



Detecting changes in essential ecosystem and biodiversity properties - towards a Biosphere Atmosphere Change Index: BACI

Deliverable 4.3: Spatialized and yearly resolved tree-ring widths for Europe



Project title:	Detecting changes in essential ecosystem and biodiversity properties - towards a Biosphere Atmosphere Change Index
Project acronym:	BACI
Grant agreement number:	640176
Main pillar:	Industrial Leadership
Topic:	EO-1-2014: New ideas for Earth-relevant space applications
Start date of the project:	1 st April 2015
Duration of the project:	48 months
Dissemination level:	Public
Responsible of the deliverable:	Flurin Babst Phone: ++48 579 516 164 Email: flurin.babst@wsl.ch
Authors and contributors:	Flurin Babst, Paul Bodesheim, David Frank, Martin Jung, Miguel Mahecha
Date of submission:	30 th March 2017

1 Introduction

Tree-ring data are an underexploited resource in Earth system science. This is partly because their spatial and temporal representation of global forest biomes is discontinuous and overlaps insufficiently with modern Earth observations (Babst et al., 2017). Improving this representation is of broad research interest because tree-rings can provide (sub-)annually resolved information on forest growth over decades to centuries – a temporal domain unmatched by other global observational resources. As systematic and repeated tree-ring sampling is impractical in many regions, we need to seek ways of estimating tree-ring growth where measurements are lacking. Progress has been made to simulate inter-annual tree-ring variability at the site level using process based modeling approaches (Breitenmoser et al., 2014; Li et al., 2014; Mina et al., 2016). Yet, the current generation of mechanistic models is dominated by carbon source activity (i.e. photosynthesis) and suffers key structural deficits that often result in a poor representation of carbon sink activities, such as radial tree growth (Fatichi et al., 2014; Zhang et al., in review). In view of these limitations, a valid alternative to mechanistic modeling is the statistical upscaling of existing measurements based on observed relationships with key environmental variables. Machine learning techniques have been successfully applied for such purposes (Jung et al., 2017) and we are exploring their capacity to upscale tree-ring data across the BACI domain. Within project task 4.3, we strive to develop gridded products of annual tree-ring variability that can easily be integrated with remotely sensed Earth observations and mechanistic model estimates of forest dynamics. For this purpose, we use publicly available tree-ring width data from the International Tree-Ring Data Bank (ITRDB) that has been homogenised and quality checked within Task 3.4 of this project. During pre-processing, the raw measurements have been standardized using a cubic smoothing spline detrending to remove long term trends that are related to distributing biomass around an ever increasing circumference as the tree ages. The resulting dimensionless indices retain annual to multi-decadal variability in radial tree growth. Our efforts to produce gridded tree-ring products are expected to increase the utility of tree-ring data for interdisciplinary research both within and outside BACI.

2 Approach

In order to obtain spatialized and yearly resolved tree-ring widths, we have decided to use a random forest regression approach since we also achieved promising results with this technique for upscaling diurnal cycles of carbon and energy fluxes, which is Task 4.1 of Work package 4 within this project. Random forest regression (Breiman, 2001) is a well studied machine learning technique to predict continuous outputs (such as the tree-ring widths) from a set of predictor variables. As predictor variables, we have used monthly average values of climate observations (variable names are given in Table 1) from the CRU TS dataset (version 3.22) as well as monthly average values of a water availability index (WAI) based on ERA-Interim harmonized data and WATCH meteorological forcings that has been used by Zscheischler et al. (2014). All predictor variables are available for the years 1901 to 2010 covering a large fraction of the tree-ring measurements that can be used to learn a regression model for estimating tree-ring increments from climate variables.

To obtain the predictor variables for the individual sites where we have in-situ measurements of tree-ring widths, the corresponding values have been extracted from the gridded products and the corresponding grid cells of each site. Because it is known that tree-ring growth for a single year does not only depend on climate of this specific year but also of the previous year, we have used for each annual tree-ring measurement the monthly predictor values of the current and the previous year. In addition, we also performed experiments with anomalies of the predictors by subtracting the mean seasonal cycles in contrast to using the plain values. Furthermore, we learned individual regression models for each genus to have more consistent chronologies in each training set and utilized only data from sites located in Europe. Since the available tree-ring chronologies are nonuniformly distributed among the different genera, we focused on the six most dominant genera with respect to

Table 1: Variables of the CRU TS dataset (version 3.22) that have been used as predictor variables in order to estimate tree-ring widths.

