Validation of simulated climate-adjusted forest growth by means of tree ring measurements

Thomas Zumbrunn and Christian Körner

Institute of Botany
Section Plant Ecology
University of Basel
Schönbeinstrasse 6
CH-4056 Basel

Final version for the Swiss Federal Office for the Environment (FOEN)

Contents

1. Summary

Due to its commitments to limitations and reductions of greenhouse gas emissions, Switzerland has to provide annual reports of its carbon stocks in forests to the United Nations (UN). The reported values are based on empirical 10-year mean values of large-scale biomass estimates by the Swiss National Forest Inventory (NFI), but the annual deviations from these mean values are estimated by means of the mechanistic ecosystem simulation model BIOME-BGC that simulates biogeochemical and hydrologic processes assuming these are a function of climate and general life-form characteristics. It was the aim of this study (1) to validate these simulation-model-based estimates of tree growth by a large sample of tree ring width measurements for Norway spruce (*Picea abies*) collected immediately after the storm Lothar in winter 2000 and (2) to suggest alternative ways of estimating the annual deviations from the empirical 10-year mean values.

The measured tree ring widths for 1650 trees from 104 locations did not correlate with the predictions of the mechanistic simulation model BIOME-BGC. Provided that tree ring width is a suitable proxy of annual tree biomass increment, the conclusion is that the calculation method for the annual greenhouse gas reporting to the UN based on the mechanistic model BIOME-BGC should be abandoned. The calculation method, as currently applied by the Swiss Federal Office for the Environment (FOEN), additionally implements a statistical modelling approach. However, the statistical models are solely used to bypass the use of BIOME-BGC for the prediction of annual biomass deviations since the latter was considered to be to difficult to handle in everyday use.

As an alternative to the currently applied calculation method, we suggest to discard the mechanistic modelling approach and to continue using a statistical modelling approach, the later being changed in two important ways. First, the response variables should be empirically obtained, i.e. measured biomass data (or proxies thereof) instead of biomass data obtained via mechanistic simulation models such as BIOME-BGC. The reason is that BIOME-BGC does not gain biomass data independently from explanatory variables and thus misleadingly overestimates the amount of explained variance by basically expressing autocorrelation. Estimates of actual biomass increments could be

obtained from the planned annual surveys of the NFI or an annual sampling specifically designed for the purpose of the greenhouse gas reporting. We replaced the simulated BIOME-BGC based biomass data with our measured tree ring widths as the response variable in the statistical models and used the same set of explanatory variables as applied in the statistical models of the FOEN. It was found that, although only a relatively small part of the variance in tree ring width is explained, the estimated prediction error is at maximum c. 4 % (average departure of predicted values from real values). Therefore, and considering the missing correspondence between the biomass data simulated by BIOME-BGC and our measurements, such models could be used as a replacement for the currently applied ones. Second, we suggest to extend the set of explanatory variables by characteristics of forest water balance. We applied a simple hydrologic model that simulates the amount of available water in forest soils based on climate data and estimates of forest evapotranspiration and soil characteristics. Simple drought criteria based on the simulated soil water availability were used as additional explanatory variables in the statistical models. In most cases, these variables did not significantly improve our models, but on the other hand, there were cases where the forest water balance based drought criteria were the most significant of all predictors. Work on drought criteria based on the hydrologic model will therefore be continued in order to be able to develop improved prediction models.

2. Introduction

2.1 Switzerland's greenhouse gas inventory and carbon fluxes in forests

With the ratification of the United Nations Framework Convention on Climate Change (UNFCCC, `http://unfccc.int/`) on 10 December 1993 and more specifically with the ratification of the Kyoto Protocol (KP, an addition to the UNFCCC treaty) on 9 August 2003, Switzerland entered greenhouse gas (GHG) emission limitation and reduction commitments. The signees of the UNFCCC and the KP have to submit an annual report on their GHG inventory in the form of the so-called National Inventory Report (NIR) which serves to control the countries' compliance with their commitments. Prior to signing the Kyoto protocol (KP) on 16 March 1998, Switzerland decided to take its forests as a potential source or sink of carbon into account (article 3.4 of the protocol):

> [...] human-induced activities related to changes in greenhouse gas emissions by sources and removals by sinks in the agricultural soils and the land-use change and forestry categories shall be added to, or subtracted from, the assigned amounts [...]

Therefore, forest management activities and their impact on forests' carbon stocks are part of Switzerland's GHG inventory and listed in the NIR (Heldstab et al. 2009).

2.2 Current calculation method for annual carbon fluxes in forests

In order to assess the carbon budget of forests on an annual basis, the annual fluxes of CO_2 have to be recorded. The difference of the yearly gain and loss of biomass in terms of wood is translated into an equivalent amount of CO_2 and is taken as an estimate of the CO_2 balance of Swiss forests (Thürig & Schmid 2008). Whereas data on the yearly

biomass loss through forestry activities are available (Pasi et al. 2009), reliable empirical estimates on biomass gain through tree growth are based on the Swiss National Forest Inventory (NFI, see Brassel & Brändli 1999) which has so far been conducted only once per decade and is only now being reorganised to implement an annual sampling scheme. Because annual rather than 10-year average based biomass estimates were needed, Thürig & Schmid (2008) developed a calculation method for the Swiss GHG inventory which allowed them to predict annual (positive and negative) deviations from the empirical 10-year mean values. They employed the mechanistic ecosystem simulation model BIOME-BGC that simulates biogeochemical and hydrologic processes assuming these are a function of climate and general life-form characteristics (as described in Thornton et al. 2002). First, they used BIOME-BGC to simulate the carbon pools for the five production regions of the NFI, and they did so separately for three altitudinal ranges, for deciduous and coniferous forests, furthermore handling the Eastern and Western Alps separately for the two lower altitudinal range. These combinations of NFI production region, altitudinal range, and forest type were termed (and will be referred to as) *strata* by Thürig & Schmid (2008). Then, they determined the size of the simulated carbon pool for each year and used these values to get estimates of yearly forest biomass gain (assuming specific carbon contents of wood). These annual values were afterwards standardised by dividing them by their 10-year mean. The resulting dimensionless values were termed *climate correction factors* (Thürig & Schmid 2008) and were suggested to be interpreted as the relative annual deviation of forest biomass from the long-term mean in a specific stratum. The idea was that an empirical 10-year biomass estimate from the NFI could be multiplied by the simulated annual climate correction factor of the corresponding production region, altitudinal range, and forest type to obtain an annual biomass value that could be reported in the GHG inventory.

Ideally, the simulated biomass values of BIOME-BGC would have been used directly for the calculation of the climate correction factors of forest growth. However, to simplify everyday application and since the handling of BIOME-BGC seems to be rather laborious, Thürig & Schmid (2008) used the biomass data simulated by BIOME-BGC in yet another (but now statistical) modelling step as the response variable in multiple linear regression models for each stratum. They used summary statistics (such as annual sum of precipitation, mean temperature during growing period, etc.) of the same climatic data that were already used for the BIOME-BGC simulations as explanatory variables in the multiple linear regression models. Subsequently, the regression models were simplified by removing non-significant predictors by an undescribed procedure (see Thürig et al. 2006,

a report with more technical details on which Thürig & Schmid 2008 is based and which we cite interchangeably). To report annual forest gross growth in the NIR (Heldstab et al. 2009), these statistical models are used to predict climate correction factors which are then in turn used to "correct" the NFI 10-year mean values of gross growth.

2.3 Validation of current calculation method

Ultimately, i.e. over periods of 10 years, the reported gross growth is based on the NFI estimates and should thus have a sound empirical basis. However, current estimation of annual deviations is based on the mechanistic simulation model BIOME-BGC and thus primarily on climatic data (Thürig & Schmid 2008). Moreover, the auxiliary statistical modelling approach of the calculation method for the GHG inventory suffers from circular reasoning because the predictive accuracy of the multiple regression equations was assessed by comparison with the output of the (mechanistic simulation) model BIOME-BGC which in turn is used as the response variable in the (statistical) multiple regression models (Thürig & Schmid 2008). Comparison of multiple regression models with regard to their predictive accuracy should, however, be based on response variables independently obtained from the explanatory variables — which was clearly not the case.

Candidates for independently derived response variables are empirical tree growth measurements. These measurements serve as proxies for actual biomass increments. For a study on the association between wood properties and windthrow damage, a large sample of wood cores has been collected (Meyer et al. 2008). In the study at hand, the annual ring widths of these wood cores serve as a measure for tree growth and were planned to be used to validate the output of the mechanistic simulation model BIOME-BGC and the auxiliary multiple regression equations, respectively.

The only data available from the currently applied GHG inventory calculation method (Thürig & Schmid 2008) are climate correction factors. Furthermore, no gold standard for the actual annual Swiss forest gross growth is available with which one could compare either the currently applied calculation method or alternative methods such as the ones suggested here. Thus we initially had to define how to validate these data by means of our measurements. Because climate correction factors as defined by Thürig & Schmid (2008) are (1) standardised values and (2) averaged over large areas, we had to define a variable, based on our tree ring width measurements, that is at least qualitatively comparable. We decided to base our comparisons on detrended annual tree ring width averaged over the same strata (i.e. combinations of NFI production region, altitudinal range and forest

type) as used by Thürig & Schmid (2008). The details of how the detrended annual tree ring widths are obtained are outlined in section 3.4.2. Validation is done with a subset of the data since Meyer et al. (2008) restricted themselves to a sampling of the most abundant tree species of Switzerland, Norway spruce (*Picea abies*), in the Swiss Jura mountains and the Swiss Plateau plus a few measurements taken in the Prealps, i.e. our results should at least be applicable to coniferous forests in these regions.

Prior to validation of the models, we describe the wood core data set in detail since such high resolution data sets on forest growth in Switzerland are rare and might thus be of general interest. Next, we use forest growth as estimated by the tree ring width measurements to validate the models developed by Thürig & Schmid (2008). Finally, we present alternative approaches for estimating the carbon inventory of Swiss forests.

3. Material and methods

3.1 Sampling procedure

Coworkers of Meyer et al. (2008) visited storm damaged forests during February and March 2000 right after the windstorm "Lothar" had hit Western and Central Europe on 26 December 1999. A sampling scheme could not be set up in advance because the primary aim was to determine chemical wood properties of damaged trees, and thus sampling had to take place while temperatures were still low to keep biological degradation of the wood as small as possible. Furthermore, only a fraction of the damaged areas had been cartographed by the public authorities before sampling started. Site selection was done arbitrarily by driving by car through parts of the Swiss Jura mountains and the Swiss Plateau and visual inspection of forests for windstorm damage. The selection of the resulting 104 forest sites was thus on the one hand biased towards accessibility by roads and on the other hand by the necessity to have been affected by the windstorm Lothar (Fig. 3.1). A whole range of studies were conducted in order to unravel potential causes of tree breakage or uprooting during Lothar (Indermühle et al. 2005). Generally, it was found that it was mostly wind speed and the occurrence of turbulances that explained most of the variance in the risk of a tree to be damaged. For the residual variance, important predictors were tree species, tree species mixture, forest height, height differences between forests, and acidity of the soil (Indermühle et al. 2005). Although we cannot exclude the possibility that some of these factors led to a bias in site selection which may be relevant for the estimates derived further below, the influence is probably negligible and thus we feel confident that the selected sites are close to a random sample of forest sites over the sampled area.

