Prediction of Need for Early Tending in Norway Spruce Plantations

NUUTTI KILJUNEN
Suonenjoki Research Station, Finnish Forest Research Institute, Juntintie 154
FIN-77600 Suonenjoki, FINLAND

Kiljunen, N. 2004. Prediction of Need for Eearly Tending in Norway Spruce Plantations. *Baltic Forestry*, 10 (2): 56–60.

To determine the need for early tending of Norway spruce (*Picea abies* L.) stands, the effects of soil texture, site vegetation type, method of soil preparation, and stoniness and wetness of the stand were analysed. The data were collected from 3-year-old Norway spruce plantations in southern Finland. Many site attributes markedly affected the need for early tending. Nevertheless, classification of individual stands according to site characteristics was poor, even with state-of-the-art machine learning classifiers. To ensure correct identification of need for early tending in Norway spruce plantations, a survey of recently established stands is recommended.

Key words: Norway spruce, tending, machine learning

Introduction

In Finland, the regimen for tending of coniferous stands and the role of work science and technology in tending have been discussed intensively (Immonen 2003, Kiljunen *et al.* 2003, Schildt, 2003, Harstela 2003). Current silvicultural instructions (Hyvän metsänhoidon… 2001) provide rough guide-lines for tending, but the effect of variation in site characteristics is poorly understood. Moreover, different tending regimens on different sites have been studied very little, and therefore the effect of these regimens is poorly known. Issues affected by the timing of tending are, for instance, work productivity, sprouting of cut broad-leaved trees, growth of sprouts, and need for clearing before first commercial thinning. The intensity of tending also affects the above-mentioned issues.

The general effects of tending and the competition caused by secondary stems in the stand are, however, well known. The height and diameter increment of Norway spruce seedlings usually decrease due to densely growing secondary broad-leaved trees that have faster early development than spruce (Walfridsson 1976, Pukkala 1981). In most spruce plantations, tending only once before the first commercial thinning may be enough; but, especially on most fertile sites and on areas suffering from excessive soil moisture, tending twice is needed. Early tending may be done either as total cleaning, removing all secondary stems, or as point cleaning, removing only secondary stems within a certain radius around the primary stems. In this paper, the term “early tending” is used for clearing spruce stands from secondary broadleaved trees 3 to 6 years after planting. The method of early tending, however, is not discussed here.

Non-industrial private forest (NIPF) owners, who dominate forest ownership in Finland, live away from their forest property more often today than previously (Karppinen *et al*. 2002.). In the future, this trend is expected to continue. The interest of NIPF landowners in their forest properties may also decrease. Thus, in many cases follow-up of the development of recently established stands may not be as easy as before. To provide decision-support for the forest owner, prediction of the need for tending in young stands should help them concentrate the follow-up inspections to the stands most likely to need tending. A survey of recently established stands is a good way to determine the need for tending but naturally, this causes additional costs. However, there are other reasons for such a survey, like quality assessment of stand establishment for determining the density of the stand. In such a quality survey, the need for early tending could easily be observed at the same time.

The main aim of this study was to test, which site factors affect the need for early tending in Norway spruce plantations. A further aim was to determine whether there are means to predict the need for early tending for an individual stand. Such information could be helpful, for example, for a stand-establishment service provider as decision-support for after-sales marketing. The basis of classification was information that can be obtained already at the time of stand establishment.

Material and methods

Survey data

The material used for the study was a data set from a survey of 3-year-old Norway spruce plantations (Saksa *et al.,* 2002). The data are measurements taken from actual plantations in NIPF forests in southern Finland. The minimum area of a stand was 0.5 hectares. The measurements were carried out on circular sample plots located systematically in the stands. The key information from the survey is presented in Table 1. In every stand, 15…20 sample plots were surveyed. The area of each sample plot was 20 m^2; i.e. the radius of the plots was 2.52 m. For every plot, the following variables were recorded: soil texture class, vegetation site type, number and height of primary and secondary stems. Vegetation site type classification is based on the Cajander's (1909) theory. The need for different management operations was also assessed, and possible weakening of the regeneration result due to stony or wet conditions was recorded. More detailed information on the data collection has been presented by Saksa *et al*. (2002).

