Applications of Bayesian Networks in Meteorology.

Rafael Cano[1], Carmen Sordo[2], and José M. Gutiérrez[2,3]

[1] Instituto Nacional de Meteorología
   CMT/CAS, Santander, Spain.
[2] Dpto. Matemática Aplicada y C.C. (grupo de IA en Meteorología),
   Universidad de Cantabria, 39005, Santander, Spain.
[3] (corresponding author): gutierjm@unican.es, http://grupos.unican.es/ai/meteo

Abstract. In this paper we present some applications of Bayesian networks in Meteorology from a data mining point of view. We work with a database of observations (daily rainfall and maximum wind speed) in a network of 100 stations in the Iberian peninsula and with the corresponding gridded atmospheric patterns generated by a numerical circulation model. As a first step, we analyze the efficiency of standard learning algorithms to obtain directed acyclic graphs representing the spatial dependencies among the variables included in the database; we also present a new local learning algorithm which takes advantage of the spatial character of the problem. The resulting graphical models are applied to different meteorological problems including weather forecast and stochastic weather generation. Some promising results are reported.

1 Introduction

The increasing availability of climate data during the last decades (observational records, radar and satellite maps, proxy data, etc.) led to the development of efficient statistical [1,2] and data mining [3] techniques in different areas of the atmospheric sciences, including climate variability and local weather forecast. Among this huge amount of data there are two important sources of information for statistical applications:

- *Climatological databases*, which collect local information (e.g., historical observations of precipitation or wind speed) from networks of stations within specific geographical regions.
- *Reanalysis databases*, which gather outputs of numerical atmospheric circulation models for long periods of time and different lead times (forecast steps). These outputs are gridded values of temperature, humidity, etc., on appropriate 3D grids covering the world, or an area of interest.

Climatological databases contain the statistical properties of the different local climatology existing in an area of study, whereas reanalysis databases contain the evolution of the atmosphere simulated by numerical models. Thus,

this information can be used to analyze problems related with the atmosphere dynamics and with the possible impacts it may cause on the climatology of local regions. Some examples of application of statistical techniques in these problems include linear regression methods [4] for weather forecast, clustering methods [5] and principal component analysis [6] for identification of representative atmospheric patterns, etc. However, these techniques deal with special problems and with particular sets of data and they both fragment the information and assume ad-hoc spatial independencies in order to simplify the resulting problem-driven models.

In this paper we introduce probabilistic graphical models (Bayesian networks) in Meteorology as a data mining technique [7,8]. Bayesian networks automatically capture probabilistic information from data using directed acyclic graphs and factorized probability functions. The graph intuitively represents the relevant spatial dependencies and the resulting factorized probability function leads to efficient inference algorithms for updating probabilities when new evidence is available. Thus, Bayesian networks offer a sound and practical methodology for discovering probabilistic knowledge in databases and for building intuitive and tractable probabilistic models which can be easily used to solve a wide variety of problems.

We deal both with the automatic construction of Bayesian networks from meteorological data and with the application of the resulting models to different interesting problems. We assume that the audience is not familiar with meteorological concepts, so the prerequisites have been kept to a minimum and special care has been made to give all the necessary meteorological background for a full understanding of the examples.

The paper is organized as follows. We start in Sec. 2 describing the sources of data used in this paper. Then, we introduce some interesting meteorological problems and show their relationship with the data (Sec. 3). These problems are used throughout the paper to illustrate the application of Bayesian networks. In Section 4 we briefly describe Bayesian networks and analyze different methods to learn these models from data; Sec. 4.4 describes probabilistic inference in these models. Finally, in Sec. 5 several applications of Bayesian networks are described.

2   Area of Study and Available Data

In this work we consider the Iberian Peninsula as the geographical area of interest, and use daily data (precipitation and maximum wind speed) from a 100 stations' network (see Fig. 1(a)) provided by the Spanish weather service (Instituto Nacional de Meteorología, INM). The data covers the period from 1959 to 2000 and is representative of the local climatology of this area. The variables are considered continuous for some applications, but are also quantized for other applications. Precipitation (*Precip*) is quantized into four different states (1="dry or very light rain",2="light rain", 3="moder-

ate rain" and 4="heavy rain"), according to the thresholds 0, 2, 10 and 20 mm/day, respectively (for instance, the event "heavy rain" corresponds to $Precip > 20mm$). Similarly, maximum wind speed (*Wind*) is quantized into three different states corresponding to the thresholds 0, 40 and 120 Km/h, respectively.
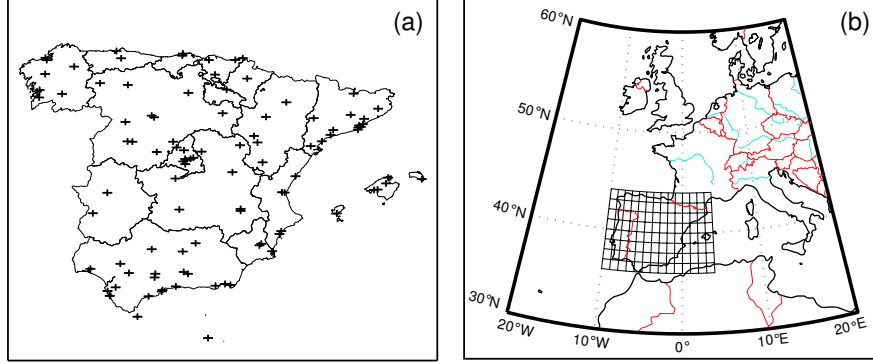

Fig. 1. (a) Geographical area of study and the location of the 100 stations considered in this work; (b) Fragment of the $1° \times 1°$ grid used by the ECMWF operative atmospheric model covering the area of interest.

