Control Chart Limits Setting when Data are Autocorrelated

Darja Noskievičová; Ing., PhD., Ass. Prof.
VŠB-Technical University of Ostrava, Faculty of Metallurgy and Material Engineering
Department of Quality Management
17. listopadu 15, 708 33 Ostrava-Poruba, Czech Republic
Darja.noskievicova@vsb.cz

Key Words: control chart, outliers analysis, setting control limits, time series analysis.

Category: Research paper

1. INTRODUCTION

Correctly set control limits in control charts is one of the main conditions for successful application of statistical process control and for meeting its basic goal, i.e. verifying statistical stability of the analysed process.

Problem of setting control limits in control charts is solved in many publications (from standards to articles and books). Algorithm of setting control limits in these publications does not distinguish between autocorrelated and nonautocorrelated data. It lies in excluding of subgroups that give the "out of control" signal from control limits computation (after revealing the existing assignable causes and realization of the corrective action). This algorithm is not wholly suitable for autocorrelated data.

This paper deals with the idea mentioned above in more detail and the proposal of methodology for setting control charts when data are autocorrelated will be applied to the selected parameter of the blast furnace process.

2. STANDARD METHODOLOGY FOR SETTING CONTROL LIMITS

The standard methodology for setting control limits consists of the following steps:
1. Data collection
2. Computation of control limits using appropriate formulae
3. Control chart construction
4. Control chart analysis
5. Process regulation
6. Control limits recalculation

Steps no. 5 and 6 are realised only when the analysis in step 4 has revealed the process nonstability (there were some points out of the limits or some nonrandom patterns in the control chart). Process regulation (step 5) consists of specifying the assignable causes of the process nonstability and acceptance and realization of adequate corrective actions. Without this step it is not recommended to go to the next step. Step 6 is obviously worked out via excluding out-of control points and control limits recomputation using the remaining points. These steps are repeated till the control chart start to signalize the process statistical stability.

In general this methodology represents the classical outliers analysis. Standard way of solving outliers presence in data set – excluding them – is not suitable when we have autocorrelative data and we use ARIMA modelling. Theoretical basis for the analysis of outliers in time series is in brief described in the next chapter.

3. OUTLIERS ANALYSIS IN TIME SERIES

Time series analysis is a part of many statistical software packages. But only some of them offer the outliers analysis (i.e. methodology for detection and assessment of possible influence of outliers).

Outliers are measurements reflecting some unusual events and disturbances that result in extraordinary patterns in the measurements that are not in accord with the rest of a time series (Box et al., 1994). Such outliers can largely affect selection of suitable model, estimation of model parameters, forecasting, properties of the model residuals including. In practice the presence of such outliers is often unknown and there is need to identify these outliers, assess their influence on the rest of time series and eliminate it.
The process for identification, assessment and the effects elimination of outliers in time series could be divided into the steps – see Fig. 1. Described algorithm follows algorithm designed in (Chang, Tiao and Chen, 1988) and mentioned in (Box et al., 1994), for more information see (Liu, 2006).

