

Valid Statistical Inference Based on Feedforward Artificial Neural Network Models

Janette F. Walde¹

Abstract

With the help of recent theoretical results, we use the estimates from neural network modelling as basis for formal statistical inference. Multi-layer perceptrons are applied to model biomass in a complex alpine terrain with limited amount of variables by combining temporal remote sensing with classical field methods from plant physiology. We test the hypothesis that the dynamics of the biomass distribution can be captured with the help of geo-registered and ortho-rectified colour images from the opposing hill slope. Therefore the network model is trained carefully and misspecification is tested by the non-linearity tests of Ramsey and of Teräsvirta, Lin and Granger. Plausibility and sensitivity analysis as well as ecological considerations in respect of content support the validity of our final model. With the help of bootstrap techniques the significance of colour patterns for modelling phytomass is demonstrated.

1 Introduction

Multi-layer perceptrons (MLP) are a widely used class for non-linear modelling. Their greatest advantage is that a-priori knowledge of the specific functional form is not required. Most applications of feedforward MLP have been concerned with the estimation of relationships between input and target variables of interest and the superiority of the performance of this approach in comparison to more classical methods, but they are not only a ‘black-box’ tool. In fact, they have the potential to significantly enhance scientific understanding of empirical phenomena subject to neural network modelling. In particular, the estimates obtained from network learning can serve as a basis for formal statistical inference. Statistical tests of specific scientific hypothesis of interest become possible. Because of the ability of MLP to extract complex nonlinear and interactive effects, the alternatives against which such tests can have power may extend usefully beyond those within reach of more traditional methods, like linear regression analysis.

¹ Department of Statistics, University of Innsbruck, Universitätsstr. 15, 6020 Innsbruck, Austria.

The paper of White and Racine (2001) put a promising theoretical foundation for valid inferences regarding input variables with the help of bootstrap techniques. We apply their theoretical results to the problem of modelling canopy development. The hypothesis that spatial distribution of colours of true-colour imageries (remote sensing data) (RGB-colours), despite topographic (stationary) variables (horizontal distance, cover of dwarf shrubs, and land-use variables), explains biomass distribution is tested in this article. As a first experiment we used multiple linear regression analysis and gave empirical evidence for the possibility to model spatial-temporal pattern of canopy structure with the help of the quoted inputs (Walde et al., 2002). Tests for neglected non-linearity identify the correctly specified model. As the obtained relationships are non-linear, the importance of the input variables is not simple measurable as the corresponding coefficient like for example in linear regression analysis, but both graphical and statistical measures identify the importance of influence factors on canopy development.

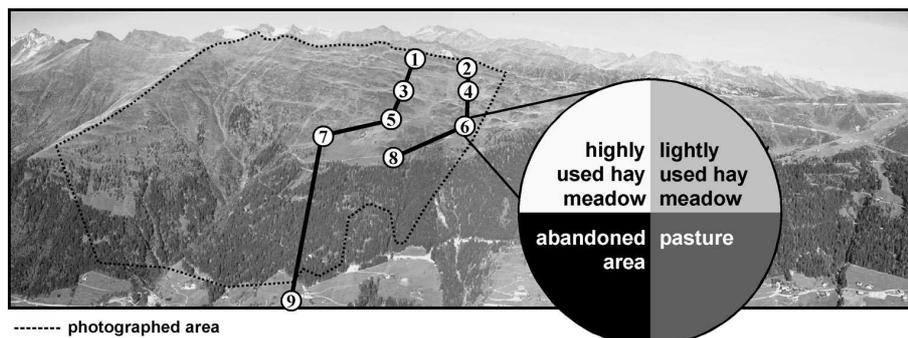


Figure 1: Along two transects from the valley till the ridge data was collected from biomass. At each point (1-9) the canopy development was ascertained to each of the most important management types.

