Biogeosciences Discuss., 6, C3741–C3744, 2009
www.biogeosciences-discuss.net/6/C3741/2009/

Biogeosciences
Discussions

Interactive comment on "Reducing impacts of systematic errors in the observation data on inversing ecosystem model parameters using different normalization methods" *by* L. Zhang et al.

Anonymous Referee #1

Received and published: 30 December 2009

Zang et al. compare different methods to reduce the impact of systematic errors of observational data on the inversion of ecosystem model parameters. They apply these methods to the inversion of photosynthetic capacity in a process based ecosystem model against artificial observations of Leaf Area Index (LAI). The artificial LAI data were created by adding different types of systematic error (fixed, proportional, fixed+proportional and binomial) and random error (Gaussian with mean zero and standard deviation proportional to true value) to LAI output of the vegetation model. Of the three method which are compared, the z-score normalization (normalization of observations by mean and standard deviation) provides posterior estimates for photosynthetic capacity that are close to the true values, which had been used to produce

the artificial data. Therefore the authors conclude that the z-score normalisation should generally be applied to observations in the context of model parameter inversion.

While the authors successfully demonstrate the effectiveness of the z-score normalization to minimize the impact of systematic errors of observations on inverted parameter values in the presented cases of artificial data, they fail to sufficiently analyse the normalization method and to discuss prerequisites, implicit assumptions and disadvantages. Additionally the presented artificial data account for only a part of probable kinds of errors, which influence an inversion against real world data.

An inversion of model parameter values against real observational data is simultaneously influenced by random and systematic errors of observations and by model error. The systematic error does not need to be similar or follow similar rules for all parts of the dataset: some data may be differently biased than others; even the direction of the bias may be different in different parts of the dataset. The examples of artificial data that are presented here, account for only one kind of systematic error in each set of data. It would be useful to analyse the impact of the normalization if different parts of the dataset are differently biased, and in combination with model error. Could the combination of different error types lead to spurious results? To which extend would these be due to model error or data error?

The model would not be optimized to reproduce the observed data values, but the observed data pattern, while the reproduced values could be quite different. The results of an inversion of model parameters against normalised observational data can no longer be directly validated against the given kind of observations, as the normalization assumes/accounts for a bias in the data.

The approach is based on a cost function in which the square difference of observations is divided by the variance of observed data (sigmaˆ2, equ. 1 and 2). The normalization (eqn. 5) is based on the same sigma (standard deviation of observations). On the other hand, the examples cited (Rayner et al. 2005 and Kaminski et al. 2002) are based on an error covariance matrix, which provides the opportunity to address a specific variance term to each single observation (sigmaˆ2_i). Lasslop et al. (2008) have shown that deriving individual estimates for the random error component may improve parameter retrieval in the inversion. Additionally they derive the random error component with respect to deviation from a model, not based on the variance of observations.

How should the posterior uncertainty ranges of the parameter values be interpreted if the observations have been normalized (fig 5, 6 and 7)? Do posterior parameter estimates have the same uncertainty ranges if they are derived without normalization, in cases where no systematic error has been added to the observations? This aspect should be analyzed. Figure 2 should be added to Figure 6 and 7, including uncertainty ranges of posterior parameter estimates based on not normalized observations.

Often different kinds of observations are used for parameter inversion. Is the normalization of observations applicable in these cases (e.g. Knorr and Kattge, 2005 or Santaren et al. 2007)?

The inversion of model parameters against observational data is often based on a Bayesian approach, including prior information of parameter estimates (e.g. Rayner et al. 2005). Is the normalization consistently applicable in a Bayesian context?

These are some aspects that would need to be analyzed or discussed before the z-score normalization might be applied to invert ecosystem model parameters against real world data. Other aspects to be analyzed may still be missing. Although the normalization of observations successfully reduced the impact of systematic errors in the presented cases of artificial data, I would therfore conclude that the manuscript does not yet provide sufficient background to apply the method in the context of real world data. The analysis of the z-score normalization method is too simplistic.

One question with respect to the description of the method: Page 10453 "... xmax and xmin are the maximum and minimum of observation or simulation respectively, xmean

and sigma are the mean and standard deviation of observation respectively." According to this description simulated xi would be normalized by observation xmean and sigma? I would guess simulated xi should be normalized by simulated xmean and sigma?

Interactive comment on Biogeosciences Discuss., 6, 10447, 2009.