North American Carbon Project (NACP) Spatial Model-Data Comparison Project (MDC) for regional and continental synthesis

1. Motivation

Available observations are localized and widely separated in both space and time, so we depend heavily on models to characterize, understand, and predict carbon fluxes at regional or global scales. The results from models differ from each other because they use different approaches (forward vs. inverse), modeling strategies (detailed process, statistical, observation based), process representation, boundary conditions, initial conditions, and driver data. We need an approach to identifying the causes of differences and deciding on which formulations and approaches best align with measurements, and why they may or may not agree with measurements.

Inverse or top-down analyses can provide estimates of carbon fluxes that are optimally consistent with measurements. Forward, or bottom-up ecosystem models can be used to test hypotheses concerning the attribution of processes in determining carbon fluxes and make projections using forcing scenarios. Examining and comparing results of inverse and forward model simulations with each other and with suitable measurements can help evaluate model strengths/weaknesses and utility, and provide multiple views of spatial and temporal patterns of fluxes, lead to understanding of processes involved, and provide a basis for making projections

Any interested party with a simulation model capable of calculating surface fluxes of carbon (also sensible and latent heat) either by employing an inverse or forward analysis (or both) for as much of the time period between 2000 to 2005 is welcome to participate. Simulations results should cover a sizable portion of the North America continent. Each additional model, even if not for the entire area or time domain, provides information that can contribute to the analyses.

2. Specific regional hypotheses that can be addressed

The Spatial MDC will identify and quantify spatial and temporal patterns of C fluxes, quantify model uncertainty and bias by comparing simulated or estimated surface fluxes and biomass to observed values at regional to continental scales for the period 2000-2005. Candidate "measurements" for regional comparisons include:

MODIS NPP, LAI (8 day, annual)

NASS crop yield (annual)

FIA based estimates of forest increment (annualized), biobass

AmeriFlux NEE (for those that haven't used these for calibration) (half-hourly, daily)

Additional available measurements? EMDI- NPP data, LIDET data, etc.?

Some specific hypotheses have been suggested by CASA (Potter et al. 2007) and CarbonTracker (Peters et al. 2007) results:

1. Do model results and observations show consistent spatial patterns in response to the 2002 drought? From measurements and model, can we infer what processes were affected by the 2002 drought?
2. What is the spatial pattern and magnitude of interannual variation in carbon sources and sinks? What are the components of carbon fluxes and pools that contribute to this variation?
3. What are the magnitudes and spatial distribution of carbon sources and sinks, and their uncertainties during the period 2000-2005?

Other questions are more diagnostic in nature and could be used to examine the magnitude of model uncertainty, such as

4. What are model sensitivities to model drivers?
5. How do similar model parameters vary across models?
6. How does process understanding vary among models, between regions, and across scales (site-region, region-continent)?

3. Project Management

The Spatial MDIP will have 2 related components – ecosystem (bottom-up) models for forward analyses, and similar models used for inversion analyses. These are complementary approaches to analyzing the synthesis questions.

The inverse analysis models will provide model results at the spatial resolution of the North America TRANSCOM regions, and monthly temporal resolution. If results are available at finer spatial and temporal resolution (for example CarbonTracker) then these can also be provided since the bottom-up models will be at finer scales. Most inverse modeling groups have indicated that 1° spatial resolution is feasible.

The forward or bottom-up models can provide analysis results at the temporal and spatial resolution of simulation runs more-or-less in hand. These will range from hourly to annual time steps with most probably being hourly, daily or twice daily (day vs. night), and 1km to county to 1° spatial resolution. MAST-DC, with help from the participants, will decide on spatial and temporal scales for aggregating/interpolating results for the planned analyses. One aggregation will be the TRANSCOM scales for comparison with inversion results. A finer target spatial resolution for comparing inverse model to forward/ecosystem models will be 1° with cells centered on half degrees. Both types of models will be asked to summarize monthly total terrestrial net ecosystem exchange at this spatial resolution. Forward ecosystem models compute many components of these net ecosystem exchange including GPP, NPP, Ra, Rh, and other diagnostic components that can be compared to observations such as LAI, aboveground biomass, crop yield, evapotranspiration, soil moisture, or soil carbon.

