



národní
úložiště
šedé
literatury

Detekce zvýšených ztrát v distribuční síti zemního plynu

Konár, Ondřej
2011

Dostupný z <http://www.nusl.cz/ntk/nusl-55974>

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

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

Datum stažení: 08.05.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní nusl.cz .

Detekce zvýšených ztrát v distribuční síti 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 spojkách

Abstrakt

Článek pojednává o problematice detekce zvýšených ztrát v distribuční soustavě. Je prezentována problematika přepravy, měření a odhadu spotřeby zemního plynu ve světě, s důrazem na situaci v ČR. Dále je prezentován projekt, v rámci něhož je zkoumána problematika detekce ztrát na vybraných uzavřených lokalitách v rámci distribuční sítě RWE GasNet, s.r.o.

Úvod

Pro obchodníky se zemním plynem představují ztráty v distribuční soustavě ztráty finanční. Proto je pochopitelné, že je snaha ztráty minimalizovat. Bohužel, jak bude podrobněji popsáno v dalších odstavcích, jsou v případě zemního plynu ztráty prakticky neměřitelné. Přesné (nebo přinejmenším uspokojivě přesné) změření výše ztrát by znamenalo osazení všech vstupů a výstupů v síti podrobným průběhovým měřením, například v denním rozlišení. Toto řešení je však v současné době velice nákladné (obnáší jednorázové náklady na pořízení měřidel, ale také dlouhodobé náklady na průběžné zpracování obrovského množství naměřených údajů). Z tohoto důvodu je obvyklé využití matematických modelů spotřeby plynu. V dalším textu je naznačena situace v ČR, popsán způsob přepravy a distribuce zemního plynu, měření jeho spotřeby a odhad spotřeby pomocí matematických modelů. Dále je popsán projekt, jehož cílem je detekovat zvýšené ztráty ve vybraných uzavřených lokalitách v rámci distribuční sítě společnosti RWE GasNet, s.r.o. V rámci tohoto projektu jsou vyvíjeny metody využití níže popsaných matematických modelů k identifikaci oblastí se zvýšenými ztrátami.

1. Přeprava a distribuce zemního plynu v ČR

1.1. Zdroje zemního plynu

V České republice existují čtyři typy „zdrojů“ zemního plynu. Jedná se o dovoz, vlastní těžbu, zásobníky a akumulaci. Hlavním zdrojem plynu je dovoz, ostatní zdroje jsou spíše podpůrné (vlastní těžba pokrývá dle Plynárenské příručky [25] řádově jednotky procent spotřeby v ČR). Průtok importovaného plynu je v průběhu roku relativně stálý, naproti tomu nerovnoměrnost spotřeby v rámci roku je obrovská. To je dáno skutečností, že nezanedbatelná část zemního plynu je spotřebována pro účely vytápění. Spotřeba v zimě pak činí až desetinásobek letní spotřeby.

Tyto výkyvy ve spotřebě je potřeba vyrovnávat. V zásadě existují dvě možnosti:

1. nerovnoměrná těžba zdrojů,
2. využití zásobníků.

V ČR pochopitelně připadá v úvahu pouze možnost 2 vzhledem ke zmíněné zanedbatelnosti vlastních těžebních zdrojů. Zásadní roli při vyrovnávání nerovnoměrnosti spotřeby v průběhu roku tedy hrají zásobníky. Plynárenská příručka [25] rozlišuje dva typy zásobníků podle způsobu využití:

sezónní zdroje – ty se vyznačují velkou uskladňovací kapacitou, těžba, resp. vtlačení, je však u těchto zásobníků relativně pomalá; jedná se především o tzv. aquiferové zásobníky (přírodní zásobníky podzemní vody jsou využity pro skladování plynu) a vytěžená ropná či plynová ložiska,

špičkové zdroje – slouží především ke krytí špičkové spotřeby (nejchladnější dny), ale také k vyrovnávání krátkodobých výkyvů; vyznačují se

menší skladovací kapacitou, ale rychlejší těžbou; jedná se např. o solné či umělé kaverny, plynovody, zásobníky zkapalněného plynu (LNG – v ČR nevyužíváno).

Zvláštním typem zásobníku je akumulace plynu v potrubí. Akumulace soustavy je definována jako celkový objem plynu v přepravní soustavě. Vzhledem k plynnému skupenství může tento objem (a spolu s ním i tlak) poměrně výrazně kolísat. Toho lze do jisté míry využít k regulaci nerovnoměrnosti spotřeby v rámci dne.

1.2. Přeprava zemního plynu

Ve světě existují dva typy přepravy zemního plynu:

1. potrubní přeprava,
2. přeprava zkapalněného plynu (LNG) tankery.

V ČR je pochopitelně využíván pouze první způsob přepravy. Přeprava pomocí tankerů je však i ve světě až na výjimky kombinována s potrubní přepravou.

Přepravní síť zemního plynu v České republice se skládá z několika kategorií plynovodů:

nízkotlaké plynovody – pracovní tlak do 5 kPa včetně, slouží pro domovní rozvody případně distribuci v menších obcích, tlak plynu není třeba před použitím ve spotřebičích dále upravovat,

středotlaké plynovody – pracovní tlak do 400 kPa včetně, využívají se při potřebě vyšší kapacity a pružnosti sítě (při připojení na středotlaký rozvod si musí odběratel opatřit regulátor k úpravě tlaku plynu na hodnotu potřebnou pro provoz spotřebičů),

vysokotlaké plynovody – pracovní tlak do 4 MPa včetně, slouží především pro vnitrostátní dálkovou dopravu plynu do jednotlivých plynofikovaných obcí,

velmi vysokotlaké plynovody – pracovní tlak nad 4 MPa, slouží především pro mezinárodní dálkovou dopravu plynu,

Dovážený plyn z Ruska vstupuje na území ČR v předávací stanici Lanžhot. Ze systému dálkové přepravy se dostává zemní plyn (ta část, která není exportována do dalších zemí) přes předávací stanice vnitrostátní soustavy. V těchto předávacích stanicích se také upravuje tlak plynu na hodnotu obvyklou v dané vnitrostátní síti. Vnitrostátní síť je zemní plyn dopravován

do měst a obcí, případně přímým odběratelům (typicky velkým průmyslovým podnikům).

