

Predikce roční spotřeby zemního plynu po ceníkových pásmech

Ondřej Konár, Marek Brabec, Ivan Kasanický,
Marek Malý, Emil Pelikán

Ústav informatiky AV ČR, v.v.i.

ROBUST 2014
Jetřichovice 20. ledna 2014

Otevřený trh s plynem

Účastníci trhu s plynem

Účastníci trhu s plynem:

- přepravci a distributoři (méně)
- obchodníci (více)
- zákazníci (nejvíce)
- operátor trhu (OTE, a.s.)
- regulátor trhu (ERÚ)

Otevřený trh s plynem

Podmínky

- Otevřený trh – více svobody, ale i více byrokracie
- Činnost distributorů je regulována
(omezení zneužití monopolního postavení)
- Mnoho regulačních výkazů
(vyhl. 59/2012 Sb. o regulačním výkaznictví)

Regulační výkazy

Objekt našeho zájmu – výkaz tarifních statistik

- Výkaz tarifních statistik (výkaz 22-Bp) obsahuje plán:
 - ① počtu zákazníků v každém ceníkovém pásmu,
 - ② celkové spotřeby v každém ceníkovém pásmu.

V obou případech na následující kalendářní rok.

- Ceníkové pásmo se přiřazuje každému zákazníkovi při každé fakturaci podle jeho spotřeby.

Formulace úlohy

Predikované veličiny

- Označme

$p = 1, \dots, P$ ceníková pásma,

$\{N_{pt}\}$ časovou řadu počtů zákazníků v pásmu p ,

$\{S_{pt}\}$ časovou řadu celkových ročních spotřeb v pásmu p .

- V roce t chceme predikovat $N_{p(t+1)}$ a $S_{p(t+1)}$ pro všechna p .
- (Zdánlivě) standardní úloha.

Problémy

... které úlohu komplikují

- 1 Veličiny N_{pt} a S_{pt} nejsou nezávislé.
- 2 Spotřeba plynu závisí na teplotě – třeba zohlednit v predikčním modelu.
- 3 Ceníkové pásmo p závisí na spotřebě zákazníka – migrace mezi pásmy.
- 4 Cyklické odečty – nepřekrývající se fakturační období.

Predikční model

Základní idea

Model má dvě úrovně:

- 1 Predikce počtu – s uvažováním migrace zákazníků mezi pásmy
- 2 Predikce spotřeby – v závislosti na počtu, za podmínky normálové teploty

Data

(aneb z čeho můžeme vařit)

- 1 Fakturační data zákaznického kmene RWE (2007–2012)
- 2 Průměrné denní teploty (skutečné i normálové)
- 3 Historické výkazy (plán i skutečnost)

Zpracování dat

Model TDD

Spotřebu Y_{ikd} zákazníka i třídy TDD k pro ve dni d odhadneme modelem TDD:

$$\hat{Y}_{ikd} = \mu_{ik} \Phi_{kd}(T_d),$$

kde

μ_{ik} je individuální parametr zákazníka i určující globální (časově nezávislou) hladinu jeho spotřeby,

$\Phi_{kd}(T_d)$ je systematická část modelu, společná pro třídu k , závislá na teplotě průměrné venkovní teplotě T_d .

Zpracování dat

Normalizace

- Fakturační spotřebu $Y_{ik\Delta}$ za období Δ transformujeme dle vzorce:

$$\hat{Y}_{ik\Omega}^{(N)} = \frac{Y_{ik\Delta}}{\sum_{d' \in \Delta} \Phi_{kd'}(T_{d'})} \sum_{d \in \Omega} \Phi_{kd}(T_d^{(N)})$$

- Položíme-li $Y_{ik\Delta} = \sum_{d \in \Delta} \hat{Y}_{ikd} = \mu_{ik} \sum_{d \in \Delta} \Phi_{kd}(T_d)$, získáme odhad $\hat{\mu}_{ik} = \frac{Y_{ik\Delta}}{\sum_{d \in \Delta} \Phi_{kd}(T_d)}$.
- $\hat{Y}_{ik\Omega}^{(N)}$ je tak odhadem spotřeby daného zákazníka za období Ω a za podmínky normálové teploty $T_d^{(N)} (\forall d \in \Omega)$.

Zpracování dat

Přřazení ceníkového pásma

- Pro každou fakturaci spočteme $\hat{Y}_{ik\Omega}^{(N)}$, kde za období Ω zvolíme posledních 365 dní před fakturačním datem.
- Na základě této hodnoty přiřadíme ceníkové pásmo (hranice z vyhlášky o pravidlech trhu s plynem).
- Pro každého zákazníka získáme historii migrace mezi ceníkovými pásmy (za podmínky normálové teploty).

