

# Statistická regulace procesů při autokorelovaných datech

Darja Noskievičová

V článku je navržena metoda ke stanovení regulačních mezí v regulačních diagramech ARIMA a ukázáno jejich použití ke statistické regulaci vysokopecního technologického procesu, kdy hodnoty charakteristického parametru procesu jsou autokorelované.

## 1. Úvod

Jednou z hlavních podmínek úspěšného použití statistické regulace procesu (*Statistical Process Control* – SPC) a naplňování jejího základního cíle, tj. ověřování statistické stability procesu, je správné stanovení regulačních mezí regulačního diagramu.

Problematice stanovení regulačních mezí je věnováno mnoho publikací, mezi jinými např. [5], [7], [8] a [10]. Algoritmus stanovení regulačních mezí uváděný v těchto publikacích však nerozlišuje mezi daty, která jsou a která nejsou autokorelovaná (tj. vzájemně závislá).

Standardní algoritmus stanovení regulačních mezí spočívá v tom, že v případě výskytu bodů mimo regulační meze regulačního diagramu (na proces s velkou pravděpodobností působí určitá vymezitelná příčina variability) jsou tyto body po stanovení a odstranění příčin neobvyklé variability vyloučeny a regulační meze přepočteny. Tento standardní algoritmus ale není vhodný v situacích, kdy jsou data autokorelovaná.

V článku je popsán postup stanovení regulačních mezí v regulačních diagramech ARIMA a je uveden příklad použití navrhovaného postupu na vybraný parametr procesu výroby surového železa.

## 2. Modely ARIMA

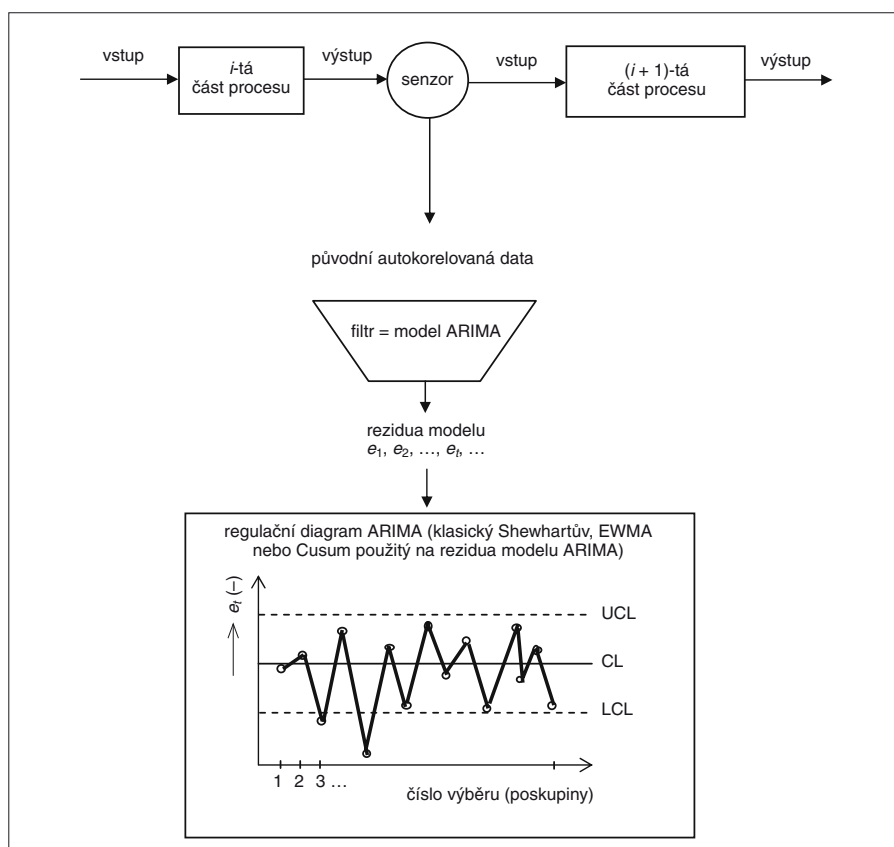
### 2.1 Základy modelování časových řad pomocí modelů ARIMA