On the other hand, we consider data which characterizes the state and evolution of the atmospheric variables. In particular, we use daily atmospheric circulation gridded fields provided by the European Center for Medium Weather Forecast (ECMWF) reanalysis project ERA-15 [9]. The fields are obtained integrating a numerical Atmospheric Circulation Model (ACM) on a global grid during the period from 1979 to 1993. To run the model for a particular date, the available observations are assimilated to obtain an initial condition for the atmosphere state (this is called the analysis); then, the integration of the model returns the predicted fields for different lead times (we shall only use the forecast corresponding to the state of the atmosphere 24 hours later). In this paper each circulation pattern is defined on the $1° \times 1°$ latitude and longitude grid covering the area of interest shown in Fig. 1(b). The patterns (data vectors) are formed considering the values of five meteorological variables (temperature, geopotential, U and V wind components, and humidity) at six vertical pressure levels (300, 500, 700, 850, 925, and 1000 mb) −the lower level corresponds to surface, whereas the higher level corresponds to approximately 10Km−. Collecting all this information, we have the following atmospheric pattern:

$$\mathbf{X_t} = (T_t^{1000}, \dots, T_t^{300}, H_t^{1000}, \dots, H_t^{300}, \dots, V_t^{1000}, \dots, V_t^{300}), \qquad (1)$$

where X_i^j denotes the j-pressure level field of variable X for date t. We get a total of 9×15 (grid) $\times 6$ (pressure levels) $\times 5$ (variables) = 4050 dimensions.

Note that the enormous size of these patterns is somehow misleading, since there exists a great correlation on all variables and space scales which makes the effective dimension much lower. In this paper, in order to simplify the treatment of atmospheric variables, we use a quantized variable S for the "state of the atmosphere" obtained by clustering the patterns (1) into 25 different classes using the k-means algorithm (further information and details about this process are given in [3,5]).

3   Some Common Problems in Meteorology

In this section we introduce some interesting problems from Meteorology which are suitable for a probabilistic treatment. We shall use later these problems to illustrate the applications of Bayesian networks, presenting some promising results and describing other interesting applications.

3.1   Statistical Weather Generators

Weather generators have been used extensively in agricultural and hydrological applications where high-spatial resolution and/or long sequential series of records are required to solve common problems [10]. Simply stated, these methods are statistical simulation techniques adapted for producing samples of meteorological variables preserving some observed statistical properties and resembling the local climatology (annual or seasonal means, variances, frequencies of events, etc.). Thus, statistical weather generators provide an alternative for filling in missing data and for producing indefinitely long series from finite observation records of a local station stored in a climatological database.

One of the most important variables for practical applications is precipitation. In this case we are interested both in the absence or presence of the event (discrete variable) and its amount or intensity (continuous). In this paper we focus on the occurrence of daily precipitation (the application of Bayesian networks to the intensity process is related with the problem presented in the following section). Markov chains were the first models applied to simulate precipitation binary series (absence/occurrence) preserving the temporal correlation and the basic statistics of a local station under study. It was shown that a first-order model could capture the persistent nature of daily precipitation (i.e., the basic correlation of the series) [11]. This led to simple and efficient weather generators for simulating series of precipitation occurrence for a single station [12]. However, when simultaneous series for a set of stations are required, the above models are inappropriate and more general frameworks are needed [13]. The simulation of spatial coherent series from multisite data is still a challenging problem.

Another challenging problem is the forecast of meteorological variables (e.g., daily precipitation) at a given location (e.g., Santander city).

3.2 Local Weather Forecast

The main tools available for weather forecast are the outputs of numerical ACMs, which provide future predicted values of atmospheric variables on a grid. However, the resolution of the global grids used in these models and the physical parametrization of sub-grid scale processes –such as cloud formation, orography, etc.– limit the skill of these models to capture variabilities of the weather on a local station (e.g., precipitation or wind speed in the city of Santander); i.e., the output fields are smoothed. This problem motivated the development of different statistical techniques to model and forecast local climatological time series: auto-regressive models [14], embedding techniques [15], neural networks [16], etc. These techniques capture the dynamics underlying the time series fitting the parameters to the historical observed data. However, it has been shown that the knowledge of statistical techniques alone is not sufficient to obtain skillful forecasts.

An alternative solution for this problem is combining the information in both climatological and reanalysis databases. In other words, statistical techniques can be applied to relate model outputs to local observations, leading to forecast models for adapting a gridded forecast to local climates in an straightforward way. For instance, given a database of atmospheric circulation patterns $\mathbf{X}_t$, used as predictors, and simultaneous historical records of a local variable Y_t (predictand), standard and modern statistical methods such as regression analysis and neural networks can be applied to obtain a model $\hat{y}_t = f(\mathbf{x}_t) + \epsilon_t$ for performing local forecast [17]. However, these models assume stationary atmosphere dynamics during the period of study (note that the global model is trained with all the available data), and this is by no means guaranteed.

Local techniques, such as the method of analogs [18] or nearest neighbors, provide a simple solution for this problem, since models are trained locally. In this case, it is assumed that similar circulation patterns lead to similar meteorological events. Thus, local predictions are derived from $\mathbf{X}_t$ by first looking for the set of its Nearest Neighbors (NNs), or analog ensemble, in a circulation reanalysis database using some desirable metric. Then, the local observations registered at the analog ensemble dates are used to obtain a forecast. Recently, some methods have substituted the analog ensemble by a discrete atmospheric circulation state variable S obtained by applying a clustering algorithm (hierarchical, k-means, etc.) to a reanalysis database (see [19,20] for details). The resulting states of the variable correspond to the different clusters ($S = c$ refers to the subset of the reanalysis database associated with cluster c). In this situation, an input pattern is assigned to the closest cluster c_t, and a local probabilistic prediction $P(Y_t = i)$ for the occurrence $Y = i$ at date t is simply obtained as the conditional probability of the phenomenon, given the cluster observations (i.e., given the value of the state variable):

$$P(Y_t = i) = P(Y = i | S = c_t), \ i = 1, \ldots, m, \tag{2}$$

where m are the possible states of variable y (four for the case of precipitation). As pointed out in [21] these models are simple naive Bayes classifiers which do not take into account the spatial dependencies of the variables.

Therefore, modeling spatial dependency among local stations is an important unsolved topic for the two problems above. Bayesian networks provide a simple and sound framework for this task.

4   Bayesian Networks. Learning from Data

In this section we briefly introduce Bayesian networks and describe the available learning algorithms to obtain these models from meteorological data. We also discuss the performance of these methods considering the spatial character of the variables involved.

4.1   Bayesian networks

Bayesian networks are known for providing a compact and simple representation of probabilistic information, allowing the creation of models associating a large number of variables [7,8]. For instance, let us consider the network of climatic stations shown in the graph in Fig. 1(a); for most of the problems described in Sec. 3, we would like to obtain a probabilistic model representing the whole uncertainty about precipitation (or wind speed) in the network. The basic idea of probabilistic networks is encoding the dependencies among a set of variables using a graphical representation (a directed acyclic graph) which is easy to understand and interpret. There is a node for each domain variable (rainfall at each of the stations), and edges connect attributes that are directly dependent on each other. The graph defines a decomposition of the high-dimensional probability density function into low-dimensional local distributions (marginal or conditional).