Model migrace mezi pásmy

Maticе pravděpodobností přechodu

- Ceníkové pásmo zákazníka i v čase t lze chápat jako náhodnou veličinu X_{it} .
- Pro X_{it} předpokládáme Markovskou vlastnost a definujeme matici

$$\mathbf{P} = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,13} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,13} \\ \vdots & & \ddots & \vdots \\ p_{13,1} & p_{13,2} & \cdots & p_{13,13} \end{pmatrix},$$

kde $p_{m,l} = p(l|m)$; $m, l = 1, \dots, 13$, značí pravděpodobnost přechodu z pásma m do pásma l .

Model migrace mezi pásmy

Predikce počtu (v ideálním případě)

- Z Markovské vlastnosti (mj.) dostaneme

$$P[X_{it} = l] = \sum_{m=1}^{13} p(l|m)P[X_{i(t-1)} = m],$$

- Ozn. $p_{it} = (P[X_{it} = 1], P[X_{it} = 2], \dots, P[X_{it} = 13])$, pak pro každé t platí

$$p_{it} = p_{i(t-1)}\mathbf{P},$$

- Označíme-li $N_{\bullet t}$ celkový počet zákazníků v čase t .
- Odhadneme $P[X_{it} = p]$ relativní četností $N_{pt}/N_{\bullet t}$.
- Tím dostaneme predikci počtů „zdarma“.

Model migrace mezi pásmy – problémy I

Vznik a zánik zákazníků

- $N_{\bullet t}$ není konstatní (zákazníci vznikají a zanikají).
- Zánik lze vyřešit definicí virtuálních pásem:
 - pásmo 0 – tam odcházejí končící zákazníci,
 - pásmo 14 – tam odcházejí zákazníci s příliš vysokou spotřebou.
- Vznik zákazníků je třeba řešit zvlášť (později).

Model migrace mezi pásmy – problémy II

Cyklické odečty – nejednotný čas

- Zákazníci jsou odečítáni v různých časech, tzn. časy $t, t + 1, \dots$ jsou pro různé zákazníky různé.
- **Řešení:** uvažujeme četnosti pásem po dnech.
- **Důsledek:**
 - 1 čas je jednotný (to je dobře),
 - 2 proces přestává být Markovský (to není dobře).
- Markovský přístup se dá použít, když se vzdáme předpokladu homogenity, tzn. máme $p_t = p_{t-1} \mathbf{P}_t$.

Transformace dat

Statistiky počtů

Pro každý den d napočteme:

- vektor počtů zákazníků v ceníkových pásmech

$$n_d = (n_{d,0}, \dots, n_{d,14})$$

- matice četností přechodu zákazníků odečtených ve dni d

$$\mathbf{N}_d^O = \begin{pmatrix} n_{d,0,0}^{(O)} & n_{d,0,1}^{(O)} & \cdots & n_{d,0,14}^{(O)} \\ n_{d,1,0}^{(O)} & n_{d,1,1}^{(O)} & \cdots & n_{d,1,14}^{(O)} \\ \vdots & & \ddots & \vdots \\ n_{d,14,0}^{(O)} & n_{d,14,1}^{(O)} & \cdots & n_{d,14,14}^{(O)} \end{pmatrix}$$

Model migrace zákazníků

Odhad matic pravděpodobností přechodu

- K diagonále matice \mathbf{N}_d^O přičteme počty zákazníků, kteří ve dni d nebyli odečtení.
- Vydělíme řádky jejich součty, tím získáme odhad pravděpodobností.
- Volitelný parametr – časová agregace – jako rozumný kompromis volíme časový krok měsíc.
- Předpokládáme roční periodicitu pravděpodobností přechodu.

Model migrace mezi pásmy

Vznik nových zákazníků

- Přírůstky odhadujeme rovněž empiricky.
- I zde předpokládáme roční periodicitu.

Model migrace mezi pásmy

Algoritmus – část 1

- Pro každý kalednářní měsíc m napočteme

$$\mathbf{N}_m = \begin{pmatrix} \sum_{d \in m} n_{d,0,0}^{(O)} + n_{m,0}^{(N)} & \sum_{d \in m} n_{d,0,1}^{(O)} & \cdots & \sum_{d \in m} n_{d,0,14}^{(O)} \\ \sum_{d \in m} n_{d,1,0}^{(O)} & \sum_{d \in m} n_{d,1,1}^{(O)} + n_{m,1}^{(N)} & \cdots & \sum_{d \in m} n_{d,1,14}^{(O)} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{d \in m} n_{d,14,0}^{(O)} & \sum_{d \in m} n_{d,14,1}^{(O)} & \cdots & \sum_{d \in m} n_{d,14,14}^{(O)} + n_{m,14}^{(N)} \end{pmatrix}$$

- Řádky matice \mathbf{N}_m vydělíme jejich součty, tím získáme matice pravděpodobností přechodu $\mathbf{P}_1, \dots, \mathbf{P}_{12}$.
- Dále napočteme průměrné počty nových zákazníků ν_1, \dots, ν_{12} .

