



národní  
úložiště  
šedé  
literatury

## **Nelineární model se smíšenými efekty pro odhad spotřeby zemního plynu**

Konár, Ondřej  
2007

Dostupný z <http://www.nusl.cz/ntk/nusl-37417>

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 02.05.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní [nusl.cz](http://nusl.cz) .

# Nelineární model se smíšenými efekty pro odhad spotřeby zemního plynu

doktorand:

MGR. ONDŘEJ KONÁR

Ústav informatiky AV ČR, v. v. i.

Pod Vodárenskou věží 2

182 07 Praha 8

konar@cs.cas.cz

školitel:

DOC. ING. EMIL PELIKÁN, CSc.

Ústav informatiky AV ČR, v. v. i.

Pod Vodárenskou věží 2

182 07 Praha 8

pelikan@cs.cas.cz

obor studia:

Inženýrská informatika v dopravě a spojích

Tato práce vznikla za podpory grantu GA AV ČR č. 1ET400300513 a výzkumného záměru ÚI AV ČR č. AV0Z10300504.

## Abstrakt

Kvalitní odhad spotřeby zemního plynu v daném časovém období (např. den, měsíc atd.) může být velice užitečný pro společnost obchodující s plynem. Důvodem je skutečnost, že spotřeba některých zákazníků není měřena na denní bázi. Takové měření není zvláště v případě malých zákazníků (díky jejich vysokému počtu a nízké spotřebě) pro obchodníky s plynem přijatelné. Zatímco průměrná délka období mezi odečty u malých zákazníků je přibližně jeden rok, pro některé účely je zapotřebí větší časové rozlišení. Odhady denních spotřeb můžeme získat pomocí různých vyvíjených statistických modelů. V oddělení nelineárního modelování ÚI AV ČR je vyvíjen statistický model zvaný Gamma, který je provozně využíván v Západočeské plynárenské, a.s. (ZČP). V tomto modelu považujeme všechny koeficienty za neznámé, ale pevné veličiny. Nabytá zkušenost ale ukazuje, že použití pevných efektů může být mírně problematické zejména v oblasti odhadu individuální roční spotřeby jako jednoho z parametrů modelu. V datech se mohou vyskytovat nezanedbatelné chyby měření a proto by odhady individuálních ročních spotřeb měly být "smrštěny" k průměru zákaznického segmentu (segmenty odpovídají předdefinovaným skupinám zákazníků s podobnými vlastnostmi). Náš nový model je formulován jako nelineární regresní model se smíšenými efekty a je rozšířením stávajícího modelu Gamma. Roční spotřeba je zde brána jako náhodná veličina s logaritmicke-normálním rozdělením. Příspěvek zobrazuje současný stav výzkumu, který zdaleka není u konce. Bude též prezentováno porovnání nového přístupu se stávajícím. V závěru jsou uvedeny některé směry dalšího výzkumu.

## 1. Úvod

Spotřeba zemního plynu domácností a malooběratelů je (narozdíl od velkooběratelů a středních odběratelů, kteří jsou měřeni po hodinách) odečítána z plynoměru v intervalu přibližně jeden rok. Odečty plynoměru probíhají z technických a ekonomických důvodů cyklicky v průběhu celého roku, tzn. že je každý den v roce odečtena část zákazníků. Spotřeby jednotlivých zákazníků jsou tedy známy za přibližně roční období, která se mezi zákazníky překrývají.

V některých situacích je ale užitečné znát přesnou spotřebu k danému dni nebo za dané období. Příkladem může být změna ceny plynu, kdy potřebujeme znát pro každého zákazníka spotřebu od předchozího odečtu k datu změny ceny a spotřebu od data změny ceny k následujícímu odečtu. Dále např. při odhadu zisků obchodníka s plynem potřebujeme znát objem prodaného plynu k danému dni. Neměřené části spotřeby je třeba

odhadnout vhodným modelem. Druhá možnost je pochopitelně provedení mimořádného odečtu, to však bývá zpravidla finančně i technicky náročnější.