For instance, the graph shown in Fig. 2 represents a directed acyclic graph where the variables are represented pictorially by a set of nodes; one node for each variable (in this paper we consider the set of nodes $\{Y_1, \ldots, Y_n\}$, where the subindex refers to the station's number). These nodes are connected by directed links, which represent a dependency relationship: If there is an arrow from node Y_j to node Y_k, we say that Y_j is a parent of Y_k, or equivalently, Y_k is a child of Y_j. The set of parents of a node Y_i is denoted as Π_i. For instance, in the inset of Figure 2 each node has a single parent ($\Pi_A = \Pi_B = \{C\}$, $\Pi_C = \{D\}$, and so on). Directed graphs provide a simple definition of independence (d-separation) based on the existence or not of certain paths between the variables (see [8] for a detailed introduction to Bayesian networks).

The dependency/independency structure displayed by the acyclic directed graph can be translated to the joint Probability Density Function (PDF) of
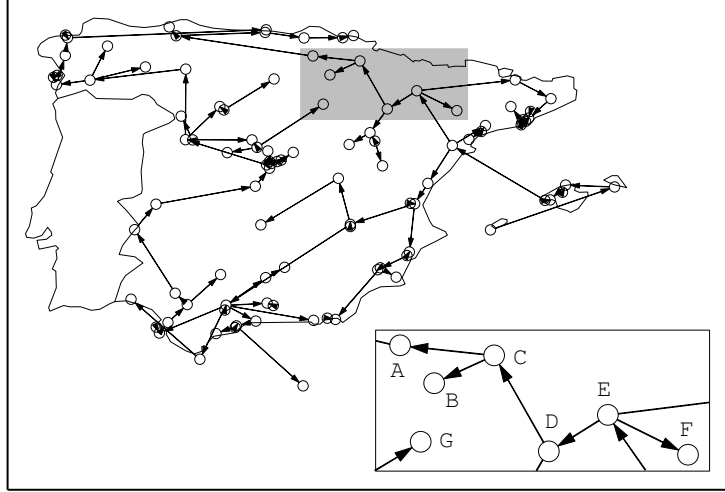
Fig. 2. Example of directed acyclic graph relating the stations shown in Fig. 1(a). The inset shows a magnification of the shaded area. The nodes shown in the inset correspond to the following cities: *A=Vitoria, B=Logroño, C=Pamplona, D=Zaragoza, E=Huesca, F=Lérida, G=Soria.*

the variables by means of a factorization as a product of conditional/marginal distributions as follows:

$$P(y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n} P(y_i | \pi_i). \tag{3}$$

Therefore, the independencies from the graph are easily translated to the probabilistic model in a sound form. Depending on the discrete or continuous character of the variables, the conditional probabilities in (3) are specific parametric families. In this paper we consider two important types of Bayesian networks:

- *Multinomial Bayesian networks.* In a multinomial Bayesian network we assume that all variables are discrete, that is, each variable has a finite set of possible values (in this paper we shall consider four states for precipitation and three states for wind speed, as described in Sec. 2). We also assume that the conditional probability of each variable given its parents is multinomial and, hence, it is specified by a probability table given by the probabilities associated with the different combinations of values of the variables involved. Note that the number of parameters required to specify the full-dependence PDF in the left hand side term of (3) is $m^n = 4^{100}$, whereas the BN involves the specification of 100 conditional probability tables, one for each variable conditioned to its unitary parents' set.

- *Gaussian Bayesian networks.* In a Gaussian Bayesian network, the variables are assumed to have a multivariate normal distribution, $N(\mu, \Sigma)$, whose joint density function is given by

$$f(x) = (2\pi)^{-n/2}|\Sigma|^{-1/2}\exp\left\{-1/2(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}, \qquad (4)$$

  where μ is the n-dimensional mean vector, Σ is the $n \times n$ covariance matrix, $|\Sigma|$ is the determinant of Σ, and μ^T denotes the transpose of μ. In a Gaussian Bayesian network this PDF is specified as in (3) by:

$$f(x_i|\pi_i) \sim N\left(\mu_i + \sum_{j=1}^{i-1}\beta_{ij}(x_j - \mu_j),\ v_i\right), \qquad (5)$$

  where β_{ij} is the regression coefficient of X_j in the regression of X_i on the parents of X_i, Π_i, and $v_i = \Sigma_i - \Sigma_{i\Pi_i}\Sigma_{\Pi_i}^{-1}\Sigma_{i\Pi_i}^T$ is the conditional variance of X_i, given $\Pi_i = \pi_i$, where Σ_i is the unconditional variance of X_i, $\Sigma_{i\Pi_i}$ is the vector of covariances between X_i and the variables in Π_i, and Σ_{Π_i} is the covariance matrix of Π_i. Note that β_{ij} measures the strength of the relationship between X_i and X_j. If $\beta_{ij} = 0$, then X_j is not a parent of X_i. Thus, the Gaussian Bayesian network is given by a collection of parameters $\{\mu_1, \ldots, \mu_n\}$, $\{v_1, \ldots, v_n\}$, and $\{\beta_{ij} \mid j < i\}$, as shown in (5).

4.2   Learning Algorithms

From a practical point of view, the application of Bayes Networks (BNs) in real-world problems depends on the availability of automatic learning procedures to infer both the graphical structure and the parameters of the conditional probabilities (the probability tables for the multinomial case, or the means, variances and regression coefficients for Gaussian networks). Several methods have been recently introduced for learning the graphical structure (structure learning) and estimating probabilities (parametric learning) from data (see [22] and references therein). Among these methods the so-called "score and search" consist of two parts:

1. A *quality measure* for computing the quality of the graphical structure and the estimated parameters for the candidate BNs.
2. A *search algorithm* to efficiently search the space of possible BNs to find the one with highest quality. Note that the number of all possible networks is huge even for a small number of variables, and so it is the search space (the search process has shown to be NP-hard in the general case).

