

Prohlášení:

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně a že všechny využité zdroje jsou citovány v Referencích.

Praze, dne 4. července 2011

Martin Isoz

Název práce: Umělé neuronové sítě pro modelování výtěžků pyrolýzy výševaroucích surovin

Abstrakt: Cílem práce bylo zhodnotit možnosti predikce výtěžků vybraných produktů ethylenové pyrolýzy modelů založených na umělých neuronových sítích. Práce využívá experimentálních dat získaných při provozních experimentech na provozních pyrolyzních pecích vybavených vlásenkovými reaktory SRT III (Lummus). Použitá data zahrnovala provozní měření pyrolýzy primárních benzinů a atmosférického plynového oleje. Složení primárních benzinů bylo známo na komponentní úrovni, složení APO bylo approximováno metodou zástupných reálných komponent. Tato složení byla převedena na zastoupení uhlíkatých fragmentů. Informace o surovině byla doplněna hodnotami hustoty, destilační křivky a molekulové hmotnosti.

Navržené modely byly založeny na dopředné neuronové síti s jednou skrytou vrstvou. Vstupy do sítě byly tvoreny fragmentovým složením vstupní suroviny případně doplněným různě volenými doplňkovými charakteristikami a procesními parametry (výstupní teplotou, zatížením a ředícím poměrem). Výstupem byly predikce výtěžků čtyř hlavních produktů pyrolýzy, ethylenu, propylenu, methanu a bilančního oleje.

Byla sledována kvalita modelu v závislosti na volbě doplňkových charakteristik a architektuře využívaných sítí. Ukázalo se, že nejlepších výsledků dosahují modely sestávající z vyššího počtu sítí specializovaných na predikci výtěžku jediného produktu. Naopak se nepodařilo prokázat nezbytnost dalších doplňkových charakteristik suroviny.

Title: Artificial neural networks for modeling steam-cracking yields of heavier pyrolysis feedstocks

Abstract: The study was aimed at evaluating the capability of artificial neural networks based models to predict the cracking yields of heavier pyrolysis feedstocks. Experimental data used in the paper were obtained from full-scale experiments carried out on industrial pyrolysis furnaces equipped with SRT III (Lummus) pyrolysis reactors. Used data contained plant measurements of primary naphthas and atmospheric gas oils pyrolysis. Composition of primary naphthas was known on componental level, while AGO composition was approximated by substitute composition method based on real components. Those compositions were transformed to simplified characterization of the mixture represented by abundance of different carbon increments. Values of density, destillation curve and molecule weight were added to the feedstock characterizations.

Proposed models were based on feedforward neural networks with one hidden layer. Network inputs constituted of increment composition of the feedstock, complementary characteristics and process parameters (such as coil outlet temperature, feed rate and dilution ratio) were included optionally. Model outputs constituted of cracking yields prediction for four main pyrolysis products: ethylene, propylene, methane and residual oil.

The quality of models in relation to the selection of complementary characteristics and neural network architecture was studied. The best results have been obtained by usage of models containing higher number of networks, that were specialized in prediction of feedstock of one product. On the other hand, the necessity of using complementary feedstock characteristics in models hasn't been proved.

Poděkování:

Chtěl bych poděkovat svému školiteli, doc. Ing. Petrovi Zámostnému, Ph.D. za trpělivost a mnoho cenných rad, své rodině za vytvoření příkladného studijního prostředí a panu doc. RNDr. Tomáši Vaňkovi, CSc. za čas, který věnoval přípravě v této práci využívaných dat.

Obsah

1	Úvod	7
2	Teoretická část	8
2.1	Základní způsoby charakterizace ropných surovin	8
2.1.1	Komponentní cesta	9
2.1.2	Skupinová cesta	9
2.2	Charakterizace složení ropných surovin pro matematické modelování	9
2.2.1	Skupinové složení - reprezentace skupin	10
2.2.2	Přímá aplikace termodynamických a fyzikálních vlastností	11
2.2.3	Náhradní složení z fyzikálních charakteristik	12
2.2.4	Slučování látek do skupin (Lumping)	13
3	Přehled dat	14
3.1	Provozní experimenty v Chemopetrol Litvínov	14
3.2	Simulace na Ústavu chemického inženýrství VŠCHT Praha	14
3.3	Příprava dat pro modelování	14
4	Koncepce matematických modelů	16
4.1	Neuronové sítě	16
4.1.1	Struktura	16
4.1.2	Učení	16
4.2	Architektura modelů	19
4.2.1	Algoritmus optimalizace architektury	19
4.2.2	Jednosítový model	20
4.2.3	Čtyřsítový model	21
5	Výsledky a diskuze	23
5.1	Optimalizace architektury modelů	23
5.2	Porovnání čtyřsítového a jednosítového modelu	25
5.3	Vliv dodatečných informací o surovině	27
5.4	Změna přesnosti s rozšířením datové sady	27
5.5	Další možnosti zpřesnění modelů	30
6	Závěr	31
7	Seznam symbolů	32
A	Algoritmus výběru sítí pro jednosítový model	35
B	Tabulky využívaných dat	37
C	Histogramy četností počtu neuronů ve skryté vrstvě pro vybrané čtyřsítové modely	41
D	Přehled průměrů a výběrových rozptylů absolutních a normovaných odchylek	46
E	Porovnání změny mezikvartilového rozpětí a střední hodnoty normovaných odchylek modelů	50
F	Výsledky predikce výtěžků pyrolýzních produktů modelem modulMDL_1	54
F.1	Učení pouze na primárních benzínech	54
F.2	Učení na všech dostupných datech	55

1 Úvod

Ethylen, který je základem výroby tří nejdůležitějších termoplastů, je pravděpodobně v současném chemickém průmyslu chemikálií číslo jedna[1]. Je tedy nasnadě, že pyrolýza ropných frakcí na ethylen a ostatní základní chemikálie, kterou vzniká ročně 97 % komerčně vyráběného ethylenu[2] je jedním z nejdůležitějších chemických procesů vůbec.

Komerční pyrolýza, neboli tepelné štěpení uhlovodíků vodní parou, je prováděna téměř výhradně ve vlásenkových reaktorech umístěných v pyrolýzních pecích. Surovina je v konvektivní části pece předehřívána a mísená s vodní parou. Vlastní pyrolýza je pak prováděna v trubce (vlásence) v radiační části pece[1].

Výtěžek ethylenu získaného pyrolýzou potom závisí jak na parametrech pyrolýzní pece tak na typu štěpené suroviny. Vhodná surovina pro pyrolýzu by měla mít poměr H/C = 2. Tomuto požadavku již nevyhovují cykloalkany s několika kondenzovanými kruhy a vůbec nevhodnou složkou nástřiku jsou aromatické uhlovodíky[1].

Nejlepší surovinou pro pyrolýzu je z pohledu výtěžku ethylenu ethan, který je ale snadno dostupný a využívaný v zemích, kde se těží ropa doprovázená ložiskovými plyny se značným obsahem ethanu a propanu (např. USA). V zemích, která takovýmito ložisky nedisponují (např. Evropa) je nejrozšířenější surovinou pro pyrolýzu lehký a střední benzín.

Nejvyššími frakcemi ropy zpracovatelnými v klasickém pyrolýzním reaktoru jsou primární petrolej a plynový olej. Práci s těžšími surovinami zabraňuje vysoký podíl aromátů v těchto frakcích způsobující takovou míru koksování, že hrozí nebezpečí propálení trubek reaktoru v místech zhoršeného přestupu tepla[1]. Přesto jsou však z mnohých důvodů stále hledány cesty zpracování těžších ropných frakcí na ethylen. Žádná z těchto technologií však ještě nedosáhla průmyslového významu[2].

Vzhledem ke svému významu je pyrolýza velmi intenzivně studována. Velké úsilí je také věnováno konstrukci matematických modelů umožňujících predikci výtěžků tohoto procesu. Do současnosti byly vyzkoušeny tři typy modelů, molekulární, mechanistické a empirické, přičemž molekulární a mechanistické typy modelů se osvědčily zejména v aplikacích na pyrolýzu lehkých ropných frakcí, u nichž je možné zjistit přesné složení suroviny. Empirické modely sice potřebují širší datovou základnu a také jsou obecně výpočetně složitější, ale lze je využít i bez znalosti přesných mechanismů reakcí probíhajících v pyrolýzní a jsou tedy použitelné i pro predikci výtěžků pyrolýzy vyšších ropných frakcí[2].

Jelikož pro využití empirických modelů není třeba znát mechanismy ve směsi probíhajících reakcí, není k jejich využití nezbytná znalost přesného složení zpracovávané suroviny. Přesto je však nutné ropnou surovinu nějakým způsobem charakterizovat. Metodám charakterizace ropných surovin je věnována Teoretická část práce.

V Experimentální části práce bylo pracováno s metodou charakterizace suroviny pomocí zástupného složení z reálných komponent vyvinutou na Ústavu chemického inženýrství VŠCHT Praha a byla zkoumána schopnost modelů založených na umělých neuronových sítích predikovat výtěžky pyrolýzy v závislosti na architektuře využívaných sítí a různých kombinacích dat, kterými byla doplnován popis suroviny získaný výše zmíněnou metodou.

2 Teoretická část

Hlavní přínos studia pyrolýzy je spatřován ve zlepšování schopnosti posuzovat kvalitu zpracovávané suroviny. K tomu je možné využít dvou různých přístupů. V prvním je na kvalitu suroviny, z důvodu nemožnosti určit skutečné složení směsi výsevroucích uhlovodíků, usuzováno z analýzy strukturních skupin (např. alkanů, alkenů, monoaromátů a polyaromátů) pomocí moderních analytických technik jako jsou plynová (GC) a kapalinová (HPLC) chromatografie nebo magnetická rezonance (NMR) či hrubějšího, skupinového popisu suroviny s využitím BMCI nebo destilačních křivek[3]. Tato cesta bude dále rozvedena v 2.1.

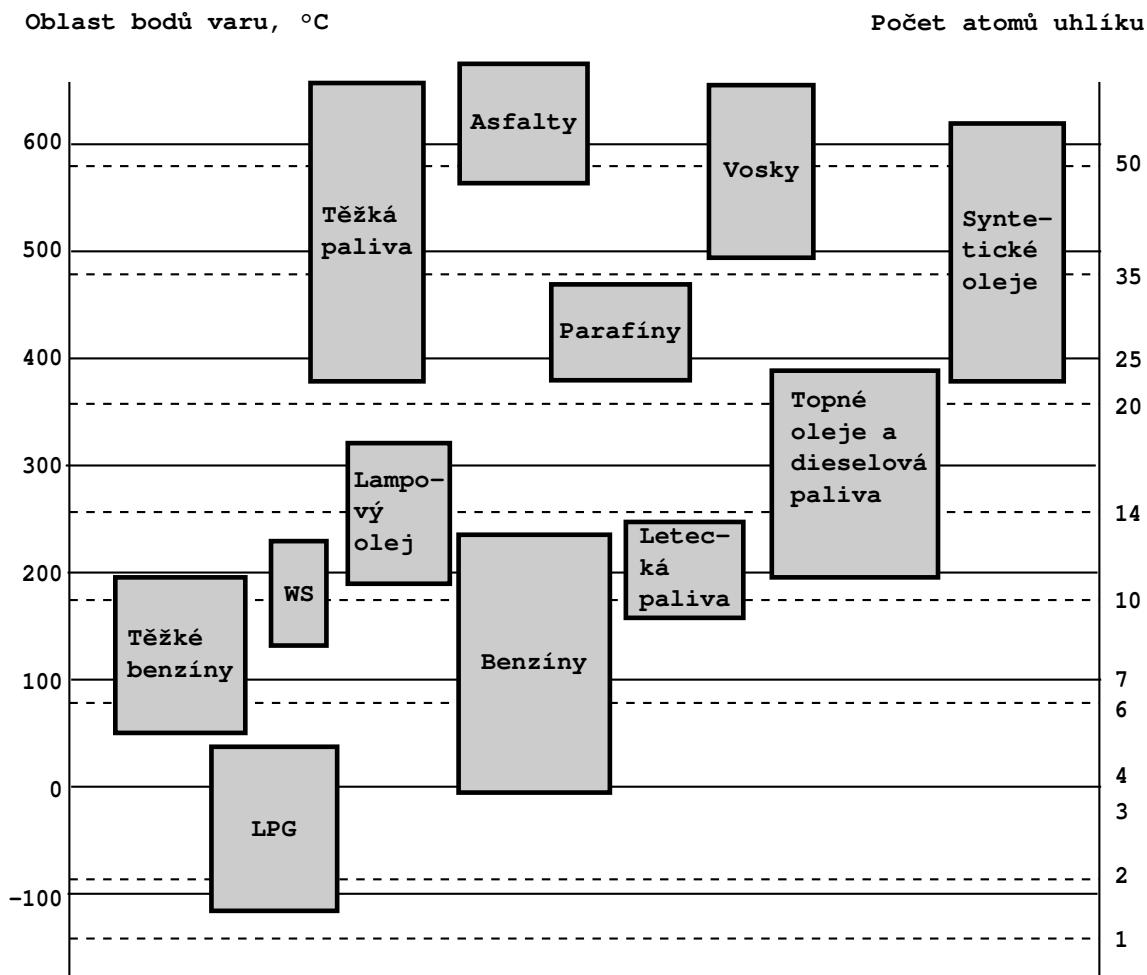
Druhý způsob určení kvality suroviny spočívá v sestrojení charakterizace suroviny a následném matematickém či fyzikálním modelování pyrolyzního procesu. Tato charakterizace by měla být v co nejlepší korelace s chemickými i fyzikálními vlastnostmi dané suroviny. Způsobem popisu ropné suroviny pro matematické modelování bude věnována kapitola 2.2.

V pyrolyze se využívají ropné frakce očištěné od všech heteroatomů a proto zde nebudu zmíněny metody sloužící k popisu podílu a povahy těchto neuhlovodíkových složek.

2.1 Základní způsoby charakterizace ropných surovin

Ropné produkty jsou na základě svého využití, bodů varu a počtu uhlíkových atomů děleny do skupin, ropných frakcí. Přehled jednotlivých skupin je na Obrázku 1.

Zařazení uhlovodíku do některé z frakcí však neposkytuje nijak výrazné vodítko na jeho fyzikální či chemické vlastnosti, jelikož tyto jsou velmi variabilní nejenom mezi frakcemi, ale i v nich. Je tedy velmi žádoucí naleznout takové metody popisu ropných surovin, které umožní danou směs charakterizovat co nejpřesněji.



Obrázek 1 – Hlavní ropné produkty, rozsahy jejich bodů vatů a počtu uhlíků. Obrázek podle [4].

2.1.1 Komponentní cesta

Při charakterizacích sloučeniny způsoby komponentní cesty je využíváno procentuálního zastoupení jednotlivých komponent směsi. Pojem komponenta je zde užíván velmi vágně, jelikož se může jednat o složku reálnou i pseudosložku. V případě reálných komponent také může být tento popis získán jak analytickými metodami jako přesné složení, tak různými algoritmy jako složení zástupné.

Vlastnosti těchto komponent jsou buď známé (velmi často v případě reálných složek) nebo je nutné je nějakým způsobem korelovat. Podrobnější popis možností charakterizace sloučeniny skrze její komponenty je vzhledem k jeho důležitosti pro matematické modelování uveden v části 2.2.3.

2.1.2 Skupinová cesta

Skupinová cesta je založena na charakterizaci suroviny pomocí procentuálního zastoupení určitých skupin, například n-alkanů či látek vroucích v jistém teplotním rozmezí. Dále tyto informace bývají doplněny dalšími ukazateli majícími význam průměrných hodnot pro celou směs.

Přehled některých nejběžnějších skupinových charakteristik suroviny bude uveden v následujícím textu.

PIONA je akronymem pro Parafiny, Isoparafiny, Olefiny, Nafteny a Aromáty. Jak je zřejmé již z názvu, jedná se o míru značící procentuální zastoupení těchto typů řetězců v surovině. Jedná se tedy o integrální charakteristiku popisující směs jako celek.

BMCI je zkratkou pro Bureau of Mining Correlation Index, jedná se o míru kvality suroviny pro zisk amorfního uhlí při pyrolýze těžkých ropných frakcí a je možné jej pro danou surovинu dopočítat ze vztahu:

$$\text{BMCI} = \frac{48.640}{\text{VABP}} + 473.7S - 456.8, \quad (1)$$

kde VABP je objemový střední bod varu (viz. [4]) a S je standardní specifická hustota (viz. 2.2.1). BMCI je dalším z doplňkových faktorů charakterizujících celou směs.

ASTM D86 je pro jednoduchost experimentálního uspořádání nejčastější metodou získávání informací o destilačním chování ropných surovin[4,5]. Využívá se pro charakterizace benzínů, kerosenů, plynových olejů a podobných ropných surovin. Pro popis těžších ropných frakcí se využívá nízkotlaká obdoba této metody, ASTM D1160.

Jedná se o destilaci za atmosférického tlaku v koloně s přibližně patnácti teoretickými patry a vysokou hodnotou refluxního poměru.

Výstupem metody je křivka závislosti bodů varu na oddestilovaném množství suroviny, která je snadno převeditelná na TBP¹ destilační křivku (příklad takové křivky je na Obrázku 2).

SIMDIST je dalším ze způsobů, jak obdržet TBP destilační křivku. Jedná se o destilaci simulovanou pomocí plynové chromatografie. Výstup metody je zde velmi podobný jako v případě ASTM D86, ale díky výrazně lepší separační schopnosti plynové chromatografie oproti destilaci se zde může počet teoretických patr kolony pohybovat mezi 100 a 10^5 .

Výsledky získané touto metodou snadno převeditelné i na ASTM D86 destilační křivku.

2.2 Charakterizace složení ropných surovin pro matematické modelování

S rozvojem výpočetní techniky a numerické matematiky v posledních desetiletích se otevřely dveře pro možnost sestavování takových empirických modelů pyrolýzních procesů, které lze využívat k modelování pyrolýzy výsevroucích, průmyslových surovin. Pro takovéto suroviny již ale není možné získat jejich přesné složení. Vyvstanul tak problém hledání takového způsobu charakterizace suroviny, který by vystihoval chemické i fyzikální vlastnosti suroviny natolik přesně, aby bylo možné na jejím základě usuzovat na procesy probíhající v reaktoru.

Takovýmto způsobem mohou být některé reprezentace skupin zmíněné v 2.1 nebo přímé využití termodynamických a fyzikálních vlastností suroviny.

¹TBP ... True Boiling Point

Většina těchto charakterizací je však v současnosti založena na sestavení zástupného složení studované suroviny. Toto zástupné složení je získáváno na základě popisu suroviny studovaného v 2.1, základem těchto metod bývá často informace o TBP destilační křivce [3,6].

Existují dvě alternativní stezky popisu suroviny zástupným složením lišící se v povaze komponentů, ze kterých je zástupné složení konstruováno. Tradičnějším přístupem je zavedení souboru pseudokomponent, hypotetických sloučenin s vhodnými fyzikálními vlastnostmi. Novější cestou prezentovanou Eckertem a Vaňkem v [7] a [8] je sestavování zástupného složení z reálných komponent, jejichž popis je dostupný ve zvolené databázi.

V některých případech, kdy i zástupné složení sestává z neúnosně vysokého počtu komponent je pak provedeno slučování komponent do skupin podle předem určených vlastností (Component lumping). Tento postup je popsán například v [6].

2.2.1 Skupinové složení - reprezentace skupin

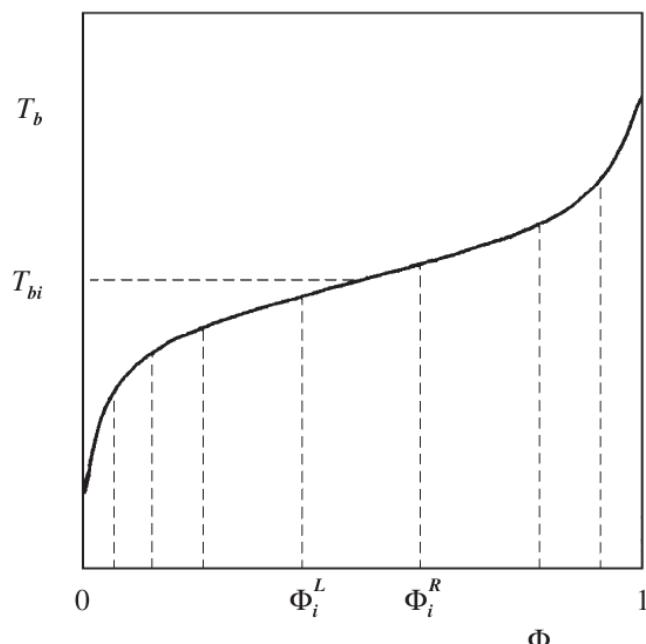
Ropa jako taková i těžší ropné frakce jsou velmi složité směsi velkého množství různých uhlovodíků, které mohou mít velmi podobné fyzikální vlastnosti. Například šestnáct z osmnácti oktanových izomerů má normální bod varu v rozmezí 12 °C[5]. Abychom mohli se surovinou pracovat, musíme jí rozdělit na jednotlivé frakce. Termínem ropná frakce se obvykle rozumí směs uhlovodíků s normálním bodem varu v jistém teplotním rozsahu. Obvykle jsou tyto rozsahy určovány podle následujícího klíče[4]:

- 10 °C pro lehké frakce s bodem varu pod 200 °C
- 15 °C pro frakce s bodem varu mezi 200 °C a 400 °C
- 20 °C pro frakce s bodem varu mezi 400 °C a 600 °C
- 30 °C pro frakce s bodem varu vyšším, než 600 °C

Obecně by měla při dělení do frakcí být zachována následující empiricky zavedená nerovnost:

$$0.020 < \frac{dT_b}{T_b} < 0.035, \quad (2)$$

kde dT_b je šířka teplotního intervalu v [K], ve kterém leží normální body varu jednotlivých sloučenin a T_b je průměrný normální bod varu dané frakce v [K].



Obrázek 2 – Ukázka závislosti normálních bodů varu T_b na oddestilovaném množství Φ pro mohasložkovou směs. Obrázek podle [5].

Rozdelením ropné suroviny do frakcí tedy „rozřežeme“ závislost normálního bodu varu na oddestilovaném množství, TBP křivku, na jednotlivé úseky, jak je zobrazeno na Obrázku 2, přičemž každý interval $\langle \Phi_i^L, \Phi_i^R \rangle$ představuje jednu frakci[7].

Střední normální bod varu pak můžeme stanovit jako aritmetický průměr krajů intervalu nebo přesněji z věty o střední hodnotě:

$$T_{bi} = \frac{1}{\Phi_i^R - \Phi_i^L} \int_{\Phi_i^L}^{\Phi_i^R} T_b(\Phi) d\Phi \quad (3)$$

Další veličinou, kterou je nutné u každé frakce určit je její specifická hustota, S_i . Tu získáme podělením hustoty dané frakce při 288 K hustotou vody při téže teplotě, neboli:

$$S_i = \frac{d_i}{999.024} \quad (4)$$

V dalším textu, je-li pracováno s pojmy normální bod varu a standardní specifická hustota, jsou tyto charakteristiky myšleny pro jednotlivé frakce a značeny T_b a S .

2.2.2 Přímá aplikace termodynamických a fyzikálních vlastností

Výroba a zpracování ropy sestává z mnoha různých procesů, jejichž návrh a konstrukce vyžadují co nejpřesnější znalosti mnoha termodynamických a fyzikálních vlastností suroviny, jako je například hustota, entalpie, tepelná kapacita, viskozita nebo fugacita. Tyto veličiny je možné spočítat z termodynamických vztahů, jako jsou například stavové rovnice nebo generalizované diagramy.

