

Metodika odhadu cílových parametrů druhého cyklu Národní inventarizace lesů České republiky (NIL2)

Jednofázové odhady na základě terénních
dat inventarizační sítě NIL1

(verze 18. března 2017)

ACNIL – č.j. UHUL/2476/2016/KM



Ing. Radim Adolt, Ph.D.

Ústav pro hospodářskou úpravu lesů Brandýs nad Labem

18. března 2017

Obsah

1	Úvod	4
2	Lokální hustota	4
2.1	Hustota pro odhad parametrů populace kmenů	5
2.2	Hustota pro odhad rozlohy území	11
2.3	Hustota pro atributové domény	11
2.4	Hustota na úrovni traktů sítě NIL1	12
3	Odhad úhrnu veličiny Y v zájmové oblasti D	13
4	Odhad podílu $R_{1,2}$ v zájmové oblasti D	14
	Metodické přílohy	19
A	Horwitz-Thompsonův teorém pro kontinua	20
A.1	Aplikace HTC pro URS	23
B	Odhad rozptylu pomocí Taylorovy aproximace	25
B.1	Rozptyl podílu, obecné řešení	26
B.2	Rozptyl podílu, design URS	28
C	Výběry inventarizačních bodů v projektech NIL	29
D	Odhady rozptylu v projektech NIL	30
E	Odhad úhrnu nestranný nepodmíněně na velikosti výběru	31
F	Geografická a atributová aditivita odhadu	34
G	Intervalový odhad a jeho statistická jistota	34

1 Úvod

Dokument předkládá metodiku jednofázových odhadů cílových parametrů projektu NIL2 (2011-2014), které jsou získávány na základě terénního šetření v inventarizační síti NIL1.

V první sekci je pojednáno o funkci lokální hustoty, která je nepostradatelnou součástí odhadu parametrů NIL prováděných pomocí Horwitz-Thompsonova teorému pro populace kontinua, viz příloha A. Popis lokálních hustot je omezen na výběrový protokol kmenů hroubí a bodový výběr. Tyto dva způsoby výběru jsou používány pro odhady úhrnů a hektarových hodnot populace kmenů (zásoba dříví celkem a na hektar porostní půdy, těžba dříví, přírůst apod.) a též pro odhad rozlohy kategorií pozemků či porostní půdy v rámci kategorie *Les*.

Následující dvě kapitoly jsou věnovány jednofázovému odhadu úhrnu a podílu. Hlavní část metodiky je doplněna přílohami, které osvětlují některé typické a podstané vlastnosti projektů NIL a poskytují odvození nejdůležitějších vztahů, z nichž se v hlavní části metodiky vychází.

V metodice jsou popsány pouze rutinně používané typy odhadů parametrů NIL. Text se zaměřuje téměř bez výhrad na design-based odhady parametrů NIL. Problematika z oblasti dendrometrie, nauky o produkci či hospodářské úpravy lesů není obsažena. Důvodem je omezený rozsah a udržení přehlednosti textu jako celku.

2 Lokální hustota

Cílem výběrového šetření Národní inventarizace lesů (NIL) je odhadnout úhrny, střední hodnoty respektive podíly úhrnů či středních hodnot ve zvolených populacích (základních souborech) a jejich částech (subpopulacích). Zvláštností NIL je velký počet základních souborů a požadavek na odhad mnoha jejich vlastností (parametrů). V předmětu zájmu NIL jsou zastoupeny jak populace diskrétních objektů, které jsou alespoň teoreticky spočítatelné (stromy, kusy mrtvého dříví), tak i populace, jejichž jednotlivé prvky je obtížné případně neopodstatněné rozlišovat. Příkladem je populace sestávající z teoreticky i prakticky nespočítatelného množství prvků - bezrozměrných lokalit, bodů, jejichž sjednocením je území s určitými vlastnostmi - například území pokryté lesem. V uvedeném příkladu jde o populaci kontinua - podmnožinu kartézského systému souřadnic \mathbb{R}^2 .

Z důvodu praktické proveditelnosti a racionality nákladů projektu je třeba hledat způsob, jak efektivně provést výběry jak z populací diskrétních objektů, tak i z populací kontinua, a na základě těchto výběrů relativně jednoduše, a pokud možno v principu stejným způsobem, vypočítat odhady cílových parametrů. Řešení spočívající v zavedení *funkce lokální hustoty* navrhl ve své dizertační práci Mandallaz [1991, sekce 3.1, od strany 18]. Základní vlastnost funkce lokální hustoty $Y(x)$ vyjadřuje rovnice (1):

$$\int_D Y(x)dx = Y, \quad (1)$$

kteřá říká, že pro uvažovanou veličinu je součet hodnot lokální hustoty na všech bodech x zájmové oblasti $D \subset \mathbb{R}^2$ (geograficky vymezené domény) roven hodnotě úhrnu Y této veličiny. Očekávaná hodnota lokální hustoty je rovna střední hodnotě veličiny v oblasti D , viz (2):

$$\mathbb{E}[Y(x)] = \frac{1}{\lambda(D)} \int_D Y(x)dx = \bar{Y}, \quad (2)$$

kde $\lambda(D)$ označuje velikost D .

Výběrové šetření lze pro každý projektem NIL uvažovaný cílový parametr navrhnout tak, abychom na základě získaných údajů mohli nadefinovat a pro každou lokalitu (inventarizační bod) vyhodnotit funkci lokální hustoty. Takto lze cílové parametry odhadovat pomocí jednotné teorie, kterou je Horwitz-Thompsonův teorém pro populace kontinua, viz příloha A od strany 20.

Vlastnostmi funkce lokální hustoty se ve své práci též podrobně zabýval Lanz [2000, kapitola 2 od strany 26]. Rozbor různých přístupů k design-based odhadu parametrů populace kmenů lze nalézt v práci Erikssonové [1995].

2.1 Lokální hustota pro odhad parametrů populace kmenů při výběru soustřednými kruhovými segmenty

V této sekci je využito formálního popisu výběrového protokolu kmenů, který publikoval Mandallaz [2007, sekce 4.2 od strany 55]. Při výběru kmenů¹ soustřednými kruhovými segmenty je na inventarizačních bodech vytyčeno několik (zpravidla dva nebo tři) soustředných kruhových segmentů označovaných K_r . Kmen i je vybrán pokud se nachází uvnitř kruhu $K_r(x) = \{y \in \mathbb{R}^2 \mid d(x, y) \leq r\}$ o poloměru r , se středem v inventarizačním bodě x . Poloha kmene je určena kolmým průmětem bodu ležícího na jeho středové ose ve výčetní výšce² do vodorovné roviny dané zvoleným souřadným systémem (S-JTSK). Označení $d(x, y)$ je použito pro Eukleidovskou vzdálenost mezi body x (střed kruhu, inventarizační bod) a y (bod náležící kruhu). Definujme náhodnou indikátorovou proměnnou $I_i(x)$ pro kmen i a inventarizační bod x :

$$I_i(x) = \begin{cases} 1 & \text{když } u_i \in K_r(x), \\ 0 & \text{když } u_i \notin K_r(x), \end{cases} \quad (3)$$

a dále N kruhů $K_i(r) = K_r(u_i)$ s konstantním poloměrem r a se středy u_i na polohách jednotlivých kmenů. Kmen i je vybrán právě tehdy a jen tehdy, když inventarizační bod x padne do kruhu $K_i(r)$:

$$I_i(x) = 1 \Leftrightarrow x \in K_i(r). \quad (4)$$

Každému kmeni i lze přiřadit kruh $K_i = K_{r_i}(u_i)$ (zónu zahrnutí do výběru), jehož poloměr závisí na vlastnostech kmene i - například na výčetní tloušťce (d_{13i}), výšce (h_i), dřevině, příslušnosti ke geograficky vymezené doméně apod. V projektech NIL závisí velikost kruhového segmentu K_i typicky na výčetních tloušťkách kmenů. V rámci inventarizační sítě NIL1 jsou pro kmeny hroubí ($d_{13i} \geq 7$ cm) používány dvě velikosti kruhových segmentů:

$$r_{iNIL2} = \begin{cases} 3.00 \text{ m} & \text{pro kmeny } 7 \text{ cm} \leq d_{13i} < 12 \text{ cm}, \\ 12.62 \text{ m} & \text{pro kmeny } d_{13i} \geq 12 \text{ cm}. \end{cases} \quad (5)$$

Pro náhodně umístěný bod x inventarizační sítě definujme lokální hustotu výběrového protokolu kmenů pomocí (6):

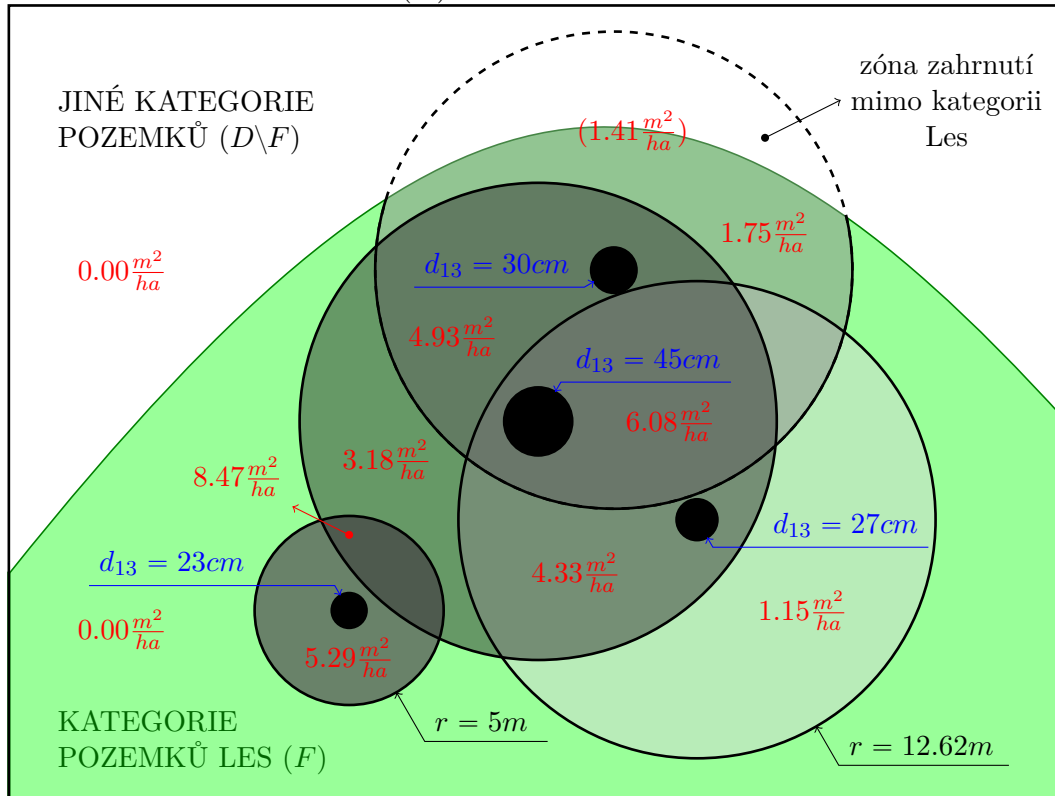
$$Y(x) = \sum_{i \in \mathcal{P}} \frac{I_i(x) Y_i}{\lambda(K_i)}, \quad (6)$$

kde $\lambda(K_i)$ je velikost kruhového segmentu K_i kmene i a \mathcal{P} je konečně velká populace kmenů vyskytujících se na přístupném a schůdném místě zájmové oblasti D a současně

¹Rozdíl mezi stromem a kmenem specifický pro pracovní postupy NIL2 je v tento okamžik nepodstatný.

²Výška 1.3 m měřená jako vzdálenost od průsečíku osy kmene a vodorovné roviny procházející bází kmene - při náklonu kmene tak měříme šikmou a vzdálenost.

ZÁJMOVÁ OBLAST -
GEOGRAFICKÁ DOMÉNA (D)



Obr. 1: Průběh funkce lokální hustoty (7) pro odhad úhrnu výčetní kruhové základny při výběru kmenů v inventarizační síti NIL2 (dva soustředné kruhy o poloměru $r_1 = 5 m$ a $r_2 = 12.62 m$ s rozhraními $d_{13, r_1} \geq 7 cm$ a $d_{13, r_2} \geq 27 cm$), situace poblíž okraje kategorie pozemků Les, hodnota $1.41 m^2 ha^{-1}$ uvedená v závorkách odpovídá nekompensované hustotě dle vzorce (6)

v rámci kategorie pozemků Les, dále označované F . Označení Y_i je použito pro hodnotu veličiny pozorované na kmeni i , jejíž úhrn za celou populaci kmenů \mathcal{P} potřebujeme odhadnout.

Takto definovaná hustota vyhovuje rovnici (1) (strana 4) pouze v případě, že jsou v terénu registrovány též kmeny vybrané některým z kruhových segmentů se středem v inventarizačním bodě mimo uvažovanou zájmovou oblast D (ČR) nebo též mimo kategorii pozemků Les, případně mimo přístupnou a schůdnou část území. Takový postup vyžaduje návštěvu, trvalou stabilizaci a provedení šetření na výrazně větším počtu inventarizačních bodů, a proto nebyl při šetřeních NIL1 ani NIL2 uplatněn. V rámci NIL1 a NIL2 byly registrovány pouze kmeny vybrané kruhovými segmenty se středem v inventarizačním bodě na území ČR, v rámci přístupné a schůdné části kategorie pozemků Les. Pro kmeny nacházející se blízko hranice státu, blízko okraje kategorie pozemků Les nebo blízko hranice přístupnosti a schůdnosti neodpovídá skutečná zóna zahrnutí kruhovému segmentu K_i . V obecném případě je dána průnikem $K_i \cap S \cap F \cap A$ (S označuje území státu, F kategorii pozemků Les a A přístupnou a schůdnou část území). To znamená, že ve jmenovateli (6) musí být pro kmeny na okraji ČR, kategorie Les či její přístupné a schůdné části uvažováno s menší zónou zahrnutí. Pausálním použitím plochy kruhového segmentu K_i v čitateli (6) vzniká tzv. okrajový efekt - podhodnocení odhadu celkového množství respektive střední hodnoty v zájmové oblasti D .

Podhodnocení může být zanedbatelné při použití malých kruhových segmentů v porovnání s průměrnými velikostmi souvislých území kategorie *Les*. V odborné literatuře je doporučováno okrajové efekty kompenzovat s upozorněním, že podhodnocení se svou velikostí může přiblížit velikosti výběrové chyby [Mandallaz, 2007, sekce 4.2, str. 60]. Vzhledem k technologické úrovni NIL2 (mapování poloh kmenů v S-JTSK, mapování okraje kategorie *Les* a nepřístupných částí území v dostatečně velkém okolí inventarizačního bodu) je použita metoda kompenzace okrajového efektu ('Measure π '), která vychází ze znalosti skutečné zóny zahrnutí všech v NIL2 vybraných kmenů [Gregoire & Valentine, 2008, sekce 7.5 na straně 224]. Lokální hustota na inventarizačním bodě může být počítána dle upraveného vzorce (7):

$$Y(x) = \sum_{i \in \mathcal{P}} \frac{I_i(x)Y_i}{\lambda(K_i \cap S \cap F \cap A)}. \quad (7)$$

Pro lokální hustotu definovanou (7) již platí (8):

$$\int_D Y(x)dx = \sum_{i \in \mathcal{P}} Y_i = Y = \lambda(D)\bar{Y}, \quad (8)$$

kde \bar{Y} je plošná střední hodnota veličiny v oblasti D o velikosti (plošné rozloze) $\lambda(D)$.

Použitím (7) v rámci teorie Horwitz-Thompsonova odhadu pro populace kontinua (viz příloha A od strany 20) lze získat nestranné odhady úhrnu různých veličin populace kmenů. Obrázek 1 schematicky zachycuje měřičskou situaci poblíž okraje kategorie pozemků *Les* včetně vizualizace hodnot lokální hustoty dle (7) pro výčetní kruhovou základnu (plocha průřezu kmene ve výčetní výšce).