1.3. Měření spotřeby zemního plynu

Údaje o množství přepraveného plynu jsou pochopitelně důležité jak pro provozovatele distribuční sítě (poplatky za přepravu), tak pro obchodníky s plynem (představa o množství spotřebovaného plynu). Předávací stanice jsou osazeny průběhovým měřením, které poskytuje hodinové hodnoty množství plynu, který těmito body protekl. Spotřeba plynu jednotlivých odběratelů je taktéž měřena, aby mohla být na základě měření provedena fakturace odebraného plynu a poplatků za přepravu. Za předpokladu konstantní akumulace (jejíž změny lze odhadnout pomocí měření tlaku v předávacích bodech) odpovídá rozdíl mezi vstupním a spotřebovaným množstvím plynu ztrátám v daném úseku distribuční soustavy.

Tyto ztráty však nelze v žádném okamžiku přesně vypočítat, a to ani když zanedbáme vliv akumulace. Problém je na straně měření spotřeby. Jednotlivá odběrná místa jsou (zpravidla dle průměrné roční spotřeby) osazena jedním ze tří typů měřidel [30]:

měřidlo typu A – průběhové měřidlo s dálkovým přenosem, naměřené hodinové hodnoty jsou průběžně odesílány na dispečink,

měřidlo typu B – průběhové měřidlo bez dálkového přenosu, naměřené hodinové hodnoty jsou zpravidla jednou měsíčně ručně vyčítány pomocí přenosného zařízení,

měřidlo typu C – bez průběhového měření, je zaznamenáno pouze celkové množství odebraného plynu, to je v určitých intervalech (jeden až osmkrát měsíčně) odečítáno, spotřeba se určí rozdílem dvou po sobě jdoucích odečtených hodnot.

Denní hodnoty ztrát by bylo možno přesně určit pouze v případě, že by všechna odběrná místa byla osazena měřením typu A, případně měřením typu B (pak by bylo možno ztráty určovat zpětně vždy na konci kalendářního měsíce). Osazení tímto typem měření je však velmi nákladné (pořizovací cena měřidla se pohybuje v desítkách tisíc korun). Proto je většina odběrných míst, především u domácností a maloodběratelů (tj. zákazníků s ročním odběrem do 630 MWh), ale i některých středních odběratelů (zákazníků s ročním odběrem mezi 630 a 4200 MWh), osazena měřením typu C.

K získání údajů o spotřebě z měřidla typu C je zapotřebí provést fyzický odečet. Ten spočívá v odečtení

stavu plynoměru pověřeným pracovníkem a zápisu do databáze. Pořízení fyzických odečtů všech zákazníků v jednom dni je velmi technicky i ekonomicky náročné. Z tohoto důvodu většina dodavatelů plynu v České (ale i Slovenské) republice přistoupila k tzv. cyklickým odečtům. V každém měsíci daného roku je odečtena určitá část zákazníků (rozložení není zcela rovnoměrné s ohledem na obtížnější přístup k plynůměrům v letních měsících z důvodů častých dovolených). Při plánování odečtových tras přitom musí být zajištěno, aby interval mezi dvěma odečty daného zákazníka nebyl delší než 18 měsíců. Typicky je délka intervalu mezi jednotlivými odečty roční, u větších zákazníků však není výjimkou ani měsíční odečítání.

Důsledkem cyklického odečítání je skutečnost, že v žádném okamžiku není přesně znám celkový objem spotřebovaného plynu v daném úseku distribuční soustavy. Z toho plyne, že nelze přesně určit ani ztráty. Pomineme-li v současné době ekonomicky neprůchodnou variantu osazení všech zákazníků měřením typu A, je pro vybrané úlohy (určení ztrát, množství nevyfakturovaného plynu, obchodní bilance atd.) nutné použít odhad. V následující kapitole je popsáno několik matematických modelů používaných k odhadu spotřeby zákazníků s měřením typu C v různých situacích.

2. Modelování spotřeby zemního plynu

Vzhledem ke komplikacím při distribuci zemního plynu popsaným v kapitole 1 (rovnoměrný dovoz, nerovnoměrný odběr, obtížné skladování, nákladné měření) je modelování spotřeby zemního plynu velmi důležitým nástrojem distribučních společností. Modely popsané v literatuře mají několik společných znaků. Typicky je součástí modelu klasifikace zákazníků na základě jejich odběratelského chování. Nejjednodušší formou je klasifikace podle výše odběru (např. maloodběratel – střední odběratel – velkoodběratel, případně jemnější ceníková pásma), dále se ale také používají klasifikátory jako je způsob užití zemního plynu (vaření, ohřev vody, vytápění, technologický odběr) případně typ podniku (výrobní prostory, služby, zemědělství atd.) a další. Volba vhodné klasifikace je náročná a závisí mimo jiné také na technických možnostech konkrétní distribuční společnosti a správnosti údajů v její databázi. Výsledné třídy by měly být pokud možno homogenní v ročním průběhu spotřeby, zároveň by však měly být od sebe zřetelně odlišitelné. Nezanedbatelným požadavkem je také dostatečné zastoupení zákazníků v každé třídě.

Modely popsané v literatuře využívají nejruznější vysvětlující proměnné. Typicky bývají zahrnuty meteo-

rologické veličiny. Již v práci [1] byla popsána závislost spotřeby na teplotě. Uvažovány jsou však i další meteorologické veličiny, jako je rychlost a směr větru, sluneční svit, atmosférický tlak či srážky. Teplotní odezva bývá často velmi komplexní, minimálně se uvažuje rozdílná teplotní závislost pro teplé a studené dny. Existuje koncepce tzv. „heating degree days“, která spočívá v zanedbání vlivu teploty na spotřebu nad určitou stanovenou mezí [12, 27]. Podobná koncepce je použita například i u modelu GAMMA používaného v ČR [13, 28]. Dalšími běžně používanými prediktory jsou kalendářní efekty jako je den v týdnu, resp. rozlišení na pracovní a nepracovní den, svátky, vánoce, velikonoce a další. Používány jsou také jiné veličiny, jako je např. cena plynu, resp. ropy, s níž je cena plynu vázaná [24].

