Climate Dynamics manuscript No.
(will be inserted by the editor)

Autoregressive Logistic Regression Applied to Atmospheric Circulation Patterns

Y. Guanche · R. Mínguez · F. J. Méndez

Received: date / Accepted: date

Abstract Autoregressive logistic regression (ALR) models have been successfully applied in medical and pharmacology research fields, and in simple models to analyze weather types. The main purpose of this paper is to introduce a general framework to study atmospheric circulation patterns capable of dealing simultaneously with: seasonality, interannual variability, long-term trends, and autocorrelation of different orders. To show its effectiveness on modeling performance, daily atmospheric circulation patterns identified from observed sea level pressure (DSLP) fields over the Northeastern Atlantic, have been analyzed using this framework. Model predictions are compared with probabilities from the historical database, showing very good fitting diagnostics. In addition, the fitted model is used to simulate the evolution over time of atmospheric circulation patterns using Monte Carlo method. Simulation results are statistically consistent with respect to the historical sequence in terms of i) probability of occurrence of the different weather types, ii) transition probabilities and iii) persistence. The proposed model constitutes an easy-to-use and powerful tool for a better understanding of the climate system.

Keywords Autoregressive logistic regression · Circulation Patterns · Simulation

R. Mínguez
Environmental Hydraulics Institute, “IH Cantabria”, Universidad de Cantabria, C/ Isabel Torres n 15, Parque Científico y Tecnológico de Cantabria Santander 39011, SPAIN
Tel.: +34 942 20 16 16
Fax: +34 942 26 63 61
E-mail: minguezsr@unican.es

1 Introduction

The study of atmospheric patterns, weather types or circulation patterns, is a topic deeply studied by climatologists, and it is widely accepted to disaggregate the atmospheric conditions over regions in a certain number of representative states. This consensus allows simplifying the study of climate conditions to improve weather predictions and a better knowledge of the influence produced by anthropogenic activities on the climate system [15–17, 29].

The atmospheric pattern classification can be achieved by using either manual or automated methods. Some authors prefer to distinguish between subjective and objective methods. Strictly speaking, both classifications are not equivalent because, although automated methods could be regarded as objective, they always include subjective decisions. Among subjective classification methods and based on their expertise about the effect of certain circulation patterns, [13] identify up to 29 different large scale weather types for Europe. Based on their study, different classifications have been developed, for instance, [10], [11] and [36] among others. To avoid the possible bias induced by subjective classification methods, and supported by the increment of computational resources, several automated classification (clusterization) methods have been developed, which may be divided into 4 main groups according to their mathematical fundamentals: i) threshold based (THR), ii) principal component analysis based (PCA), iii) methods based on leader algorithms (LDR), and iv) optimization methods (OPT). A detailed description of all these methods and their use with European circulation patterns can be found in [29].

Once the atmospheric conditions have been reduced to a catalogue of representative states, the next step is to develop numerical models for a better understanding of the weather dynamics. An appropriate modeling of weather dynamics is very useful for weather predictions, to study the possible influence of well-known synoptic patterns such as East Atlantic (EA), North Atlantic Oscillation (NAO), Southern Oscillation Index (SOI), etc., as well as to analyze climate change studying trends in the probability of occurrence of weather types, and so on. For example, [33] investigated long term trends in annual frequencies associated with weather types, demonstrating the utility of weather classification for climate change detection beyond its short-term prognosis capabilities. [26] studied the dynamics of weather types using 1st order Markovian and non-Markovian models, however seasonality is not considered. [19] introduced a seasonal Markov chain model to analyze the weather in the central Alps considering three weather types. The transition probabilities are determined using a linear logit regression model. [27] implemented a cyclic Markov chain to introduce the influence of the El Niño-Southern Oscillation (ENSO).

Generalized linear regression, and especially autoregressive logistic regression, has proved to be a promising framework for dealing with seasonal Markovian models, and not only for atmospheric conditions. Similar models have been applied successfully in medical and pharmacological research fields [35,2,30]. The main advantages of autoregressive logistic regression (ALR) are that i) it can be used to model polytomous outcome variables, such as weather types, and ii) standard statistical software can be used for fitting purposes.

The aim of this paper is twofold; firstly, to introduce autoregressive logistic regression models in order to deal with weather types analysis including: seasonality, interannual variability in the form of covariates, long-term trends, and Markov chains; and secondly, to apply this model to the Northeastern Atlantic in order to show its potential for analyzing atmospheric conditions and dynamics over this area. Results obtained show how the model is capable of dealing simultaneously with predictors related to different time scales, which can be used to predict the behaviour of circulation patterns. This may constitute a very powerful and easy-to-use tool for climate research.

The rest of the paper is organized as follows. Section 2 provides the description of Autoregressive Logistic Models. In Section 3 the model is applied to the Northeastern Atlantic, interpreting results related to the different scales, and checking the model's performance on transition probabilities and persistence. Finally, Section 4 contains a summary and discussion on model performance, possible limitations and further applications.

2 Autoregressive Logistic Model

Traditional uni- or multivariate linear regression models assume that responses (dependent variables or outcomes) are normally distributed and centered at a linear function of the predictors (independent variables or covariates). For some regression scenarios, such as the case considered in this paper, this model is not adequate because the response variable Y is categorical and its possible outcomes are associated with each weather type ($Y \in \{1,2,\ldots,n_{wt}\}$ being n_{wt} the number of weather types), which are not normally distributed. Thus the necessity to dispose of alternative regression models.

Logistic regression was originally defined as a technique to model dependent binary responses ([7,3]). The likelihood of the binary dependent outcome is expressed as the product of logistic conditional probabilities. [25] introduced the capability of dealing with transition probabilities using Markov chains, which was further explored by [35] to predict the outcome of the supervised exercise for intermittent claudication, extending the model to polytomous outcomes.

Let $Y_t;\ t=1,\ldots,n$ be the observation weather type at time t, with the following possible outcomes $Y_t \in \{1,\ldots,n_{wt}\}$ related to each weather type. Considering $X_t;\ t=1,\ldots,n$ to be a time-dependent row vector of covariates with dimensions $(1\times n_c)$, i.e. seasonal cycle, NAO, SOI, principal components of synoptic circulation, long-term trend, etc., the autoregressive logistic model is stated as follows:

$$\ln\left(\frac{\mathrm{Prob}(Y_t=i|Y_{t-1},\ldots,Y_{t-d},X_t)}{\mathrm{Prob}(Y_t=i^*|Y_{t-1},\ldots,Y_{t-d},X_t)}\right) = \alpha_i + X_t\beta_i + \sum_{j=1}^{d} Y_{t-j}\gamma_{ij};\ \forall i=1,\ldots,n_{wt}|i\neq i*, \quad (1)$$

where α_i is a constant term and β_i $(n_c\times 1)$ and γ_{ij} $(j=1,\ldots,d)$ correspond, for each possible weather type i, to the parameter vectors associated with covariates and d-previous weather states, respectively. Note that d corresponds to the order of the Markov model. The model synthesized in equation (1) provides the natural logarithm of the probability ratio between weather type i and the reference weather type i^*, conditional on covariates X_t and the d previous weather states, i.e. the odds. The left hand side of equation (1) is also known as *logit*. According to this expression, the conditional probability for any weather type is given by:

$$\mathrm{Prob}(Y_t=i|Y_{t-1},\ldots,Y_{t-d},X_t) = \frac{\exp\left(\alpha_i + X_t\beta_i + \sum_{j=1}^{d} Y_{t-j}\gamma_{ij}\right)}{\sum\limits_{k=1}^{n_{wt}} \exp\left(\alpha_k + X_t\beta_k + \sum\limits_{j=1}^{d} Y_{t-j}\gamma_{kj}\right)};\ \forall i=1,\ldots,n_{wt}. \quad (2)$$

Note that in order to make parameters unique we impose an additional condition, which fixes the parameter values related to the reference weather i^* (arbitrary chosen) to zero.

2.1 Description of the parameters

Since the purpose of this paper is to present a unique model able to reproduce different weather dynamic characteristics, including: seasonality, covariates influence, long-term trends, and Markov chains; the inclusion of these features in the model (1) will be briefly described in this subsection:

- **Seasonality:** It is known that there is a strong seasonality on weather type frequencies, for example, [19] modeled this effect for the weather in the central Alps. In their work the seasonality is introduced in the model as an autoregressive term but it could be also introduced by adding harmonic factors. Here, the seasonality is introduced in the model using harmonics as follows:

$$\pi^S = \beta_0^S + \beta_1^S \cos(wt) + \beta_2^S \sin(wt), \tag{3}$$

where π^S represents the seasonality effect on the *logit*, t is given in years, β_0^S correspond to annual mean values, and β_1^S and β_2^S are the amplitudes of harmonics, $w = 2\pi/T$ is the angular frequency. Since β_0^S is a constant term, it replaces the independent term α_i in (1). For this particular case, we choose T to be defined in years, and thus $T = 1$ and t is in annual scale. This means, for instance, that the time associated with day 45 within year 2000 is equal to $2000 + 45/365.25 = 2000.1232$. However, according to the definition of the harmonic argument ($wt = \frac{2\pi t}{T}$), t could be given in days, then T must be equal to 365.25.
Analogously to Autoregressive Moving Average (ARMA) models [4], seasonality can also be incorporated trough an autoregressive term at lag 365. Details about how to incorporate this autoregressive component are given in the autoregressive or Markov chain parameters description below.

