TG15 manuscript

David Cameron, Mike Dietze and Marcel van Oijen

2018-09-25

Contents

1 Introduction

Main issue

- Because of the difference in cost in collecting automatic measurements such as eddy covariance data and manual measurements such as soil and plant carbon stocks, there is often two orders of magnitude difference or more imbalance between the number of measurements available from different parts of the forest ecosystem.
- Since each measurement point normally counts as an independent piece of data in the Likelihood of Bayesian calibration the influence of the sparse observations can often be overwhelmed by the higher frequency data (Cameron et al (in review 2018).

- As more and more EO data becomes available (e.g. Sentinal) this issue of extremely imbalanced datasets is likely to worsen significantly.

What is the effective information content (IC) of observations (possibly include or is this really a different issue? - discuss Mike/Marcel)

- One eddy covariance tower with 17000 measurements count this as n = 1 or n = 17000 obs?
- Aggregate 5 min data to 10 min retain essentially same IC but sample size halved!
- Spectral analysis of eddy covariance NEE data two peaks: annual(seasonal) and diurnal
- Often assume in BC that each data*point provides independent information
- If biases in data or model errors are not independent

To-date solutions generally ad hoc and/or arbitrary

- Literature review…
- Ignore but then models over-fitted and uncertainty underestimated
- Apply arbitrary weights to rebalance influence of data in BC
- Thin the number of eddy covariance obs
  - throwing away useful information

Purpose of paper:

1. To identify the issue of using unbalanced data in parameter calibration of ecosystem models using artificial experiments.
2. Develop a general methodology for identifying whether and where the issue becomes a problem.
3. Start to explore the simplest modifications to the likelihood to represent model structural error and data bias to help ameliorate the issue identified in (1).

Artificial experiments:

- Developed a very simple process-besed ecosystem model (VSEM) as a testbed for experiments. Simple enough that results are easily understood. Sufficiently complex that we can be confident that the model/data issues identified here would also been seen in more complex/realistic process-based ecosystem models.
- Calibration data model's own output to control:
  1. test influence of model and data perfection/imperfection
  2. test influence of balance/unbalance of data in the calibration

2 Methods

2.1 VSEM model

- very simple toy model. not realistic but simplified form with and similar structure to many ecosystem models.

Photosynthesis equation

$$NPP = PAR \times LUE \times (1 - \exp^{(-KEXT \times C_v)}) \tag{1}$$
$$\tag{2}$$

- PAR Photosynthetically active radiation
- LUE Light use efficiency of NPP (Ra implicit)
- KEXT Beer's law light extinction coeff
- C_v Vegetation carbon

Carbon pool state equations

$$\frac{dC_v}{dt} = A_v \times NPP \qquad - \frac{C_v}{\tau_v} \tag{3}$$
$$\frac{dC_r}{dt} = (1.0 - A_v) \times NPP \qquad - \frac{C_r}{\tau_r} \tag{4}$$
$$\frac{dC_s}{dt} = \frac{C_r}{\tau_r} + \frac{C_v}{\tau_v} \qquad - \frac{C_s}{\tau_s} \tag{5}$$

- C_v, C_r and C_s : Carbon in vegetation, root and soil pools

2.2 Bayesian Calibration

- refer to forthcoming TG13 paper
- R package used BayesianTools
- DREAMzs algorithm

2.3 Idealised experiments with virtual data from VSEM

- advantage we know what the truth is.
- balanced versus unbalanced
- perfect model and model with known error
- data with and without known bias
- include additive and multiplicative parameters to represent structural errors in the model and systematic biases in the data.

3 Identifying the issue

3.1 Perfect model and balanced data

- refer to Fig. (1) and (2)
- ‘true’ parameters largely found in posterior
- posterior parameter controlling data error close to 0.1 coefficient of variance originally imposed to create observations from ‘true’ model output.
- 50% quantile red line very close to green ‘truth’ line
- posterior very narrow marked by 2.5% and 97.5% quantiles
- most data within predictive interval

3.2 Perfect model and unbalanced data

- refer to Fig. (3) and (4)
- despite unbalanced data
  - parameter values still close to true values in posterior
  - model output still close to truth even for vegetative carbon
  - although now greater uncertainty for Cv then for balanced calibration

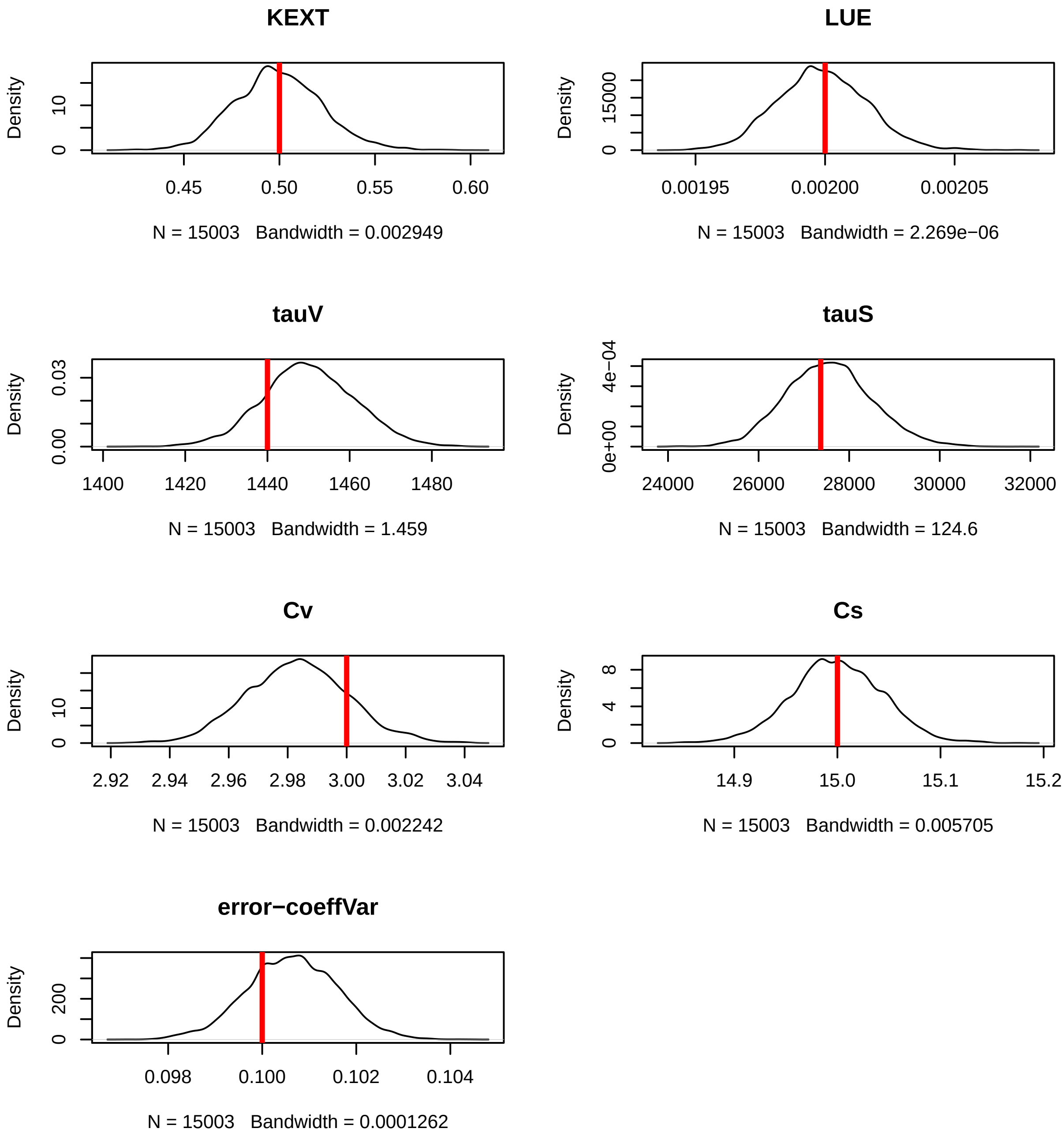

Figure 1: Perfect model, balanced data (NEE, Cv, Cs: 2048 obs). Marginal posterior distribution of model parameters and intital states. The red line marks the 'true' parameter values.

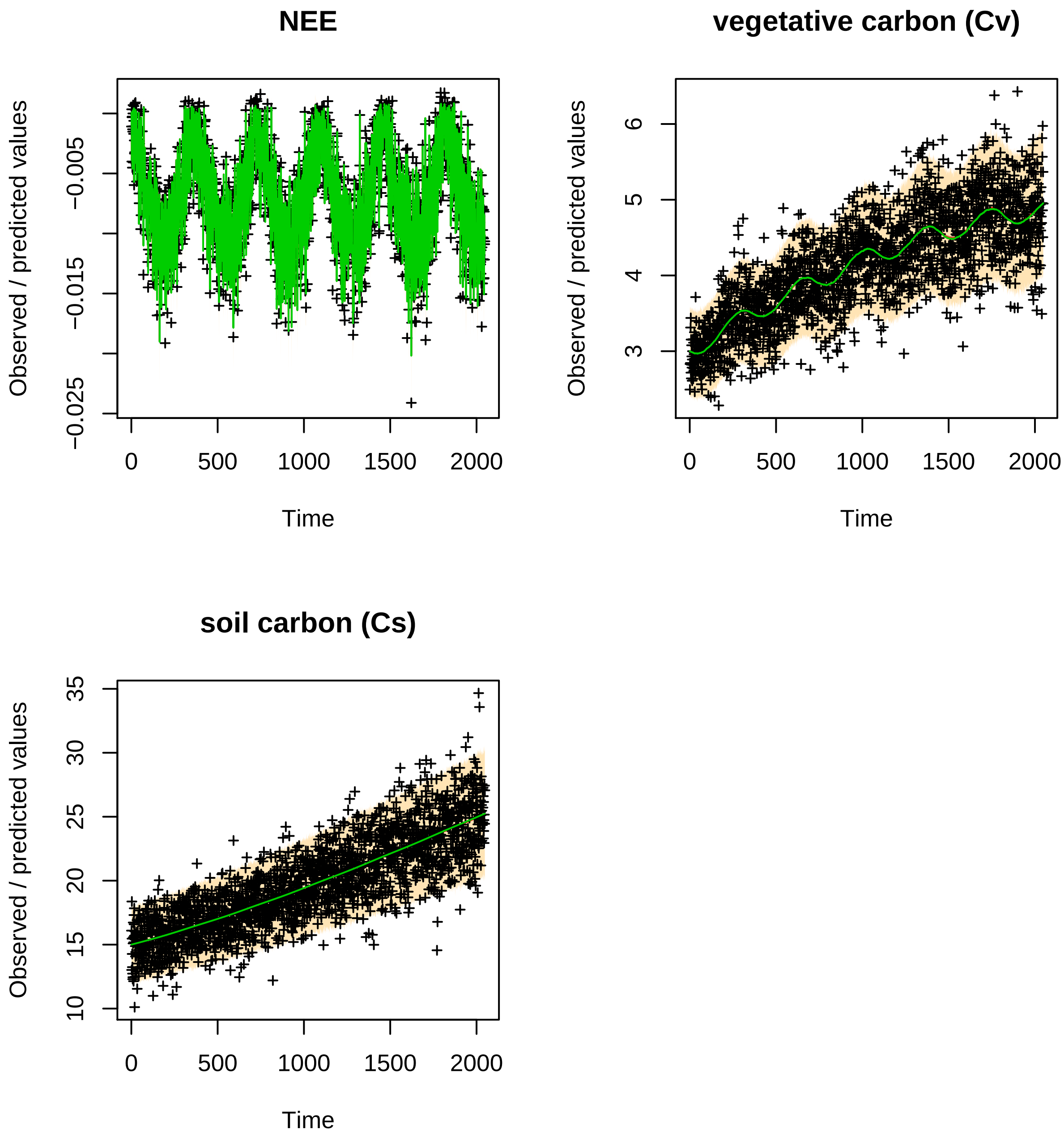

Figure 2: Perfect model, balanced data (NEE, Cv, Cs: 2048 obs). Observations included in the calibration marked with a '+'. Red line 50% quantile posterior distribution. Green line is the 'true' model output. Dark brown shading 2.5% 97.5% quantile posterior distribution. Light brown shading 2.5% 97.5% predictive interval.

