

TIME SERIES ANALYSIS OF HEAT DEMAND

Bronislav Chramcov
Jaroslav Balátě
Tomas Bata University in Zlín
Faculty of Applied Informatics
Nad Stráněmi 4511, 760 05 Zlín, Czech Republic
E-mail: chramcov@fai.utb.cz

KEYWORDS

Prediction, District Heating Control, Box-Jenkins, Control algorithms, Time series analysis.

ABSTRACT

The paper deals with an analysis of time series of heat demand course. The course of heat demand and heat consumption can be demonstrated by means of heat demand diagrams. The most important one is the Daily Diagram of Heat Supply (DDHS) which demonstrates the course of requisite heat output during the day. This analysis is utilized for building up a prediction model of the DDHS. Forecast of this heat demand course is significant for short-term and long-term planning of heat production. This forecast is very important for technical and economic considerations. The solved prediction is especially determined for the qualitative-quantitative method of hot-water piping heat output control, i.e. the Balátě System which enables to eliminate any transport delay. Its importance is increased especially for controlled systems having great time constants and great transport delay. In this paper we have focused on determination of difference order, autoregressive process order and moving average process order.

INTRODUCTION

Improvement of technological process control level can be achieved by the time series analysis in order to predict their future behaviour. We can find applications of this prediction also in the control of the Centralized Heat Supply System (CHSS), especially for the control of hot water piping heat output. Knowledge of heat demand is the base for input data for the operation preparation of CHSS. The term "heat demand" means an instantaneous heat output demanded or instantaneous heat output consumed by consumers. The term "heat demand" relates to the term "heat consumption". It expresses heat energy which is supplied to the customer in a specific time interval (generally a day or a year). The course of heat demand and heat consumption can be demonstrated by means of heat demand diagrams. The most important ones are:

- **daily diagram of heat supply** which demonstrates the course of requisite heat output during the day. (See Figure 1)
- **duration heat demand diagram** - Y-coordinates represent heat demand and distance from zero

represents duration of corresponding heat demand. Daily and yearly duration heat demand diagrams are currently known.

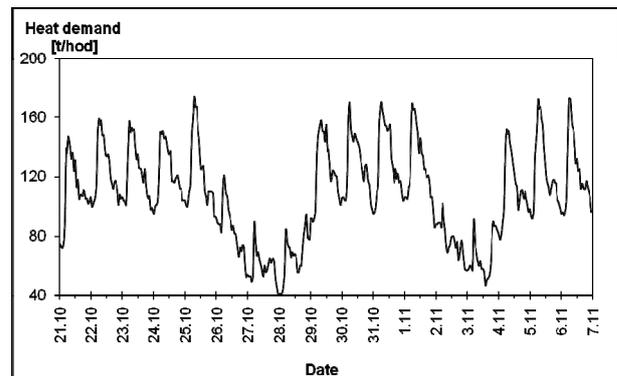


Figure 1: Time series of DDHS from the concrete locality

These diagrams are of essential importance for technical and economic considerations. Therefore forecast of the diagrams course is significant for short-term and long-term planning of heat production. It is possible to judge the question of peak sources and particularly the question of optimal load distribution between the cooperative production sources and production units inside these sources according to the time course of heat demand. The forecast of DDHS is used in this case (Balátě 1982).

FORECAST MODEL FOR CONCRETE TIME SERIES OF HEAT DEMAND

It is possible to use different solution methods for the calculation of time series forecast. (For example: solution by means of linear models, solution by means of non-linear models, spectral analysis method, neural networks etc.). In the past, lots of works were created which solved the prediction of DDHS and its use for controlling the District Heating System (DHS). Most of these works are based on mass data processing. Nevertheless, these methods have a big disadvantage that may result in out-of-date real data. From this point of view it is suitable to use the forecast methods according to the Box –Jenkins methodology (Box and Jenkins 1976). This method works with a fixed number of values which are updated for each sampling period.

This methodology is based on the correlation analysis of time series and works with stochastic models which enable to give a true picture of trend components and also that of periodic components. As this method achieves very good results in practice, it was chosen for the calculation of DDHS forecast.

Identification of time series model parameters is the most important and the most difficult phase in the time series analysis. This paper is dealing with the identification of a model of concrete time series of the DDHS. We have particularly focused on determination of difference degree as well as on obtaining a suitable order of autoregressive process and order of moving average process.