Model migrace mezi pásmy

Algoritmus – část 2

- 1 Vyjdeme z posledního pozorovaného vektoru četností
 $n_0 = (n_{0,0}, \dots, n_{0,14})$.
- 2 Za $n_{0,0}$ dosadíme ν_m pro odpovídající m .
- 3 Pro každý další měsíc t vypočteme

$$\hat{n}_t = \hat{n}_{t-1} \mathbf{P}_{m_t},$$

kde $m_t = 1, \dots, 12$ je kalendářní měsíc odpovídající měsíci t .

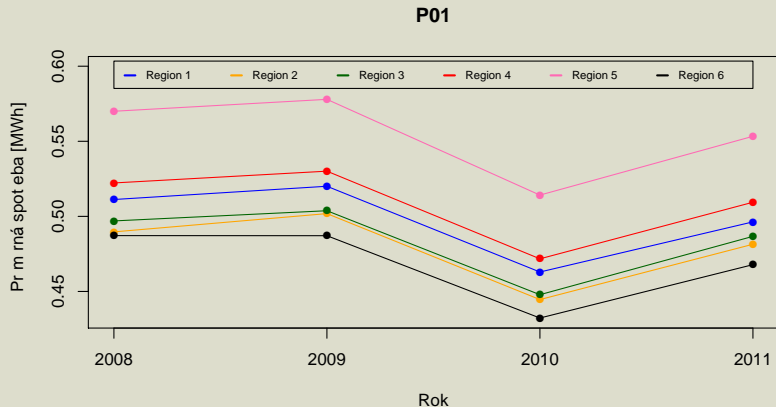
- 4 Za $\hat{n}_{t,0}$ dosadíme hodnotu $\nu_{m_{t+1}}$.
- 5 Kroky 3 až 4 opakujeme dokud potřebujeme.
- 6 Z měsíců predikovaného roku vypočteme průměrný počet pro každé pásmo.

Predikce spotřeby

- Z predikovaných počtů získáme predikce celkové spotřeby vynásobením průměrnou spotřebou v daném pásmu.
- Otázkou je, že čeho počítat průměr.
- Aktuálně používáme průměr ze všech dostupných dat (4 roky).
- Zkoušeli jsme medián, střed pásma – horší výsledky.

Predikce spotřeby

Průběh průměrných ročních spotřeb – pásmo 1 (0–1,89 MWh)



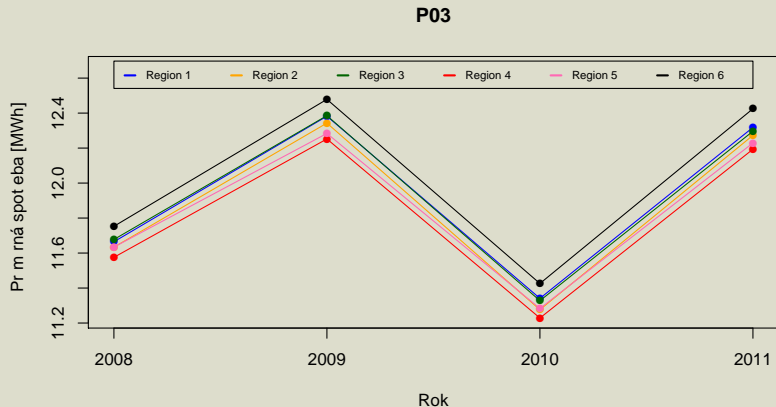
Predikce spotřeby

Průběh průměrných ročních spotřeb – pásmo 2 (1,89–7,56 MWh)