Všechny tyto vztahy ale vyžadují jisté vstupní veličiny, z nichž můžeme jnenovat například kritickou teplotu, T_C , kritický tlak, P_C , krotický molární objem, V_C , acentrický faktor, ω či molární hmotnost M_W . Bohužel jsou všechny tyto kritické veličiny v případě ropných frakcí pseudokritické a tedy neměřitelné.

Je tedy nutné nalézt vztahy, které umožní dopočet hodnot těchto veličin z měřitelných údajů, jako je například bod varu (získaný například z TBP nebo ASTM D86) nebo standardní specifická hustota směsi[9].

Vztahy pro výpočet těchto veličin jsou empirické povahy a tudíž jsou vzhledem k průmyslové důležitosti tohoto odvětví předmětem neustálého výzkumu. V textu bude uvedeno několik stručných příkladů výpočtu zmínovaných veličin dle [4], kde je základem navržených vztahů právě normální bod varu, T_b a standardní specifická hustota, S .

Normální bod varu a standardní specifická hustota jsou spolu svázány přes Watsonův charakterizační faktor, K_W dle vztahu:

$$K_W = \frac{(1.8T_b)^{\frac{1}{3}}}{S} \quad (5)$$

Tento charakterizační faktor je využíván v mnohých výpočetních metodách (například při přepočtu hodnot měřených při destilaci za sníženého tlaku na tlak atmosférický dle Maxwella a Bonnela). Pro frakce s bodem varu vyšším, než 200 °C získané destilací shodné ropné suroviny pak bývá Watsonův faktor konstantní.

Pro uhlovodíky náležející do různých skupin (dle PNA²) potom dosahuje přibližně následujících hodnot:

- $K_W \approx 13$ pro parafiny
- $K_W \approx 12$ pro nafteny (cykloalkany)
- $K_W \approx 10$ pro aromáty

Další veličiny lze poté ze standarní specifické hustoty a normálního bodu varu frakce vypočítat například metodami API založenými a vztazích definovaných Riazim a Daubertem na konci osmdesátých let. Tyto vztahy jsou využitelné pro frakce s molární hmotností mezi 70 a 300, což odpovídá počtu uhlíků 5 - 22[9]. Tyto metody by tedy byly využitelné pro práci s frakcemi, s nimiž je v této práci operováno (primární benzíny a APO).

Pro výpočet základních veličin k charakterizaci ropné frakce, M , P_C , T_C navrhují zmíněné metody vztahy:

²PNA ... Paraffins, Naphtenes, Aromatics

$$M_W = 42.965 [\exp(2.097 \cdot 10^{-4} T_b - 7.78712S + 2.08476 \cdot 10^{-3} T_b S)] \cdot (T_b^{1.260007} S^{4.98308}) \quad (6)$$

$$T_C = 189.8 + 450.6S + T_b(0.4244 + 0.1174S) + \frac{(14.410 - 100.688S)}{T_b} \quad (7)$$

$$\begin{aligned} \ln P_C &= 5.68925 - \frac{0.0566}{S} - 10^{-3} T_b \left(0.436392 + \frac{4.12164}{S} + \frac{0.213426}{S^2} \right) \\ &+ 10^{-7} T_b^2 \left(4.75794 + \frac{11.819}{S} + \frac{1.53015}{S^2} \right) - 10^{-10} T_b^3 \left(2.45055 + \frac{9.901}{S^2} \right) \end{aligned} \quad (8)$$

kde T_b je normální bod varu frakce v [K], S je standardní specifická hustota, M_W je molární hmotnost v [kg/kmol], T_C je kritická teplota v [K] a P_C je kritický tlak v [bar].

Dále lze těmito metodami dopočítat napříkladacentrický faktor.

Důvodem velké obliby API metod je také skutečnost, že jejich koeficienty jsou průběžně zpřesňovány[9].

Krom přímých výpočtů jiných termodynamických veličin bývají tyto empirické charakteristiky využívány také při získávání pseudokomponentového složení frakcí[5].

2.2.3 Náhradní složení z fyzikálních charakteristik

V současnosti nejrozšířenější přístupy k charakterizaci složení ropných surovin pro matematické modelování vychází z konstrukce zástupného složení suroviny na základě jejích měřitelných vlastností z nichž nejdůležitější je destilační křivka, která ale bývá doplněna dalšími údaji, z nichž je možné jmenovat například křivky závislosti molární hmotnosti, hustoty či viskozity na oddestilovaném množství[10].

Jak již bylo zmíněno výše, existují dva různé přístupy k sestavování zástupného složení suroviny. První část postupu je pro oba přístupy shodná, surovina je rozdělena na jednotlivé frakce tak, jak je ukázáno v 2.2.1 a pro frakce jsou buď získány další charakterizační křivky, které jsou následně převedeny na funkce ne oddestilovaného množství, ale středních normálních bodů varu tak, jak je nastíněno například v [10]. Poté se však postupy začínají lišit.

První z nich, vyvýjený již od sedesátých let minulého století a do dnešních dnů široce využívaný v simulačních programech (ASPEN PLUS, HYSYS PRO atd.) a tedy i v průmyslu je založen na nahrazování směsi systémem až sta pseudokomponent, jejichž fyzikální vlastnosti nutné pro modelování jsou dopočítávány navazujícími empirickými modely, z nichž jeden příklad je uveden v 2.2.2[7].

Tento postup má však, jak upozorňují Eckert a Vaněk v [7] několik nevýhod:

- V chemických reakcích bude velmi pravděpodobně hrát nějakou roli i chemický charakter jednotlivých komponent, avšak u pseudokomponent není možné jakkoliv definovat jejich chemismus.
- Pseudokomponenty jsou prvotně definovány pouze svým bodem varu a několika dalšími doplňkovými charakteristikami, zejména standardní specifickou hustotou, molekulovou hmotností a viskozitou. Všechny ostatní fyzikální vlastnosti nutné k modelování musí být dopočítávány na základě empirických vztahů, jejichž přesnost může být diskutabilní.
- S pseudokomponentami nelze využívat metod skupinových příspěvků (například UNIFAC) vyžadujících jako vstupní data informace o molekulární struktuře složek.
- Nelze snadno využít informací o typu směsi, například převládají-li parafiny či aromáta, či jaké typy polárních složek se ve směsi vyskytují.

Přesto je však možné po stanovení zástupného složení směsi z pseudokomponent s tímto složením kdekoliv krom procesů s chemickými reakcemi pracovat jako s reálnou směsí.

Druhý přístup navržený Eckertem a Vaňkem v [7] a v [8] a dále rozvinutý například Cuellarem a Le Rouxem v [11] nenahrazuje složení směsi uměle vytvořenými pseudosložkami, ale hledá v dostupných databázích chemických látek takové sloučeniny, jejichž kombinací je možné dosáhnout vlastností studované suroviny. Nevýhodou tohoto postupu oproti prvně zmínovanému je znatelně vyšší výpočetní náročnost a dostupnost dostatečně široké databáze chemických látek s normálními

body varu a dalšími vlastnostmi spadajícími do intervalů definovaných rozdelením suroviny na frakce.

Nezanedbatelnými výhodami prezentovanými v [7] jsou:

- Vlastnosti jednotlivých náhradních komponent jsou známé a dohledatelné v chemických databázích, není je tedy nutné dopočítávat a zavádět tak do výpočtu další chyby.
- Díky tomu, že se jedná o reálné sloučeniny, jsou v zástupném složení směsi nějakým způsobem reflektovány chemické vlastnosti původní sloučeniny a tuto charakterizaci je pak možné využít k modelování chemických dějů.
- K charakterizaci suroviny stačí nižší počet skutečných zástupných komponent, než jejich pseudokomponentních ekvivalentů - aplikace metody navržené Eckertem a Vaňkem pro pyrolýzu APO[10] pracuje s kombinací přibližně dvaceti sloučenin, kdežto například HYSYS charakterizuje ropnou surovинu až čtyřiceti pseudokomponentami[5].

Vzhledem k jejím výhodám byla metoda charakterizace ropné suroviny zástupným složením z reálných komponent využita v experimentální části práce pro predikci výtěžků pyrolýzy pomocí umělých neuronových sítí.

2.2.4 Slučování látek do skupin (Lumping)

V petrochemickém průmyslu jsou suroviny obvykle charakterizovány stovkami různých komponent, což činí simulace a matematické modelování procesů velmi náročným z důvodu jak komplexity modelů, tak jejich výpočetní náročnosti. Z tohoto důvodu je velmi často slučovat látky do skupin i v případě, že známe přesné složení suroviny.

Jako na první krok slučování můžeme pohlížet již na charakterizaci suroviny zástupným složením, avšak v průmyslové praxi je běžné, že se pracuje s více než jednou ropnou surovinou zároveň, čímž ale velmi narůstá počet dat, se kterými je nutné v průběhu simulace pracovat. Například Lin, Liebovici a Jørgensen[6] uvádějí, že běžnou průmyslovou aplikací je práce se šesti až deseti různými surovinami, čímž může počet jednotlivých frakcí pro něž je třeba nalézt zástupné složení narůst až ke jednomu tisíci.

Je tedy nasnadě, že je třeba počet dat nějakým způsobem redukovat za co nejmenší ztráty informací o surovině.

Do dnešní doby bylo vyvinuto velké množství snahy nalézt optimální způsob slučování látek do skupin, přičemž můžeme rozlišit dva různé přístupy: diskrétní, který předpokládá, že každá komponenta náleží pouze do jedné skupiny a kontinuální pracující s předdefinovanou distribuční funkcí jednotlivých komponent.

Podrobnější popis metod je dostupný například v již zmiňovaném článku [6].

3 Přehled dat

Datovým základem pro prezentovanou práci byla data měřená na průmyslových reaktorech společnosti Chemopetrol Litvínov v letech 2000 – 2003 doplněná informacemi z plynové chromatografie (GC) používaných surovin a údaji získanými simulacemi složení výševroucích surovin provedenými na Ústavu chemického inženýrství VŠCHT Praha.

3.1 Provozní experimenty v Chemopetrol Litvínov

Z provozních experimentů provedených v Chemopetrol Litvínov v letech 2000 – 2003 byly k dispozici informace o výtěžcích pyrolýzy v závislosti na datech o provozu a PIONA. Experimenty probíhaly převážně se třemi typy surovin, primárními benzíny (PrBi), atmosférickými plynovými oleji (APO) a hydrokrakovánými vakuovými destiláty (HCVD) a ve dvou typech vlásenek (SRT III a GK 6).

U primárních benzínů byla tato data doplněna údaji z GC daných surovin. Byla získána informace o skutečném složení suroviny, její střední molární hmotnosti (M_w), hustotě (d) a destilační křivce (T10, T50, T90).

V případě APO a HCVD nebyla informace o přesném složení k dispozici (a její získání je v současné době prakticky nemožné). Bylo tedy nutné hledat zástupné složení těchto surovin. K tomu byla využita metoda vyvinutá na Ústavu chemického inženýrství VŠCHT Praha[7,8].

3.2 Simulace na Ústavu chemického inženýrství VŠCHT Praha

Postupem popsaným v článcích [7,8] bylo hledáno zástupné složení APO a HCVD z reálných komponent u surovin, jejichž charakteristiky byly dostupné v experimentálních datech z Chemopetrol Litvínov.

Zde bylo nutné rezignovat na další práci s HCVD, jelikož v databázi využité k sestavení zástupného složení surovin nebylo k dispozici dostatečné množství látek s odpovídajícími vlastnostmi na to, aby jejich kombinací bylo možné získat směs s požadovanými vlastnostmi.

Celkem bylo simulací na Ústavu chemického inženýrství získána charakterizace deseti plynových olejů, což dohromady s dvaceti devíti datovými sadami pro primární benzíny poskytlo celkový soubor údajů pro třicet devět různých surovin.

3.3 Příprava dat pro modelování

Tabulka 1 – Přehled fragmentů využitých pro zjednodušenou charakterizaci suroviny

Popis fragmentu	Použitý symbol
H_2	H: vodík
CH_4	M: methan
CH_3- v alkanech a alkylech	P: primární uhlík
$-CH_2-$ v alkanech a alkylech	S: sekundární „acyklický“ uhlík
$-CH_2-$ v naftenických cyklech	S_N : sekundární „naftenický“ uhlík
$> CH-$ v isoalkanech a isoalkylech	T: terciární „acyklický“ uhlík
$> CH-$ v naftenických cyklech	T_N : terciární „naftenický“ uhlík
$> C <$ v isoalkanech a isoalkylech	Q: kvarterní „acyklický“ uhlík
$> C <$ v naftenických cyklech	Q_N : kvarterní „naftenický“ uhlík
$CH_2 =$ v alkenech	P_E : primární „alkenický“ uhlík
$CH \equiv$ v alkynech	P_Y : primární „alkynický“ uhlík
$-CH =$ v alkenech	S_E : sekundární „alkenický“ uhlík
$-C \equiv$ v alkynech	S_Y : sekundární „alkynický“ uhlík
$> C =$ v alkenech	T_E : terciární „alkenický“ uhlík
$)CH$ v aromatických jádrech	S_A : sekundární „aromatický“ uhlík
$)C-$ v aromatických jádrech	T_A : terciární „aromatický“ uhlík

Metoda vytvoření datového souboru pro modelování pyrolýzy s využitím umělých neuronových sítí vychází z postupu zmíňovaného v článku [10]. Nejprve bylo definováno sedmnáct různých frag-

mentů, jejichž kombinací bylo možno obdržet každou ze zpracovávaných sloučenin. Jejich přehlde je v Tabulce 1.

Následně byly základě reálného (u PrBi) či nalezeného zástupného (u APO) složení určeny hmotnostní zlomky jednotlivých fragmentů ve zpracovávaných surovinách. Informace o fragmentovém složení suroviny pak byla doplněna údaji o provozu reaktoru: teplotou na výstupu z vlásenky (COT), hmotnostním mástříkem suroviny na vlásenku a jejím ředícím poměrem pára/surovina.

Problémem výše zmíněného postupu je skutečnost, že získané fragmentové složení nemusí být unikátní. Nijak totiž nereflektuje velikost molekul v surovině.

Například dimethyloktan a tetramethylhexadekan budou mít shodný poměr $\text{CH}_3 - : - \text{CH}_2 -$ a tedy i stejné fragmentové složení. Obě tyto sloučeniny ale mají zcela jiné vlastnosti.

Jako řešení tohoto problému navrženo doplnění informací o fragmentovém složení a provozu reaktoru kombinací údajů o destilační křivce suroviny, její střední molární hmotnosti a hustotě, které se s velikostí molekul mění (viz. Tabulka 2).

Tabulka 2 – Přehled rozdílů průměrných doplňkových charakteristik suroviny mezi benzíny a APO

Surovina	M _w	d	T10	T50	T90
	[g mol ⁻¹]	[kg m ⁻³]	°C		
Primární benzín	93.62	704.28	97.96	197.67	292.74
APO	195.07	836.89	215.10	268.35	336.40

V práci pak byla zkoumána vhodnost tohoto doplnění původního datového souboru a byla hledána kombinace doplňkových údajů poskytující nejpřesnější výsledky.

Dále byly všechny následně využívané údaje o surovině, společně s výtežky čtyř vybraných, hlavních produktů zkombinovány do dvou tabulek poskytujících snadný přehled o využívaných datech (viz. Tabulky 7 a 8 v Příloze B).

V první tabulce jsou uvedeny informace o fragmentovém složení zkoumaných surovin a provozní data z jejich pyrrolízy jako základní soubor vstupních informací pro používané modely. Ve druhé se pak nachází přehled doplňkových údajů z GC a hmotnostní výtežky sledovaných produktů.

Nakonec byla dostupná data rozdělena do dvou sad, trénovací a validační, v poměru přibližně 6:4. Data byla rozdělena tak, aby bylo zachováno co nejrovnoměrnější rozložení výtežek jednotlivých produktů v obou sadách. V případě nesouladu byla z důvodu jejich průmyslové důležitosti upřednostňována rovnoměrnost rozložení u ethylenu a propylenu. Klíč k rozdělení dat do sad je v následující Tabulce 3.

Tabulka 3 – Rozdělení dat do sad

Typ suroviny	Datová sada	Indexy ³
Primární benzín	tréninková sada	2 , 3, 5, 6, 7, 10, 12, 14, 16, 19, 21, 22, 24, 25, 26, 27, 28
	validační sada	1, 4, 8, 11, 13, 15, 17, 18, 20, 23, 25, 29
APO	tréninková sada	se samotnými APO nelze pro nedostatek dat pracovat
	validační sada	
PrBi + APO	tréninková sada	2, 3, 5, 6, 7, 10, 12, 14, 16, 19, 21, 22, 24, 25, 26, 27, 28 30, 32, 33, 35, 37, 39
	validační sada	1, 4, 8, 11, 13, 15, 17, 18, 20, 23, 25, 29, 31, 34, 36, 38

³Indexy odpovídají pořadí pokusů tak, jak jsou uvedeny v Tabulkách 7 a 8 v Příloze B.

4 Koncepce matematických modelů

Při práci bylo k predikci výtěžků pyrolýzy středněvroucích uhlovodíků (primárních benzínů a APO) využíváno umělých neuronových sítí a metody charakterizace vstupní suroviny vyvinuté na Ústavu chemického inženýrství a Ústavu organické technologie VŠCHT Praha[10].

Byly zkoumány dva různé přístupy k problematice modelování pyrolýzy s využitím umělých neuronových sítí. V prvním byla k predikci výtěžků všech sledovaných produktů tepelného štěpení suroviny využita jediná neuronová síť. Nadále bude o tomto způsobu referováno jako o „jednosítovém modelu“. Ve druhém, zmíňovaném jako „čtyřsítový model“, pak odhad výtěžku každého z produktů obstarávají síť různé.

Dále byla sledována kvalita výstupu modelů v závislosti na doplnění charakterizace suroviny navržené Bělohlavem et al. v [10] různými informacemi z PIONA analýzy a počtu neuronů ve skryté vrstvě použitých neuronových sítí. Tímto postupem získané údaje byly následně využity pro návrh optimální architektury a rozsahu vstupních dat modelu.

4.1 Neuronové sítě

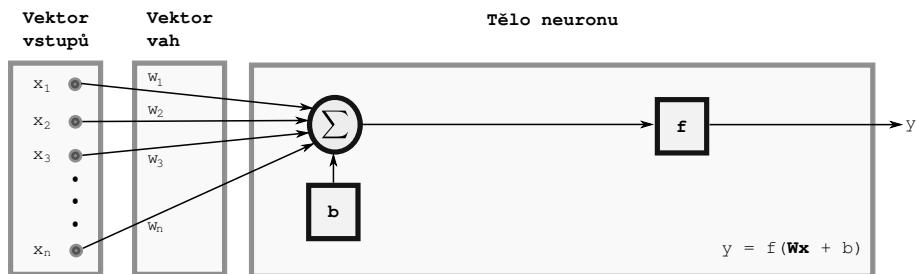
Jak jednosítový, tak čtyřsítový model jsou založeny na stejném souboru empirických modelů, umělých neuronových sítích. Jejich podobnost je však ještě hlubší, oba modely mají velmi podobnou strukturu využitých sítí a jsou trénovány stejným způsobem. V následujícím textu tedy budou stručně nastíněny principy stavby a učení umělých neuronových sítí.

4.1.1 Struktura

Umělé neuronové sítě (artificial neural networks, ANN) jsou nelineární mapovací systémy strukturě podobné lidskému nervovému systému[12]. Základní jednotkou takové sítě je umělý neuron, uzel sestávající z vektoru vstupu a vah, prahu, nelineární přenosové funkce a výstupu. Takovýto neuron potom lze zapsat jako matematickou funkci ve tvaru:

$$y = f \left(\left(\sum_{(k)} w_k x_k \right) + b \right), \quad (9)$$

kde y je výstup neuronu, \vec{x} je vektor vstupních dat, \vec{w} je vektor vah přiřazených vstupním datům, $f(\cdot)$ je nelineární přenosová funkce a b je práh (bias). Nákres umělého neuronu je na Obrázku 3.



Obrázek 3 – Schéma umělého neuronu

Celek systému je pak poskládán z určitého množství neuronů řazených ve vrstvách. Neurony v jedné vrstvě bývají shodného typu (stejný počet vstupů a stejná přenosová funkce) a nejsou propojeny mezi sebou, ale pouze s neurony v jiných vrstvách. Na Obrázku 7 je potom příklad plně propojené ANN s jednou skrytou vrstvou neuronů.

4.1.2 Učení

Neuronové sítě jsou aplikovány převážně v případech, kdy není znám vztah mezi vstupními a výstupními daty. Výstup neuronu lze měnit pouze manipulací s hodnotami vah přiřazených jednotlivým vstupům. Cílem učení neuronové sítě tedy je upravit tyto váhy tak, aby síť jako celek poskytovala co nejlepší výsledky.

V současnosti je nejvyužívanějším postupem učení ANN supervised backpropagation learning (učení zpětným vývojem s učitelem). Všechny metody založené na tomto principu využívají porovnání mezi cílovými a simulovanými daty na základě kterého různými způsoby upravují vektory vah pro jednotlivé neurony v síti.

V práci je pro trénink sítí využita Levenberg-Marquardtova (L.-M.) optimalizační metoda[13] pro backpropagation supervised learning s hodnocením založeným na průměru kvadrátů odchylek mezi simulovanými a cílovými daty (mean squared error, MSE). Úprava vektorů vah neuronů probíhá iteracně až do chvíle, než:

1. je dosaženo maximálního počtu opakování výpočtu
(epochs > max_epochs)
2. výpočet trvá příliš dlouho
(time > max_time)
3. je dosaženo cílové (minimální) hodnoty účelové funkce
(performance ≤ performance_goal)
4. gradient účelové funkce poklesne po určenou nejnižší přípustnou hodnotu
(grad ≤ min_grad)
5. adaptivní parametr μ překročí určenou nejvyšší přípustnou hodnotu
(mu ≥ mu_max)
6. počet neúspěšných validačních zkoušek překročí volenou nejvyšší přípustnou hodnotu
(fail > max_fail)

Dalším rozborem výše uvedených bodů lze zjistit, že je možné je rozdělit do tří skupin podle jejich vlivu na úspěšnost naučení sítě.

První skupinou, do které lze zařadit body 1 a 2 je ukončení tréninku z důvodu přílišné časové či výpočetní náročnosti. Síť, jejíž učení bylo ukončeno kvůli některému z těchto důvodů tedy velmi pravděpodobně nebude konkurenceschopná, jelikož k tomuto ukončení může dojít i velmi daleko od minima účelové funkce. V práci byl trénink sítě automaticky ukončován potřeboval-li více než třicet úprav vektorů vah (epochs > 30), časové kritérium nebylo zavedeno (max_time = ∞).

Do druhé skupiny lze zařadit body 5 a 6. Jedná se o sofistikovanější metody předčasného ukončení tréninku umělé neuronové sítě.

