

**Mendelova univerzita v Brně**

Lesnická a dřevařská fakulta

Ústav hospodářské úpravy lesů a aplikované geoinformatiky



Lesnická  
a dřevařská  
fakulta

**Využití moderních regresních metod pro modelování  
výškové křivky**

Disertační práce

2014/2015

Ing. Zdeněk Adamec

## Čestné prohlášení

Prohlašuji, že jsem tuto práci: „Využití moderních regresních metod pro modelování výškové křivky“ vypracoval samostatně a veškeré použité prameny a informace uvádím v seznamu použité literatury. Souhlasím, aby moje práce byla zveřejněna v souladu s § 47b zákona č. 111/1998 Sb., o vysokých školách ve znění pozdějších předpisů a v souladu s platnou Směrnicí o zveřejňování vysokoškolských závěrečných prací.

Jsem si vědom, že se na moji práci vztahuje zákon č. 121/2000 Sb., autorský zákon, a že Mendelova univerzita v Brně má právo na uzavření licenční smlouvy a užití této práce jako školního díla podle § 60 odst. 1 autorského zákona.

Dále se zavazuji, že před sepsáním licenční smlouvy o využití díla jinou osobou (subjektem) si vyžádám písemné stanovisko univerzity, že předmětná licenční smlouva není v rozporu s oprávněnými zájmy univerzity, a zavazuji se uhradit případný příspěvek na úhradu nákladů spojených se vznikem díla, a to až do jejich skutečné výše.

V Brně dne:

.....

podpis

## Poděkování

Touto cestou bych rád poděkoval mému školiteli doc. Ing. Karlu Drápelovi CSc., který mi věnoval spoustu cenných rad a zkušeností v průběhu celého doktorského studia a vedl mé kroky při psaní této práce vždy správným směrem. Dále bych chtěl poděkovat své přítelkyni Aleně Kučerové za velkou morální podporu a toleranci, kterou mi projevovala, když jsem věnoval čas této práci místo jí. V neposlední řadě bych rád poděkoval mojí rodině za trpělivost a shovívavost, kterou mi věnovali, když jsem psal tuto práci.

**Autor:** Zdeněk Adamec  
**Název práce:** Využití moderních regresních metod pro modelování výškové křivky  
**Author:** Zdeněk Adamec  
**Title of thesis:** The utilization of modern regression methods for modeling tree height-diameter relationship

**Abstrakt:**

Tato práce se zabývá využitím tří různých typů moderních regresních metod (lineární model se smíšenými efekty, zobecněný aditivní model a klasifikační a regresní stromy) pro modelování výškové křivky. Z výsledků vyplývá, že všechny tři použité metody jsou za určitých podmínek použitelné jako alternativní přístup pro modelování závislosti výšky stromu a jeho výčetní tloušťky. Smíšený model fungoval nejlépe, pokud měl náhodnou část odhadu parametru u absolutního i regresního parametru a obsahoval věk porostu (nebo jeho přirozený logaritmus) jako prediktor druhé úrovně. Zobecněný aditivní model vycházel nejlépe, pokud byla využita vyhlazovací funkce typu LOESS. Regresní strom byl schopen kvalitní predikce výšek v případě, že byly využity jako prediktory výčetní tloušťka, střední tloušťka, věk porostu a absolutní výšková bonita.

**Klíčová slova:** lineární model se smíšenými efekty, zobecněný aditivní model, klasifikační a regresní stromy, kalibrace, predikce

**Abstract:**

The topic of this thesis is utilization of three types of modern regression methods (linear mixed effects model, generalized additive model and classification and regression trees) to model height-diameter relationship. Results revealed that all three regression methods can be, under certain conditions, applicable as an alternative approach for height-diameter relationship modeling. The mixed effects model performed best, when random effects of the intercept and regression coefficient were taken into account along with the age as a second order predictor (or its natural logarithm). The generalized additive model performed best when LOESS smoothing function was applied. The regression tree method prediction was optimal when diameter at breast height, mean height, age of a forest stand and site index were included to the model as predictors.

**Keywords:** linear mixed effects model, generalized additive model, classification and regression trees, calibration, prediction

# Obsah

1	Úvod .....	7
1.1	Vývoj regresních modelů v lesnictví.....	7
1.2	Problémy s lesnickými daty .....	8
1.3	Význam modelu výškové křivky .....	11
2	Cíl práce.....	12
3	Současný stav řešené problematiky.....	14
3.1	Využití regrese v lesnictví v posledních 20 letech.....	14
3.2	Výškové křivky a jejich modely .....	16
3.2.1	Stadiální výšková křivka .....	17
3.2.2	Vývojová výšková křivka .....	20
3.2.3	Stacionární výšková křivka výběrného lesa.....	21
3.2.4	Modely výškové křivky .....	21
3.3	Smíšené regresní modely .....	24
3.4	Zobecněné aditivní modely .....	29
3.5	Klasifikační a regresní stromy .....	33
4	Metodika.....	37
4.1	Přírodní podmínky zájmového území .....	37
4.2	Terénní práce.....	38
4.3	Statistické zpracování dat.....	39
4.3.1	Výpočet běžných dendrometrických charakteristik.....	40
4.3.2	Výpočet základních statistických charakteristik.....	40
4.3.3	Průzkumová analýza dat .....	40
4.3.4	Výpočet globálního modelu výškové křivky .....	41
4.3.5	Výpočet lokálního modelu výškové křivky .....	42
4.3.6	Volba modelu .....	42
4.3.7	Výpočet smíšeného modelu výškové křivky .....	43
4.3.7.1	Tvorba smíšeného modelu .....	43
4.3.7.2	Kalibrace smíšeného modelu výškové křivky .....	47
4.3.8	Výpočet zobecněného aditivního modelu.....	51
4.3.9	Postup tvorby regresního stromu .....	55
5	Výsledky.....	57
5.1	Dendrometrické charakteristiky .....	57
5.2	Základní statistické charakteristiky.....	57
5.3	Výsledky průzkumové analýzy dat .....	57
5.4	Globální model.....	58
5.5	Lokální modely .....	59

5.6	Výběr vhodného modelu .....	60
5.7	Linearizace Pettersonovy výškové funkce .....	61
5.8	Smíšený regresní model .....	62
5.8.1	Výsledný smíšený model výškové křivky .....	62
5.8.2	Výsledky kalibrace smíšeného modelu výškové křivky .....	76
5.8.2.1	Výsledky kalibrace s výběrem stromů se střední tloušťkou .....	76
5.8.2.2	Výsledek kalibrace s výběrem stromů ve třech tloušťkových intervalech .....	78
5.9	Zobecněný aditivní model .....	82
5.9.1	Globální modely .....	82
5.9.2	Lokální modely .....	86
5.9.3	Souhrnné zhodnocení GAM modelů .....	89
5.10	Klasifikační a regresní stromy .....	90
6	Diskuze .....	95
6.1	Využití regresních metod .....	95
6.2	Model výškové křivky .....	95
6.3	Smíšený model .....	97
6.3.1	Kalibrace smíšeného modelu .....	98
6.4	GAM model .....	100
6.5	CART model .....	102
7	Praktická využitelnost hodnocených metod .....	104
7.1	Kalibrovaný smíšený model .....	104
7.2	Zobecněný aditivní model .....	105
7.3	Regresní strom .....	105
7.4	Celkové zhodnocení .....	107
8	Závěr .....	108
9	Summary .....	111
10	Seznam citované literatury .....	114
11	Seznam obrázků .....	132
12	Seznam tabulek .....	134
13	Seznam zkratk .....	135
14	Seznam příloh .....	139

# 1 Úvod

## 1.1 Vývoj regresních modelů v lesnictví

Od počátku rozvoje lesnických disciplín v 18. století jakožto samostatných vědních oborů se snaží odborníci příčinné vztahy, které se v lese vyskytují, zachytit ve zjednodušeném pojetí pomocí modelů. Jelikož se v počátcích sledoval les především z produkčního hlediska, tak je pochopitelné, že jako vůbec první lesnický model byly vytvářeny objemové tabulky. Velmi brzy se také objevily požadavky na predikci vývoje porostů, což dalo vzniknout růstovým tabulkám. Vůbec první ucelené růstové tabulky sestavil Paulsen (1795). Lesnické modelování má tedy více než 200 letou historii. I když první pokusy o sestavení různých modelů byly víceméně postaveny spíše na praktických zkušenostech a jen omezených teoretických znalostech, tak postupem času docházelo k rozvoji lesnických modelů, které byly vytvářeny na základě rozsáhlého datového materiálu zpracovaného již lépe teoreticky podloženými statistickými a biometrickými metodami. Jednou z těchto metod byla i regresní analýza. Tato vazba je naprosto přirozená, protože s vědními obory matematiky a statistiky (z nichž regrese pochází) jsou velmi úzce spjaty obory biometrie a lesnická dendrometrie, jejíž nedílnou součástí je právě i lesnické modelování. Regresní analýza se zabývá vytvořením vhodného matematického modelu studované závislosti, stanoví parametry tohoto modelu a ověřuje hypotézy o statistické významnosti a dalších důležitých vlastnostech modelu (Drápela, 2011). Za objevitele regresní analýzy je považován sir Francis Galton (1822 – 1911), který poprvé použil pojem regrese při jeho antropologických pokusech v druhé polovině 19. století. Je tedy zřejmé, že pojem regrese se objevil až poté, co se v lesnictví používala první generace růstových tabulek. Dokonce modely růstových funkcí z první poloviny 19. století (např. Hossfeld, 1822; Gompertz, 1825) lze z dnešního pohledu označit za regresní modely, i když v době jejich vzniku tento pojem ještě ani neexistoval. V průběhu 20. století již vznikaly modely, které již nebyly postaveny pouze na empirickém základě, ale byly také dobře biometricky formulované a teoreticky odvozené. Za jednoho z prvních průkopníků těchto modelů můžeme podle Fabriky a Pretzsche (2011) považovat Ernsta Assmanna, který společně s Friedrichem Franzem sestavil růstové tabulky (Assmann a Franz 1963), které splňovaly všechny výše uvedené podmínky. Porté a Bartelink (2002) uvádějí jako jeden z rozhodujících faktorů ovlivňujících vývoj lesnických modelů lepší dostupnost výkonné výpočetní techniky vybavené odpovídajícím softwarovým vybavením (včetně vizualizačních nástrojů),

což umožnilo využití výpočetně náročnějších matematických a statistických metod s vysokou vypovídací schopností. Teoretický rozvoj regresní analýzy je dnes velmi rychlý a neustále vznikají další varianty regresních modelů, eventuálně se rozšiřují možnosti těch stávajících. Proto je možné souhlasit s tvrzením Gregoirea et al. (1995), který konstatoval, že regresní analýza je dnes nejběžněji používaná statistická metoda v lesnickém modelování. Vývoj regresních metod doznal za posledních 20 let takového rozmachu (a to právě především díky rychlému vývoji výpočetní techniky), že se lze s některou variantou regresního modelu setkat v široké škále lesnického výzkumu i praxe. V dnešní době se v lesnictví můžeme setkat s regresním modelem především v empirických modelech růstu a vývoje lesa. Jsou také základní součástí růstových simulátorů dnes velmi často používaných v lesnickém výzkumu a praxi. Podle Fabriky a Pretzsche (2011) je simulátor počítačový model reálného systému, který se využívá pro účely simulace. Růstový simulátor může být definován jako soubor dílčích modelů, které společně simulují chování lesa po celou dobu jeho vývojového cyklu – tedy od obnovy porostu přes růst stromů až po těžbu. Většina dílčích modelů je založena právě na různých typech regresních technik.

## 1.2 Problémy s lesnickými daty

Využití regresních metod v lesnictví má však stále svoje omezení, a to díky problémům, které vycházejí právě z povahy dat. Podle Zhanga et al. (2005) se v lesnictví stále velmi často používá modelování pomocí lineární a nelineární regrese. V těchto typech regresí se pro odhad parametrů využívá metoda nejmenších čtverců (Calama a Montero 2005). Nicméně jeden z předpokladů použití této metody, konkrétně nezávislost mezi pozorováními, je velmi často porušena a to z důvodu časové nebo prostorové závislosti mezi daty, která má za důsledek vychýlené odhady středních chyb parametrů modelu, vychýlené odhady intervalů spolehlivosti parametrů modelu a zároveň zavádějící výsledky testů významnosti (Anselin a Griffith 1988; Searle et al. 1992; Fox et al. 2001). Vzájemná závislost je důsledkem hierarchické struktury dat, protože data pocházející ze stejných inventarizačních jednotek (strom, plocha, oblast), si jsou podobnější oproti datům z jiných jednotek (West, 1981; West et al. 1984; Neter et al. 1990; Fox et al. 2001). Pokud shrneme výše uvedené, můžeme konstatovat, že existují tři různé zdroje variability, což odpovídá tvrzení Drápely (2011), který definoval základní zdroje variability takto:



1. prostorová variabilita (konkrétní tvar závislosti se mění místo od místa v závislosti na různých faktorech, především na podmínkách prostředí),
2. časová variabilita (tvar funkce se mění s časem, typické např. pro výškové funkce),
3. hierarchická struktura dat (data jsou měřena na určitých jednotkách, které jsou součástí vyšších jednotek – např. strom – zkusná plocha nebo porost – lesní oblast, apod.). Vnoření jednotlivých úrovní měření vede k tomu, že data jsou vzájemně korelována.

Právě hierarchická struktura dat je problémem většiny lesnických využitelných modelů (např. model výškové funkce, růstové funkce, závislosti růstových veličin na faktorech prostředí a klimatu). U problému prostorové variability je ale nutné upozornit na tvrzení Legendreho (1993), který poukázal na to, že prostorová variabilita není jen pouhým statistickým problémem, ale především je důležitou vlastností ekosystému. Odhad prostorové heterogenity je nezbytný pro pochopení komplexnosti, struktury a fungování ekosystému jako takového (Zhang et al. 2008). Proto je nutné věnovat vysvětlení co největší části prostorové variability v modelu mimořádnou pozornost. Co se týče časové variability, tak víme, že mezi opakovanými měřeními na stejných jednotkách existuje silná závislost. Použijeme-li klasické regresní metody na takto časově závislá data, dochází k desinterpretaci modelu. V takovýchto případech se ve výsledcích často vyskytuje falešná statistická závislost mezi závislou proměnnou a některou z nezávisle proměnných, zapříčiněná právě časovou korelací mezi daty (Hanke a Wichern 2005). Donedávna byly tyto problémy často opomíjeny a celý datový soubor se pro účely analýzy považoval za homogenní a modelované vztahy byly vyjádřeny pouze jedním modelem, který můžeme označit jako globální, nebo se pro každou zkoumanou jednotku vytvářel samostatný model, který lze označit jako lokální. Podle Drápely (2011) je možné shrnout nevýhody obou modelů takto:

- globální model ignoruje skutečnou prostorovou (event. časovou) strukturu dat a všechna data považuje za vzájemně nezávislá a pracuje se všemi daty najednou – což způsobuje zdánlivě vysokou variabilitu dat (neboť všechny výše uvedené zdroje variability jsou zahrnuty do náhodné složky modelu) a výsledkem jsou široké intervalové odhady parametrů,
- v případě lokálního modelu platí parametry modelu jen pro experimentální jednotku (např. zkusnou plochu), pro kterou byly spočítány, resp. pro velmi omezený základní soubor (např. pro porost, ve kterém byla daná zkusná plocha

vytvořena), což znamená obtížnou možnost zobecnění (tyto modely se nedají použít např. pro nově měřené plochy, nemůžeme modelovat variabilitu mezi experimentálními jednotkami apod.) a navíc v případě velkého počtu experimentálních jednotek je nutné stanovit velké množství modelů s mnoha parametry, čemuž předchází nutnost velkého množství měření.

Odstranění nevýhod obou modelů bylo v minulosti řešeno různými způsoby (Calama a Montero 2004):

- dvoustupňovým modelem (Ferguson a Leech 1978), kdy v první fázi řešení byl daný model proveden pro každou experimentální jednotku (plochu) a v druhé fázi byly parametry modelů použity jako vysvětlující proměnné do modelů jejich závislosti na dalších, většinou porostních veličinách (počet stromů, hektarová kruhová výčetní základna, horní výška, střední tloušťka, objem středního kmene apod.), např. Krumland a Wensel (1988); Parresol (1992),
- odvozením příslušných modelů pro jednotlivé geografické oblasti nebo ekologicky definované oblasti např. Fulton (1999), Huang et al. (2000),
- využitím geostatistických metod např. Zhang et al. (2004).

Ale ani tyto metody nejsou v některých případech ideálním řešením výše uvedených problémů, a proto stále existuje snaha vytvořit modely, které by lépe dokázaly zpracovat datový materiál s komplikovanou strukturou. Tato potřeba se projevuje stále více, což lze přisuzovat tomu, že se rozvinuly nové metody sběru dat (letecké a družicové snímkování, laserové skenování apod.), které poskytují velké datové sady, jež vykazují prvky časové nebo prostorové závislosti a klasické regresní metody je neumí zpracovat. V současné době ale existují nové regresní techniky, které se umí s výše uvedenými problémy v lesnických datech velmi dobře vypořádat. Právě možnost vyřešit např. problém prostorové závislosti a heterogenity dat vede k popularizaci a stále masivnějšímu využívání širšího spektra regresních metod. Vzhledem k tomu, že vlastnosti těchto regresních technik a možnosti jejich využití v lesnictví jsou stále předmětem intenzivního výzkumu, je tato práce zaměřena na analýzu tří různých regresních metod s důrazem na řešení výše uvedených problémů v datovém materiálu. Jako příklad možné praktické aplikace bylo zvoleno modelování výškové křivky.

### 1.3 Význam modelu výškové křivky

Model výškové křivky je jedním z nejběžnějších lesnických modelů, ve kterém se nejčastěji setkáváme s výše uvedenými problémy. Jedná se o jeden z nejpoužívanějších a také nejstarších lesnických modelů. Důvody jeho vzniku a častého používání jsou velmi logické. Jelikož první modelovací pokusy směřovaly ke zjištění produkce porostu (tedy jeho zásoby) bylo nutné znát výšky a výčetní tloušťky jednotlivých stromů pro určení jejich objemu. Na rozdíl od měření tlouštěk bylo a stále je měření výšek časově náročnější (a tedy i dražší). To vedlo k tomu, že začaly být vytvářeny první modely závislosti výšky na výčetní tloušťce stromu, aby se část měřených výšek mohla nahradit výškami modelovými. Za jedny z prvních významných prací zabývajících se modelem výškové křivky lze označit např. práce Levakoviče (1935) a Wiedemanna (1936), kteří jako první publikovali regresní model vztahu výšky stromu v závislosti na jeho výčetní tloušťce. Vývoj modelu výškové funkce se neustále vyvíjí a od původních modelů, které byly například řešeny velmi jednoduchou lineární regresí, přes upravené růstové funkce a specializované nelineární výškové funkce jsme se dostali až po nejmodernější techniky, které jsou řešeny v této práci. Společným cílem všech regresních přístupů modelování výšek použitých v této práci je snaha o co nejkvalitnější predikci výšek při měření co nejmenšího množství výšek za použití co nejefektivnější metody, která bude schopna řešit problémy závislosti a heterogenity v datech.

## 2 Cíl práce

Hlavním cílem práce je ověření praktické použitelnosti vybraných regresních metod pro modelování výškových funkcí v lesnictví. Na základě praktických zkušeností a studia dostupné literatury byly jako nejperspektivnější pro tento účel zvoleny následující tři regresní modelovací techniky:

### 1. Model se smíšenými efekty

Prvním dílčím cílem je vytvoření smíšeného lineárního regresního modelu výškové křivky, který bude založen mimo prediktor první úrovně (výčetní tloušťka stromu) také na prediktorech druhého řádu (různé porostní taxační veličiny). Smíšený model (mixed effects model - MEM) bude vytvořen pro jednu výškovou funkci vybranou na základě literární rešerše a analýzy experimentálních dat. Součástí tvorby smíšeného modelu bude i tzv. kalibrovaná predikce. Kalibrovaná predikce má sloužit k tomu, aby vytvořený model predikoval výšky jednotlivých stromů s co nejmenší chybou při požadované spolehlivosti, na základě co nejmenšího počtu měřených výšek v porostu. Cílem kalibrace je zjistit, kolik stromů a s jakou tloušťkovou strukturou je nutné změřit pro dosažení požadované přesnosti predikce. Výsledný kalibrovaný model by měl tedy dosahovat pokud možno podobné přesnosti jako lokální model, ale měl by vyžadovat co nejmenší počet měřených stromů. V důsledku takto vytvořeného modelu je tedy možné snížit časovou a tedy i finanční náročnost terénního sběru dat – v tomto případě měření výšek stromů.

### 2. Zobecněný aditivní model

Druhým dílčím cílem je ověření použitelnosti zobecněného aditivního modelu (generalized additive model - GAM) výškové křivky. Důvodem pro použití tohoto modelu je obecný předpoklad, že tento typ modelu je velmi vhodný především pro modelování nelineárních vztahů, které se v lesnictví velmi často vyskytují a vztah mezi výškou stromu a výčetní tloušťkou stromu je typickým představitelem této skupiny vztahů. Budou vyzkoušeny různé typy vyhlazovacích funkcí s ohledem na praktické použití vytvořených modelů. Takto zkonstruované modely budou porovnány s globálními modely (platnými pro celé zájmové území) i s lokálními modely (jednotlivé porostní skupiny) výškových funkcí, které budou vypočítány běžnou nelineární regresí. Cílem je tedy odpovědět na otázku, jestli některý z typů GAM modelů je vhodný pro modelování nelineárního vztahu mezi výškou stromu a výčetní tloušťkou stromu.

### 3. Klasifikační a regresní stromy

Třetím dílčím cílem je využití neparametrické metody klasifikačních a regresních stromů (classification and regression trees - CART) pro popsání a vysvětlení závislosti výšky stromu na jeho výčetní tloušťce a různých porostních veličinách. Důvodem pro sestavení takového stromu je otestování, jestli může být vytvořen algoritmus, který by nebyl obecně formulovatelný pomocí matematicko-funkčně definovaného modelu, ale přesto poskytoval adekvátní výsledky. Cílem bude ověřit, jestli je možné vytvořit regresní strom, který by predikoval výšku stromu s požadovanou přesností pouze za znalosti výčetní tloušťky a některých z porostních veličin, ale bez použití parametrického modelu výškové křivky, to znamená, jestli by výčetní tloušťka a libovolné porostní veličiny mohly sloužit jako klasifikátory pro určení výšky stromu a nahradit tak klasickou výškovou křivku. I když je výsledek této techniky také nazýván modelem, tak se jedná ale o model neparametrický a lze se setkat také s označením technika bez tvorby modelu.

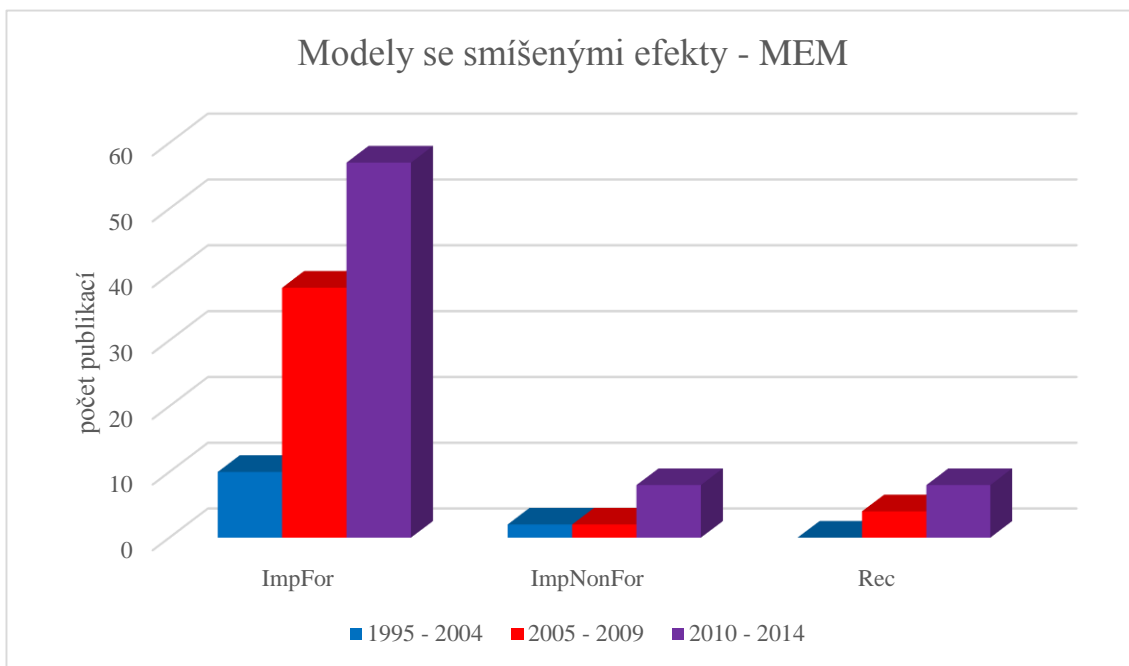
## 3 Současný stav řešené problematiky

### 3.1 Využití regrese v lesnictví v posledních 20 letech

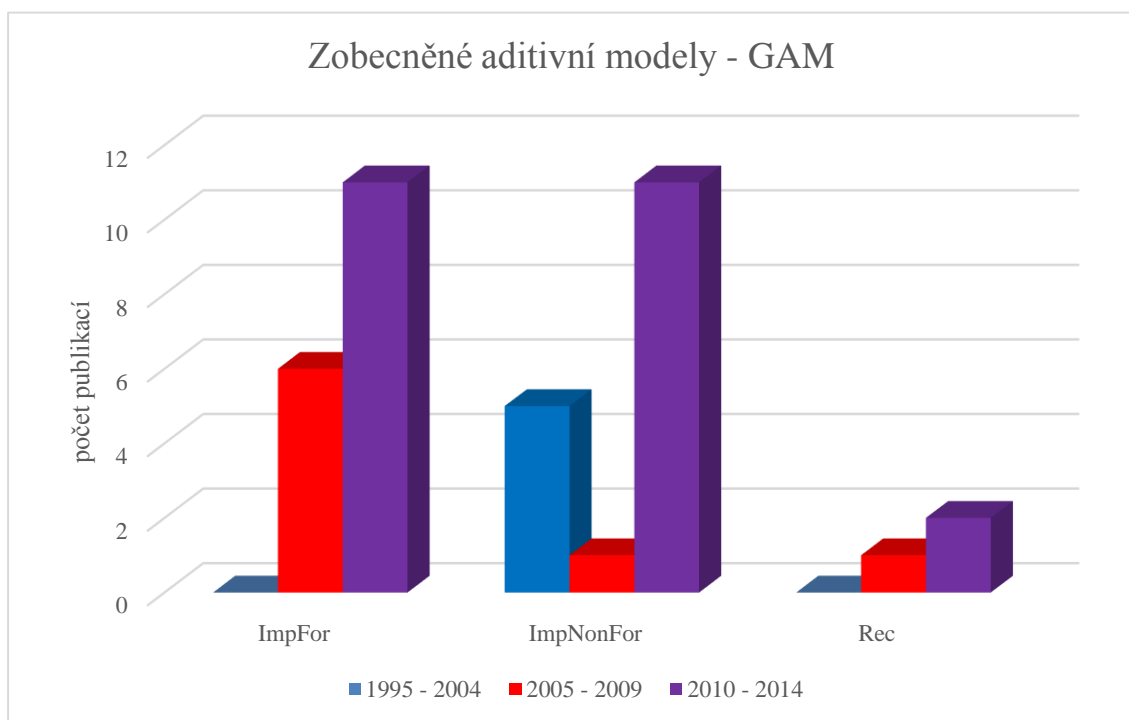
Jako kritéria využití tří regresních metod bylo použito počtu plně recenzovaných nebo impaktových článků dohledatelných ve vědeckých databázích SCOPUS nebo Web of Science za posledních 20 let rozdělených do tří kratších časových úseků. Jednalo se o období mezi lety 1995 – 2004, 2005 – 2009 a 2010 – 2014. Byly hodnoceny pouze články, které se tematicky vázaly na lesní hospodářství nebo lesnický výzkum. Jako klíčová slova byla použita: mixed effects model, generalized additive model, classification and regression trees, forest, forestry, forest management. Takto nalezené články byly členěny do tří kategorií:

- impaktové v kategorii Forestry dle databáze Web of Science (ImpFor),
- impaktové mimo kategorii Forestry dle databáze Web of Science (ImpNonFor),
- ostatní neimpaktové dohledatelné v databázi SCOPUS (Rec).

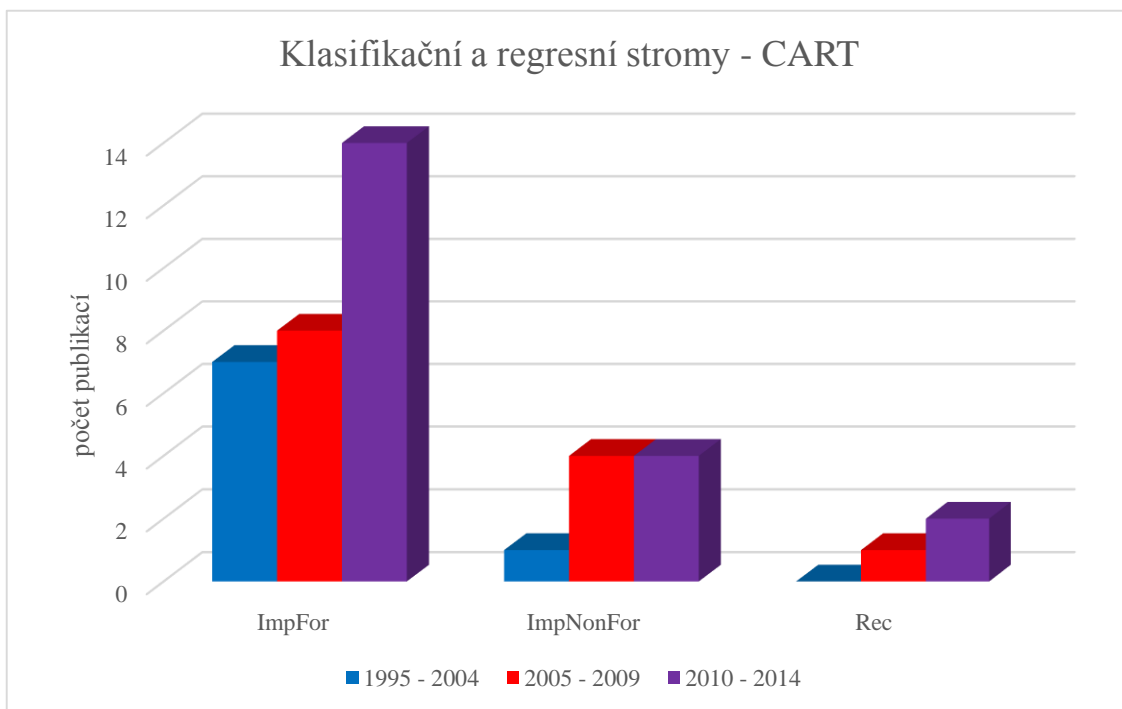
Z provedené analýzy (graficky znázorněno na obrázcích č. 1 - 4) je dobře patrné, že u všech tří metod se zvyšuje počet publikací za období mezi roky 1995 – 2014. Ve všech třech kategoriích publikací u všech metod je vidět, že nejvíce publikací bylo vydáno v posledních pěti letech, což potvrzuje zvyšující se význam těchto metod a je tedy vcelku opodstatněné, že se tyto metody nazývají v této práci jako moderní metody. Je také dobře patrné, že jednoznačně převažuje počet publikací v impaktových časopisech (bez rozdílu kategorie) nad publikacemi pouze recenzovanými. Nejvíce publikací využívá metodu modelů se smíšenými efekty bez rozdílu, jestli se jedná o modely lineární, nelineární nebo zobecněné. I když jsou všechny tři metody teoreticky známy již více než 20 let, tak jejich prudký rozmach nastal až v posledních 10 letech, což je mimo jiné jistě zapříčiněno i rychlým rozvojem výpočetních technologií a vhodných programů. Z pohledu dalšího vývoje bude zajímavé sledovat, jestli i nadále bude postupovat zvyšující se trend u všech tří metod, což by znamenalo, že se tyto metody stanou běžně užívanými metodami, které jsou akceptované především z důvodu schopnosti řešit problémy s lesnickými daty uvedenými v kapitole 1.2 nebo jestli se tento trend zastaví na úkor jednodušších lineárních a nelineárních modelů. Také bude zajímavé sledovat srovnání vzájemného poměru využívání parametrických (MEM) a neparametrických (GAM, CART) metod, protože v současnosti parametrické metody dominují, což může vyvolat domněnku, že i přes svoji vyšší výpočtovou náročnost oproti neparametrickým metodám poskytují lepší výsledky v oblasti lesnického modelování.



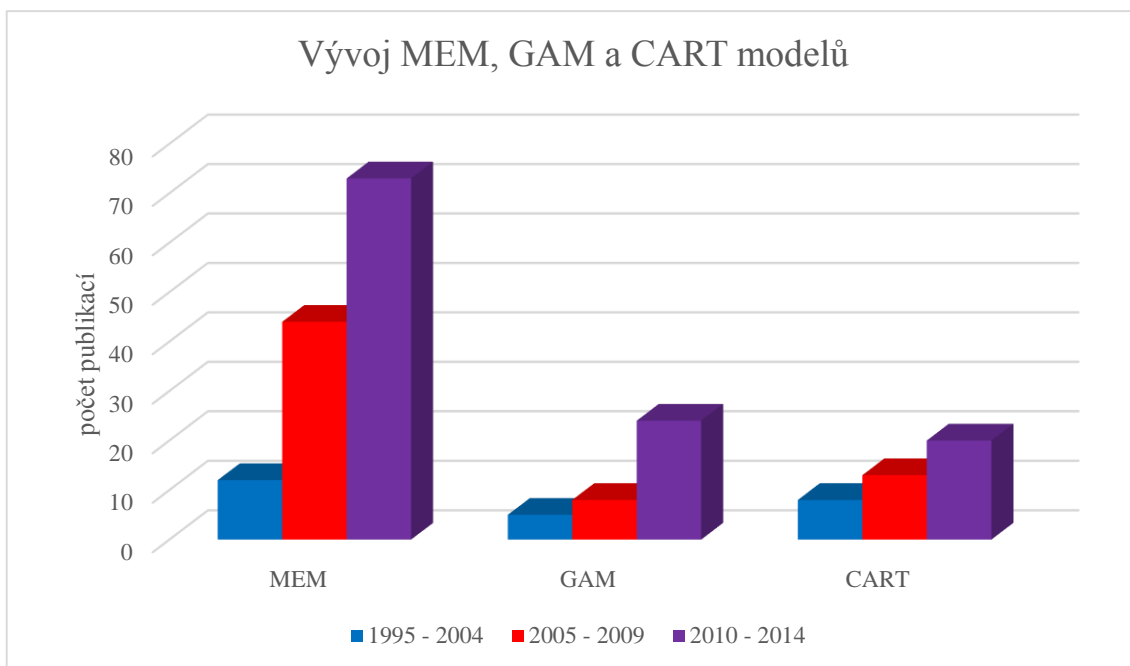
Obr. č. 1: Vývoj množství publikací o MEM v lesnictví za posledních 20 let



Obr. č. 2: Vývoj množství publikací o GAM v lesnictví za posledních 20 let



Obr. č. 3: Vývoj množství publikací o CART v lesnictví za posledních 20 let



Obr. č. 4: Vývoj počtu publikací tří různých regresních metod za posledních 20 let

### 3.2 Výškové křivky a jejich modely

Výška stromu a výčetní tloušťka jsou dvě nejčastěji měřené stromové veličiny při zjišťování stavu lesa. Zatímco měření tloušťek je jednoduché a levné, tak měření výšek stromů je časově a finančně náročné (Adame et al. 2008). Toto tvrzení platí i přesto,



že i díky měření odstupových vzdáleností pomocí laserových a ultrazvukových technologií při měření výšek se snížil čas potřebný k měření (Vargas-Larreta et al. 2009). U stromů, které nemají změřenou výšku je použita výška predikovaná pomocí modelu výškové křivky, který je vytvořen na základě již změřených výšek (Martin a Flewelling 1998; Huang et al. 1992). Predikování výšek stromů má za následek snižování nákladů na získávání dat. Z tohoto důvodu je tvorba vhodného modelu výškové křivky považována za jednu z nejdůležitějších součástí zjišťování stavu lesa (Schmidt et al. 2011). K získání dostatečně přesných odhadů výšek za pomoci modelu je nutno změřit alespoň 20-25 výšek (van Laar a Akça 2007).

Výšková křivka vyjadřuje závislost mezi výškou ( $h$ ) a výčetní tloušťkou stromů ( $d_{1,3}$ ) v porostu v určitém stádiu (věku) jeho vývoje v grafické, tabulkové anebo matematické formě. Obecně se dá výšková funkce vyjádřit vztahem

$$h = f(d_{1,3}) \quad (3.1)$$

Kvalitní výšková křivka musí splňovat následující vlastnosti (Šmelko, 2000) :

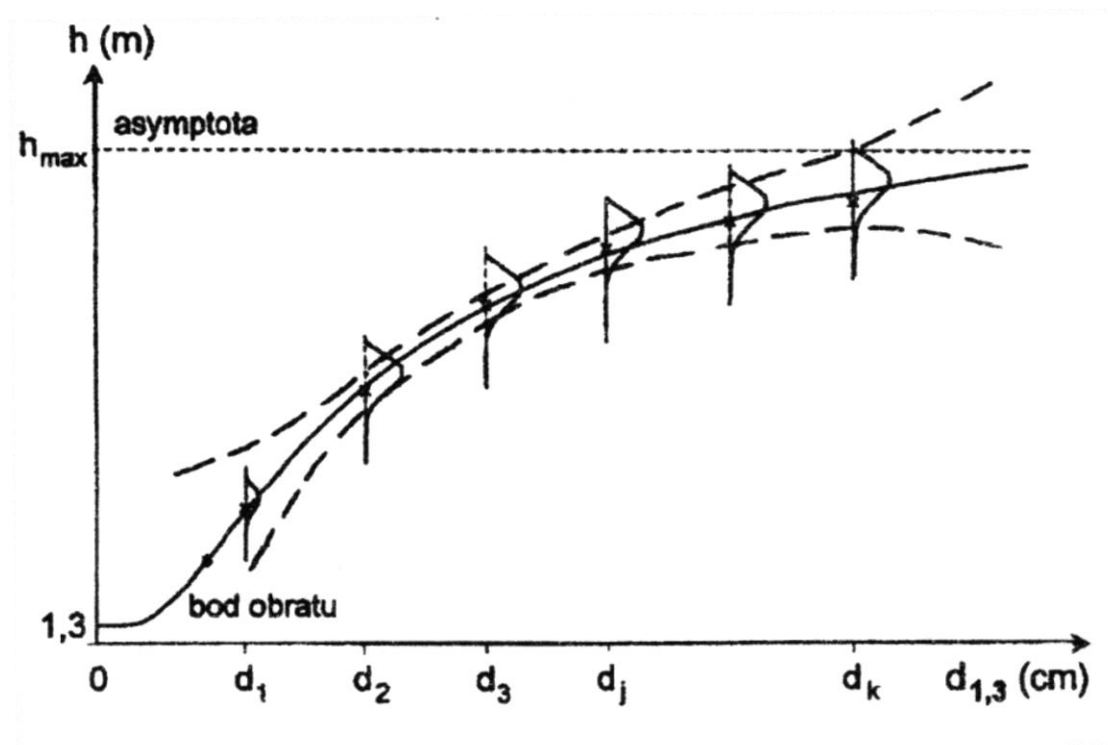
- začíná v bodě  $h = 1,3$  m (výška ve výčetní tloušťce),
- mění svůj tvar z konvexního na konkávní – má inflexní bod,
- je rostoucí (zpočátku strměji, při vyšších tloušťkách pomaleji),
- asymptoticky se přibližuje k maximální výšce  $h_{max}$ , jakou může dřevina v daných stanovištních podmínkách dosáhnout.

Dle Adame et al. (2008) je další důležitou vlastností i to, aby parametry modelu křivky byly biologicky interpretovatelné. Modely s biologicky odůvodněnými parametry poskytují kvalitnější extrapolaci modelových hodnot mimo rozsah měřených dat (Fekedulengn et al. 1999). Výšková křivka splňující všechny požadované vlastnosti je znázorněna na obrázku č. 5.

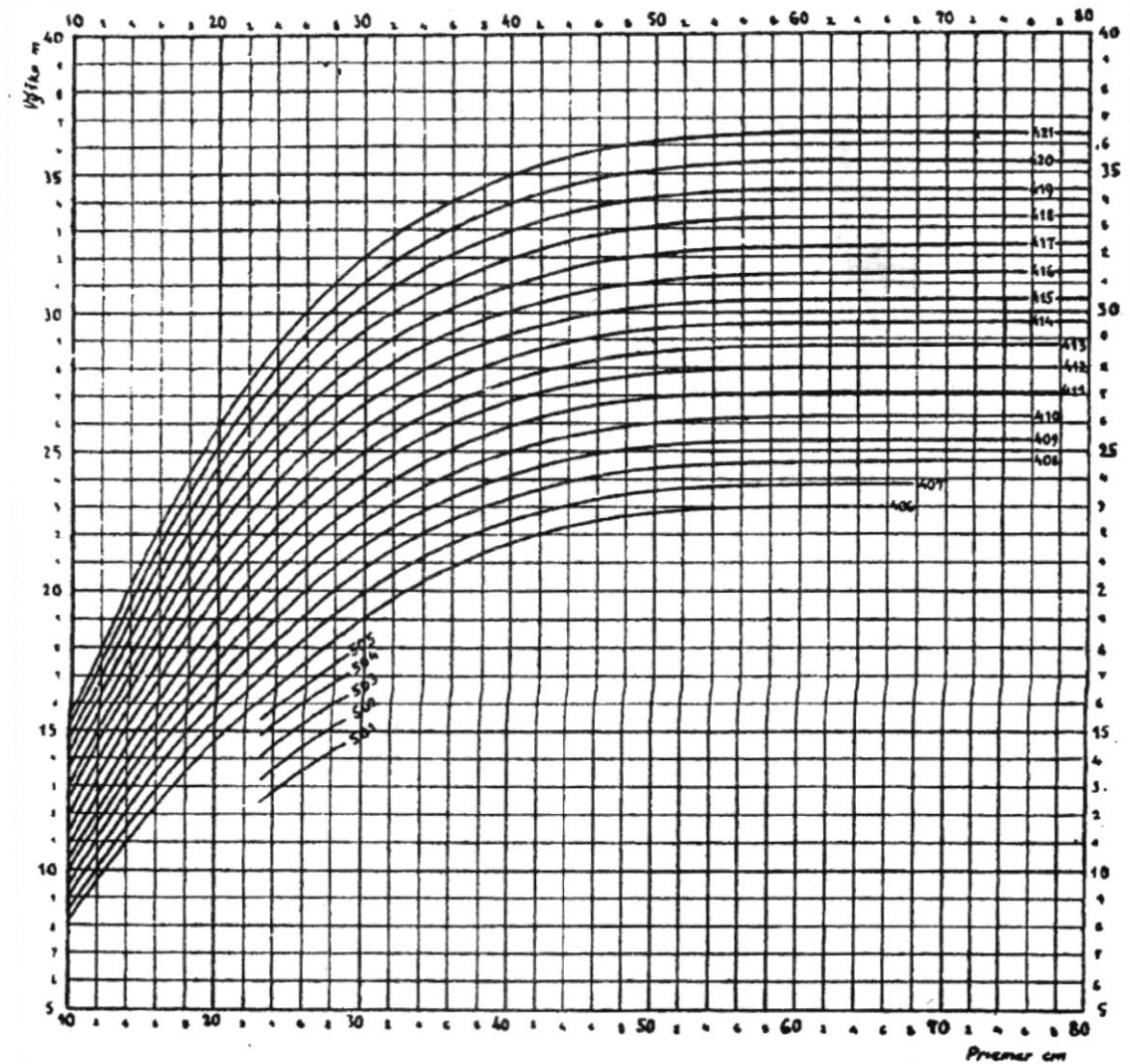
### 3.2.1 Stadiální výšková křivka

Výšková křivka vyjadřující vztah výšky a výčetní tloušťky stromu se také označuje jako stadiální výšková křivka. Stadiálnost výškové křivky znamená, že výšková křivka mění s vývojem porostu svoji polohu. S věkem porostu se postupně přesouvá nahoru a současně se prodlužuje směrem k větším tloušťkám a stává se plošší (Šmelko et al. 1992). Toto tvrzení lze historicky již najít například v pracích autorů Speidel (1893), Assman (1943), Prodan (1951), Halaj (1963, 1974) a dalších.

Právě problém stadiálnosti byl důvodem, proč se vytvořil systém jednotných výškových křivek (JVK). Tento systém měl nahradit opakované měření výšek v porostu pro konstrukci nových výškových křivek, což vyžadovalo stanovení nových hodnot parametrů, které bylo obtížné. Jedná se o ucelenou soustavu schematizovaných křivek, které modelují očekávaný průběh výškových křivek jednotlivých dřevin v dané růstové oblasti a dovolují tedy pro konkrétní porost vybrat právě jednu křivku a tou nahradit skutečnou křivku dřeviny v daném porostu (Šmelko, 2007). Jedná se tedy o soustavu křivek, které odpovídají jednotlivým stádiím vývoje stejnověkého porostu. Vývojové stádium porostu je definováno jeho střední tloušťkou (Fabrika a Pretzsch 2011). Křivky z vyšší střední tloušťkou (tedy ze staršího vývojového stádia) mají strmější počátek a plošší konec, než křivky s nižší střední tloušťkou. Na systém JVK navazuje systém jednotných objemových křivek (JOK). Ten udává objemy jednotlivých stromů v tloušťkových stupních, pro které byly stanoveny výšky právě podle systému JVK. První systém stadiálních výškových křivek sestavil Philipp (1924). V našich podmínkách je dodnes pro zjišťování zásob porostů používán systém JVK vytvořený Halajem (1955). Systém výškových křivek je ukázán na obrázku č. 6.



Obr. č. 5: Výšková křivka se znázorněním svých vlastností (Šmelko, 2007)

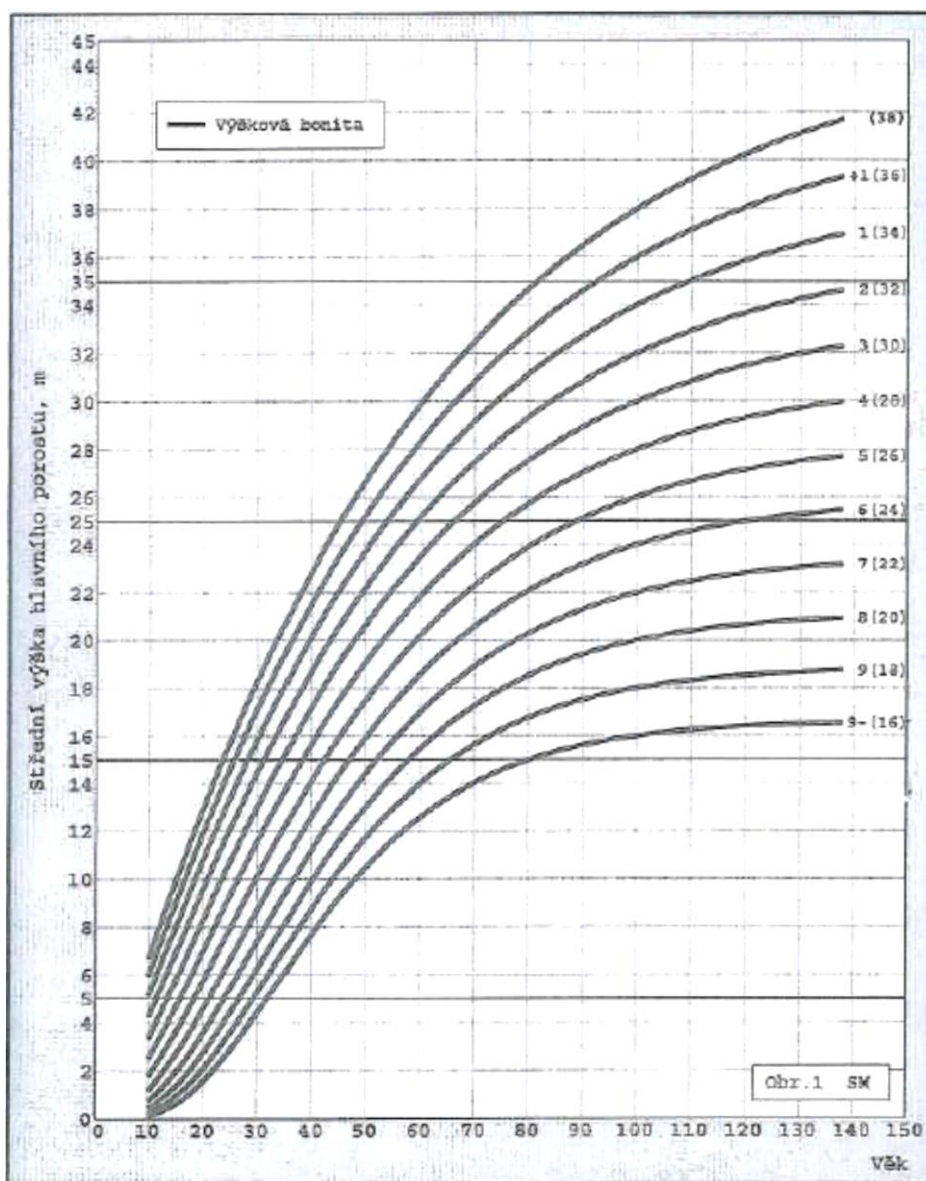


Obr. č. 6: Systém JVK pro smrk, tloušťková třída 4, střední tloušťka 23-29 cm (Halaj, 1955)

Systém jednotných výškových křivek se v praxi využívá z toho důvodu, aby se snížil celkový počet měřených výšek. Při tvorbě skutečné výškové křivky pro danou dřevinu a porost je nutné měřit několik výšek pro každý tloušťkový interval, kdežto při použití jednotné výškové křivky při dané dřevině a porostu se měří výšky pouze v tloušťkových intervalech nejbližší střední tloušťky. Šmelko (2007) uvádí, že nutný počet změřených výšek se takto sníží přibližně na polovinu, což ale stále znamená měření nejméně 10-20 stromů. Nevýhodou systému JVK je naopak to, že jsou použitelné pouze ve stejnověkových porostech.

### 3.2.2 Vývojová výšková křivka

Vývojová výšková křivka (lze ji také nazvat růstovou křivkou výšky) vyjadřuje vývoj střední výšky stejnověkého porostu v závislosti na jeho věku (Šebík a Polák 1990). Fabrika a Pretzsch (2011) ji definují stejně, ale dodávají, že může vyjadřovat také vývoj výšky jednotlivého stromu v závislosti na věku stromu. Vlastnosti vývojové výškové křivky jsou stejné jako u stadiální výškové křivky, jediný rozdíl je v tom, že počátek není při výšce 1,3 m, ale v počátku souřadnic tedy věku 0 let a výšce 0 m. Vývojová křivka se využívá při bonitaci porostu. Vývojová výšková křivka (křivky zde nejsou vykresleny až do počátku souřadnic) je uvedena na obrázku č. 7.



Obr. č. 7: Vývojové křivky středních porostních výšek na různých bonitách (Černý et al. 1996)



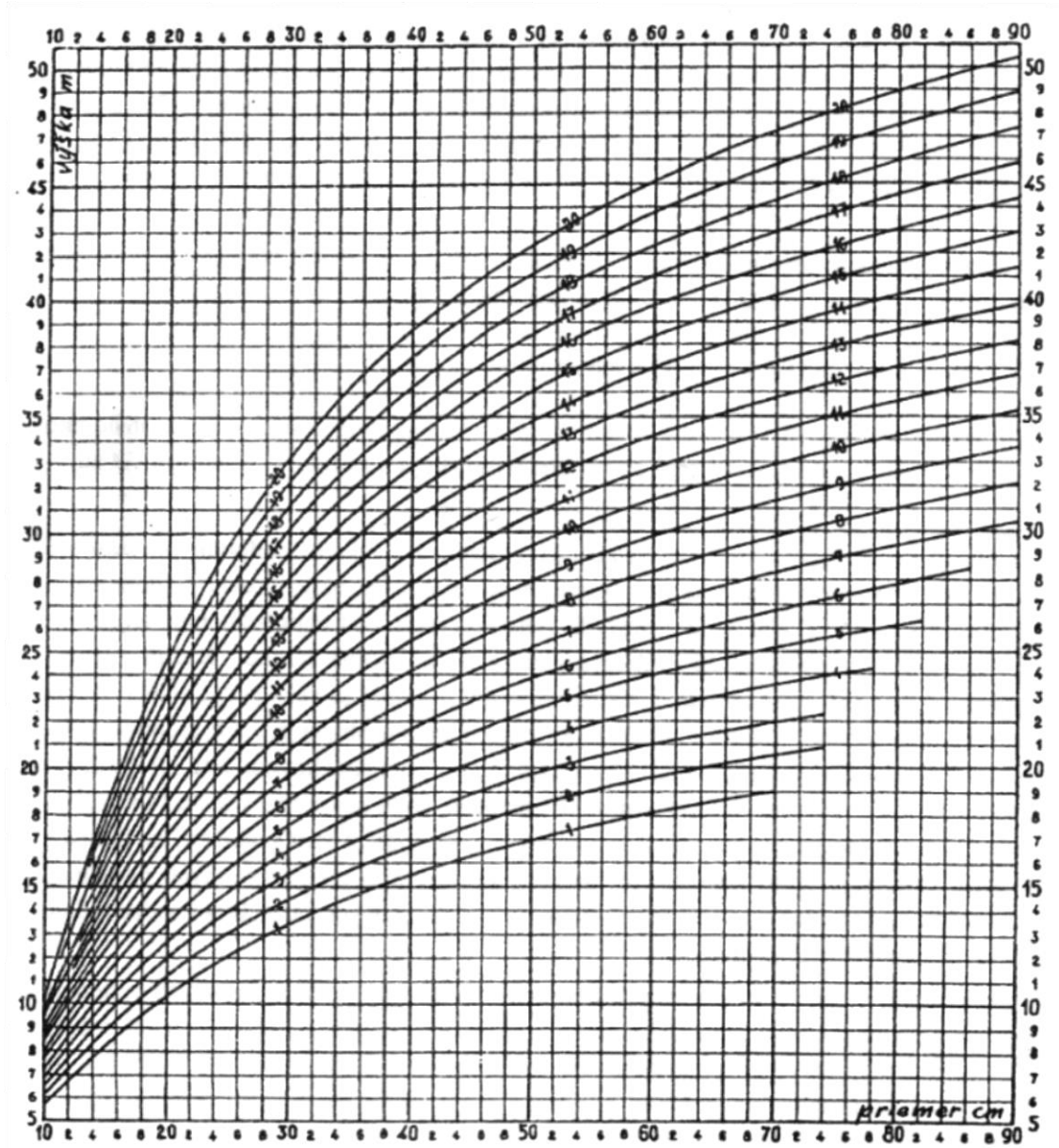
### 3.2.3 Stacionární výšková křivka výběrného lesa

Výšková křivka výběrného lesa (také nazývaná jako výškový tarif) vyjadřuje stejně jako stadiální výšková křivka závislost výšek stromů na výčetních tloušťkách. Rozdíl je pouze v tom, že tato výšková křivka není stadiální, ale je stacionární – tedy má stejný tvar i polohu po celou dobu existence porostu. Výšky v jednotlivých tloušťkových stupních se nemění v průběhu vývoje porostu (Vyskot et al. 1971). Toto platí pouze pro výběrný les v rovnovážném stavu (Šmelko et al. 1992). Podle Assmanna (1961) je však rovnovážný stav spíše výjimkou. Stacionárnost křivky je dána tím, že výběrný les obsahuje přibližně stálý poměr stromů mladých, středně starých a dospělých a tedy rozdělení výšek a tlouštěk je po celou dobu vývoje výběrného lesa víceméně neměnné. Poloha a tvar výškové křivky jsou dány kvalitou stanoviště. Proto jsou vytvářeny systémy výškových křivek (výškové tarify), které vyjadřují vliv stanoviště na růst výběrného lesa. Použití výškového tarifu je způsob, jak vyjádřit bonitu stanoviště ve výběrném lese. Systémy výškových tarifů vytvořili již na přelomu 19. a 20. století Gurnaud (1878) a Biolley (1920). V našich podmínkách sestavil Halaj (1963) výškové tarify pro oblast Slovenska. Systém výškových tarifů je na obrázku č. 8.

### 3.2.4 Modely výškové křivky

K modelování výškové křivky jsou často používány nelineární růstové funkce, které jsou upravené pro modelování výšek (přidáním konstanty 1,3) (Zhang, 1997). Ačkoliv většina z těchto funkcí může adekvátně modelovat výškové křivky, tak pokud se použijí k extrapolaci, budou predikovat chybné hodnoty (Zhang et al. 1996). Mnohem častěji se ale používají speciálně sestavené regresní modely samostatných výškových křivek nazývané jako výškové funkce. Tyto funkce jsou primárně vytvořeny pouze pro modelování výškových křivek a splňují výše uvedené vlastnosti. Šmelko (2007) doporučuje, aby výškové funkce měly menší množství parametrů a pokud je to možné, tak aby byly linearizovatelné. Otázka linearizace vychází sice spíše z požadavku na snadnější výpočet parametrů, protože jejich stanovení pro nelineární model bez možnosti využití výkonného softwaru a hardwaru bylo dříve velmi obtížné. Nicméně i v dnešní době je vhodné využívat linearizovaných modelů, protože pro lineární modely obecně existují mnohem lepší nástroje pro vlastní analýzu i interpretaci výsledků. Za jednu z prvních funkcí lze považovat funkci Levakoviče (1935), která je však velmi často a velmi úspěšně používána dodnes. Modelů výškových křivek bylo dodnes vytvořeno velké množství a jejich přehled je možné najít např. v pracích Fanga

a Baileyho (1998), Penga (1999), Huanga et al. (2000), Husche et al. (2003), Šmelka (2007), van Laar a Akça (2007). Nejčastěji se můžeme setkat s modely, které mají dva nebo tři parametry, modely s více parametry jsou spíše výjimkou.



Obr. č. 8: Systém výškových tarifů pro výběrný smrkový les (Halaj, 1963)

Model výškové křivky mimo využití při určování zásoby porostu je používán např.:

- k stanovení objemu stromu (Larsen a Hann 1987; Jayaraman a Lappi 2001),
- k popisu vertikální struktury porostu (von Gadow et al. 2001),
- k určení sociálního postavení stromu v porostu (Colbert, 2002),
- k popsání růstové dynamiky porostu (Curtis, 1967; Peng et al. 2001),

- k hodnocení poškození porostu (Parresol, 1992),
- k stanovení horní porostní výšky a s její pomocí k určení produkčního potenciálu (Huang a Titus 1993; Vanclay, 1994; Lappi, 1997; Lei a Parresol 2001).

Mnoho modelů růstu a vývoje lesa využívá výšky a tloušťky jako dvě základní vstupní proměnné, výšky jsou buď predikovány všechny nebo jen jejich část (Arney, 1985; Huang et al. 2000). Parametry výškové funkce jsou ovlivněny velkým množstvím faktorů, výchovné zásahy a kvalita stanoviště jsou jedny z nejdůležitějších. Vzájemná kombinace těchto a mnoha dalších méně důležitých faktorů způsobuje velkou prostorovou provázanost (Nanos et al. 2004). Naopak Zeide a Vanderschaaf (2002) považují za nejzřetelnější faktor hustotu porostu. Závislost mezi měřeními je důsledkem hierarchické struktury dat ve vztahu výšky a tloušťky, takže dochází k velké korelaci mezi daty (Gregoire, 1987).

Model výškové křivky může být využit na lokální úrovni nebo může být použit obecně pro větší území (Soares a Tomé 2002). Vztah výčetní tloušťky a výšky stromu je velmi variabilní mezi jednotlivými porosty a závisí na růstových podmínkách a stavu porostu (Sharma a Zhang 2004). Toto potvrzuje konstatování Curtise (1967), který říká, že vztah výšky a tloušťky se liší porost od porostu a dokonce nemusí být konstantní ani v rámci jednoho porostu v čase. Tomu odpovídá definice stadiálnosti výškové křivky. To všechno je důvodem, proč nemůže být jedna výšková křivka použita korektně na rozsáhlé území (Castedo Dorado et al. 2006). Obecně lze říci, že zahrnutí porostních veličin do modelu výškové křivky jako proměnné snižuje střední hodnotu reziduí a zvyšuje přesnost modelu (Soares a Tomé 2002; Calama a Montero 2004). Z těchto poznatků vyplývá, že pokud chceme sestavit model na lokální úrovni, tak se využívá pouze tloušťka stromu jako prediktor, ale pokud chceme použít model pro větší území, je vhodné doplnit model o další prediktory porostních nebo regionálně platných veličin. Tento model se poté nazývá regionální model výškové křivky a vyjadřuje územně širší platnost než lokální model. Tímto modelem se ve svých pracích zabývali např. Schröder a Álvarez Gonzáles (2001), López Sánchez et al. (2003), Sharma a Zhang (2004), Temesgen a von Gadow (2004), Castedo Dorado et al. (2005), Sharma a Parton (2007) a Vargas-Larreta et al. (2009) a další.

Pokud shrneme výše uvedené, tak můžeme konstatovat, že model výškové křivky má své výhody a nevýhody. Mezi jeho výhody patří:

- relativní jednoduchost modelu,

- možnost využití i pro jiný účel než jen určení zásoby porostu,
- využití v lese věkových tříd i v lese nestejnověkém.

Naopak za jeho nevýhody lze označit:

- nutnost měření relativně vysokého počtu výšek stromů pro jeho sestavení,
- stadiálnost modelu (problém s časovou závislostí),
- omezená lokální platnost (problém s prostorovou závislostí).

### 3.3 Smíšené regresní modely

K odstranění problémů v datech uvedených v kapitole 1.2 při použití metody nejmenších čtverců pramenících z různých zdrojů variability se velmi často používají modely se smíšenými efekty (často také nazývané smíšené modely), jejichž počátky lze nalézt v polovině 80. let minulého století, a které se do lesnického výzkumu dostaly díky pracím např. Biginga (1985), Lappiho (1986) nebo Gregoirea (1987). Lze se s nimi setkat také např. v zemědělství, ekologii, ekonomice, sociologii, biomedicíně a dalších vědních oborech (Calegario et al. 2005). Smíšený model je speciálním případem obecného lineárního modelu (Kangas a Haara 2012). Obecný lineární model se využívá právě v případech, kdy není splněn předpoklad o nezávislosti mezi pozorováními. Obecně lze říci, že smíšené modely jsou určitým druhem kompromisu mezi globálním a lokálním modelem – každá experimentální skupina může mít vlastní model, ale tyto modely jsou určitým způsobem provázány s globálním modelem, takže jejich parametry nemohou nabývat zcela libovolných hodnot jako u lokálních modelů a navíc je zde možné kvantifikovat variabilitu mezi experimentálními jednotkami i uvnitř nich (Drápela, 2011). Smíšené modely se skládají z pevné složky, která se týká celého zkoumaného souboru (populace) a náhodné složky, která se týká jednotlivých hierarchických úrovní souboru (Calama a Montero 2005). Smíšené modely dovolují odhadovat pevné a náhodné parametry společně, takže lokální a globální modely mohou být kombinovány do struktury pevných efektů a vnořená variabilita může být definována náhodnými parametry (Adame et al. 2008). Pevná složka odhadů parametrů ve smíšeném modelu zodpovídá za vliv jednotlivých proměnných jako v běžných regresních technikách (Yang a Huang 2011) a náhodná složka odhadů parametrů zodpovídá za heterogenitu a náhodnost v datech způsobenou známými i neznámými faktory (Vonesh a Chiinchilli 1997). Smíšené modely jsou tedy lepší než regrese s využitím metody nejmenších čtverců právě proto, že berou v potaz prostorovou a časovou závislost, čímž poskytují lepší



odhady parametrů (Budhathoki et al. 2008). Prostorová a časová závislost je brána v úvahu tím způsobem, že je definovaná kovarianční struktura náhodné složky modelu (West et al. 2007). Kovarianční struktura modelu předpokládá, že rezidua jsou nezávislá a mají konstantní rozptyl (Lhotka a Loewenstein 2011). Smíšené modely jsou také stále víc používány pro začleňování vnořené stochastické struktury do stromových růstových modelů (Lappi, 1991; Hökkä, 1997). Cílem smíšeného modelu je tedy objasnit část náhodné variability globálního modelu, ale zároveň využít lokálních modelů tak, aby vznikl jeden obecný model, který bude dostatečně flexibilní pro použití na daném území. Náhodné efekty nám tedy mohou pomoci odstranit část nekonstantnosti rozptylu, pokud existuje (Fang a Bailey 2001). Základem konstrukce smíšených modelů je model rozkladu chyby regresního modelu  $\varepsilon$  (nazývaný také model složek variability – variance component model (Mehtätalo, 2010)). V klasickém regresním modelu chyba  $\varepsilon$  vyjadřuje celkově modelem nevysvětlenou variabilitu. Podle modelu rozkladu chyby  $\varepsilon$  je každá měřená hodnota  $y_{ik}$  (která je měřená na  $k$ -tém jedinci  $i$ -té experimentální jednotky) součtem komponent podle tohoto vztahu

$$y_{ik} = \mu + u_i + e_{ik} \quad (3.2)$$

kde je

- $\mu$       fixní střední hodnota,
- $u_i$       náhodný efekt experimentální jednotky  $i$  (tj. ta část chyby  $\varepsilon$ , která je vysvětlitelná variabilitou mezi experimentálními jednotkami, tj. např. mezi plochami),
- $e_{ik}$       náhodný efekt jedince v rámci experimentální jednotky (tj. ta část chyby  $\varepsilon$ , která je vysvětlitelná variabilitou mezi jedinci v rámci experimentální jednotky  $i$ ) (Drápela, 2011).

Smíšený model je generalizací modelu rozkladu náhodné chyby  $\varepsilon$ . Základní rozdíly mezi lokálními a smíšenými regresními modely uvedl Weiss (2007) a jsou uvedeny v tabulce č. 1. Smíšený model existuje v lineární i nelineární podobě, ale velmi často se nelineární modely linearizují (např. Lappi, 1997; Mehtätalo, 2004; Kinnunen et al. 2007), protože proces tvorby smíšeného modelu je pak mnohem snadnější (Robinson a Wykoff 2004). Lineární smíšený model lze v obecné maticové formě zapsat (Gregoire et al. 1995) (zde na příkladu opakovaného měření stromů na zkusných plochách při inventarizaci lesa):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon} \quad (3.3)$$

kde je

- y**  $n$  rozměrný vektor zahrnující  $n$  pozorování závisle proměnné  $z$  na  $n_v$  stromů na  $n_u$  plochách během  $n_w$  časových period,
- X** konstrukční matice o rozměrech  $n \times m$  zahrnující nezávisle proměnné spojené s pevnou složkou modelu,
- $\beta$**   $m$  rozměrný vektor pevných parametrů (vypočítané odhady parametrů),
- Z** konstrukční matice o rozměrech  $n \times q$  pro náhodnou složku modelu,
- b**  $q$  rozměrný vektor náhodných parametrů pro úroveň strom, plocha a perioda,
- $\varepsilon$**   $n$  rozměrný vektor reziduální složky.

Tab. č. 1: Srovnání lokálního a smíšeného modelu (Weiss, 2007)

Lokální model	Smíšený model
Parametry modelu jsou interpretovány jako pevné.	Parametry modelu jsou interpretovány jako pevné nebo jako náhodné.
Nejvhodnější, pokud je cílem výzkumu srovnání modelů jednotlivých skupin.	Nejvhodnější, pokud cílem výzkumu není model pro určitou skupinu, ale charakteristika rozdělení parametrů modelu.
Vhodný, pokud proměnná definující strukturu dat je zároveň hlavním zkoumaným prediktorem.	Vhodný, pokud proměnná definující strukturu dat je především třídící proměnnou.
Vhodný pro velmi malý počet skupin.	Vhodný pro vysoký počet skupin.
Vyžaduje výpočet velkého množství parametrů (zvláště v případě modelů počítaných pro mnoho skupin).	Počet počítaných parametrů nezávisí na počtu skupin.
V případě výpočtu velkého počtu parametrů existuje tendence ke zvyšování variability odhadů parametrů.	V případě velmi malého počtu skupin mohou nastat problémy s odhadem parametrů rozdělení náhodných faktorů.

Podle Fanga a Baileyho (2001) jsou pro tvorbu smíšeného modelu nezbytné tyto tři základní kroky:

1. určení efektů parametrů (proměnných),
2. určení struktury variančně-kovarianční matice uvnitř ploch mezi jednotlivými jedinci,
3. určení struktury variančně-kovarianční matice mezi jednotlivými plochami.

Podle Voneshe a Chiinchilliho (1997) se na odhad parametrů smíšeného modelu nejčastěji používají metoda maximální věrohodnosti (maximum likelihood - ML) nebo metoda omezené maximální věrohodnosti (restricted maximum likelihood – REML). Metoda maximální věrohodnosti je metoda odhadu parametrů modelu, která pracuje s věrohodnostní funkcí. Jako věrohodnost označujeme stav, kdy se parametry modelu  $\theta$  mění a pozorování  $(x_1, x_2, \dots, x_n)$  jsou fixní. Věrohodnostní funkci lze tedy zapsat vztahem

$$L(\theta|x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n|\theta) \quad (3.4)$$

kde je

- L věrohodnostní funkce,
- $\theta$  parametry modelu,
- $x$  nezávislé pozorování  $1, \dots, n$ ,
- $f$  funkce definující model.

Metoda spočívá v nalezení takové kombinace parametrů, která bude maximalizovat věrohodnost, tedy odhady hodnoty parametrů  $\theta$  bude možné považovat za nejpravděpodobnější. Maximum této funkce se získá tak, že derivace této funkce podle parametrů  $\theta$  se bude rovnat nule a vznikne tak soustava  $n$  rovnic.

$$\frac{\partial l(\theta|x)}{\partial \theta_i} = 0 \quad \text{pro } i = 1, 2, \dots, n \quad (3.5)$$

kde je

- $\partial l$  derivace věrohodnostní funkce,
- $\theta$  parametry modelu,
- $x$  nezávislé pozorování  $1, \dots, n$ .

Odhady parametrů modelu jsou právě řešením této soustavy rovnic. Za maximálně věrohodný odhad parametru je považována ta hodnota, která odpovídá globálnímu maximu věrohodnostní funkce (Drápela, 2011). Rozdíl mezi metodou ML a REML je v tom, že odhad směrodatné odchylky rozdělení reziduí modelu  $\sigma$  je u metody ML závislý na vektoru odhadu parametrů modelu  $\beta$  a je vychýlený a u metody REML již závislý ani vychýlený není (Drápela, 2011). Jelikož se ovšem ve smíšeném regresním

modelu setkáváme se dvěma druhy parametrů, používají se pro jejich odhady odlišné metody. Pro odhad fixních parametrů se podle Mehtátala (2010) doporučuje právě metoda omezené maximální věrohodnosti REML a to z těchto hlavních důvodů:

- REML zohledňuje počet stupňů volnosti použitých při odhadu parametrů  $\beta$ . To je důležité hlavně v případě, že model má velký počet fixních parametrů ve srovnání s počtem dat,
- REML odhady prvků matice  $\mathbf{V}$  nezávisí na odhadu parametrů  $\beta$  (matice  $\mathbf{V}$  vyjadřuje rozptyl reziduí lineárního smíšeného modelu -  $\text{var}(\epsilon)$  – detailnější popis matice  $\mathbf{V}$  viz Drápela, 2011),
- REML metoda je oproti metodě maximální věrohodnosti robustnější vůči odlehlým hodnotám.

Pro odhad náhodných parametrů se využívá metoda nejlepšího lineárního nevychýleného prediktoru (best linear unbiased predictor – BLUP) dle Robinsona (1991), která je blíže vysvětlena v kapitole 4.3.7.2.

Pomocí smíšených modelů jsme schopni provádět kalibraci modelu pro konkrétní lokalitu nebo časový okamžik, což je jedním z hlavních důvodů jejich využití. Výsledkem je predikce závisle proměnné při znalosti nezávisle proměnné a vztahu specifikovaného právě modelem. Ve smíšeném modelu může být predikce prováděna dvěma způsoby (Calama a Montero 2004; Sharma a Parton 2007):

1. náhodné parametry mají hodnoty 0 a nejsou dostupné žádné hodnoty závisle proměnné – jedná se o tzv. kalibraci s pevnými efekty (kalibraci na průměr populace - population average calibration), která je specifikovaná na velikost celého základního souboru,
2. náhodné parametry jsou predikovány pro každou plochu, pro kterou jsou dostupné alespoň nějaké hodnoty závisle proměnné, samostatně – jedná se o tzv. kalibraci s náhodnými i pevnými efekty (kalibraci na skupinu - subject specific calibration), která je specifikovaná pro jednotlivé plochy nebo časové okamžiky.

Právě druhý typ predikce se provádí pomocí techniky BLUP. Tato technika vyžaduje, aby pro danou lokalitu nebo časový okamžik bylo změřeno alespoň několik jedinců na dané ploše (Calama a Montero 2005). Podle Westa et al. (2007) lze první způsob kalibrace označit také jako kalibraci nepodmíněnou a druhý způsob jako kalibraci podmíněnou. Kalibrovaný smíšený model je založen na skutečnosti, že stochastická složka variability je důsledkem společného působení různých faktorů (Calama a Montero

2005; Gutzwiller a Riffell 2007). Pokud je prováděna kalibrace na úrovni objektu (skupiny), tak je potvrzeno, že čím víc měření je pro kalibraci k dispozici, tím přesnější je predikce (Calama a Montero 2004; Yang et al. 2009; Crecente-Campo et al. 2010). Například Trincado et al. (2007) konstatovali, že proces kalibrace může zvýšit predikční schopnosti výškového modelu a není proto nutné přidávat do modelu další nezávisle proměnné. Toto tvrzení lze doplnit tvrzením Vargas-Larrey et al. (2009), který konstatoval, že pokud jsou pomocí kalibrace predikovány modelové výšky, tak je vhodnější takto vytvářet modely výškové křivky pro jednotlivé lokality rozsáhlejšího území, což je užitečné především při velkoplošné inventarizaci lesa. Právě proto, že jsou smíšené modely schopny kalibrace, jsou často navrhovány jako primární technika pro modelování růstu (Miguel et al. 2013) a lze je označit za vysoce efektivní regresní techniku (Trincado et al. 2007). Modely se smíšenými efekty se dnes využívají v různých oblastech lesnického modelování po celém světě. Se smíšenými modely se lze setkat například při modelování šířek korun (Fu et al. 2013), modelování mortality stromů (Groom et al. 2012), modelování šířky kůry (Li a Weiskittel 2011), modelování přirozené obnovy (Eerikäinen et al. 2007), modelování vývoje některé z porostních charakteristik (např. Sunanda a Jayaraman 2006; Miguel et al. 2013), modelování morfologické křivky kmene (např. Yang et al. 2009; Cao a Wang 2011; Kublin et al. 2013), modelování hustoty dřeva (Guilley et al. 2004), modelování přírůstu kruhové výčetní základny (např. Budhathoki et al. 2008; Pokharel a Dech 2012; Aakala et al. 2013), modelování přírůstu výčetní tloušťky (např. Calama a Montero 2005; Uzoh a Oliver 2008; Rohner et al. 2013), ale především při modelování výškových funkcí (např. Trincado et al. 2007; Adame et al. 2008; Schmidt et al. 2011; Castaño-Santamaría et al. 2013; Lu a Zhang 2013), kde se tato technika (podle počtu dohledatelných publikací) používá nejvíce.

### 3.4 Zobecněné aditivní modely

Prvními, kdo publikovali zobecněné aditivní modely, byli Hastie a Tibshirani (1986, 1990). Jedná se o neparametrickou metodu, která patří do skupiny zobecněných lineárních modelů (Hastie a Tibshirani 1990; Guisan et al. 2002). Zobecněný aditivní model lze dle Zhanga et al. (2005) zapsat jako:

$$Y = S_0 + \sum_{g=1}^n S_g(X_g) + \varepsilon \quad (3.6)$$

kde je

$Y$  hodnota závisle proměnné,

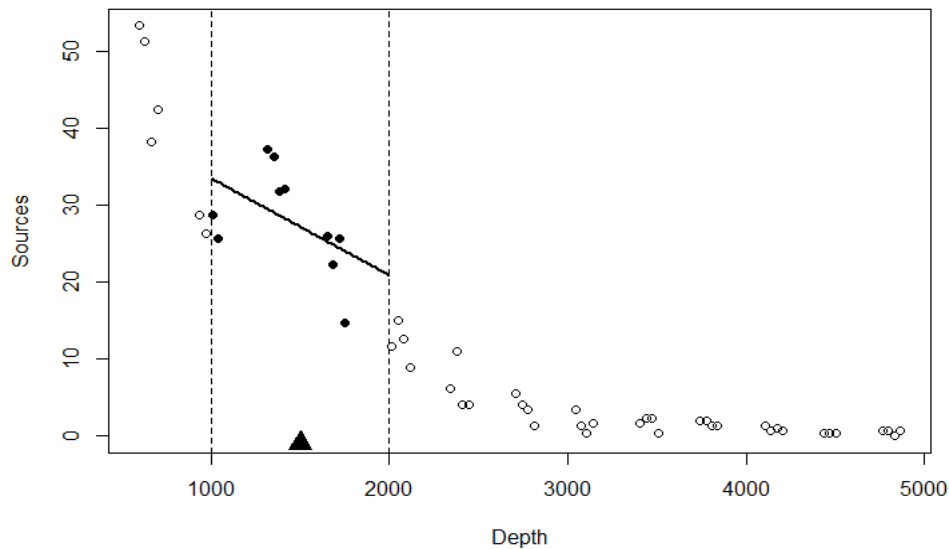
$S_0$  absolutní člen modelu,

$S_g(X_g)$  neparametrická vyhlazovací funkce  $g$ -té nezávisle proměnné  $X$  ( $g = 1, 2, \dots, n$ ),

$\varepsilon$  chyba modelu.

