Journal of Advances in Computer Research
Quarterly ISSN: 2008-6148
Sari Branch, Islamic Azad University, Sari, I.R.Iran
(Vol. 3, No. 1, February 2012), Pages: 13-22
www.jacr.iausari.ac.ir

A Surface Water Evaporation Estimation Model Using Bayesian Belief Networks with an Application to the Persian Gulf

Alireza Sadeghi Hesar[1]✉, Hamid Tabatabaee[2], Mehrdad Jalali[1]

(1) Department of Computer Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran
(2) Department of Computer Engineering, Quchan Branch, Islamic Azad University, Quchan, Iran
Alireza.sadeghi89@yahoo.com; Hamid.tabatabaee@gmail.com; Jalali@mshdiau.ac.ir

Received: 2011/12/26; Accepted: 2012/01/19

Abstract

Evaporation phenomena is a effective climate component on water resources management and has special importance in agriculture. In this paper, Bayesian belief networks (BBNs) as a non-linear modeling technique provide an evaporation estimation method under uncertainty. As a case study, we estimated the surface water evaporation of the Persian Gulf and worked with a dataset of observations (daily evaporation) in a network of 13 weather stations in the provinces of Hormozgan and Bushehr. Two major categories of methods for learning Bayesian networks are parameter learning and structure learning. In the first step, k2 search algorithm be used as a score-based method for structure learning of BBN. K2 algorithm connects weather stations to other and Makes a virtual network of stations. In the second step, Netica software be applied for parametric learning. Obtained network by k2 algorithm with the help of a probabilistic inference method (reduce gradiant) in Netica can predict the rate of evaporation. The results of the proposed method, indicating that this model has the more accuracy and reliability than existing statistical methods.

Keywords: *Evaporation, Belief Bayesian Networks, Persian Gulf, Structural Learning, Netica*

1. Introduction

Evaporation is the process by which water is converted from its liquid form to its vapor and thus transfers very big energy from land and water masses to the atmosphere. The phenomenon of evaporation affects the distribution of water in the hydrological cycle and plays a basic role in agriculture and water resource management. The rate of evaporation depends upon wind speed, tempreture and humidity. Essentially, meteorology and climatic factors estimation is very complex task. Firstly, in weather forecasting process, various topics are discussed such as weather patterns, great weather systems, climatic models and also the impacts of land and sea surface temperatures on climate**.** Secondly, the computational requirements of algorithms and training datasets must be provided. In recent decade, most of meteorology issues are solved with Acyclic Circulation Methods (ACMs). Generally, this classic methods for production of climatic and atmospheric patterns use descriptive models on short-term timescales. But today, the increasing availability of data Such as observational records, radar maps, satellite maps and approximate data, led to the development of statistical techniques and intelligent data mining in atmospheric sciences. ACMs are based on objective observations but data mining uses inferential observations. Today, many attempts have been performed to integrate classic ACMs with regression methods and machine learning techniques. Basically, this hybrid methods have been designed based on the discovery of relationships between stored observations in the meteorological databases.

One of the most important tasks in data mining is building a descriptive model of the database being mined. To do so, probability-based approaches have been considered an effective tool because of the uncertain nature of descriptive models. Unfortunately, high computational requirements and the lack of proper representation have hindered the building of probabilistic models. To alleviate the above twin problems, probabilistic graphical models have been proposed. In the past decade, many variants of probabilistic graphical models have been developed, with the simplest variant being Bayesian networks. BNN is a popular descriptive modeling technique for available data by giving an easily understandable way to see relationships between attributes of a set of records. It has been employed to reason under uncertainty, with wide different roles in the field of medicine, finance, and meteorology. Computationally, BNN provides efficient and compelet ways to represent relationships between attributes and allow reasonably fast inference of probabilities.

Fortunately, BBNs assemble suitable facilities for optimal coverage of a geographical network. This paper, evaluates performance of BBNs for surface water evaporation estimation in the Persian Gulf. proposed BBN structure is formed based on weather stations that record Observations for climatic variables. The rest of the paper is organized as follows. In the next section, several previous related works are briefly reviewed. The area of study with weather stations will be introduced in the section 3. Section 4 reviews the main concepts related to bayesian belief networks. Finally, Section 5 and 6 summarize the main results and conclusions.

