Investigation of sampling design and growing stock volume estimation in the Greek managed forests

Aristeidis Georgakis
Faculty Agriculture, Forestry and Natural Environment, School of Forestry and Natural Environment, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece, e-mail: arisgeorg@for.auth.gr

Maria J. Diamantopoulou
Faculty Agriculture, Forestry and Natural Environment, School of Forestry and Natural Environment, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece, e-mail: mdiamant@for.auth.gr

Marios Trigkas
Department of Forestry, Wood Sciences & Design, University of Thessaly, V. Griva 11, 43100, Karditsa, e-mail: mtrigkas@uth.gr

Abstract
Forest inventories (FIs) are carried out through the sampling of natural resources. Fis provide the necessary information, through the estimation of the forest-biometrical variables, for sustainable forest management. The main quantitative forest-biometrical variable that receives the greatest interest, is the growing stock volume (GSV). The Greek Forest Service, which supervises the state forests, faces the big challenge of the small area estimation (SAE) of the GSV in the managed forest stands. Motivated from the above problem, a questionnaire survey was conducted and effort was expended to record and investigate the design phase of sampling methods and the estimation phase of GSV of the forest stands/compartments.

Keywords: Questionnaire, sampling design, sample plots, forest stands, estimation phase, inventory cost

JEL classifications: C44, Q56, L73

Introduction

Forest inventories are confronted as a key source for sustainable forest management. Based on the application of the sampling methods (sampling design) applied and the volumetric variables estimations, forest inventories can be separated into two categories: a) national inventories that provide information on several qualitative and quantitative characteristics at the country and regional level and b) local management forest inventories (MFIs). The latter provide quantitative and qualitative information for the forest stands or compartments (management units). Basic quantitative forest-biometric variables of interest are growing stock volume (GSV), volume growth, diameter & volume distributions, and tree density. FIs begin with the field measurements of the sample plots and finally end with the estimation phase of the variable of interest. These estimates are calculated periodically (usually every ten years).

The first step for the accomplishment of FIs is the application of the sampling theory which includes, the determination of the area of interest (forest stands), the target variable (GSV), the sampling design, the sample size determination, and the selection of the optimum sample size and shape of the sampling units (plots). Through the sampling design, field measurements are taken that lead to volume estimations. The ultimate goal of the sampling design in the FIs is the quantitative estimation of the most valuable forest-biometric variables, such as the GSV. These variables can provide all the necessary information for the sustainable management of the forest and forest stands (Matis 2004a, Matis 2004b, Köhl, et al., 2006, Kershaw Jr et al., 2016).

Recently, research focuses on the small area estimation (SAE), which aim to provide estimates of the variable of interest, for subpopulations (small geographical areas) consisting of a small (or sometimes null) sample size (Goerndt et al., 2011, Breidenbach and Astrup 2012, Rao and Molina 2015, Mauro et al. 2017, Breidenbach et al. 2018, Georgakis 2019, Georgakis and Stamatellos 2020). These "small estimates" use estimators or models that utilize different types of auxiliary information from the forest (spatiotemporal, remote sensing). A basic assumption of the auxiliary variables is the linear correlation with the variables of interest.

The Greek MFIs are observed to face problems regarding the consistent application of the scientific principles of survey sampling. On the one hand, one reason for this problem could be the enforcement of the technical specifications guide published by the Greek Ministry of Agriculture (Ministry of Agriculture 1953, 1965), which provide tolerance to the specific requirements of inventories, making the MFIs more or less "subjective" to human expertise. The forest service, which is responsible to supervise FIs and manage forest ecosystems, faces additionally to these tasks several problems and challenges, such as understaffing, the accurate determination of the ownership of the forest land on maps, etc. The above issues, combined with the lack of adequate funding (Papadopoulos et al. 1998), often lead to qualitative degradation of management plans and therefore may cause problems in forest sustainability.

In this research, an effort is being made to reflect the valuable experience and practice of forest managers and forest practitioners concerning the Greek MFIs. An appropriate tool for this purpose was the questionnaire that was designed to obtain from respondents their practical experience. The analysis of the results aims to record the existing working method of the managers/ practitioners and point out relevant problems encountered in forestry practice, regarding the collection of the ground-truth data and the estimation phase of the GSV.

