Chapter 2
Sampling Design and Estimation Procedures

Adrian Lanz, Christoph Fischer, and Meinrad Abegg

Abstract The topic of this chapter is the statistical sampling approach of the Swiss National Forest Inventory (NFI). We start with the main populations and population parameters of interest, and how these are – under the infinite population or Monte-Carlo approach to forest inventory – re-shaped into and understood as continuous populations in space. The spatial arrangement of field data collection on a grid of permanently installed sample plots recently changed from a periodic into a continuous (annual) system, which are presented in the following sections, together with estimation procedures which make use of remote sensing and other spatial data to increase the precision of the estimates. We then emphasise on the estimation of change and change components, and depict specific solutions such as for the estimation of the average annual change and the change per unit area. Finally, the NFI system of two-stage subsampling of tally trees for stem volume estimation and the respective estimation procedures are described in this chapter.

2.1 Introduction

Swiss forests cover more than 30% of the country's territory and are stocked with about 500 million trees (with a stem diameter at breast height ≥12 cm). The sheer size of this study object inhibits full observation and enumeration of Switzerland's forest and trees. An additional challenge is that the information in demand extends beyond the current state of the forest and the stock of living trees as well as recent changes. Indeed, it includes many other aspects of forest ecosystems, such as young

A. Lanz (✉)
Statistics Lab, Swiss Federal Research Institute WSL, Birmensdorf, Switzerland
e-mail: adrian.lanz@wsl.ch

C. Fischer
Scientific Service NFI, Swiss Federal Research Institute WSL, Birmensdorf, Switzerland

M. Abegg
Scientific Service NFI, Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland

C. Fischer, B. Traub (eds.), *Swiss National Forest Inventory – Methods and Models of the Fourth Assessment*, Managing Forest Ecosystems 35,
https://doi.org/10.1007/978-3-030-19293-8_2

(regeneration) trees, standing and lying deadwood, and the occurrence of shrubs, plants and animals, as well as various site and stand characteristics, including the recreational use of forests and forest management activities.

Therefore, only a small part of the total forest area can be observed in a national survey. In fact, with the current amount of field data collected and forest area covered annually by the Swiss NFI (NFI), it would take around 40,000 years to observe Switzerland's entire forest.

2.1.1 Survey Sampling

In this situation, *statistical sampling* comes into play because randomly selected samples eliminate selection bias. The resulting samples are objective and acceptable to the public (Särndal et al. 2003). Survey sampling is a specialised field of statistics consisting of a theory and methods for extracting information from observational data to solve real-world problems (Madigan et al. 2014). Survey sampling techniques are widely used in public statistics at the national and regional levels, including in forest inventories. Statistical knowledge is essential in survey sampling. The statistician's role is to: (a) design the acquisition of data in a way that minimises bias and confounding factors, and maximises information content; (b) verify the quality of data after its collection; and (c) analyse data in a way that provides insights or information that support decision making.

A *sample survey* is built on three pillars: (a) a well-defined *population* and agreed *population parameters* of interest; (b) a *sampling design* for data acquisition; and (c) a set of *estimator algorithms* for estimating the population parameters of interest from the collected data. The three pillars interact and depend on each other. For instance, a change in the definition of the population usually gives rise to a change in the sampling design, which then requires an adaptation of the estimators. The role of the statistician is to ensure smooth interactions and to find optimal overall *survey designs*.

2.1.2 Chapter Contents

This chapter is divided into four sections. In the first section, we provide the definitions of the main *populations of interest* and the respective *population elements* with associated target variables in the case of the NFI. In the second section, we first outline the statistical theory of *probability sampling* in forest inventories in general, and then explain the *sampling design* of the NFI for the main populations of interest, i.e. trees, young (regeneration) trees and lying deadwood. The NFI implements a unique, two-stage sampling and measurement design with related procedures for estimating stem volumes and the biomass of living trees. The design is explained in detail in this section. Another focus of the second section is the spatial and temporal organisation of the field data collection on annual panels of *permanently installed sample plots*.

In the third section, the overall *estimation procedures* are presented, with emphasis on aspects that are specific to the NFI, such as: (a) the two-phase estimation procedures combining remote sensing and other spatial auxiliary data with field data for more precise estimation; (b) the two-stage estimation procedure for growing stock and biomass estimation; and (c) the implementation of area domain and sub-population estimation. The estimators are given in 20.8.2.

The *estimation of change and components of change* between inventories is paramount in the NFI. In the final section of this chapter, the panel survey system of the NFI, with periodic re-measurement of permanent plots, is described in detail. The historical context of estimating change is described, including the *méthode de contrôle* (Gurnaud 1886; Biolley 1901, 1920) established by Biolley in 1890 in Canton Neuchâtel. The NFI system with permanent plots is essentially a continuation and application of Biolley's inventory system with respect to the form of statistical sampling. The estimators and components of change are presented in detail, particularly regarding the treatment of changing area domains and the estimation of change components in the stock of living trees under the two-stage sampling design.

2.2 Sampling Design and Estimation Procedures

2.2.1 Population of Interest

Probability sampling methods always build upon a well-defined population of elements with associated target variables for which means and totals (sums) are then estimated. In the context of the NFI, it is therefore essential to have an exact definition of *forest*, but also of various population types, such as the *population of trees*, the *population of deadwood pieces*, and the *population of young trees in the forest regeneration*. These populations are briefly described in this section so that the sampling scheme used in the NFI can be understood by the reader.

2.2.2 Forest and Shrub Forest Area

The current definition of *forest*, as applied in the NFI, was published in 1976 (Mahrer 1976). The definition is designed to comply as much as possible with the legal definition of a forest according to Swiss federal law. The corresponding regulation, published in 1965, leaves space for a certain margin of discretion. This enables the use of clear and objective criteria for the NFI forest definition, which are still in agreement with the legal definition. Furthermore, it is important that the forest definition can be applied when using remote sensing alone, thereby making it possible to include inaccessible forest in NFI forest assessments. For the NFI, criteria related to the following aspects are used in the definition of forest:

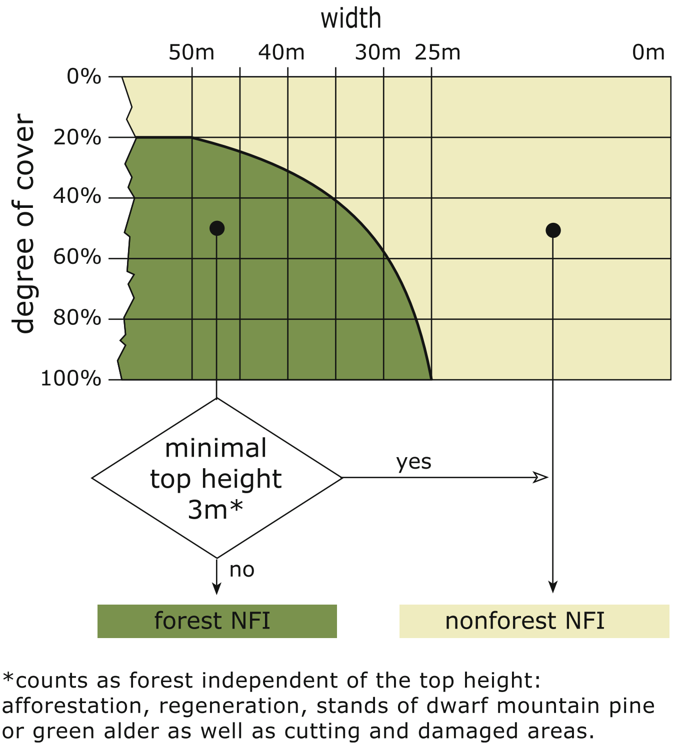

Fig. 2.1 The relationship between crown cover and width of a stocking, according to the NFI definition

- Stocking of trees
- Minimum crown cover
- Minimum height of trees
- Minimum width of stocking
- Land use

Eligible elements of a forest stocking are defined as trees of any species or members of a closed list of shrub species. Orchards are not considered because the land use is predominantly agricultural (Düggelin and Keller 2017). According to the NFI definition, a forest stocking has a minimum crown cover of 20% and a minimum width of 25 m. The required minimum width of the stocking is directly dependent on the crown cover (Fig. 2.1) and can reach up to 50 m if the crown cover is only 20%. The minimum width indirectly defines the minimum size of the stocking. The purpose of including the minimum width criterion is that linear stockings that are narrow but very long, for example bordering a brook, are not defined as forest.

The elements of the stocking must have a height of at least 3 m because this is the detectability threshold for trees in the aerial-photo interpretation. However, at the upper treeline in alpine regions there are relatively large areas covered primarily by two species, *Alnus viridis* and *Pinus mugo*, that rarely reach 3 m in height. An exception was therefore established for these two species, and they are considered stocking even if their height is <3 m. Additionally, temporarily unstocked areas, for example owing to harvesting, natural disturbances or damage, and afforested areas are counted as forest. Further, small clearings which are not clearly separable from the surrounding forested land are included as part of the forest, but only if the width of the clearing is less than 25 m. Finally, there are land-use criteria which lead to certain areas being classified as 'forest' according to the NFI forest definition, even if they are unstocked, and other areas not being considered forest even though they are stocked with trees. The following elements are not considered forest according to the NFI definition (Düggelin and Keller 2017):

- Roads and streets wider than 6 m
- Streams with a streambed wider than 6 m
- Railway tracks, cable cars, ski lifts or similar infrastructure
- Buildings with a non-forestry purpose
- Allotments or orchards
- Tree nurseries for gardening purposes
- Parks
- Rows of trees bordering a street, alleys or avenues

The following unstocked areas are considered forest:

- Forest roads up to 6 m wide
- Timber yards: small unsurfaced storage yards in permanent use, located next to a forest road and forest stand
- Recreation areas: huts, resting places, car parks and other recreation areas with a width of less than 25 m
- Tree nurseries for forestry purposes, adjacent to forest land and in small clearings up to 25 m wide
- Streams up to 6 m wide
- Erosion channels, avalanche tracks, skidding tracks or other channels with an unstocked area <25 m wide
- Small clearings of meadows, cropland or bare land

The forest area in Switzerland is divided into two subtypes of forest: *shrub forest* and *forest without shrub forest* because shrub forests are of no economic interest for the forestry industry, which was the main focus of the first NFI. Shrub forest is distinguished from the rest of the forest according to the criterion of a crown cover by shrub species of more than 3 m in height, or *Alnus viridis* or *Pinus mugo* stands, of at least two-thirds.

2.2.3 Trees

In the NFI, trees with a height $\geq$10 cm are sampled. However, the measurement procedure differs for two groups of trees: (a) *Tally trees* are living and dead, standing and lying trees and certain shrubs according to a given species list with a diameter at breast height ($d_{1.3}$) $\geq$12 cm. (b) *Young trees* consists of standing living trees, as well as certain shrubs, >10 cm in height but <12 cm in $d_{1.3}$. In NFI5, standing dead trees with a height >1.3 m will be sampled as well. The population of trees considered in the NFI only includes trees inside forest areas according to the NFI forest definition.

2.2.3.1 Tally Trees

All trees with a $d_{1.3}$ $\geq$12 cm are considered tally trees, whereas only shrubs included in a specific list (Chap. 9) are assigned to this group. This list contains all native and most non-native shrub species occurring in forests. The $d_{1.3}$ is measured on stems at 1.3 m above the highest ground level at the bottom of the tree. Tally trees can be living, dead, standing or lying. Trees and shrubs are distinguished according to a defined species list. As the NFI is based on permanent sample plots, tally trees are tracked over successive inventories – even if they leave the sample temporarily – so that their development can be assessed when they re-appear in the sample.

2.2.3.2 Young Trees

To assess regeneration, all trees, as well as shrubs of seven species (Chap. 4), with a height $\geq$10 cm but a $d_{1.3}$ <12 cm are measured. For regeneration trees <1.3 m in height, signs of game browsing is additionally assessed. Two different aspects of regeneration are evaluated using two different methods: (a) the number of individuals is estimated by counting the number of individuals in concentric circles on the regeneration plot; and (b) the relative area 'occupied' by young trees with certain characteristics of interest is estimated by assessing the young tree individual located closest to the plot centre (Schwyzer and Lanz 2010).

2.2.4 Lying Deadwood

Deadwood is important as habitat, and two different types of deadwood are thus assessed in the NFI: (a) standing and lying dead tally trees on NFI plots, and (b) pieces of lying deadwood found on line intersects. The elements of the population of lying deadwood are pieces with a minimum length of 1 m and a minimum diameter of 7 cm. Stumps are not included. The detailed measurement protocol for line-intersect sampling is described in Chap. 9 and the inference applied is described in Sect. 2.3.5.

2.2.5 Further Elements

2.2.5.1 Forest Edge Survey

Forest edges are considered important habitats for a wide range of species. In order to describe these habitats, the NFI includes an assessment of the characteristics of forest edges located close to NFI sample-plot centres (<25 m). For further details on these assessments, see Chap. 9. The data is collected along a transect 50 m in length. Due to the method of sampling and the fractal property of forest edges (Mandelbrot 1967), statistical inference of the length of forest edges in Switzerland, for instance, would require specific assumptions and estimation techniques, which are not implemented in the NFI. However, based on the available sample of approximately 1000 edge transects, it is possible to evaluate the average composition of the forest edges and to assess changes between inventories.

2.2.5.2 Forest Road Survey

The forest road survey is a census survey conducted by the NFI that provides the basis for reporting on the state of the forest road infrastructure in Switzerland. Forest truck roads are defined as roads on forest land or roads bordering forest land with a minimum width of 2.5 m and a carrying capacity of 10 t per axle. Various characteristics of the individual road sections are verified and assessed as part of the NFI interview survey with the local forest service (Chap. 10). The methods of data analysis and the way of combining the road survey data with the NFI field data changed during the course of NFI4. The earlier method is described by Paschedag and Zinggeler (2001). The new method makes use of spatially explicit data such as the NFI forest cover map (Sect. 7.2), and increased automation in the calculations.

2.3 Sampling Design

2.3.1 Population Parameters of Interest

A forest inventory, like any statistical survey, is concerned with the estimation of summary statistics, such as totals (sums) or means, and of target variables associated with the elements of the population under study. The typical elements in a forest inventory are trees located on forest land. Thus, the population bears a clear reference to space, a characteristic which plays a role not only in the definition of the population, but also in the sampling method and the mathematical notation.

The formal description of the population of interest assumes a suitable projection of the real forest landscape with trees located in the plane in $\Re^2$. Points in $\Re^2$ are denoted by ω and the area domain of interest, the forest land, is a bounded region F in $\Re^2$ with surface area λ_F.

The population P_F with N_F elements, usually trees, in F is labelled $1, \ldots, i, \ldots, N_F$ and element locations u_i are assumed to be well defined and fixed. $u_i \in F$ is true for all elements of P_F.

The element-level target variables of interest are assumed to be measurable without error. Typical target variables of interest are the stem volume of trees, the basal area of stems measured 1.3 m aboveground, and total aboveground biomass of trees. Target variables are denoted by X. If two different target variables are needed, they are referred to as X and Z. The variable $X_i \equiv 1$ for all $i \in P_F$, which is a useful, but also simple, variable for counting or estimating the number of stems in the population.

The population parameters of interest in a forest inventory are:

The total surface area of domain F, for example the forest land

$$\lambda_F = \int_{\Re^2} \imath_F(\omega) d\omega = \int_F l_F(\omega) d\omega \tag{2.1}$$

where the integral refers to Riemann integration in $\Re^2$ and

$$\imath_F(\omega) = \begin{cases} 1 & \text{if } \omega \in F \\ 0 & \text{else} \end{cases} \tag{2.2}$$

is a Bernoulli variable defined for all $\omega \in \Re^2$, indicating points in domain F.
The total of target variable X over the population P_F of elements in domain F

$$T_F^{(X)} = \sum_{i=1}^{N_F} X_i; \tag{2.3}$$

The mean spatial density of target variable X over the population P_F of elements in domain F

$$Y_F^{(X)} = \frac{1}{\lambda_F} T_F^{(X)}; \tag{2.4}$$

The ratio of the totals of two different target variables X and Z over the population P_F of elements in domain F

$$R_F^{(X/Z)} = \frac{T_F^{(X)}}{T_F^{(Z)}} = \frac{Y_F^{(X)}}{Y_F^{(Z)}}. \tag{2.5}$$

2.3.2 Infinite Population Approach

In a forest inventory covering a large area, we do not have a list of all trees for sample selection. For the sampling, a *sampling frame* $G \supseteq F$ is used, for a region of known surface area λ_G that is as large as the study area, the forest land F.

