RISK MODELING OF ACCIDENTS IN THE POWER SYSTEM OF UKRAINE WITH USING SDI DATA

Viktor Putrenko

World Data Center for Geoinformatics and Sustainable Development

National Technical University of Ukraine “Kyiv Polytechnic Institute”

37, Peremohy ave., 03056, Kyiv, Ukraine

putrenko@wdc.org.ua

Abstract

The paper presents the research of approaches to spatial risk modeling of accidents in the power system on the example of Ukraine. The study of the basic methods of mathematical modeling of accidents was conducted. The main factors that influence the occurrence of accidents were selected and analyzed. The modeling on the basis of Bayesian networks for several regions of Ukraine was carried out and verification of the results was conducted based on SDI Data.

Keywords: NSDI, risk modelling, Bayesian network, power transmission network, spatial modelling

1. INTRODUCTION

Using the data of the national spatial data infrastructure in Ukraine for the study of risk assessment of critical assets is one of the most important applied problems. Relevant and comprehensive spatial data on climatic conditions, engineering networks, and accident statistics should be provided by mapping services and local management companies for selecting support solutions and predicting emergencies.

The development of the spatial data infrastructure began with the preparation of the legal framework for the functioning of SDI in Ukraine in 2008. From that point on, a set of reforms that were and are carried out in the country transformed the understanding of spatial data infrastructure and created the mechanisms of institutional and technical implementation. The largest projects in this area should include the project of the World Bank regarding the establishment of the cadastral information system of Ukraine. The project was completed in 2012 and its key result was the cadastral map of Ukraine as the interface to the users and the provision of services. In addition, a number of data geoportals regarding administrative-territorial structure, agrarian sector, and natural resources were developed in the country. The Canadian government is supporting the development of Ukrainian SDI in the framework of the project "Laying the Foundation for a Spatial Data Infrastructure: Building Capacity within the Ukrainian Government to Support Sustainable Economic Growth", the Japanese government provides technical assistance to Ukraine, JRS and EuroGeographics by conducting trainings and seminars in Ukraine. The ecosystem of organizations has formed in Ukraine because of these efforts, which includes support and development of SDI on the national and regional levels. The main organization in the field of SDI is the State Service of Geodesy, Cartography and Cadastre, which consists of a number of subsidiaries. The main direction of political development is the integration with European organizations in the field of SDI and data harmonization under the INSPIRE directive.

Energy security manifests itself in the ability of the country to ensure the most reliable, technically safe, environmentally acceptable and reasonably sufficient energy provision for the economy and the population under the current and forecasted conditions (Shevtsov et al., 2002; Barannik, 2008). Safe functioning of the electric power system is one of the most important factors of energy security. Naturally, security is a complex political, economic, socio-economic, and scientific problem, which requires complex research of a wide range of issues.

The electric power system consists of a great number of objects and entities. The key ones are the power generating facilities and transmission networks. Hazardous situations (accidents) in the electric power system objects are usually caused by defects in the manufacture and operation of the equipment, personnel errors, and other factors leading to the forced termination of electricity supply, which constitutes a threat to the society.

The majority of the electric power transmission networks are the overhead power lines, therefore there is a threat of adverse climatic factors impact on the power transmission network components. Extreme climatic conditions lead to power lines accidents, so analysis of climate impacts on the power transmission network and prediction of consequences of such impacts are an integral part of the system security problem.

Accident statistics shows that more than a half of the failures of overhead power lines are caused by the ice and wind overloads on wires, cables and other structures (Horokhov et al., 2005). In general, they are the result of an underestimation of the actual ice and wind loads. The essential difference between overhead power lines and other power system objects is their greatest length. The total length of the transmission lines with voltage of 0.4–110(150) kV is 817.9 thousand km (Power Grid Technology Policy in Construction of High Voltage Distribution Networks, 2011).

Modeling of overhead transmission lines accidents caused by climatic factors and further forecasting of the number of accidents will be addressed in the article. The study of previous researches in the field of probabilistic accidents modeling shows that there are numerous methods of modeling, and regression analysis is used the most often. Currently, data mining methods that include neural networks, clustering algorithms, pattern recognition methods, a nearest neighbor methods, are used.

The paper addresses the development of the accidents model of the power system of Ukraine, namely overhead power lines, under the influence of climatic impacts as part of the security of the electrical energy system of Ukraine.

Current studies of impact of climatic factors on overhead power lines are limited to calculations of load of climatic factors on the overhead transmission lines, so the problem of conducting a comprehensive study of accidents probability under the influence of climatic factors is important. The novelty of the task lies in application of the existing mathematical tools to study accidents in the electric power systems under the influence of climatic factors and in use thereof to forecast the state of the electric power system of Ukraine.

