

<b>Projekt TDD–ČR</b>		<b>Číslo dokumentu: 2263</b>
<b>Vedoucí projektu:</b> doc. Ing. Emil Pelikán, CSc.	<b>Kontakt:</b> <a href="mailto:pelikan@cs.cas.cz">pelikan@cs.cas.cz</a>	<b>Kategorie:</b> výstupní dokument
<b>Řešitelé:</b> Ing. Tomáš Brabec Ing. Marek Brabec, PhD. Ing. Pavel Hrubeš, PhD. Mgr. Ondřej Konár RNDr. Marek Malý, CSc. Jiří Vondráček, prom. mat.		<b>Poslední změna:</b> 20. listopadu 2008

# Projekt TDD–ČR

## POPIS TDD MODELU

stav ke dni 15.11.2008



## **Abstrakt**

Tento dokument vznikl na základě dodatku č. II ke smlouvě o dílo č. K1-77200-210399 (21H938) mezi RWE Plynoprojektem, a.s. a Ústavem informatiky AV ČR, v.v.i. Obsahuje popis tvorby a použití modelu TDD pro výpočet spotřeby zemního plynu zákazníků s měřením typu „C“. Dokument zahrnuje stav ke dni 15.11.2008.

# Obsah

<b>1</b>	<b>Užití TDD z hlediska uživatele</b>	<b>3</b>
1.1	Získání přepočítané roční spotřeby . . . . .	3
1.2	Stanovení odhadu spotřeby pro daný den a hodinu . . . . .	4
1.2.1	Denní korekce . . . . .	4
1.2.2	Hodinová korekce . . . . .	6
1.2.3	TDD korekce . . . . .	7
1.2.4	Historická spotřeba . . . . .	7
1.2.5	Korekce odvozené z MAD . . . . .	7
1.2.6	Změny vyvolané změnou času a kalendáře . . . . .	7
1.2.7	Rozměr výsledné spotřeby . . . . .	8
1.3	Vzory a varianty možného použití TDD . . . . .	8
1.4	Postup při přechodu na regionální vyhodnocení . . . . .	9
<b>2</b>	<b>Aktualizace TDD</b>	<b>10</b>
2.1	Průběžné zpracování naměřených dat . . . . .	10
2.2	Tvorba matematického modelu . . . . .	10
2.2.1	Segmentování . . . . .	11
2.2.2	Model pro denní spotřeby . . . . .	11
2.2.3	Model pro hodinové spotřeby . . . . .	12
2.2.4	Celkový model . . . . .	13
<b>3</b>	<b>Výstupní soubory s křivkami TDD a dalšími parametry</b>	<b>14</b>
3.1	Tabulky pro přiřazení TDD . . . . .	14
3.1.1	Domácnosti . . . . .	14
3.1.2	Maloodběr a střední odběr . . . . .	15
3.2	Předávané TDD . . . . .	15

3.3	Předávaný logit . . . . .	15
3.4	Předávané parametry . . . . .	15
3.5	Dlouhodobě typická teplota . . . . .	16
<b>4</b>	<b>Poznámky k implementaci</b>	<b>17</b>
4.1	Výpočet hodinové spotřeby pomocí TDD . . . . .	17
4.2	Vzorové výpočty na reálných datech . . . . .	17
<b>5</b>	<b>Testování modelu TDD</b>	<b>19</b>
<b>6</b>	<b>Alternativní model</b>	<b>21</b>
<b>A</b>	<b>Slovník zkratk a pojmů</b>	<b>24</b>

# Kapitola 1

## Užití TDD z hlediska uživatele

### 1.1 Získání přepočítané roční spotřeby

Přepočítanou roční spotřebu získáme z historie odběrů za 3 roky zpět od aktuálního roku. Ze součtu spotřeb z odečtů v takto vymezeném období se vypočte odhad hodnoty ročního odběru zákazníka. Pro jeho stanovení postačuje znalost počátečních a konečných stavů měřidel na OM spolu se znalostí data odečtů v uvedeném období.

Pokryvá-li dokumentovaná historie zákazníka kratší období (např. pouze 2 roky), použije se maximální možné období. Pokud je však celková doba dokumentované historie odběrů zákazníka kratší nebo rovna 4 měsícům (120 dní), nebo jde-li o nové odběrné místo a není k dispozici údaj o měřené spotřebě (platí pro domácnosti a maloodběr), použije se sjednané množství ze smlouvy. Pokud není sjednané množství dodávky plynu ze smlouvy k dispozici, použije se sjednané množství podle instalovaných spotřebičů.

Výpočet přepočítané roční spotřeby probíhá ve dvou krocích:

1. Součet spotřeb (v přepočtených  $m^3$ ) za vymezené období se vydělí celkovým počtem dní tohoto období (tj. počtem dní od počátku období spotřeby příslušného prvnímu zařazenému odečtu do data posledního odečtu) a vynásobí se 365.

Je-li tato hodnota  $\leq 900 m^3$ , je považována za přepočítanou roční spotřebu daného zákazníka. V opačném případě se výpočet odhadu roční spotřeby provede postupem uvedeným v 2.kroku.

2. Součet spotřeb zjištěný ze zařazených odečtů (tj. všech odečtů z vymezeného období) se vydělí součtem koeficientů všech měsíců, které spadají do celkového období spotřeby odpovídajícího změřenému součtu spotřeb. Pro první a poslední měsíc celkového období

spotřeby se provede lineární přepočítání koeficientů měsíce na počet dní, které období z těchto měsíců zahrnuje.

Vzniklý podíl se vynásobí 100. Koeficienty měsíců byly stanoveny na základě dlouhodobých statistik odběrů jako procenta roční spotřeby pro jednotlivé měsíce roku (v pořadí leden až prosinec) takto:

16,72 14,29 11,02 7,94 3,84 1,83 1,62 1,62 5,86 6,83 10,50 17,93.

**Poznámka 1.1** V přestupném roce se za každý únorový den započítává  $\frac{1}{28}$  únorového koeficientu, tj.  $\frac{14,29}{28}$ . Koeficient pro únor v přestupném roce je tedy roven  $29 \cdot \frac{14,29}{28} \approx 14,80$ .

## 1.2 Stanovení odhadu spotřeby pro daný den a hodinu

Označíme-li spotřebu  $i$ -tého zákazníka s typem TDD  $k$  v hodině  $h$  dne  $t$  jako  $Y_{ikth}$ , počítá se její odhad z TDD modelu dle následujícího předpisu:

$$Y_{ikth} = D_{kt} H_{kth} C_{kth} p_{ik} MADKOR_k, \quad (1.1)$$

kde:

$D_{kt}$  je denní korekce,

$H_{kth}$  je hodinová korekce,

$C_{kth}$  je jádro odhadu odvozené z TDD,

$p_{ik}$  je denní průměr historické spotřeby daného zákazníka (individualizuje odhad na konkrétního zákazníka),

$MADKOR_k$  je korekce spočtená z předávaného parametru **mad**.

