

PREDIKCE V ČASOVÝCH ŘADÁCH SPOTŘEB ELEKTRICKÉ ENERGIE

RADIM KRAUT, PAVEL NEVŘIVA, DANIELA JARUŠKOVÁ

ABSTRAKT. Tento příspěvek se zabývá vytvořením a použitím predikčního algoritmu v analýze časových řad. Problematika je zaměřena na oblast energetického průmyslu. Poté, co byly analyzovány rozsáhlé datové soubory získané měřením spotřeb elektrické energie v určitém časovém intervalu, byl tvořen vlastní predikční algoritmus založený na metodách klasické matematické statistiky a teorie pravděpodobnosti ve spojení s moderními softwarovými prostředky. V článku je popsán postup provedené analýzy dat, výsledky chování, kvalitativní a kvantitativní hodnocení programu, který je určen pro prognózy hodinových spotřeb elektrické energie na základě průběhu minulých hodnot. Zejména je zde rozebrána problematika zpracování, ošetření a přípravy velkého množství hodnot měřených spotřeb elektrické energie. Bez těchto kroků patřících do oblasti tzv. počáteční analýzy by aplikace jakýchkoliv dalších metod užívaných pro analýzu časových řad či predikci byla naprosto neúčelná. Vytvořený jednoduchý predikční model užívající získanou autokorelační funkci se pak stává beta verzí dalšího modelu, který už patří do oblasti ARIMA modelování.

Abstract. This contribution concerns the creation of the algorithm for prediction of the consumption of the electricity. The method is based on classical statistical tools, however, it deeply depends on the analysis of vast real data. Aside the model itself the data and their analysis are described in details.

Резюме. В этой статье описано создание алгоритма для предсказания потребления электроэнергии. Не смотря на то что наш метод использует классические статистические методы и инструменты, он глубоко зависит от анализа крупномасштабных реальных данных. Мимо модели данные и их статистическая обработка описаны детально.

1. NÁHLED DO PROBLEMATIKY

Algoritmy lineární a nelineární predikce signálu jsou důležité v mnoha informačních a řídicích systémech. Metody obvykle používané pro predikci zahrnují regresní a autoregresní modelování signálu známé z analýzy časových řad. V poslední době se též k predikování s úspěchem používají neuronové sítě. Všechny tyto postupy vycházejí z předpokladu, že v průběhu sledování spotřeb se parametry modelu nemění.

Adaptivní technika – neboli modely s proměnlivými parametry konstruuji složky časové řady pomocí takových charakteristik, které mění v průběhu doby své hodnoty, tedy nepředpokládají stabilitu analytického tvaru ani strukturálních parametrů modelu v čase.

Námi navržený predikční algoritmus je založen na klasických statistických metodách. Při jeho vytváření jsme se snažili využít všech poznatků, které bylo možno získat z analýzy již naměřených dat – odhadu denního chodu, závislosti na teplotě

2000 *Mathematics Subject Classification.* Primary 62M20.

Klíčová slova. Časové řady, predikce, regrese, autokorelační funkce, Boxova-Jenkinsova metodologie.

Práce byla podporována granty GAČR 201/00/0769, GAČR 102/98/1259 a MSM 210000001.

a závislosti po sobě jdoucích hodnot. Při modelování této poslední závislosti byla použita Boxova-Jenkinsova metodologie.

2. ROZBOR PŘÍPRAVY A ZPRACOVÁNÍ DAT

Předmětem zpracování byly datové soubory:

- naměřené hodnoty hodinových spotřeb elektrické energie a přídatných parametrů v období od 1.1.1998 do 31.12.1998, zdrojová data ve formátu *.dbf
- naměřená minima, maxima a průměrné denní teploty z vybraných měřících míst
- soubory s telegramy HDO pro stanovené časové období

Data o spotřebách elektrické energie byla shromažďována a dodána technickým dispečinkem nejmenované společnosti. Tato data představují průměrné hodinové spotřeby elektrické energie měřené na určených územních jednotkách v průběhu několika let. Pro podrobnější statistické zpracování byla zvolena rozvodna Šumperk pro datově kompletní rok 1998.