Některé cílové parametry NIL vyžadují zjištění vlastností kmenů, které nejsou na stojících kmenech přímo měřitelné. Příkladem je zásoba dříví, jejíž odhad vychází ze zjištění objemu dřevní hmoty ve formě hroubí jednotlivých kmenů³. Rovnice (9) vyjadřuje způsob zjištění objemu hroubí daného kmene na základě predikce objemovými tabulkami:

$$Y_i = f_{ult}(d_{13i}, h_i). \quad (9)$$

Predikovaná hodnota Y_i je získána dosazením měřené výčetní tloušťky d_{13i} a výšky h_i do funkčního vztahu $f_{ult}(\cdot)$ objemových tabulek ÚLT pro danou dřevinu. V praxi naši i zahraničních NIL se chyba predikce objemu hroubí kmene objemovými tabulkami na základě *měřených* veličin (výčetní tloušťka, výška, případně další veličiny - nejčastěji horní tloušťka) zanedbává⁴ [Mandallaz, 2007, sekce 9.5, str. 166].

Měření výšky kmene je v běžných terénních podmínkách časově náročné. Proto byl vyvinut postup dvoustupňového odhadu, který je založen predikci pomocí výškové křivky:

$$h_i^* = f_h(d_{13i}), \quad (10)$$

kdy je jako prediktor používána měřená výčetní tloušťka d_{13i} . Na rozdíl od objemových tabulek se o modelu výškové křivky nepředpokládá, že je bez chyby. Chyba vznikající použitím predikce výšky namísto její měřené hodnoty se kompenzuje na základě *vzorníků druhého stupně*. Děje se tak rozšířením funkce lokální hustoty o korekční člen, viz druhý sčítanec následující rovnice:

³Objem veškeré nadzemní dřevní hmoty s tloušťkou dosahující alespoň 7 cm při měření s kůrou, nezačítává se objem pařezu.

⁴Při konsenzuálním používání konkrétních objemových tabulek jsou různé údaje na národní úrovni porovnatelné.

$$Y^*(x) = \sum_{i \in \mathcal{P}} \frac{I_i(x) Y_i^*}{\lambda(K_i \cap S \cap F \cap A)} + \sum_{i \in \mathcal{P}} \frac{\check{I}_i(x) (Y_i - Y_i^*)}{p_i * \lambda(K_i \cap S \cap F \cap A)}, \quad (11)$$

kteřý představuje odhad chyby vlivem použití modelu výškové křivky tj. odhad rozdílu $E = Y - Y^*$. První člen (11) je nestranným odhadem úhrnu Y^* zásoby dle objemových tabulek s dosazením predikovaných hodnot výšek kmenů h_i^* . Výraz $Y_i - Y_i^* = E_i$ je na konkrétním kmene zjištěný rozdíl „skutečné“ hodnoty objemu hroubí kmene (dle objemových tabulek s dosazením měřených hodnot výšky h_i) a modelové hodnoty objemu (dle objemových tabulek při dosazení predikované výšky kmene h_i^*). Rozdíly E_i jsou známé pouze na vzornících druhého stupně (změřena výška kmene). Člen $\check{I}_i(x)$ je indikátorovou proměnnou výběru vzorníků druhého stupně, pro který platí:

$$\check{I}_i(x) = 1 \Rightarrow I_i(x) = 1 \quad (12)$$

Symbol p_i označuje pravděpodobnost výběru na inventarizační ploše zaregistrovaného kmene jako vzorníku druhého stupně, na kterém se změří výška a odvodí zásoba hroubí dle objemových tabulek. Očekávaná (střední) hodnota indikátorové proměnné $\check{I}_i(x)$ podmíněně na výběru kmene jako vzorníku prvního stupně (registrace kmenů kruhovými segmenty) je rovna pravděpodobnosti p_i tj. platí:

$$\mathbb{E}[\check{I}_i(x) | I_i(x) = 1] = p_i. \quad (13)$$

Definice generalizované lokální hustoty dle (11) je převzatá a upravená definice podle Mandallaze [2007, sekce 4.4, od strany 69], který též uvádí důkaz shody její očekávané hodnoty se střední hodnotou v D :

$$\mathbb{E}[Y^*(x)] = \bar{Y}. \quad (14)$$

Rovnice (14) platí pokud je model výškové křivky (10) *externí*, což znamená, že nebyl parametrizován s použitím vzorníků NIL. První člen (11) je pro pevně daný bod $x \in D$ a za předpokladu použití externího modelu konstantou - mezi replikacemi výběru vzorníků druhého stupně se hodnota prvního členu nemění. Druhý člen (11) se s opakováním výběru vzorníků druhého stupně náhodně mění. Jeho očekávaná hodnota se shoduje s chybou E odhadu zásoby dříví způsobenou dosazením predikovaných výšek h_i^* do (9) namísto výšek měřených.

[Mandallaz, 2007, sekce 9.5, str. 165] ve své monografii prezentoval výsledek, podle kterého je optimální, aby pravděpodobnost p_i výběru vzorníků druhého stupně byla proporcionální chybě predikce objemu hroubí kmene. Na základě vlastních analýz jsme věděli, že tato chyba vzrůstá přibližně v lineárním vztahu s kruhovou základnou kmene (plocha průřezu kmene ve výčetní výšce). Proto jsme se rozhodli vybírat vzorníky druhého stupně relaskopickou metodou - výběr s pravděpodobnostmi, lépe řečeno s velikostí zóny zahrnutí, proporcionální výčetní kruhové základně kmene. Těto volbě odpovídá definice generalizované lokální hustoty, viz následující rovnice:

$$Y_{BAF}^*(x) = \sum_{i \in s_1(x)} \frac{Y_i}{\lambda(K_i \cap S \cap F \cap A)} + \sum_{i \in s_{1,2}(x)} \frac{(Y_i - Y_i^*)}{\lambda(K_{BAF,i} \cap S \cap F \cap A)}. \quad (15)$$

Symbol $\lambda(K_{BAF,i})$ je velikost zóny zahrnutí odpovídající relaskopické metodě se zvoleným faktorem BAF (Basal Area Factor). Poloměr zóny zahrnutí vzorníků druhého stupně byl při vyhodnocení NIL2 počítán dle vzorců

$$r_i = \begin{cases} d_{13i}/(2\sqrt{BAF}) \text{ m} & \text{pro kmeny } 7 \text{ cm} \leq d_{13i} < 87.4 \text{ cm} \\ 12.6156626101008 \text{ m} & \text{pro kmeny } d_{13i} \geq 87.4 \text{ cm}, \end{cases} \quad (16)$$

kde $BAF = 10^4 \sin^2(\alpha/2)$ je *relaskopický faktor* v m^2/ha a α je *záměrný úhel relaskopické pomůcky* ve stupních. Hodnota použitého relaskopického faktoru byla $12 \text{ m}^2/ha$. Symbol $s_1(x)$ označuje vzorníky prvního stupně tj. kmeny registrované kruhovými segmenty dle (5) na inventarizační ploše se středem v bodě x . Vzorníky druhého stupně vybrané kruhovými segmenty podle (16) na ploše se středem v bodě x jsou označeny $s_{1,2}(x)$.

Na rozdíl od (11) se druhý člen (15) s replikacemi výběrového designu nemění (za předpokladu použití externího modelu výškové křivky). Volbou výběrové lokality $x \in D$ dochází k současnému určení vzorníků prvního i druhého stupně. Při použití externího modelu výškové křivky je druhý člen (15) odhadem úhrnu zvláštního typu lokální hustoty (funkčního povrchu v rámci D), který je korekcí úhrnu zásoby dle objemových tabulek s dosazením modelových výšek kmenů na úhrn zásoby dle týchž tabulek, avšak s dosazením měřených výšek. Tato korekční lokální hustota se mezi replikacemi výběrového designu nemění, což je výhodné při odhadu změny zásoby a přírůstu dříví. Na dané inventarizační ploše je změna druhého členu (15) mezi inventarizačními cykly možná pouze při změně výčtu jedinců populace kmenů nebo jejich dimenzí. Naproti tomu, druhý člen (11) se v dalším inventarizačním cyklu mění navíc i vlivem odlišné volby vzorníků druhého stupně. Tím dochází k navýšení rozptylu odhadu změny zásoby a přírůstu dříví v porovnání s definicí lokální hustoty dle (15).

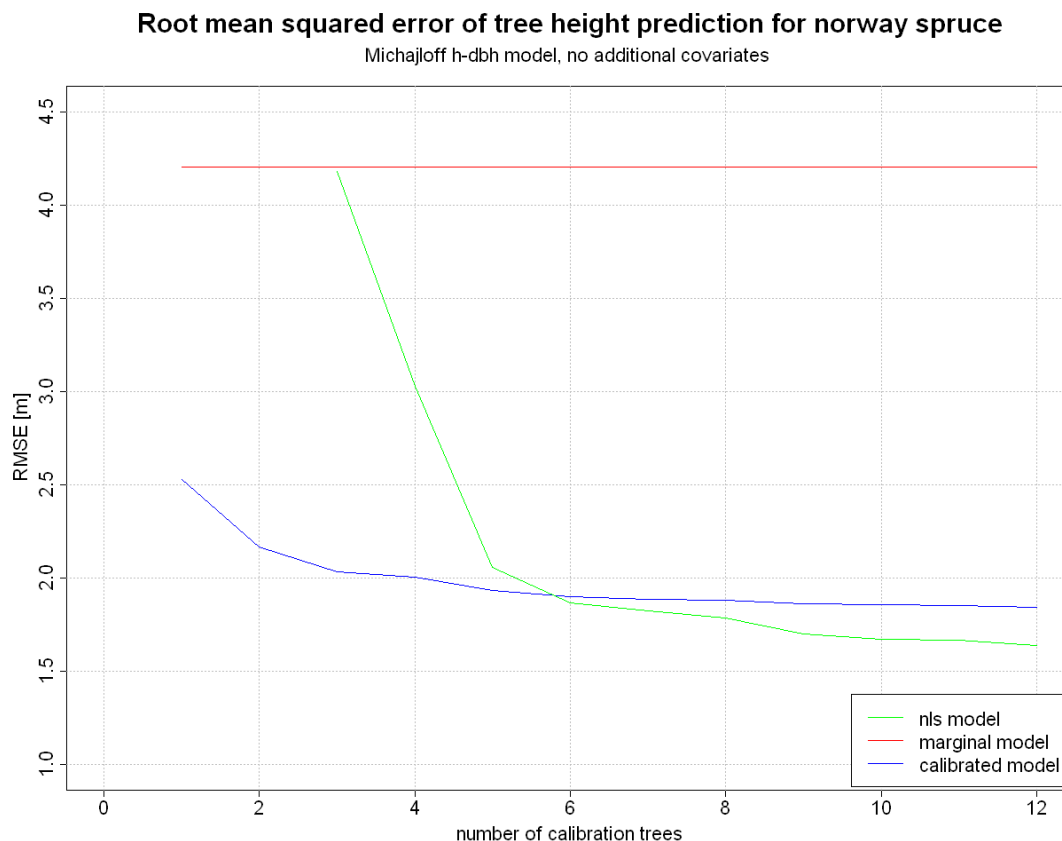
Konkrétní postup odhadu zásoby dříví NIL2 a dalších charakteristik založených na zjištění objemu hroubí jednotlivých kmenů (těžba dříví, přírůst, změna zásoby, mortalita dříví) používá hustotu dle (15) s relaskopickým faktorem $12 \text{ m}^2/ha$ a zónami zahrnutí pro vzorníky druhého stupně dle (16). Použitou výškovou funkcí je Michajlovova křivka:

$$h_i^* = a \cdot e^{-\frac{b}{d_{13i}}} + 1.3, \quad (17)$$

na jejímž základě byl odvozen nelineární model se smíšenými efekty pro normální rozdělení chyb. Náhodné efekty byly uvažovány pro oba parametry Michajlovovy funkce (a i b) na úrovni porostního segmentu⁵. Model byl parametrizován pro naše nejzastoupenější dřeviny na základě dat inventarizačního cyklu NIL1. Predikce pro dřeviny, pro které nebyl model parametrizován, se provedla dle modelu pro dřevinu s nejpodobnějšími ekologickým nároky a růstovými vlastnostmi.

Aplikace modelu při vyhodnocení NIL2 zahrnovala kalibraci výškové křivky - odhad hodnot náhodných efektů na základě jednoho *vzorníku kalibrace výškové křivky* vybraného za každou dřevinu a porostní segment na každé inventarizační ploše. Vzorníky kalibrace výškové křivky byly vybírány zcela náhodně ze seznamu kmenů dané dřeviny a porostního segmentu zaregistrovaných na inventarizační ploše (vzorníky prvního stupně). Výběr nijak nezávisel na výběru vzorníků druhého stupně pomocí relaskopické metody. V případě, že se na inventarizační ploše nenacházel žádný vhodný kandidát (např. jediný vzorník dřeviny v porostním segmentu byl silně nakloněný), byl vzorník pro kalibraci vybrán v nejbližším okolí inventarizační plochy - přednostně v rámci téhož porostního segmentu. Vždy se však muselo jednat o vzorník srovnatelného věku rostoucí ve shodných růstových podmínkách. Nebyl-li žádný kalibrační vzorník nalezen ani v blízkém okolí inventarizační plochy, použila

⁵Porostní segment je z lesnického hlediska homogenní část lesa se specifickou vnitřní strukturou vyjádřenou dřevinnou skladbou, věkovou strukturou, prostorovým uspořádáním jedinců, jejich původem a rámcovým způsobem hospodaření.



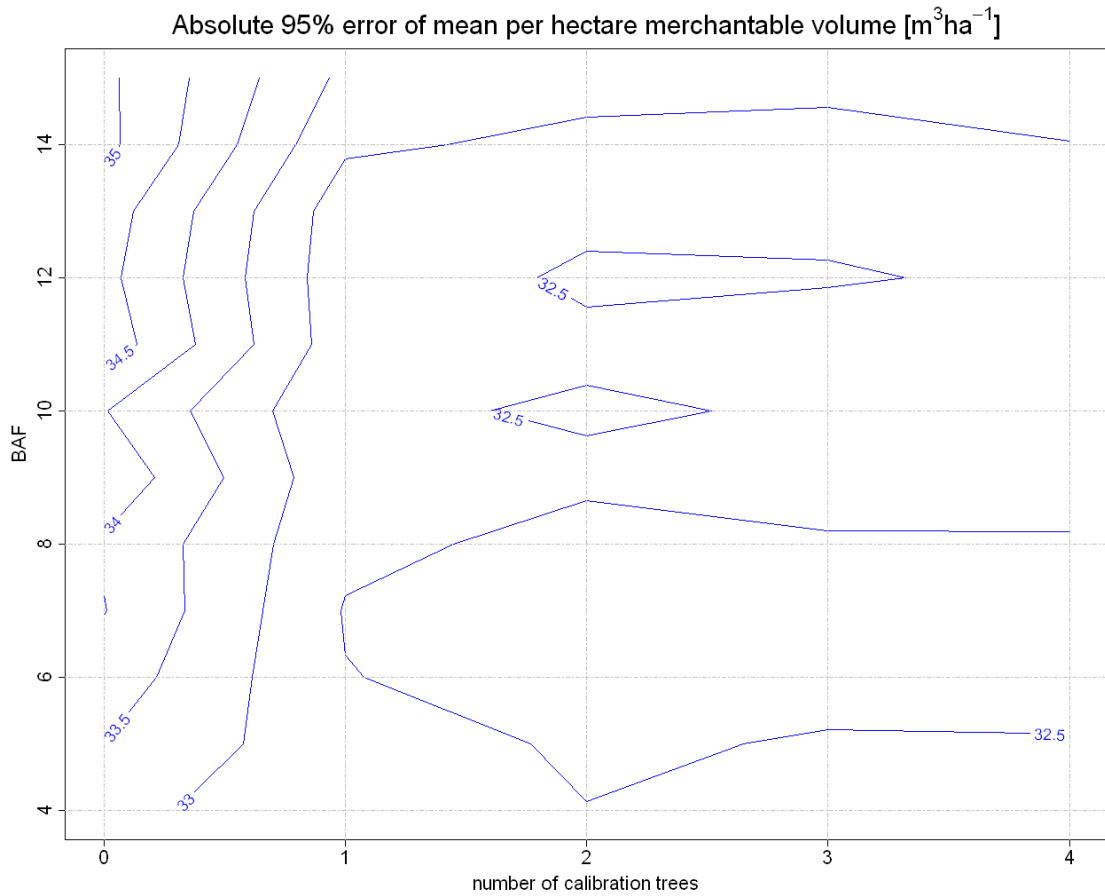
Obr. 2: Krosvalidační vyhodnocení variability chyb predikce klasického modelu Michajlovovy výškové křivky a modelu se smíšenými efekty.

se pro odvození objemu hroubí kmene marginální predikce (nulová hodnota náhodných efektů obou parametrů, populačně „průměrný“ model).