Klasická Shewhartova koncepce SPC předpokládá, že naměřená data nejsou autokorelovaná. I velmi nízký stupeň autokorelace dat způsobuje selhání klasických Shewhartových regulačních diagramů. Selhání má podobu velkého počtu bodů mimo regulační meze v regulačním diagramu. Tento jev není výjimečný v případě spojitých procesů, kde je autokorelace dat dána velkou setrvačností procesů v čase (chemické procesy, klimatologické procesy apod.). Stále častějším fenoménem se však autokorelace dat stává i v podmínkách diskretních procesů, a to zejmé-

na výrobních procesů s krátkými výrobními cykly, velkou rychlostí výroby a s vysokým stupněm automatizace výrobních, zkušebních a kontrolních operací. V těchto podmínkách lze získat data o každém produktu s tím důsledkem, že doba mezi měřeními (záznamem) dvou po sobě jdoucích hodnot sledované veličiny je velmi krátká.

nárních časových řad, založenou na teorii pravděpodobnosti.

Lineární modely AR, MA a ARMA jsou nástrojem modelování stacionárních procesů. Tyto modely mají charakteristický tvar autokorelační funkce (*Autocorrelation Function* – ACF) a parciální autokorelační funkce (*Partial Autocorrelation Function* – PACF), které jsou základ-



Obr. 1. Použití modelů ARIMA při statistické regulaci procesů

Jednou z možností, jak řešit otázku autokorelovaných dat, je stochastická koncepce modelování časových řad při použití autoregresních integrovaných modelů klouzavých průměrů, tzv. modelů ARIMA. V rámci stochastické koncepce tvorby lineárních autoregresních modelů (*autoregressive*, tzv. modely AR), modelů klouzavých průměrů (*mean average*, modely MA), smíšených modelů (modely ARMA) a modelů ARIMA, vycházející z Boxovy-Jenkinsonovy metodiky, je časová řada chápána jako realizace stochastického procesu. Boxova-Jenkinsonova metodika tak představuje moderní koncepci analýzy stacionárních a nestacio-

ním nástrojem poskytujícím informace o stochastickém procesu. Odhady ACF a PACF se využívají při identifikaci modelu časové řady.

V praxi se velmi často vyskytují procesy nestacionární. Nestacionarita procesu může být způsobena střední hodnotou měnící se v čase nebo rozptylem procesu měnícím se v čase. Jestliže po transformaci nestacionárního procesu typu „náhodná procházka“ (tzv. integrovaného procesu) při použití difference  $d$ -tého řádu lze výsledný proces popsat stacionárním modelem ARMA ( $p, q$ ), původní integrovaný proces se nazývá autoregresní integrovaný proces klouzavých průměrů řádu  $p, d, q$ , tj. ARIMA ( $p, d, q$ ).

Obecný tvar modelu ARIMA ( $p, d, q$ ) je

$$\Phi_p(B) \nabla^d x_t = \Theta_q(B) \varepsilon_t \quad (1)$$

kde

$\Phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$  je autoregresní polynom  $p$ -tého řádu,

$\Theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$  polynom klouzavých průměrů  $q$ -tého řádu,

$\nabla$  operátor zpětné diference,

$d$  řád diference,

$B$  operátor zpětného posunu ( $Bx_t = x_{t-1}$ ),

$\phi_1, \phi_2, \dots, \phi_p$  parametry autoregresního modelu,

$\theta_1, \theta_2, \dots, \theta_q$  parametry modelu klouzavých průměrů,

$\varepsilon_t$  proměnná, které se říká bílý šum a která představuje nepredikovatelnou fluktuaci v datech; má normální rozdělení se střední hodnotou rovnou nule a konstantním rozptylem  $\sigma_\varepsilon^2$  a její hodnoty jsou nekorelované.