The points ω in G are the *sampling units* from which a random sample will be selected in the survey. The challenge is to define, prior to sampling and for all sampling units, $\omega \in G$ – a *local (per unit area) density* (Mandallaz 2008) function $y_F^{(X)}(\omega)$ with regards to target variable X in domain F, with the property

$$\int_G y_F^{(X)}(\omega) d\omega = \sum\nolimits_{i=1}^{N_F} X_i = T_F^{(X)} \tag{2.6}$$

In other words: the *local density* functions need to be defined such that the population parameter $T_F^{(X)}$ of interest, the total of X over the population P_F of N_F elements in F, can be computed through Riemann integration of $y_F^{(X)}(\omega)$ over all points ω in frame G.

Once the local density function has been defined, the random sampling of population elements (trees) that form the finite, but not available, list of elements in the population can be replaced by the random sampling units (points) from the infinite, but well-defined, number of units in the sampling frame G. This concept is known as the *infinite population approach* (Mandallaz 2008; Eriksson 1995b) or *Monte-Carlo approach* (Gregoire and Valentine 2007) to sampling in forest inventories.

Local density functions can be understood as *association rules* between the sampling units $\omega \in G$ (points) and nearby population elements $u_i \in F$ (trees). In mathematical notation, the local density can be written as

$$y_F^{(\mathrm{X})}(\omega) = \sum_{i \in S(\omega)} X_i f_i(\omega) = \sum_{i=1}^{N_F} \iota_{i.S(\omega)} X_i f_i(\omega), \tag{2.7}$$

i.e. a weighted sum of target variable values X of population elements associated with ω. The term $\iota_{i.S(\omega)}$ formally describes the sampling unit at point ω. The $\iota_{i.S(\omega)}$ describe the association of population elements with sampling units: $\iota_{i.S(\omega)} = 1$ if population element i is a member of the local sample $S(\omega)$ of trees associated with sampling unit (point) ω; otherwise $\iota_{i.S(\omega)} = 0$.

$f_i(\omega)$ are element-level *extrapolation factors*, and we conclude from Eqs. (2.6) to (2.7) that a local density function should be defined such, that the condition

$$\int_G \iota_{i.S(\omega)} f_i(\omega) d\omega = \int_{A_{i.G}} f_i(\omega) d\omega = 1 \tag{2.8}$$

is true for all elements i in population P_F. $A_{i.G}$ of size $\lambda_{A_{i.G}}$ is the so-called *inclusion zone* (or inclusion field) of population element i. It includes all sampling units (points) ω in G for which i is a member of the respective local sample $S(\omega)$.

Inclusion zones and related extrapolation factors are mathematical concepts that correspond with sample-selection rules applied in field operations (duality principle). This link will be illustrated for the NFI field protocol in the next section.

Before moving on, it is important to mention that we assume fixed target variable values on population elements and fixed local densities on sampling units, which implies association rules between sampling units and population elements defined prior to sampling. The randomness on which inference is based is only in the method of selecting sampling units (points) from the sampling frame (Sect. 2.4.2).

2.3.2.1 Local Density Functions in NFI

In this section, the NFI plot configurations, the association rules between sampling units (plot centres) and population elements, and the related local density functions are explained for the most important populations of interest: trees, pieces of lying deadwood and young trees.

2.3.2.2 Area Domain Estimation

In this first section, we describe the NFI 'plot configuration' and the local density function for (forest) area estimation.

The population of interest is in this case a geographic (area) domain $F \subseteq G$ of unknown size λ_F. Area domains frequently need to be estimated in the NFI, to estimate not only Switzerland's total forest area, but also the proportion and/or surface area of various area (sub)domains, such as forest areas classified according to ownership or stand-age categories.

The population of interest is formally defined as a bounded region F in $\Re^2$ and the response for sampling units in G is simply a domain membership indicator variable $\iota_F(\omega) = 1$ for units located in F, and $\iota_F(\omega) = 0$ otherwise. Consequently, the one-to-one association rule between sampling units ω in G and population elements ω in F results in a local density function $y(\omega) = \iota_F(\omega)$.

2.3.3 *Sampling of Tally Trees*

The typical association rule, or *sampling function*, between sampling units (points) and trees in a forest inventory is called *fixed radius plot sampling*, i.e. the selection of all trees in a circular plot centred at point $\omega \in G$ of radius ρ. Therefore, for a tree positioned at point $u_i \in F$, the resulting *inclusion zone*

$$A_{i.G} = \{\omega \in G | |\omega - u_i| \leq \rho\} \quad (2.9)$$

is usually circular, centred in u_i and of size $\lambda_{A_{i.G}} = \pi\rho^2$, and the extrapolation factor is $f_i(\omega) = 1/\lambda_{A_{i.G}} = 1/\pi\rho^2$. For trees near the sampling frame's boundary, the inclusion zone may extend outside of G, so that $f_i(\omega) > 1/\pi\rho^2$.

Under *concentric (nested) fixed-radius plot sampling*, predefined *subpopulations of trees* are selected from circular plots with different radii. The subpopulations may be defined according to any criteria, for example tree species. A typical approach in forest inventories is to use larger circles for trees with large stem diameters and smaller circles for trees with smaller stem diameters.

The NFI uses two concentric circular plots with fixed sizes and the same centre points. The subpopulation of trees with a $d_{1.3} \geq 12$ cm but <36 cm is associated with a plot 200 m^2 in area, and the subpopulation with a $d_{1.3}$ >36 cm is associated with a plot 500 m^2 in area.

The local densities are calculated as a per hectare density and the *tally tree extrapolation factors* in the NFI are therefore 50 ha^{-1} for the 200 m^2 plots and 20 ha^{-1} for the 500 m^2 plots.

Slope Correction Because all area-related population parameters of interest are understood as statistics after projection to the plane, circular plots become ellipses when they are projected onto sloped terrain and a locally smooth terrain surface is assumed. To facilitate fieldwork and to avoid tree individual maximum distances between the plot centre and tree locations, the NFI radius for tally tree selection, applied parallel to the possibly sloped surface of the terrain, is defined as

$$\rho(\omega) = \frac{\rho_{\text{flat}}}{\sqrt{\cos\beta(\omega)}} \quad (2.10)$$

where $\beta(\omega)$ is the locally determined slope of the terrain, and ρ_{flat} is the standard radius of 12.61 m for the subpopulation of large trees and 7.98 m for the subpopulation of small trees.

$\rho(\omega)$ is chosen such that the surface area of the back-projected ellipse is equal to the surface area of the original circular plot ρ_{flat} in radius.

NFI Sampling Frame Boundary Correction The field protocol of the NFI leads to a *specific sampling frame for tree sampling*. The field protocol is such that only sampling units ω located on forest land are used for tally tree selection and measurement. As a consequence, trees near the forest boundary line have a lower overall probability of becoming a member of the sample, and boundary correction measures are therefore required.

The mathematical explanation for this boundary effect is related to the shape and size of tree inclusion zones. Under fixed radius plot sampling, the tree inclusion zones are bounded by circles centred at tree locations u_i. However, the circular tree inclusion zone is truncated for trees near the boundary.

The following boundary correction measures are applied in the NFI:

- For sampling units qualifying for tree sampling, i.e. located on forest land, a *reduction line*, roughly corresponding to the forest boundary, is locally determined and registered as part of the field protocol.
- The reduction line is defined in such a way that (a) no eligible tree in P_F is outside this line, and (b) the forest boundary is outside and does not cross the reduction line.
- After automated analysis of field data, the derived extrapolation factors for tally trees are then $f_i(\omega) = f_i(\omega)^{(0)}/p(\omega)$, where $f_i(\omega)^{(0)}$ is the standard extrapolation factor of tally tree i and $p(\omega)$ is the proportion of the plot, either the small or the large plot depending on the i-th tally tree $d_{1.3}$, located inside the reduction line.

This procedure of correcting tally tree extrapolation factors is, as a matter of principle, not exact (Gregoire 1982), but it is assumed to have a minor impact on the overall estimates of tree resources. Based on a modified field data collection protocol in use since NFI5, there are plans to apply tree-level boundary correction in the next NFI reports.

2.3.4 Sampling of Tariff Trees

In NFI, a subsample of *tariff trees* is randomly selected from the sample of tally trees for the more time-consuming measurement of their upper-stem diameters and stem lengths. The overall sampling scheme has been described as two-stage sampling (Mandallaz 2008). Related *two-stage estimation procedures* have been applied in NFI since NFI2 for growing stock and biomass estimation.

2.3.4.1 Motivation

Traditionally, single stem volumes are predicted by means of *tariff functions*, which predict stem volumes as a function of the stem $d_{1.3}$. Tariff functions are calibrated on the basis of section-wise measured stems. Mean tariff functions are accurate, but they are known to over-estimate or under-estimate the true stem volume of individual stems because stem forms depend on many factors, such as tree species, local growth conditions and individual growing conditions for stems in the stand. Because of these influencing factors, so-called local tariff functions under specific growth conditions and for different tree species are used in forest inventories.

In Switzerland, the need for accurate stem volume functions gained attention in connection with the development and large-scale implementation of permanent-plot forest inventories at the forest enterprise level. The starting point was in the 1960s, when a set of tree-species-specific *stem volume functions* were established that predict the stem volume as a function of the stem $d_{1.3}$, the stem upper diameter

(7 m aboveground) and the stem length (Chap. 12). This approach was further developed when the NFI began in the 1980s.

In the NFI, these stem volume functions are used to: (a) derive the stem volumes for all second-stage tariff trees in the sample, and (b) calibrate highly localised tariff functions, in which various tree, stand and site variables are used as explanatory variables. The tariff functions are then used to predict a less accurate stem volume for all tally trees in the sample.

The details and accuracy of the different volume and tariff functions are explained in Chap. 12. Briefly, the tariff-function-based stem volumes available for the first-stage sample of tally trees are moderately accurate, while the volume-function-based stem volumes available for the second-stage subsample of tariff trees are very accurate.

2.3.4.2 Poisson Sampling of Tariff Trees

In NFI1 to NFI4 (years 2009–2014), a subsample $\ddot{S}(\omega)$ of tariff trees was randomly selected from the sample $\dot{S}(\omega)$ of tally trees at ω. Individual tariff trees were selected independently and with predefined (conditional) second-stage sample inclusion probabilities p_i.

Thus, under this *Poisson sampling* of tariff trees, the *generalised (two-stage) local density function* can be written as

$$\ddot{y}_F^{(X)}(\omega) = \sum_{i \in \dot{S}(\omega)} \hat{X}_i f_i(\omega) + \sum_{i \in \ddot{S}(\omega)} \frac{R_i f_i(\omega)}{p_i} \tag{2.11}$$

where $R_i = X_i - \hat{X}_i$ denotes the difference between the true[1] (volume function) stem volume X_i and the predicted (tariff function) stem volume $\hat{X}_i$.

With single-stage sampling, the local densities $\dot{y}_F^{(X)}(\omega)$ are predefined and fixed values for all sampling units in the frame. This is not the case for the generalised local densities under two-stage sampling. For a given sampling unit, $\ddot{y}_F^{(X)}(\omega)$ is a random variable with expectation $y_F^{(X)}(\omega)$, the local single-stage density of target variable X, the true (volume function) stem volume (Mandallaz 2008).

2.3.4.3 Fixed Association of Tariff Trees

The independent Poisson selection of tariff trees in each inventory has certain disadvantages if repeatedly applied in an inventory with permanent plots, especially for change estimation but also for the simple purpose of establishing an NFI database of re-measured tariff trees.

[1]It is assumed to be true in terms of the best value available.

For this reason, the approach for selecting tariff trees was changed during NFI4. Tariff trees are now associated with sampling units ω according to predefined associations rules, as done for tally trees. No randomness is involved in the selection of tariff trees, and a tariff tree remains associated with the sampling unit over successive inventories as long as the tree is not removed from the population.

In this revised approach, the local density function at point ω can formally be written as

$$\ddot{y}_F^{(\mathrm{X})}(\omega) = \sum_{i \in \dot{S}(\omega)} \hat{X}_i \dot{f}_i(\omega) + \sum_{i \in \ddot{S}(\omega)} R_i \ddot{f}_i(\omega) \qquad (2.12)$$

where $R_i = X_i - \hat{X}_i$ denotes the difference between the true (volume function) stem volume X_i and the predicted (tariff function) stem volume $\hat{X}_i$, and $\dot{f}_i(\omega)$ and $\ddot{f}_i(\omega)$ are the respective extrapolation factors for tally trees and tariff trees.

As the sampling of tally trees is done independently from the sampling of tariff trees, the Riemann integral of the local density function $\ddot{y}_F^{(\mathrm{X})}(\omega)$ is

$$\int_G \ddot{y}_F^{(X)}(\omega)d\omega = \int_G y_F^{(\hat{X})}(\omega)d\omega + \int_G y_F^{(R)}(\omega)d\omega = T_F^{(\hat{X})} + T_F^{(R)} = T_F^{(X)} \qquad (2.13)$$

the sum of the true (volume function) stem volume X over the population P_F of trees in F. We implicitly required R_i to be available for all tariff trees, so that $\ddot{S}(\omega) \subseteq \dot{S}(\omega)$, which, however, is a restriction easily fulfilled with appropriate allocation rules.

2.3.4.4 Optimal Sampling Schemes for Tariff Trees

The rules for optimal tally and tariff tree selection are as follows (Mandallaz 2008):

- First-stage (tally) trees should be selected with sample inclusion probabilities proportional to predictions: $\lambda_{A_{i.G}}/\lambda_G \propto \hat{X}_i$,
- Second-stage (tariff) trees should be selected with sample inclusion probabilities proportional to residuals: $\lambda_{A_{i.G}}/\lambda_G p_i \propto R_i$,
- The optimal conditional second-stage (tariff) tree sample inclusion probabilities are therefore $p_i \propto R_i/\hat{X}_i$.

Under the nested (concentric) fixed-area plot sampling of the NFI, the first-stage tally tree inclusion probabilities are constant for trees with $d_{1.3}$ >= 12 cm and $d_{1.3}$ < 36 cm, and higher but again constant for tally trees with $d_{1.3} \geq$36 cm. Because the tally tree inclusion probabilities are constant within each of the two categories of tree diameters, the optimal conditional inclusion probabilities p_i of tariff trees are proportional to the residuals R_i. The increase in the residuals is roughly proportional to the true volume: $R_i \propto X_i$.

2.3.4.5 Tariff Tree Selection in NFI1

A fixed association of tariff trees was applied, which included tally trees with $d_{1.3}$ >=12 cm and $d_{1.3}$ <60 cm if the trees are located in the plot sector 0–150 gon (151 gon out of the 400 gon of the full circle), and all tally trees in the $d_{1.3}$ class ≥60 cm. Therefore, when the inclusion probability of tariff trees is written in the form $\ddot{f}_i(\omega) = \dot{f}_i(\omega)/q_i$, we get

$$q_i = \begin{cases} 151/400 & \text{if } 12 \le d_{1.3.i} \le 59 \\ 1 & \text{if } d_{1.3.i} \ge 60 \end{cases} \tag{2.14}$$

Tariff Tree Selection in NF2 to NFI4 (Years 2009–2014) The optimised Poisson sampling selection included tally trees with $d_{1.3}$ >=12 cm and $d_{1.3}$ <60 cm if the trees are located in the plot sector 0–149 gon (150 gon out of the 400 gon of the full circle) with the (conditional) inclusion probabilities p_i given below. All tally trees with $d_{1.3}$ ≥60 cm were selected as tariff trees. The resulting conditional sample inclusion probabilities of tariff trees were

$$p_i = \begin{cases} 150/400 \times 0.000015 \times (d_{1.3.i})^2 \times 50 & \text{if } 12 \le d_{1.3.i} \le 35 \\ 150/400 \times 0.000015 \times (d_{1.3.i})^2 \times 20 & \text{if } 36 \le d_{1.3.i} \le 57 \\ 150/400 & \text{if } 58 \le d_{1.3.i} \le 59 \\ 1 & \text{if } d_{1.3.i} \ge 60 \end{cases} \tag{2.15}$$

Tariff Tree Selection NFI4 (Years 2015–2017) Two aspects of tariff tree sampling were changed. First, the proportion of tariff trees was adapted so that the originally planned inclusion probabilities could be followed better. Second, a fixed association of tally trees was introduced to maximise the number of tariff trees re-measured over successive inventories.