2. PROBLEM STATEMENT

Aging of transformer substations equipment and constituent elements of transmission lines, and deteriorating climatic conditions in Ukraine lead to higher accident rate and energy losses in the power system equipment, causing an increased number of shutdowns and failures, most of which occur in the 0,4–10 kV network (Power Grid Technology Policy in Construction of High Voltage Distribution Networks, 2011). The most crucial in terms of the scale of accidents caused by climatic factors are ice formations and wind load. According to 2.5.30 (Rules for Electrical Installation, 2006), values of ice and wind loads and climatic impacts for calculation and selection of overhead power lines design are taken from the regional climatic zoning maps of Ukraine, with further adjustment based on data gathered by meteorological observation stations and observation posts regarding wind speed, ice intensity and density, temperatures, storm activity and other climatic phenomena.

2.1. History of the Problem

In recent years the number of power lines accidents increased significantly. A major accident in the Odesa region, accidents in the Zhytomyr and Volyn regions and others pointed out the necessity to review the impacts of climate loads on the power system in Ukraine and to create new zoning maps for climatic impacts.

Climatic zoning maps are maps of the territory divided into zones (areas) in terms of climatic impacts. In the latest version of Rules for Electrical Installation (2006) it was decided to use a scale with 7 areas of loads. The pre-existing methods of climatic zoning mapping envisaged consideration of phenomena with 10 years repetition period, but in the course of further studies it was found out that periods of recurrence of many phenomena were 11–13, 34–37 years (Climatic data for determining loads on overhead transmission line, 2008), therefore it was decided to take into account phenomena with 10, 25, 50-year recurrence periods. The new technique (Climatic loads, 2007) contains several innovations: weather station selection algorithm, prior data processing approach, form of the maxima distribution function, and observation data analysis approach.

2.2. Reasons for the Study

A program of scheduled transmission line reconstruction is currently being developed in Ukraine, so the problem of accident prevention arises. Ice and wind loads often cause power lines accidents. Because of the considerable weight of ice formations on wires, accidents also involve cascading collapse of towers. The main period of ice formation and sediment is the cold season, so interruption of power supply leads to the cessation of heating, and the restoration of power lines after accidents becomes much more difficult. The main method of preventing ice-caused accidents is a method of melting ice formations on wires (The Development of a US Climatology, 2002), which requires accurate knowledge of the ice deposition sizes, rate of formation and regional distribution thereof, so application of this method is considerably limited.

Thus, there is a problem to determine climatic conditions that lead to accidents. There is also a pressing issue to track ice distribution, communication and processing of real-time data to determine the need to run the ice melting process.

2.3. Input Data

To research overhead transmission lines accidents caused by climatic factors, the following input data are used:

- observations of weather stations of Ukraine regarding ice and wind events for the 1961–2012 period (Meteorological reviews, 2015) obtained in the Central Geophysical Observatory;
- geographical database of basic layers of Ukraine from the NSDI database
- x, y, z coordinates of the meteorological stations in Ukraine from the Meteorological Service.

- statistics of accidents (Development of Measures to Achieve Reliability of Power Networks under the Influence of Ice and Wind Loads in the Territory of Ukraine by Regions, 2009).

These observations of ice and wind events consist of data and characteristics of observations: the coordinates of the weather station, case number, type of sediment, beginning of icing (starting date, day, time of onset), growth duration of the icing event, the deposition parameters (diameter, thickness, weight), meteorological data at the beginning of the icing, and after reaching the maximum size (temperature, wind speed and direction). The period of ice observation runs from early autumn to late spring. Observations of wind speed include: characteristics of the weather station, wind speed (average and maximum), incidence rate by speed (number of cases for each value of the wind speed). The period of observation of wind speed lasts all year round. Data on accidents are presented in the following form: name, region, energy system district, name and characteristics of the line, date of the accident, absolute altitude of the accident area, characteristics of sediments, wind characteristics at the moment of failure, icing and wind observation data from the nearest weather station. In the above form, the accident rate data are available for the 1961–2015 period.

2.4. The Mathematical Formulation of the Problem

Nowadays studies of climate impacts are of exclusively applied nature, i.e. methodologies for determining the load on the transmission line constructions. Probabilistic modeling of overhead power line accidents under the influence of climatic factors was not performed. Thus, the challenge of constructing a model of accidents under the influence of climatic factors allows to:

– identify the main places of accidents, investigate the frequency, the main period of their occurrence, and their quantitative characteristics;

– classify accidents and conditions of their occurrence by territorial and temporal characteristics;

– identify the main climatic conditions in which accidents occur and use such data in rapid response systems;

– identify other factors that influence the occurrence of accidents, seasonality, technical state of lines, influence of the terrain;

– predict accidents;

– create maps to display the current statistics and forecasts.

In addition to the above requirements, a model must meet the requirements of adequacy and sustainability.

From the mathematical point of view, the problem of constructing models of accidents represents a problem of processing a large bulk of statistical data. Since the original data is represented as time series, any statistical methods of analysis of numerical series can be used to analyze and process them. Typically, for processing such data, methods of regression, correlation analysis and predictive analysis are used. Model construction is the task of the accident image recognition, i.e. division of all data on climate impacts into two classes: those that led to the accident and those under which the accident did not occur, and the subsequent identification of the main features of the accident. The problem requires finding the major climatic factors that affect the accident, and other impacts (seasonality, technical condition, type of terrain, etc.).

