A methodology to derive global maps of leaf traits using remote sensing and climate data

Álvaro Moreno-Martínez[a,*], Gustau Camps-Valls[b], Jens Kattge[c], Nathaniel Robinson[a], Markus Reichstein[c], Peter van Bodegom[o], Koen Kramer[d], J. Hans C. Cornelissen[e], Peter Reich[f], Michael Bahn[p], Ülo Niinemets[h], Josep Peñuelas[j], Joseph M. Craine[q], Bruno E.L. Cerabolini[g], Vanessa Minden[k], Daniel C. Laughlin[l], Lawren Sack[m], Brady Allred[a], Christopher Baraloto[n], Chaeho Byun[i], Nadejda A. Soudzilovskaia[o], Steve W. Running[a]

[a] Numerical Terradynamic Simulation Group (NTSG), College of Forestry and Conservation, University of Montana, Missoula, USA
[b] Image Processing Laboratory (IPL), Universitat de València, València, Spain
[c] Max-Planck-Institute for Biogeochemistry (MPI-BGC), Jena, Germany
[d] Wageningen Environmental Research (WUR), Wageningen University, Wageningen, the Netherlands
[e] Systems Ecology, Department of Ecological Science, Vrije Universiteit, Amsterdam, the Netherlands
[f] Department of Forest Resources, University of Minnesota, St. Paul, USA
[g] Dipartimento di Scienze Teoriche e Applicate (DiSTA), University of Insubria, Varese, Italy
[h] Department of Crop Science and Plant Biology, Estonian University of Life Sciences, Kreutzwaldi, Tartu, Estonia
[i] School of Civil and Environmental Engineering, Yonsei University, Seoul, Republic of Korea
[j] Centre for Research on Ecology and Forestry Applications (CREAF), Cerdanyola del Valles, Catalonia, Spain
[k] Landscape Ecology Group, Institute of Biology and Environmental Sciences, University of Oldenburg, Oldenburg, Germany
[l] Department of Botany, University of Wyoming, Wyoming, USA
[m] Department of Ecology and Evolutionary Biology, University of California, California, USA
[n] International Center for Tropical Botany (ICTB), Department of Biological Sciences, Florida International University, Florida, USA
[o] Conservation Biology Department, Institute of Environmental Sciences, Leiden University, Leiden, the Netherlands
[p] Plant, Soil and Ecosystem Processes, Institute of Ecology, University of Innsbruck, Innsbruck, Austria
[q] Jonah Ventures, Manhattan, USA

ARTICLE INFO

ABSTRACT

This paper introduces a modular processing chain to derive global high-resolution maps of leaf traits. In particular, we present global maps at 500 m resolution of specific leaf area, leaf dry matter content, leaf nitrogen and phosphorus content per dry mass, and leaf nitrogen/phosphorus ratio. The processing chain exploits machine learning techniques along with optical remote sensing data (MODIS/Landsat) and climate data for gap filling and up-scaling of in-situ measured leaf traits. The chain first uses random forests regression with surrogates to fill gaps in the database (> 45% of missing entries) and maximizes the global representativeness of the trait dataset. Plant species are then aggregated to Plant Functional Types (PFTs). Next, the spatial abundance of PFTs at MODIS resolution (500 m) is calculated using Landsat data (30 m). Based on these PFT abundances, representative trait values are calculated for MODIS pixels with nearby trait data. Finally, different regression algorithms are applied to globally predict trait estimates from these MODIS pixels using remote sensing and climate data. The methods were compared in terms of precision, robustness and efficiency. The best model (random forests regression) shows good precision (normalized RMSE ≤ 20%) and goodness of fit (averaged Pearson's correlation R = 0.78) in any considered trait. Along with the estimated global maps of leaf traits, we provide associated uncertainty estimates derived from the regression models. The process chain is modular, and can easily accommodate new traits, data streams (traits databases and remote sensing data), and methods. The machine learning techniques applied allow attribution of information gain to data input and thus provide the opportunity to understand trait-environment relationships at the plant and ecosystem scales. The new data products – the gap-filled trait matrix, a global map of PFT abundance per MODIS gridcells and the high-resolution global leaf trait maps – are complementary to existing large-scale observations of the land surface and

we therefore anticipate substantial contributions to advances in quantifying, understanding and prediction of the Earth system.

1. Introduction

In terrestrial ecosystems, environmental conditions and biogeochemical processes both influence and are influenced by plant communities. Historical processes such as evolution, migration and disturbance shape plants from the organismal to community level (Musavi et al., 2015). At the organismal level, plant traits, which are measurable morphological, anatomical, physiological and phenological characteristics, can influence the establishment, fitness, and survival of individuals (Westoby, 1998; Reich et al., 2007; Violle et al., 2007; Homolova et al., 2013). This definition has been recently updated to encompass also responses and effects attributes at broader scales such as population, community, and ecosystem (Reich, 2014). These traits vary widely across the ~400,000 vascular plant species (http://www.theplantlist.org//), and due to acclimation and adaptation processes vary within individual species (Turner et al., 2006; Reich et al., 2007). Standard modeling and remote sensing approaches to estimate photosynthesis, evapotranspiration and biophysical parameters such as the fraction of absorbed photosynthetically active radiation (fAPAR) and leaf area index (LAI) use plant functional types (PFTs) to include plant traits within the model (Chen et al., 1999; Myneni et al., 2002; Zhao et al., 2005; Krinner et al., 2005; Mu et al., 2011; Jiang and Ryu, 2016). In so doing however, the diversity of plant communities is simplified into a relatively few categories and key variability within individual PFTs is lost (Running et al., 1994; Wullschleger et al., 2014). Subsequently, model parameters based on plant trait properties are limited by the PFT groupings, resulting in an important source of uncertainty in many biosphere models (van Bodegom et al., 2014; Reich, 2014; Reichstein et al., 2014).