Vlastní kalibrace byla provedena postupem založeným na Taylorově aproximaci, který popisují [Trincado & Burkhart \[2006, vzorec 9, str. 674\]](#). Iterativní a přesnější postup kalibrace nelineárního smíšeného modelu uvádí [Bohora & Cao \[2014, vzorec 3, str. 63\]](#).

Z nomogramu [3](#) je patrné, že již při výběru jednoho kalibračního vzorníku za dřevinu a porostní segment, dojde k výraznému snížení vlivu BAF na velikost výběrové chyby. Chyba prakticky neklesá ani se zvyšujícím se počtem vzorníků kalibrace, což odpovídá vývoji směrodatné odchylky chyb modelu Michajlovovy křivky s měnícím se počtem kalibračních měření, viz [graf 2](#).

Volba $BAF = 12 \text{ m}^2/\text{ha}$ spolu s jedním kalibračním vzorníkem výškové křivky za každý porostní segment a každou dřevinu v něm zastoupenou vede k průměrnému počtu 5.31 měřených výšek na inventarizační ploše (výpočet na základě již sebraných dat uložených v analytické databázi). Tento počet se shoduje s dřívějším odhadem získaným v rámci pilotního projektu NIL2 (přírodní lesní oblast Středomoravské Karpaty, rok 2010).



Obr. 3: Nomogram znázorňující 95% kvantil chyby odhadu hektarové zásoby dříví jako funkci BAF výběru vzorníků druhého stupně a počtu nezávisle vybíraných vzorníků pro kalibraci výškové křivky (pilotní projekt NIL2 v přírodní lesní oblasti Středomoravské Karpaty, 2010).

2.2 Lokální hustota pro odhad rozlohy území

Při odhadu rozlohy území dané kategorie je lokální hustota rovna indikátorové proměnné $I_K(x)$ vyjadřující příslušnost inventarizačního bodu x ke kategorii K , viz rovnice (18).

$$I_K(x) = \begin{cases} 1 & \text{když } x \in K \\ 0 & \text{když } x \notin K \end{cases} \quad (18)$$

Pro různé kategorie území jsou indikátorové proměnné definovány analogicky.

2.3 Lokální hustota pro atributové domény

Odhady cílových parametrů jsou prakticky vždy členěny dle tzv. atributových domén⁶. Příkladem může být odhad úhrnu zásoby dříví členěný dle skupin dřevin. Za atributové domény označujeme části populací kterým nelze přiřadit určitou část zájmového území

⁶Kromě atributových domén rozeznáváme i domény geografické, které představují konkrétní část zájmového území D , jejíž rozlohu přesně známe. Příkladem geografických domén mohou být například kraje (NUTS3 jednotky) nebo přírodní lesní obvody. Odhady NIL jsou poskytovány pro kombinace geografických a atributových členění.

o známé rozloze. Příkladem atributové domény může být „porostní půda“, která představuje podkategorii kategorie pozemků *Les*. Neexistuje absolutně přesná mapa všech lesů a neznáme velikost území pokrytého lesem. Totéž platí i pro porostní půdu.

Odhad pro atributovou doménu je založen na úpravě lokálních hustot vynulováním příspěvků těch prvků populace, které nepatří do dané atributové domény (subpopulace). Pro názornost uveďme definici lokální hustoty (15) upravenou pro odhad zásoby dříví atributové domény „duby - všechny druhy“:

$$Y_{dub}^*(x) = \sum_{i \in s_1(x)} \frac{I_{dub,i}(x)Y_i}{\lambda(K_i \cap S \cap F \cap A)} + \sum_{i \in s_{1,2}(x)} \frac{I_{dub,i}(x)(Y_i - Y_i^*)}{\lambda(K_{BAF,i} \cap S \cap F \cap A)}. \quad (19)$$

Indikátorová proměnná $I_{dub,i}$ nabývá hodnoty 1 v případě, že kmen i je dub, jinak je nulová.

Jako další příklad uvádíme lokální hustotu pro odhad rozlohy kategorie pozemků *Les*, podkategorie porostní půda. Jedná se zde o součin dvou indikátorových proměnných:

$$I_{F,pp}(x) = I_F(x)I_{pp}(x), \quad (20)$$

kde $I_F(x)$ je indikátorová proměnná pro kategorii pozemků *Les* a $I_{pp}(x)$ je indikátorová proměnná pro porostní půdu (vylišována též v rámci kategorie pozemku OWL).

2.4 Lokální hustota na úrovni traktů sítě NIL1

Výběrové lokality (inventarizační body) jsou v inventarizační síti NIL1 uspořádány po dvojicích tj. v tzv. traktech neboli duplexech. Vzdálenost mezi inventarizačními body každého traktu činí nominálně 300 m (vzdálenost odpovídající 2D Eukleidovské vzdálenosti mezi generovanými souřadnicemi inventarizačních bodů téhož traktu). Orientace spojnice inventarizačních bodů traktu je náhodná. Podrobný popis inventarizační sítě NIL1 je uveden v monografii k projektu NIL1 [Jankovská & Štěrba, 2007, sekce 4.1, od str. 51].

Jednotlivé inventarizační body (plochy) traktu nejsou z hlediska informací na nich zjištěných vzájemně nezávislé. Hodnoty mnoha zjišťovaných veličin pozorované na inventarizačních plochách v rámci jednoho traktu jsou si v průměru více podobné, než mezi náhodně zvolenými plochami bez ohledu na jejich příslušnost k traktům.

Pro libovolně definovanou lokální hustotu vyhovující (1) a (2) je hustota na úrovni traktu sítě NIL1 definována následovně:

$$Y_{tct}(x) = \frac{Y(x_1)I_D(x_1) + Y(x_2)I_D(x_2)}{2}. \quad (21)$$

Symbol x v rámci $Y_{tct}(x)$ odpovídá vztažnému bodu traktu, za který budeme považovat první z dvojice inventarizačních bodů x_1 - v pravé části rovnice je lokální hustota zjištěná na tomto bodě označena $Y(x_1)$. Členy $I_D(x_1)$ a $I_D(x_2)$ jsou indikátorovými proměnnými, které nabývají pro příslušný bod traktu (x_1 nebo x_2) hodnotu 1 za předpokladu, že se daný inventarizační bod nachází uvnitř zájmového území (geografické domény) D . Pokud se bod nachází mimo D je hodnota indikátorové proměnné rovna 0.

Mandallaz [2007, sekce 4.3, od str. 65] publikoval postup jednofázového odhadu pro design používající trakty s pevně danou geometrií, který lokální hustotu na úrovni traktu definuje odlišně. Tento alternativní postup nebyl použit z důvodu náhodné orientace traktů v síti NIL1. Další informace k tomuto tématu lze nalézt v sekci 3.

3 Odhad úhrnu veličiny Y v zájmové oblasti D

Jednofázové odhad úhrnu Y v D (geografická doména, zájmová oblast o známé velikosti) je dle této metodiky počítán

$$\hat{Y} = \lambda(c) \sum_{\substack{x \in s \\ x \in D^+}} Y_{tct}(x), \quad (22)$$

kde $\lambda(c)$ je velikost inventarizačního čtverce sítě NIL1 (2 km×2 km tj. 400 ha). Označení $Y_{tct}(x)$ je použito pro lokální hustotu na úrovni traktu dle (21). Do výpočtu vstupují všechny trakty, jejichž výchozí inventarizační bod $x = x_1$ se nachází v zájmové oblasti rozšířené všemi směry o 300 m (obalová zóna o šířce 300 m po celém obvodu D). Tato oblast je označena D^+ .

Volba konkrétního odhadu úhrnu je diskutována v závěru přílohy C. Odhad (22) má následující vlastnosti:

- je nestranný nepodmíněně na počtu traktů s výchozím bodem v D^+ , viz příloha E
- je geograficky i atributově aditivní, viz příloha F

Odhad rozptylu (22) je počítán za pomoci následujícího vzorce:

$$\hat{V}(\hat{Y}) = \frac{\lambda^2(D^+)}{\bar{n}_{D^+} - 1 + \hat{V}(n_{D^+})/\bar{n}_{D^+}} \left\{ \widehat{Y_{tct}^2} + \hat{Y}_{tct}^2 \left[\hat{V}(n_{D^+})/\bar{n}_{D^+} - 1 \right] \right\}, \quad (23)$$

kde $\lambda(D^+)$ je velikost oblasti D^+ a \bar{n}_{D^+} značí očekávaný (střední) počet inventarizačních traktů s výchozím bodem v D^+ . Člen $\hat{V}(n_{D^+})$ je odhad variability počtu traktů s výchozím bodem v D^+ , který byl pro každou oblast získán na základě velkého počtu (2000) simulovaných výběrů v inventarizační síti NIL1. Symbol $\widehat{Y_{tct}^2}$ je podíl sumy čtverců lokální hustoty úrovně inventarizačních traktů a \bar{n}_{D^+} . Člen \hat{Y}_{tct}^2 označuje druhou mocninu podílu sumy lokálních hustot úrovně traktů a \bar{n}_{D^+} . Do výpočtu obou členů vstupují všechny trakty s výchozím bodem v D^+ (a žádné jiné).

Odhad rozptylu dle (23) je konzervativní, viz příloha D. Jeho odvození je provedeno v příloze (E). Jednofázový odhad úhrnu a jeho rozptyl metodicky vychází z Horwitz-Thompsonova teorému pro populace kontinua (HTC), který je popsán v příloze A.

Poznámky

1. Poněvadž jsou inventarizační body sítě NIL1 uspořádány v náhodně rotovaných traktech, používáme zvláštní definici funkce lokální hustoty. Ta je aritmetickým průměrem lokálních hustot úrovně inventarizační plochy vždy ze dvou bodů (inventarizačních ploch) traktu, viz sekce 2.4. Do výpočtu vstupují všechny inventarizační body, všech traktů, jejichž vztažený bod (první z dvojice) se nachází uvnitř obalové zóny D^+ vymezené rozšířením zájmové oblasti D od jejího obvodu o 300 m (pod všemi možnými úhly ale směrem vně D). Lokální hustota na inventarizačních bodech $x \in D^+, x \notin D$ (v rámci obalové zóny mimo D) je při výpočtu aritmetického průměru hustot za daný trakt nastavena na nulovou hodnotu. Díky použití obalové zóny není nutno provádět kompenzace okrajového efektu úrovně traktů - okrajový efekt nevzniká, poněvadž každý kmen (strom) může být vybrán buďto první nebo druhou plochou traktu (v závislosti na rotaci).

2. Pro design používající trakty (avšak s pevně danou geometrií - bez náhodné rotace traktu) definuje Mandallaz [2007, sekce 4.3, od str. 65] lokální hustotu na úrovni traktu odlišně. Námí zvolený postup vychází z faktu, že výběr sítí NIL1 neprobíhá pouze v rovině kartézských souřadnic S-JTSK, ale je rozšířen o úhel rotace mezi prvním a druhým bodem traktu. Podmíněně na orientaci traktu jsou zónami zahrnutí jednotlivých kmenů dva kruhy s velikostí odpovídající výčetní tloušťce. Nominální vzdálenost mezi středy kruhů činí 300 m, s danou (ale mezi replikacemi proměnlivou) orientací spojnice⁷. Funkce lokální hustoty tedy není definována pouze v rovině, ale ve trojrozměrném prostoru⁸. Řezy kolmo na souřadnou osu odpovídající úhlu rotace jsou roviny, v nichž je definována „klasická“ lokální hustota (pro konkrétní úhel rotace).

4 Odhad podílu $R_{1,2}$ v zájmové oblasti D

Podíl dvou úhrnů Y_1 a Y_2 v zájmové oblasti je dle této metodiky odhadován na základě

$$\hat{R}_{1,2} = \frac{\hat{Y}_1}{\hat{Y}_2} = \frac{\sum_{x \in A^+}^{x \in S} Y_{tct,1}(x)}{\sum_{x \in A^+}^{x \in S} Y_{tct,2}(x)}, \quad (24)$$

kde \hat{Y}_1 , \hat{Y}_2 jsou odhady úhrnů definujících samotný podíl a $Y_{tct,1}(x)$ a $Y_{tct,2}(x)$ jsou lokální hustoty na úrovni traktu počítané na principu (21), viz strana 12. Mezi čitatelem a jmenovatelem zcela vpravo (24) došlo k vykrácení výběrové váhy⁹.

Do výpočtu odhadu podílu dle vzorce (24) vstupují pouze trakty s kterýmkoli z inventarizačních bodů v oblasti A , představující rozšířenou atributovou doménu neznámé velikosti s následující definicí:

$$A = \left\{ x \in D \mid Y_1(x) \neq 0 \vee Y_2(x) \neq 0 \right\}. \quad (25)$$

Jedná se tedy pouze o trakty s výchozími body v oblasti $A^+ \subseteq D^+$, která vznikne rozšířením A vnějším směrem o obalovou zónu širokou 300 m.

Rozptyl (24) odhadneme dle vzorce

$$\hat{V}_{URS}(\hat{R}_{1,2}) = \frac{n_{A^+}^2}{(n_{A^+} - 1) \left[\sum_{x \in A^+}^{x \in S} Y_{tct,2}(x) \right]^2} \left[\widehat{Z}_{tct}^2 - \hat{Z}_{tct}^2 \right], \quad (26)$$

kde \widehat{Z}_{tct}^2 a \hat{Z}_{tct}^2 jsou aritmetický průměr čtverce a čtverec aritmetického průměru reziduální proměnné $Z_{tct}(x) = Y_{tct,1} - \hat{R}_{1,2} Y_{tct,2}(x)$ spočítané za všechny trakty s výchozím bodem v A^+ .

Odhad rozptylu (26) je odvozen z Taylorovy aproximace, viz přílohy B, B.1 a zvláště B.2. Odhad $\hat{R}_{1,2}$ není nestranným odhadem $R_{1,2}$. Pro v praxi NIL běžné velikosti výběrů (od několika desítek až stovek výše) je vychýlení $\hat{R}_{1,2}$ zanedbatelné.

⁷Zóna zahrnutí nepodmíněná úhlem rotace traktu má geometrii dvou paralelně probíhajících, vřetenovitě stočených potrubí.

⁸Trojrozměrná je už i množina všech výběrů (angl. *sample space*) o velikosti $n_{D^+} = 1$.

⁹Pokud je použit stejný typ odhadu v čitateli i jmenovateli podílu, nezáleží na konkrétní volbě výběrových vah.

Poznámky

1. Typickým příkladem odhadu podílu na základě dat sítě NIL1 je odhad hektarové zásoby dříví - podíl úhrnu celkové zásoby a rozlohy porostní půdy (oba úhrny je nutno odhadovat). Dle definice A^+ do výpočtu vstupují též trakty nezasahující do porostní půdy, za předpokladu, že lokální hustota úhrnu zásoby je na některé z ploch traktu nenulová. Tímto postupem se zamezuje vzniku okrajového efektu - nedochází ke snížení možnosti výběru kmenů v blízkosti okraje porostní půdy¹⁰.
2. Neznalost přesné rozlohy A^+ není překážkou. Musíme však být schopni rozhodnout o příslušnosti traktů k A^+ . Existují však situace, kdy žádný z inventarizačních bodů konkrétního traktu nezasahuje do A , ačkoli při jiné orientaci traktu by do A zasáhl druhý bod. Vynechání těchto traktů¹¹ nemá vliv na výpočet bodového odhadu. Odhad rozptylu vychází dle vzorce (26) poněkud vyšší - při velikostech výběru od několika desítek traktů je navýšení odhadu rozptylu zanedbatelné.
3. Pro odhad rozptylu je použita aproximace párových hustot $\pi(x_i, x_j)$ hustotou odpovídající zcela náhodnému výběru n_A^+ inventarizačních bodů v A^+ (URS). Prostorová korelace reziduální hustoty $Z_{tct}(x)$ bývá nižší, než v případě hustot $Y_{tct,1}(a)$ a $Y_{tct,2}(x)$, z nichž je počítána. Proto předpokládáme (a provedené simulace potvrzují) nižší míru nadhodnocení (konzervativnost) rozptylu $\hat{R}_{1,2}$ v porovnání s klasickým úhrnem \hat{Y} .
4. Za předpokladu, že hustota $Y_{tct,2}(x)$ je indikátorovou proměnnou, a je-li vlastnostmi výběru či charakterem populací dáno $Y_{tct,2}(x) = 0 \implies Y_{tct,1}(x) = 0$, přechází odhady $\hat{R}_{1,2}$ a $\hat{V}_{URS}(\hat{R}_{1,2})$ do podoby odpovídající odhadu aritmetického průměru a jeho rozptylu.