2.5. The Mathematical Models of Research

The model is a simplified representation of the system operation in the real world and represents the mathematical expression of the phenomenon. The research includes modeling of transmission lines accidents under the influence of climatic factors, based on pattern recognition methods. It should be noted that the meteorological observations data are not evenly distributed, which limits the use of regression and correlation analysis for modeling of power lines accidents. The basic models that can be used for accident modeling are presented below.

2.5.1. Regression models

The basic regression models used in the accident analysis are the Poisson regression model and a negative binomial distribution (Bendat and Piersol, 1989). Because of the strong nonlinearity of internal connections between variables that affect the onset of an accident, it is proposed to use methods of mining of statistical data on accidents (e.g. neural networks) instead of regression methods.

2.5.2. Data mining

The main methods of data mining are neural networks and Bayesian methods. The purpose of such network types is to create a model that correctly holds correspondence between inputs and outputs, using historical data, and, based on this correspondence, is able to make conclusions about the results, even if the desired results are not set, i.e. carry out pattern recognition (Yasnytskyi, 2005).

Probabilistic neural network. This type of networks provides a general solution to the classification problem, following the approach developed in statistics and called a Bayesian classifier. The Bayesian theory takes into account the relative probability of events and uses a priori information to improve prognosis. Using a neural network has the following advantages: no need of information about data distribution; application for multidimensional nonlinear problems; transformation of variables in the process of calculation. However, this method has some drawbacks: excessive error minimization requires a lot of computational power; individual connections between output variables and outcomes are not identified, the model is a "black box".

Bayesian approach. The Bayesian approach represents a probability model that uses a priori and a posteriori probabilities with features to predict the probability of accidents using statistical data (McCollister et al., 2007; Ma, 2006). The Poisson distribution is commonly used. The main disadvantage of this method is that an accident probability distribution must be input, but accidents are not always distributed in accordance with a certain fashion.

Nearest neighbor algorithm. The nearest neighbor algorithm is a classification method in which the class of the unknown object is determined by comparing it to all known objects stored in the database of recognized objects (Zagoruiko, 1999). The degree of similarity between objects is determined by the function called the distance function. To recognize an accident, two different functions can be used: the Euclidean distance and the metric based on the difference between values.

3. BAYESIAN BELIEF NETWORK

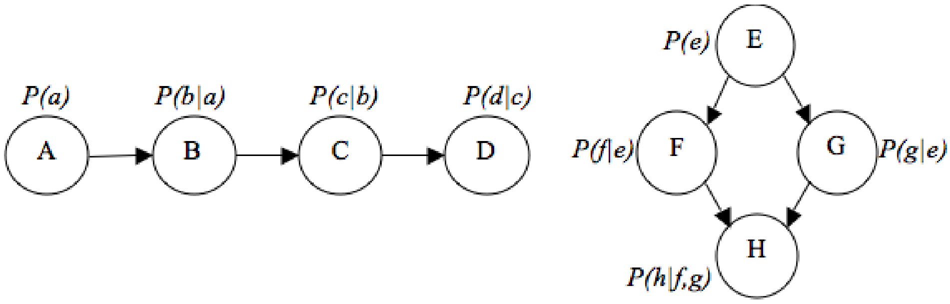
Most methods used to simulate failures require prior knowledge of the accident parameters distribution functions. In most situations, distribution data are not directly available, but instead, statistical dependency or independence between variables is known. In case of accidents, relation between the weight of ice and the number of accidents, dependency between wind speed and air temperature, and weight of ice deposits can be known. Internal connections can be represented by conditional probabilities that can be used to determine the probability of accidents under certain conditions.

3.1. The Best Option Choosing

The Bayesian network combines graphical structures (nodes representing variables and arcs expressing probabilistic dependencies between them) and corresponding conditional probabilities, which provides a comprehensive visual representation of different conditional relations between variables (Cheng et al., 2001). Local probability distributions are associated with each variable, and a set of independent conditions are represented in the network and can be directly combined to construct the overall probability distribution function for the entire network, which greatly simplifies the calculation of a posteriori probability variables.

The probabilistic structure is well illustrated (Duda et al., 2000). For example, it is necessary to determine the probability distribution for the variables d1, d2, ... to D (Fig. 1) using the table of conditional probabilities and the network topology.

Figure 1. Bayesian network structure



The distribution may be assessed by summing up the total general distribution, P (a, b, c, d) for all the variables except for d:

$$P(d) = \sum_{a,b,c} P(a,b,c,d)$$
$$= \sum_{a,b,c} P(a)P(b|a)P(c|b)P(d|c)$$
$$= \sum_{c} P(d|c) \sum_{b} P(c|b) \sum_{a} P(b|a)P(a) \qquad (1)$$

where

$$P(b) = \sum_{a} P(b|a)P(a)$$
$$P(c) = \sum_{b} P(c|b) \sum_{a} P(b|a)P(a) \qquad (2)$$

In this case, the network has a simple linear form; nonlinear networks (right side of fig.1) are calculated in a similar way.

