the 850 and 500 hPa levels at 0 h and includes, for each level, the height (A), temperature, specific humidity and U and V wind components. The notation for these variables consists of letters which identify the variable, followed by the level where it has been measured: e.g. Q_{SUP} denotes the surface humidity and A_{850} the 850 hPa height.

Other useful predictors are derived from previous variables: the daily temperature range (T_{DIF}), the temperature difference between 850 and 500 hPa levels ($T_{85/50}$) and the 500/850 hPa layer depth ($A_{85/50}$). The wind information is expressed by means of direction and velocity (d_W and v_W , respectively).

For each predictor, we calculated its anomaly series by subtracting from each daily observation the corresponding mean value, obtained through regression using the first 4 Fourier harmonic functions as covariates. These anomaly variables are denoted by the name of the corresponding variable preceded by the letter z , e.g. $z.A_{850}$.

2.2. Observed versus GCM generated variables

The first step in building the model is the definition of the potential predictor set. Since the GCM generates data corresponding to a wide area, reflecting an average situation, the values produced will not be equal to those measured in a particular observatory. For some variables the differences can occur in a certain period of the year, in other cases they can be related to the mean level, so an anomaly series could have a similar statistical behaviour, at least in certain periods.

The comparative analysis between the observed and the CGCM1 generated data begins with an exploratory analysis where we analyse, using a particular smoother (lowess) and monthly boxplot diagrams, the similarity in location and variability. From this we conclude that GCM variables do not properly reproduce the behaviour of those observed. In some cases the difference is important, see for example the results for precipitation in Fig. 1a; in other variables, such as Q_{SUP} (Fig. 1b), the boxplots show an important difference in mean position but a great similarity in dispersion.

The properties of simulated wind data are different from those of the observed data; the differences are greater at the surface level and become smaller as height increases. In Zaragoza, the prevailing surface winds are those from the northwest and, to a small extent, those from the southeast. In the simulated data the SW and NW wind frequencies are equivalent and the SE wind does not have any special relevance (Fig. 2). At upper levels, the influence of the underlying surface decreases, so at 500 hPa the difference between both distributions is less important.

This comparative exploratory analysis has been complemented with some statistical tests. For each variable and the corresponding anomaly we apply a t -test for testing mean equality, a Wilcoxon test on median equality, an F -test for checking variance equality, a Wilcoxon test applied to differences from the respective means and, finally, a Kolmogorov-Smirnov (KS) test to check the distribution equality. Both the observed and GCM simulated data show serial dependence; only the t -test takes into account the existing autocorrelation, other p -values have to be considered as a measure of the distance between sampling distributions. The sample size is 300, approximately, for each month.

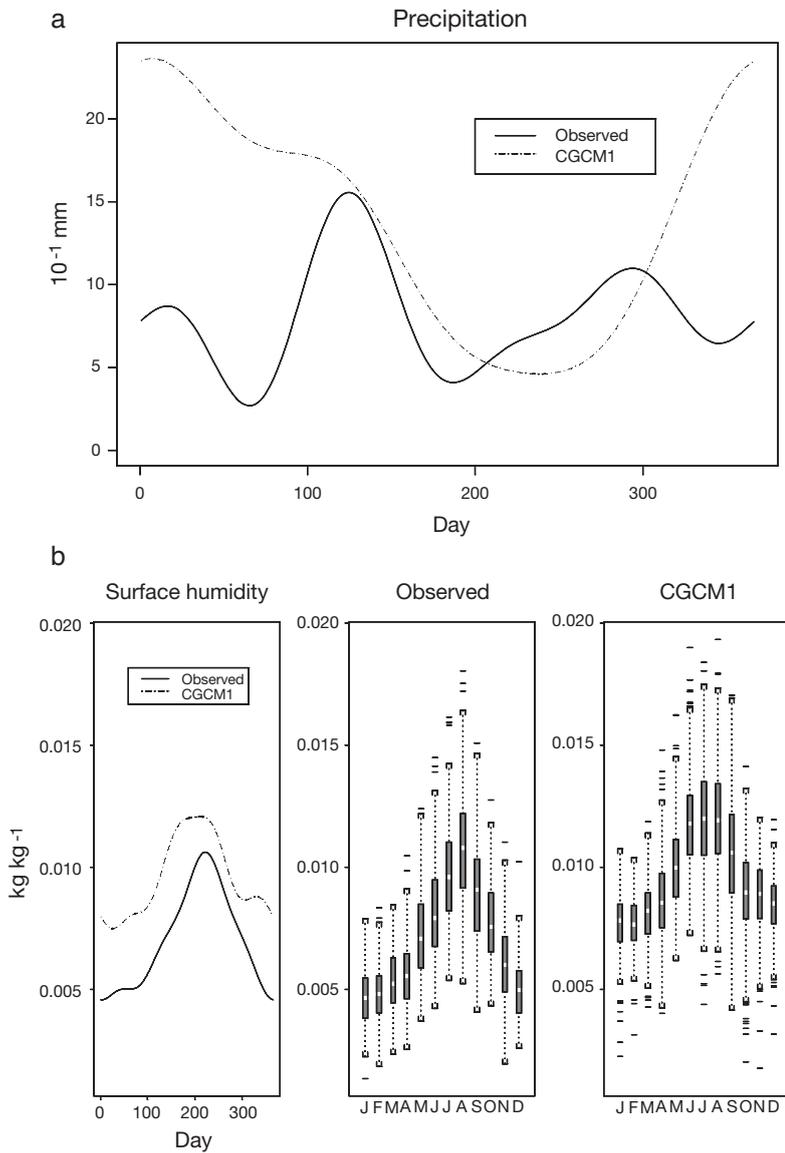


Fig. 1. Comparison of daily observed and CGCM1 generated data (1990–2000). (a) Lowess smoother of daily rainfall amounts; (b) lowess and monthly boxplots of specific humidity data at surface level

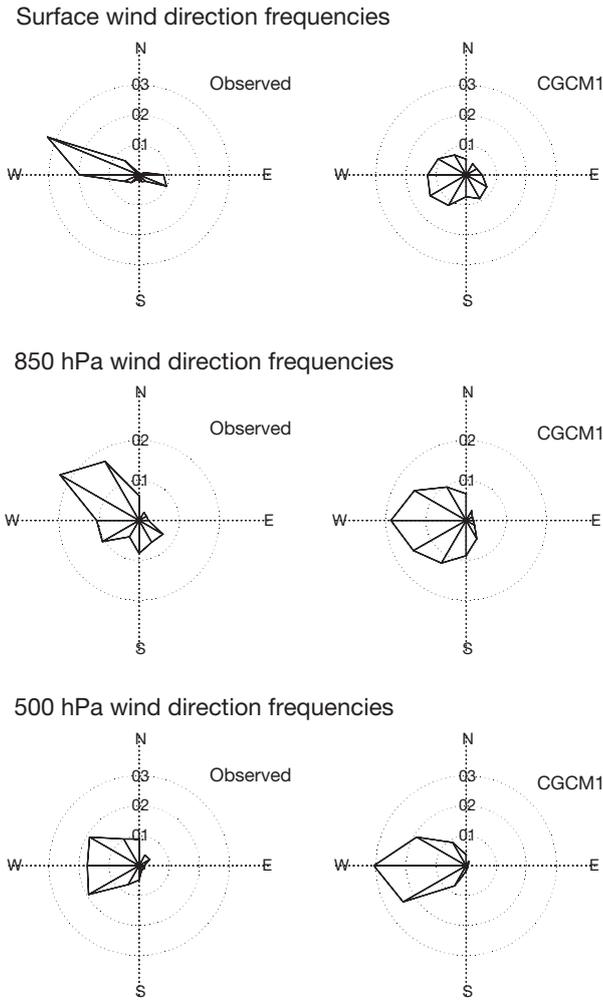


Fig. 2. Wind roses of observed and CGCM1 generated wind direction frequencies (1990–2000)

In Table 1 we show, for the different tests, the number of months with p-values less than 0.01 and 0.05. For the best reproduced variables we have analysed the similarity throughout the year; Table 2 shows, by seasons, the KS test results. They confirm that discrepancies are mainly due to location differences and to difficulties in describing the seasonal cycle; only the simulated Q_{500} and T_{850} data reproduce the observed characteristics (Fig. 3).