- **Covariates:** To introduce the effect of different covariates, the model is stated as follows:

$$\pi^C = X\beta^C = (X_1, \ldots, X_{n_c}) \begin{pmatrix} \beta_1^C \\ \vdots \\ \beta_{n_c}^C \end{pmatrix} = \sum_{i=1}^{n_c} X_i \beta_i^C, \tag{4}$$

where π^C is the covariates effect on the *logit*, X is a row vector including the values of different n_c covariates considered (SOI, NAO, monthly mean sea level pressure anomalies principal components, etc.), and β^C is the corresponding parameter vector.

- **Long-term trends:** The long-term trend is a very important issue because many authors, such as [12,17,5], perform a linear regression analysis using as predictand the probabilities of each weather type, and the time as predictor. However, mathematically speaking, this may conduct to inconsistencies, such as probabilities outside the range 0 and 1, which is not possible. To avoid this shortcoming, we use a linear regression model but for the logits, being considered as a particular case of covariate:

$$\pi^{LT} = \beta^{LT} t, \tag{5}$$

where π^{LT} represents the long-term trend effect on the *logit*, and t is given in years. The parameter represents the annual rate of change associated with the logarithm of the probability for each weather type, divided by the probability of the reference weather type, i.e. $\Delta \log \frac{p_i}{p_i^*}$. The regression coefficient β^{LT} is a dimensionless parameter, which for small values of the coefficient may be interpreted as the relative change in the odds $\frac{\delta p_i}{p_i^*}$ due to a small change in time δt. Note that (5) does not correspond to the typical trend analysis because trends are analyzed on logits. However, as numerical results show, this codification provides consistent results on long-term changes of the weather type probabilities.

- **Autoregressive or Markov chain:** The sequence of atmospheric circulation patterns can be described as a Markov chain. [19] proved that a first order autoregressive logistic model is appropriate for reproducing the weather types in the central Alps. This effect can be included in the model using the following parameterization:

$$\pi^{AR_d} = \sum_{j=1}^{d} Y_{t-j}\gamma_j, \tag{6}$$

where π^{AR_d} represents the autoregressive effect of order d on the *logit*. The order d corresponds to the number of previous states which are considered to influence the actual weather type, Y_{t-j} is the weather type on previous j-states, and γ_j is the parameter associated with previous j-state.
Note that each $Y_{t-j};\ j = 1, \ldots, d$ in (6) corresponds to a different weather type, according to the polytomous character of the variable. In order to facilitate parameter estimation using standard logistic regression techniques, the autoregressive parts must be transformed using a contrast matrix, such as the Helmert matrix [35] so that each covariate Y_{t-j} transforms into $n_{wt}-1$ dummy variables Z^{t-j}. The Helmer contrast matrix for transforming outcome Y_t into the dummy variable row vector Z^t is provided in Table 1. According to this transformation matrix, equation (6) becomes:

$$\pi^{AR_d} = \sum_{j=1}^{d} Y_{t-j}\gamma_j = \sum_{j=1}^{d} \sum_{k=1}^{n_{wt}-1} Z_k^{t-j} \gamma_{jk}. \tag{7}$$

Regarding seasonality, and according to expression (7), it can be included in the model as follows:

$$\pi^{AR_{365}} = Y_{t-365}\gamma_{365} = \sum_{k=1}^{n_{wt}-1} Z_k^{t-365} \gamma_{365,k}, \tag{8}$$

which corresponds to an autoregressive component at lag 365.

Y_t	$Z^t\ (1\times(n_{wt}-1))$					
1	-1	-1	-1	$\dots$	-1	-1
2	-1	-1	-1	$\dots$	-1	1
3	-1	-1	-1	$\dots$	2	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$n_{wt}-2$	-1	-1	$n_{wt}-3$	$\dots$	0	0
$n_{wt}-1$	-1	$n_{wt}-2$	0	$\dots$	0	0
n_{wt}	$n_{wt}-1$	0	0	$\dots$	0	0

Table 1 Helmert Contrast Matrix

Note that the prize for using standard logistic regression fitting is an increment on the number of parameters, i.e. from d to $d\times(n_{wt}-1)$.

The model can include all these effects adding the logits, i.e. $\pi=\pi^S+\pi^C+\pi^{LT}+\pi^{AR_d}$. Thus, expression (2) can be expressed as follows:

$$\text{Prob}(Y_t=i|Y_{t-1},\dots,Y_{t-d},X_t)=\frac{\exp\left(\pi_i^S+\pi_i^C+\pi_i^{LT}+\pi_i^{AR}\right)}{\sum_{k=1}^{n_{wt}}\exp\left(\pi_k^S+\pi_k^C+\pi_k^{LT}+\pi_k^{AR}\right)};\ \forall i=1,\dots,n_{wt}. \quad (9)$$

In order to deal with different time-scales within the model: annual, monthly and daily; all the parameters to be included are transformed to the lowest scale considered, i.e. daily. Thus, we require a covariate value for each day. This value may be chosen assuming a piecewise constant function over the data period (a month for monthly data, a year for yearly data, and so on), which is the one considered in this paper, or using interpolation and/or smoothing techniques, such as splines. Note that in our case, the same covariate value keeps constant for the entire month (during 30-31 days).

2.2 Data set-up

Once the mathematical modeling is defined, this section describes the data set-up from the practical perspective. Let Y correspond to the vector of weather types at different times of dimensions $(n\times1)$, so that $Y_t\in\{1,\dots,n_{wt}\}$. To deal with polytomous variables a matrix y of dimensions $(n\times n_{wt})$ is constructed as:

$$y_{tj}=\begin{cases}0 \text{ if } j\neq Y_t\\ 1 \text{ if } j=Y_t\end{cases};\ \forall j=1,\dots,n_{wt};\ \forall t=1,\dots,n. \quad (10)$$

Note that since only one weather type at a time is possible, $\sum_{j=1}^{n_{wt}}y_{tj}=1;\ \forall t$. The matrix x of dimensions $n\times(3+n_c+1+d\times(n_{wt}-1))$ includes all predictors at each of the n observations. Three parameters for seasonality (3), n_c parameters for covariates (4), one parameter for the long term trend (5), and $d\times(n_{wt}-1)$ parameters for the autocorrelation (7). The general data setup for the autoregressive logistic regression applied to weather types is provided in Table 2.

Note that the column associated with the seasonality constant term β_0 in (3), which corresponds to a column vector $(1,1,\dots,1)^T$, must be included in matrix x depending on the standard logistic model used. While some of those models automatically include this constant, others do not.

2.3 Parameter estimation

Parameter estimation is performed using the maximum likelihood estimator, which requires the definition of the likelihood function. For a given sequence of n weather types Y, the likelihood function becomes:

$$\ell(\Theta,Y,X_t)=\prod_{t=1}^{n}\prod_{i=1}^{n_{wt}}\text{Prob}(Y_t=i|Y_{t-1},\dots,Y_{t-d},X_t)^{u_{ti}}, \quad (11)$$

where Θ is the parameter matrix, and the auxiliary variable u_{ti} is equal to:

$$u_{ti}=\begin{cases}0 \text{ if } y_t\neq i\\ 1 \text{ if } y_t=i\end{cases};\ \forall i=1,\dots,n_{wt};\ \forall t=1,\dots,n. \quad (12)$$

Note that the likelihood function (11) is the product of univariate logistic functions.

An important issue for the appropriate modeling of weather types, is to decide whether the inclusion of a covariate is relevant or not. There are several tests and methods to deal with this problem, such as Akaike's information criteria or Wald's test. Further information related to logistic regression parameterization and fitting can be found in [8, 32,34].

There are several statistical software packages which are able to solve a polytomous logistic regression fitting (e.g. `SYSTAT`, `NONMEM`), but for this particular case, the function `mnrfit` in MATLAB has been used. This function estimates the coefficients for the multinomial logistic regression problem taking as input arguments matrices x and y from Table 2.