3.2. Mathematical Method

There are two main types of Bayesian networks — simple Bayesian networks and Bayesian networks with conditional probabilities.

Simple Bayesian networks. This type of Bayesian Networks assumes independence between the variables of the model. Each variable must be directly related to the output variable. Despite the simple structure, these models are commonly used in practice. Let ω = (ω (1), ..., ω (n)) T be the vector that defines n states of the system, where ω (i) takes one of the C values $\omega_1, ..., \omega_S$, P (ω), the initial probability for n states of the system. Let X = (x_1, ..., x_n) be the matrix of features, which determines the feature vector X_i observed if the system is in state ω (i), p (x | ω) — this is the conditional probability of the density function for system states ω and feature set X. Using these notations, an a posteriori probability for the system ω is calculated as follows:

$$p(\omega|X) = \frac{p(X|\omega)P(\omega)}{p(X)} = \frac{p(X|\omega)P(\omega)}{\sum_{\omega} p(X|\omega)P(\omega)}. \quad (3)$$

Although (3) is a theoretical solution, in practice the calculation of (ω | X) can be a difficult task. If each variable ω(i) can take one of the values of C, it is necessary to consider C_n possible values of ω. Some simplifications can be made if the distribution of feature vector Xi depends only on one relevant system state ω(i) and independent of the other attributes and system states. In such a case the probability density p (x | ω) is calculated based on probabilities p (x_i | ω (i)):

$$p(X|\omega) = \prod_{i=1}^{n} p\big(x_i \omega(i)\big). \quad (4)$$

Bayesian networks with conditional probabilities. This type of networks is used for graphical representation of relations between variables in the model. The table of conditional probabilities is developed for future use in forecasting. If variable B depends on variable A, then A is called an ancestor to B and B — a descendant of A. The set of all ancestors *B* is designated as *parent(B)* or *pa(B)*. The quantitative assessment of parameters θ consists of the functions of distribution of conditional probabilities p (z_i / pa (Z_t)), which are required to determine the joint distribution p (Z_1, Z_2, ... Z_n). Without prejudice to everything else, we can assume that variables Z_1, Z_2, ... Zn are arranged in set G in such a way that pa (Z_i) $\subseteq$ {Z_1, Z_2, ... Z_n}. Using the following chain of rules, the joint probability distribution P (Z) can be represented as follows:

$$P(Z_1, Z_2, \dots, Z_n) = \prod_{i=1}^{n} P(Z_1, Z_2, \dots, Z_{i-1}). \quad (5)$$

The structure of G expressly defines a set of parameters θ, which is necessary to determine the joint distribution P (Z_1, Z_2, ..., Z_n), because:

$$P(Z_1, Z_2, \dots, Z_{i-1}) = P\big(Z_i \big| pa(Z_i)\big). \quad (6)$$

This follows directly from the above definition of Bayesian networks. Thus, the total distribution can be defined as follows:

$$P(Z_1, Z_2, \dots, Z_n) = \prod_{i=1}^{n} P\big(Z_i \big| pa(Z_i)\big). \quad (7)$$

Using a Bayesian network, a posteriori probabilities of some variables on the basis of initial probabilities of other variables are calculated. Thus, for a given set of attributes, $Z_i \subset Z$, which was created for the set of values z_i, the task is to compute a posteriori probabilities, $P(Z_2 \mid Z_1 = z_1, G, \theta)$ on the set of variables Z_2 within existing values Z_i.

Use of Bayesian networks as classifiers for prediction of a specific target variable (class variable) presents a particular interest. Thus, the problem of images classification (recognition) arises. The classification process uses variable P that can assume values c_1, c_2, ... cm, and feature vector Z defined as $\{z_1, z_2, \dots z_n\}$. If an instance of Z is represented by a set of attributes $\{z_1, z_2, \dots z_n\}$, the classification purpose is to determine class c_i, that includes Z. The network performance is measured on a set of test objects by calculating the classification accuracy, i.e. the percentage of tests for which the class is correctly identified by the network.

Formally, a Bayesian network consists of a qualitative description of network structure G and quantitative probability distribution θ, which is defined over the network structure. The network model G(N, A) represents the directed acyclic graph consisting of nodes N and arcs A, where $A \subseteq N \times N$. Each node I corresponds to a discrete random variable Z_i within a limited domain ΩZ_i. A Bayesian network represents a joint probability distribution $P(Z) = P(Z_1, Z_2, \dots Z_n)$.

Network arcs represent the relations between the dependent variables. The arc from node I to node J represents a probabilistic dependence between Z_i and Z_j, and can be accurately determined by using the concept of ancestral node. Ancestor Z_i, $pa(Z_i)$, Z_i is the direct precursor to the structure of G.

A Bayesian network is encoded by a set of statements that express probabilistic independence, making each variable Z_i conditionally independent of its descendants (in the structure of G), taking into account the state of its ancestors, $pa(Z_i)$. Definition of conditional independence can be effectively determined from the network structure using the graph theory.

