Model-Data Fusion:
Observing, analyzing, and predicting our environment

Pep Canadell
GCP International Project Office
CSIRO Earth Observation Centre
Canberra, Australia

Towards a capability to analyze the natural-human system

Detection: quantifying past and current pools and fluxes

Attribution: partitioning of pools and fluxes into mechanisms that control them

Prediction: projecting the response of the natural-human system to hypothesized environmental changes and human management

General purpose of such a system

- To increase knowledge of the cycles of carbon and related entities (water, energy, nutrients) in the terrestrial biosphere, in support of the management and governance of the interactions of the biosphere with human activities which influence:
  - Global climate
  - Water resources
  - The balances of landscape nutrients and contaminants, biodiversity, and the enhanced GHGs.

- Important fluxes are:
  - Land-air exchanges (atmospheric sources and sinks)
  - Exchange with rivers and groundwater
  - Exchanges between various terrestrial pools such as biomass and soil

- Determine the main processes influence these fluxes, especially those amenable to human management.

- No single model or set of observations can supply this amount of information – hence the need for a synthesis approach.

Classic systematic Earth Observations

- Atmosphere: weather networks
- Oceans: hydrography and oceanography
- Land: geomorphology, biogeography

What has changed in recent times...

- Global Change Issues rising:
  - Greenhouse-induced climate change
  - Water shortages and imbalances
  - Land degradation
  - Soil erosion
  - Loss of biodiversity

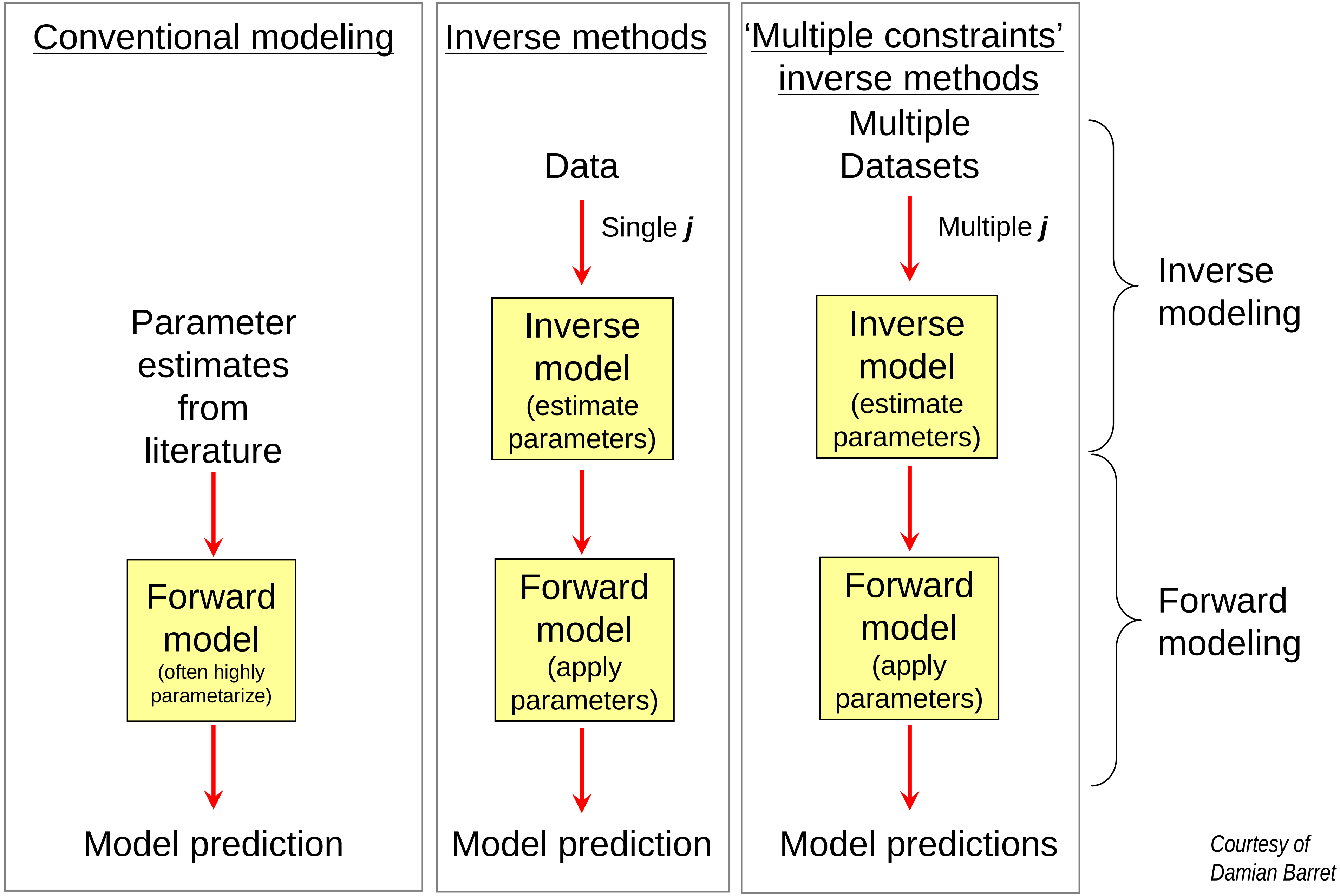
- Technological advances in sensors, satellite systems and data handling

