Some Knowledge of Prediction Methods

Prediction plays an important role in people's lives. Broadly speaking we can say that predictions, prognoses and forecasts are used every day. Short-range weather forecasts enable us to make a decision on what to wear; farms and building companies plan their working activity according to the weather forecast for the following days, while the day of a space shuttle launch or a hurricane warning also result from weather forecasts. Long-term forecasts enable us to take measures to suppress negative phenomena, such as global warming. The evaluation of the future positions of vehicles on complicated unregulated cross-roads can also be called prognosis. Prognoses can be found in all areas of human activity, in which various methods, from the simplest to the most complicated, are used, employing theories of statistical probability, fuzzy logic, neural networks, or combinations of these.

Man has been interested in foreseeing the future since the time he started to think. The evidence for this are methods which are not based on scientific knowledge or mathematical procedures, such as foreseeing the future from cards, crystal balls or the like. The most famous persons in history in this respect include Nostradamus, Sibyl and Pythia.

The difficulty of making prognoses can be classified on an increasing scale from mechanical, physical, chemical, biological to social change. Most simple is the prediction of the position of a mathematical pendulum, which is given by exact and unambiguous mathematical calculation. The prediction of the position of a physical pendulum is more complicated, with friction and wearing on the bearing making prognoses more complicated. Predicting the behaviour of a person or society is the most complicated example of prognosis. Future share prices on stock markets and the political development of countries belong among the most complicated cases. The accuracy of prognoses made often falls the more complicated the subject of investigation becomes. The weather forecast, for example, is more than 60 % accurate.

Scientific prognosis is understood as being a conscious identification activity, where the aim is to find relations between the investigated phenomena and to predict the future of the investigated objects. During the creation of prognoses it is necessary to avoid two extremes. The first case represents the prediction of future development made by computer calculation without analysis of the influence of factors on the course of past development by man. The second case represents prediction of future development made by a person or group, where the creation of a forecast can be dependent on the wishes of this person or group. Both extremes may lead to erroneous prognosis.

It is appropriate to employ mathematical procedures to form accurate prognoses, which are objective and independent of the human influence. On the other hand, however, it is necessary perform analysis, as to whether the chosen method or model is the right one, and to then make corrections to the predicted values created.

Prediction methods based on objective mathematical and statistical calculations are called quantitative prediction methods, while methods based on expert opinions are called qualitative prediction methods. The first group includes, from the simplest methods to the most complicated, Gomparts' curve, moving averages, Winters' method, the Box-Jenkins methodology, fuzzy logic and neural networks, etc. The second group includes, for example, the method of the subjective smoothing curve or the Delphi method.

Warning prognoses are very important. Phenomena which have a negative influence can be discovered and successful action taken against them. A passive prognosis can simulate the condition which would arise in the future if conditions remain unchanged. A simulation prognosis can be created not only by the prediction model, but also by models of interconnections between phenomena and processes (there is a multitude of relations which have a positive or negative effect on future development). These various areas are represented by prediction maps.

Forecasts based on scientific laws, hypotheses and theories are called predictions. The following text will use this word exclusively. Scientific forecast, prognosis and prediction may be understood as synonyms only in the most general sense.

