



MATEMATICKO-FYZIKÁLNÍ FAKULTA Univerzita Karlova

Oponentský posudek na vyhodnocení NIL-II

prof. RNDr. Jaromír Antoch, CSc.

doc. RNDr. Zdeněk Hlávka, Ph.D.

doc. RNDr. Zbyněk Pawlas, Ph.D.

Katedra pravděpodobnosti a matematické statistiky

Praha 29. srpna 2017

OBSAH

1. Úvod	1
2. Sběr dat	1
2.1. Sítě NIL1 a NIL2	2
2.2. Doporučení	2
3. Výpočet odhadů	3
3.1. Těžba	3
3.2. Připomínky k metodice	4
4. Kontrola získaných odhadů	5
5. Dokumentace	6
5.1. Doporučení	6
6. Různé	7

1. ÚVOD

Na úvod bychom chtěli zdůraznit, že NIL-II považujeme za kvalitně zpracovaný projekt. Metodika sběru, kontroly a zpracování dat jsou propracované. Použité statistické metody jsou korektní.

Dle zadání se tento posudek bude týkat především sběru dat a posouzení správnosti stanovení odhadu hlavních zjišťovaných parametrů (přírůst dříví, zásoba dříví a úbytek zásoby – těžba, plocha lesa, zastoupení dřevin) s ohledem na statistické postupy uplatněné při vyhodnocení druhého cyklu Národní inventarizace lesů České republiky (NIL-II)¹.

Dále se ve zprávě pokusíme upozornit na některá slabší místa, možné zdroje chyb a jejich pravděpodobné efekty. Zároveň jsme na řadě míst zařadili doporučení popisující, v kterých bodech by bylo dobré užívané postupy zlepšit.

Při práci na posudku jsme pracovali s následujícími podklady:

- Metodika odhadu cílových parametrů druhého cyklu Národní inventarizace lesů České republiky (NIL2), Ing. Radim Adolt, Ph.D., verze 18. března 2017.
- Metodika statistického vyhodnocení dat NIL1.
- Soubor `art_pop.qgs`, který umožňuje vizualizaci umělé populace kmenů.
- Zdrojové kódy knihovny `est4nfi`.
- Virtuální počítač s umělou populací kmenů (`artpop`).
- Virtuální počítač s vrstvou OLIL.
- Výsledky diskuse o projektu NIL-II, konané v Kroměříži 4. července 2017, jíž se zúčastnili jak autoři této zprávy, tak řada pracovníků ÚHÚL.

2. SBĚR DAT

Sběr dat probíhá na vybraných inventarizačních plochách přímo v terénu. Navštěvovány jsou ty plochy, které by mohly být zařazeny do kategorie pozemků typu *les*. Při terénním šetření i následném zpracování je velká pozornost věnována kontrole kvality měřených dat. Bylo vypracováno provádění víceúrovňových, pokud možno automatizovaných, kontrol měřených údajů nejenom přímo v terénu, ale i při jejich vkládání do databáze a následné validaci. Z našeho pohledu jsou sběr, jejich vkládání do databáze a následná víceúrovňová kontrola vyřešeny dobře. Líbil se nám též „*Management toku kvality dat pro fotogrammetrii*“, prezentovaný během naší návštěvy. Připomeňme dále, že požadavek, aby získaná data byla spolehlivá, je zásadně důležitý předpoklad pro jejich následné vyhodnocení metodami matematické statistiky, neboť bez spolehlivých dat nelze získat dostatečně přesné odhady.

Veškerá nasbíraná data jsou ukládána a spravována v SQL databázi. Z této databáze jsou pak předávána knihovně C++ `est4nfi`, pomocí které jsou prováděny vlastní výpočty

¹V celém dalším textu se snažíme rozlišovat síť NIL2 od druhé fáze inventarizace NIL-II.

odhadů cílových parametrů. Na virtuálním stroji jsme měli možnost vyzkoušet funkce knihovny `est4nfi` pro databázi se simulovanými výběry umělé populace kmenů. Na základě této zkušenosti můžeme navrženou databázi a funkční knihovnu pro výpočet odhadů pochválit. Jisté nedostatky však vidíme v dokumentaci, která navenek nepůsobí příliš přehledně.

