Ecological Indicators 81 (2017) 182–191

Contents lists available at ScienceDirect

Ecological Indicators

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

A sampling design strategy to reduce survey costs in forest monitoring

Maurizio Marchi[a], Carlotta Ferrara[b,*], Giada Bertini[a], Silvano Fares[a], Luca Salvati[b]

[a] Council for Agricultural Research and Economics, Research Centre for Forestry and Wood (CREA-FL), Viale S. Margherita 80, I-52100 Arezzo, Italy
[b] Council for Agricultural Research and Economics, Research Centre for Agriculture and Environment (CREA-AA), Via della Navicella 2-4, I-00184 Rome, Italy

ARTICLE INFO

Keywords:
ICP-Forests
Basal area increment
Forest surveys
Forest monitoring

ABSTRACT

High-quality data and long-term time series are the basis of any research activity dealing with natural resources analysis. Adequate sampling designs are fundamental to allow a robust statistical analysis to be representative of a relevant set of target variables. In this work, the sampling strategy of ICP-Forests Level II European network has been proposed to define more efficient and cost-effective procedures under the hypothesis that the average value of single-tree growth (increment) is a proxy of forest health. ICP plots have a fixed spatial structure consisting of a square of 50 × 50 m framed into 25 squared sub-plots. To estimate basal area (G) and increase over time (ΔG), two different sub-sampling methods have been implemented based on a measure of (i) the dominant layer only (i.e. a subset of the highest trees in the plot), and (ii) a random sample of squared sub-plots. While the vertical sampling procedure was performed using a progressive threshold, the horizontal sampling followed a bootstrapping procedure with random extraction without replacement. The mean absolute relative error (MARE) was used to evaluate quality of the two sub-sampling methods. Results highlighted a low predictive power with both methodologies, preventing the possibility to reduce the sampling efforts when estimating ΔG directly. In this context, the vertical sampling was strictly related to species-specific ecology, spatial structure and forest age, being influenced by vertical distribution of trees. The use of horizontal sampling for direct ΔG estimation led to systematically high errors. However, the use of horizontal sampling for total G estimation and indirect estimation of ΔG may reveal as a more effective procedure for a coherent representation of horizontal distribution of trees. Estimate ΔG as the difference between G values at time t and $t + \Delta t$ finally allows for a sensible reduction of costs with a controlled estimation error. An adequate level of MARE should be decided *a-priori* to select the number of sub-squares to be randomly sampled.

1. Introduction

Forest resources' monitoring has gained much more attention over the last decades and since Kyoto protocol (Fares et al., 2015; Patenaude et al., 2005; Schueler et al., 2014; Schulze et al., 2000), given the compelling need to assess climate change and human-related impacts on forests. Environmental assessment based on permanent experimental networks requires well-designed schemes with cost-effective sampling methods (Cochran, 1977) and acceptable accuracy (Fattorini, 2014), with the final aim to support sustainable costs-effective management of forest resources and their ecosystem services (Diehl et al., 2016; Garet et al., 2012; Kelly et al., 2015; Marchi et al., 2016; Ray et al., 2015; Salvati et al., 2016). In this view, high-quality data and low sampling efforts (i.e. monetary costs) are the two main pillars to be balanced in view of expected results and degree of uncertainty.

Dendrochronology and dendroclimatology are probably the most relevant ways to study wood formation on trees (ring widths), in relation with forest management and/or climate change impacts on forest systems (Härdtle et al., 2014; Marchi et al., 2015; Mazza et al., 2014). The most widely used proxies for forest growth and vitality are the amount of basal area and tree growth based on incremental measures (Chen and Liu, 2012; Dobbertin, 2005; Ferretti, 2013).

