PAPER • OPEN ACCESS

Precipitation forecast based on the Bayesian Network

To cite this article: S P Khabarov *et al* 2021 *IOP Conf. Ser.: Earth Environ. Sci.* **806** 012016

View the article online for updates and enhancements.

You may also like

- The entropic approach to causal correlations
  Nikolai Miklin, Alastair A Abbott, Cyril Branciard et al.

- Bayesian network models for error detection in radiotherapy plans
  Alan M Kalet, John H Gennari, Eric C Ford et al.

- Network analysis of the food–energy–water nexus in China's Yangtze River Economic Belt from a synergetic perspective
  Rihui An, Pan Liu, Lei Cheng et al.

Precipitation forecast based on the Bayesian Network

S P Khabarov, M L Shilkina and N P Vasiliev

[1]Department of Information Systems and Technologies, *St. Petersburg State Forest Technical University*, 5 Institutskiy Lane, St. Petersburg 194021, Russian Federation

*Corresponding e-mail: mchernobay@inbox.ru

Abstract. The purpose of the article is to consider an alternative approach to the forecasting meteorological events by using expert systems based on the Bayesian network (belief network). Also there is an example of building a Bayesian network capable to predict a type and a probability value of precipitation, with implementation it in the HUGIN software, considered in the article. The conclusion considers the positive and negative aspects of the presented approach.

1. Introduction
Forehanded forecasting of precipitation is extremely important for a comprehensive solution to the problems of restoration and protection of forests. This is primarily due to the growing likelihood of drought in the summer in the southern territories of Russia. This trend is primarily due to the problem of global warming and the greenhouse effect caused by the release of large amounts of methane and carbon dioxide during the thawing of permafrost. Droughts increase the fire hazard and reduce the quality of the state of forest plantations and forest areas in general. Therefore, it is necessary to take serious preventive arrangements based on long-term forecasting of precipitation to improve the quality and safety of forests and forest plantations.

Modern methods of forecasting the conditions of the atmosphere are based mainly on its mathematical models, which allow predicting the weather for the next 3-6 hours with very high accuracy [1]. However, this approach requires the connection of a huge amount of economic and computational resources, since to solve this problem, expensive data collection tools are used, such as meteorological satellites and equipment for weather stations, and powerful means of their processing, which, as a rule, are supercomputers.

There is another approach based on the use of expert systems, which can take as input data not only short-term data such as air temperature, atmospheric pressure, but also long-term factors, such as the climatic zone, season, proximity of seas and oceans, and others. In this case, it becomes possible, on the basis of statistical data or existing experience, to build the dependence of weather characteristics on the above factors and to estimate the degree of influence of these factors. It is advisable to implement this approach using an expert system based on the use of a Bayesian network, since predicting future states of the atmosphere can be burdened with the uncertainty of its current state, which will not be able to take into account expert systems based, for example, on the logic of predicates [2], or representing a neural network [3].

Bayesian network is a graph probabilistic model in which probability is defined as the degree of confidence in the truth of a judgment. This model is applied in those areas of human activity, which are characterized by the presence of uncertainty (lack of information) regarding a specific event or a

set of events, but the circumstances and the probability of these events due to these circumstances are known in advance.

A Bayesian network is an efficient, compact and intuitive knowledge representation for handling uncertainty [4]. It consists of two main parts:

- a graphical structure that defines a set of dependence and independence statements over a set of random variables representing entities of a problem domain;
- a set of conditional probability distributions specifying the strengths of the dependence relations encoded in the graphical structure. A Bayesian network $N=(X, G, P)$ over variables $X=\{X_1, …, X_n\}$, consists of a directed acyclic graph G and a set of conditional probability distributions P. It encodes a decomposition of a joint probability distribution as:

$$P(x) = \prod_{i=1}^n P\big(X_i|pa(X_i)\big), \tag{1}$$

where, $pa(X_i)$ are the parents of X_i in G. It means that the probability of any member of a joint distribution can be calculated from conditional probabilities using the chain rule (given a topological ordering of X) as follows:

$$P(X_1 = x_1,…,X_n = x_n) = \prod_{i-1}^n P(X_i = x_i|X_{i-1} = x_{i-1},…,X_1 = x_1) \tag{2}$$

Since the joint probability density function can be written as a product of the individual density functions, conditional on their parent variables, this can be written as:

$$P(X_1 = x_1, …, X_n = x_n) = \prod_{i-1}^n P\big(X_i = x_i|X_j = x_j \text{for each } X_j \text{ that is } a \text{ parent of } X_i\big) \tag{3}$$

It means the conditional independence of the variables from any of their non-descendants, given the values of their parent variables.

