Parameter estimation in the second cycle of Czech National Forest Inventory

One-phase estimators using field data of the CzNFI1 sampling grid

(Version of July 26, 2018)

ACNIL – ref. UHUL/2476/2016/KM

Ing. Radim Adolt, Ph.D.

Forest Management Institute Brandýs nad Labem

July 26, 2018

Contents

1 Introduction

This document puts forth a methodology for one-phase estimation of target parameters of the second cycle of Czech National Forest Inventory (CzNFI2, 2011-2014). Methods described herein are applied to data obtained by field survey carried out on sampling locations established during first NFI cycle (CzNFI1 sampling grid).

Section 1 describes the local density function as an indispensable component in the NFI parameter estimation based on the Horwitz-Thompson theorem for populations of the continuum (see Appendix A). Description of local density function is restricted to the sample protocols used to select stems (tress) and to the point sampling. These two sampling approaches are used to estimate the sums and per-hectare values of the population of trees (total and per-hectare of stand area growing stock, felling, increment, etc.) as well as to estimate the total area of the *Forest* category and its subcategories (e.g. area covered by stands of trees).

The two sections that follow are devoted to the one-phase estimation of total and ratio (of two totals or equivalently means). The main part of the methodology is completed with appendices, to illustrate some typical and substantial properties of the NFI projects in general and to show how the most important relationships forming the basis for the main part of the methodology are derived.

Only the routinely used estimators of NFI parameters are described in this methodology. The text focuses nearly exclusively on the design-based estimators. Issues relating to dendrometry, and forest management theory and practice are not covered – first, the extent of this booklet is too small and second, efforts have been made to maintain the clear arrangement of the text as a whole.

2 Local density

The objective of the Czech National Forest Inventory (CzNFI) sample survey is to estimate sums (totals) or mean values or ratios (of sums or mean values) in target populations (universes) and their parts (sub-populations). In general, national forest inventories (NFIs) are specific by a high number of populations and by the requirement to estimate many population properties (parameters). A typical NFI focuses both on populations of discrete objects that are countable, theoretically at least (e.g. trees, pieces of lying dead wood), and on populations where differentiation between their elements is difficult or irrelevant. Example: population consisting of a theoretically as well as practically uncountable number of elements – dimensionless sites, points, whose union is an area possessing some properties, e.g. an area which is classified as forest. This example is a population of a continuum – subset of the Cartesian coordinate system $\mathbb{R}^2$.

From the practical NFI project feasibility and cost rationality aspects, a method must be sought to effectively perform sampling both from populations of discrete objects and from populations of a continuum and, based on the data obtained by sampling, to estimate required target parameters (relatively simply and, if feasible, basically by the same procedure). A solution consisting in the introduction of a *local density function* has been proposed by Mandallaz [1991, Section 3.1, p. 18]. The very fundamental property of this local density function $Y(x)$ is expressed by Eq. (1):

$$\int_D Y(x)dx = Y, \tag{1}$$

which says that for the quantity in question, the sum of local densities on all points x of the area of interest $D \subset \mathbb{R}^2$ (geographically defined domain) is equal to the total Y of that quantity. The expected value of local density equals the mean value of the quantity within D, see (2):

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

where $\lambda(D)$ is the size of D.

A sampling protocol has to be designed for each target parameter considered by the CzNFI project so that we should be able, based on the data obtained, to define the local density function and evaluate it for each sample location (inventory point). In this manner target parameters can be estimated by using an unified theory - the Horwitz-Thompson theorem for populations of a continuum, see Appendix A, p. 24.

Properties of the local density function were also investigated in detail by Lanz [2000, Chapter 2, p. 26]. Evaluation of various approaches to design-based estimation for a population of trees can be found in the article by Eriksson [1995].

2.1 Selection of trees by circular plots of fixed size

This section uses the formal description that has been published by Mandallaz [2007, Section 4.2, p. 55]. When selecting stems[1] by circular segments, a few (typically 2 or 3) concentric circular segments, denoted K_r, are defined at each of the inventory points. Stem i s selected if it occurs inside the circle $K_r(x) = \{y \in \mathbb{R}^2 \mid d(x,y) \leqslant r\}$ having a radius r, with its centre in the inventory point x. The stem position is defined by the

[1] The difference between a tree and a stem is insignificant at this stage.

perpendicular projection of the point lying on its central axis at the breast height[2] into the horizontal plane given by the coordinate system in use (S-JTSK). The notation $d(x, y)$ is used for the Euclidean distance between points x (circle centre, inventory point) and y (a point of the circle). Let us define the random indicator variable $I_i(x)$ for stem i and inventory point x:

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

and, furthermore, N circles $K_i(r) = K_r(u_i)$ with a constant radius r and with centres u_i at the positions of the stems. Stem i is selected if and only if the inventory point x falls within the circle $K_i(r)$:

$$I_i(x) = 1 \Leftrightarrow x \in K_i(r). \tag{4}$$

Each stem i can be assigned a circle $K_i = K_{r_i}(u_i)$ (called inclusion zone), whose radius depends on the stem's properties i, e.g. its diameter at breast height (d_{13i}), total height (h_i), tree species, the geographically defined domain it belongs to, etc. In a typical NFI project the size of the circular segment K_i depends on the stems' diameter at breast height. Two circular segment sizes are used for stems above merchantable wood threshold ($d_{13i} \geqslant 7\ cm$) within the CzNFI1 sampling grid:

$$r_{i_{CzNFI2}} = \begin{cases} 3.00\ m & \text{for stems } 7\ cm \leqslant d_{13i} < 12\ cm, \\ 12.62\ m & \text{for stems } d_{13i} \geqslant 12\ cm. \end{cases} \tag{5}$$

Local density for the above described sampling protocol is defined at any point $x \in D$ as follows (6):

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

where $\lambda(K_i)$ is the size of the circular segment K_i of stem i and $\mathcal{P}$ is a finite population of stems occurring at an accessible site of the area of interest D and, at the same time, within the *Forest* land-use category (henceforth denoted F). The notation Y_i is used for the value of the quantity observed on stem i, whose sum for the entire stem population $\mathcal{P}$ is to be estimated.

Density so defined is in agreement with Eq. (1) (p. 6) only if we register also stems selected by circular segments of an inventory point located outside the territory of the Czech Republic, or beyond the *Forest* category or on its inaccessible parts. However, such an approach requires field measurements and permanent stabilisation on a considerably larger number of inventory points, and therefore was not applied during the CzNFI1 or CzNFI2 cycles: stems were only registered on plots centres of which were located within the Czech Republic's territory and, at the same time, on an an accessible part of the *Forest* category. For stems occurring near the country's boundary, near the edge of the *Forest* land category and/or near to an inaccessible area, true inclusion zones do not match the circular segment K_i (nominal inclusion zone). In general, the actual inclusion zone is given by the intersection $K_i \cap S \cap F \cap A$, where S is the territory of the country, F is the *Forest* category and A s the accessible part of D. This implies that a smaller inclusion zone must be used in the denominator of expression (6) for stems near the already mentioned

[2]The height of 1.3 m measured as the distance from the point of intersection of the stem axis and the horizontal plane passing through the stem base; so the slant distance is measured if a leaning stem is involved.

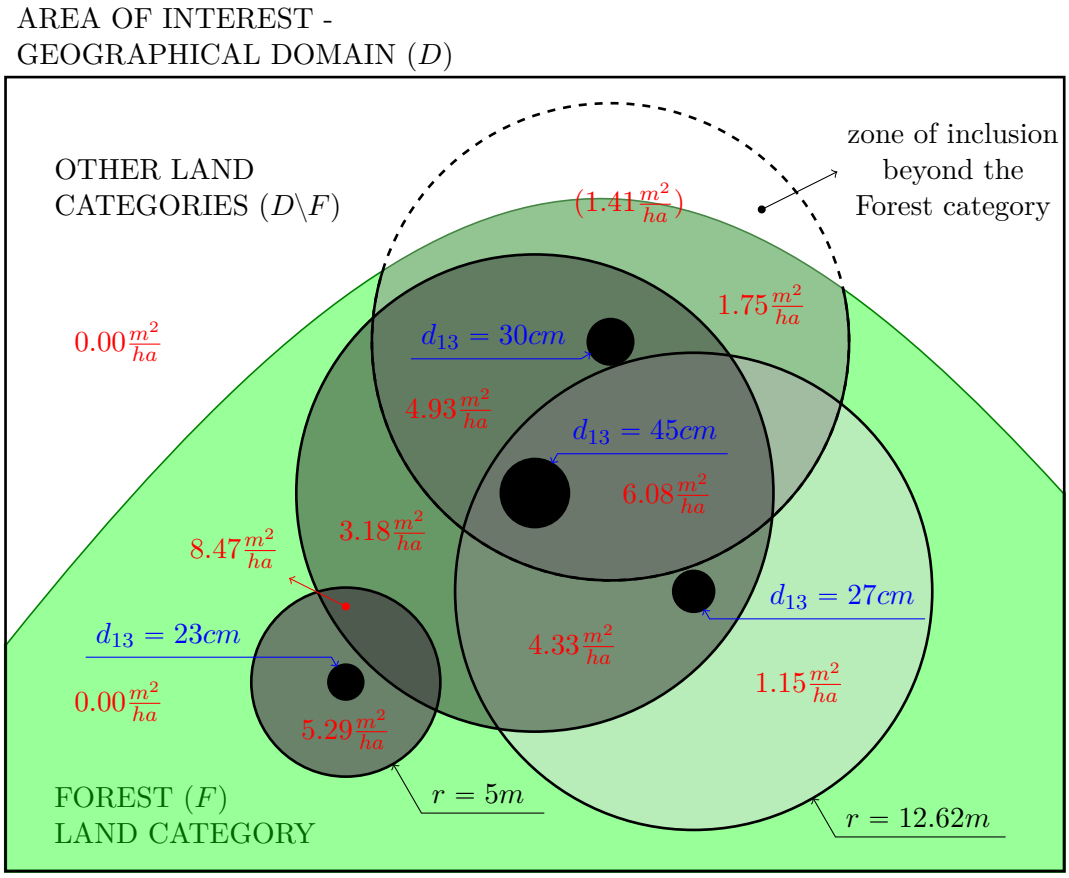

Figure 1: Diagram of the local density function (7) for the estimation of the total basal area, selection of stems by concentric circles within the CzNFI2 sampling grid (two concentric circles whose radii are $r_1 = 5\ m$ and $r_2 = 12.62\ m$ with corresponding diameter thresholds $d_{13,\,r_1} \geqslant 7\ cm$ and $d_{13,\,r_2} \geqslant 27\ cm$), situation near the edge of the *Forest* land-use category, the value of 1.41 $m^2 ha^{-1}$ given in parentheses is the uncompensated local density according to Eq. (6).

boundaries. If the area of the circular segment K_i is used without discrimination in the denominator of (6), the so called "edge effect" occurs, i.e. underestimation of the total amount or mean value of tree resource within the area of interest D.

This underestimation may be negligible if small circular segments compared to the mean sizes of the continuous areas of the *Forest* category are used. Some authors recommend the boundary effect to be compensated, warning that the underestimation may be so large that it approaches a sampling error [Mandallaz, 2007, Section 4.2, p. 60]. Taking into account the technology level of CzNFI2 (stem position mapping in S-JTSK, mapping of the edges of the Forest category and of its inaccessible parts within adequately large surroundings of the inventory point), an edge effect compensation method was used ('Measure π'), based on the knowledge of the true inclusion zone of all stems selected within CzNFI2 [Gregoire & Valentine, 2008, Section 7.5, p. 224]. The local density on an inventory point can be calculated using a modified formula (7):

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

For local density defined by Eq. (7) we obtain (8):

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

where $\bar{Y}$ is the mean area value of the quantity within area D whose size (surface area) is $\lambda(D)$.

Unbiased estimators of the total of the various stem population quantities can be constructed by using Eq. (7) within the theory of the Horwitz-Thompson estimation for populations of a continuum (see Appendix A, p. 24). Figure 1 is a schematic layout of the situation near the edge of the *Forest* land category including visualisation of the local density values as per Eq. (7) for the estimation of total basal area (total of cross sectional areas at breast height for all stems on accessible parts of *Forest* land-use category, within an arbitrarily defined D).

