INSTITUTO SUPERIOR DE AGRONOMIA
Universidade de Lisboa

ON THE USE OF THE BAYESIAN APPROACH FOR THE CALIBRATION, EVALUATION AND COMPARISON OF PROCESS-BASED FOREST MODELS

TESE APRESENTADA PARA OBTENCÃO DO GRAU DE DOUTOR EM
ENGENHARIA FLORESTAL E DOS RECURSOS NATURAIS

FRANCESCO MINUNNO

ORIENTADOR(A): Dr. JOÃO Manuel Dias dos SANTOS PEREIRA
CO-ORIENTADOR(A): Dr. Maria MARGARIDA Branco de Brito Tavares TOMÉ
CO-ORIENTADOR(A): Dr. SOFIA Cerasoli SALVATORI

JÚRI:

Presidente: Reitor da Universidade de Lisboa

Vogais: Doutor João Manuel Dias dos Santos Pereira
Professor Catedrático aposentado
Instituto Superior de Agronomia da Universidade de Lisboa;

Doutor José Miguel Oliveira Cardoso Pereira
Professor Catedrático
Instituto Superior de Agronomia da Universidade de Lisboa;

Doutora Isabel Cristina Maciel Natário
Professor Auxiliar
Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa;

Doutora Fernanda Maria doe Reis Torroaes Valente
Professor Auxiliar
Instituto Superior de Agronomia da Universidade de Lisboa;

Doutor Nuno Miguel Matias Carvalhais
Investigador
Max Planck Institute for Biogeochemistry, Germany.

LISBOA
2014

LISBOA | UNIVERSIDADE DE LISBOA

ON THE USE OF THE BAYESIAN APPROACH FOR THE CALIBRATION, EVALUATION AND COMPARISON OF PROCESS-BASED FOREST MODELS

TESE APRESENTADA PARA OBTENCÃO DO GRAU DE DOUTOR EM ENGENHARIA FLORESTAL E DOS RECURSOS NATURAIS

FRANCESCO MINUNNO

ORIENTADOR(A): Dr. JOÃO Manuel Dias dos SANTOS PEREIRA
CO-ORIENTADOR(A): Dr. Maria MARGARIDA Branco de Brito Tavares TOMÉ
CO-ORIENTADOR(A): Dr. SOFIA Cerasoli SALVATORI

JÚRI:

Presidente: Reitor da Universidade de Lisboa

Vogais: Doutor João Manuel Dias dos Santos Pereira
Professor Catedrático aposentado
Instituto Superior de Agronomia da Universidade de Lisboa;

Doutor José Miguel Oliveira Cardoso Pereira
Professor Catedrático
Instituto Superior de Agronomia da Universidade de Lisboa;

Doutora Isabel Cristina Maciel Natário
Professor Auxiliar
Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa;

Doutora Fernanda Maria doe Reis Torroaes Valente
Professor Auxiliar
Instituto Superior de Agronomia da Universidade de Lisboa;

Doutor Nuno Miguel Matias Carvalhais
Investigador
Max Planck Institute for Biogeochemistry, Germany.

LISBOA
2014

Contents

Chapter 1

Chapter 2

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

Chapter 3

"Soil water availability strongly modulates soil CO2 efflux in different Mediterranean ecosystems: Model calibration using the Bayesian approach"

Chapter 4

"Bayesian calibration, comparison and averaging of six forest models, using data from Scots pine stands across Europe"

Chapter 5

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

Chapter 6

Chapter 7

“Sobre o uso da estatística Bayesiana para a calibração, avaliação e comparação dos modelos florestais de base processual.”

Resumo

Os ecossistemas florestais têm experimentado rápidas e abruptas mudanças das condições ambientais que podem aumentar a vulnerabilidade a eventos extremos, como seca, ondas de calor, incêndios. É possível recorrer a modelos de base processual para fazer inferências sobre futuras dinâmicas ambientais, mas a robustez e a fiabilidade desses modelos depende da estrutura e da parametrização.

O objetivo principal do doutoramento foi implementar e aplicar modernas técnicas de cálculo, baseadas, sobretudo, na estatística Bayesiana, no âmbito da modelação florestal. Apresenta-se uma variedade de casos de estudo, abrangendo modelos de crescimento, modelos bioclimáticos e modelos de base processual, e demonstra-se o grande potencial do método Bayesiano na redução da incerteza dos parâmetros e dos *outputs*, bem como na avaliação dos modelos.

Finalmente, desenvolveu-se uma nova metodologia, onde se aplicou um *framework* Bayesiano em conjunto com uma análise de sensibilidade global para incrementar a estrutura dum modelo de base processual e explorar melhor o comportamento desse modelo.

Uma parte do doutoramento foi focada na redução da carga computacional para tirar o máximo proveito da estatística Bayesiana. Mostra-se como o *parameter screening* afeta as *performances* dos modelos e apresenta-se uma nova metodologia para a seleção dos parâmetros baseada na análise de correlação canónica.

Palavras-chave: modelos de base processual, estatística Bayesiana, ciclos do carbono e da agua, analise de incerteza, analise de sensibilidade global.

“On the use of the Bayesian approach for the calibration, evaluation and comparison of process-based forest models.”

Abstract

Forest ecosystems have been experiencing fast and abrupt changes in the environmental conditions, that can increase their vulnerability to extreme events such as drought, heat waves, storms, fire. Process-based models can draw inferences about future environmental dynamics, but the reliability and robustness of vegetation models are conditional on their structure and their parametrisation.

The main objective of the PhD was to implement and apply modern computational techniques, mainly based on Bayesian statistics, in the context of forest modelling. A variety of case studies was presented, spanning from growth predictions models to soil respiration models and process-based models. The great potential of the Bayesian method for reducing uncertainty in parameters and outputs and model evaluation was shown.

Furthermore, a new methodology based on a combination of a Bayesian framework and a global sensitivity analysis was developed, with the aim of identifying strengths and weaknesses of process-based models and to test modifications in model structure.

Finally, part of the PhD research focused on reducing the computational load to take full advantage of Bayesian statistics. It was shown how parameter screening impacts model performances and a new methodology for parameter screening, based on canonical correlation analysis, was presented.

Key-words: process-based models, Bayesian statistics,carbon and water cycles, uncertainty analysis, global sensitivity analysis.

Acknowledgments

At this point many names come to my mind …

First of all my supervisor Prof. João Santos Pereira for receiving me in his group and for the freedom to go on my own way.

Prof. Margarida Tomé for introducing me in the “COST Action world”, giving me the opportunity to start collaborations with many foreigner institutions and groups.

Prof. Maurizio Mencuccini that addressed me towards the Bayesian path.

I am especially grateful to my mentor Dr. Marcel van Oijen, for the long discussions and the precise answers to my chaotic questions. I am deeply thankful for the opportunity that I had of working with him.

Dr. Cameron David for the precious help in the programming tasks and the interesting discussions on Bayesian topics.

My “office mates” Dr. Alexandra Correia and Dr. Raquel Lobo do Vale for the numerous discussions and for patiently endured my noisy computer.

Josep Crous Duran, Juan Guerra, Dr. João Palma, Dr. Luis Fontes and Dr. Susanna Barreiro, for their support, company and help.

Argo & Paco for their company during the many long reflexive walks.

The PhD was founded by the Fundação para a Ciência e a Tecnologia. Scholarship reference: SFRH / BD / 65934 / 2009.

The short term scientific missions were founded by the COST Action FP0603 “Forest models for research and decision support in sustainable forest management”.

Il ringraziamento piú importante va sicuramente a Sofia, per avermi portato in questo paese meraviglioso e per appoggiare sempre le mie idee e le mie scelte.

CHAPTER 1

“Introduction”

Chapter 1

1 Introduction

Forests are essential for the life on Earth. They represent a multi-functional resource that provides many good and services such as timber for construction, fuel wood, paper, recreation. Furthermore they represent the most important oxygen suppliers on land and they are attributed a key role as a biotic sink for atmospheric carbon dioxide. During the last centuries forest ecosystems have been experiencing fast and abrupt changes in the environmental conditions, due to biotic and abiotic factors (Bonan, 2008). Understanding and predicting how these changes will affect vegetation dynamics is a challenging question in ecology and forest science (Grace, 2004), that can be adressed by forest modelling.

Fontes *et al.* (2010) reviewed the forest models used in Europe in the light of tools for forest management. They classified the models in empirical models (EM), process-based models and hybrid models (HM), analysing the strength and weaknesses of each category. EMs, based on statistical analyses, put in relation variables available form forest inventories and data collected *in situ* with target variables such as timber volume. Process-based models (PBM) provide a description of the vegetation communities by modelling the key processes that underlay plant physiology (Hartig et al., 2012) and can help drawing inferences about future dynamics of forest ecosystems. PBMs have been widely used for research purposes as well as for practical forest management (Johnsen et al., 2001; Mäkelä et al., 2000). Finally, HMs combine mechanistic components with empirical elements. However, Korzukhin *et al.* (1996) stated that all the empirical models have causal elements and *viceversa* all the mechanistic models have empirical elements and it is just a matter of the level and accuracy at which we are looking at the processes.

The reliability and robustness of dynamic vegetation models are conditional on their structure but also on their parametrisation (Minunno et al., 2013a; Van Oijen et al., 2005). PBMs are usually built using a bottom-up approach; i.e., independent experiments and/or empirical observation are used to directly determine the basic processes and to quantify parameters. However a direct parametrisation is demanding because it requires direct observations for all the parameters and it can be prohibitive when models are complex and observations scarce.

In environmental sciences the way in which models and data are used to draw inference and make predictions is changing (Clark, 2007). Large amount of data such as eddy fluxes, national forest inventory data and remotely sensed data are becoming available because of the development of new techniques. These kind of data are extremely useful for model calibration and validation because they provide information at different scales in space and time. But these data can not be used for direct parametrisation because they describe variables that depend on multiple interacting processes. This large amount of information can still be used for model calibration using an inverse approach, i.e. the model structure and parameter estimates can be inferred by comparing model outputs with the observed data.

Bayesian statistics, based on probability theory, is a logical choice for model calibration, evaluation and averaging. Using the Bayesian approach it is possible to have a deep understanding of model behaviour. By quantifying the uncertainties in the data and model parameters, identifying the key parameters and outputs and evaluating model structure, the Bayesian method can help to answer a key question that all the modellers have: “Should I trust my model?” .

The Bayesian calibration (BC) provides parameter estimates and quantifies measurement and modelling uncertainties. In Bayesian statistics the bottom-up (direct parameterisation) and the top-down (inverse parameterisation) approaches can be used in combination to improve the knowledge about parameters and model structure (Hartig et al., 2012). The bottom-up approach can be used in determining the prior, allowing the integration of different data sources in the calibration process. This approach has the merit of redressing the parameters towards realistic values and this aspect is particularly important in process-based modelling where the parameters often have a physical or a physiological meaning. In contrast, through a top-down perspective, stand variables and data collected at large temporal and spatial scales, by means of the likelihood, can be used to inform parameters that are highly variable or difficult to measure. Furthermore Bayesian statistics, by means of Bayesian model comparison (BMC), allows to evaluate multiple models in the light of their relative uncertainty and quantifies, for each model, the probability of having the correct structure (Kass and Raftery, 1995). Finally, Bayesian model averaging (BMA) allows to draw inferences using multiple models, considering the uncertainty quantified in the selecting model process (Hoeting et al., 1999). In other words, BMA uses the different model probabilities derived from a BMC to weight the predictions of different models.

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 (Ogle and Barber, 2008) because of developments of more powerful computational tools. 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).

2 Process-based forest modelling and Bayesian approach

The Bayes' theorem was formulated by reverend Thomas Bayes during the 1740s; however its modern mathematical form was given by Pierre Simon Laplace in the 1812. Nowadays the use of Bayesian statistics is becoming increasingly common in all science fields because of developments in computer science and sampling based techniques such as Markov Chain Monte Carlo simulations (MCMC) (Ogle and Barber, 2008; van Oijen et al., 2005).

A good introduction to the Bayesian statistics can be found in the book “Data analysis. A Bayesian tutorial” (Sivia and Skilling, 2006) and in “Estatística bayesiana” (Paulino et al., 2003). Gelman et al. (2004) published a more comprehensive book on Bayesian methods: “Bayesian Data Analysis”; while Clark in his book “Models for Ecological Data. An introduction” presented the emerging Bayesian approach from the prospective of an ecological and environmental modeller (Clark, 2007).

Pioneering works on the application of the Bayesian method for parameter estimations of forest models appeared in the literature in the last decade (Gertner et al., 1999; Green et al., 2000, 1999; Radtke et al., 2002). Van Oijen et al. (2005) provided an example of Bayesian calibration of a process-based forest model and showed the potential of the method, showing how data availability and data accuracy affect the uncertainty of parameters and model outputs. Xenakis et al. (2008) carried out a Bayesian calibration and a sensitivity analysis of the 3-PGN model for Scots pine in Scotland. 3-PGN is a modified version of a forest process-based model 3-PG (Landsberg and Waring, 1997; Sands and Landsberg, 2002) that have been widely used in the world for research purposes as well as for forest management. Patenaude et al. (2008) integrated remotely sensed data (i.e., hyperspectral, LiDAR, SAR) and field-based data for the calibration of 3-PG for the Corsican pine in UK.

In a report published in 2008 by the centre for ecology and hydrology, van Oijen (2008) described how the Bayesian statistics can be used for model calibration and comparison, giving few examples of simple models. In this document the Bayesian model comparison (BMC) was introduced in forestry as a tool for the evaluation of model performances. Tuomi et al. (2008) used the Bayesian model comparison to evaluate different heterotrophic soil respiration models.

Van Oijen et al. (2011) introduced a Bayesian framework consisting of model calibration, model comparison and analysis of model-data mismatch. This framework was used to evaluate four different parameter-rich process-based models of forest biogeochemistry.

The application of Bayesian statistics to forest modelling, as previously stated, has been hindered by the computational load. To reduce the computational load and take full advantage of Bayesian statistics also for complex models, it is possible to act in two directions. First, the efficiency of the sampling based techniques can be increased by using more complex algorithms (Andrieu and Thoms, 2008). Second, the number of parameters involved in the calibration can be reduced by means of parameter screening.

The simplest MCMC algorithm is the Metropolis-Hastings (*M*) random walk. Robert & Casella (2005) provided a detailed description of the algorithm; van Oijen et al. (2005) for the first time applied *M* in the context of process-based forest modelling. Recently, more sophisticated algorithms, based on the Metropolis-Hastings random walk, have been developed; a review of adaptive MCMC algorithms is provided by Andrieu and Thoms (Andrieu and Thoms, 2008). The MCMC algorithms of the Differential evolution Markov chain family (Laloy and Vrugt, 2012; Ter Braak, 2006; ter Braak and Vrugt, 2008; Vrugt et al., 2009a, 2009b) have shown good performances with extremely rich parameter models (models with more than 200 parameters). Minunno et al. (in prep.) compared the performances of six Markov chain Monte Carlo algorithms of different complexity in order to understand their suitability for the calibration of process-based forest models. The differential evolution Markov Chain with fewer chains and snooker updater algorithm (DE-MCzs) (ter Braak and Vrugt, 2008), showed more efficient performances then the other algorithms that were tested in this exercise.

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. Parameter screening has already been used with forest models to select parameters for Bayesian calibration (Van Oijen et al., 2013; van Oijen et al., 2011; Xenakis et al., 2008).

3 Objectives

The main objective of this work is to apply the Bayesian methods to forest modelling, focusing on the implementation of modern computational methods for the calibration, comparison and averaging of models used in forest research.

More in specific the main activities of this research focused on achieving the following objectives:

1. implementing the Bayesian calibration and its extensions, i.e., Bayesian model comparison and Bayesian model averaging, in the context of forest modelling;
2. developing methodologies that allow to identify strengths and weaknesses of process-based models and to evaluate modifications in model structure.
3. reducing the severity of the model calibration problem in order to extend the use of

Bayesian statistics to complex process-based forest models.

4 Thesis overview

The following chapters consists of four papers published in peer-reviewed international journals.

Chapter 2 is a work published in the journal “Environmental modelling and software”, titled: “Using a Bayesian framework and global sensitivity analysis to identify strengths and weaknesses of two process-based models differing in representation of autotrophic respiration”. In this exercise a global sensitivity analysis (Morris method) was used in combination with the Bayesian framework developed by van Oijen et al. (2011). We showed how the new framework can be used to identify strength and weaknesses of process-based models and to test improvements in the structure of a process-based model. In this work, the 3-PGN model was calibrated and enhanced using a comprehensive dataset of Eucalyptus globulus plantations in Portugal; the dataset consisted of eddy covariance and stand variable measurements. The methodology presented here has general applicability.

Chapter 3 consists of a paper published in “Agriculture, Ecosystems and Environment” and titled: “Soil water availability strongly modulates soil CO2 efflux in different Mediterranean ecosystems: Model calibration using the Bayesian approach”. Correia et al. (2012) used the Bayesian framework developed by Minunno et al., (2013) to test the effectiveness of soil bioclimatic models in estimating soil respiration on a daily and monthly time step. Field chamber measurements of soil respiration from forest and grassland sites of centre and south of Portugal were used to evaluate the models. For this work I carried out all the modelling analyses and contributed to the *results* and the *discussion* sections.

Chapter 4 is a work published in “Forest Ecology and Management” from the title: “Bayesian calibration, comparison and averaging of six forest models, using data from Scots pine stands across Europe”. Van Oijen et al. (2013) compared models of different complexity using data of Scots pine forests coming from national forest inventory data and permanent sample plot data spread across Europe. In this work, for the first time, the Bayesian model averaging (BMA) was applied to forest models and used to make predictions that take into account of both parametric and model structural uncertainty. In this paper I carried out the analyses for the 3-PG model, as well as I contributed to the *results* and *discussion* sections.

The last chapter consists of a work published in the “Journal on Uncertainty Quantification” and titled: “Selecting parameters for Bayesian calibration of a process-based model: a methodology based on canonical correlation analysis”. Minunno et al. (2013b) for the first time showed how parameter screening can impact model performances and presented a new methodology for parameter screening based on canonical correlation analysis (CCA), a multivariate technique that can be used for global sensitivity analysis. The methodology can be applied to any kind of model.

5 References

Andrieu, C., Thoms, J., 2008. A tutorial on adaptive MCMC. Stat. Comput. 18, 343–373.

Bonan, G.B., 2008. Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests. Science 320, 1444–1449.

Celpa, Associação da Industria Papeleira., 2012. Boletim estatístico, Indústria papeleira portuguesa 2011.

Clark, J.S., 2007. Models for ecological data: an introduction. Princeton University Press Princeton.

Correia, A.C., Minunno, F., Caldeira, M.C., Banza, J., Mateus, J., Carneiro, M., Wingate, L., Shvaleva, A., Ramos, A., Jongen, M., 2012. Soil water availability strongly modulates soil CO\textless sub\textgreater 2\textless/sub\textgreater efflux in different Mediterranean ecosystems: Model calibration using the Bayesian approach. Agric. Ecosyst. Environ. 161, 88–100.

Fontes, L., Bontemps, J.-D., Bugmann, H., Van Oijen, M., Gracia, C., Kramer, K., Lindner, M., Rötzer, T., Skovsgaard, J.P., 2010. Models for supporting forest management in a changing environment.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis., Second. ed.

Gertner, G.Z., Fang, S., Skovsgaard, J.P., 1999. A Bayesian approach for estimating the parameters of a forest process model based on long-term growth data. Ecol. Model. 119, 249–265.

Grace, J., 2004. Understanding and managing the global carbon cycle. J. Ecol. 92, 189–202.

Green, E.J., MacFarlane, D.W., Valentine, H.T., 2000. Bayesian synthesis for quantifying uncertainty in predictions from process models. Tree Physiol. 20, 415–419.

Green, E.J., MacFarlane, D.W., Valentine, H.T., Strawderman, W.E., 1999. Assessing Uncertainty in a Stand Growth Model by Bayesian Synthesis. For. Sci. 45, 528–538.

Hartig, F., Dyke, J., Hickler, T., Higgins, S.I., O'Hara, R.B., Scheiter, S., Huth, A., 2012. Connecting dynamic vegetation models to data – an inverse perspective. J. Biogeogr. 39, 2240–2252.

Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: a tutorial. Stat. Sci. 382–401.

ICNF,Instituto de Conservação da Natureza e das Florestas, 2013. Áreas dos usos do solo e das espécies florestais de Portugal continental, 1995/2005/2010. Resultados preliminares, v1.0.

Johnsen, K., Samuelson, L., Teskey, R., McNulty, S., Fox, T., 2001. Process models as tools in forestry research and management. For. Sci. 47, 2–8.

Kass, R.E., Raftery, A.E., 1995. Bayes Factors. J. Am. Stat. Assoc. 430, 773–795.

Korzukhin, M.D., Ter-Mikaelian, M.T., Wagner, R.G., 1996. Process versus empirical models: which approach for forest ecosystem management? Can. J. For. Res. 26, 879–887.

Laloy, E., Vrugt, J.A., 2012. High-dimensional posterior exploration of hydrologic models using multiple-try DREAM (ZS) and high-performance computing. Water Resour. Res. 48.

Landsberg, J.J., Waring, R.H., 1997. A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. For. Ecol. Manag. 95, 209–228.

Mäkelä, A., Landsberg, J., Ek, A.R., Burk, T.E., Ter-Mikaelian, M., Ågren, G.I., Oliver, C.D., Puttonen, P., 2000. Process-based models for forest ecosystem management: current state of the art and challenges for practical implementation. Tree Physiol. 20, 289–298.

Minunno, F., Van Oijen, M., Cameron, D.R., Cerasoli, S., Pereira, J.S., Tomé, M., 2013a. Using a Bayesian framework and global sensitivity analysis to identify strengths and weaknesses of two process-based models differing in representation of autotrophic respiration. Environ. Model. Softw. 42, 99–115.

Minunno, F., van Oijen, M., Cameron, D.R., Pereira, J.S., 2013b. Selecting Parameters for Bayesian Calibration of a Process-Based Model: A Methodology Based on Canonical Correlation Analysis. SIAMASA J. Uncertain. Quantif. 1, 370–385.

Ogle, K., Barber, J.J., 2008. Bayesian data—model integration in plant physiological and ecosystem ecology, in: Progress in Botany. Springer, pp. 281–311.

Patenaude, G., Milne, R., Van Oijen, M., Rowland, C.S., Hill, R.A., 2008. Integrating remote sensing datasets into ecological modelling: a Bayesian approach. Int. J. Remote Sens. 29, 1295–1315.

Paulino, C.D.M., Turkman, M.A.A., Murteira, B., 2003. Estatística bayesiana. Fundação Calouste Gulbenkian.

Radtke, P.J., Burk, T.E., Bolstad, P.V., 2002. Bayesian Melding of a Forest Ecosystem Model with Correlated Inputs. For. Sci. 48, 701–711.

Robert, C.P., Casella, G., 2005. Monte Carlo Statistical Methods. Technometrics 47, 243–243.

Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., 2004. Sensitivity analysis in practice: a guide to assessing scientific models. John Wiley & Sons.

Sands, P.J., Landsberg, J.J., 2002. Parameterisation of 3-PG for plantation grown< i> Eucalyptus globulus</i>. For. Ecol. Manag. 163, 273–292.

Sivia, D.S., Skilling, J., 2006. Data analysis a Bayesian tutorial. Oxford University Press, Oxford; New York.

Ter Braak, C.J., 2006. A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. Stat. Comput. 16, 239–249.

Ter Braak, C.J., Vrugt, J.A., 2008. Differential evolution Markov chain with snooker updater and fewer chains. Stat. Comput. 18, 435–446.

Tuomi, M., Vanhala, P., Karhu, K., Fritze, H., Liski, J., 2008. Heterotrophic soil respiration—Comparison of different models describing its temperature dependence. Ecol. Model. 211, 182–190.

Van Oijen, M., 2008. Bayesian Calibration (BC) and Bayesian Model Comparison (BMC) of process-based models: Theory, implementation and guidelines.

Van Oijen, M., Cameron, D.R., Butterbach-Bahl, K., Farahbakhshazad, N., Jansson, P.-E., Kiese, R., Rahn, K.-H., Werner, C., Yeluripati, J.B., 2011. A Bayesian framework for model calibration, comparison and analysis: Application to four models for the biogeochemistry of a Norway spruce forest. Agric. For. Meteorol. 151, 1609–1621.

Van Oijen, M., Reyer, C., Bohn, F.J., Cameron, D.R., Deckmyn, G., Flechsig, M., Härkönen, S., M. Hartig, F., Huth, A., Kiviste, A., Lasch, P., Mäkelä, A., Mette, T., Minunno, F., Rammer, W., 2013. Bayesian calibration, comparison and averaging of six forest models, using data from Scots pine stands across Europe. For. Ecol. Manag. 289, 255–268.

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

Vrugt, J.A., Ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A., Hyman, J.M., Higdon, D., 2009a. Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. Int. J. Nonlinear Sci. Numer. Simul. 10, 273–290.

Vrugt, J.A., Ter Braak, C.J., Gupta, H.V., Robinson, B.A., 2009b. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? Stoch. Environ. Res. Risk Assess. 23, 1011–1026.

Xenakis, G., Ray, D., Mencuccini, M., 2008. Sensitivity and uncertainty analysis from a coupled 3-PG and soil organic matter decomposition model. Ecol. Model. 219, 1–16.

CHAPTER 2

“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 (2013) 99–115

Contents lists available at SciVerse ScienceDirect

Environmental Modelling & Software

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

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

F. Minunno [a,*], M. van Oijen [b], D.R. Cameron [b], S. Cerasoli [a], J.S. Pereira [a], M. Tomé [a]

[a] Forest Research Center, Faculty of Agricultural Sciences, Technical University of Lisbon, Lisbon, Portugal
[b] CEH-Edinburgh, Bush Estate Penicuik, EH26 0QB, UK

ARTICLE INFO

Article history:
Received 30 January 2012
Received in revised form
10 December 2012
Accepted 13 December 2012
Available online 21 January 2013

Keywords:
3-PG
Net primary production
Respiration
Bayesian calibration
Bayesian model comparison
Morris screening
Carbon cycle
Uncertainty analysis
Global sensitivity analysis

ABSTRACT

Process-based models are powerful tools for sustainable and adaptive forest management. Bayesian statistics and global sensitivity analysis allow to reduce uncertainties in parameters and outputs, and they provide better insight of model behaviour. In this work two versions of a process-based model that differed in the autotrophic respiration modelling were analysed. The original version (3PGN) was based on a constant ratio between net and gross primary production, while in a new version (3PGN*) the autotrophic respiration was modelled as a function of temperature and biomass. A Bayesian framework, and a global sensitivity analysis (Morris method) were used to reduce parametric uncertainty, to highlight strengths and weaknesses of the models and to evaluate their performances. The Bayesian approach allowed also to identify the weaknesses and strengths of the dataset used for the analyses. The Morris method in combination with the Bayesian framework helped to identify key parameters and gave a deeper understanding of model behaviour. Both model versions reliably predicted average stand diameter at breast height, average stand height, stand volume and stem biomass. On the contrary, the models were not able to accurately predict net ecosystem production. Bayesian model comparison showed that 3PGN*, with the new autotrophic respiration model, has a higher conditional probability of being correct than the original 3PGN model.

1. 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 and 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 modelled as a fixed proportion of P_G. The understanding of the factors regulating R_{aut} is one of the most challenging questions in

---

* Corresponding author.
E-mail addresses: francescom@isa.utl.pt, checcomi@gmail.com (F. Minunno).

1364-8152/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.envsoft.2012.12.010

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 and 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 and Waring, 1992). A detailed review of the progress achieved in respiration modelling 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} modelling ($R_{maint} + R_{growth}$) has been used in many process-based models (e.g. CABALA (Battaglia, 2004); PIXGRO (Adiku et al., 2006); MAESTRO (Wang and 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 modelled 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 behaviour (Campolongo et al., 2007). 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.

2. Materials and methods

2.1. 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 (BMCs) 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.

2.2. 3PGN structure

3PGN was developed by Xenakis et al. (2008) coupling two models, 3-PG (Physiological Principles in Predicting Growth) (Landsberg and Waring, 1997) and ICBM (Introductory Carbon Balance Model) (Andrén and Kätterer, 1997). 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 vapour 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 (Landsberg, 2003). 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 (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 and Vanninen, 2000; Titus and Malcolm, 1999). Carbon decomposed from the young pools enters the old pool at a constant relative rate of humification. The fraction from each young pool that 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.

2.3. 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).

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

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.

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

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}):

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

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.

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

where W_i and r_i are dry weight and specific respiration rate, respectively, of the ith 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:

$$P_N = P_G - R_{aut} \tag{6}$$

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

Equations (4) and (5) were chosen because they required fewer parameters than other R_{maint} and R_{growth} modelling 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 and Waring, 1997).

2.4. 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).

2.4.1. 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 *E. globulus* plantation (38°38′N, 8°36′W) tended as a coppice. Originally planted in 1986 at 3 × 3 m spacing, ca. 1100 trees ha^{-1}, was 11 years old (2nd rotation) in the end of the period analysed, and ca. 20 m height. The mean annual temperature for the site is 16 °C 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).

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 biomass vs. height relationship	*aH*	–	1.9	2.8	Normal[a]
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 biomass vs. height relationship	*bW*	–	0	0.3	Weibull[a]
Defines stomatal response to VPD	*CoeffCond*	Mbar^{-1}	0.04	0.06	Uniform
Basic density	*Density*	Mg m^3	0.36	0.54	Normal
Conversion 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*[b]	–	1	3.5	Normal
Relative age to give fAge = 0.5	*rAge*	–	0.76	1	Uniform
Foliage biomass respiration rate	*rf*[b]		0.0005	0.02	Gamma
Growth respiration rate	*rg*[b]		0.2	0.3	Normal
Average monthly root turnover rate	Rttover	month^{-1}	0.012	0.018	Gamma
Woody biomass respiration rate	*rw*[b]		0.001	0.06	Gamma
Woody biomass respiration rate	*rw*[b]		0.001	0.06	Gamma
Specific leaf area at age 0	*SLA0*	m^2 kg^{-1}	10.5	14	Normal[a]
Specific leaf area for mature leaves	*SLA1*	m^2 kg^{-1}	3.7	4.4	Normal[a]
Constant in the aboveground biomass vs. diameter relationship	*StemConst*	–	1.15	1.4	Gamma[a]
Power in the aboveground biomass vs. diameter relationship	*StemPower*	–	0.5	0.55	Gamma[a]
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*	°C	32	48	Normal
Minimum temperature for growth	*Tmin*	°C	6.8	10.2	Normal
Optimum temperature for growth	*Topt*	°C	12.8	19.2	Normal
Age at which specific leaf area = (SLA0 + SLA1)/2	*tSLA*	years	1.2	2	Normal[a]
Ratio NPP/GPP	*Y*[c]	–	0.376	0.564	Normal

[a] Distributions fitted over posterior distributions.

[b] Only 3PGN* parameters.

[c] Only 3PGN parameters.

2.4.2. *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, 39°29′N, 9°13′W, 30 m a.s.l.). The mean annual temperature is 15.2 °C 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 × 3 m spacing; each seedling was supplied at planting with 200 g of a commercial fertilizer containing 14.0 g of N, 18.3 g of K and 11.6 g of P. Before planting, the soil was ploughed at 80 cm depth and 1.5 Mg ha^{-1} of dolomitic limestone (66.5% of $CaCO_3$, 32.5% of $MgCO_3$) 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 was assigned to the FR of C, F, IF; Table 1b shows the minimum and maximum values and the distributions of these parameters.

2.5. *Sensitivity analyses*

Sensitivity analyses can vary from the simplest class of the One Factor At a 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. Sensitivity analysis can be used in combination with uncertainty estimation techniques to explore the quality of parameter estimation and to identify the major sources of uncertainty in a model (Varella et al., 2010; Vezzaro and Mikkelsen, 2012). 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 and b. Factors ranged between the minimum and maximum values used for the *BC* (Table 1).

2.6. *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 $P(\theta)$. Observed data (O) that are used to update the prior distribution enter the analysis through the so-called *likelihood* function $L(\theta) = P(O|\theta)$. An updated, *posterior* distribution is then found by application of Bayes' Theorem:

$$P(\theta|O) = cP(O|\theta)P(\theta) \tag{7}$$

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

2.6.1. *Likelihood function*

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

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

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

where $\mathrm{sim}(\theta)$ 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 ith 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).

2.6.2. *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

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

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 (eqs. (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.

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

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 eqs. (1), (2) and (10) were fitted with Weibull, normal and gamma distributions and then used as prior for the *BCs* of the whole models.

2.6.3. 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 and 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 and 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 $\check{R}$ 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.

2.6.4. 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 and 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).

$$P(M_k|O) = P(O|M_k)P(M_k)/\sum P(O|M)P(M) \tag{11}$$

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 ($P(M_1) = \cdots = P(M_n)$), equation (11) becomes:

$$P(M_k|O) = P(O|M_k)/\sum P(O|M) \tag{12}$$

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

2.6.5. 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):

$$\mathrm{MSE} = \overline{(S-O)^2} = \left(\overline{S} - \overline{O}\right)^2 + (\sigma_S - \sigma_O)^2 + 2(\sigma_S\sigma_O)(1-r) \tag{13}$$

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 and Salam, 2000).

3. Results

3.1. 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 Fig. 1a–c, 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.

3.2. 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 Fig. 2. Higher values of the likelihood correspond to better model performances, while the variance of the likelihood distribution is a measure of model accuracy.

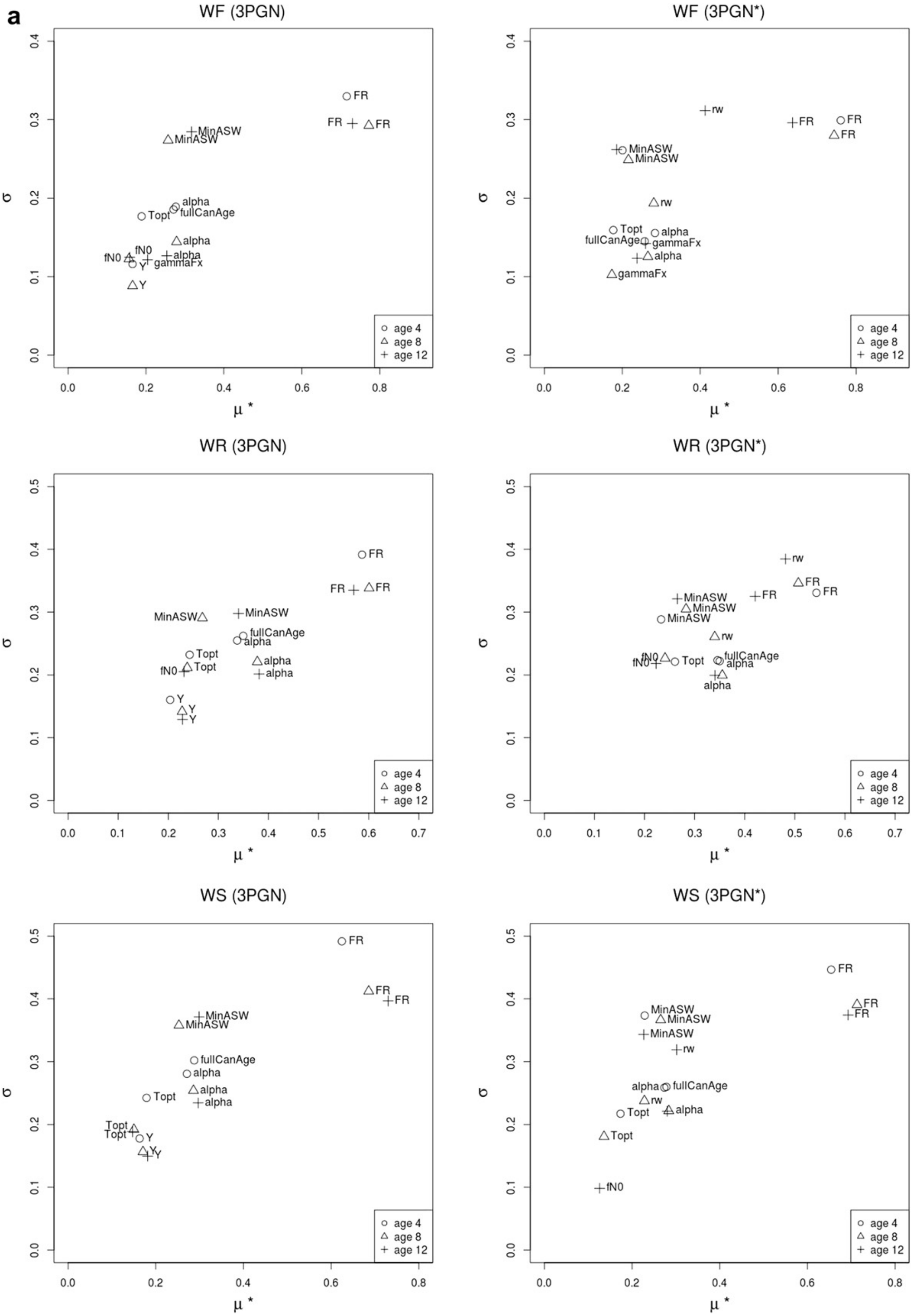


Fig. 1. a. Plots of σ vs. μ^* of the five highest sensitive parameters for foliage (WF), root (WR) and stem (WS) biomasses for 3PGN and 3PGN* outputs at age 4 (circles), 8 (triangles) and 12 (cruces). b. Plots of σ vs. μ^* of the five highest sensitive parameters for diameter at breast height (D), average stand height (H) and stand volume (V) for 3PGN and 3PGN* outputs at age 4 (circles), 8 (triangles) and 12 (cruces). c. 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).

