Model-data integration and network design for biogeochemical research: an NCAR-CSU summer school

Dave Schimel, NCAR CGD and Britt Stephens, NCAR ATD, co-chairs

Steering Committee: Dennis Ojima (CSU), Scott Denning (CSU), Tomislava Vukicevic (CSU), Scott Doney (CGD), Natalie Mahowald (CGD), Beth Holland (ACD), David Baker (ASP), Peter Thornton (CGD), Roger Dargaville (CGD), Kathy Hibbard (Univ. New Hampshire), Don Middleton (SCD), Doug Nychka (CGD/GSP), Jeff Anderson (GFDL and NCAR)

Abstract: Global biogeochemical research must increasingly address the problems of "detection" or quantification of changing fluxes to the atmosphere, and "attribution" or explanation of those fluxes in terms of specific mechanisms. Today, neither our measurement nor analysis capabilities are sufficient to meet the twin challenges of biogeochemical detection and attribution with sufficient accuracy and resolution. We propose a summer school to advance both analysis techniques (inverse and assimilation modeling) and observing system design. The Summer School will involves lectures from a broad and distinguished group of scientists on the biogeochemical cycles, current and planned measurement capability, process and data analytical modeling, and new approaches in applied math. It will have as its centerpiece a hands-on simulation exercise. Estimates of global terrestrial and oceanic fluxes will be produced from existing data and models, combined to produce flux fields with reasonable time-space variability. They will be distributed in a global simulated atmosphere using an atmospheric transport model, run at the maximum achievable resolution to produce a 4-D data set of concentrations. The participants will form competing teams to reconstruct surface fluxes. Each team will choose a measurement strategy and obtain data corresponding to that strategy by querying a referee. Each measurement will have a "price" and each team will have an identical budget in dollars to "fund" its observing program. Each team will have access to the same models. A mesoscale carbon assimilation model being developed at NCAR and Colorado State University will be available for the participants to use in network design, and to demonstrate concepts and application of data assimilation in biogeochemistry. The teams may choose any strategy they wish to estimate the "real" fluxes. The conclusions from the workshop will be summarized in a report and a series of articles on 1) design of the exercise and implications for observing system design, 2) a review of data assimilation approaches for biogeochemical cycle research, 3) possibilities for use of the simulation exercise in education. In the year following the workshop, we will make the psuedodata bases and analysis tools available and improve them to better meet the research needs of the biogeochemical research community.

Background: In the past few years, an imperative has developed to quantify the fluxes at the Earth's surface of trace gases and photochemical precursors, especially carbon dioxide, methane, carbon monoxide and reactive nitrogen species. To borrow terminology, this is the "*detection*" problem of carbon cycle research. Accompanying this is the equal imperative to understand the causes of changing fluxes, the "*attribution*" problem. This is a profoundly interdisciplinary problem, as the proximate causes of trace gas exchange lie in the details of both ecosystem and industrial metabolism, and their interactions (e.g., the impact of fertilizers on biological trace gas emission). An analytical framework based on model-data fusion techniques such as data assimilation needs to be developed to link observations with process-based knowledge.

The fluxes of gases at the land surface have a number of defining characteristics that make their estimation of the great scientific challenges in the Earth and Life sciences. First, the fluxes are heterogeneous in space and time, as a result of nonlinear, multiscaled processes and thus local measurements can only be interpreted in light of extensive data on processes and forcing (Schimel et al 1997). Second, the fluxes of all major trace gas species depend strongly on human impacts and are greatly affected by both direct human activity (land management and industry) and indirect effects (such as through nitrogen deposition, climate change and more diffuse human activities) (Schimel 1995).

Because of the complexity of integrating fluxes from *in situ* field studies via "bottom-up" techniques, increasing attention has focused on the use of the atmosphere as integrator, to provide a check on the estimation of aggregated fluxes in space and time (Vukicevic et al 2001). This approach has been used at multiple scales. First, the atmosphere can be used as an integrator of fluxes globally, and its variability in time and space provides diagnostics of fluxes on continental or ocean basic scales (Bousquet et al 2000). Spatial patterns in CO_2 are used to infer the well-known "Northern Hemisphere Sink" while temporal variations in the growth rate of CO_2 strongly suggest a link between El Nino-induced fluctuations in terrestrial rainfall and carbon exchange. Taken as a whole, the atmosphere provides a constraint on the global budget of added anthropogenic CO_2 fluxes. The availability of global CO_2 data is likely to change dramatically soon, if promising satellite techniques come on line (Engelen et al 2001). However, even in the most optimistic view, atmospheric data provide only indirect information on processes and so we must combine atmospheric data with terrestrial and oceanic process data and models.

