Ovalle-Rivera, Oriana; Van Oijen, Marcel; Läderach, Peter; Roupsard, Olivier; de Melo Virginio Filho, Elias; Barrios, Mirna; Rapidel, Bruno. 2020. **Assessing the accuracy and robustness of a process-based model for coffee agroforestry systems in Central America**.

This is a post-peer-review, pre-copyedit version of an article published in *Agroforestry Systems*, 94 (5). 2033-2051. The final authenticated version is available online at: https://doi.org/10.1007/s10457-020-00521-6.

Assessing the accuracy and robustness of a process-based model for coffee agroforestry systems in Central America

Oriana Ovalle-Rivera[1,3], Marcel Van Oijen[2*], Peter Läderach[3], Olivier Roupsard[4,5,6], Elias de Melo Virginio Filho[1], Mirna Barrios[7], Bruno Rapidel[1,8,9]

[1] CATIE, Centro Agronómico Tropical de Investigación y Enseñanza, Turrialba 30501, Costa Rica

[2] UK Centre for Ecology & Hydrology (CEH-Edinburgh), Bush Estate, Penicuik EH26 0QB, United Kingdom

[3] International Center for Tropical Agriculture (CIAT). Cali, Colombia

[4] CIRAD, UMR Eco&Sols, BP1386, CP18524, Dakar, Sénégal

[5] Eco&Sols, Univ Montpellier, CIRAD, INRAE, IRD, Institut Agro, Montpellier, France

[6] LMI IESOL, Centre IRD-ISRA de Bel Air, BP1386, CP18524 Dakar, Senegal

[7] CATIE, Centro Agronómico Tropical de Investigación y Enseñanza, Managua, Nicaragua

[8] CIRAD, UMR SYSTEM, 30501 Turrialba, Costa Rica

[9] SYSTEM, Univ Montpellier, CIHEAM-IAMM, CIRAD, INRAE, Institut Agro, Montpellier, France

Abstract: Coffee is often grown in production systems associated with shade trees that provide different ecosystem services. Management, weather and soil conditions are spatially variable production factors. CAF2007 is a dynamic model for coffee agroforestry systems that takes these factors as inputs and simulates the processes underlying berry production at the field scale. There remain, however, uncertainties about process rates that need to be reduced through calibration.

Bayesian statistics using Markov chain Monte Carlo algorithms is increasingly used for calibration of parameter-rich models. However, very few studies have employed multi-site calibration, which aims to reduce parameter uncertainties using data from multiple sites simultaneously. The main objectives of this study were to calibrate the coffee agroforestry model using data gathered in long-term experiments in Costa Rica and Nicaragua, and to test the calibrated model against independent data from commercial coffee-growing farms. Two sub-models were improved: calculation of flowering date and the modelling of biennial production patterns. The modified model, referred to as CAF2014, can be downloaded at https://doi.org/10.5281/zenodo.3608877.

Calibration improved model performance (higher R^2, lower RMSE) for Turrialba (Costa Rica) and Masatepe (Nicaragua), including when all experiments were pooled together. Multi-site and single-site Bayesian calibration led to similar RMSE. Validation on new data from coffee-growing farms revealed that both calibration methods improved simulation of yield and its bienniality. The thus improved model was used to test the effect of N fertilizer and shade in different locations on coffee yield.

Key words: Agroforestry systems, Bayesian calibration, *Coffea arabica*, Modelling, Yield

[*] Corresponding author: ORCID 0000-0003-4028-3626, mvano@ceh.ac.uk

Declarations

Acknowledgments: This study was part of the PCP Agroforestry Systems with Perennial Crops, a scientific partnership platform led by CATIE and CIRAD in Central America. We greatly acknowledge CATIE and its partners and CIRAD for facilitating the data required for the calibration of this model, and the farm owners for granting us access to their farms and records. We also thank the reviewers of the original manuscript whose comments have led to considerable improvements in the presentation.

Funding: This study was funded by the Caf'Adapt project, Fontagro/RF-1027. It was implemented as part of the CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS) with the support from CGIAR Fund Donors and through bilateral funding agreements. For details, please visit https://ccafs.cgiar.org/donors. Coffee-Flux Observatory, Llano Bonito and CATIE agroforestry trial were also supported by SOERE F-ORE-T from Ecofor, Allenvi and the French national research infrastructure ANAEE-F (http://www.anaee-france.fr/fr/), the European project CAFNET (EuropAid/121998/C/G), the Ecosfix project (ANR- 2010-STRA-003-01), the CIRAD-SAFSE project, the MACACC project (ANR-13-AGRO-0005) and the CATIE-PROCAGICA-IICA-UE project. MvO acknowledges support from BBSRC through GCRF project SEACAF (BB/S01490X/1).

Conflicts of interest/Competing interests: None.

Ethics approval: Not applicable.

Consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and material: Requested for data should be addressed to EdMVF and MB.

Code availability: The coffee agroforestry model CAF2014 can be downloaded from https://doi.org/10.5281/zenodo.3608877.

Introduction

Process-based dynamic models have been used for over 50 years to explore the effect of variation of environmental variables or agricultural practices on agronomic or environmental indicators, like crop yields or N leached to aquifers (de Wit, 1965; Bunn et al., 2014; Makowski et al., 2014). Due to their ability to explore a wide range of options, dynamic models can be used to represent and optimize management decisions for increased outputs (Dogliotti et al. 2005). Models are frequently used to assess the effect of future climate change on crop yields, as they are able to represent conditions that are difficult to observe currently. However, to simulate the impacts of future conditions adequately, scientists have to evaluate very carefully the adequacy of their models for a wide variety of current conditions, including production situations to which the models were not specifically calibrated.

Agroforestry systems combine crops with trees in the same field. As such, they can represent a solution to the challenge of producing food for a growing global population while preserving the resources used for this production, as well as other ecosystem services provided to societies, such as provision of clean water, control of soil erosion, and control of pests and weeds. For certain crops at least, production under shade trees can be as good or better than in full sun (Jose, 2009). The trees in agroforestry systems may produce goods like timber, firewood or fruits (Cerda et al., 2014), or medicine. But they are also known to protect natural resources from exhaustion, by working as safety nets for nutrients, or by mobilizing them better from the soil (Van Noordwijk et al., 1996), to regulate climate both locally and globally (Vaast et al., 2015), or to protect soil surface from crusting, runoff and erosion (Villatoro-Sánchez *et al.*, 2015).

Agroforestry systems have been used by farmers only in a limited number of cases. Such cases include perennial crops, naturally adapted to growth and reproduction as understory crops, like coffee and cocoa grown under humid climates in the tropics. They also include other crops, like dry cereals in dry climates, when soil fertility and soil water balance are enhanced by some perennial shrubs, like *Guiera senegalensis* or *Piliostigma reticulatum* (Kizito *et al.*, 2012; Yelemou *et al.*, 2013; Hernandez *et al.*, 2015) or where crops and trees explore distinct niches, as is the case for *Faidherbia albida* in West Africa (Roupsard et al., 1999). The case of coffee and cocoa, though, is particular, as those crops are mostly cultivated in agroforestry systems (Jha et al., 2014).

Success in the combined provision of goods and services by agroforestry systems depends on delicate equilibria between the plant species involved, which can oscillate between competition and facilitation depending on the species involved, their management, or the environmental conditions (Jose, 2009; De Beenhouwer *et al.*, 2013; Taugourdeau *et al.*, 2014). No combination of crop and tree species exists that can be used everywhere. Scientific knowledge has been produced for a few decades now on the processes underlying these combined provisions. Some of this research has been done on experimental sites, where long-term experiments have produced a wealth of information (Imbach *et al.*, 1989; Haggar *et al.*, 2011). However, even nowadays when interest in agroforestry is high, such experiments are few, as they require large areas of land (due to border effects of tree plantations) over long times (typically 15-30 years).

Dynamic models can be used to explore the ability of agroforestry systems to provide ecosystem services. Agroforestry models can be somewhat artificially sorted into two types. Some, of a generic nature, focus on the interactions between species, like WaNuLCAS (Van Noordwijk and Lusiana, 1998). Others are more focused on a particular crop and try to estimate the effects of shade trees on its productivity (Zuidema et al., 2005; Rahn et al., 2018). These models, whatever their type, are useful for testing hypotheses on interactions between species under different environmental conditions, and for testing the impact of environmental change scenarios on the productivity and other ecosystem services provided by agroforestry systems. They have also proven useful to elicit and nurture fruitful participatory processes between farmers and researchers on the technical management of cropping systems (Carberry *et al.*, 2004; Whitbread *et al.*, 2010; Meylan *et al.*, 2014).

Dynamic crop models simulate phenology along full crop cycles. Rodríguez et al. (2011) proposed a physiologically-based full sun coffee dynamic growth and yield model, working from coffee organ (fruiting node) to whole coffee-plant and validated in two extreme latitudinal conditions for coffee cropping, with a special effort to accurately simulate the bud, flower and fruit phenology. This model proved to be efficient at early stages of the coffee cycle (0-5 years old). Recently, Vezy et al. (2020) incorporated the reproductive modules of Rodriguez et al. (2011), including reproductive cohorts to best distribute the fruit carbon demand along the year and scaled them up to simulate ecosystem services (multi-objective calibration) of a whole agroforestry field for full rotations, but the model was parameterized and tested for only one site so far. Indeed, another model existed previously for the simulation of coffee production at the field scale in full sun and agroforestry systems, CAF2007 (van Oijen *et al.*, 2010b). It was built to simulate coffee plantations in Central America, but has not been thoroughly parameterized based on agroforestry trials, nor tested in commercial plantations, so its use has been limited so far.

Adequate parameterization of agroforestry models is a complex task. Numerous processes are closely interrelated, so it is difficult to parameterize one process without having previously parameterized other connected processes. Measurements on diverse processes in coffee agroforestry systems have been carried out in experiments and in commercial plantations for some years now (van Oijen *et al.*, 2010a; Haggar *et al.*, 2011; Charbonnier *et al.*, 2013; Meylan *et al.*, 2013; Taugourdeau *et al.*, 2014; Gagliardi *et al.*, 2015; Padovan *et al.*, 2015; Villatoro-Sánchez *et al.*, 2015; Defrenet *et al.*, 2016). This parameterization, necessary as it is to use a model with reasonable confidence, cannot be done everywhere. To avoid parameterizing the model again and again depending on its intended use, we need to assess the robustness of the parameterization process itself: to do that, we can compare site-specific and multi-site calibrations in their ability to reproduce the same sets of data (Van Oijen et al. 2013).