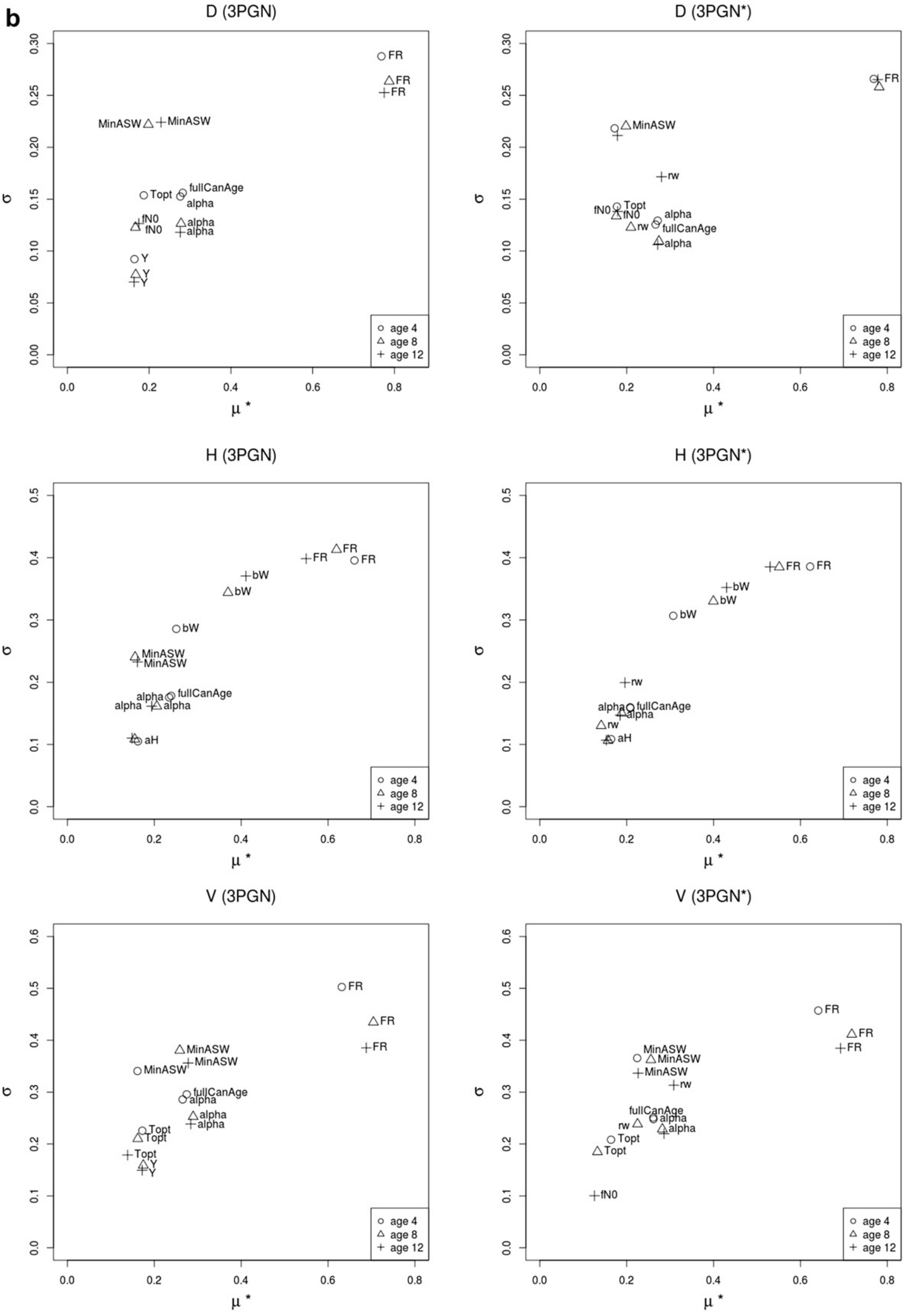


Fig. 1. (*continued*).

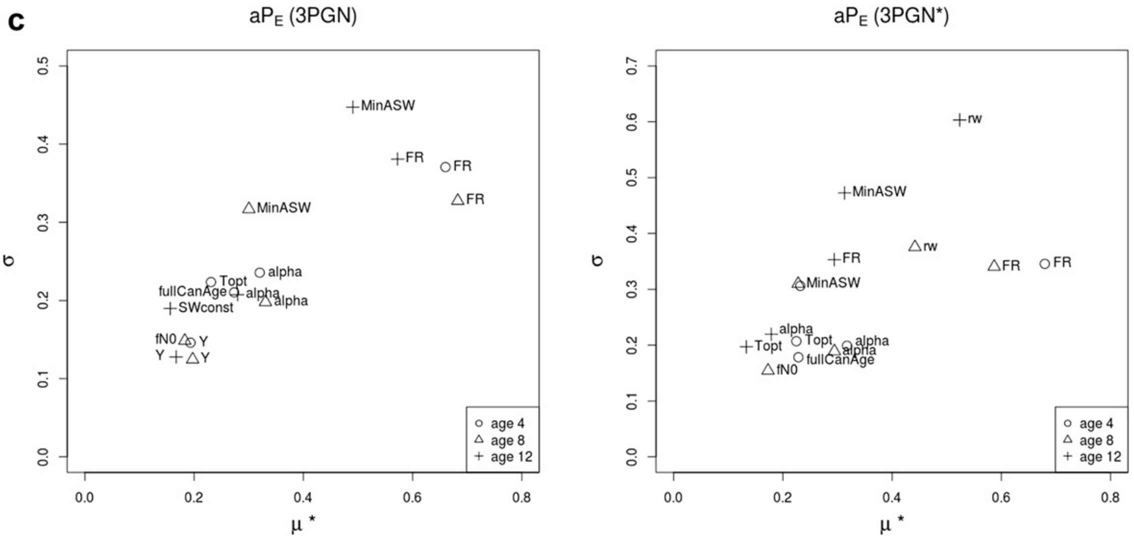


Fig. 1. (*continued*).

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*.

Fig. 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 (Fig. 3). There were differences between the posterior distributions of parameters linked with water balance and water stress (i.e., *LAIgcx*, *Blcond*, *MaxIntcpntn*, *MinASW*) (Fig. 3b), temperature and frost stress (i.e., *Topt*, *kF*) (Fig. 3c), soil parameters (i.e., *klmax*) (Fig. 3b), fertility parameters (i.e., *FR* and *fN0*) (Fig. 3c). Posterior distributions were also different for *alpha*, *tSLA* and *fullCanAge* (Fig. 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 (Fig. 3). The data were not informative for some parameters related to temperature stress (i.e., *Tmax*) (Fig. 3c), water stress (i.e., *SWpower*, *Blcond*, *LAImaxIntcptn*) (Fig. 3b), age effect on forest growth (i.e., *MaxAge*, *rAge*, *nAge*) (Fig. 3b), litterfall parameters (i.e., *gammaF0*, *tgammaF*) (Fig. 3a), stand volume (i.e., *fracBB1*, *tBB*) (Fig. 3b), stand attributes (i.e., the initial root biomass (*WR_i*) and tree density at plantation (*StemNo*)) (Fig. 3c). 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*) (Fig. 3b). The data were extremely informative for the allocation and allometric parameters (i.e., *pFS2*, *pFS20*, *StCn*, *StPw*, *pRx*, *pRn*, *aH* and *cD*) (Fig. 3a), temperature parameters (i.e., *Tmin*, *Topt*) (Fig. 3c), fertility parameters (i.e., *m0*, *FR*) (Fig. 3c), the litterfall rate at maturity (*gammaFx*) (Fig. 3a), water stress parameters (i.e., *MaxCond*, *CoeffCond*) (Fig. 3b), light use efficiency and light interception parameters (i.e., *alpha*, *k*, *SLA0*) (Fig. 3a), the age at which canopy close (*fullCanAge*) (Fig. 3a) and initial biomass (i.e., *WF_i*, *WS_i*) (Fig. 3c).

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 (Fig. 3a).

3.3. Analysis of model-data mismatch

For each output, *MSE* were calculated using the mode of the prior and posterior distributions (Fig. 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 normalised 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*.

3.4. Bayesian model comparison

Results regarding of 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.

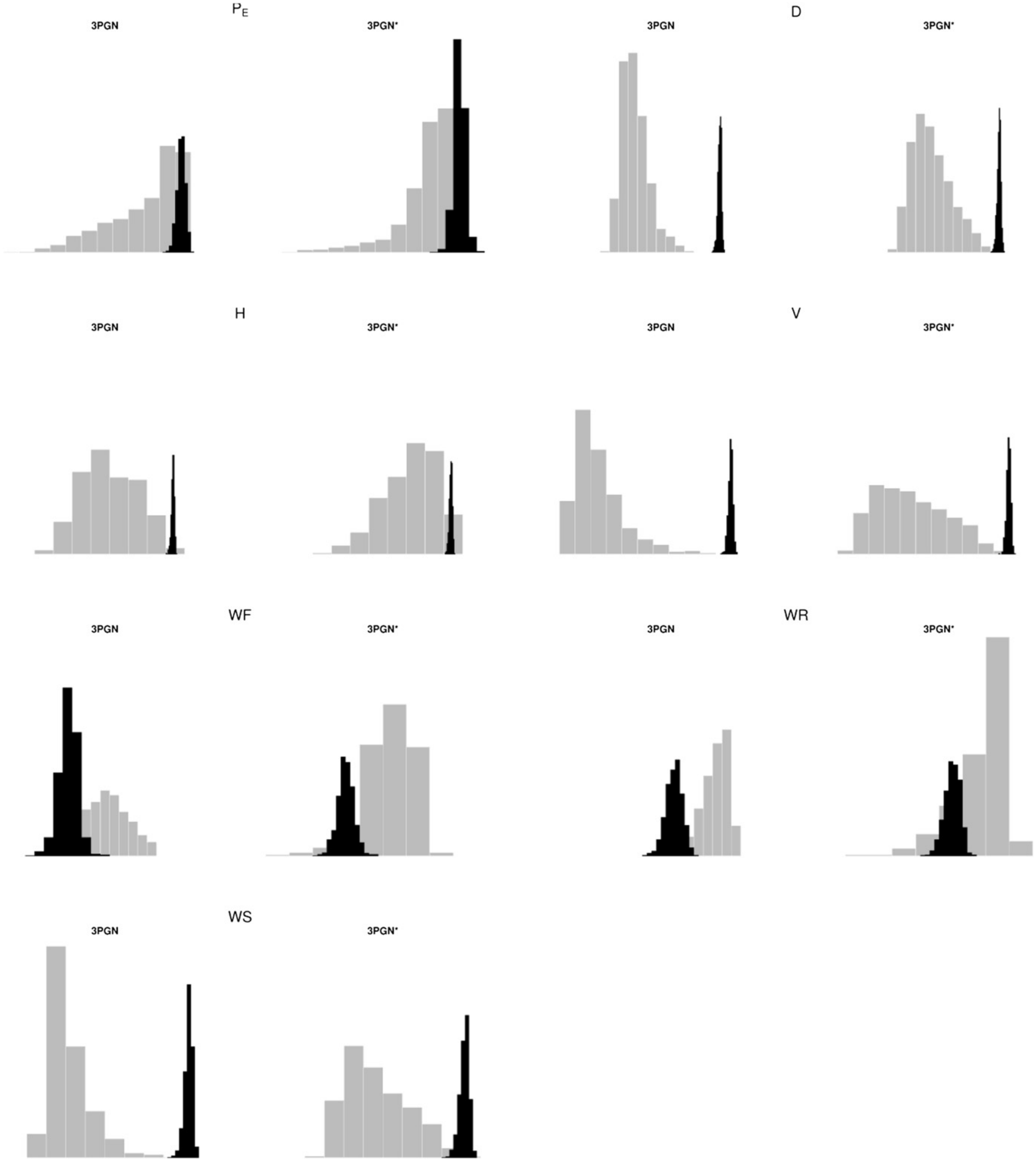


Fig. 2. Prior (grey histograms) and posterior (black histograms) distributions of log-likelihoods for the two model versions, for the seven categories of output variables.

4. 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.,

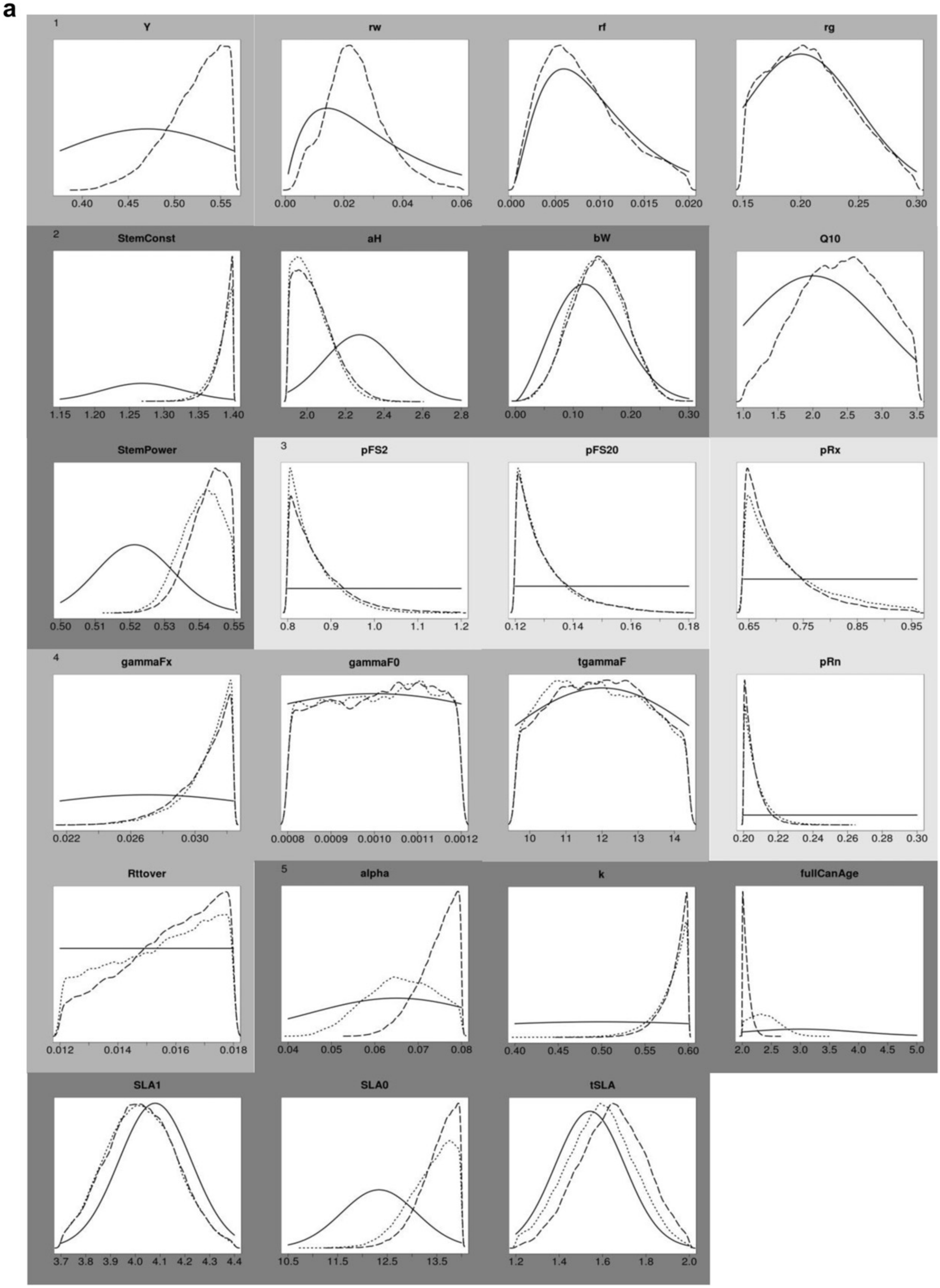


Fig. 3. a. Marginal prior distributions (continuous line) and marginal posterior distributions of 3PGN (dashed line) and 3PGN* (dotted line). Parameters are grouped as respiration parameters (group 1), allometric parameters (group 2), allocation parameters (group 3), turnover parameters (group 4), light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

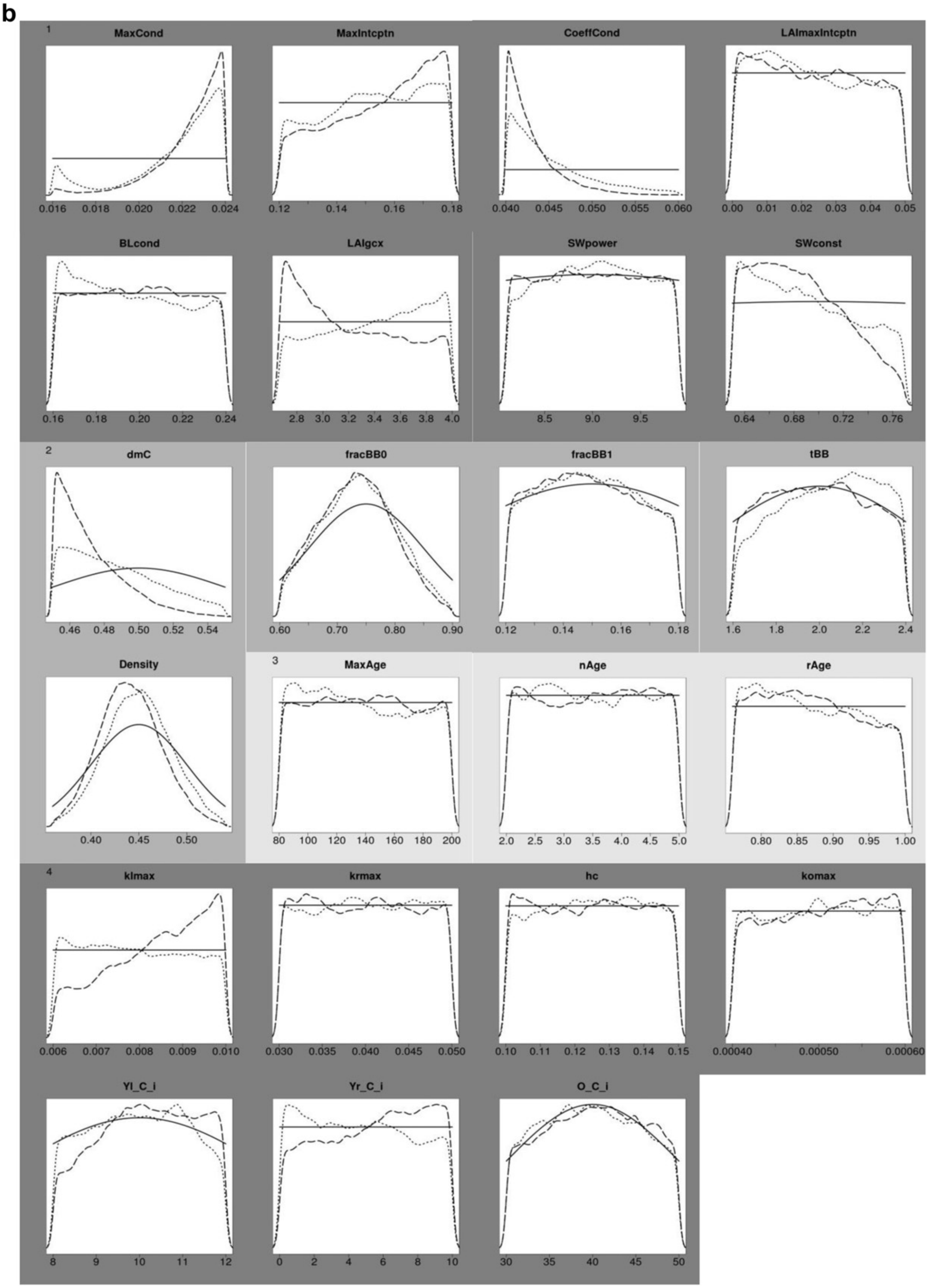


Fig. 3. (*continued*).

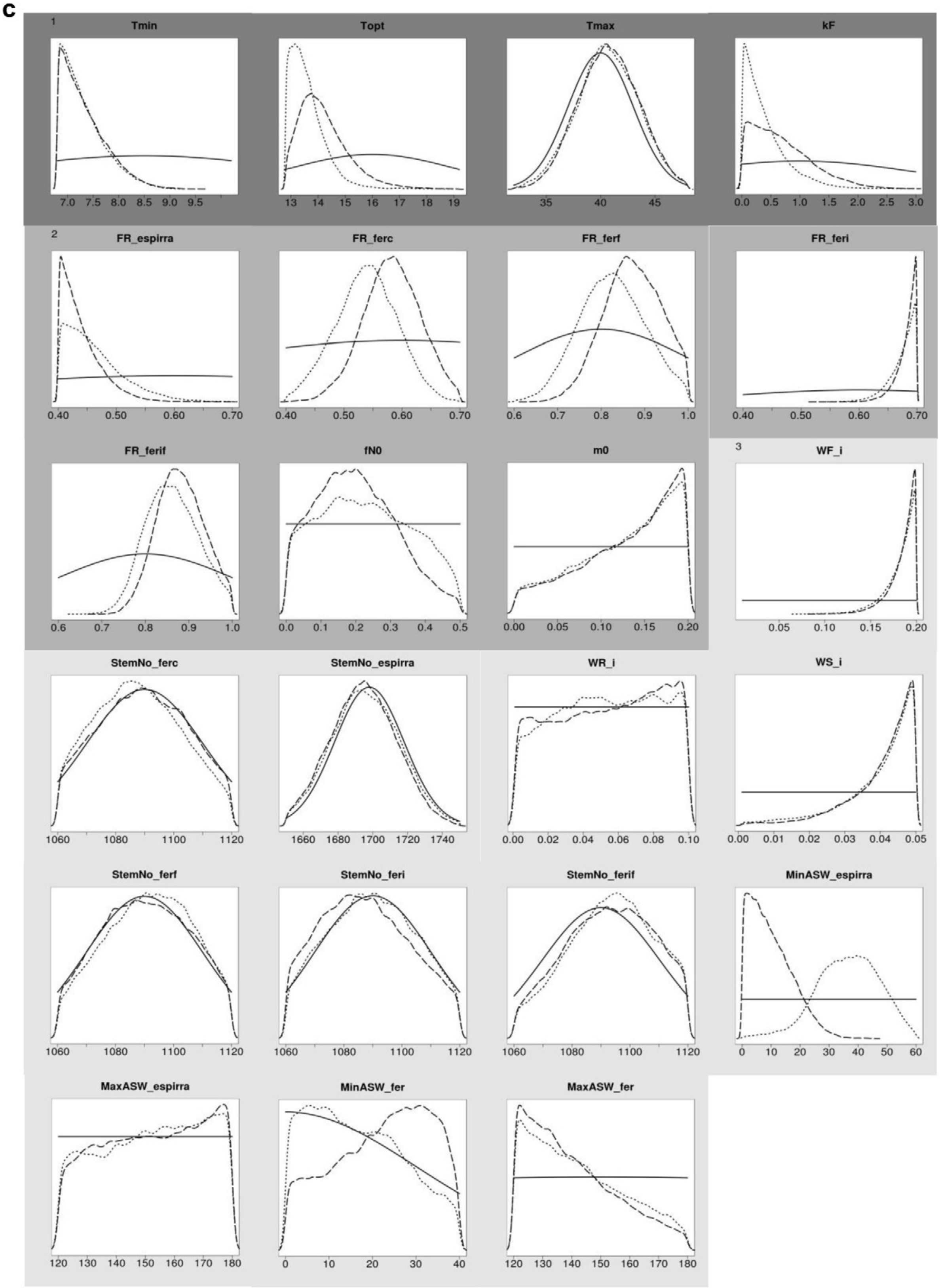


Fig. 3. (*continued*).

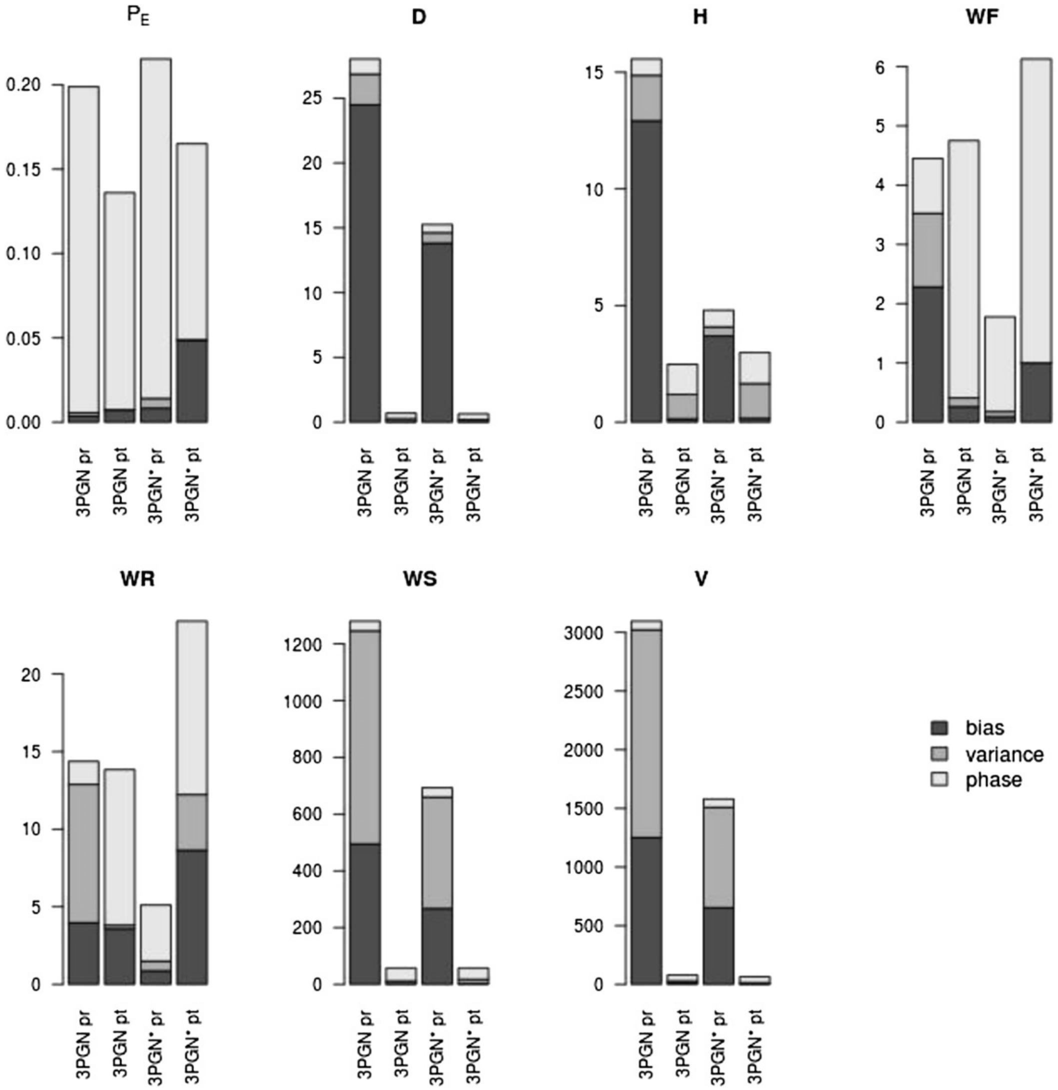


Fig. 4. Decomposition of the mean squared error associated with the modes of the prior (pr) and posterior (pt) parameter distributions, for 3PGN and 3PGN*. In squared brackets is reported the number of data for each variable.

2006) and the differential evolution Markov chain (ter Braak and Vrugt, 2008).

4.1. Uncertainty and sensitivity

Uncertainty and sensitivity analyses are fundamental processes that help to understand model behaviour. Previous works already performed sensitivity analyses of 3-PG and 3PGN, using the simplest method of the One Factor At a Time (OAT) (Almeida et al., 2004; Esprey et al., 2004; Xenakis et al., 2008) and global sensitivity analysis techniques (Song et al., 2012). 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. The minimum and the maximum values of the prior are ranges within which the parameters are meaningful, i.e. they have a biophysical sense. 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 modellers on the parameter space that is meaningful and supported by previous evidence. For instance, previous works (Almeida et al., 2004; Esprey et al., 2004; Xenakis et al., 2008; Song et al., 2012) showed that the allometric parameters of the biomass vs. stem diameter at breast height relationship (i.e., *StemConst*, *StemPower*) are between the most influent parameters on model output variables. Our analysis showed that those parameters are important, but they are not within the five most important parameters for each of the outputs considered. This is because the parameter space of *StemConst* and *StemPower* was reduced after the Bayesian calibration of eq. (1) with a high amount data (i.e., about 700 data of *D* and aboveground biomass were used for the BC of eq. (1)). We also found that the minimum available soil water is a key factor that has a strong impact on model outputs. This is a parameter that represents the availability of water from deep soil and it is

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 normalised 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).

Var.	Statistic	3PGN		3PGN*	
		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**

usually set to zero because of the difficulty in measuring it. However, in Mediterranean climates, if trees can reach deep soil water, stand production could increase significantly (Pereira et al., 2007). Therefore, more attention must be paid to this site variable, especially for sites characterized by high water stress, as we included the parameter in the calibration. The fertility rating parameter (FR) was the factor that generally had the highest impact on model outputs. FR characterizes the fertility of a site and it varies between 0 for the least fertile sites and 1 for sites that do not have nutrient limitations. Many attempts have been carried out to find out how to assign the fertility indices used in 3-PG (Landsberg and Sands, 2010). The relationship between fertility indices and the site chemistry was examined by Stape et al. (2004). Almeida et al. (2010) related FR not only to the chemical site attributes, but also to physical properties that can affect the nutrient uptake by the trees, i.e. water limitation, oxygen limitation, management limitation and topography limitation. Even though the simple approach of characterizing site fertility through a fertility rating is pragmatic, the FR assignment is not straightforward. Particular attention must be paid in the definition of the fertility rating because of the impact that this site parameter has on the model output variables, as shown by the sensitivity analysis results.

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., *fullCanAge*, *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. Xenakis et al. (2008) also carried out uncertainty analysis under a Bayesian framework and a sensitivity of 3PGN. They found that the least uncertain parameters were the most sensitive. Our results also showed that uncertainty was strongly reduced for parameters to which the model was most sensitive (i.e., fertility parameters, *alpha*, *Y*, r_w, allometric and allocation parameters), while the parameters that remained more uncertain (i.e., age stress parameters, some litterfall parameter, some soil parameter, *Blcond*, *LAImaxIntcptn*, initial tree density) had reduced impact on model outputs. Uncertainty was also reduced in parameters that have a medium-low impact on the output variables (i.e., the root partitioning parameters, *MaxIntcptn*, *kF*), meaning that the data were informative for these parameters. Data availability had strong implications on sensitivity and uncertainty analysis results. Sensitivity results are only related to the output variables considered and the posterior parameter uncertainty depended on the prior information and the data used in the BC. 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, modellers 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.

4.2. 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

(Fig. 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 and 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.

4.2.1. 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 (Tuomi et al., 2008; Correia et al., 2012), the 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 and van Oijen et al. (2013) used *BMC* to compare six models of different complexity using data of Scots pine forests. 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 (Fig. 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).

5. 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. After being Bayesian calibrated, 3PGN and 3PGN* predictions of foliage and root biomass slightly deteriorated, probably because the dataset was small and characterized by high uncertainty. Models performed poorly when used to predict net ecosystem production, 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). We are grateful to Jessica Dionne who had improved the English readability of this manuscript. We also would like to thank João Mateus, Gabriel Pita and Abel Rodriguez for providing the dataset of the Espirra forest and Luis Fontes, Alexandra Correia and Paula Soares for providing the dataset of the Furadouro experiment. Finally we wish to thank four anonymous reviewers for their comments and suggestions, from which this paper has benefited greatly.

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. http://dx.doi.org/10.1016/j.ecolmodel.2005.04.024.

Almeida, A.C., Landsberg, J.J., Sands, P.J., 2004. Parameterisation of 3-PG model for fast-growing *Eucalyptus grandis* plantations. Forest Ecology and Management 193 (1–2), 179–195. http://dx.doi.org/10.1016/j.foreco.2004.01.029.

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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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, vol. 30. Academic Press, pp. 113–175. 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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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. CO_2 efflux, CO_2 concentration and photosynthetic refixation in stems of *Eucalyptus globulus* (Labill. Journal of Experimental Botany 60 (1), 99–105. http://dx.doi.org/10.1093/jxb/ern272.

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 CO_2 efflux in different Mediterranean ecosystems: model calibration using the Bayesian approach. Agriculture, Ecosystems & Environment 161, 88–100. http://dx.doi.org/10.1016/j.agee.2012.07.025.

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. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/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., May 2010. 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. http://dx.doi.org/10.1139/x06-186. NRC Research Press Ottawa, Canada.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7 (4), 457–511. Retrieved from: http://www.jstor.org/pss/2246093. Institute of Mathematical Statistics.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis, second ed. C. & Hall/CRC, 668 pp.

Gifford, R.M., 1994. The global carbon cycle: a viewpoint on the missing sink. Australian Journal of Plant Physiology 21 (1), 1. http://dx.doi.org/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. http://dx.doi.org/10.1071/FP02083.

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

Hartig, F., Dyke, J., Hickler, T., Higgins, S.I., O'Hara, R.B., Scheiter, S., Huth, A., 2012. Connecting dynamic vegetation models to data – an inverse perspective. Journal of Biogeography 39 (12). http://dx.doi.org/10.1111/j.1365-2699.2012.02745.x.

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.

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. http://dx.doi.org/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. http://dx.doi.org/10.1016/S0304-3800(00)00420-8.

Kobayashi, K., Salam, M.U., 2000. Comparing simulated and measured values using mean squared deviation and its components. Agronomy Journal 92, 345–352.

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

Landsberg, J.J., Sands, P., 2010. Physiological ecology of forest production. In: Principles, Processes and Models (Terrestrial Ecology), vol. 4. Academic Press. Retrieved from: http://www.amazon.com/Physiological-Ecology-Forest-Production-Volume/dp/0123744601, 352 pp.

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. http://dx.doi.org/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. http://dx.doi.org/10.1046/j.1365-2486.2000.00291.x.

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. http://dx.doi.org/10.1007/s004680050224. Springer, Berlin/Heidelberg.

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 temperature. Crop Science 14 (4), 509. http://dx.doi.org/10.2135/cropsci1974.0011183X001400040005x. Crop Science Society of America.

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. http://dx.doi.org/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.

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. http://dx.doi.org/10.1126/science.263.5147.641.

Penning de Vries, F.W.T., 1974. Substrate utilization and respiration in relation to growth and maintenance in higher plants. Netherlands Journal of Agricultural Science 22, 40–44.

Penning de Vries, F.W.T., 1975. Use of assimilates in higher plants. In: Photosynthesis and Productivity in Different Environments. Cambridge University Press, pp. 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. http://dx.doi.org/10.1111/j.1365-2486.2005.001002.x.

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

Ryan, M.G., 1991. Effects of climate change on plant respiration. Ecological Applications 1 (2), 157. http://dx.doi.org/10.2307/1941808.

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.

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 Physiology 16 (3), 333–343. http://dx.doi.org/10.1093/treephys/16.3.333.

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

Sivia, D.S., 2006. Data Analysis: a Bayesian Tutorial, second ed. O. U. Press, 260 pp.

Song, X., Bryan, B.A., Paul, K.I., Zhao, G., 2012. Variance-based sensitivity analysis of a forest growth model. Ecological Modelling 247, 135–143. http://dx.doi.org/10.1016/j.ecolmodel.2012.08.005.

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. http://dx.doi.org/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. http://dx.doi.org/10.1007/s11222-008-9104-9.

Tuomi, M., Vanhala, P., Karhu, K., Fritze, H., Liski, J., 2008. Heterotrophic soil respiration—comparison of different models describing its temperature dependence. Ecological Modelling 211 (1–2), 182–190. http://dx.doi.org/10.1016/j.ecolmodel.2007.09.003.

Titus, B., Malcolm, D.C., 1999. The long-term decomposition of Sitka spruce needles in brash. Forestry 72 (3), 207–221. http://dx.doi.org/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. http://dx.doi.org/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 Physiology 25 (7), 915–927. http://dx.doi.org/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. http://dx.doi.org/10.1093/aob/mcq039.

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. http://dx.doi.org/10.1016/j.foreco.2012.09.043.

Varella, H., Guérif, M., Buis, S., 2010. Global sensitivity analysis measures the quality of parameter estimation: the case of soil parameters and a crop model. Environmental Modelling & Software 25 (3), 310–319. http://dx.doi.org/10.1016/j.envsoft.2009.09.012.

Vezzaro, L., Mikkelsen, P.S., 2012. Application of global sensitivity analysis and uncertainty quantification in dynamic modelling of micropollutants in stormwater runoff. Environmental Modelling & Software 27–28, 40–51. http://dx.doi.org/10.1016/j.envsoft.2011.09.012.

Wang, Y.P., Jarvis, P.G., 1990. Description and validation of an array model — MAESTRO. Agricultural and Forest Meteorology 51 (3–4), 257–280. http://dx.doi.org/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 Physiology 18 (2), 129–134. http://dx.doi.org/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. http://dx.doi.org/10.1016/j.ecolmodel.2008.07.020.

CHAPTER 3

“Soil water availability strongly modulates soil CO2 efflux in different Mediterranean ecosystems: Model calibration using the Bayesian approach”

Agriculture, Ecosystems and Environment 161 (2012) 88–100

Contents lists available at SciVerse ScienceDirect

Agriculture, Ecosystems and Environment

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

Soil water availability strongly modulates soil CO_2 efflux in different Mediterranean ecosystems: Model calibration using the Bayesian approach

A.C. Correia[a,*], F. Minunno[a], M.C. Caldeira[a], J. Banza[a], J. Mateus[b], M. Carneiro[a], L. Wingate[c], A. Shvaleva[d], A. Ramos[a], M. Jongen[a], M.N. Bugalho[e], C. Nogueira[a], X. Lecomte[a], J.S. Pereira[a]

[a] Centro de Estudos Florestais, Instituto Superior de Agronomia, Universidade Técnica de Lisboa, Tapada da Ajuda, 1349-017 Lisbon, Portugal
[b] Departamento de Engenharia Mecânica – Secção de Ambiente e Energia, Instituto Superior Técnico, Universidade Técnica de Lisboa, Av. Rovisco Pais, 1049-001 Lisbon, Portugal
[c] INRA UR1263, EPHYSE, F-33140, Villenave d'Ornon, France
[d] Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, P-2780-157 Oeiras, Portugal
[e] Centro de Ecologia Aplicada Prof. Baeta Neves, Instituto Superior Agronomia, Universidade Técnica de Lisboa, Tapada da Ajuda, 1349-017 Lisbon, Portugal

ARTICLE INFO

Article history:
Received 8 February 2012
Received in revised form 23 July 2012
Accepted 24 July 2012

Keywords:
Soil CO_2 efflux
Soil respiration
Mediterranean
Bayesian calibration
Empirical model
Soil moisture
Soil temperature

ABSTRACT

Soil respiration in drought prone regions is highly dependent on the precipitation regime and soil moisture conditions, which are expected to change in a global warming context. In the present study we used an extensive collection of field chamber measurements of soil respiration (R_s) from forest and grassland sites of centre and south of Portugal distributed over a 10 year period. This data were summarized and analysed with the objective to describe seasonal variability of R_s as affected by soil moisture (H_s) and soil temperature (T_s). A Bayesian framework was used to test the effectiveness of soil bioclimatic models in estimating R_s on a daily and monthly time step. R_s seasonality was similar between sites, reaching a maximum in spring and autumn and a minimum in the dry season (July–September). No differences were observed for R_s between sites with different standing biomass or soil carbon stocks either on an annual or seasonal timescale. H_s, and not T_s, was the driving factor of R_s during most of the year. T_s drove R_s response only above certain H_s limits: 10% for forest sites and 15% for grassland sites leading to a Q_{10} of 2.01, 1.61 and 1.31 for closed forests, open forests and grasslands, respectively. The Bayesian analysis showed that models using H_s as an independent variable performed better than models driven by T_s alone. Monthly estimates of R_s in grasslands can be predicted by simple climatic models based on H_s but none of them was suitable for forest ecosystems, stressing the need for a process-based approach. This study adds to the evidence that H_s controls R_s fluxes for Mediterranean ecosystems and should always be taken into account for extrapolation purposes.

1. Introduction

Soil respiration, R_s, is one of the most important fluxes in the terrestrial carbon (C) cycle. In 2008, global R_s was estimated at 98 ± 12 Pg C (Bond-Lamberty and Thomson, 2010), with an increase of 0.1 Pg C year^{-1} between 1989 and 2008, attributed to the increase in global air temperature recorded in the last decades.

Mediterranean regions in particular are presently facing an increase in the frequency and severity of droughts (Ramos et al., 2011), with potentially negative effects on the productivity of agricultural crops, pastures and forests. In such semi-arid ecosystems, plant productivity is largely controlled by climatic drivers, with water availability being among the most important. Such drought-induced decreases in primary productivity can also affect the total CO_2 efflux from soils impacting litter fall, decomposition and belowground allocation of photosynthates to roots and root turnover (Reichstein et al., 2002).

