Author-formatted manuscript prepared for J. Geo-phys. Res. Biogeosci.
doi 10.1002/2014jg002650
Date: 3 November 2014

# Balancing Multiple Constraints in Model-Data Integration: Weights and the Parameter Block Approach

**T. Wutzler**[1] **and N. Carvalhais**[1,2]

[1]Max Planck Institute for Biogeochemistry, Hans-Knöll-Straße 10, 07745 Jena, Germany
[2]Departamento de Ciências e Engenharia do Ambiente, DCEA, Faculdade de Ciências e Tecnologia, FCT , Universidade Nova de Lisboa, 2829-516 Caparica, Portugal

**Abstract.** Model data integration (MDI) studies are key to parameterize ecosystem models that synthesize our knowledge about ecosystem function. The use of diverse data sets, however, results in strongly imbalanced contributions of data streams with model fits favoring the largest data stream. This imbalance poses new challenges in the identification of model deficiencies. A standard approach for balancing is to attribute weights to different data streams in the cost function. However, this may result in overestimation of posterior uncertainty.

In this study, we propose an alternative: the parameter block approach. The proposed method enables joint optimization of different blocks, i.e., subsets of the parameters, against particular data streams. This method is applicable when specific parameter blocks are related to processes that are more strongly associated with specific observations, i.e., data streams. A comparison of different approaches using simple artificial examples and the DALEC ecosystem model is presented.

The unweighted inversion of a DALEC model variant, where artificial structural errors in photosynthesis calculation had been introduced, failed to reveal the resulting biases in fast processes (e.g., turnover). The posterior bias emerged only in parameters related to slower processes (e.g., carbon allocation) constrained by fewer data sets. On the other hand, when weighted or blocked approaches were used, the introduced biases were revealed, as expected, in parameters of fast processes.

Ultimately, with the parameter block approach, the transfer of model error was diminished and at the same time the overestimation of posterior uncertainty associated with weighting was prevented.

*Correspondence to:* T. Wutzler (twutz@bgc-jena.mpg.de)

## 1 Introduction

Evaluating the sensitivities of ecosystem responses to environmental conditions is essential to understand interactions between the terrestrial biosphere and the climate system. Terrestrial ecosystem models formalize hypotheses about the internal processes and functional responses of vegetation and soil to variations in environmental variables. The parametrization of these mechanisms is an essential step to evaluate the appropriateness of different model structures and the embedded hypotheses about ecosystem function.

Model data integration (MDI) methods, which are also referred to as model data fusion, model data assimilation, model data synthesis, or inverse modeling, present a framework to parameterize models based on the information content of observations (Richardson et al., 2010; van Oijen et al., 2005; Braakhekke et al., 2013). By minimizing the differences between modeled and observed quantities and considering the uncertainty resulting from (1) observations, (2) model parameters and drivers, and (3) model structure (Keenan et al., 2011), MDI approaches support data analysis, upscaling, prediction, and exploration of different hypotheses about ecosystem processes. MDI exercises are, therefore, very sensitive to the treatment of different sources of uncertainty, which is a prerequisite for comprehensive evaluation of the modeling structure.

The usage of multiple data streams in MDI is a promising approach to identify model structural error by inspecting the model-data mismatch (Fox et al., 2009; Williams et al., 2009; Carvalhais et al., 2010). For example, Braakhekke et al. (2014) showed that adding an extra data stream of radiocarbon to the MDI leads to a different conclusion on the relative importance of transport and soil organic matter (SOM) stabilization processes.

However, one of the challenges encountered in MDI using ecological data is the acknowledgment of imbalanced data streams, i.e., sets of measurements whose number of records

and/or the uncertainty of single records differ in orders of magnitude. In ecosystem level studies, rich data from automated high frequency measurements such as eddy covariance (EC) or from Earth observation systems are combined with sparse data from labor-intensive field work such as measurement of litterfall, soil carbon stocks, or radiocarbon data of ecosystem compartments (Richardson et al., 2010; Carvalhais et al., 2010; Braakhekke et al., 2014).

One problem associated with imbalanced data streams is that model structural uncertainties (or in short, model uncertainty) (Abramowitz et al., 2008) that are related to processes, which are strongly constrained by rich data streams, can be transferred to biases in posterior parameters of less constrained processes (best clarified with the simple example in section 3.4). Inclusion of additional sparse data streams that constrain the other processes will be of little use to counteract this transfer of structural model uncertainty unless the cost function is modified.

The underlying problem is that different sources of uncertainty are supposed to behave differently. Variance scales inversely with the number of observations (detailed in section 3.2). Uncertainty of the estimate is lower with more observations. This is a desired property for measurement uncertainty. However, there is also model uncertainty, i.e., the inability of the model structure to fully match the underlying processes. In multiple constraints approaches, often there exists a set of parameters where model predictions match one data stream and not the other, while with a different set of parameters the other data stream is matched instead. The pitfall is that this trade-off is preferentially allocated to match the rich data streams due to the scaling of variance with number of observations. Hence, errors may incorrectly emerge in parameters of less constrained processes. This may lead to a false identification of processes where the model structure needs to be improved.

Various methods have been proposed to account for model uncertainty. A simple way is to explicitly include an error term based on previous model fits (see examples in Trudinger et al., 2007; Kuppel et al., 2013). However, assigning record-specific or data stream specific weights implies that previously bad-fitting observations get low weight in the optimization. Consequently, model error in different processes can be transferred to processes constrained by the low-weight observations and, therefore, conceal other structural model limitations. Another method to deal with model uncertainty is to entirely neglect it in the cost function and assume measurement uncertainty only. This method is valid when observation errors are large and the model can flexibly adapt to the given data (e.g. interactive discussion of Braakhekke et al., 2013). An alternative method of accounting for model uncertainty, is to specify a prior distribution of combined data and model uncertainty of different data streams and integrate it to obtain a marginal distribution of model parameters (Appendix A of Kavetski et al., 2006). This method can be extended by including additional sources of uncertainties in the distribution of the model-data residuals (Schoups and Vrugt, 2010). Parameters related to combined data, input, and model uncertainty are estimated together with other model parameters as metaparameter or nuisance parameter from the fit. Such methods are suitable to quantify uncertainty of predictions but are less effective to identify inadequate model structure, as shown in section 3.4.

In summary, these methods account for model uncertainty and dealing with incomplete characterizations of model and data uncertainty. However, they do not prevent the transfer of model uncertainty to biases in parameters of less constrained processes.

Two published approaches, the Pareto-based optimizations and the Approximate Bayesian Computation (ABC), explicitly represent the inconsistencies between data streams for a given model. With Pareto-based optimization (Vincent and Grantham, 1981; Vrugt et al., 2003), all parameters are optimized against multiple constraints using several cost functions. The result, then, becomes a numeric vector instead of a scalar. The solution to the inversion problem is, in general, no longer a single "best" parameter set but consists of a Pareto set of solutions corresponding to various trade-offs between the objectives. Approximate Bayesian Computation (ABC) (Vrugt and Sadegh, 2013), uses one or multiple diagnostic application-specific summary metrics that have a better diagnostic power than a distribution based metric of the residual errors. These summary metrics are chosen without constraints of size or information content of the input data streams. Hence, ABC is robust against imbalanced data streams and is suitable to diagnose model structural inadequacy. However, it is less suitable for MDI studies, which aim at propagating the information in data streams to information on model parameters. In this paper, however, we focus on approaches that result in a single posterior distribution of the parameter set and that take formal account of the size and measurement uncertainties of the input data streams.

The objective of this paper is to explore several approaches to the problem of balancing model structural uncertainty when using imbalanced data streams. We aim at (1) making the reader aware of the problem of transfer of bias with imbalanced data streams; (2) proposing a new "parameter blocks" approach that optimizes subsets of parameters against different data streams; and (3) discussing weighting and alternative approaches by using examples of increasing complexity.

The paper is structured as follows: in section 2 the methods are explained. Sections 3.1 and 3.2 focus on the effects of weighting. Section 3.3 presents the new parameter block approach. Next a comparison of the three approaches of (1) an unweighted sum of squares, (2) a weighted sum of squares, and (3) the parameter block approach is presented using examples of increasing complexity. In section 3.4 a simple instructive example is used, while in sections 3.5 and 3.6 the consequences for the DALEC ecosystem model are studied.

Finally, the results are summarized and concluded in section 4.

## 2 Methods

### 2.1 Model Data Integration

MDI approaches rely on a combination of experimental data, a priori information, and theoretical understanding of the system. The a posteriori state of information is expressed by a joint probability density function (PDF) of the observational data ($\mathbf{o}$) and model parameters ($\boldsymbol{\theta}$) (Tarantola, 2005, eq.1.83). The posterior probability density of model parameters $\pi(\boldsymbol{\theta}|\boldsymbol{o})$ (in short posterior) is often of high interest. If a priori information and uncertainty of data are assumed to follow Gaussian distribution, the posterior information about model parameters can be expressed based on cost function $S(\boldsymbol{\theta})$ as,

$$\pi(\boldsymbol{\theta}) = K \exp[-\frac{1}{2}(S(\boldsymbol{\theta}) + S_{\mathrm{prior}}(\boldsymbol{\theta}))] \quad (1a)$$

$$S(\boldsymbol{\theta}) = (\mathbf{o} - g(\boldsymbol{\theta}))^T \mathbf{C}_D^{-1} (\mathbf{o} - g(\boldsymbol{\theta})) \quad (1b)$$

$$S_{\mathrm{prior}}(\boldsymbol{\theta}) = (\boldsymbol{\theta} - \boldsymbol{\theta}_{\mathrm{prior}})^T \mathbf{C}_M^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_{\mathrm{prior}}), \quad (1c)$$

where $K$ is a normalizing constant to ensure that the integral of $\pi$ sums to 1, $g(\boldsymbol{\theta})$ is the model prediction of observations ($\mathbf{o}$), $\mathbf{C}_D$ is the combined model and data uncertainty expressed as a covariance matrix. Prior knowledge of parameters is expressed by a multivariate Gaussian density with mean $\boldsymbol{\theta}_{\mathrm{prior}}$ and covariance matrix $\mathbf{C}_M$. The prior or data uncertainty, however, may deviate from Gaussian assumption. Although it is possible to derive formulations for other distributions as well, the multivariate normal is often the most conservative case for multidimensional parameter spaces without additional information on the specific shape or correlations among parameters or data streams (Gelman et al., 2003).