With $\ddot{f}_i(\omega) = \dot{f}_i(\omega)/q_i$, the new values are

$$q_i = \begin{cases} 5.625 \times 10^{-6} \times (d_{1.3.i})^{2.00} \times 50 & \text{if } 12 \le d_{1.3.i} \le 35 \\ 1.125 \times 10^{-8} \times (d_{1.3.i})^{3.75} \times 20 & \text{if } 36 \le d_{1.3.i} \le 59 \\ 1 & \text{if } d_{1.3.i} \ge 60 \end{cases} \tag{2.16}$$

where 5.625×10^{-6} is equal to 150/400 × 0.000015, the value used in NFI2 to NFI4 (years 2009–2014). 1.12510^{-8} was adopted to constantly increase inclusion probabilities for tariff tree selection in these $d_{1.3}$ classes. Figure 2.2 illustrates the conditional and relative inclusion probabilities of tariff trees in the different inventories, divided into diameter classes, and Fig. 2.3 shows the respective number of tariff trees it was expected would be measured in an annual sample.

The optimal proportion of tariff trees is roughly proportional to the basal area, with appropriate basal area factors defined accordingly. Therefore, an obvious

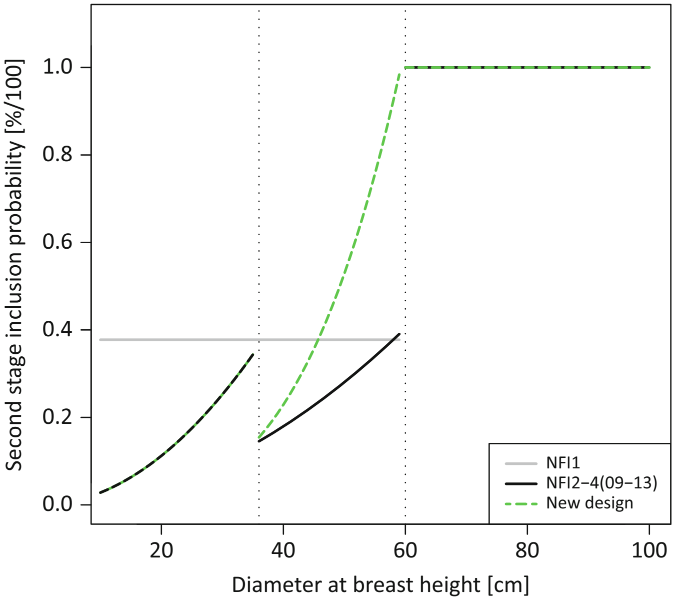

Fig. 2.2 Conditional second-stage inclusion probabilities $\boldsymbol{p_i}$ of tariff trees under Poisson sampling and relative inclusion probabilities $\boldsymbol{q_i}$ of tariff trees under a fixed association in NFI1, NFI2–NFI4 (years 2009–2014), and NFI4 (year 2015–2017)

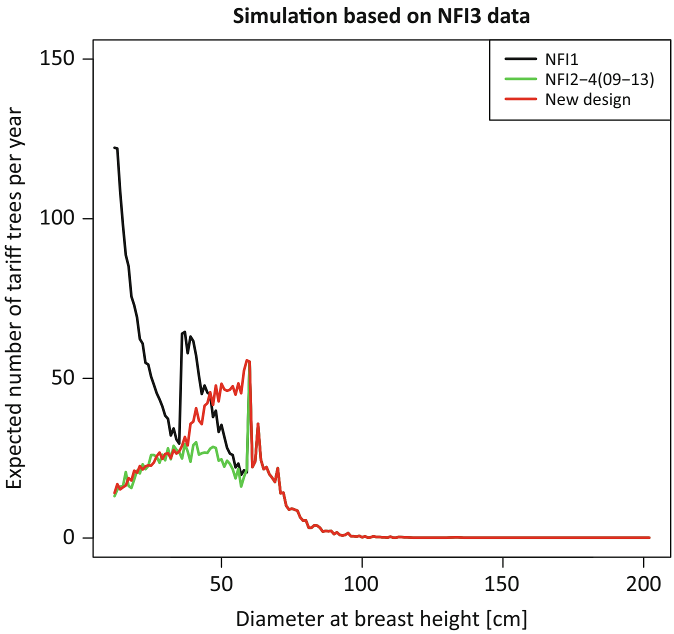

Fig. 2.3 Expected number of tariff trees in an annual panel of the NFI as a function of $\boldsymbol{d_{1.3}}$. Values are given for NFI1, NFI2–NFI4 (years 2009–2014), and NFI4 (2015–2017)

choice for the combined plot configuration of tally and tariff trees would be to overlap the fixed-radius plot sampling of tally trees with an angle-count sampling of tariff trees.

In the NFI, a concept similar to the well-known angle count sampling of tally trees was adopted for the new system of tariff tree selection. Because the tariff trees are concentrated in the plot sector between 0 and 150 gon, the new, sector sampling of tariff trees continues to favour tariff tree association in these sectors of the plots. The idea is to replace the optimal horizontal angles with optimal sectors for tariff trees selection. Using two plots of different sizes for selecting the tally tree causes a slight difficulty. To give an example: assuming the planned proportion of tariff trees in $d_{1.3}$ class 30 cm is $q = 0.25$ translates into a sector sampling of tariff trees in sector 0–100 gon, i.e. 25% of the full circle.

The resulting sector openings $s_{\text{small}.\,i}$ and $s_{\text{large}.\,i}$ for the small 200 m^2 plot and the larger 500 m^2 plot for tariff tree selection are

$$s_{\text{small}.i} \leq \begin{cases} 400 \times 5.625 \times 10^{-6} \times (d_{1.3.i})^{2.00} \times 50 & \text{if } 12 \leq d_{1.3.i} \leq 35 \\ 400 \times 5.625 \times 10^{-6} \times (35)^{2.00} \times 50 & \text{if } 36 \leq d_{1.3.i} \leq 44 \end{cases} \tag{2.17}$$

And

$$s_{\text{large}.i} \leq \begin{cases} 400 \times 1.125 \times 10^{-8} \times (d_{1.3.i})^{3.75} \times 20 \times Q_i & \text{if } 36 \leq d_{1.3.i} \leq 44 \\ 400 \times 1.125 \times 10^{-8} \times (d_{1.3.i})^{3.75} \times 20 & \text{if } 45 \leq d_{1.3.i} \leq 59 \\ 400 & \text{if } d_{1.3.i} \geq 60 \end{cases} \tag{2.18}$$

where stem diameters $d_{1.3.i}$ are down-rounded to centimetres and

$$Q_i = \frac{50}{50 - 20} \times \left(1 - \frac{5.625 \times 10^{-6} \times (35)^{2.00}}{1.125 \times 10^{-8} \times (d_{1.3.i})^{3.75}}\right). \tag{2.19}$$

is just a numerical term for trees with $d_{1.3}$ >=36 cm and $d_{1.3}$<=44 cm class trees with different sector openings are selected in the small and large NFI plots.

The inclusion probabilities and respective sector openings are given in Table 2.1 in the appendix.

Tariff Tree Non-response The number of tariff trees selected in the NFI sample is always smaller than planned, mainly because the tariff trees selected unconditionally or randomly do not qualify for upper stem and/or stem length measurements. At the estimation stage, the nominal shares q_i and the conditional inclusion probabilities p_i are adjusted for non-response. The proportion of non-response is determined in *response homogeneity groups*, which are formed for the five production regions of Switzerland, three elevation classes, the categories coniferous and broadleaved, and seven $d_{1.3}$ classes.

Boundary Correction for Tariff Trees Because of the anisotropy in the new association of tariff trees in plot sectors, the theoretically correct extrapolation factors for tariff trees can become extremely large for tariff trees near the southern border of the sampling frame. On the other hand, they can become smaller than the first-stage tally tree extrapolation factors for tariff trees near the northern border of the sampling frame.

In this sense, an isotropic association of tariff trees would clearly be preferable. The only reason an anisotropic sector association of tariff trees is currently in use in the NFI is to maximise the number of tariff trees with repeated upper stem diameter and stem length measurements. Because anisotropy is not expected in the distribution of frame boundaries, the extrapolation factors are always $\dot{f}_i(\omega)/q_i$, where the first-stage tally tree extrapolation factor $\dot{f}_i(\omega)$ may be boundary corrected.

2.3.5 Sampling Pieces of Lying Deadwood

The infinite population principles can be adapted for use in the line intersect sampling of pieces of lying deadwood on forest land (Kaiser 1983; Gregoire and Valentine 2007). Compared to tree sampling, the population elements are now *objects or particles* of arbitrary form and dimension located on forest land F. Instead of circular plots, one or more lines are used for identifying population elements associated with sampling units ω in G.

The exact solution requires: (a) the definition of position, orientation and length of transect lines associated with a sampling unit ω in G, (b) rules for the association of particles with transect lines and, therefore, for the clear-cut definition of the members of the local sample at point ω, (c) a clear field protocol for the measurements conducted on these particles; and (d) the local density function itself.

The NFI line-intersect sampling method for pieces of deadwood is one of several options used in inventory practice. It involves the following rules (Düggelin and Keller 2017):

- Three transect lines of fixed orientations 35 gon, 170 gon and 300 gon are defined for each sampling unit ω in G
- Transect lines are 10 m in length and start with an offset of 1 m from the sampling unit's position at ω. The length of a transect line is the length projected to the plane; the length applied in the field, following the inclination of the terrain, may therefore be more than 10 m (cosine formula)
- Transect lines stop at the boundary of the sampling frame for tally tree sampling (Chap. 9), and the length of this reduced transect line is registered in the field protocol
- A piece of deadwood is considered a member of the local sample at ω only if at least one of the transect lines intersects with the *central axis* of the piece. If the deadwood piece is forked or otherwise branched, all parts and branches are examined separately

- The diameter of a deadwood piece is measured perpendicular to its central axis at the point of intersection between the central axis and the transect line
- The inclination (angle) of the central axis of the deadwood piece is measured and recorded at this point of intersection during fieldwork

The local density function used in the NFI for the *volume of lying deadwood pieces* in domain F at sampling unit ω is defined as

$$y_F(\omega) = \frac{1}{m(\omega)} \sum_{h=1}^{m(\omega)} \frac{\pi^2}{8\, l_h} \sum_{i \in S_h} \frac{d_{m.i}^2}{\cos \alpha_i} \tag{2.20}$$

where $m(\omega)$ is the number of transect lines installed at point ω (usually $m(\omega) = 3$), l_h is the projected length of transect line h, and $d_{m.i}$ and α_i denote the means of two perpendicular diameter measurements and the inclination, respectively, of deadwood piece i, which is part of the local sample S_h of deadwood pieces intersected by transect line h. The estimator is an approximation. Its derivation can be found in Gregoire and Valentine (2007).

2.3.6 Sampling Young Trees

In the NFI, the sampling scheme for *young trees*, also termed a *young forest assessment,* encompasses the population of *living trees* with a height of at least 10 cm but a $d_{1.3}$ <12 cm. The scheme was changed several times between inventory cycles (Schwyzer and Lanz 2010). The common feature of all sampling schemes is that the centre point(s) for young-tree data collection is offset from the standard sampling unit's position by 10 m, and that the trees are not identified for re-measurement in a consequent inventory. The (auxiliary) plot centres for the young tree assessment are, however, considered permanent when the net change in the population of young trees was occasionally assessed between inventories.

Since NFI4, *nested (concentric) fixed-area plot sampling* has been used for the association of living young trees with sampling units, so that the local density function follows the usual rules explained in Sect. 2.3.2.1. The only deviation from standard procedures is the *offset of the plot centre for young tree sampling from the standard sampling unit location* at point ω. To obtain exact and design-unbiased estimates, boundary correction measures would be needed for young trees near the sampling frame boundary, which are currently not implemented. A correction currently applied in the NFI is the establishment of an alternative plot centre in the direction away from the sampling frame boundary in cases where the plot centre for young tree sampling is located outside the sampling frame. Such replacements approximately compensate for young forest plot centres that are not detected and established because the centre of the standard sampling unit is located outside the sampling frame. Identical frames are used for tree and young tree sampling.

In some of the earlier inventories, *two subplot centres,* offset by 10 m from the centre of the sampling unit in opposite directions, were used *for young tree sampling*. Such a sampling unit, consisting of two or more subunits, is usually understood as a cluster, and cluster sampling gives rise to certain adaptations in the mode of sampling and in the choice of estimators. However, because the information needed on young forest does not have to be so precise, the local density for a sampling unit at point ω was simply computed as the arithmetic mean of the two densities measured on the young tree subplots.

Nearest Tree Sampling The assessment of some of the young tree attributes is complex and costly. For this reason, the field protocol restricts the assessments of some of the young tree target variables and characteristics to the young tree individual positioned closest to the subplot centre. A typical target variable of interest in this type of sampling is the presence and type of damage from game browsing observed on young trees.

For this data, the local density function is defined as

$$y_F^{(C)}(\omega) = \iota_F(\omega)\, \iota_i^{(C)}(\omega) \tag{2.21}$$

where $\iota_F(\omega)$ indicates whether ω is in the domain F of interest and the indicator variables

$$\iota_i^{(C)}(\omega) = \begin{cases} 1 & \text{if } i \text{ closest to } \omega \text{ has characteristic } C \\ 0 & \text{else} \end{cases} \tag{2.22}$$

are defined with respect to characteristic C of interest, formally for the entire population P_F of N_F young trees in F.

Under such a nearest tree sampling design and with a local density function defined as above, the infinite population is actually a *Voronoi partitioning* of the domain F of interest (and $y_F^{(C)}(\omega) = 0$, if $\omega \notin F$). The young tree positions $u_i \in F$ are the seeds (centres) of associated Voronoi cells V_i of surface area λ_{V_i}. With N_F young trees in F, then $F = \cup_{i=1}^{N_F} V_i$ and $\lambda_F = \sum_{i=1}^{N_F} \lambda_{V_i}$. For a given young tree characteristic C, the total $T_F^{(C)} = \int_G y_F^{(C)}(\omega) d\omega = \lambda_{F_c}$ is the total surface area of Voronoi cells in F occupied by young trees with characteristic C.

In the NFI, a separate young tree that is nearest to plot centre is selected for several subpopulations of young trees, which are defined according to height and diameter class (Düggelin and Keller 2017).

2.4 Sampling Design of the NFI

With the definition and introduction of a sampling frame and sampling units with associated local densities, the actual population of interest and its elements, such as trees or pieces of lying deadwood, become obscured. The randomised sampling is actually implemented on the continuous plane (or universe) of points in the frame.

2.4.1 Basic Sampling Methods in Area Sampling

Horvitz-Thompson Theorem for the Continuous Universe The Horvitz-Thompson estimator plays a central role in survey sampling (Horvitz and Thompson 1952), but the theory was developed for finite populations and is mainly recognised in that context. We briefly mention an extension of the theorem to sampling from an infinite population.

A common reference is Cordy (1993), who defines *inclusion densities* $\pi(\omega)$ on the frame G, which may be thought of as a local measure of the number of points to be selected per unit area. In forest inventory practices, $\pi(\omega)$ are constant over the entire sampling frame or within sampling strata. Cordy also mentions a continuous universe version of importance sampling, which results in unequal inclusion densities.

For a sample S of m points ω in G with associated inclusion densities $\pi(\omega)$ and observed responses $y_X(\omega)$, the extended Horvitz-Thompson estimator

$$\hat{T}_G^{(X)} = \sum_{\omega \in S} \frac{y_X(\omega)}{\pi(\omega)} \tag{2.23}$$

is shown to be unbiased for the total $\hat{T}_G^{(X)} = \int_G y_X(\omega) d\omega$ of the Riemann integrable function $y_X(\omega)$ defined over G. Cordy also derives expressions for the theoretical variance $\mathbb{V}\left\langle \hat{T}_G^{(X)} \right\rangle$ and a sample-based estimator of this variance for which the pairwise inclusion densities $\pi(\omega, \omega')$ have to be calculated for the sampled points ω and ω'. The inclusion densities play a similar role to *inclusion probabilities* in finite population sampling. They describe the design according to which the sample is selected, and they are needed for design-based estimation.

