Bayesian networks application to forecast the national economies development taking into account the water factor

N Savina¹, N Kovshun², V Kostrychenko³, M Voronenko⁴ and V Koval³
¹ National University of Water Management and Environmental Engineering, Rivne, 33000, Ukraine
² Kherson National Technical University, 24 Berislavske shose, 73008 Kherson, Ukraine
³ National Academy of Sciences of Ukraine, Volodymyrska St., 54, 01601 Kyiv, Ukraine
E-mail: n.e.kovshun@nuwm.edu.ua

Abstract. The influence of this factor on the development of national economies is studied. This study uses the Bayesian network, which illustrates the interaction of indicators of water supply and water use and includes nodes that are formally represented as vectors. The most valuable result of the modelling is not the general forecast obtained with the help of the model, but the structure of the network itself, which allows to identify connections within the model and explain the reason for the emerging interdependencies. Modelling using Bayesian networks confirmed the fact that there is a direct relationship between GDP and water consumption and drainage. The obtained results confirm the possibility of achieving an increase in the overall GDP of the country with an increase in the amount of water resources used for production needs. However, this should be implemented in combination with a simultaneous reduction in the volume of return (wastewater) discharged into surface water bodies.

1. Introduction
The national wealth of each country is largely determined by the quantitative and qualitative characteristics of its resource components, among which special attention needs water resources. Their influence on the location of productive forces, large-scale involvement in production processes, importance for the population livelihood allow us to consider water resources as a priority national economy component and the functioning of the state as a whole. Balancing the relationship between economic development and water consumption is an important issue for many countries around the world and is determined not only by the factor of water supply but also by the production structure [1].

The water resources conservation problem has a planetary nature and requires mutually agreed political, environmental, economic action [2-5].

To predict the economy development, these aspects can be taken into account using various mathematical models. For example, the Kuznets ecological curve was used to balance the relationship between economic development and industrial water consumption in China [6]. The relationship between the urbanization level and water consumption at the national and subnational levels has been studied through the indirect correlation [7]. The virtual water study and its impact on the global economy is relevant today [8-10]. The water valuation, like any other natural resource, is also considered through the circular economy principles [11, 12].
In recent years, the water supply problem is exacerbated at all stages - from the global scale to the local level. This necessitates the formation of a water use system that could meet new challenges, meet the sustainable development principles, rely on innovative approaches in management, take into account Ukrainian realities and function organically as part of Ukraine national economy. Therefore, there is a need to develop the water optimal system use for Ukraine, the basis of which is the incorporation of sustainable development principles to water resources to ensure the country economic progress [13, 14]. The priority of such a system will be to rationalize the water resources distribution between public life spheres and industries, aimed at growing the national economy, socio-economic revival in the regions and improving the environment. Special attention needs to be paid to the models creation of analysis and the efficiency forecasting of water use processes, assessment of their impact on the level of national economy development.

The statistical data analysis showed that among the national economy branches of Ukraine the main water user is the industry, for which 4.28 billion cubic meters were taken from natural water bodies. Industrial enterprises used 4.20 billion cubic meters of fresh water, 96% of which are used for production purposes. The total drainage in this area is 3.58 billion cubic meters (83.7% of the water intake). The reusable and consistent use of water amounted to 33.84 billion cubic meters for industrial purposes in 2019. For the purposes of agriculture, forestry and fisheries, water intake amounted to 4.38 billion cubic meters. 18.6% of fresh water from 1.79 billion cubic meters went to production needs; total drainage was 0.42 billion cubic meters (9.6%). For the needs of housing and households were taken 2.41 billion cubic meters from natural water bodies. For production enterprises needs in this industry used 0.34 billion cubic meters (or 27.2% of fresh water volume); total drainage was 1.52 billion cubic meters (62.9%), and reused 0.05 billion cubic meters. For other national economy sectors, the volume of water use is relatively small, but they were taken into account in the modeling process using the Bayesian network. The statistical data analysis for 2019 showed that in the ranking of Ukrainian industries in GDP terms, the leading position is occupied by industry is 777.49 billion UAH; housing and household are 738.32 billion UAH; trade and public catering are 561.61 billion UAH; transport, warehousing, postal and courier activities are 447.53 billion UAH; agriculture, forestry and fisheries are 356.80 billion UAH; education is 314.17 billion UAH; health care is 119.49 billion UAH; construction is 107.43 billion UAH.