Postup výpočtu jednotlivých složek je popsán v následujících odstavcích.

### 1.2.1 Denní korekce

Denní korekce je určena vztahem

$$D_{kt} = \exp(kor\_den_{kt}) \cdot \exp(kor\_svatek_{kt}) \cdot \exp(kor\_teplota_{kt}) \cdot \exp(kor\_vanoce_{kt}) \cdot \exp(kor\_velikonoce_{kt}). \quad (1.2)$$

**Poznámka 1.2** Při implementaci vztahu (1.2) je vhodné nejprve napočítat hodnoty příslušných korekcí, tyto hodnoty sečíst a teprve nakonec použít funkci  $\exp(\cdot)$ .

Výpočet komponent denní korekce probíhá následujícím způsobem:

(a)  $kor\_den$  počítáme podle vzorce

$$\begin{aligned}
 kor\_den &= fden_1, & \text{je-li den } t & \text{neděle,} \\
 kor\_den &= fden_2, & \text{je-li den } t & \text{pondělí,} \\
 kor\_den &= fden_3, & \text{je-li den } t & \text{úterý,} \\
 kor\_den &= fden_4, & \text{je-li den } t & \text{středa,} \\
 kor\_den &= fden_5, & \text{je-li den } t & \text{čtvrtek,} \\
 kor\_den &= fden_6, & \text{je-li den } t & \text{pátek,} \\
 kor\_den &= fden_7, & \text{je-li den } t & \text{sobota,}
 \end{aligned} \tag{1.3}$$

kde  $fden_1$  je předávaný parametr  $fden1$ . Podobně s  $fden_2, \dots, fden_7$ .

(b)  $kor\_svatek$  počítáme podle vzorce

$$\begin{aligned}
 kor\_svatek &= -fsv_k, & \text{je-li den } t & \text{státem uznaným svátkem,} \\
 kor\_svatek &= fsv_k & \text{jinak,}
 \end{aligned} \tag{1.4}$$

kde  $fsv_k$  je předávaný parametr  $fsv$ .

(c)  $kor\_teplota$  počítáme podle vzorce

$$\begin{aligned}
 kor\_teplota_{kt} &= \gamma_{0k} \varphi\left(\frac{T_t - M_k}{S_k}\right) \varphi\left(\frac{T_t - N_t - m_k}{s_k}\right) - \gamma_{0k} \varphi\left(\frac{N_t - M_k}{S_k}\right) \varphi\left(\frac{-m_k}{s_k}\right) + \\
 &+ \gamma_{1k} \varphi\left(\frac{T_{t-1} - M_k}{S_k}\right) \varphi\left(\frac{T_{t-1} - N_{t-1} - m_k}{s_k}\right) - \gamma_{1k} \varphi\left(\frac{N_{t-1} - M_k}{S_k}\right) \varphi\left(\frac{-m_k}{s_k}\right),
 \end{aligned} \tag{1.5}$$

kde:

$$\varphi(x) = 1 - \frac{1}{1 + e^{-x}}, \tag{1.6}$$

$\gamma_{0k}$  je předávaný parametr  $nonlin$ ,

$\gamma_{1k}$  je předávaný parametr  $nonlin1$ ,

$m_k$  je předávaný parametr  $m$ ,

$s_k$  je předávaný parametr  $s$ ,

$M_k$  je předávaný parametr  $mm$ ,

$S_k$  je předávaný parametr  $ss$ ,

$T_t$  je průměrná denní teplota za den  $t$ ,

$N_t$  je dlouhodobě typická teplota pro den  $t$ .

**Poznámka 1.3** *Typická teplota  $N_t$  není totožná s normálovou teplotou (poskytovanou např. ČHMÚ), nýbrž vznikla vyhlazením teplotního normálu dodaného zadavatelem a řešitelem je předávána v souboru `ttyp2009.txt`.*

**Poznámka 1.4** *Funkce  $\frac{1}{1+e^{-\frac{x-m}{s}}}$ , která se opakovaně vyskytuje ve vzorcích (1.5) a (1.9) pro teplotní a hodinovou korekci, je distribuční funkce logistického rozdělení s parametry polohy  $m$  a měřítka  $s$ . Její přímý výpočet může být nestabilní pro velmi malé hodnoty argumentu  $\frac{x-m}{s}$ . Ideálně by se při výpočtu mělo použít stabilní knihovní procedury pro výpočet této funkce.*

(d) *kor\_vanoce* počítáme podle vzorce

$$\begin{aligned} \text{kor\_vanoce} &= \text{fva}_k, & \text{je-li den } t & \text{23.12., 24.12., 25.12. nebo 26.12.,} \\ \text{kor\_vanoce} &= 0 & \text{jinak,} \end{aligned} \quad (1.7)$$

kde  $\text{fva}_k$  je předávaný parametr `fva`.

(e) *kor\_velikonoce* počítáme podle vzorce

$$\begin{aligned} \text{kor\_velikonoce} &= \text{fvel}_k, & \text{je-li den } t & \text{středa, čtvrtek, pátek nebo sobota týdne} \\ & & & \text{před týdnem, v němž je Velikonoční pondělí,} \\ \text{kor\_velikonoce} &= 0 & \text{jinak,} \end{aligned} \quad (1.8)$$

kde  $\text{fvel}_k$  je předávaný parametr `fvel`.

## 1.2.2 Hodinová korekce

$$H_{kth} = \frac{\frac{\exp(L_{kth})+1}{\exp(L_{kth})+\exp(-W_{kth})}}{\sum_{j=1}^{24} \frac{1}{1+\exp(-L_{ktj}-W_{ktj})}}, \quad (1.9)$$

kde:

$W_{kth} = \text{fhod}_h$ , pokud den  $t$  je pracovním dnem (nejedná se o sobotu, neděli ani státem uznáný svátek),

$W_{kth} = -\text{fhod}_h$ , pokud den  $t$  není pracovním dnem (sobota nebo neděle nebo státem uznáný svátek či jejich libovolná kombinace).

$\text{fhod}_1$  je předávaný parametr `fhod1`, podobně s  $\text{fhod}_2, \dots, \text{fhod}_{24}$ ,

$L_{kth}$  je předávaný logit.

**Poznámka 1.5** *Ve jmenovateli vzorce (1.9) je pro jednoduchost uvedeno sčítání za 24 hodin. Ve dnech, kdy dochází ke změně času, však sčítáme přes skutečný počet hodin daného dne (tj. 23 resp. 25 hodin).*

### 1.2.3 TDD korekce

$$C_{kth} = \frac{TDD_{kth}}{100} \text{suma}_k, \quad (1.10)$$

kde:

$\text{suma}_k$  je předávaný parametr `sumpreTDD`,

$TDD_{kth}$  je hodnota předávaného TDD (vektor hodnot pro všechny hodiny roku) pro hodinu  $h$  dne  $t$  a typ TDD  $k$ .