Regional MDC Management Team

To ensure rapid and steady progress, several individuals will lead and organize the regional MDC:

Andrew Jacobson
University of Colorado and NOAA Earth System Research Laboratory

andy.jacobson@noaa.gov
303-497-4916

Mac Post
Oak Ridge National Laboratory
postwmiii@ornl.gov
865-576-3431

Regional MDC Server

We will create a central data repository that will house data, model descriptions, documentation, codes, and analysis results. MAST-DC will provide some standard software tools and assitance to help participants convert data and model output into the standard format for use in model-data comparison. We will incorporate standard security procedures to ensure only MIP participants can access the repository. In addition, we will implement a data fair use policy.

Data and Model Output Fair Use policy

The Site MDC will involve scientists from a large number of independently funded research projects. To ensure the individuals and teams that provide model output and data receive proper credit for their work, we have instituted a Fair Use Policy. The policy applies to all data and model output stored on the Regional MDC server. The Fair Use Policy is based on the Ameriflux Policy, but expanded to include all Site MDC participants:

The data and model output provided on this site are freely available and were furnished by individual scientists who encourage their use. Please kindly inform in writing (or e-mail) the appropriate participating scientist(s) of how you are using the data and of any publication plans. If not yet published, please reference the source of the data or model output as a citation or in the acknowledgments. The scientists who provided the data or model output will tell you if they feel they should be acknowledged or offered participation as authors. We assume that an agreement on such matters will be reached before publishing and/or use of the data for publication. If your work directly competes with an ongoing investigation, the scientists who provided the data or model output may ask that they have the opportunity to submit a manuscript before you submit one that uses their data or model output. When publishing, please acknowledge the agency that supported the research. We kindly request that those publishing papers using AmeriFlux data, Fluxnet Canada data, or Regional MDC model output provide reprints to the appropriate scientist providing the data or model output, and to the data archive at the Carbon Dioxide Information Analysis Center (CDIAC).

Regional MDC Email Lists

The Site MDC involves a large number of modelers, experimenters, observationalists, program managers, and other interested parties widely distributed across North America. To facilitate effective communication, we will create participant email lists to disseminate information. As required, we will create smaller email lists consisting of subsets of the full participant list to focus on specific problems or research efforts (for example, see the existing inversion modeler list on the NACP website – www.nacarbon.com/nacp).

4. Model Simulation Protocols

Regional MDC participants fall into two, not necessarily distinct, groups: model participants and data providers. The description provided below is preliminary but will be distributed early in the project and will be modified as necessary following initial discussion by all measurement and modeling participants.

Information Provided by Participants

All potential model participants should provide general descriptive information about the model they used to produce their results. This includes a short model overview (1-2 paragraphs) with a brief description, basic structure, model initialization procedure, description of driver data, web pages, and associated references. Participants should also provide a primary point of contact and, if desired, secondary points of contact for each model. Lastly, we recognize that the required inputs for each model differ, so the participants should provide a list of all inputs used by their model. In particular, vegetation type or land cover classification should be provided. This will allow some subsetting of results for comparisons by vegetation types.

Data providers should provide a short overview of their measurements, a description of how they are derived, estimates of uncertainty, and associated references.

Inputs to Model

This Regional MDC will not prescribe model driver data. Weather, phenology, soil properties, N deposition, etc. is left up to the modeling teams. For addressing the question of the absolute size of the carbon sources and sinks, initialization of carbon pools will be important and each participant should provide a description of their initialization technique. Those models with nitrogen or phosphorus biogeochemistry should initialize those variables as best suited for their model and provide a description of the procedure.

