Article

Exploring the potential of machine learning for modelling growth dynamics in an uneven-aged forest at the level of diameter classes: a comparative analysis of two modelling approaches

Gaspard Dumollard [1,*]

[1] Bern University of Applied Sciences BFH, School of Agricultural, Forest and Food Sciences HAFL; gaspard.dumollard@bfh.ch
* Correspondence: gaspard.dumollard@bfh.ch

Abstract: Growth models of uneven-aged forests on the diameter class level can support silvicultural decision making. Machine learning brings added value to the modeling of dynamics at the stand or individual tree level based on data from permanent plots. The objective of this study is to explore the potential of machine learning for modeling growth dynamics in uneven-aged forests at the diameter class level based on inventory data from practice. Two main modeling approaches are conducted and compared: i) fine-tuned linear models differentiated per diameter class, ii) an artificial neural network (multilayer perceptron) trained on all diameter classes. The models are trained on the inventory data of the Canton of Neuchâtel (Switzerland), which are area-wide data without individual tree-level growth monitoring. Both approaches produce convincing results for predicting future diameter distributions. The linear models perform better at the individual diameter class level with test R^2 typically between 50 and 70% for predicting increments in the numbers of stems at the diameter class level. From a methodological perspective, the multilayer perceptron implementation is much simpler than the fine-tuning of linear models. The linear models developed in this study achieve sufficient performance for practical decision support.

Keywords: Uneven-aged forest management; Forest growth modelling; Machine learning; Diameter distribution; Silvicultural decision support

1. Introduction

Uneven-aged forests are generally defined as forests whose stands contain several cohorts of trees of different ages, which results in a relative structural heterogeneity of these stands. Despite the diversity of uneven-aged management methods, the management of these stands is often associated with the maintenance of a targeted distribution of diameters [1]. A good understanding of the dynamics within an uneven-aged stand at the global level but also at the level of diameter classes then allows for better planning of cuts, in the sense of choosing the timing and the intensity of the next cut [2], including the intensity of the cut by diameter class. Growth models for uneven-aged stands can therefore support silvicultural decision making.

The literature on existing uneven-aged forest growth models identifies three possible resolutions for this type of model: the whole stand, diameter classes and individual trees [3,4]. It is also possible to make a distinction between empirical and mechanistic (also called process-based) models [3]. This distinction is theoretically relevant, although there is a gradient between these two archetypes, with models estimated on the basis of empirical data but whose structure is to some extent inspired by theoretical mechanisms of forest growth [5].

Stand-level models aim at predicting the increment of a key stand-level variable, typically the increment of the standing volume or basal area. The recent literature has seen the publication of several such studies based on machine learning methods and data from permanent plots [6,7,8]. In particular, some of these studies use artificial neural networks among other regressors [7,8]. All of these models achieve good predictive performance and can be useful in practice, e.g., in management systems where harvested volumes are planned on the basis of increments. However, they do not provide information on stand structure.

Models at the individual tree level aim to predict individual diameter increments. Several recent studies have approached this topic from a machine learning perspective [4,9], for example by testing several types of models including an artificial neural network [9]. These models perform well and give a maximum of details on growth within a stand, but individual monitoring data on diameter growth, i.e. data from permanent plots, are needed to train them. When this is the case, the data are usually available in large numbers, which fits well to machine learning approaches. For example, one of these studies [4] is based on 16,619 observations from the monitoring of 20 permanent plots.

Models based on diameter classes generally aim to predict increments in the number of stems per diameter class. Some studies [10,5] propose for example an empirical modeling of radial increments per diameter class based on inventory data using a linear model. These radial increments are then integrated into a process-based model to simulate the dynamics of the passage of trees between two successive classes, as well as the recruitment into the first diameter class. Numerous other studies [11,12,13,14] use an approach that is quite similar in principle but generalized and formalized. Passage rates are modeled directly (and empirically with linear models) and the increment in the number of stems in a given diameter class potentially depends on the respective numbers of stems in all other diameter classes. This type of model accounts for passages between two non-successive classes, competition effects between diameter classes, as well as the influence of the diameter distribution on recruitment. These models are formalized as matrix equation systems allowing the recursive simulation of stand dynamics, sometimes called a matrix model or transition matrix model [3]. The predictive performances of those models are limited compared to recent machine learning models developed on the stand or individual tree levels.

The canton of Neuchâtel in Switzerland is historically and professionally recognized as the Swiss high place of the selective felling forest management, which is implemented on a tree-by-tree basis (see section 2.2 for further details). In the canton of Neuchâtel, selective felling management is implemented through the so-called control method (Méthode du contrôle) [15,3,16]. This method is based on periodic and complete inventories of stands as well as on a direct control of all harvests. Some studies [10,5] use an extract of these data to develop his models. Some other studies [12,14] use data from the French Jura mountain where a similar management and controlling system prevail. These inventory data from the practice are non-experimental and are not based on individual tree monitoring, which makes their analysis more complex than data from permanent plots.

The objective of this study is to explore the potential of machine learning for modeling growth dynamics at the level of diameter classes from the available inventory data of the Canton of Neuchâtel. From a practical point of view, the aim is to determine if a model of this type can predict the evolution of the distribution of diameters that is specific to a given stand with sufficient accuracy to serve as decision support for the planning and implementation of cuts.

To achieve this goal, a machine learning workflow is implemented for data preparation and preprocessing, modeling, and evaluation of the models for their out-of-the-box predictive performances. This workflow allows to iteratively refine the data preprocessing and modeling steps. As to the modeling step, two main approaches are compared: i) an approach based on fine-tuned linear models differentiated per diameter class, analogous in spirit to previous matrix models [11], ii) an approach based on an artificial neural network trained on all diameter classes. The potential of these two types of models for predicting a future state at the stand level is analyzed in absolute terms but they are also compared with each other and with previous models found in the literature. The best model is also evaluated for its ability to be used in practice, i.e. for its ability to predict the evolution of the overall diameter distribution at the stand level, and to predict increments of important aggregated variables.

2. Materials and Methods

2.1. Geographical and climatic context in the canton of Neuchâtel

The canton of Neuchâtel, in Switzerland, is located in the Jura Mountain range. Its altitude stretches between 429m and 1552m. Its climate is humid continental [17]. In terms of altitudinal zonation, the territory of the canton is currently located in the submontane, lower montane and upper montane zones [18]. Simulations of two possible climatic futures [18] also tend to show, in the canton, a transition from the lower and upper montane zones to the submontane zone. The southern part of the Canton, located at a lower altitude and in the minority, would even enter the foothill zone.

2.2. Brief summary on forest management and forest data collection in the canton of Neuchâtel

The selection felling management that is historically applied in the canton of Neuchâtel is an uneven-aged management method where silviculture is implemented on a tree-by-tree basis. This type of management is based on periodic and situative cuts. The selection of trees to be harvested is the result of multiple determinants (wood harvesting, fostering tree vitality, fostering wood quality, continuous regeneration) but these cuts affect all diameter classes and are characterized by a fairly constant harvesting intensity that depends on the periodicity of interventions [2]. At the stand level, selective felling management results in the transformation to or the maintenance of a state of equilibrium that can be expressed in terms of diameter distribution within a given stand. A typical example of a diameter distribution at the stand level is shown in Figure 1.

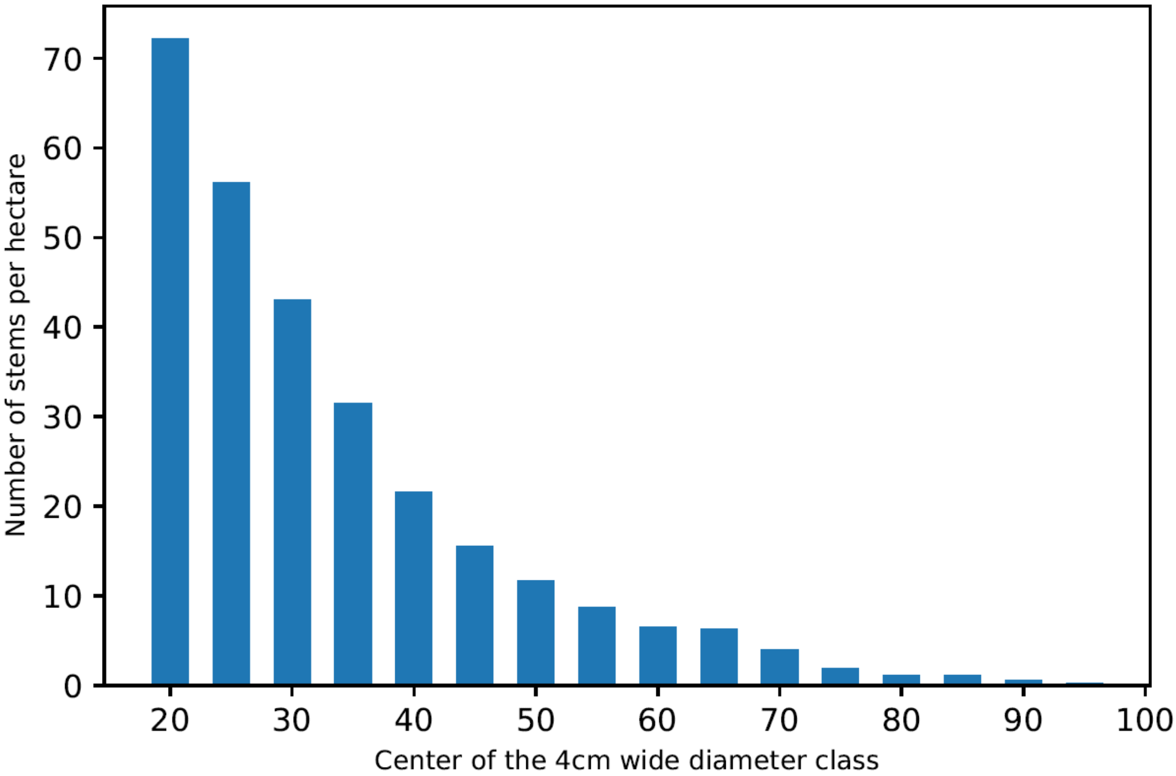

Figure 1. Example of a distribution of diameters in a forest division under selective felling management (division number 2202, inventory of 1971).

Information on the state of a stand, including the diameter distribution, also helps with the planning of the next cuts as management goals on the stand level can be expressed as a desired state of equilibrium in terms of species mixture, standing volume, stand structure [2].

Cuts take place periodically, typically every 7 to 12 years (in the Canton of Neuchâtel). The maintenance of the equilibrium is made possible by the fact that the harvesting of given trees allows for a "continuous" natural regeneration within the stand and allows for the further growth of trees in the intermediate strata.

Management by selective felling on a tree-by-tree basis is generally based on species that are shade tolerant, at least in their youth, typically but not exclusively white fir, spruce and beech. Since regeneration occurs in small patches and trees continue to grow in intermediate strata in the shade of larger trees, heliophilic species are indeed not suitable for this type of management.

In the canton of Neuchâtel, selective felling management is implemented through the so-called control method (*Méthode du contrôle*) [15,3,16] since the end of the 19th century. This method is based on several important elements: complete inventory of the stands (in particular individual diameters and species) between two successive cuts, direct control (i.e. inventory) of all harvests (as well as forced harvests), definition of parcels with fixed boundaries as management units called divisions (a term which we use hereafter in lieu of stand). The diameters are recorded by class of 5cm width and with an inventory threshold of 17.5cm. The classes are numbered from 1 to 24 and the center of the class in cm is then obtained by the following formula ((#class - 1) * 5 + 20). From the perspective of selective felling management, the purpose of collecting such systematic and detailed information is to be able to accurately characterize the silvicultural state of a division, particularly with respect to the diameter distribution, so as to better plan the next cuts. This information also allows for a detailed monitoring of forests at the division or landscape level over time and thus helps in better forest planning at the landscape level. There

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 July 2022 doi:10.20944/preprints202207.0285.v1

is no direct data on recruitment and natural mortality (excluding windfall cases) that is collected through this control method. The application of this method in Neuchâtel allowed the collection of the large data set used in this study.

2.3. Data and data preprocessing

This study is based on a standard machine learning workflow. The data collection and preprocessing steps are presented in this section. Section 2.4 presents the data sets building and features engineering steps and section 2.5 presents the modeling approaches.

The raw data consist of a record of the absolute number of stems in a division for a given inventory year, species and diameter class. These data are subjected to preprocessing which consists of the following steps.

Step 1. The absolute numbers of stems per division are converted into numbers of stems per hectare. The areas of the divisions, whose boundaries are administratively set and (persistently) fixed, are known and could be found for 98% of the divisions mentioned in the data set. The remaining divisions were discarded. Meanwhile, the main forest site of each division (in terms of area) is integrated into the data set.

Step 2. Only divisions that have been inventoried at least twice are kept, i.e. divisions for which it is possible to calculate increments over at least one period, are kept, which corresponds to 95% of the divisions kept so far. From this step on, the inventory data will be organized by couples (division, inventory year). This couple designates the state of a division at a given date, but it also designates the period between the inventory year and the date of the next inventory to which cuts and windfalls are associated (see step 4) and for which it is possible to determine increments.

Step 3. Only the couples (division, inventory year) whose species are compatible with selective felling management are retained. The three usual species in this management method are white fir, spruce and beech [16], but the presence of other species that are sufficiently tolerant to shade, in the minority or very minority, cannot be excluded. Therefore, the following criteria were defined: the majority species (in terms of basal area) must be among fir, spruce or beech, the 2nd majority species must be among fir, spruce, beech, maple, or ash, and the 3rd majority species must be among the same species or flowering ash or be another deciduous species.

