forests

Article

Effects of Choosing Different Parameterization Data in Two-Phase Forest Inventories for Standing Stock Estimation

Ambros Berger * and **Thomas Gschwantner**

Austrian Research and Training Centre for Forests, Natural Hazards and Landscape (BFW), Institute for Forest Inventory, Seckendorff-Gudent-Weg 8, 1131 Vienna, Austria; thomas.gschwantner@bfw.gv.at
* Correspondence: ambros.berger@bfw.gv.at

Abstract: The demands on national forest inventories to provide detailed information for small geographical regions are rising. Two-phase estimators are often employed to obtain forest resource estimates, yet there is little information on optimal training data selection. This study evaluates the impact of different training data on two-phase estimators, with a focus on small area estimators for standing stock and aims to develop guidelines on selecting appropriate training datasets. Linear regression models were parameterized using multiple datasets and subsets based on ecological and administrative boundaries. The models were then applied on varying scales, and their estimates and their confidence intervals were compared to each other as well as to the single-phase, purely terrestrial forest inventory. Results suggest that the different two-phase models generally yield comparable estimates but differ notably from single-phase estimates. Specifically, differences increase in smaller areas and with correspondingly smaller training datasets, suggesting a minimum of 100 data points. To ensure robust estimates, we recommend adapting training sets to local conditions and exercising caution with small training datasets and areas because implausible results may occur. Pooling appropriate datasets is the preferable solution.

Keywords: national forest inventory; model parameterization; uncertainty; small area estimation; forest stock volume

check for updates

Academic Editor: Russell G. Congalton

Received: 23 December 2024
Revised: 20 January 2025
Accepted: 24 January 2025
Published: 30 January 2025

Citation: Berger, A.; Gschwantner, T. Effects of Choosing Different Parameterization Data in Two-Phase Forest Inventories for Standing Stock Estimation. *Forests* **2025**, *16*, 259. https://doi.org/10.3390/f16020259

1. Introduction

Effective forest management depends heavily on accurate information about current timber stock and its development over time. To meet this need, forest inventories are conducted at different geographical scales. In many countries, a national forest inventory (NFI) has been implemented on the country scale. These inventories usually consist of a large number of field plots, but due to the immense area that they cover, there are still huge gaps between individual plots. The NFI plots are usually surveyed on a regular basis with usually 5 to 10 years between visits, adding a coarse temporal resolution to the coarse spatial resolution [1–3].

To bridge knowledge gaps for highly detailed, localized information, forest enterprises, other forest management units, and municipalities conduct specialized surveys for their immediate information needs in areas that contain little to no NFI field plots (Koivuniemi and Korhonen 2006). However, these surveys come with notably higher costs per hectare. A cost-effective solution is the two-phase inventory approach [4–6] which combines NFI data with more easily obtainable auxiliary data, typically derived from remote sensing, to enhance the estimate.

Remote sensing has proven to be useful for forest inventories ever since it became available, from early aerial photos in the 1920s [4,7] to state-of-the-art high-resolution digital satellite images (e.g., Landsat, Sentinel, SPOT, or Pleiades) [8] and airborne laser scanning (ALS) [9]. The availability of wall-to-wall information enables researchers to fill in the gaps between inventory field plots and generate estimates for arbitrary areas by employing two-phase models (in this case, often called small area estimators). Due to their cost efficiency, spatial coverage, and higher temporal resolution, these models are implemented in an increasing number of countries [8,10]. Their outputs are used for critical management and political decisions across local and global scales, including harvest planning, resource assessments, and carbon accounting, particularly in the context of climate change [11–14]. Consequently, numerous studies over the past 15 years have tested various data sources, input variables, and model types.

Early studies on biomass and timber stock estimation from ALS data were conducted for example by Næsset [15] in Norway and Hollaus et al. [9] in Austria. Breidenbach and Astrup [16] compared two regression estimators from mixed-effect models for above-ground biomass in Norwegian municipalities with a low number of sample plots. In addition, in Norway, Gobakken et al. [17] compared a model-dependent estimator and a model-assisted two-phase estimator using NFI plot data and ALS data of Hedmark County. They applied different models to analyze the sensitivity of the estimates to model selection. A follow-up study was conducted by Næsset et al. [18], comparing the precision of the derived estimates. Magnussen et al. [19] conducted a cross-study in Norway and Switzerland, further refining the small-area estimators and also testing mixed-effect models and model averaging.

Maack et al. [20] spatially explicitly modeled the standing stock for the entire German province Baden−Württemberg using NFI, ALS, and Landsat data, showing that each additional data source improved the models. Ståhl et al. [21] compared three different model frameworks for large-area forest surveys. They also carried out a meta-study on available studies regarding the different types of models in forest research.

Hill et al. [22] implemented the regression estimators designed by Mandallaz and Mandallaz et al. [23,24] for two small-scale management units using satellite and ALS data. In 2022, Gschwanter et al. [1] published a comprehensive overview of the history and the current state of NFIs in Europe, as well as their harmonization and the ongoing integration of remote sensing to produce spatially explicit forest resource estimates. Recently, Georgakis et al. [25] employed a linear mixed model to improve their estimates derived from a laser terrain model and satellite images on a >2000 ha test site in Greece.

In most of the mentioned studies, the data for training the models were obtained within the study area, or only one training dataset was used. The choice of adding external training data (i.e., data from outside the target area) remains underexplored, raising questions about how to select appropriate datasets to train the models. This study addresses this gap by choosing training data based on various administrative and ecological boundaries, both within and outside the respective target region, and comparing the respective results. Although more sophisticated models with nonlinear methods and artificial intelligence are being developed, this study focuses on linear regression models which are currently operationally employed by the Austrian NFI to generate standing stock estimates. The objectives are as follows: (1) to assess the impact of training data selection on the model estimates under operational conditions; and (2) to develop guidelines for selecting appropriate training datasets.

2. Materials and Methods

2.1. Overview

In order to assess the impact of training data selection, a linear regression model using NFI and remote sensing data was parameterized multiple times with identical auxiliaries, but with a different selection of data points. The selection was based on either provinces, ecological regions, or their intersection. For each model, the stock estimate and its confidence interval for each province and for the entire country were calculated, and the results were compared. In a final step, small municipalities were analyzed as well.

2.2. National Forest Inventory

Various data sources were used for this study. The target variable (standing stock of stems in m^3ha^{-1}) was obtained from the field survey of the Austrian NFI conducted between 2016 and 2021. The Austrian NFI uses "tracts", clusters of up to four circular plots, employing angle count sampling with a maximum distance of 9.77 m, which results in circles of 300 m^2. The tracts are distributed on a 4 km regular square-shaped grid across Austria using an interpenetrating panel design, which means that tracts of each year cover the entire country. Details about this survey methodology are extensively described [3,26–28].