2 Material and modelling approach

2.1 Study area

The study area is situated 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 1200-2350 m 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 steeper slopes, has been abandoned,

whereas other parts have been managed to a greater degree. This has led to a high diversity of vegetation (on the study area 21 vegetation units have been determined, Tasser et al. 2001). From the 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 Database

Biomass (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 land-use types, cf. Figure 1) that were registered with the GIS database. These investigations were accompanied by rough phytosociological assessment of the vegetation and estimation of the amount of cover deriving from dwarf shrubs. All important vegetation types were included in this inventory.

Not at each location every management type was available, and observation units with missing input values were eliminated. Hence, altogether 167 observation units were available.

The temporal heterogeneity of stand development in the study area was captured by photographs. They 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 georectified with the JUKE method (Aschenwald et al. 2001) and partitioned according to their RGB-colour values ('red', 'green', 'blue').

Topographic parameters of the study site (e.g. horizontal distance) were derived from a digital terrain model with a 5 m pixel resolution (Ostendorf et al. 1999).

2.3 Multi-layer perceptron

The RGB-colours as well as the horizontal distance from the camera position to the single pixels on the opposite hill slope, cover of dwarf shrubs and land-use indices were used as independent variables in order to model phytomass measured

on the site (Guisan et al., 2000). Three land-use indices were derived from land use types: hay utilization, pasture utilization, and mowing. ‘Hay utilization’ is defined as the quotient of all human impacts (mowing, fertilization, irrigation) and the frequency of these interferences in years, ‘pasture utilization’ is a binary variable, and ‘mowing’ is the number of cuttings per year (for details see Tappeiner et al., 1998).

As 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 (total phytomass) (1).

$$\hat{y} = f(\mathbf{x}, \hat{\theta}) = \sum_{i=0}^I w_i x_i + \sum_{j=1}^{HU} w_j^{(2)} \cdot G\left(\sum_{k=0}^I w_{kj}^{(1)} x_k\right) \quad (1)$$

$$G(x) = \frac{1}{1 + e^{-x}},$$

where I is the number of input variables, HU is the number of hidden units, $w_{..}^{(.)}$ are the weights, \mathbf{x} is the input vector, $\hat{y} = f(\mathbf{x}, \hat{\theta})$ is the network output.

Theoretical proofs show that standard multilayer feedforward network architectures are a class of universal approximators (Hornik et al. 1989). Especially in our scientific field with lack of a-priori theoretical explanations for functional form of relationships, we expect new knowledgegain 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 and cross-validation procedures to make sure to detect real relationships and not faked ones.

To test the hypotheses of misspecification, we used the RESET test and the test developed by Teräsvirta, Lin and Granger - TLG (1993). The RESET test proposed by Ramsey (1969) postulates an alternative model of the following form:

$$y_t = \hat{y}_t + a_2 \hat{y}_t^2 + a_3 \hat{y}_t^3 + \dots + a_k \hat{y}_t^k + v_t \quad \text{for some } k \geq 2, \quad (2)$$

where y_t is the observed value of the dependent variable for unit t , \hat{y}_t is the model output for unit t , a_i are parameters and v_t are the residuals.

The null hypothesis is $H_0 : a_2 = a_3 = \dots = a_k = 0$. Denoting $\hat{\mathbf{e}}$ the vector of estimated residuals from the model under the null hypothesis and $\hat{\mathbf{v}}$ the vector of estimated residuals under the alternative model, the test statistic is

$$RESET = \frac{(\hat{\mathbf{e}}^T \hat{\mathbf{e}} - \hat{\mathbf{v}}^T \hat{\mathbf{v}}) / (k - 1)}{\hat{\mathbf{v}}^T \hat{\mathbf{v}} / (n - k)}, \quad (3)$$

which is approximately $F(k-1, n-k)$ when H_0 is true (n is the sample size).

As a further checking, the test of Teräsvirta, Lin and Granger (TLG) (1993) was employed. Studies (Teräsvirta et al., 1993; Anders et al., 1996) show that this test is superior to many other existing tests including White's network test (White, 1989). The version of the Gauss-Newton regression is given in the following equation:

$$\hat{\epsilon}_t = \boldsymbol{\alpha}^T \nabla f(\mathbf{x}_t, \hat{\boldsymbol{\theta}}) + \sum_i \sum_j \sum_k \beta_{ijk} x_{it} x_{jt} x_{kt} + v_t, \quad (4)$$

where $\boldsymbol{\alpha}$ is a parameter vector, ∇ is the vector of partial derivatives with respect to the weights of the model ($\hat{\boldsymbol{\theta}}$), β_{ijk} are parameters, and v_t is the residual for unit t .