3 Case study: Weather types in the Northeastern Atlantic

In the last decade, the availability of long term databases (reanalysis, in situ measurements, satellite) allows a detailed description of the atmospheric and ocean variability all over the globe, which include the analysis and study of atmospheric patterns. To show the performance of the proposed model, Daily Sea Level Pressure (DSLP) data from NCEP-NCAR database [20] have been used. The area under study corresponds to the Northeastern Atlantic covering latitudes from 25° to 65°N and longitudes from 52.5°W to 15°E. The data record covers 55 years, from 1957 up to

					x															
											Autocorrelation									
t	Y_t	y			Seasonality		Trend	Covariates			Lag 1			Lag 2			…	Lag d		
t_1	Y_1	$y_{1,1}$	…	$y_{1,n_{wt}}$	$\cos(wt_1)$	$\sin(wt_1)$	t_1	$X_{1,1}$	…	X_{1,n_c}	$Z^{t-1}_{1,1}$	…	$Z^{t-1}_{1,n_{wt}-1}$	$Z^{t-2}_{1,1}$	…	$Z^{t-2}_{1,n_{wt}-1}$	…	$Z^{t-d}_{1,1}$	…	$Z^{t-d}_{1,n_{wt}-1}$
t_2	Y_2	$y_{2,1}$	…	$y_{2,n_{wt}}$	$\cos(wt_2)$	$\sin(wt_2)$	t_2	$X_{2,1}$	…	X_{2,n_c}	$Z^{t-1}_{2,1}$	…	$Z^{t-1}_{2,n_{wt}-1}$	$Z^{t-2}_{2,1}$	…	$Z^{t-2}_{2,n_{wt}-1}$	…	$Z^{t-d}_{2,1}$	…	$Z^{t-d}_{2,n_{wt}-1}$
t_3	Y_3	$y_{3,1}$	…	$y_{3,n_{wt}}$	$\cos(wt_3)$	$\sin(wt_3)$	t_3	$X_{3,1}$	…	X_{3,n_c}	$Z^{t-1}_{3,1}$	…	$Z^{t-1}_{3,n_{wt}-1}$	$Z^{t-2}_{3,1}$	…	$Z^{t-2}_{3,n_{wt}-1}$	…	$Z^{t-d}_{3,1}$	…	$Z^{t-d}_{3,n_{wt}-1}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
t_n	Y_n	$y_{n,1}$	…	$y_{n,n_{wt}}$	$\cos(wt_n)$	$\sin(wt_n)$	t_n	$X_{n,1}$	…	X_{n,n_c}	$Z^{t-1}_{n,1}$	…	$Z^{t-1}_{n,n_{wt}-1}$	$Z^{t-2}_{n,1}$	…	$Z^{t-2}_{n,n_{wt}-1}$	…	$Z^{t-d}_{n,1}$	…	$Z^{t-d}_{n,n_{wt}-1}$

Table 2 Data Setup for the Autoregressive Logistic Regression applied to Weather Types

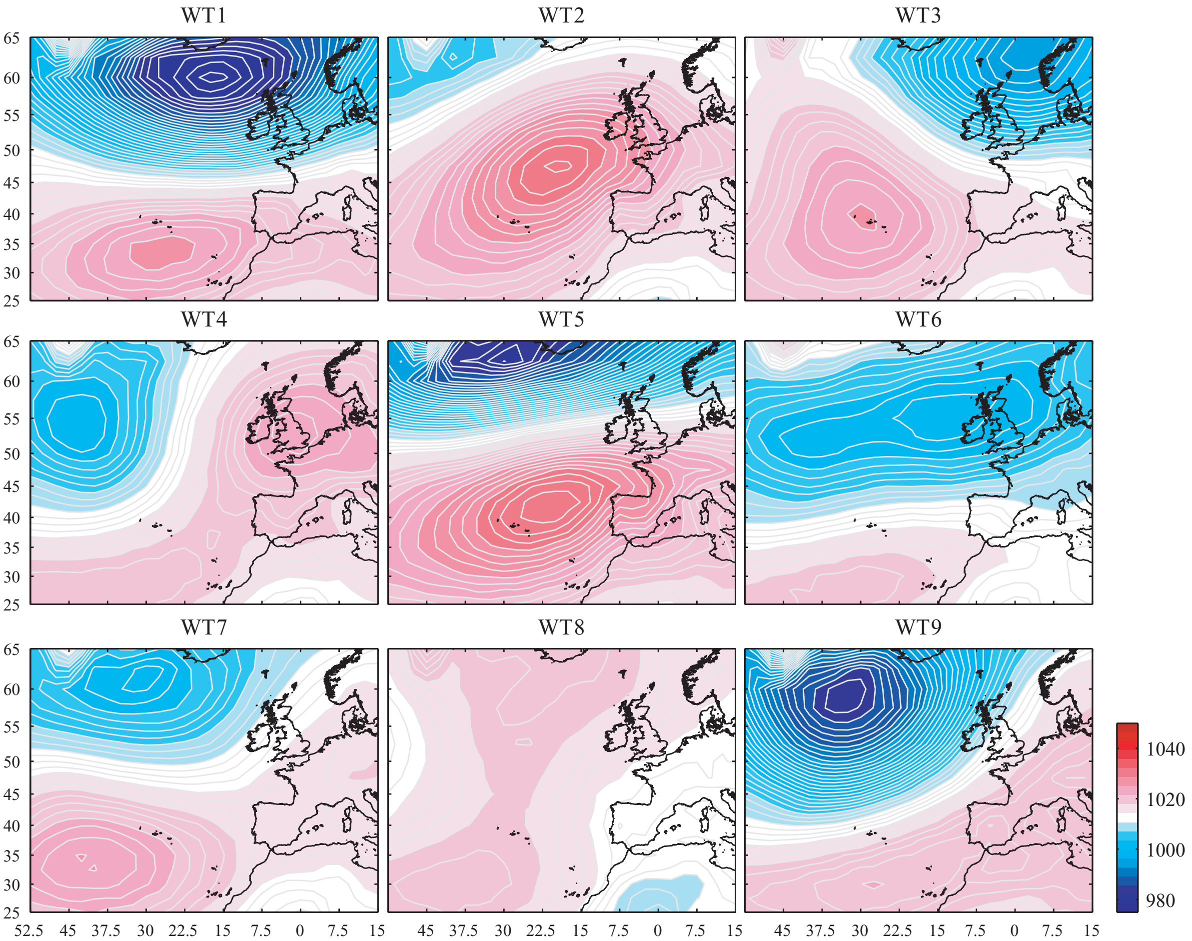

Fig. 1 DSLP synoptical patterns associated with the clusterization.

2011. Note that NCEP-NCAR data records start in 1948, however it is accepted by the scientific community that recorded data up to 1957 is less reliable [21].

The first step to apply the proposed method is data clustering. However, in order to avoid spatially correlated variables that may disturb the clusterization, a principal components analysis is applied to the daily mean sea level pressures (DSLP). From this analysis, it turns out that 11 linearly independent components represent 95% of the variability.

As proposed by several authors, such as [6,9] and [22] among others, the non-hierarchical K-means algorithm is able to classify multivariate patterns into a previously determined number of groups, eliminating any subjectivity in the classification. To reduce the likelihood of reaching local minima with the algorithm, clusterization is repeated a hundred times, each with a new set of initial cluster centroid positions. The algorithm returns the solution with the lowest value for the objective function. In this application, the daily mean sea level pressures corresponding to the 55 years of data ($n = 20088$ days), represented by 11 principal components, are classified into $n_{wt} = 9$ groups.

Note that in this particular case we select 9 weather types for the sake of simplicity, to facilitate the implementation, fit and interpretation of the model results. However, the selection of the appropriate number of clusters is an open issue not solved yet. There are authors, such as [19,6,1], that defend the use of a maximum of 10 weather types, others ([23,5,18]) claim that a higher number of weather types is required to represent the intrannual/interannual variations and seasonality appropriately. Being more specific, [6] uses only 4 weather types to represent daily precipitation scenarios, [9] classifies into 20 weather types the daily atmospheric circulation patterns, or for example, [18] uses 64 weather types to study the extreme wave height variability. This paper does not solve the problem of establishing the appropriate number of weather types, which must be decided by the user according to his/her experience. But due to the facility to implement, fit and interpret model results might help establishing a rationale criteria for solving this problem.

Figure 1 shows the 9 representative weather types obtained from the clusterization. For instance, the upper left subplot represents a synoptical circulation pattern with a low pressure center above the Britannic Islands while the Azores High remains southwestern the Iberian Peninsula, whereas the upper central subplot shows the Azores High with its center southwest of the United Kingdom.

Assigning arbitrarily an integer value between 1 and $n_{wt} = 9$, for each weather type in Figure 1, we get the time series of weather types Y, which is the input for the model.

To fit the data and according to the parameterizations given in (3)-(7), long-term trend, seasonality, covariates and a first order autoregressive Markov chain are included. Each study and location may require a pre-process to select the parameters to be included according to their influence. Related to covariates, it is worth to mention that Monthly Sea Level Pressure Anomalies fluctuations (MSLPA) have been considered. These anomalies correspond to monthly deviations from the 55-year monthly averages, which allows obtaining interannual variations. This interannual modulation can be related to well known synoptic patterns, such as EA, NAO, SOI, etc. [14], but we preferred to use the principal components of the anomalies to avoid discrepancies about what predictors should be used instead. Nevertheless, we could have used those indices within the analysis. In this case, the first 9 principal components of the monthly sea level pressure anomalies (MSLPA) that explain more than 96% of the variability are included as covariates. Figure 2 shows the spatial modes related to those principal components. Note, for instance, that the correlation between the first mode and NAO index is $r = -0.618$ and the correlation between the second mode and EA synoptic pattern is $r = 0.482$.

