PUBLISHED ARTICLE

Citation*

Aertsen W., Kint V., Van Orshoven J., Ozkan K., Muys B. 2010. Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. Ecological Modelling 221: 1119–1130

DOI: 10.1016/j.ecolmodel.2010.01.007

Authors: Wim AERTSEN[a], Vincent KINT[a], Jos VAN ORSHOVEN[a], Kürşad ÖZKAN[b], Bart MUYS[a*]

[a] Division Forest, Nature and Landscape, Katholieke Universiteit Leuven, Celestijnenlaan 200E-2411, BE-3001 Leuven, Belgium;

[b] Orman Fakultesi, Süleyman Demirel Universitesi, Isparta, Turkey.

* Corresponding author Bart Muys, Tel. +3216329721, Fax +3216329760, E-mail bart.muys@ees.kuleuven.be

Title page

a. TITLE: Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests

b. AUTHORS: Wim AERTSEN[a], Vincent KINT[a], Jos VAN ORSHOVEN[a], Kürşad ÖZKAN[b], Bart MUYS[a]

c. AFFILIATIONS: [a]Division Forest, Nature and Landscape, Katholieke Universiteit Leuven, Celestijnenlaan 200E Box 2411, BE-3001 Leuven, Belgium; [b]Orman Fakultesi, Süleyman Demirel Universitesi, Isparta, Turkey.

d. CORRESPONDING AUTHOR: Bart Muys, Tel. +3216329721, Fax +3216329760, E-mail bart.muys@ees.kuleuven.be

e. PRESENT ADDRESSES: same

f. ADDRESS FOR PROOFS: Bart Muys, Division Forest, Nature and Landscape, K.U.Leuven, Celestijnenlaan 200E Box 2411, BE-3001 Leuven, Belgium.

Abstract

Forestry science has a long tradition of studying the relationship between stand productivity and abiotic and biotic site characteristics, such as climate, topography, soil and vegetation. Many of the early site quality modelling studies related site index to environmental variables using basic statistical methods such as linear regression. Because most ecological variables show a typical non-linear course and a non-constant variance distribution, a large fraction of the variation remained unexplained by these linear models. More recently, the development of more advanced non-parametric and machine learning methods provided opportunities to overcome these limitations. Nevertheless, these methods also have drawbacks. Due to their increasing complexity they are not only more difficult to implement and interpret, but also more vulnerable to overfitting. Especially in a context of regionalisation, this may prove to be problematic. Although many non-parametric and machine learning methods are increasingly used in applications related to forest site quality assessment, their predictive performance has only been assessed for a limited number of methods and ecosystems.

In this study, five different modelling techniques are compared and evaluated, i.e. multiple linear regression (MLR), classification and regression trees (CART), boosted regression trees (BRT), generalized additive models (GAM), and artificial neural networks (ANN). Each method is used to model site index of homogeneous stands of three important tree species of the Taurus Mountains (Turkey): *Pinus brutia*, *Pinus nigra* and *Cedrus libani*. Site index is related to soil, vegetation and topographical variables, which are available for 167 sample plots covering all important environmental gradients in the research area. The five techniques are compared in a multi-criteria decision analysis in which different model performance measures, ecological interpretability and user-friendliness are considered as criteria.

When combining these criteria, in most cases GAM is found to outperform all other techniques for modelling site index for the three species. BRT is a good alternative in case the ecological interpretability of the technique is of higher importance. When user-friendliness is more important MLR and CART are the preferred alternatives. Despite its good predictive performance, ANN is penalized for its complex, nontransparent models and big training effort.

Keywords: Artificial neural networks; Boosted regression trees; Forest site classification; Generalized additive models; Multi-criteria decision analysis; Multiple linear regression; Predictive modelling

1. Introduction

In forestry, accurate estimation of site productivity is crucial for good forest resource management (Seynave et al., 2005). Productivity is very dependent on the quality of the site (i.e. the collective of physical and biotic factors present at a given location). Forest research has a long-standing tradition of studies concerning the impact of biotic and abiotic characteristics such as climate, topography, soil and vegetation on site productivity (*e.g.,* Amen, 1945). To estimate forest site quality, foresters face the problem of integrating all these site factors. Moreover, the forest itself is an important site-forming factor, which makes only approximations possible unless forest and site are considered as a complex interrelated ecosystem (Spurr and Barnes, 1980). Because of this complexity, for most areas in Europe and North America forest site quality has been derived only empirically from the tree species specific dominant height of an even-aged tree population of known age and rescaled to a reference age, termed site index (SI) (Fontes et al., 2003).

For several applications, however, it is not possible to measure this site index in a direct way, e.g. in mixed, uneven-aged stands, for stand conversion to another tree species, for afforestation of non- forested land, or because site conditions changed over time. By linking dominant height to environmental variables (Corona et al., 1998; Curt et al., 2001), landscape characteristics (Iverson et al., 1997) and understory vegetation data (Bergès et al., 2006), site quality can be estimated at non-monitored sites. Most of the early site studies predicted forest growth from one or a few environmental variables that could be measured in the field relatively easy and at low cost. Several studies have tried to model site index by coupling age and tree height measurements to abiotic site properties but with alternating success (see *e.g.*, Corona et al., 1998; Chen et al., 2002; Bergès et al., 2005). Many of these yielded low accuracy and a high degree of variation (Kayahara et al., 1998; Curt et al., 2001).

Linear regression is one of the oldest and most widely used statistical techniques for modelling site quality because of its easy use and straightforward interpretability (Curt et al., 2001; Seynaeve et al., 2005). Although a powerful approach in particular situations when appropriately applied, many ecological relations are typically non-linear. Data often have a non-constant variance distribution and many explanatory variables show collinearity. As a consequence, linear regression may not be appropriate or may lead to high unexplained variation (Guisan et al., 2002).

More recently, the development of more advanced non-parametric and machine learning techniques and the growing availability of geodatasets at high spatial resolution are opening up plenty of opportunities to predict forest site quality with greater accuracy. Despite the flexibility of these techniques to account for non-linear relationships, they are more vulnerable for overfitting the data, *i.e.* fitting noise resulting in unstable regression coefficients (Harrell et al., 1996; Guisan and Thuiller, 2005). Also the implementation, the capacity to integrate the models with other software and the interpretability of these models can become complicated and should be weighted against the improvement in accuracy and precision.

Non-parametric and machine learning techniques that may be better fit to address the mentioned problems of linear regression should be identified and their performance compared. In the domain of forest site quality assessment McKenny en Pedlar (2003) successfully used classification and regression trees (CART) to model site index from environmental variables for two boreal tree species in Canada. The performance of non-parametric techniques as CART, generalized additive models (GAM) and artificial neural networks (ANN) compared to parametric techniques was investigated by Moisen and Frescino (2002) for the prediction of several species independent forest characteristics in the

Interior Western United States. Wang et al. (2005) also evaluated these techniques for the spatial prediction of site index of Lodgepole pine in Canada. Both studies concluded that these non-parametric approaches can be more effective predictors. Boosted regression trees (BRT), an extension of CART, is a promising technique used in ecological research on species distributions and seems to be a powerful tool for all kind of ecological modelling (Leathwick et al., 2006; Guisan et al., 2006; Elith et al., 2008). Recently, a number of software programs have been developed, incorporating many of the mentioned techniques for the prediction of species distributions (Thuiller et al.,2009). Yet no such tool exists for the prediction of continuous response variables as site index. Many studies already concluded that there is no general best modelling technique, but depending on the scope and the goal of the study some techniques will probably be better suited than others in particular situations. This study can be a good guideline to acquire more insight in the strength of the different techniques to model site index.

There is no single definite test to evaluate models, and many model predictive performance measures have been formulated (Guisan and Zimmermann, 2000; Moisen and Frescino, 2002; Wang et al., 2005). Moreover, other factors such as the ecological interpretability or the user-friendliness of a technique can be of importance in making a final evaluation and ranking of site index modelling techniques (Maggini et al., 2006). Multi-criteria decision analysis (MCDA) is a family of commonly used methodologies to assist in complex decision-making situations, as it allows the consideration of multiple criteria in incommensurate units (i.e. combination of quantitative and qualitative criteria) to provide a final ranking of alternative decisions (Herath, 2004; Mendoza and Martins, 2006). The aim of this study is to compare and evaluate two statistical non-parametric (GAM, CART), one machine-learning (ANN) and one hybrid modelling techniques (BRT) for modelling site index. Although not expected to provide the best performance, multiple linear regression (MLR) is included in this study for its straightforward interpretability and as a benchmark against which other techniques can be compared. Each method is used to model site index in homogeneous stands of three important tree species of the Taurus Mountains (Turkey): *Pinus brutia* Ten. (Calabrian pine), *Pinus nigra* ssp. *pallasiana* (Arnold) K. Richt (Crimean pine) and *Cedrus libani* A. Rich. (Lebanon cedar). The specific objectives of this study are:

(1) to compare the modelling techniques with respect to their predictive performance;
(2) to rank the modelling techniques according to predictive performance and user-oriented criteria including user-friendliness and ecological interpretability.