4. APPLICATION OF THE TIME SERIES OUTLIER ANALYSIS ON SETTING CONTROL LIMITS

4.1 General algorithm for application of ARIMA modelling on SPC

To apply ARIMA modelling on SPC when data are autocorrelated firstly we must identify the most suitable model and then construct some well-known (Shewhart, CUSUM, EWMA) variable control chart for residuals of the selected ARIMA model. To be effective this procedure must result in the ARIMA model with the parameters estimates that are not biased owing to outliers and that are significant. Residuals of such model should be normally distributed, independent with constant variance (nonnormality could result from presence of outliers in time series or nonconstant variance). Only under such conditions control limits computed from the residuals of such ARIMA model can rich their goal – to offer information about the process statistical stability. For that reason procedure for the identification and assessment of time series outliers (see Fig. 1) must precede setting control limits and the control limits then need not to be recomputed as it
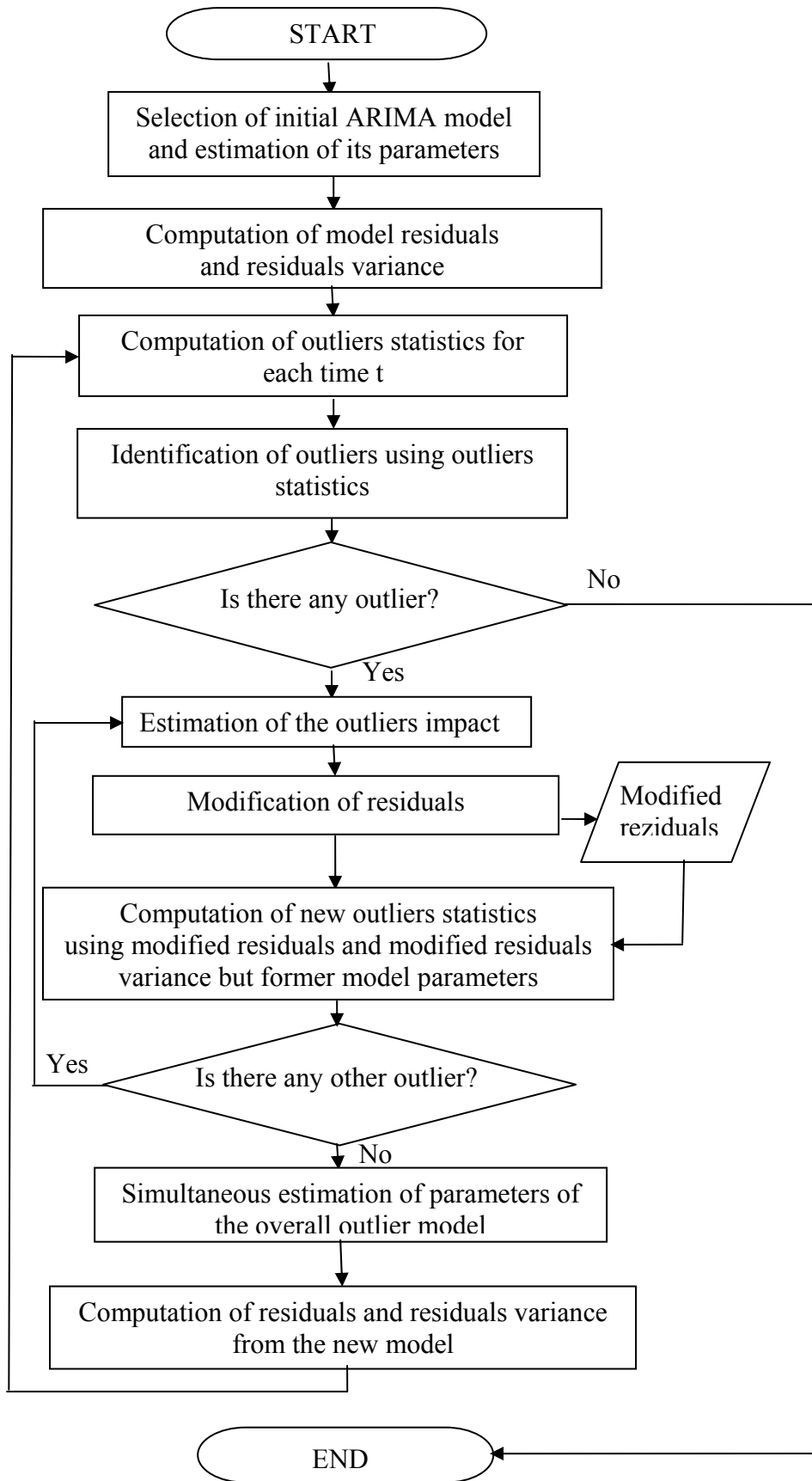
Fig. 1 Flow chart for the time series outliers identification and elimination process

