

Dynamic Bayesian Networks for Rainfall Forecasting

R. Cano

*Instituto Nacional de Meteorología,
CMT CAS, Santander, Spain
rct@inm.es,*

J.M. Gutiérrez

*Dept. of Applied Mathematics, University of Cantabria,
E-39005, Santander, Spain
gutierjm@unican.es,
<http://personales.unican.es/gutierjm>*

Abstract

In this paper we deal with the problem of forecasting local rainfall at multiple meteorological stations over the Iberian peninsula. To this aim a dynamic Bayesian network is introduced for relating rainfall to broad-scale atmospheric circulation patterns. In this way statistical historic information gathered at the available stations is combined with numerical atmospheric predictions developed at different weather services, resulting a single consensus prediction. This technique can be considered an hybrid statistical-numerical method for precipitation downscaling (predicting local values based on broad-scale grided predictions), and can be easily adapted to other meteorological variables of interest.

Keywords

Weather forecasting, downscaling, learning, expert systems, probabilistic networks, temporal modeling.

1 Introduction

Nowadays, the problem of weather forecasting 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. 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. With the aim of gaining sub-grid detail in the prediction of regional areas of interest, several methods have been proposed in the literature; these techniques are referred to as *downscaling methods*. Several downscaling procedures apply different standard statistical techniques to model the relationship between gridded atmospheric patterns and local climate variables (e.g., rainfall) at multiple local meteorological stations (regression analysis [1], hidden Markov models [2], etc.). However these techniques assume different statistical independence relationships that may neglect important information among the variables in the model.

In this paper we illustrate how Bayesian networks (BNs) and dynamical Bayesian networks (DBNs) offer a sound and practical methodology for building probabilistic models from data with. In BNs, relationships among the variables are represented by a graph, which also gives a simple factorized form of the probability distribution of the variables [3]. Similarly, DBNs use dynamical graphical graph structures to model time dependence in an intuitive way. First, we shall use BNs to capture the spatial dependence among rainfall in different meteorological stations over the Iberian peninsula. Then, we shall analyze different dynamic networks for representing temporal relationships among the rainfall BN and the atmospheric patterns given by an ACM.

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, and analyze different DBNs for this task. Finally, some conclusions and further remarks are given in Sec. 5.

2 Statistical Methods for Forecasting Rainfall

Suppose we are given a database of atmospheric circulation patterns \mathbf{x}_t (integrated by an atmospheric ACM) and the simultaneous historical precipitation records, y_t^k , at a local station of interest k (hereafter \mathbf{x}_t stands for the vector (y_1, \dots, y_t) ; analogously, the super-index denotes a particular grid point, or a particular meteorological station). 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 [4]. Each pattern is given by the daily Temperature (T), relative Humidity (H), Geopotential (Z) and U, V wind components 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 local grid covering the area of interest shown in Fig. 1(a).