Na rozdíl od zobecněných lineárních modelů, kde jejich primární parametry – regresní koeficienty – charakterizují podobu regresního modelu, nelze zobecněné aditivní modely tak jednoduše vyjádřit pomocí regresních parametrů. V GAM modelech je lineární prediktor nahrazen aditivní vyhlazovací funkcí (Lepš a Šmilauer 2000). Díky tomu, že je vyhlazovací funkce aditivní, tak je možné GAM modely interpretovat stejným způsobem jako běžné lineární modely (Zhang et al. 2008). GAM modely využívají techniku, při níž není vysvětlován vliv jednotlivých nezávisle proměnných na predikovanou veličinu samostatně, ale jako součet účinků všech nezávislých proměnných společně na závisle proměnnou (Guisan a Zimmermann 2000). Přesný zápis rovnice modelu nemusí být znám, protože data sama o sobě určí tvar modelu, což dává GAM modelům takovou míru flexibility, kterou lze při lineárním nebo nelineárním modelování jen stěží dosáhnout (Robinson et al. 2011). Právě proto jsou GAM modely často označovány jako daty řízené modely (data driven models) na rozdíl od běžných regresních modelů, které se nazývají jako funkcí řízené modely (function driven models) (Guisan et al. 2002). GAM jsou také někdy nazývané jako semiparametrické modely, které mohou mít libovolný tvar a parametry a jsou obvykle odvozeny čistě z grafické analýzy dat. Volnost při tvorbě modelu ale není úplná, protože je nutné, aby model splňoval základní požadavky na vztah mezi proměnnými (Robinson et al. 2011). Výhodou GAM modelů je tedy schopnost řešit nelineární a nemonotónní vztahy mezi závisle proměnnou a skupinou nezávisle proměnných (Zhang et al. 2005), kde závisle proměnná nemusí mít Gaussovo rozdělení (Brezger a Lang 2006). GAM modely také snižují citlivost na výskyt extrémních hodnot závisle proměnné (Byun et al. 2013). Jak již bylo řečeno, tak GAM modely využívají vyhlazovací funkce. Mezi běžně používané vyhlazovací funkce lze zařadit spline, kubický spline, B-spline, polynomický kubický spline, LOESS funkci (locally weighted scatter plot smoothing), multivariate LOESS a bivariate thin-plate spline (Venables a Ripley 1997). Ze všech uvedených vyhlazovacích funkcí se asi nejčastěji používá funkce LOESS. LOESS funkce je neparametrická funkce založená na lokálně vážené lineární regresi (Cleveland a Devlin 1988), tj. pro odhad očekávané hodnoty určitého bodu odezvy ( $y_i$ ) se použijí hodnoty prediktoru pouze

v určitém intervalu kolem  $y_i$ , přičemž váha bodů ve výpočtu je různá – nejvyšší je pro body, jejichž hodnota  $x$  je stejná jako u analyzovaného bodu a směrem k hraničním intervalu klesá. Příklad určení hodnoty závisle pomocí jádrové funkce při využití LOESS funkce je uveden na obrázku č. 9.

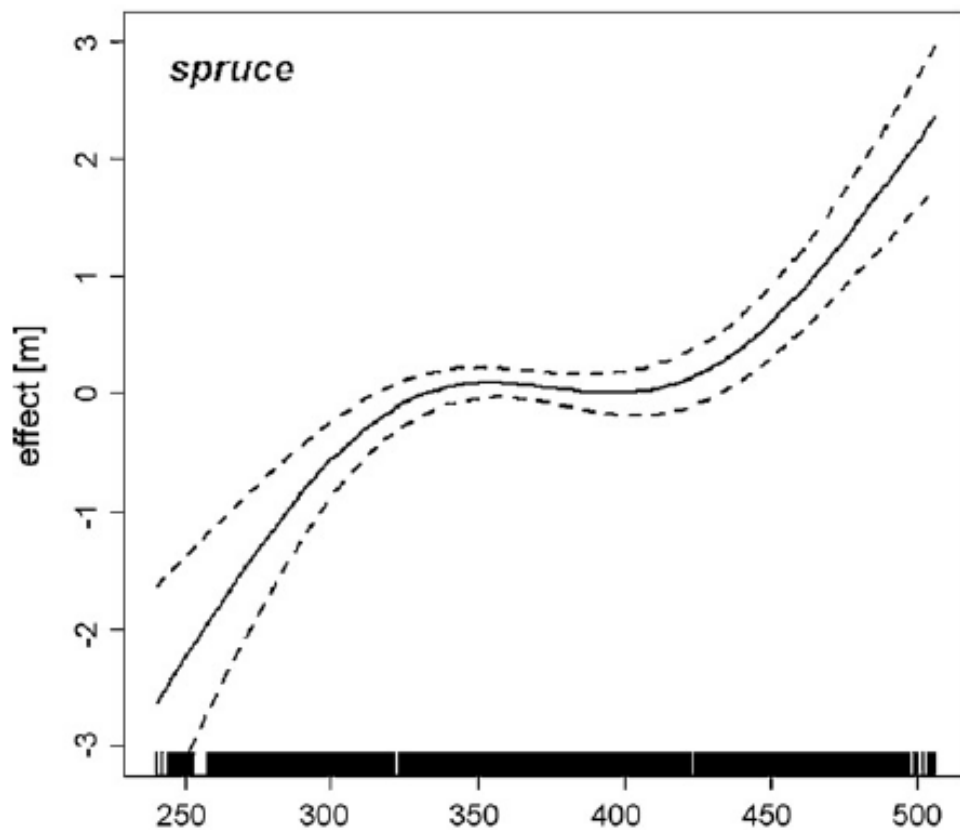


*Obr. č. 9: Princip tvorby LOESS funkce při využití jádrové funkce (Zuur et al. 2009)*

Velmi důležitá je také volba šířky intervalu, ve kterém leží body, které používáme k odhadu očekávané hodnoty odezvy. Tento interval se také nazývá šířka jádra (bandwidth) a určuje se podobně jako u GWR iterativně pomocí minimální hodnoty AIC a zjišťuje, kdy se dosáhne nejlepšího kompromisu mezi hladkostí proložení a minimalizací sumy reziduí (Drápela, 2011). Např. Sironen et al. (2003) zdůrazňuje důležitost šířky jádra, tedy počet okolních stromů, z nichž jsou hodnoty modelu počítány. V případě úzkého jádra jsou výsledné hodnoty velmi podobné výchozím, ale opakované výběry (v rámci šířky jádra) dávají velkou variabilitu výsledných odhadů modelu. Naopak velká šířka jádra poskytuje velmi „hladkou“ vyhlazovací funkci, ale odhady mohou být značně vychýlené, což potvrdil ve své práci i Altman (1992).

Co se týče hodnocení modelu, tak častým problémem GAM modelů je tzv. přeparametrizování nebo také přeúčnění (běžně se používá anglický výraz overfitting). Protože se jedná o neparametrickou metodu, tak to lze chápat tak, že výsledná křivka je příliš „pokroucená“, a to až tak, že již není ani schopna dávat odůvodněné interpolační odezvy (Robinson et al. 2011). Možností, jak určit správnou míru vyhlazení, je využít proces krosvalidace (Wahba, 1990). Nelze však obecně říci,

že by existovalo nějaké jednotné kritérium, které by rozhodlo, že zvolená míra vyhlazení je vhodná nebo ne. S tímto souvisí tvrzení Yue et al. (2012), kteří uvádí, že míra vyhlazení by měla vždy záležet na účelu, ke kterému má být daný model využit a být tak spojen s konkrétní zkoumanou problematikou. Podle Lepše a Šmilauera (2000) je však nejlepším způsobem, jak vyhodnotit GAM model, grafické vynesení odhadnuté vyhlazovací funkce. Příklad grafického výstupu zobecněného aditivního modelu je na obrázku č. 10.



*Obr. č. 10: Zobecněný aditivní model úpravy hodnoty AVB smrku v závislosti na úhrnu srážek za vegetační období (Albert a Schmidt 2010)*

GAM modely jsou doporučovány hlavně tehdy, když není kladen důraz především na přesnost modelu a na poskytování náhledu do nitra ekologických procesů, ale na kvalitu predikce (Aertsen et al. 2010). Dovolují lépe modelovat odezvy jednotlivých druhů s ohledem na environmentální faktory a výsledné modely lépe odpovídají i ekologickým teoriím ve srovnání se zobecněnými lineárními modely (Lehmann et al. 2002). Využívají se především v případech, kdy jednoduché lineární nebo nelineární modely selhávají. Mohou to být například modely, které mají složitější nelineární vztah (např. větší množství inflexních bodů, nemonotónní průběh funkce



apod.). Také se používají v případech, kdy data vykazují nelineární cyklický trend. Obecně lze říci, že GAM modely jsou výhodné, protože mají snadnou interpretovatelnost a poskytují spolehlivé predikce (Moisen et al. 2006). Také jsou doporučovány, protože jsou schopny efektivně využívat lokálních údajů (Sironen et al. 2003). V lesnictví našly GAM modely uplatnění např. při modelování velikosti koruny (Zhang et al. 2005; Fichter et al. 2013), při modelování prostorové dynamiky různých druhů dřevin (např. Falk a Mellert 2011; Kouba a Alados 2012), při modelování výškových křivek (Zhang et al. 2008; Schmidt et al. 2011), při modelování produkčních schopností stanoviště (např. Wang et al. 2005b; Albert a Schmidt 2010; Aertsen et al. 2012), při modelování tvaru kmene (Robinson et al. 2011), při modelování zdravotního stavu stromu (Brezger a Lang 2006), při modelování růstu kruhové základny nebo tloušťky stromu (Yue et al. 2012; Byun et al. 2013) nebo při analýze poškození jemných kořenů při hutnění půdy harvesterovou technologií (Schäffer et al. 2009).

### 3.5 Klasifikační a regresní stromy

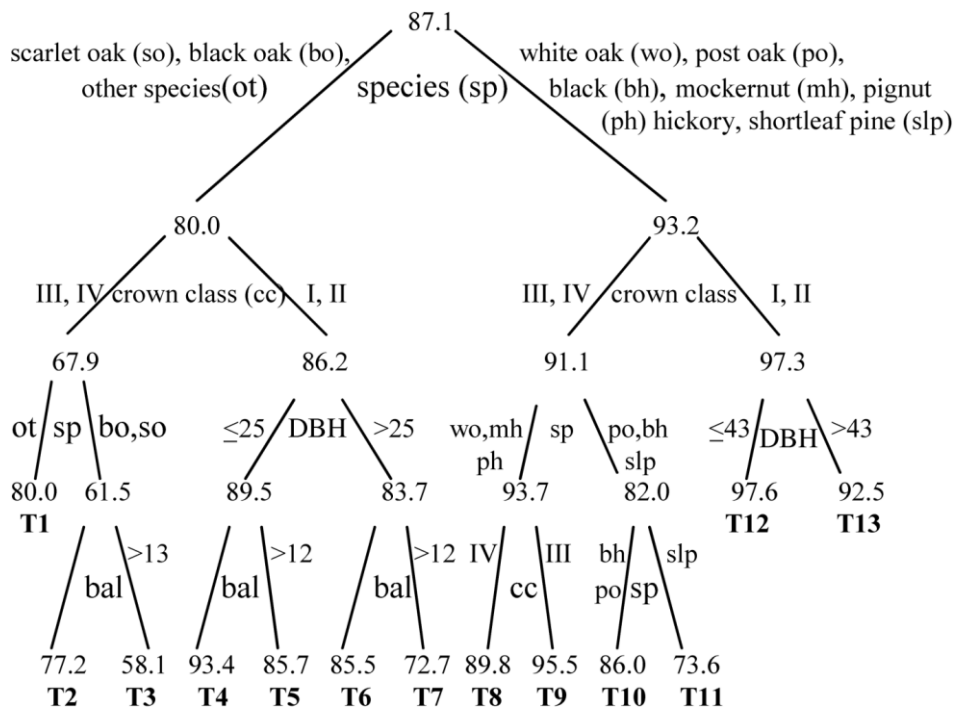
Metodu klasifikačních a regresních stromů lze sice najít poprvé v literatuře již v práci Breimana et al. (1984), ale období svého masivnějšího využití dosáhla až po roce 2000. Jedná se o neparametrickou metodu. Zde je dobré podotknout, že hlavní výhodou všech neparametrických metod je jejich schopnost efektivně využívat lokální údaje (Sironen et al. 2003). Mezi další výhody patří také to, že CART modely se dokáží mnohem lépe vypořádat s chybějícími hodnotami než jiné modely (Dobbertin a Biging 1998), což je velmi užitečná vlastnost, protože je obecně známo, že při ekologických šetřeních se velmi často objevuje problém s chybějící daty, což je problémem pro většinu běžných modelovacích přístupů (Little a Rubin 1987). Modely vytvořené pomocí metody CART jsou alternativou pro lineární a nelineární modely, velmi často nahrazují například logistickou regresi (Clark a Pregiborn 1992), protože na rozdíl od logistické regrese metoda CART nepředpokládá žádné podmínky na rozdělení reziduí (Fan et al. 2006). Podle Clarka a Pregiborna (1992) jsou CART modely jednodušší na interpretaci než lineární modely, obzvláště pokud model obsahuje směs numerických a kategoriálních proměnných. Principem metody klasifikačních a regresních stromů je binární dělení datového souboru do podskupin, které jsou postupně v hodnotách vysvětlované proměnné čím dál víc homogennější (Dobbertin a Biging 1998; Lepš a Šmilauer 2000). Pokud je vysvětlovaná proměnná kvantitativní, tak se jedná o regresní stromy a pokud je

vysvětlovaná proměnná kvalitativní, jedná se o klasifikační stromy (Moisen a Frescino 2002). Pro binární dělení se jako dělicí krok použije právě prediktor. Podle Breimana et al. (1984) je pro správné vytvoření stromu nutných těchto pět základních kroků:

1. vytvořit skupinu binárních dotazů,
2. změřit optimální velikost dělení,
3. vytvořit pravidlo pro ukončení dělení,
4. vytvořit pravidlo pro přiřazení klasifikace terminálnímu výhonu,
5. vytvořit metodu pro zpětné prořezání stromu.

Zpětné prořezání stromu je proces, při kterém se z prvotně vytvořeného stromu (obvykle dosti rozvětveného) odstraní některé větve (dělicí pravidla). Účelem prořezání je to, aby výsledný tvar stromu nebyl příliš složitý pro interpretaci, nevznikaly příliš malé skupiny dat a nedocházelo k přeürčenosti modelu. Výsledný strom musí po prořezání být ale stále schopen kvalitní predikce, což je kontrolováno pomocí různých hodnotících kritérií. Proces zpětného prořezání stromu je nejčastěji prováděn procesem krosvalidace (Fan et al. 2006). Krosvalidace je metoda, která odpovídá na otázku, jak kvalitní bude predikce pro nová data (Hendl, 2012). Je to technika, při které se vytvořený model netestuje na nových datech, ale na stávajících datech. Umožňuje tak kontrolovat kompatibilitu mezi změřenými hodnotami výběrovými a hodnotami modelovými (Scheer, 2010). Při této metodě je část změřených dat  $m_x$  vyjmuta ze souboru a jejich modelové hodnoty jsou odhadnuty pouze ze zbylého výběru o velikosti  $n-m_x$ . Hodnoty  $m_x$  se tedy nijak nepodílí na svém odhadu. Extrémním případem je krosvalidace, kdy se opakovaně vynechává pouze jedna hodnota výběrového souboru. Z rozdílu mezi měřenými a modelovými hodnotami se vypočítají průměrná chyba a rozptyl chyby. Pokud se průměrná chyba rovná nule a rozptyl jedné, tak označujeme model za nevychýlený a adekvátní. Krosvalidace je technika, která se používá pro určení optimální velikosti stromu, aby predikce hodnot byla co nejlepší (Ripley, 1996). Pokud by se prořezání neprovedlo, mohlo by dojít k přeürčenosti modelu (Moisen a Frescino 2002). Výstupy regresních a klasifikačních stromů mohou být grafické, kde kořenový systém představuje původní nedělená data, a v horní části jsou větve s listy, kde každý list představuje jednu finální skupinu dat. Doplňující informace mohou být zobrazeny na kmeni např. sumární statistické charakteristiky jednotlivých ploch apod. (De'ath a Fabricius 2000). Velmi často jsou také stromy zobrazovány přesně naopak tedy kořeny vzhůru a větvemi dolů. Při tvorbě stromu platí pravidlo, že čím menší počet listů

je vytvořen, tím lépe (Räty a Kangas 2008). Příklad klasifikačního stromu je uveden na obrázku č. 11.



Obr. č. 11: Klasifikační a regresní strom znázorňující pravděpodobnost přežití jednotlivých druhů dřevin v oblasti vysočiny Missouri Ozark v USA (Fan et al. 2006)

Legenda: so, bo, wo, po - různé druhy dubů; bh, mh, ph - různé druhy ořechovců; slp - borovice ježatá; ot - ostatní zkoumané druhy; sp - druh dřevin; cc - korunová třída; DBH - výčetní tloušťku; bal - výčetní základna stromů s větší kruhovou výčetní základnou než je průměrná.

Metoda klasifikačních a regresních stromů je ideální pro komplexní analýzu ekologických dat (a to včetně mapování), protože je neparametrická, může pracovat s kategoriálními proměnnými a je robustní vůči odlehlým hodnotám (De'ath a Fabricius 2000). Tato metoda je vhodná jak pro průzkum hierarchické struktury dat, tak i pro modelování (Baker, 1993; Rejwan et al. 1999). Jako velmi silný nástroj se jeví také v případech modelování rozdělení druhů a ekologických společenství (Prasad et al. 2006). Metoda CART je vhodná, protože (Vayssières et al. 2000; McGrath et al. 2003; Swihart et al. 2007):

- je schopna zvládnout komplexně analýzu závislosti vícenásobného datového souboru,
- je nezávislá na druhu rozdělení,
- je přesnější v generování predikcí než ekvivalentní polynomické logistické modely.

Mezi další výhody metody CART patří (De'ath a Fabricius 2000):

- flexibilita zvládnout široké spektrum závislých proměnných včetně numerických, kategoriálních a dat o pravděpodobnosti přežití,
- snadná a robustní konstrukce,
- snadná interpretace,
- schopnost pracovat s neúplnými datovými soubory bez rozdílů, jestli chybí některé konkrétní údaje nezávisle anebo závisle proměnné.

Díky těmto výhodám je metoda CART schopná doplnit, popřípadě nahradit tradiční statistické metody jako jsou například mnohonásobná regrese, analýza rozptylu, logistická regrese nebo modely přežití (De'ath a Fabricius 2000). Pokud se metoda CART používá jako vícenásobná, vzniká samostatná metoda nazvaná „náhodné lesy“ (Random forests) (Mendez a Lohr 2011), která se používá především k predikci (Díaz-Uriarte a Alvarez de Andrés 2006).

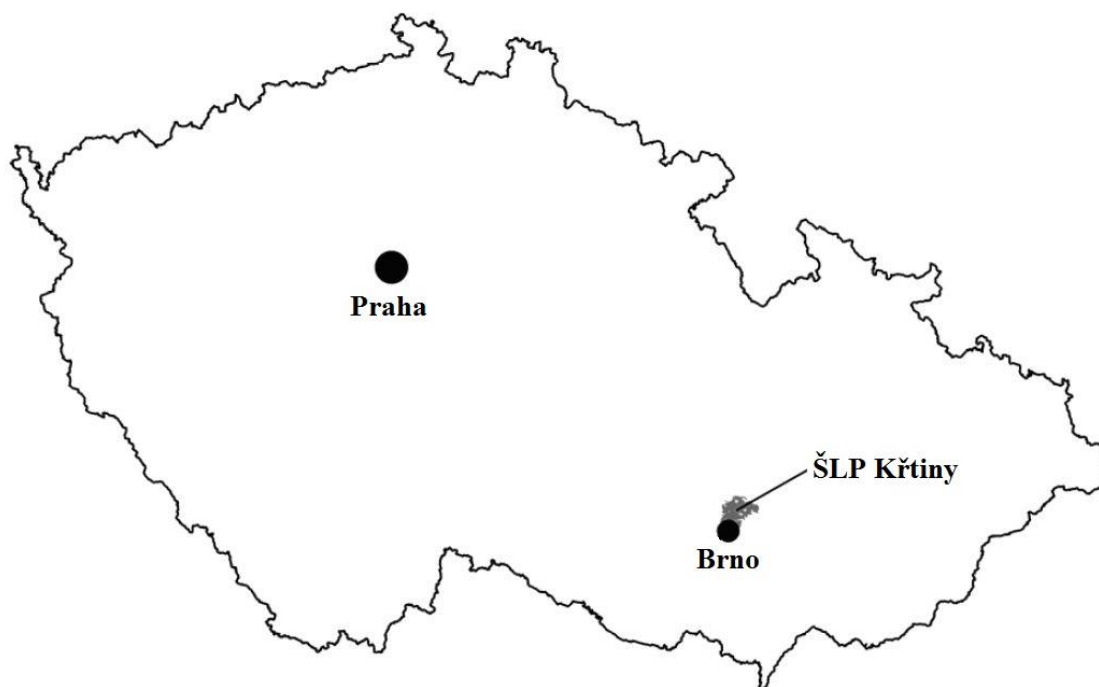
Metoda CART se v poslední době velmi rozšířila např. při modelování mortality stromů (např. Dobbertin a Biging 1998; Gottschalk et al. 1998; Kabrick et al. 2004). Můžeme se s ní setkat také při klasifikaci korun stromů (Nigh a Love 2004), ale je také vhodná pro regionalizaci různých typů globálního modelu, což potvrdili ve svých pracích např. Ráty a Kangas (2008) a Anderson et al. (2010). Svoje uplatnění našla tato metoda i ve stromových růstových modelech, protože je zde vhodná podle Chambers a Hastie (1992) právě proto, že v těchto modelech figuruje více prediktorů a pouze jedna závisle proměnná.

## 4 Metodika

Celá práce je řešena pro dřevinu smrk ztepilý (*Picea abies* (L.) Karst.). Tento druh byl zvolen jak z důvodu jeho vysokého zastoupení v podmínkách ČR (kde je hlavní hospodářskou dřevinou), tak i pro jeho vysoké zastoupení na zkoumaném území. Za zájmovou plochu byl zvolen Školní lesní podnik Masarykův les Křtiny, který je majetkem Mendelovy univerzity v Brně (dále jen ŠLP Křtiny). Vlastní metodiku lze rozdělit do dvou částí, a to do terénních prací a do statistického zpracování dat. Tato kapitola obsahuje také zkrácený popis přírodních podmínek zájmového území.

### 4.1 Přírodní podmínky zájmového území

Informace o přírodních podmínkách byly kompletně přežaty z textové části aktuálně platného Lesního hospodářského plánu pro roky 2013 – 2022 (Lesprojekt Brno, 2013). Mapa s vyznačeným zájmovým územím je na obrázku č. 12. Zájmová část ŠLP Křtiny náleží do PLO 30 Dražanská vrchovina. Nejzastoupenějším cílovým hospodářským souborem (CHS) je 45, respektive jeho varianta 44 – účelové hospodářství živných stanovišť středních poloh. V tomto CHS je v místních podmínkách typické pěstování smrku s bukem, modřínem a jedlí nebo pěstování porostů, kde je hlavní dřevinou buk. Co se týče dřevinné skladby, tak hlavními dřevinami jsou buk (33 % plochy, 31,5 % zásoby) a smrk (18,8 % plochy, 23,3 % zásoby). Zde se soustředíme pouze na popis smrkových porostů s ohledem na zaměření práce. Smrk se vyskytuje na celém území ŠLP, a to i na stanovištích pro něj nevhodných. Většina porostů smrku se nachází v severní polovině území, a to především v oblasti Proklest a v oblasti kolem Jedovnic. V těchto oblastech byl smrk zaveden uměle na začátku 20. století po větších holosečných těžbách listnatých nebo smíšených porostů. Mimo monokultury je smrk hojně zastoupen v jednotlivých až skupinových sečích. Je tedy patrné, že většina porostů je založena jako první generace smrku na bohatších stanovištích po listnatých porostech, což je důvodem toho, že smrk zde často trpí kořenovými hnilobami a václavkou. Ale na území ŠLP můžeme také najít autochtonní smrk, především v oblasti Moravského krasu. Optimální podmínky pro pěstování nachází smrk v CHS 40, 42 a 44. S ohledem na to, že smrk je zde pěstován i na nevhodných stanovištích, doporučuje se snížení jeho zastoupení na 15,2 % plochy.



Obr. č. 12: Mapa s vyznačením zájmového území

## 4.2 Terénní práce

Před vlastním sběrem dat bylo nutné zvolit správná kritéria pro výběr porostních skupin a umístění a velikosti zkusných ploch (dále jen ZP), vybrat veličiny, které budou zjišťovány na ZP, zajistit si technické vybavení k měření a vytvořit časový plán prací. Veškeré ZP se nachází na ŠLP Křtiny na polesí Habrůvka, Vranov a Bílovice nad Svitavou. ZP jsou umístěny do porostních skupin napříč téměř celým věkovým spektrem v rozmezí věků 30 - 136 let tedy od 3. do 14. věkového stupně. Mladší porostní skupiny nebyly uvažovány z důvodu obtížného měření požadovaných veličin, ale také především z důvodu, že v takto mladých porostech se často ještě pracuje s hmotou nehroubí, tudíž se neurčuje zásoba porostní skupiny, ale pracuje se pouze s veličinou plocha porostní skupiny. Z důvodu jednoduchého zaměření v terénu a lepší práce s měřičskou technikou (viz níže práce s Field-Mapem) jsou veškeré ZP kruhové. Velikost ZP byla volena různá podle věku porostní skupiny, a to následovně: porostní skupiny do 40 let věku  $250 \text{ m}^2$  ( $r = 8,92 \text{ m}$ ), porostní skupiny věku 41-80 let  $700 \text{ m}^2$  ( $r = 14,93 \text{ m}$ ) a porostní skupiny věku 81 let a více  $1200 \text{ m}^2$  ( $r = 19,54 \text{ m}$ ). Tyto velikosti ploch byly vybrány z důvodu, aby se na každé ZP nacházelo 30 - 50 stromů. Obecně je bráno jako dostačující počet stromů 15-25 na plochu (Šmelko, 2000). Tento počet byl navýšen z důvodu tvorby většího datového souboru. Pro dostatečné pokrytí variability výškové a tloušťkové struktury a zajištění reprezentativnosti byly v každé porostní skupině

umístěny dvě ZP. Více ploch nebylo plánováno, protože vzhledem k tvaru některých porostních skupin by nebylo možné umístit více než dvě ZP. Toto je další důvod, proč je navýšen počet požadovaných stromů na ZP. V každém věkovém stupni byly vybrány dvě porostní skupiny (mimo 8. věkový stupeň, kde se nepodařilo nalézt více než jednu porostní skupinu, která by vyhovovala zadaným kritériím výběru (viz níže), tudíž byl celkový počet ZP 46. Jako kritéria byly vybrány následující údaje: hospodářský soubor (dále jen HS) 441 – smrkové hospodářství živných stanovišť středních poloh, zakmenění porostní skupiny 0,8 a vyšší, zastoupení SM 80 % a vyšší, plocha porostní skupiny 0,50 ha a větší. Tato kritéria byla zvolena tak, aby na každé ZP byl zaručen dostatek požadovaných jedinců smrku. Vlastní sběr dat na ZP spočíval v zaměření GPS souřadnic středu ZP v systému jednotné trigonometrické sítě katastrální S-JTSK (ČR), zaměření polohy každého stromu vzhledem ke středu ZP (tudíž bylo možné vypočítat GPS souřadnice každého stromu), změření výčetní tloušťky stromu a výšky stromu. Vlastní měření se provádělo pomocí měřící sestavy Field-Map LT, pomocí níž byly zaměřeny polohy stromů s přesností na 1 cm a ukládána do ní veškerá data (tato sestava je primárně určena k práci na kruhových ZP a při práci s ní odpadá tedy náročná práce vytyčování ZP v terénu), výška stromu byla měřena výškoměrem Vertex laser VL 400 s přesností na 0,1 m, výčetní tloušťka stromu byla měřena dvouramennou taxační průměrkou Haglöf Mantax Precision s přesností na 1 cm a GPS souřadnice byly měřeny pomocí polního počítače s GPS přijímačem Trimble JUNO s přesností na 1m. U každé porostní skupiny byla také provedena fotodokumentace a z typologické mapy byl určen lesní typ. Po ukončení sběru dat byla veškerá data přenesena do PC a zálohována. Po tomto kroku již následovalo statistické zpracování dat. Časový plán terénních prací byl následující: leden 2011 - volba kritérií výběru porostních skupin a velikosti ZP, únor-březen 2011 - rekognoskace porostních skupin, duben-září 2011 - umístění ZP do porostních skupin a vlastní měření, říjen 2011 – kontrola a zálohování dat.

### 4.3 Statistické zpracování dat

Veškeré výpočty byly provedeny v programu Microsoft Excel, Statistica 12 (StatSoft, 2013) a open source softwaru R (R Development Core Team, 2011). Součástí výsledků jsou analýzy společné pro všechny metody a analýzy specifické pouze pro jednotlivé druhy regrese. Společnými analýzami jsou výpočet základních statistických

charakteristik, výpočet běžných dendrometrických charakteristik pro jednotlivé stromy a porostní skupiny a provedení průzkumové analýzy dat - EDA dle Tukey (1977) a Chambers et al. (1983). Veškeré statistické analýzy byly vyhodnocovány pro hladinu významnosti  $\alpha = 0,05$ , tedy se statistickou spolehlivostí 95 %.

#### 4.3.1 Výpočet běžných dendrometrických charakteristik

Dendrometrické charakteristiky byly stanoveny pro úroveň stromu a pro úroveň porostní skupiny. Na úrovni stromu to byly: objem stromu vypočítaný podle objemových rovnic dle Petráše a Pajtíka (1991) a kruhová výčetní základna stromu. Na úrovni porostní skupiny byly zahrnuty následující veličiny: střední tloušťka (vypočtená jako aritmetický průměr ze všech naměřených tloušťek), střední výška (vypočtená podle objemových rovnic pro střední tloušťku a objem středního kmene), kruhová výčetní základna/ha, objem středního kmene a počet stromů/ha. Pro všechny porostní skupiny byl určen věk (k věku uvedenému v lesním hospodářském plánu pro období 2003 – 2012 byl připočten počet let od počátku platnosti plánu, aby byl získán aktuální věk ke dni měření). Dále byla pro všechny porostní skupiny stanovena absolutní výšková bonita podle věku a střední výšky porostní skupiny z taxačních tabulek (Černý et al. 1996).

#### 4.3.2 Výpočet základních statistických charakteristik

Jako základní statistické charakteristiky zvlášť pro výšku stromu a zvlášť pro výčetní tloušťku stromu v každé porostní skupině byly vypočítány následující: střední hodnota (aritmetický průměr), interval spolehlivosti (v tabulkách uveden jako IS) střední hodnoty (pro 95 % spolehlivost), směrodatná odchylka, variační koeficient, minimální hodnota, maximální hodnota, koeficient šikmosti, koeficient špičatosti a velikost výběru. Také byly spočítány vybrané statistické charakteristiky pro celý datový soubor, z kterého se počítá globální model.

#### 4.3.3 Průzkumová analýza dat

V rámci průzkumové analýzy dat byly pro obě zkoumané taxační veličiny vytvořeny kvantil-kvantilové grafy pro všechny porostní skupiny. V těchto grafech je pro srovnání vynesena model normálního rozdělení. Také byly vytvořeny bagploty, což jsou zobecněné dvourozměrné krabicové grafy obsahující obě veličiny společně. Bagplot je tvořen z částí bag, loop a fence. Bag představuje rozsah dat v rámci



interkvartilového rozpětí (obdoba „krabice“ v krabicovém grafu), loop, který představuje rozsah neodlehých hodnot („vousy“ v krabicovém grafu) a fence je čára, která tvoří hranici odlehých a neodlehých hodnot („hradby“ v krabicovém grafu). V rámci dalších výpočtů byly všechny porostní skupiny označeny svým identifikátorem a zkratkou (PSK 1 - 23).

#### 4.3.4 Výpočet globálního modelu výškové křivky

Do globálního modelu vstupovaly hodnoty naměřené ve všech 23 porostních skupinách. Celkem se jednalo o 1590 dvojic hodnot výčetní tloušťka a výška stromu. Pro tvorbu globálního modelu bylo vybráno šest výškových funkcí, z toho pět dvouparametrických a jedna tříparametrická. Byly vybrány pouze funkce, které obecně splňují všechny podmínky na výškovou funkci uvedené v kapitole 3.2 a zároveň jsou primárně určeny pro modelování vztahu mezi výčetní tloušťkou a výškou stromu a ne jen druhotně na tento vztah upraveny, jak tomu bývá často například u funkcí růstových. Důvody pro toto rozhodnutí vycházejí ze zjištění v kapitole 3.2.4, že upravené růstové funkce využívané pro modelování výškové křivky mají špatné extrapolační vlastnosti (Zhang et al. 1996). Přehled všech vybraných funkcí je uveden v tabulce č. 2. Globální model byl spočítán pro funkce v jejich nelineárních tvarech.

Tab. č. 2: Přehled použitých výškových funkcí

Číslo funkce	Název funkce	Nelineární tvar
1	Michajlovova (Michajlov, 1943)	$h = 1,3 + a * e^{\frac{b}{d}}$
2	Pettersonova (Petterson, 1955)	$h = 1,3 + \frac{1}{\left(a + \frac{b}{d}\right)^3}$
3	Näslundova (Näslund, 1947)	$h = 1,3 + \frac{d^2}{(a + b * d)^2}$
4	Levakovičova (Huang et al. 1992)	$h = 1,3 + a * \left(\frac{d}{1 + d}\right)^b$
5	Meyerova (Meyer, 1940)	$h = 1,3 + a * (1 - e^{(-b*d)})$
6	Prodanova (Prodan, 1951)	$h = 1,3 + \frac{d^2}{a + b * d + c * d^2}$

Legenda:  $h$  - výška stromu,  $d$  - výčetní tloušťka stromu,  $a, b, c$  - parametry modelu

### 4.3.5 Výpočet lokálního modelu výškové křivky

Výpočet lokálních modelů byl proveden ve všech 23 porostních skupinách samostatně. V každé porostní skupině byly spočítány modely pro všech šest výškových funkcí, které byly použity při tvorbě globálního modelu.

### 4.3.6 Volba modelu

Ze všech použitých funkcí byla pro tvorbu smíšeného modelu a pro srovnání s aditivním modelem a modelem CART využita již pouze jedna funkce. Jako kritéria pro hodnocení modelů mezi sebou na úrovni globálního modelu byly zvoleny střední hodnota reziduí, střední chyba reziduí, směrodatná odchylka reziduí a Akaikeho informační kritérium - AIC (Akaike, 1973). Vzorce na výpočet hodnotících kritérií jsou uvedeny v tabulce č. 3. Výsledné modely byly mezi sebou také porovnány v grafickém zobrazení. Rezidua všech globálních modelů byla také porovnána proti tloušťce a standardizované tloušťce (provedena standardizace rozdělení tlouštěk, kdy průměrná tloušťka odpovídá hodnotě 0). K hodnocení byla také využita funkce mywhiskers (Mehtatalo, 2010) z prostředí jazyka R, která slouží k zobrazení středních hodnot, středních chyb a intervalů spolehlivosti reziduí modelu ve všech decilech nezávisle proměnné – v tomto případě tloušťky. Při hodnocení modelů byly vzaty v potaz jak hodnoty hodnotících kritérií, tak i praktické proložení dat při grafickém zobrazení globálního modelu.

Tab. č. 3: Hodnotící kritéria pro volbu globálního modelu

Hodnotící kritérium	Vzorec
střední hodnota reziduí	$\bar{e}_i = \frac{\sum e_i}{n}$
směrodatná odchylka reziduí	$\sigma_{e_i} = \sqrt{\frac{\sum (e_i - \bar{e}_i)^2}{n}}$
střední chyba reziduí	$SE_{e_i} = \frac{\sigma_{e_i}}{\sqrt{n}}$
Akaikeho informační kritérium	$AIC = n * \ln\left(\frac{\sum e_i^2}{n}\right) + 2 * m$

$e_i$  - hodnota rezidua,  $n$  - počet měření,  $m$  - počet parametrů modelu

Na úrovni lokálních modelů bylo použito jako rozhodovací kritérium AIC a praktické proložení dat při grafickém zobrazení na jednotlivých plochách. Při rozhodování sloužilo jako doplňující kritérium také to, jestli lze daný model linearizovat, protože software použitý k řešení smíšeného modelu poskytuje možnost detailnější analýzy a lepšího praktického zpracování pro lineární modely. Z tohoto důvodu byl vybraný model zlinearizován a jeho výsledky na globální i lokální úrovni porovnány s výsledky nelineárních modelů. Porovnání lineárních a nelineárních modelů proběhlo formou grafického porovnání.

### 4.3.7 Výpočet smíšeného modelu výškové křivky

#### 4.3.7.1 Tvorba smíšeného modelu

Vlastní metodika tvorby smíšeného modelu byla převzata od Weisse (2007) a skládá se z následujících kroků:

- vytvoření modelu 1. úrovně, kdy se řeší vztah zkoumaných veličin na úrovni jednotlivé plochy, kde se mění hodnota veličin na každém stromě na ploše – v tomto případě je tato úroveň představována vztahem výšky stromu a výčetní tloušťky stromu v rámci jedné porostní skupiny,
- vytvoření modelu 2. úrovně, kdy se model rozšiřuje o prediktory, které se vztahují na vyšší úroveň, než je jedinec na ploše. Jedná se tedy o veličiny, které popisují plochu jako celek – v tomto případě se může jednat například o věk porostu, střední taxační veličiny, počet stromů/ha, výčetní základnu porostu, zásobu/ha, bonitu apod.,
- vytvoření modelu 3. a popřípadě i vyšší úrovně, kdy se do modelu přidávají další prediktory, které jsou společné pro všechny úrovně nižší – zde by se mohlo jednat o údaje charakterizující například přírodní lesní oblast (klimatické údaje apod.) nebo i celou republiku, kdyby se model rozšířil ještě o jednu vyšší úroveň.

Pro vytvoření modelu 1. úrovně se doporučuje postupovat podle čtyř kroků, které uvádí Drápela (2011) a jsou také založeny na metodice doporučené Weissem (2007). Tento postup slouží k otestování významnosti jednotlivých členů modelu a zjištění, jestli přidání jednotlivých členů bude mít za následek snížení nevysvětlené variability. Obecný postup tvorby modelu 1. úrovně je popsán v tabulce č. 4. K testování významnosti jednotlivých členů přidávaných v jednotlivých čtyřech krocích bylo použito AIC. Jako významně lepší model se bere takový model, jehož hodnota AIC je o 7 – 10 nižší než u druhého

srovnávaného modelu (Burnham a Anderson 1998). Také byl využit na srovnávání modelů test poměrů věrohodností.

Tab. č. 4: Fáze tvorby smíšeného modelu a jejich význam (Drápela, 2011)

Fáze tvorby smíšeného modelu	Model	Význam
1. globální model	$y_{ik} = a + bx_{ik} + \varepsilon_{ik}$	základní model, který ignoruje případnou strukturu v datech, nutný pro testování přínosu smíšeného modelu
2. model průměru	$y_{ik} = a_i + \varepsilon_{ik}$ $a_i = a + u_{ai}$	model bez prediktoru, potřebný pro odhad $\tau^2$ a $\sigma^2$ ( $\varepsilon_{ik} \sim N(0, \sigma^2)$ ; $u_{ai} \sim N(0, \tau^2)$ )
3. model s náhodným absolutním členem	$y_{ik} = a_i + bx_{ik} + \varepsilon_{ik}$ $a_i = a + u_{ai}$	1. srovnáním s globálním modelem se testuje významnost struktury v datech 2. srovnáním s modelem průměru se testuje význam prediktoru
4. model s náhodným absolutním i regresním členem	$y_{ik} = a_i + bx_{ik} + \varepsilon_{ik}$ $a_i = a + u_{ai}$ $b_i = b + u_{bi}$	testuje přínos náhodného členu a prediktoru

První fází je vytvoření globálního modelu. V tomto případě byl globální model vypočítán pro lineární tvar Pettersonovy výškové funkce

$$\frac{1}{\sqrt[3]{h-1,3}} = a + b * \frac{1}{d} \quad (4.1)$$

Pro tento model byly dopočítány odhady parametrů, p-hodnoty pro hodnocení významnosti jednotlivých parametrů, p-hodnota pro hodnocení významnosti celého modelu, koeficient determinace a AIC.

Druhou fází je vytvoření nejjednodušší formy smíšeného modelu – tedy model průměru. Tento model slouží k výpočtu variančně-kovarianční struktury dat. Z výstupů lze vypočítat rozptyly náhodných efektů modelu. V tomto případě se jedná o náhodný efekt  $\varepsilon_{ik}$ , který vyjadřuje efekt stromu  $k$  na úrovni porostní skupiny  $i$  (část z celkové chyby

$\varepsilon$ , která je vysvětlitelná variabilitou stromů v rámci jedné porostní skupiny – variabilita je představována rozptylem  $\sigma^2$ ) a náhodný efekt  $u_{ai}$ , který vyjadřuje náhodný efekt mezi jednotlivými porostními skupinami  $i$  (část z celkové chyby  $\varepsilon$ , která je vysvětlitelná variabilitou mezi jednotlivými porostními skupinami – variabilita je představována rozptylem  $\tau^2$ ). Podle těchto hodnot jsme schopni usoudit, jestli je použití smíšeného modelu oprávněné například s ohledem na výskyt heterogenity v datech nebo závislosti uvnitř jednotlivých ploch (zde porostních skupin). Pokud je jeden z rozptylů výrazně vyšší než druhý, tak to poukazuje na tu úroveň, ve které je vyšší variabilita. Pomocí těchto dvou rozptylů lze také vypočítat vnitroskupinový korelační koeficient  $\rho$ , který je schopen odhalit závislost jednotlivých hodnot na úrovni plochy (v tomto případě uvnitř porostní skupiny).

$$\rho = \frac{\tau^2}{\tau^2 + \sigma^2} \quad (4.2)$$

Pro tento model byla také vypočítána hodnota AIC, která se pak využila na srovnání s dalšími modely, aby bylo možné ověřit, jestli po přidání prediktoru a dalších náhodných částí došlo ke zlepšení modelu.

Ve třetí fázi se sestaví model, ve kterém je prediktor a jehož absolutní člen má již pevnou i náhodnou část. Pro tento model byly dopočítány odhady pevných parametrů, p-hodnoty pro hodnocení významnosti jednotlivých parametrů, AIC, směrodatná odchylka náhodného efektu absolutního členu ( $\tau_a$ ) a směrodatná odchylka reziduí ( $\sigma$ ). Tento model byl porovnán s lineárním globálním modelem, aby byla otestována významnost struktury dat a také byl porovnán s modelem průměru, aby byl otestován přínos prediktoru 1. úrovně. Testování probíhalo testem poměru věrohodností a srovnáním hodnot AIC.

Čtvrtou fází je vytvoření takového modelu, který obsahuje náhodnou část absolutního i regresního členu. Po výpočtu tohoto modelu jsme již schopni zapsat variančně-kovarianční matici náhodných členů. I pro tento model byly dopočítány odhady pevných parametrů, p-hodnoty pro hodnocení významnosti jednotlivých parametrů, AIC, směrodatná odchylka náhodného efektu absolutního členu ( $\tau_a$ ), směrodatná odchylka náhodného efektu regresního členu ( $\tau_b$ ), směrodatná odchylka reziduí ( $\sigma$ ) a korelační koeficient parametrů ( $\rho$ ). Z těchto hodnot se také vypočítá kovariance mezi náhodnými efekty ( $\tau_{ab}$ ) a to jako  $\rho * \tau_a^2 * \tau_b^2$ . Výsledný model s oběma náhodnými členy lze zapsat takto:

$$\frac{1}{\sqrt[n]{h-1,3}} = (a + u_{ai}) + (b + u_{bi}) * \frac{1}{d} \quad (4.3)$$

$$u_i = \begin{bmatrix} u_{ai} \\ u_{bi} \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_a^2 & \tau_{ab} \\ \tau_{ab} & \tau_b^2 \end{bmatrix} \right) \quad (4.4)$$

Tento model byl také porovnán s modelem, který má náhodnou část pouze u absolutního členu. Porovnání bylo provedeno pomocí testu poměru věrohodností a také porovnáním podle hodnot AIC. Bylo také provedeno grafické srovnání tohoto modelu s modelem, který měl náhodný pouze absolutní člen a s modely lokálními. Dále bylo provedeno grafické srovnání obou smíšených modelů s náhodnými členy a lokálních modelů pro porovnání shody modelů. Rezidua smíšeného modelu s náhodným absolutním i regresním členem byla porovnána s rezidui globálního modelu, aby bylo možné ověřit, zda došlo ke snížení hodnot reziduí, což předpokládá právě smíšený model.

Po výpočtu smíšeného modelu první úrovně bylo ještě ověřeno, jestli mají rezidua modelu normální rozdělení. Toto bylo provedeno pomocí kvantil-kvantilového grafu (QQ grafu). Pomocí grafu reziduí proti predikovaným hodnotám bylo ověřeno, zda je splněna podmínka homoskedasticity.

Dalším krokem při tvorbě smíšeného modelu je vytvoření modelu druhé úrovně, tedy přidání do modelu další proměnné, která je společná pro celou porostní skupinu, ale odlišná mezi jednotlivými porostními skupinami. V tomto případě bylo vyzkoušeno hned několik takových proměnných a to počet stromů/ha, věk porostní skupiny, kruhová výčetní základna/ha, objem středního kmene, střední tloušťka a absolutní výšková bonita. Dále byly použity i logaritmy těchto proměnných a bylo otestováno, jestli není těsnější závislost mezi parametry modelu a těmito logaritmy než mezi parametry a vlastními hodnotami proměnných. Ke zlogaritmování se přistoupilo, ale až potom, co se prokázalo, že některé závislosti vykazují nelineární charakter. Porovnání míry závislosti parametrů modelu na proměnných a jejich logaritmech bylo otestováno pomocí testu na shodu korelačních koeficientů. Testové kritérium  $Z_k$  pro tento test se vypočítá podle vzorce

$$Z_k = \frac{|Z_{\rho_1} - Z_{\rho_2}|}{\sqrt{\frac{2}{n-3}}} \quad (4.5)$$

kde je

$n$  počet měřených hodnot,

$Z_{\rho_1}$  a  $Z_{\rho_2}$  Fisherovy transformace výběrových korelačních koeficientů.

Kritická hodnota  $Z_{\alpha/2}$  pro tento test je rovna 1,96, což je hodnota 95% kvantilu normovaného normálního rozdělení. Pokud tento test zamítne nulovou hypotézu, znamená to, že je prokazatelný rozdíl mezi korelačním koeficientem závislosti parametru modelu a porostní proměnné a korelačním koeficientem závislosti parametru modelu a logaritmu porostní proměnné. Můžeme tedy říci, že základní soubory, které jsou popsány těmito dvěma korelačními koeficienty, nejsou shodné. Do modelu druhé úrovně jako první vstoupily proměnné, které měly nejvyšší hodnoty korelačních koeficientů. Vytvořený model byl opět otestován a porovná s předchozími modely podle stejného postupu jako u modelů první úrovně. U takto vytvořeného modelu druhé úrovně byla otestována závislost hodnot parametrů modelu na zbylých porostních veličinách, a to pomocí testu významnosti korelačního koeficientu. Testové kritérium  $T_k$  pro tento test se stanoví podle vzorce

$$T_k = \frac{\rho \sqrt{n-2}}{\sqrt{1-\rho^2}} \quad (4.6)$$

kde je

- $\rho$  korelační koeficient,
- $n$  počet měřených hodnot.

Kritická hodnota  $T_{\alpha/2}$  je v tomto případě rovna 2,08 a odpovídá 95% kvantilu Studentova rozdělení pro  $n-2$  stupňů volnosti. Pokud se prokáže statisticky významná závislost u některé porostní proměnné a hodnot parametru, tak je vytvořen nový model, který musí již obsahovat i tuto proměnnou. Tento postup je nutné opakovat, dokud se stále prokazuje závislost mezi porostními proměnnými a hodnotami parametrů. Za výsledný model se bere takový, který obsahuje pouze významné proměnné a již není zjištěna žádná významná závislost mezi hodnotami parametrů tohoto modelu a jakoukoliv jinou proměnnou, která v modelu není zahrnuta.

#### 4.3.7.2 Kalibrace smíšeného modelu výškové křivky

Kalibrace smíšeného modelu se využívá ve chvíli, kdy chceme sestavený smíšený model využít na jiné plochy (porostní skupiny) nebo na jiné časové okamžiky, než pro které byl vytvořen. Na kalibraci byla využita podmíněná kalibrace (viz kapitola 3.3), která si klade za podmínku, že je nutné znát aspoň několik hodnot závisle proměnné na dané ploše nebo v daný časový okamžik. Nepodmíněná kalibrace nebyla využita. Důvodem je to,

že nevyžaduje žádné hodnoty závisle proměnné (v tomto případě výšky), což znamená, že není možné vypočítat náhodné části odhadu parametrů, takže model obsahuje pouze pevné parametry. Jedná se tedy potom o klasický globální model. Podmíněnou kalibraci lze popsat vztahem (Calama a Montero 2005):

$$E(\mathbf{y}|\mathbf{b}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} \quad (4.7)$$

kde význam jednotlivých symbolů je stejný jako v rovnici 3.3.

Při tomto typu kalibrace se pro stanovení náhodné části odhadu parametrů využívá technika BLUP. Zde je tato technika vysvětlena na odhadu náhodných částí odhadů parametrů výškové funkce. Princip lze dobře pochopit na uvažovaném vektoru  $\mathbf{h}$  o velikosti  $j$  ( $j$  je celkový počet změřených výšek). Vektor neznámých náhodných částí parametrů  $\mathbf{u}_i \sim N(0, \mathbf{D}_i)$  lze označit jako  $\mathbf{h}_1$  (je to část vektoru  $\mathbf{h}$ ) a vektor změřených výšek v dané porostní skupině  $\mathbf{y}_i$  jako  $\mathbf{h}_2$  (také část vektoru  $\mathbf{h}$ ). Princip odhadu vektor  $\mathbf{h}$  ve své práci dobře popsal Drápela (2011), který využil poznatky autorů McCulloha a Searleho (2001) a je popsán níže. Podle Drápely (2011) a McCulloha a Searleho (2001) lze tedy předpokládat, že  $E(\mathbf{h}_1) = \boldsymbol{\mu}_1$ ,  $E(\mathbf{h}_2) = \boldsymbol{\mu}_2$ ,  $\text{var}(\mathbf{h}_1) = \mathbf{V}_1$ ,  $\text{var}(\mathbf{h}_2) = \mathbf{V}_2$  a  $\text{cov}(\mathbf{h}_1, \mathbf{h}_2) = \mathbf{V}_{12}$ . Vektor  $\mathbf{h}$  lze tedy zapsat maticí

$$\mathbf{h} = \begin{pmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{pmatrix} = \begin{bmatrix} (\boldsymbol{\mu}_1) & (\mathbf{V}_1 & \mathbf{V}_{12}) \\ (\boldsymbol{\mu}_2) & (\mathbf{V}_{12} & \mathbf{V}_2) \end{bmatrix} \quad (4.8)$$

kde je

- $\boldsymbol{\mu}_1$  střední hodnota vektoru  $\mathbf{h}_1$ ,
- $\boldsymbol{\mu}_2$  střední hodnota vektoru  $\mathbf{h}_2$ ,
- $\mathbf{V}_1$  rozptyl vektoru  $\mathbf{h}_1$ ,
- $\mathbf{V}_2$  rozptyl vektoru  $\mathbf{h}_2$ ,
- $\mathbf{V}_{12}$  kovariance mezi vektory  $\mathbf{h}_1$  a  $\mathbf{h}_2$ .

Mějme tedy vektor  $\mathbf{h}_1$  a chceme predikovat hodnoty vektoru  $\mathbf{h}_2$ . Pomocí techniky BLUP můžeme stanovit vektor  $\mathbf{h}_1$  následovně:

$$\widehat{\mathbf{h}}_1 = \mathbf{u}_1 + \mathbf{V}_{12}\mathbf{V}_2^T(\mathbf{h}_2 - \boldsymbol{\mu}_2) \quad (4.9)$$



Odhad náhodných částí parametrů pomocí BLUP lze zapsat také dle vztahu (Calama a Montero 2005):

$$\hat{\mathbf{b}} = \hat{\mathbf{D}}\hat{\mathbf{Z}}^T(\hat{\mathbf{R}} + \hat{\mathbf{Z}}\hat{\mathbf{D}}\hat{\mathbf{Z}}^T)^{-1}\hat{\mathbf{e}} \quad (4.10)$$

kde je

- $\hat{\mathbf{b}}$  vektor náhodných částí odhadů parametrů,
- $\hat{\mathbf{D}}$  matice variability mezi skupinami,
- $\hat{\mathbf{Z}}^T$  konstrukční matice náhodných efektů,
- $\hat{\mathbf{R}}$  matice reziduální (vnitroskupinové) variability,
- $\hat{\mathbf{e}}$  vektor, jehož rozměr je dán počtem pozorování a jeho hodnoty jsou rezidua modelu získaného nepodmíněnou kalibrací (globálního modelu).

Když vezmeme v potaz, že  $\mathbf{V}_1 = \hat{\mathbf{D}}$ ,  $\mathbf{V}_2 = \hat{\mathbf{R}} + \hat{\mathbf{Z}}\hat{\mathbf{D}}\hat{\mathbf{Z}}^T$ ,  $\mathbf{V}_{12} = \hat{\mathbf{D}}\hat{\mathbf{Z}}^T$  a  $(\mathbf{h}_2 - \boldsymbol{\mu}_2) = \hat{\mathbf{e}}$ , tak vidíme, že rovnice 4.10 je maticovým zápisem rovnice 4.9. Podle obou tedy můžeme stanovit odhad náhodných parametrů smíšeného modelu.

V tomto případě byl použit na kalibraci smíšeného modelu výškové křivky postup, kdy byla provedena kalibrace ve dvou různých variantách.

V první variantě byly změřeny výšky 1, 2, 3, 4 anebo 5 stromů. Stromy byly vybírány tak, aby splnily podmínku střední tloušťky v intervalu  $\bar{d} \pm 2$  cm (zde byla brána za střední tloušťku hodnota aritmetického průměru všech změřených tlouštěk v porostu). Interval  $\pm 2$  cm byl zvolen z toho důvodu, že se v porostu nemusí nacházet strom s přesně vypočítanou střední tloušťkou, tak je nutné vyhledat strom s co nejbližší tloušťkou a při průměrkování naplno se často právě používá dvoucentimetrový tloušťkový interval, takže nalezení takovýchto stromů by nemělo být problematické. Střední tloušťka byla vybrána z toho důvodu, že pokud chceme sestavit výškovou křivku, tak musíme provést průměrkování ať už naplno nebo na zkusných plochách a máme tedy dostatek změřených tlouštěk pro výpočet střední tloušťky jakožto aritmetického průměru.

Druhou variantou bylo změření 3, 6 nebo 9 stromů, kdy stromy náležely do tří různých tloušťkových intervalů. Jeden interval byl v rozmezí minimální tloušťka až minimální tloušťka + 4 cm, druhý interval byl shodný s variantou jedna (tedy střední tloušťka  $\pm 2$  cm) a třetí interval byl v rozmezí maximální tloušťky - 4 cm až maximální tloušťka. Vždy se tedy jednalo o čtyřcentimetrový tloušťkový interval. Pokud byly vybírány tři stromy, tak náležel do každého intervalu jeden strom, pokud šest stromů, tak

do každého intervalu dva a pokud devět stromů, tak do každého intervalu tři stromy. Tento výběr byl zvolen, aby byly naměřeny hodnoty nejen kolem aritmetického průměru, ale také na počátku a konci výškové křivky, což by mělo napomoci kvalitnějšímu odhadu polohy a tvaru výškové funkce. Více stromů v jednotlivých intervalech bylo měřeno z důvodu lepšího zachycení variability.

Stromy, které vstupovaly do modelu, byly v rámci daných tloušťkových intervalů vybírány pomocí simulací náhodného výběru, tak aby byl odstraněn vliv subjektivity měřiče. Pro každou variantu bylo provedeno 10 simulací. Kalibrace byla provedena v šesti porostních skupinách, tak aby v každé věkové třídě byla zastoupena právě jedna porostní skupina. V rámci věkové třídy byl porost vybrán náhodně. Jedná se konkrétně o porostní skupiny s ID (identifikátor porostní skupiny) 3, 5, 11, 14, 19 a 22. Vybrané porostní skupiny se tedy výrazně liší nejen věkem, ale také ostatními taxačními veličinami. Taxační veličiny těchto porostních skupin jsou uvedeny v příloze č. 1. Pro hodnocení kvality smíšeného modelu, potažmo pro jeho srovnání s lokálními modely byla využita následující kritéria:

- koeficient determinace ( $R^2$ ),

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (4.11)$$

- střední hodnota reziduí ( $\bar{e}_i$ ) – viz tabulka č. 3,
- směrodatná odchylka reziduí ( $\sigma_{ei}$ ) – viz tabulka č. 3,
- střední chyba regresního modelu (RMSE),

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-m}} \quad (4.12)$$

- Akaikeho informační kritérium (AIC) – viz tabulka č. 3,
- střední hodnota odchylek modelových hodnot získaných smíšeným modelem od modelových hodnot lokálního modelu ( $\Delta_i$ ),

$$\Delta_i = \frac{\sum_{i=1}^n |(\hat{y}_{imm} - \hat{y}_{imn\hat{e}})|}{n} \quad (4.13)$$

kde je

$y_i$  naměřená hodnota jedince  $i$  ( $i = 1, 2, 3, \dots, n$ ),

$\hat{y}_i$  modelová hodnota jedince  $i$  ( $i = 1, 2, 3, \dots, n$ ),

- $\bar{y}_i$  průměrná hodnota ze všech jedinců  $i$  ( $i = 1, 2, 3, \dots, n$ ),
- $n$  počet měření,
- $m$  počet parametrů modelu,
- $\hat{y}_{imm}$  modelová hodnota jedince  $i$  ( $i = 1, 2, 3, \dots, n$ ) vypočítaná smíšeným modelem,
- $\hat{y}_{imm\epsilon}$  modelová hodnota jedince  $i$  ( $i = 1, 2, 3, \dots, n$ ) vypočítaná metodou nejmenších čtverců.

U směrodatné odchyly reziduí, střední chyby regresního modelu, Akaikeho informačního kritéria a střední hodnoty odchylek modelových hodnot dvou porovnávaných modelů se snažíme jejich hodnoty minimalizovat, protože čím jsou jejich hodnoty nižší, tím je model kvalitnější. Naopak co nejvyšší hodnota koeficientu determinace značí kvalitní model. Střední hodnota reziduí by se měla co nejvíce přibližovat k nule. Jako nejvhodnější způsob kalibrace byl doporučen ten, který vycházel souhrnně nejlépe podle výše uvedených kritérií.

#### 4.3.8 Výpočet zobecněného aditivního modelu

Zobecněný aditivní model (dále jen GAM model) je neparametrická metoda, která se využívá především tehdy, kdy závislost vykazuje nelineární charakter. Toto je typické právě pro výškovou křivku, a proto byl tento model zvolen k ověření využití jako alternativa k modelu výškové funkce. V GAM modelu byl otestován i přínos různých porostních veličin jako dalších proměnných stejně jako v případě smíšeného modelu. Bylo to provedeno z důvodu, že i porostní proměnné mají s výškou nelineární vztah a GAM model by mohl tuto nelinearitu být schopen analyzovat. Z porostních proměnných nebyla prokázána statisticky významná závislost výšky na absolutní výškové bonitě dřeviny a kruhové výčetní základně na hektar. Byla prokázána závislost mezi výškou stromu a věkem porostu, objemem středního kmene, střední tloušťkou a počtem stromů na ha. Mezi těmito čtyřmi nezávisle proměnnými ale existovala silná závislost (vypočítané korelační koeficienty pro všechny vzájemné závislosti byly vypočítány vždy vyšší než 0,8, respektive -0,8), takže z důvodu odstranění vlivu multikolinearity byl v modelu ponechán jako porostní proměnná pouze věk porostu, který může být chápán jako nositel ostatních proměnných. Po zanesení věku porostní proměnné do všech tří typů vyhlazovacích funkcí byl vyhodnocen přínos této porostní proměnné jako statisticky významný, ale při vizuální kontrole bylo zjištěno, že výsledná funkce má ve starších porostech klesající trend (klesání výšky s větší výčetní tloušťkou), takže došlo

k „přehlazení“ této funkce. Toto přehlazení se projevovalo i při jiných vstupních nastaveních vyhlazovacích funkcí, takže věk porostu musel být také z modelu odstraněn. Po této analýze byly tedy vytvořeny finální modely závislosti výšky stromu na jeho výčetní tloušťce. Tyto modely byly také porovnány s konkrétními modely výškových křivek vypočítanými podle Pettersonovy výškové funkce nelineární regresi. Využívaný model lze tedy zapsat rovnicí

$$h_i = \alpha + f_s(d_{1,3i}) + \varepsilon_i \quad (4.14)$$

$$\varepsilon_i \sim N(0, \sigma^2) \quad (4.15)$$

kde je

- $h_i$  modelová výška jedince  $i$ ,
- $\alpha$  absolutní člen,
- $f_s$  vyhlazovací funkce výčetní tloušťky,
- $d_{1,3i}$  výčetní tloušťka jedince  $i$ ,
- $\varepsilon_i$  reziduum jedince  $i$ ,
- $\sigma^2$  rozptyl rozdělení reziduí modelu.

Výsledná modelová výška se tedy získá tak, že se sečte hodnota absolutního členu a hodnota odečtená z vyhlazovací funkce pro danou výčetní tloušťku.

GAM modely dovolují použít různá rozdělení pro závisle proměnnou a více druhů spojovacích funkcí pro vztah mezi závisle proměnnou a nezávisle proměnnými. V tomto případě bylo uvažováno s normálním rozdělením a identickou spojovací funkcí. Důvodem bylo to, že stejné podmínky jsou využívány i pro modelování výškové křivky pomocí klasického regresního modelu výškové funkce. Byly vytvořeny dvě skupiny modelů, z nichž první byla sestavena z celého datového materiálu, a byly vytvářeny globální modely a druhá skupina vytvářela modely pro jednotlivé porostní skupiny, takže byly vypočítávány lokální modely. Pro každou tuto skupinu modelů byly použity stejné vyhlazovací funkce.

Pro tvorbu modelu byly vybrány tři nejběžněji používané vyhlazovací funkce a to konkrétně funkce LOESS, spline a kubický spline. Vyhlazovací funkce jsou neparametrické funkce, které mají za úkol proložit data křivkou, která má co nejtěsněji kopírovat průběh změřených dat bez jakéhokoliv praktického opodstatnění pouze s cílem minimalizovat míru deviance. Jedná se tedy o tzv. data driven metodu,

kdy poloha křivky je dána pouze samotnými daty a ne žádnou matematicky definovanou funkcí. Je tedy vždy nutné, aby podoba výsledného modelu byla kontrolována a byl sestaven a použit pouze takový model, který je z praktického hlediska opodstatněný a vysvětlitelný. Mezi vyhlazovacími funkcemi jsou výrazné rozdíly v metodice jejich tvorby - viz níže.

Funkce LOESS pracuje tak, že kolem každé hodnoty nezávisle proměnné se vytvoří tzv. jádro (interval hodnot o různé šířce) a pouze tímto intervalem se proloží funkce vytvořená pomocí lineární vážené regrese, z které se poté vypočítá hodnota závisle proměnné pro danou hodnotu nezávislou. Znamená to tedy, že hodnoty uvnitř jádra, které jsou dále od jeho středu, mají menší vliv na vlastní hodnotu odezvy, než hodnoty blíže středu jádra. Spojením všech odezvy vznikne finální vyhlazená křivka. Jak již bylo uvedeno v kapitole 3.4, tak optimální šířka jádra se volí na základě hodnot AIC výsledného modelu. Tento princip byl využit i v této práci, ale pouze na funkci LOESS. Zde je dobré podotknout, že je nutné kontrolovat i výsledný tvar modelu, aby měl stále reálné vysvětlení a biologické uplatnění.

Funkce spline a kubický spline pracují na odlišném principu. Všechny splinové funkce opouští princip jádra kolem jedné hodnoty, ale celý rozsah nezávisle proměnné dělí na segmenty, které jsou odděleny tzv. uzly. Hlavní rozdíl mezi funkcí spline a kubický spline je ten, že klasická spline funkce využívá pro modelování uvnitř segmentu lineární regresní model a kubický spline funkce využívá polynom 3. řádu – tedy kubický polynom. Pro spojení takto vytvořených úseků se využívá algoritmus, který je sestaven na principu derivací 1. a 2. řádu a vypočítá se pro uzel, kde se segmenty propojují stejnou hodnotou závisle proměnné a dojde tedy k propojení celé vyhlazovací funkce. Počet segmentů a uzlů se může výrazně lišit. Keelle (2008) doporučuje dělit datový soubor v rozsahu menším než 30 hodnot pomocí tří uzlů a datový soubor větší než 100 hodnot pomocí pěti uzlů. Druhou možností dělení do segmentů je opět vizuální kontrola vyhlazení ve spojení s hledáním co nejmenší hodnoty AIC. Třetí možností, která byla využita na splinové funkce v této práci, je proces tzv. zobecněné krosvalidace. Je to modifikace klasické krosvalidace, která se využívá v aditivních modelech. Důvodem je právě to, že funkce nemají samostatné parametry a celý proces je značně náročný na výpočet. Detailnější popis lze nalézt např. v práci Wooda (2006). V tomto případě byla využita verze krosvalidace, kdy se z datového souboru vypouští pouze jedno pozorování. Jedná se tedy o tzv. leave one out krosvalidaci. Tato metoda automaticky vybere nejlepší model z pohledu kvality proložení, a to tak, že porovnává validační model (s  $n-1$  pozorováními)

a celkový model (s  $n$  pozorováními) proti naměřeným hodnotám. Toto provede na více modelech, v nichž mění počet uzlů a vybere ten nejvhodnější.

Po vytvoření optimálně vyhlazených modelů byly tyto modely porovnány mezi sebou. Jako porovnávací kritéria sloužily následující veličiny:

- podíl vysvětlené deviance – tzv. pseudo  $R^2$  (Dobson, 2002)

$$\text{pseudo } R^2 = 100 * \frac{D_0 - D_r}{D_0} \quad (4.16)$$

- Akaikeho informační kritérium (AIC) – viz tabulka č. 3,
- střední hodnota reziduí ( $\bar{e}_i$ ) – viz tabulka č. 3,
- směrodatná odchylka reziduí ( $\sigma_{ei}$ ) – viz tabulka č. 3,
- střední hodnota odchylek modelových hodnot získaných aditivním modelem od modelových hodnot nelineárního modelu ( $\Delta_{ig}$ ),

$$\Delta_{ig} = \frac{\sum_{i=1}^n |(\hat{y}_{igam} - \hat{y}_{intr})|}{n} \quad (4.17)$$

- vizuální podobnost s modelem výškové křivky vytvořeného nelineární regresí,
- reálnost průběhu modelu,

kde je

$D_0$  nulová deviance,

$D_r$  reziduální deviance,

$\hat{y}_{igam}$  modelová hodnota jedince  $i$  ( $i = 1, 2, 3, \dots, n$ ) vypočítaná aditivním modelem,

$\hat{y}_{intr}$  modelová hodnota jedince  $i$  ( $i = 1, 2, 3, \dots, n$ ) vypočítaná nelineární regresí,

$n$  počet měření.

Deviance je obdoba sumy čtverců reziduí v klasických regresních modelech, která se vypočítává při použití metody maximální věrohodnosti. Nulová deviance udává celkovou devianci vypočítanou při použití nulového modelu (tedy modelu pouze s absolutním členem bez prediktoru). Vyjadřuje tedy prakticky to stejné, co celková suma čtverců v klasickém lineárním modelu. Reziduální deviance je stanovená jako míra deviance, která není vysvětlena použitím daného modelu. Má tedy velmi podobný význam jako reziduální suma čtverců v klasickém lineárním modelu.

Souhrnné zhodnocení výše uvedených kritérií rozhodlo o vybrání nejvhodnějšího modelu jak na úrovni globálního modelu, tak i na úrovni lokálních modelů. Všechny vytvořené

modely byly porovnány i graficky, a to jak mezi sebou navzájem, tak i s modelem Petersonovy výškové funkce vytvořeným nelineární regresí.

#### 4.3.9 Postup tvorby regresního stromu

Metoda CART byla vybrána, aby bylo posouzeno, zda ji lze použít jako alternativní metodu pro stanovování výšek stromů v porostních skupinách. V tomto případě byla použita metoda regresního stromu, protože výška stromu je spojitá veličina. Výstupem tzv. rozhodovacího stromu je vždy střední hodnota a rozptyl rozdělení modelových výšek pro homogenní skupinu stromů určených několika různými prediktory. Z praktického hlediska je tedy pro každý strom v této skupině stanovena stejná modelová výška. Protože byl v této práci kladen velký důraz na praktickou použitelnost metody, tak jako prediktory do regresního stromu vstupovaly pouze takové proměnné, které jsou v porostní skupině snadno měřitelné nebo je lze jednoduše vypočítat. Konkrétně se jednalo o výčetní tloušťku stromu, střední tloušťku dřeviny (zde byla brána jako průměrná tloušťka ze všech změřených stromů), věk porostní skupiny jako spojitě proměnné a absolutní výškovou bonitu dřeviny jako kategoriální proměnnou. Přínos jednotlivých proměnných se hodnotil pomocí analýzy významnosti. Tato analýza měří míru poklesu přesnosti při vynechání dané proměnné z celého stromu. Můžeme tedy říci, že ty proměnné, které mají vysokou míru významnosti, jsou pro rozhodovací strom velmi přínosné, protože jejich vypuštění by způsobilo snížení přesnosti celého výsledného modelu. V této práci je míra významnosti uváděna na stupnici 0-100. Čím vyšší je hodnota, tím se jedná o významnější proměnnou.

Vlastní tvorba regresního stromu probíhala na stejných datech jako předešlé metody, ale s tím rozdílem, že byla rozdělena na data trénovací (někdy také označovaná jako kalibrační) a validační (lze se setkat i s označením testovací) a to v poměru 860 : 730 stromům. Rozdělení bylo provedeno tak, aby obě skupiny dat měly přibližně podobný rozsah všech použitých prediktorů. Základní charakteristiky použitých proměnných v obou datových podskupinách je uveden v tabulce č. 5.

Regresní strom provádí binární dělení v jednotlivých uzlech tak, že redukuje reziduální sumu čtverců. Minimální počet hodnot v jednom uzlu byl stanoven na 50. Znamená to tedy, že pokud je v uzlu méně než 50 hodnot, tak už se dále nevětví. Neznamená to ale, že je uzel dělen dále vždy, pokud je více než 50 hodnot. Pokud strom zjistí, že je lepší výsledek zaručen i při vyšším počtu, tak se dál nedělí. Důvodem je zabránit přeurčenosti modelu a zaručit, že odhad střední hodnoty rozdělení výšek bude proveden z vyššího

počtu hodnot a bude dostatečně kvalitní. Tímto způsobem se vytvoří strom, který poskytuje nejpřesnější predikci. Tento strom je ale velmi rozvětvený a musí se ještě tzv. zpětně prořezat. V tomto případě bylo provedeno prořezání pomocí desetinásobné krosvalidace, kdy je tento proces prováděn tak, že datový materiál je rozdělen na 10 stejných dílů a na nich je provedena samotná krosvalidace. Výsledkem krosvalidace je strom, který již není tolik větvený, ale je kompromisem mezi složitostí stromu a kvalitou predikce stromu. Po vytvoření takového regresního stromu, je tento strom aplikován na validační data. Výsledná kvalita modelu se stanoví jak na trénovacích, tak i na validačních datech. Jako kvalitativní kritéria byly použity následující veličiny:

- koeficient determinace ( $R^2$ ) – viz vzorec 4.11,
- střední hodnota reziduí ( $\bar{e}_i$ ) – viz tabulka č. 3,
- střední chyba reziduí ( $SE_{ei}$ ) – viz tabulka č. 3,
- směrodatná odchylka reziduí ( $\sigma_{ei}$ ) – viz tabulka č. 3.

Také byla provedena grafická analýza reziduí, která sice není nezbytná, protože CART metoda obecně neklade žádné podmínky na rozdělení reziduí (Fan et al. 2006), ale je dobré ji provést z důvodu kontroly, zdali výsledný model není systematicky vychýlený.

*Tab. č. 5: Základní statistické charakteristiky prediktorů trénovacích a validačních dat regresního stromu*

TRÉNOVACÍ DATA				
charakteristika	d (cm)	$\bar{d}$ (cm)	T (roky)	AVB (m)
průměr	31,73	32,50	80,55	33,68
směr. odchylka	10,82	8,25	34,09	1,81
minimum	9,00	20,00	30,00	32,00
maximum	69,00	45,00	136,00	38,00
VALIDAČNÍ DATA				
charakteristika	d (cm)	$\bar{d}$ (cm)	T (roky)	AVB (m)
průměr	32,90	33,72	82,87	33,49
směr. odchylka	11,03	8,21	31,72	1,48
minimum	12,00	19,00	30,00	32,00
maximum	75,00	47,00	133,00	38,00

Legenda: d – výčetní tloušťka stromu,  $\bar{d}$  – střední tloušťka porostní skupiny, T – věk porostní skupiny, AVB – absolutní výšková bonita, směr. odchylka – směrodatná odchylka



## 5 Výsledky

### 5.1 Dendrometrické charakteristiky

Na 46 zkusných plochách ve 23 porostních skupinách bylo celkem změřeno 1590 stromů. Celková plocha, na které probíhalo měření, byla 4,06 ha. Veškeré vypočtené charakteristiky jsou uvedené v příloze č. 1. Pro přehlednost jsou zde uvedena rozpětí nejdůležitějších z nich. Střední tloušťka se pohybovala v rozmezí 18,6 cm až 47,4 cm. Střední výška měla rozpětí od 14,6 m do 36,5 m. Objem středního kmene se pohyboval mezi 0,16 m<sup>3</sup> až 2,56 m<sup>3</sup>. Absolutní výšková bonita měla rozpětí od 32 do 38 m.

### 5.2 Základní statistické charakteristiky

Všechny vypočítané statistické charakteristiky pro výčetní tloušťku jsou uvedeny v příloze č. 2 a pro výšku v příloze č. 3. Z vypočtených variačních koeficientů je dobře patrné, že variabilita tlouštěk je větší než variabilita výšek, což potvrzuje obecně uznávané tvrzení o variabilitě těchto dvou základních taxačních veličin. Také je podle hodnot koeficientu šikmosti u výšek vidět, že rozdělení výšek je ve většině případů pravostranné. Pro celý datový soubor je pro tloušťky stromů hodnota aritmetického průměru 32,3 cm, intervalový odhad střední hodnoty pro hladinu spolehlivosti 95 % je 31,8 – 32,8 cm, medián má hodnotu 31 cm, směrodatná odchylka je 10,9 cm, variační koeficient je 33,9 %, koeficient šikmosti je 0,39 a koeficient špičatosti je -0,30. Rozpětí tlouštěk stromů je 9 – 75 cm. Pro výšky stromů je aritmetický průměr 28,4 m, intervalový odhad střední hodnoty pro hladinu spolehlivosti 95 % je 28,1 – 28,7 m, medián má hodnotu 29,5 m, směrodatná odchylka je 6,4 m, variační koeficient je 22,7 %, koeficient šikmosti je -0,46 a koeficient špičatosti je -0,59. Rozpětí výšek stromů je 10 – 41,5 m. Z charakteristik celého datového souboru je patrné, že rozdělení tlouštěk i výšek stromů se sice mírně odchyluje od normálního rozdělení, ale tyto odchylky jsou malé, takže je lze označit za zanedbatelné.

### 5.3 Výsledky průzkumové analýzy dat

Průzkumová analýza dat byla provedena pomocí kvantil-kvantilových grafů a bagplotů. Všechny sestavené grafy jsou uvedeny v přílohách č. 4 a 5 (kvantilové grafy) a v příloze č. 6 (bagploty). Z výsledků průzkumové analýzy dat je patrné, že se v jednotlivých porostních skupinách data odchylují od modelového normálního rozdělení a občas

se vyskytují i mírně odlehle hodnoty, ale tyto odchylky nejsou tak významné, že by data nebyla použitelná pro další analýzy.