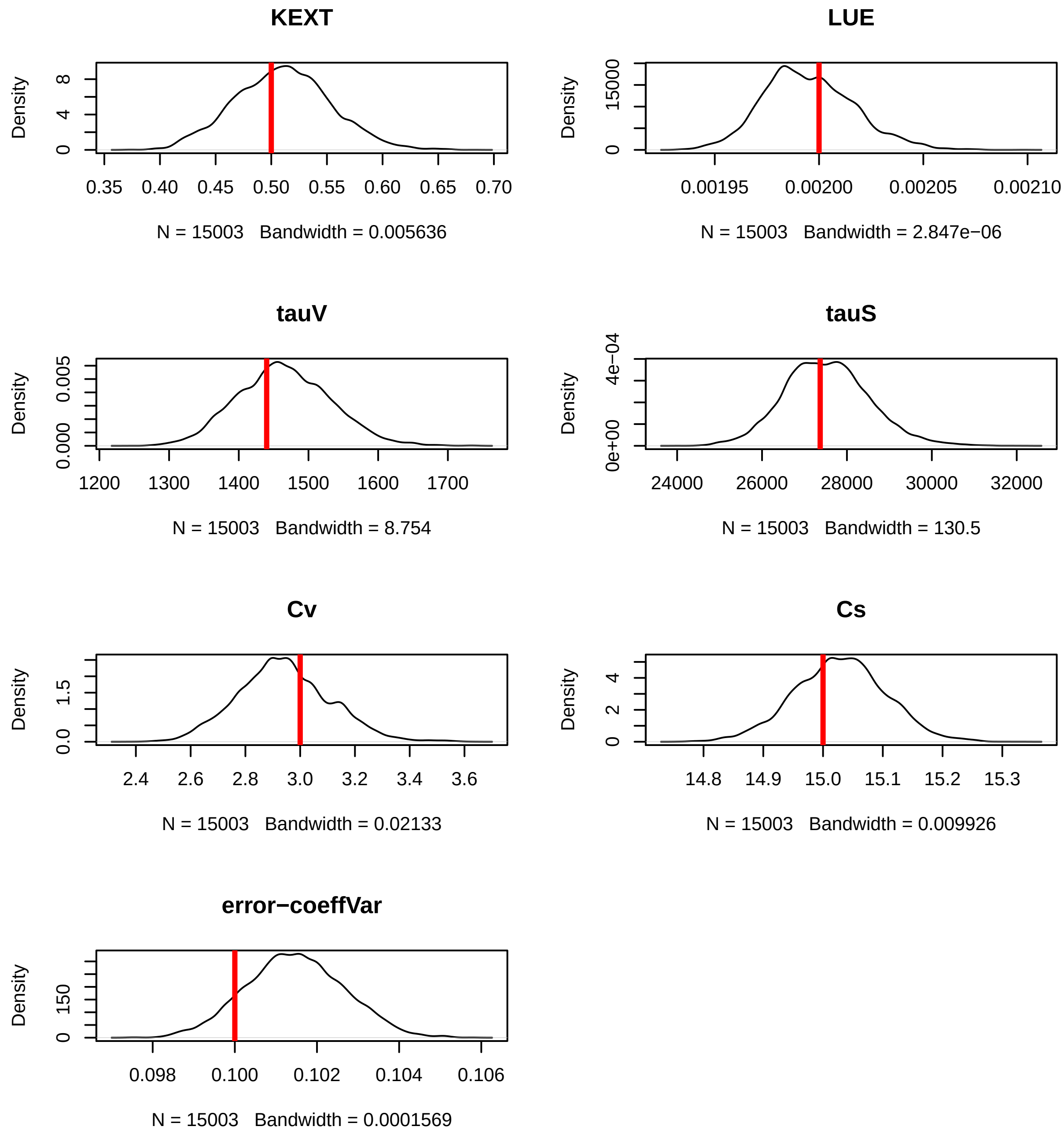

Figure 3: Perfect model, unbalanced data (NEE, Cs: 2048 obs, Cv: 6 obs). Marginal posterior distribution of model parameters and intital states. The red line marks the 'true' parameter values.

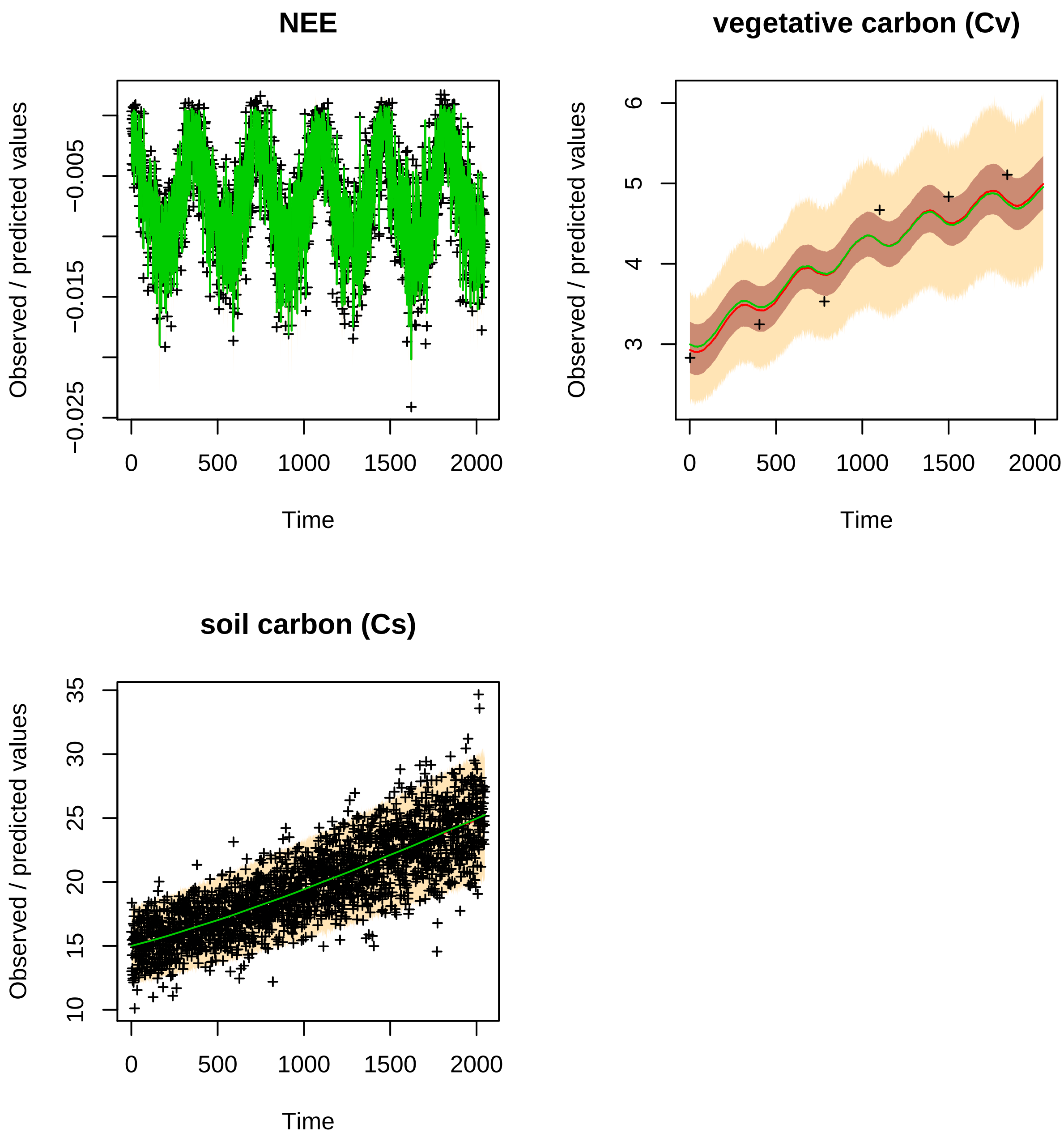

Figure 4: Perfect model, unbalanced data (NEE, Cs: 2048 obs, Cv: 6 obs). Observations included in the calibration marked with a '+'. Red line 50% quantile posterior distribution. Green line is the 'true' model output. Dark brown shading 2.5% 97.5% quantile posterior distribution. Light brown shading 2.5% 97.5% predictive interval.

3.3 Model with error and balanced data

- refer to Fig. (5) and (6)
- Root pool is effectively removed from the model by initialising pool to zero and setting allocation to roots to zero. The loss of the root pool has introduced a significant structural error to the model.
- Parameter posteriors now quite far away from 'true' values.
  - Especially parameter which controls turnover of vegetation so that rate of turnover to soil is now more than doubled.
- Parameters calibration seems to have somewhat 'absorbed' the model structural error so that
  - outputs where there was data included in the BC are still not too far away from the 'truth' line.
  - Cv now has too much variability but average increase not too bad.
  - most data still within predictive interval.

3.4 Model with error and unbalanced data

- refer to Fig. (7) and (8)
- KEXT smaller, LUE larger, Cv larger, tauS smaller, tauV much larger, Cs much larger
  - Generally parameters closer to their 'true' value. Less 'absorbing' of the model structural error.
- NEE and soil carbon pools look largely unchanged to balanced run and close to data and predictive interval.
- significant change to soil vegetation pool
  - data outside of posterior and close to one edge of predictive interval
  - departure from 'truth' line growing in time
- General sense is that six vegetation data points are being somewhat ignored by the calibration in favour of the more plentiful NEE and soil carbon data.

3.5 Perfect model and balanced data with a multiplicative bias

- refer to Fig. (9) and (10)
- soil carbon pool data multiplied by two to represent data that has a systematic bias
- Parameters KEXT larger, tauV smaller, tauS much smaller, Cs much larger
- as for model error parameter calibration 'absorb' the error
  - initial value of soil pool parameter more than double 'true' value
  - parameter controlling turnover time of soil approximately doubled to keep soil carbon pool high.
- calibrated outputs again reasonably close to observations for NEE and Cv
  - slope of soil carbon pool with time too shallow (so some data outside predictive interval) but compares well with 'true' line (influence of other more accurate NEE and vegetative carbon data?)
- data and calibrated output of soil carbon greater than 'true' line by a factor of two as might be expected to match calibration data.

3.6 Perfect model and unbalanced data with a multiplicative bias

- refer to Fig. (11) and (12)
- most parameters are far away from their 'true' values
- soil carbon improved fit to data versus previous balanced data calibration
- six vegetative carbon data points effectively ignored and overpowered by NEE and soil carbon data in the calibration.

3.7 Model with error and unbalanced data with a multiplicative bias

- refer to Fig. (13) and (14)

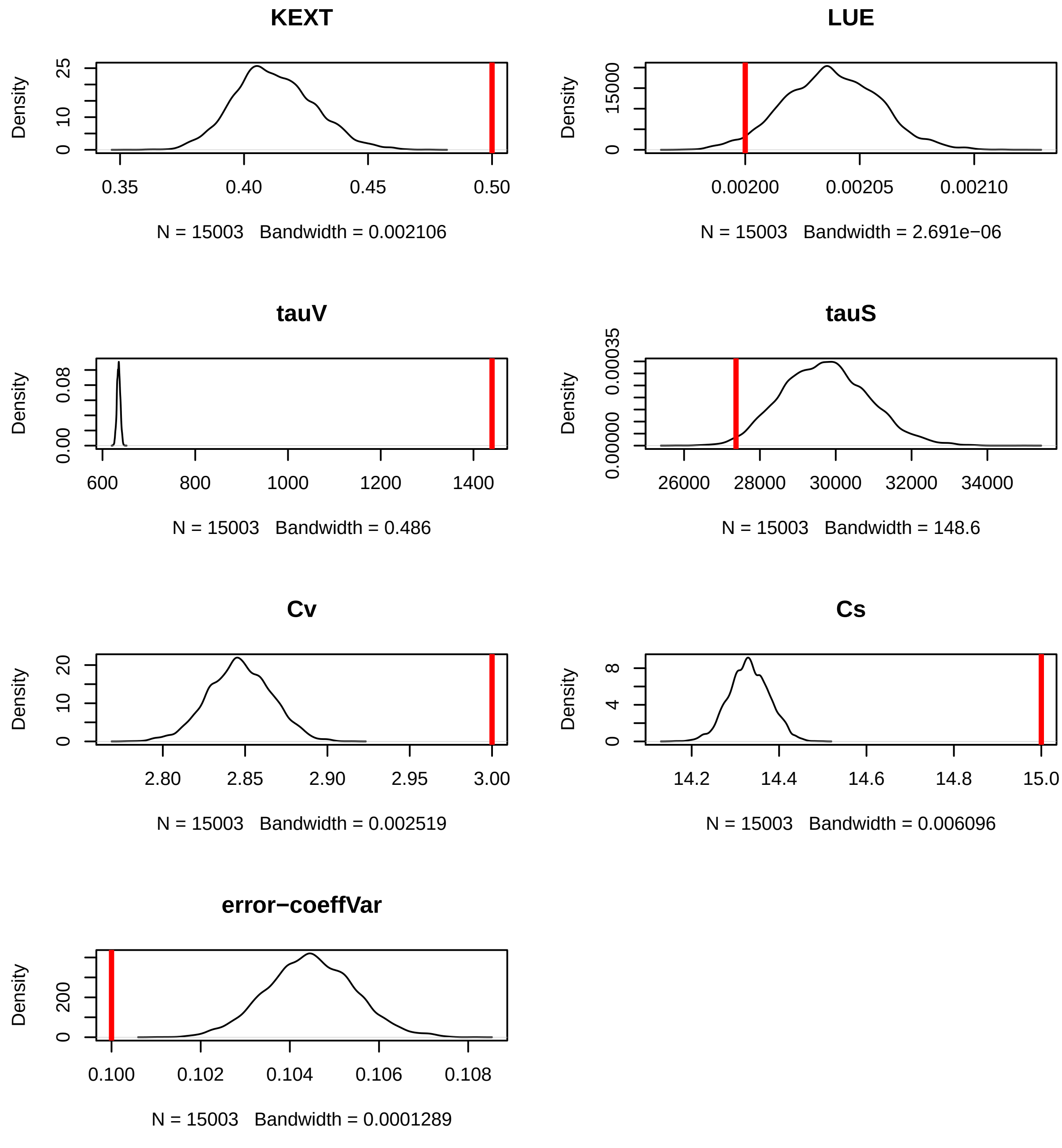

Figure 5: Model with error, balanced data. Marginal posterior distribution of model parameters and intital states. The red line marks the 'true' parameter values.

