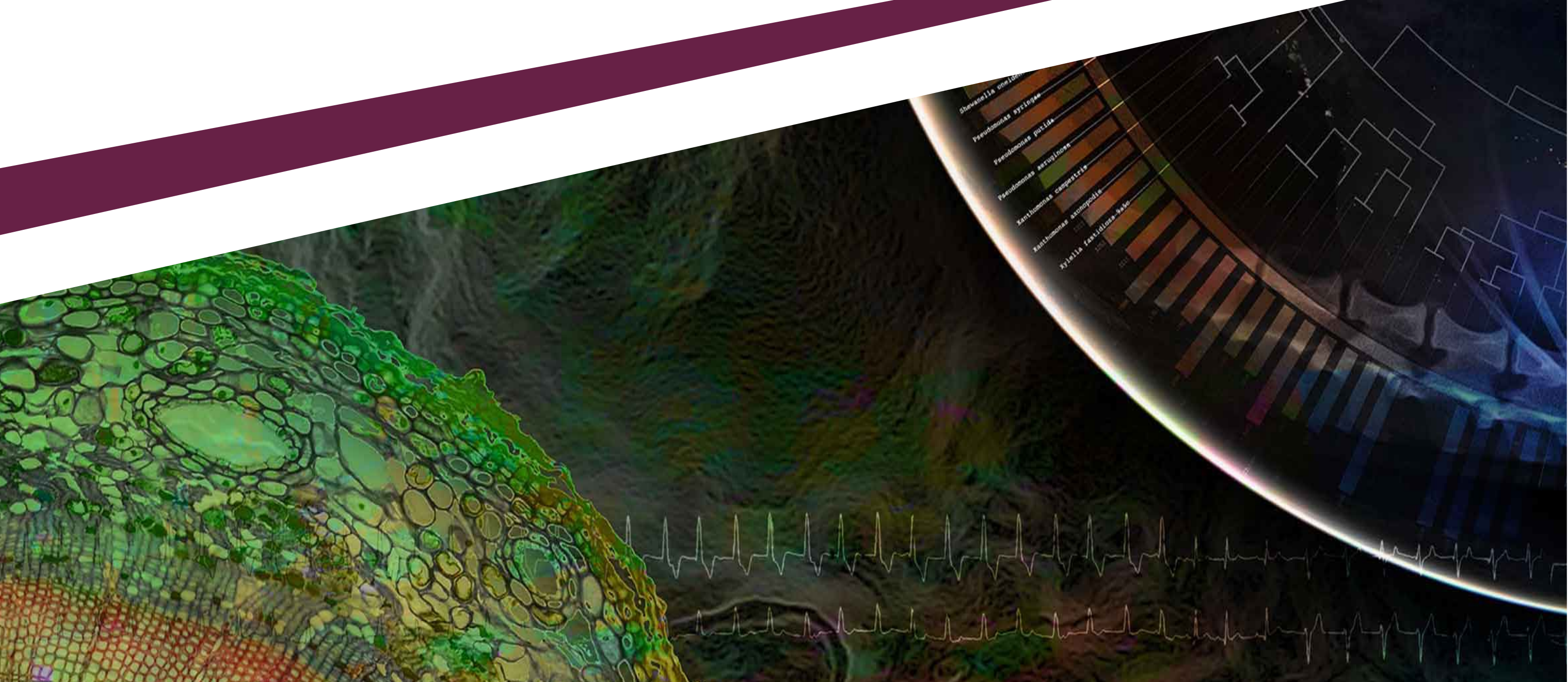
Skattningar baserade på hierarkiska urval

Kombinera data från flera källor

Anton Grafström, Åsa Ranlund, Sven Adler

Sveriges lantbruksuniversitet, SLU
Institutionen för skoglig resurshushållning
Arbetsrapport / Sveriges lantbruksuniversitet, Institutionen för skoglig resurshushållning, 558
ISBN: 1401-1204
Umeå 2023



Skattningar baserade på hierarkiska urval. Kombinera data från flera källor

Anton Grafström, https://orcid.org/0000-0002-4345-4024, Sveriges lantbruksuniversitet, Institutionen för skoglig resurshushållning,

Åsa Ranlund, https://orcid.org/0000-0002-7197-8570, Sveriges lantbruksuniversitet, Institutionen för skoglig resurshushållning,

Sven Adler, https://orcid.org/0000-0002-5414-8727, Sveriges lantbruksuniversitet, Institutionen för skoglig resurshushållning,