3.1 Model Fitting

Results obtained from the application of the proposed model to the Northeastern Atlantic are described in detail. The output given by function `mnrfit` is a matrix $\hat{p}$ of dimensions $(n_p \times (n_{wt} - 1))$ including parameter estimates by the maximum likelihood method, where n_p is the number of parameters in the model and n_{wt} is the number of weather types considered. Note that each weather type has an associated parameter except for the reference weather type, whose parameters are set to zero.

The criteria to choose the final model, i.e. the order d of the auto-regressive component, seasonality, covariates, etc. is based on statistical significance, in particular, using the likelihood ratio (LR) statistic. This statistical method is appropriate to compare nested models by comparing the deviance ratio ΔDev., which measures the change of fitting quality for two different parameterizations, and the chi-square distribution with $\Delta df = \Delta n_p \times (n_{wt} - 1)$ degrees of freedom. Note that Δn_p is the difference in terms of number of parameters for both parameterizations. Basically, it tries to check if the increment of fitting quality induced by increasing the number of parameters is justified, i.e. does the increment on fitted parameters conduct to a better model? For instance, assuming a confidence level $\alpha = 95\%$, if $\Delta\text{Dev.} > \chi^2_{0.95,\Delta df}$, the improvement achieved by adding n_p additional parameters is significant. This test allows to analyze which parameters or covariates are relevant to represent climate dynamics in a particular location.

In order to evaluate the goodness-of-fit related to the predictors, several different fits are considered. In Table 3, up to 7 nested models are compared depending on the predictors involved. In this table, the number of parameters (n_p), the deviance of the fitting (Dev.), the degrees of freedom (df) and the rate of change on deviance (ΔDev.) are provided. Model 0 is the so-called *Null* model that only takes into account an independent term (β_0). Model *I* adds the possible influence of seasonality(π^S), which according to the increment on deviance with respect to the null model $\Delta\text{Dev.} = 7417 > {\chi^2}_{95\%,16}$ is significant, confirming the hyphothesis that there is a seasonality pattern in the occurrence of the different weather types. Model *II* includes seasonality and MSLPA covariates ($\pi^S + \pi^C$), which also provide significant information. Model *III* is fitted accounting for seasonality, MSLPA covariates and long-term trend ($\pi^S + \pi^C + \pi^{LT}$). In this particular case, the increment on quality fit induced by the inclusion of an additional parameter, related to long-term trend, is not significant, i.e. $\Delta\text{Dev.} = 9 < {\chi^2}_{95\%,8}$. Models *IV* and *V* include the influence of autoregressive terms (Markov Chains, MC) with orders $d = 1$ and $d = 2$, respectively ($\pi^S + \pi^C + \pi^{LT} + \pi^{AR_d}$). Note that both autoregressive components are significant. Additionally, due to the importance of long-term

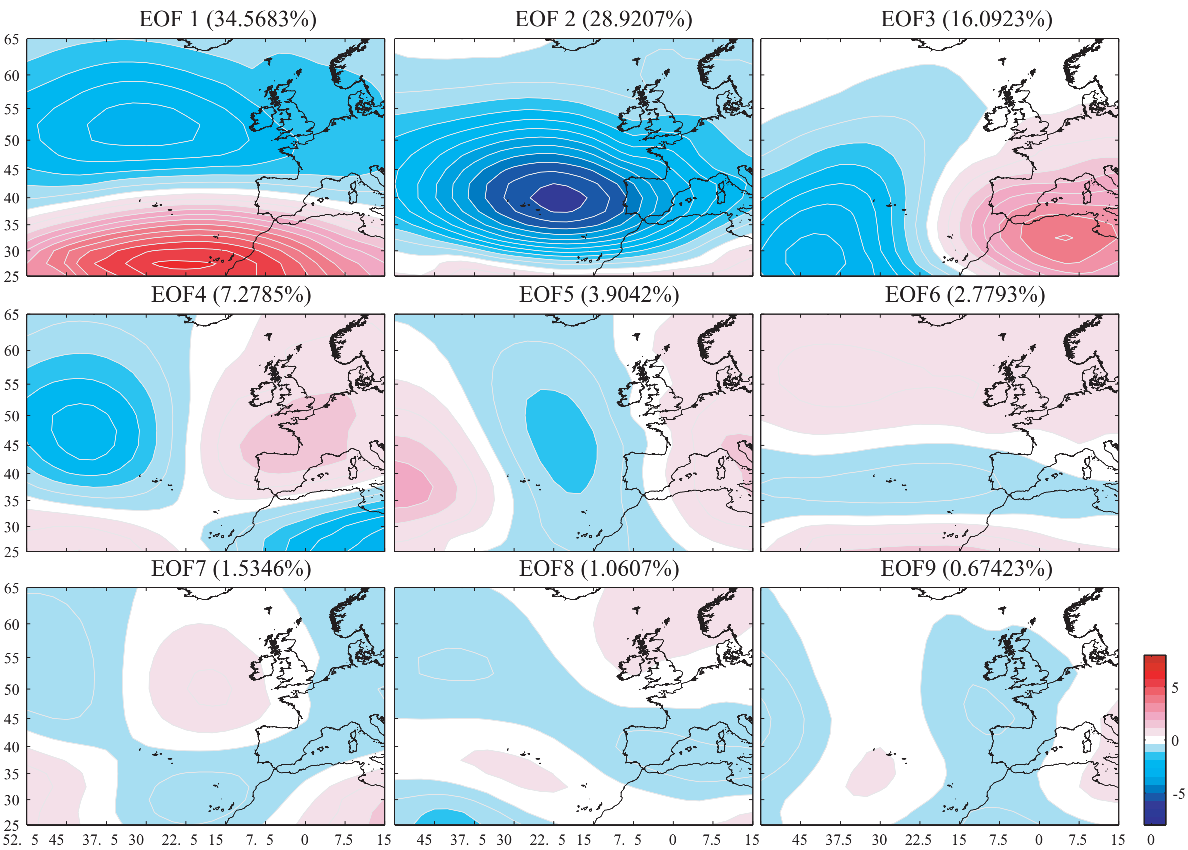

Fig. 2 MSLPA spatial modes related to the Principal Components included as covariates in the model.

changes in the probabilities of occurrence of the different weather types, a model that only takes the long term trend into account has also been fitted, model VI (π^{LT}). This additional factor is statistically significant ΔDev. $= 69 > \chi^2{}_{95\%,8}$, which means that there is a long-term evolution on the probability of occurrence related to each weather type. However, there is an inconsistency with respect to model III, where this factor is not statistically significant. The reason for this behaviour is simple, when using covariates, the long-term effects are implicitly included in the covariates and there is no reason to include additional effects not explained by those covariates.

It is important to point out that deciding which model is more appropriate for each case depends on weather dynamics knowledge of the user, and its ability to confront or contrast its feeling about which physical phenomena is more relevant, with respect to the statistical significance of the corresponding fitted model. The main advantage of the proposed method is that it provides an statistical and objective tool for deciding what information is more relevant to explain climate variability.

Note that as said in Section 2.1, the seasonality constant term β_0, which corresponds to a column vector $(1,1,\ldots,1)^T$ is automatically included in the model depending on the standard logistic model used. Using the function `mnrfit` this constant is automatically added, thus the null model (a) has $n_p = 1$ and the model fitted only with the trend (g) has $n_p = 2$.

Model	Predictors	n_p	df	Dev.	Δ Dev.	$\chi^2{}_{95\%,\Delta df}$
0	β_0	1	160696	85736		
I	π^S	3	160680	78319	7417	26.9
II	$\pi^S+\pi^C$	12	160608	68105	10214	92.8
III	$\pi^S+\pi^C+\pi^{LT}$	13	160600	68096	9	15.5
IV	$\pi^S+\pi^C+\pi^{LT}+\pi^{AR_1}$	21	160536	45937	22159	83.7
V	$\pi^S+\pi^C+\pi^{LT}+\pi^{AR_2}$	29	160472	45610	327	83.7
0	β_0	1	160696	85736		
VI	π^{LT}	2	160688	85667	69	15.5

Table 3 Fitting diagnostics for different model parameterizations, including number of parameters (n_p), the deviance of the fitting (Dev.), the degrees of freedom (df) and the rate of change on deviance (ΔDev.)

If we consider model IV, which accounts for seasonality, MSLPA covariates, long-term trend and a first order autoregressive component as predictors ($\pi^S+\pi^C+\pi^{LT}+\pi^{AR_1}$), the model has 21 parameters, $n_p = 21 = 3+n_c+1+d\times(n_{wt}-1) = 3+9+1+1\times 8$: i) three for seasonality π^S,

ii) nine for the MSLPA principal components π^C, iii) one for the long-term trend π^{LT}, and eight for the dummy variables of the first autoregressive component π^{AR_1}.