Table 1 is a contingency table of the number of couples (division, inventory year) according to their first and second majority species, before application of step 3. The table reveals that the divisions that meet the filtration criteria of step 3 are overwhelmingly in the majority, reflecting the importance of the selective felling management in the Canton of Neuchâtel. The table also reveals that most of the situations that do not meet the defined criteria correspond to couples (division, inventory year) where the 1st or 2nd majority species is oak. These situations may correspond to even-aged stands preferentially located in the south of the canton on less elevated grounds and on forest sites that are better adapted to oak.

Table 1. Contingency table of couples (division, inventory year) according to the 1st and 2nd majority species (for simplicity of representation, the 2nd majority species which had less than 10 couples for all 1st majority species combined are not reported in this table).

2nd majority species									
Spruce	Fir	Beech	Oak (sessile)	Scots pine	Maple (sycamore)	Ash	Other deciduous	Aspen	Other resinous

1st	**Fir**	1412	0	544	5	5	3	1	11	15	0
majority	**Spruce**	0	1053	444	15	6	85	7	23	8	0
species	**Beech**	201	317	0	143	19	25	5	2	0	2
	Oak (sessile)	2	14	124	0	5	0	0	0	0	8
	Scots pine	5	16	18	7	0	0	0	2	0	0
	Larch	1	0	0	0	0	0	0	0	0	0
	Other resinous	1	0	14	9	0	0	0	4	0	0
	Other deciduous	3	12	0	0	3	0	0	0	0	0
	Black pine	3	0	0	3	2	0	0	0	0	0
	Poplar	0	0	0	0	1	0	2	0	0	0
	Ash	4	0	0	1	0	0	0	0	0	0

Step 4. The inventory data are merged with the respective data on decided cuts and windfalls, which are formatted in the same way. The data on cuts and windfalls are associated with couples (division, inventory year) when the cuts and windfalls occur in this division and between the inventory year and the date of the next inventory.

Step 5. The temporal gap between two successive inventories in the same division is variable, see Figure 2.

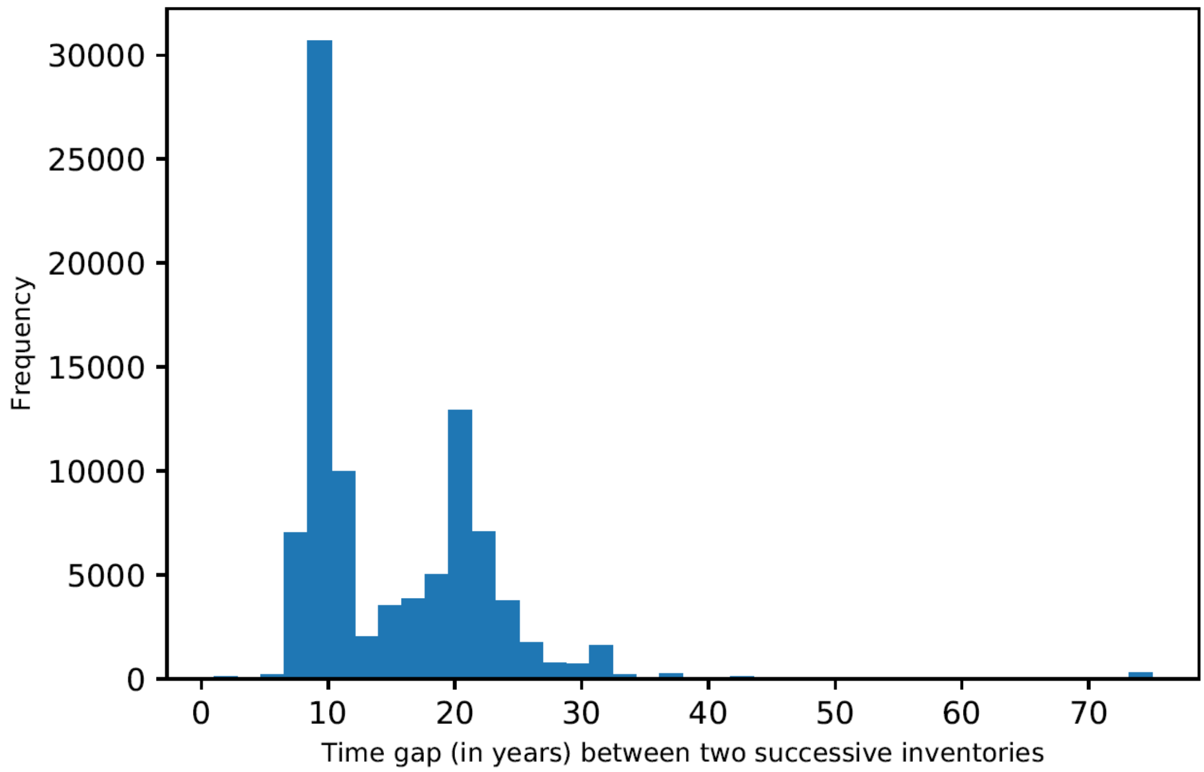

Figure 2. Histogram of the time gap between two successive inventories in a given division.

There are a large number of successive inventories that are 7 to 12 years apart but also many successive inventories that are 20 to 24 years apart. In the first case, the inventories were carried out before and after the same cut and correspond therefore to the standard application of the control method. In the second case, there are several cuts between two successive inventories. The standard application of the control method is

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 July 2022 doi:10.20944/preprints202207.0285.v1

actually historical in the Canton of Neuchâtel, but it has since been adapted and the current legislation provides for an inventory at least every 25 years for each division, which corresponds to 2 or 3 cutting cycles. In some divisions of historical or technical importance, inventories are maintained on the basis of one inventory per cutting cycle.

The goal of this study is to simulate growth dynamics recursively, i.e., we want to predict the state of a division at date $t + \Delta t$ from its state at date t. This is a pragmatic approach that is well suited to the development of the model and its use in practice. The duration Δt does not need to be constant (see Section 3) but it must be within an interval whose definition meets two constraints: i) the possibility of finding data describing dynamics over such periods, ii) the relevance of these durations Δt for a use of the model as a decision support in practice. Figure 2 shows that the available data correspond to durations from 8 to 26 years. However, this interval is too wide for model development. The diameter data are given in 5 cm diameter classes. For a relatively large annual diameter increase of 5mm.year^{-1}, a tree takes 10 years to completely cross a diameter class. Over a period of 20 years and more, with such an increase, a tree can pass through two or more diameter classes. The growth mechanisms in the sense of recursive modeling cannot therefore be the same over these two types of period. We choose to focus on durations Δt of up to 12 years. This interval allows the model to be used over a sufficiently long period to assist in planning the next cut, and it is also possible to use the models recursively to simulate dynamics over longer periods. Data corresponding to gaps of more than 12 years between two successive inventories are left out, so that only 52% of couples (division, inventory year) that had been retained until then are kept.

Step 6. Much of the harvest data is not available. All the data on the cuts, some of which are very old, have not yet been compiled and digitized. This gap in the data leads us to keep only the couples (divisions, year of inventory) for which cuts and windfalls are known, i.e. 38% of the couples that had been kept until now. In addition, a few outliers (in very small numbers) that stand out by cutting intensities that are much too strong for the selective felling management are also set aside.

Step 7. The number of trees in the upper diameter classes is small or even negligible and therefore does not allow a reliable modeling of the growth mechanisms among these classes. Therefore, the data must be truncated at a given diameter threshold. The number of trees that are beyond diameter class 16 (class 24 is the maximum class), i.e. trees with a diameter strictly greater than 97.5cm, are relatively rare compared to the lower classes, they represent about 0.07% of the trees remaining in the data set at this stage. Such a threshold is also high enough that it is not a problem for practical use of the model, so we choose this threshold to truncate the data. Nevertheless, simply ignoring trees beyond class 16 could be problematic in some cases. These trees are indeed few in number but very large and can therefore have a non-negligible impact on the growth of other trees. For this reason, only couples (division, inventory year) that meet the following criteria are retained: a maximum of 0.5 trees per ha above class 16 in the division and a maximum of 0.125 trees per ha above class 16 for cuts or windfalls. These criteria were defined partly arbitrarily but in the search for a compromise between filtering potentially biased data and keeping a sufficient volume of data. At the end of this step, 77% of the couples (division, year of inventory) that had been kept until then remain.

In the end, after this preprocessing, there are data on 580 couples (division, inventory year) and for each couple, there are data on 16 diameter classes. In spite of a very strong filtration, the volume of data remains important for a study of this type. The data is moreover complete and consistent.

2.4. Data buildings and features engineering

The objective is to develop a model to predict the gross annual increment in the number of stems per hectare and per diameter class (if there were neither cuts nor windfalls, hence the gross increments), noted ΔN_a. ΔN_a will be the endogenous variables in our

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 July 2022 doi:10.20944/preprints202207.0285.v1

models. They can be calculated as follows (with Δt the duration in years between the two successive inventories):

$$\Delta N_d = \frac{(N_{after,d} - N_d + N_{cut,d} + N_{windfall,d})}{\Delta t}. \quad (1)$$

Exogenous variables (or features) consist of primary variables stemming from the preprocessing of the data and secondary variables built upon those primary variables. Primary variables consist of the number of stems per hectare and per diameter class d at the time of the first inventory N_d and at the time of the following inventory $N_{after,d}$, as well as the number of stems cut or windfallen per hectare and per diameter class between these two inventories (respectively $N_{cut,d}$ and $N_{windfall,d}$), all of which are differentiable by species.

Following secondary variables are built (all of which are expressed per hectare):

- The numbers of stems cut and windfallen are expressed on an annual basis:

$$N_{annual\ cut,d} = \frac{N_{cut,d}}{\Delta t}, \quad (2)$$

$$N_{annual\ windfall,d} = \frac{N_{windfall,d}}{\Delta t}. \quad (3)$$

- The number of stems cut or windfallen annually in the overlying classes (as an indicator of the regulation of competition for a given diameter class):

$$N_{overlying\ annual\ cut,d} = \sum_{d'=d+1}^{16} N_{annual\ cut,d'}, \quad (4)$$

$$N_{overlying\ annual\ windfall,d} = \sum_{d'=d+1}^{16} N_{annual\ windfall,d'}. \quad (5)$$

- Variables N_d are differentiated by resinous/deciduous categories: $N_{resinous,d}$ et $N_{deciduous,d}$,
- Corresponding basal areas are calculated (from the N variables and the centers of diameter classes): G_d, $G_{resinous,d}$ et $G_{deciduous,d}$,
- The total number of stems per hectare N_{tot} and the total basal area per hectare G_{tot} are calculated too,
- The basal area overlying each diameter class is used as an indicator of the competition to which a diameter class is subject. This variable, denoted Gcum (for cumulative basal area), is defined as the sum of the basal areas of every tree larger than the considered diameter class [5,4], in our case:

$$G_{cum.d} = \sum_{d'=d+1}^{d_{max}} G_{d'}. \quad (6)$$

These variables are indicators of density, competition or competition regulation at the division or diameter class level, or indicators of the proportion of resinous trees, all of which can potentially provide additional insight into forest growth.

2.5. Modelling approaches

This study is based on two main modeling approaches: i) fine-tuned linear models, ii) an artificial neural network (a multilayer perceptron regressor). The criterion for refining and comparing models is a coefficient of determination R^2 calculated on a set of test data that are not used for training the models. This test data set accounts for 20% of total available data. Test data are selected among divisions, not at the level of triplets (division, inventory year, diameter class). Prediction performances are therefore measured on divisions that are completely unknown to the models, which is crucial. The principle of parsimony which, with equal or almost equal performance, leads to favouring the simplest models is also applied, as well as the search for models that are consistent from the point of view of forest growth.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 July 2022 doi:10.20944/preprints202207.0285.v1

Algorithms for data transformation, model training and model evaluation are derived from the sklearn library (version 0.23.1, and version 1.0.2 for the SplineTransformer). Other important libraries are pandas (version 0.24.2) and numpy (version 1.19.5).

2.5.1. Detailed method for linear modeling

Linear models are refined following a usual machine learning workflow so as to find the best possible combination of features, data transformation and (linear) regressor. This type of model provides transparent and easily interpretable results from the point of view of forest growth. Moreover, their linearity and the relatively limited number of parameters make them resilient to overfitting problems when sufficient data are available.

The best results with linear models are obtained by training a separate model for each diameter class. These models are of the following form for any diameter class $1 \leq d \leq 16$ and for each couple (division, inventory year) i:

$$\Delta N''_{d,i} = \beta_0^d + \beta_{d-2}^d \cdot N'_{d-2,i} + \beta_{d-1}^d \cdot N'_{d-1,i} + \beta_d^d \cdot N'_{d,i} + \beta_{d+1}^d \cdot N'_{d+1,i} + \cdots + \beta_{d_{\max(d)}}^d \cdot N'_{d_{\max(d)},i} + \beta_{cut}^d \cdot N'_{annual\ cut,d,i} + \beta_{windfall}^d \cdot N'_{annual\ windfall,d,i} + \varepsilon_{d,i}, \quad (7)$$

The β_j^d are the regression coefficients corresponding to the model of diameter class d. These models are parsimonious in exogenous variables. Variables were selected by backward elimination from the set of variables defined in section 2.4. The elimination of engineered features in the sense of the density and competition indicators defined in section 2.4, and the indistinction between coniferous and deciduous trees did not lead to any (or very marginal) loss of performance in terms of test R^2. These variables did not provide sufficient informational gain to be retained in the model and the application of the parsimony principle allowed for their elimination.