### 1.2.4 Historická spotřeba

$p_{ik}$  ve vzorci (1.1) představuje denní průměr historické spotřeby daného zákazníka ( $i$ -tého zákazníka s typem TDD  $k$ , pro kterého se odhad počítá). Historickou spotřebu je nutné pro každého zákazníka spočítat z jeho historických dat o spotřebě (uložených v databázi zákaznického systému).

Hodnotu  $p_{ik}$  pak dostaneme jako:

$$p_{ik} = \frac{Q_{ik}}{365}, \quad (1.11)$$

kde  $Q_{ik}$  je přepočítaná roční spotřeba vypočtená metodou, popsanou v odstavci 1.1.  $Q_{ik}$  udává průměrnou roční historickou spotřebu v  $m^3$  za 365 dní, zatímco  $p_{ik}$  udává průměrnou historickou spotřebu za den.

### 1.2.5 Korekce odvozené z MAD

$$MADKOR_k = \exp\left(\frac{(\text{mad}_k)^2}{2}\right), \quad (1.12)$$

kde  $\text{mad}_k$  je předávaný parametr `mad`.

### 1.2.6 Změny vyvolané změnou času a kalendáře

Den  $t$  ve vzorci pro výpočet spotřeby je indexován datem (např. 21.5.), hodina  $h$  pořadovým číslem hodiny uvnitř dne (1, 2, ..., 24). Předávané soubory jsou upraveny na rok 2009 (zohledňují konkrétní termíny změny času v roce 2009 a fakt, že rok 2009 není přestupný).

Předpokládáme, že i v dalších letech budou předávány aktualizované soubory na konkrétní rok. Pro úplnost uvádíme nutné úpravy při použití nyní předávaných TDD v dalších letech.

Dodaná TDD pokrývají 8760 hodin nepřestupného kalendářního roku. V přestupném roce se pro 29. únor používají hodnoty TDD odpovídající 28. únoru, tj. ve dvou po sobě následujících dnech je použita stejná sada 24 hodinových údajů. Přidáním 24 hodinových údajů za 29. únor do všech relevantních souborů (tj. `tdd2009.txt`, `logit2009.txt` a `ttyp2009.txt`) vzniknou soubory pokrývající 8784 hodin přestupného roku.

Číslování hodin odpovídá platnému zimnímu, resp. letnímu času pro rok 2009. V dalších letech budou data přechodu z letního na zimní a zpět jiná, a je tedy třeba TDD transformovat dle následujícího předpisu.

K posunu času dochází vždy poslední neděli v říjnu (přechod z letního na zimní čas, kdy se hodinky ve 3 hodiny ráno posunou zpět na 2 hodiny) a poslední neděli v březnu (přechod ze zimního času na letní, kdy se hodinky ve 2 hodiny ráno posunou na 3 hodiny). V roce 2009 ke změnám dojde 29. března 2009 a 25. října 2009.

V následujících letech se pro den, kdy v aktuálním roce dochází k přechodu z letního času na zimní, třetí hodina ve všech relevantních souborech (tj. `tdd2009.txt` a `logit2009.txt`) zdvojí. Pro den, kdy dochází k přechodu ze zimního času na letní, se třetí hodina dne vypustí. Zároveň je nutno ošetřit fakt, že stávající TDD má přechod časů v jiných konkrétních dnech (29.3., 25.10.), a to zkopírováním hodiny „29.3., 2“ a vypuštěním v pořadí druhé hodiny s nominálním označením „25.10., 3“.

### 1.2.7 Rozměr výsledné spotřeby

Výsledná spotřeba vychází ve stejných jednotkách, ve kterých vstupuje přepočtená roční spotřeba  $Q_{ik}$  (tj. typicky v  $m^3$ ).

## 1.3 Vzory a varianty možného použití TDD

Hodinovou spotřebu  $Y_{ikth}$  zákazníka  $i$  s typem TDD  $t$  v hodině  $h$  dne  $t$  vypočteme podle vzorce (1.1).

Denní odhad  $Y_{ikt\bullet}$  pro den  $t$  vznikne součtem hodinových odhadů za požadovaný den, tedy:

$$Y_{ikt\bullet} = \sum_{h=1}^{24} Y_{ikth} \quad (1.13)$$

Obdobně tomu je s odhady za týden, dekádu, měsíc, kvartál, pololetí atd. (vznikají součtem odhadů příslušných hodinových spotřeb).

Spotřebu  $Y_{ikT}$  za delší časové období  $T$  (např. měsíc, rok, ...) vypočteme sečtením hodinových hodnot za dané období, tj. podle vzorce

$$Y_{ikT} = \sum_{t \in T} \sum_{h \in t} Y_{ikth} \quad (1.14)$$

Příklad výpočtu spotřeb pro vybraného zákazníka je uveden v odstavci 4.2.

## 1.4 Postup při přechodu na regionální vyhodnocení

Při přechodu na regionální vyhodnocení je nutné „zregionalizovat“ všechny parametry TDD modelu a vyřešit řadu praktických a technických detailů s tím spojených. V takto modifikovaném modelu se poté místo celorepublikových teplot budou zadávat teploty regionální. Tato úprava zatím není součástí předávaného díla.

# Kapitola 2

## Aktualizace TDD

Aktualizace typových diagramů dodávky probíhala v roce 2008 ve dvou fázích:

1. Průběžné zpracování naměřených hodnot
2. Aktualizace parametrů matematického modelu

### 2.1 Průběžné zpracování naměřených dat

I v roce 2008 probíhala analýza naměřených hodinových spotřeb a identifikace podezřelých hodnot. Odstraňování chyb v datech trvalo po celou dobu řešení projektu a trvá dosud. O aktuálním stavu měřených dat informovala řešitelská skupina ÚI v pravidelných zprávách o kvalitě dat.

### 2.2 Tvorba matematického modelu

Konstrukce TDD je založena na statistickém modelu, který kontroluje dva aspekty:

- (i) konstrukci hladké TDD křivky jako „typického průběhu spotřeby“ pro jednotlivé hodiny celého roku,
- (ii) výpočet korekcí na odchylky od tohoto typického chodu. Ty jsou potřebné jak pro náležitý odhad v části (i) v průběhu „učení“ modelu (tedy konstrukce TDD), tak pro co možná nejpřesnější použití TDD křivek v praxi.

### 2.2.1 Segmentování

Z důvodu zachování porovnatelnosti starých a nových parametrů byly i v letošním roce zachovány typy TDD, navržené v roce 2006 a definované vyhláškou 524/2006 Sb. Typy TDD `domo2 + domo3` a `moso3 + moso4` byly v roce 2008 opět sloučeny. Formálně (z důvodu kompatibility se SW implementací i z důvodu možných změn v budoucnu) však zachováváme osm typů TDD se stejnými definicemi, pouze koeficienty sloučených typů jsou stejné.