Je-li  $\hat{x}_t$  odhad empirické hodnoty  $x_t$  získaný při použití vhodně zvoleného modelu ARIMA, rezidua tohoto modelu  $e_t = x_t - \hat{x}_t$  se budou chovat jako nezávislé náhodné proměnné pocházející z normálního rozdělení.

## 2.2 Hledání modelu ARIMA

Základní algoritmus hledání modelu ARIMA lze definovat takto:

1. Posouzení stacionarity časové řady z hlediska rozptylu a případná stabilizace časové řady z hlediska rozptylu prostřednictvím vhodné linearizující transformace.
2. Posouzení stacionarity z hlediska střední hodnoty procesu.
3. Stanovení a ověření řádu diferencování (tj. stacionarizace časové řady z hlediska střední hodnoty).
4. Volba modelu, popř. modelů AR, MA nebo ARMA pro stacionarizovanou časovou řadu, a to na základě hledání podobnosti ve tvaru výběrové ACF a PACF a teoretických ACF a PACF pro známé modely (tento krok zahrnuje stanovení řádu modelů  $p$  a  $q$ ).
5. Odhad parametrů zvoleného modelu, popř. zvolených modelů.
6. Ověření vlastností zvoleného modelu (popř. zvolených modelů) testováním jeho (jejich) reziduí. Model je považován za vhodující, pokud jeho rezidua vykazují konstantní rozptyl (jsou homoskedastická), nejsou autokorelovaná a mají normální rozdělení.
7. Volba nejlepšího modelu (v případě, že vhodujících modelů je několik, je nutné použít některé z kritérií pro volbu nejlepšího modelu, např. Akaikeho kritérium AIC, Schwartzovo kritérium SC (nejvhodnější je model, pro který dané kritérium nabývá minima).

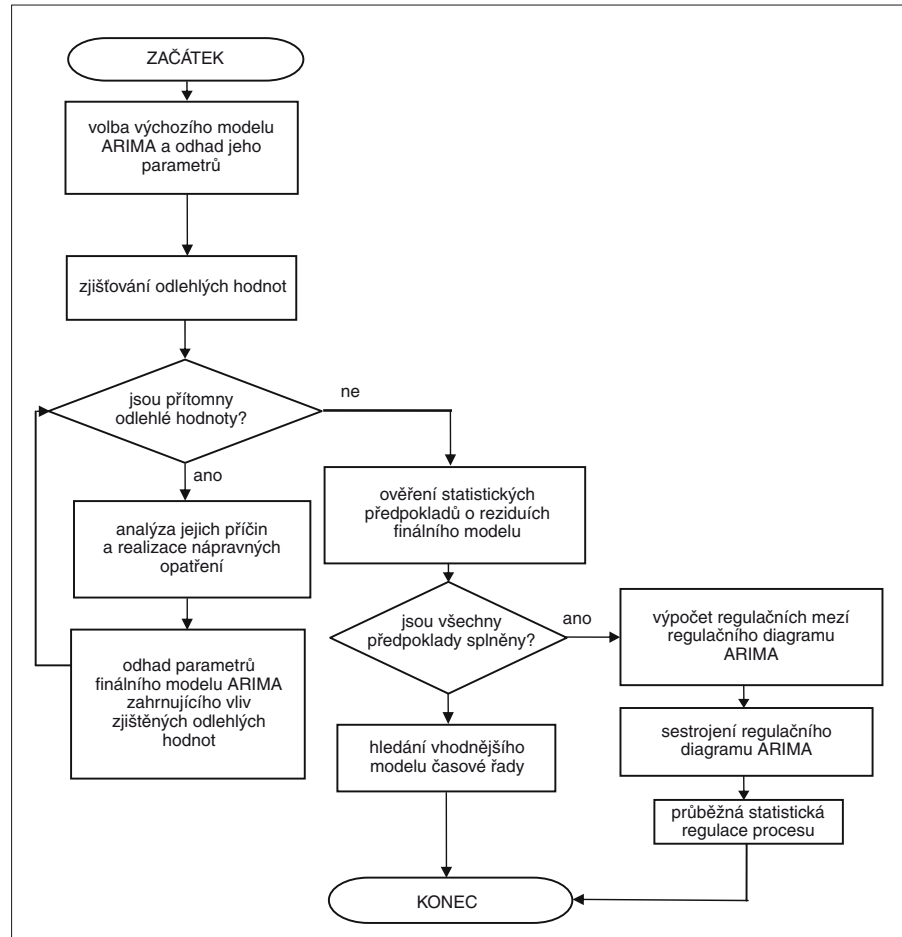
Detailní informace o modelování s použitím modelů ARIMA lze nalézt např. v [1], [2], [6] nebo [3].

