Bayesian Networks for Probabilistic Weather Prediction

Antonio S. Co£ño [1] and **Rafael Cano** [2] and **Carmen Sordo** [3] and **José M. Gutiérrez** [4]

Abstract. Several standard approaches have been introduced for meteorological time series prediction (analog techniques, neural networks, etc.). However, when dealing with multivariate spatially distributed time series (e.g., a network of meteorological stations over the Iberian peninsula) the above methods do not consider all the available information (they consider special independency assumptions to simplify the model).

In this work, we introduce Bayesian Networks (BNs) in this framework to model the spatial and temporal dependencies among the different stations using a directed acyclic graph. This graph is learnt from the available databases and allows deriving a probabilistic model consistent with all the available information. Afterwards, the resulting model is combined with numerical atmospheric predictions which are given as evidence for the model. Ef£cient inference mechanisms provide the conditional distributions of the desired variables at a desired future time. We illustrate the ef£ciency of the proposed methodology by obtaining precipitation forecasts for 100 stations in the North basin of the Iberian peninsula during Winter 1999. We show how standard analog techniques are a special case of the proposed methodology when no spatial dependencies are considered in the model.

1 INTRODUCTION

Nowadays, the problem of weather forecast is solved with the help of numerical Atmospheric Circulation Models (ACMs), which are daily integrated by different weather services on coarse-grained resolution grids covering wide geographical areas. These models provide a description of several meteorological variables (temperature, humidity, geopotential, wind components, etc.) which de£ne the predicted atmospheric pattern for a given forecast period. The spatial resolution of these models is currently constrained by both computational and physical considerations to scales of approximately 50 to 100 Km. However, meteorological phenomena such as rainfall, vary on much more local scales and therefore, ACMs do not provide a regional detailed description of such relevant phenomena. Due to this limitation, a number of different statistical and machine learning techniques have emerged in the last decade. These techniques mine the information contained in meteorological databases of historical observations to train speci£c forecast models (regression [1], hidden Markov models [2], neural networks [3, 4], etc.). The resulting models predict future outcomes of a given variable based on the past evidence collected in the database.

There have also been some attempts for combining both database information and ACMs. This is done by combining the model's predicted patterns with the information available in databases of observations (e.g., rainfall) and predictions (gridded atmospheric patterns). Therefore, sub-grid detail in the prediction is gained by post-processing the outputs of ACMs using knowledge extracted from the databases (*downscaling methods*). One of the most popular downscaling techniques is the method of analogs, which assumes that similar atmospheric patterns may lead to similar future outcomes [5]. Thus, predictions based on an atmospheric pattern can be derived from an "analog ensemble" extracted from the database. Different clustering techniques have been recently introduced to select this ensemble (see [6, 7] and references therein). However when dealing with multivariate time series, the analog technique assume different statistical independence relationships to simplify the model, neglecting important information among the variables in the database (most of them do not include spatial dependencies, and each station is predicted separately).

In this paper we illustrate how Bayesian Networks (BNs) offer a sound and practical methodology for this problem, allowing to automatically building tractable probabilistic models from data discovering the existing dependencies among the stations within the databases. These dependencies are represented by a graph, which also gives a simple factorized form of the probability distribution of the variables [8]. The resulting model provides probabilistic forecasts considering the relevant spatial-temporal information among the variables. To illustrate the performance of this technique, we shall consider a hundred meteorological stations over the Iberian peninsula. We want to remark here that Bayesian networks have been previously applied to model other meteorolgical problems; for instance in the project Hail£nder, where a network for modeling summer hail in northeastern Colorado was built [9].

In Section 2 we analyze the problems associated with statistical rainfall forecasting methods. In Sec. 3 we introduce BNs and analyze both the construction and usage steps. Section 4 describes how to incorporate information from atmospheric circulation models. Finally, some conclusions and further remarks are given in Sec. 5.

2 RAINFALL FORECAST

Suppose we are given a database of historical precipitation records, $(y_1^k, \ldots, y_t^k)$, at a local station of interest k. The geographical area of interest in this work is the Iberian Peninsula. We shall use rainfall values from 100 climatic stations provided by the Spanish National Weather Service −Instituto Nacional de Meteorología, INM− (see Figure 1(b)).