2. Related Works

Using artificial intelligent tools in estimate of evaporation has long history. In this section, several previous related works are briefly reviewed. Özgür Kişi (2006) applied neuro-fuzzy (NF) method to design a daily evaporation estimation model and evaluated effect of five different climatic variables including humidity, pressure, solar radiation, wind speed and temperature on evaporation rate in proposed model. D.R. Harp et al (2007) expanded an evaporation forecasting model using a fuzzy learning from example (OFLFE) algorithm. This research considered use of fuzzy techniques to inference under experimental uncertainty. The recorded dataset in the riparian zone of the Middle Rio Grande (New Mexico, U.S.A) applied for learning process of proposed method. Researchs of Moghaddamnia et al. (2009) and Alpaslan Yarar et al. (2009) investigated the performance of hybrid methods based on adaptive neuro-fuzzy inference system (ANFIS) and artificial neural networks (ANNs) on evaporation estimation and water level changes in lakes. Fi-John Chang et al. (2010) proposed a self organizing map neural network (SOMN) to evaluate the variability of daily evaporation based on effective local variables. The topological structure of SOMN produces a meaningful map to present the clusters of meteorological variables and so estimates the daily evaporation.

Emrah Dogan et al. (2010) estimated daily evaporation rate from measured meteorological data on Yuvacik Dam zone (Turkey) using an adaptive neuro-fuzzy inference system (ANFIS) and compared the results of proposed method with multiple linear regression (MLR) model. Nourani And Sayyah (2012) used the different ANNs such as Radial Basis Neural Network (RBNN), Multi Layer Perceptron (MLP) and Elman network for estimating daily evaporation rate. Urmia lake be considered as area of study in this research. In comparison with the previous works, this paper proposes a simpler evaporation estimation model. The following parts described the proposed mthod with details.

3. Area of Study

In this paper, Persian Gulf is as geographical area of study (See Figuer. 1). We have a database of daily records related to the evaporation of surface water at 13 synoptic meteorological stations (See Figuer. 2) published by the Persian Weather Services. The dataset cover the period from 1985 to 2010. Synoptic weather stations are responsible to measurement[1]

