DYNAMIC BAYESIAN NETWORK FOR WEATHER FORECAST AND EVALUATION OF RENEWABLE RESOURCES AVAILABILITY

CIOBANU A.*, MUNTEANU F.*, NEMEŞ C.*,ADOCHIŢEI M.
*"Gheorghe Asachi" Technical University of Iaşi, B-dul Dimitrie Mangeron, nr. 67
700050, Iaşi, Romania
ciobanu_v_alexandra@yahoo.ro

Abstract - The authors present a Bayesian network capable to estimate the weather parameters related, not only, to renewable resources: wind speed and solar irradiation. A large and systematic data base about simple and composed weather indices registered during four years, 2013-2016, was used to construct the data-driven Bayesian structure and to learn and validate its parameters. It includes 9 weather indices collected, minute by minute, by a professional Davis Instrument Pro 2 Plus weather station. The extremely large initial data base, over 1.8 million records, was discretized in 4 classes making possible to use a very simple algorithm like Bayesian search to establish the most suitable network structure fitting the data. The main and first useful results mean the probability of wind speed and solar irradiance classes. Both parameters can be transformed in electrical power considering a given wind generator and a solar panel.

Keywords: dynamic Bayesian networks, Bayesian search algorithm, weather forecast, renewable resources.

List of abbreviations

BN – Bayesian network;
DBN – dynamic Bayesian network;
BSA – Bayesian search algorithm;
Mbs – meteorological Bayesian network structure
Md – meteorological database;
M – the set of random variables included in Md;
P(Mbs/Md) – posterior probability (probability of *Mbs* given *Md*);
m – number of cases in *Md;*
C_v – *v*th case in *Md;*
Mbs_P – vector whose values denote de conditional probability assignments associated with belief network structure *Mbs;*
$f(Mbs_P/Mbs)$ – probability density function over Mbs_P given *Mbs.*

1. INTRODUCTION

The interest in dynamic Bayesian networks has grown in the last years due to the advantage of representing temporal dependencies without the need to create new variables. DBN are frequently used for forecasting natural phenomena or weather parameters.

One such system is described in paper [1] where several forecasting elements are combined with the use of BN's and the knowledge of meteorologists to predict severe weather.

Another use of DBN's is presented in paper [2], and it is related to the ability of predicting production volumes from renewable energy sources, in particular wind farms. Authors describe a technique based on Bayesian regularization for reducing model overfitting problems that may arise in forecasting of wind power generation, and their results showed that BN's display equivalent predictive performance to Neural Networks trained by Maximum Likelihood.

In paper [3], the authors used BN's to model the spatial dependencies among two meteorological variables (rainfall and temperature) for weather prediction over a specific location. Using inference algorithms it has been analyzed weather prediction by doing experiments over independent test data sets.

The authors in [4], propose a method to estimate reference evapotranspiration (ET) from limited climate data by using a Bayesian model to determine the uncertainty of different models that explain ET.

Paper [5] presents a probabilistic approach based on fuzzy Bayesian networks (FBN) to forecast the weather condition by predicting the spatio-temporal interrelationships among different climate variables.

One of the most challenging problems in weather forecasting is discussed in [6]. For rainfall forecasting, the authors have used a BN model for representing rainfall data from 21 weather stations. Using the greedy search algorithm they have been able to represent dependencies between different stations.

Other papers have been put forward by researchers employing various methods for weather forecast that include probabilistic models (Bayesian networks) [7] or neural network based techniques [8].

Besides forecasting weather parameters, DBN are used in several different domains like modeling electrical energy markets [9], forecasting short-term passenger flow in transport services [10], handwritten word recognition [11], fault detection in autonomous spacecraft [12], or in electronic equipment health diagnosis [13].

2. DYNAMIC BAYESIAN NETWORKS

Detecting temporal patterns of time-series meteorological data is a difficult task due to really unknown dependence of parameters. They are not only pure random variables but their dependencies are random also. That's why the weather forecast keeps always the "probable" attribute even for short-time intervals.

As it was mentioned above, DBN can be used to extend the ability of BN for belief calculations related to dynamically changing processes.