The test statistic is the uncentered determination coefficient of this regression times the sample size (nR_u^2) which is asymptotically χ^2 -distributed. Theoretical arguments and empirical studies (Davidson 1993) demonstrate that the F statistic (5) has better finite-sample properties than the nR_u^2 statistic based on the same Gauss-Newton regression. Hence we used the F statistic too. $USSR$ and $RSSR$ are the unrestricted and restricted sum of squared residuals, r is the number of restrictions and k the number of parameters:

$$F = [(RSSR - USSR)/r] / [USSR/(n - k)]. \quad (5)$$

To apply these two tests properly, colinearity had to be avoided by forming the principal components of $\hat{y}_t^2, \dots, \hat{y}_t^k$ ($x_{it} x_{jt} x_{kt}$), choosing the Q^* largest principal components (except the first principal component so as not to be collinear with \mathbf{x}_t), and then regressing y_t on these and \mathbf{x}_t . The resulting test statistics $RESET^*$ and F^* are $F(Q^*, n-k)$ when H_0 is true (Lee et al. 1993).

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 ($|T|=70$, $|V|=30$, $|G|=67$). Input and output variables were normalized (transformed to the interval $[-1,1]$ and $[0,1]$ respectively).

With the help of Levenberg-Marquardt learning algorithm (Bishop, 1996) the mean squared error was minimized. For optimising the network with respect to the number of hidden units we picked the network with the smallest mean squared error on the validation set.

The final model regarding the input variables was chosen with the aid of the Akaike information criterion (Akaike, 1973, 1974), the Schwartz information criterion (Schwartz, 1978) and the penalized prediction error (Bishop, 1996 (6)).

$$PE = 2 \cdot MSE + \frac{2P}{n} \cdot \sigma^2, \quad (6)$$

where MSE is the value of the mean sum-of-squares error, P is the number of parameters, n is the sample size, and σ^2 is the variance of the noise on the data (which must be estimated). Further, the mean absolute error and the mean absolute percentage error were analysed for the suitability of the model.

2.4 Significance of input variables

With the help of the optimised network, we wanted to test whether or not the colour inputs are significant. White and Racine (2001) investigated the statistic and its distribution for the null hypothesis that some input variables are not significant. In the following the most important results for this paper are summarized. The null hypothesis is framed in the following way:

$$H_0: f_1(x, w) = f_2(x, w), \quad (7)$$

where f_1 is the network which is optimised with all inputs, and f_2 is the MLP optimised with input variables except the one of interest, in our case without the colour inputs.

As statistic they used:

$$m = \int (f_1(x, w_1^*) - f_2(x, w_2^*))^2 d\mu, \quad (8)$$

where w^* are the optimal weights and μ is the probability distribution of X .

The optimal values are replaced with consistent estimators to make the statistic feasible. The resulting statistic is:

$$\hat{m}_n = \frac{1}{n} \sum_{i=1}^n (f_1(x_i, \hat{w}_1) - f_2(x_i, \hat{w}_2))^2. \quad (9)$$

Theorem 2.1 of their paper states that under suitable conditions the statistic is distributed like:

$$n \cdot \hat{m}_n \rightarrow N_2(0, C^*; M^*), \quad (10)$$

where N_2 is the distribution of n independent χ^2 -distributed variables (White 1994, Lemma 8.2). Fortunately with the help of bootstrap techniques, we can

approximate this distribution and hence obtain the rejection region of the null hypothesis.