At each site, it was attempted to arbitrarily select at least five broken, five uprooted and five undamaged "control" trees that had been located in immediate viccinity to damaged trees, i.e. 15 trees in total. Only the conifer *Picea abies*, the tree species in Switzerland which was by far most affected by Lothar (and at the same time the most abundant tree species in Switzerland), was sampled. Not all sites contained trees of all three damage

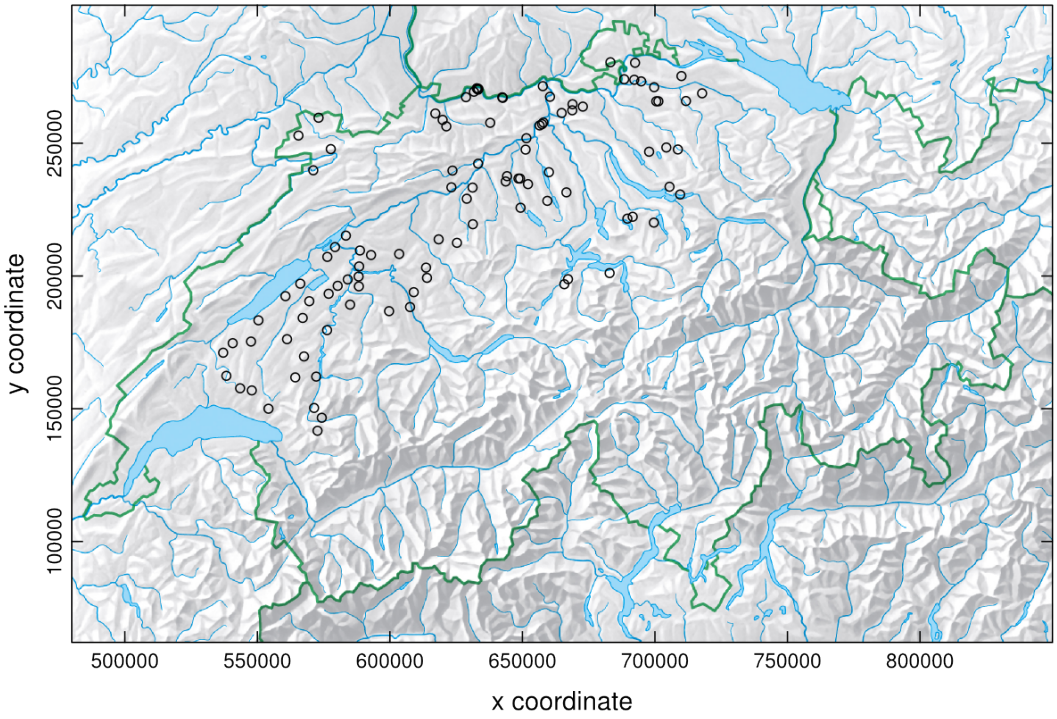

Figure 3.1 Location of the 104 sampling sites in the Swiss Jura, Plateau and Prealps. The map projection's x and y coordinate units are metres. (Relief map © 2007 Federal Office of Topography. All rights reserved.)

categories and thus median numbers for sampled broken, uprooted and control trees were 5, 2, and 7.5, respectively. At breast height (1.3 m above ground), trunk diameter was measured to the nearest centimeter and a c. 100 mm long wood core was taken with a 5 mm increment borer (Suunto Inc., Vantaa, Finland). For full details on site selection, sampling procedure and further measurements taken, consult Meyer et al. (2008).

3.2 Tree ring measurements

For the purpose the tree cores had originally been collected (wind break resistance), they were reduced at the pithward end to a standard length of 65 mm and then oven dried at 80 °C. The cores were then split in radial stem direction and one half was sanded and used for microscopic tree ring analysis. The other half was used for chemical analysis (not discussed here). Annual ring width was measured to the nearest 0.01 mm with a computerised measuring table ("LINTAB3" with TSAP software version 3.1, Rinntech.com, Heidelberg, Germany). Again, full details on methods are given in Meyer et al. (2008).

Prior to analysis, the tree ring data provided by Meyer et al. (2008) were inspected for plausibility. In total, 4 tree ring series pairs with duplicated tree and site identifiers were detected, all of which could eventually be unequivocally attributed to unique tree and site identifiers and used in the analyses. Of the initially collected 1836 cores, 1716 were usable for further processing and 1650 could be used for tree ring width analysis.

Overall, we are in the fortunate situation to have a unique data set available, with broad spatial coverage and a large number of samples within sites.

3.3 Third-party data

Apart from the data on sampling sites, trees and tree ring widths compiled and described by Meyer et al. (2008), we used several data sets from third-party sources.

First, we used the climate correction factors as estimated by Thürig & Schmid (2008) with the help of multiple regression equations for the years 1986–1999, i.e. up to the year for which tree ring width data were available. Initially, it was planned to directly use the biomass data as simulated with BIOME-BGC, however, the data are unfortunately not available anymore (S. Schmid, personal communication).

Second, we used climate data of daily resolution from several Swiss meteorological stations (1) as provided by the European Climate Assessment & Dataset (ECA&D) project (`http://eca.knmi.nl/`) made freely available by the Swiss Federal Office of Meteorology

and Climatology MeteoSwiss (see Klein Tank et al. 2002, for acknowledgements) or (2) as directly provided by MeteoSwiss (`http://www.meteo.admin.ch/`) made available solely for teaching and research purposes.

Third, we used girth tape measurements of *P. abies* diameter increment and sap flow measurements from the Swiss Canopy Crane project (Körner et al. 2005, Leuzinger & Körner 2007).

3.4 Statistical analysis

3.4.1 Characteristics of tree ring width

The tree cores were reduced to a standard length of 65 mm and thus fast growing trees having produced wider rings are represented with fewer rings in the overall sample. Due to this censoring, the further one goes back in time, the smaller the remaining sample of ring widths for the corresponding year becomes. Simultaneously, the mean of ring widths is the more underestimated the further one moves away from the sampling year 2000 and thus the most recently measured year 1999. Consequently, comparisons of ring widths between two distant time points are here only conducted using detrended measures.

To each of the 1650 censored tree ring series, a natural cubic spline with constant degrees of freedom of 5 was fitted (regardless of the length of the series). Initially, a negative exponential function was fitted to each series, but since the fit in many series was not satisfactory, this approach was discarded. In any case, we found that the particular method of detrending, and more specifically the exact parametrisation of the splines, did not alter the results decisively. The resulting trend estimates (i.e. spline functions) were used to detrend the tree ring series. In the following, *residual ring width* is defined as the measurement detrended by subtraction of the trend, and *standardised ring width* is defined as the measurement detrended by division by the trend. *Mean standardised ring width* is meant to be the standardised ring width averaged across multiple sites.

For each year, we calculated the mean for ring width, trend and standardised ring width for all 1650 measured cores. Because measurements of trees originating from the same sampling site are not independent, we applied generalised estimating equations (GEE) to obtain robust standard errors for the means. Similarly, for each year and site, we calculated the mean and standard error for the standardised ring width. Inverse distance weighted interpolation was used to estimate mean values for areas adjacent to the sampling sites.

3.4.2 Validation of current calculation method

For the validation of the climate correction factors as provided by the regression equations of Thürig & Schmid (2008), we used the standardised ring widths (see definition above) because the climate correction factors represent a standardised measure as well (simulated annual gross growth divided by 10-year mean). Thus both quantities represent relative deviations from the average trend. For each stratum for which ring width measurements were available, we calculated the mean standardised ring width. For the years 1986–1999, both mean standardised ring widths and climate correction factors were available (Thürig & Schmid 2008). We checked whether mean standardised ring widths and climate correction factors for this time period are approximately distributed according to a bivariate normal distribution. Then, we calculated Pearson product moment correlation coefficients for each stratum and tested whether the coefficients are significantly different from zero. Our null hypothesis was that the mean standardised ring widths and the correction factors do not correlate significantly. In an exemplary attempt to assess how well climate correction factors can predict mean standardised ring width, a simple linear regression model was fitted for the mean standardised ring widths of all strata for which data from at least two sites were available. The regression predictor was weighted with the inverse of the squared robust SE estimates.

3.4.3 Climatic predictors of tree ring width

To obtain reliable climatic predictors for tree ring width, multiple linear regression analysis was used to model the response variable mean standardised ring width per stratum and year.

Modelling step 1 In a first step, the same explanatory variables were used as in the calculation method of Thürig et al. (2006). Following the same notation, these are:

T annual mean temperature (°C)

$\mathbf{T}_{\text{veg}}$ mean temperature during vegetation period (April – September) (°C)

N annual precipitation (mm)

$\mathbf{N}_{\text{veg}}$ precipitation during vegetation period (April – September) (mm)

dd annual sum of degree days (sum of mean temperature of days with mean temperature $> 5\,°\text{C}$) (°C)

PP annual soil water index (potential evapotranspiration / precipitation, after Bugmann & Cramer 1998) (dimensionless)

PP_{veg} soil water index during vegetation period (April – September) (potential evapotranspiration / precipitation Bugmann & Cramer 1998) (dimensionless)

DrN drought index (after Bugmann & Cramer 1998) for coniferous forests (dimensionless)

In contrast to Thürig et al. (2006), (1) these variables were not calculated from monthly but from daily climate measurements, (2) the response variable was not a simulated variable but an independent measurement, and (3) the models were not simplified by any model selection technique. For the multiple regression analysis, only the years 1950–1999 were used (50 data points) because prior to the year 1950, the number of sites from where ring width measurements are available drops quickly. For the considered time period, only relatively few meteorological stations within the perimeter of the sites provided a complete record of measurements of daily mean temperature, precipitation, relative humidity and sunshine hours (the latter two were needed in a hydrologic model as outlined further below):

- Basel/Binningen (7.58 °E, 47.55 °N)
- Bern/Zollikofen (7.47 °E, 46.98 °N)
- Neuchâtel (6.95 °E, 47.00 °N)
- La Chaux-de-Fonds (6.80 °E, 47.08 °N)
- Luzern (8.30 °E, 47.03 °N)
- Zurich/SMA (7.47 °E, 47.38 °N)

Each sampling site was assigned to a meteorological station according to its affiliation with a specific stratum (i.e. NFI region and altitude) and, in ambiguous cases, according to the major regions as used by MeteoSwiss (the Swiss Federal Office of Meteorology and Climatology, see `http://www.meteoswiss.admin.ch/`).