The particularity of Bayesian networks is their ability to compactly encode a common probability distribution. The graphical structure of a Bayesian network model significantly improves clarity of a model, and various relations between attributes can be easily obtained. These features and the ability to encode causal relations of models make a Bayesian network a handy tool for accident modeling. Furthermore, this approach allows including knowledge into the network structure, especially in the form of causal information, which greatly simplifies the design of a Bayesian network and improves the understandability of the resulting model. The greatest advantage of Bayesian networks in comparison with other approaches is the possibility to use them as comprehensive solutions, as they can be easily used for decision-making and for selection of the next step (Nisbet, 2009).

3.3. Mathematical Accidents Model

To build a model of accidents caused by climatic factors, the following steps are required:

- identifying model variables and relations between them;
- building the structure of a Bayesian network and determining the possible values of variables and a priori probabilities based on equations;
- training a Bayesian network and refining its structure (variables and their probability);
- testing the model using data of accidents and meteorological observations;
- using the model for predicting the occurrence of accidents on the basis of information about accidents.

To simulate accidents and perform forecasting, the North-West area of the country is selected, namely the area of the Zhytomyr region, because that region is represented by the majority of climatic conditions, climatic zoning areas of ice and wind values and other climatic influences that occur in Ukraine (Meteorological reviews, 2015). Meteorological data of Zhitomir, Kyiv and Rivne regions are used for the analysis (Fig.2).

Figure 2. Map of Ukraine with selected areas for research

Definition of the variables for the model. Prediction of the accidents occurrence depends on collection and processing of data on accidents, including the selection of variables used in the structure model. The variables will be selected on the basis of knowledge about the possible causes of the accident. This includes a thorough analysis of literature and intuitive engineering knowledge.

To simulate overhead transmission lines accidents under the influence of climatic factors, it is necessary to research the statistical data on accidents and meteorological data. Table 1 presents the factors affecting accidents and their possible values.

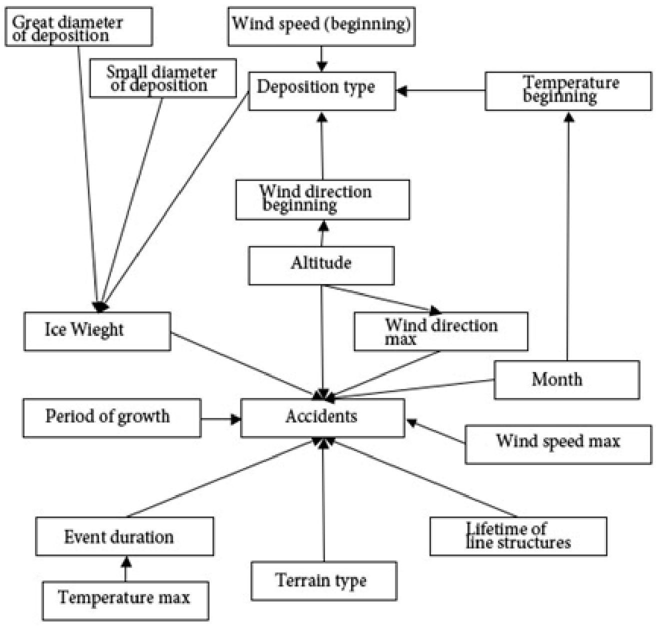
Table 1: Analysis of accidents

Factors	**Description**	**Possible values**
Year of accident	Observation period used for emergency situations 1960–2012 years	[1960 - 2015]
Month of accident	Ice and wind accidents often occur in cold seasons	1, 2, 3, 11, 12
Type of ice deposit	Ice and sediments type	ice, hoarfrost, crystalline frost, etc.
Ice growth period	Defines the period of ice growth in hours from the beginning of the ice formation to the maximum size of deposit	[1, 72]
Duration of the event	Assumes values from 1 to 150 hours	[1, 150]
Diameter	The diameter of ice deposit (large and small), divided into 10mm sectors.	[0, 10] ...> 72
Weight	Ice weight is measured in grams. Assumes value up to >256	[1, 16] ...> 256
Air temperature	Air temperature at the beginning of ice formation and at the moment it reaches the maximum size	[-25; 2]
Wind direction	Wind direction at the beginning of ice formation. Specified in the rhumbs	[1, 8]
Wind speed	Wind speed at the beginning of ice formation and after reaching the maximum size. Given in mps	[0; 25]
Extreme events	Cases of extreme values during a year	0, 1, 2

Year of structural elements installation	Determines the effects of physical state of the structures on accident occurrence	<1960 1960-1990 1990 - 2015 n/a
Classification	Contains 2 classes: event considered to be an accident or not	0,1

Accidents model. The main feature of Bayesian networks is the ability to specify relations between variables. To simulate the accident, a model (Fig. 3) was built on the basis of the initial accidents, and then in the process of training and testing some initial connections were changed.