Droughts in Mediterranean systems occur during periods of high temperature and low soil moisture content. These environmental conditions also tend to reduce the metabolic activity of living tree roots, mycorrhizae and other rhizosphere microorganisms and soil fauna leading to reductions in CO_2 losses from soils (Hanson et al., 2000). Other physical and chemical processes, like soil CO_2 degassing and transport through the soil profile and the chemical oxidation of soil minerals, are considered minor contributors to the net CO_2 efflux (Buchmann, 2000).

Soil temperature is one of the most influential and widely studied factor affecting R_s (Lloyd and Taylor, 1994; Kirschbaum, 1995; Yuste et al., 2004; Zhou et al., 2009). In general, R_s increases with T_s, and its relationship is usually described with exponential equations (see Webster et al., 2009 for examples). The temperature

---

* Corresponding author. Tel.: +351 213653345; fax: +351 213653338.
E-mail address: alexandrac@isa.utl.pt (A.C. Correia).

0167-8809/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.agee.2012.07.025

dependence of R_s, that is the factor by which a variable changes for a temperature increase of 10 °C (Q_{10}), ranges from 1.3 to 5.6 (Lloyd and Taylor, 1994), although Mahecha et al. (2010) suggest that this factor is confined to values around 1.4 ± 0.1. This Q_{10} factor, however, is only valid for a limited range of temperatures and under drought conditions temperature plays a secondary role in explaining R_s (Kirschbaum, 1995; Janssens and Pilegaard, 2003; Lenton and Huntingford, 2003; Yuste et al., 2004; Almagro et al., 2009).

In ecosystems that are prone to drought, T_s and H_s often interact to control R_s, with R_s responding to the most limiting factor. In summer, R_s significantly decreases with water stress because of a decline in plant metabolic activity and an inhibition of soil microbiological carbon decomposition (Jarvis et al., 2007). When water is not limiting, R_s generally increases with temperature (Raich and Schlesinger, 1992; Inglima et al., 2009). This partially explains why attempting to model R_s using only temperature driven variables proved ineffective in drought-prone regions (Joffre et al., 2003; Xu and Baldocchi, 2004; Almagro et al., 2009; Migliavacca et al., 2011).

While some authors argue that purely climatic driven models are sufficient for accurately predicting R_s (Raich and Schlesinger, 1992; Raich et al., 2002), others emphasise the need to separate climatic and biological effects on R_s (Reichstein et al., 2002, 2003). Although there have been many attempts to incorporate biotic variables in models, e.g. using substrate quantity and quality (Webster et al., 2009), leaf area index (Reichstein et al., 2003) and productivity (Irvine et al., 2005), climatic driven models are still widely used with success (Rey et al., 2002; Conant et al., 2004; Tang et al., 2005; Almagro et al., 2009; Webster et al., 2009; Zhou et al., 2009). This is in part because climatic driven models are simpler to use and less input demanding. Nonetheless, fitting empirical models with proxy parameters of biological activity can yield very good results when enough information is available (Migliavacca et al., 2011).

The majority of Rs data is collected using portable field chambers. One fundamental problem of ecological model application is scaling-up results from these small chambers to the stand or ecosystem level. Sporadic measurements of Rs have obvious limitations as they may not accurately capture the temporal and spatial variability or the dynamic nature of the C release from roots and microbes. In this sense models are important as they allow the estimation of these processes at larger scales. Models need to be calibrated and validated in order to apply them robustly, and one powerful tool to evaluate model applicability and performance is using Bayesian statistics. This approach allows the calibration and the quantification of uncertainty for models that contain multiple parameters (Kass and Raftery, 1995; van Oijen et al., 2011; Sivia and Skilling, 2006).

The objectives of this study were: (1) to assess the seasonal variability of R_s among vegetation types; (2) to quantify the effects of T_s and H_s on R_s in different Mediterranean ecosystems; (3) to evaluate, under a Bayesian framework proposed by van Oijen et al. (2011) and improved by Minunno et al. (submitted for publication), the performances of climate driven models in predicting R_s. We used a long-term database spanning 10 years of R_s measurements, covering grasslands and open and closed forests, using the same field-portable device, a close dynamic system EGM-1 from PP Systems (PP System, Amesbury, USA). This allowed intersite comparisons reducing systematic errors associated with the equipment (which avoided the need for cross calibration) and measurement technique (which was the same for all sites). This work is innovative because it applies a Bayesian approach for model comparison and for global sensitivity analysis to R_s models in Mediterranean regions, which had only been applied by Tuomi et al. (2008) and Hashimoto et al. (2011) in boreal and temperate regions.

2. Materials and methods

2.1. Site description

The experimental sites are located in the centre and south of Portugal. All sites are characterized by a Mediterranean climate with more than 80% of total annual precipitation occurring between October and April. Average annual precipitation (long-term averages from 1971 to 2000) is 608 mm with a mean annual temperature of 15.9 °C. Soils have sandy silt or sandy clay textures (see Table 1 for other site characteristics).

The eight sites have different characteristics that were grouped accordingly to the type of vegetation cover: closed forests (CF), open forests (OF) and Mediterranean grasslands (MG). The CF vegetation type includes a eucalyptus (*Eucalyptus globulus*) stand (CF1) with 1110 trees ha^{-1} and a young and dense evergreen cork oak (*Quercus suber*) woodland site (CF2) with 424 trees ha^{-1}. A pine (*Pinus pinea*) forest (OF1) with 120 trees ha^{-1} and two evergreen oak (*Q. suber* and *Quercus ilex*) sites with 55 (OF3) and 180 trees ha^{-1} (OF2) were included in the open forest cover type (OF). The Mediterranean grasslands (MG) included a natural pasture dominated by winter-spring C3 annuals (MG3), an improved pasture, seeded with a legume-rich seed mixture (MG2) and a C3 grassland with scattered shrub cover of *Cistus* sp. (MG1).

2.2. Soil respiration measurements

Soil respiration (R_s) was measured at each site for a minimum period of 6 months covering the growing season, at weekly or fortnightly intervals. In each site, R_s was measured with the same device, a soil respiration chamber (EGM-1, PP Systems, Hitchin, UK), that uses an infrared gas analyser to measure the rate of CO_2 accumulation inside a closed chamber system. The chamber was inserted in PVC collars that were buried 5 cm deep in the soil several days before the first measurement and left in place throughout the experimental measurement period. Measurements were carried out between 8.30 am and 8.00 pm. The number of collars used varied depending on the heterogeneity of the site, with a minimum of 2 replicates per plot with at least 3 plots per site. A total of 5933 measurements were collected. Spatial replicates were averaged and then aggregated to one value per day resulting in a total of 212 days.

R_s readings were taken every 8 s during measurement periods lasting 128 s. A linear function was then fitted to increasing CO_2 concentration and time variables. The rate of soil respiration ($\mu\text{mol m}^{-2}\text{ s}^{-1}$) was computed as the slope of the linear regression multiplied by a temperature calibration function (Tagesson, 2006). Data were excluded from analysis when linearity was not observed (when $r^2 < 0.9$). We assumed constant soil diffusivity at 0–30 cm.

2.3. Soil temperature measurements and Q_{10} calculation

Soil temperature (T_s) was recorded with a soil thermometer inserted in the mineral soil layer at 10 cm depth at the time of R_s measurement (in MG, CF2, OF2 and OF3 sites). In the remaining sites (CF1 and OF1), T_s was estimated using air temperature measured with the thermocouple inside the R_s chamber, subsequently applying correction functions derived from a regression model. To build this regression model we used half-hourly measurements of T_s and air temperature (EC5, Decagon Devices, Inc., Pullman, USA) from the Coruche site to build the regression model with a total of 23,497 records. The model used T_s at 10 cm depth as a dependent variable and air temperature as an independent variable. The regression analyses led to a multiple regression model with all parameters significantly different from zero ($P < 0.001$) at

Table 1
Sites description with site (ID), geographic coordinates (Lat/Long), location, dominant species (overstory/understory), years when soil respiration was measured (years), approximate tree age (age), tree density, average leaf area index of the overstory component (one side leaf area/total area of the plot), maximum aboveground biomass (AGB) in the herbaceous layer (understory max AGB), aboveground biomass in the trees (overstory AGB), leaf litterfall from the trees and C/N ratios at 0–10 cm soil depth.

ID	Lat/Long	Location	Overstory/understory	Year	Age	Tree density (trees ha^{-1})	LAI (m^2 m^{-2})	Understory max AGB (kg DM m^{-2})	Overstory AGB (kg DM m^{-2})	Tree litterfall (kg DM m^{-2} year^{-1})	C/N in soil
CF1	38°38′N/8°36′W	Pegões	*Eucalyptus globulus*/C3 grasslands	2002; 2006	9	1110	2.6	0.075	13.9	0.212	13.2
CF2	38°31′N/8°01′W	Mitra	*Quercus suber*/C3 grasses	2008–2009	22	430	3.8	0.057	–	0.146	14.5
OF1	38°28′N/8°38′W	Monte Novo	*Pinus pinea*/C3 grasses	2005–2006	45	120	3.2	0.066	12.4	0.271	–
OF2	39°08′N/8°19′W	Coruche	*Quercus suber*/C3 grasses and legumes	2008–2009	~60	177	3.1	0.069	5.5	0.207	20.9
OF3	38°32′N/8°00′W	Alfarrobeira	*Quercus ilex*/C3 grasses and legumes	1999	–	55	1.6	0.179	–	0.096	11.5
MG1	38°47′N/7°25′W	Vila Viçosa	C3 grasslands with disperse Cistus sp. shrubs	2006–2007	–	–	–	0.139	–	–	–
MG2	38°36′N/8°10′W	Montemor	C3 grasslands	2008	–	–	–	0.855	–	–	16.4
MG3	38°57′N/7°12′W	Elvas	C3 grasslands	2003–2005	–	–	–	0.045	–	–	9.2

95% interval confidence and an $R^2_{adj} = 0.67$. We used this model for estimating T_s in the remaining sites admitting neglectful variations in T_s derived from soil type.

The dependency of the respiratory metabolic reactions on temperature was calculated using the Van't Hoff empirical exponential equation (Lloyd and Taylor, 1994):

$R_s = ae^{bT_s}$, where R_s is the soil respiration in μmol m^{-2} s^{-1}, T_s is the soil temperature at 10 cm depth and a and b are the model coefficients. Subsequently, the temperature sensitivity of R_s, the Q_{10}, was calculated on an annual basis for each vegetation type as $Q_{10} = e^{10b}$.

2.4. Soil moisture measurements

Soil moisture (H_s) was measured at a depth of 10 cm adjacent to the soil collars at the time of R_s measurements, using calibrated soil moisture sensors: TDR (Time Domain Reflectometry, Soil Moisture Equipment Corporation, Santa Barbara, USA) was used in CF1, OF2, MG1; Theta-probe (ML2X, Delta-T Devices, Cambridge, UK) in MG3 and PR1 profile probes (Delta-T Devices, Cambridge, UK) were used in all the other sites. Measurements of H_s were reported as volume of water per volume of soil (in %). Whenever gaps in the H_s measurements were observed, the corresponding R_s measurements were excluded from the comparative analysis of H_s with R_s or T_s.

2.5. Plant biomass, litterfall and productivity

Standing biomass per hectare was estimated by measuring all tree diameters and height in 500 m^2 stands at each site. Aboveground biomass components (leaves, trunk and branch) and leaf area index (LAI) were estimated subsequently by using allometric equations specific for each tree species (Zianis et al., 2005; Correia et al., 2010) and converted on an area basis. The LAI in OF3 includes the herbaceous layer and was estimated by a leaf area meter LICOR-2000 (Reichstein et al., 2003). Herbaceous biomass was estimated by harvesting all plant material using the quadrates method (Krebs, 1989). Tree litterfall was estimated monthly or fortnightly with litter traps located randomly in each site during a period of at least one year. The above-mentioned site variables assessment was not necessarily coincident with the time period of R_s measurements.

2.6. Carbon and nitrogen content in soils

Carbon concentration in the soil was determined by either of two methods: the dry combustion method, according to International Organization for Standardization 10694, using a CNS elemental analyser (Leco CNS-2000, Michigan) or by the wet oxidation method (De Leenheer and Van Hove, 1958). Nitrogen concentration was determined by Kjeldahl digestion analysis (Digestion System 40, Kjeltec Auto 1030 Analyser, DEcator, Sweden).

2.7. Statistical analyses

To examine the differences between vegetation cover types, or seasons of the parameters measured, we used analysis of variance (one-way ANOVA). When statistically significant differences were found, differences between group means were identified by post hoc Tukey HSD tests. When ANOVA assumptions were not fulfilled, i.e. no normal distribution of the data and/or homogeneity of variances, non-parametric tests were carried out performing a comparison on ranks, with Dunn's test being used for post hoc pairwise comparisons. The Pearson Product Moment Correlation coefficient was used to display the strength of the association between pairs of variables. Multiple linear regressions were used to fit the T_s and the Q_{10} model. All statistical relationships were considered significant at $P < 0.05$. Statistical analyses were carried out using SigmaStat (SigmaStat for windows V 3, Dundas Software, Germany).

2.8. *Modelling approach*

2.8.1. *Overview*

Simple bioclimatic driven R_s models were selected for the present analysis. The criterion used was their previous application in similar temperate and Mediterranean-like ecosystems. Simple models have the advantage of being user friendly and the link between input variables and output is clear and easy to evaluate. The four bioclimatic models chosen are presented in Table 2. The Bayesian framework proposed by van Oijen et al. (2011) and improved by Minunno et al. (submitted for publication) with a global sensitivity analysis was used to reduce parametric uncertainty, to evaluate the models and to highlight their strengths and weaknesses. The global sensitivity analysis was carried out for the parameters and the inputs of the models, to better understand their behaviour. The models were calibrated and evaluated using a Bayesian approach (i.e. Bayesian calibration and Bayesian model comparison). The data were aggregated at a daily time step and split in two parts; half of the dataset was used for calibration and half for model comparison. The evaluation of the models was performed at a daily as well as monthly time steps.

2.8.2. *Sensitivity analysis (Morris method)*

Sensitivity analyses were carried out for all 4 models, using the Morris method (Morris, 1991). The Morris method is a global sensitivity analysis that evaluates the output sensitivity of a certain model to simultaneous changes in several factors. The method consists of computing basic statistics, i.e. mean (μ) and standard deviation (σ), from the distribution of a number of incremental ratios, so-called elementary effects. For more details, see Campolongo et al. (2007) and Morris (1991). The overall importance of an input factor is given by μ, and σ describes non-linear effects and interactions between factors. Campolongo et al. (2007) enhanced the Morris method by improving the sampling strategy and proposing to calculate the distribution mean 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. The factors considered in the analyses were the model's parameters ($a1$, $a2$, $a3$ and $a4$) and inputs (i.e. T_s and H_s) (Table 2). The parameters ranged between the minimum and maximum values used for the Bayesian calibration (Table 2), while model input variables varied between the maximum and minimum values measured.

2.8.3. *Bayesian framework*

Model performance was calibrated and evaluated using the Bayesian framework proposed by van Oijen et al. (2011), consisting of model calibration, model comparison, and analysis of model-data mismatch. The Bayesian theorem states that the knowledge we have about a certain model, given the data (termed as "the posterior", $P(\theta|D)$) is proportional to the probability of the data given the model (termed as "the likelihood" $P(D|\theta)$), multiplied by the previous knowledge that we had about the model (termed as "the prior" $P(\theta)$). In mathematical terms, the Bayesian theorem is expressed through the formula:

$$P(\theta|D) = cP(D|\theta)P(\theta) \tag{1}$$

where $c = p(D)^{-1}$, with the value c being fixed.

Table 2
Soil respiration (R_s in µmol m^{-2} s^{-1}) models used in this study for the Bayesian calibration. The total annual precipitation P (mm), average annual air temperature T (°C) and land use correspond to site characteristics in which the models were developed. a_i are the parameters of the models. The parameter $a2$ in M1 and the parameter $a1$ in model M2, M3 and M4 are a proxy of the base respiration values, that is soil respiration under standard conditions.

	Description	Coord.	P (mm); T (°C)	Land use	$a1$	$a2$	$a3$	$a4$	Ref
M1	$a1H_s + a2$	41°24′N 1°55′E	750; 14	Coppice of *Quercus cerris*. L with 745 trees ha^{-1}	0.08–0.22	−0.22 to 0.86	–	–	Rey et al. (2002)
M2	$a1\mathrm{e}^{a2T_s}$	41°24′N 1°55′E	750; 14	Coppice of *Quercus cerris*. L with 745 trees ha^{-1}	0.25–0.87	0.07–0.13	–	–	Rey et al. (2002)
M3	$a1\mathrm{e}^{a2\cdot T_s}\,\mathrm{e}^{a3H_s+a4H_s^2}$	36°56″N 10°22″E	559; 16	Oak–grass savanna with 194 trees ha^{-1} in California	0.0599–0.141	0.00958–0.057	13.36–28.94	−60.197 to −19.77	Tang and Baldocchi (2005)
M4	$a1\mathrm{e}^{(-((\ln(H_s/a2)/a3)^2))}\,\mathrm{e}^{(a4*(T_s-10))}$	47°07′N 5°42′E	950; 10	Mixed broadleafed mature forest with 625 trees ha^{-1} in France	2.88–3.04	0.09–0.1	0.29–0.33	1.31–1.51	Vincent et al. (2006)

The likelihood function used in this analyses was proposed by Sivia and Skilling (2006) and it is described by Eqs. (2) and (3):

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

$$R_i = \frac{M_i - D_i}{\sigma_i} \qquad (3)$$

where, M is the output from the model, D is the observed data and i varies between 1 and N (number of data). We used this likelihood function to better detect outliers (Sivia and Skilling, 2006), which are frequent in R_s measurements.

2.8.3.1. Model calibration. For the models *prior*, we used uniform distributions, with the minimum and maximum for H_s and T_s values based on literature reviews. These limits were expanded in order to get a higher variability that could underpin our data set. The models were calibrated using 50% of the data available randomly distributed, while the other 50% was used for model evaluation. For the Bayesian calibration (BC), Markov Chain Monte Carlo (MCMC) simulations were performed, using the Metropolis algorithm (Robert and Casella, 1999). The MCMC method aims to converge the sampling on the region of the parameter space with the highest probability density. A complete description of the Metropolis–Hastings algorithm is given in van Oijen et al. (2011). To assess a parameter's convergence of iterative simulations, the Gelman–Rubin criterion (Gelman and Rubin, 1992) was used.

2.8.3.2. Model comparison. Model evaluation was performed using Bayesian model comparison which is a powerful extension of the Bayesian Calibration, allowing evaluation of the models structure, on the basis of their relative likelihoods (Kass and 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.

$$P(M_k|D) = \frac{P(D|M_k)P(M_k)}{\sum P(D|M)P(M)} \qquad (4)$$

where k varies between 1 and n models. Assuming no initial preferences for either of the models ($P(\text{M1}) = \ldots = P(\text{Mn})$), Eq. (4) becomes:

$$P(M_k|D) = \frac{P(D|M_k)}{\sum P(D|M)} \qquad (5)$$

$P(M|D)$ is the "integrated likelihood" (IL) and is defined over the whole parameter space of M, i.e. $P(D|M) = \int P(D|\theta)P(\theta)d\theta$.

2.8.3.3. Analysis of model-data mismatch. Model-data mismatch was carried out using more classical methods based on the mean squared error (MSE). MSE was decomposed in three components: bias error, variance error and phase-shift error as proposed by Kobayashi and Salam (2000).

$$\text{MSE} = \overline{(M-D)^2} = (\bar{S} - \bar{D})^2 + (\sigma_M - \sigma_D)^2 + 2(\sigma_M\sigma_D)(1-r) \qquad (6)$$

where M refers to model predictions and D to the observed data. Bias error gives the mismatch between simulated versus measured data, variance error describes the ability of the model to capture data variability, phase-shift error indicates if the model is able to reproduce the pattern of data fluctuation. In addition, we calculated the normalized root mean squared error (NRMSE) for each of the 4 models.

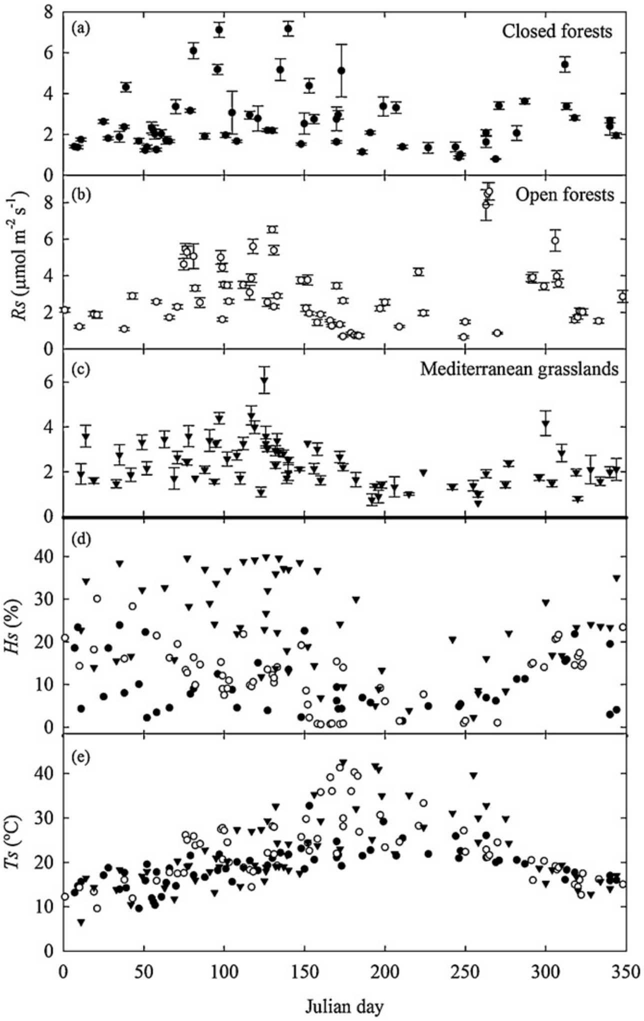


Fig. 1. Temporal dynamics of average daily soil respiration (R_s in μmol m^{-2} s^{-1}) and standard errors in the closed forest sites (a), in the open forest sites (b) and in Mediterranean grasslands sites (c); average soil water content (H_s in %) at 10 cm depth in the sites studied (d); average daily soil temperature (T_s in °C) at 10 cm depth (e) in the sites studied. Symbols: (●) closed forest, (○) open forest, (▼) Mediterranean grassland.

3. Results

3.1. Soil respiration

Seasonal R_s presented two peaks for all vegetation types: one more pronounced in spring and another one in autumn (Fig. 1a–c). Lower values were recorded in summer and in winter but no differences were observed between the three vegetation types neither on an annual ($P = 0.499$) or seasonal basis ($P = 0.222$ in autumn, $P = 0.468$ in spring, $P = 0.285$ in summer and $P = 0.150$ in winter). The highest R_s measurements were recorded in the Alfarrobeira site (OF3) in August after rain pulses with 7.9, 8.5 and 8.6 μmol m^{-2} s^{-1} (Fig. 1b). Measurement variability was highest in spring for all sites (Fig. 1a–c), and was higher in grasslands than in forested sites throughout the year.

3.2. Soil moisture and soil temperature

Similar to the fluctuations observed in R_s, soil moisture exhibited a bimodal pattern at a seasonal level following precipitation (Fig. 1d). Higher values were observed in winter and lower ones in summer. Overall, grasslands showed statistically higher H_s than forested sites. In CF, H_s was particularly low, with many measurements below 6% in spring and summer, and values below 13% in winter.

T_s at 10 cm presented maximum values during the summer months with lower values recorded in winter (Fig. 1e) following the trend in air temperatures. Closed forests presented statistically lower soil temperatures compared with the other vegetation types. Annual soil thermal amplitude decreased with increasing canopy cover and was highest in MG (min 6.7 °C–max 42.6 °C) followed by OF (min 9.6 °C–max 41.4 °C) and CF (min 9.7 °C–max 32.8 °C) (Table 3).

3.3. *Soil respiration explained by soil temperature and soil moisture*

Soil moisture was the driving factor for soil respiration during most of the year (Fig. 2a–c). Soil temperature was a poor and not statistically significant predictor of R_s on an annual timescale for all vegetation types ($P = 0.620$ for CF, $P = 0.011$ for OF and $P = 0.371$ for MG).

The lowest values of R_s were obtained in very dry soils with H_s below 5%. However, a positive trend to increase with H_s was observed in all vegetation types (Fig. 2a–c). On a yearly basis and for OF sites a positive significant correlation was found between R_s and H_s (Fig. 2b). For CF sites, this correlation, although positive, was not significant (Fig. 2a). For MG, H_s alone explained 42% of the variability in R_s (Fig. 2c). T_s drove R_s response but only under certain conditions. That is, whenever soil moisture reached a certain threshold (10% for forest sites and 15% for grasslands) R_s was observed to increase with soil temperature (Fig. 2d–f). In contrast, below these limits H_s was the driving factor of R_s (Fig. 2g–i). The temperature sensitivity of R_s (Q_{10}) was therefore calculated above these limits with values of 2.01, 1.61 and 1.31 for CF, OF and MG, respectively (Fig. 2d–f).

3.4. *Model performances*

3.4.1. *Sensitivity and uncertainty analysis*

Sensitivity analysis shows how model output is affected by changes in factors (model inputs and parameters), helping to better understand model behaviour. Sensitivity results are reported in Fig. 3. Overall, models presented a higher sensitivity to the parameters than the input variables (H_s and T_s) because we used broad parameter ranges. Between the models that depend on both T_s and H_s, M3 is more sensitive to H_s and M4 to T_s (Fig. 3).

Bayesian calibration (BC) allowed reducing the uncertainty in model parameters. The uncertainty in a parameter is expressed by the variance of its distribution (Fig. 4). In this analysis, parameter uncertainty was reduced in all models as shown in Fig. 4 where the posterior marginal distributions for the parameters of each model are reported. An exception is for M4 where *a2* remained highly uncertain for CF and MG and also *a3* just for CF.

3.4.2. *Bayesian model comparison (BMC)*

Regarding daily measurements, the M4 model presented a higher probability of being the best model for forest sites (30% for CF and 100% for OF). For MG, the models with the best probability were M1 (55%) and M3 (45%). On a monthly basis, M1 presented the highest probability of being the best model (40%) for CF, and M4 (41%) for OF. For grasslands, M1 and M3 performed better with 34% and 32% chance of being the best model. This was also confirmed by the low NRMSE values for H_s dependent models in grasslands, especially on a monthly time step (Table 4).

In Fig. 5, observed and estimated values of R_s on a monthly time step are plotted. For each vegetation type only the best model, according to BMC, was presented (M1 for CF and Mediterranean

Table 3

Seasonal average and standard errors of soil respiration rates (R_s in μmol m^{-2} s^{-1}) in each site. Annual average and standard errors and maximum and minimum (Max–Min) values for R_s, soil water content (H_s in %) and soil temperature (T_s in °C) in each site. In brackets the number of observations in each site. (–) means not determined due to inexistent or insufficient data. R_s measurements recorded during rain pulses were excluded. *MG2 was excluded from annual averages as there are only spring measurements.

	Soil respiration (R_s)						Soil moisture (*Hs*)		Soil temperature (*Ts*)	
	Winter	Spring	Summer	Autumn	Min-Max	Year	Min-Max	Average	Min-Max	Average
CF1	1.9 ± 0.2 (10)	2.1 ± 0.2 (9)	2.0 ± 0.4 (5)	2.7 ± 0.2 (9)	1.0–3.6	2.2 ± 0.1 (33)	1.5–23.4	8.7 ± 1.4 (25)	13.9–25.5	19.1 ± 0.5 (33)
CF2	2.3 ± 0.3 (10)	3.6 ± 0.4 (8)	2.1 ± 0.6 (7)	2.7 ± 1 (4)	0.8–5.4	2.7 ± 0.3 (29)	4.3–23.9	11.4 ± 1.4 (20)	9.7–32.8	18.5 ± 1 (32)
Closed Forest	2.1 ± 0.2 (20)	2.8 ± 0.3 (17)	2.1 ± 0.4 (12)	2.7 ± 0.3 (13)	–	2.4 ± 0.2 (62)	–	9.9 ± 1.0 (45)	–	18.8 ± 0.5 (65)
OF1	1.3 ± 0.2 (3)	2.8 ± 0.3 (3)	1.7 ± 0.4 (3)	1.9 ± 0.8 (3)	0.9–3.5	2.0 ± 0.3 (12)	1.0–16.3	9.8 ± 1.9 (11)	14.4–24.6	20.3 ± 1 (12)
OF2	2.4 ± 0.2 (4)	2.8 ± 0.3 (6)	2.3 ± 0.2 (3)	–	1.6–3.8	2.5 ± 0.2 (14)	7.7–30.1	17.1 ± 1.8 (14)	9.6–33.3	21.1 ± 1.9 (14)
OF3	3.5 ± 0.5 (8)	3.2 ± 0.3 (16)	1.2 ± 0.4 (8)	2.9 ± 0.3 (9)	0.7–5.5	2.8 ± 0.2 (41)	0.7–24.6	11.8 ± 1.2 (38)	10.9–41.4	23.7 ± 1.2 (47)
Open forest	2.8 ± 0.4 (15)	3.1 ± 0.2 (25)	1.6 ± 0.3 (14)	2.6 ± 0.3 (13)	–	2.6 ± 0.2 (67)	–	12.6 ± 0.9 (63)	–	22.7 ± 0.9 (73)
MG1	2 ± 0.3 (4)	2.7 ± 0.4 (9)	1.6 ± 0.2 (9)	1.8 ± 0.2 (9)	0.8–4.5	2 ± 0.2 (31)	2.3–32.0	15.8 ± 1.6 (24)	13.2–42.6	25.1 ± 1.6 (31)
MG2	–	2.8 ± 0.2 (12)	–	–	1.6–4.0	–	30.0–39.9	37.1 ± 0.7 (14)	17.6–32.1	20.6 ± 1 (14)
MG3	2.8 ± 0.3 (8)	2.5 ± 0.4 (8)	1.1 ± 0.2 (5)	2.6 ± 0.5 (4)	0.6–4.2	2.3 ± 0.2 (25)	8.6–38.5	23.8 ± 1.8 (23)	6.7–41.0	19.5 ± 1.4 (26)
Med. Grassl.	2.5 ± 0.2 (13)	2.7 ± 0.2 (29)	1.4 ± 0.2 (15)	2.0 ± 0.2 (13)	–	2.2 ± 0.1 (56)*	–	19.7 ± 1.3 (47)*	–	22.5 ± 1.1 (57)*

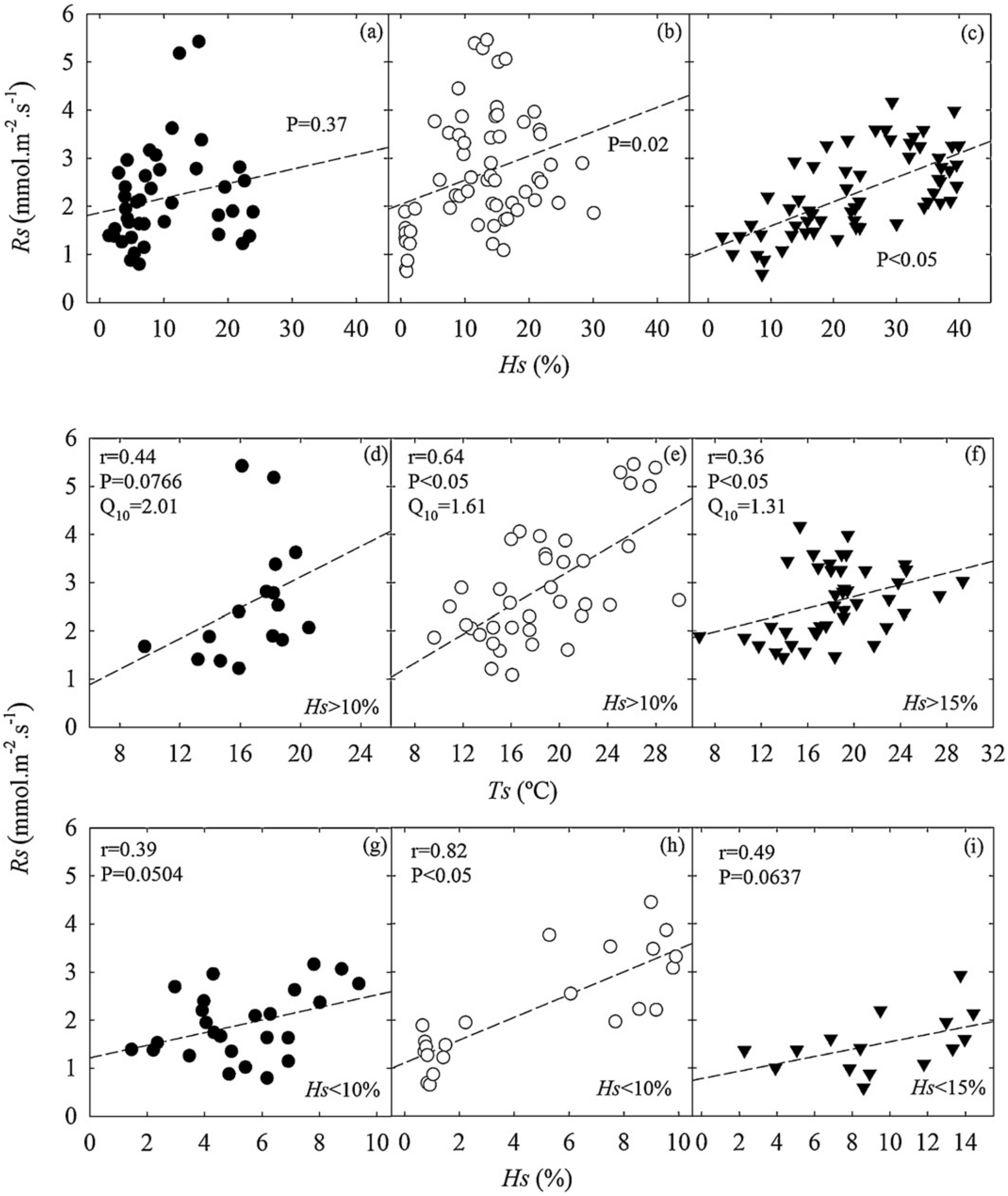


Fig. 2. Annual variation in soil respiration (R_s in μmol m^{-2} s^{-1}) correlated with soil temperature and soil water content (H_s in %) for the 3 vegetation types (a–c). Soil temperature (T_s in °C) drives soil respiration (R_s in μmol m^{-2} s^{-1}) whenever soil water content (H_s in %) is above 10% in forested sites (d and e) and above 15% in Mediterranean grasslands (f). H_s is the driving factor of R_s whenever H_s is below 10% in forested sites (g and h) and below 15% in Mediterranean grasslands (i). Symbols: (●) closed forest, (○) open forest, (▼) Mediterranean grassland.

Table 4
Model results with the Bayesian comparison. *L*% is the percentage probability of each model to be suitable to use in soil respiration estimates on a daily and monthly and time step. NRMSE is the normalized root mean squared error from the Model-data mismatch.

			M1	M2	M3	M4
Closed forests	Daily	*L*%	7.9	16.2	4.1	29.7
		NRMSE	109.8	112.7	109.9	110.2
	Monthly	*L*%	37.5	10.7	11.2	6.4
		NRMSE	125.1	153.1	93.7	120.7
Open forests	Daily	*L*%	0.0	0.0	0.0	100.0
		NRMSE	108.4	108.9	101.9	98.7
	Monthly	*L*%	8.0	5.1	30.9	40.7
		NRMSE	125.1	127.6	95.0	106.2
Mediterranean grasslands	Daily	*L*%	54.8	0.0	45.1	0.0
		NRMSE	83.8	99.7	82.2	84.1
	Monthly	*L*%	34.4	2.7	31.6	23.7
		NRMSE	47.3	83.3	44.7	47.9

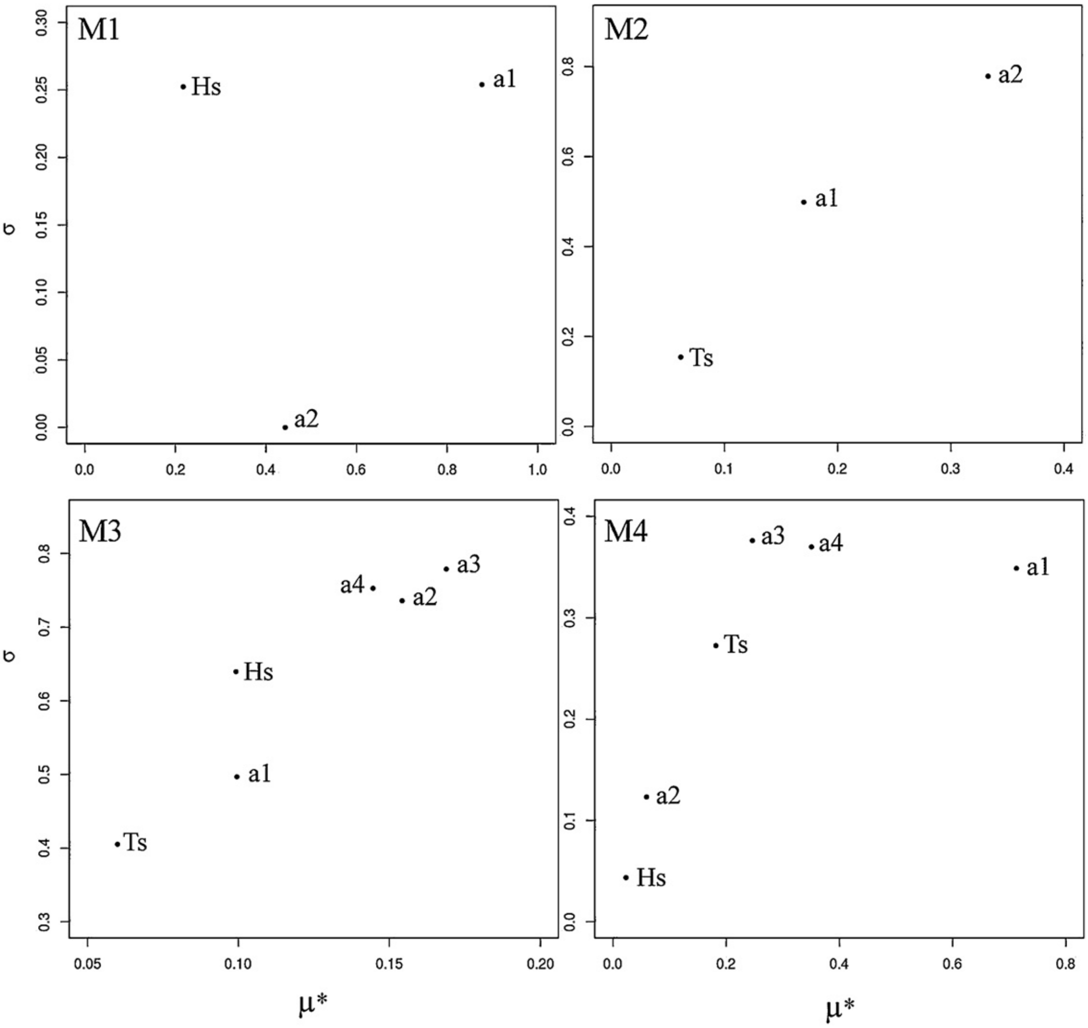


Fig. 3. Sensitivity analysis of the parameters and variables in each of the 4 models studied (M1–M4 in accordance with Table 2) where σ is the standard deviation and μ^* describes the distribution mean of the absolute values of the elementary effects. T_s is the soil temperature (in °C), H_s is the soil moisture (in %). *a1*, *a2*, *a3* and *a4* are the parameters.

grasslands and M4 for OF). The results show that M1 can be used to reliably predict R_s at monthly time steps for Mediterranean grasslands. For the forested sites, none of the models was able to accurately capture the monthly pattern of R_s and therefore they are not suitable for estimations of R_s in forest ecosystems.