DETERMINATION OF DEGREE OF DIFFERENCING

Many observed non-stationary time series exhibit certain homogeneity and can be accounted for by a simple modification of the ARMA model, the autoregressive integrated moving average (ARIMA) model. Determination of a degree of differencing d is the main problem of ARIMA model building. In practice, it seldom appears necessary to difference more than twice. That means that stationary time series are produced by means of the first or second differencing. A number of possibilities for determination of difference degree exists.

1. It is possible to use a plot of the time series, for visual inspection of its stationarity. In case of doubts, the plot of the first or second differencing of time series is drawn. Then we review stationarity of these series.
2. Investigation of estimated autocorrelation function (ACF) of time series is a more objective method. If the values of ACF have a gentle linear decline (not rapid geometric decline), an autoregressive zero is approaching 1 and it is necessary to difference.
3. Anderson (Anderson 1976) prefers to use the behaviour of the variances of successive differenced series as a criterion for taking a decision on the difference degree required. The difference degree d is given in accordance with the minimum value of variances $\sigma_z^2, \sigma_{\nabla z}^2, \sigma_{\nabla^2 z}^2, \dots$.

An example of the determination of the difference degree for our time series of DDHS is shown in this part of paper. The course of time series of DDHS contains two periodic components (daily and weekly periods). The daily period presents increase and decrease in heat demand during the day. Heat demand drop at the weekend forms the weekly period of DDHS. The general model according to Box-Jenkins enables to describe only one periodic component. From this point of view we will further consider only time series of the DDHS without Saturday and Sunday values. This time series (see Figure 2) exhibits an evident non-stationarity. It is necessary to difference. The course of time series of the first differencing is shown in Fig. 3. The differenced series looks stationary now.