¹⁰Na rozdíl od rozhraní přístupnosti a schůdnosti, a na rozdíl od kategorie pozemků *Les se porostní půda* v dostatečně velkém okolí inventarizačního bodu při terénních pracích NIL nemapuje.

¹¹Ve skutečnosti tyto trakty patří do A^+ , avšak hodnota lokálních hustot úrovně traktu je na nich nulová.

Literatura

- Anděl, J. 1978. *Matematická statistika*. Praha: SNTL - Státní nakladatelství technické literatury.
- Barabesi, L. 2003. A monte carlo integration approach to horvitz-thompson estimation in replicated environmental designs. *International journal of statistics*, **LXI**, 355–374.
- Barabesi, L., & Franceschi, S. 2010. Sampling properties of spatial total estimators under tessellation stratified designs. *Environmetrics*, (**on-line publication**).
- Bohara, B., S., & Cao, V., Q. 2014. Prediction of tree diameter growth using quantile regression and mixed effects models. *Forest ecology and management*, 62–66.
- Cochran, W. G. 1977. *Sampling techniques*. Willey series in probability and mathematical statistics - applied. John Willey & Sons.
- Cooper, C. 2006. Sampling and variance estimation on continuous domains. *Environmetrics*, **17**, 539–553.
- Cordy, C., B., & Thompson, C., M. 1995. An application of the deterministic variogram to design-based variance estimation. *Mathematical geology*, **27**, 173–205.
- Cordy, C. B. 1993. An extension of the horvitz-thompson theorem to point sampling from a continuous universe. *Statistics and probability letters*, **18**, 353–362.
- Eriksson, M. 1995. Design-based approaches to horizontal-point-sampling. *Forest science*, **41**(4), pp 890–907.
- Gregoire, T. G., & Valentine, H. T. 2008. *Sampling strategies for natural resources and the environment*. Chapman and Hall/CRC.
- Heikkinen, J. (eds. Kangas A. & Maltamo M.). 2006. *Forest inventory, methodology and applications*. Springer Verlag. Chap. Assessment of Uncertainty in Spatially Systematic Sampling, pages 155–176.
- Jankovská, Z., & Štěrbá, P. (eds). 2007. *National forest inventory in the czech republic 2001-2004: Introduction, methods, results*. Ústav pro hospodářskou úpravu lesů Brandýs nad Labem (ÚHÚL).
- Jarník, V. 1963. *Diferenciální počet i*. Nakladatelství Československé akademie věd.
- Lanz, A. 2000. *Optimal sample design for extensive forest inventories*. Ph.D. thesis, Swiss Federal Institute of Technology (ETH), Zurich.
- Li, Y. 2008. Generalized regression estimators of a finite population total using box-cox technique. *Survey methodology*, **34** (1), 79–89.
- Mandallaz, D. 1991. *A unified approach to sampling theory for forest inventory based on infinite population and superpopulation models*. Ph.D. thesis, Swiss Federal Institute of Technology (ETH), Zurich.
- Mandallaz, D. 2007. *Sampling techniques for forest inventories*. Chapman and Hall/CRC.
- Meloun, M., & Militký, J. 1998. *Statistické zpracování experimentálních dat*. 2 edn. Praha: East Publishing.

- Rao, C. R. 1965. *Linear statistical inference and its applications*. Wiley.
- Rao, J. N. K., & Singh, M. P. 1973. On the choice of estimator in survey sampling. *Australian journal of statistics*, **15**, 95–104.
- Ripley, B. D. 2004. *Spatial statistics*. John Willey & Sons.
- Sakia, R., M. 1992. The box-cox transformation technique: review. *The statistician*, **41**, 169 – 178.
- Stevens, D. L. Jr. 1997. Variable density grid-based sampling designs for continuous spatial populations. *Environmetrics*, **8**, 167–195.
- Stevens, D. L. Jr., & Olsen, A. R. 2003. Variance estimation for spatially balanced samples of environmental resources. *Environmetrics*, **14**, 593–610.
- Stevens, D. L. Jr., & Olsen, A. R. 2004. Spatially balanced sampling of natural resources. *Journal of the american statistical association*, **99**, 262–278.
- Särndal, C. E., Swensson, B., & Wretman, J. 2003. *Model assisted survey sampling*. Springer.
- Trincado, G., & Burkhart, E. 2006. A generalized approach for modeling and localizing site profile curves. *Forest science*, **52**, 670–681.

Metodické přílohy

A Horwitz-Thompsonův teorém pro populace kontinua

Nechť Y je úhrn (suma) veličiny $Y(x)$ (hodnot funkce lokální hustoty) v oblasti D , geografické doméně tj. v množině bodů x kontinua.

$$Y = \int_D Y(x) dx \quad (27)$$

Oblastí D může být jakkoli vymezená zájmová oblast, pro kterou je požadován odhad (území státu, NUTS jednotka, přírodní lesní oblast apod.). Pokud není na konkrétním místě této metodiky řečeno jinak, předpokládá se, že známe naprosto přesnou rozlohu D označovanou ve vzorcích $\lambda(D)$.

[Cordy \[1993\]](#) publikoval Horwitz-Thompsonův teorém pro populace kontinua, podle kterého lze získat nestranný odhad \hat{Y} sumy Y pomocí vzorce (28) [[Cordy, 1993](#), poslední vzorec na straně 355, teorém 1 na straně 356].

$$\hat{Y} = \sum_{x \in s} \frac{Y(x)}{\pi(x)} \quad (28)$$

Symbol $Y(x)$ představuje lokální hustotu veličiny zjištěnou na bodě x výběru s pevné velikosti n , $\pi(x)$ je hustota výběru na bodě x (z angl. *inclusion density*), která je paralelou pravděpodobnosti zahrnutí do výběru v případě výběrů z konečně velkých populací diskrétních objektů [definice viz [Cordy, 1993](#), strana 355]. Odhad (28) je nestranný pokud je funkce $Y(x)$ kladná nebo omezená a zároveň je splněna podmínka $\pi(x) > 0 \forall x \in D$ tzn., že kterýkoli bod x oblasti D může být vybrán.

Pokud je funkce lokální hustoty $Y(x)$ omezená, platí $\pi(x) > 0 \forall x \in D$ a zároveň $\int_D \frac{1}{\pi(x)} dx < \infty$, lze rozptyl odhadu \hat{Y} vyjádřit rovnicí (29) viz [Cordy \[1993, teorém 2 na straně 357\]](#)

$$\begin{aligned} \mathbb{V}_{HTC}(\hat{Y}) &= \int_D \frac{Y^2(x)}{\pi(x)} dx \\ &+ \int_D \int_D Y(x_i) Y(x_j) \left[\frac{\pi(x_i, x_j) - \pi(x_i)\pi(x_j)}{\pi(x_i)\pi(x_j)} \right] dx_i dx_j \end{aligned} \quad (29)$$

Platí-li zároveň podmínka,

$$\pi(x_i, x_j) > 0 \forall x_i, x_j \in D \quad (30)$$

lze nestranný odhad rozptylu $\mathbb{V}(\hat{Y})$ získat dle (31) popřípadě dle upraveného vzorce (32) [[Cordy, 1993, teorém 2 na straně 357](#)].

$$\begin{aligned} \hat{\mathbb{V}}_{HTC}(\hat{Y}) &= \sum_{x_i \in s} \left[\frac{Y(x_i)}{\pi(x_i)} \right]^2 \\ &+ \sum_{x_i \in s} \sum_{\substack{x_j \in s \\ x_i \neq x_j}} Y(x_i) Y(x_j) \left[\frac{\pi(x_i, x_j) - \pi(x_i)\pi(x_j)}{\pi(x_i)\pi(x_j)} \right] \end{aligned} \quad (31)$$

$$\begin{aligned}\hat{\mathbb{V}}_{HTC}(\hat{Y}) &= \sum_{x_i \in s} \left[\frac{Y(x_i)}{\pi(x_i)} \right]^2 + \sum_{x_i \in s} \sum_{\substack{x_j \in s \\ x_i \neq x_j}} \frac{Y(x_i)Y(x_j)}{\pi(x_i)\pi(x_j)} \\ &\quad - \sum_{x_i \in s} \sum_{\substack{x_j \in s \\ x_i \neq x_j}} \frac{Y(x_i)Y(x_j)}{\pi(x_i, x_j)}\end{aligned}\quad (32)$$

Srovnáním s Horwitz-Thompsonovým teorémem pro konečně velké populace disktrétních objektů (viz např. [Cochran, 1977, kapitola 9A.7 od strany 259], [Cordy, 1993, sekce 1, vzorce na straně 354], [Särndal *et al.*, 2003, kapitola 2.8 od strany 42], [Mandallaz, 2007, kapitola 2.2 od strany 4]) zjistíme, že rozdíl spočívá v nahrazení pravděpodobnosti zahrnutí do výběru π_i a potažmo $\pi_{i,j}$ hustotou výběru $\pi(x)$ respektive párovou hustotou $\pi(x_i, x_j)$ ve vzorcích (28), (31), (32), dále v nahrazení součtů určitými integrály ve vzorcích (27) a (29) a vypuštění členu $1 - \pi_i$ z čitatele prvního sčítance ve vzorcích (29), (31) a (32).

Vzorce (28) až (32) jsou odvozeny a tvrzení o nestrannosti odhadů platí pro výběrové designy s konstantní velikostí výběru. Pokud se velikost výběru může mezi realizacemi lišit, je nutné použít buďto upravené vzorce zohledňující proměnlivou velikost výběru (viz sekce E od strany 31) nebo odhady dle vzorců (28) až (32) akceptovat jako nestranné podmíněně na realizované velikosti výběru (viz příloha C od strany 29). Pro druhou z uvedených variant jsou hodnoty $\pi(x)$ a $\pi(x_i, x_j)$ vyjádřeny vzhledem k realizované velikosti výběru v D) [Cordy, 1993, sekce 3 od strany 359].

Stejně jako v případě konečných populací disktrétních objektů, lze rozptyl bodového odhadu (28) alternativně vyjádřit a odhadnout pomocí Sen-Yates-Grundyho vzorců (33) a (34) [Cordy, 1993, poslední odstavec na straně 358, pokračování na straně 359].

$$\mathbb{V}_{SYG}(\hat{Y}) = \frac{1}{2} \int_D \int_D \left[\pi(x_i)\pi(x_j) - \pi(x_i, x_j) \right] \left[\frac{Y(x_i)}{\pi(x_i)} - \frac{Y(x_j)}{\pi(x_j)} \right]^2 dx_i dx_j \quad (33)$$

$$\hat{\mathbb{V}}_{SYG}(\hat{Y}) = \frac{1}{2} \sum_{x_i \in s} \sum_{\substack{x_j \in s \\ x_j \neq x_i}} \left[\frac{\pi(x_i)\pi(x_j) - \pi(x_i, x_j)}{\pi(x_i, x_j)} \right] \left[\frac{Y(x_i)}{\pi(x_i)} - \frac{Y(x_j)}{\pi(x_j)} \right]^2 \quad (34)$$

Vzorec (34) můžeme zapsat ve výpočetně efektivnějším tvaru (35), pokud zavedeme libovolné pořadí $i = (1 \dots n)$ inventarizačních bodů $x_i \in s$, kde n představuje konstantní velikost výběru s [viz například Cochran, 1977, kapitola 9A.7, strana 260 až 261].

$$\hat{\mathbb{V}}_{SYG}(\hat{Y}) = \sum_i^{n-1} \sum_{j>i}^n \left[\frac{\pi(x_i)\pi(x_j) - \pi(x_i, x_j)}{\pi(x_i, x_j)} \right] \left[\frac{Y(x_i)}{\pi(x_i)} - \frac{Y(x_j)}{\pi(x_j)} \right]^2 \quad (35)$$

Sen-Yates-Grundyho vzorec lze aplikovat pouze pro výběry konstantní velikosti. Jejich podoba je shodná jak pro konečné populace disktrétních objektů tak pro nekonečně velké populace bodů kontinua. Důkaz ekvivalence Horwitz-Thompsonova a Sen-Yates-Grundyho vzorců pro rozptyl bodového odhadu v případě konečných populací disktrétních objektů můžeme dohledat například v práci Cochran [1977, teorém 9A.5, strana 260] a Särndal *et al.* [2003, důkaz na str. 45]. Odhady rozptylu dle (31) a (34) se v případě konkrétního výběru zpravidla liší, oba jsou však nestranné.

Jak Horwitz-Thompsonův tak i Sen-Yates-Grundyho odhad rozptylu může za určitých okolností nabývat záporných hodnot [Cochran, 1977, kapitola strana 261], které nelze

z teoretického hlediska ani prakticky (konstrukce intervalových odhadů) akceptovat. Sen-Yates-Grundyho odhady jsou však vždy kladné při splnění podmínky (36) [Särndal *et al.*, 2003, poznámka 2.8.7 na straně 47 a 48].

$$\pi(x_i)\pi(x_j) - \pi(x_i, x_j) \geq 0 \quad \forall x_i \neq x_j \in D \quad (36)$$

Garance kladných hodnot pro řadu prakticky používaných a efektivních výběrových schémat bývá důvodem k přednostnímu použití Sen-Yates-Grundyho odhadů rozptylu [Cordy & Thompson, 1995, poslední odstavec na straně 177], [Cooper, 2006, sekce 4 od strany 544]. Särndal *et al.* [2003, poznámka 2.8.7 na straně 47 a 48] konstatují, že v praxi je riziko výskytu negativních odhadů rozptylu u středně velkých a velkých výběrů poměrně malé a není třeba se této situace zvlášť obávat. Naproti tomu Stevens [1997, sekce 5 na straně 192] uvádí častý výskyt negativních hodnot Horwitz-Thompsonova odhadu rozptylu při použití designu DM-NRTS (z angl. *multiple-density nested random tessellation stratified sampling*).

Literatura uvádí, že Horwitz-Thompsonův tak i Sen-Yates-Grundyho odhad rozptylu (28) mohou být v případě některých výběrových schémat nestabilní [Cochran, 1977, kapitola 9A.7, strana 261], [Cordy & Thompson, 1995, předposlední odstavec na straně 177], [Stevens & Olsen, 2003, poslední odstavec na straně 598 pokračování na straně 599], [Stevens & Olsen, 2004, sekce 4.1, strana 269], [Cooper, 2006, první a druhý odstavec na straně 545]. Cochran [1977, kapitola 9A.7, strana 261] odkazuje na výsledky Rao a Singha [1973], kteří použitím Brewerovy metody výběru z dat 34 populací publikovaných v odborné literatuře zjistili výrazně vyšší stabilitu Sen-Yates-Grundyho v porovnání s Horwitz-Thompsonovými odhady rozptylu, které navíc vycházely často záporné. Stevens & Olsen [2003, 2004] jako důvod nestability uvádí možnost výskytu velmi nízkých hodnot $\pi(x_i, x_j) \rightarrow 0$ ve jmenovateli příslušného členu $\hat{V}_{HTC}(\hat{Y})$ respektive $\hat{V}_{SYG}(\hat{Y})$, což výrazně ovlivňuje výslednou hodnotu odhadu rozptylu. Tato situace typicky nastává při málo pravděpodobném nikoli však nemožném výběru vzájemně blízkých inventarizačních bodů některým z efektivních výběrových designů využívajících pozitivní prostorové korelace lokální hustoty. Nestabilita odhadů $\hat{V}_{HTC}(\hat{Y})$ a $\hat{V}_{SYG}(\hat{Y})$ je problémem zvláště pro výběrové designy, u kterých dochází ke skokovým změnám hustoty výběru $\pi(x)$ (např. na hranici oblastí - geograficky vymezených domén, s různou hustotou výběru inventarizačních bodů) [Stevens & Olsen, 2003, sekce 3, první odstavec na straně 599]. Dle Stevens [1997, sekce 5, strana 192] má Horwitz-Thompsonův odhad rozptylu designu DM-NRTS tendenci k nestabilitě a častému výskytu negativních hodnot zatímco Sen-Yates-Grundyho odhady rozptylu se zdají stabilní.

