Remote Sensing of Environment 183 (2016) 109–119

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Optimal resolution for linking remotely sensed and forest inventory data in Europe

Adam Moreno *, Mathias Neumann, Hubert Hasenauer

Institute of Silviculture, Department of Forest and Soil Sciences, University of Natural Resources and Life Sciences, Vienna, Peter-Jordan-Str. 82, A-1190 Wien, Austria

ARTICLE INFO

Article history:
Received 27 February 2016
Received in revised form 17 May 2016
Accepted 26 May 2016
Available online xxxx

Keywords:
NFI
Europe
Resolution
Aggregation
Optimization
Forest inventory
Remote sensing

ABSTRACT

Forests provide critical ecosystem services that ensure the sustainability of the environment and society. To manage forests on large scales, spatially explicit gridded data that describes the characteristics of these forests over the entire study area are required. There have been multiple efforts to create such data on regional and global scales. This type of gridded spatially explicit data on forest characteristics are typically done by integrating terrestrial forest inventory (NFI) and satellite-based remotely sensed data. Many studies that incorporate remotely sensed data and forest inventory data often directly compare pixels to inventory plots. The standard resolution of 0.0083° is typically used to integrate these two types of data sets. There is an assumption that, when producing gridded data sets incorporating forest inventory data, the finer the resolution the better the information. This assumption may seem intuitive, however at this resolution, in Europe, each 0.0083° cell has on average 1 NFI plot, which results in a sample with 0 degrees of freedom that represents 0.02% of the cell area. In this study, we challenge this assumption and we quantify the optimal resolution with which to compare and combine remotely sensed and NFI data from the largest collated and harmonized NFI data set in Europe including 196,434 plots. We determined that aggregating data with an original resolution of 0.0083° to between 0.0664° and 0.266° (or ×8 to ×32) produces the best agreement between these two forest inventory and remotely sensed data sets, and the lowest standard error in NFI data, and maintains the majority of the local-level spatial heterogeneity.

1. Introduction

Forests provide critical ecosystem services that ensure the sustainability of the environment and society (Costanza, Fisher, Mulder, Liu, & Christopher, 2007; Richmond, Kaufmann, & Myneni, 2007). Forests are under threat of large scale disturbances and mortality due to a changing climate (McDowell & Allen, 2015; Schröter et al., 2005; van Mantgem et al., 2009). There are, however, ways that we can manage forests that can mitigate and adapt to this change (Spittlehouse, 2005). Forest management on large scales, i.e., regional or continental, requires spatially-explicit gridded data that describe the characteristics of these forests. Multiple efforts have been made to create such data on regional and global scales (Beaudoin et al., 2014; Crowther et al., 2015; Moreno, Neumann, & Hasenauer, 2016; Simard, Pinto, Fisher, & Baccini, 2011). Such data sets require integrating terrestrial and remotely sensed data which must be derived using one resolution to make the data set consistent. The resolution chosen has an impact on the quality of the output (Blackard et al., 2008; Jenkins, Birdsey, & Pan, 2001; Wilson, Lister, & Riemann, 2012). Therefore, an optimal resolution on which to link these two independent data sets should be quantified.

Gridded, spatially-explicit data on forest characteristics are derived by integrating terrestrial national forest inventory (NFI) and satellite-based remotely sensed data. Typically, these studies use forest properties measured by satellites to extrapolate NFI data across an entire study area, with the assumption that an NFI plot represents a remotely sensed data cell covering the same location. Then, a number of different techniques, such as k-nearest neighbors, are used to match similar remotely sensed cells that do not have any underlying NFI data with those that do (Beaudoin et al., 2014; Crowther et al., 2015; Simard et al., 2011).

Additionally, remotely sensed data are used to study many aspects of the global biosphere (Justice et al., 2002). Such data can be used to measure productivity, cover type, and deforestation (Hansen et al., 2013; Justice et al., 2002; Running et al., 2004). To calibrate and validate these data sets, researchers again use terrestrial empirical observations such as those obtained from NFIs (Hasenauer, Neumann, Moreno, & Zhao, 2014; Hasenauer, Petritsch, Zhao, Boisvenue, & Running, 2012; Turner et al., 2006).

There is an assumption that when producing gridded data sets that incorporate, or are compared to, NFI data the finer the resolution the better the resulting information. This assumption may seem intuitive, however, to incorporate these two types of data together, they must be comparable spatially, thematically and temporally (Tomppo et al.,

* Corresponding author.
E-mail address: adam.moreno@boku.ac.at (A. Moreno).

http://dx.doi.org/10.1016/j.rse.2016.05.021

2008). Many studies that incorporate remotely sensed data and forest inventory data often directly compare pixels to inventory plots (Crowther et al., 2015; Gallaun et al., 2010; Simard et al., 2011). A common resolution on which these studies are done is 1 km^2 (0.0083°). Fixed area NFI plots typically have areas of 200 m^2 (Tomppo et al., 2010). Therefore, 1 plot is 0.02% of a 1 km^2 cell. One sample, that represents less than 1% of the total population, results in no confidence in the sample's (forest inventory plot) ability to describe the population (remotely sensed cell). If then new datasets are generated based on this relationship, then the output may be spurious. Studies often use this one-to-one relationship to then create national, regional or even global datasets of forest characteristics (Beaudoin et al., 2014; Crowther et al., 2015; Tomppo et al., 2008). Aggregated data that incorporate more samples within a cell may lead to more accurate and realistic results.

