Performance of modelling techniques for the prediction of forest site index: a case study for pine and cedar in the Taurus mountains, Turkey

Wim AERTSEN, Vincent KINT, Jos VAN ORSHOVEN[1], Kürşad ÖZKAN[2], Bart MUYS[3]

Forest research has a long tradition in 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 (i.e. the dominant height of a forest stand at a reference age) to environmental variables that could be measured in the field relatively easily and at low cost, using basic statistical methods such as linear regression. Because most ecological variables show a typical non-linear and a non-constant variance distribution, a big source of error remained unexplained with the use of these linear models. More recently, the development of more advanced non-linear and even non-parametric and machine learning methods provided opportunities to overcome these limits. Nevertheless, these methods have also drawbacks. Due to their increasing complexity they are not only more difficult to implement and interpret, especially the 'black box' methods, but also more vulnerable to 'overfitting'. The challenge is located in choosing the appropriate modelling technique for a specific situation.

In this study, five different modelling techniques for predicting the site index were compared and evaluated, including multiple linear regression (MLR), classification and regression trees (CART), boosted regression trees (BRT), generalized additive models (GAM), and artificial neural networks (ANN). 167 sample plots were distributed over homogeneous stands of three important tree species of the Taurus mountains of Turkey: Pinus brutia, Pinus nigra and Cedrus libani. Soil, vegetation and topographic conditions were measured in detail and related to the site index with the five earlier mentioned modelling techniques. Except for the CART-method all methods evaluated in this study showed an improved prediction performance over the traditionally used MLR. ANN showed the overall best performance, but the complexity of the model, the training and testing effort and the interpretability taken into account, BRT and especially GAM present themselves as good alternatives.

Keywords: site index, site quality assessment, yield prediction, statistical modelling, modelling techniques

Introduction

In forestry, accurate estimation of forest productivity is very important to determine the annual allowable harvest and rotation age, or to help in the selection of most appropriate species for each location (Splechtna, 2001). 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

[1] Corresponding author: Katholieke Universiteit Leuven, Belgium

[2] Corresponding author: Süleyman Demirel Üniversitesi, Isparta, Turkey

[3] Corresponding author: Katholieke Universiteit Leuven, Belgium

productivity (e.g. Amen, 1945). To produce an estimate of the forest site quality, foresters are faced with 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 & Barnes, 1980). Because of this complexity, forest site quality was for a long time in Europe and North-America only empirically derived 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, in 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 relatively easy and at low cost be measured in the field. 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 and Bergès *et al*., 2005). Many yielded low accuracy and a high degree of variation (Kayahara *et al*., 1998 and 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 situations when appropriately applied, many ecological data have a typical non-linear and a non-constant variance distribution and many explanatory variables show a strong collinear relationship. Due to these problems, the use of linear regression techniques may lead to a high unexplained variation (Guisan *et al*., 2002).

More recently, the development of more advanced non-linear, non-parametric and even machine learning techniques and the growing availability of geodatasets at high resolution, provided the domain of site quality assessment plenty of opportunities to predict forest site quality with greater accuracy. On the other hand, the implementation, interpretability and exchange capacity of these models can become very complicated which should be weighed 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) used successfully classification and regression trees (CART) to model site index from environmental variables. Moisen & Frescino (2002) and Wang *et al*. (2005) compared the performance of non-parametric methods as generalized additive models (GAM) and artificial neural networks (ANN) with parametric methods and concluded that these non-parametric approaches are more effective predictors. Boosted regression trees (BRT) is a promising technique used in ecological research on species distribution (Elith *et al*., 2008). This is an extension of CART and seems to be a powerful tool in ecological modelling (Leathwick *et al*., 2006).

In this study different site index models are developed for three common tree species in the Taurus mountains in southern Anatolia, Turkey: *Pinus brutia* Ten. (Calabrian pine), *Pinus nigra* ssp. *pallasiana* (Arnold) K. Richt (Crimean pine) and *Cedrus libani* A. Rich. (Lebanon cedar). We compare the performance of several promising modelling techniques: CART, BRT, GAM and ANN. 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 models can be compared. We also address the advantages and disadvantages of using these techniques.

