ELSEVIER

Journal of Hydrology 246 (2001) 197–208

Journal of Hydrology

www.elsevier.com/locate/jhydrol

# Analysing forest transpiration model errors with artificial neural networks

Stefan C. Dekker[a,*], Willem Bouten[a], Marcel G. Schaap[b]

[a]*Department of Physical Geography and Soil Science, Institute of Biodiversity and Ecosystem Dynamics, Faculty of Science, Universiteit van Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands*

[b]*US Salinity Laboratory, 450 West Big Springs Rd, Riverside, CA, USA*

Received 15 March 2000; revised 26 September 2000; accepted 5 March 2001

**Abstract**

A Single Big Leaf (SBL) forest transpiration model was calibrated on half-hourly eddy correlation measurements. The SBL model is based on the Penman–Monteith equation with a canopy conductance controlled by environmental variables. The model has eight calibration parameters, which determine the shape of the response functions. After calibration, residuals between measurements and model results exhibit complex patterns and contain random and systematic errors. Artificial Neural Networks (ANNs) were used to analyse these residuals for any systematic relations with environmental variables that may improve the SBL model. Different sub-sets of data were used to calibrate and validate the ANNs. Both wind direction and wind speed turned out to improve the model results. ANNs were able to find the source area of the fluxes of the Douglas fir stand within a larger heterogeneous forest without using a priori knowledge of the forest structure. With ANNs, improvements were also found in the shape and parameterisation of the response functions. Systematic errors in the original SBL model, caused by interdependencies between environmental variables, were not found anymore with the new parameterisation. After the ANNs analyses, about 80% of the residuals can be attributed to random errors of eddy correlation measurements. It is finally concluded that ANNs are able to find systematic trends even in very noisy residuals if applied properly. 

*Keywords*: Artificial Neural Networks; Forest transpiration; Penman–Monteith; Model errors

## 1. Introduction

Transpiration of water by vegetation is an important component of the energy exchange at the earth surface. Single-layer, multi-layer and 3-dimensional models exist, simulating transpiration of the vegetation at local, regional, and global scale (Raupach and Finnigan, 1988). In such models it is common to describe the vegetation as if it were a Single Big Leaf (SBL) (for example SiB by Sellers et al. (1986) and BATS by Dickinson et al. (1986)). The transpiration was experimentally measured and mathematically estimated by the Penman–Monteith equation (Monteith, 1965), from the cooling effect of the forest resulting from latent heat of evaporation of the transpired water. In a hydrological context, the most important characteristic of the SBL is its stomatal resistance to transpiration. This resistance is

* Corresponding author: The Netherlands Centre for Geo-ecological research, Environmental Sciences, Utrecht University, PO Box 80115, 3508 TC Utrecht, The Netherlands. Tel.: +31-302532500.

*E-mail address:* s.dekker@geog.uu.nl (S.C. Dekker).

0022-1694/01/$ - see front matter 
PII: S0022-1694(01)00368-7

controlled by a number of environmental conditions, which can be incorporated in the SBL model with physically based or empirical response functions (Stewart, 1988). By optimising the parameters in the response functions, the SBL model can be made to fit observations of latent heat fluxes above vegetation (Dekker et al., 2000; Huntingford, 1995). However, residuals between measurements and model results still remain after calibration. These residuals are caused by random and systematic measurement errors and model inaccuracies, and may contain information that can be used to improve the SBL model.

Artificial Neural Networks (ANNs) can be used to analyse whether any patterns occur in the residuals between measured and modelled transpiration. ANNs are a very suitable tool for this purpose because they are able to find relations in complex non-linear systems, without an a priori model concept (Hecht-Nielsen, 1991; Wijk and Bouten, 1999).

Recently, Huntingford and Cox (1997) used ANNs to detect how stomatal conductance responds to changes in the local environment and compared it with the Stewart stomatal conductance model. They concluded that the Stewart-Jarvis and ANN stomatal conductance model both perform well, although the models explain different parts of the variances. In the present study we want to test a method which is less sensitive for the chosen data set by using different sub-sets of data to calibrate and validate the ANNs. Therefore we use a data set of a Douglas fir stand in the Netherlands which was already used to model forest transpiration with a SBL model by Bosveld and Bouten (1992) and Dekker et al. (2000). We explore patterns in the residuals between observed time series of transpiration and those modelled by a calibrated SBL model for the Douglas fir stand. With ANNs, we distinguish between random errors on one hand and systematic errors or model errors on the other hand. Only systematic errors with an identifiable physical basis are used to further improve the existing SBL model. Model improvements may consist of incorporation of additional environmental variables that were not considered in the original model or may be an improved response to an environmental variable. When all relevant information is incorporated in the existing SBL model, we explore the mathematical forms of the response functions.

## 2. Materials and methods

### 2.1. Research site

The research site is located in a 2.5 ha Douglas fir stand, in a large forested area, in the central Netherlands near Garderen. The Douglas Fir forest is dense with 780 trees $ha^{-1}$ without understorey and planted in 1962. Average tree height in 1995 was 25 m, lowest living whorl 13 m, mean diameter at breast height is 0.25 m and the single sided leaf area, including stem area, ranging from 9.0 $m^2$ $m^{-2}$ to 12.0 $m^2$ $m^{-2}$ in summer (Jans et al., 1994). The forested area has different stands with dimensions of a few hectares. Most dominant species are Douglas fir, Beech, Scots Pine and Japanese Larch. The soil is a well-drained Typic Dystrochrept (Soil Survey Staff, USDA, 1975), with a forest floor of 5 cm on heterogeneous ice-pushed sandy loam and loamy sand textured river deposits. The water table is at a depth of 40 m throughout the year. The 30-year average rainfall is 834 mm $y^{-1}$ and is evenly distributed over the year, mean potential evapotranspiration is about 712 mm $y^{-1}$. Yearly transpiration reduction by water stress is low (about 5%), although short periods with considerable drought stress do occur (Tiktak and Bouten, 1994).

### 2.2. Models

Forest transpiration was modelled with the Single Big Leaf model (SBL) based on the Penman–Monteith equation (Monteith, 1965):

$$\lambda E = \frac{sR_n + \rho C_p D g_a}{s + \gamma(1 + g_a/g_s)} \quad (1)$$

where $\lambda E$ is the latent heat flux (W $m-^2$), $s$ the slope of the saturated water vapour curve (mbar $K-^1$), $R_n$ the net radiation (W $m-^2$), $\rho$ the density of air (kg $m-^3$), $C_p$ the specific heat capacity of air (J $kg-^1$ $K-^1$), $D$ the vapour pressure deficit (mbar), $\gamma$ the psychrometer constant (mbar $K^{-1}$), and $g_a$ and $g_s$ are the aerodynamic and surface conductance (m $s^{-1}$), respectively.