## 2.3 Analýza odlehlých hodnot v časových řadách

Standardní algoritmus hledání modelu ARIMA, definovaný v předchozí kapitole, lze realizovat v rámci mnoha známých softwarových produktů (např. Statgraphics). Ne-

zmiňovaném také ve [3]. Další podrobné informace o analýze časových řad z hlediska přítomnosti odlehlých hodnot lze nalézt v [6].

Prvním krokem iterativního procesu je hledání výchozího modelu ARIMA, tzn. modelu pomíjejícího možnost výskytu odleh-



Obr. 2. Vývojový diagram procesu stanovení regulačních mezí v modelech ARIMA

zahrnuje ovšem velmi významný krok, a to nalezení a analýzu odlehlých hodnot.

Odlehlé hodnoty jsou naměřené údaje, které odraží určité neobvyklé jevy a poruchy, vedoucí ke zvláštním uskupením dat, jež nejsou v souladu se zbytkem časové řady [3]. Odlehlé hodnoty mohou ovlivnit výběr správného modelu, přesnost odhadů jeho parametrů, vlastnosti reziduí modelu a také kvalitu předpovědí. Proto je třeba takové hodnoty v časové řadě odhalit, určit jejich vliv na zbytek časové řady a tento vliv zohlednit při odhadu parametrů finálního modelu ARIMA. Softwarových produktů umožňujících najít a analyzovat odlehlé hodnoty v časových řadách je podstatně méně (jako příklad lze uvést programy SPSS a SCA).

Dále popsany iterativní proces zjišťování přítomnosti odlehlých hodnot v časové řadě, určení velikosti vlivu odlehlých hodnot a jejich zohlednění při odhadu parametrů finálního modelu odpovídá algoritmu podle [4],

lých hodnot (zatím se předkládá, že časová řada neobsahuje odlehlé hodnoty). Vypočítají se hodnoty reziduí modelu  $e_t$  a odhad jejich rozptylu  $\hat{\sigma}_a^2$ . Dalším krokem je zjišťování odlehlých hodnot, a to testováním jejich přítomnosti v časové řadě. Jestliže test potvrdí přítomnost odlehlé hodnoty, je její vliv eliminován z reziduí a rezidua i odhad  $\hat{\sigma}_a^2$  jsou přepočítány. Testování odlehlých hodnot se opakuje s využitím přepočítaných hodnot reziduí a odhadu  $\hat{\sigma}_a^2$ , ale s parametry původního modelu ARIMA tak dlouho, dokud nejsou zjištěny všechny odlehlé hodnoty. Poté jsou odhadnuty parametry modelu ARIMA, který již zohledňuje zjištěné odlehlé hodnoty, a jsou vypočítána rezidua a odhad  $\hat{\sigma}_a^2$  z tohoto upraveného modelu. Celá procedura testování přítomnosti odlehlých hodnot se pak opakuje s novými rezidui a odhadem  $\hat{\sigma}_a^2$  tak dlouho, až jsou nalezeny všechny odlehlé hodnoty a jsou odhadnuty parametry finálního modelu ARIMA původní časové řady, který zohledňuje všechny zjištěné odlehlé hodnoty.