2.1. Sítě NIL1 a NIL2. V rámci NIL-II probíhal sběr dat souběžně v inventarizačních sítích NIL1 a NIL2. Připomeňme, že síť NIL1, která obsahuje cca 14,5 tisíce inventarizačních ploch, byla založena při prvním cyklu národní inventarizace lesů v letech 2001 – 2004. Některé nevýhody designu/návrhu této inventarizační sítě vedly k založení nové sítě NIL2. K tomu, aby bylo možné vyhodnotit některé v čase se vyvíjející veličiny jako přírůst dříví, změna zásoby dříví, výše těžby, apod., nebylo možné v rámci NIL-II síť NIL1 zcela opustit, a tak v síti NIL1 bylo provedeno opakované šetření. Zároveň začalo šetření na inventarizačních plochách nové sítě NIL2, obsahující cca 8,5 tisíce nových inventarizačních ploch. To bylo rozděleno do několika sezón (jarních a podzimních), ve kterých probíhal sběr dat v podsítích, které pokrývají celé území ČR.

Mezi výhody sítě NIL2 patří, že umožní lépe kombinovat informaci z pozemního šetření s fotogrammetrickým šetřením založeným na datech z katastru, případně v budoucnosti ze satelitů (zatím nepoužíváno), neboť v současné době používané jednofázové odhady tak bude možné zpřesnit dvoufázovými odhady využívajícími další zdroje informací. To neznamena, že dvoufázové odhady by nebylo možno po vhodné adaptaci použít i pro síť NIL1. Neodstranilo by to však všechny nevýhody, které se během používání sítě NIL1 objevily. Volba sítě NIL2 je proto výhodná pro kontinuální vyhodnocování cyklů NIL.

Způsob generování bodů sítě NIL2 bohužel nemůžeme podrobněji posoudit, neboť jsme jako podklad dostali pouze virtuální počítač s příkladem inventarizační sítě, která však byla očividně vygenerovaná jiným způsobem.

2.2. Doporučení.

- (1) Vzhledem k tomu, že jsou v současnosti k dispozici měření jak ze sítě NIL1 (cca 14,5 tisíce inventarizačních ploch), tak ze sítě NIL2 (cca 8,5 tisíce nových inventarizačních ploch měřených paralelně s NIL1), nabízí se možnost provést pro kontrolu porovnání odhadů na základě již existujících dat paralelně naměřených v obou těchto sítích.
- (2) Pro budoucí cykly NIL se počítá s tím, že inventarizační plochy zůstanou neměnné (založené na síti NIL2), a sesbíraná data budou doplněna fotogrammetrickými daty² vyhodnocenými na velmi husté síti, což by mělo především umožnit získání

²Připomeňme, že podle informací ÚHÚL jsou fotogrammetrická data z katastru maximálně dva roky stará.

přesnějších odhadů měnících se veličin. *Domníváme se nicméně, že by bylo nejenom možné, ale i vhodné, obměňovat část sítě NIL2 při každé inventarizaci s cílem získání přesnějších odhadů stavu lesa.* Obměňování části inventarizační sítě by také umožnilo ověřit, že odhady některých parametrů nejsou ovlivněny (systematicky vychýleny) opakovanými návštěvami inventarizačních ploch.

- (3) Výhledově uvažovat o možnosti využívat rutinně i data ze satelitů (zatím nejsou používány), pokud by byla ve vhodném tvaru a za rozumnou cenu k dispozici.

3. VÝPOČET ODHADŮ

V materiálu popisujícím metodiku odhadu cílových parametrů NIL-II jsou definovány jednofázové odhady úhrnů a podílů. Vychází se z pojmu lokální hustoty, kterou navrhl Mandallaz (1991). V dokumentu je názorně vysvětleno, jakým způsobem se provádí kompenzace na okrajové efekty. Odhady úhrnu jsou geograficky i atributově aditivní, což umožňuje provádět kontrolu při jejich vyhodnocování. Odhady hlavních zjišťovaných parametrů (přírůst dříví, zásoba dříví a úbytek zásoby – těžba, plocha lesa, zastoupení dřevin) i odhady jejich rozptylů jsou počítány správně.