Aerodynamic conductance ($g_a$) is calculated with (Monteith and Unsworth, 1990):

$$g_a = u_*^2/u \quad (2)$$

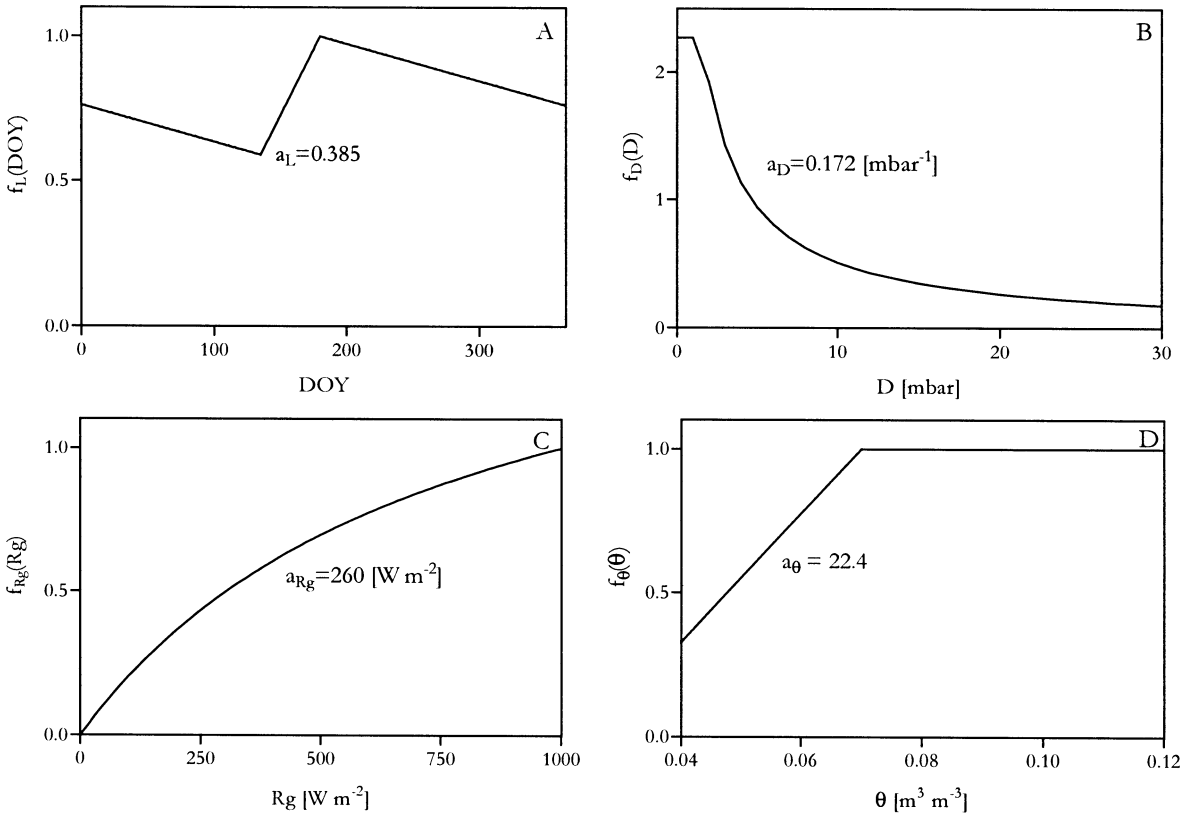


Fig. 1. Response functions to growth or leaf area index $f_L$(DOY), vapour pressure deficit $f_D(D)$, global radiation $f_{Rg}(R_g)$, temperature $f_T(T)$ and soil water content $f_\theta(\theta)$. Lines are calibrated values. Temperature response $f(_T)$ was not found and $a_T$ was fixed to zero.

where $u_*$ is the friction velocity derived from the wind profile equation under neutral conditions and $u$ is the wind speed (m/s). Friction velocity is calculated with (Monteith and Unsworth, 1990):

$$u_* = \frac{ku}{\ln\left(\frac{z-d}{z_0}\right)} \tag{3}$$

where $k$ is the von Karmann constant, $z$ the measurement height (36 m), $d$ the zero plane height taken as two thirds of the tree height (17 m), and $z_0$ is the roughness length (m). For the 1989 data set, Bosveld (1997) found $z_0$ values ranging between 1.7 and 3.6 m. Due to this large range and because the trees have grown between 1989 and 1995, $z_0$ was used as fit-parameter.

Surface conductance, $g_s$, is composed of the stomatal conductance ($g_c$) and the remaining conductance when stomata are closed ($g_0$):

$$g_s = g_c + g_0 \tag{4}$$

$g_0$ is related to culticular transport of water vapour.

For the 1989 data set, Bosveld and Bouten (1992) modelled stomatal conductance as a product of response functions of environmental variables. They found that $g_c$ depends on leaf area index ($L$), vapour pressure deficit ($D$), global radiation ($R_g$), air temperature ($T$) and volumetric soil water content ($\theta$):

$$g_c = g_{c,ref} \cdot f_L(\text{DOY}) \cdot f_D(D) \cdot f_{Rg_g}(R_g) \cdot f_T(T) \cdot f_\theta(\theta) \tag{5}$$

where the $g_{c,ref}$ is a parameter, representing the canopy conductance at reference conditions $f_i$ are reduction functions of the environmental conditions or time and DOY is Day Number of the Year. The functional shapes of the response functions, used by Bosveld and Bouten (1992), are plotted in Fig. 1.

A piece-wise linear form for the growth curve ($f_L$) was assumed. It was observed (Tiktak et al., 1991) that shoot growth starts at DOY 130 and ends at DOY 180, meaning that DOY is a surrogate for Leaf Area Index:

$$f_L(\mathrm{DOY}) = 1 - a_L(\mathrm{DOY} + 185)/315$$
$$0 \le \mathrm{DOY} \le 130$$
$$f_L(\mathrm{DOY}) = 1 - a_L(180 - \mathrm{DOY})/50$$
$$130 \le \mathrm{DOY} \le 180$$
$$f_L(\mathrm{DOY}) = 1 - a_L(\mathrm{DOY} - 180)/315$$
$$180 \le \mathrm{DOY} \le 365 \quad (6)$$

where $a_L$ is the free parameter to be optimised.