Omezení hodnoty adaptivního parametru μ souvisí přímo s L.-M. optimalizačním algoritmem. L.-M. algoritmus je kombinací Gauss-Newtonovy a gradientní metody hledání minima nelineární funkce, přičemž právě hodnota parametru μ rozhoduje o poměru příspěvků jednotlivých metod ke konečné úpravě iterované hodnoty. Je-li hodnota $\mu = 0$, L.-M. algoritmus přechází na Gauss-Newtonovu metodu. Naopak, s rostoucím μ se algoritmus mění v gradientní metodu. Bez dalšího rozpisování kompletního matematického základu L.-M. metody, lze tuto metodu zapsat:

$$\Delta \mathbf{x}_k = - [J^T(\mathbf{x}_k) J(\mathbf{x}_k) + \mu I]^{-1} J^T(\mathbf{x}_k) f(\mathbf{x}_k), \quad (10)$$

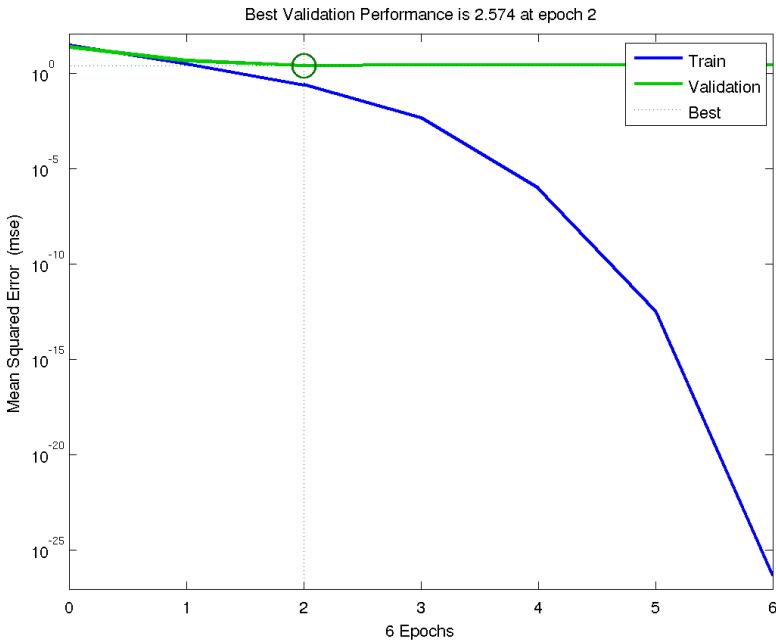
kde $\Delta \mathbf{x}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$ je změna optimalizovaného parametru (zde hodnoty vah na vstupu neuronů, viz. Obrázek 3), $J(\mathbf{x}_k)$ je Jacobijho matice minimalizované funkce, μ je adaptivní parametr, I je jednotková matice a $f(\mathbf{x}_k)$ jsou v našem případě kvadraty odchylek v jednotlivých měřených bodech.

Pro případ velmi vysokých hodnot μ ($\|\mu I\| \gg \|J^T(\mathbf{x}_k) J(\mathbf{x}_k)\|$) lze rovnici (10) přepsat do tvaru

$$\Delta \mathbf{x}_k = - \frac{1}{\mu} J^T(\mathbf{x}_k) (f(\mathbf{x}_k)), \quad (11)$$

a úprava optimalizovaného parametru začne být nepřímo úměrná velikosti parametru μ .

V průběhu vnitřního cyklu L.-M. metody je parametr μ zvyšován o určenou diskrétní hodnotu (krok) až do doby, než dojde ke snížení hodnoty účelové funkce. Poté je hodnota μ o jeden krok snížena a celý cyklus se opakuje v novém bodě. Parametr μ tedy v průběhu celého výpočtu postupně narůstá a je nutné zabránit situaci, kdy bude neúčelně omezovat rychlosť optimalizace vah sítě. Z tohoto důvodu se zavádí hodnota mu_max = 10^{10} a v okamžiku, kdy je $\mu > \text{mu_max}$ je trénink zastaven.



Obrázek 4 – Pokles hodnoty účelové funkce na trénovací a validační sadě dat v průběhu učení sítě.

Systém validačních zkoušek zabraňuje přeúčení sítě. Zajišťuje totiž, aby schopnost sítě popsat tréninková data nerostla na úkor její schopnosti zobecňovat.

V průběhu učení sítě totiž neustále klesá hodnota účelové funkce na tréninkových datech. Jak je ale vidět na Obrázku 4, hodnota účelové funkce na datech validačních klesá mnohem pomaleji a v průběhu učení může dokonce i stoupat.

Validační zkouškou rozumíme porovnání výsledků simulace na validačních datech mezi dvěma úpravami vektorů vah. V případě, že dojde ke zvýšení průměru kvadrátů odchylek je uložen záZNAM o tomto zhoršení výkonu sítě na validačních datech. Překročí-li počet neúspěšných validačních zkoušek nejvyšší přípustnou hodnotu (max_fail, v původně nastavenou na 5) je proces učení sítě ukončen a jako naučená síť je vzata ta s nejlepší hodnotou účelové funkce na validační sadě, což je dobře vidět na Obrázku 4, kde proběhlo šest epoch optimalizace vah, ale nakonec byl jako naučená síť uložen stav již po druhém optimalizačním cyklu.

Ukončení trénování sítě z důvodů patřících do druhé skupiny nemá na konkurenčeschopnost naučené sítě natolik negativní vliv jako v případě první skupiny, což je vidět i souborech vybraných sítí, kde se zvláště sítě s tréninkem ukončeným z důvodu překročení maximálního počtu neúspěšných validačních zkoušek objevují velmi často.

Do poslední, třetí skupiny patří body 3 a 4, tedy případy ukončení učení sítě z důvodu dosažení cílové hodnoty účelové funkce nebo pokles hodnoty gradientu pod nejmenší přípustnou hodnotu.

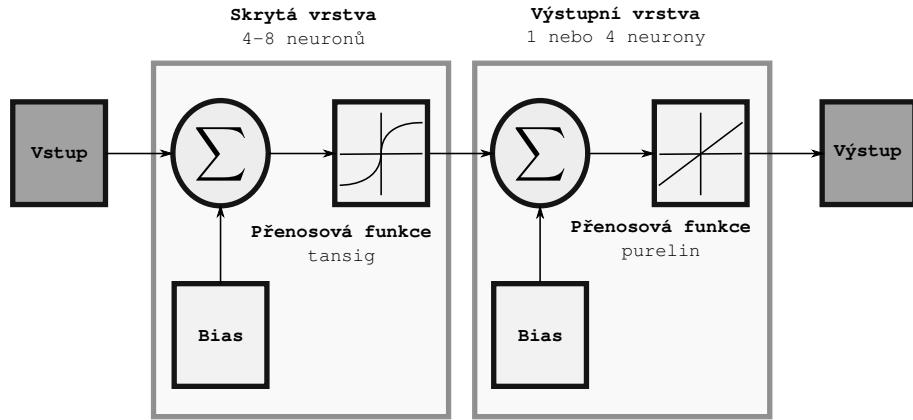
Je zřejmé, že ukončení učicího procesu sítě z důvodu dosažení cílové hodnoty účelové funkce na trénovací sadě dat je nevhodnější možný výstup optimalizačního algoritmu, avšak hodnota cílová hodnota účelové funkce (goal) je nastavena na 0. Této hodnoty může být dosaženo pouze v případě naprosté shody mezi experimentálně zjištěnými a simulovanými výtěžky pyrolýzy na trénovací datech, což je při zachování podmínky na souběžný pokles hodnoty účelové funkce na validační sadě nemožné.

Nejvhodnějším dosažitelným důvodem ukončení tréninku sítě je tedy dosažení minimální přípustné hodnoty gradientu účelové funkce, což značí přiblížení se bud' k místu, kde má účelová funkce průběh velmi podobný konstantnímu nebo nalezení lokálního minima. Jelikož je ale účelová funkce definována jako průměr čtverců odchylek mezi naměřenými a dopočtenými hodnotami je pravděpodobné, že chybouva plocha bude mít tvar vycházející z paraboloidu a nalezení hodnoty gradientu blízké nule bude znamenat přiblížení se globálnímu minimu.

Porovnáním terminačních mechanismů učení sítí mezi vybranými sítěmi bylo zjištěno, že za zde prezentovaného nastavení parametrů je nejčastějším důvodem ukončení tréninku sítě překročení maximálního přípustného počtu neúspěšných validací.

4.2 Architektura modelů

Všechny v této práci prezentované modely jsou založeny na využití ANN, se kterými bylo pracováno ve výpočetním prostředí MATLABTM, s využitím nástroje NEURAL NETWORKS TOOLBOX. Na základě doporučení v dokumentaci tohoto nástroje[14] bylo pracováno se sítěmi s jednou skrytou vrstvou neuronů a s proměnlivým počtem neuronů v této vrstvě. Jako transferová funkce ve skryté vrstvě byla použita tangenciální sigmoida. Ve výstupní vrstvě pak pak byla použita funkce lineární. Obecné schéma takovéto sítě je na Obrázku 5.



Obrázek 5 – Obecné schéma využitých sítí

Optimální architektura umělých neuronových sítí sítí pro jednosíťový i čtyřsíťový model a jejich vstupní parametry byly hledány pomocí osmi skriptů využívajících různá doplňková data. Jejich přehled je v Tabulce 4.

Tabulka 4 – Přehled skriptů

Jméno skriptu jednosíťového modelu	Jméno odpovídajícího skriptu čtyřsíťového modelu	Využitá doplňková data
netMdl_1	modulMdl_1	M, d, T10, T50, T90
netMdl_Md	modulMdl_Md	M, d
netMdl_MT	modulMdl_MT	M, T10, T50, T90
netMdl_dT	modulMdl_dT	d, T10, T50, T90
netMdl_M	modulMdl_M	M
netMdl_d	modulMdl_d	d
netMdl_T	modulMdl_T	T10, T50, T90
netMdl_nic	modulMdl_nic	-

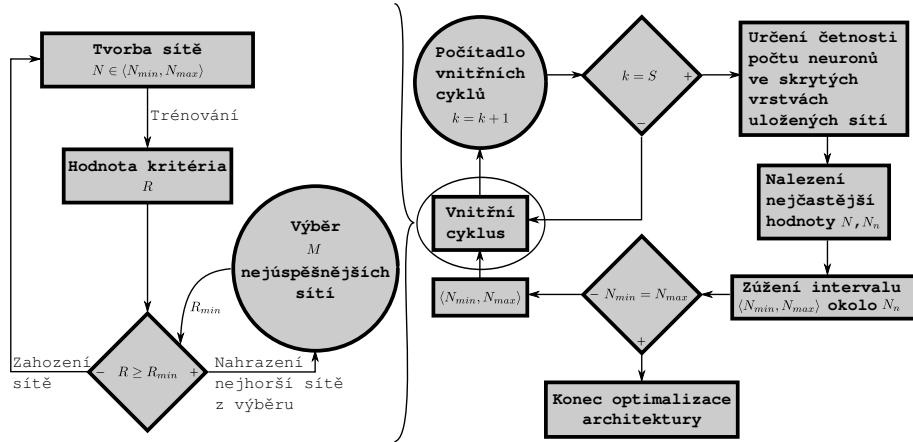
Na základě sítí užitých v [3,10,15 a 16] byl předpokládán ideální počet neuronů v jednotlivých sítích 4-6.

4.2.1 Algoritmus optimalizace architektury

Hledání optimální architektury sítě a určování vhodných doplňkových parametrů bylo prováděno ve dvou téměř oddělených fázích. V první byl pomocí níže představeného algoritmu nalezen vhodný počet neuronů ve skryté vrstvě sítě pro všechny kombinace typů modelu a vstupních parametrů (dohromady šestnáct optimalizací). Ve druhé byly porovnávány výsledky jednotlivých variant a bylo usuzováno na vhodný typ modelu a počet doplňkových charakteristik suroviny.

Optimalizace architektury sítě byla prováděna pomocí genetického algoritmu jehož schéma je na Obrázku 6. Zdrojový kód varianty algoritmu pro jednosíťový model je uveden v Příloze A. Varianta pro model čtyřsíťový se liší pouze odstraněním výpočtu váženého průměru korelačních koeficientů a není tedy nutné ji uvádět samostatně.

Celý algoritmus lze rozdělit na dvě části, vnitřní a vnější cyklus. Vnitřní cyklus obstarává výběr, dle zavedeného hodnotícího kritéria, nejúspěšnějších sítí a vnější cyklus pak na základě



Obrázek 6 – Schéma algoritmu optimalizace architektury. M je počet ukládaných sítí, „genofond“ algoritmu, N je počet neuronů ve skryté vrstvě, N_n je nejčastější hodnota počtu neuronů ve skryté vrstvě mezi vybranými sítěmi, R je hodnota kritéria úspěšnosti sítě a S je uměle definovaná hodnota úměrná počtu uložených sítí, $S \approx 10^4$.

tohoto výběru postupně upravuje šíři intervalu do kterého náleží počty navrhovaných neuronů ve skryté vrstvě sítě.

Ve vnitřním cyklu jsou zkoušeny různé sítě s náhodným počtem neuronů ve skryté vrstvě takovým, že $N \in \langle N_{min}, N_{max} \rangle$, kde N je počet neuronů ve skryté vrstvě a $\langle N_{min}, N_{max} \rangle$ je interval, ze kterého je tento počet volen. Následně je síť učena na dostupných datech trénovací sady (viz. Tabulka 3). Po naučení sítě je vyzkoušena její schopnost predikovat výtěžky pyrolýzy, je určen korelační koeficient R mezi experimentálními a simulovanými daty pro validační sadu⁴. Tento korelační koeficient je pak porovnán s nejmenším korelačním koeficientem mezi vybranými sítěmi R_{min} a v případě, že je větší, je nejhorší síť z výběru nahrazena aktuální sítí.

Úkolem vnějšího cyklu je postupně zužovat interval $\langle N_{min}, N_{max} \rangle$ až k hodnotě optimálního počtu neuronů ve skryté vrstvě pro danou kombinaci typu modelu a doplňkových vstupních parametrů suroviny. V případě, že je vyzkoušen dostatečný počet sítí, S (uměle zavedená hodnota, $S \approx 10^4$), je provedeno vyhodnocení četnosti počtu neuronů ve skryté vrstvě mezi uloženými sítěmi (v závislosti na typu modelu je uloženo 100 či 200 sítí). Následně je nalezena nejčastější hodnota, N_n a interval $\langle N_{min}, N_{max} \rangle$ je zúžen okolo této hodnoty. Algoritmus je ukončen v případě, že $N_{min} = N_{max} = N_{optimal}$. Po ukončení algoritmu jsou vybrané nejúspěšnější sítě společně s parametry jejich učení uloženy pro další zpracování.

Analogie s evolučním vývojem je zřejmá, vnitřní cyklus zkouší schopnosti jedince (neuronové sítě) přizpůsobit se prostředí (přesnost predikce výtěžku produktů pyrolýzy) a uchovává vlastnosti (počet neuronů ve skryté vrstvě) nejúspěšnějších. Vnější cyklus hledá takovou hodnotu sledované vlastnosti, která je mezi nejúspěšnějšími jedinci nejrozšířenější a upraví všechny dále zkoušené jedince tak, aby vzrostla pravděpodobnost, že budou mít právě tuto hodnotu vlastnosti (zúžení šířky intervalu $\langle N_{min}, N_{max} \rangle$).

4.2.2 Jednosítový model

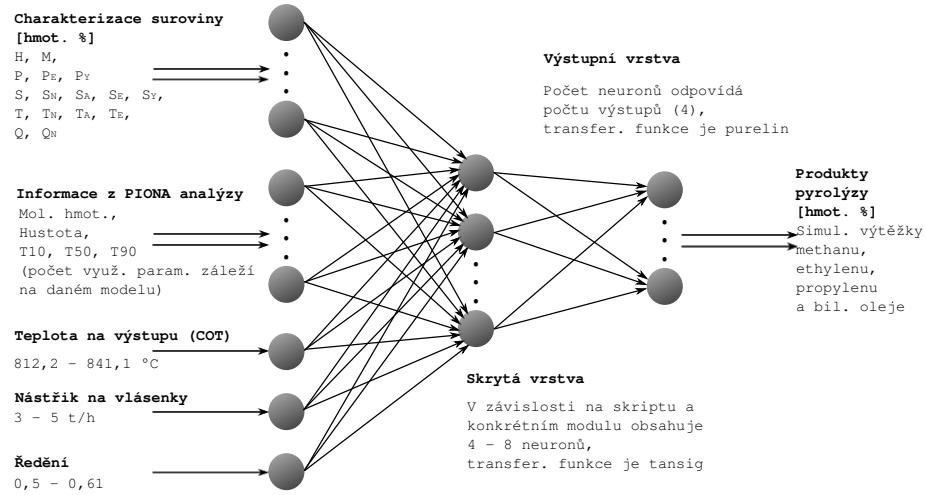
Jednodušším z přístupů k problematice je hledání takové sítě, která by byla schopná predikovat výtěžky všech sledovaných produktů. Nákres modelu sestávajícího z jedné sítě je na Obrázku 7.

Vstupní parametry jsou upravovány v závislosti na skriptu obsatarávajícím trénink sítě. Informace o fragmentovém složení[7,8,10] a o provozu (COT, nástřik na vlásenky a ředění) jsou využity vždy a v závislosti na optimalizačním skriptu jsou doplněny o údaje o molekulární hmotnosti, hustotě či destilační křivce suroviny[4].

Úspěšnost vyzkoušených sítí byla hodnocena na základě váženého průměru korelačních koeficientů mezi cílovými a dopočtenými výstupními daty pro jednotlivé suroviny. Tyto koeficienty byly počítány na validačních sadách a vzhledem k průmyslové důležitosti těchto produktů byla nejvyšší

⁴V případě jednosítového modelu je tento koeficient počítán jako vážený průměr koeficientů pro všechny produkty. Nejvyšší váha je přiřazena schopnosti přesné predikce výtěžků pyrolýzy ethylenu a propylenu. U čtyřsítového modelu tento dodatečný výpočet odpadá, jelikož je měřena schopnost sítě predikovat pouze jeden ze sledovaných produktů.

váha přiřazena koeficientům ethylenu a propylenu (úspěšnost sítě při predikci těchto produktů měla dvojnásobnou váhu oproti schopnosti predikovat výtěžky methanu a bilančního oleje).



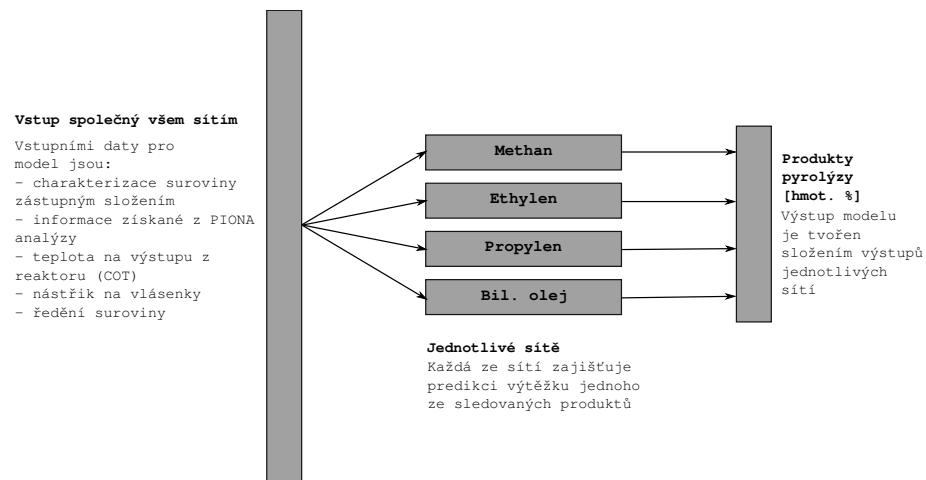
Obrázek 7 – Schéma neuronové sítě predikující výtěžky všech sledovaných produktů

Průběžně bylo ukládáno vždy 200 nejúspěšnějších sítí jako dostatečně široký soubor dat pro vytvoření histogramu počtu neuronů ve skrytých vrstvách a určení nejčastější hodnoty počtu neuronů. Vhodný rozměr skryté vrstvy byl určován na základě nejčastější hodnoty mezi vybranými sítěmi.

Počáteční odhad intervalu, do kterého náleží ideální počet neuronů byl zvolen $\langle 4,8 \rangle$. Tento interval byl pak na základě průběžných výhodnocení zužován tak, aby výpočet konvergoval k ideálnímu počtu neuronů ve skryté vrstvě (nejčastější rozměr skryté vrstvy mezi vybranými sítěmi). Algoritmus výběru vhodných sítí a postupného zužování intervalu je popsán v 4.2.1.

4.2.3 Čtyřsíťový model

Jelikož se v průběhu práce ukázalo, že jedosíťový přístup je pro pokrytí dané problematiky příliš jednoduchý a nedosahuje adekvátních výsledků, bylo rozhodnuto sestrojit o něco komplexnější model sestávající z modulů tvořených různými neuronovými sítěmi predikujícími výtěžek vždy jen jednoho sledovaného produktu. Schéma takového modelu pro sledování čtyř hlavních produktů je na Obrázku 8.

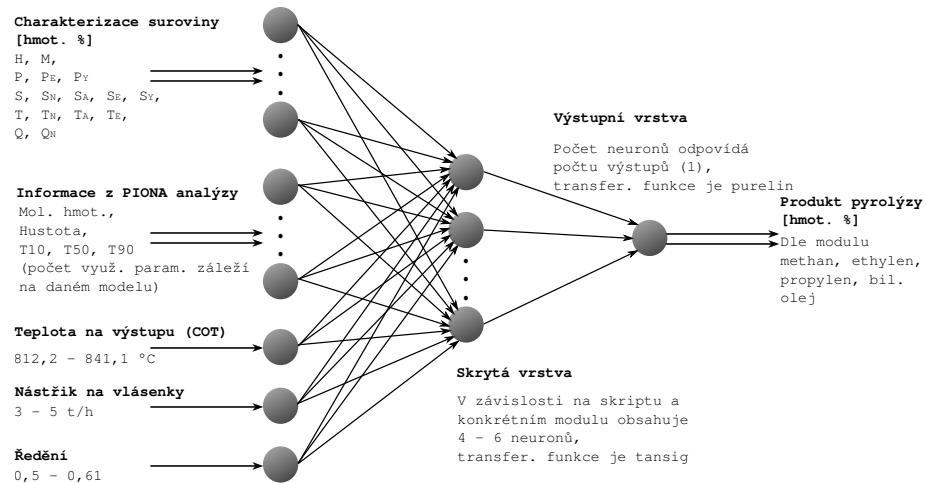


Obrázek 8 – Schéma čtyřsíťového modelu

Sítě tvořící jednotlivé moduly takového modelu pak vypadaly prakticky identicky jako síť tvořící jednosíťový model, pouze s tím rozdílem, že ve výstupní vrstvě byl jedený neuron a výstupem byly odhady výtěžku jediného produktu (viz. Obrázek 9).

Tento přístup sice poskytoval lepší výsledky, na druhou stranu však velmi vzrostl výpočetní čas nutný k optimalizaci modelu. Z tohoto důvodu byl počet průběžně ukládaných sítí snížen na 100 a úměrně byl redukován i počet sítí vyzkoušených před každým statistickým vyhodnocením. Jelikož ale bylo třeba nalézt vhodný počet neuronů ve skryté vrstvě pro čtyři místo jedné sítě, optimalizace architektury zde trvala přibližně trojnásobek času.

Optimalizační algoritmus byl velmi podobný algoritmu využitému u jednosíťového modelu ovšem nebylo zde nutno volit mezi sledovanými produkty ty důležitější. Po optimalizaci architektury jednotlivých sítí byly vybrány a zkombinovány do výsledného modelu nejúspěšnější sítě.



Obrázek 9 – Schéma sítí, které tvoří jednotlivé moduly čtyřsíťového modelu

5 Výsledky a diskuze

V následujícím textu budou prezentovány a diskutovány výsledky optimalizace architektury, rozdíly ve kvalitě mezi jednosítovým a čtyřsítovým modelem, vliv dodatečných parametrů na jednotlivé modely a změna přesnosti predikce výtěžků pyrolýzy při doplnění hlavní tréninkové sady založené na jednom typu suroviny (zde primární benzíny, PrBi – 29 datových sad) o menší soubor údajů popisujících pyrolýzu suroviny typu jiného (atmosférický plynový olej, APO – 10 datových sad).