The corresponding bootstrap statistics are:

$$\hat{A}_n^\# \equiv \sum_{t=1}^n m(X_t^\#, \hat{w}_n^\#) - \sum_{t=1}^n m(X_t^\#, \hat{w}_n) - \sum_{t=1}^n \nabla^T m(X_t^\#, \hat{w}_n) (\hat{w}_n^\# - \hat{w}_n)$$

or

(11)

$$\bar{A}_n^\# \equiv \sum_{t=1}^n m(X_t, \hat{w}_n^\#) - \sum_{t=1}^n m(X_t, \hat{w}_n) - \sum_{t=1}^n \nabla^T m(X_t, \hat{w}_n) (\hat{w}_n^\# - \hat{w}_n),$$

where $\hat{w}_n \equiv (\hat{w}_1^T, \hat{w}_2^T)^T$, $X_t^\#$ are the ‘new’ inputs drawn randomly with replacement from the sample of the original ones, $\hat{w}_n^\#$ are the corresponding weights of the optimised networks by means of the bootstrap sample.

They proved (Theorem 2.3) that $\hat{A}_n^\# \xrightarrow{d} N_2(0, C^*; M^*)$ and $\bar{A}_n^\# \xrightarrow{d} N_2(0, C^*; M^*)$.

The procedure is realized in the following steps:

1. Using original sample, solve for f_1 and f_2 to get weights
2. Draw a sample with replacement from the original sample and compute re-sampled weights
3. Compute bootstrap statistics: $\hat{A}_n^\#, \bar{A}_n^\#$
4. Replicate steps 2) and 3) 100 times
5. Compute a one sided $(1-\alpha)\%$ acceptance region (c_α)
6. Reject H_0 if the original test statistic ($n \cdot \hat{m}_n$) exceeds c_α .

3 Results

Our main hypothesis is that spatial-temporal distribution of total phytomass relates to colour patterns in the photographs. By means of descriptive analysis, the spatial differences of total phytomass show a strong elevational gradient and vary strongly with land use (Table 1). The results show also temporal change in the structure of grass communities (cp. last two columns of Table 1).

Colour values, horizontal distance (also a measure of altitude), cover of dwarf shrubs and land-use variables were used to model and consequently to explain phytomass pattern. As benchmark method we used the multiple linear regression analysis. The results demonstrate the existence of relationships between the used independent variables and total phytomass regarding the determination coefficient (cf. Table 2).

Table 1: Measured total phytomass under different land-use along an elevational gradient (for the position in the transects see Figure 1). n = number of measurements in course of the vegetation period, TP = total phytomass, s.d. = standard deviation.

	transect points	Mean altitude (m a.s.l.)	n	TP (kg m ⁻²)	
				\bar{x}	s.d.
Abandoned area	1 + 2	2098	16	0.60	0.52
	3 + 4	1998	16	0.82	0.72
	5 + 6	1834	16	0.91	1.12
	7 + 8	1742	16	1.08	1.24
Pasture	2	2088	9	0.46	0.30
	4	2007	9	0.48	0.36
	7	1773	9	0.61	0.33
Lightly used hay meadow	3 + 4	1998	18	0.31	0.76
	5 + 6	1833	17	0.44	0.62
	7 + 8	1740	17	0.44	0.91
Intensively used hay meadow	4	2007	9	0.16	0.27
	5 + 6	1835	17	0.23	0.66
	7 + 8	1748	18	0.31	1.03
	9	1214	9	0.30	0.75

Table 2: The performances of the multiple linear regression analysis and of the MLP with respect to different colour inputs are shown. Sig. is the significance of the non-linearity tests (RESET, TLG), R² is the determination coefficient. Beside the colour inputs, horizontal distance, cover of dwarf shrubs, mowing, hay utilization, and pasture were used as independent variables.