Time series forecasting methods work with patterns formed by

[1] Dept. of Applied Mathematics and Computer Science, University of Cantabria, Avda. de los Castros, 39005 Santander, Spain.
[2] Instituto Nacional de Meteorología, CMT/CAS, Spain.
[3] Dept. of Applied Mathematics and Computer Science, University of Cantabria / Instituto de Física de Cantabria CSIC, Santander, Spain.
[4] Dept. of Applied Mathematics and Computer Science, University of Cantabria, Santander, Spain. email: gutierjm@unican.es. URL: personales.unican.es/gutierjm

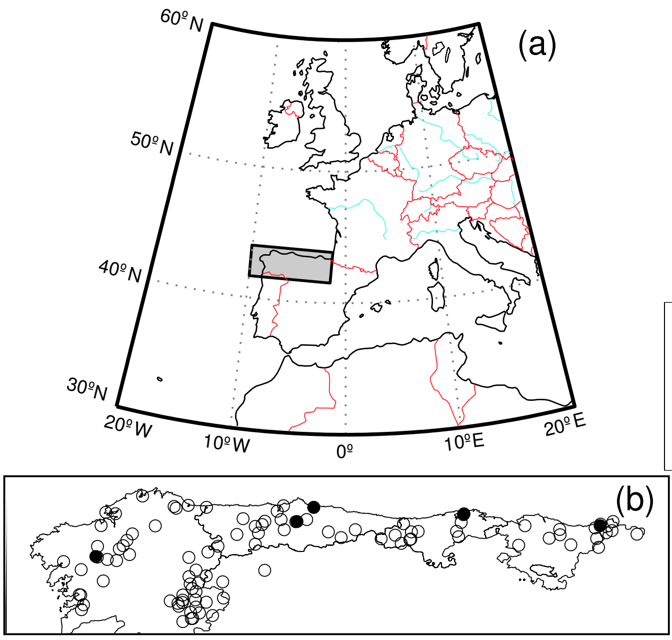

Figure 1. (a) Geographical area of study (shaded area); (b) blowup indicating the 100 stations in the North basin of the Iberian peninsula considered in this work (black circles indicate stations of the primary network).

a delayed vector $\mathbf{u_t^k} = (y^k_{t-1}, \ldots, y^k_{t-d})$ of the historical climate records, $y^k_1, y^k_2, \ldots$, of station k in the database. Given this information, standard parametric and non-parametric procedures (e.g., regession and neural networks) were applied for obtaining global or local (analog techniques) models $y^k_T = f(\mathbf{u^k_T}) + \epsilon$ for predicting station k. However, the main shortcoming of these methods when dealing with multivariate time series (such as the network of stations shown in Fig. 1) is that they assume spatial independence, neglecting important information. Several attempts to solve this problem, include multiple regression models $y^k_T = \mathbf{a}\mathbf{u_T} + b + \epsilon$ where $\mathbf{u_T} = (\mathbf{u_T}, u^1_T, \ldots, u^{100}_T)$ and feedforward networks with input $\mathbf{u_T}$. In both cases the models are useless due to large number of variables involved, requiring a unavailable huge database to avoid over£tting.

An alternative and ef£cient solution for this problem is determining the strongest dependencies among the variables $y^1, \ldots, y^{100}$, obtaining a simpler joint probability distribution with much less parameters characterizing the data. This task can be easily accomplished using Bayesian networks, which provide a simple and sound graphical framework for analyzing dependencies when dealing with uncertainty.

3 BAYESIAN NETWORKS

The basic idea of Bayesian networks (BNs) (BNs) is to reproduce the most important dependencies and independencies among a set of variables in a graphical form (a directed acyclic graph) which is easy to understand and interpret. Let us consider the subset of climatic stations shown in the graph in Fig. 2, where the variables (rainfall) are represented pictorially by a set of nodes; one node for each variable (for clarity of exposition, the set of nodes is denoted $\{y_1, \ldots, y_n\}$). These nodes are connected by arrows, which represent a cause and effect relationship. That is, if there is an arrow from node y_i to node y_j, we say that y_i is the cause of y_j, or equivalently, y_j is the effect of y_i. Another popular terminology of this is to say that y_i is a parent of y_j or y_j is a child of y_i. For example, in Figure 2, the nodes *Gijon* and *Amieva* and *Proaza* is a child of *Gijon* and *Rioseco* (the set of parents of a node y_i is denoted by π_i). Directed graphs provide a simple de£nition of independence (d-separation) based on the existence or not of certain paths between the variables (see [8] for a detailed introduction to probabilistic network models).

