Article (refereed) - postprint

Minunno, F.; van Oijen, M.; Cameron, D.R.; Cerasoli, S.; Pereira, J.S.; Tomé, M.. 2013 Using a Bayesian framework and global sensitivity analysis to identify strengths and weaknesses of two process-based models differing in representation of autotrophic respiration. *Environmental Modelling & Software* , 42. 99-115. 10.1016/j.envsoft.2012.12.010

Using a Bayesian framework and global sensitivity analysis to identify strengths and weaknesses of two process-based models differing in representation of autotrophic respiration

(Minunno[1] F., van Oijen[2] M., Cameron[2] D.R., Cerasoli[1] S., Pereira[1] J.S., Tomé[1] M.)

[1] = Forest Research Center, Faculty of Agricultural Sciences, Technical University of Lisbon, Portugal.

[2] = CEH-Edinburgh, Bush Estate Penicuik EH26 0QB, U.K.

*** = Corresponding author francescom@isa.utl.pt**

Introduction

During the last decades, forests have been experiencing fast changes in the environmental conditions, to which forest management must adapt. Process-based models (PBMs), based on eco-physiological principles, are invaluable tools for sustainable and adaptive forest management (Fontes et al., 2010). PBMs allow for the estimation of site productivity and can simulate the effects of management and environmental constraints on stand growth and the probable influence of climate change on forest productivity. Furthermore PBMs enable analyses at different spatial and temporal scales (Fontes et al., 2010). However, calibration of PBMs is often difficult because they tend to have many parameters and outputs for which only few data are available. Moreover, because models are simplifications of reality, we need to assess carefully how well their structure allows for simulation of the phenomena of interest. Bayesian statistics, based on probability theory, offers an alternative to the calibration problem and can provide parameter estimates with estimates of their uncertainty (van Oijen et al., 2005). The Bayesian approach also allows for the evaluation of model structure by quantifying the extent to which data support different models (Kass & Raftery, 1995; van Oijen et al., 2011). In addition, the increasing availability of eddy-covariance measurements with high temporal resolution (Pereira et al., 2007) provided by the Fluxnet and other regional networks, allows for calibration as well as for model validation.

In this work a Bayesian framework and a global sensitivity analysis were used in combination to test an improvement of a process-based model (3PGN (Xenakis et al., 2008)) and to study model behaviour. Two versions of 3PGN that differ in their representation of autotrophic respiration (R_{aut}) were calibrated and evaluated. 3PGN is based on a constant value of carbon–use efficiency (CUE), defined as the ratio between net primary production (P_N) and gross primary production (P_G) (Gifford, 2003); therefore, R_{aut} is modeled as a fixed proportion of P_G. The understanding of the factors regulating R_{aut} is one of the most challenging questions in ecological forest research. Many studies argue that $P_N : P_G$ is constant (Dewar et al., 1998; Gifford, 1994, 2003). Waring et al. (1998) proposed a universal value of 0.47 for most forests. More recently, van Oijen et al. (2010), using a mathematical approach based on the law of conservation of mass, showed that $P_N : P_G$ is narrowly constrained. However, owing to the difficulty in measuring carbon-use efficiency and in particular the P_G

component, methodological problems can mask variation in $P_N : P_G$ (Medlyn & Dewar, 1999), casting doubts about the existence of fixed values of the ratio between net and gross primary production. DeLucia et al. (2007), conducting a literature review, found that CUE varied between 0.23 and 0.83 across 60 different forests, with an average of 0.53.

A different approach is to model R_{aut} as the sum of two components: maintenance (R_{maint}) and growth (R_{growth}) respiration, the first being proportional to the live biomass and its temperature, the second being proportional to P_N. This theory was developed in the 1970s by McCree (1974), and many authors followed this approach (e.g., Penning de Vries, 1974, 1975; Ryan & Waring, 1992). A detailed review of the progress achieved in respiration modeling over the last decades can be found in Amthor (2000). Warmer climates should have higher respiration costs, because the maintenance respiration increases exponentially with temperature (Ryan, 1991). This kind of R_{aut} modeling ($R_{maint} + R_{growth}$) has been used in many process-based models (e.g. CABALA (Battaglia et al., 2004); PIXGRO (Adiku et al., 2006); MAESTRO (Wang & Jarvis, 1990)).

In the present work the original version of 3PGN, based on a constant $P_N : P_G$ ratio, and a new version (3PGN*), in which R_{aut} is modeled as the sum of maintenance and growth plant respiration, were calibrated and evaluated under a Bayesian framework. As proposed by van Oijen *et al.* (2011), the Bayesian framework consisted of model calibration, model comparison and analysis of model-data mismatch. Sensitivity analyses of the two model versions were also carried out to have a better insight of model behavior (Campolongo et al., 2007). For the first time a Bayesian framework and a global sensitivity analysis, Morris method (Morris, 1991), were used in combination to highlight the strengths and weaknesses of the two model versions and to evaluate their performances.

Materials and methods

Overview of the methodology

Our study used eddy-covariance data and forest measurements collected at two different sites: a CarboEurope-IP site (Espirra forest) and a field experiment (Furadouro experiment). At a first stage both models were calibrated using the full dataset (i.e., Espirra forest and Furadouro experiment). The Bayesian framework proposed by van Oijen *et al.* (2011) and the Morris method were used in combination to better understand the behaviour of the models.

Subsequently, two Bayesian model comparisons were performed to evaluate the models. The first BMC was carried out in light of the prior knowledge of the two models (*prior BMC*). Meanwhile, for the second BMC part of the dataset was used for model calibration and the rest of the data were used for model evaluation (*post BMC*). For the *prior BMC* 1000 parameter vectors were sampled from the prior distributions of the two model versions. The models were run with the sampled parameter sets and the distributions of model outputs were used in a Bayesian model comparison. For the *prior BMC* the models were compared in light of the full dataset (i.e., Espirra forest and Furadouro experiment). For the second Bayesian model comparison, the models were calibrated with the Furadouro experiment data and then compared using the Espirra forest dataset.

3PGN Structure

3PGN was developed by Xenakis *et al.* (2008) coupling two models, 3-PG (Physiological Principles in Predicting Growth) and ICBM (Introductory Carbon Balance Model). The resulting model structure was comprehensively described by Xenakis *et al.* (2008) – only a brief outline is given here.

A detailed description of 3-PG was provided by Landsberg and Waring (1997) and by Sands and Landsberg (2002). 3-PG is composed of five sub-models. One is used to calculate the productivity of the stand and another is used for partitioning biomass between different organs (foliage, roots and stem). The other three sub-models are used to determine the changes in stem number, soil water balance and variables of interest to forest managers, such as stand timber volume (V, m^3 ha^{-1}), mean diameter at breast height (D, cm) and stand basal area.

3-PG is based on the principle that the net primary production of a stand is primarily determined by radiation interception. P_G is calculated by multiplying the fraction of the photosynthetically active radiation absorbed by the stand (Φ_{aPAR}) with canopy quantum efficiency (α_c). Φ_{aPAR} is calculated using Beer's law. The canopy quantum efficiency is calculated by multiplying a theoretical maximum canopy quantum efficiency (*alpha*) with an array of site and physiological modifiers that vary between 0 and 1 (functions of atmospheric vapor pressure deficit, air temperature, frost, water balance, age and fertility rating (*FR*)). P_N is calculated as a constant fraction (Y) of P_G (Law et al., 2000; Waring et al., 1998). The carbon allocation routine sub model is based on allometric equations, on a single-tree basis. A fraction of P_N is allocated below-ground by a root allocation coefficient that is affected by soil fertility. The remaining biomass is partitioned between the aboveground organs as a function of diameter at breast height and foliage: stem ratio.

The 3-PG model has been applied to many different species and sites and it is widely used in research as well as by companies to assess forest growth and site productivity. Fontes et al. (2006) parameterized 3-PG for Portuguese plantations of *Eucalyptus globulus*, Labill., demonstrating that carbon allocation of *E. globulus* in Portugal differs strongly from allocation patterns in Australian plantations.

A complete description of ICBM is provided by Andrén and Kätterer (1997) and Kätterer and Andrén (Kätterer & Andrén, 1999, 2001). ICBM/2N considers three pools of C and three pools of N in the soil, consisting of different forms of organic matter: the "young labile" pool, that includes small tree detritus (such as litterfall and root turnover), a "young refractory" pool, that includes coarse woody detritus (coarse root, branches and stems) and an "old" pool, that includes the recalcitrant organic matter. Each pool has a different decomposition rate that varies along the year with environmental conditions (i.e., temperature and soil water content), but does not change during stand development (Mäkelä & Vanninen, 2000; Titus & Malcolm, 1999). Carbon decomposed from the young pools enters the old pool at a constant relative rate of humification. The fraction from each young poolthat is decomposed but not humified is considered as respiratory loss. Similarly, decomposition losses take place from the "old" pool. The sum of all the out-fluxes from the three pools gives the heterotrophic respiration. The nitrogen balance is based on fixed C:N ratios and the size of the C fluxes and pools.

In 3PGN, the biomass losses of the stand (litterfall, root turnover, death of trees, but excluding tree harvesting), calculated by 3-PG, are the inputs for ICBM/2N. The latter model is used to calculate the heterotrophic respiration, but not the site fertility parameter (*FR*) of 3-PG. As in the original version of 3-PG (Landsberg and Waring, 1997), the *FR* parameter was site specific. In this work, five different FRs were parameterised for each site by means of Bayesian calibration.

The two versions of 3PGN

In the two 3PGN versions used in this work, tree diameter D was calculated as a function of total aboveground dry biomass (*i.e.*, leaves included).

$$D = StCn * W_{\mathrm{abv}}^{\ StPw} \tag{1}$$

where W_{abv} is the aboveground biomass (kg per tree) and $StCn$ and $StPw$ are regression coefficients.

Because average stand height (H) is an important stand variable, a new equation for the calculation of H was introduced.

$$H = aH * W_{\mathrm{abv}}^{bW} \tag{2}$$

where aH and bW are regression coefficients.

The two model versions used in this work calculate autotrophic respiration (R_{aut}) in different ways. In the old version (3PGN), R_{aut} is proportional to photosynthesis. In the new version (3PGN*), R_{aut} is the sum of respiration for maintenance (R_{maint}) and for growth (R_{growth}):

$$R_{\mathrm{aut}} = R_{growth} + R_{maint} \tag{3}$$

Maintenance respiration is assumed to be a function of biomass and average temperature (T_{av}) and it follows different specific rates for the woody (r_w) and foliage (r_f) tissues. In the woody pool the branches, stem and the root biomass were included.

$$R_{maint} = \sum W_i\, r_i\, Q_{10}^{(\mathrm{Tav}-20)/10} \tag{4}$$

where W_i and r_i are dry weight and specific respiration rate, respectively, of the *i*th plant pool (woody or foliage); Q_{10} determines the temperature responsiveness of respiration.

Growth respiration is calculated as:

$$R_{growth} = r_g * (P_G - R_{maint}) \tag{5}$$

where r_g is the fraction of growth discarded as respiration (Penning de Vries 1975).

Finally, P_N is calculated as:

$$P_N = P_{\mathrm{G}} - R_{\mathrm{aut}} \quad (6)$$

When the calculated $R_{maint} + R_{growth}$ exceed P_G total R_{aut} is set equal to P_G.

The equations 4 and 5 were chosen because they required fewer parameters than other R_{maint} and R_{growth} modeling approaches (Amthor, 2000; Ryan et al., 1996). With the insertion of the new equations, just three additional parameters were entered into the model, maintaining model simplicity, in agreement with the idea on which 3-PG was developed (Landsberg, 2003; Landsberg & Waring, 1997).