Acronym	Explanation of the variable	Unit
cld	cloud cover	%
dtr	diurnal temperature range	°C
frs	ground frost frequency	d
pet	potential evapotranspiration	mm d ⁻¹
pre	precipitation	mm
tmn	near-surface temperature minimum	°C
tmp	near-surface temperature	°C
tmx	near-surface temperature maximum	°C
vap	vapour pressure	hPa
wet	wet day frequency	d

the number of available samples (each with more than 100 sites), namely *Abies (fir)*, *Fagus (beech)*, *Larix (larch)*, *Picea (spruce)*, *Pinus (pine)*, and *Quercus (oak)*. Qualitative results are obtained using the predictions of a cross-validation scheme, more precisely a leave-one-site-out cross-validation for each genus. The quality measures that we have used to measure the performance are the Nash-Sutcliffe model efficiency (MEF), the root-mean-square error (RMSE), and Pearson correlation coefficient (PCC) between observations and predictions.

3 Preliminary results

With the plain values of the predictor variables mentioned in the previous section, we were able to obtain a model efficiency of 0.2824 for all chronologies of all genera based on cross-validation. This roughly corresponds to 28 % explained variance on average within each time series of tree-ring increments. However, results clearly differ for individual genera, which can be observed from the summary of performances in Table 2. For example, a model efficiency of 0.3263 has been obtained for *Quercus* with plain values of predictors, while the accuracy with respect to model efficiency of estimated tree-ring increments for *Larix* is only 0.1496. Thus, the quality of the predictions varies considerably depending on the genus of the trees that are considered. The same holds also for the root-mean-square error and the correlation coefficient as can be seen by comparing the corresponding values in Table 2.

Since the tree-ring increments in the chronologies are relative values with respect to an estimated age trend, we also decided to use relative values of the predictor variables. By subtracting the mean seasonal cycle of each predictor variable obtained by averaging over the years 1901 to 2000, only the resulting anomalies have been used to learn and evaluate the regression models. In the end, it turned out that the results achieved with these anomalies have been the best among all experiments. This is reflected in a model efficiency of 0.3367 among all genera and also larger accuracies for individual genera, e.g., 0.3552 for *Quercus* and 0.3761 for *Fagus* as listed in Table 2. Again, poorer results are obtained for *Larix*. This is probably due to the influence of the larch budmoth. The larch budmoth is a widespread defoliator that feeds off fresh larch needles and periodically reaches outbreak levels (Esper et al., 2007). These events are demarcated in the wood and partly obscure the obtained relationships between radial tree growth and climate.

Additionally, we have tried further prediction approaches but we could not improve the results obtained with the anomalies of the predictor variables explained in the previous section. For example, we have used both plain values and anomalies of the predictors but prediction accuracies stayed on the same level. The same holds for incorporating estimates of Gross Primary Production (GPP) at

Table 2: Cross-validation results obtained from a leave-one-site-out strategy for each genus individually and random forest regression models with MEF being the Nash-Sutcliffe model efficiency (optimal value: 1), RMSE being the root-mean-square error (optimal value: 0), and PCC being Pearson correlation coefficient (optimal value: 1).

Genus	Performance with plain values of predictors			Performance with anomalies of predictors		
	MEF	RMSE	PCC	MEF	RMSE	PCC
Abies	0.3258	0.1468	0.5809	0.3719	0.1417	0.6111
Fagus	0.3202	0.1782	0.5710	0.3761	0.1707	0.6153
Larix	0.1496	0.2048	0.3908	0.2286	0.1951	0.4806
Picea	0.3049	0.1277	0.5701	0.3544	0.1231	0.6009
Pinus	0.2321	0.1308	0.4976	0.2952	0.1253	0.5504
Quercus	0.3263	0.1362	0.5811	0.3552	0.1332	0.5981
All	0.2824	0.1451	0.5412	0.3367	0.1395	0.5835