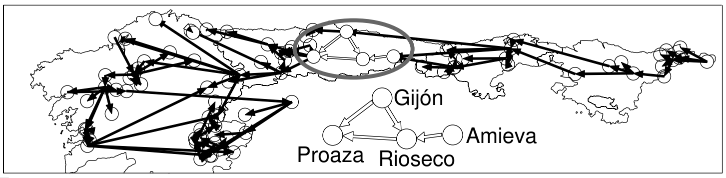

Figure 2. Directed graph associated with 100 stations in the North basin of the Iberian peninsula (the nodes are displayed maintaining their geographical disposition). For the sake of simplicity only four stations are labelled in the £gure: *Gijón*, *Proaza*, *Rioseco*, and *Amieva*.

The dependency/independency structure displayed by an acyclic directed graph can be also expressed in terms of a the Joint Probability Distribution (JPD) factorized as a product of several conditional distributions as follows:

$$Pr(y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n} P(y_i|\pi_i). \tag{1}$$

Therefore, the independencies from the graph are easily translated to the probabilistic model in a sound form. For instance, the JPD of a BN de£ned by the graph given in Fig. 2 requires the speci£cation of 100 conditional probability tables, one for each variable conditioned to its parents' set. Table 3 shows one of these probability tables. Hereafter we shall consider rainfall discretized into three different states (0="no rain", 1="weak rain", 2="heavy rain"), associated with the thresholds 0, 2, and 10 mm, respectively.

Table 1. Conditional Probability table of the node *Rioseco*, given the parent's set $\Pi = \{Gijon, Amieva\}$.

	Rioseco's state		
π	0	1	2
{0, 0}	0.82	0.10	0.08
{0, 1}	0.51	0.37	0.12
{0, 2}	0.44	0.31	0.25
{1, 0}	0.33	0.52	0.15
{1, 1}	0.15	0.63	0.22
{1, 2}	0.26	0.33	0.41
{2, 0}	0.38	0.44	0.18
{2, 1}	0.18	0.51	0.31
{2, 2}	0.17	0.30	0.53

3.1 Learning Bayesian Networks from Data

In addition to the graph structure, a BN requires that we specify the conditional probability of each node given its parents. However, in many practical problems, we do not know neither the complete topology of the graph, nor some of the required probabilities. For this reason, several methods haven been recently introduced for learning the graphical structure (structure learning) and estimating probabilities (parametric learning) from data (see [8, 10] for a review).

A learning algorithm consists of two parts:

1. A *quality measure*, which is used for computing the quality of the candidate BNs. This is a global measure, since it measures both the quality of the graphical structure and the quality of the estimated parameters.
2. A *search algorithm*, which is used to ef£ciently search the space of possible BNs to £nd the one with highest quality. Note that the number of all possible networks, even for a small number of variables and, therefore, the search space is huge.

Among the different quality measures proposed in the literature the basic idea of Bayesian quality measures is to assign to every BN a quality value that is a function of the posterior probability distribution of the available data $D = \{y_t^1, \ldots, y_t^{100}\}$ (with the index t running daily from 1979 to 1993), given the BN (M, θ) with network structure M and the corresponding estimated probabilities θ. The posterior probability distribution $p(M, \theta|D)$ is calculated as follows:

$$p(M, \theta|D) = \frac{p(M, \theta, D)}{p(D)} \propto p(M)p(\theta|M)p(D|M, \theta), \quad (2)$$

Geiger and Heckerman [11] consider multinomial networks and assume certain hypothesis about the prior distributions of the parameters, leading to the quality measure

$$\begin{aligned} log\, p(M) \quad &+ \quad \sum_{i=1}^{n} \left[\sum_{k=1}^{s_i} \left[\log \frac{\Gamma(\eta_{ik})}{\Gamma(\eta_{ik} + N_{ik})} \right.\right. \\ &+ \quad \left.\left. \sum_{j=0}^{r_i} \log \frac{\Gamma(\eta_{ijk} + N_{ijk})}{\Gamma(\eta_{ijk})} \right] \right], \end{aligned}$$

where n is the number of variables, r_i is the cardinal of the i-th variable, s_i the number of realizations of the parent's set Π_i, η_{ijk} are the "a priori" Dirichlet hyper-parameters for the conditional distribution of node i, N_{ijk} is the number of realizations in the database consistent with $y_i = j$ and $\pi_i = k$, N_{ik} is the number of realizations in the database consistent with $\pi_i = k$ and Γ is the $gamma$ function.

