LINEÁRNÍ MODELY PREDIKCE SPOTŘEBY PLYNU

Doc. Ing. Aleš Procházka, CSc, Ing. Jaroslav Stříbrský
Ústav počítačové a řídicí techniky, VŠCHT Praha

Ing. Zdeněk Kánský
Ústřední plynárenský dispečink, Praha

Souhrn: *Problematika lineární a nelineární predikce je základní otázkou řešenou v řadě informačních a řídicích systémů. Specifickým problémem je přitom předpověď chování časových řad s významnou periodickou složkou, která je v případě spotřeby plynu dána kolísáním jeho odběru s periodou jednoho dne, týdne a roku. Předmětem příspěvku je ukázka možností analýzy a predikce takovéto časové řady spotřeby plynu pomocí základních autoregresních modelů s určením příslušných intervalů spolehlivosti. Závěr příspěvku je věnován zmínce o alternativním způsobu predikce pomocí umělých neuronových sítí.*

Klíčová slova: Modelování signálů, zpracování časových řad, predikce signálů, optimalizace, intervaly spolehlivosti

1 Úvod

Modelování a predikce signálů představuje mezioborovou vědní disciplínu využívající počítačové informační systémy a stejný matematický aparát pro aplikace v různých inženýrských disciplínách založených na zpracování signálů a obrazů. Teoretické základy tohoto oboru zahrnují popis *lineárních systémů* pomocí diferenčních rovnic a autoregresních (AR) modelů [5] a případně *nelineárních systémů* založených na využití umělých neuronových sítí [4]. Ve všech případech je podstatná *optimalizace parametrů* modelu tak, aby se jeho výstupní hodnoty co nejvíce blížily hodnotám pozorovaným a umožnily tak predikci chování reálného systému. Pozorovanou veličinou může být přitom spotřeba plynu nebo časový průběh jiné fyzikální, inženýrské či ekonomické veličiny. Kritériem shody je v řadě případů minimalizace součtu čtverců odchylek mezi vypočtenými a pozorovanými hodnotami pomocí *metody nejmenších čtverců* [1, 2, 3, 6, 7, 8].

Příspěvek je věnován vybraným metodám analýzy pozorovaných časových řad a popisu systému pomocí diferenčních rovnic. Optimalizace parametrů modelu aplikovaná na reálná data pozorovaná s periodou jednoho dne a dále dvou hodin umožňuje rovněž výpočet intervalů spolehlivosti predikce. Příslušné výsledky jsou uvedeny pro oba tyto případy.