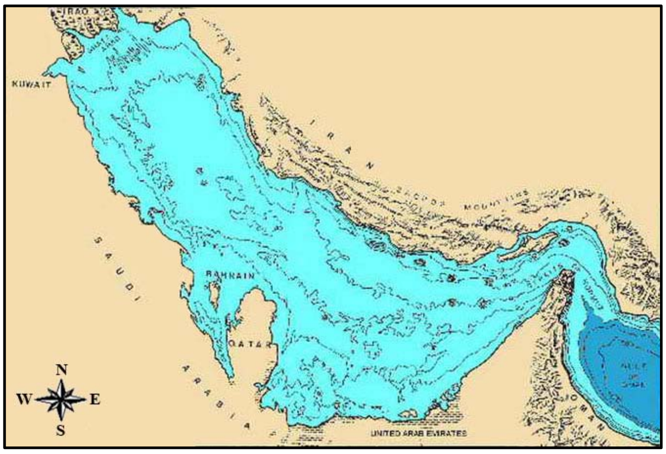


Figure 1. Geographical area of study – Persian Gulf

1. This measurement is done in two ways: Synop (Every three hours) and Metar (Hourly)

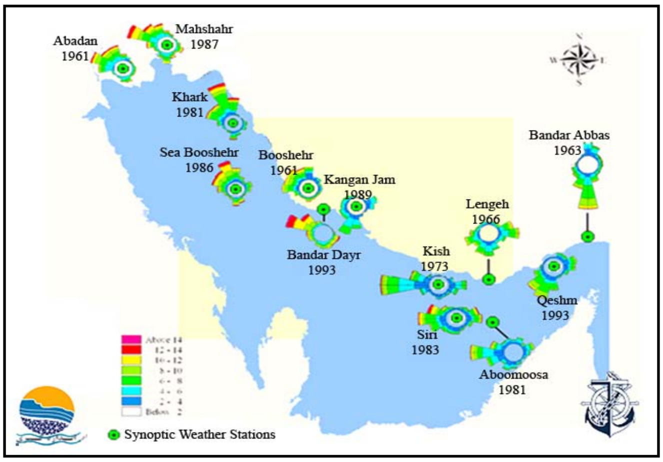


Figure 2. Synoptic weather stations in Booshehr and Hormozgan provinces

4. Bayesian Belief Networks

Bayesian belief networks is named based on studies of Thomas Bayes (1702–1761) in the field of probability theory. His studies led to the production of Bayes' rule Which is expressed as follows:

$$P\ (b|a) = \frac{P\ (a|b) * P(b)}{P(a)} \tag{1}$$

Basically, BBNs are known for providing a compact and simple representation of probabilistic information, allowing the creation of models associating a large number of variables. A BBN is composed of three basic concepts, including DAG[1], CPT[2] and JPD[3]. BBN use a DAG for discovery and represent of dependences between variables. variables and dependencies are represented as the nodes and edges of DAG. Different values of the random variables are determined based on their possible states. If there is a causal relationship between A and B variables, the edge is a line that connects A to B. Each node has a CPT associated with it, and provides the it's probability being in a particular state, given any combination of parent states. When evidence is entered for a node in the network, the fundamental rule for probability calculus and Bayes' rule can be used to propagate this evidence through the network. A BBN factorizes a JPD over a set of variables based on product of conditional probability distribution. The JPD is factored as:

$$P_r\ (y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n} P\ (y_i/\pi_i) \tag{2}$$

Consider the following simple example that indicates some of the properties of BBNs by DAG and with CPTs for each node when the variables are discrete (consists of four random variables).

1. Directed Acyclic Graph
2. Conditional Probability Table
3. Joint Probability Distribution

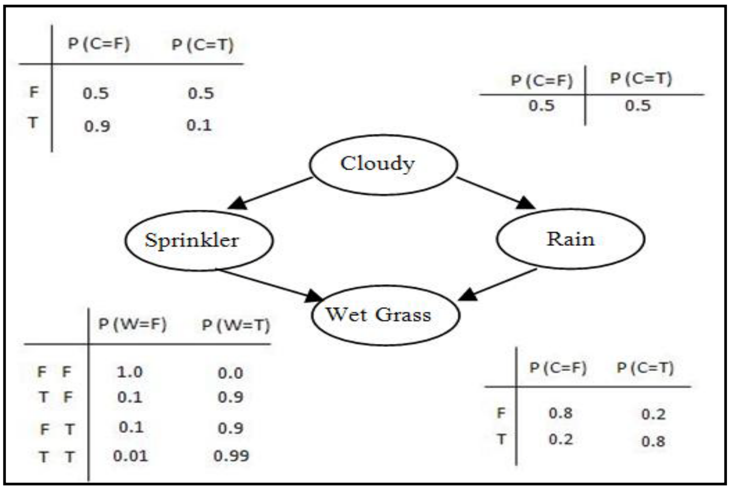


Figure 3. A simple of BBN

and the atmospheric parameters production and send them in telecommunication networks according to the rules and regulations of the World Meteorological Organization simultaneously all over the world. Before any processing on the data set, Continuous variables should be converted into discrete variables using a classification method. For evaporation variable, four different classes are considered: class1 = no evaporation (evaporation < 6mm), class 2 = Low evaporation (6mm< evaporation < 30mm), class 3 = moderate evaporation (30mm < evaporation < 60mm), class 4, high evaporation (evaporation > 60mm).

Figure 3 shows conditional distributions of nodes using CPTs. For the continuous distributions different methods are used. For example, Wet Grass Node has two parents. Probability of this node is dependent on the Sprinkler and Rain as parent nodes.. The degree of relationship is shown in CPT rows of Wet Grass. Cloudy Node has no parents, so probability will be determined using initial values.