The K2 is a simple greedy search algorithm for £nding a high quality Bayesian network [12]. This algorithm starts with a network with no links, and assumes that the nodes are ordered. For each variable Y_i, the algorithm adds to its parent set Π_i the node that is lower numbered than Y_i and leads to a maximum increment in the chosen quality measure. The process is repeated until either adding new nodes does not increase the quality or a complete network is attained.

Taking advantage of the network's decomposability, the contribution of the variable Y_i with parent set Π_i to the quality of the network is given by

$$\sum_{k=1}^{s_i} \log \frac{\Gamma(\eta_{ik})}{\Gamma(\eta_{ik} + N_{ik})} + \sum_{j=0}^{r_i} \log \frac{\Gamma(\eta_{ijk} + N_{ijk})}{\Gamma(\eta_{ijk})}. \quad (3)$$

This leads to a simple iterative learning algorithm.

For instance, the graph in Fig. 2 and the corresponding probabilities (such as the one given in Table 3) were obtained applying the $K2$ learning algorithm with the GH quality measure to a database of precipitation records covering the period 1979-1993 for the network of stations shown in Fig. 1 (we used an implementation of this algorithm provided in the Matlab BNToolbox [13]).

3.2 Inference

Once a model describing the relationships among the set of variables has been selected, it can then be used to answer queries when evidence becomes available. Before any information is known about the rainfall at the different stations, there is an initial or *a priori* marginal probability for precipitation at each station k, $P(y^k = i), i = 0, 1, 2$. These initial probabilities can be ef£ciently calculated taking advantage of the independence relationships encoded in the graph (see [8] for a detailed description of inference methods in BNs). For instance, Table 2 shows the initial probabilities of some stations. From this table we can see the rain regimes on the geographical area of study are quite similar due to the correspondence with a single hydrographic basin.

	Stations (initial probability $P(y^k)$)				
State	Coruña	Santiago	Santander	Bilbao	Gijon
0	0.579	0.550	0.564	0.567	0.505
1	0.249	0.279	0.290	0.268	0.313
2	0.172	0.171	0.156	0.165	0.182

Table 2. Marginal distributions of some variables of the BN in Fig. 2.

Now, as soon as we receive some information e, the above probabilities $P(y^k)$ may change as a result of this new evidence or knowledge. The way by which the new probabilities $P(y^k|e)$ are calculated is called uncertainty or evidence propagation. There are several methods for uncertainty propagation in the literature. Some of these methods are exact and others are approximate (see [8] for details). For instance, Table 3 shows the effect produced by different pieces of evidence on the probability of rain at different stations. Comparing the probabilities in Table 3 to the corresponding initial probabilities in Table 2 we can see that the in¤uence of evidence is more strong in those stations associated with nodes which are dependent on the evidence variables.

	Stations ($P(y^k|Coruna = 2)$)				
State	Coruña	Santiago	Santander	Bilbao	Gijon
0	**0.00**	<u>0.01</u>	0.57	0.58	0.51
1	**0.00**	<u>0.14</u>	0.31	0.25	0.30
2	**1.00**	<u>0.85</u>	0.12	0.17	0.19

	Stations ($P(y^k|Bilbao = 2)$)				
State	Coruña	Santiago	Santander	Bilbao	Gijon
0	0.58	0.55	<u>0.12</u>	**0.00**	<u>0.49</u>
1	0.25	0.28	<u>0.26</u>	**0.00**	<u>0.37</u>
2	0.17	0.17	<u>0.62</u>	**1.00**	<u>0.24</u>

Table 3. Conditional probability distributions given the evidences $Coruna = 2$ and $Malaga = 2$ of some variables of the BN in Fig. 2. Evidence has been boldfaced and probabilities signi£cantly changing with evidence have been underlined.

3.3 Validation of the Bayesian Network Forecast Model

To check the quality of BN in a simple case, we shall apply this methodology to a nowcasting problem. In this case we are given a forecast in a given subset of stations and we need to infer a prediction for the remaining stations in the network. To this aim, consider we are given predictions in the £ve stations of the primary network shown as black circles in Fig. 1(b). These predictions shall be plugged in the network as evidences, obtaining the probabilities for the remaining stations in the secondary network.