Základní varianta odhadu rozptylu je dle této metodiky založena na aproximaci párové hustoty $\pi(x_i, x_j) = n(n-1)/\lambda(D)$, což odpovídá zcela náhodnému výběru inventarizačních bodů v D (viz příloha A.1, strana 23). Při této aproximaci jsou Horwitz-Thompsonův a Sen-Yates-Grundyho rozptyly ekvivalentní a totéž platí pro hodnoty jejich odhadů [Särndal *et al.*, 2003, poznámka 2.8.4 na straně 47]. Podmínka (36) platí pro jakoukoli dvojici inventarizačních bodů, což vylučuje možnost získání záporných odhadů rozptylu. Konstantní hodnota používané aproximace párové hustoty spolu s běžnou velikostí výběru v uvažovaných doménách jsou zárukou akceptovatelné stability odhadů rozptylu založených na tomto teorému.

A.1 Aplikace Horwitz-Thompsonova teorému pro zcela náhodný výběr inventarizačních bodů v zájmové oblasti D

Při zcela náhodném výběru (zkratka URS z angl. *Uniform Random Sampling*) pevného počtu n inventarizačních bodů v zájmové oblasti D pro hustotu a sdruženou hustotu platí (37), (38) [viz například Stevens & Olsen, 2003, první odstavec, strana 599].

$$\pi(x) = \pi(x_i) = \frac{n}{\lambda(D)} \quad (37)$$

$$\pi(x_i, x_j) = \frac{n(n-1)}{\lambda^2(D)} \quad (38)$$

Dosazením do (28) získáme pro bodový odhad úhrnu \hat{Y} vzorec (39).

$$\hat{Y}_{URS} = \sum_{x \in s} \frac{Y(x)}{\pi(x)} = \lambda(D) \frac{\sum_{x \in s} Y(x)}{n} = \lambda(D) \hat{Y} \quad (39)$$

Pro rozptyl $\mathbb{V}(\hat{Y}_{URS})$ platí vzorce (40) až (46).

$$\mathbb{V}(\hat{Y}_{URS}) = \int_D \frac{Y^2(x)}{n \lambda^{-1}(D)} dx - \frac{1}{n} \int_D \int_D Y(x_i) Y(x_j) dx_i dx_j \quad (40)$$

$$= \frac{\lambda^2(D)}{n \lambda(D)} \int_D Y^2(x) dx - \frac{\lambda^2(D)}{n \lambda^2(D)} \int_D Y(x) dx \int_D Y(x) dx \quad (41)$$

$$= \frac{\lambda^2(D)}{n} \left\{ \mathbb{E}[Y^2(x)] - \mathbb{E}^2[Y(x)] \right\} = \frac{\lambda^2(D)}{n} (\bar{Y}^2 - \bar{Y}^2) \quad (42)$$

$$= \frac{\lambda^2(D)}{n} \mathbb{V}[Y(x)] = \frac{\lambda^2(D)}{n} \mathbb{E} \left\{ \left(Y(x) - \mathbb{E}[Y(x)] \right)^2 \right\} \quad (43)$$

$$= \frac{\lambda^2(D)}{n} \int_D [Y(x) - \bar{Y}]^2 f(x) dx \quad (44)$$

$$= \frac{\lambda(D)}{n} \int_D [Y(x) - \bar{Y}]^2 dx \quad (45)$$

$$= \lambda^2(D) \mathbb{V}(\hat{Y}) \quad (46)$$

Vzorec (40) získáme dosazením (37) a (38) do (29) (strana 20). Výraz v hranatých závorkách uvnitř druhého integrálu (29) se po dosazení (37) a (38) zredukuje na $-n^{-1}$. Ve vzorci (41) je provedeno násobení členů podíly rovnými 1, čehož je využito při úpravě do podoby (42) a dále (43) pomocí známého výpočtového tvaru pro rozptyl náhodné veličiny [viz například Anděl, 1978, vzorec 12 na straně 14]. Druhý výraz (43) obsahuje rozptyl ve tvaru střední hodnoty druhých mocnin odchylek $Y(x)$ od střední hodnoty \bar{Y} tj. v podobě druhého centrálního momentu. Při úpravě do podoby (44) (a dříve též při úpravě do podoby prvního výrazu (42)) se využívá vlastností (de-facto definice) střední hodnoty spojitě náhodné veličiny [viz například Anděl, 1978, vzorec 9, strana 13], kde $f(x) = \lambda^{-1}(D)$ je hustota pravděpodobnosti spojitě rovnoměrného rozdělení (při URS je možnost výběru inventarizačního bodu ve všech částech D stejná). Vzorec (45) získáme prostým vykrácením $\lambda(D)$ v čitateli (44). Poslední uvedený vzorec (46) využívá vztahu mezi rozptylem náhodné veličiny a rozptylem odhadu její střední hodnoty při zcela náhodném výběru (výběr kteréhokoli inventarizačního bodu nijak nezávisí na inventarizačních bodech již dříve vybraných) [viz například Anděl, 1978, věta 11, vzorec 31, strana 32].

Vzorec pro odhad rozptylu \hat{Y} získáme postupnými úpravami (47) až (54) výchozího obecného vzorce (32) (strana 21) s dosazením hustot (37) a (38).

$$\begin{aligned} \hat{V}(\hat{Y}_{URS}) &= \sum_{x \in s} \left[\frac{\lambda(D)Y(x)}{n} \right]^2 + \sum_{x_i \in s} \sum_{\substack{x_j \in s \\ x_i \neq x_j}} \frac{\lambda^2(D)Y(x_i)Y(x_j)}{n^2} \\ &\quad - \sum_{x_i \in s} \sum_{\substack{x_j \in s \\ x_i \neq x_j}} \frac{\lambda^2(D)Y(x_i)Y(x_j)}{n(n-1)} \end{aligned} \quad (47)$$

$$\begin{aligned} &= \frac{\lambda^2(D)}{n-1} \left\{ (n-1) \sum_{x \in s} \left[\frac{Y(x)}{n} \right]^2 \right. \\ &\quad \left. + (n-1) \sum_{x_i \in s} \sum_{\substack{x_j \in s \\ x_i \neq x_j}} \frac{Y(x_i)Y(x_j)}{n^2} - \sum_{x_i \in s} \sum_{\substack{x_j \in s \\ x_i \neq x_j}} \frac{Y(x_i)Y(x_j)}{n} \right\} \end{aligned} \quad (48)$$

$$\begin{aligned} &= \frac{\lambda^2(D)}{n-1} \left\{ n \sum_{x \in s} \left[\frac{Y(x)}{n} \right]^2 - \sum_{x \in s} \left[\frac{Y(x)}{n} \right]^2 \right. \\ &\quad \left. + \sum_{x_i \in s} \sum_{\substack{x_j \in s \\ x_i \neq x_j}} \frac{Y(x_i)Y(x_j)}{n} - \sum_{x_i \in s} \sum_{\substack{x_j \in s \\ x_i \neq x_j}} \frac{Y(x_i)Y(x_j)}{n^2} \right. \\ &\quad \left. - \sum_{x_i \in s} \sum_{\substack{x_j \in s \\ x_i \neq x_j}} \frac{Y(x_i)Y(x_j)}{n} \right\} \end{aligned} \quad (49)$$

$$= \frac{\lambda^2(D)}{n-1} \left\{ \sum_{x \in s} \frac{Y^2(x)}{n} - \sum_{x_i \in s} \sum_{x_j \in s} \frac{Y(x_i)Y(x_j)}{n^2} \right\} \quad (50)$$

$$= \frac{\lambda^2(D)}{n-1} \left\{ \sum_{x \in s} \frac{Y^2(x)}{n} - \sum_{x \in s} \frac{Y(x)}{n} \sum_{x \in s} \frac{Y(x)}{n} \right\} \quad (51)$$

$$= \frac{\lambda^2(D)}{(n-1)} \left\{ \sum_{x \in s} \frac{Y^2(x)}{n} - 2 \sum_{x \in s} \frac{Y(x)}{n} \hat{Y} + \hat{Y}^2 \right\} \quad (52)$$

$$= \frac{\lambda^2(D)}{(n-1)} \sum_{x \in s} \frac{Y^2(x) - 2Y(x)\hat{Y} + \hat{Y}^2}{n} \quad (53)$$

$$= \frac{\lambda^2(D)}{n(n-1)} \sum_{x \in s} [Y(x) - \hat{Y}]^2 \quad (54)$$

Úpravy až po (49) jsou pouze základní algebraické operace. Vzorec (50) rezultuje vzájemným vyrušením třetího a pátého členu a sloučením druhého a čtvrtého členu (49) do jednoho dvojitého součtu (viz indikace druhého součtu přes všechny prvky výběru v rámci (50) versus čtvrtý člen (49) s druhým součtem vynechávajícím aktuální prvek prvního součtu). Úprava druhého členu ve vzorci (51) je triviální, je zde obsažena pouze pro zdůraznění shody obou činitelů s odhadem střední hodnoty $\hat{Y} = \hat{Y}\lambda(D)^{-1}$. Ve vzorci (51) je výraz ve složených závorkách upraven přičtením a odečtením \hat{Y}^2 . Úprava do tvaru (52) využívá rovnosti $\sum_{x \in s} \hat{Y} = n\hat{Y}$, kde suma probíhá přes n prvků výběru s a odhad střední hodnoty \hat{Y} vystupuje jako konstanta. Ke vzorci (53) dospějeme jednoduchou úpravou (52)

spočívající ve vytknutí n^{-1} před sumu a vyjádření čitatele jako druhé mocniny rozdílu dvou čísel.

B Odhad rozptylu pomocí Taylorovy aproximace

Některé cílové parametry NIL jsou dány nelineární funkcí (55) dvou odhadovaných parametrů (náhodných veličin), které zde označíme Y_1, Y_2 .

$$Y_{1,2} = f(Y_1, Y_2) \quad (55)$$

Hodnoty parametrů Y_1 a Y_2 jsou neznámé a odhadují se v rámci samotného šetření. Hodnota $Y_{1,2}$ se odhaduje jako $\hat{Y}_{1,2}$ dle (56).

$$\hat{Y}_{1,2} = f(\hat{Y}_1, \hat{Y}_2) \quad (56)$$

Analytické vyjádření rozptylu $\mathbb{V}(\hat{Y}_{1,2})$ odhadu $\hat{Y}_{1,2}$ jakožto nelineární funkce náhodných veličin je zpravidla komplikované. Proto se využívá metody Taylorova rozvoje s cílem aproximovat $\hat{Y}_{1,2}$ lineární funkcí, jejíž rozptyl je dán (62). Aproximace $\hat{Y}_{1,2}$ v okolí (Y_1, Y_2) se zanedbáním členů druhého a vyšších řádů je vyjádřena (57) [Jarník, 1963, kapitola XII, §1, od strany 289], [Särndal *et al.*, 2003, kapitola 5.5, vzorec 5.5.7, 5.5.8, strana 173 až 174].

$$\hat{Y}_{1,2} \approx \hat{Y}_{1,2}^{(apx)} = Y_{1,2} + a(\hat{Y}_1 - Y_1) + b(\hat{Y}_2 - Y_2) \quad (57)$$

$$a = \left. \frac{\partial f}{\partial \hat{Y}_1} \right|_{Y_1, Y_2} \quad \dots \text{parciální derivace } f \text{ dle } \hat{Y}_1 \text{ v bodě } (Y_1, Y_2) \quad (58)$$

$$b = \left. \frac{\partial f}{\partial \hat{Y}_2} \right|_{Y_1, Y_2} \quad \dots \text{parciální derivace } f \text{ dle } \hat{Y}_2 \text{ v bodě } (Y_1, Y_2) \quad (59)$$

Rozptyl bodového odhadu $\hat{Y}_{1,2}$ je pomocí (57) aproximován rozptylem $\hat{Y}_{1,2}^{(apx)}$ viz (60). Dle (61) aproximujeme také odhad $\hat{\mathbb{V}}(\hat{Y}_{1,2})$.

$$\mathbb{V}(\hat{Y}_{1,2}) \approx \mathbb{V}(\hat{Y}_{1,2}^{(apx)}) \quad (60)$$

$$\hat{\mathbb{V}}(\hat{Y}_{1,2}) \approx \hat{\mathbb{V}}(\hat{Y}_{1,2}^{(apx)}) \quad (61)$$

Rozptyl $\mathbb{V}(\hat{Y}_{1,2}^{(apx)})$ je díky lineární povaze bodového odhadu $\hat{Y}_{1,2}^{(apx)}$ dán (62) [Anděl, 1978, věta 4, vzorec 16 na stranách 28 a 29, věta 2, vzorec 14, strana 14], [Särndal *et al.*, 2003, kapitola 5.5, vzorec 5.5.3, strana 172].

$$\mathbb{V}(\hat{Y}_{1,2}^{(apx)}) = a^2\mathbb{V}(\hat{Y}_1) + b^2\mathbb{V}(\hat{Y}_2) + 2ab\mathbb{C}(\hat{Y}_1, \hat{Y}_2) \quad (62)$$

Hodnoty parciálních derivací a (58) a b (59) nejsou známy, poněvadž neznáme přesné hodnoty parametrů Y_1 a Y_2 tj. polohu bodu, v němž mají být parciální derivace vyhodnoceny. Odhad rozptylu dle (63) proto používá \hat{a} a \hat{b} tj. parciální derivace $\hat{Y}_{1,2}$ v bodě (\hat{Y}_1, \hat{Y}_2) , což je v případě dostatečně velkých výběrů zdůvodněno tím, že \hat{a} a \hat{b} jsou konzistentními odhady a a b za předpokladu, že \hat{Y}_1 a \hat{Y}_2 jsou konzistentními odhady Y_1 a Y_2 [Särndal *et al.*, 2003, kapitola 5.5, poslední odstavec na straně 174].

$$\hat{\Psi}(\hat{Y}_{1,2}^{(apx)}) \approx \hat{a}^2 \hat{\Psi}(\hat{Y}_1) + \hat{b}^2 \hat{\Psi}(\hat{Y}_2) + 2\hat{a}\hat{b}\hat{C}(\hat{Y}_1, \hat{Y}_2) \quad (63)$$

B.1 Rozptyl podílu $\hat{R}_{1,2}$, obecné řešení pomocí Taylorovy aproximace

Řada cílových parametrů NIL může být vyjádřena jako podíl $R_{1,2}$ úhrnů Y_1 a Y_2 nebo ekvivalentně středních hodnot \bar{Y}_1 a \bar{Y}_2 dle (64). Typickým příkladem je hektarová střední zásoba hroubí tj. podíl celkové zásoby a rozlohy *porostní půdy* (jeden ze tří druhů pozemku v rámci kategorie *Les*). Odhad $\hat{R}_{1,2}$ je získán jako podíl odhadů úhrnu \hat{Y}_1 a \hat{Y}_2 nebo ekvivalentně jako podíl odhadů středních hodnot $\hat{\bar{Y}}_1$ a $\hat{\bar{Y}}_2$ viz (64) a (65).

$$R_{1,2} = \frac{Y_1}{Y_2} = \frac{\bar{Y}_1}{\bar{Y}_2} \quad \hat{R}_{1,2} = \frac{\hat{Y}_1}{\hat{Y}_2} = \frac{\hat{\bar{Y}}_1}{\hat{\bar{Y}}_2} \quad (64)$$

$$\hat{Y}_1 = \lambda(D)\hat{\bar{Y}}_1 \quad \hat{Y}_2 = \lambda(D)\hat{\bar{Y}}_2 \quad (65)$$

Rozptyl $R_{1,2}$ jakožto nelineární funkci můžeme vyjádřit a odhadovat s využitím Taylorovy aproximace (viz sekce B). Parciální derivace (58), (59) a jejich odhady jsou pro funkci podílu dány (66) a (67).