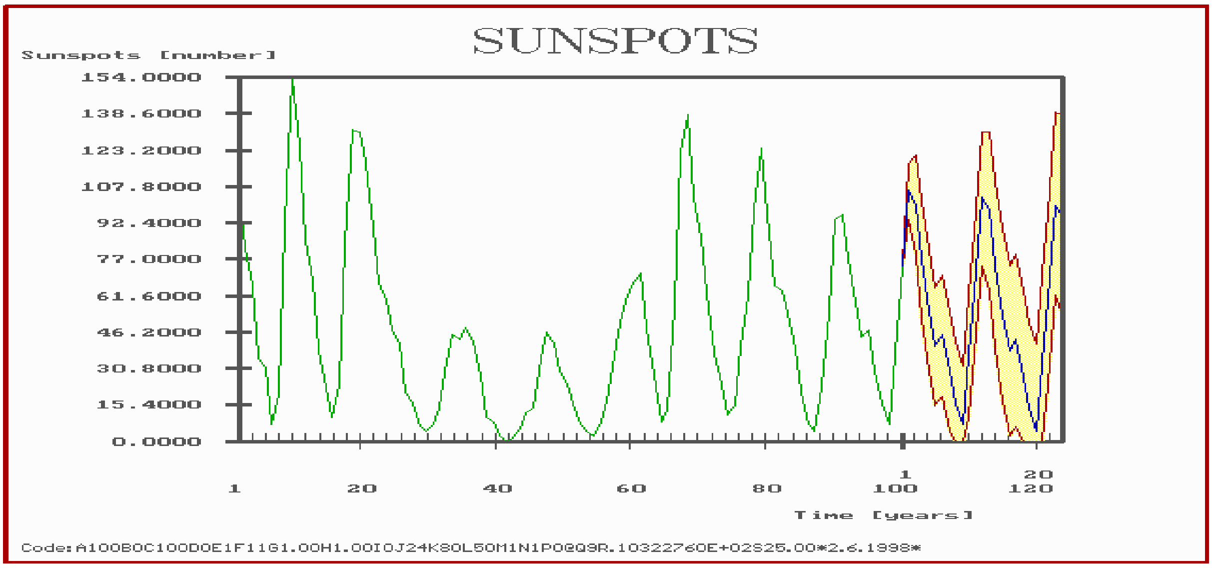
The Box-Jenkins methodology will be mentioned further on. This methodology works with time series, i.e. the sequences of measured values in time. The Box-Jenkins model for prediction can be written in the form

$$\Phi_P(B^S)\,\phi_p(B)\nabla_s^D\,\nabla^d\,z_t = \Theta_Q(B^S)\theta_q(B)a_t$$

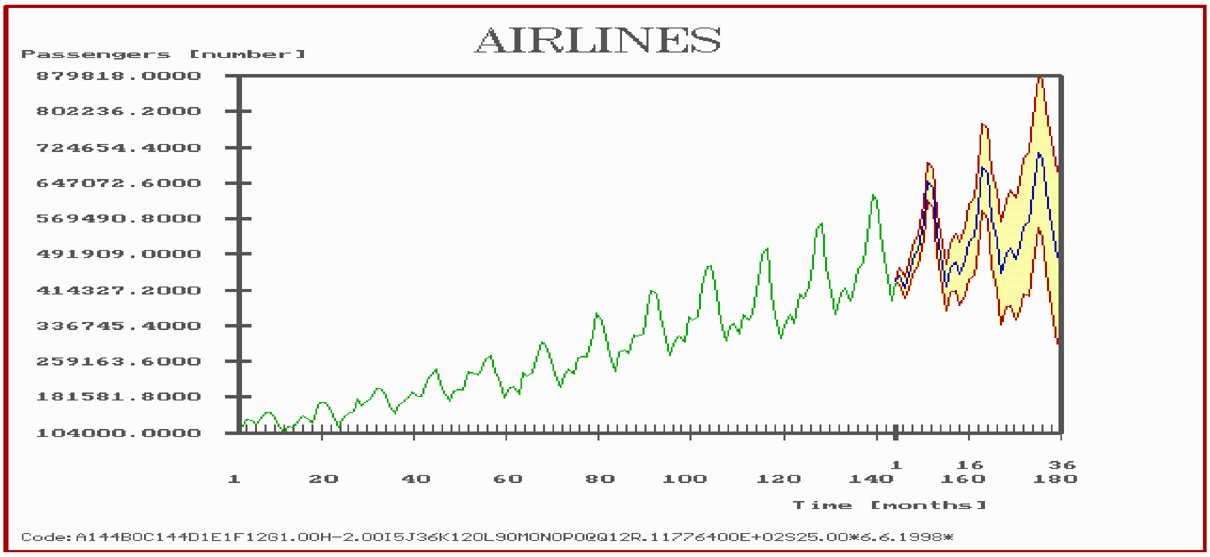
where the predicted value is calculated on the basis of the past development of the time series by means of auto-regressive integrated moving average models. Φ_P and ϕ_p are the seasonal and non-seasonal auto-regressive parameters, Θ_Q and θ_q are the seasonal and non-seasonal moving average parameters, B^S and B are the seasonal and non-seasonal backward operators, ∇_s^D and ∇^d are seasonal and non-seasonal differential operators, z_t is the time series values, a_t is a random error, and t is time. The Box-Jenkins methodology is described in article 5.