We considered the precipitation time series corresponding to the period December 1999, and January and February 2000 (90 validation days). Predictions for each station are treated as individual forecasts and the skills of the different stations are combined into the £nal outcome. As a standard validation procedure for probabilistic forecasts, we use the Brier Skill Score (BSS [14]), which compares our forecast method against a reference prediction systems (climatic values) (see [15] for a detailed description of validation measures for probabilistic forecast):

$$BSS = 1 - \frac{BS_p}{BS_c}, \qquad (4)$$

where the Brier Score of the prediction is computed as $BS_p = \sum_{i=1}^{100}\sum_{j=1}^{m}(p_{ij} - \hat{p}_{ij})^2$, where $\hat{p}_{ij}$ is the predicted probability at station i for case j ($P(y^i = j)$), and BS_c is the Brier Score given by the associated climatology (p_{ij} is given by the frequency of precipitation case j in station i for the analyzed period). Positive values of the BSS indicate larger skill of the prediction method when compared with the climatic forecast; on the contrary, negative values indicate a poor performance of the method.

Figs. 3(a) and (b) show the BSS values for the 90 validation days corresponding to the BN model forecasts associated with the binary events $Pp > 2mm$ and $Pp > 10mm$, respectively. From this £gure we can see a predominance of positive skill values with mean 0.42 and 0.33, respectively. On the other hand, Figs. 3(c) and (d) show the Relative Operating Characteristics (ROC) curves (hit rate vs. false alarms). From these £gures we can see how the BN provides a computationally ef£cient and skilful nowcasting prediction system for extrapolating predictions on an heterogeneous spatial medium.

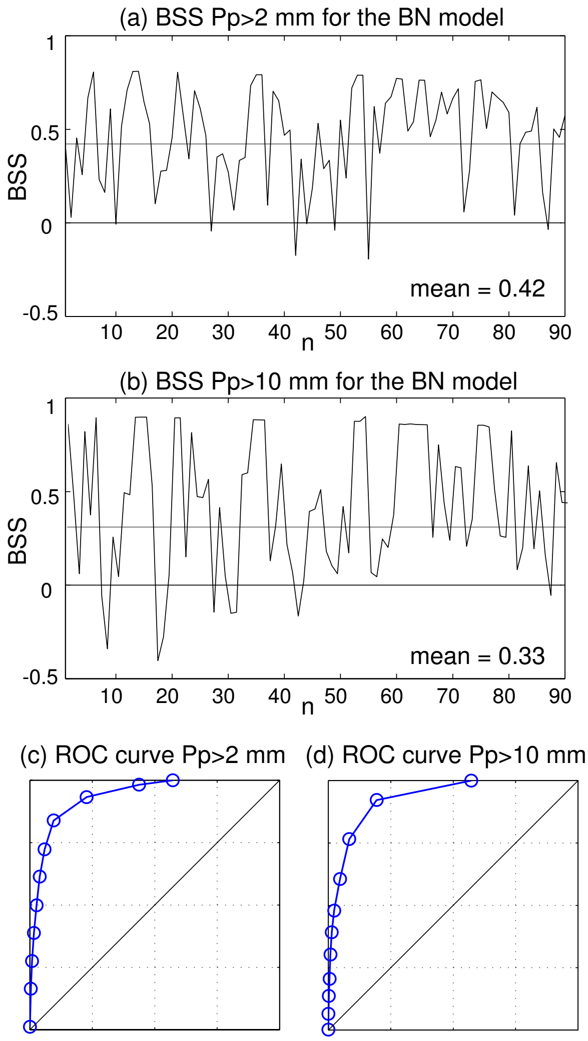

Figure 3. (a) Validation BSS (Brier Skill Score) for $Pp > 2mm$ of the BN forecast model relative to the climatology, the period DJF 1999 (90 days); (b) BSS for $Pp > 10mm$; (c) and (d) ROC curves. Precipitation values at £ve stations of the primary network are used as evidence.