Colour Inputs	Multiple Linear Regression				MLP						
	R ² (%)	RESET	TLG	Sig.	R ² (%)			RESET	TLG	Sig.	
	overall	k=2	k=6, Q [*] =1	Sig.	overall	T	V	G	k=2	k=6, Q [*] =1	Sig.
red	70.02	0.074	0.037	0.045	78.22	79.18	76.58	78.03	0.981	0.216	0.182
green	69.64	0.146	0.042	0.000	80.39	82.55	76.45	77.92	0.310	0.101	0.221
blue	69.59	0.150	0.043	0.000	78.95	85.71	74.02	79.70	0.001	0.058	0.146
red, green	69.87	0.054	0.016	0.002	77.25	79.69	75.48	76.61	0.059	0.453	0.361
red, blue	69.81	0.096	0.030	0.002	79.72	81.32	78.80	79.16	0.858	0.207	0.410
<i>green, blue</i>	69.60	0.155	0.044	0.005	84.77	92.66	82.07	81.60	0.171	0.313	0.630

With the help of both statistical tests (RESET and Teräsvirta, Lin and Granger test), the null-hypothesis of a well specified model had to be rejected regarding the linear model. Concerning the MLP, the results of the non-linearity tests depend on the used inputs. The input variable ‘blue’, as sole colour input, is not sufficient to capture the non-linearity in the data. Interestingly, this model as well as the approach with ‘red’ and ‘green’ as colour inputs required the highest number of hidden units. Given the data limitation we should employ parsimonious models. The models with the colour input ‘red’ and the model with the inputs ‘green’ and ‘blue’ are the most parsimonious models with respect to the number of parameters (two and three hidden units respectively). The two non-linearity tests agree that we can accept the null hypothesis of no misspecification for these models.

The Akaike information criterion as well as the penalized prediction error are smaller for the model with the inputs ‘green’ and ‘blue’. The Schwartz information criterion favour a little bit more the MLP with just the colour ‘red’ as input. Furthermore, the smaller mean absolute error and the smaller mean absolute percentage error indicate the suitability of this model (cf. Table 3). Consequently, we chose the approach with the colour inputs ‘green’ and ‘blue’ for the following analyses, although the big difference between the determination coefficient in the training set and the determination coefficient in the generalization set may be a sign of overfitting (Table 2).

Table 3: With the help of the used criteria, the MLP with the colour inputs ‘green’ and ‘blue’ was chosen as appropriate model.

<i>Criteria</i>	<i>Colour Input</i>	
	<i>‘red’</i>	<i>‘green’ and ‘blue’</i>
Akaike information criterion	0.385	0.222
Schwartz information criterion	0.814	0.875
Penalized prediction error	0.01767	0.01294
Mean absolute error	0.0725	0.0608
Mean absolute percentage error	1.777	1.586

3.1 Plausibility analysis

In order to find the importance of each input variable, we optimised the network again but without the input variable of interest. The decrease of the determination coefficient in the generalisation set defines the importance of the excluded input variable.

The explanation by the model decreases especially if ‘mowing’ is not taken into account (cf. Figure 2). The number of human activities per year thus has a crucial effect on phytomass distribution within the project area. Cover of dwarf shrubs already comes second. If this variable lacks, the coefficient decreases by about 6%. The time-dependency, captured by the colour inputs ‘green’ and ‘blue’, is the third important relationship. The variables ‘horizontal distance’ and ‘hay utilization’ contain considerably less explaining information. Accordingly to these results, the fact whether an area is pastured or not has very little impact on phytomass distribution.

We investigated further the plausibility of the functional relationship between phytomass and the independent variables. For this purpose, 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. Apart from scanning the variable of interest, all other variables were kept constant at their median. We did not have many observations units with values greater than 40

regarding the variable cover of dwarf shrubs. Therefore we used all original input values of these cases to calculate the output value.

