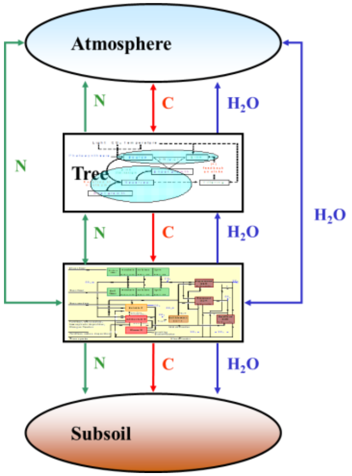
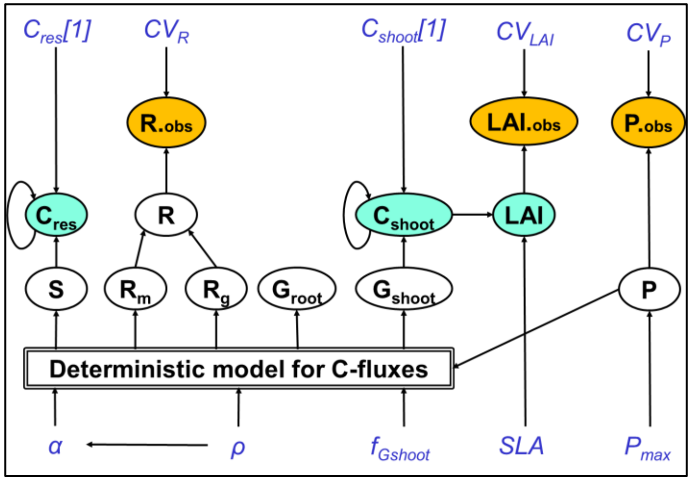
Bayesian Methods for Ecological and Environmental Modelling: DIAGNOSING MODEL WEAKNESSES





Marcel van Oijen
CEH-Edinburgh, 2019-09-12

Contents

1. Introduction
2. Comparison of prior and posterior
3. Analysing posterior model-data mismatch

1. Introduction

1.1 What can BC & BMC not do?

- BC tells us about our parameters: what their values probably are
- BMC tells us about the structure of our models: which model is more plausible than others.

But …

- BC does not tell us why the most probable parameter values sometimes look strange
- BMC does not tell us whether the most plausible model could be improved, or how.

However …

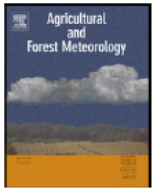
- Comparing prior & posterior distributions can be informative, as can analysis of model-data mismatch

1.2 A three step-procedure

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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



A Bayesian framework for model calibration, comparison and analysis: Application to four models for the biogeochemistry of a Norway spruce forest

M. van Oijen[a,*], D.R. Cameron[a], K. Butterbach-Bahl[b], N. Farahbakhshazad[c,d], P.-E. Jansson[c], R. Kiese[b], K.-H. Rahn[b], C. Werner[b,e], J.B. Yeluripati[f]

[a] Centre for Ecology and Hydrology, CEH-Edinburgh, Bush Estate, Penicuik EH26 0QB, United Kingdom
[b] Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research, Atmospheric Environmental research (IMK-IFU), Kreuzeckbahnstr. 19, 82467 Garmisch-Partenkirchen, Germany
[c] Department of Land and Water Resources Engineering, Royal Institute of Technology, 100 44 Stockholm, Sweden
[d] Swedish Secretariat for Environmental Earth Systems Sciences (SSEESS), The Royal Swedish Academy of Sciences, Stockholm, Sweden
[e] LOEWE Biodiversity and Climate Research Centre (BiK-F), Frankfurt, Germany
[f] School of Biological Sciences, University of Aberdeen, Cruickshank Building, St Machar Drive, Aberdeen AB24 3UU, United Kingdom

ARTICLE INFO

Article history:
Received 7 September 2010
Received in revised form 22 June 2011
Accepted 27 June 2011

Keywords:
Carbon cycle
Nitrogen cycle
NO
N_2O
Uncertainty analysis
Water cycle

ABSTRACT

Four different parameter-rich process-based models of forest biogeochemistry were analysed in a Bayesian framework consisting of three operations: (1) Model calibration, (2) Model comparison, (3) Analysis of model–data mismatch.