To align with the NFI sampling framework, data from individual plots were aggregated to the tract level using weighted means for all variables relevant to the modelling. In the NFI, the share of the forest area of the total plot area is measured in tenths of a plot which was used weight. Therefore, each tract provides one data point, and the respective weight is the sum of forest area tenths of all four plots within the tract, ranging from 1 to 40.

2.3. Normalized Digital Surface Model

A digital terrain model (DTM) of Austria was derived from an ALS campaign that was carried out over the span of several years [29]. There is a permanent flight campaign for aerial photos that covers the entire area of Austria every three years. These images are used to create a digital surface model (DSM) employing 3D image matching [30]. For this study, the DSM and the DTM were aggregated to a 1 m resolution. Subtracting the DTM from the DSM yields a normalized digital surface model (nDSM), known also as the vegetation height. Despite the time difference between the field and the remote sensing surveys, the data were treated as if they had occurred simultaneously.

A plot was removed if, compared to the previous NFI, 30% of the stock had been removed and the flight campaign had occurred before the field survey. A second rule was to remove plots if the data showed a severe mismatch between the calculated vegetation height in the remote sensing data and the timber stock volume recorded in the field. Specifically, the stock in m^3/ha had to be larger than the average vegetation height in m multiplied by 8 minus 100 and smaller than the average vegetation height multiplied by 80 plus 150. The first rule removed 13% of the plots, and the second rule removed another 1% of the remaining data. This also covered cases when removals happened after the field survey and the flight campaign even after that. Apart from removals, the mismatch can also be due to a shift in the plot location. After the respective plots were excluded, the data were aggregated to the tract level. The final sample included 3485 tracts (see also Table 2 which provides an overview of the data distribution later in the article).

2.4. Tree Species Map

Species information was derived from Sentinel-2 satellite images with a 10 m resolution. A neural network algorithm was used to assign occurrence probabilities to more than 20 classes of tree species or tree species compositions [31]. A full list of these classes is provided in Supplementary Table S1. The species information was aggregated to compute the proportion of deciduous species at the pixel level. The digital forest map published by the Austrian NFI [32] was used to obtain the forest boundaries. The remote sensing data were intersected with the forest map, and only pixels within the forest area were used to derive the auxiliary variables.

2.5. Model Development

Hill et al. [22] published the R-package forest inventory and a comprehensive documentation. For their package, they use the formulas for two-phase estimators and the respective variances developed and published by Daniel Mandallaz and his scientific group over several years [23,24]. A two-phase estimator consists of two data collection phases. The first phase is a large-scale survey (in our case, wall-to-wall) for the model's auxiliary variables, and the second phase (in our case, the field survey of the NFI) measures the target variable which cannot be observed directly in the first phase. These two datasets are then combined with a model (in our case, a linear regression).

The stock models were created in the statistical software R 4.1.1 using the lm function. As the first step, the standing stock per hectare on the NFI plot was plotted against the average vegetation height on the plot according to nDSM, with a distinction between conifer-dominated and deciduous-dominated tracts. Lowess lines were added to visualize trends within each group, as shown in Figure 1.

Figure 1. Standing stock against nDSM for two species groups (conifer- and deciduous-dominated). Lowess lines indicate the trends for both groups.

2.6. Basic Model

It is apparent from the correlation in Figure 1 that nDSM is a good predictor for stock—as has been found by multiple other studies, e.g., [9,16]. Notably, the relationship for conifers appears to be nonlinear. Generalized additive models were used to assess the influence of the auxiliary variables. This revealed that most relationships are linear, as indicated by the respective polynomial orders. Only ndsm showed a relevant nonlinear trend, and therefore, ndsm2 was introduced into the model. Table 1 summarizes the final variable selection. All auxiliary variables were highly significant except dss, which was kept in the model because its interaction term with ndsm2 was highly significant.

Following the selection of auxiliary variables, the two-phase estimator for national-level standing stock was implemented as:

$$\hat{y}_{2p} = b_0 + \sum_{i=1}^{6} b_i \cdot v_i, \tag{1}$$

and was referred to as the country model (CM) version. Here, b_i are the regression coefficients calculated by the *lm* function, v_i are the auxiliary variables from Table 1, and y is the target variable, standing stock. The subscripts *2p* and *1p* indicate two-phase and one-phase estimators, respectively, in Formulas (1) and (2). The values of b_i are listed in Supplementary Table S2. A residual analysis and a ten-fold cross-validation confirmed that the model was appropriate.

Table 1. Overview of the auxiliary variables in the volume stock model.

Variable	Description
ndsm	The average nDSM
ndsm2	The square of the average nDSM
dss	The share of deciduous species
ndsm2xdss	The product of the square of the nDSM and the share of deciduous species
slope	The average slope
dtm	The height above sea level as the average of the DTM

The model results and the respective confidence intervals were compared to the regular one-phase NFI estimates based purely on field data. Using the R^2 of the model, the standard deviation for the two-phase estimate can be derived as the square root of the two-phase variance, which is as follows [33]:

$$Var(\hat{y}_{2p}) = Var(\hat{y}_{1p}) \cdot \left(1 - R^2\right) \quad (2)$$

This simple formula can be used because the remote sensing information is a full survey. The "one-phase NFI estimate" in this study is not exactly the same value as the published, official NFI results. The values were adapted to account for the fact that the forest map includes areas that are not included in the NFI survey, for example inaccessible forest.

2.7. Large Scale: Provinces and Growth Regions

Accurate standing stock information is also required for smaller areas, such as Austria's nine provinces, presented in Figure 2. The Austrian NFI has traditionally reported results at this level. There are several options to obtain results in this case. For the most part, the provinces are large enough to contain sufficient field plots to create independent models. Therefore, the first options are either completely separate models (version PS) or a single model for the entire country, with bias correction to adjust the estimates for the individual provinces (version PB). The results of PS and PB differ because in PB all data points influence the solutions for all provinces and the province-specific adjustments are not as big as the differences between the independent models of PS. Both options were carried out and compared. Version PS is basically the extension of the single-phase NFI estimates for the provinces because only data from within the respective province are used.

For the separate models, the same formulas used for the country-wide estimate apply, each having their individual R^2. The independent variables were kept the same across all models for comparability reasons, even if they were not significant in one of the models.