Among the different quality measures proposed in the literature, Bayesian quality measures have a firm theoretical foundation. Each Bayesian net $B = (M, \theta)$, with network structure M and the corresponding estimated probabilities θ, is assigned a function of its posterior probability distribution given the

available data D. This posterior probability distribution $p(B|D)$ is calculated as follows:

$$p(B|D) = p(M, \theta|D) = \frac{p(M, \theta, D)}{p(D)} \propto p(M)p(\theta|M)p(D|M, \theta), \qquad (6)$$

In the case of multinomial networks, assuming certain hypothesis about the prior distributions of the parameters and a uniform initial probability for all models, then the following quality measure is obtained [23]:

$$p(B|D) \propto \sum_{i=1}^{n} \left[\sum_{k=1}^{s_i} \left[\log \frac{\Gamma(\eta_{ik})}{\Gamma(\eta_{ik} + N_{ik})} + \sum_{j=0}^{r_i} \log \frac{\Gamma(\eta_{ijk} + N_{ijk})}{\Gamma(\eta_{ijk})} \right] \right], (7)$$

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 $\pi_i = k$, and N_{ik} is the number of those consistent with $\pi_i = k$. Note that all the parameters in (7) can be easily computed from data.

In the case of Gaussian networks, a similar quality measure can be obtained assuming normal-Wishart distribution for the parameters (see [24] for a detailed description). Other quality measures have been introduced in the literature for multinomial and Gaussian networks but, for the sake of simplicity, in this paper we shall restrict to the above introduced measures.

Among the proposed search strategies, the $K2$ algorithm is a simple greedy search algorithm for finding a high quality Bayesian network in a reasonable time [25]. This algorithm takes advantage of the network's decomposability (which allow computing the separate contribution of each variable Y_i to the quality of the network (7)) and the non-unique representation of dependency models using graphs (equivalent graphs can be obtained reversing some of their links). The $K2$ iterative algorithm assumes that the nodes are ordered (to avoid cycles) and starts with a network with no links. 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 quality measure. The process is repeated until either adding new links does not increase the quality or a complete network is attained. A maximum number of parents for each node can also be enforced during the search. For instance, the graph in Fig. 2 was obtained applying the $K2$ learning algorithm to the quantized precipitation data, considering at most one single parent for each node (the database of precipitation records covers the period 1979-1993 for the network of stations shown in Fig. 1(a)). This Bayesian network do not provide an efficient model of the problem, but gives us a benchmark for comparing the performance of more realistic Bayesian networks trained with the data. For instance, Figure 3(a) shows a directed graph obtained increasing the threshold of the algorithm to two parents.
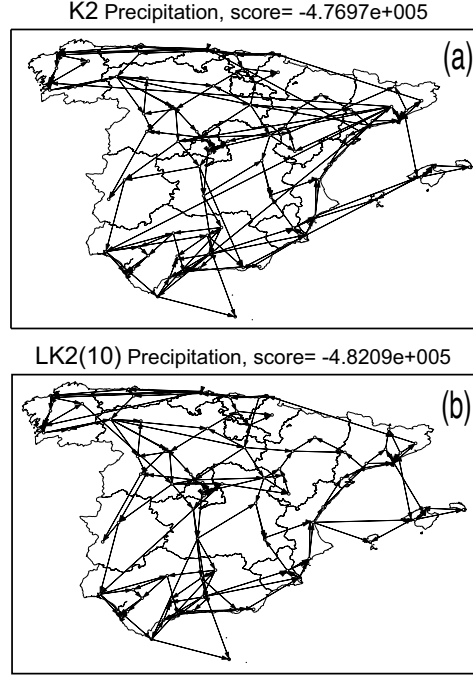
K2 Precipitation, score= -4.7697e+005

LK2(10) Precipitation, score= -4.8209e+005

Fig. 3. Directed acyclic graphs for precipitation obtained applying (a) the $K2$ algorithm and (b) the local $K2$ LK2 with 10 neighbors as candidate parents. In both cases, a maximum of two parents for each node are considered.

The $K2$ learning algorithm is simple and useful. However, further research has to be done in order to develop efficient learning methods for dealing with a large number of variables (stations) and huge amounts of data. Some preliminary works have recently appeared introducing promising ideas such as modularity, or data partitioning (see [26]). The spatial character of the meteorological problem presented in this paper can be taken into account for modifying the $K2$ algorithm to increase the speed and the meteorological significance of the obtained graphs. We call the resulting method the *local K2 learning algorithm* or simply *LK2*.

4.3   The Local K2 Learning Algorithm

In order to increase the efficiency of the $K2$ search strategy we modify the set of candidate parents of node Y_i, $\{1, \ldots, i-1\}$, to include only those nodes with similar climatology to the climatology of Y_i. This is done by computing the correlation of the observed records in different stations and obtaining the k-nearest neighbors for each station (the larger the parameter k, the more similar the algorithm to the original $K2$). This modification reduces the complexity of the search process, since now the set of candidate

parents is of constant size. This fact is illustrated in Fig. 4 which shows the CPU time of the learning process versus the number of nodes for three different learning algorithms (the standard $K2$ and local $K2$ with five and ten neighbors, respectively). The savings in terms of computational time are clearly shown in this figure. Figure 3(b) shows the directed graph obtained using $Lk2$ algorithm with ten neighbors ($LK2(10)$) for precipitation. The score of the local graph is similar to the one obtained using the $K2$ algorithm (Fig. 3(a)). Moreover, from visual inspection, the meteorological significance of the $LK2$ graph is higher.
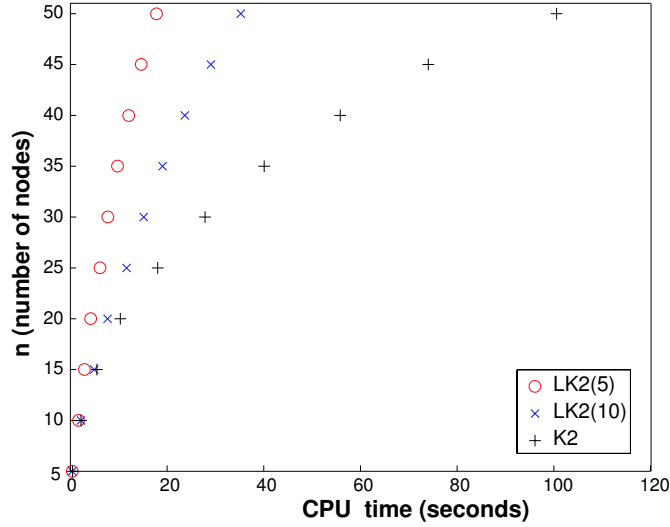

Fig. 4. Problem size (number of nodes) versus CPU time for three different learning algorithms: the standard $K2$ and local $K2$ with five $LK2(5)$ and ten $LK2(10)$ neighbors, respectively (Pentium IV 1.6 GHz).