The measurements made to parameterize the agroforestry models concern complex processes, measurement methods are frequently delicate and their results often come with significant uncertainties. These uncertainties need to be taken into account in the parameterization process. Methods for including probability distributions for measurements, parameters and outputs do exist, based on Bayesian statistics, and these methods have proven their suitability to complex processes and related models (Van Oijen et al., 2005; Van Oijen 2017). Bayesian calibration has been implemented in different models for specific sites. Multi-site calibration is a relatively new method for calibration of process-based models such as the VSD model, which simulates chemical solution of soil and nitrogen pools in natural and semi-natural ecosystems (Reinds *et al.*, 2008), the BASFOR forest model (Van Oijen *et al.*, 2013), and the BASGRA_N grassland model (Höglind *et al.*, 2020). We followed the procedure described by Van Oijen *et al.* (2005) which makes it possible to calibrate the parameters that influence the model processes based on data measured in the field while accounting for uncertainties in measurements and modelling.

This paper reports how the CAF2007 coffee agroforestry model was modified (and renamed to CAF2014), parameterized using data gathered over the course of several years at multiple sites, validated under commercial conditions for coffee in Central America, and applied to address challenges associated with the management of coffee tree plantations regarding the effect of shade and fertilization dose and distribution at different sites and altitudes in Costa Rica and Nicaragua.

2. Materials and methods

2.1 Study area

The study was carried out in the coffee-growing regions of Nicaragua and Costa Rica. The climatic conditions in these coffee growing regions have been analyzed and clustered in four different climatic zones shown in Figure 1, mainly related to the rainfall-temperature combinations from the WorldClim historical weather data base (Läderach et al. 2017).

Climatic zone 1 is characterized by cold and dry weather with an annual average precipitation of 1,544 mm and a mean annual temperature of 20°C. These conditions were only found in Nicaragua. *Climatic zone 2* is cold and humid with an annual average precipitation of 2,503 mm and a mean annual temperature of 19°C, present in both countries in some of the best producing regions, Jinotega and Matagalpa in Nicaragua and Tarrazú in Costa Rica. *Climatic zone 3* is characterized by being hot and humid with an annual average precipitation of 2,886 mm and a mean annual temperature of 23°C, mostly present in Costa Rica (Turrialba) and marginally in Nicaragua. *Climatic zone 4* is dry and hot, with an annual average precipitation of 1,688 mm and an annual mean temperature of 23°C, mainly present in Nicaragua (Masatepe, the oldest coffee producing region in Nicaragua, is a typical example of it), and almost restricted to the Nicoya peninsula in Costa Rica.

2.2 Sites used for model calibration

Twelve sites were used for calibration, representing three of the four climatic zones (Table 1). The sites were located at four different locations:

a. The CATIE long-term agroforestry experiment in Turrialba, Costa Rica (six sites - Zone 3) planted in 2000: Six of the calibration sites were located in the canton of Turrialba in the province of Cartago in Costa Rica, at 600 m above sea level. Haggar et al. (2011) described this location as one of low altitude with humid weather. Six sites were selected for calibrating the model with different intensities of management (quantities of fertilizers and other inputs), different densities and species of shade tree.

b. The Llano Bonito coffee-growing farm in San Pablo de León Cortés in Tarrazú, Costa Rica (single site - Zone 2): The calibration site was located at a coffee-growing farm in the region of Los Santos at 1,620 m above sea level near the central mountain range in Costa Rica. The selected farm has shade predominantly from Erythrina trees and some from musaceae (Meylan, 2012). The coffee field was gradually replanted conform local farming practice.

c. The Coffee-Flux observatory at the Aquiares farm in Cartago, Costa Rica (single site - Zone 3): The final Costa Rican calibration site was at the Aquiares farm which is located 10 km northwest of Turrialba at an average altitude of 1,100 m above sea level. 98% of the selected site area is cultivated with the Caturra coffee cultivar with shade from tall free-growing Erythrina trees (no pruning or thinning). The general management practices varied from year to year. The data for calibration in Aquiares were obtained from Charbonnier 2013, Taugourdeau *et al.,* 2014, Defrenet et al., 2016 and Kinoshita *et al.,* 2016.

d. The CATIE long-term coffee agroforestry trial in the low and dry zone in Masatepe, Nicaragua (four sites – Zone 4). The sites were located in the Pacific Center for Training and Regional Services (UNICAFE) with two repetitions planted in 2000. The sites were planted with the Pacas coffee variety (genetically very similar to the Caturra variety) with different management intensities. Two sites were in the shade predominantly from *Inga edulis trees* and two other sites were in full sun (Table 1).

2.3 Field data used for calibration

Seventeen variables were used for calibrating the model. These were variables that the model calculated and for which also measurements were available, but not all variables at all sites, as data had been collected primarily for other purposes. Information was available about coffee productivity at all sites, but data on average soil carbon content were only collected at 92% of the sites (Table 2). Data on the content of carbon in the above-ground portion of the coffee plants were available for 50% of sites. The leaf area indices of the coffee and shade trees as well as the content of carbon in the trunk and coffee leaves were measured more rarely.

Additionally, we had access to historical data on coffee flowering dates in the agroforestry trial in Turrialba. From prior simulations, we knew that flowering date was not predicted accurately. We used these data to modify the subroutine of the model that calculates the onset of flowering, which is essential as all other phenological stages are based on this flowering date (see next section).

2.4 From CAF2007 to CAF2014

2.4.1 Original version of the model

CAF2007 is a basic dynamic process model for simulating managed coffee full sun or agroforestry fields at a daily time step (van Oijen *et al.*, 2010b). Two vegetation layers are distinguished: shade tree and coffee. CAF2007 was designed to assist in taking decisions associated with management strategies such as fertilizer dose, shade tree density and species, pruning and thinning schedule. The model is also able to simulate the response of the system to environmental change (climate, atmospheric CO_2). The model simulates growth, yield and other services associated with specific tree species, taking into account the main processes occurring in plants and soil. These include the processes that contribute to the C-, N- and water-balance of the system. The model is generic by nature but it has thus far been calibrated only for the edapho-climatic conditions, coffee and tree genotypes and management conditions that are typical of Central America.

The model takes into consideration environmental inputs including radiation, precipitation, temperature, [CO_2], water, and nitrogen. The behavior of the simulated agroforestry system is constrained by soil properties, weather conditions, and individual site management. CAF2007 simulates the effects of shade trees on coffee through competition for light, water, and nutrients, and it takes into account the contribution of pruning and thinning to organic matter in the litter layer (van Oijen *et al.*, 2010b).

The model has 104 parameters, 70 of which are calibrated. Prior information for estimating parameter values was obtained from reviews of literature (van Oijen *et al.*, 2010a) including dissertations, project reports, data collections, and interviews with farmers. We now describe two modifications of the model, which led to a new model version that we refer to as CAF2014.

2.4.2 Model modifications for flowering

In the original model, flowering was triggered by daily rainfall exceeding a certain threshold, set at 10 mm by default, as soon as it occurred in the calendar year (Van Oijen 2010b). We modified this to better simulate actual flowering dates in regions where flowering is grouped and occurs after a significant period of water shortage: flowering now starts on the first day of the year on which the product of the amount of daily rainfall and the Julian day is greater than 1,000. This means that it can take 100 days after January 1 for flowering to occur with a daily rainfall of 10 mm to induce flowering or just 10 days of 100 mm rain. We used multi-annual time-series of flowering dates observed at the Aquiares farm experiment to check the ability of this new routine to improve the simulation of coffee flowering dates (Figure 2). The modification reduced RMSE for flowering date from 41.5 to 26.0. Further increases in prediction quality may be achievable, but it would require the writing of a new, complex model that takes into account soil water content, temperature and day length. We considered that the model in its new form was sufficiently accurate for our purposes, and consistent with our limited knowledge on the triggering of coffee flowering.

2.4.3 Model modifications for biennial production

In current full sun and moderate shade systems, years with high yields and low leaf-area index (LAI) tend to alternate with years with low yields but high LAI (Carvalho et al. 2020). The original CAF2007 model did not simulate a biennial pattern of coffee productivity. To incorporate this widely occurring phenomenon, the sink strength of the coffee beans is now inversely related to previous year's sink strength. This small change leads to biennial variation of simulated coffee yields which matches observations as shown in Figure 3. In the absence of data on bean sink strength, the inclusion of this modification in the model was not tested independently of the whole model.

2.4.4 Initialization and inputs of CAF2014

We refer to the model formed by modifying the flowering and bean sink algorithms of CAF2007 as CAF2014. This new model version is freely downloadable from https://doi.org/10.5281/zenodo.3608877, and a description of model structure can be found in a paper by Rahn et al. (2018), who carried out a parameter sensitivity analysis of CAF2014 for application in Uganda and Tanzania. To run the model, the initial values of state variables must be specified, as must be the site management practices and weather conditions. Data to meet these model information requirements were compiled for each of the experimental sites and coffee-growing farms in the study.

- Model initialization. Four values of initial carbon content in different plant parts are needed for shade trees, and four values for coffee trees. Seven initial values (primarily the contents of N and C) are needed for the soil.
- Management. Three parameters for coffee management (first day of pruning, pruning interval, and pruned biomass fraction), six for shade tree management (first day of pruning, pruning interval, pruned biomass fraction, thinning data, thinned biomass fraction, and initial tree density), and two for soil fertility management (date of application and dose of soil fertilizer).
- Weather. Six daily variables: minimum and maximum temperature (°C), wind speed (m s^{-1}), global radiation (MJ m^{-2} d^{-1}), atmospheric vapour pressure (kPa), and precipitation (mm d^{-1}).