An alternative implementation of the bias correction is to introduce the provinces as a categorical variable (or factor) into the model. Applying this model to the individual provinces directly yields the respective results. In this case, the R^2 of the model is the R^2 of the global model and is the same for all provinces. Finally, for comparison purposes, the estimates for the provinces $\hat{y}_{pi}$ were used to calculate country estimates $\hat{y}_c$ by using a weighted mean with the share of each province's forest area relative to the country forest area as weights w_i so that the weights sum to 1:

$$\hat{y}_c = \sum_{i=1}^{9} w_i \cdot \hat{y}_{pi} \tag{3}$$

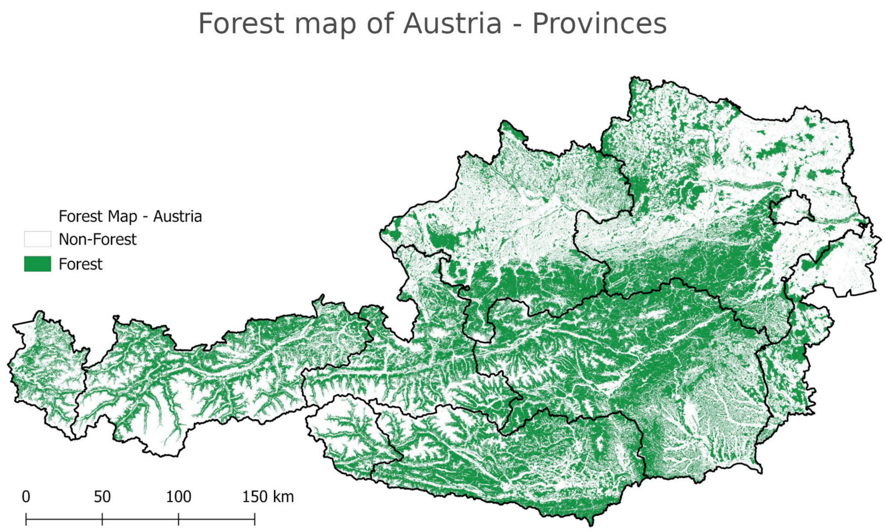

Figure 2. The forest map of Austria overlayed with the boundaries of the provinces.

The variance of those country estimates was calculated by assuming that the province estimates are independent—so all the covariances are 0—and then using a weighted sum with the same w_i as in the previous formula:

$$Var(\hat{y}_c) = \sum_{i=1}^{9} {w_i}^2 \cdot Var(\hat{y}_{pi}) \tag{4}$$

Austria is also divided into nine distinct natural regions which are called Hauptwuchsgebiete (main growth regions; [34]). As they are ecologically defined, they are expected to be better suited for stratification than political boundaries such as provinces. It is evident in Figure 3 that their boundaries match the forest land structure better than the administrative boundaries of the provinces in Figure 2.

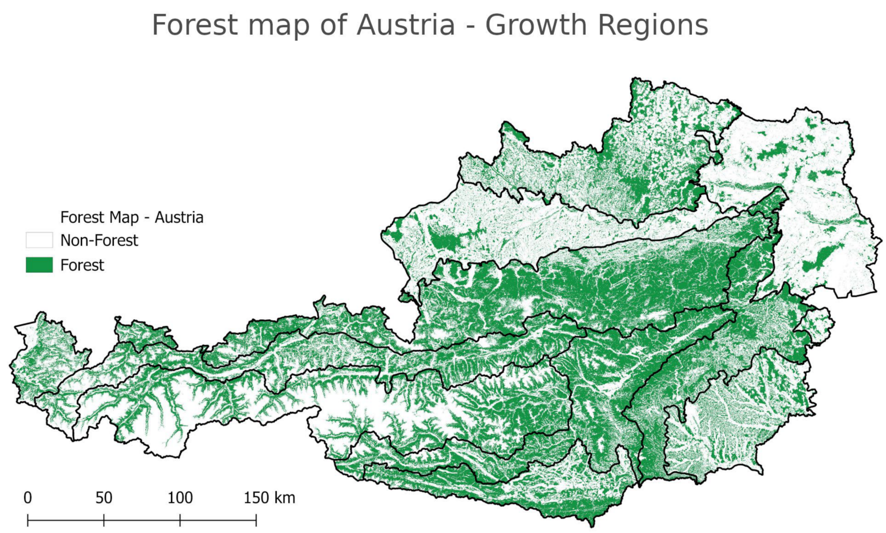

Figure 3. The forest map of Austria overlayed with the boundaries of the growth regions.

Thus, stocks for the provinces were calculated in three additional ways. First, the provinces and the growth regions were intersected to create 34 disjoint regions because not all growth regions occur in all provinces. Each province contains parts of two to six different growth regions. The provinces, growth regions, and their intersections differ greatly in size, and consequently, in the number of data points they contain. Table 2 provides an overview of the overlaps between provinces and growth regions, as well as the distribution of data points across intersections.

Table 2. Number of data points in provinces and growth regions as well as their intersections.

		Growth Region									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Sum**
Province	**1**	0	0	0	0	34	0	0	90	0	124
	2	95	0	185	0	4	203	0	0	0	487
	3	0	0	0	215	123	0	30	124	211	703
	4	0	1	0	184	0	0	123	0	166	474
	5	114	77	0	97	0	0	14	0	0	302
	6	64	89	221	134	188	0	0	153	0	849
	7	158	135	29	112	0	5	0	0	0	439
	8	0	21	0	76	0	0	0	0	0	97
	9	0	0	0	0	6	0	0	4	0	10
	Sum	431	323	435	818	355	208	167	371	377	3485

In the next step, different versions for stock models were created based on the growth regions. One version was a set of separate models for each intersection (version IS), another consisted of one model per growth region (version GS), and the third was a model parameterized on the entire country (version GB). The latter used bias correction as PB for the provinces, and consequently, GB also differs from GS. Table 3 provides an overview of the different model versions investigated in this study.

Table 3. Overview of the different model versions.

Model	**Description**
CM	Full country model, using all available data
PS	Separate models for each province, using only data within the province
PB	One model using all available data, including bias correction for provinces
GS	Separate models for each growth region; aggregated to the province level
GB	One model using all available data, including bias correction for growth region; aggregated to the province level
IS	Separate models for each intersection of growth region and province, if possible; aggregated to the province level

If fewer than 10 data points were available in the intersection, a synthetic model trained on the entire growth region was used. This means that no bias correction was used due to the lack of data within that intersection on which a bias correction could be based. For the variance estimation, the entire training dataset was used. Consequently, these estimators underestimated the variance but no other reliable information was available with which the estimate for the variance could be improved [22].

