EnviroInfo 2002 (Wien)
Environmental Communication in the Information Society - Proceedings of the 16th Conference
 ISBN: 3-9500036-7-3

Identifying Significant Determinants for Canopy Development on an Alpine Test Site by means of Artificial Neural Networks

Janette F. Walde[1] and Erich Tasser[2]

Abstract

We develop a methodical approach to model structural canopy development (biomass, leaf area) of different grassland stands in the North Italian landscape. Classic punctual field measures are linked to temporal-spatial high-resolution remote sensing data using artificial neural networks in order to generate large-scale forecasts for canopy development. We find empirical evidence to support our claim that RGB (red, green, blue) colour values can contribute to a better understanding of canopy development over time and in space. We provide graphical and statistical measures to identify the form and the importance of influence factors on canopy development. This approach allows us to scale up the plot-level measurements to landscape-level measurements (e.g. from biomass data to a biomass map).

1. Introduction

Detailed studies of canopy development (biomass, leaf area) can only be conducted at a limited number of locations. If we want to assess the dynamics and spatial differences on a larger regional scale (as this is important for land and water resource management), we must use spatial technologies like remote sensing. Several authors (e.g. Shroder et al. 1995) have already employed such approaches and some (e.g. Peddle et al. 1999) place emphasis on forest biomass development. However, these field assessments and detailed remote sensing tools have a limited temporal frequency. The spatio-temporal quantification of ecosystem processes, therefore, remains a big challenge (Ostendorf et al. 2001).

In this article we investigate the hypothesis that the spatial distribution of colours of true-colour images (remote sensing data), despite topographic (stationary) vari-

[1] Department of Statistics, University of Innsbruck, Universitätsstr. 15, 6020 Innsbruck, Austria; email: janette.walde@uibk.ac.at

[2] European Academy Bozen/Bolzano, Scientific area of "Alpine Environment", Drususallee 1, 39100 Bozen, Italy; email: erich.tasser@eurac.edu

ables, explains biomass and leaf area distribution. As a first experiment Tasser et al. (2001b) used multiple linear regression analysis to find empirical evidence in support of a model for the spatial-temporal pattern of canopy structure with the help of the quoted inputs. The functional form of this relationship is not known a priori and furthermore cannot be assumed to be linear. Hence, artificial neural networks are used for modelling. With the help of tests for neglected nonlinearity we show that our model is correctly specified. Insignificant parameters are identified and consequently eliminated. We demonstrate that this procedure leads to an improvement of the model and to reliable results. As the gained relationship is non-linear, the effects of the input variables are not measurable simply by the corresponding regression coefficients as is the case in linear regression analysis. We provide both graphical and statistical measures to identify the form and the importance of influence factors on canopy development.

2. Materials and Methods

2.1 Study Area

The study area is in the upper Passeier Valley (South Tyrol, Italy) in the central eastern Alps. It is a narrow, v-shaped valley with fairly steep slopes. It extends over approximately 3 km² and ranges from 1200-2350m in altitude. The area of interest is located on the south exposed hill slope, comprising the alpine meadows of the farmers of the village Walten and the adjoining forest below. A great part of the meadows, especially on the steeper slopes, has been abandoned. Other parts have been managed to a greater degree. This has led to a high diversity of vegetation (Tasser et al. in press discovered 21 vegetation units in the region). From a scientific point of view, the differences in colour distribution and canopy phytomass, which are connected with the patchy vegetation and land-use distribution, are the most interesting aspect of this landscape.

2.2 Methods

Our modelling approach is based on 3 steps:

Field data collection: Structural development (photosynthetically active phytomass and total phytomass) of typical stands were analysed at nine times during the vegetation period (04/23/97, 05/16/97, 05/27/97, 06/15/97, 07/16/97, 07/30/97, 09/15/97, 09/29/97, and 10/29/97) along two altitudinal transects (1240 – 2200 m a.s.l.) with 21 permanently marked locations within the study (9 transect points and 1-4 different management types) that were permanently marked and registered with the GIS database. Altogether 167 observations were available. Each observation was accompanied by a rough phytosociological assessment of the vegetation and an estimation of the amount of cover deriving from grasses, herbs and dwarf shrubs. All important vegetation types were included in this inventory, so that a representative predicate about the canopy development of the entire area is possible.