Generally, the joint probability of DAG shown in figure using the chain rule is:

$P(C, S, R, W) = P(C) * P(S|C) * P(R|C, S) * P(W|C, S, R).$

This dependence using conditional independence relationships can be written as follows:

$P\ (C, S, R, W)\ =\ P(C)\ * P(S|C)\ * P(R|C)\ *\ P(W|S, R).$

Basically, the conditional independence relationships represent the complex and complete joint probability by reduce requerment storage spaces. The full joint requires O(2^n) bytes, but the modified form requires O(n 2^k) space to represent that k is the maximum fan-in of a node and n is the number of binary nodes.

5. Results And Discussion

The BBNs construction process can be separated two major steps: parameter learning and structure learning. Since the BBNs are statistical model and the parameter learning is a method for learning in statistics, it can also be used in Bayesian networks. Basically, parameter learning is calculation of the conditional probabilities for the obtained structure of BNN and parameters in a BBN are the probabilistic values in the CPTs. The most effective methods for parameter learning are reduce gradiant and maximize likelihood estimation (MLE).

The main objective of the structure learning is finding the best structure for BBN that is compatible with existing dataset and is optimal from the point of complexity and construction tine. The structure learning includes two different approaches of constraintbased learning and score-based learning. In constraint-based learning the network structure is achieved using conditional independence relations between variables. But score-base learning assigns a score to each possible topology and tries to maximize it using metric scoring functions. The finding optimal structure for BBNs is a NP-Hard problem. In the context of meteorology problems although BBNs can be used for weather forecasting but there are very high scenarios to connect the weather stations in a geographical area. To solve this problem, Greedy search algorithms such as K2 search, hill-climbing and tabu search are a common choice. Generally, greedy search algorithms are based on score-based methods that offer scale or metric solutions. This methods evaluate all of the possible relationships between nodes in the general space, and determine a instance with the maximum ranking.

5.1 Structure Learning

In this paper, K2 search algorithm as a structure learning method is used. K2 algorithm (See Figure. 4) is a greedy algorithm that obtains the best structure through a iterative process among all possible arangement.

```
K2 Algorithm

procedure K2;
{Input: A set of n nodes, an ordering on the nodes,
 an upper bound u on the
number of parents a node may have,
 and a database D containing m cases.}
{Output: For each node, a printout of the parents of the node.}
for i:= 1 to n do
 π_i := φ;
 P_old := f(i,πi )
 OKToProceed := true;
 While OKToProceed and |π_i| < u do
   let z be node in Pred(π_i) - π_i that maximizes f(i, π_i ∩{z});
   P_new := f(i, π_i ∩{z});
   if P_new > P_old then
   P_old := P_new ;
   π_i := π_i ∩ {z};
   else OKToProceed := false;
   end {while};
write("Node: ", xi, " Parent of xi: ", π_i);
end {for};
end{K2};
```

Figure 4. K2 algorithm psudo code

Before the execution of scoring function, the variables must to be ordered and a fixed and limited amount to be considered for the parents of each node. Usually in specific applications such as weather forecasting, the experts of field determine order of nodes and amount of parents for each node. this procedure prevents from generating loops in the DAG and final score of network will obtains by multiplying the individual score of nodes. The scoring function of K2 algorithm for i nodes is in this format:

$$f(\mathrm{i}\,, \pi_{\mathrm{i}}) = \prod_{j=1}^{|\Phi_i|} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} \alpha_{ijk}\,! \qquad (3)$$

Table 1, presents details of the scoring function.

Table 1. Detailes of scoring function

	The current node
	The number of states that can have i
	The parents of i
$\lvert\varphi_i\rvert$	The number of values within the CPT of i
α	The number of cases in the dataset in which i has its k^{th} value and π_i have their j^{th} value in the CPT
N_{ij}	The sum of the α_{ijk} for each state of i

In this section, the nodes of DAG are weather synoptic stations (See Figuer. 2) that record evaporation of each region. Each node of DAG represens a geographical area. Four states are defined for each node that was fully expressed in the second section. A set of nodes, an ordering on the nodes, an upper bound u on the number of parents a node may have, and a database D containing m cases are inputs of K2 algorithm and a printout of the parents of the node for each nodes is output of K2 algorithm. With Considering a maximum of 3 parents for each node or U=4 and 3-nearest neighbors for each nodes or K=3, K2 algorithm produces the network shown in figure 5.