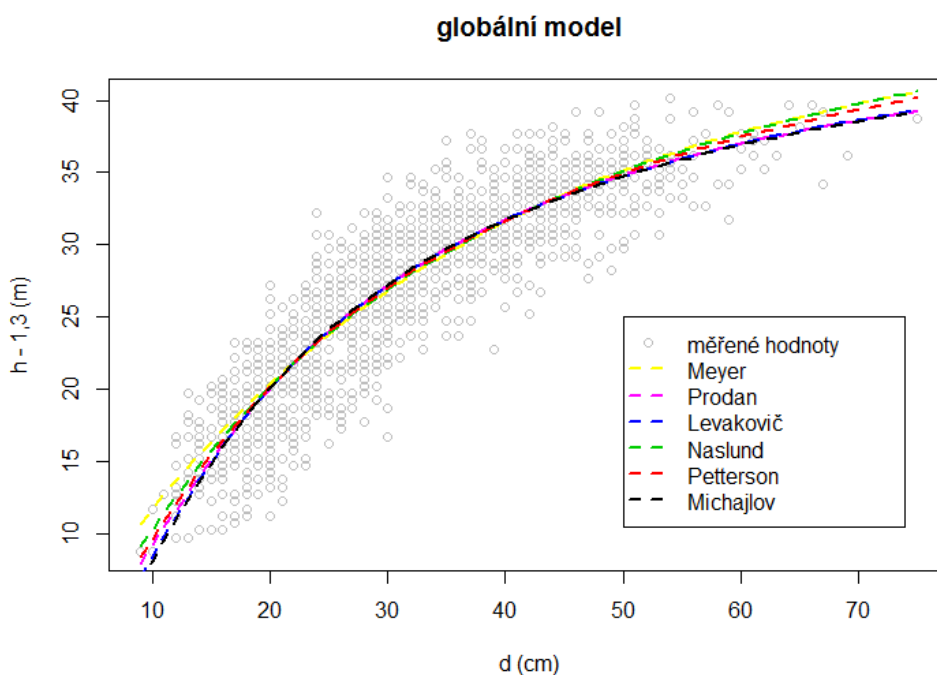
## 5.4 Globální model

Všech šest globálních modelů bylo úspěšně vyřešeno a hodnoty jednotlivých parametrů modelů jsou uvedeny v tabulce č. 6. Grafické zobrazení všech globálních modelů je znázorněno na obrázku č. 13. Z obrázku je patrné, že všechny modely velmi dobře prokládají data a mezi jednotlivými modely jsou pouze malé rozdíly v poloze výsledné výškové křivky. Je patrné, že všechny modely jsou dobře použitelné pro daný datový materiál.

Tab. č. 6: Odhady parametrů globálních modelů

Funkce	Parametr <i>a</i>	Parametr <i>b</i>	Parametr <i>c</i>
Michajlovova	50,041 *** (0,398)	18,253 *** (0,250)	
Pettersonova	0,264 *** (0,001)	2,062 *** (0,029)	
Näslundova	1,793 *** (0,025)	0,133 *** (0,001)	
Levakovičova	50,595 *** (0,408)	18,928 *** (0,259)	
Meyerova	45,493 *** (0,643)	0,029 *** (0,001)	
Prodanova	5,983 *** (0,739)	0,289 *** (0,050)	0,021 *** (0,001)

Legenda: první hodnota v buňce je vlastní hodnota odhadu parametru; \* - \*\*\* udává významnost odhadu parametru při různé hladině významnosti: \* - p-hodnota < 0,05; \*\* - p-hodnota < 0,01; \*\*\* - p-hodnota < 0,001; hodnota v závorce udává střední chybu odhadu.



Obr. č. 13: Srovnání globálních modelů šesti výškových funkcí

## 5.5 Lokální modely

Ze všech použitých modelů výškových funkcí byl podle hodnoty AIC v osmi porostních skupinách jako nejlepší vybrán model Michajlovovy funkce, v sedmi porostních skupinách model Prodanovy funkce, ve třech porostních skupinách modely Meyerovy a Näslundovy funkce a ve dvou porostních skupinách model Pettersonovy funkce. Model Levakovičovy funkce se neprojevil jako nejlepší podle hodnoty AIC ani v jednom případě. Výsledné hodnoty AIC pro všechny modely ze všech 23 porostních skupin jsou uvedeny v tabulce č. 7 – nejnižší (tedy nejlepší) hodnota AIC je zvýrazněna tučně. Grafické znázornění všech lokálních modelů je uvedeno v příloze č. 9.

Tab. č. 7: Vypočtené hodnoty AIC pro nelineární lokální modely výškových funkcí

PSK	Funkce (zkrácené názvy)					
	Michajlov	Petterson	Näslund	Levakovič	Meyer	Prodan
1	194,85	196,22	196,94	195,32	196,18	<b>192,45</b>
2	137,19	<b>137,17</b>	137,18	137,19	137,18	139,17
3	<b>146,78</b>	147,57	148,09	146,98	148,74	148,75
4	<b>159,31</b>	159,98	160,34	159,55	159,40	159,52
5	<b>228,80</b>	230,22	231,08	229,19	231,52	230,45
6	266,06	264,85	<b>264,32</b>	265,73	264,64	263,79
7	370,61	371,68	372,25	370,93	370,73	<b>370,02</b>
8	258,69	258,63	258,73	258,63	<b>257,86</b>	260,52
9	210,68	<b>210,64</b>	210,65	210,66	211,14	212,64
10	366,56	367,89	368,71	366,86	366,36	<b>365,48</b>
11	275,38	277,07	278,03	275,77	277,81	<b>274,75</b>
12	<b>267,60</b>	267,95	268,15	267,68	268,05	269,20
13	182,86	182,72	<b>182,67</b>	182,82	182,80	184,61
14	293,16	292,84	292,68	293,06	295,06	<b>289,13</b>
15	<b>295,88</b>	297,04	297,77	296,07	296,10	296,43
16	298,02	298,66	298,98	298,15	297,74	<b>297,31</b>
17	<b>278,18</b>	278,26	278,30	278,20	278,89	280,21
18	<b>238,54</b>	238,74	238,85	238,59	238,81	240,18
19	337,16	337,02	<b>336,99</b>	337,12	337,27	338,96
20	379,12	379,40	379,56	379,17	<b>378,81</b>	380,11
21	307,01	307,24	307,37	307,08	<b>305,63</b>	308,07
22	277,98	277,27	276,93	277,85	277,36	<b>273,30</b>
23	<b>242,98</b>	243,01	243,05	242,99	243,37	245,03

Legenda: PSK – porostní skupina

(hodnoty AIC pro nejlepší model jsou zvýrazněny tučně)

## 5.6 Výběr vhodného modelu

Pro globální model všech použitých výškových funkcí byly spočítány hodnotící kritéria uvedená v kapitole 4.3.6. Výsledné hodnoty jsou uvedeny v tabulce č. 8. Z výsledných hodnot kritérií je jasně patrné, že mezi všemi funkcemi se nejlépe jeví Prodanova funkce. V příloze č. 7 jsou zobrazeny rezidua všech modelů oproti tloušťce, respektive standardizované tloušťce. Dále jsou v příloze č. 8 zobrazeny vlastnosti reziduí podle decilů vypočítané funkcí mywhiskers. Z grafů v přílohách č. 7 a 8 je dobře patrné, že rezidua globálních modelů mají velmi podobný průběh, což potvrzuje značnou vyrovnanost jednotlivých funkcí.

U lokálních modelů se sice podle hodnot AIC projeví nejlépe modely podle Prodana a Michajlova, ale je nutné zdůraznit, že všech šest modelů se od sebe lišilo podle hodnot AIC pouze nepatrně, a to ve všech porostních skupinách. Pokud tedy budeme vycházet z výše uvedeného tvrzení Burnhama a Andresona (1998) o odlišnosti modelů podle hodnot AIC, tak je jasné, že z lokálních modelů nemůže být ani jeden použitý model označen jako nejlepší. Z grafického zobrazení lokálních modelů je ale dobře patrné, že Prodanova funkce není vhodná pro použití v porostní skupině č. 14 a 22 (viz příloha č. 9), protože na těchto dvou stanovištích nesplňuje požadavky na kvalitní výškovou funkci (funkce musí být neustále rostoucí).

Z pohledu možnosti linearizace vyplývá, že Prodanova funkce již má lineární tvar, Meyerova funkce není linearizovatelná a všechny ostatní funkce jsou linearizovatelné.

Po zvážení všech výše uvedených kritérií byla vyloučena pro další využití Prodanova funkce, protože nesplňovala požadavky na výškovou funkci na některých plochách, dále pak Meyerova funkce, protože nebyla linearizovatelná a také Levakovičova funkce, protože nevyšla v absolutních jednotkách AIC v porovnání lokálních modelů ani v jednom případě jako nejlepší. Zbylé tři funkce – tedy Michajlovova, Pettersonova a Nāslundova – vyšly vyrovnaně na úrovni lokálních modelů, praktické použitelnosti i linearizace, ale podle všech hodnotících kritérií na úrovni globálního modelu se jednoznačně nejlépe projevila Pettersonova výšková funkce, tudíž byla právě ona vybrána jako funkce, která bude dále v práci využívána k různým regresním analýzám.

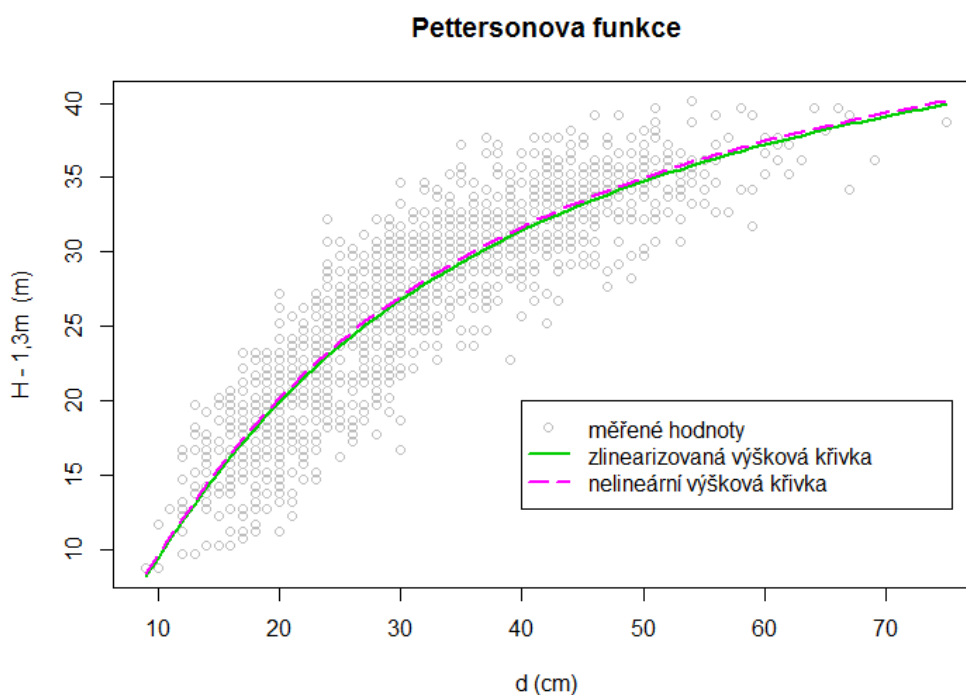
Tab. č. 8: Výsledky srovnání globálních modelů výškových funkcí

Funkce	Střední hodnota reziduí	Směrodatná odchylka reziduí	Střední chyba reziduí	AIC
Michajlovova	0,01033	2,81006	0,07047	7802,8
Pettersonova	-0,00058	2,80230	0,07028	7794,0
Näslundova	-0,00706	2,80809	0,07042	7800,6
Levakovičova	0,00851	2,80692	0,07039	7799,3
Meyerova	-0,03133	2,82442	0,07083	7819,2
Prodanova	0,00720	2,79580	0,07011	7788,7

Legenda: AIC – Akaikeho informační kritérium

### 5.7 Linearizace Pettersonovy výškové funkce

Potvrzení správnosti výběru Pettersonovy funkce je doloženo linearizací této funkce a porovnáním nelineárního a lineárního modelu na všech 23 porostních skupinách a také na globálním modelu. Na lokální ani na globální úrovni se neprojeví žádné výrazné odchylky těchto dvou tvarů funkce. Výsledky srovnání nelineárního a lineárního tvaru Pettersonovy funkce na lokálním modelu jsou uvedeny v příloze č. 10. Srovnání globálního nelineárního a lineárního modelu je zobrazené na obrázku č. 14.



Obr. č. 14: Srovnání nelineárního a lineárního globálního modelu Pettersonovy výškové funkce

## 5.8 Smíšený regresní model

### 5.8.1 Výsledný smíšený model výškové křivky

První fází při tvorbě modelu bylo vypočítat globální model. Výsledný spočítaný model lze zapsat rovnicí

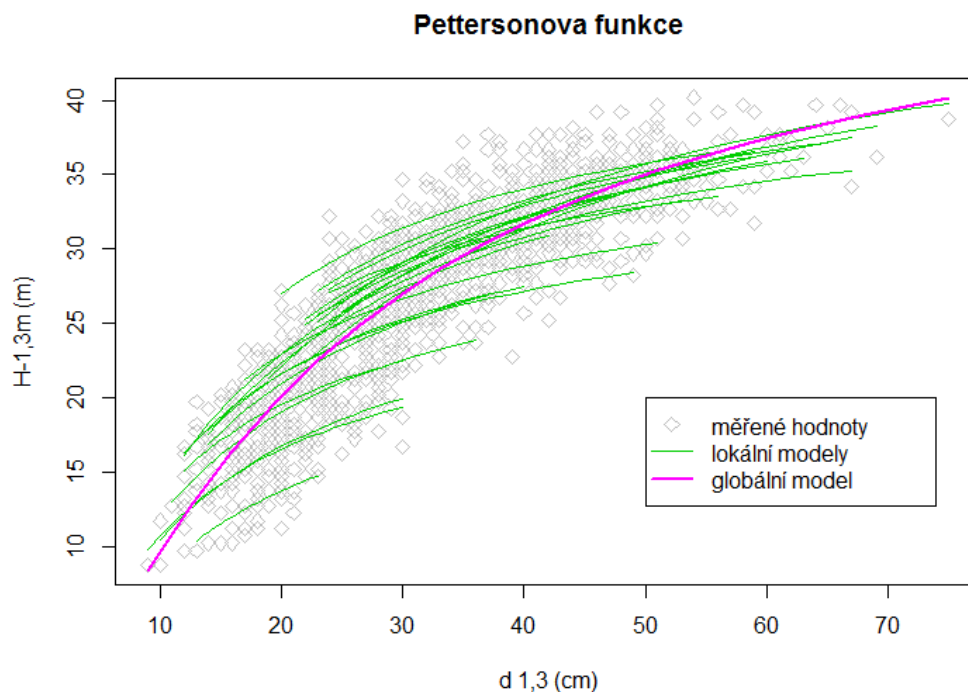
$$\frac{1}{\sqrt[3]{h-1,3}} = 0,2648 + 2,0848 * \frac{1}{d} \quad (5.1)$$

Výsledný model byl jako celek statisticky významný (p-hodnota < 0,001), oba dva odhady parametrů byly statisticky významné (p-hodnota < 0,001), odhady parametrů jsou uvedeny ve výše uvedené rovnici 5.1. Koeficient determinace měl hodnotu 0,781, takže celý model vysvětlil více než 78 % variability měřených výšek. Hodnota AIC byla -8861,3. Směrodatná odchylka reziduí nabyla hodnoty 0,0149. Výsledný lineární globální model je zobrazen na obrázku č. 15 a pro srovnání jsou zde uvedeny i lokální lineární modely Petersonovy funkce. Z grafu je patrné, že globální model příliš dobře nepostihuje odchylky jednotlivých lokálních modelů, a to především u stromů s tloušťkou přibližně do 35 cm.

Druhou fází byl výpočet modelu průměru. Z výsledků vyplývá, že rozdělení náhodných efektů má rozdělení  $\varepsilon_{ik} \sim N(0; 0,0138^2)$  a  $u_{ai} \sim N(0; 0,0317^2)$ . Jelikož je hodnota rozptylu  $\tau^2$  vyšší než 0, tak je jasné, že se v datech vyskytuje heterogenita, což jen potvrzuje volbu smíšeného modelu jako vhodného řešení. Druhým důvodem, proč je vhodné použít smíšený model, je to, že vnitroskupinový korelační koeficient byl vypočítán ve výši 0,840, což potvrzuje přítomnost závislosti mezi daty uvnitř jednotlivých porostních skupin, a to znamená porušení jedné z podmínek pro použití lineární regrese. Při porovnání rozptylů uvnitř porostních skupin  $\sigma^2$  a mezi porostními skupinami  $\tau^2$  bylo zjištěno, že rozptyl uvnitř porostních skupin je menší než rozptyl mezi skupinami. To znamená, že je vhodné do modelu zařadit i prediktory 2. úrovně. Hodnota absolutního členu byla určena 0,3399 a na základě hodnocení p-hodnoty je tento člen považován za statisticky významný (p-hodnota = 0). AIC má hodnotu -8955,7.

Ve třetí fází byl vypočítán model s náhodným absolutním členem a již obsahoval prediktor 1. úrovně. Odhady parametrů byly stanoveny pro absolutní člen 0,2952 a pro regresní člen 1,2624. Oba dva parametry byly statisticky významné (p-hodnota = 0). Směrodatná odchylka náhodného efektu absolutního členu byla 0,0192 a směrodatná odchylka reziduí byla 0,0076. AIC mělo hodnotu -10863,2. Při porovnání tohoto modelu s globálním modelem vyšla p-hodnota testu poměru věrohodností menší než 0,0001,

takže model s náhodným absolutním členem je jednoznačně lepší než globální model. Při testu poměru věrohodností mezi modelem průměru a modelem s náhodným absolutním členem vyšla p-hodnota menší než 0,0001, což potvrzuje hypotézu, že zanesení prediktoru 1. úrovně (v tomto případě výčetní tloušťky) výrazně zlepšilo model. Obě dvě tato tvrzení potvrzuje i výrazně nižší hodnota AIC v porovnání s hodnotami AIC obou srovnávaných modelů.



*Obr. č. 15: Srovnání globálního modelu a lokálních modelů Pettersonovy výškové funkce po linearizaci*

Čtvrtá fáze spočívala ve výpočtu modelu s náhodným absolutním i regresním členem. Oba dva parametry byly podle p-hodnoty (p-hodnota = 0) statisticky významné a jsou zapsány v níže uvedené rovnici 5.2. Vypočítaná směrodatná odchylka náhodného efektu absolutního členu byla 0,0194, směrodatná odchylka náhodného efektu regresního členu 0,1726 a směrodatná odchylka reziduí 0,0075. Korelační koeficient nabyl hodnoty -0,282. Kovarianční člen kovarianční matice náhodných členů měl hodnotu -0,0000032. Hodnota AIC byla -10881,2. Model lze zapsat pomocí rovnic

$$\frac{1}{\sqrt[3]{h-1,3}} = (0,2944 + u_{ai}) + (1,2919 + u_{bi}) * \frac{1}{d} + \varepsilon_{ik} \quad (5.2)$$

$$\mathbf{u}_i = \begin{bmatrix} u_{ai} \\ u_{bi} \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0,0003764 & -0,0000032 \\ -0,0000032 & 0,0297739 \end{bmatrix} \right) \quad (5.3)$$

$$\varepsilon_{ik} \sim N(0, 0,0000558) \quad (5.4)$$

Při porovnání s modelem, který měl náhodnou část pouze u absolutního členu, bylo zjištěno, že model s náhodným absolutním i regresním členem je výrazně lepší. Rozhodnutí vychází z testu poměru věrohodností, jehož p-hodnota byla menší než 0,0001. Stejný výsledek ukazuje u porovnání hodnot AIC, které je pro model s pouze náhodným absolutním členem -10863,2 a pro model s náhodným absolutním i regresním členem -10881,2. Z grafického porovnání smíšených modelů s náhodným pouze absolutním členem a s oběma náhodnými členy a s lokálními modely je dobře patrné, že k velmi dobré shodě dochází u modelů lokálních a modelu s oběma náhodnými členy. Model, který má náhodný pouze absolutní člen, se na některých plochách odchyluje dost výrazně od lokálního modelu. Je tedy jasné, že přidání náhodné části i regresnímu členu výrazně zlepšilo model. Výsledky toho grafického srovnání jsou uvedeny v příloze č. 11. V tabulce č. 9 jsou uvedeny odhady parametrů a dalších charakteristik čtyř dílčích modelů první úrovně. V tabulce č. 10 jsou uvedeny hodnoty náhodných částí parametrů smíšeného modelu s náhodným absolutním i regresním členem a také konečné hodnoty parametrů pro jednotlivé porostní skupiny.

*Tab. č. 9: Odhady parametrů a dalších charakteristik dílčích modelů první úrovně*

<b>Dílčí model</b>	<b>Globální model</b>	<b>Model průměru</b>	<b>Model s náhodným absolutním členem</b>	<b>Model s náhodným absolutním i regresním členem</b>
$a$	0,2648	0,3999	0,2952	0,2944
$b$	2,0848	-----	1,2624	1,2919
$\sigma_{u_{ai}}$	-----	0,0317	0,0192	0,0194
$\sigma_{u_{bi}}$	-----	-----	-----	0,1726
$\varepsilon_{ik}$	0,0149	0,0138	0,0076	0,0001
<b>AIC</b>	-8861,3	-8955,7	-10863,2	-10881,2

Legenda:  $a$  – pevná část odhadu absolutního parametru,  $b$  – pevná část odhadu regresního parametru,  $\sigma_{u_{ai}}$  – směrodatná odchylka odhadu náhodné části absolutního parametru,  $\sigma_{u_{bi}}$  – směrodatná odchylka odhadu náhodné části regresního parametru,  $\varepsilon_{ik}$  - směrodatná odchylka reziduí, AIC – Akaikeho informační kritérium

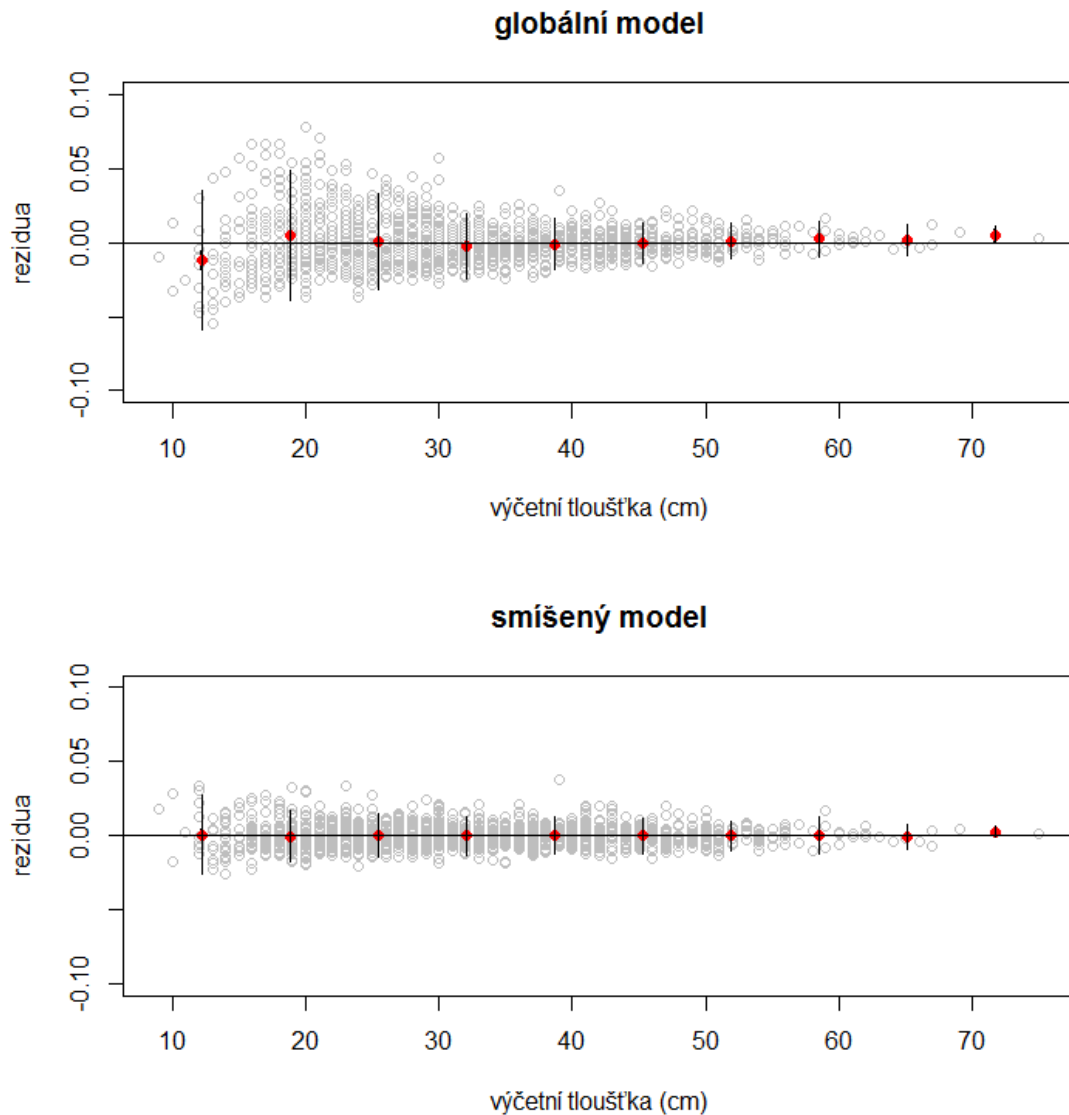


Tab. č. 10: Hodnoty náhodných složek a celkových hodnot parametrů smíšeného modelu

PSK	$u_{ai}$	$u_{bi}$	$a$	$b$
1	0,0359	-0,0469	0,3304	1,2449
2	0,0501	0,1750	0,3445	1,4669
3	0,0297	0,0435	0,3242	1,3354
4	0,0235	-0,2264	0,3180	1,0655
5	0,0177	-0,0415	0,3122	1,2503
6	0,0049	-0,0230	0,2994	1,2689
7	0,0080	-0,1519	0,3024	1,1399
8	0,0103	-0,1988	0,3047	1,0931
9	0,0028	-0,1661	0,2973	1,1258
10	-0,0083	0,0207	0,2862	1,3126
11	-0,0083	0,0819	0,2862	1,3738
12	-0,0122	-0,0044	0,2823	1,2875
13	-0,0139	-0,0354	0,2806	1,2565
14	-0,0059	-0,1635	0,2885	1,1284
15	-0,0185	0,2612	0,2759	1,5530
16	-0,0120	0,1849	0,2745	1,4768
17	-0,0086	-0,0444	0,2859	1,2475
18	-0,0143	-0,0643	0,2801	1,2276
19	-0,0126	-0,0284	0,2819	1,3203
20	-0,0137	0,1216	0,2808	1,4135
21	-0,0140	-0,1781	0,2805	1,1138
22	-0,0165	0,1758	0,2780	1,4677
23	-0,0162	0,2517	0,2782	1,5436

Legenda: PSK – porostní skupina,  $u_{ai}$  – náhodná část odhadu absolutního parametru,  $u_{bi}$  – náhodná část odhadu regresního parametru,  $a$  – celkový odhad absolutního parametru,  $b$  – celkový odhad regresního parametru.

Na obrázku č. 16 jsou znázorněna rezidua globálního a smíšeného modelu. Z jejich srovnání je jasně vidět, že smíšený model přispěl k výraznému snížení hodnot reziduí a také přiblížení jejich středních hodnot více k nule. QQ graf reziduí potvrdil normalitu rozdělení reziduí, drobné odchýlení je pouze u nejvyšších hodnot. Z grafu reziduí je dobře patrné, že rezidua splňují podmínku homoskedasticity, protože jejich rozsah je u všech tloušťek přibližně konstantní, pouze u vyšších hodnot je díky nižší četnosti dat bodové pole méně husté. Oba grafy jsou v příloze č. 12.



*Obr. č. 16: Srovnání reziduí globálního a smíšeného modelu*

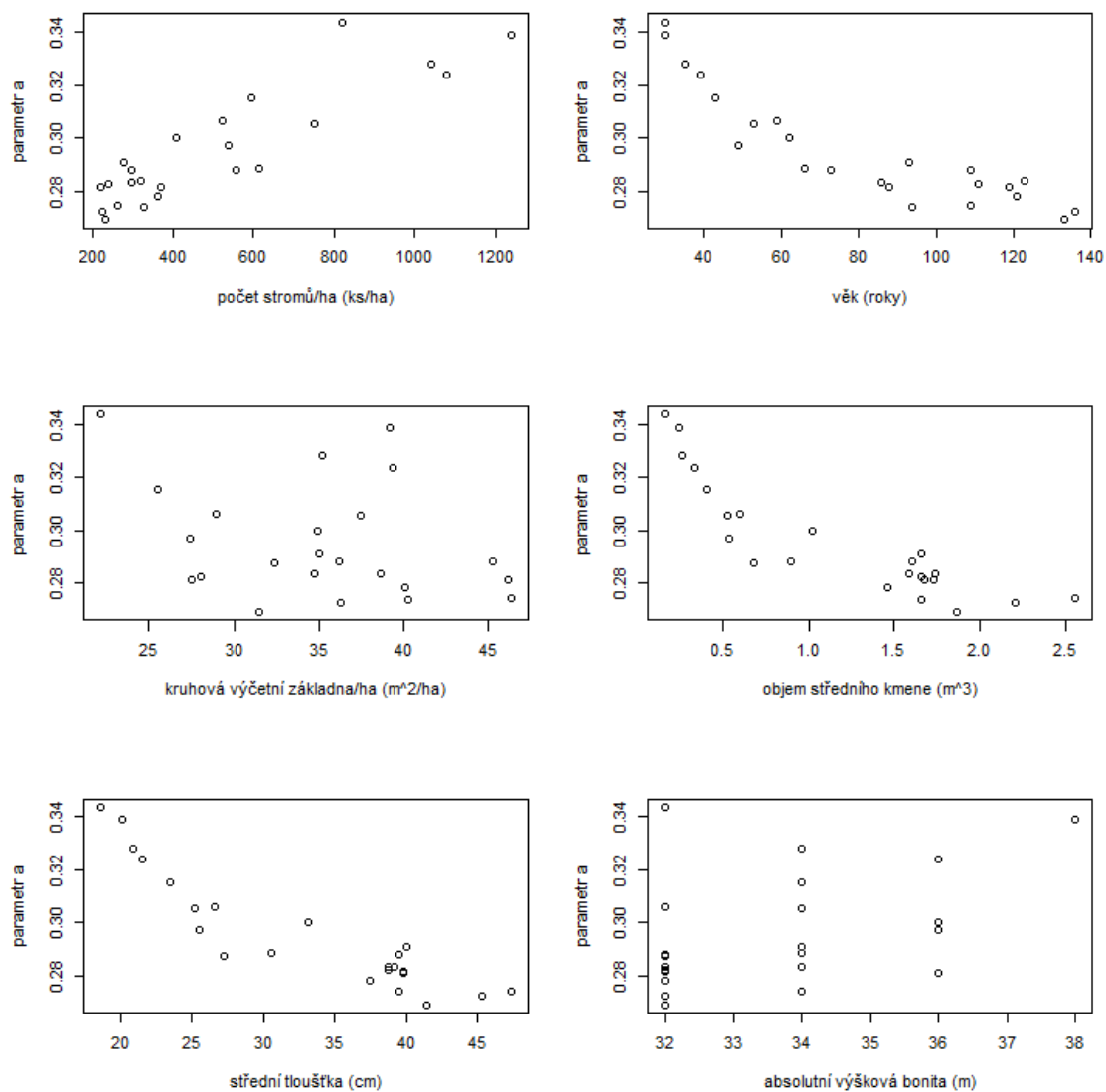
Legenda: červené body - střední hodnota reziduí, tenké černé čáry - intervaly spolehlivosti reziduí.

Dalším bodem tvorby modelů bylo vytvoření modelu druhé úrovně přidáním porostních proměnných. Na obrázcích č. 17 až 20 je znázorněna závislost jednotlivých proměnných a jejich logaritmů na parametrech modelu. Z grafů je jasně patrné, že závislost mezi parametrem  $a$  a porostními veličinami i jejich logaritmy je viditelně těsnější než mezi parametrem  $b$  a porostními veličinami i jejich logaritmy. Hodnoty korelačních koeficientů závislosti parametrů modelu na porostních veličinách a jejich logaritmech jsou uvedeny v tabulce č. 11. Součástí tabulky č. 11 jsou i testová kritéria pro shodu korelačních koeficientů. Z výsledků tohoto testu je jasně patrné, že je prokazatelný rozdíl v hodnotách korelačních koeficientů pouze u závislosti věku nebo jeho logaritmu

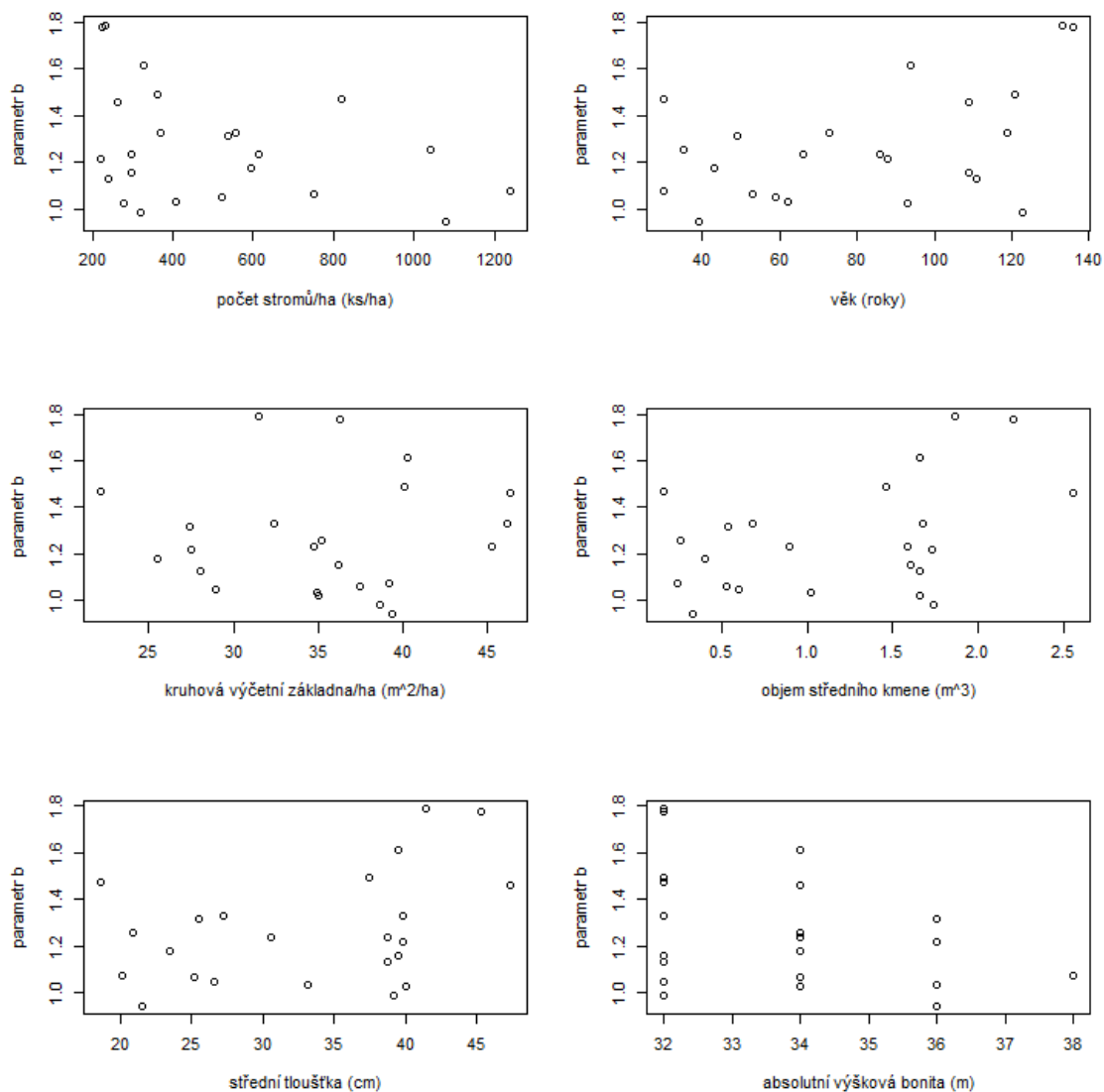
a parametru  $a$  i  $b$  a závislosti objemu středního kmene nebo jeho logaritmu a parametru  $b$ . Jako porostní proměnné, které budou vstupovat do modelu druhé úrovně jako první, byly tedy vybrány věk porostní skupiny ( $T$ ) pro parametr  $b$  a logaritmus věku porostní skupiny ( $\ln T$ ) pro parametr  $a$ . Důvodem jsou nejvyšší hodnoty korelačních koeficientů.

Tab. č. 9: Hodnoty korelačních koeficientů závislosti hodnot parametrů modelu první úrovně a porostních proměnných (a jejich logaritmů) a testových kritérií testu na shodu korelačních koeficientů

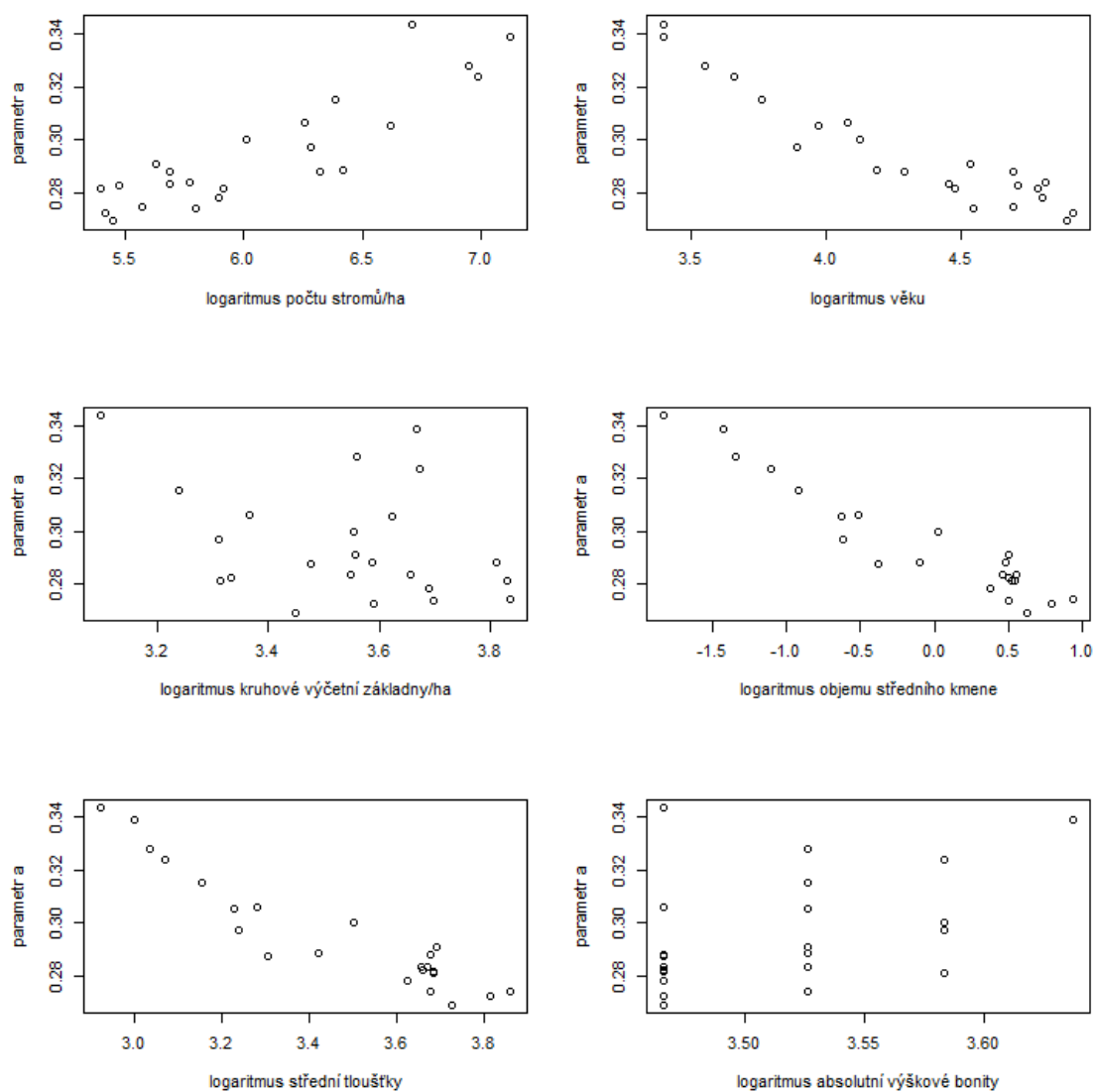
Proměnná	Parametr $a$		Parametr $b$	
	Korelační koeficient	Testové kritérium $Z_k$	Korelační koeficient	Testové kritérium $Z_k$
počet stromů/ha	0,8877		0,3510	
kruhová výčetní základna/ha	0,3189		0,0224	
střední tloušťka	0,9023		0,3902	
věk porostu	0,8859		0,4574	
absolutní výšková bonita	0,4026		0,3802	
objem středního kmene	0,8717		0,4162	
logaritmus počtu stromů/ha	0,8785		0,2592	
logaritmus kruhové výčetní základny/ha	0,3395	0,5803	0,0015	0,5887
logaritmus střední tloušťky	0,9296	0,7690	0,3499	1,1352
logaritmus věku porostu	0,9437	1,9881	0,3848	2,0450
logaritmus absolutní výškové bonity	0,3969	0,1605	0,3802	0,0000
logaritmus objemu středního kmene	0,9311	1,1673	0,3055	3,1183
kritická hodnota $z_{\alpha/2} = 1,96$ ; $H_0 = \rho_1 = \rho_2$ , $H_1 = \rho_1 \neq \rho_2$				



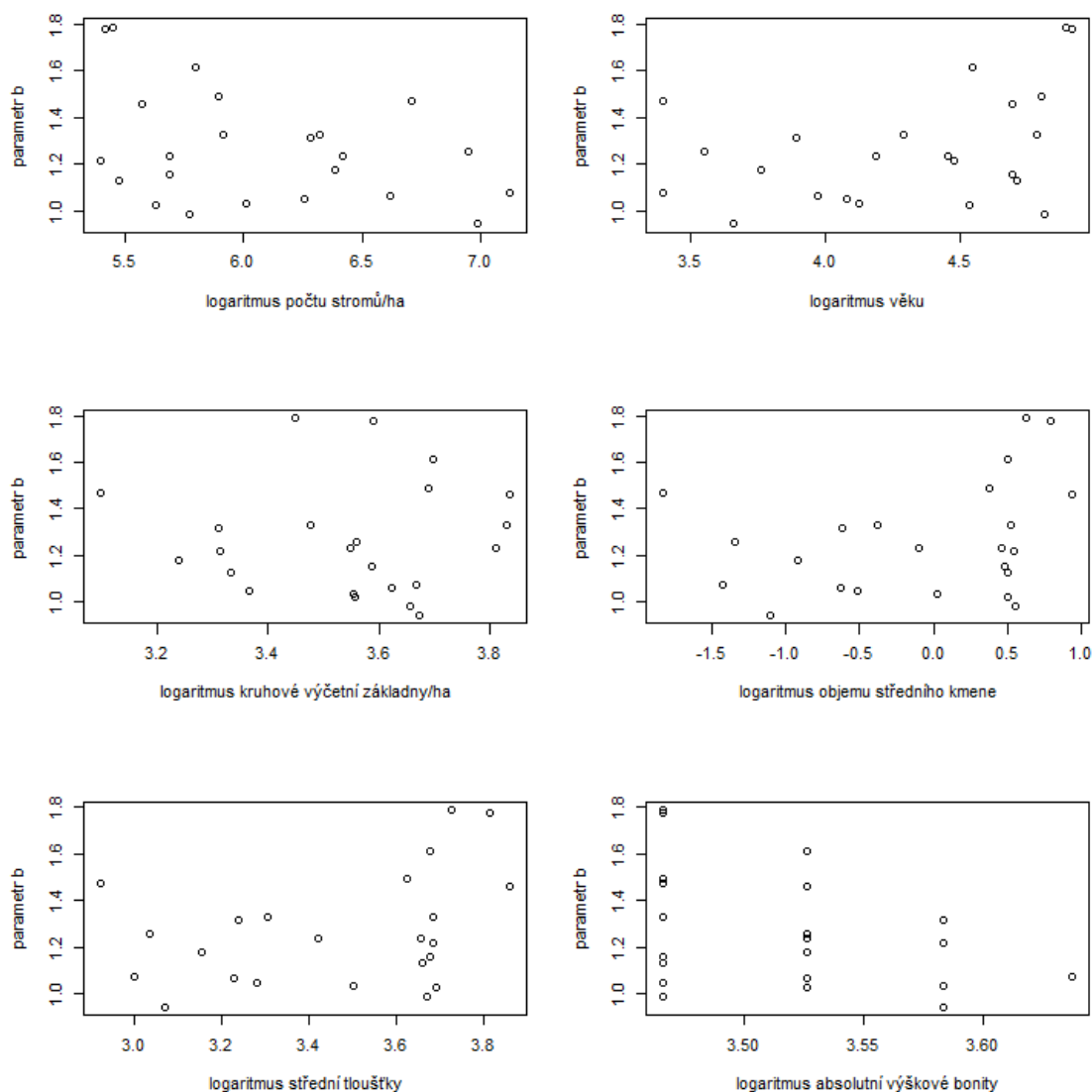
Obr. č. 17: Grafy závislosti parametru  $a$  modelu první úrovně na porostních veličinách



Obr. č. 18: Grafy závislosti parametru  $b$  modelu první úrovně na porostních veličinách



Obr. č. 19: Grafy závislosti parametru  $a$  modelu první úrovně na logaritmech porostních veličin



Obr. č. 20: Grafy závislosti parametru  $b$  modelu první úrovně na logaritmech porostních veličin

Model druhé úrovně, kde je přidán věk porostu (logaritmus věku porostu pro parametr  $a$ ) má všechny parametry statisticky významné ( $p$ -hodnota = 0 u absolutního členu a regresních členů tloušťky a logaritmu věku,  $p$ -hodnota = 0,0047 u věku). Směrodatná odchylka náhodného efektu absolutního členu vyšla 0,0055, směrodatná odchylka náhodného efektu regresního členu 0,1759 a směrodatná odchylka reziduí 0,0075. Korelační koeficient vyšel -0,066. Kovarianční člen kovarianční matice náhodných členů byl vypočítán na hodnotu -0,0000001. Hodnota AIC byla -10927,7. Výsledný model je možné zapsat rovnicemi

$$\frac{1}{\sqrt[3]{h-1,3}} = (0,4572 - 0,0381 * \ln T + u_{ai}) + (1,1109 + 0,0025 * T + u_{bi}) * \frac{1}{d} + \varepsilon_{ik} \quad (5.5)$$

$$\mathbf{u}_i = \begin{bmatrix} u_{ai} \\ u_{bi} \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0,0000303 & -0,0000001 \\ -0,0000001 & 0,0309408 \end{bmatrix} \right) \quad (5.6)$$

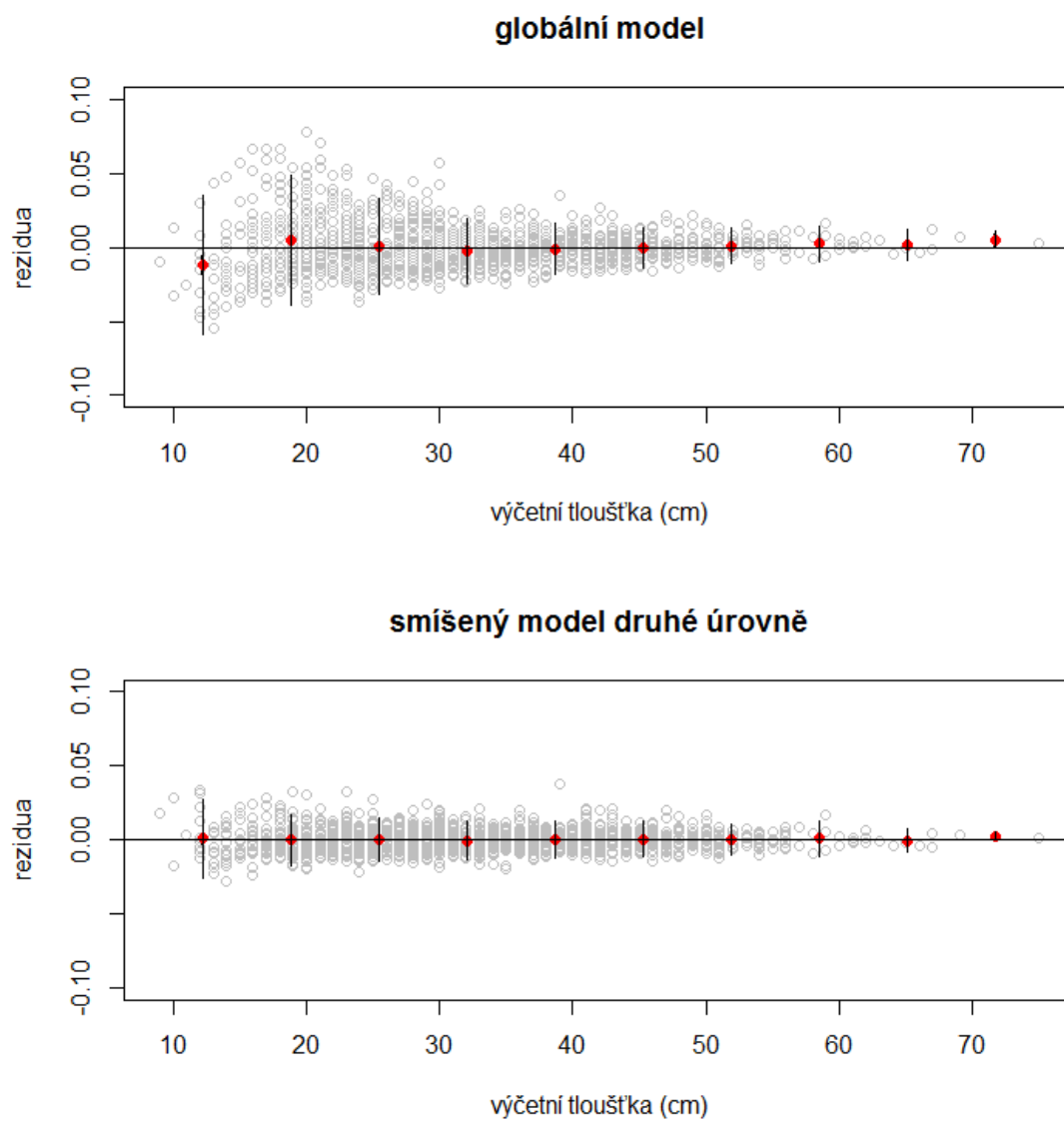
$$\varepsilon_{ik} \sim N(0, 0,0000563) \quad (5.7)$$

Při porovnání s modelem první úrovně, který měl náhodnou část u absolutního i regresního členu, bylo zjištěno, že model druhé úrovně se zahrnutím porostní proměnné věku (a jeho logaritmu) je výrazně lepší. Tento výsledek je dán testem poměru věrohodností, jehož p-hodnota byla menší než 0,0001. Bylo také opět provedeno srovnání reziduí globálního modelu a smíšeného modelu druhé úrovně. Výsledek je uveden na obrázku č. 21. Byl proveden test významnosti korelačních koeficientů závislosti hodnot parametrů modelů druhé úrovně a zbylých porostních veličin. Z výsledků uvedených v tabulce č. 12 je patrné, že mezi hodnotami parametrů modelu druhé úrovně a ostatními porostními proměnnými není závislost. Tyto výsledky jsou patrné i z obrázku č. 22 a 23, kde jsou znázorněny závislosti hodnot parametrů modelu druhé úrovně a porostních veličin. Hodnoty náhodných částí odhadů parametrů a celkové hodnoty odhadů parametrů jsou uvedeny v tabulce č. 13. Výsledný smíšený model je tedy takový, který obsahuje náhodnou část u absolutního i regresního členu, obsahuje jako prediktor druhé úrovně věk porostu respektive logaritmus věku porostu u absolutního členu. Grafická podoba finálního smíšeného modelu v porovnání s lokálními modely je zobrazena v příloze č. 13.

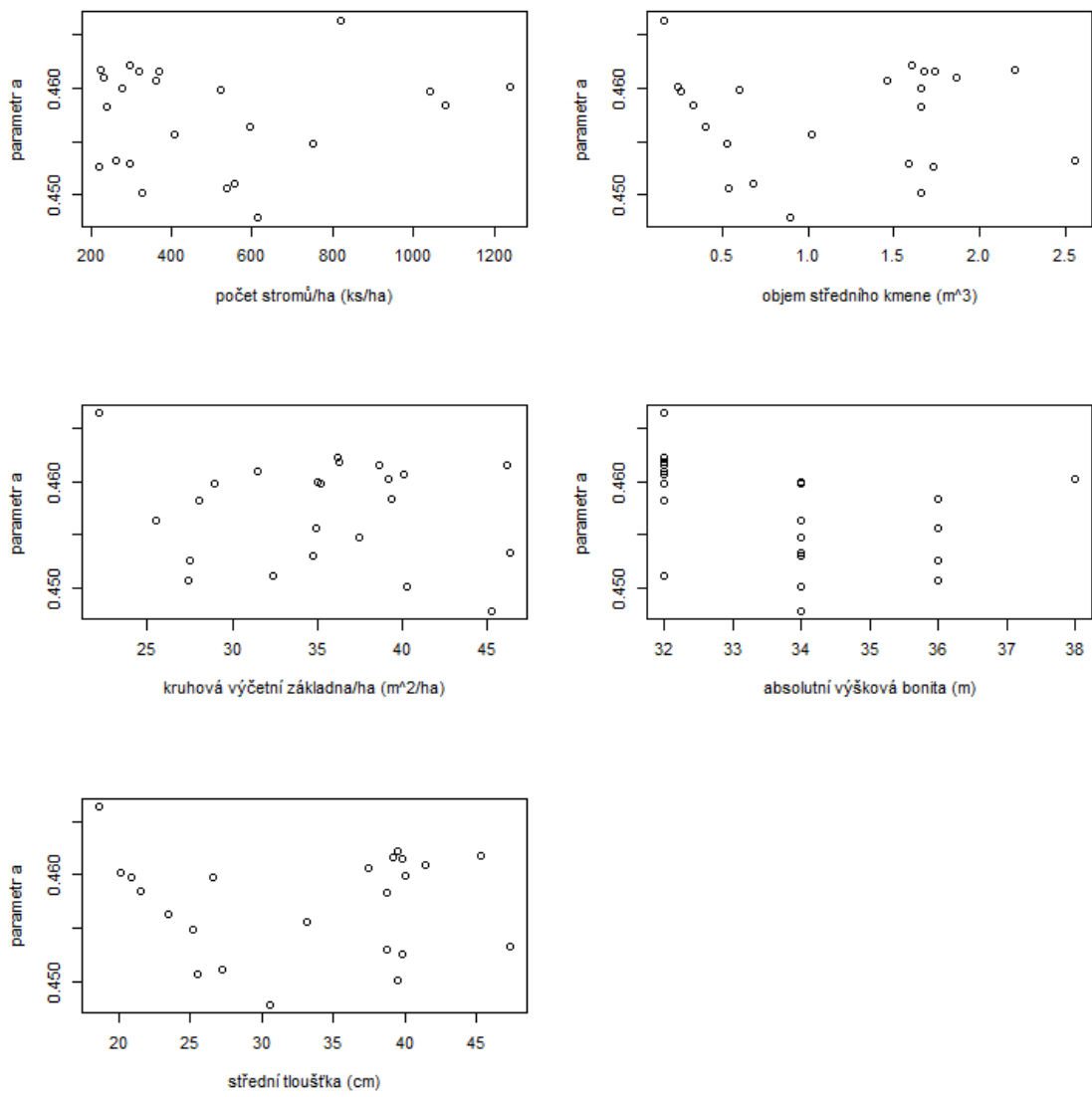
*Tab. č. 10: Hodnoty korelačních koeficientů závislosti parametrů modelu druhé úrovně a porostních veličin a hodnoty testových kritérií testů významnosti těchto korelačních koeficientů*

Proměnná	Parametr a		Parametr b	
	Korelační koeficient	Testové kritérium $T_k$	Korelační koeficient	Testové kritérium $T_k$
počet stromů/ha	0,1081	0,4985	0,1327	0,6137
kruhová výčetní základna/ha	0,1696	0,7887	0,0971	0,4469
střední tloušťka	0,0423	0,1942	0,0806	0,3709
absolutní výšková bonita	0,4012	2,0074	0,2519	1,1928
Objem středního kmene	0,0133	0,0611	0,0291	0,1337
Kritická hodnota $T_{\alpha/2} = 2,08$ ; $H_0 = \rho_1 = 0$ , $H_1 = \rho_1 \neq 0$				

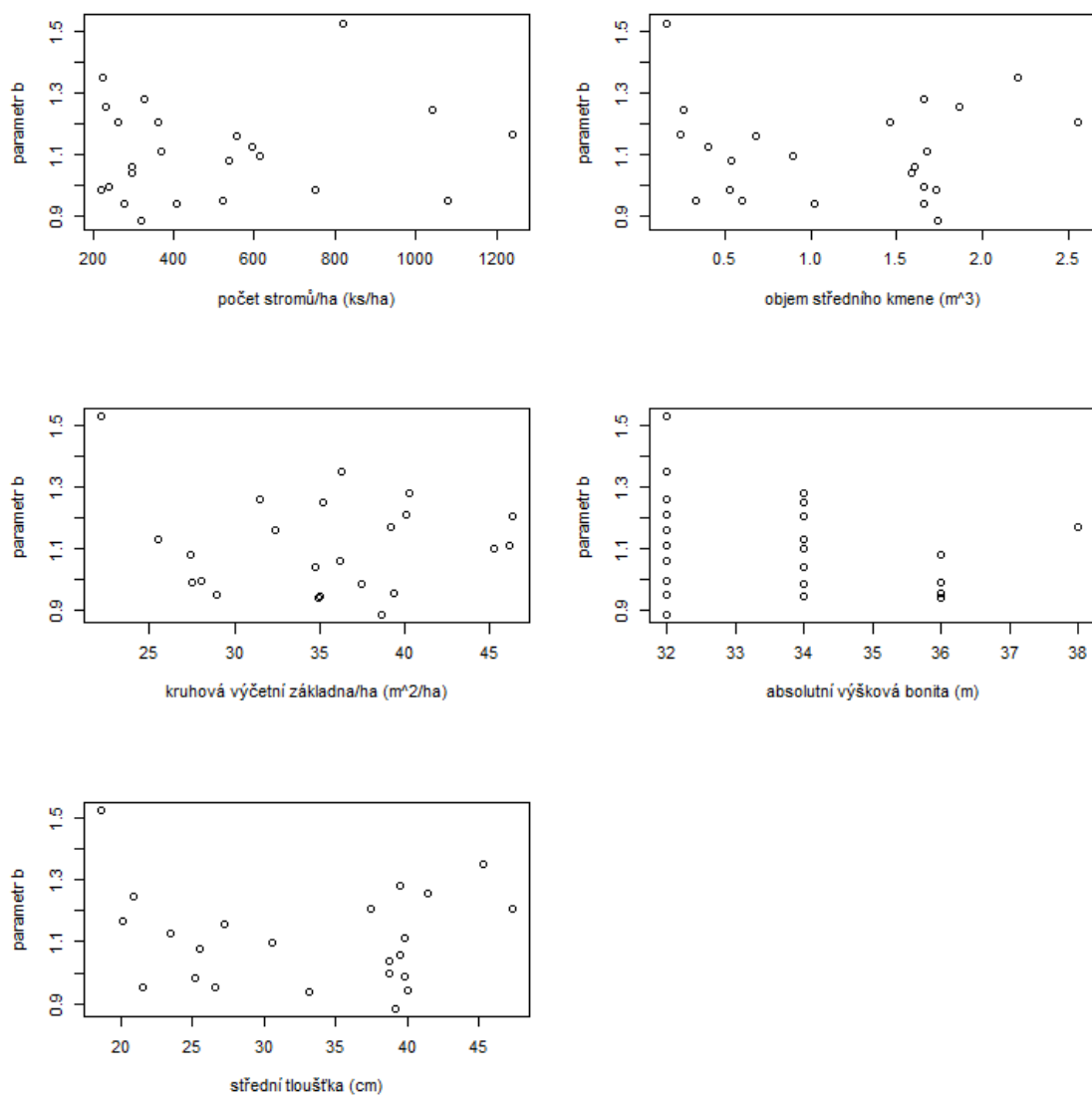




*Obr. č. 21: Srovnání reziduí globálního modelu a smíšeného modelu druhé úrovně*  
 Legenda: červené body - střední hodnota reziduí, tenké černé čáry - intervaly spolehlivosti reziduí.



Obr. č. 22: Grafy závislosti parametru *a* modelu druhé úrovně a porostních veličin



Obr. č. 23: Graf závislosti parametru  $b$  modelu druhé úrovně a porostních veličin

Tab. č. 11: Hodnoty náhodných složek a celkových hodnot parametrů smíšeného modelu druhé úrovně (zahrnut faktor věku)

PSK	$u_{ai}$	$u_{bi}$	$a$	$b$
1	0,0029	0,0563	0,4601	1,1673
2	0,0091	0,4163	0,4663	1,5273
3	0,0025	0,1358	0,4597	1,2467
4	0,0011	-0,1590	0,4583	0,9519
5	-0,0009	0,0161	0,4563	1,1271
6	-0,0065	-0,0323	0,4507	1,0786
7	-0,0024	-0,1270	0,4548	0,9840
8	0,0025	-0,1596	0,4598	0,9513
9	-0,0016	-0,1719	0,4556	0,9389
10	-0,0094	-0,0130	0,4478	1,0979
11	-0,0061	0,0476	0,4511	1,1586
12	-0,0042	-0,0722	0,4530	1,0386
13	-0,0047	-0,1236	0,4526	0,9873
14	0,0027	-0,1690	0,4599	0,9419
15	-0,0070	0,1699	0,4501	1,2809
16	-0,0040	0,0954	0,4533	1,2063
17	0,0049	-0,0516	0,4622	1,0592
18	0,0010	-0,1148	0,4582	0,9961
19	0,0043	-0,0008	0,4615	1,1101
20	0,0034	0,0969	0,4606	1,2079
21	0,0043	-0,2269	0,4615	0,8841
22	0,0037	0,1470	0,4609	1,2580
23	0,0045	0,2402	0,4617	1,3512

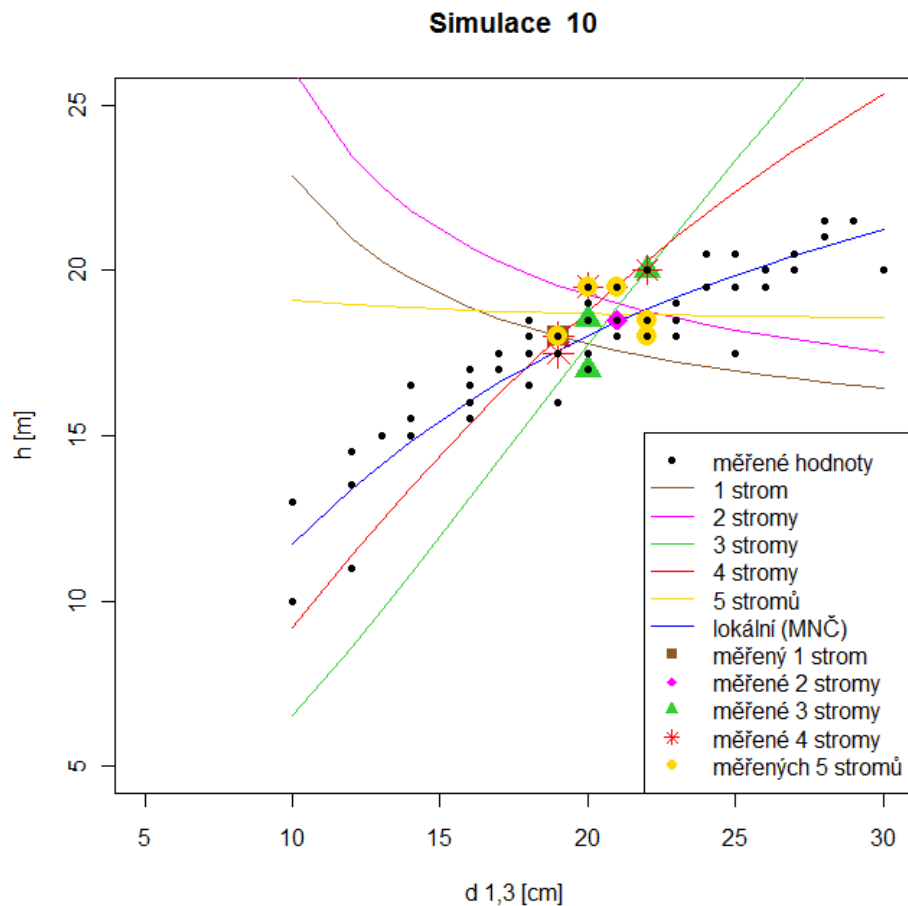
Legenda: PSK – porostní skupina,  $u_{ai}$  – náhodná část odhadu absolutního parametru,  $u_{bi}$  – náhodná část odhadu regresního parametru,  $a$  – celkový odhad absolutního parametru,  $b$  – celkový odhad regresního parametru

## 5.8.2 Výsledky kalibrace smíšeného modelu výškové křivky

### 5.8.2.1 Výsledky kalibrace s výběrem stromů se střední tloušťkou

První varianta kalibrace, která byla prováděna tak, že byly měřeny stromy se střední tloušťkou  $\bar{d} \pm 2$  cm v počtu 1 – 5 stromů, se projevila jako nevyhovující. Důvodem je to, že stromy byly koncentrovány pouze do jednoho úzkého intervalu a čím nižší byly odchylky jednotlivých tlouštěk od střední tloušťky, tím více se model odchyloval od lokálního modelu a přibližoval se více modelu průměru. Často se také stalo, že pokud byl vybrán jako tenčí strom nadúrovňový jedinec (v rámci daného tloušťkového intervalu s vyšší výškou) a jako tlustší strom podúrovňový jedinec (v rámci daného tloušťkového

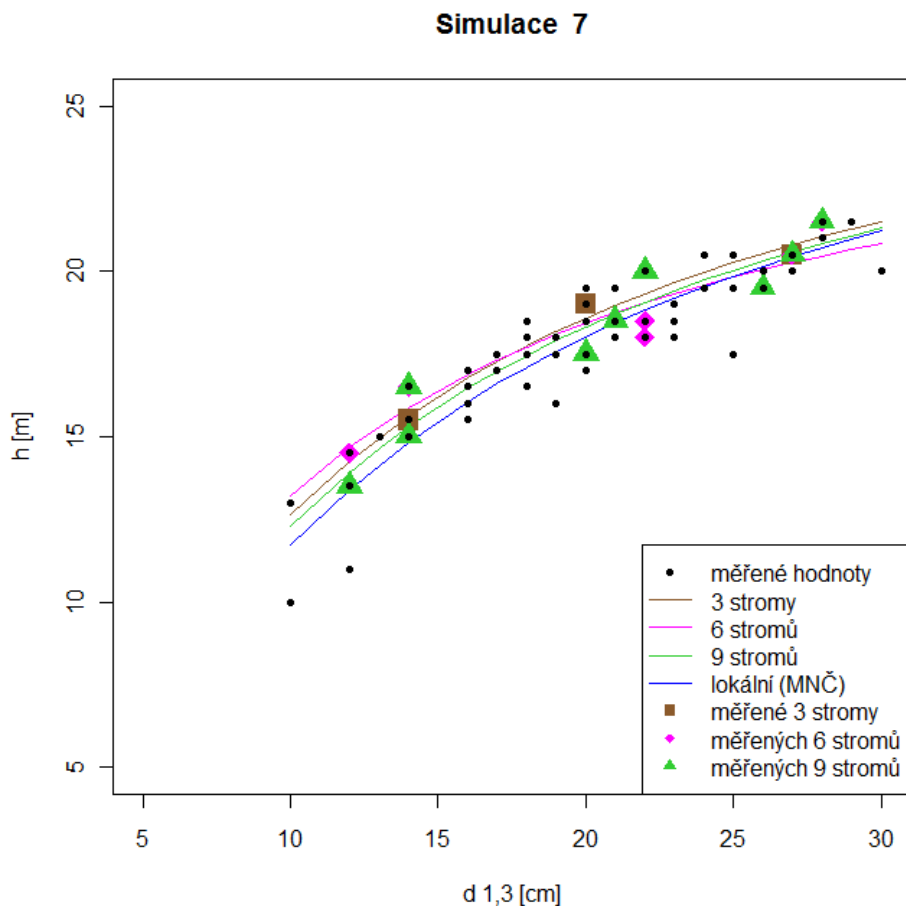
intervalu s nižší výškou), tak výsledný model vyšel proti smyslu lokálního modelu jako klesající křivka a nemohl být tedy ani považován za model, který splňuje předpoklady kladené na výškovou křivku. Toto nastávalo i v případech, kdy byl měřen vyšší počet stromů (4 nebo 5). Lze tedy konstatovat, že kalibrace provedená i na vyšším počtu stromů, které nejsou vybrány náhodně, ale podle střední tloušťky porostu, se prokázala jako většinou nevyhovující a pro běžné praktické použití se nedá doporučit. Grafické výsledky všech provedených 60 kalibrací (šest porostních skupin a v každé 10 simulací výběru stromů) jsou uvedeny v přílohách č. 14, 16, 18, 20, 22 a 24. Z důvodu velmi nevyhovujících výsledků kalibrace nejsou v této práci uvedeny žádné výsledné hodnoty hodnotících kritérií uvedených v kapitole 4.3.7.2, protože při jejich výpočtu nabývaly zcela neadekvátních hodnot (např. příliš vysoké hodnoty AIC v porovnání s lokálním modelem, nízké hodnoty koeficientu determinace nebo příliš vysoké střední hodnoty odchylek modelových hodnot získaných smíšeným modelem od modelových hodnot lokálního modelu). Příklad zcela nevyhovující kalibrace v porostní skupině č. 3 je uveden na obrázku č. 24.



*Obr. č. 24: Grafický výsledek nevyhovující kalibrace při výběru 1-5 stromů se střední tloušťkou v porostní skupině č. 3*

### 5.8.2.2 Výsledek kalibrace s výběrem stromů ve třech tloušťkových intervalech

Druhá varianta kalibrace, při níž byl proveden výběr stromů k měření ve třech tloušťkových intervalech (oblast nejnižších, nejvyšších a průměrných tlouštěk) a to v počtu tří, šesti nebo devíti stromů (jeden, dva nebo tři stromy na tloušťkový interval) se prokázala jako významně lepší než varianta 1. Kompletní výsledky pro jednotlivé porostní skupiny v grafickém zobrazení jsou uvedeny v přílohách č. 15, 17, 19, 21, 23 a 25. V přílohách č. 26 až 31 jsou tabelovány vypočítané hodnoty kritérií pro hodnocení kvality kalibrace pro jednotlivé porostní skupiny. Z výsledků je patrné, že žádná z možností různého počtu stromů se nedá nazvat jako absolutně nevyhovující. Je možné najít simulace, kde se výběr pouze tří stromů téměř shoduje s lokálním modelem, čemuž odpovídají i hodnotící kritéria (např. porostní skupina 3, simulace 9). Stejně tak je tomu i u výběru šesti nebo devíti stromů. Příklad kalibrací s velmi podobným (a kvalitním) výsledkem je uveden na obrázku č. 25.

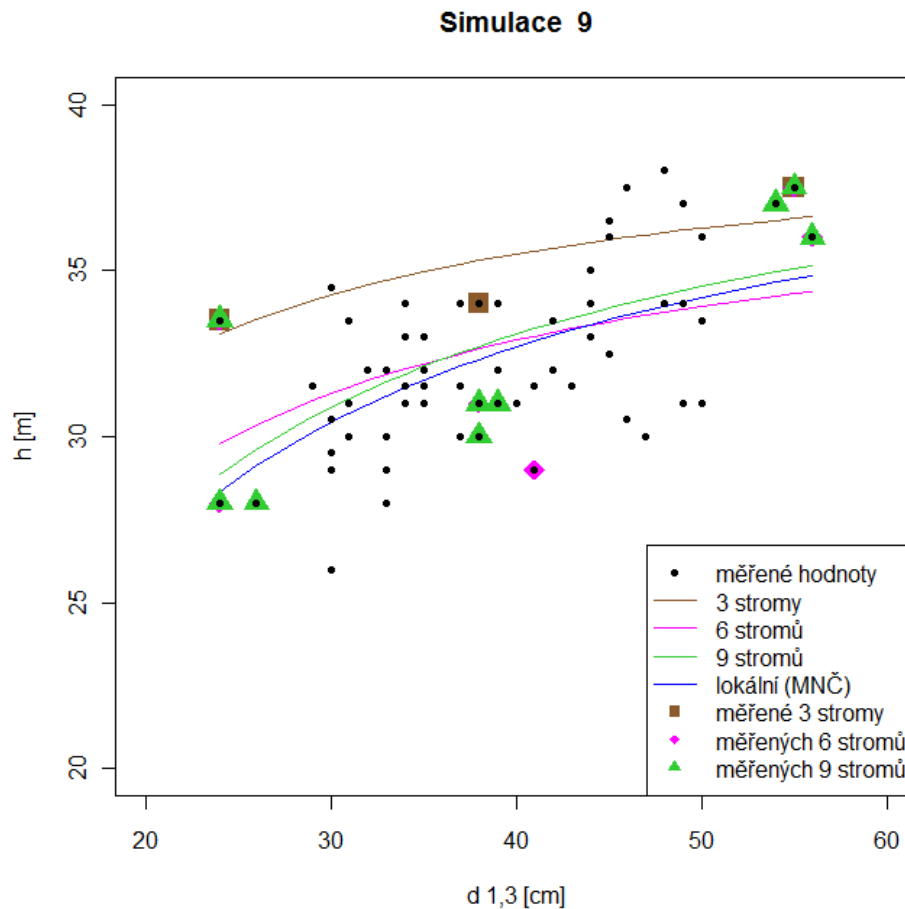


Obr. č. 25: Grafický výsledek vyhovující kalibrace při výběru 3-9 stromů ve třech tloušťkových intervalech v porostní skupině č. 3

Lze ale říci, že s rostoucí velikostí výběru se stává model méně citlivý na výběr stromu v rámci tloušťkového intervalu. Pokud je totiž u modelu se třemi stromy vybrán strom, který je silně nadúrovňový nebo podúrovňový, tak je celý model vychýlen od lokálního právě směrem k tomuto stromu. Při výběru šesti stromů se tento vliv již zmenšuje a při devíti stromech je tento vliv ještě výrazněji potlačen – viz obrázek č. 26. Na tomto obrázku je vidět, že při kalibraci s výběrem tří stromů (z nichž nejtenčí a nejtlustší lze označit za nadúrovňové) dochází k vychýlení a tedy i kalibrovaný model bude poskytovat nadhodnocené výšky. Naopak při výběru šesti stromů jsou vybrány dva stromy podúrovňové, ale výsledný model je výrazně lepší než při použití tří stromů. Nejlépe v tomto případě vychází model vytvořený kalibrací s devíti změřenými stromy, přestože čtyři z nich jsou výrazně nadúrovňové a dva podúrovňové. Můžeme tedy konstatovat, že větší velikost výběru využitá pro kalibraci (za určitých podmínek rozvržení těchto jedinců po tloušťkovém rozpětí) je schopna eliminovat vliv stromů, které se nachází dál od lokální výškové křivky a mohou při malých výběrech způsobit výrazné vychýlení modelu (výrazně nadúrovňové nebo podúrovňové stromy). Pokud budeme hledat nejlepší kalibraci v rámci jednotlivých porostních skupin, tak zjistíme, že ve všech vychází nejlépe podle šesti hodnotících kritérií kalibrace, která používala údaje z devíti změřených stromů. Výjimku lze nalézt pouze u směrodatné odchylky reziduí, která ve třech porostních skupinách vyšla nižší u jiných variant počtu stromů. Souhrnně lze tedy tento způsob kalibrace označit za nejlepší.

Pokud se podíváme na hodnocení kalibrace co nejvíce z praktického hlediska, tak je nejdůležitějším ukazatelem kvality a použitelnosti modelu kritérium střední hodnota odchylek modelových hodnot získaných smíšeným modelem od modelových hodnot lokálního modelu. Čím menší tato odchylka bude, tím menších chyb se budeme poté dopouštět při následné predikci výšek. Při volbě vhodného způsobu kalibrace je nutné ale rozlišovat mezi dvěma praktickými rovinami, jestli chceme využít predikované výšky pro zjišťování zásob pomocí objemových tabulek nebo jestli chceme predikovat výškovou strukturu například pro generování celého porostu v růstovém simulátoru a využít ji následně jako počáteční stav pro predikci jeho budoucího vývoje či k jiným vědeckým účelům, kde je vyžadována vyšší přesnost. Pokud jde o zjišťování zásob, tak stačí modelovat výšky s přesností na 1 m, protože to je nejnižší rozlišovací detail, který objemové tabulky mají. Pokud se jedná o úroveň modelování výšek v růstovém simulátoru, tak je nutné se dostat na co nejpřesnější hodnoty, protože výška

stromu jako taková vstupuje do dalších částí simulátoru jako proměnná a může tedy ovlivnit další vývoj jednotlivého stromu nebo porostu.