The response function for $D$ ($f_D$) is:

$$f_D(D) = \frac{1}{1 + a_D(D - D_r)} \quad (7)$$

where $a_D$ (mbar$^{-1}$) is the free parameter and $D_r$ (mbar) a reference $D$, here chosen at 4.6 mbar at which $f_D$ becomes 1. For $D < 1.5$ mbar the response function was set to $f_D(D = 1.5$ mbar).

The light response function ($f_{Rg}$) is described with:

$$f_{Rg}(R_g) = \frac{R_g(1000 - a_{Rg})}{R_g(1000 - 2a_{Rg}) + a_{Rg}1000} \quad (8)$$

where $a_{Rg}$ (W m$^{-2}$) is the free parameter and 1000 is the maximum radiation (W m$^{-2}$).

For the temperature response ($f_T$), Jarvis (1976) used a function that is forced to zero at $T = 0$ and $T = 40$°C while the optimum temperature ($T_{OPT}$) is a free parameter. A disadvantage is that the function can not set to 'no response' and therefore Bosveld and Bouten (1992) used:

$$f_T(T) = 1 - a_T + a_T\left(\frac{40 - T}{40 - T_{OPT}}\right)^{2 - T_{OPT}/20}\left(\frac{T}{T_{OPT}}\right)^{T_{OPT}/20}$$
$$0° \le T \le 40° \quad (9)$$

where $a_T$ is the free parameter and $T_{OPT}$ is set to 25°C.

The soil water content ($f_\theta$) is described with:

$$f_T(\theta) = 1 \quad \theta \ge 0.072$$
$$f_T(\theta) = 1 - a_\theta(0.072 - \theta) \quad \theta < 0.072 \quad (10)$$

where $a_\theta$ is the free parameter and 0.072 represents the so called reduction point, e.g. the starting level at which soil water stress occurs. This reduction point was found with the soil water model SWIF (Tiktak and Bouten, 1992,1994).

In summary, the SBL model has eight parameters. One parameter, $z_0$ is used to calculate $g_a$, $g_0$ accounts for canopy conductance when the stomata are closed, $g_{c,ref}$ is used to scale the five response functions, which together contain five parameters.

### 2.3. Measurements and data processing

Transpiration was calculated from measured half-hourly latent heat fluxes minus the forest floor evaporation. Only periods with a dry canopy were selected to avoid evaporation fluxes of intercepted rain. In total, 4048 half-hourly measurements remained in 1995. The latent heat flux was measured at 30 m above the forest floor with a fast response Ly-$\alpha$ hygrometer and a sonic anemometer-thermometer system (Bosveld et al., 1998). With half hourly measurements, the random error amounts to 15% of the flux (Bosveld and Bouten, 1992) with an additional offset of 5 W m$^{-2}$. The forest floor evaporation was simulated with the model of Schaap and Bouten (1997), who used a Penman–Monteith approach where surface resistance depends on the water content of the forest floor. For the same forest they measured and modelled a maximum forest floor evaporation of 25 W m$^{-2}$.

Half-hourly values of meteorological driving variables were measured by the Royal Meteorological Institute of the Netherlands (KNMI) on a 36 m high guyed mast. Short wave incoming radiation was measured with a CM11 Kipp solarimeter. Ambient temperature and humidity were measured with ventilated and shielded dry bulb and wet bulb sensors at 18 m above the forest floor. Wind speed was measured with a three-cup anemometer at 36 m above the forest floor. The soil water model SWIF, calibrated on soil water content measurements of the same forest, was used to simulate daily water contents of the forest floor and mineral soil. To obtain representative water contents of the root zone, the

Table 1
Minimum, maximum and mean values of the measured environmental variables

| | Minimum | Maximum | Mean |
|---|---|---|---|
| DOY | 101 | 293 | 190 |
| $R_g$ (W m$^{-2}$) | 0 | 953 | 223 |
| $R_n$ (W m$^{-2}$) | −80 | 710 | 135 |
| $D$ (mbar) | 0 | 32.2 | 6.1 |
| $T$ (°C) | 0.3 | 31.2 | 16.7 |
| $u$ (m s$^{-1}$) | 0.4 | 6.9 | 2.9 |
| $\lambda E$ (W m$^{-2}$) | −29 | 370 | 59 |
| $\theta$ (m$^3$ m$^{-3}$) | 0.053 | 0.118 | 0.083 |
| SoilEvap (W m$^{-2}$) | 0 | 23 | 5.6 |

simulated vertical water content profile was weighted with the root density. Range and mean values of measured environmental variables are shown in Table 1.

### 2.4. Analysis with artificial neural networks

The type of Artificial Neural Networks (ANNs) applied is a feed-forward back propagation (Haykin, 1994; Hecht-Nielsen, 1991) with three layers, an input, a hidden and an output layer. The number of input and output nodes corresponds to the number of input and output variables, while the number of hidden nodes depends on the complexity of the relations between input and output variables. At each neuron, the input values are biased and weighed by model parameters. A sigmoid transfer function for the hidden layer and a linear transfer function for the output layer provide the non-linear capabilities of the ANN. A properly calibrated neural network is able to approximate any continuous (non-linear) function (Haykin, 1994; Hecht-Nielsen, 1991), therefore neural networks are well-suited to explore the residuals between model predictions and observations. The neural network parameters were optimised with the Levenberg–Marquardt algorithm (Marquardt, 1963; Demuth and Beale, 1994) which minimises the root mean squared errors (RMSE) between measurements and model results.

When calibrating ANNs one has to cope with the flexible structure, local minima, overtraining and the high sensitivity to sets of calibration and test data (Morshed and Kaluarachchi, 1998). Problems with local minima were solved by initialising the model 20 times with different initial parameter values. Sensitivity analyses proved that 20 initialisations was enough. Problems with overtraining and high sensitivity to outliers were solved by using different sub-sets of data. The total data set was divided in independent sets for calibration and validation. The calibration data sets were randomly drawn and contain 67% of the total data set. An ANN was calibrated on a calibration set and tested on the corresponding validation data set. In total 30 calibrations-validations were carried out. The best run of the 20 initialisations was selected. Mean and standard deviation were calculated from these best runs of the 30 sets.