When we compare the KS test results for the anomaly and the variable series, we find that distributions are closer for the anomalies. The KS and F -test p-values are greater than 0.05 at least in 6 months for $z.Q_{SUP}$, $z.Q_{500}$, $z.P_{SUP}$, $z.A_{850}$, $z.U_{850}$ and $z.V_{850}$. The KS test produces p-values less than 0.01 in more than 6 months for $z.T_{MAX}$, $z.T_{850}$, $z.A_{85/50}$ and for the anomalies of surface wind components.

When we analyse these results seasonally, we see that there are variables, such as $z.P_{SUP}$, $z.A_{850}$, $z.Q_{850}$, $z.Q_{500}$, $z.V_{850}$ or $z.V_{500}$, where the KS test gives p-values less than 0.05 for 1 month, at most, in each season; for other variables, $z.Q_{SUP}$ for example, this level of agreement is lost in some seasons (in spring in the case of $z.Q_{SUP}$).

To sum up, although in general the CGCM1 data distributions are not similar to the observed ones in Zaragoza, we have found some GCM anomaly series which acceptably reproduce the observed characteristics and, therefore, could be considered as potential predictors for a downscaling model.

Table 1. Monthly comparison of observed and CGCM1 daily data distributions (1990–2000). n/m: number of months where the p-value is less than 0.01 (n) or 0.05 (m). KS: Kolmogorov-Smirnov test

	t-test	Wilcoxon	F-test	Dispersion	KS
T_{MAX}	9/11	12/12	10/11	10/10	12/12
T_{MIN}	6/7	7/10	7/9	6/6	11/12
T_{DIF}	10/10	10/10	8/9	5/7	10/10
T_{850}	2/5	6/7	8/8	9/11	11/11
T_{500}	6/12	9/11	11/11	9/9	11/12
$T_{85/50}$	6/8	9/9	6/8	7/8	10/11
Q_{SUP}	12/12	12/12	1/2	1/1	12/12
Q_{850}	9/10	10/10	6/8	5/8	9/11
Q_{500}	1/2	3/4	4/5	1/3	2/4
P_{SUP}	9/9	11/11	5/6	4/5	9/11
A_{850}	7/8	9/9	2/3	2/2	9/9
A_{500}	7/8	8/9	12/12	8/12	11/12
$A_{85/50}$	4/8	10/10	9/11	7/8	11/11
U_{SUP}	6/7	7/9	11/11	10/11	11/12
V_{SUP}	11/11	11/11	5/5	6/6	12/12
U_{850}	4/5	6/6	4/6	5/6	6/7
V_{850}	8/9	9/10	4/6	0/2	10/10
U_{500}	4/7	8/9	10/10	8/8	10/11
V_{500}	4/6	6/8	7/9	5/8	8/9
$z.T_{MAX}$	0/0	0/1	10/10	10/11	9/10
$z.T_{MIN}$	0/0	0/1	8/9	7/8	6/6
$z.T_{DIF}$	0/0	0/1	8/9	5/7	0/5
$z.T_{850}$	0/0	1/1	9/10	11/11	8/9
$z.T_{500}$	0/0	0/2	11/11	9/10	3/6
$z.T_{85/50}$	0/0	0/1	6/8	7/8	2/6
$z.Q_{SUP}$	0/1	3/3	2/2	1/2	1/3
$z.Q_{850}$	0/0	0/0	7/8	6/8	0/2
$z.Q_{500}$	0/0	0/0	4/5	1/3	0/1
$z.P_{SUP}$	0/0	0/0	4/6	4/5	1/3
$z.A_{850}$	0/0	0/0	2/3	1/2	0/0
$z.A_{500}$	0/0	0/0	11/11	10/10	3/8
$z.A_{85/50}$	0/0	2/3	9/11	7/8	7/8
$z.U_{SUP}$	0/0	0/2	11/11	11/11	8/9
$z.V_{SUP}$	0/0	0/2	5/5	6/6	7/8
$z.U_{850}$	0/0	0/2	5/6	5/6	2/3
$z.V_{850}$	0/0	1/3	4/6	0/2	2/3
$z.U_{500}$	0/0	0/0	10/10	8/8	3/7
$z.V_{500}$	0/0	0/0	7/9	5/8	0/1

Table 2. Seasonal analysis of KS test p-values corresponding to the comparison of the observed and CGCM1 daily data (1990–2000). n/m: number of months where the p-value is less than 0.01 (n) or 0.05 (m)

	Global	Winter (DJF)	Spring (MAM)	Summer (JJA)	Autumn (SON)
$z.T_{MAX}$	9/10	3/3	3/3	2/2	1/2
$z.T_{MIN}$	6/6	0/0	1/1	2/2	3/3
$z.T_{DIF}$	0/5	0/0	0/3	0/0	0/2
$z.P_{SUP}$	1/3	0/0	1/1	0/1	0/1
$z.Q_{SUP}$	1/3	0/0	1/2	0/0	0/1
$z.T_{850}$	8/9	2/3	3/3	2/2	1/1
$z.A_{850}$	0/0	0/0	0/0	0/0	0/0
$z.Q_{850}$	0/2	0/0	0/1	0/0	0/1
$z.T_{500}$	3/6	1/2	0/1	0/1	2/2
$z.A_{500}$	3/8	1/3	0/1	0/1	2/3
$z.Q_{500}$	0/1	0/0	0/1	0/0	0/0
$z.T_{85/50}$	2/6	0/0	0/1	1/3	1/2
$z.A_{85/50}$	7/8	2/3	2/2	0/0	3/3
$z.U_{SUP}$	8/9	2/2	1/2	2/2	3/3
$z.V_{SUP}$	7/8	2/3	1/1	3/3	1/1
$z.U_{850}$	2/3	0/0	0/0	1/2	1/1
$z.V_{850}$	2/3	0/0	1/1	1/1	0/1
$z.U_{500}$	3/7	0/1	1/2	1/3	1/1
$z.V_{500}$	0/1	0/1	0/0	0/0	0/0

3. STATISTICAL MODEL

A logistic regression model gives the rainfall occurrence probability for each day, $P(Y_t = 1|\mathbf{X}_t)$, in the conditions stated by the predictor set, denoted by \mathbf{X}_t , which contains the relevant information corresponding to Days t , $t + 1$ and those prior to Day t .

The amount model, which explains the precipitation collected on a wet day, is also a generalized linear model (GLM) with a Gamma error and a logarithmic link. The Gamma distribution is characterised by a location parameter, the expected value, and a dispersion one ν , the square of the variation coefficient. When $\nu = 1$ this distribution corresponds to the Exponential model and if $\nu > 1$, there is not a modal value and the density function shows a strong concentration near zero. According to a recommendation by McCullagh & Nelder (1989), we estimate the ν parameter, considered as constant in the model, by the moment method.

The occurrence and quantity sub-models describe the rainfall phenomenon completely.

3.1. Criteria applied for building the model

In order to increase the explained variability, the linear predictor includes not only information about Day t and the previous days, but also about Day $t + 1$, since the future use of the model allows this possibility.

Our objective is to obtain simple linear predictors which allow a physical interpretation; thus, we use the original variables, or simple transformations, not applying statistical techniques to reduce the predictor matrix dimension.

The modelling process is an iterative procedure consisting of formulating and estimating different structures and analysing their proposals in contrast to data. In order to select the best model, we apply a statistical test which compares, on each occasion, the deviance of 2 nested models; the significance level used is 0.05 except for removing harmonic terms when another of a lower order is not significant, in this case a level 0.01 is used.

Due to the large number of predictors available, a 2-step procedure is applied. First, a full model is built by the addition of terms which significantly improve the fitting. More precisely:

- We introduce the Fourier harmonic functions which allow us to model complex periodic patterns using a few parameters. We denote $S_i t$ and $C_i t$ as the values of sine and cosine variables corresponding to i th order harmonic function evaluated at Day t .
- We then introduce the possible climate covariates; if one is retained, the corresponding quadratic effect is considered, and so on. If a variable enters the model, the variables concerning the previous and following days are also analysed for entering.