A Bayesian network can serve as a knowledge integration and representation tool for supporting decision making under uncertainty. The information captured by a Bayesian network may originate from a range of different sources. For instance, the conditional probability distributions may be defined based on expert knowledge assessment or subjective estimates, mathematical expressions relating parent configurations to states of the child, or estimated from data. This makes a Bayesian network an excellent knowledge integration framework.

2. Methods and Materials
As a rule, building a Bayesian Network starts with defining the objects of the system and their connections. Weather characteristics depend on many factors such as long-term (proximity of seas and oceans, height above sea level, relief, etc.), and short-term (atmospheric pressure, wind speed and direction, season and others). To make a forecast of precipitation (we predict the type and probability of precipitation), the following objects were selected (the given list of factors is not exhaustive and can be supplemented):

- Atmosphere pressure;
- Proximity of seas and oceans;
- Climatic zone;
- Season;
- Climate;
- Cloudiness;
- Air temperature;
- Type of precipitation.

Before determining the degree of dependences of each of the factors, it is necessary to determine the states which they may take, and which play a role in predicting the types and probability of precipitation.

Suppose that precipitation can fall only in the form of rain, snow, or not at all. Rain is most likely when the air temperature is above five degrees Celsius, and snow is most likely if it is lower. The

probability of precipitation in our model depends on atmospheric pressure and cloud cover. The higher atmospheric pressure causes the higher probability of an anticyclone characterized by clear and calm weather, and the lower probability of precipitation. For this reason, only two states can be distinguished for atmospheric pressure: below or above the norm (760 mm Hg). Speaking about the effect of cloudiness, it should be noted that some types of clouds rather indicate the absence of precipitation [5], however, in our model we assume that the probability of precipitation depends not on the type of clouds, but on the percentage of the sky covered by them: 85% or more - high cloudiness (the probability of precipitation is higher), lower - low (the chance of precipitation is lower).

The climate-forming factors of the created model are the proximity of seas and oceans and the climatic zone. We will choose the states which the climatic zone can take on the basis of the classification of climatic zones by Boris Pavlovich Alisov [6], in which, however, we will combine the equatorial and subequatorial zones into the equatorial, as well as the polar and subpolar into the polar. The reason for the unification is the similarity of climatic conditions. The close location of the sea or ocean is a distance within 500 km. This choice is justified by the difference in climates, for example, between St. Petersburg and Tver, the distance between which is approximately 500 km. Let us distinguish the types of climates on the basis of the classification of climates by B P Alisov [6].

The dependence of air temperature on climate and season is obvious, but the situation is different with cloudiness. In this case, there is an assumption that precipitation falls with different abundance at different seasons. Based on this, we can conclude about the average cloud level for a particular season.

