Environmental Modelling & Software 73 (2015) 14–26

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

On the site-level suitability of biomass models

César Pérez-Cruzado [a], [*], Lutz Fehrmann [a], Paul Magdon [a], Isabel Cañellas [b], Hortensia Sixto [b], Christoph Kleinn [a]

[a] *Chair of Forest Inventory and Remote Sensing, Georg-August-Universität Göttingen, D-37077 Göttingen, Germany*
[b] *Selviculture and Forest Management Department, CIFOR-INIA, E-28040 Madrid, Spain*

ARTICLE INFO

Article history:
Received 17 March 2015
Received in revised form
16 July 2015
Accepted 31 July 2015
Available online 12 August 2015

Keywords:
Biomass and carbon inventory
Accuracy
Estimation uncertainty
Model localization
Monitoring

ABSTRACT

Tree biomass estimates in environmental studies are based on allometric models, which are known to vary with species, site, and other forest characteristics. The UNFCCC published a guideline to evaluate the appropriateness of biomass models before application, but it misleads the concept of model suitability and does also allow the selection of models with systematic deviations in the predictions. Here we present an alternative approach based on non-parametric techniques. The approach was tested for pure stands, but this methodology is likewise applicable to mixed forests. The proposed tests perform well in rejecting a model if the predictions for the targeted population are systematically deviant. It is demonstrated that the suitability of an allometric model is a matter of accuracy. The proposed method also allows localizing the model. The presented approach can improve the transparency of global forest monitoring systems and can be implemented with relatively small effort.

1. Introduction

Forest biomass has developed towards one of the most relevant target variables for forest monitoring, especially in context of reporting requirements on the national and international level. The program on Reducing Emissions from Deforestation and forest Degradation (REDD+) is but one example showing that the estimation of biomass and carbon stocks is an essential basis for results-based payments when countries or projects shall be compensated for their efforts in avoiding/reducing deforestation and forest degradation (Gibbs et al., 2007). The estimation of forest biomass stocks and their changes for large areas is a complex task that usually involves forest mensuration, statistical sampling in the context of forest inventories, statistical modelling and remote sensing analysis.

As a target variable, biomass cannot be directly determined in the field. Thus, allometric biomass models are an indispensable core component of carbon accounting systems. Developed based on a number of destructively sampled trees, they establish the relationship between measurable variables like tree diameter (usually measured at breast height, *dbh*), sometimes tree height and the dry weight. A general recommendation for the required sample size allowing stable estimates of the model parameters is 50 trees (Sileshi, 2014), even though other authors pointed out that, in the case of large area volume estimates, at least 100 trees would be necessary to reduce the effects of model prediction uncertainty on large area estimates (McRoberts and Westfall, 2014). However, this is also dependent on the given variability in the tree population under study. Given that destructive sampling for biomass is a time consuming and expensive exercise, many published models are based on relatively small datasets. The large costs for destructive sampling also restrict the model validation – as part of the model building with independent data from the same site – which requires additional samples. Therefore model validation with an independently selected sample taken from the same site is rarely found in the context of biomass assessments, making cross-validation by using re-sampling techniques the most common way of solving this problem (i.e. Vanclay and Skovsgaard, 1997).

Depending on the scope of the study and the datasets used, the biomass models are built either for specific tree species (Pérez-Cruzado and Rodríguez Soalleiro, 2011), tree species groups (Muukkonen, 2007), specific forest types (Brown and Lugo, 1992), site conditions (Ketterings et al., 2001), geographic regions (Fehrmann and Kleinn, 2006; Shaiek et al., 2011), and in the case of forest plantations also for specific genetic varieties (Tuskan and Rensema, 1992). However, the variability of different conditions particularly in large area forest monitoring and the large costs of destructive sampling lead to the widespread use of generalized

---

[*] Corresponding author.
E-mail address: cperez@gwdg.de (C. Pérez-Cruzado).

http://dx.doi.org/10.1016/j.envsoft.2015.07.019

models (Chave et al., 2005), such as the biomass models per forest type listed in IPCC guidelines for LULUCF projects (Watson et al., 2000).

If forest biomass shall be estimated from individual tree data collected from a forest inventory, the identification of a suitable model from the literature is a major source of uncertainty (e.g. Ketterings et al., 2001; Chave et al., 2004). Predictions from different models, even if developed for the same species, species group or geographical context, might produce significantly different estimates. In the case that various published candidate models are available, there remains the question of how to select the most suitable one for the specific conditions of a particular study. Model evaluation is an indispensable step in the iterative process of developing environmental models (i.e. Denzer, 2005; Jakeman et al., 2006), but this issue has traditionally been addressed by the model builder's instead of the end-user's point of view.