4 CONNECTING WITH NUMERICAL ATMOSPHERIC MODELS

In the above sections we have illustrated the use of BNs by only considering historical data. However, from an operative point of view, it is necessary to connect the above models with the atmospheric patterns $\mathbf{y_t}$, provided by an ACM [16]. Then, suppose that besides the historical precipitation records, x_t^k, we are given a simultaneous database of atmospheric circulation patterns $\mathbf{y_t}$ (integrated by an atmospheric ACM). In this paper we shall use the daily atmospheric patterns given by the ECMWF reanalysis project ERA-15 covering the period from 1979 to 1993 [17]. Each pattern is given by the values of several fundamental and derivate meteorological variables (temperature, geopotential, convective and large scale precipitation, etc.) at six pressure levels (from 300 to 1000 mb). The geographical area of interest in this work is the Iberian Peninsula. Therefore, we restrict the reanalysis to the $1^\circ \times 1^\circ$ lat and lon grid covering the area of interest shown in Fig. 4.

Since we are interested in rainfall forecast, we shall use the gridded forecasts of total precipitation given by the operative ECMWF model (these values are obtained by adding both the convective and the large scale precipitation outputs). The forecasts are obtained 24 hours ahead; therefore, they give a numeric estimation of the future precipitation pattern (one day ahead) on a coarse-grained resolution grid.

The dependency among ACM gridded patterns and local observatories' values can be established by including the grid points as new nodes of the network and using the training algorithm above described to obtain the new structure and parameters using both historical and reanalysis databases. The BN resulting from this process is shown in Fig. 5.

Now, we can use as evidence the gridded precipitation patterns valid one day ahead, and obtain the probabilities of precipitation in the hundred observatories considered in this work. Fig. 6 shows the BSS and ROC curves obtained in this case. Note this hybrid numerical-statistical technique is fed with predictions of the operative model and, therefore, can be directly linked to the weather fore-

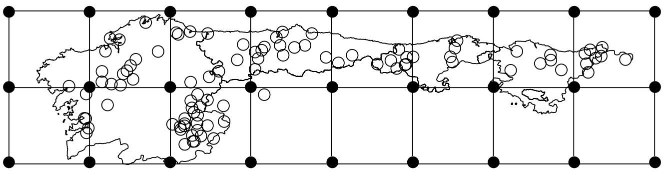

Figure 4. $1^\circ \times 1^\circ$ lat and lon grid of the ACM reanalysis model covering the region under study.

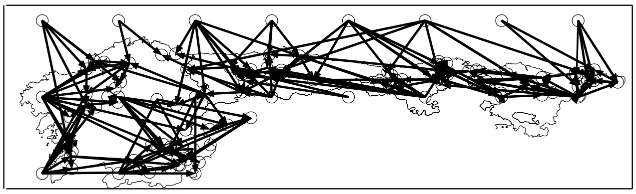

Figure 5. Bayesian network of precipitation grid points and local precipitation at the network of local stations.

cast products developed diary in weather institutes.

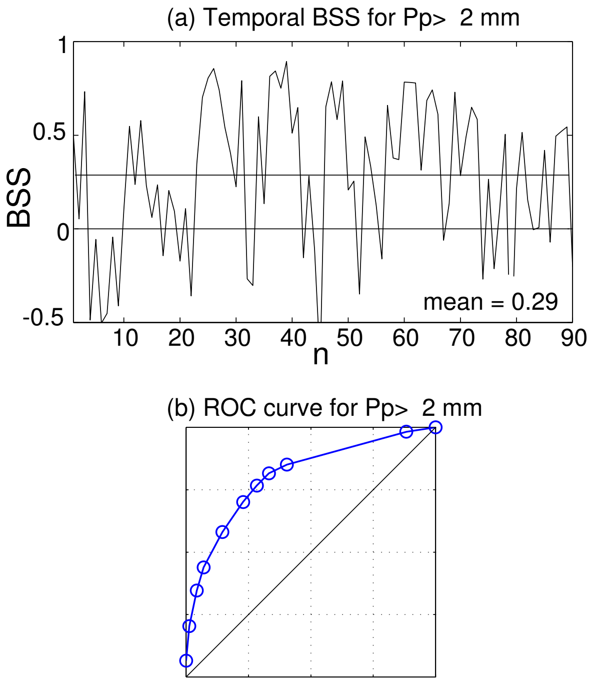

Figure 6. BSS and ROC curve for the validation of the forecast model the period DJF 1999 (90 days). Precipitation gridded patterns are used as evidence.