Materials and Methods

Study area

The study area (55,000 ha) was the Ağlasun forest district (37°33′N, 30°32′E, 350–2200 m above sea level) in southern Anatolia. 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 content and stoniness vary with topography. Most soils can be classified as leptosols, regosols or cambisols (FAO *et al*., 1998), depending on shallowness and stoniness.

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 the three species that are the most important for forestry purposes: *Pinus brutia*, *Pinus nigra* and *Cedrus libani*.

Data collection

Data were collected in the summer periods of 2005 and 2006. To maximize spatial variation in the dataset, transects were laid out throughout the study area, principally oriented from valley to ridge, perpendicular to the contour lines. Along those transects, one hundred and sixty-seven 20 m × 20 m plots were selected and mapped using altimeter and GPS.

The position of the plot 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 & 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 collected at random locations inside each plot. Subsamples were combined 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).

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).

Model description

Multiple linear regression (MLR)

The most widely used technique in site quality assessment is undoubtedly linear regression. Where early studies predicted site index from a single variable relationship (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.), stepwise as well as backward MLR techniques were tested on each species for selecting the most important predictor variables from the suite of environmental data. Models can be described in the following way:

$$SI_{100} = ß_0 + ß_t x_t + ß_s x_s + ß_v x_v + ß_* x_* + \varepsilon \qquad [1]$$

where SI_{100} is site index at the age of 100 years; x_t, x_s, x_v, x_* are variable vectors corresponding to topographic, edaphic and vegetative factors as well as combination variables, the $ß$ represent model coefficients and ε is the additive error term (Corona *et al*., 1998). Variables of nominal or ordinal scale have been recoded into dummy variables (0-1 values) for correct analysis (Field, 2005).

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 (McKenny and Pedlar, 2003). The predicted value of a 'terminal' node is the average of the response values in that node (Moisen & Frescino, 2002). 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. 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 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 (The MathWorks Inc.). This results generally 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 & Pedlar, 2003).

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. 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, and their predictive performance is superior to most traditional modelling methods (Elith *et al*., 2008). Despite it apparent good predictive power, this technique is not so much used in ecological research (Moisen *et al*., 2006).

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). Fitting the models was carried out with the *gbm.step* function, where a simplification of the model to reduce the number of explanatory variables was done with the *gbm.simplify* function. The most appropriate learning rate (lr), tree complexity (tc) and bag fraction (bf) were determined by trial and error.

Generalized additive models (GAM)

GAM is a non-parametric extension of Generalized Linear Models (GLM), which is in turn an extension of the MLR (Moisen *et al*., 2006). GAM uses transformation techniques that are independent for each predictor variable, which are counted together to calculate the response variable (Guisan & 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 (Austin, 2002; Lehmann *et al*., 2003). A limitation of the smoothed functions obtained from GAM, is their inability to extrapolate outside the range of the data used to build the model. Therefore, values of the validation data set that were outside the range of the model-building data set, are assigned the maximum/minimum value of the respective variable in the model-building data set (Frescino *et al*., 2001).

The analysis of GAM was carried out in R version 2.7.0 with the GRASP (Generalized Regression Analysis and Spatial Prediction)-software (Lehmann *et al*., 2003), an extension for the R package *gam* that combines the algorithms of GAM with spatial predictions. A simple Gaussian family was specified as a link function for the continuous response data. Predictor variables entered the models individually using a smoothing spline with only 2 degrees of freedom to avoid 'overfitting'.

Artificial neural networks (ANN)

ANN belong to the 'machine learning' techniques. Because in reality links are not always linear or exponential, this technique makes links between several variables, without worrying about the form of these links. 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 of response variable. 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. On the other hand there are also some drawbacks of ANN: it is usually treated as a 'black-box', with which the weights are uninterpretable due to the presence of hidden layers and the non-linearity of the activation function. There is also time required to train and test

neural networks. The learning curve is steep and only developers with experience will become more efficient using this technique (Changhui Peng & Xuenzhi Wen, 1999).

In this study several supervised feed-forward neural networks were trained with one hidden layer containing several hidden nodes. Different training algorithms were tested of which 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 (The MathWorks Inc.).