Long-term data from extensive surveys within environmental monitoring networks support forest research. The sampling layout is generally designed with a probabilistic approach, where all the sampling units have the same probability of extraction (Fattorini, 2014). Among the experimental networks monitoring forests and the surrounding environment, the ICP-Forests (http://icp-forests.net/) is progressively gaining attention thanks to its pan-European coverage. ICP-Forests is a European monitoring network established in 1985 under the Convention on Long-range *trans*-boundary Air Pollution (CLRTAP) of the United Nations Economic Commission for Europe (UNECE). The ICP-Forests initiative includes 42 Countries monitoring selected forest attributes in view of climate change and human-related disturbances,

* Corresponding author.
E-mail addresses: carlotta.ferrara@crea.gov.it, carlottaferrara@gmail.com (C. Ferrara).

http://dx.doi.org/10.1016/j.ecolind.2017.05.011
Received 14 December 2016; Received in revised form 2 May 2017; Accepted 3 May 2017
Available online 06 June 2017
1470-160X/

and in accordance with two different assessment schemes and analysis' levels. The extensive network (Level I) is composed of approximately 6000 observation plots and is addressed to gain information about spatio-temporal variation in forest conditions. The spatial distribution of plots was derived from a non-random sampling scheme where each plot is positioned on the nodes of the pre-existing systematic transnational 16 × 16 km grid covering the whole of Europe. The intensive network (Level II) is currently structured on nearly 500 plots in selected forest ecosystems aiming at clarifying the relationships between environmental factors and forest conditions. The monitoring network is based on a preferential design (Ferretti, 2013), a suite of case studies representative of the most diffused forest-types within each country. The layout of a Level II plot is standardized across all member states in Europe. A Level II plot is materialized on the ground as a permanent square 50 × 50 m, divided into 25 sub-squares (10 × 10 m each). While mensuration survey (i.e. diameter at breast height or dbh) is undertaken for each tree across the whole plot every 5 years, some other attributes (e.g. ground vegetation) are evaluated on a lower number of sub-squares. Similarly, the temporal resolution can vary, depending on the expected rate of change. Few surveys are based on data collected continuously (e.g. meteorology) or annually (e.g. crown conditions), others have a lower temporal resolution, ranging from 2 (foliar chemistry) to 10 years (soil chemistry).

In such a framework, the ICP-Forests network (including Level I and II) represents one of the most important sources of information for forest research at European level (Allegrini et al., 2009). Thanks to its wide spectrum of long-term data, from the air pollution to forest growth, many studies on forest health and vitality across Europe can be performed (Eichhorn, 2007).

In forest inventories, a longer sampling period or a lower sampling intensity may reduce survey costs, but may negatively affect the representativeness and reliability of data (Arabatzis and Burkhart, 1992; Westfall et al., 2016). In this paper, an exploratory analysis of forest parameters from Italian ICP-Forests Level II plots has been carried out, aiming at reducing the amount of time invested in sampling, measuring relevant variables on fewer sub-squares or less often, without a significant loss of information. The estimation error of basal area (G) per plot and of its periodic increment (ΔG) has been here tested using two different sub-sampling strategies, each of them working on the vertical or horizontal structure of 14 selected ICP-Forests plots. Based on these results, we proposed a novel sampling strategy for field works to reduce costs of operations with a controlled estimation error.

2. Materials and methods

The Italian ICP-extensive network (Level I) was established in 1985 with 243 sampling units covering the whole of forest area. Afterwards, the intensive network (Level II) was established in 1995 under the CONECOFOR (CONtrolli ECOsistemi FORestali) programme. Differently from the Level I, the sampling design of Level II followed a non-probabilistic scheme with 20 plots in total (Petriccione, 2008). Between 1999 and 2003, 11 new plots were added in northern and central Italy, including new forest categories as holm oak forests, floodplain forests, and high elevation Norway spruce and European larch forests. Consequently, the complete dataset, made by 31 permanent plots, was designed to cover a wide spectrum of case-studies, ranging from beech high forests, spruce forests, beech and oak coppice forests, and transitory crops, i.e. coppice forests undergoing conversion into high forest (Bertini et al., 2011). During the following years and between 2005 and 2015, due to increased survey costs and reduced technical staff, a lower number of plots have been gradually monitored over time. Consequently, the complete time-series is currently available for 14 plots only and the present study was performed on this data subset (Fig. 1).