Once the parameter estimates for the models $\hat{\Theta}$ are known, the predicted probabilities $\hat{p}$ for the multinomial logistic regression model associated with given predictors $\tilde{x}$ can be easily calculated. This task can be performed using the MATLAB function `mnrval`, which receives as arguments the estimated parameters $\hat{\Theta}$ and the covariate values $\tilde{x}$. In addition, confidence bounds for the predicted probabilities related to a given confidence level ($\alpha = 0.99, 0.95, 0.90$) can be computed under the assumption of normally distributed uncertainty. Note that these probabilities $\hat{p}$ correspond to the probability of occurrence for each weather type according to the predictor values $\tilde{x}$.

These probabilities allow direct comparison with the empirical probabilities from the data, and the possibility to simulate random sequences of weather types. The graphical comparison between fitted model and observed data can be done in different time scales, aggregating the probabilities of occurrence within a year, year-to-year or for different values of the covariates (MSLPA).

– **Seasonality** To analyze the importance of seasonality, Figure 3 shows the comparison of the probabilities of occurrence for each weather type within a year. Color bars represent cumulative empirical probabilities, and black lines represent the same values but given by the fitted model I, which only accounts for seasonality using harmonics (panel above in Figure 3), and also using an autoregressive term at lag 365 (panel below in Figure 3). For each day within a year the bars represent cumulative probabilities of occurrence of all the 9 weather types, which are calculated for each day using the 55 data associated with each year. Note that there is a clear seasonal pattern which is captured by the model using harmonics, being circulation patterns 4, 7 and 8 the most likely weather types during the summer, while groups 1, 6 and 9 are more relevant during the winter. Comparing both ways of accounting for seasonality, the harmonic (panel above of Figure 3) is capable of reproducing the seasonal behavior better than the autocorrelation term at lag 365 (panel below of Figure 3).
This seasonal variation through the years is also shown in Figure 4. In this particular case color bars represent cumulative monthly probabilities. Note that the model (black line) repeats the same pattern all over the years since we are using fitting results associated with model I. Analogously to the previous Figure 3, it is observed a clear seasonal pattern. For example, in the lower part of the graph it is observed how weather types 1 and 2, mostly related to winter and summer, respectively, change the occurrence probability depending on the season within the year. The same behavior is observed in the upper part of the graph related to weather types 3 and 9.

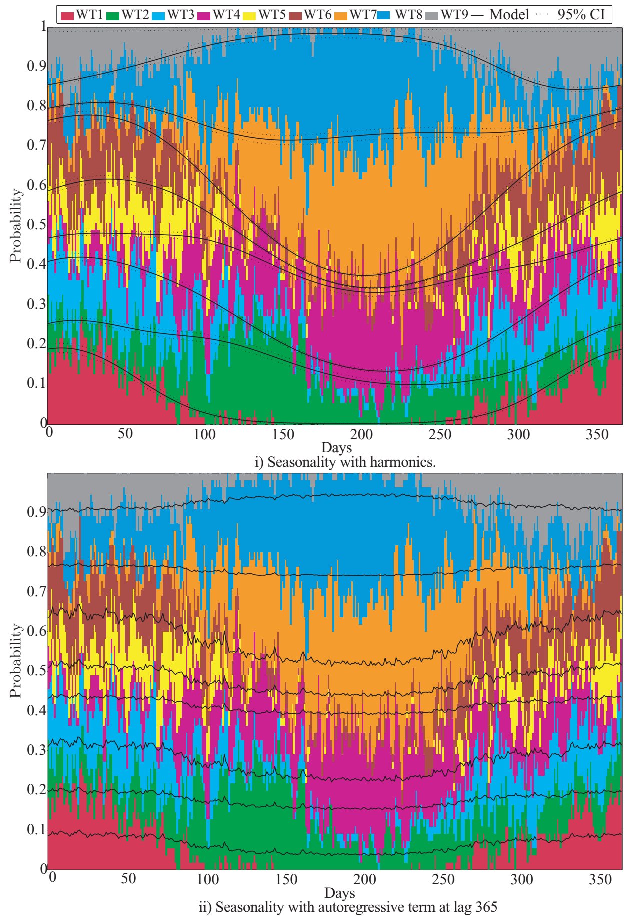

Fig. 3 Model fitting diagnostic plot considering seasonality: i) using harmonics (Model I), and ii) using an autoregressive term at lag 365.

– **Mean Sea Level Pressure Anomalies (MSLPA)** Although model I reproduces and explains the important seasonality effect, it can be observed in Figures 3 and 4 that there are important fluctuations and discrepancies between the empirical data and the model on a daily and monthly basis, respectively. If model IV including seasonality, MSLPA covariates, an autoregressive component of order $d = 1$ and long-term trend ($\pi^S + \pi^C + \pi^{AR_1} + \pi^{LT}$) is considered, results are shown in Figures 5 and 6 . The fitted model now explains all fluctuations both on the daily and monthly scale.
Note that once the noise on daily and monthly probabilities is explained by those additional factors, the consideration of seasonality through the 365-lag autoregressive model also provides similar diagnostic fitting plots, i.e. model $IV : \pi^S + \pi^C + \pi^{AR_1} + \pi^{LR} \equiv \pi^{AR_{365}} + \pi^C + \pi^{AR_1} + \pi^{LR}$.

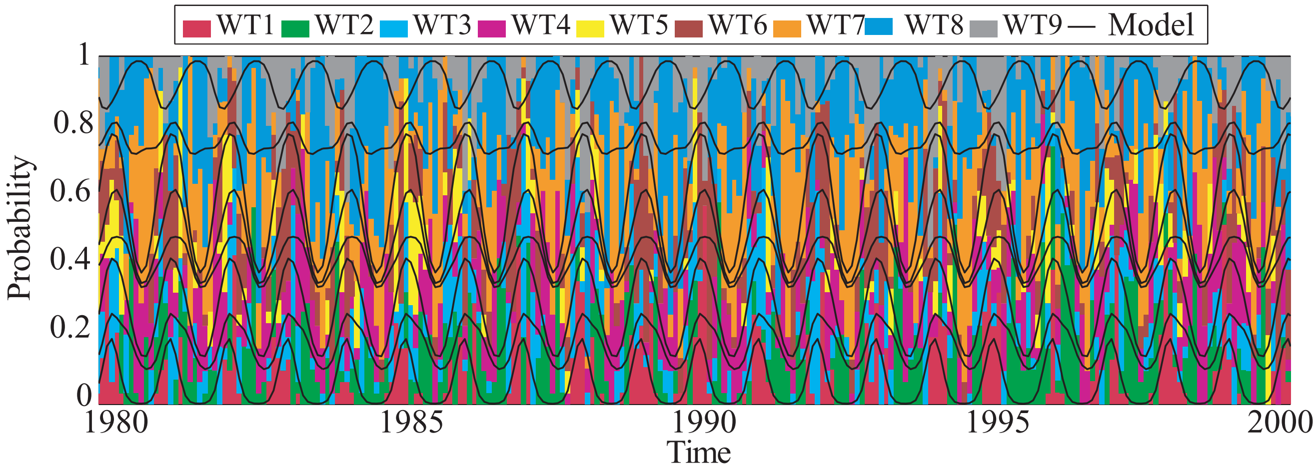

Fig. 4 Evolution of the monthly probabilities of occurrence during 20 years and comparison with the seasonal fitted model *I* (black line).

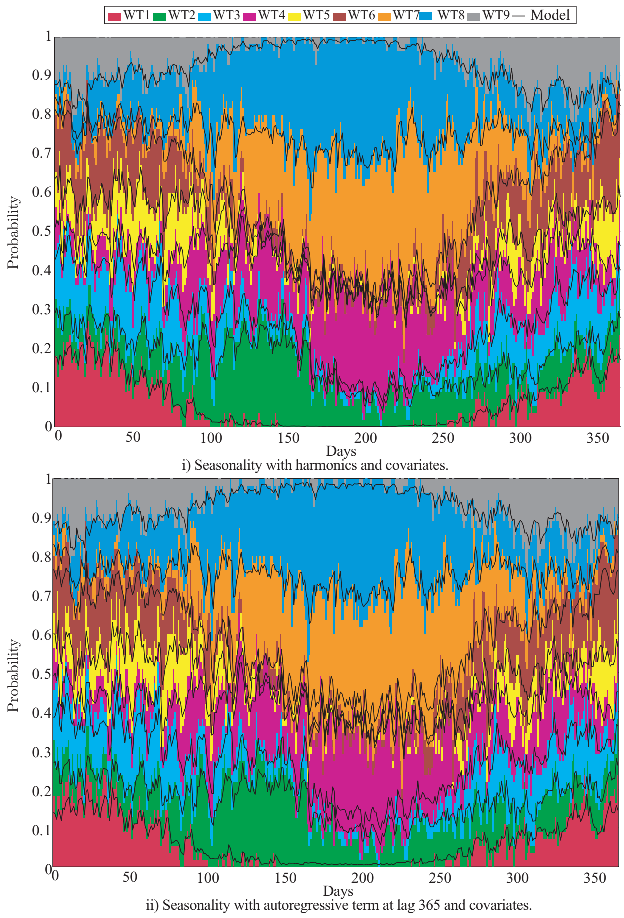

Fig. 5 Model fitting diagnostic plot considering model *IV*: i) using harmonics (Model *I*), and ii) using an autoregressive term at lag 365.

