nFIESTA (new Forest Inventory ESTimation and Analysis)

Estimation methods

Radim Adolt[*], Jiří Fejfar[*], and Adrian Lanz[†]

[*] *Forest Management Institute Brandýs nad Labem (ÚHÚL). Czech Republic*
[†] *Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf. Switzerland*

Task T2.3.1 Common data analysis and output delivery system

---

updated on February 6, 2019

Contents

1 Introduction

In this document estimation methods implemented by the nFIESTA software are described. These methods were used within the Diabolo task T2.3.1 case study which explored possible effects of using selected Copernicus remote sensing products (Forest Type and Tree Cover Density maps, version 2012) as a source of auxiliary information to improve accuracy of biomass estimates. The case study concerned four European countries - Czech Republic, France, Germany and Switzerland.

In the second chapter the attention is paid to the Horvitz-Thompson theorem for continuous populations (HTC) [Cordy, 1993]. The use of this theorem can be considered as an innovative part of the task T2.3.1 of the Diabolo project and also the nFIESTA approach. It addresses theoretical issues arising from the integration of NFI (National Forest Inventory) plot data collected by many countries which use different sampling designs. This integration is necessary in order to produce estimates of forest attributes for arbitrary study areas possibly encroaching boundaries between the countries (or sampling strata). This generic approach can be applied virtually anywhere i.e. its use is not restricted to Europe only.

A special section on the derivation of so called inclusion (or sampling) densities follows immediately after the HTC introduction. Inclusion densities are indispensable for the HTC theorem application. They play the same role as the inclusion probabilities do in the Horvitz-Thompson theorem for finite populations i.e. they determine the sampling design itself as well as statistical properties of the estimators in terms of unbiasedness and precision (design-based variance). NFIs of particular countries can provide plot data with relative sampling weights attached. A method how these relative weights can be translated to inclusion densities has been proposed and implemented in nFIESTA. This method allows for an unbiased estimation of totals, ratios (of totals) and also their variances. The HTC theorem can be used for parameter estimation in case of non-uniform sampling designs e.g. the importance sampling [Gregoire & Valentine, 2008, p. 94], which was chosen as the nFIESTA's implicit approximation of any non-uniform sampling design.

The third chapter describes how target parameters defined as totals of an attribute over a study area (the estimation cell D) are calculated by nFIESTA. Estimation cells are arbitrarily shaped regions which are entirely covered within the union of all sampling strata i.e. $D \subseteq \cup_{j=1}^{J} \mathcal{S}_j$.

The first section of the third chapter is devoted to single-phase estimator(s) which only use data available at sample plots visited during the field NFI surveys. Single-phase estimators are a standard solution, which may be used for a vast number of target parameters. In the section which follows an implementation of the modified direct estimator as a special class of GREG (generalised regression estimator) [Rao, 2003, p. 13] is described.

This estimator uses not only data coming from (field) sample plots but also auxiliary data obtained from an external source. This strategy often leads to significant improvements of precision.

Depending on user settings (choice of parametrisation area and its relation to the estimation cell) and depending on the presence or absence of sample plots within the estimation cell itself the nFIESTA implementation results in an appropriate type of the GREG or the synthetic estimator (no data in the estimation cell, parametrisation is done in a larger region encompassing the estimation cell) [Mandallaz, 2012, p. 10].

The last chapter documents a generic implementation of a ratio estimator which is composed by two total estimators of potentially different type (single-phase or a kind of GREG using auxiliaries). Technically the two types can even be combined in the nominator and denominator of the ratio. However, from the practical point of view, the combinations between the nominator and denominator should follow certain constraints in order not to loose to much of the potential precision.

Ratio estimators are very important because many target parameters can be defined as a ratio of two totals (or means). Mean above ground biomass per hectare of forest (the exact map of which does not exist) was considered as one of the target parameters within the T2.3.1 case study.

2 The continuous Horvitz-Thompson Theorem

Cordy [1993] published the Horvitz-Thompson Theorem (HTC) for populations of a continuum, according to which an unbiased estimator $\hat{t}_{\mathcal{S}}$ of the total $t_{\mathcal{S}}$ of variable y in a sampling stratum $\mathcal{S}$ is defined by

$$\hat{t}_{\mathcal{S}} = \sum_{x \in s} \frac{y(x)}{\pi(x)}, \tag{1}$$

where The symbol $y(x)$ denotes local density of the quantity observed at point x of sample s of fixed size n, $\pi(x)$ is the inclusion density at point x which is an analogy to the inclusion probability when sampling finitely large populations of discrete objects [p. 355][Cordy, 1993]. The term local density was introduced by Mandallaz [1991] as a main building block of the infinite population approach to forest inventory. The infinite sampling approach to forest inventory was also studied by Eriksson [1995].

Estimator (1) is unbiased if the function $y(x)$ is positive or bounded and, at the same time, the condition $\pi(x) > 0 \; \forall \, x \in \mathcal{S}$ is satisfied, (i.e. any point x of the sampling stratum $\mathcal{S}$ can be sampled).

Furthermore if the local density function $y(x)$ is bounded and conditions $\pi(x) > 0 \; \forall x \in \mathcal{S}, \int_{\mathcal{S}} \frac{1}{\pi(x)} dx < \infty$ and $\pi(x, x') > 0 \; \forall \, x, x' \in \mathcal{S}$ hold an unbiased estimator of variance $\mathbb{V}(\hat{t}_{\mathcal{S}})$ can be obtained by using Eq. (2) or the modified Eq. (3) [Cordy, 1993, p. 357].

$$\hat{\mathbb{V}}(\hat{t}_{\mathcal{S}}) = \sum_{x \in s} \left[\frac{y(x)}{\pi(x)} \right]^2 + \sum_{x \in s} \sum_{\substack{x' \in s \\ x \neq x'}} y(x) y(x') \left[\frac{\pi(x, x') - \pi(x)\pi(x')}{\pi(x, x')\pi(x)\pi(x')} \right] \tag{2}$$

$$\hat{\mathbb{V}}(\hat{t}_{\mathcal{S}}) = \sum_{x \in s} \left[\frac{y(x)}{\pi(x)} \right]^2 + \sum_{x \in s} \sum_{\substack{x' \in s \\ x \neq x'}} \frac{y(x) y(x')}{\pi(x)\pi(x')} - \sum_{x \in s} \sum_{\substack{x' \in s \\ x \neq x'}} \frac{y(x) y(x')}{\pi(x, x')} \tag{3}$$

where $\pi(x, x')$ is a pairwise inclusion density [Cordy, 1993, p. 355].

The variance of the total estimator $\hat{t}_{\mathcal{S}}$ can alternatively be expressed and estimated by using Sen-Yates-Grundy (SYG) formulas [Cordy, 1993, pp. 358-359]. Although the SYG approach has several advantages it has not been implemented because of larger computational overhead. Unlike the SYG the variance calculation according to Eq. (2) or (3) can usually be performed faster by leaving out all sample locations (plots or clusters) with zero values of local density.

2.1 HTC and cluster sesigns

Unlike single plot a cluster design allocates a set of sample plots at each anchor point x. The positioning of individual plots of the cluster is determined by a geometric template (set of positional vectors with respect to the anchor point). The cluster orientation can be randomized between anchor points, but usually a fixed orientation is used. Cluster anchor points are generated within a sampling frame $\mathcal{F}$ the definition of which is implicitly given in terms of the following two conditions:

- At least one plot of any cluster with an anchor point in $\mathcal{F}$ is located in the sampling stratum $\mathcal{S} \subseteq \mathcal{F}$. This condition is relaxed for designs with randomly rotated clusters - it suffices if there is an orientation for which at least one plot derived from the anchor point $x \in \mathcal{F}$ would fall into $\mathcal{S}$.
- Each point x within the sampling stratum $\mathcal{S}$ can be selected by a plot pertaining to a cluster with anchor point in $\mathcal{F}$.