2. Problem Statement

In this study, the Bayesian network illustrates the water supply interaction and water use (X) and includes nodes that are formally represented as vectors. Parameters (X) have three states: minimum, average and maximum. The target node (Y) has two states: minimum and maximum. It is this node, being the network resulting node, shows whether there is a normal or increased load on GDP. In the research process, a model was developed that analyzes water resources and the relationship with GDP, and set the following tasks: first, to establish the relationship between the integrated indicator of gross domestic product (Y) and the parameters that characterize water use: the volume of water taken from natural water bodies (X1), the volume of fresh water used, total (X2), the water used volume, including for production needs (X3), the total drainage volume (X4), the reversible volume, reusable and sequential use of water (X10); secondly, to establish whether there is a connection between parameters X4 and X10.

To solve this problem, it will be appropriate to assess the water use impact of and water supply indicators on the country's GDP with the help of Bayesian networks (BNs). Obviously, the indicators of water intake from natural sources, general water disposal, the discharged water amount (including untreated ones), as well as factors of reuse should not be considered separately from each other, but in aggregate with GDP. A model analyzing water resources should take into account all the processes associated with the water exploitation, as well as the relationship with GDP. Since BNs are causal chains, that is, they show causal relationships between "parent and child" nodes, we can use them to predict the most likely outcome. With each change in the input data, the BN will give a new forecast of the events development.
3. Method

There are two methods for quantifying the situation occurring risk: using a priori statistics and using modeling.

The disadvantage of applying traditional statistics is that it analyzes data from the past and does not take into account the situation development and the introduction of new data. Situational risk modeling seems to be a more progressive method, since the situation is assessed before a particular incident occurs. For example, on the theoretical basis of Kuznets curve ecological curve, the complex influence of socio-economic factors on water use in the aggregate and by industries is estimated [15].

The most valuable modeling result is not a general forecast obtained with the help of the model, but the structure of the BN itself, which makes it possible to single out connections within the model and explain the reason for the emerging interdependencies. Moreover, every time we change the input data, we will receive a new numerical result at the output, that is, in a situational analysis, each node can be considered as a measure of control over the state of the system, which can be investigated.

BNs are a powerful and flexible tool for analyzing economic data and research results [16, 17]. The methods for developing a forecasting tool can be divided into 2 approaches: regression models and machine learning algorithms. Risk modeling using BN has advantages over regression-based approaches, and in this article we will focus on three of them:

1. creating network structures in which the relationship between variables can be easily established;
2. the ability to apply Bayes' theorem to assess trends at an individual level;
3. easy transformation of Bayesian models into decision making models.

Bayesian network (or Bayesian network, belief network) is a graph probabilistic model, which is a set of variables and their Bayesian probabilistic dependencies. For example, a Bayesian network can be used to calculate the likelihood that a patient is sick from the presence or absence of symptoms number, based on data on the relationship between symptoms and disease. The mathematical apparatus of Bayesian networks was developed by the American scientist Jude Pearl in 1988 [18].

The observed events cannot be unambiguously described as direct consequences of strictly deterministic causes. And therefore, in practice, a probabilistic description of phenomena is widely used. There are several reasons for this:

- the presence of fatal errors in the process of experimentation and observations;
- the complete description impossibility of the system structural complexities under study;
- uncertainties due to the observations scope finiteness.

There are certain difficulties on the probabilistic modeling way, which can be conditionally divided into two groups:

- technical (computational complexity);
- calculating the variables values probabilities of interest to us in a Bayesian network, provided that we have at our disposal some information about the other variables values, is the statistical inference task. It is solved using conditional probabilities. The conditional probability concept forms the basis of the Bayesian approach to uncertainty analysis. This formula means “assuming what happened (and everything else that has nothing to do with A). The occurrence joint probability of events A and B is determined by the total probability formula:

   \[ P(A, B) = P(A | B) \cdot P(B) \]  

If we have information about the dependent variables (consequences), and the essence of the study is to determine the comparative probabilities of the initial variables (causes), then Bayes' theorem comes to the rescue. Let there be a conditional probability \( P(A | B) \) of a certain event occurring A, provided that an event has occurred B. Bayes' theorem gives a solution for the inverse problem - what is the probability of an earlier event occurring B if it is known that a later event has occurred A. More precisely, let be a \( A_1, ..., A_n \) set (full group) of incompatible mutually exclusive events (or
alternative hypotheses). Then the posterior probability $P(A_j \mid B)$ of each of their events $A_j$, provided that an event has occurred $B$, is expressed in terms of the prior probability $P(A_j)$:

$$P(A_j \mid B) = \frac{P(A_j) \cdot P(B \mid A_j)}{P(B) = \sum_{j=1}^{n} P(A_j) \cdot P(B \mid A_j)}$$ (2)

The inverse probability $P(B \mid A_j)$ is called the likelihood, and the denominator $P(B)$ in Bayes' formula is the evidence.