Na přání Ministerstva zemědělství České republiky se nicméně zastavme podrobněji u parametru *těžba*.

3.1. Těžba. Na základě existujících dat lze v současné době odhad těžby získat pouze v síti NIL1 spočtením vytěžených stromů. Tento údaj umí ÚHÚL získat velmi přesně. Ze studia podkladů a zevrubné diskuse v Kroměříži vyplynulo, že postup použitý v ÚHÚL je podle našeho názoru korektní a že získaný odhad odpovídá skutečnosti i přesto, že při výpočtu je nutné používat jisté aproximace a odhady. Na druhé straně je třeba si uvědomit, že:

- Odhad objemu dřeva u vytěženého stromu je v metodice ÚHÚL založen pouze na průměru a výšce naměřené v síti NIL1.
- Okamžik těžby není známý, takže použitý odhad přírůstu v polovině sledovaného období může mírně nadhodnocovat skutečnost.
- Z předložené dokumentace nám nebylo jasné, jak se při odhadu těžby započítávají inventarizační plochy, které se změnilly z položky *les* na *neles*. Z diskuse v Kroměříži však vyplynulo, že tyto plochy jsou započítány korektně.

Nejenom odbornou veřejnost, ale i MZe ČR a tisk, zaujal nesoulad statistik poskytovaných Lesy České republiky (LČR) a odhady získanými v rámci šetření NIL, jež se týkají těžby. Domníváme se, že tento nesoulad může být způsoben tím, že LČR při prodeji dřeva používají jinou metodiku, z níž je výše těžby stanovována, než ÚHÚL v rámci šetření NIL, takže, lidově řečeno, každá strana měří „něco trochu jiného“, a tím pádem i deklaruje „něco trochu jiného“.

Stále nám není jasné, zda dříví vykazované LČR jako prodané odpovídá veškeré těžbě, která se měří v rámci projektu NIL. Obáváme se, že tomu díky rozdílným metodikám tak není. Pokud by se měly tyto údaje v budoucnu smysluplně porovnávat, je třeba zajistit jejich srovnatelnost. Toho lze dosáhnout nejlépe tak, že ÚHÚL a LČR budou:

- Měřit stejné údaje a používat stejnou metodiku, nebo, alespoň;
- Na menším vzorku prozkoumají souvislosti a závislosti mezi odhady NIL s údaji vykázanými pro týž vzorek LČR, a dohodnou se na metodice pro přepočítání (kalibraci).

Možná opatření tedy jsou:

- (1) Sjednotit metodiku způsobu měření těžby na straně LČR podle NIL.
- (2) Vyjasnit případné rozdíly mezi používanými definicemi parametrů *celková těžba*, *prodané dřevo*, atd.
- (3) Měřit průměr a výšku stromů prodávaných LČR, což by mohlo zajistit srovnatelnost údajů ÚHÚL a LČR po korektním přepočtu pomocí vhodné kalibrační křivky. Při výpočtu těžby metodikou ÚHÚL i LČR v menších územních celcích (např. krajích) by šlo ověřit, zda je poměr získaných údajů stabilní nebo zda se údaje v některých krajích liší.
- (4) Začít pracovat na metodice měření dříví u odběratele, apod.

3.2. Připomínky k metodice. Používané jednofázové odhady získané z dat jsou matematicky popsány v dokumentu *Metodika odhadu cílových parametrů druhého cyklu Národní inventarizace lesů České republiky (NIL2)*. Tento materiál obsahuje potřebná teoretická odvození design-based odhadů cílových parametrů NIL. K jeho obsahu uvádíme upozornění na některé nepřesnosti a terminologické záležitosti.