The causal dependencies of the selected factors can be represented in the form of a diagram (figure 1), where the ovals with outgoing arrows are parent nodes, with incoming ones - child nodes, and the states of the nodes are written in the rectangles.
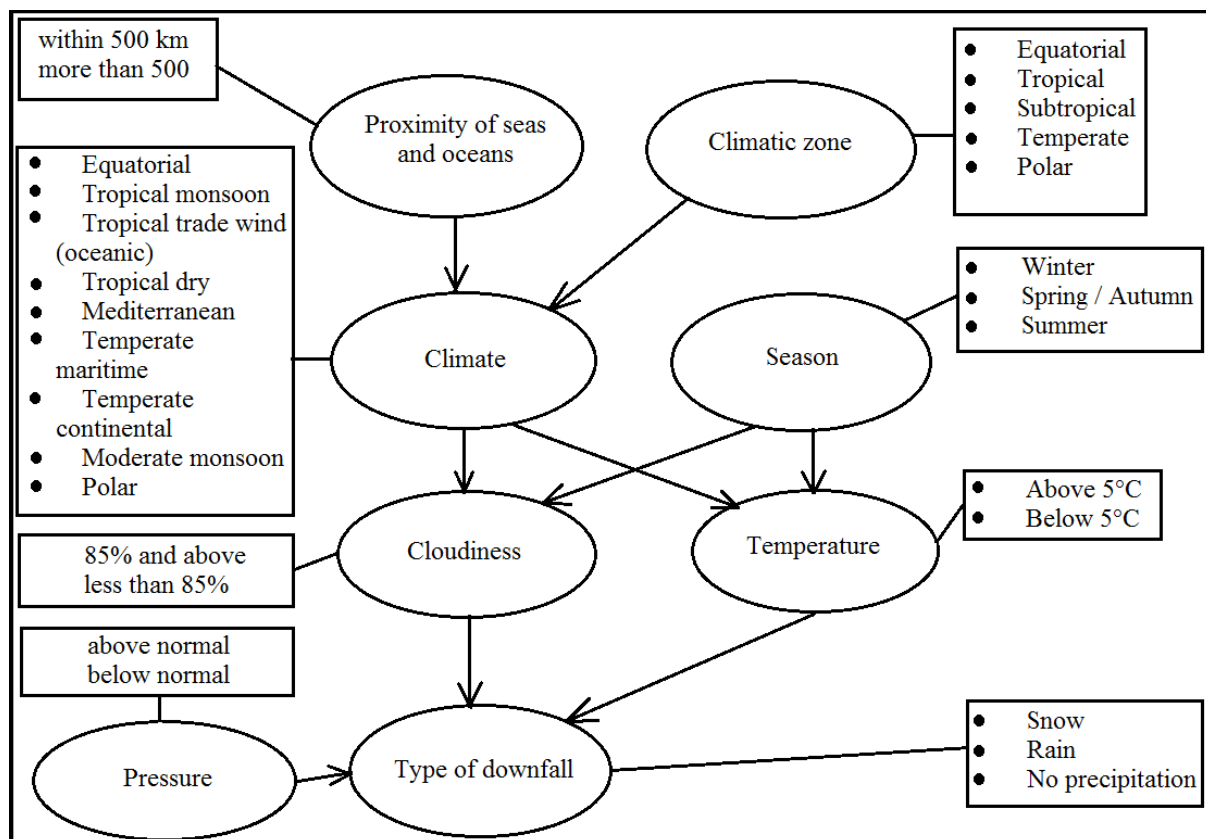
Figure 1. Factors, their states and dependencies.

As a software tool that implements the developed model and allows to automate calculations, we use the HuginLite 8.8 program.

In order for the developed qualitative model to turn into a real Bayesian network, it is necessary to fill in the table of conditional probabilities of the schema nodes. As an example, consider the table of conditional probabilities of the "Precipitation Type" node (table 1). The table shows that the names of the columns are the states of the parent nodes, and the names of the rows are the states of the current one. At the intersections, we set the conditional probabilities of the states of the current node, depending on the state of the parent ones. For nodes that have no parent, the values of a priori (independent) probabilities are filled.

Table 1. Probability values for the graph "Precipitation type".

	Below normal				Above normal			
	Above 85%		Below 85%		Above 85%		Below 85%	
	$\geq 5℃$	$<5℃$	$\geq 5℃$	$<5℃$	$\geq 5℃$	$<5℃$	$\geq 5℃$	$<5℃$
Snow	0.01	0.98	0.1	0.8	0.01	0.8	0	0
Rain	0.98	0.01	0.8	0.1	0.8	0.17	0.05	0.05
No precipitation	0.01	0.01	0.1	0.1	0.19	0.03	0.95	0.95

The considered table of probability values for the graph "Precipitation type" in the HuginLite interface is as shown in figure 2.

Figure 2. The table of probability values for the graph "Precipitation type" in HuginLite.

3. Results and Discussion
Having implemented the developed scheme in HuginLite 8.8 with the completion of the tables of conditional probabilities, let's consider the test the resulting Bayesian network by predicting precipitation in St. Petersburg in the summer.

For this, it is necessary to determine such factors as the proximity of the seas and oceans, climatic zone, season and atmospheric pressure.

- St. Petersburg is located on the coast of the Gulf of Finland, for this reason the state of the node "Proximity of seas and oceans" set the value "within 500 km" with the prior probability of 0.99.
- In accordance with the map of climatic zones B P Alisov, St. Petersburg is located in the temperate zone. For this reason, we assign the "temperate" attribute of the "climatic zone" node a probability value of 1.
- Suppose we don't know anything about the pressure. In this case, we set the probability value of both attributes of the "pressure" node equal to 0.5.

For the "season" node let us make the assumption that spring is already over. In this case, we assign to the "spring/autumn" attribute the probability value of 0.8, and to the "summer" attribute the probability value of 0.2.