In principle, there are two techniques for a time depending process:

- Time slice approach consisting in capturing the evolving process states in time steps, as depicted in Fig. 1;
- Decomposing the BN in identical models or sub-models multiplied over each time step as detailed in Fig. 2.

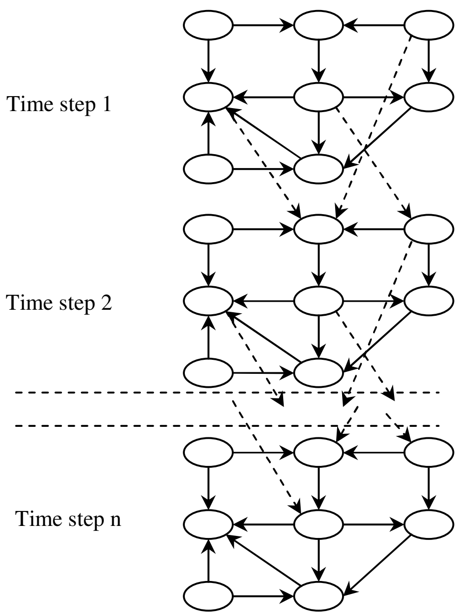

Fig. 1. Dynamic Bayesian network modeling a time steps process evolution

In this paper the authors used an adapted version of DBN compared to what it was presented in [14] and implemented in Genie software package [15], according to Fig. 2.

2.1. Weather database

The information about the weather, targeting the wind speed and solar irradiation as renewable resources, was registered minute by minute during 4 years, between 2013 and 2016.

A professional Davis Instrument weather station type Vantage Pro2 Plus was used to collect the information about 9 weather parameters, simple or composed, are presented in table 1.

The station recorded over 1.8 million data rows with some inevitable errors or missing values.

There was necessary to clean the data, replace with zero, average values or near similar values using different robust techniques to avoid wrong records on data chains while keeping an acceptable accuracy of final results [16]. A discretization procedure was necessary with a view to find the Bayes network structure.

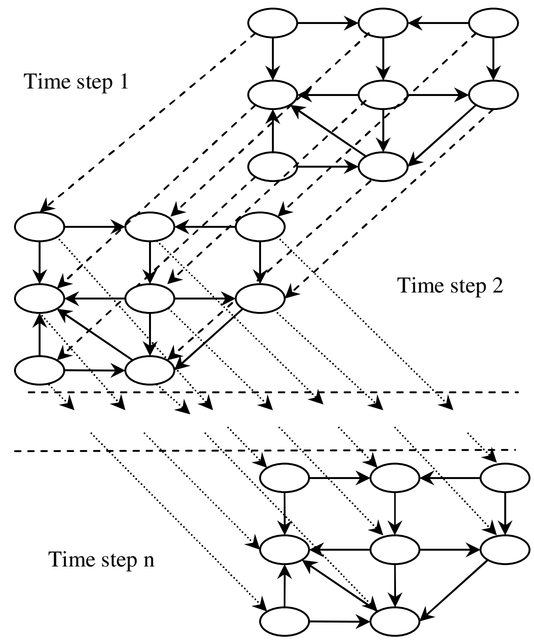

Fig. 2. Multiplied temporal Bayesian network

Table 1. Indices recorded for weather forecast

Weather parameter and acronym in DBN (random variable)	**Accuracy**	**Range**	**Units**
Evapotranspiration – **ET (ET)**	5%	0 – 19999.9	mm
Barometric pressure – **Bar (B)**	0.8 mm Hg 1.0 mb	410 - 820 mm Hg 540 – 1100 mb	mm Hg mb
Outside humidity – **Out humid (OU)**	3%	0% - 100%	%
Solar irradiance – **Solar rad (SR)**	5%	0 - 1800	W/m^2
Outside temperature – **Temp out (TO)**	0.5	-40 - +65	^{0}C
Wind speed – **Wind speed (WS)**	5%	1 – 67	m/s
Heat index – **Heat index (HI)**	1.5	-40 - +74	^{0}C
THW index – **THW index (THW)**	2	-68 - +64	^{0}C
THSW index – **THSW index (THSW)**	2	-68 - +64	^{0}C

2.2. Bayesian searching algorithm

There are many learning methods to find a suitable Bayesian network structure fitting a set of database cases. One of the simplest is Bayesian Search Algorithm (BSA) clearly presented in [17]. The essence of this algorithm is based on finding the most probable Bayes network structure given a database and calculating $P(Mbs|Md)$.