3.4.3. Analysis of model-data mismatch

Results from the analysis of model-data mismatch with the decomposition of the mean square error (MSE) led to a better understanding of the models strengths and weaknesses by giving additional information on bias, variance and phase error, which was not provided by the BMC (Fig. 6). At a daily time step, none of the models were able to reliably estimate R_s. Poor performances were obtained also for both forested sites at a monthly time step. On the other hand, for grasslands, the monthly pattern of R_s was accurately predicted by the models dependent on H_s (M1) and/or T_s (M3 and M4).

For forest sites, at a daily time step, no significant differences were observed in the decomposition of the MSE. However, total MSE for OF for models M3 and M4 were slightly lower. The main component of MSE, at a daily time step, is the variance error. This means that there is a great variability in the data on a daily basis. An exception is M3 and M4 for OF, which show a higher phase error (that is a lower pattern correlation between the data).

In MG, at the daily time step, M1, M3 and M4 showed an MSE significantly lower than the other models, with the variance error being the main component for the T_s dependent model (M2), while the phase error is the highest component in M1, M3 and M4.

At a monthly time step, the differences between models performances are significant for all vegetation types. In general, the models that depend on both T_s and H_s (M3 and M4) had lower MSE, with the exception of M4 for CF. In addition, M1 had a low MSE for CF and MG. Overall the MSE was lower at a monthly compared to a daily time step. Bias error (that is the variation related to the average) was a significant component of MSE for forest sites at a monthly time step, contrasting with the very low bias error for MG sites. At a monthly time step, the variance error was very high for T_s dependent models in MG.

4. Discussion

4.1. Rates and seasonality of soil respiration

Soil respiration has two main contributing components: (i) an autotrophic component produced by roots (root carbohydrates and exudates) and rhizosphere and (ii) a heterotrophic component that result from the decomposition of organic materials by soil microorganisms (Epron et al., 1999). The relative contribution of these

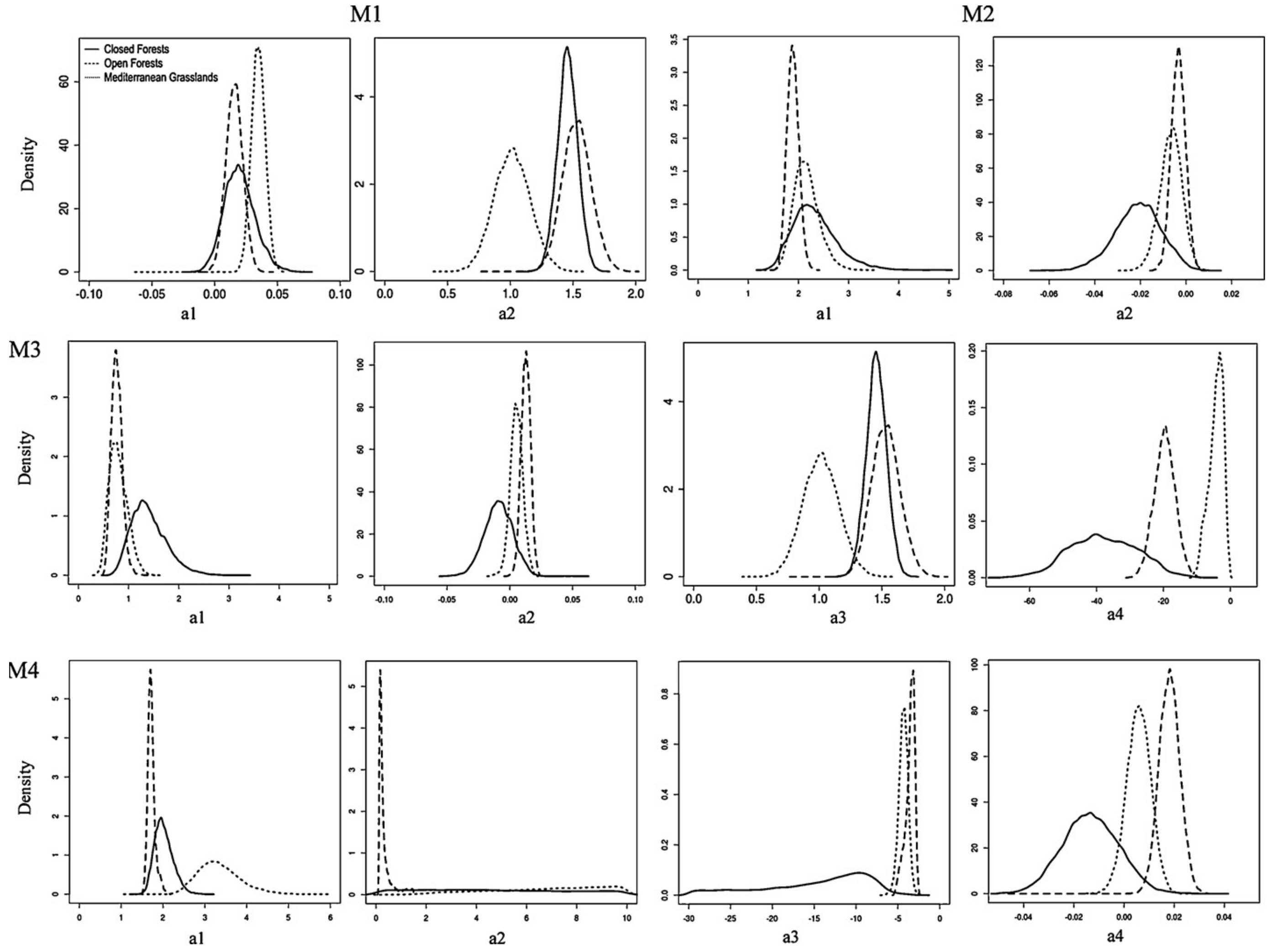


Fig. 4. Bayesian calibration results. Parameter posterior marginal distributions of the four models tested (M1–M4 in accordance with Table 2). Lines correspondence: solid line – closed forests, dashed line – open forests and dotted line – Mediterranean grasslands. Axes represent: *y* – density and *x* – range of parameters values.

components to R_s can differ along the year and between ecosystems.

The seasonality of R_s observed in the study sites followed the plant growth period and both H_s availability and the optimum temperatures for plant growth. Maximum rates also coincide with the peak of maximum photosynthesis in spring for Mediterranean ecosystems (see Fig. 5 from Pereira et al., 2007 for the seasonal variation in gross primary productivity) and also with adequate microclimatic conditions for soil microbial decomposition. For oak dominated sites (CF2, OF2 and OF3), this is also the peak of maximum leaf fall and the onset of herbaceous senescence, both events providing a higher soil organic matter availability for

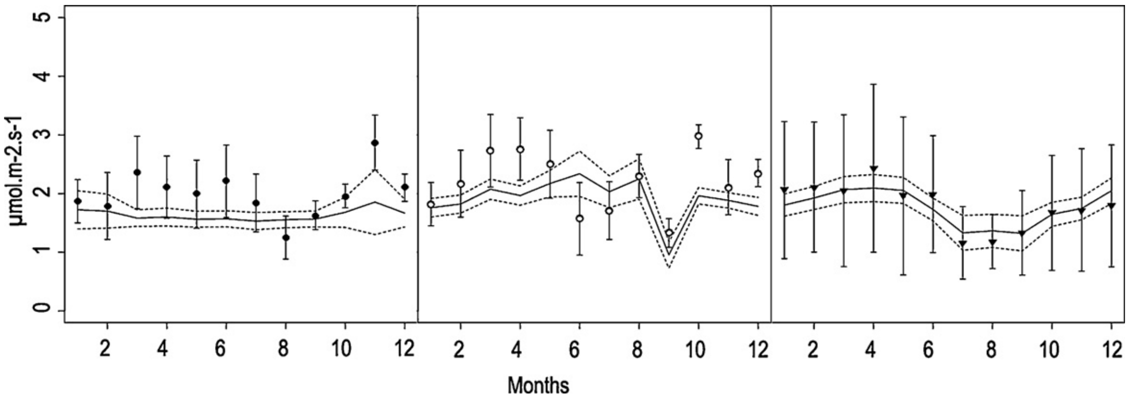


Fig. 5. Monthly observed and estimated R_s (in µmol m^{-2} s^{-1}) for each vegetation type. Estimated values represented by the solid line with upper and lower dash line as the 25 and 75% percentile, using the best model: M1 for closed forests and Mediterranean grasslands and M4 for open forests. Observed monthly averages represented by the symbols: (●) closed forests, (○) open forests, (▼) Mediterranean grassland, with the correspondent standard error in vertical bars.

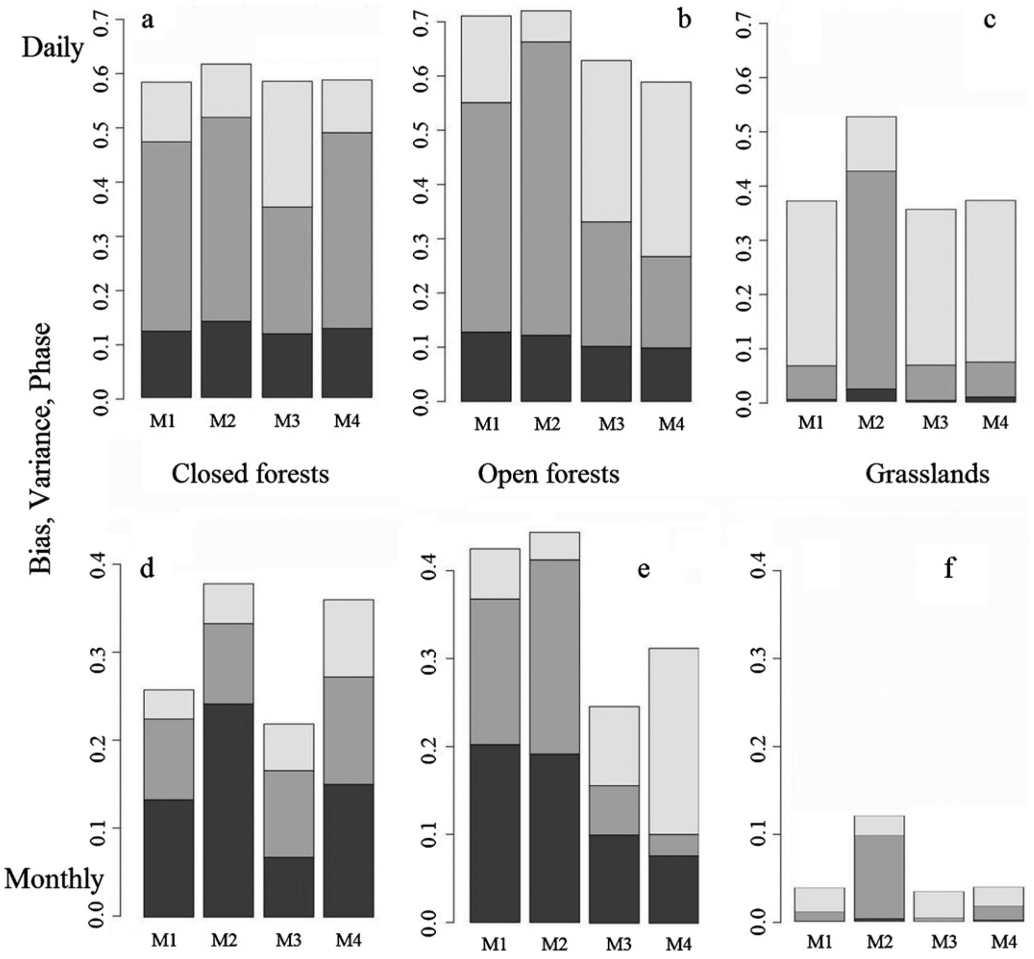


Fig. 6. Means squared error decomposition in the 3 errors: bias (dark grey), variance (medium grey), phase (light grey) and a daily and monthly time step. CF, closed forests; OF, open forests; MG, Mediterranean grasslands.

decomposition (Caritat et al., 2006). In spring, soil water is generally not limiting and higher air temperatures and radiation stimulates photosynthesis and the consequent allocation of photosynthates for new root growth (Hibbard et al., 2005). The combined effect of both autotrophic and heterotrophic respiration may account for the higher spring CO_2 effluxes.

In summer, the decrease in R_s is in accordance with the onset of the dry period followed by tree growth decline and senescence of the vegetation in annual grasslands of Mediterranean climates (Baldocchi et al., 2010). Concomitantly, water stress can limit access to labile carbon substrates and inhibit the activity of microbes in the top soil layers where the substrate for respiration is commonly abundant.

It is likely that the increase of R_s in autumn results more from a stimulation of the soil heterotrophic response and/or release of inorganic carbon from soils than from an autotrophic response, as the rain at the end of the dry season occurs before annual grassland and tree root re-growth (Pereira et al., 2007).

Although we observe a coupling between plant annual life cycle and the seasonally pattern of R_s, we do not have seasonal detailed information regarding plant productivity to provide a quantitative explanation for the similarity in R_s rates in such contrasting vegetation types. For example, in the coppiced eucalypt site (CF1), spring photosynthesis was twice that of the open evergreen woodlands (OF3) and grassland sites (Pereira et al., 2007), but this was not reflected in higher R_s suggesting that root respiration is probably a small fraction of total soil CO_2 efflux. This has been found in other studies in Mediterranean ecosystems. Unger et al. (2010), using stable carbon isotope techniques, found a root contribution to total soil CO_2 efflux of 15–28%, a much lower percentage than the estimates of up to 85% for the heterotrophic component. Others studies report values in the range of 14–29% in a Mediterranean grassland (Gavrichkova et al., 2010) and of 23% for a *Quercus cerris* forest in Italy (Rey et al., 2002).

The low contribution of root respiration to total soil CO_2 efflux may reduce the relevance of autotrophic activity in explaining R_s seasonality. However, this does not explain the effect of productivity and substrate availability related with litter decomposition in the between site variability of R_s. Thus a more detailed description of the seasonal variations in substrate availability, either through changes in productivity or transfers to decomposable pools should be included in order to increase model accuracy.

4.2. *The effect of vegetation cover on soil microclimate*

In our study, the similarity of R_s rates in forests and grasslands, either on an annual or seasonal time step contrasts with the findings of other studies (e.g. Hibbard et al., 2005). The higher R_s in forest ecosystems is explained by higher input of organic matter from tree litterfall, consequently stimulating heterotrophic respiration (Hibbard et al., 2005; Tedeschi et al., 2006), and the contribution of growth and maintenance respiration of tree roots (Jackson et al., 1996; Bond-Lamberty et al., 2004; Hogberg et al., 2008). However, the literature from drought prone regions on this subject is contradictory. Studies on R_s (with at least one year of R_s measurements) comparing contrasting vegetation types in semi-arid

Table 5
Average soil respiration (R_s in μmol m^{-2} s^{-1}) measurements for forests and grasslands collected in studies from the Mediterranean region.

Site	Lat/Long	P (mm)	T (°C)	R_s chamber type	R_s (μmol m^{-2} s^{-1})	Ref
Oak–grass savanna (open areas)	36°56′N 10°22′E	556	16.3	Li-6400	0.5	Tang and Baldocchi (2005)
Degraded Steppe alpha grass	36°50′N 2°15′W	220	18	PP systems	0.8	Rey et al. (2011)
Natural Steppe alpha grass	36°56′N 2°1′W	220	18	PP systems	1.1	Rey et al. (2011)
Olive grove	38°05′N 1°47′W	370	15.5	Li-6400	1.1	Almagro et al. (2009)
Oak–grass savanna (under trees)	36°56′N 10°22′E	559	16.3	Li-6400	1.6	Tang and Baldocchi (2005)
Abandoned field with shrubs	38°05′N 1°47′W	370	15.5	Li-6400	1.7	Almagro et al. (2009)
Forest of Alleppo pine and shrubs	38°05′N 1°47′W	370	15.5	Li-6400	2.1	Almagro et al. (2009)
Holm oak forest with shrubs	41°13′N 0°55′E	658	12	PP systems	2.3	Asensio et al. (2007)
Coppiced oak chronosequence:1 year	42°24′N 11°55′E	755	14	PP systems	3.5	Tedeschi et al. (2006)
Coppiced oak chronosequence: 5 years	42°24′N 11°55′E	755	14	PP systems	4.2	Tedeschi et al. (2006)
Coppiced oak chronosequence: 17 years	42°24′N 11°55′E	755	14	PP systems	6.6	Tedeschi et al. (2006)

ecosystems report similar R_s rates between grasslands and forests (Epron et al., 1999; Tedeschi et al., 2006; Rey et al., 2011) but others not (Almagro et al., 2009) (see annual average values from these studies in Table 5). Taking into account the low contribution of root respiration in explaining within site R_s variability and the fact that there are obvious differences in substrate quality and quantity in each site (Table 1), we stress that the similarity in R_s rates may be explained by plant-mediated effects, creating soil microclimate thresholds for microbial decomposition.

First, the reduced canopy cover in grassland sites and exposure to direct sunlight warms the soil and may induce an increase in microbial decomposition (Lloyd and Taylor, 1994; Katterer et al., 1998; Yuste et al., 2007). Consistently higher T_s and H_s values were found in grasslands as compared with forests. The combined effect of higher T_s and higher H_s could probably contribute to higher values of R_s, because H_s did not limit R_s in grassland. The high leaf-fall, rapid root turnover rates and generally lower C/N in grasslands (Gavrichkova et al., 2010) may also contribute to the R_s enhancement but the data collected in these sites do not allow any statistical conclusion. Secondly, forested sites, especially CF sites, presented consistently lower H_s values (Table 3 and Fig. 1) which may be a consequence of a higher water uptake by tree roots in all soil layers and throughout the year as compared to grasslands. In this case, it is likely that the activity of microorganisms may be somewhat inhibited by the low H_s also.

In conclusion, a number of factors may be concomitantly influencing R_s and only a more detailed site-specific seasonal analysis on R_s related to plant productivity, leaf litterfall and soil substrate availability could help disentangle these responses.

The R_s average values found in this study are within the range of values obtained for other Mediterranean systems (Table 5). Although the data collected in the literature do not allow for any conclusions on the differences between grasslands and forests, a positive and significant correlation between R_s and total average annual site precipitation could be found ($r = 0.64$, $P = 0.007$) which underlines the importance of integrating water related variables in models that aim to predict annual variability of R_s in drought-prone regions.

4.3. Sensitivity of soil respiration to soil temperature and moisture

The temperature sensitivity of R_s (Q_{10}) decreased with tree cover from closed forests (2.01), to open forests (1.61) to grasslands (1.31) and was only significant above a certain H_s threshold: 10% in forests and 15% in grasslands. These soil moisture intervals are similar to the ones obtained at other sites in drought-prone regions (Davidson et al., 2006; Almagro et al., 2009). The Q_{10} values also fall within the range of values reported for other ecosystem types (Raich and Schlesinger, 1992; Kirschbaum, 1995; Jarvis et al., 2007; Almagro et al., 2009). The Q_{10} decrease from forests to grasslands is probably explained by a higher fraction of lignin compounds in forests that decomposes more slowly but is also more sensitive to increasing temperatures (Leifeld and Fuhrer, 2005). It is likely that a co-mixture of plant/soil factors may be affecting the relationship between R_s and T_s (Tedeschi et al., 2006) also at a seasonal level, as soil organic matter availability changes seasonally due to the balance between input of organic matter, stabilisation and mineralisation. From a system dynamic perspective, this would imply a site dependency of the base respiration, that is, soil respiration under standard conditions (see Table 2) with both substrate availability and quality. However, this analysis was not possible to perform because of data limitation. A more detailed approach would be to use for example, the methodology proposed by Mahecha et al. (2010) that specifically addresses the direct responses of R_s to temperature versus long-term organic matter dynamics described by the basal respiration rate.

4.4. Bayesian modelling

A comprehensive analysis of the different models was performed using the Bayesian framework, where the uncertainty in parameter and model output was significantly reduced by the calibration process. The sensitivity and uncertainty analysis provided a better understanding of model behaviour that, together with the BMC and model-data mismatch, allowed evaluating models by highlighting their weaknesses and strengths.

The models behaved differently according to vegetation type and with respect to the time step considered. In a general overview, the climatic-driven models were not suitable for R_s estimates in forested sites, either on a daily or monthly time step. As previously suggested by other authors (Reichstein et al., 2003; Migliavacca et al., 2011), there are a series of non-climatic processes driving soil CO_2 efflux in forests, e.g. plant phenology, microbial growth dynamic and soil physical processes, which cannot be captured in the over-simplistic models tested here.

On the other hand, a good performance was achieved for grassland sites, using H_s driven models on a monthly time step. This is probably caused by the straightforward correlation between grassland productivity and available water in the soil, that consequently affects the soil CO_2 efflux. Both the exploratory analysis and the Bayesian framework agreed on the relevance of H_s (with or without the integration of T_s in the model) in explaining R_s in grasslands. For example, M1 (driven by H_s) and M3 (driven by T_s and H_s) performed better than the other models. Although M4 is also driven by H_s and T_s, it performed worse than M3 because of a higher sensitivity to T_s than H_s.

In general, the model performances were worst at a daily time step, reflecting model limitations in catching the high variability of daily measurements. The improvement in models performance for a monthly time step resulted from the adjustment of the temporal scale, that reduced errors for R_s estimates and led to a smoother seasonal pattern. Other types of errors can be attributed to equipment and different users, although in this study these errors were minimized. A more mechanistic approach is required for daily predictions of R_s for these vegetation types, but supported by background studies, e.g. on the partitioning between soil, plants and microbial components. Although there has been an increasing number of studies on this subject in the last decade (Hanson et al., 2000; Hogberg et al., 2002; Bond-Lamberty et al., 2004; Tang and Baldocchi, 2005; Tang et al., 2005; Zhou et al., 2009; Unger et al., 2010), the processes controlling soil CO_2 efflux in arid and semi-arid regions clearly need more investigation. It must be stressed that although the increase in model complexity would theoretically reproduce biological processes in a more accurate way, it will hamper the calibration processes as it increases the uncertainty associated with the parameters and variables added. Therefore a compromise must be reached between model complexity and parameter uncertainty and the research aims, including different constraint variables, as these may enable a more comprehensive evaluation of model function.

5. Conclusions

In conclusion, this study reinforces the importance of soil moisture in predicting soil respiration in drought-prone regions and should always be taken into account in upcalling exercises. Our study concluded that purely climatic driven models – dependent only on H_s or in conjunction with T_s were able to accurately predict R_s for grassland sites on a monthly time step. In contrast, forest sites presented a higher degree of variability, probably related to plant physiological processes that could not be captured by the over-simplistic models used. The high variability observed in R_s daily measurements for all vegetation types underpins the difficulty in modelling R_s at a higher temporal resolution (daily). The similarity in R_s at seasonal and annual time scales between forests and grasslands is in disagreement with other studies where such comparisons have been previously performed. This may be attributed to the effect of the vegetation on soil bioclimatic factors: forest sites presented consistently lower H_s and lower T_s that may be limiting microorgabial activity, root respiration and the net soil CO_2 efflux. The temperature sensitivity of R_s (Q_{10}) varies between vegetation types, decreasing from CF to OF to MG, stressing the need to use species specific Q_{10} in soil process models. The H_s threshold for Q_{10} also varies between vegetation types: above 10% in forested sites and 15% in grasslands. Further model improvements should be addressed in future research integrating variables or processes aimed at describing the high variability observed in R_s for forest sites.

Acknowledgements

The research was carried out with financial support from the following projects:

O Sequestro de Carbono e a Gestão Florestal Sustentável no Sul de Portugal (AGRO 8.1 I&D-543), CARBOEUROPE-IP (GOCE-CT-2003-505572), MIND (EVK2-CT-2002-00158), NITROEUROPE-IP (FP6-2004-Global3-017841), IMECC (FP6 2007-2011-026188), Effects of deer on productivity and nutrient cycling of a Mediterranean oak woodland (FCT-POCTI 63322/2004). Fundação para a Ciência e Tecnologia funded the following grants: Correia A.C. (SFRH/BD/39058/2007), Shvaleva A. (SFRH/BPD/43643/2008) and Minunno F. (SFRH/BD/65934/2009). We thank also to the following institutions for providing field work facilities: Portucel/Aliança Florestal Company (Herdade da Espirra), Fundação Casa de Bragança, Herdade de Comonte, Herdade da Machoqueira do Grou, Herdade da Mitra. The authors wish to thank the people who collaborated in field work namely Cathy Kurz-Besson, Filipe Costa e Silva, Gabriel Pita, Mário Abrantes, Mafalda Evangelista, Marta Bastos, Marta Lópes Ramos, Pedro Almeida and Raquel Lobo do Vale. We also acknowledge Peter Fay and Stephan Unger for the English review and two anonymous reviewers for the manuscript improvements.

References

Almagro, M., Lopez, J., Querejeta, J.I., Martinez-Mena, M., 2009. Temperature dependence of soil CO_2 efflux is strongly modulated by seasonal patterns of moisture availability in a Mediterranean ecosystem. Soil Biol. Biochem. 41, 594–605.

Asensio, D., Penuelas, J., Ogaya, R., Llusia, J., 2007. Seasonal soil and leaf CO_2 exchange rates in a Mediterranean holm oak forest and their responses to drought conditions. Atmos. Environ. 41, 2447–2455.

Baldocchi, D.D., Ma, S.Y., Rambal, S., Misson, L., Ourcival, J.M., Limousin, J.M., Pereira, J., Papale, D., 2010. On the differential advantages of evergreenness and deciduousness in Mediterranean oak woodlands: a flux perspective. Ecol. Appl. 20, 1583–1597.

Bond-Lamberty, B., Wang, C.K., Gower, S.T., 2004. A global relationship between the heterotrophic and autotrophic components of soil respiration? Global Change Biol. 10, 1756–1766.

Bond-Lamberty, B., Thomson, A., 2010. Temperature-associated increases in the global soil respiration record. Nature 464, 579–582.

Buchmann, N., 2000. Biotic and abiotic factors controlling soil respiration rates in *Picea abies* stands. Soil Biol. Biochem. 32, 1625–1635.

Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. Environ. Model. Softw. 22, 1509–1518.

Caritat, A., Garcia-Berthou, E., Lapena, R., Vilar, L., 2006. Litter production in a *Quercus suber* forest of Montseny (NE Spain) and its relationship to meteorological conditions. Ann. For. Sci. 63, 791–800.

Conant, R.T., Dalla-Betta, P., Klopatek, C.C., Klopatek, J.A., 2004. Controls on soil respiration in semiarid soils. Soil Biol. Biochem. 36, 945–951.

Correia, A.C., Tomé, M., Pacheco, C.A., Faias, S., Dias, A.C., Freire, J., Carvalho, P.O., Pereira, J.S., 2010. Biomass allometry and carbon factors for a Mediterranean pine (*Pinus pinea* L.) in Portugal. Forest Syst. 19, 418–433.

Davidson, E.A., Janssens, I.A., Luo, Y.Q., 2006. On the variability of respiration in terrestrial ecosystems: moving beyond Q_{10}. Global Change Biol. 12, 154–164.

De Leenheer, L., Van Hove, J., 1958. Détermination de la teneur en carbone organique dessols. Etude critique des méthodes titrimétriques. Pedologie 8, 39–77.

Epron, D., Farque, L., Lucot, E., Badot, P.M., 1999. Soil CO_2 efflux in a beech forest: dependence on soil temperature and soil water content. Ann. For. Sci. 56, 221–226.

Gavrichkova, O., Moscatelli, M.C., Kuzyakov, Y., Grego, S., Valentini, R., 2010. Influence of defoliation on CO_2 efflux from soil and microbial activity in a Mediterranean grassland. Agric. Ecosyst. Environ. 136, 87–96.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 7, 457–472.

Hanson, P.J., Edwards, N.T., Garten, C.T., Andrews, J.A., 2000. Separating root and soil microbial contributions to soil respiration: a review of methods and observations. Biogeochemistry 48, 115–146.

Hashimoto, S., Morishita, T., Sakata, T., Ishizuka, S., Kaneko, S., Takahashi, M., 2011. Simple models for soil CO_2, CH_4, and N_2O fluxes calibrated using a Bayesian approach and multi-site data. Ecol. Model. 222, 1283–1292.

Hibbard, K.A., Law, B.E., Reichstein, M., Sulzman, J., 2005. An analysis of soil respiration across northern hemisphere temperate ecosystems. Biogeochemistry 73, 29–70.

Hogberg, P., Nordgren, A., Agren, G.I., 2002. Carbon allocation between tree root growth and root respiration in boreal pine forest. Oecologia 132, 579–581.

Hogberg, P., Hogberg, M.N., Gottlicher, S.G., Betson, N.R., Keel, S.G., Metcalfe, D.B., Campbell, C., Schindlbacher, A., Hurry, V., Lundmark, T., Linder, S., Nasholm, T., 2008. High temporal resolution tracing of photosynthate carbon from the tree canopy to forest soil microorganisms. New Phytol. 177, 220–228.

Inglima, I., Alberti, G., Bertolini, T., Vaccari, F.P., Gioli, B., Miglietta, F., Cotrufo, M.F., Peressotti, A., 2009. Precipitation pulses enhance respiration of Mediterranean ecosystems: the balance between organic and inorganic components of increased soil CO_2 efflux. Global Change Biol. 15, 1289–1301.

Irvine, J., Law, B.E., Kurpius, M.R., 2005. Coupling of canopy gas exchange with root and rhizosphere respiration in a semi-arid forest. Biogeochemistry 73, 271–282.

Jackson, R.B., Canadell, J., Ehleringer, J.R., Mooney, H.A., Sala, O.E., Schulze, E.D., 1996. A global analysis of root distributions for terrestrial biomes. Oecologia 108, 389–411.

Janssens, I.A., Pilegaard, K., 2003. Large seasonal changes in Q_{10} of soil respiration in a beech forest. Global Change Biol. 9, 911–918.

Jarvis, P., Rey, A., Petsikos, C., Wingate, L., Rayment, M., Pereira, J., Banza, J., David, J., Miglietta, F., Borghetti, M., Manca, G., Valentini, R., 2007. Drying and wetting of Mediterranean soils stimulates decomposition and carbon dioxide emission: the Birch effect. Tree Physiol. 27, 929–940.

Joffre, R., Ourcival, J.M., Rambal, S., Rocheteau, A., 2003. The key-role of topsoil moisture on CO_2 efflux from a Mediterranean *Quercus ilex* forest. Ann. For. Sci. 60, 519–526.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.

Katterer, T., Reichstein, M., Andren, O., Lomander, A., 1998. Temperature dependence of organic matter decomposition: a critical review using literature data analyzed with different models. Biol. Fertil. Soils 27, 258–262.

Kirschbaum, M.U.F., 1995. The temperature-dependence of soil organic-matter decomposition, and the effect of global warming on soil organic-C storage. Soil Biol. Biochem. 27, 753–760.

Krebs, C.J., 1989. Ecological Methodology. Harper & Row Publishers, London.

Kobayashi, K., Salam, M.U., 2000. Comparing simulated and measured values using mean squared deviation and its components. Agron. J. 92, 345–352.

Leifeld, J., Fuhrer, J., 2005. The temperature response of CO_2 production from bulk soils and soil fractions is related to soil organic matter quality. Biogeochemistry 75, 433–453.

Lenton, T.M., Huntingford, C., 2003. Global terrestrial carbon storage and uncertainties in its temperature sensitivity examined with a simple model. Global Change Biol. 9, 1333–1352.

Lloyd, J., Taylor, J.A., 1994. On the temperature dependence of soil respiration. Funct. Ecol. 8, 315–323.

Mahecha, M.D., Reichstein, M., Carvalhais, N., Lasslop, G., Lange, H., Seneviratne, S.I., Vargas, R., Ammann, C., Arain, M.A., Cescatti, A., Janssens, I.A., Migliavacca, M., Montagnani, L., Richardson, A.D., 2010. Global convergence in the temperature sensitivity of respiration at ecosystem level. Science 329, 838–840.

Migliavacca, M., Reichstein, M., Richardson, A.D., Colombo, R., Sutton, M.A., Lasslop, G., Tomelleri, E., Wohlfahrt, G., Carvalhais, N., Cescatti, A., Mahecha, M.D., Montagnani, L., Papale, D., Zaehle, S., Arain, A., Arneth, A., Black, T.A., Carrara, A., Dore, S., Gianelle, D., Helfter, C., Hollinger, D., Kutsch, W.L., Lafleur, P.M., Nouvellon, Y., Rebmann, C., da Rocha, H.R., Rodeghiero, M., Roupsard, O., Sebastia, M.T., Seufert, G., Soussana, J.F., van der Molen, M.K., 2011. Semiempirical modeling of abiotic and biotic factors controlling ecosystem respiration across eddy covariance sites. Global Change Biol. 17, 390–409.

Minunno, F., van Oijen, M., Cameron, D.R., Cerasoli, S., Pereira, J.S., Tomé, M. Using a Bayesian framework and a global sensitivity analysis to explore strengths and weaknesses of two process-based models differing in representation of autotrophic respiration. Environ. Model. Softw., submitted for publication.

Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics 33 (2), 161–174.

Pereira, J.S., Mateus, J.A., Aires, L.M., Pita, G., Pio, C., David, J.S., Andrade, V., Banza, J., David, T.S., Paco, T.A., Rodrigues, A., 2007. Net ecosystem carbon exchange in three contrasting Mediterranean ecosystems – the effect of drought. Biogeosciences 4, 791–802.

Raich, J.W., Schlesinger, W.H., 1992. The global carbon-dioxide flux in soil respiration and its relationship to vegetation and climate. Tellus Ser. B: Chem. Phys. Meteorol. 44, 81–99.

Raich, J.W., Potter, C.S., Bhagawati, D., 2002. Interannual variability in global soil respiration, 1980–1994. Global Change Biol. 8, 800–812.

Ramos, A.M., Trigo, R.M., Santo, F.E., 2011. Evolution of extreme temperatures over Portugal: recent changes and future scenarios. Clim. Res. 48, 177–192.

Reichstein, M., Tenhunen, J.D., Roupsard, O., Ourcival, J.M., Rambal, S., Dore, S., Valentini, R., 2002. Ecosystem respiration in two Mediterranean evergreen Holm Oak forests: drought effects and decomposition dynamics. Funct. Ecol. 16, 27–39.

Reichstein, M., Rey, A., Freibauer, A., Tenhunen, J., Valentini, R., Banza, J., Casals, P., Cheng, Y.F., Grunzweig, J.M., Irvine, J., Joffre, R., Law, B.E., Loustau, D., Miglietta, F., Oechel, W., Ourcival, J.M., Pereira, J.S., Peressotti, A., Ponti, F., Qi, Y., Rambal, S., Rayment, M., Romanya, J., Rossi, F., Tedeschi, V., Tirone, G., Xu, M., Yakir, D., 2003. Modeling temporal and large-scale spatial variability of soil respiration from soil water availability, temperature and vegetation productivity indices. Global Biogeochem. Cycles 17, 15.

Rey, A., Pegoraro, E., Tedeschi, V., De Parri, I., Jarvis, P.G., Valentini, R., 2002. Annual variation in soil respiration and its components in a coppice oak forest in Central Italy. Global Change Biol. 8, 851–866.

Rey, A., Pegoraro, E., Oyonarte, C., Were, A., Escribano, P., Raimundo, J., 2011. Impact of land degradation on soil respiration in a steppe (*Stipa tenacissima* L.) semi-arid ecosystem in the SE of Spain. Soil Biol. Biochem. 43, 393–403.

Robert, C.P., Casella, G., 1999. Monte Carol Statistical Methods. Spring Verlag, New York.

Sivia, D.S., Skilling, J., 2006. Data Analysis, a Bayesian Tutorial. Oxford University Press, New York.

Tagesson, T., 2006. Calibration and Analysis of Soil Carbon Efflux Estimates with Closed Chambers at Forsmark and Laxemar. Department of Physical Geography and Ecosystem Analysis, Lund University.

Tang, J.W., Baldocchi, D.D., 2005. Spatial–temporal variation in soil respiration in an oak-grass savanna ecosystem in California and its partitioning into autotrophic and heterotrophic components. Biogeochemistry 73, 183–207.

Tang, J.W., Misson, L., Gershenson, A., Cheng, W.X., Goldstein, A.H., 2005. Continuous measurements of soil respiration with and without roots in a ponderosa pine plantation in the Sierra Nevada Mountains. Agric. For. Meteorol. 132, 212–227.

Tedeschi, V., Rey, A., Manca, G., Valentini, R., Jarvis, P.G., Borghetti, M., 2006. Soil respiration in a Mediterranean oak forest at different developmental stages after coppicing. Global Change Biol. 12, 110–121.

Tuomi, M., Vanhala, P., Karhu, K., Fritze, H., Liski, J., 2008. Heterotrophic soil respiration – comparison of different models describing its temperature dependence. Ecol. Model. 211, 182–190.

Unger, S., Máguas, C., Pereira, J.S., Aires, L.M., David, T.S., Werner, C., 2010. Disentangling drought-induced variation in ecosystem and soil respiration using stable carbon isotopes. Oecologia 163, 1043–1057.

van Oijen, M., Cameron, D.R., Butterbach-Bahl, K., Farahbakhshazad, N., Jansson, P.E., Kiese, R., Rahn, K.H., Werner, C., Yeluripati, J.B., 2011. A Bayesian framework for model calibration, comparison and analysis: application to four models for the biogeochemistry of a Norway spruce forest. Agric. For. Meteorol. 151, 1609–1621.

Vincent, G., Shahriari, A.R., Lucot, E., Badot, P.M., Epron, D., 2006. Spatial and seasonal variations in soil respiration in a temperate deciduous forest with fluctuating water table. Soil Biol. Biochem. 38, 2527–2535.

Webster, K.L., Creed, I.F., Skowronski, M.D., Kaheil, Y.H., 2009. Comparison of the performance of statistical models that predict soil respiration from forests. Soil Sci. Soc. Am. J. 73, 1157–1167.

Xu, L.K., Baldocchi, D.D., 2004. Seasonal variation in carbon dioxide exchange over a Mediterranean annual grassland in California. Agric. For. Meteorol. 123, 79–96.

Yuste, J.C., Janssens, I.A., Carrara, A., Ceulemans, R., 2004. Annual Q_{10} of soil respiration reflects plant phenological patterns as well as temperature sensitivity. Global Change Biol. 10, 161–169.

Yuste, J.C., Baldocchi, D.D., Gershenson, A., Goldstein, A., Misson, L., Wong, S., 2007. Microbial soil respiration and its dependency on carbon inputs, soil temperature and moisture. Global Change Biol. 13, 2018–2035.

Zhou, T., Shi, P.J., Hui, D.F., Luo, Y.Q., 2009. Global pattern of temperature sensitivity of soil heterotrophic respiration ($Q(10)$) and its implications for carbon-climate feedback. J. Geophys. Res. Biogeosci. 114, 9.