2. Material and methods

2.1 Study area

The study area (55 000 ha) covers the Ağlasun forest district (37°33′N, 30°32′E, 350–2200 m above sea level) in southern Anatolia, Turkey. The region has a cold and sub-humid Mediterranean climate with pronounced winter precipitation and summer drought (Paulissen et al., 1993). Limestone is the predominating parent material. Locally also conglomerates and sandstones are present. Soil depth, moisture regime and stoniness vary with topography. Most soils can be classified as leptosols, regosols or cambisols (FAO et al., 1998), depending on shallowness and stoniness (Fontaine et al., 2007).

The study area is covered for 53% by Mediterranean mountain forests mainly composed of *Quercus coccifera* (Kermes oak) (11 000 ha), *Pinus brutia* (Calabrian pine) (10 500 ha), *Juniperus* spp. (6000 ha) and *Pinus nigra* (Crimean pine) (2500 ha). Some relic stands of *Cedrus libani* (Lebanon cedar) (about 900 ha) forest occur as well (Fontaine et al., 2007). The study focuses on three tree species with expected distribution and corresponding site quality needs along a height gradient: *Pinus brutia*, *Pinus nigra* and *Cedrus libani*.

Pinus brutia is a characteristic species of the eastern Mediterranean basin and is ecologically and economically one of the most important tree species of Turkey. Its typical elevation range is between 0 and 1500 meter above sea level (a.s.l.) with a mean annual temperature between 12 and 20°C and a mean annual precipitation between 400 and 2000 mm (Boydak, 2004).

Pinus nigra subsp. *pallasiana* also occupies the eastern Mediterranean basin but is typical for higher elevations. Its optimum is located between 1000 and 1200 m a.s.l.. The species occurence in humid conditions in Greece and in drier environments in Turkey illustrates its ecological flexibility (Quézel, 1980; Fontaine et al., 2007).

Cedrus libani is significant from the historical, cultural, aesthetic, scientific and economic perspectives and is presently found primarily in the Taurus Mountains of Turkey. Socio-economic problems associated with grazing and other land uses have reduced its historical distribution drastically, but the almost inaccessible topography of the Taurus Mountains has prevented the species from becoming locally extinct. *Cedrus libani* occurs generally between 800 and 2100 m a.s.l. with a mean annual temperature ranging from 6 to 12°C and a mean annual precipitation between 600 and 1200 mm (Boydak, 2003).

2.2 Data collection

Data were collected in the summers of 2005 and 2006. To maximize spatial variation in the dataset, transects were established throughout the study area, principally oriented from valley to ridge, perpendicular to the contour lines, according a random-stratified sampling design. Due to the limited number of appropriate cedar forests in the Ağlasun forest district, additional cedar forests in neighbouring districts were selected (i.e. Kasnak National Forest (37° 44' N, 30° 49' E), Prof. Dr. Bekir Sıtkı Evcimen Taurus cedar Protection forest, Senirkent (38° 05' N, 30° 41' E) and Gölishar forest (36° 53' N, 29° 27' E)). Along those transects, 167 plots (20 m × 20 m) were established at random intervals, covering contrasting topographic situations: 65 plots of *Pinus brutia*, 46 of *Pinus nigra* and 56 of *Cedrus libani*. Plot locations were mapped using altimeter and GPS (Fig. 1). The mean distance between neighbouring plots was 614 m (with a minimum of 71 m), and no relevant spatial dependency has been observed.

Environmental variables collected as a basis for modelling site index are summarized in Table 1. The position of the plots in the landscape along the vertical gradient, soil surface roughness and landform were recorded at sight. Surface stoniness (%) and soil depth were assessed using the rod penetration method (Eriksson and Holmgren, 1996) at 10 random locations in each plot. Slope (%) was measured using a clinometer. Aspect was recorded as the azimuth (θ) measured from true north and transformed to a radiation index using the equation TRASP = $[1 - \cos((\pi/180)(\theta\text{-}30))]/2$. This assigns a value of zero to land oriented in a north-northeast direction (typically the coolest and wettest orientation) and a value of one on the hotter, drier south-southwesterly slopes (Moisen and Frescino, 2002). The depth of the ectorganic horizon was measured and separated into three sublayers if present (litter–fermentation–humus).

To quantify nutrient availability, five topsoil samples (0–10 cm) were randomly collected inside each plot. Samples were mixed and analyzed in the laboratory. Soil texture was determined using the Bouyoucos hydrometer method (Bouyoucos, 1962), soil acidity (pH) was measured in distilled water, total inorganic carbonate was assessed with the Shiebler calcimeter method (Allison and Moodie, 1965) and total soil organic matter was assessed with the Walkley–Black wet oxidation method (Allison, 1965).

The species composition of each plot (woody and herbaceous) was recorded as species cover using the Braun–Blanquet scale. Plots were assigned to plant communities according to Fontaine et al. (2007).

Tree height was assessed by means of a Blume–Leiss clinometer and tree age by counting growth rings on core samples obtained using a Pressler increment borer. Site index (SI) was obtained by recalculation of the dominant height to the reference age of 100 years by means of site index tables (Kalipsiz, 1963; Öktem, 1987; Palahí et al., 2008).

2.3 Modelling techniques

2.3.1 Multiple linear regression (MLR)

The most widely used technique in site quality assessment is linear regression. Where early studies predicted site index from a single variable (single linear regression), recent studies mostly combine many predictor variables into multiple linear regression models (MLR) which leads to a higher accuracy. Using Matlab 7.5.0 (The MathWorks Inc., Natick), stepwise as well as backward MLR techniques were tested on each studied species for selecting the most important predictor variables from the suite of environmental data. Variables of nominal or ordinal scale are recoded into dummy variables (0-1 values) for correct analysis (Field, 2005).

2.3.2 Classification and regression trees (CART)

CART encompasses a non-parametric regression technique, that ‘grows’ a decision tree based on a binary partitioning algorithm that recursively splits the data until groups are either homogeneous or contain not less observations than a user-defined threshold. The predicted value of a ‘terminal’ node is the average of the response values in that node (Breiman et al., 1984). CART is a popular technique because it represents information in a way that is intuitive and easy to visualize. Preparation of candidate predictors is simplified because predictor variables can be of any type (numeric, binary, categorical, etc.), model outcomes are unaffected by monotone transformations and differing scales of measurement among predictors. Regression trees are insensitive to outliers, and can accommodate missing data in predictor variables by using surrogates (Breiman et al., 1984). The hierarchical structure of a regression tree means that the response to one input variable depends on values of inputs higher in the tree, so interactions between predictors are automatically modelled.

Regression trees were built with the *classregtree*-function of the statistics toolbox of Matlab 7.5.0. This generally results in an over-complex decision tree that needs to be 'pruned' in order to convey only the most important information (i.e. the nodes that explain the largest amount of deviance) (McKenny and Pedlar, 2003).

2.3.3 Generalized additive models (GAM)

GAM is a non-parametric extension of Generalized Linear Models (GLM), which is in turn an extension of the MLR (Hastie and Tibshirani, 1990). GAM uses transformation techniques that are independent for each predictor variable, which are counted together to calculate the response variable (Guisan and Zimmerman, 2000). This allows exploration of shapes of species response curves to environmental gradients, and allows the fitting of statistical models in better agreement with ecological theory (Frescino et al., 2001; Austin, 2002; Lehmann et al., 2003).

GAM were constructed using R version 2.7.0 (R Development Core Team, 2006) with the GRASP (Generalized Regression Analysis and Spatial Prediction) software, an extension for the R package *gam* that combines the algorithms of GAM with spatial predictions (Lehmann et al., 2003). A simple Gaussian family was specified as a link function for the normally distributed response data. Predictor variables entered the models individually using a smoothing spline with only 2 degrees of freedom to avoid overfitting.

2.3.4 Boosted regression trees (BRT)

BRT is a combination of statistical and machine learning techniques. It is one of several techniques that aim to improve the performance of a single model by fitting many models and combining them for prediction (Schapire, 2003). BRT uses two algorithms: regression is from the CART group of models, and boosting builds and combines a collection of models. This method deals with each of these components in turn (Elith et al., 2008). Boosting is a method for improving model accuracy, based on the idea that it is easier to find and average many rough rules of thumb than to find a single, highly accurate prediction rule (Schapire, 2003). Fitting multiple trees in BRT overcomes the biggest drawback of single tree models: their relatively poor predictive performance. Although BRT models are complex, they can be summarized in ways that give powerful ecological insight (Elith et al., 2008). Despite its apparent good predictive power, this technique is not so much used in ecological research.

Boosted regression trees were developed in R version 2.7.0, with the help of the BRT-extension for the *gbm* package (Ridgeway, 2006), developed by Elith et al. (2008). Models were fitted using the *gbm.step* function, and the model was simplified by reducing the number of explanatory variables with the *gbm.simplify* function.

2.3.5 Artificial neural networks (ANN)

ANN belongs to the ‘machine learning’ techniques. This technique makes links without worrying about their form, as in reality links between variables are also not always linear or exponential. In fact it mimics the human brain’s problem solving process. The network consists of several layers of nodes (neurons) that are in connection with each other. Every node is connected to the nodes in the next layer. In the input layer, the predictor variables are inserted; the output layer delivers one or more predictive values for the response variable(s). In between there are one or more hidden layers and the network is trained using an iterative method to adjust the weights of the connections between the units.