2.5 Bayesian calibration

The values of model parameters are generally poorly constrained and the consequences of these uncertainties for model outputs must be quantified. We can represent such parameter uncertainties of process-based models by means of prior probability distributions, and use measurements on the model's output variables to calibrate the model within a Bayesian framework (Kennedy & O'Hagan 2001, Van Oijen et al. 2005, Van Oijen 2017).

2.5.1 Selection and prioritization of parameters to be calibrated

Some parameters values were known or directly measurable. These included geographic parameters and other parameters well documented in scientific literature. We did not include these parameters in the model calibration. Also not included in the calibration were parameters that had no significant impact on the results of the model, as shown in a sensitivity analysis by Remal (2009). Therefore, only those parameters were calibrated that had a significant impact on the results of the model and were not measured directly. Depending on each site, the number of parameters ranges from 63 to 67: 26 tree parameters, 13-17 soil parameters (depending on whether there was information available from a soil analysis at the site), and 24 coffee parameters. The sites with the largest number of calibration parameters were those for which there was no initial soil analysis available.

2.5.2 Bayesian calibration

Every Bayesian calibration begins by assigning a *prior* probability distribution to the model's parameters. The prior distribution for CAF2014 consisted of wide beta probability distributions based on literature review and other information (Van Oijen et al. 2010a). The calibration itself consists of using data on model output variables to update the parameter distribution, by application of Bayes' Theorem. We assumed independent measurement errors, represented by zero-centered Gaussian probability distributions with a coefficient of variation of 0.3. After all data are used, the updated distribution is referred to as the *posterior* parameter distribution. The method that we used for the calibration was Markov chain Monte Carlo sampling (MCMC) by means of the Metropolis algorithm (Robert & Casella 1999, Van Oijen et al. 2005). The R-code for the Metropolis algorithm is provided together with CAF2014 code at https://doi.org/10.5281/zenodo.3608877. The algorithm produces a representative sample from the posterior parameter distribution by a walk through 'parameter space'. Each proposed next step of the walk, i.e. each proposed new parameter vector, is accepted or rejected based on the product of the prior probability for that parameter vector and the *likelihood* of the data given CAF2014's outputs for the parameter vector. In this way, Bayesian calibration combines prior information with new data. For the calibrations reported here, we used Markov chains of length 100,000. Trace plots of the chains – showing how parameters values changed over the 100,000 iterations, were inspected to assess convergence visually. Based on this, an initial burn-in phase of 10,000 iterations was discarded from the final sample.

2.5.3 Types of calibration

We carried out both *single-site and multi-site* calibrations (Reinds et al. 2008). In the single-site calibrations, all calibrated parameters were considered to be site-specific. A separate MCMC was thus run for each site of Table 1, leading to twelve different site-specific posterior parameter distributions. In multi-site calibrations, data from multiple sites were used simultaneously in one MCMC, and posterior parameter estimates were assumed to apply to all sites involved. Two types of multi-site calibration were carried out: '*cluster*' calibration using subsets of sites close to each other (this was done for Turrialba and for Masatepe) and '*generic*' calibration which included all twelve sites of Table 1. Therefore a total of 15 different calibrations were carried out:

- 12 single-site calibrations (one for each site),
- 2 cluster calibrations (a six-site calibration for Turrialba and a four-site calibration for Masatepe),
- 1 generic calibration (for all twelve sites simultaneously).

2.5.4 Calibration evaluation

To estimate the goodness of fit of the model to data, the root mean square error (RMSE) was calculated for the mode of the posterior parameter distribution. The number of measurements observed vs. the number of simulated measurements was taken into account. The RMSE is defined as the square root of the sum of the squared differences between observed and simulated values divided by the total number of values. Values close to zero indicate a good model fit to the data.

$$\text{RMSE}=\sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i}-X_{model,i})^2}{n}},$$

where n = number of observations in the sample, $X_{obs,i}$= values observed for the "i"-th instance, and $X_{model,i}$ = are the values modelled for the "i"-th instance.

The validity of the RMSE value is limited in that this indicator assumes that data measured are accurate, which is contradictory to the Bayesian calibration principle that affirms that all values, including measured data, are associated with an uncertainty represented by a distribution of probabilities. The interpretation of RMSE must therefore be taken with some caution; in our study, we will focus on its use for the detection of systematic bias in the modelling outputs and possible correction.

2.6 Sites used for model validation

For validation purposes, information was compiled from non-experimental sites in Nicaragua (Table 3) where yield and climatic data could be collected accurately. Historical data were compiled from farmer-surveys and climatic data from weather stations near the farms for running the model. These included input data for driving the model such as weather, management of coffee plants, trees and soil as well as data on coffee yields to compare with model outputs.

One site was taken from each farm, planted with the Caturra coffee variety where shade comes predominantly from Inga trees using different coffee tree management practices. The coffee-growing farms were located in three climatic zones:

- Climatic zone 2 (cold and humid) was represented by the Solingalpa farm located in Jinotega. The farm has steep slopes (25%) planted with the Caturra coffee variety. Shade comes predominantly from guaba (*Psidium sp.*) trees with selective management practices.
- El Rosal farm is in Climatic zone 4 (dry and hot). It is located in Carazo department in Nicaragua, and represented by the "Las Negras" site. This site has shade predominantly from Erythrina trees with presence of the Catrenic coffee variety and management of shade and coffee trees.
- Lastly, the Hammonia and La Pinedita farms in the department of Matagalpa in Nicaragua represented Climatic zone 1 (dry and cold), to further challenge the robustness of model predictions.

2.7 Sensitivity analysis

To assess model behaviour under a wider range of conditions than were present in the study sites, we analysed the sensitivity of the calibrated model to various management options regarding fertilization and shade. The calibration sites differed in many respects (weather, shade management, fertilizer use etc.), so cannot be compared directly. The sensitivity analysis standardized fertilization to analyse shade response differences between sites (Table 5), and it standardized shade management to analyse fertilization impact differences (Table 6).

3. Results

The study results were first broken down into individual and multi-site calibrations using the modified model. The model was then validated using information of coffee-growing farms located in different climate clusters. We finally ran simulations of coffee-growing sites with the calibrated model, as a preliminary assessment of the capacity of the model to evaluate different management practices and site conditions.

3.1 Calibration of the CAF2014 model

In the first stage, the calibration was performed separately for each of the twelve sites listed in Table 1 (single site calibrations), then for Turrialba and Masatepe (cluster calibrations) and lastly, for the set of all sites (generic calibration). For each calibration, 100,000 MCMC iterations were carried out. Figure 3 shows examples of model simulations after cluster calibration for two of sites in Turrialba with different levels of fertilization and shade.

3.1.1 Single-site calibration

Measured production data were available for between 10 and 11 years for all sites, with the exception of Llano Bonito (only two years). Maximum measured production values of coffee beans dry matter in Turrialba, Aquiares, and Masatepe were 5.74, 4.3, and 5 tons DM cherry y^{-1}, respectively. The single-site calibrated model simulated maximum production values in Turrialba, Aquiares, and Masatepe of 5.81, 4.8, and 2.27, respectively.

Figure 4 shows simulated coffee production compared against measured production and the relevant determination coefficients (R^2) for each of the calibrated sites. We can globally observe that all Turrialba experiments were adequately simulated, with acceptable R^2, ranging from 0.54 to 0.71. More importantly, there seems to be no clear bias, overestimations and underestimations seem to compensate each other. On the other hand, although low production levels in Masatepe were correctly estimated, high productions are not, and this is particularly clear in the full sun intensive management site, where the best production was measured at 5 tons ha^{-1}, in 2005-2006, while the production simulated did not exceed 2.3 tons ha^{-1}. In Aquiares the model overestimates most harvests on average by 0.7 t DM cherry ha^{-1} y^{-1}. It has, however, a good fit with an R^2 value of 0.71 (Figure 4).

A comparison of the individually calibrated and uncalibrated sites (Fig. 5) indicates that the RMSE for coffee production (t DM cherry ha^{-1} y^{-1}) improves at the majority of the calibrated sites with an improvement in RMSE that ranges from 0.22 to 1.84. Several sites in Turrialba exhibit a good fit with low RMSE. Llano Bonito exhibits a high RMSE from the calibration of the coffee production. This is due to the low number of measured production data. This is also the case for sites in full sun with high conventional management practices in Turrialba and Masatepe before calibration, but RMSE was greatly reduced by calibration.

3.1.2 Multi-site calibration (by cluster and generically)

Figure 5 shows the RMSE for model simulations of coffee production in t DM cherry ha^{-1} y^{-1}, for each of the sites, after different calibration efforts: (1) no calibration, (2) generic calibration, (3) cluster calibration, (4) single-site calibration. The highest RMSE values are found in the case of no calibration, confirming that a calibrated model fits measured data better. On average, RMSE improves by 0.91.

The progression of RMSE from generic via cluster to individual calibration is uneven: it generally decreases but this evolution is not systematic: at Aquiares, surprisingly, the single-site calibration shows higher RMSE than generic calibration, but both calibrations show rather low RMSE.

Coefficients of determination (R^2) for cluster and generic calibrations are shown in Figure 6. Turrialba and Masatepe exhibit a similar R^2 value of 0.54-0.56. Some of the high harvest values simulated at both sites are underestimated. Generic calibration yields an R^2 of 0.64. The underestimations of the model at high productivity remain, but are not systematic.

Table 4 shows average coffee production as simulated following the three types of calibration. There were no significant differences versus measured production for any site with the exception of the Masatepe-3, the Nicaraguansite in full sun with high fertilization.

3.2 Validation of the CAF2014 model

Production simulations using the generically calibrated model exhibit low RMSE values and a good determination coefficient. Figure 7 shows an R^2 of 0.55 for the four validation farms, whose data had not been used for any model calibration. The model underestimated some of the high harvests, while the other harvests exhibit a good fit.

As the results from generic calibration were shown to perform adequately for calibration sites, without any dramatic increase in RMSE compared to cluster calibration or single-site calibration, we decided to use the generically calibrated model for the following simulations.

