

**MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
SUMY STATE UNIVERSITY
UKRAINIAN FEDERATION OF INFORMATICS**

**PROCEEDINGS
OF THE V INTERNATIONAL SCIENTIFIC
CONFERENCE
ADVANCED INFORMATION
SYSTEMS AND TECHNOLOGIES**

AIST-2017
(Sumy, May 17–19, 2017)



**SUMY
SUMY STATE UNIVERSITY
2017**

Moving average method in epidemic process forecasting

P. E. Piletskiy, D. I. Chumachenko

National Aerospace University “Kharkiv Aviation Institute”, Ukraine, pavel1997pileckii@gmail.com

Abstract – In this study, the application of the moving average method to the forecasting of the epidemic process is considered. As an example, the method is used to predict the incidence of ixodes tick borreliosis. To automate the calculation of the forecast, a software package was developed. The accuracy of the results was verified on real statistical data on morbidity in the Kharkiv region.

Keywords – moving average methods, forecasting, epidemic process, ixodes tick borreliosis.

I. INTRODUCTION

Among the most widespread infectious diseases in the world, a group of natural focal is distinguished, the pathogens or carriers of which are directly related to their habitat [1]. One of the most common diseases of this group in Ukraine is ixodes tick borreliosis (ITB), a group of infectious transmissible natural focal infections transmitted by ixodes ticks. Clinically, the disease occurs with a predominant lesion of the skin, nervous system, musculoskeletal system, heart muscle and is characterized by a tendency to chronic as well as latent flow. In the United States, a specialized center for the control and prevention of diseases (CDC) in Atlanta annually registers more than 5,000 cases. In European countries, according to German scientists, the number of cases can be up to 8-10 thousand per year. Intensive morbidity rates for borreliosis in France are 39.4 per 100 thousand people, in Bulgaria – 36.6 [2].

The aim of given paper is to calculate the forecast of incidence of ITB.

II. SIMULATION OF EPIDEMIC PROCESS

Simulation should be considered as the most important tool for understanding the epidemic process. The wide use of the terms “model”, “mathematical model” and models themselves in epidemiology marks a step forward in comparison with the classical methodological guidelines. Simulation of the epidemic process has its own history, with its main pages are linked with mathematical models. Mathematical modeling in epidemiology is a formal description of the main elements of the mechanism of the epidemic process with the help of a system of relationships, formulas, functions, equations, etc [3]. Depending on how deeply described in the mathematical terms elements (factors, indicators) characterize the epidemic process, several classes of modeling are distinguished:

- a formal approximation, consisting in transferring knowledge of the mathematical description of outwardly similar phenomena from other areas (for example, wave oscillations) to the epidemic process;
- formal extrapolation (mainly incidence curves), which gives satisfactory results only if the factors forming the epidemic process under consideration are approximately constant;
- substantial modeling of an epidemic process with discrete or continuous flow.

Each of these modeling classes operates with its own specific set of mathematical tools that have certain limitations and indications for use. At the same time, models belonging to the same class provide a certain level of impact in the study of the epidemic process. Consequently, there is a close relationship between the task and the modeling method.

In epidemiology, modeling is used for research purposes, to predict the nature of the epidemic process and to determine the strategy for health services.

In the construction of the epidemiological model, there are several stages:

- establishment of the structure of the model based on the collected actual data on the parameters of the epidemic process (susceptibility, stability, incubation period, duration of the disease, bacteriocarrier, duration of immunity, etc.);
- mathematical formulation of the model;
- simulation on the computer of a number of variants of the epidemic process with the inclusion of various conditions that affect the spread of infection, with the purpose of choosing the optimal one.

The advantage of the modeling method is that it is possible to accurately track the sources and routes of transmission of intestinal pathogens acting in these conditions. Everything remains as in a natural epidemic process.

III. STATISTICAL METHODS OF FORECASTING

Statistical methods of forecasting are a scientific and educational discipline, the main tasks of which are the development, study and application of modern mathematical and statistical prediction methods on the basis of objective data; development of the theory and practice of probabilistic statistical modeling of expert methods of forecasting; methods of forecasting in a risk

environment and combined forecasting methods using jointly economical-mathematical and econometric (both mathematical-statistical and expert) models. The scientific base of statistical prediction methods is applied statistics and decision theory.