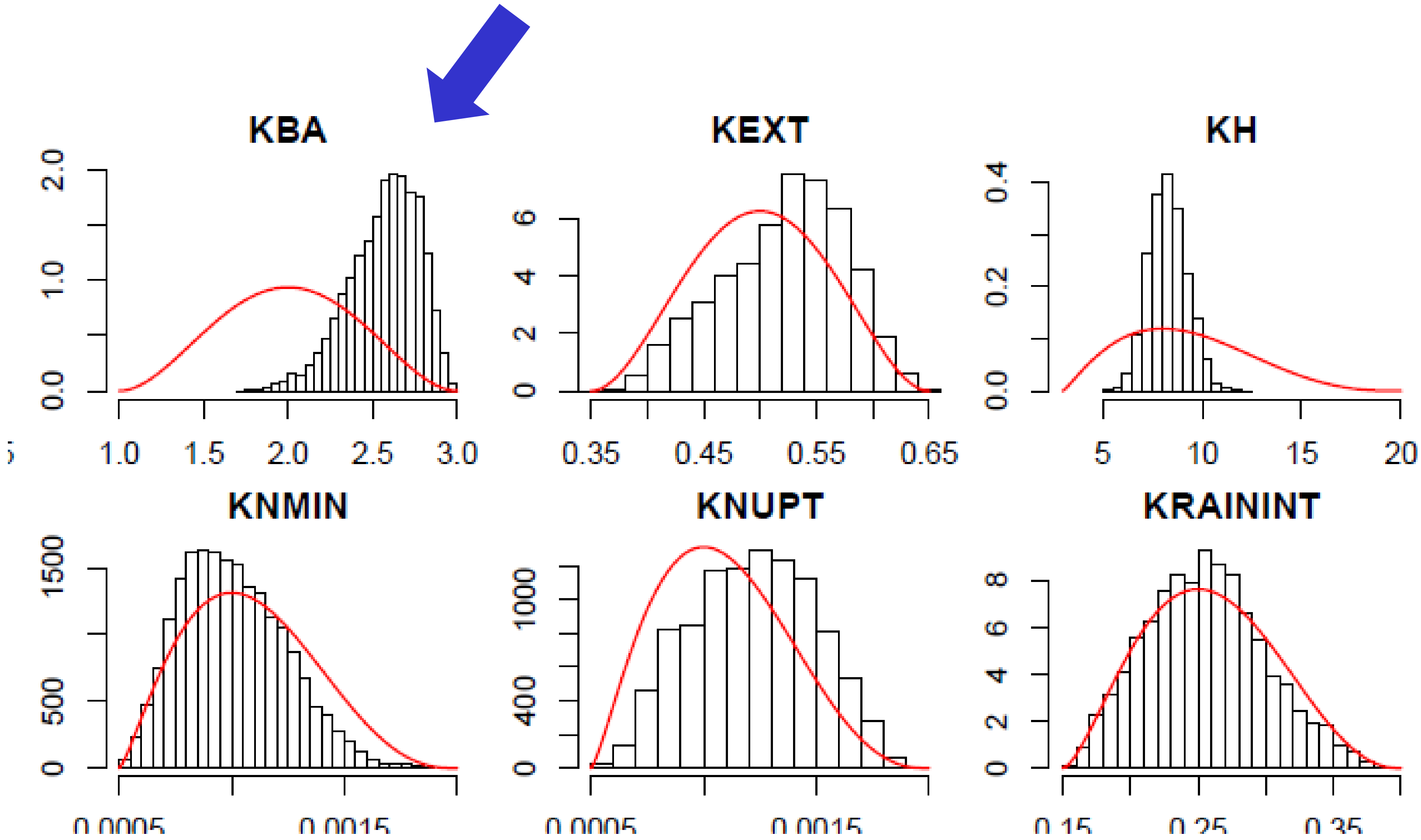
Data were available for four output variables common to the models: soil water content and emissions of N_2O, NO and CO_2. All datasets consisted of time series of daily measurements. Monthly averages and quantiles of the annual frequency distributions of daily emission rates were calculated for comparison with equivalent model outputs. This use of the data at model-appropriate temporal scale, together with the choice of heavy-tailed likelihood functions that accounted for data uncertainty through random and systematic errors, helped prevent asymptotic collapse of the parameter distributions in the calibration.