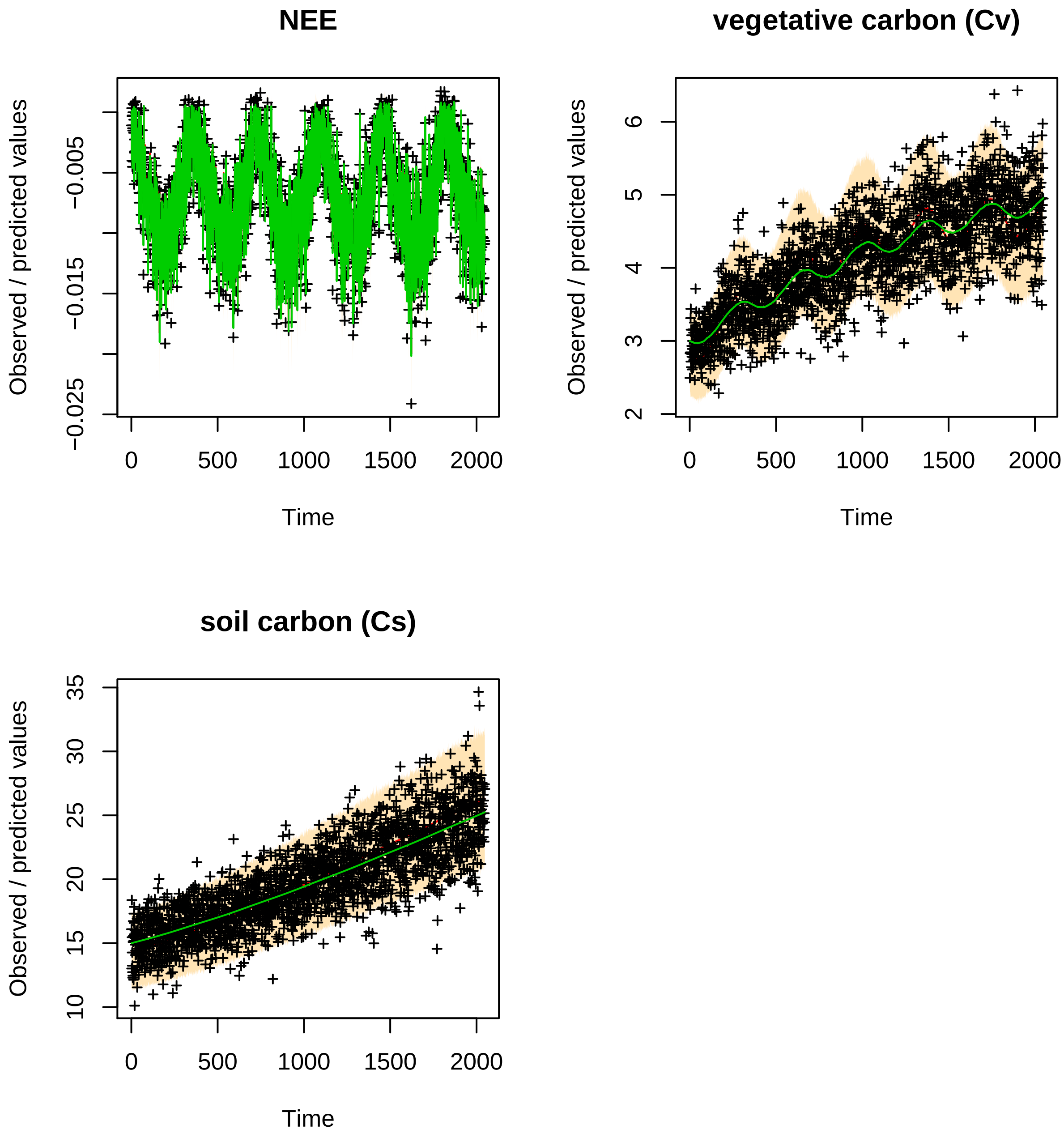

Figure 6: Model with error, balanced data. Observations included in the calibration marked with a '+'. Red line 50% quantile posterior distribution. Green line is the 'true' model output. Dark brown shading 2.5% 97.5% quantile posterior distribution. Light brown shading 2.5% 97.5% predictive interval.

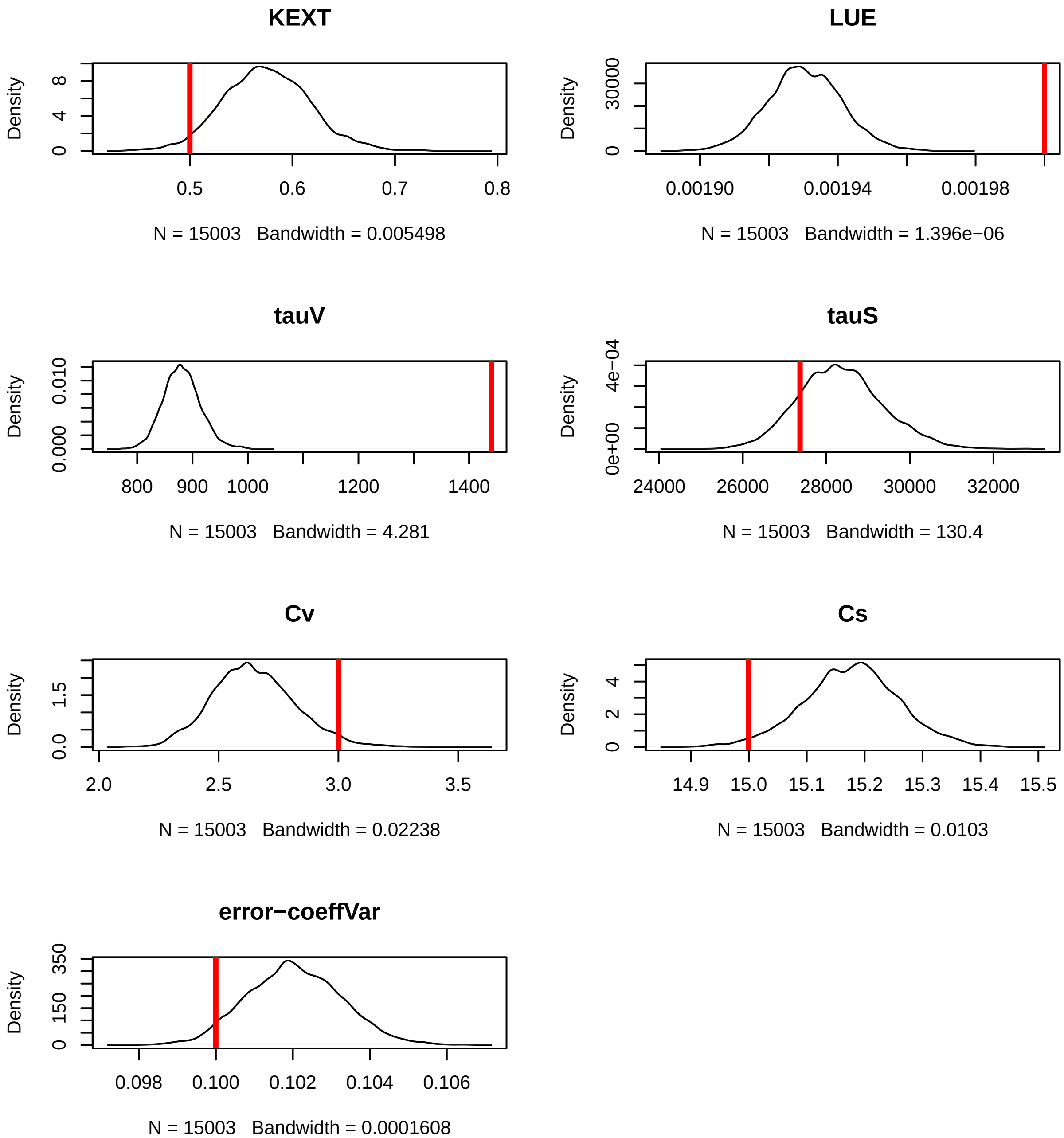

Figure 7: Model with error, unbalanced data (NEE, Cs: 2048 obs, Cv: 6 obs). Marginal posterior distribution of model parameters and intital states. The red line marks the 'true' parameter values.

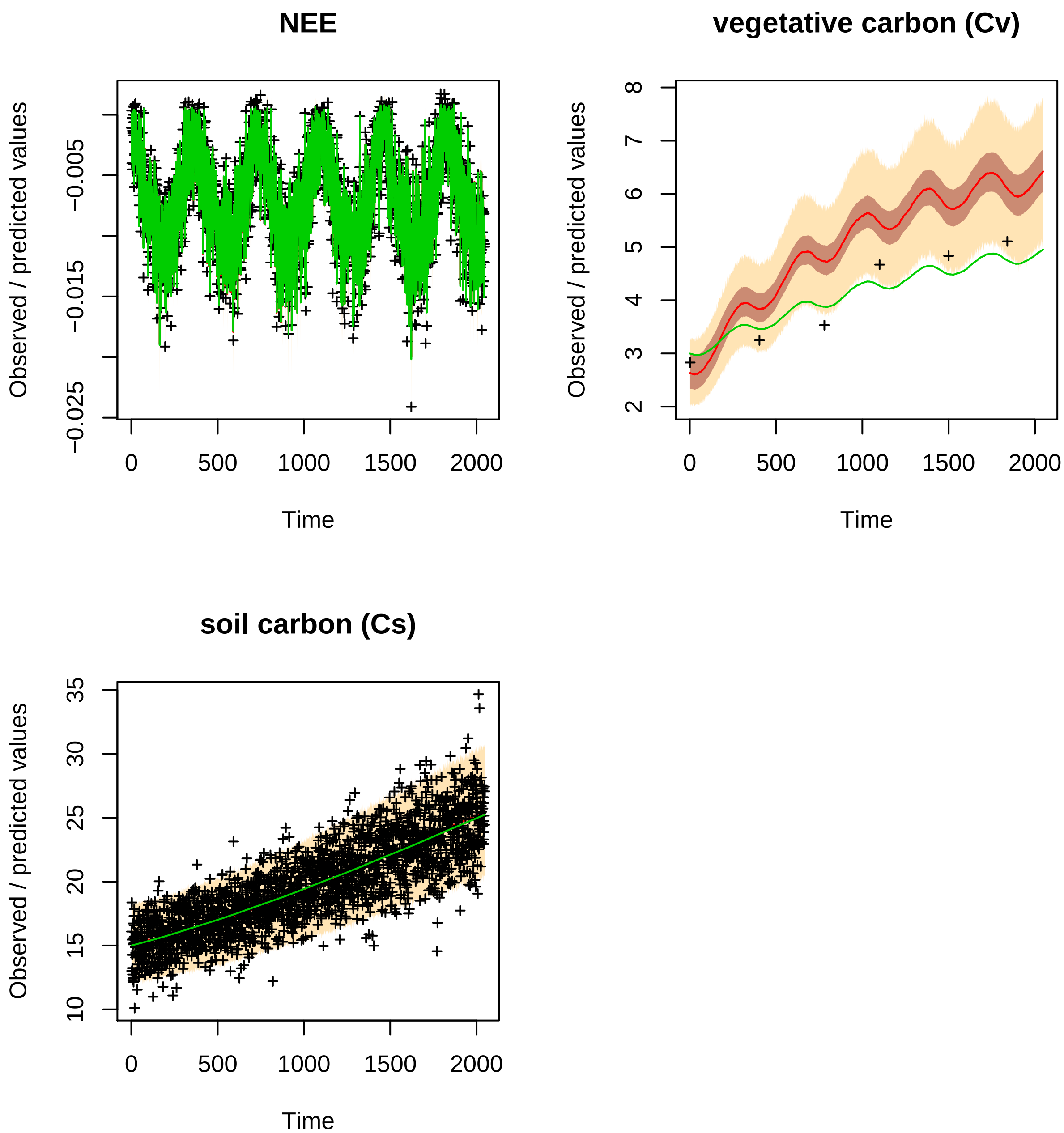

Figure 8: Model with error, unbalanced data (NEE, Cs: 2048 obs, Cv: 6 obs). Observations included in the calibration marked with a '+'. Red line 50% quantile posterior distribution. Green line is the 'true' model output. Dark brown shading 2.5% 97.5% quantile posterior distribution. Light brown shading 2.5% 97.5% predictive interval.