$S(\boldsymbol{\theta})$ is directly related to the negative log-likelihood. Hence, by minimizing $S(\boldsymbol{\theta})$, maximum likelihood estimates of $\boldsymbol{\theta}$ can be calculated. By explicitly solving eq. 1 or drawing a sample by MC simulation from this PDF, any statistics about the posterior can be estimated. Note, however, that the posterior estimate depends on the knowledge or estimate of the model and data uncertainties (matrix $\mathbf{C}_D$).

With multiple data streams, the covariance matrix $\mathbf{C}_D$ is composed of blocks for given data streams and zero correlation between observations of different streams. If all the observation errors are independent (also within block), the cost function $S(\boldsymbol{\theta})$ simplifies to a sum of the misfits $S_k$ for a data stream $k$.

$$S(\boldsymbol{\theta}) = \sum_k S_k(\boldsymbol{\theta}) \quad (2a)$$

$$S_k(\boldsymbol{\theta}) = \sum_i \frac{(o_i - g(\boldsymbol{\theta}))^2}{\sigma_i^2} \quad (2b)$$

$$\sigma_i^2 = \sigma_{o,i}^2 + \sigma_{m,i}^2 \approx \sigma_{o,i}^2, \quad (2c)$$

where $\sigma_i^2$ is the sum of variances of measurement errors, $\sigma_{o,i}^2$, and model structural errors, $\sigma_{m,i}^2$, for observation and prediction $i$, respectively. The uncertainty is often approximated as measurement uncertainty only. Note that even if independence of observations is violated by autocorrelation, the above formula (2) can be used with adjusted effective variance (Emery et al., 2007).

To estimate posterior parameters and their uncertainty (probability density function, PDF), equations 2 are used as a cost function with an adaptive Monte Carlo Markov Chain (MCMC) sampling for all unweighted scenarios of this study. Prediction uncertainty is estimated by running the model based on parameters sampled from posterior parameter distribution. Specifically, Differential Evolution Monte Carlo sampling is used by adopting past states (ter Braak and Vrugt, 2008) with four independent populations consisting of four chains each.

### 2.2 DALEC Model and Observation Data from Howland Forest

For the results presented in section 3.5, the simulations from a simple process-based ecosystem model of carbon dynamics, the Data Assimilation Linked Ecosystem Carbon (DALEC) (Williams et al., 2005), are used. Carbon, assimilated by plants through photosynthesis, cycles through different vegetation and soil pools based on simulated allocation, litterfall, and decomposition processes. These processes are driven by meteorological forcing and controlled by parameters describing the sensitivities and the compartmental carbon cycling within and between plants and soil.

The DALEC model has been used in the study of ecosystem dynamics (Williams et al., 2005), to evaluate different optimization algorithms (Fox et al., 2009), to investigate the role of multiple constraints approaches in parameter estimation (Richardson et al., 2010), and to study the separation of model from driver errors in predictive uncertainty (Spadavecchia et al., 2011).

DALEC considers a labile pool in vegetation, which is replenished by carbon from leaves before senescence and fuels the leaf production in the next spring. The labile pool is largely decoupled from forest biomass dynamics (Chuter, 2013).

Here we rely on the inversion of 15 parameters and initial conditions for the DALEC model (Table 1) to explore the impacts of imbalanced data streams on estimations of parameters and the resulting inferences. The other nine parameters

and initial conditions required by the DALEC model were fixed at the parameter values from previous own model inversions.

To obtain the the results in section 3.6, DALEC parameters were inverted by fitting to 10-years observations of eddy covariance-based net ecosystem exchange (NEE) (Hollinger et al., 2004; Hollinger and Richardson, 2005), soil respiration, and litterfall at the Howland forest. The data set, compiled for an intercomparison study (T. F. Keenan, personal communication 2013), was previously used in a study by Richardson et al. (2010). Parameter $\sigma_i$ in eq. 2c was set as measurement uncertainty.

The observation uncertainties increased with the magnitude of the measurements (T. F. Keenan, personal communication 2013). They are expressed as,

$$\begin{aligned}\sigma_{\mathrm{NEE}} &= \begin{cases} 0.12\ \mathrm{NEE}^2 + 0.19\ \mathrm{NEE} + 0.16, & \text{if NEE} > 0. \\ -0.19\ \mathrm{NEE} + 0.76, & \text{otherwise.} \end{cases} \\ \sigma_{\mathrm{Litterfall}} &= 0.2\,\mathrm{Litterfall} \\ \sigma_{\mathrm{SResp}} &= 0.02 + 0.35\,\mathrm{SResp} \end{aligned} \tag{3}$$

## 3 Results and Discussion

### 3.1 Why Not to Weight - Inflating Posterior Uncertainty

In order to increase the relative influence of sparse data streams in the cost function, the summation terms in cost function (eq. 2a) can be replaced by weighted sums (4a), with the weights $w_k$ inversely proportional to the number of observations in data stream $k$.

$$S(\boldsymbol{\theta}) = \sum_k w_k S_k(\boldsymbol{\theta}) \tag{4a}$$

$$w_k = 1/n_k \tag{4b}$$

In effect, this changes the prescribed uncertainty of the model-observation residuals $\sigma_i^2$ as,

$$S_{w,k} = w_k S_k = \sum_i w_k \frac{(o_i - g(\boldsymbol{\theta}))^2}{\sigma_i^2} = \sum_i \frac{(o_i - g(\boldsymbol{\theta}))^2}{\sigma_{w,i}^2}, \tag{5a}$$

$$\text{with } \sigma_{w,i}^2 = \frac{\sigma_i^2}{w_k} \tag{5b}$$

Hence, weighting by the inverse of the number of observations ($w_k = 1/n_k$) has the same effect as prescribing an increase of measurement variance by a factor of $n_k$.

The effect is demonstrated with a basic example of a single downweighted data stream. Artificial observations ($n_k = 9$) were sampled from a normal distribution around 0 with a variance $\sigma_i^2 = 9$ in three scenarios. The single parameter of the model, $\boldsymbol{\theta} = (m)$, corresponded to the single model prediction, namely the mean of observations $g(m) = m$.

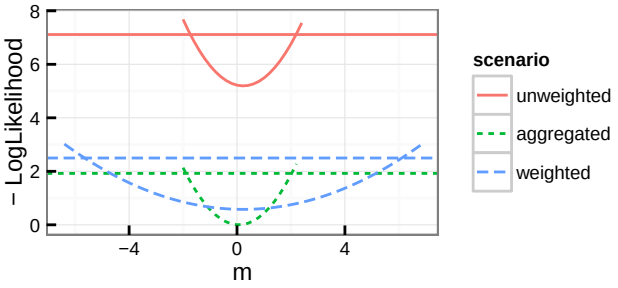


Fig. 1: Log-likelihood profiles, $1/2S(m)$, (Hilborn and Mangel, 1997) of the mean of observations. Horizontal lines depict the level of the minimum $+1.92$. Above this level parameter values differ significantly ($\alpha = 5\%$) from the best estimate. Scenario descriptions are given in text section 3.1. Note the flattening of the surface and the increased parameter confidence interval with weighting of the data streams.

The shape of the negative log-likelihood (i.e., half the cost function) as a function of the observation mean, $1/2S(m)$, was examined with the following scenarios.

*Unweighted*: All the observations were part of an unweighted cost function (eq. 2a). The shape of the corresponding log-likelihood function is displayed as the solid red line in fig. 1.

*Aggregated*: Only the mean of the observations was part of the cost function. The corresponding standard error was calculated by usual error propagation: $\sigma_m = \sigma_i/\sqrt{n} = \sqrt{9}/\sqrt{9}$. The likelihood profile (short dashed green line in fig. 1) was lower than that in the unweighted scenario but had the same curvature. As expected, the range of confidence interval, i.e., where the likelihood curve is below the horizontal line, was the same as in the unweighted scenario ( $\pm 2\sigma_m$).

*Weighted*: All the observations were part of a cost function, but the data streams were weighted by the number of observations $w_1 = 1/n$ (eq. 4a). The corresponding likelihood profile (long dashed blue line in fig. 1) was much flatter than the other two cases. Hence, the confidence interval of the estimated parameter was much larger. This is caused by the inflation of variance of observations by a factor of $n = 9$.