Colour classification and topographic variables: In order to capture the spatial-temporal heterogeneity of stand development in the study area, photographs were taken from the opposing hill slope at the same time as the canopy development was analysed. As a capturing system a 35 mm single lens reflex camera, namely a Pentax A-3 with Takumar 28 – 80 mm objective, was used with Kodak Elite Chrome 100 colour slide films. To be able to take photographs automatically an electronic timer was added to the camera. The lens aperture and the exposure time were set to automatic to allow adjustment to different weather conditions. By means of the timer one photograph was taken at noon (12.00h CET). The horizontal distance from the camera position to the area of interest ranged between 1250 m and 2350 m. Nine photographs corresponding to the harvest dates were georectifed with the JUKE method (Aschenwald et al. 2001) and partitioned according to their RGB-colour values. Topographic parameters of the study site were derived from a digital terrain model with a 5m pixel resolution (Ostendorf et al. 1999).

Spatial extrapolation of leaf area: The RGB-colours as well as the horizontal distance from the camera position to the single pixels on the opposite hill slope, and land-use indices were used as independent variables in order to model phytomass measured on the site (Guisan et al. 2000). Three ordinal land-use indices were derived from land use types: hay utilization, pasture utilization, and a dummy variable for “cut” or “uncut” fields (for details see Tappeiner et al. 1998).

As an analysing tool, we used a fully connected, three-layered perceptron (MLP) with an additional linear connection from the input layer to the output layer and a single output unit corresponding to our dependent variable (photosyntetically active phytomass, total phytomass) Theoretical proofs show that standard multilayer feed-forward network architectures are a class of universal approximators. Especially in our scientific field, which lacks a-priori theoretical explanations for the appropriate functional form of the relationships between the variables, we expect that the method will lead to new knowledge about connections and influences of variables. The challenge using this highly flexible method is to handle the huge amount of degree of freedoms in an appropriate way (especially in comparison to our limited amount of data) and hence get reliable and interpretable results. We established our MLP with the help of statistical methods to make sure to detect real relationships and not faked ones. To test the hypotheses of misspecification and of the necessity for an additional hidden unit respectively, we used the RESET test (Ramsey 1969) and the test developed by Teräsvirta, Lin and Granger (TLG, 1993). Studies (Teräsvirta et al. 1993, Anders et al. 1996) show that the latter test is superior to many other comparable tests including White’s network test.

The whole data set was divided randomly into three disjoint samples: training set (T) to optimise the network, validation set (V) to control the error during learning process, and generalisation set (G) to estimate the quality of the model (|T|+|V|=100, |G|=67). Input and output variables were normalized (transformed to the interval]-1,1[and]0,1[respectively). With the help of the Levenberg-Marquardt learning algorithm the mean squared error was minimized. For optimising the network with respect to the number of hidden units we picked the network with the highest determination coefficient on the validation set. To supplement the training process, we tested wether or not an additional hidden unit is necessary by means of the TLG test.

3. Results

We hypothesize that spatio-temporal distribution of stand structure relates to colour pattern in the photographs. The spatial differences of stand structure showed a strong elevational gradient and varied strongly with land use. The data exhibited also significant temporal change in the structure of grass communities. Colour values, horizontal distance, and land-use variables were used to explain phytomass pattern. The variables of the colour distributions were highly multicollinear therefore we decided to use only one colour in our models. We used the linear regression as the benchmark method. The results show significant relationships between independent variables and photosynthetically active phytomass and total phytomass. From the literature we know that low red values indicate a high concentration of chlorophyll, hence, we were not surprised that we gained the best results using this colour as input for our linear model.

Previous descriptive analysis of residuals gave evidence that the assumption of a linear underlying relationship does not apply to our problem. We tested whether or not our data appears to be generated by a linear model against the alternative that they are nonlinearly related. Both tests clearly reject H_0 at a significance level of 1%. To apply the RESET test properly, collinearity had to be avoided by forming the principal components (RESET* in Table 1). This nonlinearity was investigated further using the neural network model. The perfomance of the optimised MLP was significantly higher than the one of the linear regression (Table 1). As the statistical tests indicate, no further nonlinearity in the data is expected. Contrary to the linear regression results, the nonlinear model shows a different behaviour with respect to the colour input. The MLP ranked 'blue' higher than 'red'. Chlorophyll absorbs red and blue spectra and reflects green. Consequently the variable 'red' contains similar information as the variable 'blue', nevertheless the latter was more suitable as input for the MLP. For our further analyses we used the model with the highest performance.