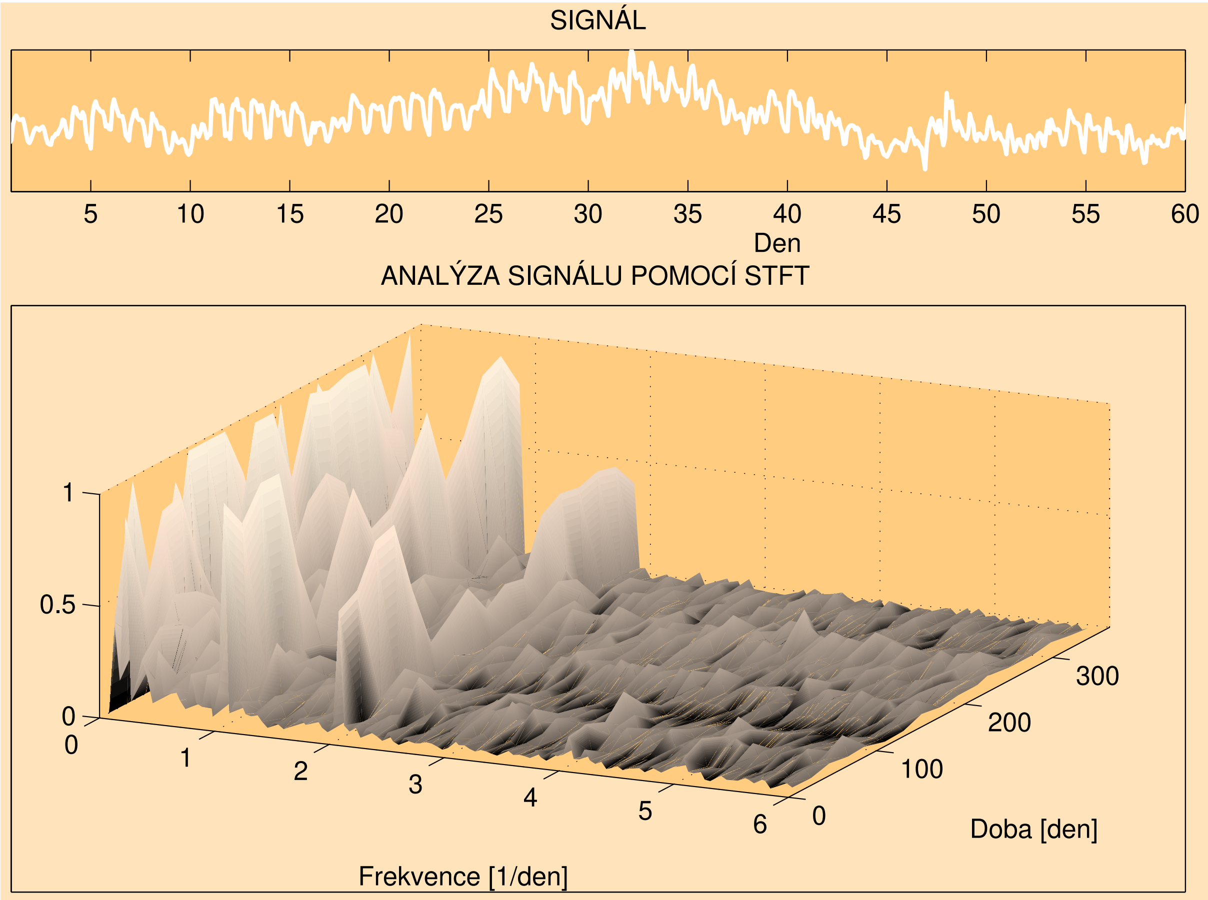
2 Diskrétní Fourierova transformace a analýza signálů

Číslicové metody rozboru pozorovaných časových řad představují základní matematický nástroj pro jejich následný popis, modelování a predikci. Pro časovou řadu $\{x(n)\}_{n=0}^{N-1}$ lze aplikovat diskrétní Fourierovu transformaci (FFT) a určit posloupnost

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\, kn2\pi/N} \qquad (1)$$

pro $k = 0, 1, \cdots, N-1$. Extrémy této funkce detekují periodické komponenty daného signálu.

Obr. 1 presentuje průběh části spotřeby plynu v prvých dvou měsících roku 1998 vykazující zřejmou periodicitu délky jednoho dne a jednoho týdne. Tuto periodicitu lze matematicky určit aplikací FFT na pozorovanou časovou řadu podle vztahu (1) a posouzením extrémů této funkce. V případě krátké diskrétní Fourierovy transformace (STFT) se tento vztah aplikuje na signál po jednotlivých částech a umožňuje tak posouzení případných časových změn frekvenčních složek signálu. Obr. 1 ukazuje příslušné výsledky, které jasně detekují periodicitu s periodou 7 dní tj. frekvencí 1/7 [1/den] a frekvencí 1 [1/den] s extrémy i na vyšších harmonických složkách. Jelikož detekce periodických složek signálu není v časové oblasti v řadě případů možná, představuje analýza signálu v oblasti frekvenční často jedinou možnost pro zjištění periodických složek signálu.



Obrázek 1: Spektrální analýza časové řady délky 365 dní pro vzorkovací periodu délky 2 hodin a spotřebu plynu v období od 1. ledna do 31. prosince 1998

3 Modelování signálu

Znalost periodických komponent signálu má velký význam pro jeho další efektivní modelování a predikci, a to v případě modelů lineárních i nelineárních.