Experimental sites and data acquisition

The data used for model calibration and evaluation were collected at two sites: Espirra and Furadouro. The Espirra forest dataset consisted of measurements of net ecosystem production (P_E, Mg C ha^{-1} y^{-1}), mean stand height (H, m) and mean stand diameter at breast height (D, cm). The dataset from the Furadouro experiment consisted of measurements of foliage (WF, Mg of dry mass (DM) ha^{-1}), stems (WS, Mg DM ha^{-1}) and roots (WR, Mg DM ha^{-1}), stand volume, mean stand height and mean diameter at breast height. The whole dataset consisted of 305 data points between the seven output variables considered (i.e., P_E, D, H, V, WF, WR and WS).

Espirra forest

The carbon fluxes, from which P_E was derived, were measured by eddy covariance (Aubinet et al., 1999; Baldocchi, 2003) in Espirra (Pereira et al. 2007). This CarboEurope-IP site is a 300 ha *Eucalyptus globulus* plantation (38º38`N, 8º36`W) tended as a coppice. Originally planted in 1986 at 3 x 3 m spacing, ca. 1100 trees ha^{-1}, was 11 years old (2nd rotation) in the end of the period analyzed, and ca. 20 m height. The mean annual temperature for the site is 16ºC whereas the mean annual precipitation is 709 mm, more than 80% of which occurs from October to April.

The flux data were collected between October 2002 and December 2005 at the half hourly scale. Net ecosystem production data were aggregated at monthly time step and used for model calibration and validation. Flux data quality control followed the CarboEurope-IP recommendations; gap filling was performed according to Reichstein et al. (2005).

Furadouro experiment

The mensurational data used for model calibrations were collected in a field experiment installed from 1986 to 1992 at Quinta do Furadouro (Óbidos, Portugal, 39º29`N, 9º13`W, 30 m a.s.l.). The mean annual temperature is 15.2 ºC and the mean annual precipitation is 607 mm, but less than

10% occurs between May and September. Three months old *E. globulus* seedlings were planted at 3 x 3 m spacing; each seedling was supplied at planting with 200g of a commercial fertilizer containing 14.0g of N, 18.3g of K and 11.6g of P. Before planting, the soil was ploughed at 80 cm depth and 1.5 Mg ha^{-1} of dolomitic limestone (66.5% of CaCO3, 32.5% of MgCO3) was applied.

The experimental design consisted of three treatments and a control. The treatments were daily irrigation from April to October (I), application of a pelleted fertilizer in March and October of each year (F) and daily irrigation as in I, combined with a liquid fertilizer solution once a week (IF). No fertilization (except the initial amount at plantation) and irrigation were supplied to the control (C).

The differences in soil nitrogen concentration between C and I were due to some amount of N contained into the irrigation water; while the different amounts of nutrient in F and IF resulted both from the influence of irrigation water and from different application rates. For these reasons, the fertility rate parameter of 3-PG was calibrated independently for each treatment. Different prior were assigned to the FR of C, F, IF; Table 1b shows the minimum and maximum values and the distributions of these parameters.

Sensitivity analyses

Sensitivity analyses can vary from the simplest class of the One Factor At the Time (OAT) to global sensitivity. While OAT quantifies model output variation in relation to changes of one factor at a time, global sensitivity analyses evaluate model's output sensitivity to simultaneous changes in several factors. In this work the global sensitivity method proposed by Morris (1991) was adopted. This method is a good compromise between efficiency and accuracy and it is particularly well-suited when a high number of factors are considered and/or the model is costly to compute (Campolongo et al. 2007).

The method consists of computing basic statistics, i.e., mean (μ) and standard deviation (σ), from the distribution of a number of incremental ratios, called Elementary Effects. μ gives the overall importance of an input factor, while σ describes non-linear effects and interactions between factors. For a more detailed analysis of this methodology see Campolongo et al. (2007) and Morris (1991).

Campolongo et al. (2007) enhanced the Morris method improving the sampling strategy and proposed to calculate the mean of the distribution of the absolute values of the elementary effects (μ*). μ* is calculated to solve the problem of non-monotonic models, where the effects of opposite signs could mask the importance of a factor.

For the sensitivity analyses of 3PGN and 3PGN* we considered the following output: stem, foliage and root biomasses, average stand diameter at breast height, average stand height, stand volume and annual net ecosystem productivity (aP_E). Because output sensitivity to the factors could vary across stand development, the sensitivity was computed at different stand ages (i.e., at four, eight and twelve years). For the sensitivity analysis the environmental data (weather, soil, management) of the Espirra forest were used as drivers for the models. The factors involved in the analysis consisted in the parameters and the site variables reported in Table 1a-b. Factors ranged between the minimum and maximum values used for the *BC* (Table 1).

Bayesian framework

Model calibration and comparison were carried out using a Bayesian approach. Bayesian statistics is part of probability theory and it requires that beliefs about parameter values and models be expressed as probability distributions. Our initial information about plausible parameter values, and about which model is correct, is expressed in the *prior* distribution P(θ). Observed data (*O*) that are used to update the prior distribution enter the analysis through the so-called *likelihood* function L(θ) = P(*O*|θ). An updated, *posterior* distribution is then found by application of Bayes' Theorem:

$$P(\theta|O) = c\ P(O|\theta)\ P(\theta) \qquad (7)$$

where, $c = p(O)^{-1}$. The value c is fixed, and usually it is not necessary to compute it explicitly.

Likelihood function

The likelihood function (L) used was proposed by Sivia (2006) and it is described by the equations (8) and (9):

$$P(O|\theta) = \prod_{i=1}^{N} \frac{1}{\sigma_i\sqrt{2\pi}} \frac{1-exp\left(-R_i^2/2\right)}{R_i^2} \qquad (8)$$

$$R = (\mathrm{sim}(\theta) - O)/\sigma \qquad (9)$$

where, sim(θ) is the output from the model for parameter values θ, N is the number of data points and σ is the uncertainty about the random error of the *i*th data point.

This likelihood was chosen because it is heavy-tailed, so it puts less weight on the outliers that can occur in eddy covariance measurements (Sivia 2006, Van Oijen et al. 2011).

Prior distribution for the parameters

Table 1 shows the types of distribution and their bounds that were used for the prior marginal distributions of the parameters. The prior was assigned using different sources of information: literature, measurements and posteriors from previous Bayesian calibrations.

For parameters for which knowledge is scarce the uniform distribution was chosen. The truncated Gaussian distribution was assigned to many of the other parameters, using information derived from literature. Those distributions were also quite uninformative (not too peaked).

The prior distributions of the woody and foliage specific respiration rates (r_w and r_f, respectively) of eq. 5 were fitted with gamma distributions, on the basis of spot gas exchange measurements collected at the Nicolaus site, close to the Espirra forest (Cerasoli et al., 2009).

Since data were available to calibrate the allometric equations (eq. 1 and 2) and the 3PGN equation to calculate the specific leaf area (*SLA*) as time function (eq. 10), Bayesian calibrations were carried out independently for those equations.

$$SLA(t) = SLA1 + (SLA0 - SLA1) * e^{-(\ln 2)(t / tSLA)^2} \quad (10)$$

where t is the stand age, *SLA0* is the specific leaf area at age 0, *SLA1* is the specific leaf area for mature leaves, *tSLA* is the age at which *SLA* = (*SLA0* + *SLA1*)/2.

After Bayesian calibration (*BC*), the posterior distributions of the parameters of eq. 1, 2 and 10 were fitted with Weibull, normal and gamma distributions and then used as prior for the *BCs* of the whole models.

Bayesian calibration (BC)

Bayesian calibration revises the state of knowledge about parameter values using new data. Process based models are not analytically solvable and they need to be run to quantify the likelihood. Therefore, to summarize the posterior distribution as a sample, from which we can calculate summary statistics like the posterior mean, we used the version of Markov Chain Monte Carlo (*MCMC*), known as the Metropolis-Hastings random walk (Robert & Casella, 2004). The *MCMC* method aims to converge the sampling on the region of the parameter space with highest probability density. A complete description of the Metropolis-Hastings algorithm is given in van Oijen et al. (2005).

To optimise the MCMC-algorithm, some preliminary calibrations were carried out, varying the chain length and the scale of the proposal distribution, in order to achieve efficient convergence of the Markov chain. BCs were carried out with a chain length of 100,000 and 500,000 and the burn-in was 40% of the chain length. To assess convergence of iterative simulations, the Gelman-Rubin criterion (Gelman & Rubin, 1992) was used. This method consists in comparing at least two independent simulated sequences, checking if the variance within the chains is comparable with the variance between the chains. To monitor convergence, the potential scale reduction ($\check{R}$) is estimated; $\check{R}$ tends to 1 when we have a good inference about the target distribution. Gelman et al. (2004) stated that for the majority of the cases a value of 1.1 for Ř is acceptable, but in some cases a higher level of precision may be more appropriate. Three chains were considered to evaluate convergence; after the BCs, all chains, discarding the burn in, were joined and treated as a unique sample from the target distribution.

Bayesian model comparison (BMC)

Bayesian model comparison is a powerful extension of *BC* that allows for the evaluation of different model structures on the basis of their relative likelihoods (Kass & Raftery 1995; van Oijen et al. 2011). In this case the Bayesian theorem is not applied over the parameter space of a single model but over a set of models (M) (van Oijen et al. 2005).

$$P(M_k|O) = P(O\,|M_k)\,P(M_k) \,/ \sum P(O\,|M)\,P(M) \quad (11)$$

where k varies between 1 and n models. In our application, with just two model versions being compared, n=2.

Assuming no initial preference for either of the models ($P(M_1) = \ldots = P(M_n)$), equation 11 becomes:

$$P(M_k|O) = P(O|M_k) / \sum P(O|M) \qquad (12)$$

$P(O|M)$ is the “integrated likelihood” (IL) which is defined over the whole parameter space of M, i.e., $P(O|M) = \int P(O|\theta)P(\theta)d\theta$.

Analysis of model-data mismatch

The Bayesian model comparison treats models as black boxes, giving just indication about which model is more plausible (van Oijen et al. 2011). The mismatch of simulated vs. observed data can also be evaluated using more classical methods that allow identifying model weakness. For each of the seven outputs considered, normalised root mean squared error (*NRMSE*) and squared correlation coefficient (r^2) were calculated across the range of prior and posterior distributions.

Moreover, for the modes of the prior and posterior distributions we calculated the mean squared error (*MSE*) of each output. *MSE* was decomposed in three components as suggested by Kobayashi and Salam (2000):

$$MSE = \overline{(S-O)^2} = (\bar{S}-\bar{O})^2 + (\sigma_S - \sigma_O)^2 + 2(\sigma_S\sigma_O)(1-r) \qquad (13)$$

where S are model predictions and O are the observed data, σ_S and σ_O are their respective standard deviations, and r is the correlation between simulated and observed data.

The first component of *MSE* is a measure of the average deviation of the simulations from the data (i.e., bias error), the second element indicates if the model is able to catch the variability of the data (i.e., variance error) and the third element expresses the ability of the model to reproduce the pattern of the fluctuations among the data (i.e., phase shift error) (Kobayashi & Salam, 2000).

Results

Sensitivity analyses

The Morris method allowed for the identification of the key parameters for each of the model output across the stand development. Note that the sensitivity analysis was contingent on the parameter space considered for the Bayesian calibration, because parameters varied between the minimum and maximum values used in the BC. Part of the results about sensitivity analysis are reported in Figure 1a-b, where μ* and σ of the five factors at which model outputs are most sensitive (highest μ*) are plotted for each year considered (i.e., 4, 8, 12). More comprehensive results were difficult to report in graphs and tables because of the high number of parameters, therefore general results were only discussed in the text.