Materials and Methods

The questionnaire was selected as an appropriate and effective research tool for the easy and quick extraction of the necessary information (Bradburn et al. 2004, Brancato 2006) of FI practice. The research population concerns both the foresters from the forest service and the freelances who are working in the field by setting up sample plots in order to assess the GSV of the Greek forests. The respondents were approached by a) internet correspondence (by sending an email), b) by personal communication (in person and by telephone communication), and c) posting the questionnaire on a well-known website dealing with forest and environmental issues (https://dasarxeio.com).

The questionnaire consists of 30 questions/surveys. The first 29 belong to the broad category of closed-ended, including a) the Likert scale (Likert 1932), b) multiple choice answers and d) yes/no answers. The last question was open-ended and gave to the participants the opportunity to express their personal views and suggestions. The questionnaire is divided into three thematic sections, the questions include a) the determination of the problem, b) the sampling design used in FI (shape and size of sample plots, sampling intensity & sample size, instruments) and c) the estimation phase procedure of the GSV.

A total of 61 replies were received from which after a coherent examination of the questionnaire three replies have not met the requirements and thus, excluded from the analysis. Finally, 58 answers were used in the analysis, which reflects the maximum frequency of answers per question. The use of the Likert scale covered the majority of questions and was used to rank the agreement or disagreement to a statement. Five-choice Likert response options were used from 1-5, 1. Not at all (totally Disagree), 2. Little (Disagree), 3. Enough (Neutral), 4. Very (Agree) and 5. Very much (Strongly agree). The internal validity of the questionnaire was tested by detecting any systematic errors. The validity of its content was tested by five expert practitioners by calculating the validity ratio of its content. In addition, it has been answered by practice experts from all over the population. Finally, the reliability of the questionnaire was checked by implementing an internal coherence test by a group of questions and in aggregate using the Cronbach's Alpha test (Cronbach 1951).

The current frequentist analysis is based on the mode of the most frequent replies. When there was more than one mode, the maximum one was selected. The mode was considered to provide more meaningful information over the average or median measurements of the central tendency. The different number of replies per question led to the use of the absolute and relevant frequencies of the answers. The answers are recorded automatically in the encoded format of spreadsheets (Google Sheets). The data was processed and analyzed through the IBM SPSS 2019 and the Microsoft Office Excel 2019 packages. The Likert scale graphs were produced with the "likert" library (Bryer et al. 2016) of the R statistical language (RCore Team 2021). For the analysis of the Likert results, the stacked bar plot was the preferred method for the presentation of the results. The five options are separated from the middle (third) which acts as a divider between the extreme "left" (1. Not at all, 2. Little) and the extreme "right-wing" (4. Very and 5. Very much) responses. The mid-range value carries the relative frequency in the responses, while the two other frequency values are provided on either side of the middle option/response. If the middle (third response) is not considered neutral, then the medium response can be given positive or negative weight, accordingly.

Results

Questionnaire analysis and validation

The period of data collection through the questionnaire took place from June to July 2020. The 58 responses (n=58) from participants, of which 60% (n=35) of them were self-employed foresters, 33% (n=19) were foresters in a forest service and the 7% (n =4) were foresters-researches working at a university or research institution of the country. Out of the 13 administrative regions of Greece, 10 (n= 58) participated. Central Macedonia and Thessaly contributed significantly to the survey with 36% and 19% respectively of the total responses *(*Figure 1*)*. The questionnaire was checked by five expert practitioners ($N = 5$), with a fraction of validity of content: $(n_i-N/2)/(N/2)$ where n_i is the number of specialists that characterized the questions as "necessary" equal to one since all questions were considered "necessary" by the experts. The questionnaire has high external validity and thus provides the possibility of generalization. Reliability analysis of the questionnaire was applied through the internal coherence test for each group of questions and the whole questionnaire. The test gave a value of the Cronbach’s Alpha test, a = 0.924 which indicates a "very high" reliability of the questionnaire. When the Cronbach’s Alpha test, a, was calculated per group of questions, it gave a value of 0.877 and 0.886 (high reliability) for the groups of "General Questions" and the "Inventory Method - Sampling Plan", respectively, while it gave low reliability for the question group of the "stand/compartment volume estimation". An interpretation that could be given is the existence of a great homogeneity and "tendency" for the first two groups, while in the third group there is a great heterogeneity of responses.