Model evaluation

A critical consideration in the evaluation of modelling techniques is the selection of fair means of comparing their outcome. 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. 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. On small data sets cross-validation has generally been accepted as superior to split-sample techniques (Stone, 1974; Drummond *et al*., 2003).

There is, moreover, no universal measure of 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 data, and $\tilde{Q}$ is the mean of the modelled values.

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 means of comparing 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).

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

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 & Smith, 2007).

$$CE = 1 - \frac{\sum_{i=1}^{n} (Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{n} (Q_i - \overline{Q})^2} \qquad -\infty < CE \le 1 \qquad [3]$$

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 & Smith, 2007). The relative root mean square error (RRMSE) is calculated by dividing the RMSE by the mean observed data.

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

$$RRMSE = \sqrt{\frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{n}} \cdot \frac{1}{\overline{Q}} \qquad 0 \le RRMSE \qquad [5]$$

Some modelling techniques are overparameterized, which may lead to uncertainty in parameter estimation and consequently to uncertainty in model predictions. The use of information criteria as 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 (*p*), and the number of data points used in its calibration (*m*). 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).

$$AIC = m\ln(RMSE) + 2p \qquad 0 \le AIC \qquad [6]$$

$$BIC = m\ln(RMSE) + p\ln(m) \qquad 0 \le BIC \qquad [7]$$

Results and Discussion

A total of 15 SI-models were built, using 5 techniques for each of the 3 species. All models were critically investigated for confounding factors and checked whether the basic assumptions were met. A summary of the variables that were used to build each of these models is presented in Table 1. The number of predictor variables entering the models is ranging from one tot six, and also the predictor variables selected by each technique can vary much.

The example of *Pinus nigra* shows that longitude is selected by every technique as a predictor variable. Even though, longitude is an indirect gradient, indicating a gradient that is not (well) covered by the variables measured in this study, with a decline in SI from east to west (Eq. [8], Fig. 1 & Fig. 2).

$$SI_{(P.\,nigra)} = 100.7 - 2.7 * \text{Longitude} - 0.3 * \%\ \text{Lime in the soil} - 2.3 * \text{rough soil (dummy)} \qquad [8]$$

Probably, in this case, longitude is a proxy for a maritime influence: a humid wind blown through the 'Kovada channel'-valley from the south-eastern to the western part of the study area, of which the air becomes drier along its way. The use of indirect gradients as predictive parameters has the drawback that the predictions are less physiologically 'mechanistic' than with models based on only resource and direct gradients, and so less *general* and applicable over large areas (Guisan & Zimmerman, 2000; Leathwick *et al*., 2006; Elith et al, 2008).

Besides, the partial response curves of the BRT- (Fig. 1) and GAM-model (Fig. 2) show that SI has an almost quadratic response to the pH of the soil with an optimum around 7.4, a variable that does not appear in the MLR- or CART-model. This confirms the strength of these more advanced techniques to model natural phenomena with non-linear relationships between the response and the set of explanatory variables (Guisan *et al*., 2002).

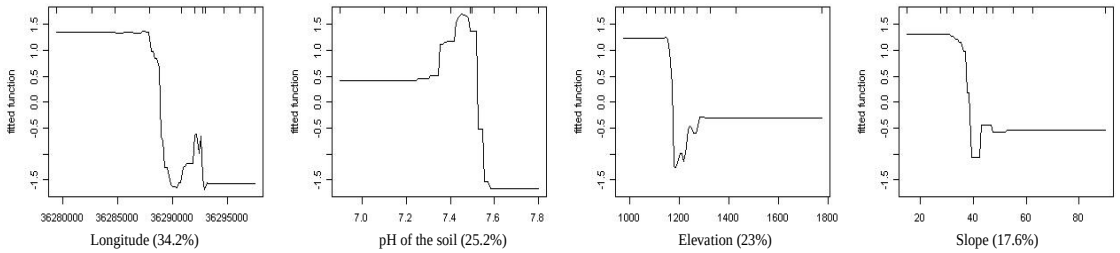


Fig. 1. Partial dependence plots of the four most influential variables in the BRT-model for predicting the SI of *Pinus nigra*. The relative contribution of each predictor is reported between brackets. Rug plots at inside top of plots show distribution of sites along that variable, in deciles.