The weighting essentially ignores the additional certainty that can be acquired by using more observations. This results in an overly wide posterior distribution of parameters and a large uncertainty envelope of predictions. While this is obvious in the simple example shown here, the same effect can be expected to hold true in a more complex setting involving multiple imbalanced data streams. In addition, the choice of the weights also introduces some subjectivity because the number of records depends on preaggregation of measure-

Table 1: Description and prior probability distribution of the optimized DALEC model parameters and initial conditions. The decay rates have been inferred at log-transformed scale. For each parameter, a normal distribution was prescribed so that the 95% confidence range matched the minimum and maximum limits. For values outside these limits, the prior was set to zero, i.e., values were rejected, resulting in a truncated Gaussian prior.

| | Unit | Minimum | Maximum | Trans | Description |
|---|---|---|---|---|---|
| Cfmax | $gC/m^2$ | 100 | 500 | | Maximum foliage carbon pool |
| Clab0 | $gC/m^2$ | 50 | 500 | | Initial value of labile carbon |
| Et | $1/^\circ C$ | 0.01 | 0.2 | | Parameter in exponential temperature rate modifier |
| Fg | 1 | 0.4 | 0.9 | | Fraction of GPP used for growth respiration |
| Fll | 1 | 0.1 | 0.5 | | Fraction of C in leaf loss transferred to litter |
| Flr | 1 | 0.01 | 0.5 | | Fraction of labile transfers respired |
| Fnf | 1 | 0.5 | 0.9 | | Fraction of NPP allocated to foliage |
| Fnrr | 1 | 0.1 | 0.5 | | Fraction of NPP2 allocated to roots |
| Lfall | $^\circ C$ | 10 | 20 | | Minimum daily temperature causing leaf fall |
| Lout | $^\circ C$ | 150 | 250 | | Value of Growing degree days causing leaf out |
| Pr | 1 | 5 | 15 | | Nitrogen use efficiency parameter (a1) in the canopy model |
| pRaBelow | 1 | 0.1 | 0.5 | | Autotrophic respiration from below ground |
| Tf | 1/day | 1E-4 | 0.2 | log | Turnover rate of foliage |
| Tl | 1/day | 1E-6 | 0.01 | log | Turnover rate of litter |
| Tlab | 1/day | 1E-4 | 0.1 | log | Turnover rate of labile carbon |

ments and sampling frequency (e.g., the Unweighted versus Aggregated scenarios above). From a perspective of MDI, assuming a perfect model, weighting yields a conservative posterior parameter uncertainty without detracting from the ability of converging to the same optimum. In case of an imperfect model with conflicting data streams, however, the optimum solution is sensitive to the weighting as shown in the following section.

### 3.2 Why to Weight - Balancing Model Uncertainty

The main objective for weighting is to balance the allocation of model uncertainty across data streams. The model-data mismatch caused by structural uncertainty of model is additional to measurement uncertainty and can show up in different data streams depending on parameter values.

Consider that all observations are independent and refer to the same time. Then we can use the mean and standard error of the mean instead of the original observations in the cost function. Hence, the variance decreases by the factor of the number of records as

$$\overline{o_i} = \frac{1}{n}\sum_i o_i \tag{6a}$$

$$\sigma^2(\overline{o_i}) = \frac{1}{n^2}\sum_i \sigma_i^2 = \frac{1}{n}\overline{\sigma_i^2} \tag{6b}$$

The uncertainty is smaller with larger number of observation records. Unfortunately, in an unweighted inversion, also the model structural uncertainty is smaller for a data stream with more records. Therefore, the model parameters will be derived in a way that minimizes the mismatch with the richest data, resulting in an allocation of the structural model-data mismatch to sparse data streams. Only a small part of model uncertainty will be allocated to parameters of processes that are constrained by rich data streams, but a large part will be allocated to parameters of processes that are constrained by sparse data streams only. The actual model structural errors, however, are obviously not distributed according to richness of data streams. Eventually, with unweighted scenarios, the model error or bias in well-constrained processes gets transferred to bias in model parameters of less constrained processes. This effect is demonstrated in section 3.4.

To balance model uncertainty across data streams, it is necessary to counteract the scaling of variance with the number of records (eq. 6b). This can be achieved by multiplying the variance by the number of records of the respective data stream, i.e., specifying weights $w_k = 1/n_k$ in equation 4a (see section 3.1). This, ultimately, corroborates the practice of weighting of different data streams.

Thus, in calibration studies with multiple imbalanced data streams, there is an inevitable trade-off between good estimates of posterior uncertainty (without weighting) versus balancing the structural model error (with weighting). Hence, we suggest not to use the weighting approach if model uncertainty is small compared to measurement uncertainty, and to use the weights $1/n_k$ if model uncertainty is of concern.

Are there better methods to deal with this conflicting situation? In the following section, we propose an approach to address this question.

### 3.3 The parameter block Approach

One approach to prevent the transfer of model uncertainty to parameters of less constrained processes is to optimize different subsets of parameters against different data streams. For example, optimizing parameters of photosynthesis and respiration against rich NEE and respiration data streams, but optimizing parameters of slow allocation processes against sparse data streams of litterfall measurements, and changes in wood and soil carbon stocks. The parameter vector is divided into two or more blocks, i.e., subsets, and each block is inverted against a different cost function (eq. 7) as

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_{\mathrm{rich}}, \boldsymbol{\theta}_{\mathrm{sparse}}) \tag{7a}$$

$$\pi(\boldsymbol{\theta}_{\mathrm{rich}}|\boldsymbol{\theta}_{\mathrm{sparse}}) = f(S_{\mathrm{rich}}(\boldsymbol{\theta}), S_{\mathrm{sparse}}(\boldsymbol{\theta})) \tag{7b}$$

$$\pi(\boldsymbol{\theta}_{\mathrm{sparse}}|\boldsymbol{\theta}_{\mathrm{rich}}) = f(S_{\mathrm{sparse}}(\boldsymbol{\theta})), \tag{7c}$$

where equation (7a) depicts the division of the parameter vector into two blocks. Note that the parameters in block $\boldsymbol{\theta}_{\mathrm{sparse}}$ in equation (7c) are not constrained by the rich data stream in order to prevent the transfer of bias. For the parameters constrained by the rich data (7b), the sparse data can be included. With independent Gaussian distribution of observations, function $f$ in equation (7) is defined (according to eqs. 1a and 2a) as follows:

$$f(S_1, \ldots S_l) = K \exp\left[-\frac{1}{2}\left(\sum_{k=1\ldots l} S_k(\boldsymbol{\theta}) + S_{\mathrm{prior}}(\boldsymbol{\theta})\right)\right] \tag{8}$$

It is possible to carry out the inversions of the blocks separately (e.g. Peylin et al., 2013). First, a full inversion of one parameter block against the rich data streams is performed, while keeping other parameters fixed. Next follows an inversion of the other parameter block against the other data streams. After the second inversion, however, the posterior of the first parameter block might have changed due to the change in values of parameters in the second block. The procedure, therefore, needs to be iterative. Despite the iteration, uncertainty due to correlations between parameters in different blocks might be neglected.

In this study, we use a Block-at-a-Time Metropolis Hastings Algorithm with Monte Carlo sampling. Chib and Greenberg (1995) show the feasibility of drawing in succession from each of the conditional densities and associated cost functions. Before doing the Metropolis step, the log-likelihood of the current state in parameter space should be recalculated if other parameter values have changed. Therefore, in each step of the Monte Carlo walk, a block can be updated based on one cost function, followed by an update of another block based on a different cost function (Fig. 2).

The correlations between the parameters in different blocks are also captured because all blocks are updated in every step, and a change in one block immediately influences the successive updates of other blocks.

**repeat**
  **for** each parameter block $i$ **do**
    **if** $\boldsymbol{\theta}_{di}$ changed since last update of block $i$ **then**
      recalculate current posterior density $\pi_i(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{di})$
    **end if**
    propose new $\hat{\boldsymbol{\theta}}_i$ and calculate $\hat{\pi}_i(\hat{\boldsymbol{\theta}}_i|\boldsymbol{\theta}_{di})$
    **if** $\hat{\pi}_i/\pi_i > r_{unif}$ **then**
      set $\boldsymbol{\theta}_i = \hat{\boldsymbol{\theta}}_i$ and $\pi_i = \hat{\pi}_i$
    **end if**
  **end for**
  record current state $\boldsymbol{\theta} = \bigcup \boldsymbol{\theta}_i$
**until** converged to limiting distribution

Fig. 2: Essential loop of the Block-at-a-Time Metropolis Algorithm. $\boldsymbol{\theta}_{di}$ comprises all parameters from other parameter blocks that the calculation of the posterior density $\pi_i(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{di})$ depends on. Parameters in $\boldsymbol{\theta}_{di}$ may have been updated in other blocks. $r_{\mathrm{unif}}$ is a random number from a uniform distribution between 0 and 1, corresponding to the standard MCMC update rule (Metropolis et al., 1953; Chib and Greenberg, 1995).

By specifying a different cost function for a parameter block that does not depend on the rich data streams, the transfer of model uncertainty to these parameters can be prevented. However, the information of the rich streams for the given parameter block is not used, and, hence, the posterior uncertainty may be larger compared to the case when information of all data stream is used for all parameters. The parameter block approach is therefore applicable when certain processes and the associated parameters are related to observations from particular data streams.