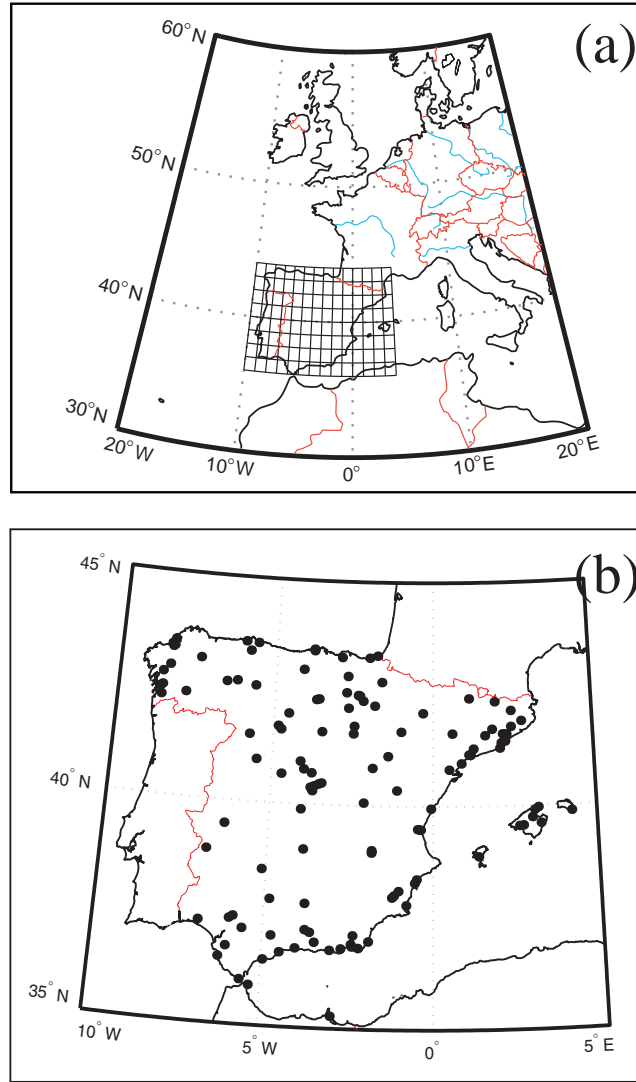


Figure 1: (a) High resolution limited area grid with 1.0° (approx. 80 Km) for both latitude and longitude. (b) Primary network of 112 climatic stations of the INM.

In regarding the local climate records database, we shall use rainfall values from the primary network of 112 climatic stations from the the Spanish National Weather Service –Instituto Nacional de Meteorología, INM– (see Figure 1(b)).

Given this information, a standard statistical procedure for obtaining local predictions is fitting the linear model $y_T^k = \mathbf{a}\mathbf{x}_T + b + \epsilon$ using paired samples $(\mathbf{x}_n, y_n^k), n = 1, \dots, t$ of data. This allows adapting a gridded prediction of present weather to the local climate in an straightforward way (see, e.g. [5]). Other authors suggests more general nonlinear models using modern nonparametric techniques (feedforward neural networks [6], etc.). The main shortcoming of these methods is that they assume spatial independence of precipitation records at different stations. However, in practise, there is a strong spatial correlation among different stations and, therefore, the above models neglect an important information which may spoil their practical results.

This problem could be solved by considering a multiple regression model $y_T^k = \mathbf{a}\mathbf{z}_T + b + \epsilon$, where $\mathbf{z}_T = (\mathbf{x}_T, y_T^1, \dots, y_T^{112})$. However, such a model is useless due to larger number of variables involved which involves the joint probability distribution (JPD) of a large set of variables $P(\mathbf{x}_T, y_T^1, \dots, y_T^{112})$.

An alternative and efficient solution for this problem is determining the strongest dependencies among the variables y^1, \dots, y^{112} , obtaining a simpler joint probability distribution. This task can be easily accomplished using Bayesian networks, which provide a graphical framework for analyzing dependencies when dealing with uncertainty.

3 Bayesian Networks

The basic idea of Bayesian networks (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 $\{x_1, \dots, x_n\}$). These nodes are connected by arrows, which represent a cause and effect relationship. That is, if there is an arrow from node x_i to node x_j , we say that x_i is the cause of x_j , or equivalently, x_j is the effect of x_i . Another popular terminology of this is to say that x_i is a parent of x_j or x_j is a child of x_i . For example, in Figure 2, the node *Valladolid* is a parent of *Leon* and a child of *Burgos* and *Segovia* (the set of parents of a node x_i is denoted by π_i).