To find out the importance of each input variable we optimised the network in the same procedure as described above but without the input variable of interest. The decrease of the determination coefficient on the generalisation set defines the importance of the excluded input variable (Fig. 1).

Table 1: Performance of both models and significance for nonlinearity tests with respect to colour input.

Col-ourIn-put	Linear Regression						MLP						
	R^2 (%)			RESET p value	RESET* p value	TLG p value	R^2 (%)				RESET p value	RESET* p value	TLG p value
	overall	T+V	G	k=2	k=6, Q*=1		overall	T	V	G	k=2	k=6, Q*=1	
red	51.1	48.9	54.0	0.001	0.000	0.002	72.4	80.0	68.1	67.6	0.541	0.461	0.449
blue	50.4	49.2	52.1	0.001	0.000	0.001	75.8	84.5	72.4	69.6	0.614	0.741	0.706
green	51.0	49.1	52.0	0.005	0.000	0.001	68.9	72.2	69.8	65.8	0.003	0.006	0.309

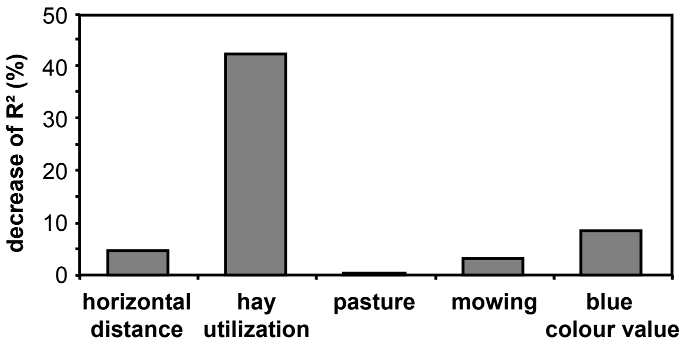

Fig. 1: Decrease of the determination coefficient (R^2) due to the missing input variable of interest. Hay utilization = quotient of all human impacts (mowing, fertilization, irrigation) and the frequency of these interferences in years; mowing = numbers of cuttings per year.

Model quality decreases especially if hay utilization is not taken into account. The number of human activities (mowing, fertilization, irrigation) per year thus has a crucial effect on phytomass distribution within the project area. Blue colour value already comes second. If this variable lacks, the coefficient decreases by about 10%. The variables „horizontal distance“ and „cutting“ (cutted yes/no) contain considerably less explaining information. According to these results, the fact whether an area is pastured or not has very little impact on phytomass distribution.

For the purpose of analysing the form of the dependence, we varied each metric variable from its minimum to its maximum in steps of hundredth of its range (ordinal variables were varied exactly according to their parameter values). All other variables were kept constant at their mean or median respectively. The change in the dependent variable due to the variation of one independent variable is shown in Fig. 2. The results show that a change in independent variables may have different, non-linear impact on the explanation of the corresponding phytomass. However, not the independent variables themselves lead to changes in phytomass, but rather the direct or indirect effects on vegetation and resource availability of plants in ecosystems under different land use.

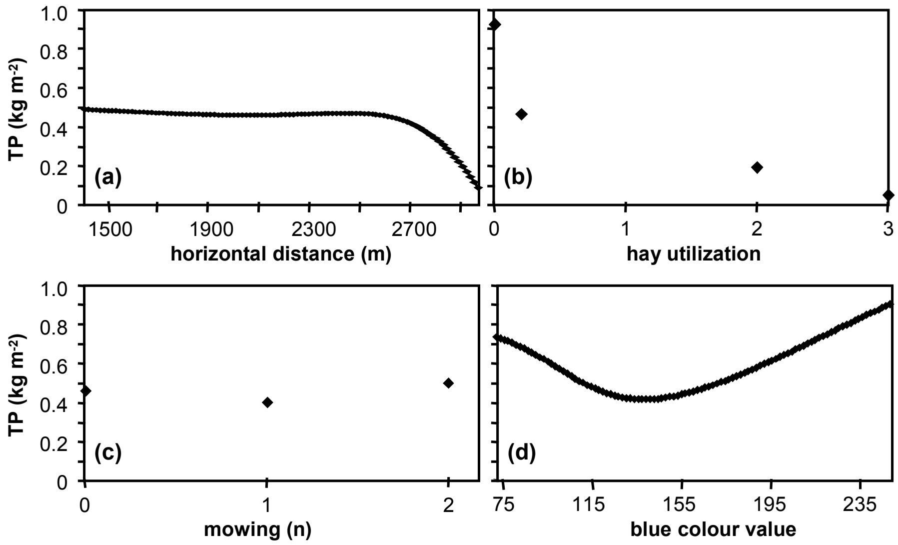

Fig. 2: The change in the total phytomass (TP) due to the variation of one independent variable.