Modelling step 2 In a second step, a natural cubic spline with constant degrees of freedom of 5 was fitted to the girth tape data set (from the Swiss Canopy Crane Project, see Körner et al. 2005, Leuzinger & Körner 2007) pooled over the years 2001–2005 to

obtain an estimate of instantaneous ring increment of *P. abies* over the course of a sesason. The spline function was artificially bounded by the interval 1 March to 31 October (because the exact start and end points of diameter increment are unknown) and set to zero at increments below 0 mm. The function was used to weigh the explanatory variables that only cover the vegetation period to adjust for differential growth over time:

$\mathbf{T}_{\text{veg weighted}}$ mean temperature during vegetation period (April – September) (°C)

$\mathbf{N}_{\text{veg weighted}}$ precipitation during vegetation period (April – September) (mm)

$\mathbf{PP}_{\text{veg weighted}}$ soil water index during vegetation period (April – September) (potential evapotranspiration / precipitation Bugmann & Cramer 1998) (dimensionless)

These variables were used instead of T_{veg}, N_{veg} and PP_{veg} (see modelling step 1) in the regression models.

Modelling step 3 In a third step, the daily climate data were plugged into a hydrologic model developed by Leuzinger & Körner (2009, includes a complete specification of the model) which simulates the amount of available soil water in a forest. From the temporal progression of water availability, characteristics for possible drought situations were derived and used as additional explanatory variables in the regression models (on top of the variables used in modelling step 1). Two additional predictors were defined:

SW cumulative soil water content during growing season (April – September) (mm)

$\mathbf{SW}_{\text{prev}}$ cumulative soil water content during previous growing season (April – September) (mm)

Because, except for daily climate data, many parameters were not available for the sampling sites, the measurements taken at the site described by Leuzinger & Körner (2009) were used throughout all strata. These notably include estimates of evapotranspiration and soil depth. Climate data from 1940-1949 were used to calibrate the model and only estimates of soil water content from 1950-1999 were used for the two additional predictors defined above.

Model comparison The models fitted in each of the three steps were compared by means of the Akaike's "an information criterion" (AIC) in order to assess whether the use of weights in the second step or the addition of drought criteria in the third step contribute to goodness-of-fit. The goodness-of-fit of models was further assessed and compared by

means of the coefficient of determination and the adjusted coefficient of determination. To assess predictive accuracy, leave-one-out cross-validation was applied in order to obtain estimates of absolute prediction error (expected value of the absolute difference between future and predicted responses).

3.4.4 Autocorrelation

Spatial autocorrelation due to multiple measurements per site was taken into account by using generalised estimating equations (GEE) in order to obtain robust standard errors of estimates of means. On the other hand, temporal autocorrelation was not considered explicitly in our analyses. Implicitly modelling temporal autocorrelation might improve model fitting resulting in better estimates of the variance explained by explanatory variables. However, our inferential analyses are based on summary statistics such as the mean so that autocorrelative patterns probably cancel each other out.

3.4.5 Software

All analyses were conducted with R (R Development Core Team 2010, version 2.10.1). Code from packages dplR (Bunn 2008, version 1.2.3) and pspline (version 1.0-14) was adapated for detrending. Packages gee (version 4.13-14) and geepack (version 1.0-16) were used to obtain robust standard errors. Package boot (version 1.2-41) was used for the leave-one-out cross-validation. Plots were produced with package lattice (Sarkar 2008, version 0.17-26). Throughout the plots, statistics of standardised ring width are indicated in dark green and statistics of climate correction factors (Thürig & Schmid 2008) are displayed in bright red.

4. Results

4.1 Characteristics of tree ring width

The 1650 measured tree cores were all reduced to a length of 65 mm. Therefore younger and fast growing trees tend to be represented with fewer, larger ring widths in these outer 65 mm than older and slow growing trees since the total basal area increment is distributed over a smaller or larger circumference, respectively. Consequently, cores with large ring widths are underrepresented and cores with narrow ring widths are overrepresented the further one goes back in time. The probability distribution of the oldest year still included in the tree core is given in Fig. 4.1. The median of the oldest represented year is 1966, i.e. 50 % of the cores include rings back to at least this year and cover a range of 34 years.

The mean ring width across all sites over the whole measurement period is given in Fig. 4.2c. It is accompanied by the mean of the individual series' trend lines based on natural cubic smoothing splines. Due to the overrepresentation of cores with narrow ring widths in the earlier measurement period, both the mean ring width and the mean trend forcibly increase over time (Fig. 4.2c). The mean of the standardised ring width is not affected by this problem, and therefore it is used to validate the climate correction factors by the assessment of correlation coefficients (Fig. 4.2d, see section 4.2). It is apparent from the number of sites and trees for which measurements were available (Fig. 4.2a,b) that the erratic trajectories of mean width, trend and standardised ring width in early years is due to too small sample sizes and consequently too large standard errors. Prior to 1893, the sample size consists of only a single tree and thus no standard errors can be calculated.

For each site and year, mean ring width was calculated. Inverse distance weighted interpolation was applied to these site specific year-wise mean values to illustrate their spatial and temporal variation (Fig. 4.3). Only the values for 1976–1999 are depicted. The period is of interest because it is delimited by the very dry year 1976 and the years for which the climate correction factors were available (1986–1999, see section 4.2). Note that interpolation in areas far off any sampling site becomes erratic.

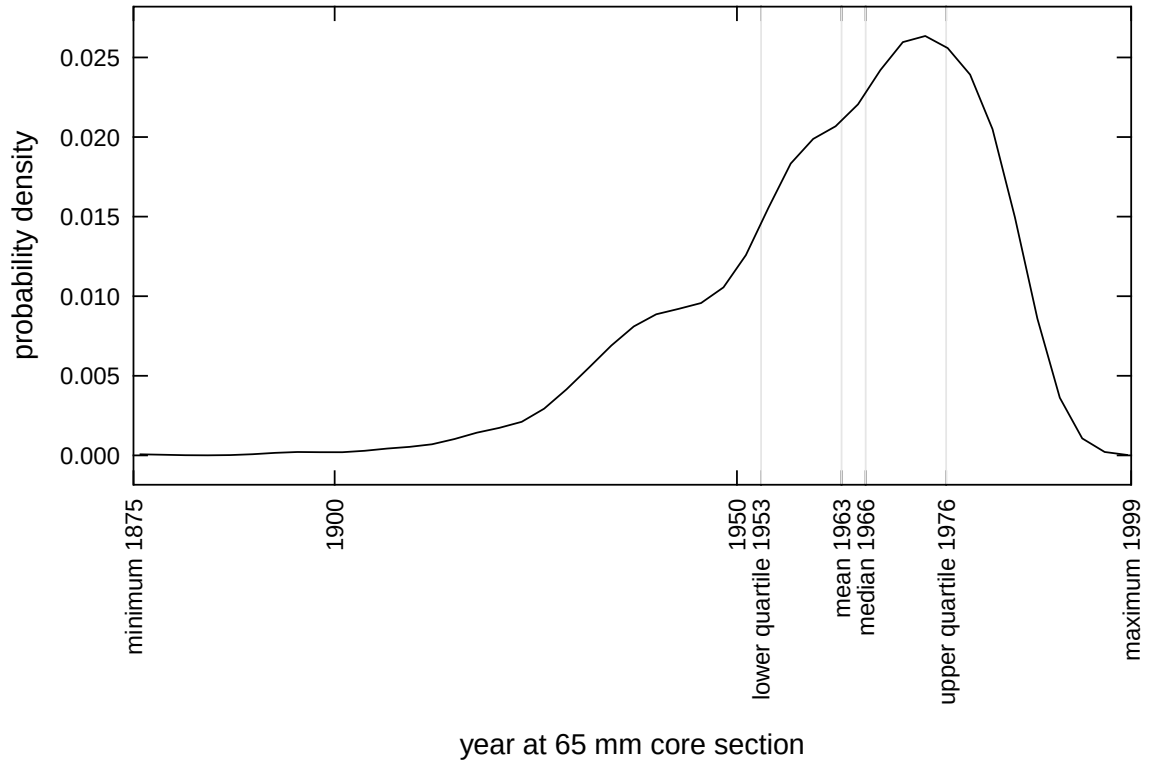

Figure 4.1 Probability distribution (kernel density estimate) of the year of the oldest ring present in each tree core (at the 65 mm core section). The vertical lines indicate the minimum, lower quartile, mean, median, upper quartile and maximum oldest ring year in the whole sample of tree cores.