Some of the CzNFI target parameters are defined as sums or means of certain stem properties that are not directly measurable on standing (growing) stems (trees in the common meaning). Total growing stock which is defined as a sum of merchantable timber volumes of individual stems[3] is an example. Equation (9) is a way to identify merchantable volume of a stem based on a prediction by using the so-called volume tables:

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

The predicted value Y_i is obtained by inserting the measured diameter at breast height d_{13i} and stem (tree) height h_i into the functional relation $f_{ult}(..)$ of the ÚLT volume tables for a given tree species. In practical national forest inventory both in the Czech Republic and abroad, the error of stem's merchantable volume prediction by using volume tables based on the *measured* quantities (diameter at breast height, total tree height & some other quantities, typically an additional measurement of upper stem diameter at certain height) is neglected[4] [Mandallaz, 2007, Section 9.5, p. 166].

Field measurements of tree height are very time-consuming in routine field survey conditions. A two-step estimation method based on tree height prediction by using a dendrometric model has been developed to alleviate this:

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

where the measured diameter at breast height d_{13i} is used as the predictor. In contrast to model of volume tables, the height curve model is not expected to be free from error. The error arising from the use of the height prediction instead of its measured value is compensated based on *second-stage sample stems*. For this, the local density function is extended with a correction term – see the second summand in the following equation:

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

which is an estimator of the error due to the use of the height curve model, i.e. estimator of the difference $E = Y - Y^*$. The first term in Eq. 11 is the unbiased estimator of the total stock Y^* based on the volume tables using the predicted stem heights h_i^*. The term $Y_i - Y_i^* = E_i$ s the difference between the "true" stem merchantable timber volume

[3] Volume of all above-ground timber mass whose thickness, measured over-bark, is no smaller than 7 cm; stump volume is not included.

[4] If specific volume tables are consensually used at the national level, different sources of information about growing stock are comparable.

found for a particular stem (according to the volume tables when inserting the measured diameter at breast height d_{13i} and the measured height h_i values) and the model volume (based on measured d_{13i} but predicted stem height h_i^*). The differences E_i are only known on the second-stage sample stems (stem height measured). The $\ddot{I}_i(x)$ term is the indicator variable of second-stage sample stems, for which we have:

$$\ddot{I}_i(x) = 1 \;\Rightarrow\; I_i(x) = 1. \tag{12}$$

The symbol p_i is the probability of selection of a stem registered at the inventory plot (1st-stage sample stem) also as a second-stage sample stem, for which the height is measured and timber volume is derived based on volume tables. The expected (mean) value of the indicator variable $\ddot{I}_i(x)$ conditionally on the inclusion of the given stem into the 1st-step sample (stems registered within circular segments) is equal to probability p_i in other words, we have:

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

The definition of the generalised local density according to Eq. (11) is the adopted and modified definition by Mandallaz [2007, Section 4.4, p. 69], who also presents a proof of identity of its expected value with the mean value in D:

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

Equation (14) applies if the height curve model (10) is *external*, which means that it has not been parameterised by using actual sample stems. The first term in Eq. (11) is constant for a fixed point $x \in D$ assuming that an external model is used – the first term value does not change among replicated (hypothetical or true) second-stage samples of stems. The second term in Eq. (11) varies at random with repetitions of the second stage of sampling. Its expected value is identical with error E of the total growing stock estimator arising from the insertion of the predicted stem heights h_i^* instead of the measured ones into Eq. (9).

[Mandallaz, 2007, Section 9.5, p. 165] presented in its monograph a result according to which it is optimal if the probability p_i of inclusion of a given stem into the second-stage sample is proportional to the stem's merchantable volume prediction error. We knew based on our own analyses that this error increases roughly linearly with the stem's basal area (stem's cross section area at breast height). Therefore, we decided to select the second-stage stems by so-called relascopic method (angle count sampling) – sampling with a probability, or more precisely with the inclusion zone size, proportional to the stem's basal area. Consistent with this choice is the following definition of the generalised local density, see the equation:

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

Here $\lambda(K_{BAF,i})$ is the size of the zone of inclusion corresponding to the relascopic method with a selected basal area factor (BAF). For CzNFI2, the radius of the inclusion zone of the second-stage sample stems was calculated by using the formulas:

$$r_i = \begin{cases} d_{13i}/(2\sqrt{BAF})\ m & \text{for stems } 7\ cm \leqslant d_{13i} < 87.4\ cm \\ 12.6156626101008\ m & \text{for stems } d_{13i} \geqslant 87.4\ cm, \end{cases} \tag{16}$$

where $BAF = 10^4 \sin^2(\alpha/2)$ is the *relascopic factor* in m^2/ha a α is the *relascopic device's angle of sight* (deg). The relascopic factor value used was 12 m^2/ha. The symbol $s_1(x)$ denotes first-step sample, i.e. stems registered by circular segments according to Eq. (5) on the inventory plot with its centre in point x. The set of second-stage stems selected by circular segments according to Eq. (16) is denoted $s_{1,2}(x)$.

In contrast to Eq. (11), the second term in Eq. (15) does not change with replications of the second-stage of sampling (assuming the use of an external height curve model). First and second stage sample stems are identified simultaneously by selection of the sample point $x \in D$. If the external height curve model is used, the second term in Eq. (15) is an estimator of the total of a specific local density type (functional surface within D), which is the correction of the total growing stock based on volume tables using predicted stem heights to the total growing stock based on the same volume tables but using measured stem heights. The corrective local density does not change when passing from one sampling design replication to another, which is convenient when estimating change of growing stock and growth. On a given sample plot, change of the second term in Eq. (15) between two repeated NFI surveys (inventory cycles) is only possible when changing the enumeration of the individuals in the stem population and/or their dimensions. By contrast, the second term in Eq. (11) changes in the next inventory cycle also due to the differences in the sets of second-stage stems (assuming the positions of sample plots are kept among inventory cycles). This brings about an increase in the variance of the estimated change of growing stock and growth as compared to the local density definition according to Eq. (15).

The specific procedure to estimate the growing stock in CzNFI2 and of additional parameters based on determination of the timber volumes of the individual stems (total and per hectare estimates of felling, growth, mortality and growing stock change) uses the local density according to Eq. (15) with the relascopic factor 12 m^2/ha and inclusion zones for second-stage stems according to Eq. (16). The Michajloff curve served us to capture the diameter - height relationship:

$$h_i^* = a \cdot e^{-\frac{b}{d_{13i}}} + 1.3. \tag{17}$$

Based on this function a nonlinear mixed-effects model was fitted for the normal distribution of errors. Random effects were considered for the two parameters (a and b) at the stand segment level[5]. The model was parametrised for our most abundant tree species based on data from the first inventory cycle (CzNFI1). Prediction for tree species for which the model had not been parametrised was performed based on the model for a tree species with the closest environmental demands and growth properties.

Application of the model for the CzNFI2 evaluation included diameter-height curve calibration. Prediction of the so-called random effect values was based on a pair of diameter at breast height and stem height measurements taken on one single *diameter-height curve calibration sample stem* selected for each represented tree species and stand segment on each sample plot. Sample stems for diameter-height curve calibration were selected at random from the list of stems of the tree species in question in the given the stand segment, registered on the sample plot (set of first stage-sample stems). This selection was absolutely independent of the selection of second-stage stems conducted by the relascopic method. If no adequate candidate for calibration measurement was present among first-stage sample stems (e.g. the single stem for particular tree species within the particular

[5]A stand segment is a homogeneous part of a forest with a specific internal structure characterised by the tree species composition, age structure, spatial arrangement of the individuals, their origin and framework management patterns.

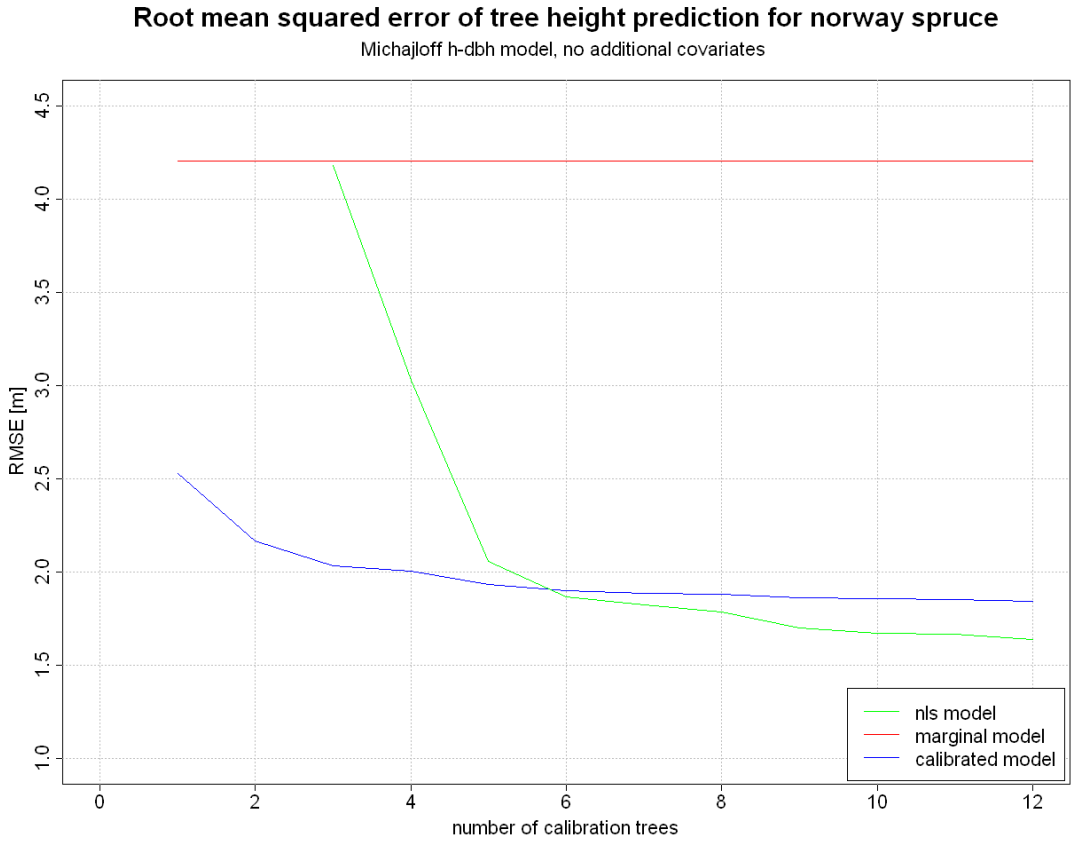

Figure 2: Cross-validation of the prediction error variability for the conventional Michajloff diameter-height curve model (nls) and a corresponding mixed-model (marginal and calibrated prediction).

stand segment leaned appreciably), the calibration stem was selected subjectively in the nearest surroundings of the sample plot – preferentially within the same stand segment. Always, however, the calibration stem selected outside the plot perimeter had to be of comparable age and growing in identical site conditions. If no calibration sample was found in the nearest surroundings either, the stem timber volume was derived by using marginal prediction of stem height (zero value of random effects of the two parameters i.e. a population 'average' model for particular species was used to predict stem heights).

The calibration itself was performed by applying a procedure based on Taylor's approximation as described by Trincado & Burkhart [2006, formula 9, p. 674]. A more precise iterative calibration procedure for the nonlinear mixed model has been published by Bohora & Cao [2014, formula 3, p. 63].

It is clear from nomogram 3 that even if a single calibration stem per tree species and stand segment is selected, the effect of the BAF on the sampling error diminishes appreciably. The increasing number of calibration samples is accompanied by nearly no decrease in the sampling error, which is consistent with the development of the root mean squared error (RMSE) of the Michajloff curve model with a varying number of calibration measurements (see diagram 2).