It is relevant to point out how the inclusion of MSLPA allows explaining the monthly fluctuations on the probabilities of occurrence associated with the different weather types (see Figure 6). These results confirm that model *IV* is capable of reproducing and explaining the weather dynamics accurately, both on a daily and monthly basis. Using this model we manage to model atmospheric processes on both the short and the long term, using a combination of short-term sequencing through autocorrelation terms and long-term correlations included implicitly through seasonality, covariates and long-term variations.

To further explore the influence of the MSLPA on the occurrence probability for each weather type, Figure 7 shows the probability of occurrence of each weather type conditioned to the value of the MSLPA principal components (PC_i; $i = 1,\ldots,9$) included as covariates. Color bars represent the cumulative empirical probabilities from data, and the black lines are fitted model probabilities.

According to results shown in Figure 7, the presence or absence of a weather type may be related with the value of the PC anomaly. For instance, in the subplot associated with the first principal component (upper left subplot), negative values of the principal component imply an increment on the occurrence of weather types 1 (red), 6 (maroon) and 9 (grey); while for positive values the most likely weather types are 2 (green), 3 (light blue) and 5 (yellow). On the other hand, for negative values of the second principal component, the dominant weather type is the blue one (8), prevailing weather types 1 and 5 for positive values of the PC. Finally, for the third principal component, the behavior is different; the lowest values of this principal component indicate a higher likelihood of weather types 4 and 9, while higher values increase the probability of occurrence of weather type 3.

Note that according to the low variance explained by principal component from 4 to 9, we could be tempted to omit them from the analysis. To check whether these

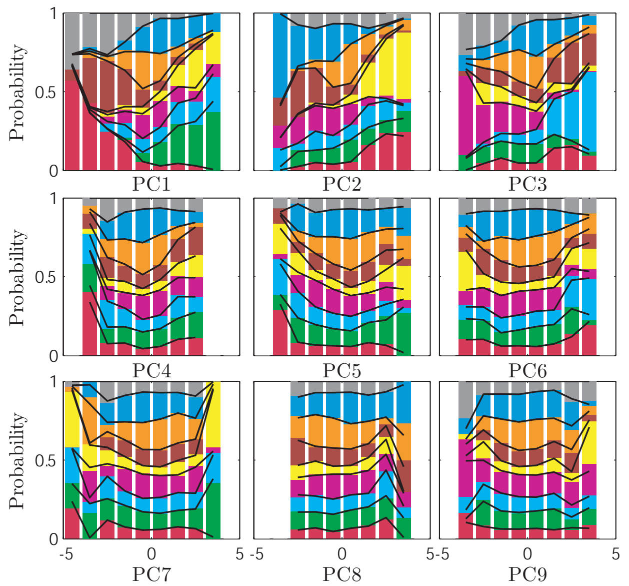

Fig. 7 Evolution of the probabilities of occurrence of each weather type conditioned to the principal component value associated with fitted model *IV* (black line).

covariates improve significantly the quality of the fit, we have included the principal components one at a time, and check the likelihood ratio (LR) statistic. Table 4 provides the results from the analysis. Note that although it is clear that the most relevant information is given by the first three principal components, which represent important increments on deviance, the remainder covariates also improve the quality of the model from an statistical viewpoint. For this particular case, all principal components are statistically significant on a 95% confidence level.

Model	df	Dev.	Δ Dev.	$\chi^2_{95\%,8}$
0	160696	85736		
PC_1	160688	81308	4428	15.5
PC_2	160680	77429	3879	15.5
PC_3	160672	75137	2292	15.5
PC_4	160664	74912	225	15.5
PC_5	160656	74790	122	15.5
PC_6	160648	74729	61	15.5
PC_7	160640	74650	79	15.5
PC_8	160632	74551	99	15.5
PC_9	160624	74533	18	15.5

Table 4 Fitting diagnostics related to the principal components associated with MSLPA, including the deviance of the fitting (Dev.), the degrees of freedom (df) and the rate of change on deviance (ΔDev.)

– **Trends** Finally, in order to show the possible influence of a long-term trends, results associated with model VI, which only accounts for long term trends, are shown in Figure 8. Color bars represent the annual probability

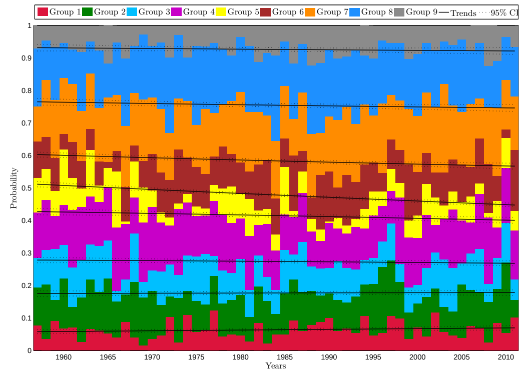

Fig. 8 Annual probabilities of occurrence for each weather type and comparison with model VI fitting results (black line) in the period 1957 – 2011.

of occurrence for each year (55 data record) associated with the 9 established weather types. The black line represents the model fitting (model VI in Table 3). Note that we do not present results associated with model *IV* because the long term trend is not statisticcally significant in that model, because long-term effects are implicitly accounted for through the covariates.

	WT_1	WT_2	WT_3	WT_4	WT_5	WT_6	WT_7	WT_8
Trend($\times 10^{-2}$)	0.09	*-0.4*	**-0.51**	**-0.48**	**-1.33**	0.25	-0.07	-0.16
$\sigma_{Trend}(\times 10^{-2})$	0.24	0.21	0.22	0.20	0.24	0.21	0.19	0.19

Table 5 Fitting parameters associated with model VI including long-term trends, and their corresponding standard error. Values in bold are statistically significant at 95% confidence level and values in cursive are significant at 90% confidence level.

The parameters for the trends and their corresponding standard errors are provided in Table 5. Note that statistically significant trends at 95% confidence levels are boldfaced, while trends which are statistically significant at 90% confidence level are in italics. According to results given in this table the following observations are pertinent:

– The reference weather type is weather type number 9. That is the reason why there is no parameter related to this case. Note that it is a typical winter weather type.
– The coefficients may be interpreted as the relative change in the odds due to a small change in time δt, i.e. the percentage of change in odds between weather types 5 and 9 during one year is approximately equal to -1.33%.
– Weather types 4, 7 and 8, which represent summer weather types, decrease with respect to type 9. This means that weather types related to winter are increasing its occurrence probability. This result is

consistent with recent studies about the increment of wave climate severity, which is linked to weather types during the winter season.

- Note that weather type 1, also typical during winter, slightly increases the odds with respect to type 9. Confirming the increment of occurrence related to winter weather types.

3.2 Monte Carlo Simulations

Once the model has been fitted and the $\hat{p}$ matrix is obtained, synthetic sequences of weather types can be generated through Monte Carlo method. In this particular case, since we require the knowledge of the covariate values during the simulation period, 55 years of daily data series ($n = 20088$) are sampled using the original covariates. In order to obtain statistically sound conclusions according to the stochastic nature of the process, the simulation is repeated 100 times. The results obtained are validated with a threefold comparison against the original sequence of weather types: i) occurrence probabilities of WT, ii) transition probability matrix between WT and iii) persistence analysis of WT.

- **Occurrence Probabilities**
  The probabilities of occurrence of the 9 groups for the 100 simulations, against the empirical probability of occurrence from the 55-year sample data, are shown in Figure 9. Note that results are close to the diagonal, which demonstrates that the model simulations are capable of reproducing the probability of occurrence associated with weather types appropriately.
- **Transition Probabilities Matrix**
  The transition probabilities express the probability of changing from group i to group j between consecutive days. Thus, in the case of having 9 weather types, the transition matrix (T) has dimensions 9×9, and each cell $T_{i,j}$ is the probability of changing from weather type i to weather type j ([31]). The diagonal of the transition matrix T corresponds to the probability of staying in the same group. The transition matrix is calculated for each of the 100 simulated samples. Figure 10 shows the scatter plot related to the $9 \times 9 = 81$ elements of transition matrix, including its uncertainty due to the simulation procedure, against the empirical transition probabilities obtained from the initial data set. The model is able to reproduce correctly the transitions between circulation patterns within the sequence. In this particular case, the points with probabilities in the range $0.6 - 0.8$ are those representing the probability of staying in the same group (diagonal of the transition matrix).
- **Persistence Analysis**

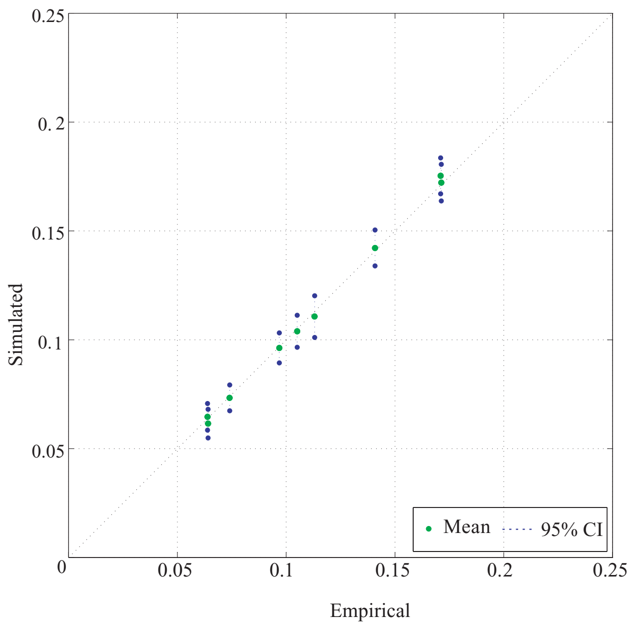

Fig. 9 Scatter plot of the empirical occurrence probabilities associated with the weather types versus Monte Carlo simulation results.