Under a uniform distribution of random points in G and with m independently selected points ω, the inclusion densities are constant over G and given by $\pi(\omega) = m/\lambda_G$, with joint densities $\pi(\omega, \omega') = m(m-1)/\lambda_G^2$.

The Infinite Population Approach to Forest Inventories In forest inventories sampling units are usually distributed with constant density, although sometimes densities vary among strata. The original description of the infinite population approach starts with an assumed sample of only one random point ω, uniformly distributed in G (Mandallaz 2008; Eriksson 1995b).

Then,

$$t_X(\omega) = \lambda_G y_X(\omega) = \hat{T}_G^{(X)} = \hat{T}_F^{(X)} = \sum_{i=1}^{N_F} \frac{X_i \iota_{i.S(\omega)}}{\pi_i}. \tag{2.24}$$

turns out to be the Horvitz-Thompson estimator of the total $T_F^{(X)}$ of target variable X in F, and also of target variable X in G, because $G \supseteq F$, by definition.

$y_X(\omega)$ is the local density and $t_X(\omega)$ is the local density expanded to totals, without explicit reference to the area domain F of interest, $\pi_i = 1/\lambda_G f_i(\omega)$ are sample inclusion probabilities, and $\iota_{i.\ S(\omega)}$ sample membership indicator variables for elements in P_F.

The expectation of the $t_X(\omega)$ for a random point ω in G is, by construction, the total $T_G^{(X)}$ of X in G, which is equal to the population parameter of interest, the total of X in F.

The theoretical variance of $\hat{T}_G^{(X)}$ can be given using the Horvitz-Thompson theorem. Then the joint inclusion probabilities $\pi_{i\,i'} = \lambda_{A_{i.G} \cap A_{i'.G}}$ have a physical interpretation as surface areas of overlapping inclusion zones between pairs of trees. Because most second-order inclusion probabilities are zero, the Horvitz-Thompson variance estimator is not available for samples originating from the random selection of a single point.

However, with a replicate sample of m independently and uniformly distributed random points, a design-unbiased estimator of the true variance of $\hat{T}_G^{(X)} = \lambda_G \hat{Y}_G^{(X)}$ $= \frac{\lambda_G}{m} \sum_{j=1}^{m} y_X(\omega_j)$ is immediately available and given by

$$\hat{\mathbb{V}}\left\langle \hat{T}_G^{(X)} \right\rangle = \frac{\lambda_G^2}{m} \frac{\sum_{j=1}^{m} \left(y_X(\omega_j) - \hat{Y}_G^{(X)} \right)^2}{m-1}. \tag{2.25}$$

Systematic Sampling In forest inventories, sampling units are usually selected in a systematic way and require the definition of two non-collinear vectors in $\mathfrak{R}^2$ which define a partitioning of G (actually $\mathfrak{R}^2$) into fundamental cells of known size λ_C and with the same shape. Under aligned systematic sampling, the selection of a single random point in one of the fundamental cells determines the position of the sampling units in all cells.

Systematic sampling usually leads to sample inclusion probabilities $\pi_i > 0$ for all elements in the population, but the second-order sample inclusion probabilities $\pi_{i\,i'} = 0$ for most pairs of population elements. The immediate consequence is that

$$\hat{T}_G^{(X)} = \lambda_C \sum_{\omega_j \in G} y_X(\omega_j) \tag{2.26}$$

is a design-unbiased point estimate of the true total $T_G^{(X)}$ of X over frame G, but – as a matter of mathematical principle – there is no design-unbiased estimator for the variance of $\hat{T}_G^{(X)}$ available under systematic sampling.

The advantage of systematic sampling over independent random point sampling is the even spread of the sample over the entire study area, as well as the balanced coverage of area subdomains of interest, with the number of sampling units proportional to the size of the subdomains.

The NFI approach to producing estimates of the unknown variance under systematic sampling is to assume that the sampling units are independently and uniformly distributed over G. It is often argued, and has been demonstrated in simulation studies for different populations, that the independent random point estimators tend to overestimate the true variance for systematic sampling. The implicit assumption of independent sampling units may, however, be reasonable because silvicultural forest treatments in Switzerland tend to be small scale and the spatial correlation ranges observed in geostatistical case studies are low.

Cluster Sampling and Stratification For cluster sampling, specific instructions are need for sampling units at the sampling frame boundary, and also for subplot centres off-set from the centre of sampling units as in the NFI young tree survey. For young tree sampling, plot centres are relocated to avoid unintentional manipulation of the population prior to observation and measurement in the field. In earlier NFIs, two separate subplots, i.e. true clusters, were used for assessing young trees. The three transect lines used in the assessment of lying deadwood can also be considered clusters. In both cases, the NFI estimation procedure is to simply pool together the two plots or three lines to form a single unit, partially ignoring the theoretically correct solutions for the treatment of clusters at sampling frame boundaries (Gregoire and Valentine 2007; Mandallaz 2008).

The density of sampling units (points) in NFI is constant throughout all of Switzerland. Stratification has been proposed occasionally to reduce the markedly higher costs of accessing field plots in mountainous regions (Lanz 2000). On the other hand, the information required for forest policy and management in the protective forests in these regions is particularly important. Hence, a uniform distribution of sampling units is largely accepted with the current requirements for information. Post-sampling stratification is, however, applied in the computation of the estimates (Sect. 2.5.2).

2.4.2 Sampling Frames of the NFI

The NFI sampling frame covers the entire surface area of the country and includes water in the form of lakes, rivers and streams, and large areas of unproductive land, such as rocks and glaciers in the mountains.

This sampling frame, which includes approximately 65% non-forest land, is primarily needed for estimating the forest area. Important gains in precision could be achieved mathematically if the sampling frame were reduced to include less non-forest land, for example only 10%. No such reduced sampling frame has been available for the NFI so far, mainly because the landscape in Switzerland is very fragmented.

For cost reasons, the NFI uses a *specific sampling frame for field-data collection*, which is limited to points located in forests or shrub forests. The manual aerial-photo interpretation still covers the entire country, and is used mainly to detect new forest plots and plots clearly located outside forest land. These plots are then eliminated from the (annual) list of sampling units to be visited by field crews for data collection. Protection against omission errors is built into the aerial-image interpretation to guarantee the visit of sampling units in cases of ambiguity. An average of approximately 5% of sampling units visited by field crews turn out to be located on non-forest land.

At the estimation stage of the inventory, all sampling units located outside forests are assigned a local density of $y_X(\omega) = 0$.

Inaccessible Sampling Units Sampling units ω located on forest land and not accessible for field-data collection remain in the sample, with a local density set to $y_X(\omega) = 0$ for all target variables except for forest area estimation.

A boundary between parts of the forest that are accessible for data collection and those that are inaccessible may cross a plot located on forest land, in which case the sampling unit is treated as: (a) an inaccessible field plot without field-data collection if the sampling unit centre ω is located in the inaccessible part of the plot; or (b) a sampling unit with field-data collection if the sampling unit centre is located in the accessible part of the plot. In case (b), the boundary between the accessible and inaccessible parts of the plot is considered a reduction line (Sect. 2.3.3).

The estimates in the standard result tables in NFI, such as growing stock or biomass, always refer to parts of the forest that are accessible for data collection Inaccessible sampling units could be treated as non-response, but have not so far in NFI for two main reasons: (a) imputation is unreliable because auxiliary (replacement) data may be lacking; and (b) effects of omitting these plots are considered marginal because many of these inaccessible plots are located on unproductive land and topographically very exposed sites, such as steep slopes and rocky terrain.

2.4.3 NFI Panels for Terrestrial Data Collection

The square grid used for field-data collection in the NFI has a density of one sampling unit per 2 km^2, resulting in a total of approximately 20,000 sampling units, of which around 7500 are located on forest land. The plot centres were permanently installed during the first inventory cycle, NFI1 (1983–1985), and the field-data collection was repeated in NFI2, NFI3 and NFI4. Statistically, the inventory may be understood as a *panel or longitudinal survey* with one single random event that defined the orientation and position of the original grid in 1983.

The chronology of field-data collection is illustrated graphically in Fig. 2.4. The inventories, NFI1 to NFI3, were carried out during three distinct cycles, each of which ran for 3 years and included field-data collection on the full set of sampling units. A new, continuous mode of field-data collection was introduced in NFI4 in which the original panel is split into nine annual panels in the form of interpenetrating grids, each covering the entire country, and field data is collected from all sampling units of an annual panel within one calendar year.

2.4.4 NFI Panels for Auxiliary Data Collection

In the NFI, manual stereo-image interpretation has always been used to exclude plots that are clearly ‘non-forest’ from the field-data collection. This minimises the very high cost of accessing plot centres in some of the mountainous regions in the country.

In NFI2, a densified square grid with eight sampling units per 2 km^2 was installed for the manual stereo-image interpretation and the assessment of auxiliary data. These units have been used in a two-phase estimation procedure to partly mitigate

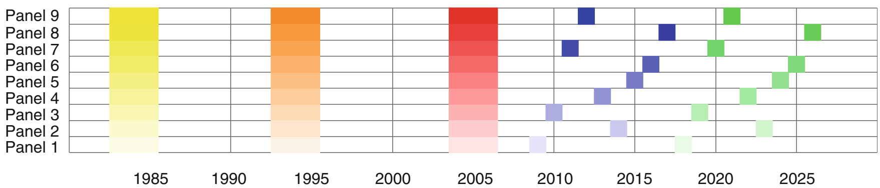

Fig. 2.4 Graphical representation of the inventory cycles for field-data collection in the NFI with the calendar years of the field-data collection on the x-axis. The colours yellow, orange, red, blue and green refer to NFI1 to NFI5, respectively, and the different shades of each colour to the nine annual panels

increased sampling error due to the 50% reduction in the original sampling grid. Manual stereo-image interpretation was repeated in NFI3 on the same densified grid.

Currently, the auxiliary data used in the two-phase estimation procedures is collected periodically and derived semi-automatically over the entire country (Chap. 7). The relevant aspects of integrating the auxiliary data into the estimation procedures are: (a) GPS measurements during field-data collection to identify the exact position of sampling unit centres; (b) identification and production of suitable response data from the vast repositories of auxiliary raw data; (c) calibration of optimised regression models for the standard output production; and (d) analyses and case studies for the custom-tailored estimation results for small areas.

These aspects of auxiliary data integration are outlined in Sect. 2.5.2.

Annual Panels The system of field data collection in NFI4 was made with the following objectives and considerations in mind:

- Change the NFI survey into a Project with a continuous, regular budget
- Promote the availability of timely information, if needed
- Remain within the same overall budget as in the periodic system
- Give more priority to estimating change
- Increase flexibility so as to be able to adapt field protocols at any time
- Maintain the existing panels of permanent plots

At the statistical sampling design level, two main aspects were investigated: (1) the selection of sampling units for annual panels from the existing square grid of terrestrial plots in order to minimise that the travel time and distance between sampling units. A rotating system allows the concentration of the sampling in a different region of the country each year and thus should reduce the cost of travel to a minimum. Using annual interpenetrating square grids requires far more travelling since each grid covers the entire country. With nine annual panels, this corresponds to a plot density of one plot per 18 km^2. The travel costs involved in maintaining the basic idea of annual interpenetrating panels, each covering the entire country, can be reduced by selecting clusters of two to nine neighbouring sampling units from the existing grid to form panels consisting of systematically distributed clusters. The within-cluster density corresponds then to the original density of one plot per 2 km^2, while the cluster density is one cluster per 32 km^2 for a cluster size of two plots and one cluster per 162 km^2 for a cluster size of nine plots. With the larger cluster size, however, sizeable regions may not be covered at all in specific years.

An interpenetrating, non-clustered square grid configuration was finally chosen for the NFI (Fig. 2.5). The relative position of the annual square sampling grids was selected such that the spatial spread of sampling units is maximised when subsequent annual sampling grids are pooled together at the estimation stage of the inventory. For example, if the positions of the annual panels 3 and 6 were to be interchanged, the pooled sampling units of annual panels 1–3 would be arranged in lines with short distances between sampling units in the North-South direction and large distances between sampling units in the East-West direction. The pooled sampling units of annual panels 4–6 would then be clearly arranged in clusters of three sampling units.

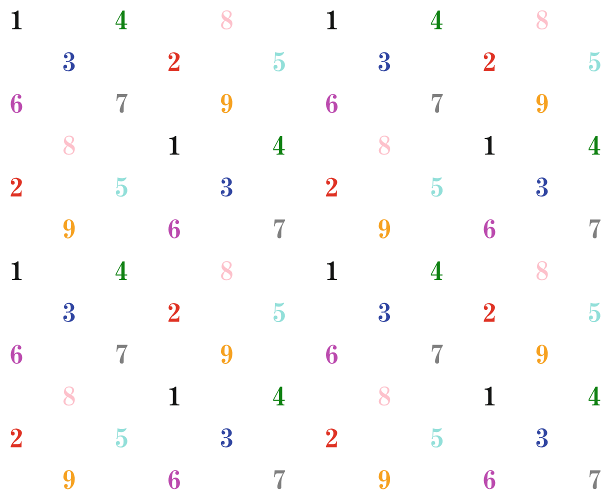

Fig. 2.5 Layout of the nine annual panels of the NFI

Under the periodic system, field data from about 7000 sampling units was collected every 10 years. Under the new annual system, assuming an approximately equal overall budget, field data from approximately 700 sampling units can be collected every year. The final decision was to maintain the entire sample of about 7000 permanent plots and to create 9 annual panels with about 7500 plots re-measured every year. The resulting interval between plot re-measurements is 9 years.

A major point of discussion has been the length of time between the re-measurements of the permanent plot. With long intervals in-between, more changes in the population remain unobserved, and the time when a specific change happens cannot be detected precisely. Many European NFIs have switched to annual inventory systems with permanent plots that are re-measured every 5 years.

Remeasuring the Swiss permanent plots every 5 years would mean ignoring about 50% of the original sample as field data, which could only be collected from about 700 plots per year under the budget constraints. Hence, the trade-off is essentially between reducing the existing sample of permanent plots by 50% or more, and detecting changes less precisely because the plots are measured less frequently. In the NFI, the interval between plot measurements has always been ten (or more) years. Moreover, various (external) data users expressed a strong interest in maintaining as many permanent plots as possible. Therefore, it was decided to create nine annual panels.

Pooled Samples With annual panels that each cover the entire study area, one option is to pool consecutive years together into a larger sample for the estimations. The general effects of pooling panels and some important considerations are:

- The precision of estimates increases with the number of pooled annual panels, but at the price of a less precise assignment of the estimates to a specific point in time (calendar year)

- The trade-off between the two effects depends heavily on effective and/or supposed changes in the population, which may differ markedly between target variables and area domains of interest
- Pooled samples tend to smooth peaks of cyclic change in the population and detect trends later than annual panels
- Consecutive (unpooled) annual panels may suggest changes in the population, whereas in fact only random sampling effects are observed

Since the introduction of the annual system of field-data collection, the NFI produced intermediate results (NFI4a) using the pooled sample of the three annual panels from 2009 to 2011, intermediate results (NFI4b) using the pooled sample of the five annual panels from 2009 to 2013, and finally, results (NFI4) from the pooled sample of the nine annual panels from 2009 to 2017.

2.5 Estimation Procedures

2.5.1 Basic Estimators

Under the systematic sampling design of the NFI,

$$\hat{T}_G^{(X)} = \lambda_C \sum_{j \in G} y_X(\omega_j) \tag{2.27}$$

is an exact design-unbiased estimator of the total $T_G^{(X)}$ of the local density function $y_X(\omega)$ over G (Mandallaz 2008), and therefore the total of target variable X in P_F, where λ_C denotes the size (surface area) of the basic cell of the systematic grid.

The number of sampling units m_G in the sampling frame G is a random number under systematic sampling, with two consequences. First, the estimator

$$\hat{Y}_G^{(X)} = \frac{1}{m_G} \sum_{j \in G} y_X(\omega_j) \tag{2.28}$$

for the true mean spatial density $Y_G^{(X)} = T_G^{(X)} / \lambda_G$ of $y(\omega)$ over G is not strictly design-unbiased because $\hat{Y}_G^{(X)}$ is a ratio of two random variables. Second, conditioning on the realised sample size m_G, $\hat{T}_G^{(X)}$ is biased whenever $\frac{\lambda_G}{\lambda_C} \neq m_G$.

For this reason, $\hat{T}_G^{(X)} = \lambda_G \hat{Y}_G^{(X)}$ has been used to estimate totals since NFI2.