Uvažujme opět časovou řadu $\{x(n)\}_{n=0}^{N-1}$ a její lineární model ve tvaru

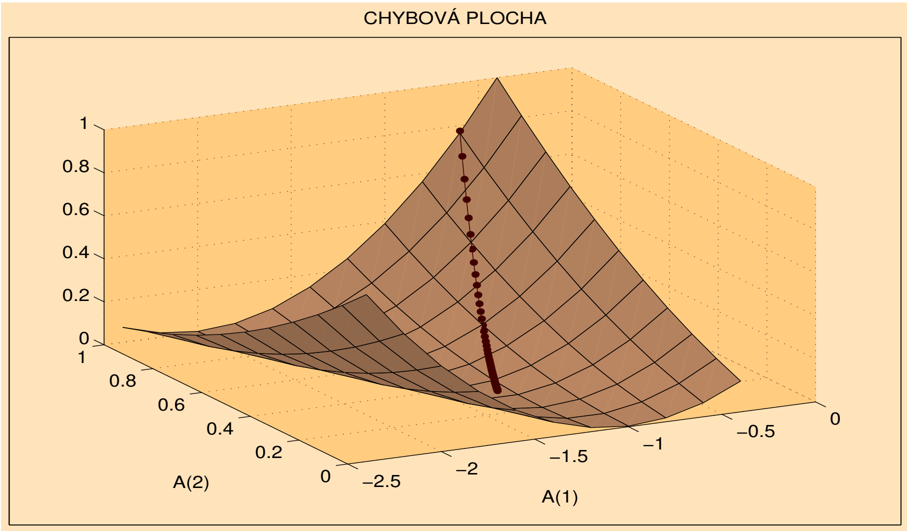
$$\begin{aligned} x(n) \quad &+ \quad a_1 x(n-1) + \cdots + a_{na} x(n-na) = \\ &= \quad e(n) + c_1 e(n-1) + c_2 e(n-2) + \cdots + c_{nc} e(n-nc) \end{aligned} \qquad (2)$$

kde hodnoty $\{a_1, a_2, \cdots, a_{na}\}$ a $\{c_1, c_2, \cdots, c_{nc}\}$ představují neznámé vektory koeficientů modelu a $\{e(n)\}_{n=0}^{N-1}$ representuje náhodné hodnoty, které mohou být použity pro matematický model pozorované posloupnosti.

Základní otázkou směřující k predikci pozorované posloupnosti hodnot je vhodná volba struktury matematického modelu, kterou je v tomto případě počet jeho koeficientů $na + nc$ a případně vhodný výběr těch koeficientů, které významně ovlivňují přesnost modelu [1]. Další otázkou je optimalizace těchto parametrů tak, aby daný model co nejlépe popisoval chování reálné časové řady. Obr. 2 ukazuje průběh chybové plochy, která je pro případ $nc = 0$ a dvou neznámých parametrů a_1, a_2 dána funkcí dvou proměnných

$$S(a_1, a_2) = \sum_{n=2}^{N-1} e(n)^2 = \sum_{n=2}^{N-1} (-a_1 x(n-1) - a_2 x(n-2) - x(n))^2 \tag{3}$$

Cílem optimalizace je přitom nalezení takových hodnot a_1, a_2, které definují minimum na této ploše. Obr. 2 ukazuje průběh řešení pomocí gradientní metody pro daný úsek dat spotřeby plynu a zvolené výchozí hodnoty konstant a_1, a_2, které se v průběhu iteračního řešení blíží k hledaným hodnotám diferenční rovnice. Tato metodika je významná zvláště v případě nelineárních modelů vycházejících z aplikace umělých neuronových sítí.



Obrázek 2: Užití gradientní metody pro nalezení minima součtu čtverců odchylek mezi určenými a danými hodnotami spotřeby plynu pro model řádu $na = 2$

4 Lineární predikce časové řady

Koeficienty diferenční rovnice modelu časové řady umožňují její užití pro predikci chování dané posloupnosti hodnot a následný rekursivní výpočet intervalů spolehlivosti této predikce.