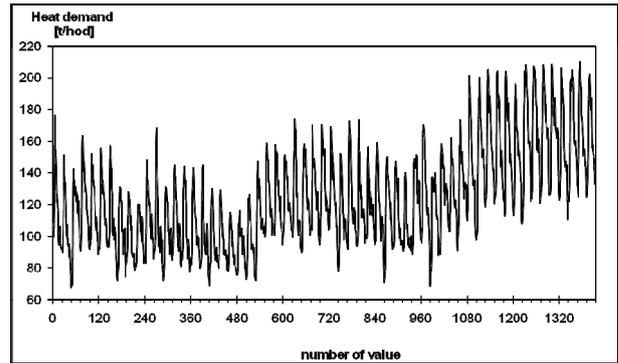


Figure 2: The course of DDHS without values of Saturday and Sunday

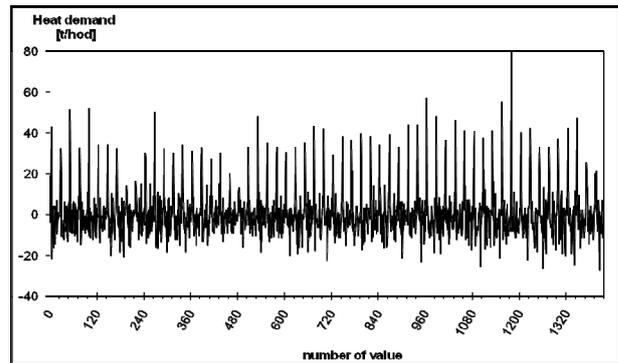


Figure 3: The course of DDHS after the first differencing

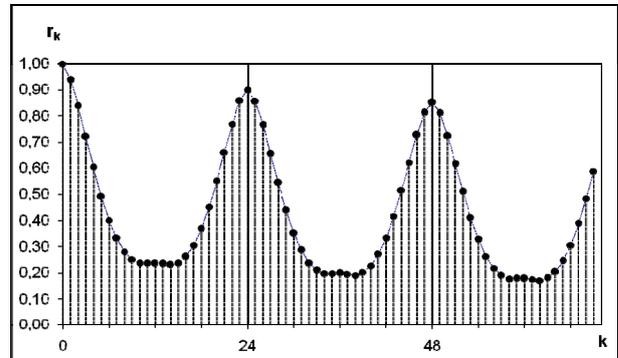


Figure 4: The course of estimated ACF of DDHS

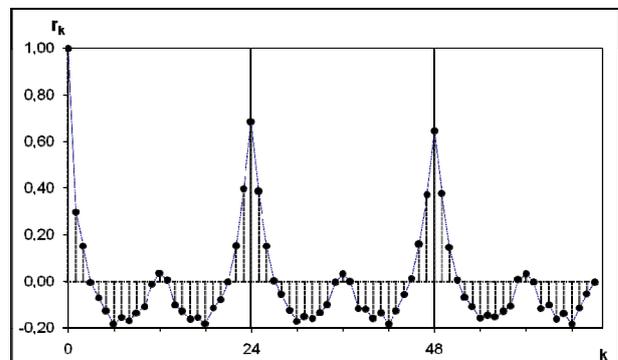


Figure 5: The course of estimated ACF of the DDHS after the first differencing

This fact is confirmed by the sample ACF (see Figure 4 and Figure 5). It is also possible to use the estimated variance of non-differenced and differenced series for making a comparison. In our case we have obtained the following results (1):

$$\hat{\sigma}_z^2 = 967.7; \hat{\sigma}_{\nabla z}^2 = 119.9; \hat{\sigma}_{\nabla^2 z}^2 = 168.2. \quad (1)$$

From these results and Figures above, it is suitable to choose the first differencing of the DDHS ($d=1$).

In practice, many time series have seasonal components. These series exhibit periodic behaviour with a period s . Therefore it is necessary to determine a degree of seasonal differencing - D . The seasonal differencing is marked by ∇_s^D . In seasonal models, necessity of differencing more than once occurs very seldom. That means $D=0$ or $D=1$. It is possible to decide on the degree of seasonal differencing on the basis of investigation of sample ACF. If the values of ACF at lags $k*s$ achieve the local maximum, it is necessary to make the first seasonal differencing ($D=1$) in the form $\nabla_s z_t$.

As our time series of the DDHS exhibits an obvious seasonal pattern, it is necessary to make an analysis of the seasonal differencing of our time series, as well. The course of the ACF sample evidences the seasonal pattern (see Figure 4 or Figure 5). These functions have their local maxima at lags 24, 48, That represents a seasonal period of 24 hours by a sampling period of 1 hour. On the basis of the executed analysis, it is necessary to make the first differencing and also the first seasonal differencing of the DDHS in the form (2).

$$\nabla \nabla_{24} z_t = z_t - z_{t-1} - z_{t-24} + z_{t-25} \quad (2)$$

The course of differenced DDHS in the form (2) is shown in Figure 6.

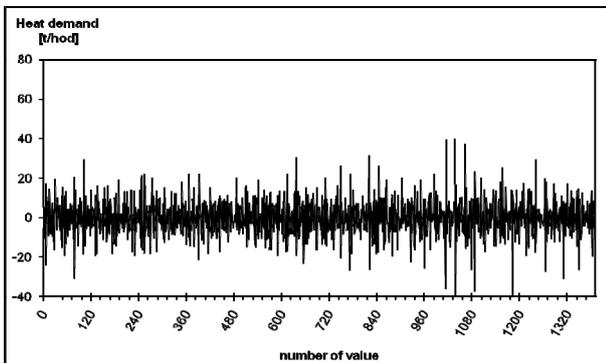


Figure 6: The course of DDHS after the first differencing and after the first seasonal differencing

For the comparison, it is possible to calculate the variances of time series and differenced series according to (Anderson 1976). The results are shown in the form (3).

$$\begin{aligned} \hat{\sigma}_z^2 &= 967.7; \hat{\sigma}_{\nabla z}^2 = 119.9; \hat{\sigma}_{\nabla^2 z}^2 = 168.2; \\ \hat{\sigma}_{\nabla_{24} z}^2 &= 150.5; \hat{\sigma}_{\nabla \nabla_{24} z}^2 = 72.7 \end{aligned} \quad (3)$$

DETERMINATION OF AUTOREGRESSIVE PROCESS ORDER AND MOVING AVERAGE PROCESS ORDER

After differencing the time series, we have to identify the order of autoregressive process $AR(p)$ and order of moving average process $MA(q)$. The traditional method consists in comparing the observed patterns of the sample autocorrelation and partial autocorrelation functions with the theoretical autocorrelation and partial autocorrelation function patterns. These theoretical patterns are shown in Table 1.