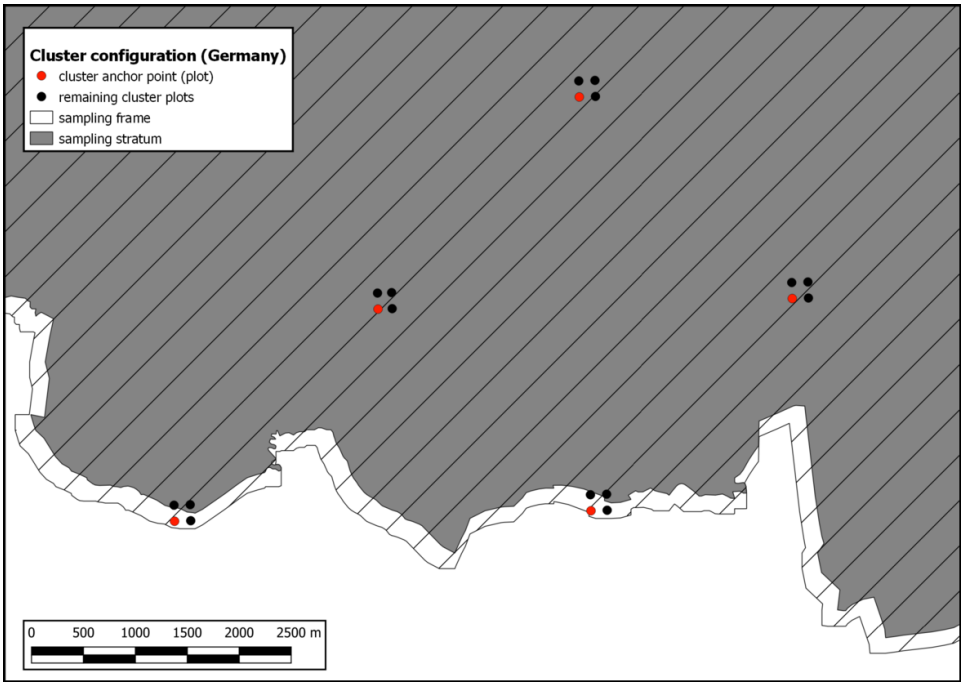

Figure 1: *Visualisation of sampling stratum and sampling frame (German NFI). Sampling stratum is shown as a (dark) gray polygon covered by the sampling frame shown as hatched area encroaching the borders of stratum up to 150 m (size of the square-like cluster) in south-west and south direction.*

In case of single-plot designs the sampling frame exactly corresponds to the sampling stratum i.e. we have $\mathcal{F} = \mathcal{S}$. Otherwise ($\mathcal{S} \subseteq \mathcal{F}$, cluster designs) the sampling frame need to be reconstructed by nFIESTA based on the geometry of sampling stratum and a geometrical representation of a cluster (the later also specifies the position of the cluster anchor point). Figure 1 shows an example of possible relation between sampling stratum and sampling frame taking the respective cluster geometry into account.

For designs using clusters the HTC theorem shown in the preceding section has to be reformulated with respect to $\mathcal{F}$ (sampling frame) rather than $\mathcal{S}$ (sampling stratum). In fact this is always possible and advantageous. In case clusters are not used we simply have $\mathcal{F} = \mathcal{S}$.

Further adjustments are needed in terms of the local density $y(x)$ which has to be replaced by the cluster level local density:

$$y_t(x) = \frac{\sum_{i=1}^{k} I_{i\mathcal{S}}(x) y_i(x)}{k} \tag{4}$$

where the (classical plot level) density $y_i(x)$ is observed for particular sample plot denoted by index i, belonging to the cluster established at anchor

point x. The symbol k is used for the nominal (and fixed) number of sample plots in each cluster pertaining to the given sampling stratum (irrespective on the number of plots located in the forest, inside or outside $\mathcal{F}$ or any other par of it). The term $I_{i\mathcal{S}}(x)$ indicates whether the sample plot i (of a cluster established at anchor point x) is found (value 1) or not in the sampling stratum $\mathcal{S}$ (value 0).

Notes:

- The position x corresponds to the cluster anchor point i.e. the position from which the cluster had been constructed. In general the position of an anchor point does not need to correspond to a plot center.
- The multiplication of local density by $I_{i\mathcal{S}}(x)$ is practically not needed, because plots outside the respective sampling stratum are either not uploaded to the nFIESTA system or their local densities are zeroed before the upload (the later option is needed in case of designs with randomized cluster orientation at each anchor point).
- The local density at cluster level $yt(x)$ is defined in way which differs from the approach suggested by Mandallaz [2007, p. 65], p. 65] and considered by Ene *et al.* [2016, p. 17]. With the definition of cluster level local density used by nFIESTA the mean spatial densities (per hectare values) can be estimated as ratios of two total estimators - both can be single-phase, of the regression type (GREG), or the single-phase and regression types can be mixed in the nominator and denominator of the ratio.

2.2 Derivation of inclusion densities

In the estimation system (its database part) a relative sampling weight $\chi(x)$ is attached to each combination of panel and sample location x (single plot or cluster depending on design). These relative weights can be intuitively understood as a way to express shares on the whole sampling frame area represented by particular sample locations. If all sample locations within a panel represent equally large regions a relative weight of one is used for all of them.

The use of relative weights has been established by the E-Forest system and NFI data providers from many European countries already got accustomed to this approach. Therefore it is supposed that this practice will be preserved.

The purpose of this section is to describe how the relative sampling weights can be transformed into sampling densities compliant to the Horvitz-Thompson theorem for continuous (infinite) populations (HTC). This is a

crucial step to be able to use this universal theorem for statistical estimations on the basis of an integrated database of many (not only) European NFIs.

Despite all formulas of this section are given in a form corresponding to a cluster design, they can be used also for single plot designs noting that in this case $\mathcal{S} = \mathcal{F}$. Complementary information concerning mathematical definition and properties of the inclusion densities $\pi(x)$ and $\pi(x, x')$ can be found in [Cordy, 1993, pp. 355-356].

Inclusion densities at sampling frame level

Following Cordy [1993] the inclusion density $\pi(x)$ for any fixed sample size design is given by

$$\pi(x) = \sum_{i=1}^{n_{\mathcal{F}}} f_i(x), \tag{5}$$

where $f_i(x)$ is a probability density function defined for each point x in $\mathcal{F}$ (sampling frame, a continuum i.e a bounded set of infinitely many points). This function describing the stochastic process of selection of the $i-th$ out of the fixed number of $n_{\mathcal{F}}$ sample elements (i.e. points) in $\mathcal{F}$.

The pairwise inclusion density is defined

$$\pi(x, x') = \sum_{i=1}^{n_{\mathcal{F}}} \sum_{j \neq i} f_{ij}(x, x'), \tag{6}$$

where $f_{ij}(x, x')$ is a joint probability density of selecting a pair of distinct sample elements i and j simultaneously.

Now, under the (still very generic) assumption that i) all $n_{\mathcal{F}}$ sample points are selected using the same inclusion density function $f(x)$ (i.e no stratification is used) and that ii) the particular sample points are selected independently and without replacement, we can write

$$\pi(x) = n_{\mathcal{F}} f(x), \tag{7}$$

and

$$\pi(x, x') = n_{\mathcal{F}}(n_{\mathcal{F}} - 1) f(x) f(x'). \tag{8}$$

Note that because of the independent selection of sample points at positions x and x' we have

$$f(x, x') = f(x) f(x'). \tag{9}$$

The probability density function $f(x)$ has the property

$$\int_{\mathcal{F}} f(x)dx = 1 \tag{10}$$

and it can be expressed with respect to a positive and bounded function $\zeta(x)$ defined in $\mathcal{F}$ as it follows

$$f(x) = \frac{\zeta(x)}{\int_{\mathcal{F}} \zeta(x)dx}. \tag{11}$$

Without any impact on any derivations made here the probability density $f(x)$ can be redefined using the total area $\lambda(\mathcal{F})$ of the sampling frame. Doing so we get

$$f(x) = \frac{c\ \zeta(x)}{c \int_{\mathcal{F}} \zeta(x)dx}, \tag{12}$$

where c is a constant defined by

$$c = \frac{\lambda(\mathcal{F})}{\int_{\mathcal{F}} \zeta(x)dx}. \tag{13}$$

In general, the construction of relative sampling weights should fulfill the condition that the larger the relative weight $\chi(x)$ the larger the share of the particular sample point x on the total of $\zeta(x)$ over the whole $\mathcal{F}$. This can be expressed by following equation

$$\chi(x) = \frac{const}{n_{\mathcal{F}}\, f(x)\ \int_{\mathcal{F}} \zeta(x)dx} = \frac{const}{n_{\mathcal{F}}\, \zeta(x)}, \tag{14}$$

where *const* is a positive constant.