Dále je možné dělit modely podle způsobu jejich využití:

1. Odhad individuální spotřeby:

- rozpočet známé spotřeby do kratších časových úseků (například při změně ceny plynu, ke které došlo mezi dvěma řádnými odečty měřidla typu C),
- odhad spotřeby za určitý časový úsek v minulosti (například v případě, kdy je ze zákona nutno vystavit zákazníkovi fakturu, ale nejsou k dispozici údaje z měření, pak je třeba odhadnout spotřebu od posledního odečtu do okamžiku fakturace),
- předpověď individuální spotřeby v budoucnu.

2. Odhad spotřeby větších skupin zákazníků:

- odhad nevyfakturovaného plynu (celkové množství plynu, které bylo spotřebováno, ale nebylo fakturováno),
- bilance v soustavě (odhad ztrát či jejich rozpočet mezi jednotlivé účastníky trhu),
- predikce celkové spotřeby celého zákaznického kmene.

Rozpočet (případ 1a) je úloha s největší dostupností potřebných údajů. Z tohoto důvodu je svým způsobem jednodušší než zbývající úlohy. Problémy 1b a 1c se liší především v dostupnosti hodnot vysvětlujících proměnných, např. průměrné teploty vzduchu. Ve světové literatuře je nejvíce publikací věnováno predikci celkové spotřeby (2c) [9, 10, 12, 17, 21–24], méně pak tvorbě tzv. typových diagramů dodávky (TDD), tj. úloze 2b [11].

Různé úlohy pochopitelně mohou někdy být řešeny pomocí stejného modelu, obecně však nelze zkonstruovat model, který optimálně řeší všechny. Použití jednotného

přístupu k řešení různých problémů je vždy kompromisem. Nicméně vzhledem k přirozené tendenci minimalizovat náklady je tento požadavek v praxi běžný.

K řešení úlohy 1a se dříve v ČR používala tzv. otovpová křivka [29]. Jednalo se o sadu dvanácti měsíčních koeficientů, která udávala empiricky zjištěný podíl daného měsíce na spotřebě plynu při vytápění. Pro zákazníky s malou spotřebou, u nichž se vytápění plynem nepředpokládalo, se rozpočet prováděl rovnoměrně. V dnešní době se používá k rozpočtu používá model TDD [30]. Ten se nyní užívá i k náhradě odečtu (úloha 1b). Úloha predikce individuálních spotřeb (1c) byla řešena pro zákazníky s měřením typu B např. v publikaci [3].

Pro odhad nevyfakturované složky plynu (2a) byl konstruován nelineární regresní model GAMMA [8, 13, 28]. Pro účely zúčtování odchylek (2b) slouží poněkud složitější model TDD třídy GAM (generalized additive models) [4, 5, 30]. Řešení problematiky odhadu ztrát se v ČR teprve připravuje. Predikcí celkové spotřeby (2c) se zabýval např. systém ELVÍRA [18, 19], který byl založen na metodách analýzy časových řad.

V následujících odstavcích jsou podrobněji popsány modely GAMMA a TDD pro odhad spotřeby zemního plynu vyvíjené v ÚI AV ČR ve spolupráci s účastníky trhu s plynem v ČR.

2.1. Model GAMMA

Model GAMMA [13, 28] byl primárně určen pro odhad nevyfakturovaného plynu [8]. O možném využití k dalším praktickým úlohám pojednává článek [2]. Je založen na odhadech individuálních spotřeb, které jsou posléze agregovány po stanovených třídách zákazníků. Celkové množství nevyfakturovaného plynu se určí jako součet individuálních odhadů zákazníků v dané zóně. Model byl v letech 2004 až 2009 rutinně používán v Západočeské plynárenské, a.s. (dále jen ZČP). Od roku 2010 přebrala model firma RWE GasNet, s.r.o. (dále jen RWE), která uvažuje o jeho použití pro odhad ztrát v tzv. uzavřených lokalitách. Tato problematika je podrobněji popsána v kapitole 3.

Odběrná místa jsou klasifikována dle typu klienta (domácnost, malooběr) a dle charakteru odběru (vaření, ohřev vody, vytápění, technologický odběr). Uvažovány jsou všechny kombinace prvních tří charakterů (celkem 7 tříd pro domácnosti a 7 tříd pro malooběr) a dvě třídy pro technologický odběr (čistě technologický a v kombinaci s vytápěním). Celkem máme tedy k dispozici 16 zákaznických tříd. Vybrané parametry modelu jsou společně všem zákazníkům dané třídy.

Základním časovým rozlišením modelu GAMMA je den, typické použití je však pro odhad spotřeby za delší časové období (1 až 18 měsíců). Jako vysvětlující proměnné se používají průměrné denní teploty vzduchu v daném regionu a také dlouhodobý teplotní normál. Model je relativně jednoduchý s ohledem na to, že pro odhad parametrů nejsou a nikdy nebyla k dispozici denní data. Pro odhad parametrů se využívá kombinace mimořádných měsíčních odečtů náhodně vybraných cca 1700 zákazníků a dále údajů z řádných (zpravidla ročních) odečtů celého zákaznického kmene (očištěného od podezřelých hodnot).

Pro spotřebu Y_{ikd} zákazníka i třídy k ve dni d je definován následující nelineární regresní model:

$$Y_{ikd} = \mu_{ik}\Phi_{kd} + \varepsilon_{ikd}, \quad (1)$$

kde

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

Φ_{kd} je systematická část modelu, společná pro třídu k ,

ε_{ikd} je náhodná složka, u níž předpokládáme nulovou střední hodnotu a rozptyl úměrný střední hodnotě denní spotřeby, tj. členu $\mu_{ik}\Phi_{kd}$.