- New earth system acknowledge allows us to test new hypothesis of the interconnectedness of its components including humans activities.

Motivations of the model-data synthesis approaches include:

1. Model testing and data quality control (data uncertainty)
2. Interpolation of spatially and temporally sparse observations
3. Inference from available observations of quantities which are not directly observables (C stores and fluxes over large areas)
4. Forecasting (prediction forward in time on the basis of past and current observations).

What is different from the more traditional modeling?



Courtesy of Damian Barret

Classic modeling on Net Ecosystem Productivity

Light use efficiency

Carbon allocation

Q_{10} - Temperature sensitivity to soil Respiration

Gross Primary Production

↓ $R_{aut.}$

Net Primary Productivity

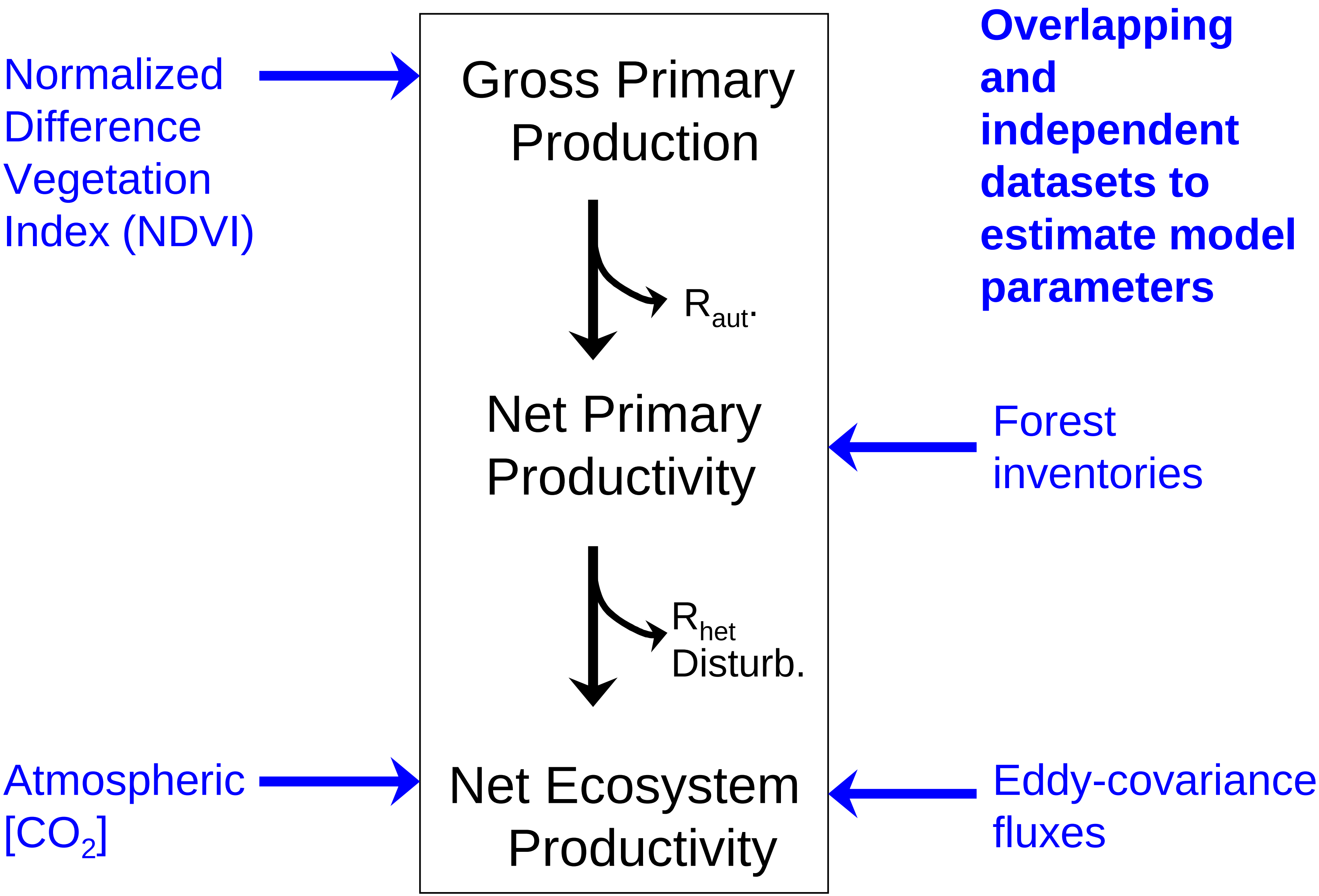
↓ R_{het} Disturb.