Druhým uceleným souborem dat, který byl použit, jsou průměrné denní teploty, měřené rovněž v určených lokalitách.

Částečně bylo provedeno ošetření naměřených dat od deterministických složek (předem známé, smluvně plánované významnější odběry elektrické energie, hlášené výpadky provozu velkých výrobních podniků a firem atd.).

Vzhledem k tomu, že data byla v dodané formě prakticky nepoužitelná k další analýze, byla vytvořena databázová aplikace pro výběr, zpracování a přípravu těchto dat. Toto databázové prostředí určené ke spolupráci s predikčním algoritmem bylo vyvinuto v SW Borland Delphi a to v lokální formě na PC v laboratoři VŠB-TU a pracuje pod operačním systémem Windows.

V další fázi byla navržena testovací verze výše zmíněného algoritmu, který spolupracuje s vytvořenou bankou dat a má sloužit jako zdroj krátkodobé předpovědi spotřeby elektrické energie. Spojením těchto dvou hlavních částí řešení vznikla ucelená aplikace.

Vlastnosti databázové aplikace

- Přehledně v tabulce zobrazí naměřené datové údaje pro požadované datum.
- Nástroje dovolují zobrazit pouze určité položky (činné výkony, řezy, průměry...).
- Pro zvolené období a dané měřící místo generuje soubor dat, který je možno zobrazit v přehledném grafu, nebo uložit do souboru požadovaném formátu.
- Generovaný soubor je možno využít pro další zpracování (predikce provedena v jiných aplikacích, další grafické a matematické analýzy atp.).

Dílní problémy při zpracování dat

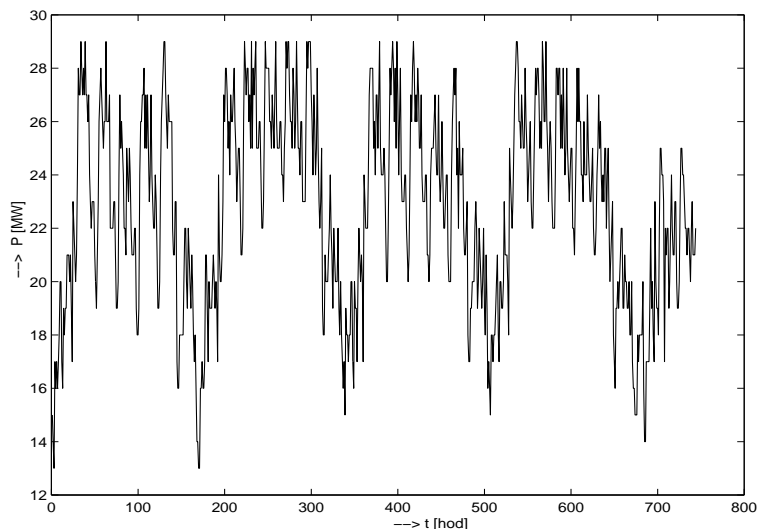
- Původní dodané datové soubory byly velmi obsáhlé a nebylo možno je v dané formě zpracovat kvůli technickým problémům (spojených s používaným softwarem). Pro hlavní řešenou část byla celá řada rozdělena na jednotlivé kalendářní měsíce i když by asi bylo výhodnější využít pro řešení celého projektu rozdělení na tzv. „topnou“ a „netopnou“ sezónu – bude diskutováno dále.
- Velmi obtížně, a ne zcela úspěšně, proběhlo ošetření zdrojových dat od předem známých deterministických hodnot. Zde šlo především o neochotu poskytovatele podkladových údajů dodat ucelené a kvalitní informace a také o nedostatek znalostí pro podrobné a přesné dekódování telegramů hromadného dálkového ovládání (HDO).