After reaching a full model, a simplifying procedure which uses the t -statistic absolute value as the criterion is applied to eliminate the non-significant terms. A term is not removed if a higher order one in the same variable is significant.

3.2. Fitted statistical models

A model useful for projecting the rainfall pattern must reproduce the observed pattern when the covariate values generated by the GCM for the current climate are used as input. So, the usual tools used for model checking must be complemented with an analysis of fitted values obtained using the observed and the CGCM1 generated covariates. In this way we can discriminate between problems attributable to the model from those resulting from an inadequate reproduction of the observed local climate by the GCM.

In a first attempt, we used as predictors the anomalies whose KS p-values were greater than 0.01 in at least 6 of the months. The selection was based on the KS test because it compares the global distribution and not only partial aspects of it. This criterion is not very strict, so it allowed us to use a wide number of covariates and to achieve a good fitting. No wind component anomalies of any level were used in the model.

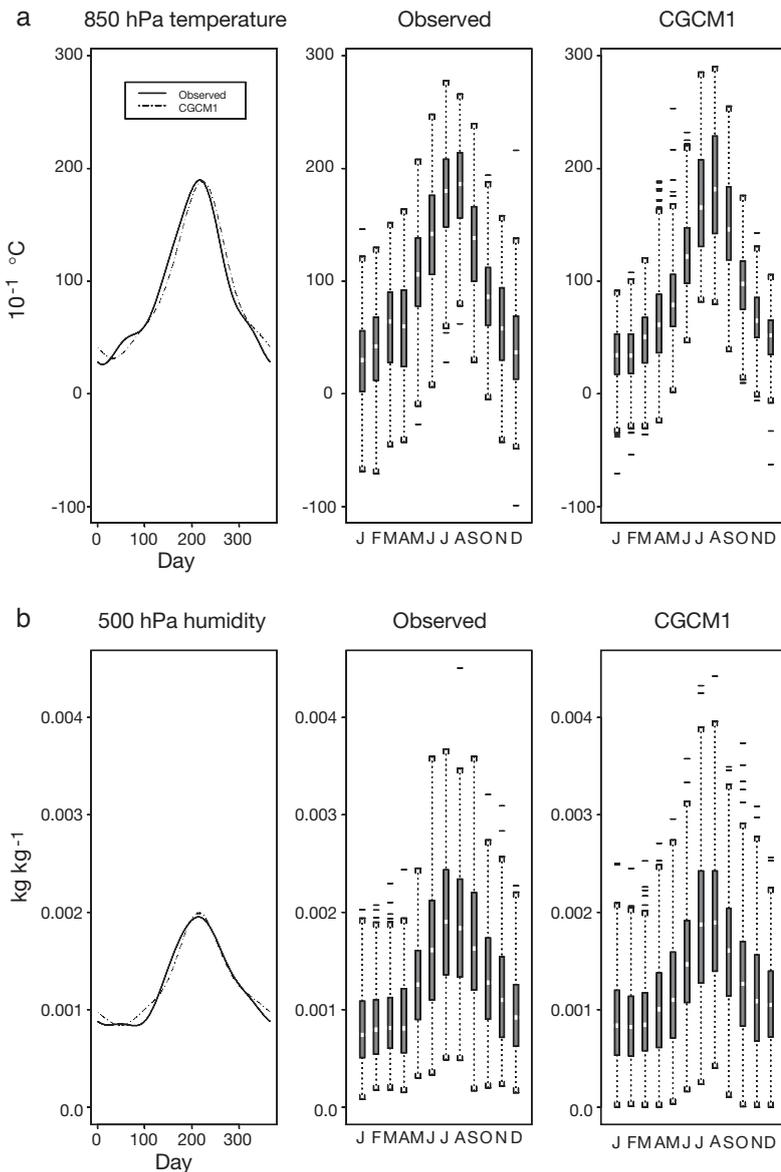


Fig. 3. Lowess and monthly boxplots of observed and CGCM1 generated 1990–2000 daily data. (a) 850 hPa temperature; (b) 500 hPa specific humidity

This model, not shown, provided an adequate fitting of rainfall characteristics when it was used with the observed data but strong discrepancies appeared using the GCM generated data: the rainfall frequency in the spring months and the annual total amount were underestimated by 30% and a larger bias appeared in the spring rainfall amount that was estimated to be lower than the summer amount.

A sensitivity analysis showed us that GCM values of $z.T_{\text{MIN}}$, $z.T_{\text{DIF}}$, $z.A_{500}$ and $z.Q_{\text{SUP}}$, necessary for achieving a good fit, were causing the bias. This led us to use them only in those seasons where they showed satisfactory behaviour regarding observed data. The final

rule was to employ these covariates in seasons where the KS p-values were greater than 0.05 in at least 2 months (Table 2). Exceptionally, $z.T_{\text{DIF}}$, whose autumn KS p-values were greater than 0.05 only in 1 month and greater than 0.01 in all of them, was used. The model resulting from applying this rule, denoted the Down model, is described in the next subsection.

3.2.1. Description of the Down Model

The linear predictors of occurrence and quantity sub-models, denoted Down–Oc and Down–Q respectively, are shown in Table 3. Both contain harmonic terms up to the 4th order and interaction terms of climate covariates with seasons; these terms are denoted by combining the season and covariate names.

The linear predictor of Down–Oc contains 29 terms, 9 of which are main (non interaction) terms. Fig. 4 shows the sign-opposite effect of $z.P_{\text{SUP}}$ in Days t and $t-1$ and also the different effect of $z.T_{\text{DIF}}$ in winter from the rest of the year. We point out that the $z.Q_{\text{SUP}}$ effect doubles in winter with respect to summer or autumn.

The Down–Q model is simpler; its main effects are related to the upper air variables $z.A_{850}$, $z.Q_{500}$ and $z.T_{85/50}$. No information about surface pressure or any particular term concerning the autumn season appears in this model.

3.2.2. Goodness of fit

The harmonic terms adequately fit the seasonal cycle of rainfall frequency (Fig. 5) and amounts (Fig. 6); the correlation coefficient between observed and fitted monthly mean values is 0.98 for Down–Oc and 0.93 for Down–Q (Tables 4 & 5).

Down–Oc explains, for the estimation period, 39% of response variability, a value similar to those obtained by Beckmann & Buishand (2001, 2002) for different European observatories. Our model correctly classifies 78.6% of the days, using the smoothed relative rainfall frequency as a threshold (Table 4). For the validation data set, a similar proportion of explained variability (32.6%) is globally obtained, but this value decreases appreciably in some seasons (13 and 10% in spring and autumn, respectively).

Table 3. Linear predictor terms of the occurrence (Down–Oc) and quantity (Down–Q) sub-models of the Down model

