Minunno, F.; van Oijen, M.; Cameron, D.R.; Pereira, J.S. 2013. **Selecting parameters for Bayesian calibration of a process-based model: a methodology based on canonical correlation analysis**. *SIAM/ASA Journal on Uncertainty Quantification*, 1 (1). 370-385. 10.1137/120891344

Selecting parameters for Bayesian calibration of a process-based model: a methodology based on canonical correlation analysis.

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

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

[2] = Centre for Ecology and Hydrology, CEH-Edinburgh, Bush Estate, Penicuik EH26 0QB, U.K.

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

Introduction

In recent decades, forests have started to experience significant changes in environmental conditions (Fontes et al., 2010). Ecosystems have to adapt to variations in the mean climatic variables, but also the increased risk of extreme events such as drought, heat waves, storms, late frost and flooding (Lindner et al., 2010). The impact of environmental changes on forest function varies between different regions. In some areas positive effects on forest growth are expected because of longer growing seasons, nitrogen deposition and higher content of CO_2 in the air. Conversely, forest productivity is likely to decline in areas that are more vulnerable to drought and fire. Therefore, it is essential to adapt forest management to these changing conditions. Process-based models are flexible tools that can support forest management under abiotic and biotic changes. Their use in forest research as well as in practical forest management has significantly increased (Johnsen et al., 2001). The reliability and robustness of models are conditional on their structure but also on their parameterisation (van Oijen et al., 2011; Minunno et al., 2013). Hence the calibration of model parameters is a key stage in the model building process.

Bayesian calibration (BC), based on probability theory, is a logical choice for model calibration (van Oijen et al., 2005). BC provides parameter quantification and quantifies uncertainties in model input and output. Even though Bayes' theorem was formulated in the 18th century, its practical application has been hindered by the fact that it is often computationally demanding. Nowadays the use of Bayesian statistics is becoming increasingly common in the environmental sciences because of developments in computers and sampling based techniques such as Markov Chain Monte Carlo simulation (MCMC). However, the use of the Bayesian approach is still limited in forest research, especially for computationally expensive models with many parameters (van Oijen *et al.*, 2005). To reduce the computational load and take full advantage of Bayesian statistics it is possible to act in two directions. First, one can increase the efficiency of the sampling based techniques by using more complex algorithms (Andrieu & Thoms, 2008). Second, the number of parameters involved in the calibration can be reduced by means of parameter screening. Of these two options, screening is the most simple and straightforward. Parameter screening aims to rank model parameters according to their impact on the output (Saltelli et al., 2004). In this case, sensitivity analysis plays a key role in

identifying the most important parameters. In earlier studies, screening was used with forest models to select parameters for Bayesian calibration (van Oijen et al., 2013; van Oijen et al., 2011; Xenakis et al., 2008), however it has not been investigated if the BC of a parameter subset, instead of the full parameter set, affects model performance. Furthermore, parameter and output uncertainties may not be fully quantified. The method used for parameter selection can also affect parameter screening results. A key difference is that between local sensitivity analysis (LSA) and global sensitivity analysis (GSA). LSA quantifies model output variations in relation to changes of one parameter at a time at a specified point of the parameter space. GSA explores the full parameter space and evaluates the model's output sensitivity to simultaneous changes in several parameters, thus better characterising the behaviour of the model. The Fourier Amplitude Sensitivity Test method (Cukier et al., 1973), the Sobol' (Sobol', 1990) and Saltelli's method (Saltelli et al., 2010) are common GSAs, all of which can be computationally demanding for parameter-rich models. The Morris method (Campolongo et al., 2007; Morris, 1991) is a GSA that has already been used to screen parameters of a process-based forest model (van Oijen et al., 2011). Canonical correlation analysis (CCA) is a technique that can be used for GSA (Hair et al., 1998); CCA is a multivariate technique and thus has the advantage of evaluating multiple output variables simultaneously giving a quantitative measure of model output sensitivity. However, CCA has never been used to quantify the output variable sensitivity of forest models.

The main objectives of this work were:

1. Determine whether choosing a subset of parameters for calibration impacts the performance of a process-based forest model.

2. Provide a methodology for parameter screening based on canonical correlation analysis.

The model used for this exercise was 3PGN (Xenakis et al., 2008) and the analyses were carried out using a comprehensive dataset of *Eucalyptus globulus* plantations in Portugal.

Materials and methods

3PGN Structure

3PGN was developed by Xenakis *et al.* (2008) coupling two models, 3PG (Physiological Principles in Predicting Growth) and ICBM (Introductory Carbon Balance Model). The first model simulates forest growth while the second computes soil carbon and nitrogen balances. The combination of the two models permits analysis at ecosystem level.

In 3PG the gross primary production (P_G) is calculated by multiplying photosynthetically active radiation absorbed by the stand (aPAR) with a light-use efficiency (α). aPAR is determined through Beer's law, while α depends on atmospheric vapour pressure deficit, air temperature, frost events, soil water balance, tree age and site fertility. Net primary production (P_N) is a constant fraction of P_G and the biomass is allocated to the tree organs: root, stem and foliage. 3PG simulates stand attributes, such as stand timber volume, mean diameter at breast height, average stand height, basal area and mean

annual growth increment. A detailed description of the model is provided in Landsberg and Waring (1997) and in Sands and Landsberg (2001).

ICBM considers three pools of C and three pools of N in the soil, storing different forms of organic matter. Small tree detritus (from litterfall and root turnover) accumulates in a "young labile" pool, coarse woody detritus (i.e., coarse root, branches and stems) accumulates in a "young refractory" pool and the recalcitrant organic matter accumulates in an "old" pool. Each pool has a decomposition rate that varies with soil moisture and soil temperature. The sum of the outflows from the different pools represents the heterotrophic respiration. A complete description of ICBM is provided by Andrén and Kätterer (1997) and Kätterer and Andrén (1999; 2001).