V oddělení nelineárního modelování byl vyvinut nelineární regresní model spotřeby zemního plynu, nazvaný Gamma, [4, 6]. Model je provozně využíván v Západočeské plynárenské, a.s. Zkušenosti s provozem tohoto modelu byly popsány např. v [1, 2]. Cílem tohoto příspěvku je popsat současný stav vývoje modelu, konkrétně jeho modifikace na bázi *nelineárních modelů se smíšenými efekty*, [3, 5].

## 2. Model Gamma

Zákazníci jsou tříděni podle tzv. *typu klienta* (domácnost nebo malooběratel), který víceméně souvisí s výší odběru, a podle tzv. *typu smlouvy*. Ten je dán způsobem užívání zemního plynu, přičemž se uvažují kombinace těchto způsobů využití: vaření, ohřev vody (TUV),

vytápění a (pouze pro maloodběr) technologie. Celkem je uvažováno 16 zákaznických segmentů (7 pro domácnosti a 9 pro maloodběr). Parametry modelu se odhadují zvlášť pro každý segment zákazníků.

Model se mírně liší pro neotopové a otopové segmenty (tj. segmenty neobsahující, resp. obsahující, zákazníky, kteří využívají zemní plyn k vytápění). Spotřebu  $Y_{ikt}$  zákazníka  $i$  z neotopového segmentu  $k$  ve dni  $t$  modelu jeme vztahem

$$Y_{ikt} = \mu_{ik}(\Psi_t e^{-\gamma_k \varphi(T_t, N_t)} + p_k) + \varepsilon_{ikt}, \quad (1)$$

kde  $\mu_{ik}$  je individuální parametr zákazníka  $i$  (v podstatě násobek jeho roční spotřeby),  $\Psi(t)$  je sezónní složka modelu společná pro všechny segmenty (odhadnutá v minulosti z agregovaných dat za celou ZČP, v současnosti k ní přistupujeme jako k pozorované),  $\gamma_k$  je parametr určující míru teplotní závislosti spotřeby v segmentu  $k$ ,  $p_k$  je stálá složka spotřeby zákazníků ze skupiny  $k$ ,  $T_t$  je skutečná a  $N_t$  normálová teplota ve dni  $t$ . Funkce  $\varphi$  je rozdíl teplot (skutečné a normálové) oříznutý pro hodnoty vyšší než 14°C. K tomuto oříznutí přistupujeme proto, že teplotní závislost spotřeby je porovávána pouze v teplotách nižších než 14°C. Člen  $\varepsilon_{ikt}$  je považován za náhodnou veličinu s nulovou střední hodnotou a rozptylem úměrným systematické složce spotřeby. Nejsou kladeny žádné předpoklady o rozdělení  $\varepsilon_{ikt}$ .

Pro otopové segmenty je model mírně modifikován:

$$Y_{ikt} = \mu_{ik} [I_t(\Psi_t e^{-\gamma_k \varphi(T_t, N_t)} + p_k) + (1 - I_t)q_k] + \varepsilon_{ikt}, \quad (2)$$

kde  $q_k$  je konstanta určující průměrnou výši spotřeby v “létě” a  $I_t = 1$ , je-li průměrná teplota za poslední tři dny (tj. za dny  $t, t-1, t-2$ ) nižší než 14°C. V opačném případě je  $I_t = 0$ .

Navrhovaná modifikace Gamma modelu spočívá ve znárodnění individuálního parametru  $\mu_{ik}$ . V současné fázi je tato modifikace spojena i se změnou předpokladů o náhodném členu  $\varepsilon_{ikt}$ . Pro jednoduchost budeme v následujícím textu uvažovat pouze modifikaci modelu (1). Modifikace modelu (2) pro otopové segmenty je zcela analogická.

### 3. Model se smíšenými efekty

#### 3.1. Popis modelu

Pro spotřebu  $Y_{ikt}$  zákazníka  $i$  z (neotopového) segmentu  $k$  ve dni  $t$  uvažujeme nový model

$$Y_{ikt} = (\mu_k e^{b_{ik}})(\Psi_t e^{-\gamma_k \varphi(T_t, N_t)} + p_k) + \varepsilon_{ikt}, \quad (3)$$

kde  $\mu_k$  je pevná část individuální složky (společná pro celý segment),  $b_{ik}$  je náhodná veličina, u níž předpokládáme normální rozdělení s nulovou střední hodnotou a rozptylem  $\sigma_b^2$ , z čehož plyne logaritmicke-normální rozdělení nového individuálního “parametru”  $\mu_k e^{b_{ik}}$ . U náhodného členu  $\varepsilon_{ikt}$  předpokládáme rovněž normální rozdělení s nulovou střední hodnotou a rozptylem  $\sigma^2$ .