Let consider the following assumptions and notations:

- BSA compares two or more Bayesian network structures (Mbs_i, Mbs_j, Mbs_k, …) generated by the same data base Md and calculates the conditional probability ratio according to fundamental rule of probability:

$$\frac{P(Mbs_i|Md)}{P(Mbs_j|Md)} = \\ = \frac{P(Mbs_i, Md)/P(Md)}{P(Mbs_j, Md)/P(Md)} = \frac{P(Mbs_i, Md)}{P(Mbs_j, Md)} \quad (1)$$

- The unique database M includes the meteorological variables:

$$M \equiv \{ET, B, OU, SR, TO, WS, H, THSW, THW\} \quad (2)$$

which are the conditional dependent variables of *Mbs*.

- A database row of variables means a case.
- The cases are assumed to be conditional independent and this is expressed by the equation
-

$$P(Mbs, Md) = \int_{Mbs_P} \left[\prod_{v=1}^{m} P(C_v|Mbs, Mbs_P) \right] \times \\ \times f(Mbs_P|Mbs) P(Mbs) dMbs_P \quad (3)$$

- The continuous recorded variables were discretized in 4 classes without missing values.

Based on the above mentioned assumptions, as it was demonstrated in [17],

$$P(Mbs, Md) = P(Md) \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}! \quad (4)$$

where

- n is the number of discrete x_i variables;
- every variable has r_i values (v_{i1}, v_{i2}, …., v_{iri});
- q_i is the unique instantaneous value of the set of parents π_i of the variable x_i in *Md*;
- N_{ijk} is the number of cases in *Md* in which the variable x_i has the value v_{ik} while the set of variables π_i is instantiated as w_{ij}; consequently,

$$N_{ij} = \sum_{k=1}^{r_j} N_{ijk} \quad (5)$$

The BSA find the optimal structure reducing the number of possible structures considering equal priors *P(Mbs) = C* on *Mbs* and maximizing *P(Mbs|Md)*. Eq. 4 becomes

$$P(M_{bs}, M_d) = C \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}! \quad (6)$$

Maximum of $P(M_{bs}|M_d)$ means maximum of the second product:

$$\max_{M_{bs}} P(M_{bs}, M_d) = \\ = C \prod_{i=1}^{n} \max_{\pi_i} \left[\prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}! \right] \quad (7)$$

A more general expression of eq. (7) is fully presented in [17].

3. DATA-DRIVEN BAYESIAN NETWORK STRUCTURE

Fig. 3 shows a sample of 10 data cases of the 9 non-discretized meteorological variables of *Md*. To apply the BSA it was necessary to use discretized variable values.

The authors' tests proved the BN structure based is depending on year or even on the month of the year which means a real difficulty in establishing a final one. The solution was to ask climate experts opinion and to validate finally the structure according with data and minimize the errors.

Anyway, establish a suitable structure for weather forecast remains an extreme complex task associated with errors.

1	Date	Time	Temp out	Out humid	Wind speed	Heat index	THW index	THSW index	Bar	Solar rad	ET
2	1/1/2016	00:01	-6.80	69.00	0.90	-6.90	-8.20	-10.70	771.70	0.00	0.00
3	1/1/2016	00:02	-6.80	69.00	1.30	-7.00	-9.40	-11.90	771.80	0.00	0.00
4	1/1/2016	00:03	0.00	69.00	0.90	0.00	0.00	0.00	771.70	0.00	0.00
5	1/1/2016	00:04	-6.80	69.00	0.40	-7.00	-7.00	-9.50	771.70	0.00	0.00
6	1/1/2016	00:05	-6.80	69.00	0.40	-7.00	-7.00	-9.50	771.70	0.00	0.00
7	1/1/2016	00:06	-6.80	69.00	0.40	-7.00	-7.00	-9.50	771.70	0.00	0.00
8	1/1/2016	00:07	-6.80	71.00	0.90	-7.00	-8.20	-10.70	771.70	0.00	0.00
9	1/1/2016	00:08	-6.80	71.00	1.30	-7.00	-9.40	-11.90	771.70	0.00	0.00
10	1/1/2016	00:09	-6.90	71.00	1.30	-7.10	-9.40	-11.90	771.70	0.00	0.00

a)