*Obr. č. 26: Vliv velikosti výběru na kvalitu kalibrace při výběru nadúrovňových nebo podúrovňových stromů*

Pro všechny porostní skupiny byla vypočítaná průměrná hodnota středních hodnot odchylek modelových hodnot získaných smíšeným modelem od modelových hodnot lokálního modelu včetně 95 % intervalu spolehlivosti. Pro tyto hodnoty byly také vypočítány průměry a intervaly spolehlivosti ze všech porostních skupin. Výsledky jsou uvedeny v tabulce č. 14. Z výsledků vyplývá, že z pohledu praktické použitelnosti v obecném měřítku není vhodné vůbec využívat kalibraci s využitím změření tří stromů, protože v jednotlivých porostních skupinách se průměrná hodnota středních hodnot odchylek pohybovala i přes 1 m, což je velmi velké vychýlení. Toto tvrzení je podloženo i průměrnými hodnotami za všechny porostní skupiny. Sice je možné mezi simulacemi nalézt případ, kde tato kalibrace dobře funguje, ale to je způsobeno konkrétním výběrem stromů v dané porostní skupině. Kalibrace s využitím šesti změřených stromů je využitelná pro stanovení zásoby porostu pomocí objemových tabulek, protože její průměr



středních hodnot odchylek se pohybuje okolo 65 cm a ani jeho interval spolehlivosti nepřekračuje 1 m, což je právě nejmenší výškový krok v objemových tabulkách. Výjimkou je pouze porostní skupina č. 22. Ovšem střední hodnoty odchylek jsou stále příliš velké, aby byla tato kalibrace použitelná pro modelování výškové struktury porostu. Kalibrace, která využívá devíti změřených stromů, poskytuje jednoznačně nejlepší výsledky, protože průměrná hodnota středních hodnot odchylek dvou použitých modelů vychází méně než 50 cm a ani její interval spolehlivosti nepřesahuje 70 cm. Z 60 provedených simulací tohoto typu kalibrace je střední hodnota odchylky ve 43 případech menší než 50 cm a v 18 případech dokonce menší než 30 cm. Je tedy patrné, že tento typ kalibrace lze díky své malé střední hodnotě odchylek dvou porovnávaných modelů doporučit pro modelování výškové struktury porostu např. pro generování v růstovém simulátoru nebo i pro jiné vědecké účely. Z praktického hlediska lze tedy konstatovat, že je dostačujících 6 - 9 změřených stromů v závislosti na dovolené střední hodnotě odchylky smíšeného a lokálního modelu. Změřené stromy by měly být v intervalech  $\langle d_{\min}; d_{\min} + 4 \text{ cm} \rangle$ ,  $\langle \bar{d} - 2 \text{ cm}; \bar{d} + 2 \text{ cm} \rangle$  a  $\langle d_{\max} - 4 \text{ cm}; d_{\max} \rangle$ .

*Tab. č. 12: Výsledné průměrné hodnoty i s jejich intervaly spolehlivosti středních hodnot odchylek modelových hodnot získaných smíšeným modelem od modelových hodnot lokálního modelu pro jednotlivé porostní skupiny*

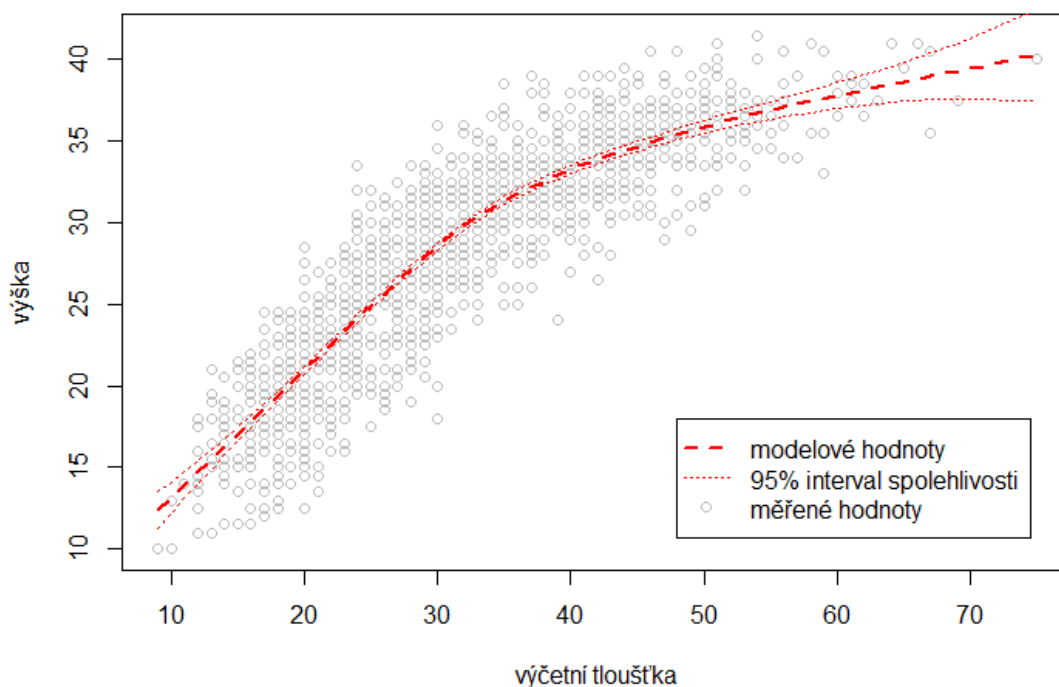
Počet změřených stromů	$\Delta_i$	Porostní skupina						
		3	5	11	14	19	22	průměr
3	průměr	0,5493	0,7515	1,1447	1,0316	0,6936	1,3151	0,9143
	DH IS	0,3427	0,4979	0,6939	0,4793	0,3803	0,8079	0,6044
	HH IS	0,7560	1,0052	1,5954	1,5839	1,0067	1,8223	1,2242
6	průměr	0,5418	0,3393	0,6648	0,7645	0,5895	1,1148	0,6691
	DH IS	0,4127	0,1488	0,4550	0,5696	0,4366	0,7252	0,3958
	HH IS	0,6708	0,5298	0,8746	0,9594	0,7425	1,5045	0,9424
9	průměr	0,4004	0,2392	0,6319	0,6420	0,2987	0,4934	0,4509
	DH IS	0,3500	0,1463	0,4854	0,4022	0,2057	0,2247	0,2743
	HH IS	0,4508	0,3321	0,7783	0,8817	0,3916	0,7621	0,6275

Legenda: DH IS – dolní hranice 95% intervalu spolehlivosti průměru, HH IS – horní hranice 95% intervalu spolehlivosti průměru,  $\Delta_i$  – střední hodnota odchylek modelových hodnot získaných smíšeným modelem od modelových hodnot lokálního modelu.

## 5.9 Zobecněný aditivní model

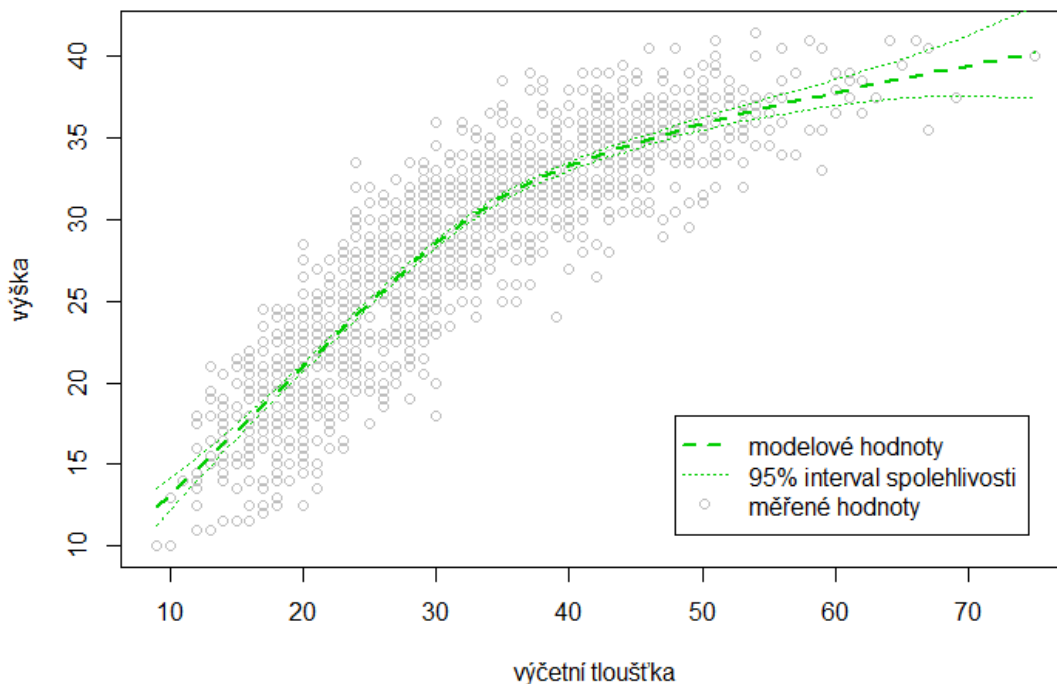
### 5.9.1 Globální modely

Nejkvalitnější modely (z pohledu vyhlazení), které využívaly splinové funkce, byly vybrány pomocí krosvalidace. Výsledný model s klasickým splinem měl hodnotu odhadu absolutního členu 28,3877 ( $p$ -hodnota  $< 0,001$ ), aditivní účinek výčetní tloušťky vyšel také statisticky významný ( $p$ -hodnota  $< 0,001$ ) a to při míře vyhlazení (udanou odhadnutými stupni volnosti) 4,578, směrodatná odchylka reziduí nabyla hodnoty 2,7657. Tento model vysvětlil 81,6 % deviance. Výsledný model je zobrazen na obrázku č. 27.



Obr. č. 27: GAM model s využitím spline funkce

Model s kubickým splinem se ve výsledcích velmi podobal předchozímu modelu. Odhad absolutního členu byl 28,3877 ( $p$ -hodnota  $< 0,001$ ), statisticky významný aditivní účinek výčetní tloušťky ( $p$ -hodnota  $< 0,001$ ) byl dosažen při míře vyhlazení 4,655 a směrodatná odchylka reziduí vyšla 2,7655. Vysvětlená míra deviance byla také 81,6 %. Tento model je zobrazen na obrázku č. 28.

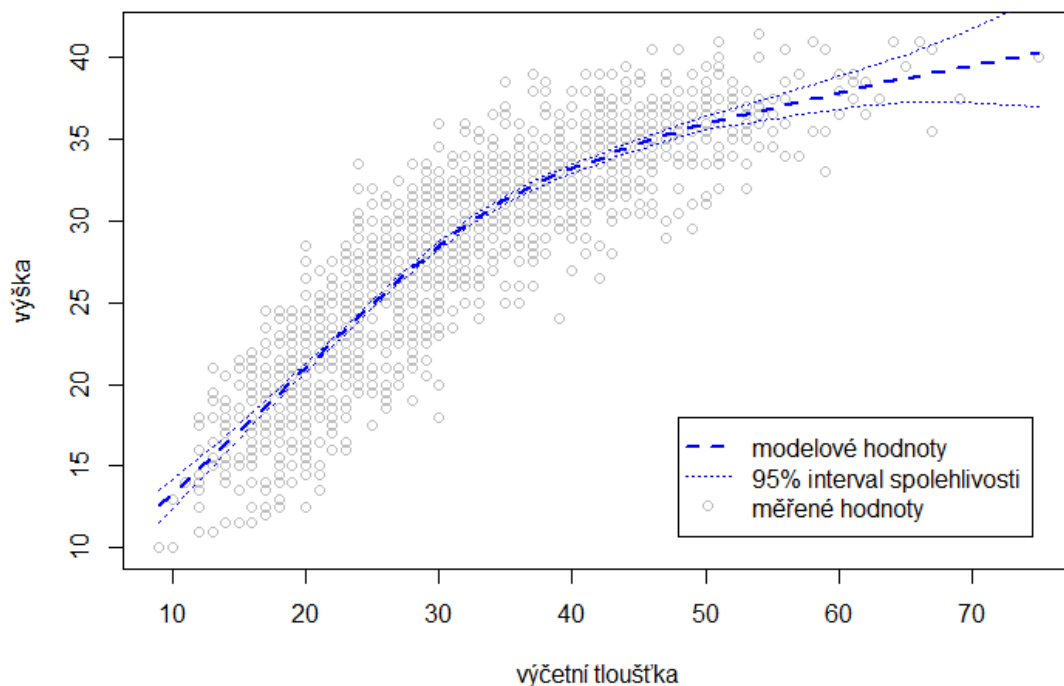


Obr. č. 28: GAM model s využitím kubické spline funkce

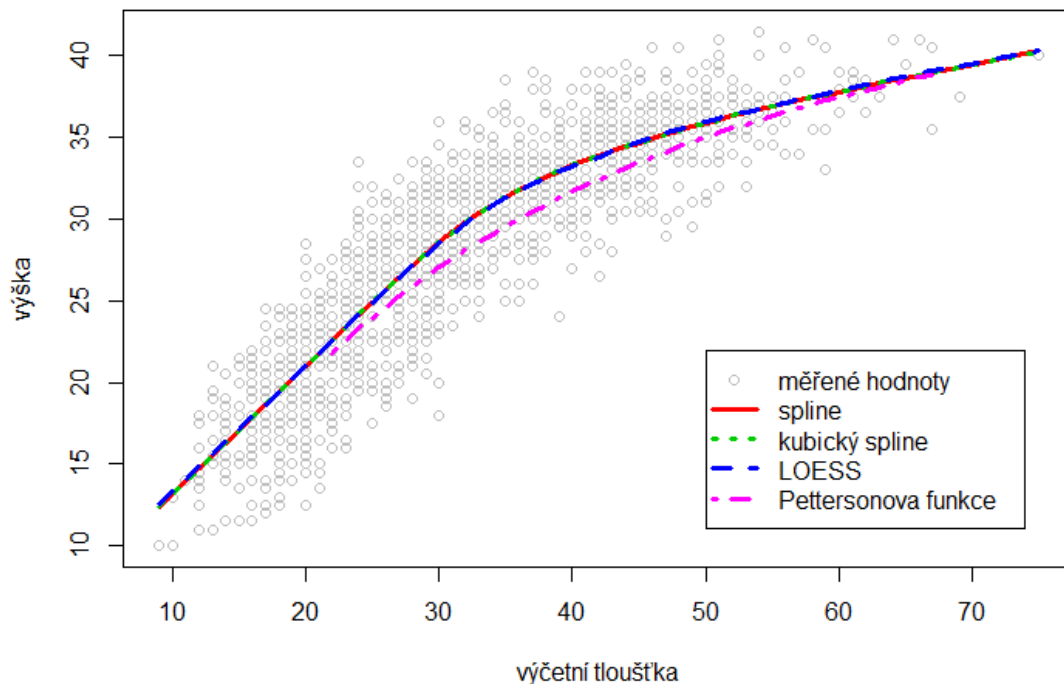
U modelu, kde byla vybrána vyhlazovací funkce LOESS, byla vybrána ta varianta, která měla šířku jádra 0,35, čemuž odpovídala hodnota AIC 7757,59, která byla ze všech testovaných modelů s LOESS funkcí nejnižší. Zároveň grafické znázornění poskytlo adekvátně vhodný výsledek, aniž by docházelo k přílišnému přeurčení modelu. Odhad absolutního členu byl 11,8306 (p-hodnota < 0,001), aditivní účinek výčetní tloušťky vyšel také statisticky významný (p-hodnota < 0,001) a byl dosažen při míře vyhlazení 4,3 a směrodatná odchylka reziduí vyšla 2,7628. Vysvětlená míra deviance byla 81,6 %. Finální model je na obrázku č. 29.

Všechny tři vytvořené modely, které využívaly různé typy vyhlazovacích funkcí, jsou společně zobrazeny na obrázku č. 30. Součástí obrázku je pro porovnání i model výškové křivky dle Pettersona vypočítaný pomocí nelineární regrese. Z obrázku je patrné, že všechny tři uvedené modely by se daly použít jako alternativa pro výškovou křivku, protože mají reálný průběh odpovídající klasické výškové křivce. Nižší poloha Pettersonovy výškové funkce oproti GAM modelům ve středové části je dána tím, že průběh funkce je řízen jejími parametry a tedy i jejich prvotním odhadem. Na rozdíl od toho GAM modely jsou řízeny pouze měřenými daty a více kopírují jejich vlastní průběh. Tyto modely lze považovat za lepší, protože mají výrazně nižší hodnoty AIC

než Pettersonova funkce, která měla AIC 7794,02. O kvalitě GAM modelů vypovídají i hodnoty středních hodnot a směrodatných odchylek reziduí, které vycházejí nižší než v případě Pettersonovy funkce. Tato funkce má střední hodnotu reziduí -0,0006 a směrodatnou odchylku reziduí 2,8023. Střední hodnoty odchylek jednotlivých GAM modelů od nelineárního modelu vykazují nízké hodnoty. Znamená to tedy, že Pettersonova funkce podhodnocuje predikované výšky proti výškám získaným GAM modely. Tyto hodnoty odchylek jsou ale akceptovatelné, a to zvláště v případě, že bude predikce výšek prováděna například s přesností na 1 m. Pokud by byl ovšem požadavek na kvalitu predikce vyšší, tak by s ohledem na výše uvedená kvalitativní kritéria bylo vhodnější používat pro přesnější predikci výšek GAM modely.



Obr. č. 29: GAM model s využitím funkce LOESS



Obr. č. 30: Srovnání různých GAM modelů a modelu Pettersonovy výškové křivky vytvořeného nelineární regresí

V tabulce č. 15 jsou uvedeny výsledky porovnávacích kritérií pro jednotlivé GAM modely. Z hodnot v tabulce je patrné, že všechny tři sestavené modely vycházejí ve všech kritériích téměř stejně. Také bylo vytvořeno několik grafů pro kontrolu všech tří modelů, a to konkrétně graf reziduí proti výčetní tloušťce, histogram a kvantil-kvantilový graf reziduí a graf měřených výšek proti modelovým výškám. Všechny tyto grafy jsou uvedeny v příloze č. 32 a poskytují velmi podobné výsledky. Všechny grafy jasně ukazují, že použití GAM modelů bylo vhodnou metodou, protože rezidua vykazují normální rozdělení, vykazují homoskedasticitu a modelové odezvy velmi dobře odpovídají hodnotám naměřeným. Jak již bylo uvedeno výše, tak střední hodnoty reziduí GAM modelů mají sice bodové odhady nižší než střední hodnota modelu vypočteného nelineární regresí, ale hypotéza o shodě středních hodnot reziduí nebyla zamítnuta ( $p$ -hodnota  $> 0,05$ ). Byl také proveden  $\chi^2$  test na shodu všech tří GAM modelů a byla potvrzena nulová hypotéza, že všechny modely lze považovat za shodné (ve všech vzájemných porovnáních byly hodnoty  $p > 0,05$ ). Souhrnně lze tedy říci, že všechny tři použité typy modelu jsou vhodné pro modelování výškové křivky globálního modelu a typ vyhlazovací funkce nemá vliv na výslednou kvalitu modelu. Všechny tři typy

vyhlazovacích funkcí lze tedy doporučit jako alternativní řešení ke klasickým výškovým křivkám, ale s upozorněním, že musí být vždy kontrolováno, aby sestavený aditivní model měl reálný průběh funkce při grafickém zobrazení a nedocházelo k přeurenění modelu, což se dá specifikovat tak, že je dovoleno, aby výsledná funkce měla pouze 1 inflexní bod a přitom byla stále rostoucí.

*Tab. č. 13: Kritéria pro hodnocení různých GAM modelů*

<b>Použitá vyhlazovací funkce</b>	<b>Pseudo <math>R^2</math> (%)</b>	<b>AIC</b>	<b>Střední hodnota reziduí</b>	<b>Směrodatná odchylka reziduí</b>	<b><math>\Delta_{ig}</math> (m)</b>
spline	81,6	7754,74	0,00000	2,7657	0,3453
kubický spline	81,6	7754,73	0,00000	2,7655	0,3459
LOESS	81,6	7757,59	0,00000	2,7628	0,3167

Legenda: Pseudo  $R^2$  (%) – podíl vysvětlené deviance, AIC – Akaikeho informační kritérium,  $\Delta_{ig}$  – střední hodnota odchylek modelových hodnot získaných aditivním modelem od modelových hodnot nelineárního modelu.

### 5.9.2 Lokální modely

V tabulce č. 16 jsou uvedeny výsledné hodnoty srovnávacích kritérií pro jednotlivé porostní skupiny. Výsledné grafické znázornění všech lokálních GAM modelů a lokálního modelu Petersonovy výškové funkce vypočítaného pomocí nelineární regrese jsou uvedeny v příloze č. 33. Z dosažených výsledků je patrné, že splinové funkce mají v některých případech problém vytvořit model, který by splňoval podmínky na kvalitní výškovou funkci a to proto, že dochází k přeurenění modelu. Toto je ale patrné pouze v grafickém znázornění modelu, protože podle vypočtených kritérií toto nelze poznat. Naopak je možné takovýto model hodnotit jako velmi kvalitní podle jeho výsledných hodnot srovnávacích kritérií a zároveň i lepší než GAM model s použitou LOESS funkcí. Dobrým příkladem přeurenění jsou GAM modely se splinovými funkcemi v porostní skupině č. 13, kde s rostoucí výčetní tloušťkou výška klesá nebo v porostní skupině č. 23, kde GAM modely oscilují kolem nelineárního modelu – viz obrázek č. 31. Z celkového počtu 23 porostních skupin byl GAM model s klasickým splinem nepoužitelný v pěti případech, GAM model s kubickým splinem dokonce v osmi případech. Lze to přisuzovat principu metody výběru nejlepšího modelu pomocí krosvalidace, protože tato metoda skutečně hledá model, který nejlépe prokládá datové pole bez ohledu na jeho praktický smysl. Naopak při použití GAM modelu s funkcí LOESS vykazovala hodnotící kritéria často mírně horší výsledky než splinové funkce, ale ve všech porostních skupinách se tento model svým chováním velmi přiblížil klasické

výškové funkci – viz obrázek č. 31 a příloha č. 33. Je to dáno tím, že výsledný model je vybrán kompromisně na základě hodnoty AIC a vizuální kontroly grafického průběhu modelu. U této funkce bylo možné vždy vybrat model, který bude prakticky dobře napodobovat průběh výškové funkce na úkor nižších hodnot ostatních hodnotících kritérií, ale tyto rozdíly nebyly nikdy natolik odlišné, že by model byl označen za celkově nepoužitelný. Protože všechny vytvořené GAM modely s LOESS funkcí použily šířku jádra větší než 0,60 (hodnota uvedená v závorce u LOESS funkce v tabulce č. 16), tak lze konstatovat, že pro použití GAM modelu s LOESS funkcí jako modelu výškové křivky na lokální úrovni není vhodné používat nižší hodnoty šířky jádra, protože v takovýchto případech dochází k přeurčenosti vytvořeného modelu.

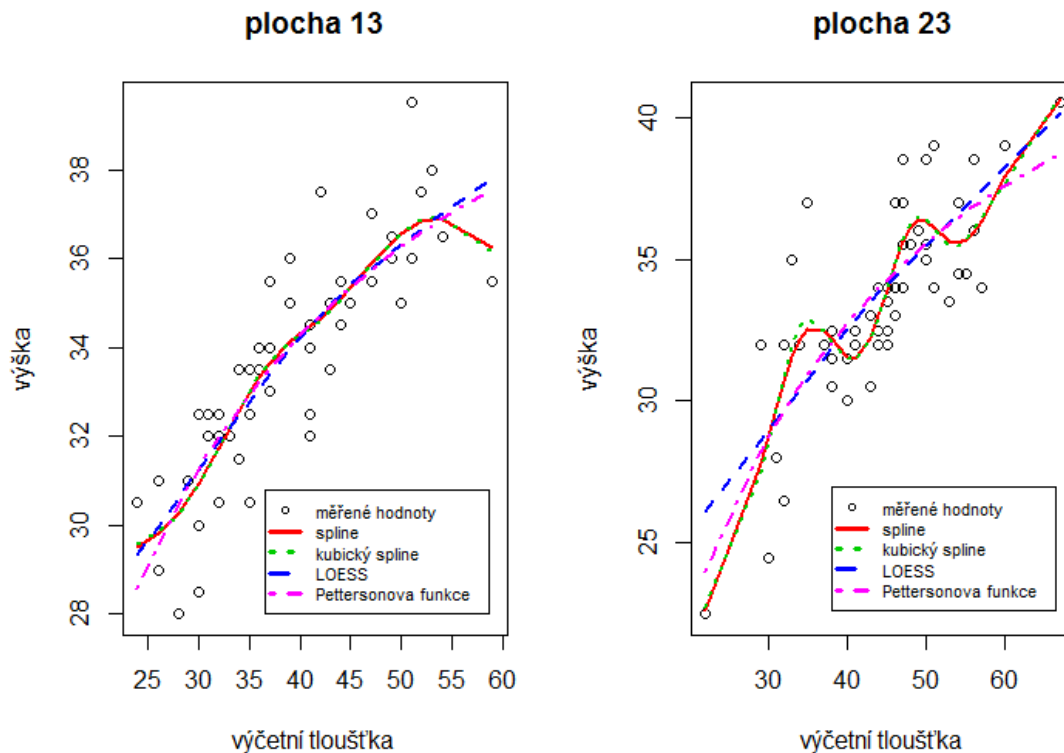
*Tab. č. 14: Kritéria pro srovnání lokálních GAM modelů*

Porostní skupina	Použitá vyhlazovací funkce	Pseudo $R^2$ (%)	AIC	Střední hodnota reziduí	Směrodatná odchylka reziduí	$\Delta_{ig}$ (m)
1	spline	68,6	194,97	0,0000	1,0887	0,0068
	kubický spline	68,6	194,95	0,0000	1,0885	0,0068
	LOESS (0,70)	67,9	195,97	0,0000	1,1001	0,0063
2	spline	42,6	137,34	0,0000	1,2155	0,0019
	kubický spline	42,6	137,34	0,0000	1,2155	0,0019
	LOESS (0,75)	42,8	139,68	0,0000	1,2140	0,0011
3	spline	85,3	150,48	0,0000	0,9454	0,0017
	kubický spline	85,3	150,51	0,0000	0,9460	0,0017
	LOESS (0,65)	84,9	151,55	0,0000	0,9582	0,0023
4	spline	74,8	160,57	0,0000	0,9906	0,0044
	kubický spline	81,3	155,68	0,0000	0,8542	0,0121
	LOESS (0,75)	74,3	160,77	0,0000	1,0010	0,0040
5	spline	86,3	231,72	0,0000	0,9176	0,0063
	kubický spline	86,3	231,74	0,0000	0,9180	0,0063
	LOESS (0,60)	85,9	232,61	0,0000	0,9319	0,0050
6	spline	76,2	262,44	0,0000	1,3463	0,0151
	kubický spline	76,2	262,44	0,0000	1,3463	0,0151
	LOESS (0,85)	76,4	263,63	0,0000	1,3409	0,0122
7	spline	74,7	368,21	0,0000	1,2990	0,0156
	kubický spline	74,6	368,47	0,0000	1,3012	0,0148
	LOESS (0,80)	70,0	377,74	0,0000	1,4145	0,0054
8	spline	75,4	260,88	0,0000	1,3662	0,0057
	kubický spline	75,4	260,89	0,0000	1,3666	0,0057
	LOESS (0,95)	74,7	261,98	0,0000	1,3893	0,0075
9	spline	62,5	213,43	0,0000	1,4805	0,0049
	kubický spline	62,5	213,43	0,0000	1,4804	0,0049
	LOESS (0,85)	62,5	213,48	0,0000	1,4791	0,0041

10	spline	74,6	363,12	0,0000	1,8864	0,0224
	kubický spline	74,5	363,28	0,0000	1,8903	0,0218
	LOESS (0,85)	72,7	366,51	0,0000	1,9564	0,0124
11	spline	88,5	269,38	0,0000	1,2589	0,0189
	kubický spline	88,6	269,14	0,0000	1,2549	0,0194
	LOESS (0,80)	86,7	275,69	0,0000	1,3545	0,0059
12	spline	71,2	269,25	0,0000	1,4976	0,0085
	kubický spline	71,2	269,16	0,0000	1,4959	0,0087
	LOESS (0,80)	69,2	271,23	0,0000	1,5483	0,0046
13	spline	76,0	182,60	0,0000	1,2100	0,0062
	kubický spline	76,3	182,35	0,0000	1,2029	0,0069
	LOESS (0,85)	72,7	184,29	0,0000	1,2912	0,0034
14	spline	41,5	288,23	0,0000	1,9991	0,0109
	kubický spline	41,5	288,23	0,0000	1,9991	0,0109
	LOESS (0,95)	42,1	288,64	0,0000	1,9891	0,0150
15	spline	82,1	298,50	0,0000	1,5088	0,0076
	kubický spline	82,2	298,40	0,0000	1,5074	0,0078
	LOESS (0,75)	81,7	299,40	0,0000	1,5289	0,0056
16	spline	60,1	299,66	0,0000	2,4324	0,0116
	kubický spline	60,0	299,70	0,0000	2,4348	0,0114
	LOESS (0,80)	58,8	300,29	0,0000	2,4693	0,0091
17	spline	68,5	273,68	0,0000	1,4473	0,2159
	kubický spline	69,3	271,65	0,0000	1,4292	0,2187
	LOESS (0,95)	57,3	282,67	0,0000	1,6851	0,0089
18	spline	58,3	241,33	0,0000	1,8662	0,0033
	kubický spline	58,4	241,27	0,0000	1,8633	0,0037
	LOESS (0,95)	56,5	242,39	0,0000	1,9053	0,0045
19	spline	64,5	339,13	0,0000	1,5617	0,0063
	kubický spline	68,3	338,16	0,0000	1,4746	0,0150
	LOESS (0,95)	63,9	339,61	0,0000	1,5743	0,0081
20	spline	53,6	380,35	0,0000	2,0619	0,0054
	kubický spline	59,0	379,67	0,0000	1,9385	0,0225
	LOESS (0,95)	53,0	380,50	0,0000	2,0741	0,0040
21	spline	58,1	305,99	0,0000	1,6750	0,0103
	kubický spline	58,1	305,96	0,0000	1,6749	0,0103
	LOESS (0,85)	57,6	306,00	0,0000	1,6852	0,0099
22	spline	57,8	272,45	0,0000	2,6356	0,0136
	kubický spline	57,8	272,45	0,0000	2,6356	0,0136
	LOESS (0,95)	58,4	272,79	0,0000	2,6159	0,0187
23	spline	72,0	236,27	0,0000	1,8403	0,0283
	kubický spline	73,0	234,86	0,0000	1,8089	0,0299
	LOESS (1,00)	58,5	246,38	0,0000	2,2415	0,0082

Legenda: Pseudo  $R^2$  (%) – podíl vysvětlené deviance, AIC – Akaikeho informační kritérium,  $A_{ig}$  – střední hodnota odchylek modelových hodnot získaných aditivním modelem od modelových hodnot nelineárního modelu.





Obr. č. 31: GAM modely a nelineární model v porostní skupině 13 a 23

### 5.9.3 Souhrnné zhodnocení GAM modelů

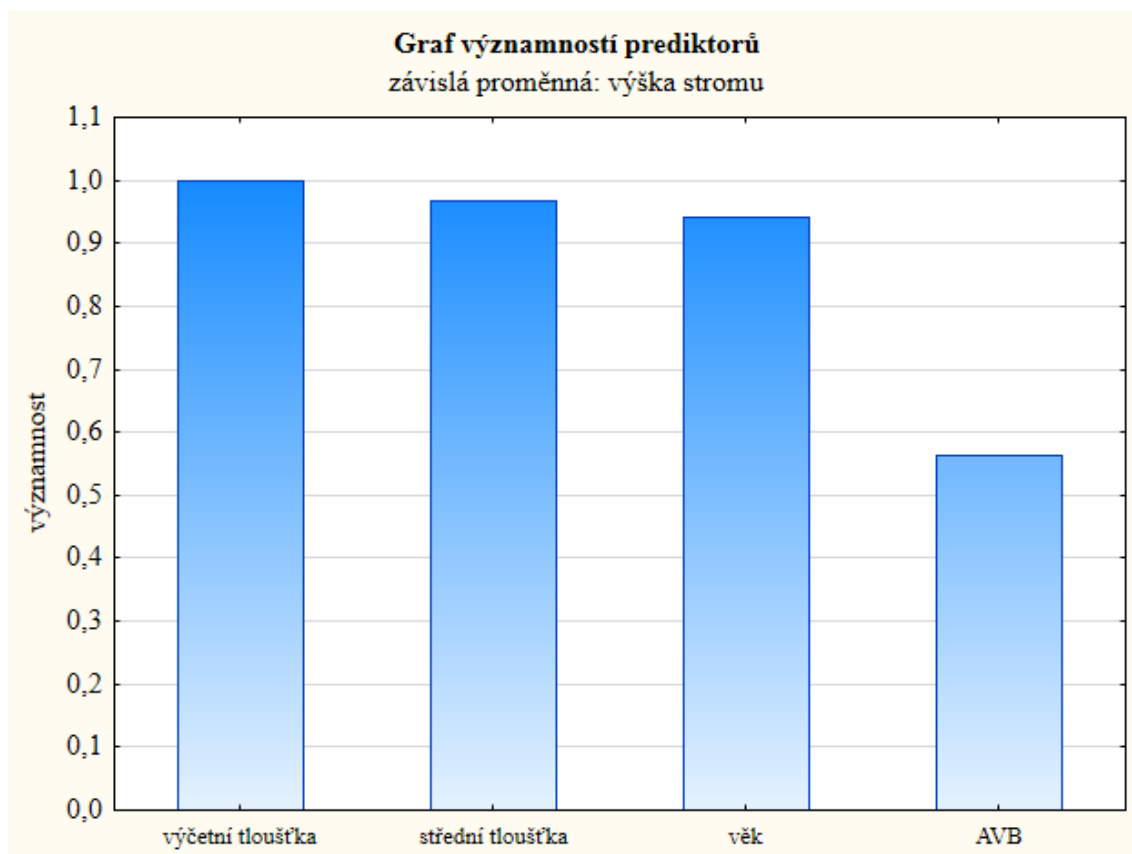
Zhodnocení použitelnosti GAM modelů jako alternativy k modelování výškové křivky lze shrnout do několika bodů:

- při použití GAM modelů na datový materiál většího rozsahu (nejčastěji globální model) lze konstatovat, že nejsou statisticky i reálně prokazatelné rozdíly ve výsledcích modelů, které používají různé vyhlazovací funkce (spline, kubický spline, LOESS), a to z důvodu, že dostatečná velikost datového souboru je schopna eliminovat přeúčnění modelu,
- při použití GAM modelů na datové soubory o menším rozsahu (na úrovni lokálního modelu) lze doporučit jako vhodný GAM model ten, který využívá vyhlazovací funkci LOESS a to při volbě větší šířky jádra, aby nedocházelo k přeúčnění modelu,
- výsledný model nehodnotit pouze podle jeho numerických výsledků, ale vždy kontrolovat především grafické zobrazení modelu s ohledem na jeho reálný průběh a finální model vybrat jako kompromis mezi hladkostí průběhu modelu a jeho výslednými kvalitativními kritérii,

- dosažený bias (vychýlení) GAM modelů (vyjádřené střední hodnotou odchylek modelových hodnot získaných aditivním modelem od modelových hodnot nelineárního modelu) je dostačující pro predikci výšek s přesností 1 m a lze tedy GAM modely doporučit jako nástroj pro velmi přesnou predikci.

## 5.10 Klasifikační a regresní stromy

Výsledný regresní strom vytváří homogenní skupiny stromů se stejnou výškou, a to v závislosti na výčetní tloušťce stromu, střední tloušťce dřeviny, věku porostní skupiny a absolutní výškové bonitě dřeviny. Významnost jednotlivých proměnných je znázorněna na obrázku č. 32. Je z něj patrné, že největší přínos má pro regresní strom výčetní tloušťka, pak střední tloušťka, věk a nakonec bonita dřeviny. Po prvotním sestavení stromu a jeho prořezání pomocí krosvalidace má výsledný strom 19 uzlů a 20 finálních homogenních skupin. Výsledné hodnoty kvalitativních kritérií jsou uvedeny v tabulce č. 17. Z výsledků je patrné, že kvalita sestaveného modelu pomocí regresního stromu je vysoká, protože jak na trénovacích, tak i na validačních datech byla vysvětlena velká část variability výšek stromů a všechny vypočítané statistické charakteristiky reziduí jsou na velmi nízkých hodnotách. I když hodnoty získané z validačních dat nejsou tak vysoké ( $R^2$ ) a nízké (charakteristiky reziduí) jako z trénovacích, tak i takové výsledky je možné hodnotit jako velmi dobré. Z kvantil-kvantilových grafů obou skupin reziduí je dobře vidět, že rezidua mají normální rozdělení. Výsledné grafy jsou uvedeny na obrázku č. 33. Z grafů reziduí obou skupin proti predikovaným hodnotám je dobře vidět, že rezidua se pohybují kolem nuly a model tedy není vychýlený. Také můžeme konstatovat, že v reziduích není žádný viditelný konkrétní tvar, který by mohl znamenat např. problém s heteroskedasticitou. Výsledné grafy jsou uvedeny na obrázku č. 34. Z grafů porovnávajících měřené a predikované hodnoty výšek z obou podskupin dat je vidět, že modelové hodnoty velmi dobře odpovídají měřeným hodnotám. Výsledné grafy jsou znázorněny na obrázku č. 35.

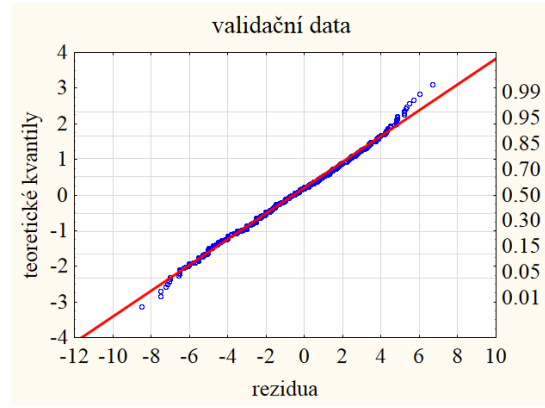
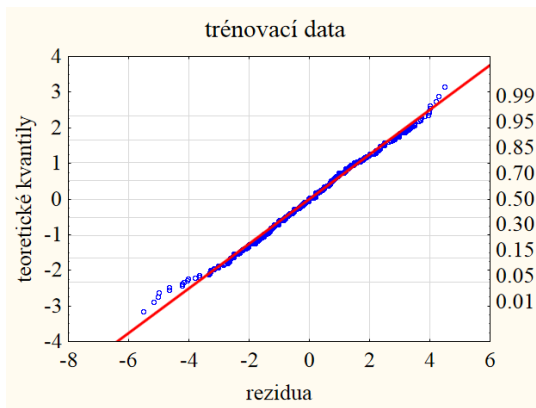


Obr. č. 32: Významnost jednotlivých prediktorů pro regresní strom

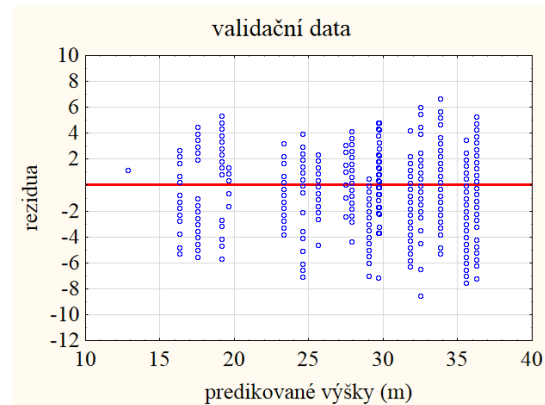
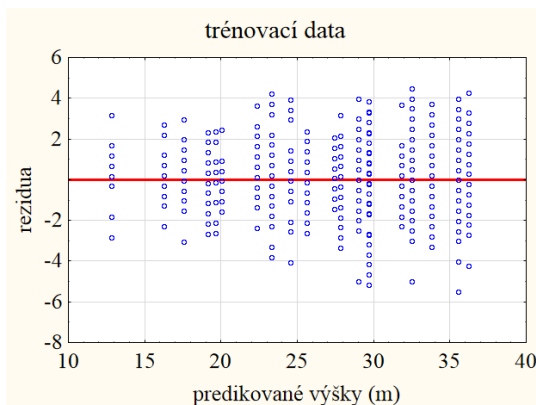
Legenda: AVB – absolutní výškový bonita.

Tab. č. 15: Výsledné hodnoty kvalitativních kritérií regresního stromu

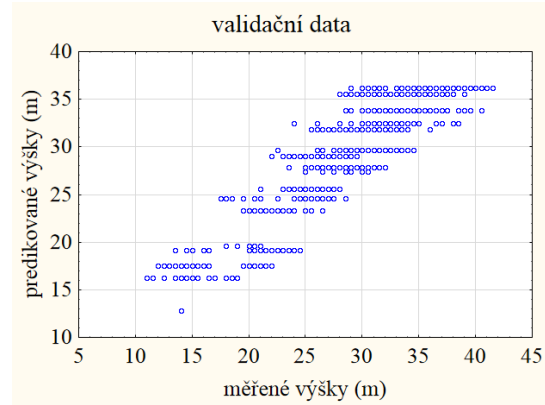
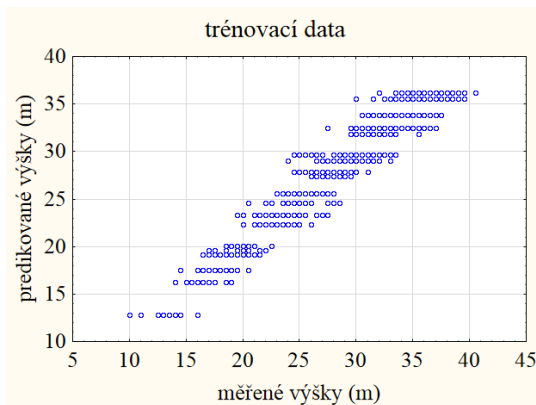
Hodnotící kritérium	Trénovací data	Validační data
koeficient determinace	0,9400	0,8040
střední hodnota reziduí	0,0000	-0,5848
střední chyba reziduí	0,0542	0,1016
směrodatná odchylka reziduí	1,5901	2,7459



Obr. č. 33: Kvantil-kvantilový graf reziduí trénovacích a validačních dat



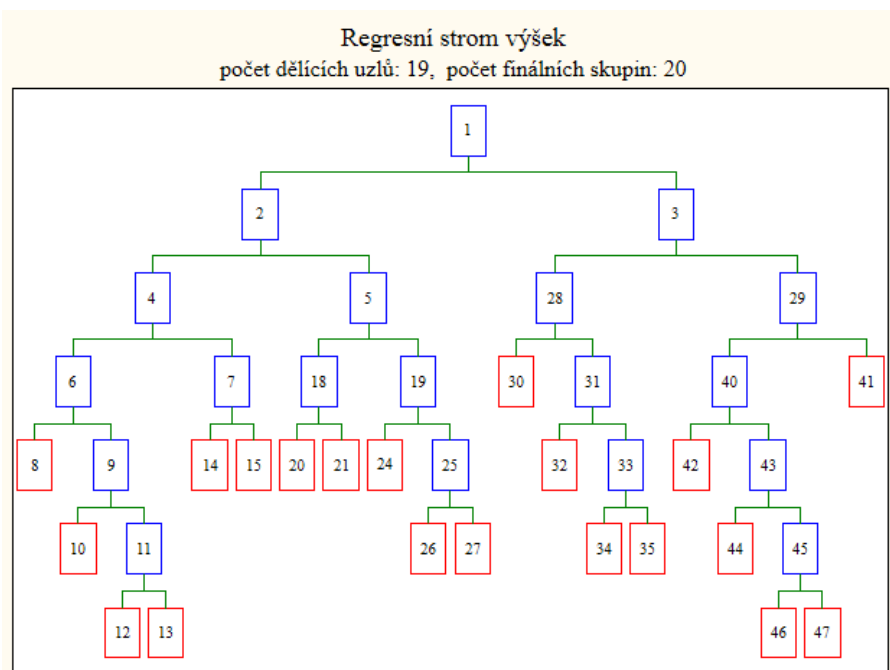
Obr. č. 34: Grafy reziduí proti predikovaným hodnotám trénovacích a validačních dat



Obr. č. 35: Grafy predikovaných hodnot proti měřeným hodnotám trénovacích a validačních dat

Celý výsledný regresní strom je uveden na obrázku č. 36. Strom je uveden pouze schematicky s čísly uzlů a finálních skupin. Hodnoty netvoří přesně jdoucí pořadí, protože některé uzly byly odstraněny procesem krosvalidace. Vlastní hodnoty počtu měření, středních hodnot výšek, rozptylů výšek, druhu dělicího kritéria a jeho přesné hodnoty jsou uvedeny v tabulce č. 18. Hodnoty dělicích kritérií jsou vždy nastaveny tak, aby v levé větvi byly jedinci s hodnotou menší nebo rovnou tomuto kritérii a v pravé větvi

jedinci s hodnotou větší než toto kritérium. Z výsledků je patrné, že s rostoucím věkem a výčetní tloušťkou roste i predikovaná střední výška, takže i pomocí této metody je možné modelovat reálný vztah výšky stromu s jeho tloušťkou v různých vývojových fázích. Ve dvou uzlech je konstatováno, že stromy, které pocházejí z porostů s vyšší střední tloušťkou, mají nižší výšky, což sice odporuje obecnému pravidlu, že s rostoucí střední tloušťkou rostou i výšky stromů, ale po identifikaci porostů, ve kterých se tyto stromy nacházely, bylo zjištěno, že porosty s vyšší střední tloušťkou mají nižší zakmenění, takže mezi stromy v těchto porostech nejsou tak silné konkurenční vztahy. Můžeme z toho tedy vyvozovat, že stromy koncentrují svůj růst více do tloušťkového přírůstu a proto mají nižší výšky než stromy v porostech s vyšším zakmeněním, kde je maximum růstu směřováno do výškového přírůstu, aby stromy měly dostatek prostoru pro zachycení slunečního záření nutného pro tvorbu asimilátu při fotosyntéze. Co se týče bonity, tak ta byla využita pouze v jednom uzlu a výsledky ukazují, že stromy s nižší AVB mají vyšší výšky než stromy s vyšší AVB. Toto by se mohlo zdát jako chyba, ale opět po identifikaci těchto 46 stromů bylo zjištěno, že se jedná sice o stromy s porostů s nižší AVB, ale konkrétní stromy byly silně nadúrovňové a výrazně přesahovaly svojí tloušťkou střední tloušťku porostu, takže je regresní strom správně zařadil do jedné společné skupiny díky využití více prediktorů než pouze AVB. Metodu regresních stromů lze tedy označit za použitelnou alternativu pro modelování výšek stromů.



Obr. č. 36: Regresní strom výšek stromů

Legenda: modré obdélníky - dělicí uzly, červené obdélníky – finální skupiny, zelené čáry – větve dělicích kritérií.

Tab. č. 16: Hodnoty počtu měření, středních hodnot a rozptylů výšek a dělicích kritérií regresního stromu

Číslo uzlu	Levá větev k uzlu	Pravá větev k uzlu	Počet hodnot v uzlu	$\mu_h$ v uzlu (m)	$\sigma_h^2$ v uzlu (m <sup>2</sup> )	Dělicí kritérium	Hodnota dělicího kritéria
1	2	3	860	27,97	42,10	střední tloušťka	30,1
2	4	5	380	22,09	19,14	věk	48
4	6	7	197	19,09	8,40	výčetní tloušťka	20,5
6	8	9	93	17,04	5,37	výčetní tloušťka	12,5
8			11	12,82	3,24		
9	10	11	82	17,60	2,94	výčetní tloušťka	16,5
10			26	16,29	1,39		
11	12	13	56	18,21	2,49	věk	39
12			41	17,54	1,40		
13			15	20,07	0,80		
7	14	15	104	20,92	4,00	věk	39
14			47	19,16	1,35		
15			57	22,37	1,52		
5	18	19	183	25,33	10,51	výčetní tloušťka	23,5
18	20	21	78	22,54	4,89	výčetní tloušťka	16,5
20			16	19,63	2,52		
21			62	23,29	2,75		
19	24	25	105	27,40	4,61	AVB	32
24			46	29,01	2,97		
25	26	27	59	26,14	2,29	výčetní tloušťka	30,5
26			42	25,62	1,82		
27			17	27,44	1,08		
3	28	29	480	32,62	11,32	výčetní tloušťka	32,5
28	30	31	126	28,99	8,26	výčetní tloušťka	23,5
30			16	24,56	4,90		
31	32	33	110	29,64	5,48	věk	75
32			23	27,85	2,75		
33	34	35	87	30,11	5,13	střední tloušťka	39,3
34			18	31,81	2,03		
35			69	29,67	5,00		
29	40	41	354	33,91	6,07	výčetní tloušťka	46,5
40	42	43	269	33,18	4,98	věk	75
42			25	29,70	2,22		
43	44	45	244	33,53	3,90	střední tloušťka	39,3
44			45	35,52	4,26		
45	46	47	199	33,08	2,72	výčetní tloušťka	40,5
46			111	32,51	2,60		
47			88	33,80	1,95		
41			85	36,22	2,46		

Legenda:  $\mu_h$  – střední hodnota výšek,  $\sigma_h^2$  – rozptyl výšek, AVB – absolutní výšková bonita.

## 6 Diskuze

### 6.1 Využití regresních metod

V práci jsou řešeny tři různé regresní metody, které v posledních letech získaly na oblibě, a jejich využití je ukázáno na příkladu výškové funkce smrku. Rozvoj regresních metod zažil ale v posledních 20 – 25 letech tak rychlý vývoj, že podchytit všechny rychle se rozvíjející metody je nad rámec této práce. Z dalších regresních metod, které by mohly být označeny jako perspektivní pro lesnictví, to jsou například:

- metoda geograficky vážené regrese - GWR, která se zabývá geostatistickým pojetím odhadu parametrů modelu (např. práce autorů Fotheringham et al. (1998, 2002), Zhang et al. (2004), Wang et al. (2005a), Zhang et al. (2008)),
- metoda umělých neuronových sítí - ANN, což je metoda založená na strojovém učení, jejíž hlavní výhodou je to, že nemusí znát přesný tvar nebo rovnici vztahu mezi proměnnými (např. práce autorů Moisen a Frescino (2002), Aertsen et al. (2010, 2011), Özçelik et al. (2013), Castaño-Santamaría et al. (2013)),
- zobecněné modely se smíšenými efekty GMEM, což jsou modely se smíšenými efekty, které nevyžadují u závisle proměnné normální rozdělení (např. práce autorů Eerikäinen et al. (2007), Groom et al. (2012)).

Pro využití v lesnický orientovaných modelech připadají do úvahy ještě další varianty regresních technik, ale metody použité v práci nebo výše uvedené v přehledu se považují za nejperspektivnější v oblasti lesnického modelování.

### 6.2 Model výškové křivky

Model výškové křivky je jedním z nejstarších a zároveň jedním z nejpoužívanějších lesnických modelů. Pro modelování byly voleny různé typy funkcí, od nejjednodušších lineárních vztahů (např. Zhang et al. 2004) a alometrických rovnic (např. Trincado et al. 2007), přes upravené růstové funkce (např. Zhang, 1997) až po speciálně pro tento vztah sestavené výškové funkce (např. Näslund, 1947). V současné době existují desítky funkcí, které lze využít k modelování výškové křivky. Žádnou z těchto funkcí ale nelze označit za nejlepší (Mehtätalo, 2004).

Také se dnes nepoužívají jen funkce, které obsahují pouze dvě proměnné (tedy výšku a tloušťku), ale využívají se i modely, které zahrnují další veličiny porostního charakteru (např. Temesgen a von Gadow (2004), Sharma a Parton (2007) a Vargas-Larreta et al. (2009)). Hlavními důvody pro zanesení porostní proměnné do modelu je to, že se snažíme

snížit podíl nevysvětlené variability a vysvětlit změny modelu v průběhu dynamického vývoje porostu (Özçelik et al. 2013). I když se tento typ modelu výškové křivky hojně využívá hlavně v posledních 20 letech, tak vůbec první model, který obsahoval porostní veličinu, sestavil již Wiedemann (1936). Jeho model obsahoval jako porostní veličinu střední tloušťku porostu, takže byl schopen sestavit soustavu výškových křivek v závislosti na střední tloušťce v podobě, která byla základem pro sestavení systému jednotných výškových křivek, jak jej známe dnes. Porostními veličinami, které nejčastěji vstupují do modelu, jsou věk porostu, např. Curtis (1967), horní porostní výška, např. Soares a Tomé (2002) a střední tloušťka, např. Schröder a Álvarez Gonzáles (2001). Pokud bychom chtěli ale sestavit model výškové křivky s využitím věku porostu jako porostní proměnné, tak musíme souhlasit s konstatováním Huanga a Tituse (1993), kteří upozornili na to, že takovýto model nelze uplatnit v nestejnověkém lese, protože věk jako porostní veličinu zde nemůžeme uvažovat. Jelikož se nestejnověké lesy svojí strukturou nejčastěji blíží lesům výběrným, tak je v nich vhodným modelem právě model stacionární výškové křivky, který se využívá i ve výběrném lese. Další odlišný přístup zvolili ve své práci Huang et al. (2000), kteří neobohatili model o žádnou další proměnnou, ale rozdělili území na menší oblasti, které měly stejné ekologické podmínky, a pro tyto oblasti sestavili modely. Dokázali tedy závislost parametrů modelu na ekologických podmínkách stanoviště. Aktuálním tématem je také provázanost odhadů parametrů v závislosti na prostorovém umístění zájmových lokalit. Tento problém je řešen různými geostatistickými metodami a ve své práci se s ním zabírali např. Zhang et al. (2004). Ti využili pro modelování závislosti odhadů parametrů výškové funkce na poloze stromů v porostu GWR metodu. Zjistili, že využití GWR vysvětluje velkou část prostorové variability v datech, což je hlavní výhoda této metody. Výsledné modely výškových křivek jsou kvalitnější než modely sestavené pomocí metody nejmenších čtverců. Také sestavili mapu odhadů parametrů modelů v závislosti na poloze jednotlivých stromů. Nevýhodou této metody je ale nutnost znát přesnou polohu jednotlivých stromů, což zvyšuje náročnost inventarizačních prací v terénu.

V této práci byla pro daná data vybrána jako nejvhodnější Petersonova funkce. Tato funkce není nijak komplikovaná z hlediska parametrů, ale splňuje všechny podmínky na výškovou funkci kladené (tyto jsou uvedeny v kapitole 3.2). Splňuje i požadavek Šmelka (2007), aby byla linearizovatelná. Tato funkce je velmi často používaná a pro svoje kvality ji doporučují např. autoři Kramer a Akça (1995), Van Laar a Akça (2007), Pretzsch (2009) nebo Drápela (2011). Je nutné ale podotknout,



že ani ostatní funkce se neprojeví jako výrazně nevyhovující (s výjimkou Prodanovy funkce na ploše 14 a 22), takže je možné konstatovat, že všechny použité funkce jsou vhodné pro modelování výškové křivky. Toto potvrzuje i konstatování Fanga a Baileyho (1998), kteří tvrdí, že pokud se použijí na stejná data funkce, které mají stejný počet parametrů a je použita stejná metoda odhadu parametrů (podle nich metoda nejmenších čtverců), tak by výsledné modely měly poskytovat velmi podobné výsledky, a to jak z pohledu predikce výšek, tak i kvalitativních kritérií modelu.

### 6.3 Smíšený model

Finální smíšený model obsahuje náhodnou část odhadu u absolutního i u regresního parametru a také prediktor druhé úrovně (věk na úrovni regresního parametru, respektive přirozený logaritmus věku na úrovni absolutního parametru. Tento model měl podstatně nižší směrodatnou odchylku reziduí než globální model a z obrázku č. 21 je také patrné, že tento model odstranil problém heteroskedasticity v datech. K totožným závěrům dospěl ve své práci i Drápela (2011), který vytvořil smíšený model pro borovici z dat Finské národní inventarizace lesů. Byl tedy splněn cíl pro úspěšné použití smíšeného modelu daný prací autorů Fanga a Baileyho (2001), kteří konstatovali, že cílem smíšeného modelu je objasnit část náhodné variability globálního modelu, odstranit část nekonstantního rozptylu a sestavit natolik flexibilní model využívající lokálních modelů, aby byl použitelný na daném území. Výslednou kvalitu modelu lze tedy shrnout tvrzením Zhaoa et al. (2013), že modely se smíšenými efekty jsou silným a efektivním nástrojem k analýze skupinovitě uspořádaných dat.

Jelikož byl jako prediktor druhé úrovně zvolen věk porostu, tak může být tento model použit pouze v lese věkových tříd a ne v lesích nestejnověkových. Existují i práce, které ve výsledném modelu obsahují i jiné porostní veličiny a to např. střední tloušťku (Schmidt et al. 2011), horní výšku a horní tloušťku (Crecente-Campo et al. (2010), Özçelik et al. (2013) a Castaño-Santamaría et al. (2013)) nebo kombinace více porostních veličin (např. Sharma a Parton (2007), Adame et al. (2008), Temesgen et al. (2014)). Je možné nalézt i práce přímo zaměřené na nestejnověké porosty, kde je jako porostní veličina využita střední nebo horní tloušťka a horní výška např. Vargas-Larreta et al. (2009) nebo Castaño-Santamaría et al. (2013). Důvodem, proč je téměř ve všech uvedených pracích uvažována jako porostní proměnná horní výška porostu může být to, že tuto veličinu lze využít pro vyjádření bonity stanoviště jako ukazatele produkčního potenciálu stanoviště (Eerikäinen, 2003). Lze tedy souhlasit s tvrzením, že zahrnutí

porostních veličin do modelu zlepšuje výslednou kvalitu modelu (Soares a Tomé 2002; Calama a Montero 2004).

### 6.3.1 Kalibrace smíšeného modelu

Kalibrace smíšeného modelu je jedním z hlavním důvodů, proč se model se smíšenými efekty využívá. V praxi se osvědčila nejvíce podmíněná kalibrace, která vyžaduje znalost několika hodnot závisle proměnné. I v této práci byla k predikci výšek využita podmíněná kalibrace. Byly vyzkoušeny dvě varianty výběru stromů sloužících ke kalibraci. Varianta 1 testovala výběr 1 – 5 stromů v intervalu  $\pm 2$  cm od střední tloušťky a varianta 2 testovala výběr 3, 6 a 9 stromů ve třech čtyřcentimetrových tloušťkových intervalech s rozmezí minima tloušťky, střední tloušťky a maxima tloušťky. Tyto varianty nejsou stanoveny žádnou obecně uznávanou metodikou, ale autor práce využil své zkušenosti pro stanovení výběru v kombinaci s inspirací z dostupných prací jiných autorů. Obecně je možné konstatovat, že každý autor smíšeného modelu výškové funkce ve své práci testoval jiný výběr stromů pro účely podmíněné kalibrace. Například Adame et al. (2008) a Kangas a Haara (2012) vybírali náhodně 1 – 4 stromy, Sharma a Parton (2007), Vargas-Larreta et al. (2009) a Schmidt et al. (2011) vybírali náhodně pouze 1 strom, Trincado et al. (2007) vybrali náhodně 1 – 3 stromy. Drápela (2011) ve své práci otestoval náhodný výběr 1, 3 a 5 stromů a výběr 1, 3 a 5 stromů, které náležely do tloušťkových intervalů, které vznikly rozdělením tloušťkového rozpětí do pěti 20 % kvantilů. Zřejmě nejsložitější metodika výběru stromů se objevuje v pracích autorů Crecente-Campo et al. (2010), Özçelik et al. (2013) a Castaño-Santamaría et al. (2013), kteří hodnotili výběr 1 - 10 stromů náhodně vybraných, 1 -10 stromů s největší tloušťkou, 1 - 10 stromů s nejmenší tloušťkou, 1 - 10 stromů se střední tloušťkou a 3, 6 a 9 stromů v tloušťkových stupních s nejmenší, průměrnou a největší tloušťkou.

Z výsledků této práce je patrné, že výběr 1 - 5 stromů se střední tloušťkou není schopen kvalitní podmíněné kalibrace modelu, protože predikované hodnoty výšek jsou buď velmi odlišné od predikcí lokálního modelu, anebo zcela chybné. K podobným závěrům došli ve své práci i Crecente-Campo et al. (2010), kteří zjistili, že při využití 1 – 10 stromů se střední tloušťkou má výsledný model vysokou střední hodnotu reziduí a velmi vysokou hodnotu střední chyby regresního modelu, takže jeho využití také nedoporučili. Castaño-Santamaría et al. (2013) a Özçelik et al. (2013), kteří také hodnotili kalibraci se stromy se střední tloušťkou, došli sice k podobným závěrům, ale jejich výsledné modely neměly tak vysoké hodnoty středních hodnot reziduí a středních chyb

regresního modelu, takže použití těchto modelů zcela nezamítli. Autoři prací shodně doporučují použití spíše stromů s nižšími tloušťkami, protože při nich dosáhli lepších výsledků.

Co se týče druhé varianty výběru stromů použité v této práci, tak její výsledky potvrzují, že použití tlouštěk z celého tloušťkového rozpětí (respektive minima, středu a maxima) je jednoznačně lepší a se zvyšujícím se počtem stromů v jednotlivých tloušťkových intervalech se přesnost predikce zlepšuje. Za dostatečně přesné výsledky kalibrované predikce je možné považovat ty, které vznikly na základě měření šesti a devíti stromů. Výsledky plně korespondují s pracemi Crecente-Campa et al. (2010), Drápely (2011), Castaño-Santamaría et al. (2013) a Özçelika et al. (2013), kteří všichni testovali kalibrace, pro které byly stromy vybírány z celého rozpětí tlouštěk. I jejich výsledky potvrzují, že použití více tloušťkových intervalů pro výběr stromů zlepšuje predikční schopnosti modelu. Všichni autoři se shodují, že čím více je změřených stromů pro kalibraci (bez ohledu na to, v jakém tloušťkovém intervalu se stromy nacházejí), tím je predikce přesnější (nižší hodnota střední chyby reziduí, střední chyby regresního modelu i průměrné hodnoty odchylek lokálního a smíšeného modelu). Stejně tvrzení lze nalézt i v pracích Trincada et al. (2007), Adame et al. (2008), Kangas a Haara (2012).

Všichni autoři, kteří využili ve své práci jak podmíněnou, tak i nepodmíněnou kalibraci se shodují, že podmíněná kalibrace je vždy výrazně lepší ve všech hodnocených kritériích (nejčastěji střední chyba regresního modelu a střední hodnota reziduí) a to dokonce i v případech, kdy je ke kalibraci využit pouze jeden náhodně vybraný strom (např. Sharma a Parton (2007), Vargas-Larreta et al. (2009) a Schmidt et al. (2011)).

Pokud doporučené počty stromů k podmíněné kalibraci porovnáme s počty doporučenými pro tvorbu lokálního modelu (20-25 výšek) autory Van Laar a Akça (2007) nebo počty obecně doporučovanými (3-5 pro každý tloušťkový stupeň) pro tvorbu modelu výškové křivky, využitelné pro stanovení zásoby porostu pomocí objemových tabulek, zjistíme, že smíšený model s podmíněnou kalibrací povede ke snížení nákladů na měření výšek v porostu, což dělá tuto metodu vysoce efektivní a užitečnou (Crecente-Campo, 2010). Nicméně je nutné brát v potaz připomínku autorů Kangas a Haara (2012), kteří upozorňují na to, že každé další měření výšek je nákladné (a tedy i při kalibraci) a je nutné, aby zlepšení modelu, které změření těchto výšek přinese, mělo přínos minimálně ve stejné výši, jako jsou právě náklady na změření výšek.

## 6.4 GAM model

Vztahy mezi závisle proměnou a nezávisle proměnnými jsou někdy tak složité, že je nutné vynaložit velké úsilí, aby byla nalezena správná matematická funkce, která je schopna popsat tento vztah (Robinson et al. 2011). Navíc mohou do tohoto vztahu vstupovat i další proměnné (např. porostní, klimatické či geografické), takže výsledkem může být tak silně nelineární vztah, že nemusí být ani běžnými funkcemi jednoduše popsitelný.

V této práci byly sestaveny GAM modely výškové křivky se třemi různými typy vyhlazovacích funkcí. Jednalo se konkrétně o funkce LOESS, spline a kubický spline. Tyto vyhlazovací funkce jsou doporučovány jako vhodné pro modelování různých lesnických modelů např. i autory Venables a Ripley (1997), Moisen a Frescino (2002), Wang et al. (2005b), Zhang et al. (2008), Falk a Mellert (2011) a Yue et al. (2012). Drápela (2011) označil funkce LOESS a kubický spline za vůbec nejpoužívanější.

Volba optimální míry vyhlazení pro splinové funkce byla prováděna pomocí zobecněné krosvalidace. V případě globálního modelu se tato metoda plně osvědčila a lze ji plně doporučit, ale na lokálním modelu v několika případech nedokázala zabránit přeurčenosti modelu. Právě přeurčenost označili Robinson et al. (2011) za jeden ze základních problémů GAM modelů. Krosvalidaci, jako velmi dobrou metodu, ve svých pracích doporučují také autoři Albert a Schmidt (2010) nebo Falk a Mellert (2011). Falk a Mellert (2011) ale zároveň upozornili na to, že je nutné brát jako důležité kritérium také ekologickou srozumitelnost modelu, což ve své práci zdůrazňují také Austin et al. (2006).

Výsledkem ověření použitelnosti GAM modelů jako alternativy pro výškovou funkci je konstatování, že GAM model je vhodný pro modelování výškové křivky, a to jak na úrovni globálního, tak i lokálního modelu. Tato metoda zatím není pro tuto problematiku příliš používaná, ale podobné závěry lze nalézt v pracích autorů Zhanga et al. (2008) a Schmidta et al. (2011). Zhang et al. (2008) sestavili GAM model výškové křivky s využitím jednoduché splinové funkce pro různé druhy dřevin pro oblast Ontaria v Kanadě. Konstatovali, že metoda GAM je vhodná pro modelování závislosti výšky a výčetní tloušťky stromu. V rámci svých výsledků potvrdili také hypotézu, že GAM modely mají menší střední hodnoty reziduí než klasické regresní modely počítané metodou nejmenších čtverců. V této práci tato hypotéza statisticky potvrzena nebyla, i když bodové odhady středních hodnot reziduí ji mohou naznačovat. Schmidt et al. (2011) vytvořili ve své práci GAM model výškové křivky pro borovicí

lesní (*Pinus sylvestris* L.) v Estonsku. Jedním z jejich závěrů je ten, že GAM model je velmi vhodný pro modelování výškové křivky. Rozdíl oproti GAM modelům použitým v této práci byl pouze v tom, že autoři pro jejich model využili penalizovaný kubický spline, který nepracuje s různým počtem uzlů, ale s pevně daným vysokým počtem uzlů a při krosvalidaci se hledá optimální hodnota tzv. vyhlazovacího parametru  $\lambda$ . Obě výše uvedené práce docházejí tedy ke stejným závěrům, jaké byly vyvozeny autorem této práce. Výsledky uvedené v této práci byly potvrzeny pouze pro základní vztah výšky a tloušťky stromu bez uvažování dalších proměnných. I když je obecně uvažováno, že GAM model by měl být schopen pracovat např. i s porostními proměnnými, tak v tomto případě nebyla tato hypotéza potvrzena. Důvodem pro její zamítnutí byla nekvalitní predikce, která neodpovídala podmínkám na predikci pomocí výškové funkce, a to vždy na části datového souboru. Toto se nepodařilo odstranit ani použitím modelů s jinými vyhlazovacími kritérii (šířka jádra apod.).

Jako jeden z výsledků této práce je uvedeno konstatování, že GAM modely jsou doporučeny jako nástroje pro velmi přesnou predikci. Toto se shoduje s tvrzením autorů i jiných prací zabývajících se GAM modely v lesnictví. Například Zhang a Gove (2005) tvrdí, že GAM modely jsou schopny lepší predikce díky svojí robustnosti a flexibilitě. Stejně tak Moisen et al. (2006) poukazují na kvalitní predikci a jednoduchou interpretovatelnost GAM modelů. Aertsen et al. (2010) doporučují GAM modely v případech, kdy je kladen důraz právě na kvalitu predikce. Robinson et al. (2011) doporučují GAM modely vhodné pro lesnické modelování ve chvíli, kdy odhad parametrů modelu není prvořadým úkolem, ale naopak kdy cílem je co nejpřesnější predikce. Robustnost GAM modelů potvrzuje také konstatování Wang et al. (2005b), kteří ve své srovnávací studii dospěli k závěru, že GAM modely jsou mnohem adaptabilnější na extrémní hodnoty při tvorbě modelu než jiné nelineární a neparametrické metody, což ve své práci potvrzují i Byun et al. (2013).

Souhrnně lze konstatovat, že GAM modely dovolují vytvářet takové modely, které jsou schopny dobře popisovat biologické děje (Frescino et al. 2001; Austin, 2002; Lehmann et al. 2003) a jsou schopny popsat složité nelineární vztahy, které se velmi často vyskytují v ekosystému (Byun et al. 2013). Vhodné příklady modelování biologických dějů lze nalézt i v lesnictví. Zhang et al. (2008), kteří modelovali výškovou křivku pomocí GAM modelu, dospěl k závěrům, že lokální model sestavený pomocí GAM je stejně kvalitní jako model vypočítaný metodou nejmenších čtverců. Toto plně odpovídá výsledkům, které byly zjištěny i v této práci. Schmidt et al. (2011) konstatovali,

že GAM technika je vhodná pro modelování výškové křivky a jejich model obsahoval i porostní proměnnou střední tloušťky, takže byl schopen sestavit i systém výškových křivek. Robinson et al. (2011) otestovali GAM model pro tvorbu morfologické křivky kmene šesti druhů dřevin a označili ji jako vhodnou metodu pro modelování tohoto silně nelineárního vztahu. Byun et al. (2013) dokázali pomocí GAM vytvořit model silně nelineárního vztahu radiálního růstu borovice hustokvěté (*Pinus densiflora* Sieb. Et Zucc.) a dubu (*Quercus* spp.) v závislosti na věku, množství srážek, teplotě a indexu vlhkosti klimatu. Jejich výsledky plně odpovídají tvrzení o vhodnosti této metody pro silně nelineární vztahy mezi proměnnými. Wang et al. (2005b) využili GAM metodu k sestavení modelu bonity borovice pokroucené široolisté (*Pinus contorta* var. *latifolia* (Engelm.) Critchf.) v závislosti na šesti klimatických charakteristikách, nadmořské výšce a geografické poloze. GAM model porovnali s nelineární regresí, metodou neuronových sítí a stromovým modelem. Právě GAM model a neuronové sítě označili ve své práci za výrazně lepší než klasický nelineární model, a to především z toho důvodu, že pomocí metody GAM byli schopni sestavit jednodušší model, který měl lepší predikční schopnosti než nelineární model.

## 6.5 CART model

V této práci byl sestaven regresní strom pro určování výšek stromů v závislosti na výčetní tloušťce a různých porostních veličinách. Důvodů pro rozhodnutí otestovat tuto metodu jako alternativu pro modelování výšek je celá řada. Fan et al. (2006) označili tuto metodu jako neparametrickou, která neklade žádné požadavky na rozdělení reziduí. Gómez et al. (2012) dodávají, že nevyžaduje žádné konkrétní rozdělení vstupních dat, je schopna zpracovat silně nelineární vztahy mezi proměnnými a je robustní vůči odlehlým hodnotám ve vstupních datech. Podle autorů De'atha a Fabriciuse (2000) a Shifleye et al. (2006) je velmi vhodná pro popis struktury dat stejně jako pro predikci. Z dosažených výsledků je patrné, že se jedná o rychlou a efektivní metodu, která díky svojí relativní jednoduchosti může nalézt široké uplatnění v oblasti modelování lesa. Jelikož bylo na sestavení regresního stromu využito více proměnných, které jsou silně korelované (např. věk porostu a střední tloušťka), tak by se dal očekávat problém multikolinearity. Tento problém ale nebyl v průběhu analýzy prokázán jako významný, takže lze metodu označit jako robustní vůči multikolinearitě. Toto tvrzení plně odpovídá výsledkům práce Shifleye et al. (2006), kteří konstatovali, že metoda CART se dokáže

vypořádat s problémem multikolinearity, a to tak, že jako důležitější proměnné bere ty, které jsou v hierarchické struktuře stromu umístěny výše a výsledný model více závisí na významu jednotlivých proměnných v modelu než na jejich vzájemné korelaci, takže nedochází k vychýleným odhadům parametrů. Aertsen et al. (2010) ve své práci tvrdí, že interakce mezi proměnnými je vysvětlována zcela automaticky, vzhledem k hierarchické struktuře regresního stromu, protože hodnota závisle proměnné stanovená podle hodnoty nezávisle proměnné je zároveň závislá na hodnotách nezávisle proměnných použitých výše ve stromu. Právě velmi jednoduše popsatelné interakce mezi proměnnými dělají z této metody nástroj s dobrou interpretací výsledného stromu především v případě, kdy počet nezávisle proměnných stoupá a jiné typy modelů jsou již velmi složitě interpretovatelné (De'ath a Fabricius 2000).

I přes výše uvedené výhody této metody je nutné upozornit i na její rizika. Mezi nejzávažnější patří fakt, že metoda je často volena ve chvíli, kdy je kladen důraz pouze na co největší uživatelskou přijatelnost (Aertsen et al. 2011) a Elith et al. (2008) uvádějí, že někdy až příliš zjednodušuje reálné vztahy mezi veličinami.

Souhrnně lze výsledky použití této metody jako alternativního modelovacího přístupu shrnout pomocí tvrzení Gómeze et al. (2012), kteří konstatují, že díky jednoduché struktuře (danou malým počtem dělicích pravidel a koncových skupin) klasifikační a regresní stromy představují praktický a šetrný nástroj pro klasifikaci lesních porostů v rámci lesního hospodaření. Titiž autoři zároveň dodávají, že stromové modely jsou snadno interpretovatelné a aplikovatelné a při splnění několika málo statistických podmínek je lze označit za vhodnou metodu pro odhady různých veličin v lesnictví. Toto se plně potvrdilo i v rámci této práce, protože výsledný strom byl velmi jednoduše sestrojen a má i snadnou intepretaci a velmi dobře predikuje výšky stromů v lesních porostech. Stejně tak označili tuto metodu jako vhodnou pro lesnické modelování autoři Dobbertin a Biging (1998), Fan et al. (2006) a Shifley et al. (2006), kteří ji použili k odhadu mortality stromů, Aertsen et al. (2010), kteří ji využili k predikci bonity anebo Räty a Kangas (2008), kteří takto určovali výtvarnici vzorníkových stromů v rámci Finské národní inventarizace lesa.

## 7 Praktická využitelnost hodnocených metod

### 7.1 Kalibrovaný smíšený model

Hlavní praktický přínos metody spočívá v úspoře času (a tedy i snížení ekonomických nákladů) při měření výšek. Velmi jednoduše se dá vypočítat úspora času oproti měření výšek pro sestavení výškové křivky při zjišťování zásoby porostu pomocí objemových tabulek. Můžeme uvažovat, že v lesním porostu se bude vyskytovat v průměru 15 tloušťkových stupňů, z toho 5 bude mít nižší zastoupení četností a 10 vyšší zastoupení četností. Budou změřeny tři výšky pro stupně z nižší četností a pět výšek pro stupně z vyšší četností. Celkový počet je 65 změřených výšek. Pokud použijeme kalibrační postup stanovený v této práci, tak při změření šesti výšek snížíme potřebný čas na necelých 10 % a při změření devíti výšek na 14 %. Když uvážíme, že vyhledání daného stromu a změření jeho výšky trvá průměrně dvě minuty, tak v takovémto vzorovém porostu by se jednalo o úsporu 112 – 118 minut, tedy téměř dvou hodin. Tato úspora nastane v případě, kdy je pro danou oblast sestaven smíšený model. Ten by mohl být vytvořen při zařizování nového LHP na reprezentativním vzorku porostních skupin. Také by mohla k jeho sestavení být využita data z Národní inventarizace lesů (NIL) nebo ze statistické provozní inventarizace, která se dnes začíná více využívat nejen na území národních parků, ale také u soukromých vlastníků.

Metoda kalibrace poskytuje mírně odchylené hodnoty oproti lokálním modelům, ale pokud porovnáme míru odchýlení při změření šesti a devíti výšek (vyjádřenou průměrem středních odchylek smíšeného a lokálního modelu) zjištěnou v této práci s úsporou času uvedenou výše, tak je možné konstatovat, že tato metoda je velmi dobrým kompromisem mezi přesností stanovení modelových výšek a časovou náročností měření výšek a může tak sloužit jako plnohodnotná alternativa pro modelování výškové křivky při stanovování zásoby porostu. Pokud by měly takto stanovené výšky sloužit i k jiným účelům (např. modelování výšek na výzkumných plochách nebo v růstových simulátorech), tak záleží na uživateli tohoto modelu, jakou bude požadovat přesnost. Pokud bude požadovaná přesnost odpovídat výsledkům této práce, tak je možné modelovat tímto způsobem výšky i pro jiné účely. Pokud bude požadovaná přesnost vyšší, pak se jeví jako možné řešení navýšení počtu měřených výšek a otestování nutného počtu výšek pro požadovanou přesnost. Avšak při nutnosti měřit vyšší počet výšek, který se bude blížit počtu měřenému na lokálním modelu, využití této metody ztrácí význam.



## 7.2 Zobecněný aditivní model

Hlavní přínos této metody je především na úrovni lokálního modelu, kdy může být použita jako alternativní řešení pro klasické parametrické funkce. Častým problémem je otázka volby správné výškové funkce. Dle Mehtátala (2004) ale nelze žádnou z funkcí označit za nejlepší, ale při výběru záleží na konkrétní situaci. Dle výsledků se na lokální úrovni nejlépe osvědčila vyhlazovací funkce LOESS, takže můžeme konstatovat, že při stanovování modelu výškové křivky na lokální úrovni může být použit GAM model s vyhlazovací funkcí LOESS namísto klasické výškové funkce a to především v případech, kdy je kladen důraz na co největší kvalitu predikce. Může tedy sloužit nejen k určování výšek pro stanovení zásoby porostu pomocí objemových tabulek, ale také například pro modelování chybějících výšek na výzkumných plochách, při generování výškové struktury porostu nebo při modelování výšek v růstových simulátorech. Při využití GAM modelu se neprovádí odhad parametrů modelu, takže se odstraní problém s volbou prvotních odhadů, které ovlivňují výsledný tvar výškové křivky a tedy i její celkovou kvalitu z pohledu použitelnosti.

Co se týče využití na úrovni globálního modelu, tak výsledky této metody poskytují adekvátní výsledky k metodě nejmenších čtverců. Bohužel se nepodařilo sestavit správně fungující model, který by obsahoval porostní proměnnou, takže globální GAM model bude mít stejné obecné nedostatky jako globální model sestavený pomocí metody nejmenších čtverců. Nelze ho tedy použít k regionalizaci např. až na úroveň porostní skupiny.