$$a = \frac{1}{Y_2} \quad b = -\frac{Y_1}{Y_2^2} = -\frac{R_{1,2}}{Y_2} \quad (66)$$

$$\hat{a} = \frac{1}{\hat{Y}_2} \quad \hat{b} = -\frac{\hat{Y}_1}{\hat{Y}_2^2} = -\frac{\hat{R}_{1,2}}{\hat{Y}_2} \quad (67)$$

Podíl $R_{1,2}$ aproximujeme Taylorovým rozvojem se zanedbáním členů druhého a vyšších řádů (viz (68)).

$$\hat{R}_{1,2} \approx \hat{R}_{1,2}^{(apx)} \quad (68)$$

Aproximaci $\hat{R}_{1,2}^{(apx)}$ můžeme na základě (57) a (66) psát ve tvaru (69). Vzorec (72) platí obecně bez ohledu na typ odhadu \hat{Y}_1 a \hat{Y}_2 .

$$\hat{R}_{1,2}^{(apx)} = R_{1,2} + \frac{1}{Y_2}(\hat{Y}_1 - Y_1) - \frac{R_{1,2}}{Y_2}(\hat{Y}_2 - Y_2) \quad (69)$$

$$\begin{aligned} &= R_{1,2} + \frac{\hat{Y}_1}{Y_2} - \frac{R_{1,2}\hat{Y}_2}{Y_2} + \left[\frac{R_{1,2}\hat{Y}_2}{Y_2} - \frac{Y_1}{Y_2} \right] \\ &\quad - \frac{Y_1}{Y_2} + \left[\frac{Y_1}{Y_2} - \frac{R_{1,2}\hat{Y}_2}{Y_2} \right] + \frac{R_{1,2}Y_2}{Y_2} \end{aligned} \quad (70)$$

$$= R_{1,2} + \frac{1}{Y_2}(\hat{Y}_1 - R_{1,2}\hat{Y}_2) - R_{1,2} + R_{1,2} \quad (71)$$

$$= R_{1,2} + \frac{1}{Y_2}(\hat{Y}_1 - R_{1,2}\hat{Y}_2) \quad (72)$$

Další úpravy (73) až (75) jsou provedeny pro případ, kdy obojí \hat{Y}_1 i \hat{Y}_2 jsou jednofázovými odhady.

$$\hat{R}_{1,2}^{(apx)} = R_{1,2} + \frac{1}{Y_2} \sum_{x \in s} \frac{Y_1(x) - R_{1,2}Y_2(x)}{\pi(x)} \quad (73)$$

$$= R_{1,2} + \frac{1}{Y_2} \sum_{x \in s} \frac{Z(x)}{\pi(x)} \quad (74)$$

$$= R_{1,2} + \frac{1}{Y_2} \hat{Z} \quad (75)$$

Veličina $Z(x)$ je reziduální proměnná definovaná (76), \hat{Z} dle (77) označuje jednofázový odhad (28) úhrnu $Z(x)$ v oblasti D .

$$Z(x) = Y_1(x) - R_{1,2}Y_2(x) \quad (76)$$

$$\hat{Z} = \sum_{x \in s} \frac{Y_1(x) - R_{1,2}Y_2(x)}{\pi(x)} = \sum_{x \in s} \frac{Z(x)}{\pi(x)} = \lambda(D)\hat{Z} \quad (77)$$

Střední hodnota odhadu \hat{Z} je nulová, jak je patrné z rovnice (78).

$$\mathbb{E}(\hat{Z}) = \bar{Z} = \bar{Y}_1 - R_{1,2}\bar{Y}_2 = \bar{Y}_1 - \frac{\bar{Y}_1}{\bar{Y}_2}\bar{Y}_2 = \bar{Y}_1 - \bar{Y}_1 = 0 \quad (78)$$

Rozptyl $\mathbb{V}(\hat{R}_{1,2})$ aproximujeme $\mathbb{V}(\hat{R}_{1,2}^{(apx)})$ dle (79), který získáme jako rozptyl (75) tj. rozptyl součtu konstanty $R_{1,2}$ a součinu konstanty Y_2^{-1} a odhadu úhrnu \hat{Z} [Anděl, 1978, aplikace věty 2, vzorec 14, strana 14].

$$\mathbb{V}(\hat{R}_{1,2}) \approx \mathbb{V}(\hat{R}_{1,2}^{(apx)}) = \frac{1}{Y_2^2} \mathbb{V}(\hat{Z}) \quad (79)$$

Poněvadž není známa skutečná hodnota podílu $R_{1,2}$ používá se pro výpočet $\hat{\mathbb{V}}(\hat{R}_{1,2}^{(apx)})$ namísto $Z(x)$ respektive \hat{Z} aproximací $Z_o(x)$ a \hat{Z}_o dle (80) a (81). Nahrazení $R_{1,2}$ odhadem $\hat{R}_{1,2}$ pro dostatečně velké výběry zanedbáváme vzhledem ke konzistenci $\hat{R}_{1,2}$, která vyplývá konzistence \hat{Y}_1 a \hat{Y}_2 .

$$Z_o(x) = Y_1(x) - \hat{R}_{1,2}Y_2(x) \quad (80)$$

$$\hat{Z}_o = \sum_{x \in s} \frac{Y_1(x) - \hat{R}_{1,2}Y_2(x)}{\pi(x)} = \sum_{x \in s} \frac{Z_o(x)}{\pi(x)} = \lambda(D)\hat{Z}_o \quad (81)$$

$$\hat{Z} = \lambda(D)\hat{Z} \approx \hat{Z}_o = \lambda(D)\hat{Z}_o \quad (82)$$

Vzorec (83) je výslednou aproximací odhadu rozptylu podílu dvou jednofázových odhadů.

$$\hat{\mathbb{V}}(\hat{R}_{1,2}) \approx \hat{\mathbb{V}}(\hat{R}_{1,2}^{(apx)}) \approx \frac{1}{\hat{Y}_2^2} \hat{\mathbb{V}}(\hat{Z}_o) \quad (83)$$

Odhad rozptylu $\hat{\mathbb{V}}(\hat{Z}_o)$ získáme dle (31), (32) nebo v některých případech vhodněji dle (34), (35).

Ekvivalentně lze rozptyl podílu dvou jednofázových odhadů vyjádřit pomocí (84) [Särndal *et al.*, 2003, podrobnosti odvození v sekci 5.6 na straně 176],

$$\hat{\mathbb{V}}(\hat{R}_{1,2}) \approx \hat{\mathbb{V}}(\hat{R}_{1,2}^{(apx)}) \approx \frac{1}{\hat{Y}_2^2} \left[\hat{\mathbb{V}}(\hat{Y}_1) + \hat{R}_{1,2}^2 \hat{\mathbb{V}}(\hat{Y}_2) - 2\hat{R}_{1,2} \hat{\mathbb{C}}(\hat{Y}_1, \hat{Y}_2) \right] \quad (84)$$

kde $\hat{\mathbb{V}}(\hat{Y}_1)$, $\hat{\mathbb{V}}(\hat{Y}_2)$ jsou odhady rozptylů jednofázových úhrnů v čitateli (\hat{Y}_1) a jmenovateli (\hat{Y}_2) podílu $\hat{R}_{1,2}$ a $\hat{\mathbb{C}}(\hat{Y}_1, \hat{Y}_2)$ je kovariance uvedených jednofázových odhadů. Vzorec (84) přímo vyplývá ze vzorce (63) (sekce B, strana 26) s dosazením \hat{a} a \hat{b} dle (67).

B.2 Rozptyl podílu, řešení pomocí Taylorovy aproximace při zcela náhodném výběru inventarizačních bodů v zájmové oblasti D

Stejně jako v obecném případě, též při zcela náhodném výběru pevného počtu n inventarizačních bodů v zájmové oblasti D , pro podíl dvou úhrnů a jeho odhad platí vzorec (64) (viz sekce B.1, strana 26). Za účelem odhadu rozptylu $\hat{R}_{1,2}$ je výhodné provést aproximaci (85) Taylorovým rozvojem (viz sekce B.1 od strany 26).

$$\mathbb{V}_{URS}(\hat{R}_{1,2}) \approx \mathbb{V}_{URS}(\hat{R}_{1,2}^{(apx)}) \quad (85)$$

Pro rozptyl $\mathbb{V}_{URS}(\hat{R}_{1,2}^{(apx)})$ platí (86) až (91).

$$\mathbb{V}_{URS}(\hat{R}_{1,2}^{(apx)}) = \frac{1}{Y^2} \mathbb{V}_{URS}(\hat{Z}) \quad (86)$$

$$= \frac{\lambda^2(D)}{Y^2} \mathbb{V}_{URS}(\hat{Z}) \quad (87)$$

$$= \frac{\lambda^2(D)}{nY^2} \mathbb{V}_{URS}[Z(x)] \quad (88)$$

$$= \frac{\lambda(D)}{nY^2} \int_D [Z(x) - \bar{Z}]^2 dx \quad (89)$$

$$= \frac{\lambda(D)}{nY^2} \int_D Z^2(x) dx \quad (90)$$

$$= \frac{\lambda(D)}{nY^2} \int_D [Y_1(x) - R_{1,2}Y_2(x)]^2 dx \quad (91)$$

Vzorec (86) je analogií vzorce (79) v sekci B.1 (strana 27). Vzorce (87) až (89) vyjadřují $\mathbb{V}_{URS}(\hat{Z})$ analogicky ke vzorcům (46), (43) a (45) v sekci A.1. Využitím (78) (sekce B.1, strana 27) můžeme z rozdílu v hranatých závorkách vzorce (89) vypustit nulovou hodnotu \bar{Z} . Tím získáme (90). Závěrečnou úpravu do podoby (90) provedeme dosazením (76) (sekce B.1, strana 27) za $Z(x)$.

Odhad rozptylu $\hat{\mathbb{V}}_{URS}(\hat{R}_{1,2})$ aproximujeme dle (92) (sekce B.1, strana 28). Výsledné aproximativní vzorce jsou dány (93) a (94).

$$\hat{\mathbb{V}}_{URS}(\hat{R}_{1,2}) \approx \hat{\mathbb{V}}_{URS}(\hat{R}_{1,2}^{(apx)}) \quad (92)$$

$$\hat{\mathbb{V}}_{URS}(\hat{R}_{1,2}^{(apx)}) \approx \frac{\lambda^2(D)}{n(n-1)\hat{Y}_2^2} \sum_{x \in s} Z_c^2(x) \quad (93)$$

$$\approx \frac{\sum_{x \in s} [Y_1(x) - \hat{R}_{1,2}Y_2(x)]^2}{n(n-1)\hat{Y}_2^2} \quad (94)$$

Vzorec (93) vyplývá ze vzorce (83) (sekce B.1, strana 27) a vzorce (54) (příloha A.1, strana 24) s aproximací $\hat{Z} \approx 0$ - viz (78) (sekce B.1, strana 27). Vzorec (93) využívá rovnosti $\hat{Y}_2 = \hat{Y}_2 \lambda^{-1}(D)$.

C Výběry inventarizačních bodů v projektech NIL

Pro inventarizaci lesů typické (quasi-)systematické popřípadě prostorově stratifikované výběry využívají čtvercové nebo obdélníkové (méně často trojúhelníkové popřípadě šestiúhelníkové) inventarizační sítě s náhodně zvoleným počátkem (méně často též s náhodně zvolenou orientací). Síť bezzbytku dělí výběrový rámec (sampling frame) $\mathcal{F} \supseteq D$ na tzv. inventarizační bloky stejné velikosti i tvaru, které navazují jeden na druhý a nepřekrývají se. Pro NIL je typické, že výběrový rámec nebývá shodný se zájmovou oblastí D (geografickou doménou). S použitím speciální topologie jej můžeme definovat způsobem, který garantuje konstantní velikost výběru v \mathcal{F} [Stevens, 1997, sekce 3.1 od strany 172], [Mandallaz, 2007, sekce 5.6 od strany 92]. To si lze představit tak, že část roviny \mathbb{R}^2 ve tvaru čtverce nebo obdélníku obsahující doménu D i samotnou inventarizační síť (před náhodným posunem) nejprve svineme tak, aby došlo k navázání dvou protilehlých okrajů po celé jejich délce. Získáme tak válcovou plochu (rotačního válce) kterou v dalším kroce transformujeme napojením okrajových hran (hrany protilehlých podstav rotačního válce). Tím vznikne torus (anuloid), jehož povrch obsahuje doménu D i s inventarizační sítí. Pro libovolný náhodný vektor posunu inventarizační sítě je pak zajištěno, že síť nevystoupí z výběrového rámce \mathcal{F} s pevnou velikostí a umístěním vzhledem k doméně D . Při náhodných posunech inventarizační síť - obrazně řečeno - klouže po povrchu toru.

Typicky jsou požadovány odhady pro celou řadu geografických domén vytvářejících hierarchická členění (NUTS jednotky, jinak vymezené geografické oblasti). Nejsou-li přírodní podmínky zkoumaného území velmi různorodé volí se definice rámce \mathcal{F} tak, aby obsahoval sjednocení všech uvažovaných domén, kterým je v případě NIL celé území státu. Při tomto pojetí, je území státu geografickou doménou nejvyšší uvažované úrovně - nikoli samotným výběrovým rámcem.

Při systematickém výběru jsou inventarizační body umístěny do středu (respektive na průsečíky linií) zatímco při prostorově stratifikovaném výběru zcela náhodně dovnitř každého z inventarizačních bloků (prostorových strát). Na rozdíl od výběrového rámce \mathcal{F} nelze v typickém případě doménu D bezzbytku dělit na vzájemně navazující a nepřekrývající se inventarizační bloky. Části některých inventarizačních bloků většinou zasahují mimo doménu D . Výsledkem náhodného posunu inventarizační sítě vůči D , popřípadě náhodného generování polohy inventarizačních bodů (výchozích bodů traktů) uvnitř každého z bloků, je proměnlivé množství inventarizačních bodů uvnitř domény D . Velikost výběru v D je při použití zde uvažovaných výběrových designů nekonstantní.

Vzorce pro bodové odhady a jejich rozptyl, tak jak jsou uvedeny v sekci A, platí a učiněné odhady jsou nestranné pro výběry konstantní velikosti (viz druhý odstavec na straně 21). Pevnou velikost výběru lze zajistit formálním ztotožněním zájmové oblasti D s výběrovým rámcem \mathcal{F} . Při současném předefinování funkce lokální hustoty tak, aby v oblastech vně D , ale v rámci \mathcal{F} měla tato funkce paušálně nulovou hodnotu. Dosáhneme tak stavu, kdy úhrn kterékoli veličiny v rámci D bude roven úhrnu téže veličiny v rámci \mathcal{F} . Nestranný odhad úhrnu v \mathcal{F} je zároveň nestranným odhadem úhrnu v zájmové oblasti D . Touto alternativou, ačkoli je relevantní pro některé typy používaných inventarizačních sítí [Stevens, 1997], se zde dále nezabýváme. Zdůvodnění je jednak ve vyšší výpočetní náročnosti (do výpočtu odhadů pro stát i jeho podoblasti vstupují pokaždé všechny inventarizační body), ale též v navýšení odhadu rozptylu při použití výpočtových vzorců

založených na předpokladu zcela náhodného rozmístění inventarizačních bodů v \mathcal{F} . Faktem však, je že při odhadu pro atributové domény je nahrazování lokálních hustot (či veličin z nichž, je lokální hustota následně odvozena) nulami standardní postup, viz sekce 2.3.

Dle Cordyho [1993, sekce 3 od strany 359] lze k nekonstantní velikosti výběru v D přistoupit dvojím způsobem:

- pojetí nestrannosti bodového odhadu a jeho rozptylu podmíněně na velikosti výběru (conditional inference);
- nahrazením hodnot hustoty výběru a párové hustoty výběru inventarizačních bodů (výchozích bodů traktů) jejich očekávanými hodnotami (přes všechny možné velikosti výběru) - při této variantě jsou odhady a jejich rozptyl nestranné nepodmíněně na realizované velikosti výběru (unconditional inference, sekce E od strany 31)

Odhady nestranné podmíněně na realizované velikosti výběru mají lepší statistické vlastnosti - nižší rozptyl, očekávaná hodnota chyby podmíněně na velikosti výběru (conditional bias) je rovna nule. Poněvadž je však při výpočtu odhadů v geografických doménách použito jiných (doménově specifických) výběrových vah než při odhadu pro celek, nejsou odhady úhrnu geograficky aditivní.