	Down–Oc			Down–Q ($v = 1.443$)		
	Coef.	SE	Coef./SE	Coef.	SE	Coef./SE
β_0	-1.808	0.110	-16.470	3.030	0.087	34.750
S1	-0.165	0.112	-1.475	0.046	0.089	0.512
C1	0.328	0.133	2.467	-0.097	0.112	-0.872
S2	-0.491	0.099	-4.945	-0.237	0.080	-2.944
C2	-0.207	0.116	-1.778	-0.409	0.101	-4.040
S3	0.237	0.098	2.428	4.26e-3	0.081	0.052
C3	0.306	0.100	3.052	0.260	0.083	3.139
S4	0.276	0.097	2.842	0.162	0.078	2.079
C4	-0.179	0.095	-1.892	-0.011	0.078	-0.144
$z.P_{SUPt}$	-0.018	2.39e-3	-7.507			
$z.P_{SUPt}^2$	-3.37e-5	1.13e-5	-2.978			
$z.P_{SUPt-1}$	9.47e-3	3.75e-3	2.524			
$z.A_{850t-1}$	-7.26e-3	3.37e-3	-2.154			
$z.A_{850t+1}$				-9.36e-3	1.48e-3	-6.314
$z.A_{850t+1}^2$				-2.22e-5	1.08e-5	-2.047
$z.Q_{850t+1}$	135.10	75.99	1.778			
$z.Q_{850t+1}^2$	-9.59e+4	2.65e+4	-3.618			
$z.Q_{500t}$	538.4	141.7	3.798	336.8	100.8	3.343
$z.Q_{500t+1}$	615.5	149.2	4.125			
$z.T_{500t+1}$	-0.018	2.75e-3	-6.430			
$z.T_{85/50t+1}$				-4.4e-3	2.15e-3	-2.045
$z.T_{85/50t+1}^2$				-1.43e-4	4.12e-5	-3.475
Winter						
$z.T_{MINt}$	-0.024	7.16e-3	-3.414			
$z.T_{DIFt}$	-0.027	5.22e-3	-5.205	-0.016	3.44e-3	-4.532
$z.T_{DIFt}^2$	2.48e-4	1.13e-4	2.198			
$z.Q_{SUPt}$	1063.0	210.4	5.050			
Spring						
$z.T_{MINt}$	0.027	5.81e-3	4.591			
$z.A_{500t}$				2.76e-3	1.20e-3	2.309
$z.A_{500t+1}$	-4.91e-3	1.76e-3	-2.795			
Summer						
$z.T_{DIFt}$	-0.024	6.04e-3	-3.969			
$z.Q_{SUPt}$	447.2	87.0	5.141	172.9	71.6	2.416
$z.A_{500t+1}$	-0.012	2.99e-3	-4.074	-4.39e-3	1.94e-3	-2.267
Autumn						
$z.T_{DIFt}$	-0.026	4.63e-3	-5.694			
$z.Q_{SUPt}$	489.5	89.9	5.443			

The variability explained by Down–Q is, globally, 17.4% and it decreases to 13.6% in spring, when $z.Q_{SUP}$ and $z.T_{DIF}$ are excluded from the model. The result improves on that obtained by Beckmann & Buishand (2001) for the Spanish observatories they study, where a maximum of 13.3% is attained. Our value is close to that obtained by these authors in Austrian and Swiss observatories (20%). As an additional check, we classified each observed amount into the 5 intervals defined by 10th, 25th, 75th and 90th percentiles of the distribution fitted to the corresponding day. According to this criterion the fitting is adequate: the frequency distribution is close to the expected probabilities with the exception of quantities below the first decile.

The degree of achievement of Down–Q in validation ensemble is less satisfactory. The explained variability is low in some seasons because of the great influence of certain data; when these data are eliminated, better values are obtained: in summer, 9.9% is attained if we disregard 4 days (5 June 1998, 1 July 1998, 1 and 9 June 2000) and in winter we attain 18.4% if the period 30 January 1998 to 3 February 1998 is eliminated.

3.2.3. Model simulation

In order to analyse the degree of achievement attained in describing some important daily rainfall proper-

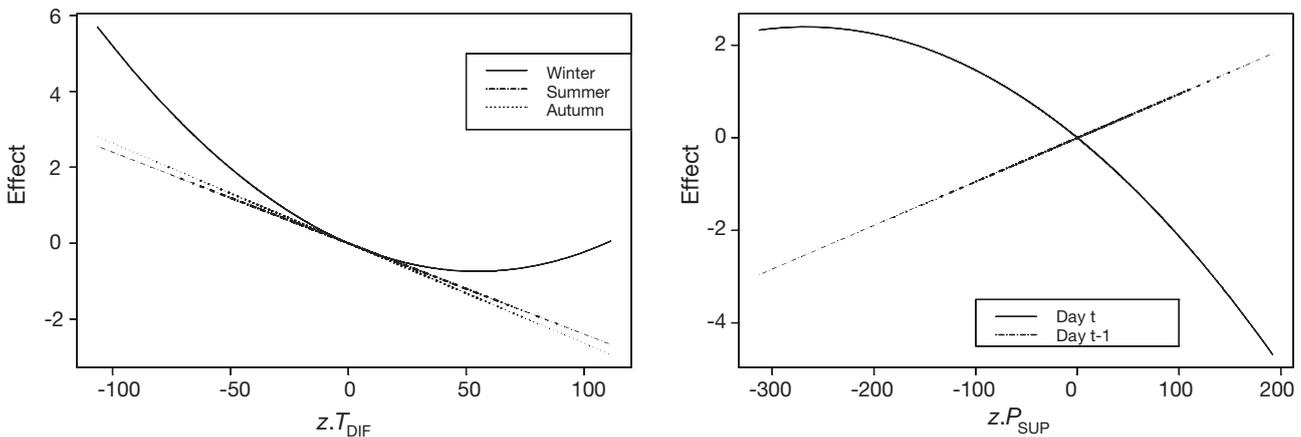


Fig. 4. Different effects of $z.T_{DIF}$ and $z.P_{SUP}$ in the linear predictor of the Down–Oc model

ties, we simulated, using the Down model, 20 rainfall series with the same length as the observed one and we applied to them the diagnostic checks proposed by Wilby et al. (1998) (Table 6). In our opinion, these checks are incomplete and can be misleading because of their aggregation level, which does not take into

account the phenomenon seasonality and does not reflect the model behaviour with regard to daily characteristics. For this reason we compare the run-length and the amount distributions monthly.

The occurrence model is found to be adequate; the KS test (results not shown) is statistically significant at

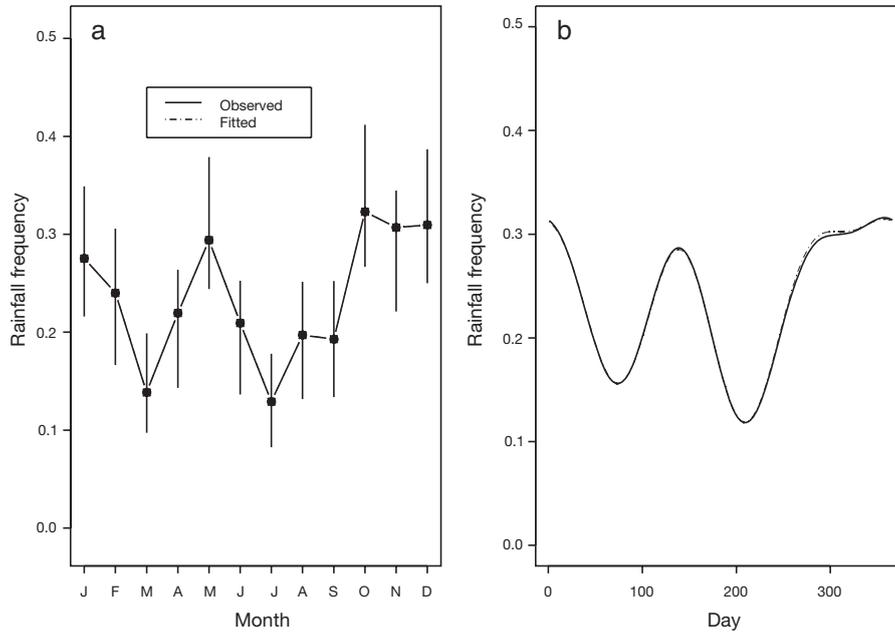


Fig. 5. Wet day observed relative frequencies and fitted probabilities using the Down–Oc model and 1990–1997 observed covariates. (a) Monthly means; (b) lowess of daily values

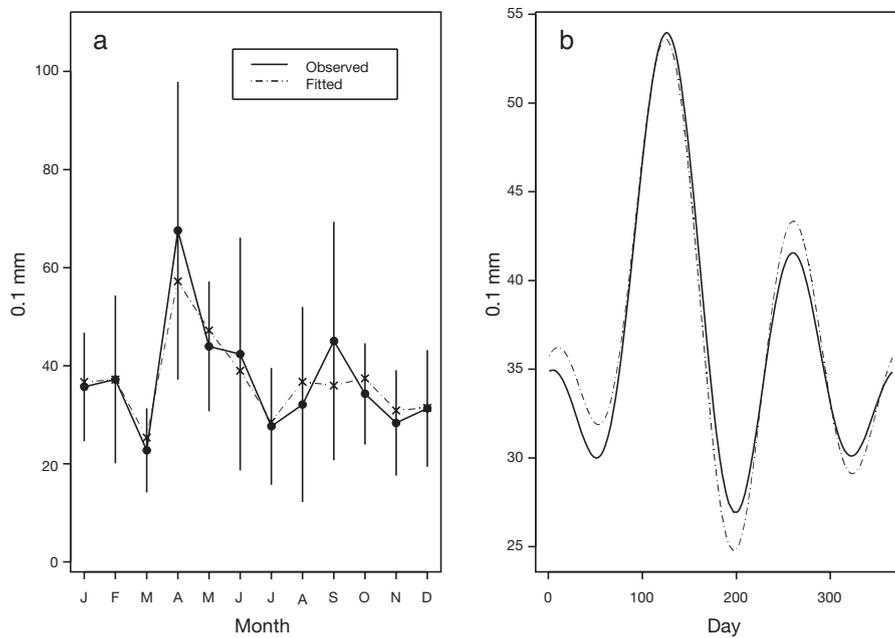


Fig. 6. Observed rainfall amounts and fitted values using the Down–Q model and 1990–1997 observed covariates. (a) Monthly means; (b) lowess of daily values