Some basic variables (numbers of stems per diameter class N_d) could also be eliminated. When modeling the increment in the number of stems for a diameter class d (ΔN_d), among underlying classes, only the two preceding ones ($d-1$ and $d-2$) provide relevant information. Some stems in these classes may indeed enter class d during the growth periods considered in the data, typically 7 to 12 years. Trees in classes lower than $d-2$ do not have the time to grow as much (at least 10cm in diameter) over a period of 7 to 12 years. The corresponding variables could therefore be eliminated since they only captured statistical noise (overfitting). In the same way, it is not always necessary to keep all the variables corresponding to numbers of stems in classes higher than a given class, i.e. d_{max} is not necessarily equal to 16 in all cases (see equations 7). d_{max} depends on the model for a given diameter class d. An interpretation of these d_{max} values is given below after the detailed presentation of results. Finally, among all the variables on the numbers of stems cut or windfallen, the only ones that provide relevant information on the dynamics of a given diameter class are the numbers of stems cut or windfallen in this same diameter class. This is at least the case for the growth periods modelled here (typically 7 to 12 years). Over longer periods, these variables could certainly play a role.

Despite their relative simplicity, linear models could be significantly improved by transforming the variables. The exogenous variables are each subjected to a standard scaling (noted by an apostrophe ' in equations 7). This usual transformation is obtained by removing mean and scaling to unit variance. The endogenous variable is subjected to a Yeo-Johnson transformation [19]. This transformation is noted by a double-apostrophe '' in equations 7. This non-linear transformation allows to reduce the skewness of a distribution and solves the heteroscedasticity problem that was visible on the residuals plots. This transformation is similar to the Box-Cox transformation [20] which is better known but it accepts negative values as an argument, which is necessary in our case. The Yeo-Johnson transformation improved the performance of the models of classes 1 to 12 (as a reminder, classes are 5cm wide and the center of a class in cm = (#class - 1) * 5 + 20). For classes 13 to 16, this transformation did not improve performances, and even deteriorated them. The transformation was therefore applied to classes 1 to 12. These two

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 July 2022 doi:10.20944/preprints202207.0285.v1

transformations (Standard scaling of the exogenous variables and Yeo-Johnson transformation of the endogenous variable) were performed separately for each model. B-splines transformations [21,22] on the exogenous variables were also tested. This type of non-linear transformation, related to the Generalized additive modelling (GAM) approach, allows for non-linearity effects between the respective exogenous variables and the endogenous variable. The use of this transformation did not produce strong and unambiguous results (performances were slightly better for half of the diameter classes and slightly worse for half of the diameter classes with no particular order among these diameter classes). By principle of parsimony and to keep a common structure to our models, this transformation was left aside.

Finally, best performances were obtained using a Huber loss function [23,24]. This loss function is quadratic below a given error threshold but linear above this threshold, which allows to limit the weight of points that stand out (e.g. residual outliers). Moreover, a regularization term of type L2 is added to the loss function [25], which allows to control the overfitting in a systematic way. Note that this regularization term makes the models non-linear in the strict sense. A Stochastic Gradient Descent regressor [25] is used to train the model and the hyperparameters related to the Huber loss function and the regularization term are determined by gridsearch and cross-validation.

2.5.2. Detailed method with multilayer perceptrons

The use of a multilayer perceptron [26,27], which is a nonlinear model, allows a more generic and flexible modeling approach. A multilayer perceptron with at least one hidden layer (hence multilayer) can indeed serve as a universal approximator according to the universal approximation theorem [28], in the sense that it can approximate with a finite number of neurons any continuous function (on particular subdomains, see [28] for further details). The scikit-learn library [27] provides such a regressor ready-to-use. In this case, potential exogenous variables can be all pooled together because this model takes into account non-linear effects and interaction effects. However, this type of model, usually based on a large number of parameters (at least larger than linear models), requires large volumes of data to avoid overfitting. In addition, results (the values of parameters) are harder to interpret. The volume of data used in this study is comparatively large but it is not large in the sense of big data. One of the objectives of this second modeling approach is to analyze the potential of the multilayer perceptron for this modeling problem based on a large but probably limiting data set.

The multilayer perceptron regressor was trained to directly predict the gross numbers of stems (without the removal effect from cuts or windfalls) by diameter class at the end of a growth period of duration Δt: $N_{after\ gross,d}$. This direct approach, without going through increments, resulted in better predictive performances. Only one model was trained instead of separate models for each diameter class. Therefore, the training of this model is based on 9280 entries, which is a comparatively large volume of data for this type of application.

The exogenous variables presented in section 2.4 were all retained and several variables were even added: the diameter (center of the class) to distinguish the class under discussion, the length of the growth period Δt which differs from one couple (division, inventory year) to another, as well as the forest site. The model hyperparameters : the number of hidden layers, the number of neurons in hidden layers, and the regularization parameter [27], are tuned by gridsearch and cross-validation. The model is trained twice on two different train-test-splits. The best topology (among those tested) had two hidden layers with 10 neurons each. The optimal value of the regularization parameter (L2) optimally defined by the gridsearch is 100 in both cases, an intermediate value among the possible values.

3. Results

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 July 2022 doi:10.20944/preprints202207.0285.v1

3.1. Results of the linear modeling

The performances obtained for the different linear models are presented in Table 2. As the train-test-split is random, it is repeated 10 times and the table therefore presents averaged results. Table 2 gives the test R^2s for the prediction of transformed annual increments of the number of stems per hectare $\Delta N''_{predict,d,i}$. The table also shows the test R^2s for predicting gross (i.e. if neither cuts nor windfalls occur) stem counts at the end of the growth period, i.e., based on the errors between the $N_{after\ gross,d,i}$ and the $N_{predict\ after\ gross,d,i}$ [6]:

$$N_{after\ gross,d,i} = N_{after,d,i} + N_{cut,d,i} + N_{windfall,d,i}, \tag{8}$$

$$N_{predict\ after\ gross,d,i} = N_{d,i} + \Delta N_{predict,d,i} \cdot \Delta t. \tag{9}$$

Table 2. Performances of the models measured by R2 coefficients for respective prediction of $\Delta N''_{d,i}$ and $N_{after\ gross,d,i}$ (the overall R2 is measured on the whole data set when the models are trained on the whole data set).

Diameter class	**Diameter range**	**Test R^2 for the prediction of $\Delta N''_{d,i}$**	**Test R^2 for the prediction of $N_{after\ gross,d,i}$**	**Overall R^2 for the prediction of $N_{after\ gross,d,i}$**
1	[17.5 ;22.5[	13.6%	82.4%	82.0%
2	[22.5;27.5[	55.2	93.5	93.2
3	[27.5;32.5[	64.5	93.4	93.5
4	[32.5;37.5[	71.6	92.9	92.4
5	[37.5;42.5[	66.9	91.3	91.7
6	[42.5;47.5[	60.5	89.2	91.4
7	[47.5;52.5[	54.7	90.3	92.0
8	[52.5;57.5[	49.0	88.7	90.8
9	[57.5;62.5[	51.2	90.6	90.9
10	[62.5;67.5[	56.1	90.4	90.9
11	[67.5;72.5[	58.1	89.6	89.7
12	[72.5;77.5[	61.5	88.0	88.0
13	[77.5;82.5[	60.6	84.2	85.3
14	[82.5;87.5[	50.5	74.5	79.7
15	[87.5;92.5[	53.8	69.8	72.9
16	[92.5;97.5[	57.3	65.9	70.8
All classes combined			90.3%	96.7%

The plots of the residuals (corresponding to one of the train-test-split) for the 16 models are presented in Figure A1 in the appendix A. A review of the histograms giving the distributions of these residuals (not shown here) shows Gaussian type distributions centered in 0.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 July 2022 doi:10.20944/preprints202207.0285.v1

The results are satisfying except for class 1 ($17.5cm \leq d < 22.5cm$). Since there is no data on the number of stems for diameter classes below 17.5cm, it is impossible to precisely predict the number of trees that will enter class 1 during the growth period, i.e. the recruitment rate of our set of models, hence the limited performance. For the other models, test R^2s range typically from 50 to 70%. These models manage to predict most of the variance of the $\Delta \boldsymbol{N''}_{\boldsymbol{d,i}}$ based on the initial state of a division $N'_{d,i}$ (and on cuts $N'_{annual\ cut,d,i}$ and windfalls $N'_{annual\ windfall,d,i}$) for cases that were not used for training the model. Using these predicted increments (after inverse transformation) to predict stem counts at the end of the growth period also produces satisfying results, mechanically better than the prediction of increments since the initial stem counts are known and largely explain the stem counts 7 to 12 years later. All diameter classes combined, the models predict future stem numbers with a test R^2 of 90.3%. This is a figure that can be compared to the result obtained with the artificial neural network approach (see below).

The models are then re-trained on the whole data set to determine the final regression coefficients. Table 2 presented above shows corresponding R^2s, which are in this case training R^2s. Those training R^2s are very close to the test R^2s presented before, showing an absence of overfitting, which is normal given that the linear models have relatively few variables and have been trained on a fairly large data set (580 observations each). However, those training R^2s are slightly higher, which tends to show that using the whole data set for training the models could further improve performances. The regression coefficients corresponding to the training on the whole data set are presented in Table A1 in the Appendix and are shown in Figure 3 as a heatmap for ease of exposition and interpretation. Because of the transformations undergone by the variables (Yeo-Johnson and standard scaling and for each diameter class separately), the coefficients are difficult to interpret in absolute terms, but their sign can be interpreted, as can their relative values within a given model.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 July 2022 doi:10.20944/preprints202207.0285.v1

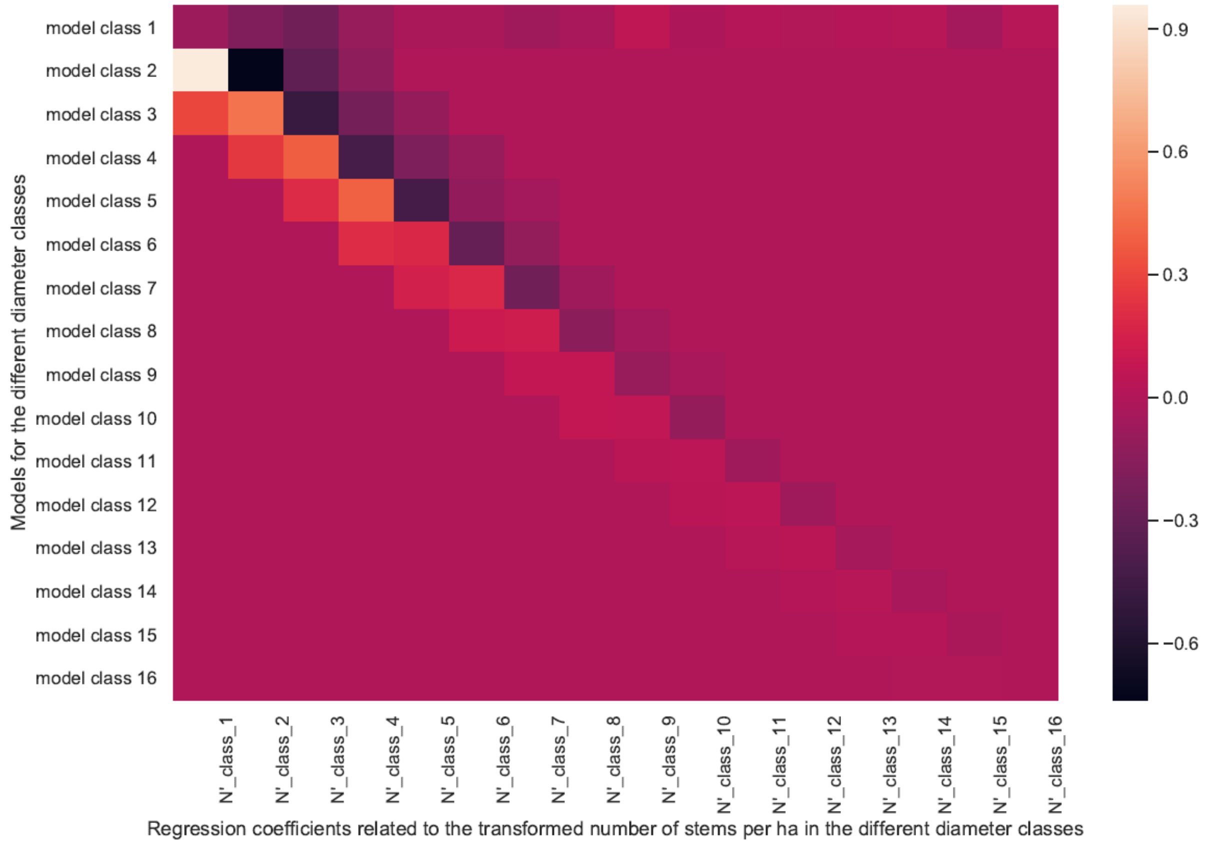

Figure 3. Heatmap of the regression coefficients for the models given by equations 7 (the values are given in Table A1 in the Appendix).