### 3.4 Comparison: Artificial Model

First, the effects of different specifications of cost functions on posterior parameters and predictions are compared by using a simple instructive example. We modeled the observations of two data streams $\mathbf{y}_{\mathrm{rich}}$ ($n_{\mathrm{rich}} = 1000$) and $\mathbf{y}_{\mathrm{sparse}}$ ($n_{\mathrm{sparse}} = 10$), with observed covariates $\boldsymbol{x}_{\mathrm{rich}}$ and $\boldsymbol{x}_{\mathrm{sparse}}$, respectively, (eq. 9) using the parameters $\boldsymbol{\theta} = (a,b,c)$ as

$$\hat{y}_{i,\mathrm{rich}}(a,b,c) = a x_{1,\mathrm{sparse}} + b(x_{i,\mathrm{rich}} - \mathrm{c}) \tag{9a}$$

$$\hat{y}_{i,\mathrm{sparse}}(a,b) = a x_{i,\mathrm{sparse}} + b\overline{x}_{\mathrm{rich}}/10 \tag{9b}$$

Note that the prediction in the sparse data stream, $\mathbf{y}_{\mathrm{sparse}}$, mainly depends on covariates, $\boldsymbol{x}_{\mathrm{sparse}}$, of sparse data stream and only uses a single aggregated value, $\overline{x}_{\mathrm{rich}}$, of the covariates of the rich data stream. The rich data stream, for instance, may represent a short-term measurement series of soil respiration, and the sparse data stream may represent annual measurements of export of dissolved organic carbon from soil or any other less frequently measured inventory. The covariates $\boldsymbol{x}_{\mathrm{rich}}$ can be daily temperature time series, while the aggregated value $\overline{x}_{\mathrm{rich}}$ can be mean annual temperature.

Table 2: Scenarios of model inversions.

| Scenario | Description |
|---|---|
| unweighted | Unweighted sum of rich and sparse streams |
| estUnc | Estimate residual uncertainty from fit |
| weighted | Sum of rich and sparse streams weighted by $1/n_k$ |
| blocked | Sampling parameter blocks using Metropolis |

The covariates $\boldsymbol{x}_{\mathrm{sparse}}$ may track slow changes in soil carbon stocks, with $x_{1,\mathrm{sparse}}$ being the specific value for the year when the respiration time series is measured.

In order to demonstrate the transfer of model uncertainty, a bias variable $c$ is introduced. This variable may represent a bias in measurement $\boldsymbol{x}_{\mathrm{rich}}$ or a conceptual difference between measurement and its meaning in the model.

True dependent variables $\boldsymbol{y}^*$ were generated using model (9) with parameters $a=1$ and $b=2$, and bias variable $c=0.3$. The covariates $\boldsymbol{x}_{\mathrm{rich}}$ and $\boldsymbol{x}_{\mathrm{sparse}}$ were sampled from uniform distributions as $\sim\mathcal{U}(0.7,1)$ and $\sim\mathcal{U}(0.5,1.5)$, respectively. Next, artificial measurements, $\boldsymbol{y}$, were generated by adding Gaussian noise to $\boldsymbol{y}^*$ with standard deviation of 6% and 2% of the mean of $\boldsymbol{y}^*_{\mathrm{rich}}$ and $\boldsymbol{y}^*_{\mathrm{sparse}}$, respectively.

Using this artificial data, the posterior of parameters $a$ and $b$ were sampled by different scenarios of model inversions (Table 2), all using the same model that slightly differs from the data-generating model by using the fixed bias variable $c=0$. A flat uniform unbounded prior, specifically $S_{\mathrm{prior}}(a,b)=0$, was used. As the model error was introduced in eq. 9a, a bias in predictions of $\hat{y}_{i,\mathrm{rich}}$ was expected.

The estUnc scenario tried to deal with the problem of transfer of bias by estimating residual uncertainties, $\sigma_{Rich}$ and $\sigma_{Sparse}$ that include the model structural error (eq. 2c) from the model-data fits. Specifically, these two parameters were treated in a separate parameter block and were sampled each step from a conjugate Scaled-inverse-$\chi^2$ distribution with a zero prior ($\nu_0=0$) (Gelman et al., 2003, p. 51).

With the blocked scenario, parameter $a$ was updated based on a cost function of only the sparse data stream (eq. 10c). Similarly, parameter $b$ was optimized based on a cost function of only the rich data stream (eq. 10d).

$$S_{\mathrm{sparse}}(a,b)=\sum_i \frac{\left(y_{i,\mathrm{sparse}}-\hat{y}_{i,\mathrm{sparse}}(a,b,0)\right)^2}{\sigma^2_{i,\mathrm{sparse}}} \quad (10a)$$

$$S_{\mathrm{rich}}(a,b)=\sum_i \frac{\left(y_{i,\mathrm{rich}}-\hat{y}_{i,\mathrm{rich}}(a,b)\right)^2}{\sigma^2_{i,\mathrm{rich}}} \quad (10b)$$

$$\pi_{\mathrm{sparse}}(a|b)=f\left(S_{\mathrm{sparse}}(a,b)\right) \quad (10c)$$

$$\pi_{\mathrm{rich}}(b|a)=f\left(S_{\mathrm{rich}}(a,b)\right), \quad (10d)$$

with $f$ defined in eq. 8 and the hats denoting model predictions (eq. 9).

By sampling the predictive posterior, i.e., forward simulations using samples of the parameter posterior, the median model response and the 95% model prediction interval were inferred.

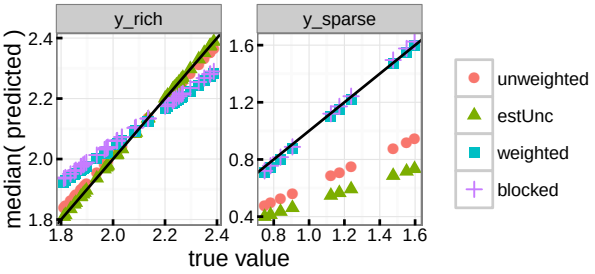


Fig. 3: Scatterplot of median of predictions against true values for the artificial case. The black line denotes the 1:1 correspondence. Symbols represent different inversion scenarios (Table 2). Note that without weighting or parameter blocks the prescribed model structural error in calculation of $\mathbf{y}_{\mathrm{rich}}$ got transferred to $\mathbf{y}_{\mathrm{sparse}}$.

Fig. 3 displays the scatterplots of median predictive posterior against true values. Ideally, the mismatch between model and data should manifest in the predictions in rich data stream $\mathbf{y}_{\mathrm{rich}}$ due to introduction of model error in eq. 9a. In the estUnc scenario, the introduced model error was compensated by a worse fit in the sparse data stream, i.e., the model error was transferred to the sparse data stream. Fitting against both rich and sparse data streams using unweighted observation uncertainties (unweighted scenario in Fig. 3) still shows the error in sparse data stream. This suggests that the second data stream did not prevent the transfer of model error from rich data stream. Hence, results falsely identified that model errors are related with the processes constrained by sparse data stream $\mathbf{y}_{\mathrm{sparse}}$ (eq. 9). Only after weighting the data streams or using parameter blocks, the model error was correctly allocated to rich data stream $\mathbf{y}_{\mathrm{rich}}$. With the weighted approach, however, the parameter uncertainties were larger (Fig. B-1).

### 3.5 Comparison: DALEC Model With Artificial Data

Inversions of the DALEC model parameters (Section 2.2) were tested using artificial data. Two rich data streams (NEE and soil respiration) with 2000 observations, and a sparse data stream (litterfall) with 10 observations were generated by adding Gaussian noise to model predictions using prescribed parameter values.

In the parameter blocks scenario, three parameters related to allocation ($\mathrm{Fnf},\mathrm{Fll},\mathrm{Cfmax}$) out of 15 optimized parameters (Table 1), were optimized using a cost function that only considered the misfit to sparse litterfall data stream but did not consider the rich NEE and soil respiration data streams (eq. 11c).

$$\boldsymbol{\theta}_{\mathrm{sparse}} = (\mathrm{Fnf}, \mathrm{Fll}, \mathrm{Cfmax}) \tag{11a}$$

$$\pi(\boldsymbol{\theta}_{\mathrm{rich}}|\boldsymbol{\theta}_{\mathrm{sparse}}) = f(S_{\mathrm{NEE}}, S_{\mathrm{SResp}} S_{\mathrm{Litterfall}}) \tag{11b}$$

$$\pi(\boldsymbol{\theta}_{\mathrm{sparse}}|\boldsymbol{\theta}_{\mathrm{rich}}) = f(S_{\mathrm{Litterfall}}), \tag{11c}$$

with $f$ defined by eq. 8.

In addition to the model-data misfit and prior knowledge on model parameters (Table 1), prior knowledge about the development of the labile pool (Appendix A) was included.

First it was checked, whether the different inversion settings (Table 2) could recover the prescribed true parameter set. When the original data-generating DALEC model was used, the posterior distributions were correctly centered around the true prescribed parameters.

Next, an artificial model error was introduced to the calculation of gross primary production (GPP) in DALEC. Specifically, the leaf mass per area (LMA) value, that is used to translate leaf area index to leaf biomass, was changed from 100 to 20 ($\mathrm{gCm^{-2}}$).

During the inversion of the biased model, posterior density of the parameters differed between different inversion scenarios (Fig. 4, Table B-1, and Fig. B-2). With the unweighted cost function, parameters related to temperature, such as Et and Lfall, were constrained very well, whereas parameters related to allocation, such as Fll, had a large bias. However, with the explicit introduction of error in GPP calculation, the errors were expected to be more related to fast NEE-related process. In the weighted approach, confidence ranges were large and mostly encompass the true parameter value. The parameter block approach achieved a balance between the two scenarios. Temperature related parameters (Et and Lfall) were constrained near their true values, while ranges of most other parameters encompassed the true parameter values. In all of the inversion scenarios, there was a bias in parameter Flr, the fraction of labile transfers respired to make up for the introduced error in GPP calculation.