Uvažujme diferenční rovnici (2) ve tvaru

$$x(n) + a_1 x(n-1) + \cdots + a_{na} x(n-na) = e(n) \tag{4}$$

Pro zvolený časový okamžik k_0 lze stanovit odhad následných hodnot ze vztahů

$$\begin{aligned} x(k_0+1) &= -a_1 x(k_0) - a_2 x(k_0-1) - \cdots - a_{na} x(k_0+1-na) \\ x(k_0+2) &= -a_1 x(k_0+1) - a_2 x(k_0) - \cdots - a_{na} x(k_0+2-na) \\ &\cdots \end{aligned} \tag{5}$$

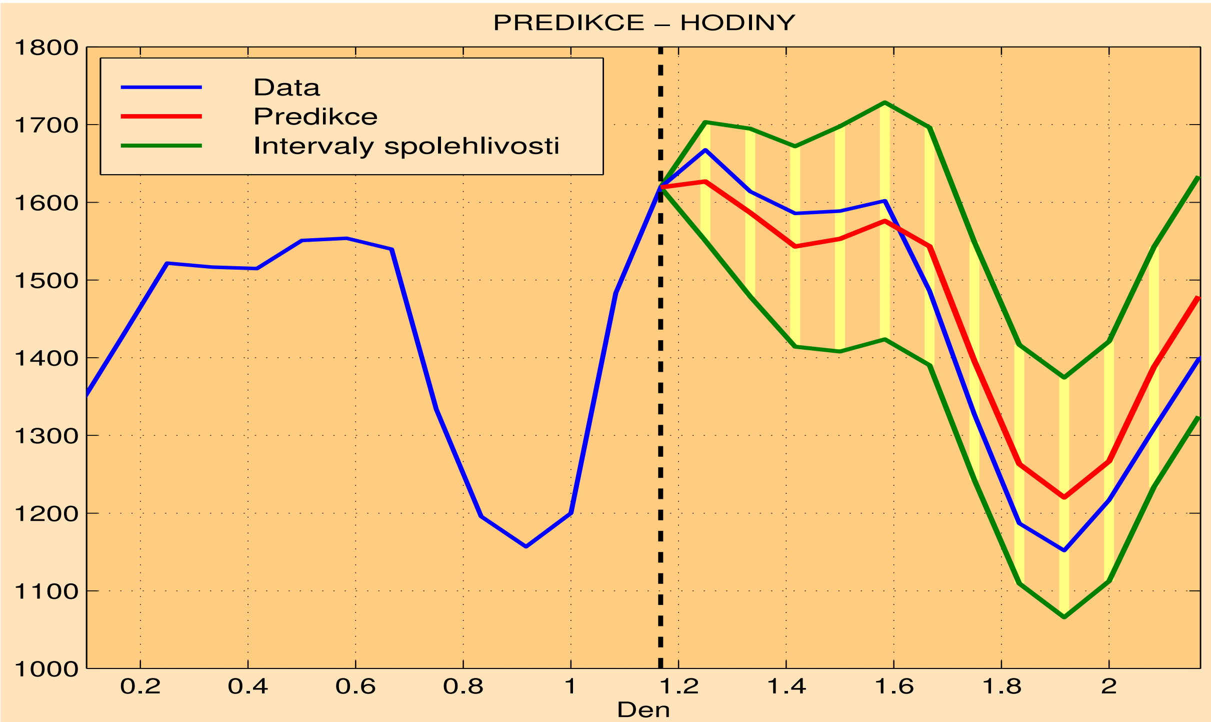
ve kterých představují všechny hodnoty $x(k_0+m), m = 1, 2, \cdots$ odhady této veličiny.

K hodnotám predikce podle vztahu (5) lze dále určit rozpětí, ve kterém se skutečná hodnota se zvolenou pravděpodobností vyskytne. Statistický výpočet těchto intervalů spolehlivosti vychází ze zvolené pravděpodobnosti [1] a modelu systému.

Pro případ dat spotřeby plynu pozorovaných s periodou vzorkování $T_s = 2$ hodiny detekuje diskrétní Fourierova transformace základní periodicitu této řady délky 12 vzorků. Výchozí autoregresní model ve tvaru rovnice (4) je tedy možno volit s řádem $na = 12$. Následná identifikace vede pro zvolný úsek pozorování ke konstantám

$$\mathbf{a} = [1.0000, -0.9223, -0.0159, 0.1949, 0.2305, -0.3025, -0.1184, \\ 0.1641, 0.2807, -0.2599, -0.1667, 0.0121, -0.0934]$$

Predikce reálného signálu representujícího spotřebu plynu určená tímto modelem je uvedena na obr. 3. Na témže grafu jsou dále vyznačeny meze, ve kterých se predikovaná veličina s 50 procentní pravděpodobností vyskytne. Uvedené intervaly spolehlivosti predikce závisí jednak na řádu modelu, na zvolené kritické hodnotě pravděpodobnosti a na délce predikce.