ANN has the advantage over other statistical techniques that it is more accurate, particularly when the problem or task addressed is either poorly defined or misunderstood. It is also faster than other techniques when the problem is extremely complex; and it does not require a priori knowledge of underlying process or assumptions of the structure of the target function. There are also some drawbacks of ANN: it is a 'black-box' method, in which the weights are not interpretable due to the presence of hidden layers and the non-linearity of the activation function.

In this study several supervised feed-forward neural networks were trained with one hidden layer containing several hidden nodes. Different training algorithms were tested including Backpropagation, Quasi-Newton and Levenberg-Marquardt. To contribute to the problem of overfitting, an 'early-stopping' mechanism was applied on the training process when the minimal error on a separate validation set was obtained. Neural networks were built and trained with the Neural Network toolbox of Matlab 7.5.0.

2.4 Model evaluation

A critical consideration in the evaluation of models is the selection of fair means to compare their outcome. There is no universal measure of model performance and the metrics that are chosen should correspond to the particular needs of each individual application. Single measurements are mostly insufficient. The use of multiple measurements of performance is a common and more objective occurrence (Dawson et al., 2007). Several global measures were selected to assess the predictive performance of the models and calculated with HydroTest, a web-based toolbox of evaluation metrics (Dawson et al., 2007). In the following equations, Q_i is the observed value, $\hat{Q}_i$ is the modelled value (with i=1 to n data points), $\overline{Q}$ is the mean of the observed values, $\tilde{Q}$ is the mean of the modelled values and p is the number of parameters used in the model.

The most commonly used criterion of model performance has been the coefficient of determination (R^2) (Pearson, 1896). However, a number of authors have concluded that R^2 is not a good measure to compare different models because it only informs on how well the model fits the data used to build the model, and not on how well it performs on external data. Overfitting is often the result (Cerrato and Blackmer, 1990; Drummond et al., 2003).

$$R^2 = \left[\frac{\sum_{i=1}^{n} (Q_i - \overline{Q})(\hat{Q}_i - \tilde{Q})}{\sqrt{\sum_{i=1}^{n} (Q_i - \overline{Q})^2 \sum_{i=1}^{n} (\hat{Q}_i - \tilde{Q})^2}} \right]^2 \qquad 0 \leq R^2 \leq 1 \qquad [1]$$

Better suited to evaluate the goodness-of-fit is the coefficient of efficiency (CE) (Nash and Sutcliffe, 1970), because CE is sensitive to additive and proportional differences between simulations and observations. However, like R^2, CE is overly sensitive to extreme values because it squares the values of paired differences (Harmel and Smith, 2007).

$$CE = 1 - \frac{\sum_{i=1}^{n} (Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{n} (Q_i - \overline{Q})^2} \qquad -\infty < CE \leq 1 \qquad [2]$$

The root mean square error (RMSE) is a well accepted absolute goodness-of-fit indicator for continuous response variables, which describes the difference in observed and predicted values in the appropriate units (Harmel and Smith, 2007). The relative root mean square error (RRMSE) is calculated by dividing the RMSE by the mean observed data.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{n}} \qquad 0 \le RMSE \qquad [3]$$

$$RRMSE = \sqrt{\frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{n}} \cdot \frac{1}{\overline{Q}} \qquad 0 \le RRMSE \qquad [4]$$

Some modelling techniques are over-parameterized, which may lead to uncertainty in parameter estimation and consequently to uncertainty in model predictions. The purpose of using Akaike information criteria (AIC) or Bayesian information criteria (BIC) is to find an optimal trade-off between an unbiased approximation of the underlying model and the loss of accuracy caused by estimating a number of parameters, and the number of data points used in its calibration. These criteria combine some measure of fit with a penalty term to account for model complexity, and therefore tend to result in more parsimonious models (Senthil Kumar et al., 2005; Dawson et al, 2007).

$$AIC = n\ln(RMSE) + 2p \qquad [5]$$

$$BIC = n\ln(RMSE) + p\ln(n) \qquad [6]$$

Finally, the adjusted coefficient of determination (adjusted R^2) is a modification of R^2 that also adjusts for the number of explanatory terms used in a model. Unlike R^2, adjusted R^2 increases only if the new term improves the model more than would be expected by chance. Adjusted R^2 can be negative, and will always be less than or equal to R^2.

$$\text{adjusted } R^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1} \qquad -\infty < \text{adjusted } R^2 \leq 1 \qquad [7]$$

2.5 Model predictive performance

A common way to measure the predictive performance on a test set is by means of a 'split sample', in which a subsample of the observation data is withheld from training and used to measure the accuracy of prediction. In small data sets, this single measure can be quite misleading and very dependent on the validation subset, and cross-validation has generally been accepted to be superior to the split-sample techniques (Stone, 1974; Drummond et al., 2003; Maggini et al., 2006). In this study 10-fold cross-validation is used to assess model predictive performance. In 10-fold cross-validation, the data are divided into 10 subsets of equal size. The regression technique is then applied 10 times, each time leaving out one of the subsets and using that subset to compute the prediction accuracy. Predictive performance is quantified by calculating model evaluation measures on the predicted values for cross-validation.

2.6 Other model evaluation criteria

Model predictive performance measures are not the only criteria for evaluating modelling techniques, also qualitative criteria as 'ecological interpretability' and 'user-friendliness' can be of importance in the evaluation. 'Ecological interpretability' is referring to the degree in which the model incorporates the relative importance of predictor variables and how site index is changing with changes in predictor variable(s). This is particularly important for understanding and checking the ecological soundness of the model. 'User-friendliness' is referring to the simplicity of the technique, the statistical and technical background necessary to apply the technique and the simplicity to upscale the results to full-coverage site index maps.

2.7 Multi-criteria decision analysis (MCDA)

MCDA is applied to make a final ranking of site index modelling techniques. MCDA techniques provide solutions to problems involving multiple and conflicting objectives. Analytic hierarchy process (AHP) is a powerful and flexible MCDA technique for dealing with complex problems where both quantitative (*e.g.* predictive performance) and qualitative (*e.g.* ecological interpretability) aspects need to be considered (Saaty, 1980), and is therefore chosen as the appropriate technique for this study. AHP compares alternatives pair-wise and provides an overview of the complex relationships between decision elements (i.e. criteria and alternatives). The resulting rankings of alternatives are both transitive and complete (Gilliams et al., 2005). The essence of the process is the decomposition of a complex problem into a hierarchy with the goal (objective) at the top of the hierarchy, criteria and sub-criteria at levels and sub-levels of the hierarchy, and decision alternatives at the bottom of the hierarchy (Fig. 2). Elements at given hierarchy levels are compared in pairs to assess their relative preference with respect to each of the elements at the next higher level. The method computes and aggregates their eigenvectors until the composite final vector of weight coefficients for alternatives is obtained. The entries of the final weight coefficients vector reflect the relative importance (value) of each alternative with respect to the goal stated at the top of the hierarchy (Pohekar and Ramachandran, 2004). MCDA-analysis was performed with Super Decisions 2.0.8 software.

Four hierarchical levels were defined for the MCDA; with three main criteria at the second level: predictive performance, ecological interpretability and user-friendliness (Fig. 2). Where ecological interpretability and user-friendliness are not further divided into sub-criteria, model evaluation is split into three uncorrelated performance measures (i.e. RMSE, AIC and adjusted R^2), after combined correlation and principal component analysis.

Because the determination of the weights can vary very much depending on end-user preferences, two different scenarios were developed: a scientific scenario and a forest management planning scenario. Scenarios are implemented by defining different weights at the second level (Table 2). Since it is obvious that the models should be a good representation of the reality, predictive performance is given in both scenarios the highest weight. The scientific scenario defines further weights from the point of view of a typical researcher, for whom the ecological interpretability of model outcome is more important than the user-friendliness of the technique. The planning scenario, on the other hand, assumes that for forestry or restoration purposes the main interest lies in the possibility of applying models in a straightforward way to new field situations, and hence in the user-friendliness of the technique. At the third level predictive performance is subdivided into three uncorrelated sub-criteria: RMSE, AIC and adjusted R^2. Whereas the RMSE is an absolute goodness-of-fit indicator, both AIC and adjusted R^2 are indices considering also model complexity. For this reason, half of the weight is given to RMSE while the other half is divided over AIC and adjusted R^2. The weights at this level remain identical under both scenarios. For both scenarios MCDA is applied to each species separately, and for all species together (keeping all previous weights the same, and with identical weights for the individual tree species). A sensitivity analysis reveals how the preferences are changing with changing weights.

3. Results

A total of 15 SI-models were built, using 5 modelling techniques for each of the 3 species. All models were critically investigated for confounding factors and collinearity between explanatory variables and checked whether all basic assumptions were met. The three studied species clearly differ in site quality needs, as expressed by the different models (Table 3). Only easting and soil organic matter content seem to be common predictors for site index for all species, whereas *Pinus brutia* further responds to e.g. landscape position, *Pinus nigra* to soil pH and *Cedrus libani* to average soil depth. Elevation is important for both *Pinus* species, while *Cedrus* is more affected by slope.