Below a general overview of the sensitivity analysis results is given. The fertility rate parameter (*FR*) had a strong impact on all the outputs of both models. The parameters related to the autotrophic respiration were also key factors; in particular, the $P_N : P_G$ ratio (*Y*) for 3PGN and the woody biomass respiration rate for 3PGN* had a high influence on all the output variables. 3PGN* outputs resulted also quite sensitive to Q_{10} and r_g, while less sensitive to r_f. Both models were highly sensitive to the light use efficiency parameter (i.e., *alpha*), the optimum temperature for growth (i.e., *Topt*) and the minimum available soil water (*minASW*). In the first part of stand development model outputs were highly influenced by the age at which canopy close (i.e., fullCanAge) To a lesser extent, model outputs were sensitive to parameters related to fertility (i.e., *fN0*), allometric parameters (i.e., StemPower and StemConst), allocation parameters (i.e., *pRx, pRn, pFS2, pFS20*), parameters related to temperature stress (i.e., *Tmin, Tmax*), parameters and variables related to water stress (i.e., *maxAWS, MaxCond, CoeffCond*), soil parameters (i.e., *klmax*) and other parameters like the litterfall rate at maturity (*gammaF1*). Low impact on model outputs was given by factors related to age stress (i.e., *nAge, MaxAge, rAge*), root turnover, soil parameters and variables (i.e., *hc, komax, krmax, O_C_i, Yl_C_i*), the initial biomass of stem and root (i.e., *WS, WR*), frost days.

Bayesian calibration

Bayesian calibration allowed for the updating of the joint probability distribution for the model parameters in light of the data used (i.e., Furadouro experiment, Espirra forest). Using *MCMC*-algorithms, convergence must be reached by all the parameters to obtain an accurate sample for the posterior distribution. For *BCs* of 100,000 chain length, the $\check{R}$ factor, calculated over three chains, assumed values lower than 1.1 for all the parameters (data not showed). However, almost 20% of the parameters did not assume the same marginal posterior distribution over the three chains. $\check{R}$ was lower or close to 1.03 for the *BCs* of 500,000 chain length. In this case all parameter marginal posterior distributions were similar over the three chains (data not showed). The *BCs* with different chain lengths showed that 500,000 chain length and $\check{R}$ factors lower than 1.03 were proper to reach a good convergence for parameter rich process-based models.

The likelihood distributions of the two model versions, for each output, before and after *BC*, are presented in Figure 2. Higher values of the likelihood correspond to better model performances, while the variance of the likelihood distribution is a measure of model accuracy.

BC significantly shifted the likelihood towards higher values for *D, H, WS* and *V*. This means that, after calibration, the models better simulated those variables. Posterior likelihoods of 3PGN and 3PGN* were also higher for P_E, even though, for this output the likelihood improvements were less pronounced. On the contrary, *WF* and *WR* likelihoods decreased after *BC*.

Figure 3 shows prior and posterior marginal distributions of all the parameters.

For the two model versions, in most of the cases, parameter posterior distributions were very similar (Figure 3). There were differences between the posterior distributions of parameters linked with water balance and water stress (i.e., *LAIgcx, Blcond, MaxIntcpntn, MinASW*) (Figure 3a), temperature and frost stress (i.e., *Topt, kF*) (Figure 3a), soil parameters (i.e., *klmax*) (Figure 3b), fertility parameters (i.e., *FR* and *fN0*) (Figure 3b). Posterior distributions were also different for *alpha*, *tSLA* and *fullCanAge* (Figure 3a).

From marginal posterior distribution it is possible to understand parameter uncertainty in light of the data used for *BC*; if the posterior variance is lower than the prior variance the data were informative for the parameter. The data used for *BCs* allowed for the reduction in the uncertainty of about 70% of the parameters (Figure 3). The data were not informative for some parameters related to temperature stress (i.e., *Tmax*) (Figure 3a), water stress (i.e., *SWpower, Blcond, LAImaxIntcptn*) (Figure 3a), age effect on forest growth (i.e., *MaxAge, rAge, nAge*) (Figure 3b), litterfall parameters (i.e., *gammaF0, tgammaF*) (Figure 3a), stand volume (i.e., *fracBB1, tBB*) (Figure 3b), stand attributes (i.e., the initial root biomass (*WR_i*) and tree density at plantation (*StemNo*)) (Figure 3b). Uncertainty also underpinned many parameters of the soil decomposition model (ICBM/2N): the decomposition rates of the different soil pools (i.e., *krmax, komax, hc*) and the initial soil carbon contents (i.e., *Yr_C_i, Yl_C_i, O_C_i*) (Figure 3b). The data were extremely informative for the allocation and allometric parameters (i.e., *pFS2, pFS20, StCn, StPw, pRx, pRn, aH* and *cD*) (Figure 3a), temperature parameters (i.e., *Tmin, Topt*) (Figure 3a), fertility parameters (i.e., *m0, FR*) (Figure 3b), the litterfall rate at maturity (*gammaFx*) (Figure 3a), water stress parameters (i.e., *MaxCond, CoeffCond*) (Figure 3a), light use efficiency and light interception parameters (i.e., *alpha, k, SLA0*) (Figure 3a), the age at which canopy close (*fullCanAge*) (Figure 3a) and initial biomass (i.e., *WF_i, WS_i*) (Figure 3b).

Upon examination of the posterior distribution of the parameters related to the autotrophic respiration, it is shown that the data were highly informative to *Y* (in 3PGN) and r_w (in 3PGN*), moderately informative to Q_{10} and r_f and uninformative to r_g (Figure 3a).

Analysis of model-data mismatch

For each output, *MSE* were calculated using the mode of the prior and posterior distributions (Figure 4). *BC* allowed for the reduction, to a varying extent, of the phase, variance and bias error of *D, V* and *WS*. Bias error for H was also strongly reduced, while phase and variance error slightly increased. WF *MSEs* increased after calibration, especially for 3PGN*; the highest component of WF *MSE* was the phase error. WR *MSE* slightly decreased for 3PGN, because *BC* reduced the variance error but increased the phase error. Instead, after *BC*, all WR *MSE* components significantly increased in 3PGN*. *BC* decreased the *MSE* of net ecosystem production, but the phase error remained quite high.

For each model, 1000 parameter vectors were sampled from the prior and posterior distributions to calculate the coefficient of correlation, the slopes and the normalized root mean squared error for the comparisons between the predicted and the observed data. Table 3 shows the mean r^2, slopes and *NRMSE* for both prior and posterior of the seven outputs.

The coefficient of correlation was high for all the output apart for P_E. Even if r^2*, the slope* and *NRMSE* of P_E improved after the calibration, the models were not able to reproduce the net ecosystem productivity pattern over the months. *BC* significantly improved all the statistics (i.e., r^2, slopes and *NRMSE*) for *D, WS* and *V*. This being said, model performances worsened for *WF* and *WR*.

Bayesian model comparison

Results regarding the Bayesian model comparison are summarised in Table 2 where the log-transformed integrated likelihood values are presented for the *prior BMC* and *post BMC*. The highest integrated likelihood indicates the most plausible model. The percentage probability of a model of being correct is obtained dividing the integrated likelihood of each model by the sum of the integrated likelihoods. In the *prior BMC* the integrated likelihood showed that the 3PGN* model had a probability of 84% of being the superior model. Also results from the *post BMC* supported the new model version, in this case 3PGN* had a 99% probability of being the superior model.

Discussion

For the first time in this work we showed how the Bayesian framework proposed by van Oijen et al. (2011) can be used to improve the structure of a process-based model. Furthermore the framework was strengthened with a global sensitivity analysis, to better explore strengths and weaknesses of the model. These techniques can be applied to any kind of model, simpler or more complicated than 3PGN. However, the use of the Bayesian framework for model of higher complexity can be hampered by computational limitations. In particular future works should search to increase the efficiency of the Bayesian calibration to reduce the computational costs. The BC efficiency can be increased reducing the number of parameters involved in the calibration by means of parameter screening or using more effective MCMC algorithms such as the delayed rejection adaptive Metropolis (Haario et al., 2006) and the differential evolution Markov chain (ter Braak & Vrugt, 2008).

Uncertainty and sensitivity.

Uncertainty and sensitivity analyses are fundamental processes that help to understand model behavior. Even though previous works (Esprey et al. 2004; Xenakis et al., 2008) already performed sensitivity analyses of 3-PG and 3PGN, using the simplest method of the One Factor At a Time (OAT) screening techniques, this was the first attempt to study 3PGN sensitivity using a global method. In this work the Morris sensitivity analysis was performed within the parameter space defined by the prior, instead of varying the parameters values of a certain fix percentage (Esprey et al. 2004; Xenakis et al. 2008). The minimum and the maximum values of the prior are ranges within which the parameters are meaningful. The prior represents the state of knowledge about the parameters before the calibration and it contains information coming from different sources such as literature, experimental data or previous Bayesian calibrations. In this way sensitivity analysis permits to focus the attention of the modelers on the parameter space that is meaningful and supported by previous evidence.

The sensitivity-analysis carried out at different ages helped to understand how the impact of the factors on model outputs varies across the rotation. Some of the parameters are more influential on the outputs at the beginning of the rotation (i.e., *fullCan Age, Topt*), while others, like the parameters related to water stress (i.e., *minASW, MaxCond, Swconst*), had a higher impact on the outputs at the end of the rotation. These results imply that having a dataset that spans across the stand development is crucial to achieving a good calibration of the models. For all 3PGN* output variables, the sensitivity to the wood respiration rate increased at the end of the rotation and this parameter became the most influential one, because r_w is related to the biomass that increases with age. For this reason, particular attention must be given to the parameterization of r_w. Furthermore, we are not considering in the autotrophic respiration model the percentage of the wood that do not contribute to R_{aut} (i.e., heartwood), because *E. globulus* plantations are usually managed with a 13 year rotation and the trees do not present heartwood at this stage or it is negligible. If the new version of 3PGN is applied to different species and to different Eucalyptus management, the percentage of heartwood must be taken into account.

The uncertainty in both parameters and model predictions was significantly reduced by the calibration. The degree of parameter uncertainty varied across the parameters but was similar between the two models. P_E measurements were particularly useful for model calibration because they reflected the seasonal variability of stand growth and for this reason they were more informative for the physiological parameters. Eddy-covariance data reduced the uncertainty of parameters related to the photosynthetic activity like water stress, light use efficiency and temperature stress parameters, while the biometric data (i.e., *D, H, V, WF, WR,* and *WS*) were mainly informative for parameters related to the allometry and the carbon allocation routine.

In the future, to reduce the uncertainty of parameters that remained less certain, modelers can work on the prior of those parameters or can use, in a future calibration, output variables that are highly sensitive to the uncertain parameters. Model simplification can also be considered if the parameters do not affect any of the output variables of interest.

The 3PGN and 3PGN* outputs characterized by the highest uncertainty a posteriori were foliage and root biomasses. These were the variables with fewest measurements, so the biomass datasets should be enriched correspondingly to decrease the degree of uncertainty.

Bayesian calibration and model-data mismatch:

In *BC* the bottom-up and the top-down approaches can be used in combination to improve the knowledge about parameters (Hartig et al., 2012). The bottom-up approach can be used in determining the prior, as we did for the respiration rates (i.e., r_w and r_f) and the parameters of the allometric equations. This approach allows for the integration of different data sources in the calibration process and it has the merit of redressing the parameters towards realistic values. In contrast, using a top-down approach, stand variables like *D*, *H* or *V* can be used, by means of the likelihood, to inform parameters that are highly variable or difficult to measure.