Time series generally consist of deterministic and stochastic components. The deterministic time series component can be further divided into a trend component (increasing, decreasing or unchanging) and a periodicity, repeating component (for example sinusoidal, rectangular, saw-tooth, etc.). The stochastic component is a random component, i.e. a component with is not repeated, has no trend and of which we often do not know the origin.

A prediction can be made for any time series in various fields. The most widely-known cases include the prediction of the activity of the sun (number of sunspots from 1770 to 1869, yearly sampling, 100 samples, prediction for 24 years) and the prediction of numbers of airline passengers (number of passengers from 1949 to 1966, month sampling, 144 samples, prediction for 3 years). In both these cases the deterministic component is evidently significant (periodicity component, the second picture also has an increasing trend component). The prediction is calculated on the basis of previous values and prediction values with, in addition to prediction values, a 95 % upper and lower limit drawn (a 95% probability limit means that 95 % of future values will fall inside this interval). As there is, in both these cases, a significant deterministic component the probability band is "narrow", and the prediction can be used in practice.



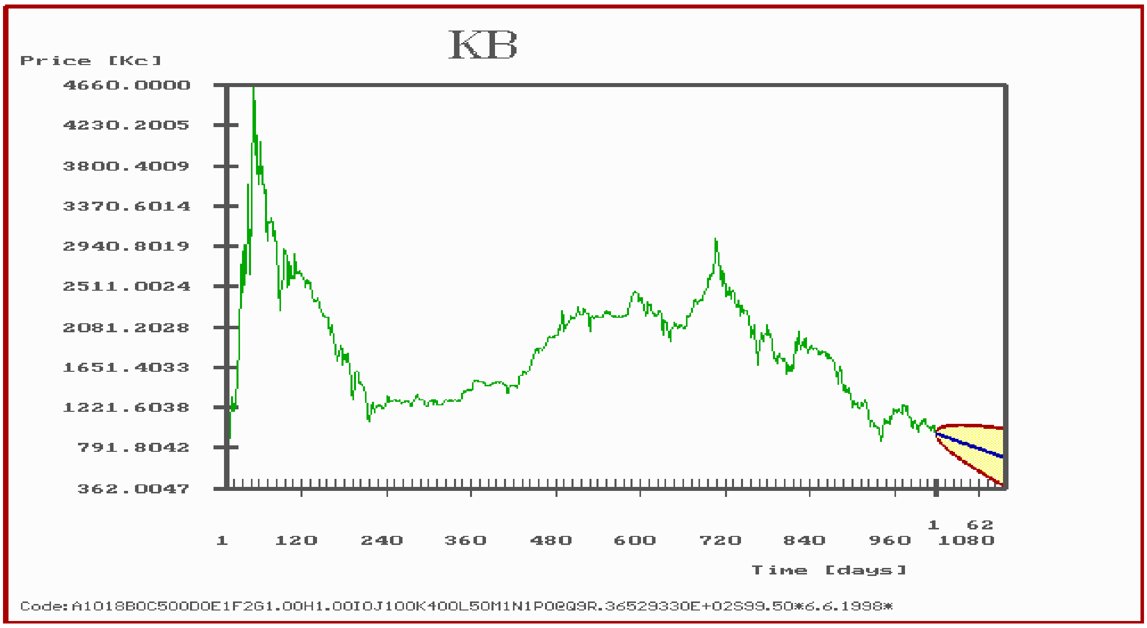
Prediction of the occurrence of sunspots