The choice of $BAF = 12\ m^2/ha$ along with one diameter-height curve calibration mea-

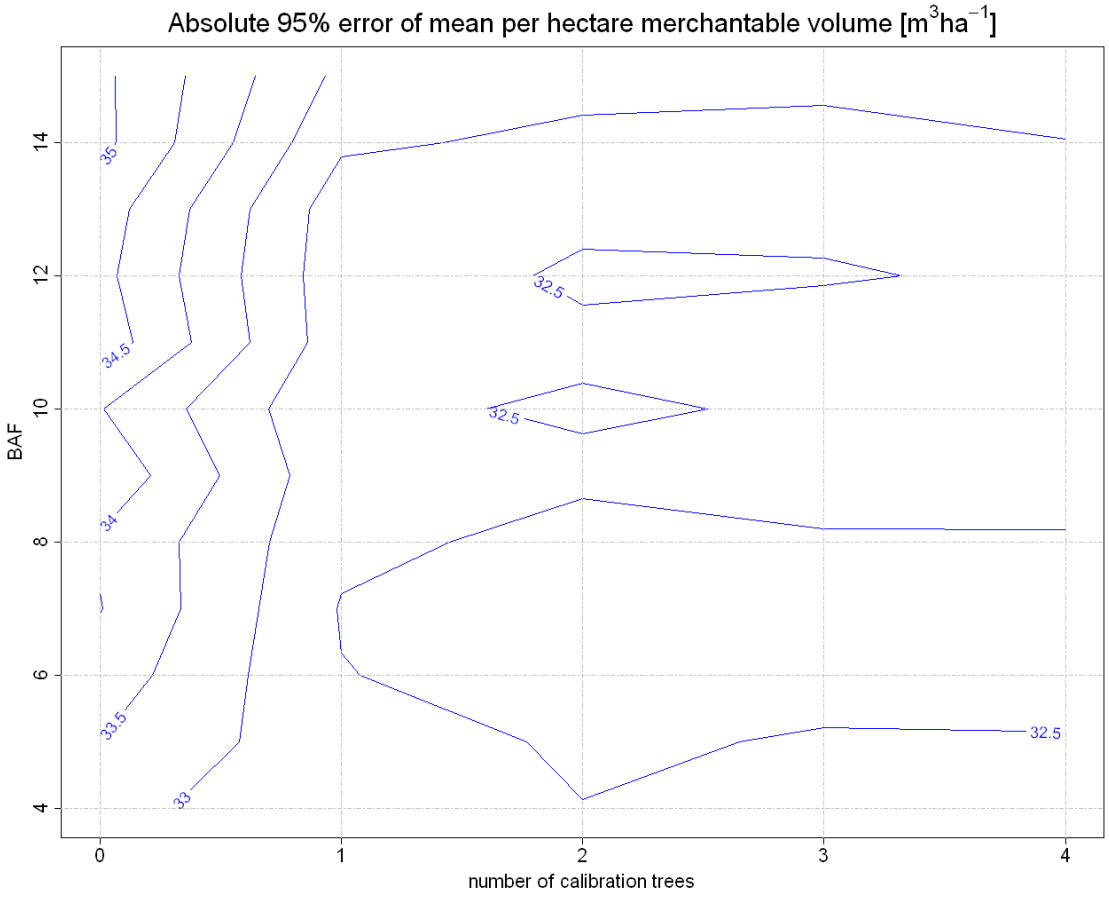

Figure 3: Nomogram showing the a half 95% confidence interval of the per hectare growing stock estimate as a function of the BAF used for the selection of second-stage stems and of the number (per stand segment) of independently selected diameter-height curve calibration stems (CzNFI2 pilot project in the natural forest region of the Central Moravian Carpathians, 2010).

surement for each stand segment and each tree species represented in the stand segment leads to a mean number of 5.31 heights measured per sample plot (calculation based on the full extent of CZNFI2 data stored in the analytical database). This number is consistent (almost identical) with the previous estimate obtained within the CzNFI2 pilot project (natural forest region of the Central Moravian Carpathians, 2010).

2.2 Local density for point sampling

When estimating the area of a territory of a given properties e.g a land belonging to an arbitrarily defined land-use category denoted K, local density is equal to the indicator variable $I_K(x)$, see formula (18).

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

Indicator variables for any other categories (subsets of D) are defined analogously.

2.3 Local density for attribute domains

Target parameter estimates are nearly always categorised in what is called attribute domains[6]. The total growing stock estimate divided into the tree species groups can be an example. The term 'attribute domain' may also be used in geographical context - to denote a (sub)population to which a certain part (of unknown size) of the area of interest D can be attributed. *Forest* land-use category as well as its subcategory 'stand area' are typical examples. No absolutely precise map of *Forest* category exists, nor do we know the exact area of the corresponding part of D. This holds also for the 'stand area'.

An estimator for an attribute domain is based on a modification of the local densities by elimination of contributions from those population elements that do not fall in the specific attribute domain (subpopulation). As an illustration, the definition of the local density (15) modified for the purpose of estimation of total growing stock of 'oak' attribute domain is shown here:

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

The value of the $I_{oak,i}$ indicator variable is 1 if the i-th stem is oak and 0 otherwise.

The next example is the local density for the estimation of the area of 'stand area' subcategory of the *Forest* land-use category. This is the product of two indicator variables:

$$I_{F,sa}(x) = I_F(x) I_{sa}(x), \tag{20}$$

where $I_F(x)$ is the indicator variable for the *Forest* land category and $I_{sa}(x)$ s the indicator variable for 'stand area'. The use of two indicator categories is justified by the fact that 'stand area' may be distinguished independently from the *Forest* category - it is also considered within *Other Wooded Land* and *Other Land with Tree Cover* land-use categories.

2.4 Local density at the CzNFI1 tract level

Sampling locations (inventory points, sample plots) are arranged in pairs or duplexes, called tracts, in the CzNFI1 sampling grid. The nominal distance between the inventory points of a tract is 300 m (corresponding to the 2D Euclidean distance between the generated coordinates of the inventory points of the same tract). Lines connecting the inventory points of a tract are oriented at random. A detailed description of the CzNFI1 sampling grid is presented in the CzNFI1 project monograph [Jankovská & Štěrba, 2007, Section 4.1, p. 51].

Inventory points (sample plots) of the same tract are not mutually independent with respect to the information gained from them. The values of a number of quantities observed at sample plots within a tract are, in the average, more similar than between sample plots selected at random irrespective of their assignment to the tracts.

For an arbitrarily defined local density satisfying Eqs. (1) and (2) the density at the level of a tract of the CzNFI1 grid is defined as follows:

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

[6]Geographic domains are also used in addition to attribute domains. A geographic domain is a specific part of the area of interest D, whose size is precisely known. Administrative regions (NUTS3 units) and natural forest regions are examples of geographic domains. CzNFI estimates are provided for combinations of geographic and attribute divisions.

The symbol x in $Y_{tct}(x)$ denotes a reference point of the tract, for which we take the first from the pair of inventory points (x_1), $I_D(x_1)$ and $I_D(x_2)$ are indicator variables, whose values are 1 if the corresponding inventory point (x_1 or x_2) lies within the area of interest (geographic domain) D and 0 otherwise.

Mandallaz [2007, Section 4.3, p. 65] published a one-phase estimation procedure for a design using tracts with a fixed geometry, defining local density at the tract level in a different manner. This alternative procedure has not been applied in view of the random orientation of the tracts in the CzNFI1 grid and the requirement for geographically additive estimators. Additional information on this topic can be found in Section 3.

3 Estimator of total Y within an area of interest D

The one-phase estimator of the total of Y within D (geographic domain, area of interest of a known size) is calculated by this methodology as

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

where $\lambda(c)$ is the size of the inventory block (square) in the CzNFI1 grid (2 km×2 km, i.e. 400 ha). The symbol $Y_{tct}(x)$ denotes local density at the tract level, see Eq. (21). The calculation encompasses all tracts whose reference inventory point $x = x_1$ lies within the area of interest expanded by 300 m in all directions (buffer zone 300 m wide along the circumference of D). This area is denoted D^+.

Selection of this specific total estimator is discussed in the concluding part of Appendix C. The estimator (22) possesses following properties:

- It is unbiased unconditionally on the number of tracts with the reference point in D^+, see Appendix E;
- It is additive in the geographical as well as attribute domain sense, see Appendix F

The variance estimate for Eq. (22) is calculated as:

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

where $\lambda(D^+)$ is the size of the area D^+ and $\bar{n}_{D^+}$ is the expected (mean) number of inventory tracts with their reference points in D^+. The term $\hat{\mathbb{V}}(n_{D^+})$ is the estimated variance of the number of tracts with reference points in D^+, which was obtained for each definition of area of interest D under consideration based on a large number (2000) of simulated samples using the design of the CzNFI1 sampling grid.

The symbol $\widehat{\overline{Y^2_{tct}}}$ is the ratio of the sum of squares of tract level local densities to the expected sample size $\bar{n}_{D^+}$. The term $\hat{\bar{Y}}^2_{tct}$ is the square of the ratio of the sum of tract level local densities to the $\bar{n}_{D^+}$. The calculation of the two ratios encompasses all tracts with reference points in D^+ (and no others).

The variance estimator according to Eq. (23) is conservative, see Appendix D. It is derived in Appendix (E). The one-phase estimator of the total and its variance is methodologically based on the Horwitz-Thompson theorem for the population of a continuum (HTC), described in Appendix A.

Notes

1. Since the inventory points in the CzNFI1 grid are arranged in tracts that are rotated at random, we use a specific definition of the local density function. This is the arithmetic mean of the two plot-level local densities for each tract, see Section 2.4. The calculation encompasses all sample plots of all tracts whose reference point (the first in the pair) is located inside the buffer zone denoted D^+ defined as the area of interest D expanded by 300 m along its circumference (under any conceivable angle but in the direction outwards of D). Local density on the inventory points $x \in D^+, x \notin D$ (within the buffer zone but beyond D) is set to zero when calculating the arithmetic mean of the densities for the given tract. Owing to the use of the

buffer zone, the edge effect at the tract level does not need not be compensated – this effect does not occur because each stem (tree) may be selected either by the first or by the second sample plot of a tract (depending on the rotation).

2. For a design that uses tracts (but with a fixed geometry, with no random tract rotation), Mandallaz [2007, Section 4.3, p. 65] defined local density at the tract level in a different way. The different procedure used by this methodology is based on the fact that the CzNFI1 grids were not only selected with respect to the planar Cartesian coordinate system S-JTSK; in fact, the angle of rotation between the first and second points of a tract was also included. The inclusion zones of the individual stems, conditionally on the tract orientation, are two circles whose size corresponds to measured diameter at breast height. The nominal distance between the circle centres is 300 m, with a defined (but variable between the replications) orientation of the interconnecting line[7]. Hence, the local density function is not defined in a plane only, it is defined in the three-dimensional space[8]. Sections perpendicular to the coordinate axis corresponding to the tract rotation angle are planes in which the 'conventional' local density is defined (for a specific angle of rotation).

[7] An inclusion zone that is not conditional on the tract rotation angle has the geometry of two parallel coiled pipes.

[8] Actually, the sample space for samples having just one element i.e. $n_{D^+} = 1$ is also three-dimensional.

4 Estimator or ratio $R_{1,2}$ within the area of interest D

According to this methodology, the ratio of two totals Y_1 and Y_2 within the interest area is estimated as

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

where $\hat{Y}_1$ and $\hat{Y}_2$ are total estimators defining the ratio itself and $Y_{tct,1}(x)$ and $Y_{tct,2}(x)$ are local densities at the tract level calculated on the principle (21), see page 14. The sampling weights in the numerator and denominator on the right-hand side of Eq. (24) cancelled each other out[9].

The calculation of the ratio estimate by Eq. (24) encompasses only tracts with any of the inventory points in area A, representing the expanded attribute domain of unknown size with the following definition:

$$A \;= \Big\{ x \in D \mid Y_1(x) \neq 0 \vee Y_2(x) \neq 0 \Big\}. \tag{25}$$

Hence, only tracts with the initial points within the region $A^+ \subseteq D^+$, which is obtained by expansion of A outwards by an buffer zone 300 m wide are involved.

Variance (24) is estimated by using the formula

$$\hat{\mathbb{V}}_{URS}\big(\hat{R}_{1,2}\big) = \frac{n_{A^+}^2}{(n_{A^+} - 1)\left[\sum_{\substack{x \in s \\ x \in A^+}} Y_{tct,2}(x)\right]^2} \left[\widehat{Z_{tct}^2} - \hat{\bar{Z}}_{tct}^2\right], \tag{26}$$

where $\widehat{Z_{tct}^2}$ and $\hat{\bar{Z}}_{tct}^2$ are the arithmetic mean of the square and the square of the arithmetic mean, respectively, of the residual variable $Z_{tct}(x) = Y_{tct,1}(x) - \hat{R}_{1,2} Y_{tct,2}(x)$ calculated for all tracts with their initial points in A^+.

Variance estimator (26) has been derived from Taylor's approximation, see Appendices B, B.1 and, in particular, B.2. The estimator $\hat{R}_{1,2}$ is not an unbiased for $R_{1,2}$. The bias of $\hat{R}_{1,2}$ is negligible for sample sizes occurring practically in the CzNFI (from tens to hundreds upwards).

Notes

1. A typical example of a ratio estimation is the per hectare of stand area growing stock – the ratio of the total growing stock and the extent of stand area (both totals are unknown and must be estimated). Due to the definition of A^+ the calculation also encompasses tracts that do not encroach on the stand area provided that the local density for growing stock is non-zero for that particular tract. This procedure avoids the occurrence of the edge effect – the chance to select a stem inside but near the edge of stand area is not reduced[10].