Model behaviour and how it was affected by calibration was analysed by quantifying the normalised RMSE and r^2 for the different output variables, and by decomposition of the MSE into contributions from bias, phase shift and variance error. The simplest model, BASFOR, seemed to underestimate the temporal variance of nitrogenous emissions even after calibration. The model of intermediate complexity, DAYCENT, simulated the time series well but with large phase shift. COUP and MoBiLE-DNDC were able to remove most bias through calibration.

The Bayesian framework was shown to be effective in improving the parameterisation of the models, quantifying the uncertainties in parameters and outputs, and evaluating the different models. The analysis showed that there remain patterns in the data – in particular infrequent events of very high nitrogenous emission rate – that are unexplained by any of the selected forest models and that this is unlikely to be due to incorrect model parameterisation.

2. Comparison of prior and posterior

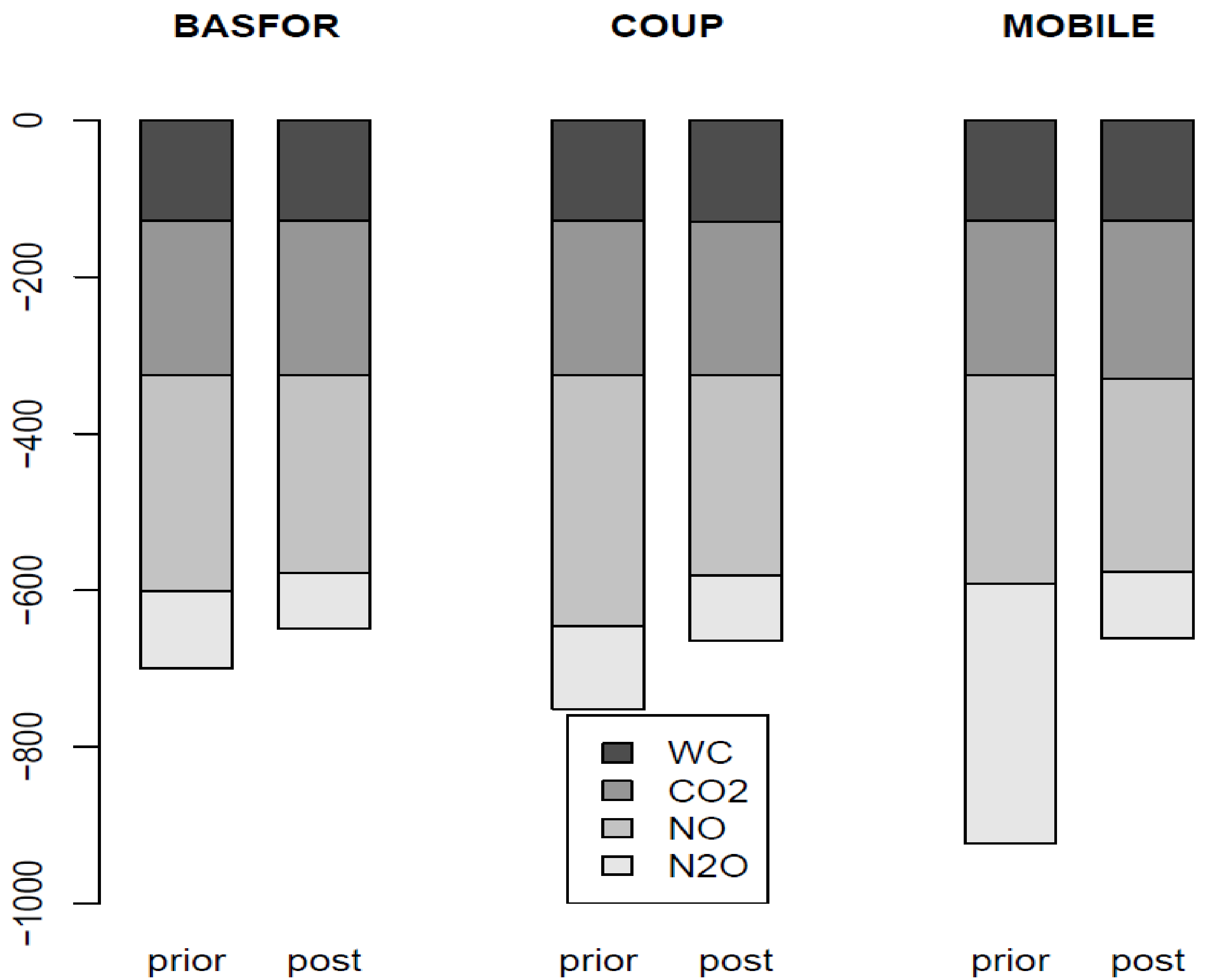
2.1 EXAMPLE: BC giving strange posterior pdf’s