- Náležitá pozornost byla věnována měsícům, ve kterých dochází k výraznějším změnám celkové spotřeby, ať už z důvodu rychlé změny venkovní teploty (přechody mezi čtyřmi ročními obdobími) nebo, jak už bylo dříve řečeno, měsícům s výraznými výkyvy v důsledku Vánoc, Velikonoc a přelomu roku. Měsíce byly z hlediska závislosti na teplotě rozděleny přibližně na dvě poloviny. Do jedné skupiny patřily měsíce s výraznější závislostí na teplotě (větší korelační koeficient), které byly zařazeny do tzv. „topné sezóny“ a do druhé skupiny pak spadaly měsíce s menší teplotní závislostí tzv. skupina „netopné sezóny“. Podle tohoto rozdělení pak bylo prováděno ošetření na teplotu.
- Určitá komplikace nastala i v případě přípravy souborů obsahujících údaje o naměřených venkovních teplotách. Naše zpracování bylo provedeno pro hodinové spotřeby elektrické energie v jednotlivých dnech kalendářního měsíce, resp. kalendářního roku, avšak teploty byly měřeny a zaznamenány jen jako průměrné, minimální a maximální hodnoty v daném dni. Řešením bylo užití průměrné teplotní hodnoty a její rozložení do 24 hodin každého dne, čímž mohlo dojít při dalším zpracování k určitým nepřesnostem, které však zatím nebyly podrobeny další analýze.

3. VYTVOŘENÍ STOCHASTICKÉHO MODELU

Průběh spotřeby elektrické energie na Severní Moravě se stejně jako v celé ČR mění v průběhu roku hlavně v závislosti na změnách počasí. V časovém horizontu jednoho týdne je spotřeba rozdílná v průběhu pracovních dní a o víkendu. V jednodenním horizontu má spotřeba elektrické energie svůj charakteristický průběh, který je odlišný během všedních a během víkendových dní.

Vzhledem k tomu, že soubory hodinových spotřeb byly značně rozsáhlé, celý soubor dat byl rozdělen na jednotlivé měsíce a každý měsíc zpracováván odděleně. Obrázek 1 ukazuje průběh hodinové spotřeby v měsíci březnu.