Bayesian calibration of the two model versions significantly reduced uncertainty in the outputs and parameters. Calibration improved the probability distributions of P_E, *D, H, V* and *WS, i.e.* the posterior likelihood distribution means were shifted towards higher values and the standard deviations were strongly reduced (Figure 2). The analyses on *MSE* confirmed the effectiveness of the calibration, with the posterior *MSE* being much lower than the prior *MSE,* for the majority of outputs. On the one hand, the highest reductions in MSE were achieved for the data that were more certain (i.e., *D, H*, stand volume and stem biomass) and with a high number of measurements. On the other hand, *MSE* just slightly decreased for the net ecosystem productivity and increased for foliage and root biomasses. Other works already demonstrated the weakness of 3-PG in predicting foliage biomass and LAI (Sands & Landsberg 2002). In our case, model failure could be explained by the fact that *WR* and *WF* were the data characterised by the highest uncertainty and the lowest number of measurements, therefore *WF* and *WR* had smaller weight on the likelihood than the other data. *BC* results suggested that foliage and root biomass dataset should be improved to better test if the models are able to reliably reproduce those data, otherwise model structure should be improved.

The decomposition of *MSE* provided additional useful information about model performances and structure in light of the data used. The models were not able to reliably reproduce P_E measurements. In fact, even if the models had really low bias and variance error, i.e. the models were able to capture the mean and the magnitude of the fluctuation among the measurements, they failed to simulate the pattern of the fluctuation (phase shift error), because r was low. In other words, the models are not able to reproduce the seasonal pattern of net ecosystem production. Model failure in predicting P_E can be explained by systematic and/or random errors in the measurements, a problem in the model structure or wrong settings of some parameter bounds in the prior. We expect the error to be mainly in model structure, as 3PGN was probably too simple to respond to all environmental changes that affect net ecosystem production, as shown by Minunno *et al.* (2010). Furthermore, the P_E dataset was characterized by one year of intensive drought (year 2005-2006) and simple models like 3PGN and 3PGN* are ill designed to capture forest responses to extreme events.

Bayesian model comparison

In ecological modelling there is a lively discussion about model complexity. Simple models are not able to reproduce the intricacies of the ecological mechanisms, while complicated models are theoretically closer to real processes. Nevertheless, it is difficult to calibrate parameter rich models, because of lack of data or the difficulty in measuring variables related to the parameters. This is not a negligible aspect as simple models with well-known parameters might perform better than complicated ones. There is a need to find a compromise between model complexity and parameter uncertainty, in accordance with the amount of data that are available. Therefore, model implementation should always take into account these two aspects. *BMC* is such a method that allows for the evaluation of model performances across their whole parameter distribution, in light of the data used. Even if this method has already been applied in ecological sciences, this application to parameter rich forest process-based models is still a novelty. As far as we are aware, only van Oijen *et al.* (2011) already implemented *BMC* for the evaluation of four biogeochemical models in a Norway spruce forest, while this work uses, for the first time, Bayesian model comparison to evaluate improvements in model structure.

The Bayesian model comparison of 3PGN and 3PGN* showed that the new version of the model performed better, even though it increased model complexity, adding three new parameters. Although 3PGN* autotrophic respiration model is slightly more complicated than the $P_N : P_G$ ratio used in 3PGN, 3PGN* parameter uncertainty is not necessarily higher. In fact, wood and foliage respiration rates might be easier to measure than the $P_N : P_G$ ratio, because of the difficulty to reliably measure P_G.

The marginal posterior distributions of the parameters that are common to the two models gave additional information about model structure. Posterior distributions (Figure 3) are not significantly different for the majority of the parameters, however the parameters that assumed significantly different marginal posterior distribution between the two model versions were the parameters at which the output variables are most sensitive (i.e., *alpha*, *MinASW*, *Topt*, *fullCanAge* and *FR*). Therefore, the new autotrophic model produced strong changes to the 3PGN structure, because the autotrophic respiration parameters, in particular *Y* for 3PGN and r_w for 3PGN*, have strong influence on the model output variables.

In conclusion, our results supported the new version of 3PGN. It should be noted, however, that models are always incorrect because they are a simplification of real processes and model performances cannot be discussed in an absolute manner (Oreskes et al., 1994). Thus, our analyses and probabilities of correctness must be considered as indicative information towards plausible model structures (van Oijen et al. 2011).

Conclusions

In this work, different methods (i.e., *BC*, *BMC*, *MSE*-decomposition and the Morris method) were used in combination for the first time to evaluate improvements in the structure of a process-based model. Our results showed that the new version of the 3PGN model, with the new algorithm for autotrophic respiration based on maintenance and growth respiration, has a higher conditional probability of being correct. Overall, the three operations of the Bayesian framework (Bayesian calibration, Bayesian model comparison and the analysis of model-data mismatch) in combination with the Morris method, allowed us to reduce uncertainties in parameters and outputs, and identify the weaknesses of the two 3PGN versions. Furthermore, the Bayesian approach allowed to identify the weaknesses and strengths of the dataset used, making possible the improvement and optimization of future data collection.

The analyses on model-data mismatch showed that both versions of the model are able to reliably predict average stand diameter at breast height, average stand height, stand volume and stem biomass. However, the models were unable to accurately predict foliage and root biomass, probably because the dataset was small and characterized by high uncertainty. Net ecosystem production was also not well predicted, because of uncertainty in the data but also due to model structural errors.

The efficiency of the *MCMC* algorithm should be enhanced to reduce the chain length and make the process less time consuming. In our study with process-based models rich in parameters, good convergence of all parameters is reached when the potential scale reduction ($\check{R}$) assumes values close to 1.03.

Acknowledgements

We thank the EU for support through COST Action FP603 and for a short-term scientific mission of F. M. to CEH. We also thank the FCT-Portugal for F.M. PhD fellowship (SFRH/BD/65934/2009).

References

Adiku, S., Reichstein, M., Lohila, A., Dinh, N., Aurela, M., Laurila, T., Lueers, J., et al. (2006). PIXGRO: A model for simulating the ecosystem CO exchange and growth of spring barley. *Ecological Modelling*, *190*(3-4), 260-276. doi:10.1016/j.ecolmodel.2005.04.024

Almeida, A. C., Siggins, A., Batista, T. R., Beadle, C., Fonseca, S., & Loos, R. (2010). Mapping the effect of spatial and temporal variation in climate and soils on Eucalyptus plantation production with 3-PG, a process-based growth model. *Forest Ecology and Management*, *259*(9), 1730-1740. doi:10.1016/j.foreco.2009.10.008

Amthor, J. (2000). The McCree–de Wit–Penning de Vries–Thornley Respiration Paradigms: 30 Years Later. *Annals of Botany*, *86*(1), 1-20. doi:10.1006/anbo.2000.1175

Andrén, O., & Kätterer, T. (1997). ICBM: the introductory carbon balance model for exploration of soil carbon balances. *Ecological Applications*, *7*(4), 1226-1236. doi:10.1890/1051-0761(1997)007[1226:ITICBM]2.0.CO;2

Aubinet, M., Grelle, A., Ibrom, A., Rannik, Ü., Moncrieff, J., Foken, T., Kowalski, A. S., et al. (1999). Estimates of the Annual Net Carbon and Water Exchange of Forests: The EUROFLUX Methodology. In A. H. F. Research & D. G. R. B. T.-A. in Ecological (Eds.), (Vol. Volume 30, pp. 113-175). Academic Press. Retrieved from http://www.sciencedirect.com/science/article/pii/S0065250408600185

Baldocchi, D. (2003). Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. *Global Change Biology*, *9*(4), 479-492. doi:10.1046/j.1365-2486.2003.00629.x

Battaglia, M. (2004). CABALA: a linked carbon, water and nitrogen model of forest growth for silvicultural decision support. *Forest Ecology and Management*, *193*(1-2), 251-282. doi:10.1016/j.foreco.2004.01.033

Campolongo, F., Cariboni, J., & Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, *22*(10), 1509-1518. doi:10.1016/j.envsoft.2006.10.004

Cerasoli, S., McGuire, M. A., Faria, J., Mourato, M., Schmidt, M., Pereira, J. S., Chaves, M. M., et al. (2009). CO2 efflux, CO2 concentration and photosynthetic refixation in stems of *Eucalyptus globulus* (Labill.). *Journal of experimental botany*, *60*(1), 99-105. doi:10.1093/jxb/ern272

DeLucia, E. H., Drake, J. E., Thomas, R. B., & Gonzalez-Meler, M. (2007). Forest carbon use efficiency: is respiration a constant fraction of gross primary production? *Global Change Biology*, *13*(6), 1157-1167. doi:10.1111/j.1365-2486.2007.01365.x

Dewar, R. C., Medlyn, B. E., & McMurtrie, R. E. (1998). A mechanistic analysis of light and carbon use efficiencies. *Plant, Cell and Environment*, *21*(6), 573-588. doi:10.1046/j.1365-3040.1998.00311.x

Esprey, L., Sands, P., & Smith, C. W. (2004). Understanding 3-PG using a sensitivity analysis. *Forest Ecology and Management*, *193*(1-2), 235-250. doi:10.1016/j.foreco.2004.01.032

Fontes, L., Bontemps, J.-D., Bugmann, H., Van Oijen, M., Gracia, C., Kramer, K., Lindner, M., et al. (2010, May). Models for supporting forest management in a changing environment. *Forest Systems*. Retrieved from http://84.88.0.49/index.php/IA/article/view/9315

Fontes, L., Landsberg, J. J., Tomé, J., Tomé, M., Pacheco, C. A., Soares, P., & Araujo, C. (2006). Calibration and testing of a generalized process-based model for use in Portuguese eucalyptus plantations. *Canadian Journal of Forest Research*, *36*(12), 3209-3221. NRC Research Press Ottawa, Canada. doi:10.1139/x06-186

Gelman, A, Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian Data Analysis.* (C. & Hall/CRC., Ed.) (Second., p. 668).

Gelman, A, & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*(4), 457-511. Institute of Mathematical Statistics. Retrieved from http://www.jstor.org/pss/2246093

Gifford, R. M. (1994). The Global Carbon Cycle: a Viewpoint on the Missing Sink. *Australian Journal of Plant Physiology*, *21*(1), 1. doi:10.1071/PP9940001

Gifford, R. M. (2003). Plant respiration in productivity models: conceptualisation, representation and issues for global terrestrial carbon-cycle research. *Functional Plant Biology*, *30*(2), 171. doi:10.1071/FP02083

Haario, H., Laine, M., Mira, A., & Saksman, E. (2006). DRAM: Efficient adaptive MCMC. *Statistics and Computing*, *16*(4), 339-354. doi:10.1007/s11222-006-9438-0

Hartig, F., Dyke, J. G., Hickler, T., Higgins, S., O'hara, R. B., Scheiter, S., & Huth, A. (2012). Connecting dynamic vegetation models to data - an inverse perspective. *Journal of Biogeography*, *submitted*. Retrieved from http://eprints.soton.ac.uk/337710/

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *430*(90), 773-795. American Statistical Association. Retrieved from http://www.jstor.org/pss/2291091

Kobayashi, K., & Salam, M. U. (2000). Comparing Simulated and Measured Values Using Mean Squared Deviation and its Components. *Agronomy Journal*, *92*, 345-352.

