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

Deliverable 3.1: Synthesis dataset of plant traits and phenology characterized by uncertainty estimation

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:	1st April 2015
Duration of the project:	48 months
Dissemination level:	Public
Responsible of the deliverable	Jens Kattge, Julia Joswig Phone: +49 3641 576226 Email: {jkattge, jjoswig}@bgc-jena.mpg.de
Date of submission:	(V1) 30.09.2017; (V2) 18.01.2018

Summary

This deliverable describes the efforts to generate plant trait data for the purpose of “upscaling” and “interpretation” in WP4 and WP8. The scientific focus is on providing well-curated, clean, and usable data at various aggregation levels in order to have a very flexible setup for further ingestion into the scientific analyses.

Aims

The objective of this deliverable is consolidated ground based plant trait information for allowing a further integration with EO data. Plant traits mediate between the abiotic environment, and the ecosystem functioning. Therefore the abiotic environment (encoded e.g. in climate conditions or soil properties) is thought to be one of the defining factors for actual species assembly and trait distributions (Carmona et al. 2015, Kraft et al. 2015). Plant traits have shown to be related to the environment (e.g. Reich et al. 2004, Moles et al. 2014), but often only a small set of environmental variables were tested. Rarely any type of soil information was used for explaining or up-scaling plant traits, despite of the well-known relevance for explaining plant functional traits (Maire et al. 2015). In the course of the project, we identified this issue as a major gap in current research activities. In particular, if we think of the potential of the new Sentinel data a good spatial constraint of plant traits would be very helpful. Our plans have therefore shifted from seeking a phonological signature in plant traits, towards first exploring the spatial patterns and environmental covariations in-depth. The data prepared here serve primarily the goal to extract and prepare data that allow us to relate multiple environmental signals to a large number of functional traits. We extracted data of 17 plant functional traits from the TRY database (www.try-db.org, Kattge et al. 2011).

Data

We prepared a synthesis dataset of plant traits for preparing the plant trait up-scaling exercises of WP4, and more generically relevant for the characterization of ecosystem function and services. The data product is based on original trait observations from more than 150 original datasets (see original data references) compiled in the TRY database version 3.0 (Kattge et al. 2011a,b; www.try-db.org). To be able to include the maximum number of species and traits, we used the gap-filled version of TRY, with completely predicted data based on probabilistic models by a hierarchical Bayesian probabilistic matrix factorization algorithm (PMFb HPMF, Shan et al. 2012, Schrodt et al. 2015). In total, we included 156,834 individual observations from 16,855 species, 4,065 genera and 368 families. For a subset of the data that is intended to be directly used in upscaling exercises, we excluded observations that were not geo-referenced.

Specifically, the synthesis dataset comprises four individual “datasets” (or levels of data post-processing for facilitating the subsequent analyses):

1) Gap-filled trait predictions
2) Species and regional mean traits
3) Categorical traits look-up table

4) Environmental site information for geo-referenced trait observations

The underlying TRY data provide trait information for about 15% of the currently existing 350000 plant species. In combination the four datasets allow aggregating traits of individual plants to characterise the trait distributions of plant species and of plant functional types (PFTs) - based on categorical traits and/or environmental information. Spatial extrapolation to global scale is possible either in combination with range maps of plant species (e.g. MapOfLife *https://mol.org*) or PFTs (e.g. Poulter et al. 2011) and/or spatial clustering and trait environmental correlations of geo-referenced trait data.

Data Use Agreements

The underlying parts of the synthesis dataset have not been collected by the BACI project. Yet, BACI progressed these data BACI and we work in close collaboration with the original data contributors to generate a free data set by 2018. In view that the BACI team is not the original “owner” of the TRY data, also the synthesis dataset provided here can only be used in the specific context of the BACI project and must not be distributed beyond BACI. If any part of the synthesis dataset will be used in the context of BACI, a dedicated data request needs to be addressed to TRY (TRY Data Request 2435: Jens Kattge, jkattge@bgc-jena.mpg.de). Before publication of analyses based on the synthesis dataset the providers of original data sources to TRY need to be invited for collaboration and original data sources need to be cited.