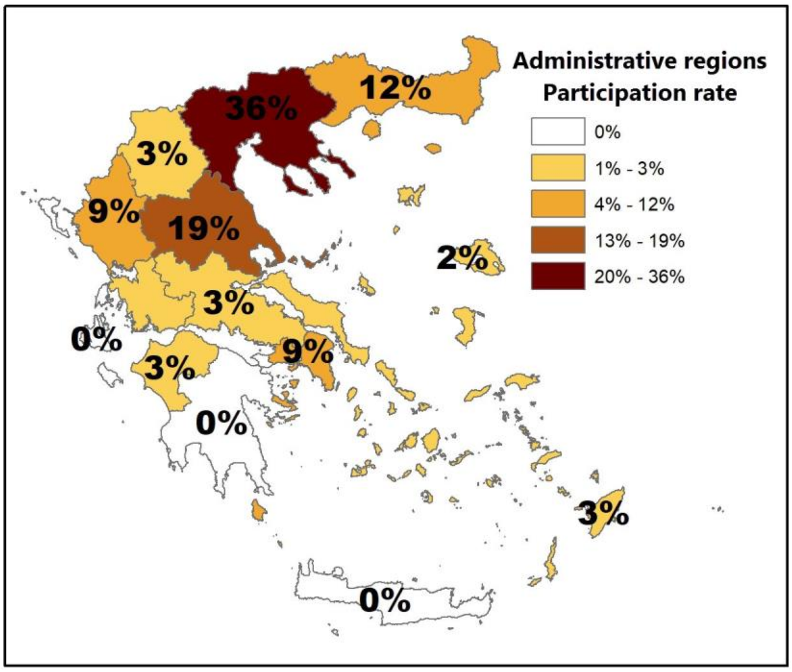

Figure 1: Participation rates by region (respondents' workplace)

Problem identification

The first question tried to identify the practitioner's experience about the current method of FIs (sampling design, field data collection, and estimation of GSV) and their suggestion for potential improvements. The analysis of the survey suggests that the currently applied FI methods should be improved. More specifically, 97% of respondents chose *Enough - Very much*, and 67% of them chose *Very - Very much* (n=58 replies). The GSV remains a major forest variable for estimation, after a relevant question about the most important variables in the FIs. The last relevant thematic question, "*Where do you propose to give more research weight? In the estimates of the GSV at forest stand level or in the estimates of the GSV at the sample plot level?*" the respondents show the priority for accurate estimates of GSV, constituting a strong indication that the main problem of FI located on the estimation phase (n = 58 replies).

FI method - Sampling design

The existing protocol of field data collection, according to a relevant question, seems to provide accurate estimates of the GSV at the sample plot level, while it implies that the estimation problem is concentrated in the management units (forest stands). The following Figure 2 reveals the problem of subjective sampling methods (category *B*) preference on the contrast with statistically sound methods (category *A*). The majority of forest practitioners prefer both the circular and the rectangular shape of the *sample plot*. Also, the vast majority (*Very - Very much*, 81%) choose sample plots of 0,1 Ha and sometimes combination with 0.5, 1.5, or 2 Ha.