Red lines: Prior pdf.

Black histograms: Posterior pdf after using data from Scots pine in Estonia.

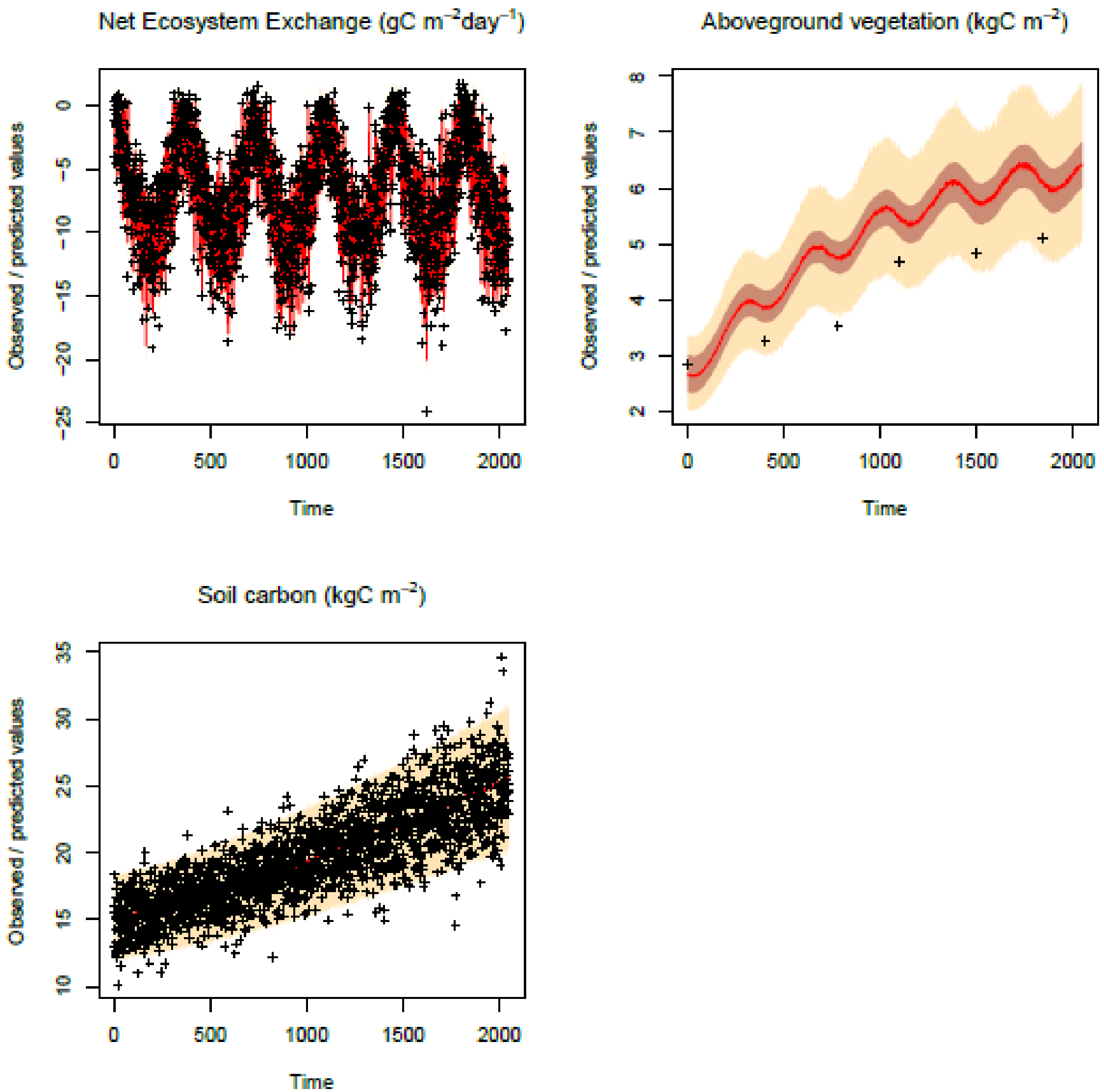
2.2 Analysis of model-data mismatch before/after BC: logL