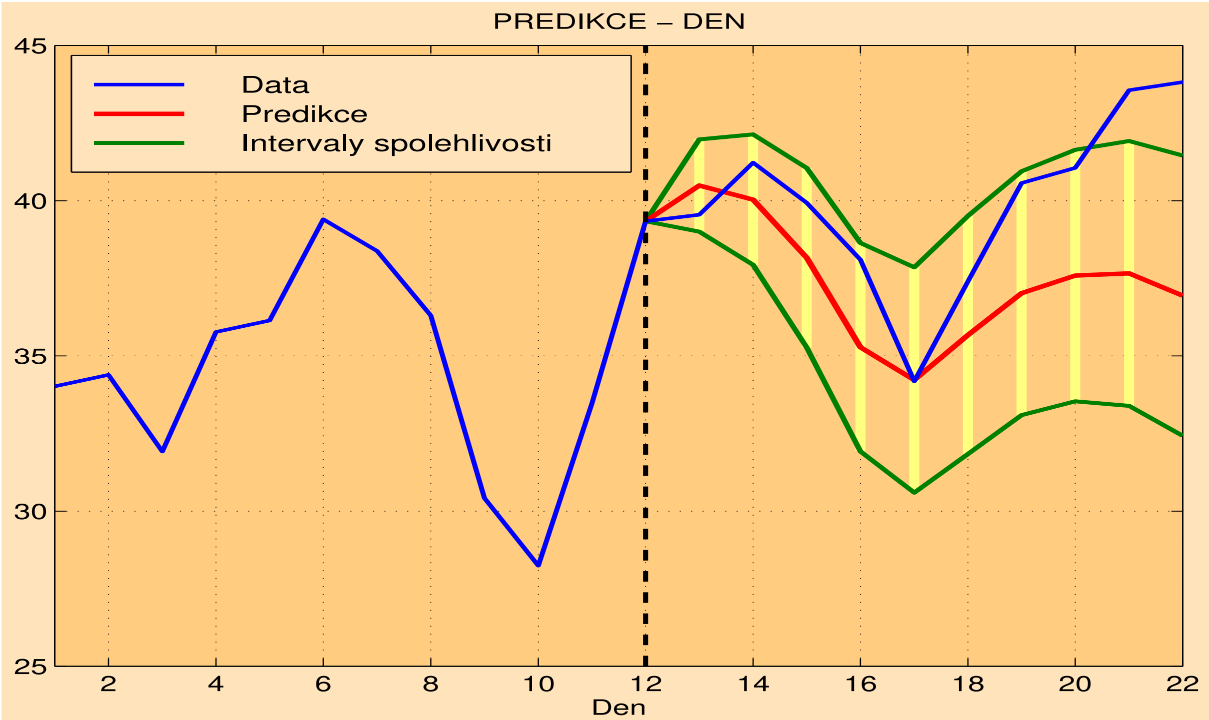
Obrázek 3: Predikce reálného signálu spotřeby plynu pro vzorkovací periodu délky $T_s = 2$ hodin a 50 % intervaly spolehlivosti této predikce pro autoregresní model řádu 12

Podobná metoda byla použita pro data representující spotřebu plynu s periodou vzorkování $T_s = 1$ den. Výsledné koeficienty modelu ve formě vektoru

$$\mathbf{a} = [1.0000, -1.3362, 0.6361, -0.2425, 0.0049, 0.2113, -0.4795, 0.2085]$$

umožňují predikci signálu uvedenou na obr. 4 spolu s příslušnými intervaly spolehlivosti. Výsledky několika testů ukazují, že predikce pro více než 3 dny dopředu je ve většině případů již velice nespolehlivá. Jednodenní predikce však ve většině případů poskytuje výsledky velice dobré se značně úzkými intervaly spolehlivosti.