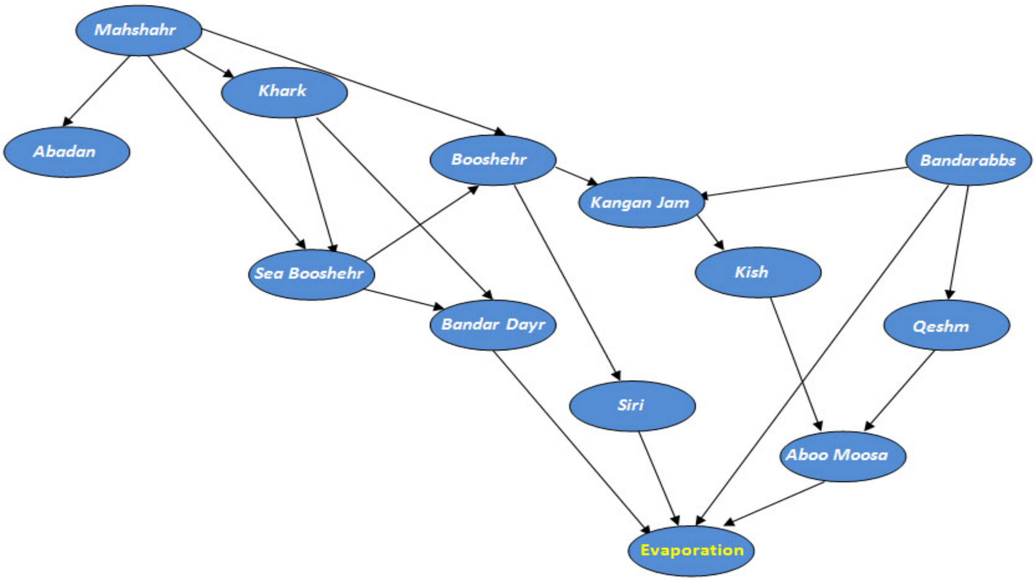


Figure 5. Proposed BBN for connecting weather stations in Persian Gulf

5.2 Parameter Learning

The obtained Bayesian network structure in the structural learning step is given to Netica[1] software for parametric learning. The Netica software makes CPTs based on the type of variables, number of states and obtained structure. And then using reduce the gradient technique as a parametric learning method, computes values of CPTs. The obtained Values at the nodes of final level are considered as predicted resultes. Exprimental data set for learning process cover the period from 1985 to 2010. Figure 6

1. Netica Software is a Bayesian Simulator

shows the output of Netica software for proposed BBN and Figure 7 shows the the inference method of Netica for last node.