The consequences of different parameter estimates can be better seen in the predictive posterior, i.e., model predictions using estimated parameter values (Fig. 5). With the unweighted cost function, the use of rich data streams of NEE and soil respiration results in a good fit despite the introduction of error in the model. The litterfall data, however, did not fit well even though it was a part of inversion. Bias has been transferred to the sparse data stream. The weighted and parameter block approaches both achieved a better balance of the misfit between these imbalanced data streams.

### 3.6 Comparison: Real World Ecosystem Case

The DALEC model has been inverted using different scenarios (Table 2) against real observation data. The DALEC model and the 10 year observation data at the Howland forest are described in (section 2.2). The setup of the cost function and the inversion was identical to the experiments described in section 3.5 with the nonbiased model.

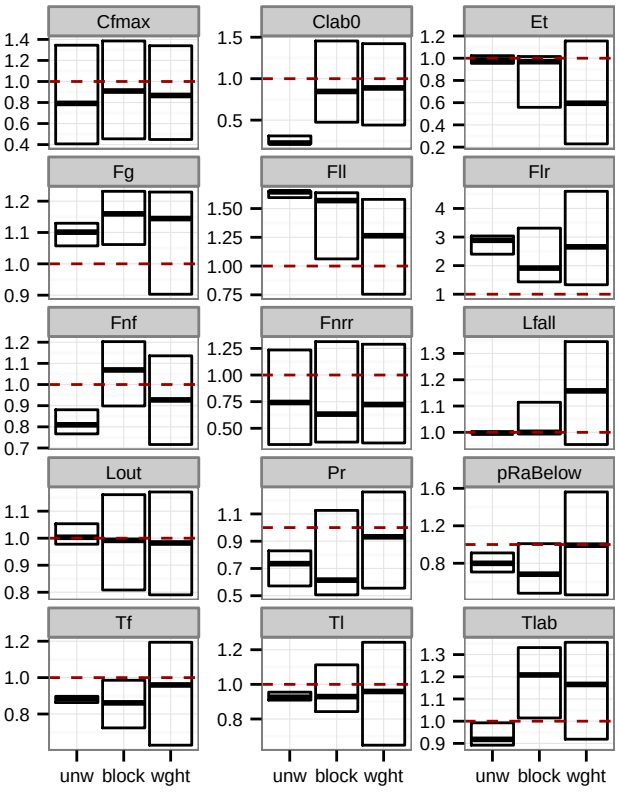


Fig. 4: Posterior density statistics of inverting the biased DALEC model against artificial data. Parameters have been normalized by their true values: $p/p_{\mathrm{true}}$. Bars represent the 95% confidence range and solid lines the best estimate. The dashed line helps to see if the range encompasses the true value at 1. Columns represent different scenarios of model inversion (unweighted, blocked, and weighted as described in Table 2).

There was a conflict between the NEE and litterfall data given the DALEC model structure (Fig. 6). When the model uncertainties were not balanced (unweighted scenario), a high confidence in predictions of NEE and soil respiration was achieved only at a cost of a clear bias in predictions of leaf litterfall. Seven out of 10 litterfall observations were outside the confidence interval of model predictions. Weighting the data streams by the number of records substantially increased the uncertainty in model predictions, whereas the marginal parameter posterior distributions were not constrained much compared to the prior distributions (Table B-2 and Fig. B-3). The parameter block approach achieved a balance between the two cases. The bias in litterfall predictions decreased, and the uncertainty in predictions of NEE and soil respiration was moderate.

To study the consequences of differences in posterior parameters, the model was run for a 40 year period into the future using a repeated time series of forcing data (Figure 7). In the unweighted scenario, the predictions of the range of NEE,

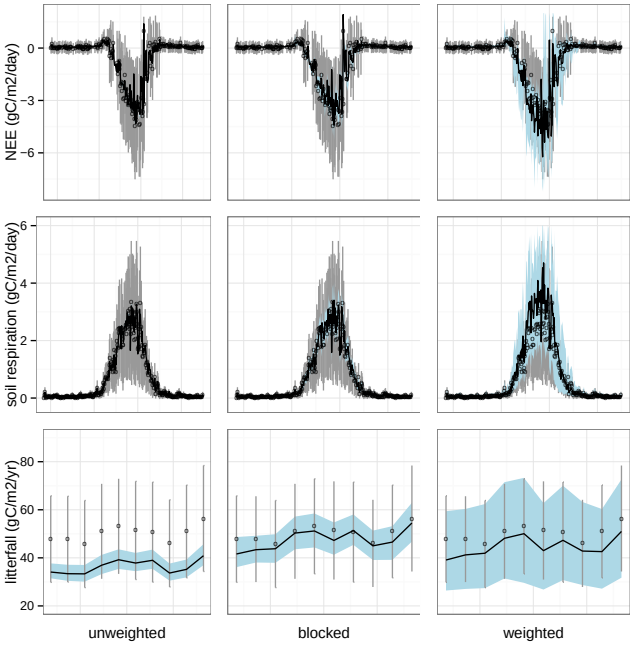


Fig. 5: Predictive posterior of inverting the biased DALEC model with artificial data over time. NEE and soil respiration are displayed against days of the tenth year and litterfall against the 10 years. Open dots and grey bars represent observations and their 95% confidence interval. The black lines and the blue bands represent the median and 95% confidence interval of population model predictions. Columns represent different scenarios of model inversion (Table 2).

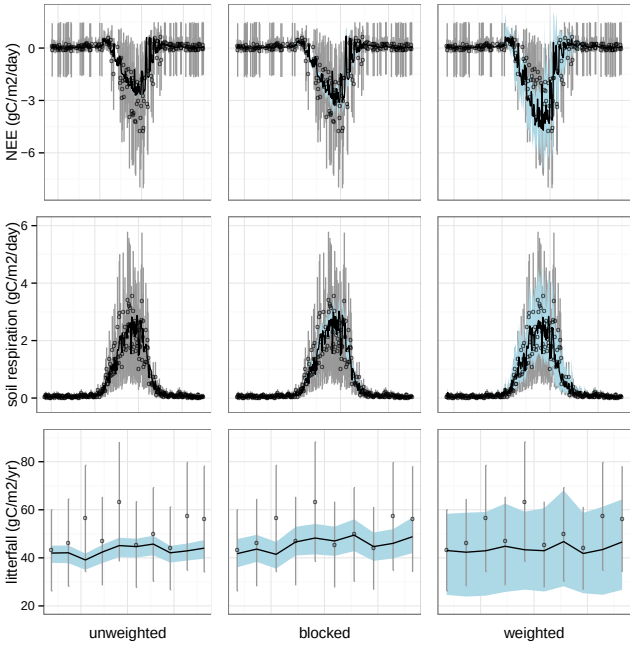


Fig. 6: Predictive posterior of inverting the DALEC model with data from Howland forest using alternative inversion scenarios (Table 2), similar to Figure 5.

respiration, and leaf litterfall decreased with time. In the weighted scenario, the uncertainty bounds of the model predictions were quite large, allowing for a wide range of possible ecosystem responses. Finally, in the parameter block scenario, an intermediate response was predicted. The range of NEE was at the lower limits of the predictions in the weighted scenario.

This demonstrates how large imbalance of data streams forces the model to fit the richest NEE data and allocates mismatches to other processes. The high resolution of the time series might convey information about site and event-specific details of the processes that were neither intended to be modeled exactly, nor likely to occur in the same way across different sites and times. Hence, there is a good chance that the model parameters are overfitted to the NEE data (Ginzburg and Jensen, 2004).

Another important finding of this example was, that significant differences in multidimensional parameter space are hard to detect by visualizing the parameter distribution alone. The differences were much clearer when the predictions of different parameter sets in a realm outside the calibration period were evaluated.

It is unclear why the DALEC model was not able to match

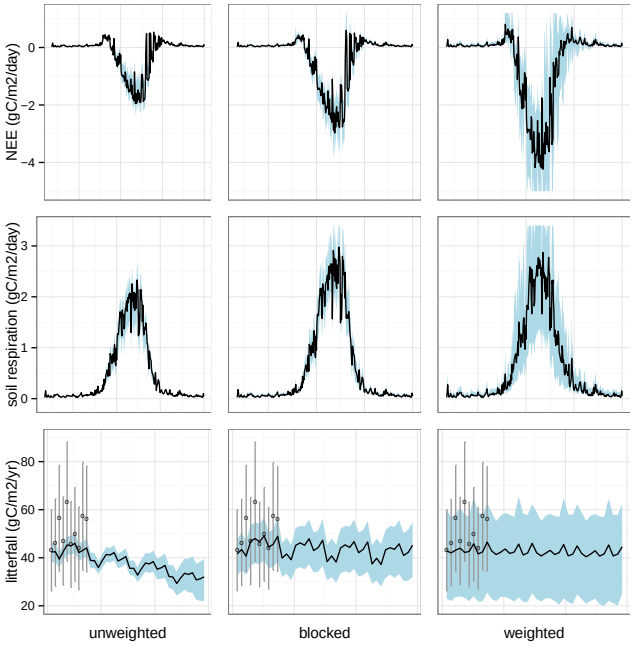


Fig. 7: Predictive posterior of inverting the DALEC model with data at the Howland forest. Similar to figure 6, but showing future prediction for year 40 instead of year 10 and litterfall for 40 years. Note that many of the uncertainties with the weighted scenarios are larger than the range of $y$ axis.

all the observations. Although a detailed analysis is beyond the scope of this study, there are a few hints. Model parameters were inverted such that there were a few years where the carbon investments into leaves were not fully compensated by carbon fixation. Consequently, this led to a modeled depletion of the labile pool and the inability to simulate appropriate new leaf growth in spring. Such mismatch could emerge from a systematic overestimation of measured litterfall.

