Supplement of Biogeosciences Discuss., 11, 9215–9247, 2014
http://www.biogeosciences-discuss.net/11/9215/2014/
doi:10.5194/bgd-11-9215-2014-supplement

Biogeosciences
Discussions
Open Access

Supplement of

Using atmospheric observations to evaluate the spatiotemporal variability of CO_2 fluxes simulated by terrestrial biospheric models

Y. Fang et al.

Correspondence to: Y. Fang (yyfang@stanford.edu)

S1 Terrestrial biospheric models (TBMs)

Four TBMs participating in the North American Carbon Program Regional Interim Synthesis (NACP RIS) project (Huntzinger et al., 2012) are used in the analyses presented in this work. These TBMs were selected because their NEE fluxes are available at 3-hourly and 1°×1° resolution. The four models are the CASA coupled with the Global Fire Emissions Database (CASA-GFED, Van Der Werf et al., 2006), Simple Biosphere (SiB3, Baker et al., 2008), Organising Carbon and Hydrology In Dynamic Ecosystems (ORCHIDEE, Krinner et al., 2005) and Vegetation Global Atmosphere and Soil (VEGAS2, Zeng et al., 2005). The runs used here represent "off the shelf" simulations, and are therefore not based on a standardized protocol. Any differences in their performance can therefore be driven not only by structural differences, but also by differences in initial conditions, spin up, driver data etc. Table S1 provides a brief summary of the key features of each model, with more detail available in Huntzinger et al. (2012).

Table S1. Terrestrial Biospheric Models (TBMs) evaluated and their phenology, resolution, photosynthetic and soil carbon decomposition formulations

Model	Phenology	Native temporal resolution	Native spatial resolution	Photosynthetic formulation	# Plant functional types	# Soil pools
CASA-GFED	Diagnostic	Monthly	1°	Light Use Efficiency	3	5
ORCHIDEE	Prognostic	30 min	0.5°	Enzyme Kinetic	12	8
SiB3	Diagnostic	Hourly	1°	Enzyme Kinetic	14	0
VEGAS2	Prognostic	Daily	1°	Light Use Efficiency	4	6

S2 Covariance matrices ($\mathbf{R}$ and $\mathbf{Q}$) used in the synthetic and real data experiments

The model-data mismatch covariance matrix **R** describe the expected magnitude of discrepancies between the observed and modeled CO_2 concentrations. These errors are assumed here to be uncorrelated in space and time, and **R** is therefore a diagonal matrix with individual elements representing variances (σ_R^2) that vary across measurement towers and months. The prior flux covariance matrix $\mathbf{Q}$ characterizes the spatially- and temporally-correlated flux deviations from the model of the trend, and is modeled using a covariance function that varies as a function of the separation distance between flux times and location, as in Gourdji et al. (2012):

$$\mathbf{Q} = \sigma_Q^2 \underbrace{\left[exp\left(-\frac{\boldsymbol{h}_t}{l_t}\right)\right]}_{\substack{temoral\\covariance}} \otimes \underbrace{\left[exp\left(-\frac{\boldsymbol{h}_s}{l_s}\right)\right]}_{\substack{spatial\\covariance}} \tag{S1}$$

where σ_Q^2 is the asymptotic variance of flux deviation in space and time, $\boldsymbol{h}_t$ and $\boldsymbol{h}_s$ represent the separation lags between estimation locations in space and time, respectively, and l_s and l_t are the spatial and temporal correlation length parameters. The variance and correlation parameters vary across months. Temporal correlations are only assumed across days for the same times of the day, and not within days, so as not to risk smoothing out the diurnal variability.

S2.1 Covariance matrices ($\mathbf{R}$ and $\mathbf{Q}$) used in the real data experiments

The covariance parameters for **R** and $\mathbf{Q}$ for the RD-one-$\boldsymbol{\xi\varepsilon}$ and RD-all-$\boldsymbol{\xi\varepsilon}$ experiments (see Section 4 and Figure 2 in the main text) are estimated using Restricted Maximum Likelihood approach (*e.g.*, Gourdji et al., 2010; Gourdji et al., 2012; Michalak et al., 2004), which minimizes the negative log-likelihood of the available atmospheric measurements with respect to the covariance parameters in **R** and $\mathbf{Q}$. The corresponding objective function for a given candidate model $\mathbf{X}_c$ is (Kitanidis, 1995):

$$
\begin{aligned}
\mathrm{L} = \ln|\boldsymbol{\Sigma}| + \ln|(\mathbf{HX}_c)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{HX}_c| \\
+[\mathbf{z}^{\mathrm{T}}(\boldsymbol{\Sigma}^{-\mathbf{1}} - \boldsymbol{\Sigma}^{-\mathbf{1}}\mathbf{HX}_c((\mathbf{HX}_c)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{HX}_c)^{-1}(\mathbf{HX}_c)^{\mathrm{T}}\boldsymbol{\Sigma}^{-\mathbf{1}})\mathbf{z}]
\end{aligned}
\tag{S1}
$$

where all variables are as defined in Section 3 of the main document.