The number of predictor variables entering the models is ranging from one to six, while also the predictor variables selected by each technique are not identical. Model output of the *Pinus nigra* models obtained with MLR, BRT and GAM are given as an example in equation [8], Fig. 3 and Fig. 4 respectively.

$$SI_{(P.\ nigra)} = 100.7 - 2.7 \cdot 10^{-4} * \text{Easting} - 0.3 * \%\ \text{Lime in the soil} - 2.3 * \text{Rough soil} \qquad [8]$$

The measures of performance are summarized for each model in Table 4. Better model performance is realized with lowest (R)RMSE, AIC and BIC values and with R^2, adjusted R^2 and CE closest to unity. A distinction is made between the values for model calibration and values for 10-fold cross-validation, while for the evaluation of the performance only the predictive performance, i.e. the validation values, are taken into account. By comparing the predictive performance of all models, similar trends can be observed for each species. The best goodness-of-fit, i.e. lowest values for RMSE and the highest R^2 and CE, is obtained by the ANN-models, followed by the GAM and BRT-models respectively. MLR and especially CART are scoring worse for these indicators. When the complexity of the models is taken into account, which is the case for the AIC, BIC and adjusted R^2 evaluation measures, ANN suddenly performs extremely poor in most cases. GAM is still performing very well for *Pinus nigra* and *Cedrus libani* and only worse for *Pinus brutia*, while the opposite is observed for BRT. The less complex models like MLR and CART score relatively better when model performance is penalised for the complexity of the model.

Plotting the model residuals versus the predicted values learns that for all the models the residuals are randomly distributed and no trends or bias are observed. For all techniques but especially for CART and BRT the range of the predicted values, i.e. the difference between the maximum and the minimum value, is narrowed very much in comparison to the range of the observed values (Table 5).

Scores of the modelling techniques for the quantitative sub-criteria, *i.e.* RMSE, AIC and adjusted R^2, at the bottom level of the MCDA, are a rescaling of their quantitative outcomes, which is species dependent. For the qualitative criteria 'ecological interpretability' and 'user-friendliness' a relative ranking was made of the alternatives on a [1-10] scale, based on the experiences of the authors, and rescaled into relative importance vectors (Table 6, see discussion section for details). Final MCDA rankings of the modelling techniques for the different species under the two scenarios are expressed on a relative scale compared to the preferred technique (Fig. 5). For the scientific scenario, GAM seems to be the overall best modelling technique, where BRT is a good alternative technique in case of *Pinus brutia* and *Pinus nigra*, but less for *Cedrus libani,* while in latter case CART and MLR are the best alternatives. For the planning scenario the preferences are less consistent between the different tree species. Due to the high weight of user-friendliness in this scenario, easy applicable techniques as MLR and CART score remarkably better than in case of the scientific scenario. Nevertheless, GAM is still scoring the best for *Pinus nigra* and in the overall situation, and scoring moderately good for *Pinus brutia* and *Cedrus libani*. Unexpectedly, ANN reaches the best score for *Pinus brutia*; apparently the high predictive performance for this species outweighs the penalty for user-friendliness and ecological interpretability in this case.

The performances of the different techniques for *Pinus brutia* are very similar (Fig. 5) and consequently the relative ranks may change considerably with different weights. The sensitivity analysis reveals that ANN remains the preferred technique as long as the weight of the ecological interpretability is below 0.2; for weights higher than 0.2 GAM becomes the preferred technique, whereas CART scores the best for weights of user-friendliness higher than 0.45. Conversely, the models for *Pinus nigra* are rather insensitive to changes in weights at this level. The preferences remain identical under both scenarios (Fig. 5). Only in case the weight of the user-friendliness would rise above 0.5, CART would be preferred above GAM. For the *Cedrus libani* models preferences change little between the two scenarios. GAM remains the preferred model as long as the weight of the user-friendliness is lower than 0.35; for weights higher than 0.35 CART becomes the preferred technique. This is exactly the point that defines the planning scenario, which explains the little difference in preference between GAM and CART for this scenario. ANN would only become the preferred technique here in case the weight of the predictive performance would rise above 0.8.

4. Discussion

4.1 Predictive performance

Based on our data, non-parametric techniques outperform MLR for predicting site index. Only CART performed for all species worse than MLR, which was also observed by Moisen and Frescino (2002) in predicting other forest characteristics. Leathwick et al. (2006) concluded from their study on modelling demersal fish species richness that due to their capability for fitting interactions among predictor variables, BRT appears to offer considerable performance gains over modelling techniques as GAM. Also Moisen et al. (2006) found for the prediction of basal area that, although the predictions were poor, BRT-like models (stochastic gradient boosting in their study) performed better and obtained more stable results than GAM. Our study cannot confirm these findings, as based on most evaluation measures GAM models are performing better than BRT models. This may be due to the fact that BRT models, together with CART, tend to overfit stronger (cf. the difference between evaluation measures for the calibration and validation, Table 4) and to restrict the range of model predictions (cf. Table 5) more than other techniques. Nevertheless, the predictive success of ANN models in terms of goodness-of-fit, i.e. R^2, CE and RMSE, is always the highest of all modelling techniques, which makes ANN at first sight the most suited technique for predicting SI. However, when model complexity is taken into account (AIC, BIC and adjusted R^2), ANN is penalized for its complex models.

At first sight the overall performance of all models seems to be rather weak (Table 4). There are many potential sources of error in the data sets used for modelling, including measurement errors, sampling bias, limitations in field data collection, genetic variability, etc. These errors may be affecting the overall accuracies of the models (Moisen et al., 2006). All the models, except the CART-model for *Pinus nigra*, performed better on the validation data than simply predicting the sample mean (as indicated by a positive CE). With R^2 values for the best models ranging from 0.55 to 0.84, the results look satisfactory compared to other studies with R^2 values ranging from 0.4 to 0.8 (McKenny and Pedlar, 2003). Also the predictive performances of this study, with RRMSE's for cross-validation ranging between 14 and 21%, are comparable or better than those found in other studies (Corona et al., 1998; Chen et al., 2002; McKenny and Pedlar, 2003; Szwaluk and Strong, 2003).

4.2 Ecological interpretability

The application of different techniques is expected to result in models which may differ considerably, as they are based on different algorithms. Both the number but also the type of explanatory variables can vary strongly (Table 3). However, the example of *Pinus nigra* shows that easting is selected by every technique as a predictor variable (Table 3), leading to a decline in SI from east to west (Eq. [8], Fig. 3 and Fig. 4). Easting is however an indirect variable, indicating a regional gradient not (well) covered by the variables measured in this study. Probably, in this case, easting is a proxy for maritime influence: a humid wind blown through the 'Kovada channel'-valley from the south-eastern to the western part of the study area, by which the air becomes drier along its way. The use of indirect gradients as predictive parameters has the drawback that the predictions are less 'eco-mechanistic' compared to predictions by models which are based on resource and direct gradients only, and so less *general* and applicable over large areas (Guisan and Zimmerman, 2000; Leathwick et al., 2006; Elith et al., 2008).

While the predictive success can be very high, it does not mean automatically that the shape of a response curve for an environmental predictor is ecologically rational (Austin, 2007). Both MLR and CART are techniques that are easy and straightforward to interpret, but too simple to describe many real-world situations (Elith et al., 2008). The recognized strength of more advanced techniques as GAM and BRT to model natural phenomena with non-linear relationships is confirmed by the SI-models of *Pinus nigra*. The partial dependence plots of the BRT (Fig. 3) and GAM model (Fig. 4) indicate an almost quadratic response of SI to soil pH with an optimum around 7.4, a variable that does not appear in the MLR or CART model. The GAM partial dependence plots for *P. nigra* (Fig. 4) together with the Gaussian link function used, could give the impression that a second order polynomial regression would also be able to fit the same quadratic response. Second order polynomial regression models have been built for this situation but they showed even no predictive improvement over the first order MLR and so the GAM is still preferred (data not shown). Where GAM seems to smoothly model the important ecological relations, BRT partial dependence plots often show a more erratic course. The unexpected little peaks are sampling data dependent and often difficult to explain or understand (Fig. 3). On the other hand, clear threshold values present in the data, are more explicitly represented in the BRT partial dependence plots by a sudden jump, and can be interesting for ecological interpretation. ANN is a “black-box” technique, with weights that are uninterpretable due to the presence of hidden layers and the non-linearity of the activation function. Its ecological interpretability is therefore poor. Foresters not inclined to make use of them when other, more easily understandable models are available (Changhui Peng and Xuenzhi Wen, 1999). Based on these experiences each technique is scored for the criterion ecological interpretability which is further used in the multi-criteria decision analysis (Table 6).

4.3 User-friendliness

Due to recent developments towards more integrated software packages, encompassing several modelling techniques and increased computational capacity, the user-friendliness of the applied techniques is improving. Nevertheless, important differences in user-friendliness between techniques are still present. Moisen et al. (2002) considered already the computation run time in the discussion about the suitability of different techniques. Also the theoretical background needed to apply the techniques or the easiness to upscale the information to develop site index maps can be of importance. CART is probably the easiest technique used in this study. Predictor variables can be of any type and the technique is little influenced by outliers or missing data in the predictor set (Elith et al., 2008). MLR is also straightforward, but requires some more data preparation as nominal and ordinal data need to be transformed. Also the underlying assumptions should be controlled for every model. The more advanced statistical techniques as GAM and BRT require a broader statistical background and running these models is more time consuming. Thanks to the software packages GRASP (Lehmann et al, 2003) and *gbm* (Ridgeway, 2006) developed for R, spatial predictions are facilitated.