Outputs from Model

Each model will save a core set of required variables and a secondary set of optional variables. Participants will convert their output to standard units and file format. Table 1 is provided for guidance. This is taken from the ALMA netCDF convention mentioned below. Variables of interest include NEE for inverse and forward ecosystem models. For the purposes of being able to diagnose the causes of C sources and sinks, forward ecosystems models should also, at minimum provide NPP, Rh and if possible GPP, Ra. Not all models calculate the full set of these but should provide what they can with NPP and NEE being the most important. The models should also provide total live biomass (TotLivBiom) and total soil carbon (TotSoilCarb). If there are multiple pools these should be saved for potential use during the comparison. If the models can provide sensible heat, and latent heat; and prognostic variables like soil temperature these should be saved and potentially contributed if they look like they are useful in an emerging analysis.

Table 1. Model outputs - core set of required variables and secondary set of optional variables. Allowances will be made for models that don't calculate some of these.

Variable Name	**Description**	**Definition**	**Units**	**Positive Dir. (Traditional)**	**Positive Dir. (Mathematical)**	**Priority**
GPP	Gross Primary Production	Net assimilation of carbon by the vegetation	Kg/m2/s2	Downward	Upward	Mandatory
NPP	Net Primary Production	Carbon assimilation by photosynthesis	Kg/m2/s2	Downward	Upward	Mandatory
NEE	Net Ecosystem Exchange	Sum of all carbon fluxes exchanged between the surface and the atmosphere	Kg/m2/s2	Upward	Upward	Mandatory
AutoResp	Autotrophic Respiration	Autotrophic respiration includes maintenance respiration and growth respiration	Kg/m2/s2	Upward	Upward	Recommended
HeteroResp	Heterotrophic Respiration	Total flux from decomposition of organic matter	Kg/m2/s2	Upward	Upward	Recommended
TotSoilCarb	Total Soil Carbon	Total soil and litter carbon content integrated over the entire soil profile	Kg/m2	-	-	Recommended
TotLivBiom	Total Living Biomass	Total carbon content ofthe living biomass	Kg/m2	-	-	Recommended

For each model, output the data at the finest time scale and the finest temporal scale the model computes should be provided. Having finer resolution up-front gives us more flexibility in the types of analysis we can do. It also provides the greatest amount of information. We expect that most comparisons will be made at monthly time step. For Inverse models TRANSCOM regions will be the nominal spatial scale. Most forward ecosystem models probably can produce results at a 1° spatial resolution and monthly time step. This is the likely temporal and spatial resolution we will use for comparison to inverse models and can be aggregated to TRANSCOM regions. Finer temporal intervals are possible with most models and finer spatial scales of most data sets will be available for comparisons. Most of the towers estimate half-hour fluxes, so the models that can should save half-hour averages for at least the tower pixels. For a list of sites, see the AmeriFlux web site http://public.ornl.gov/ameriflux/. Some observations are annual, such as crop biomass/yield or forest wood increment, so the models will output "snapshot" or instantaneous values at a particular time consistent with the observation time. The participant should convert all sub-daily model output to Greenwich Mean Time (GMT).

The best file format for providing spatial model output is CF compliant netCDF. An example will be provided on the MAST-DC ftp site. Instructions for producing such a file from ASCII files and appropriate meta-data will be developed if necessary. These instructions provide additional guidance on what information will be required for model output. Software tools for most computer platforms for producing and viewing netCDF files is available. See http://www.unidata.ucar.edu/software/netcdf/ for general information and http://cf-pcmdi.llnl.gov/ for a complete description the CF convention for netCDF files. As a concrete example, please see ftp://ftp.cmdl.noaa.gov/ccg/co2/carbontracker/regions.nc. This CF-1.0 and COARDS-compliant netCDF file not only defines the grid but also gives Transcom 22 region definitions on that grid. We will also accept netCDF files with the ALMA convention (see http://web.lmd.jussieu.fr/~polcher/ALMA/dataex_main.html). If a data or modeling team has difficulty in translating files to netCDF contact us and we'll do what we can with limited MAST-DC resources to help. Do not assume that if your files are not in netCDF format that you are eliminated from participating.