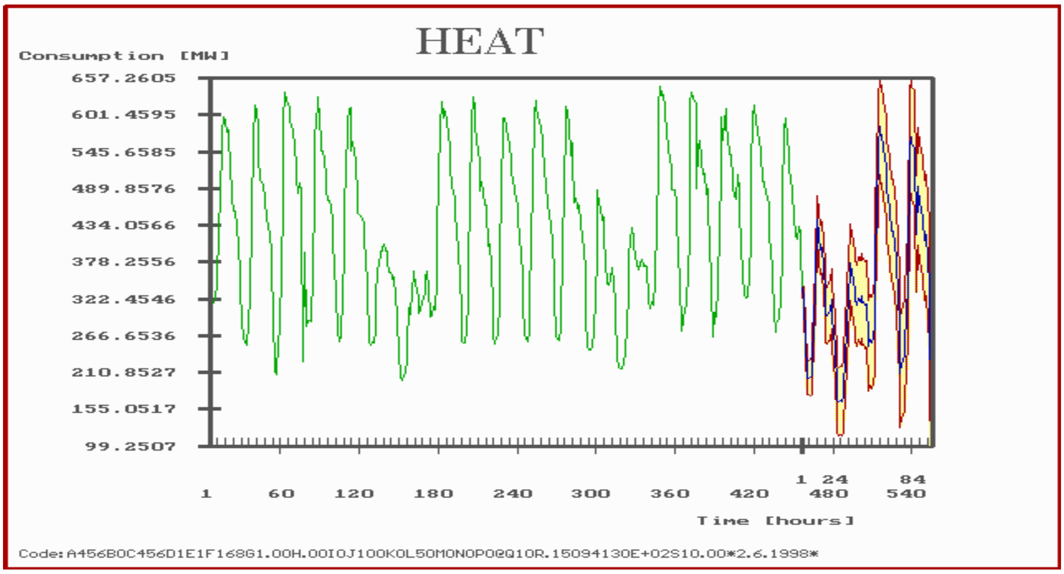
Prediction of numbers of airline passengers

Another example gives a predication of the price of Komercní banka shares (share price on the stock market in Czech crowns from 22.6.93 to 28.5.98, daily sampling, 1018 samples, prediction for 100 days). The values are taken from the beginning of trading in Komercní banka shares. This series shows a significant time series stochastic component and a trend component. As the stochastic component is significant the prediction of future values with a 95 % probability band is “wide“ and the prediction, therefore, does not produce results usable in practice. Predictions based on this method can be used merely to determine future trends.

Another example presents heat consumption predictions (heat consumption prediction for the city of Brno in MW from 1.00 a.m. on 6.2.78 to midnight on 22.2.78, hourly sampling, 456 samples, prediction for 100 hours). Different variants of use of the Box-Jenkins methodology are suggested to obtain more precise prediction. A method called superposition of models was proposed and is described in articles 9 and 10. This method makes prediction more precise and, therefore, its calculation gives good results even for a longer prediction time, and a “narrow“ probability band is calculated. This time series contains daily and weekly deterministic components and trend and stochastic components. The picture shows higher heat consumption on working days, lower consumption on Saturdays and lowest consumption on Sundays. The picture shows that the method of superposition of models reflects the mentioned phenomena. This prediction can be used for all time series of such types as the consumption of heat, electricity and water.