In Earth system modeling, methods are being developed to improve PFT approaches, such as refining PFT categories and/or making the PFTs more spatiotemporally dynamic (Poulter et al., 2011). An alternative approach is to model the continuous spatial variability of plant traits themselves (Yang et al., 2015; Musavi et al., 2016; van Bodegom et al., 2014; Díaz et al., 2016; Madani et al., 2014). This can be done with the use of plant trait databases through establishing empirical trait-environment relationships and trait covariation (Wullschleger et al., 2014; Verheijen et al., 2015). There are a number of global traits databases containing in-situ trait observations of a comprehensive suite of plant traits for numerous species around the globe (Kattge et al., 2011; Reichstein et al., 2014; Díaz et al., 2016). These extensive databases are continually evolving and growing and provide the foundation for making broader and spatially explicit inferences of plant traits. Spatializing plant traits however, is not without substantial challenges. First, despite the large number of species included in trait databases, they are sparse compared to the overall richness and diversity of species globally (Jetz et al., 2016). Second, the large trait databases are amalgamations of many individual datasets, and contain numerous gaps. Third, the in-situ trait observations are temporally disjointed, meaning they come from a wide range of years depending on when measurements were made. Finally, these observations are made at the individual plant scale, and not necessarily representative of the variability at coarser scales.

Attempts to spatialize plant traits fall into two general categories: biogeographical and remote sensing based approaches. Biogeographical approaches attempt to extrapolate local trait measurements across different spatial scales by relating traits to abiotic factors, assuming that these factors (i.e., climate and soils) constrain the structure and function of natural ecosystems (Niinemets, 2001; Kattge et al., 2011; Reichstein et al., 2014; Díaz et al., 2016; Madani et al., 2018). For

example, van Bodegom et al. (2014) generated global trait maps by relating traits to gridded soil and climate data. Using only these environmental drivers, they were able to explain up to 50% of the global variation of plant traits. These approaches, however, do not take into account actual measured vegetation dynamics and are limited by the coarser resolution of the input data. Remote sensing approaches, on the other hand, can capitalize on higher resolution observations of actual vegetation dynamics. The estimation of plant traits from optical remote sensing is often done through physical radiative transfer models (RTMs) or empirical approaches (Haboudane et al., 2004; Mulla, 2013). RTMs attempt to explicitly define the complex interactions between the radiation and the vegetation canopy properties, these models could be inverted to retrieve biophysical variables from leaf/canopy reflectances (Jacquemoud and Baret, 1990; Dawson et al., 1998; Jacquemoud et al., 2000; Houborg et al., 2007; Stuckens et al., 2009). The combined use of RTMs with satellite data from airborne and satellite-based platforms (Liang, 2005; Baret and Buis, 2008; Berger et al., 2018) has allowed the successful retrieval of vegetation traits at different spatial and temporal scales (e.g. chlorophyll content, (Houborg et al., 2007; Zhang et al., 2005), water content, (Houborg et al., 2007; Zarco-Tejada et al., 2003), and others like leaf dry matter content and specific leaf area, (Ali et al., 2016; Feret et al., 2008)). However, applying RTMs across broad spatiotemporal extents is challenging as parameterizing RTMs across a wide range of growth forms, biomes and ecosystems is challenging (Berger et al., 2018). Furthermore, RTMs are generally based on single scene reflectance values and do not consider climatic variables that are valuable proxies for various plant traits. Alternatively, empirical approaches relating in-situ observations of plant traits to remote sensing data have been successful at mapping localized gradients of plant traits. These approaches have limited broader applications as in-situ data are often scarce or incomplete. Recent studies have combined remote sensing and biogeographical approaches (Butler et al., 2017) to obtain global maps of leaf traits at a very low spatial resolution (0.5° × 0.5° grid). The main limitation of these approaches is that, until now, they have utilized static remote sensing PFT maps for the spatialization of traits, being restricted to the simplicity of the PFTs, and not fully exploiting the full potential of optical remote sensing data (spatial and temporal variability of spectral responses), responses that can be invaluable in the estimation of key plant traits.

In this manuscript, we present and validate a combined remote sensing and biogeographic approach to spatializing estimates of key leaf traits. We integrate plant traits databases, remotely sensed data, and climatological data resulting in spatialized global maps of leaf traits at an unprecedented spatial resolution (500 m), that can be incorporated into other Earth system's models. We capitalize on the extensiveness of traits databases, the growing archive of satellite remote sensing data at multiple resolutions through time, global climatological data, and the advent of high-performance cloud computing technologies specifically designed for remote sensing applications (e.g. Google Earth Engine), combined with machine learning models for gap filling, classification and spatializing. We develop these methods for a selected set of 5 key leaf traits: Specific Leaf Area (SLA; ratio of leaf area per unit dry mass), Leaf Dry Matter Content (LDMC), Leaf Nitrogen Content per leaf dry mass (Leaf Nitrogen Concentration, LNC), Leaf Phosphorus Content per leaf dry mass (Leaf Phosphorus Concentration, LPC), and Leaf Nitrogen to Phosphorus ratio (LNPR). SLA is a key trait of the leaf economics spectrum reflecting the trade-off between leaf longevity and carbon gain (Wright et al., 2004; Díaz et al., 2016). SLA is thus indicative for different plant life strategies with respect to fast versus slow return of carbon investments (Reich, 2014). Some authors have