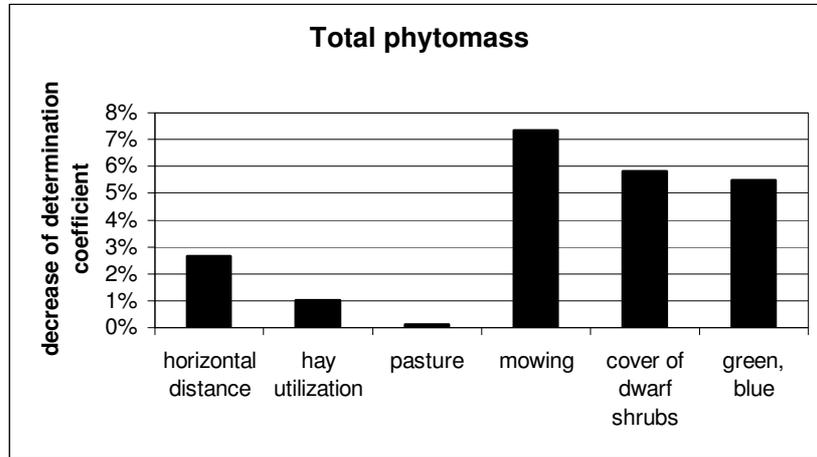


Figure 2: Decrease of the determination coefficient (R^2) due to the missing input variable of interest for the dependent variable total phytomass. Hay utilization = quotient of all human impacts (mowing, fertilization, irrigation) and the frequency of these interferences in years; mowing = numbers of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

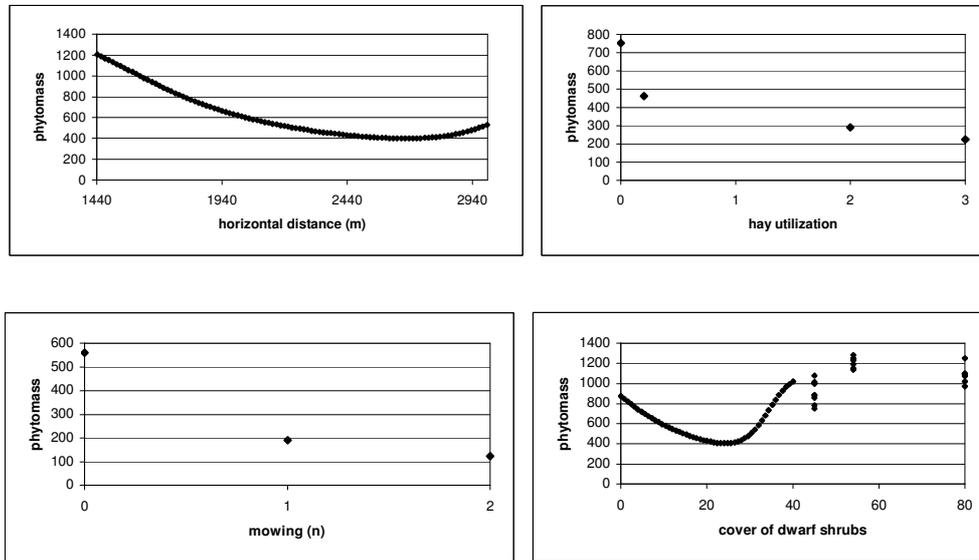


Figure 3: The change in the total phytomass due to the variation of one independent variable using the MLP.

The change in the dependent variable due to the variation of one independent variable is shown in Figure 3. Changes in independent variables have different, non-linear impacts on 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.

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. 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).

The empirical results can be supported with theoretical considerations and consequently are plausible and point to a carefully trained, valid non-linear model.

Finally, we want to test whether or not the colour inputs have a statistically significant influence.

3.1 Statistical inference with respect to colour inputs

We calculated the bootstrap statistics both by means of the data sets in the generalization set and with the entire sample. Limited to the generalization set, we obtain significance of the colour inputs (‘green’ and ‘blue’) at a 5%-level. In contrast, we can reject H_0 only at an approximately 10%-level using all data sets (Table 4).

White and Racine (2001) did not use our distinction of training set, validation set, and generalization set in their derivations. Due to the severe problem of overfitting especially with the available limited amount of data, we suspect that the separation of the data in these disjoint sets is necessary to get reliable results and furthermore a better power and an accurate level of the proposed test. For the future work we want to show this hypothesis.

Table 4: Empirical significance of the colour inputs calculated with the help of the bootstrap approach.

	<i>Empirical significance</i>	
	G	T+V+G
$\hat{A}_n^\#$	0.044	0.108
$\overline{A}_n^\#$	0.044	0.102

However, even the results do not agree completely, they support the hypothesis that spatial-temporal distribution of stand structure relates significantly to colour pattern in the photographs.