Table 1. Key information in the data.

Proportion of stands determined to require early tending	0.59
Number of stands surveyed	1473
Total survey area, ha	2618.1
Average size of stand, ha	1.79

Before the analyses, the data were processed to remove unusable cases, which were those collected on sites prepared by an unknown method. The sample-plot data were converted to create the variables for the stand-level analysis (Table 2). Peatlands were classified according to the forest site types with the most similar yield potential. For testing differences between different attributes in relation to the determined need for early tending, cross-tabulation was used. To test significances between attribute classes, a chi-square test was used.

Table 2. Variables of the survey data used in sample plot and stand levels. Boolean variables are indicated by (B).

Need for early tending (B)
Dominating site type class
Dominating soil texture class
Stony area >20% (B)
Wet area >30% (B)
Main method of soil preparation

Machine learning and classification

To classify the data according to the observed need for early tending, naļve Bayes (NB) and J4.8 machine-learning algorithms were used. The software used was WEKA 3-3-4 (Waikato Environment for Knowledge Analysis), developed at the University of Waikato, New Zealand (Kirkby 2003). WEKA is an open-source software issued under GNU licence at www.cs.waikato.ac.nz/~ml/weka.

For every classification run, the data set was split in half. One half, the training data, was used for constructing the model for classification and the other half for testing the constructed models. The testing data set was picked out randomly from the whole data set, but was the same for every machine-learning run.

A naļve Bayesian network (Figure 1) is a simple structure in which all random variables representing observable data have a single common parent - the class variable. In this case the class variable was the observed need for early tending. The naļve Bayes model assumes that given a class $C=j$, the attributes X_k are independent (Hastie *et al.* 2001):

$$f_j(X) = \prod_{k=1}^{p} f_{jk}(X_k).$$

In general, this assumption is not true. In the data used in this study, for example, the soil texture and method of soil preparation are strongly dependent on each other. Although the individual estimates of class density may be biased, the bias still might not affect the posterior probabilities as much (Hastie *et al.* 2001).

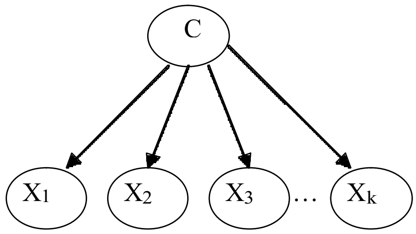

Figure 1. The principle of naïve Bayes.

In the machine learning of the NB structure, the classifier learns the conditional probability of each attribute X_k given the label C in the training data. Classification is then done by applying Bayes rule to compute the probability of C given the particular instantiation of $X_1, \ldots, X_k$ (Friedman & Goldszmidt 1996). The learner estimates the required probabilities by calculating the corresponding frequencies observed in the training set.

The J4.8 classifier, also used in this study (Witten & Frank 2000) is actually an improved Java coded version of the widely used C4.5 classifier (Quinlan

1993). The principles of the algorithms are the same (Witten & Frank 2000). C4.5 and J4.8 produce pruned decision trees having the attributes *Xi* in the stem and branches, and predicted classes as values in the leaves.

The algorithm examines each attribute by evaluating the likelihood that it will improve the overall decision performance of the tree. J4.8 is based on a node-splitting entropy function that works by reducing the degree of randomness in the current node. While constructing the tree of J4.8, the following entropy function is minimized when constructing the tree:

$$E(n) = \sum_{j=1}^{C} - p_j(C = C_j \mid n) \log_2 p_j(C = C_j \mid n)$$

where

n=set of cases

$E(n)$ = entropy in the set n

C=class variable

P_j=probability that a case belongs to class c_j

Decision trees are pruned by replacing a whole subtree by a leaf node (Quinlan 1993). The subtree is replaced if a decision rule establishes that the expected rate of error in the subtree is greater than that in a single leaf. Thus, those parts of the tree that do not contribute to the accuracy are removed.

