APPLICATION OF A GEOSTATISTICAL KALMAN SMOOTHER TO THE ESTIMATION OF MONTHLY GRIDSCALE FLUXES OF CARBON DIOXIDE

A.M. Michalak[1], K. Mueller[1], S. Gourdji[1], C. Humphriss[1], L. Bruhwiler[2], K. Schaefer[2], and P.P. Tans[2]

[1] *Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI 48109-2125; anna.michalak@umich.edu*

[2] *NOAA Climate Monitoring and Diagnostics Laboratory, 325 Broadway, Boulder, CO 80305;*

ABSTRACT
Inverse modeling methods are now commonly used for estimating surface fluxes of carbon dioxide, using atmospheric mass fraction measurements combined with a numerical atmospheric transport model. Michalak et al. [2004] recently developed a geostatistical approach to flux estimation that takes advantage of the spatial and/or temporal correlation in fluxes and does not require prior flux estimates. In this work, a geostatistical implementation of a fixed-lag Kalman smoother is developed and applied to the recovery of gridscale carbon dioxide fluxes for 1997 – 2001 using data from the NOAA-CMDL Cooperative Air Sampling Network.

INTRODUCTION AND METHODOLOGY
The goal of this work is to estimate monthly carbon dioxide fluxes using data from the NOAA-CMDL Cooperative Air Sampling Network without using a prior estimate of the flux distribution. We first develop a geostatistical implementation of a fixed-lag Kalman smoother to decrease the computational cost of gridscale geostatistical inversions. Second, we incorporate information on auxiliary environmental variables to further constrain fluxes. Third, we apply the developed methods to the recovery of monthly-averaged carbon dioxide fluxes for 1997-2001.

Geostatistical inverse modeling
The geostatistical approach to inverse modeling is a Bayesian approach in which the prior probability density function is based on an assumed form for the spatial and/or temporal correlation of the surface fluxes to be estimated. This differs from the traditional Bayesian approaches, where the prior information is in the form of initial surface flux estimates. Geostatistical flux estimates are not subject to some of the limitations of traditional Bayesian inversions, such as potential biases created by the choice of prior fluxes and aggregation error resulting from the use of large regions with prescribed flux patterns. The geostatistical approach is also ideally suited to inversions at fine spatial scales. The objective function used in the solution of a linear geostatistical inverse problem is:

$$L_{s,\beta} = (\mathbf{z} - \mathbf{Hs})^T \mathbf{R}^{-1} (\mathbf{z} - \mathbf{Hs}) + (\mathbf{s} - \mathbf{X}\beta)^T \mathbf{Q}^{-1} (\mathbf{s} - \mathbf{X}\beta)$$

where $\mathbf{H}$ is a matrix of sensitivities of the observations $\mathbf{z}$ to the discretized unknown surface flux distribution $\mathbf{s}$, $\mathbf{R}$ is the model-data mismatch covariance matrix, $\mathbf{X}\beta$ is the model of the mean of the flux distribution**,** where $\mathbf{X}$ contains known information and β are unknown drift coefficients (e.g. the fluxes can have a constant but unknown mean), and the flux covariance matrix $\mathbf{Q}$ is based on a spatial and/or temporal correlation structure. The inverse problem involves solving for both β and $\mathbf{s}$, and the form of the solution is therefore different from the classical Bayesian setup [*Michalak et al.* 2004]. No prior flux estimates are used, and this term in the objective function is replaced by a description of the spatial (and/or temporal) covariance of deviations of fluxes from their mean behavior.

Fixed-lag Kalman smoother
Gridscale inversions become computationally prohibitive as the period for which fluxes are estimated increases. In order to address this issue, we develop a geostatistical implementation of a fixed-lag Kalman smoother [*Bruhwiler et al.* 2005]. The Kalman smoother allows for the sequential estimation of fluxes using an appropriate portion of available observations, while taking into account the inferred temporal covariance between fluxes estimated for various times. The form of the solution developed in Bruhwiler et al. [2005] was compatible with the classical Bayesian approach. For the case of monthly flux estimates, independently obtained flux estimates (typically from flux inventories) are used as prior information the first time a given month of fluxes is estimated, and the latest (*a posteriori*) estimate is updated in the subsequent steps using additional months of atmospheric data. In the geostatistical approach, the system needs to account for the unknown components of the model of the mean (β) in obtaining the first estimate of a given month's fluxes, but needs to use the latest (*a posteriori*) estimates for subsequent estimates of a given month's fluxes. This requires a substantial modification to the form of the Kalman smoother because each step through the smoother involves both months being estimated for the first time (with no

prior flux estimate), and months being estimated for at least the second time (with the latest flux estimates used as priors). The details of the mathematical development are presented in an upcoming paper [*Michalak et al*. in prep.]