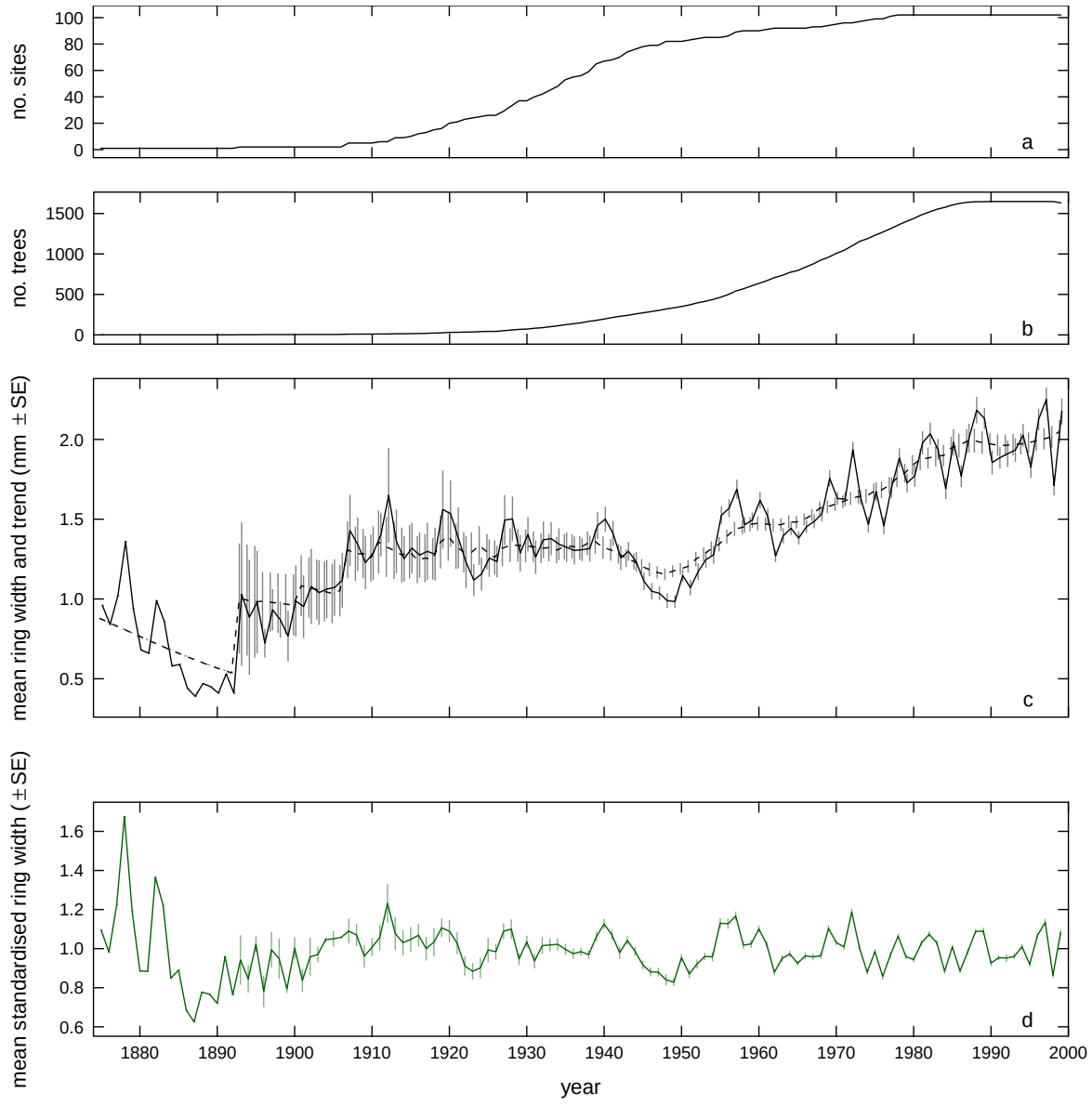

Figure 4.2 Number of sites (a) and number of trees (b) from which ring width measurements were available, mean ring width and mean trend (c, continuous and hatched line, respectively), and mean standardised ring width (d, green line). The robust standard errors (SE) are indicated with vertical bars and were estimated with generalised estimating equations (GEE) adjusting for multiple measurements within the same sampling site.

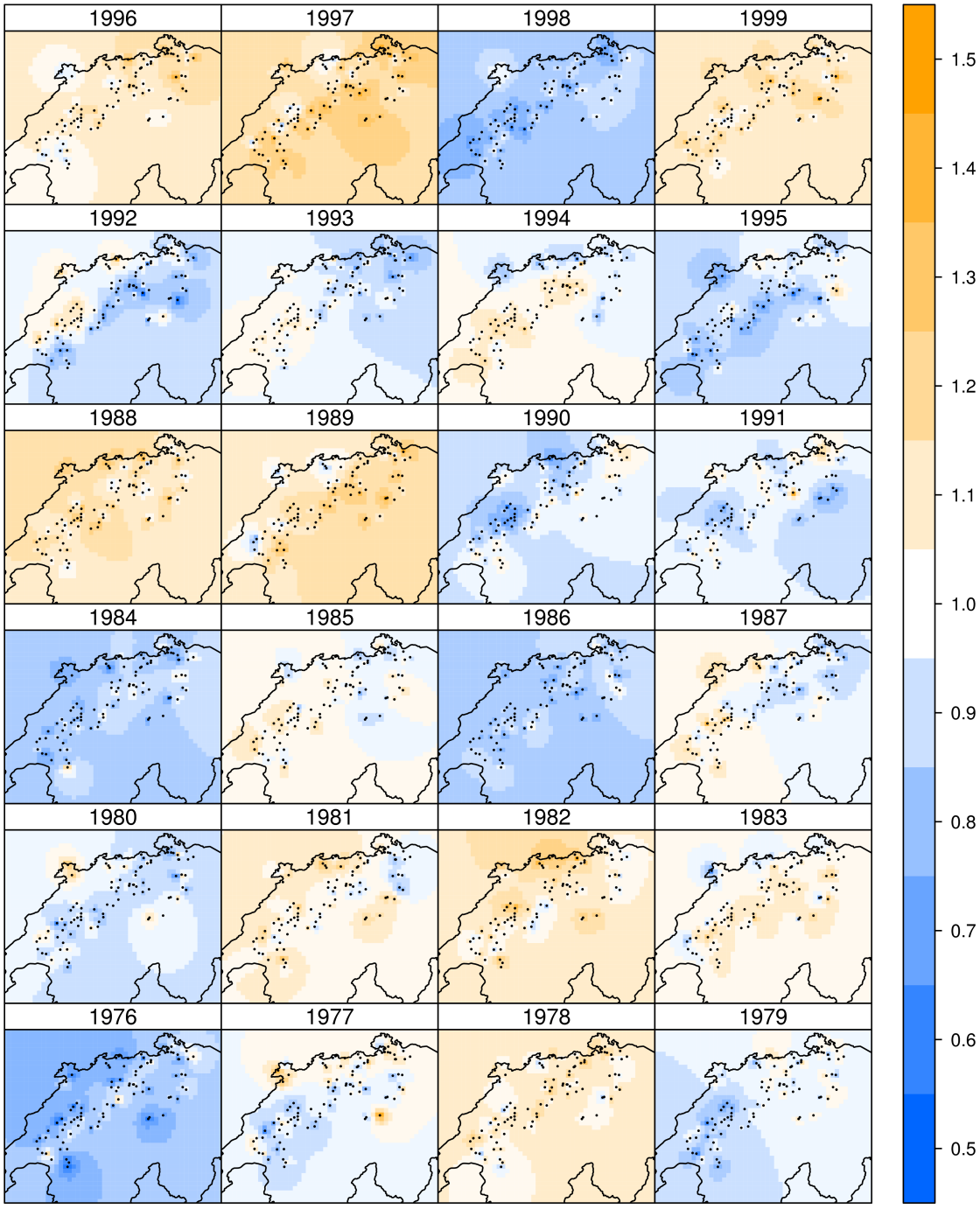

Figure 4.3 Mean standardised ring width per sampling site for the years 1976–1999 (blue indicates below average ring width increment and orange indicates above average ring width increment). Each dot marks a sampling site. Spatial prediction was done with inverse distance weighted interpolation. Note that interpolation in areas far off any sampling site becomes erratic.

4.2 Validation of current calculation method

Since we only had access to the climate correction factors estimated by multiple regression for the period 1986–2007 (Thürig & Schmid 2008) and ring width data were only available until 1999, validation of the climate correction factors was limited to the period 1986–1999, i.e. 14 years. For the six strata for which ring width measurements for at least two sites were available, the temporal trajectories of the mean standardised ring width and the climate correction factors are displayed next to each other in Fig. 4.4. For none of the strata, the pair of trajectories is in phase. The amplitude of standardised ring width tends to be smaller than the amplitude of the climate correction factors.

The values of Fig. 4.4 are plotted against each other in Fig. 4.5 to assess the degree to which standardised ring width and climate correction factors are associated. In none of the strata, a significant correlation resulted. In the stratum with the most reliable estimates of ring width (Plateau below 600 m a.s.l., the stratum with the largest number of sampling sites), the Pearson product moment correlation coefficient virtually dropped to zero. The results are so distinct that one can foreclose the inevitable conclusion: The validation of the (simulated) climate correction factors by means of the (measured) tree ring widths turned out negative. The two different approaches of estimating actual annual forest biomass increment, i.e. measurements (standardised ring width) vs. simulated values (climate correction factor), do not in the slightest describe one and the same.

The simple linear regression model for the mean standardised ring widths of all strata for which data from at least two sites were available is shown in Fig. 4.6. It is clear that prediction of mean standardised ring width with climate correction factors is futile because the 95 % confidence bands for prediction (1) span a very wide range along the ordinate and (2) run almost in parallel to a horizontal line (Fig. 4.6). The model thus does not help at all to discriminate different values of mean standardised ring width based on different values of climate correction factor. This is a further corroboration of the conclusion that the validation of the climate correction factors failed.

4.3 Climatic predictors of tree ring width

Four strata comprised a reasonable amount of sampling sites:

stratum 1 Jura, ≤ 600 m a.s.l. (16 sites)

stratum 7 Plateau, ≤ 600 m a.s.l. (40 sites)

stratum 9 Plateau, 600-1200 m a.s.l. (26 sites)

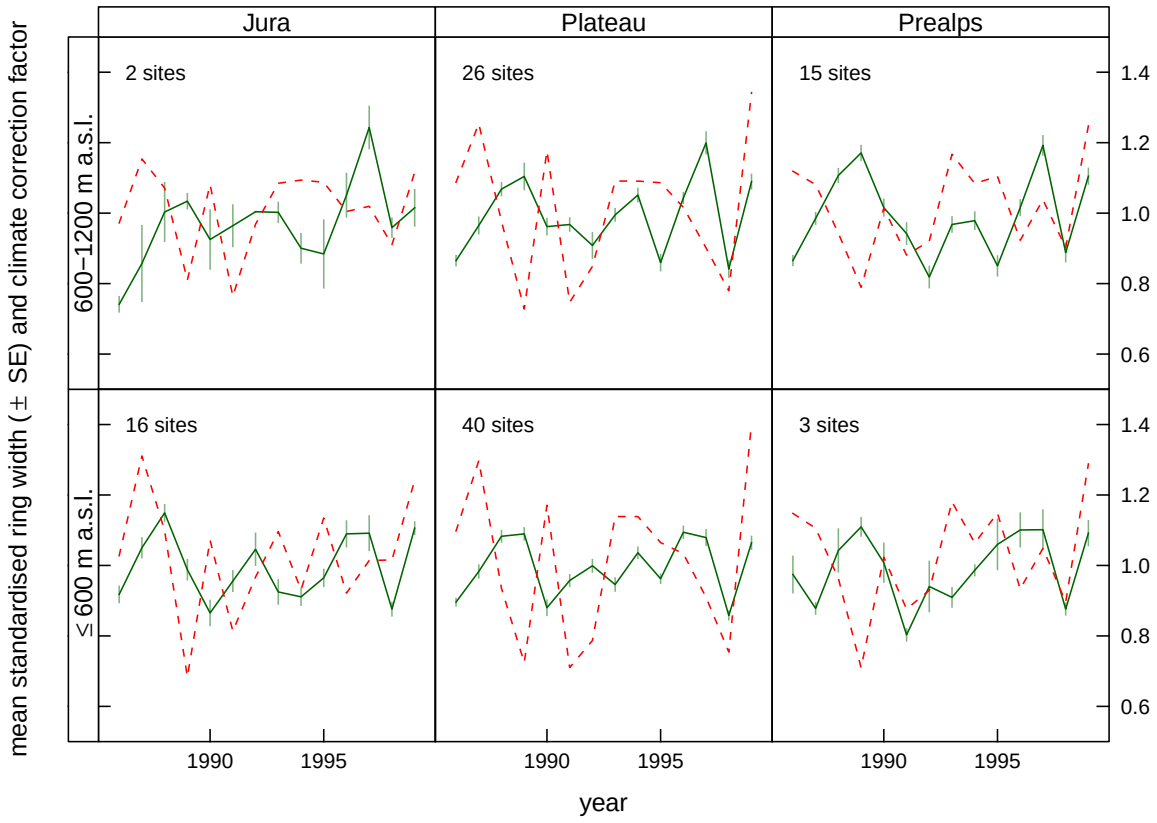

Figure 4.4 Trajectories of mean standardised ring width (green, continuous line, ± SE) and climate correction factor (red, hatched line, after Thürig et al. 2006, Thürig & Schmid 2008) from 1986–1999.