The simplest methods for reconstructing the dependencies used for prediction come from a given time series, that is, a function defined at a finite number of points on the time axis. The time series is often considered within the framework of a certain probability model, other factors (independent variables) are introduced, in addition to time, for example, the volume of the money supply. The time series can be multidimensional. The main tasks to be solved are interpolation and extrapolation. The method of least squares in the simplest case (a linear function of one factor) was developed by K. Gauss in 1794-1795. Preliminary transformations of variables may be useful, for example, logarithm. The method of least squares is most often used for several factors. The method of least modules, splines, and other extrapolation methods are used less often, although their statistical properties are often better.

Estimation of the accuracy of the forecast (in particular, using confidence intervals) is a necessary part of the prediction procedure. Usually, probability-statistical models for restoring dependence are used, for example, they construct the best forecast using the maximum likelihood method. Parametric (usually based on the model of normal errors) and non-parametric estimations of forecast accuracy and confidence limits for it (based on the central limit theorem of probability theory) are developed. Heuristics are also used, not based on probability-statistical theory, for example, the method of moving averages.

Modern statistical methods of forecasting also include models of exponential smoothing, moving average autoregressions, systems of econometric equations based on both parametric and nonparametric approaches [4].

In this paper, the forecasting process is implemented using the moving average method. This method is used to smooth and predict time series.

IV. STATISTICAL METHODS OF FORECASTING

Moving average method allows to reveal tendencies of change of actual values of parameter Y in time and to predict future values of Y [5]. The received model can be effectively used in cases if for values of predicted parameter the established tendency in dynamics is observed. This method is not so effective in cases when such a tendency is violated, for example, in natural disasters, military actions, public unrest, with a sudden change in the parameters of the internal or external situation (inflation level, commodity prices); At the radical change in the plan of activity of the company, suffering losses.

The main idea of the moving average method is to replace the actual levels of the time series under study by their average values that cancel out the random oscillations. Thus, as a result, we obtain a smoothed series of values of the parameter under study, which allows us to more clearly identify the main trend of its change.

When developing a prediction of a population of tick-borne mites, the moving average method, based on observations over 3 (or 4) previous years, is more effective than methods based on long-term observations (over 10 years or more). This is due to the fact that as a result of the application of 3-year. The moving average each of the 3 values of the indicator (over the three years) is responsible for one-third of the forecast value. With a 10-year moving average, the values of each of the indicators of the same last three years are only for one-tenth of the forecast.

Unfortunately, there is no rule that allows us to select the optimal number m of the moving average members. However, it can be noted that the smaller m, the stronger the forecast responds to the oscillations of the time series, and vice versa, the larger m, the more the forecasting process becomes more inertial. In practice, the value of m is usually taken in the range from 2 to 10. If there are enough elements of the time series, the acceptable value for m can be determined, for example, as follows:

- set a few preliminary values m;
- smooth out the time series using each given value m;
- calculate the average prediction error.
- select the value of m, which corresponds to the smallest error.

V. AUTOMATION AND PROGRAM REALIZATION

To automate the prediction of the incidence of ITB using C # programming language, a software package has been developed that allows calculating prognosis morbidity based on existing statistical data in real time. The basic configuration of the software package includes data from 2004 to 2015. The data for the years include intensive incidence rates per 100,000 population, the population's readability for tick bites, the results of studies of ticks removed from humans, the presence of *Borrelia* and the proportion of ticks infected by *borrelia* collected on the flag, the numbers of mites.

To start calculating the forecast, it is necessary to enter the years for which data are available (at least 3 years, inclusive) and for how many years ahead it is necessary to carry out the forecast (Fig. 1). Then you need to enter the data for each year or select the available values for the period 2004 – 2016 and carry out the forecast (Fig. 2). After entering the data for one year, you need to click on the "Add data" button. After filling in all the values for each year, you must click "Forecast" to complete the forecast.