Temp out	Out humid	Wind speed	Heat index	THW index	THSW index	Bar	Solar rad	ET
s02_n10_n2	s07_68_76	s02_0_1	s03_n7_0	s02_n10_n5	s02_n13_n7	s08_769_773	s01_below_102	s01_below_
s02_n10_n2	s07_68_76	s02_0_1	s03_n7_0	s02_n10_n5	s02_n13_n7	s08_769_773	s01_below_102	s01_below_
s03_n2_5	s07_68_76	s02_0_1	s03_n7_0	s03_n5_2	s04_n1_3	s08_769_773	s01_below_102	s01_below_
s02_n10_n2	s07_68_76	s01_below_0	s03_n7_0	s02_n10_n5	s02_n13_n7	s08_769_773	s01_below_102	s01_below_
s02_n10_n2	s07_68_76	s01_below_0	s03_n7_0	s02_n10_n5	s02_n13_n7	s08_769_773	s01_below_102	s01_below_
s02_n10_n2	s07_68_76	s01_below_0	s03_n7_0	s02_n10_n5	s02_n13_n7	s08_769_773	s01_below_102	s01_below_
s02_n10_n2	s07_68_76	s02_0_1	s03_n7_0	s02_n10_n5	s02_n13_n7	s08_769_773	s01_below_102	s01_below_
s02_n10_n2	s07_68_76	s02_0_1	s03_n7_0	s02_n10_n5	s02_n13_n7	s08_769_773	s01_below_102	s01_below_
s02_n10_n2	s07_68_76	s02_0_1	s02_n11_n7	s02_n10_n5	s02_n13_n7	s08_769_773	s01_below_102	s01_below_
s02_n10_n2	s07_68_76	s02_0_1	s02_n11_n7	s02_n10_n5	s02_n13_n7	s08_769_773	s01_below_102	s01_below_

b)

Fig. 3 A sample of the continuously (a) and discretized (b) recorded data-cases of meteorological variables of *Md*

The BN structure based on 4 years full database is shown in Fig. 4.

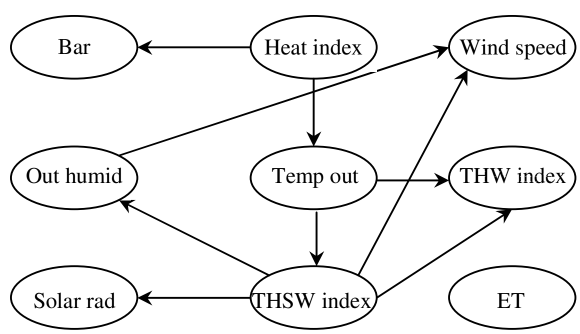

Fig. 4 The BN structure generated using BSA

BSA showed the conditional independence of ET from the rest of variables while Heat index is a marginal one.

The conditional probabilities of variables classes are presented in Fig. 5 for the initial time step.

An important aspect in constructing a data-driven BN is the error checking involving the opposite procedure: fitting the resulting BN structure to the data. There are many techniques for this.

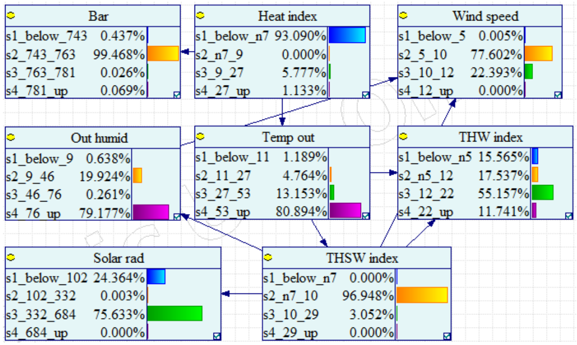

Fig. 5 Conditional probabilities of meteorological variables classes for the first time step

A simple one is to randomly check the results offered by BN as a result of considering the evidence of n-1 variables and see if the n^{th} (THSW in Fig. 6) has the same value as in the data base accordingly.