Finally, the estimates for the provinces were calculated using weighted means analogously to Equation (3). For each of the 34 intersections, the variance was obtained as well, and the estimate for the variance of the province can be calculated analogously to Equation (4). Again, country estimates were derived from the province estimates.

2.8. Small Scale: Municipalities

On the other end of the spectrum, Austria is divided into 7850 cadastral municipalities, 7773 of which contain forested areas as identified in the forest map used for this study. However, these municipalities are so small that none contain sufficient data points for model training. To address this, all six developed versions of the stock model were applied as synthetic models across these cadastral municipalities. The results were compared. Pair-wise, the standard deviation of the weighted differences between the estimates was calculated, the weight being the municipality's share of the total forest area. Thus, the weights sum to one which simplifies the formula for the standard deviation to:

$$sd(x) = \sqrt{\frac{n}{n-1}\sum_{j=1}^{n} w_j (x_j - \overline{x})^2}, \tag{5}$$

where n is the number of municipalities containing forest, w_j are the weights, x is the vector of the differences, x_j are the elements of x, and $\bar{x}$ is the weighted mean of x.

3. Results

3.1. Large-Scale Estimates

The results of the different modelling versions for the nine provinces and the entire country, the NFI estimates, and all respective confidence intervals are summarized in Figure 4. The black symbols show the estimate, while the according smaller gray symbols denote the respective 95% confidence interval. CM has no confidence intervals in the provinces because it is not a proper estimate. The CM value is biased and is included in the figure to illustrate how much the bias correction changes the estimate. The red horizontal lines show the estimates (full lines) and confidence intervals (dashed lines) of the one-phase NFI. The values of all estimates are presented in Table 4 and Supplementary Table S3 contains all estimates and all confidence intervals displayed in Figure 4.

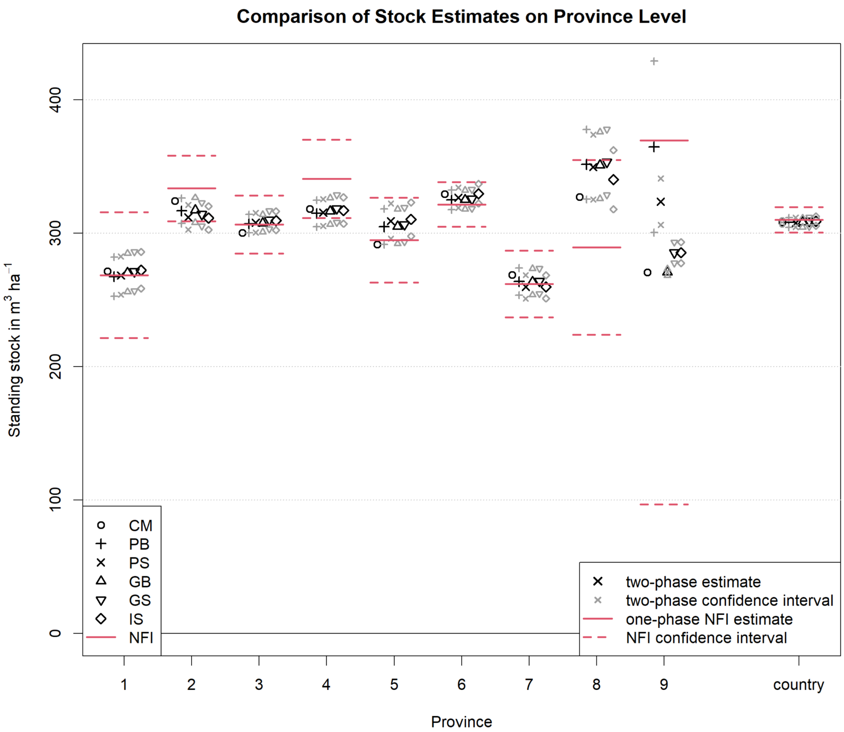

Figure 4. Two-phase estimates for the standing stock of different versions for the provinces and the entire country. The red lines indicate the one-phase NFI estimates.

Table 4. Standing stock estimates (in m^3ha^{-1}) of all versions for provinces and the entire country, rounded to one decimal place.

Province	NFI	CM	PB	PS	GB	GS	IS
1	268.6	271.4	267.5	268.3	270.4	271.3	272.3
2	333.7	324.1	316.8	312.0	317.4	314.2	311.5
3	306.6	300.3	307.2	307.9	307.5	310.0	309.4
4	340.8	318.2	314.8	315.5	316.6	318.4	317.1
5	294.9	291.6	304.9	309.1	305.1	306.4	310.4
6	321.5	329.3	325.0	326.7	325.1	325.7	329.7
7	262.1	268.8	263.8	259.8	263.6	264.1	259.8
8	289.5	327.3	351.7	349.7	351.0	353.3	340.2
9	369.7	270.8	364.8	323.7	270.8	285.4	285.4
Country	310.0	308.1	308.1	307.8	308.4	309.1	308.9

The overall stock lies between 307.8 and 309.1 m^3ha^{-1} for all versions, whereas the single-phase NFI estimate is 310.0 m^3ha^{-1}. Even though the two-phase estimates are all lower, they fall well within the confidence interval of the NFI estimate, as well as within each other's confidence interval. The estimates for the entire country are very precise, with a standard deviation of only 0.89 m^3ha^{-1} for the CM and the standard deviations ranging from 1.7 to 1.8 m^3ha^{-1} for the versions that first calculated province estimates.

At the province level, the picture is more diverse. With the exception of province 9, again, all two-phase estimates lie within the confidence intervals of the others and within the confidence intervals of the NFI estimate, but sometimes near the edge (provinces 2, 4, and 8). As such, in several cases, the confidence interval of the two-phase estimate does not contain the NFI estimate.

3.2. R^2 of Models

The models were parameterized on diverse regions and thus vary in quality. The model for the entire country reaches an R^2 of 0.752, and Figure 5 presents an overview of the R^2-values for the models developed for PS, GS, and IS. There are no boxplots for CM, PB, and GB because each consists of just one model.