Although this still remains more complicated and time consuming than for the MLR and CART models of which the results can be directly implemented in most GIS packages. ANN is the most complicated technique as it is based on artificial intelligence. Discovering the suitable number of nodes and layers by training the networks, for optimizing accuracy and generalization power, can be a big effort. The learning curve is steep and only developers with experience will become more efficient applying this technique (Changhui Peng and Xuenzhi Wen, 1999). Spatial predictions in GIS software are still complicate. Based on these experiences each technique is scored for the criterion user-friendliness which is further used in the multi-criteria decision analysis (Table 6).

Multi-criteria decision analysis

The understanding of the interrelationship between ecological theory, statistical theory and performance of statistical models is a complex issue. The assessment of ecological models may not depend solely on the prediction success (Austin, 2007). Even if the predictive performance is high, this does not necessarily mean that the relation is ecologically rational. Moreover different performance indices can result in opposite outcomes, as is shown in this study. Multi-criteria decision analysis was therefore applied and indicates GAM as the overall best modelling technique, for both scenarios and within a wide range of weightings. Only in very specific situations where very low importance is given to ecological interpretability (<0.2 for *Pinus brutia* or <0.15 for *Cedrus libani*) other techniques have a slight advantage over GAM. GAM is a flexible method offering both good model performance and good ecological insight and is therefore the preferred technique for modelling site index. Wang et al. (2005) concluded in a comparable study that GAM presented a better fit and better adaptability to extreme observations than other nonlinear and nonparametric techniques. The bad scores for ecological interpretation, user-friendliness and performance measures which account for model complexity make the MCDA ranking of ANN in most cases very low. ANN would probably perform better on very complex and large datasets, where its benefits over the other techniques would become greater than its drawbacks.

As expected BRT showed great potential for predictive modelling of site index, although this was not the case in all situations. Scores for ecological interpretability and user-friendliness were similar to those of GAM, but for the predictive performance BRT is slightly worse than GAM. Nevertheless, in another context where the predictive performance is of less interest, and analysis only serve to investigate the ecological relations between variables in a sample population, BRT can probably be preferred over GAM because of its capability for fitting interactions among predictor variables and its better fit of the calibration data (Table 4), explaining more of the variance. While overfitting is often seen as a problem in statistical modelling, it can enable an accurate description of the relationships in the data, provided that the overfitting is appropriately controlled (Elith et al., 2008).

Finally, despite the advantages of GAM over MLR to model non-linear relationships between response and predictor values, it should be noted that in case only linear relationships are existent or of importance, MLR models should be preferred over GAM models because of the lower risk of overfitting and the fewer degrees of freedom consumed for fitting the model.

5. Conclusions

Five modelling techniques were compared and evaluated for predicting the site index of three tree species in the Taurus Mountains of Turkey. Based on a multi-criteria decision analysis that simultaneously evaluated ‘Predictive performance’, ‘Ecological interpretability’ and ‘User-friendliness’ of the models, GAM is the preferred technique for modelling site index of these species. BRT is a good second choice in case the ecological interpretability of the technique is of high importance. When user-friendliness is more important MLR and CART are the preferred alternatives. ANN scores poor in most cases. Despite its very high goodness-of-fit ANN is penalized for its complex, nontransparent models and big training effort. Although in an MCDA, the determination of the criteria and their weights remains a more or less subjective matter, the outcome of the different scenarios, the sensitivity analysis and the consistency of our results over three species having clearly different site requirements suggests that also for other species and in other forest ecosystems GAM should be preferred for site index modelling.

Acknowledgements

This research was supported by the Research Fund K.U.Leuven (OT/07/046) and a Concerted Action of the Flemish Government (GOA 02/2) in the framework of the Sagalassos Archaeological project. Special thanks go to the Turkish Forest Administration offices at Burdur and Ağlasun for granting access permission to the forests and for providing valuable forest inventory data.

References

Allison, L.E., 1965. Organic carbon: Walkley-Black method. In: Agronomy Monographs 9, Part 2. Methods of analysis. American Society of Agronomy, Madison, Wisconsin, USA, pp. 1367-1378.

Allison, L.E., Moodie, C.D., 1965. Carbonate: Volumetric Calcimeter Method. In: Agronomy Monographs 9, Part 2. Methods of analysis. American Society of Agronomy, Madison, Wisconsin, USA, pp. 1389-1392.

Amen, J.T., 1945. Prediction of site index for Yellow-poplar from soil and topography. J.

Forest. 43, 662-668.

Austin, M., 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. Ecol. Model. 200, 1-19.

Austin, M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecol. Model. 157, 101-118.

Bergès, L., Chevalier, R., Dumas, Y., Franc, A., Gilbert, J.M., 2005. Sessile oak (*Quercus Petraea* Liebl.) site index variations in relation to climate, topography and soil in even-aged high-forest stands in northern France. Ann. Forest Sci. 62, 391-402.

Bergès, L., Gegout, J.C., Franc, A., 2006. Can understory vegetation accurately predict site index? A comparative study using floristic and abiotic indices in Sessile oak (*Quercus Petraea* Liebl.) stands in northern France. Ann. Forest Sci. 63, 31-42.

Bouyoucos, G.J., 1962. Hydrometer method improved for making particle size analyses of soils. Agron. J. 54, 464-465.

Boydak, M., 2003. Regeneration of Lebanon Cedar (*Cedrus libani* A. Rich.) on karstic lands in Turkey. Forest Ecol. Manage. 178, 231-243.

Boydak, M., 2004. Silvicultural characteristics and natural regeneration of *Pinus brutia* Ten. - A review. Plant Ecol. 171, 153-163.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and regression trees. Wadsworth International Group, Belmont.

Cerrato, M.E., Blackmer, A.M., 1990. Comparison of models for describing corn yield response to nitrogen-fertilizer. Agron. J. 82, 138-143.

Changhui Peng, Xuezhi Wen, 1999. Recent applications of artificial neural networks in forest resource management: an overview. Environmental decision support systems and artificial intelligence. Aaai Workshop 15-22

Chen, H.Y.H., Krestov, P.V., Klinka, K., 2002. Trembling aspen site index in relation to environmental measures of site quality at two spatial scales. Can. J. Forest Res. 32, 112-119.

Corona, P., Scotti, R., Tarchiani, N., 1998. Relationship between environmental factors and site index in Douglas-fir plantations in central Italy. Forest Ecol. Manage. 110, 195-207.

Curt, T., Bouchaud, M., Agrech, G., 2001. Predicting site index of Douglas-fir plantations from ecological variables in the Massif Central area of France. Forest Ecol. Manage. 149, 61-74.

Dawson, C.W., Abrahart, R.J., See, L.M., 2007. Hydrotest: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. Environ. Model. Softw. 22, 1034-1052.

Drummond, S.T., Sudduth, K.A., Joshi, A., Birrell, S.J., Kitchen, N.R., 2003. Statistical and neural methods for site-specific yield prediction. Trans. Asae 46, 5-14.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77, 802-813.

Eriksson, C.P., Holmgren, P., 1996. Estimating stone and boulder content in forest soils - evaluating the potential of surface penetration methods. Catena 28, 121-134.

FAO, IRIC, ISSS, 1998. World Reference Base for soil resources. Food and Agricultural Organisation of the United Nations, Rome.

Field, A. 2005. Discovering statistics using SPSS. SAGE Publications Ltd, London, 779 pp.

Fontaine, M., Aerts, R., Ozkan, K., Mert, A., Gulsoy, S., Suel, H., Waelkens, M., Muys, B., 2007. Elevation and exposition rather than soil types determine communities and site suitability in Mediterranean mountain forests of southern Anatolia, Turkey. Forest Ecol. Manage. 247, 18-25.

Fontes, L., Tome, M., Thompson, F., Yeomans, A., Luis, J.S., Savill, P., 2003. Modelling the Douglas-fir (*Pseudotsuga Menziesii* (Mirb.) Franco) site index from site factors in Portugal. Forest. 76, 491-507.

Frescino, T.S., Edwards, T.C., Moisen, G.G., 2001. Modeling spatially explicit forest structural attributes using generalized additive models. J. Veg. Sci. 12, 15-26.

Gilliams, S., Raymaekers, D., Muys, B., Van Orshoven, J., 2005. Comparing multiple criteria decision methods to extend a geographical information system on afforestation. Comput. Electron. Agric. 49, 142-158.

Guisan, A., Edwards, T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol. Model. 157, 89-100.

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J.M.C., Aspinall, R., Hastie, T., 2006. Making better biogeographical predictions of species' distributions. J. Appl. Ecol. 43, 386-392.

Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecol. Lett. 8, 993-1009.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135, 147-186.

Harmel, R.D., Smith, P.K., 2007. Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. J. Hydrol. 337, 326-336.

Harrell, F.E., Lee, K.L., Mark, D.B., 1996. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat. Med. 15, 361–387.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. Chapman and Hall, London.

Herath, G., 2004. Incorporating community objectives in improved wetland management: the use of the analytic hierarchy process. J. Environ. Manage. 70, 263-273.

Iverson, L.R., Dale, M.E., Scott, C.T., Prasad, A., 1997. A GIS-derived integrated moisture index to predict forest composition and productivity of Ohio forests (USA). Landsc. Ecol. 12, 331-348.