Although they should be known to the respective country providing the NFI data the total $\int_{\mathcal{F}} \zeta(x)dx$ and values of $\zeta(x)$ are not accessible in the nFIESTA system. This is why probability density $f(x)$ neither the inclusion densities $\pi(x)$ and $\pi(x, x')$ can be derived from relative weights $\chi(x)$ based on Eq. (14). Nevertheless taking the Eq.(12) into account it is obvious that the previous assumption about relative weights also implies that the larger the relative weight the larger the representative area $1/\pi(x)$ of sample point x and the larger the share of this representative area on the total sampling frame area. Therefore we can also write

$$\chi(x) = \frac{const}{n_{\mathcal{F}}\, f(x)\, \lambda(\mathcal{F})} = \frac{const}{c\, n_{\mathcal{F}}\, \zeta(x)}. \tag{15}$$

Up to the *const* and c the values of all terms in Eq. (15) are available in the nFIESTA database and the key probability density $f(x)$ function can be expressed as it follows

$$f(x) = \frac{const}{\chi(x)\, n_{\mathcal{F}}\, \lambda(\mathcal{F})}. \tag{16}$$

The remaining question is how the *const* should be defined so that the relative weights according to Eq. (15) and (14) are the same or at least as close as possible while the probability density $f(x)$ can be evaluated in practice. We can make a step forward by combining Eqs. (10) and (15) to get

$$\int_{\mathcal{F}} f(x)dx = \frac{const}{n_{\mathcal{F}}\lambda(\mathcal{F})} \int_{\mathcal{F}} \frac{1}{\chi(x)}dx = 1, \tag{17}$$

which implies

$$\int_{\mathcal{F}} \frac{1}{\chi(x)}dx = \frac{n_{\mathcal{F}}\lambda(\mathcal{F})}{const} \tag{18}$$

and

$$\mathbb{E}\left[\frac{1}{\chi(x)}\right] = \frac{n_{\mathcal{F}}\lambda(\mathcal{F})}{const\,\lambda(\mathcal{F})} = \frac{n_{\mathcal{F}}}{const}. \tag{19}$$

The expected value of the reciprocal relative sampling weights is unknown in the general case but it can be estimated from the sample[1]:

$$\hat{\mathbb{E}}\left[\frac{1}{\chi(x)}\right] = \frac{\sum_{x\in s}\chi(x)\frac{1}{\chi(x)}}{\sum_{x\in s}\chi(x)} = \frac{n_{\mathcal{F}}}{W_{\mathcal{F}}}. \tag{20}$$

This estimator can be put into Eq. (19) which leads to $const = W_{\mathcal{F}}$, where the $W_{\mathcal{F}}$ is used for the sum of relative weights over the whole sample. Substituting $W_{\mathcal{F}}$ into Eq. (16) we get

$$f(x) = \frac{W_{\mathcal{F}}}{\chi(x)\,n_{\mathcal{F}}\,\lambda(\mathcal{F})}, \tag{21}$$

and using Eqs. (7) and (8) we come to the final expressions for inclusion densities, which are implemented by nFIESTA:

$$\pi(x) = \frac{W_{\mathcal{F}}}{\chi(x)\,\lambda(\mathcal{F})}, \tag{22}$$

and

$$\pi(x,x') = \frac{(n_{\mathcal{F}}-1)\,W_{\mathcal{F}}^2}{\chi(x)\,\chi(x')\,n_{\mathcal{F}}\,\lambda^2(\mathcal{F})}. \tag{23}$$

[1] According to E-Forest specification of the relative sampling weights $\sum_{x\in s}\chi(x)y(x)/\sum_{x\in s}\chi(x)$ is an unbiased estimator of the mean of y in $\mathcal{F}$.

Notes:

- For uniform sampling designs (having a constant relative sampling weight attached to all clusters) Eqs. (22) and (23) take the form corresponding to Uniform Random Sampling (URS, both $\pi(x)$ and $\pi(x, x')$ are constant over $\mathcal{F}$). The probability density is defined as $f(x) = 1/\lambda(\mathcal{F})$ while for the underlying function we have $\zeta(x) = 1$ everywhere within $\mathcal{F}$.

- In case of a non-constant probability density $f(x)$ and the respective function $\zeta(x)$ we have the so called importance sampling as presented by [Cordy, 1993, p. 356]. The inclusion densities $\pi(x)$ and $\pi(x, x')$ also vary over $\mathcal{F}$.

- Pairwise densities $\pi(x, x')$ for sample points x and x' belonging to different sampling strata are not needed because the corresponding summands of Eqs. (2) and (3) are zero in this case. Pairs of sample points from different strata can be skipped during the variance calculation.

- Single-phase estimates using densities derived at sampling frame level are unbiased conditionally on actual sample size in $\mathcal{F}$.

- The sum of separate single-phase estimators for arbitrary but complementary subsets of the whole sampling stratum exactly matches the estimator for the stratum as a whole. In case of estimators using auxiliaries the additivity is limited to the intersection of stratum and parametrisation area D_+ (sum of total estimates for complementary subsets of this intersection exactly matches the estimate for the intersection itself).

Inclusion densities at estimation cell level

Estimators using the inclusion densities as defined in the preceding section suffer from variance inflation due to uncontrolled (varying) sample size in the intersection of particular estimation cell D and sampling stratum $\lambda(\mathcal{F}_{jD}) \ll \lambda(\mathcal{F}_j)$. The variance increase becomes more pronounced as the relative cell size with respect to stratum size decreases.

This effects are well known from literature on finite sampling [Särndal *et al.* , 2003, pp. 283-284, 396-397]. Cordy [1993, p. 361] himself advices to use inclusion densities derived for particular sample sizes - as an alternative to unconditional inference admitting that the sample size vary within the study area (this is also the case of our estimation in cells $D \subset \mathcal{S}$).

Following the proposition of [Cordy, 1993] inclusion densities can be defined with respect to the extent of and the sample size in the sampling frame corresponding to one or the other of the two options:

- an intersection $\mathcal{F}_{jD}$ of particular stratum $\mathcal{S}_j$ and estimation cell D
- an intersection $\mathcal{F}_{jD_+}$ of particular stratum $\mathcal{S}_j$ and parametrisation region $D_+ \supseteq D$.

The modification of the basic approach described in preceding section consists in the uses of W_{jD} or W_{jD_+} and $\lambda(\mathcal{F}_{jD})$ or $\lambda(\mathcal{F}_{jD_+})$ instead of W and $\lambda(\mathcal{F})$ within Eqs. (22) and (23), where

$$W_{jD} = \sum_{\substack{x \in s \\ x \in \mathcal{F}_{jD}}} \chi(x), \tag{24}$$

and

$$W_{jD_+} = \sum_{\substack{x \in s \\ x \in \mathcal{F}_{jD_+}}} \chi(x). \tag{25}$$

By deriving inclusion densities for strata and cell intersections the inflation of variance due to uncontrolled sample size is ruled out completely (use of W_{jD} and $\lambda(\mathcal{F}_{jD})$) or at least potentially reduced (use of W_{jD_+} and $\lambda(\mathcal{F}_{jD_+})$). Therefore the estimator based on these densities have better statistical properties (unbiasedness conditional on the sample size in strata and cell intersections). However, this also means that the geographical additivity of totals is also gone.

Notes:

- This approach makes no sense for strata and cell (or parametrisation region) intersections in which less than two sampling locations exist. In such a case inclusion densities have to be derived at the whole stratum level which may lead to mixed inference within one estimation cell (or parametrisation region). For some strata and cell intersections we can have unbiasedness conditional on the number of sample locations at the whole sampling stratum level, for the rest an unbiasedness conditional on sample size in the stratum and cell intersection.
- Conditioning on the sample size in $\mathcal{F}_{jD}$ makes sense for single-phase estimators or for the direct GREG estimator if $D = D+$ i.e. the model parametrisation is performed using plots from the estimation cell only (no extended parametrisation area is used). Otherwise inclusion densities are to be derived at the whole sampling frame level or at the level of $\mathcal{F}_{jD+}$.
- The areas $\lambda(\mathcal{F}_{jD}))$ and $\lambda(\mathcal{F}_{jD_+}))$ are determined by GIS operations as part of the configuration of nFIESTA for the use of particular sampling strata, estimation cells and parameterisation areas.