Because the NFI sampling frame extends over the entire country, the mean spatial density $Y_G^{(X)}$ of a local density $y_X(\omega)$ over the entire frame G is almost never of interest. Many population parameters of interest can be expressed, however, as the ratio of the totals of two local densities $y_X(\omega)$ and $y_Z(\omega)$, both defined over the entire frame G. In this case

$$\hat{R}_G^{(X/Z)} = \frac{\hat{T}_G^{(X)}}{\hat{T}_G^{(Z)}} \tag{2.29}$$

is used to make an asymptotic design-unbiased estimate of the unknown population parameter $R_G^{(X/Z)}$.

Post-sampling Strata The NFI uses stratified estimators for country estimates because the standard output tables are formatted so that estimates are always presented for the entire country and for a set of the main regions of interest, which are usually the five production regions of Switzerland. Other sets of primary interest used in the output tables could be cantons, biogeographic regions and forest districts.

The estimates within these post-sampling strata of known surface area are produced with the estimators given above. To maintain *additivity of regional estimates with the estimate for the country* within output tables, the post-stratified estimator

$$\hat{T}_G^{(X)} = \sum_{h=1}^{H} \hat{T}_h^{(X)} = \sum_{h=1}^{H} \lambda_h \hat{Y}_h^{(X)} = \sum_{h=1}^{H} \frac{\lambda_h}{m_h} \sum_{j \in h} y_X(\omega_j) \tag{2.30}$$

is used for estimating totals, and the combined ratio estimator

$$\hat{R}_G^{(X/Z)} = \frac{\sum_{h=1}^{H} \hat{T}_h^{(X)}}{\sum_{h=1}^{H} \hat{T}_h^{(Z)}} \tag{2.31}$$

for estimating ratios $R_G^{(X/Z)}$.

The disadvantage of this approach is that the estimates for the entire country vary (slightly) numerically in the output tables according to the set of regions used for the repartitioning of the results. The numerical differences between the estimates are, however, very small because the strata and sample sizes are reasonably large, and all estimates remain design-unbiased. Technically, the minimum differences in the estimates reflect the different sources of auxiliary data used for estimation. For total growing stock, for instance, for which the relative standard error of the estimates with different sets of post-sampling strata is around 1%, the difference between estimates is less than 0.1%.

Variance Estimators The NFI variance estimators are those derived under the assumption of an independent distribution of the sampling units (Sect. 2.4.1). No alternatives have yet been implemented. A few, mostly undocumented, attempts have confirmed that the spatial auto-correlation between sampling units is very low for most target variables. The NFI standard error estimates can be assumed to have a tendency to overestimate the true error under the systematic sampling. This makes it less likely that stakeholders will interpret the results over-optimistically. The estimators are documented in Sect. 20.8.2 and follow the recommendations in the literature.

2.5.2 Use of Auxiliary Data

The NFI uses auxiliary data made available on a dense grid in an estimation procedure known as *two-phase (double) sampling for stratification*. The current sampling grid from which auxiliary data is derived has a density of one sampling unit per hectare and includes sampling units located both within and outside the forest.

The primary auxiliary data is a recently produced forest cover map (Chap. 7). Additional data includes: (a) vegetation height characteristics retrieved from digital stereo images (Chap. 7), where the main variables are means, medians and upper quantiles derived on circular plots 500 m^2 in area; (b) the proportion of coniferous trees in the total canopy cover, recently derived from four-band aerial images (Chap. 7) and (c) other spatial data, such as elevation, aspect and slope, obtained from digital terrain models. All raw data is provided by swisstopo, the Swiss Federal Office of Topography, and is prepared for use in the standard estimation procedures of the NFI.

Regression-tree modelling was used to calibrate the post-strata categories. The growing stock estimation was optimized (Pulkkinen et al. 2018). The estimation procedure can be understood as model-assisted estimation with an ANOVA-type regression model in a two-phase sampling framework, where the first-phase sample of sampling units is selected uniformly over the entire territory of Switzerland, and the second-phase subsample of sampling units with field data is a systematic subsample of the first-phase sample. In the derivation of the estimators, the assumption is a simple random sampling (without replacement) of second-phase plots from the first-phase sample of plots with auxiliary data. The inferential framework and estimators were introduced in NFI2 (Köhl 1994, 2001). Slight adaptations to the methods used for model building and for variance estimation have been implemented recently for NFI4. The estimators are given in Sect. 20.8.2.

A few aspects are worth noting:

- The two-phase estimation procedure is a so-called *model-assisted estimation technique* (Särndal et al. 2003; Mandallaz 2008), which leads to approximate design-unbiased estimates regardless of how well the model fits the data.
- Positional matching of sampling units for field-data collection and for auxiliary data collection is assumed and a homogeneous source of auxiliary data must be used. Modern automatic classification systems usually meet this assumption better than earlier manual image interpretation, and the assumption is becoming easier to fulfil as technology progresses.

- The primary gain in the context of the NFI is explicit and implicit total forest area estimation, where implicit forest area estimation is involved whenever the total of a target variable over the entire forest is estimated. This procedure divides the territory into a post-stratum of *presumed forest* and a post-stratum of *presumed non-forest*. This markedly increases the precision of the total estimates, even for target variables, such as lying deadwood, that are otherwise not strongly correlated with the auxiliary data.
- Auxiliary data related to vegetation height further increases the precision of the total and mean spatial density estimates for target variables, such as growing stock and biomass, which are clearly related to vegetation height. While gains in precision are achieved when estimating growing stock over the entire forest area, the gain may be marginal if the growing stock is estimated over certain area subdomains, such as private forest, or for certain subpopulations, such as beech trees, if auxiliary data indicating these subdomains and subpopulations is not available. For example, the relative standard errors of the single-phase estimates of total growing stock for the entire country in forest and private forest is 1.5% and 2.7%, respectively. Under two-phase sampling, the sampling error is reduced by 40% for all forest land, but by only 15% for the private forest land.
- The double sampling post-sampling strata have been optimised for growing stock estimation, but they are used for virtually all target variables and standard NFI outputs for the following reasons: (a) estimates for area subdomains, such as private and public forests, or for subpopulations, such as broadleaved and coniferous trees, would otherwise not remain numerically additive; (b) the effort to calibrate and adopt a model for each target variable is considerable; and (c) auxiliary data for area subdomains and subpopulations are not readily available.

2.5.3 Domain Estimation

Estimates for parts of the population distinguished according to certain criteria are regularly provided in the standard result tables of the NFI. From a statistical point of view, the related terminology and methods are not always unambiguous. In this section, we therefore provide a short overview of specific aspects of domain estimation in forest inventories.

Subpopulations We use the term *subpopulation* at the population element level, i.e. for population elements (trees) with a certain characteristic. Hence, when estimating growing stock, separate estimates are often required for the subpopulations of living broadleaved and living coniferous trees. In the infinite population approach to forest inventories, the actual population and population elements of interest become obscured at the estimation stage of the survey.

Technically, the handling of subpopulations is straightforward with subpopulation indicator variables applied at the element level in the derivation of plot-level local densities. There is no immediate difficulty in the estimation of totals and spatial means of rare subpopulations unless the local density is zero for most sampling units. The sample size itself remains constant.

A special case is the estimation of the mean of target variables at the population element level, for example the mean stem volume (of a given species). Under the infinite population approach, the estimator is a ratio, with an estimate for the total volume (of a given species) in the numerator and an estimate for the total number of stems (of a given species) in the denominator.

Area Subdomains In survey planning in general, planned and unplanned domains are commonly distinguished. Planned domains are domains of high interest and relevance for optimised allocation of sampling efforts, whereas unplanned domains correspond to domains of lower interest and relevance at the estimation stage of the survey.

The main area domains of interest in the NFI are the five production regions in Switzerland. The surface area of each of these regions is known, and independent estimates are produced for each of them. Stratified estimators are then used to combine these estimates to get an overall estimate for the entire country.

Most area domains of interest are of unknown size, but the membership of plot centres with area domains is known. The local density of target variables with respect to area domain is simply set to $y_D^{(X)}(\omega) = \iota_D(\omega)\, y_F^{(X)}(\omega)$, where $y_F^{(X)}(\omega)$ denotes the local density of target variable X with respect to the forest land F and $\iota_D(w)$ indicates plots with centre in area domain D.

The approach applied in the NFI involves associating the entire plot, with all its population elements (trees), with area domains through a point decision at the plot centre. This approach is robust, cost efficient in the field, and facilitates computations and algorithms at the estimation stage of the survey. In many national forest inventories, a more complex field protocol is used: sample plots are partitioned into subplots, and associations between area domains and population elements (trees) are observed and recorded for each subplot. The advantage of this approach is higher accuracy in domain estimation, at least in theory. In practice it comes at the price of increased complexity at the sampling and estimation stages of the inventory. Therefore, the system applied in the NFI has so far been left unchanged.

2.5.4 Sampling Error and Confidence Intervals

All NFI estimates in the standard output tables are accompanied by the corresponding *standard error* estimate, which is defined as the square root of the estimated variance of the estimate.

Non-sampling Errors Non-sampling errors, such as measurement and registration errors or frame imperfection effects, are not included in the reported standard error of the estimates. However, the magnitude of some of these errors is known from the approximately 5% of field plots independently re-measured by a second field team (Chap. 21). Moreover, the accuracy of some of the models used in NFI, such as single-stem volume functions, has been investigated (Chap. 12).

In this context, the two-stage sampling and estimation procedure of the NFI for stem volume (growing stock and biomass) is worth noting. In contrast to many other (national) forest inventories, the NFI subsample of tariff trees is used not only to fit tariff functions, but also to correct bias in tariff functions. The two-stage estimates are only at first glance less precise than the single-stage estimates, in which inference is based on the large sample of tally trees and associated stem volume predictions. Although the two-stage sampling inference is based on a much smaller subsample of tariff trees, the stem volumes are accurate. In this sense, the two-stage estimation for growing stock can be understood as a method that integrates the remaining uncertainties associated with single-stem volume models into the reported sampling error.

Confidence Intervals The standard output tables of the NFI contain the estimate $\hat{\theta}$ of the unknown population parameter θ together with a second value, the *absolute standard error* $\sqrt{\hat{\mathbb{V}}\langle\hat{\theta}\rangle}$ or the *relative standard error* $\frac{\sqrt{\hat{\mathbb{V}}\langle\hat{\theta}\rangle}}{\hat{\theta}} \times 100$ of the estimate.

The absolute standard error $\sqrt{\hat{\mathbb{V}}\langle\hat{\theta}\rangle}$ is an estimate of the unknown standard deviation, according to which the estimates $\hat{\theta}$ would be distributed under reiterations of the survey in the same (unchanged) population and with the same sampling design. It is not possible from a single survey to determine whether the estimate $\hat{\theta}$ is greater or less than the true value θ for the population parameter of interest, and whether $\hat{\theta}$ is close to θ. However, we can be confident at a $1 - \alpha$ level – again under repeated surveys – that the proportion of surveys for which the true value $\hat{\theta}$ is within the lower and upper bounds of the respective *1 − α confidence interval* $\left[\hat{\theta} - z_{1-\alpha/2}\sqrt{\widehat{V\langle\hat{\theta}\rangle}}, \hat{\theta} + z_{1-\alpha/2}\sqrt{\widehat{V\langle\hat{\theta}\rangle}}\right]$ is $1 - \alpha$, where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution, so that $z_{1-\alpha/2} = 2$ is frequently used when calculating the 95% confidence interval.

2.6 Estimation of Change

The permanent plot (annual panel) sampling design of the NFI facilitates the estimation of *net change* and *components of change* between inventories. However, there are some specific issues in the derivation of variables and in the interpretation of results that need special attention and are discussed in this section.

Net change is the general technical term for the difference between two population states. More precisely, it is the difference between the value of a population parameter of interest, for example the total growing stock in a forest, in two consecutive inventories.

Under the infinite population approach, net change in the total of a target variable over a sampling frame G is understood as

$$\Delta_T = T_2 - T_1 = \int_G y_2(\omega)d\omega - \int_G y_1(\omega)d\omega \tag{2.32}$$

$$= \int_G (y_2(\omega) - y_1(\omega))d\omega = \int_G \Delta_y(\omega)d\omega \tag{2.33}$$

Where $\Delta_y(\omega) = y_2(\omega) - y_1(\omega)$ denotes the difference at point ω between the local density of the target variable in the second inventory at t_2 and the local density of the target variable in the first inventory at t_1.

$\Delta_y(\omega)$ is only available if data collection is repeated at exactly the same sample points (*inventory with permanent plots*). However, Δ_T may alternatively be estimated with two independent samples, with data collected at t_1 and t_2, resulting in the estimates T_1 and T_2 (*inventories with temporary plots*).

2.6.1 Time Schedule Effects

In the NFI, as in many other national and large-area forest inventories, the temporal sequence of plot measurements within and between inventory cycles is complex. NFI field-data collection starts in March and ends in October to take into account the large proportion of mountain forest in Switzerland. In NFI1, NFI2 and NFI3 field data was collected over three successive years in a periodic sampling system. Data collection in NFI4 was done in a continuous (annual) sampling system spread over nine consecutive years (Fig. 2.4). The annual panel field data is again collected between March and October, and it is planned to continue using this approach in NFI5 and beyond.

Each annual panel with data collected in two successive inventory cycles can be used to estimate change between plot re-measurements (Fig. 2.6). The transition from periodic to annual sampling led to the number of years between inventories NFI3 and NFI4 being different for different annual panels. For instance, annual panel 5 could be used to estimate the average NFI1 state for the combined years 1983–1985, the NFI5 state for 2024, or intermediate NFI2, NFI3 and NFI4 states (Fig. 2.6). It could also be used to estimate the average change between NFI1 and NFI2 for the combined years 1983–1985 to 1993–1995, the average change between NFI3 and NFI4 for the combined years 2004–2006 to 2015, or the intermediate changes between NFI2 and NFI3, as well as between NFI4 and NFI5.

For the NFI4 output tables, between three and five annual panels were pooled for the intermediate reporting and nine for the final reporting. This is an intuitive

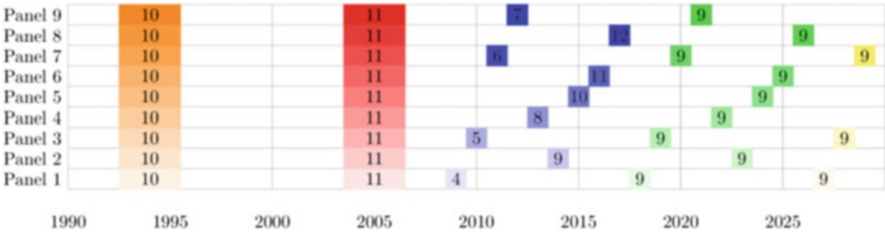

Fig. 2.6 Graphic representation of the inventory cycles for field-data collection with measurement of change in the NFI with the calendar years on the x-axis. The colours yellow, orange, red, blue and green refer to NFI1 to NFI5, respectively, and the different shades of each colour to the nine annual panels. The approximate number of years since the last inventory cycle is given for each cell

approach to produce estimates that include the full extent of the field data sets collected during the annual inventory panel.

The result tables provide estimates of *states in individual inventory cycles* and of *change between inventory cycles* without any clear-cut associations with points in time. This can be problematic for users, e.g. some international agencies, requiring estimates for particular calendar years. Estimates of change are, furthermore, typically expressed as average *annual changes* when analysing and comparing forest growth between regions and over time.

2.6.2 Dates and Calendar Years

In the NFI, the exact dates of field-data collection are known for each sampling unit. This information can be used to calculate the mean date of a given annual panel or of pooled annual panels. This mean date is, however, of little practical use and is therefore not computed or published as part of the NFI standard output routines.

Changes in target variables for a population rarely occur continuously and linearly over time, which is why results for a specific date are always biased, irrespective of the averaging method used. Moreover, the complete data set from a representative sample is needed to produce estimates.

Even estimates produced with data from a single annual panel refer to an average *state in that calendar year* and never to a specific date, and have, so far, not been published in the NFI. This situation extends to estimates produced with data collected over multiple years. The NFI solution is to label result tables with: (a) the inventory cycle for which the plot measurements were used, and (b) the calendar years during which the field data was collected.