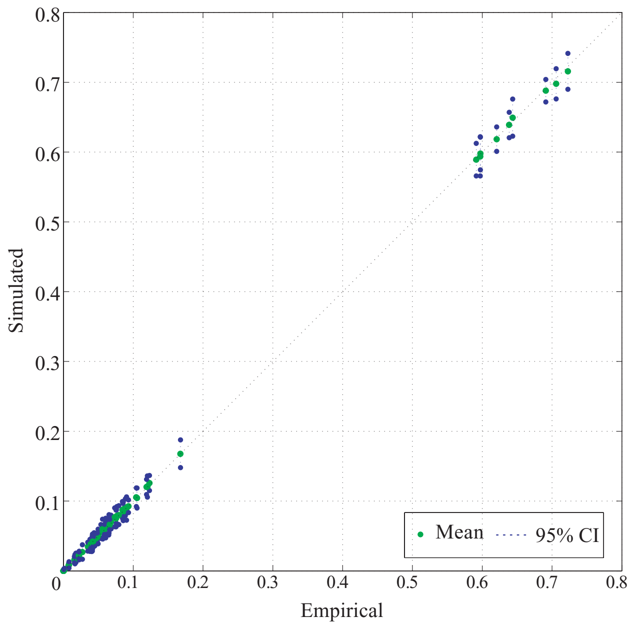

Fig. 10 Scatter plot of the empirical transition probabilities between weather types versus Monte Carlo simulation results.

Finally, a persistence analysis is performed over the simulated samples in order to check the ability of the model to reproduce weather dynamics. The correct reproduction of the weather types persistence is very important for many climate related studies, because it may be related to length of droughts, heat waves, etc. Figure 11 shows the empirical cumulative distributions of the persistence associated with each weather type. Note that the average empirical distribution (green line) is

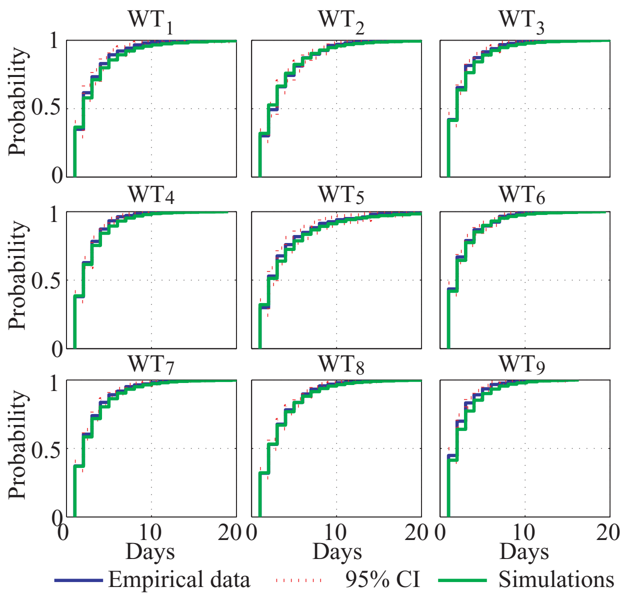

Fig. 11 Empirical cumulative distribution of the persistence for the 9 groups related to: i) historical data and ii) sampled data using Monte Carlo method.

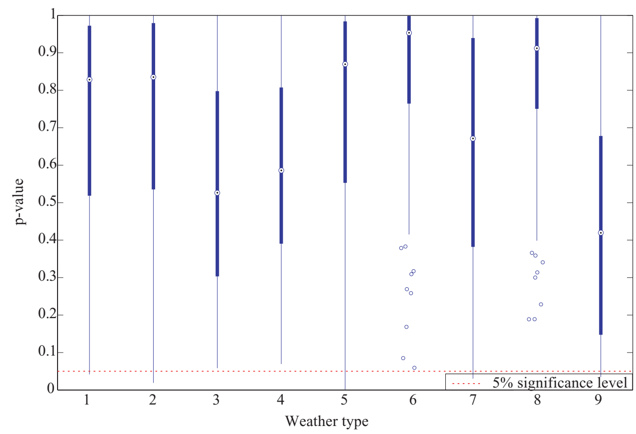

Fig. 12 Box plot associated with the *p*-values from the 100 tests for each weather type.

very close to the one related to the historical sample data (blue line) for all cases. This blue line stays between the 95% confidence intervals (red dotted line) related to the 100 simulations. To further analyze the performance on persistence from an statistical viewpoint, we perform a two-sample Kolmogorov-Smirnov ([24]) goodness-of-fit hypothesis test between the original data and each sampled data. This test allows determining if two different samples come from the same distribution without specifying what that common distribution is. In Figure 12 the box plots associated with the *p*-values from the 100 tests for each weather type are shown. Note that if the *p*-value is higher than the significance level (5%) the null hypothesis that both samples come from the same distribution is accepted. Results shown in Figure 12 prove that for most of the cases the persistence distributions from the Monte Carlo simulation procedure come from the same distribution as the persistence distribution from the historical data. For all the weather types the interquartile range (blue box) is above the 5% significance level (red dotted line). These results confirm the capability of the model to reproduce synthetic sequences of weather types coherent in term of persistence.

4 Conclusions

This work presents an autoregressive logistic model which is able to reproduce weather dynamics in terms of weather types. The method provides new insights on the relation between the classification of circulation patterns and the predictors implied. The advances with respect to the state-of-the-art can be summarized as follows:

– The availability of the model to include autoregressive components allows the consideration of previous time steps and its influence in the present.
– The models allows including long-term trends which are mathematically consistent, so that the probabilities associated with each weather type always range between 0 and 1.
– The proposed model allows to take into account simultaneously covariates of different nature, such as MSLPA or autoregressive influence, where the time scales are completely different.
– The capability of the model to deal with nominal classifications enhances the physical point of view of the problem.
– The flexibility of the proposed model allows the study of the influence of any change in the covariates due to long-term climate variability.

On the other hand, the proposed methodology presents a weakness in relation with the data required for fitting purposes, because a long-term data base is needed to correctly study the dynamics of the weather types.

Although further research must be done on the application of the proposed model to study processes that are directly related with weather types, such as marine dynamics (wave height, storm surge, etc.) or rainfall, this method provides the appropriate framework to analyze the variability of circulation patterns for different climate change scenarios ([28]).

Acknowledgements This work was partially funded by projects "AMVAR" (CTM2010-15009), "GRACCIE" (CSD2007-00067, CONSOLIDER-INGENIO 2010), "IMAR21" (BIA2011-2890) and "PLVMA" (TRA2011-28900) from the Spanish Ministry MICINN, "MARUCA" (E17/08) from the Spanish Ministry MF and "C3E" (200800050084091) from the Spanish Ministry MAMRM. The support of the EU FP7 Theseus

"Innovative technologies for safer European coasts in a changing climate", contract ENV.2009-1, n. 244104, is also gratefully acknowledged. Y. Guanche is indebted to the Spanish Ministry of Science and Innovation for the funding provided in the FPI Program (BES-2009-027228). R. Mínguez is also indebted to the Spanish Ministry MICINN for the funding provided within the "Ramon y Cajal" program.