- nFIESTA can work in both modes i.e. inclusion densities derived at the whole sampling frame level, or at the $\mathcal{F}_{jD}$ or $\mathcal{F}_{jD_+}$ level depending on user preference - additivity of totals versus better statistical properties of estimators. This is achieved through the user configuration of particular estimation task.

3 Estimation of totals

This section describes the way how target parameters defined as totals of an attribute over a study area called cell D are estimated by nFIESTA. The cells are arbitrarily shaped regions entirely within the union of all sampling strata i.e. $D \subset \cup_j^J \mathcal{S}_j$. The definitions of sampling strata (inclusing their geometrical representation by digital maps) as well as plot data belonging to them are available in nFIESTA.

Obviously some cells may be located completely within one sampling stratum but often the cells will encroach to several sampling strata. Unlike the preceding section on the the Horvitz-Thompson theorem the following part of the document deals with this and also other practical aspects of estimation for arbitrary cells.

3.1 Single-phase estimation

Single-phase estimators do not use any type of auxiliary information to improve their accuracy (reduce their variance). Most typically such estimators use data collected on sample plots through a dedicated field survey. Based on the HTC theorem (see chapter 2), the single-phase estimator $\hat{t}_y$ of total t_y of an attribute y in the estimation cell D is defined by

$$\hat{t}_y = \sum_{j=1}^{J} \sum_{x \in s_{2j}} \frac{y_{tjD}(x)}{\pi_j(x)}, \tag{26}$$

where the index j is used to iterate over J sampling strata, $\pi_j(x)$ is the corresponding inclusion density (see section 2.2) and $y_{tjD}(x)$ is the cluster level local density evaluated for anchor point x belonging to the (terrestrial) sample s_{2j} of stratum $\mathcal{S}_j$. The cluster level density itself is given by

$$y_{tjD}(x) = \frac{\sum_{i=1}^{k_j} I_{i\mathcal{S}_j}(x) I_{iD}(x) y_i(x)}{k_j}, \tag{27}$$

where k_j is the nominal number of sample plots per each cluster belonging to the sampling stratum $\mathcal{S}_j$, index i is used to iterate over all plots of the cluster established at anchor point x, $y_i(x)$ is the plot level local density on the plot i, $I_{iD}(x)$ is an indicator variable coding whether the position of sample plot i is within D (1) or not (0), likewise $I_{i\mathcal{S}_j}(x)$ is an indicator variable coding whether the plot lies inside stratum $\mathcal{S}_j$ (1) or not (0).

The indicator $I_{i\mathcal{S}_j}(x)$ is practically not needed because plots outside the respective sampling stratum are either not uploaded to nFIESTA or their local densities are zeroed before the upload (the later option is needed in case of designs with randomized cluster orientation at each anchor point). Consequently the cluster level local density can be simplified also noting that within the nFIESTA database the anchor points x themselves identify the respective sampling strata to which they belong so the use of j index in $y_{tjD}(x)$, $\pi_j(x)$ and $\pi_j(x, x')$ can be skipped:

$$y_{tD}(x) = \frac{\sum_{i=1}^{k_j} I_{iD}(x) y_i(x)}{k_j}, \tag{28}$$

and also the formula for $\hat{t}_y$ can be simplified to

$$\hat{t}_y = \sum_{x \in s_2} \frac{y_{tD}(x)}{\pi(x)}, \tag{29}$$

where the sum iterates over the sample $s_2 = \cup_{j=1}^{J} s_{2j}$ being a union of samples from all strata represented in given cell D.

Following the HTC (see chapter 2) the variance of $\hat{t}_y$ can be estimated by

$$\begin{aligned} \hat{\mathbb{V}}(\hat{t}_y) = &\sum_{j=1}^{J} \sum_{x \in s_{2j}} \left[\frac{y_{tD}(x)}{\pi(x)} \right]^2 \\ &+ \sum_{j=1}^{J} \sum_{x \in s_{2j}} \sum_{\substack{x' \in s_{2j} \\ x \neq x'}} y_{tD}(x) y_{tD}(x') \left[\frac{\pi(x, x') - \pi(x)\pi(x')}{\pi(x, x')\pi(x)\pi(x')} \right] \end{aligned} \tag{30}$$

which, unlike the Eq. (29), follows the structure of strata in order to speed up the computation. Knowing that $\pi(x, x') = \pi(x)\pi(x')$ for anchor points x and x' belonging to different strata (a consequence of independence of samples between strata) the value of the respective terms within the second (double) summation of Eq. (30) would be zero if x and x' coming from different strata entered the calculation.

Notes:

- In nFIESTA a linkage between sample plots and estimation cells as well as between sample plots and clusters is maintained which simplifies the practical calculation of $\hat{t}_y$ by taking only plots found inside of particular cell D and clusters and strata corresponding to these plots (no indicator values are needed for the practical calculation).

- Similarly also Eq. (30) is calculated more efficiently by taking only sample plots found in the estimation cell. All sample plots with local density equal to zero are skipped because such zero plots do not contribute to cluster densities and in addition zero cluster densities do not contribute to variance estimator according to Eq. (30).
- The formulas of this section are in a form corresponding to cluster designs. For simple plot designs these formulas can still be used (single plots are handled as clusters having just one plot i.e. $k_j = 1$).

3.2 Modified direct GREG estimator

This type of estimator(s) uses auxiliary information available and known anywhere in the estimation cell D as well as in the corresponding parameterisation region $D+ \supseteq D$. Conceptually similar estimator called small area estimator has been proposed by Mandallaz [2007, p. 119].

Under the condition $D = D+$, i.e. if the parametrisation of the regression model takes place at the level of the estimation cell itself, the formulas of this section can be used without any modification. They simply take the form of the (Generalised) Regression Estimator (GREG) known from the literature on finite population sampling e.g. [Rao, 2003, p.13] or [Särndal *et al.* , 2003, p. 225].

In addition the estimators of this section are formulated in a way which makes it also possible to calculate estimates even for a (potentially small) geographical domain D that actually does not contain any sample plots (or tracts). In such a case, the respective formulas (point as well as variance estimators) take the from corresponding to a synthetic estimator (model dependent estimation framework).

Modified direct GREG for single-plot designs

Let's define $\hat{t}_{y,md}$ - the modified direct GREG estimator of a population total t_y, which uses a linear model fitted *in an extended parametrization area* $D_+ \supseteq D$:

$$\hat{t}_{y,md} = \sum_{\substack{x \in s_2 \\ x \in D_+}} \frac{y(x)}{\pi(x)} \left\{ I_D(x) + (\mathbf{t_x} - \hat{\mathbf{t}}_\mathbf{x})' [\mathbf{X}_+ \mathbf{\Sigma}_+ \mathbf{\Pi}_+ \mathbf{X}'_+]^{-1} \frac{\mathbf{X}(x)}{\sigma^2(x)} \right\} \tag{31}$$

$$= \sum_{\substack{x \in s_2 \\ x \in D_+}} w(x) \tilde{g}(x) y(x) \tag{32}$$

$$= \sum_{\substack{x \in s_2 \\ x \in D_+}} \tilde{w}(x) y(x), \tag{33}$$

where the meaning of used symbols is the following:

D is the estimation cell. A digital map of D is stored in nFIESTA so the total area of D is exactly known. Although the plot coordinates are not provided by countries nFISTA has the information whether particular sample point x lies inside D or not.

D_+ is the parametrization area. A relation $D \subseteq D_+$ always holds. In addition D_+ is always enclosed within the union of all sampling strata considered, which means that condition $D_+ \subseteq \cup_{i=1}^{k} S_i$ must also be fulfilled (S_i denotes the $i-th$ of k strata). Parametrisation areas are defined (manually) by the user of nFIESTA.

s_2 is a subset of points $x \in D_+$ on which field survey has been performed (the so called second phase sample). In case of wall-to-wall (map-like) auxiliary data, the first phase sample is a census, so we know the values of a wall2wall auxiliary variable(s) anywhere in D_+.

x is a particular (dimensionless) sample point $x \in D_+$

$\pi(x)$ is the inclusion density function evaluated at the position of sample point x. Here it is supposed that the definition of inclusion density changes from stratum to stratum. Moreover, inclusion density is not necessarily a constant function within particular strata. More on inclusion densities can be found in chapter 2 and particularly in section 2.2.