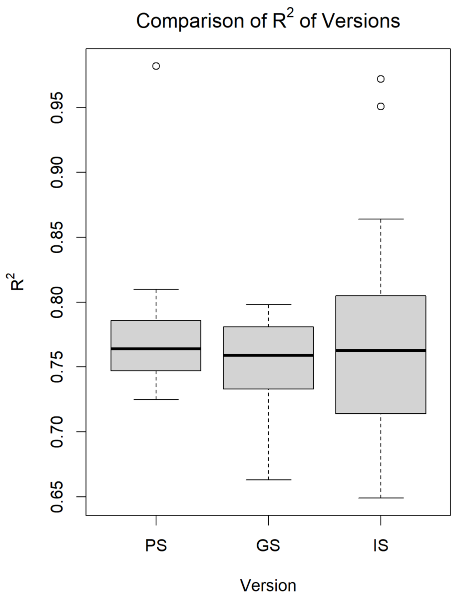

Figure 5. Boxplots of the R^2-values of the versions PS, GS, and IS.

No statistically significant difference can be observed between the medians of the three versions. They are all greater (but not significantly greater) than those of CM. The circles show outliers with extremely high R^2, stemming from models with less than 30 data points each. A full list of the model coefficients for all versions, their R^2 and the respective number of data points is detailed in Supplementary Table S2.

3.3. Results on Municipalities

Figure 6 presents the results for the synthetic application of all versions for the cadastral municipalities. The straight line is the identity line, and the standard deviations in m^3ha^{-1} from Equation (5) are presented in the upper-right half of the matrix.

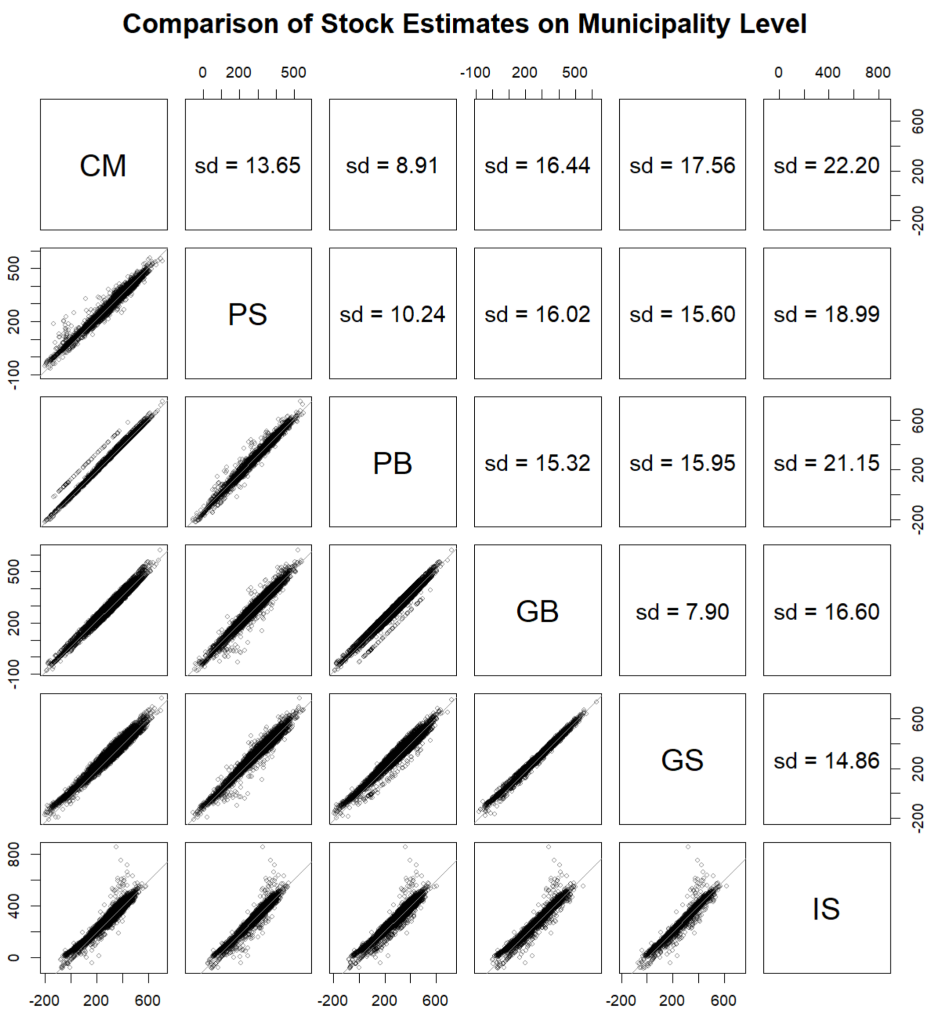

Figure 6. Comparison of the stock estimates of the different versions in m^3ha^{-1} for the cadastral municipalities. The line indicates the identity line, and the standard deviation is an indicator of dissimilarity between two versions.

The standard deviations in Figure 6 serve as a measure of dissimilarity between the versions. The mean of the standard deviations associated with a version reflects the degree of similarity in estimates produced by that version relative to other versions. While results across all versions are highly correlated, notable differences are evident in some municipalities and sometimes the estimated stock per hectare is even negative. The means of the standard deviations and the number of negative estimates per version are presented in Table 5. Version IS exhibits the highest mean SDs by a large margin, followed by CM. In contrast, the remaining versions exhibit relatively similar mean SD values that range from 14.3 to 14.9.

Table 5. Average standard deviation and number of negative estimates for the different versions.

Version	CM	PB	PS	GB	GS	IS
Mean of SDs	15.8	14.3	14.9	14.5	14.4	18.8
Negative estimates	76	71	44	64	27	27

4. Discussion

4.1. Model Performance

With an R^2 of about 0.75, the model CM performs on par with the best model in Maack et al. [20]. The unexplained variance could stem from the following aspects: (i) the inherent limitations of independent variables in fully describing vegetation characteristics; and (ii) uncertainties associated with field sampling methods at the single plot and single tree level, which were analyzed for the Austrian NFI by Berger et al. [28] and Berger et al. [35], respectively. Additional uncertainties stem from the processing and correction procedures applied to the dss and nDSM, as well as from the minor inaccuracies in the exact coordinates of the field plots and the remote sensing data. Given these constraints, achieving notably higher R^2 values with current technologies and data is unlikely.

At the national scale, all tested versions yield consistent estimates, with minor deviations between the NFI and two-phase estimates. The deviations were expected due to the following factors: (i) differences in timing of field and remote sensing data collection; and (ii) the exclusion of certain field plots from the model, for example plots without species information or those affected by disturbances such as harvesting or calamities between the remote sensing and the field surveys. Despite these possibilities, the national-level estimates align remarkably well.