Kätterer, T., & Andrén, O. (1999). Long-term agricultural field experiments in Northern Europe: analysis of the influence of management on soil carbon stocks using the ICBM model. *Agriculture, Ecosystems & Environment*, *72*(2), 165-179. doi:10.1016/S0167-8809(98)00177-7

Kätterer, T., & Andrén, O. (2001). The ICBM family of analytically solved models of soil carbon, nitrogen and microbial biomass dynamics — descriptions and application examples. *Ecological Modelling*, *136*(2-3), 191-207. doi:10.1016/S0304-3800(00)00420-8

Landsberg, J. J. (2003). Modelling forest ecosystems: state of the art, challenges, and future directions. *Canadian Journal of Forest Research*, *33*(3), 385-397. NRC Research Press Ottawa, Canada. doi:10.1139/x02-129

Landsberg, J. J., & Sands, P. (2010). *Physiological Ecology of Forest Production, Volume 4: Principles, Processes and Models (Terrestrial Ecology)* (p. 352). Academic Press. Retrieved from http://www.amazon.com/Physiological-Ecology-Forest-Production-Volume/dp/0123744601

Landsberg, J. J., & Waring, R. H. (1997). A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. *Forest Ecology and Management*, *95*(3), 209-228. doi:10.1016/S0378-1127(97)00026-1

Law, B. E., Waring, R. H., Anthoni, P. M., & Aber, J. D. (2000). Measurements of gross and net ecosystem productivity and water vapour exchange of a Pinus ponderosa ecosystem, and an evaluation of two generalized models. *Global Change Biology*, *6*(2), 155-168. doi:10.1046/j.1365-2486.2000.00291.x

McCree, K. J. (1974). Equations for the Rate of Dark Respiration of White Clover and Grain Sorghum, as Functions of Dry Weight, Photosynthetic Rate, and Temperature1. *Crop Science*, *14*(4), 509. Crop Science Society of America. doi:10.2135/cropsci1974.0011183X001400040005x

Medlyn, B. E., & Dewar, R. C. (1999). Comment on the article by R.H. Waring, J. J. Landsberg and M. Williams relating net primary production to gross primary production. *Tree physiology*, *19*(2), 137-138.

Minunno, F., Xenakis, G., Perks, M. P., & Mencuccini, M. (2010). Calibration and validation of a simplified process-based model for the prediction of the carbon balance of Scottish Sitka spruce (

Picea sitchensis) plantations. *Canadian Journal of Forest Research*, *40*(12), 2411-2426. doi:10.1139/X10-181

Morris, M. D. (1991). Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, *33*, 161-174. The American Society for Quality Control and The American Statistical Association. Retrieved from http://www.jstor.org/pss/1269043

Mäkelä, A., & Valentine, H. T. (2001). The ratio of NPP to GPP: evidence of change over the course of stand development. *Tree Physiol*, *21*(14), 1015-1030. doi:10.1093/treephys/21.14.1015

Mäkelä, A., & Vanninen, P. (2000). Estimation of fine root mortality and growth from simple measurements: a method based on system dynamics. *Trees*, *14*(6), 316-323. Springer Berlin / Heidelberg. doi:10.1007/s004680050224

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the Earth sciences. *Science (New York, N.Y.)*, *263*(5147), 641-646. doi:10.1126/science.263.5147.641

Robert, C., & Casella, G. (2004). Monte Carlo Statistical Methods. *Book*, *96*(453). Springer.

Penning de Vries, F. W. T. (1974). Substrate utilization and respiration in relation to growth and maintenance in higher plants. *Neth. J. Agric. Sci.*, *22*, 40-44.

Penning de Vries, F. W. T. (1975). Use of assimilates in higher plants. Photosynthesis and Productivity in Different Environments. *Cambridge University Press*, 459-479.

Pereira, J. S., Mateus, J. A., Aires, L. M., Pita, G., Pio, C., David, J. S., Andrade, V., et al. (2007). Net ecosystem carbon exchange in three contrasting Mediterranean ecosystems ? the effect of drought. Retrieved from http://hal-insu.archives-ouvertes.fr/hal-00297720/

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., et al. (2005). On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biology*, *11*(9), 1424-1439. doi:10.1111/j.1365-2486.2005.001002.x

Ryan, M. G. (1991). Effects of Climate Change on Plant Respiration. *Ecological Applications*, *1*(2), 157. doi:10.2307/1941808

Ryan, M. G., Hubbard, R. M., Pongracic, S., Raison, R. J., & McMurtrie, R. E. (1996). Foliage, fine-root, woody-tissue and stand respiration in Pinus radiata in relation to nitrogen status. *Tree Physiol*, *16*(3), 333-343. doi:10.1093/treephys/16.3.333

Ryan, M. G., & Waring, R. H. (1992). Maintenance Respiration and Stand Development in a Subalpine Lodgepole Pine Forest. *Ecology*, *73*(6), 2100-2108. The Ecological Society of America. Retrieved from http://www.jstor.org/pss/1941458

Sands, P., & Landsberg, J. (2002). Parameterisation of 3-PG for plantation grown *Eucalyptus globulus*. *Forest Ecology and Management*, *163*(1-3), 273-292. doi:10.1016/S0378-1127(01)00586-2

Sivia, D. S. (2006). *Data Analysis: A Bayesian Tutorial*. (O. U. Press., Ed.) (2nd ed., p. 260 pp.).

Stape, J. L., Ryan, M. G., & Binkley, D. (2004). Testing the utility of the 3-PG model for growth of with natural and manipulated supplies of water and nutrients. *Forest Ecology and Management*, *193*(1-2), 219-234. doi:10.1016/j.foreco.2004.01.031

ter Braak, C. J. F., & Vrugt, J. A. (2008). Differential Evolution Markov Chain with snooker updater and fewer chains. *Statistics and Computing*, *18*(4), 435-446. doi:10.1007/s11222-008-9104-9

Titus, B., & Malcolm, D. C. (1999). The long-term decomposition of Sitka spruce needles in brash. *Forestry*, *72*(3), 207-221. doi:10.1093/forestry/72.3.207

van Oijen, M, Cameron, D. R., Butterbach-Bahl, K., Farahbakhshazad, N., Jansson, P.-E., Kiese, R., Rahn, K.-H., et al. (2011). A Bayesian framework for model calibration, comparison and analysis: Application to four models for the biogeochemistry of a Norway spruce forest. *Agricultural and Forest Meteorology*, *151*(12), 1609-1621. doi:10.1016/j.agrformet.2011.06.017

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*(7), 915-927. doi:10.1093/treephys/25.7.915

van Oijen, M, Schapendonk, A., & Höglind, M. (2010). On the relative magnitudes of photosynthesis, respiration, growth and carbon storage in vegetation. *Annals of botany*, *105*(5), 793-797. doi:10.1093/aob/mcq039

Vanninen, P., & Mäkelä, A. (2005). Carbon budget for Scots pine trees: effects of size, competition and site fertility on growth allocation and production. *Tree Physiol*, *25*(1), 17-30. doi:10.1093/treephys/25.1.17

Wang, Y. P., & Jarvis, P. G. (1990). Description and validation of an array model — MAESTRO. *Agricultural and Forest Meteorology*, *51*(3-4), 257-280. doi:10.1016/0168-1923(90)90112-J

Waring, R. H., Landsberg, J. J., & Williams, M. (1998). Net primary production of forests: a constant fraction of gross primary production? *Tree Physiol*, *18*(2), 129-134. doi:10.1093/treephys/18.2.129

Xenakis, G., Ray, D., & Mencuccini, M. (2008). Sensitivity and uncertainty analysis from a coupled 3-PG and soil organic matter decomposition model. *Ecological Modelling*, *219*(1-2), 1-16. doi:10.1016/j.ecolmodel.2008.07.020

Table(s)

Table 1a. Symbols, units, minimum and maximum values and prior distributions for the 3PGN and 3PGN* parameters calibrated for *Eucalyptus globulus* in Portugal.

Parameter description	Symbols	Units	Min	Max	Prior distr.
Constant in the aboveground biomoss vs. height relationship	*aH*	—	1.9	2.8	Normal*
Canopy quantum efficiency	*alpha*	mol C * MJ^{-1}	0.04	0.08	Normal
Canopy boundary layer conductance	*BLcond*	$m*s^{-1}$	0.16	0.24	Uniform
Power in the aboveground biomoss vs. height relationship	*bW*	—	0	0.3	Weibull*
Defines stomatal response to VPD	*CoeffCond*	$Mbar^{-1}$	0.04	0.06	Uniform
Basic density	*Density*	$Mg*m^{3}$	0.36	0.54	Normal
Convertion of fresh biomass to dry biomass	*dmC*	—	0.45	0.55	Normal
Value of fNutr when FR = 0	*fN0*	—	0	0.5	Uniform
Branch and bark fraction at age 0	*fracBB0*	—	0.6	0.9	Normal
Branch and bark fraction for mature stands	*fracBB1*	—	0.12	0.18	Normal
Age at canopy cover	*fullCanAge*	years	2	5	Normal
Litterfall rate at t = 0	*gammaF0*	$month^{-1}$	0.0008	0.0012	Normal
Maximum litterfall rate	*gammaFx*	$month^{-1}$	0.0216	0.0324	Normal
Humification coefficient	*hc*	—	0.1	0.15	Uniform
Extinction coefficient for absorption of PAR by canopy	*k*	—	0.4	0.6	Normal
Days of production lost per frost day	*kF*	days	0	3	Normal
Decomposition rate constant for the ‘‘young and labile’’ pool per month	*klmax*	$month^{-1}$	0.006	0.01	Uniform
Decomposition rate constant for the ‘‘old’’ pool	*komax*	$month^{-1}$	0.0004	0.0006	Uniform
Decomposition rate constant for the ‘‘young and refractory’’ pool per month	*krmax*	$month^{-1}$	0.03	0.05	Uniform
LAI for maximum canopy conductance	*LAIgcx*	—	2.664	3.996	Uniform
LAI for maximum rainfall interception	*LAImaxIntcptn*	—	0	0.05	Uniform
Value of m when FR = 0	*m0*	—	0	0.2	Uniform
Maximum stand age used in age modifier	*MaxAge*	years	80	200	Uniform
Maximum canopy conductance	*MaxCond*	$m*s^{-1}$	0.016	0.024	Uniform
Maximum proportion of rainfall evaporated from canopy	*MaxIntcptn*	—	0.12	0.18	Uniform
Power of relative age in function for fAge	*nAge*	—	2	5	Uniform
Foliage–stem partitioning ratio @ D = 2 cm	*pFS2*	—	0.8	1.2	Uniform
Foliage–stem partitioning ratio @ D = 20 cm	*pFS20*	—	0.12	0.18	Uniform
Maximum fraction of NPP to roots	*pRn*	—	0.2	0.3	Uniform
Minimum fraction of NPP to roots	*pRx*	—	0.64	0.96	Uniform
Q10	*Q10***	—	1	3.5	Normal
Relative age to give fAge = 0.5	*rAge*	—	0.76	1	Uniform
Foliage biomass respiration rate	*rf***		0.0005	0.02	Gamma
Growth respiration rate	*rg***		0.2	0.3	Normal
Average monthly root turnover rate	Rttover	$month^{-1}$	0.012	0.018	Gamma

Table 1a. (Concluded)

Parameter description	Symbols	Units	Min	Max	Prior distr.
Woody biomass respiration rate	*rw***		0.001	0.06	Gamma
Woody biomass respiration rate	*rw***		0.001	0.06	Gamma
Specific leaf area at age 0	*SLA0*	$m^2 * kg^{-1}$	10.5	14	normal*
Specific leaf area for mature leaves	*SLA1*	$m^2 * kg^{-1}$	3.7	4.4	normal*
Constant in the aboveground biomass vs. diameter relationship	*StemConst*	—	1.15	1.4	gamma*
Power in the aboveground biomass vs. diameter relationship	*StemPower*	—	0.5	0.55	gamma*
Moisture ratio deficit for fq = 0.5	*SWconst*	—	0.63	0.77	normal
Power of moisture ratio deficit	*SWpower*	—	8.1	9.9	normal
Age at which fracBB = (fracBB0 + fracBB1)/2	*tBB*	years	1.6	2.4	normal
Age at which litterfall rate has median value	*tgammaF*	years	9.6	14.4	normal
Maximum temperature for growth	*Tmax*	ºC	32	48	normal
Minimum temperature for growth	*Tmin*	ºC	6.8	10.2	normal
Optimum temperature for growth	*Topt*	ºC	12.8	19.2	normal
Age at which specific leaf area = (SLA0 + SLA1)/2	*tSLA*	years	1.2	2	normal*
Ratio NPP/GPP	*Y****	—	0.376	0.564	normal

* distributions fitted over posterior distributions

** only 3PGN* parameters

*** only 3PGN parameters

Table 1b. Symbols, units, minimum and maximum values and prior distributions for the 3PGN site variables used in this work.