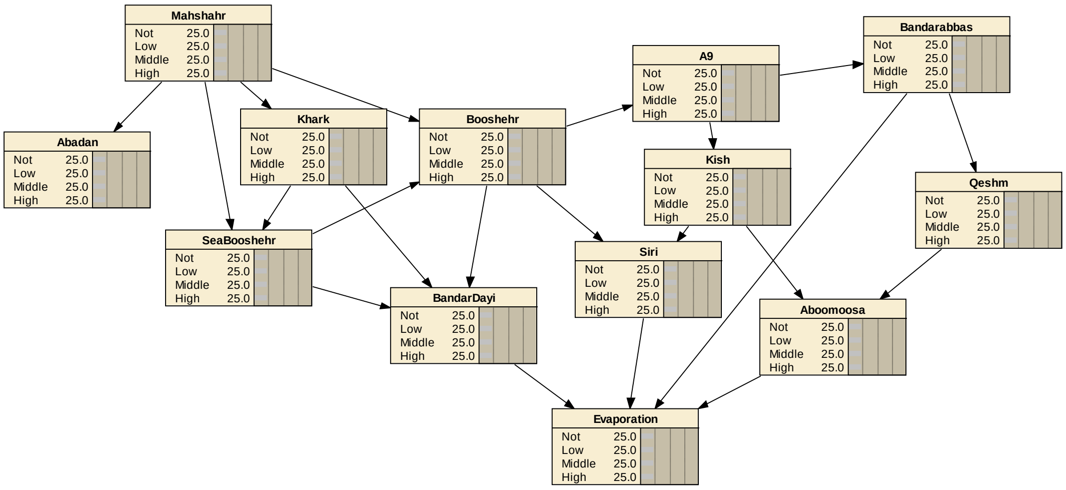
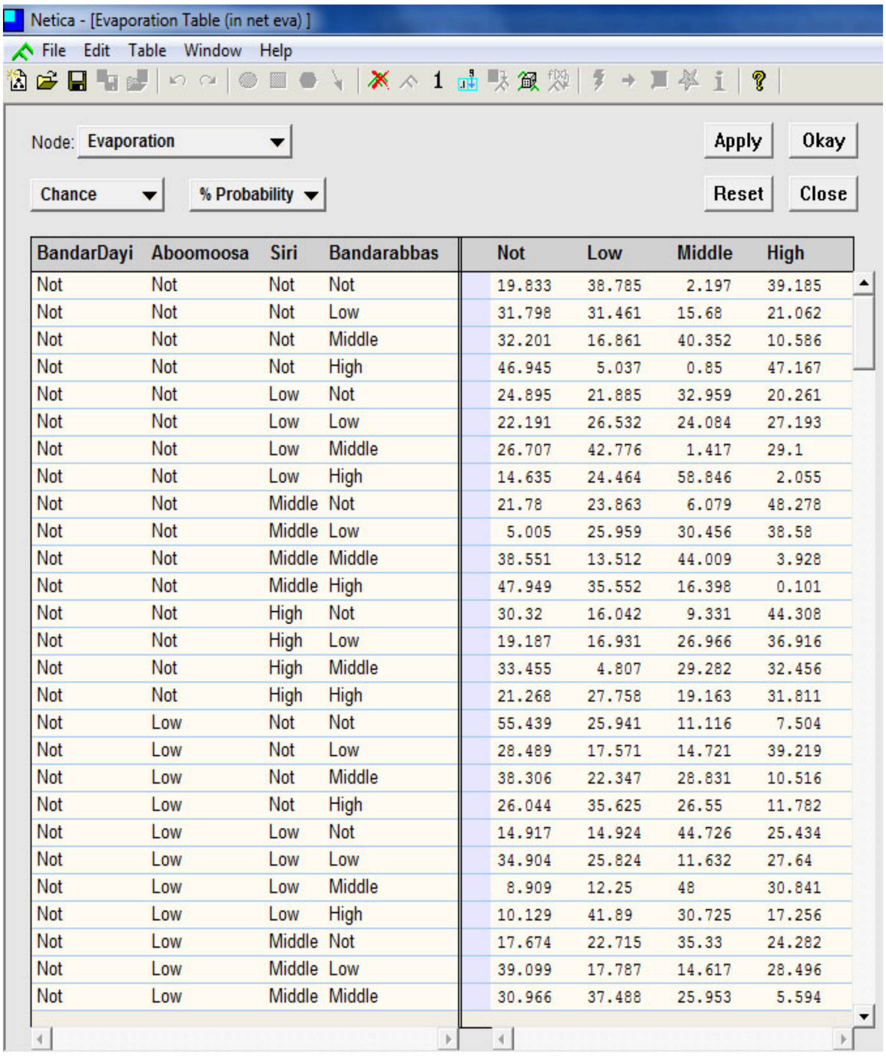