At the provincial level, the estimates across all versions consistently fall within the 95% confidence interval of the one-phase estimate, although minor differences between provinces exist. This suggests that the versions agree on the overall stock but sometimes allocate the stock differently across different provinces. However, all model versions generally agree on whether the one-phase estimate is too small, too large, or appropriate. While the country-wide NFI estimate is highly accurate (with a much smaller confidence interval in the two-phase system), the estimates for some provinces (and particularly even smaller areas) can be improved with the two-phase system.

The fact that the versions produce similar results on the province level indicates that training data from locations that differ considerably from the area of interest neither improve nor worsen the estimate, provided that appropriate bias corrections are implemented and no influential points of leverage are present. Depending on the specific data structure, these data points from different locations can impact the model's R^2 value and thus alter the confidence interval. Therefore, caution is warranted when using synthetic estimators because it is impossible to add a bias correction, as there are no data available for this purpose [10,22,36].

4.2. Small Datasets

For provinces with fewer data points, more differences emerge between the model versions. Province 9, which has the smallest forest area and fewest field plots, shows unusual effects. Due to the small number of field plots, the NFI confidence interval is huge, and the upper bound is not visible in Figure 4.

For version PS in province 9, only ten data points were available to parameterize a model with seven coefficients. This leads to a vastly inflated R^2 value of almost 0.98, which clearly indicates overfitting (a similar effect is observed in two IS models; see Supplementary Table S2 and Figure 5). The estimate aligns with other estimates, but the confidence interval is severely underestimated due to the overestimated R^2. In contrast, the confidence interval for version PB is much larger and a lot more plausible. The confidence intervals of the versions relying on growth regions are even narrower because subdividing the already small province further into two main growth regions requires the use of a synthetic estimator. These estimators contain no ground truth information in the area of interest. Consequently, the variation between field plots is unavailable, and the confidence are underestimate [22]. Similar effects were observed in three other provinces. However, these provinces are larger, and because the subdivisions of the provinces are summed up using the forest area as weight, the effect of the small subdivisions becomes negligible when aggregated at the province level. Nonetheless, the estimates and confidence intervals for province 9 remain unreliable and are retained solely for completeness and illustrative purposes.

4.3. Growth Regions and Municipalities

It was expected that the models trained on the main growth regions would be trained on more homogenous data and thus have a higher R^2, but this could not be confirmed. Possibly, the auxiliary variables altitude above sea level and slope captured much of the variability between the growth regions. In fact, the model with the lowest R^2 is based on growth regions, and overall, the versions demonstrate comparable quality. An exception is CM, which has a known and uncorrected bias at the province level. For the other version, no clear preference emerges as they generally performed similarly.

The means of the standard deviations in Table 5 indicate that the province (PB, PS)- and growth region (GB, GS)-based models are quite similar to each other and distinct from CM and particularly IS. CM has no bias correction and therefore misses details, while some IS models rely on very small training datasets and produce unreliable results. The other four models are based on datasets that were each selected in a different way but are of comparable size (triple digits, except for province 9). Despite the different selection regime for training data, they produce the most similar results, implying that a sweet spot for dataset size has been found and any of the four models can be applied in practice (except PS for province 9), allowing for the consideration of local characteristics while still using sufficient data for robust estimation.

Another possible indicator for choosing training data and model type is the number of negative estimates at the municipality level. In this regard, the individual models have a slight advantage, in particular GS and IS. However, the municipalities with negative estimates all have low average vegetation height, potentially including young regenerating forest, temporarily unstocked forest land, and limited forest area (all under 10 ha). For comparison, the average forest area in Austrian municipalities is approximately 500 ha, and the median is 181 ha.

Supplementary Table S2 shows that the large majority of the models have a negative intercept; for CM, it is ca. -85 m^3ha^{-1}, and others have triple digits. On the other hand, the coefficient for ndsm (vegetation height—the auxiliary variable with the most influence) is positive, which yields a positive timber stock for a regularly stocked forest. However, in the case of none or low vegetation, the negative intercept may dominate the result for this particular area. Some of the negative values may also be caused by local data errors which can have a large effect if the total forest area of the municipality is small. If negative values need to be avoided on the municipality level, the modelling approach needs to be changed. For example, nonlinear models that can only asymptotically approach 0 could be used. However, in exchange, one would lose the advantages of linear regression model, for example the very efficient calculation of estimates and easily obtainable, symmetrical confidence intervals.

The comparison of version PB with others (Figure 6) reveals a line of points parallel to the main data cloud. These points are associated with province 9, where the bias correction in PB is derived from just 10 data points. In contrast, PS utilizes these points to parameterize the entire model, leading to the observed scatter around the main point cloud. This discrepancy suggests that in cases of limited data points, pooling geographically adjacent provinces to create a single combined model would be more appropriate to generate reliable outputs. The value of a comprehensive NFI becomes evident in this context as it offers an invaluable and unique source of data, where diverse forest conditions and locations are surveyed in a consistent and homogeneous way, allowing the use of data from geographically distant but structurally and ecologically similar forest plots, thus enhancing the reliability and generalizability of model predictions.

Version IS, trained on the most split up datasets, exhibits deviations from the other versions in the scatterplot matrix. This could be due to the fragmentation and the resulting small intersections of provinces and growth regions, a similar issue as observed with province 9 in PS. Province 9 does not cause issues in the models based on growth regions, because it is split up, and due to the even further reduced number of plots per fragment, synthetic models trained on bigger datasets are used. The versions GB and GS have no such problems because the minimum number of used data points is 167 (see Table 2).

When considering all of the above, version GS is most preferred. However, the suitability of other versions should not be discounted, as they may perform equally well or even surpass GS in certain contexts. Specifically, localized models often adapt better to local characteristics, but this study suggests that, to optimize model performance, a minimum of 75 to 100 data points should be targeted. For example, province 8 with 97 data points still exhibited more variation than ideal. Most critical are the models with 30 data points or less, for example the three outliers visible in Figure 5 which are the result of overfitting. Using such models leads to false confidence in the results.

In this study, the mentioned number of 75 to 100 tracts corresponds to an area of roughly 1000 km^2 of forest area. However, for countries with different landscape structure, forest conditions or sampling regimes compared to Austria, these numbers may need to be adjusted. However, there is limited literature from other countries that offers insights into such modelling thresholds. Most studies referenced utilized datasets containing over 600 plots, often in the thousands. An exception is a study of Breidenbach and Astrup [16], in which they worked with only 145 plots distributed over 14 municipalities, each containing from 1 to 35 plots. This smaller dataset constrained their ability to construct separate models and led to the use of mixed-effect model estimators, which resembled synthetic estimators. Their findings reinforced the notion that more data are generally beneficial, particularly for smaller municipalities where bias correction remains challenging.