Metadata, methods and processing

The synthesis dataset of plant traits is a data product based on original trait observations compiled in the TRY database version 3.0 (Kattge et al. 2011a,b; www.try-db.org). The synthesis dataset comprises four individual datasets:

1) Gap-filled trait predictions
2) Species mean traits
3) Categorical traits look-up table
4) Environmental site information

The individual parts of the synthesis dataset are related to each other via unique identifiers of traits (TraitID), species (AccSpecID) and observations (ObservationID, combining different measurements on the same entity). The unique identifiers are based on the TRY database and consistent across all individual datasets of the synthesis dataset.

The synthesis dataset of plant traits provided here will be linked to a synthesis dataset of plant phenology, to be developed in the context of BACI, via consolidated species names, geo-references and sampling date. Species names and geo-references are consolidated within the plant trait synthesis dataset. Information on sampling date and other variables is available via data request 2435 to the TRY database (request PI: Jens Kattge).

1) Gap-filled trait predictions

For preparing the synthesis of plant traits we need to be able to include the maximum number of species and traits. Given the scarcity of the original data, we used the gap-filled version, based on trait data extracted from the TRY database (version 3.0). A hierarchical Bayesian probabilistic matrix factorization algorithm (PMFb HPMF, Shan et al. (2012), Fazayeli et al. (2014) and Schrodt et al. (2015) served to predict all trait values. The gap-filled trait predictions datasets was selected and pre-processed as described in Kattge et al. (2011a) and Diaz et al. (2016).

The gap-filled data product provides two matrices of predicted trait values (one matrix with mean values and one matrix with standard deviations) for 632,938 observations of 18 numerical traits with obvious relevance for ecosystem function and good data coverage in the TRY database. Both estimates, means and standard deviations, are calculated from the probability density distribution sampled by BHPMF for each predicted trait value. The standard deviation provides the uncertainty of the mean estimate. A quality check of predicted trait values is provided by plotting root mean standard error for predicted values versus observed data against standard deviation of predicted values (Rmse_std_plot.png). This plot indicates that predicted means with small standard deviation (low uncertainty) are on average close to the observed trait records.

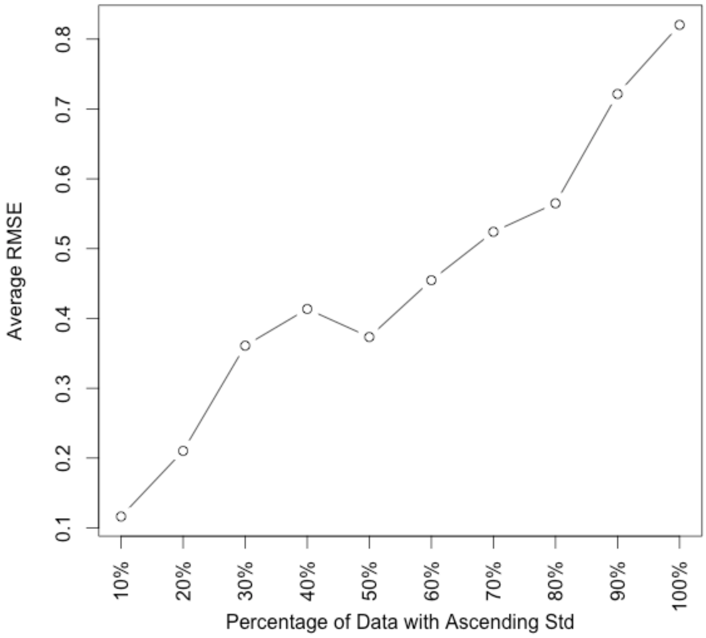

Figure 1: Average RMSE of predicted means versus observed trait records plotted against the percent of data with ascending standard deviation. This analysis indicates that the standard deviation (uncertainty of the prediction) is a good predictor for the precision of the prediction: a small standard deviation indicates high precision. (For a more detailed explanation of the plot see Fazayeli et al. (2014).)

The two datasets are provided as comma-delimited values in text format (.csv). The mean trait values are in standardized units, the standard deviations are provided in log- and z-transformation without unit.