Pro každý z těchto segmentů zákazníků byl vytvořen typový diagram dodávky (TDD) a určeny příslušné parametry korekce. Domácnosti jsou pro účely zařazení k typu TDD klasifikovány podle výše ročního odběru (do  $900\text{ m}^3$  a nad  $900\text{ m}^3$ ) a podle kódu odběru (kódy S01-S03 z tabulky popisující charakter odběrného místa). Zákazníci maloodběru a středního odběru byli klasifikováni podle charakteru využití a charakteru odběrného místa.

### 2.2.2 Model pro denní spotřeby

Stejně jako v roce 2007 byl model pro výpočet TDD byl konstruován ve dvou krocích. V prvním kroku se uvažovaly pouze denní hodnoty spotřeby pro všechny měřené zákazníky, na jejichž základě byl navržen a odhadnut tzv. denní model pro spotřebu  $Y_{ikt\bullet}$   $i$ -tého zákazníka v  $k$ -tém segmentu ve dni  $t$

$$Y_{ikt\bullet} = D_{kt}C_{kt\bullet}p_{ik}MADKOR_k \quad (2.1)$$

kde:

$D_{kt}$  je denní korekce, která zahrnuje korekci na den v týdnu, na státem uznaný svátek, speciální efekt Vánoc a Velikonoc a korekci na průměrnou denní teplotu a její odchylky od dlouhodobě typické hodnoty  $N_t$  (viz příloha A),

$C_{kt\bullet}$  je hladké jádro denního odhadu spotřeby a tvoří základ TDD,

$p_{ik}$  je denní průměr historické spotřeby daného zákazníka (individualizuje odhad na konkrétního zákazníka),

$MADKOR_k$  je korekce založená na MAD (median absolute deviation), kde MAD byla spočtena z reziduí na logaritmické škále.

Prokládání celého modelu probíhá najednou pomocí penalizovaného splinu (implementovaného v podobě LME - lineárního modelu s náhodnými efekty). Používají se přitom data

dostupná pro daný typ TDD (tedy za všechny dny a zákazníky, u nichž spotřeba nechybí nebo není evidentně neplatná). V letošním roce byla z optimalizace vyřazena data z 31.12.2004 a starší. Při tomto proložení jsou odhadnuty neznámé parametry. Jádrem celého TDD v denním rozlišení je přitom křivka  $C_{kt\bullet}$ . Na rozdíl od ostatních členů, které korigují průběh spotřeby na „odlišnosti“ dané kalendářem a teplotou, tato funkce modeluje „typický“ průběh a je tak základním kamenem pro konstrukci TDD křivky.

Dle našich analýz tvoří vztah současné a historické spotřeby úzké hrdlo celého přístupu. Vylepšení odhadů historických spotřeb (např. založené na historických datech většího rozsahu) by mohlo vést k podstatnému vylepšení predikčních schopností modelu a celého přístupu. Jakkoli je absolutní výše historické spotřeby důležitá pro „ukotvení“ individuálních predikcí, tvar „dynamiky“ neboli chování časově proměnlivých členů v modelu (2.1) jí naštěstí ovlivněno příliš není. Odhad denní TDD křivky tvořené normovanou křivkou  $C_{kt\bullet}$  ani tvar korekce na teplotu a kalendářové efekty tedy nejsou příliš ovlivněny.

### 2.2.3 Model pro hodinové spotřeby

Jak již bylo zmíněno v části 2.1, byl v prvním kroku vyvinut model pro denní hodnoty spotřeb. V tomto odstavci se budeme věnovat druhému kroku postupu, tj. rozpočtu těchto denních součtů do hodinových hodnot. Denní spotřeby  $Y_{ikt\bullet}$  rozpočítáváme pomocí hodinového modelu, v němž se modelují podíly

$$\pi_{kth} = \frac{\sum_{i \in k} Y_{ikth}}{\sum_{i \in k} \sum_{j \in t} Y_{iktj}} \quad (2.2)$$

spotřeb v jednotlivých hodinách dne po logistické transformaci jako:

$$\frac{1}{1 + \exp(-\pi_{kth})} = a_k + b_{kh} + c_{kth} + d_{kth} + \varepsilon_{kth}, \quad (2.3)$$

kde:

$a_k$  je absolutní člen, závislý na typu TDD,

$b_{kh}$  je průměrný efekt hodiny  $h$  uvnitř dne pro typ TDD  $k$ ,

$c_{kth}$  je změna hodinového profilu vyvolaná odlišností dynamiky spotřeby v pracovním a nepracovním dni,

$d_{kth}$  je člen popisující hladkou změnu hodinového profilu v průběhu roku,

$\varepsilon_{kth}$  je náhodná normálně rozdělená chyba.

Prokládání hladké části i parametrických korekcí probíhá najednou – v tzv. GAM modelu (Generalized Additive Model).

## 2.2.4 Celkový model

Syntézou denního modelu (2.1) a hodinového modelu (2.3) byl vytvořen celkový model TDD pro výpočet spotřeby v jednotlivých hodinách daného roku (1.1).

# Kapitola 3

## Výstupní soubory s křivkami TDD a dalšími parametry

Z důvodů zaručení správného přečtení všech předávaných souborů jsou tyto předávány ve formátu TXT (text oddělený tabulátory). Výjimkou je soubor `vypocet2009.xls` se vzorovým výpočtem včetně příslušných vzorců.

### 3.1 Tabulky pro přiřazení TDD

Tabulky pro přiřazení TDD zákazníkovi jsou dodávány v textových souborech (přípona TXT). Hodnoty v řádcích jsou odděleny tabulátory, řádky jsou odděleny znakem „enter“ (0x0D+0x0A).

#### 3.1.1 Domácnosti

Pro domácnosti je přiřazovací tabulka dodávána v souboru `prirazeniD0.txt`. Soubor má 3 sloupce a 85 řádků. První řádek je hlavička s názvy sloupců, druhý až osmdesátý pátý řádek obsahuje charakteristiky zákazníka a přiřazovaný typ TDD.

První sloupec obsahuje kód charakteru odběru zákazníka (tzv. R-kód), druhý sloupec obsahuje identifikátor roční spotřeby (`do900`, `nad900`), třetí sloupec obsahuje typ TDD (`domo1` – `domo4`). Názvy typů TDD odpovídají názvům uvedeným v hlavičkách ostatních předávaných souborů (`tdd2009.txt`, `logit2009.txt`, `koef2009.txt`). V tabulce jsou uvedené všechny povolené kombinace R-kódu a roční spotřeby.

### 3.1.2 Maloodběr a střední odběr

Pro maloodběr a střední odběr je tabulka dodávána v souboru `prirazeniM0S0.txt`. Soubor má 2 sloupce a 541 řádků. První řádek je hlavička s názvy sloupců, druhý až pětistý čtyřicátý první řádek obsahuje charakteristiky zákazníka a přiřazovaný typ TDD.