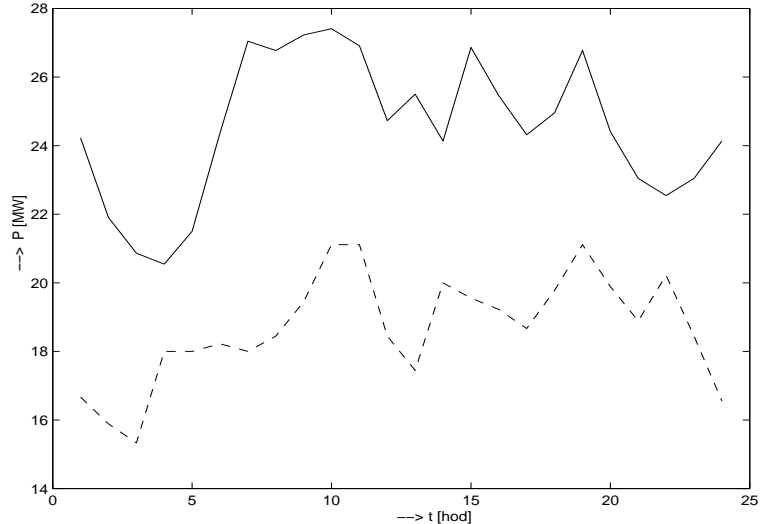


Obr. 1 *Hodinové spotřeby v měsíci březnu.*

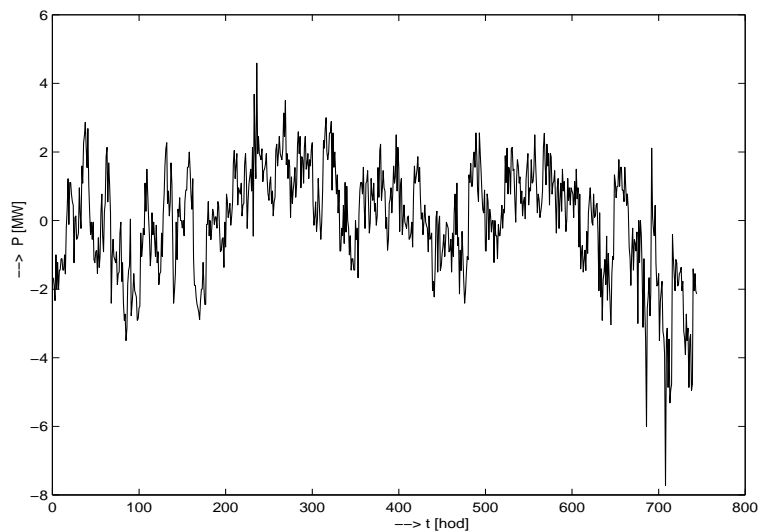
V každém měsíci byly dny rozděleny na všední a víkendové. Poté byly spočteny odděleně pro všední i víkendové dny hodinové průměry, a tak získán odhad denního chodu spotřeb. Tento odhadnutý denní chod může sloužit jako primární nástroj

k predikci v daném měsíci. Na obrázku 2 je možno vidět denní chod spotřeby ve všední dny a o víkendech v měsíci březnu.

Je zřejmé, že i takto jednoduchý model obsahuje $24 \cdot 2 \cdot 12$ parametrů. V budoucnu by bylo třeba vzít do úvahy, zda některé měsíce nemají podobný denní chod a zda by se jejich sloučením nemohl počet parametrů snížit. V každém případě by se měl provést podrobnější rozbor denního chodu spotřeby.



Obr. 2 Denní chod spotřeby ve všední dny a o víkendech v měsíci březnu.



Obr. 3 Březnová řada spotřeb očištěná od denního chodu.

4. JEDNODUCHÝ PREDIKČNÍ ALGORITMUS

Jednoduchý predikční algoritmus vychází z toho, že řada očištěná od denního průběhu, to je řada v_t , kde byl od spotřeb odečten průměr odpovídající dané hodině, je již stacionární řada. Pro takto očištěnou řadu byla spočtena výběrová autokorelační funkce:

$$r_k = \frac{\frac{1}{n} \sum_{i=1}^{n-k} (v_{i+k} - \bar{v})(v_i - \bar{v})}{\frac{1}{n} \sum_{i=1}^n (v_i - \bar{v})^2}, \quad k = 1, \dots, m.$$

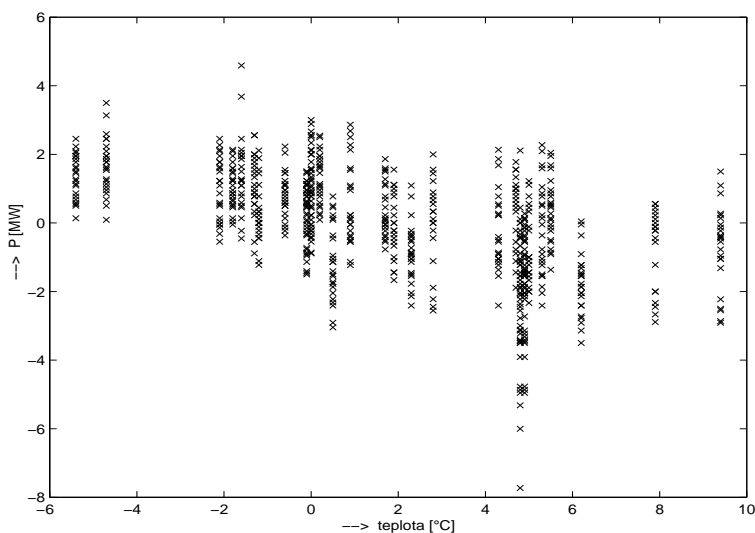
Predikce o k hodin dopředu byla počítána podle vztahů:

$$\begin{aligned} \hat{v}_{t+k} &= v_t \cdot r_k, \\ \hat{X}_{t+k} &= \hat{v}_{t+k} + m_{hod}, \end{aligned}$$

kde m_{hod} je průměr odpovídající dané hodině. Tento algoritmus byl navržen Prof. Nevřivou. Jeho výhodou je velká jednoduchost. Pro předpověď o jeden krok dopředu je tento algoritmus stejný jako algoritmus pro $AR(1)$ posloupnost, jestliže je autoregresní parametr odhadován momentovou metodou.