The parameter block approach is applicable in situations when certain processes, and their associated parameters, can be related to corresponding observations. When the relationship between parameters and observations is not clear, then choice of the parameter blocks versus the cost functions may, however, introduce some subjectivity. Thus, the resulting posterior parameter distribution may not be directly comparable to other studies. However, the strength of the parameter block approach is the ability to explore the source of structural uncertainties of models while still propagating uncertainty from observations to parameters and at the same time counteracting the transfer of model uncertainty to parameters of processes constrained by sparse data streams.

## 4 Conclusions

The effects of different treatments of multiple imbalanced data streams have been studied with several examples of increasing complexity. The findings are common across all examples. Effects of model structural uncertainty are allocated to different processes depending on the size of the data streams. In the unweighted approach, or when uncertainties are estimated from fits, parameters are optimized such that the rich data streams fit well and additional sparse data streams have only a very small influence. Bias due to model structural error is transferred to the parameters of processes constrained by observation with sparse data only.

In the biased version of the DALEC model, unweighted inversion against artificial observations did not result in the expected bias in parameters of strongly constrained fast processes, such as NEE, to which the artificial bias was introduced. Instead, bias was transferred to parameters of less constrained slow processes, such as litterfall.

The transfer of model structural uncertainty has large consequences in model predictions and identification of model process that requires improvements. In future, with the availability of high-frequency ecosystem measurements and a necessity of abstractions in model development, the issue can be expected to be discussed more often.

Therefore, it is important to heed the following conclusions.

1. Model inversion against imbalanced data streams leads to allocation or transfer of model structural uncertainty from parameters of more constrained processes to parameters of less constrained processes.
2. Weighting the data streams by number of observations allocates model uncertainty equally across processes constrained by different data streams. When only model uncertainty is of concern, weighting by the number of observation records is reasonable even though it may strongly inflate the posterior uncertainties.
3. With both weighting and parameter block approaches, the best estimate of parameters is expected to be closer to true values than that with the unweighted approach. With weighting though, the posterior depends more strongly on the choice of informative priors. The parameter block approach can constrain the posterior density better than the weighting approach.

We present a new 'parameter block' approach as an alternative way to address the issues with the use of imbalanced data streams in model-data integration. Specifically, optimization of different parameter subsets against different data streams helps to counteract the problem of transfer of model uncertainty.

*Acknowledgements.* We thank Trevor F. Keenan and David Hollinger for permission to use the intercomparison Howland data and model setup for demonstrating the uncertainty issues. We thank Andrew Richardson, Carlos Sierra, Maarten Braakhekke, Sujan Koirala, and two reviewers for valuable comments on the manuscript. This work was supported by the QUASOM ERC starting grant (ERC-2007-StG-208516).

## References


Abramowitz, G., Leuning, R., Clark, M., and Pitman, A.: Evaluating the performance of land surface models, Journal of climate, 21, 5468–5481, doi:abs/10.1175/2008JCLI2378.1, 2008.

Braakhekke, M. C., Wutzler, T., Reichstein, M., Beer, C., Hoosbeek, M. R., Kruijt, B., Schrumpf, M., Schoening, I., , and Kabat, P.: Modelling the vertical soil organic matter profile using 210Pbex measurements and Bayesian inversion, Biogeosciences, 10, 399–420, doi:10.5194/bg-10-399-2013, 2013.

Braakhekke, M. C., Beer, C., Schrumpf, M., Ekici, A., Ahrens, B., Hoosbeek, M. R., Kruijt, B., Kabat, P., and Reichstein, M.: The use of radiocarbon to constrain current and future soil organic matter turnover and transport in a temperate forest, Journal of Geophysical Research: Biogeosciences, doi: 10.1002/2013JG002420, 2014.

Carvalhais, N., Reichstein, M., Ciais, P., Collatz, G. J., Mahecha, M. D., Montagnani, L., Papale, D., Rambal, S., and Seixas, J.: Identification of vegetation and soil carbon pools out of equilibrium in a process model via eddy covariance and biometric constraints, Global Change Biology, 16, 2813–2829, doi: 10.1111/j.1365-2486.2010.02173.x, 2010.

Chib, S. and Greenberg, E.: Understanding the Metropolis-Hastings algorithm, American Statistician, 49, 327335, 1995.

Chuter, A. M.: A Qualitative Analysis of the Data Assimilation Linked Ecosystem Carbon Model, DALEC, Ph.D. thesis, University of Surrey, Faculty of Engineering and Physical Sciences, 2013.

Emery, A., Valenti, E., and Bardot, D.: Using Bayesian inference for inverse modeling, Measurement Science and Technology, 18, 19–29, 2007.

Fox, A., Williams, M., Richardson, A., Cameron, D., Gove, J., Quaife, T., Ricciuto, D., Reichstein, M., Tomelleri, E., Trudinger, C., et al.: The REFLEX project: Comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data, Agricultural and Forest Meteorology, 149, 1597–1615, 2009.

Gelman, A., Carlin, J., Stern, H., and Rubin, D.: Bayesian Data Analysis. 2003, Boca Raton (FL): Chapman Hall, 2003.

Ginzburg, L. R. and Jensen, C. X.: Rules of thumb for judging ecological theories, Trends in Ecology & Evolution, 19, 121–126, doi:10.1016/j.tree.2003.11.004, 2004.

Hilborn, R. and Mangel, M.: The ecological detective. confronting models with data, Princeton Univesity Press, Princeton, NJ., 1997.

Hollinger, D. Y. and Richardson, A. D.: Uncertainty in eddy covariance measurements and its application to physiological models, Tree Physiology, 25, 873–885, doi:10.1093/treephys/25.7.873, 2005.

Hollinger, D. Y., Aber, J., Dail, B., Davidson, E. A., Goltz, S. M., Hughes, H., Leclerc, M. Y., Lee, J. T., Richardson, A. D., Rodrigues, C., and et al.: Spatial and temporal variability in forest-atmosphere CO2 exchange, Global Change Biology, 10, 1689–1706, doi:10.1111/j.1365-2486.2004.00847.x, 2004.

Kavetski, D., Kuczera, G., and Franks, S. W.: Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, Water Resources Research, 42, doi:10.1029/2005WR004368, 2006.

Keenan, T. F., Carbone, M. S., Reichstein, M., and Richardson, A. D.: The model–data fusion pitfall: assuming certainty in an uncertain world, Oecologia, 167, 587–597, doi:10.1007/s00442-011-2106-x, 2011.

Kuppel, S., Chevallier, F., and Peylin, P.: Quantifying the model structural error in carbon cycle data assimilation systems, Geoscientific Model Development, 6, 45–55, doi:10.5194/gmd-6-45-2013, 2013.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E.: Equation of state calculations by fast computing machines, The journal of chemical physics, 21, 1087–1092, 1953.

Peylin, P., Macbean, N., Bacour, C., Maignan, F., Chevallier, F., Ciais, P., Thum, T., et al.: Optimization of the process-based global model, ORCHIDEE, using multiple data streams (in-situ FluxNet NEE, LE and biomass, satellite NDVI, and atmospheric CO2 data), in: EGU General Assembly Conference Abstracts, vol. 15, p. 11951, http://adsabs.harvard.edu/abs/2013EGUGA..1511951P, 2013.

Richardson, A., Williams, M., Hollinger, D., Moore, D., Dail, D., Davidson, E., Scott, N., Evans, R., Hughes, H., Lee, J., Rodrigues, C., and Savage, K.: Estimating parameters of a forest ecosystem C model with measurements of stocks and fluxes as joint constraints, Oecologia, pp. –, doi:10.1007/s00442-010-1628-y, 2010.

Schoups, G. and Vrugt, J. A.: A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors, Water Resources Research, 46, doi:10.1029/2009WR008933, 2010.

Spadavecchia, L., Williams, M., and Law, B.: Uncertainty in predictions of forest carbon dynamics: Separating driver error from model error, Ecological Applications, 21, 1506–1522, doi: 10.1890/09-1183.1, 2011.

Tarantola, A.: Inverse Problems Theory, Methods for Data Fitting and Model Parameter Estimation, SIAM, Philadelphia, 2005.

ter Braak, C. and Vrugt, J.: Differential Evolution Markov Chain with snooker updater and fewer chains, Statistics and Computing, 18, 435–446, 2008.

Trudinger, C., Raupach, M., Rayner, P., Kattge, J., Liu, Q., Pak, B., Reichstein, M., Renzullo, L., Richardson, A., Roxburgh, S., et al.: OptIC project: An intercomparison of optimization techniques for inverse modeling in terrestrial biogeochemical models, J. Geophys. Res, 112, G02 027, 2007.

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

Vincent, T. L. and Grantham, W. J.: Optimality in parametric systems, Wiley New York, 1981.

Vrugt, J. A. and Sadegh, M.: Toward diagnostic model calibration and evaluation: Approximate Bayesian computation, Water Resources Research, pp. n/a–n/a, doi:10.1002/wrcr.20354, 2013.

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W., and Sorooshian, S.: Effective and efficient algorithm for multiobjective optimization of hydrologic models, Water Resources Research, 39, 1214, doi:10.1029/2002WR001746, 2003.

Williams, M., Schwarz, P. A., Law, B. E., Irvine, J., and Kurpius, M. R.: An improved analysis of forest carbon dynamics using data assimilation, Global Change Biology, 11, 89–105, 2005.