[Van Oijen et al. 2011]

3. Analysing posterior model-data mismatch

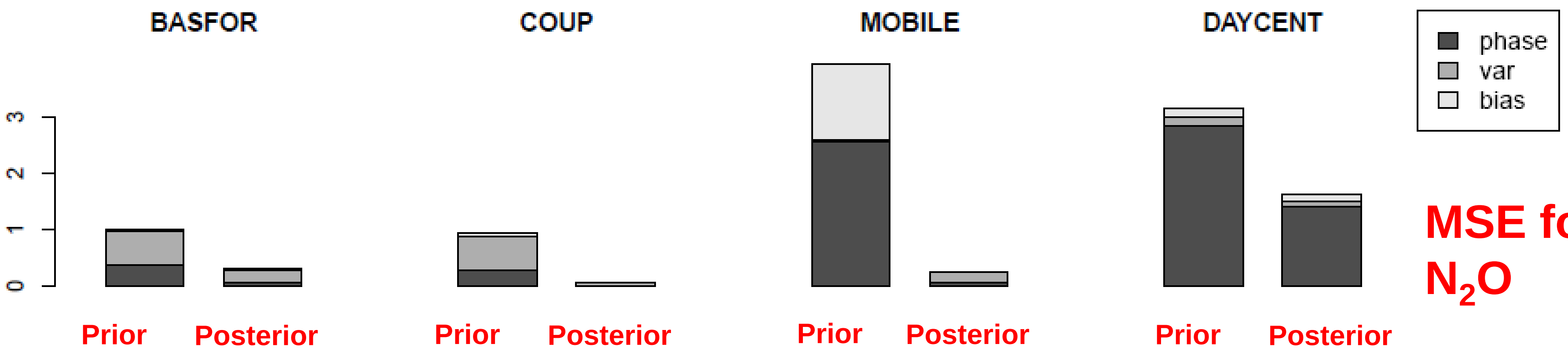
3.1 Structural model error



3.2 Analysis of model-data mismatch before/after BC: MSE

[Kobayashi & Salam, 2000, Agron. J. 92: 345-352]

$$\boxed{\begin{aligned} MSE = \overline{(M-D)^2} \;&=\; 2\sigma_M\sigma_D(1-R) \;+\; (\sigma_M-\sigma_D)^2 \;+\; (\overline{M}-\overline{D})^2 \\ &=\; phase \;+\; variance \;+\; bias \end{aligned}}$$



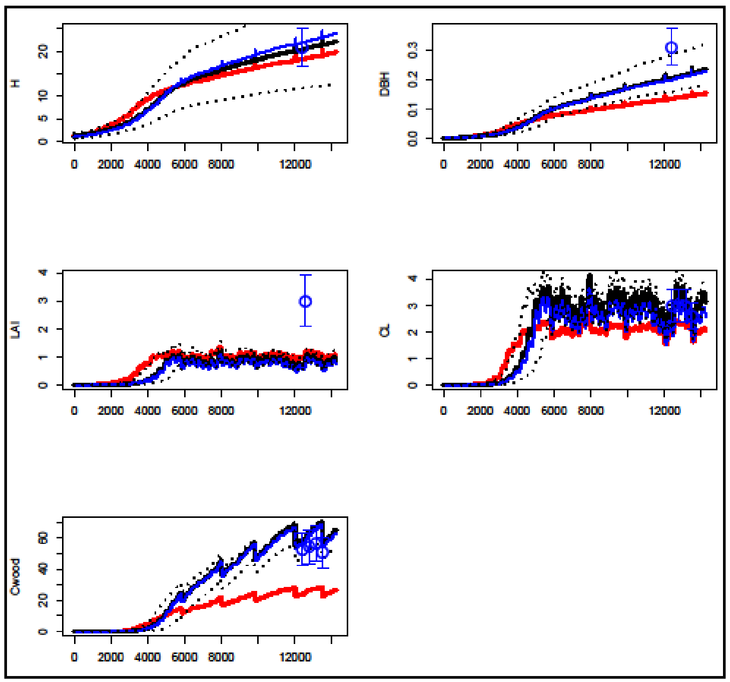
MSE for N_2O

[Van Oijen et al. 2011]