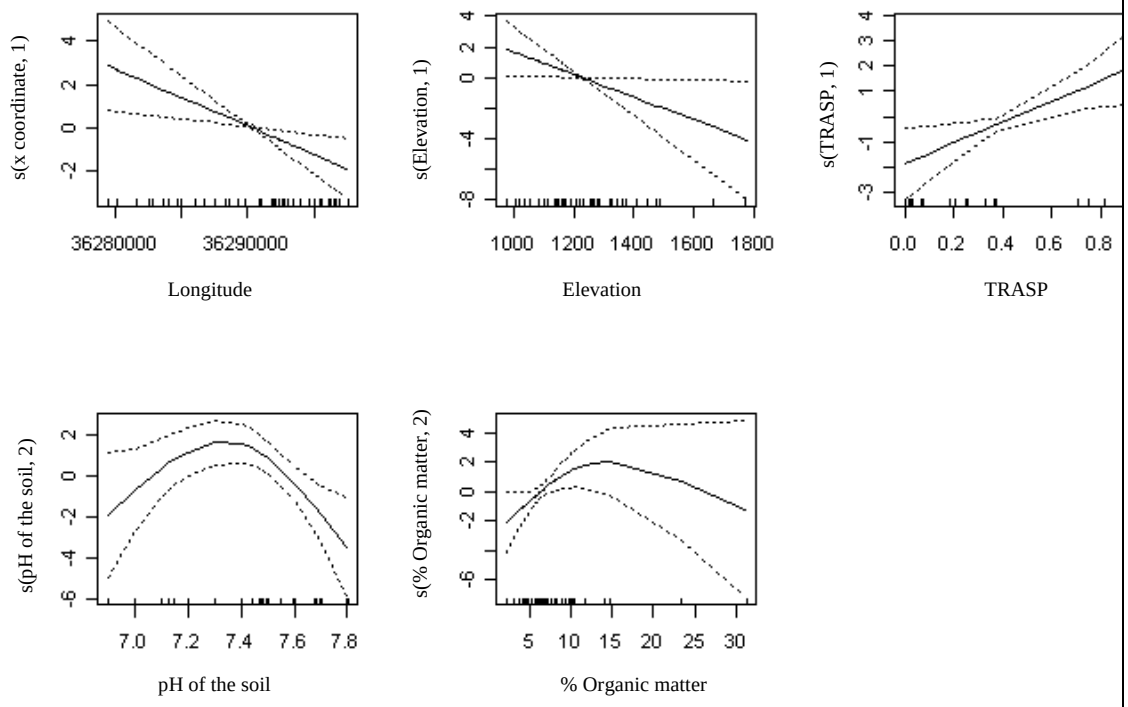


Fig. 2. Partial dependence plots for the five variables in the GAM-model for predicting the SI of Pinus nigra (full line). Dashed lines represent upper and lower point wise twice-standard-error curves. Rug plots at inside bottom of plots show distribution of sites along that variable.

Table 1. Overview of the predictor variables entering the different models.		
Tree species	**Modelling technique**	**Variable(s) entered into the model**
Pinus brutia (n= 65)	MLR	Longitude, Latitude, Elevation, Valley (dummy), Eu-Mediterranean vegetation type (dummy)
	CART	Longitude, % Organic matter
	BRT	Longitude, Elevation, Thickness of leaf layer
	GAM	Longitude, Latitude, Elevation, Landscape position, Vegetation type, % Loam
	ANN	Longitude, Elevation, Landscape position
Pinus nigra	MLR	Longitude, % Lime in the soil, Rough soil (dummy)
(n=46)	CART	Longitude, % Organic matter
	BRT	Longitude, Elevation, pH of soil, Slope
	GAM	Longitude, Elevation, pH of soil, % Organic matter, TRASP
	ANN	Longitude, Elevation, pH of soil, % Organic matter, TRASP, % Lime in the soil
Cedrus libani	MLR	Longitude, Slope, Average soil depth
(n=56)	CART	Slope
	BRT	Slope, % Organic matter, Average soil depth
	GAM	Slope, Average soil depth
	ANN	Longitude, Slope, Average soil depth, % Organic Matter