U GAM modelu je ale nutné upozornit na to, že výsledný model je řízen daty a musí být kontrolován, aby měl i svoji biologickou opodstatněnost (tedy aby výška stromu s jeho tloušťkou rostla). Splnění této podmínky zaručuje jeho praktickou použitelnost při modelování výšek.

## 7.3 Regresní strom

Metoda CART by mohla být v současné době v praxi využitelná jako alternativa pro modelování výšek při stanovování zásob porostů pomocí metody objemových tabulek. Při této metodě se stanovuje výška buď odečtením z výškového grafikonu sestaveného ručně měřičem, nebo se pro každou tloušťku (resp. tloušťkový stupeň) vypočítá výška ze sestavené výškové funkce. Jelikož jsou ale v objemových tabulkách uvedeny výšky v 1 m intervalu, tak se v praxi postupuje tak, že se modelové výšky

(získané jedním ze způsobů uvedeným výše) zaokrouhlí právě na celé metry. Velmi často se tak stává, že stromy i v několika sousedních tloušťkových stupních mají stejnou modelovou výšku. Pokud bychom si takto získané výšky vynesli do grafu proti jejich tloušťkám, tak bychom viděli v datech jasný trend, který by měl schodovitý charakter. A právě stejný schodovitý charakter poskytuje i metoda regresních stromů. Z praktického pohledu modelování výšek má tato metoda několik nedostatků:

- vyžaduje relativně velký vstupní datový materiál (Fan et al. 2006),
- není vhodná v případech, kdy máme změřené výšky a tloušťky například pouze v jedné porostní skupině.

Na druhé straně má svoji hlavní výhodu především v tom, že je vhodná pro zjemnění globálního modelu na lokální úroveň (porostní skupina), a to tak, že se do regresního stromu zanesou mimo výčetní tloušťky ještě další proměnné. Tyto proměnné jsou rozdílné pro jednotlivé porostní skupiny (věk, bonita, střední tloušťka, počet stromů/ha apod.). Právě zanesením těchto proměnných do modelu jsme schopni změnit použití globálního modelu, který je sestaven pouze na základě výšky a výčetní tloušťky (např. na úrovni lesního majetku, kraje, přírodní lesní oblastí apod.) na aplikaci až na lokální úroveň (porostní skupina). Ráty a Kangas (2008) a Anderson et al. (2010) označili ve svých pracích právě regionalizaci globálního modelu za jednu z hlavních výhod použití CART metody.

V praxi to znamená, že pokud nasbíráme dostatečně velký datový soubor pro globální úroveň (např. úroveň majetku) v hierarchické struktuře jak stromových, tak i porostních nebo i jiných proměnných vyššího řádu, tak jsme schopni poté sestavit regresní strom pro stanovení výšek stromů a takto vytvořený regresní strom poté používat v jednotlivých porostních skupinách. Z pohledu taxátora to znamená, že pokud bude sestaven regresní strom pro dané zájmové území, tak poté v konkrétní porostní skupině pouze změří tloušťky a potřebné porostní veličiny a všechny výšky stromů může určit právě za pomoci regresního stromu. Dojde tedy k časové úspoře. Změření vstupního datového materiálu je sice také časově i finančně nákladné, ale zde se opět otevírá možnost pro další využití dat získaných například v rámci NIL nebo statistické provozní inventarizace.

Je ale také nutné podotknout, že se jedná o metodu, která sdružuje stromy s různými tloušťkami do homogenních skupin, takže v praxi není vhodná v případech, kdy je cílem stanovit co nejpřesněji výšku pro jednotlivé stromy, což se děje často například při stanovení výškové struktury porostní skupiny v rámci chybějících výšek na výzkumných plochách nebo například v růstových simulátorech.

## 7.4 Celkové zhodnocení

Pokud shrneme výše uvedené praktické závěry, tak můžeme konstatovat, že všechny tři testované metody jsou využitelné pro modelování výškové křivky s ohledem na požadovanou míru přesnosti a územní působnost. Za hlavní výhodu můžeme považovat to, že tyto metody se dobře vypořádávají s problémem prostorové variability a hierarchické struktury v datech. U smíšeného modelu a modelu regresního stromu jsou hlavními výhodami regionalizace globálního modelu a snížení počtu měřených výšek v porostní skupině. U aditivního modelu se jedná především o velmi kvalitní predikci bez nutnosti odhadovat parametry modelu. Hlavní nevýhodou smíšeného modelu a modelu regresního stromu je potřeba většího datového materiálu pro vlastní sestavení modelu (obvykle pro rozsáhlejší území), který je pak používán pro kalibraci v konkrétní lokalitě. Tento problém je v podmínkách ČR řešitelný využíváním dat např. NIL nebo statistické provozní inventarizace. Nevýhodou širšího použití GAM modelu je jeho neschopnost provádět regionalizaci globálního modelu.

Nárůst širšího využití těchto metod lze v praxi očekávat v souvislosti s rozvojem moderních technologií pro sběr dat v lesnictví, a to především v oblasti bezkontaktního měření taxačních veličin pomocí leteckých snímků nebo laserového skenování. Výstupem těchto technologií jsou velké datové sady, jejichž zpracování pomocí zkoumaných metod by mělo vést k efektivnějšímu zjišťování stavu lesa při požadované míře přesnosti, a to na různé rozlišovací úrovni od modelování jednotlivých stromů, porostních skupin až například po celou PLO.

## 8 Závěr

Disertační práce řešila téma využití moderních regresních metod pro modelování výškové křivky. Hodnoceny byly vybrány lineární model se smíšenými efekty, zobecněný aditivní model a klasifikační a regresní stromy. Celá práce byla zpracována pro dřevinu smrk. Jako zájmová oblast byl zvolen Školní lesní podnik Masarykův les Křtiny, který slouží jako účelové zařízení Mendelovy univerzity v Brně. Jednalo se o 23 porostních skupin v rozsahu věku 30 – 136 let, které splňovaly určitá kritéria, jejichž společným cílem bylo to, aby v porostní skupině bylo možné naměřit dostatek jedinců pro vytvoření modelu výškové křivky. Současně byly v těchto porostních skupinách zjištěny ještě porostní taxační veličiny jako například střední tloušťka, hektarová kruhová výčetní základna, absolutní výšková bonita a jiné. Celkem bylo změřeno 1590 stromů.

Všechny tři zvolené metody byly hodnoceny jako potencionální alternativy pro modelování výšek namísto výškové funkce.

U lineárního modelu se smíšenými efekty byla kompletní analýza provedena na modelu Pettersonovy výškové funkce. Tato funkce byla vybrána ze šesti hodnocených výškových funkcí, protože splnila všechny požadované podmínky na výškovou funkci a souhrnně se mezi hodnocenými funkcemi jevila jako nejlepší. Byly sestaveny různé typy smíšených modelů a tyto byly testovány vzájemně mezi sebou. Jako výsledný model byl vybrán takový, který zahrnoval náhodnou část odhadu parametru jak u absolutního členu, tak i u regresního členu. Nejednalo se ale pouze o model první úrovně, ale o model druhé úrovně. Výsledný model obsahoval jako prediktor druhé úrovně přirozený logaritmus věku porostní skupiny u absolutního členu a věk porostní skupiny u regresního členu. Tento model splnil požadavky kladené na smíšený model, protože odstranil část nekonstantního rozptylu, objasnil velkou část variability globálního modelu a vytvořil výsledný model, který efektivně využívá lokálních údajů a je zároveň plně použitelný pro zkoumané území.

Sestavený model byl použit k podmíněné kalibraci výškové křivky. Byly otestovány dvě varianty kalibrace. První varianta prováděla kalibraci při změření 1-5 stromů v intervalu  $\pm 2$  cm od střední tloušťky. Druhá varianta vybírala 1, 2 nebo 3 stromy ve třech tloušťkových stupních, takže bylo měřeno celkem 3, 6 nebo 9 stromů. Tloušťkové stupně byly čtyřcentimetrový interval z nejnižších tlouštěk, interval  $\pm 2$  cm od střední tloušťky a čtyřcentimetrový interval z nejvyšších tlouštěk. Po provedené kalibraci bylo zjištěno, že první varianta není z praktického pohledu vůbec využitelná,

protože predikce jsou příliš vychýlené a výsledný model velmi často ani nemá reálný smysl, protože například s rostoucí výčetní tloušťkou modelovaná výška klesá a není tak splněna základní podmínka pro kvalitní model výškové křivky. U druhé varianty bylo konstatováno, že tato již poskytuje výrazně lepší výsledky a ve všech případech měla reálné opodstatnění, ale při měření pouze tří stromů nabývaly průměrné odchylky příliš vysokých hodnot (přes 1 m), i když v jednotlivých případech mohlo docházet k dobré predikci. Tato nevyrovnanost byla odůvodněna vysokým vlivem nadúrovňových nebo podúrovňových stromů na výslednou podobu výškové křivky. Proto kalibrace při měření pouze tří stromů nebyla doporučena. Kalibrace, která byla provedena při změření šesti a devíti stromů poskytovala výrazně lepší a vyrovnanější výsledky a bylo konstatováno, že při vyšším počtu měřených stromů se eliminuje vliv nadúrovňových a podúrovňových stromů na výsledný tvar a pozici výškové křivky. Při použití šesti změřených stromů byla tato metoda doporučena například pro modelování výšek při určování zásoby porostu pomocí objemových tabulek, protože v tomto případě nepřekročila střední hodnota průměrné odchylky modelů 1 m. V případě kdy bylo změřeno devět stromů, střední hodnota průměrné odchylky modelů nepřekročila 50 cm. Tato subvarianta byla tedy doporučena v případech, kdy je kladen důraz na co nejpřesnější predikci výšek s malou dovolenou průměrnou odchylkou, a to například při stanovování chybějících výšek na výzkumných plochách nebo při generování výšek v růstových simulátorech. Souhrnně lze konstatovat, že pro praxi použitelných výsledků lze dosáhnout, pokud bude prováděna druhá varianta kalibrace při změření šesti nebo devíti stromů v závislosti na požadované míře přesnosti.

Zobecněný aditivní model byl ve finální podobě sestaven pro závislost výšky na výčetní tloušťce. Byly otestovány tři typy vyhlazovacích funkcí, a to konkrétně spline, kubický spline a LOESS. Vytvoření GAM modelu proběhlo na úrovni globální i lokální. Na úrovni globálního modelu bylo zjištěno, že není prokazatelný rozdíl mezi modely sestavenými pomocí různých vyhlazovacích funkcí. Na úrovni lokálního modelu již ale bylo konstatováno, že jednoznačně za nejlepší můžeme považovat ty modely, které jsou sestaveny s využitím LOESS funkce, protože u ostatních typů funkcí se vyskytl problém s přeureností modelu a výsledné modely tak ztratily svoje biologické opodstatnění. Souhrnně můžeme tedy zobecněné aditivní modely označit za vhodnou alternativu pro model výškové křivky, ale s podmínkou, že musí být kladen velký důraz na výslednou kontrolu modelu s ohledem na jeho přeurenost a reálný průběh křivky. Je tedy nutné hledat kompromis mezi mírou vyhlazenosti modelu a jeho praktickou

použitelností, který lze najít nejlépe, pokud se bude vždy provádět grafická interpretace výsledného modelu.

Metoda klasifikačních a regresních stromů byla využita pro predikci výšek v závislosti na výčetní tloušťce a různých porostních veličinách. Jelikož byla výška modelována jako spojitá proměnná, byl sestaven regresní strom. Jako významné porostní proměnné byly ve stromu ponechány střední tloušťka, věk porostní skupiny a absolutní výšková bonita. Výsledkem regresního stromu jsou predikované výšky pro homogenní skupiny vzniklé binárním dělením nezávisle proměnných, takže výsledný trend nabývá dojem schodovité regrese. Obdobný trend poskytuje modelování výšek potřebných pro zjištění zásoby porostu pomocí objemových tabulek, protože i zde jsou výšky stanoveny schodovitě (v praxi jsou modelové výšky zaokrouhlovány pouze na celé metry). Výsledky regresního stromu byly ověřeny i na validačních datech a i v tomto případě bylo dosaženo obdobně dobrých výsledků. Díky této shodě je možné konstatovat, že regresní strom je metoda využitelná v praxi pro stanovování výšek při zjišťování zásob porostů pomocí objemových tabulek.

Souhrnným závěrem celé práce může být konstatování, že všechny tři testované moderní regresní metody mohou být při splnění určitých podmínek použity teoreticky i prakticky jako alternativní metody pro modelování výšek stromů, které budou využitelné k různým účelům.

## 9 Summary

The topic of the dissertation thesis was the utilization of modern regression methods for modeling tree height-diameter relationships (H-D model). Linear mixed effects model, generalized additive model and classification and regression trees were chosen as evaluated modern regression techniques. In this thesis Norway spruce was chosen as a tree species for testing regression methods. The study was situated in the Training Forest Enterprise “Masarykův les” Křtiny, which is a special-purpose facility of Mendel University in Brno, where 23 plots were measured. The age of the forest stand was between 30 to 136 years. All plots were selected according to certain special criteria (minimal area, stand density etc.). The reason for these criteria was that in the forest stand must be enough trees to create a model of tree height-diameter relationship. In the plots there were different mensurational variables (for example mean diameter, basal area, site index etc.) were measured too. 1590 trees were measured. All three chosen methods were evaluated as a potential alternative for modeling tree height-diameter relationship.

Linear mixed effects model were analysed with H-D model of Petterson function. This function was chosen between 6 evaluated functions because it had the best comprehensive evaluation by all criteria. Different types of mixed effects models were fitted and compared to each other. The final model had got the random part of an intercept and a regression coefficient too. The model included second order variable. It was the forest stand age for the regression coefficient and a natural logarithm of the forest stand age for the intercept. This model eliminated the problem with heteroskedasticity and cleared up a big part of the variability of global model. Local data are effectively used by the final model. This model is completely applicable for study area.

A model was used for conditional calibration of H-D model. Two variants of the calibration were tested. In the first calibration 1-5 trees were measured. The diameter at breast height (DBH) of these trees was  $\pm 2$  cm of mean DBH. In the second calibration trees were measured in three diameter classes. In every class 1 or 2 or 3 trees were measured, so overall 3-9 trees were measured. Range of the diameter classes was  $\langle d_{\min}; d_{\min} - 4 \text{ cm} \rangle$ ,  $\langle \bar{d} - 2 \text{ cm}; \bar{d} + 2 \text{ cm} \rangle$  and  $\langle d_{\max} - 4 \text{ cm}; d_{\max} \rangle$ . It was found that first variant of calibration isn't usable from the practical point of view. The reasons for that are very biased predictions and unrealistic fitted H-D models (in few examples fitted height declined with rising DBH). The Second variant of the calibration has got better and realistic results of fitted H-D models. Measuring only three trees acquiring

average of too high deviation (over 1m, although in individual cases good prediction could appear (only in few models this mean deviation was smaller). This was due to dominant and subdominant trees which influence the final position of H-D model. That was the reason why the calibration with only three measured trees wasn't recommended. In case of calibration with 6 or 9 measured trees fitted H-D models were much better. It means that the measurement of more trees eliminate the influence of dominant and subdominant trees to the position and shape of H-D model. When the 6 trees were measured this type of calibration was recommended for example for fitting the heights in the standing volume determination by volume tables because in this case mean deviation of the models wasn't bigger than 1 m. Type of calibration with 9 measured trees had got mean deviation of the models smaller than 0.5 m. It was recommended as convenient calibration for the fitting of missing heights in the research plots or for the generating of heights in growth simulators. The reason for that is small mean deviation of the models and more accurate prediction. Calibration approach of the linear mixed effects model could be used in practice in case of that will be used second variant of the calibration with 6 or 9 measured trees according to required accuracy.

Final generalized additive model fitted relationship between height of a tree and diameter at breast height of a tree. Three smoothing function were tested: spline, cubic spline and LOESS. The model was fitted for local and global level too. At the global level statistically significant difference wasn't determined between models fitted by different type of smoothing functions. At the local level this difference was determined. The best models at the local level were fitted by smoothing function LOESS. Some models, which were fitted by other types of smoothing functions were overfitted. This problem caused that the final curves weren't realistic. GAM could be a good alternative for modeling of height-diameter relationship but only in case that final model must be controlled for overfitting and realistic course. The compromise between smoothing of the model and its practical applicability is needed be found. It means that every model must be graphically interpreted.

Classification and regression trees technique was used to predict tree heights in relationship with DBH and different stand variables. Because the height of a tree is continuous variable a regression tree was used. Mean diameter, age of a forest stand and site index as a stand variables and DBH as a tree variable were significant variables in the final regression tree. Fitted heights are predicted for homogenous groups. These groups are the results of independent variables binary division.



The final appearance of the model is a stairway regression. Absolutely the same stairway trend is in the fitting of the heights in the standing volume determination by volume tables (fitted heights are rounded to 1 m in practice). The results of the regression tree were verified by validation data. The validation was successful. This makes it possible to say, that a regression tree is convenient for the fitting of the heights in the standing volume determination by volume tables.

The general conclusion of the thesis is that all three tested modern regression methods can be used (under certain conditions) as a convenient alternative approach for modeling height-diameter relationships in different ways in the forestry.

## 10 Seznam citované literatury

- AAKALA, T., FRAVER, S., D'AMATO, A. W., PALIK, B. J., 2013. Influence of competition and age on tree growth in structurally complex old-growth forests in northern Minnesota, USA. *Forest Ecology and Management*, 308. 128-135.
- ADAME, P., DEL RÍO, M., CAÑELLAS, I., 2008. A mixed nonlinear height–diameter model for pyrenean oak (*Quercus pyrenaica* Willd.). *Forest Ecology and Management*, 256 (1-2). 88-98.
- AERTSEN, W., KINT, V., MUYS, B., VAN ORSHOVEN, J., 2012. Effects of scale and scaling in predictive modelling of forest site productivity. *Environmental Modelling & Software*, 31. 19-27.
- AERTSEN, W., KINT, V., VAN ORSHOVEN, J., MUYS, B., 2011. Evaluation of modelling techniques for forest site productivity prediction in contrasting ecoregions using stochastic multicriteria acceptability analysis (SMAA). *Environmental Modelling & Software*, 26. 929-937.
- AERTSEN, W., KINT, V., VAN ORSHOVEN, J., OZKAN, K., MUYS, B., 2010. Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. *Ecological Modelling*, 221 (8). 1119-1130.
- AKAIKE, H., 1973. Information theory and an extension of the maximum likelihood principle. In: PETROV B. N., CSAKI F.: Proceedings of the 2nd International symposium on information theory, Budapest, Akademiai Kiado. 268-281.
- ALBERT, M., SCHMIDT, M., 2010. Climate-sensitive modelling of site-productivity relationships for Norway spruce (*Picea abies* (L.) Karst.) and common beech (*Fagus sylvatica* L.). *Forest Ecology and Management*, 259 (4). 739-749.
- ALTMAN, N. S., 1992. Introduction to kernel and nearest-neighbour nonparametric regression? *The American Statistician*, 46 (3). 175-185.
- ANDERSON, M. G., FERREE, C. E., OLIVERO, A. P., ZHAO, F., 2010. Assessing floodplain forests: Using flow modeling and remote sensing to determine the best places for conservation. *Natural Areas Journal*, 30 (1). 39-52.
- ANSELIN, L., GRIFFITH, D. A., 1988. Do spatial effects really matter in regression analysis? *Papers - Regional Science Association*, 65:11-34.

- ARNEY, J. D., 1985. A modeling strategy for the growth projection of managed stands. *Canadian Journal of Forest Research*, 15 (3). 511 – 518.
- ASSMAN, E., 1943. Untersuchungen über die Höhenkurven von Fichtenbeständen. *Allgemeine Forst- und Jagdzeitung*, 119. 77-88, 105-123, 133-151.
- ASSMANN, E., 1961. *Waldetragskunde. Organische Produktion, Struktur, Zuwachs und Ertrag von Waldbeständen*. München, BLV Verlagsgesellschaft, 490 s.
- ASSMANN, E., FRANZ, F., 1963. *Vorläufige Fichten – Ertragstafel für Bayern*. München, Institut für Ertragskunde der Forstlichen Forschungsanstalt, 104 s.
- AUSTIN, M. P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, 157 (2-3). 101-118.
- AUSTIN, M. P., BELBIN, L., MEYERS, J. A., DOHERTY, M. D., LUOTO, M., 2006. Evaluation of statistical models for predicting plant species distributions: role of artificial data and theory. *Ecological Modelling*, 199 (2). 197-216.
- BAKER, F.A., 1993. Classification and regression tree analysis for assessing hazard of pine mortality caused by *Heterobasidion annosum*. *Plant Disease*, 77 (2). 136-139.
- BIGGING, G.S., 1985. Improved estimates of site index curves using a varying-parameter model. *Forest Science*, 31 (1). 248-259.
- BIOLLEY, H. H., 1920. *L'aménagement des forêts par la méthode expérimentale et spécialement la méthode du Contrôle*. Paris, Attinger, 90 s.
- BREIMAN, J., FRIEDMAN, J. H., OLSHEN, R. A., STONE, C. G., 1984. *Classification and Regression Trees*. Belmont, California, USA, Wadsworth International Group, 368 s.
- BREZGER, A., LANG, S., 2006. Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics and Data Analysis*, 50 (4). 967-991.
- BUDHATHOKI, C.B., LYNCH, T.B., GULDIN, J.M., 2008. Nonlinear mixed modeling of basal area growth for shortleaf pine. *Forest Ecology and Management*, 255 (8-9). 3440-3446.
- BURNHAM, K. P., ANDERSON, D. R., 1998. *Model selection and inference: A practical information-theoretic approach*. New York, USA, Springer, 355 s.

- BYUN, J. G., LEE, W. K., KIM, M., KWAK, D. A., KWAK, H., PARK, T., BYUN, W. H., SON, Y., CHOI, J. K., LEE, Y. J., SABOROWSKI, J., CHUNG, D. J., JUNG, J. H., 2013. Radial growth response of *Pinus densiflora* and *Quercus* spp. to topographic and climatic factors in South Korea. *Journal of Plant Ecology*, 6 (5). 380-392.
- CALAMA, R., MONTERO, M., 2004. Interregional nonlinear height-diameter model with random coefficients for stone pine in Spain. *Canadian Journal of Forest Research*, 34 (1). 150–163.
- CALAMA, R., MONTERO, G., 2005. Multilevel linear mixed model for tree diameter increment in stone pine (*Pinus pinea*): a calibrating approach. *Silva Fennica*, 39 (1). 37-54.
- CALEGARIO, N., DANIELS, R. F., MAESTRI, R., NEIVA, R., 2005. Modeling dominant height growth based on nonlinear mixed-effects model: a clonal *Eucalyptus* plantation case study. *Forest Ecology and Management*, 204 (1). 11-20.
- CAO, Q., WANG, J., 2011. Calibrating fixed- and mixed-effects taper equations. *Forest Ecology and Management*, 262 (4). 671-673.
- CASTAÑO-SANTAMARÍA, J., CRECENTE-CAMPO, F., FERNÁNDEZ-MARTÍNEZ, J. L., BARRIO-ANTA, M., OBESO, J. R., 2013. Tree height prediction approaches for uneven-aged beech forests in northwestern Spain. *Forest Ecology and Management*, 307. 63-73.
- CASTEDO DORADO, F., BARRIO ANTA, M., PARRESOL, B. R., ÁLVAREZ GONZÁLEZ, J. G., 2005. A stochastic height-diameter model for maritime pine ecoregions in Galicia (northwestern Spain). *Annals of Forest Science*, 62 (5). 455-465.
- CASTEDO DORADO, F., DIÉGUEZ-ARANDA, U., BARRIO ANTA, M., SÁNCHEZ RODRÍGUEZ, M., VON GADOW, K., 2006. A generalized height-diameter model including random components for radiata pine plantations in northwestern Spain. *Forest Ecology and Management*, 229 (1-3). 202-213.
- CLARK, L. A., PREGIBORN, D., 1992. Tree-based models. In CHAMBERS, J. M., HASTIE, T. J. *Statistical models in S*. Pacific Grove, California, USA, Wadsworth & Brooks, 377-419 s.
- CLEVELAND W. S., DEVLIN S. J., 1988. Locally-weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83 (403). 596-610.

- COLBERT, K. C., LARSEN, D. R., LOOTENS, J. R., 2002. Height-diameter equations for thirteen midwestern bottomland hardwood species. *Northern Journal of Applied Forestry*, 19 (4). 171-176.
- CRECENTE-CAMPO, F., TOMÉ, M., SOARES, P., DIÉGUEZ-ARANDA, U., 2010. A generalized nonlinear mixed-effects height-diameter model for *Eucalyptus globulus* L. in northwestern Spain. *Forest ecology and management*, 259 (5). 943-952.
- CURTIS, R. O., 1967. Height-diameter and height-diameter-age equations for second-growth douglas-fir. *Forest science*, 13 (4). 365-375.
- ČERNÝ M., PAŘEZ J., MALÍK Z., 1996. Růstové a taxační tabulky hlavních dřevin České republiky. Jílové u Prahy, IFER - Ústav pro výzkum lesních ekosystémů s.r.o., 245 s.
- DE'ATH, G., FABRICIUS, K. E., 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology*, 81 (11). 3178-3192.
- DÍAZ-URIARTE, R., ALVAREZ DE ANDRÉS, S., 2006. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7. 3.
- DOBBERTIN, M., BIGING, G.S., 1998. Using the non-parametric classifier CART to model forest tree mortality. *Forest Science*, 44 (4). 507-516.
- DOBSON, A. J., 2002. *Introduction to Generalized Linear Models*. Second Edition. Boca Raton, Florida, USA, Chapman&Hall/CRC Press, 221 s.
- DRÁPELA, K., 2011. Regresní modely a možnosti jejich využití v lesnictví. Habilitační práce. Brno, Mendelova univerzita v Brně, 235 s.
- EERIKÄINEN, K., 2003. Predicting the height-diameter pattern of planted *Pinus kesiya* stands in Zambia and Zimbabwe. *Forest Ecology and Management*, 175 (1-3). 355-366.
- EERIKÄINEN, K., MIINA, J., VALKONEN, S., 2007. Models for the regeneration establishment and the development of established seedlings in uneven-aged, Norway spruce dominated forest stands of southern Finland. *Forest Ecology and Management*, 242 (2-3). 444-461.
- ELITH, J., LEATHWICK, J. R., HASTIE, T., 2008. A working guide to boosted regression trees. *Journal of Animal Ecology*, 77 (4). 802-813.
- FABRIKA, M., PRETZSCH, H., 2011. *Analýza a modelovanie lesných ekosystémov*. Zvolen, Technická univerzita vo Zvolene, 599 s.

- FALK, W., MELLERT, K. H., 2011. Species distribution models as a tool for forest management planning under climate change: risk evaluation of *Abies alba* in Bavaria. *Journal of Vegetation Science*, 22 (4). 621-634.
- FAN, Z., KABRICK, J. M., SHIFLEY, S. R., 2006. Classification and regression tree based survival analysis in oak-dominated forests of Missouri's Ozark highlands. *Canadian Journal of Forest Research*, 36 (7). 1740-1748.
- FANG, Z., BAILEY, R. L., 1998. Height-diameter models for tropical forest on Hainan islands in Southern China. *Forest Ecology and Management*, 110 (1-3). 315-327.
- FANG, Z., BAILEY, R. L., 2001. Nonlinear mixed effects modeling for slash pine dominant height growth following intensive silvicultural treatments. *Forest Science*, 47 (3). 287-300.
- FEKEDULEGN, D., MAC SIURTAÍN, M. P., COLBERT, J. J., 1999. Parameter estimation of nonlinear growth models in forestry. *Silva Fennica*, 33 (4). 327-336.
- FERGUSON, I. S., LEECH, J. W., 1978. Generalized least squares estimation of yield functions. *Forest Science*, 24 (1). 27-42.
- FICHTER, A., STURM, K., RICKERT, C., VON OHEIMB, G., HÄRDTLE, E., 2013. Crown size-growth relationships of European beech (*Fagus sylvatica* L.) are driven by the interplay of disturbance intensity and inter-specific competition. *Forest Ecology and Management*, 302. 178-184.
- FOTHERINGHAM, A. S., BRUNSDON, C. A., CHARLTON, M. E., 1998. Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*, 30 (11). 1905-1927.
- FOTHERINGHAM, A. S., BRUNSDON, C. A., CHARLTON, M. E., 2002. Geographically weighted regression: The analysis of spatially varying relationships. New York, USA, John Wiley & Sons, 269 s.
- FOX, J. C., ADES, P. K., BI, H., 2001. Stochastic structure and individual-tree growth models. *Forest Ecology and Management*, 154 (1-2). 261-276.
- FRESCINO, T. S., EDWARDS, T. C., MOISEN, G. G., 2001. Modeling spatially explicit forest structural attributes using generalized additive models. *Journal of Vegetation Science*, 12 (1). 15-26.

- FU, L., SUN, H., SHARMA, R. P., LEI, Y., ZHANG, H., TANG, S., 2013. Nonlinear mixed-effects crown width models for individual tree of Chinese fir (*Cunninghamia lanceolata*) in south-central China. *Forest Ecology and Management*, 302. 210-220.
- FULTON, M. R., 1999. Patterns in height–diameter relationships for selected tree species and sites in eastern Texas. *Canadian Journal of Forest Research*, 29. 1445-1448.
- GÓMEZ, C., WULDER, M. A., MONTES, F., DELGADO, J. A., 2012. Modeling forest structural parameters in the mediterranean pines of central Spain using QuickBird-2 imagery and classification and regression tree analysis (CART). *Remote Sensing*, 4 (1). 135-159.
- GOMPERTZ, B., 1825. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of Royal Society of London*, 115. 513-585.
- GOTTSCHALK, K. W., COLBERT, J. J., FEICHT, D. L., 1998. Tree mortality risk of oak due to gypsy moth. *European Journal of Forest Pathology*, 28 (2). 121-132.
- GREGOIRE, T. G., 1987. Generalized error structure for forestry yield models. *Forest Science*, 33 (2). 423-444.
- GREGOIRE, T. G., SCHABENBERGER, O., BARRETT, J. P., 1995. Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent plot measurements. *Canadian Journal of Forest Research*, 25 (1). 137-156.
- GROOM, J. D., HANN, D. W., TEMESGEN, H., 2012. Evaluation of mixed-effects models for predicting Douglas-fir mortality. *Forest Ecology and Management*, 276. 139-145.
- GUILLEY, E., HERVÉ, J. C., NEPVEU, G., 2004. The influence of site quality, silviculture and region on wood density mixed model in *Quercus petraea* Liebl. *Forest Ecology and Management*, 189 (1-3). 111-121.
- GUISAN, A., EDWARDS JR., T. C., HASTIE, T., 2002. Generalized linear and generalized additive models in studies of species distributions: Setting the scene. *Ecological Modelling*, 157 (2-3). 89-100.
- GUISAN, A., ZIMMERMANN, N. E., 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, 135 (2-3). 147-186.

- GURNAUD, A., 1878. Cahier l'aménagement pour l'application de la méthode par contenance exposée sur la forêt des Eperous. Paris, 160 s.
- GUTZWILLER, K. J., RIFFELL, S. K., 2007. Using statistical models to study temporal dynamics of animal-landscape relations. In BISSONETTE, J. A., STORCH, I. Temporal dimensions of landscape ecology: wildlife responses to variable resources. New York, USA, Springer. 93-118 s.
- HALAJ, J., 1955. Tabuľky jednotných hmotových kriviek pre určovanie hmoty porastov. Bratislava, SVPL, 220 s.
- HALAJ, J., 1963. Tabuľky na určovanie hmoty a prírastku porastov. Bratislava, SVPL, 328 s.
- HALAJ, J., 1974. Výskum konštrukcie rastových tabuliek pre hlavné dreviny ČSSR. Etapa: Prebierky smrekových porastov ČSSR. Čiastková záverečná správa. Zvolen, VÚLH, 94 s.
- HANKE, J.E., WICHERN, D.W., 2005. Business Forecasting. Eight edition. Upper Saddle River, New Jersey, USA, Prentice Hall, 576 s.
- HASTIE, T., TIBSHIRANI, R. J., 1986. Generalized additive models. Statistical Science, 1 (3). 297-318.
- HASTIE, T., TIBSHIRANI, R. J., 1990. Generalized additive models. New York, USA, Chapman and Hall, 335 s.
- HENDL, J., 2012. Přehled statistických metod. Analýza a metaanalýza dat. 4. rozšířené vydání. Praha, Portál, 736 s.
- HÖKKÄ, H., 1997. Height-diameter curves with random intercepts and slopes for trees growing on drained peatlands. Forest Ecology and Management, 97 (1). 63-72.
- HOSSFELD, J W., 1822. Mathematik für Forstmänner, Ökonomen und Cameralisten. Gotha, Hennings, 472 s.
- HUANG S., PRICE D., TITUS S. J., 2000. Development of ecoregion-based height-diameter models for white spruce in boreal forests. Forest Ecology and Management, 129 (1-3). 125-141.
- HUANG. S., TITUS, S. J., 1993. An index of site productivity for uneven-aged and mixed species stands. Canadian Journal of Forest Research, 23 (3). 558-562.



- HUANG, S., TITUS S. J., WIENS, D. D., 1992. Comparison of nonlinear mixed height-diameter function for major Alberta tree species. *Canadian Journal of Forest Research*, 22 (9). 1297-1304.
- HUSCH, B., BEERS, T. W., KERSHAW, J. A., 2003, *Forest mensuration*. 4<sup>th</sup> edition. Hoboken, New Jersey, USA, John Wiley & Sons, Inc., 443 s.
- CHAMBERS, J. M., CLEVELAND, W. S., KLEINER, B., TUKEY, A., 1983. *Graphical methods for data analysis*. Belmont, Duxbury press, 395 s.
- CHAMBERS, J. M., HASTIE, T. J., 1992. *Statistical models in S*. Pacific Grove, California, USA, Wadsworth & Brooks, 608 s.
- JAYARAMAN, K., LAPPI, J., 2001. Estimation of height-diameter curves through multilevel models with special references to even-aged teak stands. *Forest Ecology and Management*, 142 (1-3).155-162.
- KABRICK, J. M., SHIFLEY, S. R., JENSEN, R. G., FAN, Z., LARSEN, D. R., 2004. Factors associated with oak mortality in Missouri Ozark Forests. In *Proceedings of the 14<sup>th</sup> Central Hardwood Forest Conference*. Wooster, Ohio, USDA For. Serv. Gen. Tech. Rep. GTR-NE-316. 27-35.
- KANGAS, A., HAARA, A., 2012. Comparison of nonspatial and spatial approaches with parametric and nonparametric methods in prediction of tree height. *European Journal of Forest Research*, 131 (6). 1771-1782.
- KEELE, L. J., 2008. *Semiparametric Regression for the Social Sciences*. Chichester, John Wiley & Sons Ltd., 230 s.
- KINNUNEN, J., MALTAMO, M., PÄIVINEN, R., 2007. Standing volume estimates of forests in Russia: how accurate is the published data? *Forestry*, 80 (1). 53-64.
- KOUBA, Y., ALADOS, C. L., 2012. Spatio-temporal dynamics of *Quercus faginea* forests in the Spanish Central Pre-Pyrenees. *European Journal of Forest Research*, 131 (2). 369-379.
- KRAMER, H., AKÇA, A., 1995. *Leitfaden zur Waldmesslehre*. Frankfurt a. Main, Sauerländer, J. D., 298 s.
- KRUMLAND, B. E., WENSEL, L. C., 1988. A generalized height-diameter equation for coastal California species. *Western Journal of Applied Forestry*, 3 (4). 113-115.

- KUBLIN, E., BREIDENBACH, J., KÄNDLER, G., 2013. A flexible stem taper and volume prediction method based on mixed-effects B-spline regression. *European Journal of Forest Research*, 132 (5-6). 983-997.
- LAPPI, J., 1986. Mixed linear models for analysing and predicting stem form variation of scots pine. *Communicationes Instituti Forestalis Fenniae*, 134. 1-69.
- LAPPI, J., 1991. Calibration of height and volume equations with random parameters. *Forest Science*, 37 (3). 781-801.
- LAPPI, J., 1997. A longitudinal analysis of height/diameter curves. *Forest Science*, 43 (4). 555-570.
- LARSEN, D. R., HANN, D. W., 1987. Height-diameter equations for seventeen tree species in southwest Oregon. Research paper 49. Corvallis, Forest Research Laboratory, Oregon State University, 16 s.
- LEGENDRE, P., 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology*, 74 (6). 1659-1673.
- LEHMANN, A., SIMONS, J. W., SHRINER, S. A., FRANZREB, K. E., 2002. Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs*, 72 (3). 445-463.
- LEHMANN, A., OVERTON, J. M., LEATHWICK, J. R., 2003. GRASP: generalized regression analysis and spatial prediction. *Ecological Modelling*, 160 (1-2). 165-183.
- LEI, Y., PARRESOL, B. R., 2001. Remarks on height-diameter modelling. Research Note SRS-10. Asheville, North Carolina, USA, USDA Forest Service, Southern Research Station, 8 s.
- LEPŠ, J., ŠMILAUER, P., 2000. Mnohorozměrná analýza ekologických dat, [Online] Dostupné na World Wide Web: <<http://regent.jcu.cz/skripta.pdf>>.
- LESPROJEKT BRNO A.S., 2013. Textová část LHP pro LHC ŠLP Masarykův les Křtiny s platností 1.1.2013 – 31.12.2022. Brno, Lesprojekt Brno a. s., 622 s.
- LEVAKOVIČ, A., 1935. Analytical form of growth laws. *Glasnik za sumske pokuse (Zagreb)*, 4. 189-282.

- LHOTKA, J. M., LOEWENSTEIN, E. F., 2011. An individual- tree diameter growth model for managed uneven-aged oak-shortleaf pine stands in the Ozark Highlands of Missouri, USA. *Forest Ecology and Management*, 261 (3). 770-778.
- LI, R., WEISKITTEL, A. R., 2011. Estimating and predicting bark thickness for seven conifer species in the Acadian Region of North America using a mixed-effects modeling approach: comparison of model forms and subsampling strategies. *European Journal of Forest Research*, 130 (2). 219-233.
- LITTLE, R. J. A., RUBIN, D. B., 1987. *Statistical analysis with missing data*. New York, USA, Wiley, 408 s.
- LÓPEZ SÁNCHEZ, C. A., GORGOSO VARELA, J., CASTEDO DORADO, F., ROJO ALBORECEDA, R., RODRÍGUEZ SOALLEIRO, R., ÁLVAREZ GONZÁLES, J. G., SÁNCHEZ RODRÍGUEZ, F., 2003. A height-diameter model for *Pinus radiata* D. Don in Galicia (Northwest Spain). *Annals of Forest Science*, 60 (3). 237-245.
- LU, J., ZHANG, L., 2013. Evaluation of structure specification in linear mixed models for modeling the spatial effects in tree height-diameter relationships. *Annals of Forest Research*, 56 (1). 137-148.
- MARTIN, F., FLEWELLING, J., 1998. Evaluation of tree height prediction models for stand inventory. *Western Journal of Applied Forestry*, 13 (4). 109-119.
- MCCULLOCH, C. E., SEARLE, S. R., 2001. *Generalized, Linear and Mixed Models*. New York, USA, Wiley-Interscience, 325 s.
- MCGRATH, M. T., DESTEFANO, S., RIGGS, R. A., IRWIN, L. L., ROLOFF, G. J., 2003. Spatially explicit influences on northern goshawk nesting habitat in the interior Pacific North-west. *Wildlife Monographs*, 154. 1-63.
- MEHTÄTALO, L., 2004. A longitudinal height-diameter model for Norway spruce in Finland. *Canadian Journal of Forest Research*, 34 (1). 131-140.
- MEHTÄTALO, L., 2010. *Forest biometrics with examples in R. Lecture notes*. Joensuu, Finland, University of Eastern Finland, School of Forest Sciences, 193 s.
- MENDEZ, G., LOHR, S., 2011. Estimating residual variance in random forest regression. *Computational Statistics and Data Analysis*, 55 (11). 2937-2950.

- MEYER, H. A., 1940. A mathematical expression for height curves. *Journal of Forestry*, 38 (5): 415-420.
- MIGUEL, S., GUZMÁN, G., PUKKALA, T., 2013. A comparison of fixed- and mixed-effects modelling in tree growth and yield prediction of an indigenous Neotropical species (*Centrolobium tomentosum*) in a plantation system. *Forest Ecology and Management*, 291. 249-258.
- MICHAJLOV, I., 1943. Zahlenmäßiges Verfahren für die Ausführung der Bestandeshöhenkurven. *Forstwissenschaftliches Centralblatt. und Tharandter Forstliches Jahrbuch*, 6. 273-279.
- MOISEN, G. G., FREEMAN, E. A., BLACKARD, J. A., FRESCINO, T. S., ZIMMERMANN, N. E., EDWARDS JR., T. C., 2006. Predicting tree species presence and basal area in Utah: A comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. *Ecological Modelling*, 199 (2). 176-187.
- MOISEN, G.G., FRESCINO, T.S., 2002. Comparing five modelling techniques for predicting forest characteristics. *Ecological Modelling*, 157 (2-3). 209-225.
- NANOS, N., CALAMA, R., MONTERO, G., GIL, L., 2004. Geostatistical prediction of height/diameter models. *Forest Ecology and Management*, 195 (1-2). 221-235.
- NÄSLUND, M., 1947. Functions and Tables for Computing the Cubic Volume of Standing Trees: Pine, Spruce and Birch in Southern Sweden and in the Whole of Sweden, Report 36, Stockholm, Sweden, National Forest Research Institute, 81 s.
- NETER, J., WASSERMAN, W., KUNTER, M., 1990. *Applied Linear Statistical Models*. Third ed. Homewood, Illinois, USA, Richard D. Irwin Inc.
- NIGH, G. D., LOVE, B. A., 2004. Predicting crown class in three western conifer species. *Canadian Journal of Forest Research*, 34 (3). 592-599.
- ÖZÇELİK, R., DIAMANTOPOULOU, M. J., CRECENTE-CAMPO, F., ELER, U., 2013. Estimating Crimerian juniper tree height using nonlinear regression and artificial neural network models. *Forest Ecology and Management*, 306. 52-60.
- PARRESOL, B. R., 1992. Baldcypress height–diameter equations and their prediction confidence interval. *Canadian Journal of Forest Research*, 22 (9). 1429-1434.

- PAULSEN, J. C., 1795. Kurze praktische Anleitung zum Forstwesen. Verfaßt von einem Forstmanne. Detmold, Hrsg. Von Kammerrat G. F. Führer, 152 s.
- PENG, C., 1999. Nonlinear height-diameter models for nine boreal forest tree species in Ontario. OFRI – Report no. 155. Sault Ste. Marie, Ontario, Canada, Ontario forest research institute, Ministry of natural resources, 28 s.
- PENG, C., ZHANG, L., HUANG, S., ZHOU, X., PARTON, J., WOODS, M., 2001. Developing ecoregion-based height-diameter models for jack pine and black spruce in Ontario. OFRI – Report no. 159. Sault Ste. Marie, Ontario, Canada, Ontario forest research institute, Ministry of natural resources, 18 s.
- PETTERSON, H., 1955. Barrskogens volymproduktion. Meddelanden från Statens skogsforskningsinstitut, 45 (1A). 1-391.
- PETRÁŠ, R., PAJTIK, J., 1991. Sústava česko-slovenských objemových tabuliek drevín. Lesnícky časopis, 37 (1). 49-56.
- PHILIPP, K., 1924. Hilfstabellen für die Forsttaxatoren. Forstabteilung des Badischen Finanzministeriums. Karlsruhe, Badenia A. G für Verlag und Druckerei.
- POKHAREL, B., DECH, J. P., 2012. Mixed-effects basal area increment models for tree species in the boreal forest of Ontario, Canada using an ecological land classification approach to incorporate site effects. Forestry, 85 (2). 255-270.
- PORTÉ, A., BARTELINK, H. H., 2002. Modelling mixed forest growth: a review of models for forest management. Ecological Modelling, 150 (1-2). 141-188.
- PRASAD, A. M., IVERSON, J. V., LIAW, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems, 9 (2). 181-199.
- PRETZSCH, H., 2009. Forest Dynamics, Growth and Yield: From Measurement to Model. Berlin Heidelberg, Springer-Verlag, 664 s.
- PRODAN, M., 1951. Messung der Waldbestände. Frankfurt a. M., J. D. Sauerländer's Verlag, 260 s.
- RÄTY, M., KANGAS, A., 2008. Localizing general models with classification and regression trees. Scandinavian Journal of Forest Research, 23 (5). 419-430.

- R DEVELOPMENT CORE TEAM, 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, dostupné na World Wide Web: <<http://www.R-project.org/>>.
- REJWAN, C., COLLINS, L. J., BRUNNER, L. J., SHUTER, B. J., RIDGWAY, M. S., 1999. Tree regression analysis on the nesting habitat of smallmouth bass. *Ecology*, 80 (1). 341-348.
- RIPLEY, B. D., 1996. Pattern recognition and neural networks. Cambridge, UK, Cambridge University Press, 416 s.
- ROBINSON, A. P., LANE, S. E., THERIEN, G., 2011. Fitting forestry models using generalized additive models: a taper model example. *Canadian Journal of Forest Research*, 41 (10). 1909-1916.
- ROBINSON, A. P., WYKOFF, W. R., 2004. Imputing missing height measures using a mixed-effects modelling strategy. *Canadian Journal of Forest Research*, 34 (12). 2492-2500.
- ROBINSON, G. K., 1991. That BLUP is a good thing: The estimation of random effects. *Statistical Science*, 6 (1). 15-32.
- ROHNER, B., BUGMANN, H., BIGLER, C., 2013. Estimating the age-diameter relationship of oak species in Switzerland using nonlinear mixed-effects models. *European Journal of Forest Research*, 132 (5-6). 751-764.
- SEARLE, S. R., CASELLA, G., MCCULLOCH, C. E., 1992. Variance components. New York, John Willey, 501 s.
- SHARMA, M., ZHANG, S. Y., 2004. Height-diameter models using stand characteristics for *Pinus banksiana* and *Picea mariana*. *Scandinavian Journal of Forest Research*, 19 (5). 442-451.
- SHARMA, M., PARTON, J., 2007. Height-diameter equations for boreal tree species in Ontario using a mixed-effects modeling approach. *Forest Ecology and Management*, 249 (3). 187-198.
- SCHÄFFER, J., VON WILPERT, K., KUBLIN, E., 2009. Analysis of fine rooting below skid trails using linear and generalized additive models. *Canadian Journal of Forest Research*, 39 (11). 2047-2058.

- SHIFLEY, S. R., FAN, Z., KABRICK, J. M., JENSEN, R. G., 2006. Oak mortality risk factors and mortality estimation. *Forest Ecology and Management*, 229 (1-3). 16-26.
- SCHEER, L'., 2010. *Biometria*. Zvolen, Technická univerzita vo Zvolene, 334 s.
- SCHMIDT, M., KIVISTE, A., VON GADOW, K., 2011. A spatially explicit height-diameter model for Scots pine in Estonia. *European Journal of Forest Research*, 130 (2). 303-315.
- SCHRÖDER, J., ÁLVAREZ GONZÁLES, J. G., 2001. Comparing the performance of generalized diameter-height equations for maritime pine in Northwestern Spain. *Forstwissenschaftliches Centralblatt*, 120 (1). 18-23.
- SIRONEN, S., KANGAS, A., MALTAMO, M., KANGAS, J., 2003. Estimating individual tree growth with nonparametric methods. *Canadian Journal of Forest Research*, 33 (3). 444-449.
- SOARES, P., TOMÉ, M., 2002. Height-diameter equation for first rotation eucalypt plantations in Portugal. *Forest ecology and management*, 166 (1-3). 99-109.
- SPEIDEL, H., 1893. Aus Theorie und Praxis der Forstbetriebseinrichtung. *Allgemeine Forst- und Jagdzeitung*, 69. 145-159, 181-199.
- STATSOFT, INC., 2013. STATISTICA (data analysis software system) version 12, Tulsa.
- SUNANDA, C., JAYARAMAN, K., 2006. Prediction of stand attributes of even-aged teak stands using multilevel models. *Forest Ecology and Management*, 236 (1). 1-11.
- SWIHART, R. K., GOHEEN, J. R., SCHNELKER, S. A., RIZKALLA, C. E., 2007. Testing the generality of patch and landscape-level predictors of tree squirrel occurrence at a regional scale. *Journal of Mammalogy*, 88 (3). 564-572.
- ŠEBÍK, L., POLÁK, L., 1990. *Náuka o produkcii lesa*. 1. vydání. Bratislava, Príroda a. s., 322 s.
- ŠMELKO, Š., 2000. *Dendrometria*. Zvolen, Technická univerzita vo Zvolene, 399 s.
- ŠMELKO, Š., 2007. *Dendrometria*. Vysokoškolská učebnica. Zvolen, Technická univerzita vo Zvolene, 400 s.
- ŠMELKO, Š., WENK, G., ANTANAITIS, V., 1992. *Rast, štruktúra a produkcia lesa*. Bratislava, Príroda a.s., 342 s.

- TEMESGEN, H., VON GADOW, K., 2004. Generalized height-diameter models - an application for major tree species in complex stands of interior British Columbia. *European Journal of Forest Research*, 123 (1). 45-51.
- TEMESGEN, H., ZHANG, C. H., ZHAO, X. H., 2014. Modelling tree height-diameter relationships in multi-species and multi-layered forests: A large observational study from Northeast China. *Forest Ecology and Management*, 316. 78-89.
- TRINCADO, G., VANDERSCHAAF, C. L., BURKHART, H. E., 2007. Regional mixed-effects height-diameter models for loblolly pine (*Pinus taeda* L.) plantations. *European Journal of Forest Research*, 126 (2). 253-262.
- TUKEY, J. W., 1977. *Exploratory data analysis*. Reading, Massachusetts, USA, Adison-Wesley, 670 s.
- UZOH, F. C. C., OLIVER, W. W., 2008. Individual tree diameter increment model for managed even-aged stands of ponderosa pine throughout the western United States using a multilevel linear mixed effects model. *Forest Ecology and Management*, 256 (3). 438-445.
- VANCLAY, J. K., 1994. *Modelling Forest Growth and Yield – Application to Mixed Tropical Forests*. Wallingford, UK, CAB International, 330 s.
- VAN LAAR, A., AKÇA, A., 2007. *Forest mensuration. Managing Forest Ecosystems. Volume 13*. Dordrecht, Holandsko, Springer, 383 s.
- VARGAS-LARRETA, B., CASTEDO-DORADO, F., ALVAREZ-GONZALEZ, F. J., BARRIO-ANTA, M., CRUZ-COBOS, F., 2009. A generalized height-diameter model with random coefficients for uneven-aged stands in El Salto, Durango (Mexico). *Forestry*, 82 (4). 445-462.
- VAYSSIÉNES, M., P., PLANT, R., E., ALLEN-DIAZ, B., H., 2000. Classification trees: An alternative non-parametric approach for predicting species distribution. *Journal of Vegetation Science*, 11 (5). 679-694.
- VENABLES, W. N., RIPLEY, B. D., 1997. *Modern Applied Statistics with S-Plus*. 2nd Edition. New York, Springer, 497 s.
- VONESH, E. F., CHINCHILLI, V. M., 1997. *Linear and nonlinear models for the analysis of repeated measurements*. Boca Raton, Florida, USA, Chapman & Hall/CRC, 560 s.



- VON GADOW, K., REAL, P., ÁLVAREZ-GONZALES, J. G., 2001. Modelización del crecimiento y la evolución de los bosques. Vienna, Austria, IUFRO World Series volume 12, 242 s.
- VYSKOT, M., DOLEŽAL, B., JURČA, J., KORF, V., KORPEL, Š., MACHÁČ, D., POLÁK, L., PRIESOL, A., ŘEHÁK, J., WOLF, J., 1971. Základy růstu a produkce lesů. Praha, Státní zemědělské nakladatelství, 440 s.
- WAHBA, G., 1990. Spline models for observational data. Philadelphia, Pennsylvania, USA, Society for Industrial and Applied Mathematics (SIAM), 169 s.
- WANG, Q., NI, J., TENHUNEN, J., 2005a. Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems. *Global Ecology and Biogeography*, 14 (4). 379-393.
- WANG, Y., RAULIER, F., UNG, C. H., 2005b. Evaluation of spatial predictions of site index obtained by parametric and nonparametric methods - A case study of lodgepole pine productivity. *Forest Ecology and Management*, 214 (1-3). 201-211.
- WEISS, J. *Environmental studies 562: Statistics for Environmental Science*. Lecture notes. [Online] citováno 3. dubna 2007. Dostupné z World Wide Web: <<http://www.unc.edu/courses/2007spring/enst/562/001/docs/lectures.htm>>
- WEST, B., WELCH, K. B., GALECKI, A. T., 2007. *Linear Mixed Models: A Practical Guide using Statistical Software*. Boca Raton, Florida, USA, Chapman & Hall/CRC, 376 s.
- WEST, P. W., 1981. Simulation of diameter growth and mortality in regrowth eucalypt forest of Southern Tasmania. *Forest Science*, 27 (3). 603-616.
- WEST, P. W., RATKOWSKY, D. A., DAVIS, A. W., 1984. Problems of hypothesis testing of regressions with multiple measurements from individual sampling units. *Forest Ecology and Management*, 7 (3). 207-224.
- WIEDEMANN, E., 1936. Über die vereinfachung der Höhenermittlung bei der Vorratsaufnahme. *Mitt. Aus Forstwirtschaft und Forstwissenschaft*, 7. 387-412.
- WOOD, S. N., 2006. *Generalized Additive Models: An Introduction with R*. Boca Raton, Florida, USA, Chapman and Hall/CRC, 410 s.

- YANG, Y., HUANG, S., TRINCADO, G., MENG, S. X., 2009. Nonlinear mixed-effects modeling of variable-exponent taper equation for lodgepole pine in Alberta, Canada. *European Journal of Forest Research*, 128 (4). 415-429.
- YANG, Y., HUANG, S., 2011. Comparison of different methods for fitting nonlinear mixed forest models and for making predictions. *Canadian Journal of Forest Research*, 41 (8). 1671-1686.
- YUE, C., KOHNLE, U., KAHLE, H. P., KLÄDTKE, J., 2012. Exploiting irregular measurement intervals for the analysis of growth trends of stand basal area increments: A composite model approach. *Forest Ecology and Management*, 263. 216-228.
- ZEIDE, B., VANDERSCHAAF, C., 2002. The effect of density on the height-diameter relationship. In *General Technical Report, SRS-48*. Asheville, North Carolina, USA, USDA, Forest Service, Southern Research Station. 463-466.
- ZHANG, L., 1997. Cross-validation of non-linear growth function for modelling tree height-diameter relationships. *Annals of Botany*, 79 (3). 251-257.
- ZHANG, L., BI, H., CHENG, P., DAVIS, C. J., 2004. Modeling spatial variation in tree diameter-height relationships. *Forest Ecology and Management*, 189 (1-3). 317-329.
- ZHANG, L., GOVE, J. H., 2005. Spatial assessment of model errors from four regression techniques. *Forest Science*, 51 (4). 334-346.
- ZHANG, L., GOVE, J. H., HEATH, L. S., 2005. Spatial residual analysis of six modeling techniques. *Ecological Modelling*, 186 (2). 154-177.
- ZHANG, L., MA, Z., GUO, L., 2008. Spatially assessing model errors of four regression techniques for three types of forest stands. *Forestry*, 81 (2). 209-225.
- ZHANG, L., MOORE, J. A., NEWBERRY, J. D., 1996. Evaluation of the prediction performance for eight nonlinear height-diameter equations. In *Proceedings of the 1995 Society of American Foresters Convention*. Bethesda, Maryland, USA, SAF Publication. SAF-96-01. 447-448.
- ZHAO, L., CHUNMING, L., TANG, S., 2013. Individual-tree diameter growth model for fir plantations based on multi-level linear mixed effects models across southeast China. *Journal of Forest Research*, 18 (4). 305-315.

ZUUR, A. F., IENO, E. N., WALKER, N. J., SAVELIEV, A. A., SMITH, G. M., 2009. *Mixed Effects Models and Extensions in Ecology with R*. New York, Springer, 574 s.

## 11 Seznam obrázků

Obr. č. 1: Vývoj množství publikací o MEM v lesnictví za posledních 20 let.....	15
Obr. č. 2: Vývoj množství publikací o GAM v lesnictví za posledních 20 let.....	15
Obr. č. 3: Vývoj množství publikací o CART v lesnictví za posledních 20 let.....	16
Obr. č. 4: Vývoj počtu publikací tří různých regresních metod za posledních 20 let.....	16
Obr. č. 5: Výšková křivka se znázorněním svých vlastností (Šmelko, 2007) .....	18
Obr. č. 6: Systém JVK pro smrk, tloušťková třída 4, střední tloušťka 23-29 cm (Halaj, 1955) .....	19
Obr. č. 7: Vývojové křivky středních porostních výšek na různých bonitách (Černý et al. 1996) .....	20
Obr. č. 8: Systém výškových tarifů pro výběrný smrkový les (Halaj, 1963) .....	22
Obr. č. 9: Princip tvorby LOESS funkce při využití jádrové funkce (Zuur et al. 2009).	31
Obr. č. 10: Zobecněný aditivní model úpravy hodnoty AVB smrku v závislosti na úhrnu srážek za vegetační období (Albert a Schmidt 2010) .....	32
Obr. č. 11: Klasifikační a regresní strom znázorňující pravděpodobnost přežití jednotlivých druhů dřevin v oblasti vysočiny Missouri Ozark v USA (Fan et al. 2006)	35
Obr. č. 12: Mapa s vyznačením zájmového území .....	38
Obr. č. 13: Srovnání globálních modelů šesti výškových funkcí .....	58
Obr. č. 14: Srovnání nelineárního a lineárního globálního modelu Pettersonovy výškové funkce.....	61
Obr. č. 15: Srovnání globálního modelu a lokálních modelů Pettersonovy výškové funkce po linearizaci .....	63
Obr. č. 16: Srovnání reziduí globálního a smíšeného modelu .....	66
Obr. č. 17: Grafy závislosti parametru a modelu první úrovně na porostních veličinách .....	68
Obr. č. 18: Grafy závislosti parametru b modelu první úrovně na porostních veličinách .....	69
Obr. č. 19: Grafy závislosti parametru a modelu první úrovně na logaritmech porostních veličin.....	70
Obr. č. 20: Grafy závislosti parametru b modelu první úrovně na logaritmech porostních veličin.....	71
Obr. č. 21: Srovnání reziduí globálního modelu a smíšeného modelu druhé úrovně.....	73
Obr. č. 22: Grafy závislosti parametru a modelu druhé úrovně a porostních veličin .....	74

Obr. č. 23: Graf závislosti parametru b modelu druhé úrovně a porostních veličin.....	75
Obr. č. 24: Grafický výsledek nevyhovující kalibrace při výběru 1-5 stromů se střední tloušťkou v porostní skupině č. 3.....	77
Obr. č. 25: Grafický výsledek vyhovující kalibrace při výběru 3-9 stromů ve třech tloušťkových intervalech v porostní skupině č. 3.....	78
Obr. č. 26: Vliv velikosti výběru na kvalitu kalibrace při výběru nadúrovňových nebo podúrovňových stromů.....	80
Obr. č. 27: GAM model s využitím spline funkce.....	82
Obr. č. 28: GAM model s využitím kubické spline funkce.....	83
Obr. č. 29: GAM model s využitím funkce LOESS.....	84
Obr. č. 30: Srovnání různých GAM modelů a modelu Pettersonovy výškové křivky vytvořeného nelineární regresí.....	85
Obr. č. 31: GAM modely a nelineární model v porostní skupině 13 a 23.....	89
Obr. č. 32: Významnost jednotlivých prediktorů pro regresní strom.....	91
Obr. č. 33: Kvantil-kvantilový graf reziduí trénovacích a validačních dat.....	92
Obr. č. 34: Grafy reziduí proti predikovaným hodnotám trénovacích a validačních dat	92
Obr. č. 35: Grafy predikovaných hodnot proti měřeným hodnotám trénovacích a validačních dat.....	92
Obr. č. 36: Regresní strom výšek stromů.....	93

## 12 Seznam tabulek

Tab. č. 1: Srovnání lokálního a smíšeného modelu (Weiss, 2007).....	26
Tab. č. 2: Přehled použitých výškových funkcí.....	41
Tab. č. 3: Hodnotící kritéria pro volbu globálního modelu .....	42
Tab. č. 4: Fáze tvorby smíšeného modelu a jejich význam (Drápela, 2011).....	44
Tab. č. 5: Základní statistické charakteristiky prediktorů trénovacích a validačních dat regresního stromu .....	56
Tab. č. 6: Odhady parametrů globálních modelů.....	58
Tab. č. 7: Vypočtené hodnoty AIC pro nelineární lokální modely výškových funkcí ...	59
Tab. č. 8: Výsledky srovnání globálních modelů výškových funkcí.....	61
Tab. č. 9: Odhady parametrů a dalších charakteristik dílčích modelů první úrovně .....	64
Tab. č. 10: Hodnoty náhodných složek a celkových hodnot parametrů smíšeného modelu .....	65
Tab. č. 11: Hodnoty korelačních koeficientů závislosti hodnot parametrů modelu první úrovně a porostních proměnných (a jejich logaritmu) a testových kritérií testu na shodu korelačních koeficientů.....	67
Tab. č. 12: Hodnoty korelačních koeficientů závislosti parametrů modelu druhé úrovně a porostních veličin a hodnoty testových kritérií testů významnosti těchto korelačních koeficientů .....	72
Tab. č. 13: Hodnoty náhodných složek a celkových hodnot parametrů smíšeného modelu druhé úrovně (zahrnut faktor věku) .....	76
Tab. č. 14: Výsledné průměrné hodnoty $i$ s jejich intervaly spolehlivosti středních hodnot odchylek modelových hodnot získaných smíšeným modelem od modelových hodnot lokálního modelu pro jednotlivé porostní skupiny .....	81
Tab. č. 15: Kritéria pro hodnocení různých GAM modelů.....	86
Tab. č. 16: Kritéria pro srovnání lokálních GAM modelů.....	87
Tab. č. 17: Výsledné hodnoty kvalitativních kritérií regresního stromu .....	91
Tab. č. 18: Hodnoty počtu měření, středních hodnot a rozptylů výšek a dělicích kritérií regresního stromu .....	94

## 13 Seznam zkratek

$a, b, c$  – parametry modelu

AIC – Akaikeho informační kritérium

ANN – metoda umělých neuronových sítí

AVB – absolutní výšková bonita

$\mathbf{b} - q$  rozměrný vektor náhodných parametrů pro úroveň strom, plocha a perioda

$\hat{\mathbf{b}}$  – vektor náhodných částí odhadů parametrů

BLUP – nejlepší lineární nevychýlený prediktor

CART – klasifikační a regresní stromy

CHS – cílový hospodářský soubor

ČR – Česká republika

DBH – anglické označení pro výčetní tloušťku stromu

DH – dolní hranice

$d_{\max}$  – největší tloušťka v porostní skupině

$d_{\min}$  – nejmenší tloušťka v porostní skupině

$D_r$  – reziduální deviance

$D_0$  – nulová deviance

$d_{1,3}$  – tloušťka stromu ve výčetní výšce

$\hat{\mathbf{D}}$  – matice variability mezi skupinami

$\bar{d}$  – střední tloušťka porostní skupiny

$e$  – exponenciální funkce

EBLUP – empirický lineární nevychýlený prediktor

EDA – průzkumová analýza dat

$e_i$  – reziduum modelu

$e_{ik}$  – náhodný efekt jedince v rámci experimentální jednotky  $i$

$E(\mathbf{y}|\mathbf{b})$  – odhad závisle proměnné při podmíněné kalibraci

$\hat{\mathbf{e}}$  – vektor, jehož rozměr je dán počtem pozorování a jeho hodnoty jsou rezidua modelu získaného nepodmíněnou kalibrací

$\bar{e}_i$  – střední hodnota reziduí

$f$  – obecné označení matematické funkce

$f_s$  – vyhlazovací funkce výčetní tloušťky

GAM – zobecněný aditivní model

GMEM – zobecněné modely se smíšenými efekty

GPS – globální polohový systém  
 GWR – geograficky vážená regrese  
 $h$  – výška stromu  
 $\mathbf{h}$  – náhodný vektor o velikosti  $j$   
 $\mathbf{h}_1$  – vektor neznámých náhodných částí parametrů  
 $\mathbf{h}_2$  – vektor změřených výšek v dané porostní skupině  
 H-D model – anglické označení pro model výškové křivky  
 HH – horní hranice  
 $h_{max}$  – maximální výška stromu  
 HS – hospodářský soubor  
 $H_0$  – nulová hypotéza  
 $H_1$  – alternativní hypotéza  
 ImpFor – časopis s impakt faktorem v kategorii „Forestry“  
 ImpNonFor – časopis s impakt faktorem mimo kategorii „Forestry“  
 IS – interval spolehlivosti  
 JOK – jednotná objemová křivka  
 JVK – jednotná výšková křivka  
 LHP – lesní hospodářský plán  
 L – věrohodnostní funkce  
 $\ln$  – přirozený logaritmus  
 $\ln T$  – přirozený logaritmus věku porostní skupiny  
 LOESS – druh vyhlazovací funkce založený na lokálně vážené regresi  
 $m$  – počet parametrů modelu  
 MEM – model se smíšenými efekty  
 ML – maximální věrohodnost  
 $m_x$  - část výběrových dat  
 $n$  – velikost výběrového souboru  
 N – normální rozdělení  
 NIL – Národní inventarizace lesů  
 PLO – přírodní lesní oblast  
 Pseudo  $R^2$  – podíl vysvětlené deviance  
 PSK – porostní skupina  
 QQ graf – kvantil-kvantilový graf  
 Rec – plně recenzovaný časopis bez impakt faktoru dohledatelný v databázi SCOPUS



REML – omezená maximální věrohodnost

RMSE – střední chyba regresního modelu

ROC – označení křivky vyjadřující citlivost modelu

$\hat{\mathbf{R}}$  – matice reziduální (vnitroskupinové) variability

$R^2$  – koeficient determinace

SCOPUS – označení celosvětové databáze vědeckých publikací

$SE_{ei}$  – střední chyba reziduí

$S_g(X_g)$  – neparametrická vyhlazovací funkce  $g$ -té nezávisle proměnné  $X$  v aditivním modelu

S-JTSK – systém jednotné trigonometrické sítě katastrální

SM – smrk ztepilý

$S_0$  – absolutní člen aditivního modelu,

ŠLP – školní lesní podnik

$T$  – věk porostní skupiny

$T_k$  – testové kritérium testu významnosti korelačních koeficientů

$T_{\alpha/2}$  – kritická hodnota testu významnosti korelačních koeficientů

$u_{ai}$  – náhodná část absolutního parametru

$u_{bi}$  – náhodná část regresního parametru

$u_i$  – náhodný efekt experimentální jednotky  $i$

$\mathbf{V}$  – matice vyjadřující variabilitu modelu

$\mathbf{V}_1$  – rozptyl vektoru  $\mathbf{h}_1$

$\mathbf{V}_2$  – rozptyl vektoru  $\mathbf{h}_2$

$\mathbf{V}_{12}$  – kovariance mezi vektory  $\mathbf{h}_1$  a  $\mathbf{h}_2$

$\mathbf{X}$  – konstrukční matice o rozměrech  $n \times m$  zahrnující nezávisle proměnné spojené s pevnou složkou modelu

$\mathbf{y}$  –  $n$  rozměrný vektor zahrnující  $n$  pozorování závisle proměnné z  $n_v$  stromů na  $n_u$  plochách během  $n_w$  časových period

$y_i$  – naměřená hodnota jedince  $i$  ( $i=1,2,3,\dots,n$ )

$\hat{y}_{igam}$  – modelová hodnota jedince  $i$  ( $i=1,2,3,\dots,n$ ) vypočítaná aditivním modelem

$y_{ik}$  – měřená hodnota závisle proměnné na jedinci  $k$  na experimentální jednotce  $i$

$\hat{y}_{imm}$  – modelová hodnota jedince  $i$  ( $i=1,2,3,\dots,n$ ) vypočítaná smíšeným modelem

$\hat{y}_{imm\check{c}}$  – modelová hodnota jedince  $i$  ( $i=1,2,3,\dots,n$ ) vypočítaná metodou nejmenších čtverců

$\hat{y}_{intr}$  – modelová hodnota jedince  $i$  ( $i=1,2,3,\dots,n$ ) vypočítaná nelineární regresí

$\bar{y}_i$  – průměrná hodnota ze všech jedinců  $i$  ( $i=1,2,3,\dots,n$ )

$\hat{y}_i$  – modelová hodnota jedince  $i$  ( $i=1,2,3,\dots,n$ )  
 $\mathbf{Z}$  – konstrukční matice o rozměrech  $n \times q$  pro náhodnou složku modelu  
 $Z_k$  – testové kritérium testu na shodu korelačních koeficientů  
 $ZP$  – zkusná plocha  
 $Z_{\alpha/2}$  – kritická hodnota testu na shodu korelačních koeficientů  
 $Z_{\rho 1}, Z_{\rho 2}$  – Fisherovy transformace výběrových korelačních koeficientů  
 $\hat{\mathbf{Z}}^T$  – konstrukční matice náhodných efektů  
 $\boldsymbol{\beta}$  –  $m$  rozměrný vektor pevných parametrů (vypočítané odhady parametrů)  
 $\Delta_i$  – střední hodnota odchylek modelových hodnot získaných smíšeným modelem od modelových hodnot lokálního modelu  
 $\Delta_{ig}$  – střední hodnota odchylek modelových hodnot získaných aditivním modelem od modelových hodnot lokálního modelu  
 $\boldsymbol{\varepsilon}$  –  $n$  rozměrný vektor reziduální složky  
 $\varepsilon_{ik}$  – variabilita nevysvětlitelná modelem  
 $\theta$  – parametry modelu  
 $\lambda$  – vyhlazovací parametr  
 $\mu$  – fixní střední hodnota  
 $\mu_h$  – střední hodnota výšek  
 $\boldsymbol{\mu}_1$  – střední hodnota vektoru  $\mathbf{h}_1$   
 $\boldsymbol{\mu}_2$  – střední hodnota vektoru  $\mathbf{h}_2$   
 $\rho$  – korelační koeficient  
 $\sigma_{ei}$  – směrodatná odchylka reziduí  
 $\sigma_h^2$  – rozptyl výšek  
 $\sigma^2$  – rozptyl vysvětlený v rámci porostní skupiny  
 $\tau_a$  – směrodatná odchylka náhodného efektu absolutního členu  
 $\tau_{ab}$  – kovariance náhodných efektů  
 $\tau_b$  – směrodatná odchylka náhodného efektu regresního členu  
 $\tau^2$  – rozptyl vysvětlený mezi porostními skupinami

## 14 Seznam příloh

Příloha č. 1: Dendrometrické charakteristiky porostních skupin

Příloha č. 2: Základní statistické charakteristiky pro výčetní tloušťku v porostních skupinách

Příloha č. 3: Základní statistické charakteristiky pro výšku v porostních skupinách

Příloha č. 4: Kvantil-kvantilové grafy tlouštěk stromů v jednotlivých porostních skupinách

Příloha č. 5: Kvantil-kvantilové grafy výšek stromů v jednotlivých porostních skupinách

Příloha č. 6: Bagploty výšek a výčetních tlouštěk stromů v jednotlivých porostních skupinách

Příloha č. 7: Grafy reziduí globálních modelů zobrazených proti tloušťce a standardizované tloušťce

Příloha č. 8: Grafy středních hodnot, středních chyb a intervalů spolehlivosti reziduí zobrazených proti tloušťce a standardizované tloušťce pro jednotlivé decily globálních modelů

Příloha č. 9: Zobrazení lokálních modelů šesti výškových funkcí v jednotlivých porostních skupinách

Příloha č. 10: Srovnání nelineárních a lineárních lokálních modelů Pettersonovy výškové funkce na všech porostních skupinách

Příloha č. 11: Srovnání lokálního modelu, modelu s náhodným absolutním členem a modelu s náhodným absolutním i regresním členem ve všech porostních skupinách

Příloha č. 12: Grafy reziduí smíšeného modelu s náhodným absolutním i regresním členem

Příloha č. 13: Srovnání lokálního modelu a smíšeného modelu druhé úrovně, kde je zahrnut faktor věku ve všech porostních skupinách

Příloha č. 14: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 1 – 5 vybraných stromů se střední tloušťkou (výběr simulován) v porostní skupině č. 3

Příloha č. 15: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 3 – 9 vybraných stromů ve třech tloušťkových intervalech (výběr simulován) v porostní skupině č. 3

Příloha č. 16: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 1 – 5 vybraných stromů se střední tloušťkou (výběr simulován) v porostní skupině č. 5

Příloha č. 17: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 3 – 9 vybraných stromů ve třech tloušťkových intervalech (výběr simulován) v porostní skupině č. 5

Příloha č. 18: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 1 – 5 vybraných stromů se střední tloušťkou (výběr simulován) v porostní skupině č. 11

Příloha č. 19: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 3 – 9 vybraných stromů ve třech tloušťkových intervalech (výběr simulován) v porostní skupině č. 11

Příloha č. 20: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 1 – 5 vybraných stromů se střední tloušťkou (výběr simulován) v porostní skupině č. 14

Příloha č. 21: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 3 – 9 vybraných stromů ve třech tloušťkových intervalech (výběr simulován) v porostní skupině č. 14

Příloha č. 22: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 1 – 5 vybraných stromů se střední tloušťkou (výběr simulován) v porostní skupině č. 19

Příloha č. 23: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 3 – 9 vybraných stromů ve třech tloušťkových intervalech (výběr simulován) v porostní skupině č. 19

Příloha č. 24: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 1 – 5 vybraných stromů se střední tloušťkou (výběr simulován) v porostní skupině č. 22

Příloha č. 25: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 3 – 9 vybraných stromů ve třech tloušťkových intervalech (výběr simulován) v porostní skupině č. 22

Příloha č. 26: Hodnotící kritéria kalibrace smíšeného modelu druhé úrovně při změření 3 – 9 vybraných stromů v porostní skupině č. 3

Příloha č. 27: Hodnotící kritéria kalibrace smíšeného modelu druhé úrovně při změření 3 – 9 vybraných stromů v porostní skupině č. 5

Příloha č. 28: Hodnotící kritéria kalibrace smíšeného modelu druhé úrovně při změření 3 – 9 vybraných stromů v porostní skupině č. 11

Příloha č. 29: Hodnotící kritéria kalibrace smíšeného modelu druhé úrovně při změření 3 – 9 vybraných stromů v porostní skupině č. 14

Příloha č. 30: Hodnotící kritéria kalibrace smíšeného modelu druhé úrovně při změření 3 – 9 vybraných stromů v porostní skupině č. 19

Příloha č. 31: Hodnotící kritéria kalibrace smíšeného modelu druhé úrovně při změření 3 – 9 vybraných stromů v porostní skupině č. 22

Příloha č. 32: Grafy pro hodnocení kvality různých typů globálních GAM modelů

Příloha č. 33: Srovnání lokálních GAM modelů různých typů a lokálního modelu Petersonovy funkce vypočítaného nelineární regresí

*Příloha č. 1: Dendrometrické charakteristiky porostních skupin*

identifikátor porostní skupiny	označení porostní skupiny	věk (2011) (roky)	zastoupení SM (%)	střední tloušťka (cm)	střední výška (m)	kruhová výčetní základna/ha (m <sup>2</sup> )	počet stromů/ha (ks)	objem středního kmene (m <sup>3</sup> )	absolutní výšková bonita (m)	lesní typ
1	199Aa3	30	100	20,1	17,9	39,2	1240	0,24	38	4B6
2	203Ca3	30	93	18,6	14,6	22,2	820	0,16	32	3B6
3	178Da3a	35	100	20,8	18,5	35,2	1040	0,26	34	4H1
4	167Ba3a	39	100	21,5	21,5	39,3	1080	0,33	36	3B2
5	20Ca4a	43	100	23,4	21,9	25,5	593	0,40	34	3B2
6	131Ea4	49	94	25,5	25,0	27,4	536	0,54	36	3B2
7	167Aa5	53	80	25,2	25,1	37,4	750	0,53	34	3B2
8	129Ea5	59	80	26,6	25,5	29,0	521	0,60	32	3B2
9	198Ba6	62	87	33,1	28,8	34,9	407	1,02	36	4H1
10	189Aa6	66	89	30,6	29,3	45,2	614	0,90	34	4S1
11	146Ca7	73	80	27,2	27,7	32,4	557	0,68	32	4S1
12	181Aa8	86	80	38,7	33,2	34,8	296	1,59	34	4S1
13	341Da8	88	95	39,8	34,2	27,5	221	1,73	36	3H2
14	200Aa9	93	89	40,0	32,8	35,0	279	1,66	34	3B2
15	178Ba9	94	96	39,5	33,5	40,3	329	1,66	34	4S6
16	178Aa10	109	94	47,4	36,5	46,3	263	2,56	34	4H1
17	178Ba10	109	97	39,5	32,6	36,2	296	1,61	32	4B1
18	173Ca11	111	91	38,8	34,3	28,0	238	1,66	32	3B2
19	184Da11	119	100	39,8	33,4	46,1	371	1,68	32	4S1
20	184Aa12	121	98	37,5	32,5	40,0	363	1,46	32	4S1
21	105Ba12a	123	97	39,2	35,2	38,7	321	1,74	32	3H2
22	194Ca13	133	85	41,4	34,4	31,4	233	1,87	32	3B2
23	185Aa13	136	80	45,3	34,5	36,3	225	2,21	32	4S1

*Příloha č. 2: Základní statistické charakteristiky pro výčetní tloušťku v porostních skupinách*