### 3. Návrh algoritmu stanovení regulačních mezí v regulačních diagramech ARIMA

#### 3.1 Obecný algoritmus použití modelů ARIMA při SPC

Modelování s využitím modelů ARIMA je jednou z možností, jak při použití metody SPC vyřešit problém s autokorelovanými daty. Nejdříve je třeba najít nejvhodnější model ARIMA a pak použít některý ze známých regulačních diagramů (Shewhartův, *Cumulative Sum* – Cusum, *Exponentially Weighted Moving Average* – EWMA) na rezidua tohoto modelu. Takto vytvořený regulační diagram se označuje jako regulační diagram ARIMA. Schematicky je uvedený postup znázorněn na obr. 1.

Aby popsána procedura byla efektivní, musí vést k modelu, jehož odhady parametrů jsou statisticky významné a nejsou zatíženy vlivem odlehlých hodnot. Rezidua takového modelu by měla mít normální rozdělení, být neautokorelovaná, s konstantním rozptylem. Pouze za těchto podmínek budou regulační meze vypočítané z reziduí daného modelu ARIMA plnit svůj účel – tj. umožní posoudit statistickou stabilitu analyzovaného procesu. Z tohoto důvodu musí být procedura odhalení a určení vlivu odlehlých hodnot součástí procesu hledání vhodného modelu ARIMA, a musí tedy předcházet stanovení regulačních mezí. Regulační meze pak již není třeba přepočítávat, jak je to obvyklé podle standardního algoritmu stanovení regulačních mezí.

#### 3.2 Algoritmus stanovení regulačních mezí v regulačních diagramech ARIMA

Na obr. 2 je znázorněn algoritmus stanovení regulačních mezí v modelech ARIMA. Algoritmus v sobě zahrnuje kroky již popsané v kap. 2.2 a kap. 2.3.

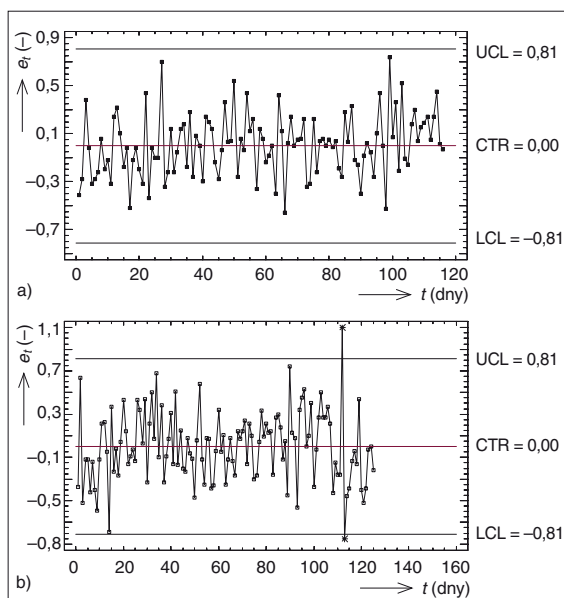
Po nalezení výchozího modelu ARIMA, který nebere v potaz možné odlehlé hodnoty, a odhadu jeho parametrů musí být zjištěny odlehlé hodnoty a ohodnocen jejich vliv na zbytek časové řady. Po zjištění odlehlé hodnoty musí být stanovena příčina jejího výskytu a přijato adekvátní nápravné opatření, které zajistí, aby v budoucnu tento faktor na proces nepůsobil. Po nalezení finálního modelu, zahrnujícího vliv všech zjištěných odlehlých hodnot, se verifikují statistické vlastnosti reziduí tohoto modelu. Mají-li rezidua modelu požadované statistické vlastnosti (viz kap. 3.1), je možné považovat nalezený model za přijatelný a lze přejít k výpočtům regulačních mezí a střední přímkou z reziduí daného modelu ARIMA, zkonstruovat tento regulační diagram a zahájit průběžnou regulaci daného procesu, popř. využít informace z regulačního diagramu k další analýze jeho variability. Jestliže však rezidua nemají předpokládané statistické vlastnosti, je třeba hledat vhodnější model časové řady, lépe vyhovující je-