#### 3.2. Odhad parametrů

Parametry modelu odhadujeme ve dvou krocích:

1. Odhadneme parametry  $\beta_k = (\mu_k, \gamma_k, p_k)^T$  a  $b_{ik}$  metodou penalizovaných nelineárních nejmenších čtverců (PNLS), tj. minimalizací funkce

$$\sum_{i \in k} \left[ \|Y_{ik} - f_{ik}(\beta_k, b_{ik})\|^2 + (\hat{\Delta} b_{ik})^2 \right], \quad (4)$$

kde  $Y_{ik}$  je vektor spotřeb  $i$ -tého zákazníka ze segmentu  $k$  za příslušné období (které je ovlivněno dostupností dat – chybějící pozorování jsou vynechána),  $f_{ik}(\beta_k, b_{ik})$  je vektor hodnot regresní funkce

$$f_{ikt}(\beta_k, b_{ik}) = (\mu_k e^{b_{ik}})(\Psi_t e^{-\gamma_k \varphi(T_t, N_t)} + p_k)$$

za totéž období a  $\hat{\Delta}$  je aktuální odhad poměru  $\Delta = \sigma_b^2 / \sigma^2$ .

2. S využitím odhadů  $\hat{\beta}_k$  a  $\hat{b}_{ik}$  získaných v předchozím kroku odhadneme rozptyly  $\sigma_b^2$  a  $\sigma^2$  pomocí výběrových rozptylů  $\hat{\sigma}_b^2 = \text{var}(\hat{b}_{ik})$  a

$$\hat{\sigma}^2 = \text{var} \left[ Y_{ik} - f_{ik}(\hat{\beta}_k, \hat{b}_{ik}) \right]$$

a zopakujeme předchozí krok s aktualizovaným odhadem poměru  $\Delta$ .

Minimalizaci funkce (4) lze snadno převést na problém nelineárních nejmenších čtverců tak, že položíme  $\tilde{Y}_{ik} = (Y_{ik}, 0)^T$  a  $\tilde{f}_{ik}(\beta_k, b_{ik}) = (f_{ik}(\beta_k, b_{ik}), \hat{\Delta} b_{ik})^T$  a pak minimalizujeme funkci

$$\sum_{i \in k} \|\tilde{Y}_{ik} - \tilde{f}_{ik}(\beta_k, b_{ik})\|^2. \quad (5)$$

K tomu je již možné použít libovolný hotový software, v našem případě se jednalo o *Optimization toolbox* k *MATLABu* (konkrétně o funkci `lsqnonlin`).

Optimalizaci začínáme PNLS s počátečním odhadem  $\hat{\Delta} = 1$  a počátečními hodnotami parametrů  $\beta$ ,  $b_{ik}$  takovými, aby odpovídaly parametrům z poslední verze stávajícího Gamma modelu. Vyše uvedené dva kroky opakujeme, dokud rozdíl mezi po sobě následujícími odhady

nejsou menší než stanovená mez (v této fázi byla mez nastavena na  $10^{-7}$  pro 5 po sobě následujících odhadů). Rozdíly jsou ve většině segmentů dostatečně malé již v prvních pěti krocích. Z toho důvodu (také s ohledem na výpočetní náročnost) byl maximální počet iterací nastaven na poměrně nízkou hodnotu 20.

### 3.3. Použitá data

Výše uvedené metody vyžadují data v denním rozlišení. Byla proto použita data z náhodného výběru cca 700 zákazníků z celé ČR, jejichž spotřeby byly měřeny v období délky dvou let (říjen 2004 - září 2006) v rámci projektu "Tvorba typových diagramů dodávky" (TDD) organizovaném Českou plynárenskou unií. Data obsahují nemalé množství chybějících (nebo chybných) hodnot, které byly pro účely odhadu parametrů vypuštěny. Tento datový soubor je vhodný pro výzkumné účely. Pro případný provoz modelu je však nutné model upravit tak, aby mohly být parametry odhadovány z dat agregovaných za větší časové úseky. K dispozici máme např. soubor cca 1700 měsíčně odečítaných zákazníků získaný v rámci grantu GA AV ČR č. 1ET400300513. Minimálně odhad náhodného parametru  $b_{ik}$  je však nutné v případě provozního využití počítat přímo z pravidelných (zpravidla ročních) odečtů daného zákazníka. Pořizování dat s vyšší frekvencí odečtu (i např. měsíčních) je totiž finančně i technicky příliš náročné.