První sloupec obsahuje kód charakteru odběru zákazníka (tzv. R-kód), druhý sloupec obsahuje typ TDD (*moso1 – moso4*). Názvy typů TDD odpovídají názvům uvedeným v hlavičkách ostatních předávaných souborů (`tdd2009.txt`, `logit2009.txt`, `koef2009.txt`). V tabulce jsou uvedené všechny povolené hodnoty R-kódu.

## 3.2 Předávané TDD

TDD jsou předávány v souboru `tdd2009.txt`. Hodnoty v řádcích jsou odděleny tabulátory, řádky jsou odděleny znakem „enter“ (0x0D+ 0x0A).

Soubor má 11 sloupců a 8761 řádků. První řádek je hlavička, obsahující názvy sloupců, další řádky obsahují hodnoty TDD pro jednotlivé hodiny v roce. První sloupec obsahuje den, druhý měsíc a třetí hodinu (1-24), čtvrtý až jedenáctý sloupec obsahují hodnoty normovaného typového diagramu dodávky pro daný typ TDD. Název příslušného typu TDD je obsažen v prvním řádku.

## 3.3 Předávaný logit

V souboru `logit2009.txt` jsou obsaženy hodnoty tzv. logitu. Hodnoty v řádcích jsou odděleny tabulátory, řádky jsou odděleny znakem „enter“ (0x0D+ 0x0A).

Soubor má 11 sloupců a 8761 řádků. První řádek je hlavička, obsahující názvy sloupců, další řádky obsahují hodnoty logitu pro jednotlivé hodiny v roce. První sloupec obsahuje den, druhý měsíc a třetí hodinu (1-24), čtvrtý až jedenáctý sloupec obsahují hodnoty logitu pro daný typ TDD. Název příslušného typu TDD je obsažen v prvním řádku.

## 3.4 Předávané parametry

Ostatní parametry TDD modelu jsou předávány v souboru `koef2009.txt`. Jsou to koeficienty určené k přepočtu TDD na skutečné podmínky. Hodnoty v řádcích jsou odděleny tabulátory, řádky jsou odděleny znakem „enter“ (0x0D+0x0A).

Soubory mají 43 sloupců a 9 řádků. První řádek je hlavička, obsahující názvy jednotlivých parametrů (zmiňované v dokumentaci k TDD modelu), druhý až devátý řádek obsahují hodnoty příslušných parametrů pro daný typ TDD. První sloupec je hlavička, obsahující názvy typů TDD.

### 3.5 Dlouhodobě typická teplota

Hodnoty dlouhodobě typické teploty jsou dodávány v souboru `ttyp2009.txt`. Hodnoty v řádcích jsou odděleny tabulátory, řádky jsou odděleny znakem „enter“ (0x0D+ 0x0A).

Soubor má 4 sloupce a 366 řádků. První řádek je hlavička. V prvním sloupci je identifikace dne, ve druhém identifikace měsíce, třetí sloupec obsahuje hodnoty dlouhodobých typických teplot (`ttyp`) a čtvrtý sloupec obsahuje hodnoty lagovaných (o jeden den zpožděných) dlouhodobých typických teplot (`ttyp1`).

**Poznámka 3.1** *Soubor `ttyp2009.txt` obsahuje stejnou typickou teplotu jako loňský soubor `ttyp2008.txt`.*

# Kapitola 4

## Poznámky k implementaci

### 4.1 Výpočet hodinové spotřeby pomocí TDD

Odhad hodinové spotřeby pomocí TDD se bude počítat podle vzorce (1.1) v kapitole 1.2. Při implementaci vztahu (1.2) je vhodné nejprve napočítat hodnoty příslušných korekcí, tyto hodnoty sečíst a teprve nakonec použít funkci  $\exp(\cdot)$ .

Při výpočtu teplotní korekce podle vzorce (1.5) je třeba dát pozor na možný problém při výpočtu funkce  $\varphi(x) = 1 - \frac{1}{1+e^{-x}}$ . Její přímý výpočet může být nestabilní pro velmi malé hodnoty argumentu  $x$ . Doporučujeme použít libovolné stabilní knihovní procedury pro výpočet distribuční funkce logistického rozdělení

$$F_{log}(x) = \frac{1}{1 + e^{-x}}, \quad (4.1)$$

z níž pak vypočteme funkci  $\varphi(x)$  podle vztahu

$$\varphi(x) = 1 - F_{log}(x). \quad (4.2)$$

### 4.2 Vzorové výpočty na reálných datech

Pro ověření správnosti implementace uvádíme následující příklad:

Uvažujme zákazníka s typem TDD *domo1*, s přepočtenou roční spotřebou  $103 \text{ m}^3$ . Pro 1. hodinu dne 1. 10. 2008 (od 0:00 do 0:59 SELČ) vypočteme pomocí TDD odhad hodinové spotřeby. Vstupní parametry jsou uvedeny v tabulce 4.1.

Vypočtená spotřeba uvažovaného zákazníka v 1. hodině 1.10.2008 je  $0.0046 \text{ m}^3$ . Hodnoty všech komponent vzorce (1.1) jsou také přiloženy v souboru `vypocet2009.xls`.

Ze souboru tdd2009.txt		
1. hodina normovaného TDD	TDD	0.00398383
Ze souboru logit2009.txt		
logit pro 1. hodinu	logit	-4.042148989
Parametry ze souboru koef2009.txt		
korekce na den - středa	fden4	-0.067929265
korekce na svátek - není svátek	fsv	-0.067327468
parametr teplotní korekce $\gamma_{0k}$	nonlin	0.27326524
parametr teplotní korekce $\gamma_{1k}$	nonlin1	0.375407662
parametr teplotní korekce $m$	m	100
parametr teplotní korekce $s$	s	1
parametr teplotní korekce $M$	mm	11.48448546
parametr teplotní korekce $S$	ss	7.21164358
parametr TDD korekce $suma_k$	sumpreTDD	460.1754
parametr korekce MAD	mad	0
parametr 1. hodiny	fhod1	0.026932206
Ze souboru ttyp2009.txt		
Normálová současná teplota	ttyp	10.89944598
Skutečná teplota dodaná externě – nebyla dodána, použita vzorová teplota		
Skutečná současná teplota		12.3
Ze souboru ttyp2009.txt		
Lagovaná (zpožděná) normálová teplota	ttyp1	11.05235457
Skutečná teplota dodaná externě		
Lagovaná (zpožděná) skutečná teplota		10.5