Pozn: Pro zachování přehlednosti je část grafů a tabulek umístěna mimo text, do Příloh. Zde jsou pouze ukázky.

5.1 Optimalizace architektury modelů

Získané optimalizované architektury jsou obecně v dobré shodě s předpokládanými výsledky (viz. 4.2). Pro většinu sítí byl navržen počet neuronů ve skryté vrstvě mezi 4 a 6. Nejčastější navrhovaná hodnota byla 5. Stručný přehled skriptů s výsledky optimalizace sítí je uveden v Tabulce 5.

Tabulka 5 – Přehled výsledků optimalizace architektury sítí

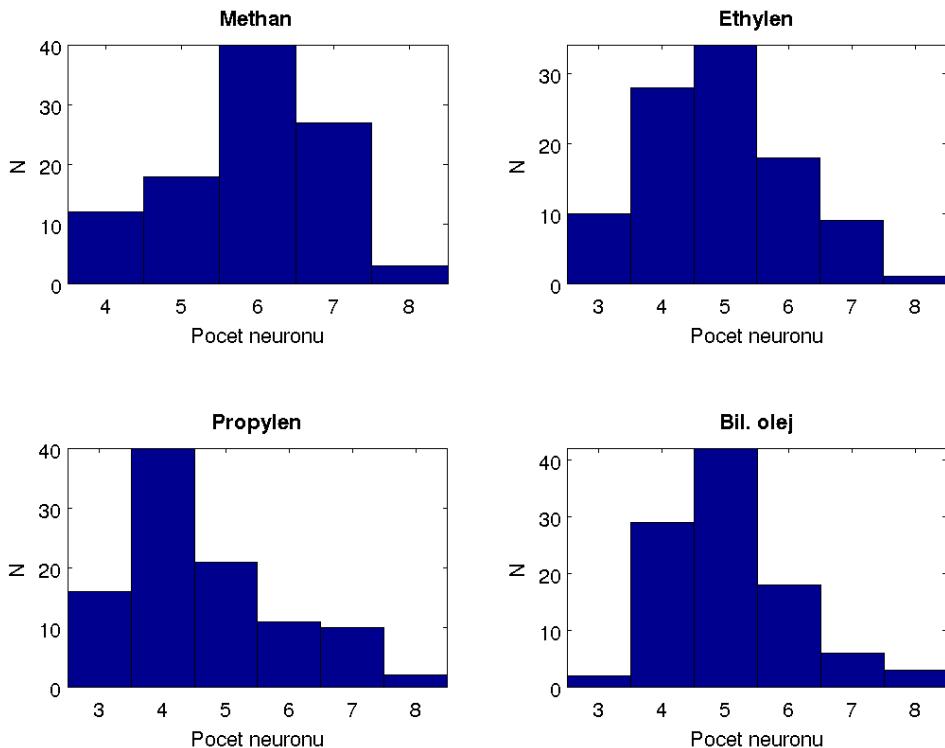
Jméno modelu	Počet sítí	Počet neuronů ve skryté vrstvě sítí ⁵	
		PrBi	PrBi + APO
netMdl_1	1	6	5
modulMdl_1	4	6, 5, 4, 5	6, 5, 6, 5
netMdl_Md	1	5	7
modulMdl_Md	4	6, 4, 5, 6	7, 4, 6, 5
netMdl_MT	1	6	5
modulMdl_MT	4	6, 5, 4, 5	7, 4, 6, 4
netMdl_dT	1	5	6
modulMdl_dT	4	6, 4, 5, 6	6, 4, 6, 4
netMdl_M	1	6	5
modulMdl_M	4	6, 5, 4, 5	7, 5, 5, 4
netMdl_d	1	5	6
modulMdl_d	4	6, 4, 5, 6	5, 5, 5, 4
netMdl_T	1	6	4
modulMdl_T	4	6, 5, 4, 5	6, 5, 5, 5
netMdl_nic	1	5	6
modulMdl_nic	4	6, 4, 5, 6	6, 5, 5, 4

Je zvláštní, že pro jedinou síť jednosítového modelu je navrhován obdobný počet neuronů jako pro jednotlivé sítě modelu čtyřsítového. Očekávaný by zde totiž byl spíše vyšší počet neuronů reflektující vyšší počet výstupních parametrů a tudíž i vyšší nároky na schopnosti sítě. Jelikož je ale zachován stejný počet neuronů ve skryté vrstvě, jednosítový model poskytuje méně přesné výsledky než model čtyřsítový.

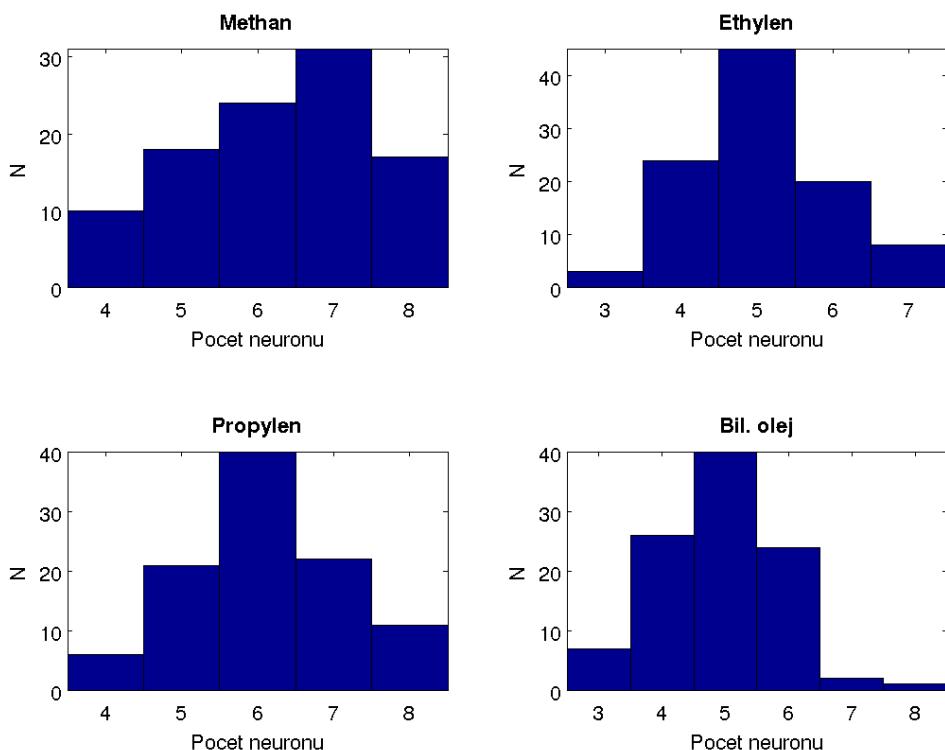
Toto chování se nezměnilo, ani když bylo při trénování sítě zabráněno ukončení kvůli překročení maximálního počtu neúspěšných validačních zkoušek či vypršení časového limitu. Zdá se, že plně propojená síť není k predikci výtěžku většího množství produktů vhodná.

Čtyřsítový model lze interpretovat i jako jedinou síť, která má ve skryté vrstvě počet neuronů odpovídající sumě neuronů ve skrytých vrstvách sítí kterými je tvořen, avšak skrytá a výstupní vrstva nejsou plně propojeny. Jelikož čtyřsítový (nepropojený) model poskytuje lepší výsledky, než-li model jednosítový (plně propojený) a zároveň je nasnadě předpokládat, že je mezi výtěžky

⁵Je-li použito více sítí, počty neuronů řazený v pořadí: síť pro predikci výtěžku methanu, síť pro predikci výtěžku ethylenu, síť pro predikci výtěžku propylenu a síť pro predikci výtěžku bilančního oleje.



Obrázek 10 – Porovnání četnosti počtu neuronů ve skryté vrstvě algoritmem vybraných sítí pro čtyřsíťový model ModulMDL_1 navržený pro predikci pouze primárních výtěžků pyrolyzy benzínů.



Obrázek 11 – Porovnání četnosti počtu neuronů ve skryté vrstvě algoritmem vybraných sítí pro čtyřsíťový model ModulMDL_1 navržený pro predikci primárních výtěžků pyrolyzy benzínů i plynových olejů.

jednotlivých produktů nějaká souvislost, bylo by nadále vhodné experimentovat i s různou mírou propojenosti skryté a výstupní vrstvy. Toto je však spojeno s enormním náruštěm počtu zkoušených sítí nutných pro optimalizaci architektury sítě a případné zlepšení vlastností výsledného modelu by velmi pravděpodobně nebylo úměrné vynaloženému výpočetnímu času.

Dále je zajímavé pozorovat změny navrženého počtu neuronů se změnou rozsahu tréninkových dat. Jak je vidět v Tabulce 5, na Obrázcích 10 a 11 na předchozí straně či na grafech v příloze C, v případě čtyřsítového modelu počet neuronů ve skryté vrstvě s rozšířením datové sady většinou roste u sítí predikujících výtěžky methanu, ethylenu a propanu. U sítě určené k odhadu výtěžku bilančního oleje tento počet neuronů naopak vždy klesá či zůstává stejný.

To je pravděpodobně zapříčiněno buď řádovým rozdílem mezi podily bilančního oleje při pyrolyze APO a PrBi, který v kombinaci s některou z dodatečných charakteristik suroviny zjednoduší nalezení vztahu mezi vstupními a výstupními daty, nebo poměrně velkým rozptylem (chybou) měřených dat pro olej. Kvůli větší chybě sítě nemůže dobře zobecňovat a MSE na validační sadě začne velmi brzy růst. Odlišnosti mezi výtěžky zbylých produktů jsou méně významné, přesto však nezanedbatelné, čímž náročnost jejich rozlišení zřejmě naopak narůstá.

5.2 Porovnání čtyřsítového a jednosítového modelu

Prvotní snahou bylo sestrojit co nejjednodušší model poskytující dostatečně přesné výsledky s co největší spolehlivostí. Ukázalo se však, že jednoduší jednosítový model nedokáže tato kritéria splnit. Byl tedy navržen komplikovanější model sestávající ze čtyř různých sítí specializovaných na predikci výtěžku vždy jen jednoho ze sledovaných produktů.

Největší nevýhodou čtyřsítového modelu je výrazný nárůst výpočetního času potřebného k optimalizaci jeho architektury. Vhodné množství neuronů ve skryté vrstvě je totiž nutné hledat čtyřikrát namísto jednou.

I když u specializované sítě je její učení méně časově náročné, než u sítě obstarávající všechny produkty, doba nutná k optimalizaci architektury modelu vzrostla téměř na trojnásobek⁶. Tato činnost je však jednorázová a na další využití modelu již nemá vliv. Čas potřebný k učení jednotlivých modelů a doba trvání simulace se pohybuje v řádu setin až desetin vteřin.

Přesto je ale nutné mít na paměti, že úspěšnost naučení sítě je vždy do jisté míry dílem náhody a tudíž je nutné při každé adaptaci modelu na novou datovou sadu proces učení vícekrát opakovat. Zde se již vyšší výpočetní náročnost čtyřsítového modelu projeví citelněji.

Jako kritérium přesnosti odhadu byl zvolen aritmetický průměr absolutních⁷ a normovaných⁸ odchylek mezi naměřenou a dopočtenou hodnotou výtěžků jednotlivých produktů na validační sadě dat. K vypočtení tohoto průměru byly použity vždy všechny uložené sítě pro daný model (viz. 4.2.2 a 4.2.3). Čím byla hodnota průměru odchylek nižší, tím byl model považován za úspěšnější.

Na spolehlivost modelů pak bylo usuzováno z šířky mezikvartilového rozpětí těchto odchylek a jejich rozptylů. Na Obrázku 12 jsou krabicové diagramy zobrazující mezikvartilové rozpětí a průměr absolutních odchylek pro modely využívající všech dostupných dodatečných charakteristik suroviny. Tabulky s přehledem průměrných odchylek a jejich rozptylů pro všechny modely jsou pak v Příloze D.

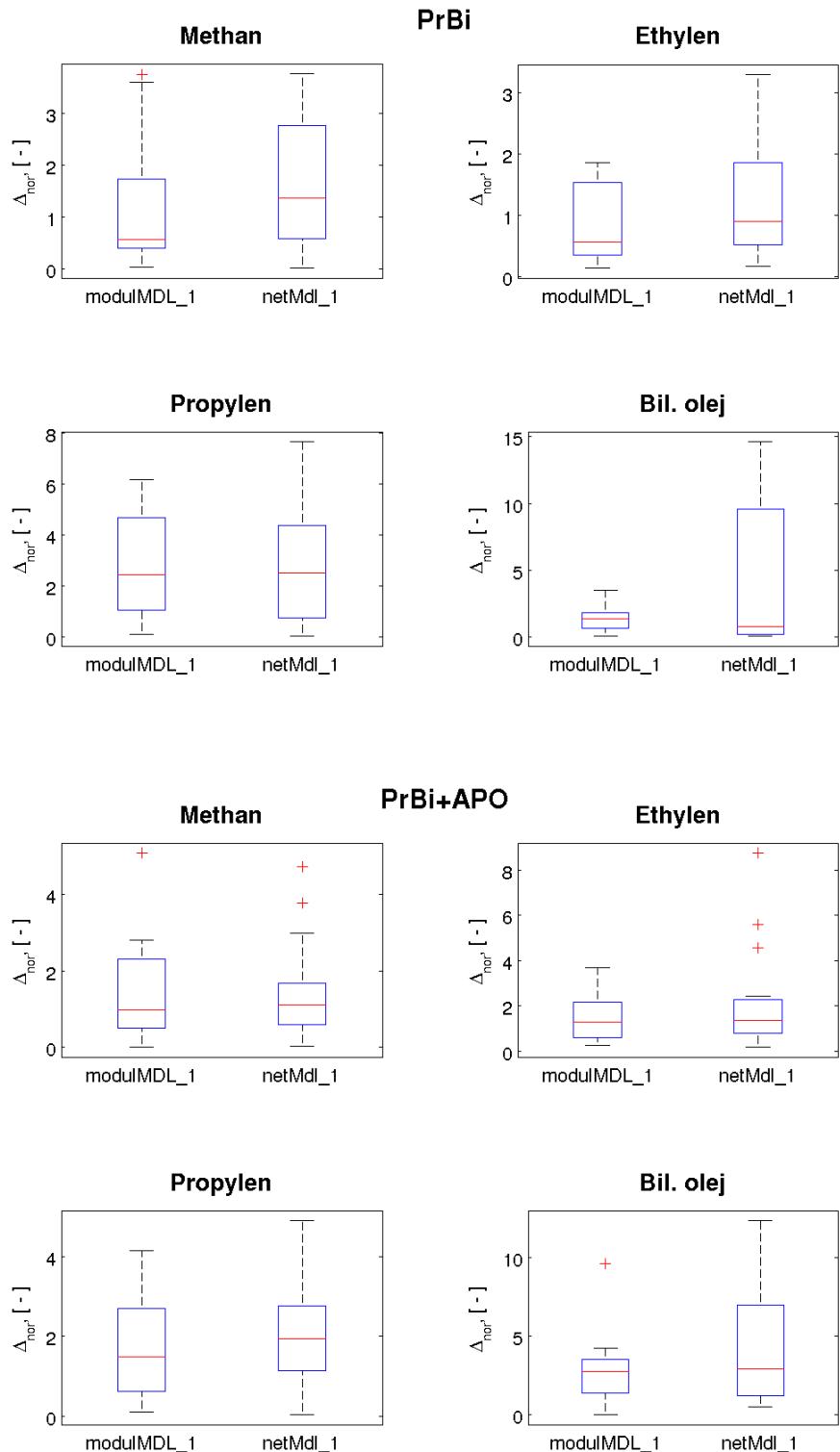
Jak je vidět převážně na rozptylu normovaných odchylek, který je vzhledem k malým rozdílům v absolutních odchylkách citlivější mírou, než rozptyl odchylek absolutních, čtyřsítový model je znatelně spolehlivější než model jednosítový. Rozdíl v přesnosti modelu pak již není, zvláště v případě predikce výtěžku propylenu tak dramatický. Čtyřsítový model dosahuje lepších výsledků, zejména při rozšíření datové sady o jiný typ suroviny.

Další výhodou čtyřsítového modelu je, že není nastaven tak, aby upřednostňoval přesnost pouze u některých typů produktů (viz. 4.2.2). Ačkoliv tedy nejsou rozdíly ve schopnosti modelů předpovídat výtěžky ethylenu a propylenu tak markantní, u ostatních produktů dosahuje jednosítový model horších výsledků. To je případ zejména bilančního oleje, u nějž je predikce výtěžku obecně nejméně spolehlivá.

⁶Na počítací s procesorem Intel Core 2 Duo o frekvenci 1,8 GHz trvalo vyhodnocení jednoho z osmi skriptů něco přes deset hodin namísto necelých čtyř u jednosítového modelu. Výsledný časový poměr byl cca 85:30 hodinám ve prospěch jednodušího modelu.

⁷Absolutní ochylky byly dopočítávány dle vzorce $\Delta_{abs} = |vyp.hodnota - exp.hodnota|$.

⁸Normované odchylky byly dopočítávány dle vzorce $\Delta_{nor} = \frac{|vyp.hodnota - exp.hodnota|}{\max(exp.hodnota) * 1.25} \cdot 100$.



Obrázek 12 – Krabicové diagramy porovnávající výběrové střední hodnoty a mezikvartilové rozpětí normovaných odchylek mezi simulovanými a naměřenými daty pro modely využívající všechna dostupná doplňková data.

5.3 Vliv dodatečných informací o surovině

Surovina byla na vstupu modelu vždy charakterizována svým fragmentovým složením a informacemi o provozu reaktoru. Tato data pak byla doplňována údaji o hustotě, molární hmotnosti a destilační křivce suroviny v různých kombinacích a byl porovnáván vliv těchto dodatečných parametrů na přesnost předpovědi výtěžku produktů.

Hlavním důvodem doplnění fragmentového složení a údajů o provozu o tyto dodatečné parametry je neschopnost fragmentového složení reflektovat velikost molekuly. Popisuje totiž pouze poměr uhlíkových skupin ze kterých se surovina skládá a ten může být shodný pro několik sloučenin významně se lišících velikostí a tedy majících různé vlastnosti pro pyrolýzu.

Bylo tedy předpokládáno, že vliv dodatečných informací bude v případě trénování a užití modelu pouze pro predikci pyrolýzy primárních benzínů velmi malý a s rozšířením simulovaných dat o plynové oleje vzroste.

Jak je ale vidět na grafech v Příloze E, vliv dodatečných charakteristik suroviny je u čtyřsíťového modelu zanedbatelný i při práci se dvěma typy suroviny. To je zřejmě zapříčiněno rozdílným fragmentovým složením u obou typů surovin. V Tabulce 6 jsou uvedeny průměrné hmotnostní zlomky fragmentů využitých ke zjednodušené charakterizaci suroviny.

Tabulka 6 – Průměrné rozdělení fragmentů pro Primární benzíny a APO

Surovina	Fragment [hmot. %]									
	P	S	S _N	T	T _N	Q	Q _N	S _E	S _A	T _A
Primární benzín	35.1	32.0	16.8	5.4	4.7	0.2	0.1	0.1	4.3	1.2
APO	9.5	43.9	29.2	1.4	6.3	0.1	0	0	8.0	1.6

Z Tabulky 6 je zřejmé, že k charakterizaci reálných komponent zastupujících složení primárních benzínů je v případě zpracovávaných dat nutná vyšší variabilita fragmentů. Navíc je mezi primárními benzíny a plynovými oleji znatelný rozdíl v zastoupení primárních a sekundárních „alifatických“ uhlovodíků a teké je mezi PrBi a APO výrazný rozdíl v hodnotách teplot na výstupu vlásenky (průměrně 798 °C u APO oproti 827°C u PrBi). To je pak zřejmě při trénování sítě již samo o sobě dostatečně výrazným ukazatelem na typ suroviny a významnost informací o molární hmotnosti, hustotě či destilační křivce suroviny klesá.

Je však velmi pravděpodobné, že při doplnění simulovaných dat o další typ suroviny nebo i při rozšíření stávajících datových sad (zvláště APO) by se tyto rozdíly ve fragmentovém složení smazaly a důležitost dodatečných charakteristik suroviny by vzrostla.

V případě jednosíťového modelu je vliv doplnění charakterizace suroviny o další údaje značnější a zejména sítě využívající kombinaci všech dostupných informací či údajů o molární hmotnosti suroviny dosahují nadprůměrných výsledků. Tento vývoj by mohl být ukazatelem na vhodné doplňkové parametry pro čtyřsíťový model v případě větší diversifikace zpracovávaných surovin.

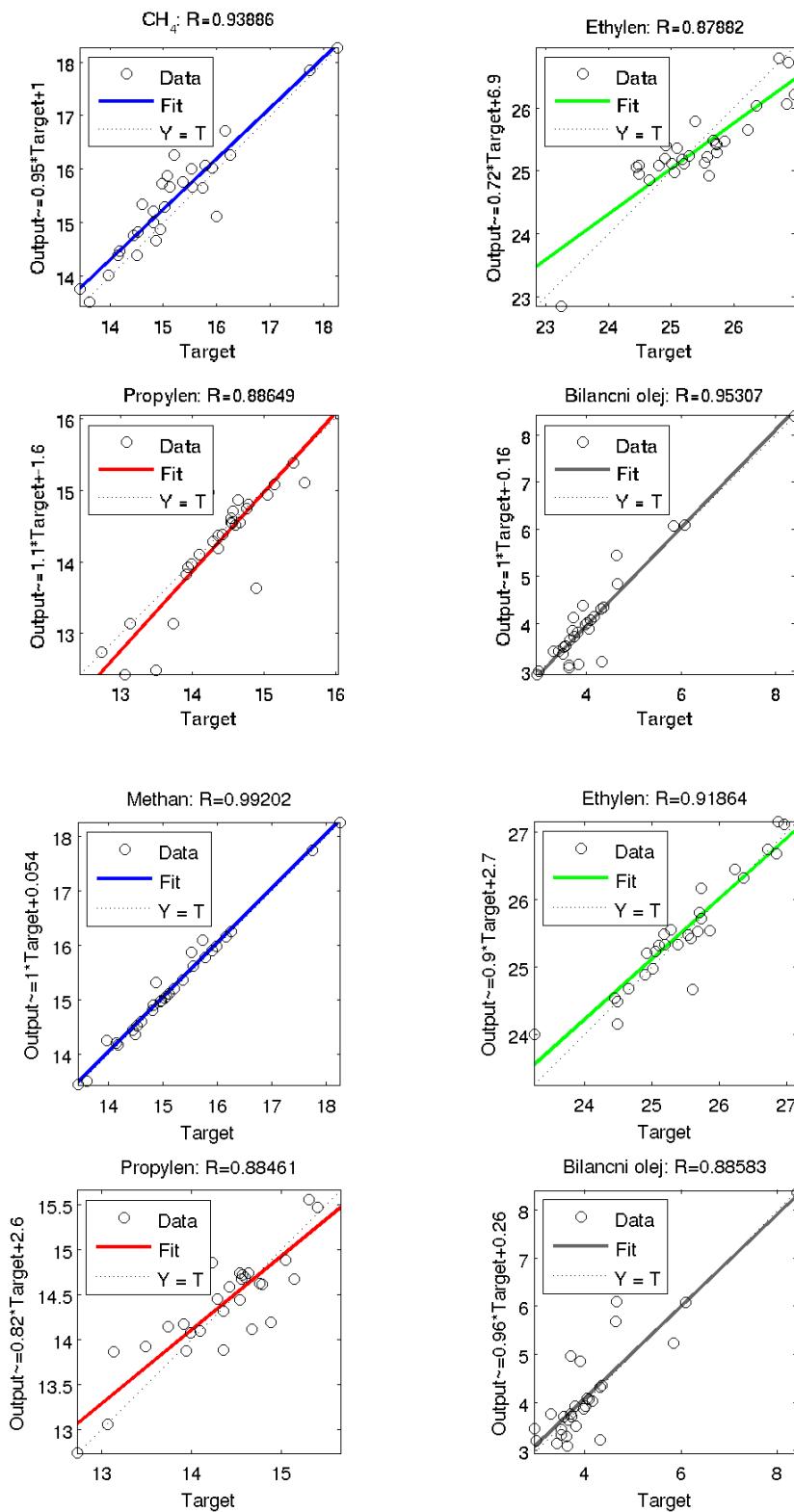
5.4 Změna přesnosti s rozšířením datové sady

Většinu dat dostupných pro modelování tvořily údaje o primárních benzínech. Úloha informací o pyrolýze plynových olejů byla spíše doplňková – poměr datových sad byl téměř 3:1 ve prospěch primárních benzínů. V práci tedy byl zkoumán vliv seskupení všech dostupných dat na přesnost modelu tvořeného neuronovými sítěmi a schopnost takového modelu predikovat výtěžky pyrolýzy suroviny pro kterou nebylo dostupné dostatečné množství dat k naučení specializovaného modelu.