Use of auxiliary data and model of the mean

In Michalak et al. [2004], land and ocean fluxes were assumed to tend toward different but constant means. Because deviations from these means were correlated, this simple model of the mean still allowed for complex flux patterns. In the current work, we model the carbon flux mean using objective auxiliary variables known to be associated with the global carbon cycle (e.g. precipitation, temperature, land use, population density, Normalized Difference Vegetation Index (NDVI)) (see Fig. 1). This allows the model greater flexibility, while still not requiring the posterior fluxes to follow any pre-prescribed flux patterns. The model of the mean is constrained by a covariance matrix which assumes separate correlation structures over land and ocean. In addition, carbon fluxes over the land are influenced by different auxiliary variables than those over the ocean. A second pass at this geostatistical inverse modeling approach will use more "subjective" variables in the model of the mean, such as the output of biospheric models and extrapolated measurements of ocean carbon dioxide concentrations. This approach will be closer to the traditional Bayesian approach.

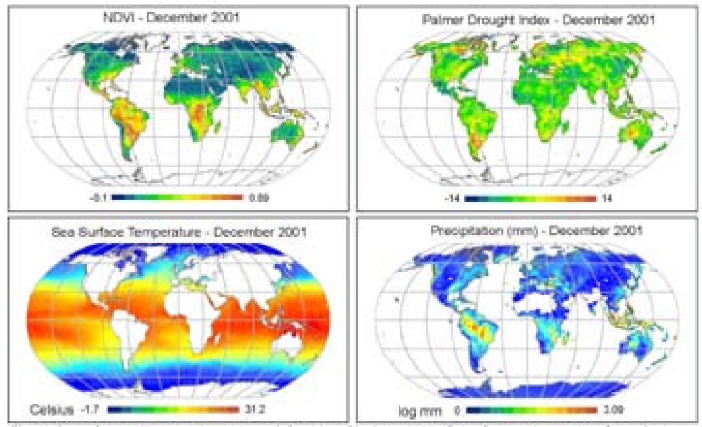

Fig. 1: Sample auxiliary environmental data used in inversion (See electronic version for color).

SAMPLE SETUP

In the presented setup, the sensitivity of the atmospheric measurements to surface fluxes is calculated using an adjoint implementation of the TM3 atmospheric transport model [*Rödenbeck et al.* 2003]. Data from 39 sites from the NOAA-CMDL cooperative air sampling network are used to constrain the fluxes. Surface fluxes are estimated first using a constant mean model [*Michalak et al.* 2004], and second using a model that contains both a constant component and components that take into account auxiliary environmental data. For both cases, the Kalman smoother is used with the geostatistical inverse modeling approach to estimate global monthly carbon dioxide fluxes from 1997 through 2001 on a 3.75° latitude by 5.0° longitude grid.

RESULTS AND DISCUSSION

Results indicate that even the constant mean model yields flux estimates that agree well with independent flux information for well-constrained areas of the globe (e.g. temperate North America). The complex mean model provides additional information at smaller scales, and allows for sharper discontinuities in the flux distributions, corresponding to changes in environmental regimes (e.g. different land cover types). Because these estimates do not take into account bottom-up flux estimates that are typically used as prior information in Bayesian inversions, they provide an independent source of information. These estimates are particularly interesting in interpreting (i) carbon fluxes in regions that are less well constrained by atmospheric data and where Bayesian inversions largely revert to prior estimates with little reduction in uncertainty, and (ii) interannual flux variability that is not captured by bottom-up estimates (e.g. ENSO effects on ocean fluxes). One critical factor to consider is the trade-off between including additional information in the inversion setup to reduce the flux uncertainty versus relying more heavily on the atmospheric data. The ultimate setup should always depend on the available data, the specific scientific question to be answered, and the assumptions that are considered acceptable for a given application.

REFERENCES

Bruhwiler, L.M.P., A.M. Michalak, W. Peters, D.F. Baker, and P. Tans (2005), An Improved Kalman Smoother for Atmospheric Inversions, *Atmos. Chem. Phys. Disc.*, 5, 1891-1923.

Michalak, A.M., L. Bruhwiler, and P.P. Tans (2004), A Geostatistical Approach to Surface Flux Estimation of Atmospheric Trace Gases, *J. Geophys. Res.*, 109, D14109, doi:10.1029/2003JD004422.

Rödenbeck, C., S. Houweling, M. Gloor, and M. Heimann (2003), CO_2 flux history 1982-2001 inferred from atmospheric data using a global inversion of atmospheric transport, *Atmos. Chem. Phys.* 3, 1919-1964.