3PGN was chosen because it is a simple process-based model; it has about 50 parameters and initial constants. It has a monthly time step, so that the model is sufficiently efficient to easily perform Bayesian calibrations using the full parameter set or different parameter subsets.

Experimental sites and data acquisition

The data used for this exercise were collected at Espirra forest and the Furadouro experiment. 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 forest is a 300 ha *Eucalyptus globulus* plantation (38º38`N, 8º36`W) tended as a coppice (Pereira et al., 2007). The mean annual temperature for the site is 16ºC and the mean annual rainfall is 709 mm. About 80% of the precipitation occurs between October and April. Espirra forest is a CarboEurope-IP site where fluxes of H_2O and CO_2 have been measured by eddy covariance, following the Fluxnet protocols (Aubinet et al., 1999; Baldocchi, 2003). Flux data quality control followed the CarboEurope-IP recommendations; gap filling and partitioning of P_E to gross primary production and ecosystem respiration was performed according to Reichstein et al. (2005). To allow model calibrations the P_E data were monthly averaged.

The Furadouro experiment dataset consisted of foliage (*WF*, Mg of dry mass (DM) ha^{-1}), stem (*WS*, Mg DM ha^{-1}) and root (*WR*, Mg DM ha^{-1}) biomasses, stand volume (*V*, m^3 ha^{-1}), *H* and *D*. The data were collected from a *E. globulus* plantation at Quinta do Furadouro (Óbidos, Portugal, 39º29`N, 9º13`W, 30 m a.s.l.) from 1986 to 1992. The mean annual temperature is 15.2 ºC and the mean annual precipitation is 607 mm, of which less than 10% occurs between May and September. The experimental design consisted of three treatments and a control. In the first treatment, daily irrigation was supplied from April to October (*I*), in the second treatment a pelleted fertilizer was applied in March and October of each year (*F*), in the third treatment the daily irrigation as in *I* was combined with a liquid fertilizer solution (*IF*). No fertilization and irrigation were supplied to the control (*C*).

Table 1 shows the number of measurements for each data type and their relative uncertainty expressed through the coefficient of variation (CV).

Table 1. Data types, number of measurements and coefficient of variation used for the calibration of 3PGN for Eucalyptus globulus in Portugal.

Data type	Number of data	Coefficient of variation
P_E	38	0.3
D	36	0.1
H	35	0.2
V	32	0.3
WS	20	0.2
WF	20	0.3
WR	12	0.4

Canonical correlation analysis

Canonical correlation analysis (CCA) is a multivariate technique that aims to find the relationship between two sets of variables. Therefore, CCA is particularly useful for sensitivity analyses of process-based models that have many parameters and multiple outputs. Canonical correlation analysis was introduced by Hotelling in 1936 (Hotelling, 1936). As with many multivariate techniques, the application of CCA has recently increased with the availability of computer programs that facilitate its implementation.

A detailed description of the method can be found in Hair et al. (1998), while here we provide a brief outline. CCA computes relationships between linear combinations of dependent and independent variables. The linear composites are called *canonical variates*, while the relationships between them are the *canonical functions*. An array of canonical functions is developed to maximize the correlation (*canonical correlation*) between two linear composites, one for the dependent and one for the independent variables. Therefore each function is developed using pairs of canonical variates; the maximum number of functions is equal to the number of variables in the smallest set. The first pair of canonical variates has the highest correlation between the dependent and independent variables and it accounts for the maximum variance in the set of variables. The second pair of canonical variates is then derived maximizing the correlation between the two sets of variables, based on the residual variance. Successive canonical functions are computed on the basis of the remaining variance and the canonical correlations become smaller as new pairs are extracted.

CCA results can be interpreted through three measures: canonical weights, canonical loadings and canonical cross-loadings (see Hair et al. (1998) for further details). Because we were interested in analyzing the sensitivity of model output (dependent variables) to the parameters (independent variables), we examined the canonical cross-loadings. Canonical cross-loadings are the correlations between individual variables, dependent or independent, and their opposite canonical variates. Our interest was in quantifying the impact of each parameter on the set of model outputs. We therefore examined the canonical cross-loadings between the individual parameters and the composite output variates.

Table 2. Symbols, units, minimum and maximum values for the 3PGN parameters calibrated for Eucalyptus globulus in Portugal.

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

Table 2. (Concluded)

Parameter description	Symbols	Units	Min	Max
Specific leaf area at age 0	*SLA0*	m^2*kg^{-1}	10.5	14
Specific leaf area for mature leaves	*SLA1*	m^2*kg^{-1}	3.7	4.4
Constant in the aboveground biomass vs. diameter relationship	*StemConst*	—	1.15	1.4
Power in the aboveground biomass vs. diameter relationship	*StemPower*	—	0.5	0.55
Moisture ratio deficit for fq = 0.5	*SWconst*	—	0.63	0.77
Power of moisture ratio deficit	*SWpower*	—	8.1	9.9
Age at which fracBB = (fracBB0 + fracBB1)/2	*tBB*	years	1.6	2.4
Age at which litterfall rate has median value	*tgammaF*	years	9.6	14.4
Maximum temperature for growth	*Tmax*	ºC	32	48
Minimum temperature for growth	*Tmin*	ºC	6.8	10.2
Optimum temperature for growth	*Topt*	ºC	12.8	19.2
Age at which specific leaf area = (SLA0 + SLA1)/2	*tSLA*	years	1.2	2
Ratio NPP/GPP	*Y*	—	0.376	0.564
Fertility rating for the Espirra plot	*FR_Espirra*	—	0.4	0.7
Fertility rating for the C plot	*FR_C*	—	0.4	0.7
Fertility rating for the F plot	*FR_F*	—	0.6	1
Fertility rating for the I plot	*FR_I*	—	0.4	0.7
Fertility rating for the IF plot	*FR_IF*	—	0.6	1

The advantages of using canonical correlation analysis for model sensitivity analysis are that CCA can calculate the relationships between multiple sets of variables, parameters and outputs, and it also provides quantitative information (i.e., canonical cross-loadings) about model output sensitivity to individual parameters. The main limitation is that CCA is a linear analysis so any non-linearity in the model is not considered, although higher order non-linear terms can be included. CCA finds the linear combinations of terms but the terms themselves can be appropriately transformed in order to convert non-linear relationships to linear forms.