Z důvodu požadavku aditivity odhadů úhrnu (definice aditivity je uvedena v příloze F) byly pro vyhodnocení NIL2 na základě dat inventarizační sítě NIL1 použity odhady nestranné nepodmíněně (druhá výše jmenovaná varianta). Vzhledem k velkým očekávaným velikostem výběru v ČR i v uvažovaných podoblastech lze očekávat prakticky zanedbatelné zvýšení rozptylu a chybu úhrnu podmíněnou rozdíly realizovaných a očekávaných velikostí výběru.

Přibližně systematické rozmístění inventarizačních bodů v síti NIL1 ve spojení s poměrně kompaktním tvarem geografických domén výrazně snižují variabilitu počtu inventarizačních bodů v geografických doménách. Hodnoty podílu směrodatné odchylky¹² počtu inventarizačních traktů v geografické doméně (rozšířené o obalovou zónu o velikosti 300 m) a očekávaného počtu traktů v geografické doméně se pohybují od 0.28 % pro Jihočeský kraj až po 1.59 % pro Hlavní město Prahu. Na úrovni celé ČR byla hodnota variačního koeficientu odhadnuta na 0.14 %. Nevýhody odhadu úhrnu nestranného nepodmíněně se tak tím spíše dostávají na prakticky zanedbatelnou úroveň.

D Odhady rozptylu v projektech NIL

Zcela náhodný výběr inventarizačních bodů v zájmové oblasti D zpravidla není vhodným postupem v porovnání se systematickými a prostorově stratifikovanými výběry, které zajišťují rovnoměrnější pokrytí D inventarizačními body [Cochran, 1977, kapitoly 8.9, 8.10 a 8.13, strany 219, 221, 227], [Cordy & Thompson, 1995, první odstavec na straně 174], [Stevens, 1997, sekce 3, strana 171], [Barabesi, 2003, druhý odstavec, strana 356], [Ripley, 2004, kapitola 3 od strany 19], [Heikkinen, 2006, kapitola 10, od strany 155], [Gregoire & Valentine, 2008, sekce 3.2.2, od strany 49].

Na nepříliš vzdálených inventarizačních bodech lze z důvodu často se vyskytující pozitivní prostorové korelace, očekávat podobné hodnoty sledované veličiny. Výběrem relativně blízkých inventarizačních bodů dochází ke snížení množství informace, kterou by bylo možné získat, pokud by na bodech, v průměru vzájemně více vzdálených, byly pozorovány hodnoty rozdílnější. Efektivní výběrové designy proto nějakým způsobem omezují ($\pi(x_i, x_j) \rightarrow 0$ pro velmi blízké body $x_i, x_j \in D$), nebo přímo vylučují možnost výběru

¹²Odhadnuté na základě simulovaných výběrů dle designu sítě NIL1.

vzájemně blízkých inventarizačních bodů. Tím je zajištěna větší variabilita hodnot lokální hustoty na inventarizačních bodech v rámci výběru a současně dochází ke snížení variability bodových odhadů mezi výběry tj. ke zvýšení přesnosti odhadu cílového parametru.

V krajním případě není splněna podmínka (30) (sekce A, strana 20). Horwitz-Thompsonovy ani Sen-Yates-Grundyho vzorce pak nelze použít k získání nestranných odhadů rozptylu a nezbývá než hledat vhodné alternativy. S tímto omezením se typicky setkáme u systematických a prostorově stratifikovaných výběrů s jedním bodem v každém z pevně daných strát angl. *tesselation stratified sampling* ve zkratce TS [Cordy & Thompson, 1995], popřípadě CTS (z angl. *congruent tessellation stratified sampling*) [Stevens, 1997] nebo TSS [Barabesi & Franceschi, 2010], ale také v případě inventarizační sítě NIL1 odvozené od systematického výběru (quasi-systematický výběr).

Nestrannost odhadu většinou nebývá striktním požadavkem. V praxi se používá řada odhadů (nejen rozptylu), které, ačkoli nejsou nestranné, mají jiné přednosti například menší rozptyl respektive střední kvadratickou chybu [Anděl, 1978, kapitola XV.1, strana 252]. Mezi odhady mají zvláštní postavení odhady asymptoticky nestranné a konzistentní. Zjednodušeně řečeno, očekávaná hodnota těchto odhadů se při vzrůstající velikosti výběru přibližuje skutečné hodnotě parametru v populaci (ne zcela ustálené definice viz. [Rao, 1965, sekce 5c.1, strana 344], [Cochran, 1977, kapitola 2.4, strana 21], [Särndal *et al.*, 2003, kapitola 5.3 od strany 166], [Gregoire & Valentine, 2008, sekce 2.2.1, strana 19]). Pokud neexistují, nebo nejsou-li známy nestranné, konzistentní odhady rozptylu vhodné pro konkrétní výběrový design a vlastnosti populace, lze v některých případech použít alespoň konzervativní odhady, které skutečný rozptyl nadhodnocují. Intervalový odhad konstruovaný na základě konzervativního odhadu rozptylu (při správném postupu odpovídajícím rozdělení pravděpodobnosti bodového odhadu) zachytí skutečnou hodnotu odhadovaného parametru s pravděpodobností větší než nominální $(1 - \alpha)$. Statistická jistota těchto konzervativních intervalových odhadů je tedy vyšší než nominální.

Z hlediska aplikací je ideální, aby nestrannost, konzistence případně konzervativnost odhadu rozptylu vyplývaly z použitého designu a nebyly nijak podmíněny vlastnostmi populace tj. vlastnostmi funkce lokální hustoty $Y(x)$. Bohužel, pro designy nesplňujících podmínku (30) (sekce A, strana 20) je takový stav buďto přímo vyloučen (nestranné odhady), nebo jde o jev spíše výjimečný. Výhodou však zůstává, pokud je alespoň některá z pozitivních vlastností odhadu podmíněna takovými vlastnostmi lokální hustoty, které lze pro danou populaci reálně očekávat - například předpoklad kladné prostorové korelace lokální hustoty při výběrových šetřeních v oblasti životního prostředí.

Pro výběry nesplňující podmínku (30) je běžně používána aproximace $\pi(x_i, x_j) = n(n-1)/\lambda(D)$ odpovídající zcela náhodnému výběru inventarizačních bodů v D viz (38) (příloha A.1, strana 23). Odhad rozptylu parametru θ populace získaný s touto aproximací je ve většině případů NIL (pozitivní prostorová korelace lokální hustoty), nikoli však zákonitě (možná periodicitu funkce lokální hustoty), konzervativní tj. platí

$$\mathbb{E}[\hat{V}_{URS}(\hat{\theta})] > \mathbb{V}(\hat{\theta}) . \quad (95)$$

E Odhad úhrnu nestranný nepodmíněně na velikosti výběru

Při tomto přístupu k nekonstantní velikosti výběru v D se vychází ze znalosti případně aproximace očekávané hustoty $\bar{\pi}(x)$ výběru na bodě x a párové hustoty $\bar{\pi}(x_i, x_j)$, které ve vzorcích (28), (29), (31), (32) nahradí hustoty $\pi(x)$ a $\pi(x_i, x_j)$ viz Cordy [1993, sekce 3, strana 359].

$$\bar{\pi}(x) = \mathbb{E}[\pi_{n_D}(x)] = \sum \alpha_{n_D} \pi_n(x) \quad (96)$$

$$\bar{\pi}(x_i, x_j) = \mathbb{E}[\pi_{n_D}(x_i, x_j)] = \sum \alpha_{n_D} \pi_n(x_i, x_j) \quad (97)$$

Sumy ve vzorcích (96) a (97) jsou přes všechny možné velikosti výběru, α_{n_D} je pravděpodobnost, že při realizaci výběrového designu získáme výběr o velikosti n_D (právě n_D inventarizačních bodů padne dovnitř D).

Očekávanou hodnotu $\bar{\pi}(x)$ lze v některých případech získat i bez explicitní znalosti pravděpodobností α_{n_D} . Pro neinformativní výběrové designy, kdy výběr bodu x nijak nezávisí na hodnotě hustoty $Y(x)$ ani na veličinách s $Y(x)$ korelovaných, platí

$$\pi_{n_D}(x) = n_D \lambda^{-1}(D) . \quad (98)$$

Výraz $\lambda(D)$ představuje velikost (plochu, Lebesgueovu míru) domény D . Po vytknutí konstanty $\lambda^{-1}(D)$ můžeme místo (96) psát

$$\bar{\pi}(x) = \lambda^{-1}(D) \sum \alpha_{n_D} n_D = \bar{n}_D \lambda^{-1}(D) , \quad (99)$$

kde \bar{n}_D je očekávaná (střední) hodnota velikosti výběru. Pro \bar{n}_D zjevně platí

$$\bar{n}_D = \lambda(D) \lambda^{-1}(c) , \quad (100)$$

kde $\lambda(c)$ je pevná velikost inventarizačního bloku. Dosazením (100) do (99) dojde k vykrácení $\lambda(D)$. Očekávanou hodnotu hustoty výběru bodu x lze ve shodě s intuitivní představou získat dle (101) jako převrácenou hodnotu pevné velikosti inventarizačního bloku

$$\bar{\pi}(x) = \lambda(c)^{-1} . \quad (101)$$

Pro některé v praxi často používané inventarizační sítě nesplňuje hustota $\pi_{n_D}(x_i, x_j)$ podmínku (30) (sekce A), proto je zpravidla používána aproximace odpovídající zcela náhodnému výběru. Je-li velikost výběru v zájmové oblasti D nekonstantní, můžeme hustotu $\bar{\pi}(x_i, x_j)$ aproximovat 102

$$\bar{\pi}(x_i, x_j) \approx \bar{n}_D (\bar{n}_D - 1 + \mathbb{V}(n_D) / \bar{n}_D) \lambda^{-2}(D) . \quad (102)$$

Vychází se z konstrukce, kdy je pro každý z hypoteticky opakovaných výběrů náhodně velikosti n_D uvažována párová hustota $\pi_{n_D}(x_i, x_j) = n_D(n_D - 1) / \lambda^2(D)$. Pro očekávanou hodnotu párové hustoty $\bar{\pi}(x_i, x_j)$ (přes všechny možné velikosti výběru) pak můžeme psát:

$$\bar{\pi}(x_i, x_j) \approx \mathbb{E} \left[\frac{n_D(n_D - 1)}{\lambda^2(D)} \right] = \frac{\mathbb{E}(n_D^2) - \mathbb{E}(n_D)}{\lambda^2(D)} = \frac{\mathbb{E}(n_D^2) - \bar{n}_D}{\lambda^2(D)} . \quad (103)$$

Očekávaná hodnota $\mathbb{E}(n_D)$ je dána (100). Očekávanou hodnotu $\mathbb{E}(n_D^2)$ vyjádříme s pomocí rovnice (104) viz. Anděl [1978, vzorec 12 na straně 14]

$$\mathbb{V}(n_D) = \mathbb{E}(n_D^2) - \mathbb{E}^2(n_D) . \quad (104)$$

Rozptyl $\mathbb{V}(n_D)$ počtu inventarizačních bodů vybraných v D sice přesně neznáme, ale pro zájmové oblasti vymezené mapou jej, stejně jako očekávanou hodnotu $\mathbb{E}(n_D^2)$, můžeme velmi přesně odhadnout na základě simulovaných výběrů. Dosazením do (103) získáme aproximaci (102).

Přibližnou hodnotu $\mathbb{E}(n_D^2)$ můžeme odvodit také na základě Taylorovy aproximace funkce n_D^2 v okolí $\mathbb{E}(n_D^2)$ viz (105) a (106)

$$n_D^2 \approx \mathbb{E}(n_D^2) + 2\sqrt{\mathbb{E}(n_D^2)} \left\{ n_D - \sqrt{\mathbb{E}(n_D^2)} \right\}, \quad (105)$$

$$\begin{aligned} \mathbb{E}(n_D^2) &\approx \mathbb{E}(n_D^2) + 2\sqrt{\mathbb{E}(n_D^2)}\mathbb{E}(n_D) - 2\mathbb{E}(n_D^2) \\ &\approx \mathbb{E}^2(n_D) = \bar{n}_D^2. \end{aligned} \quad (106)$$

Výsledek odpovídá řešení rovnice (104) se zanedbáním hodnoty rozptylu $\mathbb{V}(n_D)$. Aproximace \bar{n}_D^2 podhodnocuje $\mathbb{E}(n_D^2)$ právě o hodnotu rozptylu $\mathbb{V}(n_D)$. Dosazením \bar{n}_D^2 za $\mathbb{E}(n_D^2)$ do (103) získáme aproximaci (107)

$$\bar{\pi}(x_i, x_j) \approx \frac{\bar{n}_D(\bar{n}_D - 1)}{\lambda^2(D)}. \quad (107)$$

Vzhledem k relativně velkým výběrům NIL a vzhledem k použití inventarizačních sítí, které rovnoměrně pokrývají zájmové oblasti, se nabízí efekt podhodnocení $\mathbb{E}(n_D^2)$ zanedbat. Důsledkem je však neznámé snížení konzervativnosti odhadů rozptylu pod úroveň, která by odpovídala samotnému nahrazení párové hustoty hustotami pro náhodný výběr n_D bodů. Proto při vyhodnocení NIL dle této metodiky pracujeme s aproximací párové hustoty (102) s tím, že rozptyl $\mathbb{V}(n_D)$ je odhadován pomocí simulovaných výběrů.

Vzorce odvozené pro zcela náhodný výběr pevného počtu n bodů v D (viz příloha A.1 od strany 23) upravíme na základě (99) a (102) do podoby zohledňující nekonstantní velikost výběru následovně:

$$\hat{Y} = \sum_{\substack{x \in s \\ x \in D}} \frac{Y(x)}{\bar{\pi}(x)} = \lambda(D) \frac{\sum_{x \in s} Y(x)}{\bar{n}_D} = \lambda(c) \sum_{x \in s} Y(x) = \lambda(D) \hat{Y}, \quad (108)$$

$$\hat{\mathbb{V}}(\hat{Y}) = \frac{\lambda^2(D)}{\bar{n}_D - 1 + \hat{\mathbb{V}}(n_D)/\bar{n}_D} \left\{ \widehat{Y}^2 + \hat{Y}^2 \left[\hat{\mathbb{V}}(n_D)/\bar{n}_D - 1 \right] \right\}. \quad (109)$$

Vzorec (109) získáme použitím (99) a (102) v rámci odvození odhadu rozptylu ve vzorcích (47) až (51) (příloha A.1, strana 23). Symbol \widehat{Y}^2 je podíl sumy čtverců lokální hustoty a očekávaného počtu \bar{n}_D inventarizačních bodů v D . Člen \hat{Y}^2 označuje druhou mocninu podílu sumy lokálních hustot a \bar{n}_D .

$$\mathbb{E}(\hat{Y}) = Y \quad (110)$$

Odhadu rozptylu je při aproximaci 102 pro (quasi-)systematické nebo prostorově stratifikované sítě a za předpokladu pozitivní prostorové korelace hodnot lokální hustoty konzervativní nepodmíněně na velikosti výběru v D , tj. platí

$$\mathbb{E}[\hat{\mathbb{V}}(\hat{Y})] > \mathbb{V}(\hat{Y}). \quad (111)$$

Odhady nestranné nepodmíněně na velikosti výběru v D jsou vhodné zejména v případě, kdy je potřeba porovnat výběrové designy na základě očekávané přesnosti. Při této úloze totiž nelze vycházet z konkrétní velikosti výběru v D , protože ta při budoucím, reálném provedení výběru vůbec nemusí nastat. Z hlediska vyhodnocení již realizovaných výběrových šetření jsou tyto odhady vhodné tam, kde je očekávána nebo požadována aditivita odhadů, viz příloha F od strany 34.