Tabulka 4.1: Příklad na reálných datech – zadávané parametry TDD modelu

$D$	0.868191585
$H$	1.020091584
$C$	0.018332606
$p$	0.282191781
$MADKOR$	1

Tabulka 4.2: Příklad na reálných datech – vypočtené komponenty TDD modelu

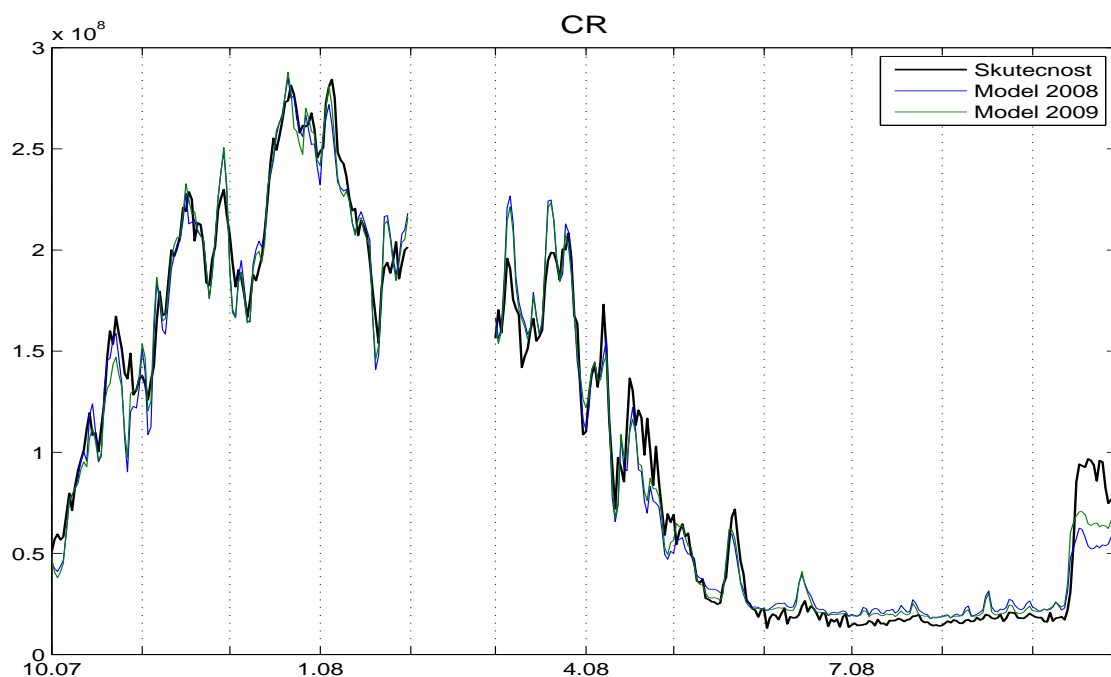
# Kapitola 5

## Testování modelu TDD

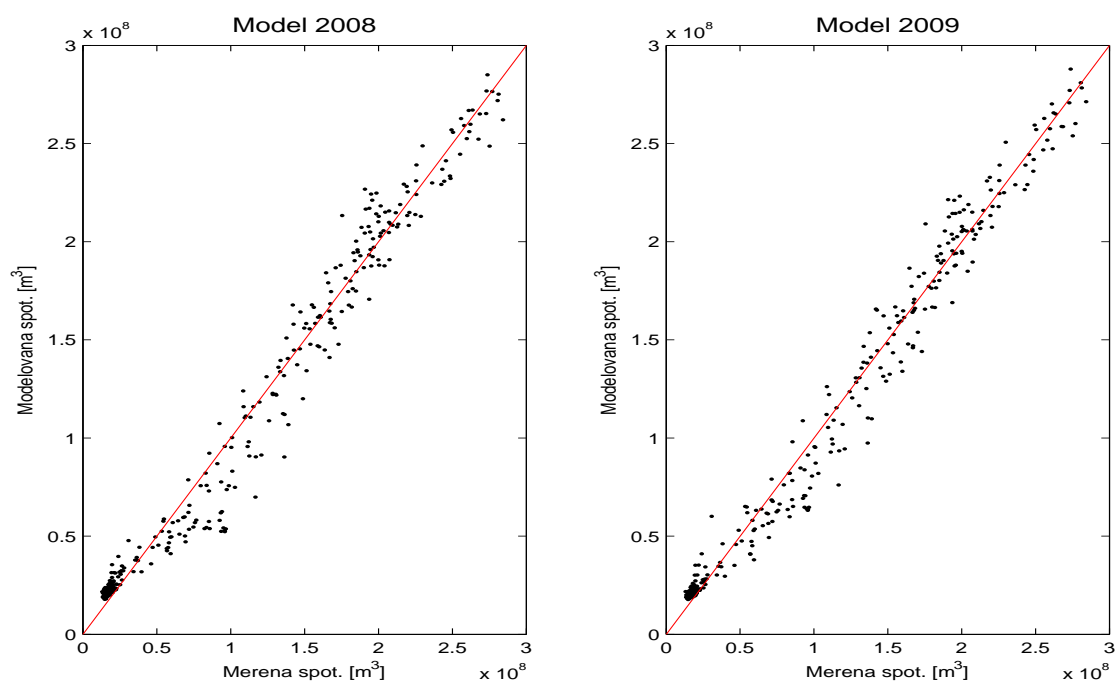
Model byl testován na datech ze systému SAO. Dostupná byla data za období 1.10.2007 – 30.9.2008, tj. za období délky 1 roku. Na obrázku 5.1 vidíme průběhy nového a starého modelu srovnané s měřenou spotřebou (data z měsíce února 2008 nebyla dodána), jiný pohled přináší obrázek 5.2, na němž vidíme modelové a měřené hodnoty bodově proti sobě. Hodnoty by se měly v ideálním případě pohybovat kolem červené čáry (rovnost modelu a měření). V tabulce 5.1 jsou uvedeny tvarové, relativní celkové a celkové nepřesnosti obou modelů vypočtené z dat ze systému SAO. U verze 2009 je pozorovatelné zlepšení jak z hlediska nepřesností, tak z hlediska průběhu spotřeb.