$I_D(x)$ is a an indicator function coding whether a particular sample point x belongs to D ($I_D(x) = 1$) or not ($I_D(x) = 0$).

$\mathbf{t_x}$ is a column-vector of dimension $p \times 1$ (p corresponds to number of rows) holding exactly known totals of p auxiliary variables within D.

$\hat{\mathbf{t}}_\mathbf{x}$ is a column-vector of dimension $p \times 1$ (p corresponds to number of rows) of single-phase estimates of totals of p auxiliary variables within D, under the given sampling design defined by $\pi(x)$ and $\pi(x, x')$ i.e. single and pairwise inclusion densities, see chapter 2.

$\mathbf{X}_+$ is a $p \times n_{D_+}$ matrix of p auxiliary variables on n_{D_+} sample points in D_+.

$\mathbf{\Pi}_+$ is an $n_{D_+} \times n_{D_+}$ diagonal matrix of sampling weights. All non-diagonal elements of Π_+ are equal to zero. Diagonal elements of $\mathbf{\Pi}_+$ correspond to inverse values of inclusion density $\pi(x)$ at particular sample point $x \in D_+$.

$\mathbf{X}(x)$ is a $p \times 1$ column-vector of p auxiliary variables observed on particular sample point x (belonging to s_2 and also D_+).

$y(x)$ is a value of local density function based on field survey data observed on sample point $x \in D_+$. It is supposed, that any potential edge effect at the plot level has been compensated.

$\sigma^2(x)$ is the anticipated (error) variance of local density $y(x)$ on particular sample point denoted by x. In many (practical) cases we can write $\sigma^2(x) = \sigma^2$ i.e. the variance is constant in the whole $D+$. If this is the case, terms $\mathbf{\Sigma}_+$ and $\sigma^2(x)$ can be left out from Eq. (31).

$\mathbf{\Sigma}_+$ is an $n_{D_+} \times n_{D_+}$ diagonal matrix of weights. All non-diagonal elements of $\mathbf{\Sigma}_+$ are zero wheareas diagonal ones correspond to inverse values of $\sigma^2(x)$.

$w(x)$ is a sampling weight of particular sample point $x \in D_+$. It is defined as a reciprocal value of inclusion density $\pi(x)$.

$\tilde{g}(x)$ is a regression or calibration weight of sample point $x \in D_+$ as defined by Eq. (34).

$\tilde{w}(x)$ is a product of sampling $w(x)$ and regression weight $\tilde{g}(x)$ corresponding to sample point $x \in D_+$.

The modified direct GREG estimator $\hat{t}_{y,md}$ shown in Eq. (32) is defined as a sum of products of local densities $y(x)$, sampling weights $w(x)$, and regression or g-weights $\tilde{g}(x)$ defined by

$$\tilde{g}(x) = I_D(x) + \left(\mathbf{t_x} - \hat{\mathbf{t}}_\mathbf{x}\right)' \left[\mathbf{X}_+ \mathbf{\Sigma}_+ \mathbf{\Pi}_+ \mathbf{X}'_+\right]^{-1} \frac{\mathbf{X}(x)}{\sigma^2(x)}. \tag{34}$$

Using matrix notation $\hat{t}_{y,md}$ can be expressed by

$$\hat{t}_{y,md} = \tilde{\mathbf{G}}_+ \mathbf{\Pi}_+ \mathbf{Y}'_+, \tag{35}$$

where $\tilde{\mathbf{G}}_+$ is an $1 \times n_{D_+}$ row-vector of regression weights (n_{D_+} corresponds to the number of sample points in D_+) and $\mathbf{Y}_+$ is an $1 \times n_{D_+}$ row-vector of local densities corresponding to the set of sample points in D_+. The vector of regression weights itself is defined as

$$\tilde{\mathbf{G}}_+ = \mathbf{I_D} + \left(\mathbf{t_x} - \hat{\mathbf{t}}_\mathbf{x}\right)' \left[\mathbf{X}_+ \mathbf{\Sigma}_+ \mathbf{\Pi}_+ \mathbf{X}'_+\right]^{-1} \mathbf{X}_+ \mathbf{\Sigma}_+, \tag{36}$$

where $\mathbf{I_D}$ is an $1 \times n_{D_+}$ row-vector of indicator variables $I_D(x)$ observed on sample points in D_+.

Calibration property: When applied to the auxiliary variables defining the g-weights the Modified direct GREG estimator according to Eqs. (31), (32), (33) and (35) equals the exactly known vector $\mathbf{t_x}$ of auxiliary totals over D. This property is derived in an equation that follows:

$$\hat{\mathbf{t}}'_{\mathbf{x,md}} = \tilde{\mathbf{G}}_+ \mathbf{\Pi}_+ \mathbf{X}'_+, \tag{37}$$

$$= \mathbf{I_D} \mathbf{\Pi}_+ \mathbf{X}'_+ + (\mathbf{t_x} - \hat{\mathbf{t}}_\mathbf{x})' [\mathbf{X}_+ \mathbf{\Sigma}_+ \mathbf{\Pi}_+ \mathbf{X}'_+]^{-1} \mathbf{X}_+ \mathbf{\Sigma}_+ \mathbf{\Pi}_+ \mathbf{X}'_+, \tag{38}$$

$$= \hat{\mathbf{t}}'_\mathbf{x} + (\mathbf{t_x} - \hat{\mathbf{t}}_\mathbf{x})' \mathbf{I_{n_{D_+}}}, \tag{39}$$

$$= \mathbf{t}'_\mathbf{x} \tag{40}$$

Variance of the modified direct GREG: Let's define empirical residuals1 $\tilde{e}(x)$ as follows

$$\tilde{e}(x) = y(x) - \tilde{y}_p(x), \tag{41}$$

$$\tilde{e}(x) = y(x) - \mathbf{X}'(x) \hat{\boldsymbol{\beta}}_+, \tag{42}$$

where $\tilde{y}_p(x)$ is the empirical prediction and $\hat{\boldsymbol{\beta}}_+$ a $p \times 1$ column vector of regression coefficients estimated in D_+ by using

$$\hat{\boldsymbol{\beta}}_+ = [\mathbf{X}_+ \mathbf{\Sigma}_+ \mathbf{\Pi}_+ \mathbf{X}'_+]^{-1} \mathbf{X}_+ \mathbf{\Sigma}_+ \mathbf{\Pi}_+ \mathbf{Y}'_+ \tag{43}$$

or

$$\hat{\boldsymbol{\beta}}_+ = \tilde{\boldsymbol{G}}_{\beta_{t+}} \mathbf{\Pi}_+ \mathbf{Y}'_+ \tag{44}$$

where $\tilde{\boldsymbol{G}}_{\beta_{t+}}$ is a matrix of dimension $p \times n_{D_+}$. Using matrix notation, the $1 \times n_{D_+}$ vector of residuals $\tilde{\mathbf{E}}_+$ can be obtained as

$$\tilde{\mathbf{E}}_+ = \mathbf{Y}_+ - \mathbf{X}'_+ \hat{\boldsymbol{\beta}}_+. \tag{45}$$

Variance of $\hat{t}_{y,md}$ can be estimated as the variance of single-phase total of products of g-weights and empirical residuals over the whole parametrization area:

$$\hat{\mathbb{V}}(\hat{t}_{y,md}) = \hat{\mathbb{V}}(\hat{t}_{\omega,D_+}) \tag{46}$$

where

$$\hat{t}_{\omega,D_+} = \sum_{\substack{x \in s2 \\ x \in D_+}} \frac{\tilde{g}(x)\tilde{e}(x)}{\pi(x)} = \sum_{\substack{x \in s2 \\ x \in D_+}} w(x)\omega(x) = \tilde{\mathbf{G}}_+ \mathbf{\Pi}_+ \tilde{\mathbf{E}}'_+ \tag{47}$$

The variance $\hat{\mathbb{V}}(\hat{t}_{\omega,D_+})$ can be evaluated according to Eq. (30) in section 3.1 by substituting $y_{tD}(x)$ by $\omega(x)$.