Results

Cross-tabulation

The site classes had some effect on the need for early tending (Table 3). The significance of the differences was p=0.06. The need for tending decreased slightly as the fertility of the site decreased. The vegetation site classes are from the poorest to the most fertile: VT = *Vaccinium* type, MT = *Myrtillus* type, OMT = *Oxalis-Myrtillus* type and OMaT = *Oxalis-Majanthemum* type (Cajander 1909). The number of instances in the highest fertility class (OMaT) was small, making inference more difficult.

Table 3. Need for early tending in different site classes.

	OMaT	OMT	MT	VT
Not needed	4	161	434	10
Needed	4	289	570	14
Sum	8	450	1004	24
Proportion early tending needed	0.50	0.64	0.57	0.58

Soil texture classes differed significantly (p<0.01) in relation to the need for early tending (Table 4). In mineral soils the finer the texture was, the higher was the proportion of plantations needing early tending. Peatlands also had high proportion of stands needing early tending.

Table 4. Need for early tending in different soil textures.

	Coarse	Medium	Fine-grained	Peat
Not needed	17	447	126	19
Needed	23	583	233	38
Sum	40	1030	359	57
Proportion of early tending needed	0.58	0.57	0.65	0.67

The need for early tending also differed significantly (p<0.01) among stands with different methods of soil preparation (Table 5). Mounding resulted in the largest proportion of stands with a need for early tending. The mounded sites in this data are mostly ditch-mounded; however, no distinction from other mounding methods is made. Patch-scarification and disc-trenching differed little from each other and from stands with no mechanical site preparation.

Table 5. Need for early tending with different methods of soil preparation.

	No preparation	Patching	Disc-trenching	Mounding
Not needed	41	191	334	39
Needed	62	270	401	135
Sum	103	461	735	174
Proportion of early tending needed	0.60	0.59	0.55	0.78

Machine learning

The prior probabilities of NB model learned from the training data are shown in Tables 6 and 7. Not all the joint probabilities of such a number of variables cannot be presented concisely and thus not presented here. Their clear presentation actually requires the use an interactive system like a computer programme.

Table 6. Prior probabilities in the training data split for cases where early tending was not needed (41.1% of the cases).

Soil preparation	no_preparation:6.1 patch: 32.3 disc-trench: 51.6 mound:10.0
Site type	OMaT: 0.6 OMT: 22.6 MT: 73.9 VT: 2.9
Soil texture	coarse: 2.6 medium: 74.8 fine: 19.0 peat: 3.5
Stones>20%	False: 98.7 True: 1.3
Wet>30%	False: 99.7 True: 0.3

Table 7. Prior probabilities in the training data split for cases where early tending was needed (58.9% of the cases).

Soil preparation	no_preparation: 6.9 patch: 32.5 disc-trench: 42.6 mound: 18.0
Site type	OMaT: 0.7 OMT: 32.5 MT: 64.1 VT: 2.8
Soil texture	coarse: 2.8 medium: 66.6 fine: 26.5 peat: 4.1
Stones>20%	False: 98.1 True:1.9
Wet>30%	False: 98.1 True:1.9

```
WET>30% = FALSE
|   SITE_TYPE = OMaT: 1 (3.0/1.0)
|   SITE_TYPE = OMT : 1 (207.0/69.0)
|   SITE_TYPE = MT
|   |   PREPARATION = NO_PREP: 1 (30.0/13.0)
|   |   PREPARATION = PATCH
|   |   |   TEXTURE = COARSE: 0 (4.0/1.0)
|   |   |   TEXTURE = MEDIUM: 0 (116.0/56.0)
|   |   |   TEXTURE = FINE  : 1 (36.0/13.0)
|   |   |   TEXTURE = PEAT  : 0 (2.0/1.0)
|   |   PREPARATION = DISC_TR
|   |   |   STONES>20% = FALSE
|   |   |   |   TEXTURE = COARSE: 1 (6.0/2.0)
|   |   |   |   TEXTURE = MEDIUM: 0 (194.0/96.0)
|   |   |   |   TEXTURE = FINE  : 1 (31.0/13.0)
|   |   |   |   TEXTURE = PEAT  : 1 (1.0)
|   |   |   STONES>20% = TRUE: 1 (5.0/1.0)
|   |   PREPARATION = MOUND: 1 (75.0/24.0)
|   SITE_TYPE = VT
|   |   PREPARATION = NO_PREP: 1 (1.0)
|   |   PREPARATION = PATCH  : 0 (3.0/1.0)
|   |   PREPARATION = DISC_TR: 1 (11.0/4.0)
|   |   PREPARATION = MOUND  : 0 (4.0/2.0)
WET>30% = TRUE: 1 (7.0)

Number of Leaves  :     18
Size of the tree :      25
```