On the other hand, Figure 5 compares the performance of the standard $K2$ algorithm versus $LK2$ with different number of neighbors $LK2(k)$ when applied to the precipitation data. From this figure we can see the exponential convergence of the score as a function of k, increasing from the threshold given by an empty graph, to the threshold given by the $K2$ algorithm. Thus, substantial saving in computing time can be achieved with a minor loss of score using a small set of neighbors as candidate parents.

4.4   Probabilistic Inference

When some evidence becomes available (e.g., we know that the event "heavy rain" is occurring in Madrid and Coruña $e = \{Madrid = 4, Coruña = 4\}$),
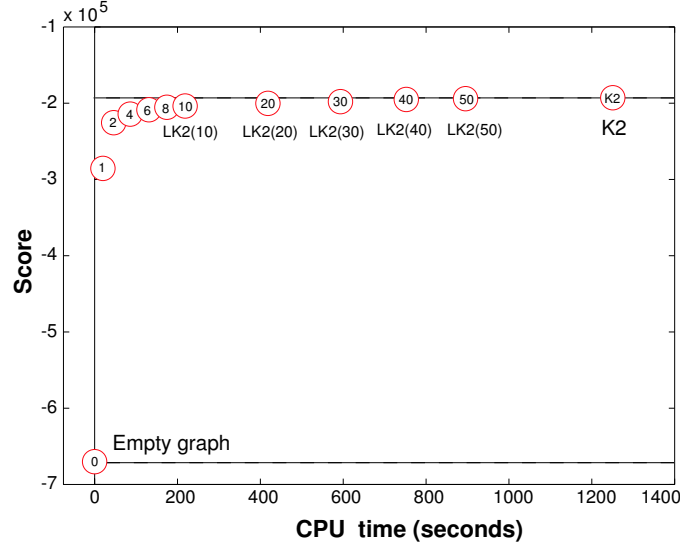
Fig. 5. Comparing the performance of the standard $K2$ algorithm with several local $K2$ algorithms with different neighbors $LK2(k)$ (Pentium IV 1.6 GHz).

Bayesian networks can be used for computing conditional probabilities of the form $P(y_i|e)$. This task can be efficiently done by taking advantage of the independencies encoded in the graph, so probabilistic answers to problems can be obtained in reasonable time.

For instance, Table 1 shows the probabilities of the stations *Santander*, *Zaragoza* and *Soria* (see Fig. 2 for their relative geographical positions) computed by propagating different pieces of evidence using three different graphs: the one obtained using the standard $K2$ algorithm (Fig. 3(a)), the given by $LK2(10)$ with a maximum of one parent per node (Fig. 2) and the one corresponding to $LK2(10)$ with a maximum of two parents per node (Fig. 3(b)). Table 1(a) shows the initial or *a priori* marginal probabilities for each station. From this table we can see that all the graphs provide similar marginal probabilities (all of them have been learnt from the same data). However, this is not the case for the probabilities conditioned to different pieces of evidence. From Tables (b)-(d) we observe a clear difference between the one-parent $LK2(10)$ sparse graph and the other two more dense Bayesian networks with closer scores shown in Fig. 3 (the mean difference between the probabilities provided by these two graphs is less than 3%).

5   Applications of Bayesian Networks

In this section we present some illustrative applications of Bayes networks to some topics related with the problems presented in Sec. 3.

Table 1. (a) Marginal probabilities for three nodes: Santander, Soria and Zaragoza, and (b)-(d) conditional probabilities given different pieces of evidence: $e_1 = \{Madrid = 4\}$, $e_2 = \{Coruña = 4\}$, and $e_3 = \{Madrid = 4, Coruña = 4\}$. Three different Bayesian networks have been considered to propagate the evidence: $K2$ refers to the standard $K2$ algorithm, whereas $LK2_1$ and $LK2_2$ refer to the local $K2$ algorithm with 10 neighboring nodes and a maximum of 1 and 2 parents per node, respectively.

(a) Initial probability, $P(y_k)$

State	Santander			Soria			Zaragoza		
	$K2$	$Lk2_1$	$Lk2_2$	$K2$	$Lk2_1$	$Lk2_2$	$K2$	$Lk2_1$	$Lk2_2$
0	0.502	0.502	0.502	0.613	0.613	0.616	0.741	0.736	0.741
1	0.211	0.211	0.211	0.223	0.215	0.216	0.163	0.158	0.162
2	0.250	0.250	0.250	0.155	0.162	0.158	0.091	0.099	0.092
3	0.037	0.037	0.037	0.009	0.010	0.010	0.005	0.006	0.005

(b) Conditional probability, $P(y_k|Madrid = 4)$

State	Santander			Soria			Zaragoza		
	$K2$	$Lk2_1$	$Lk2_2$	$K2$	$Lk2_1$	$Lk2_2$	$K2$	$Lk2_1$	$Lk2_2$
0	0.357	0.478	0.285	0.156	0.147	0.156	0.331	0.570	0.322
1	0.220	0.211	0.224	0.277	0.296	0.267	0.330	0.235	0.329
2	0.357	0.269	0.409	0.518	0.505	0.523	0.316	0.182	0.324
3	0.065	0.041	0.081	0.049	0.051	0.054	0.023	0.013	0.024

(c)(Conditional probability, $P(y_k|Coruña = 4)$

State	Santander			Soria			Zaragoza		
	$K2$	$Lk2_1$	$Lk2_2$	$K2$	$Lk2_1$	$Lk2_2$	$K2$	$Lk2_1$	$Lk2_2$
0	0.152	0.264	0.150	0.339	0.412	0.291	0.597	0.673	0.563
1	0.183	0.264	0.181	0.331	0.295	0.330	0.237	0.188	0.251
2	0.502	0.425	0.506	0.311	0.275	0.353	0.157	0.130	0.176
3	0.162	0.106	0.163	0.019	0.017	0.026	0.009	0.009	0.010