2.6.3 Vegetation Periods

For any forest inventory, the *average annual growth, or increment, of trees* is important information. In Switzerland, the length and dates of the growing season for trees vary considerably from one calendar year to the next, but also depend on the vegetation zones and various other site and stand characteristics.

For a given sampling unit, field-data collection may take place before, during or after the in-situ growing season of tally trees in a given calendar year. The exact date of the plot measurements is known, but the time interval between re-measurements, expressed in calendar years (or months or days), does not necessarily represent the number of tree-growing seasons between field visits. This effect is amplified in the NFI, where field-data collection between inventory cycles is not synchronised to a constant time interval between plot re-measurements for all sampling units. A sampling unit measured in May in one inventory may be re-measured on a different date of the year in the next inventory.

To estimate the average annual growth of trees, the *number of vegetation periods* $\Delta_v(\omega)$ *between plot re-measurements* is needed. In the NFI, this value is approximated by assuming that: field data collected before June corresponds to the period in a given calendar year before the start of the tree-growing season; a quarter of the growing season will have happened already if field data is collected in June, half in July and three-quarters in August. The growing season of trees will have almost finished if field data is collected later than August.

Using an approximation for a target variable like growing stock seems reasonable when deriving the number of vegetation periods and transforming the observed total change between two inventories into an average annual change, or more precisely, into an average change per growing season. The situation is much less clear for other target variables, such as estimates of area domain sizes, for example the proportion of private forest, and quantities of deadwood. Values for these variables can change at any time during the calendar year. The average annual change in these variables may therefore be better defined as the number of calendar years between the field visits than as the number of vegetation periods. To maintain consistency among target variables, however, NFI annual changes are always annualised with the number of vegetation periods between plot re-measurements.

2.7 Net Change with Independent Samples

Here we briefly describe one method for estimating the net change between arbitrary calendar years, which is not currently in routine use in the NFI. The basis for the estimation is the continuous mode of field-data collection in which a representative sample is obtained each calendar year.

Estimates of states for any calendar year since 2009 and estimates of net changes between any two calendar years are, in principle, readily available. Using this method for estimating change with independent samples has two basic effects:

- The resulting state estimates are not very precise because the annual samples are small.
- The two samples used for net change estimation are independent of each other, with field data collected on different annual panels. As a result, (a) estimates of change components, such as gains and losses, are not available, and (b) the precision of the net change estimates is relatively low.

The second point follows from the general statistical principle that the variance of a difference between two independent random variables, here the state estimates $\hat{T}_1$ and $\hat{T}_2$ in the first and second inventory, is given by

$$\mathbb{V}\langle\hat{\Delta}_T\rangle = \mathbb{V}\langle\hat{T}_2 - \hat{T}_1\rangle = \mathbb{V}\langle\hat{T}_2\rangle + \mathbb{V}\langle\hat{T}_1\rangle - 2\mathbb{C}\langle\hat{T}_1, \hat{T}_2\rangle \tag{2.34}$$

where $\mathbb{C}\langle\hat{T}_1, \hat{T}_2\rangle$ is the co-*variance between the two state estimates*. $\mathbb{C}\langle\hat{T}_1, \hat{T}_2\rangle$ is zero in the case of independent samples (with temporary plots) and positive in the case of an annual panel survey (with permanent plots).

2.7.1 Pooled Annual Panels

Field data collected over two or more annual panels is frequently combined into a single pooled sample for estimation in national forest inventories. This has the advantage of increasing the precision of the estimates because of the size of the pooled sample is larger. The disadvantage is the increased fuzziness in the assignment of the results to a specific point in time, for example a calendar year. From a mathematical point of view, the optimal trade-off depends on the target variable, as well as other considerations such as:

- Estimates based on pooled annual panels cannot be expected to be reliable for specific calendar years if the population abruptly changes, for example following a massive windthrow of trees or extensive forest fires
- A pooled sample may provide reasonable estimates for specific calendar years if there is a linear or no change in the population
- In a long-term monitoring program such as a national forest inventory, the average (smoothed) states are likely to be considered sufficient for requirements such as deciding on national policy
- With pooled annual panels, emerging trends cannot be detected early.

We conclude that the NFI sampling system with nine annual panels offers various ways for detecting, monitoring and understanding states and trends in the population, although this may be occasionally at the price of extensive and complex data analysis. As already mentioned, the reports from the NFI are produced with the entire sample of the inventory cycle. Thus, nine annual panels are pooled together for the NFI4 results.

2.7.2 Average Annual Change

The number of years between states, and therefore the transformation of net change between two inventories into an average annual change, is obvious with independent samples. Instead of using calendar years, it may be reasonable to approximate the number of calendar years between the inventories by the *number of vegetation periods*, as explained in Sect. 2.6.3. With the calculated mean dates $\bar{a}_1$ and $\bar{a}_2$ of the respective state estimates and the mean number of vegetation periods $\Delta_v = \bar{a}_2 - \bar{a}_1$ between the two inventories, the *average annual change* is $\hat{\Delta}_T/\Delta_v$ with variance $\mathbb{V}\langle\hat{T}_2 - \hat{T}_1\rangle/\Delta_v^2$.

2.8 Estimation of Change Using Annual Panels

Estimating net change from a sample of permanent, re-measured sampling units is very efficient because the co-variance between the two state estimates in Eq. 2.34 is maximised. In practice, the co-variance and variance terms are not estimated separately because a representative sample of the difference $\Delta_y(\omega) = y_2(\omega) - y_1(\omega)$ between the local density of the target variable in the second inventory at t_2 and the local density of the target variable in the first inventory at t_1 is directly measured for sampling units ω in the annual panel.

Net change can also be estimated over three or more inventories. To estimate net change over three inventories, for instance, it is treated simply as the plot-level difference, which is the sum of two already derived changes $y_2(\omega) - y_1(\omega)$ between the inventories at t_2 and t_1 and of $y_3(\omega) - y_2(\omega)$ for the inventories at t_3 and t_2.

2.8.1 The Role of Area Domains

NFI estimates are always produced according to *reference domains* and *area domains*. A typical reference domain is the *accessible forest without shrub forest*. A set of area domains partitions the reference domain into subcategories of interest, such as forest ownership categories.

Reference and area domains play an intuitive and prominent role in the description of forest states and net changes. In the estimation of change, there are different options for the treatment of reference and area subdomains, depending on the user's information needs.

We introduce

$$\iota_t^{(e)}(\omega) = \begin{cases} 1 & \text{if } \omega \in e \\ 0 & \text{else} \end{cases} \tag{2.35}$$

to indicate whether, in inventory t, the sampling unit at point ω is inside or outside some area domain e.

In the NFI, the method usually used to derive the local density of the net change in a target variable between two inventories t_2 and t_1 is

$$\varDelta_y(\omega) = \iota_1^{(r)}(\omega)\,\iota_2^{(r)}(\omega)\,\iota_2^{(e)}(\omega)\,(y_2(\omega) - y_1(\omega)) \tag{2.36}$$

where r denotes the reference domain, such as the accessible forest without shrub forest, and e denotes some area domain category, such as private forest. The resulting estimates refer to the net change in the target variable in areas that are considered accessible forest without shrub forest in both inventories, according to the ownership categories observed in the second inventory.

Another method for deriving net change in sampling units is

$$\Delta_y(w) = \iota_2^{(r)}\,(\omega)\iota_2^{(e)}(\omega)y_2(\omega) - \iota_1^{(r)}(\omega)\iota_1^{(e)}(\omega)y_1(\omega) \tag{2.37}$$

This method is used to estimate the *overall net change according to reference and area domains*. $\Delta_y(w)$ is zero for sampling units that lie outside the reference domain in both inventories. The change estimates include all target variable gains and losses due to increases and decreases in the geographic extent of the reference and area domains. In the NFI, this method is optionally applied in net change estimation.

Both of the approaches described above, and their resulting estimates, make intuitive sense and can be correctly understood by the user if clearly specified in output tables.

Components of Change in Area Domain There are no technical difficulties in estimating change in target variables due to gains and losses in the reference or area domain between inventories. For instance

$$\Delta_y(w) = \iota_2^{(r)}\,(\omega)\iota_1^{(r)}(\omega)\iota_2^{(e)}(\omega)\Big(1 - \iota_1^{(e)}(\omega)\Big)y_2(\omega) \tag{2.38}$$

generates local densities used for estimating the overall net change in the target variable with respect to area domain e that is due to an increase in the area extent of e between inventories. Because result tables tend to become quite complex and large, these components of change at the area domain level are normally not calculated and reported in the NFI standard output routines.

Notes on Estimating Area Domain Changes First, it is conceptually important to distinguish between target variable change caused by *area domain (plot-level) change* between inventories, described in this section, and target variable change caused by *population element (tree-level) change* between inventories, covered in Sect. 2.9.

Reference and area domains of the NFI are always defined for the point ω, the plot centre, which is why target variable gains and losses according to reference and area domains always include all population elements associated with the sampling unit at ω. In contrast, the field protocol of a forest inventory may allow the partitioning of sampling units – usually sectors or segments of circular plots – into different area domains, such as land-use classes or forest ownership categories.

Both approaches are common. The advantages of the NFI approach are that the field protocol is relatively simple and field-data collection is restricted to sampling units with their centres in forests. The subplot or plot segments approach, on the other hand, requires field measurements on sampling units at the forest boundary with their centre outside forests. An elaborate protocol is therefore required to register area-domain proportions and the association of population elements (trees) with area domains. The downside of the NFI method is that the association of population elements (trees) with area domains is relatively coarse. However, this is only the case for sampling units partitioned into different area domains. In a national forest inventory, differences in the effects of using the two approaches for obtaining overall estimates can be assumed to be negligible.

2.8.2 Average Change Per Unit Area

For comparison reasons, the net change between inventories is often expressed as an average change per unit area, i.e. an average density of the target variable per hectare of forest.

While the mean spatial density parameter is easily computed and clearly defined when estimating states, this is not necessarily the case when estimating change. Because the reference domain and the area domains of interest may change between inventories, *there is no unambiguous definition and understanding of the average change per unit area between inventories*.

Such estimates are produced in the NFI, but only with local densities defined according to Eq. 2.36. In this case, net change is estimated for the land that is part of the reference domain in both inventories according to the area-domain categories used in the second inventory. The disadvantage of this approach is that changes in the target variable due to area gains and losses of the reference domain is not included in the change estimates.

2.8.3 Average Annual Change

In this section, we describe two approaches for calculating the *average annual change*. The first method was applied in the NFI under the periodic inventory system (NFI1 to NFI3) and is known from the forest inventory literature. The second method possibly copes better with the range of time intervals between plot re-measurements in NFI4 became wider than in NFI3 with the transition from the periodic to the continuous (annual) inventory system.

Traditional Estimator The traditional estimator for the average annual change requires a simple modification of the local density function

$$\Delta_y^{(MOR)}(\omega) = \frac{\Delta_y(\omega)}{\Delta_v(\omega)} \tag{2.39}$$

where $\Delta_y(\omega)$ is the local density of the change in the target variable between the two inventories at point ω, and $\Delta_v(\omega)$ is the number of vegetation periods between the two inventories at point ω.

A conspicuous property of this local density function $\Delta_y^{(MOR)}(\omega)$ is that the implicit mean number of vegetation periods between the two inventories $\Delta_v^* = \hat{\Delta}_T/\hat{\Delta}_{T/a}*$ varies depending on the target variable for which the change components are estimated. $\hat{\Delta}_T$ denotes the estimate for the *total change between inventories*, for example the mean of plot values under simple random point sampling, and $\hat{\Delta}^*_{T/a}$ denotes the estimate of the *average annual change between inventories* with the local density function $\Delta_y^{(MOR)}(\omega)$ and under the same sampling design.

Transition Estimator The aims of the estimator of average annual change are: (a) to use the same mean number of vegetation periods between inventories for all target variables, and (b) to use an average number of vegetation periods between inventories, which has an intuitive interpretation.

With a set of H post-sampling strata, the *mean number of vegetation periods* for plots located in an area domain e is calculated as

$$\Delta_v = \frac{\sum_{h=1}^{H} \lambda_h n_h^{-1} \sum_{j=1}^{n_h} \iota_e(\omega) \Delta_v(\omega)}{\sum_{h=1}^{H} \lambda_h n_h^{-1} \sum_{j=1}^{n_h} \iota_e(\omega)} \tag{2.40}$$

where λ_h is the surface area and n_h the number of sampling units in post-sampling stratum h, and $\iota_e(\omega)$ is the area domain indicator variable. The resulting estimator of the average annual change is $\hat{\Delta}_{T/a} = \hat{\Delta}_T/\Delta_v$. Because Δ_v is a sampling and measurement design constant, the variance of $\hat{\Delta}_{T/a}$ is $V\langle\hat{\Delta}_{T/a}\rangle = V\langle\hat{\Delta}_T\rangle/\Delta_v^2$.

Discussion The traditional and the new estimator are identical if the number of vegetation periods is the same for all plots ($\Delta_v(\omega) = \Delta_v$), as it is (approximately) between NFI1 and NFI2, NFI2 and NFI3, and NFI4 and NFI5. Between NFI3 and NFI4, the new estimator has the advantage that the number of vegetation periods is always the same for a given area domain of interest and intuitively interpreted as *the total change in the target variable measured over all plots in the respective area domain of interest, divided by the total number of vegetation periods over which this change has been measured on these plots*.

2.9 Change in the Population of Trees

In Sect. 2.8, we discussed the estimation of change and change components at the plot level. However, forest inventories, especially those with permanent plots, also require estimations of changes in the population of trees. In this section we describe methods for estimating these changes.

2.9.1 Definitions and Notation

The definition, analysis and estimation of specific change components between the state of a population of trees at t_1 and t_2, with $t_1 < t_2$, requires a tree-level analysis of change effects.

Population The population is defined as consisting of living and dead trees, and the notation is a bit tricky. A population survivor tree is a tree that is eligible in both inventories. It is not necessarily a living tree; it may be dead at t_2, or at t_1 and t_2.

We use $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2$ to denote the *population of eligible trees* at t_1 and t_2, and define:

- The subpopulations of living trees at time points t_1 and t_2 *as* $\mathcal{P}_{l.1}, \mathcal{P}_{l.2}$, and of dead trees as $\mathcal{P}_{d.1}, \mathcal{P}_{d.2}$, with $\mathcal{P}_1 = \mathcal{P}_{l.1} \cup \mathcal{P}_{d.1}$ and $\mathcal{P}_2 = \mathcal{P}_{l.2} \cup \mathcal{P}_{d.2}$
- $\mathcal{S}_{12} = \mathcal{P}_1 \cap \mathcal{P}_2$, the *subpopulation of survivor trees* between t_1 and t_2 with
  - $\mathcal{S}_{l.12} = \mathcal{P}_{l.1} \cap \mathcal{P}_{l.2}$, the subpopulation of trees living at both time points,
  - $\mathcal{S}_{d.12} = \mathcal{P}_{d.1} \cap \mathcal{P}_{d.2}$, the subpopulation of dead trees at both time points,
  - $\mathcal{M}_{12} = \mathcal{P}_{l.1} \cap \mathcal{P}_{d.2}$, the subpopulation of *mortality trees*, i.e. living at t_1 and dead at t_2, with $\mathcal{S}_{12} = \mathcal{S}_{l.12} \cup \mathcal{S}_{d.12} \cup \mathcal{M}_{12}$
- $\mathcal{D}_{12} = \mathcal{P}_1 \backslash \mathcal{P}_2$, the subpopulation of cut and mortality trees between t_1 and t_2, with
  - $\mathcal{D}_{12}^{(A)}$, the subpopulation of trees lost between t_1 and t_2 because they are located inside the domain of interest at t_1 and outside at t_2, e.g. because of a shift in the domain boundary,

– $\mathcal{D}_{12}^{(L)}$, the subpopulation of trees lost due to (natural) in-situ destruction (decomposition) between t_1 and t_2,
– $\mathcal{D}_{12}^{(U)}$, the subpopulation of trees lost because of removal (and usage) from the population between t_1 and t_2,
– $\mathcal{D}_{l.12} = \mathcal{P}_{l.1} \backslash \mathcal{P}_{l.2} = \mathcal{D}_{l.12}^{(A)} \cup \mathcal{D}_{l.12}^{(L)} \cup \mathcal{D}_{l.12}^{(U)} \cup \mathcal{M}_{12}$, the subpopulation of trees lost from the population $\mathcal{P}_{l.1}$ of living trees,
– $\mathcal{D}_{d.12} = \mathcal{P}_{d.1} \backslash \mathcal{P}_{d.2} = \mathcal{D}_{d.12}^{(A)} \cup \mathcal{D}_{d.12}^{(L)} \cup \mathcal{D}_{d.12}^{(U)}$, the subpopulation of trees lost from the population $\mathcal{P}_{d.1}$ of dead trees, with $\mathcal{D}_{12} = (\mathcal{D}_{1.12} \cup \mathcal{D}_{d.12}) \backslash \mathcal{M}_{12} = \mathcal{D}_{12}^{(A)} \cup \mathcal{D}_{12}^{(L)} \cup \mathcal{D}_{12}^{(U)}$

- $\mathcal{E}_{12} = \mathcal{P}_2 \backslash \mathcal{P}_1$, the *subpopulation of ingrowth trees* between t_1 and t_2, with

– $\mathcal{E}_{12}^{(A)}$, the subpopulation of trees located outside the domain of interest at t_1 and inside at t_2,
– $\mathcal{E}_{12}^{(I)}$, the subpopulation of trees reaching population eligibility, such as the minimum dbh threshold, between t_1 and t_2,
– $\mathcal{E}_{l.12} = \mathcal{P}_{l.2} \backslash \mathcal{P}_{l.1} = \mathcal{E}_{l.12}^{(A)} \cup \mathcal{E}_{l.12}^{(I)}$, the subpopulation of newly eligible trees in population $\mathcal{P}_{l.2}$ of living trees
– $\mathcal{E}_{d.12} = \mathcal{P}_{d.2} \backslash \mathcal{P}_{d.1} = \mathcal{E}_{d.12}^{(A)} \cup \mathcal{E}_{d.12}^{(I)} \cup \mathcal{M}_{ld.12}$, the subpopulation of newly eligible trees in population $\mathcal{P}_{d.2}$ of dead trees, with $\mathcal{E}_{12} = \mathcal{E}_{l.12} \cup \mathcal{E}_{d.12} = \mathcal{E}_{12}^{(A)} \cup \mathcal{E}_{12}^{(i)}$.