5. ZÁVISLOST NA TEPLOTĚ

Ze zkušenosti je známo, že výše teploty významně ovlivňuje spotřebu elektrické energie. V našem případě jsme však zpracovávali jednotlivé měsíce odděleně, a tudíž se vliv teploty neprojevil tak výrazně. Nejvýrazněji se projevil v měsících na přelomu sezón (např. v březnu), kde na počátku bývá ještě chladněji, zatímco ke konci měsíce se již otepluje a spotřeba je tudíž nižší.

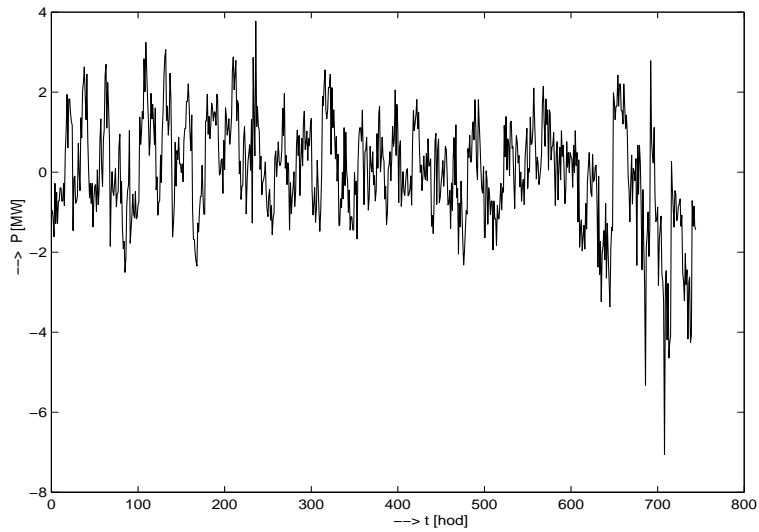


Obr. 4 Závislost mezi teplotou a odpovídajícími očištěnými hodinovými hodnotami spotřeb.

Měsíce byly z hlediska závislosti na teplotě rozděleny přibližně na dvě poloviny. Jednu skupinu tvořily měsíce s výraznější závislostí na teplotě (vyšším korelačním koeficientem), které byly zařazeny do tzv. „topné sezóny“ a do druhé skupiny pak spadaly měsíce s menší teplotní závislostí tzv. skupina „netopné sezóny“. Pro měsíce v „topné sezóně“ jsme použili model jednoduché lineární regrese, kde průměrná denní teplota v odpovídajícím dni T_t hrála roli nezávisle proměnné:

$$v_t = a + b \cdot T_t + er_t;$$

$\{er_t\}$ jsou náhodné chyby. Metodou nejmenších čtverců jsme získali odhady \hat{a} a \hat{b} parametrů a a b a spočetli jsme residua $Y_t = v_t - \hat{a} - \hat{b} \cdot T_t$. Pro měsíc březen jsme získali odhady $\hat{a} = 0.4432$ a $\hat{b} = -0.2325$.



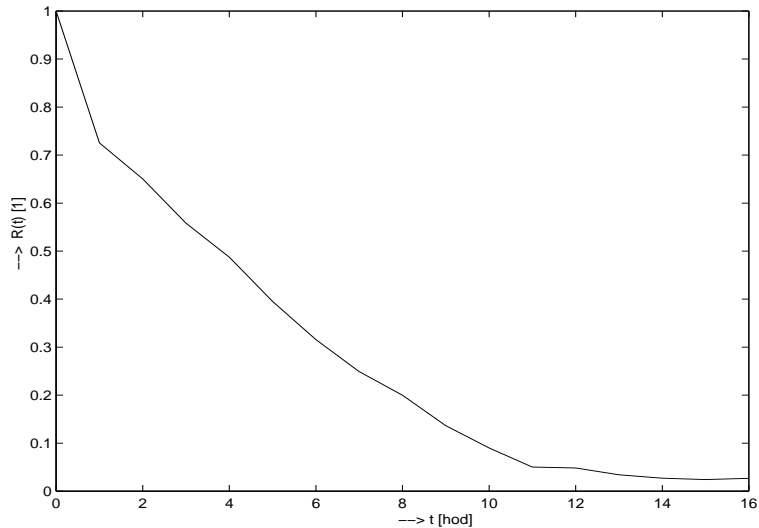
Obr. 5 Residua $\{Y_t\}$ pro měsíc březen.