Canonical correlation analysis was performed to calculate the influence of 3PGN parameters on seven output variables (foliage, stem and root biomasses, net ecosystem production, mean stand diameter at breast height, mean stand height and stand volume). Model outputs were predictions at the end of the rotation (12 years stand age). The analyses were made for the Espirra forest and for the four treatments, i.e., *I*, *F*, *FI* and *C*, of the Furadouro experiment. For the sensitivity analyses parameter values varied between the minimum and maximum values shown in Table 2.

CCA was performed between the full parameter set (51 independent variables) and the seven outputs over the five sites (35 dependent variables). Parameter vectors were created using Latin hypercube sampling, to efficiently sample the whole parameter space, and the 35 outputs were calculated for each parameter vector. To ensure that canonical correlation analysis results were not

specific only to the sample, but could be generalized, we increased the sample number until achieving similar values of the canonical cross-loadings of subsequent CCAs. A sample of 50000 parameter vectors was required to generalize the CCA results.

CCA produced a matrix **C** of canonical cross-loadings with 35 columns of canonical variates and 51 rows of 3PGN parameters. The cross-loadings expressed the importance of parameters for each canonical variate, so in **C** each parameter had 35 canonical cross-loadings. The highest cross-loading of each parameter was selected and a ranking of parameters was created. The most important parameters had the highest cross-loadings.

Bayesian calibrations

The calibration of 3PGN, using the Espirra and Furadouro data, was carried out by means of the Bayesian method. Bayesian calibration updates the current state of knowledge about parameter values, expressed as a joint probability distribution (*prior* distribution), using new data. The data, by means of the *likelihood* function (L(θ)), are used to modify the prior uncertainty. The updated joint probability distribution for the parameters is the *posterior* distribution.

For the likelihood function, the Gaussian distribution is the most common choice. However, the Gaussian assigns very low likelihoods when a large mismatch between the observed and simulated data occurs, so its use is not recommended in the presence of outliers. Sivia (2006) proposed the likelihood function of Eq. 3 that gives less weight to outliers because of its slowly decaying Cauchy-like tails.

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

where, sim(θ) is the output from the model for the parameter vector θ, N is the number of data points, σ is the uncertainty about the random error of the i-th data point and R_i = (sim(θ) - O_i)/σ_i (O are the observed data).

The Sivia likelihood was used here because outliers can occur in eddy covariance measurements. A uniform prior was assigned to all 3PGN parameters; the parameter bounds were the same as those reported in Minunno et al. (2013) and are shown in Table 2.

The Bayesian calibration was carried out by means of Markov Chain Monte Carlo sampling (MCMC), using the Metropolis-Hastings random walk. A complete description of the algorithm, in the context of forest modelling, is given by van Oijen et. al. (2005). For the BC, three chains of 500,000 iterations were computed. Convergence of iterative simulations was assessed through the Gelman-Rubin test (Gelman & Rubin 1992). After convergence, the three chains were joined together and treated as a single sample from the posterior distribution.

Four Bayesian calibrations were performed, one using the full 3PGN parameter set (denoted as 'p100%') and the other three using parameter subsets selected by means of CCA. The full parameter set consisted of 51 parameters, five parameters were site-specific and related to site fertility (FR

parameters), while the remaining 46 were common over the sites. CCA was used to rank the 46 common parameters according to their influence on model output. From the ranking created through CCA, three parameter subsets (p25%, p50% and p75%) were created representing the 25%, 50% and 75% most important parameters. The five site-specific parameters were always included in the calibrations because the model is highly sensitive to FR (Esprey et al. 2004; Minunno et al., 2013). For the calibration of the parameter subsets, average values between parameter minimum and maximum were assigned to the parameters not involved in the calibrations.

Results

CCA

Table 3 lists the 3PGN parameters ranked using the highest cross-loading value of each parameter; the five site-specific parameters were not included in the list. The 12 most important parameters are those reported in the first column of Table 3 (p25% subset). The parameters in the first and second column were used for the p50% BC. The parameters of the first, second and third columns represent the parameter subset used in the p75% BC. The 25% of the parameters that were most important were related to allometric equations, light use efficiency, decomposition rates and autotrophic respiration, wood density, litterfall and temperature stress. The sets consisting of the 50% and 75% most important parameters included parameters related to water stress, allocation routines, specific leaf area and frost stress. According to CCA screening, the least important parameters (column four of Table 3) were parameters related to age stress, initial soil carbon content and decomposition rates of humified organic matter.

Bayesian calibrations

Each of the Bayesian calibrations carried out (i.e., p25%, p50%, p75%, p100%) generated a joint probability distribution of the parameter sets involved in the calibration process. By means of the Kolmogorov-Smirnov test for each parameter it was tested if the marginal posterior distributions obtained by p25%, p50% and p75% were statistically different from the marginal distributions achieved through the p100% calibration.