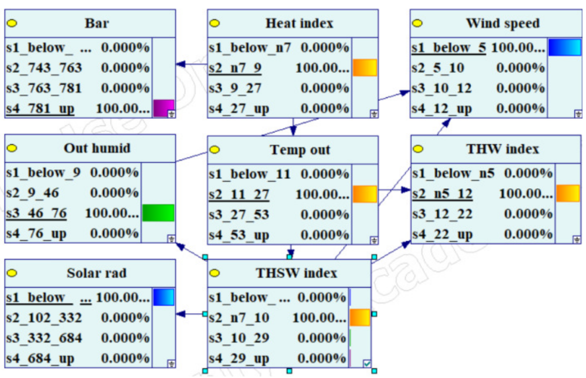

Fig. 6 Randomly setting the evidence for n-1 variables and checking if the n^{th} (e.g. THSW) has the same value like in data base

A more elegant method is to use specific methods for error checking based on:

- general accuracy which is a numerical (%) result showing how many records are correctly estimated by the network from the total; for the BN in Fig. 4, maximum accuracy is 0.8553 for wind speed between 10-12 [m/s] (class s3) and 0.5416 for solar irradiance higher than 684 [W/m^2] (class s4);
- confusion matrix [19] shows the same result in terms of the number of records correctly and incorrectly classified;
- Receiver Operating Characteristic - ROC for each of the states of each of the class variables [19]. Fig. 7 shows the ROC for the class s3 of wind speed and associated value for AUC – Area Under ROC while Fig. 8 includes the same information but for class s4 of solar irradiance;
- calibration characteristic could an important measure of performance of a probabilistic model [19].

3.1. Dynamic Bayesian network analysis

A time step analysis for the BN is a complex one from computing time and computer resources even in the case of a relative simple BN like in Fig. 4.

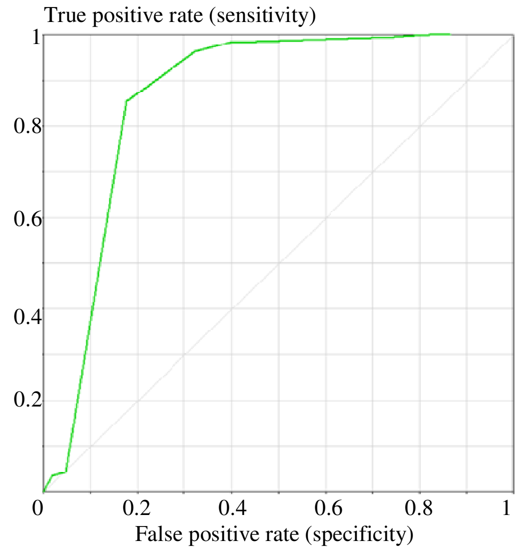

Fig. 7 ROC for class s3 [10-12 m/s] of wind speed and AUC=0.7298

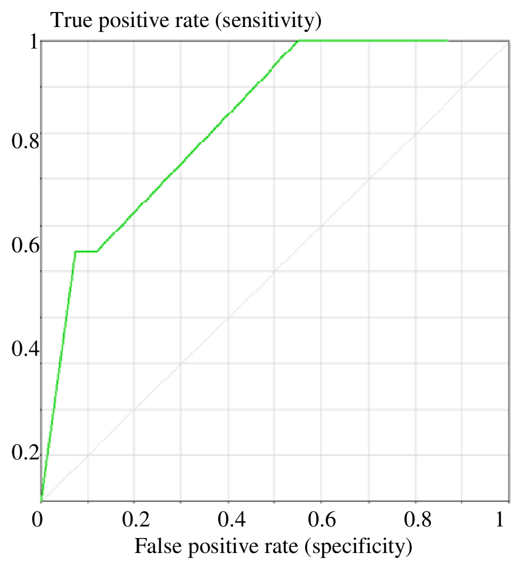

Fig. 8 ROC for class s4 [up 684 W/m^2] of solar irradiance and AUC=0.6974

The DBN was analyzed for a few months due to computer limitations, especially the memory [18]. The authors used an Intel® Core™ i3-2120 CPU @ 3.30 GHz, 64-bit operating system, x64-based processor with 4.00 GB installed RAM.