3.3 Analysis of model-data mismatch before/after BC: MSE

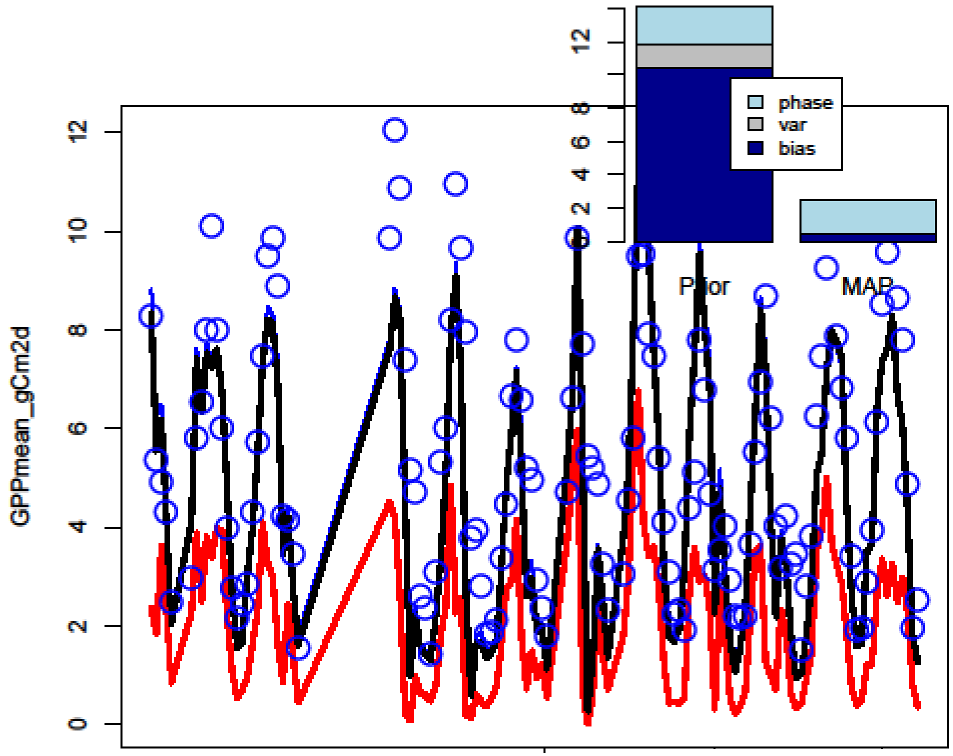
$$
\begin{aligned}
MSE = \overline{(M-D)^2} &= 2\sigma_M\sigma_D(1-R) + (\sigma_M-\sigma_D)^2 + (\overline{M}-\overline{D})^2 \\
&= phase + variance + bias
\end{aligned}
$$

TREE GROWTH



1970-2010 **1970-2010**

GPP (EDDY COVARIANCE)



1997 **2003** **2008**

Mode of prior
Mode of posterior (MAP)

3.4 Can probability theory help in diagnosis?

Diagnosing model weaknesses

Decomposing mean-squared 'error' (MSE) between a modelled and measured time series (Kobayashi & Salam 2000, van Oijen et al. 2011)

$$\mathrm{MSE} = \overline{(\mathrm{M}-\mathrm{D})^2} = (\bar{\mathrm{M}}-\bar{\mathrm{D}})^2 + (\sigma_{\mathrm{M}}-\sigma_{\mathrm{D}})^2 + 2(\sigma_{\mathrm{M}}\sigma_{\mathrm{D}})(1-r)$$

Difference	Mean	Variance	Phase
Error	Process missing?	Feedback missing?	Linked state variable missing?