Georgakis et al. [25] had to cluster the forest management units in order to successfully apply their model because they contained mostly fewer than three plots. The clusters had an average of 11.6 plots which is substantially lower than what is suggested in this article, but they never created models for single clusters and they had over 200 plots available in total. Despite the more sophisticated method (mixed-effects model and data preprocessing), they reached an R^2 of 0.71 which is lower than that of CM. This is likely due to the smaller sample size and, even more importantly, because their auxiliary variable "vegetation height" was obtained from satellite images instead of aerial photography.

For very small datasets, model estimates can become unreliable, also due to issues of extrapolation. Ideally, the training data should cover the full range of conditions present in the target area, a goal that becomes increasingly difficult to achieve in small datasets. The modelling approach presented here underscores the importance of using a proper probability sample [21]. Given that the choice of training data can significantly affect model outcomes, transparency and rigor in the modelling process are essential to ensure robustness and reliability of the results.

5. Conclusions

The two-phase estimators are gaining prominence in the field of forest inventory and monitoring due to their demonstrated capacity to enhance the precision of the estimates, especially for smaller areas. Moreover, these methods enable NFIs to extend their analytical capabilities and services both spatially and temporally to address evolving management and policy needs. Generally, there are several legitimate ways to parameterize models for estimating standing stock. When applied to sufficiently large and representative datasets, these approaches typically yield comparable results. However, the presence of influential outliers can cause significant deviations in outcomes. The authors recommend using training datasets that comprise at least 75 to 100 data points, ensuring they cover the full range of environmental conditions relevant to the target area. For very small areas, it may become necessary to take local particularities into account and choose the training dataset accordingly to avoid artifacts such as negative estimates. However, such cases are rare, and excessive manual selection risks introducing bias. Therefore, any modifications to the training dataset must be well justified and transparently documented. This study demonstrates that pooling geographically distant data points from ecologically similar areas offers a robust alternative to relying on small or incomplete datasets. It shows the flexibility of two-phase estimation approaches in adapting to diverse and challenging data constraints, making them an invaluable tool for forest resource monitoring and management.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/f16020259/s1, Table S1: Species in the tree species map, Table S2: Coefficients of all models, Table S3: All estimates at province and country level including their confidence intervals.

Author Contributions: Conceptualization, A.B. and T.G.; methodology, A.B.; software, A.B.; validation, A.B. and T.G.; formal analysis, A.B.; writing—original draft preparation, A.B.; writing—review and editing, A.B. and T.G.; visualization, A.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data underlying this article will be made available on reasonable request to the corresponding author due to GDPR restrictions, and where necessary, in aggregated form.

Acknowledgments: We thank our colleagues at the Austrian NFI who collect and process the data. Without their hard work, this study would not have been possible.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gschwantner, T.; Alberdi, I.; Bauwens, S.; Bender, S.; Borota, D.; Bosela, M.; Bouriaud, O.; Breidenbach, J.; Donis, J.; Fischer, C.; et al. Growing Stock Monitoring by European National Forest Inventories: Historical Origins, Current Methods and Harmonisation. *For. Ecol. Manag.* **2022**, *505*, 119868. [CrossRef]
2. Vidal, C.; Alberdi, I.A.; Hernández Mateo, L.; Redmond, J.J. *National Forest Inventories: Assessment of Wood Availability and Use*; Springer: Cham, Switzerland, 2016. [CrossRef]
3. Tomppo, E.; Gschwantner, T.; Lawrence, M.; McRoberts, R.E. *National Forest Inventories: Pathways for Common Reporting*; Springer: Dordrecht, The Netherlands, 2010. [CrossRef]
4. Loetsch, F.; Haller, K.E.; Brünig, E.F. Vol. 1: Statistics of Forest Inventory and Information from Aerial Photographs. In *Forest Inventory*; BVL Verlagsgesellschaft: München, Germany, 1964.
5. Cochran, W.G. *Sampling Techniques*, 3rd ed.; Wiley: New York, NY, USA, 1977; ISBN 978-0-471-16240-7.