Williams, M., Richardson, A. D., Reichstein, M., Stoy, P. C., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D. Y., Kattge, J., Leuning, R., Luo, Y., Tomelleri, E., Trudinger, C. M., and Wang, Y. P.: Improving land surface models with

FLUXNET data, Biogeosciences, 6, 1341–1359, doi:10.5194/bg-6-1341-2009, 2009.

## Appendix A Prior Knowledge on Changes of the Labile Pool

The predictions based on inversions with or without prior of the labile pool matched the observations for the calibration period almost equally well. However, in future predictions without prior of the labile pool, the systems' ability to produce enough new leaves in spring diminished after about 30 years, which lead to a strong net carbon source despite identical forcing.

An examination of the dynamics revealed that a discrepancy between NEE and litterfall data streams was resolved by a consistent depletion of the labile pool over the years. Even though the effects were negligible during the observation period, they became larger after a sufficient depletion of labile pool after about 20 years. This was a clear case of a good fit for wrong reasons.

Hence, the knowledge (that labile pool should not consistently change over the observation period) was included by penalty in the cost function. In the following, the details of this penalty component in cost function are explained.

First, the consistent change of labile pool over the 10 year observation period was calculated by a regression of pool size at yearday 365 against year. Next, the slope was normalized by the mean and maximum allowed change of pool size.

$$s_r = abs\left(\frac{s}{c_{\mathrm{mean}}\, c_{\mathrm{limit}}}\right), \tag{A1}$$

where $s$ is the regression slope, $c_{\mathrm{mean}}$ is the mean of labile pool across the observation period, and $c_{\mathrm{limit}}$ is a prescribed maximum relative change (a value of 0.8% was used). At these prescribed maximum change limits ($s_r = 1$), the cost penalty, $S_{\delta\mathrm{CLab}}$, was set to twice the expected misfit, i.e., number of records in the richest data stream, $p_{\delta\mathrm{CLab}}$ (eq. A2). Within the limits ($s_r < 1$), a function that is steep at the edges but very flat inside was used. Outside the limits ($s_r > 1$), a slowly increasing function was used to help the inversion to find the near zero slope range (Fig. A1).

$$S_{\delta\mathrm{CLab}} = p_{\delta\mathrm{CLab}} \begin{cases} s_r^{10} & \text{for } s_r < 1 \\ 1 + \log(s_r) & \text{else} \end{cases} \tag{A2}$$

In a previous attempt, we tried to reject all parameter vectors that yielded larger slopes However, this was not found to be effective because it rejected large regions in the parameter space hampering the burnin of the inversion.

The cost $S_{\delta\mathrm{CLab}}$ was added to the cost of model-data misfit (eq. 8). This penalty helped the inversion to converge to an optimum, in which the labile pool was not consistently depleted. For the best parameters, the cost of this penalty was an order of magnitude smaller that the cost of model-data misfit and hence did not distort the posterior parameter probabilities.

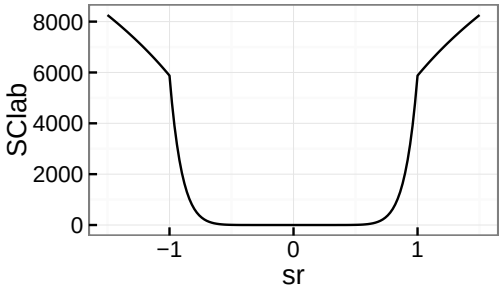


Fig. A1: Cost for consistent relative change of labile pool $s_r$ during observation period. The cost at $abs(s_r) = 1$, i.e., at the limits of allowed change, corresponds to twice the expected misfit with the best parameter estimate.

## Appendix B Supplement Figures and Tables

More detailed and complex figures and tables are presented here. They are a bit harder to read than the figures in the text, but help those readers who are interested in comparing the presented results to other studies. Figure B-1 presents the posterior parameters in addition to posterior predictions of Figure 3. Figure B-2 and table B-1 detail the non-normalized parameter distributions in addition to Figure 4 for the DALEC example with artificial data. Figure B-3 and table B-2 do so for the DALEC example with observation data from Howland forest.

Fig. B-1: Violin plots (probability density versus parameter value) of posterior densities of the artificial example using the model with error in calculation of the observations in the rich data stream. The black dashed vertical lines denote the true values. In addition to scenarios in table 2, a fit to the rich data stream only and a Gibbs sampling (Gelman et al., 2003) (blocked G scenario) are included. The latter sampled parameters in each step directly from PDFs given the data instead of using a MCMC update. This was possible because of the simple model structure. It yields almost the same results as the parameter-blocks approach (blocked scenario) and, hence, further validates the parameter-blocks approach.

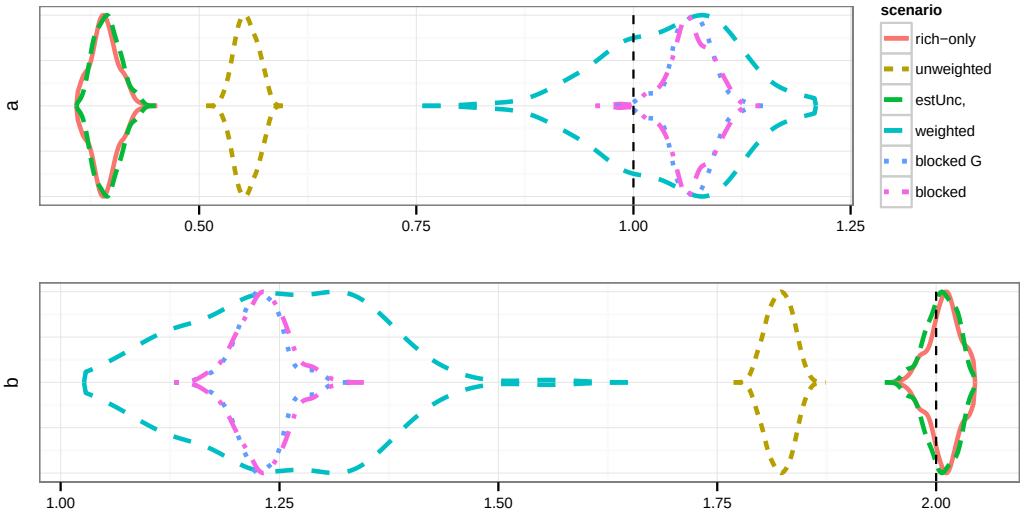

Fig. B-2: Posterior densities of inverting the biased DALEC model against artificial data. Black vertical lines represent the true prescribed parameter values. The colored vertical lines represent the maximum likelihood estimate for the respective inversion scenario. Note that although the model bias was introduced into the GPP calculation, the parameters of these fast processes, such as $E_t$ are not biased but the bias shows up in slow process parameters such as $Clab0$ or $Fll$ with the unweighted scenario.

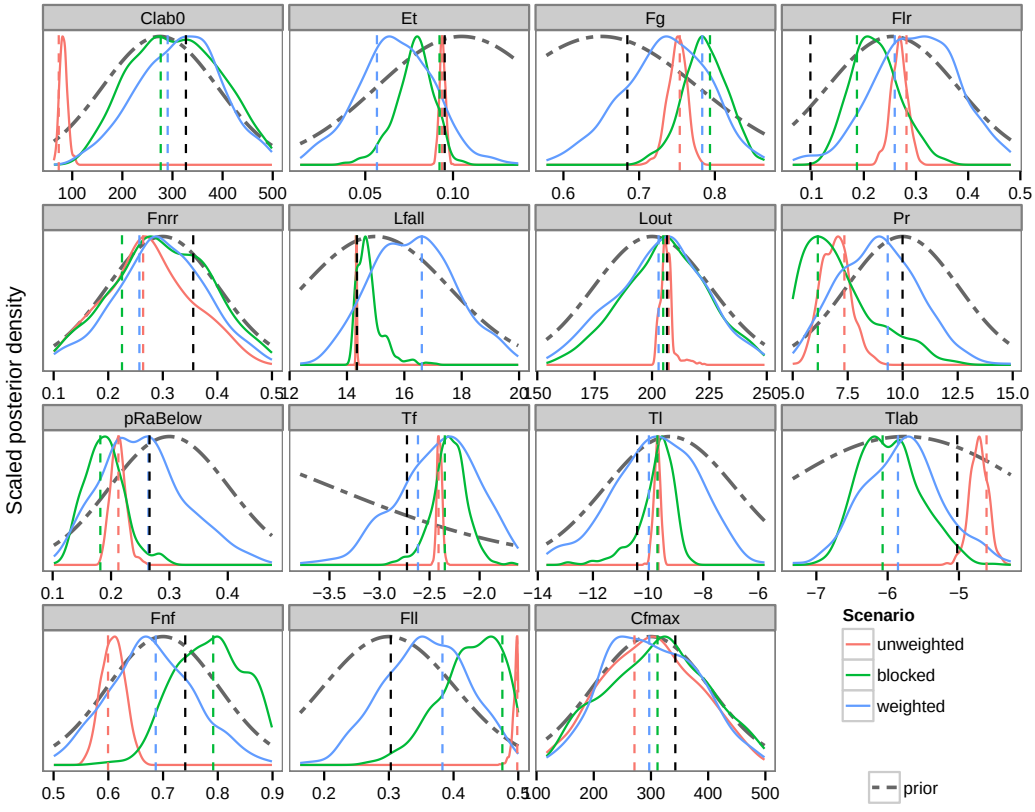


Table B-1: Posterior parameter statistics of the biased DALEC model inverted against artificial observations. Turnover rates (Tf, Tl,Tlab) are log-transformed. MLE represent the parameter vector with maximum likelihood. In the case of the parameter-block approach, this is the vector with the highest minimum rank across likelihoods of the different blocks.