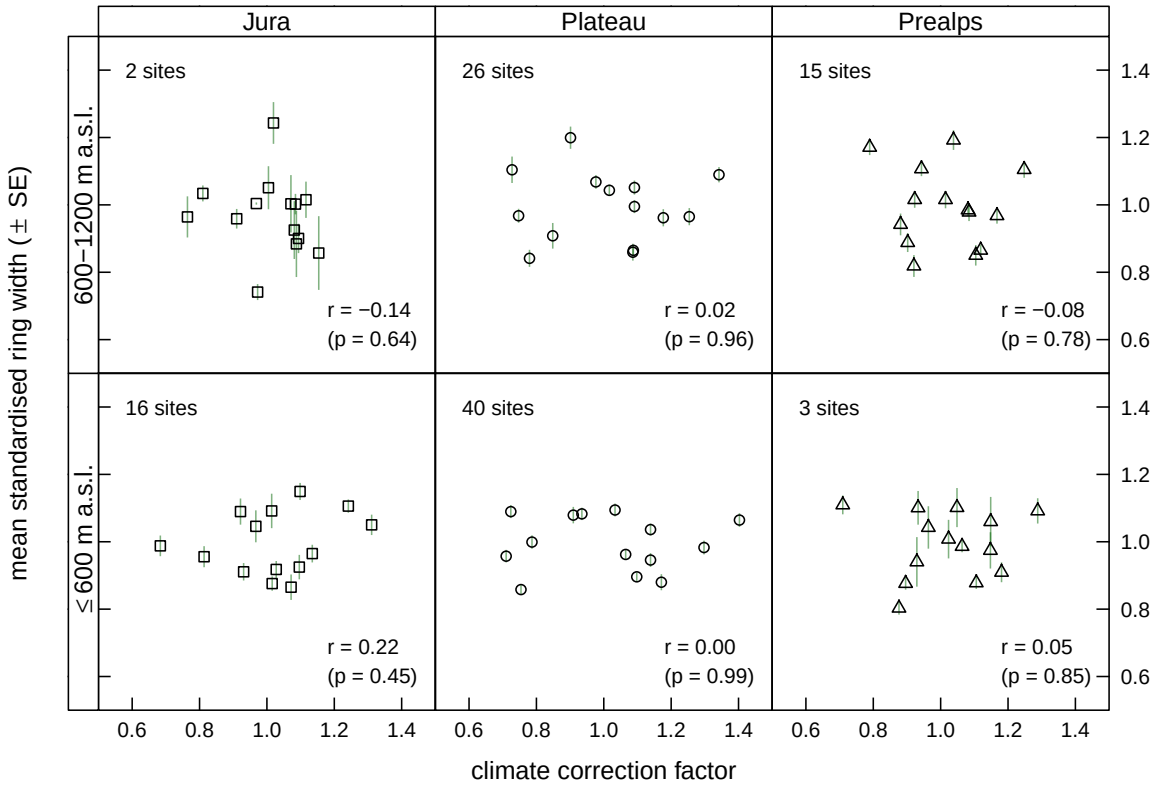

Figure 4.5 Mean standardised ring width (± SE) and climate correction factor (after Thürig et al. 2006, Thürig & Schmid 2008) from 1986–1999 plotted against each other for each stratum for which measurements were taken in at least 2 sites. Pearson product moment correlation coefficients r and p values for the hypothesis test that the coefficient is significantly different from zero are given in the lower right corner of each panel. Note that all correlation coefficient estimates are far from being significantly different from 0 and that in the stratum with the best data base (Plateau below 600 m a.s.l.), the correlation coefficient drops close to 0.

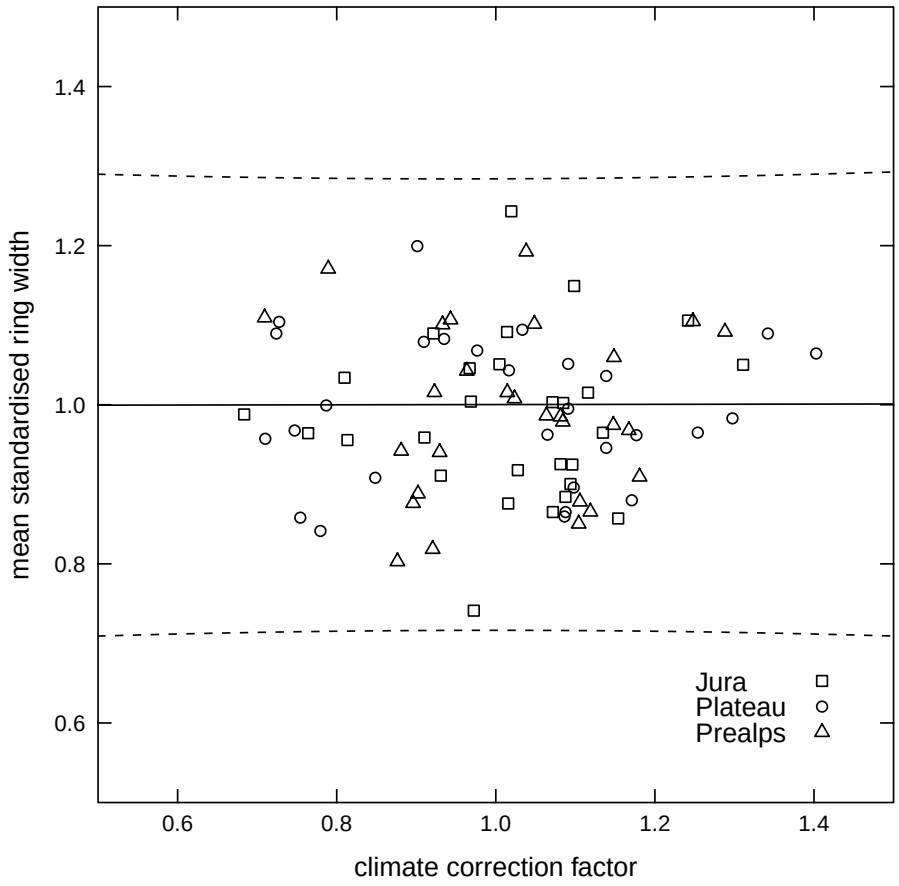

Figure 4.6 Mean standardised ring width (± SE) and climate correction factor (after Thürig et al. 2006, Thürig & Schmid 2008) from 1986–1999 plotted against each other including all strata for which at least two measurements were available. The straight line indicates a simple linear regression fit, the upper and lower hatched lines indicate the 95 % confidence bands for predicting mean standardised ring width with climate correction factors. For the prediction, it is assumed that the SE of the mean of future observations equals the mean of SEs in the sample. Prediction is basically futile because the model is essentially equivalent to a horizontal line. Furthermore, assuming mean SEs in future data points, the confidence interval of prediction would be very large.

stratum 15 Prealps, 600-1200 m a.s.l. (15 sites)

Thus in the following, only the regression models for these four strata will be considered.

Modelling step 1 The coefficients of the regression models for the response variable standardised ring width and the same explanatory variables as used in the original calculation method (Thürig et al. 2006, Thürig & Schmid 2008) are displayed in Table 4.1. Different goodness-of-fit criteria (AIC, R^2, adjusted R^2) and the cross-validation prediction error estimates are given in Table 4.4 for comparison of models of different strata and for models with alternative combinations of explanatory variables (see below).

The low coefficients of determination (between 0.086 and 0.252) indicate that only up to c. 25 % of the whole variation in the response (mean standardised ring width) are explained by the multiple regression models (Table 4.4). Nevertheless, given that the values of mean standardised ring width lie always in the range of 0.7–1.3 (compare Fig. 4.6), the estimated prediction error based on leave-one-out cross-validation is rather small, i.e. at most 0.0215 / 0.7 or c. 3.1 % off the real value (Table 4.4). Thus, given that future values of mean standardised ring width would be distributed in a similar manner as the values used to fit the models, the prediction error can be regarded as minor.

Modelling step 2 The girth tape data were used to obtain a spline function which in a very coarse way describes the 5-year average of instantaneous ring width increment (Fig. 4.7). Taking this spline function and standardising it by the maximum value yields values in the range of 0–1. Using these as weights in the multiple regression equations to adjust for differential growth over the course of the year did improve the model fitting slightly (Table 4.2).

Again, the coefficients of determination and thus the amounts of explained variance are low but a bit higher than in the standard models (modelling step 1) and lie between 0.137 and 0.260 (Table 4.4). The prediction error estimates are also slightly smaller and are, again given that the values of mean standardised ring lie always in the range of 0.7–1.3 (compare Fig. 4.6), at maximum 0.0194 / 0.7 or c. 2.8 % off the real value (Table 4.4). This is a slight improvement in predictive accuracy.