Individuální parametr μ_{ik} je odhadován metodou vážených nejmenších čtverců s využitím řádných odečtů daného zákazníka až tři roky do historie. Délka využívané historie je také jedním z parametrů modelu specifických pro zákaznickou třídu. V minulosti bylo také experimentováno s odhadem parametru μ_{ik} pomocí metod NLME (nelineární modely s smíšenými efekty) [14, 15]. Tento přístup však nakonec nebyl v provozu využit z důvodu výrazně vyšší výpočetní náročnosti.

Systematická část Φ_{kd} má následující tvar:

$$\Phi_{kd} = Z_{kd}(\Psi_d \cdot \exp\{-\gamma_k f(T_d, N_d)\} + p_k) + (1 - Z_{kd})q_k, \quad (2)$$

kde

Z_{kd} je indikátorová proměnná, je rovna 1, jestliže je průměrná teplota ve třech posledních dnech ($d, d-1, d-2$) nižší než 14°C (indikuje zimní období) nebo je třída k tzv. „neotopová“ (zákazníci nevyužívají plyn k vytápění), v opačném případě má Z_{kd} hodnotu 0,

Ψ_d je sezónní složka s roční periodicitou společná všem třídám,

γ_k je parametr udávající míru teplotní závislosti,

$f(T_d, N_d)$ je funkce průměrné denní teploty T_d a normálové teploty N_d tvaru

$$f(T_d, N_d) = T_d - N_d, \quad (3)$$

p_k je stálá (nesezónní) složka spotřeby v „zimním“ období,

q_k je stálá složka spotřeby v „letním“ období (pouze pro „otopové“ třídy).

Indikátorová proměnná Z_{kd} slouží k „přepínání“ zimního a letního provozu u zákazníků využívajících plyn k vytápění. U těchto zákazníků se ukázalo, že použití jednotného tvaru po celý rok způsobuje neuspokojivou přesnost odhadu spotřeby v letních měsících. Použití třídních průměrů teplot místo pevného časového určení částečně řeší problém tzv. přechodových období, tj. období začátku a konce topné sezóny, které nastávají v každém roce jindy. Tato období jsou zároveň velice citlivá, neboť je zde zvýšená variabilita ve spotřebě mezi zákazníky daná rozdílným zákaznickým chováním.

Sezónní složka Ψ_d byla jednorázově odhadnuta neparametricky z měsíčních hodnot vstupu do distribuční sítě ZČP. Měsíční hodnoty byly posléze interpolovány pomocí polynomu. Tak vznikly denní hodnoty Ψ_d . Tento odhad je následně používán jako vysvětlující proměnná v nelineárním regresním modelu daném rovnicemi (1) a (2), tj. je pro daný den v roce považován za pevnou konstantu. Důvodem tohoto postupu byla skutečnost, že v době návrhu modelu nebyla k dispozici jiná data s dostatečně jemným časovým rozlišením.

Tvar teplotní funkce $f(T_d, N_d)$ byl zvolen pro jednoduchost a snadnou interpretovatelnost. Jak skutečně, tak normálové teploty jsou však před výpočtem shora ořezány v hodnotě 14°C. Důvodem je experimentální zjištění, že přibližně v této hodnotě zaniká teplotní závislost spotřeby.

Ostatní parametry modelu, tj. γ_k , p_k a q_k jsou odhadovány metodou nejmenších absolutních odchylek s využitím údajů z mimořádných měsíčních měření. V případě nedostatečného množství mimořádných měření se použijí řádné odečty. Tato metoda byla zvolena, neboť poskytovala stabilnější odhady než tradičnější metoda nejmenších čtverců.

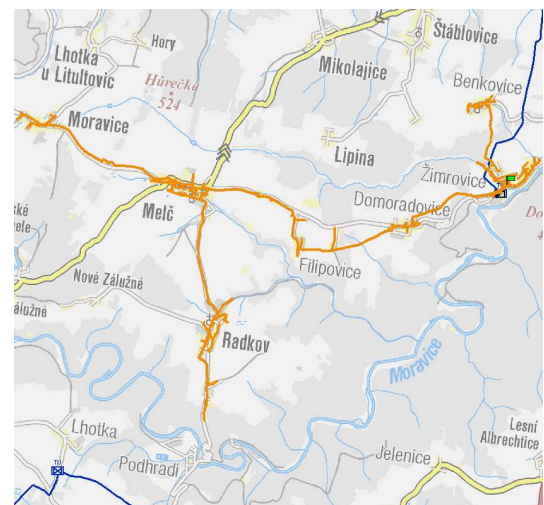
Modularita modelu umožňuje provádět dílčí změny bez výraznějšího narušení struktury, a tudíž i procesu odhadu parametrů. Chceme-li například změnit tvar teplotní odpovědi modelu, lze prostě zaměnit funkci f jinou vhodnou funkcí. Při vývoji modelu byl kladen důraz

na maximální přesnost odhadu celkového nevyfakturovaného plynu, tj. odhadu spotřeby relativně velkého celku za relativně dlouhé období. Pro použití k detekci zvýšených ztrát může být nutné provést určité modifikace.

3. Detekce zvýšených ztrát

V současné době probíhá ve spolupráci s distribuční společností RWE GasNet, s.r.o., vývoj metodiky detekce anomálního časového průběhu ztrát v uzavřené lokalitě s využitím modelu GAMMA popsaného v odstavci 2. Příkladem atypického průběhu může být krátkodobé zvýšení ztrát (např. při havárii) případně dlouhodobě vyšší ztráty (např. při ilegálním odběru). Součástí projektu je získání údajů z několika desítek uzavřených oblastí v rámci celé její distribuční sítě (prakticky celá ČR s výjimkou Prahy a Jihočeského kraje).