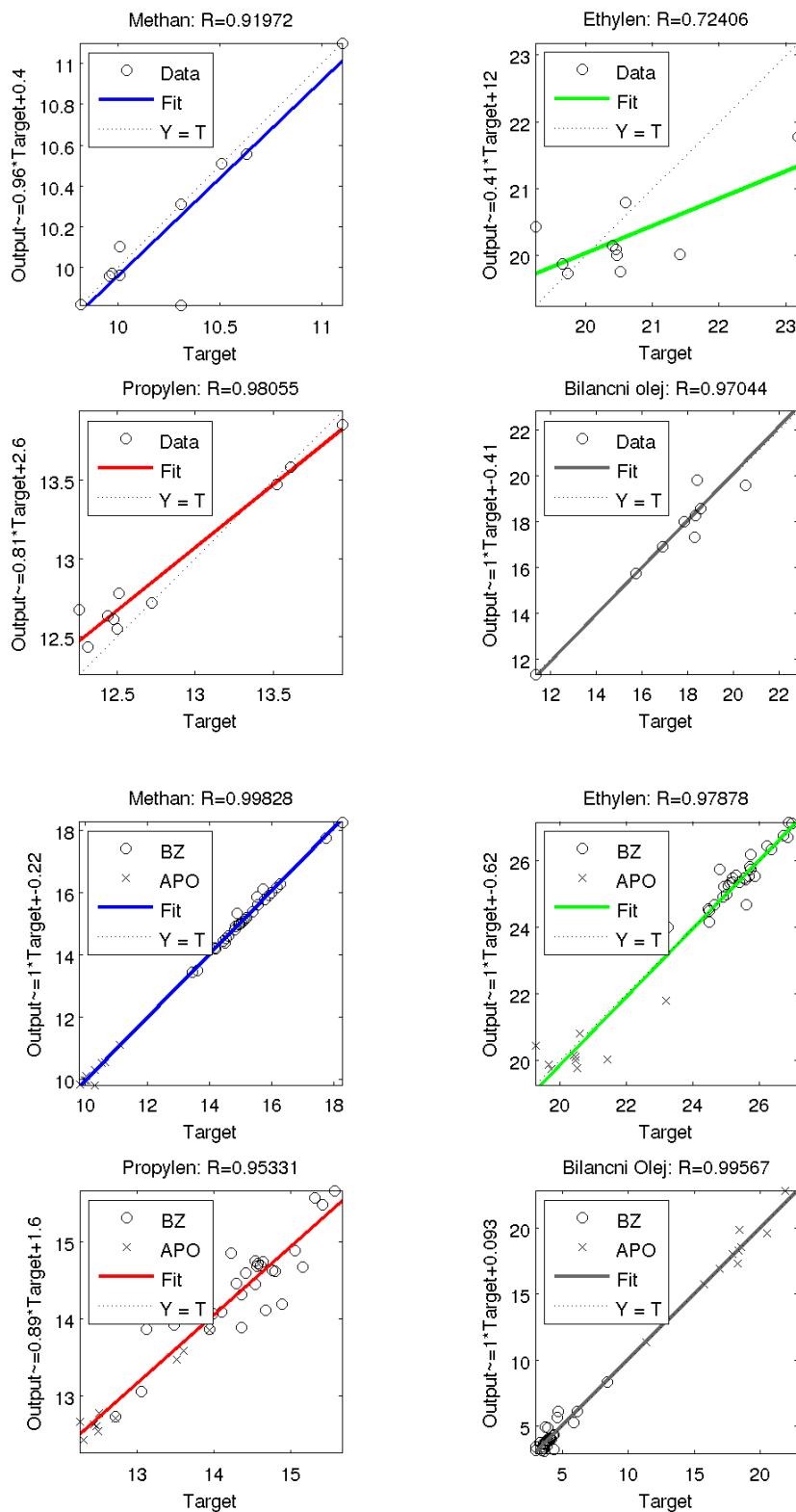
ANN jsou obecně velmi nedoporučované k extrapolaci dat [14,12], avšak ukázalo se, že i přes malý rozsah informací o plynových olejích je možné sítě trénované na souhrnu dat využít k predikci pyrolýzy této suroviny.

Na Obrázku 13 obsahujícím porovnání korelačních koeficientů mezi experimentálními a simulovanými daty při predikci výtěžků pyrolýzy primárních benzínů čtyřsíťovým modelem modulMDL_1 můžeme vidět, že při rozšíření datové základny o jiný typ suroviny nedojde k žádné signifikantní ztrátě přesnosti odhadu. Naopak, z krabicových diagramů v Příloze E lze usoudit, že se velmi často přesnost a spolehlivost modelu při předpovědi výtěžku pyrolýzy původní suroviny zlepší.

Na následujícím Obrázku 14 jsou proti sobě zobrazeny predikované a naměřené hodnoty výtěžků APO a všech dat dohromady a korelační koeficienty mezi nimi. Při simulaci byl využit čtyřsíťový model modulMDL_1. Z grafů je zřejmé, že model nedosahuje zvláště při predikci výtěžků ethylenu



Obrázek 13 – Regresní grafy, zobrazena jsou simulované hodnoty v závislosti na experimentálních pro jednotlivé produkty a hodnoty korelačních koeficientů mezi nimi. Zobrazeny jsou všechny datové sady pro PrBi. Horní čtyři grafy reprezentují výsledky čtyřsíťového modelu modulMDL_1 trénovaného pouze na PrBi, spodní čtyři pak výsledky toho samého modelu trénovaného na všech dostupných datech. Hodnoty na osách odpovídají hmot. % výtěžku daného produktu.

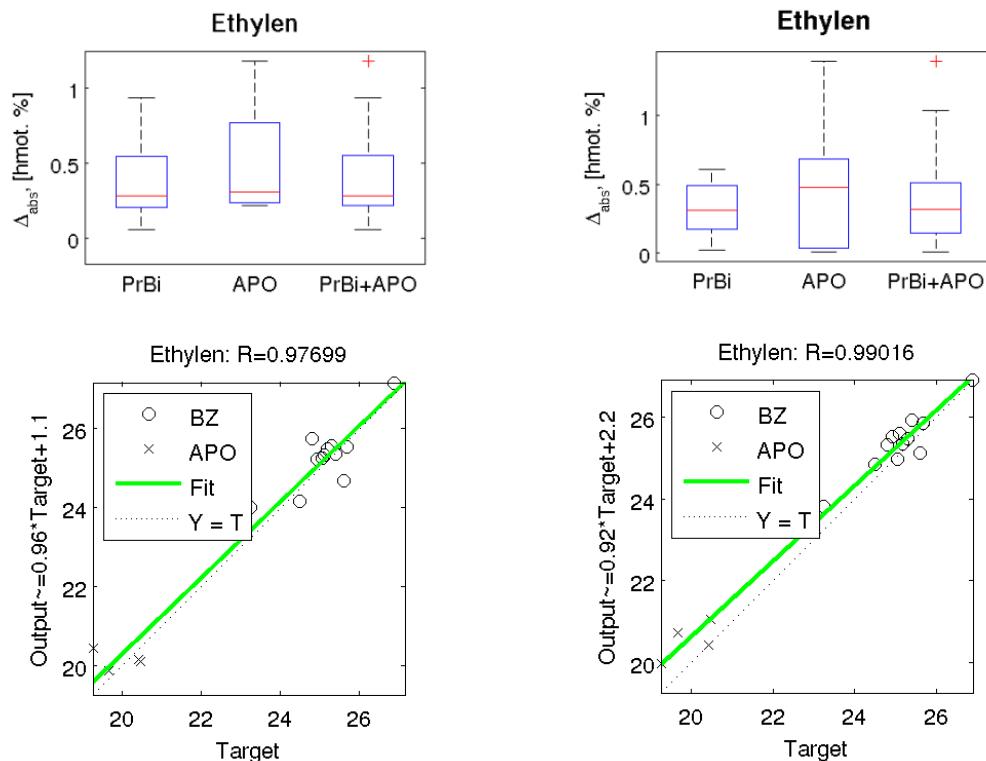


Obrázek 14 – Regresní grafy, zobrazena jsou simulované hodnoty v závislosti na experimentálních pro jednotlivé produkty a hodnoty korelačních koeficientů mezi nimi. Zobrazeny jsou výsledky čtyřsíťového modelu modulMDL_1 trénovaného na všech dostupných datech. Na horní čtyři grafy ukazují schopnost modelu predikovat vítězky APO, spodní potom všech dostupných dat. Hodnoty na osách odpovídají hmot. % výtěžku daného produktu.

bezchybných výsledků, ale jak lze vidět na tabulkách v Příloze F, absolutní odchyly odchodu výtěžku ve většině případů nepřekročí jedno hmotnostní procento. Nevezmeme-li v potaz bilanční oleje, u nichž jsou nepřesnosti u všech využitých modelů největší, zůstane většina absolutních odchylek menších než půl hmotnostního procenta.

5.5 Další možnosti zpřesnění modelů

Vzhledem k vysokému počtu uložených sítí pro jednotlivé kombinace vstupních parametrů je možné pro predikci výtežků pyrolyzy s výhodou využít všech těchto vyzkoušených sítí. Na Obrázku 15 je zobrazeno porovnání krabicových a regresních grafů pro nejúspěšnější čtyřsíťový model využívající všech dostupných dodatečných charakteristik a průměrný odhad všech uložených čtyřsíťových modelů využívajících tuto kombinaci dodatečných parametrů. Jak je vidět, provedením simulace s využitím všech pro danou kombinaci vstupních parametrů a typu modelu dojde k částečnému vzájemnému vyrušení odlehlych hodnot a zvýšení spolehlivosti modelů. Nejlépe je tato skutečnost viditelná při zobrazení pouze validačních sad dat.



Obrázek 15 – Porovnání krabicových a regresních grafů pro nejúspěšnější čtyřsítový model využívající všech dostupných dodatečných charakteristik (vlevo) a průměrný odhad všech uložených čtyřsítových modelů využívajících tuto kombinaci dodatečných parametrů (vpravo). Zobrazena jsou data z predikce výtěžku ethylenu.

Zjevnou nevýhodou využití všech dostupných uložených modelů je výrazný nárůst výpočetní doby. K jedné simulaci je potom potřeba 100 či 200 násobek strojového času (v závislosti na typu modelu). Jedna simulace s pomocí čtyřsíťového modelu poté trvá přibližně sedm vteřin⁹. Vzhledem k tomu, že zlepšení vlastností modelů není nijak závratné je ekonomičnost tohoto postupu značně diskutabilní.

⁹Na počítači s procesorem Intel Core 2 Duo o frekvenci 1,8 GHz.

6 Závěr

V práci bylo zjištěno několik skutečností ohledně možností predikovat výtěžky pyrolýzy středněvroucích uhlovodíků pomocí umělých neuronových sítí.

1. Většina architektur sítí navrhovaných optimalizačním algoritmem obsahovala 5 neuronů ve skryté vrstvě a to jak pro jednosíťový, tak pro čtyřsíťový model.
2. Čtyřsíťový model dosahoval stabilně lepších výsledků, než model jednosíťový a to i v případě, že byl ve skryté vrstvě sítě uměle zachován počet neuronů odpovídající součtu počtu neuronů ve skrytých vrstvách jednotlivých sítí modelu čtyřsíťového. Plně propojená síťová architektura zřejmě není pro predikci výtěžků pyrolýzy vhodná.
Dále by bylo vhodné experimentovat s různými stupni propojenosti skryté a výstupní vrstvy modelu. To by ale bylo vzhledem k počtu možných variant časově velice náročné.
3. Vliv dodatečných parametrů na přesnost predikce byl zanedbatelný a to dokonce i v případě práce s více než jedním typem suroviny. To bylo velmi pravděpodobně zapříčiněno rozdílným fragmentovým složením a provozních vlastností obou zpracovávaných typů surovin a s rozšířením datové základny by důležitost doplňkových informací o surovině vzrostla.
4. U jednosíťového modelu se doplňkové informace projevovaly výrazněji. Úspěšnější byly zejména sítě využívající všechny dostupné informace a informaci o molární hustotě suroviny.
5. Model založený na ANN lze s úspěchem použít k modelování výtěžek pyrolýzy více typů surovin a to i v případě, že je k popisu jednoho z těchto typů pouze neúplný soubor experimentálních dat.
6. Výsledky modelů je možné dále mírně zlepšit využitím všech dostupných uložených sítí, ale tento postup je velmi náročný na výpočetní čas a nepřináší markantní zlepšení.

7 Seznam symbolů

d	hustota, [kg m ⁻³]
Δ_{abs}	absolutní odchylky, [hmot. %]
Δ_{nor}	normalizované odchylky, [-]
Φ	oddestilované množství, [hmot. %]
k	počítadlo vnitřních cyklů algoritmu, [-]
K_W	Watsonův charakterizační faktor, [-]
M	počet algoritmem ukládaných sítí, [-]
M_W	molární hmotnost, [g mol ⁻¹]
N	počet neuronů ve skryté vrstvě sítě, [-]
P_C	kritický tlak, [bar]
R	kritérium úspěšnosti sítě, [-]
S	standardní specifická hustota, [-]
T_b	normální bod varu, [K]
T_C	kritická teplota, [K]

Reference

- [1] Pašek, J.; Reiner, S. *Surovinová základna výroby organických látok*, 1st ed.; VŠCHT Praha: Praha, 1991.
- [2] Grantom, R. L.; Royer, D. J. A10. *Ullmann's Encyclopedia of Industrial Chemistry*, 5th ed.; VCH: New York, 1987; Chapter Ethylene, pp 47–68.
- [3] Eckert, E.; Belohlav, Z.; Vanek, T.; Zamostny, P.; Herink, T. ANN modelling of pyrolysis utilising the characterisation of atmospheric gas oil based on incomplete data. *Chem. Eng. Sci.* **2007**, *62* (18), 5021–5025.
- [4] Wauquier, J., Ed *Petroleum Refining*, 1st ed.; Institut Français du Pétrole Publications: Paris, 1994.
- [5] Eckert, E. Potřebujeme Pseudosložky?. *Chem. Listy* **2001**, *95*, 368–373.
- [6] Lin, B.; Liebovici, C. F.; Jørgensen, S. B. Optimal component lumping: Problem formulation and solution techniques. *Comput. Chem. Eng.* **2008**, *32* (6), 1167–1172.
- [7] Eckert, E.; Vanek, T. New approach to the characterisation of petroleum mixtures used in the modelling of separation processes. *Comput. Chem. Eng.* **2005**, *30* (3), 343–356.
- [8] Eckert, E.; Vanek, T. Extended Utilization of the Characterization of Petroleum Mixtures Based on Real Components. *Chem. Pap.* **2005**, *59* (6a), 428–433.
- [9] Aladwani, H. A.; Riazi, M. R. Some Guidelines fo Choosing a Characterisation Method for Petroleum Fractions in Process Simulators. *Chem. Eng. Research and Design* **2005**, *83* (A2), 160–166.
- [10] Belohlav, Z.; Zamostny, P.; Herink, T.; Eckert, E.; Vanek, T. A novel approach for the prediction of hydrocarbon thermal cracking product yields from the substitute feedstock composition. *Chemical Engineering & Technology* **2005**, *28* (10), 1166–1176.
- [11] Cuellar, F. D. R.; Le Roux, G. A. C. A Method of Representation of Petroleum by a Real Components Substitute Mixture. *Comput. Aided. Chem. Eng.* **2009**, *27*, 345–350.
- [12] Reed, R. D.; Marks II, R. J. *Neural Smithing*, 1st ed.; MIT Press: Cambridge, Massachusetts, 1998.
- [13] Singh, V.; Gupta, I.; Gupta, H. O. ANN-based estimator for distillation using Levenberg–Marquardt approach. *Eng. App. of Art. Int.* **2007**, *20* (2), 249–259.
- [14] Mathworks Inc. Neural Networks Toolbox documentation, 2011. Multilayer Networks and Backpropagation Training. <http://www.mathworks.com/help/toolbox/nnet/ug/bss33r9.html> (accessed May 11, 2011).
- [15] Nabavi, R.; Niaezi, A.; Salari, D.; Towfighi, J. Modeling of thermal cracking of LPG: Application of artificial neural network in prediction of the main product yields. *J. Anal. Appl. Pyrolysis* **2007**, *80* (1), 175–181.
- [16] Zhu, Q.; Li, Ch. Dimensionality Reduction with Input Training Neural Network and Its Application in Chemical Process Modelling. *Chinese Journal of Chemical Engineering* **2006**, *14* (5), 597–603.
- [17] Gao, G. Y.; Wang, M.; Ramshaw, C.; Li, X. G.; Yeung, H. Optimal operation of tubular reactors for naphtha cracking by numerical simulation. *Asia-Pac. J. Chem. Eng.* **2009**, *4*, 885–892.
- [18] Van Geem, K. M.; Hudebine, D.; Ramshaw, C.; Reyniers, M. F.; Wahl, F.; Verstræt, F.; Marin, G. B. Molecular reconstruction of naphtha steam cracking feedstocks based on commercial indices. *Comp. and Chem. Eng.* **2007**, *31*, 1020–1034.
- [19] Belohlav, Z.; Zamostny, P.; Herink, T. The kinetic model of thermal cracking for olefins production. *Chem. Eng. Process.* **2003**, *42* (6), 461–473.

- [20] Kumar, P.; Kunzru, D. Modeling of Naphtha Pyrolysis . *Ind. Eng. Chem. Process Des. Dev.* **1985**, *24* (3), 774–782.
- [21] Sadramelia, S. M.; Green, A. E. S. Modeling of Naphtha Pyrolysis. *J. Anal. Appl. Pyrolysis* **2005**, *73* (2), 305–313.

Přílohy

A Algoritmus výběru sítí pro jednosíťový model

Výběr sítí je založen na váženém průměru korelačních koeficientu na validační sadě dat proti jednotlivým produktům pyrolýzy. Nejvyšší váha je přiřazena schopnosti predikovat výtežky ethylenu a propylenu. Data byla dělena mezi validační a trénovací sadu v poměru 60:40, testovací sada nebyla pro nedostatek dat využívána.

```
while L >= 0 && toc < doba_trvani %bud zuzim interval nebo dojde cas

    %sit bude konstruovana s nahodnym pocitem neuronu z predem urceneho
    %intervalu, nasledne bude na zaklade uspesnosti danych siti interval
    %postupne zuzovan

    Nn = lower + round(coef*rand(1));% Nn_first in <4;8>
    net = newfit(data_sources,data_targets,Nn);

    net.trainParam.showWindow = false;%nechci videt krasne okenko
    net.trainParam.showCommandLine = true;%necham si udelat vypis do cmw
    net.trainParam.epochs = 30;%snizeni max. poctu iteraci (zrychleni vypoctu)

    % Rozdeleni dat na trenovaci, validacni a testovaci sadu
    % defaultni je 0.7/0.15/0.15
    net.divideFcn = 'divideind'; % vyberu si funkci na rozdeleni dat
    net.divideParam.trainInd = trainInd;
    net.divideParam.valInd = valInd;
    net.divideParam.testInd = testInd;

    %ulozim si net a tr pro analyzu treninkovych dat a dalsi zpracovani
    [net tr]= train(net,data_sources,data_targets);

    %protahne siti vstupni data - kvuli regresi (posouzeni kvality site)
    data_outputs = sim(net,data_sources);

    %vyberu za dat validacni sadu, a urcim vhodnost site podle prumerneho
    %R na jednotlivych produktech
    vOut = data_outputs(:,tr.valInd);
    vTarg = data_targets(:,tr.valInd);

    data_outputs = sim(net,data_sources);

    % rozdelim data na jednotlive produkty
    methan_vOut = vOut(1,:);
    methan_vTarg = vTarg(1,:);

    ethylen_vOut = vOut(2,:);
    ethylen_vTarg = vTarg(2,:);

    propylen_vOut = vOut(3,:);
    propylen_vTarg = vTarg(3,:);

    olej_vOut = vOut(4,:);
    olej_vTarg = vTarg(4,:);

    %urci korelacni koeficient (nejvyssi vaha je priazena validaci
    %ethylenu a propylenu)

    %dopoctu korelacni koeficienty na jednotlivych sadach dar
```

```

Rv_meth = corrcoef(methan_vTarg,methan_vOut);
Rv_eth = corrcoef(ethylen_vTarg,ethylen_vOut);
Rv_prop = corrcoef(propylen_vTarg,propylen_vOut);
Rv_olej = corrcoef(olej_vTarg,olej_vOut);

Rv_meth = Rv_meth(1,2);
Rv_eth = Rv_eth(1,2);
Rv_prop = Rv_prop(1,2);
Rv_olej = Rv_olej(1,2);

Rr = mean([Rv_meth Rv_eth Rv_eth...
           Rv_prop Rv_prop Rv_olej]);

%rozhodnu, zda se mi vyplati ulozit soucasnou sit
if Rr > min(R)
    j = find(R == min(R),1);%kdyby se nahodou indexy rovnaly

    %ulozi soucasnou sit do cell
    nets{1,j} = net;

    %ulozim i treninkove parametry pro vykresleni
    trs{1,j} = tr;

    %prepisu hodnotu korelacniho koeficientu a pocet jejich neuronu v NumNeur
    R(j) = Rr;
    NumNeur(j) = net.layers{1}.dimensions;
end

clear net % vycistim si pracovni promennou

%statisticke vyhodnoceni provadim kazdych stat_koef*NN vyzk. siti
if mod(k,stat_koef*NN) == 0
    %v pripade rovnosti zrejme brat mensi - mode
    M = mode(NumNeur);
    lower = M - L;
    if lower <= 0 %nemuzu mit pocty neuronu <= 0
        lower = 1;
    end
    coef = L*2;
    L = L-1;

    %skocim do slozky, do ktere ukladam data
    cd './nets'
    cd(day);cd(dir);

    %ulozim si histogram ulozenych siti
    hist(NumNeur,min(NumNeur):max(NumNeur));% ctnosti v jednotl. bodech
    saveas(gcf,['hist_NumNeur', mat2str(hist_count)],'png');
    close(gcf);

    %zvysim pocitadlo histogramu o 1
    hist_count = hist_count + 1;

    %vratim se do pracovni slozky
    cd ...;cd ...;cd ...;
end
k = k+1;%pocitadlo, kolik bylo vyzkouseno siti
end