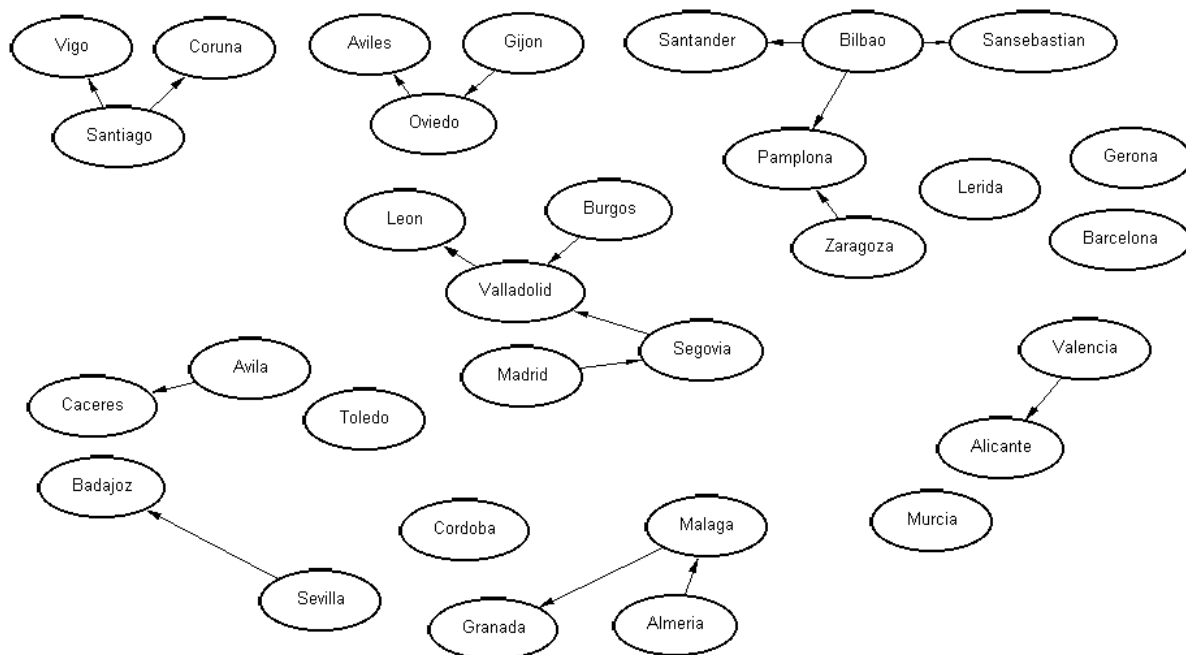


Figure 2: Directed graph associated with 31 main stations over the Iberian peninsula (the nodes are displayed maintaining their spatial geographical disposition).

Directed graphs provide a simple definition of independence (d-separation) based on the existence or not of certain paths between the variables in the graph (see [3] for a detailed introduction to probabilistic network models). BNs are probabilistic models consisting of a directed acyclic graph, and a JPD of the set of variables with the dependency/independency structure displayed by the graph. This is facilitated by a well known result in probability theory which states that any JPD of the variables can be expressed as a product of several conditional distributions as follows:

$$Pr(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | x_{i+1}, \dots, x_n). \quad (1)$$

This result has at least two implications:

1. The JPD is determined by the product of smaller distributions.
2. The JPD can be represented graphically. Variables are represented by nodes and for each of the conditional distributions $P(x_j | x_{j+1}, x_{j+2}, \dots, x_n)$, a link is drawn from each of the variables $x_{j+1}, x_{j+2}, \dots, x_n$ to variable x_j .

The JPD factorized according a BN is given by the simpler factorization

$$Pr(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | \pi_i), \quad (2)$$

since some of the variables are independent in the graph and, therefore, some of the conditional probabilities in (1) can be simplified.

A BN is defined as a directed acyclic graph, together with a factorization of the JPD of the variables given by (2). For instance, the JPD of a BN defined by the graph given in Fig. 2 requires the specification of 31 conditional probabilities, one for each variable conditioned to its parents' set. Fig. 3 shows some of these probability tables, considering rainfall discretized into four different states (0="no rain", 1="weak rain", 2="moderate rain", 3="heavy rain"), associated with the thresholds 0, 1, 10, and 50 l/m^2 , respectively.