Justification of the variance estimator: Multiplying both sides of Eg. (40) by vector of the population (true) regression coefficients $\boldsymbol{\beta}_+$ we get

$$\hat{\mathbf{t}}'_{\mathbf{x,md}}\boldsymbol{\beta}_+ = \mathbf{t}'_{\mathbf{x}}\boldsymbol{\beta}_+ = t_{y_p} \tag{48}$$

where t_{y_p} is the true total of population level predictions in D. By substitution into the left hand site of Eq. (48) we get

$$\begin{aligned}
\hat{\mathbf{t}}'_{\mathbf{x,md}}\boldsymbol{\beta}_+ &= \mathbf{I_D}\boldsymbol{\Pi}_+\mathbf{X}'_+\boldsymbol{\beta}_+ + && (49)\\
&\quad + (\mathbf{t_x} - \hat{\mathbf{t}}_{\mathbf{x}})'\left[\mathbf{X}_+\boldsymbol{\Sigma}_+\boldsymbol{\Pi}_+\mathbf{X}'_+\right]^{-1}\mathbf{X}_+\boldsymbol{\Sigma}_+\boldsymbol{\Pi}_+\mathbf{X}'_+\boldsymbol{\beta}_+, && (50)\\
&= \tilde{\mathbf{G}}_+\boldsymbol{\Pi}_+\mathbf{Y}'_{\mathbf{p}+}, && (51)\\
&= \hat{t}_{y_p,md} && (52)
\end{aligned}$$

where

$$\mathbf{Y}'_{\mathbf{p}+} = \mathbf{X}'_+\boldsymbol{\beta}_+ \tag{53}$$

is the $n_{D_+} \times 1$ vector of theoretical predictions, i.e. predictions which are based on the true (but unknown) vector of regression parameters $\boldsymbol{\beta}_+$. Looking at Eqs. (51) and (35) it can be seen immediately that the former corresponds to the $\hat{t}_{y_p,md}$ i.e. the modified direct GREG estimator of theoretical predictions $y_p(x)$ over D, where

$$y_p(x) = \mathbf{X}'(\mathbf{x})\boldsymbol{\beta}_+. \tag{54}$$

Based on Eqs. (48) and (52) it is obvious that the variance of $\hat{t}_{y_p,md}$ is zero because the estimator always equals the total of theoretical predictions t_{y_p} in D. In the following part it will be explained how this property can be used to derive an approximate variance estimator for the modified direct GREG.

Let's express $\hat{t}_{y,md}$ in terms of theoretical predictions $y_p(x)$ and (also theoretical) residuals $e(x)$ based on the following relations

$$\begin{aligned}
y(x) &= y_p(x) + e(x) && (55)\\
\mathbf{Y}_+ &= \mathbf{Y}_{\mathbf{p}+} + \mathbf{E}_+ && (56)
\end{aligned}$$

Substituting the right hand side of Eq. (56) into Eq. (35) we arrive at

$$\begin{aligned}
\hat{t}_{y,md} &= \tilde{\mathbf{G}}_+\boldsymbol{\Pi}_+\left[\mathbf{Y}'_{\mathbf{p}+} + \mathbf{E}'_+\right], && (57)\\
&= \tilde{\mathbf{G}}_+\boldsymbol{\Pi}_+\mathbf{Y}'_{\mathbf{p}+} + \tilde{\mathbf{G}}_+\boldsymbol{\Pi}_+\mathbf{E}'_+ && (58)\\
&= \hat{t}_{y_p,md} + \hat{t}_{e,md} && (59)
\end{aligned}$$

Consequently for the variance of $\hat{t}_{y,md}$ we can write

$$\mathbb{V}(\hat{t}_{y,md}) = \mathbb{V}(\hat{t}_{y_p,md}) + \mathbb{V}(\hat{t}_{e,md}) + \mathbb{C}(\hat{t}_{y_p,md}, \hat{t}_{e,md}). \tag{60}$$

Taking Eqs. (48) and (51) into account we have

$$\mathbb{V}(\hat{t}_{y_p,md}) = \mathbb{V}(t_{y_p}) = 0 \tag{61}$$

and further also

$$\mathbb{C}(\hat{t}_{y_p,md}, \hat{t}_{e,md}) = \mathbb{C}(t_{y_p}, \hat{t}_{e,md}) = 0 \tag{62}$$

The variance $\mathbb{V}(\hat{t}_{y,md})$ is then given as the variance of the modified direct GREG estimator of the total of theoretical residuals in D.

$$\mathbb{V}(\hat{t}_{y,md}) = \mathbb{V}(\hat{t}_{e,md}). \tag{63}$$

Obviously and unfortunately this result can't be used for practical work because we have expressed the variance of modified diret GREG by the variance of another modified direct GREG. In addition, theoretical residuals are generally unknown. In practice, the estimator $\hat{\mathbb{V}}(\hat{t}_{y,md})$ of the variance of $\hat{t}_{y,md}$ shown in Eq. (46) is obtained by replacing the vector of theoretical residuals $\mathbf{E}_+$ by sample residuals $\tilde{\mathbf{E}}_+$ and neglecting the actual dependence of g-weights $\tilde{\mathbf{G}}_+$ on particular sample(s). In other words it is supposed that the product $\tilde{g}(x)\tilde{e}(x)$ in Eq. (47) depends only on the particular sample point x and not on any other point(s), which were actually selected or could be selected from D_+ if we repeated the sampling (in the design-based sense). Under this assumption $\hat{t}_{\omega,D_+}$ is a one-phase estimator and its variance can be calculated accordingly i.e. according to Eq. (30) in section 3.1.

Zero estimate of total residuals: If there is a constant column vector ν fulfilling condition expressed by Eq. (64), the SREG and synthetic estimators are equal at the level of D_+. In other words the estimator $\hat{t}_{e,D_+}$ of the total of residuals in D_+ equals zero.

$$\sigma^2(x) = \nu' \mathbf{X}(x) \tag{64}$$

Proof and further details can be found in [Rao, 2003, chapter 2, page 14] or in [Särndal *et al.* , 2003, chapter 6, pages 231 and 232]. The proof given by Särndal *et al.* [2003] is now reproduced for shake of completeness of this methodology. Let's start with the equation

$$\hat{t}_{e,D_+} = \hat{t}_{y,D_+} - \hat{t}_{y_p,D_+} \tag{65}$$

showing the one-phase total estimator of regression residuals as the difference of two total estimators - one for the study variable y and the other for its

predictions y_p. The three one-phase total estimators are calculated for D_+ i.e. the region where the respective regression model is parameterized.
It can be shown that under the condition (64) the one-phase estimator $\hat{t}_{y_p,D_+}$ equals $\hat{t}_{y,D_+}$:

$$\begin{aligned}
\hat{t}_{y_p,D_+} &= \left(\sum_{\substack{x \in s_2 \\ x \in D_+}} \frac{\mathbf{X}'(x)}{\pi(x)} \right) \hat{\boldsymbol{\beta}}_+ \qquad (66) \\
&= \left(\sum_{\substack{x \in s_2 \\ x \in D_+}} \frac{\nu' \mathbf{X}(x) \mathbf{X}'(x)}{\sigma^2(x)\pi(x)} \right) \hat{\boldsymbol{\beta}}_+ \\
&= \nu' \left(\sum_{\substack{x \in s_2 \\ x \in D_+}} \frac{\mathbf{X}(x)\mathbf{X}'(x)}{\sigma^2(x)\pi(x)} \right) \left(\sum_{\substack{x \in s_2 \\ x \in D_+}} \frac{\mathbf{X}(x)\mathbf{X}'(x)}{\sigma^2(x)\pi(x)} \right)^{-1} \sum_{\substack{x \in s_2 \\ x \in D_+}} \frac{\mathbf{X}(x)Y(x)}{\sigma^2(x)\pi(x)} \\
&= \nu' \sum_{\substack{x \in s_2 \\ x \in D_+}} \frac{\mathbf{X}(x)Y(x)}{\sigma^2(x)\pi(x)} = \sum_{\substack{x \in s_2 \\ x \in D_+}} \frac{\nu' \mathbf{X}(x) Y(x)}{\sigma^2(x)\pi(x)} \\
&= \sum_{\substack{x \in s_2 \\ x \in D_+}} \frac{Y(x)}{\pi(x)} = \hat{t}_{y,D_+}
\end{aligned}$$