```

B Tabulky využívaných dat

Tabulka 7 – Tabulka dat využitých při modelování, Fragmentové složení a Data z provozu

Pokus	Fragmentové složení											Data z provozu			
	P	S	S _N	T	T _N	Q	Q _N	P _E	S _E	T _E	S _A	T _A	COT [°C]	nástrík na vlásenku [t h ⁻¹]	ředění [–]
PrBi1104	0,344	0,323	0,163	0,052	0,050	0,002	0,001	0,000	0,000	0,000	0,046	0,014	826,50	4,100	0,500
PrBi1804	0,347	0,324	0,161	0,053	0,049	0,001	0,001	0,000	0,000	0,000	0,048	0,014	826,80	4,100	0,600
PrBi2004	0,341	0,323	0,164	0,052	0,052	0,002	0,001	0,000	0,000	0,000	0,048	0,015	826,90	4,100	0,600
PrBi2704	0,321	0,305	0,166	0,048	0,065	0,001	0,001	0,000	0,000	0,000	0,065	0,023	821,80	3,770	0,600
PrBi0205	0,339	0,311	0,173	0,053	0,052	0,001	0,001	0,000	0,000	0,000	0,053	0,016	826,80	3,940	0,600
PrBi0405	0,353	0,329	0,153	0,056	0,047	0,002	0,001	0,000	0,000	0,000	0,045	0,013	827,00	3,920	0,600
PrBi1605	0,342	0,320	0,166	0,051	0,050	0,001	0,001	0,000	0,000	0,000	0,052	0,016	829,70	4,100	0,600
PrBi2905_1	0,356	0,323	0,167	0,056	0,046	0,002	0,001	0,000	0,000	0,000	0,038	0,011	812,20	3,300	0,600
PrBi2905_2	0,356	0,323	0,167	0,056	0,046	0,002	0,001	0,000	0,000	0,000	0,038	0,011	822,10	3,300	0,600
PrBi2905_3	0,356	0,323	0,167	0,056	0,046	0,002	0,001	0,000	0,000	0,000	0,038	0,011	832,40	3,300	0,600
PrBi3005_1	0,350	0,326	0,158	0,053	0,045	0,001	0,001	0,000	0,000	0,000	0,049	0,014	813,70	3,600	0,600
PrBi3005_2	0,350	0,326	0,158	0,053	0,045	0,001	0,001	0,000	0,000	0,000	0,049	0,014	824,00	3,600	0,600
PrBi3005_3	0,350	0,326	0,158	0,053	0,045	0,001	0,001	0,000	0,000	0,000	0,049	0,014	834,00	3,600	0,600
PrBi3105_1	0,345	0,319	0,184	0,052	0,048	0,001	0,001	0,000	0,000	0,000	0,038	0,010	815,70	3,900	0,600
PrBi3105_2	0,345	0,319	0,184	0,052	0,048	0,001	0,001	0,000	0,000	0,000	0,038	0,010	825,90	3,900	0,600
PrBi3105_3	0,345	0,319	0,184	0,052	0,048	0,001	0,001	0,000	0,000	0,000	0,038	0,010	836,00	3,900	0,600
PrBi1306	0,368	0,335	0,150	0,064	0,038	0,002	0,001	0,000	0,001	0,000	0,032	0,008	828,00	4,060	0,600
PrBi1406	0,356	0,325	0,166	0,058	0,044	0,002	0,001	0,000	0,000	0,000	0,038	0,010	827,90	4,070	0,600
PrBi1506	0,373	0,317	0,150	0,062	0,043	0,004	0,001	0,000	0,000	0,000	0,037	0,012	828,00	4,070	0,600
PrBi2206	0,353	0,314	0,181	0,055	0,049	0,001	0,001	0,000	0,000	0,000	0,036	0,008	825,60	4,000	0,610
PrBi1107	0,377	0,331	0,153	0,065	0,036	0,002	0,001	0,000	0,000	0,000	0,026	0,005	840,00	4,030	0,600
PrBi1807	0,356	0,305	0,177	0,057	0,047	0,003	0,001	0,000	0,000	0,000	0,039	0,010	830,10	3,900	0,500
PrBi1907	0,353	0,309	0,177	0,053	0,049	0,002	0,001	0,000	0,000	0,000	0,042	0,012	832,20	3,900	0,500
PrBi0611	0,358	0,324	0,163	0,057	0,045	0,003	0,001	0,000	0,000	0,000	0,039	0,010	830,00	3,000	0,550
PrBi0602	0,351	0,312	0,175	0,050	0,038	0,002	0,001	0,004	0,008	0,005	0,042	0,010	830,00	3,800	0,500

Pokračování na další straně

Pokračování z předchozí strany																
Pokus	Fragmentové složení											Data z provozu				
	P	S	S _N	T	T _N	Q	Q _N	P _E	S _E	T _E	S _A	T _A	COT [°C]	nástr. na vls. [t h ⁻¹]	ředění [-]	
88	PrBi0702	0,347	0,311	0,187	0,049	0,042	0,002	0,001	0,001	0,010	0,003	0,038	0,009	841,10	4,600	0,500
	PrBi1302	0,354	0,321	0,168	0,054	0,046	0,002	0,001	0,000	0,000	0,000	0,042	0,011	825,60	5,000	0,500
	PrBi0305	0,350	0,321	0,170	0,052	0,048	0,001	0,001	0,000	0,000	0,000	0,044	0,012	829,29	4,140	0,502
	PrBi2606	0,332	0,299	0,179	0,045	0,056	0,002	0,001	0,001	0,012	0,003	0,051	0,016	827,02	3,691	0,500
	APO1906	0,214	0,336	0,211	0,052	0,082	0,012	0,000	0,000	0,000	0,000	0,068	0,026	795,30	4,000	0,750
	APO2006	0,074	0,427	0,362	0,012	0,070	0,000	0,000	0,000	0,000	0,000	0,046	0,009	800,10	4,000	0,750
	APO1307	0,084	0,488	0,225	0,009	0,045	0,000	0,000	0,000	0,000	0,000	0,125	0,023	802,10	4,100	0,800
	APO2007	0,070	0,402	0,402	0,012	0,078	0,000	0,000	0,000	0,000	0,000	0,029	0,005	801,40	4,400	0,750
	APO2407	0,085	0,485	0,230	0,009	0,047	0,000	0,000	0,000	0,000	0,000	0,121	0,022	798,90	4,200	0,750
	APO2507	0,082	0,484	0,234	0,007	0,048	0,000	0,000	0,000	0,000	0,000	0,123	0,023	799,00	4,200	0,750
	APO2607	0,068	0,396	0,420	0,011	0,082	0,000	0,000	0,000	0,000	0,000	0,020	0,004	799,10	4,200	0,700
	APO3005	0,109	0,431	0,300	0,017	0,072	0,000	0,000	0,000	0,000	0,000	0,060	0,011	791,76	3,831	0,745
	APO2201	0,093	0,501	0,206	0,011	0,038	0,000	0,003	0,000	0,002	0,002	0,123	0,023	795,17	3,764	0,750
	APO2803	0,067	0,443	0,330	0,000	0,066	0,000	0,000	0,000	0,000	0,000	0,080	0,015	795,76	4,097	0,750

Tabulka 8 – Tabulka dat využitých při modelování, Údaje z GC a Výtežky sledovaných produktů

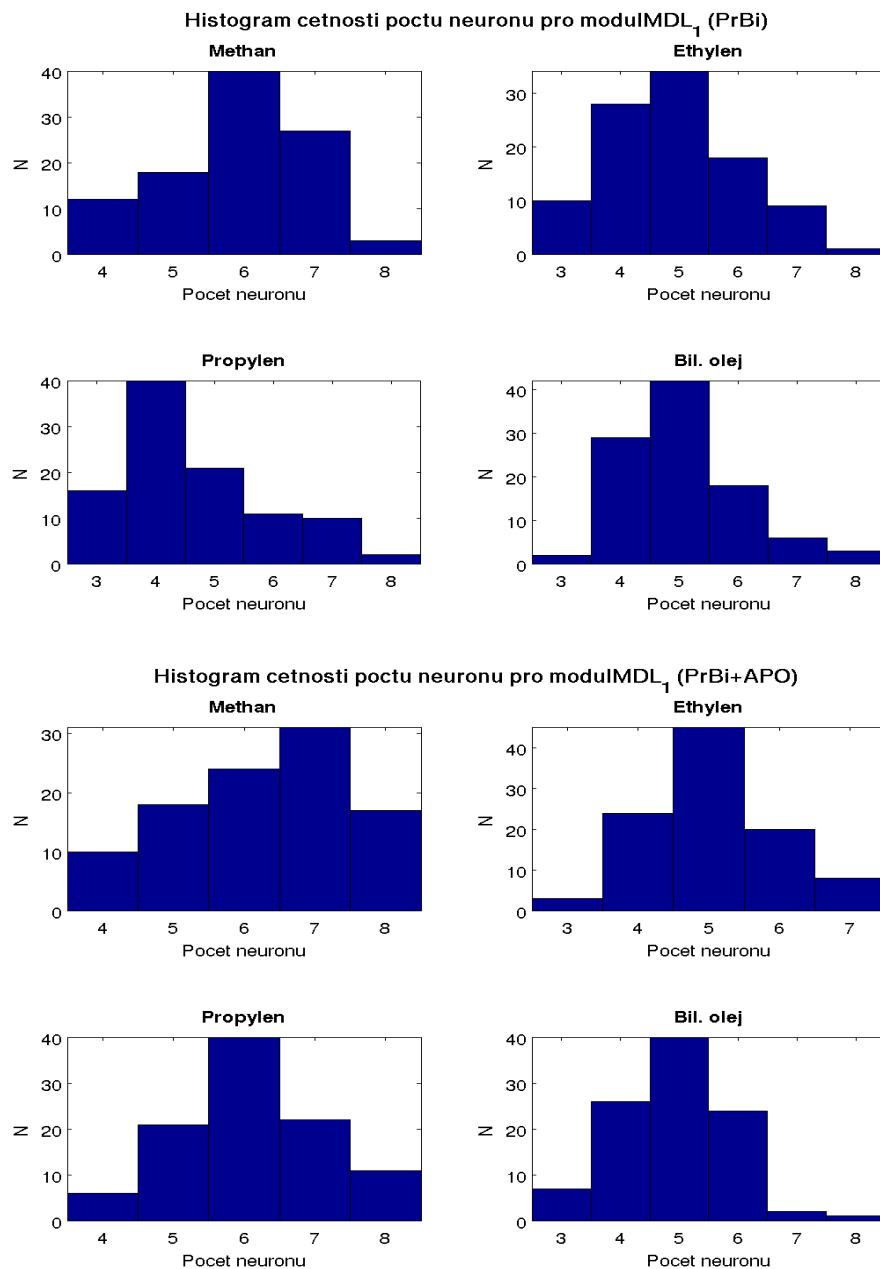
Pokus	Údaje z GC					Výtežky			
	M_w [g mol ⁻¹]	d [kg m ⁻³]	T10	T50 [°C]	T90	Methan	Ethylen [hmot. %]	Propylen	Bil. olej
PrBi1104	94,882	705,277	96,908	197,366	303,476	15,53	25,60	14,89	3,91
PrBi1804	95,780	709,502	96,908	209,156	303,476	15,03	25,86	15,15	3,56
PrBi2004	97,221	712,474	96,908	209,156	303,476	15,37	25,71	14,76	4,15
PrBi2704	105,068	732,257	145,886	245,372	331,700	14,87	23,25	13,74	5,85
PrBi0205	96,200	712,395	96,908	209,156	306,338	14,45	25,58	14,79	3,98
PrBi0405	95,838	707,752	96,908	197,366	303,476	15,91	25,73	14,56	3,65
PrBi1605	96,104	711,850	96,908	209,156	303,476	14,52	25,20	14,29	3,80
PrBi2905_1	92,522	700,576	96,908	194,090	291,974	13,97	25,05	15,57	2,94
PrBi2905_2	92,431	700,621	96,908	194,090	291,614	14,82	25,73	14,60	3,82
PrBi2905_3	91,634	698,636	96,908	193,604	282,416	15,78	26,85	13,99	3,73
PrBi3005_1	93,482	703,878	96,908	197,330	298,346	13,61	24,49	15,31	3,29
PrBi3005_2	94,116	706,554	96,908	197,330	303,476	14,18	25,01	14,42	4,01
PrBi3005_3	94,394	707,285	96,908	197,330	303,476	14,97	25,68	13,49	4,64
PrBi3105_1	92,514	703,084	96,908	194,090	289,904	13,44	24,45	15,41	2,98
PrBi3105_2	91,940	701,482	96,908	194,090	282,416	14,15	25,29	14,64	3,50
PrBi3105_3	91,947	701,400	96,908	194,090	282,416	15,12	26,23	13,91	4,35
PrBi1306	92,350	696,048	96,908	194,090	267,728	14,95	25,39	15,05	3,63
PrBi1406	93,208	701,790	96,908	197,096	282,416	14,61	25,10	14,57	3,62
PrBi1506	92,788	698,391	82,112	194,090	303,476	14,81	24,65	14,54	3,51
PrBi2206	91,731	699,700	96,908	194,090	277,160	14,50	24,92	14,23	4,32
PrBi1107	89,266	688,000	96,908	177,296	245,372	15,99	26,97	14,36	3,41
PrBi1807	91,717	698,300	96,908	194,090	289,904	15,20	24,90	14,67	3,70
PrBi1907	92,367	702,000	96,908	195,386	291,614	15,55	25,18	14,36	3,71
PrBi0611	92,955	701,400	96,908	194,090	288,392	18,26	26,36	13,94	4,08
PrBi0602	89,157	697,790	93,308	177,296	281,048	16,26	26,88	14,10	4,31
PrBi0702	90,348	700,874	96,908	177,296	278,636	17,75	26,72	12,73	6,09
PrBi1302	92,463	702,013	96,908	194,090	289,904	16,16	25,54	14,54	4,04

Pokračování na další straně

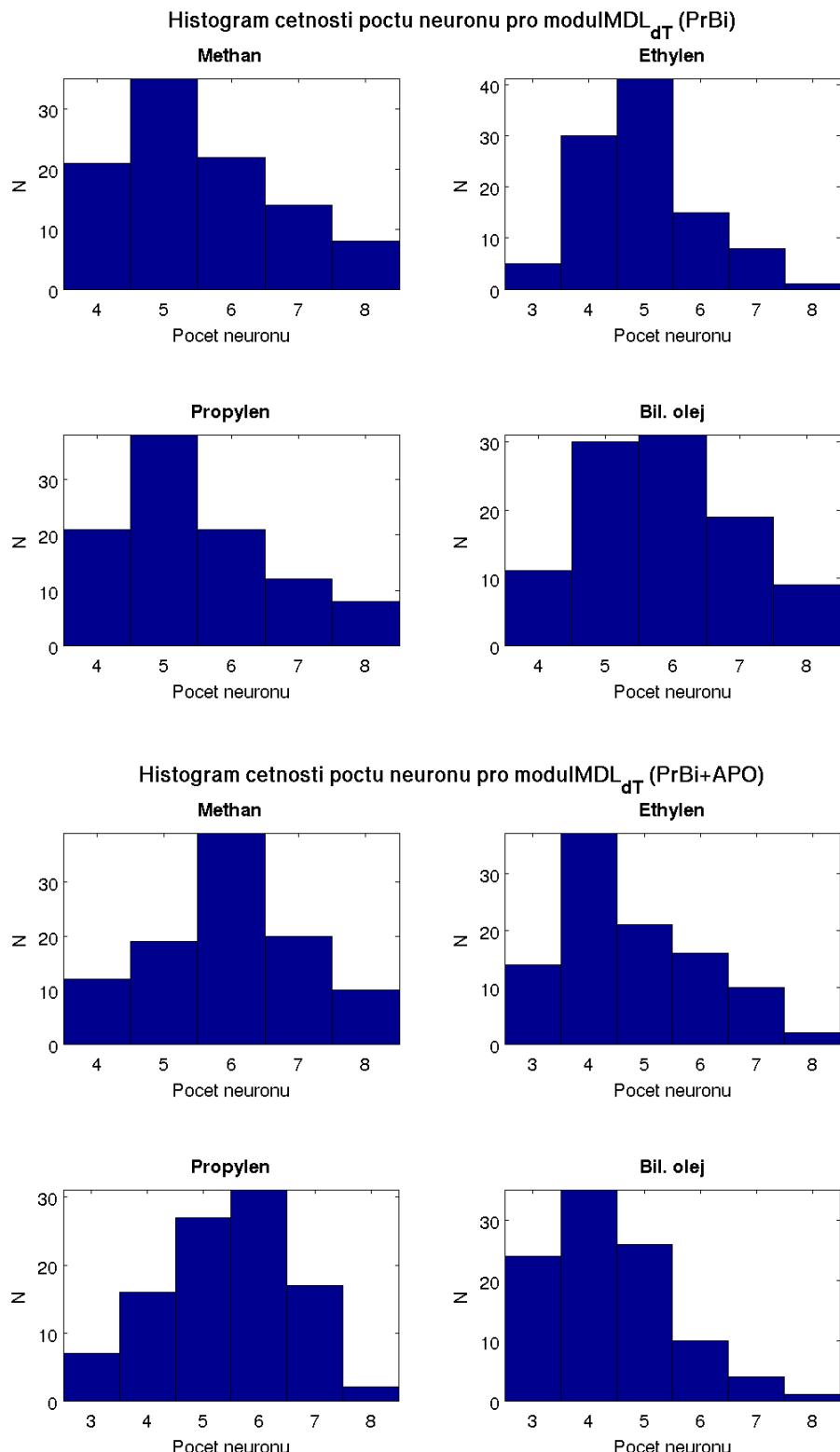
Pokračování z předchozí strany

Pokus	Údaje z GC					Výtěžky			
	M_w [g mol ⁻¹]	d [kg m ⁻³]	T10	T50 [°C]	T90	Methan	Ethylen [hmot. %]	Propylen	Bil. olej
PrBi0305	93,133	704,374	96,908	197,096	298,346	15,07	24,49	13,13	8,41
PrBi2606	97,417	718,444	96,908	213,674	313,916	15,73	24,80	13,06	4,66
APO1906	193,406	837,000	218,000	266,500	335,500	9,82	19,73	12,72	18,56
APO2006	193,827	837,000	219,000	267,000	335,500	10,01	19,66	12,26	20,53
APO1307	194,694	836,900	219,500	268,000	339,000	10,31	23,18	13,94	11,36
APO2007	193,240	835,900	213,500	266,000	335,000	9,96	20,52	12,50	17,87
APO2407	192,936	835,400	212,500	265,500	334,500	10,01	20,46	12,51	18,31
APO2507	191,794	834,900	211,500	264,000	332,000	9,97	20,47	12,48	18,35
APO2607	191,794	834,900	211,500	264,000	332,000	10,31	20,41	12,44	18,43
APO3005	185,100	828,000	197,500	254,000	330,500	11,10	20,60	13,61	15,74
APO2201	207,667	846,000	223,500	285,500	346,500	10,63	19,25	12,31	21,87
APO2803	206,216	842,900	224,500	283,000	343,500	10,51	21,41	13,52	16,92

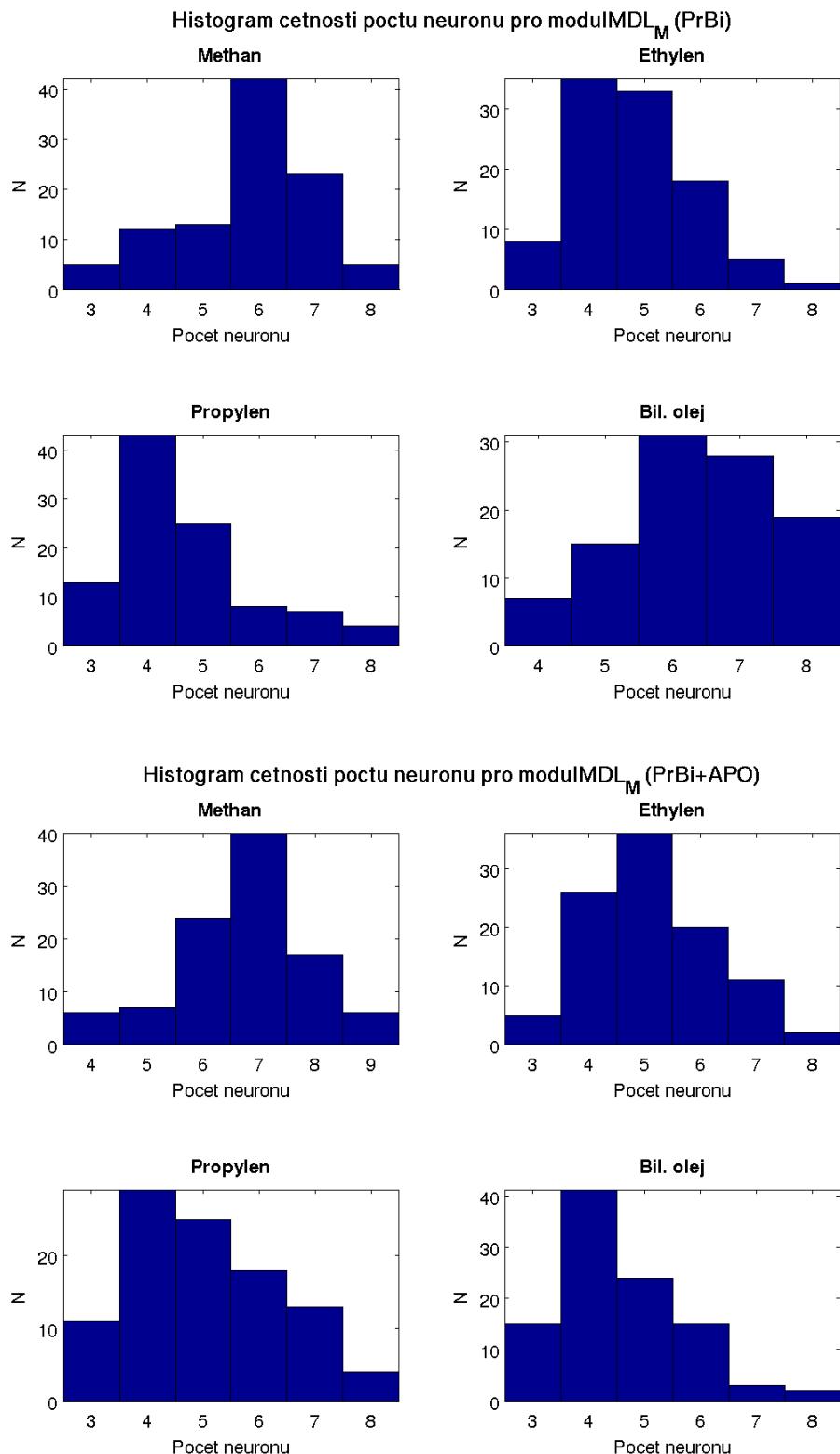
C Histogramy četnosti počtu neuronů ve skryté vrstvě pro vybrané čtyřsíťové modely