Utgivare:	Sveriges lantbruksuniversitet, Institutionen för Skoglig resurshushållning
Utgivningsår:	2023
Utgivningsort:	Umeå
Serietitel:	Arbetsrapport / Sveriges lantbruksuniversitet, Institutionen för skoglig resurshushållning
Delnummer i serien:	558
ISSN:	1401-1204
Nyckelord:	miljöövervakning, statistik, design, skattningar

Sammanfattning

En flexibel design för miljöövervakning baserad på heltäckande data från fjärranalys har utvecklats inom Nationell Inventering av Landskap i Sverige (NILS). Denna design syftar till att effektivt skatta tillstånd och upptäcka förändringar över tid. NILS nya system använder olika stickprovstätheter för att inventera vanliga och ovanliga fenomen i gräsmarker, lövskogar, fjäll och havsstränder. Målet är att kombinera data för bättre skattningar av mindre vanliga men viktiga fenomen och få förbättrad övervakning över tid. Statistisk teori har utvecklats för att optimera användningen av data från olika stickprovstätheter inom inventeringarna.

Nyckelord: miljöövervakning, statistik, design, skattningar

Abstract

A flexible design for environmental monitoring, based on comprehensive data from remote sensing, has been developed within the National Inventory of Landscapes in Sweden (NILS). This design aims to effectively assess state and detect changes over time. NILS's new system utilizes different sampling densities to survey common and rare phenomena in grasslands, deciduous forests, mountains, and coastal zones. The objective is to combine data for improved estimations of less common but significant phenomena and achieve enhanced monitoring over time. Statistical theory has been developed to optimize the utilization of data from various sampling densities within the inventories.

Keywords: environmental monitoring, statistics, design, estimators

Innehåll

1 Introduktion

Samtidigt som vår miljö förändras, förändras förutsättningarna för miljöövervakning: Nya mål tillkommer, de ekonomiska förutsättningarna skiftar och teknisk utveckling skapar nya möjligheter. Det är en utmaning för miljöövervakning, där en grundpelare också är att bygga stabila tidsserier av data med hög kvalitet som är jämförbara över tid. Sammantaget ställer det höga krav på miljöövervakning att både vara flexibel och bidra med kontinuitet av miljödata. Där tidigare miljöövervakningsprogram varit bra på att bygga stabila tidsserier av miljödata behöver ny miljöövervakning ta höjd för hur både miljön och miljöövervakningens förutsättningar kommer att förändras. Genom att från början använda en flexibel design för miljöövervakning ökar våra chanser att i framtiden kunna samla in data med kontinuitet, även i en föränderlig omvärld. Arbetet med att skapa ett mer flexibelt system för miljöövervakning har pågått sedan 2019 inom Nationella Inventeringar av Landskapen i Sverige (NILS) och bygger på erfarenheter från både mer traditionell miljöövervakning och utvecklingsprojekt.

Tillgången till heltäckande information baserad på data från fjärranalys, till exempel satellitbilder, LiDAR-data och flygbilder, erbjuder stora möjligheter för att skapa effektiva inventeringar. Teorin för att nyttja denna typ av information i designen av en inventering är relativt välutvecklad. Riksskogstaxeringens tillfälliga utlägg använder sedan 2018 en sådan design, se Grafström et al (2017). En design som nyttjar heltäckande information blir mer effektiv för att skatta tillstånd, men är precis som ett permanent urval inte anpassad för att på ett effektivt sätt följa förändring över längre tid. För att ett urval ska vara effektivt över längre tid krävs att det uppdateras i takt med att populationen förändras.

Att kunna nyttja ny och bättre heltäckande information när den blir tillgänglig, utan att designen förlorar sin styrka att upptäcka förändring, ställer stora krav på inventeringens design. Om ett nytt oberoende urval görs ökar variansen dramatiskt på en skattning av en förändring mellan tidigare urval och det nya urvalet, vilket gör det i princip omöjligt att upptäcka en relativt stor förändring i populationen. Här ligger också den största utmaningen för teoretisk utveckling: Att hitta ett system för miljöövervakning som är förhållandevis enkelt, samtidigt som det är tillräckligt flexibelt och effektivt för att följa upp förändring över tid. NILS nya system av inventeringar är unikt på så sätt att det har utvecklats med syftet att på bästa möjliga sätt möta dessa krav.