Table 4. Goodness-of-fit measures for the Down–Oc model in the estimation period (1990–1997); values obtained in the validation period (1998–2000) are given within parentheses. AIC, BIC: Akaike and Bayesian information criteria

	Deviance	df		
Null model	2349.76	2152		
Fitted model	1475.52	2123		
AIC: 1518.83 BIC: 1683				
Hosmer-Lemeshow stat.: 9.15 (19.74), p-value: 0.33 (0.01)				
Explained variance (%): 39.61 (32.62)				
	Spring	Summer	Autumn	Winter
	36.08 (23.01)	38.48 (40.72)	40.66 (30.42)	40.57 (39.47)
Lag 1 residual autocorrelation: 0.08 (0.10)				
Correctly classified days (%) Global: 78.63 (80.46)				
	Spring	Summer	Autumn	Winter
	79.96 (75.23)	79.04 (87.58)	76.89 (77.14)	78.6 (81.92)
		Day $t-1$:	Dry	Wet
Day t	Dry	82.01 (86.48)	58.02 (67.09)	
	Wet	75.56 (66.67)	85.65 (77.97)	
Monthly wet day frequencies: observed vs. fitted				
Correlation coef.: 0.98 (0.72) Efficiency coef.: 0.96 (0.23)				
Index of agreement: 0.99 (0.79) Error mean: 0.45 % (14.71 %)				

the 5 % level for February wet runs only. For rainfall quantities, the test is statistically significant at the 1 % level for January, December and spring months, and at the 5 % level from September to November. The simulated distributions show greater dispersion in April and May while the opposite is true in January and March for observed data.

We conclude that the Down model adequately represents many of the characteristics of observed rainfall. The occurrence model reproduces the monthly frequencies and the dry and wet run distributions. With regard to amounts, the model reproduces the monthly mean values well and the distributions are not distinguishable at the 5 % level in summer months and at the 1 % level in autumn months; some discrepancies are found in the other seasons.

3.3. Evaluation of the Down model using CGCM1 data

An overview of the model performance is displayed in Table 7, which shows, for the period 1990–2000, the

annual and seasonal mean number of rainfall days and the mean accumulated amount, observed and fitted, using as input (1) the observed covariates and (2) the CGCM1 generated ones for this period. The expected number of rainfall days is obtained by adding the fitted probabilities; for the expected amount we add the products of daily fitted probabilities by fitted amounts. The errors are calculated as a percentage of the observed value.

The maximum error observed in rainfall frequencies using CGCM1 data is about 8 % and occurs in spring (Fig. 7); the discrepancy between the lowess smoother applied to the observed and the estimated frequencies is small.

Satisfactory results are also obtained when the aforementioned diagnostic checks are applied to the 20 simulated series obtained using the 1990–2000 CGCM1 generated data (Table 6). With regard to dry and wet run persistence (results not shown), the KS test suggests that only June dry runs are significantly different at the 5 % level, the simulated runs being a little longer; no difference is detected in wet run distributions.

The monthly mean values of simulated amounts are close to the ones observed (Fig. 8), with a maximum discrepancy in September. Nevertheless, the KS test

Table 5. Goodness-of-fit measures for the Down–Q model

	Deviance	df		
Null model	938.16	529		
Fitted model	733.7	511		
AIC : 5408.46 BIC: 5485.34				
Explained variance (%): 17.44 (4.93)				
	Spring	Summer	Autumn	Winter
	13.58 (10.44)	23.3 (0)	10.75 (22.67)	22.7 (0)
Lag 1 residual autocorrelation: 0.01 (–0.01)				
Monthly amount means: observed vs. fitted				
Correlation coef. 0.93 (0.59) Efficiency coef. 0.83 (–0.20)				
Index of agreement 0.94 (0.72) Error mean 1.20 % (21.7 %)				
Observed amount classification (%), defined by fitted distribution quantiles (q)				
$<q_{0.10}$	$[q_{0.10}, q_{0.25}]$	$[q_{0.25}, q_{0.75}]$	$(q_{0.75}, q_{0.90}]$	$>q_{0.90}$
5.67 (9.09)	14.93 (19.48)	55.20 (50.65)	14.56 (9.09)	9.64 (11.69)
Classification results when $Y_{t-1} = 0$				
6.87 (8.51)	17.18 (19.15)	51.55 (54.26)	15.81 (8.51)	8.59 (9.57)
Classification results when $Y_{t-1} = 1$				
4.20 (10.00)	12.18 (20.00)	59.66 (45.00)	13.03 (10.00)	10.92 (15.00)
% days into	Spring	Summer	Autumn	Winter
$(q_{0.10}, q_{0.90})$	80.99 (77.14)	89.58 (75.86)	84.47 (82.35)	84.77 (84.21)

Table 6. Diagnostic measures (Wilby et al. 1998) for comparing observed and Down model simulated series for the period 1990–2000. Right column: results obtained for 2090–2100 projection

Diagnostic statistic	Observed	— Model simulated data —		
		Observed covar.	GCM data 1990–2000	GCM data 2090–2100
Mean wet day amount	36.67	37.02	33.66	38.47
SD wet day amount	54.99	58.33	52.96	69.22
Median wet day amount	16.00	16.38	15.00	15.70
Maximum daily precipitation	413.0	946.4	1159.0	1842.0
$p_{00} = P(Y_t = 0 Y_{t-1} = 0)$	0.826	0.826	0.833	0.827
$p_{11} = P(Y_t = 1 Y_{t-1} = 1)$	0.434	0.417	0.418	0.430
$p_w = P(Y_t = 1)$	0.235	0.230	0.223	0.232
Mean wet-spell length	1.79	1.72	1.72	1.75
Mean dry-spell length	5.88	5.73	5.60	5.80
Frequency of dry-spells > 10 d	0.142	0.153	0.171	0.142

rejects the equality hypothesis at the 5% level in 10 months and at the 1% level in 6 (results not shown), suggesting that the model Down-Q must be improved.

To sum up, the Down model is able to generate, using CGCM1 data as input, rainfall series which are indistinguishable from the observed data with regard to the frequency properties; for the amounts, the model produces similar monthly mean values but it is not able to reproduce the distributions. Keeping these limitations in mind, we will use the Down model for analysing changes in rainfall properties in a climate change scenario.