Tab. 1. Typy odlehlých hodnot v časových řadách v softwaru SPSS verze 15

Typ odlehlé hodnoty	Charakteristika
aditivní	odlehlá hodnota, která má vliv na jediné pozorování
posun úrovně	odlehlá hodnota, která posune od určitého místa v časové řadě všechna pozorování o určitou konstantu
inovační	odlehlá hodnota, která působí jako přídavek k šumu v určitém místě časové řady; u stacionárních časových řad tento druh odlehlé hodnoty ovlivní několik pozorování; u nestacionárních časových řad může ovlivňovat všechna pozorování počínaje určitým místem v časové řadě
přechodná	odlehlá hodnota, jejíž účinek ubývá exponenciálně k nule
lokální trend	lokální odlehlá hodnota, kterou v určitém místě časové řady začíná trend
aditivní shluk	skupina dvou nebo více aditivních po sobě jdoucích odlehlých hodnot

Tab. 2. Odhady parametrů finálního modelu ARIMA zahrnujícího vliv odlehlých hodnot pro analyzovaný parametr vysokopecního procesu v roce 1 – výrobní metoda A

Odhad parametrů modelu	Zařazení konstanty	Nalezené odlehlé hodnoty/jejich typy	Odhad vlivu odlehlé hodnoty
konstanta = 2,807	ano	48/aditivní	0,904
$\phi_1 = 0,805$		70/aditivní	0,751
		79/přechodná:	
		....- velikost	0,981
		....- faktor úbytku	0,726
		97/inovační	0,738
		109/aditivní	0,735



Obr. 3. Regulační diagram ARIMA – výrobní metoda A: a) rok 1, b) rok 2

jímu charakteru (lineární či nelineární model volatility, lineární model časové řady s dlouhou pamětí, nelineární model s proměnlivými režimy apod.).

### 4. Použití navrženého algoritmu, stanovení regulačních mezí modelu ARIMA

#### 4.1 Základní charakteristika úlohy

V následujícím textu je ukázáno použití navrženého algoritmu ke statistické analýze denních hodnot vybraného parametru vysokopecního procesu. Během analyzovaného ob-

dobí (rok 1 a rok 2) byly použity dvě výrobní metody, lišící se přídatným palivem (necht jsou označeny jako metoda A a metoda B). Cílem analýzy bylo porovnat statistickou stabilitu procesu z pohledu vybraného parametru při použití metody A a při použití metody B.

Pro metodu A i metodu B byl vhodný model ARIMA nalezen na základě analýzy dat z roku 1. Pro hledání modelu ARIMA, včetně analýzy odlehlých hodnot, byl použit statistický software SPSS verze 15, který rozlišuje šest typů odlehlých hodnot v časových řadách (tab. 1).

Po verifikaci reziduí nejvhodnějšího modelu z nich byl sestaven Shewhartův regulační diagram pro individuální hodnoty (regulační diagram ARIMA). Tento regulační diagram

byl poté použit k analýze stability vysokopecního procesu při metodě A a B v roce 2 a výsledky byly využity k porovnání statistické stability obou metod.

#### 4.2 Analýza výrobní metody A

V prvním kroku byla verifikována data z hlediska jejich autokorelace. Grafy časové řady a ACF, stejně jako testy náhodnosti, potvrdily hypotézu, že hodnoty sledovaného parametru vysokopecního procesu jsou autokorelované. Proto bylo rozhodnuto modelovat daný parametr procesu s použitím modelu ARIMA.