As the reported fitting statistics of a model, (e.g.: the coefficient of determination R^2, the root mean square error *RMSE* or the Akaike Information Criterion *AIC*) provide information on how well the model fits the data that were collected for model development, they consequently refer to the underlying tree population exclusively and do not allow any conclusions about the suitability of the same model when applied to a different population. Even for the same population, adequate estimates are not necessary obtained when the model is applied at different scale than the one considered in the developing phase (Gertner, 1984). It is, therefore, not surprising that these general fitting statistics were found to be incomplete or tested with inadequate model diagnosis in many cases (Sileshi, 2014). Nevertheless, other studies exclusively evaluated the contribution of model uncertainty to the final overall precision of volume estimates (McRoberts and Westfall, 2014) and conclude that this model uncertainty is negligible compared to the simple random sampling error in observational studies when sample size is bigger than 100 trees. However, there is agreement that the model selection might have a huge influence on the final estimates.

In order to reduce this uncertainty induced by the model selection on a statistically sound basis, an evaluation of the suitability of the model in the target population is indispensable. This implicitly assumes that the quality of a given model does not only depend on the quality of its production, but also on the characteristics of the population in which it should be applied. Similar to a model validation that is usually performed in the context of model development, which itself is based on independently selected samples from the same source population (Jakeman et al., 2006), the evaluation of the suitability of an existing model against a different target population would require sampling additional trees. Even if this is related to additional costs, it may well be justified if the accuracy of biomass estimation and the transparency of the employed methodology are to be improved. The model that is finally applied to the individual tree data collected in the field affects the subsequent plot-level estimates, through which it also affects all map products created by combining field and remote sensing data used to predict biomass on larger areas in the form of wall-to-wall maps.

A validation sample from the local population may not only be used to evaluate the suitability of available biomass models, but could also serve to adjust models and thus improving the fit of the model to the local population. Examples on such “model localization” approaches can be found for growth and yield models (Leary and Hamlin, 1987; Zhang et al., 1997; Martin, 2001). The topic of spatial dependence of localization parameters of large areas models to local conditions was recently addressed by using different techniques (Räty and Kangas, 2007, 2008, 2010; Räty et al., 2011), but the main problem of these techniques when applied to REDD+ projects is the lack of observational studies in the implementation countries. Studies on model localization to individual tree biomass models by using independently gathered destructive samples appear to be scarce. Local biomass data are usually taken to fit an entirely new model instead of trying to utilize it for localizing existing models.

There are, to the best knowledge of the authors, no references in the scientific literature on forest biomass monitoring that give an indication of the minimum size of the local dataset that is required to perform a statistically reliable test on the suitability of a model in an “unknown” population. In an approved IPCC methodology (UNFCCC, 2011) for practical application, a sample size of $n = 10$ trees is recommended that shall cover the *dbh* range of the local population, however, the scientific bases of this recommendation remains unclear. The references in the “grey” literature in this respect are also scarce, even for the cases where it has an evident practical applicability: the MRV (Measurement, Reporting, Verification) process within REDD+ projects. As most of the ongoing REDD+ projects are bilateral agreements between the donor and the implementation countries, the technical documents of MRV are not usually available for the broad audience. An exception on that is the MRV process of the REDD+ program in Guyana (Brown et al., 2014), where an extremely small number of only four trees was used to evaluate the suitability of a biomass model for the whole country.

An important question, even in the case that a suitability check sample is available for the local population, is which statistical property is defining the suitability of the base model in the local population: accuracy or precision. Ideally, a suitable model would produce accurate (not significantly deviated), and precise (undispersed) predictions. When applied to the same population where the training data was gathered, the base model is expected to be: i) accurate, as, for example, ordinary least squares estimators are unbiased under different variance structures (Judge et al., 1985), and ii) with a precision proportional to the inherent variability of the local population, which is quantified by means of the prediction uncertainty. Whereas the hypothetical target of accuracy for a suitable model is clear (zero deviation), the precision target is related to the inherent variability in the local population, which is unknown. That is why the working hypothesis of the authors is that accuracy is the key aspect in judging model suitability. Accuracy testing has received attention in forest monitoring research since long (Freese, 1960), but the standard parametric techniques for testing accuracy have found to be sensitive to lack of normality in the data and not constant variance structures (Gregoire and Reynolds, 1988).

If a suitability check sample from the local population is available, the usual graphical and statistical tests for model validation may be applied to evaluate its suitability for that particular local population. A comprehensive compilation of the literature on forestry models validation can be found in (Gregoire, 2007). Even if graphical techniques proved to be very effective in the context of model validation (Piñeiro et al., 2008), their subjectivity is reason enough to combine them with statistical validity tests, when methodological transparency is required. The issue of quantitative vs qualitative model performance measurement for environmental models was addressed by Bennett et al. (2013), who proposed a classification of the quantitative methods based on common characteristics as: 1) those which directly compare model output to observed data as a whole, 2) those which combine individual observed and modelled values, 3) those which preserve data pattern, and those which measure model performance based on 4) parameter values, and 5) transformation of the data to a different domain.