The BN given in Fig. 4 was generated using the full database records. The corresponding unrolled BN is depicted in Fig. 9 and it was obtained following the technique indicated in Fig. 2 for 5 monthly time steps.

Solar irradiance and wind speed were the target variables for which the corresponding dynamic class probabilities are showed in Fig. 10.

These conditional time-dependent probabilities of renewable resources allow computing the corresponding power generated by given renewable sources.

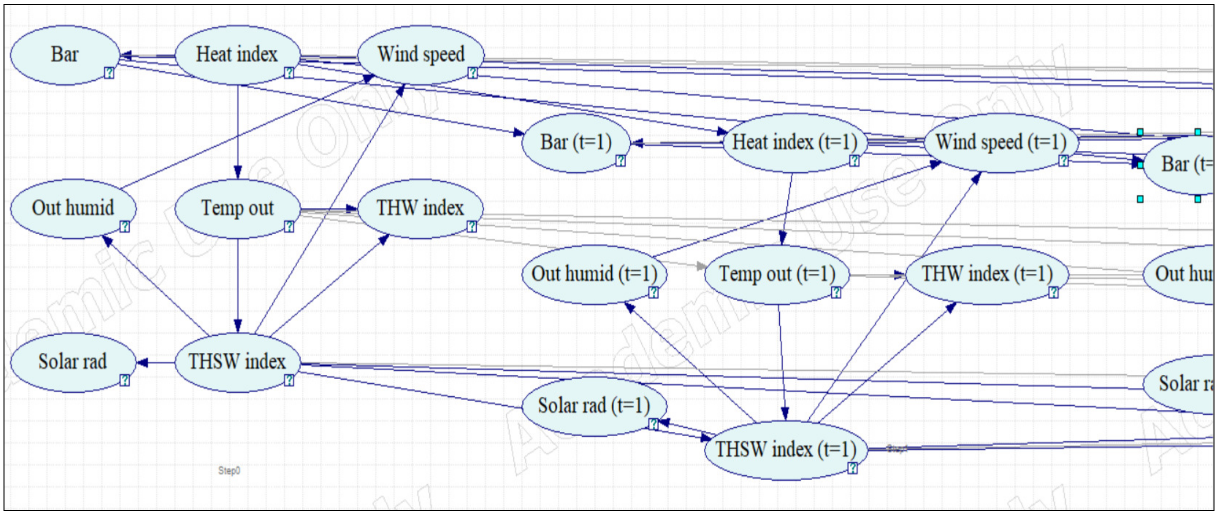

Fig. 9 The unrolled BN (first time step)

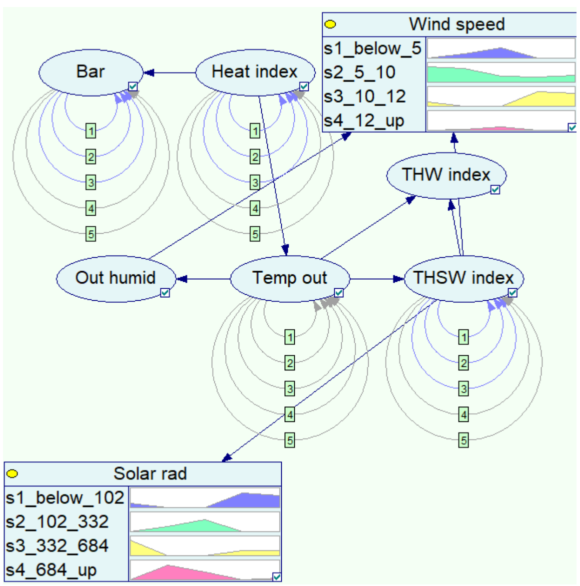

Fig. 10 The dynamic conditional probabilities for wind speed and solar irradiance generated by DBN

Details about the probabilities dynamic are given in Fig. 11 and Fig. 12 for wind speed and solar irradiance respectively.