For diameter classes greater than 1 (diameters greater than 22.5cm), a common pattern is visible. This pattern is similar to the results presented in [13], despite the differences due to our data transformations. This pattern is also clearly visible when displaying the coefficients for a given model on a graph, see for example Figure 4 and Figure 5.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 July 2022 doi:10.20944/preprints202207.0285.v1

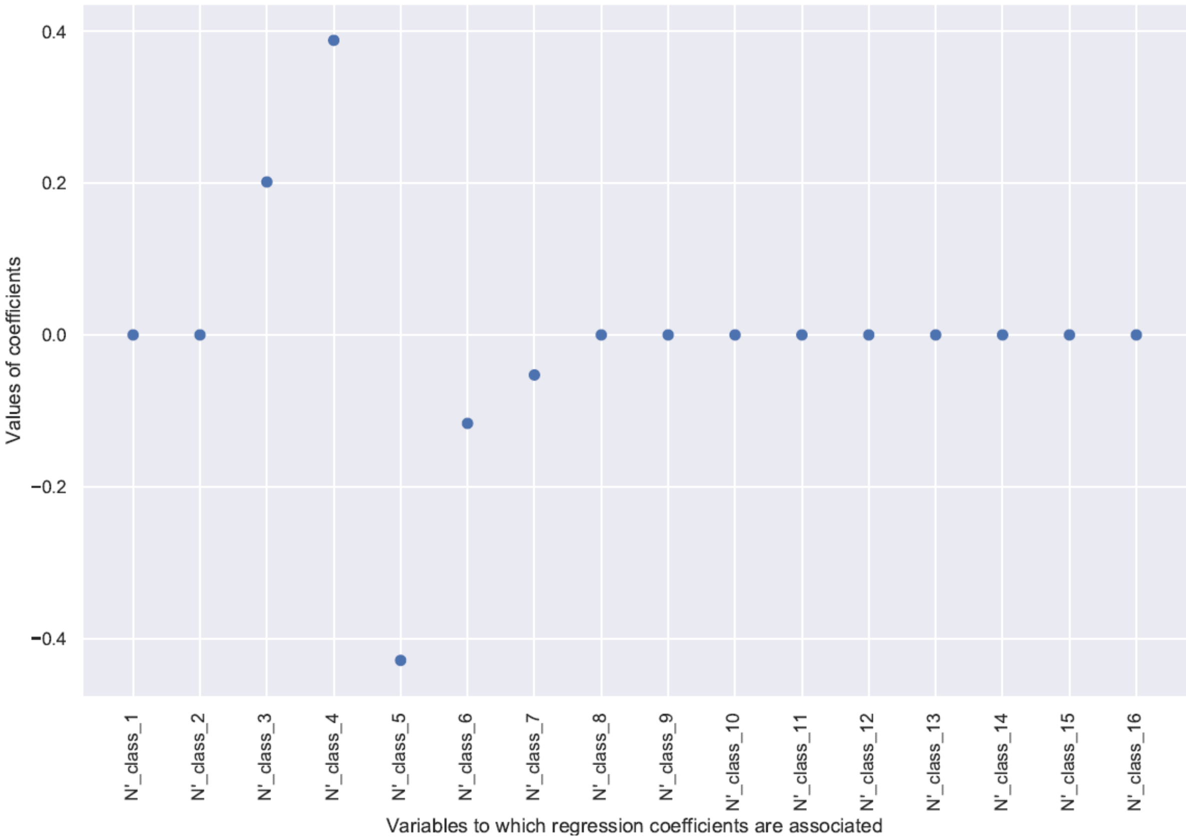

Figure 4. Regression coefficients for the model of diameter class 5, associated with transformed variables N_d'.

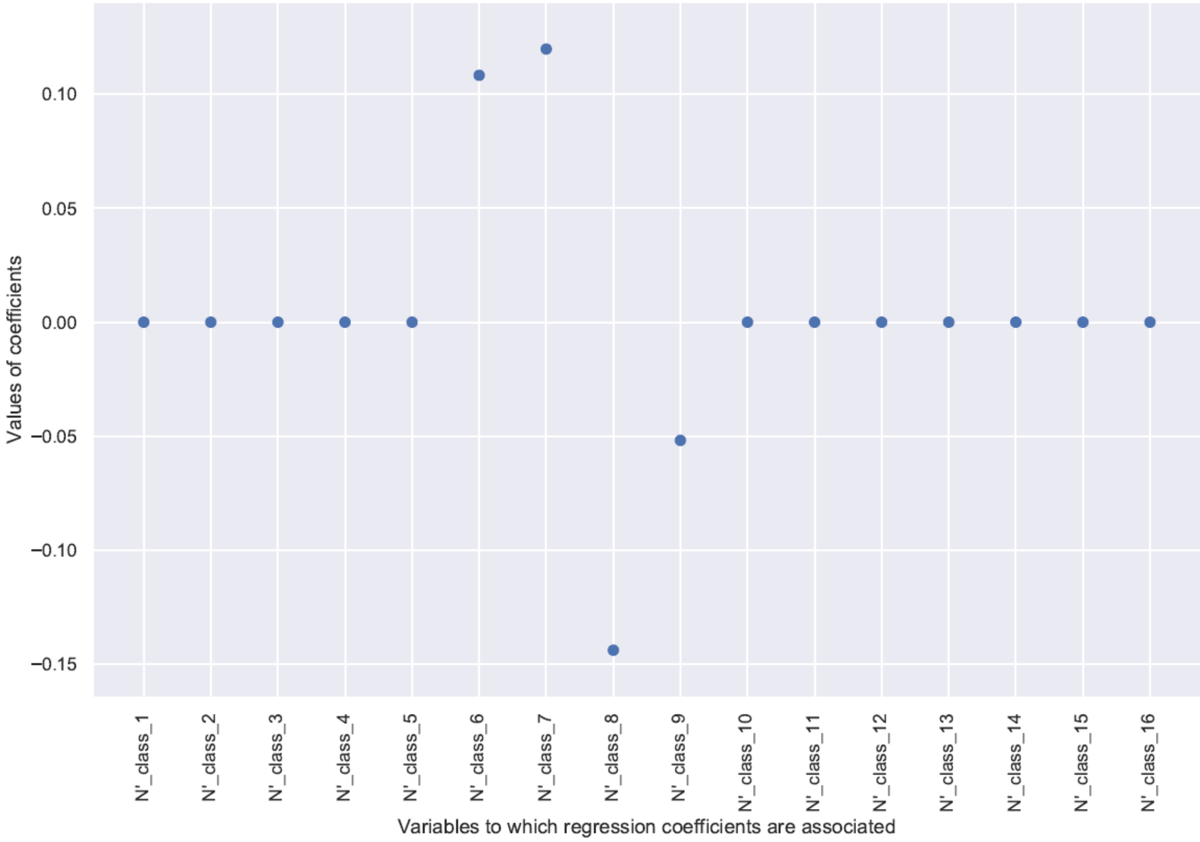

Figure 5. Regression coefficients for the model of diameter class 8, associated with transformed variables N'_d.

The effects of the respective variables are comparable within the same model despite very different numbers of stems depending on the class (high for small classes to very low for large classes) because these variables have been scaled to the unit variance. The increment in the number of stems per ha ΔN_d of a given diameter class d is positively influenced by the number of stems per ha in the two directly underlying classes $(d-1)$ and $(d-2)$. These are the trees whose growth will allow them to move from class $(d-1)$ to d or from class $(d-2)$ to d during the growth period. The fact that the increments are expressed on an annual basis does not change the fact that they are trained and therefore applicable over periods typically between 7 and 12 years.

The number of stems in class d, on the other hand, has a negative influence on the increment of class d. This effect corresponds to the trees that will leave class d and move to the overlying classes during the growth period. The effect of the overlying classes is negative but more difficult to interpret. In general, and somewhat surprisingly, only the directly overlying classes seem to have a slight negative influence on the increment of the number of stems in a given class. It can be assumed that there is some competition between close diameter classes provided that there is some spatial proximity between trees of these different classes. This can indeed be the case in small and intermediate classes. It is more uncommon for the largest trees whose neighborhood at their height is generally more sparse (those trees have already been fostered by previous cuts). Classes much larger than a given diameter class did not seem to play a clear and strong role on the increment in that class, either because of a lack of strong competition between trees of very different diameters (e.g. due to spatial differentiation) or because this competition effect is relatively constant across divisions and that without variability, the models fail to quantify this effect. For this reason, the corresponding variables were not included in the model.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 July 2022 doi:10.20944/preprints202207.0285.v1

The threshold for not including these variables in the model, d_{max} in equations 7, depends on a given diameter class, hence the notation $d_{max}(d)$.

For class 1, the situation is different for the reasons already stated, which have to do with the lack of information on the underlying classes. In this case, the model performed better by keeping all variables corresponding to the overlying classes. Figure 6 shows the value of the regression coefficients associated with variables N'_d (see equations 7) for the diameter class 1 model.

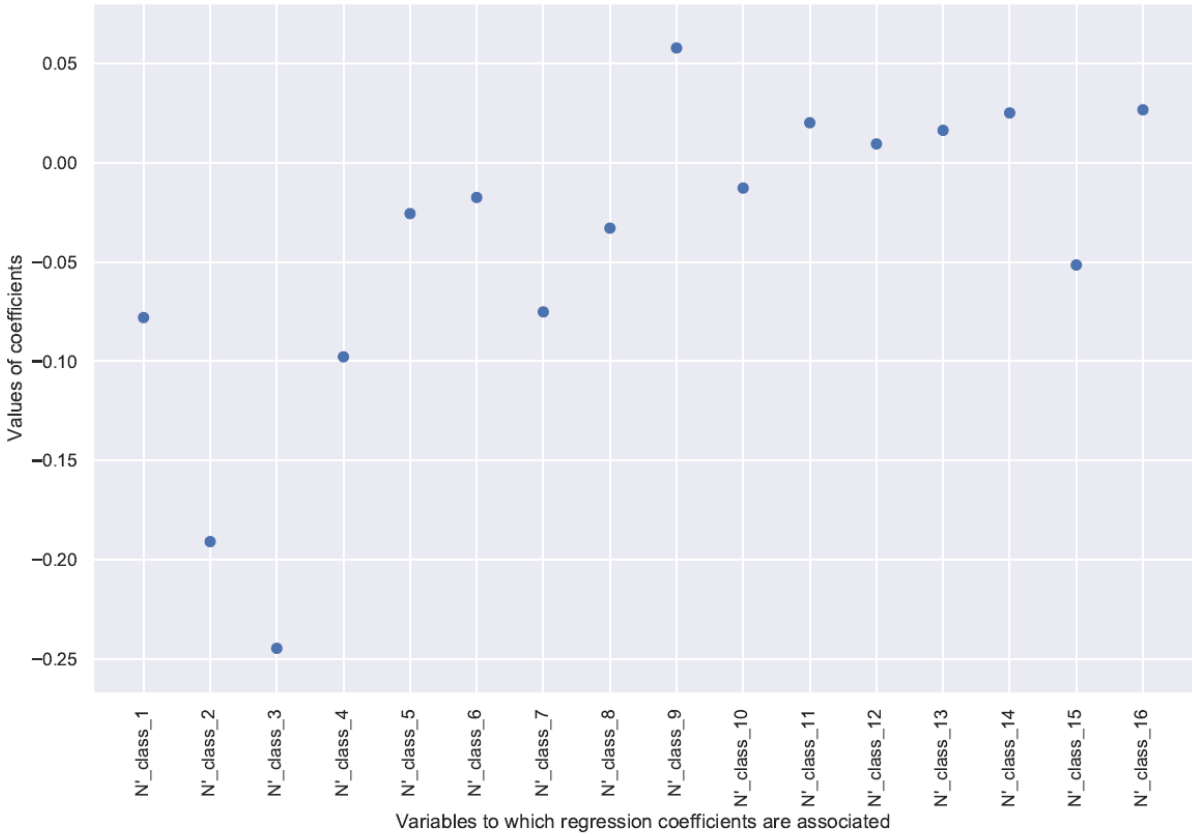

Figure 6. Regression coefficients for the model of diameter class 1, associated with transformed variables N'_d.

The number of stems in class 1 has a negative effect on the increment in the same class. This effect is explained by the trees that will leave this diameter class during the growth period. The effect of the number of stems in the directly overlying diameter classes is also negative. This could be explained by the existing competition between the trees of these diameter classes and the trees of class 1. In contrast, the numbers of stems in the larger diameter classes do not appear to have a significant effect on the increment of class 1. The latter two findings could be explained by the hypotheses made above.

Figure 7 shows the regression coefficients associated with the variable $N'_{annual\ cut,d}$ (see equations 7) for the different models.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 July 2022 doi:10.20944/preprints202207.0285.v1

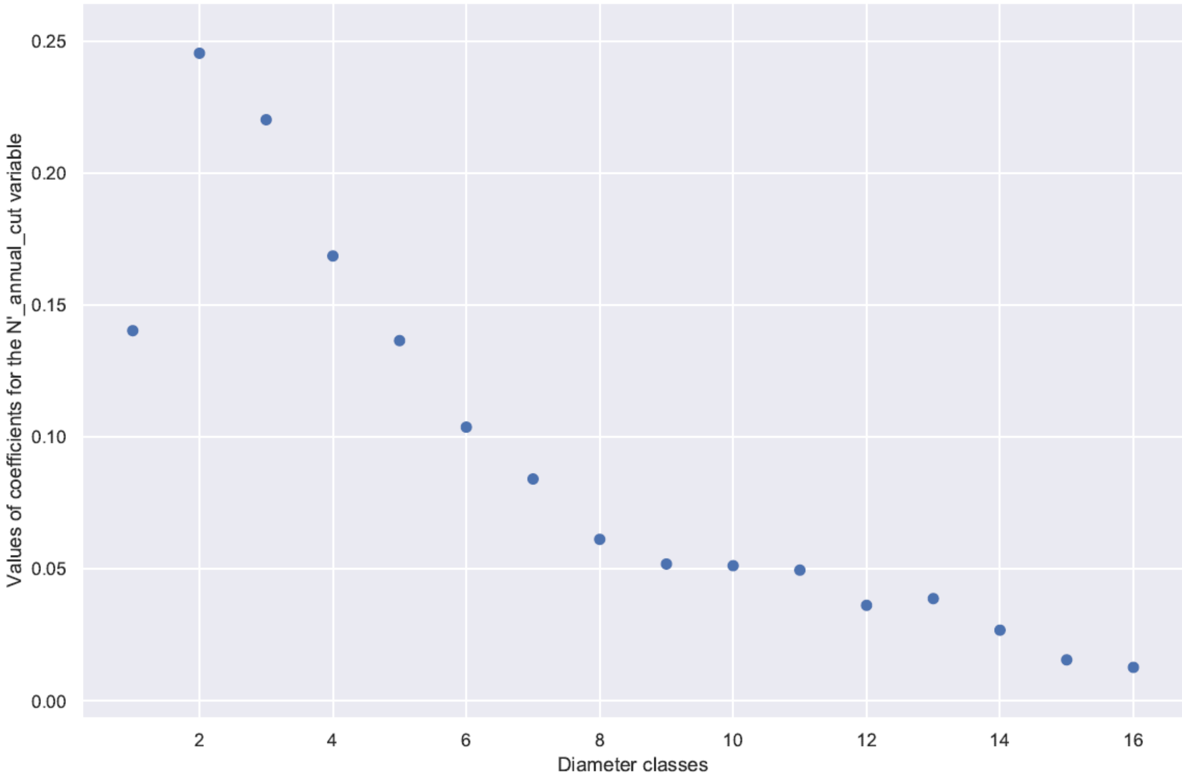

Figure 7. Regression coefficients corresponding to the transformed variable $N'_{annual\ cut,d}$.