Table 7. Comparison between observed values and different estimations, using the Down model, of the annual and seasonal precipitation frequency and amount for the periods 1990–2000 and 2090–2100. Also shown are results directly obtained from the CGCM1 model

	Annual	Winter	Spring	Summer	Autumn
Observed and fitted values using observed data					
Wet day frequency					
Observed	82.1	22.0	22.5	15.0	22.8
Fitted	82.4	22.3	20.8	14.7	24.5
% change	0.36	1.1	-7.4	-1.8	7.8
Rainfall amount (10^{-1} mm)					
Observed	2920.0	648.0	906.0	518.0	850.0
Fitted	2950.0	760.0	900.0	507.0	786.0
% change	1.2	17.2	-0.66	-2.1	-7.6
Fitted values using 1990–2000 GCM data					
Wet day frequency					
Fitted	80.8	21.2	20.7	14.6	24.2
% change	-1.6	-3.6	-8.0	-2.3	6.2
Rainfall amount (10^{-1} mm)					
Fitted	2740.0	608.0	856.0	485.0	791.0
% change	-6.2	-6.2	-5.6	-6.3	-7.0
Projected values for 2090–2100					
Wet day frequency					
Projected	84.8	25.1	13.2	13.6	32.9
% change/observed	3.2	14.1	-41.3	-9.3	44.5
% change/GCM 1990–2000	4.9	18.4	-36.2	-6.8	36.0
Rainfall amount (10^{-1} mm)					
Projected	3260.0	745.0	764.0	686.0	1070.0
% change/observed	11.7	14.9	-15.7	32.4	25.9
% change/GCM 1990–2000	19.0	22.5	-10.7	41.4	35.3
CGCM1 output rainfall data					
Wet day frequency					
GCM 1990–2000	240.9	80.7	75.8	36.5	47.8
GCM 2090–2100	227.3	81.2	67.5	32.0	46.6
% change	-5.62	0.6	-10.9	-12.4	-2.5
Rainfall amount (10^{-1} mm)					
GCM 1990–2000	4999.9	2022.6	1512.2	636.1	827.8
GCM 2090–2100	4745.3	2162.9	1102.1	629.0	851.5
% change	-5.07	6.94	-27.1	-1.13	2.86

4. PROJECTED RAINFALL PATTERN FOR THE PERIOD 2090–2100 UNDER AN IS92a SCENARIO

In this section we analyse the rainfall pattern predicted by the Down model for the period 2090–2100 in the area represented by the Zaragoza observatory; the input is the CGCM1 projected data under the IS92a scenario.

The climate projection proposed by CGCM1 indicates an increase of about 4°C in T_{MAX} and T_{MIN} mean values at the surface level and greater increases at higher levels: 4.5°C at 850 hPa and 6°C at 500 hPa; the standard deviation would increase 10% at the surface level for both T_{MAX} and T_{MIN} and also for the temperature at the 850 hPa level.

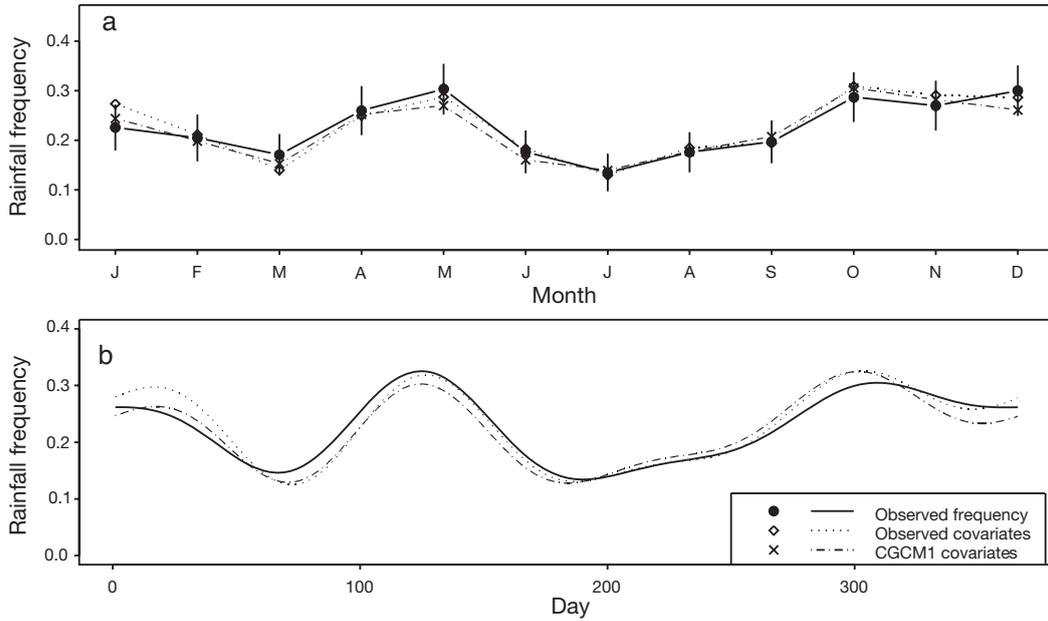


Fig. 7. Wet day observed relative frequencies and rainfall probabilities fitted by the Down model using (1990–2000) observed and CGCM1 generated covariates. (a) Monthly means; (b) lowess of daily values

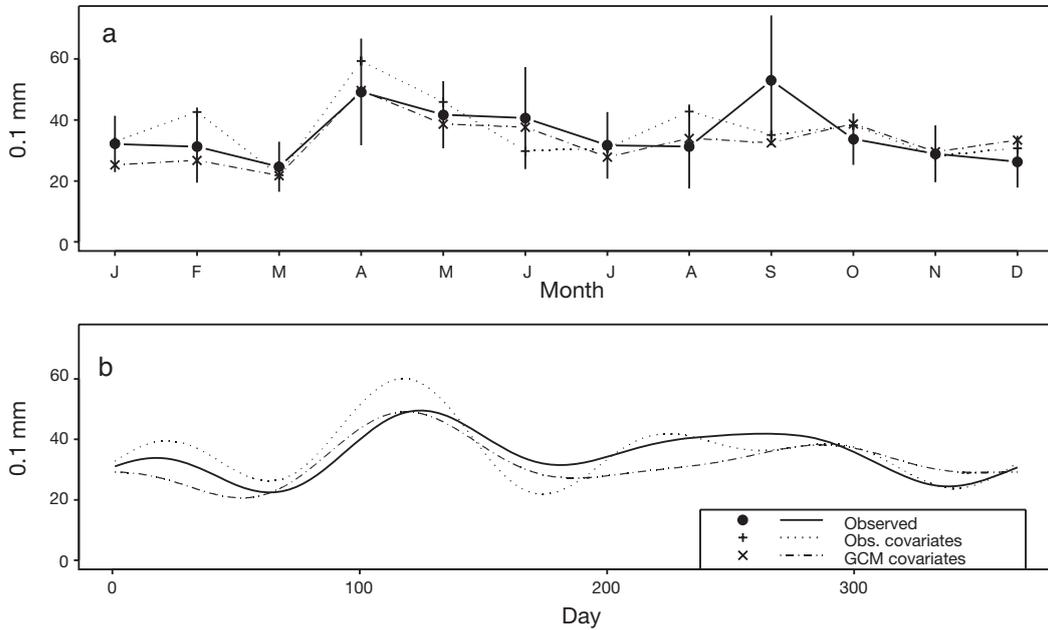


Fig. 8. Observed amounts and Down fitted values using (1990–2000) observed and CGCM1 generated covariates. (a) Monthly means; (b) lowess of daily values

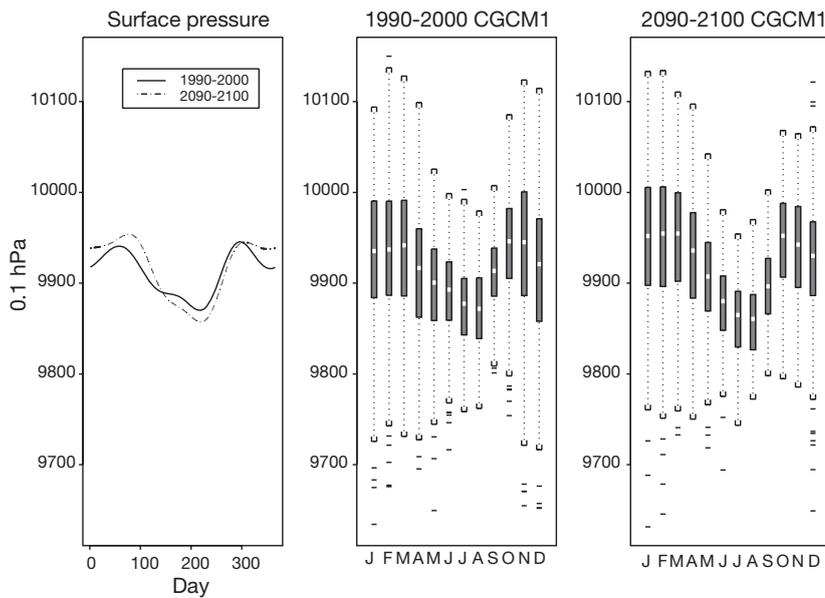


Fig. 9. Lowess and monthly boxplots of surface pressure data generated by CGCM1 for the periods 1990–2000 and 2090–2100

According to this projection, the pressure distribution would not experience significant changes in location and variability (Fig. 9). The specific humidity would keep its seasonal profile and increase the mean level at the 3 heights, the rise being greater at higher levels: 53.1% at 500 hPa (Fig. 10), 26.3% at 850 hPa and 22.5% at the surface level; its variability would also increase in all the cases. This scenario has already been described: the rise of temperature would increase the evapotranspiration and, consequently, the water vapor in the atmosphere.