Figure 6. Output of Netica software



BandarDayi	Aboomoosa	Siri	Bandarabbas	Not	Low	Middle	High
Not	Not	Not	Not	19.833	38.785	2.197	39.185
Not	Not	Not	Low	31.798	31.461	15.68	21.062
Not	Not	Not	Middle	32.201	16.861	40.352	10.586
Not	Not	Not	High	46.945	5.037	0.85	47.167
Not	Not	Low	Not	24.895	21.885	32.959	20.261
Not	Not	Low	Low	22.191	26.532	24.084	27.193
Not	Not	Low	Middle	26.707	42.776	1.417	29.1
Not	Not	Low	High	14.635	24.464	58.846	2.055
Not	Not	Middle	Not	21.78	23.863	6.079	48.278
Not	Not	Middle	Low	5.005	25.959	30.456	38.58
Not	Not	Middle	Middle	38.551	13.512	44.009	3.928
Not	Not	Middle	High	47.949	35.552	16.398	0.101
Not	Not	High	Not	30.32	16.042	9.331	44.308
Not	Not	High	Low	19.187	16.931	26.966	36.916
Not	Not	High	Middle	33.455	4.807	29.282	32.456
Not	Not	High	High	21.268	27.758	19.163	31.811
Not	Low	Not	Not	55.439	25.941	11.116	7.504
Not	Low	Not	Low	28.489	17.571	14.721	39.219
Not	Low	Not	Middle	38.306	22.347	28.831	10.516
Not	Low	Not	High	26.044	35.625	26.55	11.782
Not	Low	Low	Not	14.917	14.924	44.726	25.434
Not	Low	Low	Low	34.904	25.824	11.632	27.64
Not	Low	Low	Middle	8.909	12.25	48	30.841
Not	Low	Low	High	10.129	41.89	30.725	17.256
Not	Low	Middle	Not	17.674	22.715	35.33	24.282
Not	Low	Middle	Low	39.099	17.787	14.617	28.496
Not	Low	Middle	Middle	30.966	37.488	25.953	5.594

Figure 7. Inference method in Netica Software for Evaporation Node

6. Conclusions

Evaporation phenomena has special importance in agriculture and water resources management. In this paper, a Bayesian Belief Network (BBN) for surface water evapoaration estimation is presented. BBNs as probabilistic graphical models provide efficient facilities for reseaning under uncertainty. Proposed BBN uses K2 algorithm for structural learning step and uses Netica software for parametric learning step. Netica software apply reduce gradiant technique for inference. The case study is estimation of surface water evaporation in Persian Gulf. Analysis and Comparison of the actual values with estimated values for two different intervals of time (the first decade of October and first decade of November 2011) is evident in the following diagrams that reflects acceptable accuracy of proposed method (in around 90/14% of cases). The Bayesian Belief Networks can be used more often for many application of meteorology and weather forecasting like rainfall, wind speed or soil moisture estimation. As future works, we are interested to investigate the use of other greedy based algorithms such as tabu search, hill climbing and greedy thick thinning for searching the state space in structural learning.

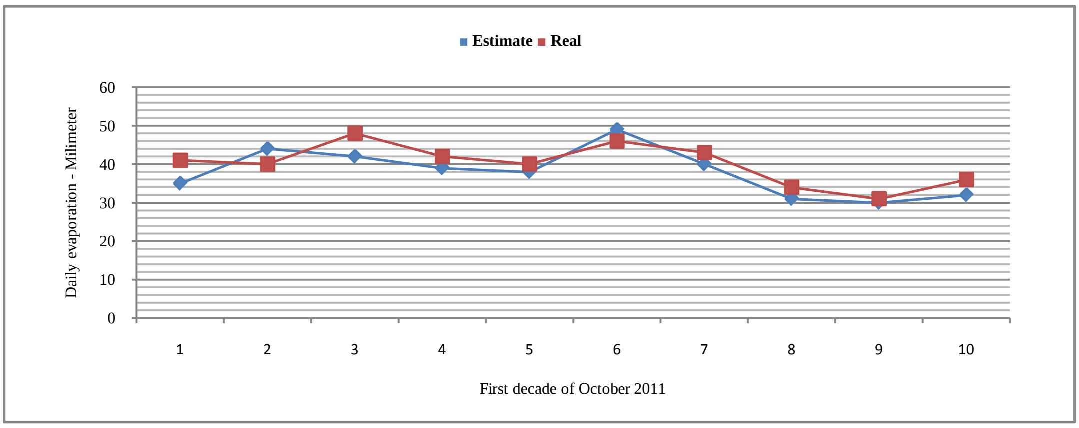


Figure 8. Comparison of predicted values with actual values (first decade of October 2011)