(d) Conditional probability, $P(y_k|Madrid = 4, Coruña = 4)$

State	Santander			Soria			Zaragoza		
	$K2$	$Lk2_1$	$Lk2_2$	$K2$	$Lk2_1$	$Lk2_2$	$K2$	$Lk2_1$	$Lk2_2$
0	0.127	0.264	0.095	0.077	0.068	0.064	0.285	0.535	0.266
1	0.170	0.205	0.161	0.256	0.279	0.235	0.341	0.250	0.341
2	0.520	0.425	0.540	0.603	0.591	0.621	0.348	0.200	0.364
3	0.183	0.106	0.204	0.064	0.062	0.079	0.026	0.014	0.028

5.1   Spatial Consistency in Stochastic Weather Generators

In the previous section, exact inference techniques have been applied to obtain the conditional probabilities of nodes in a Bayesian network given some evidence. There are also alternative methods for computing the probabilities approximately. The basic idea behind these methods is generating a sample of realizations from the joint PDF of the variables, and computing approximate values for the probabilities from the sample (as frequencies observed in the sample). The simulation is facilitated by the factorization given by the Bayesian network, since each node can be simulated independently (according to the values of the parents), following an appropriate ancestral ordering (see [8] for details). One of the most intuitive and simple simulation methods operates in a forward manner generating the instantiations one variable at a time; a variable is sampled only after all its parents have been sampled [27]. If the obtained realization match the evidence (if any), then it is included in the sample; otherwise it is rejected. The algorithm ends when the required number of acceptable instantiations is obtained.

The above algorithm provide us a simple method for generating stochastic weather from a Bayesian network in an spatially consistent manner. Instead of simulating values independently for each variable (as in standard weather generator methods) we simulate spatial realizations taking into account the constraints imposed by the dependencies in the graph. For the sake of simplicity, we illustrate the application of simulation algorithms in the case of discrete variables (precipitation), though a similar scheme is applicable to the continuous case. Figure 6 shows a sample of rain values (1, 2, 3, or 4) generated for 200 consecutive days for three different stations. We can see that the generated rain values are spatially consistent.
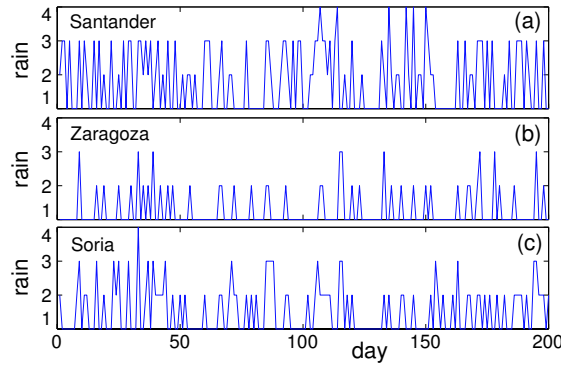

Fig. 6. Simulating a spatial time series of precipitation for 200 days applying the acceptance-rejection method to the marginal distribution (the simulation order is the one used for the $K2$ algorithm): (a) Santander, (b) Zaragoza, (c) Soria.

We can also simulate weather consistent with different evidences (wet or rainy episodes). For instance, Figs. 7 show precipitation simulations obtained when wet and rainy events are assumed to be occurring simultaneously at the staions *Madrid* and *Coruña* (note that this flexibility is not proper of standard weather generators [10]). These figures also exhibit spatial consistency, but now the simulated weather is adapted to the observed evidence.
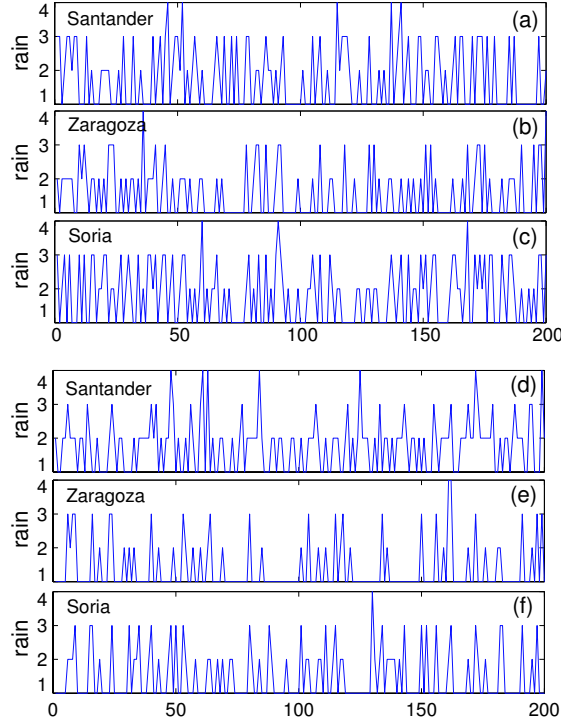

Fig. 7. Simulating a spatial time series of precipitation for 200 days considering the evidence *Madrid=4, Coruña=4* (figures (a)-(c)), and *Madrid=1, Coruña=1* (d)-(f).

5.2   Filling Missing Data.

Another interesting application of Bayesian networks is filling missing data. Missing values are usually present in observation records and some applications require complete data. Missing value estimation is a particular type of forecast problem which is extremely difficult for locations with high spatial variance. Missing data must be recovered preserving the main characteristics of the original data (distribution, mean, extremes, etc.). Standard techniques for this problem use regression models associated with neighboring stations