Table 1: Behaviour of theoretical autocorrelation and partial autocorrelation function

Model	ACF	PACF
AR(p)	Tails off	Cuts off after p
MA(q)	Cuts off after q	Tails off
ARMA(p,q)	Tails off	Tails off

The expression *Tails off* in Table 1 means that the function decreases in an exponential, sinusoidal or geometric fashion, approximately, with a relatively large number of nonzero values. Conversely, *Cuts off* implies that the function truncates abruptly with only a very few nonzero values. The standard errors of the ACF and PACF samples are useful in identifying nonzero values. As a general rule, we would assume an autocorrelation or partial autocorrelation coefficient to be zero if the absolute value of its estimate is less than twice its standard error. Order of autoregressive operator p and order of moving average operator q are not usually high; therefore only 20 values of ACF and PACF will be enough to compute. The ACF and PACF samples for the time series of the DDHS are shown in Figure 7 and Figure 8. For assistance in interpreting these functions, two-standard-error limits are plotted in the graphs. We can see from Figure 7 that the ACF sample cuts off after one lag, while the PACF sample tails off with a sinusoidal decline. Thus, we would tentatively identify the underlying model of our series to be a stationary MA(1) model.

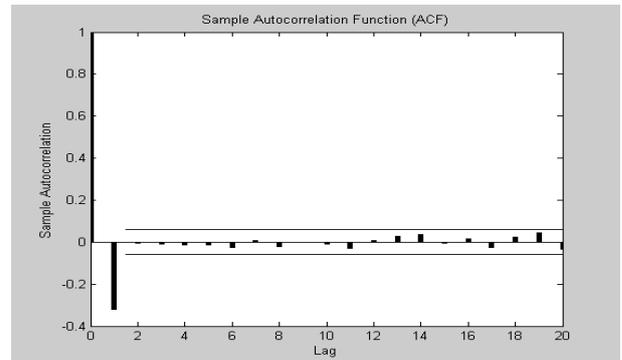


Figure 7: The course of ACF of differencing DDHS

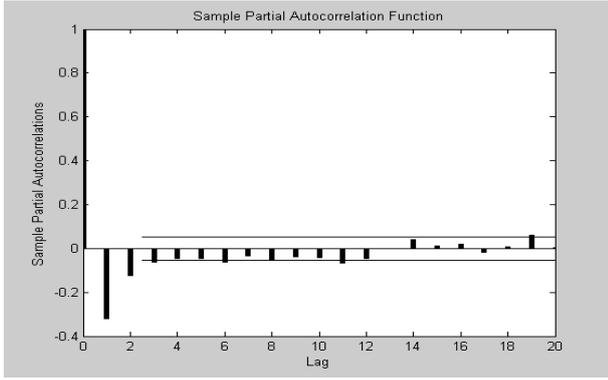


Figure 8: The course of PACF of differencing DDHS

The order of model is usually difficult to determine on the basis of the ACF and PACF. This method of identification requires a lot of experience in building up models. From this point of view it is more suitable to use the objective methods for the tests of the model order.

Criteria for model order

A number of procedures and methods exists for testing the model order (Cromwell 1994). These methods are based on the comparison of the residuals of various models by means of special statistics. In this paper, the Akaike Information Criterion (AIC) in the form (4), Bayesian Information Criterion (BIC) (5) and Schwarz test in the form (6) are used for testing.

$$AIC(p, q) = n \cdot \ln \hat{\sigma}_a^2 + 2(p + q) \quad (4)$$

$$BIC(p, q) = n \cdot \ln \hat{\sigma}_a^2 - (n - p - q) \ln \left(1 - \frac{p+q}{n} \right) + (p + q) \ln n + (p + q) \ln \left(\frac{\frac{\hat{\sigma}_z^2}{\hat{\sigma}_a^2} - 1}{p + q} \right) \quad (5)$$

$$SC(p, q) = n \cdot \ln \hat{\sigma}_a^2 + (p + q) \cdot \ln n \quad (6)$$

Where n is a number of residuals, $\hat{\sigma}_a^2$ is an estimate of the residual variance, $\hat{\sigma}_z^2$ is the variance associated with z_t .

The test results of the model order of the DDHS are included in the next tables (see Tables 2, 3 and Table 4). The tables represent the values of the particular criteria in dependence on model order (p, q).