Figure 3: Conditional probability tables $P(Santiago)$, $P(Madrid)$, $P(Almeria)$, $P(Leon|Valladolid)$, and $P(Segovia|Madrid)$ corresponding to the BN given by the graph in Fig. 2.

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 recently introduced for learning the graphical structure and estimating probabilities from data (we do not discuss this issue here, see [3, 7] for a review). For instance, Fig. 4 shows a database of precipitation records, starting at 1/1/1985. Note that the database includes some missing values (aprox. 15%) corresponding to mistakes and lost information while elaborating the information.

BARCELONA	GERONA	SANSEBASTIAN	BILBAO	SANTANDER	GIJON	AVILES	OVIEDO	CORUNA	SANTIAGO	VIGO	BURGOS	VALLADOLID	AVILA	SEGOVIA
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	2	3	2	3	3	3	2	2	3	1	1	0	1
0	1	2	2	1	1	2	2	0	1	0	0	0	0	0
0	0	2	1	2	2	2	3	2	2	1	NA	0	0	0
0	0	2	2	2	2	1	2	0	1	0	1	0	0	0
NA	1	0	0	0	0	0	0	0	NA	0	0	0	0	0
0	0	0	NA	1	1	0	0	3	3	3	2	2	NA	0
1	1	NA	NA	0	0	1	NA	3	3	3	2	2	2	1
2	2	1	NA	1	NA	0	0	2	2	3	0	1	NA	0
0	0	NA	1	1	1	1	1	2	3	3	2	2	1	1
1	0	1	1	1	1	1	1	2	3	3	3	2	2	2
0	0	0	NA	0	1	1	1	2	3	3	1	2	0	0
1	0	0	0	1	0	1	NA	2	3	2	2	2	2	2
0	1	1	1	1	1	1	0	2	3	3	1	0	0	0
0	0	0	NA	0	0	0	0	2	3	3	1	1	0	0
0	0	0	0	0	0	0	0	NA	2	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 4: Database of precipitation records covering a period of five years (1985-1990). Symbol “NA” denotes missing values.

The graph in Fig. 2 and the corresponding probabilities (as those given in Fig. 3) were obtained using the learning algorithm for partial data (missing values) described in [8]. This algorithm consists of a three phase mechanism. The three phases are: drafting, thickening and thinning. In the first phase, this algorithm computes the mutual information of each pair of nodes as a measure of closeness and creates a draft based on this information. In the second phase, the algorithm adds links when the pairs of nodes cannot be d-separated. In the third phase, each link of the current graph is examined using independence tests and are removed if the two nodes of the edge can be d-separated (this results in a sparse graph). Once the graph is known, the required probabilities can be estimated from data using the EM algorithm. The resulting probabilistic model keeps all the dependency structures embedded into the data, up to a threshold value given by the user for the mutual information; below this value, no link will be added between couples of nodes. For instance, the graph in Fig. 2 only includes the strongest couples when applying the algorithm.

From this graph we can see how strong links (strong dependencies) are established between

geographically close stations (we can even distinguish different basins from this figure). If the same algorithm is applied with half threshold value, the graph in Fig. 5 is obtained. We shall use this last graph in the following sections.