The marginal posterior parameter distributions that were different from the marginal posterior distributions of p100% are shown in Figure 1. The parameters that were common over the sites are ordered according to the highest cross-loading value as listed in Table 3; while the distributions of the site specific parameters (FRs) are plotted at the end of Figure 1. Because the marginal posterior distributions contain only part of the information about the posterior, ignoring parameter interactions, in Table 4 the parameter correlations with absolute value higher than 0.24 are reported for each calibration.

For some of the parameters (i.e., *rho1, pRn, LAIgx, MaxCond, pRx*), the marginal posterior distributions were the same for the different calibrations (data not shown). Significant differences were found between the marginal distributions of other parameters (i.e., *StemPower, bW, alpha, aH, fracBB1,*

Y, Tmin, pFS2, klmax, k, fullCanAge, tBB) (Figure 1). For almost all the parameters, no significant differences were found between the marginal posterior parameter distributions of p75% and p100%, while the differences increased when smaller set of parameters were calibrated. Furthermore, results showed that parameter uncertainty decreased when the number of parameters involved in the calibrations was reduced, but at the same time the interactions between parameters increased (Table 4). In fact, the second order correlations between parameters were similar in p100% and p75% (Table 4), while in p25% and p50% more correlations with absolute value higher than 0.24 were found. The parameter *alpha* interacted with many others in p25%, while, in the other calibrations, *alpha* was significantly correlated with just *Y*. We did not quantify higher order correlations between the parameters that are likely to decrease when parameters are discarded from the calibration.

Table 3. List of parameters ranked on the basis of the canonical correlation analysis. The highest cross-loadings of each parameter with the output canonical variates are shown ('Hc-l').

	p25%			p50%	
Ranking	parameter	Hc-l	ranking	Parameter	Hc-l
1	rho1	0.779	13	Y	0.358
2	StemPower	0.775	14	LAIgcx	0.343
3	pRn	0.654	15	Tmin	0.311
4	bW	0.627	16	MaxCond	0.301
5	alpha	0.596	17	pFS2	0.254
6	aH	0.588	18	StemConst	0.23
7	gammaF1	0.587	19	pRx	0.216
8	fN0	0.541	20	Klmax	0.214
9	Topt	0.506	21	K	0.198
10	pFS20	0.496	22	m0	0.166
11	fracBB1	0.488	23	fullCanAge	0.144
12	gammaR	0.448	24	tBB	0.142

	p75%			p100%	
Ranking	parameter	Hc-l	ranking	Parameter	Hc-l
25	CoeffCond	0.139	36	tgammaF	0.017
26	SLA1	0.117	37	Yl_C_i	0.012
27	tSLA	0.106	38	nAge	0.011
28	krmax	0.09	39	O_C_i	0.01
29	kF	0.068	40	Yr_C_i	0.009
30	fracBB0	0.059	41	gammaF0	0.009
31	Tmax	0.047	42	rAge	0.008
32	BLcond	0.038	43	LAImaxIntcptn	0.008
33	MaxIntcptn	0.036	44	Komax	0.008
34	SLA0	0.035	45	MaxAge	0.007
35	hc	0.019	46	dmC	0.007

Figure 1. Marginal posterior distributions of 3PGN parameters for the four calibrations carried out. Only those parameters for which the cumulative distribution functions had a Kolmogorov-Smirnov distance of at least 0.05 from p100% are shown.