## 4. Porovnání obou modelů

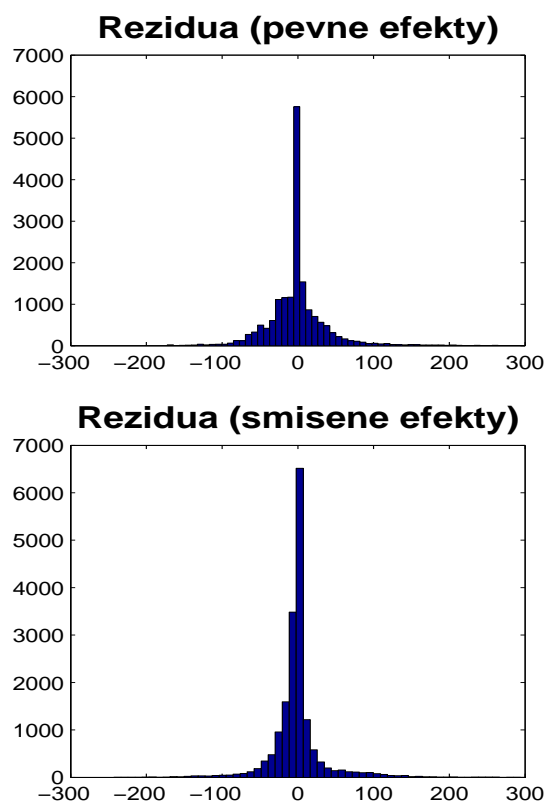
### 4.1. Rozptyl reziduí

Rozptyl reziduí z modelu se smíšenými efekty je dle očekávání nižší než rozptyl reziduí původního modelu Gamma. Ke snížení rozptylu došlo ve všech zákaznických segmentech. Na celém vzorku zákazníků došlo ke snížení rozptylu o cca 21 %. Rozptyly reziduí z obou modelů v jednotlivých segmentech zákazníků jsou uvedeny v tabulce 1.

Na obrázku 1 jsou histogramy reziduí z obou modelů pro vybraný segment zákazníků. I zde je viditelné snížení rozptylu, byť je třeba přiznat, že v ostatních segmentech toto není z histogramů patrné v takové míře.

Typ	Užití ZP	Rozptyl rez.		Rozdíl (%)
		Gamma	NLME	
D	O	24.7043	16.6759	32.4977
D	T	25.3028	10.6976	57.7215
D	O+T	77.9285	60.8689	21.8913
D	O+V	20.0326	13.0108	35.0518
D	V+T	1.7415	0.681	60.8984
D	O+T+V	24.9684	20.8241	16.5981
M	T	510.0277	331.0669	35.0884
M	O+T	636.4539	527.0364	17.1917
M	V	73.5877	61.4379	16.5107
M	O+V	241.1401	193.8942	19.5927
M	V+T	625.567	625.5272	0.0064
M	O+T+V	803.7484	603.7964	24.8774
M	Tc	1914.7845	1420.1211	25.8339
M	Tc+O	1328.3279	1158.1101	12.8144
Total		479.2681	378.6474	20.9947

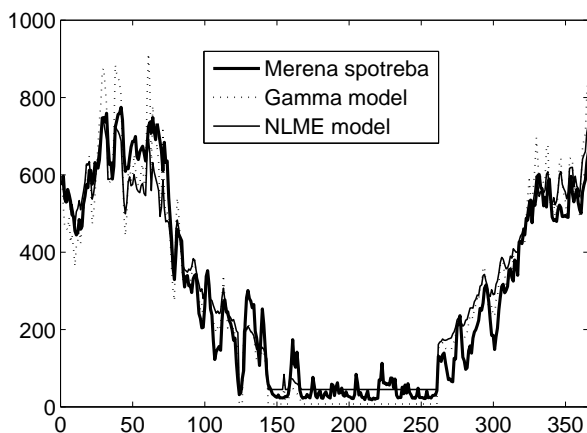
**Tabulka 1:** Rozptyl reziduí ve vybraných segmentech, D – domácnost, M – malooběr, O – vytápění, T – ohřev vody, V – vaření, NLME – model se smíšenými efekty, Gamma – model s pevnými efekty.