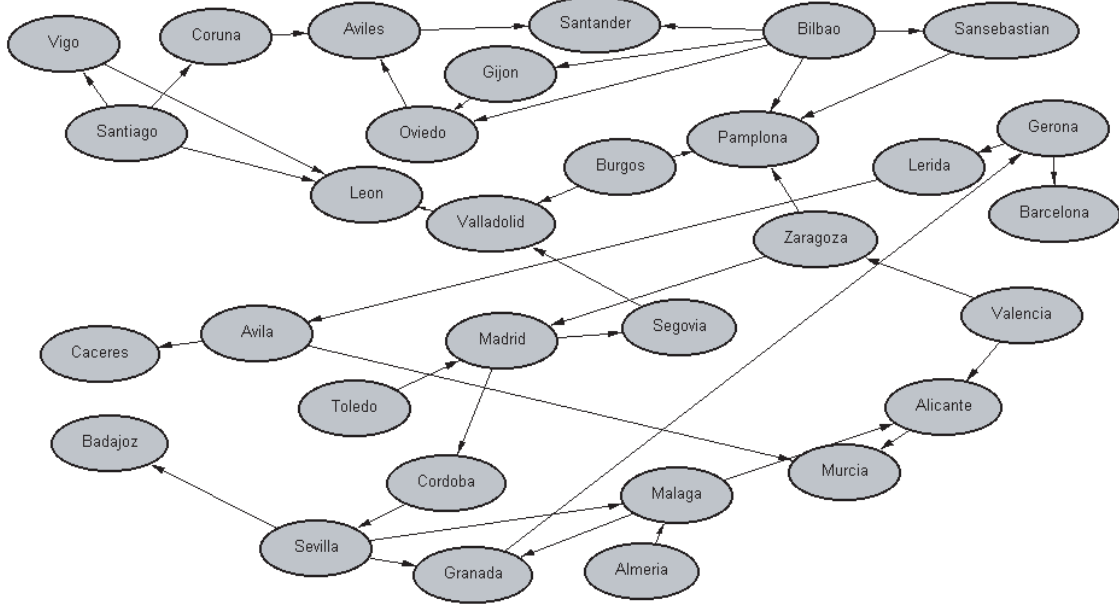


Figure 5: Directed graph obtained with a low mutual information threshold value.

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 , $P(x_i = k), k = 0, 1, 2, 3$. These initial probabilities can be efficiently calculated taking advantage of the independence relationships encoded in the graph (see [3] for a detailed description of inference methods in BNs). For instance, Table 1 shows the initial probabilities of some nodes. From this table we can see the different rain regimes on the geographical area of study. The most rainy station is *Santiago* with almost half probability of rain. On the other hand, *Malaga* has less than 10% probability of rain.

Now, as soon as we receive some information e , the above probabilities $P(x_i)$ may change as a result of this new evidence or knowledge. The way by which the new probabilities $P(x_i|e)$ 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 [3] for details). For instance, Table 2 shows how different pieces of evidence produce changes of different conditional probabilities (those associated with nodes which are dependent on the evidence variables). The effect of the

Stations (initial probability $P(x_i)$)								
State	Coruña	Santiago	Santander	Bilbao	Madrid	Valladolid	Barcelona	Málaga
0	0.579	0.550	0.584	0.587	0.808	0.788	0.842	0.901
1	0.209	0.189	0.240	0.218	0.131	0.152	0.050	0.061
2	0.174	0.172	0.150	0.161	0.057	0.056	0.041	0.028
3	0.038	0.089	0.027	0.034	0.004	0.004	0.021	0.010

Table 1: Initial marginal probability distributions of some variables of the BN in Fig. 5.

evidence on the probabilities of other diseases and symptoms can be seen by comparing the probabilities in Table 2 to the corresponding initial probabilities in Table 1.

Stations ($P(x_i Coruna = 3)$)								
State	Coruña	Santiago	Santander	Bilbao	Madrid	Valladolid	Barcelona	Málaga
0	0.000	<u>0.001</u>	<u>0.519</u>	0.587	0.808	0.788	0.842	0.901
1	0.000	<u>0.047</u>	<u>0.265</u>	0.218	0.131	0.152	0.050	0.061
2	0.000	<u>0.254</u>	<u>0.174</u>	0.161	0.057	0.056	0.041	0.028
3	1.000	<u>0.699</u>	<u>0.042</u>	0.034	0.004	0.004	0.021	0.010

Stations ($P(x_i Malaga = 3)$)								
State	Coruña	Santiago	Santander	Bilbao	Madrid	Valladolid	Barcelona	Málaga
0	0.579	0.550	0.584	0.587	0.808	0.788	<u>0.771</u>	0.000
1	0.209	0.189	0.240	0.218	0.131	0.152	<u>0.109</u>	0.000
2	0.174	0.172	0.150	0.161	0.057	0.056	<u>0.077</u>	0.000
3	0.038	0.089	0.027	0.034	0.004	0.004	<u>0.043</u>	1.000

Table 2: Conditional probability distributions given the evidences $Coruna = 3$ and $Malaga = 3$ of some variables of the BN in Fig. 5. Conditional probabilities different from the initial ones are underlined. Evidence has been boldfaced.