Zianis, D., Muukkonen, P., Makipaa, R., Mencucinni, M., 2005. Biomass and stem volume equations for tree species in Europe. Silva Fenn. Monogr. 4, 63.

CHAPTER 4

“Bayesian calibration, comparison and averaging of six forest models, using data from Scots pine stands across Europe”

Forest Ecology and Management 289 (2013) 255–268

Contents lists available at SciVerse ScienceDirect

Forest Ecology and Management

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

Bayesian calibration, comparison and averaging of six forest models, using data from Scots pine stands across Europe

M. van Oijen[a,*], C. Reyer[b], F.J. Bohn[c], D.R. Cameron[a], G. Deckmyn[d], M. Flechsig[b], S. Härkönen[e], F. Hartig[c], A. Huth[c], A. Kiviste[f], P. Lasch[b], A. Mäkelä[g], T. Mette[h], F. Minunno[i], W. Rammer[j]

[a] Centre for Ecology and Hydrology, CEH-Edinburgh, Bush Estate, Penicuik EH26 0QB, United Kingdom
[b] Potsdam Institute for Climate Impact Research, Telegrafenberg, P.O. Box 601203, Potsdam, Germany
[c] UFZ – Helmholtz-Centre for Environmental Research, Department of Ecological Modeling, Permoserstr. 15, 04318 Leipzig, Germany
[d] Plant and Vegetation Ecology, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk/Antwerpen, Belgium
[e] Finnish Forest Research Institute, PL 68, FI-80101 Joensuu, Finland
[f] Institute of Forestry and Rural Engineering, Estonian University of Life Sciences, Kreutzwaldi 5, 51014 Tartu, Estonia
[g] Department of Forest Sciences, University of Helsinki, P.O. Box 27, FI-00014 Helsinki, Finland
[h] Forest Growth and Yield Science, Technical University of Munich, 85354 Freising, Germany
[i] Institute of Agronomy, Forest Research Centre, Tapada da Ajuda, 1349-017 Lisbon, Portugal
[j] Institute of Silviculture, University of Natural Resources and Life Sciences (BOKU), Vienna, Austria

ARTICLE INFO

Article history:
Received 16 May 2012
Received in revised form 26 September 2012
Accepted 28 September 2012
Available online 28 November 2012

Keywords:
Dynamic modelling
Forest management models
Growth prediction
National forest inventories
Permanent sample plots
Uncertainty

ABSTRACT

Forest management requires prediction of forest growth, but there is no general agreement about which models best predict growth, how to quantify model parameters, and how to assess the uncertainty of model predictions. In this paper, we show how Bayesian calibration (BC), Bayesian model comparison (BMC) and Bayesian model averaging (BMA) can help address these issues.

We used six models, ranging from simple parameter-sparse models to complex process-based models: 3PG, 4C, ANAFORE, BASFOR, BRIDGING and FORMIND. For each model, the initial degree of uncertainty about parameter values was expressed in a prior probability distribution. Inventory data for Scots pine on tree height and diameter, with estimates of measurement uncertainty, were assembled for twelve sites, from four countries: Austria, Belgium, Estonia and Finland. From each country, we used data from two sites of the National Forest Inventories (NFIs), and one Permanent Sample Plot (PSP). The models were calibrated using the NFI-data and tested against the PSP-data. Calibration was done both per country and for all countries simultaneously, thus yielding country-specific and generic parameter distributions. We assessed model performance by sampling from prior and posterior distributions and comparing the growth predictions of these samples to the observations at the PSPs.

We found that BC reduced uncertainties strongly in all but the most complex model. Surprisingly, country-specific BC did not lead to clearly better within-country predictions than generic BC. BMC identified the BRIDGING model, which is of intermediate complexity, as the most plausible model before calibration, with 4C taking its place after calibration. In this BMC, model plausibility was quantified as the relative probability of a model being correct given the information in the PSP-data. We discuss how the method of model initialisation affects model performance. Finally, we show how BMA affords a robust way of predicting forest growth that accounts for both parametric and model structural uncertainty.

1. Introduction

Ecological models are built for a variety of purposes. One general motivation is trying to integrate our understanding of the processes underlying natural phenomena. At a time when the earth system is subject to substantial changes in land use and climate, however, it also becomes of increasing importance to be able to make quantitative predictions, supported by a quantification of uncertainty, about the future of our ecosystems.

Forest ecosystems are a prominent example where quantitative predictions are of particular ecological and economic importance, but for which there is considerable uncertainty because different modelling approaches, models and parameters are available (Mäkelä et al., 2012). We focus here on weather-sensitive dynamic models, which simulate the growth of forest stands over time. Dynamic models that have been considered for forest management range from fairly simple, parameter-sparse empirical models to complex models with many parameters (Fontes et al., 2010). None

---

[*] Corresponding author. Tel.: +44 131 4458567; fax: +44 131 4453943.
E-mail address: mvano@ceh.ac.uk (M. van Oijen).

0378-1127/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.foreco.2012.09.043

of these models has found widespread application across Europe, which may be due to problems of parameterisation and a lack of knowledge about the generalisability of the models. Given the increasing availability of forest data from National Forest Inventories (NFIs) and Permanent Sample Plots (PSPs), and other data sources, however, it can be hoped that limitations of dynamic forest models with respect to data availability can be substantially reduced in the future (Hartig et al., 2012). These data can help in parameterisation and evaluation of the models, if we can find robust ways of comparing models and accounting for measurement and modelling uncertainties. In this paper, we use methods based on probability theory, more specifically Bayesian calibration (BC), Bayesian model comparison (BMC) and Bayesian model averaging (BMA), to address these issues. A strength of these methods is that they can be applied to any type of model. Although we do restrict our focus here to dynamic, weather-sensitive models, we have included models of widely differing structure, complexity and data needs, providing a broad practical test of the methods.

Bayesian methods have been used before to calibrate the parameter distributions of dynamic forest models, starting with the work of Green et al. (1999), but application to parameter-rich process-based models is still rare (Luo et al., 2009). The use of BMC to compare and evaluate dynamic forest models – or any other vegetation models – is a more recent application. Van Oijen et al. (2011) included BMC in their analysis of four models for forest biogeochemistry and Fu et al. (2012) used BMC to identify the most plausible models for predicting tree budburst. Here we present, as far as we know, the first applications of BMC and BMA to dynamic forest growth models that include both parameter-sparse semi-empirical models and complex process-based models with many parameters. Using NFI- and PSP-data on Scots pine (*Pinus sylvestris* L.) from four European countries, we compared the results of calibration and testing of these models using the combined dataset with the results where the same methods were applied to within-country data only. The purpose of this was to assess whether the models would be most effectively calibrated and applied at smaller or larger spatial scales. Similar comparisons of Bayesian approaches applied locally and generically have been made for a simple soil ionic concentration model by Reinds et al. (2008) and for a model of N_2O-emissions in crops by Lehuger et al. (2009).

We ask the following questions:

- How effective are local stand data in reducing uncertainties about forest model parameters in a Bayesian framework?
- Are the considered dynamic models for Scots pine sufficiently general to allow a generic calibration to data from across Europe, or should models be calibrated on a country-by-country basis?
- How effective is Bayesian model comparison in identifying plausible predictive models, and what are the main distinguishing characteristics of forest models that are selected?
- Does Bayesian model averaging lead to improved predictions compared to individually calibrated models?

Although these questions, as well as the models and data used, are focused on forestry in Europe, our methodology is unrestrictedly general. BC, BMC and BMA, and the contrasts made between within- and cross-country applications, can be applied to any other combination of data sets and models in the environmental sciences.

2. Materials and methods

2.1. Overview of methodology

Our study used six models and 12 data sets which originated from forest measurements in four European countries (Table 1). The data were from National Forest Inventory (NFI) sites and from sites with Permanent Sample Plots (PSPs). From all sites we retrieved environmental data (weather, soil and management) and tree growth data (height and diameter). These data were used by all models to the extent of each model's input data requirements (Table 2). Fig. 1 is a flow chart that shows how the data were used in the consecutive stages of the study. The environmental data from the NFI-sites were used as drivers for model application to those sites. Each model was run multiple times for each NFI-site, to assess the impact of parameter uncertainty on model outputs. We refer to this step as 'prior uncertainty quantification' (prior UQ) because no data of tree growth had been used at this point for improvement of parameter values. The distributions of model outputs generated by this prior UQ were used in a Bayesian model comparison (prior BMC) to quantify the relative plausibility of each model before calibration. These differences in model plausibility were then used as weights in Bayesian model averaging (BMA), thus producing an averaged prediction to which all six models contributed differently. Next, the NFI-data were used for Bayesian calibration of the parameters of the different models. The calibration was carried out both per country and generically using data from all NFI-sites. The calibrated models were then applied to the PSP-sites using local environmental data. At this stage, we again carried out uncertainty quantification, now termed 'posterior UQ' because the model parameter distributions were already informed by the NFI-data. Finally, the results from the posterior UQ were compared with measurements from the PSP-sites for a posterior Bayesian model comparison, again accompanied by BMA. In the rest of this section, we describe data, models and statistical methods in more detail.

Table 1
Data. Each row represents one of the twelve measurement sites. If multiple values of stem number are shown, they refer to changes over the period of measurement. The rightmost column gives the total number of data points at the site, for tree height and diameter combined.

Country	Site name	Site code	Site type	Lat. (°)	Long. (°)	Plot size (m^2)	Mean temp. (°C)	Mean precip. (mm y^{-1})	Age at last obs. (y)	Stem number (ha^{-1})	# Data
Austria	Point 1	A1	NFI	48.31°	14.79°	1200	7.6	855	∼64	554–526	4
	Point 2	A2	NFI	48.51°	15.70°	1200	9.2	466	∼66	1772–1363	4
	PSP	A3	PSP	48.51°	15.70°	1500	9.2	466	59	790–690	4
Belgium	Hechtel	B1	NFI	51°17′	5°31′	1000	9.9	812	67	400–380	4
	Pijnven	B2	NFI	51°17′	5°31′	1000	9.9	819	66	520–393	4
	Brasschaat	B3	PSP	51°18′	4°31′	20000	9.9	811	79	538–362	6
Estonia	EST-1	E1	PSP	57°51′	25°55′	1963	5.4	629	70	428–402	6
	EST-2	E2	PSP	57°59′	25°38′	1257	5.4	632	67	796–692	6
	EST-3	E3	PSP	57°35′	25°17′	1963	5.3	625	73	652–667	6
Finland	NFI-1	F1	NFI	61°58′	27°40′	100–300	2.8	534	75	899	4
	NFI-2	F2	NFI	63°50′	24°39′	100–300	2.2	442	55	1067	4
	Vesijako	F3	PSP	61°20′	25°2′	1000	3.5	521	79	8700–1710	14

Table 2
Models. Each row represents one of the six models. The weather variables driving the models include radiation, temperature, precipitation, wind speed and atmospheric humidity (BASFOR), or a subset of those (3PG, 4C, ANAFORE, BRIDGING, FORMIND). The rightmost column shows whether models simulated forest growth from planting or were initialised using the earliest measurements at each site. IBM = Individual-Based Model requiring specification of size and position of each tree.

Model	Time step	Environmental variables	Number of state variables	Number of parameters (# in calibration)	Initialisation
3PG	Monthly	Weather	9	51 (48)	Planting date
4C	Daily–yearly	Weather, soil conditions, N-deposition, CO_2	15	46 (43)	First measurement
ANAFORE	Half-hourly	Weather, soil conditions, N-deposition, CO_2	26	146 (138)	First measurement
BASFOR	Daily	Weather, N-deposition, CO_2, soil conditions	14	48 (41)	Planting date
BRIDGING	Yearly	Weather	5	38 (13)	First measurement
FORMIND	Yearly	Weather	IBM	42 (4)	First measurement

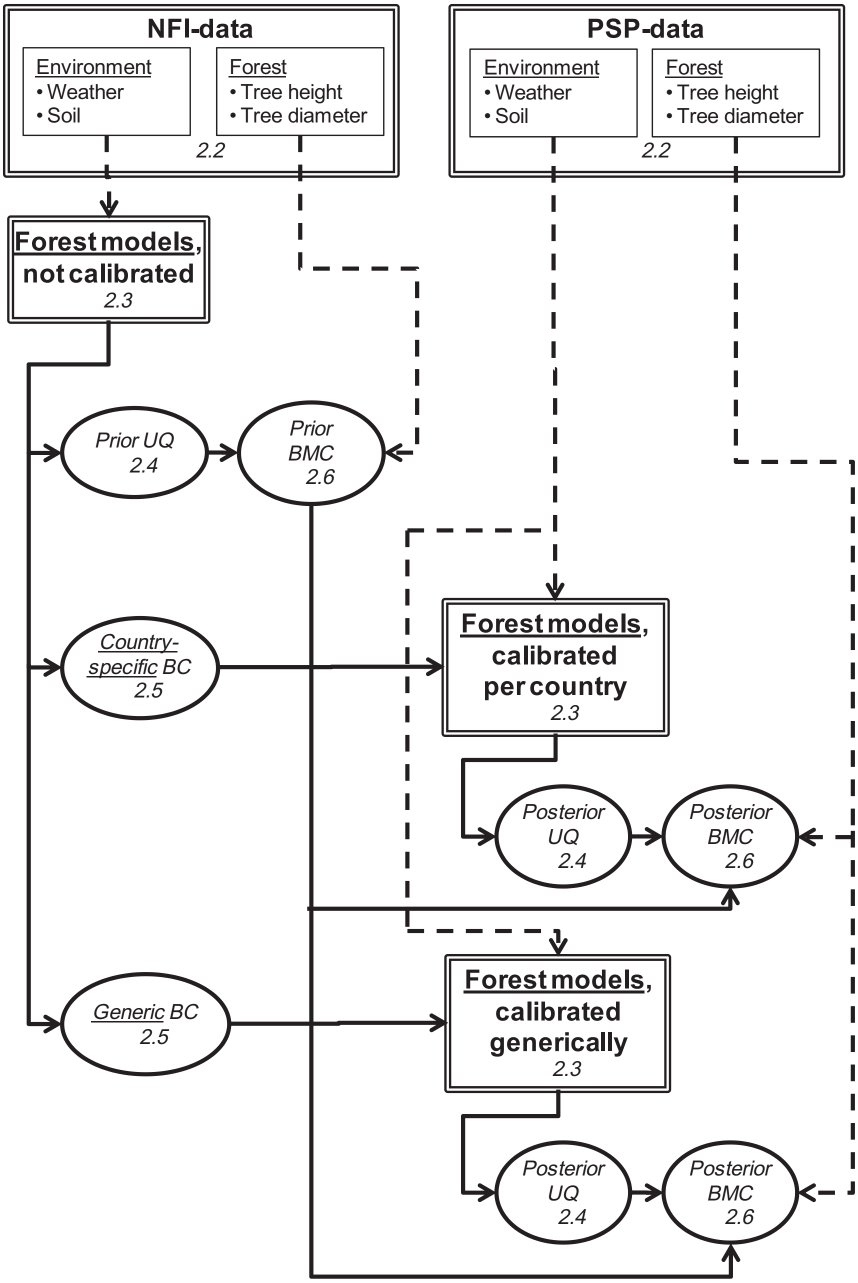


Fig. 1. Flow chart of the study. The numbers within icons (2.2–2.6) indicate in which paragraph of Section 2 further explanation can be found.

2.2. Data

Data of twelve even-aged *P. sylvestris* stands were assembled from four European countries (Table 1). From each country, two NFI sites and one PSP-site were selected for this study. An exception was Estonia, for which NFI-data were not available and three PSPs were used. For ease of reference, we used a site-code for each site consisting of the first letter of the country's name, followed by 1 or 2 for the NFI-sites and 3 for the PSP-site (Table 1), except for Estonia where the numbers refer to the three

PSPs. For model calibration, we only used data from the sites coded 1 or 2, whereas for model comparison and averaging the data from sites with code number 3 were used. The data used were for mean tree height and stem diameter at 1.3 m above ground, which were available from all sites. Data on stem number and tree age were used as uncalibrated inputs. All sites provided several measurements for the different variables (between two and seven), separated by intervals of at least 5 years (Fig. 2). We now briefly describe the sites in each country.

2.2.1. Austria

The NFI-plots A1 and A2 are part of the Austrian Forest Inventory grid consisting of ~10,000 points. The plots are 100% *P. sylvestris* and the soils are classified as Semipodsol and Cambisol with soil depths exceeding 0.3 m and field capacity around 36%. They are located at different altitudes in the "Waldviertel", a region in Lower Austria north of the Danube. A1 lies about 300 m higher than A2 and is cooler and drier. On both sites, measurements were taken in two years (1987 & 2000 and 1989 & 2002). The sample consisted for each plot of a combined angle count measurement (for trees >10.5 cm diameter) and a circle with a fixed radius (for trees <10.5 cm). Height measurements were done for a subset of trees of the angle count measurement; the other heights were calculated. Nothing is known about management history or planting time, except that no management occurred during the period of measurements.

The selected PSP-site, A3, was established in 1970 and measured every 5 years. The site is maintained by the Austrian Federal Forest Office BWF (http://bfw.ac.at/) and is located near A2 with similar soil properties. It is a pure *P. sylvestris* stand with a size of 1500 m^2 and a stem number of 790 ha^{-1} in 1980.

Climate data for the NFI- and PSP-sites were provided from nearby weather stations of the Austrian weather service ZAMG (Central Institute for Meteorology and Geodynamics).

All three stands reached heights of about 18 m at an age of about 60 years. However, they differ significantly in diameter (207–324 mm), with lower values at high stem number.

2.2.2. Belgium

The Belgian plots B1 and B2 are NFI's of the ANB (Agentschap Natuur en Bos, 'Forest and Nature Agency'), situated in the Campine region of north-eastern Belgium, were established in 1937 and 1942 respectively and regularly thinned since then from the original 12,500 trees ha^{-1}. B1 is situated on loamy sand, and data from 2000 and 2004 were available; thinning during this period reduced stem number from 400 to 380 ha^{-1}. B2 is situated on sandy soil close to B1 and data from 2000 and 2008 were available. Thinning during this period reduced stem number from 520 to 393 ha^{-1}. The data were obtained from 40 × 25 m sample plots.

The PSP-site, B3, "De Inslag", is a mixed patchy coniferous/deciduous forest located in Brasschaat also in the Belgian Campine region. The site is part of the European Carboeurope-IP network and is a level-II observation plot of the European network program (ICP-II forests) for intensive monitoring of forest ecosystems (EC-UN/ECE, 1996), managed by the Flemish Research Institute for Nature and Forest (INBO). Here we only focus on one particular even-aged Scots pine stand planted in 1929 and described by Curiel Yuste et al. (2005). In this experimental stand, stem number was 556 ha^{-1} in 1997. In November 1999, a thinning was performed reducing the stem number to 377 ha^{-1} and further thinned to 362 ha^{-1} in 2002. The soil is loamy sand, moderately wet, with a distinct humus and iron B-horizon (Baeyens et al., 1993) and is classified as Umbric Regosol. Although the Belgian plots are on relatively sandy soils, soil water table is quite high (0.7–1.1 m) and soil fertility is high due to high nitrogen deposition (30–40 kg N ha^{-1} year^{-1}).

Despite similar age (66–67 years) and stem number (380–390 ha^{-1}), the two NFI-plots had quite different heights (18.4 and 23.2 m) and diameter (271 and 293 mm) indicating differences in site quality. The PSP-site was older and had lower tree number; height was intermediate but diameter was greater than at the NFI-plots.

2.2.3. Estonia

The Estonian plots E1, E2 and E3 belong to the Estonian Forest Research Plots Network which consists of more than 700 PSP and

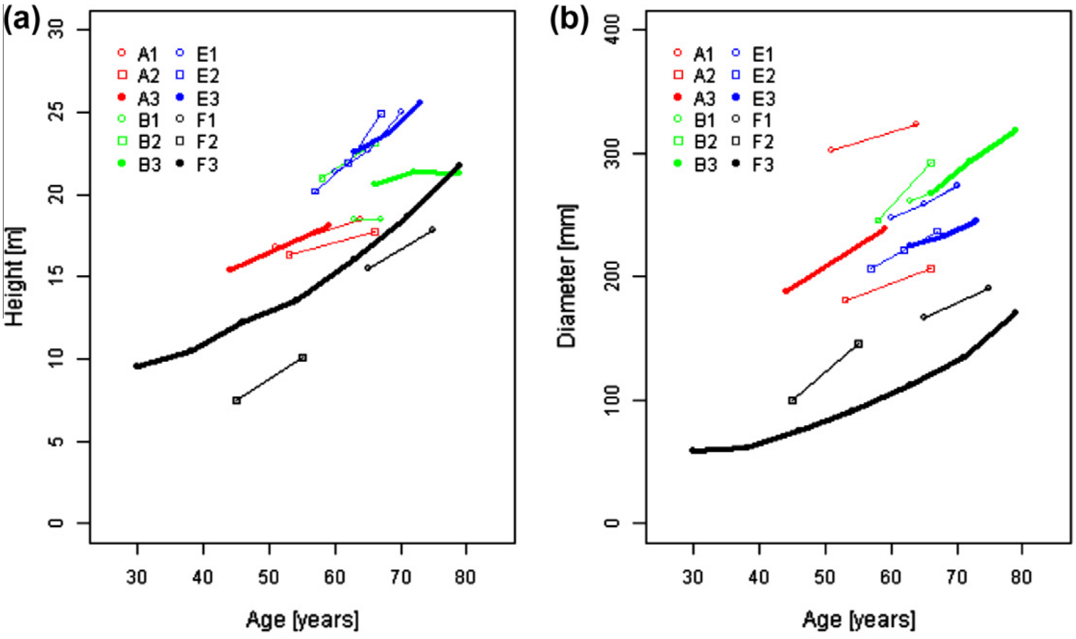


Fig. 2. (a) Mean tree height vs. stand age as observed at the twelve forest sites. (b) Idem for stem diameter. Site-codes (A1 to F3) are explained in Table 1.

are maintained by the Estonian University of Life Sciences (Sims et al., 2009). These plots were established at the observation sites of the European network programme ICP Forest Level I plots. The plots, established in 2000, are circular with radii of 25, 20 and 25 m, respectively and were re-measured in 2005 and 2010. The plots have not been thinned during that period, but earlier management history is unknown. On each plot, the diameter at breast height was assessed for each tree. Tree height and height to crown base were measured in every fifth tree. All three plots are dominated by Scots pine (more than 90% of total volume), but there is a small mixture of Silver birch (*Betula pendula*) and Norway spruce (*Picea abies*). The plots are located in southern Estonia where mean effective temperature sum is about 1650° days. The plots are on sandy soils on glaciofluvial deposits with sufficient water availability belonging to WRB 2006 soil units Gleyic Podzol, Histic Podzol and Albic Podzol respectively. The vegetation types of the plots are Rhodococcum, drained Polytrichum-Nyrtillus, and Rhodococcum. The basal area of the plots reached 24.8, 33.7, and 31.8 $m^2\,ha^{-1}$ at stand ages 70, 67, and 73 years, with average heights of 25.2, 24.7, and 25.6 m and volumes of 285, 384, and 374 $m^3\,ha^{-1}$. Differences in diameter (237–274 mm) were larger than height differences, with largest values reached at the lowest stem number.

2.2.4. Finland

The Finnish plots F1 and F2 are permanent NFI sample plots located in Southern Finland established by the Finnish Forest Research Institute. They have been measured in 1985 and 1995. The plots have not been thinned during that period. The earlier treatment history is unknown. The plot size varied according to the stem diameter at breast height, being 100 m^2 when the diameter was under 10.5 cm, and otherwise 300 m^2. The trees with diameter smaller than 4.5 cm were measured only if they were expected to survive until the next measuring date. Diameter at breast height and tree species were recorded from all the tally trees. Heights, crown base heights and crown widths were measured from the sample trees, which include the trees that were located in a circular area around the sample plot mid-point, where the circle radius is half of the original sample plot radius.

The Finnish plot F3 is a control plot with no thinnings in a permanent thinning experiment of the Forest Research Institute at Vesijako in southern Finland. The experiment was established in 1948 in a pine stand sown in 1918, and it was followed until 1997. The site is fairly fertile with adequate moisture for pine. The plot has a small mixture of birch (*Betula* spp.), less than 10% of basal area. Plot size was 1000 m^2, and all trees were numbered on this plot and measured for breast height diameter in a total of seven measurements. For height (and crown base height in the two most recent measurements), 21–67 trees were chosen as sample trees. The final heights of 17.8 m (75 years, NFI 1), 10.1 m (55 years, NFI 2) and 21.8 m (79 years, PSP) indicate that despite the age difference, the site conditions at NFI 2 were probably less favourable (cf. Fig. 2a). The comparatively low stem number and the high diameter, and the fact that no mortality occurred, suggest that the NFI plots were thinned at some point before the surveys. In contrast, at the PSP-site only self-thinning occurred leading to high stem numbers and low diameters.

2.3. Models

We used six different forest models in the assessment, ranging from simple semi-empirical models to parameter-rich process-based models (Table 2). All models are able to predict mean tree height and mean stem diameter. Some of the models are able to simulate variation between individual trees as well, but the corresponding predictions were not tested against data. Four of the models are initialised at the first measurement date, i.e. they require the earliest observed values of mean tree height and/or diameter to quantify the model's initial constants (Table 2). This reduces the number of data available for Bayesian calibration. The remaining two models, 3PG and BASFOR, include state variables that are difficult to estimate from mean height and stem diameter only, such as nitrogen pools in soil and trees, and it was therefore decided to initialise them from planting. These two models therefore have more data available for calibration, but their predictions of forest growth may already start deviating from observations before the first measurement date. We shall now briefly describe each model, referring to earlier publications for more detail. Each model description finishes with an account of how the prior probability distribution for the model's parameters was set by the respective modellers. The role of these probability distributions in uncertainty quantification and Bayesian calibration is explained in Sections 2.4–2.5.

2.3.1. 3PG

3PG calculates the dynamics of biomass in different organs (foliage, roots and stem) and simulates the soil water balance and variables of interest to forest managers, such as stand timber volume, mean diameter at breast height, stand basal area and mean annual growth increment. Gross primary production (GPP) is calculated by multiplying photosynthetically active radiation absorbed by the stand with a light-use efficiency that changes with environmental conditions. Light absorption is calculated using Beer's law, while the light-use efficiency varies in dependence of atmospheric vapour pressure deficit, air temperature, the presence of frost, soil water balance, tree age and site fertility. Net primary productivity (NPP) is calculated as a constant fraction of GPP (Law et al., 2000; Waring et al., 1998). Carbon allocation is based on allometric equations, applied on a single-tree basis. The fraction of NPP allocated below-ground decreases with soil fertility. Site fertility is expressed through a site specific reduction factor (FR) that varies between 0 (for the least fertile sites) and 1 (for sites that do not have nutrient limitations). The remaining NPP is partitioned between the aboveground organs as a function of stem diameter at breast height. The diameter at breast height and the average stand height are calculated through allometric functions of average aboveground biomass per tree. 3PG has been applied to various different species and sites and is widely used in research as well as by companies to assess forest growth and site productivity. Detailed descriptions of 3PG were provided by Landsberg and Waring (1997) and Sands and Landsberg (2002).

Before this study, Landsberg et al. (2005) tested the performance of 3PG for Scots pine in Finland, using a modified carbon allocation routine. Xenakis et al. (2008) coupled 3PG with ICBM/2N (Introductory Carbon Balance Model (Andren and Katterer, 1997)) a soil matter decomposition model. The new model, 3PGN, was calibrated and tested for Scots pine plantations in Scotland. The information from these two previous studies was utilised to construct the prior, using truncated Gaussian distributions. For each parameter, the prior mean was set to the average of the values used in Landsberg et al. (2005) and Xenakis et al. (2008). The bounds of the prior were set at ±30% of the mean value. The site fertility parameters were also included in the BCs and BMCs; the FRs ranged between 0 and 1, while the prior mean was 0.5. For all parameters, the prior was kept quite uninformative (i.e. high variance and wide ranges), reflecting the fact that the 3PG-modeller in the current study did not have previous experience with Scots pine.

2.3.2. 4C

The forest model 4C (FORESEE–FORESt Ecosystems in a changing Environment) has been developed to simulate the impact of changing environmental conditions on forest ecosystems. It is climate sensitive and calculates physiological processes on the tree and stand level depending on the process in question in daily to yearly time steps (Bugmann et al., 1997; Suckow et al., 2001). Establishment, growth and mortality of tree cohorts are explicitly modelled at the patch scale on which horizontal homogeneity is assumed. Cohorts of trees compete for light, water and nutrients (Bugmann et al., 1997). Every cohort develops specific values for fine root, foliage, stem biomass, etc. and species-specific parameters steer the physiological processes for each species. Photosynthetic rate is calculated after Haxeltine and Prentice (1996) and a constant fraction of GPP is lost to respiration (Landsberg and Waring, 1997). The resulting NPP thus depends on environmental conditions and is allocated according to the principles of the pipe model (Shinozaki et al., 1964) and of the functional balance (Davidson, 1969) and organ-specific, constant senescence rates. In this allocation model, height growth is decoupled from diameter growth, with high degrees of intra-canopy shading leading to extra height growth. Nitrogen limitation has been calculated dynamically. When the tree water demand of a cohort exceeds the plant available water in the soil, the canopy conductance and ultimately NPP of that cohort is reduced. 4C requires daily meteorological variables, a soil description including physical and chemical parameters as well as a forest stand description. For further details of model processes and recent model applications, see Suckow et al. (2001), Lasch et al. (2005), Seidl et al. (2008) and Reyer et al. (2010).

The prior distribution for all parameters of 4C was uniform with boundaries at ±50% of the initial (standard 4C) value, reflecting large uncertainty about parameter values. The selection of the parameters to be calibrated was restricted to species-specific parameters that could be informed by Scots Pine data, giving a total of 43 parameters amenable to calibration.

2.3.3. ANAFORE

ANAFORE (ANAlysing FORest Ecosystems) is a stand-scale, mechanistic forest model that dynamically simulates the fluxes of carbon, water and nitrogen through the ecosystem (Deckmyn et al., 2008). The forest stand is described as consisting of trees of different size cohorts (e.g. dominant, co-dominant and suppressed trees), either of the same or of different species (deciduous or coniferous). Half-hourly carbon and water fluxes are modelled at the leaf, tree and stand level from half-hourly, daily or monthly climate data. In addition to total growth and yield, the model simulates allocation changes in crown size, DBH-height ratio, root-shoot ratio and even the daily evolution of tracheid or vessel biomass and radius, parenchyma and branch development. From these data, early and late wood biomass, wood tissue composition and density are calculated to allow wood quality estimation. Simulation of the labile carbon stored in the living tissues allows for simulation of trans-seasonal and trans-yearly effects, and simulation of the long-term effects of environmental stresses on growth. A detailed soil model including fungal, bacterial and mycorrhizal effects on SOM degradation and aggregate formation is included (Deckmyn et al., 2009). Model initialisation was at the first measuring point. Because ANAFORE needs a detailed tree description – not available for most sites – allocation as observed at the Belgian sites was used throughout (% heartwood, branch biomass, crown length). Crown width was set to fill the site.

The prior distribution for the parameters was uniform with boundaries at ±10% of the initial value, reflecting measured data (mainly on the Belgian Brasschaat site) and data from literature as described in Deckmyn et al. (2008). Although ANAFORE was calibrated for Scots pine before this study, this was only for Belgian stands and the uncertainty concerning parameterisation across Europe is large, so the same prior was used.

2.3.4. BASFOR

The BASic FORest simulator, BASFOR, is a deterministic daily time step forest model used for simulating coniferous or deciduous forests. The model simulates carbon and nitrogen cycling in trees, soil organic matter and litter. It simulates the response of trees and soil to radiation, temperature, precipitation, humidity, wind speed, atmospheric CO_2 and N-deposition, and thinning regime. The model has 14 state variables, representing carbon and nitrogen pools in trees and soil, and 48 parameters which include the initial constants of the state variables. Besides time series for the state variables, output may be produced of NPP, tree height, stem diameter, ground cover, LAI, N-mineralisation and other tree and soil variables. BASFOR is built from well known process representations. Light absorption is calculated by Beer's law. GPP is calculated as light absorption times a light-use efficiency (LUE). NPP is calculated as a fixed ratio of GPP. LUE is temperature-, CO_2- and water-dependent and may be reduced if insufficient nitrogen is taken up by the plants. Potential nitrogen uptake scales with root system surface area. Actual nitrogen uptake is the minimum of demand, determined by tissue N-concentration, and potential uptake. Allocation of assimilates follows allometric rules, but water stress may limit leaf area index (LAI). Turnover of tree and soil components proceeds at temperature-dependent relative rates.

The model structure was described by Van Oijen et al. (2005), more recent model applications are reported by Van Oijen and Thomson (2010) and Van Oijen et al. (2011), and the model is now also in use as the tree component of an agroforestry model (Van Oijen et al., 2010). The prior for BASFOR was constructed from beta-distributions for the individual parameters, with ranges and modes based on literature as described before (Levy et al., 2004; Van Oijen et al., 2005, 2011).

2.3.5. BRIDGING

The BRIDGING model (Valentine and Mäkelä, 2005) was developed to bridge the gap between process-based and empirical approaches to modelling tree growth by formulating a process-based model that can be fitted and applied in an empirical mode. Tree growth in the model is based on carbon balance, and its allocation is consistent with pipe model theory and an optimal control model of crown development (Mäkelä and Sievanen, 1992). These provide a framework for expressing the components of tree biomass in terms of tree height, crown height and stem cross-sectional area, the growth of which is regulated by photosynthesis and respiration. The parameters of the model comprise physiological rates and morphological ratios and can be estimated from lower-level process models or direct measurements. In the empirical mode, the original parameters are combined into a set of fewer, aggregate parameters which can be estimated from inventory type data using statistical procedures. Here, we calculate the photosynthesis and respiration parameters from lower-level models of stand productivity (Mäkelä et al., 2008) and canopy structure (Duursma and Makela, 2007) using a procedure proposed by Härkönen et al. (2010). The productivity model is driven by daily data of global radiation, vapour pressure deficit and air temperature, while field data on inventory variables (stand-level mean values of height, diameter, crown base height and crown width, stocking density or basal area, and site fertility) are used for parameterising canopy structure. These parameters are given fixed, deterministic values. The parameters related to growth of tree height and basal area are employed in their aggregate form and estimated using the Bayesian approach with the given inventory data.

The BRIDGING model has 38 different parameters, of which the 13 parameters relating to the dynamic growth of tree height and basal area were used in the calibration. Uniform distributions were used throughout. Parameters left out of the calibration included structural relationships, which were calculated directly based on the measured stand data, biomass estimates, and light-use efficiency estimates. The uniform distributions were mainly quantified based on earlier pipe model studies (Mäkelä, 1997; Mäkelä and Vanninen, 2001; Vanninen and Mäkelä, 2005; Valentine and Mäkelä, 2005; Palmroth et al., 1999; Duursma and Makela, 2007).

2.3.6. FORMIND

FORMIND is an individual-based, spatially semi-explicit gap-type model (Köhler and Huth, 1998; Ruger et al., 2007). Spatially semi-explicit means that the modelled plot (in this case 1 ha) is divided into 20 × 20 m gaps. Tree individuals are assigned to one of these gaps, but do not have an explicit position within gaps. As in classical gap models, tree crowns are assumed to cover the gap uniformly in horizontal direction at a certain height, depending on the size of the trees. The vertical stratification through the different crown heights of the trees and the differences in light climate that result from that for each individual tree are important determinants of the predicted community dynamics. NPP is calculated as the difference between GPP and respiration. GPP of each individual tree depends on the available light at crown top, temperature and soil water content. The temperature dependence follows a hump shape. A reduction due to insufficient soil water occurs below a threshold and GPP is completely reduced if soil water content falls below the permanent wilting point. Additionally, maintenance respiration has a temperature dependence following the Q10-approach (Gutiérrez and Huth, 2012). The model was initialised for each site at the first recorded year with the observed number of trees, all of the same observed average diameter, randomly distributed over the modelled area of one hectare.

The marginal prior probability distributions for FORMIND were all uniform. Parameters were excluded from the calibration that were either unrelated to those model outputs that were compared to calibration data, or for which there were other parameters already under calibration that acted on the model outputs in a similar way. Based on this premise, four parameters were selected for calibration. These included the two parameters that determine the diameter-height relationship, the main growth parameter that determines the maximum growth rate under full light, and the wilting point, which is the determinant of how strongly the plants react to water stress. The other parameters were fixed according to literature data. For each of the calibration parameters, flat and relatively wide priors were chosen reflecting large uncertainty about parameter values.

2.4. Uncertainty quantification (UQ)

Predictive uncertainty (i.e. uncertainty regarding model outputs) was quantified for each model at three stages in our study: before any parameter calibration had been carried out (prior UQ), and after country-specific and generic calibration (posterior UQ) (Fig. 1). In each case, the UQ consisted of running the model 1001 times, using a sample of that length from the parameter distribution for the model.

For each model, the prior parameter uncertainty – before any of the NFI- or PSP-data had been used for calibration – was expressed in the form of a probability distribution. This was done by each modelling group separately, no standardisation of priors being attempted (see Section 2.3). To derive from that the prior predictive uncertainty, we used a sample consisting of the mode of this parameter distribution plus 1000 other parameter vectors sampled from the prior distribution using Latin Hypercube Sampling to ensure good coverage of parameter space. This prior UQ was carried out for all 12 sites.

To assess the posterior predictive uncertainty, i.e. the uncertainty resulting from the reduced parameter uncertainty after country-specific or generic Bayesian calibration (see Section 2.5), we used the mode of the posterior parameter distribution, i.e. the Maximum A Posteriori (MAP) parameter vector, and again 1000 other parameter vectors that were selected by equidistant subsampling from the parameter chains generated in the calibration. Posterior UQ was carried out only for PSP-sites because the data from those sites had not been used in the calibration.

2.5. Bayesian calibration (BC)

Bayesian calibration was carried out as documented in other recent forest model studies (Van Oijen et al., 2005, 2011) and we shall give only a brief outline here. The method starts by expressing uncertainty about the model's parameter values in a so-called prior parameter distribution, $\mathrm{P}(\theta)$. In this notation, θ represents the full parameter vector of a model, so $\mathrm{P}(\theta)$ is a multivariate distribution. All modellers in this study assigned prior distributions without any correlations between different parameters, so $\mathrm{P}(\theta)$ could be written as the product of independent distributions for the individual parameters. By comparing model predictions with NFI-data, D, we can derive a likelihood value $\mathrm{P}(D|\theta)$ for each possible parameter value (see Section 2.6), which can be interpreted as a relative "goodness-of-fit" measure for this parameter (Hartig et al., 2012). Bayes' formula then allows us to combine both pieces of information (prior and likelihood) into one posterior parameter distribution. The formula states that:

$$\mathrm{P}(\theta|D) \propto \mathrm{P}(\theta) \cdot \mathrm{P}(D|\theta),$$

i.e. that posterior probability is proportional to prior times likelihood $\mathrm{P}(D|\theta)$. To derive a likelihood function, we made the assumption, for all models and measurements, that measurement errors were normally distributed with a coefficient of variation of 20%. The fairly high value of 20% was chosen to account for multiple factors affecting the measurements, including instrument error, demographic stochasticity of the tree populations, and environmental heterogeneity. No correlations between measurement errors were assumed, so our likelihood function could be written as the product of independent Gaussian functions of the difference between data D and model output $M(\theta)$:

$$\begin{aligned}\mathrm{P}(D|\theta) &= \text{Probability of measurement error equal to } D - M(\theta)\\ &= \prod_{i=1}^{n} \varphi(D_i - M_i(\theta); 0, (0.2D_i)^2),\end{aligned}$$

where the i-subscripts index the n data points and the corresponding model outputs, and where φ denotes a Gaussian probability density function with given mean and variance.