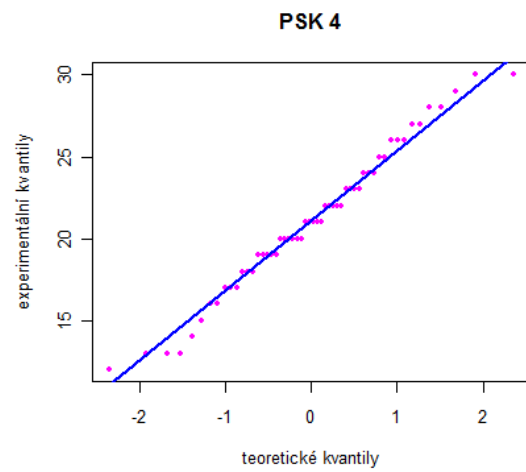
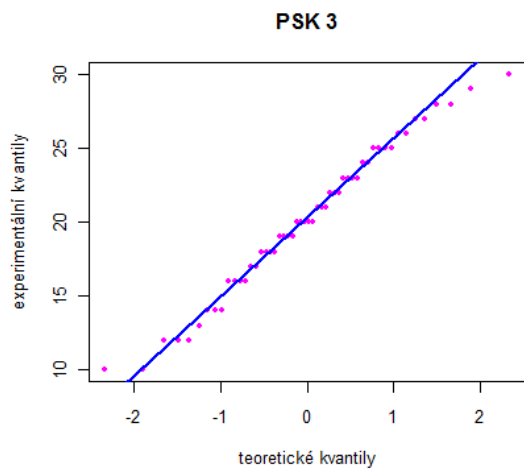
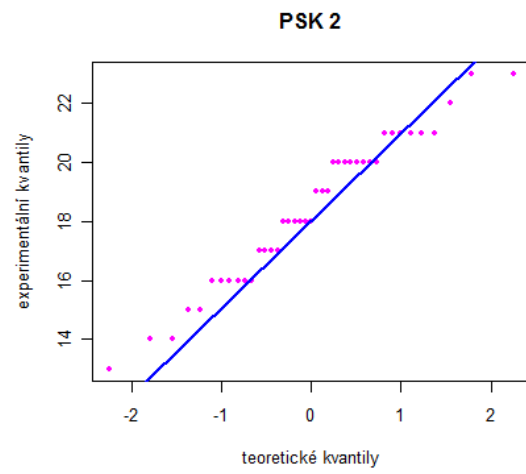
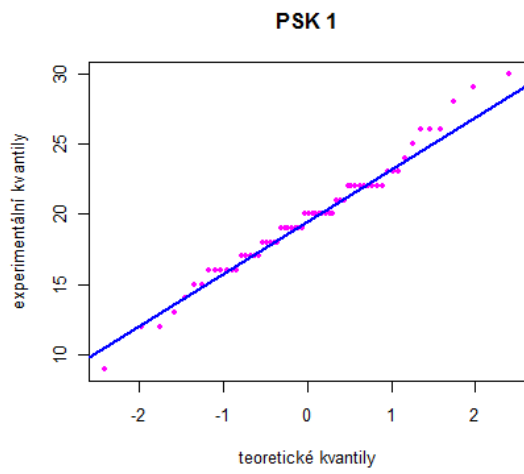
identifikátor porostní skupiny	označení porostní skupiny	počet měření	střední hodnota (cm)	dolní IS (95%) (cm)	horní IS (95%) (cm)	minimum (cm)	maximum (cm)	směrodatná odchylka (cm)	variační koeficient (%)	koeficient šikmosti	koeficient špičatosti
1	199Aa3	62	19,6	18,6	20,7	9,0	30,0	4,1	20,9	0,15	0,56
2	203Ca3	41	18,4	17,6	19,2	13,0	23,0	2,5	13,6	-0,19	-0,70
3	178Da3a	52	20,1	18,7	21,5	10,0	30,0	5,1	25,4	-0,12	-0,66
4	167Ba3a	54	21,1	19,9	22,3	12,0	30,0	4,5	21,3	0,02	-0,42
5	20Ca4a	83	22,8	21,6	24,0	11,0	36,0	5,4	23,7	-0,14	-0,28
6	131Ea4	75	24,9	23,7	26,2	14,0	37,0	5,5	22,1	0,20	-0,67
7	167Aa5	105	24,6	23,4	25,7	14,0	40,0	5,7	23,2	0,23	-0,55
8	129Ea5	73	25,7	24,1	27,3	12,0	49,0	6,8	26,5	0,64	1,54
9	198Ba6	57	32,3	30,4	34,2	17,0	51,0	7,1	22,0	0,07	-0,13
10	189Aa6	86	29,5	27,7	31,3	12,0	53,0	8,3	28,1	0,53	0,24
11	146Ca7	78	26,2	24,5	27,8	14,0	42,0	7,4	28,2	0,22	-1,07
12	181Aa8	71	37,7	35,6	39,7	22,0	63,0	8,7	23,1	0,59	0,40
13	341Da8	53	38,9	36,7	41,2	24,0	59,0	8,3	21,3	0,40	-0,61
14	200Aa9	67	39,2	37,3	41,1	24,0	56,0	7,8	19,9	0,24	-0,67
15	178Ba9	79	38,4	36,2	40,5	19,0	59,0	9,4	24,5	0,54	0,66
16	178Aa10	63	46,2	43,5	48,9	23,0	75,0	10,8	23,4	0,00	0,00
17	178Ba10	71	38,6	36,7	40,5	23,0	67,0	8,0	20,7	0,34	1,35
18	173Ca11	57	37,9	35,7	40,1	23,0	65,0	8,2	21,6	0,35	0,96
19	184Da11	89	39,1	37,5	40,7	22,0	56,0	7,4	18,9	0,00	-0,53
20	184Aa12	87	36,8	35,3	38,3	25,0	60,0	7,2	19,6	0,87	0,76
21	105Ba12a	77	38,4	36,6	40,2	20,0	56,0	7,9	20,6	-0,07	-0,51
22	194Ca13	56	40,5	38,2	42,8	24,0	59,0	8,7	21,5	0,19	-0,76
23	185Aa13	54	44,5	42,2	46,8	22,0	67,0	8,5	19,1	-0,15	0,45

*Příloha č. 3: Základní statistické charakteristiky pro výšku v porostních skupinách*

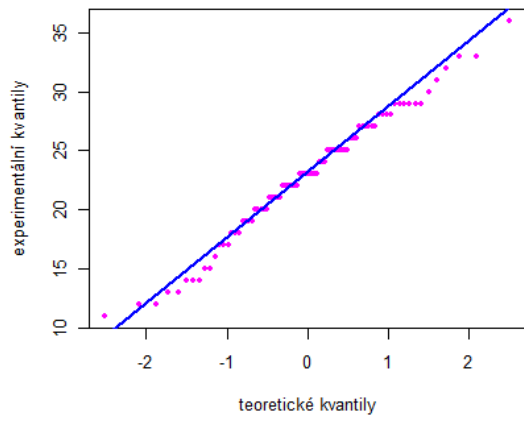
identifikátor porostní skupiny	označení porostní skupiny	počet měření	střední hodnota (m)	dolní IS (95%) (m)	horní IS (95%) (m)	minimum (m)	maximum (m)	směrodatná odchylka (m)	variační koeficient (%)	koeficient šikmosti	koeficient špičatosti
1	199Aa3	62	17,5	17,0	18,0	10,0	21,0	1,9	10,9	-1,20	2,90
2	203Ca3	41	14,4	13,9	14,9	11,0	26,5	1,6	11,1	-0,47	-0,75
3	178Da3a	52	17,7	17,0	18,4	10,0	21,5	2,5	14,1	-1,03	1,39
4	167Ba3a	54	21,0	20,5	21,5	14,0	24,5	2,0	9,5	-1,07	2,12
5	20Ca4a	83	21,1	20,6	20,7	12,5	26,0	2,5	11,8	-1,21	1,96
6	131Ea4	75	24,3	23,6	24,9	18,0	30,5	2,8	11,5	0,08	-0,55
7	167Aa5	105	24,5	24,0	25,0	17,0	29,5	2,6	10,6	-0,56	0,04
8	129Ea5	73	24,8	24,1	25,4	18,0	29,5	2,8	11,3	-0,76	0,16
9	198Ba6	57	28,3	27,6	28,9	22,0	33,0	2,4	8,5	-0,56	0,20
10	189Aa6	86	28,4	27,6	29,2	17,5	34,5	3,7	13,0	-0,94	0,82
11	146Ca7	78	26,4	25,6	27,3	17,0	33,0	3,7	14,0	-0,50	-0,49
12	181Aa8	71	32,5	31,8	33,2	25,5	37,5	2,8	8,6	-0,70	0,22
13	341Da8	53	33,7	33,0	34,3	28,0	39,5	2,5	7,4	-0,10	-0,15
14	200Aa9	67	32,4	31,8	33,0	26,0	38,0	2,6	8,0	0,24	-0,18
15	178Ba9	79	32,4	31,6	33,2	22,5	38,5	3,6	11,1	-0,71	0,08
16	178Aa10	63	35,7	34,7	36,6	22,5	41,5	3,8	10,6	-1,10	1,84
17	178Ba10	71	32,1	31,5	32,7	20,5	37,0	2,6	8,1	-1,31	4,77
18	173Ca11	57	33,7	33,0	34,5	26,0	39,5	2,9	8,6	-0,78	0,73
19	184Da11	89	32,9	32,3	33,4	25,5	37,5	2,6	7,9	-0,69	0,38
20	184Aa12	87	31,9	31,2	32,5	26,5	39,0	3,0	9,4	0,20	-0,79
21	105Ba12a	77	34,7	34,2	35,3	28,5	39,5	2,6	7,5	-0,31	-0,54
22	194Ca13	56	33,4	32,3	34,5	24,0	41,0	4,1	12,3	0,08	-0,55
23	185Aa13	54	33,8	32,8	34,7	22,5	40,5	3,5	10,4	-0,84	1,70



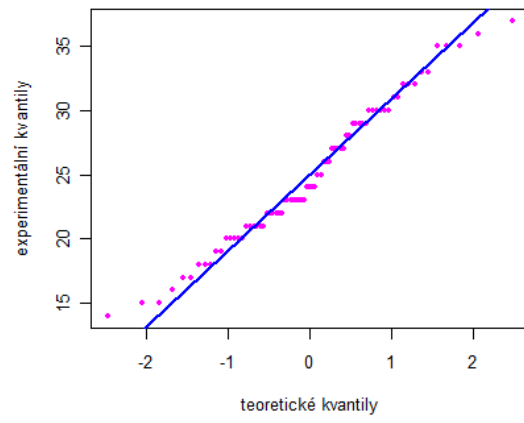
*Příloha č. 4: Kvantil-kvantilové grafy tloušťek stromů v jednotlivých porostních skupinách*



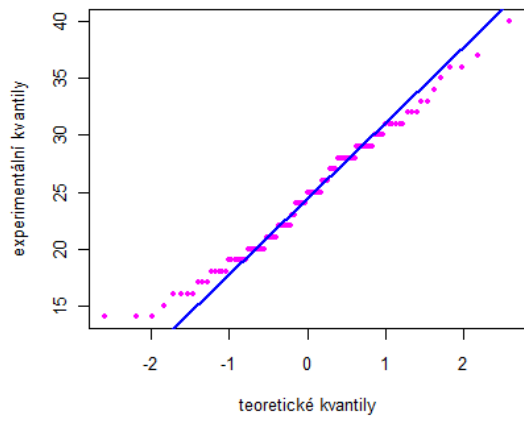
**PSK 5**



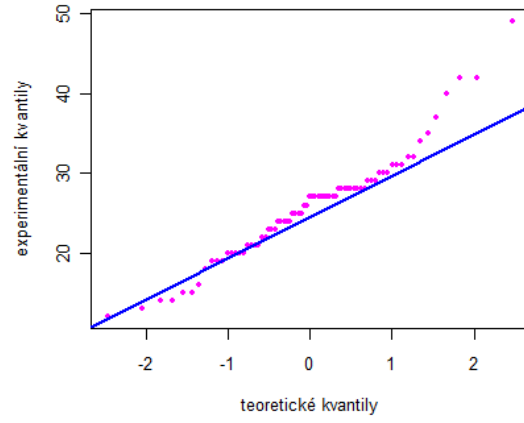
**PSK 6**



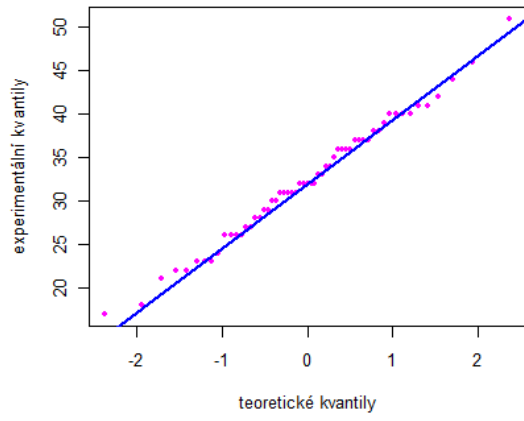
**PSK 7**



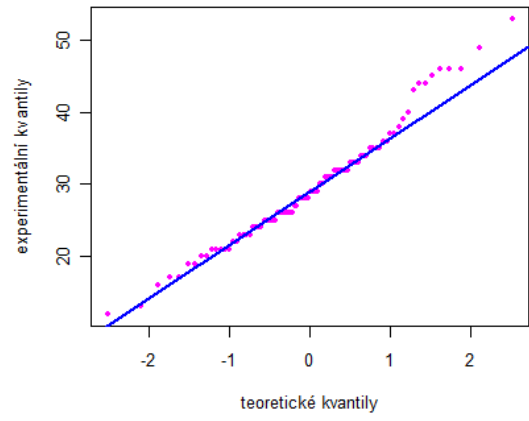
**PSK 8**



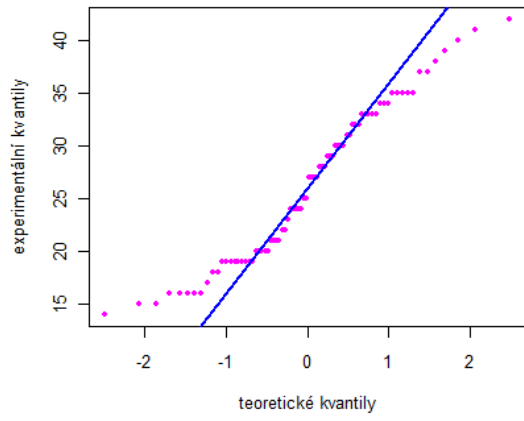
**PSK 9**



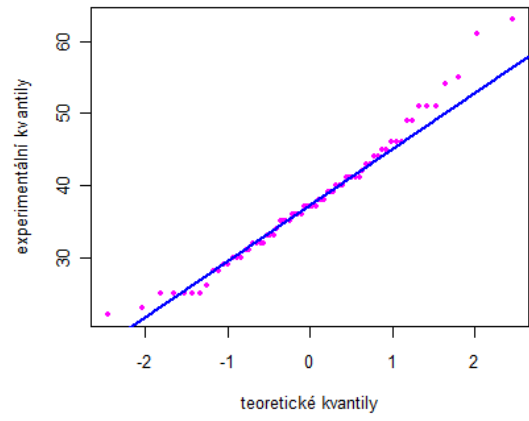
**PSK 10**



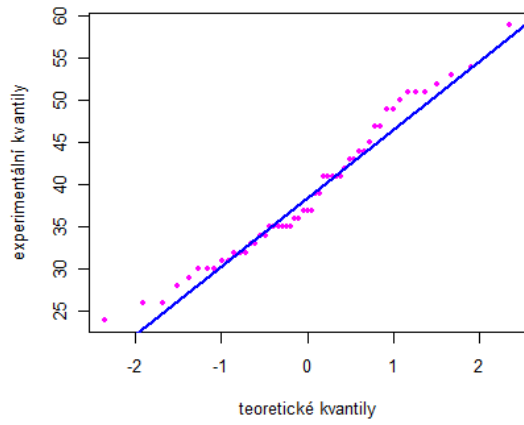
**PSK 11**



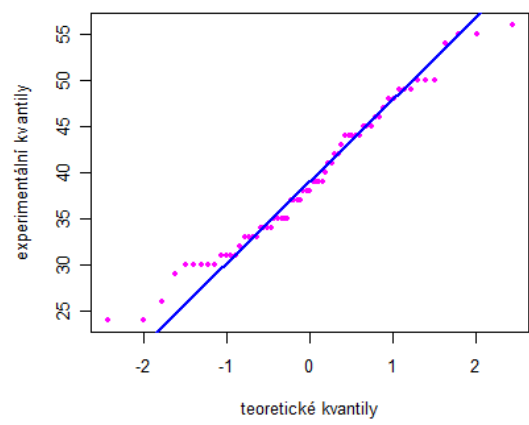
**PSK 12**



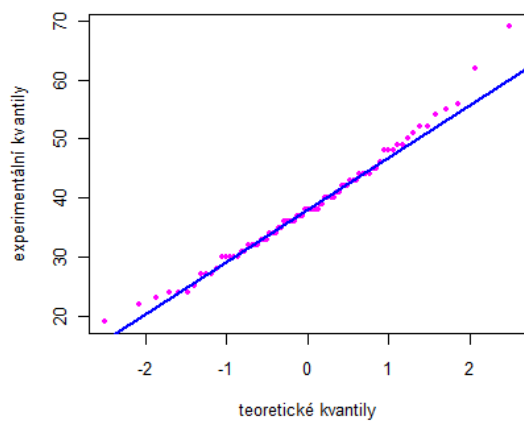
**PSK 13**



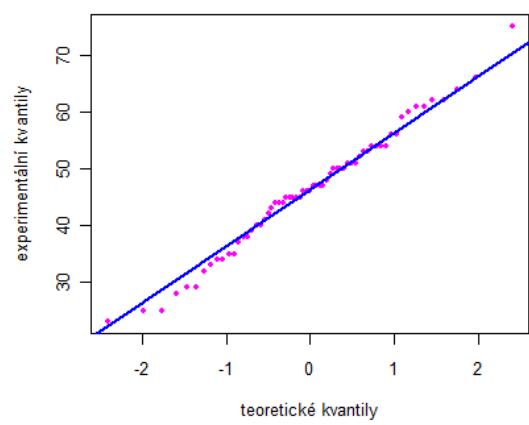
**PSK 14**



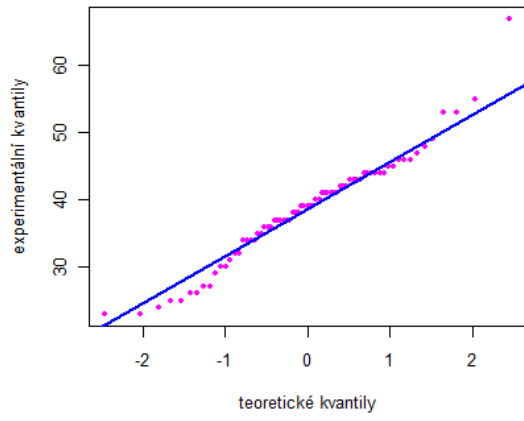
**PSK 15**



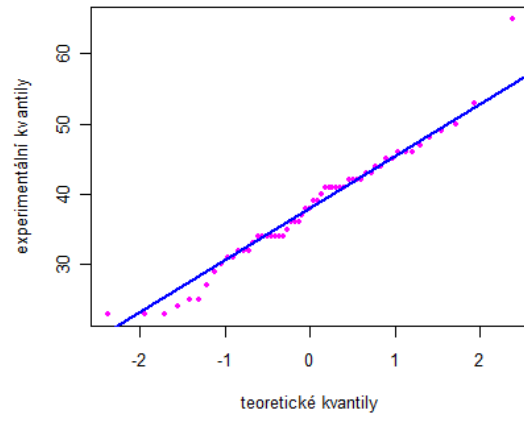
**PSK 16**



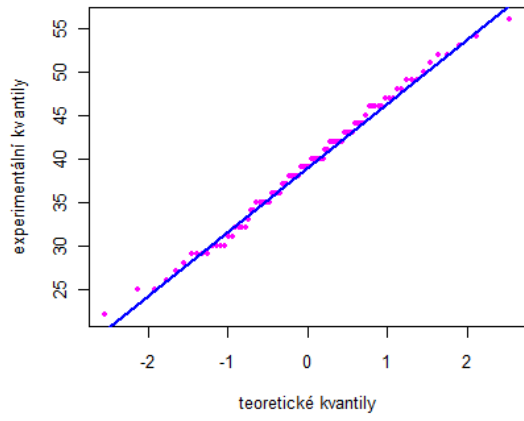
**PSK 17**



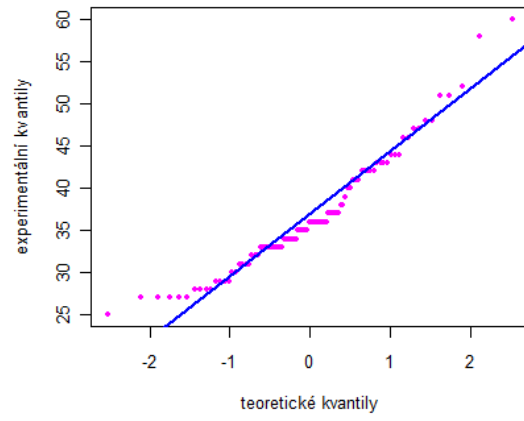
**PSK 18**



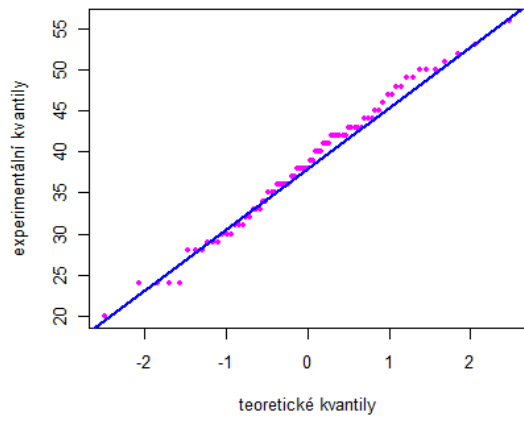
**PSK 19**



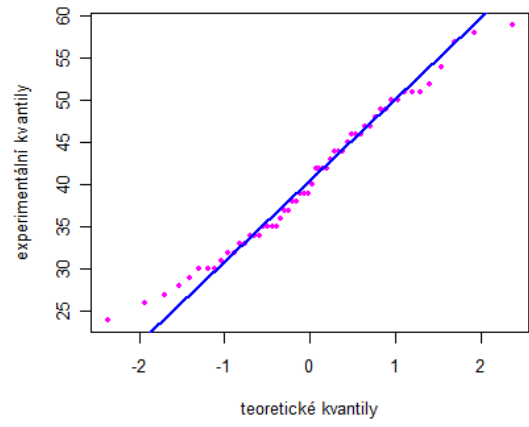
**PSK 20**



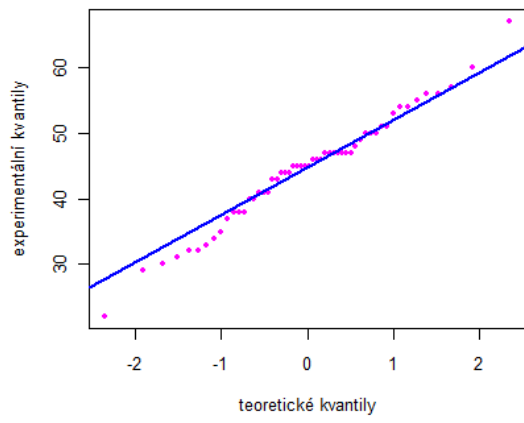
**PSK 21**



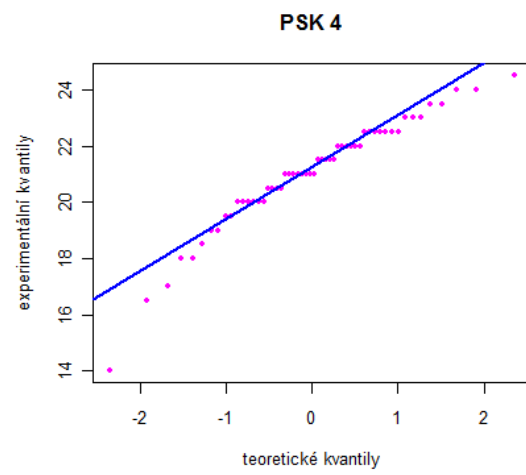
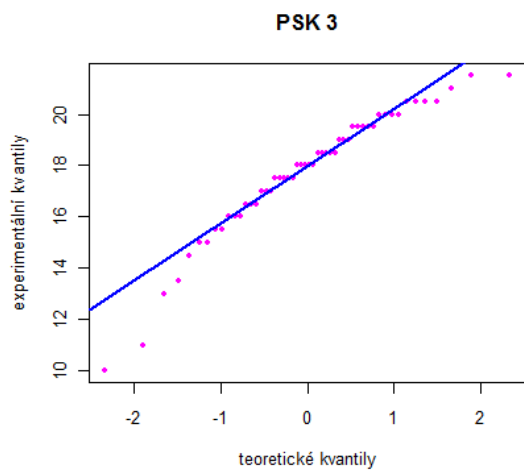
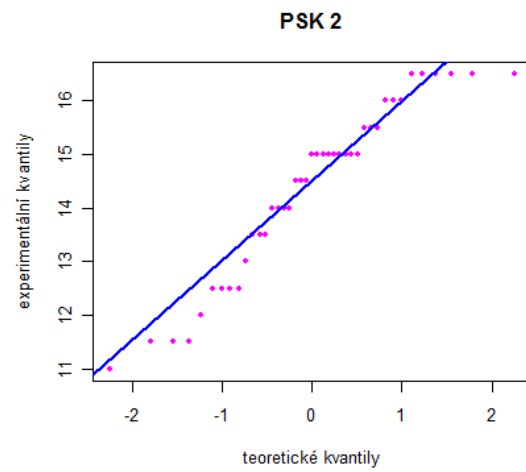
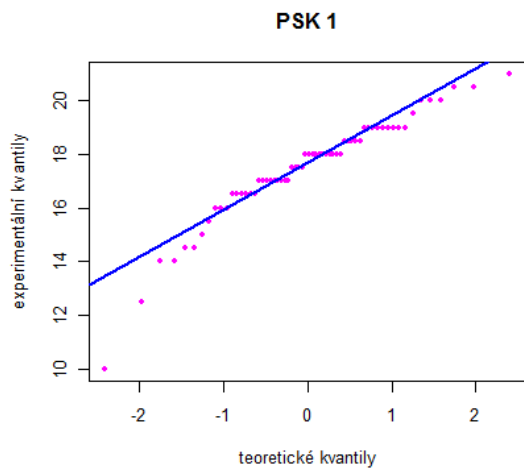
**PSK 22**



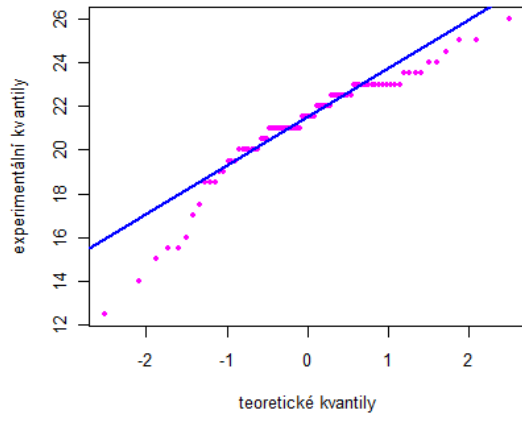
**PSK 23**



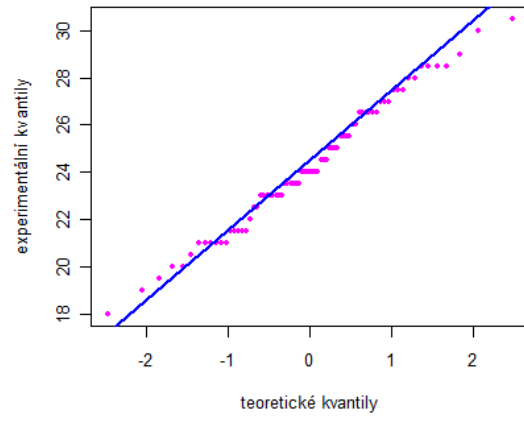
*Příloha č. 5: Kvantil-kvantilové grafy výšek stromů v jednotlivých porostních skupinách*



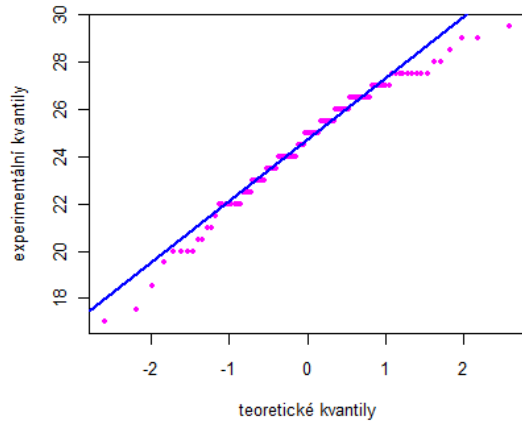
**PSK 5**



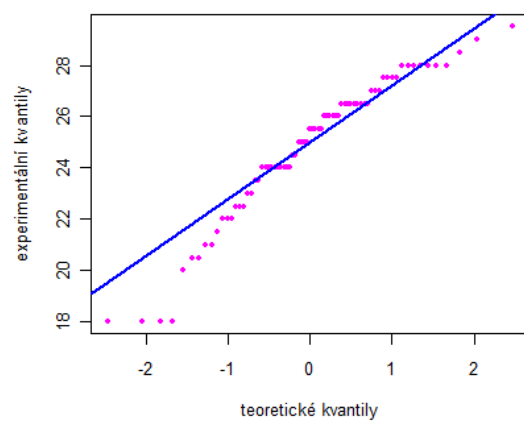
**PSK 6**



**PSK 7**

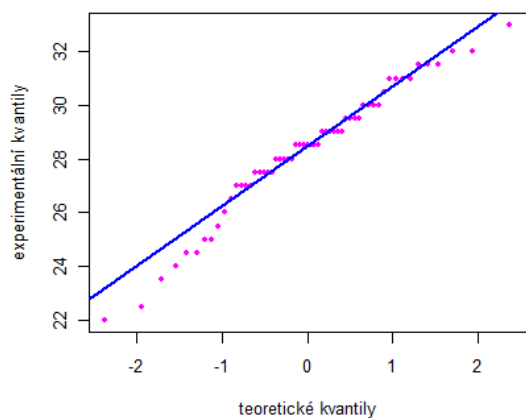


**PSK 8**

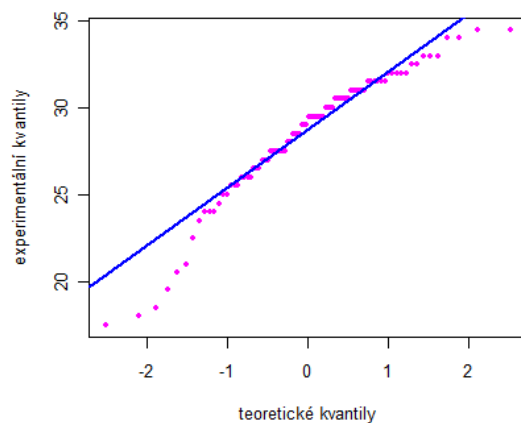




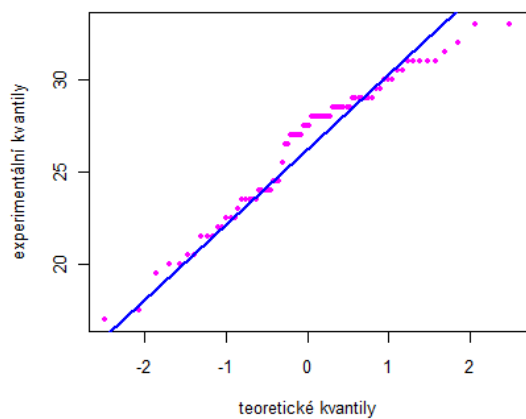
**PSK 9**



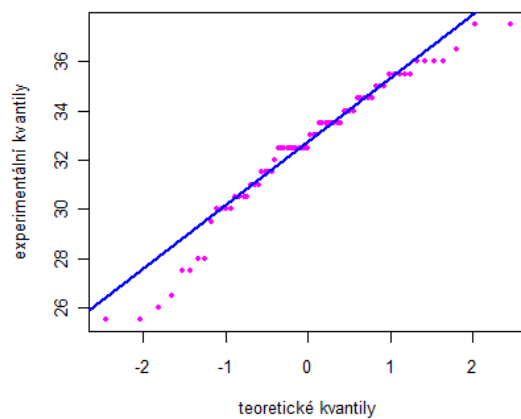
**PSK 10**



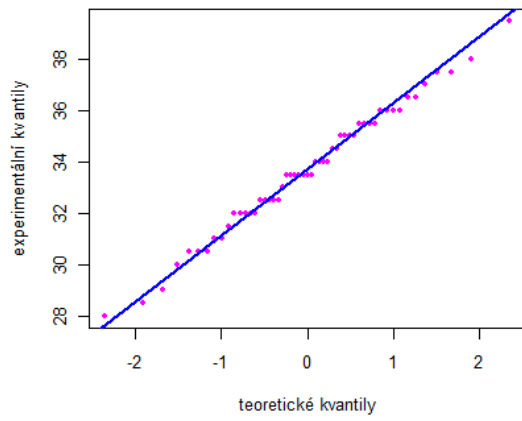
**PSK 11**



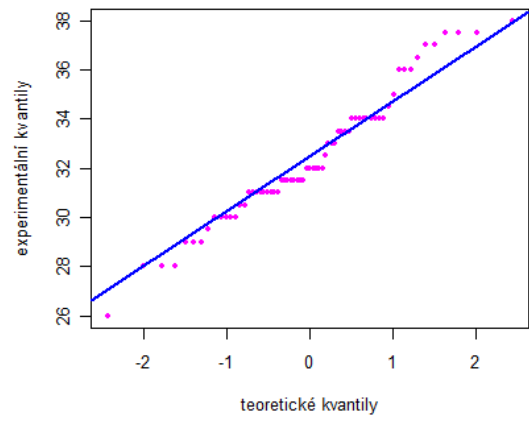
**PSK 12**



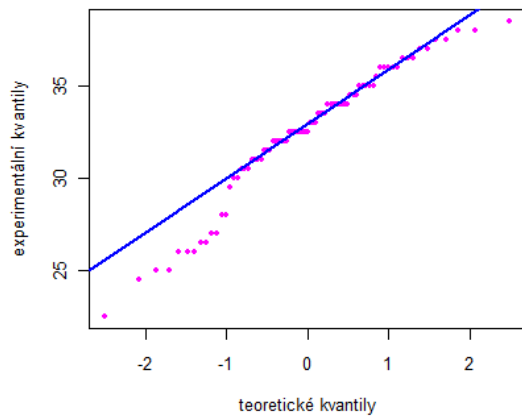
**PSK 13**



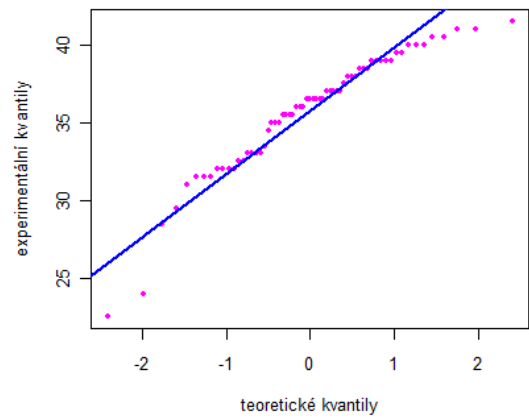
**PSK 14**



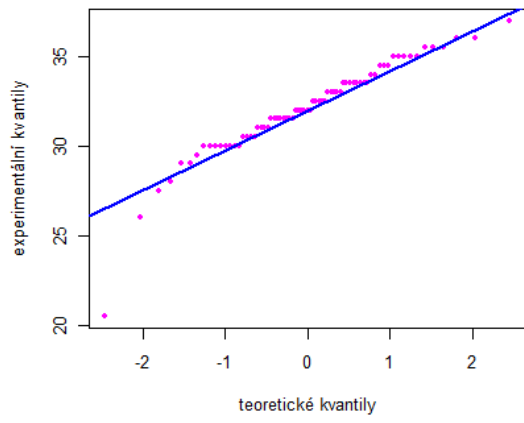
**PSK 15**



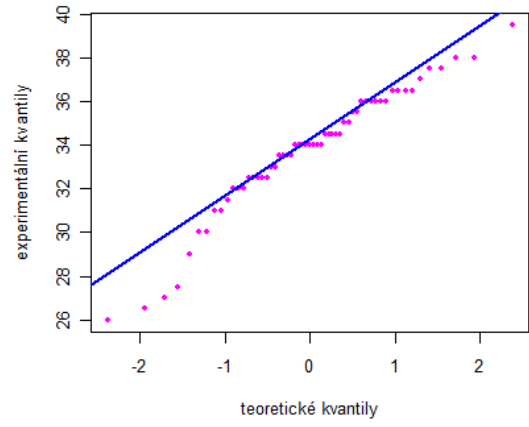
**PSK 16**



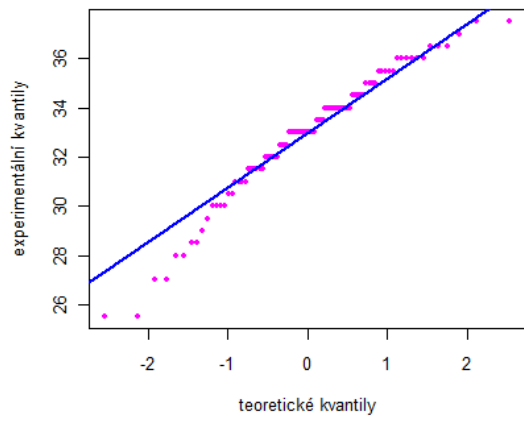
**PSK 17**



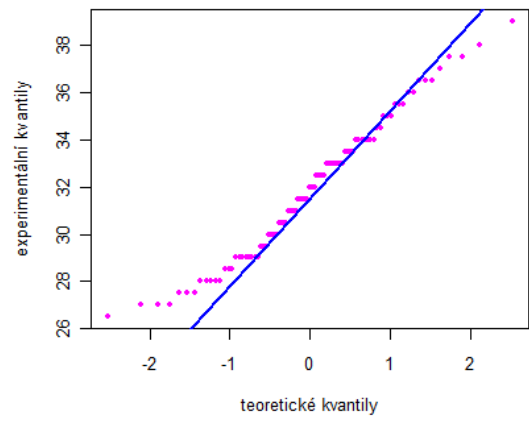
**PSK 18**



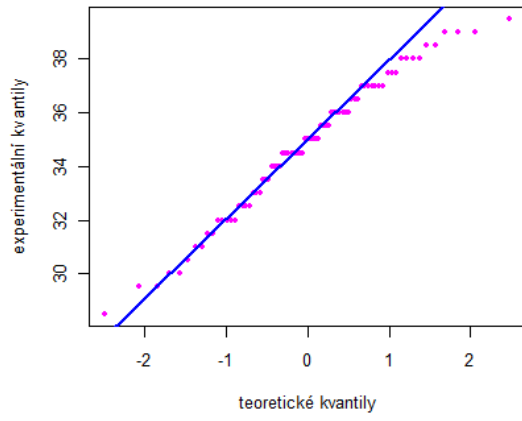
**PSK 19**



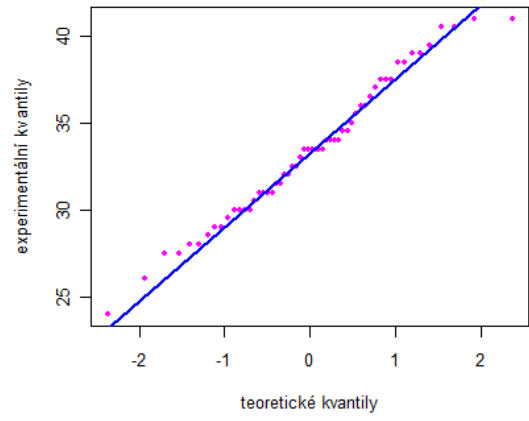
**PSK 20**



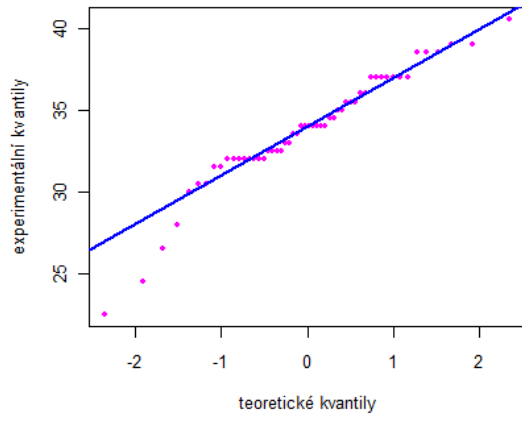
**PSK 21**



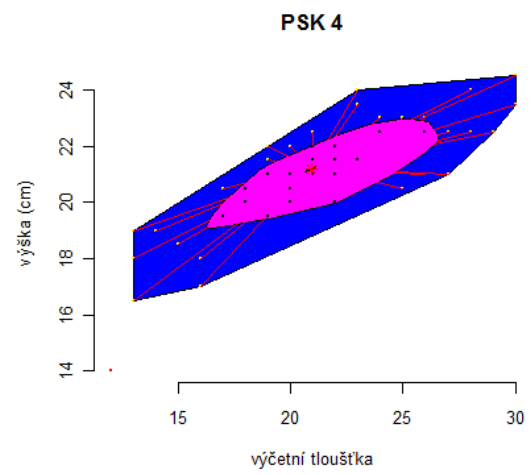
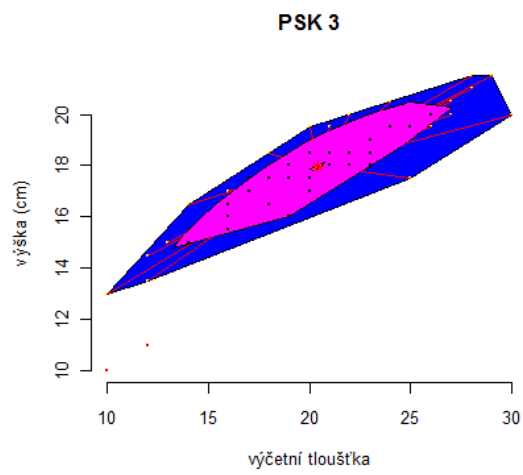
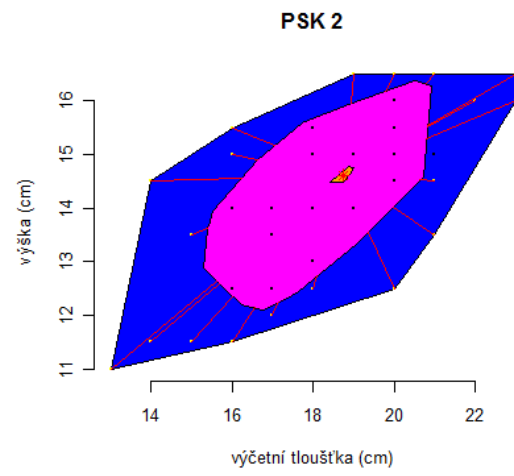
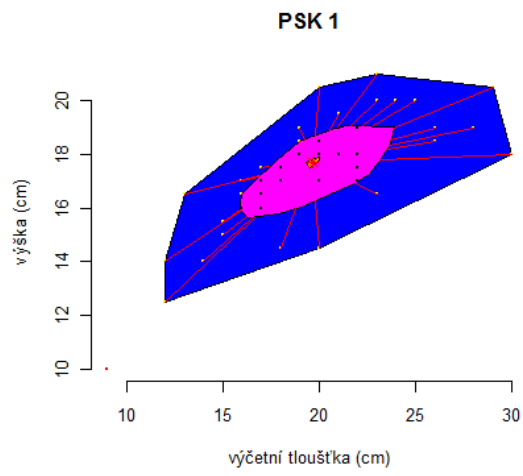
**PSK 22**

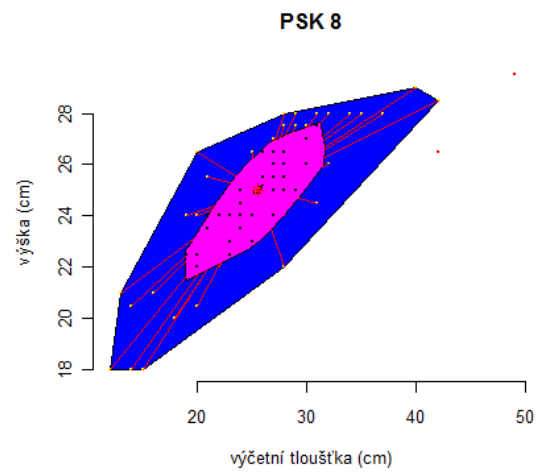
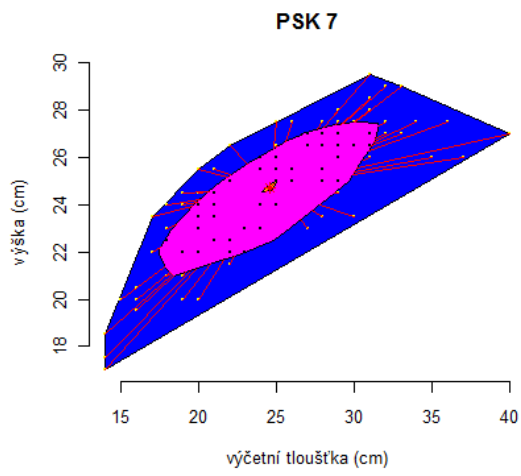
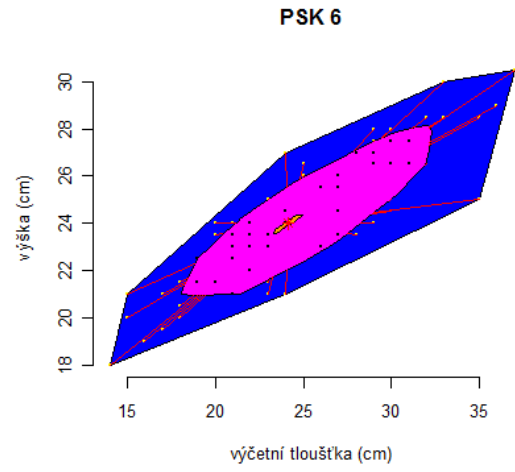
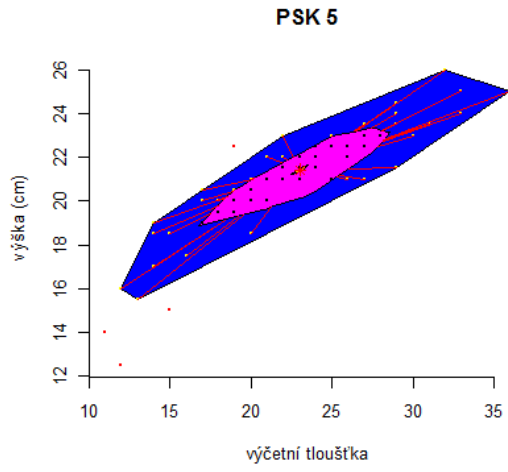


**PSK 23**

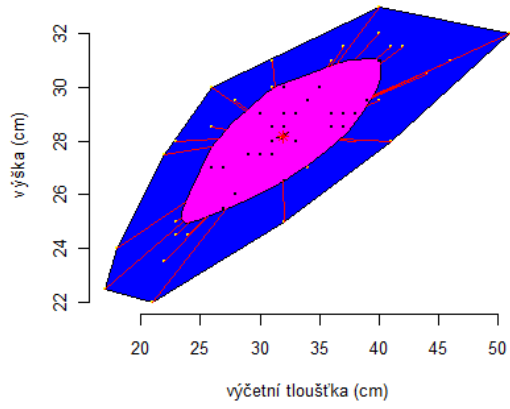


*Příloha č. 6: Bagploty výšek a výčetních tloušťek stromů v jednotlivých porostních skupinách*

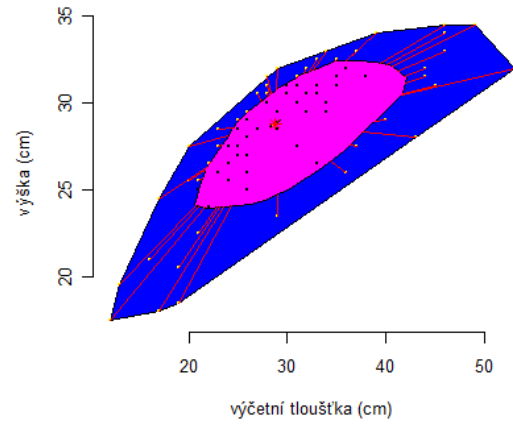




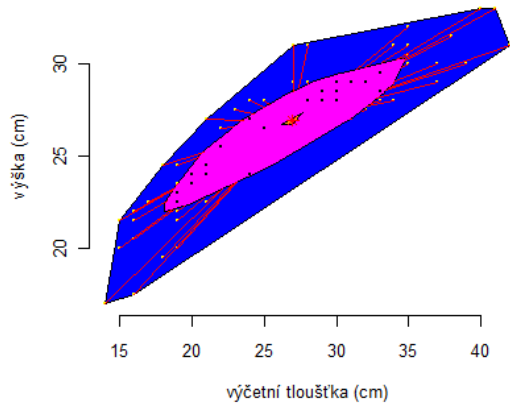
**PSK 9**



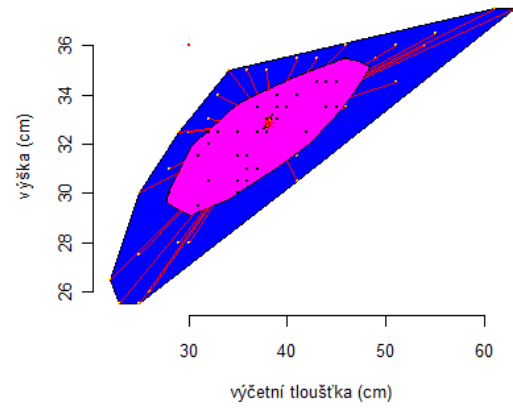
**PSK 10**



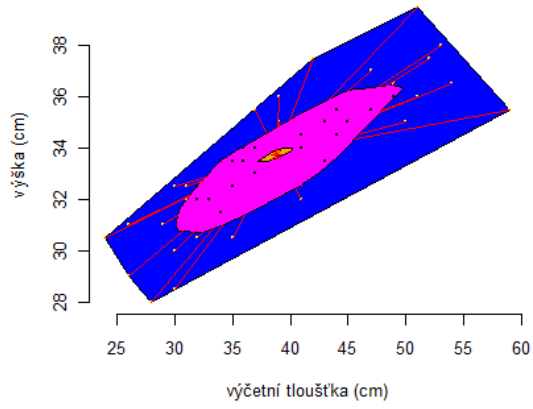
**PSK 11**



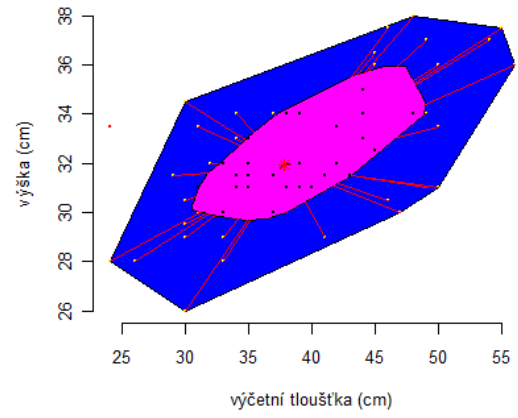
**PSK 12**



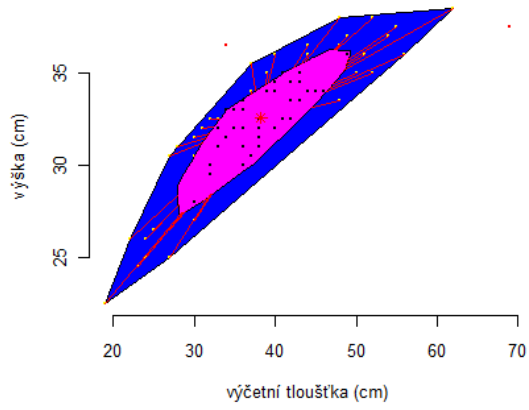
**PSK 13**



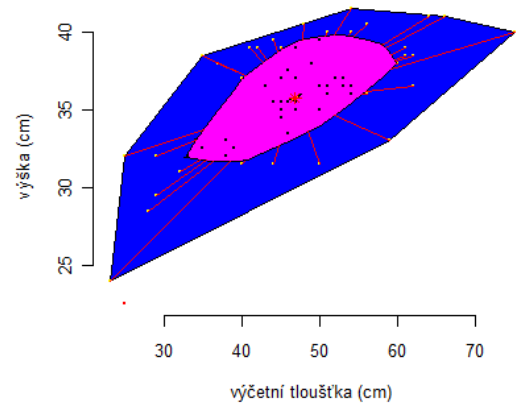
**PSK 14**



**PSK 15**

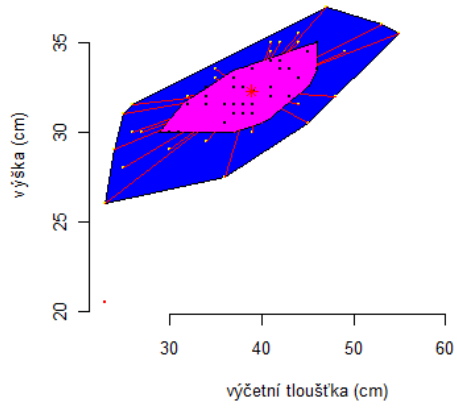


**PSK 16**

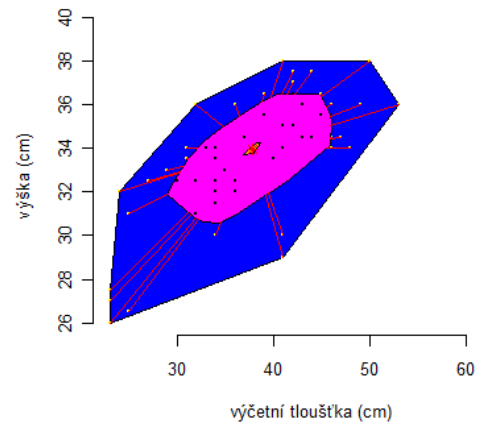




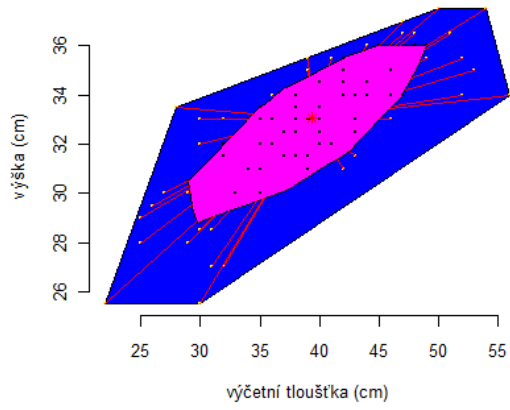
**PSK 17**



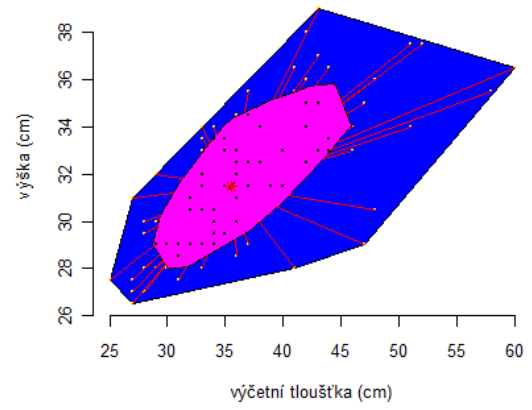
**PSK 18**

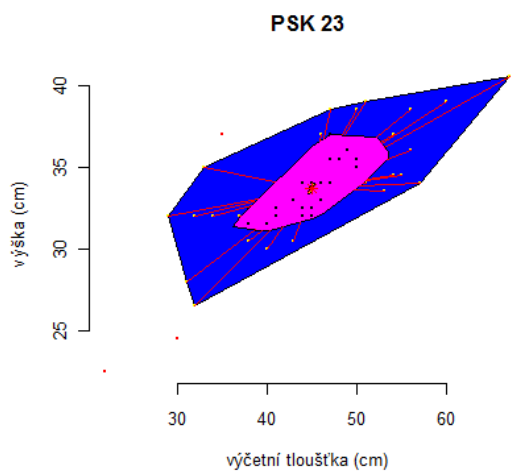
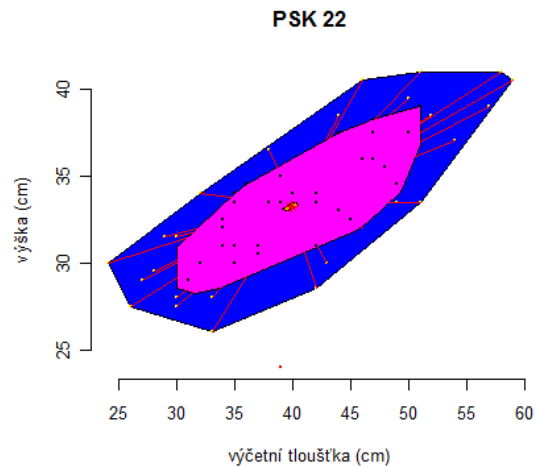
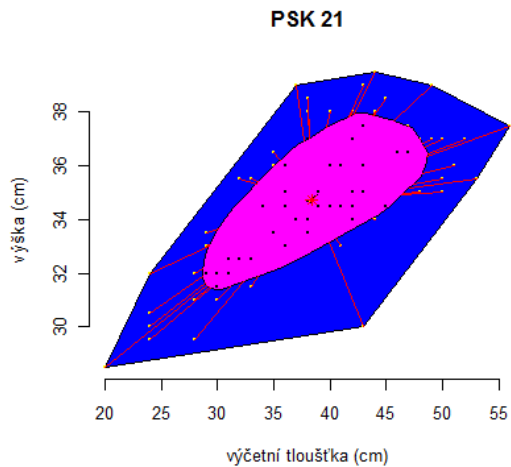


**PSK 19**



**PSK 20**



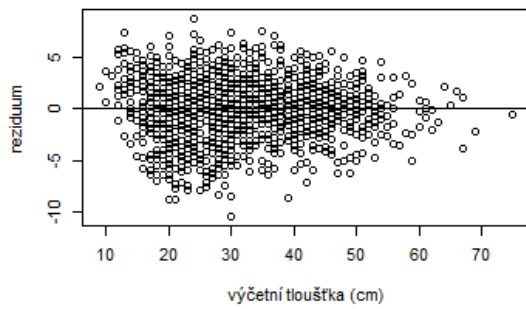


**Legenda:**

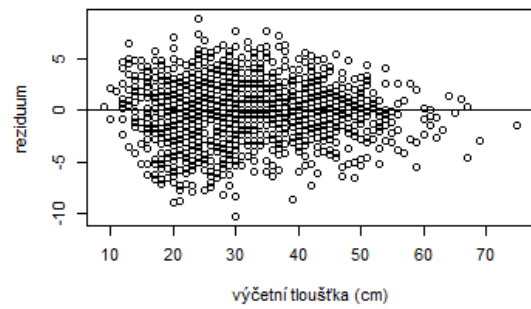
- medián \*
- intervalový odhad mediánu
- interkvartilové rozpětí
- rozsah neodlehých hodnot
- neodlehlé hodnoty •
- odlehlé hodnoty •

*Příloha č. 7: Grafy reziduí globálních modelů zobrazených proti tloušťce a standardizované tloušťce*

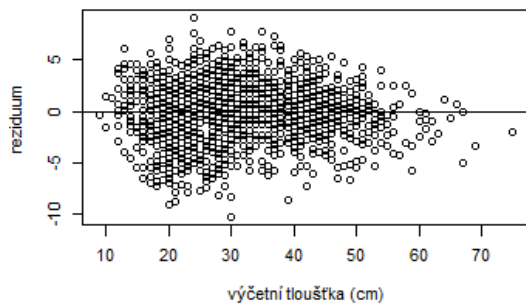
**Michajlovova funkce**



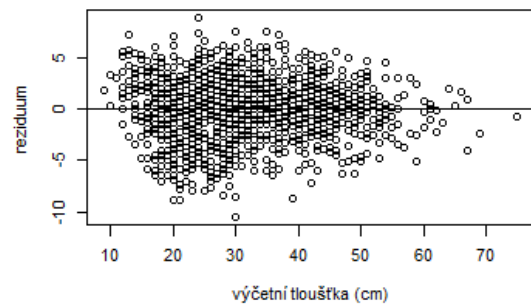
**Pettersonova funkce**



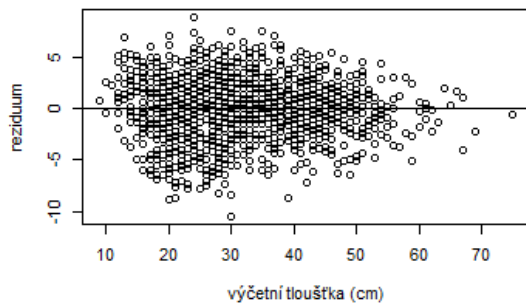
**Näslundova funkce**



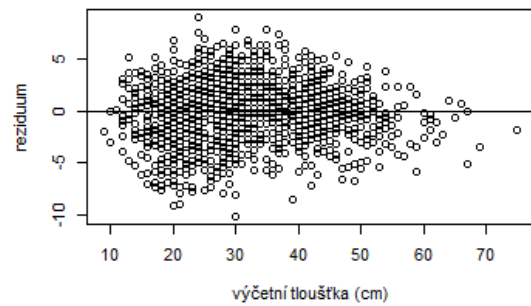
**Levakovičova funkce**



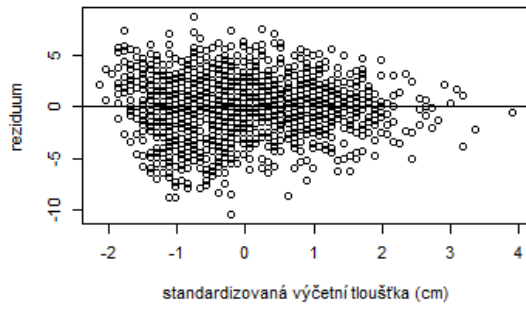
**Prodanova funkce**



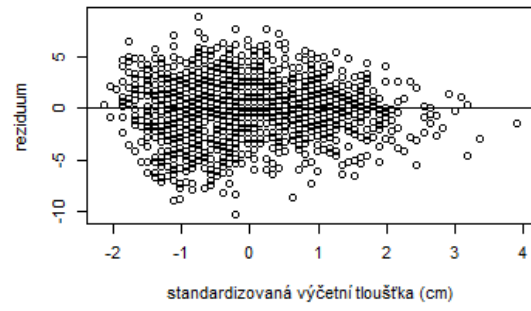
**Meyerova funkce**



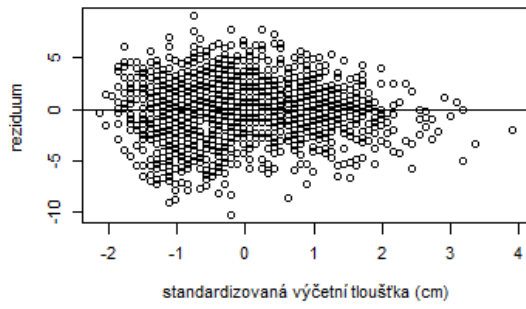
**Michajlova funkce**



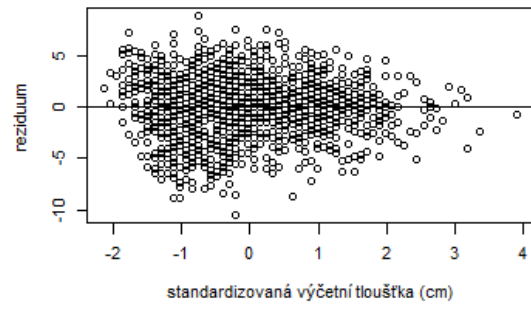
**Pettersonova funkce**



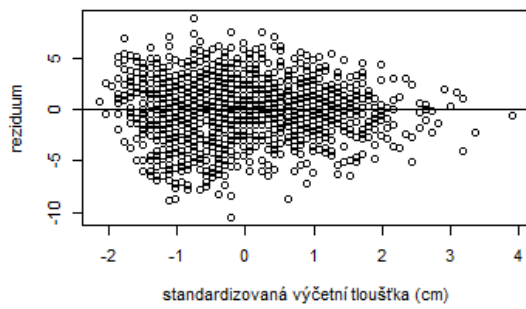
**Näslundova funkce**



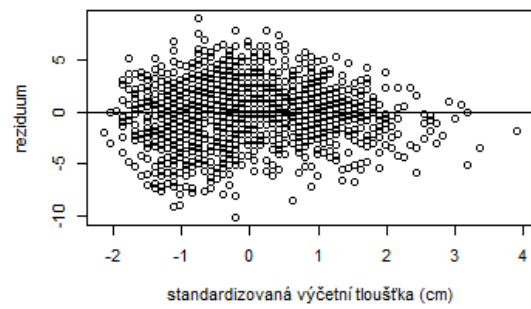
**Levakovičova funkce**



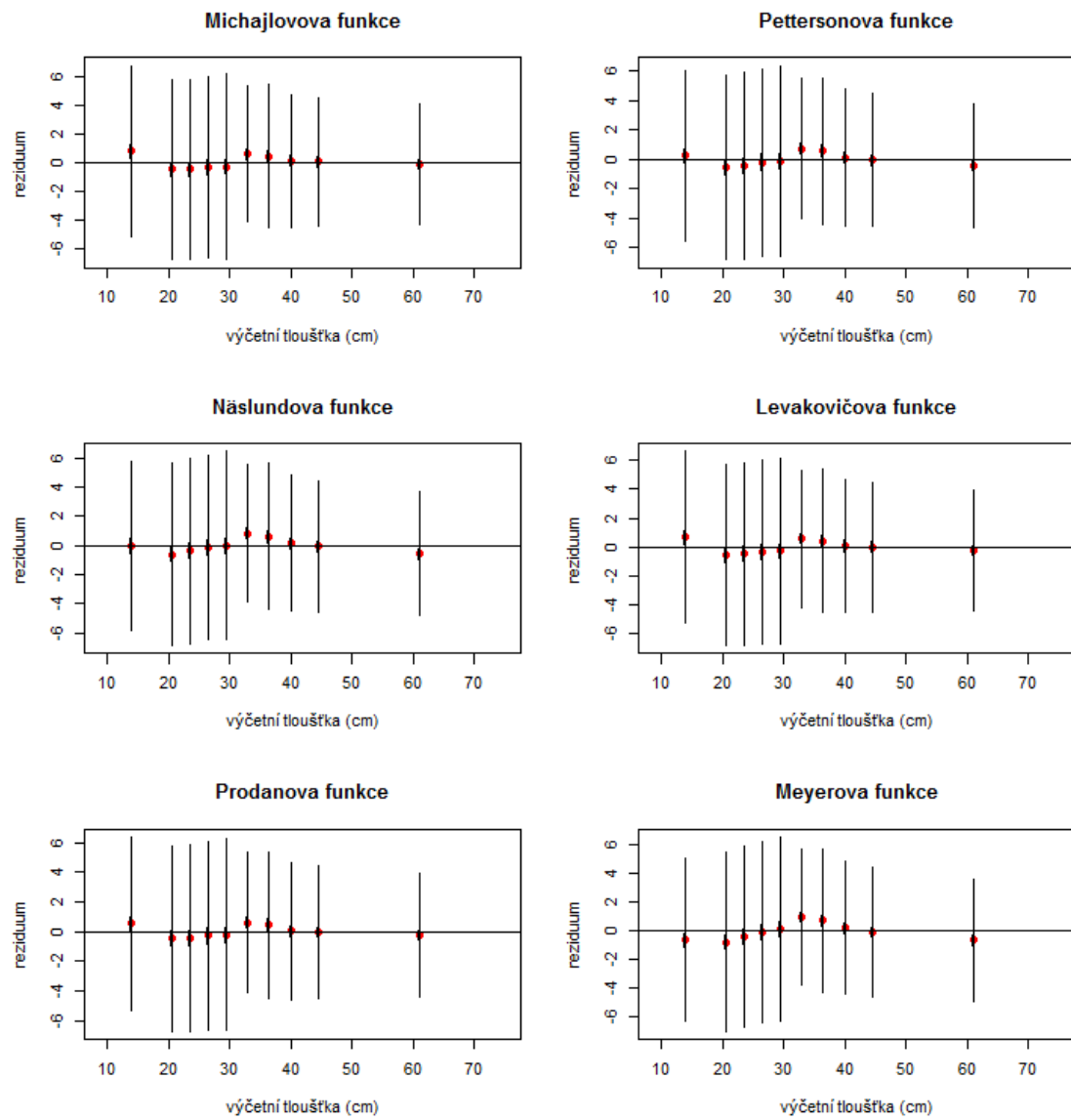
**Prodanova funkce**

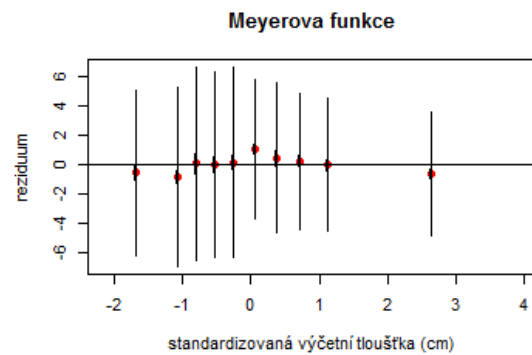
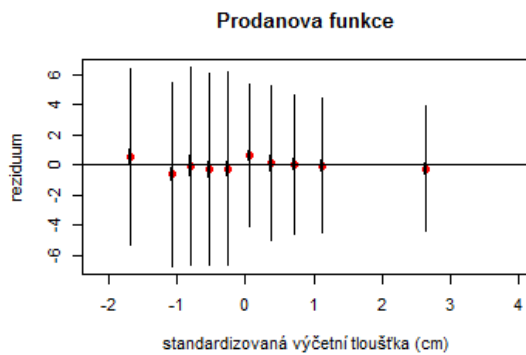
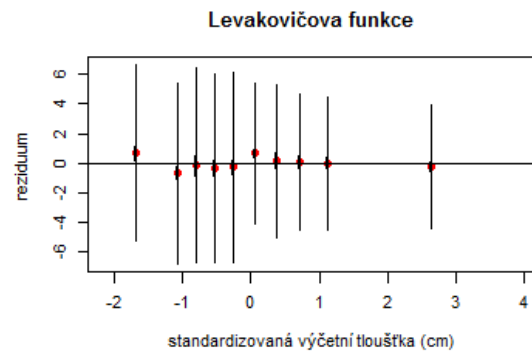
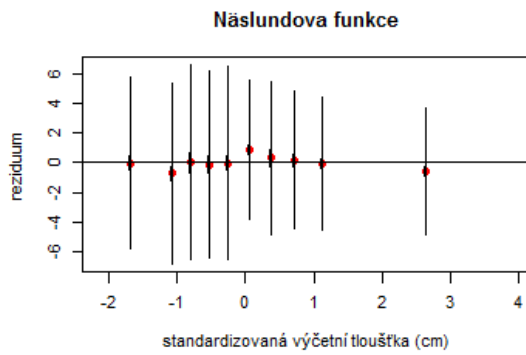
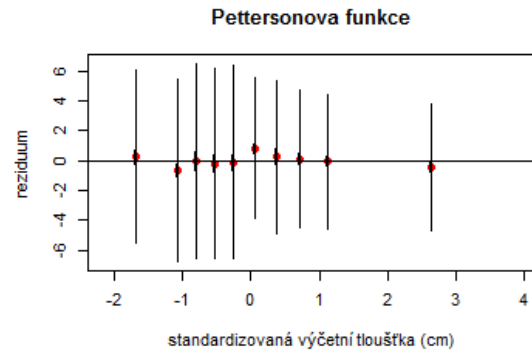
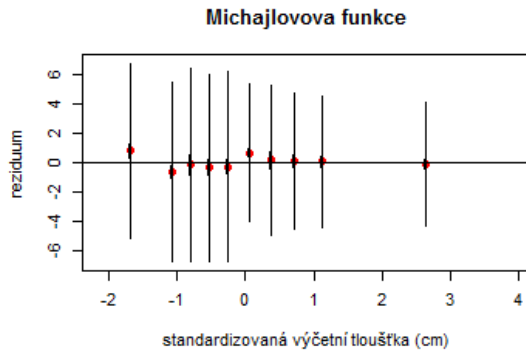


**Meyerova funkce**



Příloha č. 8: Grafy středních hodnot, středních chyb a intervalů spolehlivosti reziduí zobrazených proti tloušťce a standardizované tloušťce pro jednotlivé decily globálních modelů

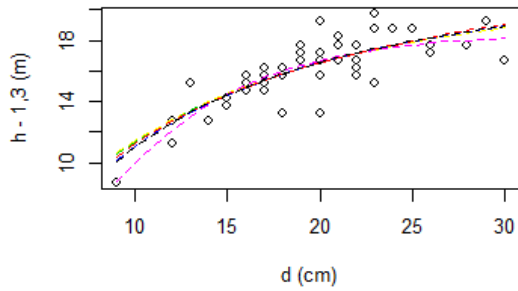




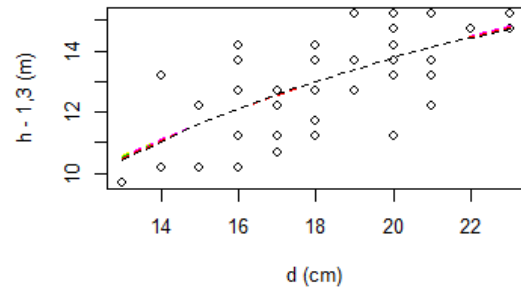
**Legenda:** střední hodnota reziduí •  
 střední chyba reziduí   
 interval spolehlivosti reziduí

*Příloha č. 9: Zobrazení lokálních modelů šesti výškových funkcí v jednotlivých porostních skupinách*