In order to complete the above BN, besides of the historical precipitation records, y_t^k , we can also consider the simultaneous series of states of the atmosphere \mathbf{x}_t , provided by an ACM. In this way, the probability distribution over the 31 stations factorized according to the graph will also depend on a new variable AS (Atmospheric State):

$$P(y^1, \dots, y^{31}, AS) = P(y^1, \dots, y^{31}|AS)P(AS) = \prod_{i=1}^{31} P(y^i|\pi^i, AS)P(AS). \quad (3)$$

Therefore, the resulting directed graph contains a new directed link from variable AS to all other variables in the network, as shown in Fig. 6. The state variable can be discretized by performing a clustering algorithm (e.g., the k-means algorithm) to the ERA-15 database, associating a specific state to each of the clusters resulting from this process.

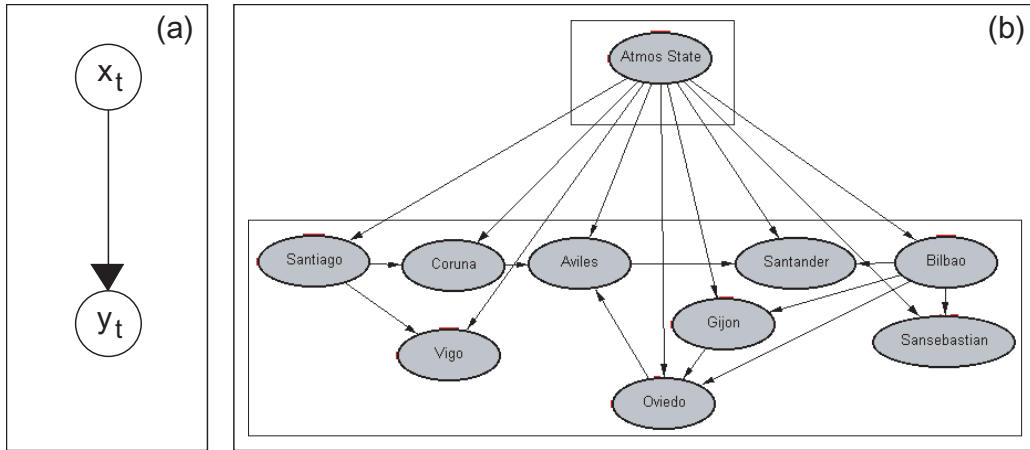


Figure 6: Adding a state variable to the model shown in Fig. 5. (a) Schematic model. (b) Complete model (only the North basin is displayed in the figure).

4 Dynamic Bayesian Networks for Rainfall Forecast

Fig. 7 shows a dynamic Bayesian network (DBN) which represents the evolution over time of variables y and x ; only three time slices are shown, since we shall only consider first order interactions between the variables at different times. In this case we have two different types of links. Contemporary relationships are established between variables in the same time slice; we assume that all the contemporary relationships are given by a directed acyclic graph which is invariant over time (we consider the graph described in Fig. 6). On the other hand, since the value of x_t will be always available from an ACM, we only consider non-contemporary relationships between each rainfall variable y_t^k and the same variable at the next time step y_{t+1}^k . The graph given in Fig. 7 represents these relationships.