5 CONCLUSIONS AND FUTURE WORK

We have introduced probabilistic networks and show their applicability for local weather forecasting and downscaling. The preliminary results presented in this paper only illustrate how such models can be built and how they can use for performing inference (obtaining conditional probabilities of nodes given some evidence). Further analysis is still needed for determining the practical operative ef£ciency of these models; £rst experiments are being promising. We are currently working on adapting existing learning algorithms for dealing with this special problem, where the evidence is always given in the same subset of variables.

More information about this project can be obtained from the Web page http://grupos.unican.es/ai/meteo

Acknowledgments

The authors are grateful to the Instituto Nacional de Meteorología (INM) for providing us with the necessary data for this work. The authors are also grateful to the University of Cantabria, INM and the Comisión Interministerial de Ciencia y Tecnología (CICYT) (Project REN2000-1572) for partial support of this work.

REFERENCES

[1] W. Enke and A. Spekat, 'Downscaling climate model outputs into local and regional weather elements by classi£cation and regression', *Climate Research*, **8**, 195–207, (1997).

[2] E. Bellone, J.P. Hughes, P. Guttorp, 'A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts'. *Climate research*, **15(1)**, 1–15, (2000).

[3] M.W. Gardner and S.R. Dorling, 'Arti£cial Neural Networks (the Multilayer Perceptron) - A Review of Applications in the Atmospheric Sciences', *Journal of Applied Meteorology*, **39**, 147–159, (1998).

[4] W. W. Hsieh and B. Tang, 'Applying Neural Network Models to Prediction and Data Analysis in Meteorology and Oceanography', *Bulletin of the American Meteorological Society*, **79(9)**, 1855– 1870, (1998).

[5] E.N. Lorenz, 'Atmospheric Predictability as Revealed by Naturally Occurring Analogues', *Journal of the Atmospheric sciences*, **26**, 636–646, (1969).

[6] T. Cavazos, 'Using Self-Organizing Maps to Investigate Extreme Climate Event: An Application to Wintertime Precipitation in the Balkans', *Journal of Climate*, **13**, 1718–1732, (2000).

[7] A.B. McBratney and A.W. Moore . 'Application of Fuzzy Sets to Climatic Classi£cation', *Agricultural and Forest Meteorology*, **35**, 165–185, (1985).

[8] E. Castillo, J. M. Gutiérrez, and A. S. Hadi, *Expert Systems and Probabilistic Network Models*, Springer-Verlag, New York, 1997. (Free Spanish version URL: personales.unican.es/gutierjm).

[9] B. Abramson, J.M. Brown, W. Edwards, A. Murphy, and R.L. Winkler, 'Hail£nder: A Bayesian System for Forecasting Severe Weather'. *International Journal of Forecasting*, **12**, 57–71, (1996).

[10] D. Heckermean, 'A Tutorial on Learning with Bayesian Networks', *Tech. Rep. MSR-TR-95-06*, Microsoft Research, March (1996).

[11] D. Geiger and D. Heckerman, 'A Characterization of the Dirichlet Distribution with Application to Learning Bayesian Networks', *Proceedings of the Eleventh Conference on Uncertainty in Arti£cial Intelligence*. Morgan Kaufmann Publishers, San Francisco, 196–207, (1995).

[12] G.F. Cooper, and E., Herskovitz, 'A Bayesian method for the induction of probabilistic networks from data', *Machine Learning*, **9**, 309–347, (1992).

[13] K.P. Murphy, 'The Bayes net toolbox for Matlab', *Computing Science and Statistics*, **33**, 110–127, (2001). URL: www.cs.berkeley.edu/ murphyk/Bayes/bayes.html

[14] G.W. Brier, 'Veri£cation of forecasts expressed in terms of probabilities' *Monthtly Weather Review*, **78**, 13–22, (1950).

[15] D. S. Wilks, *Statistical methods in the atmospheric sciences: an introduction*, Academic Press, 1995.

[16] R. Cano, F.J. López, A.S. Co£ño, J.M. Gutiérrez and M.A. Rodríguez, 'Aplicación de Mtodos de Clasi£cación al Downscaling Estadístico', in *Actas del V Simposio Nacional de Prediccin*, Instituto Nacional de Meteorología, 235–240, (2001).

[17] ECMWF Re-Analysis. URL: www.ecmwf.int/research/