Figure 3. The initial network of connections between factors



The network shown in Figure 3 presents nodes and arcs. The nodes represent model variables, whereas the arcs define relations between variables. As it is stated, occurrence of accidents in the network depends on the following factors: ice weight, event duration and period of sediments growth, type of terrain, lifetime, wind speed during the maximum deposition, month of the accident, altitude and wind direction at the beginning and at the moment when the maximum ice size is reached. The variables were selected by analyzing information on the accident rate.

About 20,000 events were used for simulation, 1,500 accident event records in the northwestern region of the country was processed. The initial probabilities for modeling were selected by analyzing the incidence of each variable of the model.

3.4. Mathematical Solution for Forecasting

Forecasting in a Bayesian network is based on the classical Bayesian classifier i.e. a statistically optimal classifier that minimizes the risk of misclassification (Duda et al., 2000). Any classifier attributes each observed data vector x to one of the predefined classes ωi, i = 1,2, ... n, where n is the number of possible classes. The effectiveness of many classifiers is limited by the number of data elements that vector x can contain, and by the number of possible classes n. The classical Bayesian classifier implements the Bayesian rule of conditional probabilities, for which probability P (ωi | x) that x belongs to class ωi is calculated as follows:

$$P(\omega_i|x) = \frac{P(\omega|c_i)P(c_i)}{P(\omega)} \quad (8)$$

where P (ωi | x) is conditional probability x of given set ωi, P (ωi) is the probability of getting data from class ωj and

$$P(\omega) = \sum_{j-1}^{n} P(x|\omega_j)P(\omega_j). \quad (9)$$

The Bayesian formula shows that observing the value of x it is possible to convert probability P (ωj) into an a posteriori probability P (ωj | x) — the probability that the system state is ωj, provided that the value of x was defined.

Suppose there are two classes (the accident occurred or not). If there is observation x, for which P (w1 | x) is higher than P (w2 | x), the natural tendency is to expect that an accident will happen. Conversely, if P (w2 | x) in higher than P (w1 | x) we shall consider that an accident will not occur. Thus, to classify input vector x, it is necessary to fulfill the following condition:

$$P(x|c_i)P(c_i) > P(x|c_j)P(c_j) \quad (10)$$

Verification of the model. The Bayesian network model was tested using a cross-validation technique. In this method, the training set is divided into two groups — a set for evaluation (which is used to assess probabilistic models), and a set for checking (used to evaluate the performance of the constructed probabilistic model). The cross-validation method allows comparing the performance verification tests of the selected model on testing and training data sets.

4. THE SIMULATION RESULTS

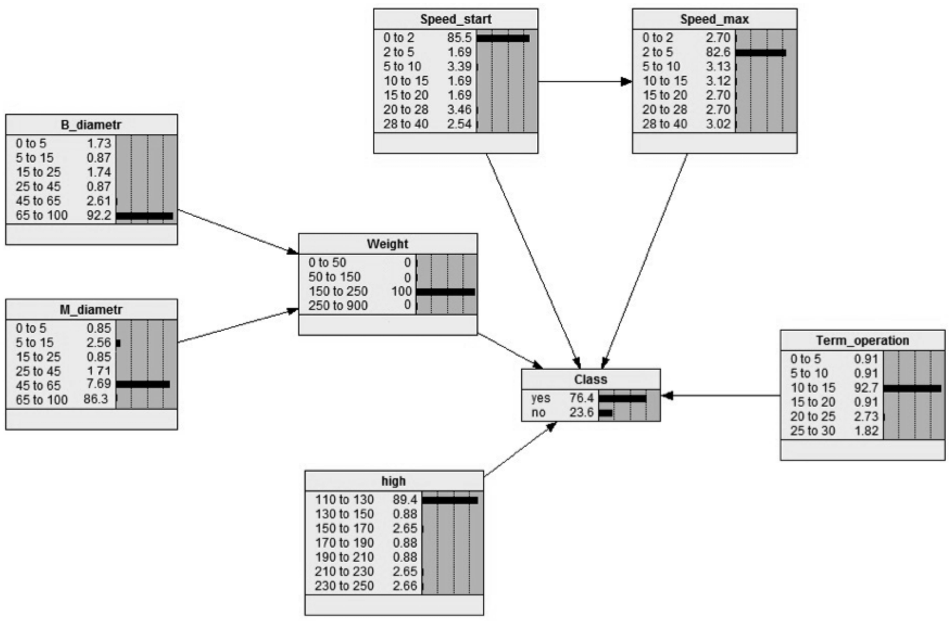
4.1. Test Examples and Model Tests

To simulate accidents, the following variables were used: ice weight, duration of the event and period of deposits growth, type of terrain, constructions lifetime, wind speed during the maximum deposition size, month of event occurrence, altitude and wind direction at the beginning and after reaching the maximum size of deposition. During the simulation, the initial relations between the variables of the model have changed. In particular, the factors that affect the accident rate were highlighted and other factors were discarded.

The remaining factors of the model and impact on accidents include: deposit diameters, ice weight, wind speed at the beginning of the event and after reaching the maximum size, altitude and lifetime of overhead lines constructions.