Střední chyba predikce pro vzorkování $T_s = 2$ hodiny a $T_s = 1$ den je pro různé řády modelu a rozdílnou délku predikce tabelována v tabulkách 1 a 2. Přesnost predikce roste s rostoucím



Obrázek 4: Predikce reálného signálu spotřeby plynu pro vzorkovací periodu délky $T_s = 1$ den a 50 % intervaly spolehlivosti této predikce pro autoregresní model řádu 7

řádem modelu a je nepřímo úměrná délce predikce. Snížení těchto chyb lze očekávat zařazením dalších veličin (teploty) do modelu systému a využitím nelineárních modelů založených na využití neuronových sítí,

Uvedený postup lze dále modifikovat v případě časové řady s významnou periodickou složkou zjednodušením výsledného modelu a využitím pouze těch předchozích hodnot časové řady, které predikci významně ovlivňují. Příslušný výběr lze zajistit aplikací metody hlavních komponent s využitím znalostí o periodických složkách signálu a rozkladem signálu na jeho složku periodickou, náhodnou a složku trendu.

Problematiku predikce časových řad lze dále řešit pomocí umělých neuronových sítí se zvolenou strukturou a nelineárními přenosovými funkcemi. Podobně jako v uvedeném případě

Řád	*Délka predikce [hod]*				
	2	*4*	*6*	*8*	*10*
2	6.0	10.7	14.6	17.1	18.0
4	5.8	10.1	13.9	16.4	17.3
6	5.7	10.0	14.0	16.7	17.6
8	5.5	8.7	11.7	13.4	13.7
10	4.5	6.2	7.5	8.5	8.8
12	4.4	6.3	7.9	8.9	9.1
14	3.7	5.0	6.2	7.1	7.6

Tabulka 1: Střední chyba (v procentech) pro predikci dat spotřeby plynu s periodou vzorkování $T_s = 2$ hodiny v závislosti na řádu modelu a délce predikce

Řád	*Délka predikce [dny]*				
	1	*2*	*3*	*4*	*5*
2	8.4	13.0	16.0	18.1	18.6
4	8.2	12.4	14.6	15.9	16.5
6	7.9	11.8	14.2	15.8	16.5
8	7.3	10.5	12.9	14.7	15.7
10	7.2	10.4	12.4	14.1	14.9
12	7.2	10.2	12.3	14.0	14.9
14	7.1	10.2	12.1	13.9	14.8

Tabulka 2: Střední chyba (v procentech) pro predikci dat spotřeby plynu s periodou vzorkování $T_s = 1$ den v závislosti na řádu modelu a délce predikce

lineárním se nejprve provádí optimalizace konstant modelu a hledání takových koeficientů, pro které existuje globální minimum na chybové ploše analogické k ploše uvedené na obr. 2. Výsledný nelineární model lze pak použít k predikci časové řady.

5 Závěr

Příspěvek presentuje základní obecné matematické postupy analýzy, modelování a predikce stacionárních i nestacionárních časových řad se stanovením příslušných intervalů spolehlivosti. Postupy presentované v příspěvku pro modely lineární lze dále modifikovat pro případ modelů nelineárních, které v řadě případů využívají struktury umělých neuronových sítí. Tyto postupy budou presentovány v příspěvku následujícím.

Reference

[1] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis: Forecasting and Control.* Prentice Hall, Engelwood Cliffs, U.S.A., third edition, 1994.

[2] P. Denbigh. *System Analysis and Signal Processing.* Addison Wesley, 1998.

[3] M. H. Hayes. *Statistical Digital Signal Processing and Modeling.* John Wiley & Sons, Inc., New York, 1996.

[4] S. Haykin. *Neural Networks.* IEEE Press, New York, 1994.

[5] L. Ljung. *System Identification.* Prentice Hall, Englewood Cliffs, N.J., 1987.

[6] B. Mulgrew, P. Grant, and J. Thompson. *Digital Signal Processing, Concepts and Applications.* Macmillan Press, Ltd, Houndmills, Hampshire, U.K., 1999.

[7] A. Procházka, J. Uhlíř, P. J. W. Rayner, and N. G. Kingsbury, editors. *Signal Analysis and Prediction.* Birkhauser, Boston, U.S.A., 1998.

[8] J. Uhlíř and P. Sovka. *Číslicové zpracování signálů.* Vydavatelství ČVUT, Praha, 1995.