Figure 7 shows the gross effect of cuts in a given diameter class on the increment in number of stems in that same class, i.e., the indirect effect of cuts in a class on the growth of trees in that class but without the direct effect of tree removal. The indirect effect of cuts is positive for all diameter classes. Comparing diameter classes with each other is less obvious because increments (*a fortiori* transformed according to Yeo-Johnson) do not mean the same thing from one class to another; an increment in the number of stems will necessarily be lower in the larger diameter classes. However, it may be that intra-class competition is indeed weaker in the larger diameter classes because the larger trees are more spatially separated from each other (they have been fostered by previous cuts), compared to the smaller classes. The pattern of regression coefficients associated with the variable $N'_{annual\ windfall,d}$ is similar to that in Figure 7.

3.2. Results of the modeling with the multilayer perceptron

Even though a unique model is trained for all diameter classes, its performances are disaggregated by diameter class in order to compare the results with those of the linear models. Results are presented in Table 3.

Table 3. Comparison of performances for predicting $N_{after\ gross,d,i}$ between the linear models and the multilayer perceptron.

Diameter class	**Diameter range**	**Test R^2 linear models (reminder)**	**Test R^2 multilayer perceptron model**
1	[17.5;22.5[	82.4%	85.0%

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 July 2022 doi:10.20944/preprints202207.0285.v1

2	[22.5;27.5[	93.5	91.7
3	[27.5;32.5[	93.4	91.0
4	[32.5;37.5[	92.9	84.0
5	[37.5;42.5[	91.3	73.6
6	[42.5;47.5[	89.2	74.0
7	[47.5;52.5[	90.3	74.4
8	[52.5;57.5[	88.7	73.5
9	[57.5;62.5[	90.6	77.9
10	[62.5;67.5[	90.4	71.8
11	[67.5;72.5[	89.6	66.7
12	[72.5;77.5[	88.0	65.3
13	[77.5;82.5[	84.2	34.0
14	[82.5;87.5[	74.5	1.6
15	[87.5;92.5[	69.8	< 0
16	[92.5;97.5[	65.9	< 0
All classes combined		90.3%	96.5%

The analysis of results by diameter class shows that the multilayer perceptron results in lower and more heterogeneous performance than the linear models. The performance is close or even better for some diameter classes (the lower diameter classes for which a large volume of data is available) and it is very inferior in the higher diameter classes. For classes 14, 15, and 16, for which there is very limited data, the multilayer perceptron is apparently unable to predict future stem numbers. In fact, the performance by diameter class depends on the topology, especially for higher classes (except for classes 15 and 16 for which it never works). In the case presented in Table 3, the model with 2 hidden layers of 10 neurons each is relatively complex, which gives good results in the lower classes but produces overfitting and thus bad results in the higher classes. With simpler topologies, results are better in the higher classes but inferior in lower classes, so there is a trade-off. The performance of the multilayer perceptron seen on the whole data set is however better than that of the linear models. As the multilayer perceptron has been trained on the whole data set, its objective is to minimize the errors on the whole data set, which leads to this better global result.

3.3. *Results of linear modeling at the forest division scale*

Linear models perform better for predictions at the diameter class level, which are important from a silvicultural point of view. In particular, linear models also produce good results for larger diameter classes. In addition, due to their simplicity and transparency, linear models are much easier to interpret in terms of forest growth, which facilitates the quality control. Linear models could be used as a decision support tool for the planning of cuts in forests under selective felling management.

The performance of these models is also convincing when tested in real situations. The linear models are applied on whole test divisions (completely unknown to the models) to predict the increment of the number of stems in each diameter class. These increments are then used to calculate the raw stem counts (i.e. without tree removal from cuts or windthrows) at the end of the growth period and can be compared to the actual values. Figures 8, 9, 10 and 11 show the results for four couples (division, inventory year).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 July 2022 doi:10.20944/preprints202207.0285.v1

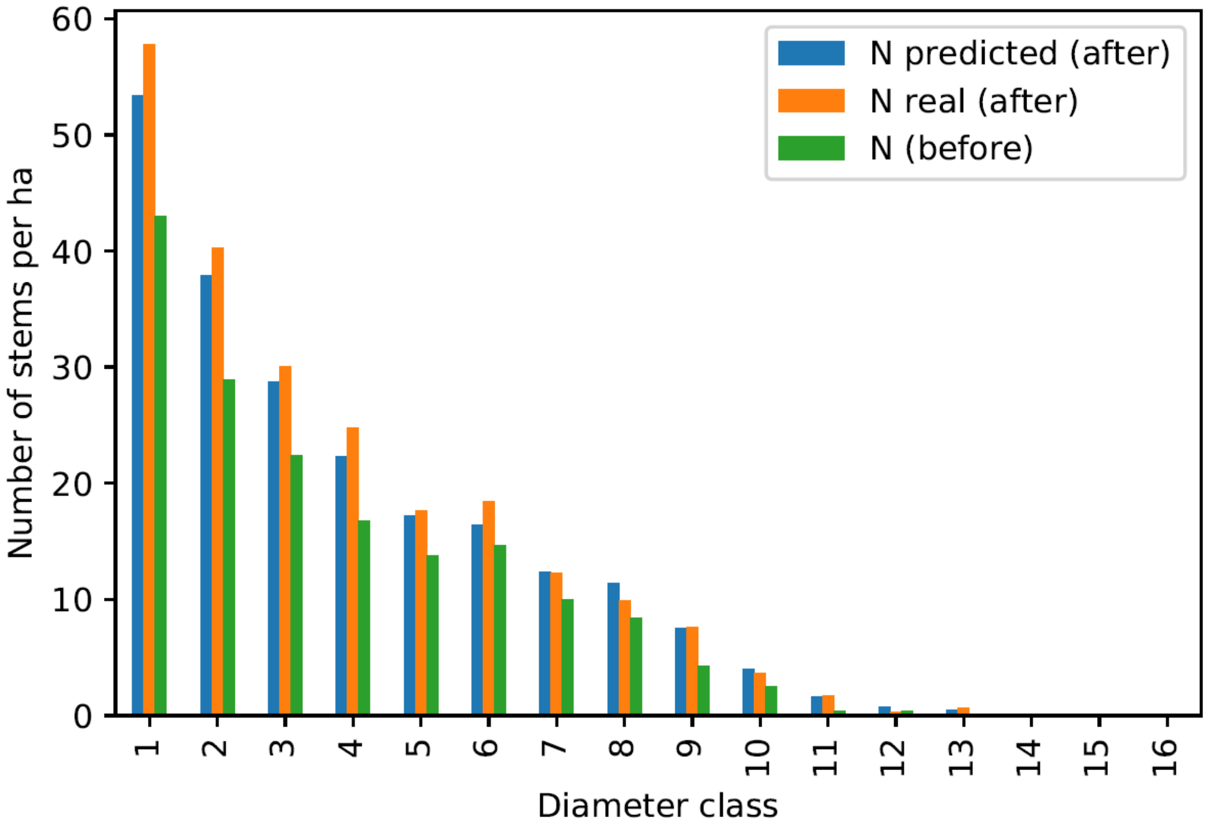

Figure 8. Overall prediction quality control for the division 1974 inventoried in the year 2003.

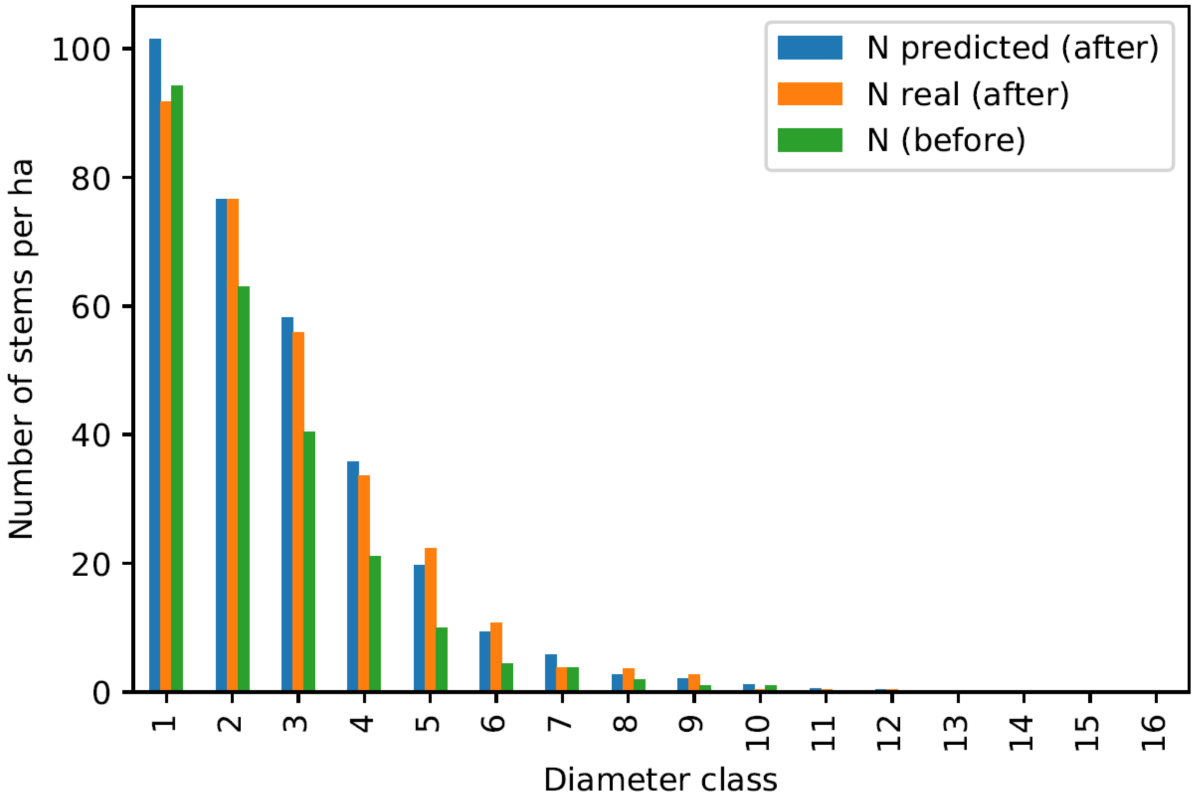

Figure 9. Overall prediction quality control for the division 2240 inventoried in the year 1947.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 July 2022 doi:10.20944/preprints202207.0285.v1

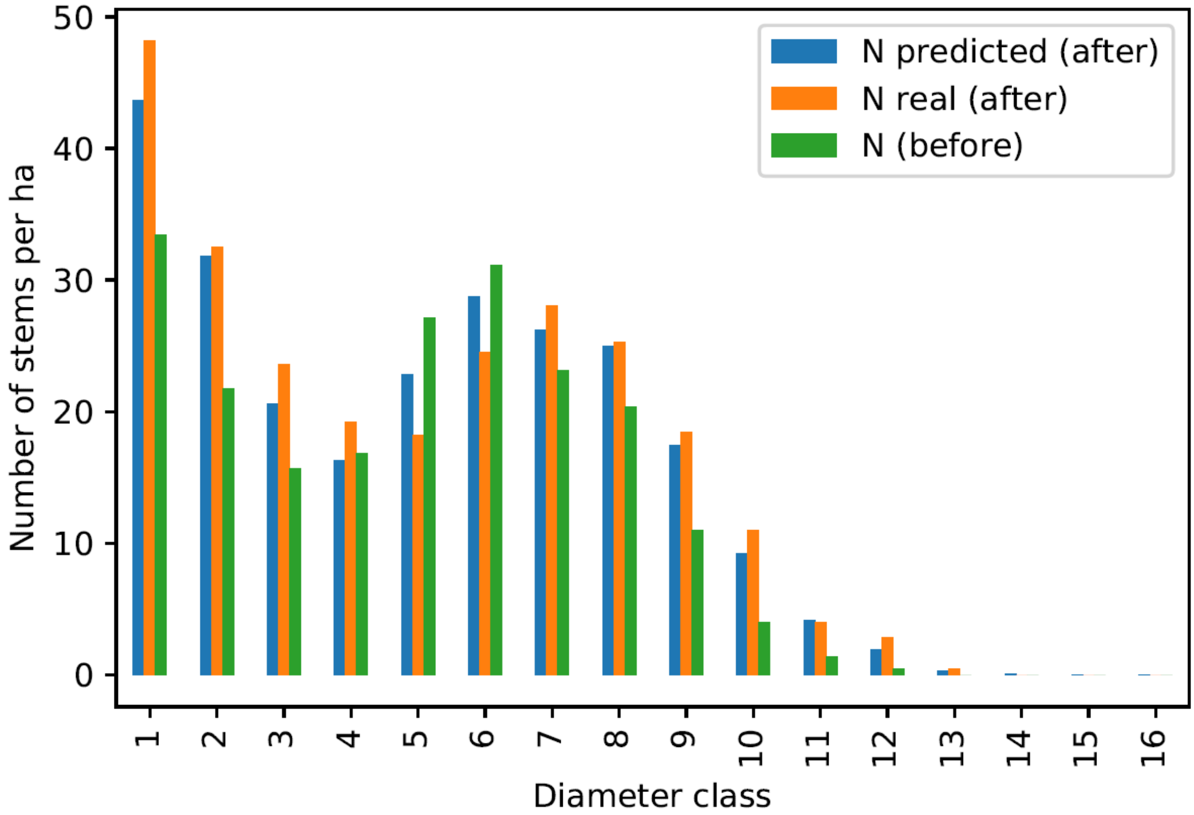

Figure 10. Overall prediction quality control for the division 809 inventoried in the year 1989.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 July 2022 doi:10.20944/preprints202207.0285.v1