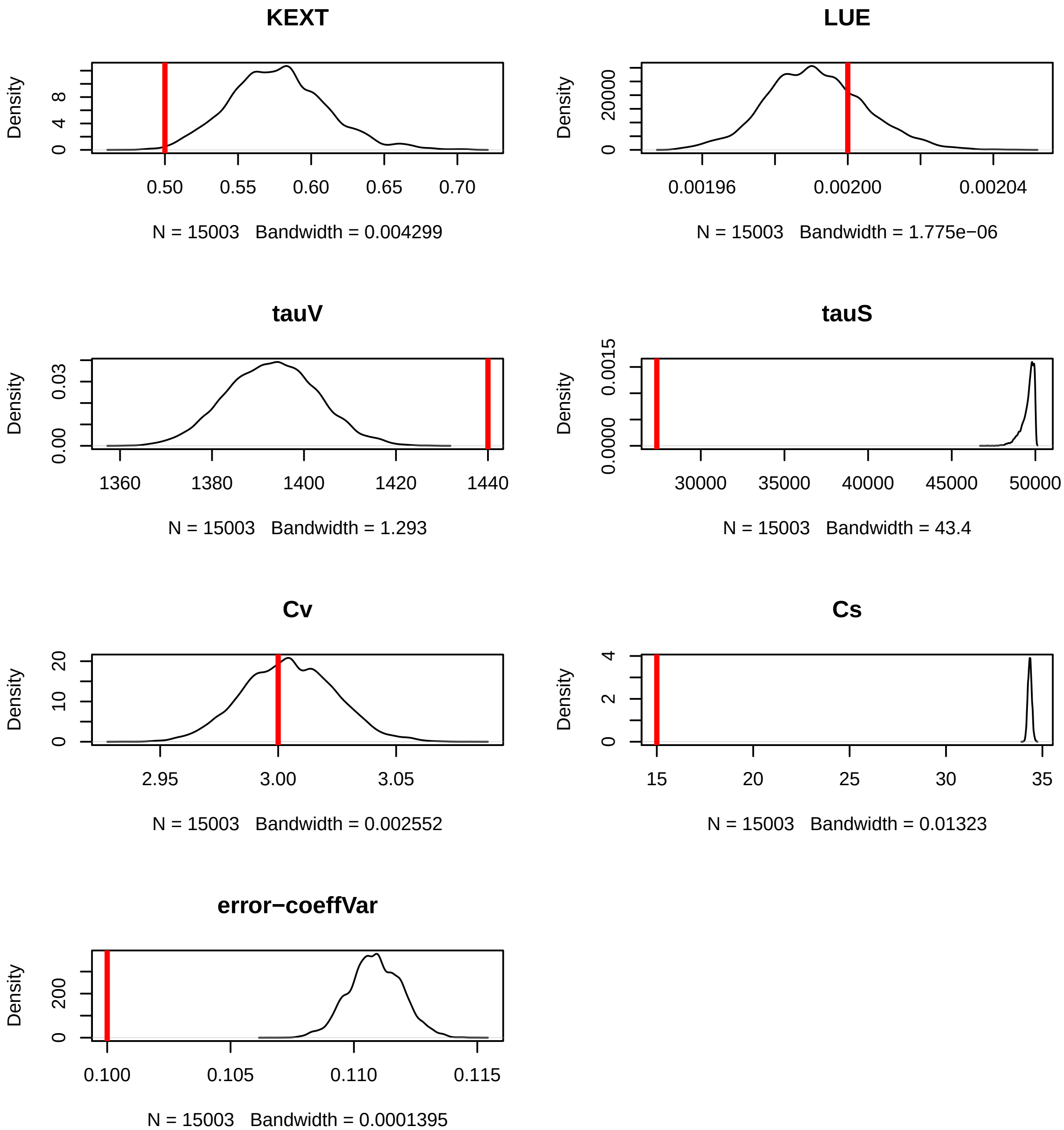

Figure 9: Perfect model and balanced data with a multiplicative bias. Marginal posterior distribution of model parameters and intital states. The red line marks the 'true' parameter values.

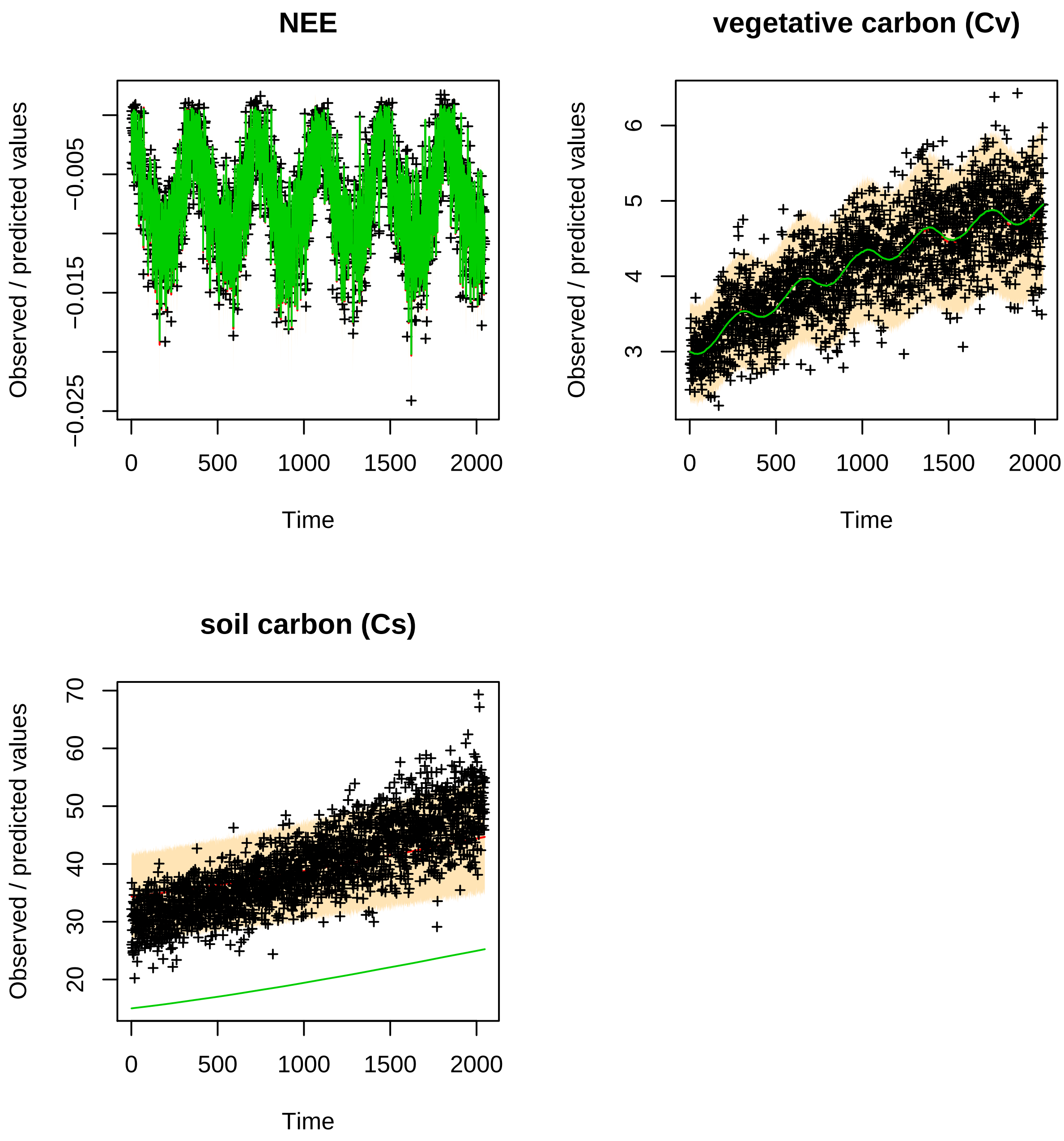

Figure 10: Perfect model and balanced data with a multiplicative bias. Observations included in the calibration marked with a '+'. Red line 50% quantile posterior distribution. Green line is the 'true' model output. Dark brown shading 2.5% 97.5% quantile posterior distribution. Light brown shading 2.5% 97.5% predictive interval.

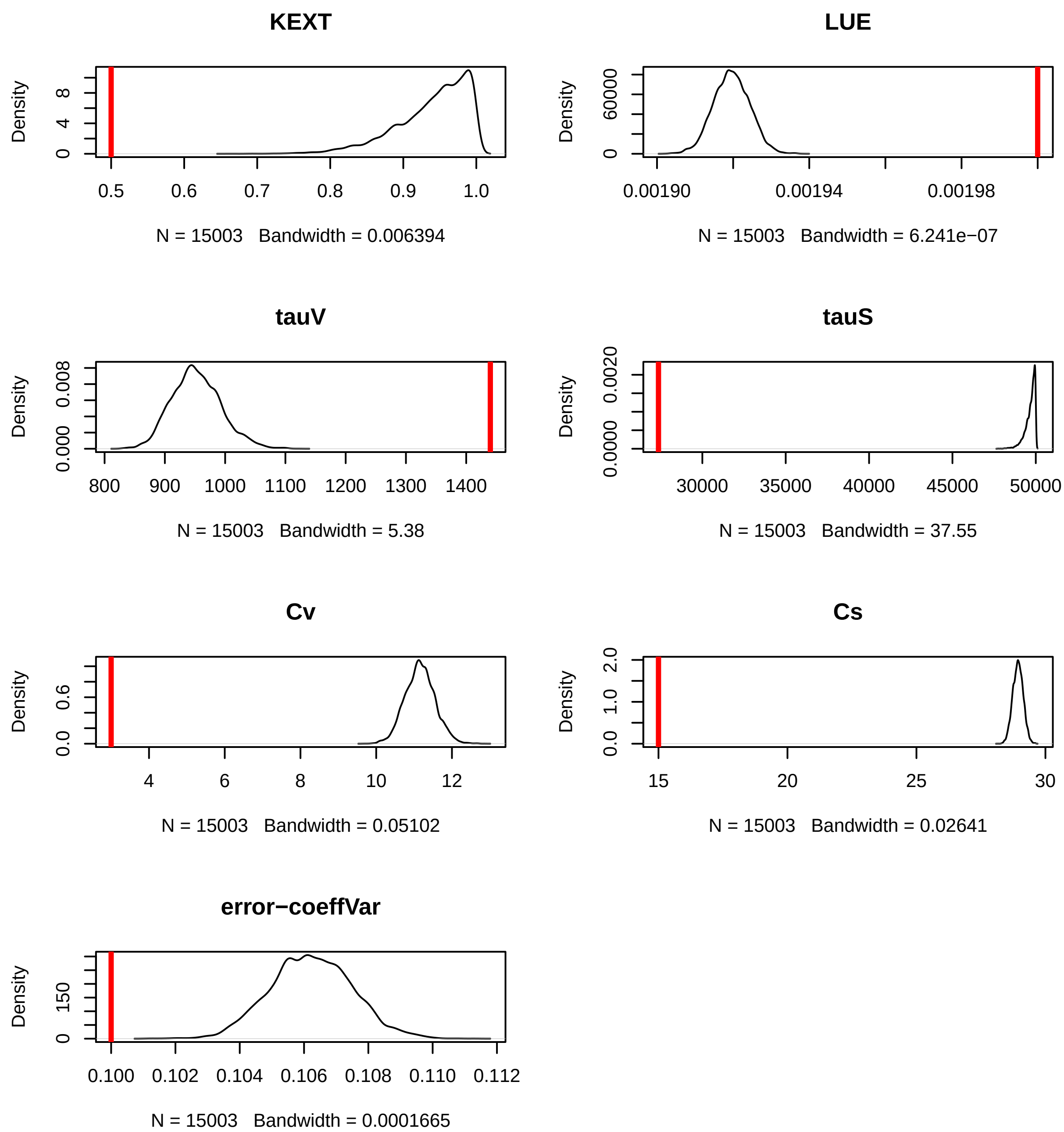

Figure 11: Perfect model and unbalanced data with a multiplicative bias. Marginal posterior distribution of model parameters and intital states. The red line marks the ‘true’ parameter values.