with complete data archives. A Gaussian Bayesian network provide a global and robust alternative for this problem. An advantage of Bayesian networks is that any piece of information can be plugged in as evidence, obtaining an estimation of the missing values. For instance, Fig. 8 shows the values of maximum wind speed (Wind) estimated from a Gaussian Bayesian network considering as evidence the real values occurred at Madrid and Coruña (left column) and Madrid, Coruña, Ranón and Sondika stations (right column). The Gaussian network has been trained considering the Wind values observed in the stations shown in Fig. 1(a) during the period 1959-2000.
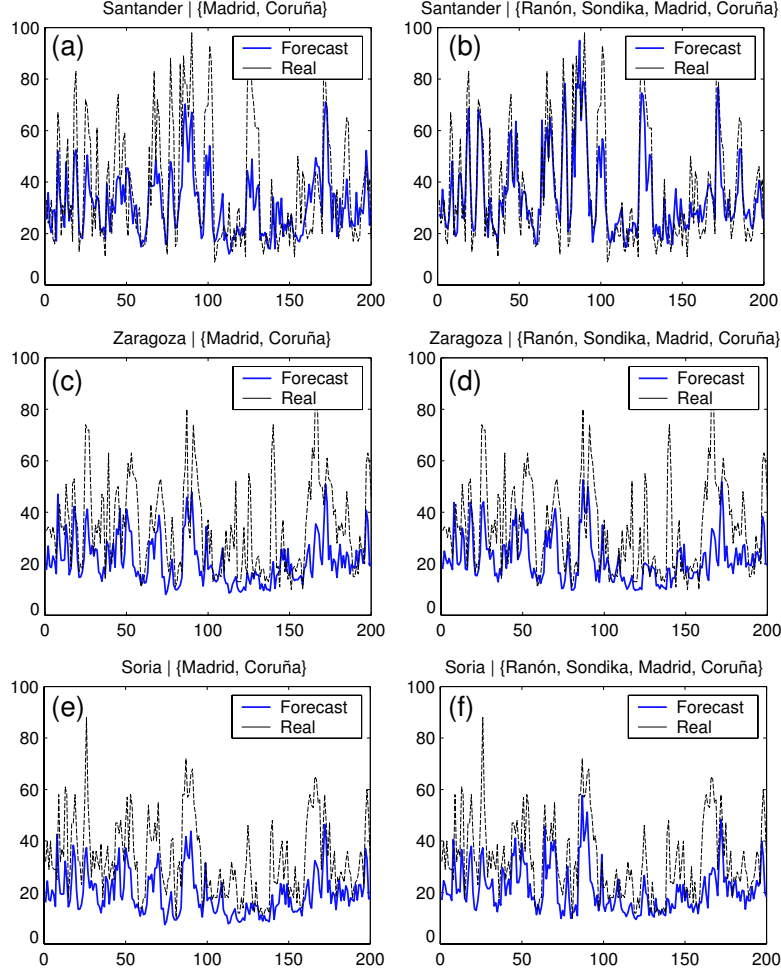
Fig. 8. Conditional mean of the normal distributions of Wind for Santander (a),(b), Zaragoza (c),(d), and Soria (e),(f). Evidences are the values observed at: Madrid and Coruña (left column), Ranón, Sondika, Madrid and Coruña (right column).

   The estimated values are closer to the real ones when the evidence is informative for the corresponding stations (from Figs. 8(a) and (b) we see that Sondika and Ranón provide valuable information for Wind at Santander). This technique can also be used to remove incoherent data from the database (the one leading to probabilities inconsistent with the observations of neighboring stations), and even for short-time automatic forecast (nowcasting) using the real-time observations available at the forecast time [28].

5.3   Local Weather Forecast.

As mentioned in Section 3.2, analog probabilistic forecast methods can be considered a particular type of naive classifiers, where each predictand variable is only conditioned to the atmospheric state S. Given the state of the atmosphere provided by a numerical ACM and considering four states for precipitation (dry, light rain, moderate rain, heavy rain), the naive Bayesian classifier computes the probability for each state at a local station as shown in (2). However, once again, this model does not guarantee the spatial consistency of the results (see [21] for details). This problem can be solved by using the Bayesian network to represent the spatial dependencies among the variables (instead of considering conditional independence of the stations, given S). Figure 9(left column) shows the observed values and the predicted probabilities for two months (January and February 1991) for the event "dry" in Santander, Zaragoza, and Soria using the Bayesian network shown in Fig.3(a).
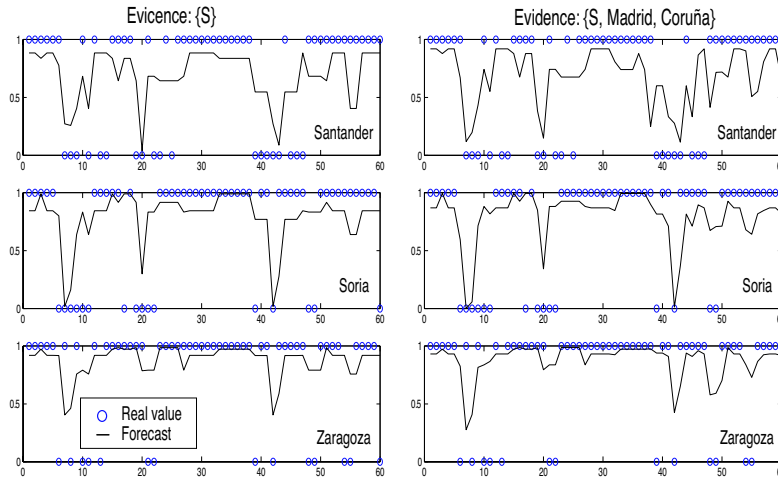

Fig. 9. Observations of the event "dry" in Santander, Zaragoza, and Soria (circles indicate the absence "0" or the occurrence "1"). Forecasts obtained with evidences: S (left column); S and the values observed in Madrid and Coruña (right column).

On the other hand, if we consider a continuous variable (such as Wind), we can condition each of the continuous nodes of the Gaussian network to the discrete atmospheric state, obtaining a conditioned Gaussian network [30]. In this case, the local forecast can be obtained as the conditional mean of the variables, given the state of the atmosphere predicted by a numerical ACM. Figure 10 shows the resulting values when considering 25 (left column) and 50 (right column) states for the atmospheric variable. This figure shows the sensitive dependence of the forecast on the number of states chosen for the atmospheric variable.

The above examples are illustrative and can not be considered competitive models in Meteorology. Further research is still needed to investigate the operative performance of these models.
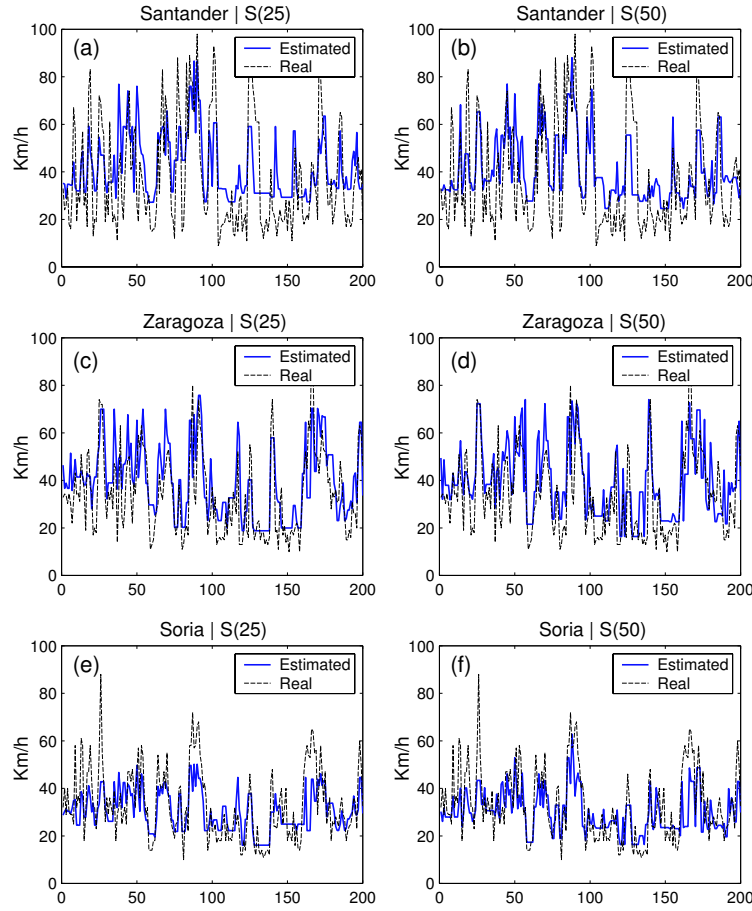