The different measures of performance are summarized for each species in Table 2. Better model performance is realized if the values of (R)RMSE, AIC and BIC are closer to zero, and R^2 and CE are closer to unity. A distinction is made between the measures based on the training set and the validation set, while for the evaluation of the performance only the predictive performance, i.e. the measures on the validation set, are taken into account. By comparing the performance of all techniques, similar trends can be observed for each species. The best performance, i.e. lowest values for RMSE and the highest R^2 and CE, is obtained with the ANN-models, followed by the GAM and BRT-models respectively. Based on these evaluation measures, CART always performed worse than MLR. On the other hand when the complexity of the models is taken into account, which is the case in the AIC and BIC evaluation measures: ANN, which are the most complex models, suddenly perform in most cases very bad, while GAM still perform very good for *Pinus nigra* and *Cedrus libani* but worse for *Pinus brutia*, and BRT perform very good for *Pinus brutia* but worse for *Pinus nigra* and *Cedrus libani*. The more simple models, MLR and CART, score relatively better when model performance is penalised for the complexity of the model.

At first sight the overall performance of the models seems to be rather weak. Moisen *et al*. (2006) stated already that the modelling of continuous response variables is a difficult task. Moreover, 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). On the other hand all-but-one models, the CART-model for *Pinus nigra*, performed better on the validation data than simply predicting the sample mean (as indicated by a positive CE) and also compared to other studies, the results look still satisfactory. Moreover, McKenny and Pedlar (2003) noted in comparable studies modelling SI, R^2 values ranging from 0.4 to 0.8, which is not better than the best models of this study with R^2 values from 0.55 to 0.84. Also the predicting performance taken into account, with RRMSE's ranging from 14 to 21%, are the results of this study comparable or better than other studies modelling SI (Corona *et al*., 1998; Chen *et al*., 2002; McKenny & Pedlar, 2003; Szwaluk & Strong, 2003).

Table 2. Performance indices computed for all developed 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	
	Training	Validation	Training	Validation	Training	Validation	Training	Validation	Training	Validation
P. brutia										
R^2	0.42	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	82.91	95.73	72.48	94.40	72.52	90.71	77.45	137.28	81.08	92.31
BIC	93.78	106.60	76.83	98.75	81.22	99.41	90.50	123.74	107.18	118.40
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
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^2 = Coefficient of determination, CE = Coefficient of Efficiency, RMSE = Root mean squared error, RRMSE = relative RMSE, AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion .

Finally, Austin (2007) stated that the understanding of the interrelationship between ecological theory, statistical theory and performance of statistical models is a complex issue; and that evaluating ecological models may not depend solely on the prediction success. Even if the prediction success is high, this does not necessary mean that the shape is ecologically rational. Although the predictive success of ANN is in most cases very high, which makes the technique very suitable for predicting SI, it is not suitable for understanding the driving processes and mechanisms causing the differences in SI. Moreover, the selection of appropriate predictor variables for ANN was also based on the variable selection of the other techniques, which makes ANN especially useful for modelling SI in combination with other techniques, of which the relations are interpretable in a direct way (MLR, CART) or by means of partial dependence plots (BRT, GAM) (Fig. 1 & Fig. 2).

Conclusions

Five modelling techniques were compared and evaluated for predicting the site index for three tree species in the Taurus mountains of Turkey. Except for the CART-method all non-parametric methods evaluated in this study showed an improvement over the traditionally used MLR. Although CART can be an interesting technique to discover important relations or interaction, the predictive performance of the technique was for all species weak compared to the other techniques. BRT and GAM performed remarkably better than MLR, and provide by means of partial dependence plots a powerful insight in the underlying relations between site index and the explanatory variables. ANN finally showed the best predictive performance, but has the drawback that it remains on the one hand time consuming to train and test the neural network adequately and on the other hand it is not self-explanatory, which makes it difficult to really understand the relations between input and output variables.

Acknowledgements

This research was supported by a Concerted Action of the Flemish Government (GOA 02/2) in the framework of the Sagalassos Archaeological project and by the Research Fund K.U. Leuven. 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, Agronomy Monographs 9, Part 2. Methods of Analysis. American Society of Agronomy, Madison, Wisconsin, USA: 1367–1378.