NILS nya system för riktade inventeringar bygger på en hierarki med olika stickprovstätheter och har implementerats i NILS fjäll-, havsstrand-, gräsmarks- och lövskogsinventeringar (Adler m.fl. 2020, Adler m.fl. 2021, Ranlund m.fl. 2021). Gemensamt för alla dessa fyra inventeringar är att de ska inventera både relativt vanliga fenomen och sådant som är mer ovanligt. Vanliga fenomen inventeras i ett glest stickprov och ovanligare fenomen inventeras i tätare stickprov. Det behövs en metod för att nyttja data från alla stickprovstätheter för att bättre skatta dessa fenomen samt att kombinera data mellan inventeringar.

Eftersom flexibilitet kan komma med en kostnad i komplexitet försöker vi i den här rapporten reda ut vilka skattningar som är aktuella under olika förutsättningar inom den nya designen. Syftet är att möjliggöra användandet av (allt) insamlat data i förbättrade skattningar av mindre vanliga men viktiga fenomen, genom att kombinera data inom NILS nya system av inventeringar samt Riksskogstaxeringen.

Vi har i denna rapport tagit fram statistisk teori för skattningar som på bästa sätt nyttjar tillgängliga data från flera stickprovstätheter inom en eller flera inventeringar.

2 Skattningar med sannolikhetsurval

Sannolikhetsurval används för att välja ett urval (en delmängd) från en population och sedan dra slutsatser om populationen baserat på urvalet, utan att behöva mäta hela populationen. De skattningsformler som vi presenterar här bygger alla på ett sannolikhetsurval från den population vi är intresserad av. Ett grundkrav vid sannolikhetsurval är att alla enheter i populationen ska ha en positiv sannolikhet att väljas (för en ändlig population). En skattningsformel kallas även för en estimator och den används för att skatta en populationsparameter. Vanligast är att vi vill skatta en populationstotal (summan över enskilda förekomster eller integralen) av en variabel. En populationstotal kan till exempel vara arealen skogsmark eller totalt antal björkar (över en viss diameter) inom ett visst område. För en ändlig population av N enheter (som björkarna) kan totalen skrivas som

$$Y = \sum_{i=1}^{N} y_i,$$

där y_i är värdet för enhet i i populationen. Om björkar är vår population och vi vill skatta antalet är $y_i = 1$ för alla björkar (och $Y = N$).

Om vi vill skatta en areal har vi (oftast) en kontinuerlig population och en kontinuerlig urvalsram, som då innehåller ett icke uppräkneligt oändligt antal punkter $\mathbf{x}$. En total kan då skrivas som

$$Y = \int_{\mathbf{x} \in F} y(\mathbf{x}) d\mathbf{x},$$

där F betecknar urvalsramen (området där vi slumpar ut punkter). Om vi vill skatta total areal skogsmark låter vi $y(\mathbf{x}) = 1$ om punkten $\mathbf{x}$ ligger i skogsmark och $y(\mathbf{x}) = 0$ annars. Då är Y total areal skogsmark inom urvalsramen F.

Eftersom vi vanlitvis inte kan mäta hela populationen använder vi ett sannolikhetsurval och skattar totalen med en estimator. En ofta använd estimator kallas för Horvitz-Thompson (HT) estimatorn. Denna estimator av totalen Y betecknar vi här som $\widehat{Y}$ och den är för en ändlig population definierad som

$$\widehat{Y} = \sum_{i \in S} \frac{y_i}{\pi_i}, \tag{1}$$

där $\pi_i = \Pr(i \in S)$ är sannolikheten att enhet i väljs till stickprovet S och y_i är värdet på den intressanta variablen för enhet i. Vi kan även skriva skattningen på formen

$$\widehat{Y} = \sum_{i=1}^{N} \frac{y_i}{\pi_i} I_i, \tag{2}$$

där $I_i = 1$ om enhet i tillhör stickprovet S och $I_i = 0$ annars. Formen (2) är en enklare form matematiskt än (1) då vi i (2) alltid summerar över hela populationen. Givet att vi kan mäta y_i för de utvalda enheterna utan fel och att $\pi_i > 0$ för alla enheter är HT-estimatorn väntevärdesriktig, vilket betyder att den skattar rätt värde på populationsparametern Y i genomsnitt över upprepade urval. Vi kan också säga att HT-estimatorn inte har något systematiskt fel (bias). För en kontinuerlig population kan estimatorn skrivas på formen

$$\widehat{Y} = \sum_{S} \frac{y(\mathbf{x})}{\pi(\mathbf{x})}, \tag{3}$$

där $\pi(\mathbf{x})$ är samplingintensiteten i punkten $\mathbf{x}$. Samplingintensiteten kan till exempel vara $\pi(\mathbf{x}) = 1/10000$ för alla punkter $\mathbf{x}$ i urvalsramen. Det kan betyda att vi till stickprovet i genomsnitt väljer en punkt per 10000 m^2 (eller en punkt per hektar). För att (3) ska vara väntevärdesriktig krävs att $y(\mathbf{x})$ mäts utan fel samt att mängden punkter där $\pi(\mathbf{x}) > 0, \mathbf{x} \in F$ har samma storlek som F.