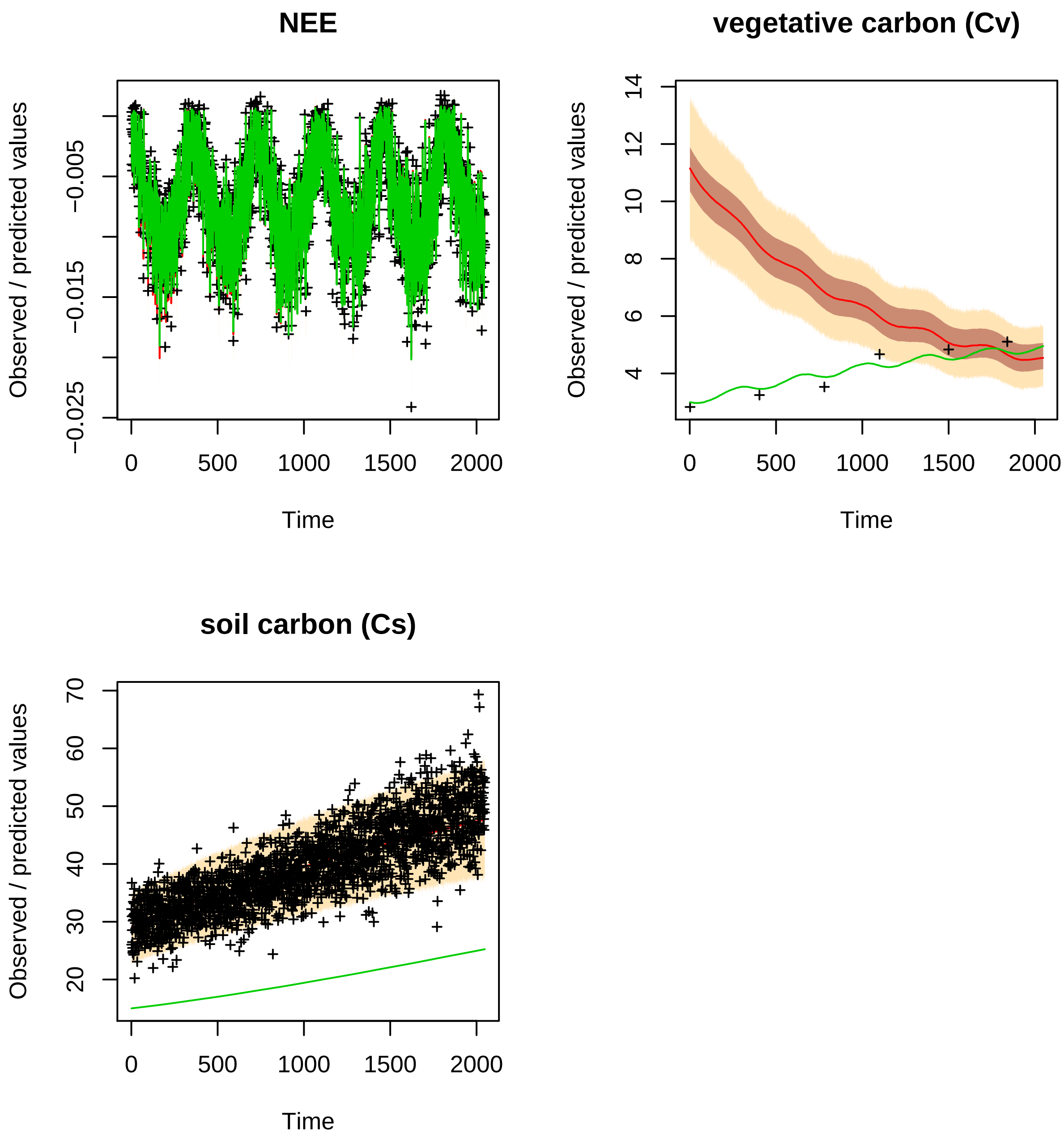

Figure 12: Perfect model and unbalanced data with a multiplicative bias. Observations included in the calibration marked with a '+'. Red line 50% quantile posterior distribution. Green line is the 'true' model output. Dark brown shading 2.5% 97.5% quantile posterior distribution. Light brown shading 2.5% 97.5% predictive interval.

- similar to calibration with bias in data above.
- vegetative carbon perhaps very slightly closer to the data than when only data bias was present. This may indicate compensating errors in the model and data.

4 Diagnosing the issue

4.1 Comparing model output with virtual data as truth.

Moving on from identifying the issue in the previous section, here we develop a tool for helping to diagnose at what point and to what extent having unbalanced data in Bayesian calibration (BC) becomes an issue when models and data are imperfect.

This is done by running a number of calibrations with perfect and imperfect models where the quantity and imbalance of data used increases with each calibration. Here we chose an increasing power series of two (2^3,2^4 ... 2^{11}) for the increase in the quantity of calibration data; eight calibrations in all. In the balanced data BC case, quantities of NEE, vegetative carbon and soil carbon data included in the BC all increased in tandem in each subsequent calibration. For the unbalanced BC case, NEE and soil carbon data increased as before but the quantity of vegetative carbon data included in the BC was held fixed at six data points for each of the eight calibrations. After running the calibrations the VSEM was rerun with the maximum a posteriori (MAP) vector and the RMS difference with the 'true' data was calculated and plotted (Fig. 15).

The figure shows broad similarity in results except for vegetative carbon case when the model has an error and where there is an imbalanced in calibration data. In general, the RMS difference has a tendency to go down as the quantity of data included in calibration increases. There is also a marked grouping of results with the perfect model getting closer to the data than the model with the error, as might be expected. For NEE and soil carbon with an imperfect model, the unbalanced calibration gets closer to the data than the balanced calibration especially as the quantity of calibration data increases. This is in marked contrast to vegetative carbon where RMS differences increase significantly as quantity of calibration data increases when the model has an error and when there is an imbalanced in calibration data. This increase in RMS difference for vegetative carbon occurs in tandem with the decreases noted already from NEE and soil carbon. This signature of increasing RMS difference for the low quantity data output versus the decreasing RMS difference for the high quantity can be used to diagnose when large imbalances in calibrations data with imperfect models and data start to become an issue. In this case, it appears after the quantity of data included in the calibration exceeds 32 but this will be different for each model, likelihood function and for each dataset used in calibrations.

4.2 Comparing model output against "obervations"

The diagnosis made in the previous section had the benefit of access to the 'true' data and a perfect model. Unfortunately this is never the case for real world ecological model calibrations. Therefore, here we have repeated the previous graph Fig.(15) with just the imperfect model and the imbalanced calibration, but with RMS differences now calculated against observations (NEE: 2048 points, vegetative carbon: 6 points, soil carbon: 2048 points) (Fig. 16). While there are clear differences in the RMS values versus the previous graph, as might be expected, the broad-scale signature of increasing RMS difference for vegetative carbon and decreasing RMS difference for NEE and soil carbon is retained. As before, this graph can be used to diagnose when the imbalanced in data is starting to interact with the erroneous model. In this case, as before, this occurs for a data quantity greater than 32.

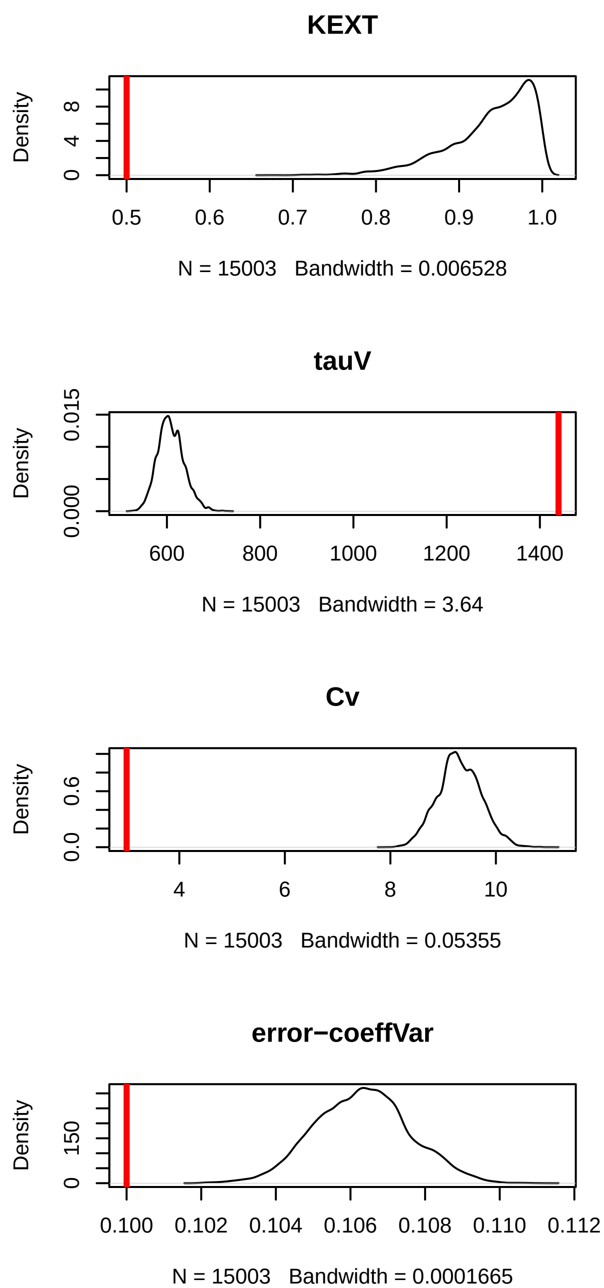

Figure 13: Model with error and unbalanced data with a multiplicative bias. Marginal posterior distribution of model parameters and intital states. The red line marks the 'true' parameter values.

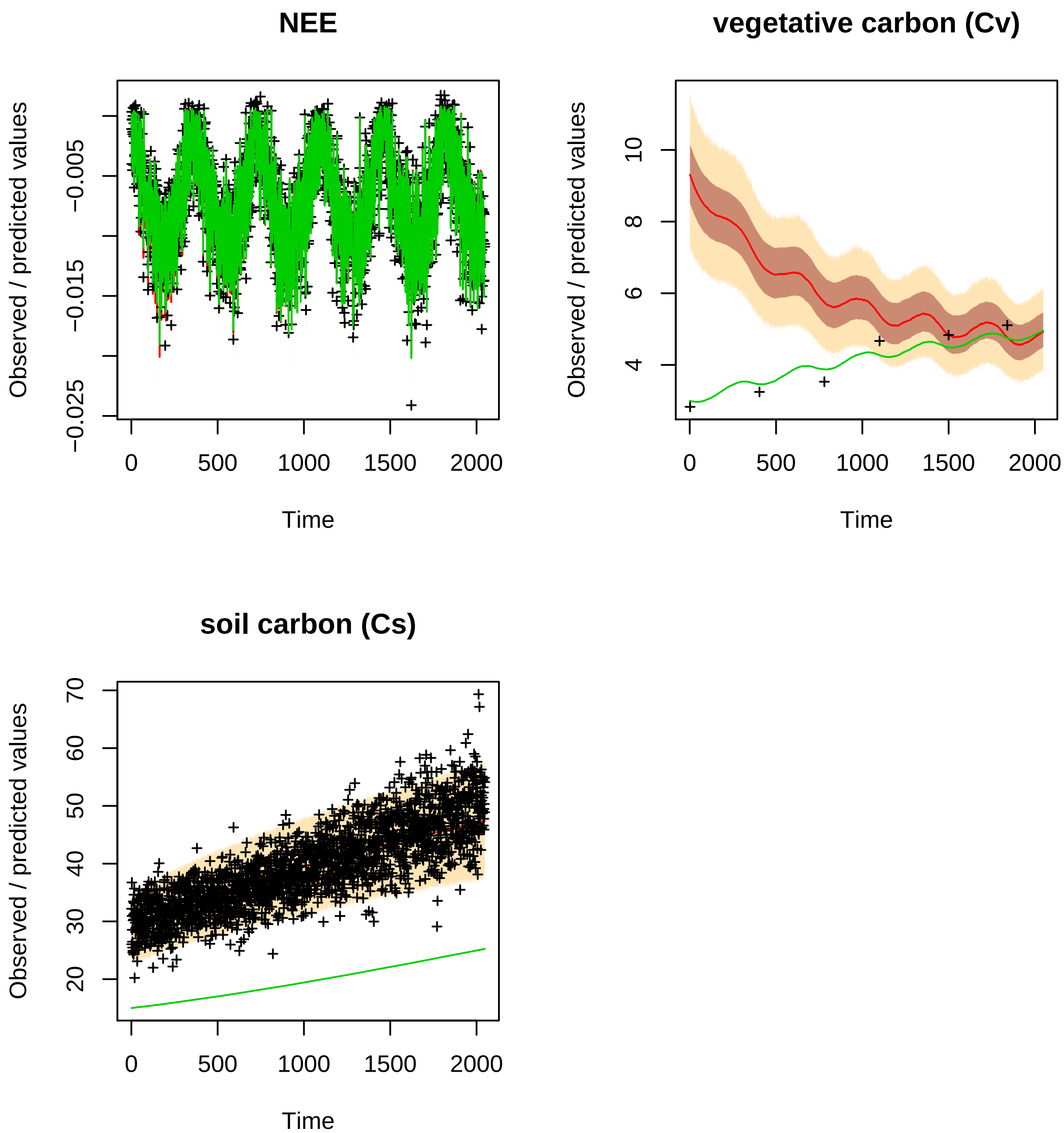

Figure 14: Model with error and unbalanced data with a multiplicative bias. Observations included in the calibration marked with a '+'. Red line 50% quantile posterior distribution. Green line is the 'true' model output. Dark brown shading 2.5% 97.5% quantile posterior distribution. Light brown shading 2.5% 97.5% predictive interval.