a monthly time scale obtained from global vegetation models also known as TRENDY models (Sitch et al., 2015). Interestingly, seasonal aggregation of the monthly values leads to worse prediction performances for both setups, using plain values or the anomalies of the predictor variables. Thus, individual monthly values seem to be important for achieving more accurate estimations. The same conclusion can be drawn from our experiments in which we reduced the number of input dimensions for random forest training by applying principal component analysis to all the predictor variables. With different numbers of principal components (5, 10, . . . , 50) taken into consideration, we always achieved inferior performance compared to using all monthly values of all variables.

Furthermore, we analyzed the auto-correlation within the chronologies as well as within our predictions and the residuals. As one example, we have focused on beech trees (*Fagus*). The largest auto-correlation (about 30% on average) in the corresponding chronologies has been observed for a lag of one year, whereas larger lags show negligible auto-correlation on average as it can be observed from Figure 1. Although we have also incorporated monthly values of the previous year for each predictor variable in our cross-validation experiments, it turned out that the auto-correlation with a lag of one year is still present in the residuals of our predictions. In fact, it is not even reduced but on a comparable level. Thus, we plan to work on this issue in the next period of the project in order to improve the accuracies of the predictions. One idea is to investigate modern deep learning methods from the machine learning literature that are designed for time series predictions and directly model memory and lag effects, e.g., LSTM models (Hochreiter and Schmidhuber, 1997; Greff et al., 2015).

Encouragingly, with our current random forest regression models we are already able to obtain spatialized and yearly resolved tree-ring increments. From the chronologies of in-situ measurements at individual sites, we have learned a random forest regression model for each genus using all available chronologies of the corresponding genus. To produce first dense products, we have used a grid with 0.5° spatial resolution and extracted the predictor variables also used in our cross-validation experiments for each grid cell in Europe (improving the spatial resolution will be part of future work). Based on the regression models for each genus as well as the predictor variables for each grid cell and each year, spatialized tree-ring increments can be estimated. Of course, it only makes sense to predict tree-ring widths for grid cells where the trees of the corresponding genus occur. Therefore, we made use of the tree species distribution maps from the *European Atlas of Forest Tree Species* (de Rigo et al., 2016). The high-resolution data with grid cell sizes of 1km^2 has been aggregated to decide whether we predict the tree-ring increment for the larger grid cell at 0.5° spatial resolution. In particular, we have treated the large grid cell as a valid one for a genus if at least in one smaller grid cell of corresponding species distribution maps, the presence of the trees is at least 5%. Thus,