At the other extreme, the variations in CO_2 concentrations at the local (meters to kilometers) scale and on the turbulent time-scale are used via the eddy correlation technique to infer fluxes over specific ecosystems (Falge et al 2001). A global network of eddy correlation sites has been developed. Locally, eddy correlation towers provide information on the relationship between drivers such as radiation and soil moisture and photosynthesis and respiration fluxes for specific ecosystems. A few towers have long and reliable enough records to begin providing clues about interannual variability and “slow” processes of carbon accumulation (Goulden et al 1995). Eddy correlation and related techniques also provide information about many other trace species as well and considerable data, for example, for methane and ozone.

At the third scale, mesoscale studies of CO_2 have been initiated, using a variety of atmospheric budgeting techniques (Gloor et al 2001). While tower-scale studies provide integration at the ecosystem scale, difficult scaling problems lie between the ecosystem and global scales. Tall towers and aircraft can provide integration at the landscape or regional scale to help constrain integration using bottom-up techniques, and to detect other problems intrinsic to surface flux measurements. Aircraft data measure regional patterns of concentration and flux, but detect the convolved signals of complex surface flux fields, heterogeneous vertical mixing and entrainment and free tropospheric transport. Thus, the surface fluxes must be retrieved using strong constraints from atmospheric motion, observed or modeled.

A Summer School: Getting closure on trace gas fluxes, both by closing budgets and by attributing fluxes to specific processes requires an unprecedented degree of scientific integration. This is a large problem involving the physical and biological sciences and human dimensions equally. We propose a summer school to advance both analysis techniques (inverse and assimilation modeling) and observing system design. The workshop currently has sponsorship from NCAR’s new Biogeosciences Program and the Colorado State-centered NSF Integrated Research Challenge grant on regional carbon modeling (PI Dennis Ojima, BIO Directorate), with several other requests pending.

We propose a Summer School to be held in Boulder for two weeks in June of 2002. The Summer School will bring together an interdisciplinary group of scientists interested in biogeochemistry to develop a new strategy for carbon cycle research. The meeting will bring together researchers interested in models and data analysts, measurement networks, and new measurement capability development. We will begin with overview talks on the state of the art in measurement techniques for fluxes and concentrations, on the status of global flux and concentration networks, on estimation techniques (inverse, assimilation, Kalman Filter, etc), for atmosphere, land and ocean systems. We will discuss model-data fusion at ecosystem, landscape-continental and global scales, taking advantage of the questions, measurement approaches and models at those scales. In addition to data assimilation and inverse approaches based on atmospheric models, we will also review applications of these techniques to ecosystem process models and data such as eddy correlation fluxes and results from experimental manipulation studies such as free-air carbon dioxide enrichment studies. We will brainstorm, on the one hand, new measurements and how we might learn from them, and on the other, analysis strategies and what new measurements they would require. The goal is to identify candidate new measurement and analysis strategies.

Participants will include invited scientists representing: 1) different scales of measurement and modeling, 2) the relevant “domains”, atmosphere, ocean, land and direct human dimensions (fossil and land use fluxes), plus applied mathematicians, 3) a range of approaches including instrument design, network design, process studies, process and global modeling and inverse and assimilation techniques. We will also invite a few individuals who have been prominent in innovative earth sciences education (e.g. Arthur Few). Applications and nominations of students and post-docs will be taken and we hope to have 20-30 student participants. Both U.S. and international participants are expected.

The centerpiece of the Summer School will be a hands-on exercise to explore new approaches in observations, network design and analysis. Estimates of global terrestrial and oceanic fluxes will be produced from existing data and models, combined to produce flux fields with reasonable time-space variability. They will be distributed in a global simulated atmosphere using an atmospheric transport model, run at the maximum achievable resolution (80-100 km) to produce a 4-D data set of concentrations. This database will serve as the basis for the exercise. A series of access tools will be developed to allow students to query the database via a referee. The tools will sample the simulation in ways that correspond to flux towers, Tall Towers, aircraft flights, permanent monitoring stations (the CMDL network) and a satellite retrieval. Within the US we will allow queries of soil and vegetation carbon stocks based on VEMAP estimates. Each technique will have a specified precision and accuracy, and a dollar value.