A Bayesian network defines different variables, dependencies between them, and conditional probabilities of these dependencies. BN can use this information to calculate the probability of various possible causes of the accident event. Conditional probabilities were calculated based on dependencies represented in the Bayesian network model that is shown in Fig. 8. The model shows that there are three nodes that require calculation of conditional probabilities. These are nodes C, F and H.

Figure 4. Bayesian network for prediction of overhead transmission line accidents



Node C: ice weight. The ice deposition weight for each case depends on the actual deposit parameters. The conditional probability for the weight of ice (C) and for large and small diameters (A and B) was calculated by formula (11).

$$P(C|A,B) = \frac{P(A,B,C)}{P(A)P(B)} \quad (11)$$

Node F: wind speed after reaching the maximum deposition size. Wind speed after reaching the maximum ice size depends on wind speed at the beginning of ice creation. The conditional probability for the wind speed at the ice maximum size (F), where the initial speed is (E), is calculated by formula (12).

$$P(F|E) = \frac{P(F,E)}{P(E)} \quad (12)$$

Node H: accident occurrence. The accident occurrence depends on the parameters shown in Fig. 4. The conditional probability of an accident is calculated by formula (13)

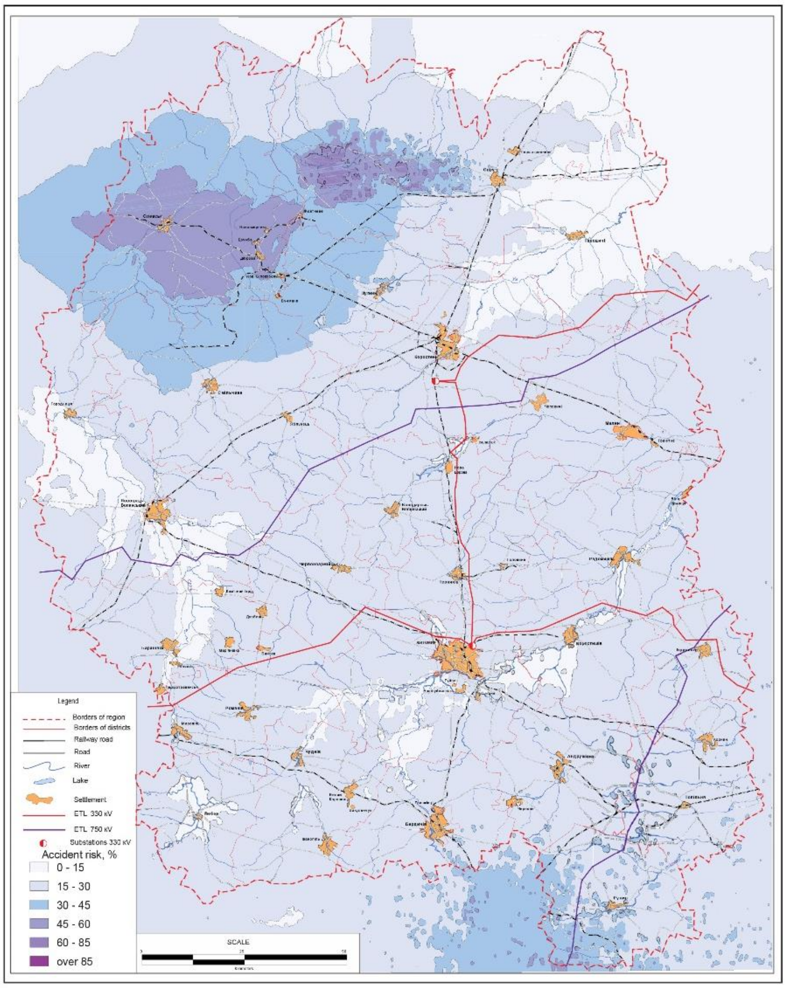
$$P(H|A \dots G) = \frac{P(A \dots H)}{P(C|A,B)P(E)P(F|E)P(G)P(D)} \quad (13)$$

4.2. Prediction Accuracy

The forecasting accuracy was 78%. Data used for forecasting were not included into the training set. Such prediction accuracy is quite high for the network. During forecasting, errors in determination of accident classes are not equivalent, i.e. a false statement that an accident will occur is less important than the false statement that the accident will not occur. Since the follow-up steps to prevent accidents are carried out by a power system management worker, a false alarm will not cause undesirable results. Therefore, error class definition is considered separately (class 1 — accident and class 0 — no accident). The above prediction accuracy corresponds to the lowest precision — error of non-recognition of an accident.

For the network training, a data set of 5,000 records was used, including 2,400 records that constituted a training sample, and 1,200 records were used for cross-testing of the model, and other records — for general model testing.

Figure 5. Map of accident risks in Zhytomyr region



5. DISCUSSION

The increase in the number of cases affects to the distribution of probability weights that the expected event occurs. Therefore, the simulation of results are very dependent from the selection volume and accumulation of new cases. Results obtained using a Bayesian network can be considered of valid only for a predetermined space-time interval. Shifts in climatic conditions enable the distribution simulation results only in a limited area with similar conditions. The methods of spatial multi regression more may be suitable for a large variability in the changes of factors possibly.