4 Conclusion

We modelled biomass 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 hill slope) with classical field methods from plant physiology. The non-linear approach was based on multi-layer perceptrons. Faced with the limited amount of data, we trained the network model carefully and tested misspecification. Additionally, plausibility analyses and ecological considerations in respect of content support the validity of our final model. With the help of sensitivity analysis, and bootstrap techniques the significance of colour patterns for modelling phytomass was demonstrated. We showed that multi-layer perceptrons can be used for enhancing scientific understanding of relationships.

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 Walde was financed by the Austrian Science Fund (FWF).

I am indebted to Dr. Erich Tasser (European Academy Bozen/Bolzano, Scientific area of "Alpine Environment", Bozen, Italy) who analysed the results with respect to their ecological plausibility.

References

- [1] Akaike, H. (1973): Information theory and an extension of the maximum likelihood principle. In B.N. Petrov and F. Csaki (Eds.): *Second International Symposium on Information Theory*. Akademiai Kiado, 267-281.
- [2] Akaike, H. (1974): A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **AC-19**, 716-723.
- [3] Anders, U. and Korn, O. (1996): Model selection in neural networks. *ZEW Discussion Paper No. 96-21*, Mannheim 1996.
- [4] Aschenwald, J., Leichter, K., Tasser, E., and 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**, 885-893.
- [5] Bishop, C.M. (1996): *Neural Networks for Pattern Recognition*. New York: Oxford University Press.
- [6] Davidson, R. and MacKinnon, J.G. (1993): *Estimation and Interference in Econometrics*. New York: Oxford University Press.
- [7] Guisan, A. and Zimmermann, N.E. (2000): Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147-186.
- [8] Hornik, K., Stinchcombe, M., and White, H. (1989): Multilayer feedforward networks are universal approximators. *Neural Networks*, **2**, 359-366.
- [9] Lee, T.H., White, H., and Granger, C.W.J. (1993): Testing for neglected nonlinearity in time series models. A comparison of neural network methods and alternative tests. *Journal of Econometrics*, **56**, 269-290.
- [10] Ostendorf, B., Mayr, V., and Tappeiner, U. (1999): GIS-contents and goals. In A. Cernusca, U. Tappeiner, and N. Bayfield (Eds.): *Land-Use Changes in European Mountain Ecosystems*. Berlin: Blackwell, 180-187.
- [11] Ramsey, J.B. (1969): Tests for specification errors in classical linear leastsquares regression analysis. *Journal of the Royal Statistical Society, Series B*, **31**, 350-371.
- [12] Schwarz, G. (1978): Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461-464.
- [13] Tappeiner, U., Tasser, E., and 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.
- [14] Tasser, E., Aschenwald, J., and Tappeiner, U. (2001): Geostatistical approaches for landscape modelling with regard to the effects of site factors and land-use changes. Submitted to *Landscape modelling*.

-
- [15] Tasser, E. and Tappeiner, U. (in press): The impact of land-use changes in time and space on vegetation distribution in mountain areas. *Applied Vegetation Science*.
- [16] Teräsvirta, T., Lin, C.F., and Granger, C.W.J. (1993): Power of the neural network linearity test. *Journal of Time Serie Analysis*, **14**, 209-220.
- [17] Walde, J. and Tasser, E. (2002): Environmental communication in the information society. In W. Pillmann and K. Tochtermann (Eds.): Identifying significant determinants for canopy development on an Alpine test site by means of artificial neural networks. International Society for Environmental Protection, Vienna, 86-95.
- [18] White, H. (1989): An additional hidden unit test for neglected nonlinearity in multilayer feedforward networks. *Proceedings of the International Joint Conference on Neural Networks*,. **2**, New York, 451-455.
- [19] White, H. (1994): *Estimation, Inference and Specification Analysis*. New York : Cambridge University Press.
- [20] White, H. and Racine, J. (2001): Statistical inference, the bootstrap, and neural network modelling with application to foreign exchange rates. *IEEE Transactions on Neural Networks: Special Issue on Neural Networks in Financial Engineering*, **12**, 657-673.