Uzavřenou lokalitou se rozumí část distribuční soustavy, která má jeden nebo více měřených vstupů a výstup pouze u koncových odběratelů. Typicky se jedná o několik menších obcí (řádově 500-1000 odběratelů). Příklad takové lokality je v podobě mapy uveden na obrázku 1.



Obrázek 1: Příklad uzavřené lokality.

Veškerý plyn, který vteče do oblasti by měl být spotřebován. Rozdíl mezi spotřebou plynu a přítokem do soustavy představuje ztráty v soustavě. Problém je v tom, že spotřebu (a tudíž ztráty) na rozdíl od vstupu do soustavy nelze v praxi změřit v dostatečném časovém rozlišením, jak bylo popsáno v kapitole 1.

Odhad denních ztrát získáme pomocí odhadu spotřeb (např. modelem GAMMA) tak, že od měřeného

vstupního objemu v daném dni odečteme měřenou spotřebu velkých zákazníků a odhadnutou spotřebu malých zákazníků. Výsledkem je pro každou oblast řada denních odhadů ztrát.

3.1. Data

Pro řešení úlohy jsou k dispozici následující datové soubory:

1. Data z vybraných uzavřených oblastí:
 - a) denní měřené objemy vstupujícího plynu,
 - b) denní měřené spotřeby všech velkoobdobatelů a středních odběratelů,
 - c) fakturační odečty všech odběratelů v dané oblasti (typicky roční interval).
2. Průměrné denní teploty v jednotlivých krajích (zdroj ČHMÚ) od roku 1999.
3. Podpůrná data:
 - a) mimořádné měsíční odečty cca 1700 zákazníků ze západních Čech,
 - b) fakturační odečty zákaznického kmene RWE.

Objem zpracovávaných dat je poměrně rozsáhlý. Kromě toho jsou data v poměrně syrovém stavu. Zkušenosti s dosavadní implementací modelu GAMMA v západních Čechách ukazují, že model může sloužit také jako nástroj k automatizovanému hledání chyb v datech. To však pochopitelně nevede k odstranění všech chyb a je třeba kombinovat s „ručním“ hledáním.

Příkladem využití modelu GAMMA pro hledání chyb v datech je čištění dat z mimořádných měsíčních odečtů. Označíme-li Y_{im} měsíční spotřebu zákazníka i v měsíci m , \hat{Y}_{im} její odhad modelem GAMMA a M celkový počet měřených měsíců, zavedeme pro každého zákazníka a každý měsíc penalizaci

$$P_{im} = \begin{cases} 1 & \Leftrightarrow \left(\frac{|Y_{im} - \hat{Y}_{im}|}{Y_{im}} > 2 \right) \& \\ & \left(|Y_{im} - \hat{Y}_{im}| > 0,75 \frac{1}{M} \sum_{t=1}^M Y_{it} \right) \\ 0 & \text{jinak.} \end{cases} \quad (4)$$

Jako podezřelého pak označíme zákazníka, pro kterého platí $\sum_{m=1}^M P_{im} > \frac{M}{2}$. Validita měření u podezřelých zákazníků je následně ověřována poskytovatelem dat.

3.2. Řešení

Hlavním úkolem je porovnávat odhady ztrátových procesů v jednotlivých měřených oblastech. Cílem je

vytipovat oblasti se zvýšeným objemem ztrát a ty pak osadit průběhovými měřeními, a to ve všech bodech. Primárně bude vyměněno měřidlo na vstupu z důvodu vyloučení jeho případné systematické chyby. Pokud bude po určeném čase stav přetrvávat, budou průběhovými měřidly osazena všechna odběrná místa, aby bylo možno ztráty změřit. Poté bude následovat vyhodnocení reálných (naměřených) ztrát oproti jejich odhadům pomocí modelu GAMMA.

Úloha detekce ztrát je komplikována několika problémy, které vyžadují nestandardní postupy:

1. Chyba odhadu není a nikdy nebude pozorována. V daném dni je totiž měřen pouze vstup do soustavy, nikoli výstup. Spotřeba zákazníků v daném dni je odhadována modelem GAMMA. Rozdíl mezi měřeným vstupem a odhadnutou spotřebou představuje odhad ztrát.
2. Do hry vstupuje velmi mnoho proměnných, které mohou více či méně ovlivnit průběh odhadu ztrát:
 - a) různá skladba zákazníků v jednotlivých lokalitách (počet i spotřeba),
 - b) různé odběratelské zvyklosti v různých regionech (např. různá pracovní doba, různé nároky na teplotu v bytě, v domě apod.),
 - c) různé povětrnostní podmínky (chladnější a teplejší regiony),
 - d) vlastní chyba použitého matematického modelu a její variabilita (v čase i mezi regiony).

Tyto faktory je třeba brát v úvahu a snažit se odfiltrovat jejich vliv před vlastní diagnostikou.

3. Pojem „zvýšené ztráty“ není zcela jasně definován. Před hledáním detekčních metod bude třeba dobře rozmyslet, co vlastně chceme detekovat. V zásadě jde o problém detekce odlehých pozorování, ovšem s tím, že máme dva druhy odlehlostí:
 - a) odlehlost uvnitř lokality (např. při krátkodobém černém odběru, úniku při havárii apod.),
 - b) odlehlost mezi lokalitami (např. při dlouhodobém černém odběru, ale také např. při odlišné skladbě zákazníků).

Není vyloučeno, že bude třeba použít rozdílných metod pro detekci různých typů odlehlostí.

4. První výsledky

V současné době jsou k dispozici pro testování údaje z 9 uzavřených lokalit (v budoucnu by se měl tento počet rozšířit na řádově desítky lokalit). Z důvodu důvěrnosti používaných dat budou v dalším textu lokality rozlišeny číselnými kódy. V těchto lokalitách byla odhadnuta spotřeba všech zákazníků modelem GAMMA a porovnána s celkovým vstupem do soustavy (po odečtení průběhově měřených zákazníků).