To estimate the posterior distributions, we used a Markov Chain Monte Carlo (MCMC) algorithm (Metropolis et al., 1953; Van Oijen et al., 2005). Convergence of the MCMC was verified both visually – by inspection of the parameter trace plots – and by calculation of the Gelman-Rubin statistic (Gelman and Rubin, 1992).

2.6. Bayesian model comparison (BMC) and calculation of NRMSE

Bayesian model comparison relies on the same probabilistic ideas as BC, but now the probability distribution to be informed by the data is not that for the parameters but for the models themselves (Kass and Raftery, 1995). A key strength of BMC is that it evaluates models not at one single parameter vector value but takes into account parameter uncertainty (Tuomi et al., 2008). The formal need for this coverage of parameter uncertainty is seen when we write out Bayes' Theorem as applied to model comparison:

$$\mathrm{P}(M|D) \propto \mathrm{P}(M) \cdot \mathrm{P}(D|M),$$

where following the law of total probability:

$$\mathrm{P}(D|M) = \int \mathrm{P}(D|M(\theta))\mathrm{P}(\theta)d\vartheta.$$

So each model's parameter uncertainty, and not only the best value, determines how much support a model receives. Among other things, this provides a natural safeguard against overfitting using overly flexible models. P(D|M) is referred to as the 'integrated likelihood', or also the 'marginal likelihood' as it is calculated by marginalizing out the uncertain influence of the model's parameters. We assumed that each model had the same prior probability of 1/6 before any data were used. Application of the models to the NFI-sites, in the prior UQ, provided 1000 model results which were used to derive each model's integrated likelihood for those data. The posterior probability for each model was then calculated as the model's integrated likelihood divided by the sum of the integrated likelihoods for all models (Kass and Raftery, 1995). A similar procedure was applied at the next applications of BMC, where the integrated likelihoods of the models were calculated for the PSP-data after the models had been calibrated on the NFI-data. These posterior BMC's were carried out after both country-specific and generic BC.

Additionally, we calculated a standard goodness-of-fit measure, the normalised root mean squared error (NRMSE), for model predictions at PSP-sites. This was done for both the prior and posterior parameter distributions. In contrast to the calculation of the integrated likelihood, the NRMSE had to be calculated separately for height and diameter as its calculation involves a normalisation by the average of the measurements:

$$\mathrm{NRMSE} = \frac{1}{\bar{D}}\sqrt{\frac{1}{n_c \times 1000}\sum\nolimits_{c=1}^{n_c}\sum\nolimits_{\theta=1}^{1000}(M_c(\theta) - D_c)^2},$$

where n_c is the number of countries from which PSP-data were used, D_c are the measured values, $\bar{D}$ is the average of the measurements across the n_c countries, θ indexes the 1000 parameter vectors sampled from prior or posterior distribution and $M_c(\theta)$ is model prediction for country c using parameter vector θ. In the case of the prior and generic posterior parameter distribution, the calculation of NRMSE uses $n_c = 4$, but in the case of country-specific posteriors, NRMSE is calculated first per country ($n_c = 1$) followed by averaging of the four errors to arrive at an estimate of overall NRMSE.

2.7. Bayesian model averaging (BMA)

Bayesian model averaging uses the different model probabilities P(M), derived in preceding BMC, to calculate a weighted probability distribution for model outputs (Hoeting et al., 1999; Kass and Raftery, 1995):

$$\mathrm{P}(y) = \sum_{m=1}^{6}\mathrm{P}(M^{(m)})\mathrm{P}(y|M^{(m)}),$$

where P(y) is the averaged output distribution, P($M^{(m)}$) is the probability for model m as derived from the BMC, and P (y|$M^{(m)}$) is the output distribution for model $M^{(m)}$. Expanding the last term gives:

$$\mathrm{P}(y) = \sum\nolimits_{m=1}^{6}\mathrm{P}(M^{(m)})\int \mathrm{P}(y|M^{(m)}, \theta^{(m)})\mathrm{P}(\theta^{(m)})d\theta^{(m)},$$

which shows that the BMA accounts for both overall model structural uncertainty, P($M^{(m)}$), and each individual model's parameter uncertainty, P($\theta^{(m)}$). In this study, BMA was applied after both prior and posterior BMC, with P($\theta^{(m)}$) representing prior and posterior parameter uncertainty, respectively. The same model output samples used in BMC were used for BMA as well, but subsampled with sample size proportional to P($M^{(m)}$). The BMA-forecasts thus produced were compared against the measurements at the PSP-sites. Note that in this procedure only the prior BMA was subjected to a fully out-of-sample test of predictive capacity of the model averaging.

3. Results

3.1. Uncertainty quantification before and after Bayesian calibration

The first quantity calculated was the prior predictive uncertainty, that is, the model uncertainty before any data were used for calibration. Table 3 shows summary statistics of the prior predictive distributions for the NFI-sites: the value of mode of the prior plus the 5% and 95% quantiles. Figs. 3 and 4 depict the ranges between the 5% and 95% quantiles for the PSP-sites. The prior output ranges – delimited by the 5% and 95% quantiles – were generally widest for the three most parameter-rich models, i.e. ANAFORE, BASFOR and 3PG.

Bayesian calibration (BC) was carried out both per individual country and generically, so samples from five different posterior parameter distributions were produced for each model. Our results show that generic Bayesian calibration reduced parameter uncertainty in all models except ANAFORE, with average reductions in the standard deviation of marginal parameter distributions (i.e. for individual parameters) ranging from 1% to 13%. These averages were invariably the result of a majority of parameters being hardly affected by the BC and a small number with strongly reduced uncertainty, with maximum reductions in standard deviation for individual parameters ranging from 6% to 83% across all models (data not shown). The results of country-specific BC were similar but with generally lower reductions in uncertainty.

Figs. 3 and 4 show predictive uncertainty after calibration for mean height and diameter. With respect to output uncertainty, measured as the distance between the 5% and 95% quantiles, the results for country-specific and generic BC were quite similar (Figs. 3 and 4). BC reduced tree height uncertainty in all models, but most in 3PG and BASFOR and least in BRIDGING. For stem diameter, 3PG and BASFOR again saw large uncertainty reductions but otherwise the results differed markedly from those for tree height, with ANAFORE and BRIDGING seeing no clear reductions in predictive uncertainty and FORMIND even becoming worse at B3, E3 and F3.

3.2. Bayesian model comparison before and after calibration

The predictions of the uncalibrated models for the NFI-sites, generated as part of the prior UQ reported in the previous paragraph, were compared against the corresponding NFI-data in a prior Bayesian model comparison (BMC) (Fig. 5). Despite the fact that the data tended to fall between the 5% and 95% quantiles of each model's prior uncertainty ranges (Table 3), the Bayesian model comparison still assigned very different prior probabilities to the different models. The most parameter-rich model, ANAFORE, and the two models initialised at planting, 3PG and BASFOR, had prior probabilities orders of magnitude lower than the other three models. BRIDGING and, to slightly lesser extent, 4C achieved the highest integrated likelihoods (Fig. 5).

The posterior BMC, in which models outputs after calibration were compared with measurements at PSP-sites, showed smaller differences between model probabilities and slightly altered the ranking of the models (Fig. 5). The posterior BMC assigned the highest probability to 4C, followed by BRIDGING and FORMIND with 3PG thereafter.

Similar ranking can be observed in the values of NRMSE (Fig. 6), which like the integrated likelihoods of the models were calculated

Table 3
Prior predictions by six models of final tree height (m) and stem diameter (mm) on twelve sites. Site-codes (A1, A2, etc.) are explained in Table 1. For each combination of model and variable, the first row shows the predictions using the mode of the prior parameter distribution, and the second gives the range (5–95% quantiles). The upper two rows show the measured values for comparison.

Source	Variable	A1	A2	A3	B1	B2	B3	E1	E2	E3	F1	F2	F3
Data	Height	18.5	17.7	18.1	18.4	23.2	21.3	25.0	24.9	25.6	17.8	10.1	21.8
	Diameter	324	207	239	271	293	319	274	237	245	191	146	170
3PG	Height	52.4	21.0	28.4	28.6	28.8	32.8	40.7	32.7	36.0	30.2	23.5	19.5
		21.3–145	10.7–45.0	13.5–62.1	13.1–66.9	13.5–67.6	14.3–82.2	17.7–102	15.4–78.9	16.3–88.5	14.1–68.0	11.5–47.6	9.3–43.6
	Diameter	622	211	303	301	305	356	462	357	400	325	241	194
		337–1476	140–403	195–568	178–607	188–599	201–760	287–960	227–749	248–865	205–646	156–430	110–407
4C	Height	21.6	20.9	20.7	19.6	23.1	24.5	22.5	20.7	21.8	16.7	12.5	26.0
		15.9–29.1	15.6–27.2	14.3–29.9	17.8–25.0	20.0–30.1	19.2–32.6	20.0–29.3	19.0–25.4	21.3–26.0	14.4–22.2	7.6–20.9	10.2–45.3
	Diameter	381	267	284	287	297	352	288	254	244	205	161	340
		291–430	191–298	191–344	267–305	250–322	263–398	243–320	211–271	224–271	170–233	120–201	139–495
ANAFORE	Height	30.2	27.6	28.5	19.4	25.4	46.9	29.0	28.7	24.7	26.7	20.5	48.0
		23.9–59.2	17.4–59.1	18.3–59.2	18.9–23.1	23.3–33.6	31.4–59.0	18.8–52.0	20.5–51.6	18.5–59.2	20.3–49.5	10.0–46.6	22.4–59.3
	Diameter	457	185	330	309	323	457	471	355	376	280	238	219
		335–481	182–195	222–331	299–323	303–344	417–516	277–426	210–326	241–364	245–314	206–436	89–237
BASFOR	Height	25.9	14.6	18.9	22.5	18.9	21.2	18.0	17.9	19.0	16.4	14.6	13.1
		12.6–48.1	1.4–36.2	1.7–40.2	10.8–41.6	1.4–36.9	5.8–39.9	7.8–33.9	7.8–33.4	8.3–35.6	2.5–31.1	2.2–27.9	3.1–24.7
	Diameter	229	98	144	186	144	170	133	132	145	115	97	82
		131–319	3–221	3–261	103–259	3–220	31–244	52–190	49–189	62–208	6–170	4–143	9–119
BRIDGING	Height	18.2	17.5	18.2	19.2	21.8	22.6	22.7	21.4	23.9	17.5	11.5	12.9
		17.5–18.8	17.0–18.1	17.0–19.4	18.9–19.6	21.5–22.2	22.0–23.2	22.1–23.3	20.9–22.0	23.3–24.5	16.6–18.4	10.0–13.0	12.1–16.8
	Diameter	423	261	305	312	331	353	320	271	279	226	210	265
		375–442	229–273	261–321	296–321	302–349	327–363	290–334	245–282	255–289	200–237	175–225	233–388
FORMIND	Height	26.6	21.0	22.1	22.0	20.9	22.1	20.9	18.5	19.8	16.0	11.0	8.0
		16.0–32.4	12.0–26.3	12.5–29.1	14.8–26.4	15.1–26.0	16.0–27.6	14.3–25.9	13.0–22.7	13.5–24.5	11.2–19.6	8.2–13.1	6.3–9.1
	Diameter	352	251	270	268	250	270	250	210	230	170	100	63
		302–362	190–264	201–288	260–273	250–273	270–305	250–251	210–212	230–232	170–170	100–102	56–78

as averages for the whole parameter distribution. For all models except ANAFORE, the values of NRMSE for mean height and diameter were markedly reduced by BC but with little difference between country-specific and generic BC.

3.3. Bayesian model averaging before and after calibration

The weighted average predictions of the models for the PSP-sites, using prior and posterior model probabilities as weights, are included in Figs. 3, 4 and 6. The prior BMA, which was based on model probabilities derived from NFI-data without any model calibration, showed robust out-of-sample predictive capacity for the PSP-sites, as shown by low NRMSE-values for both output variables (Fig. 6). In the case of tree height, only the BRIDGING model had lower NRMSE, whereas for stem diameter only 4C had clearly lower error. Also, predictive uncertainty from the prior BMA was moderate, with at least half of the models showing larger uncertainty ranges for all combinations of variable and site except stem diameter at F3.

Predictions from posterior BMA were also compared against the measurements at PSP-sites (Figs. 3, 4 and 6). In contrast to the tests of prior BMA, and despite the fact that only NFI-data were used in model calibration, these were in-sample tests of predictive capacity because PSP-data had been used to calculate the model probabilities. Prediction using posterior BMA was less of an improvement compared to most individual models than was the case for prior BMA (Figs. 3, 4 and 6).

4. Discussion

4.1. Model performance before and after Bayesian calibration on NFI-data

If forest models are to be useful in management, their predictions must be sufficiently accurate and precise. A quantification of model accuracy for growth is given in Table 3, where the predictions for the modes of prior parameter distributions can be compared against measurements. The same table also provides information about predictive uncertainty, in the form of the 5% and 95% quantiles of model predictions. The results show that only the BRIDGING model had high *a priori* predictive accuracy for mean tree height with low accompanying uncertainty at all sites except F3. For stem diameter, none of the uncalibrated models was very precise – BRIDGING, 4C and FORMIND did best – and only BRIDGING and FORMIND had low uncertainties throughout. The balance of accuracy and precision for the NFI-sites was such that the prior Bayesian model comparison assigned 55% prior probability to BRIDGING and 42% to 4C.

One reason for the prior success of BRIDGING and 4C, and to lesser extent FORMIND, was that these models were initialised for each site at the first date of measurement. The models were thus started off with values of mean tree height and stem diameter correct for the site, and with fewer years of growth remaining to be predicted than what was asked from models initialised at planting, such as 3PG and BASFOR. The advantage of late model initialisation – having less time to deviate from true on-site growth patterns – apparently weighed heavier than that of 3PG and BASFOR being able to process more detailed information about the site conditions. Furthermore, information about the early management history of sites, such as the tree thinning regime, tends to be less reliable than information for the measurement periods. Late initialisation, however, does not always improve predictive performance, as demonstrated by the results for ANAFORE. In the case of ANAFORE, a highly detailed model, there was a large suite of other state variables besides mean height and diameter that needed to be initialised, and for which no good information was available for most sites so default model settings could not be adjusted. While some models may be designed to run with stand-level information such as typically provided by NFIs, other models may perform better if more detailed initialisation data

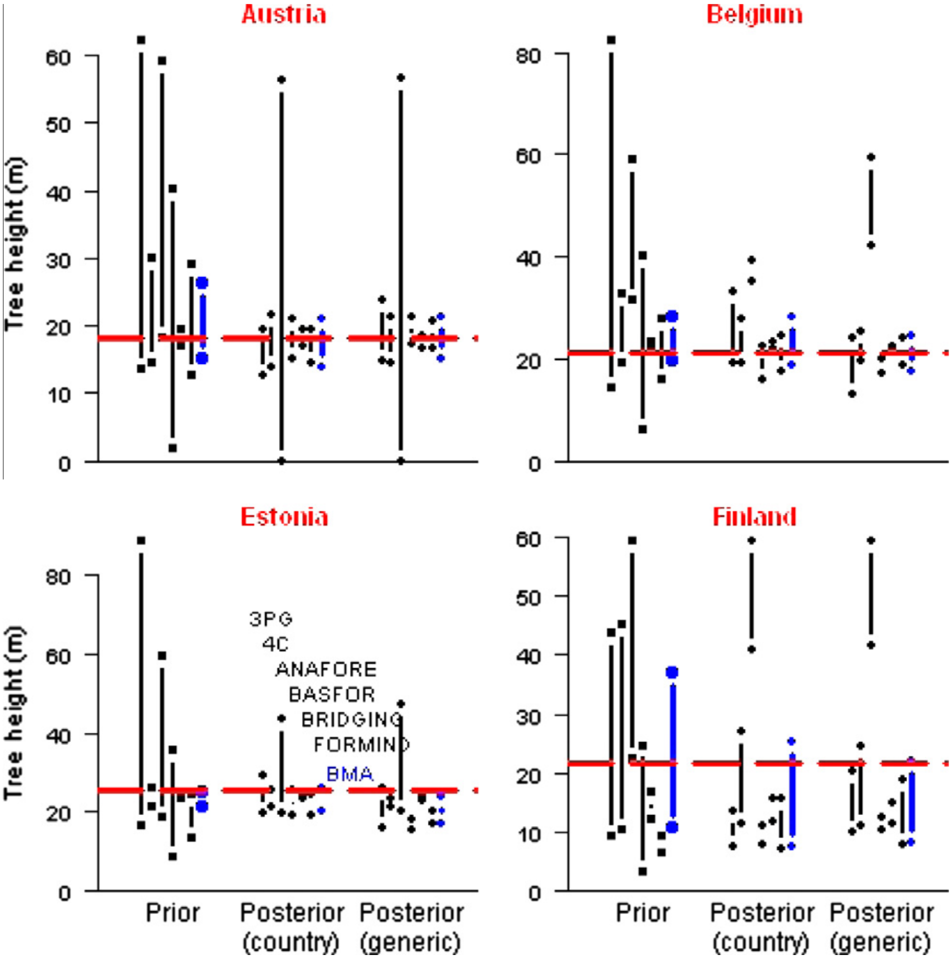


Fig. 3. Model output uncertainty for final mean tree height at the PSP-sites A3, B3, E3 and F3. Vertical bars show the central 90% of distributions. For each country, the three clusters of bars show prior and posterior (country-specific, generic) predictions. The seven bars in each cluster are for the six models plus the Bayesian Model Averaging result, in the order indicated in the bottom-left panel. The dashed horizontal lines indicate observed values, which were not used for model calibration.

are available. In this study, the most complex model, ANAFORE was clearly overparameterised in relation to the very limited data. We also note that BRIDGING and 4C might have been rated best if initialisation values would have been estimated rather than being set *a priori* – but that was not investigated in this study.

These comparisons of the prior performance of the different models were inevitably also affected by how the prior parameter distributions were defined. Different methods for quantifying prior parameter distribution of a process-based forest model, PnET-II, were discussed by Radtke et al. (2001). The prior distributions in our study were set independently by each modelling group, using the information available to them from literature and from previous experience with their model. This partly explains why some models, such as 3PG, showed wider prior output ranges than other models.

To restrict the influence of subjective prior parameterisation, it is therefore important to compare differences in model performance after all models have been calibrated for the tree species under study. Both country-specific and generic Bayesian calibration on NFI-data markedly increased the accuracy and precision of prediction for the PSP-sites by all models except the most complex and parameter-rich model, ANAFORE (Figs. 3 and 4). After these general improvements, the 4C model performed best (Fig. 5), but note that the differences in model initialisation method again affected the results, and that the strength of the data was probably still not sufficient to completely overrule the effect of prior choice after calibration. Also note that the assessments of model performance and plausibility in this study are restricted to predictions for mean tree height and stem diameter. If data from other variables, such as above-and belowground biomass and wood quality, had been used, model evaluation would likely have yielded different results.

4.2. Spatial differences in model performance

All models had the poorest predictions of mean tree height for the Finnish PSP-site. That site, F3, had an atypically high stem number (Table 1), which may have contributed to comparatively strong height growth at relatively small diameter despite advanced age (Fig. 2). Most models apparently struggled to simulate this growth pattern, irrespective of model complexity. The problems with this site largely persisted after calibration.

Sites within a single country are likely to be more similar in tree provenance, soil type and climate than sites in different parts of Europe. Therefore, the performance of models at a given PSP-site was expected to be best after calibration exclusively on the two NFI-sites from the same country, as opposed to model performance after generic calibration on all NFI-sites. However, the two types of calibration led to predictions of similar integrated likelihood and NRMSE (Figs. 5 and 6). It should be noted that this somewhat surprising result is partly explained by the fact that we had fewer data available per country, so the likely greater relevance of data used in within-country calibration was offset by the low weight of evidence from using data from 2 NFI-sites as compared to 8 in generic

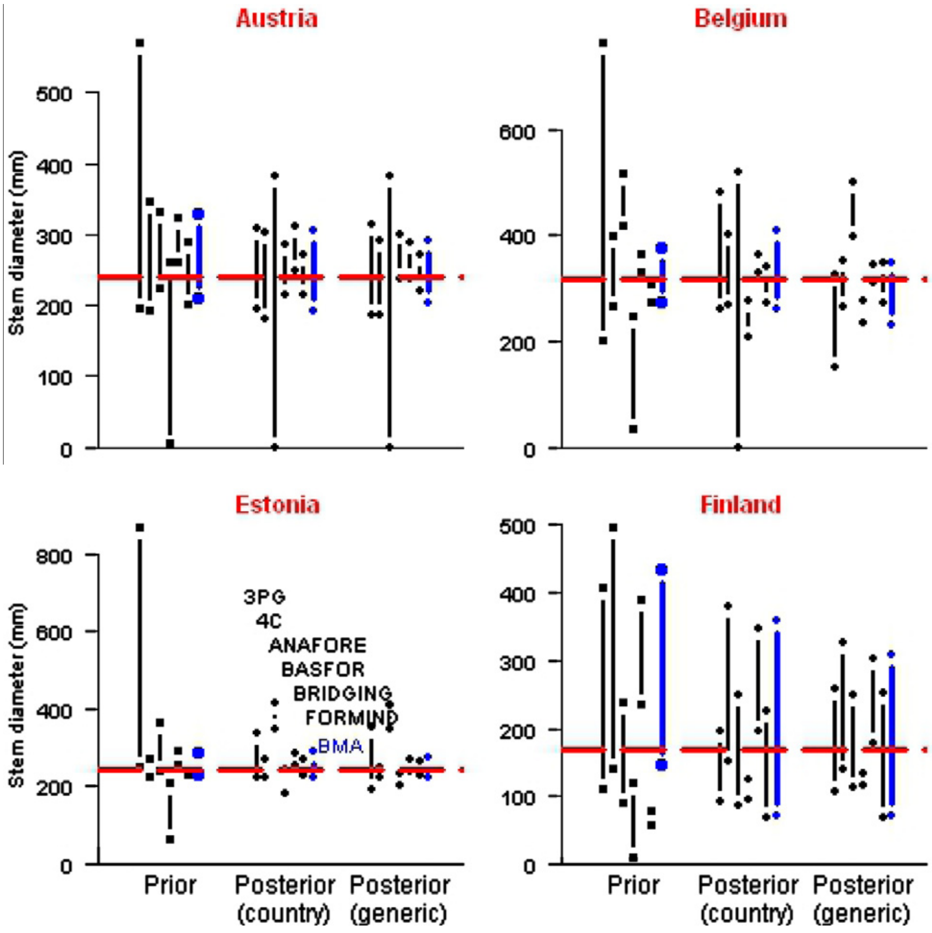


Fig. 4. Model output uncertainty for final mean stem diameter at the PSP-sites A3, B3, E3 and F3. The lay-out of the figure is the same as for Fig. 3.

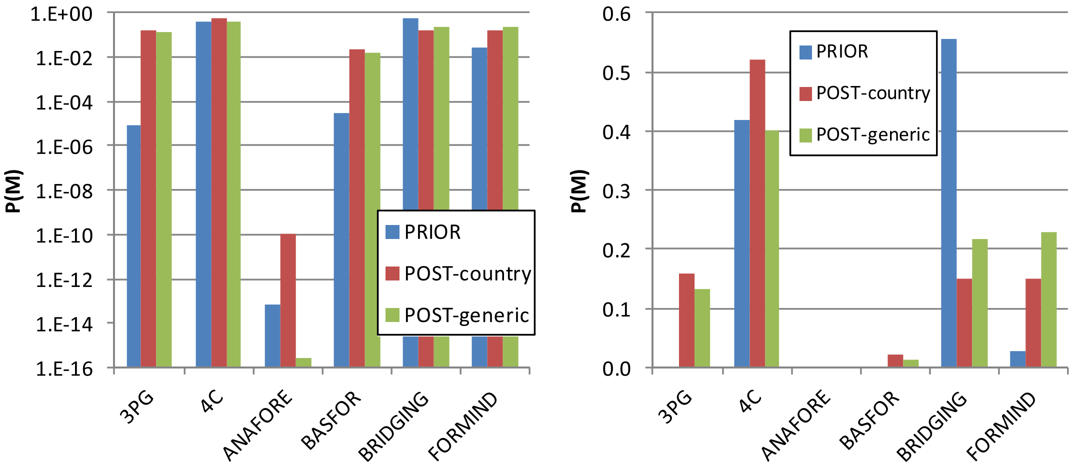


Fig. 5. Prior and posterior model probabilities, derived from the integrated likelihoods of NFI and PSP-measurements. Left: logarithmic scale; right: absolute scale.

BC. Still, it can be conjectured that the considered models are sufficiently general to provide a useful generic parameterisation for Scots pine in Europe, although a future study with larger numbers of NFI-sites per country would be needed to test this hypothesis rigorously. The extra sites should be chosen to cover spatial variation in tree genotypes and geographical conditions. Such increased spatial coverage would also be needed if we want to move from assessing model predictive capacity at site-level to country-wide upscaling.

4.3. Quantifying and reducing uncertainties

The extent to which Bayesian calibration can reduce parameter uncertainties of a model depends both on the structure of the model and on the prior distribution assigned by the modeller. In the present study, Bayesian calibration reduced parameter and output uncertainty of all models except the parameter-richest one, ANAFORE. Likewise, the Bayesian model comparison was able to identify which models were most plausible by calculating the

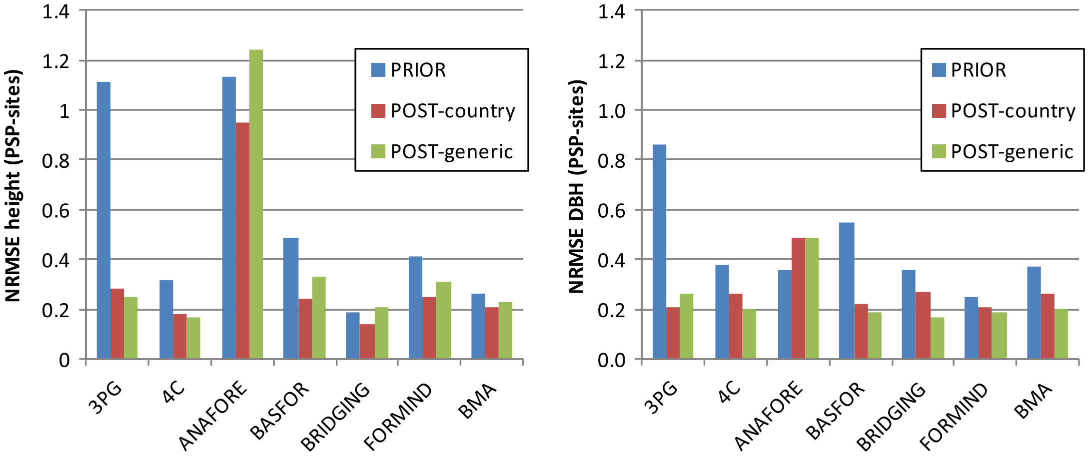


Fig. 6. Normalised RMSE, derived from simulations at PSP-sites using samples from prior and posterior parameter distributions. Left: tree height, right: diameter at breast height. The rightmost three bars in both panels are the result of Bayesian Model Averaging (BMA).

integrated likelihood for each model at different stages in the study. The integrated likelihood accounts for parameter uncertainty (by integrating over its distribution) and is a natural way of combining diverse measurements in one model comparison criterion. This is in contrast to the commonly used NRMSE, which has to be calculated for every variable separately. Another potential advantage of the integrated likelihood over other measures, such as NRMSE and squared correlation coefficient, r^2, is that the integrated likelihood can account for different levels of uncertainty about measurement error for different data points. However, that did not play a role in the present study because all height and diameter data were assumed to have the same degree of uncertainty.

4.4. Impact of the choices of prior distribution

As discussed in Sections 4.2–4.4, the choices made to set the prior probability distributions for the parameters of the different models affected our results to some degree, in particular in the early stages of the analysis where the prior predictive performance of the models was quantified and compared. Because prior distributions for structurally different models cannot be set in a standardised way, and were based on the expertise of the responsible modellers, this introduced a subjective element in the study. This included model-specific choices about parameter-screening, i.e. which of a model's parameters to include in the Bayesian calibration. This subjectivity concerning the prior parameter distribution is unavoidable, to some extent, in any application of Bayesian methodology. However, the procedure we applied here, where all models were calibrated on the same data (NFI) and were subsequently compared against the same independent data (PSP) removed much of the effect of the choice of prior (Figs. 3 and 4). We therefore suggest that Bayesian model comparisons are most useful after such standardisation.

4.5. On the use of multiple models

The use of BMC is formally conditional on one of the models being 'correct' – which is never truly the case in environmental modelling – so we should use the results from the BMC as a guide towards finding the most plausible model in the set of six rather than as formal model probabilities. The results suggest that the 4C model should be recommended as the model of choice for a forest manager who wants to select a single model to help estimate future productivity out of the six models in this study. We believe that for the forest scientist the results are less clear-cut because the Bayesian probabilities do not by themselves explain what makes one model structure more plausible than another. The Bayesian model comparison largely treats the models as black boxes characterised by their input–output relationships. In a previous Bayesian forest model comparison (Van Oijen et al., 2011) it was therefore recommended that after the BC of all models, and their BMC, a detailed analysis should be carried out of the model-data mismatch remaining after calibration. It was recommended in particular to decompose likelihoods into terms for individual output variables and to decompose mean squared errors (MSEs) into terms for bias, variance mismatch and phase-shift (Kobayashi and Salam, 2000). However, in our study with only two output variables and extremely short time-series, these decompositions are not informative. To allow such detailed study of model-data mismatch – and therefore to help explain the results presented here – we would need more detailed data sets, e.g. long time-series of annual data.

Another natural follow-up to BMC, and one that was carried out in this study, is calculating forecasts using Bayesian model averaging (BMA, e.g. Kass and Raftery, 1995). In BMA, no single model is selected for making predictions; instead the probability distributions for the individual model predictions are averaged using as weights the model probabilities determined by the BMC. Because BMA integrates parameter and model structural uncertainty, it is less prone to underestimation of predictive uncertainty than the common practice of selecting and using only a single 'best' model. In the present study, the out-of-sample predictive capacity of BMA was very good, as shown by the NRMSE-values for both output variables in the prior BMA. This is not exceptional; BMA has been reported to have higher forecasting skill than each individual model in other fields, such as medical prognosis (Hoeting et al., 1999) and climate prediction (Min and Hense, 2006). We found that the predictive performance of posterior BMA was only average. However, this was a partly within-sample test - with model probabilities (but not parameters) informed by the PSP-data - so this should be repeated with independent data.

5. Conclusions

- Bayesian calibration successfully reduced uncertainties in parameters and predictions of five out of six forest models.
- Calibrating models separately for each country did not clearly improve within-country predictive capacity compared to generic calibration. This might change when more data become available per country.

- Bayesian model comparison using NFI- and PSP-data identified the 4C model, which is of moderate complexity but mechanistic, as the most plausible forest model after calibration.
- The main caveat to the results is the issue of model initialisation: how it is carried out and which data are available for it. This study suggests that models are favoured that are initialised using on-site measurements of tree growth, unless model complexity requires more data for such initialisation than are available. But model ranking might have been different if more data, or data from other variables than mean tree height and stem diameter, would have been available for use.
- For a detailed analysis of model-data mismatch, NFI-data are insufficient, but information from PSPs not used in this study, such as single tree data, could be used.
- BMA afforded good out-of-sample forecasts of forest productivity and may be a promising tool for forest management, of sufficient accuracy and precision whilst not underestimating uncertainties.

Acknowledgements

We thank the EU for support of all participants through COST Action FP603 and for support of M.v.O. in IP Carbo-Extreme (FP7, GA 226701). We also thank the national forestry services in Austria, Belgium and Finland for providing the NFI- and PSP-data. The Estonian Meteorological and Hydrological Institute provided climate data and the Estonian Environment Information Centre provided soil data. F.H. acknowledges support from ERC Advanced Grant 233066.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.foreco.2012.09.043.

References

Andren, O., Katterer, T., 1997. ICBM: the introductory carbon balance model for exploration of soil carbon balances. Ecological Applications 7 (4), 1226–1236.

Baeyens, L., van Slycken, J., Stevens, D., 1993. Description of the Soil Profile in Brasschaat. Institute of Forestry and Game Management, Geraardsbergen, Belgium.

Bugmann, H., Grote, R., Lasch, P., Lindner, M., Suckow, F., 1997. A new forest gap model to study the effects of environmental change on forest structure and functioning, impacts of global change on tree physiology and forest ecosystems. In: Forestry Sciences. Springer, Dordrecht, pp. 255–261.

Curiel Yuste, J., Konopka, B., Janssens, I.A., Coenen, K., Xiao, C.W., Ceulemans, R., 2005. Contrasting net primary productivity and carbon distribution between neighboring stands of *Quercus robur* and *Pinus sylvestris*. Tree Physiology 25 (6), 701–712.

Davidson, R.L., 1969. Effect of root/leaf temperature differentials on root/shoot ratios in some pasture grasses and clover. Annals of Botany 33, 561–569.

Deckmyn, G., Verbeeck, H., Op de Beeck, M., Vansteenkiste, D., Steppe, K., Ceulemans, R., 2008. ANAFORE: a stand-scale process-based forest model that includes wood tissue development and labile carbon storage in trees. Ecological Modelling 215 (4), 345–368.

Deckmyn, G., Mali, B., Kraigher, H., Torelli, N., Op de Beeck, M., Ceulemans, R., 2009. Using the process-based stand model ANAFORE including Bayesian optimisation to predict wood quality and quantity and their uncertainty in Slovenian Beech. Silva Fennica 43 (3), 523–534.

Duursma, R.A., Makela, A., 2007. Summary models for light interception and light-use efficiency of non-homogeneous canopies. Tree Physiology 27 (6), 859–870.

Fontes, L., Bontemps, J.D., Bugmann, H., Van Oijen, M., Gracia, C., Kramer, K., Lindner, M., Rotzer, T., Skovsgaard, J.P., 2010. Models for supporting forest management in a changing environment. Forest Systems 19, 8–29.

Fu, Y.H., Campioli, M., Van Oijen, M., Deckmyn, G., Janssens, I., 2012. Bayesian comparison of six different temperature-based budburst models for four temperate tree species. Ecological Modelling 230, 92–100.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7, 457–511.

Green, E.J., MacFarlane, D.W., Valentine, H.T., Strawderman, W.E., 1999. Assessing uncertainty in a stand growth model by Bayesian synthesis. Forest Science 45 (4), 528–538.

Gutiérrez, A.G., Huth, A., 2012. Successional stages of primary temperate rainforests of Chiloé Island, Chile. Perspectives in Plant Ecology, Evolution and Systematics 14, 243–256.

Härkönen, S., Pulkkinen, M., Duursma, R., Mäkelä, A., 2010. Estimating annual GPP, NPP and stem growth in Finland using summary models. Forest Ecology and Management 259 (3), 524–533.

Hartig, F., Dyke, J., Hickler, T., Higgins, S., O'Hara, R.B., Scheiter, S. and Huth, A., 2012. Connecting dynamic vegetation models to data – an inverse perspective. Journal of Biogeography. doi.org/10.1111/j.1365-2699.2012.02745.x.

Haxeltine, A., Prentice, I.C., 1996. A general model for the light-use efficiency of primary production. Functional Ecology 10, 551–561.

Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: a tutorial. Statistical Science 14 (4), 382–417.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. Journal of the American Statistical Association 90 (430), 773–795.

Kobayashi, K., Salam, M.U., 2000. Comparing simulated and measured values using mean squared deviation and its components. Agronomy Journal 92 (2), 345–352.

Köhler, P., Huth, A., 1998. The effects of tree species grouping in tropical rainforest modelling: simulations with the individual-based model FORMIND. Ecological Modelling 109 (3), 301–321.

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.

Landsberg, J., Makela, A., Sievanen, R., Kukkola, M., 2005. Analysis of biomass accumulation and stem size distributions over long periods in managed stands of *Pinus sylvestris* in Finland using the 3-PG model. Tree Physiology 25 (7), 781–792.

Lasch, P., Badeck, F.W., Suckow, F., Lindner, M., Mohr, P., 2005. Model-based analysis of management alternatives at stand and regional level in Brandenburg (Germany). Forest Ecology and Management 207 (1–2), 59–74.

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.

Lehuger, S., Gabrielle, B., Van Oijen, M., Makowski, D., Germon, J.C., Morvan, T., Hénault, C., 2009. Bayesian calibration of the nitrous oxide emission module of an agro-ecosystem model. Agriculture, Ecosystems & Environment 133 (3–4), 208–222.

Levy, P.E., Wendler, R., Van Oijen, M., Cannell, M.G.R., Millard, P., 2004. The effects of nitrogen enrichment on the carbon sink in coniferous forests: uncertainty and sensitivity analyses of three ecosystem models. Water, Air and Soil Pollution: Focus 4, 67–74.

Luo, Y., Weng, E., Wu, X., Gao, C., Zhou, X., Zhang, L., 2009. Parameter identifiability, constraint, and equifinality in data assimilation with ecosystem models. Ecological Applications 19 (3), 571–574.

Mäkelä, A., 1997. A carbon balance model of growth and self-pruning in trees based on structural relationships. Forest Science 43, 7–24.

Mäkelä, A., Sievanen, R., 1992. Height growth strategies in open-grown trees. Journal of Theoretical Biology 159 (4), 443–467.

Mäkelä, A., Vanninen, P., 2001. Vertical structure of Scots pine crowns in different age and size classes. Trees 15, 385–392.

Mäkelä, A., Pulkkinen, M., Kolari, P., Lagergren, F., Berbigier, P., Lindroth, A., Loustau, D., Nikinmaa, E., Vesala, T., Hari, P., 2008. Developing an empirical model of stand GPP with the LUE approach: analysis of eddy covariance data at five contrasting conifer sites in Europe. Global Change Biology 14 (1), 92–108.

Mäkelä, A., del Río, M., Hynynen, J., Hawkins, M.J., Reyer, C., Soares, P., Van Oijen, M., Tomé, M., 2012. Using forest growth models for estimating indicators of sustainable forest management. Forest Ecology and Management 285, 164–178.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. Journal of Chemical Physics 21, 1087–1092.

Min, S.-K., Hense, A., 2006. A Bayesian approach to climate model evaluation and multi-model averaging with an application to global mean surface temperatures from IPCC AR4 coupled climate models. Geophysical Research Letters 33 (8), L08708.

Palmroth, S., Berninger, F., Nikinmaa, E., Lloyd, J., Pulkkinen, P., Hari, P., 1999. Structural adaptation rather than water conservation was observed in Scots pine over a range of wet to dry climates. Oecologia 121, 302–309.

Radtke, P.J., Burk, T.E., Bolstad, P.V., 2001. Estimates of the distributions of forest ecosystem model inputs for deciduous forests of eastern North America. Tree Physiology 21 (8), 505–512.

Reinds, G.J., van Oijen, M., Heuvelink, G.B.M., Kros, H., 2008. Bayesian calibration of the VSD soil acidification model using European forest monitoring data. Geoderma 146 (3–4), 475–488.