Modelling step 3 Using the soil water content as simulated by the hydrologic model of Leuzinger & Körner (2009), two additional predictors were defined, namely cumulative amount of soil water during the current growing season and the cumulative amount of soil water during the previous growing season. Example output of the model is given in

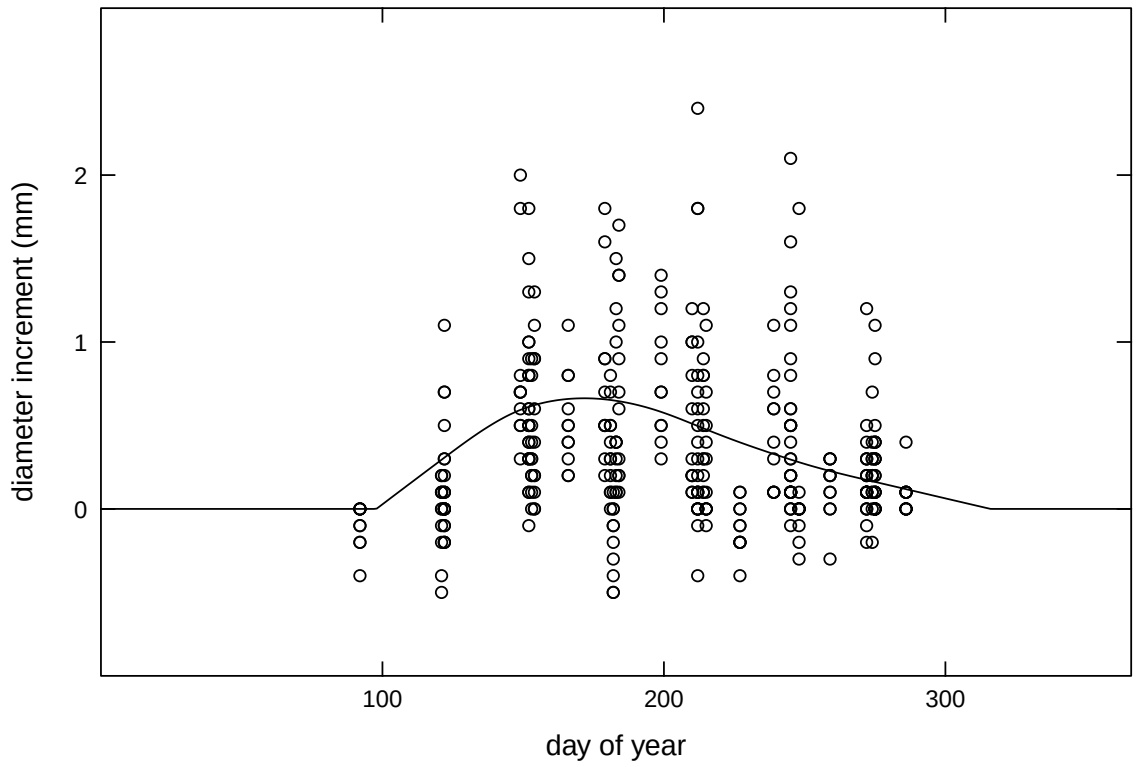

Figure 4.7 Consecutive diameter increment (i.e. difference between current and previous measurements) of 9 *P. abies* individuals from the Swiss Canopy Crane project (Körner et al. 2005, Leuzinger & Körner 2007) pooled over 5 years (2001–2005). The smoothing function is a natural cubic spline fit through the data points bounded by 1 March and 31 October and set to zero at increments below 0 mm.

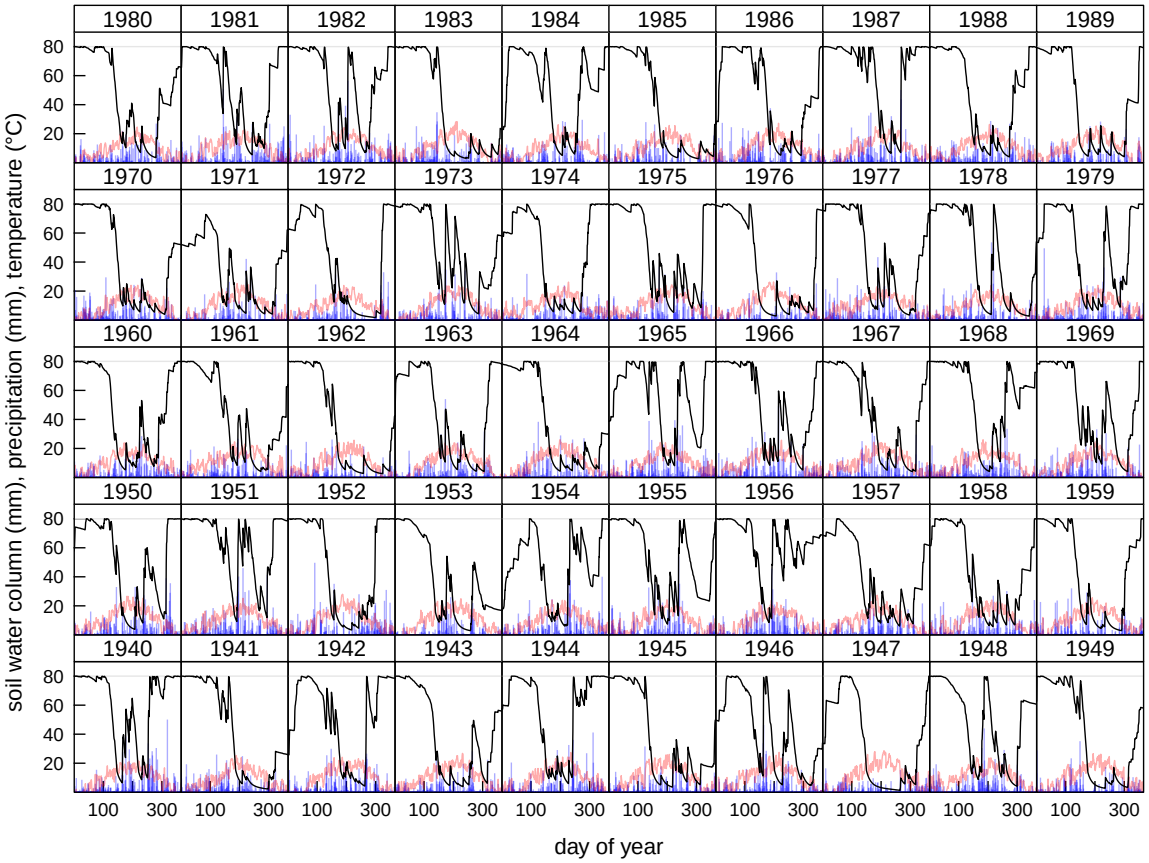

Figure 4.8 Example output of the hydrologic model developed by Leuzinger & Körner (2009) driven with data from the MeteoSwiss station Basel/Binningen (7.583° E, 47.550° N, 316 m a.s.l.) for the period 1 January 1940 to 31 Dezember 1999. The solid black line indicates the simulated soil water content (mm). Daily precipitation (blue bars) and daily mean temperature (red line) are two of the input variables and drawn in the background. For the regression models, cumulative soil water content of the current and previous growing season were extracted for the period 1950–1999. The period 1940–1949 was used to calibrate the model.

Fig. 4.8 to illustrate how the two additional predictors were derived. Note that the model was calibrated during the first decade and the simulated values produced during this period were discarded. Adding the predictors to the regression models, whose coefficients are listed in Table 4.3, did not substantially improve model fitting either, rather on the contrary.

The coefficients of determination might seem to be higher than for the ordinary models (modelling step 1), but as two more predictors were used in the models, the adjusted coefficients of determination should be used to compare models of modelling steps 1 and 3 (since adding variables to a model can only inflate but not deflate the amount of explained variance). It becomes clear when penalising for the number of explanatory variables used in the model that the goodness-of-fit of models of modelling step 3 is not consistently higher than of the models of modelling step 1 (Table 4.4). The maximum prediction error estimates are actually the highest of all modelling steps, again given that the values of mean standardised ring width lie always in the range of 0.7–1.3 (compare Fig. 4.6), resulting in 0.0276 / 0.7 or c. 3.9 % off the real value (Table 4.4).

As already mentioned above, a comparison of the models fit in the three steps for the four selected strata is given in Table 4.4. For each model, Akaike's "an information criterion" (AIC), the coefficient of determination (R^2), the adjusted coefficient of determination (R^2 adjusted), the leave-one-out cross-validation prediction error estimate (CV prediction error) are given. The first three are used to assess goodness-of-fit and the last one to assess predictive accuracy.

Table 4.1 Modelling step 1: coefficients of regression models for strata 1, 7, 9, and 15.

	stratum 1 (AIC = -57.41)			stratum 7 (AIC = -83.07)			stratum 9 (AIC = -96.55)			stratum 15 (AIC = -92.65)		
	estimate	SE	*p*	estimate	SE	*p*	estimate	SE	*p*	estimate	SE	*p*
(intercept)	1.6578	0.9880	0.10	-0.4780	1.0259	0.64	1.8296	0.8943	0.05	0.6396	1.1455	0.58
T	-0.0400	0.0880	0.65	-0.0741	0.0647	0.26	-0.0684	0.0561	0.23	-0.0545	0.0600	0.37
T_{veg}	0.0064	0.0500	0.90	0.0003	0.0432	0.99	-0.0033	0.0361	0.93	-0.0155	0.0499	0.76
N	-0.0001	0.0008	0.91	0.0007	0.0007	0.34	-0.0007	0.0005	0.20	-0.0003	0.0008	0.69
N_{veg}	-0.0009	0.0009	0.33	-0.0000	0.0007	0.96	-0.0002	0.0006	0.72	0.0008	0.0010	0.44
dd	0.0001	0.0003	0.58	0.0000	0.0002	0.91	0.0003	0.0002	0.10	0.0001	0.0002	0.64
PP	-0.0364	0.0628	0.57	0.1041	0.1350	0.45	-0.1077	0.0789	0.18	-0.0912	0.1703	0.60
PP_{veg}	-0.0229	0.0340	0.50	-0.0106	0.0569	0.85	-0.0233	0.0367	0.53	0.0742	0.1062	0.49
DrN	0.3026	0.7847	0.70	1.2892	0.6314	0.05	0.8201	0.5491	0.14	0.7425	0.5030	0.15

Table 4.2 Modelling step 2: coefficients of regression models for strata 1, 7, 9, and 15. The predictors T_{veg}, N_{veg}, and PP_{veg} were weighted with the spline function obtained from the diameter increment data in Fig. 4.7 and used to replace the unweighted predictors of modelling step 1.

	stratum 1 (AIC = -60.24)			stratum 7 (AIC = -84.09)			stratum 9 (AIC = -97.08)			stratum 15 (AIC = -94.17)		
	estimate	SE	*p*	estimate	SE	*p*	estimate	SE	*p*	estimate	SE	*p*
(intercept)	2.6257	0.9788	0.01	-0.2765	1.0542	0.79	1.8890	0.9412	0.05	0.2067	1.2119	0.87
T	0.0601	0.0825	0.47	-0.0650	0.0619	0.30	-0.0607	0.0557	0.28	-0.0618	0.0565	0.28
$\mathbf{T}_{\text{veg weighted}}$	-0.2619	0.2107	0.22	-0.0213	0.1683	0.90	-0.0813	0.1482	0.59	-0.0787	0.1482	0.60
N	-0.0008	0.0008	0.29	0.0004	0.0007	0.51	-0.0008	0.0005	0.11	-0.0002	0.0006	0.73
$\mathbf{N}_{\text{veg weighted}}$	0.0002	0.0018	0.90	0.0005	0.0015	0.72	0.0000	0.0011	0.97	0.0020	0.0015	0.20
dd	0.0001	0.0002	0.56	0.0001	0.0002	0.80	0.0003	0.0002	0.07	0.0002	0.0002	0.42
PP	-0.0409	0.0620	0.51	0.0704	0.1244	0.57	-0.1184	0.0730	0.11	-0.0596	0.1420	0.68
$\mathbf{PP}_{\text{veg weighted}}$	-0.0080	0.0163	0.62	0.0023	0.0308	0.94	-0.0055	0.0186	0.77	0.0489	0.0456	0.29
DrN	-0.9776	0.8257	0.24	1.0717	0.6657	0.12	0.7067	0.5708	0.22	0.5809	0.4883	0.24