4.1. Kritérium

Proces detekce musí být při případném provozním použití plně automatický (nelze spoléhat na ruční procházení stovek až tisíců průběhů spotřeb). Z toho důvodu je klíčová volba kritéria pro hodnocení ztrát v dané lokalitě. V zásadě je třeba vytvořit uspořádání lokalit podle zvoleného kritéria a posléze věnovat zvýšenou pozornost „nejhorším“ lokalitám.

Přirozenou cestou je využívat odchylky modelovaných spotřeb od měřeného vstupu (které představují odhad ztrát v dané lokalitě). Vzhledem k tomu, že prioritní je odhalit ztráty dlouhodobé, je vhodné v první fázi volit kritéria zohledňující chování na delším časovém úseku.

Dále je třeba volit mezi kritérii absolutními a relativními. Výhodou absolutních kritérií, jako např.

$$Err_1(l) = \frac{1}{|\tau|} \sum_{t \in \tau} |\hat{Y}_{tl} - Y_{tl}|, \quad (5)$$

kde

\hat{Y}_{tl} je odhad celkové spotřeby neprůběhově měřených zákazníků za den t v lokalitě l ,

Y_{tl} je měřená hodnota vstupu do lokality l ve dni t po odečtení spotřeb všech průběhově měřených zákazníků,

τ je vyhodnocované období (v současné době celé měřené období od 1.6.2007 do 31.8.2010),

$|\tau|$ je počet dní období τ ,

je přímá vazba na výši případných finančních ztrát (i přes relativně komplikovaný systém cen plynu pro různé zákazníky lze přinejmenším odhadnout, o kolik peněz by společnost přicházela za předpokladu, že by odhad ztrát byl přesný). Distribuční společnosti však typicky zajímá také vztah k nějakému celku. To vede k použití různých typů relativních kritérií.

Například kritérium

$$Err_2(l) = 100 \cdot \frac{\sum_{t \in \tau} |\hat{Y}_{tl} - Y_{tl}|}{\frac{1}{L} \sum_{l=1}^L \sum_{t \in \tau} Y_{tl}}, \quad (6)$$

kde L je celkový počet vyhodnocovaných lokalit, vede ke stejnému uspořádání lokalit jako kritérium (5), hodnoty kritéria jsou však vztažené k celkové průměrné denní spotřebě všech vyhodnocovaných lokalit. Pochopitelně z hlediska uspořádání není hodnota normovací konstanty důležitá, normování je použito s ohledem na interpretovatelnost výsledků.

Kritérium (6) však „znevýhodňuje“ velké lokality, u kterých lze očekávat vyšší odchylky již z toho důvodu, že mají celkově vyšší spotřebu. Nejvíce budou tedy penalizovány lokality s nejvyššími odchylkami mezi všemi lokalitami. Tomu lze předejít užitím kritéria

$$Err_3(l) = 100 \cdot \frac{\sum_{t \in \tau} |\hat{Y}_{tl} - Y_{tl}|}{\sum_{t \in \tau} Y_{tl}}, \quad (7)$$

které posuzuje průměrnou denní odchylku vzhledem k průměrné denní spotřebě v dané lokalitě. Toto kritérium pochopitelně naopak více penalizuje lokality malé, kde je nízký základ díky nízké spotřebě. Nejvíce jsou tedy penalizovány lokality s vysokými odchylkami v rámci časového průběhu dané lokality.

Vzhledem k tomu, že odhad modelem GAMMA má pochopitelně také svou chybu, která je neměřitelná a promítá se do odhadu ztrát, lze uvažovat také o kritériích, která alespoň částečně vliv této chyby potlačují. Například lze zvolit lokalitu λ minimalizující kritérium 7 a posléze užít kritérium

$$Err_4(l) = \frac{1}{|\tau|} \sum_{t \in \tau} |(\hat{Y}_{tl} - Y_{tl}) - (\hat{Y}_{t\lambda} - Y_{t\lambda})|, \quad (8)$$

resp. jeho relativní podobu

$$Err_5(l) = \frac{\sum_{t \in \tau} |(\hat{Y}_{tl} - Y_{tl}) - (\hat{Y}_{t\lambda} - Y_{t\lambda})|}{\sum_{t \in \tau} Y_{tl}}. \quad (9)$$

4.2. Vliv chyby modelu

K posouzení vlivu chyby modelu byly vedle využití výše popsaných kritérií (8) a (9) provedeny experimenty s různými verzemi modelu GAMMA. Kromě poslední provozní verze modelu GAMMA určené k odhadu nevyfakturovaného plynu v Západočeské plynárenské, a.s. (ze září 2009 – ozn. jako verze 2.4.0) byly testovány další verze (se stejnou strukturou, odlišné pouze volbami při optimalizaci parametrů), a to:

verze 2.4.1.0 – optimalizace standardním způsobem (s využitím dat ZČP),

verze 2.4.1.1 – optimalizace s využitím dat ze zákaznického kmene RWE GasNet,

verze 2.4.1.2 – regionální optimalizace parametrů.

Kritériem pro porovnání jednotlivých verzí je relativní chyba odhadu posledního odečtu všech zákazníků, tj.

$$\frac{\sum_i \hat{Y}_i}{\sum_i Y_i}, \quad (10)$$

kde

Y_i je poslední měřená spotřeba zákazníka i ,

\hat{Y}_i je odhad spotřeby zákazníka i modelem GAMMA za období odpovídající spotřebě Y_i .

Vypočtené hodnoty kritéria (10) pro jednotlivé testované verze jsou uvedeny v tabulce 1. Ukazuje se dílčí zlepšení v jednotlivých po sobě jdoucích verzích.

Tabulka 1: Porovnání testovaných verzí modelu GAMMA na kmenových datech podle kritéria (10).

Verze	2.4.0	2.4.1.0	2.4.1.1	2.4.1.2
Přesnost [%]	97,67	98,15	98,53	98,69

K otestování vlivu verze modelu na porovnání lokalit lze využít například kritérium (6), jehož hodnoty jsou uvedeny v tabulce 2. Ač pozorujeme drobné (ve srovnání s celkovou hodnotou) rozdíly v hodnotách kritéria (6), výsledné pořadí lokalit se nemění. Zdá se tedy, že vliv verze modelu bude přinejmenším v první fázi vyhodnocování zanedbatelný.