6. Mandallaz, D. *Sampling Techniques for Forest Inventories*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2007; ISBN 978-0-429-14461-5.
7. Schreuder, H.T.; Gregoire, T.G.; Wood, G.B. *Sampling Methods for Multiresource Forest Inventory*; Wiley: New York, NY, USA, 1993; ISBN 978-0-471-55245-1.
8. McRoberts, R.; Tomppo, E. Remote Sensing Support for National Forest Inventories. *Remote Sens. Environ.* **2007**, *110*, 412–419. [CrossRef]
9. Hollaus, M.; Wagner, W.; Schadauer, K.; Maier, B.; Gabler, K. Growing Stock Estimation for Alpine Forests in Austria: A Robust Lidar-Based Approach. *Can. J. For. Res.* **2009**, *39*, 1387–1400. [CrossRef]
10. Dettmann, G.T.; Radtke, P.J.; Coulston, J.W.; Green, P.C.; Wilson, B.T.; Moisen, G.G. Review and Synthesis of Estimation Strategies to Meet Small Area Needs in Forest Inventory. *Front. For. Glob. Chang.* **2022**, *5*, 813569. [CrossRef]
11. Breidenbach, J.; Granhus, A.; Hylen, G.; Eriksen, R.; Astrup, R. A Century of National Forest Inventory in Norway—Informing Past, Present, and Future Decisions. *For. Ecosyst.* **2020**, *7*, 46. [CrossRef] [PubMed]
12. Alberdi, I.; Bender, S.; Riedel, T.; Avitable, V.; Boriaud, O.; Bosela, M.; Camia, A.; Cañellas, I.; Castro Rego, F.; Fischer, C.; et al. Assessing Forest Availability for Wood Supply in Europe. *For. Policy Econ.* **2020**, *111*, 102032. [CrossRef] [PubMed]
13. Baycheva-Merger, T.; Sotirov, M.; Holmgren, S.; Selter, A. Institutional and Actor-Oriented Factors Constraining Expert-Based Forest Information Exchange in Europe: A Policy Analysis from an Actor-Centred Institutionalist Approach. *Forests* **2018**, *9*, 129. [CrossRef]
14. Alberdi, I.; Vallejo, R.; Álvarez-González, J.G.; Condés, S.; González-Ferreiro, E.; Guerrero, S.; Hernández, L.; Martínez-Jauregui, M.; Montes, F.; Oliveira, N.; et al. The Multi-Objective Spanish National Forest Inventory. *For. Syst.* **2017**, *26*, e04S. [CrossRef]
15. Næsset, E. Airborne Laser Scanning as a Method in Operational Forest Inventory: Status of Accuracy Assessments Accomplished in Scandinavia. *Scand. J. For. Res.* **2007**, *22*, 433–442. [CrossRef]
16. Breidenbach, J.; Astrup, R. Small Area Estimation of Forest Attributes in the Norwegian National Forest Inventory. *Eur. J. Forest Res.* **2012**, *131*, 1255–1267. [CrossRef]
17. Gobakken, T.; Næsset, E.; Nelson, R.; Bollandsås, O.M.; Gregoire, T.G.; Ståhl, G.; Holm, S.; Ørka, H.O.; Astrup, R. Estimating Biomass in Hedmark County, Norway Using National Forest Inventory Field Plots and Airborne Laser Scanning. *Remote Sens. Environ.* **2012**, *123*, 443–456. [CrossRef]
18. Næsset, E.; Gobakken, T.; Bollandsås, O.M.; Gregoire, T.G.; Nelson, R.; Ståhl, G. Comparison of Precision of Biomass Estimates in Regional Field Sample Surveys and Airborne LiDAR-Assisted Surveys in Hedmark County, Norway. *Remote Sens. Environ.* **2013**, *130*, 108–120. [CrossRef]
19. Magnussen, S.; Mandallaz, D.; Breidenbach, J.; Lanz, A.; Ginzler, C. National Forest Inventories in the Service of Small Area Estimation of Stem Volume. *Can. J. For. Res.* **2014**, *44*, 1079–1090. [CrossRef]
20. Maack, J.; Lingenfelder, M.; Weinacker, H.; Koch, B. Modelling the Standing Timber Volume of Baden-Württemberg—A Large-Scale Approach Using a Fusion of Landsat, Airborne LiDAR and National Forest Inventory Data. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *49*, 107–116. [CrossRef]
21. Ståhl, G.; Saarela, S.; Schnell, S.; Holm, S.; Breidenbach, J.; Healey, S.P.; Patterson, P.L.; Magnussen, S.; Næsset, E.; McRoberts, R.E.; et al. Use of Models in Large-Area Forest Surveys: Comparing Model-Assisted, Model-Based and Hybrid Estimation. *For. Ecosyst.* **2016**, *3*, 5. [CrossRef]
22. Hill, A.; Massey, A.; Mandallaz, D. The *R* Package Forestinventory: Design-Based Global and Small Area Estimations for Multiphase Forest Inventories. *J. Stat. Soft.* **2021**, *97*, 1–40. [CrossRef]
23. Mandallaz, D. A Three-Phase Sampling Extension of the Generalized Regression Estimator with Partially Exhaustive Information. *Can. J. For. Res.* **2014**, *44*, 383–388. [CrossRef]
24. Mandallaz, D.; Breschan, J.; Hill, A. New Regression Estimators in Forest Inventories with Two-Phase Sampling and Partially Exhaustive Information: A Design-Based Monte Carlo Approach with Applications to Small-Area Estimation. *Can. J. For. Res.* **2013**, *43*, 1023–1031. [CrossRef]
25. Georgakis, A.; Papageorgiou, V.E.; Stamatellos, G. A New Approach to Small Area Estimation: Improving Forest Management Unit Estimates with Advanced Preprocessing in a Multivariate Fay–Herriot Model. *For. Int. J. For. Res.* **2024**. [CrossRef]
26. Gabler, K.; Schadauer, K. *Methods of the Austrian Forest Inventory 2000/02—Origins, Approaches, Design, Sampling, Data Models, Evaluation and Calculation of Standard Error*; BFW-Berichte; Schriftenreihe des Bundesforschungs- und Ausbildungszentrums für Wald, Naturgefahren und Landschaft: Wien, Austria, 2008.
27. Gschwantner, T.; Berger, A.; Büchsenmeister, R.; Hauk, E. Austria. In *National Forest Inventories*; Vidal, C., Alberdi, I.A., Hernández Mateo, L., Redmond, J.J., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 135–157. ISBN 978-3-319-44014-9.
28. Berger, A.; Gschwantner, T.; Schadauer, K. The Effects of Truncating the Angle Count Sampling Method on the Austrian National Forest Inventory. *Ann. For. Sci.* **2020**, *77*, 16. [CrossRef]
29. Bundesamt für Eich- und Vermessungswesen WMS Höhe DGM ALS. 2020. Available online: https://doi.org/10.48677/99D7B0CC-7AA8-4217-B56F-667EF52E9C24 (accessed on 13 March 2020).

30. Mandlburger, G.; Wenzel, K.; Spitzer, A.; Haala, N.; Glira, P.; Pfeifer, N. Improved Topographic Models via Concurrent Airborne Lidar Anddense Image Mathcing. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **2017**, *IV-2/W4*, 259–266. [CrossRef]
31. Schadauer, T.; Karel, S.; Loew, M.; Knieling, U.; Kopecky, K.; Bauerhansl, C.; Berger, A.; Graeber, S.; Winiwarter, L. Evaluating Tree Species Mapping: Probability Sampling Validation of Pure and Mixed Species Classes Using Convolutional Neural Networks and Sentinel-2 Time Series. *Remote Sens.* **2024**, *16*, 2887. [CrossRef]
32. BFW Waldkarte. Available online: https://waldinventur.at/?x=1486825&y=6059660&z=7.14944&r=0&l=1111#/map/1/mWaldkarte/Bundesland/erg9 (accessed on 31 March 2023).
33. Särndal, C.-E.; Swensson, B.; Wretman, J. *Model Assisted Survey Sampling*; Springer Series in Statistics; Springer: New York, NY, USA, 1992; ISBN 978-0-387-40620-6.
34. Kilian, W.; Starlinger, F.; Müller, F. Die forstlichen Wuchsgebiete Österreichs - Eine Naturraumgliederung nach waldökologischen Gesichtspunkten. In *FBVA-Berichte*, 1st ed.; Forstliche Bundesversuchsanstalt: Wien, Austria, 1993.
35. Berger, A.; Gschwantner, T.; McRoberts, R.E.; Schadauer, K. Effects of Measurement Errors on Individual Tree Stem Volume Estimates for the Austrian National Forest Inventory. *For. Sci.* **2014**, *60*, 14–24. [CrossRef]
36. McRoberts, R.E.; Næsset, E.; Gobakken, T. Inference for Lidar-Assisted Estimation of Forest Growing Stock Volume. *Remote Sens. Environ.* **2013**, *128*, 268–275. [CrossRef]