| | | unweighted | | | | blocked | | | | weighted | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | true | MLE | median | 2.5% | 97.5% | MLE | median | 2.5% | 97.5% | MLE | median | 2.5% | 97.5% |
| Fg | 0.68 | 0.75 | 0.75 | 0.72 | 0.77 | 0.79 | 0.79 | 0.73 | 0.84 | 0.78 | 0.74 | 0.62 | 0.84 |
| Fnf | 0.74 | 0.60 | 0.61 | 0.57 | 0.65 | 0.79 | 0.79 | 0.67 | 0.89 | 0.69 | 0.68 | 0.53 | 0.84 |
| Fnrr | 0.36 | 0.26 | 0.27 | 0.12 | 0.44 | 0.23 | 0.30 | 0.13 | 0.47 | 0.26 | 0.30 | 0.13 | 0.46 |
| Tf | -2.73 | -2.41 | -2.40 | -2.45 | -2.35 | -2.35 | -2.30 | -2.69 | -1.97 | -2.62 | -2.41 | -3.26 | -1.71 |
| Tl | -10.40 | -9.62 | -9.67 | -9.93 | -9.45 | -9.66 | -9.59 | -11.57 | -8.76 | -9.97 | -9.67 | -12.92 | -6.74 |
| Et | 0.10 | 0.09 | 0.09 | 0.09 | 0.10 | 0.09 | 0.08 | 0.05 | 0.10 | 0.06 | 0.07 | 0.02 | 0.11 |
| Pr | 9.99 | 7.35 | 6.89 | 5.71 | 8.29 | 6.14 | 6.76 | 5.06 | 11.25 | 9.32 | 8.92 | 5.55 | 12.60 |
| Lout | 206.53 | 207.24 | 206.09 | 201.94 | 217.57 | 204.87 | 206.05 | 166.94 | 239.79 | 202.86 | 203.53 | 163.25 | 241.86 |
| Lfall | 14.35 | 14.31 | 14.33 | 14.30 | 14.37 | 14.35 | 14.68 | 14.26 | 15.99 | 16.61 | 16.33 | 13.68 | 19.30 |
| Fll | 0.30 | 0.50 | 0.50 | 0.48 | 0.50 | 0.48 | 0.43 | 0.32 | 0.50 | 0.38 | 0.36 | 0.23 | 0.48 |
| Tlab | -5.02 | -4.61 | -4.73 | -4.99 | -4.48 | -6.07 | -6.01 | -6.69 | -5.10 | -5.86 | -5.79 | -6.81 | -4.62 |
| Flr | 0.10 | 0.28 | 0.27 | 0.23 | 0.30 | 0.19 | 0.22 | 0.14 | 0.32 | 0.26 | 0.30 | 0.13 | 0.45 |
| Cfmax | 342.59 | 271.18 | 297.65 | 139.26 | 460.76 | 311.48 | 308.29 | 155.71 | 474.51 | 297.00 | 295.35 | 153.53 | 459.05 |
| pRaBelow | 0.27 | 0.21 | 0.21 | 0.19 | 0.24 | 0.18 | 0.19 | 0.13 | 0.27 | 0.26 | 0.25 | 0.12 | 0.42 |
| Clab0 | 326.84 | 73.17 | 83.03 | 69.74 | 101.11 | 276.53 | 309.64 | 155.09 | 475.41 | 290.42 | 312.97 | 144.09 | 464.84 |

Fig. B-3: Posterior densities of inverting the DALEC model using Howland observational data.

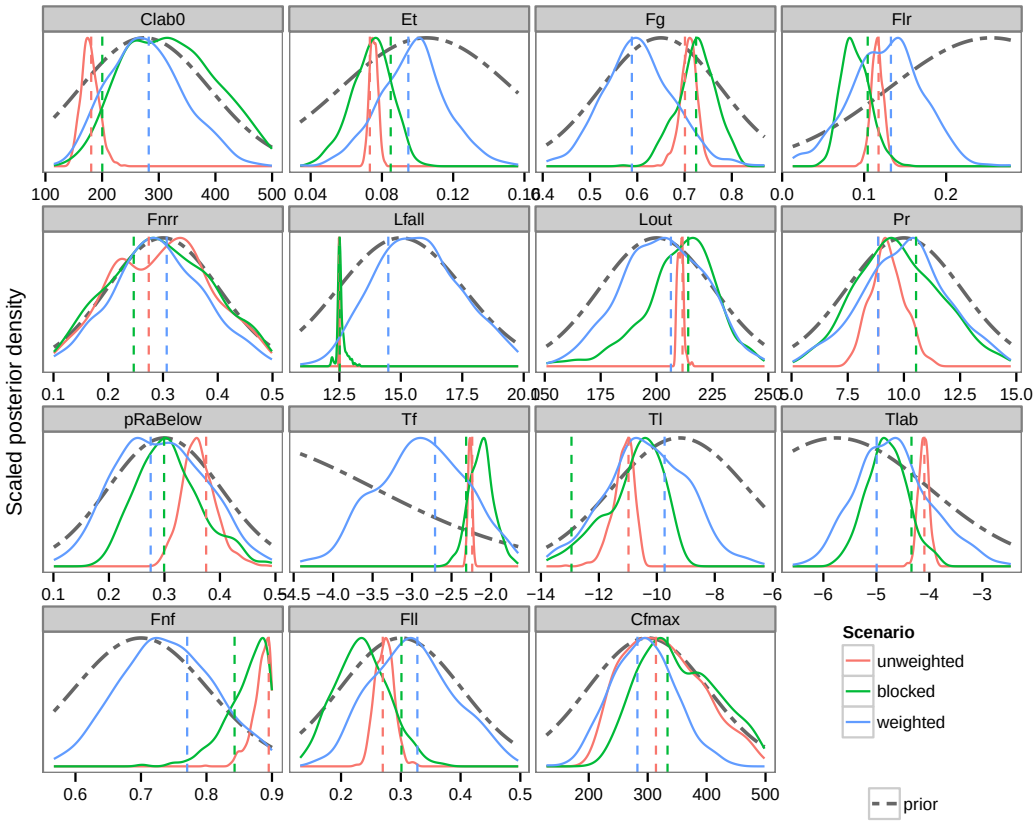


Table B-2: Posterior parameter statistics of the DALEC model inverted against observations at Howland forest. Same as table B-1, unless the true value is not known.

| | unweighted | | | | blocked | | | | weighted | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MLE | median | 2.5% | 97.5% | MLE | median | 2.5% | 97.5% | MLE | median | 2.5% | 97.5% |
| Fg | 0.70 | 0.71 | 0.67 | 0.74 | 0.72 | 0.73 | 0.64 | 0.80 | 0.59 | 0.60 | 0.46 | 0.76 |
| Fnf | 0.90 | 0.89 | 0.84 | 0.90 | 0.84 | 0.87 | 0.78 | 0.90 | 0.77 | 0.74 | 0.60 | 0.88 |
| Fnrr | 0.27 | 0.30 | 0.13 | 0.47 | 0.25 | 0.29 | 0.12 | 0.47 | 0.31 | 0.29 | 0.14 | 0.46 |
| Tf | -2.24 | -2.27 | -2.32 | -2.22 | -2.32 | -2.12 | -2.42 | -1.79 | -2.71 | -2.87 | -4.03 | -1.80 |
| Tl | -10.98 | -11.13 | -12.06 | -10.59 | -12.95 | -10.67 | -13.37 | -9.41 | -9.73 | -10.34 | -13.22 | -7.43 |
| Et | 0.07 | 0.08 | 0.07 | 0.08 | 0.08 | 0.07 | 0.05 | 0.09 | 0.09 | 0.10 | 0.06 | 0.13 |
| Pr | 8.86 | 9.25 | 7.88 | 10.88 | 10.54 | 9.57 | 6.14 | 13.51 | 8.85 | 10.13 | 6.24 | 13.92 |
| Lout | 211.62 | 210.59 | 208.70 | 213.90 | 214.16 | 212.21 | 176.79 | 240.30 | 206.42 | 203.39 | 166.33 | 239.61 |
| Lfall | 12.49 | 12.49 | 12.48 | 12.50 | 12.51 | 12.53 | 12.26 | 13.06 | 14.49 | 15.67 | 12.52 | 19.26 |
| Fll | 0.27 | 0.27 | 0.24 | 0.30 | 0.30 | 0.23 | 0.15 | 0.32 | 0.33 | 0.32 | 0.20 | 0.45 |
| Tlab | -4.09 | -4.11 | -4.30 | -3.91 | -4.34 | -4.82 | -5.55 | -4.02 | -4.99 | -4.68 | -5.92 | -3.14 |
| Flr | 0.12 | 0.12 | 0.10 | 0.13 | 0.10 | 0.09 | 0.06 | 0.13 | 0.13 | 0.13 | 0.04 | 0.22 |
| Cfmax | 314.04 | 322.98 | 222.25 | 469.65 | 333.69 | 348.85 | 242.19 | 482.79 | 282.58 | 293.49 | 191.10 | 408.44 |
| pRaBelow | 0.38 | 0.36 | 0.30 | 0.43 | 0.30 | 0.30 | 0.21 | 0.44 | 0.28 | 0.29 | 0.14 | 0.45 |
| Clab0 | 180.75 | 178.05 | 151.00 | 211.23 | 200.31 | 321.19 | 187.13 | 476.17 | 282.32 | 275.25 | 159.91 | 409.99 |