Site variable description	Symbols	Units	min	max	prior distr.
Fertility rating for the Espirra site	*FR_espirra*	—	0.4	0.7	normal
Fertility rating for the ferc site	*FR_ferc*	—	0.4	0.7	normal
Fertility rating for the ferf site	*FR_ferf*	—	0.6	1	normal
Fertility rating for the feri site	*FR_feri*	—	0.4	0.7	normal
Fertility rating for the ferif site	*FR_ferif*	—	0.6	1	normal
Maximum available soil water for the Espirra forest	*MaxASW_espirra*	mm*ha^{-1}	120	180	uniform
Maximum available soil water for the Furadouro experiment	*MaxASW_fer*	mm*ha^{-1}	120	180	normal
Minimum available soil water for the Espirra forest	*MinASW_espirra*	mm*ha^{-1}	0	60	uniform
Minimum available soil water for the Furadouro experiment	*MinASW_fer*	mm*ha^{-1}	0	40	normal
Initial carbon in the old pool	*O_C_i*	kg*ha^{-1}	30	50	normal
Tree density at the Espirra site	*StemNo_espirra*	trees*ha^{-1}	1650	1750	normal
Tree density at the ferc site	*StemNo_ferc*	trees*ha^{-1}	1060	1120	normal
Tree density at the ferf site	*StemNo_ferf*	trees*ha^{-1}	1060	1120	normal
Tree density at the feri site	*StemNo_feri*	trees*ha^{-1}	1060	1120	normal
Tree density at the ferif site	*StemNo_ferif*	trees*ha^{-1}	1060	1120	normal
Initial foliage biomass	*WF_i*	kg*ha^{-1}	0.01	0.2	uniform
Initial root biomass	*WR_i*	kg*ha^{-1}	0.001	0.1	uniform
Initial stem and branches biomass	*WS_i*	kg*ha^{-1}	0.001	0.05	uniform
Initial carbon in the young labile pool	*Yl_C_i*	kg*ha^{-1}	8	12	normal
Initial carbon in the young refractory pool	*Yr_C_i*	kg*ha^{-1}	0	10	uniform

Table 2. Results of the Bayesian model comparison of 3PGN and 3PGN*. The table shows the log-transformed integrated likelihood values for the *prior BMC* and *post BMC*.

	3PGN	3PGN*
prior BMC	-640.6	-638.94
post BMC	-71.58	-65.71

Table 3. Comparison of data with model outputs: squared correlation coefficient (r^2) and normalized root mean square error (NRMSE). The table shows the distribution means of statistics induced by prior and posterior parameter distributions. In bold: posterior values that are improvements over the prior (r increased, NRMSE reduced).

		3PGN		3PGN*	
Var.	Statistic	Prior	Post.	Prior	Post.
P_E	*r^2*	0.11	**0.25**	0.15	**0.26**
	slope	0.27	**0.47**	0.33	**0.52**
	NRMSE	118.4	**101.8**	121.3	**109.3**
D	*r^2*	0.89	**0.98**	0.91	**0.98**
	Slope	1.51	**0.88**	1.24	**0.9**
	NRMSE	124.7	**20.4**	91.6	**14.8**
H	*r^2*	0.96	0.93	0.92	**0.93**
	slope	1.65	**1.25**	1.38	**1.29**
	NRMSE	78.9	**33.2**	44.7	**31.2**
WF	*r^2*	0.92	0.71	0.83	0.68
	slope	1.36	**0.89**	1.03	0.88
	NRMSE	52	55.4	33.1	57.2
WR	*r^2*	0.99	0.95	0.96	0.93
	slope	1.5	**0.83**	1.06	0.94
	NRMSE	28	38.4	17.3	31.3
WS	*r^2*	0.94	**0.97**	0.91	**0.96**
	slope	3.53	**1.09**	2.15	**1.15**
	NRMSE	86.2	**19.5**	63.4	**21.5**
V	*r^2*	0.94	**0.98**	0.95	**0.98**
	slope	3.74	**1.06**	2.14	**1.1**
	NRMSE	89.3	**14**	63.9	**17.1**

Figure(s)

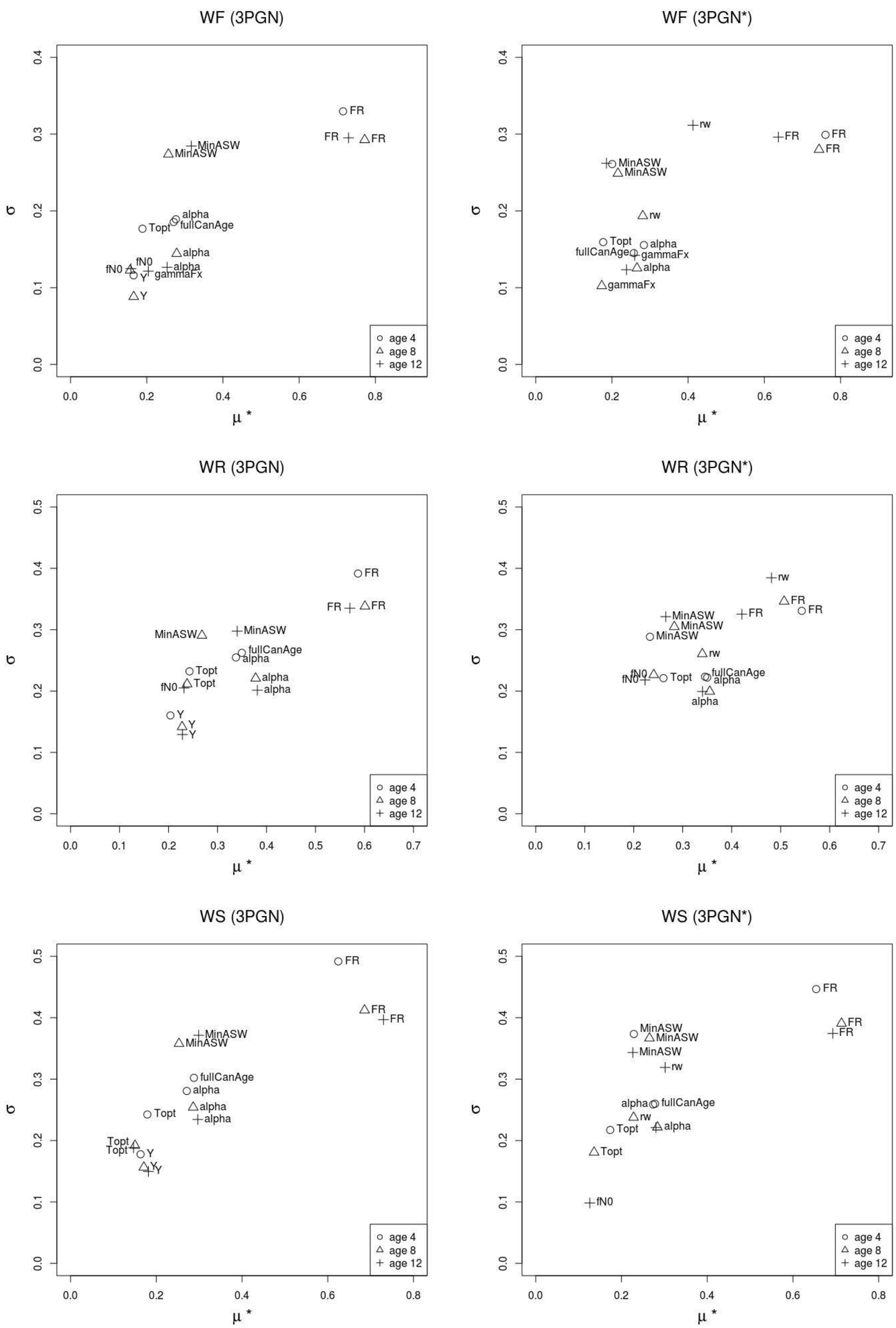

Figure 1a. Plots of σ vs. μ* of the five highest sensitive parameters for foliage (WF), root (WR) and stem (WS) biomasses for 3PGN and 3PGN* outputs at age 4 (circles), 8 (triangles) and 12 (cruces).

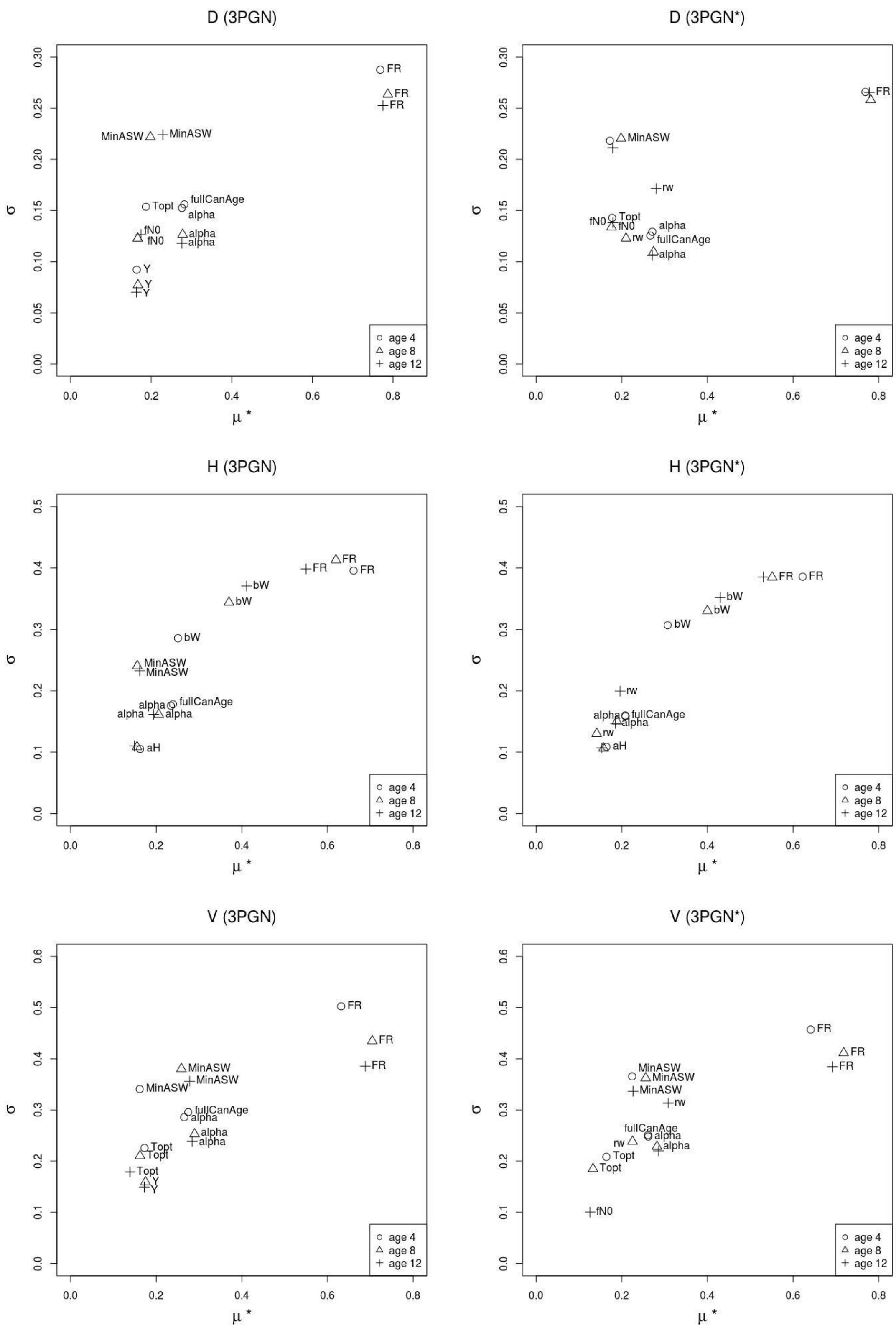

Figure 1b. Plots of σ vs. μ* of the five highest sensitive parameters for diameter at breast height (D), average stand height (H) and stand volume (V) for 3PGN and 3PGN* outputs at age 4 (circles), 8 (triangles) and 12 (cruces).