Hence, the total estimator $\hat{t}_{e,D_+}$ must be zero, and the proof is completed.
Condition (64) holds for some frequently used linear models used for model-assisted estimation:

- *models with an intercept term and constant variance* $\sigma^2(x) = \sigma^2$
- *ratio estimator* i. e. models with just one auxiliary variable $x(x)$ without the intercept and variance $\sigma^2(x) = \sigma^2 x(x)$
- *post-stratification* i.e. models including one or more categorial auxiliary variables with an error variance being linear combination of these

Modified direct GREG for cluster designs

In case sampling plots are clustered (grouped) into so called clusters of a fixed geometry (but not necessarily orientation) the approach outlined above should be modified. Unless cluster level edge-effects are compensated at each single geographical domain D (cell) separately, the estimation using above formulas would lead to biased estimators.
Unfortunately an explicit compensation of edge-effects due to clusters (e.g. by mirroring, walk-through, vector-walk or similar method) would lead to

geographically non-additive estimates of totals. In other words the estimate of total for the whole D_+ would not match the sum of estimates for any of its complementary geographical partitions. However, geographical additivity is very appreciated from the end user[2] as well as analyst perspective[3].
The modified approach described bellow retains the geographical additivity of totals at the level of D_+ s well as other good properties of the modified direct GREG (i.e. calibration property). At this point, it is also worth to mention that all clusters having at least one sample plot located in D_+ enter the below described calculations.

Tract level local density is defined by Eq. (28) in section 3.1. The same definition will also be used in this and following sections but let's define also the (tract) density y_{t,D_+}

$$y_{tD_+}(x) = \frac{\sum_{i=1}^{k_j} I_{iD_+}(x) y_i(x)}{k_j}, \tag{67}$$

where the indicator variable I_{iD_+} codes whether the plot i lies in D_+ or not. Finally, for designs using clusters the modified direct GREG estimator can be expressed by

$$\begin{aligned}\hat{t}_{y,mdt} = {} & \mathbf{I_{tD}} \boldsymbol{\Pi}_+ \mathbf{Y'_{tD}} + \\ & + (\mathbf{t_x} - \hat{\mathbf{t}}_\mathbf{x})' [\mathbf{X_{tD_+}} \boldsymbol{\Sigma}_+ \boldsymbol{\Pi}_+ \mathbf{X'_{tD_+}}]^{-1} \mathbf{X_{tD_+}} \boldsymbol{\Sigma}_+ \boldsymbol{\Pi}_+ \mathbf{Y'_{tD_+}},\end{aligned} \tag{68}$$

or equivalently by

$$\hat{t}_{y,mdt} = \mathbf{I_{tD}} \boldsymbol{\Pi}_+ \mathbf{Y'_{tD}} + \boldsymbol{\Delta}'_{\hat{\mathbf{t}}_\mathbf{x}} \tilde{\boldsymbol{G}}_{\beta_{t+}} \boldsymbol{\Pi}_+ \mathbf{Y'_{tD_+}}, \tag{69}$$

where

$\mathbf{I_{tD}}$ is an $1 \times n_{D_+}$ row vector of indicator values $I_{tD}(x)$ coding whether particular cluster with reference point (origin) at sample point x belongs to D or not. Value 1 codes the situation when at least one plot of cluster is found within D. In any other case the value of $I_{tD}(x)$ is zero.

$\mathbf{Y_{tD}}$ is an $1 \times n_{D_+}$ row vector of cluster level local densities $y_{tD}(x)$

$\mathbf{X_{tD_+}}$ is a $p \times n_{D_+}$ matrix composed by n_{D_+} column vectors $\mathbf{X_{tD_+}}(x)$ of dimension $p \times 1$, the elements of which correspond to cluster level densities of p auxiliaries, defined in direct analogy to $y_{tD_+}(x)$ (i.e. local densities are zeroed on plots outside D_+ before the aggregation at cluster level is performed).

[2] The sum of estimates over all mutually exclusive but complementary subsets of D_+ exactly equals the modified direct GREG estimate for D_+.

[3] A possibility to implement specific quality control measures namely to check whether additivity between subsets and the whole is fulfilled.

$\mathbf{Y_{tD_+}}$ is an $1 \times n_{D_+}$ row vector of cluster level local densities $y_{tD_+}(x)$

$\tilde{\boldsymbol{G}}_{\boldsymbol{\beta_{t+}}}$ is a $p \times n_{D_+}$ matrix used just for typing convenience

The vector $\hat{\mathbf{t}}_{\mathbf{x}}$ of one-phase estimators of totals of the auxiliaries in D is calculated according to

$$\hat{\mathbf{t}}_{\mathbf{x}} = \mathbf{I_{tD}} \boldsymbol{\Pi}_+ \mathbf{X'_{tD}}, \tag{70}$$

where $\mathbf{X_{tD}}$ is $p \times n_{D_+}$ matrix composed by n_{D_+} column vectors (dimension $p \times 1$) of auxiliary densities $\mathbf{X_{tD}}(x)$, the elements of which are defined in a direct analogy to $y_{tD}(x)$ (i.e. for plots outside D local densities are set to zero before cluster level density is calculated).

Variance of $\hat{t}_{y,mdt}$ can be estimated as variance of a one-phase estimator of total $\hat{t}_{\phi,D_+}$ of a density $\phi(x)$ in D_+

$$\hat{\mathbb{V}}(\hat{t}_{y,mdt}) = \hat{\mathbb{V}}(\hat{t}_{\phi,D_+}). \tag{71}$$

Estimator $\hat{t}_{\phi_t,D_+}$ is expressed by

$$\hat{t}_{\phi_t,D_+} = \sum_{\substack{x \in s_2 \\ x \in D_+}} \frac{\phi(x)}{\pi(x)} = \sum_{\substack{x \in s_2 \\ x \in D_+}} w(x)\phi(x) \tag{72}$$

and the corresponding density $\phi(x)$ by

$$\phi(x) = I_{tD}(x)\tilde{e}_{tD}(x) + \boldsymbol{\Delta}'_{\hat{\mathbf{t}}_{\mathbf{x}}} \tilde{\boldsymbol{G}}_{\boldsymbol{\beta_{t+}}}(x)\tilde{e}_{tD_+}(x), \tag{73}$$

where $\tilde{\boldsymbol{G}}_{\boldsymbol{\beta_{t+}}}(x)$ is a $p \times 1$ column vector of the matrix $\tilde{\boldsymbol{G}}_{\boldsymbol{\beta_{t+}}}$ corresponding to a cluster established at sample point x. The cluster-level residuals $\tilde{e}_{tD}(x)$ and $\tilde{e}_{tD_+}(x)$ are defined by

$$\begin{aligned} \tilde{e}_{tD}(x) &= \frac{\sum_{i=1}^{k_j} I_{iD}(x)\tilde{e}_i(x)}{k_j} \\ &= y_{tD}(x) - \mathbf{X'_{tD}}(x)\hat{\boldsymbol{\beta}}_{\boldsymbol{t+}} = y_{tD}(x) - y_{ptD}(x) \end{aligned} \tag{74}$$

and

$$\begin{aligned} \tilde{e}_{tD_+}(x) &= \frac{\sum_{i=1}^{k_j} I_{iD_+}(x)\tilde{e}_i(x)}{k_j} \\ &= y_{tD_+}(x) - \mathbf{X'_{tD_+}}(x)\hat{\boldsymbol{\beta}}_{\boldsymbol{t+}} = y_{tD_+}(x) - y_{ptD_+}(x), \end{aligned} \tag{75}$$

where $\tilde{e}_i(x)$ is used for plot level residuals

$$\tilde{e}_i(x) = y_i(x) - \mathbf{X'}_i(x)\hat{\boldsymbol{\beta}}_{\boldsymbol{t+}}. \tag{76}$$

The terms $y_{ptD}(x)$ and $y_{ptD+}(x)$ are the corresponding cluster-level model predictions.
In Eqs. (74), (75) and (76) the index i identifies a particular plot, $y_i(x)$ is plot level local density, $\mathbf{X}_i(x)$ is $p \times 1$ column vector of auxiliary local densities $x_i(x)$ sampled at plot i, $\hat{\boldsymbol{\beta}}_{\boldsymbol{t+}}$ is $p \times 1$ column vector of empirical regression coefficients obtained through