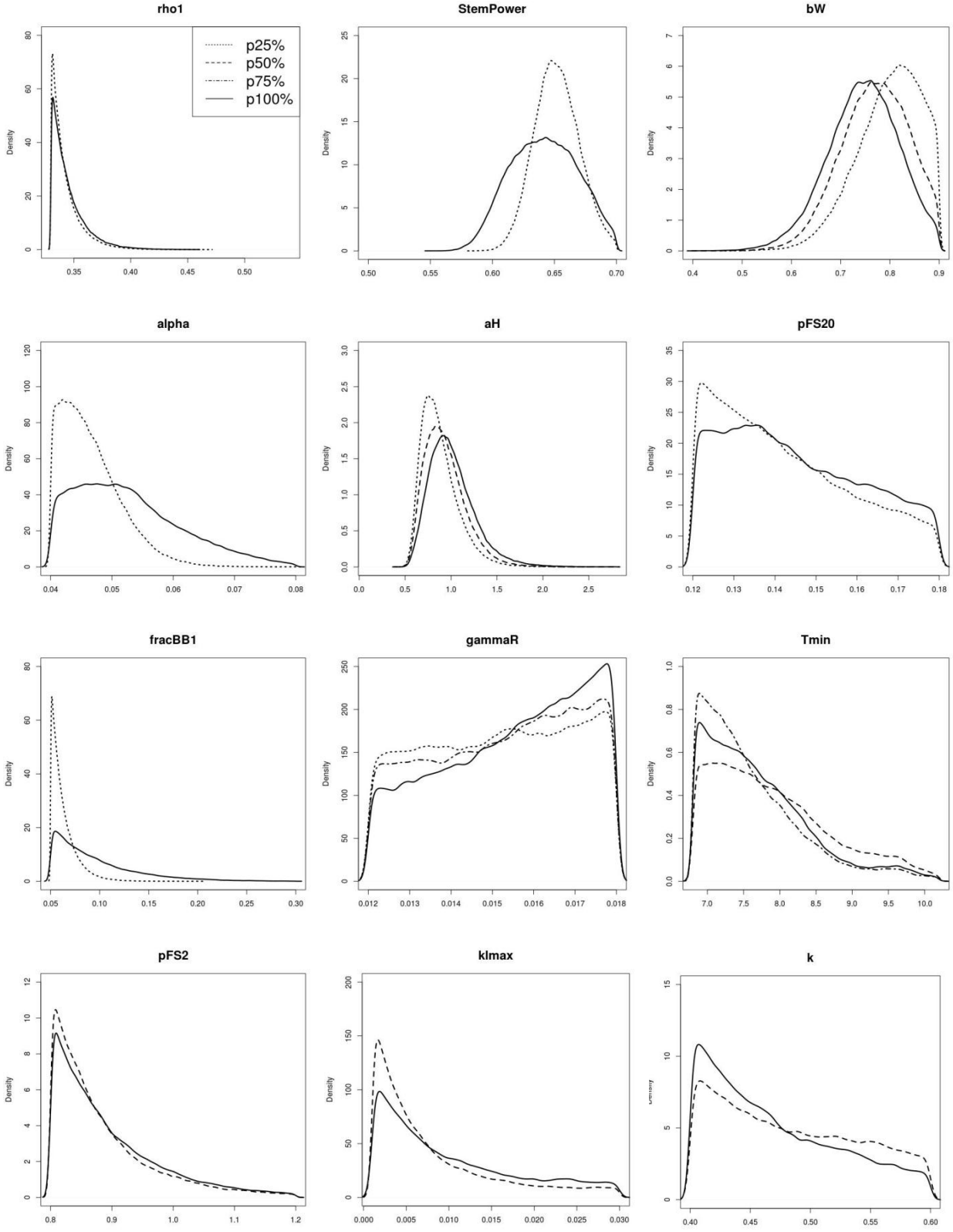

Figure 1. (Concluded).

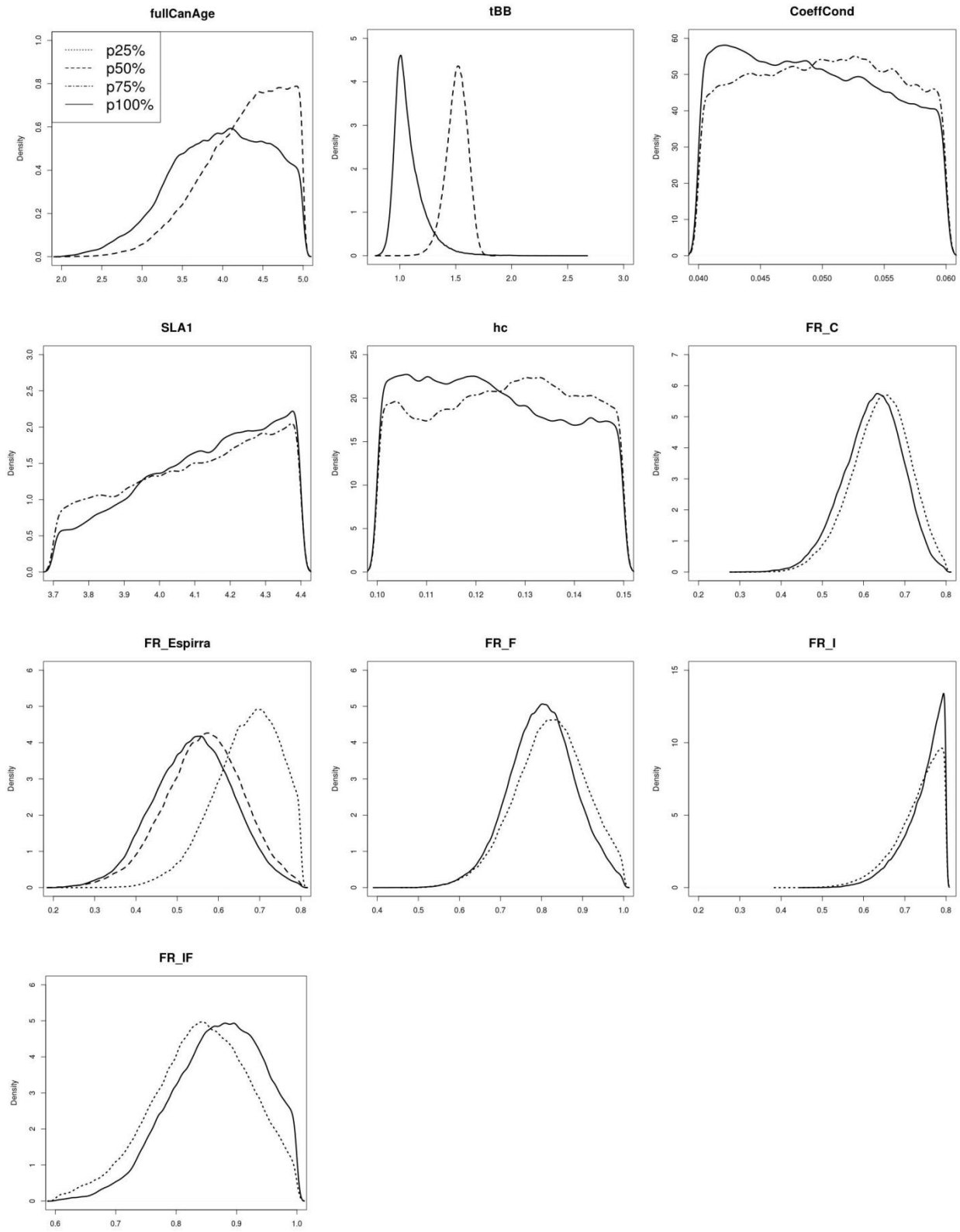

Table 4a. Pearson product-moment correlation coefficients (r) of 3PGN parameters higher than 0.24 or lower than -0.24, for p100% and p75%.