Po nalezení výchozího modelu a odhadu jeho parametrů byla provedena analýza odlehlých hodnot a zohledněn jejich vliv při odhadu parametrů finálního modelu ARIMA. Jako nejlepší model byl identifikován model ARIMA (1, 0, 0). Bližší charakteristika tohoto modelu, včetně identifikace a odhadu vlivu odlehlých hodnot, je uvedena v tab. 2.

Dále byly diskutovány příčiny výskytu odlehlých hodnot a byla zvažována nápravná opatření. Poté byla ověřena normalita, míra autokorelace a stálost hodnot reziduí finálního modelu. Na základě výsledků vybraných testů lze předpokládat, že rezidua mají normální rozdělení, konstantní rozptyl a nejsou autokorelovaná.

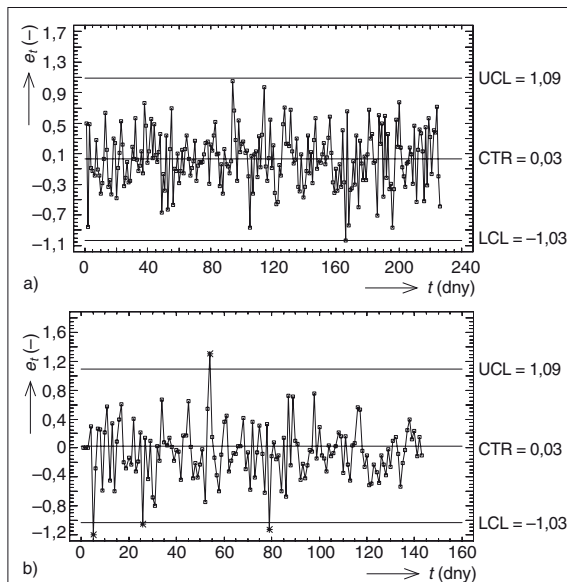
Protože lze předpokládat, že rezidua daného modelu mají vlastnosti, které musí mít regulovaná veličina, má-li být použit klasický Shewhartův regulační diagram pro individuální hodnoty, je nyní možné z hodnot reziduí vypočítat regulační meze a střední průmku a zkonstruovat regulační diagram ARIMA (obr. 3a). Z uvedeného regulačního diagramu je patrné, že proces lze pokládat za statisticky stabilní (žádný bod není mimo regulační meze), a proto lze tyto meze použít k ověřování statistické stability procesu v následujícím období (obr. 3b). Jak je patrné z obr. 3b, v roce 2 nelze analyzovaný proces pokládat za statisticky stabilní.

### 4.3 Analýza výrobní metody B

Analýza sledovaného parametru vysokopecního procesu při použití výrobní metody B byla provedena stejným způsobem jako analýza metody A.

Jako nejlepší model analyzované časové řady hodnot parametru při použití výrobní metody B byl nalezen z dat za rok 1 model ARIMA (0, 1, 2). Odlehlá hodnota nyní nebyla zjištěna žádná. Poté byl sestaven regulační diagram ARIMA pro rezidua uvedeného modelu (obr. 4a). Z regulačního diagramu

je patrné, že proces lze pokládat za statisticky stabilní (žádný bod není mimo regulační meze), a proto je možné regulační meze použít k ověřování statistické stability procesu v následujícím období (obr. 4b). Jak lze vyčíst z obr. 4b, v roce 2 nelze analyzovaný proces pokládat za statisticky stabilní.



Obr. 4. Regulační diagram ARIMA – výrobní metoda B: a) rok 1, b) rok 2

### 4.4 Porovnání statistické stability výrobních metod A a B

Z porovnání obr. 3b a obr. 4b lze vyvodit závěr, že výrobní metoda B je méně stabilní než metoda A (čtyři body mimo regulační meze u metody B v porovnání se dvěma body mimo meze u metody A).

### 5. Závěr

V článku je navržena metoda určená ke stanovení regulačních mezí v regulačních diagramech ARIMA. Navržený algoritmus byl porovnán se standardním postupem pro stanovení regulačních mezí a byl ověřen na vybraném parametru vysokopecního procesu, přičemž cílem bylo posoudit statistickou stabilitu tohoto procesu při použití dvou různých výrobních metod.