### 2.5. Approach and presentation of results

This study followed two main steps. In the first step the residuals of the whole SBL model were examined with ANNs. First the parameters of the SBL model were calibrated on the eddy correlation data using the Simplex algorithm (Press et al., 1988). Due to collinearity between the parameters, as found by Dekker et al. (2001), we used different initialisations to find the best fit. Subsequently, the residuals between the predicted and measured transpiration fluxes were analysed with ANNs to investigate if there are systematic deviations, which were correlated with the environmental variables. The ANN analyses of the residuals were carried out using wind direction (WD), $u$, $R_g$, $R_n$, $D$, $T$, DOY and $\theta$ as input. Results of this first step are presented in Section 3.1.

In the second step, the goal was to establish improved response functions of the $g_c$ function (Eq. (5)) to predict optimal $g_c$ by using only information of the transpiration measurements. The disadvantage of using response-functions, as defined in Eqs. (6)–(10), is that some parts of these functions are pre-set. On the contrary, ANNs do not use a-priori functions and find the best fit only based on the data and not based on these pre-set functions. To establish these optimal response functions, the $g_c$-function is first degraded by putting all the original reduction functions in Eq. (5) to 1.0 and $g_0$ to 0.0. Because no reduction functions are present in this version of the SBL model, the residuals between model outcome ($y_1$) and observations are large and are very likely correlated with one or more environmental variables. To this end we use an iterative approach based on ANN analyses of the

Table 2
Root Mean Squared Error (RMSE) and optimised free parameters of the canopy conductance model with different wind sectors

| Wind sector | 0–360 | 15–125 |
|---|---|---|
| RMSE (W $m^{-2}$) | 26.41 | 21.85 |
| $g_{c,ref}$ (mm $s^{-1}$) | 18.12 | 13.8 |
| $g_0$ (mm $s^{-1}$) | 0.50 | 0.55 |
| $a_L$ | 0.385 | 0.320 |
| $a_D$ (mbar$^{-1}$) | 0.172 | 0.129 |
| $a_{Rg}$ (W $m^{-2}$) | 260 | 283 |
| $a_T$ | 0 | 0 |
| $a_\theta$ | 22.4 | 30.0 |

residuals between the SBL model and the observed transpiration. In the first iteration, only the $g_{c,ref}$ parameter of the SBL model is recalibrated on the data-set while all reduction functions were set to 'no response'. Five ANN analyses are carried out to establish the response of the residuals to variations of $R_g$, $D$, $T$, DOY and $\theta$. The strongest response is selected and added to the predicted transpiration by the SBL model ($y_2$) and the predicted offset is transposed to $g_0$. Subsequently, $g_s$ can be found by inverting the Penman–Monteith equation with $y_1$ as transpiration flux. To obtain the functional shape of the new response, this $g_s(y_1)$ is divided by the $g_s(y_2)$. After describing the response function in appropriate mathematical terms, it is incorporated in the SBL model of the first iteration. The SBL model is subsequently recalibrated for $g_{c,ref}$, $g_0$ and the parameter of the response function and once again the residuals are analysed with ANN after which a response function is established. This iterative improvement is carried out until no meaningful improvement of the SBL model is obtained. Results of this second step are shown in Section 3.2.

Table 3
Root Mean Squared Error (RMSE) of the SBL model with different wind sectors, and improvements of fits by ANN in % of the original RMSE. ANN_$u$ means that an ANN analyses with wind speed ($u$) as input gives an improvement with respect to the original SBL model

| Wind sectors (°) | 0–360 | 15–125 |
|---|---|---|
| RMSE_SBL (W $m^{-2}$) | 26.41 | 21.85 |
| ANN_$u$ (%) | 0.8 | 0.5 |
| ANN_$WD$ (%) | 1.8 | 0.2 |
| ANN_$\theta$ (%) | 1.1 | 1.2 |
| ANN_$D$ (%) | 0.3 | 1.0 |
| ANN_$R_g$ (%) | 0.3 | 0.4 |
| ANN_$R_n$ (%) | 0.3 | 0.4 |
| ANN_$T$ (%) | 0.3 | 0.2 |
| ANN_DOY (%) | 0.2 | 0.1 |

## 3. Results and discussion

### 3.1. Systematic deviations of the residuals

Eight parameters of the SBL model were calibrated to fit the measurements (first column of Table 2). This calibration shows that no temperature response could be identified and therefore $a_T$ was fixed to zero. Bosveld and Bouten (1992) found also no temperature response for the 1989 data set. If using a similar type of temperature function and not the function of Jarvis (1976) that can not be set to 'no response', maybe 'no temperature response' could also be identified for other forests. The shapes of the four remaining response functions are plotted in Fig. 1.

Improvements in model fit of the eight ANN analyses are shown in Table 3 as percentages of the original model fit. The ANNs with WD as input showed the strongest improvement. This response together with the $u$ response is further evaluated. Fig. 2(a) and (b) shows residuals against $u$ and WD. In these figures, a positive residual means that the SBL model underestimates the measurements. A clear systematic trend is not visible because of large random errors. Fig. 2(c) and (d) shows the trend found by the ANNs. Dashed lines are the standard deviations, calculated from the best 30 ANNs, representing the reliability of the trend. Responses that vary with wind speed and direction reflect the variations in forest structure and species. Bosveld (1997) determined different roughness lengths from wind profile relations for every 30° wind sector for the 1989 data set. He found deviant values in the sectors 210–330°, which he attributed to other tree species. However, another roughness length does not lead to other transpiration values because the mean $g_a$ is 70 times smaller than $g_s$. Therefore we must focus on a source area of only one species. Fig. 2(d) shows a constant residual in the wind sector 15–185° and tends to confirm a homogeneous forest structure in that direction. With the data of this sector only, the ANN