3.3 Additional simulations using the generically calibrated CAF2014 model

Simulations were carried out with the generically calibrated model to show the effect of shade and fertilization at different sites and altitudes in Costa Rica and Nicaragua (Tables 5, 6). The results reveal that production varies depending on the altitude and weather conditions at each site. Production in a hot and dry area (Masatepe) is lower in the sun than in the shade, in contrast to the wetter conditions (Turrialba and Llano Bonito) where shade reduces production (Table 5). But shade has a positive effect on productivity in the drier conditions. In contrast, in the more humid Costa Rican areas, production decreases by 10 to 22% in the shade from *Inga edulis* trees. The model simulations show that this is due to the fact that the tree crown diameter grows at a faster rate in humid zones than in dry zones.

Two virtual experiments were run to explore fertilization effects, the most expensive input in coffee production in Central America (Meylan *et al.*, 2013), related to dose and fractionation (Table 6). The dose that simulates the largest production in three fractions is 300 kg N ha^{-1} y^{-1}. At higher doses, the production did not increase anymore; most of this additional N was lost. Simulations using different fractionation of this fertilization rate showed that the effects of higher fractionations were real (with one exception), but minimal, probably less than the labour cost of an additional application. The days of N application were optimized in each experiment.

4. Discussion

We started from CAF2007, a simple dynamic model of coffee agroforestry systems (van Oijen *et al.*, 2010b), and modified the algorithms for two processes that were simulated inaccurately, i.e. blossoming date and biennial oscillation of cherry production. We then proceeded to calibrate the new model, CAF2014 using measurements from contrasting environmental conditions and management regimes. A Bayesian method was used for the calibration, for a total of 12 experimental sites. We found few differences between calibrations performed for each site separately (leading to site-specific estimates for coffee, tree and soil parameters), by cluster (Turrialba- and Masatepe-specific parameters), or generically for the complete dataset. The generically calibrated model was able to account for most of the variation in independent yield data from commercial plantations, the model was thus considered to be robust. We finally found that the modelled effects of N fertilization were not as strong as expected, and the effects of shade depended mainly on local humidity.

4.1 Single-site vs. multi-site model calibration

Single-site and multi-site calibrations revealed that the model exhibits very similar fits regardless of whether it is calibrated for a single site, for clusters, or for all sites together. The RMSE values are very similar at any of the sites regardless of the procedure, and always lower than the RMSE values of the uncalibrated model. This result is encouraging because it suggests that parameter values for coffee ecophysiology have limited variability in the studied region, which facilitates broadscale model application across Costa Rica and Nicaragua without a need for additional calibration. The finding is consistent with the narrow genetic base of cultivated *Coffea arabica* in the Western hemisphere (Sousa et al. 2017). The RMSE values for shaded sites for coffee production in Costa Rica and Nicaragua (t DM cherry ha^{-1} y^{-1}) were low and the R^2-values were high. Strong model performance for these sites may have been aided by the availability of good information on initial constants, site management, and *a priori* distribution of parameters. A remarkable feature of the calibrated model is that it accounts very well for the very high interannual variability in yields that was observed on all sites. Model predictions always accounted for more than 50% of interannual variation, and for about half the sites this reached about 70% (Figures 4 and 6). So the calibrated model can reproduce patterns of alternating high- and low-yielding years, i.e. alternate bearing (see also Figure 3). The absolute values of yield were underestimated in some years with high yields, in particular for Masatepe (e.g. Figure 4i). This site is in Climatic zone 4, which is dry and hot, so CAF2014 may be overestimating the impacts of water deficiency. The calibrated model also had a relatively high RMSE-value for the Llano Bonito site where shade was provided by *Erythrina poeppigiana* trees that were pollarded twice or thrice each year (see Figure 5). CAF2014 uses allometric equations to establish the relationship between tree branch biomass and crown area – and this relationship may conceivably be disrupted by the frequent pollarding. Quick re-growth of branches of this tree species after pollarding is generally observed, initiated by rapid mobilization of reserves from trunks (Nygren *et al.*, 1993). A new, Erythrina-specific tree submodel would be required to model the pollarding response, possibly based on the earlier work by Nygren et al. (1993, 1996). This is considered for future modifications of the model.

4.2 Model testing against independent data and sensitivity analysis

Our tests against independent data from three of the climatic zones corroborated that the model behaves well under different management and biophysical conditions (Fig. 7). The tests were carried out using the posterior mode from generic calibration; no site-specific information was used to adjust parameter values. Overall, the comparison of model estimates and production rates at commercial farms showed the same qualities and defects as the calibration results. The model correctly estimated low production rates, but underestimated high production rates. It is possible that the control of weeds, pests and diseases as well as the reliability of the data themselves differed between the experimental calibration sites and the commercial testing sites, but detailed information on the growing conditions at the commercial farms is lacking. Nevertheless, the model again ranked high- and low-yielding years for the most part correctly, leading to an R^2 value of 0.55 (Fig. 7). It thus seems that the alternate bearing pattern of coffee may largely be explained from factors that were present in the model, i.e. interannual variation in weather conditions and the negative lag-effects of high reproductive sinks on sink strengths in succeeding years – conform theories of carbon allocation in woody plants (Génard *et al.*, 2008). The flowering date, the modelling of which was modified in CAF2014, also affects the balance between the sources and sinks of carbohydrates, as allocation patterns change dramatically after flowering. We note however that our new implementation of biennial sink patterns was not highly mechanistic, so there remains significant scope for model improvement. This is complicated because of the difficulty of measuring sink strength directly and because of the complicated interannual dynamics of reserves in perennial woody plants. It does constitute an important research question because alternate bearing is a phenomenon common to a large number of species of fruit trees (Monselise and Goldschmidt, 1982). In future model development, CAF2014 may benefit from incorporating the equations of Rodriguez et al. (2011) for the dynamics of cohorts of reproductive organs and reserve compartment, as was done by Vezy et al. (2020) in their DynACof model. That would constitute a more mechanistic simulation of sink competition between leaf and reproductive compartments than we attempted here, but it would increase model complexity. Moreover, the method still needs independent testing across sites in multiple climatic zones (only one site was used by Vezy et al. 2020) and Bayesian multi-site calibration following the approach that we developed here.

Our findings suggest that our model can be used in Central America, because the calibrations at experimental sites exhibited a good and relatively robust fit, which was confirmed through validation. Moreover, the sensitivity analysis provided plausible conclusions with respect to management: least yield loss from shading at low altitude ((Table 5) and little benefit from fertilizer above 200 kg N ha^{-1} y^{-1}, both of which are consistent with the literature (e.g. Beer et al. 1998, Meylan et al. 2017). The calibrated model may thus become a useful tool for various stakeholders, such as farms and policymakers, to support decisions regarding issues like climate change, fertilization efficiency, use of tree species for shade, and other management practices. The model can also provide estimates of other ecosystem services, including water-, carbon- and nitrogen-retention, but the quality of model predictions for those variables requires additional data to allow further testing of the model beyond the yield estimates that we focused on here.

5. Conclusions

We were able to calibrate the CAF2014 coffee agroforestry model for farms in Costa Rica and Nicaragua that span different climatic zones, soils, shading practices and management conditions. Interannual variability was well accounted for by the model. Whereas simulation of coffee production (t DM cherry ha^{-1} $year^{-1}$) using the original model underestimated production, the modified and calibrated model showed realistic production rates, decreasing RMSE and increasing R^2. Simulations were improved for coffee production in three climatic zones, including one that had not been included for calibration. However, the model still underestimates very high production rates at some sites. Coffee models implemented thus far have allowed providing an assessment of the niche-range over which the species is distributed and comparing the ability of crops to face climate changes in the future. The calibrated CAF2014 model makes it possible to simulate coffee production yields in agroforestry systems, thus enabling estimates of the costs and benefits of implementing the system as well as the impacts of climate change, elevated CO_2, fertilization and pruning of coffee plants and trees - estimates that empirical suitability models are not able to provide (Ovalle-Rivera *et al.*, 2015). The model may thus be used as a tool for exploring different adaptation scenarios in the face of current and future problems of coffee growers, as shown in our preliminary study of the effects of N fertilizer and shade in different locations on coffee productivity.

References

Beer, J., Muschler, R., Kass, D., Somarriba, E., 1998. Shade management in coffee and cacao plantations. Agrofor. Syst. 38, 139–164.

Bunn C, Läderach P, Ovalle Rivera O, Kirschke D, 2014. A bitter cup: climate change profile of global production of Arabica and Robusta coffee. Climatic Change 129: 89–101 DOI 10.1007/s10584-014-1306-x

Carberry, P., Gladwin, C., Twomlow, S., 2004. Linking simulation modelling to participatory research in smallholder farming systems. In: Delve, R.J., Probert, M.E. (Eds.), Modelling nutrient management in tropical cropping systems. ACIAR Proceedings No. 114, pp. 32-46.

Carvalho, H.F., Galli, G., Ferrao, L.F.V., Nonato, J.V.A., Padilha, L., Maluf, M.P., de Resende Jr, M.F.R., Guerreiro Filho, O., Fritsche-Neto, R., 2020. The effect of bienniality on genomic prediction of yield in arabica coffee. Euphytica 216, 101: 1-16.

Cerda, R., Deheuvels, O., Calvache, D., Niehaus, L., Saenz, Y., Kent, J., Vilchez, S., Villota, A., Martinez, C., Somarriba, E., 2014. Contribution of cocoa agroforestry systems to family income and domestic consumption: looking toward intensification. Agrofor Syst 88, 957-981.

Charbonnier, F., 2013. Mesure et modélisation des bilans de lumière, d'eau, de carbone et de productivité primaire nette dans un système agroforestier à base de caféier au Costa Rica. Université de Lorraine, Nancy, France, p. 230.

Charbonnier, F., Roupsard, O., le Maire, G., Guillemot, J., Casanoves, F., Lacointe, A., Vaast, P., Allinne, C., Audebert, L., Cambou, A., Clement-Vidal, A., Defrenet, E., Duursma, R.A., Jarri, L., Jourdan, C., Khac, E., Leandro, P., Medlyn, B.E., Saint-Andre, L., Thaler, P., Van den Meersche, K., Aguilar, A.B., Lehner, P., Dreyer, E., 2017. Increased light-use efficiency sustains net primary productivity of shaded coffee plants in agroforestry system. Plant Cell and Environment 40, 1592-1608.