is obvious in standard algorithm for setting control limits (see chapter 2). Algorithm for setting control limits when using ARIMA modelling is described in the next chapter.
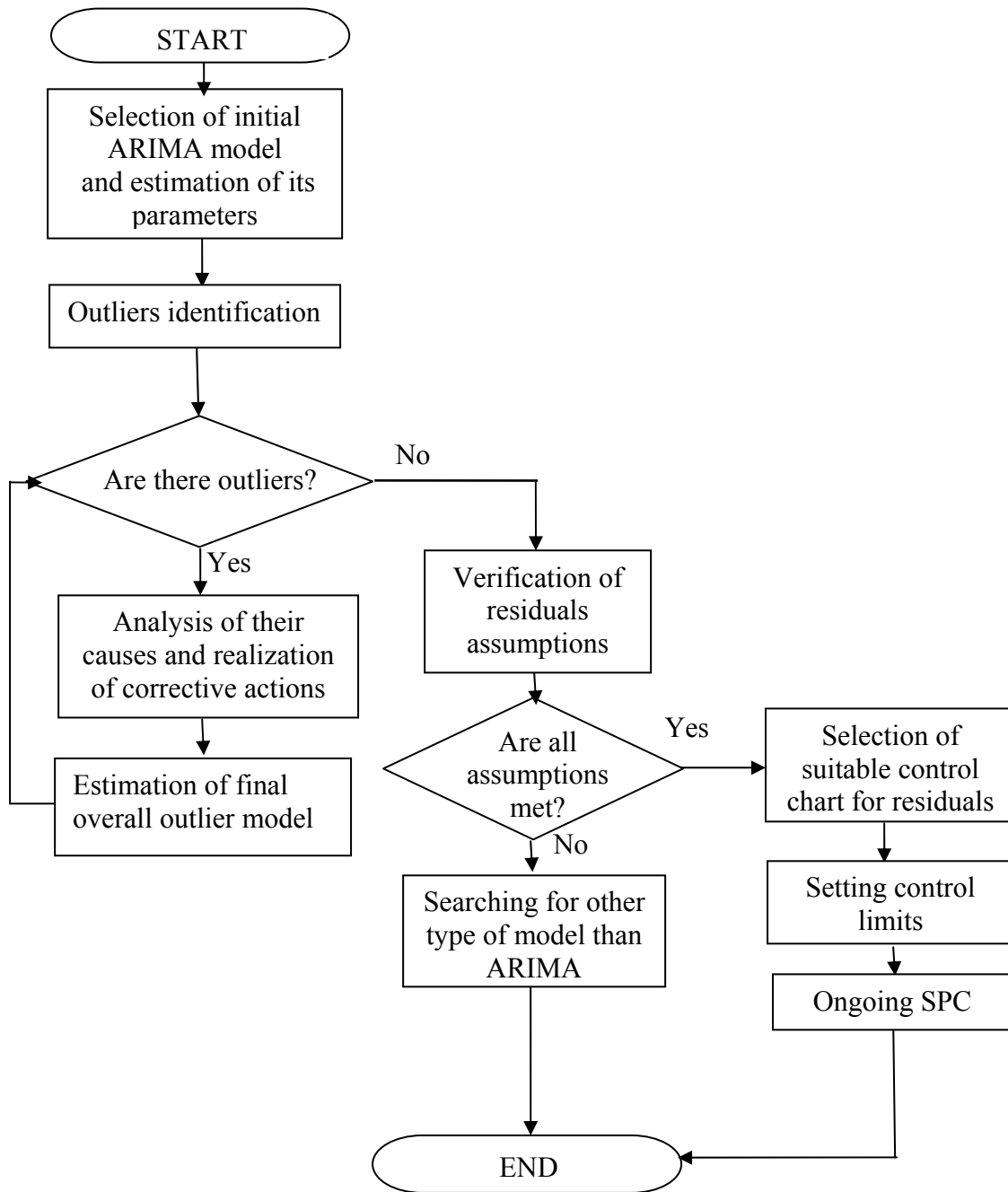
```
                          ┌─────────────┐
                          │    START    │
                          └─────────────┘
                                 │
                                 ▼
                    ┌──────────────────────┐
                    │  Selection of initial │
                    │     ARIMA model       │
                    │  and estimation of its│
                    │     parameters        │
                    └──────────────────────┘
                                 │
                                 ▼
                    ┌──────────────────────┐
                    │ Outliers identification│
                    └──────────────────────┘
                                 │
                                 ▼
                      ◇ Are there outliers? ◇ ─── No ───┐
                                 │                       │
                                Yes                      ▼
                                 │            ┌──────────────────────┐
                    ┌──────────────────────┐ │    Verification of    │
                    │   Analysis of their   │ │      residuals        │
                    │ causes and realization│ │     assumptions       │
                    │  of corrective actions│ └──────────────────────┘
                    └──────────────────────┘             │
                                 │                        ▼
                    ┌──────────────────────┐   ◇ Are all assumptions met? ◇ ── Yes ──┐
                    │   Estimation of final │             │                          │
                    │  overall outlier model│            No                          ▼
                    └──────────────────────┘             │            ┌──────────────────────┐
                                                          ▼            │     Selection of      │
                                             ┌──────────────────────┐ │   suitable control    │
                                             │  Searching for other  │ │  chart for residuals  │
                                             │  type of model than   │ └──────────────────────┘
                                             │       ARIMA           │             │
                                             └──────────────────────┘             ▼
                                                          │            ┌──────────────────────┐
                                                          │            │   Setting control     │
                                                          │            │       limits          │
                                                          │            └──────────────────────┘
                                                          │                        │
                                                          │                        ▼
                                                          │            ┌──────────────────────┐
                                                          │            │     Ongoing SPC       │
                                                          │            └──────────────────────┘
                                                          ▼                        │
                                                 ┌─────────────┐                   │
                                                 │     END     │◄──────────────────┘
                                                 └─────────────┘
```