Figure 2. The unpruned J4.8 tree.

The unpruned J4.8 tree is shown in Figure 2. When the pruning option was applied, the algorithm classified all test instances as needing tending (Table 8). The stem of the decision tree was Wet>30%, indicating the influence of moisture on the need for tending. The first figure in each leaf is the number of cases belonging to that leaf in the training data set, and the second figure is the number of misclassified cases in the training data. Zeros as the number of misclassifications are not shown.

Table 8. Confusion matrixes of the classification methods used in the test data split.

Actual class	Classification by different methods					
	Naïve Bayes		J4.8 pruned		J4.8 unpruned	
	Not needed	needed	Not needed	needed	Not needed	needed
Not needed	94	205	0	299	147	152
Needed	116	322	0	438	173	265

The highest classification accuracy was achieved by J4.8 with the pruning option, despite its lacking ability to recognize any stands not needing early tending. The Kappa statistic indicating the quality of classification result was highest with unpruned J4.8, although its accuracy was weaker than with pruning option (Table 9). J4.8 also had the highest precision for identifying the stands needing early tending. The poorer accuracy of unpruned J4.8 and NB compared to pruned J4.8 is a result of classifying many stands wrongly, i.e. as not needing early tending (Table 9).

Table 9. Accuracy of the classifiers in the test data split.

	Naïve Bayes	J4.8 pruned	J4.8 unpruned
Correctly classified instances, %	56.4%	59.4%	55.9%
Kappa statistic	0.05	0.05	0.10
Precision for no need for tending	44.8%	0%	45.9%
Precision for need for tending	61.1%	59.4%	63.5%

Discussion and conclusions

Pre-selection of potentially relevant attributes is critical, as omitting a key feature or including irrelevant ones can both lead to poor classification accuracy (Garner *et al*. 1995). Although there were evident differences in the effect of soil properties on the need for early tending, classification of individual stands according to their need for early tending was not successful. The accuracy of both methods of classification used in this study was poor. Some states of the attributes resulted in different learned classifiers than what were expected. For example, the plantations established on mounded soil were often related to need for early tending, although this differs from current practical experience. This is caused most probably by biased data. The mounded areas were mostly sites where drainage was seen as necessary and the soil was prepared by ditch-mounding. Another reason having some influence on the poor classification result might have been variation inside the stands.

In comparative studies of classification methods, the results achieved by C4.5 and NB have been similar (Friedman & Goldszmidt (1996). For the amount of data used in this study, the computation performance of both algorithms used was good. Other machine learning methods often used for classification, for example, tree-augmented networks (TAN) have usually performed the same way as C4.5 and NB. Therefore, it's unlikely to expect any advantage of using them in this problem.

In conclusion, although some attributes influence the need for early tending in a young Norway spruce plantation, in a large set of survey data these influences become thoroughly mixed up. The aim of regeneration of Norway spruce must be to achieve a good regeneration result and at the same time be prepared for early tending. To detect the need for early tending, a survey of young stands should be carried out a few years after stand establishment. The stand-establishment service provider could offer such a survey as part of an after-sales marketing system.

References

Cajander, A. 1909. Ueber Waldtypen. Acta Forestalia Fennica, 1: 1-175.

Friedman, N. & Goldszmidt, M. 1996. Building classifiers using Bayesian network. In: Proceedings of the Thirteenth National Conference on Artificial Intelligence, The AAAI Press, Menlo Park, CA., 1277-1284.