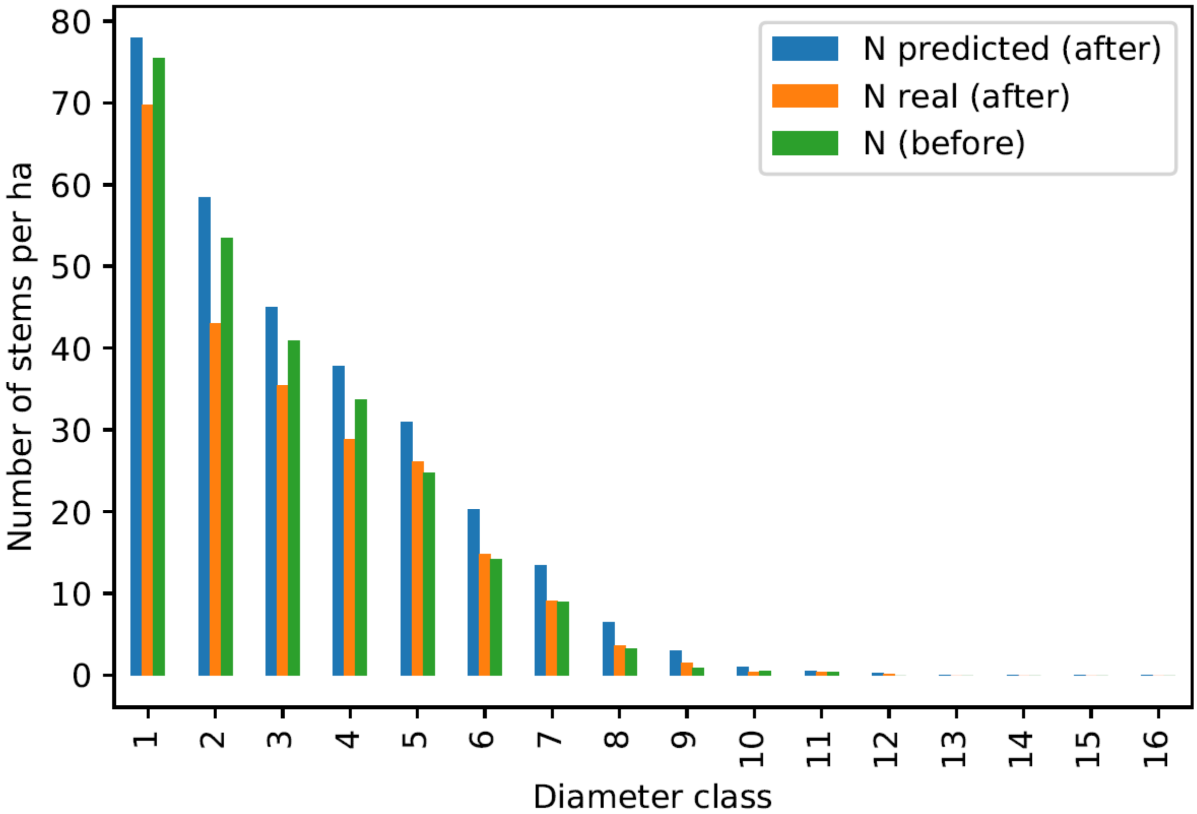

Figure 11. Overall prediction quality control for the division 1878 inventoried in the year 1983.

Figures 8, 9 and 10 show quite contrasting situations for which the models produce good results. The predictions for the 1st diameter class are rather approximate due to the lower performance of this model. However, results are particularly convincing for intermediate classes, which correspond to the best performing models. Fig. 11 shows an example where the models produce poor results, which happens for a small minority of the divisions tested.

Linear models also perform well for predicting increments of some aggregate variables at the division level, e.g., number of stems per ha N, basal area per hectare G, and standing volume V. Prediction tests are this time performed on all couples (division, inventory year) to keep the number of points large. This is not a problem as the overfitting of linear models is extremely limited.

For each couple (division, inventory year), the linear models are used to predict the ΔN and ΔG at the end of the corresponding growth period. The predicted values are then compared to the actual values. For ΔN, an R^2 of 52.3% is obtained, or a mean absolute error of 13.2 stems.ha^{-1}. For ΔG, an R^2 of 42.5% is obtained, or a mean absolute error of 1.37 m^2.ha^{-1}. Figures 12 and 13 show the corresponding residuals plots.

 doi:10.20944/preprints202207.0285.v1

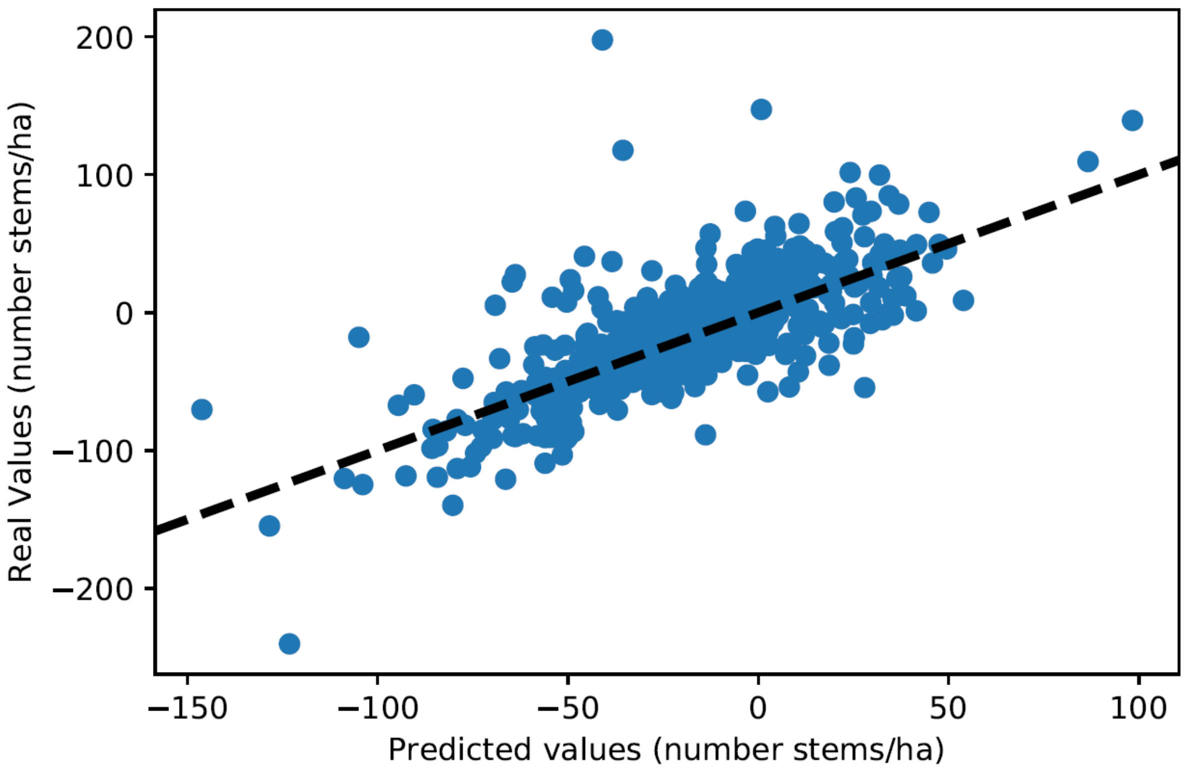

Figure 12. Residuals plot for ΔN predictions at the level of couples (division, inventory year).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 July 2022 doi:10.20944/preprints202207.0285.v1

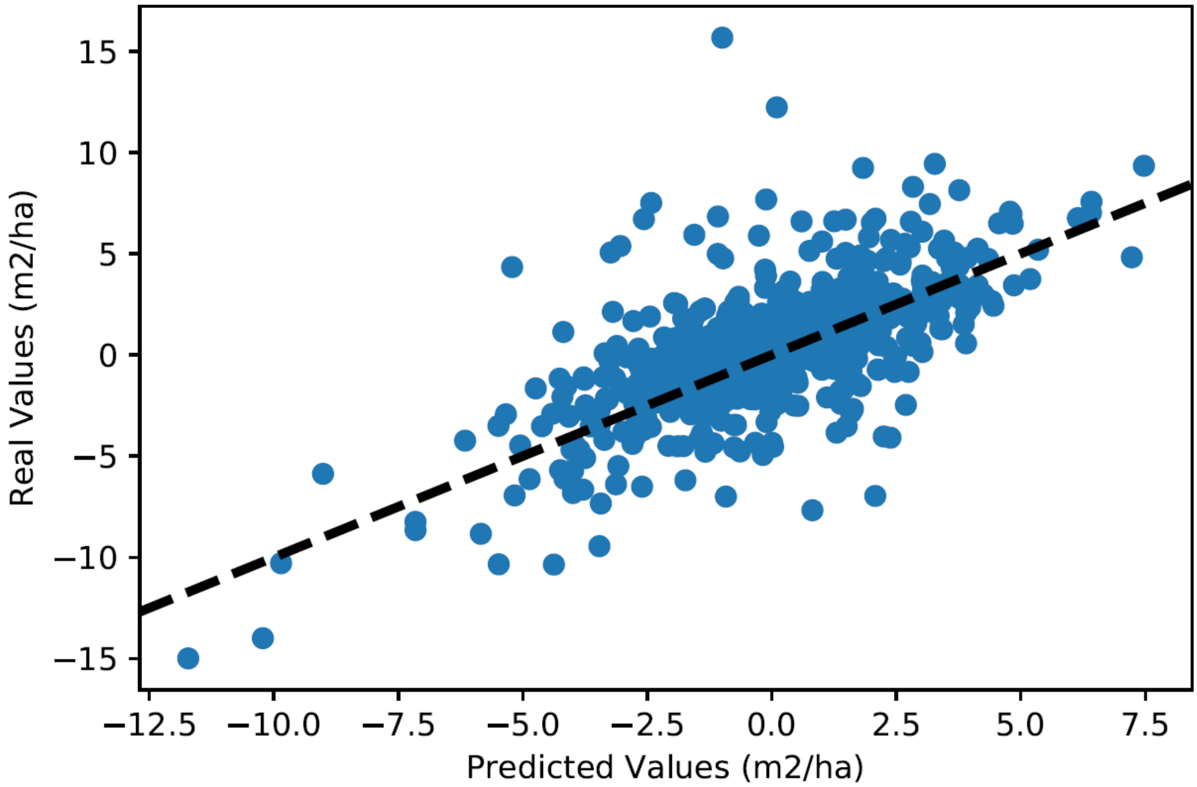

Figure 13. Residuals plot for ΔG predictions at the level of couples (division, inventory year).

In selective felling management practice, variables N, G and V are often disaggregated by large diameter classes, in the Canton of Neuchâtel, those classes are: small woods (17.5cm to 32.5cm), medium woods (32.5cm to 52.5cm) and large woods (> 52.5cm). Disaggregating the variables into these three classes provides very concise information about the structure of a division. The predictive capacity of our models for ΔG in these three respective classes is tested. For small woods, we obtain an R^2 of 66.2%, i.e. a mean absolute error of 0.46 $m^2.ha^{-1}$. For medium woods, we obtain an R^2 of 78.9%, that is a mean absolute error of 0.68 $m^2.ha^{-1}$. For large woods, we obtain an R^2 of 41.9%, that is a mean absolute error of 0.73 $m^2.ha^{-1}$. Figures 14, 15 and 16 show the corresponding residuals plots.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 July 2022 doi:10.20944/preprints202207.0285.v1

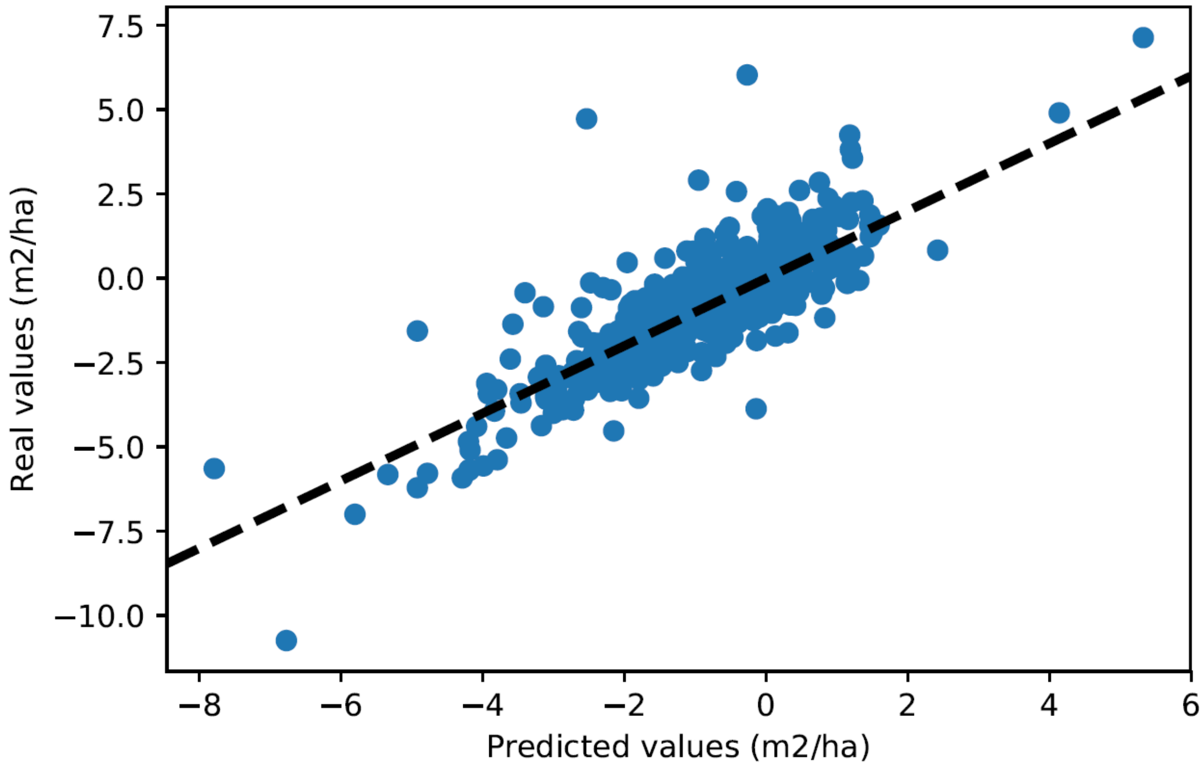

Figure 14. Residuals plot for ΔG predictions at the level of couples (division, inventory year) for the small woods class.