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

Amen, J.T. 1945. Prediction of site index for yellow-poplar from soil and topography. Journal of Forestry, 43: 662-668.

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

Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological Modelling, 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, Annals of Forest Science, 62: 391–402.

Berges, 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. Annals of Forest Science, 63: 31-42.

Bouyoucos, G.J. 1962. Hydrometer method improved for making particle size analyses of soils, Agronomy Journal, 54: 464–465.

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

Cerrato, M. E. & Blackmer, A. M. 1990. Comparison of models for describing corn yield response to nitrogen-fertilizer. Agronomy Journal, 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. Canadian Journal of Forest Research, 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 Ecology and Management, 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 Ecology and Management, 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. Environmental Modelling & Software, 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. Transactions of the ASAE, 46(1): 5-14.

Elith, J., Leathwick, J. R. & Hastie, T. 2008. A working guide to boosted regression trees. Journal of Animal Ecology, 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. 781p

Fontaine, M., Aerts, R., Özkan, K., Mert, A., Gülsoy, S., Süel, 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 Ecology and Management, 247: 18-25.

Fontes, L., Tomé, M., Thompson, F., Yeomans, A., Luis, J.S. & Savill, P. 2003. Modelling the Douglas-Fir site index from site factors in Portugal. Forestry, 76: 491-507.

Frescino, T. S., Edwards, T. C. & Moisen, G. G. 2001. Modelling spatially explicit forest structural attributes using Generalized Additive Models. Journal of Vegetation Science, 12: 15-26.

Guisan, A., Edwards, T. C. & Hastie, T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling, 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. Journal of Applied Ecology, 43: 386-392.

Guisan, A. & Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. Ecological Modelling, 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. Journal of Hydrology, 337: 326-336.

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). Landscape Ecology, 12: 331-348.

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

Kayahara, G.J., Klinka, K. & Marshall, P.L. 1998. Testing site index site-factor relationships for predicting *Pinus contorta* and *Picea engelmannii* × *P. glauca* productivity in central British Columbia, Canada, Forest Ecology and Management, 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. Marine Ecology-Progress Series, 321:267-281.

Lehmann, A., Overton, J. M. & Leathwick, J. R. 2002. Grasp: Generalized Regression Analysis and Spatial Prediction. Ecological Modelling, 157:189-207.

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 Ecology and Management, 175: 497-507.

Moisen, G. G., Freeman, E. A., Blackard, J. A., Frescino, T. S., Zimmermann, N. E., and 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. Ecological Modelling, 199: 176-187.

Moisen, G. G. & Frescino, T. S. 2002. Comparing five modelling techniques for predicting forest characteristics. Ecological Modelling, 157: 209-225.

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

Öktem, E. 1987. Kizilçam (In Turkish), Forest Research Institute P.K. 24, Bahçelievler - Ankara, Turkey.

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. Annals of Forest Science, 65: 501.

Paulissen, E., Poesen, J., Govers G. & De Ploey, J. 1993. The physical environment at Sagalassos (Western Taurus, Turkey). A reconnaissance survey. In: M. Waelkens and J. Poblome, Editors, Sagalassos II. Report of the Third Excavation Campaign of 1992, Leuven, 229–247.

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

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

Schapire, R. 2003. The boosting approach to machine learning – an overview. MSRI Workshop on Nonlinear Estimation and Classification, 2002 (eds D.D.Denison, M. H. Hansen, C. Holmes, B. Mallick & B. Yu). Springer, New York.

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. Hydrological Processes, 19(6): 1277-1291.

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. Canadian Journal of Forest Research, 35: 1669-1678.

Splechtna, B.E. 2001. Height growth and site index models for Pacific silver fir in southwestern British Columbia. B.C. Journal of Ecosystems and Management, 1(1): 1-14.

Spurr, S.H. & Barnes, B.V. 1980. Forest ecology. John Wiley & Sons, United States of America. 687p.

Stone, M. 1974. Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society Series B-Statistical Methodology, 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 Ecology and Management, 176: 13-24.

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 Ecology and Management, 214: 201-211.