Tabulka 2: Porovnání testovaných verzí modelu GAMMA na dostupných uzavřených lokalitách - kritérium (6).

Lokalita	verze modelu			
	2.4.0	2.4.1.0	2.4.1.1	2.4.1.2
8	128,25	127,68	129,91	131,00
7	78,61	75,32	76,10	77,35
9	44,29	41,70	40,79	39,56
4	29,41	27,05	27,32	27,32
5	25,74	23,35	23,04	22,85
1	24,01	21,98	21,85	21,81
6	10,35	9,82	9,75	9,66
3	9,70	9,49	9,36	9,13
2	5,19	5,02	4,89	4,74

4.3. Vyhodnocení

Jak bylo ukázáno v předchozím odstavci, volba verze modelu má na vyhodnocení ztrát v jednotlivých lokalitách zanedbatelný vliv. Pro vyhodnocení dostupných

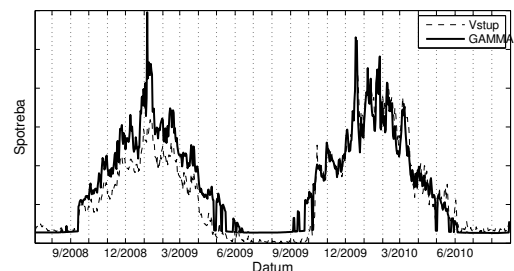
lokalit proto zvolíme jednu z dostupných verzí. S ohledem na nejlepší výsledky na kmenových datech použijeme regionální verzi modelu (2.4.1.2).

Porovnání podle kritéria (6) poskytuje tabulka 2. Kritérium (5) dává stejné uspořádání, proto je z úsporných důvodů vynecháno. Porovnání podle zbylých kritérií uvedených v odstavci 4.1 je uvedeno v tabulce 3.

Tabulka 3: Porovnání testovaných lokalit podle kritérií (7), (8) a (9).

L.	$Err_3(l)$	L.	$Err_4(l)$	L.	$Err_5(l)$
7	48,730	8	3435,258	2	95,322
8	28,115	7	1268,978	3	49,797
3	25,838	9	479,103	7	43,227
2	23,602	2	59,767	6	27,201
1	21,886	3	327,209	8	27,104
4	20,512	4	312,672	1	16,246
6	16,505	1	307,173	4	11,708
9	16,209	6	293,452	9	9,710
5	12,386	5	0,000	5	0,000

Při pohledu na pořadí podle prvních tří kritérií figurují v popředí (jakožto nejhorší) lokality 8 a 7. V lokalitě 8 je však pozorováno podezřelé chování vstupních hodnot, jak ukazuje obrázek 2. Zdá se, že cca od října 2008 do října 2009 vypadlo měření na jednom nebo více ze čtyř vstupních bodů do lokality. V této lokalitě je proto zapotřebí prověřit kvalitu vstupních měřidel. V lokalitě 7 lze pozorovat obdobný problém (obrázek vynechán z úsporných důvodů). Vynecháme-li tyto lokality, lze věnovat zvýšenou pozornost lokalitám 4 a 5 podle kritéria (6), případně 2 a 3 podle dalších kritérií. V současné době bylo na základě prezentovaných výsledků rozhodnuto o osazení lokalit 7 a 4 průběhovými měřeními ve všech vstupních i výstupních bodech a k revizi průběhového měření vstupu v lokalitě 8.



Obrázek 2: Porovnání časového průběhu vstupu a odhadu celkové spotřeby modelem GAMMA v lokalitě 8 za dostupné časové období.

5. Závěr

Byla uvedena problematika přepravy, měření a odhadu spotřeby zemního plynu ve světě a (zejména) v ČR. Dále byl prezentován projekt, v rámci něhož je ve spolupráci pracovních skupin ÚI AV ČR, v.v.i., a RWE GasNet, s.r.o., řešena problematika detekce zvýšených ztrát v uzavřených lokalitách.

Projekt je v počáteční fázi, kdy probíhá vývoj metodiky vyhodnocování odhadů ztrát. Jsou testována různá kritéria hodnocení ztrátových procesů a na základě nich vytipovávány lokality s největšími problémy. Ty mohou být způsobeny jak chybami v předávaných údajích (například výpadky měření na vstupu, chyby v průběhových měřeních spotřeb velkých zákazníků apod.), tak skutečnými zvýšenými ztrátami. Také pochopitelně mohou hrát roli další zatím nepředvídané vlivy.

Uvedení do praxe (rutinní zpracování stovek až tisíců uzavřených lokalit v rámci celé distribuční sítě) musí kromě masivního čištění zákaznické databáze předcházet ještě minimálně otestování na měřených údajích o ztrátách, což obnáší osazení vybraných lokalit průběhovými měřeními (všechny vstupy i odběrná místa). Tyto lokality budou vybrány na základě výsledků současné fáze projektu.

Poděkování

Autor studie děkuje svému školiteli Emilu Pelikánovi za revizi textu a cenné věcné i formální připomínky, dále pak celé řešitelské skupině projektu GAMMA.