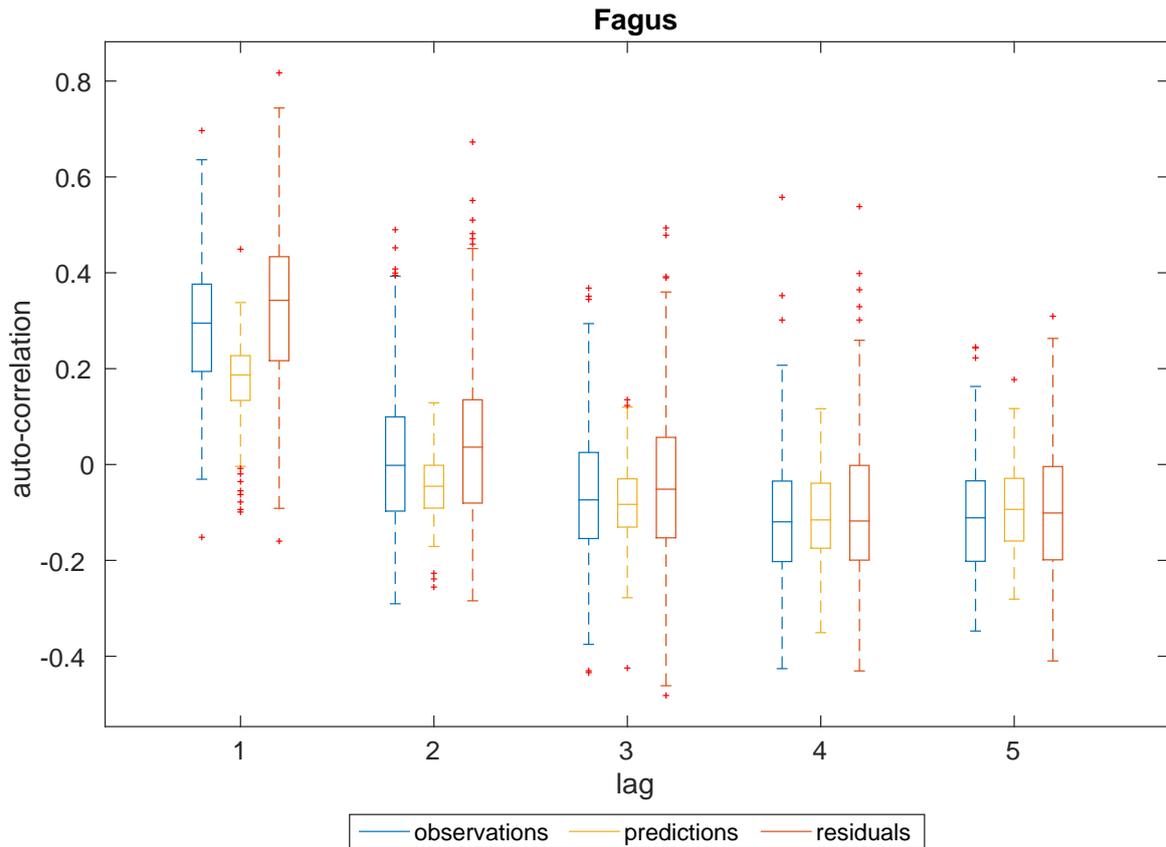


Figure 1: Auto-correlation in observations, predictions, and residuals for *Fagus* chronologies up to a lag of five years.

we could restrict our estimations to a smaller set of grid cells compared to whole Europe. Some example maps are shown in Figure 2. We have selected two genera (*Fagus* and *Quercus*) and show corresponding maps for the years 1990, 1995, and 2000. Nevertheless, appropriate maps can also be computed for other years if data for the predictor variables is available (we have climate data from the CRU TS dataset since 1901) and also for other genera. To ensure plausible results, it is recommended to select one of the six dominant genera that have been used in the cross-validation analysis to obtain a reasonable size of the training set for the regression models. The latter is also the reason why we currently work on a genus level and not on a species level, because some species are only represented by chronologies of few sites.

In addition, one can also observe from Figure 2 that trees of different genera respond differently to the same climate conditions when comparing maps of the same year, e.g., looking at the same regions in both the map for *Fagus* of the year 2000 and the map for *Quercus* of the year 2000. Thus, it is important to make appropriate distinctions on the genus or species level. We also verified this observation based on the chronologies from the in-situ measurements. In particular, we identified 57 locations in our data set where we have a tree ring chronology for both beech trees (*Fagus*) and oak trees (*Quercus*). For each pair of chronologies, we computed the correlation coefficient taking only those years into account, where measurements are available in both time series to handle the case in which one chronology ends earlier than the other one. The distribution of the resulting correlation coefficients is shown as a histogram in Figure 3. Out of the 57 pairs of chronologies, 45 pairs obtain a value smaller than 0.45, which indicates no clear correlation and at most a rather weak connection. This analysis supports our strategy of utilizing regression models from machine learning that are trained individually for each genus.