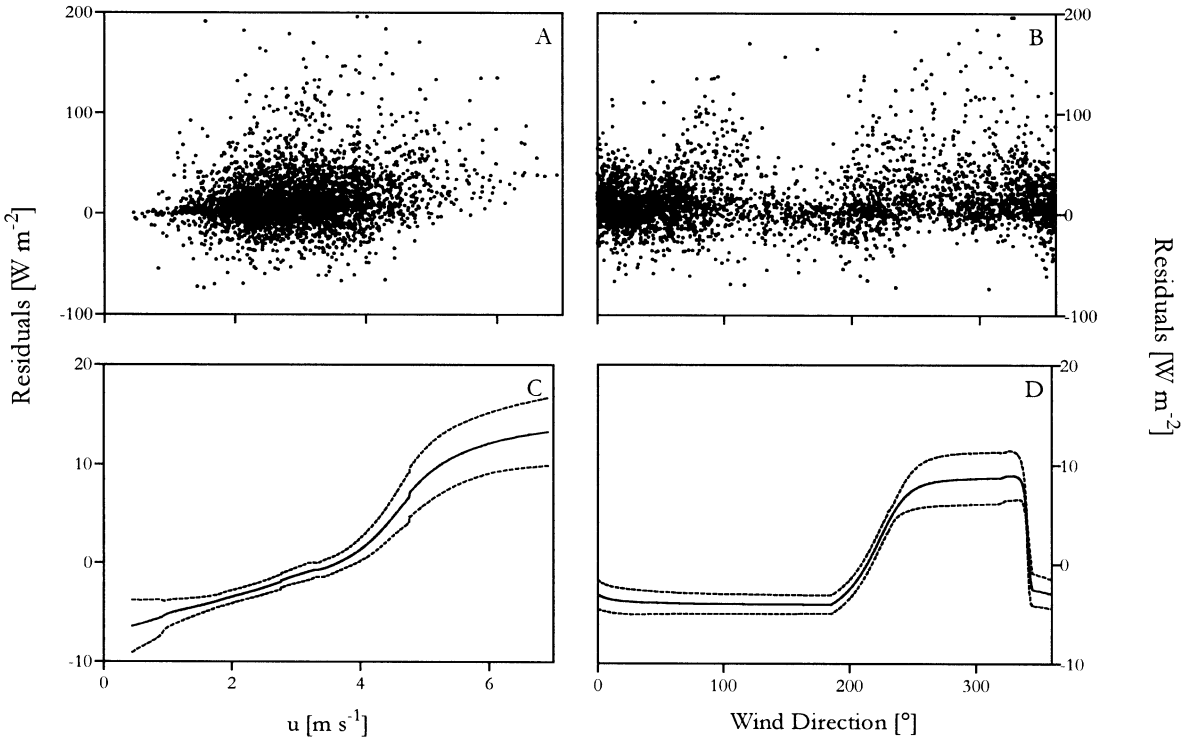


Fig. 2. Residuals between model results and measured transpiration against wind speed ($u$) and wind direction (a and b). Systematic residuals against $u$ and wind direction found by the ANNs (c and d). Dashed lines are standard deviations of the 30 best runs. A positive fit means that the original model underestimates the measurements. The scale of the a and b figures are 10 times larger.

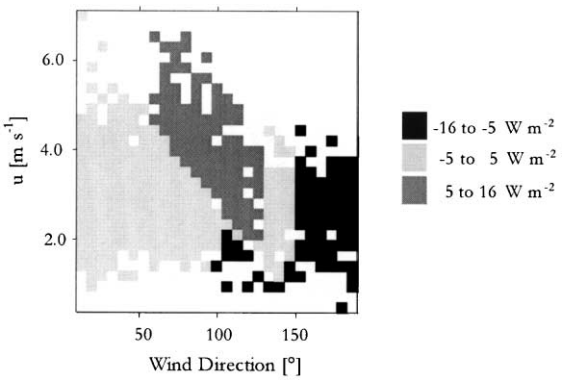


Fig. 3. Systematic errors found by the ANN with wind direction and wind speed ($u$) as input using the data with wind direction between 15 and 185°. A positive value means that the SBL model underestimates the measurements.

analysis was repeated again with $u$ and WD as input at the same time. Still an improvement of 1.6% was found. The response found by the ANN (Fig. 3) corresponds with the characteristics of the forest stand. The sectors above 125° are dominated by Scots Pine. The sector 50–125° has the largest fetch of the Douglas fir although Fig. 3 shows that the conditions are not constant against wind speed. In the SBL model, stability corrections of the boundary layer were ignored because $g_a$ has only a very small effect on transpiration. However, during unstable conditions the source area is much smaller because the cut off in the boundary layer is much steeper. Unstable conditions mainly appear with low $u$ meaning that the fetch will be smaller during these circumstances. Stable conditions will mainly appear during night with low $u$ leading to a large fetch. Information of stable conditions can not be identified because during night no transpiration and therefore no information in the data is available.

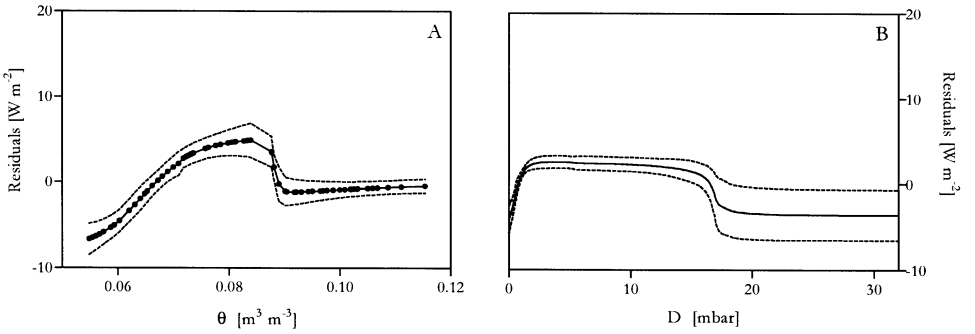


Fig. 4. Systematic trends of residuals against water content ($\theta$) and vapour pressure deficit ($D$) found by ANN on the reduced data set. Dashed lines are standard deviations of the 30 best runs. A positive value means that the original model underestimates the measurements. Dots are mean daily values.

Bouten et al. (1992) found wetter soil conditions for the wind sector 50–125° at about 150 m distance from the meteorological tower. The fetch will be that large during neutral conditions. Neutral conditions mainly appear during high $u$. The underestimation by the model between 50 and 125° and high $u$ (Fig. 3) can possibly be caused by these wetter soil conditions. As a result of this analysis, it is shown that variations in forest structure can be derived from transpiration observations.

To reduce the effect of forest structure heterogeneity and with a focus on the source area of Douglas fir only we used data from the 15–125° wind sector for further analyses resulting in a reduced data set of 1633 measurements. The SBL model was calibrated again (Table 2, column 2) and the ANN analyses were repeated (Table 3, column 2). Only a small improvement in WD remained, indicating that the forest structure is sufficiently homogeneous in the selected wind sector. The remaining improvement in $u$ is caused by different lengths of the fetch caused by unstable and neutral conditions of the atmosphere.