Measurements

There are many types of measurements that can be compared against model output. Examples include MODIS-NPP, MODIS-LAI, EMDI-NPP, crop yield, forest inventory based biomass and growth, NEE at AmeriFlux locations, soil respiration measurements, etc. Investigators compiling and maintaining these observations and measurements have be invited to participate in the project planning, execution, and publication of findings (see Table 3). Spatial data should be prepared in the same fashion as model output described in the previous section.

5. Intercomparison Methods and Analysis

Model-data comparison

Comparisons between model spatial patterns and data spatial patterns are often examined visually. Difference plots can also be useful. One very useful diagrammatic form, termed "Taylor diagram", can convey information about the pattern similarity between a model and observations. This same type of diagram can be used to illustrate the relative accuracy among a number of model variables or different observational data sets. One additional advantage of the "Taylor diagram" is that there is no restriction placed on the time or space domain considered. A good description of the Taylor diagram can be found at
http://www.ipsl.jussieu.fr/~jmesce/Taylor_diagram/taylor_diagram_definition.html

Variograms, EOFs, Getis statistic of spatial autocorrelation (Wulder and Boots 1998, Int. J. Remote Sens. 19, 2223–2231), etc., for examining the spatial scale of synoptic patterns will also be considered.

Other model data comparisons have used a wide variety of point-by-point comparison statistics. There is a large literature in other disciplines on comparing models and data. Some references are: Janssen and Heuberger 1995, Ecol Modelling 83:55-66; Legates and McCabe 1999, Water Resources Research 35:233-241; Wilmott 1982 BAMS 63:1309-1313; Fox 1981 BAMS 62:599-609.

Model-Model comparisons

While the emphasis will be on model data comparisons, there is much to be learned by comparing models directly. In addition to comparing the performance of each model to flux measurement or observations, there is an analysis of how models compare with one another. This will be of particular interest for comparing forward ecosystem models with inverse models. The output can be aggregated up into biome types or regions (similar to CarbonTracker) and then the compared. Do they agree? Is there more agreement between models for particular biome types? etc. This would also allow for the inclusion of regional scale models into the study (i.e., included in the comparison of certain sub-regions). The variograms would also be a useful in comparing the spatial variability included in the different models.

Model and Data Participants

Table 2 and Figure 1 list the potential models participating as of 04/11/2008.

Table 2. Ecosystem or forward modeling participants in the NACP Regional/Continental model data comparison project as of May 8, 2008.

Model	Contact	Region
SiBCASA	Denning/Shaeffer	NA, MCI
TECOR	Luo/Zhou	NA
VEGAS	Zeng	NA, MCI
CLM-CN	Thornton	NA
CLM-CASA’	Hoffman/Thornton	NA
Biome-BGC	Turner/Law	Oregon-California
ISAM	Jain	NA
EPIC	Izaurralde	MCI
ED	Moorcroft, Dietze	New England
CASA	Potter	NA, MCI
TOPS/Biome-BGC	Nemani	NA, MCI
ECOSYS	Grant	MCI, NA
CN-CLASS	Altaf Arain	NA
GTEC/LoTEC	Post,King, Gu, Riciutto	MCI, NA
CENTURY+MODIS	Ogle	MCI
ORCHIDEE	Piao, Ciais, Viovy	NA
EDCM	Liu/Bliss	NA, MCI
CBM-CFS3	Kurz	Canada
ISOLSM	Riley	ARM-CART/MCI
CLASS-CTEM	Peng	NA
DAYCENT	Parton	MCI, Continental US
MC1	Nielson	NA
SSiB2	Sahoo	NA
TEM	McGuire	NA
TEM-TCM	Tian	NA

Invitations to NACP investigators including modelers and data providers to join in this project were sent in early February by the coordinators and posted on the NACP and MAST-DC websites. All investigators, even those not formally part of the NACP, are welcome to participate and do not have to wait for an invitation to contact the project coordinators to be included in the planning and execution of the project.