6. CONCLUSIONS

The paper addresses the research of overhead power lines accidents in the power systems of Ukraine under the influence of climatic factors. The article presents the construction of a model of accidents under the influence of climatic impacts and prediction of emergencies onsets based on data from NSDI. Pattern recognition techniques, namely the Bayesian network, were used to simulate accidents. This method is based on calculation of a posteriori probabilities of model variables. As a result, a model of accidents under the influence of climatic factors was built, which constitutes a Bayesian network with given conditional probabilities and independent variables of the model.

The following key results were obtained using the accident event prediction model: wind speed and weight of ice deposits are the main causes of accidents among the climatic factors, and other influences (among the selected variables) are lifetime of line structures and altitude; two groups of characteristics that cause accidents are revealed: large ice mass in the absence of strong wind pressure, and strong wind load with medium (in terms of the model) deposition weight. The predicted places of accidents occurrence coincide with the areas where accidents occurred in 80% of cases.

The model allows to predict the onset of an accident. The maps created based on SDI data correspond to the events of accidents that were not used for model construction, but only for validation thereof. For the further development of the study, the following steps are proposed: inclusion into the model of other variables and relations between them to facilitate consideration of more factors and improve the prediction accuracy; extension of the results obtained in the study to all the regions of Ukraine, considering topography and climatic and other factors. The latter field of work requires additional research in climatology, because climatic conditions of certain parts of the country are exceptionally particular (the Carpathian regions). Further studies require application of state-of-the-art techniques in the field of restoration of damaged and missing information (Neill at al., 2009) and multidimensional discretization strategies that will reduce information losses, thus increasing the effectiveness of the models using the spatial database from SDI warehouses.

REFERENCES

Barannik, V. (2008). *Comprehensive Evaluation Technique and Ways to Ensure Energy Independence*, PhD Thesis, Kyiv.

Bendat, J. and A. Piersol (1989). Random Data: Analysis and Measurement Procedures, Moscow: Mir.

Cheng, J., and R. Greiner (2001). Learning Bayesian Belief Network Classifiers: Algorithms and System, London, Springer-Verlag: 141-151.

Climatic Data for Determining Loads on Overhead Transmission Lines. Methodology of Processing. (2008). GRIFRE, (Regulation of the Ministry of Energy of Ukraine).

Climatic Loads on Overhead Transmission Lines Based on Topographic Features. Methodology (2007), GRIFRE, (Regulation of the Ministry of Energy of Ukraine).

Duda, O. and P. Hart (2000). *Pattern Classification (2nd Edition.),* John Wiley and Sons, New York.

Development of Measures to Achieve Reliability of Power Networks under the Influence of Ice and Wind Loads in the Territory of Ukraine by Regions. Book 1 (2009). “Selenergoproekt”, Kiev, 110 p.

Horokhov, E.V., M.I. Kazakevich, S.V. Turbin, and Y. Nazim (2005). Ice and Wind Impacts on Overhead Transmission Lines, Donetsk.

Ma, J. (2006). Bayesian Multivariate Poisson Regression for Models of Injury Count, *Statistical Methods and Crash Prediction Modeling*: 34–59.

McCollister, G. and C. Pflaum (2007). A Model to Predict the Probability of Highway Rail Crossing Accidents, *Journal of Rail and Rapid Transit*, # 221, pp. 321 - 329.

Meteorological Reviews: Republican Hydro-Meteorological Administration of Ukraine, 1961–2015 (2015). Kyiv.

Neill, D., G. Cooper, K. Das, X. Jiang and J. Schneider (2009). Bayesian Network Scan Statistics for Multivariate Pattern Detection, Statistics for Industry and Technology: 221-249.

Nisbet, R. (2009). Handbook of Statistical Analysis and Data Mining Applications, Pacific Capital Bankcorp NA Santa Barbara.

Power Grid Technology Policy in Construction of High Voltage Distribution Networks. Part I, Technical Strategy in Construction and Operation of Electrical Distribution Networks (2011). Separated subdivision "Scientific and Technical Center of Energy Systems”, Ukrenergo Kyiv.

Rules for Electrical Installations. Chapter 2.5 “Overhead Power Lines with Voltage over 1 kV and up to 750 kV” (2006). Official Journal. Kyiv: -GRIFRE: Ministry of Fuel and Energy of Ukraine.

Shevtsov, A., M. Zemlynicnuy, V. Barannik (2002). Ukraine's Energy Security: Strategy and Mechanisms, Dnipropetrovsk.

The Development of a U.S Climatology of Extreme Ice Loads (2002). US Department of Commerce NOAA / NESDIS National Climatic Data Center Asheville, (Technical Report).

Yasnytskyi, L. (2005). Introduction to Artificial Intelligence. Moscow: Publishing Center "Academy".

Zagoruiko, N. (1999). Applied Methods of Data and Knowledge Analysis, Novosibirsk: Mathematics.