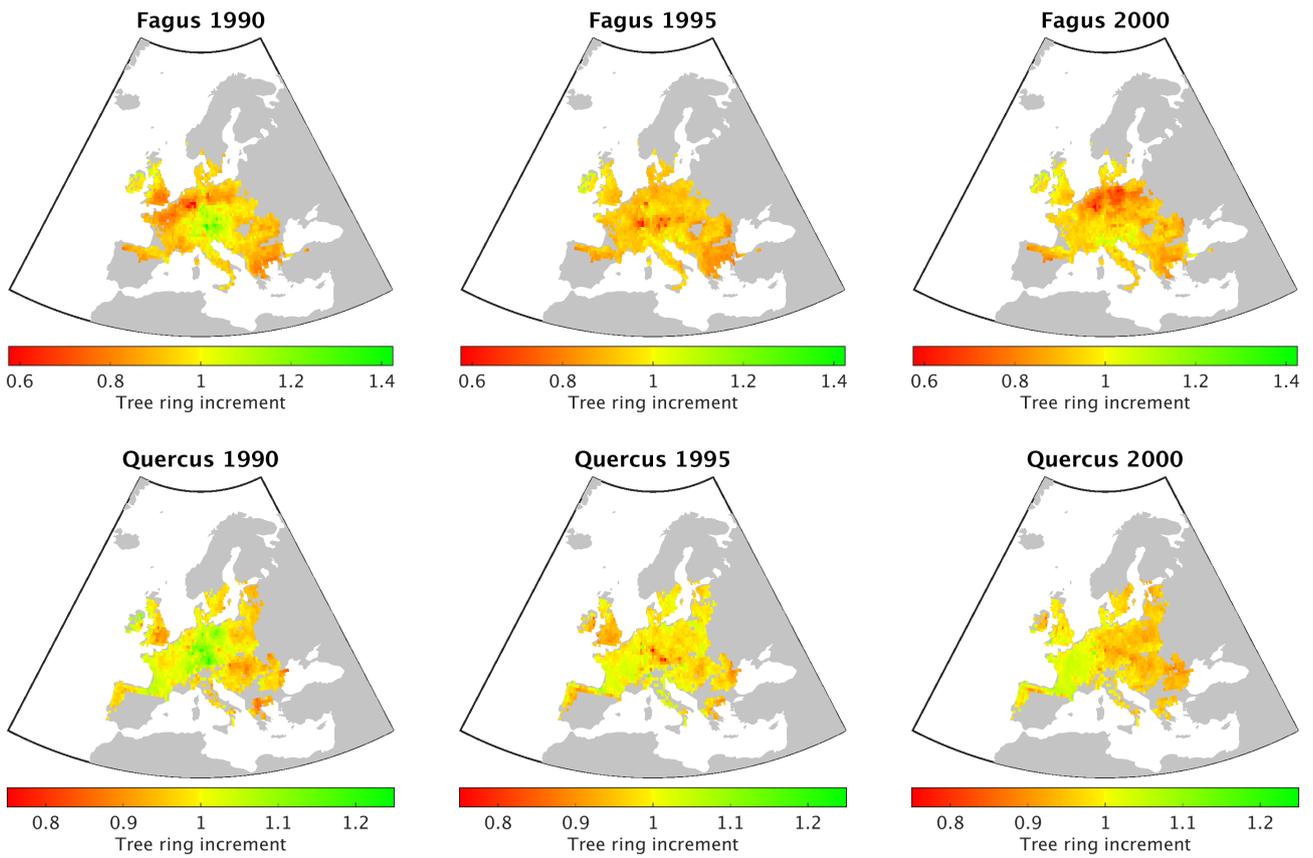


Figure 2: Estimated tree ring increments of *Fagus* and *Quercus* at 0.5° spatial resolution in Europe are visualized for the years 1990, 1995, and 2000.

Histogram of correlation coefficients for paired chronologies from *Fagus* and *Quercus* at the same site (57 pairs in total)