Pro „topnou sezónu“ označují tedy Y_t residua, zatímco pro „netopnou sezónu“ $Y_t = v_t$.

6. ZÁVISLOST MEZI PO SOBĚ JDOUCÍMI HODNOTAMI

Závislosti mezi po sobě jdoucími hodnotami v časové řadě residuí $\{Y_t\}$ jsme modelovali pomocí ARIMA posloupností. Vhodný model byl vybrán

- pomocí analýzy autokorelační a parciální autokorelační funkce – viz Boxova-Jenkinsova metodologie,
- pomocí informačního Akaikeho kritéria.

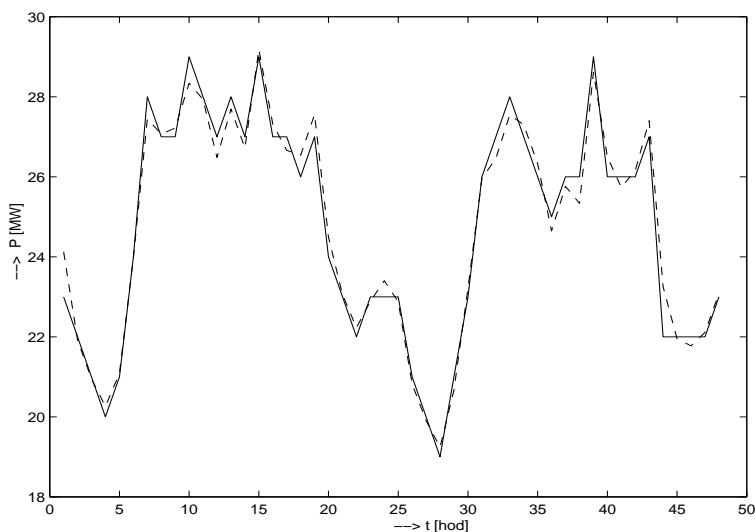


Obr. 6 Autokorelační funkci pro březnová residua $\{Y_t\}$.

Odhad parametrů vhodného ARIMA modelu byl proveden pomocí podmíněné metody nejmenších nelineárních čtverců, která je součástí Statgraphicsu 4.2. Pro březnová residua byl například vybrán model ARIMA(0,1,1):

$$Y_t = Y_{t-1} + e_t - 0.41308 \cdot e_{t-1}.$$

Postupně vytvořený stochastický model může být pak použit pro predikci. Obrázek 7 ukazuje původní řadu spolu s řadou, která vznikne predikcí o jeden krok dopředu, pro období od pondělí 2. 3. 1998 do úterý 3. 3. 1998. Je třeba však poznamenat, že shoda na obrázku je příliš optimistická, protože tatáž data byla použita na vytvoření modelu a pak opět použita pro predikování, což je v rámci matematické statistiky nepřipustné. Přesto však tento obrázek dává jistý názor o tom, kolik informace se nám podařilo z dat vytěžit.



Obr. 7 Původní a predikovaná řada pro období 2. – 3. 3. 1998.

7. ZÁVĚR

Kvalita predikce je dána za prvé množstvím informace, kterou má statistik k dispozici při vytváření modelu, a za druhé kvalitou modelu, který byl k predikování vybrán.

V našem případě nám výrazným způsobem chyběla informace o smluvních dodávkách elektrické energie. V předem dohodnutých dobách totiž dochází k dočasné předem domluvené spotřebě v podnicích a firmách. O těchto náhlých nárůstech případně poklesech jsme měli jen velmi sporé informace.