Literatura

- [1] H. G. Berrisford, *The relation between gas demand and temperature: A study in statistical demand forecasting*. Operational Research Quarterly 16 (2), 229-246, 1965.
- [2] L. Blahová, J. Bečvář, M. Pihera, E. Pelikán a O. Konár, *Využití matematického modelování spotřeby zemního plynu*. Plyn: odborný měsíčník pro plynárenství 88 (10), 216-218, 2008.
- [3] M. Brabec, O. Konár, E. Pelikán a M. Malý, *A Nonlinear Mixed Effects Model for the Prediction of Natural Gas Consumption by Individual Customers*. International Journal of Forecasting 24 (4), 659-678, 2008.
- [4] M. Brabec, O. Konár, M. Malý, E. Pelikán a J. Vondráček, *A Statistical Model for Natural Gas Standardized Load Profiles*. Journal of the Royal Statistical Society Series C – Applied Statistics 58 (1), 123-139, 2009.
- [5] M. Brabec, M. Malý, E. Pelikán a O. Konár, *Statistical Model of Segment-Specific Relationship Between Natural Gas Consumption and Temperature in Daily and Hourly Resolution*. In *Natural Gas*. Rijeka: Sciyo, 393-416, 2010.
- [6] M. Brabec, M. Malý, E. Pelikán a O. Konár, *Consumption Model Calibration and Related Statistical Problems*. In *Applied Economics, Business and Development*. Athens: WSEAS Press, 141-146, 2009.
- [7] M. Brabec, M. Malý, E. Pelikán a O. Konár, *Statistical Calibration of the Natural Gas Consumption Model*. WSEAS Transactions on Systems 8 (7), 902-912, 2009.
- [8] J. Čermáková, J. Matějovic, O. Naxerová, J. Bečvář, M. Brabec, T. Brabec, O. Konár, M. Malý, E. Pelikán, M. Šimůnek a J. Vondráček, *Matematické modelování spotřeby zemního plynu zákazníků bez průběhového měření*. Plyn: odborný měsíčník pro plynárenství 85 (2), 34-37, 2005.
- [9] D. Ivezic, *Short-term natural gas consumption forecast*. FME Transactions 34, 165-169, 2006.
- [10] R. Fildes, A. Randall, and P. Stubbs, *One day ahead demand forecasting in the utility industries: Two case studies*. Journal of the Operational Research Society 48 (1), 15-24, 1997.
- [11] B. Geiger and M. Hellwig, *Gaslastprofile für nicht leistungsgemessene Kunden*. BWK – Das Energie Fachmagazin 55, 72-78, 2003.
- [12] S. Gil and J. Deferrari, *Generalized model of prediction of natural gas consumption*. Transactions of the ASME 126, 90-97, 2004.
- [13] O. Konár, *Matematické modelování spotřeby zemního plynu domácností a maloodběratelů*. In *Sborník ROBUST 2006*. Praha: JČMF, 151-158, 2006.
- [14] O. Konár, *Model spotřeby zemního plynu s náhodnými parametry*. In *Projekty GA AV ČR „Informační společnost“*. Praha: Ústav informatiky AV ČR. [Seminář projektů GA AV ČR Informační společnost 1ET400300513 a 1ET400300414, Beroun, 23.04.2007-24.04.2007, CZ].
- [15] O. Konár, *Nelineární model se smíšenými efekty pro odhad spotřeby zemního plynu*. In *Doktorandské dny 07*. Praha: Ústav informatiky AV ČR, v. v. i. & MATFYZPRESS, 2-7, 2007.
- [16] O. Konár, *Detekce zvýšených ztrát v distribuční síti zemního plynu*. In *ROBUST 2010*. Kniha abstraktů. Praha: JČMF, 22-22, 2010.
- [17] F. K. Lyness, *Gas demand forecasting*. Statistician 33, 9-21, 1984.

- [18] E. Pelikán, K. Eben, M. Šimůnek, M. Kolman a J. Hais, *Predikční systém Elvíra pro predikci spotřeby zemního plynu*. Plyn: odborný měsíčník pro plynárenství 80 (11), 253-255, 2000.
- [19] E. Pelikán, K. Eben a M. Šimůnek, *Predikce spotřeby elektrické energie a zemního plynu*. AUTOMA, odborný časopis pro automatizační techniku 7 (4), 15-18, 2001.
- [20] E. Pelikán, M. Brabec, M. Malý, O. Konár a J. Vondráček, *Zkušenosti se statistickým modelováním spotřeby zemního plynu*. In ROBUST 2008. Praha: JČMF, 315-341, 2009.
- [21] P. Potočník, M. Thaler, E. Govekar, I. Grabec, and A. Poredoš, *Forecasting risks of natural gas consumption in Slovenia*. Energy Policy 35, 4271-4282, 2007.
- [22] P. Potočník, E. Govekar, and I. Grabec, *Short-term natural gas consumption forecasting*. Proceedings of the 16th IASTED International Conference on Applied Simulation and Modelling, Palma de Mallorca, August 29-31, 2007 (A Publication of IASTED, 581). ACTA Press, Anaheim, 35-357, 2007.
- [23] E. F. Sánchez-Úbeda and A. Berzosa, *Modeling and forecasting industrial end-use natural gas consumption*. Energy Economics 29, 710-742, 2007.
- [24] J. Suykens, P. Lemmerling, W. Favoreel, B. De Moor, M. Crepel, and P. Briol, *Modelling the Belgian gas consumption using neural networks*. Neural Processing Letters 4, 157-166, 1996.
- [25] A. Svoboda a kol., *Plynárenská příručka – 150 let plynárenství v Čechách a na Moravě*. Praha: GAS s.r.o., 1997.
- [26] I. Vajk and J. Hetthéssy, *Load forecasting using nonlinear modelling*. Control Engineering Practice 13 (7), 895-902, 2005.
- [27] E. Valor, V. Meneu, and V. Caselles, *Daily Air Temperature and Electricity Load in Spain*. Journal of Applied Meteorology 40 (8), 1413 – 1421, 2001.
- [28] J. Vondráček, E. Pelikán, O. Konár, J. Čermáková, K. Eben, M. Malý a M. Brabec, *A Statistical Model for the Estimation of Natural Gas Consumption*. Applied Energy 85 (5), 362-370, 2008.
- [29] *Vyhláška č. 524/2006 Sb. o pravidlech pro organizování trhu s plynem a tvorbě, přiřazení a užití typových diagramů dodávek plynu (příloha 9) se změnami: 184/2007 Sb., 321/2007 Sb., 354/2008 Sb.*. Energetický regulační úřad, 2006-2008.
- [30] *Vyhláška č. 365/2009 Sb. o Pravidlech trhu s plynem*. Energetický regulační úřad, 2009.