p100%		p75%	
parameters	*R*	Parameters	*r*
aH, bW	-0.94	aH, bW	-0.94
tBB, fracBB0	-0.89	tBB, fracBB0	-0.9
StCn, StPw	-0.8	StCn, StPw	-0.82
Y, alpha	-0.53	Y, alpha	-0.53
fN0, FR_Espirra	-0.37	fN0, FR_Espirra	-0.39
fN0, FR_C	-0.29	fN0, FR_C	-0.3
alpha, FR_F	-0.28	alpha, FR_F	-0.27
fCanAge, dmC	-0.26	alpha, k	-0.26
alpha, k	-0.24	fCanAge, FR_Espirra	-0.25
Topt, FR_Espirra	-0.27		

Table 4b. Pearson product-moment correlation coefficients (r) of 3PGN parameters higher than 0.24 or lower than -0.24, for p50% and p25%.

p50%		p25%	
parameters	*R*	Parameters	*r*
aH, bW	-0.93	aH, bW	-0.92
StCn, StPw	-0.8	alpha, Topt	-0.44
tBB, fracBB0	-0.54	alpha, StPw	-0.4
Y, alpha	-0.53	alpha, FR_I	-0.33
fN0, FR_Espirra	-0.34	alpha, FR_IF	-0.32
fN0, FR_C	-0.33	alpha, FR_F	-0.31
Topt, Tmin	0.27	alpha, FR_C	-0.31
alpha, FR_F	-0.26	StPw, FR_F	-0.29

The log-likelihood distribution (logL) associated with the posterior parameter distribution provides a measure of model fit and output uncertainty. We used the logL for this purpose because it integrates information about all different model output variables into one measure. Better model fit corresponds to higher values of logL and a high variance of logL is indicative of high variability of model output for the posterior parameter sample, i.e. high output uncertainty. In this case the likelihood is proportional to the joint posterior distribution because a uniform prior was used. Figure 2 shows logL distributions of p25%, p50%, p75% and p100%, giving an idea of how model fit changes for the different calibrations; in other words Figure 2 shows to which extent parameter screening affects the calibration process. Model fit decreased when fewer parameters were included in the calibrations. logL assumed lowest values for the calibrations that involved just 25% of parameters, while the highest logL values were achieved by p75% and p100%. Model output uncertainty decreased for smaller sets of parameters, while p75% and p100% were characterized by the highest uncertainty in model output. The log-likelihood of the calibrations with 75% parameters and with the whole parameter set had similar

distributions (Figure 2). In order to provide a more direct measure of model performance, the normalised root mean squared errors (NRMSE) were calculated. From the posterior distribution of each calibration 1000 parameter vectors were sampled. The outputs generated from these samples were averaged and used to calculate the NRMSEs; Table 5 shows the prediction errors aggregated for each output variable. p75% and p100% had similar NRMSEs for all outputs; while the errors of p25% and p50% for predictions of P_E, *V*, *WF*, *WR* and *WS* were significantly higher than those of p75% and p100%, meaning that there was a model fit degradation when smaller subsets of parameters were calibrated.

According to our results, the parameters that had a highest canonical cross-loading lower than 0.02 could be discarded from the calibration, because they did not affect model fit. In fact, in addition to the parameters reported in the fourth column of Table 3, the 35th parameter (*hc*) was also not influential on the likelihood and the NRMSEs (data not shown).

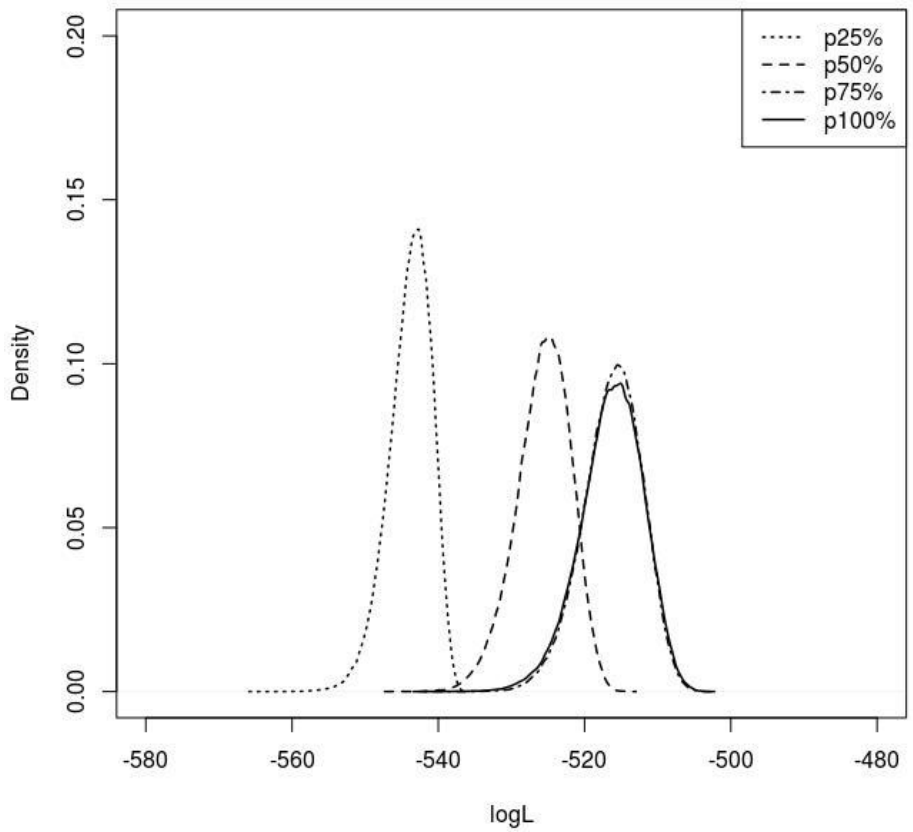

Figure 2. log-likelihood (logL) distributions for the four calibrations carried out.