Fig.2   Flow chart of algorithm for setting control limits when using ARIMA modelling

4.2 Algorithm for setting control limits when using ARIMA modelling

In this chapter algorithm for setting control limits when using ARIMA modelling is described.

After identification of the initial ARIMA model and estimation of its parameters the outliers identification and assessment is to be realized (see Fig. 2).   When some outlier is identified its cause must be searched for and adequate corrective action must be realized. When final overall outlier model is identified and no outliers are resent residuals from this model should be verified. When they are normally distributed, independent with constant variance central line and control limits for the selected control chart can be computed from residuals and ongoing statistical process control can start.

When residuals do not meet all assumptions we must try to identify some different time series model than ARIMA.

5. APPLICATION OF SUGGESTED ALGORITHM FOR SETTING CONTROL LIMITS

Suggested algorithm will be shown on the analysis of the selected output parameter of the blast furnace process, i.e. the amount of H_2 in the output blast furnace gas (in %). During the analysed period (2004 and 2005 years) there were applied two production methods different in additional fuel (let us mark these different methods A and B). The comparison of stability of these two production methods from the point of view of the output parameters (portion of H_2 in the blast furnace gas including) has been set as a partial goal of this statistical analysis. For the both methods fitting model was identified and estimated using data from 2004. After residuals verification control limits for classical Shewhart control chart for individuals were computed using residuals of this model. These control limits were used for ongoing process control in 2005.

5.1 Application of SPC on the method A

For the first time data autocorrelation was verified. Tests and graphs confirmed that data are autocorrelated (see Tab. I and Fig. 3). For that reason it was decided to apply time series modelling and to identify and estimate the fitting ARIMA model, outliers identification and assessment including.