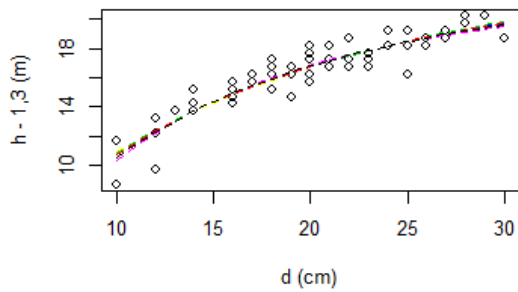
**Plocha 1**



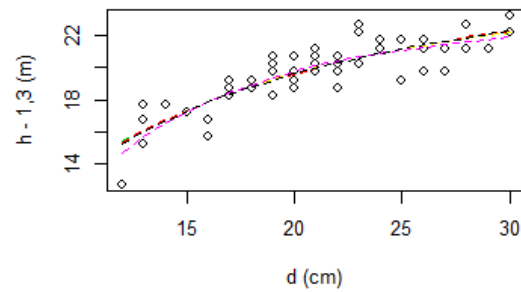
**Plocha 2**



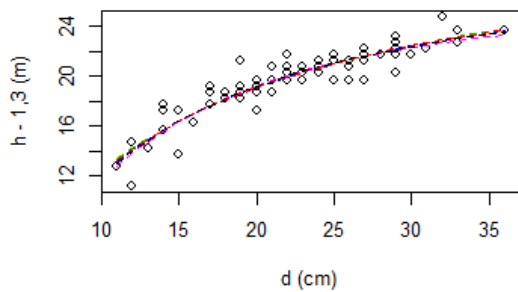
**Plocha 3**



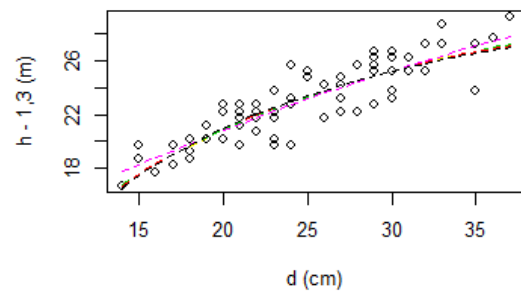
**Plocha 4**



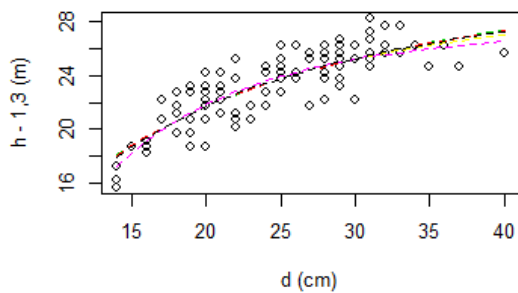
**Plocha 5**



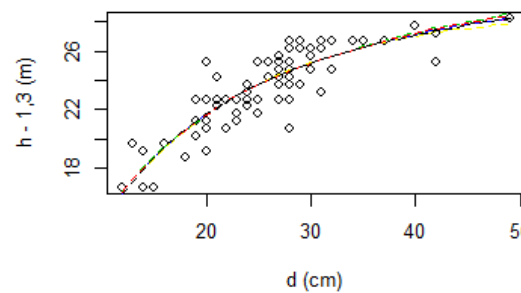
**Plocha 6**



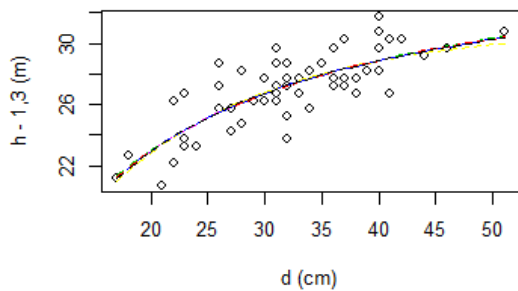
**Plocha 7**



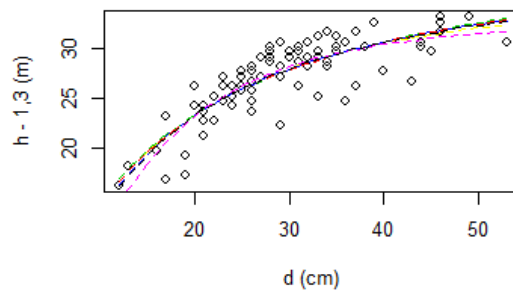
**Plocha 8**



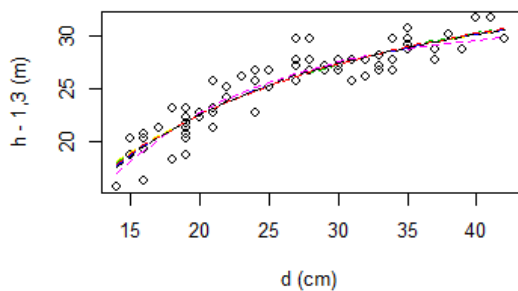
**Plocha 9**



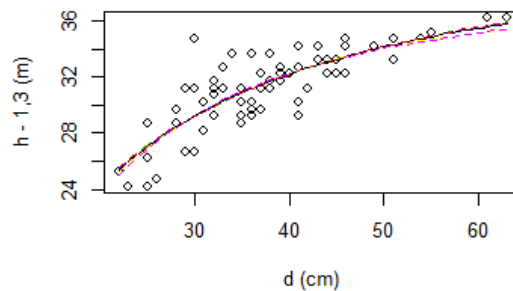
**Plocha 10**



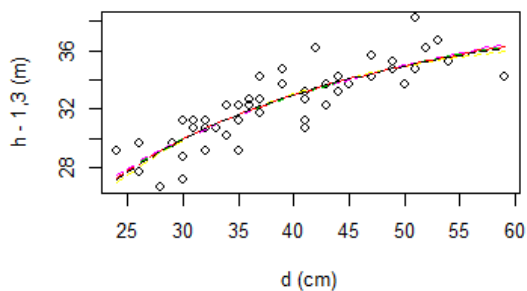
**Plocha 11**



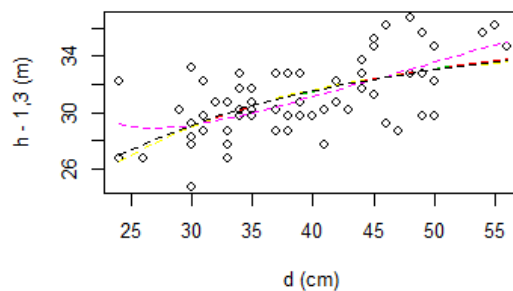
**Plocha 12**



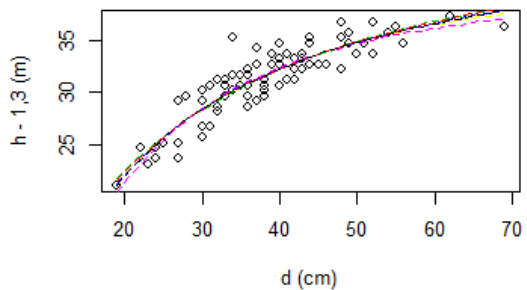
**Plocha 13**



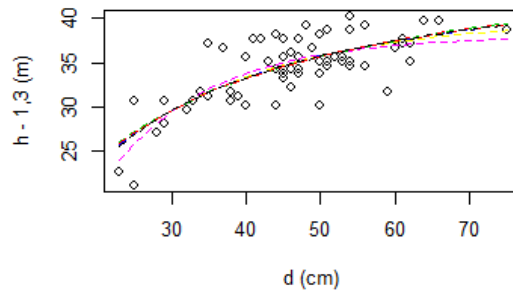
**Plocha 14**



**Plocha 15**

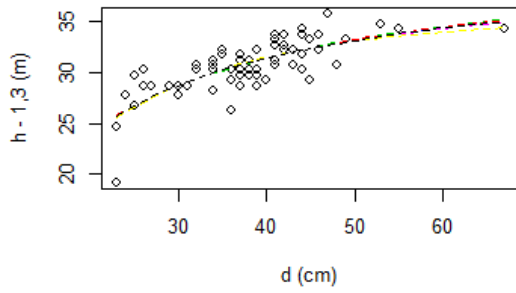


**Plocha 16**

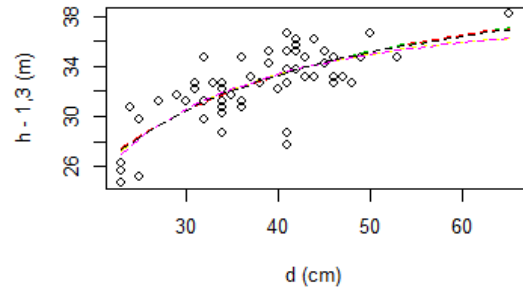




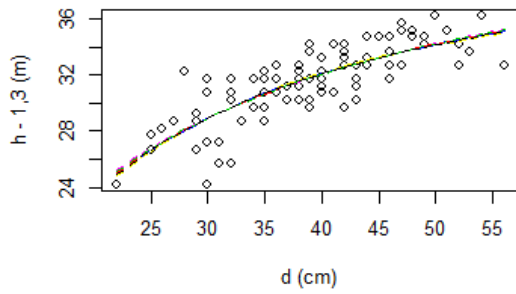
**Plocha 17**



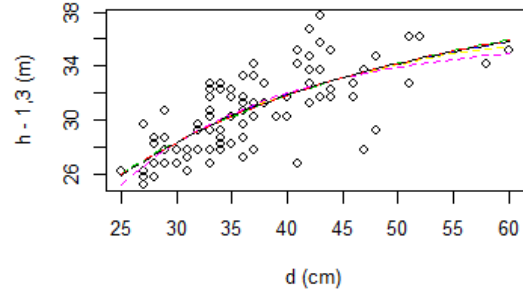
**Plocha 18**



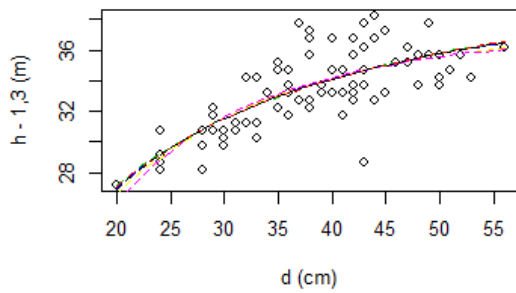
**Plocha 19**



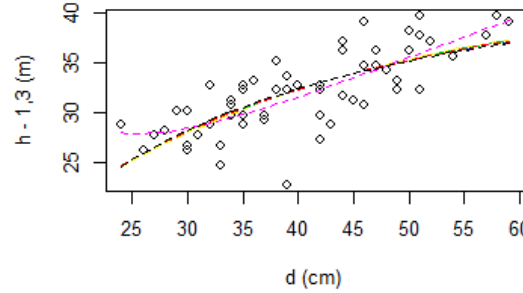
**Plocha 20**



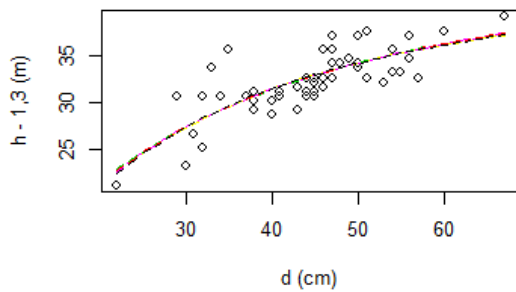
**Plocha 21**



**Plocha 22**



**Plocha 23**



**Legenda:**

Michajlovova funkce - - - - -

Pettersonova funkce - - - - -

Näslundova funkce - - - - -

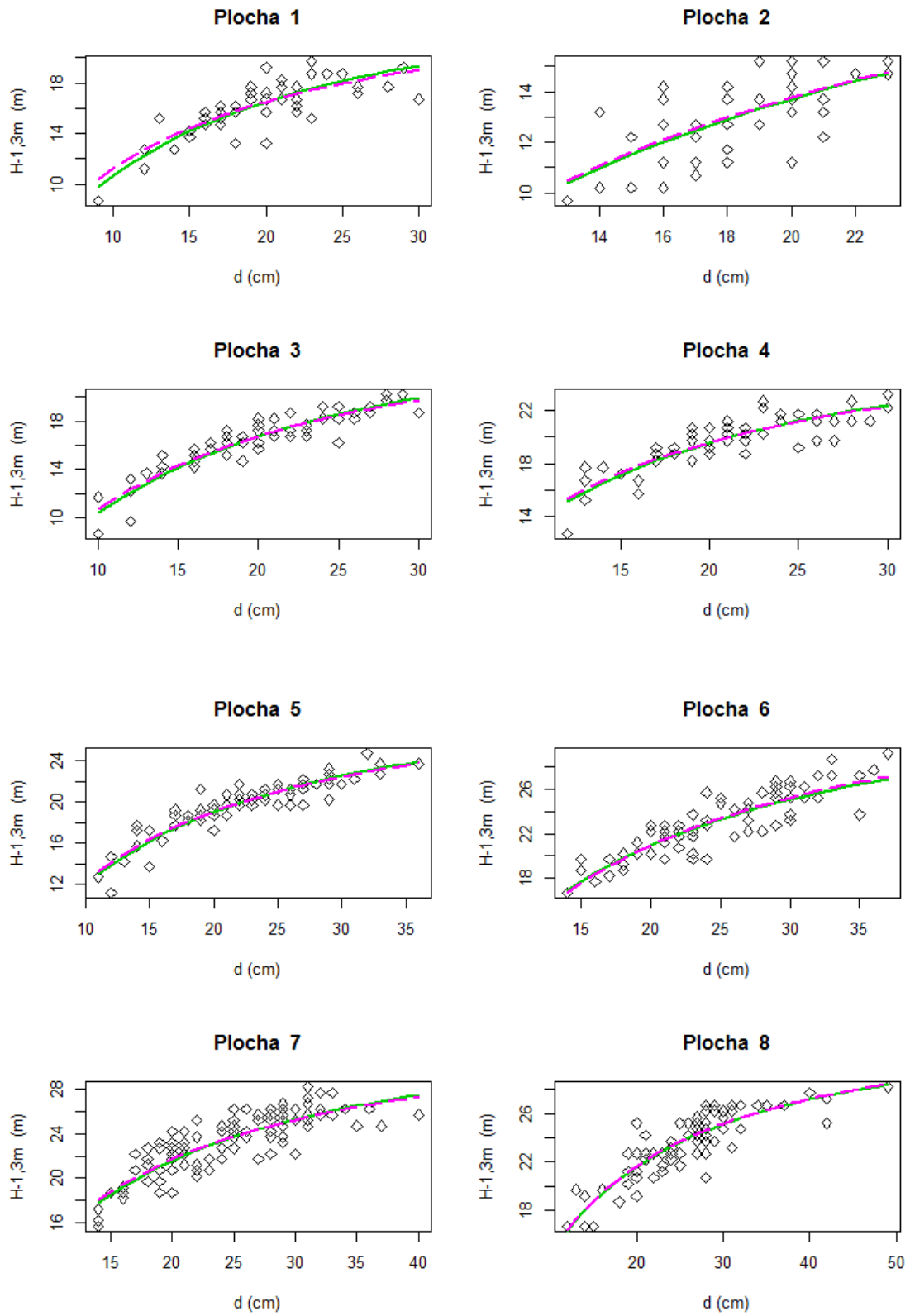
Levakovičova funkce - - - - -

Meyerova funkce - - - - -

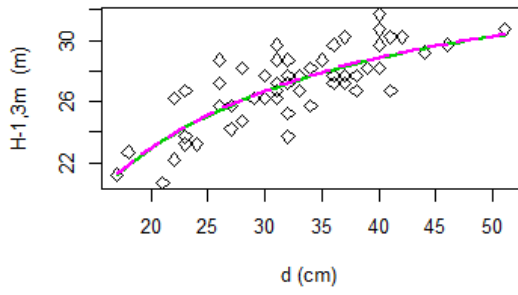
Prodanova funkce - - - - -

měřené hodnoty ◊

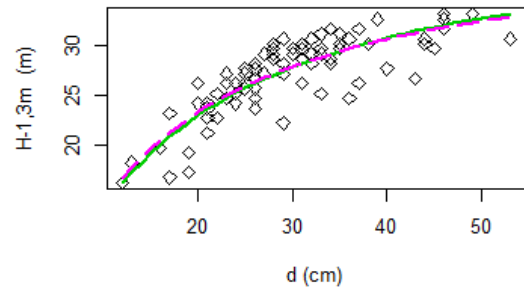
*Příloha č. 10: Srovnání nelineárních a lineárních lokálních modelů Pettersonovy výškové funkce na všech porostních skupinách*



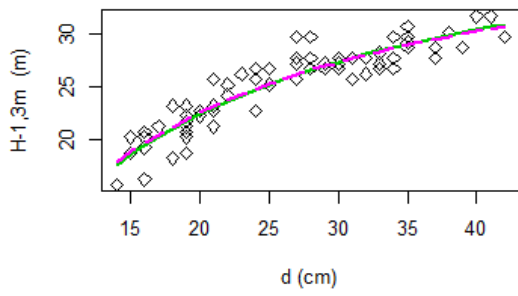
**Plocha 9**



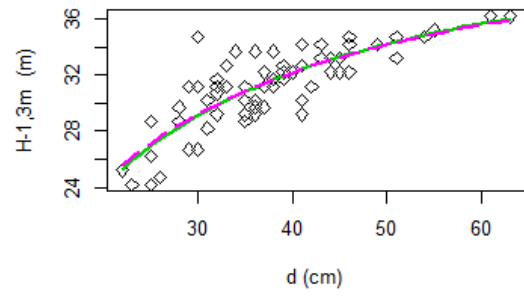
**Plocha 10**



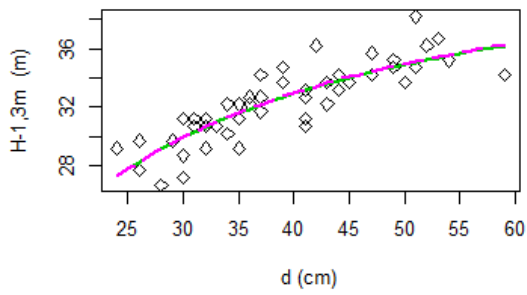
**Plocha 11**



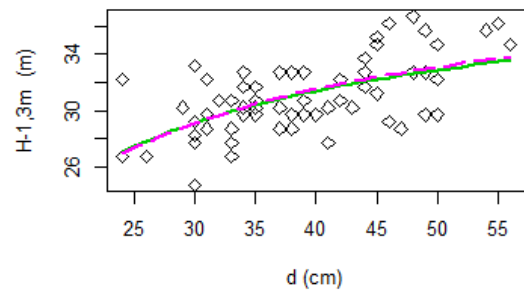
**Plocha 12**



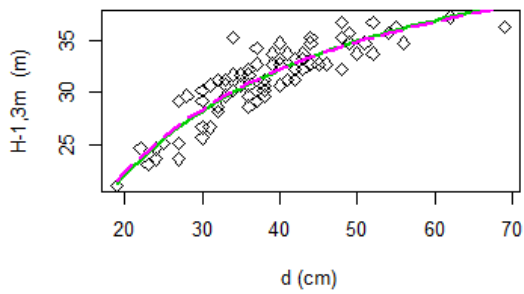
**Plocha 13**



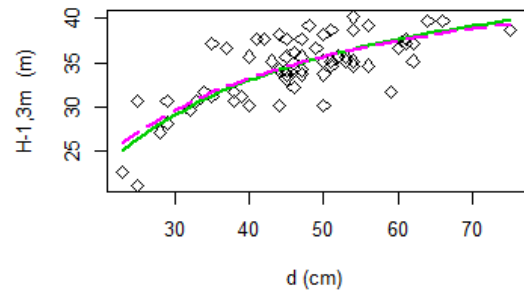
**Plocha 14**



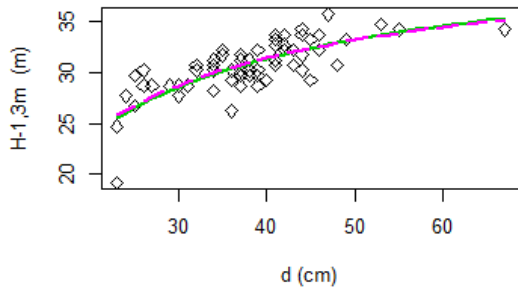
**Plocha 15**



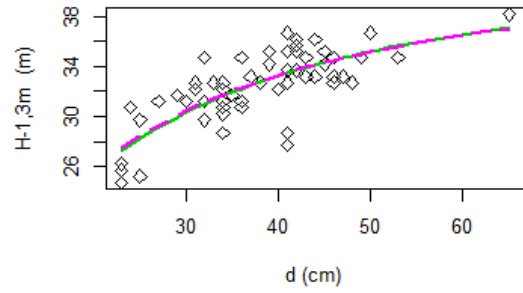
**Plocha 16**



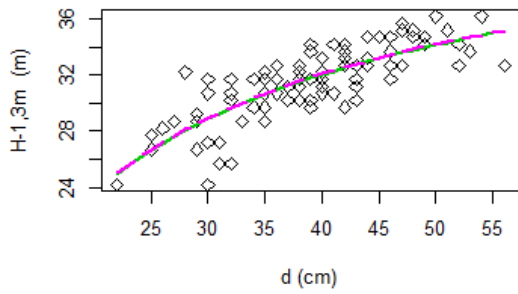
**Plocha 17**



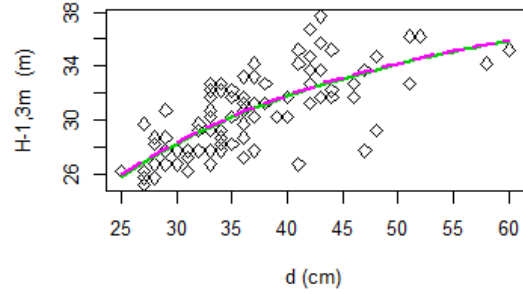
**Plocha 18**



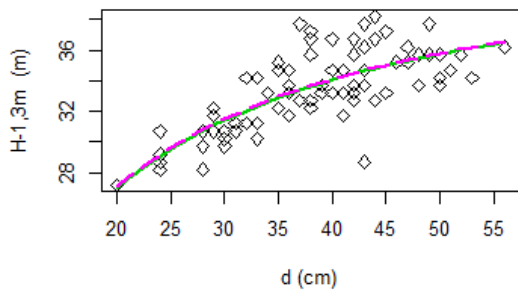
**Plocha 19**



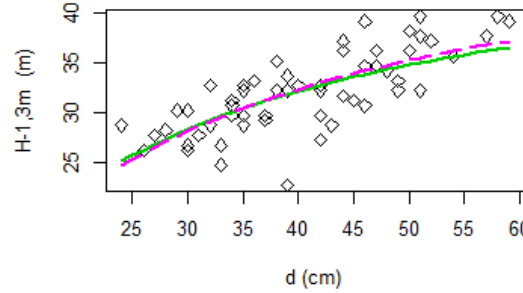
**Plocha 20**



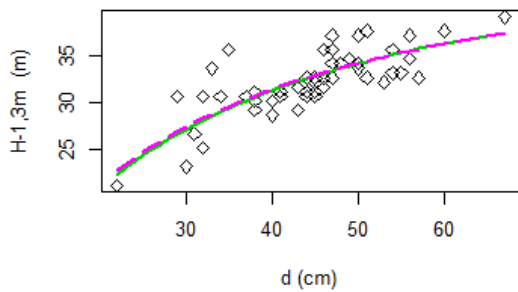
**Plocha 21**



**Plocha 22**



**Plocha 23**



**Legenda:**

měřené hodnoty



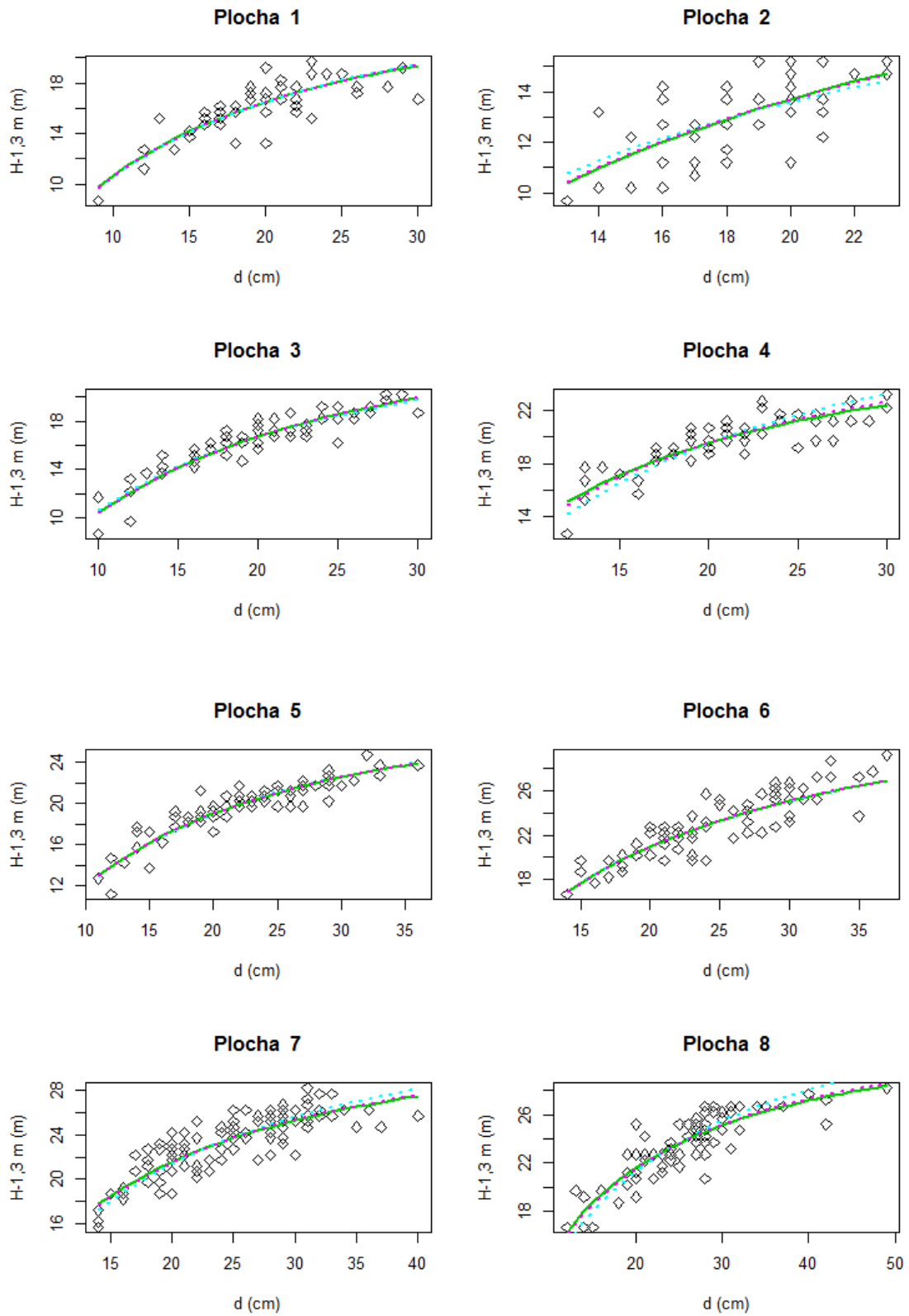
lineární model



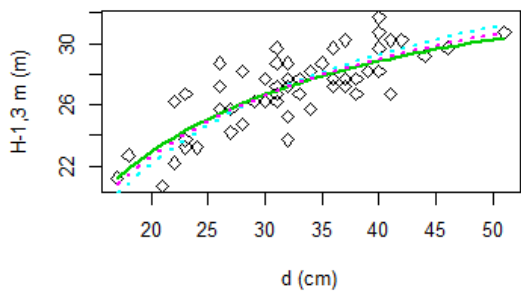
nelineární model



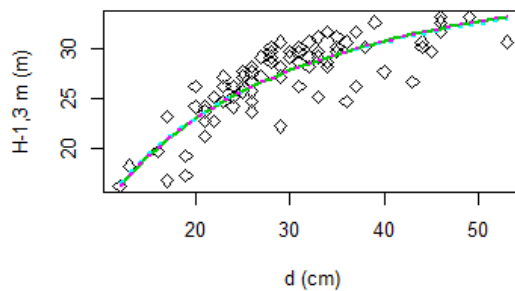
*Příloha č. 11: Srovnání lokálního modelu, modelu s náhodným absolutním členem a modelu s náhodným absolutním i regresním členem ve všech porostních skupinách*



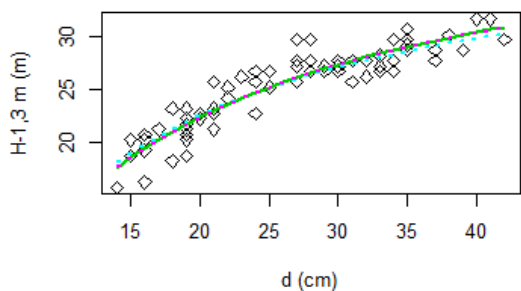
**Plocha 9**



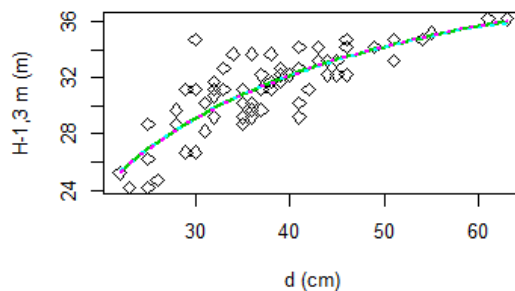
**Plocha 10**



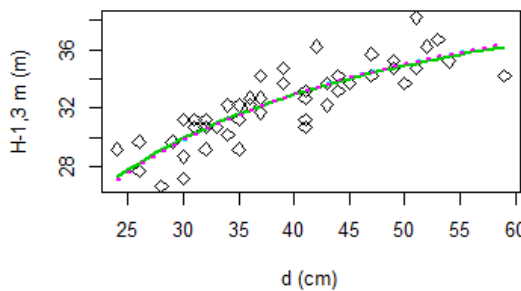
**Plocha 11**



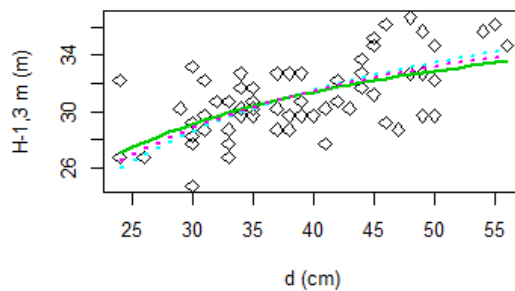
**Plocha 12**



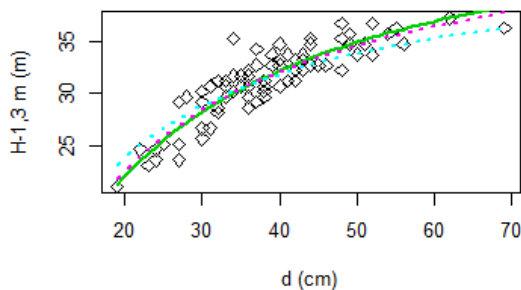
**Plocha 13**



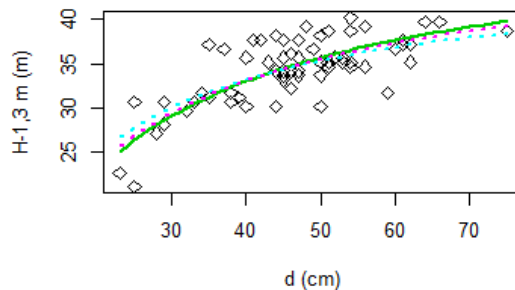
**Plocha 14**



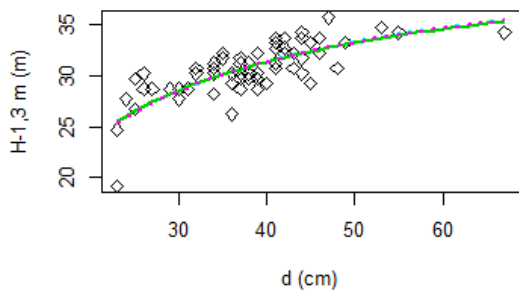
**Plocha 15**



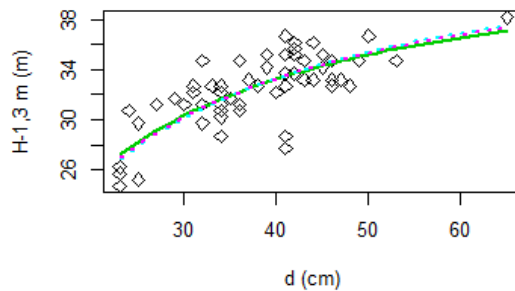
**Plocha 16**



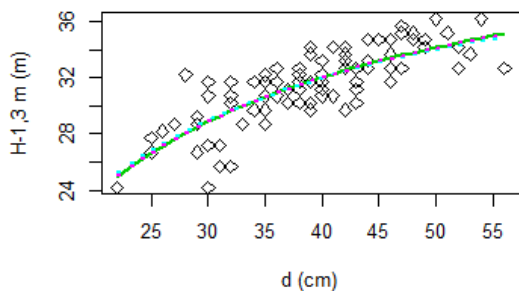
**Plocha 17**



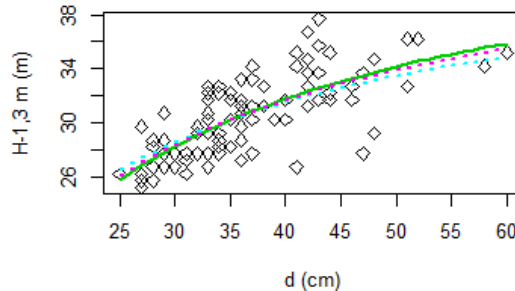
**Plocha 18**



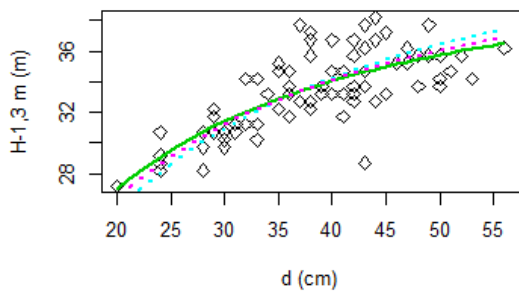
**Plocha 19**



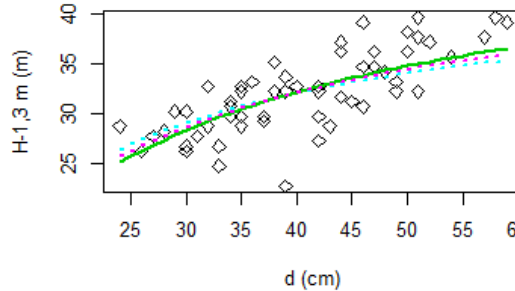
**Plocha 20**



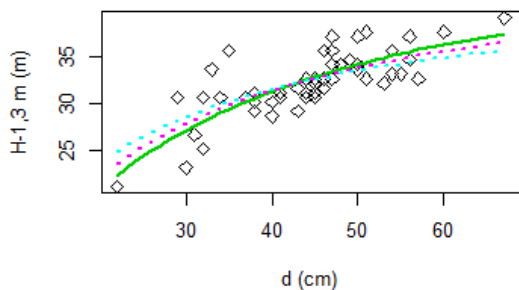
**Plocha 21**



**Plocha 22**



**Plocha 23**

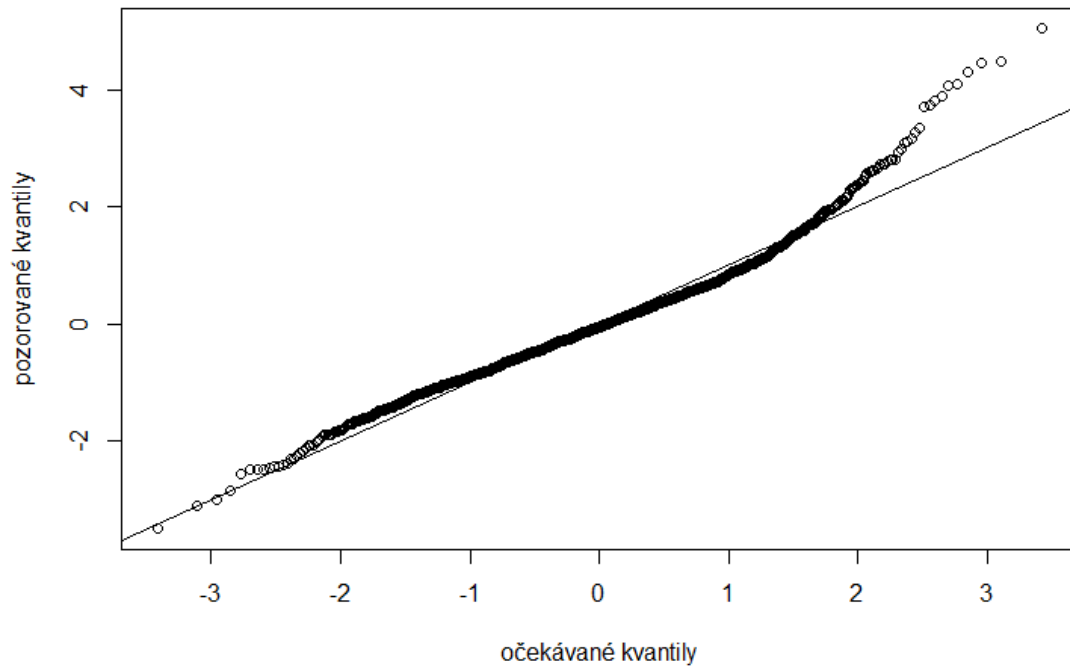


**Legenda:**

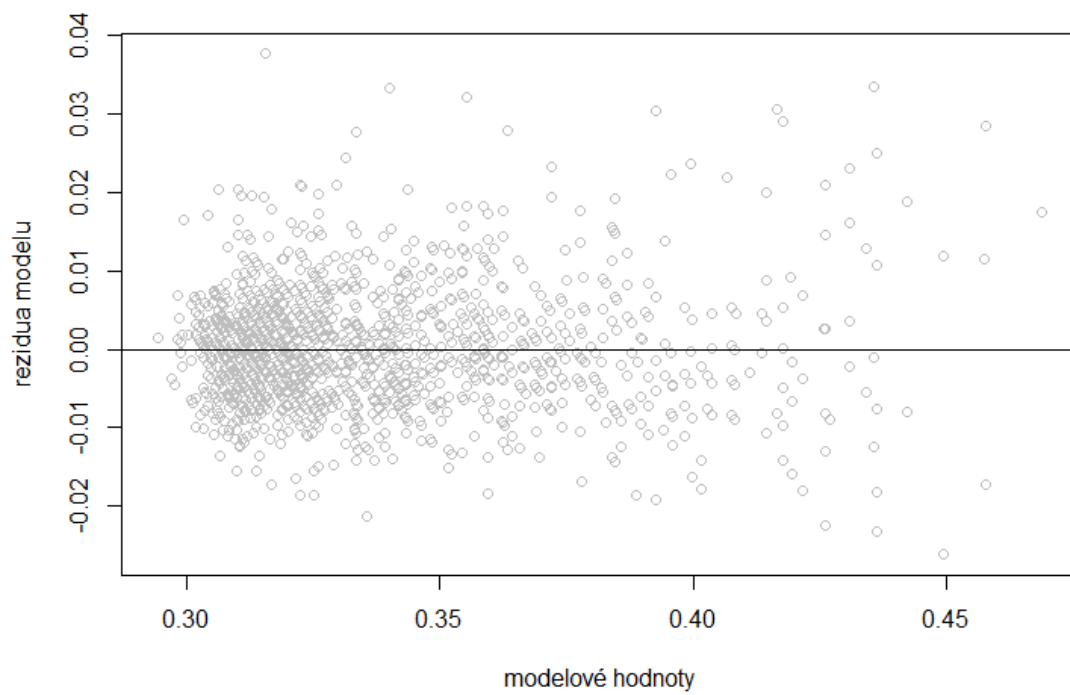
- měřené hodnoty
- lokální model
- model s náhodným absolutním členem
- model s náhodným absolutním i regresním členem

Příloha č. 12: Grafy reziduí smíšeného modelu s náhodným absolutním i regresním členem

normální pravděpodobnostní Q-Q graf reziduí

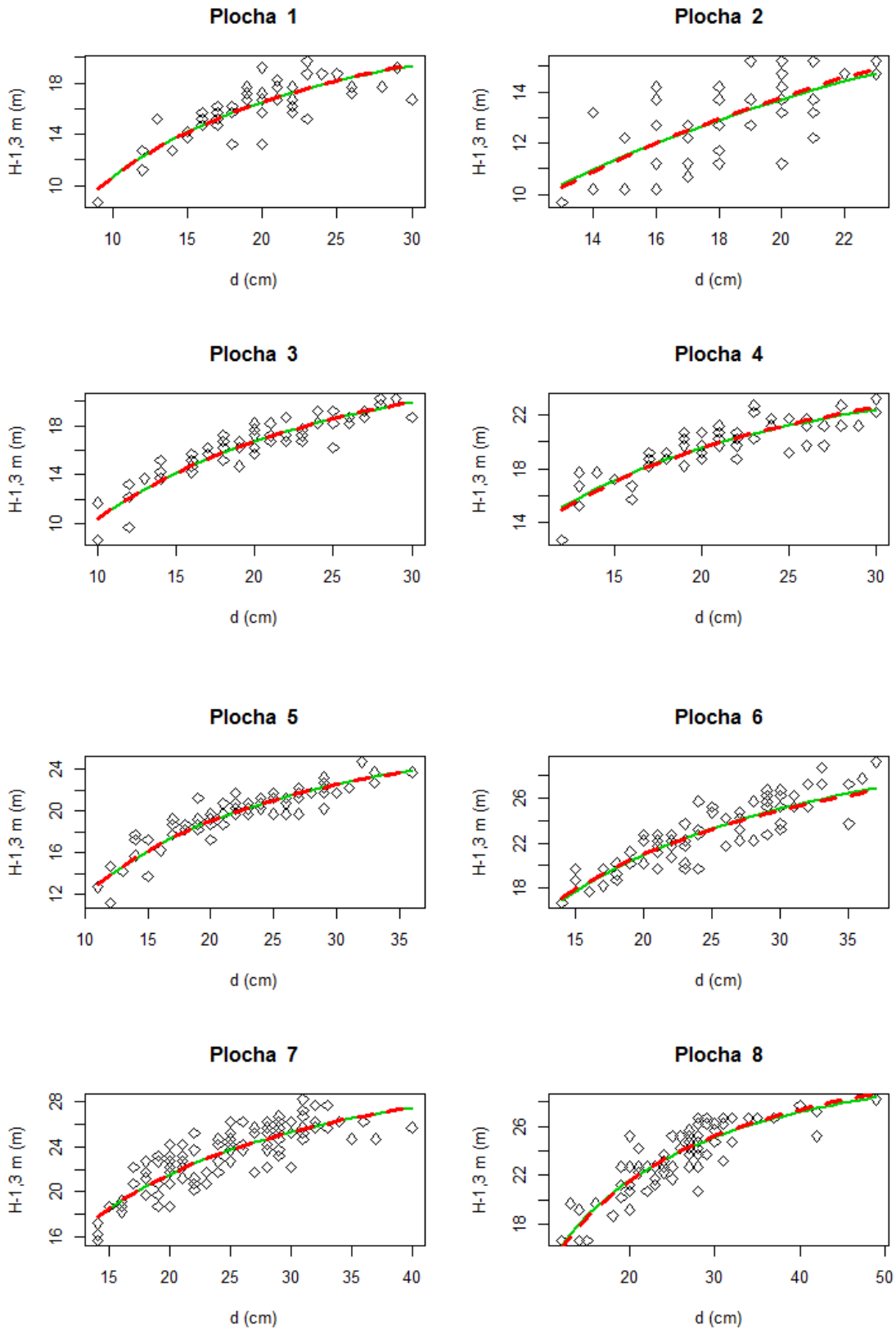


graf závislosti reziduí na modelových hodnotách

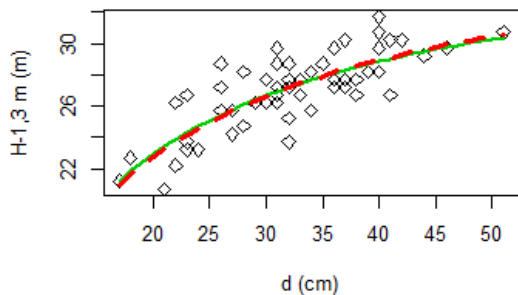




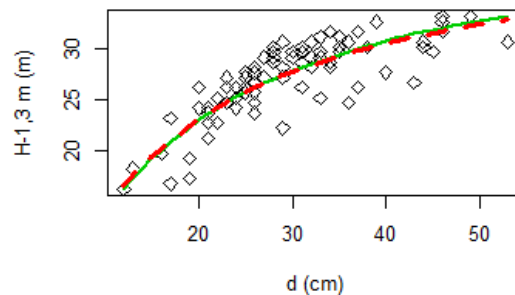
Příloha č. 13: Srovnání lokálního modelu a smíšeného modelu druhé úrovně, kde je zahrnut faktor věku ve všech porostních skupinách



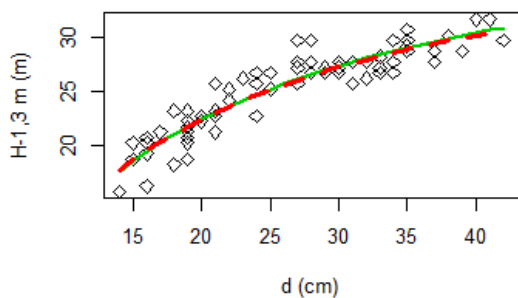
**Plocha 9**



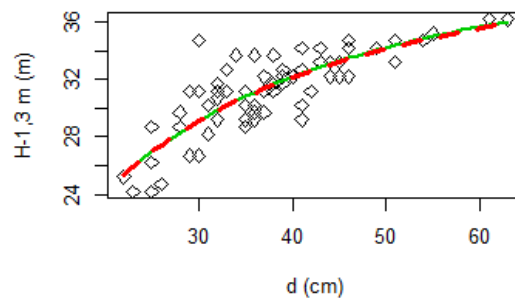
**Plocha 10**



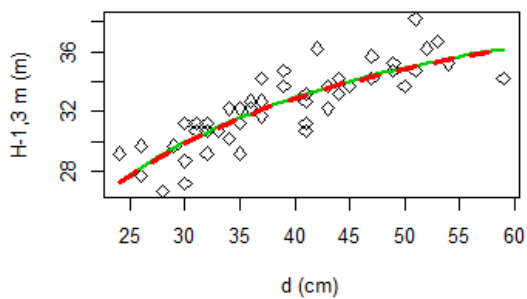
**Plocha 11**



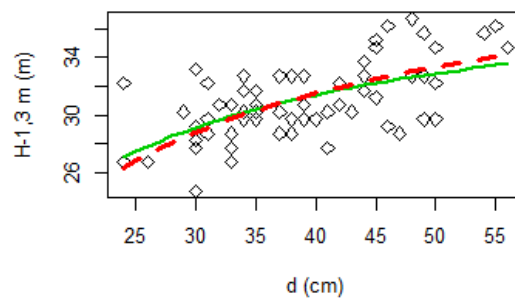
**Plocha 12**



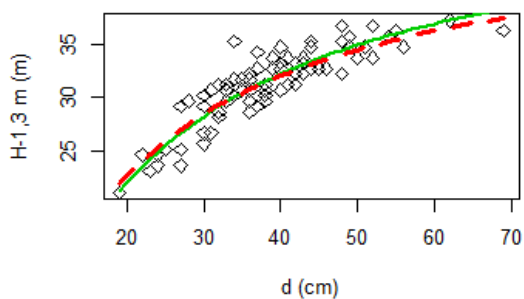
**Plocha 13**



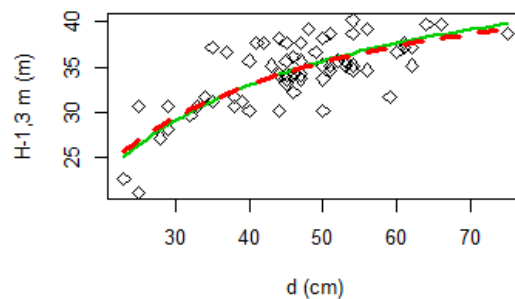
**Plocha 14**



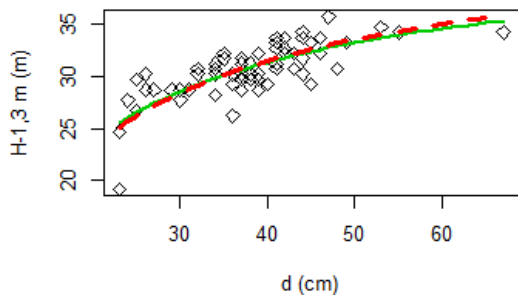
**Plocha 15**



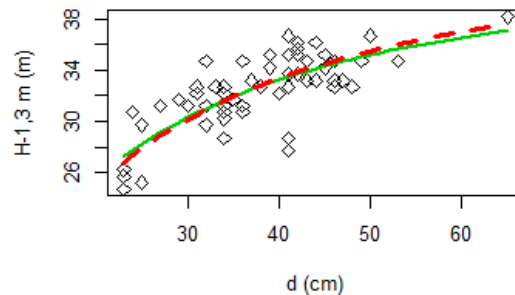
**Plocha 16**



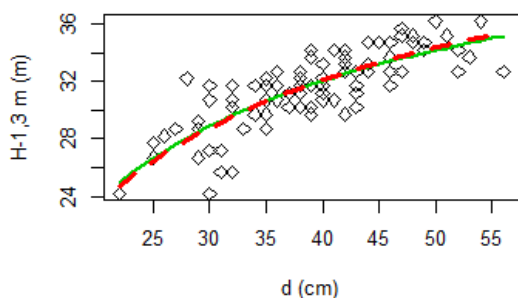
**Plocha 17**



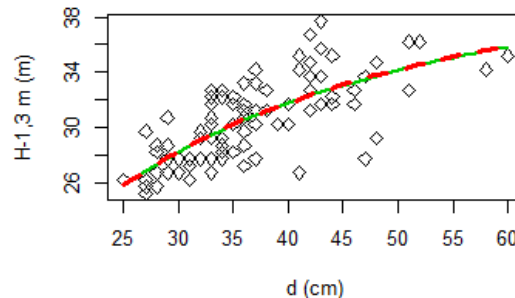
**Plocha 18**



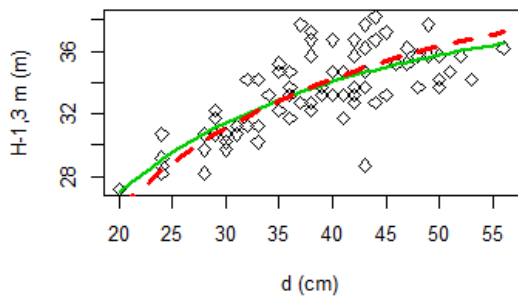
**Plocha 19**



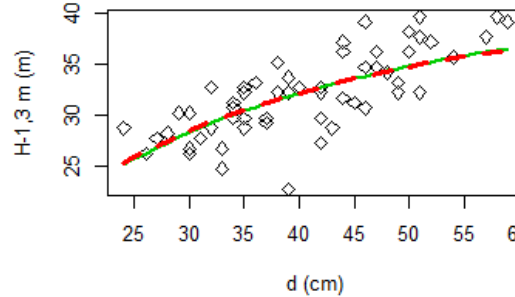
**Plocha 20**



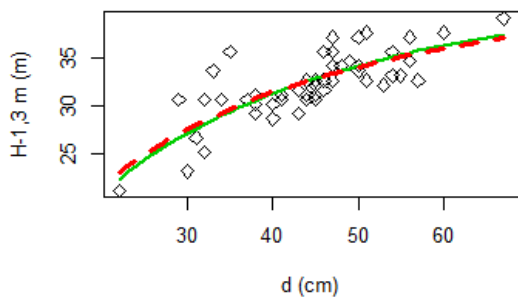
**Plocha 21**



**Plocha 22**



**Plocha 23**



**Legenda:**

měřené hodnoty



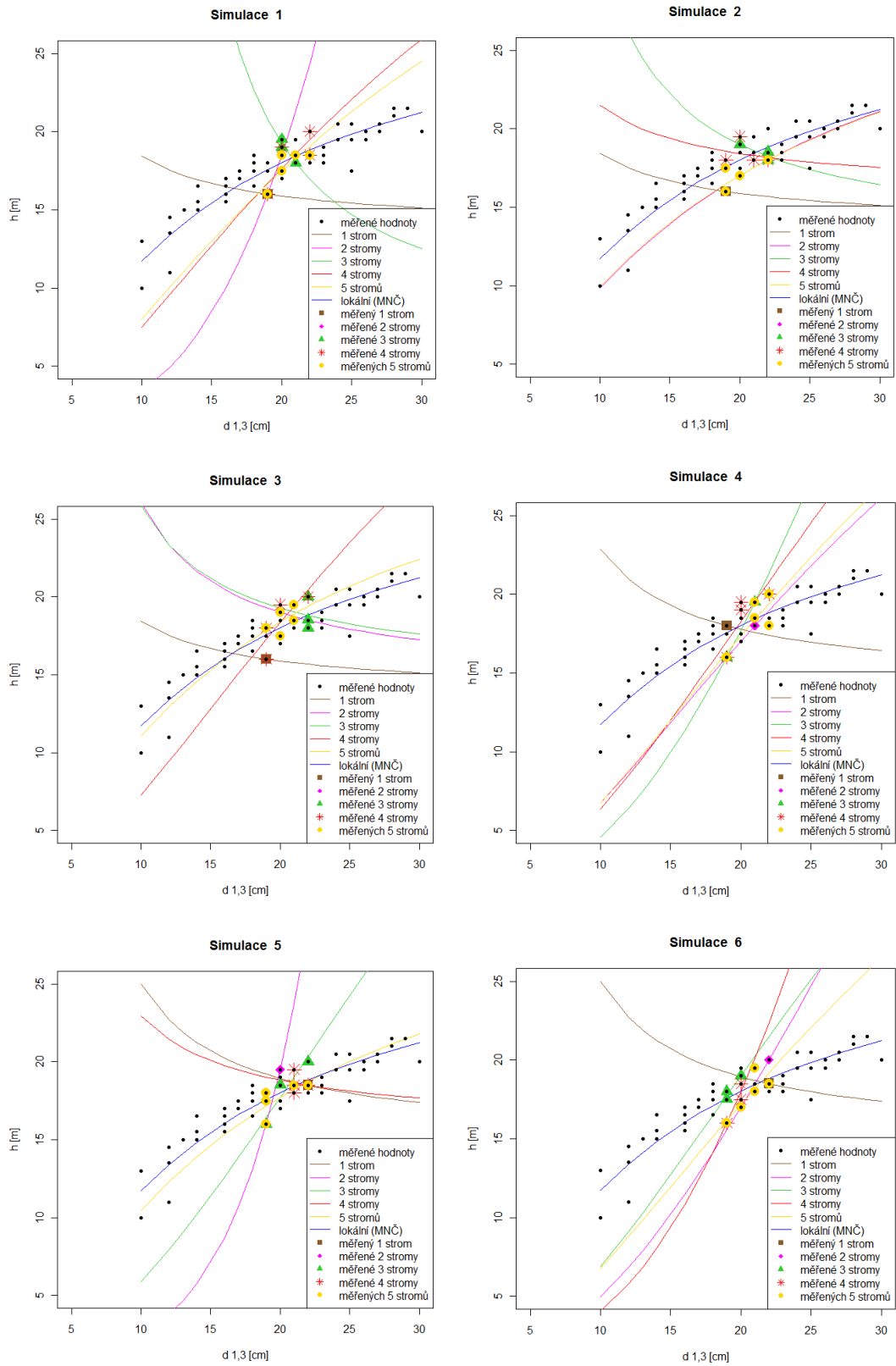
lokální model



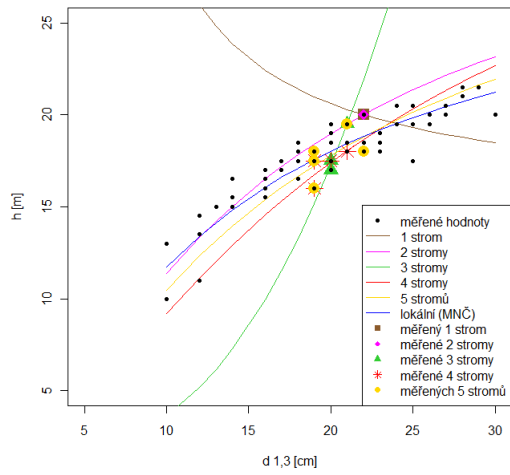
smíšený model druhé úrovně



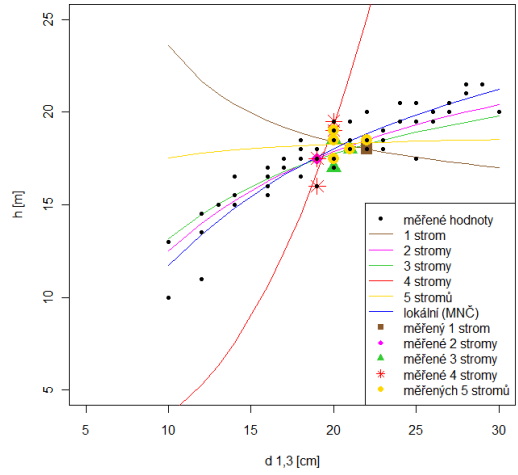
*Příloha č. 14: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 1 – 5 vybraných stromů se střední tloušťkou (výběr simulován) v porostní skupině č. 3*



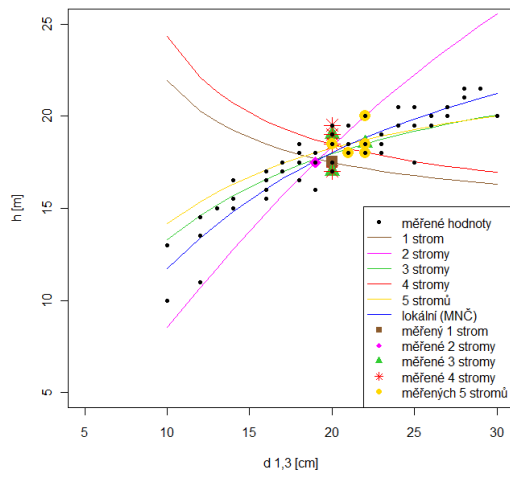
Simulace 7



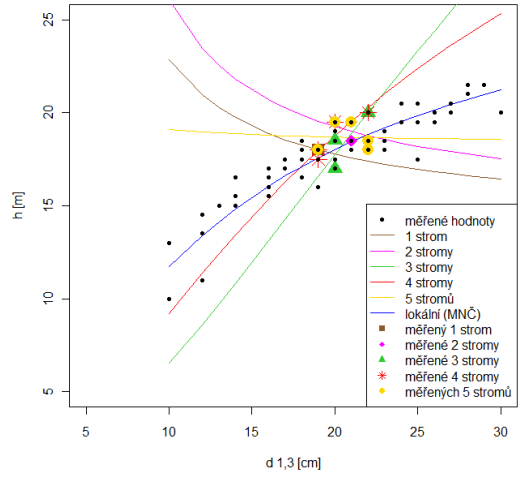
Simulace 8



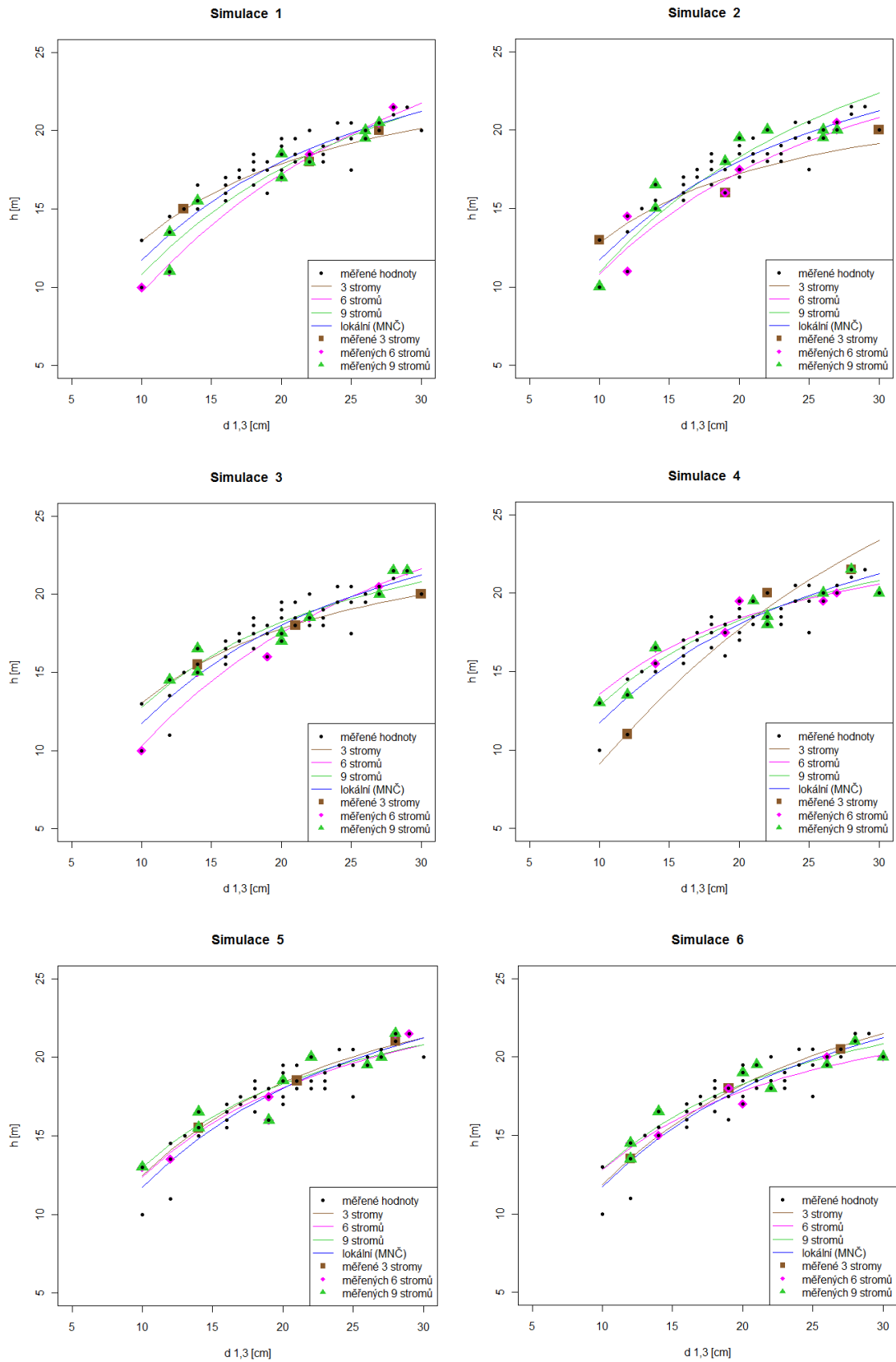
Simulace 9



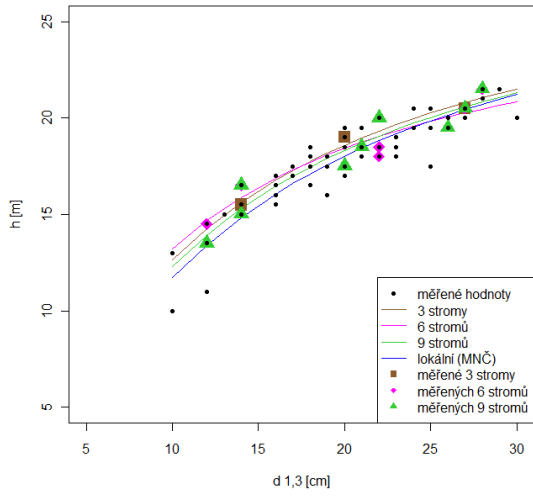
Simulace 10



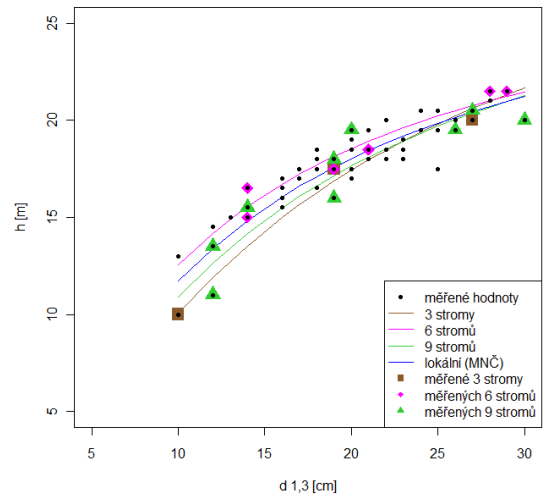
*Příloha č. 15: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 3 – 9 vybraných stromů ve třech tloušťkových intervalech (výběr simulován) v porostní skupině č. 3*



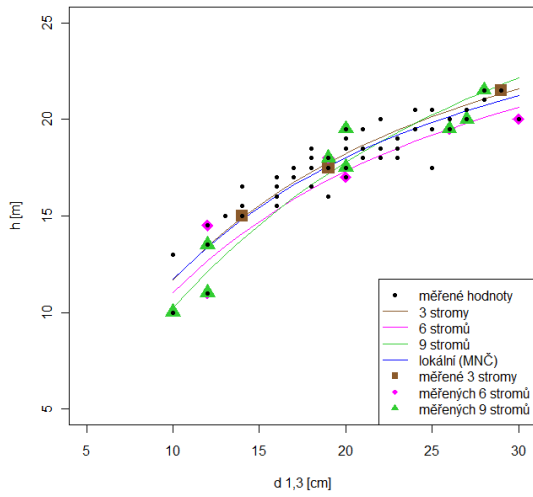
Simulace 7



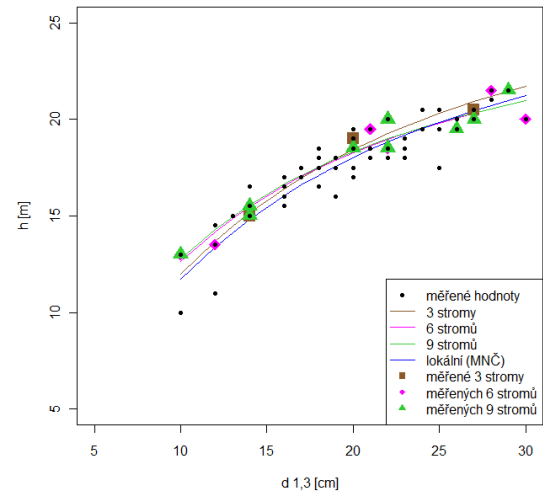
Simulace 8



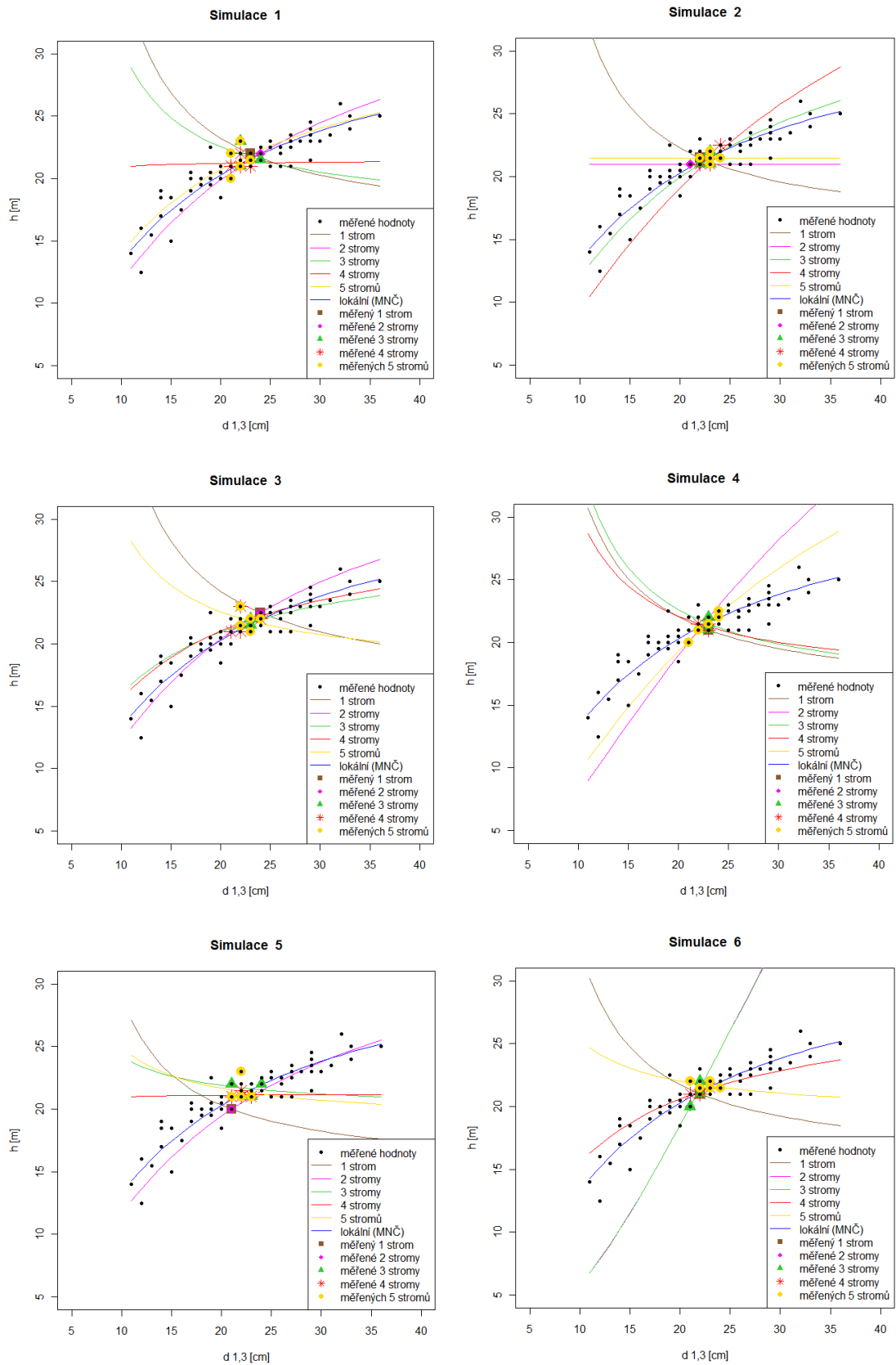
Simulace 9



Simulace 10

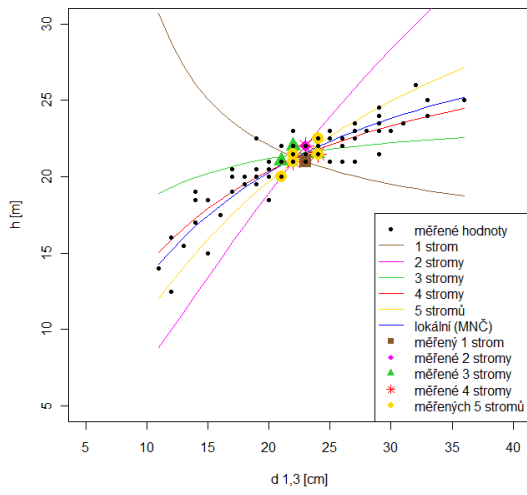


Příloha č. 16: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 1 – 5 vybraných stromů se střední tloušťkou (výběr simulován) v porostní skupině č. 5

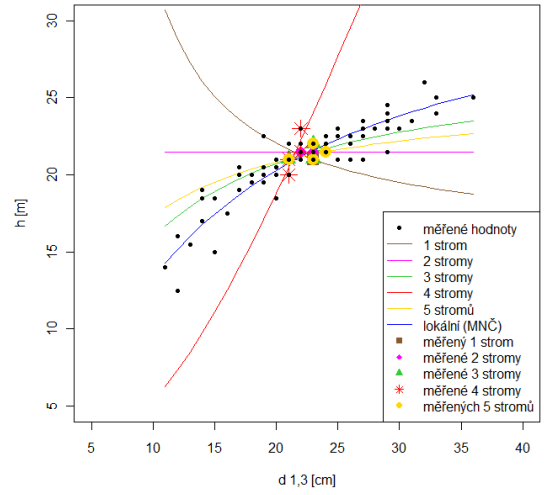




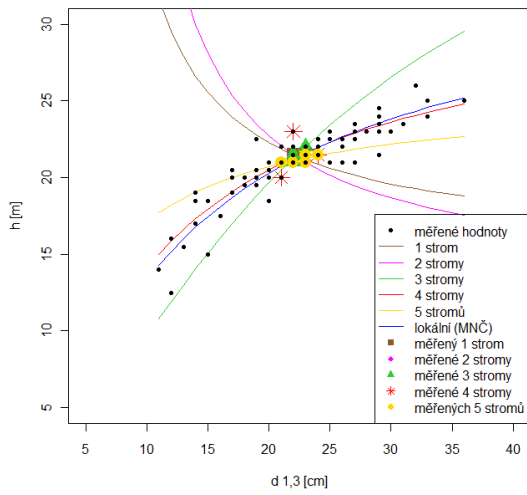
Simulace 7



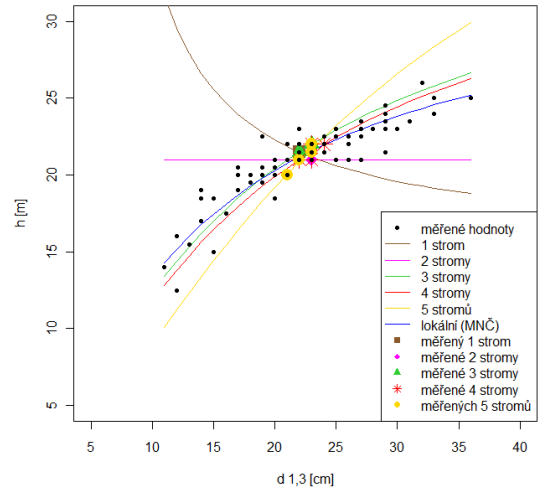
Simulace 8



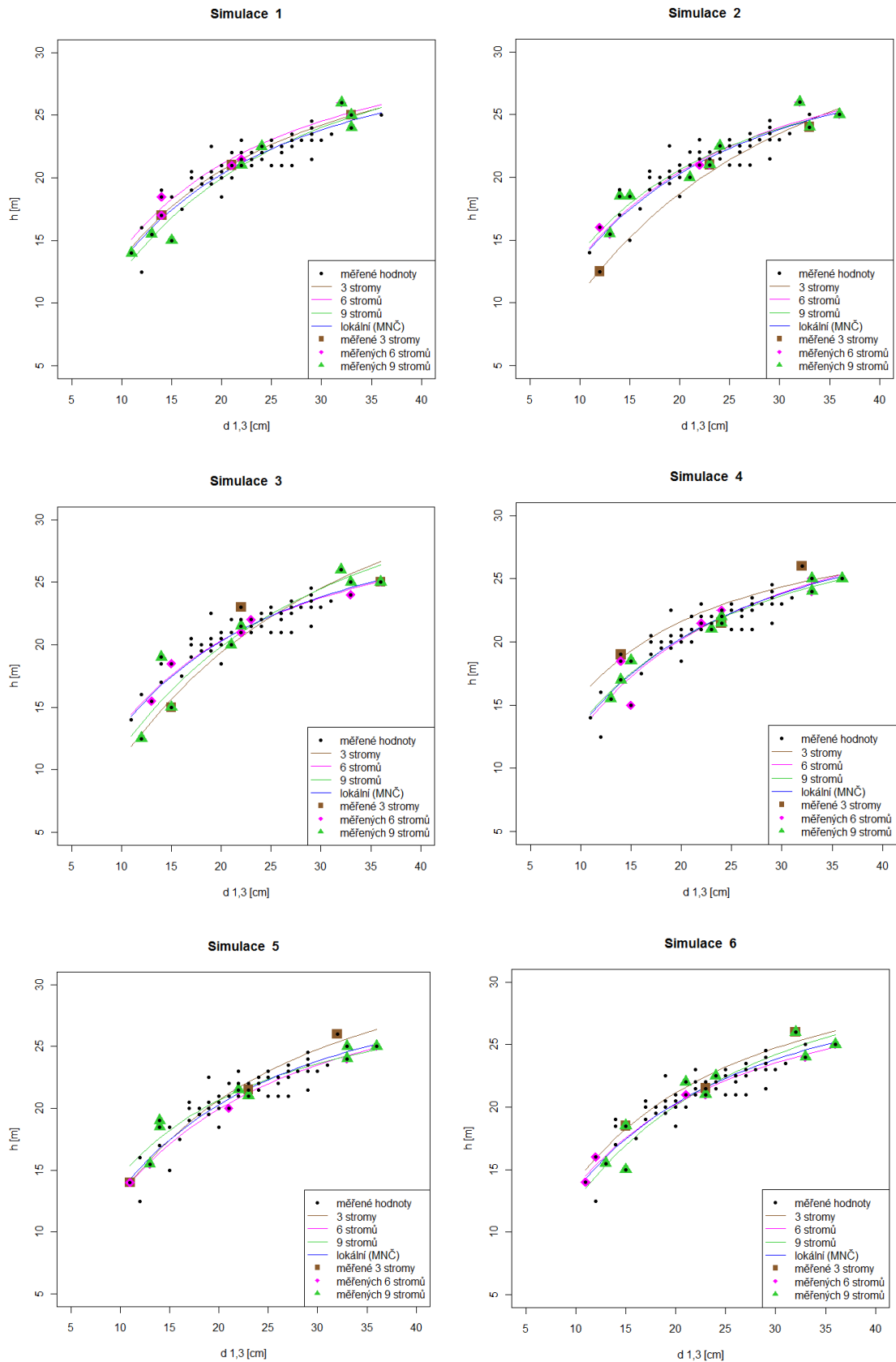
Simulace 9



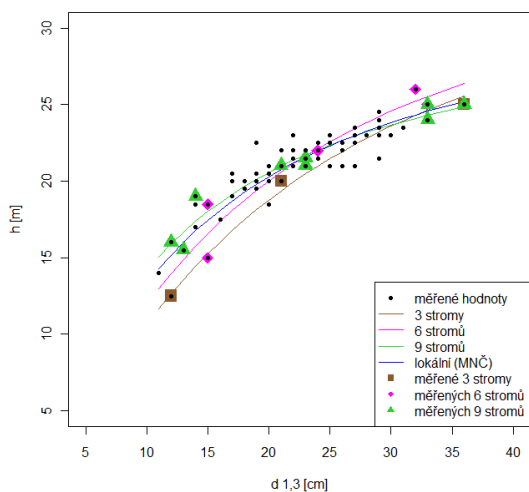
Simulace 10



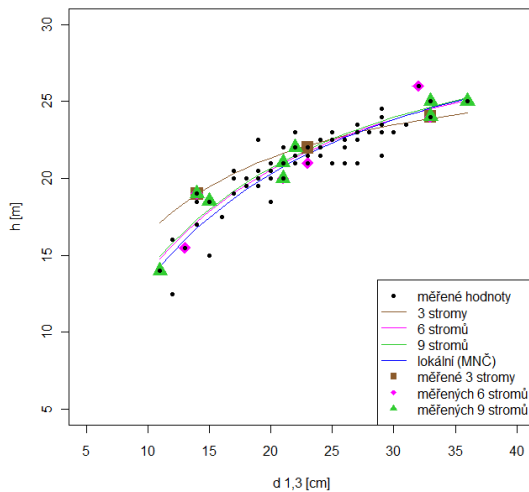
*Příloha č. 17: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 3 – 9 vybraných stromů ve třech tloušťkových intervalech (výběr simulován) v porostní skupině č. 5*