Fig. 10. Mean of the conditional normal distributions of maximum wind speed for Santander (a),(b), Zaragoza (c),(d), and Soria (e),(f).

Acknowledgments

References

1. von Storch H., Navarra A. (1995). Analysis of Climate Variability. Applications of Statistical Techniques, second edition. Springer Verlag
2. von Storch H., Zwiers F.W. (1999). Statistical Analysis in Climate Research. Cambridge University Press
3. Cofiño A.S., Gutiérrez J.M., Jakubiak B., Melonek M. (2003) Implementation of data mining techniques for meteorological applications. In Realizing Teracomputing (W. Zwieflhofer and N. Kreitz editors), World Scientific Publishing, 256–271
4. Wilby R.L., Wigley T.M.L. (1997) Downscaling general circulation model output. A review of methods and limitations. Progress in Physical Geography **21**, 530–548
5. Enke W., Spekat A. (1997) Downscaling climate model outputs into local and regional weather elements by classification and regression. Climate Research **8**, 195–207
6. Kutzbach J. (1967) Empirical eigenvectors of sea-level pressure, surface temperature and precipitation complexes over North America. Journal of Applied Meteorology **6**, 791-802
7. Pearl J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann
8. Castillo E., Gutiérrez J. M., Hadi A. S. (1997). Expert Systems and Probabilistic Network Models. Springer-Verlag, New York (Free Spanish version `http://personales.unican.es/gutierjm`)
9. ECMWF ERA-15 Reanalysis project `http://www.ecmwf.int/research/era/ERA-15/Project/`
10. Wilby R.L., Wilks D.S. (1999) The weather generation game. A review of stochastic weather models. Progress in Physical Geography **23**, 329–357
11. Gabriel K.R., Neumann J. (1962) A Markov chain model for daily rainfall occurrence at Tel Aviv. Quarterly Journal of the Royal Meteorological Society **88**, 90–95
12. Richardson C.W. (1981) Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resources Research **17**, 182–190
13. Wilks D.S. (1999) Multisite downscaling of daily precipitation with a stochastic weather generator. Climate Research **11**, 125–136

14. Zwiers F., von Storch, H. (1990) Regime dependent auto-regressive time series modeling of the Southern Oscillation. Journal of Climate **3**, 1347–1360
15. Pérez-Muñuzuri V., Gelpi I.R. (2000) Application of nonlinear forecasting techniques for meteorological modeling. Annales Geohysicae **18** 1349–1359
16. Gardner M.W., Dorling S.R. (1998) Artificial neural networks (the multilayer perceptron). A review of applications in the atmospheric sciences. Journal of Applied Meteorology **39** 147–159
17. Klein W.H., Glahn, H.R. (1974) Forecasting local weather by means of model output statistics. Bulletin of the American Meteorological Society **55**, 1217–1227
18. Lorenz E.N. (1969) Atmospheric Predictability as Revealed by Naturally Occurring Analogues. Journal of the Atmospheric sciences **26**, 636–646
19. Cano R., López F.J., Cofiño A.S., Gutiérrez J.M., and Rodríguez M.A. (2001) Aplicación de métodos de clasificación al downscaling estadístico. In Actas del V Simposio Nacional de Predicción. Instituto Nacional de Meteorología, 235–240
20. Gutiérrez J.M., Cofiño A.S, Cano R., Rodríguez M.A. (2003) Clustering methods for statistical downscaling in short-range weather forecast. Submitted.
21. Cano R., Cofiño A.S., Gutiérrez J.M., Sordo C. (2002) Probabilistic networks for statistical downscaling and spatialisation of meteorological data Geophysical Research Abstracts **4**, 194
22. Jordan M.I. (1998). Learning in Graphical Models. MIT Press.
23. Geiger D., Heckerman, D. (1995) A characterization of the Dirichlet distribution with application to learning Bayesian networks. Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers, 196–207
24. Geiger D., Heckerman, D. (1994) Learning gaussian networks. In Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers, 235–243
25. Cooper G.F., Herskovitz E. (1992) A Bayesian method for the induction of probabilistic networks from data. Machine Learning **9**, 309–347
26. Neil M., Fenton N., and Nielson L. (2000) Building large-scale Bayesian networks. The Knowledge Engineering Review **15**, 257–284
27. Henrion M. (1988) Propagation of uncertainty by logic sampling in Bayes' networks. In Uncertainty in Artificial Intelligence 2 (Lemmer, J. F. and Kanal, L. N., editors). North Holland, 149–164
28. Michaelides S.C., Pattichis C.S., Kleovoulou G. (2001) Classification of rainfall variability by using artificial neural networks. International Journal of Climatology **21**, 1401–1414
29. Cofiño A.S., Cano R., Sordo C., Gutiérrez J.M. (2002) Bayesian networks for probabilistic weather prediction. In Proceedings of the 15th European Conference on Artificial Intelligence, 695–700
30. Peña J.M., Lozano J.A., Larrañaga P. (2001) Performance evaluation of compromise conditional Gaussian networks for data clustering. International Journal of Approximate Reasoning **28**, 23–50
31. Murphy K.P. (2001) The Bayes net toolbox for Matlab. Computing Science and Statistics **33**. `http://www.cs.berkeley.edu/ murphyk/Bayes/bnt.html`