Count	**TraitID**	**Trait Name (and unit)**
1	1	Leaf area (mm2)
2	4	Stem dry mass per stem fresh volume (stem specific density, SSD, wood density; mg/mm3)
3	11	Leaf area per leaf dry mass (specific leaf area, SLA; mm2/mg)
4	13	Leaf carbon (C) content per leaf dry mass (mg/g)
5	14	Leaf nitrogen (N) content per leaf dry mass (mg/g)
6	15	Leaf phosphorus (P) content per leaf dry mass (mg/g)
7	18	Plant height (m)
8	26	Seed dry mass (mg)
9	27	Seed length (mm)
10	47	Leaf dry mass per leaf fresh mass (Leaf dry matter content, LDMC; g/g)
11	50	Leaf nitrogen (N) content per leaf area (g/m2)
12	56	Leaf nitrogen/phosphorus (N/P) ratio (g/g)
13	78	Leaf nitrogen (N) isotope signature (delta 15N) (per mill)
14	138	Seed number per reproduction unit
15	163	Leaf fresh mass (mg)
16	169	Stem conduit density (vessels and tracheids) (1/microm2)
17	237	Dispersal unit length (mm)
18	282	Wood vessel element length (micro m)

2) *Species mean traits and aggregation*

The species mean traits database is based on trait data extracted from the TRY database (version 3.0), and selected and pre-processed as described in Kattge et al. (2011a) and Diaz et al. (2016). The dataset provides mean trait values for 48526 species and the following traits:

TraitID	**Trait Name**
1	Leaf area (mm2)
4	Stem dry mass per stem fresh volume (stem specific density, SSD, wood density; mg/mm3)
11	Leaf dry mass per leaf area (LMA, 1/SLA, mg/mm2)
14	Leaf nitrogen (N) content per leaf dry mass (mg/g)
18	Plant height (m)
26	Seed dry mass (mg)

Species mean trait values have been calculated on log-transformed data to improve normality of trait distributions and are provided here after back transformation to the original scale (Kattge et al. 2011a). The file format is an Excel file format (xlsx).

Local trait distributions can be thought of as the potential trait space (Carmona et al. 2016), that plants need to possess to survive. Our assumption is that relevant trait information to assess ecosystems is located in the main, average signal. We plan to aggregate the trait data to environmentally homogeneous regions (Olson et al. 2001) and extract the average species trait value per region. This average trait value of regional trait distributions should serve as a measure to compare across regions. Consequently we exclude the signal of neutral processes for species assembly (Chase 2014). For insuring model applicability, trait data will be log-transformed (Kattge et al. 2011).

3) *Categorical traits look-up table*

The categorical traits look-up table provides data relevant for attribution of plant species to plant functional types (PFTs). The look-up table is an extension of the TRY - Categorical Traits Dataset, which is publicly available at https://www.try-db.org/TryWeb/Data.php#3.

The categorical traits look-up table in the trait synthesis dataset provides information for 69297 species and the following traits:

TraitID	Trait	Categories
38	Plant woodiness	woody, non-woody
42	Plant growth form	tree, shrub, herb, grass
43	Leaf type	broad-leaved, needle-leaved
37	Leaf phenology	deciduous, evergreen
22	Photosynthetic pathway	C3, C4, CAM

The categorical traits look-up table is provided in Excel file format (xlsx).

4) Environmental site information

Environmental information for geo-referenced trait observations provides long term mean site characteristics for climate and soil derived from global datasets (also publicly available at the TRY website https://www.try-db.org/TryWeb/Data.php#6).

The environmental site information contains the data file (TRY30_site_climate_soil_2015_02_18.xlsx Excel 2010 format), documentation files (pdf format), pictures (png format), and a data use agreement (pdf format).

In total, we included 156,834 individual observations from 16,855 species, 4,065 genera and 368 families.

Results

First results for spatial extrapolation of relevant traits to global scale were presented at the Fall Meeting of the American Geophysical Union (AGU) December 2016 in the session 'Plant traits and biogeochemical cycles' (abstract submitted by Butler, Datta, Reich, Banerjee, Kattge et al.) and published in the context of a BACI publication in PNAS (Butler et al. 2017).