Figure 1. Potential participants in the NACP Regional/Continental model data comparison project – inverse models.

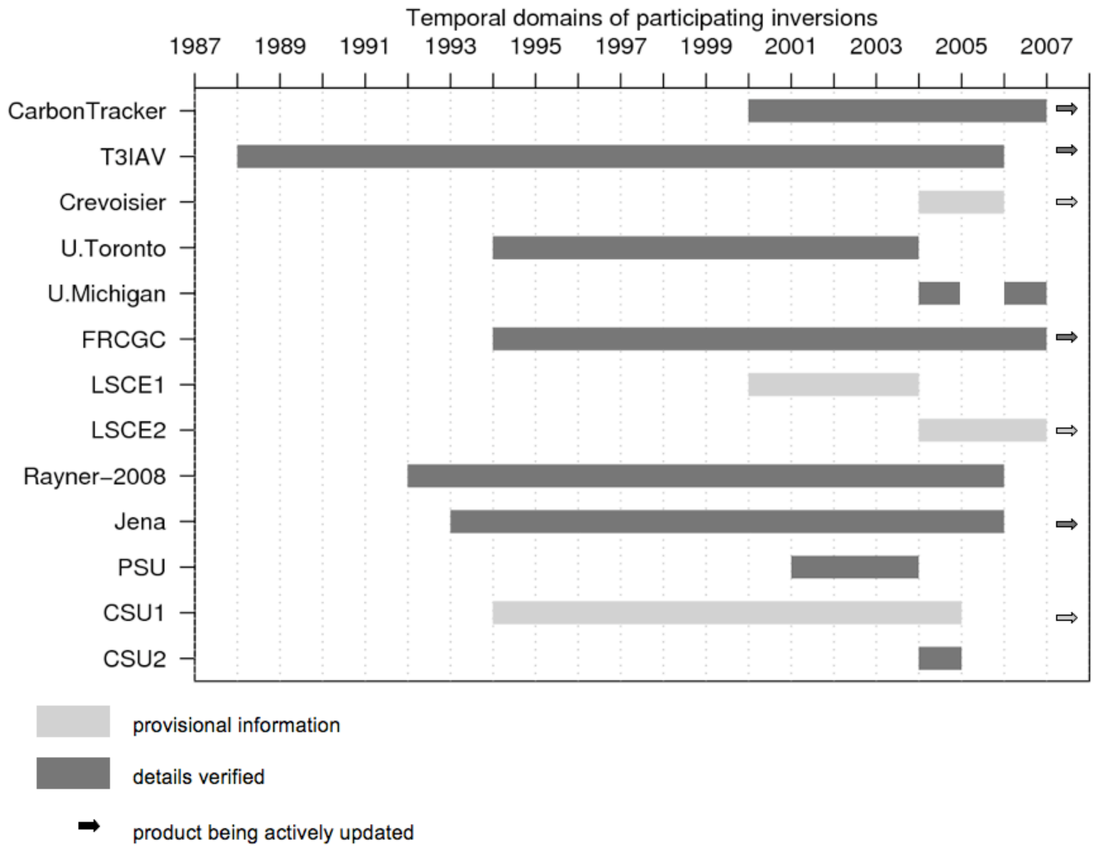

Table 3a. Spatial data providers participating in the NACP Regional/Continental model data comparison project as of May 8, 2008.