[9] The particular selection of the sampling weights is irrelevant if the same estimator is used in the numerator and in the denominator of the ratio.

[10] In contrast to accessibility and in contrast to the *Forest* land-use category, stand area in adequately large surroundings of the inventory point is not mapped during the CzNFI field work.

2. Ignorance of the precise A^+ is not an obstacle. However, we must be able to decide on the membership of the tracts to A^+. Situations exist, though, when none of the inventory points of a specific tract encroaches on A, although the second point would, was the tract orientation different. Omission of such tracts[11] has no effect on the point estimate calculation. The variance estimate calculated by Eq. (26) is somewhat higher but the increase is negligible if tens of tracts and more are sampled.

3. The variance is estimated by approximating the pair densities $\pi(x_i, x_j)$ by a density corresponding to a uniform random point sampling (URS) in n_A^+. Spatial correlation of the residual density $Z_{tct}(x)$ is usually lower than in the case of the densities $Y_{tct,1}(x)$ and $Y_{tct,2}(x)$, from which it is calculated. Therefore, we assume (and confirmed many times by simulated sampling) a lower extent of overestimation (conservativeness) of the variance estimator for $\hat{R}_{1,2}$ compared to the variance estimator for conventional total $\hat{Y}$.

4. If the density $Y_{tct,2}(x)$ is an indicator variable and, due to the properties of the sample or to the nature of the populations, the relation $Y_{tct,2}(x) = 0 \implies Y_{tct,1}(x) = 0$ holds true, the estimators $\hat{R}_{1,2}$ and $\hat{\mathbb{V}}_{URS}\big(\hat{R}_{1,2}\big)$ are transformed into a form corresponding to the arithmetic mean and its variance, respectively.

[11] In fact, the tracts fall in A+, the value of the local density at the tract level, however, is zero on them.

References

Anděl, J. 1978. *Matematická statistika.* Praha: SNTL - Státní nakladatelství technické literatury.

Barabesi, L. 2003. A monte carlo integration approach to horvitz-thompson estimation in replicated environmental designs. *International journal of statistics*, **LXI**, 355–374.

Barabesi, L., & Franceschi, S. 2010. Sampling properties of spatial total estimators under tessellation stratified designs. *Environmetrics*, **(on-line publication)**.

Bohora, B., S., & Cao, V., Q. 2014. Prediction of tree diameter growth using quantile regression and mixed effects models. *Forest ecology and management*, 62–66.

Cochran, W. G. 1977. *Sampling techniques.* Willey series in probability and mathematical sttaistics - applied. John Willey & Sons.

Cooper, C. 2006. Sampling and variance estimation on continuous domains. *Environmetrics*, **17**, 539–553.

Cordy, C., B., & Thompson, C., M. 1995. An application of the deterministic variogram to design-based variance estimation. *Mathematical geology*, **27**, 173–205.

Cordy, C. B. 1993. An extension of the horwitz-thompson theorem to point sampling from a continuous universe. *Statistics and probability letters*, **18**, 353–362.

Eriksson, M. 1995. Design-based approchaes to horizontal-point-sampling. *Forest science*, **41**(4), pp 890–907.

Gregoire, T. G., & Valentine, H. T. 2008. *Sampling strategies for natural resources and the environment.* Chapman and Hall/CRC.

Heikkinen, J. (eds. Kangas A. & Maltamo M.). 2006. *Forest inventory, methodology and applications.* Springer Verlag. Chap. Assessment of Uncetainty in Spatially Systematic Sampling, pages 155–176.

Jankovská, Z., & Štěrba, P. (eds). 2007. *National forest inventory in the czech republic 2001-2004: Introduction, methods, results.* Ústav pro hospodářskou úpravu lesů Brandýs nad Labem (ÚHÚL).

Jarník, V. 1963. *Diferenciální počet i.* Nakladatelství Československé akademie věd.

Lanz, A. 2000. *Optimal samle design for extensive forest inventories.* Ph.D. thesis, Swiss Federal Institute of Technology (ETH), Zurich.

Li, Y. 2008. Generalized regression estimators of a finite population total using box-cox technique. *Survey methodology*, **34 (1)**, 79–89.

Mandallaz, D. 1991. *A unified approach to sampling theory for forest inventory based on infinite population and superpopulation models.* Ph.D. thesis, Swiss Federal Institute of Technology (ETH), Zurich.

Mandallaz, D. 2007. *Sampling techniques for forest inventories.* Chapman and Hall/CRC.

Meloun, M., & Militký, J. 1998. *Statistické zpracování experimentálních dat.* 2 edn. Praha: East Publishing.

Rao, C. R. 1965. *Linear sttaistical inference and its applications.* Wiley.

Rao, J. N. K., & Singh, M. P. 1973. On the choice of estimator in survey sampling. *Australian journal of statistics*, **15**, 95–104.

Ripley, B. D. 2004. *Spatial statistics.* John Willey & Sons.

Sakia, R., M. 1992. The box-cox transformation technique: review. *The statistician*, **41**, 169 – 178.

Stevens, D. L. Jr. 1997. Variable density grid-based sampling designs for continuous spatial populations. *Environmetrics*, **8**, 167–195.

Stevens, D. L. Jr., & Olsen, A. R. 2003. Variance estimation for spatialy balanced samples of environmental resources. *Environmetrics*, **14**, 593–610.

Stevens, D. L. Jr., & Olsen, A. R. 2004. Spatialy balanced sampling of natural resources. *Journal of the american statistical association*, **99**, 262–278.

Särndal, C. E., Swensson, B., & Wretman, J. 2003. *Model assisted survey sampling.* Springer.

Trincado, G., & Burkhart, E. 2006. A generalized approach for modeling and localizing ste profile curves. *Forest science*, **52**, 670–681.

Methodological appendices

A The Horwitz-Thompson theorem for the population of a continuum

Denote Y the total (sum) of the quantity $Y(x)$ (local density function values) in an area D being a geographical domain, i.e. a set of points x of a continuum.

$$Y = \int_D Y(x)dx \tag{27}$$

The area D can be any defined area of interest for which the estimate is required (territory of a country, NUTS unit, natural forest region, ...). Unless stated otherwise at a specific site of this methodology, it is assumed that the size of D, referred to as $\lambda(D)$, is absolutely precisely known.

Cordy [1993] published the Horwitz-Thompson theorem for populations of a continuum, according to which an unbiased estimator $\hat{Y}$ of the sum Y is defined by formula (28) [Cordy, 1993, formula on p. 355, theorem 1 on p. 356].

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

The symbol $Y(x)$ denotes local density of the quantity observed at point x of sample s of fixed size n, $\pi(x)$ is inclusion density at point x which is an analogy to the inclusion probability when sampling finitely large populations of discrete objects [see Cordy, 1993, p. 355]. The estimator (28) is unbiased if the function $Y(x)$ is positive or bounded and, at the same time, the condition $\pi(x) > 0 \ \forall x \in D$ s satisfied, which means that any point x of the area D may be sampled.

If the local density function $Y(x)$ is bounded, $\pi(x) > 0 \ \forall x \in D$ applies and, simultaneously, $\int_D \frac{1}{\pi(x)} dx < \infty$, then the variance of $\hat{Y}$ can be expressed by Eq. (29), see Cordy [1993, theorem 2 on p. 357]:

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

If, at the same time, the condition,

$$\pi(x_i, x_j) > 0 \ \forall x_i, x_j \in D \tag{30}$$

is satisfied, the unbiased estimator of variance $\mathbb{V}(\hat{Y})$ can be obtained by using Eq. (31) or the modified Eq., see (32) [Cordy, 1993, theorem 2 on p. 357].

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

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

When comparing with the Horwitz-Thompson theorem for finite populations of discrete objects (see, e.g. [Cochran, 1977, Chapter 9A.7, p. 259], [Cordy, 1993, Section 1, formulas on p. 354], [Särndal *et al.* , 2003, Chapter 2.8, p. 42], [Mandallaz, 2007, Chapter 2.2, p. 4]) we find that the difference consists in substituting the inclusion probability π_i or $\pi_{i,j}$ by sampling density $\pi(x)$ or the pair-wise density $\pi(x_i, x_j)$ in Eqs. (28), (31) and (32), and in substituting the sums by certain integrals in Eqs. (27) and (29) and eliminating the term $1 - \pi_i$ from the numerator of the first addend in Eqs. (29), (31) and (32).

Eqs. (28) to (32) were derived and the unbiased nature statement holds for sampling designs with a constant sample size. If the sample size can differentiate from one realisation of sampling to the other, then either modified formulas taking the variable sample size into account must be used (see Section E, p. 40) or the estimators based on Eqs. (28) to (32) must be adopted as unbiased conditionally on the sample size (see Appendix C, p. 36). For the latter option the $\pi(x)$ and $\pi(x_i, x_j)$ values are expressed with respect to the actual sample size in D [Cordy, 1993, Section 3, p. 359].

As in the case of finite populations of discrete objects, the variance of the point estimator (28) can be alternatively expressed and estimated by using Sen-Yates-Grundy formulas (33) and (34) [Cordy, 1993, last paragraph on p. 358, continued on p. 359].

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

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

Eq. (34) can be modified to obtain the computationally more effective formula (35) if an arbitrary sequence $i = (1 \ldots n)$ of the inventory points $x_i \in s$ is introduced (n is the constant size of sample s) [see, e.g. Cochran, 1977, Chapter 9A.7, pp. 260 to 261].

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

Sen-Yates-Grundy formulas can only be applied to designs generating a fixed sample size. Their form for finite populations of discrete objects and for infinitely large populations of points of a continuum is identical. A proof of equivalence between the Horwitz-Thompson and the Sen-Yates-Grundy formulas for the variance of a point estimator in case of finite populations of discrete objects can be found in publications by Cochran [1977, theorem 9A.5, p. 260] and Särndal *et al.* [2003, proof on p. 45]. The variance estimates obtained by (31) and (34) are usually different for a particular sample, both, however, are unbiased.

Both the Horwitz-Thompson and the Sen-Yates-Grundy variance estimators can attain negative values in certain circumstances [Cochran, 1977, chapter p. 261], which is unacceptable from the theoretical as well as practical point of view. The Sen-Yates-Grundy estimates, however, are always positive if the condition (36) is met [Särndal *et al.* , 2003, note 2.8.7 on pp. 47 and 48].

$$\pi(x_i)\pi(x_j) - \pi(x_i, x_j) \ >= \ 0 \ \forall x_i \ \neq \ x_j \ \in \ D \tag{36}$$

The guarantee of positive values for a number of practically used and effective sampling schemes is the reason why the Sen-Yates-Grundy variance estimator is often given preference [Cordy & Thompson, 1995, last paragraph on p. 177], [Cooper, 2006, Section 4, p. 544]. Särndal *et al.* [2003, note 2.8.7 on pp. 47 and 48] note that in practical circumstances, the hazard of occurrence of negative variance estimates for medium and large samples is reasonably low and this situation need not be feared very much. By contrast, Stevens [1997, Section 5 on p. 192] notes that the Horwitz-Thompson variance estimates are frequently negative if the DM-NRTS (multiple-density nested random tessellation stratified sampling) design is used.

It has been noted that both the Horwitz-Thompson and Sen-Yates-Grundy estimators of variance of (28) can be unstable in some sampling schemes [Cochran, 1977, Chapter 9A.7, p. 261], [Cordy & Thompson, 1995, last but one paragraph on p. 177], [Stevens & Olsen, 2003, last paragraph on 598 continued on p. 599], [Stevens & Olsen, 2004, Section 4.1, p. 269], [Cooper, 2006, first and second paragraphs on p. 545].

Cochran [1977, Chapter 9A.7, p. 261] refers to the results by Rao and Singh [1973], who using Brewer's method to sample data from 34 published populations found that the Sen-Yates-Grundy variance estimators are considerably more stable than the Horwitz-Thompson estimators which, in addition, are frequently negative. According to Stevens & Olsen [2003, 2004] the instability is due to the possible occurrence of very low $\pi(x_i, x_j) \to 0$ values in the appropriate term $\hat{\mathbb{V}}_{HTC}(\hat{Y})$ or $\hat{\mathbb{V}}_{SYG}(\hat{Y})$, which affects the final variance estimate appreciably. This situation occurs typically for the not very likely but not impossible sampling of nearby points by using some of the effective sampling designs using positive spatial local density correlation. Instability of the $\hat{\mathbb{V}}_{HTC}(\hat{Y})$ and $\hat{\mathbb{V}}_{SYG}(\hat{Y})$ estimators is a problem particularly for sampling designs involving jump changes of sampling density $\pi(x)$ (e.g. at the boundary of areas – geographically defined domains with different sampling densities) [Stevens & Olsen, 2003, Section 3, first paragraph on p. 599]. According to Stevens [1997, Section 5, p. 192] the Horwitz-Thompson variance estimator for the DM-NRTS design has a tendency to instability and frequent occurrence of negative values while the Sen-Yates-Grundy estimator seems to be stable.