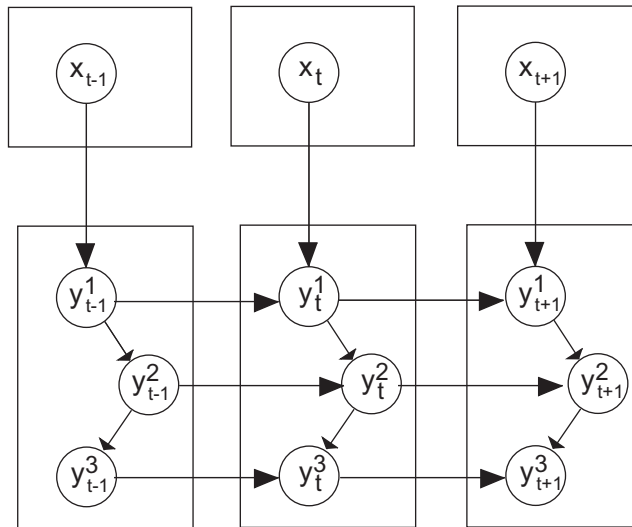


Figure 7: Sketch of a DBN considering only the variables at time $t - 1$, t , and $t + 1$.

Therefore, from the above graphical structure we have:

$$P(y_t^k | \mathbf{y}_{t-1,1} x_{t-1}) = P(y_t^k | y_{t-1}^k, \pi_t^k, x_t) \quad (4)$$

Taking into account that for every instant t , that some values from \mathbf{y}_{t-1} and x_t are available as evidence at time t , rainfall forecasting will be obtained by computing the probabilities (4) according to the graph, conditioned on the observed evidence in each case. We started performing some experiments to check the efficiency of this method for operative forecasting; the results will be soon published elsewhere.

5 Conclusions and Future Work

We have introduced dynamical 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 efficiency of these models; first experiments are being promising. The models presented in this paper can be also very useful to perform sensitivity analysis of the conditional probabilities of nodes, depending on the parameters of the model (initial marginal or conditional probabilities of different stations or atmospheric patterns). This could be done by generalizing the result found in Bayesian networks for conditional means and variances of the nodes given the evidence [9].

Acknowledgments

The authors are grateful to Marco Ramoni and Paola Sebastiani, the Open University (<http://kmi.open.ac.uk/projects/bkd>); some of the results in this paper have been obtained using their software “Bayesian Knowledge Discoverer”. The authors are also grateful to the University of Cantabria, the Instituto Nacional de Meteorología and the Comisión Interministerial de Ciencia y Tecnología (CICYT) (Project REN2000-1572) for partial support of this work.

References

- [1] J. Billet, M. Delisi, B.G. Smith, and C. Gates Use of Regression Techniques to Predict Hail Size and the Probability of Large Hail Weather and Forecasting, **12**, 154-164, 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] E. Castillo, J. M. Gutiérrez, and A. S. Hadi, *Expert Systems and Probabilistic Network Models* (Springer-Verlag, 1997).
- [4] ECMWF Re-Analysis (ERA) Project. <http://www.ecmwf.int/research/era/index.html>
- [5] w. Enke and A. Spekat Downscaling climate model outputs into local and regional weather elements by classification and regression. *Climate Research*, **8**, 195-207, 1997.
- [6] M.W. Gardner and S.R. Dorling Artificial Neural Networks (the Multilayer Perceptron) - A Review of Applications in the Atmospheric Sciences. *Journal of Applied Meteorology*, **39**, 147-159, 1998.
- [7] D. Heckerman A Tutorial on Learning with Bayesian Networks. Tech. Rep. MSR-TR-95-06, Microsoft Research, March 1996.
- [8] J. Cheng, D.A. Bell and W. Liu. An algorithm for Bayesian belief network construction from data, *Proceedings of AI-STAT'97*, Ft. Lauderdale, Florida, 1997.
- [9] Castillo, E., Gutiérrez, J. M., and Hadi, A. S. (1996), Sensitivity Analysis in Discrete Bayesian Networks. *IEEE Transactions on Systems, Man and Cybernetics*, **27**(4) 412-42.