Secondly, a central result is that the species mean traits dataset is the data basis of the BACI publication by Diaz et al. (2016) in Nature. In this paper we introduced 'The global spectrum of plant form and function'. Amongst other applications, species mean traits are used to characterise ecosystem functional properties (Reichstein et al. 2014, Musavi et al. 2015) and to identify changes of ecosystem functional properties when shifts in species composition are detected along spatial gradients (e.g. Swenson et al. 2012) or in time (e.g. Newbold et al. 2015).

In BACI we thirdly envisage aggregating the data according to ecoregions sensu (Olsen et al. 2001) and understand at the global scale, how plant trait patterns are related to environmental conditions.

References:

Butler, E. E. et al. (2017). Mapping local and global variability in plant trait distributions. Proc.Natl.Acad.Sci.USA, 114(51), E10937-E10946. doi:10.1073/pnas.1708984114.

Carmona, C. P. et al. (2015). Traits Without Borders: Integrating Functional Diversity Across Scales. Trends in Ecology & Evolution. 31(5) .doi: http://dx.doi.org/10.1016/j.tree.2016.02.003

Chase, J. M. (2014): Spatial scale resolves the niche versus neutral theory debate Journal of Vegetation Science 25 , 319–322

Díaz, S., et al. (2016). The global spectrum of plant form and function. Nature, 529, 167-171. doi:10.1038/nature16489.

Fazayeli, F., et al. (2014). Uncertainty quantified matrix completion using Bayesian Hierarchical Matrix factorization. 13th International Conference on Machine Learning and Applications (ICMLA), 312-317. doi:10.1109/ICMLA.2014.56.

Kattge, J., et al.(2011a). TRY - a global database of plant traits. Global Change Biology, 17(9), 2905-2935. doi:10.1111/j.1365-2486.2011.02451.x.

Kattge, J., et al. (2011b). A generic structure for plant trait databases. Methods in Ecology and Evolution, 2(2), 202-213. doi:10.1111/j.2041-210X.2010.00067.x

Kraft, N. B. et al. (2015). Community assembly, coexistence and the environmental filtering metaphor. Functional Ecolgy. 29. 592–599. doi: 10.1111/1365-2435.12345

Maire, V. et al. (2015). Global effects of soil and climate on leaf photosynthetic traits and rates. Global Ecology and Biogeography. 24, 706–717. doi: 10.1111/geb.12296
Moles, A.T. et al. (2014). Which is a better predictor of plant traits: temperature or precipitation? Journal of Vegetation Science, 25, 1167-1180. doi: 10.1111/jvs.12190

Musavi, T. S., et al. (2015). The imprint of plants on ecosystem functioning: A data-driven approach. International Journal of Applied Earth Observation and Geoinformation, 43, 119-131. doi:10.1016/j.jag.2015.05.009.

Newbold, T., et al. (2015). Global effects of land use on local terrestrial biodiversity. Nature, 520: 7545, pp. 45-50. doi:10.1038/nature14324.

Olson et al. (2001) Terrestrial Ecoregions of the World: A New Map of Life on Earth. BioScience. 51, 933-938.

Poulter, B., et al. (2011) Plant functional type mapping for earth system models. Geosci. Model Dev., 4, 993–1010

Reichstein, M., et al. (2014). Linking plant and ecosystem functional biogeography. Proc.Natl.Acad.Sci.USA, 111(38), 13697-13702. doi:10.1073/pnas.1216065111.

Schrodt, F., et al. (2015). BHPMF – a hierarchical Bayesian approach to gap-filling and trait prediction for macroecology and functional biogeography. Global Ecology and Biogeography, 24(12), 1510-1521. doi:10.1111/geb.12335.

Shan, H., et al. (2012). Gap filling in the plant kingdom - Trait prediction using hierarchical probabilistic matrix factorization. In J. Langford (Ed.), Proceedings of the International Conference for Machine Learning (ICML) (pp. 1303-1310). Edinburgh: International Conference on Machine Learning.

Swenson N.G., et al. (2012). The biogeography and filtering of woody plant functional diversity in North and South America. Global Ecology and Biogeography 21: 798–808.