Kalipsiz, A., 1963. Türkiyede Karaçam (*Pinus nigra* Arnold) mesçerelerinin tabii bünyesi ve verim kudreti üzerine arastirmalar (In Turkish). Istanbul Üniversitesi Orman Fakultesi, Istanbul.

Kayahara, G.J., Klinka, K., Marshall, P.L., 1998. Testing site index site-factor relationships for predicting *Pinus contorta* and *Picea engelmannii x P. glauca* productivity in central British Columbia, Canada. Forest Ecol. Manage. 110, 141-150.

Leathwick, J.R., Elith, J., Francis, M.P., Hastie, T., Taylor, P., 2006. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. Mar. Ecol.-Prog. Ser. 321, 267-281.

Lehmann, A., Overton, J.M., Leathwick, J.R., 2003. GRASP: generalized regression analysis and spatial prediction. Ecol. Model. 160, 165-183.

Maggini, R., Lehmann, A., Zimmermann, N.E., Guisan, A., 2006. Improving generalized regression analysis for the spatial prediction of forest communities. J. Biogeogr. 33, 1729-1749.

Mckenney, D.W., Pedlar, J.H., 2003. Spatial models of site index based on climate and soil properties for two boreal tree species in Ontario, Canada. Forest Ecol. Manage. 175, 497-507.

Mendoza, G.A., Martins, H., 2006. Multi-criteria decision analysis in natural resource management: a critical review of methods and new modelling paradigms. Forest Ecol. Manage. 230, 1-22.

Moisen, G.G., Freeman, E.A., Blackard, J.A., Frescino, T.S., Zimmermann, N.E., Edwards, T.C., 2006. Predicting tree species presence and basal area in Utah: a comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. Ecol. Model. 199, 176-187.

Moisen, G.G., Frescino, T.S., 2002. Comparing five modelling techniques for predicting forest characteristics. Ecol. Model. 157, 209-225.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models 1: a discussion of principles. J. Hydrol. 10(3), 282-290.

Öktem, E., 1987. Kizilçam (In Turkish). Ormancilik Arasterma Enstitüsü (Forest Research Institute) P.K. 24, Bahçelievler, Ankara.

Palahí, M., Pukkala, T., Kasimiadis, D., Poirazidis, K., Papageorgiou, A.C., 2008. Modelling site quality and individual-tree growth in pure and mixed *Pinus brutia* stands in North-East Greece. Ann. Forest Sci. 65, 501-514.

Paulissen, E., Poesen, J., Govers, G., De Ploey, J., 1993. The physical environment at Sagalassos (Western Taurus, Turkey). A reconnaissance survey. Waelkens, M. and Poblome, J., Sagalassos II. Report of the Third Excavation Campaign of 1992, Leuven.

Pearson, K., 1896. Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. Philosophical Transactions of the Royal Society of London 187, 253-318.

Pohekar, S.D., Ramachandran, M., 2004. Application of multi-criteria decision making to sustainable energy planning - a review. Renew. Sustain. Energy Rev. 8, 365-381.

Quézel, P., 1980. Biogéographie et écologie des coniferes sur le pourtour méditerranéen. In: Pesson, P. (Ed.), Actualités d'écologie forestière. Gauthier-Villars, Paris, pp. 205–255.

R Development Core Team, 2006. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna.

Ridgeway, G., 2006. Generalized boosted regression models. Documentation on the R Package 'gbm', version 1·5–7. Available at: http://www.i-pensieri.com/gregr/gbm.shtml

Saaty, T.L., 1980. The Analytic Hierarchy Process. McGraw-Hill, NewYork, 287 pp.

Schapire, R.E., 2003. The Boosting Approach to Machine Learning: an Overview. Nonlinear Estim. Classif. 171, 149-171.

Senthil Kumar A.r., Sudheer K.p., Jain S.k., Agarwal P.k., 2005. Rainfall-runoff modelling using artificial neural networks: comparison of network types. Hydrol. Process. 19(6), 1277-91.

Seynave, I., Gegout, J.C., Herve, J.C., Dhote, J.F., Drapier, J., Bruno, E., Dume, G., 2005.

Picea abies site index prediction by environmental factors and understorey vegetation: a two-Scale approach based on survey databases. Can. J. Forest Res. 35, 1669-1678.

Spurr, S.H., Barnes, B.V., 1980. Forest ecology. John Wiley & Sons, New York, 687 pp.

Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. J. R. Stat. Soc., B 36, 111-147.

Szwaluk, K.S., Strong, W.L., 2003. Near-surface soil characteristics and understory plants as predictors of *Pinus contorta* site index in Southwestern Alberta, Canada. Forest Ecol. Manage. 176, 13-24.

Thuiller, W., Lafourcade, B., Engler, R., Araújo, M.B., 2009. BIOMOD – a platform for ensemble forecasting of species distributions. Ecography 32, 369-373 (version 0).

Wang, Y.H., Raulier, F., Ung, C.H., 2005. Evaluation of spatial predictions of site index obtained by parametric and nonparametric methods - a case study of Lodgepole pine productivity. Forest Ecol. Manage. 214, 201-211.

Tables

Table 1. Summary of the continuous site characteristics and levels of the factor variables stratified according to the three studied tree species

	Pinus brutia (n=65)			*Pinus nigra* (n=46)			*Cedrus libani* (n=56)		
Variable	Mean (S.D.)	Min.	Max.	Mean (S.D.)	Min.	Max.	Mean (S.D.)	Min.	Max.
Site index (m)	20.28 (4.45)	12.50	34.33	21.24 (4.10)	13.84	30.75	20.70 (4.75)	11.26	31.00
Easting* (m)	292130 (8412)	270577	302160	290235 (5103)	279456	297528	268863(45450)	179794	309263
Northing* (m)	4164409 (4188)	4157659	4172002	4167419 (3596)	4161088	4173650	4158382 (42512)	4080436	4220520
Elevation (m)	974 (255)	340	1345	1228 (162)	976	1775	1442 (147)	1170	1775
TRASP	0.55 (0.32)	0.02	0.98	0.45 (0.35)	0.00	0.98	0.41 (0.37)	0.01	1.00
Slope (%)	40 (21)	5	95	43 (17)	15	90	40 (17)	5	90
Surface stoniness (%)	28 (22)	0	80	38 (26)	0	80	55 (18)	20	90
Ectorganic horizon (cm)	2.5 (1.7)	0.2	8.0	2.2 (1.9)	0.2	8.5	1.4 (1.1)	0.0	3.5
Litter layer (cm)	1.3 (0.7)	0.2	4.0	1.1 (0.7)	0.2	3.0	0.6 (0.4)	0.0	2.0
Fermentation layer (cm)	0.7 (0.7)	0.0	3.0	0.7 (0.8)	0.0	4.0	0.5 (0.5)	0.0	2.0
Humus layer (cm)	0.4 (0.6)	0.0	3.0	0.5 (0.7)	0.0	3.0	0.3 (0.4)	0.0	1.0
Average soil depth (cm)	49 (28)	10	120	39 (22)	8	80	26 (11)	8	50
Sand (%)	35.78 (14.43)	3.19	74.15	40.24 (14.60)	14.06	85.08	40.42 (12.85)	21.7	78.2
Loam (%)	24.09 (5.86)	8.18	42.03	24.49 (5.45)	8.18	37.15	23.86 (5.40)	9.0	33.1
Clay (%)	40.44 (12.83)	17.67	78.69	35.27 (11.21)	5.41	55.21	34.02 (9.73)	5.4	50.8
pH	7.40 (0.24)	6.80	7.90	7.42 (0.24)	6.90	7.80	7.37 (0.56)	4.9	7.9
Total lime (%)	6.10 (9.80)	0.00	54.59	4.19 (5.87)	0.00	18.66	2.90 (4.43)	0.0	27.7
Organic matter (%)	6.40 (2.17)	1.61	10.99	7.67 (5.06)	2.20	31.23	7.97 (3.35)	2.7	26.6
Landscape position	Ridge, Upper slope, Middle slope, Lower slope, Valley								
Surface roughness	Flat, Flat-rough, Rough, Rough-rocky, Rocky								
Landform	Linear, Undulating, Convex, Concave								
Plant community	Eu-Mediterranean, Supra-Mediterranean with thin litter layer, Supra-Mediterranean with thick litter layer, Dry mountainous Mediterranean, Humid mountainous Mediterranean								
Geology	Limestone, Alluvium, Conglomerate, Micrite, Other								

* Universal Transverse Mercator (UTM) zone 36N

Table 2. Weights of the second level criteria under two scenarios for the multi-criteria decision analysis for the best technique for modelling site index

	Weights	
	Scientific scenario	Planning scenario
Model evaluation	0.50	0.50
Ecological interpretability	0.35	0.15
User-friendliness	0.15	0.35

Table 3. Overview of the predictor variables selected by the site index models developed with five techniques

Tree species	Modelling technique	Variable(s) selected by the model
Pinus brutia	MLR	Easting, Elevation, Ridge (dummy)
(n= 65)	CART	Easting, % Organic matter
	BRT	Easting, Elevation, Thickness of the litter layer
	GAM	Easting, Northing, Elevation, Landscape position, Plant community, % Loam
	ANN	Easting, Elevation, Landscape position
Pinus nigra	MLR	Easting, % Lime in the soil, Rough soil (dummy)
(n=46)	CART	Easting, % Organic matter
	BRT	Easting, Elevation, pH of soil, Slope
	GAM	Easting, Elevation, pH of soil, % Organic matter, TRASP
	ANN	Easting, Elevation, pH of soil, % Organic matter, TRASP, % Lime in the soil
Cedrus libani	MLR	Easting, Slope, Average soil depth
(n=56)	CART	Slope
	BRT	Slope, % Organic matter, Average soil depth
	GAM	Slope, Average soil depth
	ANN	Easting, Slope, Average soil depth, % Organic Matter

Table 4. Performance indices of all SI-models for the three tree species and five modelling techniques: multiple linear regression (MLR), classification and regression trees (CART), boosted regression trees (BRT), generalized additive models (GAM) and artificial neural networks (ANN). Best model performance for every evaluation measure, based on the validation data, is highlighted in bold.