Tab. I Tests for randomness for % H_2 in 2004 – production method A

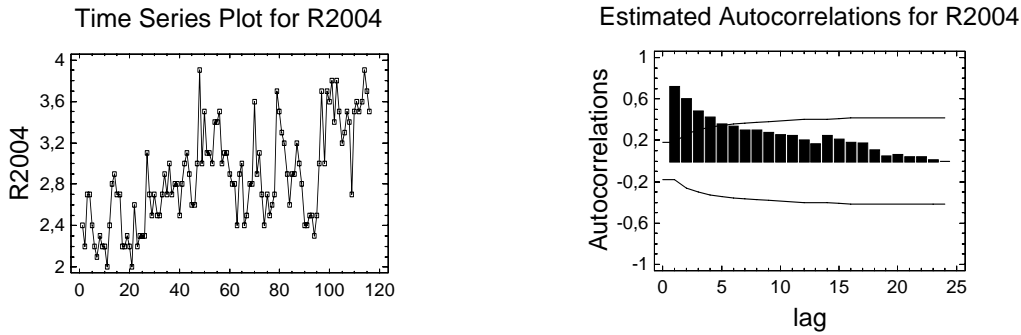| Test for randomness	P-value
Runs above and below median	5,75E-9
Runs up and down	0,0001028
Box-Pierce	0

Fig. 3 Time series plot and ACF for % H_2 in 2004 – production method A

As the best model for this time series there was identified ARIMA (1,0,0). More information about model we can find in tab.II.

Tab.II Final model parameters for % H_2 2004 – production method A

Parameters estimation	P-value of t-test	Constant incorp.	Outliers	Outlier estimate	P-value of t-test
Constant = 2,807	0,000	Yes	48 Additive	0,904	0,000
ϕ_1= 0805	0,000		70 Additive	0,751	0,000
			79 Transient		
			Magnitude	0,981	0,000
			Decay factor	0,726	0,001
			97 Innovational	0,738	0,007
			109 Additive	0,735	0,001

P-values say us that all estimates of the model parameters estimated outliers effects including are significant and we could suppose that values of parameters and outliers effects are not equal to zero. Causes of outliers were discussed and possible corrective actions were considered.

Tab.III   Results of verification of residuals

Verified assumption	Test	P-value
Normality	X^2	0,49
	Kolmogorov	0,65
	Kolmogorov-Smirnov	> 0,1
	Anderson-Darling	0,55
	Skewness	0,89
	Kurtosis	0,22
Autocorrelation	Runs above and below median	0,34
	Box-Pierce	0,78
Constant variance	Bartlett	0,78

In the next step residuals from this final overall outlier model were verified. Tests confirmed that residuals are normally distributed, independent and with constant variance. (see Tab. III).

Results of the residuals verification of the final ARIMA model enabled us to use these residuals for setting control limits and verifying process stability. We applied Shewhart

control chart for individuals on these residuals (see Fig. 4). We can see that process in 2004 can be considered statistical stable (in control) and that control limits were correctly set and can be applied to the process in a future (see Fig.5).

As we can see in 2005 the process (technology A) could not be considered in control. It reflects that all discussed corrective actions were not been actually realized.
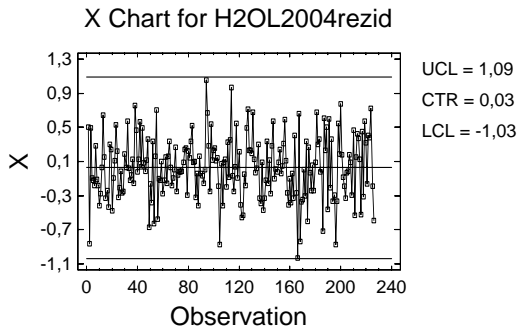
X Chart for residuals of H2 model

X Chart for rezidPREPN2005
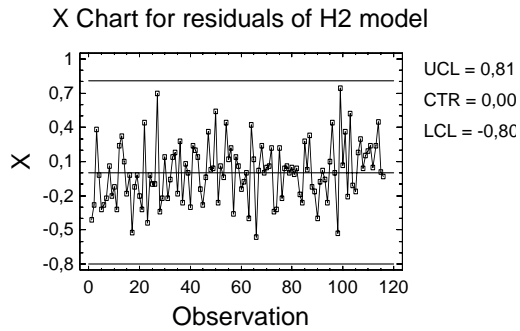
2004

2005

Fig.4 Shewhart control chart for residuals of ARIMA model for % H₂ (technology A)