De Beenhouwer, M., Aerts, R., Honnay, O., 2013. A global meta-analysis of the biodiversity and ecosystem service benefits of coffee and cacao agroforestry. Agr Ecosyst Environ 175, 1-7.

de Wit, C.T., 1965. Photosynthesis of leaf canopies. Agricultural Research Report, 663. Wageningen University, Wageningen.

Dogliotti, S., van Ittersum, M.K., Rossing, W.A.H., 2005. A method for exploring sustainable development options at farm scale: a case study for vegetable farms in South Uruguay. Agr Syst 86, 29-51.

Gagliardi, S., Martin, A.R., Virginio Filho, E.d.M., Rapidel, B., Isaac, M.E., 2015. Intraspecific variation in leaf traits along a shade tree biodiversity gradient in coffee (Coffea arabica) agroforestry systems of Costa Rica. Agr Ecosyst Environ 200, 151-160.

Génard, M., Dauzat, J., Franck, N., Lescourret, F., Moitrier, N., Vaast, P., Vercambre, G., 2008. Carbon allocation in fruit trees: from theory to modelling. Trees 22, 269-282.

Haggar, J., Barrios, M., Bolaños, M., Merlo, M., Moraga, P., Munguia, M., Ponce, A., Romero, S., Soto, G., Staver, C., Virginio, E., 2011. Coffee agroecosystem performance under full sun, shade, conventional and organic management regimes in Central America. Agrofor Syst 82, 285-301.

Hernandez, R.R., Debenport, S.J., Leewis, M., Ndoye, F., Nkenmogne, K.I.E., Soumare, A., Thuita, M., Gueye, M., Miambi, E., Chapuis-Lardy, L., Diedhiou, I., Dick, R.P., 2015. The native shrub, *Piliostigma reticulatum*, as an ecological "resource island" for mango trees in the Sahel. Agriculture Ecosystems & Environment 204, 51-61.

Höglind, M., Cameron, D., Persson, T., Huang, X., Van Oijen, M. 2020. BASGRA_N: A model for grassland productivity, quality and greenhouse gas balance. Ecological Modelling 417: 108925. https://doi.org/10.1016/j.ecolmodel.2019.108925.

Imbach, A., Fassbender, H., Beer, J., Borel, R., Bonnemann, A., 1989. Sistemas agroforestales de Café (Coffea arabica) con Laurel (*Cordia Alliodora*) y Café con Poro (*Erythrina poeppigiana*) en Turrialba, Costa Rica. VI. Balances hidricos e ingreso con lluvias y lixiviacion de elementos nutritivos. Turrialba 39, 400-414.

Jha, S., Bacon, C.M., Philpott, S.M., Mendez, V.E., Laderach, P., Rice, R.A., 2014. Shade coffee: Update on a disappearing refuge for biodiversity. BioScience 64, 416-428.

Jose, S., 2009. Agroforestry for ecosystem services and environmental benefits: An overview. Agrofor Syst 76, 1-10.

Kennedy, M.C., O'Hagan, A. 2001. Bayesian calibration of computer models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63: 425–64.

Kizito, F., Dragila, M.I., Sene, M., Brooks, J.R., Meinzer, F.C., Diedhiou, I., Diouf, M., Lufafa, A., Dick, R.P., Selker, J., Cuenca, R., 2012. Hydraulic redistribution by two semi-arid shrub species: Implications for Sahelian agro-ecosystems. J Arid Environ 83, 69-77.

Läderach, P., Ramirez–Villegas, J., Navarro-Racines, C., Zelaya1, C., Martinez–Valle, A., Jarvis, A., 2017. Climate change adaptation of coffee production in space and time. Clim. Ch. 141: 47-62.

Meylan, L., 2012. Design of cropping systems combining production and ecosystem services: developing a methodology combining numerical modeling and participation of farmers. Fonctionnement des Ecosystèmes Naturels et Cultivés. Montpellier Supagro, Montpellier, p. 145.

Meylan, L., Gary, C., Allinne, C., Ortiz, J., Jackson, L., Rapidel, B. 2017. Evaluating the effect of shade trees on provision of ecosystem services in intensively managed coffee plantations. Agric Ecosys Env 245: 32-42.

Meylan, L., Merot, A., Gary, C., Rapidel, B., 2013. Combining a typology and a conceptual model of cropping system to explore the diversity of relationships between ecosystem services: The case of erosion control in coffee-based agroforestry systems in Costa Rica. Agr Syst 118, 52-64.

Meylan, L., Sibelet, N., Gary, C., Rapidel, B., 2014. Combining a numerical model with farmer participation for the design of sustainable and practical agroforestry systems. IIIrd World Congress of Agroforestry. ICRAF, Delhi, India, 10-14 Feb 2014.

Monselise, S.P., Goldschmidt, E.E., 1982. Alternate bearing in fruit trees. Horticultural Reviews. John Wiley & Sons, Inc., pp. 128-173.

Nygren, P., Kiema, P., Rebottaro, S., 1996. Canopy development, CO_2 exchange and carbon balance of a modeled agroforestry tree. Tree Physiol 16, 733-745.

Nygren, P., Rebottaro, S., Chavarria, R., 1993. Application of the pipe model-theory to nondestructive estimation of leaf biomass and leaf-area of pruned agroforestry trees. Agrofor Syst 23, 63-77.

Ovalle-Rivera, O., Läderach, P., Bunn, C., Obersteiner, M., Schroth, G., 2015. Projected shifts in *Coffea arabica* suitability among major global producing regions due to climate change. PLoS ONE 10, e0124155.

Padovan, M.P., Cortez, V.J., Navarrete, L.F., Navarrete, E.D., Barrios, M., Munguía, R., Centeno, L.G., Deffner, A.C., Vílchez, J.S., Vega, C., Costa, A.N., Brook, R.M., Rapidel, B., 2015. Root distribution and water use in a coffee shaded with *Tabebuia rosea* Bertol. and *Simarouba glauca* DC. compared to full sun coffee in sub-optimal environmental conditions. Agrofor Syst 89, 857-868.

Rahn E et al., Vaast, P., Läderach, P., van Asten, P., Jassogne, L., Ghazoul, J. 2018 Exploring adaptation strategies of coffee production to climate change using a process-based model. Ecol Model 371:76–89.

Reinds, G.J., Van Oijen, M., Heuvelink, G.B.M., Kros, H., 2008. Bayesian calibration of the VSD soil acidification model using European forest monitoring data. Geoderma 146, 475-488.

Remal, S., 2009. Conceptual and numerical evaluation of a plot scale, process-based model of coffee agroforestry systems in Central America: CAF2007. Production Végétales Durables. SupAgro, Monptellier, p. 40.

Rodríguez, D., Cure, J.R., Cotes, J.M., Gutierrez, A.P., Cantor, F., 2011. A coffee agroecosystem model: I. Growth and development of the coffee plant. Ecol Model 222, 3626-3639.

Roupsard, O., Ferhi, A., Granier, A., Pallo, F., Depommier, D., Mallet, B., Joly, H.I., Dreyer, E., 1999. Reverse phenology and dry-season water uptake by *Faidherbia albida* (Del.) A. Chev. in an agroforestry parkland of Sudanese west Africa. Funct Ecol 13, 460-472.

Sousa, T.V., Caixeta, E.T., Alkimim, E.R., de Oliveira, A.C.B., Pereira, A.A., Zambolim, L., Sakiyama, N.S., 2017. Molecular markers useful to discriminate *Coffea arabica* cultivars with high genetic similarity. Euphytica 213, 75: 1-15.

Taugourdeau, S., le Maire, G., Avelino, J., Jones, J.R., Ramirez, L.G., Quesada, M.J., Charbonnier, F., Gómez-Delgado, F., Harmand, J.-M., Rapidel, B., Vaast, P., Roupsard, O., 2014. Leaf area index as an indicator of ecosystem services and management practices: an application for coffee agroforestry. Agr Ecosyst Environ 192, 19-37.

Vaast, P., Harmand, J.-M., Rapidel, B., Jagoret, P., Deheuvels, O., 2015. Coffee and cocoa production in agroforestry—A climate-smart agriculture model. In: Torquebiau, E. (Ed.), Climate Change and Agriculture Worldwide. Editions Quae, Paris, France, pp. 209-224.

Van Noordwijk, M., Lawson, G., Soumaré, A., Groot, J.J.R., Hairiah, K., 1996. Root distribution of trees and crops: competition and/or complementarity. In: Ong, C.K., Huxley, P.W. (Eds.), Tree-Crop Interactions: A Physiological Approach. CAB International, Wallingford, UK, pp. 319-364.

Van Noordwijk, M., Lusiana, B., 1998. WaNuLCAS, a model of water, nutrient and light capture in agroforestry systems. Agrofor Syst 43, 217-242.

Van Oijen, M., 2017. Bayesian methods for quantifying and reducing uncertainty and error in forest models. Current Forestry Reports 3: 269-280. https://doi.org/10.1007/s40725-017-0069-9 .

Van Oijen, M., Dauzat, J., Harmand, J.-M., Lawson, G., Vaast, P., 2010a. Coffee agroforestry systems in Central America: I. A review of quantitative information on physiological and ecological processes. Agrofor Syst 80, 341-359.

Van Oijen, M., Dauzat, J., Harmand, J.-M., Lawson, G., Vaast, P., 2010b. Coffee agroforestry systems in Central America: II. Development of a simple process-based model and preliminary results. Agrofor Syst 80, 361-378.

Van Oijen, M., Reyer, C., Bohn, F.J., Cameron, D.R., Deckmyn, G., Flechsig, M., Härkönen, S., Hartig, F., Huth, A., Kiviste, A., Lasch, P., Mäkelä, A., Mette, T., Minunno F., Rammer, W. 2013. Bayesian calibration, comparison and averaging of six forest models, using data from Scots pine stands across Europe. Forest Ecology and Management 289: 255–68. https://doi.org/10.1016/j.foreco.2012.09.043 .