In the present methodology of CzNFI2 target parameters estimation based on data collected by field survey in the CzNFI1 sampling grid, the variance estimator used is based on an approximation of the pair density $\pi(x_i, x_j) = n(n-1)/\lambda(D)$, corresponding to uniform random point sampling in D (see Appendix A.1, p. 27). The Horwitz-Thompson and the Sen-Yates-Grundy variances are equivalent in this approximation, as are the values of their estimates [Särndal *et al.* , 2003, note 2.8.4 on p. 47]. The condition (36) applies to any pair of inventory points, thereby ruling out the possibility of obtaining negative variance estimates. A constant value of the pair density approximation used along with the usual CzNFI sample sizes in the considered domains are a guarantee of an acceptable stability of the variance estimators based on this theorem.

A.1 Application of the Horwitz-Thompson theorem to uniform random point sampling within the area of interest D

For uniform random sampling (URS) of a fixed number n of inventory points within the area of interest D, the density and pair-wise density obey Eqs. (37) and (38), respectively [see, e.g., Stevens & Olsen, 2003, first paragraph, p. 599].

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

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

Insertion in Eq. (28) gives formula (39) for the point estimate of the total $\hat{Y}$.

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

Formulas in Eq. (40) to (46) apply to the variance $\mathbb{V}\big(\hat{Y}_{URS}\big)$.

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

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

$$= \frac{\lambda^2(D)}{n} \Big\{ \mathbb{E}\big[Y^2(x)\big] - \mathbb{E}^2\big[Y(x)\big] \Big\} = \frac{\lambda^2(D)}{n} (\overline{Y^2} - \bar{Y}^2) \tag{42}$$

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

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

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

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

Formula (40) is obtained by substituting (37) and (38) in (29) (p. 24). The term in the brackets inside the second integral of (29) by using (37) and (38) is thereby reduced to $-n^{-1}$. In Eq. (41) the terms are multiplied by fractions which are equal to 1, which is made use of when rearranging to (42) and further to (43) by using the well-known calculation formula for the variance of a random quantity [see, e.g., Anděl, 1978, Eq.12 on p.14]. The second expression (43) contains the variance in the form of the mean value of the squares of the deviations $Y(x)$ from the mean value $\bar{Y}$, i.e. in the form of the second central moment. When rearranging to (44) (and formerly also when rearranging to the first expression of (42)), the property (de-facto definition) of the mean value of a continuous quantity is used [see, e.g., Anděl, 1978, Eq. 9, p. 13], where $f(x) = \lambda^{-1}(D)$ is the probability density of the continuous uniform distribution (for URS, the possibility of sampling an inventory point is identical in all parts of D). Formula (45) s obtained by simply cancelling out $\lambda(D)$ in the numerator of (44). The last formula specified, Eq. (46) uses the relation between the variance of a random quantity and the variance of the estimator of its mean value for uniform random sampling (selection of any arbitrary point

is fully independent of the previously sampled inventory points) [see, e.g., Anděl, 1978, proposition 11, formula 31, p.32].

A formula to estimate variance of $\hat{Y}$ is obtained by consecutive modifications (47) to (54) of the initial general formula (32) (p. 25) by inserting the densities (37) and (38).

$$\hat{\mathbb{V}}(\hat{Y}_{URS}) = \sum_{x\in s}\left[\frac{\lambda(D)Y(x)}{n}\right]^2 + \sum_{x_i\in s}\sum_{\substack{x_j\in s\\ x_i\neq x_j}}\frac{\lambda^2(D)Y(x_i)Y(x_j)}{n^2}$$
$$- \sum_{x_i\in s}\sum_{\substack{x_j\in s\\ x_i\neq x_j}}\frac{\lambda^2(D)Y(x_i)Y(x_j)}{n(n-1)} \tag{47}$$
$$= \frac{\lambda^2(D)}{n-1}\left\{(n-1)\sum_{x\in s}\left[\frac{Y(x)}{n}\right]^2\right.$$
$$\left. + (n-1)\sum_{x_i\in s}\sum_{\substack{x_j\in s\\ x_i\neq x_j}}\frac{Y(x_i)Y(x_j)}{n^2} - \sum_{x_i\in s}\sum_{\substack{x_j\in s\\ x_i\neq x_j}}\frac{Y(x_i)Y(x_j)}{n}\right\} \tag{48}$$
$$= \frac{\lambda^2(D)}{n-1}\left\{n\sum_{x\in s}\left[\frac{Y(x)}{n}\right]^2 - \sum_{x\in s}\left[\frac{Y(x)}{n}\right]^2\right.$$
$$+ \sum_{x_i\in s}\sum_{\substack{x_j\in s\\ x_i\neq x_j}}\frac{Y(x_i)Y(x_j)}{n} - \sum_{x_i\in s}\sum_{\substack{x_j\in s\\ x_i\neq x_j}}\frac{Y(x_i)Y(x_j)}{n^2}$$
$$\left. - \sum_{x_i\in s}\sum_{\substack{x_j\in s\\ x_i\neq x_j}}\frac{Y(x_i)Y(x_j)}{n}\right\} \tag{49}$$
$$= \frac{\lambda^2(D)}{n-1}\left\{\sum_{x\in s}\frac{Y^2(x)}{n} - \sum_{x_i\in s}\sum_{x_j\in s}\frac{Y(x_i)Y(x_j)}{n^2}\right\} \tag{50}$$
$$= \frac{\lambda^2(D)}{n-1}\left\{\sum_{x\in s}\frac{Y^2(x)}{n} - \sum_{x\in s}\frac{Y(x)}{n}\sum_{x\in s}\frac{Y(x)}{n}\right\} \tag{51}$$
$$= \frac{\lambda^2(D)}{(n-1)}\left\{\sum_{x\in s}\frac{Y^2(x)}{n} - 2\sum_{x\in s}\frac{Y(x)}{n}\hat{\hat{Y}} + \hat{\hat{Y}}^2\right\} \tag{52}$$
$$= \frac{\lambda^2(D)}{(n-1)}\sum_{x\in s}\frac{Y^2(x) - 2Y(x)\hat{\hat{Y}} + \hat{\hat{Y}}^2}{n} \tag{53}$$
$$= \frac{\lambda^2(D)}{n(n-1)}\sum_{x\in s}\left[Y(x) - \hat{\hat{Y}}\right]^2 \tag{54}$$

The modifications up to Eq. (49) are basic algebraic operations only. Formula (50) results from the third and fifth terms cancelling each other out and the second and fourth terms in Eq. (49) being combined into one double-sum (see indication of the second sum over all elements in the sample within Eq. (50) versus the forth term in Eq. (49) with the second sum leaving out the current element of the first sum). Modification of the second term in Eq. (51) is trivial, it is shown here only to emphasise the agreement of the two factors with the mean value estimator $\hat{\hat{Y}} = \hat{Y}\lambda(D)^{-1}$. In Eq. (51) the term in

the braces is modified by adding and subtracting $\hat{\bar{Y}}^2$. Rearrangement to Eq. (52) uses the equality $\sum_{x \in s} \hat{\bar{Y}} = n\hat{\bar{Y}}$, where the summation is over n elements of sample s and the mean $\hat{\bar{Y}}$ is a constant here. Eq. (53) is attained by simple rearrangement of Eq. (52) by factoring out n^{-1} from the sum and expressing the numerator as the square root of the difference of two numbers.

B Variance estimation based on Taylor's approximation

Some CzNFI target parameters are given by the nonlinear function (55) of two estimated parameters (random quantities), denoted Y_1, Y_2.

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

The Y_1 and Y_2 values are unknown and are estimated within the survey. The $Y_{1,2}$ value is estimated as $\hat{Y}_{1,2}$ according to (56).

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

Analytical expression of variance $\mathbb{V}(\hat{Y}_{1,2})$ of the estimator $\hat{Y}_{1,2}$ as a nonlinear function of random quantities is usually a complex issue. Therefore, the Taylor's series approach is used with a view to approximating $\hat{Y}_{1,2}$ by a linear function whose variance is given by Eq. (62). Approximation of $\hat{Y}_{1,2}$ in the neighbourhood of (Y_1, Y_2) neglecting the second and higher order terms is expressed by Eq. (57) [Jarník, 1963, Chapter XII, par. 1, p. 289], [Särndal *et al.* , 2003, Chapter 5.5, formulas 5.5.7, 5.5.8, pp. 173 to 174].

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

$$a = \left.\frac{\partial f}{\partial \hat{Y}_1}\right|_{Y_1, Y_2} \quad \ldots \text{partial derivation of } f \text{ with respect to } \hat{Y}_1 \text{ in point } (Y_1, Y_2) \tag{58}$$

$$b = \left.\frac{\partial f}{\partial \hat{Y}_2}\right|_{Y_1, Y_2} \quad \ldots \text{partial derivation of } f \text{ with respect to } \hat{Y}_2 \text{ in point } (Y_1, Y_2) \tag{59}$$

By using Eq. (57) the variance of point estimator $\hat{Y}_{1,2}$ is approximated by variance of $\hat{Y}_{1,2}^{(apx)}$, see Eq. (60). Eq. (61) is used to also approximate the variance estimator $\hat{\mathbb{V}}(\hat{Y}_{1,2})$.

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

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

Owing to the linear nature of Eq. (57) the variance of $\hat{Y}_{1,2}^{(apx)}$ is given by Eq. (62) [Anděl, 1978, proposition 4, formula 16 on pp. 28 and 29, proposition 2, formula 14, p. 14], [Särndal *et al.* , 2003, Chapter 5.5, formula 5.5.3, p. 172].

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

The values of the partial derivatives a (58) and b (59) are unknown because we do not know the precise values of the parameters Y_1 and Y_2, i.e. the position of the point in which the partial derivatives are to be evaluated. Therefore, the variance estimator according to Eq. (63) uses $\hat{a}$ and $\hat{b}$, i.e. partial derivatives of $\hat{Y}_{1,2}$ at point $(\hat{Y}_1, \hat{Y}_2)$, which is justified (for adequately large samples) by the fact that $\hat{a}$ and $\hat{b}$ are consistent estimators of a and b provided that $\hat{Y}_1$ and $\hat{Y}_2$ are consistent estimators of Y_1 and Y_2 [Särndal *et al.* , 2003, Chapter 5.5, last paragraph on p. 174].

$$\hat{\mathbb{V}}\left(\hat{Y}_{1,2}^{(apx)}\right) \approx \hat{a}^2\hat{\mathbb{V}}(\hat{Y}_1) + \hat{b}^2\hat{\mathbb{V}}(\hat{Y}_2) + 2\hat{a}\hat{b}\hat{\mathbb{C}}(\hat{Y}_1, \hat{Y}_2) \tag{63}$$

B.1 Variance of a ratio, general solution using Taylor's approximation

A number of CzNFI target parameters can be expressed as the $R_{1,2}$ ratio of the totals Y_1 and Y_2 of the equivalently mean values $\bar{Y}_1$ and $\bar{Y}_1$ according to Eq. (64). A typical example is the per hectare growing stock, i.e. the ratio of the total growing stock to the size of the *stand area* (one of the three subclasses within the *Forest* category). The $\hat{R}_{1,2}$ estimator is obtained as the ratio of the estimators $\hat{Y}_1$ and $\hat{Y}_2$ or equivalently as the ratio of the estimators of the mean values $\hat{\bar{Y}}_1$ and $\hat{\bar{Y}}_1$, see Eqs. (64) and (65).

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

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

Variance of $R_{1,2}$ as a non-linear function can be expressed and estimated by using Taylor's approximation (see Section B). The partial derivatives (58), (59) and their estimators are given for the function of the ratio by Eqs. (66) and (67).

$$a = \frac{1}{Y_2} \qquad b = -\frac{Y_1}{Y_2^2} = -\frac{R_{1,2}}{Y_2} \tag{66}$$

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

The $R_{1,2}$ ratio is approximated with a Taylor series neglecting the 2nd and higher order terms (see Eq. (68)).

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

Based on Eqs. (57) and (66), approximation $\hat{R}_{1,2}^{(apx)}$ can be written in the form of Eq. (69). Formula (72) applies generally irrespective of the type of the $\hat{Y}_1$ and $\hat{Y}_2$ estimators.