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{t+}} = [\mathbf{X}_{\mathbf{tD}_+} \boldsymbol{\Sigma}_+ \boldsymbol{\Pi}_+ \mathbf{X}'_{\mathbf{tD}_+}]^{-1} \mathbf{X}_{\mathbf{tD}_+} \boldsymbol{\Sigma}_+ \boldsymbol{\Pi}_+ \mathbf{Y}'_{\mathbf{tD}_+} \tag{77}$$

or equivalently by

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{t+}} = \tilde{\boldsymbol{G}}_{\beta_{\boldsymbol{t+}}} \boldsymbol{\Pi}_+ \mathbf{Y}'_{\mathbf{tD}_+}. \tag{78}$$

The elements of $\boldsymbol{\Sigma}_+$ i.e. the diagonal matrix of inverse variances can be set in a way which guarantees zero value of the total estimate $\hat{t}_{e,D_+}$ of residuals in the whole D_+. To achieve this property each diagonal element of $\boldsymbol{\Sigma}_+$ is set to

$$\boldsymbol{\Sigma}_+, \mathbf{j}(x,x) = \frac{k_j^2}{\sum_{i=1}^{k_j} I_{iD_+}(x)} = \frac{k_j^2}{m(x)}, \tag{79}$$

where $m(x)$ corresponds to the number of plots of the respective cluster which are located inside the parametrization area D_+. The symbol k_j is again the nominal cluster size (number of plots) in stratum j. The justification is based on the assumption that the variance of cluster level errors of the regression model follows

$$\sigma^2(x) = \frac{\sigma^2}{k_j^2} \sum_{i=1}^{k_j} I_{iD_+}(x), \tag{80}$$

noting that each cluster-level prediction $y_{ptD_+}(x)$ is influenced by a sum of up to k_j random errors which are (for simplicity reasons) supposed to be mutually independent with a constant variance σ^2. The term (variance) σ^2 is not part of the $\boldsymbol{\Sigma}_+(x,x)$ elements because it is a constant that cancels out between the inverse matrix and the following term in Eq. (43). In other words it has no influence on the estimate of parameters of the internally used linear regression model.

Notes:

1. For the commonly used regression models (e.g. linear model with intercept, or a combination of metric variable(s) and post-stratification without intercept) a constant vector ν according to Eq. (64) can be designed combining zeros and σ^2/k_j. In the special case of post-stratification the term σ^2/k_j is used at every position of ν.

2. This construction of $\mathbf{\Sigma}_+$ and the existence of corresponding vectors ν guarantees that the estimate of total residuals for the whole D_+ is always zero even if clusters are used. Hence we have an exact (not only approximate) equality of model-assisted and synthetic estimators of total.

3. For design that use no clusters $\mathbf{\Sigma}_+$ should be constructed as an identity matrix unless there is a (good) reason to use other specific weights.

4. The formulas of this section can directly be applied even if samples come from several strata, with different sampling designs, which may or may not use clusters or single plots.

5. According to this methodology it is not necessary to calculate total estimates separately for each stratum encroaching the boundary of given cell. However, concerning variance estimation, nFIESTA implementation follows structure of strata, see Eq. (30) in section 3.1.

4 Estimation of ratios

A number of target parameters can be expressed as the $R_{1,2}$ i.e. a ratio of two totals t_1 and t_2 (or equivalently mean values) according to Eq. (81).

$$R_{1,2} = \frac{t_1}{t_2} \tag{81}$$

The estimator $\hat{R}_{1,2}$ of $R_{1,2}$ can be obtained as the ratio of two total estimators $\hat{\theta}_1$ (for t_1) and $\hat{\theta}_2$ (for t_2).

$$\hat{R}_{1,2} = \frac{\hat{\theta}_1}{\hat{\theta}_2}. \tag{82}$$

Estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ can be of the same or of different types with respect to using or not using auxiliaries. In other words each of the two may correspond either to single-phase estimator of total according to Eq. (29) in section 3.1 or to the modified direct estimator (using wall-to-wall auxiliaries) expressed by Eq. (69) in section 3.2.

For shake of variance estimation the ratio $R_{1,2}$ can be approximated by $R_{1,2}^{(apx)}$ applying Taylor's linearisation technique while neglecting the 2nd and higher order terms

$$R_{1,2}^{(apx)} = R_{1,2} + \frac{1}{t_2} \sum_{x \in s_2} \frac{z(x)}{\pi(x)}, \tag{83}$$

$$= R_{1,2} + \frac{\hat{t}_z}{t_2}, \tag{84}$$

where $\hat{t}_z$ is a single-phase total of a residual variable $z(x)$ the definition of which depends on the type of estimator in the nominator and denominator of $\hat{R}_{1,2}$. Following options are possible:

$$z(x) = y_{tD}^{(1)}(x) + y_{tD}^{(2)}(x)\hat{R}_{1,2}, \tag{85}$$

$$z(x) = \phi_1(x) + \phi_2(x)\hat{R}_{1,2}, \tag{86}$$

$$z(x) = \phi_1(x) + y_{tD}^{(2)}(x)\hat{R}_{1,2}, \tag{87}$$

$$z(x) = y_{tD}^{(1)}(x) + \phi_2(x)\hat{R}_{1,2}, \tag{88}$$

where the first is used when both $\hat{\theta}_1$ as well as $\hat{\theta}_2$ are single-phase estimators of total, the second in case both estimators correspond to the modified direct GREG, the third if $\hat{\theta}_1$ is a modified direct while $\hat{\theta}_2$ is a single-phase estimator, and the last option is an analogy to the third with the estimator types just switched between the nominator and denominator of $\hat{R}_{1,2}$. Local densities corresponding to the nominator and and denominator are distinguished by indexes (1 for the nominator, 2 for the denominator). The terms ϕ_1 and ϕ_2 respectively are defined by Eq. (73) in section 3.2.

The variance $\mathbb{V}(\hat{R}_{1,2})$ is estimated by an approximate formula

$$\hat{\mathbb{V}}(\hat{R}_{1,2}) \approx \hat{\mathbb{V}}\left(\hat{R}_{1,2}^{(apx)}\right) \approx \frac{1}{\hat{\theta}_2^2}\hat{\mathbb{V}}(\hat{t}_z), \tag{89}$$

while $\hat{\mathbb{V}}(\hat{t}_z)$ is calculated according to Eq. (30) in section 3.1 by substituting $y_{tD}(x)$ by the residual local density $z(x)$ of the corresponding type.

Notes:

- Except the first case represented by Eq. (85) variance calculation involves all sample plots (as well as clusters and strata to which these plots belong) found in the parametrisation area D_+. Only in the case of the single-phase ratio the calculation is restricted to plots located in the estimation cell.
- Although nFIESTA implements all options of $\hat{R}_{1,2}$ with respect to the type of estimator in the nominator and the denominator it is not recommended to combine single-phase and modified direct GREG estimators. Furthermore, when two modified direct estimators define the ratio it is not recommended to use (very) different working models in the nominator and denominator. In all these cases a significant loss of precision is expected because the covariance between these (very) different estimators is decreasing.

- The precision of a ratio defined through single-phase estimators is practically the same or even better compared to a ratio of two (modified direct) GREGs. However because the (modified direct) GREG estimators of total are often more precise that their single-phase counterparts this type of ratio estimator can be useful at least from the user perspective - because it is consistent with the two GREGs which are likely to be published. In addition, if the working models in the nominator and denominator of the ratio are similar or even the same, the precision may be at least as good as the precision of the alternative single-phase ratio.

References

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.

Ene, L., T., Lanz., A., Adolt, R., & Hervé, Jean-Christophe. 2016. *Report on a general and flexible estimation procedure associating nfi field data with auxiliary data from various remote sensing sources and maps.* Tech. rept. DIABOLO - Distributed, Integrated and Harmonised Forest Information for Bioeconomy Outlooks (European Union's Horizon 2020 research and innovation programme, grant agreement No 633464).

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.

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.

Mandallaz, D. 2012. *Design-based properties of some small-area estimators in forest inventory with two-phase sampling.* Tech. rept. ETH (Available from http://e-collection.library.ethz.ch).

Rao, J. N. K. 2003. *Small area estimation.* Wiley.

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