Van Oijen, M., Rougier, J., Smith, R., 2005. Bayesian calibration of process-based forest models: bridging the gap between models and data. Tree Physiol 25, 915-927.

Vezy, R., le Maire, G., Christina, M., Georgiou, S., Imbach, P., Hidalgo, H.G., Alfaro, E.J., Blitz-Frayret, C., Charbonnier, F., Lehner, P., Loustau, D., Roupsard, O., 2020. DynACof: A process-based model to study growth, yield and ecosystem services of coffee agroforestry systems. Environmental Modelling & Software 124, 104609.

Villatoro-Sánchez, M., Le Bissonnais, Y., Moussa, R., Rapidel, B., 2015. Temporal dynamics of runoff and soil loss on a plot scale under a coffee plantation on steep soil (Ultisol), Costa Rica. J Hydrol 523, 409-426.

Whitbread, A., Robertson, M.J., Carberry, P.S., Dimes, J.P., 2010. Farming systems simulation can aid the development of more sustainable smallholder farming systems in Southern Africa. Eur J Agron 32, 51-58.

Yelemou, B., Dayamba, S.D., Bambara, D., Yameogo, G., Assimi, S., 2013. Soil carbon and nitrogen dynamics linked to Piliostigma species in ferugino-tropical soils in the Sudano-Sahelian zone of Burkina Faso, West Africa. Journal of Forestry Research 24, 99-108.

Zuidema, P.A., Leffelaar, P.A., Gerritsma, W., Mommer, L., Anten, N.P.R., 2005. A physiological production model for cocoa (Theobroma cacao): model presentation, validation and application. Agr Syst 84, 195-225.

Table 1. Sites used for model calibration: Turrialba (9.8962N; 83.6673W, 610 masl), Aquiares (9.9383N; 83.7279W, 1100 masl) and Llano Bonito (9.6707N; 84.0951W, 1620 masl) in Costa Rica, and Masatepe (11.9008N; 86.1461W, 467 masl) in Nicaragua.

Site	Climatic zone	Shade	Fertilization (kg N ha^{-1} y^{-1})	Shade tree pruning	Annual shade tree thinning
Turrialba-1	3	*Terminalia amazonia*	280	Regulated	80%
Turrialba-2	3	*Terminalia amazonia*	150	Regulated	80%
Turrialba-3	3	*Erythrina poeppigiana*	280	Drastic	50%
Turrialba-4	3	*Erythrina poeppigiana*	150	Regulated	50%
Turrialba-5	3	Full sun	280	-	-
Turrialba-6	3	Full sun	150	-	-
Aquiares	3	*Erythrina poeppigiana*	260	Unregulated	Without
Llano Bonito	2	Mainly *E. poeppigiana*	300	Regulated	20%
Masatepe-1	4	Mainly *Inga edulis*	144	Regulated	61%
Masatepe-2	4	Mainly *Inga edulis*	73	Regulated	66%
Masatepe-3	4	Full sun	144	-	-
Masatepe-4	4	Full sun	73	-	-

Table 2. Output variables for calibration of the CAF2014 model. The frequency indicates the percentage of sites where a variable was measured.

Variable	Identifier	Unit	Frequency
Average content of C in the soil	Csoilave	t C ha^{-1}	92%
Average leaf area index of coffee trees	LAIave	m^2 m^{-2}	17%
Average leaf area index of shade trees	LAIT	m^2 m^{-2}	17%
C in the above-ground portion of coffee plants	CT	kg C m^{-2}	50%
C in the above-ground portion of shade trees	CTT	kg C m^{-2}	50%
C in coffee leaves in full sun	CL(1)	kg C m^{-2}	8%
C in coffee leaves in the shade.	CL(2)	kg C m^{-2}	8%
C in coffee trunks in full sun	CW(1)	kg C m^{-2}	8%
C in coffee trunks in the shade	CW(2)	kg C m^{-2}	8%
Coffee productivity *	harvDMav_year	t DM ha^{-1} y^{-1}	100%
Leaf area index in full sun	LAI(1)	m^2 m^{-2}	8%
Leaf area index in the shade	LAI(2)	m^2 m^{-2}	8%
N in the soil	Nsoilave	t N ha^{-1}	75%
Shade area	SA	m^2 m^{-2}	50%
Tree crown area	CAtree	m^2	17%
Tree height	h	m	33%
Water content in the soil	WC_F	m^3 H_2O m^{-3}	25%

* split into "under the sun" and "in the shade" at Llano Bonito

Table 3. Coffee-growing farms in three climatic zones in Nicaragua used for validating the CAF2014 model.

Climatic zone	Farm	Coordinates / Altitude (masl)	Main shade sp.	Soil fertilization (kg N ha^{-1})	Shade tree pruning*	Shade tree thinning**
1	Hammonia	12.99N 85.92W / 1237	guaba	80	Once	No
1	La Pinedita	12.92N 85.90W / 917	guaba	116	None	50%
2	Solingalpa	13.03N 85,91W / 1368	guaba	182	10%	60%
4	El Rosal	11.88N 86.20W / 588	Erythrina	136	10%	50%

Selective coffee tree pruning * Fraction for each tree pruning ** Fraction of thinning

Table 4. Average production (10-11 years) in t DM ha^{-1} y^{-1} for calibration sites in Turrialba, Costa Rica, and in Masatepe, Nicaragua.

Site	Measured production	Simulated production (Single-site calibration)	*p-value*	Simulated production (Cluster calibration)	*p-value*	Simulated production (Generic calibration)	*p-value*
Turrialba-1	2.25	2.22	0.94	2.19	0.91	1.68	0.25
Turrialba-2	1.58	1.73	0.73	1.56	0.98	1.53	0.92
Turrialba-3	3.03	2.82	0.81	2.88	0.70	2.53	0.36
Turrialba-4	1.90	2.31	0.33	2.68	0.14	2.33	0.33
Turrialba-5	3.53	3.22	0.65	2.47	0.13	2.60	0.16
Turrialba-6	2.87	3.27	0.53	2.47	0.54	2.57	0.63
Masatepe-1	1.59	1.29	0.4	1.19	0.23	1.40	0.25
Masatepe-2	1.47	1.24	0.49	1.25	0.51	1.26	0.92
Masatepe-3	2.28	1.55	0.14	1.26	**0.045**	1.31	**0.05**
Masatepe-4	1.78	1.55	0.51	1.30	0.18	1.31	0.198

Table 5. Effect of shade on coffee production at three different altitudes with fertilization fractionated in three doses of 150 kg N ha^{-1} y^{-1}.

Altitude (m above sea level)	**Site**	**Production average over 11 years** (t DM ha^{-1} y^{-1})	
		Sun	**Shade/Inga edulis**
453	Masatepe	1.61	1.68 (42% shade)
600	Turrialba	3.00	2.70 (53% shade)
1620	Llano Bonito	3.16	2.47 (56% shade)

Table 6. Effect on coffee production of fertilization in different fractions and at different doses for Turrialba and Masatepe.

Dose (kg N ha^{-1} y$^{-1)}$	**Fractionation** (day of year)	**Production average over 11 years** (t DM ha^{-1} y^{-1})		
		Turrialba		**Masatepe**
		Full sun	Erythrina trees	Full sun
50	(135,289,350) Turrialba (185, 256, 276) Masatepe	2.61	2.52	1.46
100		2.85	2.73	1.54
200		3.09	2.90	1.66
300		3.17	2.97	1.73
400		3.20	2.95	1.76
300	(135,289) Turrialba (165,275) Masatepe	3.09	2.92	1.72
	(135,289,350) Turrialba (185,256,276) Masatepe	3.17	2.97	1.73
	(135,210,289,350) Turrialba (165,215,275,300) Masatepe	3.17	2.94	1.73

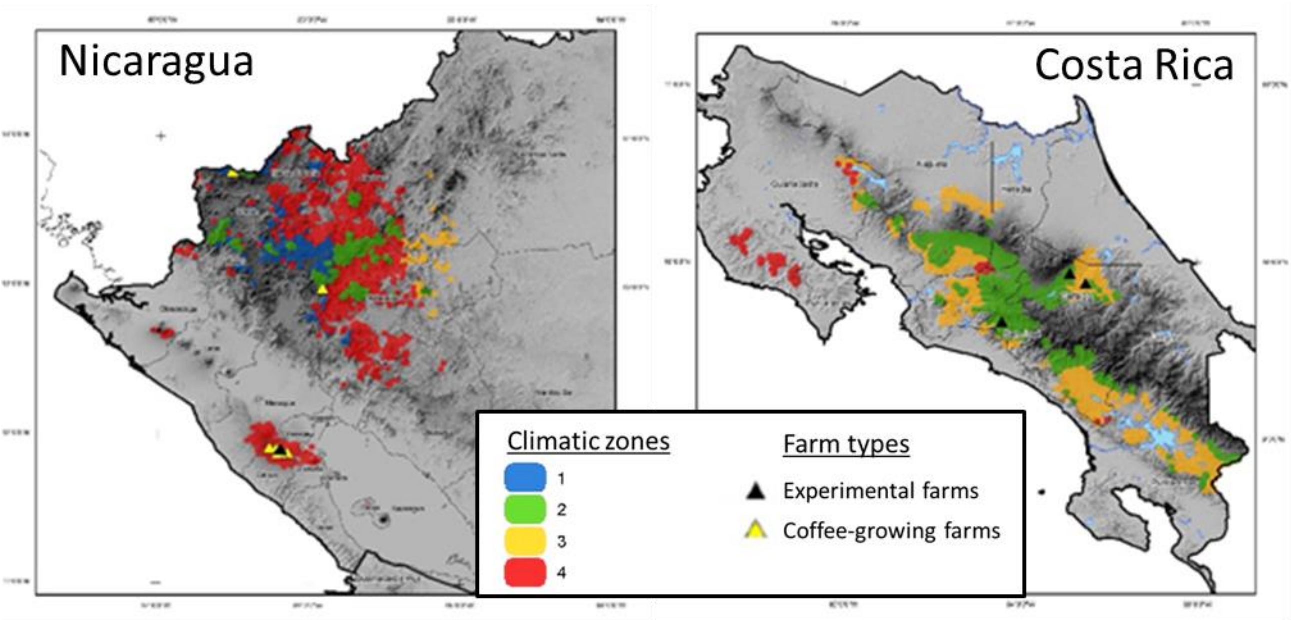

Figure 1. The coffee growing regions of Nicaragua and Costa Rica. Four climatic zones: 1 = cold and dry; 2 = cold and humid ; 3 = hot and humid; 4 = hot and dry. Triangles indicate the experimental and coffee-growing farms from which data were used for model simulations.