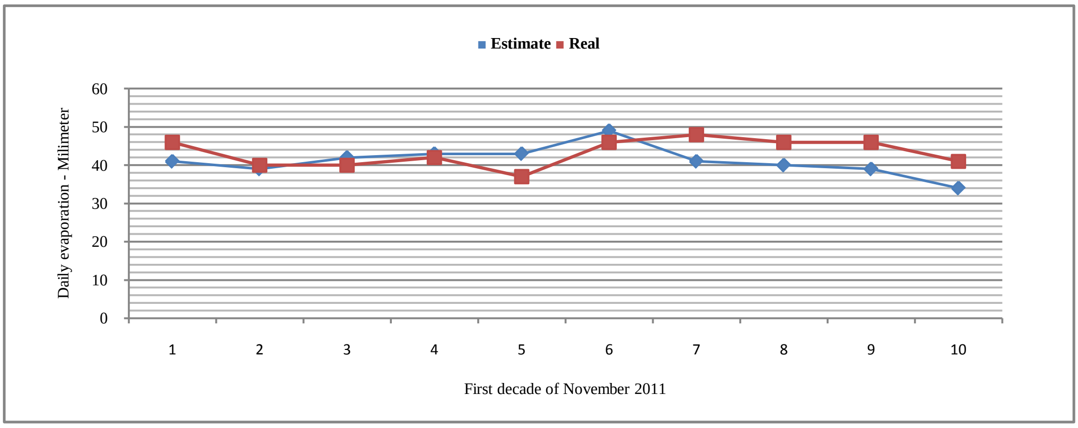


Figure 9. Comparison of predicted values with actual values (first decade of November 2011).

7. References

[1] A.Cofino, R.Cano, C.Sordo and Jose M. Gutierrez, "Bayesian Networks for Probabalistic Weather Prediction," Proc. of the 15th European Conference on Artificial Intelligence, IOS Press, 2000, pp. 695-700.

[2] R.Cano, C.Sordo and M.J.Guti´errez, "Applications of Bayesian Networks in Meteorology Advances in Bayesian Networks," in Advances in Bayesian Networks, Gámez et al. eds., Springer, 2004, pp. 309-327.

[3] M.Kent, H.Le, M.Tadross and A.Potgieter, "Weather Forecasting With Bayesian Network," Ph.d, Department of Computer Science University of Cape Town , Cape Town, South Africa, 2008.

[4] B.Lee and J.Joseph," Learning a probabalistic model of rainfall graphical models", Quarterly Journal of the Royal Meteorological Society, 1998, pp.90–96.

[5] K.Korb and A.Nicholson, Bayesian Artificial Intelligence, Chapman & Hall/CRC Press LLC, 2004, pp. 75-83.

[6] A.Khanteymoori, M.M.Homayounpour and M.B.Menhaj, "A Bayesian Network Based Approach For Data Classification Using Structural Learning,"Communications in Computer and Information Science, Springer, 2009, pp. 25-32.

[7] E.Lamma, F.Riguzzi and S.Storari," Improving the K2 Algorithm Using Association Rule Parameters," Journal of Modern Information Processing, Elsevier, 2006, pp. 207-217.

[8] D.Heckerman, A tutorial on learning with Bayesiannetworks. In M. I. Jordan (Ed.), Learning in Graphical Models, MIT Press, 1999, pp.301-354.

[9] D. Heckerman, D.Geiger and D.M.Chickering, “Learning Bayesian networks: the combination of knowledge and statistical data. Machine Learining,” Vol. 20, No. 3, 1995, pp.197-243.

[10] E.Gyftodimos and P.Flach, "Hierarchical Bayesian networks: an approach to classification and learning for structured data, Methods and Applications of Artificial Intelligence," Proc. Third Hellenic Conference on AI, SETN 2004, Samos, Greece, 2004.

[11] S. Russell and P.Norvig, "Artificial intelligence, a modern approach", New York: Prentice Hall, 2003.

[12] N.Friedman, and M.Goldszmidt, “Learning Bayesian networks with local structure,” Proc. The Twelfth Conference on Uncertainty in Artificial Intelligence, 1996, pp. 252-262.