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

$$= R_{1,2} + \frac{\hat{Y}_1}{Y_2} - \frac{R_{1,2}\hat{Y}_2}{Y_2} + \left[\frac{R_{1,2}\hat{Y}_2}{Y_2} - \frac{Y_1}{Y_2}\right] - \frac{Y_1}{Y_2} + \left[\frac{Y_1}{Y_2} - \frac{R_{1,2}\hat{Y}_2}{Y_2}\right] + \frac{R_{1,2}Y_2}{Y_2} \tag{70}$$

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

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

Additional modifications (73) to (75) are made for the case where both $\hat{Y}_1$ and $\hat{Y}_2$ are one-phase estimators.

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

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

$$= R_{1,2} + \frac{1}{Y_2}\hat{Z} \tag{75}$$

The quantity $Z(x)$ is a residual variable defined by Eq. (76) while $\hat{Z}$ in Eq. (77) denotes the one-phase estimator (28) of the total of $Z(x)$ within the area D.

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

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

The mean value of the estimator $\hat{\bar{Z}}$ is zero, as is clear from Eq. (78).

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

The variance $\mathbb{V}\left(\hat{R}_{1,2}\right)$ is approximated by $\mathbb{V}\left(\hat{R}_{1,2}^{(apx)}\right)$ according to Eq. (79) obtained as the variance of (75), i.e. variance of the sum of the constant $R_{1,2}$ and the product of the constant Y_2^{-1} and the estimator of the total $\hat{Z}$ [Anděl, 1978, application of proposition 2, formula 14, p. 14].

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

Since the true value of the $R_{1,2}$ ratio is unknown, the $Z_\circ(x)$ and $\hat{Z}_\circ$ approximations according to Eqs. (80) and (81) are used instead of $Z(x)$ and $\hat{Z}$ respectively, to calculate $\hat{\mathbb{V}}\left(\hat{R}_{1,2}^{(apx)}\right)$. The substitution of $R_{1,2}$ by the $\hat{R}_{1,2}$ is neglected for adequately large samples in view of the consistency of $\hat{R}_{1,2}$, following from the consistency of $\hat{Y}_1$ and $\hat{Y}_2$.

$$Z_\circ(x) = Y_1(x) - \hat{R}_{1,2}Y_2(x) \tag{80}$$

$$\hat{Z}_\circ = \sum_{x\in s}\frac{Y_1(x) - \hat{R}_{1,2}Y_2(x)}{\pi(x)} = \sum_{x\in s}\frac{Z_\circ(x)}{\pi(x)} = \lambda(D)\hat{\bar{Z}}_\circ \tag{81}$$

$$\hat{Z} = \lambda(D)\hat{\bar{Z}} \approx \hat{Z}_\circ = \lambda(D)\hat{\bar{Z}}_\circ \tag{82}$$

Eq. (83) is the resulting approximative variance estimator of the ratio.

$$\hat{\mathbb{V}}\left(\hat{R}_{1,2}\right) \approx \hat{\mathbb{V}}\left(\hat{R}_{1,2}^{(apx)}\right) \approx \frac{1}{\hat{Y}_2^2}\hat{\mathbb{V}}\left(\hat{Z}_\circ\right) \tag{83}$$

The $\hat{\mathbb{V}}\left(\hat{Z}_\circ\right)$ variance estimator is obtained according to Eqs. (31), (32) or, in some cases, more appropriately according to Eqs. (34), (35).

Analogously, the variance estimator of the ratio of two one-phase estimators can be expressed by means of Eq. (84) [Särndal *et al.* , 2003, details of the derivation in Section 5.6 on p. 176],

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

where $\hat{\mathbb{V}}\big(\hat{Y}_1\big)$, $\hat{\mathbb{V}}\big(\hat{Y}_2\big)$ are variance estimators of the one-phase totals in the numerator $(\hat{Y}_1)$ and denominator $(\hat{Y}_2)$ of the ratio estimator $\hat{R}_{1,2}$ and $\hat{\mathbb{C}}\big(\hat{Y}_1,\hat{Y}_2\big)$ is the estimator of covariance of the two totals. Eq. (84) follows directly from Eq. (63) (Section B, p. 30) in which $\hat{a}$ and $\hat{b}$ are inserted in accordance with Eq. (67).

B.2 Variance of the ratio for uniform random point sampling

As in the general case, formulas (64) apply to the ratio of the two totals and to its estimator also in the case of uniform random sampling of a fixed number of n inventory points within the area of interest D (see Section B.1, p. 31). Approximation (85) with Taylor's series is convenient when estimating the variance $\hat{R}_{1,2}$ (see Section B.1 p. 31).

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

Eqs. (86) to (91) apply for $\mathbb{V}_{URS}\Big(\hat{R}_{1,2}^{(apx)}\Big)$.

$$\begin{aligned}
\mathbb{V}_{URS}\Big(\hat{R}_{1,2}^{(apx)}\Big) &= \frac{1}{Y^2}\mathbb{V}_{URS}\big(\hat{Z}\big) && (86)\\
&= \frac{\lambda^2(D)}{Y^2}\mathbb{V}_{URS}\big(\hat{\bar{Z}}\big) && (87)\\
&= \frac{\lambda^2(D)}{n\,Y^2}\mathbb{V}_{URS}\big[Z(x)\big] && (88)\\
&= \frac{\lambda(D)}{n\,Y_2^2}\int_D \big[Z(x)-\bar{Z}\big]^2 dx && (89)\\
&= \frac{\lambda(D)}{n\,Y_2^2}\int_D Z^2(x)dx && (90)\\
&= \frac{\lambda(D)}{n\,Y_2^2}\int_D \big[Y_1(x)-R_{1,2}Y_2(x)\big]^2 dx && (91)
\end{aligned}$$

Eq. (86) is analogous to Eq. (79) in Section B.1 (p. 32). Eqs. (87) to (89) describe $\mathbb{V}_{URS}\big(\hat{Z}\big)$ analogously to Eqs. (46), (43) and (45) in Section A.1. Eq. (78) (Section B.1, p. 32) is used to eliminate the zero value of $\bar{Z}$ from the difference in the brackets in Eq. (89) and so to obtain Eq. (90). The final rearrangement to Eq. (90) is obtained by substituting Eq. (76) (Section B.1, p. 32) for $Z(x)$.

The variance estimator $\hat{\mathbb{V}}_{URS}\big(\hat{R}_{1,2}\big)$ is approximated according to Eq. (92) (Section B.1, p. 33). The resulting approximate formulas are shown in Eqs. (93) and (94).

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

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

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

Eq. (93) follows from Eq. (83) (Section B.1, p. 32) and Eq. (54) (Appendix A.1, p. 28) with the approximation $\hat{\tilde{Z}} \approx 0$ - see Eq. (78) (Section B.1, p. 32). Eq. (93) makes use of the equality $\hat{\tilde{Y}}_2 = \hat{Y}_2\lambda^{-1}(D)$.

C Aspects of NFI sampling related to sample size

The (quasi-)systematic and/or spatially stratified sampling that is typical for (national) forest inventories uses square or rectangular (or, less frequently, triangular or hexagonal) inventory grids with their origin selected at random (and, less frequently, with their orientation selected also at random). The grid divides the sampling frame $\mathcal{F} \supseteq D$ completely into congruent inventory blocks possessing the same size and shape, one bordering but not overlapping the other.

It is typical of NFI that the sampling frame is usually not identical with the area of interest D (the geographical domain). It can be defined by using special topology in a manner guaranteeing a constant sample size in $\mathcal{F}$ [Stevens, 1997, Section 3.1, p. 172], [Mandallaz, 2007, Section 5.6, p. 92]. This can be visualized so that a square-shaped or rectangular part of the plane $\mathbb{R}^2$ containing the domain D and the inventory grid itself (prior to a random displacement) is first coiled so that the opposite sides are interconnected along the whole length. The cylindrical surface (rotary cylinder) so obtained is transformed by connecting the side edges (edges of the opposite bases of the rotary cylinder). A torus whose surface contains the domain D along with the inventory grid is thereby obtained. It is ensured for any random inventory grid displacement vector that the grid will not leave the sampling frame $\mathcal{F}$ with a fixed size and location with respect to the domain D. When displaced at random, the inventory grid slides (figuratively speaking) over the torus surface.

Typically, estimates are required for a number of geographical domains creating hierarchic structures (NUTS units, arbitrarily defined geographical areas). Unless the natural conditions of the area examined are very diverse, the definition of the frame $\mathcal{F}$ is selected so that it encompasses the union of all the domains considered – this is the whole area of the country in the case of the CzNFI. In this concept the area of the country is the highest-level geographical domain considered rather than the sampling frame as such.

For systematic sampling, the inventory points are located in the centre (or at the line intersection points) whereas in spatially stratified sampling they are located entirely at random inside of each inventory block (spatial strata). In contrast to the sampling frame $\mathcal{F}$, the domain D cannot be (in the typical case) completely divided into mutually neighbouring and not overlapping congruent inventory blocks. Parts of some inventory blocks largely reach beyond the domain D. The result of a random displacement of the inventory grid with respect to D, or random generation of the inventory point positions (initial points of the tracts) within each block – is a variable number of inventory points inside geographical domain D. The sample size in D is non-constant among possible replications of the sampling process when the sampling designs considered here are applied.

The formulas for point estimators and their variances given in Section A apply, and the estimators are unbiased, for constant-size samples (see the second paragraph on p. 25). A fixed sample size can be ensured by formally identifying the area of interest D with the sampling frame $\mathcal{F}$ while redefining the local density function so that it is always zero when lying beyond D (but still within $\mathcal{F}$). So a state is reached where the total of any quantity within D equals the total of that quantity within $\mathcal{F}$. An unbiased estimator of the total within $\mathcal{F}$ is also an unbiased estimator of the total within the area of interest D. This variant, although relevant to some designs [Stevens, 1997], is not dealt with here in more detail, first because of the higher demands on the computations (the calculation of the estimates for the country and for its sub-regions encompasses all of the inventory points) and second, because of the further increase of conservativeness of the variance estimator based on the assumption of uniform random point sampling in $\mathcal{F}$. Nevertheless, putting

local densities equal to zero for all other attribute domains except the actually considered is a standard approach, see Section 2.3.

According to Cordy [1993, Section 3, p. 359], the non-constant sample sizes in D can be approached in two ways:

- by understanding the unbiased nature of the point estimator and its variance as conditional on the sample size (conditional inference);
- by substituting the sampling densities and pair-wise densities of inventory points (or reference points of tracts) by their expected values (over all the possible sample sizes that may be generated by the design); in this variant the estimators and their variance are unbiased unconditionally of the existing sample size (unconditional inference, Section E, p. 40)

The estimator that is unbiased conditionally on the actual sample size possesses more favourable statistical properties – lower variance, the expected bias conditionally on the sample size (conditional bias) is equal to zero. Since, however, the sampling weights used during the estimate calculations for the geographical domains (domain-specific weights) are different from those used in the calculation of the estimate for the whole, this estimator of total is not geographically additive.

Taking into account the additivity requirement for the estimates of the total (additivity definition is presented in Appendix F), the unconditionally unbiased estimators (the latter of the above two variants) were used for CzNFI2 calculations based on the CzNFI1 inventory grid data.

In view of the expected sample sizes within the country as well as within the subregions, the increase in the estimator's variance and its bias that is conditional on the difference between the actual and expected sample size are acceptably low. Moreover, the approximately systematic distribution of tracts within the CzNFI1 grid along with the relatively compact shape of the geographical domains reduce the variability in the number of the inventory points within the geographical domains considerably when compared to uniform random point sampling. The ratios of the standard deviation[12] of the inventory tracts in a geographical domain (expanded by the 300 m buffer zone) to the expected number of tracts in the geographical domain (coefficient of variation of the number of sample points in domain) range from 0.28 % for the South-Bohemian region to 1.59 % for the region of the Capital Prague. The coefficient of variation at the level of the Czech Republic was estimated to 0.14 %. The shortcomings of the unconditionally unbiased estimator thus become reasonably negligible in the particular case of CzNFI.

[12] Estimated based on simulated samples for the CzNFI1 sampling grid.

D Variance estimation in a typical NFI setting

Uniform random point sampling within the area of interest D is usually not as suitable as the systematic and spatially stratified sampling approaches, ensuring more uniform coverage of D with inventory points [Cochran, 1977, Chapters 8.9, 8.10 and 8.13, pp. 219, 221, 227], [Cordy & Thompson, 1995, first paragraph on p. 174], [Stevens, 1997, Section 3, p. 171], [Barabesi, 2003, second paragraph on p. 356], [Ripley, 2004, Chapter 3, p. 19], [Heikkinen, 2006, Chapter 10, p. 155], [Gregoire & Valentine, 2008, Section 3.2.2, p. 49].