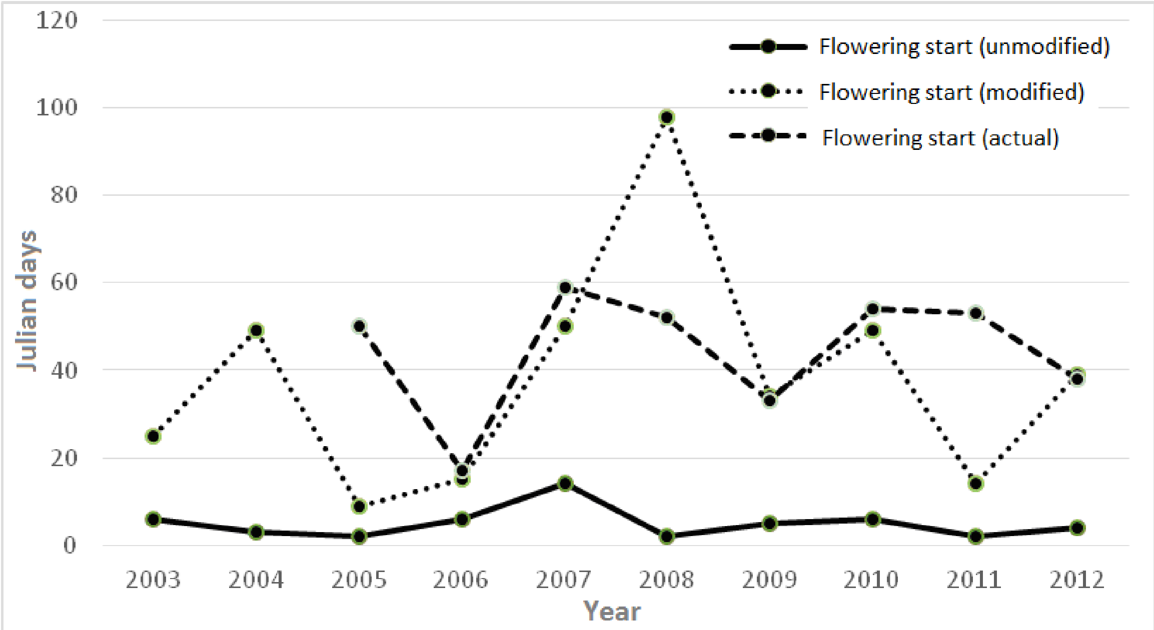

Figure 2. Start date of flowering simulated with the original unmodified model, the modified model (CAF2014) and actually observed flowering in Turrialba, Costa Rica.

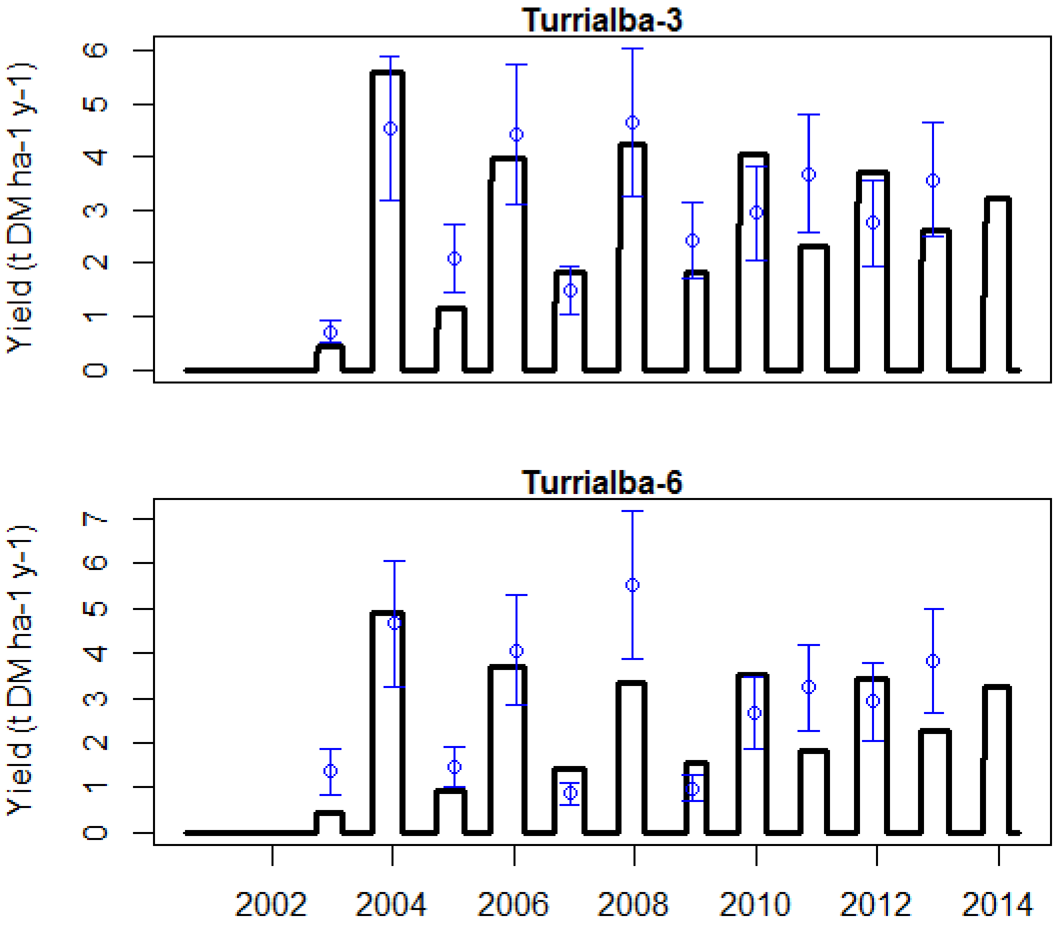

Figure 3. Coffee production at two sites in Costa Rica. Fertilization rate was high in Turrialba-3 (280 kg N ha^{-1} y^{-1}) and intermediate in Turrialba-6 (150 kg N ha^{-1} y^{-1}). Blue circles and error bars: measurements. Black lines: simulations using the posterior mode from cluster calibration, showing cumulative yield within each calendar year.

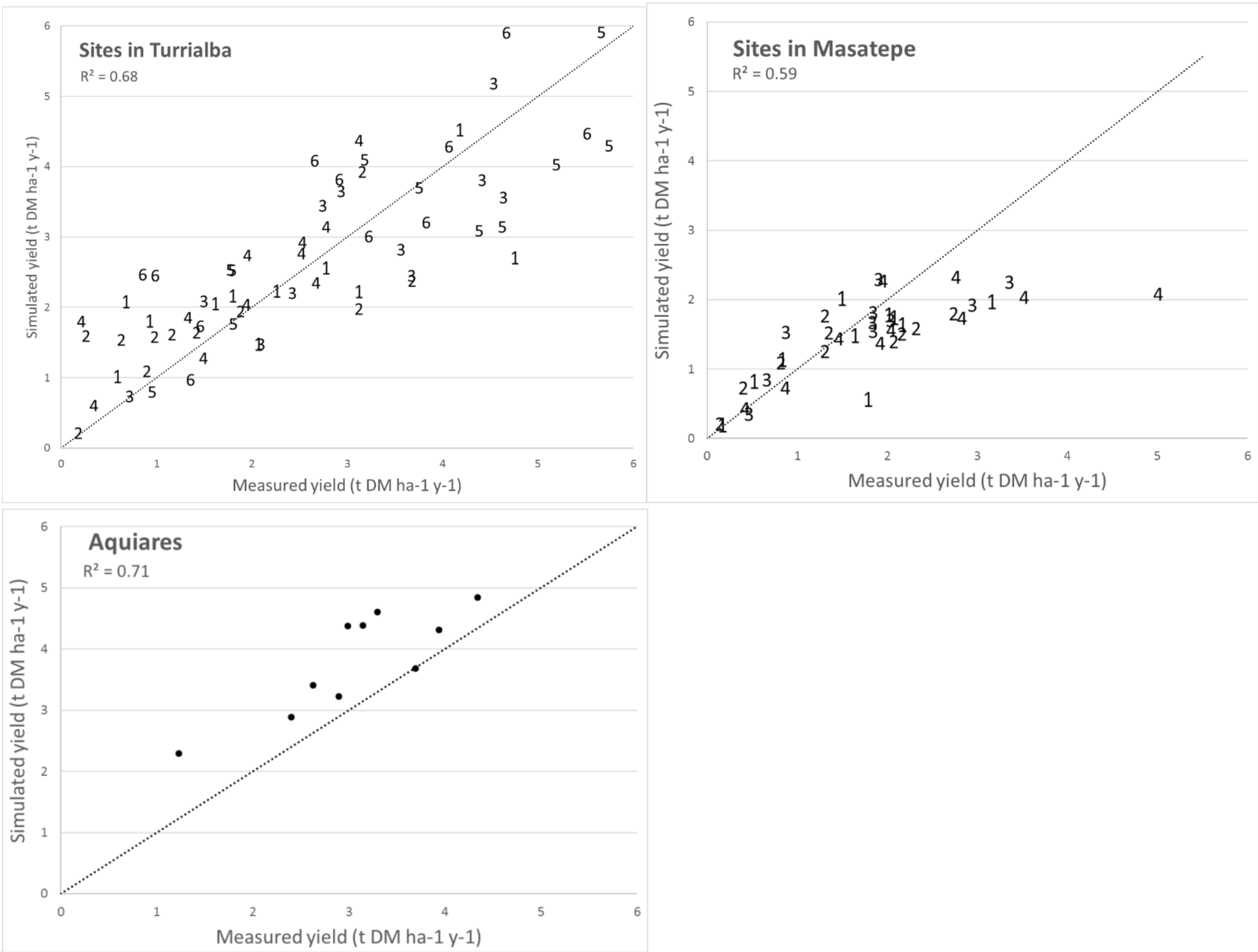

Figure 4. Simulated vs. measured coffee productivity (t DM ha^{-1} y^{-1}) at each calibrated site. The simulated yields are from the posterior mode after single-site calibration. The Llano Bonito site is not shown because it has data for only two years of production. The digits on the top two panels identify the six different sites in Turrialba and the four sites in Masatepe (see Table 1).