The annual number of rainy days for 2090–2100 forecasted by the Down model is similar to the currently observed values. However, a strong reduction is projected for spring months, a significant increase in September and other minor ones in August and October (Fig. 11).

The analysis of 20 simulated trajectories allows us to evaluate the change in wet and dry run length distributions. The results (not shown) support the seasonal profile already discussed; the KS test asserts the significant difference, at the 5% level, of the dry run length distribution in April, May, June and September. For wet runs, the Wilcoxon test states that significant differences exist, at the 5% level, between February and May and also

for September; in March, the third quartile is about 1, which means that 75% of the forecasted rainy spells will be formed by isolated days.

There are 2 periods, spring (especially April) and August, where an increase in precipitation intensity is forecasted (Fig. 12). In other months, the simulated mean values are similar to those currently observed. The diagnostic checks shown in Table 6 do not detect any discrepancies between observed and forecasted regimes, because the most important difference is associated with the seasonal cycle.

The forecasted rainfall pattern is summarised in Table 7. The changes in annual rainfall frequency and in total amount would be minor but the seasonal cycle would change significantly. Rainfall frequency would fall by 37% in spring, with respect to values proposed by the model for the cur-

rent period; the number of wet days proposed would be similar to the number currently observed in summer. Spring is the only season where the model forecasts an amount decrease; the drop is less important than the one corresponding to the frequency because an increase in the rainfall intensity is forecasted. The model proposes an annual cycle with a wet phase in autumn and winter; in this period the rainfall frequency would be greater than the current values and

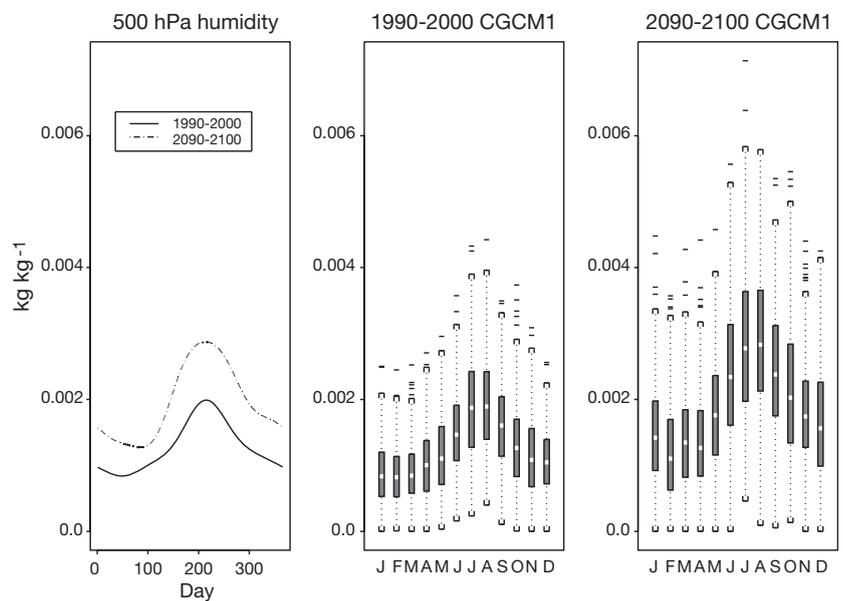


Fig. 10. Lowess and monthly boxplots of 500 hPa specific humidity data generated by CGCM1 for the periods 1990–2000 and 2090–2100

the total amount would increase by 20%. In contrast, the dry period, between March and August, would show a minor frequency of wet days and a total amount similar to the current values.

The comparison of the precipitation change projected by the downscaling model with that obtained directly from the GCM output (see the bottom of Table 7) shows

some agreement on the summer frequency and also on the direction of the change proposed for the spring frequency and for winter and spring amounts; nevertheless, the relevant change of the seasonal cycle projected by the Down model is not forecasted by CGCM1.

We do not know of any downscaling study which has obtained precipitation projections for the middle

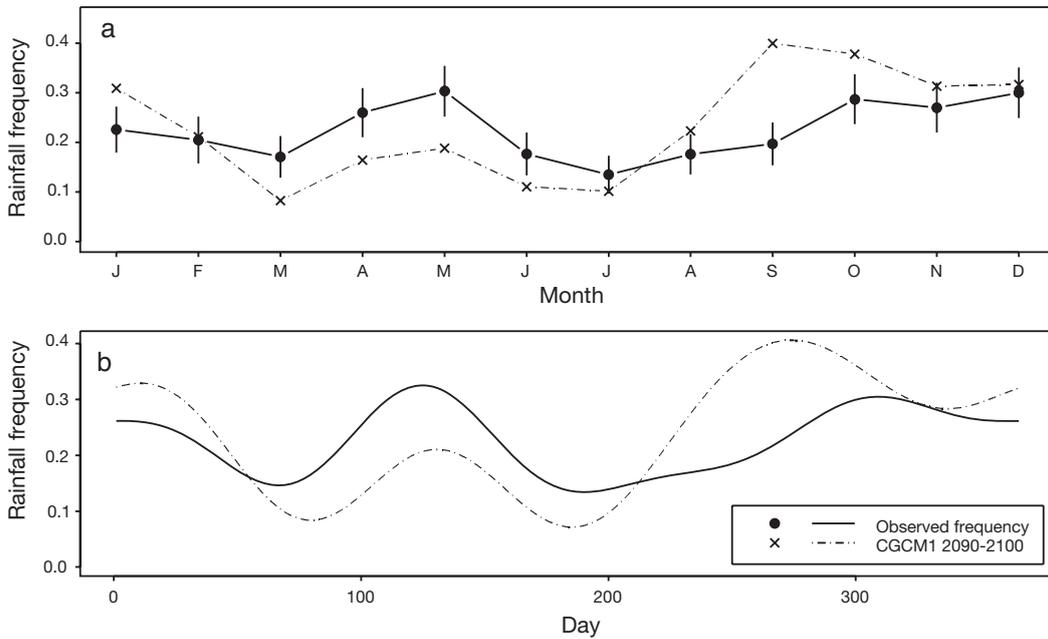


Fig. 11. Wet day observed relative frequencies and rainfall probabilities fitted by the Down model using (2090–2100) CGCM1 generated covariates. (a) Monthly means; (b) lowess of daily values

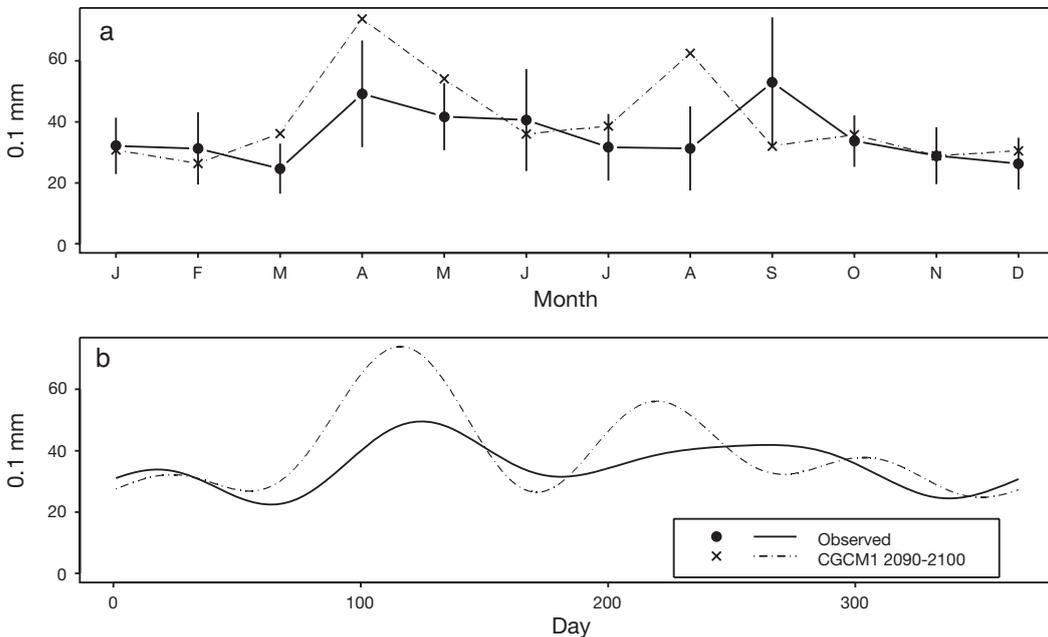


Fig. 12. Observed amounts and Down projected values for the period 2090–2100. (a) Monthly means; (b) lowess of daily values