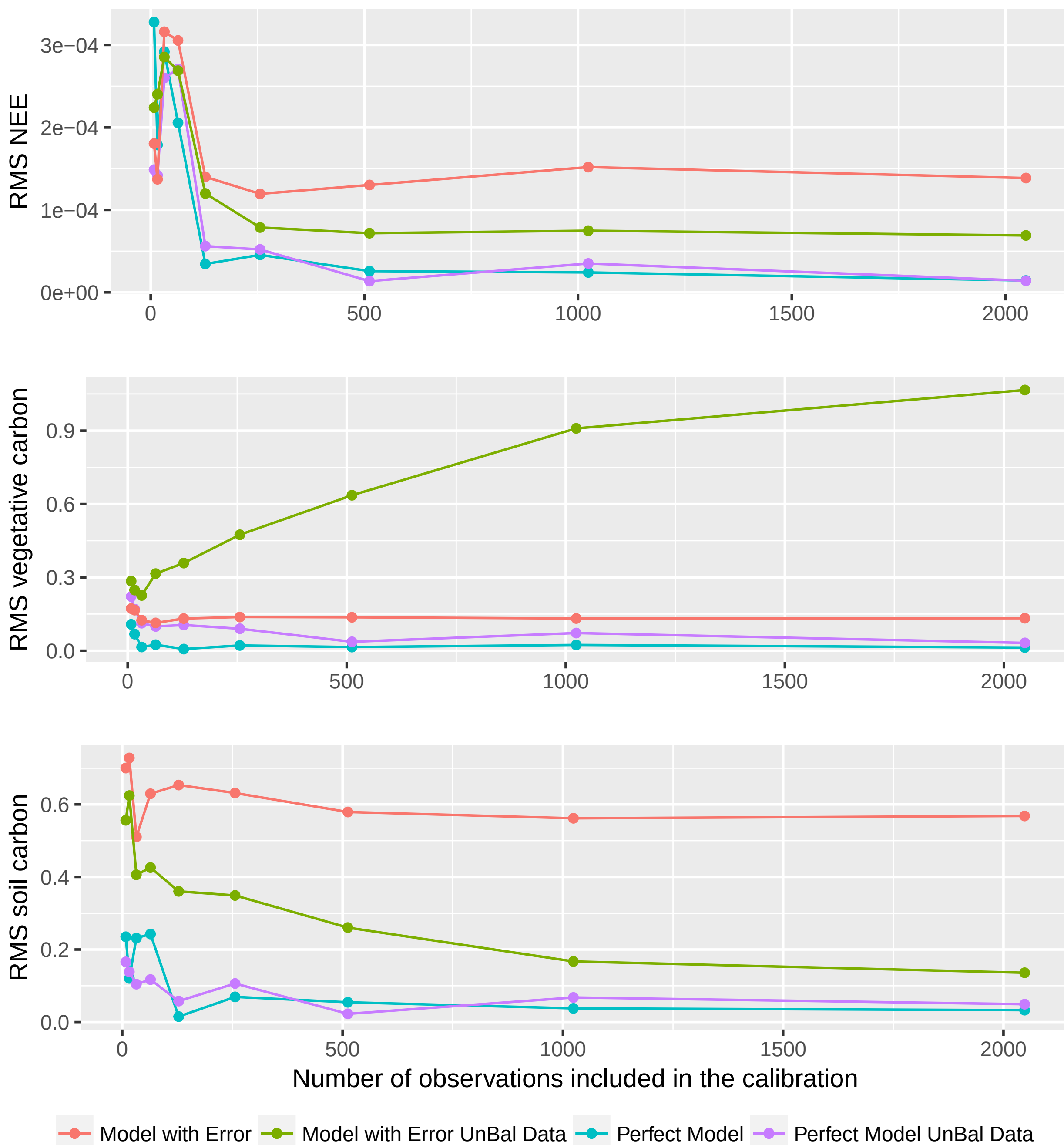

Figure 15: Each point in the three graphs (NEE, vegetative carbon, and soil carbon) represents the RMS difference between the VSEM model and the 'truth' run with different maximum a posteriori (MAP) vectors. The MAP vector at each point is obtained from a Bayesian calibration (BC) where the quantity of data included in the BC increases in a sequence along the x-axis following the exponentiation of base two. For the balanced calibration case (red and cyan) vegetative carbon data increases in tandem with NEE and soil carbon. For the unbalanced calibration case (green and purple) the quantity of vegetative carbon data is held fixed at six data values for each point along the x-axis. The VSEM model is either 'perfect' (cyan and purple) or has a known error (red and green) relative to the 'true' data that was derived from it.

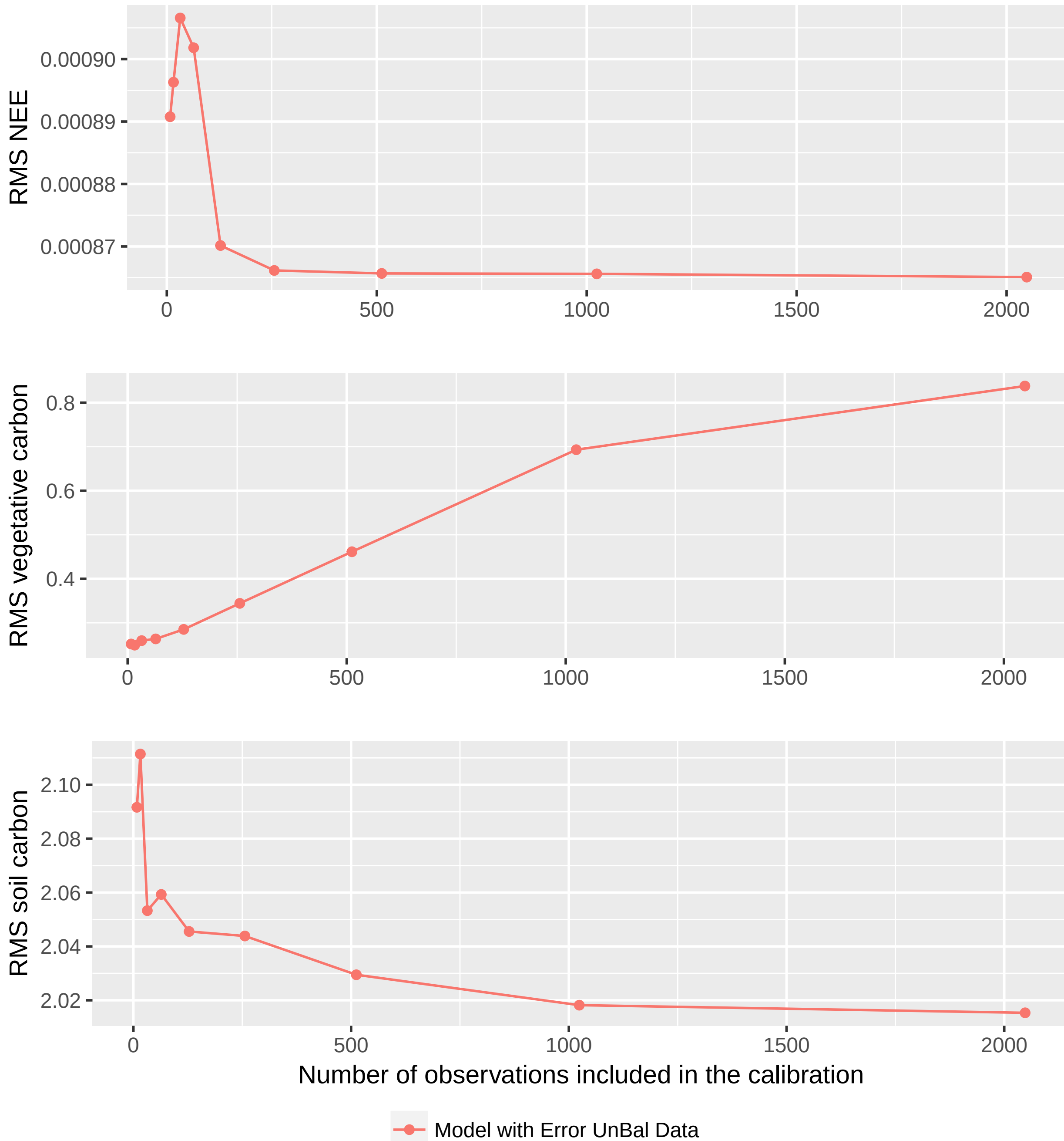

Figure 16: Each point in the three graphs (NEE, vegetative carbon, and soil carbon) represents the RMS difference between the VSEM model and virtual observations run with different maximum a posteriori (MAP) vectors. The MAP vector at each point is obtained from a Bayesian calibration (BC) where the quantity of data included in the BC for NEE and soil carbon increases in a sequence along the x-axis following the exponentiation of base two. The quantity of vegetaive carbon data is held fixed at six for all points in the graphs. The VSEM model used has a known error relative to the virtual observations that was derived from it.

5 Changes to the Likelihood to represent model and data errors

5.1 Model with error and unbalanced perfect data with additive and multiplicative parameters to represent model error.

- refer to Fig. (17) and (18)
- KEXT, LUE, Cv, Cs and error-coeffVar are now significantly closer to the 'true' values. tauS is not but is much more uncertain.
- vegetative carbon much improved
  - 5 out of 6 data points are now inside the posterior confidence interval
  - 50% quantile line now much closer to the 'true' line.

5.2 Perfect model and and unbalanced data with a multiplicative bias and additive and multiplicative parameters to represent the bias.

- refer to Fig. (19) and (20)
- many model parameters (KEXT, LUE, tauV, tauS, initial Cv) much closer to 'true' values
- multiplicative bias multiplication parameter modmultCs centred around 2.25 which is close to the multiplication factor applied to the data.
- NEE and soil carbon close to the data and within the predictive interval.
- vegetative carbon pool much improved with all data points covered by the posterior credible interval.

5.3 Model with error and and unbalanced data with a multiplicative bias and additive and multiplicative parameters to represent model error and the data bias.

- refer to Fig. (21) and (22)
- as above many parameters improved (KEXT, LUE, tauS, initial Cv)
- modmultCs value centered ~1.8 compromise between value in Fig. (19) for bias only and Fig. (17) model error only.
- vegetative pool much improved with
  - 5 out of 6 data points within posterior credible interval
  - 50% quantile line now much closer to the 'true' line.
  - similar to Fig. (18)

6 Discussion

6.1 Identifying the issue with unbalanced dataset BC

- Unbalanced data are not an issue though uncertainty is larger.
- For a model with a significant structural error or systematic bias in the calibration data parameters 'absorb' the error so that model output is not too far away from the data.
- With a model structural error or data with a systematic bias general sense is that sparse data are somewhat ignored in the BC in favour of the plentiful data.
- This is what we often observe with unbalanced datasets in BC. For example, Cameron et al (2018) but results here make it apparent that the issue is the presence of the structural error in the model rather than that the calibration data are unbalanced.