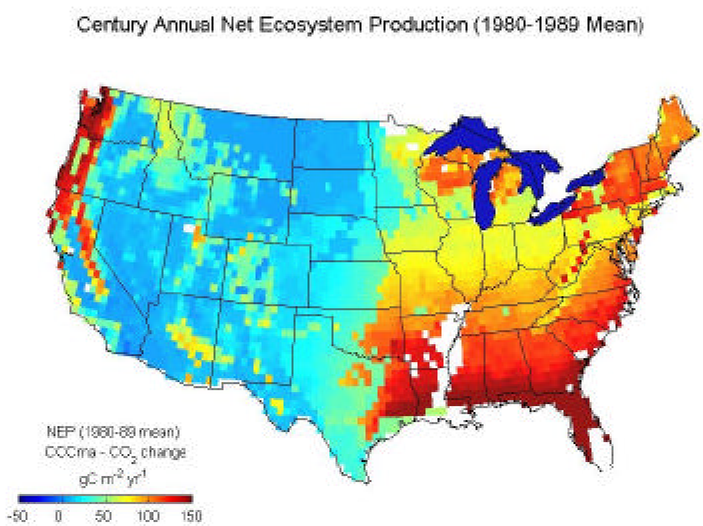

Figure 1 Net ecosystem exchange for natural and agricultural ecosystems of the US 48 States. These fluxes will be combined with similar, global, fluxes as input to a transport model to produce a simulated atmosphere.

The participants will be encouraged to use any tools they choose to analyze the psuedodata. We will encourage all participants to port their preferred analysis tools to NCAR prior to the meeting, or bring them on laptops if they do not require large computers. NCAR staff will support the participants in porting code to NCAR during the three months prior to the workshop, to make sure that it works during the two-week

summer school. In addition, we are developing a unique tool for regional analysis. We will use the NCAR Atmospheric Chemistry Division's HANK mesoscale chemical transport model, together with its adjoint, as the basis for a regional carbon dioxide assimilation system (Vukicevic and Hess 2000). This system, which will be in prototype for the workshop, will use atmospheric concentration measurements as input, and as output will report surface fluxes. It can be run with grid resolutions from ~1 to100 km, and is forced with either model-generated or analysed winds. HANK is based on MM5 and HANK and its adjoint use parameterizations for boundary layer mixing and convection that are consistent with MM5. The team developing this simple regional carbon assimilation system includes Dave Schimel, Peter Hess (the author of HANK), Tomislava Vukicevic (developer of the HANK adjoint), and NCAR post-docs David Baker and Roger Dargaville. The assimilation system will use either a variational approach (Vukicevic et al 2001) or a Kalman Filter (Anderson et al 1999).
The participants will form two or more teams (Red and Blue) competing to reconstruct global and North American fluxes. Each team will choose a measurement strategy and obtain data corresponding to that strategy by querying the referee. Each team will have an identical budget in dollars to "fund" its chosen observing program. Each team will have access to the same global and mesoscale atmospheric models, and an ecosystem model, and these models will differ from those used to produce the data set. The teams may choose any strategy they wish to estimate the "real" fluxes, including pilot studies, followed by intensives, distributed networks with long time series or a combination of approaches. Participants on each team will be selected so that the teams are roughly balanced in terms of experience and technical skills, both teams will include participants experienced in atmospheric, ecosystem, inverse and data assimilation modeling as well as statisticians from NCAR's Geophysical Statistics Project. NCAR will provide computing equally to both teams.

The competition will address network design, observing strategies and data analysis approaches. This type of concentrated exercise provides important results and a foundation for the participants to continue with the process of experimental design and analysis. The conclusions from the workshop will be summarized in a report and a series of articles, which will be submitted to a peer reviewed journal for rapid publication. We anticipate an immediate report in EOS and longer papers as well. One paper will address the design and use of the simulation exercise (this could be submitted together with the data base to Earth Interactions), one paper will focus on advances in network design and measurement strategy, and a third will summarize advances in assimilation approaches. A final paper, if possible, will describe the educational aspects of the Summer School and the simulation exercise. This last paper could be submitted to Bulletin of the AMS, or to Ecological Applications. We expect that this type of system will be a great tool for teaching global biogeochemical cycles and will schedule a discussion during the workshop to get the participant's ideas on how best to pursue educational applications.

During the second year, based on experienced gained in the Summer School, we will follow up as needed. We expect to develop an improved version of the atmospheric carbon dioxide data set, and we also will develop versions of the atmospheric data set with specific types of sources and sinks added in. We will consider several hypotheses about land-use and disturbance -driven sources and sinks not included in the initial data sets. These terms could include a major western US fire season, recovery from Gyspie Moth in the Northern tier states (as is suggested in some satellite analyses) or a perturbation to global fluxes from ENSO effects on the ocean. Additional exercises using these alternate answers would ask the question, could a network optimized using the basic fluxes detect and correctly quantify surprise changes to the fluxes. A steering committee meeting, plus additional key participants will occur late in the second year to complete publications and reporting back to the communities and agencies.

We will also continue to improve the analysis tools developed for the Summer School. We will make the code for the regional assimilation model available, and anticipate many groups will build on this and apply it in innovative ways. The access tools for the data set will be integrated into a web-based solution so that users can query the large atmospheric CO_2 psuedodata. Part-time support for an Associate Scientist to work with the participants and new users is requested for the second year.