### Poděkování

Článek vznikl v rámci národního výzkumného projektu CEZ MSM 6198910019 *Procesy snižování emisí CO<sub>2</sub> – DECO<sub>x</sub> procesy*.

### Literatura:

- [1] ARLT, J.: *Moderní metody modelování ekonomických časových řad*. Grada Publishing, Praha, 1999.
- [2] ARLT, J. – ARLTOVÁ, M.: *Ekonomické časové řady*. Grada Publishing, Praha, 2007.
- [3] BOX, G. E. P. – JENKINS, G. M. – REINSEL, G. L.: *Time Series Analysis. Forecasting and Control*. Prentice Hall, Englewood Cliffs, New Jersey, 1994.
- [4] CHANG, I. – TIAO, G. C. – CHEN, C.: *Estimation of Time Series Parameters in the Presence of Outliers*. *Technometrics*, 1988, 30, pp. 193–204.
- [5] GRIFFITH, G. K.: *Statistical Process Control Methods for Long and Short Runs*. ASQC Quality Press, Milwaukee, 1996.
- [6] LIU, L. M.: *Time series analysis and forecasting*. Scientific Computing Associates, Corp., Villa Park, 2006.
- [7] MITRA, A.: *Fundamentals of Quality Control and Improvement*. Macmillan Publishing Company, New York, 1993.
- [8] MONTGOMERY, D. C.: *Introduction to Statistical Quality Control*. John Wiley, New York, 2001.
- [9] NOSKIEVIČOVÁ, D.: *Analýza vybraných ukazatelů vysokopecního procesu pomocí Boxovy-Jenkinsovy metodiky*. Zpráva z řešení dílčího úkolu v rámci projektu CEZ MSM 6198910019 *Procesy snižování produkce CO<sub>2</sub> – DECO<sub>x</sub> procesy*. VŠB-TUO, Ostrava, 2006.
- [10] WADSWORTH, H. M. et al.: *Modern Methods for Quality Control and Improvement*. John Wiley, New York, 1986.

doc. Ing. Darja Noskiewiczová, CSc.,  
VŠB – Technická univerzita Ostrava,  
Fakulta metalurgie a materiálového  
inženýrství,  
katedra kontroly řízení jakosti  
(darja.noskiewiczova@vsb.cz)

Lektorovala: Ing. Věra Pelantová, Ph.D.

## ► Dresser Masoneilan členem FDT Group

Společnost Dresser Masoneilan, významný výrobce průmyslových regulačních ventilů, se v září 2008 stala členem FDT Group. Tato skutečnost významně posílila všeobecné přijetí FDT mezi dodavateli ventilů: nyní jsou všichni významnější světoví výrobci regulačních a bezpečnostních ventilů, s je-

dinou výjimkou, členy FDT Group. Je to jasný důkaz toho, že odborníci přijali FDT jako nástroj, který významně usnadňuje dohled nad ventily, jejich diagnostiku a prediktivní údržbu. Proto má FDT silnou podporu také ze strany koncových uživatelů, např. společností Saudi Aramco, Shell nebo Chevron. Také WIB, mezinárodní sdružení uživatelů přístrojové techniky se sídlem v Nizozemí, označilo FDT za nejlepší nástroj pro správu a údržbu provozní přístrojové techniky.

Sandro Esposito, product marketing manager společnosti Dresser Masoneilan, prohlásil, že od členství v FDT si slibují expanzi na trhu moderních inteligentních regulačních a bezpečnostních ventilů. Poptávka po takových ventilech, s dálkovou diagnostikou a funkcemi umožňujícími prediktivní údržbu, roste, protože napomáhají snižovat celkové výrobní náklady, zkracovat dobu nutných odstávek a předcházet poruchám.

(Bk)

krátké zprávy