Dataset	**Contact**	**Region**
FIA (biomass, growth)	Heath	MCI, US
MODIS GPP, NPP, LAI	Running, Morisette, Zhao	NA
NEE (CO_2, H_2O)	Participating AmeriFlux, Fluxnet investigators	NA
NASS crop yield	Ogle	US
Soil C	West	US
NDVI	Zeng	NA
MODIS albedo	Schaaf	NA
PAR record	Liang	NA

In addition to these datasets that will be used directly in comparison with model results, there are investigators working on additional data sets that may be useful in resolving issues in forward-inversion model comparisons. These include temporal and spatially resolved CO_2 emissions, river transport of DIC and DOC, transport and fate of agricultural crops, forest disturbance, and fire emissions.

Table 3b. Additional datasets for Regional/Continental synthesis.

River DOC, DOC	Raymond	MCI, US Continental
CO_2 Emissions	Gurney, Marland	US Continental, NA
Fire Emissions	Randerson	US
Agricutural transport	West	US Continental
Forest disturbance	Masek, Goward	US Continental

Schedule

Synthesis Protocol send to participants – February 2008 (completed)

Prospectus to NACP for Funding – February 2008 (completed, $75K approved by DOE)

Observational data sent to MAST-DC – May 2008

Model results sent to MAST-DC – May 2008

Analysis of model-data comparison – June through September 2008

Regional MDC Workshop – September 2008

Draft papers for publication – September through December 2008

Present results at NACP All-Scientist meeting – February 2009

Finalize papers for publication, submit – March 2009

References

Fox 1981 BAMS 62:599-609.

Janssen and Heuberger 1995, Ecol Modelling 83:55-66.

Legates and McCabe 1999, Water Resources Research 35:233-241.

Peters, W., A.R. Jacobson, C. Sweeney, A.E. Andrews, T.J. Conway K. Masarie, J.B. Miller, L.M. Bruhwiler, G. Petron, A.I. Hirsch, D.E.J. Worthy, F.R. van der Werf, J.T. Randerson, O.O. Wennberg, M.C. Krol, and P.P. Tans. 2007 PNAS 104:18925-18930.

Potter, C., S. Klooster, A. Huete, V. Genovese 2007. Terrestrial carbon sinks for the United States predicted from MODIS satellite data and ecosystem modeling. Earth Interations 11:13-21.

Wilmott 1982 BAMS 63:1309-1313.

Wulder, M., Boots, B., 1998. Local spatial autocorrelation characteristics of remotely sensed imagery assessed with the Getis statistic. Int. J. Remote Sens. 19, 2223–2231.

Appendix 1. Variables describing the carbon cycle at the surface.

(from the ALMA netCDF convention - http://web.lmd.jussieu.fr/~polcher/ALMA/descriptions.html#carbon)

GPP:

The total Net assimilation of carbon by the vegetation. This variable is given by the mean gross assimilation minus the dark respiration. It should be averaged over all vegetation types within a grid cell.

NPP:

Net primary production must be equal to GPP - AutoResp. Averaged over all vegetation types within a grid cell. l

NEE:

Net Ecosystem Exchange sums all carbon fluxes exchanged between the surface and the atmosphere. It represents at least AutoResp + HeteroResp - GPP. But outgoing NEE can also contain others fluxes like for instance CO_2 from fires. By convention it is negative when the outgoing flux (GPP) is greater than the incoming flux. This variable should also be averaged over all vegetation types within a grid cell

AutoResp:

The total autotrophic respiration includes maintenance respiration and growth respiration. It can include others terms for very advanced models which for instance calculate respiration from ion uptake. This variable must be positive in the traditional sign convention and be a deficit for the surface.

HeteroResp:

The total flux from decomposition of organic matter. This include fluxes from soil and litter. It must also be positive in the traditional sign convention and be a deficit for the surface.

TotSoilCarb:

Total soil and litter carbon content integrated over the entire soil profile. This term must contain only carbon from dead material (not from roots for instance).

TotLivBiom:

Total carbon content of living biomass. This include above and below ground biomass (e.g leaves, fine roots, coarse roots, heathwood, sapwood etc...). This variable needs to be averaged over all vegetation types within a grid cell.