After starting the program, the result presented on figure 3 is obtained. The probability of "climate", "cloudiness", "precipitation" and "temperature" nodes are the result of the program calculation. The numbers in front of the state names are the probabilities of these conditions, expressed as a percentage.
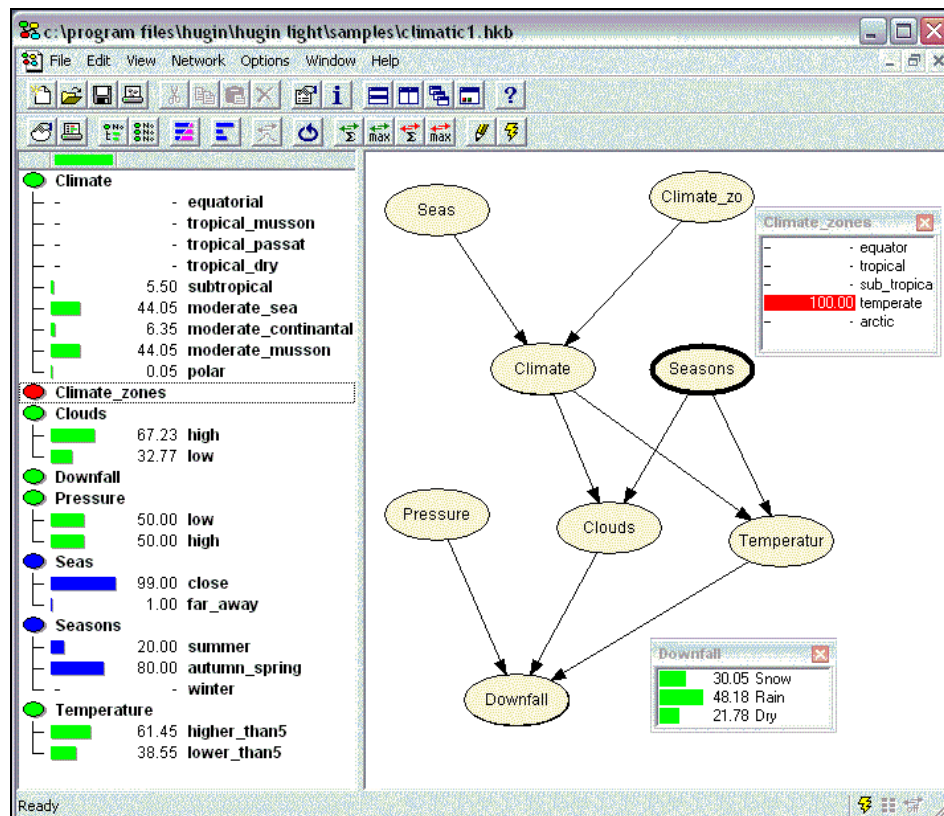
**Figure 3.** Calculating the probability of precipitation.

Based on raw data, the Bayesian model of representation knowledge suggests that rain in the late spring in St. Petersburg should be expected with a probability of about 48%, no precipitation is possible with probability of almost 22%, and the probability of snowfall - about 30%.

A feature of expert systems based on the Bayesian network is the ability to evaluate the graph in the inverse direction. In the context developed Bayesian network, this means that by setting the probability conditions for factors such as precipitation type, air temperature, season and proximity to the sea, the program is able to assume a climatic zone for which the specified conditions will be characteristic. Let's demonstrate this by setting the probabilities equal to 1 for:

- The "type of precipitation" node is set to "rain" value with probability 0.85 and to "without precipitation" value with probability 0.15;
- The "temperature" node is set to "5 ℃ and above" value with probability 1;
- The "season" node is set to "summer" value with probability 1;
- The "pressure" node is set to "above the norm" value with probability 1;
- The "the proximity of the seas and oceans "node is set to "within 500 km" value with probability 0.75;
- The "cloudiness" node is set to "low cloudiness" value with probability 1.