Table 4.3 Modelling step 3: coefficients of regression models for strata 1, 7, 9, and 15. In addition to the ordinary models (modelling step 1), two additional predictors representing characteristics of soil water availability as estimated by a hydrologic model are used (SW, cumulative soil water content, and SW_{prev}, cumulative soil water content of previous year).

	stratum 1 (AIC = -51.11)			stratum 7 (AIC = -90.37)			stratum 9 (AIC = -91.78)			stratum 15 (AIC = -89.42)		
	estimate	SE	*p*	estimate	SE	*p*	estimate	SE	*p*	estimate	SE	*p*
(intercept)	1.6825	1.0646	0.12	-0.9796	0.9407	0.30	1.1950	1.0103	0.24	0.0502	1.1965	0.97
T	-0.0345	0.0929	0.71	-0.0860	0.0592	0.15	-0.0789	0.0573	0.18	-0.0583	0.0605	0.34
T_{veg}	0.0028	0.0533	0.96	0.0088	0.0395	0.83	0.0096	0.0385	0.80	-0.0140	0.0509	0.78
N	-0.0001	0.0008	0.91	0.0007	0.0006	0.25	-0.0005	0.0005	0.35	-0.0004	0.0008	0.66
N_{veg}	-0.0008	0.0011	0.44	0.0001	0.0007	0.91	-0.0000	0.0006	0.96	0.0010	0.0010	0.32
dd	0.0001	0.0003	0.64	-0.0000	0.0002	0.90	0.0003	0.0002	0.15	0.0001	0.0002	0.62
PP	-0.0336	0.0666	0.62	0.1071	0.1213	0.38	-0.0880	0.0810	0.28	-0.0973	0.1707	0.57
PP_{veg}	-0.0237	0.0357	0.51	0.0042	0.0515	0.94	-0.0158	0.0377	0.68	0.1063	0.1098	0.34
DrN	0.2666	0.8223	0.75	1.3873	0.6189	0.03	0.9488	0.5806	0.11	0.6959	0.5147	0.18
SW	-0.0945	0.4252	0.83	0.1076	0.2123	0.62	-0.0153	0.2327	0.95	0.1536	0.2445	0.53
$\mathbf{SW}_{prev}$	0.0812	0.2148	0.71	0.3864	0.1193	0.00	0.1782	0.1276	0.17	0.2056	0.1260	0.11

Table 4.4 Comparison of the three modelling steps for the four selected strata. Akaike's "an information criterion" (AIC), the coefficient of determination (R^2), the adjusted coefficient of determination (R^2 adjusted), and the leave-one-out cross-validation prediction error estimate (CV prediction error) are given for each model.

stratum	modelling step	AIC	R^2	R^2 adjusted	CV prediction error
1	1	-57.4	0.086	-0.092	0.0215
1	2	-60.2	0.137	-0.032	0.0194
1	3	-51.1	0.091	-0.148	0.0276
7	1	-83.1	0.205	0.050	0.0114
7	2	-84.1	0.221	0.069	0.0112
7	3	-90.4	0.404	0.247	0.0095
9	1	-96.6	0.252	0.106	0.0086
9	2	-97.1	0.260	0.115	0.0083
9	3	-91.8	0.288	0.101	0.0089
15	1	-92.6	0.160	-0.004	0.0090
15	2	-94.2	0.185	0.026	0.0089
15	3	-89.4	0.223	0.018	0.0093

5. Discussion

Based on a broadly spaced dendrological sample of 1650 spruce trees, we constructed multiple linear regression models to predict (a proxy of) annual forest biomass increment in Switzerland outside the Alps based on standard meteorological data. Tree ring width does not seem to be unequivocally driven by a single climatic variable. Nevertheless, the predictive power of the models is in the range of about 3 %, which makes them applicable to the estimation of forest biomass increment, not solely in the context of the greenhouse gas inventory. Our data set for this most important tree species in central Europe covers a wide spectrum of locations and includes very dry (e.g. the year 1976) and very wet periods. The patterns seen in spruce may, at least trendwise, hold for many other co-occuring tree species. Handling such data inevitably incurs a number of difficulties and sources of errors. Our sample was restricted to the outer 65 mm of stems of tall, c. 100 year old trees, and thus, trees with high vigour in recent decades ended up with fewer rings captured than slow growing trees.

5.1 Temporal and spatial patterns of tree ring width

Even though the tree core sample was limited to a single (albeit the most common) tree species, covered only the outer 65 mm of the trunks and was biased towards storm affected sites, it provides valuable insights into temporal and spatial patterns of tree growth in the Swiss Jura mountains and Plateau. As initially mentioned, thicker ring widths are overrepresented in more recent years, be it because trees with thicker rings were younger at the time of sampling and the secondary growth was distributed over a smaller area or be it because trees with thicker rings of comparable age actually grew faster. We took this fact into account by detrending each ring width series and working only with dimensionless, detrended ring width (which we call standardised ring width throughout the report). The temporal evolution of the overall mean of standardised ring width as shown in Fig. 4.2d makes it clear that from the year 1950 onwards, when a couple of hundred measurements were available, the annual negative and positive deviances of ring

width were rather synchronous across sites. This becomes evident when one considers the very small standard errors and thus the high precision of the estimates of the mean. When one moves from the overall mean to the site specific mean standardised ring width, spatial differences within a given year emerge (Fig. 4.3). Although some sites largely departed from the rest of the sites, on the whole, as was already implied by the precise estimates of the overall mean, the majority of sites was surprisingly synchronous with regard to negative and positive annual deviations, at least in close spatial proximity.

5.2 Failed validation of current calculation method

We tested projections derived from a mechanistic forest growth model based on classical input variables such as photosynthesis, respiration, allocation rules and nutrient feedback (Thürig & Schmid 2008) against real data from the Swiss Jura and Plateau as represented by measured tree ring widths. In essence, the validation of the currently applied calculation method of the carbon stocks of Swiss forests for the greenhouse gas inventory was based on the hypothesis that, within a stratum, detrended ring width (i.e. mean standardised ring width) should correlate with the corresponding climate correction factor (i.e. mean standardised simulated gross growth). (A stratum is a combination of NFI production region, altitudinal range, and forest type as defined by Thürig & Schmid (2008)). The significance tests for the Pearson correlation coefficients clearly showed that in none of the strata for which data from multiple sites were available any association could be found (Fig. 4.5). Not even a slight trend could be detected, and in the stratum with the most profound data base (Pleateau below 600 m a.s.l. with 40 sites, smallest SE), the corelation coefficient basically dropped to 0. This is even more astonishing because the sample size of 14 data points per correlation is rather low and thus chances that spurious correlations are found may be higher due to the violation of the assumption of a bivariate normal distribution of data points. In the same vein, the absence of a significant correlation means that it is basically impossible to predict mean standardised ring width with climate correction factors and vice versa (compare Fig. 4.6).

The hypothesis that mean standardised ring width and climate correction factors should correlate is based on the assumption that tree ring width is a suitable proxy of tree biomass increase (as has been shown for *P. abies* e.g. by Mäkinen et al. 2002). Given this is the case to a measurable extent and in view of the fact that the estimates of mean standardised ring width are based on a very large sample of 1650 trees collected over a relatively large area (Meyer et al. 2008), we conclude that the mechanistic ecosystem

simulation model BIOME-BGC failed to reliably predict carbon stocks. Although the accordance of BIOME-BGC predictions with measurements has been tested for a few sites in Switzerland (Schmid et al. 2006), the model's predictive accuracy might nevertheless not be sufficient (see appendix A for a critique).

5.3 Statistical modelling of climate correction factors and tree ring width

The conclusion from the previous section is very explicit: climate correction factors as defined by Thürig & Schmid (2008) do not adequately reflect reality and should therefore not be used to predict forest carbon stocks for the Swiss greenhouse gas inventory. But why were Thürig & Schmid (2008) and coworkers confident to use climate correction factors nonetheless? The quintessence of our critique was already uttered in section 2.3 and shall not be repeated. For an in-depth critique of the modelling approach based on BIOME-BGC, we refer you to appendix A.

As a consequence of the failed validation of the climate correction factors as a central part of the calculation method for the greenhouse gas inventory (Heldstab et al. 2009), we propose that future calculation methods should use measurement data instead of data generated from climate data driven mechanistic simulation models. Ideally, such measurements are conducted on an annual basis in order to obtain reliable up-to-date estimates of Swiss forest carbon stocks. One could obtain measurements from the NFI (Brassel & Brändli 1999) which presumably changes its sampling scheme from a 10-year to an annual survey in the near future. Alternatively, an annual sampling scheme specifically tailored to the needs of the greenhouse gas inventory could be developed, covering the whole area of Switzerland and stratifying to forest type and possibly other variables. The sampling scheme could include tree ring width measurements by taking cores as used in the study at hand as well as other measurements of tree biomass in order to obtain data on the quality of tree rings as a proxy for biomass increment. In essence, no modelling effort would be needed at all given the measurements are a good proxy for tree biomass increment.

Nevertheless, in case measurements cannot be obtained in due course, one would have to resort to modelling. Measurement data could be used to build statistical models for the prediction of future annual gross growth of Swiss forests. It was not the aim of this report to present a comprehensive solution but to suggest a couple of modelling approaches which might be worth being refined or built upon. In the following, the different approaches are

discussed successively.

To check whether mean standardised ring width could be explained by climatic data, three approaches were taken (see Table 4.4 for a comparison of different measures of goodness-of-fit and predictive accuracy). First of all, a multiple regression approach analogous to the one of Thürig & Schmid (2008) was chosen (Tables 4.1 and 4.4). In contrast to Thürig et al. (2006), the explanatory variables were calculated from daily climate data, no model simplification technique was applied and, most importantly, the response variable was the mean standardised ring width instead of the climate correction factors. Due to these differences (particularly in the modelled response variables) and due to the potential problems with the multiple regression models for the climate correction factors (as outlined extensively in appendix A), a direct comparison of goodness-of-fit criteria of the models presented here and in Thürig et al. (2006) is not meaningful, but nonetheless instructive. R^2 values of the models for mean standardised ring width are considerably lower, as one would expect for ecological data. The very small adjusted R^2 values could possibly be increased by model reduction, but we intentionally did not use any such technique in order to keep the comparison of the different modelling approaches simple. Because the ultimate goal consists in building prediction models, the estimates of the cross-validation prediction error are a more meaningful measure to be considered here. The (absolute) cross-validation prediction error is the expected value of the absolute difference between the future and predicted responses. Since we applied leave-one-out cross-validation, a potential bias of the estimates should be negligible. From Fig. 4.6 it is apparent that all stratum specific mean standardised ring width values are in the range of 0.7–1.3. Considering this range, the prediction error estimates in Table 4.4 seem to be reasonably small to justify the use of such models for prediction purposes, e.g. for reports such as the NIR (the expected prediction error over all considered models would thus vary between c. 0.6–3.9 %).