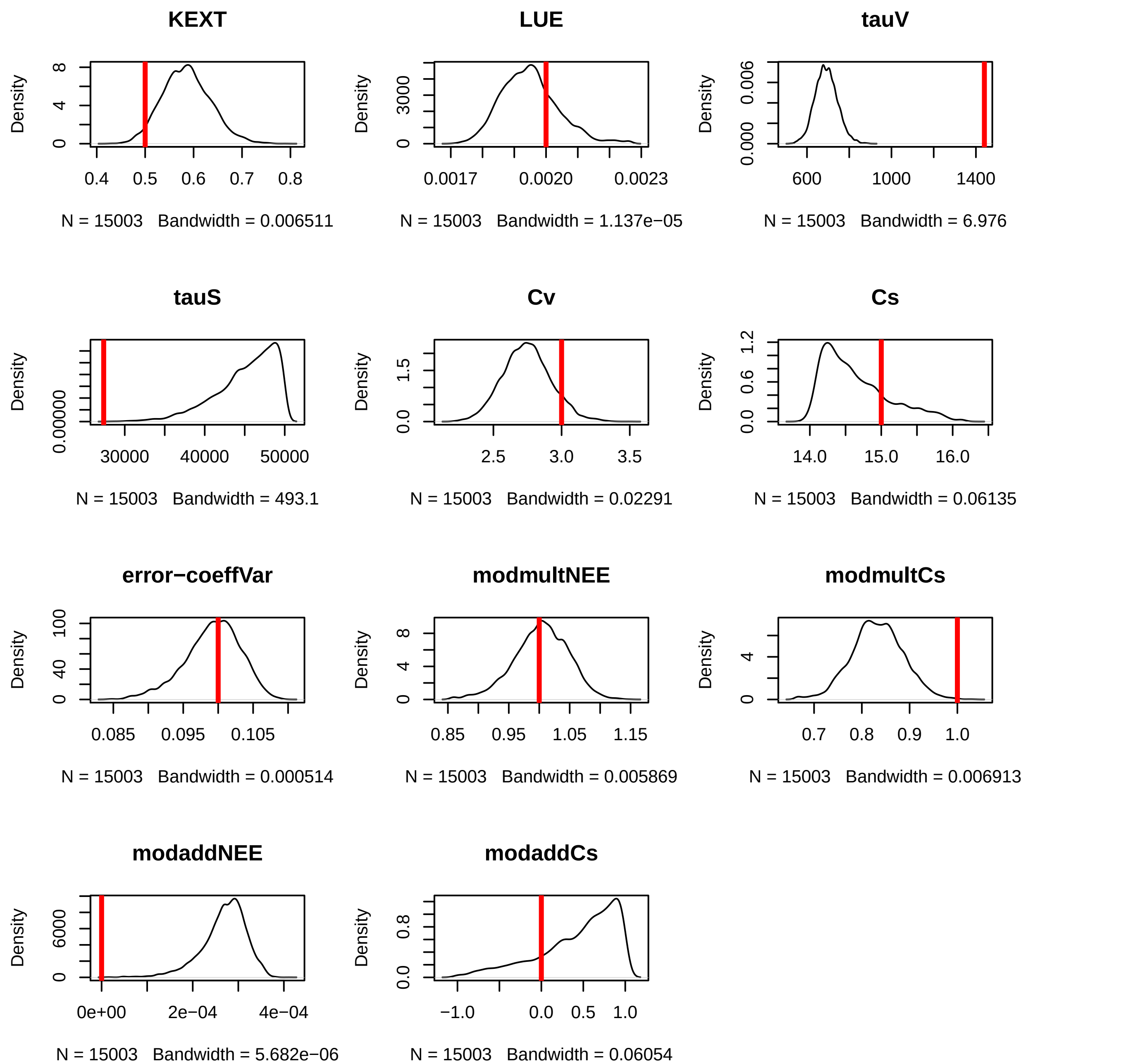

Figure 17: Model with error and unbalanced data with additive and multiplicative parameters to represent model error. Marginal posterior distribution of model parameters and intital states. The red line marks the 'true' parameter values.

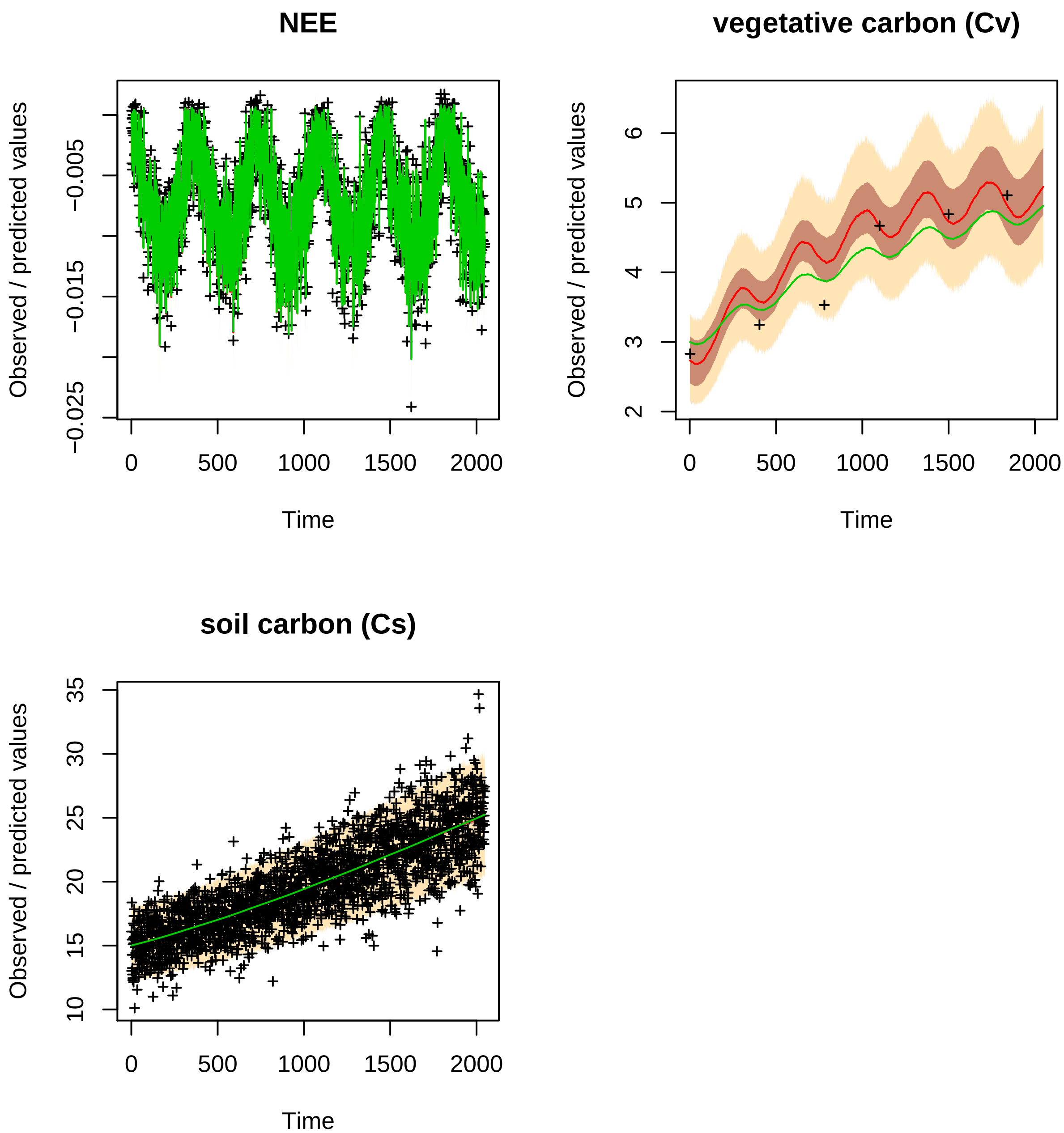

Figure 18: Model with error and unbalanced data with additive and multiplicative parameters to represent model error. Observations included in the calibration marked with a '+'. Red line 50% quantile posterior distribution. Green line is the 'true' model output. Dark brown shading 2.5% 97.5% quantile posterior distribution. Light brown shading 2.5% 97.5% predictive interval.

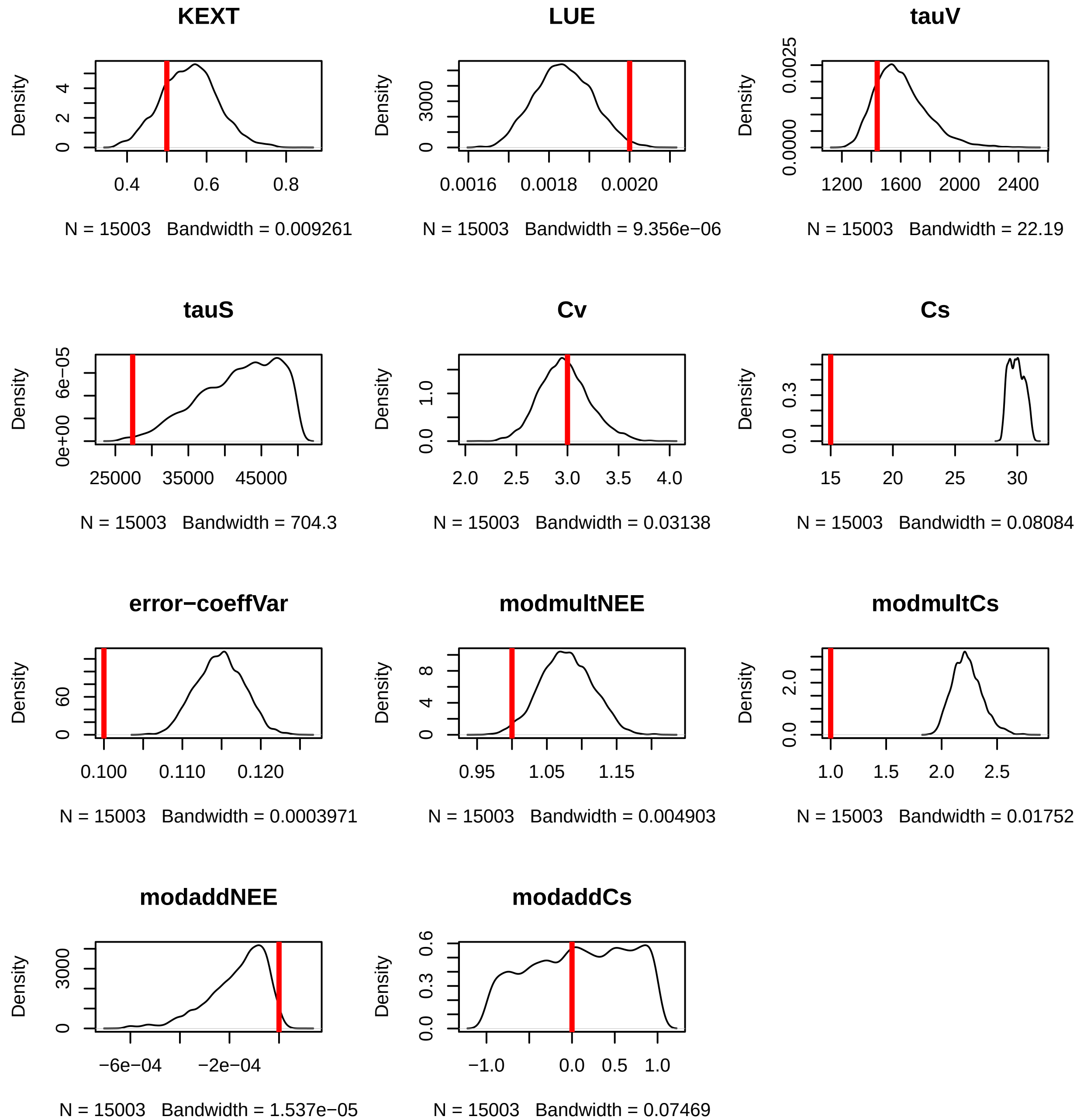

Figure 19: Perfect model and and unbalanced data with a multiplicative bias and additive and multiplicative parameters to represent the bias. Marginal posterior distribution of model parameters and intital states. The red line marks the ‘true’ parameter values.