Due to the frequently occurring positive spatial correlation, similar values of the quantity examined can be expected at not very distant inventory points. Sampling of relatively close inventory points brings about reduction in the amount of information that might be gained if more different values were observed at points that are more distant in the average. This is why effective sampling designs restrict in some manner ($\pi(x_i, x_j) \to 0$) occurrence of very close sample points $x_i,\ x_j \in D$, or even directly rule out the possibility of sampling close inventory points. This ensures a higher local density variability at the inventory points within the sample and, at the same time, reduces the variability of the point estimates between the samples, i.e., improves the accuracy of the target parameter estimator.

In the extreme (but relatively frequent) case, condition (30) (Section A, p. 24) fails to be satisfied. Neither the Horwitz-Thompson nor the Sen-Yates-Grundy formulas are then applicable to estimate variance unbiasedly, and so, appropriate alternatives must be sought. This limitation is typically encountered with systematic and spatially stratified samples with one point in each of the fixed strata – tessellation stratified sampling, TS [Cordy & Thompson, 1995], or congruent tessellation stratified sampling, CTS [Stevens, 1997], or TSS [Barabesi & Franceschi, 2010] or in the case of an CzNFI1 inventory grid derived from systematic sampling (quasi-systematic sampling).

An unbiased nature of the estimator is usually not a strict requirement. A number of estimators (not only of variance) are used in practice, which, although not unbiased, have other assets, e.g. lower variance or mean square error [Anděl, 1978, Chapter XV.1, p. 252]. Asymptotically unbiased and consistent estimators hold a privileged position among all types of estimators. With some simplification, the expected value of the estimator approaches the true value of the parameter in the population as the sample size is increased (for a not fully adopted definition see [Rao, 1965, Section 5c.1, p. 344], [Cochran, 1977, Chapter 2.4, p. 21], [Särndal *et al.* , 2003, Chapter 5.3, p 166], [Gregoire & Valentine, 2008, Section 2.2.1, p. 19]).

If unbiased, consistent variance estimators suitable for a specific sampling design and population properties do not exist or are not known, conservative estimators overestimating the true variance can sometimes be used. Confidence intervals constructed based on a conservative variance estimator (which corresponds to the point estimator's probability distribution if a correct procedure is applied) will include the true value of the parameter in question with a probability higher than nominal ($1 - \alpha$, among hypothetical or real repetitions of sampling process). Hence, the statistical certainty of such conservative confidence interval is higher than nominal.

From the application point of view, it is ideal if the unbiased, consistent (and conservative) nature of the variance estimator follows from the design and is not conditional on the properties of the population, i.e. of the local density function $Y(x)$. Unfortunately, this situation is either impossible (unbiased estimators) or rather rare for designs that do not satisfy condition (30) (Section A, p. 24). It is still convenient, though, if some of the favourable properties of the estimator at least is conditional on such local density prop-

erties as can actually be expected for the population in question – e.g. the assumption of positive spatial correlation of local density in environmental sampling.

The approximation $\pi(x_i, x_j) = n(n-1)/\lambda(D)$ corresponding to uniform random sampling of the inventory points in D, see (38) (Appendix A.1, p. 27), is conventionally used where the sampling does not satisfy condition (30). The variance estimator of the population's parameter θ obtained by using this approximation is conservative in the majority of NFI cases[13] (positive spatial local density correlation) i.e. we have

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

[13]But not necessarily - e.g. when periodicity of local density function exists and it interferes with the spacing of sampling grid.

E Unconditionally unbiased total estimator

This approach to the non-constant sample size in D is based on the knowledge or approximation of the expected sample density $\bar{\pi}(x)$ at point x and of the density $\bar{\pi}(x_i, x_j)$, which will replace the densities $\pi(x)$ and $\pi(x_i, x_j)$ in Eqs. (28), (29), (31) and (32), see Cordy [1993, Section 3, p. 359].

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

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

The sums in Eqs. (96) and (97) run over all the possible sample sizes and α_{n_D} is the probability that a sample whose size is n_D (which means that precisely n_D inventory points fall within D) is obtained during the implementation of the sampling design.

The expected $\bar{\pi}(x)$ value can sometimes also be obtained without explicitly knowing the probability α_{n_D}. For non-informative sampling designs where sampling of point x is absolutely independent of the density $Y(x)$ as well as of quantities correlating with $Y(x)$ we have

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

The term $\lambda(D)$ represents the size (area, Lebesgue measure) of the domain D. On factoring out the constant $\lambda^{-1}(D)$, (96) can be written as:

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

where $\bar{n}_D$ is the expected (mean) sample size. It will be clear that

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

where $\lambda(c)$ is the fixed inventory block size. If (100) is substituted in (99), $\lambda(D)$ is cancelled out. The expected point x sampling density can be obtained (as felt intuitively) according to Eq. (101) as the inversion of the fixed inventory block size:

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

Condition (30) (Section A) fails to be satisfied by the density $\pi_{n_D}(x_i, x_j)$ of some inventory grids that are frequently used in practice, and then the approximation for uniform random sampling is used. If the size of the sample in the area of interest D is non-constant, then the density $\bar{\pi}(x_i, x_j)$ can be approximated by

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

The derivation is based on a scheme where the pair density $\pi_{n_D}(x_i, x_j)$ according to Eq. (103) is considered for each hypothetically repeated sample of random size n_D.

$$\pi_{n_D}(x_i, x_j) = n_D(n_D - 1)\lambda^2(D) \tag{103}$$

The expected pair density $\bar{\pi}(x_i, x_j)$ (over all the possible sample sizes) can then be written as:

$$\bar{\pi}(x_i, x_j) \approx \mathbb{E}\left[\frac{n_D(n_D - 1)}{\lambda^2(D)}\right] = \frac{\mathbb{E}(n_D^2) - \mathbb{E}(n_D)}{\lambda^2(D)} = \frac{\mathbb{E}(n_D^2) - \bar{n}_D}{\lambda^2(D)}. \tag{104}$$

The expected value $E(n_D)$ is given by Eq. (100). The expected value $\mathbb{E}(n_D^2)$ is expressed by using Eq. (105) (see Anděl [1978, formula 12 on p. 14].

$$\mathbb{V}(n_D) = \mathbb{E}(n_D^2) - \mathbb{E}^2(n_D). \tag{105}$$

Although the variance $\mathbb{V}(n_D)$ of the number of the inventory points sampled in D unknown in general, it can be estimated very precisely for the areas of interest that are defined by an existing map by means of simulated sampling. This also applies to the expected value $\mathbb{E}(n_D^2)$. Approximation (102) is obtained by inserting in Eq. (104).

An approximate value of $\mathbb{E}(n_D^2)$ can also be derived based on Taylor's linearisation of the function n_D^2 in the neighbourhood of $\mathbb{E}(n_D^2)$ – see (106) and (107)

$$n_D^2 \approx \mathbb{E}(n_D^2) + 2\sqrt{\mathbb{E}(n_D^2)}\left\{n_D - \sqrt{\mathbb{E}(n_D^2)}\right\}, \tag{106}$$

$$\begin{aligned}\mathbb{E}(n_D^2) &\approx \mathbb{E}(n_D^2) + 2\sqrt{\mathbb{E}(n_D^2)}\mathbb{E}(n_D) - 2\mathbb{E}(n_D^2)\\ &\approx \mathbb{E}^2(n_D) = \bar{n}_D^2.\end{aligned} \tag{107}$$

The result corresponds to the solution of Eq. (105) where the variance $\mathbb{V}(n_D)$ is neglected. The approximation $\bar{n}_D^2$ underestimates $\mathbb{E}(n_D^2)$ by the $\mathbb{V}(n_D)$ variance value. Substitution of $\bar{n}_D^2$ for $\mathbb{E}(n_D^2)$ in Eq. (104) leads to the approximation (108)

$$\bar{\pi}(x_i, x_j) \approx \frac{\bar{n}_D(\bar{n}_D - 1)}{\lambda^2(D)}. \tag{108}$$

In view of the relatively large CzNFI samples and of the use of inventory grids that cover the areas of interest uniformly, it seems to be the straightforward approach to disregard the effect of underestimation of $\mathbb{E}(n_D^2)$ This, however, would result in unknown reduction of the conservativeness of the variance estimator to below the level that would correspond to the substitution of the pair density by densities for random sampling of n_D points itself. Therefore, we use the pair density approximation (102) when evaluating the CzNFI by this method while estimating the variance $\mathbb{V}(n_D)$ by means of simulated sampling.

The formulas derived for uniform random sampling of a fixed number of n points in D (see Appendix A.1 p. 27) are modified as follows based on Eqs. (99) and (102) to gain a form taking into account the non-constant sample size:

$$\hat{Y} = \sum_{\substack{x\in s\\ x\in D}} \frac{Y(x)}{\bar{\pi}(x)^{x\in s}} = \lambda(D)\frac{\sum_{x\in s} Y(x)}{\bar{n}_D} = \lambda(c)\sum_{x\in s} Y(x), \tag{109}$$

$$\hat{\mathbb{V}}\big(\hat{Y}\big) = \frac{\lambda^2(D)}{\bar{n}_D - 1 + \hat{\mathbb{V}}(n_D)/\bar{n}_D}\left\{\widehat{\widehat{Y^2}} + \hat{\hat{Y}}^2\Big[\hat{\mathbb{V}}(n_D)/\bar{n}_D - 1\Big]\right\}. \tag{110}$$

Eq. (110) is obtained by using Eqs. (99) and (102) when deriving the variance estimator in Eqs. (47) to (51) (Appendix A.1, p. 27).

The term $\widehat{\widehat{Y^2}}$ is the ratio of the sum of squares of local densities to the expected sample size $\bar{n}_D$, whereas $\hat{\hat{Y}}^2$ is the square of the ratio of the sum of tract level local densities to the $\bar{n}_D$.

$$\mathbb{E}\big(\hat{Y}\big) = Y \tag{111}$$

The variance estimator, when applying approximation 102 or a (quasi-)systematic or spatially stratified grid and assuming positive spatial correlation of the local density values, is conservative unconditionally of the sample size in D, in other words, we have

$$\mathbb{E}\Big[\hat{\mathbb{V}}\big(\hat{Y}\big)\Big] > \mathbb{V}\big(\hat{Y}\big). \tag{112}$$

Estimators that are unbiased unconditionally of the size of the sample in D are appropriate especially where the sampling designs are to be compared with respect to the expected precision. In fact, this problem cannot be solved by starting from a particular size of the sample in D because this may not occur at all during the actual future sampling. As regards evaluation of existing samples, such estimators are appropriate where additivity of the estimates is expected or required (see Appendix F, p. 44).

The statement on the unbiased nature of the point estimator of the total does not apply to a specific sample size $n_D \neq \bar{n}_D$ which is a drawback compared to the estimators that are unbiased (conservative) conditionally on the sample size. This drawback, however, is theoretical only in the conditions related to the CzNFI1 sampling grid (low variability of sample size in D, strong conservativeness of the variance estimator) – the bias and variance increase levels are practically negligible. Other comments concerning the CzNFI can be found in the concluding paragraphs of Appendix C, p. 36

F Geographical and attribute additivity of an estimator

The geographical additivity of an estimator is expressed by Eq. (113)

$$\hat{\theta}_{\bigcup_{i=1}^{k} D_i} = \sum_{i=1}^{k} \hat{\theta}_{D_i}, \tag{113}$$

according to which the estimate calculated for any geographical domain, which is the union $\bigcup_{i=1}^{k} D_i$ of the sub-domains $D_1, D_2 \ldots D_k$ is precisely equal to the sum of estimates calculated individually for each of the k sub-domains.

Attribute additivity is defined analogously:

$$\hat{\theta}_{\bigcup_{j=1}^{J} A_j} = \sum_{j=1}^{j} \hat{\theta}_{A_j}, \tag{114}$$

where A_j represent the attribute domains (e.g. tree species groups, type of forest stand).

The total estimator proposed by this methodology is geographically as well ass attribute-additive. Disturbance of the attribute-additivity of the ratio estimator is a logical consequence of attribute categorisation of the estimator in the denominator – additivity is only met within the attribute categorisation subgroups defined by the denominator's common attribute domain. Example: timber stock per hectare of stand area categorised by the tree species (numerator of the ratio) and by land tenure category (denominator of the ratio), where the estimators by the tree species are additive only within each ownership category but not across of them.