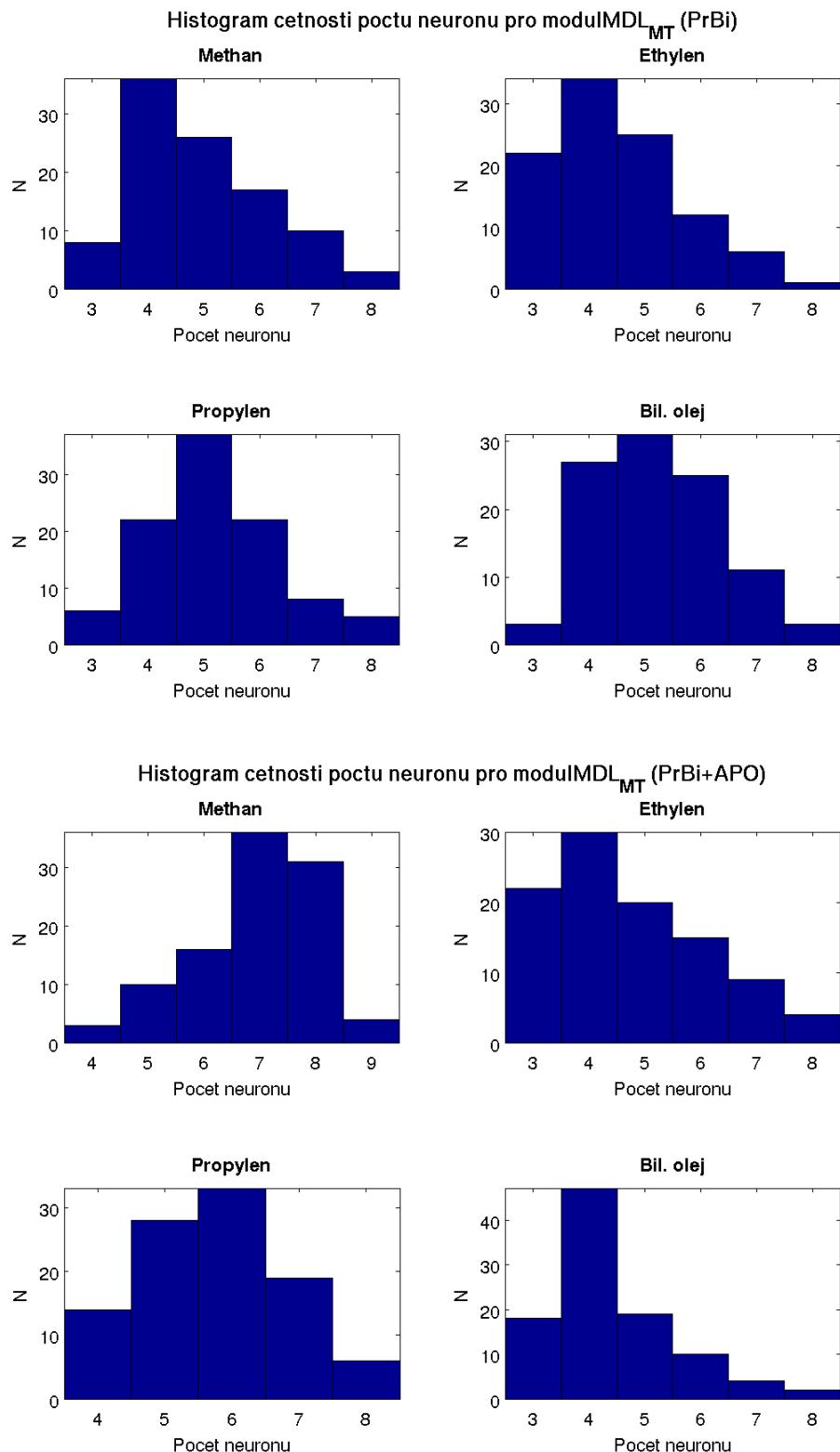
Obrázek 16 – Porovnání četností počtu neuronů ve skryté vrstvě algoritmem vybraných sítí pro čtyřsíťový model ModulMDL-1. Horní grafy reprezentují počty neuronů ve skrytých vrstvách sítí navržených pro predikci pouze primárních výtěžků pyrolýzy benzínů, spodní pak výsledek optimalizace architektury pro modelování pyrolýzy primárních benzínů i plynových olejů.



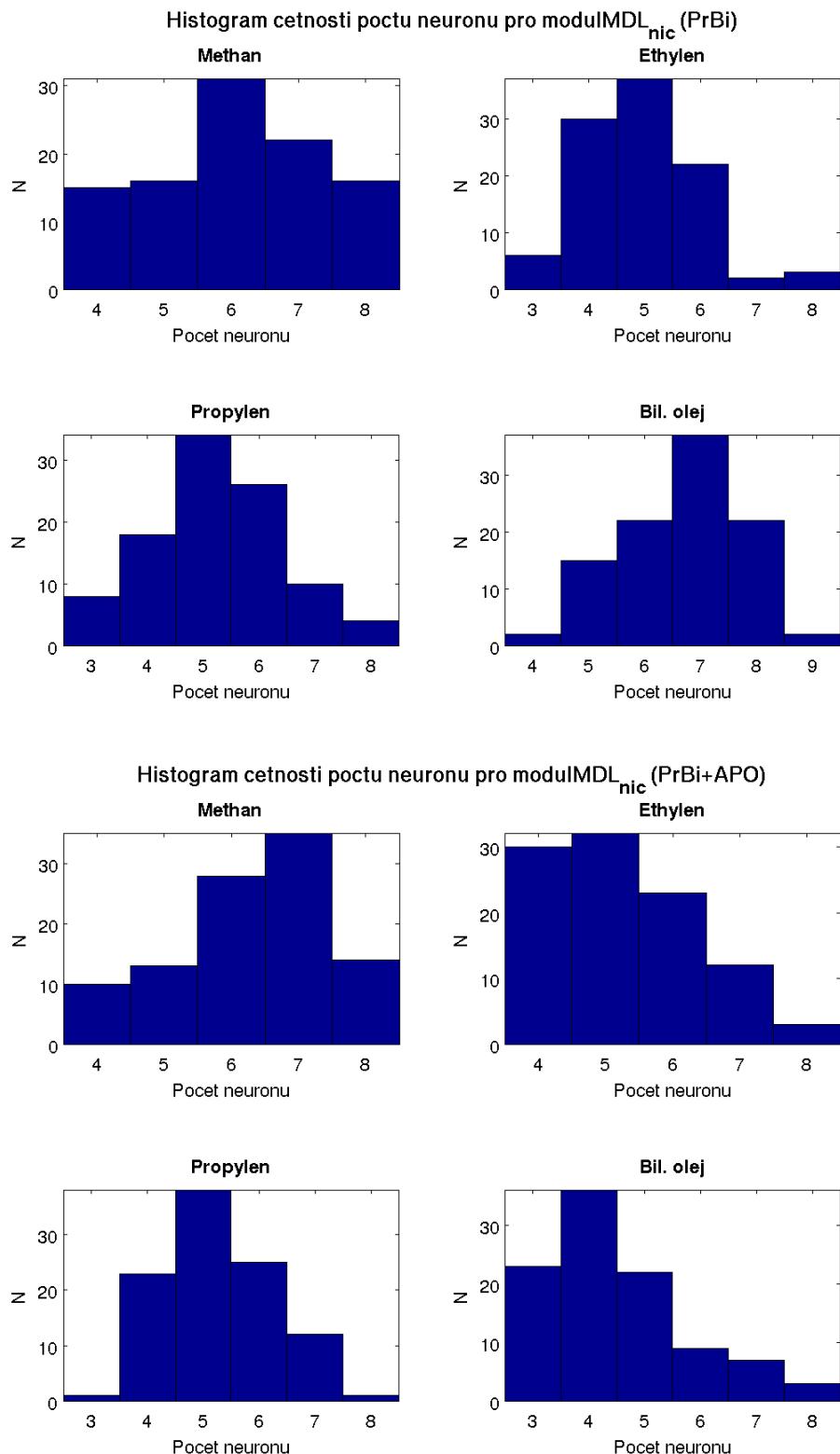
Obrázek 17 – Porovnání četností počtu neuronů ve skryté vrstvě algoritmem vybraných sítí pro čtyřsíťový model ModulMDL-dT. Horní grafy reprezentují počty neuronů ve skrytých vrstvách sítí navržených pro predikci pouze primárních výtěžků pyrolyzy benzínů, spodní pak výsledek optimalizace architektury pro modelování pyrolyzy primárních benzínů i plynových olejů.



Obrázek 18 – Porovnání četností počtu neuronů ve skryté vrstvě algoritmem vybraných sítí pro čtyřsíťový model ModulMDL_M. Horní grafy reprezentují počty neuronů ve skrytých vrstvách sítí navržených pro predikci pouze primárních výtěžků pyrolyzy benzínů, spodní pak výsledek optimalizace architektury pro modelování pyrolyzy primárních benzínů i plynových olejů.



Obrázek 19 – Porovnání četností počtu neuronů ve skryté vrstvě algoritmem vybraných sítí pro čtyřsíťový model ModulMDL-MT. Horní grafy reprezentují počty neuronů ve skrytých vrstvách sítí navržených pro predikci pouze primárních výtěžků pyrolyzy benzínů, spodní pak výsledek optimalizace architektury pro modelování pyrolyzy primárních benzínů i plynových olejů.



Obrázek 20 – Porovnání četností počtu neuronů ve skryté vrstvě algoritmem vybraných sítí pro čtyřsíťový model ModulMDL_{nic}. Horní grafy reprezentují počty neuronů ve skrytých vrstvách sítí navržených pro predikci pouze primárních výtěžků pyrolýzy benzínů, spodní pak výsledek optimalizace architektury pro modelování pyrolýzy primárních benzínů i plynových olejů.

D Přehled průměrů a výběrových rozptylů absolutních a normovaných odchylek

Zde jsou prezentovány tabulky s průměrnými absolutními a normovanými odchylkami a jejich rozptyly vypočtené na všech ukládaných sítích (100 sítí/produkt pro modulMDL a 200 sítí/model pro netMdl).

Tabulka 9 – Methan, porovnání průměrů a rozptylů Δ_{abs} a Δ_{nor} pro jednotlivé modely a datové sady

Skript	Rozsah dat	$\overline{\Delta_{abs}}$ [hmot. %]	$s^2(\Delta_{abs})$	$\overline{\Delta_{nor}}$ [-]	$s^2(\Delta_{nor})$
modulMDL_1	PrBi	0.2453	0.0041	1.2068	2.3845
modulMDL_d	PrBi	0.3776	0.0008	1.8577	0.4499
modulMDL_dT	PrBi	0.2185	0.0005	1.0749	0.2919
modulMDL_M	PrBi	0.4788	0.0026	2.3555	1.5049
modulMDL_Md	PrBi	0.3022	0.0032	1.4868	1.8495
modulMDL_MT	PrBi	0.3418	0.0022	1.6815	1.2932
modulMDL_nic	PrBi	0.5685	0.0107	2.7972	6.2593
modulMDL_T	PrBi	0.2461	0.0028	1.2106	1.6699
netMdl_1	PrBi	0.3365	0.0042	1.6554	2.4624
netMdl_d	PrBi	1.4081	0.2619	6.9281	153.4719
netMdl_dT	PrBi	0.8834	0.3563	4.3464	208.7845
netMdl_M	PrBi	0.5153	0.0697	2.5355	40.8519
netMdl_Md	PrBi	0.5051	0.0598	2.4853	35.0260
netMdl_MT	PrBi	0.5194	0.0273	2.5554	15.9684
netMdl_nic	PrBi	0.7582	0.1457	3.7306	85.3551
netMdl_T	PrBi	0.5974	0.0543	2.9393	31.7996
modulMDL_1	PrBi+APO	0.2883	0.0047	1.4187	2.7289
modulMDL_d	PrBi+APO	0.2341	0.0014	1.1518	0.8353
modulMDL_dT	PrBi+APO	0.2698	0.0035	1.3273	2.0267
modulMDL_M	PrBi+APO	0.2652	0.0082	1.3048	4.8254
modulMDL_Md	PrBi+APO	0.3392	0.0151	1.6689	8.8523
modulMDL_MT	PrBi+APO	0.2582	0.0039	1.2702	2.2655
modulMDL_nic	PrBi+APO	0.3319	0.0056	1.6329	3.2783
modulMDL_T	PrBi+APO	0.2842	0.0020	1.3984	1.1951
netMdl_1	PrBi+APO	0.2939	0.0043	1.4459	2.5030
netMdl_d	PrBi+APO	0.5287	0.0540	2.6013	31.6138
netMdl_dT	PrBi+APO	0.6445	0.0282	3.1711	16.4994
netMdl_M	PrBi+APO	1.1303	0.0973	5.5614	57.0397
netMdl_Md	PrBi+APO	0.7073	0.0847	3.4797	49.6273
netMdl_MT	PrBi+APO	0.3285	0.0065	1.6163	3.8042
netMdl_nic	PrBi+APO	0.3919	0.0132	1.9283	7.7608
netMdl_T	PrBi+APO	0.5301	0.0181	2.6081	10.6138

Tabulka 10 – Ethylen, porovnání průměrů a rozptylů Δ_{abs} a Δ_{nor} pro jednotlivé modely a datové sady

Skript	Rozsah dat	$\overline{\Delta_{abs}}$ [hmot. %]	$s^2(\Delta_{abs})$	$\overline{\Delta_{nor}}$ [-]	$s^2(\Delta_{nor})$
modulMDL_1	PrBi	0.2823	0.0019	0.8401	0.1511
modulMDL_d	PrBi	0.2717	0.0013	0.8087	0.1024
modulMDL_dT	PrBi	0.3226	0.0031	0.9602	0.2428
modulMDL_M	PrBi	0.3755	0.0045	1.1175	0.3545
modulMDL_Md	PrBi	0.2399	0.0013	0.7141	0.1037
modulMDL_MT	PrBi	0.4333	0.0036	1.2897	0.2858
modulMDL_nic	PrBi	0.3260	0.0019	0.9703	0.1486
modulMDL_T	PrBi	0.3366	0.0069	1.0017	0.5379
netMdl_1	PrBi	0.4309	0.0125	1.2825	0.9845
netMdl_d	PrBi	1.0700	0.4344	3.1846	34.0862
netMdl_dT	PrBi	0.3417	0.0025	1.0170	0.1996
netMdl_M	PrBi	0.4161	0.0116	1.2384	0.9136
netMdl_Md	PrBi	0.3487	0.0022	1.0377	0.1721
netMdl_MT	PrBi	0.3195	0.0010	0.9510	0.0790
netMdl_nic	PrBi	0.3496	0.0002	1.0406	0.0181
netMdl_T	PrBi	0.5313	0.0203	1.5812	1.5903
modulMDL_1	PrBi+APO	0.5086	0.0137	1.5138	1.0744
modulMDL_d	PrBi+APO	0.4541	0.0130	1.3515	1.0198
modulMDL_dT	PrBi+APO	0.7583	0.0190	2.2568	1.4882
modulMDL_M	PrBi+APO	0.4722	0.0134	1.4054	1.0553
modulMDL_Md	PrBi+APO	0.7802	0.0628	2.3220	4.9251
modulMDL_MT	PrBi+APO	0.7984	0.0237	2.3763	1.8634
modulMDL_nic	PrBi+APO	0.3514	0.0034	1.0458	0.2646
modulMDL_T	PrBi+APO	0.4700	0.0439	1.3989	3.4438
netMdl_1	PrBi+APO	0.7266	0.3043	2.1625	23.8780
netMdl_d	PrBi+APO	0.5044	0.0346	1.5013	2.7168
netMdl_dT	PrBi+APO	0.7203	0.1058	2.1439	8.3016
netMdl_M	PrBi+APO	0.7380	0.1109	2.1965	8.7043
netMdl_Md	PrBi+APO	0.7762	0.1047	2.3100	8.2164
netMdl_MT	PrBi+APO	0.7674	0.3874	2.2840	30.3978
netMdl_nic	PrBi+APO	0.4880	0.1345	1.4523	10.5489
netMdl_T	PrBi+APO	0.6415	0.2851	1.9093	22.3684

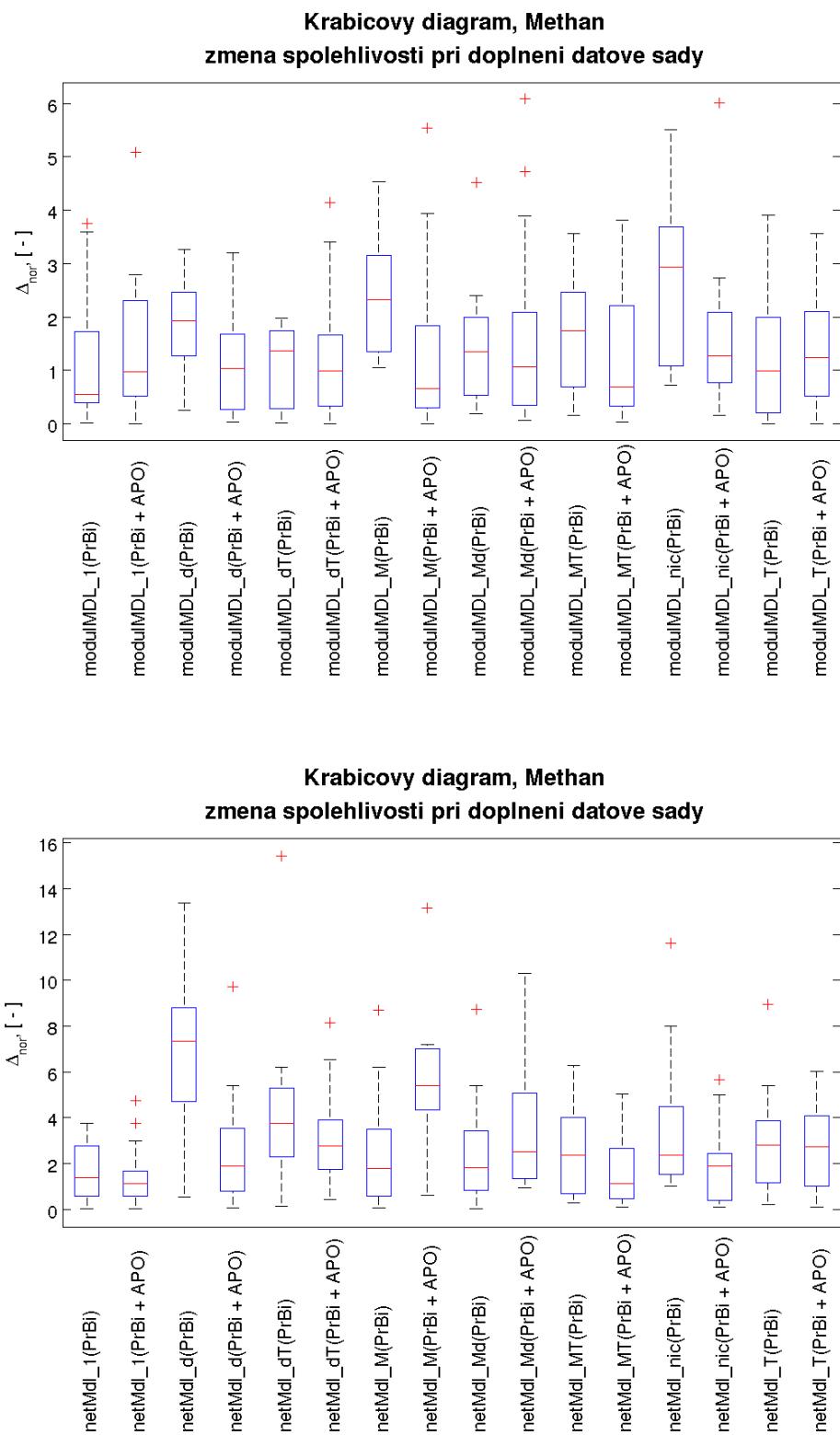
Tabulka 11 – Propylen, porovnání průměrů a rozptylů Δ_{abs} a Δ_{nor} pro jednotlivé modely a datové sady

Skript	Rozsah dat	$\overline{\Delta_{abs}}$ [hmot. %]	$s^2(\Delta_{abs})$	$\overline{\Delta_{nor}}$ [-]	$s^2(\Delta_{nor})$
modulMDL_1	PrBi	0.5267	0.0191	2.7062	13.2848
modulMDL_d	PrBi	0.3005	0.0022	1.5438	1.5509
modulMDL_dT	PrBi	0.4741	0.0030	2.4358	2.0702
modulMDL_M	PrBi	0.3196	0.0042	1.6421	2.9216
modulMDL_Md	PrBi	0.3396	0.0052	1.7448	3.5998
modulMDL_MT	PrBi	0.3359	0.0112	1.7256	7.8148
modulMDL_nic	PrBi	0.2783	0.0101	1.4301	7.0471
modulMDL_T	PrBi	0.4693	0.0300	2.4114	20.9415
netMdl_1	PrBi	0.5430	0.0410	2.7901	28.5753
netMdl_d	PrBi	0.4546	0.0158	2.3360	11.0246
netMdl_dT	PrBi	0.3996	0.0030	2.0532	2.0583
netMdl_M	PrBi	0.3564	0.0231	1.8313	16.0888
netMdl_Md	PrBi	0.4248	0.0169	2.1827	11.7680
netMdl_MT	PrBi	0.3955	0.0081	2.0321	5.6671
netMdl_nic	PrBi	0.4838	0.0629	2.4860	43.8378
netMdl_T	PrBi	0.4794	0.0213	2.4632	14.8769
modulMDL_1	PrBi+APO	0.3381	0.0034	1.7374	2.3878
modulMDL_d	PrBi+APO	0.3142	0.0030	1.6145	2.0897
modulMDL_dT	PrBi+APO	0.3343	0.0040	1.7177	2.7865
modulMDL_M	PrBi+APO	0.3395	0.0048	1.7445	3.3468
modulMDL_Md	PrBi+APO	0.2752	0.0035	1.4141	2.4527
modulMDL_MT	PrBi+APO	0.3989	0.0022	2.0498	1.5289
modulMDL_nic	PrBi+APO	0.3021	0.0014	1.5521	1.0080
modulMDL_T	PrBi+APO	0.3437	0.0043	1.7658	2.9801
netMdl_1	PrBi+APO	0.3986	0.0032	2.0482	2.2105
netMdl_d	PrBi+APO	0.4715	0.0097	2.4225	6.7939
netMdl_dT	PrBi+APO	0.5009	0.0046	2.5735	3.2072
netMdl_M	PrBi+APO	0.2945	0.0047	1.5130	3.3092
netMdl_Md	PrBi+APO	0.4650	0.0060	2.3893	4.1848
netMdl_MT	PrBi+APO	0.3723	0.0043	1.9128	3.0002
netMdl_nic	PrBi+APO	0.4327	0.0043	2.2231	3.0214
netMdl_T	PrBi+APO	0.4947	0.0070	2.5416	4.8960

Tabulka 12 – Bilanční olej, porovnání průměrů a rozptylů Δ_{abs} a Δ_{nor} pro jednotlivé modely a datové sady