The joint probability is the most complete statistical description of the observed data. Thus, the Bayesian network (Fig. 1) consists of the following concepts and components:
- a lot of random variables and directional connections between them;
- each variable can take one of a finite set of mutually exclusive values;
- variables together with connections form a directed graph without cycles;
- each descendant variable $A$ with ancestor variables $B_1, \ldots, B_n$ is assigned a table of conditional probabilities $P(A \mid B_1, \ldots, B_n)$.

Fig. 1. Static Bayesian model, which is built using the algorithm of structural learning TAN of the form Bar Chart.
The designed static BN has qualitative and quantitative components. The qualitative component is the network structure, its conceptual model, derived from the experts' knowledge. The quantitative component is conditional probability distributions that define the exact relationships between states and variables.

Learning a Bayesian network to determine its structure involves using the following algorithms:
1. Bayesian Search
2. Conditional path search (PC)
3. Search by main graph
4. Algorithm of greedy thick thinning (GTT or Greedy)
5. Naive Bayes Supplemented Tree Algorithm (TAN)
6. Advanced Naive Bayes (ANB)
7. Naive Bayes Algorithm (NB)

A. Dempster and Drudzel [19] experimentally confirmed that the PC algorithm is sufficiently resistant to multivariance. They performed a search for base graphs using the PC algorithm and evaluated various base graphs using a Bayesian search approach. The greedy thick thinning (GTT) structure learning algorithm is based on the Bayesian search approach and was described in the papers by Cheng [20]. The tree structured learning algorithm naive Bayesian augmented (TAN) is based on the Bayesian search approach described and thoroughly researched in the works of Friedman [21].

4. Initial data
Having input indicators, it is necessary to build a static Bayesian network that allows to determine the influence of the input data on the country's GDP integral indicator. In our study, we work with the Genie2.3 Academic software environment and we use a set of interrelated statistical data, consisting of 11 indicators in 11 areas for the period 2018-2019.

We design the necessary nodes with probabilities of 0.33 in each state for nodes X₁-X₁₀ with three states and with probabilities of 0.5 in each state of the resulting node Y. Overall accuracy for the BN is 40.7%, the result accuracy is 66.8%.

The simulation results do not contradict common sense. This fact confirms that the model is adequate to the processes under study. Since the model was developed to assess trends, and not to obtain accurate results, it has some limitations, including those inherent in the BN in general:
1. Incompleteness of the model due to the absence of some nodes can affect the result. For example, the node “Financing of water supply authorities” was not included in the model, because at the moment there is no data on the relationship between the level of water supply financing and the GDP level.
2. Greater robustness of the model leads to low the output sensitivity. This serves as a signal for the need to introduce a procedure for "weighing" the nodes.
3. The statistical data were used for modeling. These data are characterized by insufficient information, which is acceptable for practical analysis.

5. Discussion of experimental results
The study shows that the parameters X₁, X₂, X₃, X₄ in the state of "minimum" network are insensitive, and in the state of "maximum" the integrated value of Y in some nodes (X₁, X₃, and X₁₀) increases, and in nodes X₂ and X₄ - decreases. Thus, the model indicates a GDP growth of 3% (from 68% to 71%) depending on the water amount taken from natural water bodies (Fig. 2). The relationship between X₁ and GDP can be explained as follows: as water abstraction from water natural bodies increases to a maximum, GDP will increase. However, the stock of GDP shown by our model is insignificant (3%), so it is not enough to simply increase the water volume intake to significantly increase GDP.
Fig. 2. The modeling result with increasing water intake from natural water bodies.

Next, the model showed GDP growth of 8% (from 68% to 76%) as a result of increasing the water use for production needs to a maximum (Fig. 3).

The simulation result showed a decrease in GDP by 1% (from 68% to 67%) due to an increase in the use of fresh water and by 4% (from 68% to 64%) as a result of an increase in total drainage to a maximum (Fig. 4).

Calculations also showed a 2% increase in GDP (from 68% to 70%) with a decrease in the volume of circulating, reusable and consistent use of the country water resources (Fig. 5).

Fig. 3. The modeling result with increasing use of water resources for production needs.
In addition, the constructed model showed that when changing any parameters of $X$, in particular the total drainage ($X_4$), the volume of reversible, reusable and sequential use of water ($X_{10}$) decreases by 11-14% from the initial 35% to 23-21%.