We did not specify the length of time between t_1 and t_2 and related effects of trees becoming intermediate members of the population for some periods, such as from t_e to t_d, with $t_1 < t_e < t_d < t_2$ between t_1 and t_2. The population is periodically observed with permanently installed sample plots, which means such intermediate changes in population are, by definition, not detected and not included in the resulting change estimates.

Sample of Permanent Plots In a forest inventory with permanent plots, the tally trees have to be classified during field-data collection according to the definitions of population survivor, population cut and mortality, and population ingrowth trees. The first-inventory tally trees must be re-identified in the second inventory.

The classification of the tally trees into subsamples is similar (in terminology) to the classification of population trees into subpopulations, with some artefact trees (Eriksson 1995a). We use a different notation for trees in the sample of permanent plots. For a given sampling unit at ω or the entire sample of permanent plots, the samples $P_1(\omega)$ and $P_2(\omega)$ of tally trees at t_1 and t_2 can be partitioned into the following subsamples:

- $S_{12}(\omega)$ sample survivor trees registered at t_1 and t_2,
- $C_{12}(\omega)$ sample cut trees removed from the population at t_d, with $t_1 < t_d < t_2$,

- $M_{12}(\omega)$ sample mortality trees removed from the population of living trees, and transferred to and remaining in the population of dead trees at t_d, with $t_1 < t_d < t_2$.
- $I_{12}(\omega)$ sample ingrowth trees, not fulfilling population eligibility at t_1 (population ingrowth tree) and immediate association with the sampling unit when population eligibility is reached at $t_1 < t_e < t_2$,
- $O_{12}(\omega)$ sample ongrowth trees, not fulfilling population eligibility at t_1 (population ingrowth tree), and either (a) reaching first population eligibility and later sampling unit association at $t_1 < t_e < t_2$ ($O_{12.\ a}(\omega)$), or (b) immediate association with the sampling unit at $t_1 < t_e < t_2$, together with population eligibility ($O_{12.\ b}(\omega)$),
- $N_{12}(\omega)$ sample nongrowth trees, fulfilling population eligibility at time t_1 (population survivor tree), but reaching sampling unit association later at time $t_1 < t_e < t_2$.

Note that sample ongrowth trees exist under angle-count sampling but not under the NFI plot configuration with two nested, concentric circles of fixed size.

2.9.2 Change Components

The conceptual difference between subpopulations of trees relevant for describing and defining change and change components in the population is subtle, and the related subsamples of trees are observed and measured on permanent plots.

Definitions The forest inventory in Switzerland has a long tradition and unique understanding of change assessment and relevant change components. The pioneering work, both in theory and in practice, is the *méthode de contrôle* (Biolley 1901).

Biolley introduced periodic, full-census tree measurements in forests in the Jura valleys of Canton Neuchâtel about 130 years ago. He divided the forest into compartments (*divisions*), each with a surface area of approximately 10–20 ha. Measurements were taken with callipers and have continued to be taken every 5–15 years ever since. The diameters of any removed trees are assessed at the time of felling and a local tariff function applied for estimating stem volumes. The inventory is restricted to trees with $d_{1.3} \geq 17.5$ cm.

The analysis of change, especially growth, starts from the simple fact that the state in the second inventory is the result of the gains and losses that have occurred since the first inventory, where gains refer to either the growth of trees already measured in the first inventory and to the ingrowth of new trees, while losses arise from tree felling, removal and mortality.

In our symbolic notation, Y represents the total of some target variable or the per hectare mean density, for example the stem volume of trees, in the compartment. The subscript denotes the respective change component with a notation referring to the subpopulations of trees introduced in Sect. 2.8.

Biolley's change components of interest are

$$\begin{array}{ll}
\text{net change } (\textit{augmentation/diminuition}) & \hat{=} Y_{\mathcal{P}_{l.2}} - Y_{\mathcal{P}_{l.1}} \\
\text{depletion } (\textit{matériel exploité}) & \hat{=} Y_{\mathcal{D}_{l.12}} \\
\text{eligibility } (\textit{passage à la futaie}) & \hat{=} Y_{\mathcal{E}_{12}} \\
\text{total growth } (\textit{accroissement total}) & \hat{=} Y_{\mathcal{P}_{l.2}} - Y_{\mathcal{P}_{l.1}} + Y_{\mathcal{D}_{l.12}}
\end{array}$$

as well as gross increment without eligibility (*accroissement du matériel initial*), defined as $Y_{\mathcal{P}_{l.2}} - Y_{\mathcal{P}_{l.1}} + Y_{\mathcal{D}_{l.12}} - Y_{\mathcal{E}_{l.12}}$.

Depletion is defined at the moment of loss and *eligibility* at the moment of gain. In other words, *depletion* is the amount of wood at the moment t_d of loss, and *eligibility* is the amount of wood at the moment t_e of trees reaching *eligibility*, which is the $d_{1.3}$ threshold of 17.5 cm in the case of Biolley's inventory system. The English terms have been taken from Eriksson (1995a) and Mandallaz (2008). These authors use the same definitions for the components of change as Biolley.

Beers and Schmid-Haas Components of Change In an inventory with permanent plots, the change components are derived for each sampling unit, based on the classification of tally trees into 'survivor', 'cut', 'mortality', 'ingrowth', 'ongrowth' and 'nongrowth' trees. The original definition and estimation of change components with permanent plot inventories seems to come from Beers (1962). Schmid-Hass adopted and developed the system of conducting forest inventories with permanent plots (*Kontrollstichprobe*) for Switzerland more than 50 years ago (Schmid-Haas 1983). The system is in wide use for inventories at the forest enterprise (regional) level and uses a single fixed-area circle for tree association. The relevant components of change are, derived at plot level,

$$\begin{array}{ll}
\text{net change} & \hat{=} y_{P_{l.2}}(\omega) - y_{P_{l.1}}(\omega) = y_{l.2} - y_{l.1} \\
\text{cut and mortality} & \hat{=} y_{C_{l.12}}(\omega) + y_{M_{12}}(\omega) = c_{l.1} + m_1 \\
\text{ingrowth} & \hat{=} y_{I_{l.12}}(\omega) = i_{l.2} \\
\text{gross increment} & \hat{=} y_{S_{l.2}}(\omega) - y_{S_{l.1}}(\omega) + y_{I_{l.12}}(\omega) = s_{l.2} - s_{l.1} + i_{l.2}
\end{array}$$

The respective subset of tally trees is given in the subscript and follows the notation given in Sect. 2.9.1.

There are two specific problems with periodic re-measurement of tally trees in permanent plot sampling, for which various solutions have been proposed (Grosenbaugh 1958; Martin 1982; Van Deusen et al. 1986; Roesch et al. 1989, 1991; Eriksson 1995a). First, the definitions for *cut and mortality*, *ingrowth* and *gross increment* are not exactly identical to the change components introduced by Biolley, i.e. *depletion*, *eligibility* and *total growth*. The reason is that neither the exact time points t_d and t_e of depletion and eligibility, nor the value of the target variables at these time points, are known under periodic sampling. Therefore, *cut* *and mortality* underestimates *depletion* and *gross increment* underestimates *total growth* by the (unknown) amount of growth in the subpopulation of cut and mortality trees between t_1 and t_d.

A second problem arises under sampling designs in which the extrapolation factors of sample trees change over time due to an increment in $d_{1.3}$. This is the case for angle-count sampling, as well as the nested (concentric) fixed-area plot design of the NFI.

Eriksson's Components of Change Eriksson's components of change are using the same abbreviated notation as before:

net increase (net change)	$\hat{=} y_{l.2} - y_{l.1}$
depletion (cut and mortality)	$\hat{=} c_{l.d} + m_d$
growth of cut and mortality	$\hat{=} c_{l.d} + m_d - c_{l.1} - m_1$
eligibility (ingrowth)	$\hat{=} i_{l.e}$
growth of ingrowth	$\hat{=} i_{l.2} + o_{b.l.2} - i_{l.e} - o_{b.l.e}$
survivor growth	$\hat{=} s_{l.2} + n_{l.2} + o_{a.l.2} - s_{l.1}$

where *total growth* is equal to the sum of *survivor growth*, *growth of cut and mortality* and *growth of ingrowth* (Eriksson 1995a). The target variable values $c_{l.d}$ and m_d at the time of cut and mortality are not observed and are assumed to be predicted with a growth model.

Eriksson's method provides, as in the methods of Biolley, Beers and Schmid-Hass, additive components in the sense that, at sample plot or domain level, *net increase* is numerically equal to the sum of *total growth* and *eligibility*, minus *depletion*.

This method is appealing because no back-prediction of target variables is needed for *nongrowth trees* (see below) and that additivity is guaranteed over more than two successive measurements.

NFI Components of Change A disadvantage of Eriksson's approach is that there is a relatively large jump in the local density whenever a nongrowth tree appears in the sample. This has a negative effect on the precision of the overall estimates of change components, i.e. of total growth. In the NFI approach, this effect is mitigated by backdating the growth of nongrowth trees as follows

net change	$\hat{=} y_{l.2} - y_{l.1}$
cut and mortality	$\hat{=} c_{l.d} + m_d$
growth of cut and mortality	$\hat{=} c_{l.d} + m_d - c'_{l.1} - m'_1$
ingrowth	$\hat{=} i_{l.2}$
survivor growth	$\hat{=} s_{l.2} + n_{l.2} - s'_{l.1} - n'_{l.1}$

In addition, *gross increment* is defined as the sum of survivor growth, growth of cut and mortality, and ingrowth.

The target variable values $c_{l.d}$ and m_d at time t_d of loss for cut and mortality trees, and the target variable value $n'_{l.1}$ for nongrowth trees at time t_1 are not known. Appropriate models for tree-level predictions of basal area, and the related volume and biomass, have been developed in the NFI (Chap. 12).

The prime symbol in $c'_{l.1}$ and m'_1, as well as $s'_{l.1}$ and $n'_{l.1}$ indicates that these values are calculated with values taken from t_1 and extrapolation factors taken from t_d and t_2, respectively.

The change components *cut and mortality*, *ingrowth*, *survivor growth* and *gross increment* can be considered unbiased, but they are not numerically exactly additive with *net change*. In other words,

$$y_{l.2} - y_{l.1} \neq s_{l.2} + n_{l.2} - s'_{l.1} - n'_{l.1} + i_{l.2} - c_{l.d} - m_d. \tag{2.41}$$

In practice, the non-additivity of these components of change is not relevant.

A disadvantage and challenge of this approach is that models for the back-prediction of nongrowth trees are needed.

Subpopulations of Living and Dead Trees Change components are traditionally defined only for the subpopulation of living trees. In Fig. 2.7, we graphically show an immediate extension to the subpopulation of dead trees, at least to the subpopulation of standing dead trees.

Lying dead trees are defined and measured in the NFI. At the moment of status change, however, trees may change position, leaving or newly appearing in the sample. The assessment of change components in the subpopulation of dead trees is, therefore, less precise than in the subpopulation of living trees. An extension of the NFI4 field protocol introduces all relevant parameters for, in principle, unbiased estimation of change components in the subpopulation of living and dead trees.

2.9.3 Derivation of Change Components for Tariff Trees

Under two-stage sampling, the general form of local densities for the estimation of change and change components is (compare with Eqs. 2.11 and 2.12)

$$y_2 - y_1 = \left(\hat{y}_2 - \hat{y}_1\right) + \left(r_2 - r_1\right) \tag{2.42}$$

where the notation has been simplified slightly. $\hat{y}$ denotes a local density based on tariff function stem volumes, and r a local density based on the residual between volume and tariff function stem volumes, available for the subset of second-stage tariff trees only. The subscripts indicate the two successive inventories.

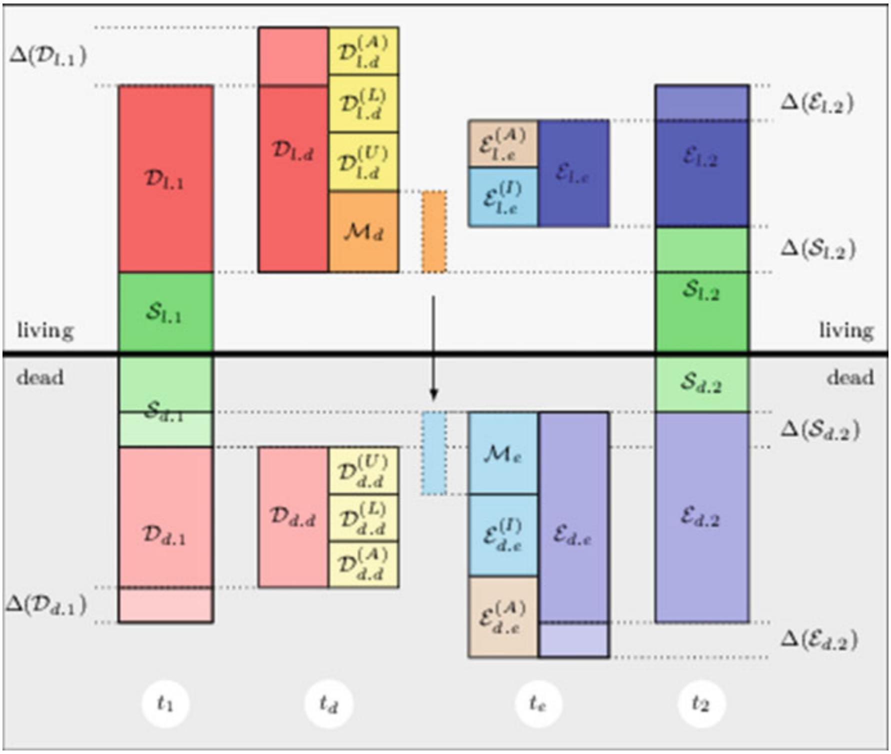

Fig. 2.7 Graphic representation of the subpopulations of trees relevant for estimating change between two inventories at t_1 and t_2. The subpopulations are defined in Sect. 2.9.1. Δ denotes the increment (or decrement) within subpopulations for certain variables, such as basal area, stem volume and biomass. The block in the middle, which changes the status from living (orange) to dead (cyan), refers to mortality trees, which remain in the overall population and are losses in the subpopulation of living trees and gains in the subpopulation of dead trees

Change at Tally Tree and Tariff Tree Level The full expression of change between two inventories under two-stage sampling is

$$y_2 - y_1 = \left(\sum_{i \in \dot{S}_{1\cap 2}} \left(\frac{\hat{X}_{2.i}}{\pi_{2.i}} - \frac{\hat{X}_{1.i}}{\pi_{1.i}} \right) \right) + \left(\sum_{i \in \dot{S}_{2\setminus 1}} \frac{\hat{X}_{2.i}}{\pi_{2.i}} \right) - \left(\sum_{i \in \dot{S}_{1\setminus 2}} \frac{\hat{X}_{1.i}}{\pi_{1.i}} \right)$$
$$+ \left(\sum_{i \in \ddot{S}_{1\cap 2}} \left(\frac{R_{2.i}}{\pi_{2.i} p_{2.i}} - \frac{R_{1.i}}{\pi_{1.i} p_{1.i}} \right) \right) + \left(\sum_{i \in \ddot{S}_{2\setminus 1}} \frac{R_{2.i}}{\pi_{2.i} p_{2.i}} \right) - \left(\sum_{i \in \ddot{S}_{1\setminus 2}} \frac{R_{1.i}}{\pi_{1.i} p_{1.i}} \right). \tag{2.43}$$

The first three sums represent changes in the first-stage sample of tally trees, with $\dot{S}_{1\cap 2} = \dot{S}_1 \cap \dot{S}_2$ denoting the set of first-phase tally trees remaining in the sample (and in the population) between the two inventories, and $\dot{S}_{2\backslash 1}$ and $\dot{S}_{1\backslash 2}$ denoting new and lost members of the first-stage sample of tally trees. The same basic subsets of trees remaining in the sample, becoming new members of the sample, or disappearing from the sample also occur in the second-stage sample of tariff trees. These three subsets are denoted by $\ddot{S}_{1\cap 2}$, $\ddot{S}_{2\backslash 1}$, and $\ddot{S}_{1\backslash 2}$, respectively, and it should be kept in mind that $\ddot{S}_1 \subseteq \dot{S}_1$ and $\ddot{S}_2 \subseteq \dot{S}_2$ are always fulfilled by definition.