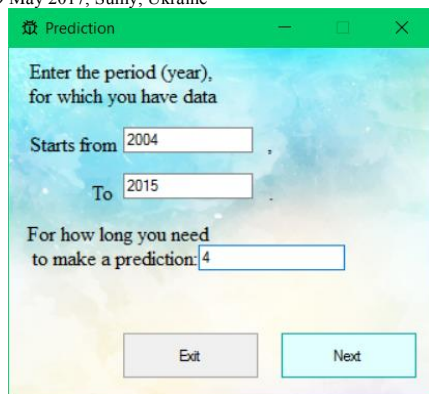


Figure 1. Initial window of the program

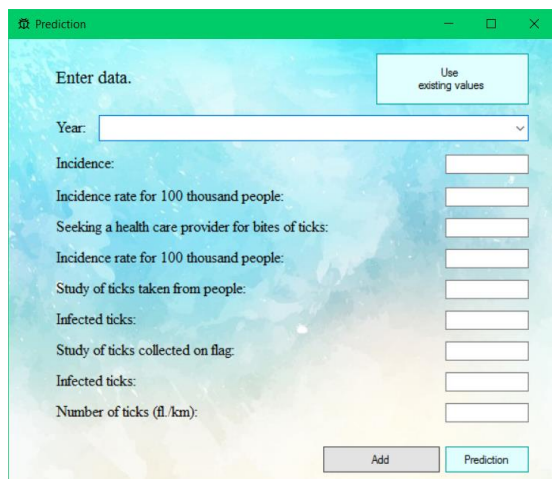


Figure 2. The data entry window for forecasting

Next, the software automatically calculates the forecast. The results are presented in the form of graphs. It is possible to look at the exact value by hovering the mouse over the point of interest.

VI. RESULTS

Based on the calculated prognosis, it can be concluded that there is a further increase in the number of cases of ITB with a certain three-year cycle: an increase in the number of cases of ITB within 2 years with a slight decrease for the following year. On the basis of the forecast, the further increase in the number of requests for medical assistance in connection with tick bites is traced. There is no clear cyclicity. The general tendencies of the growth of incidence of Lyme disease and the population's access to medical care for bites by ticks for 100 thousand people are traced. Based on the calculated prognosis, it is possible to assume a slight decrease and stabilization of the number of ticks in natural habitats. Figure 1 shows the relationship between the level of infection by borrelia ticks in nature and ticks removed from humans. As can be seen from the figure, there is no direct connection.

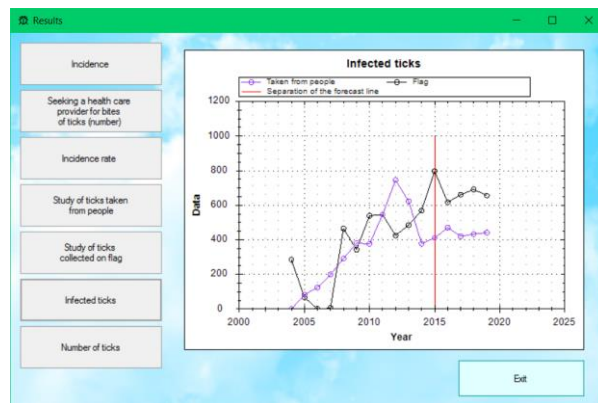


Figure 3. The level of ticks infection by Borrelia

The accuracy of calculated forecast has been verified on real statistics of morbidity by ITB in Kharkiv region.

CONCLUSIONS

In the framework of this study, an analysis of the incidence of ITB in the Kharkov region was carried out from 2000 to 2016. A software package has been developed that allows calculating the predicted incidence of ITB on the basis of the moving average method. The adequacy of the checked prognosis was verified on real statistics on the incidence of ITB.

Thus, the conducted research shows the continuing unstable epidemic situation with regard to the ITB, which dictates the need for a plan of preventive measures, the main goal of which is to reduce the incidence of people in the ITB. Virtual verification of the effectiveness of such events will be the next step in our study. Also a further area of the study is a quantitative assessment of the impact of external influences, such as public awareness, climate change, etc., on the incidence of ITB.

REFERENCES

- [1] M. Gulia-Nuss, A. B. Nuss and others, "Genomic insights into the Ixodes scapularis tick vector of Lyme disease", *Nature Communications*, Vol. 7, Iss. 10507, pp. 1–13, 2016.
- [2] *ICD-10. International Statistics Classification Of Diseases and Related Health Problems, 10th Revision, Volume 2: Instruction manual, 2010 Edition*, World Health Organization, 2011, 201 p.
- [3] Yu. Chernyshev, T. Chumachenko, D. Chumachenko, A. Tovstik, "System of Simulation of Epidemic Diseases Spreading", *Proceedings of East West Fuzzy Colloquium 2012 (19th Zittau Fuzzy Colloquium, September 5 – 7, 2012)*. Institut für Prozesstechnik Prozessautomatisierung und Messtechnik, 2012. – pp.154 – 161.
- [4] O. V. Yarovaya, D. I. Chumachenko, *Programming and numerical methods. Part 2: Numerical methods using MatLab and Mathcad: guidance manual for laboratory works*. Kharkiv: National Aerospace University «KhAI», 2017, 92 p.
- [5] S. M. Harmsen, Y. H. Chang, S. J. Hattrup, "Simple Moving Average: A Method of Reporting Evolving Complication Rates", *Orthopedics*, Vol. 39, Iss. 5, 2016.