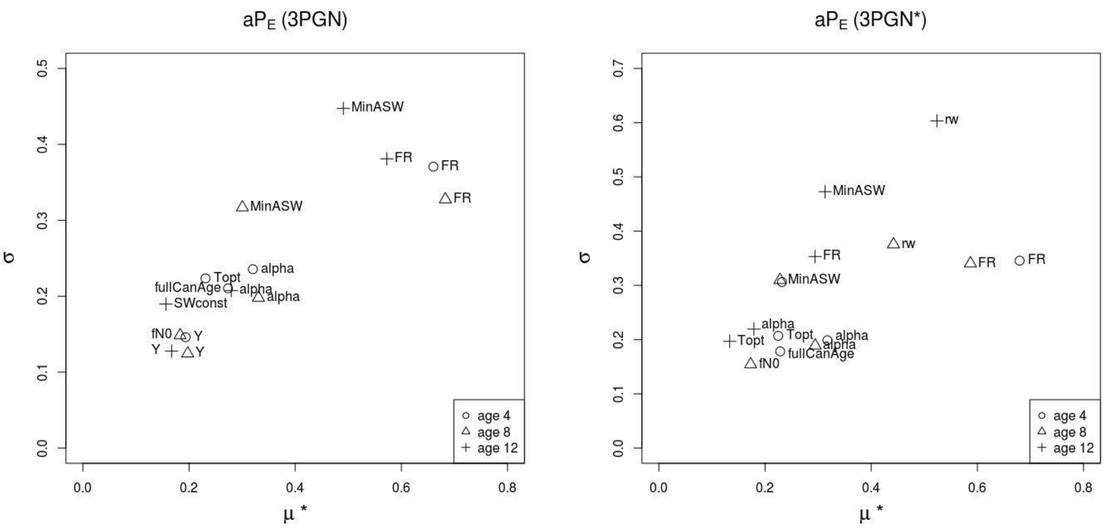

Figure 1c. Plots of σ vs. μ* of the five highest sensitive parameters for the annual net ecosystem production (aP_E) for 3PGN and 3PGN* outputs at age 4 (circles), 8 (triangles) and 12 (cruces).

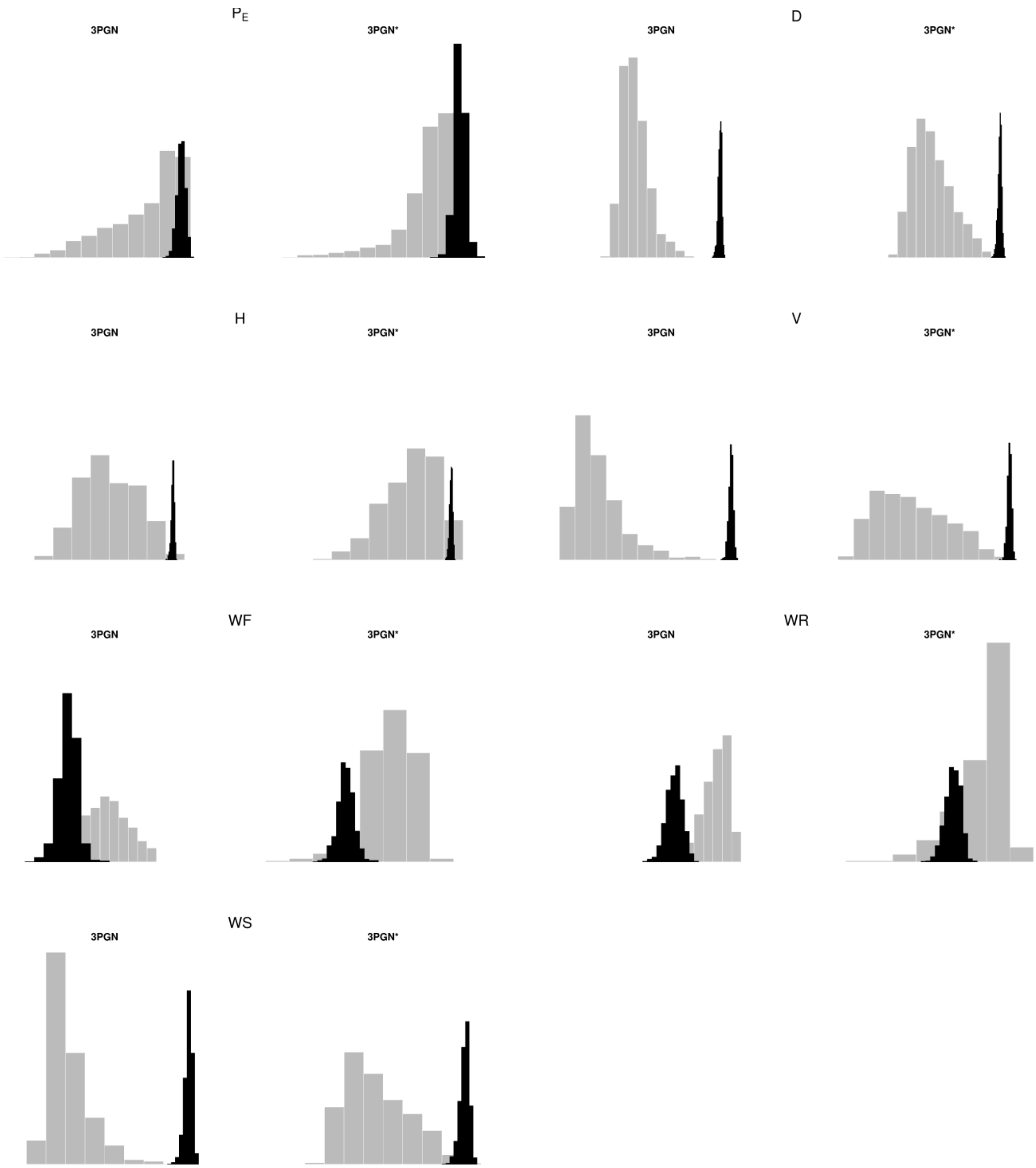

Figure 2. Prior (grey histograms) and posterior (black histograms) distributions of log-likelihoods for the two model versions, for the seven categories of output variables.

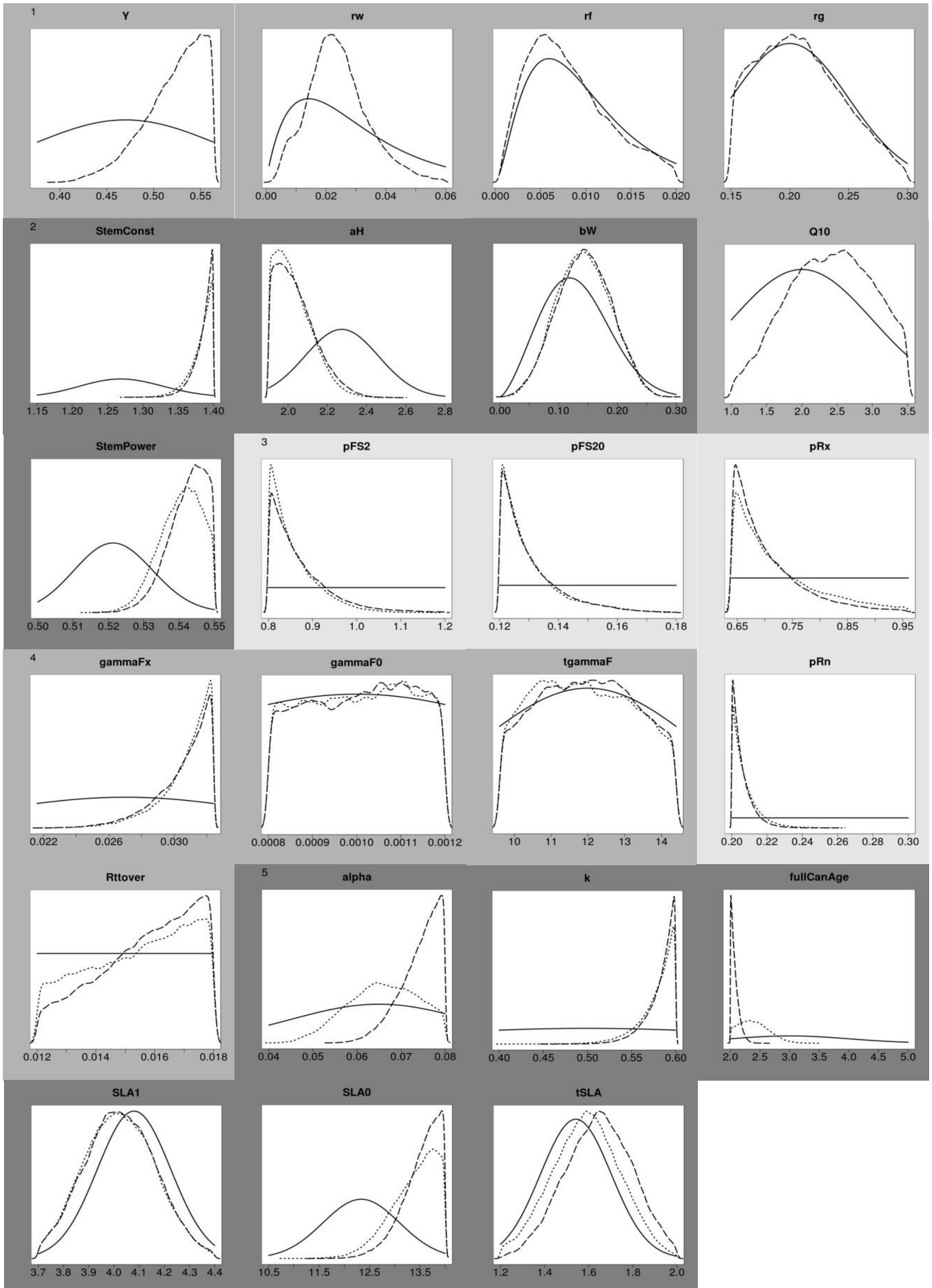

Figure 3a. Marginal prior distributions (continuous line) and marginal posterior distributions of 3PGN (dashed line) and 3PGN* (dotted line). Parameters are grouped as respiration parameters (group 1), allometric parameters (group 2), allocation parameters (group 3), turnover parameters (group 4), light use efficiency and light interception parameters (group 5).