Table 5. Normalised root mean squared errors of the four calibrations (i.e., p25%, p50%, p75% and p100%) for each data type.

	P_E	*D*	*H*	*V*	*WF*	*WR*	*WS*
p100%	96.9	20.6	32.4	52.6	58.2	44.8	59.1
p75%	96.8	20.4	33	51.9	57.8	44.5	58.6
p50%	102.8	19.9	33.4	57.4	65	48.6	63.4
p25%	109.4	19.5	30.2	89.3	66.4	50.4	65.8

Discussion

The impact of parameter screening on Bayesian calibration

During the last decades, the use of Bayesian statistics has increased substantially in biological science (van Oijen et al., 2013; Correia et al., 2012; Ogle & Barber, 2008). However, Bayesian calibration of parameter-rich models, like process-based forest models, is still challenging. The calibration process is computationally demanding and can be prohibitive when many parameters are involved, simulated time periods are long, or the time step of the model is short. The practice of limiting the BC to just a subset of model parameters (Xenakis et al., 2008; Van Oijen et al., 2011) is one potential solution for Bayesian calibration of complex models. For the first time, we investigated if this practice has an impact on model performance and model uncertainty. Our results showed that parameter screening for BC must be carried out carefully. We found that only parameters to which the model was least sensitive could be excluded from the calibration without strongly affecting the *a posteriori* behaviour of the model. When the calibration was limited to a more restricted subset (p25% or p50%), some influential parameters ended up being fixed at constant values; also parameter interactions were changed (Table 4). As a consequence, we obtained different joint posterior probability distributions from the p25% and p50% calibrations than from p100%, which must be considered the best parameterization of 3PGN, taking into account uncertainty about all parameters. Even though the same model was used, the likelihood distributions of p25% and p50% assumed lower values than the likelihood distributions of p75% and p100% (Figure 2) and the NRMSEs of p25% and p50% were higher than the prediction errors of p75% and p100% (Table 5). Hence the exclusion of parameters from BC can lead to a reduction of model performance and an underestimation of the uncertainty associated with model predictions.

As in all Bayesian approaches, the choice of the prior can influence the calibration (Efron, 2013). ‘Strong’ priors could influence the range of likelihood-values that are sampled by the MCMC, especially when the dataset used for calibration is characterized by a low number of measurements and/or high measurement uncertainty. An extreme situation would be that of a highly informative prior, where all parameters are considered to be known with high accuracy and precision. That would effectively shield the joint parameter distribution from subsequent calibration. We examined here the more common situation in process-based vegetation modelling, where considerable parameter uncertainty exists and calibration is required. The values assigned to the parameters that were removed from the calibration process in p75%, p50% and p25% could affect model fit. In this exercise we choose the average value between the minimum and maximum of each parameter. The more those

values were distant from the maximum *a posteriori* parameter vector of p100% the more the model fit is expected to deteriorate; but it depends also on the sensitivity of the model to the parameters.

Our results showed that parameter screening can be done if limited to the least important parameters. In fact, p75% had nearly the same joint posterior distribution and the same likelihood distribution as p100%. Hence it may be possible to reduce the computational load of BC by excluding from the calibration process those parameters that have negligible influence on model output variables. Reducing dimensionality is attractive because convergence of an MCMC requires that all parameters have converged to their marginal distribution, and correlations between parameters may hamper convergence if the proposal distribution is not adaptive (Gill, 2008). Roberts et al. (1997) proved, albeit for Gaussian distributions rather than process-based models such as 3PGN, that the optimal acceptance rate in an MCMC decreases with dimensionality of the distribution. Parameter screening may therefore accelerate a BC carried out by means of MCMC, but the screening does pose a risk: parameters that are not important for some outputs could have strong impact on other outputs or could become more influential in different conditions (e.g. different environmental conditions or stand age). For instance in our study, the age related stress parameters (i.e., nAge, rAge and MaxAge) are the parameters to which 3PGN was least sensitive. But sensitivity analyses were carried out considering model outputs at 12 years, an age at which Eucalyptus plantations are commonly cut in Portugal. The impact of those parameters on model outputs could increase when simulating old stands. Sensitivity analysis is a key process that should always be carried out over the parameter space before the calibration, this will help modelers to better interpret model behavior in representing the natural processes.

In addition, the method presented here of comparing model performance using the full parameter set and different parameter subsets, selected by means of sensitivity analysis, could also be used for model structure simplification if a comprehensive and complete dataset is available.

On the use of CCA for parameter screening

This work is the first attempt to use canonical correlation analysis to quantify the parameter sensitivity of a process-based forest model, and to use the results in parameter screening. Our results about model sensitivity are in agreement with other studies that have already explored the sensitivity of 3PGN, through local sensitivity analyses (Xenakis *et al.*, 2008) and global sensitivity analysis (Minunno *et al.*, 2013). However the aim of this work was not to explore model sensitivity but to provide a method for parameter screening by means of canonical correlation analysis, in order to increase the efficiency of the BC. Results showed that model performances are strongly affected by the parameter selection used in the calibration, so it is important to find a robust and reliable method for parameter screening. The Morris method (Campolongo et al., 2007; Morris, 1991) is a GSA that has already been used to screen parameters of a process-based forest model (van Oijen et al., 2011). This method is efficient, requires a relatively small number of runs, and is therefore particularly suitable for parameter-rich models. But Morris screening is applied to one output variable at a time and it is not straightforward to obtain a parameter ranking that relates to all outputs of the model. The procedure presented in this study based on canonical correlation analysis is an alternative to Morris screening, because it is not too computationally demanding and provides an overall ranking in relation to all

outputs of the model. CCA is particularly suitable when several inputs and outputs are involved in the analyses, and is most appropriate for models that do not appear to be especially non-linear in the parameters.