Additivity is usually preferred from the NFI result user's point of view. An asset of additive estimators is in the fact that check mechanisms of the evaluation process can be applied. If the resulting estimates are found not to be geographically additive or attribute-additive, this implies that errors must have occurred during the data preparation process or during the calculation itself. Additivity check is very useful when large numbers of geographical and attribute domains are evaluated (which is the case with the CzNFI).

G Confidence interval and its statistical certainty

The statistical certainty of a confidence interval $CI = [CI^-; CI^+]$ is identical with probability P in Eq. (115), where CI^- and CI^+ are the lower and upper confidence limits, respectively. The design-based approach defines this probability as the expected proportion of the confidence intervals which, if the sampling is repeated with the same design (one confidence interval being constructed based on each of the replicated samples), encompass the true value of the unknown parameter θ of the population. In general, it is required that the true statistical certainty of the confidence interval corresponds to the chosen nominal value $1 - \alpha$, where α (level of significance) is the proportion of the confidence intervals that do not cover θ - the true value of the population's parameter to be estimated.

$$P(CI^- \leqslant \theta \leqslant CI^+) = 1 - \alpha \tag{115}$$

Confidence intervals are constructed by this methodology as follows:

$$CI = \hat{\theta} \pm z_{(1-\alpha/2)} \sqrt{\hat{\mathbb{V}}(\hat{\theta})}, \tag{116}$$

where $\hat{\theta}$ is the point estimator of the unknown population parameter θ, $\hat{\mathbb{V}}(\hat{\theta})$ is the estimator of variance $\mathbb{V}(\hat{\theta})$ and $z_{(1-\alpha/2)}$ is the $(1 - \alpha/2)$ quantile of the standardised normal distribution. According to Särndal et al. [2003, Chapter 2.11, p. 56] the confidence interval (116) will contain the unknown value of the parameter θ for approximately $(1-\alpha)\%$ percent cases of sample repetition if the following conditions are met::

1. Probability distribution for the point estimator $\hat{\theta}$ is approximately normal, with the expected (mean) value θ and variance $\mathbb{V}(\hat{\theta})$.

2. A consistent estimator $\hat{\mathbb{V}}(\hat{\theta})$ of variance $\mathbb{V}(\hat{\theta})$ exists.

The first condition requires that the ratio (117) below has approximately standardised normal distribution $\mathcal{N}(0, 1)$

$$\frac{\hat{\theta} - \theta}{\sqrt{\mathbb{V}(\hat{\theta})}}. \tag{117}$$

An approximation of the probability distribution of the point estimator by normal distribution can be justified based on validity of the central limit theorem. In certain, very general conditions, the sums of independent random quantities possess asymptotically normal distribution [Rao, 1965, Section 2c.5, p. 126], [Anděl, 1978, Chapter X.3, p. 184]. The point estimator (28) and many other estimators are frequently such sums, and so their distribution for adequately large samples can be approximated by normal distribution. However, no generally valid rule exists based on which one can decide if a sample is large enough for use of the normal distribution approximation [Cochran, 1977, Chapter 2.15, p. 39], [Särndal *et al.* , 2003, notes 2.11.2 and 2.11.3 on p. 57], [Gregoire & Valentine, 2008, Chapter 2.3, p. 31]. Cochran [1977, Chapter 2.15, p. 42] suggests that crude identification of cases where one cannot rely on normal approximation can be made by using the approximate relation (118):

$$n_{min} > 25 G_1^2, \tag{118}$$

where G_1 is Fisher's measure of skewness and n_{min} is the minimum required sample size. The procedure is applicable to populations where the deviation from normality is primarily due to the distribution asymmetry. According to Cochran, Eq. (118) is designed so that the 95% confidence intervals do not miss the true mean value in more than 6 % cases.

This criterion was tested by simulated estimation of the total and of the per hectare growing stock, respectively, and of the total forest area. Inventory points based on the CzNFI1 grid design were newly generated for each of the simulations (2000 replications of the CzNFI1 sampling grid in total). The first model population was the timber stock based on the database of forest management plans covering the entire area of the Czech Republic (approximately 2.7 million hectares of forest land). The other model population was a map of occurrence of land in the *Forest* category, obtained within the first CzNFI cycle by digitisation of orthophoto maps produced by aerial survey (OLIL map, also about 2.7 million hectares of area within the Czech Republic, 96% thematic accuracy). The estimates were calculated at the levels of the Czech Republic as a whole and of its administrative regions and districts, and also for the "conifers" and "broad-leaved trees" categories where growing stock was involved.

Where an adequately large sample size was indicated by criterion (118) the observed statistical certainty was never lower than the minimum of 94 % (declared by Cochran) for any of the combinations of the parameter assessed and the geographical/attribute domain. An inadequately low sample size was detected in districts where the attribute domains in question were little abundant, such as the estimate of the total growing stock of conifers in the Břeclav administrative district or the estimate of the stock of broad-leaved trees in the Žďár nad Sázavou administrative district (NUTS3 level). Statistical certainty, though, did not drop below the 94 % limit in over 80 % of cases where an inadequately low sample size was detected by the mentioned criterion. We attribute this fact to the conservativeness (upward bias) of the variance estimator, see Appendix C.

Where doubts exist as to the appropriateness of the procedure using normal approximations, an alternative approach is available, based on local density transformation performed with a view to making the distribution closer to normal distribution. Once a suitable local density transformation has been performed, the point and confidence intervals can be calculated and then transformed back for them to apply to the initial scale of the variable.

The topic of transformations to normal distribution and their use for constructing confidence intervals can be found in the monograph by Meloun and Militký [1998, Section 2.1.4, p. 85, and Section 2.1.5, p. 90]. Sakia [1992] published an overview devoted to Box-Cox transformations, which are among the most frequently used transformations. Li [2008] discussed the use of Box-Cox transformations in conjunction with the generalised regression estimator.

Estimates calculated based on the transformed quantities are not additive in domains among which various local density transformations have been performed (this concerns both geographical and attribute additivity, see Appendix F). This is why transformations are not routinely used during NFI evaluations.

An alternative that has been proposed by Gregoire & Valentine [2008] consists in the construction of a conservative confidence interval based on Eq. (119) whose basis is in Chebyshev's theorem:

$$P\left[\hat{\theta} - k\sqrt{\mathbb{V}(\hat{\theta})} \leqslant \theta \leqslant \hat{\theta} + k\sqrt{\mathbb{V}(\hat{\theta})}\right] \geqslant 1 - \frac{1}{k^2} \,. \tag{119}$$

The upper and lower limits of the confidence interval are calculated by using Eq. (120),

which, like Eq. (116), uses the estimator $\hat{\mathbb{V}}(\hat{\theta})$ in place of the true variance $\mathbb{V}(\hat{\theta})$.

$$CI_t = \hat{\theta} \pm k\sqrt{\hat{\mathbb{V}}(\hat{\theta})} \ . \tag{120}$$

This construction of the confidence interval does not require any assumptions regarding the probability distribution of the point estimator to be satisfied. When comparing the value of $k \doteq 4.47$ approximately corresponding to a 95% statistical certainty ($\alpha = 0.05$) with the quantile of the standardised normal distribution $z_{1-\alpha/2} \doteq 1.96$ used in Eq. (116), it is evident at first glance that the price to be paid for the general nature of the solution is appreciably wider confidence intervals.

Somewhat narrower confidence intervals can be determined based on the Vysochanskij-Petunin inequality:

$$P\left[\hat{\theta} - k\sqrt{\mathbb{V}(\hat{\theta})} \leqslant \theta \leqslant \hat{\theta} + k\sqrt{\mathbb{V}(\hat{\theta})}\right] \geqslant 1 - \frac{2}{3k^2} \ , \tag{121}$$

where the upper and lower confidence interval limits are also calculated from Eq. (120), in which, however, the appropriate k value (2.98 for a 95% statistical certainty) is inserted. The width of the confidence interval is at the level of two-thirds of the width of the interval constructed based on Chebyshev's theorem. This approach is only applicable to parameters for which uni-modal probability distribution with a finite variance of the point estimator can be assumed. We suppose that such properties exist in the majority, if not all, of the NFI target parameters and their estimators.

A common drawback of the approaches based on Chebyshev's theorem and on the Vysochanskij-Petunin inequality is the insertion of the variance estimate in place of the true variance into Eq. (120), the true variance being unknown in practical NFI settings. The required statistical certainty is not guaranteed during potential underestimation of the variance. In fact, however, the NFI variance estimators (for the one-phase estimators of the total in particular) are usually overestimating the true variance of the point estimator (see Appendix D). We also assume that they are consistent (see later). On the other hand, typical CzNFI sample sizes in typical geographical domains (the whole country or NUTS3 administrative regions) are at the level of hundreds to a few thousand inventory points. So we assume that the confidence intervals based on the normal approximation (116) are adequate for the CzNFI and the associated target parameters and estimators.

From the first condition formulated on p. 46 it also follows that the estimator $\hat{\theta}$ should be unbiased. A biased estimator means that a non-zero difference exists between the expected $\mathbb{E}(\hat{\theta})$ and true θ values of the population parameter for the specific sample size. This difference is the bias [Anděl, 1978, Chapter XV.1, p. 254]. Särndal *et al.* [2003, Chapter 5.2, p. 163] quantified the effect of bias of a point estimator on the true statistical certainty of the confidence interval (116). The authors found that for nominal statistical certainty $(1 - \alpha) = 0.95$ and a 10 % bias relative to the point estimator's standard deviation, the true statistical certainty is 0.9489. The true statistical certainty does not decrease below 0.9210 even if the bias is rather large, between 10 % and 50 % of the point estimator's standard deviation. A similar overview had been published before by Cochran [1977, Chapter 1.8, p. 12]. Since the estimator (28) is unbiased, bias can occur in its use only due to some systematic errors of the sampling and/or other deficiencies of the data collection and processing system.

The other condition, i.e. consistency of the variance estimator, is necessary because the true value of the variance $\mathbb{V}(\hat{\theta})$ is usually unknown when constructing the confidence interval. Hence, we depend on the approximation by normal distribution although the

estimator $\hat{\mathbb{V}}(\hat{\theta})$ is used in the denominator of (117) in place of the true variance $\mathbb{V}(\hat{\theta})$. Särndal *et al.* [2003, Chapter 2.11, p. 56] explain the significance of consistency by means of Eq. (122).

$$\frac{\hat{\theta}-\theta}{\sqrt{\hat{\mathbb{V}}(\hat{\theta})}} = \frac{\hat{\theta}-\theta}{\sqrt{\mathbb{V}(\hat{\theta})}}\left[\frac{\mathbb{V}(\hat{\theta})}{\hat{\mathbb{V}}(\hat{\theta})}\right]^{1/2} \tag{122}$$

When assuming consistency of $\hat{\mathbb{V}}(\hat{\theta})$, the ratio $\left[\mathbb{V}(\hat{\theta})/\hat{\mathbb{V}}(\hat{\theta})\right]^{1/2}$ approaches unity with a high probability for adequately large samples. This tendency is also stressed by the square root in Eq. (122). The error arising from the substitution of the variance $\mathbb{V}(\hat{\theta})$ by its estimator $\hat{\mathbb{V}}(\hat{\theta})$ is then neglected for adequately large samples. Although unbiased estimators are frequently consistent as well, this is not a rule. Therefore, one should remember that the mere fact that a variance estimator is unbiased is not a sufficient condition for constructing the confidence interval according to relation (116), the statistical certainty of which should correspond to the nominal value of $(1-\alpha)$.

Statistical certainty is closely related to the conservativeness of the variance estimator. This is frequently expected just based on certain assumptions regarding the population, which, however, may actually not be met. Depending on some population properties that are not always fully known in practice, the selected variance estimator may be too conservative or, on the contrary, too anti-conservative, of which we may not be fully aware when interpreting the result. In the former case, the confidence interval obtained is too wide compared to the width corresponding to the nominal statistical certainty $(1-\alpha)$ – the calculated target parameter precision is lower than the true precision. In the latter case, on the contrary, the confidence interval width is insufficient – the true statistical certainty does not reach the nominal value $(1-\alpha)$ and the target parameter estimator seems to be more precise than it actually is. Sometimes it may be more appropriate to use a slightly anti conservative variance estimator than to use a strongly conservative estimator – the choice usually depends on the application, on whether the assumptions under which the particular estimator is (anti)conservative are or are not met, and, of course, on additional properties of the variance estimator (guarantee of positive values, stability).