region of the Ebro valley, so in order to establish some comparison with the results obtained we discuss the projections proposed by other authors for 2 nearby areas. Beckmann & Buishand (2001), using ECHAM-OPYC3 model data, estimated the change in winter and summer seasonal mean values of the rainfall frequency and amount, between the periods 1968–1997 and 2070–2099, for the Salto de Bolarque area (Iberian South Meseta), located 220 km south-southwest of Zaragoza. They projected a decrease in the number of wet days, especially in summer, and a relatively smaller decrease in both seasonal amounts because of the effect of the change in the geopotential height at 850 hPa, which would raise the wet day amount. Our results for the Zaragoza area agree with the frequency decrease forecasted for summer and with the increase, in both seasons, of rainfall intensity.

Sumner et al. (2003) analysed the rainfall episodes which generate amounts greater than 5 mm in at least 5 % of the considered stations, for the Spanish Mediterranean coastal region. Using the data generated for the period 2080–2099 by the ECHAM-OPYC3 model, they project for the Ebro delta area, 180 km southeast of Zaragoza, an increase in convective situations at the end of the summer and autumn and no changes in the annual amount, 2 characteristics which we find in the projection for the Zaragoza area.

5. CONCLUSIONS

The method proposed is useful for obtaining long-term projections for the daily rainfall pattern in a climate change scenario at a local scale. The current Down model is able to simulate rainfall series from which we can obtain conclusions about the seasonal characteristics and, also, about certain aspects of the daily behaviour, such as dry and wet run persistence. The inaccuracy of the quantity model does not allow us, for the moment, to be confident about conclusions on other properties such as characterisation of extreme events. Thus, we have restricted our analysis to rainfall frequency and to seasonally aggregated results for the amounts. These results could probably be improved by using another GCM with a better spatial resolution or providing a more informative output.

Our method has some advantages compared to other approaches: (1) it is able to use the GCM daily output directly; (2) it analyses the way the GCM reproduces the observed climatology at the scale where the projection is made, the daily scale; (3) the regression model provides a more flexible approach than the models based on the atmospheric state. It allows us to use any variable well described by the GCM instead of being limited to pressure-related information. The

presence in the model of other variables let us take into account different aspects of climate change, which increases the method sensitivity. Moreover, our model generates probabilities and expected amounts varying in a continuous range whereas models based on the climate state provide values only for the defined situations.

The modelling process has proved the relevance of variables such as the specific humidity at the surface and 500 hPa levels, as stated by Beckmann & Buishand (2001, 2002), and also of the temperature daily range, a variable not used in previous works. The suitability of the model for climate long-term projection has been proved by analysing the results obtained using the GCM generated variables corresponding to the present climate as input; this checking is important since the posterior application requires the minimal bias from the model.

As in all prediction methods, this procedure needs the established relationship for the current climate between rainfall and atmospheric variables to stay in the climate change scenario. In favor of the Down model we can say that the apparent inflexibility of its seasonal profile, described by harmonic terms that do not change in the period 2090–2100, has not prevented the projection of an important change in rainfall seasonality.

Statistical modelling supposes that non-explained variability follows a random behaviour. As we have seen, the models built could not include relevant information and were forced to employ anomalies as covariates because the corresponding variables were not properly described by the GCM. The solution to this problem is out of the scope of our research; the bias associated will tend to disappear when future GCM versions that properly describe a greater number of atmospheric variables are available. The projecting procedure proposed will still be useful and the reliability of the results will improve.

Acknowledgements. The authors thank the Instituto Nacional de Meteorología (INM) of Spain for providing the observed data and the CCCma for providing CGCM1 output data. This work was sponsored by the Spanish Ministry of Science and Technology with the grant REN2002-00009.

LITERATURE CITED

- Abaurrea J, Asín J (2003) A method for statistical downscaling of local daily rainfall. Contribution to the ISI International Conference on Environmental Statistics and Health, Santiago de Compostela, Spain, July 16–18, 2003
- Asín J (2003) Modelling daily precipitation in a Generalized Linear Models frame. PhD dissertation, University of Zaragoza
- Beckmann BR, Buishand TA (2001) KNMI contribution to the

- European project WRINCLE: downscaling relationships for precipitation for several European sites. Koninklijk Nederlands Meteorologisch Instituut, De Bilt
- Beckmann BR, Buishand TA (2002) Downscaling relationships for precipitation for the Netherlands and North Germany. *Int J Climatol* 22:15–32
- Beersma JJ, Buishand TA (2003) Multi-site simulation of daily precipitation and temperature conditional on the atmospheric circulation. *Clim Res* 25:121–133
- Boer GJ, Flato GM, Reader MC, Ramsden D (2000a) A transient climate change simulation with greenhouse gas and aerosol forcing: experimental design and comparison with the instrumental record for the 20th century. *Clim Dyn* 16:405–425
- Boer GJ, Flato GM, Reader MC, Ramsden D (2000b) A transient climate change simulation with greenhouse gas and aerosol forcing: projected climate for the 21st century. *Clim Dyn* 16:427–450
- Buishand TA, Klein Tank AMG (1996) Regression model for generating time series of daily precipitation amounts for climate change impact studies. *Stochastic Hydrol Hydraul* 10:87–106
- Cavazos T (2000) Using self-organizing maps to investigate extreme climate events: an application to wintertime precipitation in the Balkans. *J Climate* 13:1718–1732
- Goodess CM (2000) The construction of daily rainfall scenarios for Mediterranean sites using a circulation-type approach to downscaling. PhD dissertation, University of East Anglia, Norwich
- Hughes JP, Guttorp P, Charles SP (1999) A non-homogeneous hidden Markov model for precipitation occurrence. *Appl Stat* 48:15–30
- McCullagh P, Nelder JA (1989) *Generalized linear models*. Chapman & Hall, London
- Palutikof JP, Goodess CM, Watkins SJ, Holt T (2002) Generating rainfall and temperature scenarios at multiple sites: examples from the Mediterranean. *J Climate* 15:3529–3548
- Stehlik J, Bardossy A (2002) Multivariate stochastic downscaling model for generating daily precipitation series based on atmospheric circulation. *J Hydrol* 256:120–141
- Stern RD, Coe R (1984) A model fitting analysis of daily rainfall data. *J R Statist Soc A* 147:1–34
- Sumner GN, Romero R, Homar V, Ramis C, Alonso S, Zorita E (2003) An estimate of the effects of climate change on the rainfall of Mediterranean Spain by the late twenty first century. *Clim Dyn* 20:789–805
- Wilby RL, Wigley TML, Conway D, Jones PD, Hewitson BC, Main J, Wilks DS (1998) Statistical downscaling of general circulation model output: a comparison of methods. *Water Resour Res* 34:2995–3008
- Wilby RL, Hay LE, Leavesley GH (1999) A comparison of downscaled and raw GCM output: implications for climate change scenarios in the San Juan River basin, Colorado. *J Hydrol* 225:67–91

*Editorial responsibility: Otto Kinne,
Oldendorf/Luhe, Germany*

*Submitted: March 25, 2004; Accepted: March 13, 2005
Proofs received from author(s): May 13, 2005*