The minimum value of BIC and SC is for $p=0$ and $q=1$, while the minimum value of AIC is for $p=2$ and $q=1$. AIC for $p=0$ and $q=1$ is very close to the minimum value. According to these results and on the basis of the general theorem (small values of model order), it is obvious that we would tentatively identify our time series as the MA(1) process. This fact is also confirmed by the analysis of ACF and PACF samples described above. Adequacy of this tentative model may be examined by means of Portmanteau test.

Table 2: Values of Akaike Information Criterion

$p \backslash q$	0	1	2	3	4	5	6	7
0	1688	1647.657	1649.6	1649.2	1652.2	1650.2	1652.2	1654.2
1	1653.1	1649.6	4357.9	1651.7	2530.5	1652.3	1654.2	4409.1
2	1651.6	1647.655	1651.1	1792.6	1906.5	1654.2	4389.6	3979.6
3	1651.4	1652.4	2032.4	1942.2	2222.5	1655.6	1657.3	2262.8
4	1653.1	2391.6	1704.9	1659.4	1656	1657.3	3175.7	2304.2
5	1652.6	1654.1	1656.8	1675.3	1656.3	2486.3	1934.2	2043.5
6	1653.3	1655.9	1926.3	1749	2412.5	2049.4	2024.3	2984.5
7	1654.9	1657	1659	2469.5	2493.4	2027	3869	N/A

Note: Value N/A in the tables means that the criterion is not available.

Table 3: Values of Bayesian Information Criterion

$p \backslash q$	0	1	2	3	4	5	6	7
0	N/A	1654.7	1662.4	1667.2	1675.1	1677.7	1684.1	1690.4
1	1660.2	1662.4	4369.4	1674.6	2530.9	1684.2	1690.4	4431.7
2	1664.4	1665.6	1674	1818	1933.7	1690.4	4412.2	4003.9
3	1669.4	1675.3	2053.8	1968.7	2244.6	1695.9	1701.6	2289
4	1676	2403.4	1735.9	1695.4	1696.3	1701.5	3199.9	2329.6
5	1680.2	1686	1692.9	1715.2	1700.6	2490.4	1976.7	2084.3
6	1685.3	1692.1	1959.9	1790.8	2428.1	2087.4	2065.9	3010.4
7	1691.1	1697.3	1703.3	2477.3	2495.1	2068.5	3899.2	N/A

Table 4: Values of Schwarz Criterion Test

$p \backslash q$	0	1	2	3	4	5	6	7
0	1688	1651.5	1657.5	1661	1667.9	1669.8	1675.8	1681.7
1	1657	1657.5	4369.7	1667.4	2550.2	1675.8	1681.7	4440.5
2	1659.4	1659.3	1666.8	1812.3	1930.1	1681.7	4421	4015
3	1663.2	1668.1	2052.1	1965.8	2250	1687	1692.7	2302.1
4	1668.9	2411.2	1728.5	1686.9	1687.5	1692.6	3215	2347.4
5	1672.3	1677.7	1684.3	1706.8	1691.7	2525.6	1977.4	2090.7
6	1676.9	1683.4	1957.8	1784.4	2451.8	2092.6	2071.5	3035.6
7	1682.5	1688.4	1694.4	2508.8	2536.6	2074.2	3920.1	N/A

Diagnostic checking – Portmanteau test

If the fitted model is adequate, it should transform the observations to a white noise process. Thus, a logical method of diagnostic checking is to compute the residuals and then estimate and examine their autocorrelation function. If the model is appropriate, then the residual sample autocorrelation function should not differ significantly from zero for all lags greater than one. Adequate model can be tested by means of Portmanteau test. We may obtain an indication of whether the first K residual autocorrelation considered together indicate adequacy of the model. This indication may be obtained through an approximate chi-square test of model adequacy. We consider the test statistic in the form (7), which is approximately distributed as chi-

square with $K-p-q$ degrees of freedom if the model is appropriate.

$$Q = (n - d - s) \cdot \sum_{k=1}^K r_k^2(\hat{a}) \quad (7)$$

Where n is a number of residuals, d is the difference degree, s is the season period, $r_k^2(\hat{a})$ is value of ACF sample of residuals.

If the model is inadequate, the calculated value of Q will be too large. Thus we should reject the hypothesis of model adequacy if Q exceeds an appropriately small upper tail point of the chi-square distribution with $K-p-q$ degrees of freedom. Further details of this test are in (Box and Pierce, 1970).