En estimator är en slumpvariabel som kan anta olika värden för olika stickprov. Det genomsnittliga värdet på $\widehat{Y}$ över upprepade urval kallas väntevärdet av $\widehat{Y}$ och betecknas ofta med $E\left(\widehat{Y}\right)$, där E står för "Expected value". För HT-estimatorn gäller alltså att $E\left(\widehat{Y}\right) = Y$. En slumpvariabel $\widehat{Y}$ har också en varians (då den kan anta olika värden för olika urval). Variansen betecknas ofta med $V\left(\widehat{Y}\right)$. Denna varians är (nästan alltid) okänd, men kan ofta skattas med hjälp av observationerna i stickprovet.

3 Kombinera skattningar från olika källor

Givet att det går att skatta en populationstotal Y med två olika skattningar $\widehat{Y}_1$ och $\widehat{Y}_2$ kan en kombinerad skattning göras med linjärkombination

$$\widehat{Y} = w_1 \widehat{Y}_1 + w_2 \widehat{Y}_2, \tag{4}$$

där $0 \leq w_1 \leq 1$ är en fix vikt för skattningen $\widehat{Y}_1$ och $w_2 = 1 - w_1$ är vikten för skattningen $\widehat{Y}_2$. Givet att $\widehat{Y}_1$ och $\widehat{Y}_2$ är väntevärdesriktiga skattningar av Y gäller då att

$$E\left(\widehat{Y}\right) = w_1 E\left(\widehat{Y}_1\right) + w_2 E\left(\widehat{Y}_2\right) = w_1 Y + (1 - w_1) Y = Y, \tag{5}$$

dvs. $\widehat{Y}$ är (också) en väntevärdesriktig skattning av Y. Variansen för skattningen $\widehat{Y}$ ges av

$$V\left(\widehat{Y}\right) = w_1^2 V\left(\widehat{Y}_1\right) + w_2^2 V\left(\widehat{Y}_2\right) + 2 w_1 w_2 C\left(\widehat{Y}_1, \widehat{Y}_2\right), \tag{6}$$

där $C\left(\widehat{Y}_1, \widehat{Y}_2\right)$ betecknar kovariansen mellan $\widehat{Y}_1$ och $\widehat{Y}_2$. Om $\widehat{Y}_1$ och $\widehat{Y}_2$ är oberoende (t.ex. skattningar baserade på två oberoende stickprov/inventeringar) gäller att kovariansen är noll och variansen för $\widehat{Y}$ förenklas till

$$V\left(\widehat{Y}\right) = w_1^2 V\left(\widehat{Y}_1\right) + w_2^2 V\left(\widehat{Y}_2\right). \tag{7}$$

I fallet med oberoende skattningar får $\widehat{Y}$ lägst varians om vikterna (w_1 och w_2) väljs omvänt proportionellt mot variansen. Det betyder att de optimala vikterna är

$$w_1 = \frac{V\left(\widehat{Y}_2\right)}{V\left(\widehat{Y}_1\right) + V\left(\widehat{Y}_2\right)}, \quad w_2 = \frac{V\left(\widehat{Y}_1\right)}{V\left(\widehat{Y}_1\right) + V\left(\widehat{Y}_2\right)}. \tag{8}$$

Problemet är (oftast) att varianserna $V\left(\widehat{Y}_1\right)$ och $V\left(\widehat{Y}_2\right)$ inte är kända och därför kan inte w_1 och w_2 beräknas. Det är emellertid ganska vanligt att skattade varianser används i stället, men det riskerar att leda till systematiskt fel för den resulterande skattningen. Speciellt när målvariabeln har en skev fördelning (t.ex. många nollor) blir skattning och variansskattning kraftigt korrelerade. Alltså ett lågt utfall på skattning tenderar att ge ett lågt utfall på variansskattning. Om då två skattningar vägs samman läggs störst vikt på den skattning som fått ett lågt utfall, vilket leder till ett negativt systematiskt fel. Se Grafström et al. (2019) för mer information.