Pro konkrétní velikost výběru $n_D \neq \bar{n}_D$ tvrzení o nestrannosti bodového odhadu úhrnu neplatí, což je jistou nevýhodou v porovnání s odhady nestrannými (konzervativními) podmíněně na velikosti výběru. V podmínkách odhadů v síti NIL1 (nízká variabilita velikosti výběru v D , silná konzervativnost odhadu rozptylu) zůstává tato nevýhoda pouze v teoretické rovině - hodnoty vychýlení a navýšení rozptylu jsou prakticky zanedbatelné. Další komentář týkající se NIL je uveden v závěru sekce C od strany 29.

F Geografická a atributová aditivita odhadu

Geografickou aditivitu odhadu vyjadřuje rovnice (112)

$$\hat{\theta}_{\bigcup_{i=1}^k D_i} = \sum_{i=1}^k \hat{\theta}_{D_i}, \quad (112)$$

podle níž je odhad vypočtený pro libovolnou geografickou doménu danou sjednocením $\bigcup_{i=1}^k D_i$ domén $D_1, D_2 \dots D_k$ přesně roven součtu odhadů vypočtených jednotlivě pro každou z k uvažovaných domén.

Atributová aditivita je definována analogicky:

$$\hat{\theta}_{\bigcup_{j=1}^j A_j} = \sum_{j=1}^j \hat{\theta}_{A_j}, \quad (113)$$

kde A_j značí jednotlivé atributové domény (například skupiny dřevin, či kategorií přirozenosti lesního porostu).

Atributově aditivní jsou všechny odhady úhrnu a velká část odhadů podílu počítaných dle této metodiky. Porušení aditivity podílu je logickým důsledkem atributového třídění odhadu ve jmenovateli - aditivita je splněna pouze v rámci podskupin atributového členění vymezených společnou atributovou doménou jmenovatele. Příkladem je zásoba dříví na hektar porostní půdy členěná dle dřevin (čítatel podílu) a kategorií vlastnictví lesa (jmenovatel podílu), kdy jsou odhady dle dřevin aditivní pouze v rámci jednotlivých kategorií vlastnictví.

Aditivita bývá preferována z pohledu koncového uživatele výsledků NIL. Výhodou aditivních odhadů je možnost implementace kontrolních mechanismů procesu vyhodnocení. Nejsou-li výsledné odhady geograficky či atributově aditivní, muselo dojít k chybám při přípravě dat či samotném výpočtu. Kontrola aditivity je velmi přínosná při velkém počtu vyhodnocovaných oblastí a atributových domén (případ NIL).

G Intervalový odhad a jeho statistická jistota

Statistická jistota intervalového odhadu $CI = [CI^-; CI^+]$ se shoduje s pravděpodobností P ve vzorci (114), kde CI^- a CI^+ představují spodní respektive horní mez odhadu. Design-based přístup tuto pravděpodobnost definuje jako očekávaný podíl intervalových odhadů, které při opakovaném provedení výběru se stejným designem (na základě každého z uvažovaných výběrů je konstruován jeden intervalový odhad) zachytí skutečnou hodnotu neznámého parametru θ populace. Obecným požadavkem je, aby skutečná statistická jistota intervalového odhadu odpovídala zvolené nominální hodnotě $1 - \alpha$, kde α (hladina významnosti) představuje podíl intervalových odhadů, které nezachytí skutečnou hodnotu parametru θ .

$$P(CI^- \leq \theta \leq CI^+) = 1 - \alpha \quad (114)$$

Intervalové odhady podle této metodiky jsou konstruovány následovně:

$$CI = \hat{\theta} \pm z_{(1-\alpha/2)} \sqrt{\hat{V}(\hat{\theta})} \quad (115)$$

kde $\hat{\theta}$ představuje bodový odhad neznámého parametru populace θ , $\hat{V}(\hat{\theta})$ je odhad rozptylu $V(\hat{\theta})$ a $z_{(1-\alpha/2)}$ je $(1 - \alpha/2)$ kvantil standardizovaného normálního rozdělení. Dle Särndala a kol. [2003, kapitola 2.11, strana 56] bude intervalový odhad (115) obsahovat neznámou hodnotu parametru θ pro přibližně $(1 - \alpha)\%$ případů opakování výběru, pokud jsou splněny následující podmínky:

1. Rozdělení pravděpodobnosti bodového odhadu $\hat{\theta}$ je přibližně normální s očekávanou (střední) hodnotou θ a rozptylem $V(\hat{\theta})$.
2. Existuje konzistentní odhad $\hat{V}(\hat{\theta})$ rozptylu $V(\hat{\theta})$.

První podmínka požaduje, aby podíl (116) měl přibližně standardizované normální rozdělení $\mathcal{N}(0, 1)$

$$\frac{\hat{\theta} - \theta}{\sqrt{V(\hat{\theta})}} \quad (116)$$

Aproximaci rozdělení pravděpodobnosti bodového odhadu rozdělením normálním lze zdůvodnit platností centrálních limitních vět. Za určitých, velmi obecných podmínek mají součty nezávislých, náhodných veličin asymptoticky normální rozdělení [Rao, 1965, sekce 2c.5, strana 126], [Anděl, 1978, kapitola X.3, strana 184]. Bodový odhad dle (28) a řada dalších jsou velmi často takovými součty, proto lze jejich rozdělení pro dostatečně velké výběry aproximovat normálním. Neexistuje však obecně platné pravidlo pro posouzení, zda je velikost výběru dostatečná pro použití normální aproximace [Cochran, 1977, kapitola 2.15, od strany 39], [Särndal *et al.*, 2003, poznámky 2.11.2 a 2.11.3 na straně 57], [Gregoire & Valentine, 2008, kapitola 2.3, strana 31]. Cochran [1977, kapitola 2.15, strana 42] navrhuje přibližný vztah (117) pro hrubou identifikaci případů, kdy na normální aproximaci nelze spoléhat

$$n_{min} > 25G_1^2, \quad (117)$$

kde G_1 je Fisherova míra šikmosti a n_{min} je minimální požadovaná velikost výběru. Postup je použitelný pro populace, jejichž odchylka od normality je dána především asymetrií rozdělení. Vzorec (117) je dle Cochranu navržen tak, aby 95% intervalové odhady neminuly skutečnou polohu střední hodnoty ve více jak 6% případech.

Použití tohoto kritéria bylo testováno na simulovaných odhadech celkové a hektarové zásoby dříví, a dále plochy lesa. Při simulacích byly opakovaně vybírány inventarizační body dle designu sítě NIL1 (celkem 2000 simulací výběru v síti NIL1). První modelovou populací byla zásoba dříví dle databáze lesních hospodářských plánů a osnov pokrývající celé území ČR (přibližně 2,7 mil. ha). Druhou byla mapa výskytu kategorie pozemků *Les*, která vznikla v rámci prvního cyklu NIL digitalizací ortofotomap získaných leteckým snímkováním (mapa OLIL, taktéž zhruba 2,7 mil. ha území v rámci ČR, 96% tematická správnost). Odhady byly počítány pro úroveň ČR, krajů i okresů, v případě zásoby dříví též pro atributové domény „jehličnany“ a „listnáče“.

Pokud kritérium (117) indikovalo dostatečnou velikost výběru, nebyla pro žádnou z kombinací odhadovaného parametru, geografické a atributové domény zjištěna nižší než Cochranem deklarovaná minimální statistická jistota 94%. Nedostatečná velikost výběru

byla detekována při odhadech v okresech pro málo zastoupené atributové domény - například odhad celkové zásoby jehličnanů v okrese Břeclav, či odhad zásoby listnáčů v okrese Žďár nad Sázavou. Ve více než 80% případů, kdy byla detekována nedostatečná velikost výběru, nedošlo k poklesu statistické jistoty pod 94%. Tento výsledek přičítáme současně konzervativnosti (navýšení) odhadu rozptylu, viz příloha C.

Existují-li pochybnosti o správnosti postupu využívajícího normální aproximace, nabízí se alternativní řešení založené na transformaci hodnot lokální hustoty s cílem přiblížení jejího rozdělení pravděpodobnosti k rozdělení normálnímu. Po provedení vhodné transformace lokální hustoty lze vypočítat bodové a intervalové odhady a tyto následně zpětně transformovat tak, aby odpovídaly původní proměnné.

Problematiku transformací na normální rozdělení a jejich využití pro konstrukci intervalových odhadů lze dohledat v práci Melouna a Militkého [1998, sekce 2.1.4 a sekce 2.1.5 od strany 85 a 90]. Sakia [1992] publikoval přehled věnovaný Box-Cox transformacím, které patří mezi nejčastěji používané transformace vůbec. Li [2008] se ve své práci zabýval použitím Box-Cox transformace ve spojení s generalizovaným regresním odhadem.

Odhady počítané na základě transformovaných veličin nejsou aditivní v doménách, mezi nimiž byly provedeny různé transformace lokálních hustot (týká se geografické i atributové aditivity, viz příloha F. Proto nejsou transformace při vyhodnocení NIL rutinně používány.

Alternativa, kterou navrhují Gregoire & Valentine [2008], spočívá v konstrukci konzervativního intervalového odhadu na základě vzorce (118) vycházejícího z Čebyševova teorému

$$P \left[\hat{\theta} - k\sqrt{\mathbb{V}(\hat{\theta})} \leq \theta \leq \hat{\theta} + k\sqrt{\mathbb{V}(\hat{\theta})} \right] \geq 1 - \frac{1}{k^2}. \quad (118)$$

Horní a dolní mez intervalového odhadu jsou počítány dle vzorce (119), který stejně jako (115) používá odhad $\hat{\mathbb{V}}(\hat{\theta})$ na místo skutečné hodnoty rozptylu $\mathbb{V}(\hat{\theta})$

$$CI_t = \hat{\theta} \pm k\sqrt{\hat{\mathbb{V}}(\hat{\theta})}. \quad (119)$$

Tato konstrukce intervalového odhadu nevyžaduje splnění žádných předpokladů o rozdělení pravděpodobnosti bodového odhadu. Porovnáním hodnoty $k \doteq 4.47$ přibližně odpovídající statistické jistotě 95% ($\alpha = 0.05$) s kvantilem standardizovaného normálního rozdělení $z_{1-\alpha/2} \doteq 1.96$ používaným v rámci (115) je na první pohled zřejmé, že cenou za obecnost řešení jsou podstatně širší intervalové odhady.

Poněkud užší meze intervalového odhadu lze stanovit na základě teorému Vysochanského a Petunina

$$P \left[\hat{\theta} - k\sqrt{\mathbb{V}(\hat{\theta})} \leq \theta \leq \hat{\theta} + k\sqrt{\mathbb{V}(\hat{\theta})} \right] \geq 1 - \frac{2}{3k^2}, \quad (120)$$

kdy horní a dolní mez intervalového odhadu jsou počítány taktéž dle vzorce (119), ale s dosazením odpovídající hodnoty k (2.98 pro 95% statistickou jistotu). Šířka intervalového odhadu je zde na úrovni 2/3 šířky intervalu konstruovaného dle Čebyševova teorému. Tento postup je možný pouze pro parametry, u nichž lze předpokládat unimodální rozdělení pravděpodobnosti s konečným rozptylem. Domníváme se, že tyto vlastnosti má většina - ne-li všechny - odhady NIL.

Společným „nedostatkem“ jak postupu na základě Čebyševova, tak i Vysochanského-Petuninova teorému, je dosazení odhadnuté hodnoty rozptylu do (119) namísto hodnoty skutečné, kterou při výběrovém šetření NIL neznáme. Při potenciálním podhodnocení

odhadu rozptylu není garantována požadovaná statistická jistota. Faktem je, že odhady rozptylu NIL (zejména jednofázové odhady úhrnu) bývají nadhodnocené, viz příloha D. Domníváme se také, že jsou konzistentní (viz dále). Na druhou stranu, běžné velikosti výběru NIL ve standardně uvažovaných geografických doménách (kraje) se pohybují v řádu stovek až několika tisíc inventarizačních bodů. Proto předpokládáme, že intervalové odhady na základě normální aproximace (115) jsou v případě NIL a jí publikovaných odhadů vyhovující.

Z první podmínky formulované na straně 35 dále vyplývá, že odhad $\hat{\theta}$ má být nestranný. Není-li tomu tak, jedná se o odhad vychýlený, což znamená, že při konkrétní velikosti výběru existuje nenulový rozdíl očekávané $\mathbb{E}(\hat{\theta})$ a skutečné hodnoty parametru θ populace. Tento rozdíl nazýváme vychýlením odhadu [Anděl, 1978, kapitola XV.1, strana 254] (angl. bias). Särndal *et al.* [2003, kapitola 5.2, od strany 163] kvantifikovali vliv vychýlení bodového odhadu na skutečnou statistickou jistotu intervalového odhadu (115). Autoři konstatují, že při nominální statistické jistotě $(1 - \alpha) = 0.95$ a 10% vychýlení vzhledem ke směrodatné odchylce bodového odhadu činí skutečná statistická jistota 0.9489. Ještě při poměrně vysokých hodnotách vychýlení mezi 10% až 50% směrodatné odchylky bodového odhadu neklesá skutečná statistická jistota intervalového odhadu pod 0.9210. Srovnatelný přehled publikoval již dříve Cochran [1977, kapitola 1.8, od strany 12]. Poněvadž je odhad (28) nestranný, může při jeho správném použití dojít k vychýlení pouze z důvodu systematických chyb šetření a jiných nedostatků sběru a zpracování dat.

Druhá podmínka požadující konzistenci odhadu rozptylu je nezbytná, poněvadž při konstrukci intervalového odhadu zpravidla neznáme skutečnou hodnotu rozptylu $\mathbb{V}(\hat{\theta})$. Na aproximaci normálním rozdělením tak spoléháme ačkoli je ve jmenovateli (116) použit odhad $\hat{\mathbb{V}}(\hat{\theta})$ místo $\mathbb{V}(\hat{\theta})$. Särndal *et al.* [2003, kapitola 2.11, strana 56] význam konzistence vysvětlují pomocí (121).

$$\frac{\hat{\theta} - \theta}{\sqrt{\hat{\mathbb{V}}(\hat{\theta})}} = \frac{\hat{\theta} - \theta}{\sqrt{\mathbb{V}(\hat{\theta})}} \left[\frac{\mathbb{V}(\hat{\theta})}{\hat{\mathbb{V}}(\hat{\theta})} \right]^{1/2} \quad (121)$$

Za předpokladu konzistence $\hat{\mathbb{V}}(\hat{\theta})$ a pro dostatečně velké výběry se $[\mathbb{V}(\hat{\theta})/\hat{\mathbb{V}}(\hat{\theta})]^{1/2}$ s vysokou pravděpodobností blíží jedné. Tato tendence je dále podtržena odmocninou ve vzorci (121). Chybu z nahrazení rozptylu $\mathbb{V}(\hat{\theta})$ jeho odhadem $\hat{\mathbb{V}}(\hat{\theta})$ tak při dostatečné velikosti výběru zanedbáváme. Ačkoli bývají nestranné odhady často zároveň i konzistentní, nejedná se o pravidlo. Proto je třeba upozornit, že samotná nestrannost odhadu rozptylu není postačující podmínkou pro konstrukci intervalového odhadu dle (115), jehož statistická jistota má odpovídat nominální hodnotě $(1 - \alpha)$.

Statistická jistota úzce souvisí s konzervativností odhadu rozptylu. Ta je často očekávána pouze na základě určitých předpokladů o populaci, které ale nemusí být ve skutečnosti vůbec splněny. V závislosti na v praxi ne vždy zcela známých vlastnostech populace může být zvolený odhad rozptylu až příliš konzervativní nebo naopak antikonzervativní, čehož si při interpretaci výsledků nemusíme být vědomi. V prvním případě získáme neúměrně široký intervalový odhad v porovnání s šířkou, která by odpovídala nominální statistické jistotě $(1 - \alpha)$ - uváděná přesnost cílového parametru je nižší než skutečná. Ve druhém případě je šířka intervalového odhadu naopak nedostatečná - skutečná statistická jistota nedosahuje nominální hodnoty $(1 - \alpha)$ a odhad cílového parametru se jeví přesnější než ve skutečnosti je. Někdy může být vhodnější použít mírně antikonzervativní než silně konzervativní odhad rozptylu - volba zpravidla závisí na aplikaci, na případném splnění či nesplnění předpokladů, za kterých je ten který odhad (anti)konzervativní a samozřejmě i na dalších vlastnostech odhadu rozptylu (garance kladných hodnot, stabilita).