Beyond the difference in plot size versus cell size, there are other hurdles involved in combining these two datasets across countries that are often overlooked. Methods to obtain NFI data differ across countries (Tomppo et al., 2010). Some countries use fixed area plots, which give a specific size to every sample plot. While other countries use angle count sampling, which determines which trees are counted in a sample using the tree diameter and the distance to the plot center (Bitterlich, 1952). Each sampling technique will produce a different description of a forest (Motz, Sterba, & Pommerening, 2010). Also, the country-level sampling design through space and time varies by country, from a regularly spaced grid that is the same every year, to randomized points that change every year (Tomppo et al., 2010). All of these factors affect the confidence in the NFI data at different scales. The variance in confidence through scales has an effect on the reliability of data that is produced by incorporating remotely sensed and NFI data (Seidl et al., 2013).

There are also political hurdles that hinder the use NFI data spatially. In Europe, there is no coherent NFI database from which to obtain all inventory data for the entire continent (Neumann et al., 2016). Researchers must, therefore, obtain data from each country, individually. When obtaining this data, most countries will not provide the exact location of the inventory plots out of concern for compromising the samples. These countries then provide data with a falsified location within a certain radius or within a certain grid cell (Moreno et al., 2016). The perfect link between remotely sensed data and NFI data would be to use fine resolution remotely sensed data that covers the NFI plot and only the NFI plot. However, because NFI data are given with falsified locations, this is not possible. The lack of an overarching and open NFI system or database for Europe has hindered the ability of researchers to understand how NFI data behave throughout the continent on different scales.

Additionally, inventory sample plots are meant to be used as aggregations to derive average results for a region, and not as single data points (Tomppo et al., 2010). The minimum aggregation on which NFI data is released by national organizations typically depends on the variation of the variable of interest, the sample design grid and the desired confidence interval and a single plot provides an undefined confidence interval. Therefore, to improve the confidence in NFI data and the comparability with remotely sensed data, both datasets must be aggregated until an acceptable error/confidence is reached. The optimal resolution at which to compare these two types of data and the benefits and draw-backs of aggregation are currently not quantified on the continental scale for Europe. An optimal resolution with which to combine remotely sensed and NFI data will justify a resolution that is not arbitrarily chosen based on data limitations or assumptions, but on which resolution produces the most accurate results. This will, in turn improve our confidence in the European and global scale data on forests that can be used to inform forest managers and policy makers on how best to improve forest stewardship that will benefit the environment and society, today and into the future.

In this paper, we quantify the optimal aggregation step at which to compare NFI and remotely sensed data in Europe. We accomplish this by assessing NFI data from 11 countries in Europe along with different gridded data sets. The objectives of this paper are:

1. Assess the agreement of remotely sensed data at their original resolution with NFI data
2. Quantify the loss of information with aggregation
3. Assess how aggregation affects agreement between remotely sensed and NFI data sets
4. Quantify the standard error of NFI variables at various aggregation steps
5. Determine the optimal resolution on which to combine remotely sensed and NFI data

2. Data

We use NFI data in conjunction with 4 remotely sensed land cover products.

2.1. Forest inventory data

We use national forest inventory (NFI) data from 11 European countries, Austria (Gabler and Schadauer 2008), Belgium (region Flanders) (Wouters et al. 2008), Czech Republic, Finland (Tomppo and Tuomainen in Tomppo et al., 2010), France (Nikolas et al. in Tomppo et al., 2010), Germany (Kandler, 2009), Norway (Tomppo et al., 2010), Poland, Romania (Marin et al. in Tomppo et al., 2010), Spain (Alberdi et al. in Tomppo et al., 2010), and Regional or Provincial Forest inventory (RFI, PFI) from 5 provinces in Italy (Trento, Sicily, Umbria, Piemonte and Liguria) (Neumann et al., 2016). This is currently the largest harmonized plot level forest inventory data set of Europe and includes 196,434 plots (Neumann et al., 2016). We use only data taken between 2000–2010 so as to match the same time period as the remotely sensed data products we use with no resamples in our dataset. We chose these countries because of accessibility and because they cover a latitudinal gradient throughout the continent. Datasets have been collated and harmonized by Neumann et al. (2016). The plot locations were falsified by the respective national organization responsible for the forest inventory to avoid revealing the location of the sample plots. The plot location was either altered into a random direction not to exceed more than 100 m or the plot locations were re-projected onto the center of the MODIS land cover grid (0.0083°) (Fig. 1).

Each NFI system has a different sampling density, arrangement of the sample plots and sampling method (Table 1). The sampling method has an effect on the uncertainty in results (Bergseng, Ørka, Næsset, & Gobakken, 2014; Hasenauer & Eastaugh, 2012; Hasenauer et al., 2012). Some countries in our dataset use angle count sampling (ACS) while the majority use fixed area plots (FAP) (Table 1) (Bitterlich, 1952).

Basal area factor for ACS determines the trees sampled based on their size and distance to the center of a plot. Both the basal area factor for ACS and the plot area for FAP vary in our NFI data. The size of the subplots range from 250 m^2 in Norway to almost 2000 m^2 in Spain. Six countries have their plots aligned in clusters, 5 countries as single plots on each grid point. When arranged in clusters, the number of plots varies between 2 to 18 plots in each cluster. All NFI systems use a systematic sample plot grid with constant distance between grid points. The grid distance varies by country and ranges from 0.5 km (provinces Sicily and Piemonte in Italy) to more than 10 km in Northern Finland. In Finland and Romania the grid distance also changes within the country which leads to a varying number of samples within these countries spatially (Table 1, Fig. 1).

2.2. MODIS land cover

Moderate Resolution Imaging Spectroradiometer (MODIS) land cover-type product (MCD12Q1; Hansen et al., 2002) is a global land