Då vi misstänker att målvariabeln har en skev fördelning kan det vara bättre att välja fixa vikter. Förslagsvis kan vikter väljas proportionella mot stickprovsstorlek (t.ex. inventerad areal), om designerna inte skiljer sig alltför mycket. Givet att de två skattningarna är väntevärdesriktiga räcker det att vikterna som används är fixa och summerar till ett för att kombinationen ska vara väntevärdesriktig. För att minimera variansen på kombinationsskattningen bör vikterna dock inte ligga långt ifrån de optimala vikter som presenteras i formel (8). Det kan i vissa fall krävas omfattande simuleringar (inklusive vissa antaganden) för att hitta vikter som är optimala.

4 Kombinera skattningar från hierarkiska urval

I föregående avsnitt så skattas samma total med olika inventeringar och så kombinerar vi de två skattningarna. Här skattas åtminstone delvis olika saker, olika delmängder av Y, som sedan kombineras för att få hela bilden av det fenomen vi är intresserade av.

4.1 Olika hierarkiska urval inom en inventering

Inom NILS design används trakter som samplingenhet. En trakt kan bestå av 196 provytor som i ett första skede klassas via manuell flygbildstolkning eller via automatisk klassning. Dessa klasser ger sedan urvalsklasser (strata) på trakten. Enligt en viss tilldelning väljs ett antal provytor från de urvalsklasser som ska inventeras i fält. En klass (ej aktuellt för denna inventering) inventeras inte i fält då det eftersökta fenomenen sannolikt inte finns i denna klass. Alltså utesluts allt som klassas som ej aktuellt från fältinventering. För att minimera risken att utesluta intressanta fenomen har överklassning av intressanta fenomen tillämpats.

Här vill vi skatta totalen av en variabel som finns i flera olika urvalsklasser, där vissa klasser inventerats i tätare stickprov och andra i glesare (del) stickprov. Ett glesare stickprov är en delmängd av ett tätare. De skattningar som presenteras här korrigerar för att fenomen av intresse återfinns i flera urvalsklasser. Alltså, det intressanta fenomenet som ska skattas har funnits i urvalsklasser som inte inventeras i fält i tätare stickprovet. Trots detta kan vi använda allt data i skattning av det intressanta fenomenet. Vi skattar total av variabeln inom varje urvalsklass för sig och sedan lägger vi ihop resultaten. Teorin fungerar för flera täthetsnivåer.

En generell lösning för K st urvalsklasser presenteras här. Låt Y_k vara total av intressant variabel i klass $k = 1, 2, ..., K$. Populationstotalen blir nu

$$Y = \sum_{k=1}^{K} Y_k,$$

som skattas med

$$\widehat{Y} = \sum_{k=1}^{K} \widehat{Y}_k.$$

Variansen för $\widehat{Y}$ är

$$V\left(\widehat{Y}\right) = \sum_{k=1}^{K}\sum_{\ell=1}^{K} C\left(\widehat{Y}_k, \widehat{Y}_\ell\right).$$

En skattning av variansen blir

$$\widehat{V}\left(\widehat{Y}\right) = \sum_{k=1}^{K}\sum_{\ell=1}^{K} \widehat{C}\left(\widehat{Y}_k, \widehat{Y}_\ell\right).$$

Då samma trakter ingår i skattningarna finns möjligtvis ett beroende mellan $\widehat{Y}_k$ och $\widehat{Y}_\ell$, även fast urvalen (och skattningarna) inom en trakt är oberoende för olika urvalsklasser. Uttrycket går därför inte att förenkla till en summa av varianserna för de enskilda skattningarna.

Skattningen av kovariansen $C\left(\widehat{Y}_k, \widehat{Y}_\ell\right)$ kan göras på olika sätt. Ett alternativ (via koordination med oberoende trakter) är

$$\widehat{C}\left(\widehat{Y}_k, \widehat{Y}_\ell\right) = \frac{a_F^2 n_{k\ell}}{n_k n_\ell (n_{k\ell}-1)} \sum_{S_{k\ell}} \left(\hat{Y}_k(\mathbf{x}) - \frac{1}{n_{k\ell}}\sum_{S_{k\ell}} \hat{Y}_k(\mathbf{x})\right)\left(\hat{Y}_\ell(\mathbf{x}) - \frac{1}{n_{k\ell}}\sum_{S_{k\ell}} \hat{Y}_\ell(\mathbf{x})\right),$$