The power generated by wind source is given by:

$$P_{wind} = \frac{1}{2} \rho \cdot c \cdot A \cdot v^3 \ [W] \tag{8}$$

where c is the generator efficiency, ρ is density air [kg/m^3], A is the generator area perpendicular to the wind [m^2] and v is the wind speed [m/s].

For a solar panel, the power generated is given by:

$$P_{solar} = S \cdot r \cdot H \cdot P_r \ [W_p] \tag{9}$$

where S is total solar panel area [m^2], r is solar panel yield or efficiency [%], H is the solar irradiance [W/m^2], P_r is performance ratio (range between 0.5 and 0.90.

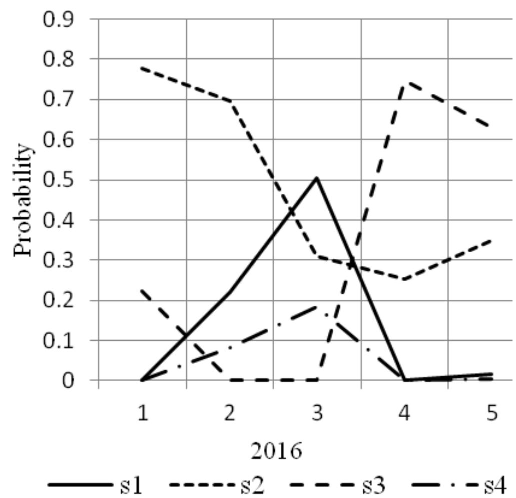

Fig.11 Conditional probability for wind speed classes during the first five months in 2016

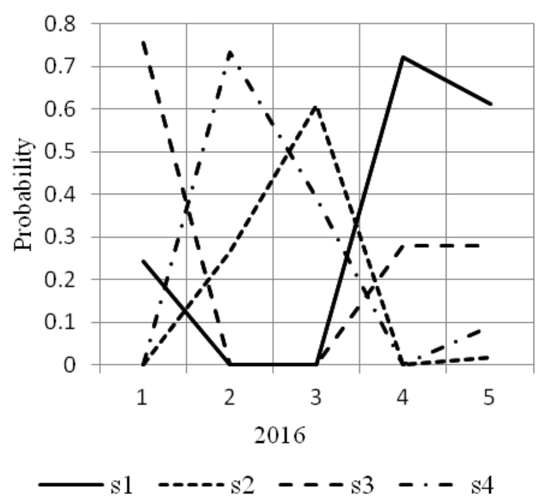

Fig.12 Conditional probability for solar irradiance classes during the first five months in 2016

Using Eq. 8 and Eq. 9 together with probability distributions of renewable resources class variables we can get the corresponding distributions of power (and the corresponding energy) generated by renewable sources.

4. CONCLUSIONS

DBN are a suitable instrument for weather forecast and the associated renewable resources availability. The time distribution probabilities for wind speed and solar irradiance allow calculating the expected power/energy generated.

Future work has to be dedicated to a more precise date-driven BN structure and DBN' better adjusted time intervals network.

More sophisticated algorithms like PC, NPC, Essential Search Graph, Greedy Thick Thinning, etc., fitting the database are available.

A computer with high resources, from which at least 8-16 GB RAM is necessary for a yearly, or more, extended DBN analysis.

The aim is to find a good BN structure for weather forecast (which is not a simple task) and estimate the available energy from renewable sources adapted to costumers needs.

ACKNOWLEDGMENT

This work was supported by Romanian Ministry of Research and Innovation, CCCDI - UEFISCDI, project number PN-III-P2-2.1-CI-2017-0823, in the frame of PNCDI III.

REFERENCES

[1]. Abramson, B., Brown J., Edwards W., Murphy A., Winkler R.L. 1996. Hailfinder: A Bayesian System for Forecasting Severe Weather. International Journal of Forecasting 12, 57-71.