Statistical index	MLR		CART		BRT		GAM		ANN	
	Calibration	Validation	Calibration	Validation	Calibration	Validation	Calibration	Validation	Calibration	Validation
P. brutia										
R^2	0.52	0.33	0.58	0.22	0.64	0.35	0.62	0.43	0.70	**0.60**
CE	0.52	0.28	0.58	0.17	0.63	0.35	0.62	0.42	0.70	**0.58**
RMSE	3.07	3.74	2.87	4.02	2.70	3.57	2.74	3.36	2.41	**2.86**
RRMSE	0.15	0.18	0.14	0.20	0.13	0.18	0.13	0.17	0.12	**0.14**
AIC	78.91	91.73	72.48	94.40	72.52	**90.71**	77.45	137.28	81.08	92.31
BIC	85.43	**98.25**	76.83	98.75	81.22	99.41	90.50	123.74	107.18	118.40
R^2adj	0.50	0.26	0.57	0.20	0.62	0.31	0.58	0.37	0.64	**0.50**
P. nigra										
R^2	0.31	0.11	0.21	0.03	0.57	0.20	0.56	0.33	0.84	**0.42**
CE	0.31	0.09	0.21	-0.33	0.55	0.19	0.54	0.33	0.84	**0.41**
RMSE	3.38	3.87	3.60	4.67	2.72	3.65	2.75	3.32	1.61	**3.12**
RRMSE	0.16	0.18	0.17	0.21	0.13	0.17	0.13	0.15	0.07	**0.14**
AIC	62.01	68.22	60.90	72.90	54.02	67.50	56.61	**65.19**	70.02	100.32
BIC	67.49	**73.71**	62.73	74.73	61.33	74.81	65.75	74.34	113.91	144.21
R^2adj	0.26	0.04	0.20	0.01	0.53	0.12	0.50	**0.25**	0.66	-0.24
C. libani										
R^2	0.34	0.27	0.28	0.21	0.44	0.26	0.34	0.30	0.74	**0.42**
CE	0.34	0.27	0.28	0.20	0.43	0.26	0.34	0.30	0.74	**0.40**
RMSE	3.82	4.03	3.98	4.20	3.55	4.05	3.83	3.95	2.39	**3.64**
RRMSE	0.18	0.19	0.19	0.20	0.17	0.20	0.18	0.19	0.12	**0.18**
AIC	81.10	83.99	79.38	82.39	76.96	84.35	79.15	**80.91**	72.85	96.36
BIC	87.18	90.06	81.41	**84.42**	83.03	90.42	83.20	84.96	97.16	120.66
R^2adj	0.30	0.23	0.27	0.20	0.41	0.22	0.31	**0.27**	0.67	0.25

R^2 = Coefficient of determination, CE = Coefficient of Efficiency, RMSE = Root mean squared error, RRMSE = relative RMSE, AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion, R^2adj = adjusted R^2.

Table 5. Minimum, maximum and the range of site index values (dominant height at 100 years, in meters) as observed in the field data and modelled from environmental variables with five modelling techniques

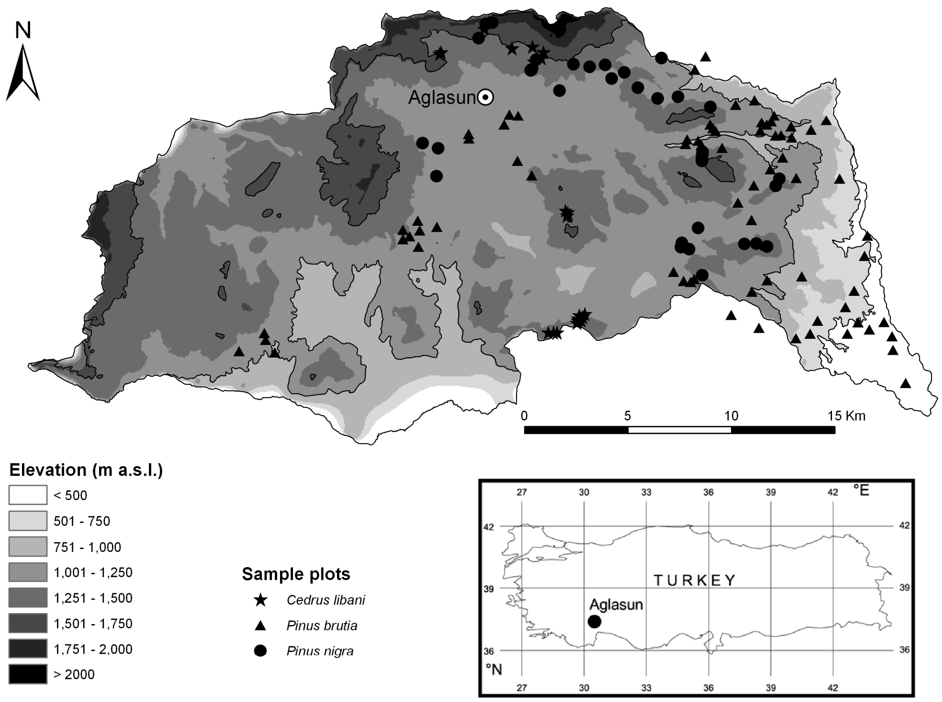
	Pinus brutia			*Pinus nigra*			*Cedrus libani*		
	minimum	maximum	range	minimum	maximum	range	minimum	maximum	range
Observed	12.50	34.33	21.83	13.84	30.75	16.91	11.26	31.00	19.74
Modelling technique									
ANN	12.81	34.33	21.52	13.84	28.20	14.36	15.98	31.04	15.07
BRT	15.16	26.81	11.65	17.17	26.88	9.70	16.34	24.64	8.30
CART	17.18	31.43	14.25	19.87	23.80	3.93	17.81	22.86	5.06
GAM	12.64	29.93	17.29	14.82	27.28	12.45	12.76	26.77	14.01
MLR	15.03	34.33	19.31	15.15	26.05	10.90	12.72	26.95	14.23

Table 6. Scores of the modelling techniques for the qualitative sub-criteria of the multi-criteria decision analysis for the best technique for modelling site index

	Scores[1]	
	Ecological interpretability	User- friendliness
ANN	0.05	0.07
BRT	0.25	0.17
CART	0.20	0.33
GAM	0.30	0.17
MLR	0.20	0.27

[1] Scores are based on the authors experiences explained in the discussion section

759 **Figures**



760

761 **Figure1.** Location of the sample plots in the Ağlasun forest district of southern Anatolia,

762 Turkey (bottom inset). Sample plots are labelled according to the dominant tree species.

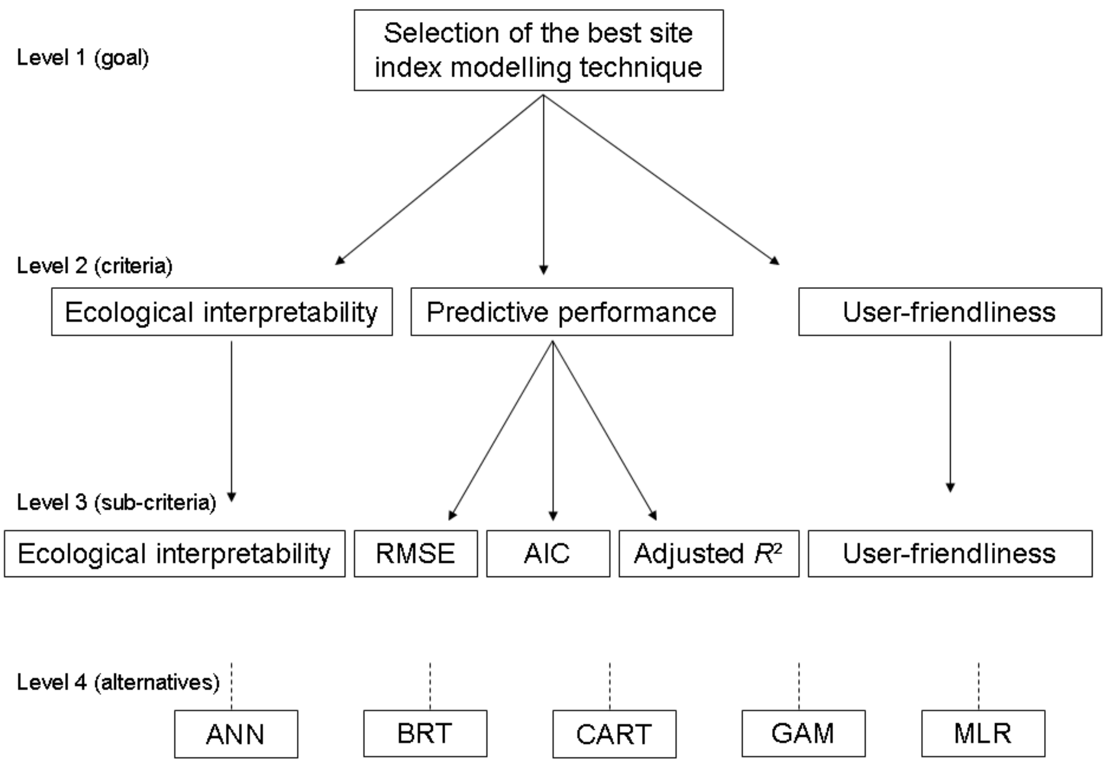


Figure2. Hierarchical structure of the multi-criteria decision analysis to evaluate the suitability of five modelling techniques for predicting site index.