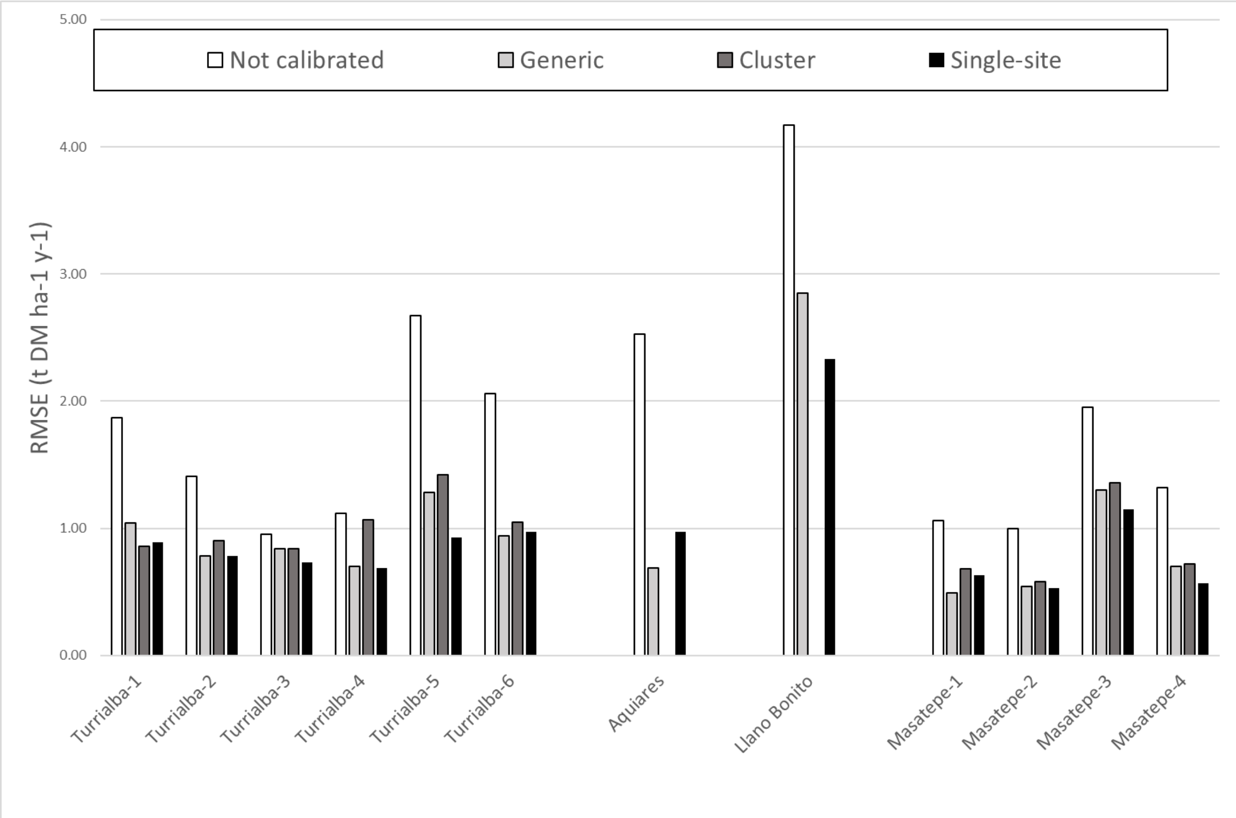

Figure 5. RMSE values for coffee production (t DM ha^{-1} y^{-1}) from 12 sites after different calibrations in Costa Rica and Nicaragua. See Table 1 for details about the sites.

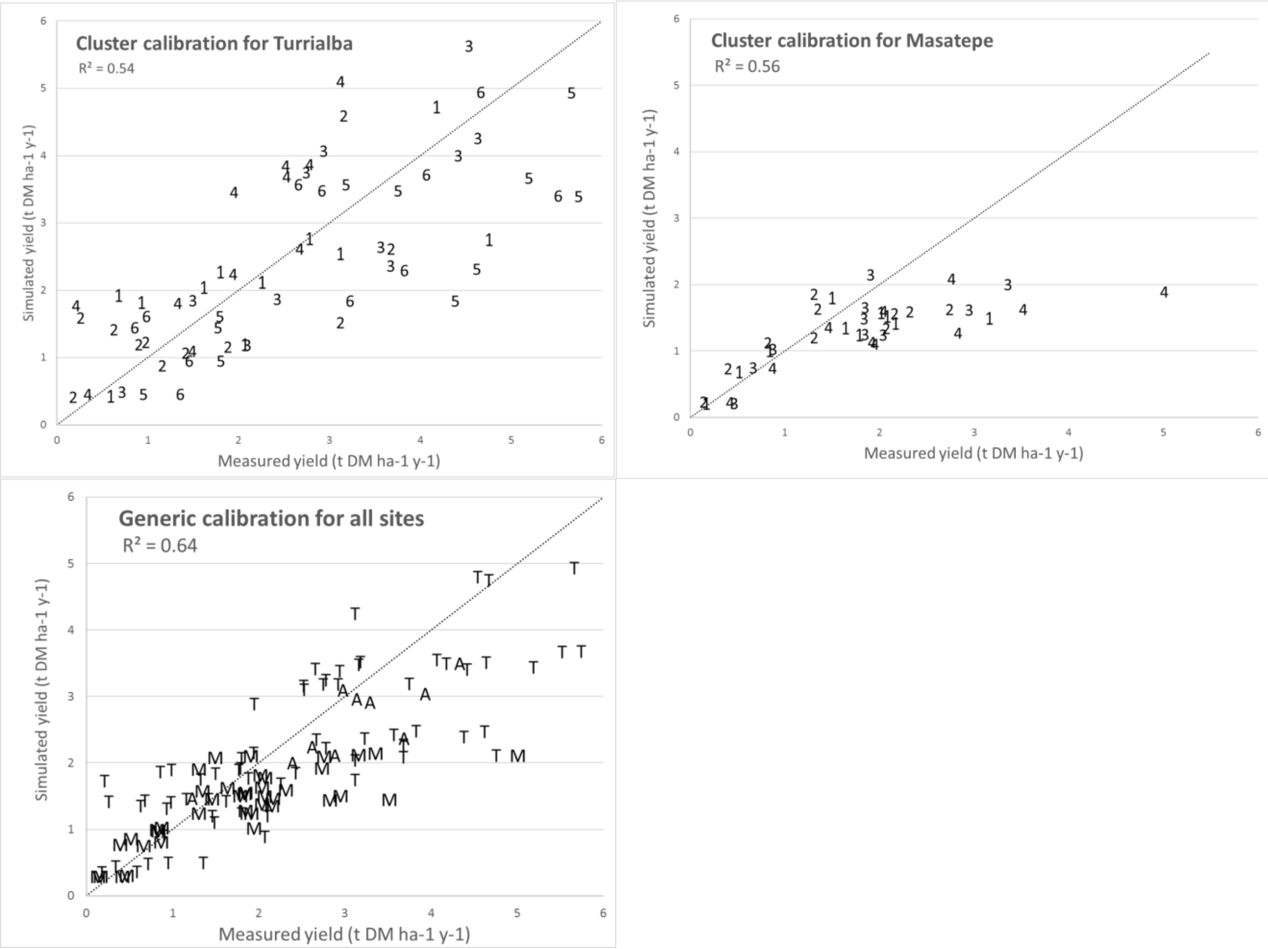

Figure 6. Simulated vs. measured coffee productivity (t DM ha^{-1} y^{-1}) after two types of multi-site calibration. Top two panels show results for the posterior mode from cluster calibration, the bottom panel for the posterior mode from generic calibration. The digits on the top panels identify the six different sites in Turrialba and the four sites in Masatepe (see Table 1). The letters in the bottom panel identify the Aquiares site and the Turrialba and Masatepe clusters.

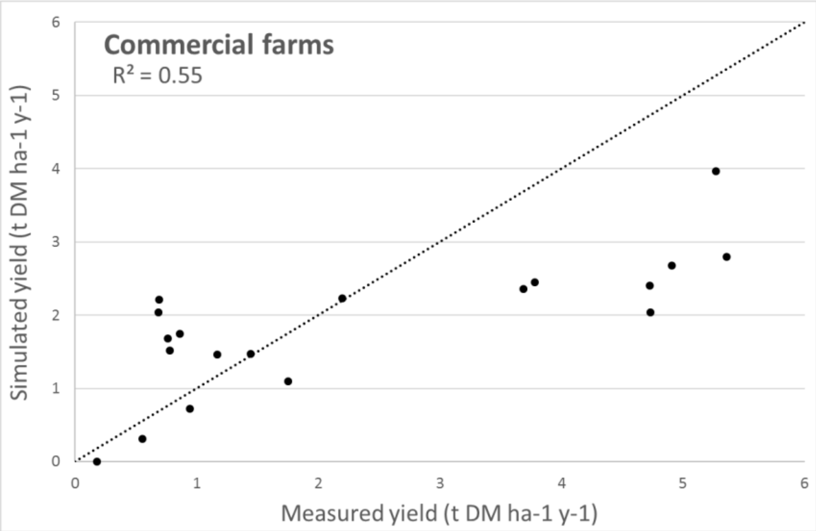

Figure 7. Simulated vs. measured coffee productivity (t DM ha^{-1} y^{-1}) for commercial farms in Nicaragua.