Because the $\mathbf{R}$ and $\mathbf{Q}$ parameters depend on the candidate model of the trend $\mathbf{X}_c$, and the selection of the model of the trend is affected by $\mathbf{R}$ and $\mathbf{Q}$ (Eq. 6-7), the model selection and parameter optimization proceed iteratively. The final optimized $\mathbf{R}$ and $\mathbf{Q}$ for each experiment are henceforth denoted as $\mathbf{R}_{RML}$ and $\mathbf{Q}_{RML}$. Note that for the RD-one-$\boldsymbol{\xi\varepsilon}$ experiments, different $\mathbf{R}_{RML}$ and $\mathbf{Q}_{RML}$ are obtained based the $\mathbf{X}_c$ that includes biome-month combinations for each individual TBM, and these are themselves different from the single $\mathbf{R}_{RML}$ and $\mathbf{Q}_{RML}$ obtained for the RD-all-$\boldsymbol{\xi\varepsilon}$ experiment based on the $\mathbf{X}_c$ that include biome-month combinations from all four TBMs.

S2.2 Covariance matrices (R and Q) used in the synthetic data experiments

For the SD-one-ØØ experiments that do not consider model-data mismatch, all variances in $\mathbf{R}$ are set to a nominal value of $\sigma_R^2 = 0.01\ \mathrm{ppm}^2$ for all towers and all months. The remaining synthetic data experiments (SD-one-Ø$\boldsymbol{\varepsilon}$, SD-one-$\boldsymbol{\xi\varepsilon}$ and SD-all-$\boldsymbol{\xi\varepsilon}$) all include realistic model-data mismatch errors, and the variances in $\mathbf{R}$ are set to be equal to those used in the analogous real data experiments (Section S2.1).

For synthetic data experiments with no additional spatiotemporal variability added to the underlying flux field (SD-one-ØØ and SD-one-Ø$\boldsymbol{\varepsilon}$), the variance of flux deviations from a trend including all TBM biome-month combinations is technically zero, whereas the variance/covariance of flux deviations from a trend that includes none of the TBM biome-month combinations would be equal to that of the full underlying flux field. This second setup represents a more conservative assumption, *i.e.*, does not prescribe a priori that the variability in the candidate TBM represents the true underlying variability. Consistent with this setup, the parameters of the matrix $\mathbf{Q}$ are set to those representing the full variability of the underlying fluxes, where these parameters are obtained by minimizing the negative log likelihood of the fluxes (Gourdji et al., 2010; Gourdji et al., 2008; Mueller et al., 2008):

$$\mathrm{L_Q} = \ln|\mathbf{Q}| + \ln|\mathbf{X}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{X}| + \frac{1}{2}[\boldsymbol{s}^{\mathrm{T}}(\mathbf{Q}^{-1} - \mathbf{Q}^{-1}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Q}^{-1})\boldsymbol{s}] \quad \text{(S2)}$$

Here, $\mathbf{X}$ is a simply column of ones, such that the covariance parameters represent the correlation structure of the full flux field. $\mathbf{Q}$ estimated using this method is referred as $\mathbf{Q}_{krig}$, and is different for each TBM.

For synthetic data experiments with the presence of spatially-correlated flux residuals (SD-one-$\boldsymbol{\xi\varepsilon}$ and SD-all-$\boldsymbol{\xi\varepsilon}$), the $\mathbf{Q}$ applied is $\mathbf{Q}_{RML}$ derived from the RD-all-$\boldsymbol{\xi\varepsilon}$ experiment.