Based on the given probabilities, the program determines equatorial climatic zone with a probability of 20%, tropical -24%, subtropical - 36%, moderate - 17% and arctic - 3%, these results are presented on figure 4.
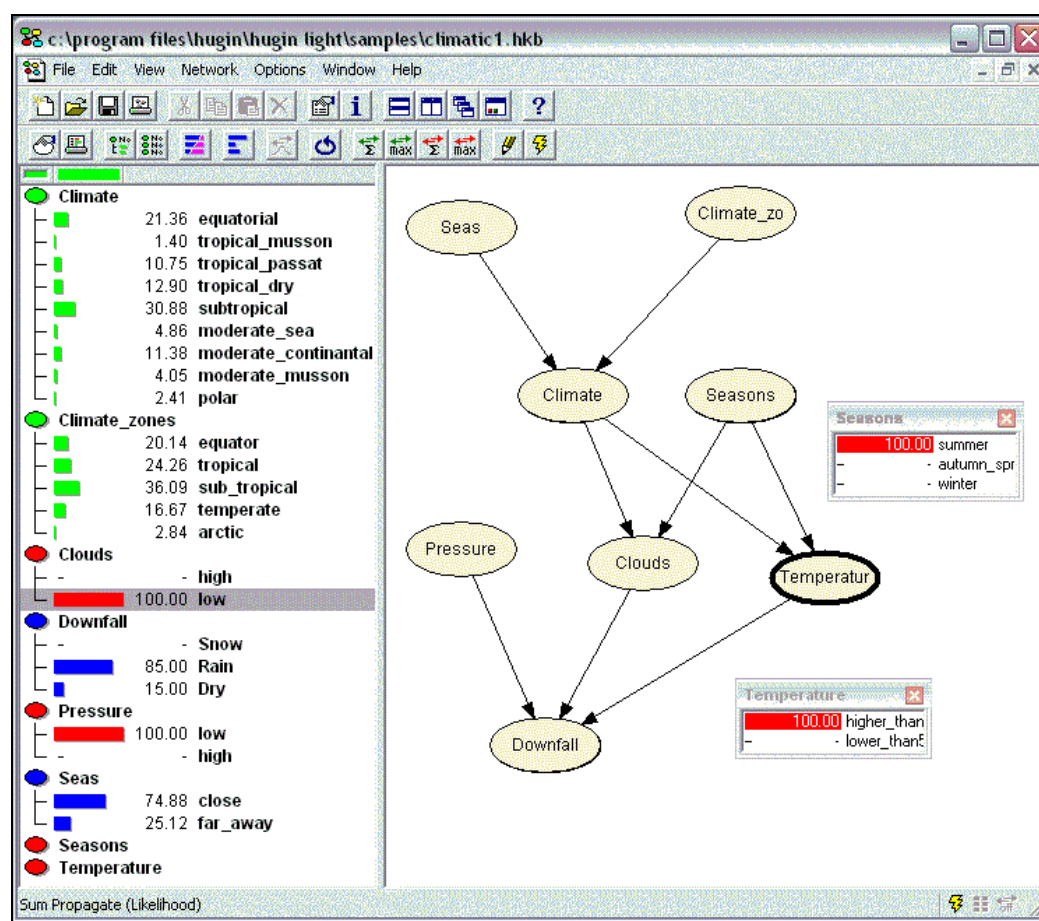
Figure 4. Identification of the climatic zone.

4. Summary

The presented approach to forecasting precipitation is capable of significantly reduce spending on expensive equipment. It can also be applied when the computing resources are insufficient for making a forecast using mathematical models of the atmosphere. The disadvantage of the described approach is the lack of accuracy, due to which it cannot be applied in cases where a high accuracy is required (for example, in aviation). However, the accuracy of forecasting precipitation based on Bayesian network can be increased by adding new factors to the model, increasing the number of states and adjustments of conditional probability tables. Also the approach can be used as a complement numerical forecasting methods.

References

[1]   Verbitskaya E M and Romansky S O 2015 Application of High-Performance Systems for Numerical Weather Prediction on the Russian Far East Territory *Bulletin PNU* **3(38)** 131
[2]   Khabarov S P and Pushkareva L G 2019 *Representation of knowledge in information systems. Using the PIE framework for Database and Knowledge Base design* (St. Petersburg: SPbSFTU) p 66 *available at:* https://e.lanbook.com/book/117636

[3]   Zayats A M and Khabarov S P 2019 Construction of neural network classification of
        Fisheririses based on JavaScript *News of the Saint Petersburg State Forest Technical
        Academy* **226** 233

[4]   Kjærulff U B and Madsen A L 2013 *Bayesian Networks and Influence Diagrams – A Guide to
        Construction and Analysis* (New York: Springer Publishing Company) p 319

[5]   Bespalov D P,   Devyatkin A M,   Dovgalyuk Yu A,   Kondratyuk V I,   Kuleshov Yu V,
        Svetlova T P, Suvorov S S and Timofeev V I 2011 *Atlas of clouds / Federal Service for
        Hydrometeorology and Environmental Monitoring (Roshydromet), Voeikov A I Geophysical
        Observatory* (St. Petersburg: D'ART) p 248

[6]   Khromov S P 2006 *Meteorology and climatology* (Moscow: Moscow State University Press
        Nauka) p 582