Reyer, C., Lasch, P., Mohren, G.M.J., Sterck, F.J., 2010. Inter-specific competition in mixed forests of Douglas fir (*Pseudotsuga menziesii*) and common beech (*Fagus sylvatica*) under climate change – a model-based analysis. Annals of Forest Science 67 (8), 11.

Ruger, N., Gutierrez, A.G., Kissling, W.D., Armesto, J.J., Huth, A., 2007. Ecological impacts of different harvesting scenarios for temperate evergreen rain forest in southern Chile – a simulation experiment. Forest Ecology and Management 252 (1–3), 52–66.

Sands, P.J., Landsberg, J.J., 2002. Parameterisation of 3-PG for plantation grown *Eucalyptus globulus*. Forest Ecology and Management 163 (1–3), 273–292.

Seidl, R., Rammer, W., Lasch, P., Badeck, F.W., Lexer, M.J., 2008. Does conversion of even-aged, secondary coniferous forests affect carbon sequestration? A simulation study under changing environmental conditions. Silva Fennica 42 (3), 369–386.

Shinozaki, K., Yoda, K., Hozumi, K., Kira, T., 1964. A quantitative analysis of plant form – the pipe model theory. I. Basic analysis. Japanese Journal of Ecology 14, 97–105.

Sims, A., Kiviste, A., Hordo, M., Laarmann, D., von Gadow, K., 2009. Estimating tree survival: a study based on the Estonian forest research plots network. Annales Botanici Fennici 46 (4), 336–352.

Suckow, F., Badeck, F.W., Lasch, P., Schaber, J., 2001. Nutzung von Level-II-Beobachtungen für Test und Anwendungen des Sukzessionsmodells FORESEE. Beiträge für Forstwirtschaft und Landschaftsökologie 35, 84–87.

Tuomi, M., Vanhala, P., Karhu, K., Fritze, H., Liski, J., 2008. Heterotrophic soil respiration – comparison of different models describing its temperature dependence. Ecological Modelling 211 (1–2), 182–190.

Valentine, H.T., Mäkelä, A., 2005. Bridging process-based and empirical approaches to modeling tree growth. Tree Physiology 25 (7), 769–779.

Van Oijen, M., Thomson, A., 2010. Toward Bayesian uncertainty quantification for forestry models used in the United Kingdom greenhouse gas inventory for land use, land use change, and forestry. Climatic Change 103 (1), 55–67.

Van Oijen, M., Rougier, J., Smith, R., 2005. Bayesian calibration of process-based forest models: bridging the gap between models and data. Tree Physiology 25 (7), 915–927.

Van Oijen, M., Dauzat, J., Harmand, J.-M., Lawson, G., Vaast, P., 2010. Coffee agroforestry systems in Central America. II. Development of a simple process-based model and preliminary results. Agroforestry Systems 80 (3), 361–378.

Van Oijen, M., Cameron, D.R., Butterbach-Bahl, K., Farahbakhshazad, N., Jansson, P.E., Kiese, R., Rahn, K.H., Werner, C., Yeluripati, J.B., 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.

Vanninen, P., Mäkelä, A., 2005. Carbon budget for Scots pine trees: effect of size, competition and site fertility on growth allocation and production. Tree Physiology 25, 17–30.

Waring, R.H., Landsberg, J.J., Williams, M., 1998. Net primary production of forests: a constant fraction of gross primary production? Tree Physiology 18 (2), 129–134.

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.

CHAPTER 5

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

SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. 1, pp. 370–385

Selecting Parameters for Bayesian Calibration of a Process-Based Model: A Methodology Based on Canonical Correlation Analysis[*]

F. Minunno[†], M. van Oijen[‡], D. R. Cameron[‡], and J. S. Pereira[†]

Abstract. Bayesian statistics is becoming increasingly common in the environmental sciences because of developments in computers and sampling-based techniques for parameter estimation. However, the use of the Bayesian approach is still limited in forest research, especially for models with many parameters. Some studies have used parameter screening to make the calibration of a computationally expensive model possible. In this paper we introduce a new methodology for parameter screening, based on canonical correlation analysis. Furthermore we show how parameter screening impacts the performance of a process-based model. The methodology presented here can be generally applied and is particularly suitable for complex process-based models because it is not computationally demanding and is easy to implement. It provides an overall ranking in relation to all outputs of the model, as opposed to common sensitivity methods that analyze one model output variable at a time. We found that parameter screening can be used to reduce the computational load of Bayesian calibration, but only the least important parameters should be excluded from the calibration if we do not want to affect model performance. In this exercise, 25% of the parameters of a process-based forest model could be excluded from the calibration without affecting model performance. When calibration was limited to a more restricted number of parameters, model performance significantly deteriorated.

Key words. sensitivity analysis, parameter screening, canonical correlation analysis, Bayesian calibration

AMS subject classifications. 62F07, 62F15

DOI. 10.1137/120891344

1. Introduction.

In recent decades, forests have started to experience significant changes in environmental conditions [10]. Ecosystems have to adapt to variations in the mean climatic variables, but also to the increased risk of extreme events such as drought, heat waves, storms, late frost, and flooding [19]. 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 [15]. The reliability and robustness of models are conditional on their structure but also on their parameterization [31, 20]. Hence the calibration of model

---

[*] Received by the editors September 13, 2012; accepted for publication (in revised form) July 3, 2013; published electronically October 24, 2013. This research was supported by the EU through COST Action FP603.
http://www.siam.org/journals/juq/1/89134.html

[†] Faculty of Agricultural Sciences, Forest Research Center, Technical University of Lisbon, Tapada da Ajuda, Lisbon 1349-017, Portugal (francescom@isa.utl.pt, jspereira@isa.utl.pt). The work of the first author was partially supported by the EU through a short-term scientific mission to CEH, and by the FCT-Portugal through a Ph.D. fellowship (SFRH/BD/65934/2009).

[‡] CEH-Edinburgh, Bush Estate, Penicuik EH26 0QB, UK (mvano@ceh.ac.uk, dcam@ceh.ac.uk).

parameters is a key stage in the model building process.

Bayesian calibration (BC), based on probability theory, is a logical choice for model calibration [33]. 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. Today 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 [33]. 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 [2]. Second, the number of parameters involved in the calibration can be reduced by means of parameter screening. Of these two options, screening is the simpler and more straightforward. Parameter screening aims to rank model parameters according to their impact on the output [27]. 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 [32, 31, 34]; however, it has not been investigated whether 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 characterizing the behavior of the model. The Fourier amplitude sensitivity test method [7], the Sobol' [30] method, and Saltelli's method [26] are common GSAs, all of which can be computationally demanding for parameter-rich models. The Morris method [5, 21] is a GSA that has already been used to screen parameters of a process-based forest model [31]. Canonical correlation analysis (CCA) is a technique that can be used for GSA [13]; 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 the following:

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 CCA.

The model used for this exercise was 3PGN [34], and the analyses were carried out using a comprehensive dataset of *Eucalyptus globulus* plantations in Portugal.

2. Materials and methods.

2.1. 3PGN structure. 3PGN was developed by Xenakis, Ray, and Mencuccini [34], who coupled 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 the 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 vapor 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 by Landsberg and Waring [18] and by Sands and Landsberg [28].

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 roots, 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 [1] and by Kätterer and Andrén [17, 16].

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.

2.2. 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 (38°38′N, 8°36′W) tended as a coppice [23]. The mean annual temperature for the site is 16°C, 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 [4, 3]. Flux data quality control followed the CarboEurope-IP recommendations; gap filling and partitioning of P_E to gross primary production and ecosystem respiration were performed according to Reichstein et al. [24]. To allow model calibrations the P_E data were averaged monthly.

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 an *E. globulus* plantation at Quinta do Furadouro (Óbidos, Portugal, 39°29′N, 9°13′W) from 1986 to 1992. The mean annual temperature was 15.2 °C, and the mean annual precipitation was 607 mm, of which less than 10% occured 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), and 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

Data types, number of measurements, and coefficient of variation used for the calibration of 3PGN for E. 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

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

2.3. 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. CCA was introduced by Hotelling in 1936 [14]. 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. [13], 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 variables 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 by 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. [13] 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.

The advantages of using CCA for model sensitivity analysis are that CCA can calculate the relationships between multiple sets of variables, parameters, and outputs, and that 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 nonlinearity in the model is not considered, although higher order nonlinear terms can be included. CCA finds the linear combinations of terms, but the terms themselves can be appropriately transformed in order to convert nonlinear relationships to linear forms.

CCA 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 CCA results were not only specific 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.

2.4. Bayesian calibrations. The calibration of 3PGN, using the Espirra and Furadouro data, was carried out by means of the Bayesian method. BC 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 ($L(\theta)$), 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 (see [29]) proposed the likelihood function of (2.1) that gives less weight to outliers because of its slowly decaying Cauchy-like tails:

$$L(\theta) = \prod_{i=1}^{N} \frac{1}{\sigma_i \sqrt{2\pi}} \frac{1 - \exp(-R_i^2/2)}{R_i^2}, \tag{2.1}$$

where N is the number of data points, π is the uncertainty about the random error of the ith data point, and $R_i = (\mathrm{sim}(\theta)$ - $O_i)/\pi_i$ ($\mathrm{sim}(\theta)$ is the output from the model for the parameter vector θ, and O are the observed data).

The likelihood proposed by Sivia was used here because outliers can occur in eddy covariance measurements. A uniform prior was assigned to all 3PGN parameters; the parameter

Table 2

Symbols, units, and minimum and maximum values for the 3PGN parameters calibrated for E. globulus *in Portugal.*

Parameter description	Symbols	Units	Min	Max
Constant in aboveground biomass vs. height rel.	aH	-	1.9	2.8
Canopy quantum efficiency	α	mol C * MJ^{-1}	0.04	0.08
Canopy boundary layer conductance	BLcond	m*s^{-1}	0.16	0.24
Power in aboveground biomass vs. height rel.	bW	-	0	0.3
Stomatal response to vapor pressure deficit (VPD)	CoeffCond	Mbar^{-1}	0.04	0.06
Wood density	ρ	Mg*m^{3}	0.36	0.54
Conversion 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 coeff. 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
Leaf area index (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
Max. 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 at D = 2 cm	pFS2	-	0.8	1.2
Foliage-stem partitioning ratio at D = 20 cm	pFS20	-	0.12	0.18
Maximum fraction of net primary production (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
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 aboveground biomass vs. diameter rel.	StemConst	-	1.15	1.4
Power in aboveground biomass vs. diameter rel.	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	°C	32	48
Minimum temperature for growth	Tmin	°C	6.8	10.2
Optimum temperature for growth	Topt	°C	12.8	19.2
Age at which specific leaf area = (SLA0 + SLA1)/2	tSLA	years	1.2	2
Ratio NPP/GPP (gross primary production)	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
Initial young labile carbon pool	Yl_C_i	Mg C ha^{-1}	8	12
Initial young refractory carbon pool	Yr_C_i	Mg C ha^{-1}	0	3
Initial old carbon pool	O_C_i	Mg C ha^{-1}	30	50

bounds were the same as those reported in Minunno et al. [20] and are shown in Table 2.

The BC was carried out by means of MCMC sampling, using the Metropolis–Hastings random walk. A complete description of the algorithm, in the context of forest modeling, is given by Van Oijen, Rougier, and Smith [33]. For the BC, three chains of 500000 iterations were computed. Convergence of iterative simulations was assessed through the Gelman–Rubin test [11]. After convergence, the three chains were joined together and treated as a single sample from the posterior distribution.

Four BCs 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 [20, 9]. For the calibration of the parameter subsets, average values between parameter minimum and maximum were assigned to the parameters not involved in the calibrations.

3. Results.

3.1. 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 columns 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 (last column of Table 3) were parameters related to age stress, initial soil carbon content, and decomposition rates of humified organic matter.

3.2. 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 we tested whether 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.

Table 3

List of parameters ranked on the basis of CCA. The highest cross-loadings (HCL) of each parameter with the output canonical variates are shown.

p25%			p50%		
Ranking	Parameter	HCL	Ranking	Parameter	HCL
1	ρ	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	α	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	Rttover	0.448	24	tBB	0.142

p75%			p100%		
Ranking	Parameter	HCL	Ranking	Parameter	HCL
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

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, α, 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 sets 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 α interacted with many others in p25%, while in the other calibrations, α 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.

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

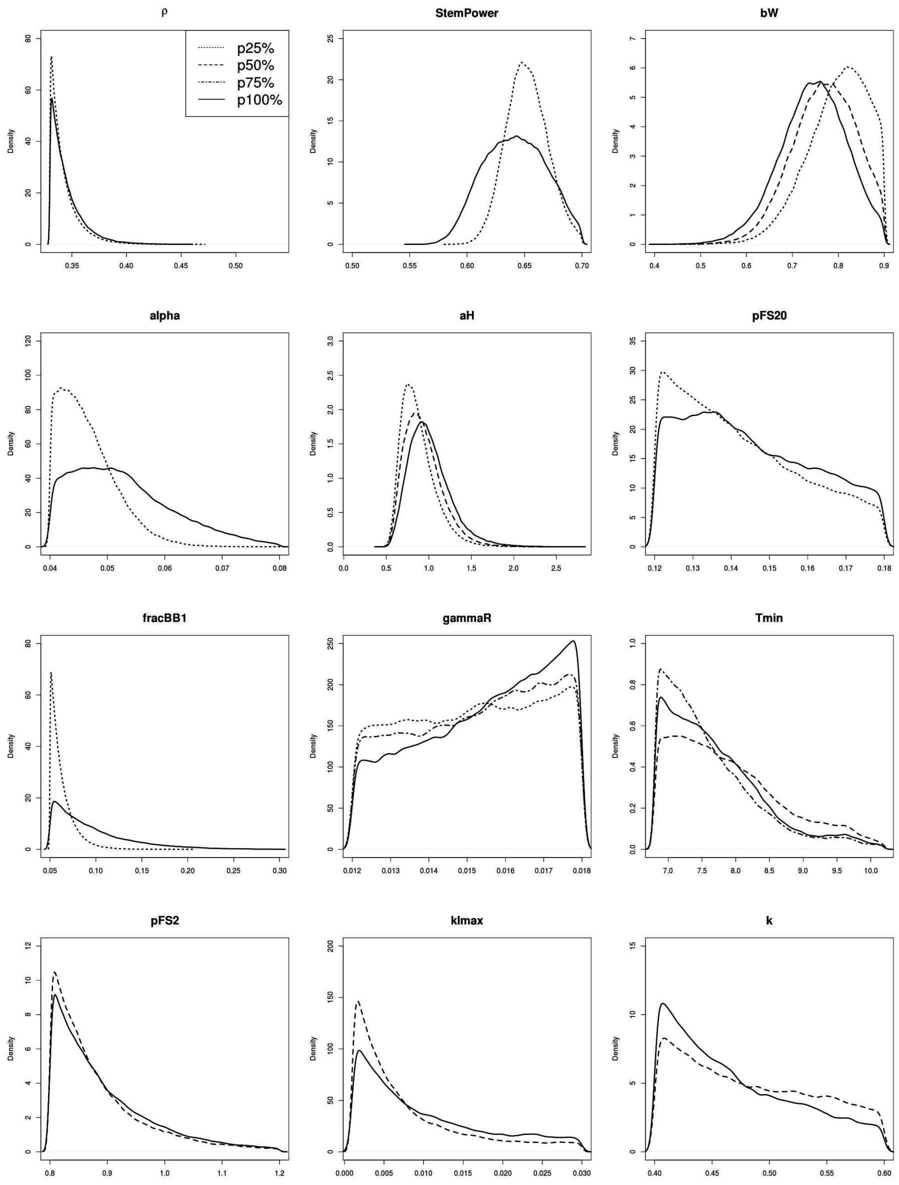


Figure 1. *Marginal posterior distributions of* 3*PGN parameters for the four calibrations carried out. Only those parameters for which the cumulative distribution functions had a Kolmogorov–Smirnov distance of at least* 0.05 *from* p100% *are shown.*

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%,

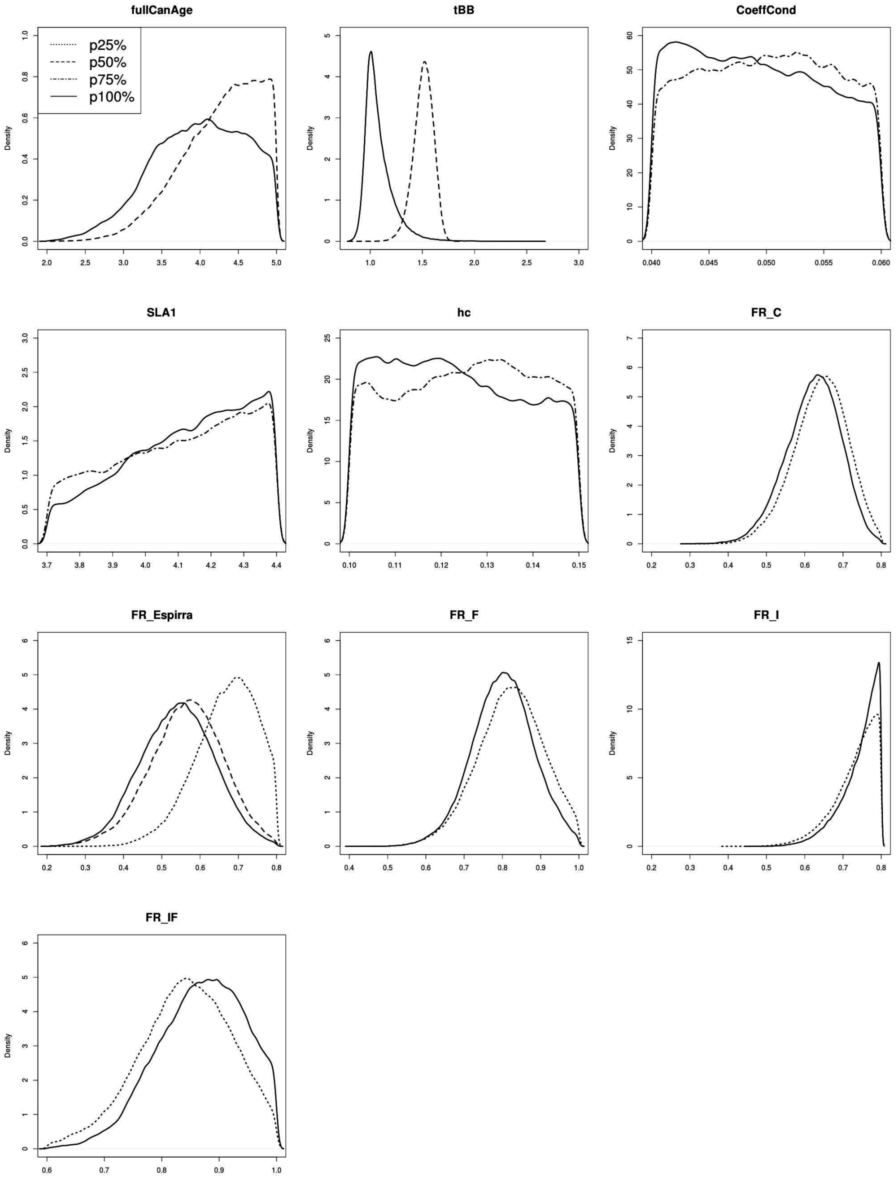


Figure 1. *(cont.)*

and p100%, giving an idea of how model fit changes for the different calibrations; in other words, Figure 2 shows to what 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

Table 4

Pearson product-moment correlation coefficients (r) of 3PGN parameters higher than 0.24 or lower than −0.24, for p100%, p75%, p50%, and p25%.

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, α	−0.53	Y, α	−0.53
fN0, FR_Espirra	−0.37	fN0, FR_Espirra	−0.39
fN0, FR_C	−0.29	fN0, FR_C	−0.3
α, FR_F	−0.28	α, FR_F	−0.27
fCanAge, dmC	−0.26	α, K	−0.26
α, K	−0.24	fCanAge, FR_Espirra	−0.25
Topt, FR_Espirra	−0.27		

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

parameter set had similar distributions (Figure 2). In order to provide a more direct measure of model performance, the normalized root mean squared errors (NRMSEs) 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 PE, 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 last column of Table 3, the 35th parameter (hc) was also not influential on the likelihood and the NRMSEs (data not shown).

4. Discussion.

4.1. The impact of parameter screening on Bayesian calibration.

During the last decade, the use of Bayesian statistics has increased substantially in biological science [32, 6, 22]. However, BC of parameter-rich models, such as 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 [31, 34] is one potential solution for BC of complex models. For the first time, we investigated whether this

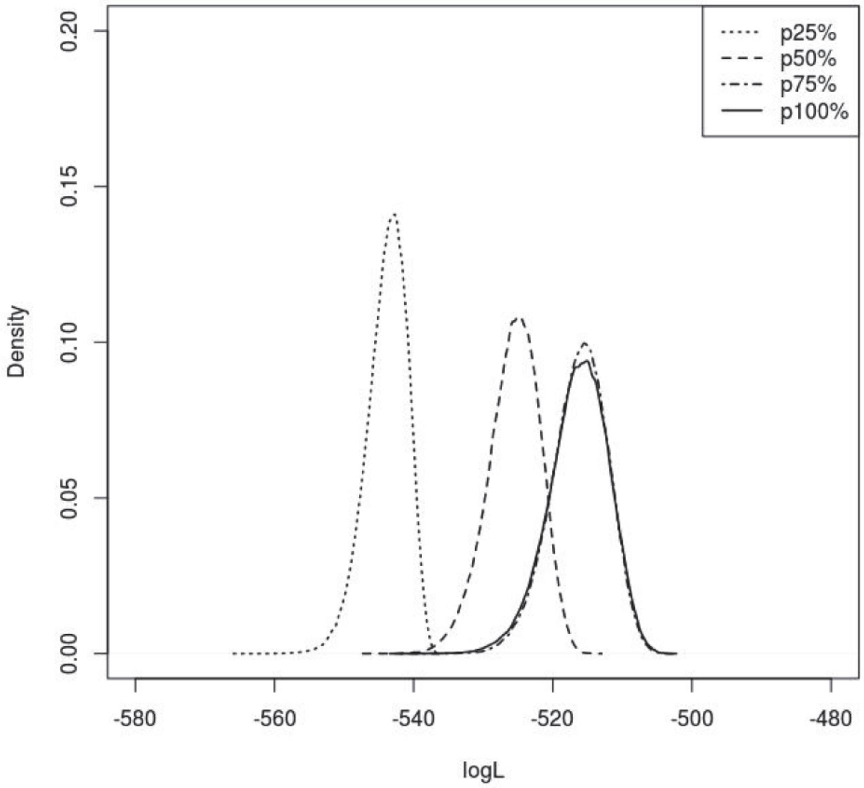


Figure 2. log$L$ *distributions for the four calibrations carried out.*

Table 5

NRMSEs 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

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 behavior 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 [8]. "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 modeling, 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 chose 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 was expected to deteriorate, but this 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 [12]. Roberts, Gelman, and Gilks [25] 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 a 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.

4.2. On the use of canonical correlation analysis for parameter screening. This work is the first attempt to use CCA 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 LSA [34] and GSA [20]. However, the aim of this work was not to explore model sensitivity but to provide a method for parameter screening by means of CCA 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 [5, 21] is a GSA that has already been used to screen parameters of a process-based forest model [31]. 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 CCA, 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 nonlinear in the parameters.

The number of parameters that can be left out of 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 GSA 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.

5. Conclusions.

We introduced a new methodology for parameter screening based on CCA. 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 analyze the sensitivity of only 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 BC 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.

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

REFERENCES

[1] O. Andrén and T. Kätterer, *ICBM: The introductory carbon balance model for exploration of soil carbon balances*, Ecol. Appl., 7 (1997), pp. 1226–1236.

[2] C. Andrieu and J. Thoms, *A tutorial on adaptive MCMC*, Stat. Comput., 18 (2008), pp. 343–373.

[3] M. Aubinet, A. Grelle, A. Ibrom, Ü. Rannik, J. Moncrieff, T. Foken, A. S. Kowalski, et al., *Estimates of the annual net carbon and water exchange of forests: The EUROFLUX methodology*, Adv. Ecol. Res., 30 (1999), pp. 113–175.

[4] D. D. Baldocchi, *Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: Past, present and future*, Global Change Biol., 9 (2003), pp. 479–492.

[5] F. Campolongo, J. Cariboni, and A. Saltelli, *An effective screening design for sensitivity analysis of large models*, Environ. Model. Softw., 22 (2007), pp. 1509–1518.

[6] A. C. Correia, F. Minunno, M. C. Caldeira, J. Banza, J. Mateus, M. Carneiro, L. Wingate, et al., *Soil water availability strongly modulates soil CO_2 efflux in different Mediterranean ecosystems: Model calibration using the Bayesian approach*, Agr. Ecosyst. Environ., 161 (2012), pp. 88–100.

[7] R. I. Cukier, C. M. Fortuin, K. E. Shuler, A. G. Petschek, and J. H. Schaibly, *Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients.* I. *Theory*, J. Chem. Phys., 59 (1973), pp. 3873–3878.

[8] B. Efron, *Bayes' theorem in the* 21*st century*, Science, 340 (2013), pp. 1177–1178.

[9] L. J. Esprey, P. J. Sands, and C. W. Smith, *Understanding* 3-*PG using a sensitivity analysis*, For. Ecol. Manage., 193 (2004), pp. 235–250.

[10] L. Fontes, J.-D. Bontemps, H. Bugmann, M. Van Oijen, C. Gracia, K. Kramer, M. Lindner, T. Rötzer, and J. P. Skovsgaard, *Models for supporting forest management in a changing environment*, Forest Syst., 19 (2010), pp. 8–29.

[11] A. Gelman and D. B. Rubin, *Inference from iterative simulation using multiple sequences*, Statist. Sci., 7 (1992), pp. 457–511.

[12] J. Gill, *Is partial-dimension convergence a problem for inferences from MCMC algorithms?*, Polit. Anal., 16 (2008), pp. 153–178.

[13] J. F. Hair, R. L. Tatham, R. E. Anderson, and W. C. Black, *Canonical correlation analysis*, in Multivariate Data Analysis, 5th ed., Prentice Hall, Englewood Cliffs, NJ, 1998, chapter 8.

[14] H. Hotelling, *Relations between two sets of variates*, Biometrika, 28 (1936), pp. 321–377.

[15] K. Johnsen, L. Samuelson, R. Teskey, S. McNulty, and T. Fox, *Process models as tools in forestry research and management*, Forest Sci., 47 (2001), pp. 2–8.

[16] T. Kätterer and O. Andrén, *Long-term agricultural field experiments in Northern Europe: Analysis of the influence of management on soil carbon stocks using the ICBM model*, Agr. Ecosyst. Environ., 72 (1999), pp. 165–179.

[17] T. Kätterer and O. Andrén, *The ICBM family of analytically solved models of soil carbon, nitrogen and microbial biomass dynamics—Descriptions and application examples*, Ecol. Model., 136 (2001), pp. 191–207.

[18] J. J. Landsberg and R. H. Waring, *A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning*, For. Ecol. Manage., 95 (1997), pp. 209–228.

[19] M. Lindner, M. Maroschek, S. Netherer, A. Kremer, A. Barbati, J. Garcia-Gonzalo, R. Seidl, et al., *Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems*, For. Ecol. Manage., 259 (2010), pp. 698–709.

[20] F. Minunno, M. Van Oijen, D. R. Cameron, S. Cerasoli, J. S. Pereira, and M. Tomé, *Using a Bayesian framework and global sensitivity analysis to identify strengths and weaknesses of two process-based models differing in representation of autotrophic respiration*, Environ. Model. Softw., 42 (2013), pp. 99–115.

[21] M. D. Morris, *Factorial sampling plans for preliminary computational experiments*, Technometrics, 33 (1991), pp. 161–174.

[22] K. Ogle and J. J. Barber, *Bayesian data—Model integration in plant physiological and ecosystem ecology*, in Progress in Botany, Springer, Springer-Verlag, Berlin, Heidelberg, 2008, pp. 281–311.

[23] J. S. Pereira, J. A. Mateus, L. M. Aires, G. Pita, C. Pio, J. S. David, V. Andrade, et al., *Net ecosystem carbon exchange in three contrasting Mediterranean ecosystems—The effect of drought*, Biogeosciences, 4 (2007), pp. 791–802.

[24] M. Reichstein, E. Falge, D. Baldocchi, D. Papale, M. Aubinet, P. Berbigier, C. Bernhofer, et al., *On the separation of net ecosystem exchange into assimilation and ecosystem respiration: Review and improved algorithm*, Global Change Biol., 11 (2005), pp. 1424–1439.

[25] G. O. Roberts, A. Gelman, and W. R. Gilks, *Weak convergence and optimal scaling of random walk Metropolis algorithms*, Ann. Appl. Probab., 7 (1997), pp. 110–120.

[26] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola, *Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index*, Comput. Phys. Comm., 181 (2010), pp. 259–270.

[27] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto, *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, John Wiley & Sons, Chichester, UK, 2004.

[28] P. J. Sands and J. J. Landsberg, *Parameterisation of* 3-*PG for plantation grown Eucalyptus globulus*, For. Ecol. Manage., 163 (2002), pp. 273–292.

[29] D. S. Sivia and J. Skilling, *Data Analysis: A Bayesian Tutorial*, Oxford University Press, Oxford, UK, 2006.

[30] I. M. Sobol', *Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates*, Math. Comput. Simulation, 55 (2001), pp. 271–280.

[31] M. Van Oijen, D. R. Cameron, K. Butterbach-Bahl, N. Farahbakhshazad, P.-E. Jansson, R. Kiese, K.-H. Rahn, C. Werner, and J. B. Yeluripati, *A Bayesian framework for model calibration, comparison and analysis: Application to four models for the biogeochemistry of a Norway spruce forest*, Agr. Forest Meteorol., 151 (2011), pp. 1609–1621.

[32] M. Van Oijen, C. Reyer, F. J. Bohn, D. R. Cameron, G. Deckmyn, M. Flechsig, S. Härkönen, et al., *Bayesian calibration, comparison and averaging of six forest models, using data from Scots pine stands across Europe*, For. Ecol. Manage., 289 (2013), pp. 255–268.

[33] M. Van Oijen, J. Rougier, and R. Smith, *Bayesian calibration of process-based forest models: Bridging the gap between models and data*, Tree Physiol., 25 (2005), pp. 915–927.

[34] G. Xenakis, D. Ray, and M. Mencuccini, *Sensitivity and uncertainty analysis from a coupled* 3-*PG and soil organic matter decomposition model*, Ecol. Model., 219 (2008), pp. 1–16.

CHAPTER 6

“Conclusions”

Chapter 6

1 Conclusions

In this work, modern computational methods were implemented and used in the context of forest modelling for a variety of case studies, spanning from growth prediction models to soil respiration models and process-based models. The great potential of the Bayesian method for reducing uncertainty in model parameters and outputs was shown. Furthermore Bayesian statistics, used in combination with other techniques, helped to evaluate changes in forest models.

Modifications in the structure of process-based models can be theoretically correct, but can be misleading if the data used to drive or calibrate the model are inadequate. The methodology presented in Chapter 2 uses in combination the Bayesian calibration, the Bayesian model comparison, the error decomposition analysis and the Morris method to rigorously test changes in the structure of PBMs. The case study of the 3PGN model was shown; the results had two main implications: one is methodological and more relevant from a modeler's perspective, while the second is related to the application of 3PGN for the modelling of *Eucalyptus globulus* forests in Portugal, being more relevant from a user's perspective.

From the methodological point of view, thanks to the different analysis carried out, it was possible to identify the strength and weaknesses of the two model versions. This methodology is applicable to any kind of model, from simple parameter sparse empirical models to complex process-based models and it allows to have a deep and complete understanding of model behaviour. Furthermore the methodology makes possible the improvement and optimization of future data collection. In fact, knowing which are the most uncertain parameters and knowing the sensitivity of model outputs, data collection can be designed focusing on the variables that are more significant for the reduction of parameter uncertainty, lowering the costs of field surveys.

The second main conclusion is that the new algorithm for autotrophic respiration implemented in 3PGN, turned the model more robust and reliable in predicting the growth and the carbon sequestration of *Eucalyptus globulus* plantations in Portugal. A correct and effective modelling of these ecosystems is highly relevant for the forestry sector in Portugal. According to the preliminary results of the 6th forest national inventory (ICNF,Instituto de Conservação da Natureza e das Florestas, 2013) the area covered by Eucalyptus in the Continental Portugal, in 2010, is about 812 millions of hectares, corresponding to the 26% of forest total area. Eucalyptus is used by the pulp and paper industry, which is responsible for a turnover of over 2 billions of euros (Celpa, Associação da Industria Papeleira., 2012). The high competitiveness of the manufacturing industry of Eucalyptus makes this an economically important forest species. The increase in the annal availability of wood and pulp can be achieved by an increment of the area of the species or increasing the productivity of existing or new stands. Forest management plays a key role in optimizing and improving the productivity of forests. 3PGN is a useful tool for forest management and it is already used by one of the leading groups of printing and writing paper manufacturer (i.e., the Portucel Soporcel group) to perform analysis at strategical and operational level. The 3PGN model can now be used at regional scale to estimate the carbon sequestration and the productivity of Eucalyptus forests in Portugal and to assess the impact of climate change on the functionality of these ecosystems (Minunno *et al.*, in prep.(a)). Future work should also take in consideration the calibration of 3PGN for different Eucalyptus clones commonly used in the country. Modelling the growth of distinct clones will allow forest managers to chose the variety that is more productive in specific environmental conditions and also is more suitable under future climatic changes. Finally 3PGN should be calibrated for the other two most economically important forest species in Portugal, namely *Quercus suber* and *Pinus pynaster*.

In Chapter 3, the methodology previously presented was applied to simple empirical models. In this study the ability of bioclimatic models in assessing soil respiration of Mediterranean ecosystems was tested. Results reinforced the importance of soil moisture as one of the main factors driving soil respiration in Mediterranean region. In fact, the models that were more sensitive to this environmental variable better reproduced soil carbon fluxes. Simple empirical models can be used to simulate soil respiration in Mediterranean grassland at a monthly time step, but they are not reliable for forest ecosystems. Future works should test if causal models that reproduce the plant physiology processes are more appropriate to simulate soil carbon fluxes at finer time scales and for more complex ecosystems.

In the fourth chapter, it was shown how model parameter assessment, the quantification of model output uncertainties and model evaluation can be addressed by means of Bayesian calibration (BC), Bayesian model comparison (BMC) and Bayesian model averaging (BMA). In this study we specifically looked at the ability of forest models of predicting forest growth, but this approach can be extended to any ecological problem.

Process-based models are developed with the intention of having a causal structure and being generally applicable, but they are demanding in terms of input and parameters, which leads necessarily to uncertainty in the projections. In this chapter it was faced the issue of evaluating the performances of forest models using regional specific calibrations and generic parameter estimates. The idea was to test if the considered dynamic models for Scots pine are sufficiently general, so a generic calibration to data from across Europe would be better than country-by-country calibrations. Calibrating models separately for each country did not clearly improve the within-country predictive capacity of the models; however, future studies with larger number of sites and data would be needed to confirm or not our results.

Finally, the main novelty of this work was introducing the Bayesian Model averaging in forest research. BMA, taking into accounts for both parametric and model structural uncertainty, affords a robust way of modelling natural ecosystems.

Part of the PhD work focused on providing methodologies and tools that can increase the efficiency of the Bayesian computations in order to expand their application to more complex problems. One of the weaknesses of the Bayesian approach lays in the fact that it is computationally costly. In Chapter 5 we presented a methodology for parameter screening aimed at reducing the parameter space to be calibrated, focusing only on the parameters that are more influential on model output. The methodology is generally applicable, easy to implement and particularly suitable for complex process-based models, but it has the limit that the likelihood is not integrated in the screening method. In fact, from a Bayesian perspective we are mostly interested in the impact that parameter changes have on the likelihood and not directly on model output. Future work should investigate how to improve the screening method provided here, including the likelihood in this methodology.

The efficiency of Bayesian calibration can be increased also using more effective sampling algorithms. Many Markov chain Monte Carlo algorithms have been developed with the aim of efficiently exploring the parameter space to provide a representative estimate of the posterior distribution without an extremely large number of samples. Future works should test if the use of more complex MCMCs would bring a significant reduction of the computational costs for the calibration of process-based forest models (Minunno *et al.*, in prep.(b)).

The use of the Bayesian approach will continue to increase in the future because it gives to environmental modellers a better awareness of the robustness and reliability of their predictions and inferences. Furthermore, by means of the inverse parameterisation, it is possible to integrate, in the biogeochemical models, a huge amount of information coming from the increasing availability of multiple sources of observational data (i.e., eddy covariance fluxes, forest inventories, permanent sample plots and remote sensing). A new source of data useful for improving our knowledge about forest ecosystems comes from Unmanned Air Vehicles (UAVs). Aerodrones are already used in the military, in cinematography and agriculture; but they are fairly new to forestry. The usage of UAVs for agricultural and forest environments allows the gathering of data for monitoring in a fast and inexpensive way. Drones can be equipped with cameras, sensors and GPS providing useful data for forest survey. All these kind of data are extremely useful for model calibration and validation because they provide information at different scales in space and time. Integrating this informations in a Bayesian framework for forest model calibration should be take in consideration in the future.

Models are always a simplification of the real system that we aim to reproduce (Mäkelä et al., 2000) and forest models span from simple equations to complex parameter-rich models (Fontes et al., 2010); so how should we choose a model? Modellers should find a compromise between complexity and simplicity in order to accurately reproduce the processes but at the same time reducing the sources of uncertainty. The Bayesian model comparison provides a logical and robust answer in choosing the appropriate model by combining parameter uncertainty, model structure and data availability.

Ultimately, the importance of the prior distribution and the great potential that it has in the contest of process-based modelling calibration must be emphasised. Efron (2013) showed how the choice of the prior can affect Bayesian inferences. Through the prior it is possible to lead parameter estimates toward realistic values. Even the simplest prior can enclose basic but at the same time fundamental information about the parameters. Consider the more common situation in vegetation modelling, where high parameter uncertainty exists so the uniform prior distribution is used, by defining a maximum and minimum the modeller constrains the parameter values in a range that is biologically meaningful. Furthermore comparing the distributions *a posteriori* with those *a priori*, will provide important information about model structure. If the posterior distribution is strongly pushed toward the minimum or the maximum value of a parameter, it is likely that there is some structural problem in the way the processes are represented. Using multiple sources of information, gathered thorough literature searches, expert knowledge, observations or previous BCs, it is possible to build robust and strong prior that can improve the calibration. For instance, many physiological measurements are becoming available, providing relevant details about the vegetation processes occurring at fine ecological level. By means of Bayesian statistics forest modellers can integrate all the data and information available at different spatial and temporal scales enhancing the representation of the ecosystem dynamics.

2 References

Celpa, Associação da Industria Papeleira., 2012. Boletim estatístico, Indústria papeleira portuguesa 2011.

Efron, B., 2013. Bayes' Theorem in the 21st Century. Science 340, 1177–1178.

Fontes, L., Bontemps, J.-D., Bugmann, H., Van Oijen, M., Gracia, C., Kramer, K., Lindner, M., Rötzer, T., Skovsgaard, J.P., 2010. Models for supporting forest management in a changing environment.

ICNF,Instituto de Conservação da Natureza e das Florestas, 2013. Áreas dos usos do solo e das espécies florestais de Portugal continental, 1995/2005/2010. Resultados preliminares, v1.0.