The number of parameters that can be left out from the calibration process is highly case-specific and depends on the model and the data; but the methodology introduced here is generally applicable. However, we are far from finding an optimal solution for parameter screening, and there are still a number of issues to be considered. CCA can rank the relationships between parameters and model outputs by means of the highest canonical cross-loadings. But from a Bayesian perspective we are mostly interested in the impact that parameter changes have on the likelihood. Unfortunately, involving the likelihood in the sensitivity analysis is not straightforward, because of the difficulty in quantifying the change in the likelihood in a robust way. The main problem is that the likelihood function tends to be highly peaked in parameter space and any global sensitivity analysis technique may not sample the area of high likelihood intensively enough. We therefore did not carry out a GSA of the likelihood itself but instead a CCA involving all model output variables as composite variates. Future work should investigate how to improve the screening method provided here, integrating the likelihood in a methodology that can be generally applied.

Conclusions

We introduced a new methodology for parameter screening, based on canonical correlation analysis. This methodology can be generally applied and is particularly suitable for complex process-based models because it is not computationally demanding and is easy to implement. Furthermore it provides an overall ranking in relation to all outputs of the model, as opposed to common GSA-methods that can only analyse the sensitivity of one model output variable at a time.

We applied the screening method to a process-based forest model to select parameters that could be excluded from calibration. We used Bayesian calibration and quantified, for the first time, the impact of parameter screening on calibration and subsequent performance of a process-based forest model. In this case study, about 25% of 3PGN parameters could be excluded from the calibration without affecting model performance. The percentage of parameters that can be excluded without significantly influencing the results will vary with the model and the observations used.

Acknowledgments

We thank Jonathan Rougier who suggested to us the idea of combining Bayesian calibration of parameter-rich models with canonical correlation analysis. 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).

Bibliography

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

Andrieu, C. & Thoms, J. (2008). A tutorial on adaptive MCMC. *Statistics and Computing*, 18: 343-373.

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. and D. G. R. B. T.-A. in E. Research (Ed.), (Vol. 30, pp. 113-175). Academic Press.

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

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

Correia, A. C., Minunno, F., Caldeira, M. C., Banza, J., Mateus, J., Carneiro, M., Wingate, L., et al. (2012). Soil water availability strongly modulates soil CO2 efflux in different Mediterranean ecosystems: Model calibration using the Bayesian approach. *Agriculture, Ecosystems & Environment*, *161*(null), 88–100. doi:10.1016/j.agee.2012.07.025

Cukier R.I., Fortuin C.M., Shuler K.E., Petschek A.G., and Schaibly J.H. 1973. Study of the Sensitivity of coupled reaction systems to uncertainties in rate coefficients. I. Theory. *Journal of Chemical Physics* 59: 3873–3878.

Efron, B. (2013). Bayes' Theorem in the 21st Century. *Science*, *340*(6137), 1177–1178. doi:10.1126/science.1236536.

Esprey, L., Sands, P., & Smith, C. . (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 10). Models for supporting forest management in a changing environment. *Forest Systems*.

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*(4), 457-511. Institute of Mathematical Statistics.

Gill, J. (2008). Is partial-dimension convergence a problem for inferences from MCMC algorithms? Political Analysis 16: 153-178.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate Data Analysis*. (Joseph F Hair, R. E. Anderson, R. L. Tatham, & W. C. Black, Eds.) *International Journal of Pharmaceutics* (Vol. 131, p. 816). Prentice Hall. doi:10.1016/j.ijpharm.2011.02.019

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, *28*, 312– 377.

Johnsen K., Samuelson L., Teskey R., McNulty S., & Fox T. (2001). Process Models as Tools in Forestry Research and Management. *Forest Science*, *47*(1), 7. Society of American Foresters.

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., & 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

Lindner, M., Maroschek, M., Netherer, S., Kremer, A., Barbati, A., Garcia-Gonzalo, J., Seidl, R., et al. (2010). Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems. *Forest Ecology and Management*, *259*(4), 698-709. doi:10.1016/j.foreco.2009.09.023

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. doi:10.1016/j.envsoft.2012.12.010

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.

Ogle, K., & Barber, J. J. (2008). Bayesian Data—Model Integration in Plant Physiological and Ecosystem Ecology . In U. Lüttge, W. Beyschlag, & J. Murata (Eds.), *Progress in Botany* (Vol. 69, pp. 281-311). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-540-72954-9

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. *Biogeosciences*, *4*(5), 791–802.

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

Roberts, G.O.; Gelman, A.; Gilks, W.R. (1997). "Weak convergence and optimal scaling of random walk Metropolis algorithms". *Annals of Applied Probability*, 7 (1): 110-120

Saltelli, A., S. Tarantola, F. Campolongo, M. Ratto (2004). Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models. John Wiley & Sons.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., & Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, *181*(2), 259–270. doi:10.1016/j.cpc.2009.09.018

Sands, P., & Landsberg, J. (2001). 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.).

Sobol', I. M. 1993. Sensitivity Analysis for nonlinear mathematical models. *Mathematical Modeling & Computational Experiment 1*: 407-414

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

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), 1621-1609. doi:10.1016/j.agrformet.2011.06.017

Van Oijen, M., Reyer, C., Bohn, F. J., Cameron, D. R., Deckmyn, G., Flechsig, M., Härkönen, S., et al. (2013). Bayesian calibration, comparison and averaging of six forest models, using data from Scots pine stands across Europe. *Forest Ecology and Management*, *289*, 255–268. doi: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*(7), 915-927. doi:10.1093/treephys/25.7.915