Při vytváření modelu jsme zpracovávali data o spotřebě pro jednotlivé měsíce odděleně. Hlavní důvod pro toto rozdělení byl čistě technický. Používali jsme totiž pro zpracování starou verzi Statgraphicsu, který neumožňuje zpracovávat delší časové řady. Ve skutečnosti je však toto rozdělení umělé. Výhodou rozdělení řady na menší úseky je lepší pochopení chování řady. Nevýhodou je na druhé straně neúměrný nárůst počtu parametrů. Připomeňme, že pro predikci je třeba znát 24 hodinových průměrů pro všední dny, 24 hodinových průměrů pro víkendové dny, dále 2 parametry lineární regrese pro závislost na teplotě (pouze pro měsíce v „topné sezóně“) a dále parametry příslušné ARIMA posloupnosti.

Po teoretické stránce je náš postup při vytváření modelu jednoduchý. V praxi však situace není jednoduchá. Před tím, než bylo započato s vytvářením modelu musela být provedena počáteční analýza, která umožní další zpracování (vyloučit svátky i různé problémy s kalendářem, vyloučit několik jasně chybných pozorování, atd.). Praktický problém spočívá i v tom, že je velmi obtížné najít software, kde by bylo možno provést celou analýzu. Příprava a kontrola dat v praxi tvoří více než polovinu vlastního zpracování dat.

Náš model je jenom jedním z možných modelů, které lze k predikování použít. Bylo by jistě zajímavé se celým problémem zabývat hlouběji, vytvořit několik alternativních modelů a jejich kvalitu porovnat.

LITERATURA

- [1] Anderson O.D.: *Time Series Analysis and Forecasting, The Box-Jenkins approach*. Butterworth & Co, Great Britain, 1976.
- [2] Arlt J.: *Moderní metody modelování ekonomických časových řad*. GRADA Publishing, Praha, 1999.
- [3] Brockwell P., Davis R. *Time Series: Theory and Methods*. Springer Verlag, second edition 1991.
- [4] Cipra T.: *Analýza časových řad s aplikacemi v ekonomii*. SNTL/ALFA, Praha, 1986.
- [5] Hindls R., Kaňoková J., Novák I.: *Metody statistické analýzy pro ekonomy*. Management Press, Praha 1997.
- [6] Kraut R., Gelnar P.: *Determining the Accuracy of the Prediction Algorithm for Making Electric Consumption Forecasts*. Nostradamus'99, 2nd International Prediction Conference held at the Faculty of Technology in Zlín, Czech Republic, October 7–8, 1999.
- [7] Kraut R., Gelnar P., Srovnal V.: *Prediction of Electrical Energy Consumption*. PDS 2000, IFAC workshop on Programmable Devices and Systems PDS 2000, VSB – TU Ostrava, February 8–9, 2000.
- [8] Kraut R.: *New Prediction Algorithm of Electric Energy Consumption*, PhD Workshop, Cybernetics & Informatics Eurodays Young Generation Viewpoint, Mariánská, September 26–30, 2000.
- [9] Kraut R., Nevřiva P.: *Prediction of Electric Energy Consumption*, Fourth International Conference on Technical Informatics CONTI 2000, October 12–13, 2000, Timisoara, Romania.
- [10] Kraut R., Nevřiva P., Jarušková D.: *Time series analysis in prediction of electric energy consumption*, IWCIT'01, VSB–TU Ostrava, September 19–21, 2001.
- [11] Shumway R.: *Applied Statistical Time Series Analysis*. Prentice Hall. New Jersey 1998.

VŠB - TECHNICKÁ UNIVERZITA OSTRAVA, FEI, 17. LISTOPADU 15, 708 33 OSTRAVA 8
E-MAIL: radim.kraut@vsb.cz, pavel.nevriva@vsb.cz

ČVUT V PRAZE, STAVEBNÍ FAKULTA, KATEDRA MATEMATIKY, THÁKUROVA 7, 166 29 PRAHA 6
E-MAIL: jarus@mat.fsv.cvut.cz