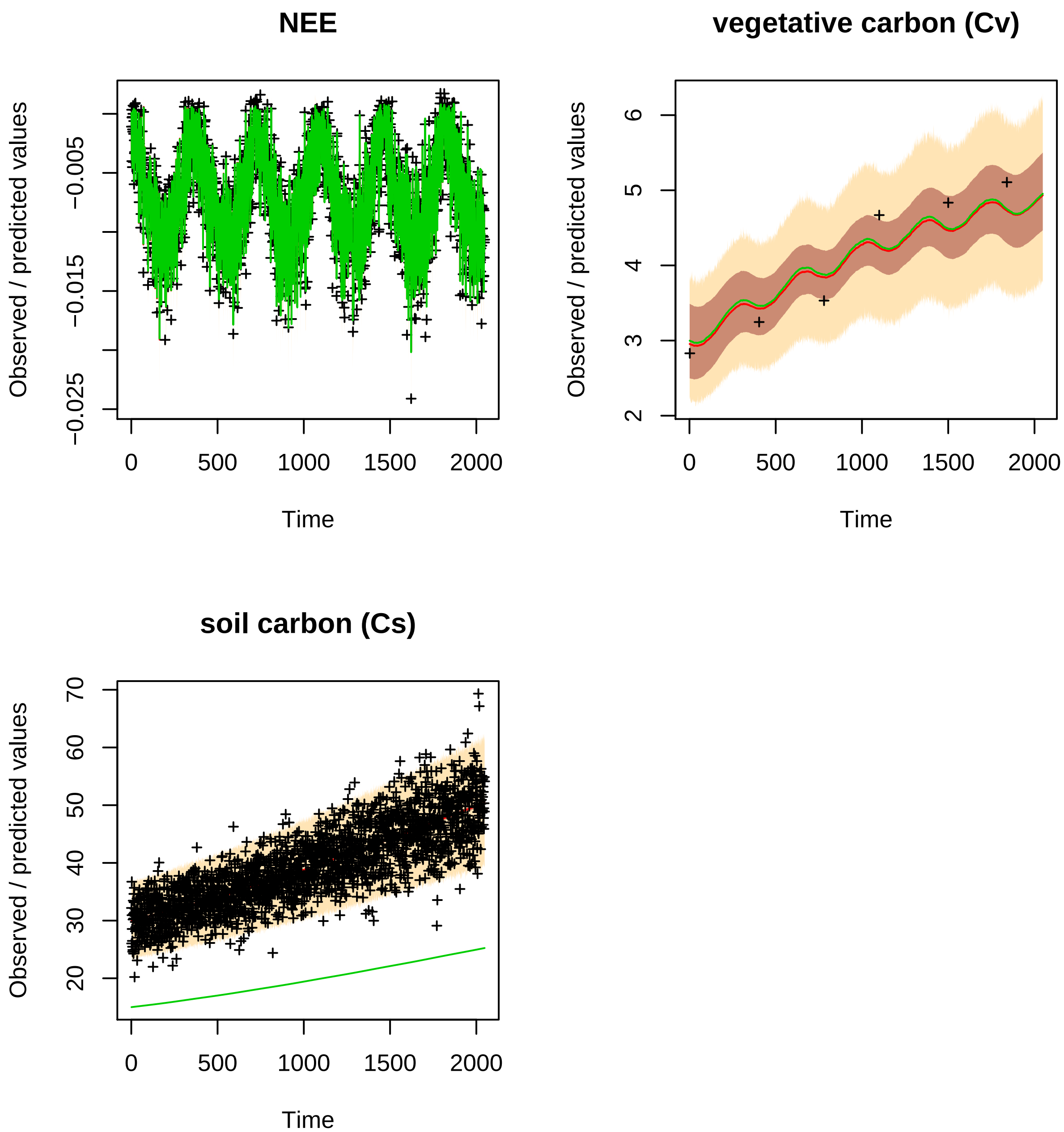

Figure 20: Perfect model and and unbalanced data with a multiplicative bias and additive and multiplicative parameters to represent the bias. Observations included in the calibration marked with a '+'. Red line 50% quantile posterior distribution. Green line is the 'true' model output. Dark brown shading 2.5% 97.5% quantile posterior distribution. Light brown shading 2.5% 97.5% predictive interval.

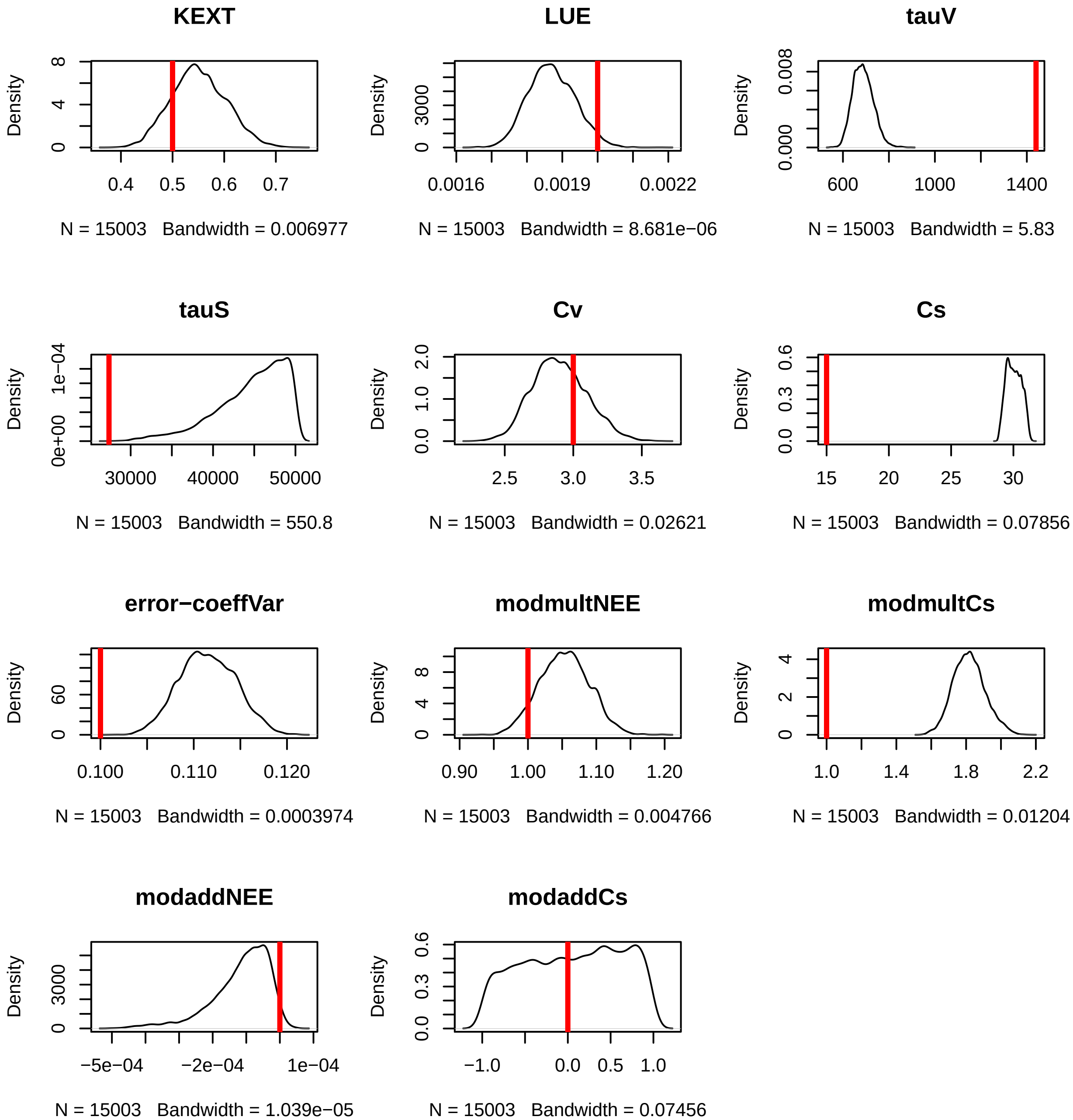

Figure 21: Model with error and and unbalanced data with a multiplicative bias and additive and multiplicative parameters to represent model error and the data bias. Marginal posterior distribution of model parameters and intital states. The red line marks the 'true' parameter values.

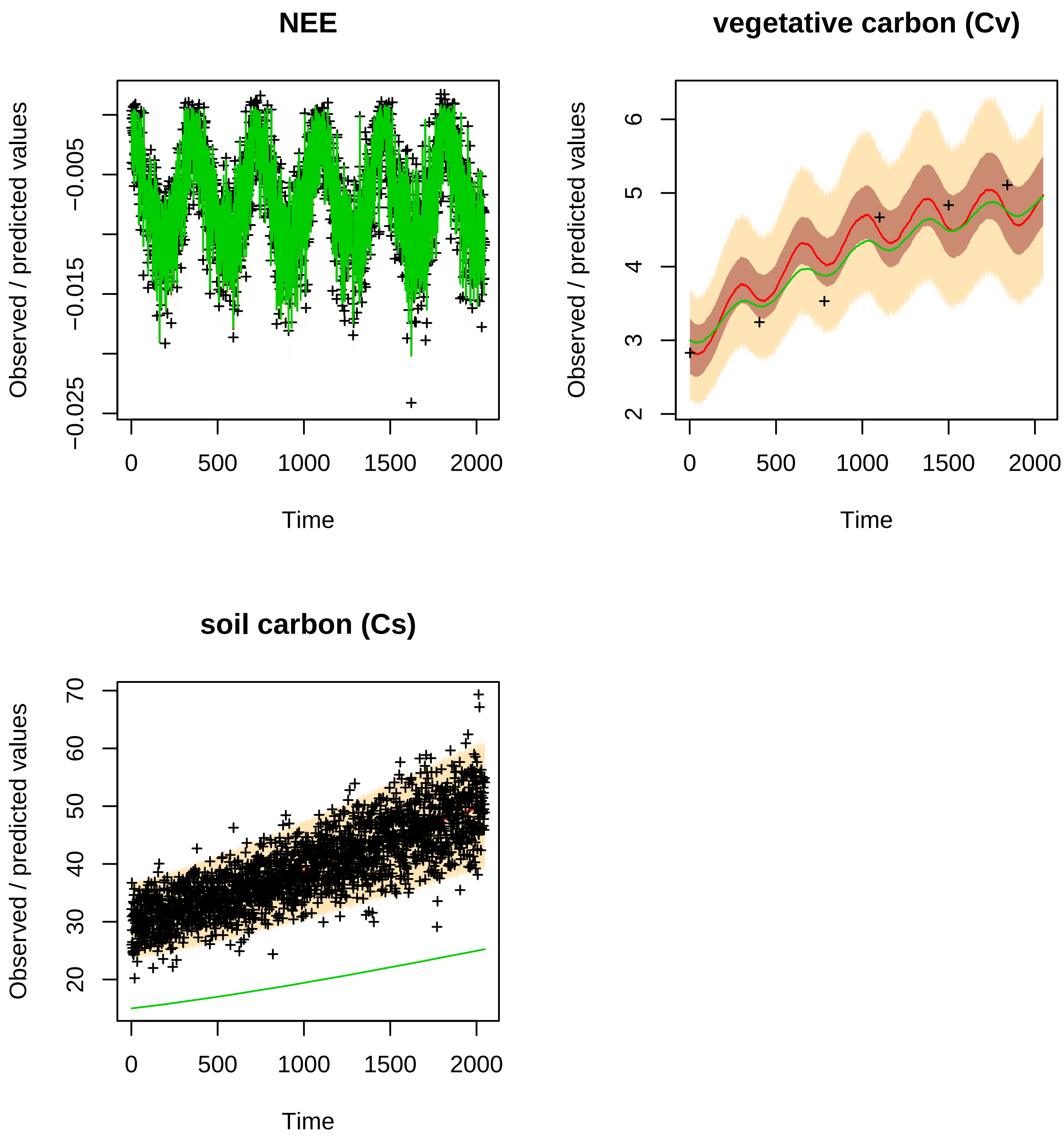

Figure 22: Model with error and and unbalanced data with a multiplicative bias and additive and multiplicative parameters to represent model error and the data bias. Observations included in the calibration marked with a '+'. Red line 50% quantile posterior distribution. Green line is the 'true' model output. Dark brown shading 2.5% 97.5% quantile posterior distribution. Light brown shading 2.5% 97.5% predictive interval.

6.2 Diagnostic tool introduced

- Have created a methodology and graph which can be used in many applications to diagnose and help judge the severity of the issue.

6.3 Representing model and data error in BC helps to alleviate the issue

- In this very simple example we were able to demonstrate a significant improvement by including terms in the likelihood to represent model and data error.
- In more real-world applications representing model and data error in BC will be much more challenging but the analysis demonstrated here shows how to deal with the issue of calibrating with unbalanced datasets without resorting to ad hoc methods.