4. Discussion

The determination coefficient decreases especially if hay utilization is not taken into account. The number of times the area is mown, fertilization and irrigation thus prove to be determinant factors for structure and composition of canopies (Tappeiner et al 1998). At the beginning of the vegetation period, hay meadows have very little above-ground phytomass, as plant matter is taken away once or twice a year by the farmer. As annual mean phytomass thus decreases. The more intensively an area is managed, i.e. the higher the „hay utilization“ value, the lower the phytomass according to the results of our study. This is mainly due to the shift of species initiated by nitrogen supply and increased mowing. New species appear while others are crowded out. The increase in nutrients deprives numerous plant species which only flourish on poor soils of the conditions they need to survive. Such species accordingly disappear, either because they cannot cope with the effects of the fertiliser or because they are suppressed as a result of stimulated growth in competitor species. Meadows that are mown regularly but receive no fertiliser become low-nutrient meadows. That process is accompanied by a reduction in the height of the plants and above-ground biomass. This is also shown by the relation between the "mowing" variable and phytomass. But, in case of continuing reduced cultivation (mowing every 1-3 years), the area becomes colonised by dwarf shrubs (Tasser et al. 1999, in press). With the increase of dwarf shrubs, i.e. also of lignified plant matter within the canopy, phytomass increases considerably. The highest phytomass values are thus found in abandoned areas, i.e. where the „hay utilization" value is 0. The major part of abandoned areas is covered by the dwarf-shrub rich variant of plant communities (Tasser et al. in press).

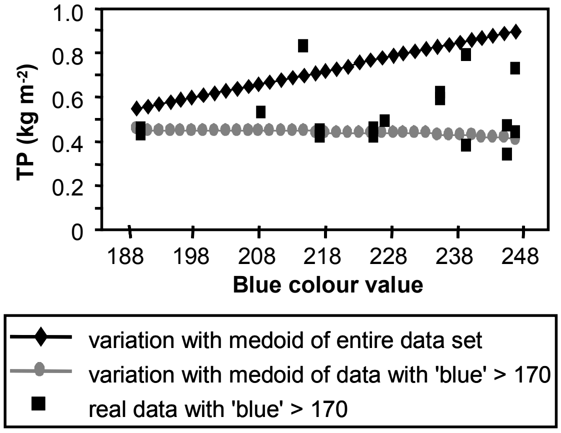

Fig. 3: Corrected functional form of the dependence of phytomass on 'blue'

Blue colour value can explain 10% of phytomass variability. Thus this variable, besides land use, is extremely important. Plant physiological response such as pigment, moisture, and growth factor changes has been found to produce measurable changes in spectral reflectance. As Tasser et al. (2002) could show, the blue colour value increases with increasing altitude if taken as one year average. Thus the decrease of phytomass is explained by the increase of blue colour value. The driving force behind it is the climate and the climate-related growth factors. Fig. 2 shows an increase in phytomass due to 'blue' which is theoretically not explainable. We investigated this part of the curve further by analysing the real data with these problematic blue values. It turned out that the medians of three out of four input variables were different from the one used in the analysis for Fig. 2. We set the medoid to its proper values and did the analysis again. The corresponding results are shown in Fig. 3.

The variable „horizontal distance“ contains considerably less explaining information. The relation is constant in large parts. Only at a distance of about 2640m or more there is an extreme decrease in phytomass. This is probably due to the site's topography. The most remote areas are the crest areas of the opposite slope. These crest areas have considerably worse climate conditions. Less precipitations, more wind, lower temperatures and long periods without snow in winter have an impact on plant growth (Körner 1999, Tappeiner et al. 1998). Thus phytomass decreases noticeably when approaching the crest.

5. Conclusion

We could show a new way to simulate biomass and leaf area distribution in a complex alpine terrain with limited amount of variables by combining temporal remote sensing (geo-registered and ortho-rectified colour images from the opposing hillslope) with classical field methods from plant physiology. The modelling approach was based on non-linear statistical models and on the use of Geographical Information System. Remarkable features of our models are simplicity, comprehensibility and clarity. The high degree of abstraction leads to the fact that the results can easily be generalised and are suitable for simulations. However, such models cannot lay any definite truth claim, as only selected key factors have been used for decision-making.