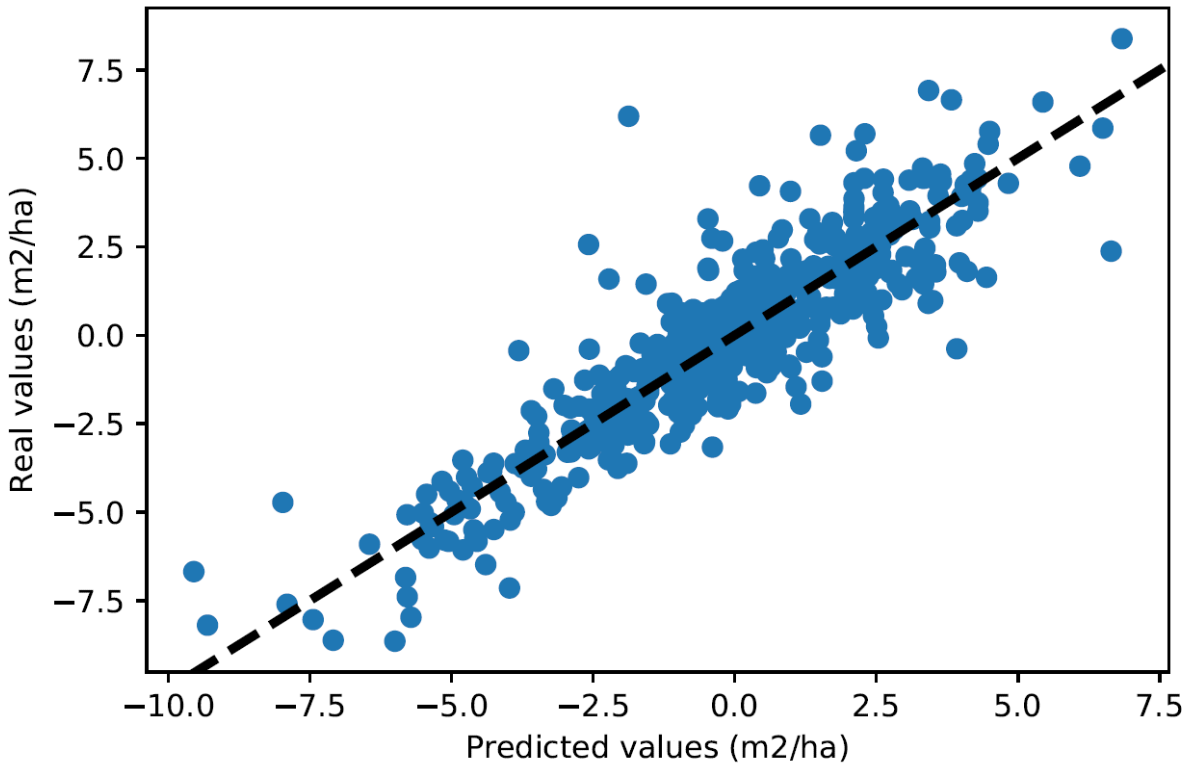

Figure 15. Residuals plot for ΔG predictions at the level of couples (division, inventory year) for the medium woods class.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 July 2022 doi:10.20944/preprints202207.0285.v1

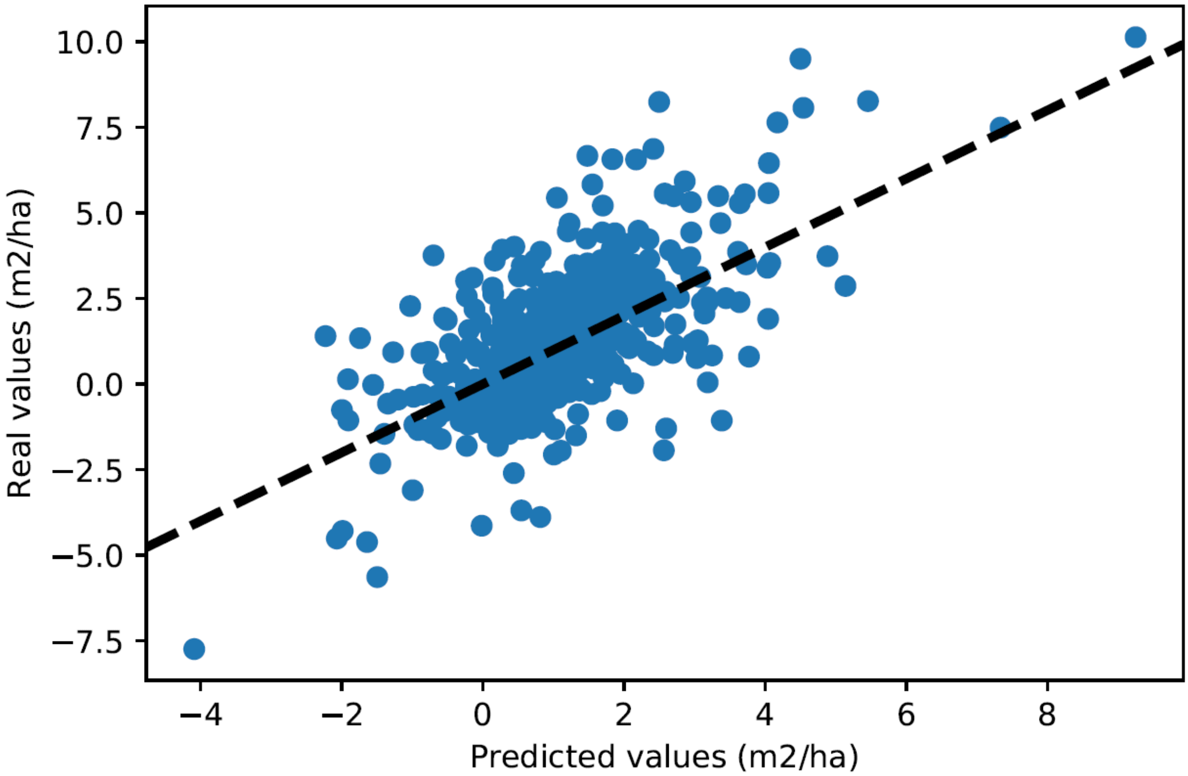

Figure 16. Residuals plot for ΔG predictions at the level of couples (division, inventory year) for the large woods class.

4. Discussion

4.1. *Absolute and comparative performance of the models developed in this study*

The two modeling approaches followed in this study for modeling increments in the numbers of stems by diameter class both produce compelling results, which differ from each other. The linear models achieve better performance at the level of individual diameter classes for most classes and equally good performance for the others. For the linear models, test R^2 coefficients are typically between 50 and 70% (excluding class 1). These good performances are made possible by training a specific model for each diameter class (a form of data panelization).

The implementation of a machine learning workflow, facilitated by the technologies offered by the scikit-learn library and supported by the Python language, allowed to significantly improve the performance of linear models by iteratively improving the selection and transformation of the variables, as well as the selection of a regression algorithm. Train-test-splits also allowed an unbiased evaluation of models' performances in terms of predictive capabilities, which is rarely proposed in inferential statistical studies.

The multilayer perceptron provides the best performance at the global level for predicting the future number of stems with a test R^2 of 96.5%. This is due to the fact that only one model was trained on the whole data set and not independently for each diameter class. In terms of performance by diameter class, there is an interesting trade-off in the topology of the multilayer perceptron. A more flexible model (with more hidden layers and more neurons) allows to reproduce the performance of the linear models for the lower diameter classes for which there are many data (few zero values) but degrades the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 July 2022 doi:10.20944/preprints202207.0285.v1

performance in the higher diameter classes (many 0-values in these data) due to an overfitting. The volume of data was considered insufficient to train a multilayer perceptron per diameter class as there would only be 580 observations per class. From a methodological point of view, the implementation of the multilayer perceptron was much simpler than the fine-tuning of linear models.

4.2. *Comparison of performances with a literature benchmark*

Both types of models developed in this study do better than previously developed linear models for predicting dynamics at the diameter class level. For example, the model [12] for predicting the probabilities of transition of trees between size classes achieves an R^2 (on the whole data set) of 40% or less. Our approach, differentiated by diameter class and based on a larger number of variables (notably the initial diameter distribution) produces these better results. It is not possible to directly compare our results with the results of models predicting individual tree diameter increment, but we can still note that the orders of magnitude of the test R^2 obtained are similar to ours, 57% for the best model in [4] and 53% for the best model in [9].

Aggregating the values predicted by our models for predicting increments of the number of stems per ha, basal area per hectare, or standing volume at the division level, produces relevant results for the practice of forest management. Nevertheless, the prediction performance is inferior to the performance of models specialized on this approach and based on machine learning methods. One model in [8] based on an artificial neural network achieves a test R^2 of 76% and another model in [7], also based on an artificial neural network, achieves a test R^2 of 94% (versus 87% with linear regressions). This better performance is enabled by specialized models that directly predict the aggregate increment but probably also by the fact that these models are based on data from permanent plots. It may be interesting in the future to use the data from this study for direct prediction of increment at the stand level to explore its potential.

4.3. *Assessment of the use of data from the Canton of Neuchâtel*

Our results show that the inventory data from the Canton of Neuchâtel, which are non-experimental data from practice, generate valuable information to train convincing growth models at the diameter class level. The use of this type of data is an additional challenge compared to the use of data from permanent plots. Because increments are not tracked on the individual tree level, the calculation of increments is more approximate, for example because of the lack of data on mortality. The lack of information on stems below 17.5cm and of direct data on recruitment in the first diameter classes makes it difficult to model their dynamics. From this point of view, the performance of our models is limited. Another particularity of the data collected by the Canton of Neuchâtel compared to permanent plots is related to the inventory areas. The divisions used for this purpose are each about ten hectares in size instead of a few tens of ares at most for the plots. Data at such a large scale are necessarily averaged, resulting in a probable reduction of variability in the data and reduced information for model training compared to the permanent plots.

The fact that the volume of data provided by the Canton of Neuchâtel was so large was clearly an added value for our study. The volume of data was clearly sufficient to train our linear models with a remarkable absence of overfitting. This large volume of data was also crucial for the training of the multilayer perceptron which reached convincing performances. Nevertheless, the volume of data remained limiting for the latter, especially in the upper diameter classes.

5. Conclusions

The linear models developed in this study achieve sufficient performance for practical use. The predictions from these models are division-specific in that the predicted dynamics depend on the initial diameter distribution and take into account competition

 doi:10.20944/preprints202207.0285.v1

between diameter classes to some extent. Linear models also have the advantage of being relatively transparent and therefore convincing for practitioners (coefficients can be directly interpreted by a professional knowledgeable in forest growth). Linear models, at least in this study, also predict fewer outliers than the multilayer perceptron.

Linear models can be used to predict the growth dynamics of a division over a period of 7 to 12 years. They can possibly be used recursively to predict dynamics over longer periods, but it is not recommended to simulate more than two or three growth periods, especially because the prediction of recruitment is limiting. The models can typically be used for better planning and implementation of the next cut in a given division, at least in terms of analyzing the role of cuts on the diameter distribution in that division. The model is not well suited to understanding the role of cuts on recruitment.

Before the model can be widely used in practice, it must first be tested under real conditions, for example on a current case study in the Canton of Neuchâtel. To do so, the model should be further developed by adding a 17th diameter class to take into account the effect of growth on class 16 (the maximum class in our model) or by adding outlier control mechanisms, for example by prohibiting negative values for the number of stems or prohibiting a positive increment when the two or three underlying diameter classes are empty. The model will also need to be usable from a prototype GUI suitable for use in practice.

Funding: This study was financed by the Swiss Federal Office for the Environment (contract 00.0059.PZ / 5933E7220). The author alone is responsible for the content of this article, the Federal Office for the Environment cannot be held responsible.

Data Availability Statement: The data used in this study were provided as a courtesy by the Canton of Neuchâtel (Switzerland). Some of the data are publicly available on the canton's geoshop (sitn.ne.ch/geoshop), free of charge when used for scientific purposes. Some data were not on the geoshop but were provided on motivated request and as a courtesy by the *Service de la faune, des forêts et de la nature* of the Canton of Neuchâtel.

