Predictive models of forest development in Slovenia

Marko Debeljak¹, Bernard Ženko¹, Bernard Ženko²

Keywords: predictive clustering trees, forest development, growing stock

1. Introduction

In the area of ecological data analysis there is an increasing demand for methods and tools based on novel approaches from machine learning and information theory that would complement classic statistical methods. This would significantly increase the number of tasks that can be addressed with data analysis and provide higher quality of the analysis results. Data mining, for example, uses machine learning methods that employ approaches from classical statistics as well as information theory. Machine learning tools have been successfully used for data analysis and learning of qualitative and quantitative models from the data. Models can be written in a human readable form (e.g., decision rules and trees, equations) or in a form that can only be used for predicting new examples (e.g., neural networks, support vector machines, etc.). For the purpose of the analysis of ecological data, decision trees are frequently the method of choice. Due to their hierarchical structure, models in the form of decision trees are easy to interpret and can be used to predict values of the target variable that can be simple or structured (e.g., a vector, a hierarchy, etc.). We present a pilot study of applying data mining techniques in order to analyse the forest in Slovenia and to develop scenarios of its future development that could be used in forest management. We have focused on the analysis of the time dynamics of growing stock changes and developed models that enable us to investigate the factors that influence the total growing stock in a forest, as well as to predict the growing stock in the future.

2. Methods

The basis of our study was the Silva-SI database (Poljanec, 2008). The database comprises data from 21052 permanent forest compartments in five periods from year 1970 to 2008 (1970, 1980, 1990, 2000 and 2008). Depending on the share of state owned forest in a given compartment, we have formed two groups: "state owned" forest, which is 100% owned by the state (5237 compartments), and "private owned" forest, where the share of state owned forest is below 100% (15815 compartments). We have modelled the dynamics of growing stock changes within the compartments with predictive clustering trees (Blockeel et al., 1998) implemented in the Clus data mining system (Blockeel & Struyf, 2002). Models for predicting the total growing stock in each forest compartment in year 2018 were learned with model trees in the Weka data mining suite (Witten & Frank 2005). Besides the models for the total growing stock, we have also built separate models for different thickness classes (A, B and C), but due to space constraints, we at this point only present the models for the total growing stock in "private" forests.

¹ Jožef Stefan Institute, Department of knowledge technologies, Jamova cesta 39, 1000 Ljubljana, Slovenia. tel.: 01 4773124, e-mail: marko.debeljak@ijs.si
² University of Ljubljana, Biotechnical Faculty.
3. Results

The structure of the model describing the dynamics of growing stock changes within the compartments (Fig.1) shows that the most important attributes are phytogeographic region (an area with uniform climatic conditions and having a distinctly recognisable type of vegetation) and growing region coefficient (the quality of forest growing site ranging from 1 (the lowest) to 17 (the highest)) followed by bedrock and altitude. The trends of growing stock changes are all increasing. The model evaluation shows that going from the first period (1970) to the present the errors (absolute mean error, root mean squared error) are decreasing, while the correlation coefficient remains roughly the same through all periods.

![Diagram of the model describing the dynamics of growing stock changes in "private" forests.](image)

Fig. 1: The model describing the dynamics of the growing stock changes in "private" forests.

Descriptive models for total growing stock in different forest compartments for specific periods from year 1970 to 2008 show that the single most important factor affecting growing stock in the selected period is the growing stock in the preceding period. Given the absence of data before year 1970, the model of growing stock in year 1970 does not include the attribute for the "growing stock in the preceding period"; the most important attribute in this case is phytogeographic region. Evaluation of this model, however, gives significantly lower correlation coefficient and suggests that the model is of a limited value.

The amount of the total growing stock in year 2018 was assessed by extrapolating the trends from the leaves of the model tree learned for the year 2008 (Fig.2). We have validated this methodology by extrapolating the model for the growing stock in year 2000 to the year 2008 and compared these predictions with the real data for the year 2008. The differences between the predicted and real values were small (the largest difference was 9.2%), so we believe this approach can give us sound estimates of the growing stock in the future.
The coefficients of extrapolated linear models LM1, LM2 and LM3 for year 2018 suggest further increase of growing stock, provided that the forest management will not significantly change in the next decade. The largest increase can be expected in compartments that had low growing stock in 2008, while the smallest increase will be in compartments with already high growing stock (Table 1).

Table 1: Predicted growing stock for "private" forest in year 2018.

<table>
<thead>
<tr>
<th></th>
<th>LM1</th>
<th>LM2</th>
<th>LM3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extrapolation function coefficients; y=kx+n (k, n)</td>
<td>9.48; -18815</td>
<td>5.84; -11464</td>
<td>1.83; -3336.6</td>
</tr>
<tr>
<td>Average true wood stock in year 2008</td>
<td>224.4</td>
<td>280.0</td>
<td>351.7</td>
</tr>
<tr>
<td>Predicted wood stock in year 2008</td>
<td>224.5</td>
<td>279.6</td>
<td>351.7</td>
</tr>
<tr>
<td>Predicted wood stock in year 2018</td>
<td>319.3</td>
<td>338.1</td>
<td>370.1</td>
</tr>
<tr>
<td>Growth index (year 2008 =100)</td>
<td>142.3</td>
<td>120.8</td>
<td>105.2</td>
</tr>
</tbody>
</table>

4. Conclusions

Based on the obtained results we can conclude that data mining methods can be applied on the database *Silva-SI* with the purpose of induction of models describing the dynamics of growing stock changes and models for predicting the growing stock in the future. The proposed modelling approach allows for the optimization of the model structure according to the characteristics of the data, thus allowing the identification of data clusters and determines the most influential factors and interactions between them. We can envision that using qualitative decision support modelling would enable us to link the presented results of data mining to the existing domain expert knowledge, which could result in an efficient and user-friendly forest management decision support system.
References