The estimation of change components is a bit more complex than immediately revealed by Eq. (2.43). *First-stage tally trees* include:

- Tally trees in the first sum of Eq. (2.43), which are trees that have survived in the sample (and in the population);
- Tally trees in the second sum of Eq. (2.43), which fall into the categories ingrowth, ongrowth and nongrowth trees. Various options exist regarding how these trees should be treated in change estimation;
- Tally trees in the third sum of Eq. (2.43), which are losses from the sample. Basically, these are trees that have been removed from the population through cut and mortality.

Likewise, second-stage tariff trees include:

- Tariff trees in the fourth sum of Eq. (2.43), which are trees that have survived in the sample (and in the population);
- Tariff trees in the fifth sum of Eq. (2.43), which are trees in the second inventory that may be population survivor or population ingrowth trees;
- Tariff trees in the sixth sum of Eq. (2.43), which have become lost since the first inventory and may be population survivor or population cut or mortality trees.

Survivor Growth The growth of population survivor trees is of major interest because most of the overall population growth is in these trees.

By specifying the different subsets of tally and tariff trees in the sample, the *difference in the residual part for population survivor trees* turns out to be

$$(r_2 - r_1)_{|\text{survivors in population}} = \sum_{\substack{i \in \dot{S}_{1\cap 2} \\ i \in \ddot{S}_{1\cap 2}}} \left(\frac{R_{2.i}}{\pi_{2.i} p_{2.i}} - \frac{R_{1.i}}{\pi_{1.i} p_{1.i}} \right)$$
$$+ \sum_{\substack{i \in \dot{S}_{1\cap 2} \\ i \in \ddot{S}_{2\backslash 1}}} \frac{R_{2.i}}{\pi_{2.i} p_{2.i}} - \sum_{\substack{i \in \dot{S}_{1\cap 2} \\ i \in \ddot{S}_{1\backslash 2}}} \frac{R_{1.i}}{\pi_{1.i} p_{1.i}} + \sum_{\substack{i \in \dot{S}_{2\backslash 1} \\ i \in \ddot{S}_{2\backslash 1}}} I_{P_{1\cap 2}.i} \frac{R_{2.i}}{\pi_{2.i} p_{2.i}} \qquad (2.44)$$

where the two subscripts under the sigma sign indicate the respective subsamples of tally and tariff trees, and indicates nongrowth tally trees (population survivor trees).

The fourth sum of Eq. (2.44) refers to residuals after selecting nongrowth tally trees as tariff trees.

The components in the second and third sums in Eq. (2.44) only occur when the subselection of tariff trees is not permanent, i.e. when a tariff tree chosen (by chance) in the first inventory is not automatically re-measured as a tariff tree in the second inventory. In the NFI, this is the case for tally trees with $d_{1.3}$ <60 cm. The selection of second-stage tariff trees in inventories NFI2, NFI3 and NFI4 (years 2009–2014) was carried out independently from the selection in the previous inventory (Sect. 2.3).

These components lead to an increased variance in the change estimates because even a considerable volume residual for an unusually formed stem can remain quite stable between inventories. It is thus of negligible consequence for inference, but it can carry considerable weight if observed in only one of the two inventories.

If tariff trees are selected independently on occasions one and two, another estimator may be used. We know that $P\langle i \in \ddot{S}_{1\cap 2}\rangle = p_{1.i}p_{2.i}$ and the residual component for population survivor trees can be estimated separately for the subset of sample survivor and nongrowth trees as

$$(r_2 - r_1)_{|\text{survivors in population}} = \sum_{\substack{i \in \dot{S}_{1\cap 2} \\ i \in \ddot{S}_{1\cap 2}}} \frac{R_{2.i} - R_{1.i}}{\pi_{2.i}p_{1.i}p_{2.i}} + \sum_{\substack{i \in \dot{S}_{2\backslash 1} \\ i \in \ddot{S}_{2\backslash 1}}} I_{P_{1\cap 2}.i} \frac{R_{2.i} - \tilde{R}_{1.i}}{\pi_{2.i}p_{2.i}} \tag{2.45}$$

Only re-measured tariff trees are used in the first sum, whereas back-predicted, tree-level residuals $\tilde{R}_{1.i}$ are assumed for nongrowth tariff trees in the second sum. In practice, this estimator has a larger variance than the estimator given in Eq. (2.44) because the subset of tariff tress selected in both inventories is small.

The NFI has a model for $\tilde{R}_{1.i}$ back-prediction, and the residual component for population survivor trees is then calculated according to

$$\begin{aligned}(r_2 - r_1)_{|\text{survivors in population}} = &\sum_{\substack{i \in \dot{S}_{1\cap 2} \\ i \in \ddot{S}_{1\cap 2}}} \frac{R_{2.i} - R_{1.i}}{\pi_{2.i}p_{2.i}} \\ &+ \sum_{\substack{i \in \dot{S}_{1\cap 2} \\ i \in \ddot{S}_{2\backslash 1}}} \frac{R_{2.i} - \tilde{R}_{1.i}}{\pi_{2.i}p_{2.i}} + \sum_{\substack{i \in \dot{S}_{2\backslash 1} \\ i \in \ddot{S}_{2\backslash 1}}} I_{P_{1\cap 2}.i} \frac{R_{2.i} - \tilde{R}_{1.i}}{\pi_{2.i}p_{2.i}}.\end{aligned} \tag{2.46}$$

The first two sums provide together a prediction for the residual component in sample survivor trees, and the third sum provides the same prediction for nongrowth trees. Added together, these three sums provide a prediction for the change in the residual component in population survivor trees. This approximate estimator has the tendency to slightly underestimate the true change in the residual component because of the modelled (smoothed) back-predictions $\tilde{R}_{1,i}$. The estimator is, therefore, robust with a lower variance than the estimators in Eqs. 2.44 and 2.45.

Overall, for two-stage change estimation and tariff-tree selection, we conclude that:

- The residual correction in the estimation of change (growth) is, in principle, needed to ensure the state estimates with the estimates of change are unbiased (and numerically additive);
- The precision of the estimation change in the residual component depends on the method of second-stage tariff tree selection;
- The efficiency of the design-based estimators increases with a larger proportion of re-measured tariff trees;
- For small domains and small sample sizes, back-predicting residuals as, well as omitting the estimation of change in the residuals and restricting change estimation to first-stage trees, may be an option.

Appendix

Table 2.1 Tariff tree inclusion probabilities

d	p_old	p_new	Sector small	Sector small rounded	Sector large	Sector large rounded
12	0.04050	0.04050	16.200	16		
13	0.04753	0.04753	19.013	19		
14	0.05513	0.05513	22.050	22		
15	0.06328	0.06328	25.313	25		
16	0.07200	0.07200	28.800	29		
17	0.08128	0.08128	32.513	33		
18	0.09113	0.09113	36.450	36		
19	0.10153	0.10153	40.613	41		
20	0.11250	0.11250	45.000	45		
21	0.12403	0.12403	49.613	50		
22	0.13613	0.13613	54.450	54		
23	0.14878	0.14878	59.513	60		
24	0.16200	0.16200	64.800	65		
25	0.17578	0.17578	70.313	70		
26	0.19013	0.19013	76.050	76		
27	0.20503	0.20503	82.013	82		
28	0.22050	0.22050	88.200	88		
29	0.23653	0.23653	94.613	95		
30	0.25313	0.25313	101.250	101		
31	0.27028	0.27028	108.113	108		

(continued)

Table 2.1 (continued)

d	p_old	p_new	Sector small	Sector small rounded	Sector large	Sector large rounded
32	0.28800	0.28800	115.200	115		
33	0.30628	0.30628	122.513	123		
34	0.32513	0.32513	130.050	130		
35	0.34453	0.34453	137.813	138		
36	0.14580	0.15428	137.813	138	10.980	11
37	0.15401	0.17098	137.813	138	22.110	22
38	0.16245	0.18896	137.813	138	34.099	34
39	0.17111	0.20829	137.813	138	46.987	47
40	0.18000	0.22904	137.813	138	60.817	61
41	0.18911	0.25126	137.813	138	75.631	76
42	0.19845	0.27502	137.813	138	91.473	91
43	0.20801	0.30039	137.813	138	108.387	108
44	0.21780	0.32744	137.813	138	126.417	126
45	0.22781	0.35623			142.492	142
46	0.23805	0.38683			154.734	155
47	0.24851	0.41932			167.730	168
48	0.25920	0.45377			181.509	182
49	0.27011	0.49025			196.100	196
50	0.28125	0.52883			211.534	212
51	0.29261	0.56960			227.840	228
52	0.30420	0.61263			245.050	245
53	0.31601	0.65799			263.195	263
54	0.32805	0.70576			282.305	282
55	0.34031	0.75604			302.414	302
56	0.35280	0.80889			323.555	324
57	0.36551	0.86440			345.759	346
58	0.37500	0.92265			369.060	369
59	0.37500	0.98373			393.493	393
60	1.00000	1.00000			400.000	400

New (p new) and old (p old) tariff tree inclusion probabilities according to diameter (d) classes with respective exact sector openings on the small (sector small) and on the large (sector large) plot, both also with a rounded to integer version

References

Beers TW (1962) Components of forest growth. J For 60(4):245–248

Biolley B (1901) Le jardinage cultural. Schweiz Z Forstwes 52(6):97–104

Biolley H (1920) L'Aménagement des forêts par la méthode expérimentale et spécialement la méthode du contrôle. Attinger Frères, Neuchâtel

Cordy CB (1993) An extension of the Horvitz—Thompson theorem to point sampling from a continuous universe. Stat Probab Lett 18(5):353–362

Düggelin C, Keller M (eds) (2017) Schweizerisches Landesforstinventar – Feldaufnahme Anleitung 2017. Eidgenössische Forschungsanstalt für Wald Schnee und Landschaft, Birmensdorf

Eriksson M (1995a) Compatible and time-additive change component estimators for horizontal-point-sampled data. For Sci 41(4):796–822

Eriksson M (1995b) Design-based approaches to horizontal-point-sampling. For Sci 41(4):890–907

Gregoire T (1982) The unbiasedness of the mirage correction procedure for boundary overlap. For Sci 28:504–508

Gregoire T, Valentine H (2007) Sampling strategies for natural resources and the environment, Applied Environmental Statistics. Chapman & Hall, Boca Raton

Grosenbaugh L (1958) Point-sampling and line-sampling: probability theory, geometric implications, synthesis. Occ. Paper 160, USDA For. Serv. South. For. Exp. Stn

Gurnaud A (1886) La Sylviculture française et la méthode du Contrôle. Imprimerie Paul Jacquin, Besançon

Horvitz DG, Thompson DJ (1952) A generalization of sampling without replacement from a finite universe. J Am Stat Assoc 47:663. https://doi.org/10.2307/2280784

Kaiser L (1983) Unbiased estimation in line-intersect sampling. Biometrics 39:965–976

Köhl M (1994) Statistisches Design für das zweite Schweizerische Landesforstinventar: ein Folgeninventurkonzept unter Verwendung von Luftbildern und terrestrischen Aufnahmen, Mitteilungen der Eidgenössischen Forschungsanstalt für Wald, Schnee und Landschaft, vol 1. Eidgenössische Forschungsanstalt für Wald Schnee und Landschaft

Köhl M (2001) Inventory concept NFI2. In: Brassel P, Lischke H (eds) Swiss National Forest Inventory: methods and models of the second assessment. Swiss Federal Research Institute WSL, Birmensdorf. http://www.lfi.ch/publikationen/publ/methoden-en.php

Lanz A (2000) Optimal sample design for extensive forest inventories. Doctoral Thesis, ETH Zürich, Switzerland

Madigan D, Bartlett P, Bühlmann P et al (2014) Statistics and science – a report of the London workshop on the future of the statistical sciences

Mahrer F (1976) Abgrenzungsnormen zur Erfassung der Waldfläche im Schweizerischen landesforstinventar (LFI). In: Beiträge zur Bestimmung des Begriffes „Wald". Bericht Eidg. Anstalt für das Forstliche Versuchswesen, 167, Birmensdorf, Switzerland

Mandallaz D (2008) Sampling techniques for forest inventories. Chapman & Hall/CRC, New York

Mandelbrot B (1967) How long is the coast of Britain? Statistical self-similarity and fractional dimension. Science 156(5):636–638

Martin GL (1982) A method for estimating ingrowth on permanent horizontal sample points. For Sci 28(1):110–114

Paschedag I, Zinggeler J (2001) Forest transportation system survey. In: Brassel P, Lischke H (eds) Swiss National Forest Inventory: methods and models of the second assessment. Swiss Federal Research Institute WSL, Birmensdorf, pp 162–196

Pulkkinen M, Ginzler C, Traub B, Lanz A (2018) Stereo-imagery-based post-stratification by recursive partitioning in Swiss National Forest Inventory. Remote Sens Environ 213:182–194

Roesch FA, Green EJ, Scott CT (1989) New compatible estimators for survivor growth and ingrowth from remeasured horizontal point samples. For Sci 35(2):281–293

Roesch FA, Green EJ, Scott CT (1991) Compatible basal area and number of trees estimators from remeasured horizontal point samples. For Sci 37(1):136–145

Särndal CE, Swensson B, Wretman J (2003) Model assisted survey sampling. Springer Science & Business Media, New York

Schmid-Haas P (1983) Swiss Continuous Forest Inventory: twenty years' experience. In: Bell JF, Atterbury T (eds) Renewable resource inventories for monitoring changes and trend. Proc., 15–19 August 1983, Corvallis, OR, SAF 83–14. pp 133–140

Schwyzer A, Lanz A (2010) Verjüngungserhebung im schweizerischen Landesforstinventar. In: Jahrestagung der Sektion Forstliche Biometrie und Informatik im Deutschen Verband der Forstlichen Forschungsanstalten (22.) und der AG Ökologie und Umwelt in der Internationalen Biometrischen Gesellschaft – Deutsche Region, Klaus Römisch, TU Desden Fachrichtung Forstwissenschaften Tharandt. Arne Nothdurft, Forstliche Versuchs- und Forschungsanstalt Baden-Württemberg (FVA) Freiburg. Uwe Wunn, Forschungsanstalt für Waldökologie und Forstwissenschaft Reinland-Pfalz in Lippstadt, Göttingen, Die Grüne Reihe. ISSN: 1860-4064

Van Deusen PC, Dell TR, Thomas CE (1986) Volume growth estimation from permanent horizontal points. For Sci 32(2):415–422