The ANN response to $\theta$ and $D$, which show the largest improvements, are plotted in Fig. 4. The soil water trend in Fig. 4(a) shows that the model underestimates the transpiration at $\theta$ between 0.067 and 0.088 $m^3\ m^{-3}$, overestimates at $\theta < 0.067\ m^3\ m^{-3}$ and predicts well at $\theta > 0.088\ m^3\ m^{-3}$. The model underestimation can not be caused by a wrong initial soil water stress point because in that case we should expect an overestimation of the model between 0.072 and 0.088 $m^3\ m^{-3}$. Therefore, this systematic error must be caused by the interplay of environmental variables that lead to the evolution of $\theta$. This interplay is caused by coupled environmental conditions, which are available in these kind of monitoring data sets, as pointed out by Huntingford and Cox (1997); Dekker et al. (2000). The relation of the SBL model residuals and $D$ shows a shift at 17 mbar (Fig. 5(b)). However, ANNs responses were not conclusive at higher water vapour deficits as reflected by wide uncertainty ranges. A further reduction of the data set was therefore not considered.

### 3.2. Optimisation of canopy conductance responses

To reduce effects as shown in Fig. 5(a) and (b) caused by interplays between environmental variables and being not dependent on pre-set forms of response functions (Eqs. (6)–(10)), ANNs were used to establish improved functions. In this step, improved response functions were established with an iterative approach to predict optimal $g_s$. In the first iteration, only the free parameter $g_{c,ref}$ was recalibrated on the reduced data set of Douglas fir (Table 4, first column) while all response functions were set to 1.0. The Root Mean Squared Error (RMSE) between modelled and measured transpiration was large (41.2 W $m^{-2}$). With the ANN analyses it was shown that the $R_g$ response caused the strongest reduction in the RMSE indicating that it is the most important controlling factor in stomatal behaviour. The residual fit found of this ANN analyses is plotted in Fig. 5(a), dashed lines are again the standard deviations, calculated from the best 30 ANNs, representing the reliability of the trend. Fig. 5(b) (left $y$-axis) shows the response

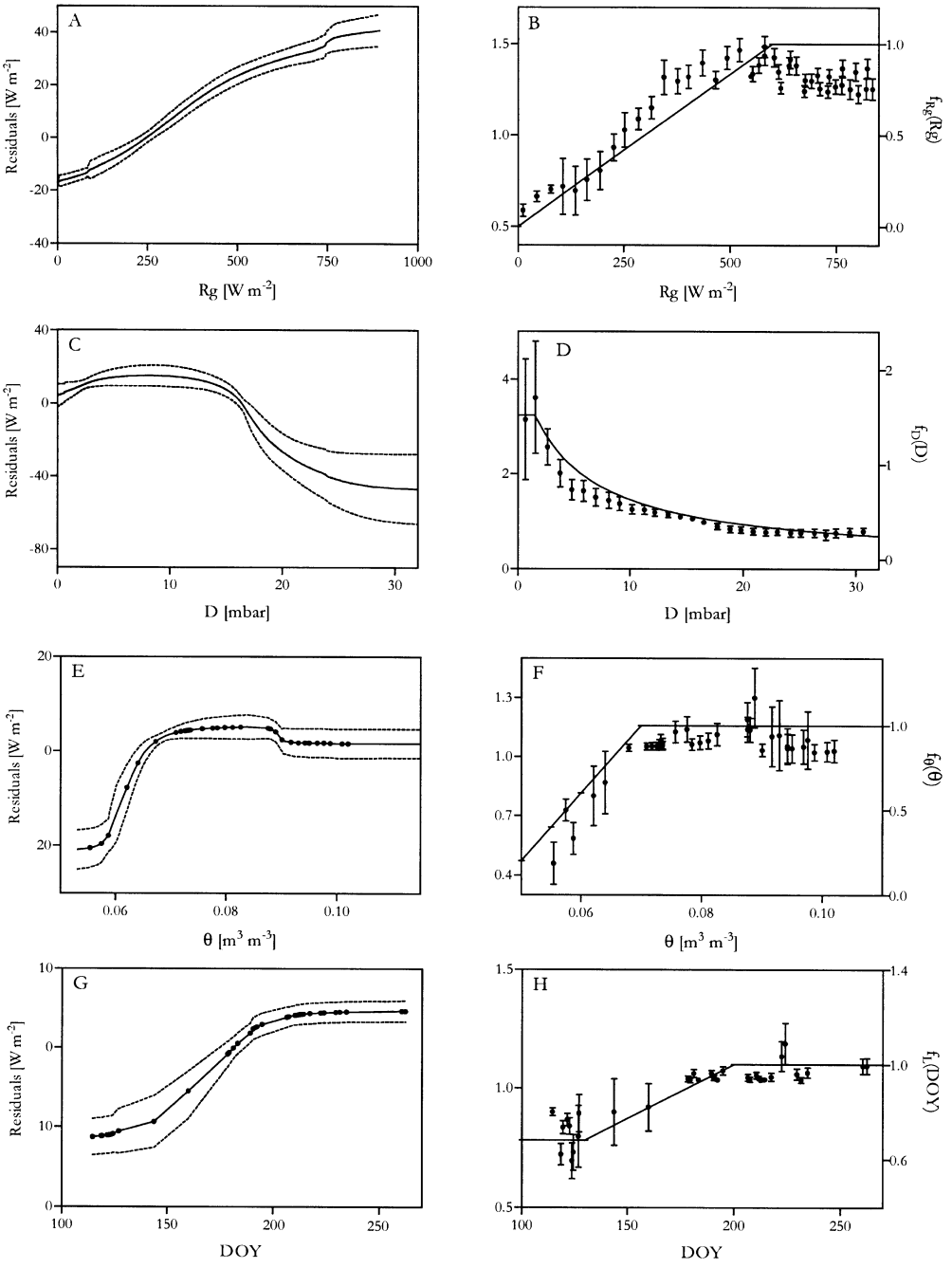


Fig. 5. (a) Shows the ANN fit to $R_g$ during iteration 1. Dashed lines are standard deviation values of the 30 best runs. (b) Shows the light response of the bulk stomatal conductance model by inverting the Penman–Monteith equation with uncertainties (left *y*-axis). Solid line is the functional shape used as response function (right *y*-axis). (c) and (d) Function with *D* (Iteration 2); (e) and (f) function with $\theta$ (Iteration 3); and (g) and (h) function with DOY (Iteration 4). Dots in (e) and (g) are daily mean values containing minimal five half-hourly measurements during daytime.

function of the conductance, which is calculated for 30 classes of $R_g$. The $R_g$ response function shows a linear trend between 0 and 600 W m$^{-2}$ with a slightly decreasing response at values above 600 W m$^{-2}$. This decreasing response is caused by the interference of $D$. A high $D$, which is correlated to a high $R_g$, causes a lower response. As this response is not used yet we neglected the decrease in the $R_g$ response. As $R_g$ response, we used a piece-wise linear function, with a maximum at $R_g \geqq 600$ W m$^{-2}$ ($a_{Rg,max}$). Minimum $R_g$ response was found at 0.5. To conform to commonly used response functions we rescaled the light response between zero and one (right axis in Fig. 5(b)) while the remaining conductance during night-time is optimised with $g_0$. Jarvis (1976); Stewart (1988) both used a non-linear light response curve as shown in Fig. 1(c), which was suggested by plant physiological studies carried out under controlled conditions. Our analysis, however, does not support a non-linear light response curve for this forest.