där a_F betecknar arealen av urvalsramen F och $\widehat{Y}_k$ har formen

$$\widehat{Y}_k = \frac{a_F}{n_k}\sum_{S_k} \widehat{Y}_k(\mathbf{x}).$$

Urvalet $S_{k\ell}$ är snittet av S_k och S_ℓ och $n_{k\ell}$ är storleken (antalet trakter) i $S_{k\ell}$. Skattningen $\widehat{Y}_k(\mathbf{x})$ skattar här traktmedel per arealenhet för trakten på koordinat $\mathbf{x}$. Alltså är $\widehat{Y}_k(\mathbf{x})$ en HT-skattning av totalen på trakten delat på hela traktens areal.

Det går även att ta fram en skattning av kovarianserna som tar hänsyn till att urvalet är rumsligt balancerat. Det är dock mer komplext då antalet grannar vid beräkning av lokala medel troligtvis bör vara proportionellt mot storleken på urvalet/snittet (antal trakter) som kommer att variera. Bäst är då troligtvis att sätta antalet grannar till ca 4 st för det minsta urvalet (snittet) och sedan ta proportionellt fler för det större urvalet. En sådan skattning får en liknande form som tidigare.

$$\widehat{C}\left(\widehat{Y}_k, \widehat{Y}_\ell\right) = \frac{a_F^2 n'_{k\ell}}{n_k n_\ell (n'_{k\ell}-1)} \sum_{S_{k\ell}} \left(\hat{Y}_k(\mathbf{x}) - \frac{1}{n'_{k\ell}} \sum_{S_{k\ell}(\mathbf{x})} \hat{Y}_k(\mathbf{x}')\right)\left(\hat{Y}_\ell(\mathbf{x}) - \frac{1}{n'_{k\ell}} \sum_{S_{k\ell}(\mathbf{x})} \hat{Y}_\ell(\mathbf{x}')\right)$$

där $n'_{k\ell}$ är antalet grannar i lokalt medel och $S_{k\ell}(\mathbf{x})$ innehåller de närmsta $n'_{k\ell}$ grannarna till trakten $\mathbf{x}$ i snittet $S_{k\ell}$ (mätt i de variabler urvalet är spritt). Denna kovariansskattning kommer från en approximativ skattning vid koordination med rumsligt balanserade urval via SCPS (Spatially Correlated Poisson Sampling). Se Zhao (2021).

4.2 Samma hierarkiska urval men olika inventeringar

När data samlats in i samma trakter från olika uppsättningar av urvalsklasser, t.ex. från olika inventeringar, kan det finnas ett överlapp mellan inventeringarna som behöver hanteras för att det ska gå att kombinera skattningarna mellan dem. För NILS gräsmarks- och lövskogsinventeringar gäller att de har varsina uppsättningar av urvalsklasser men de inventeras i helt eller delvis samma trakter. Skillnaden här, jämfört med förra delen, har att göra med att skattningar ska kombineras mellan olika uppsättningar av urvalsklasser som har inventerats i samma trakter.

Givet att det går att skatta (hela) totalen Y var för sig med två olika inventeringar kan en kombinerad skattning göras enligt formel (4) med varians som ges av (6). Här betecknar $\widehat{Y}_1$ skattningen av Y med inventering 1 och $\widehat{Y}_2$ skattningen av Y med inventering 2. Kovariansen kan skattas med överlappet (gemensamma trakter), men det blir kovarianser att beräkna mellan alla kombinationer av urvalsklasser.

Med K urvalsklasser i inventering 1 och G klasser i inventering 2 fås totalt $K+G$ klasser. Varians/kovarians ska beräknas med största möjliga urval för kombinationen. Med

$$\widehat{Y}_1 = \sum_{k=1}^{K} \widehat{Y}_{1k} \text{ och } \widehat{Y}_2 = \sum_{\ell=1}^{G} \widehat{Y}_{2\ell},$$

fås

$$C\left(\widehat{Y}_1, \widehat{Y}_2\right) = C\left(\sum_{k=1}^{K} \widehat{Y}_{1k}, \sum_{\ell=1}^{G} \widehat{Y}_{2\ell}\right) = \sum_{k=1}^{K} \sum_{\ell=1}^{G} C\left(\widehat{Y}_{1k}, \widehat{Y}_{2\ell}\right).$$

Kovariansen $C\left(\widehat{Y}_{1k}, \widehat{Y}_{2\ell}\right)$ skattas därefter på samma sätt som $C\left(\widehat{Y}_k, \widehat{Y}_\ell\right)$.