The sample autocorrelation function of the residuals for the model MA(1) for our time series of DDHS is shown in Figure 9. None of the 20 autocorrelations exceeds twice their standard errors. Furthermore, the chi-square statistic applied to the first 20 autocorrelations ($K=20$) is $Q=20.1995$, with 19 degrees of freedom. Comparing this value with a 5 percent value chi-square variable with 19 degrees of freedom, we find out that $\chi_{0.05,19}^2 = 30.14$ and so we would conclude that there is no strong evidence to reject the model.

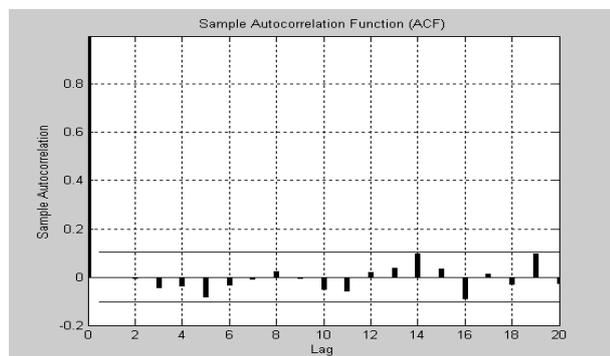


Figure 9: The ACF sample of residuals of DDHS

CONCLUSION

This paper presents the method for building up the model of time series of DDHS. The model is used for prediction of heat demand. This prediction of DDHS is necessary for the control in the Centralized Heat Supply System (CHSS), especially for the qualitative-quantitative control method of hot-water piping heat output – the Balátě System (Balátě 1982). The modelling is based on the Box-Jenkins methodology. The time series analysis was made for the DDHS from the concrete locality.

REFERENCES

Anderson, O. D. 1976. “*Time Series Analysis and Forecasting – The Box-Jenkins approach*”. London and Boston: Butterworths, 180 p., ISBN: 0-408-70926-X.

Balátě, J. 1982. “*Design of Automated Control System of Centralized Heat Supply*”. Thesis of DrSc (Doctor of Science) Work. Faculty of Mechanical Engineering of TU Brno, Brno.

Box G.E.P. and G.M. Jenkins. 1976. “*Time Series Analysis-Forecasting and Control*”. USA: Holden Day, 575 p., ISBN 0-8162-1104-3.

Box G.E.P. and D.A. Pierce. 1970. “Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models”. *Journal of the American Statistical Association*, vol. 64.

Cromwell, J.B., W.S. Labys and M. Terraza. 1994. “*Univariate Tests for Time Series Models*”. SAGE Publication, Thousand Oaks, California, ISBN 0-8039-4991-X.

ACKNOWLEDGEMENT

This work was supported in part by the Grant Agency of the Czech Republic under grant No: 101/06/0920 and in part by the Ministry of Education of the Czech Republic under grant No. MSM7088352102.

AUTHOR BIOGRAPHIES



BRONISLAV CHRAMCOV was born in Uherské Hradiště, Czech Republic, in 1975. He studied Automatization and control technology at the Faculty of Technology in Zlin of the University of Technology in Brno, and he took his degree in 1998. In 2006 he graduated his doctoral degree from the Faculty of Applied Informatics of Thomas Bata University in Zlin. His doctoral thesis was focused on the utilization of time series prediction for control of technological process. He is working now as a lecturer at the Faculty of Applied Informatics of Thomas Bata University in Zlin. His research activities are focused on Control Algorithms for District Heating Systems, Time Series Forecast in Energy or Using of Fuzzy Logic for Time Series Forecast. His email address is: chramcov@fai.utb.cz.



JAROSLAV BALÁTĚ was born in Kolín, Czech Republic in 1931. He graduated from the Faculty of Mechanical Engineering of Slovak Technical University Bratislava. He worked in Power and Heating Plant Brno for 12 years as a process engineer, head of automation and measurement department. In 1966 he started to work at TU Brno-Faculty of Mechanical Engineering gradually as an Assistant Professor and a Professor (he became Professor of Technical Cybernetics at the Czech Technical University Prague, Faculty of Mechanical Engineering in 1986). He gradually managed the Department of Hydraulic Machines and Equipment and the Department of Automation. At present, he is working at the Faculty of Applied Informatics of Thomas Bata University in Zlin. His research activities are focused on Technical cybernetics, Theory of automatic control, Application of automatic control on the energy processes, District heating control for large heat supply systems. His email address is: balate@fai.utb.cz.