In the second iteration, the SBL model was optimised with $g_{c,ref}$, $g_0$ and $a_{Rg,max}$. Results of calibration and the ANN analyses are shown in Table 4. Strongest residual fit was found with $D$ (Fig. 5(c)). The high uncertainties at $D > 25$ mbar were caused by the limited number of measurement points (29). The response function (Fig. 5(d)) shows a similar shape as the original one (Fig. 1(b)). The high uncertainty in the first constant part is caused by low transpiration fluxes. The uncertainty at high $D$ seems small, 0.04 (Fig. 5(d), right axis), but the fluxes are high, resulting in a high uncertainty as shown in Fig. 5(c).

In the third iteration, the SBL model was optimised with $g_{c,ref}$, $g_0$, $a_{Rg,max}$ and $a_D$. Results of calibration and ANN analyses are shown in Table 4. Strongest response was found with $\theta$ (Fig. 5(e) and (f)). Because $\theta$ is constant during the day, only daily average values were plotted in Fig. 5(f). Although there is some scatter in the conductance plot, the soil water stress response curve is almost identical to the original one and the reduction point was also found at 0.072 m$^3$ m$^{-3}$. Moreover, an irrational shape as shown in Fig. 4(a) was not found anymore.

In the fourth iteration, the model was optimised with $g_{c,ref}$, $g_0$, $a_{Rg,max}$, $a_D$ and $a_\theta$ (results shown in Table 4). Strongest ANN response was found with DOY (Fig. 5(g) and (h)). From these growth curves, we assume that shoot growth starts at DOY 130 and ends around DOY 200. A linear decrease after DOY 200 as shown in Fig. 1(d) was not found. A systematic trend before DOY 130 could not be found due to a lack of data. Therefore constant values are assumed before DOY 130 and after DOY 200, while the steepness of the change between DOY 130 and 200 was used as free parameter.

In the last iteration, the model was optimised with $g_{c,ref}$, $g_0$, $a_{Rg,max}$, $a_D$, $a_\theta$ and $a_L$ (Table 3, column 5). No clear improvements could be found by including $T$ in

Table 4
Results of ANN analyses of five iterations. For each iteration, the calibrated parameter values and the Root Mean Squared Errors (RMSE) of the SBL model between modelled and measured transpiration is given. Five ANN analyses are carried out to establish the response of the residuals of this calibrated SBL model to variations of $R_g$, $D$, $T$, DOY and $\theta$. RMSE errors of these ANN fits are shown in the last five lines. Bold value is strongest response and new mathematical function of this variable is incorporated in the SBL model. Then the iteration is repeated by recalibrating the parameters

| | Iteration 1 | Iteration 2 | Iteration 3 | Iteration 4 | Iteration 5 |
|---|---|---|---|---|---|
| $g_{c,ref}$ (mm s$^{-1}$) | 3.8 | 3.5 | 13.2 | 13.4 | 16.7 |
| $g_0$ (mm s$^{-1}$) | – | 0.91 | 0.66 | 0.67 | 0.68 |
| $a_{Rg,max}$ (W m$^{-2}$) | – | 590 | 578 | 595 | 592 |
| $a_D$ (mbar$^{-1}$) | – | – | 0.181 | 0.159 | 0.191 |
| $a_\theta$ | – | – | – | 0.360 | 0.358 |
| $a_L$ | – | – | – | – | 0.353 |
| RMSE_SBL (W m$^{-2}$) | 41.2 | 34.6 | 25.4 | 23.0 | 20.8 |
| RMSE_$R_g$ (W m$^{-2}$) | **35.3** | 33.2 | 25.3 | 22.9 | 20.7 |
| RMSE_$D$ (W m$^{-2}$) | 40.5 | **30.2** | 25.2 | 22.8 | 20.7 |
| RMSE_$\theta$ (W m$^{-2}$) | 40.0 | 32.6 | **24.1** | 22.8 | 20.7 |
| RMSE_$DOY$ (W m$^{-2}$) | 41.0 | 33.4 | 25.3 | **22.5** | 20.7 |
| RMSE_$T$ (W m$^{-2}$) | 41.1 | 32.3 | 25.4 | 22.9 | 20.7 |

the canopy conductance model which was also found in the first calibration of the original SBL model (Table 3).

In comparison with the first analyses, presented in Section 3.1, the irrational shape part of the $\theta$ curve (Fig. 4(a)) and the shift at 17 mbar of the $D$ curve (Fig. 4(b)) from the first analysis were not found anymore. This justifies the conclusion that both systematic errors were caused by interdependencies among environmental variables meaning that the iterative approach, presented in this study, leads to a set of stomatal conductance response functions without these interdependencies. Moreover, pre-set forms of response functions, also for instance the temperature function of Jarvis (1976), can cause interplays to other variables causing apparent sensitivities.

The improvement in model fit, from 26.41 to 21.85 due to the reduction of the forest structure heterogeneity and from 21.85 to 20.82 W $m^{-2}$, due to the new parameterisation may seem small. However the random error of half-hourly eddy correlation measurement was estimated at a RMSE of 16.7 W $m^{-2}$ by Bosveld and Bouten (1992), 80% of the total error. As a result only an error of 4.1 W $m^{-2}$ remains to be explained.

This remaining error can be caused by measurement errors of the environmental conditions, model errors of soil evaporation and soil water or by the wetter soil conditions at larger distance, as shown in Fig. 3.

Improvements of the SBL model with this data set are not foreseen. As pointed out before, high uncertainties in the ANN response was found at high $D$ (Fig. 5(c)) and in the ANN response before DOY 130 (Fig. 5(g)). Both uncertainties were caused by a lack of data, meaning that these functions can be better estimated with more measurements during these specific conditions.