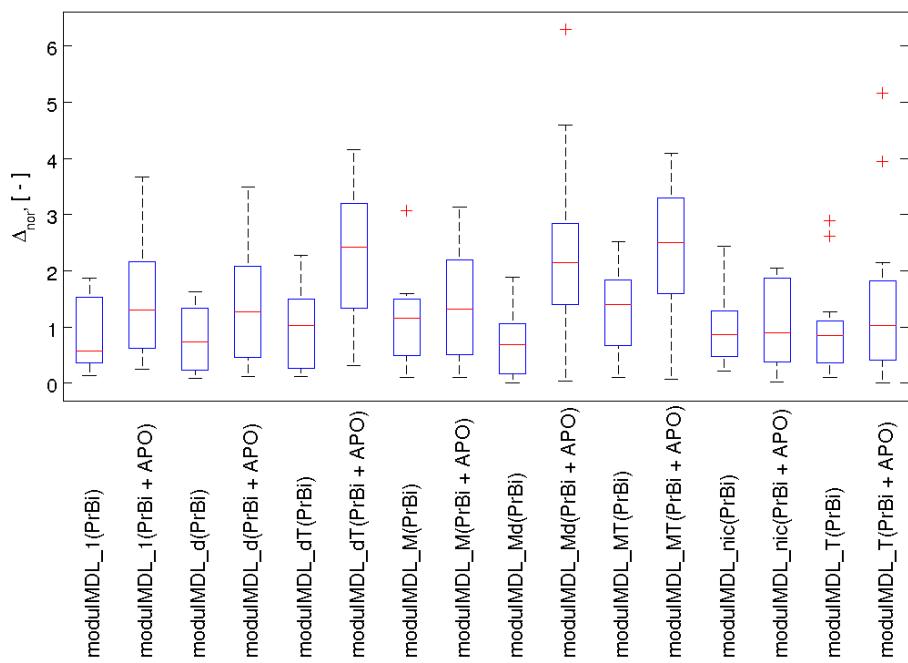
Skript	Rozsah dat	$\overline{\Delta_{abs}}$ [hmot. %]	$s^2(\Delta_{abs})$	$\overline{\Delta_{nor}}$ [-]	$s^2(\Delta_{nor})$
modulMDL_1	PrBi	0.3782	0.0047	1.3836	0.8435
modulMDL_d	PrBi	1.8447	0.0450	6.7477	8.0648
modulMDL_dT	PrBi	0.5199	0.0382	1.9018	6.8398
modulMDL_M	PrBi	0.5924	0.0149	2.1670	2.6679
modulMDL_Md	PrBi	1.3393	0.9222	4.8990	165.1089
modulMDL_MT	PrBi	0.4844	0.0234	1.7717	4.1855
modulMDL_nic	PrBi	0.3234	0.0056	1.1831	0.9951
modulMDL_T	PrBi	0.8605	0.5271	3.1477	94.3749
netMdl_1	PrBi	1.1952	4.4950	4.3721	804.8115
netMdl_d	PrBi	2.2785	20.9725	8.3348	3755.0380
netMdl_dT	PrBi	1.0728	2.4982	3.9244	447.2998
netMdl_M	PrBi	1.0357	0.2587	3.7885	46.3183
netMdl_Md	PrBi	0.9877	1.1571	3.6129	207.1736
netMdl_MT	PrBi	1.1705	0.1607	4.2818	28.7707
netMdl_nic	PrBi	1.1360	3.3603	4.1555	601.6415
netMdl_T	PrBi	0.7757	0.0493	2.8376	8.8228
modulMDL_1	PrBi+APO	0.7561	0.1278	2.7656	22.8888
modulMDL_d	PrBi+APO	0.7849	0.3039	2.8713	54.4087
modulMDL_dT	PrBi+APO	0.8977	1.0207	3.2839	182.7490
modulMDL_M	PrBi+APO	0.8634	0.5333	3.1583	95.4769
modulMDL_Md	PrBi+APO	0.7738	0.5120	2.8306	91.6635
modulMDL_MT	PrBi+APO	0.9119	0.4192	3.3359	75.0513
modulMDL_nic	PrBi+APO	0.8158	0.2316	2.9841	41.4625
modulMDL_T	PrBi+APO	0.8522	0.2499	3.1174	44.7446
netMdl_1	PrBi+APO	1.1941	1.0213	4.3680	182.8557
netMdl_d	PrBi+APO	1.3319	1.6917	4.8722	302.8959
netMdl_dT	PrBi+APO	1.2384	2.1505	4.5300	385.0388
netMdl_M	PrBi+APO	1.8467	11.1191	6.7551	1990.8233
netMdl_Md	PrBi+APO	1.4692	2.0538	5.3744	367.7182
netMdl_MT	PrBi+APO	1.3133	1.3671	4.8041	244.7699
netMdl_nic	PrBi+APO	1.4997	2.3464	5.4859	420.1195
netMdl_T	PrBi+APO	1.2611	1.9201	4.6131	343.7934

E Porovnání změny mezikvartilového rozpětí a střední hodnoty normovaných odchylek modelů

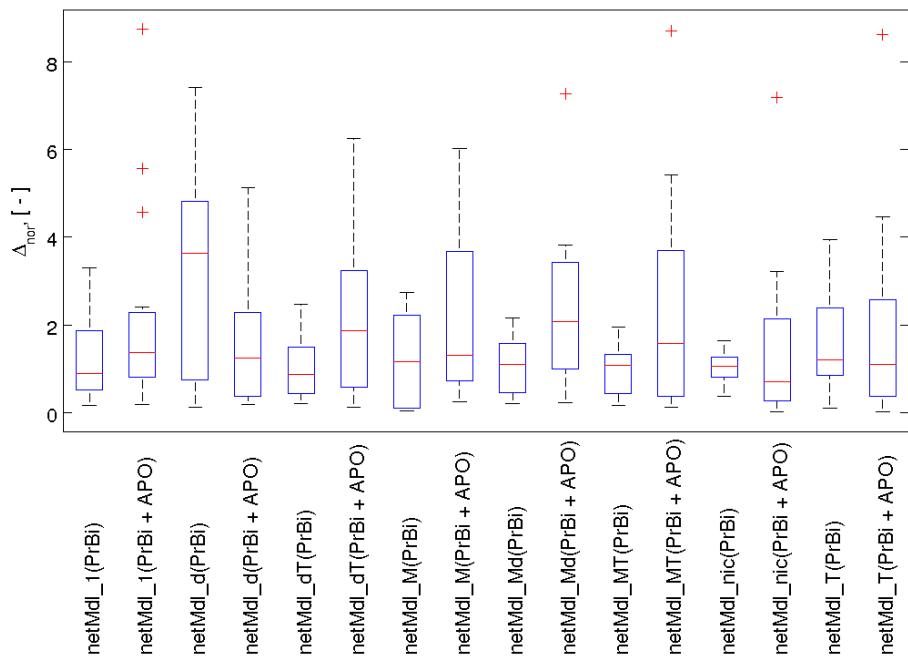


Obrázek 21 – Methan, porovnání změny střední hodnoty a mezikvartilového rozpětí normovaných odchylek při rozšíření datové sady a při použití různých doplňkových charakteristik suroviny.

Krabicovy diagram, Ethylen
zmena spolehlivosti pri doplneni datove sady

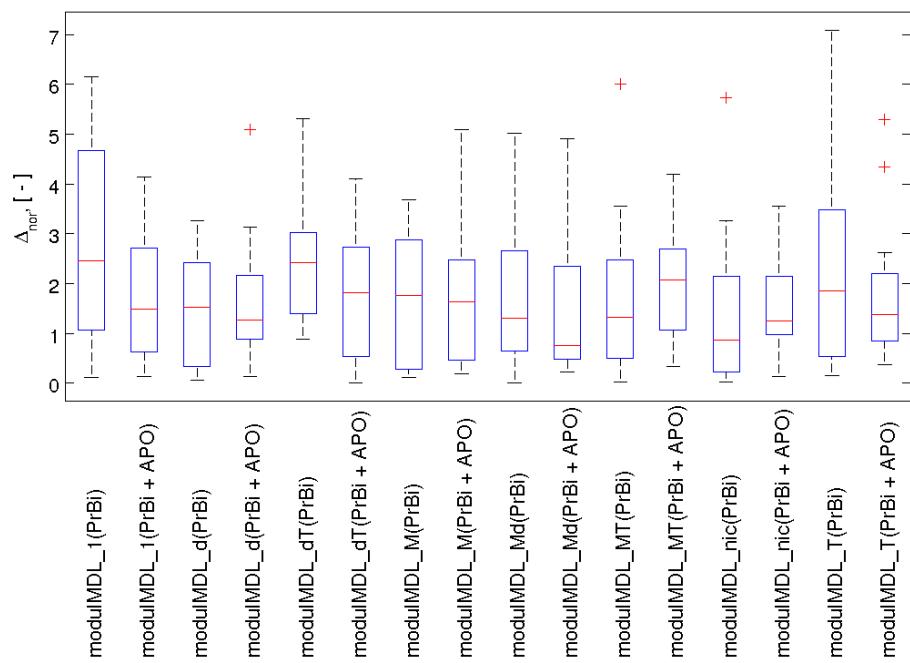


Krabicovy diagram, Ethylen
zmena spolehlivosti pri doplneni datove sady

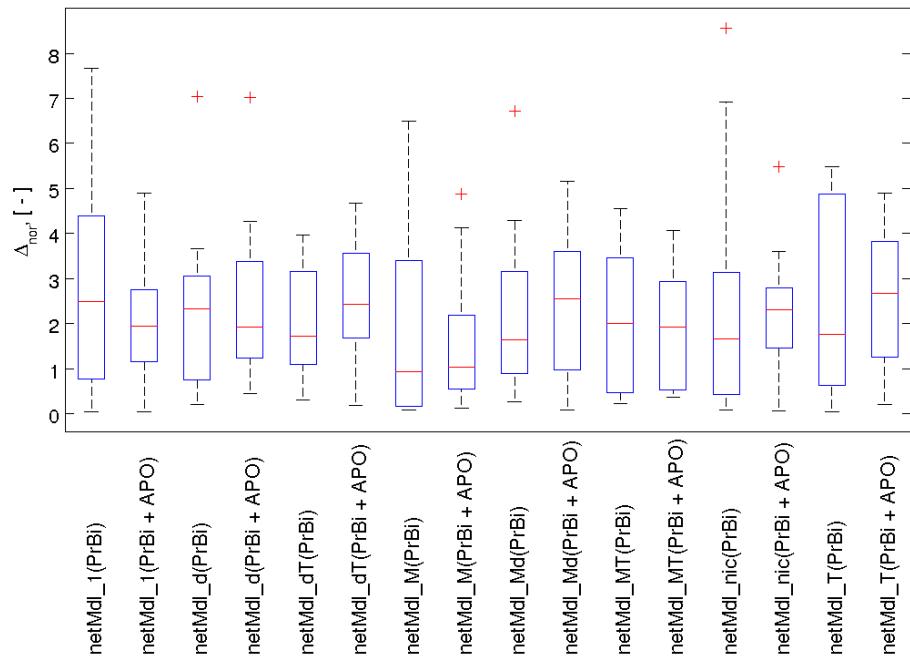


Obrázek 22 – Ethylen, porovnání změny střední hodnoty a mezikvartilového rozpětí normovaných odchylek při rozšíření datové sady a při použití různých doplňkových charakteristik suroviny.

Krabricovy diagram, Propylen
zmena spolehlivosti pri doplneni datove sady

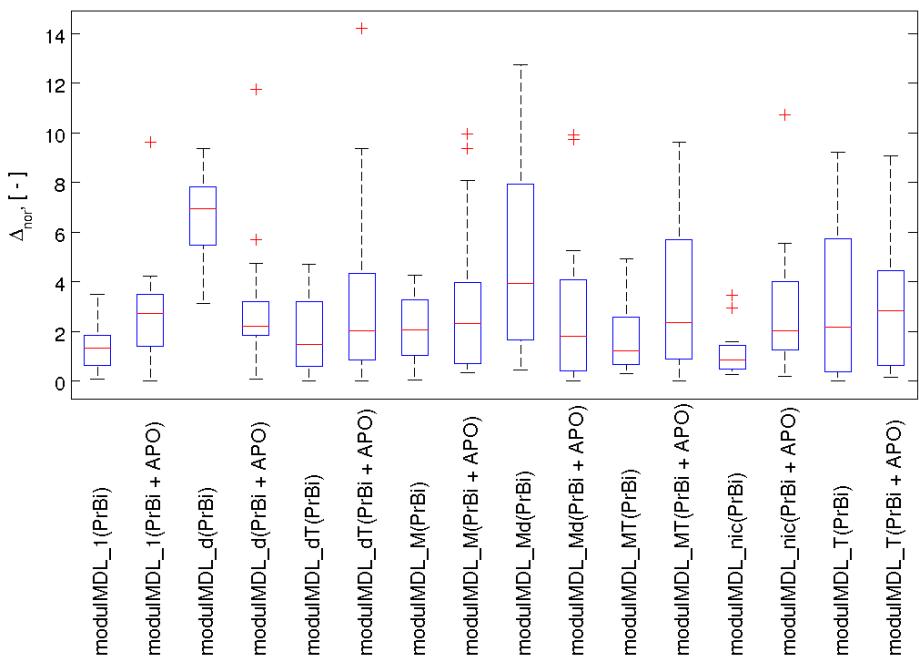


Krabricovy diagram, Propylen
zmena spolehlivosti pri doplneni datove sady

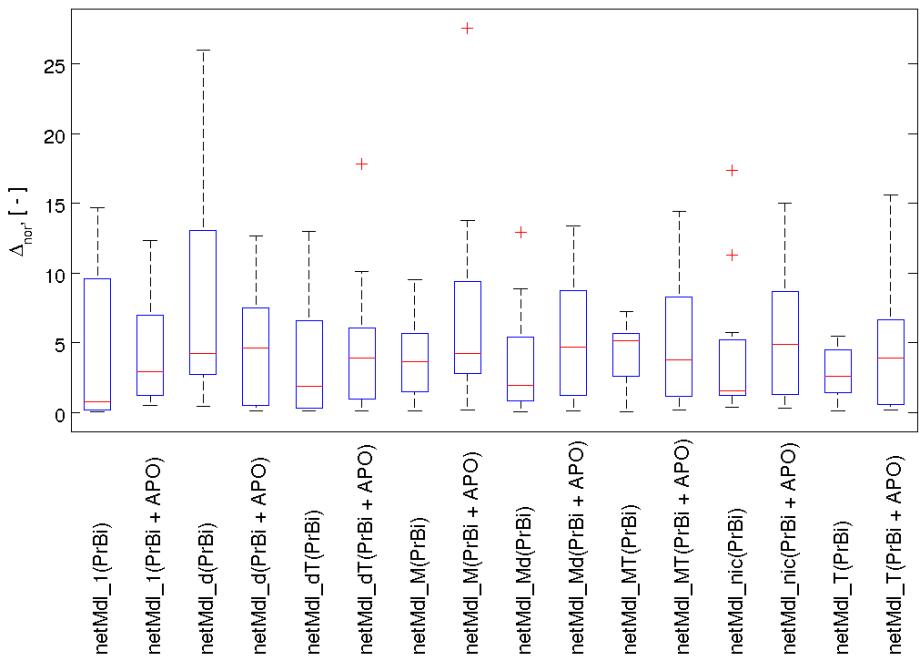


Obrázek 23 – Propylen, porovnání změny střední hodnoty a mezikvartilového rozpětí normovaných odchylek při rozšíření datové sady a při použití různých doplňkových charakteristik suroviny.

Krabicovy diagram, Bil. olej
zmena spolehlivosti pri doplneni datove sady



Krabicovy diagram, Bil. olej
zmena spolehlivosti pri doplneni datove sady



Obrázek 24 – Bilanční olej, porovnání změny střední hodnoty a mezikvartilového rozpětí normovaných odchylek při rozšíření datové sady a při použití různých doplňkových charakteristik suroviny.

F Výsledky predikce výtěžků pyrolýzních produktů modelem modulMDL_1

F.1 Učení pouze na primárních benzínech

Tabulky zde prezentované byly získány porovnáním simulovaných a naměřených dat získaných odhadováním pyrolýzních produktů primárních benzínů. Celkem bylo dostupných 29 datových sad, z nichž bylo k učení využito 18, zbylých 11 bylo použito k validaci učebního procesu a následnému porovnání úspěšnosti sítí. V tabulkách zobrazované hodnoty byly získány simulací na validační sadě dat.

Tabulka 13 – Methan, Porovnání hodnot naměřených **Tabulka 14** – Ethylen, Porovnání hodnot naměřených a dopočtených modelem modulMdl_1 a dopočtených modelem modulMdl_1

Exp. val. [hmot. %]	Sim. val. [hmot. %]	Δ_{abs} [hmot. %]	Δ_{nor} [–]	Exp. val. [hmot. %]	Sim. val. [hmot. %]	Δ_{abs} [hmot. %]	Δ_{nor} [–]
15.530000	15.790762	0.260762	1.282960	25.600000	24.757990	0.842010	2.505981
14.870000	14.723117	0.146883	0.722670	23.250000	24.039498	0.789498	2.349695
13.970000	13.869651	0.100349	0.493724	25.050000	24.623952	0.426048	1.268001
13.610000	13.327876	0.282124	1.388062	24.490000	24.593247	0.103247	0.307282
14.970000	14.535400	0.434600	2.138254	25.680000	25.861455	0.181455	0.540045
14.150000	14.164829	0.014829	0.072958	25.290000	25.410343	0.120343	0.358165
14.950000	15.226733	0.276733	1.361541	25.390000	25.704185	0.314185	0.935075
14.610000	14.950714	0.340714	1.676332	25.100000	25.543247	0.443247	1.319186
14.500000	14.117159	0.382841	1.883597	24.920000	25.064713	0.144713	0.430693
15.550000	15.670106	0.120106	0.590929	25.180000	25.096549	0.083451	0.248367
16.260000	16.346002	0.086002	0.423133	26.880000	26.945043	0.065043	0.193580
15.730000	16.410064	0.680064	3.345949	24.800000	25.276596	0.476596	1.418439

Tabulka 15 – Propylen, Porovnání hodnot naměřených a dopočtených modelem modulMdl_1

Tabulka 16 – Bilanční olej, Porovnání hodnot naměřených a dopočtených modelem modulMdl_d

Exp. val. [hmot. %]	Sim. val. [hmot. %]	Δ_{abs} [hmot. %]	Δ_{nor} [–]	Exp. val. [hmot. %]	Sim. val. [hmot. %]	Δ_{abs} [hmot. %]	Δ_{nor} [–]
14.890000	14.164650	0.725350	3.726912	3.910000	4.170606	0.260606	3.563846
13.740000	14.199343	0.459343	2.360144	5.850000	6.455565	0.605565	8.281226
15.570000	14.787542	0.782458	4.020336	2.940000	3.558870	0.618870	8.463180
15.310000	15.540483	0.230483	1.184243	3.290000	3.554501	0.264501	3.617107
13.490000	13.101483	0.388517	1.996236	4.640000	4.065943	0.574057	7.850347
14.640000	14.488306	0.151694	0.779414	3.500000	3.685016	0.185016	2.530139
15.050000	15.311280	0.261280	1.342478	3.630000	3.483372	0.146628	2.005173
14.570000	14.777553	0.207553	1.066427	3.620000	3.559126	0.060874	0.832459
14.230000	14.957441	0.727441	3.737655	4.320000	3.551914	0.768086	10.503741
14.360000	14.547329	0.187329	0.962511	3.710000	4.071020	0.361020	4.937023
14.100000	13.951764	0.148236	0.761648	4.310000	4.677954	0.367954	5.031844
13.060000	12.687802	0.372198	1.912388	4.660000	4.055274	0.604726	8.269752

F.2 Učení na všech dostupných datech

Tabulky zde prezentované byly získány porovnáním simulovaných a naměřených dat získaných odhadováním pyrolýzních produktů primárních benzínů a atmosférického plynového oleje. Celkem bylo dostupných 39 datových sad (29 + 10), z nichž bylo k učení využito 23, zbylých 16 bylo použito k validaci učebního procesu a následnému porovnání úspěšnosti sítí. V tabulkách zobrazované hodnoty byly získány simulací na validační sadě dat.

Tabulka 17 – Methan, Porovnání hodnot naměřených a dopočtených modelem modulMdl_1

Tabulka 18 – Ethylen, Porovnání hodnot naměřených a dopočtených modelem modulMdl_1

Exp. val. [hmot. %]	Sim. val. [hmot. %]	Δ_{abs} [hmot. %]	Δ_{nor} [-]	Exp. val. [hmot. %]	Sim. val. [hmot. %]	Δ_{abs} [hmot. %]	Δ_{nor} [-]
15.530000	15.868812	0.338812	1.666970	25.600000	24.662390	0.937610	2.790507
14.870000	15.326607	0.456607	2.246531	23.250000	23.997547	0.747547	2.224843
13.970000	14.250540	0.280540	1.380271	25.050000	25.230306	0.180306	0.536626
13.610000	13.504350	0.105650	0.519805	24.490000	24.154895	0.335105	0.997336
14.970000	14.983904	0.013904	0.068410	25.680000	25.521853	0.158147	0.470676
14.150000	14.214960	0.064960	0.319604	25.290000	25.554898	0.264898	0.788386
14.950000	14.963716	0.013716	0.067485	25.390000	25.337084	0.052916	0.157489
14.610000	14.610180	0.000180	0.000884	25.100000	25.322566	0.222566	0.662398
14.500000	14.373015	0.126985	0.624773	24.920000	25.204547	0.284547	0.846867
15.550000	15.613514	0.063514	0.312492	25.180000	25.485671	0.305671	0.909734
16.260000	16.259755	0.000245	0.001208	26.880000	27.150840	0.270840	0.806071
15.730000	16.101638	0.371638	1.828476	24.800000	25.729195	0.929195	2.765461
10.010000	10.101611	0.091611	0.450731	19.660000	19.872916	0.212916	0.633679
10.010000	9.964365	0.045635	0.224526	20.460000	20.102329	0.357671	1.064496
10.310000	9.812141	0.497859	2.449493	20.410000	20.152094	0.257906	0.767577
10.630000	10.556651	0.073349	0.360881	19.250000	20.429710	1.179710	3.511043

Tabulka 19 – Propylen, Porovnání hodnot naměřených a dopočtených modelem modulMdl_1

Tabulka 20 – Bilanční olej, Porovnání hodnot naměřených a dopočtených modelem modulMdl_d

Exp. val. [hmot. %]	Sim. val. [hmot. %]	Δ_{abs} [hmot. %]	Δ_{nor} [-]	Exp. val. [hmot. %]	Sim. val. [hmot. %]	Δ_{abs} [hmot. %]	Δ_{nor} [-]
14.890000	14.192104	0.697896	3.585849	3.910000	4.858342	0.948342	3.469015
13.740000	14.146613	0.406613	2.089213	5.850000	5.242871	0.607129	2.220866
15.570000	15.662998	0.092998	0.477830	2.940000	3.461279	0.521279	1.906829
15.310000	15.565786	0.255786	1.314251	3.290000	3.764022	0.474022	1.733963
13.490000	13.918304	0.428304	2.200661	4.640000	5.683103	1.043103	3.815649
14.640000	14.740383	0.100383	0.515775	3.500000	3.321398	0.178602	0.653324
15.050000	14.890293	0.159707	0.820590	3.630000	3.098278	0.531722	1.945026
14.570000	14.719363	0.149363	0.767441	3.620000	3.299722	0.320278	1.171572
14.230000	14.856936	0.626936	3.221250	4.320000	3.215042	1.104958	4.041912
14.360000	13.883678	0.476322	2.447382	3.710000	4.960964	1.250964	4.576000
14.100000	14.093916	0.006084	0.031260	4.310000	4.307755	0.002245	0.008214
13.060000	13.057351	0.002649	0.013610	4.660000	6.093758	1.433758	5.244657
12.260000	12.669630	0.409630	2.104717	20.530000	19.577859	0.952141	3.482911
12.510000	12.775122	0.265122	1.362220	18.310000	17.320436	0.989564	3.619804
12.440000	12.633761	0.193761	0.995561	18.430000	19.816555	1.386555	5.071990
12.310000	12.433048	0.123048	0.632234	21.870000	22.824988	0.954988	3.493327