Implementation: The meeting will be organized by co-Chairs Britt Stephens and Dave Schimel (NCAR). A local committee will assist, and will also take primary responsibility for developing the simulation exercise for the Summer School. It will include Dennis Ojima, Scott Denning (CSU), Tomislava Vukicevic (CSU), Scott Doney, Natalie Mahowald, Beth Holland, David Baker, Peter Thornton, Phil Rasch and Roger Dargaville, together with Kathy Hibbard from the Carbon Joint Project. The organization of travel, lecture halls, audiovisual and accommodations will be conventional and will be organized by NCAR, with support from NCAR CGD support staff (Susan Chavez) with assistance from the Advanced Studies Program. The venue will be Boulder, Colorado, to allow support from NCAR's computing resources.

NCAR's scientific computing division (SCD) will provide a dedicated 16-processor machine for the duration of the Summer School and additional processors if necessary. Computing and data support will be provided to the participants and broader community after the workshop as well. SCD will also assist in other ways, and we will integrate this exercise with the Community Data Portal to provide access and display/data management tools. The meeting and required computing labs will either be in NCAR facilities or on the CU Campus (a fast link is available). We anticipate substantial involvement of NCAR's SOARS students in the Summer School.

Deliverables

Deliverables from this project will include:

1) A new generation of scientists from many disciplines trained in model-data fusion techniques and observing system issues,
2) Evaluation of current and planned observing systems and analytical approaches is a formal framework,
3) An integrated 4-D data base for research and educational applications,

4) Tutorial material and simulations for model-data fusion analyses,
5) Publications documenting the results

References cited

Anderson, J. L., and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. Monthly Weather Review, 127(12), p 2741-2758.

Bousquet P, Peylin P, Ciais P, Le Quere C, Friedlingstein P, and Tans PP. Regional changes in carbon dioxide fluxes of land and oceans since 1980. SCIENCE 290 (5495): 1342-1346

Engelen RJ, Stephens GL, and Denning AS. The effect Of CO2 variability on the retrieval of atmospheric temperatures. GRL 28 (17): 3259-3262

Falge E, Baldocchi D, Olson R, Anthoni P, Aubinet M, Bernhofer C, Burba G, Ceulemans R, Clement R, Dolman H, Granier A, Gross P, Grunwald T, Hollinger D, Jensen NO, Katul G, Keronen P, Kowalski A, Lai CT, Law BE, Meyers T, Moncrieff H, Moors E, Munger JW, Pilegaard K, Rannik U, Rebmann C, Suyker A, Tenhunen J, Tu K, Verma S, Vesala T, Wilson K, and Wofsy S. Gap filling strategies for defensible annual sums of net ecosystem exchange. Ag. Forest Met. 107 (1): 43-69

Gloor M, Bakwin P, Hurst D, Lock L, Draxler R, and Tans P. What is the concentration footprint of a tall tower? J. Geophys. Res.: 106 (D16): 17831-17840

Goulden ML, Wofsy SC, Harden JW, Trumbore SE, Crill PM, Gower ST, Fries T, Daube BC, Fan SM, Sutton DJ, Bazzaz A, and Munger JW. Sensitivity of boreal forest carbon balance to soil thaw. Science 279 (5348): 214-217

Pacala SW, Hurtt GC, Baker D, Peylin P, Houghton RA, Birdsey RA, Heath L, Sundquist ET, Stallard RF, Ciais P, Moorcroft P, Caspersen JP, Shevliakova E, Moore B, Kohlmaier G, Holland E, Gloor M, Harmon ME, Fan SM, Sarmiento JL, Goodale CL, Schimel D, and Field CB. Consistent land- and atmosphere-based US carbon sink estimates. Science 292: 2316-2320

Vukicevic T, Braswell BH, and Schimel D. A diagnostic study of temperature controls on global terrestrial carbon exchange. Tellus B: 53: 150-170

Vukicevic, T., and P. G. Hess, 2000: Analysis of tropospheric transport in t he Pacific basin using the adjoint technique. J. Geophys. Res., 105: 7213-7230.

Schimel DS, Emanuel W, Rizzo B, Smith T, Woodward FI, Fisher H, Kittel TGF, McKeown R, Painter T, Rosenbloom N, Ojima DS, Parton WJ, Kicklighter DW, McGuire AD, Melillo JM, Pan Y, Haxeltine A, Prentice C, Sitch S, Hibbard K, Nemani R, Pierce L, Running S, Borchers J, Chaney J, Neilson R, Braswell BH. Continental scale variability in ecosystem processes: Models, data, and the role of disturbance. Ecol. Monogr. 67 (2): 251-271