- Ve vzorci (2) je x na levé straně rovnoměrně náhodně vybraný bod, vzhledem ke kterému se uvažuje střední hodnota, takže by se spíš hodilo psát jako X , aby se odlišil od integrační proměnné na pravé straně.

Podobných nesrovnalostí je v textu více a doporučujeme proto materiál systematicky projít a sjednotit značení. Například tak, že náhodné veličiny a vektory budou značeny velkými písmeny, odpovídající měření písmeny malými, odhady pomocí stříšek, apod.

- V prvním členu vzorců (15) a (19) má nejspíš vystupovat Y_i^* .
- Jak bylo podrobně diskutováno v Kroměříži, vypovídací schopnost obrázku 3 na straně 11 (konturový graf) a obrázků jemu podobných (jež pracovníci ÚHÚL prezentovali), je především na „okrajích“ značně diskutabilní. Doporučujeme jejich konstrukci přepracovat.
- V popisku obrázku 3 není jasné, co se myslí 95 procentním kvantilem chyby.

- U odhadu (23) by se mělo zmínit, jakým způsobem se získá očekávaný počet \bar{n}_{D+} .
- V (58) a (59) by parciální derivace neměly být podle \hat{Y}_1 a \hat{Y}_2 , nýbrž podle proměnných y_1 a y_2 funkce $f(y_1, y_2)$.
- Stálo by zato zmínit, jakým způsobem se odhaduje kovariance jednofázových odhadů, tedy jak se určí $\hat{C}(\hat{Y}_1, \hat{Y}_2)$ ve vzorcích (63) a (84).
- Na straně 32₁₃ má být (102).
- Vzorec (110) je „osamělý“ a potřebuje doplnit výkladem.
- Práce Vysochansky-Petuni není uvedena v seznamu literatury.
- Na straně 35 by intervalový odhad měl obsahovat θ ve $100(1 - \alpha)\%$ případů.
- Úpravy výrazů v přílohách jsou podrobně rozepsány. Někdy je to až zbytečně podrobně, odvození by se na některých místech dala provést jednodušeji. Na druhé straně, některé důležité vzorce autor přebíral z literatury, a bez hledání v ní je lze zkontrolovat pouze s obtížemi.
- Text by potřeboval řádnou jazykovou korekturu jak obecné, tak odborné češtiny. Tak například, ve statistice je zvykem užívat výraz *spolehlivost*, nikoliv *statistická jistota*. Podobně je lepší používat výraz *horní/dolní konfidenční mez*, apod.
- Text by též potřeboval na řadě míst korekturu typografickou. Tak například, text po vysazených vzorcích by neměl odskakovat, což lze nejjednodušeji zařídit systematickým používáním příkazu `\noindent`, apod.

4. KONTROLA ZÍSKANÝCH ODHADŮ

Je zjevné, že statistická analýza dat získaných při inventarizaci lesa je komplexní problém. Pro získání korektních odhadů je třeba:

- (1) Porozumět teorii, na základě které se počítají odhady potřebných parametrů.
- (2) Sesbírat potřebná data, provést jejich důkladnou kontrolu a uchovávat je v přehledném a přístupném formátu.
- (3) Naprogramovat teoreticky odvozené odhady.
- (4) Napočítat potřebné odhady ze sesbíraných dat.

Při praktické implementaci takto komplexního a rozsáhlého systému je vyloučení chyb velice komplikované. Asi jediným řešením je soustavná kontrola, která případné chyby umožní co nejdříve najít a odstranit:

- Pro teoretickou část se zdá být přínosná zejména spolupráce se zahraničními pracovišti, která se také zabývají inventarizací lesa.
- Sběr dat je v ÚHÚL průběžně kontrolován přímo v SQL databázi pomocí interního databázového systému MediaWiki.

V případě kontroly celého systému vyhodnocení dat může být užitečná počítačová simulace, tj. opakovaný výpočet odhadů pro náhodně generované populace stromů a porovnání nasimulovaného rozdělení získaných odhadů se „skutečností“ (která je u počítačově generovaných dat známá).