Garner, S., Cunningham, S., Holmes, G., Nevill-Manning, C. & Witten, I. 1995. Applying a Machine Learning Workbench: Experience with Agricultural Databases. Proceedings of Workshop on Applications of Machine Learning, ML'95. Tahoe City, 8 p.

Harstela, P. 2003. Taimikonhoidon vaikutus kuusen laatuun ja tuottoon [Effect of tending on quality and yield of spruce]. Metsätieteen aikakauskirja, 2/2003: 143-152. (in Finnish)

Hastie, T., Tibshirani, R. & Friedman, J. 2001. The Elements of Statistical Learning. Data mining, Inference, and Prediction. Springer Series in Statistics, 533 p.

Hyvän metsänhoidon suositukset. 2001. [Guide-lines for good forest management] Metsätalouden kehittämiskeskus Tapio. Julkaisusarja, 13: 95 p. (in Finnish)

Immonen, K. 2003. Avauspuheenvuoro: Taimikonhoidon tilanne ja kehittämistarpeet [Opening address: Situation and needs for development in tending]. Kehittyvä puuhuolto 2003. Seminaarijulkaisu. Metsäteho, 44-47. (in Finnish)

Karppinen, H., Hänninen, H. & Ripatti P. 2002. Suomalainen metsänomistaja 2000 [Finnish forest landowner 2000]. Finnish Forest Research Institute, Research Papers 852, 83 p. (in Finnish)

Kiljunen, N., Harstela, P. & Kaila, S. 2003. Ajoitus, kustannukset ja puun tuotto taimikonhoidossa [Timing, costs and yield in relation to tending]. Kehittyvä puuhuolto 2003. Seminaarijulkaisu. Metsäteho, p. 50-54. (in Finnish)

Kirkby, R. 2002. WEKA Explorer User Guide for Version 3-3-4. University of Waikato. 12 p.

Pukkala, T. 1981. Nuoren viljelysekametsikön tuotos [Yield in young mixed artificially established stand]. Metsänarvioimistieteen laudaturtyö. Helsingin yliopiston metsänarvioimistieteen laitos, 86 p. (in Finnish)

Quinlan, J.R. 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 302 p.

Saksa, T., Särkkä-Pakkala, K. & Smolander, H. 2002. Työkalu metsänuudistamisen laatutyöhön [A tool for quality management in forest regeneration]. Metsätieteen aikakauskirja, 1/2002: 29-34. (in Finnish)

Schildt, J. 2003. Avauspuheenvuoro: Puuntuotannon kannattavuus taimikonhoidon kehittämisen lähtökohtana [Opening address: Profitability of wood production as starting point for developing tending]. Kehittyvä puuhuolto 2003. Seminaarijulkaisu. Metsäteho, p. 48-49. (in Finnish)

Walfridsson, E. 1976. Lövets konkurrens i barrkulturer [Competition of broad-leaved trees in coniferous plantations]. Skogen, 63: 631-633. (in Swedish)

Witten, I. & Frank, E. 2000. Data Mining. Practical machine learning tools and techniques with Java implementations. Morgan Kaufmann Publishers, 371 p.

Received 12 March 2004

АНАЛИЗ ПОТРЕБНОСТИ В РАННЕМ УХОДЕ В МОЛОДЫХ НАСАЖДЕНИЯХ ЕЛИ ОБЫКНОВЕННОЙ

Н. Кильюнен

Для выявления потребности в раннем уходе в молодняках ели обыкновенной (*Picea abies* L.) проводился анализ воздействия структуры почвы, типа леса, метода обработки почвы, а также каменистости и влажности насаждения. Данные собраны в культурах ели 3-х летнего возраста на юге Финляндии. Многие факторы среды значительно влияли на потребность в раннем уходе. Тем не менее, классификация отдельных насаждений по признакам местопроизрастания была затруднена несмотря на применение современных алгоритмов классификации. Для точного определения потребности в раннем уходе в насаждениях ели обыкновенной рекомендуется обследование новых заложенных культур.

Ключевые слова: ель обыкновенная, уход, алгоритм классификации