**Obrázek 1:** Histogram reziduí stávajícího modelu Gamma a modelu se smíšenými efekty – segment malooběr-technologie

#### 4.2. Časový průběh odhadů

Na obrázku 2 vidíme porovnání časových průběhů původního Gamma modelu a nového modelu se smíšenými efekty. Lze pozorovat mírné vylepšení odhadu v případě modelu se smíšenými efekty. Jedná se o segment otopový. Průběh v ostatních otopových segmentech je podobný. V neotopových segmentech je vylepšení rovněž pozorovatelné, avšak rozdíl je o něco větší. Na uvedený obrázek je však třeba pohlížet s určitou rezervou, neboť v této chvíli není zcela jasné, jaký podíl na zlepšení má volba modelu a jaký volba optimalizačních metod (parametry původního Gamma modelu byly odhadovány metodou váženého nejmenšího součtu absolutních hodnot reziduí, narozdíl od penalizovaných nejmenších čtverců použitých v případě modelu se smíšenými efekty.)



**Obrázek 2:** Porovnání modelových spotřeb a měřené spotřeby za rok 2005 (na ose  $x$  jsou dny tak, že 1=1.1.2005), segment domácnosti-vytápění

#### 5. Závěr a výhledy

Byl prezentován nelineární regresní model pro odhad spotřeby zemního plynu domácností a maloodběratelů a především jeho modifikace na bázi nelineárního regresního modelu se smíšenými efekty. Oba přístupy (pevné a smíšené efekty) byly porovnány dle vybraných kritérií, přičemž bylo pozorováno určité vylepšení odhadu v případě modelu se smíšenými efekty.

V rámci dalšího výzkumu je třeba prošetřit zejména ná-

sledující oblasti:

1. Úprava metod odhadování tak, aby mohla být využívána data z měsíčních, resp. ročních odečtů (alespoň pro individuální parametr zákazníka).
2. Analýza předpokladu o konstantním rozptylu reziduí, prozkoumat možnosti proměnlivosti rozptylu v čase či závislosti na výši spotřeby.
3. Analýza možností modifikace sezónní složky modelu (v současné době považované za pozorovanou).
4. Analýza teplotní závislosti spotřeby (případná modifikace závislostní funkce).

#### Literatura

- [1] J. Čermáková, J. Matějovic, O. Naxerová, J. Bečvář, M. Brabec, T. Brabec, O. Konár, M. Malý, E. Pelikán, M. Šimůnek, J. Vondráček "Matematické modelování spotřeby zemního plynu zákazníků bez průběhového měření", *Plyn*, č. 2/2005, ISSN 0032-1761, pp. 34–37, 2005.
- [2] J. Čermáková, O. Volfová-Naxerová, J. Bečvář, "Stanovení objemu nevyfakturovaného plynu pomocí virtuální fakturace a matematického modelu GAMMA", *Plyn*, č. 4/2005, ISSN 0032-1761, pp. 76–79, 2005.
- [3] M. Davidian, D. M. Giltinan, "Nonlinear Models for Repeated Measurement Data", Chapman & Hall, London, 1995.
- [4] O. Konár, "Matematické modelování spotřeby zemního plynu domácností a maloodběratelů", *sborník konference ROBUST 2006*, ISBN 80-7015-073-4, pp. 151–158, 2006.
- [5] J. C. Pinheiro, D. M. Bates, "Mixed-Effects Models in S and S-PLUS", Springer-Verlag, New York, 2000.
- [6] J. Vondráček, E. Pelikán, O. Konár, J. Čermáková, K. Eben, M. Malý, M. Brabec, "A statistical model for the estimation of natural gas consumption", *accepted for publication in Applied Energy*