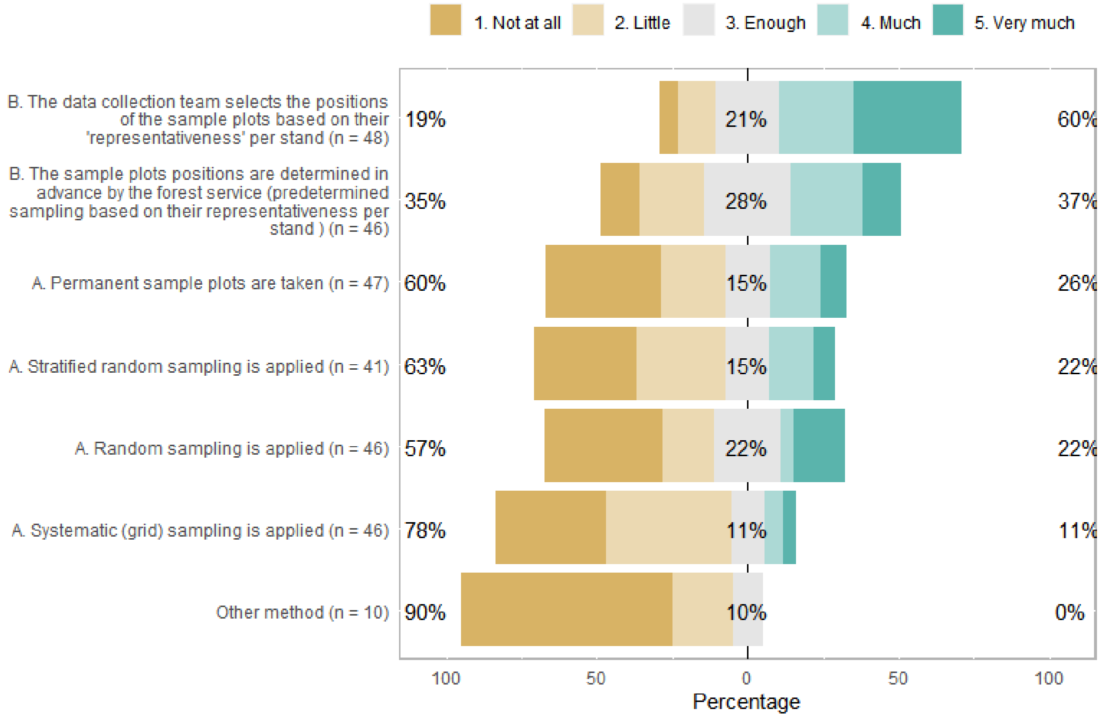

Figure 2: Sampling in forest practice

Another important finding of this research reveals the small usage of Bitterlich plots, 52% (n=58) of practitioner does not use at all or little (12%). This should be further investigated, knowing already very well the great efficiency of estimating the basal area of the trees just by counting them under a certain basal area factor. *Remote sensing data* indicate that can help to allocate the sample plots visually, the majority of 89% (n = 57) of participants use *aerial photographs and satellite images* for this purpose. Furthermore, 81% (n=58) of practitioners replied that record the geographical coordinates of the sample plots centers. These records may be proved very beneficial, particularly if the positional accuracy is smaller than 5 meters radius because FIs can be enhanced from the related RS information and by the use of model-based of model-assisted estimates. Findings indicate that all the sampling designs in MFIs include a form of stratification because only the purely forest areas have a probability to include sampling units. Lastly, for this thematic topic, the respondents do not see it as a cost-effective solution to reduce the sample plot size by half and duplicate simultaneously the sample size. The highest inventory cost concerns the different plot locations and not so much the size of the plots.

The estimation phase of GSV in the MFIs stands

In the estimation phase, different methods for the estimation of GSV are used by the practitioners. The two most important estimation methods that closed were a) the estimation of GSV from the mean of the sample plots and the extrapolation to the forest stand level, and b) by giving different weight on the sample plot "participation" of the forest stand. Both of them face the problem of the extremely small sample size, usually one or two sample plots per stand. Additionally, the last one is subject to personal experience about the "weights of participation" in the estimation phase. Finally, when serious deviations in the estimation of the GSV occur, then the next action is the inspection of the accurate position of the sample plots and secondly, new sample plots are set up to moderate larger deviations from the expected estimations.

Discussion

This research allowed the foresters who work on the MFIs, to describe the current situation and express their opinions, via questionnaire, on the running sampling design and estimation procedure based on their own experience (Bradburn et al. 2004; Brancato 2006). The results indicate that particular emphasis is needed to be paid in the estimation phase of forest biometric parameters, such as GSV, at the forest stand level, while the measurements of the sample plots are not considered as a key problem. However, it is important to accurately record the sample plot variables and their locations.

Both the questionnaire analysis and the forest practitioners identified the existing problem of the GSV estimation. Important errors can be introduced in the estimations because of the subjective selection of FIs methods. In forest practice, a subjective way of obtaining the sample plots (Koivuniemi and Korhonen 2006) and estimating the GSV of forest stands usually is preferred. The main reasons leading to this approach are the relative freedom in how to choose the location of the sample plots, the small sample size suggested by the sampling design and the uppermost reason is the subjective way of estimating or extrapolating the sample plot data in the level of the forest stands. All of the above can be improved with additional funding, modernization of protocols (technical specifications), and the centralized supervision of FIs.