Tidigare skattning (4) bygger på antagandet att det går att skatta (hela) totalen Y var för sig med två olika inventeringar. Om detta inte är uppfyllt kan (eventuellt) ett alternativ vara att använda en estimator på formen

$$\widehat{Y} = \widehat{Y}_1 + \widehat{Y}_2 - \widehat{Y}_{12}, \tag{9}$$

där vi tänker oss att $\widehat{Y}_1$ skattar en del av totalen Y och $\widehat{Y}_2$ skattar en del (som eventuellt överlappar delvis med den del $\widehat{Y}_1$ skattar) samt att $\widehat{Y}_{12}$ skattar eventuellt överlapp mellan de två skattningarna. Ett exempel på denna situation har uppstått i inventeringarna av gräsmarker och lövskogar i NILS nya designsystem. Trädklädda betesmarker är en heterogen naturtyp där särskilt ytor med hög krontäckning är svåra att klassificera korrekt genom fjärranalys. Det mesta av de trädklädda betesmarkerna kommer med i gräsmarksinventering men för en del trädklädda betesmarker går det inte att se betet under träden, de klassa då som skog och kan inkluderas enbart till lövskogsinventeringen. Ytterligare en del av de trädklädda betesmarkerna inventeras i både gräsmarks- och lövskogsinventeringarna, antingen för att det utifrån fjärranalys inte går att avgöra om det förekommer bete eller ej eller för att provytan är delad mellan gräsmark och lövskog. Eventuellt finns en mindre del trädklädda betesmarker som inte inventeras i någon av inventeringarna, det skulle i så fall gälla barrträdsdominerade trädklädda betesmarker där tecken på bete inte går att urskilja vid fjärranalys.

Båda inventeringarna tillämpar överklassning för att inte riskera att något av intresse hamnar i den klass (ej aktuellt) som inte inventeras. Inventeringarna har olika klasser (som alltså kan överlappa). I det glesaste stickprovet inventeras alla klasser för både gräsmarker och lövskogar (förutom det som är klassat som ej aktuellt i respektive inventering). Det betyder att en skattning på formen (9) kan vara ett alternativ. Skattningarna

$$\widehat{Y}_1 = \sum_{k=1}^{K} \widehat{Y}_{1k} \text{ och } \widehat{Y}_2 = \sum_{\ell=1}^{G} \widehat{Y}_{2\ell},$$

är i princip som tidigare men vi tänker nu att de inte var för sig skattar hela Y eftersom olika delar (av totalen Y) kan tillhöra klasser som inte inventeras fullt ut i en enskild inventering. Kovariansen mellan $\widehat{Y}_1$ och $\widehat{Y}_2$ beräknas (och skattas) som tidigare.

Vi har dock att variansen av en skattning på formen (9) ges av

$$V\left(\widehat{Y}\right) = C\left(\widehat{Y}, \widehat{Y}\right) = C\left(\widehat{Y}_1 + \widehat{Y}_2 - \widehat{Y}_{12}, \widehat{Y}_1 + \widehat{Y}_2 - \widehat{Y}_{12}\right) =$$

$$= V\left(\widehat{Y}_1\right) + V\left(\widehat{Y}_2\right) + V\left(\widehat{Y}_{12}\right) + 2\left(C\left(\widehat{Y}_1, \widehat{Y}_2\right) - C\left(\widehat{Y}_1, \widehat{Y}_{12}\right) - C\left(\widehat{Y}_2, \widehat{Y}_{12}\right)\right).$$

För att kunna beräkna variansen behöver vi komma åt kovarianserna $C\left(\widehat{Y}_1, \widehat{Y}_{12}\right)$ och $C\left(\widehat{Y}_2, \widehat{Y}_{12}\right)$. För att göra det behöver vi först specificera hur $\widehat{Y}_{12}$ ser ut. Vi antar att ytor inventeras likvärdigt i de båda inventeringarna oberoende av vilken klass de tillhör. Vi kan då skatta överlappet för varje klass (k) i första inventeringen och får $\widehat{Y}_{12k\cdot}$. Denna skattning använder för en trakt alla ytor som har klass k i första inventeringen och som kan väljas till båda inventeringarna, samt valts ut för inventering i någon av inventeringarna Det betyder att som inklusionssannolikhet används sannolikheten att väljas till minst en av inventeringarna, se avsnitt 5. På motsvarande sätt kan $\widehat{Y}_{12\cdot\ell}$ beräknas med de ytor som har klass ℓ i andra inventeringen. Vi har