In the next step, the three climatic variables that only integrated data over the putative vegetation period (April – September) were replaced by variants where each daily measurement was weighted by a standardised growth curve obtained from *P. abies* diameter increment data (Fig. 4.7). As can be seen from a comparison of Tables 4.1 and 4.2, this partial replacement of explanatory variables led to a decrease of AIC values in all four considered strata implying a better goodness-of-fit. It also resulted in a slight improvement with regard to cross-validation prediction error (Table 4.4). Consequently, it might be worthwile to investigate how climatic predictors could be weighted in order to better reflect tree growth.

In the last step, the ordinary models (Table 4.1) were extended with two additional explanatory variables, namely the cumulative soil water content of the current and previous growing period as simulated by the hydrologic model of Leuzinger & Körner (2009) (Table 4.3). For an adequate application of the hydrologic model, many parameters apart from daily climate data would be needed. However, many characteristics of the sampling sites (soil depth, measurements of evapotranspiration, vegetation period, etc.) were only available for the site described in Leuzinger & Körner (2009) (i.e. Basel/Binningen) and were thus also used for the other sampling sites. This might partly explain why the extended models of 3 out of the 4 considered strata have worse goodness-of-fit measures than the ordinary models (Table 4.4). The notable exception is the extended model for stratum 7, the stratum with the most reliable data base (40 sampling sites): cumulative soil water content of the previous vegetation period was a very good predictor of mean standardised ring width (Table 4.3) and contributed to an increased goodness-of-fit over the original model and the model with weighted predictors (Table 4.4). A thorough assessment is however pending: Is this only a spurious finding? Or can the regression models be substantially improved by refining the application of the hydrologic model and by defining more meaningful drought criteria derived thereof?

Regardless of which of the presented modelling approaches one considers, the estimated prediction error would lie somewhere near 3 %, which seems very acceptable indeed, realising that the climate correction factors (Thürig & Schmid 2008) represent, as has been demonstrated, nothing close to actual measurements of forest biomass and that past greenhouse gas inventory calculations were based on exactly such climate correction factors.

Acknowledgements We thank Fabian Meyer for the collection and preparation of the tree ring data set; Jan Esper and Dave Frank for recommendations regarding data analysis; Stéphanie Schmid and Esther Thürig for providing insight into their calculation method for the Swiss Greenhouse Gas Inventory; Roland Asshoff, Manuel Mildner and Susanna Riedl for providing the girth tape data; Sebastian Leuzinger for access to the source code of his hydrologic model; David M. Lawrence for access to the source code of his radiation estimation algorithm; Erika Hiltbrunner for tips on sources of long term meteorological data; MeteoSwiss (the Swiss Federal Office of Meteorology and Climatology) for providing climate data; and Nele Rogiers for the project draft, the climate correction factors and helpful comments.

This work was financially supported by the Swiss Federal Office for the Environment (FOEN).

A. Critique of modelling approach based on BIOME-BGC (by Thomas Zumbrunn)

In this appendix, I elaborate on some of the potential problems inherent in the GHG calculation method put forward by Thürig & Schmid (2008), described in detail in Thürig et al. (2006) and used for the NIR (Heldstab et al. 2009). A consideration of these problems is important when evaluating the relative merits of the criticised calculation method and alternative approaches.

A.1 Assessment of predictive accuracy of BIOME-BGC

The predictive accuracy of the model BIOME-BGC has been validated with measurements at 19 selected sites in Switzerland (Schmid et al. 2006) and it was argued that BIOME-BGC is sufficiently well calibrated for the purposes of GHG reporting (Thürig & Schmid 2008). Schmid et al. (2006) present statistical tests on the mean deviation of the model from measured growth and conclude that the mean underestimation in the colline region is not significantly different from zero. Yet from their figures it can be derived that the mean absolute percentual deviation is c. 23 % with a 95 % confidence interval of [13 %; 32 %] (assuming an approximate normal distribution of the 19 absolute percentage values). Therefore, without averaging over multiple sites (for which the model either under- or overestimates growth), significant deviations may result. The statistical maxim that "absence of evidence is not evidence of absence" also holds true here, i.e. one cannot show that there is a difference of zero between two populations, and therefore the statistical tests applied by Schmid et al. (2006) are not very helpful in assessing the predictive accuracy.

A.2 Interpretation of multiple regression models

The way Thürig & Schmid (2008) interpret multiple linear regression models relies, amongst others, on the assumption that the response variable is obtained independently from the explanatory variables. If this type of independence is not given, interpretation of the models may become erroneous. As illustrated in Thürig et al. (2006), this assumption was violated. The response variable is basically a construct, albeit a very complex one (via the simulation model BIOME-BGC), of climatic variables which are highly correlated with the explanatory variables (Thürig et al. 2006). Thus it is by no means surprising that Thürig et al. (2006) obtain relatively high values for the coefficient of determination R^2 (proportion of explained variance) for their multiple regression models. However, the conclusion that the high values for R^2 indicate high predictive accuracy of the models is deceptive. The missing independence between response variable and explanatory variables as outlined above prbabably artificially inflates R^2 values. Other reasons might be a misinterpretation of the implications of the coefficient of determination. The coefficient of determination is only a measure of how well a model fits the very data set with which its coefficients were estimated. Thus R^2 can change dramatically e.g. by taking a subset of the data set or adding/removing influential points. Furthermore, the R^2 value can only grow but not shrink when explanatory variables are added to the model. To be able to compare the goodness-of-fit of models with different constellations of explanatory variables, one could use the adjusted coefficient of determination (adjusted R^2) or information criteria such as Akaike's "an information criterion" (AIC) which penalise for increasing model complexity. The regression models in Thürig et al. (2006), after having been reduced by an unknown model selection procedure, contain between 1 to 5 explanatory variables (plus an intercept each). Due to the different complexity of the models, the comparison of goodness-of-fit of these models by means of R^2 values is not valid (as mentioned, measures such as adjusted R^2 or AIC should be used instead). Moreover, the 34 models in Thürig et al. (2006) do not only differ in the constellation of explanatory variables but also in the number of data points that were used for model fitting. For instance, only 27 data points were used for the models of strata 5 and 6 with initially 8 explanatory variables (plus an intercept) in the full models (i.e. before the application of model selection procedures). The number of data points that are needed to obtain reliable estimates of all model coefficients is thus too small. In the given example, initially merely three data points were used per coefficient estimate, and thus the subsequent model reduction procedures might have led to erroneous results. Such an overspecification of regression models is a further source of artificial inflation of (unadjusted) R^2 values since the coefficient of determina-

tion reaches a value of 1 and the residuals all become zero when the number of model coefficients equals the number of data points (because the residual degrees of freedom drop to zero).

References

Brassel P, Brändli U (eds.) (1999) Schweizerisches Landesforstinventar: Ergebnisse der Zweitaufnahme 1993-1995. Paul Haupt

Bugmann H, Cramer W (1998) Improving the behaviour of forest gap models along drought gradients. For Ecol Manage 103:247–263

Bunn AG (2008) A dendrochronology program library in R (dplR). Dendrochronologia 26:115–124, ISSN 1125-7865

Heldstab J, Bolliger R, Schellenberger A (2009) Switzerland's Greenhouse Gas Inventory 1990–2007. National Inventory Report 2009. Tech. rep., Federal Office for the Environment (FOEN)

Indermühle M, Raetz P, Volz R (2005) Lothar. Ursächliche Zusammenhänge und Riskoentwicklung. Synthese des Teilprogramms 6. Umwelt-Materialien Nr. 184, Bundesamt für Umwelt, Wald und Landschaft (BUWAL)

Klein Tank AMG, Wijngaard JB, Können GP, Böhm R, Demarée G, Gocheva A, Mileta M, Pashiardis S, Hejkrlik L, Kern-Hansen C, Heino R, Bessemoulin P, Müller-Westermeier G, Tzanakou M, Szalai S, Pálsdóttir T, Fitzgerald D, Rubin S, Capaldo M, Maugeri M, Leitass A, Bukantis A, Aberfeld R, van Engelen AFV, Forland E, Mietus M, Coelho F, Mares C, Razuvaev V, Nieplova E, Cegnar T, López JA, Dahlström B, Moberg A, Kirchhofer W, Ceylan A, Pachaliuk O, Alexander LV, Petrovic P (2002) Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. Int J Climatol 22:1441–1453

Körner C, Asshoff R, Bignucolo O, Hättenschwiler S, Keel SG, Peláez-Riedl S, Pepin S, Siegwolf RTW, Zotz G (2005) Carbon flux and growth in mature deciduous forest trees exposed to elevated CO_2. Science 309:1360–1362

Leuzinger S, Körner C (2007) Water savings in mature deciduous forest trees under elevated CO_2. Global Change Biol 13:2498–2508

Leuzinger S, Körner C (2009) Rainfall distribution is the main driver of runoff under future CO_2-concentration in a temperate deciduous forest. Global Change Biol

Meyer F, Paulsen J, Körner C (2008) Windthrow damage in *Picea abies* is associated with physical and chemical stem wood properties. Trees 22:463–473

Mäkinen H, Nojd P, Isomäki A (2002) Radial, height and volume increment variation in *Picea abies* (L.) Karst. stands with varying thinning intensities. Scand J For Res 17:304–316

Pasi T, Gross D, Schmutz A (2009) Jahrbuch Wald und Holz 2008. Umwelt-Wissen Nr. 0904. Tech. rep., Federal Office for the Environment FOEN

R Development Core Team (2010) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Sarkar D (2008) Lattice. Multivariate data visualisation with R. Springer

Schmid S, Zierl B, Bugmann H (2006) Analyzing the carbon dynamics of central European forests: comparison of Biome-BGC simulations with measurements. Reg Environ Change 6:167–180

Thornton P, Law B, Gholz H, Clark K, Falge E, Ellsworth D, Goldstein A, Monson R, Hollinger D, Falk M, Chen J, Sparks J (2002) Modeling and measuring the effects of disturbance history and climate on carbon and water budgets in evergreen needleleaf forests. Agric For Meteorol 113:185–222

Thürig E, Kaufmann E, Schmid S, Bugmann H (2006) Treibhausgas Inventar: Waldkennzahlen und jährlicher Klimaeinfluss. Tech. rep., Swiss Federal Institute for Forest, Snow and Landscape Research and Chair of Forest Ecology, Department of Environmental Sciences, ETH Zurich

Thürig E, Schmid S (2008) Jährliche CO_2-Flüsse im Wald: Berechnungsmethode für das Treibhausgasinventar. Swiss For J 159:31–38