The forest inventories, up until recently, were mainly based on the exclusive selection and implementation of an appropriate sampling design for a specific type of forest (Matis 2004b, Köhl et al. 2006, Kershaw Jr et al. 2016). The modern trend of estimating forest parameters includes also the utilization of three-dimensional remote sensing data (crown/tree height from LiDAR or photogrammetry) and the use of statistical models, such as SAE models (Goerndt et al. 2011, Breidenbach and Astrup 2012, Rao and Molina 2015, Mauro et al. 2017, Breidenbach et al. 2018, Georgakis 2019, Georgakis and Stamatellos 2020). The above methodology applied extensively to the national FIs but gradually can be applied in the MFIs as well.

Funding
"This research has been financially supported by the General Secretariat for Research and Technology (GSRT) and the Hellenic Foundation for Research and Innovation (HFRI) (Scholarship Code: 1319)"

Acknowledgments
From this position, we express our warm thanks to all participants for the valuable time they spend. Warm thanks to our fellow foresters freelancers for their significant participation in research, to colleagues who serve in the forest service, as well as researchers from universities/research institutes in our country.

References

Bradburn, N., Sudman, S. & Wansink, B. (2004). Asking Questions, The Definitive Guide to Questionnaire Design - For Market Research, Political Polls and Social and Health Questionnaires. San Francisco: Jossey-Bass.
Brancato, S., Macchia, M., Signore, G., Simeoni, G., Blanke, K., Koerner, T. et al. (2006). Handbook
Breidenbach, J., & Astrup, R. (2012). Small area estimation of forest attributes in the Norwegian National Forest Inventory. European Journal of Forest Research, 131(4), 1255-1267. doi:10.1007/s10342-012-0596-7
Breidenbach, J., Magnussen, S., Rahlf, J., & Astrup, R. (2018). Unit-level and area-level small area estimation under heteroscedasticity using digital aerial photogrammetry data. Remote Sensing of Environment, 212, 199-211. doi:https://doi.org/10.1016/j.rse.2018.04.028
Bryer, J., & Speerschneider, K. (2016). Package 'likert'.
Cronbach, L. (1951). Coefficient Alpha and The Internal Structure of Tests. Psychometrika, 16(3), 297-334.
Georgakis, A. (2019). Small Area Estimation in Forest Inventory. Το άρθρο εντάχθηκε στα πρακτικά του Seventh International Conference on Environmental Management, Engineering, Planning and Economics (CEMEPE 2019) And SECOTOX Conference, Mykonos island, Greece. http://cemepe7.civil.auth.gr/
Georgakis, A., & Stamatellos, G. (2020). Sampling Design Contribution to Small Area Estimation Procedure in Forest Inventories. Modern Concepts & Developments in Agronomy, 7(1). doi:10.31031/MCDA.2020.07.000654
Goerndt, M. E., Monleon, V. J., & Temesgen, H. (2011). A comparison of small-area estimation techniques to estimate selected stand attributes using LiDAR-derived auxiliary variables. Canadian Journal of Forest Research, 41(6), 1189-1201.
IBM Corp. Released (2019). IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp
Kershaw Jr, J. A., Ducey, M. J., Beers, T. W., & Husch, B. (2016). Forest Mensuration (5th ed.): John Wiley & Sons.
Köhl, M., Magnussen, S., & Marchetti, M. (2006). Sampling methods, remote sensing and GIS multiresource forest inventory: Springer, Berlin, Heidelberg.
Koivuniemi, J., & Korhonen, K. (2006). Inventory by Compartments. In (pp. 271-278).
Likert, R. (1932). A technique for the measurement of attitudes. Archives of Psychology, 22, 5-55.
Matis, K. (2004a). Forest Biometrics: II. Dendrometry (in Greek) (2 ed. Vol. 2). Thessaloniki, Greece: Pegasus.
Matis, K. (2004b). Sampling of Natural Resources. Thessaloniki, Greece: Pegasus.
Mauro, F., Monleon, V. J., Temesgen, H., & Ford, K. R. (2017). Analysis of area level and unit level models for small area estimation in forest inventories assisted with LiDAR auxiliary information. PloS one, 12(12), 14. doi:10.1371/journal.pone.0189401
of Recommended Practises for Questionnaire Development and Testing in the European Statistical System. ISTAT.
R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/index.html
Rao, J. N., & Molina, I. (2015). Small area estimation: John Wiley & Sons, Inc.