$$\widehat{Y}_{12} = \sum_{k=1}^{K} \widehat{Y}_{12k\cdot} = \sum_{\ell=1}^{G} \widehat{Y}_{12\cdot\ell},$$

vilket betyder att vi kan skatta överlappet genom att antingen aggregera över klasserna i första inventeringen eller över klasserna i andra inventeringen. För kovarianserna får vi

$$C\left(\widehat{Y}_1, \widehat{Y}_{12}\right) = C\left(\sum_{k=1}^{K} \widehat{Y}_{1k}, \sum_{\ell=1}^{K} \widehat{Y}_{12\ell\cdot}\right) = \sum_{k=1}^{K}\sum_{\ell=1}^{K} C\left(\widehat{Y}_{1k}, \widehat{Y}_{12\ell\cdot}\right)$$

respektive

$$C\left(\widehat{Y}_2, \widehat{Y}_{12}\right) = C\left(\sum_{k=1}^{G} \widehat{Y}_{2k}, \sum_{\ell=1}^{G} \widehat{Y}_{12\cdot\ell}\right) = \sum_{k=1}^{G}\sum_{\ell=1}^{G} C\left(\widehat{Y}_{1k}, \widehat{Y}_{12\cdot\ell}\right),$$

där $C\left(\widehat{Y}_{1k}, \widehat{Y}_{12\ell\cdot}\right)$ och $C\left(\widehat{Y}_{1k}, \widehat{Y}_{12\cdot\ell}\right)$ skattas på samma sätt som $C\left(\widehat{Y}_k, \widehat{Y}_\ell\right)$.

Notera att skattningen (9) inte är en känt förekommande skattning och att det inte är garanterat att den ger ett positivt utfall när en positiv parameter ska skattas. Mer praktiska test behöver göras för att avgöra om en skattning på denna form är lämplig. Det kan om möjligt vara bättre att införa ett harmoniserat system av klasser i olika inventeringar.

5 Inkludera fältinformation från provytor på samma trakt, men som valts till annan inventering

Eftersom gräsmarks- och lövskogsinventeringarna till stor del genomförs i samma trakter förekommer det att provytor får en positiv inklusionssannolikhet för båda inventerarna men bara väljs för fältinventering i den ena. För att i sådana fall kunna inkludera fältinformation även till den inventering som provytan inte valts för fältbesök till, kan inklusionssannolikheten för provyta i beräknas utifrån sannolikheten att den kommer med i någon av inventeringarna, enligt

$$\pi_i = 1 - (1 - \pi_{1i})(1 - \pi_{2i}), \tag{10}$$

där π_{1i} är sannolikheten att provyta i väljs till inventering 1 och π_{2i} är sannolikheten att provyta i väljs till inventering 2.

Referenser

Adler, S., Christensen, P., Gardfjell, H., Grafström, A., Hagner, Å., Hedenås, H. & Ranlund, Å. (2020). Ny design för riktade naturtypsinventeringar inom NILS och THUF. Arbetsrapport 513. Sveriges lantbruksuniversitet, Institutionen för skoglig resurshushållning, Umeå.

Adler, S., Hedenås, H., Hagner, Å., Ranlund, Å. och Christensen, P., (2022). Utvärdering av NILS fjällinventering 2021. Arbetsrapport/Sveriges lantbruksuniversitet, Institutionen för skoglig resurshushållning, (532).

Grafström, A., Zhao, X., Nylander, M., & Petersson, H. (2017). A new sampling strategy for forest inventories applied to the temporary clusters of the Swedish national forest inventory. Canadian Journal of Forest Research, 47(9), 1161-1167.

Grafström, A., Ekström, M., Jonsson, B. G., Esseen, P. A., & Ståhl, G. (2019). On combining independent probability samples. *Survey Methodology*, 45(2), 349-364.

Ranlund, Å., Sjödin, M., Press, A., Gardfjell, H., Hedenås, H., Hagner, Å., Forsman, H., Christensen, P., Andersson, M., & Adler, S. (2021). Metodbeskrivning : 2020 års inventeringar av gräsmarker och lövskogar. Arbetsrapport 530. Institutionen för skoglig resurshushållning, Sveriges lantbruksuniversitet.

Zhao, X. (2021). Design-based sampling methods for environmental monitoring (No. 2021: 51). Avhandling. Institutionen för skoglig resurshushållning, Sveriges lantbruksuniversitet, Umeå.