Mäkelä, A., Landsberg, J., Ek, A.R., Burk, T.E., Ter-Mikaelian, M., Ågren, G.I., Oliver, C.D., Puttonen, P., 2000. Process-based models for forest ecosystem management: current state of the art and challenges for practical implementation. Tree Physiol. 20, 289–298.

Minunno, F., Palma, J., in prep. (a). The impact of the climate change on *Eucalyptus globulus* plantations in Portugal. Bayesian model averaging of two process-based models.

Minunno, F., van Oijen, M., Cameron, D., in prep. (b). A comparison of six MCMC algorithms for the Bayesian calibration of process-based forest models.

CHAPTER 7

“Appendices”

Appendix – I

MCMC code in R: the Metropolis algorithm

```
    nChain <- as.integer(100000)
    nBI    <- 0.5*nChain
    pChain <- matrix(0, nrow=nChain, ncol=npar)
    logLChain <- matrix(0, nrow=nChain)
    covPar <- matrix(0, nrow=npar, ncol=npar)

    # Start the chain using uniform random values in [parmin, parmax].
    pValues <- runif(npar, min=parmin/abs(parmod), max=parmax/abs(parmod))

    #-----
    # First call and ru the model in fortran.
    Sim_pValues <- pValues*abs(parmod)
    .Fortran('changepars', Sim_pValues)
    logL0 <- 0.0
    y     <- matrix(1,144,3)
    output        <- .Fortran('espirra',logL0,y)[[2]]
    Sims_data <- output

    # Calculate likelihood
    logL0         <- flogL(Sims_data,obs_data,obs_data_s)

    #Calculate prior probability
    logPrior0  <-    sum(sum(dtnorm(pValues[norm], mean=parar1[norm]/abs(parmod[norm]),
sd=parar2[norm]/abs(parmod[norm]),lower=parmin[norm]/abs(parmod[norm]),
upper=parmax[norm]/abs(parmod[norm]), log = T)),
                    sum(dweibull(pValues[post_wei], scale=parar1[post_wei],
    shape=parar2[post_wei], log = T)),
                    sum(dunif(pValues[uni], min=parmin[uni]/abs(parmod[uni]),
max=parmax[uni]/abs(parmod[uni]), log=T)),
                    sum(dnorm(pValues[post_norm],  parar1[post_norm], parar2[post_norm], log
= T)),
                    sum(dgamma(pValues[post_gam], shape=parar1[post_gam],rate=
parar2[post_gam], log = T)))

    psetMAP    <- pValues
    logMAP     <- logPrior0 + logL0
    pChain[1,] <- pValues
    logLChain[1] <- logL0

    vcovProp <- diag( (0.01*(parmax - parmin)/abs(parmod))^2 )
    diag(scalProp) <- scal_Chain*2.4^2/npar
    covPar <- vcovProp
    kc     <- 0

    #start loop
    for (j in 2:nChain)
```

```
              {
              if (j%%100 == 0) print(j)

              candidatepValues  <- mvrnorm(n=1, pValues, scalProp*covPar, tol=1e-6,
empirical=FALSE)
              reflectionFromMin <- pmin(0.,candidatepValues-parmin/abs(parmod))
              reflectionFromMax <- pmax(0.,candidatepValues-parmax/abs(parmod))
              candidatepValues  <- candidatepValues - 2.*reflectionFromMin - 2.*reflectionFromMax

              Sim_candidatepValues <- candidatepValues*abs(parmod)
              .Fortran('changepars', Sim_candidatepValues)

              logL1 <- 0.0

          y     <- matrix(1,144,3)
          output         <- .Fortran('espirra',logL1,y)[[2]]
          Sims_data <- output

          # Calculate likelihood
          logL1        <- flogL(Sims_data,obs_data,obs_data_s)

          #Calculate prior probability
          logPrior1  <-     sum(sum(dtnorm(Sim_candidatepValues[norm],mean =
parar1[norm]/abs(parmod[norm]), sd=parar2[norm]/abs(parmod[norm]),
lower=parmin[norm]/abs(parmod[norm]), upper=parmax[norm]/abs(parmod[norm]), log = T)),
                            sum(dweibull(Sim_candidatepValues[post_wei], scale=parar1[post_wei],
          shape=parar2[post_wei], log = T)),
                            sum(dunif(Sim_candidatepValues[uni], min=parmin[uni]/abs(parmod[uni]),
max=parmax[uni]/abs(parmod[uni]), log=T)),
                            sum(dnorm(Sim_candidatepValues[post_norm],  parar1[post_norm],
parar2[post_norm], log = T)),
                            sum(dgamma(Sim_candidatepValues[post_gam],shape=parar1[post_gam],
rate= parar2[post_gam], log = T)))

              logalpha <- logPrior1 + logL1 - (logPrior0 + logL0)
              if (log(runif(1,0,1)) < logalpha)
                {
                pValues   <- candidatepValues
                logPrior0 <- logPrior1
                logL0     <- logL1

                if (j > nBI)
                  {
                  if ((logPrior0 + logL0) > logMAP)
                    {
                    logMAP  <- logPrior0 + logL0
                    psetMAP <- pValues
                    }
                  kc <- kc + 1
                  }
                }
```

```
           pChain[j,] <- pValues
           logLChain[j] <- logL0
          }
```

MCMC code in R: Differential Evolution Markov Chain algorithm with fewer chains and snooker updater (DE-MCzs)

```
        DE_MC.ZS <- function(Npop = 3, Z, FUN, X= matrix(Z[,1:Npop], ncol = Npop), CR= 1.0,
F = 2.38, pSnooker= 0.1, pGamma1 = 0.1, n.generation = 10, n.thin = 1, n.burnin = 0, eps.mult
=0.2,eps.add = 0, ...){
        M0 = mZ = ncol(Z)
        Npop = ncol(X)
        Npar = nrow(X)
        Npar12  =(Npar - 1)/2  # factor for Metropolis ratio DE Snooker update
        F2 = F/sqrt(2*Npar)
        F1 = 1.0
        accept = rep(NA,n.generation)
        iseq = 1:Npop
        rr = NULL
        r_extra = 0
        logfitness_X = apply (X, 2, FUN, ...)
        posterior <- matrix(NA,n.generation,Npop)
        pChain_r <- array(NA,dim=c(n.generation,Npar,Npop))
        pChain <- matrix(NA,((n.generation-n.burnin)*Npop),Npar)
        logLChain <- rep(NA, ((n.generation-n.burnin)*Npop))

        for (iter in 1:n.generation) {
                if (iter%%100 == 0) print(c(iter,logfitness_X))
          accepti = 0
          for (i in iseq){
           # select to random different individuals from Z in rr, a 2-vector
           if ( runif(1)< pSnooker ) {
           # DE-Snooker update
            # if (Npop >1) { z =  X[,sample(iseq[-i],1)]} else { # no real advantage and precludes
parallel computing
             rr = sample(1:mZ, 3, replace = FALSE)
             z = Z[,rr[3]]
            # }                                  # no real advantage and precludes parallel computing
             x_z = X[,i] - z
             D2 = max(sum(x_z*x_z),1.0e-300)
             gamma_snooker = runif(1, min=1.2,max=2.2)
             #gamma_snooker =1.7
             projdiff = sum((Z[,rr[1]] -Z[,rr[2]]) *x_z)/D2  # inner_product of difference with x_z /
squared norm x_z
             x_prop = X[,i] + (gamma_snooker * projdiff) * x_z
             x_z = x_prop - z
             D2prop = max(sum(x_z*x_z), 1.0e-30)
             r_extra = Npar12 * (log(D2prop) - log(D2))   # Npar12  =(Npar - 1)/2  # extra term in
logr for accept - reject ratio
```

```
    } else {
    # DE-parallel direction update
       if ( runif(1)< pGamma1 ) { gamma_par = F1 # to be able to jump between modes
        } else {
        gamma_par = F2 * runif(Npar, min=1-eps.mult, max=1+eps.mult)    # multiplicative error to be applied to the difference
         # gamma_par = F2
       }
       rr = sample(1:mZ, 2, replace = FALSE)
       if (eps.add ==0) {  # avoid generating normal random variates if possible
         x_prop = X[,i] + gamma_par * (Z[,rr[1]]-Z[,rr[2]]) } else {
         x_prop = X[,i] + gamma_par * (Z[,rr[1]]-Z[,rr[2]])  +  eps.add*rnorm(Npar,0,1)
       }
       r_extra = 0
      }

     logfitness_x_prop = FUN(x_prop,  ...)
     logr =  logfitness_x_prop - logfitness_X[i]
    # print(c(logfitness_X[i], logfitness_x_prop ,logr,r_extra))
     if (!is.na(logr) & (logr + r_extra)> log(runif(1)) ){
        accepti = accepti+1
        X[,i] = x_prop
        logfitness_X[i] = logfitness_x_prop
     }
     pChain_r[iter,,i] <- X[,i]
  } # i loop
  accept[iter] = accepti
  if (!(iter%%n.thin) ){
   Z = cbind(Z,X)
   mZ = ncol(Z)
    }
posterior[iter,] <- logfitness_X

} # n.generation
  lChain <- (n.generation-n.burnin)
  for (ij in 1:Npop){
    pChain[(lChain*(ij-1)+1):(lChain*ij),] <- pChain_r[(n.burnin+1):n.generation,,ij]
    logLChain[(lChain*(ij-1)+1):(lChain*ij)] <- posterior[(n.burnin+1):n.generation,ij]}

list(Draws= Z[,-(1:(M0 + Npop* floor(n.burnin/n.thin)))] , accept.prob= accept/Npop, X.final = X, logfitness.X.final = logfitness_X, logLChain = logLChain, pChain_r =pChain_r)
}
```

Appendix – II

Chapter 2: “Using a Bayesian framework and global sensitivity analysis to identify strengths and weaknesses of two process-based models differing in representation of autotrophic respiration”

Results of model calibration

MCMC algorithm: *Metropolis algorithm*
Chain length: 100,000
Number of parameters: 66
Convergence test: Gelman & Rubin 1992 (R) calculated using the R package “coda”
Number of chains: 3
Highest value of R test: 1.37

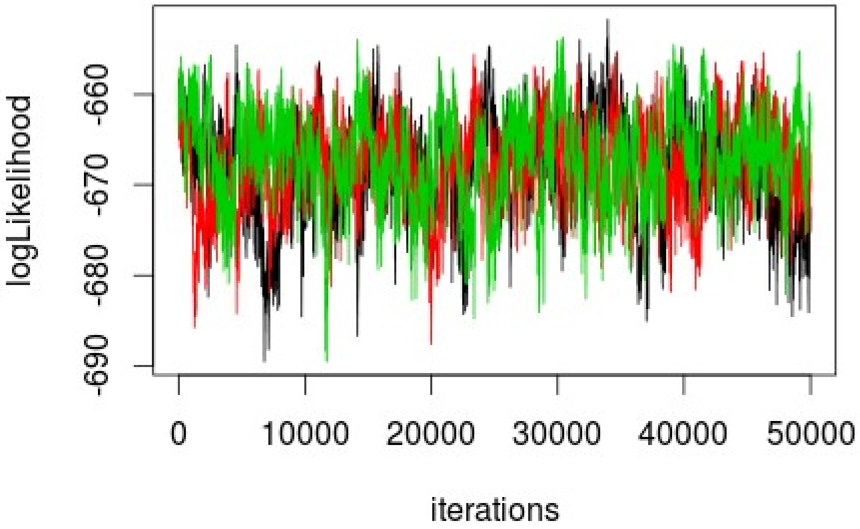
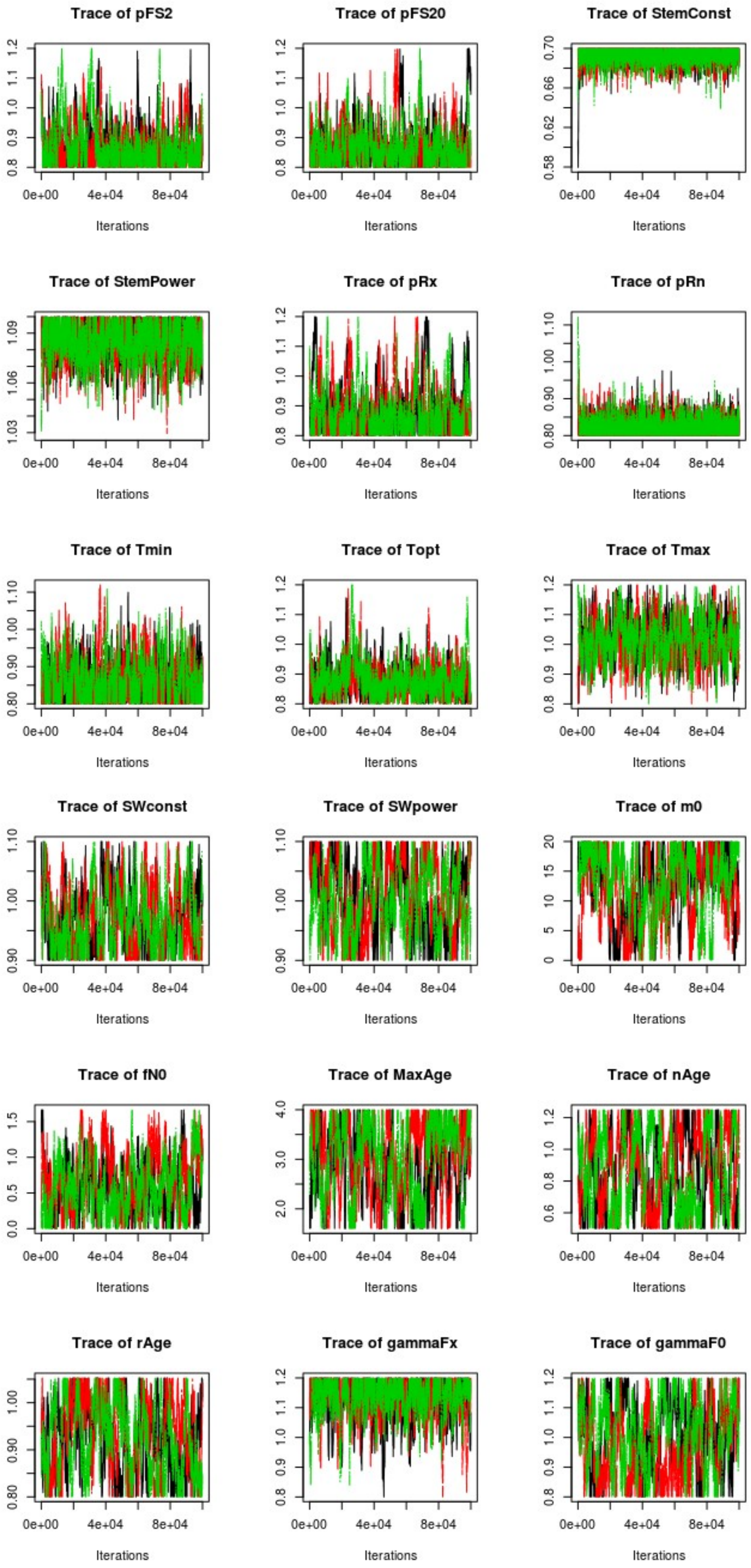


Figure 1. Trace-plot of the likelihood in logarithmic scale. Each colour corresponds to a different chain.


Trace of pFS2

Trace of pFS20

Trace of StemConst

Trace of StemPower

Trace of pRx

Trace of pRn

Trace of Tmin

Trace of Topt

Trace of Tmax

Trace of SWconst

Trace of SWpower

Trace of m0

Trace of fN0

Trace of MaxAge

Trace of nAge

Trace of rAge

Trace of gammaFx

Trace of gammaF0

Figure 2. Trace-plots of the 66 parameters. Each colour corresponds to a different chain.

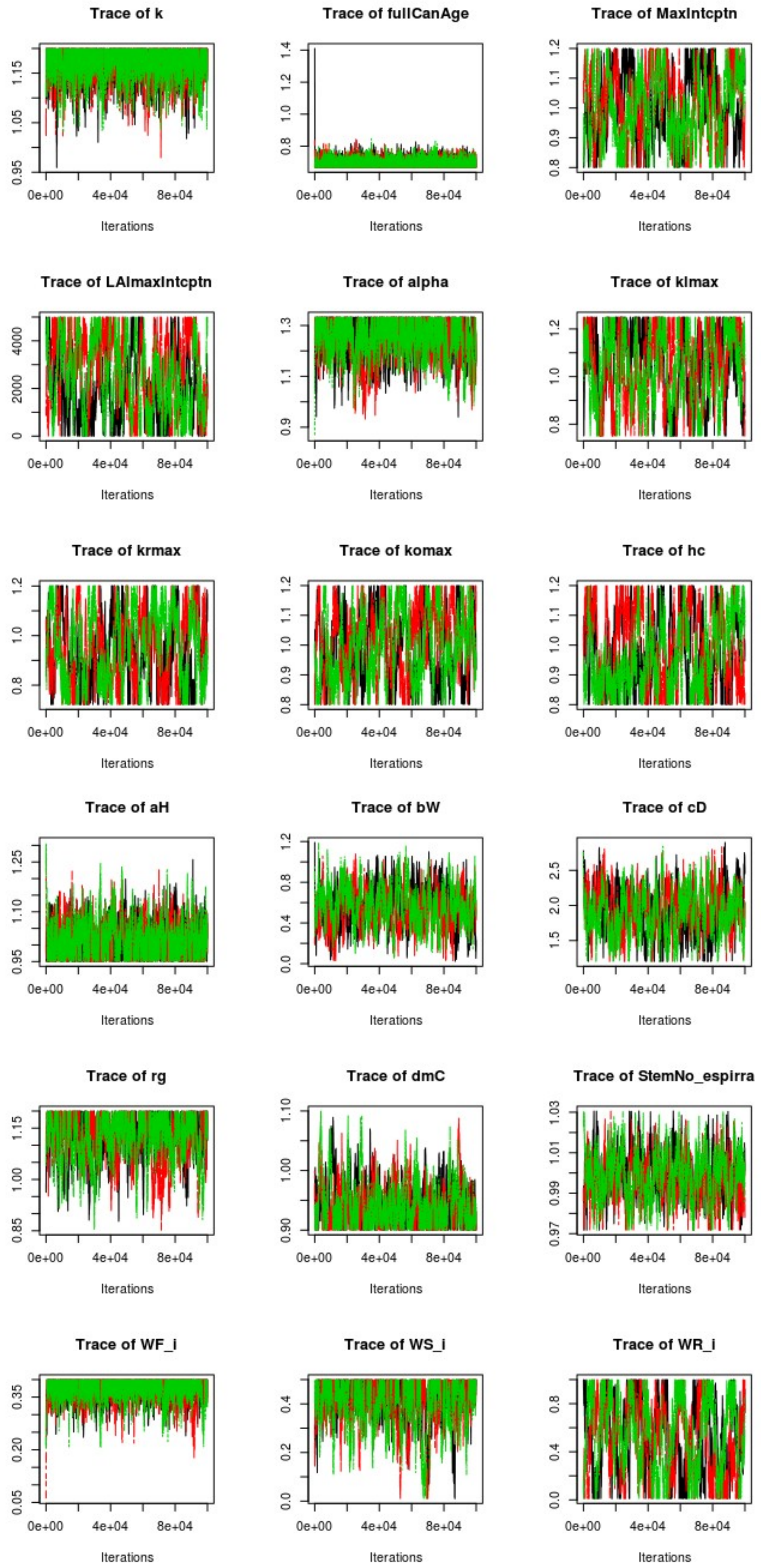


Figure 2. *Continued*

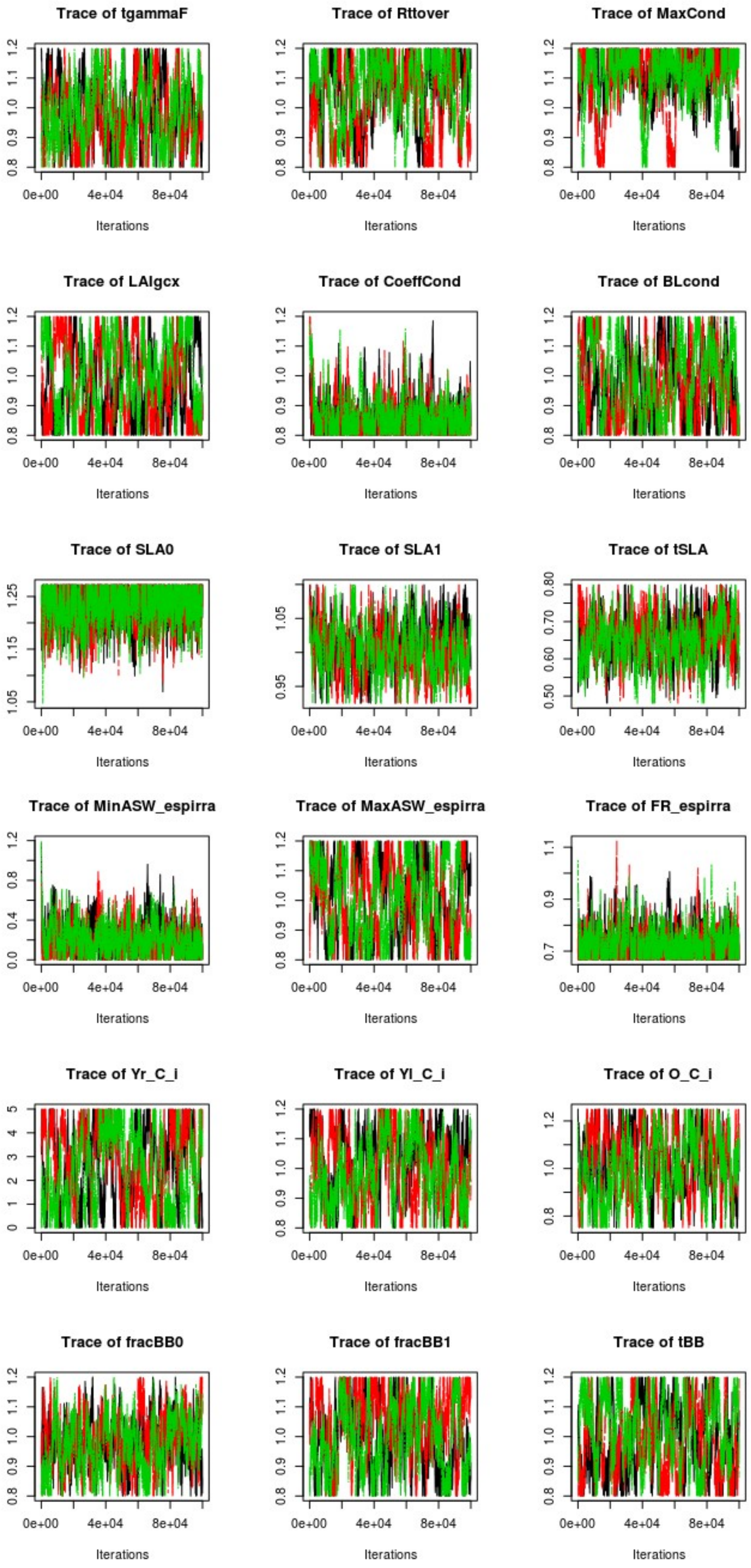


Figure 2. *Continued*

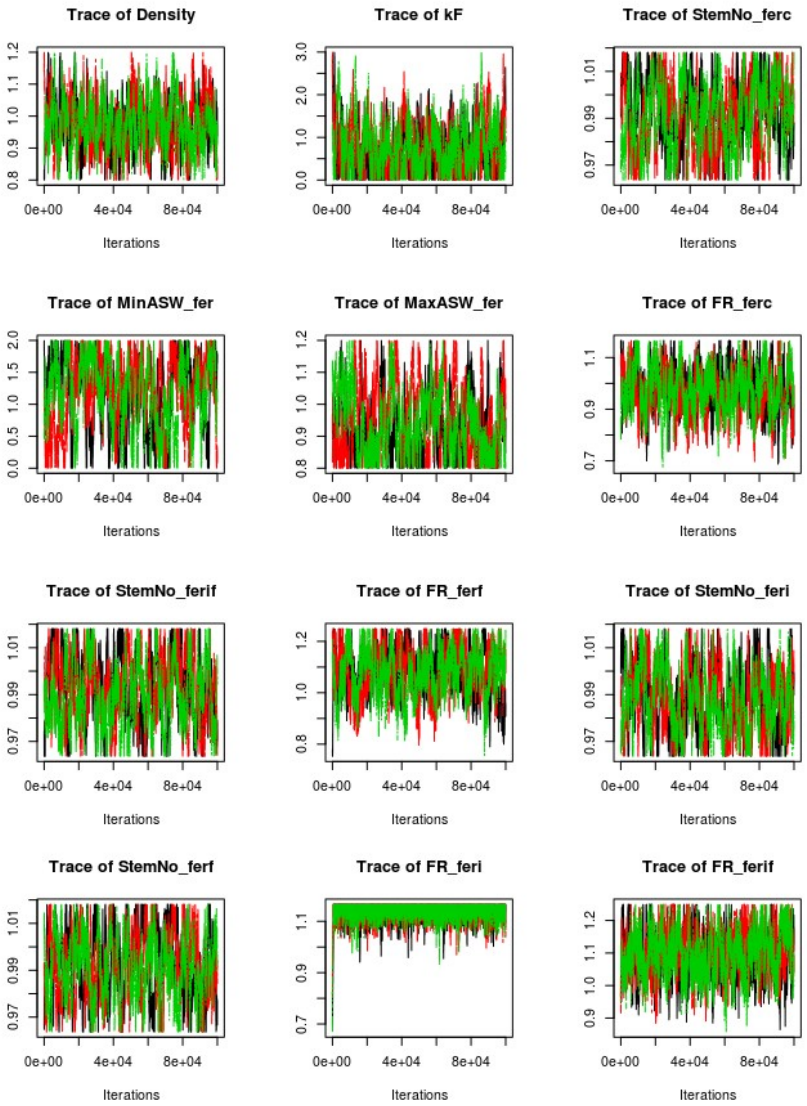


Figure 2. *Continued*

Appendix – III

Chapter 5: “Selecting parameters for Bayesian calibration of a Process-based model: a methodology based on canonical correlation analysis.”

Results of model calibration using 100% of the parameter set

MCMC algorithm: *DE-MCzs*
Chain length: 500,000
Number of chains: 3
Number of parameters: 51
Convergence test: Gelman & Rubin 1992 (R) calculated using the R package “coda”
Highest value of R test: 1.05

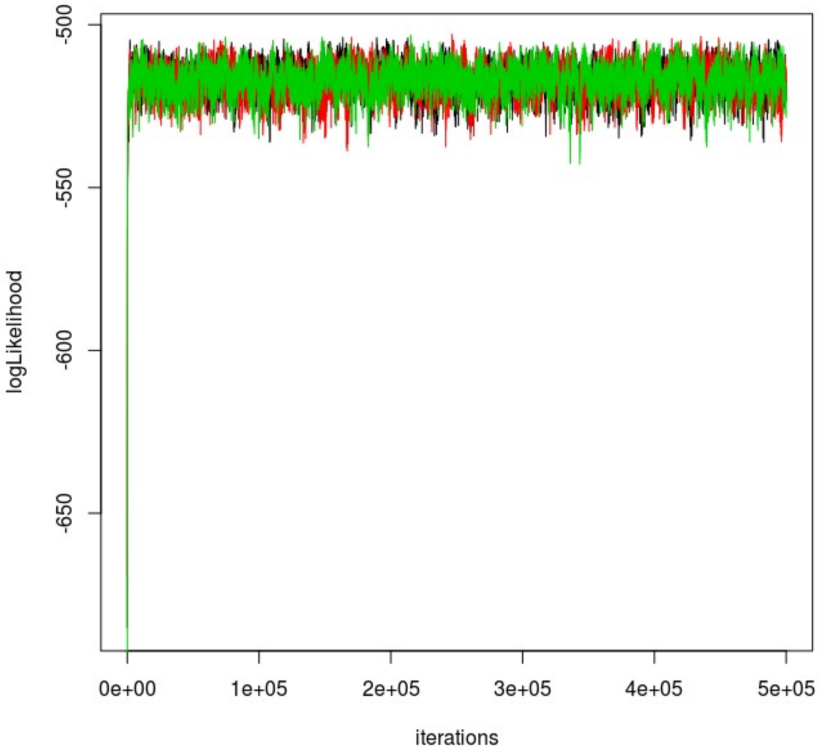


Figure 1. Trace-plot of the likelihood in logarithmic scale. Each colour corresponds to a different chain.

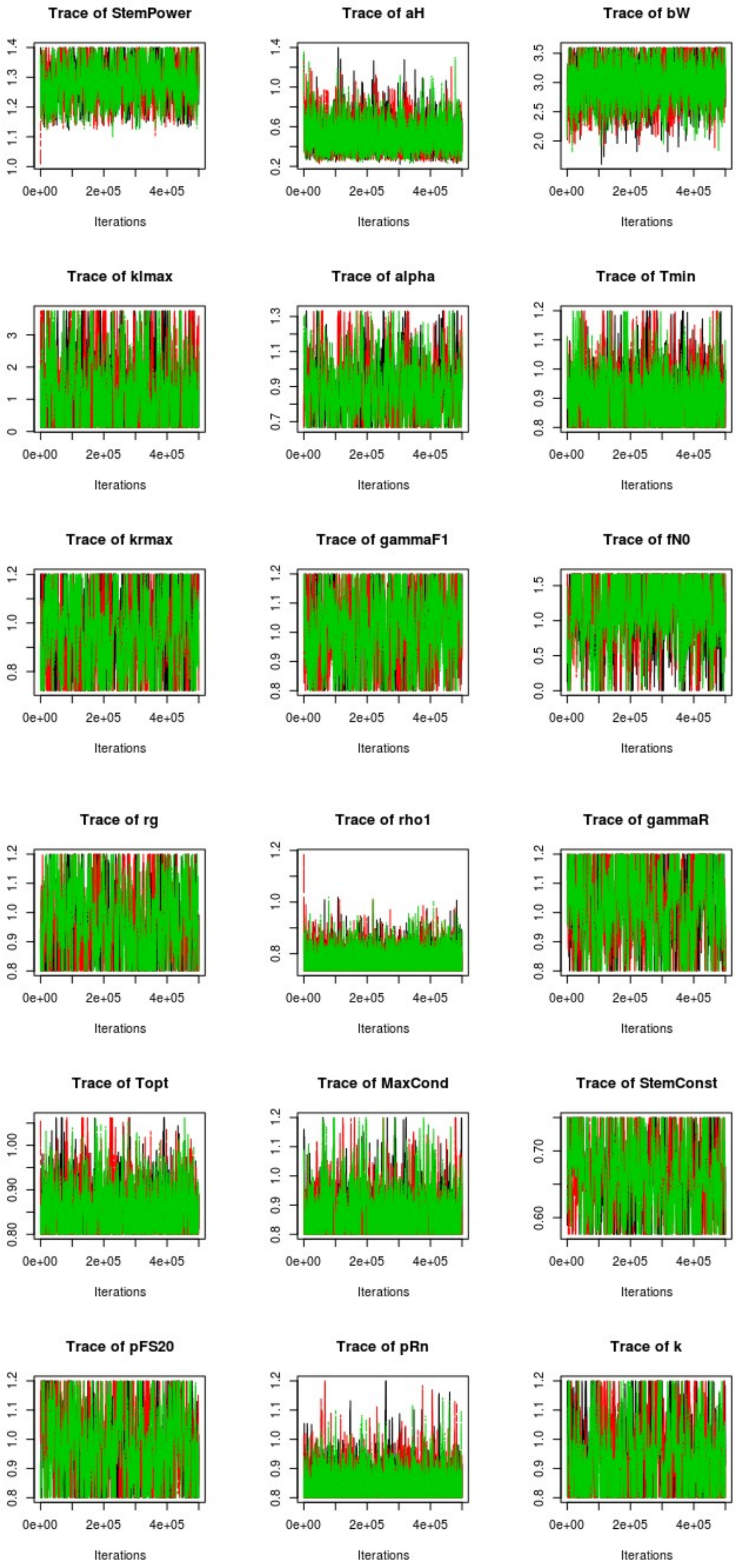


Figure 2. Trace-plots of the 51 parameters. Each colour corresponds to a different chain.

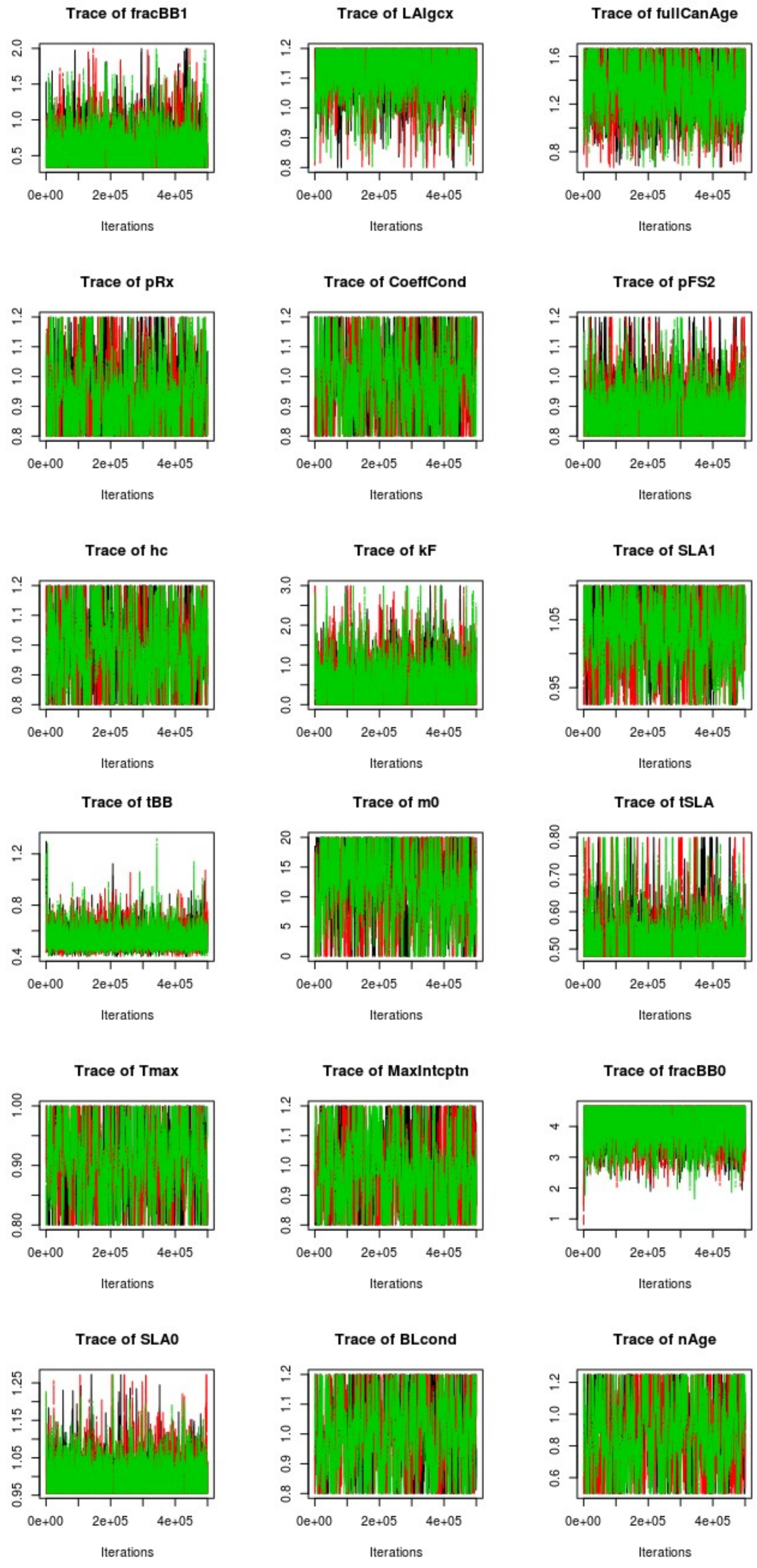


Figure 2. *Continued*

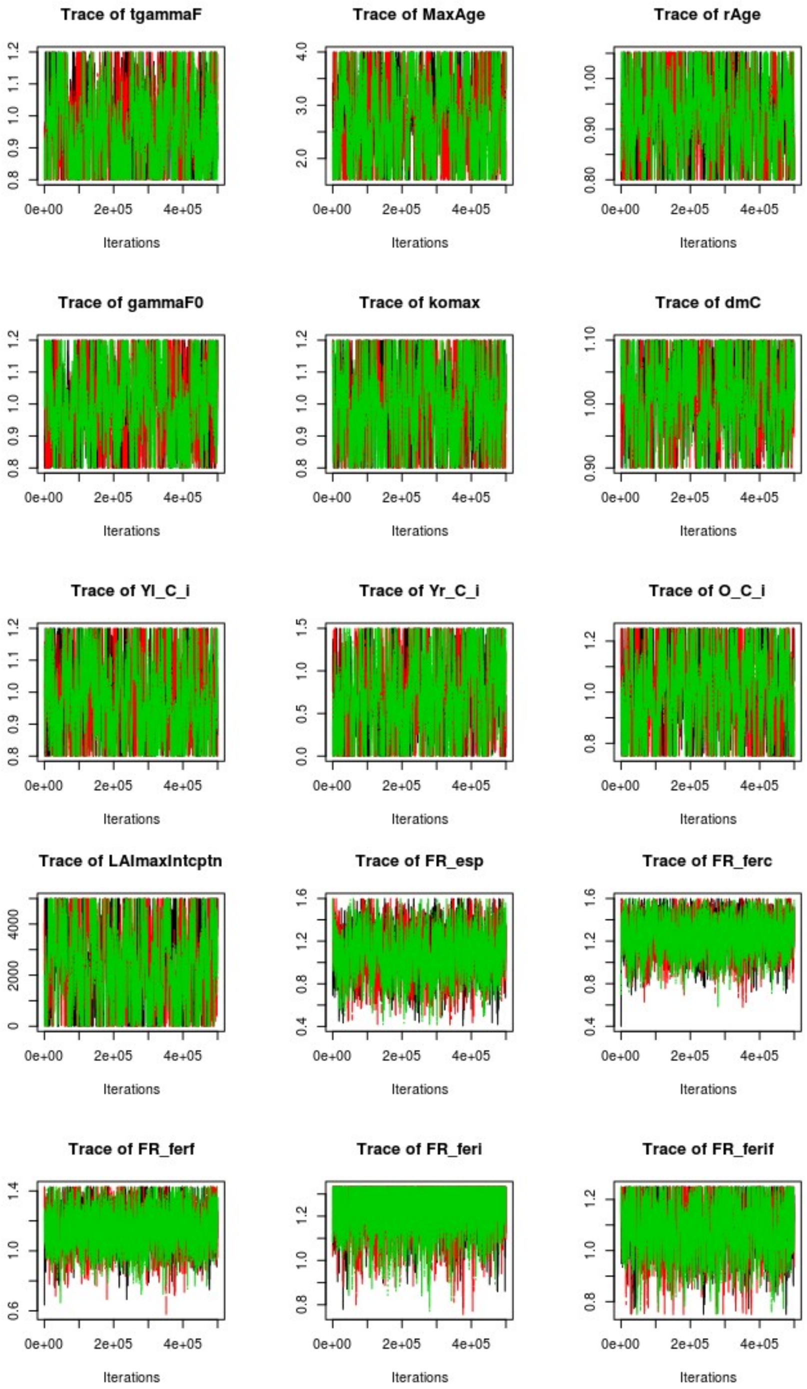


Figure 2. *Continued*