S3 Flux residuals (ξ) and model-data mismatch errors (ε) in the synthetic data experiments

In all synthetic data experiments, measurements are generated as $\mathbf{z} = \mathbf{H}(\mathbf{s}_{\mathrm{TBM}} + \boldsymbol{\xi}) + \boldsymbol{\varepsilon}$, in which, $\mathbf{s}_{\mathrm{TBM}}$ is simulated NEE from a TBM, $\boldsymbol{\xi}$ represents any spatiotemporally-correlated flux residuals beyond the variability represented by the TBM, and $\boldsymbol{\varepsilon}$ represents the model-data mismatch errors.

For the SD-one-ØØ experiments, ε is a zero vector. In the SD-one-Ø$\boldsymbol{\varepsilon}$, SD-one-$\boldsymbol{\xi\varepsilon}$, and SD-all-$\boldsymbol{\xi\varepsilon}$ experiments, $\boldsymbol{\varepsilon}$ is a randomly-generated vector of independent normally-distributed values with variances corresponding to the diagonal elements (σ_R^2) of $\mathbf{R}_{RML}$ for the analogous real data experiment and a mean of 0.

When no additional spatiotemporally correlated flux residuals are included (SD-one-ØØ and SD-one-Ø$\boldsymbol{\varepsilon}$), $\boldsymbol{\xi}$ is a zero vector. In all SD cases that include realistic flux deviations (SD-one-$\boldsymbol{\xi\varepsilon}$, and

SD-all-$\boldsymbol{\xi\varepsilon}$), $\boldsymbol{\xi}$ is a randomly-generated vector of normally-distributed values with a covariance structure equal to $\mathbf{Q}_{RML}$ from the RD-all-$\boldsymbol{\xi\varepsilon}$ experiment.

References

Baker, I. T., Prihodko, L., Denning, A. S., Goulden, M., Miller, S., and Da Rocha, H. R.: Seasonal drought stress in the Amazon: Reconciling models and observations, Journal of Geophysical Research: Biogeosciences, 113, G00B01, 2008.

Gourdji, S. M., Hirsch, A. I., Mueller, K. L., Yadav, V., Andrews, A. E., and Michalak, A. M.: Regional-scale geostatistical inverse modeling of North American CO2 fluxes: a synthetic data study, Atmos. Chem. Phys., 10, 6151-6167, 2010.

Gourdji, S. M., Mueller, K. L., Schaefer, K., and Michalak, A. M.: Global monthly averaged CO2 fluxes recovered using a geostatistical inverse modeling approach: 2. Results including auxiliary environmental data, Journal of Geophysical Research: Atmospheres, 113, D21115, 2008.

Gourdji, S. M., Mueller, K. L., Yadav, V., Huntzinger, D. N., Andrews, A. E., Trudeau, M., Petron, G., Nehrkorn, T., Eluszkiewicz, J., Henderson, J., Wen, D., Lin, J., Fischer, M., Sweeney, C., and Michalak, A. M.: North American CO2 exchange: inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion, Biogeosciences, 9, 457-475, 2012.

Huntzinger, D. N., Post, W. M., Wei, Y., Michalak, A. M., West, T. O., Jacobson, A. R., Baker, I. T., Chen, J. M., Davis, K. J., Hayes, D. J., Hoffman, F. M., Jain, A. K., Liu, S., Mcguire, A. D., Neilson, R. P., Potter, C., Poulter, B., Price, D., Raczka, B. M., Tian, H. Q., Thornton, P., Tomelleri, E., Viovy, N., Xiao, J., Yuan, W., Zeng, N., Zhao, M., and Cook, R.: North American Carbon Program (NACP) regional interim synthesis: Terrestrial biospheric model intercomparison, Ecological Modelling, 232, 144-157, 2012.

Kitanidis, P. K.: Quasi-Linear Geostatistical Theory for Inversing, Water Resources Research, 31, 2411-2419, 1995.

Krinner, G., Viovy, N., De Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C.: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, Global Biogeochemical Cycles, 19, GB1015, 2005.

Mueller, K. L., Gourdji, S. M., and Michalak, A. M.: Global monthly averaged CO2 fluxes recovered using a geostatistical inverse modeling approach: 1. Results using atmospheric measurements, Journal of Geophysical Research: Atmospheres, 113, D21114, 2008.

Van Der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Kasibhatla, P. S., and Arellano Jr, A. F.: Interannual variability in global biomass burning emissions from 1997 to 2004, Atmos. Chem. Phys., 6, 3423-3441, 2006.

Zeng, N., Mariotti, A., and Wetzel, P.: Terrestrial mechanisms of interannual CO2 variability, Global Biogeochemical Cycles, 2005. 19, 2005.