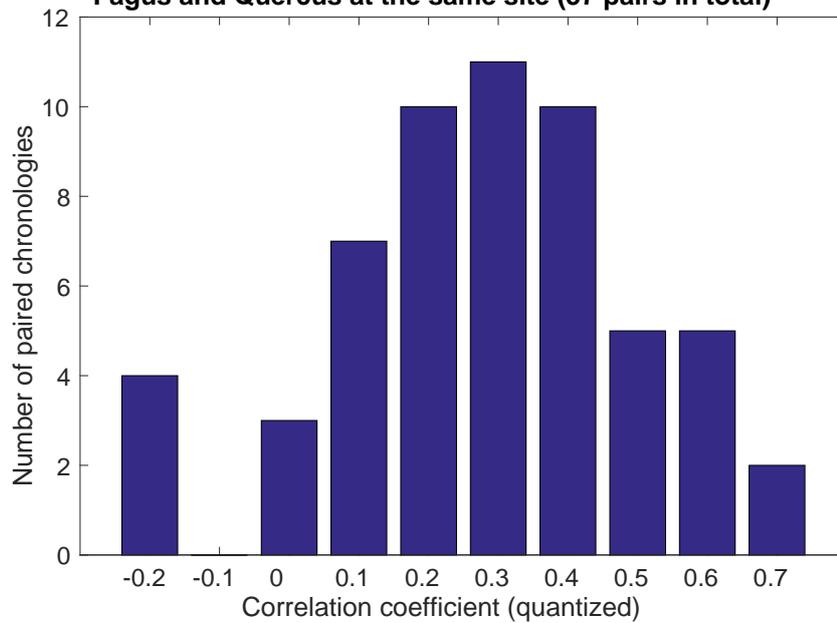


Figure 3: Distribution of correlation coefficients computed from pairs of chronologies consisting of one chronology for beech trees (*Fagus*) and one chronology for oak trees (*Quercus*) both sampled at the same site.

References

- F. Babst, B. Poulter, P. Bodesheim, M. D. Mahecha, and D. C. Frank. Improved tree-ring archives will support earth-system science. *Nature Ecology & Evolution*, 1:1–2, 2017.
- L. Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.
- P. Breitenmoser, S. Brönnimann, and D. C. Frank. Forward modeling of tree-ring width and comparison with a global network of tree-ring chronologies. *Climate of the Past*, 10:437–449, 2014.
- D. de Rigo, G. Caudullo, T. Houston Durrant, and J. San-Miguel-Ayanz. The European atlas of forest tree species: modelling, data and information on forest tree species. In *European Atlas of Forest Tree Species*. Publication Office of the European Union, Luxembourg, 2016.
- J. Esper, U. Büntgen, D. C. Frank, D. Nievergelt, and A. Liebhold. 1200 years of regular outbreaks in alpine insects. *Proceedings of the Royal Society of London B: Biological Sciences*, 274:671–679, 2007.
- S. Fatichi, S. Leuzinger, and C. Körner. Moving beyond photosynthesis: From carbon source to sink-driven vegetation modeling. *New Phytologist*, 201:1086–1095, 2014.
- K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber. LSTM: A search space odyssey. *CoRR*, abs/1503.04069, 2015. URL <http://arxiv.org/abs/1503.04069>.
- S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9:1735–1780, 1997.
- M. Jung, M. Reichstein, C. R. Schwalm, C. Huntingford, S. Sitch, A. Ahlström, A. Arneth, G. Camps-Valls, P. Ciais, P. Friedlingstein, F. Gans, K. Ichii, A. K. Jain, E. Kato, D. Papale, B. Poulter, B. Raduly, C. Rödenbeck, G. Tramontana, N. Viogy, Y.-P. Wang, U. Weber, S. Zaehle, and N. Zeng. Compensatory water effects link yearly global land co2 sink changes to temperature. *Nature*, 541: 516–520, 2017.
- G. Li, S. P. Harrison, I. C. Prentice, and D. Falster. Simulation of tree-ring widths with a model for primary production, carbon allocation and growth. *Biogeosciences*, 11:6711–6724, 2014.
- M. Mina, D. Martin-Benito, H. Bugmann, and M. Cailleret. Forward modeling of tree-ring width improves simulations of forest growth responses to drought. *Agricultural and Forest Meteorology*, 221:13–33, 2016.
- S. Sitch, P. Friedlingstein, N. Gruber, S. D. Jones, G. Murray-Tortarolo, et al. Recent trends and drivers of regional sources and sinks of carbon dioxide. *Biogeosciences*, 12:653–679, 2015.
- Z. Zhang, F. Babst, V. Bellassen, et al. Climate sensitivity of observed radial tree growth and simulated net primary productivity from ecosystem models across europe. *Ecosystems*, in review.
- J. Zscheischler, M. Reichstein, S. Harmeling, A. Rammig, E. Tomelleri, and M. D. Mahecha. Extreme events in gross primary production: a characterization across continents. *Biogeosciences*, 11: 2909–2924, 2014.