Acknowledgements: This work was funded by the 4th framework of the EU, project ECOMONT (ENV4-CT95-0179). Field studies were partly carried out within the framework of the INTERREG-II project INTEGRALP. The manpower of Janette F. Walde was financed by the Austrian Science Fund (FWF).

Bibliography

Anders, U., Korn, O. (1996): Model Selection in Neural Networks. ZEW Discussion Paper No. 96-21, Mannheim 1996

Aschenwald J., Leichter K., Tasser E., Tappeiner U. (2001): Spatio-Temporal Landscape Analysis in Mountainous Terrain by Means of Small Format Photography: a Methodological Approach. *IEEE Transactions on Geoscience and Remote Sensing* 39/4, 885-893

Guisan A., Zimmermann N.E. (2000): Predictive habitat distribution models in ecology. *Ecological Modelling* 135**,** 147–186

Körner C. (1999): Alpine plant life: functional plant ecology of high mountain ecosystems. Springer, Berlin.

Ostendorf B., Mayr V., Tappeiner U. (1999): GIS-contents and goals. In: *Land-Use Changes in European Mountain Ecosystems* (eds. by A. Cernusca, U. Tappeiner, N. Bayfield) Blackwell, Berlin, 180-187

Ostendorf, B., Hilbert, D.W., Köstner, B., Tappeiner, U., Tasser, E. (2001): The importance of understanding spatial pattern for scaling up plot-level matter and energy fluxes to regional scales. *SAMS* 41, 391-407

Peddle D.R., Hall F.G., LeDrew E.F. (1999): Spectral Mixture Analysis and Geometric-Optical Reflectance Modeling of Boreal Forest Biophysical Structure. *Remote Sensing of Environmental* 67**,** 288–297

Ramsey, J.B. (1969): Tests for Specification Errors in Classical Linear Leastsqares Regression Analysis. *Journal of the Royal Statistical Society*, Series B, 31, 350-371.

Shroder, J.F., Bishop J.M. (1995): Geobotanical assessment in the Great Plains, Rocky Mountains and Himalaya. *Geomorphology* 13, 101-119

Tappeiner U., Tasser E., Tappeiner G. (1998): Modelling vegetation patterns using natural and anthropogenic influence factors: preliminary experience with a GIS based model applied to an Alpine area. *Ecological Modelling* 113**,** 225-237

Tappeiner, U., Tasser, E., Mayr, V. & Ostendorf, B. (1999): Research area "Passeier Valley". In: *Land-Use Changes in European Mountain Ecosystems* (eds. by A. Cernusca, U. Tappeiner, N. Bayfield) Blackwell, Berlin, 48-61

Tasser E., Aschenwald J.F., Tappeiner U. (2001a): Geostatistical approaches for landscape modelling with regard to the effects of site factors and land-use changes. Submitted to *Landscape modelling*

Tasser, E., Prock, S., Mulser, J. (1999): Impact of land-use on the vegetation along an Eastern Alpine transect. In: *Land-Use Changes in European Mountain Ecosystems* (eds. by A. Cernusca, U. Tappeiner, N. Bayfield) Blackwell, Berlin, 235-246.

Tasser, E., Tappeiner, U. (in press): The impact of land-use changes in time and space on vegetation distribution in mountain areas. *Applied Vegetation Science*.

Tasser, E., Walde, J., Ostendorf, B., Schmid, P., Tappeiner, U. (2002) An interdisciplinary approach for evaluation of temporal-spatial heterogeneity of colour distribution and canopy development on an Alpine test site. In: Interdisciplinary Mountain Research (eds. by R. Bottarin, U. Tappeiner) Blackwell, Berlin

Tasser, E., Walde, J.F., Ostendorf, B., Schmid, P., Tappeiner, U. (2001b): An interdisciplinary approach for evaluation of temporal-spatial heterogeneity of colour distribution and canopy development on an Alpine test site. Interdisciplinary Mountain Research (eds. R. Bottarin, U. Tappeiner, F. Ruffini), in press.

Teräsvirta, T., Lin, C.F., Granger, C.W.J. (1993): Power of the Neural Network Linearity Test. *Journal of Time Serie Analysis*, 14, 209-220