Nepřesnost	Model 2008	Model 2009
Tvarová [%]	9.05	7.83
Rel. celková [%]	8.94	7.81
Celková [ $m^3$ ]	9585922	8371506

Tabulka 5.1: Nepřesnosti pro verze 2008 a 2009 na datech ze systému SAO



Obrázek 5.1: Porovnání měřené a modelové spotřeby na datech ze systému SAO



Obrázek 5.2: Porovnání měřené a modelové spotřeby na datech ze systému SAO - pohled model vs. měření

# Kapitola 6

## Alternativní model

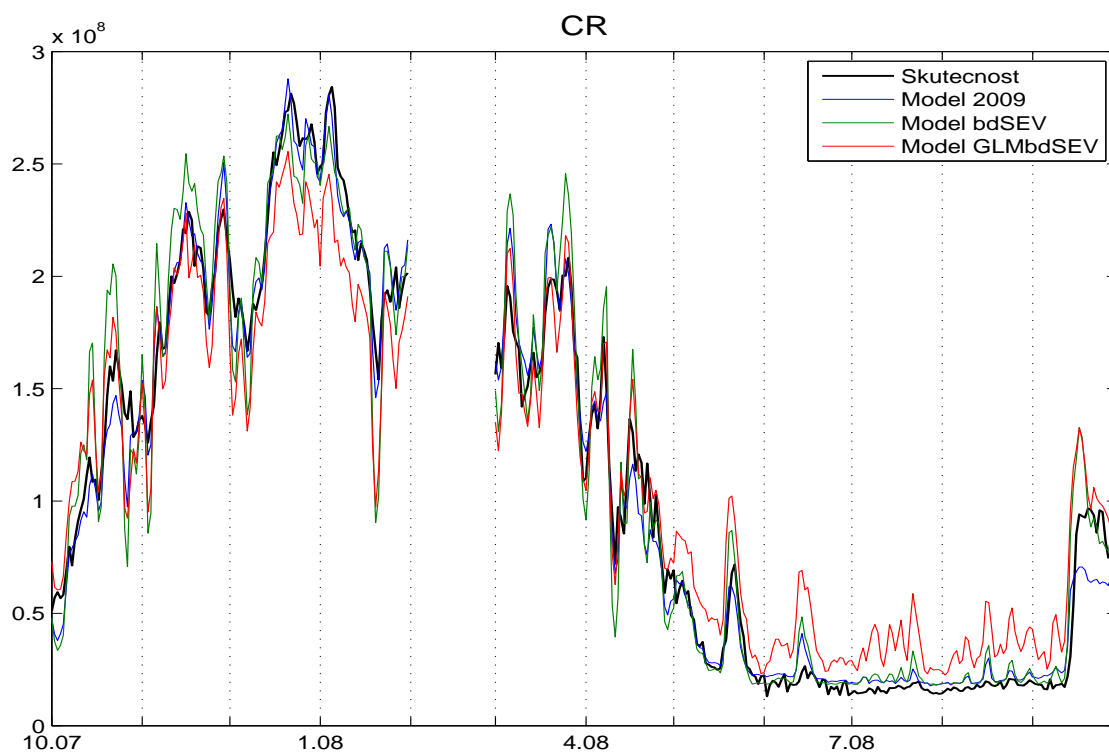
Analýzou chování TDD modelu vzhledem k vnějším teplotám, kalendářním a sezónním jevům, bylo zjištěno problematické chování modelu v „přechodových obdobích“ (začátek a konec topné sezóny). Tento jev je pozorovatelný zejména na datech ze systému SAO za září 2008 a je dle názoru řešitelské skupiny ÚI (částečně) způsoben přítomností typické části standardního modelu (teplotní normál, hladká TDD křivka). Jako možné řešení problémů v „přechodových obdobích“ řešitelé zahájili vývoj alternativního modelu s pozměněnou teplotní korekcí, který neobsahuje „normálové části“ (tj. hladkou TDD křivku a teplotní normál).

Alternativní model se od původního TDD modelu liší absencí hladké části  $C_{kth}$  a jiným (mnohem jednodušším) tvarem teplotní korekce  $kor\_teplota$ . Navrhovaný alternativní model je strukturálně jednodušší a obsahuje méně parametrů než původní model TDD. Neobsahuje hladkou část (jádro TDD - nahrazeno konstantou) ani teplotní normál. Model je schopen citlivěji reagovat na průběh skutečné teploty, což by mělo mít pozitivní vliv v přechodových obdobích. Typický průběh spotřeby lze prostřednictvím alternativního modelu získat stejně jako v původním modelu dosazením normálové teploty za teplotu skutečnou.

Parametry alternativního modelu mohou být pro testovací účely odhadovány dvěma různými postupy lišícími se ve způsobu využití historických ročních spotřeb. Tyto postupy jsou označeny jako *bdSEV* a *GLMbdSEV*. Verze *GLMbdSEV* modeluje přímo denní spotřeby bez normování na roční spotřebu, čímž se více přibližuje k datům z průběhových měření. Verze *bdSEV* modeluje podíly jednotlivých denních spotřeb z historické roční spotřeby. Implementačně se obě verze liší pouze hodnotami předávaných parametrů, nikoli strukturou modelu.

Průběh obou verzí alternativního modelu je porovnán se standardním modelem (verze

2009) na obrázku 6.1. Z obrázku 6.1 je vidět vylepšené chování alternativního modelu v září 2008. Na optimalizaci alternativního modelu se stále pracuje a pokud se zadavatel rozhodne, bude možné jej po řádném otestování implementovat pro rok 2010.



Obrázek 6.1: Porovnání měřené a modelové (standardní a alternativní modely) spotřeby na datech ze systému SAO

# Shrnutí a závěr

Z uživatelského hlediska zůstává použití modelu TDD stejné jako v minulém roce. Nová verze se liší pouze hodnotami předávaných parametrů.

Při kontrole kvality modelu na datech ze systému SAO byly pozorovány problémy v „přechodových obdobích“ (začátek a konec topné sezóny). Jako reakci na tyto problémy začala řešitelská skupina ÚI vyvíjet alternativní model, který neobsahuje normálové složky TDD modelu (teplotní normál a hladké jádro TDD). Výkonnost prvních verzí alternativního modelu je prezentována v sekci 6. Model je nadále vyvíjen a v případě zájmu jej bude možné po řádném otestování implementovat v příštím roce.

V budoucnu (ideálně již v roce 2010) by bylo vhodné k přepočtu roční spotřeby využívat TDD příslušného typu korigovaný regresory na skutečné historické podmínky. S využitím TDD k výpočtu plánované roční spotřeby bylo počítáno již od počátku projektu a vzhledem k implementaci a operativnímu použití modelu TDD v systému SAO je již podle názoru řešitelů ÚI přechod k novému způsobu možný.

# Příloha A

## Slovník zkratk a pojmů

V této příloze jsou vysvětleny zkratky a pojmy, týkající se problematiky TDD, které jsou použity v tomto dokumentu.