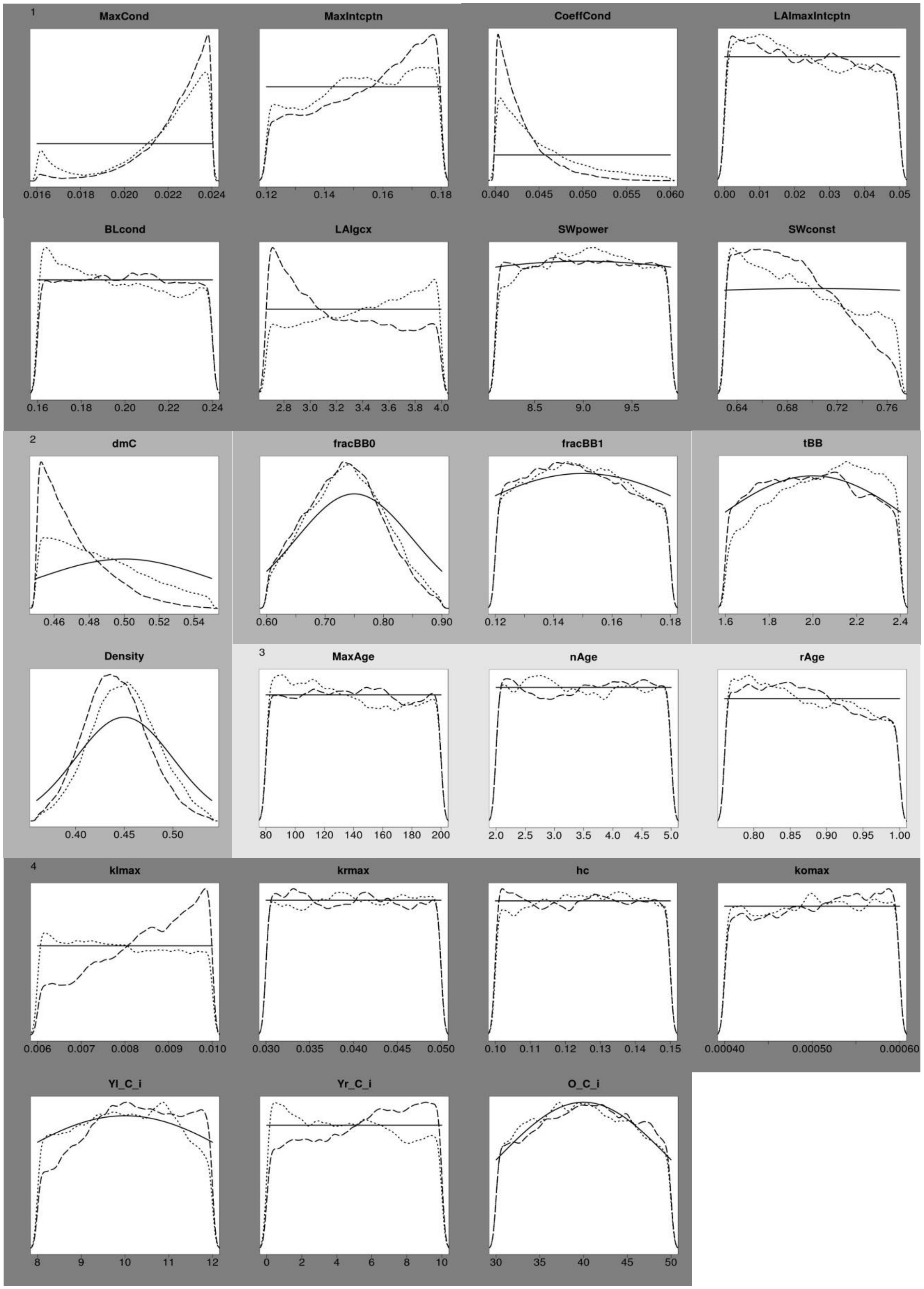

Figure 3b. Marginal prior distributions (continuous line) and marginal posterior distributions of 3PGN (dashed line) and 3PGN* (dotted line). Parameters are grouped as water stress parameters (group 1), volume and density parameters (group 2), age stress parameters (group 3) and soil parameters (group 4).

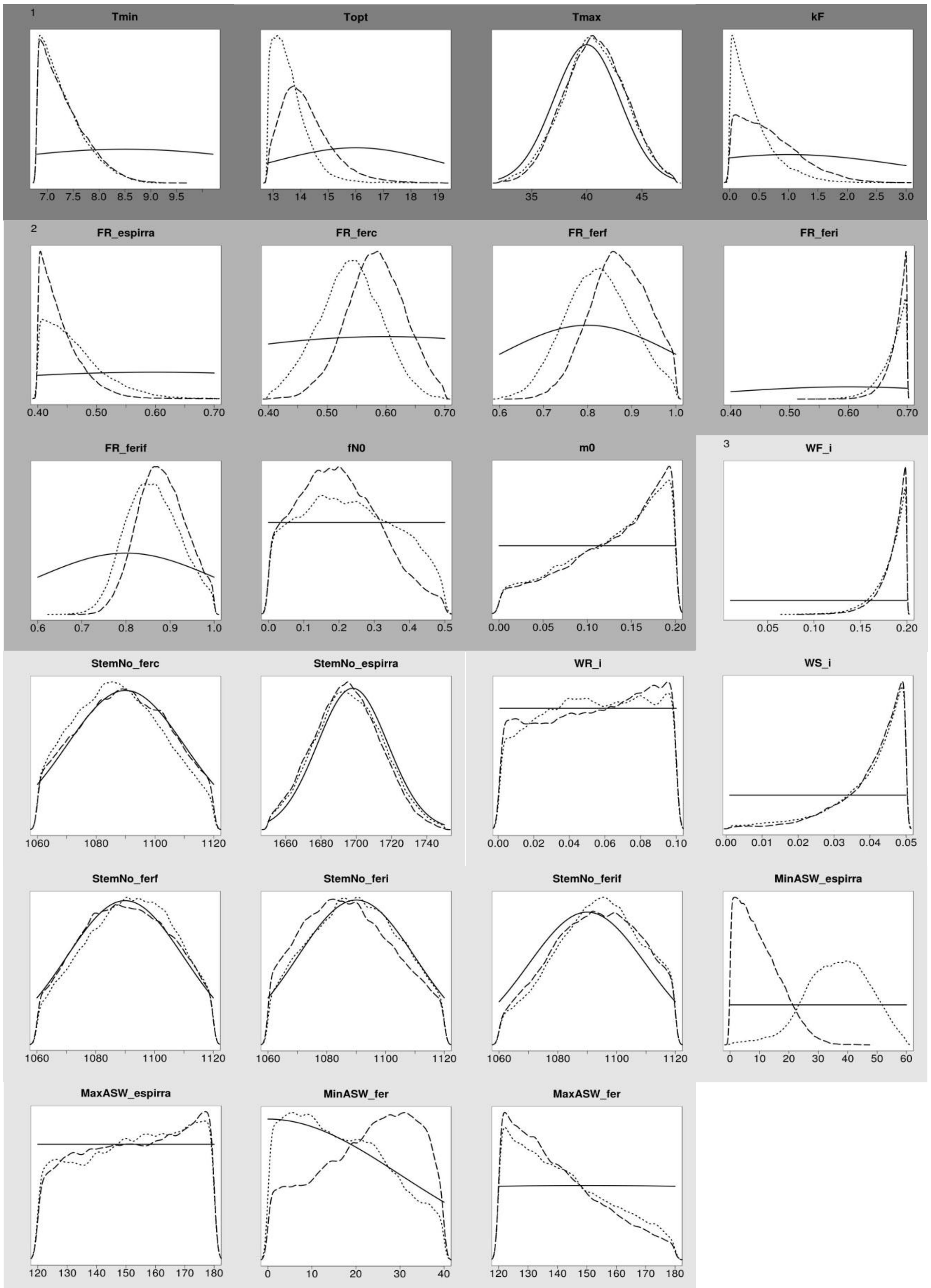

Figure 3c. Marginal prior distributions (continuous line) and marginal posterior distributions of 3PGN (dashed line) and 3PGN* (dotted line). Parameters are grouped as temperature and frost stress parameters (group 1), fertility parameters (group 2), site parameters (group 3).

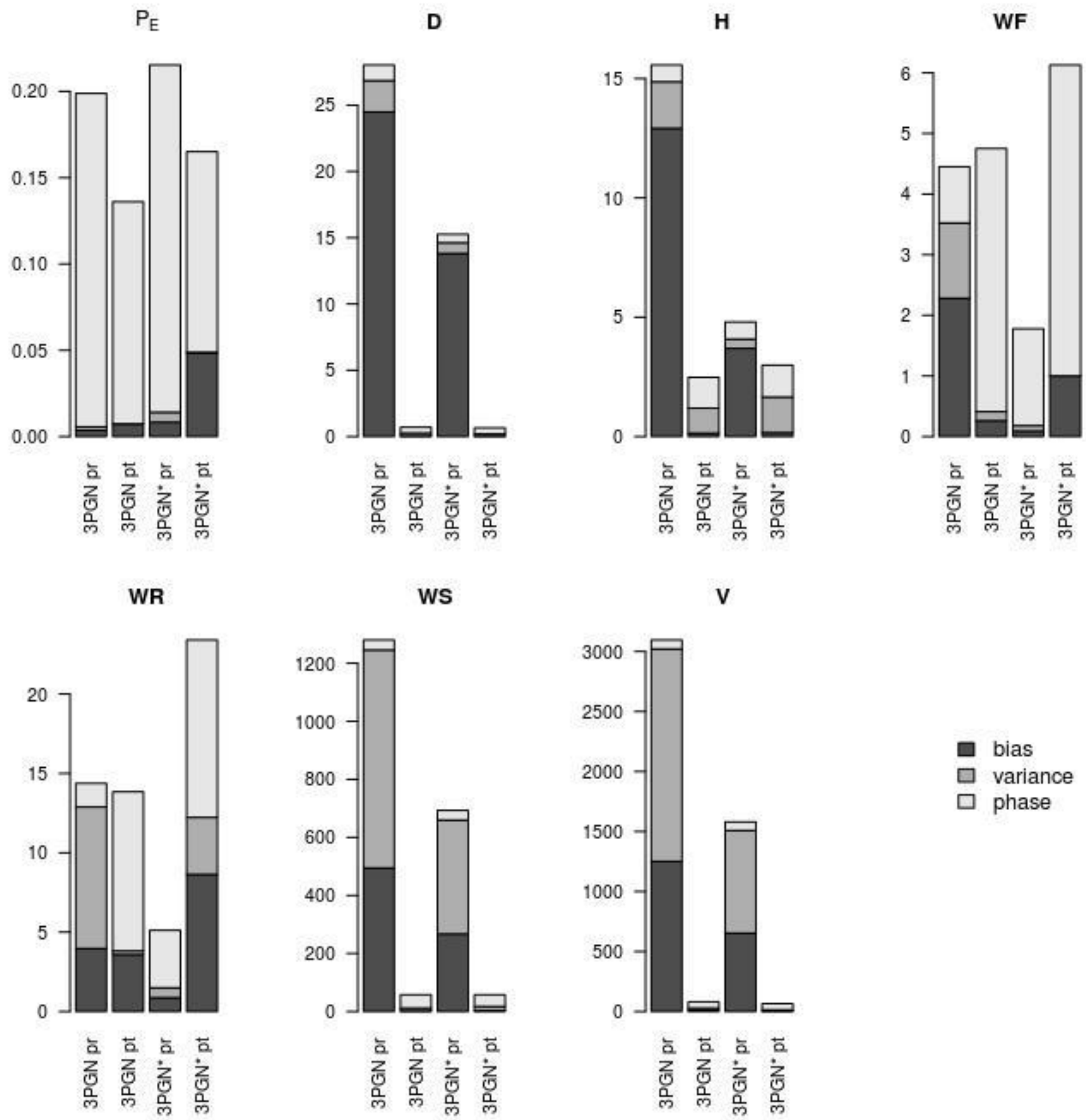

Figure 4. Decomposition of the mean squared error associated with the modes of the prior (pr) and posterior (pt) parameter distributions, for 3PGN and 3PGN*. In squared brackets are reported the number of data for each variable.