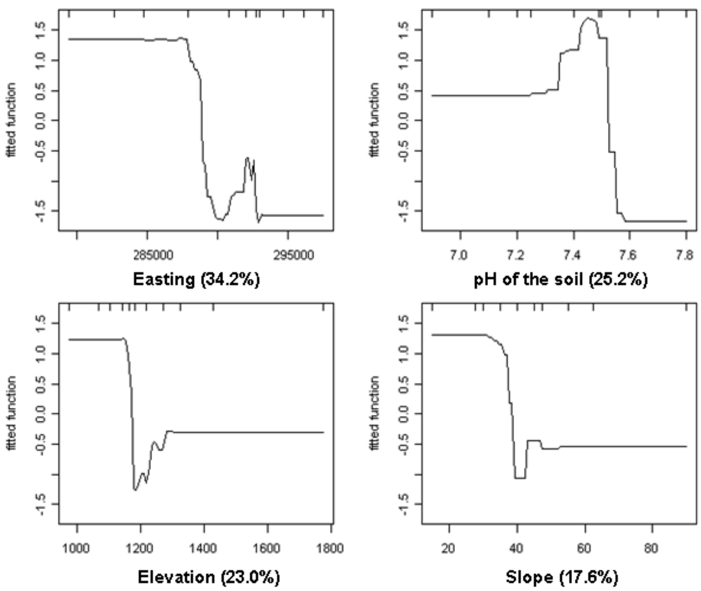


Figure 3. Partial dependence plots of the four predictor variables in the BRT-model for predicting the site index of *Pinus nigra*. The relative contribution of each predictor is reported between brackets. Rug plots at inside top of graph show distribution of sample sites along that variable, in deciles.

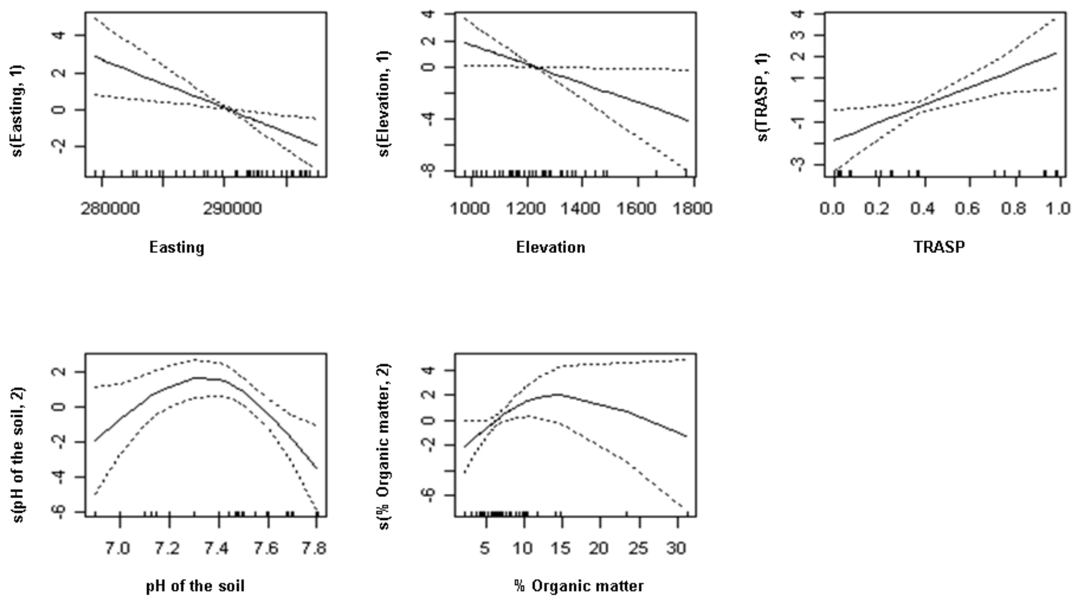


Figure 4. Partial dependence plots of the five predictor variables in the GAM-model for predicting the site index of *Pinus nigra* (full line). Dashed lines represent upper and lower twice-standard-error curves. Rug plots at inside bottom of graphs show distribution of sample sites along that variable.

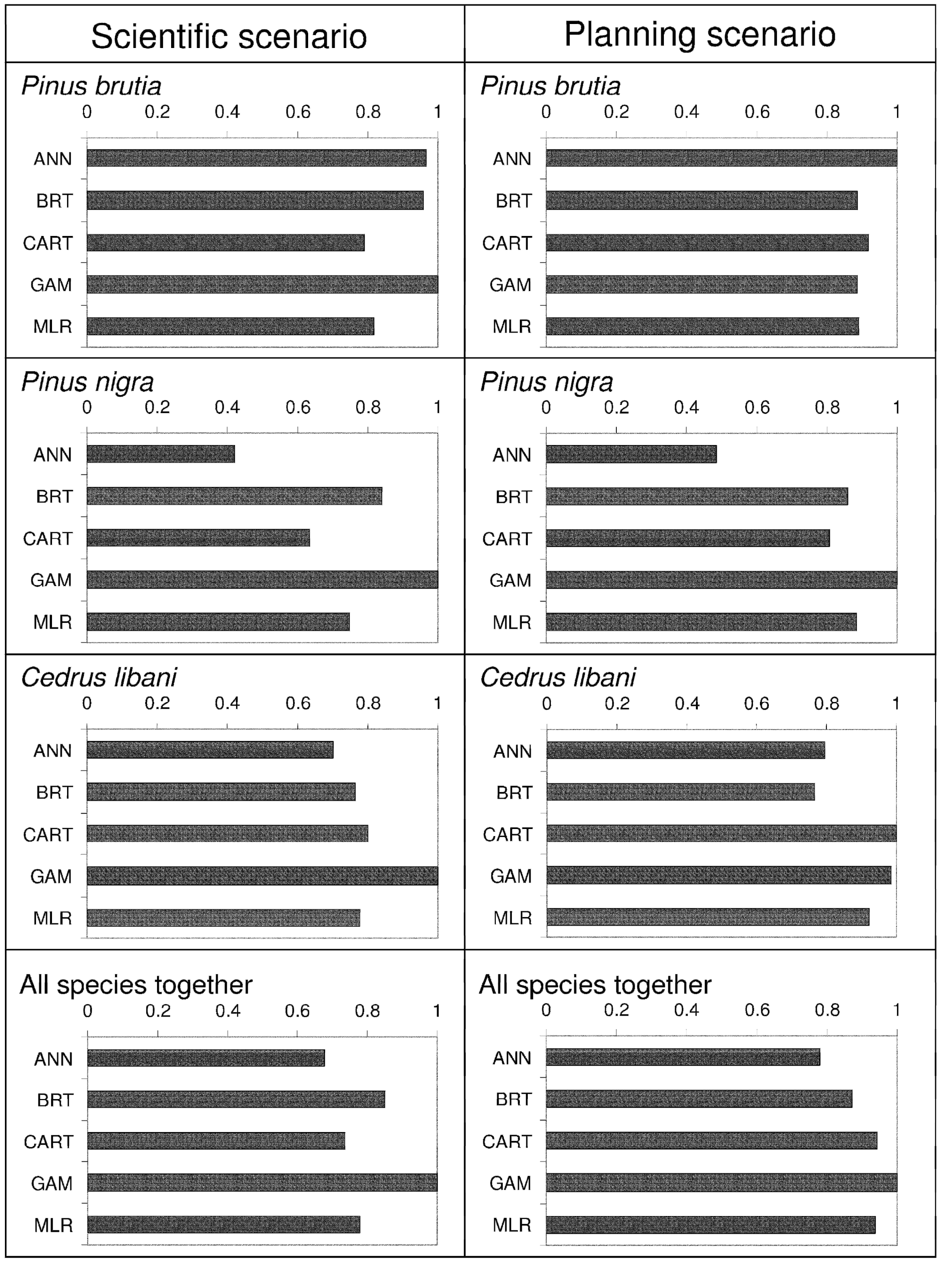


Figure 5. Results of the multi-criteria decision analysis for the suitability of five modelling techniques for predicting site index. Analysis is carried out for two scenarios for the three tree species separately and also all species together. The preferred technique is given as a value of 1, while the performance of the other techniques is expressed relatively to the best technique.