Acknowledgments: I thank Christian Rosset from the Bern University of Applied Sciences BFH for his feedbacks on the project design and acquisition.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

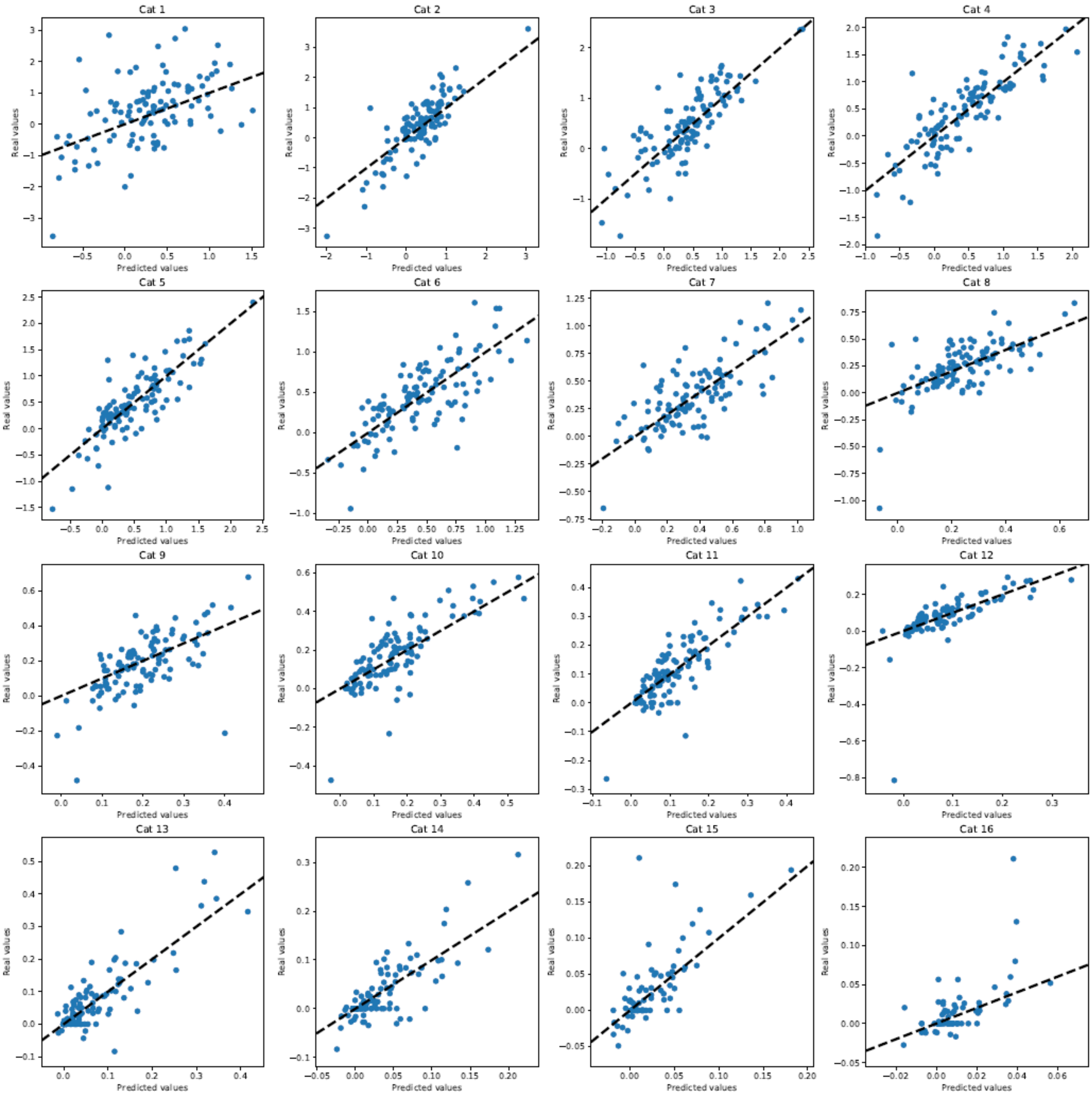

Figure A1. Residuals plots for linear models differentiated by diameter class presented in equations 7 (classes from 1 to 16 where class center in cm = (#class - 1) * 5 + 20).

Table A1. Regression coefficients for linear models trained on the whole data set (class center in cm = (#class - 1) * 5 + 20).

Model	Model class 1	Model class 2	Model class 3	Model class 4	Model class 5	Model class 6	Model class 7	Model class 8	Model class 9	Model class 10	Model class 11	Model class 12	Model class 13	Model class 14	Model class 15	Model class 16
Diameter range	[17.5;22. 5[	[22.5;27. 5[	[27.5;32. 5[	[32.5;37. 5[	[37.5;42. 5[	[42.5;47. 5[	[47.5;52. 5[	[52.5 ;57.5[	[57.5;62. 5[	[62.5;67. 5[	[67.5;72. 5[	[72.5;77. 5[	[77.5;82. 5[	[82.5;87. 5[	[87.5;92. 5[	[92.5;97. 5[
Intercept	0.31567 783	0.43808 498	0.47340 118	0.46051 455	0.52286 354	0.42056 252	0.35215 925	0.25705 117	0.20267 967	0.151799 39	0.104450 36	0.071826 93	0.059078 5	0.028896 7	0.015386 93	0.008068 61

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 July 2022 doi:10.20944/preprints202207.0285.v1

N_class_1	-0.08857224	0.95262586	0.30478168	0	0	0	0	0	0	0	0	0	0	0	0	0
N_class_2	-0.19883118	-0.70449874	0.44469622	0.26800507	0	0	0	0	0	0	0	0	0	0	0	0
N_class_3	-0.2479029	-0.31537605	-0.47945268	0.33349209	0.19441854	0	0	0	0	0	0	0	0	0	0	0
N_class_4	-0.09295262	-0.12675351	-0.22661129	-0.36086538	0.38562311	0.21856427	0	0	0	0	0	0	0	0	0	0
N_class_5	-0.03021211	0	-0.10632899	-0.21420664	-0.44715699	0.19482813	0.14910786	0	0	0	0	0	0	0	0	0
N_class_6	-0.02551455	0	0	-0.11447264	-0.12177319	-0.30391304	0.18257123	0.10812181	0	0	0	0	0	0	0	0
N_class_7	-0.09059655	0	0	0	-0.05773986	-0.1078003	-0.24606661	0.11924328	0.06907012	0	0	0	0	0	0	0
N_class_8	-0.04804268	0	0	0	0	0.00040884	-0.06378476	-0.14591235	0.06992039	0.07125811	0	0	0	0	0	0
N_class_9	0.03877847	0	0	0	0	0	0	-0.05121676	-0.09001726	0.06730688	0.04144019	0	0	0	0	0
N_class_10	-0.03040491	0	0	0	0	0	0	0	-0.03252369	-0.10502199	0.05030268	0.04166453	0	0	0	0
N_class_11	0.00132175	0	0	0	0	0	0	0	0	0	-0.06807874	0.04787912	0.02745242	0	0	0
N_class_12	-0.00334377	0	0	0	0	0	0	0	0	0	0	-0.06532824	0.03742049	0.01877528	0	0
N_class_13	0.01200967	0	0	0	0	0	0	0	0	0	0	-0.00304633	-0.03685599	0.02548971	0.01338672	0
N_class_14	0.01124542	0	0	0	0	0	0	0	0	0	0	0	0	-0.03028762	0.01327143	0.00454961

N_class_15	-0.06703446	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.01833805	0.0064068
N_class_16	0.02649344	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.00812474
N_annual_cut	0.1378631	0.2531186	0.21802826	0.1655776	0.13114236	0.10657015	0.08537938	0.05979587	0.05113959	0.05160261	0.04996778	0.03751677	0.03917284	0.02788062	0.01534358	0.01277216
N_annuel_windfall	0.09950984	0.04107477	0.03255623	0.02292689	0.00990127	0.03090872	0.02697962	0.0147016	0.00961194	0.0032046	0.01048788	0.00560738	0.00773745	0.00713012	0.00603813	0.00204978

References

1. Bettinger, P., Boston, K., Siry, J. P., & Grebner, D. L. (2017). Estimation and projection of stand and forest conditions. *Forest management and planning, 2nd edn. Academic Press, London*, 87-111. doi.org/10.1016/B978-0-12-809476-1.00004-7
2. Susse, R., Allegrini, C., Bruciamacchie, M., Burrus, R. (2011). *Management of irregular forests: developing the full potential of the forest*. Azur Multimedia. Saint Maime, France (translated from French by Morgan, P.)
3. Peng, C. (2000). Growth and yield models for uneven-aged stands: past, present and future. *Forest ecology and management*, *132*(2-3), 259-279. doi.org/10.1016/S0378-1127(99)00229-7
4. Ou, Q., Lei, X., & Shen, C. (2019). Individual tree diameter growth models of larch–spruce–fir mixed forests based on machine learning algorithms. *Forests*, *10*(2), 187. doi.org/10.3390/f10020187
5. Schütz, J. P. (2006). Modelling the demographic sustainability of pure beech plenter forests in Eastern Germany. *Annals of Forest Science*, *63*(1), 93-100. doi.org/10.1051/forest:2005101
6. Eskelson, B.N.I., Barrett, T.M. & Temesgen, H. 2009. Imputing mean annual change to estimate current forest attributes. Silva Fennica 43(4): 649–658. doi.org/10.14214/sf.185
7. Hamidi, S. K., Zenner, E. K., Bayat, M., & Fallah, A. (2021a). Analysis of plot-level volume increment models developed from machine learning methods applied to an uneven-aged mixed forest. *Annals of Forest Science*, *78*(1), 1-16. doi.org/10.1007/s13595-020-01011-6
8. Bayat, M., Bettinger, P., Hassani, M., & Heidari, S. (2021). Ten-year estimation of Oriental beech (Fagus orientalis Lipsky) volume increment in natural forests: a comparison of an artificial neural networks model, multiple linear regression and actual increment. *Forestry: An International Journal of Forest Research*, *94*(4). 598-609. doi.org/10.1093/forestry/cpab001
9. Hamidi, S. K., Weiskittel, A., Bayat, M., & Fallah, A. (2021b). Development of individual tree growth and yield model across multiple contrasting species using nonparametric and parametric methods in the Hyrcanian forests of northern Iran. *European Journal of Forest Research*, *140*(2), 421-434. doi.org/10.1007/s10342-020-01340-1
10. Schütz, J. P. (1975). Dynamique et conditions d'équilibre de peuplements jardinés sur les stations de la hêtraie a sapin. *Journal forestier suisse*, *126*, 637-670.
11. Lu, H. C., & Buongiorno, J. (1993). Long-and short-term effects of alternative cutting regimes on economic returns and ecological diversity in mixed-species forests. *Forest Ecology and Management*, *58*(3-4), 173-192. doi.org/10.1016/0378-1127(93)90144-C

12. Buongiorno, J., Peyron, J. L., Houllier, F., & Bruciamacchie, M. (1995). Growth and management of mixed-species, uneven-aged forests in the French Jura: implications for economic returns and tree diversity. *Forest science*, *41*(3), 397-429. doi.org/10.1093/forestscience/41.3.397
13. Buongiorno, J. (2001). Quantifying the implications of transformation from even to uneven-aged forest stands. *Forest ecology and management*, *151*(1-3), 121-132. doi.org/10.1016/S0378-1127(00)00702-7
14. Rollin, F., Buongiorno, J., Zhou, M., & Peyron, J. L. (2005). Management of mixed-species, uneven-aged forests in the French Jura: from stochastic growth and price models to decision tables. *Forest Science*, *51*(1), 64-75. doi.org/10.1093/forestscience/51.1.64
15. Pardé, J. (1991). La méthode du Contrôle, d'hier à aujourd'hui. *Revue forestière française*, *43*(3), 185-202. Doi.org/10.4267/2042/26198
16. Favre, L. A., & Oberson, J. M. (2002). 111 années d'application de la méthode du contrôle à la forêt de Couvet| 111 years of application of the method of control in the community forest of Couvet. *Schweizerische Zeitschrift fur Forstwesen*, *153*(8), 298-313. doi.org/10.3188/szf.2002.0298
17. Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific data*, *5*(1), 1-12. doi.org/10.1038/sdata.2018.214
18. Allgaier Leuch, B., Streit, K., Stillhard, J., Cioldi, F., & Brang, P. (2018). Zukunft der Fichte im Schweizer Mittelland. *Wald und Holz*, *98*(3), 31-34.
19. Yeo, I. K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, *87*(4), 954-959. doi.org/10.1093/biomet/87.4.954
20. Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211-243. doi.org/10.1111/j.2517-6161.1964.tb00553.x
21. Eilers, P., & Marx, B. (1996). Flexible Smoothing with B-splines and Penalties. *Statistical science*, *11*(2), 89-121. doi.org/10.1214/ss/1038425655
22. scikit-learn (2021b). Spline Transformer – scikit-learn 1.0.2 documentation. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.SplineTransformer.html#sklearn.preprocessing.Splin Transformer
23. Huber, Peter J. (1964). "Robust Estimation of a Location Parameter". Annals of Statistics. 53 (1): 73–101. doi.org/10.1214/aoms/1177703732
24. scikit-learn (2021c). Huber regressor – scikit-learn 1.0.2 documentation. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.HuberRegressor.html
25. scikit-learn (2021d). SGD regressor – scikit-learn 1.0.2 documentation. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDRegressor.html
26. Haykin, S. (2004). Neural networks: A comprehensive foundation (Second edition). Prentice Hall, Upper Saddle River.
27. scikit-learn (2021a). Neural network models (supervised) — scikit-learn 1.0.2 documentation. https://scikit-learn.org/stable/modules/neural_networks_supervised.html#regression
28. Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, *2*(5), 359-366. doi.org/10.1016/0893-6080(89)90020-8