[2]. R. Mbuvha, M. Jonsson, N. Ehn and P. Herman, "Bayesian neural networks for one-hour ahead wind power forecasting," 2017 IEEE 6th International Conference on Renewable Energy Research and Applications (ICRERA), San Diego, CA, 2017, pp. 591-596. doi: 10.1109/ICRERA.2017.8191129

[3]. A. Nandar, "Bayesian network probability model for weather prediction," 2009 International Conference on the Current Trends in Information Technology (CTIT), Dubai, 2009, pp. 1-5. doi: 10.1109/CTIT.2009.5423132

[4]. S. Hern´ndez, L. Morales and P. Sallis, "Estimation of Reference Evapotranspiration Using Limited Climatic Data and Bayesian Model Averaging," 2011 UKSim 5th European Symposium on Computer Modeling and Simulation, Madrid, 2011, pp. 59-63. doi: 10.1109/EMS.2011.81

[5]. M. Das and S. K. Ghosh, "A probabilistic approach for weather forecast using spatio-temporal inter-relationships among climate variables," 2014 9th International Conference on Industrial and Information Systems (ICIIS), Gwalior, 2014, pp. 1-6. doi: 10.1109/ICIINFS.2014.7036528

[6]. A. Sharma and M. K. Goyal, "Bayesian network model for monthly rainfall forecast," 2015 IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kolkata, 2015, pp. 241-246. doi: 10.1109/ICRCICN.2015.7434243

[7]. C. Rodriguez Rivero, J. Pucheta, M. Herrera, V. Sauchelli and S. Laboret, "Time Series Forecasting Using Bayesian Method: Application to Cumulative Rainfall," in IEEE Latin America Transactions, vol. 11, no. 1, pp. 359-364, Feb. 2013. doi: 10.1109/TLA.2013.6502830

[8]. S. Cai, W. W. Hsieh and A. J. Cannon, "A comparison of bayesian and conditional density models in probabilistic ozone forecasting," 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, 2008, pp. 2310-2314. doi: 10.1109/IJCNN.2008.4634117

[9]. K. Dehghanpour, H. Nehrir, J. Sheppard and N. Kelly, "Agent-based modeling in electrical energy markets using dynamic Bayesian networks," 2017 IEEE Power & Energy Society General Meeting, Chicago, IL, USA, 2017, pp. 1-1. doi: 10.1109/PESGM.2017.8274243

[10]. J. Roos, S. Bonnevay and G. Gavin, "Dynamic Bayesian networks with Gaussian mixture models for short-term passenger flow forecasting," 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Nanjing, 2017, pp. 1-8. doi: 10.1109/ISKE.2017.8258756

[11]. N. Ghanmi, A. M. Awal and N. Kooli, "Dynamic Bayesian networks for handwritten Arabic word recognition," 2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR), Nancy, 2017, pp. 104-108. doi: 10.1109/ASAR.2017.8067769

[12]. D. Codetta-Raiteri and L. Portinale, "Dynamic Bayesian Networks for Fault Detection, Identification, and Recovery in Autonomous Spacecraft," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 45, no. 1, pp. 13-24, Jan. 2015. doi: 10.1109/TSMC.2014.2323212

[13]. H. Xie, J. Shi, W. Lu and W. Cui, "Dynamic Bayesian networks in electronic equipment health diagnosis," 2016 Prognostics and System Health Management Conference (PHM-Chengdu), Chengdu, 2016, pp. 1-6. doi: 10.1109/PHM.2016.7819945.

[14]. J. Hulst, "Modelling physiological processes with dynamic Bayesian networks". MSc thesis. TU Delft, 2006.

[15]. www.bayesfusion.com

[16]. Rousseeuw, P. J., & Hubert, M., "Robust statistics for outlier detection". Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(1), 73-79, 2011.

[17]. Cooper, G. F., Herskovits E., "A Bayesian method for the induction of probabilistic networks from data". Machine Learning, no. 9, pp. 309-347, 1992. Kluwer Academic Publishers, Boston.

[18]. *** Genie package software. User Manual.

[19]. Fawcett, Tom (2006). "An Introduction to ROC Analysis". Elsevier - Pattern Recognition Letters. 27 (8): 861–874. doi:10.1016/j.patrec.2005.10.010.