Ještě lepší kontrolu by bylo možno zajistit zveřejněním naměřených dat, neboť lze očekávat, že případné chyby v programech a/nebo metodických postupech by byly rychle objeveny, pokud by se získaná data začala používat při psaní kvalifikačních prací a odborných článků těmi, které daná problematika z jakéhokoliv důvodu zajímá, ale nepatří do řešitelského kolektivu NIL.

5. DOKUMENTACE

Postup použitý při výpočtu odhadů by měl být nejen správný, ale měl by být i kvalitně zdokumentován, aby bylo možné získané výsledky ověřit a správně interpretovat. Například zpracování tohoto posudku bylo poněkud komplikováno předloženou dokumentací, jejíž webovou část lze označit za nepřehlednou až chaotickou. Jak jsme viděli během diskuse v Kroměříži, vyznají se v ní pouze ti, kteří ji každodenně používají.

Ačkoliv jsme měli k dispozici větší množství materiálů, chyběl nám uspořádaný, systematický a jasný popis jednotlivých kroků NIL. Většina nejasností se nakonec vyjasnila až při naší červencové návštěvě ÚHÚL v Kroměříži, kde nám řešitelé ukázali, že *interně* jsou veškeré podklady poměrně přehledně uspořádány pomocí SQL a MediaWiki.

5.1. Doporučení.

- (1) Struktura dokumentace jak pro interní použití, tak pro odbornou veřejnost, a hledání v ní by měla být jasně vysvětlena již na začátku kořenové struktury, aby případný zájemce nemusel, tak jako nyní, používat metodu „řízeného náhodného hledání“.
- (2) V budoucnu by bylo vhodné zveřejnit získaná data, samozřejmě po rozumné anonymizaci a s výjimkou přesných souřadnic inventarizačních míst. Pro jejich případné využití však bude nutné výrazně vylepšit i veřejně přístupnou část dokumentace celého projektu. Jasně a stručně by mělo být popsáno:
 - Definice základních užívaných pojmů a jejich parametrizace. Předejde se tím například tomu, že ÚHÚL a Lesy České republiky jinak definují, každý po svém, tak klíčové pojmy jakými je těžba, dorost, vykácený les, apod.
 - Způsob generování bodů v sítích NIL1 a NIL2, a jejich použití v jednotlivých etapách vyhodnocení.
 - Metodika sběru dat a vlastní průběh inventarizace.
 - Popis naměřených dat (vyexportovaných z databáze SQL).

- Metodika statistického vyhodnocení a způsob výpočtu odhadů cílových parametrů včetně odhadu chyb.
- Praktický příklad ilustrující výpočet nejdůležitějších odhadů.

K jednotlivým bodům již alespoň částečná dokumentace existuje, ale měly by se vyjasnit i návaznosti jednotlivých kroků, a veškerá dokumentace by měla být přístupná na jednom místě.

6. RŮZNÉ

- (1) Pro další připravované národní inventarizace lesů ÚHÚL plánuje používat metodiku dvoufázových odhadů kombinujících data z měření „na místě“ v rámci sítě NIL2 s daty fotogrammetrickými, vyhodnocenými ve „velmi husté“ síti bodů. To by mělo vedle zvýšení přesnosti odhadů snížit náklady na výjezdy. Na druhé straně bude třeba za přístup k fotogrammetrickým datům patrně platit. V každém případě doporučujeme připravovanou metodiku nechat posoudit dříve než bude naprogramována v SQL. Předejde se tím řadě potenciálních problémů.
- (2) Uvážit pro budoucnost vedle fotogrammetrických dat větší využití dat satelitních. Samozřejmě za podmínky, že by byla ve vhodném tvaru a za rozumnou cenu k dispozici.

prof. RNDr. Jaromír Antoch, CSc.

doc. RNDr. Zdeněk Hlávka, Ph.D.

doc. RNDr. Zbyněk Pawlas, Ph.D.

Katedra pravděpodobnosti a matematické statistiky MFF UK

Praha 29. srpna 2017