## 4. Conclusions

Artificial Neural Networks (ANN) show trends in residuals between results of a forest transpiration model (SBL) and eddy correlation measurements that were related to both wind speed and wind direction. They were able to localise the source area of the fluxes of the Douglas fir stand within a larger heterogeneous forest without using a priori knowledge of the forest structure. After restricting the data set to wind sections with homogeneous forest, the response functions of the canopy conductance model were also analysed with ANNs in an iterative approach. The analysis led to a piece-wise linear light response curve with saturation at 600 W $m^{-2}$ while only small changes for the other functions were found. Systematic errors in the original model were caused by interdependencies between environmental variables. These errors were not found anymore with the new parameterisations, and new functional forms of the response functions. The method presented here, that used different subsets of data to calibrate and validate the ANNs, is able to trace systematic trends even in very noisy residuals.

## Acknowledgements

The authors thank F. C. Bosveld from the Royal Meteorological Institute of the Netherlands for providing the meteorological data of 1995. We also thank Koos Verstraten and Guda van der Lee for critical comments on the text of an earlier draft and two reviewers for critically reading of this paper. The investigations were in part supported by the Earth Life Sciences and Research Council (ALW) with financial aid from the Netherlands Organisation for Scientific Research (NWO) and the University of Amsterdam.

## References

Bosveld, F.C., 1997. Derivation of fluxes from profiles over a moderately homogeneous forest. Boundary Layer Meteorology 84, 289–327.

Bosveld, F.C., Bouten, W., 1992. Transpiration dynamics of a Douglas fir forest. II: Parametrization of a single big leaf model. PhD thesis W. Bouten: Monitoring and modelling forest hydrological processes in support of acidification research,. University of Amsterdam, pp. 163–180.

Bosveld, F.C., Vliet, J.G.v.d., Monna, W.A.A., 1998. The KNMI Garderen experiment, micrometeorological observations 1988–1989. Instruments and data sets. TR-208. KNMI de Bilt, 53 pp.

Bouten, W., Heimovaara, T.J., Tiktak, A., 1992. Spatial patterns of throughfall and soil water dynamics in a Douglas fir stand. Water Resources Research 28 (12), 3227–3233.

Dekker, S.C., 2000. On the information content of forest

transpiration measurements for identifying canopy conductance model parameters. PhD thesis: Modelling and monitoring forest evapotraspiration: behaviour, concepts and parameters, University of Amsterdam, pp. 53–68.

Dekker, S.C., Bouten, W., Verstraten, J.M., 2000. Modelling forest transpiration from different perspectives. Hydrological Processes 14 (2), 251–260.

Dekker, S.C., Bouten, W., Bosveld, F.C., 2001. On the information content of forest transpiration measurements for identifying canopy conductance model parameters. Hydrological Processes (in press).

Demuth, H., Baele, M., 1994. Neural network Toolbox for use with Matlab. The Mathworks, Inc., Natick.

Dickinson, R.E., Henderson-Sellers, A., Kennedy, P.J., Wilson, M.F., 1986. Biosphere-Atmosphere Transfer schemes for the NCAR Community Climate Model, Tech. Note NCAR/RN-275 + STR, National Centre for Atmospheric Resource Boulder, CO.

Haykin, S., 1994. Neural Networks, a Comprehensive Foundation. Macmillan College Publishing Company, New York.

Hecht-Nielsen, R., 1991. Neurocomputing. Addison-Wesley Publishing Company Inc, New York.

Huntingford, C., 1995. Non-dimensionalisation of the Penman–Monteith model. Journal of Hydrology 170, 215–232.

Huntingford, C., Cox, P.M., 1997. Use of statistical and neural network techniques to detect how stomatal conductance responds to changes in the local environment. Ecological Modelling 97, 217–246.

Jans, W.W.P., Roekel, G.M.v., Orden, W.H.v., Steingröver, E.G., 1994. Above ground biomass of adult Douglas fir. A data set collected in Garderen and Kootwijk from 1986 onwards. 94/1:1-59, IBN-DLO, Wageningen, The Netherlands.

Jarvis, P.G., 1976. The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field. Phil. Trans. R. Soc. Lond. B. 273, 593–610.

Marquardt, D.W., 1963. An algorithm for least-squares estimation of non-linear parameters. J. Soc. Ind. Appl. Math. 11, 411–431.

Monteith, J.L., 1965. Evaporation and environment. In: Fogg, G.E. (Ed.), The State and movement of water in living organisms. 19th Symp. Soc. Exp. Biol. Cambridge University Press, London, pp. 205–235.

Monteith, J.L., Unsworth, M.H., 1990. Principles of Environmental Physics. Arnold, London.

Morshed, J., Kaluarachchi, J.J., 1998. Application of artificial neural network and genetic algorithm in flow and transport simulations. Advances in Water Resources 22 (2), 145–158.

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., 1988. Numerical Recipes. Cambridge University Press, Cambridge.

Raupach, M.R., Finnigan, J.J., 1988. 'Single-layer models of evaporation from plant canopies are incorrect but useful, whereas multilayer models are correct but useless': Discussion. Australing Journal of Plant Physiology 15, 705–716.

Schaap, M.G., Bouten, W., 1997. Forest floor evaporation in a dense Douglas fir stand. Journal of Hydrology 193, 97–113.

Sellers, P.J., Mintz, Y., Sud, Y.C., Dalcher, A., 1986. A simple biosphere model (SiB) for use in general circulation models. Journal Atmospheric Science 45, 505–531.

Soil Survey Staff, 1975. Soil taxonomy: a basic system of soil classification for making and interpreting soil surveys. Washington, D.C: Soil Conservation Service U.S. Department of Agriculture. Agriculture handbook, no. 436, 754 p.

Stewart, J.B., 1988. Modelling surface conductance of pine forest. Agricultural and Forest Meteorology 43, 19–35.

Tiktak, A., Bouten, W., 1992. Modelling soil water dynamics in a forested ecosystem III: model description and evaluation of discretization. Hydrological Processes 6, 455–465.

Tiktak, A., Bouten, W., 1994. Soil water dynamics and long-term water balances of a Douglas fir stand in the Netherlands. Journal of Hydrology 156, 265–283.

Tiktak, A., Bouten, W., Jans, W., Olsthoorn, A.F.M., 1991. Temporal dynamics of shoot extension and fine root activity. In: Evers, P. (Ed.), CORRELACI, Identification of traditional and air pollution related stress factors in a dense Douglas Fir ecosystem: the ACIFORN Stands, Dorschkamp Rep. 623, IBN-DLO, 299, pp. 548–550.

Wijk, M.T.v., Bouten, W., 1999. Water and carbon fluxes above European coniferous forests modelled with artificial neural networks. Ecological Modelling 120, 181–197.