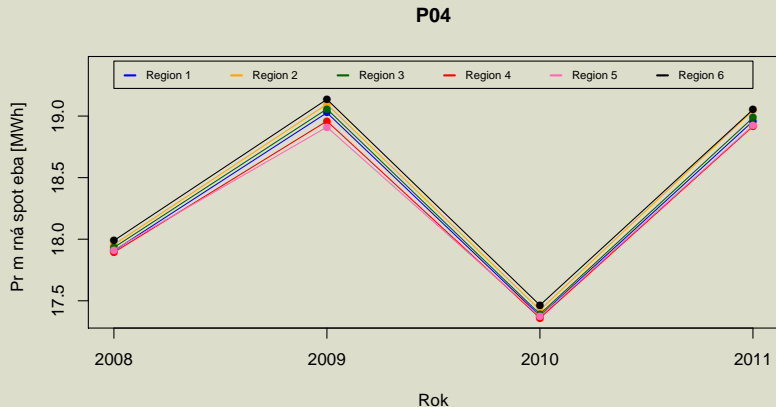
Predikce spotřeby

Průběh průměrných ročních spotřeb – pásmo 3 (7,56–15 MWh)



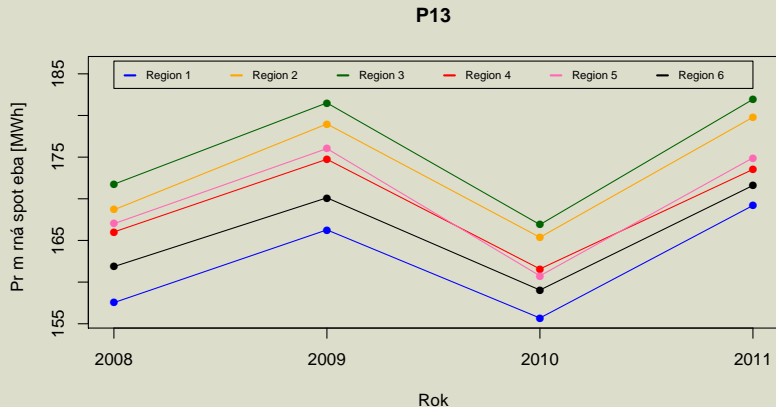
Predikce spotřeby

Průběh průměrných ročních spotřeb – pásmo 4 (15–20 MWh)



Predikce spotřeby

Průběh průměrných ročních spotřeb – pásmo 13 (63–630 MWh)



Vyhodnocení

Porovnávané modely

- Při vyhodnocení porovnáme následující modely:
 - 1 náš predikční model s „celorepublikovými“ maticemi pravděpodobností přechodu
 - 2 náš predikční model s regionálními maticemi pravděpodobností přechodu
 - 3 naivní model „zopakuj, co bylo před 3 lety“
 - 4 expertní odhad (z historických výkazů RWE)
- Pro každý model napočítáme relativní chybu predikce v procentech

$$100 \left(\frac{\hat{Y}}{Y} - 1 \right)$$

Vyhodnocení

Relativní predikční chyby [%] – počty zákazníků

Pásmo	Rep.	Reg.	Naivní	Expertní
P01	0,47	0,32	0,74	1,95
P02	-3,4	-2,88	-11,52	-0,05
P03	-1,22	-0,38	-11,78	-20,11
P04	1,12	1,41	-5,78	-2,72
P05	1,16	1,66	-1,83	3,72
P06	3,29	3,45	4,66	13,59
P07	4,08	3,92	13,96	23,29
P08	5,78	5,37	24,12	32,24
P09	8,25	7,85	33,34	39,36
P10	10,03	10,13	39,64	48,21
P11	7,92	8,4	38,53	44,18
P12	9,18	9,49	29,87	37,49
P13	2,93	1,88	6,56	16,15
Celkem	0,7	0,85	-0,62	0,84

Vyhodnocení

Spotřeby

Pásmo	Rep.	Reg.	Naivní	Expertní
P01	0,15	0,33	3,49	29,27
P02	-7,42	-6,9	-17,77	10,45
P03	-4,33	-3,52	-16,32	-8,28
P04	-2,35	-2,06	-11,03	3,06
P05	-2,32	-1,83	-7,3	7,31
P06	-0,27	-0,09	-1,16	11,92
P07	0,48	0,33	7,6	16,3
P08	2,07	1,69	17,1	18,26
P09	4,42	4,05	25,78	17,33
P10	6,2	6,32	31,79	20,06
P11	4,08	4,55	30,66	20,04
P12	5,34	5,66	22,68	15,17
P13	-0,78	-1,74	0,46	13,56
Celkem	-0,8	-0,85	0,12	9,75

Možnosti do budoucna

Co by šlo ještě vylepšit?

- Změna způsobu přiřazení ceníkového pásma – stabilizace průměrných spotřeb.
- Pomocí simulace (s využitím průběhových měření) se pokusit „natáhnout“ data a pak:
 - zkusit parametrický model pro matice pravděpodobností přechodu,
 - zkusit lepší odhad vývoje průměrných ročních spotřeb.

Děkuji za pozornost.