$a_k$  – absolutní člen v *hodinovém modelu*,

$b_{kh}$  – průměrný efekt hodiny  $h$  uvnitř dne pro typ TDD  $k$  v *hodinovém modelu*,

$C_{kth}$  – jádro *modelu TDD* odvozené z *předávaného TDD*,

$c_{kth}$  – korekce *hodinového modelu* na pracovní/nepracovní den,

$C_{kt\bullet}$  – jádro *denního modelu* (není součtem  $C_{kth}$  přes jednotlivé hodiny dne),

$D_{kt}$  – *denní korekce* v *modelu TDD*,

$d_{kth}$  – změna *hodinového profilu* během roku v *hodinovém modelu*,

DO, DOM – domácnosti,

$\exp(\cdot)$  – exponenciála o základu  $e$ ,

$F_{log}(x)$  – distribuční funkce logistického rozdělení,

GAM – Generalized Additive Model; statistický model použitý pro odhad parametrů *hodinového modelu*,

$H_{kth}$  – *hodinová korekce* v *modelu TDD*,

$kor\_den$  – komponenta  $D_{kt}$ ; korekce *modelu TDD* na pořadí dne v týdnu,

*kor\_svatek* – komponenta  $D_{kt}$ ; korekce modelu TDD na svátek,

*kor\_teplota* – komponenta  $D_{kt}$ ; korekce modelu TDD na teplotu,

*kor\_vanoce* – komponenta  $D_{kt}$ ; korekce modelu TDD na vliv Vánoc,

*kor\_velikonoce* – komponenta  $D_{kt}$ ; korekce modelu TDD na vliv Velikonoc,

$L_{kth}$  – hodnota předávaného logitu pro hodinu  $h$  dne  $t$  a typ TDD  $k$ ,

**LME** – Lineární model s náhodnými efekty; statistický model používaný pro odhad parametrů modelu TDD od roku 2006,

$m_k$  – parametr pro výpočet teplotní korekce,

$M_k$  – parametr pro výpočet teplotní korekce,

**MAD** – Median Absolute Deviation; alternativa směrodatné odchylky (Standard Deviation) založená na mediánu,

**MO** – maloodběr,

$MADKOR_k$  – korekce modelu TDD odvozené z MAD,

$N_t$  – dlouhodobě typická teplota pro den  $t$ ,

**OM** – odběrné místo; tj. zákazník, jehož spotřebu modelujeme pomocí TDD,

$p_{ik}$  – denní průměr historické spotřeby  $i$ -tého zákazníka s typem TDD  $k$ ,

$Q_{ik}$  – přepočtená roční spotřeba zákazníka  $i$  s typem TDD  $k$ ,

$s_k$  – parametr pro výpočet teplotní korekce,

$S_k$  – parametr pro výpočet teplotní korekce,

**SO** – střední odběr,

$suma_k$  – parametr pro výpočet  $C_{kth}$  z TDD, odpovídá předávanému parametru **sumpreTDD**,

$T_t$  – průměrná denní teplota v ČR pro den  $t$ ,

**TDD** – typový diagram dodávky,

**TUV** – teplá užitková voda; zákazník používá přístroj pro ohřev vody pomocí zemního plynu,

$W_{kth}$  – parametr pro výpočet hodinové korekce,

$Y_{ikth}$  – odhad hodinové spotřeby zákazníka  $i$  s typem TDD  $k$  pro hodinu  $h$  dne  $t$  počítaný modelem TDD,

$Y_{ikt\bullet}$  – odhad denní spotřeby modelované denním modelem,

$Y_{ikT}$  – odhad spotřeby zákazníka  $i$  s typem TDD  $k$  za období  $T$  (např. den, týden, měsíc atd.) počítaný modelem TDD,

$Z$  – konstanta pro přepočtení předávaného TDD na aktuální (jiný než předávaný) rok,

$\gamma_{0k}, \gamma_{1k}$  – koeficienty pro výpočet teplotní korekce,

$\varepsilon_{kth}$  – náhodná chyba v hodinovém modelu,

$\varphi(\cdot)$  – pomocná funkce pro výpočet teplotní korekce,

**časovost odběru** – dvojice indikátorů určující, kdy zákazník odebírá plyn (pracovní den, víkend),

**denní korekce** – korekce předávaného TDD na vlivy daného dne (tj. pořadí dne v týdnu, průměrná denní teplota v ČR, svátek,

$Q_{ik}$  – přepočtená roční spotřeba zákazníka  $i$  s typem TDD  $k$ , Vánoce, Velikonoce),

**denní model** – první fáze tvorby modelu TDD; statistický model pro výpočet transformované denní spotřeby; slouží pouze pro odhadování parametrů nikoli pro přímé použití v praxi,

**dlouhodobě typická teplota** – typická teplota (odvozená vyhlazením dlouhodobého normálu dodaného zadavatelem) předávaná řešitelem v souboru `ttyp2009.txt`,

**hodinová korekce** – korekce předávaného TDD na vlivy dané hodiny (tj. pořadí hodiny ve dni, pracovní/nepřacovní den), obsahuje též změnu hodinového vlivu během roku,

**hodinový model** – druhá fáze tvorby modelu TDD; rozpočet namodelovaných denních spotřeb do jednotlivých hodin dne,

**charakter odběru** – čtveřice indikátorů (tj. veličin nabývajících hodnot 0 nebo 1), určující způsob využití zemního plynu, jednotlivé hodnoty indikují využití pro vaření, ohřev TUV, otop a technologické využití,

**charakter využití OM** – jeden z 11 typů využití odběrného místa (byt, administrativní prostory, hotely a restaurace, výrobní prostory, školy, prodejny, nemocnice, zimní technologický odběr, letní technologický odběr, kotelny, ostatní drobné odběry),

**koeficienty měsíců** – váhy jednotlivých měsíců pro získávání *přepočtené roční spotřeby* zákazníků s roční spotřebou vyšší než  $900 \text{ m}^3$ ; koeficienty vychází z tzv. „otopové křivky“,

**lagovaná teplota** – zpožděná teplota; hodnota teploty (resp. dlouhodobě typické teploty) z předchozího dne,

**model TDD** – model pro výpočet hodinové spotřeby daného zákazníka, popsáný vzorcem (1.1); vznikl syntézou *denního a hodinového modelu*,

**předávané TDD** – 8760 hodnot, předávaných v souboru `tdd2009.txt`; jedná se o bezrozměrné hodnoty, které určují podíl roční spotřeby zákazníka daného typu v jednotlivých hodinách roku; tyto hodnoty jsou různé pro jednotlivé *typy TDD*,

**předávaný logit** – 8760 hodinových hodnot, předávaných v souboru `logit2009.txt`; tyto hodnoty jsou různé pro jednotlivé *typy TDD*,

**předávaný parametr** – parametr, předávaný v souboru `koef2009.txt`; parametry mohou být různé pro různé typy TDD (v souboru `koef2009.txt` je typ TDD označen v hlavičce řádků), název parametru je v souboru `koef2009.txt` uveden v hlavičce sloupců,

**přepočtená roční spotřeba** – vstup do TDD modelu; počítá se z historických spotřeb zákazníka za poslední tři roky pomocí metodiky popsané v odstavci 1.1,

**segment zákazníků** – skupina zákazníků s danými vlastnostmi (např. *charakter využití OM*, *charakter odběru*, roční spotřeba atd.); termín je používán obecně pro skupiny definované libovolnými vlastnostmi,

**typ TDD** – *segment zákazníků*, určený *charakterem využití OM*, *charakterem odběru*, *časovostí odběru* a *přepočtenou roční spotřebou*; celkem je 8 typů TDD (4 pro domácnosti, 4 pro maloodběr a střední odběr).