Fig.5 Shewhart control chart for residuals of ARIMA model for % H₂ (technology A)

X Chart for H2OL2004rezid

X Chart for h2ol2005PREPREZID
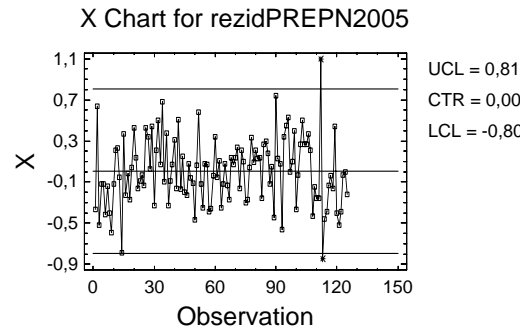
2004

2005

Fig.6 Shewhart control chart for residuals of ARIMA model for % H₂ (technology B)
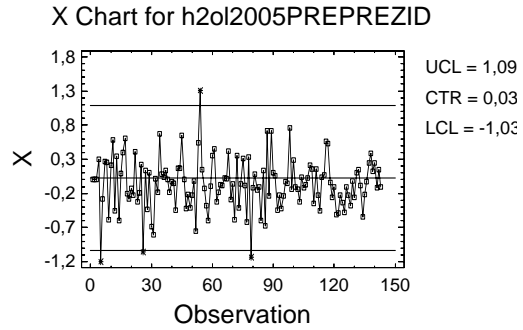
Fig.7 Shewhart control chart for residuals of ARIMA model for % H₂ (technology B)

5.2 Application of SPC on the method B

Analysis of the portion of H₂ in the blast furnace gas by the method B (technology using the other additional fuel) was done in the same way as the previous one.

As the best final model for the time series of data from 2004 year there was identified ARIMA (0,1,2). There were no outliers identified.

Shewhart control chart for individuals for residuals from this model has been constructed (see Fig. 6). Fig. 6 shows that the process (technology B) in 2004 can be considered statistical stable (in control) and that control limits were correctly set and can be applied to the process in a future (see Fig. 7). As we can see on Fig. 7 in 2005 the process

(technology B) could not be considered in control. It reflects the same as the technology A i.e. all discussed corrective actions have not been actually realized.

5.3 Comparison of statistical stability of technology A and B

When we compare Fig. 5 and Fig. 7 we can see that technology B is less stable than technology A (4 points out of limits as compared to 2 points out of limits). In addition expressed with standard deviation of original H_2 measurements variation of technology B is larger then variation of technology A.

6. CONCLUSIONS

This paper dealt with the suggestion of methodology for setting control limits when data are autocorrelated and ARIMA modelling is used. The suggested algorithm was compared to the traditional one and it was applied to the selected output parameter of the blast furnace process.

The suggested algorithm will be used when possible for the analysis of other input and output parameters of the blast furnace process as a part of statistical analysis (Noskievičová, 2006) realized in the frame of the national research project focused on reduction of CO_2 production - $DECO_x$ processes.

7. REFERENCES

Box, G.E.P., Jenkins, G.M., Reinsel, G.L., (1994), *Time Series Analysis. Forecasting and Control,* Prentice Hall, Englewood Cliffs, New Jersey.

Chang, I., Tiao, G. C., Chen, C., (1988), "Estimation of Time Series Parameters in the Presence of Outliers", *Technometrics,* 30, 193-204.

Liu, L.M., (2006), *Time series analysis and forecasting.* Scientific Computing Associates, Corp., Villa Park.

Noskievičová, D., (2006), "The Analysis of Selected Blast Furnace Process Indicators using Box-Jenkins Methodology", Report on the Subproject Solving in the Frame of the Research Project CEZ MSM 6198910019 *Reduction of CO_2 Production - $DECO_x$ Processes,* VŠB-TUO, Ostrava. (In Czech)