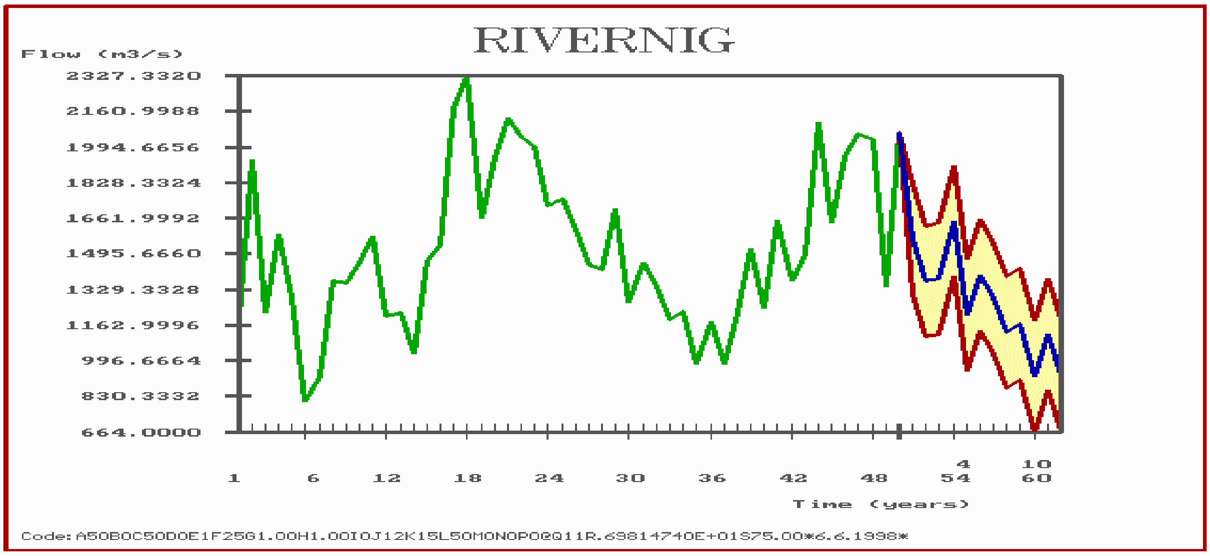
Prediction of the price of Komercní banka shares



Prediction of heat consumption in Brno

The next picture shows prediction of the flow of water in the Niger river in Coulicouro in Mali, (water flow in m^3/s from 1908 to 1957, yearly sampling, 50 samples, prediction for 12 years). The picture shows a significant periodic sinusoidal component and a stochastic component.

The Box-Jenkins methodology is suitable for time series prediction. Its main advantage over other methods is the small number of values necessary for prediction calculation, its rapid adaptability to changing conditions, and the good results of prediction of curves with a significant deterministic component. Its disadvantage can be seen from prediction of curves with a significant stochastic component, where the calculated band of probability disables the use of the results of prediction. The question is, whether prediction systems based on neural networks allowing adaptation using back-propagation, genetic algorithms or other methods will provide usable predictions.



Prediction of the flow of water in the Niger river

As has been mentioned above predictions can be made on any time series. The question remains as to whether the values produced by predication can be used in practice. Besides the given predictions (numbers of sunspots, numbers of airline passengers, the price of bank shares, heat consumption, the flow of water in a river) I could mention other examples of predictions, chosen at random, such as consumer prices, food prices, the prices of goods and raw materials, energy consumption, company turnover, economic development, changes to exchange rates, milk production in cows, television and radio viewing and listening figures, numbers of ill people, the popularity of politicians, population trends for the world and individual countries, criminality, air and water pollution, global warming, rainfall, etc. Interesting series for climatic prediction include the serial number of the first day of cherry tree blossom, the serial number of first day of lake freezing, the thickness between individual year rings of trees, and time prediction of the activity of geysers in Yellowstone park in the USA.

The use of prediction is evident from the given examples. Accurate prediction can uncover negative phenomena, and prediction may form part of a control system. It can be said that prediction enables us to make better decisions. Good prediction will bring higher quality, as we “see“ beyond the present, even at the price of a certain probability.

The methodology of statistical prediction methods described plays an important role in prediction, although investigation is necessary to discover whether other methods, for example neural networks, fuzzy logic and other nameless methods, whether simple or complicated, multi-purpose or single-purpose, provide better results than the procedure described. This is a continuous and complex process, which requires the work of many experts, co-operation between them and a mutual exchange of information.