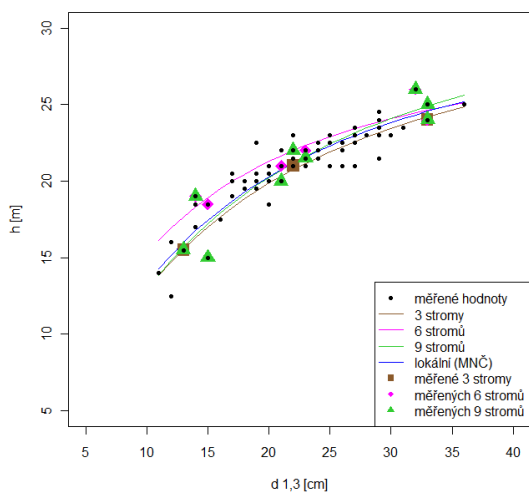
Simulace 7



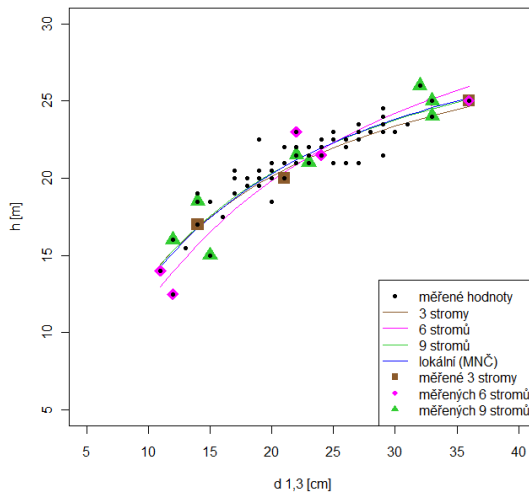
Simulace 8



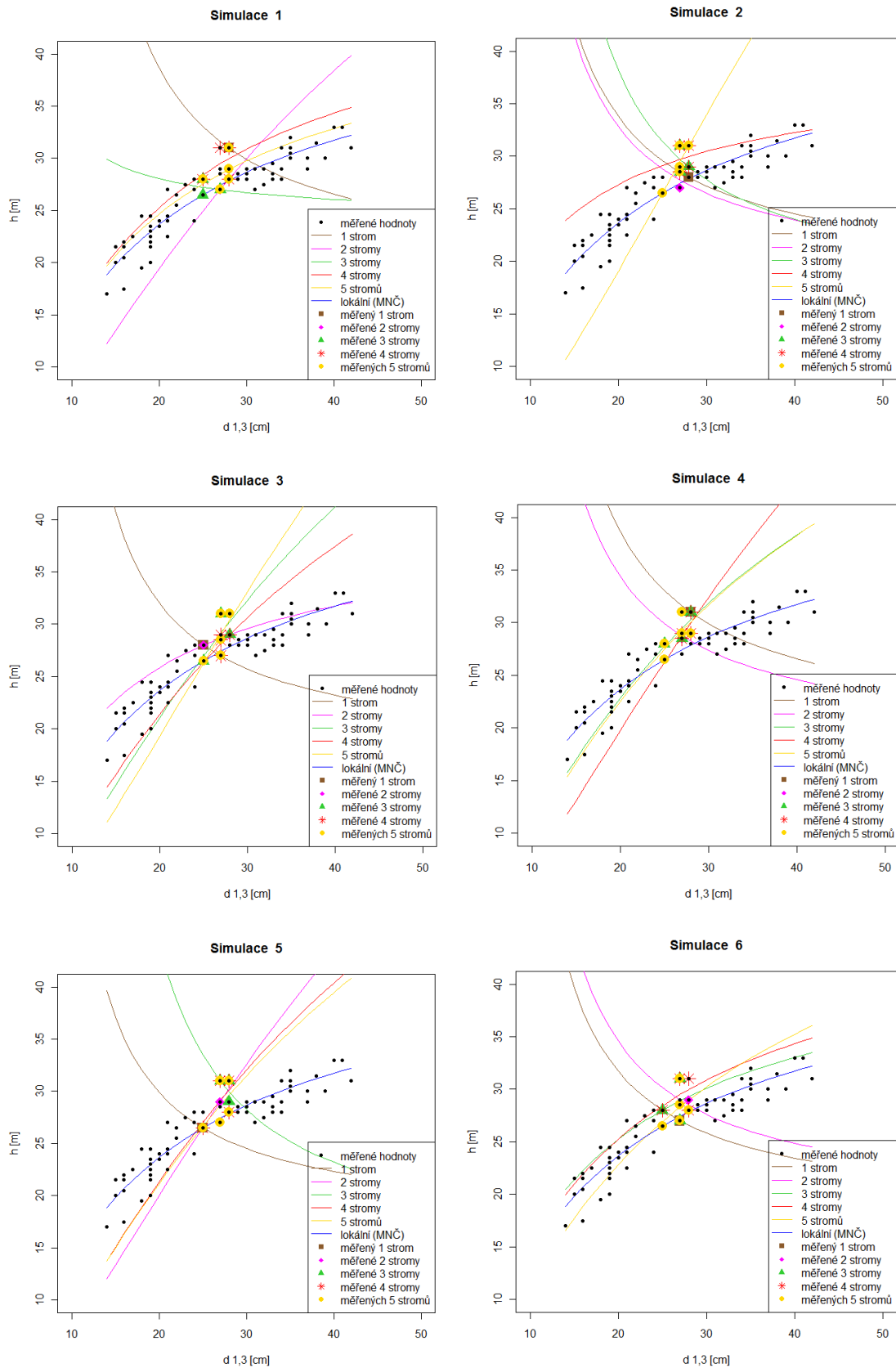
Simulace 9



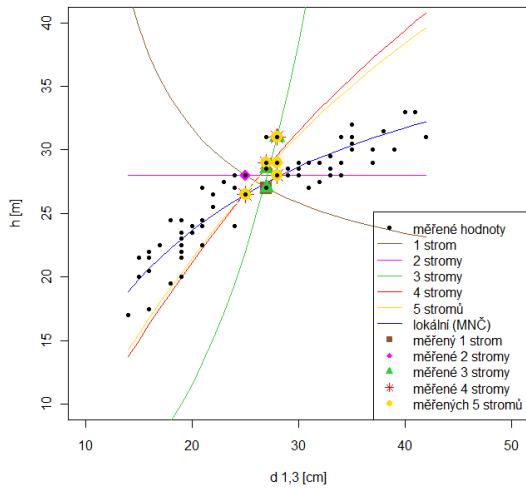
Simulace 10



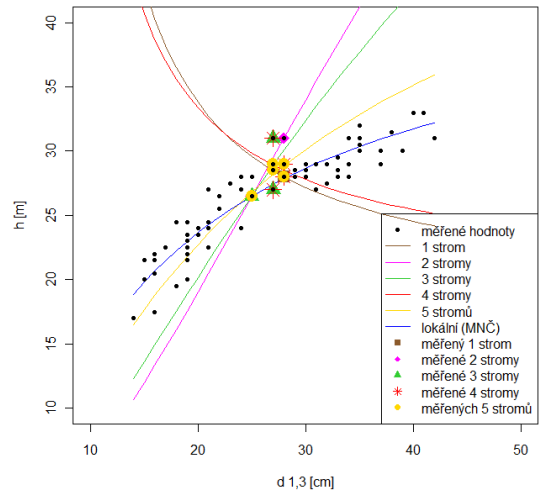
Příloha č. 18: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 1 – 5 vybraných stromů se střední tloušťkou (výběr simulován) v porostní skupině č. 11



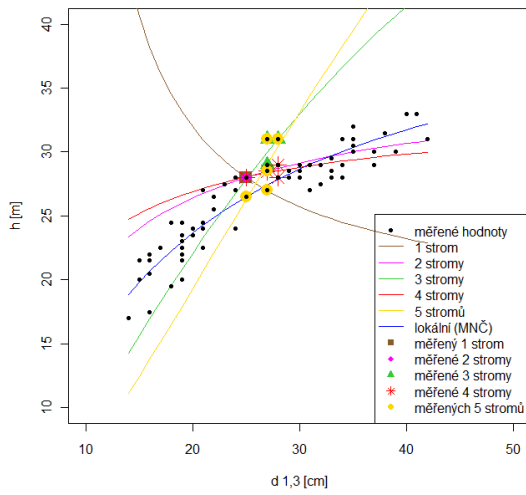
Simulace 7



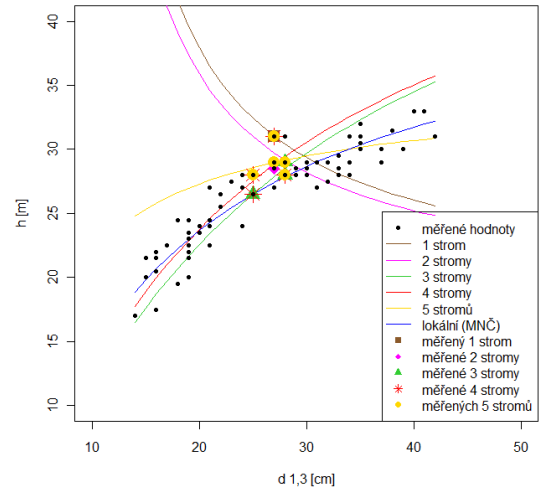
Simulace 8



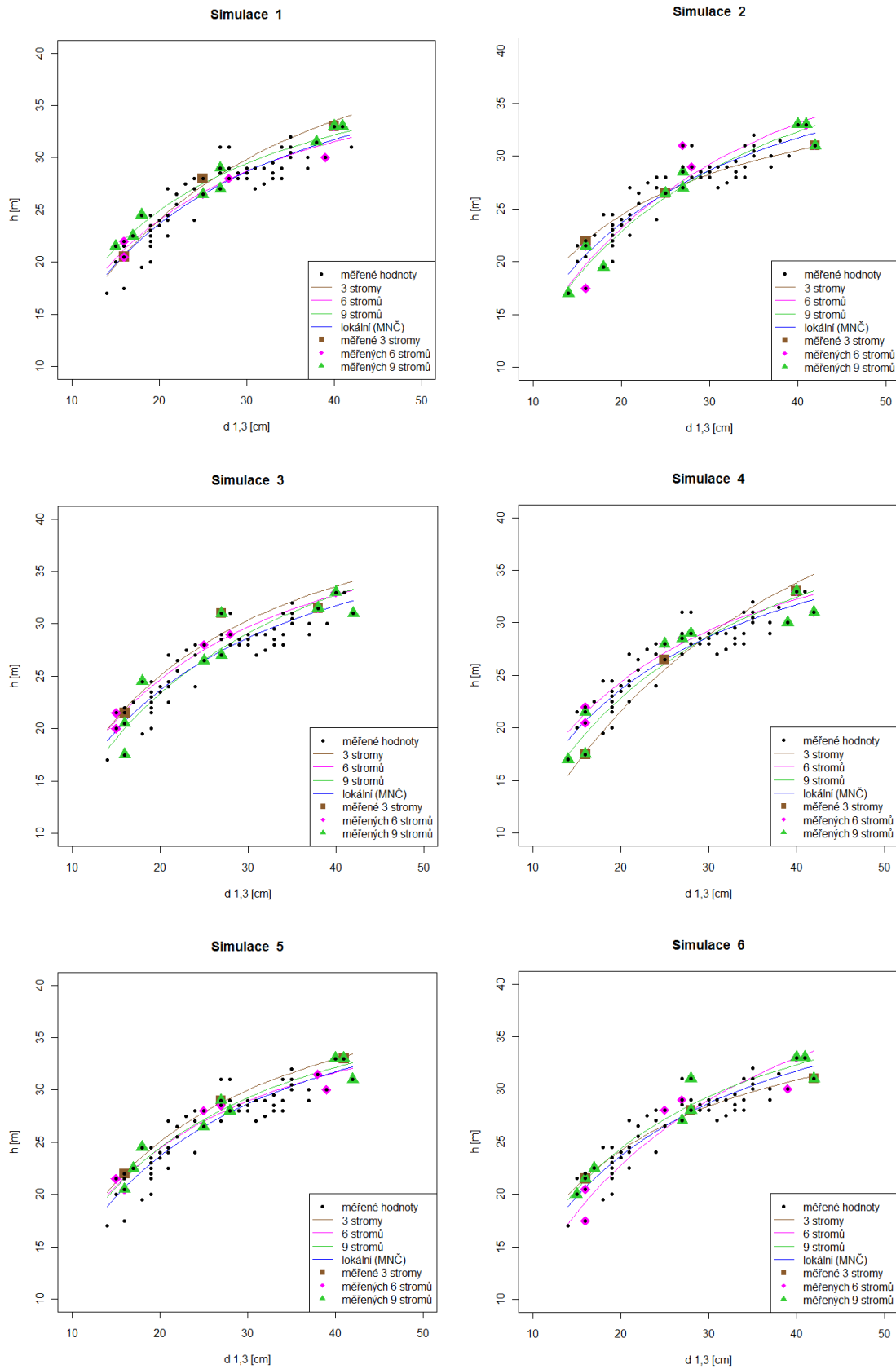
Simulace 9



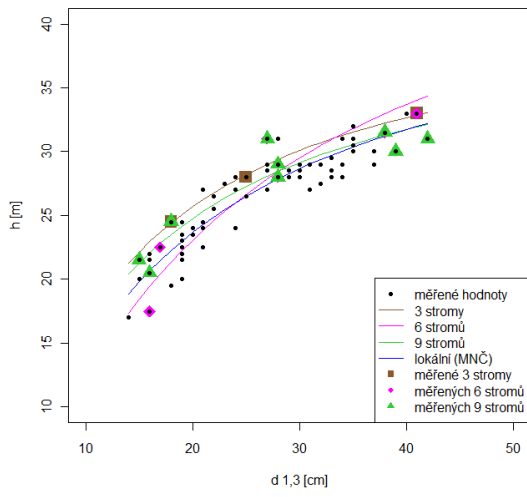
Simulace 10



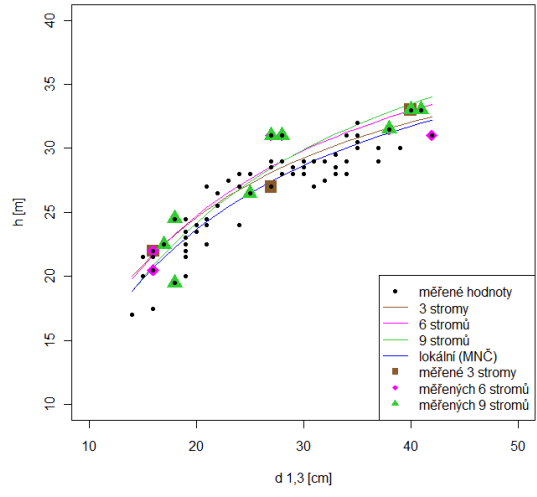
*Příloha č. 19: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 3 – 9 vybraných stromů ve třech tloušťkových intervalech (výběr simulován) v porostní skupině č. 11*



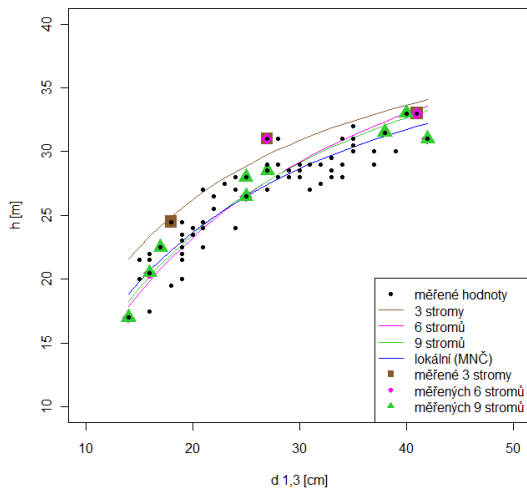
Simulace 7



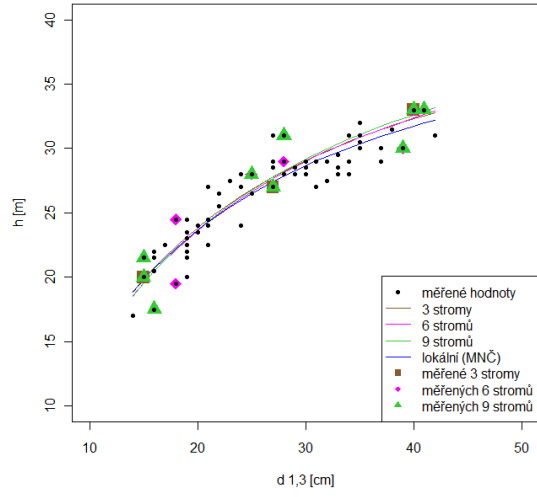
Simulace 8



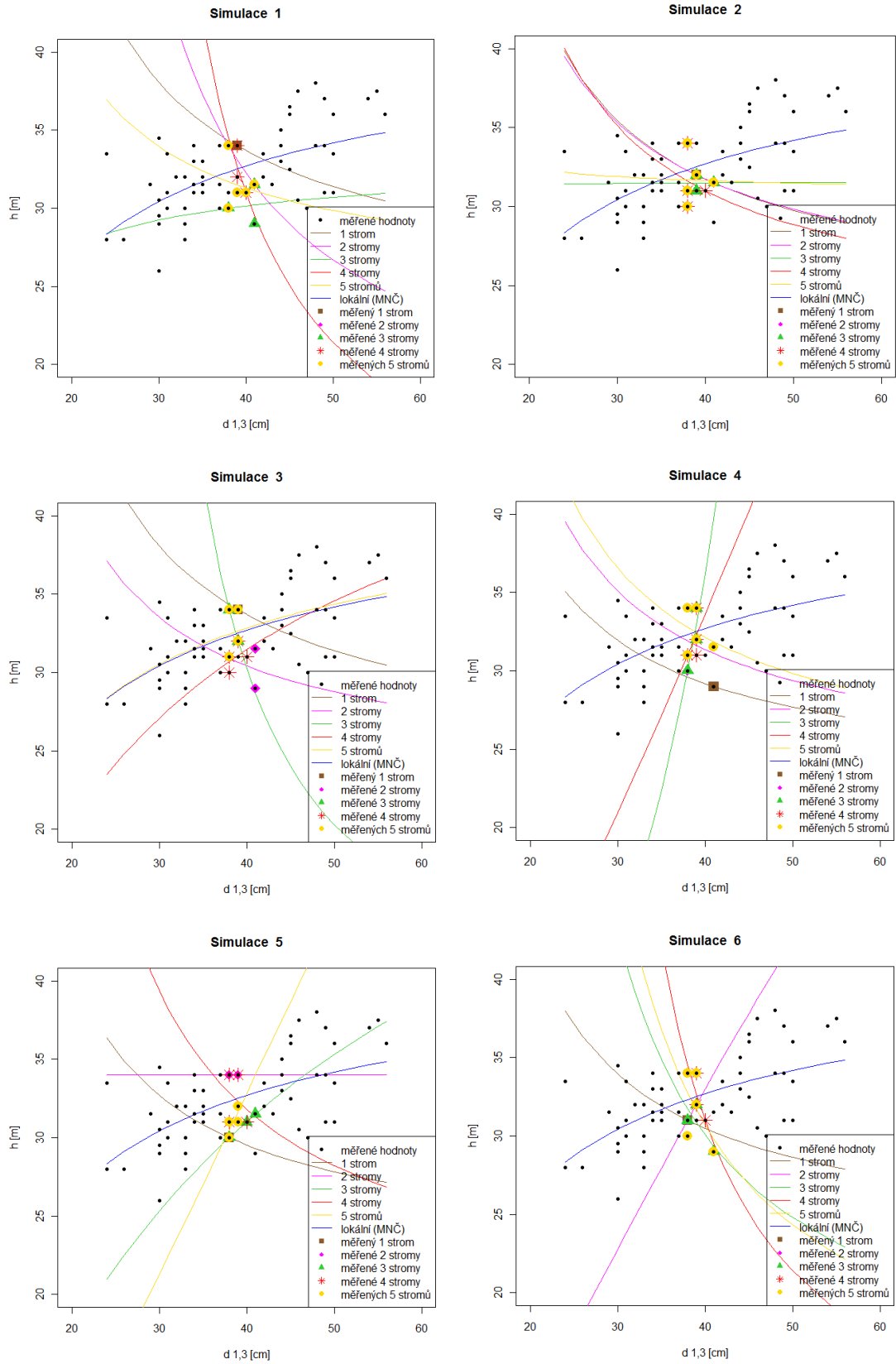
Simulace 9



Simulace 10

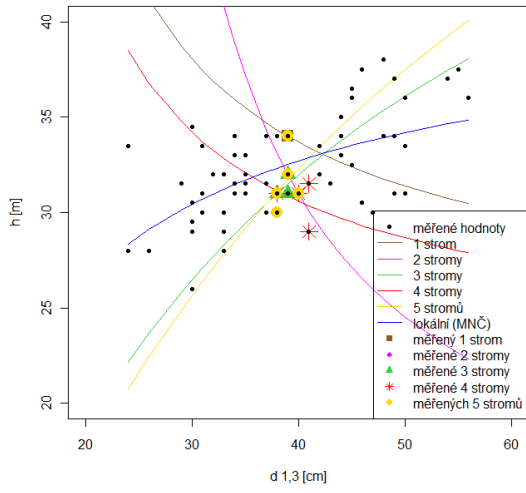


*Příloha č. 20: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 1 – 5 vybraných stromů se střední tloušťkou (výběr simulován) v porostní skupině č. 14*

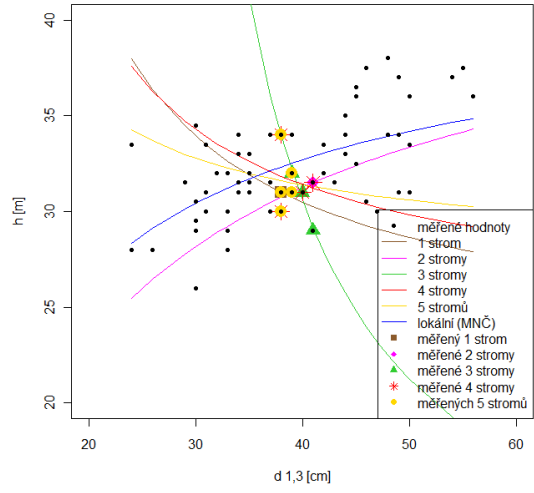




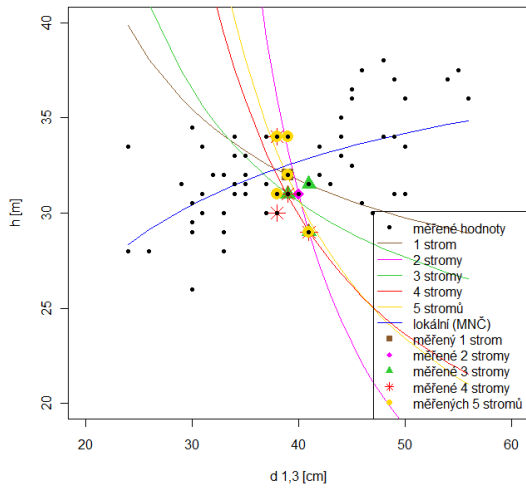
Simulace 7



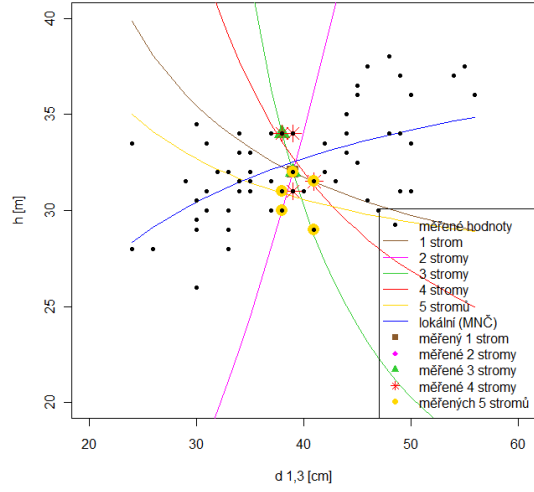
Simulace 8



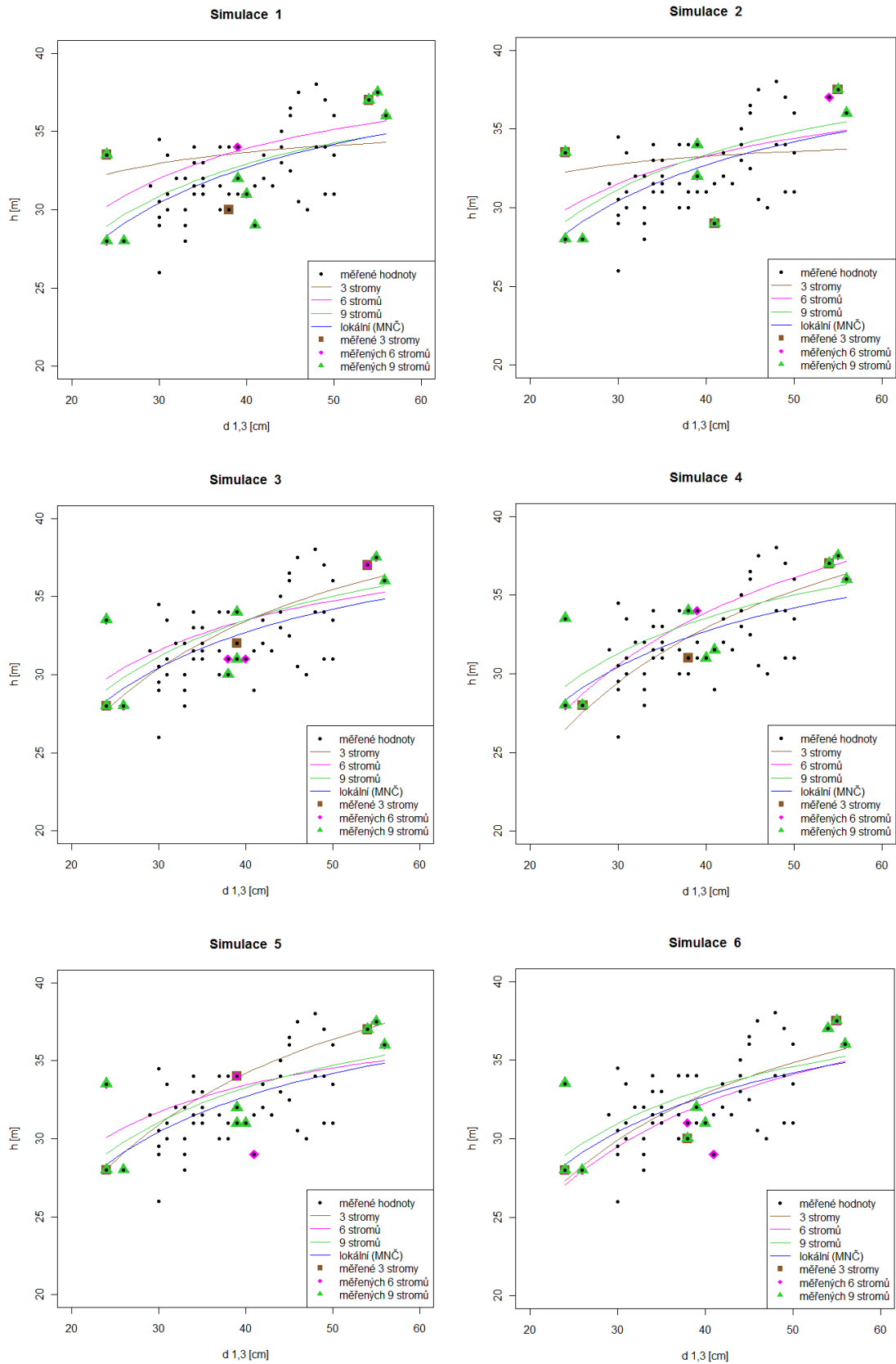
Simulace 9



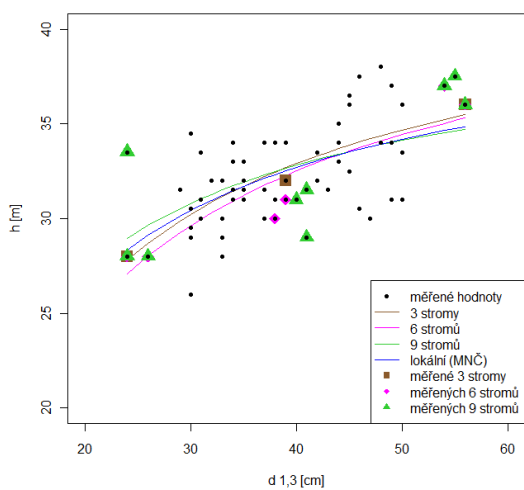
Simulace 10



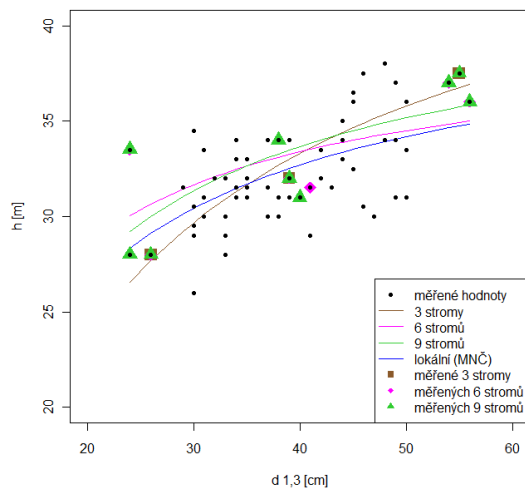
*Příloha č. 21: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 3 – 9 vybraných stromů ve třech tloušťkových intervalech (výběr simulován) v porostní skupině č. 14*



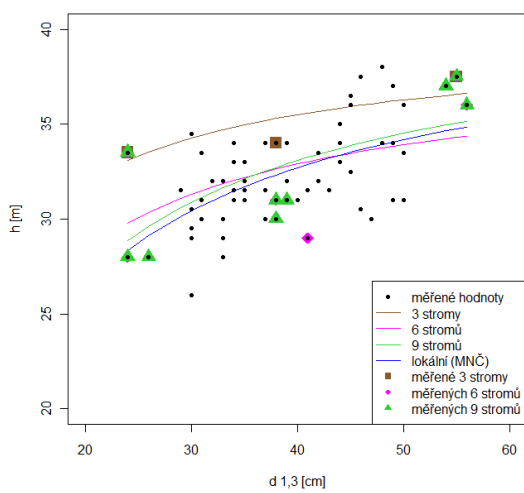
Simulace 7



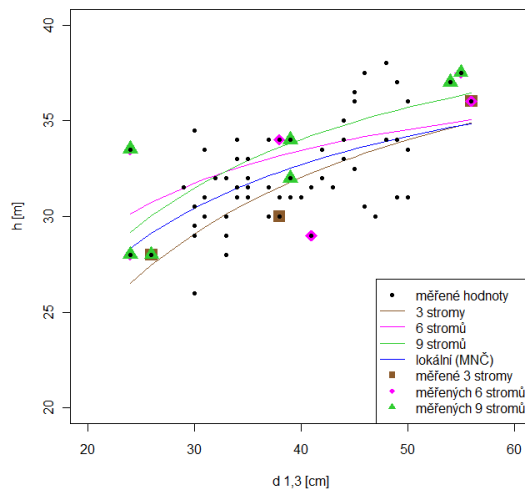
Simulace 8



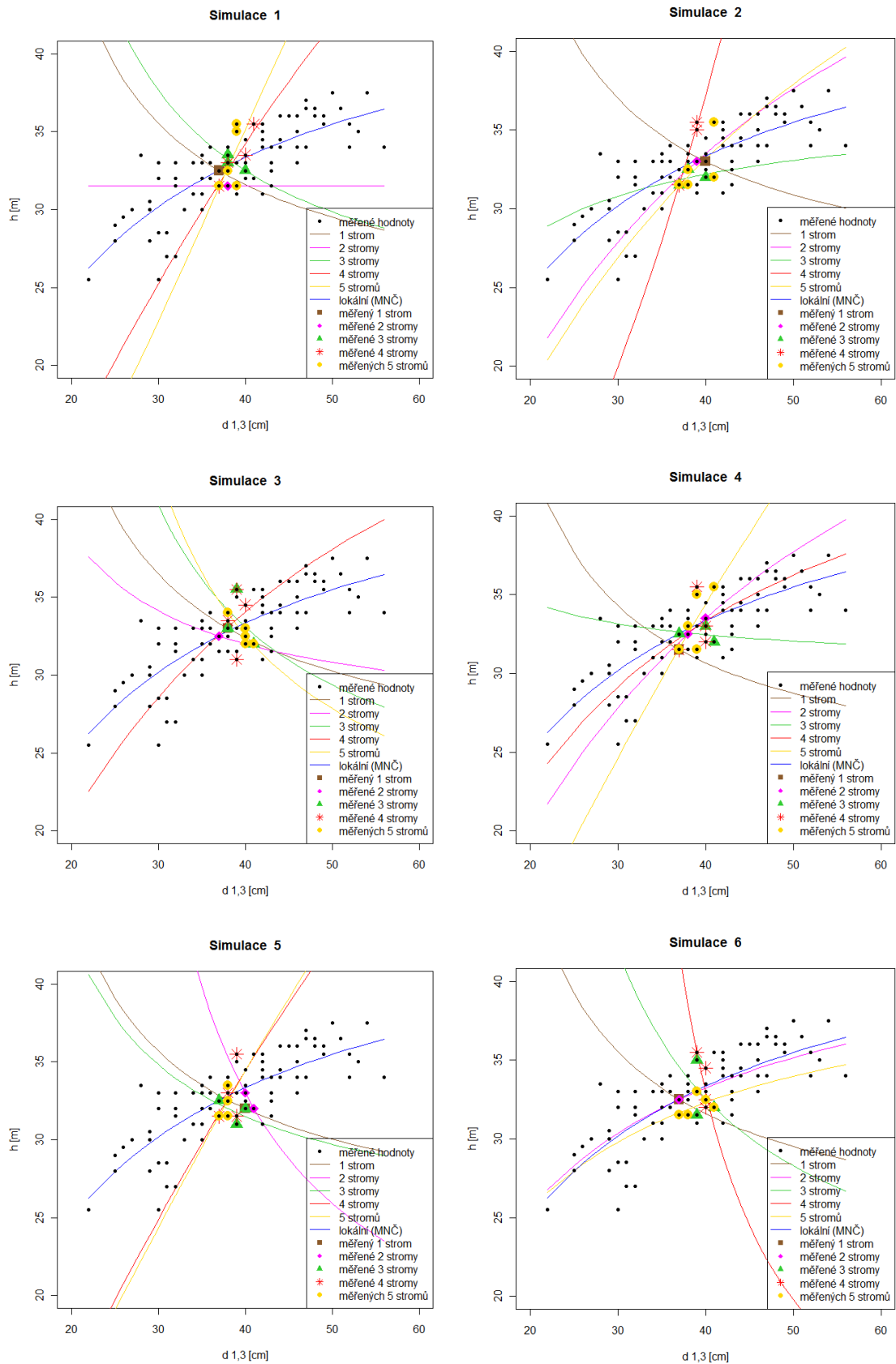
Simulace 9



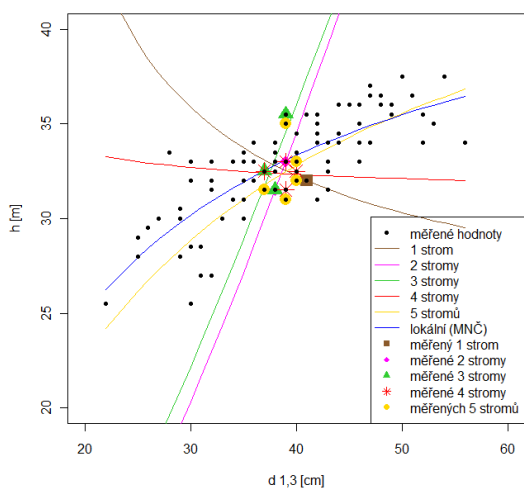
Simulace 10



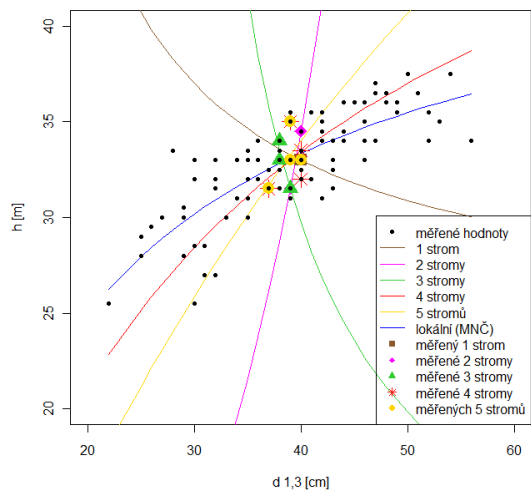
Příloha č. 22: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 1 – 5 vybraných stromů se střední tloušťkou (výběr simulován) v porostní skupině č. 19



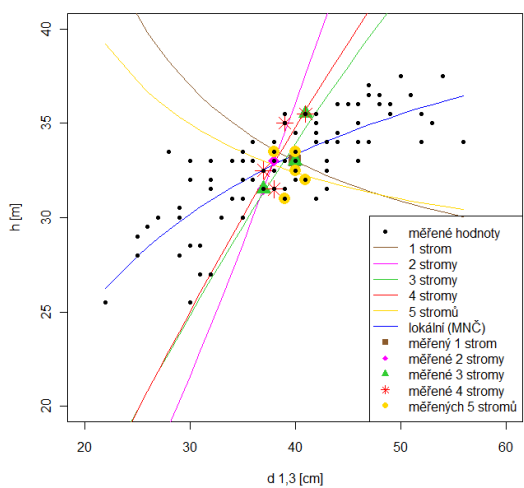
Simulace 7



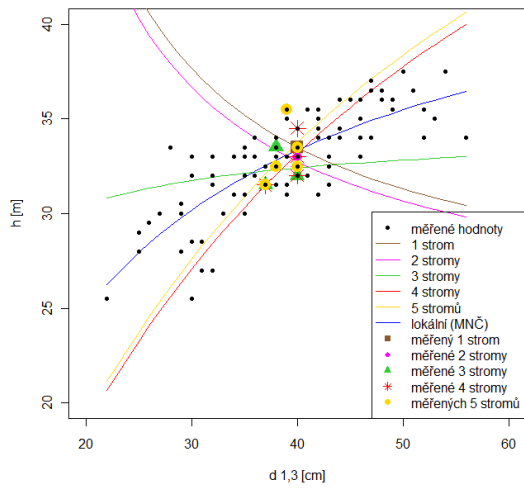
Simulace 8



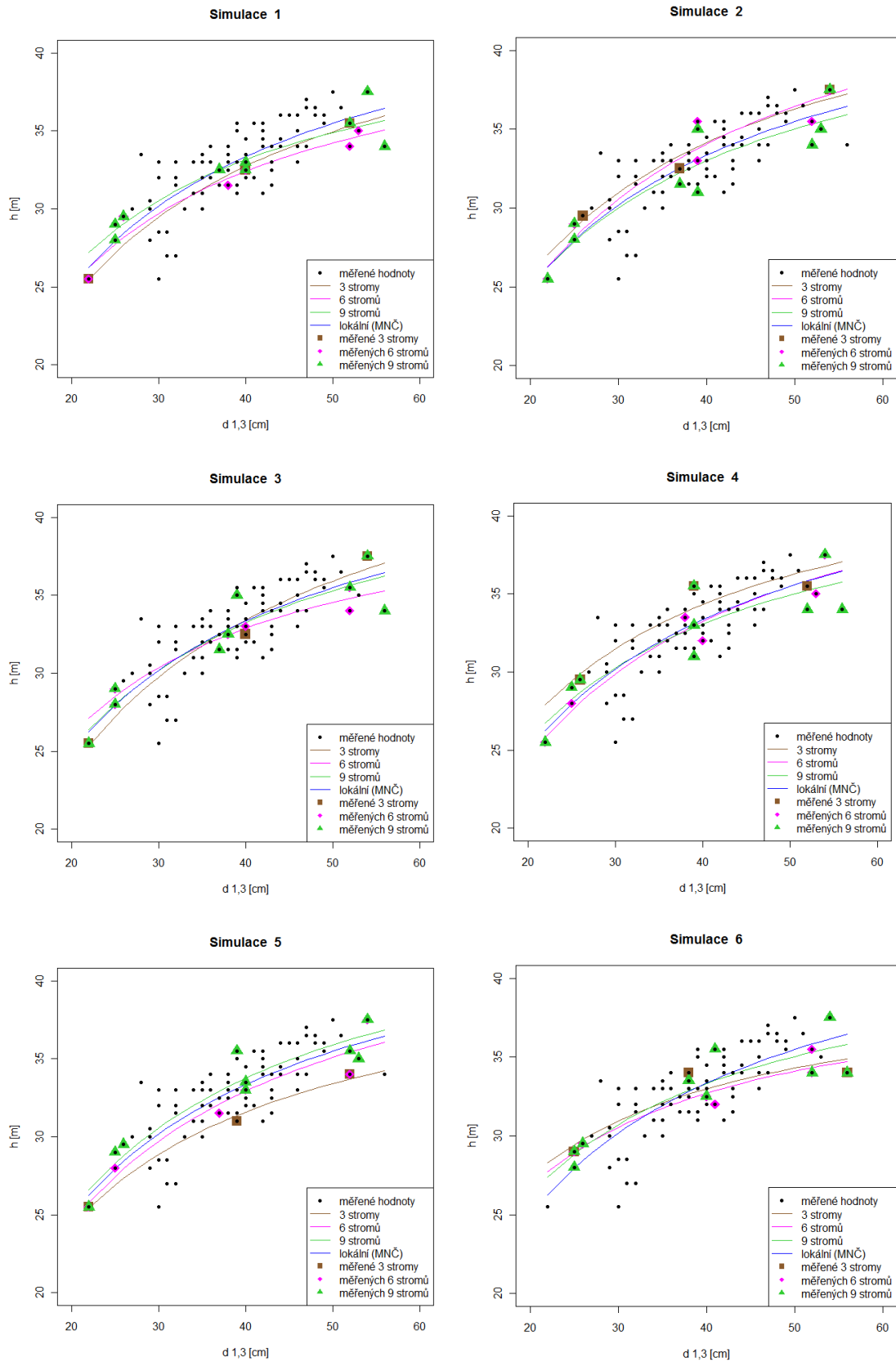
Simulace 9



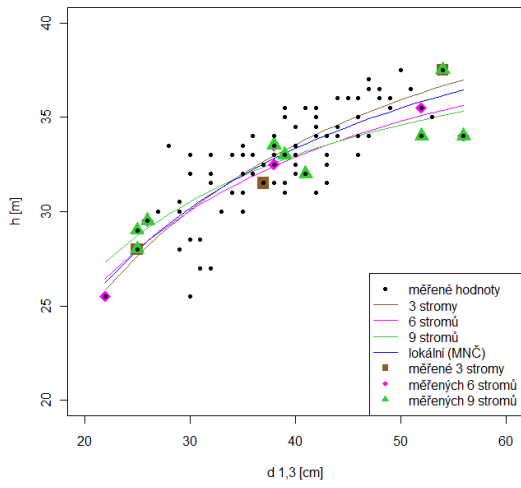
Simulace 10



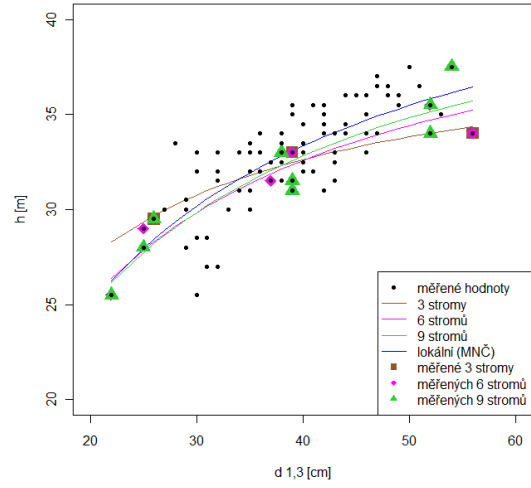
*Příloha č. 23: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 3 – 9 vybraných stromů ve třech tloušťkových intervalech (výběr simulován) v porostní skupině č. 19*



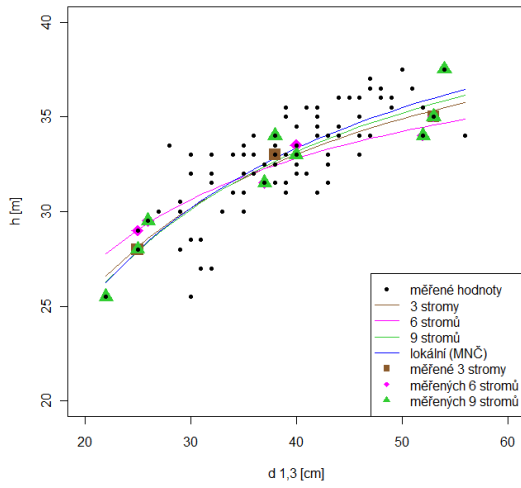
Simulace 7



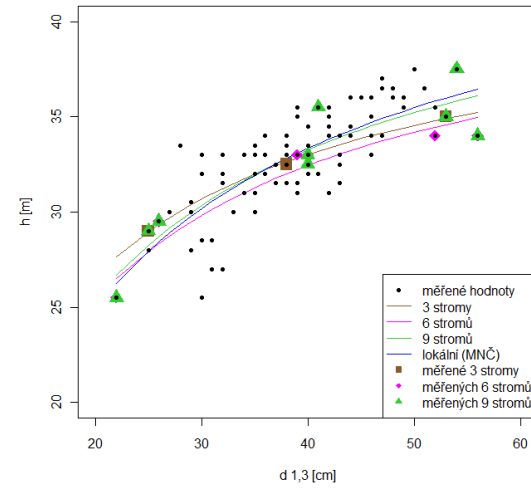
Simulace 8



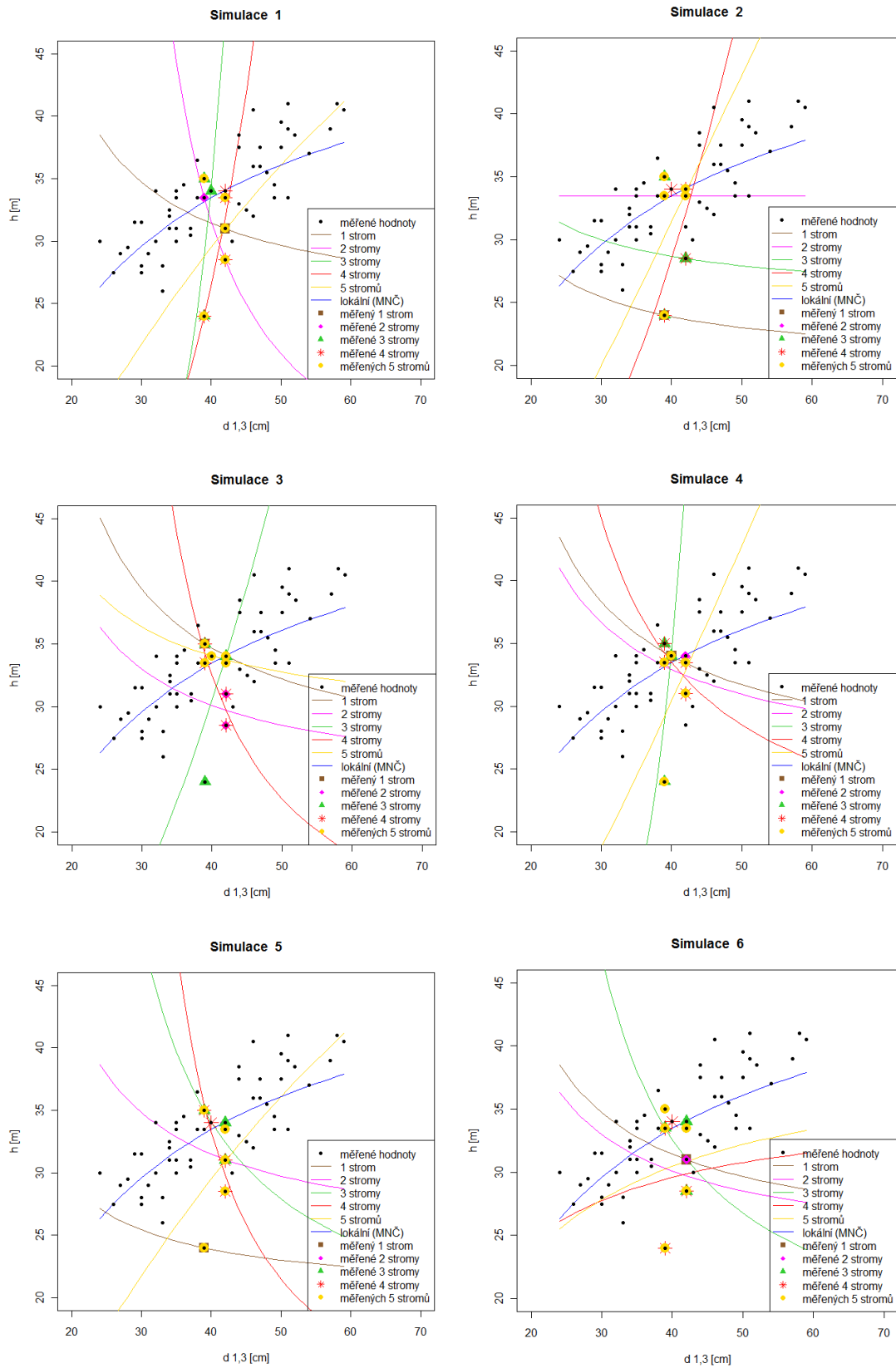
Simulace 9



Simulace 10

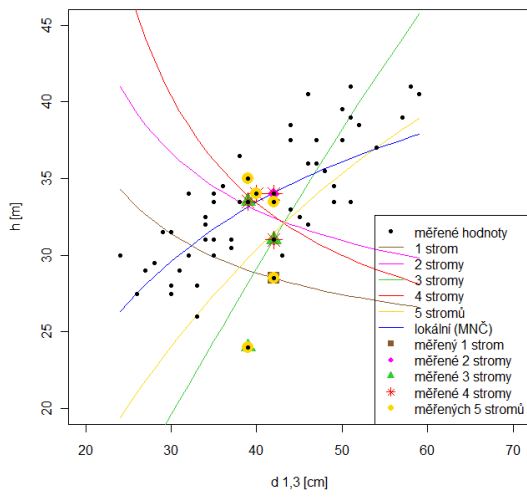


*Příloha č. 24: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 1 – 5 vybraných stromů se střední tloušťkou (výběr simulován) v porostní skupině č. 22*

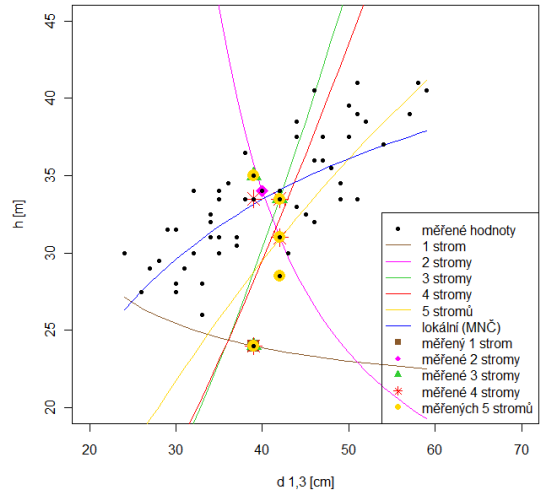




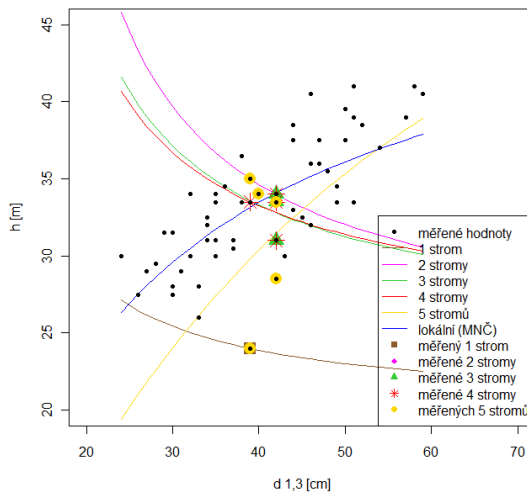
Simulace 7



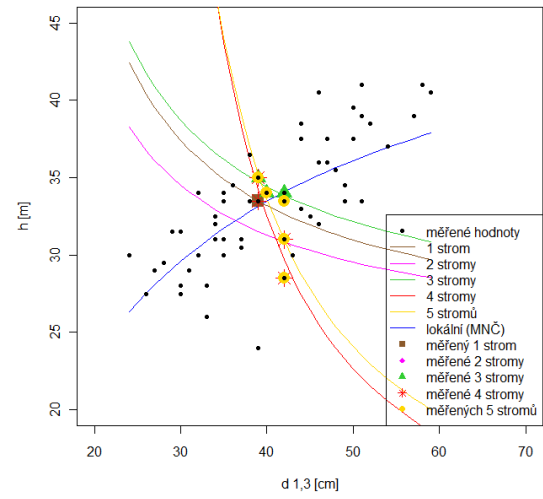
Simulace 8



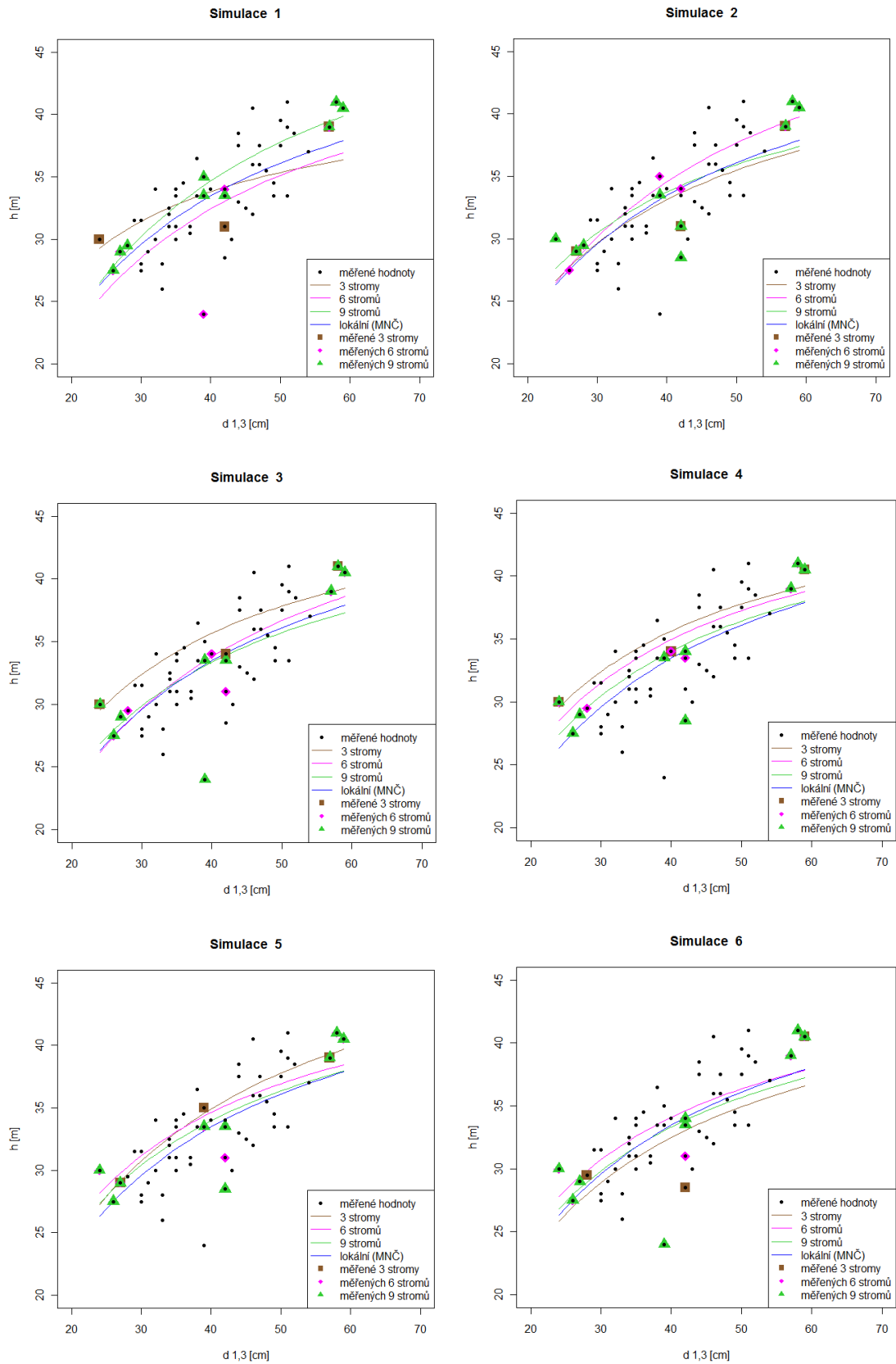
Simulace 9



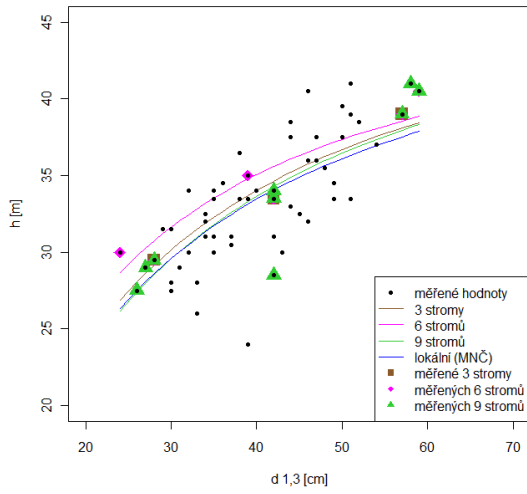
Simulace 10



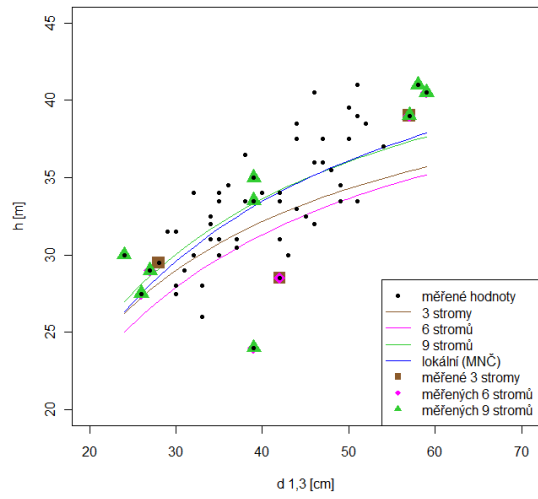
Příloha č. 25: Výsledky kalibrace smíšeného modelu druhé úrovně při změření 3 – 9 vybraných stromů ve třech tloušťkových intervalech (výběr simulován) v porostní skupině č. 22



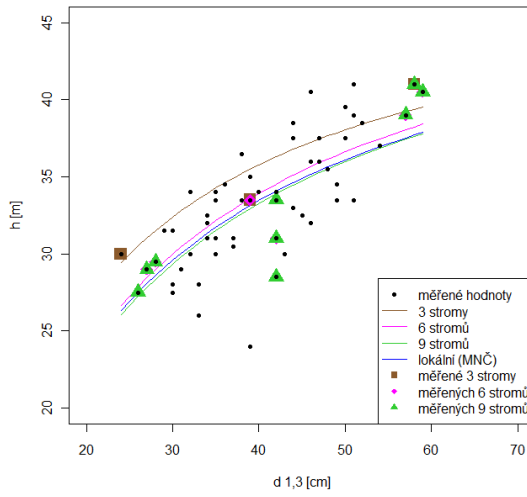
Simulace 7



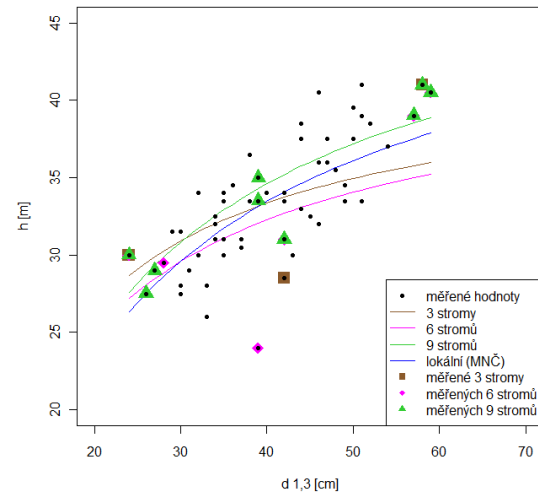
Simulace 8



Simulace 9



Simulace 10









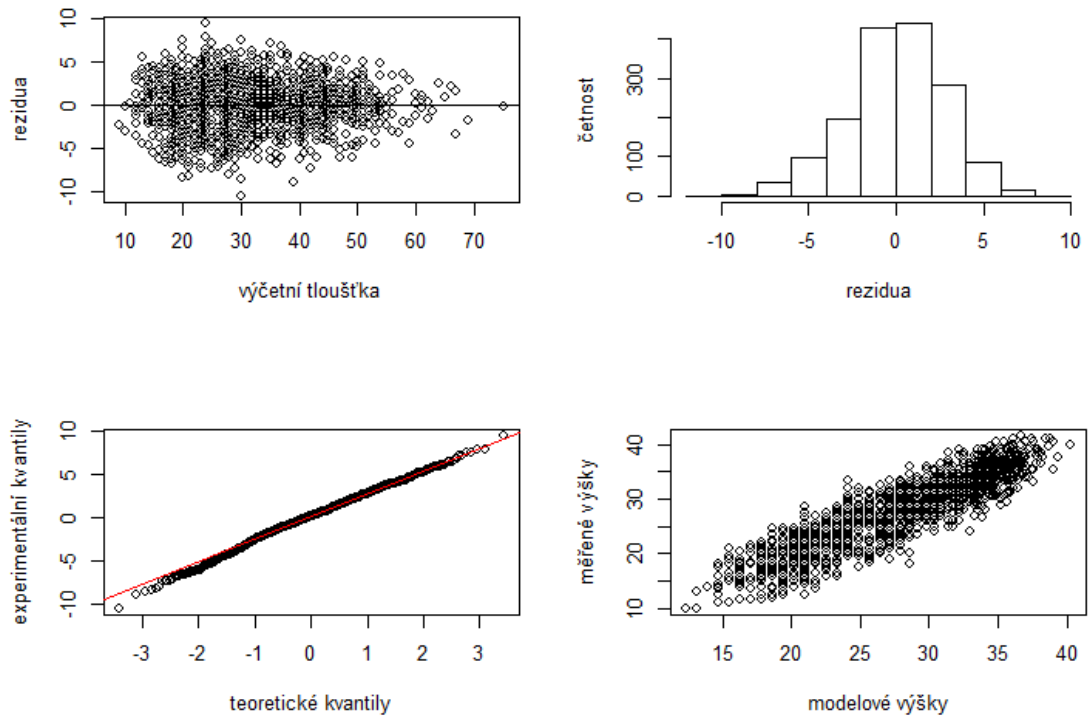




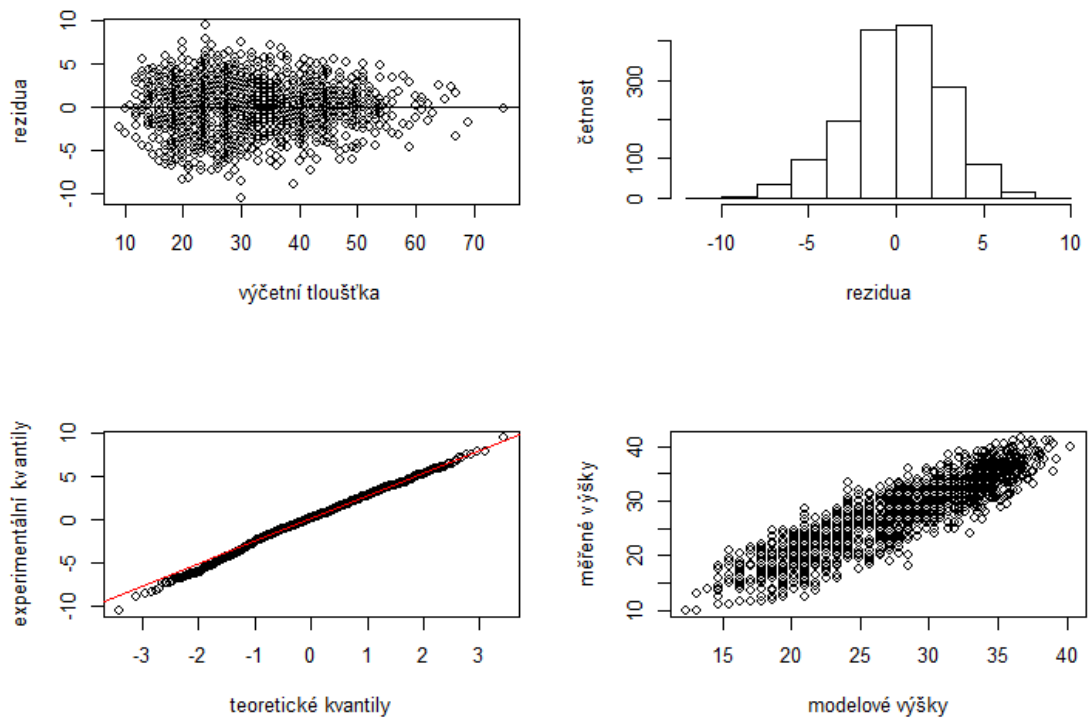




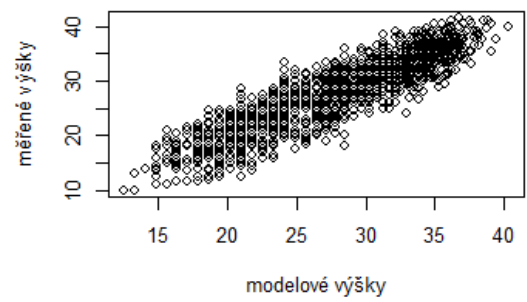
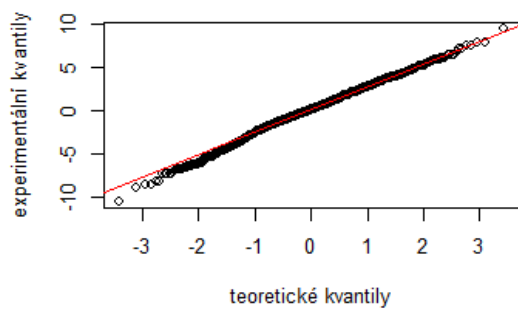
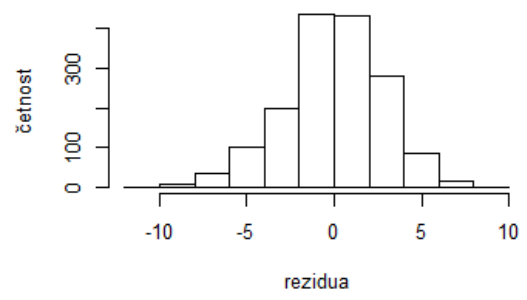
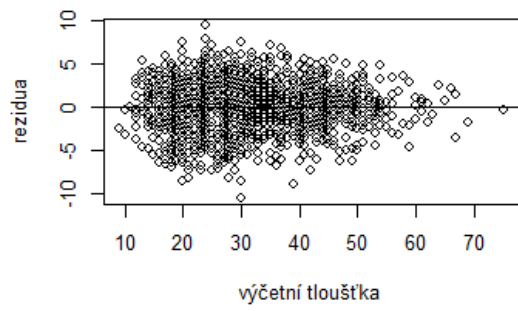
Příloha č. 32: Grafy pro hodnocení kvality různých typů globálních GAM modelů



Grafy pro GAM model s použitou spline funkcí

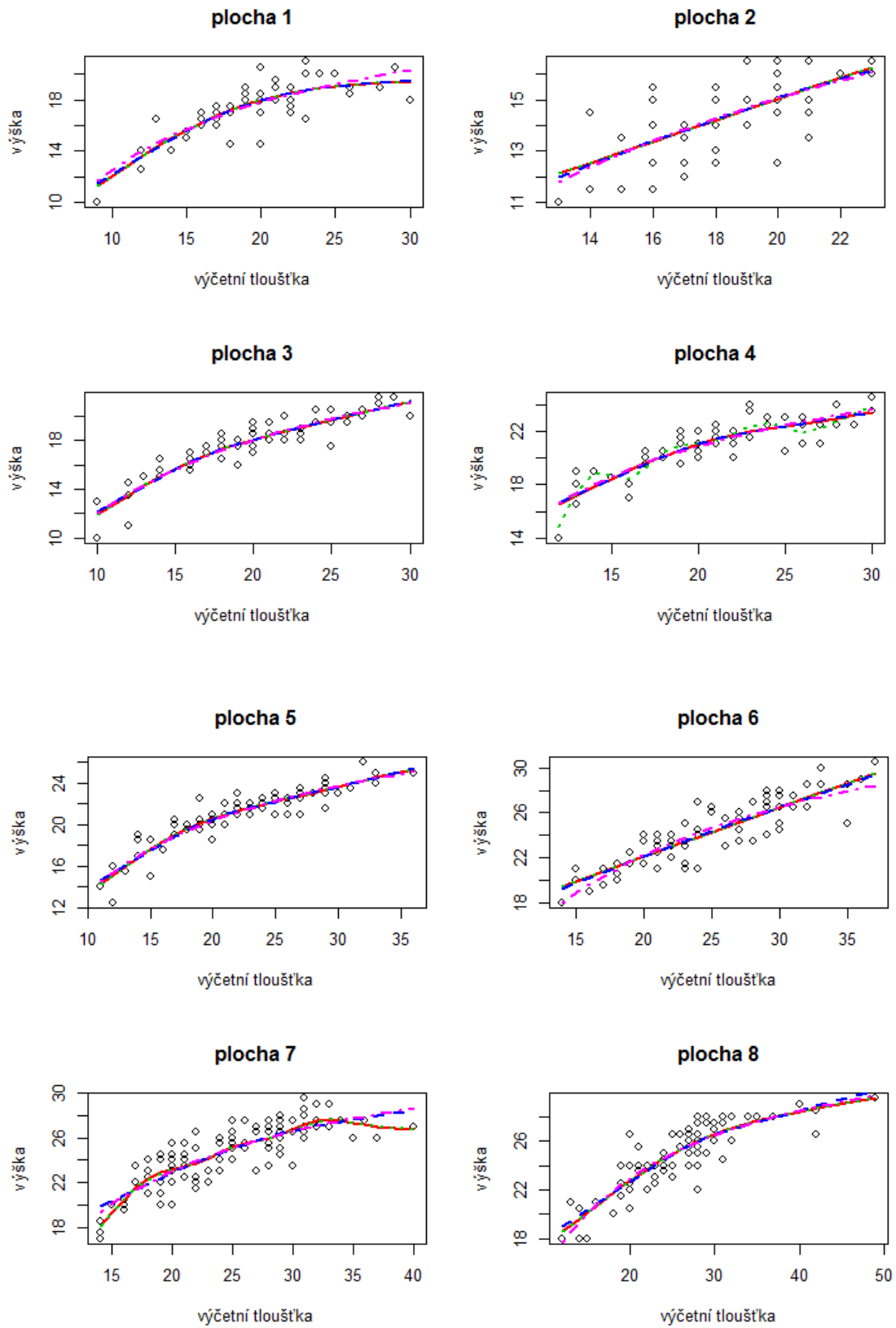


Grafy pro GAM model s použitou kubickou spline funkcí

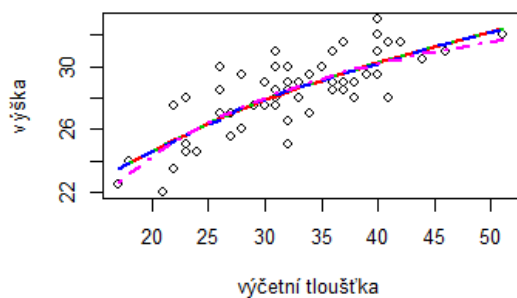


Grafy pro GAM model s použitou funkcí LOESS

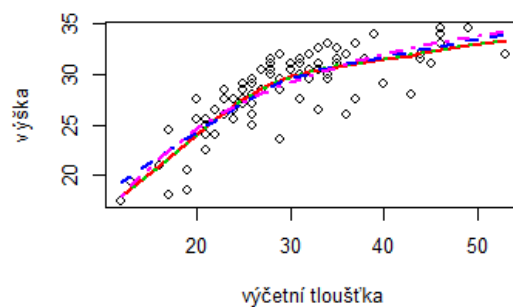
Příloha č. 33: Srovnání lokálních GAM modelů různých typů a lokálního modelu Petersonovy funkce vypočítaného nelineární regresi



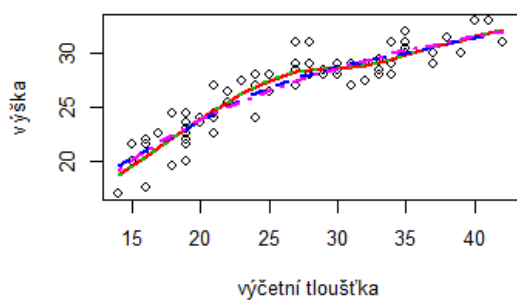
**plocha 9**



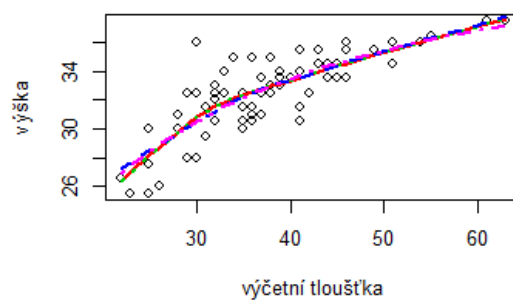
**plocha 10**



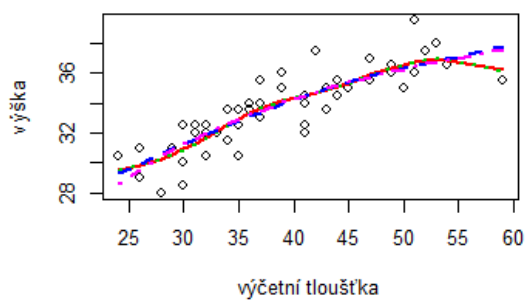
**plocha 11**



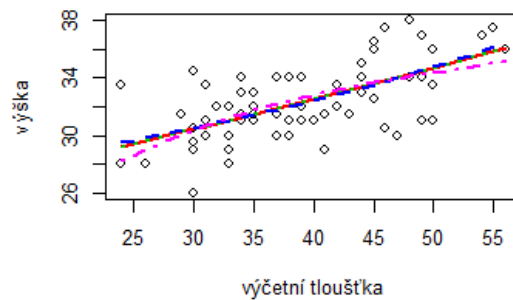
**plocha 12**



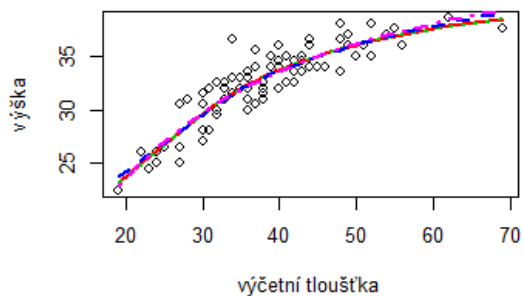
**plocha 13**



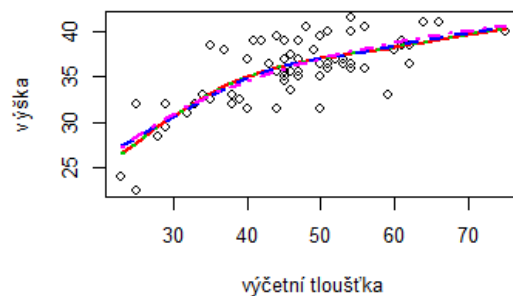
**plocha 14**

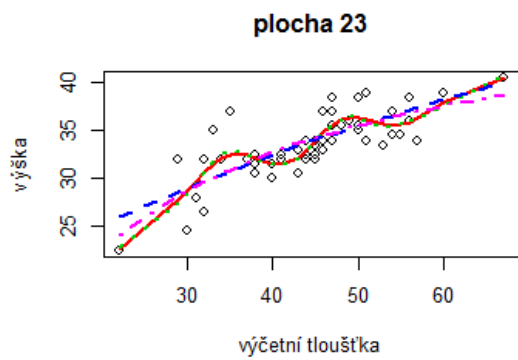
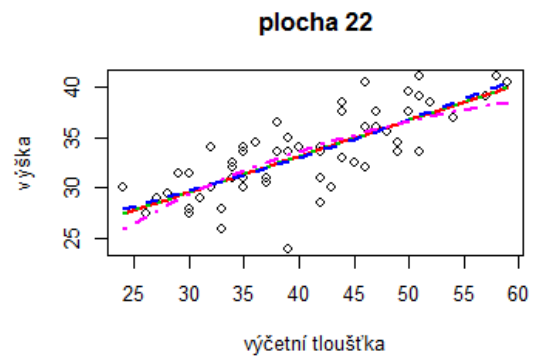
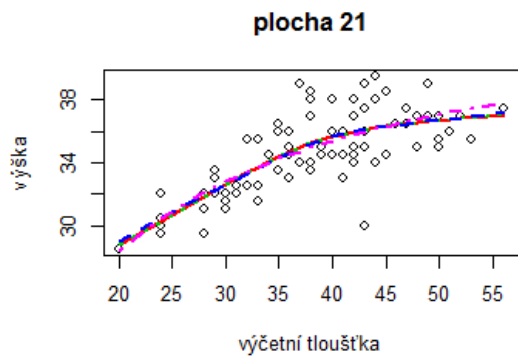
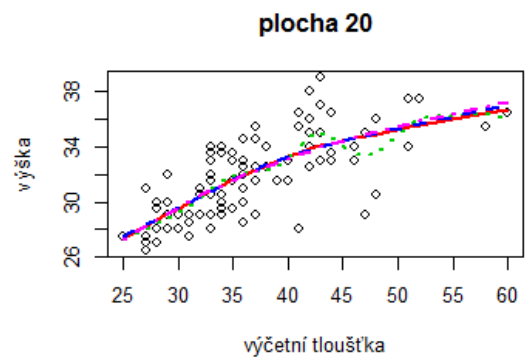
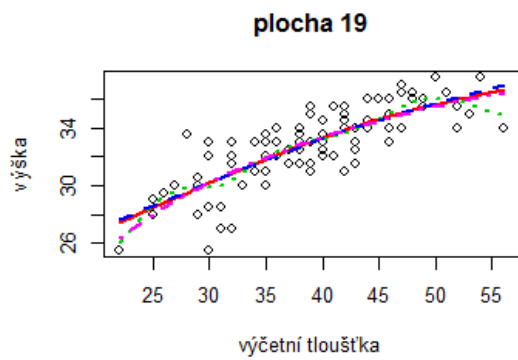
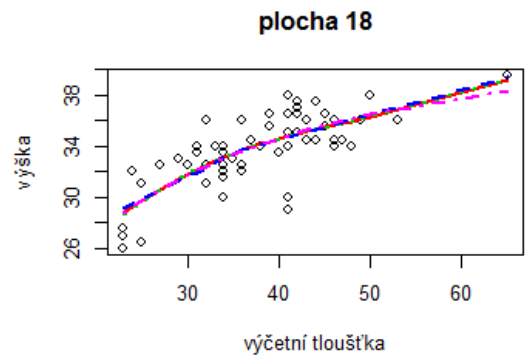
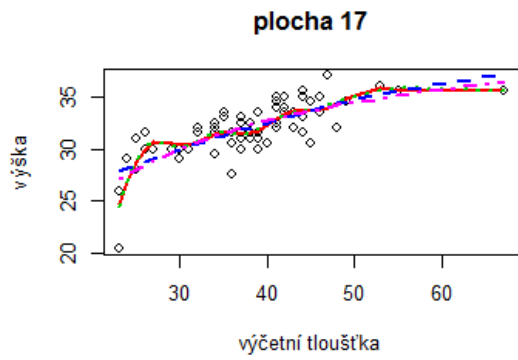


**plocha 15**



**plocha 16**





**Legenda:**

měřené hodnoty



spline



kubický spline



LOESS



Petterssonova funkce