6. Conclusions

Thus, modeling using Bayesian networks confirmed the fact that there is a direct relationship between GDP and water consumption and drainage. The obtained results confirm the possibility of achieving an increase in the overall level of GDP of the country with an increase in the water resources amount used for production needs. However, this should be implemented in combination with a simultaneous reduction in the volume of return (wastewater) discharged into surface water bodies. When individual nodes interact (for example, parameters $X_3$ and $X_4$ in their maximum state), parameter $X_{10}$ decreases from 35% to 10%. GDP is growing by 31% and reaches its maximum value (99%). By changing the conditions to the opposite, we get the parameter $X_{10}$ at the level of 19%, and GDP will be the lowest (17% compared to the initial level of 68%).
References
[1] Anríquez G and Melo O 2018 The socio-economic context of Chilean water consumption and water markets growth: 1985–2015 (Springer, Cham)
[2] Abu-Zeid M A 1998 Water and sustainable development: the vision for world water, life and the environment. Water policy 1(1) 9-19
[3] Agboola J and Braimoh A K 2009 Strategic partnership for sustainable management of aquatic resources. Water resources management 23(13) 2761-2775
[4] European Environmental Agency 2012 EEA report no 12/2012. European Environmental Agency. Climate change, impacts and vulnerability in Europe — an indicator-based report.
[5] Aznar-Sánchez J A, Belmonte-Ureña L J, Velasco-Muñoz J F and Manzano-Agugliaro F 2018 Economic analysis of sustainable water use: A review of worldwide research. Journal of Cleaner Production 198 1120-1132
[6] Yue Z, Alun G U & Bolin P 2017 Relationship between industrial water consumption and economic growth in China based on environmental Kuznets curve. Energy Procedia 105 3557-3564
[7] Wang Q, Wang X, Liu Ya and Li R 2021 Urbanization and water consumption at national-and subnational-scale: The roles of structural changes in economy, population, and resources. Sustainable Cities and Society 103 272
[8] Shao L, Guan D, Wu Z, Wang P & Chen G Q 2017 Multi-scale input-output analysis of consumption-based water resources: method and application. Journal of Cleaner Production 164 338-346
[9] Zhang X, Liu J, Zhao X, Yang H, Deng X, Jiang X & Li Y 2019 Linking physical water consumption with virtual water consumption: Methodology, application and implications. Journal of Cleaner Production 228 1206-1217
[10] Wu X D, Guo J L, Li C H, Shao L, Han M Y and Chen G Q 2019 Global socio-hydrology: an overview of virtual water use by the world economy from source of exploitation to sink of final consumption. Journal of Hydrology 573 794-810
[11] Voulvoulis N 2018 Water reuse from a circular economy perspective and potential risks from an unregulated approach. Current Opinion in Environmental Science & Health 2 32-45
[12] Kovshun N, Kostrychenko V, Semeniuk K, Filipishyna L, & Antonova L 2021 Multi-vector activity in the introduction of integrated administration of environmentally safe nature resources use. E3S Web of Conferences 255 01028
[13] Kovshun N, Savina N, Diallo M A, Kushnir N, Zhiwei D and Zoshchuk V 2021 Principles of creating a system of sustainable water use in Ukraine. E3S Web of Conferences 280 10007
[14] Yankovyi O, Koval V, Lazorenko L, Poberezhets O, Novikova M and Gonchar V 2021 Modeling sustainable economic development using production functions. Estud. Econ. Apl. 39(5). http://dx.doi.org/10.25115/eea.v39i5.5090
[15] Hao Y, Hu X AND Chen H 2019 On the relationship between water use and economic growth in China: New evidence from simultaneous equation model analysis. Journal of Cleaner Production 235 953-965
[16] Voronenko M 2020 Dynamic Bayesian Networks Application for Economy Competitiveness Situational Modelling. Advances in Intelligent Systems and Computing 1293 210-224
[17] Lytvynenko V, Naumov O, Voronenko M, Krejci J, Naumova L, Niskyten D and Savina N 2020 Dynamic bayesian networks application for evaluating the investment projects effectiveness. Lecture Notes in Computational Intelligence and Decision Making 1246. https://doi.org/10.1007/978-3-030-54215-3 20
[18] Pearl J 1988 Probabilistic Reasoning in Intelligent Systems (San Francisco, CA: Morgan Kaufmann)
[19] Dempster A P, Laird N M and Rubin D B 1977 Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (Methodological) 39(1) 1-22
[20] Cheng J and Druzdzel M J 2000 AIS-BN: An adaptive importance sampling algorithm for evidential reasoning in large Bayesian networks. Journal of Artificial Intelligence Research 13 155-188

[21] Friedman N 1998 The Bayesian structural EM algorithm. Artificial Intelligence (UAI ’98) (San Francisco, CA: Morgan Kaufmann)