References

1. van den Besselaar, E., Klein Tank, A., van der Schrier, G.: Influence of circulation types on temperature extremes in Europe. Theoretical and Applied Climatology **99**, 431–439 (2009)
2. Bizzotto, R., Zamuner, S., De Nicolao, G., Karlsson, M.O., Gomeni, R.: Multinomial logistic estimation of Markov-chain models for modelling sleep architecture in primary insomnia patients. Pharmacokinet Pharmacodyn **37**, 137–155 (2010)
3. Bonney, G.E.: Logistic regression for dependent binary data. Biometrics **43**(4), 951–973 (1987)
4. Box, G.E.P., Jenkins, G.M., Reinsel, G.C.: Time Series Analysis: Forecasting and Control. Prentice-Hall International, New Jersey, NJ (1994)
5. Cahynova, M., Huth, R.: Enhanced lifetime of atmospheric circulation types over europe: fact or fiction? Tellus **61A**, 407–416 (2009)
6. Corte-Real, J., Xu, H., Qian, B.: A weather generator for obtaining daily precipitation scenarios based on circulation patterns. Climate Research **13**, 61–75 (1999)
7. Cox, D.R.: The analysis of multivariate binary data. Journal of the Royal Statistical Society. Series C (Applied Statistics) **21**(2), 113–120 (1972)
8. Dobson, A.J.: An Introduction to Generalized Linear Models, second edn. Chapman & Hall/CRC, Florida (2002)
9. Esteban, P., Martín-Vide, J., Mases, M.: Daily atmospheric circulation catalogue for Western Europe using multivariate techniques. International Journal of Climatology **26**, 1501–1515 (2006)
10. Gerstengarbe, F.W., Werner, P.: Katalog der Grosswetterlagen Europas(1881-1998) nach Paul Hess und Helmuth Brezowsky. Potsdam Institut für Klimafolgenforschung, Postdam, Germany (1999)
11. Gerstengarbe, F.W., Werner, P.: Katalog der Grosswetterlagen Europas(1881-2004) nach Paul Hess und Helmuth Brezowsky. Potsdam Institut für Klimafolgenforschung, Postdam, Germany (2005)
12. Goodess, C.M., Jones, P.D.: Links between circulation and changes in the characteristics of Iberian rainfall. International Journal of Climatology **22**, 1593–1615 (2002)
13. Hess, P., Brezowsky, H.: Katalog der Grosswetterlagen Europas(Catalog of the European Large Scale Weather Types). Ber. Dt. Wetterd, in der US-Zone 33, Bad Kissingen, Germany (1952)
14. Hurrell, J., Kushnir, Y., Ottersen, G., Visbeck, M.: The North Atlantic Oscillation: Climate Significance and Environmental Impact, geophysical monograph series 134. American Geophysical Union, Washington, DC (2003)
15. Huth, R.: A circulation classification scheme applicable in GCM studies. Theoretical and Applied Climatology **67**, 1–18 (2000)
16. Huth, R.: Disaggregating climatic trends by classification of circulation patterns. International Journal of Climatology **21**, 135–153 (2001)
17. Huth, R., Beck, C., Philipp, A., Demuzere, M., Ustrnul, Z., Cahynová, M., Kyselý, K., Tveito, O.E.: Classification of atmospheric circulation patterns. Trends and Directions in Climate Research **1146**, 105–152 (2008)
18. Izaguirre, C., Menéndez, M., Camus, P., Méndez, F.J., Mínguez, R., Losada, I.J.: Exploring the interannual variability of extreme wave climate in the northeast atlantic ocean. Ocean Modelling **59–60**, 31–40 (2012). DOI http://dx.doi.org/10.1016/j.ocemod.2012.09.007
19. Jordan, P., Talkner, P.: A seasonal Markov chain model for the weather in the central Alps. Tellus **52A**, 455–469 (2000)
20. Kalnay, E.M., Kanamitsu, R., Kistler, W., Collins, D., Deaven, L., Gandin, M., Iredell, S., Saha, G., White, J., Woollen, Y., Zhu, M., Chelliah, W., Ebisuzaki, W., Higgins, J., Janowiak, K.C., Mo, C., Ropelewski, J., Wang, A., Leetmaa, R., Reynolds, R., Jenne, R., Joseph, D.: The NCEP/NCAR 40-year reanalysis project. Bulletin of the American Meteorological Society **77**, 437–470 (1996)
21. Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., Fiorino, M.: The NCEP-NCAR 50-year reanalysis: Monthly means CD-Rom and documentation. Bulletin of the American Meteorological Society **82**, 247–268 (2001)
22. Kyselý, K., Huth, R.: Changes in atmospheric circulation over Europe detected by objective and subjective methods. Theoretical and Applied Climatology **85**, 19–36 (2006)
23. Maheras, P., Tolika, K., Anagnostopoulou, C., Vafiadis, M., Patrikas, I., Flocas, H.: On the relationships between circulation types and changes in rainfall variability in Greece. International Journal of Climatology **24**, 1695–1712 (2004)
24. Massey, F.J.: The Kolmogorov-Smirnov test for goodness of fit. Journal of the American Statistical Association **46**, 68–78 (1951)
25. Muenz, L.R., Rubinstein, L.V.: Models for covariate dependence of binary sequences. Biometrics **41**(1), 91–101 (1985)
26. Nicolis, C., Ebeling, W., Baraldi, C.: Markov processes, dynamic entropies and the statistical prediction of mesoscale weather regimes. Tellus A **49**, 108–118 (1997)
27. Pasmanter, R.A., Timmermann, A.: Cyclic Markov chains with an application to an intermediate ENSO model. Nonlinear Processes in Geophysics **10**, 197–210 (2003)
28. Pastor, M.A., Casado, M.J.: Use of circulation types classifications to evaluate AR4 climate models over the euro-atlantic region. Climate Dynamics (2012)
29. Philipp, A., Bartholy, J., Beck, C., Erpicum, M., Esteban, P., Fettweis, X., Huth, R., James, P., Joudain, S., Kreienkamp, F., Krennert, T., Lykoudis, S., Michalides, S.C., Pianko-Kluczynska, K., Post, P., Rasilla, D., Schiemann, R., Spekat, A., Tymvios, F.S.: Cost733cat- A database of weather and circulation type classification. Physics and Chemistry of the Earth **35**, 360–373 (2010)
30. Plan, E., Elshoff, J.P., Stockis, A., Sargentini-Maier, M.L., Karlsson, M.O.: Likert pain score modelling: A Markov model and an autoregressive continuous model. Clinical Pharmacology and Therapeutics **91**, 820–828 (2012)
31. Polo, I., Ullmann, A., Roucou, P., Fontaine, B.: Weather regimes in the euro-atlantic and mediterranean sector, and relationship with west african rainfall over the 1989-2008 period from a self-organizing maps approach. Journal of Climate **24**, 3423–3432 (2011)
32. Rodríguez, G.: Lecture Notes on Generalized Linear Models, first edn. Princeton University, New Jersey (2007)
33. Stefanicki, G., Talkner, P., Weber, R.: Frequency changes of weather types in the alpine region since 1945. Theoretical and Applied Climatology **60**, 47–61 (1998)
34. Vidakovik, B.: Statistics for Bioengineering Sciences, first edn. Springer, New York (2011)
35. de Vries, S.O., Fiedler, V., Kiupers, W.D., Hunink, M.G.M.: Fitting multistate transition models with autoregressive logistic regression: Supervised exercise in intermittend claudication. Medical Decision Making **18**, 52–60 (1998)

36. Werner, P., Gerstengarbe, F.W.: Katalog der Grosswetterlagen Europas(1881-2009) nach Paul Hess und Helmuth Brezowsky. Potsdam Institut für Klimafolgenforschung, Postdam, Germany (2010)

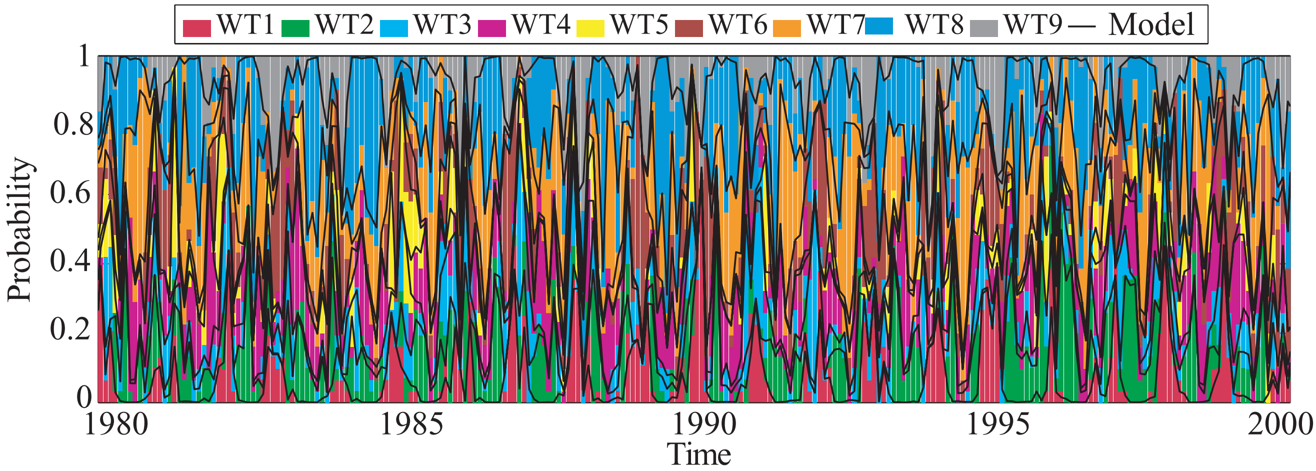

Fig. 6 Evolution of the monthly probabilities of occurrence during 20 years and comparison with the seasonal fitted model *IV* (black line).