The availability of the full measurement (census) on each plot allowed to compile the testing dataset with a tree-level information, derived from the most recent surveys carried out in 2000, 2005 and 2010 (during this analysis, 2015 was still under post-processing phase). Working on the basic structure of an ICP-Forests Level II plot, the diameter at breast height (dbh), the social rank, and the spatial position (relative coordinates) of each tree across the plot have been collected and used to calculate basal area (G) in each sub-square (10 × 10 m) for the above-mentioned time-spans. Then, G increment (ΔG) was derived for two time-spans (2005–2010 and 2000–2010), as the difference between surveys. These two time-slices were selected to evaluate the two sampling procedures stability across time (i.e. 5 and 10 years), and to test a possible reduction of sampling efforts and costs. In particular, ΔG estimation has been performed by (i) measuring fewer trees every 5 years, i.e. the canonical period of ICP surveys, and (ii) measuring fewer trees at a wider time-scale (10 years). These sub-sampling hypotheses were tested working on the vertical and horizontal forest structure.

The first approach was developed under the hypothesis that the dominant layer (i.e. the vertical storey made by the tallest trees) concentrates most of wood increment (basal area, tree height, tree volume). This assumption is based on a well-known ecological dynamics stating that, in case of competition, tree physical prevalence (tree height, crown development) corresponds to a different growth trend (Becagli et al., 2013; Cook and Kalriukstis, 1990; Vanderwel and Purves, 2014). A high proportion of the total G increment was expected to be concentrated in a reduced amount of trees, depending on the vertical structure. The sampling procedure was applied using a decreasing tree height threshold that includes a progressive higher number of trees in the samples up to the total. For the tested intervals, this sample was used to evaluate the amount of ΔG for each plot, estimated not at the tree-level but rather at plot-level. That is, for a given proportion of trees sampled, there is no requirement that trees sampled at time t be the exact same trees sampled at time $t + \Delta t$. Actually, unequal height growth rates can change the height rankings by which trees are selected for sampling. With this procedure, the total G was not analysed due to obvious result connected to the sampling theory (i.e. the amount of G included within the selected trees will always be proportionally lower than total G). The purpose of this first method was to identify whether fewer trees could be measured across the plot spatial structure to obtain an unbiased estimation of total ΔG in the plot (tree-level approach). In case of an effective costs' reduction, a height threshold could be quantified as a proportion of total height of the tallest tree in the plot.

The second method was built on the plot horizontal structure, given the mandatory geometric structure of the ICP-Forests plot. G and ΔG were estimated by means of a bootstrapping procedure with random extraction without replacement. This procedure was implemented using an increased number of sub-squares from 1 to 24 per tested plot. To derive the amount of G and ΔG for the whole plot, an expansion factor w was used, calculated as:

$$w = \frac{25}{n}$$

where n is the number of randomly sampled sub-squares. To avoid calculation biases due to random extraction across a variable horizontal structure (i.e. unequal number of trees and G per sub-squares), the procedure was repeated 100k times for each n value, averaging the results. This second method aimed to define a relation between ΔG (or G) expected estimation error and the sub-squares number to be measured. The estimation quality was assessed by means of Mean Absolute Relative Error (MARE) and standard error (SE), as follows:

$$MARE = \frac{|\gamma' - \gamma|}{\gamma} SE = \frac{\sigma}{\sqrt{(n)}}$$

where γ' and γ represent the estimated and measured G or ΔG values, σ and n are respectively the sample standard deviation and size. To analyse and model the relationship between MARE and the sample size to be used (trees of sub-squares), three non-linear models were tested taking into account G or ΔG coefficient of variation (CV) within vertical