Net Ecosystem Productivity

Parameter optimization to fit the model to test data

← Eddy flux data to validate results

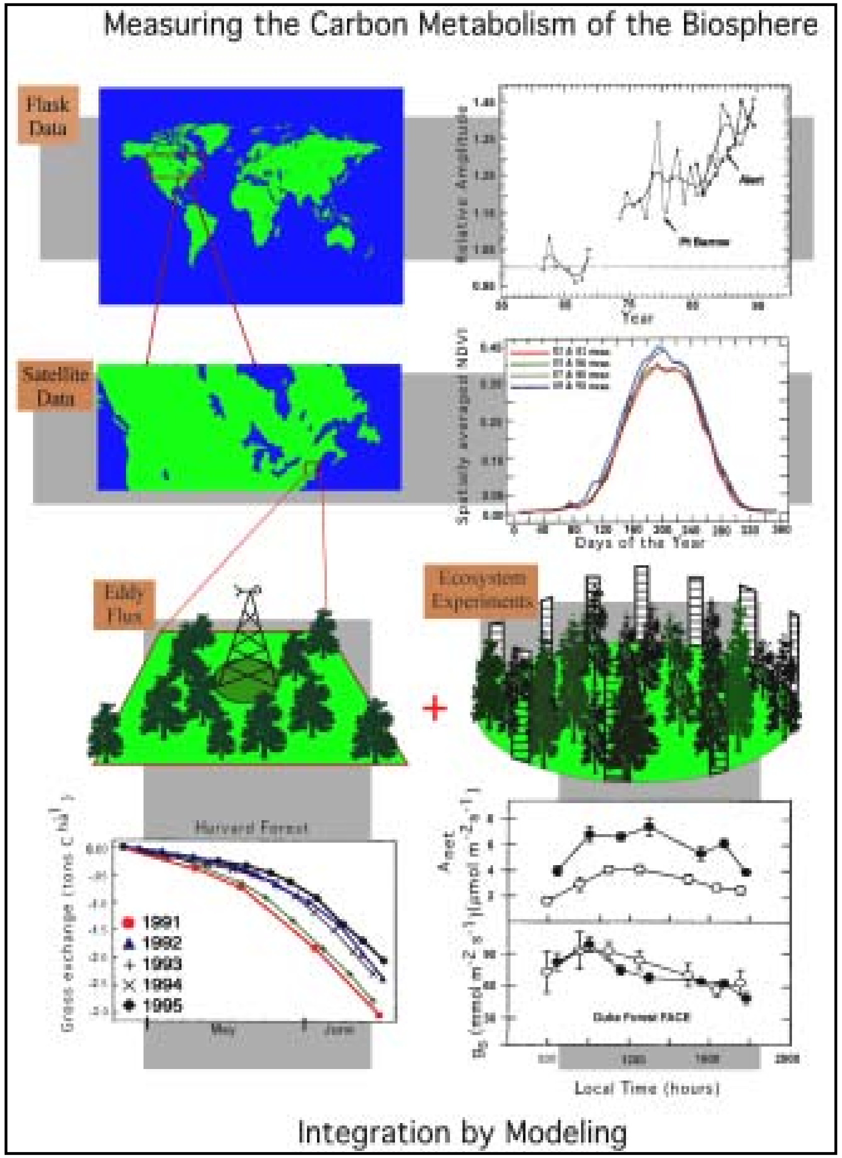
Multiple constraints on Net Ecosystem Productivity



Requirements for Model-Data Synthesis

- A **model** or sets of models (starting with terrestrial biogeochemical models).
- **Data** or information:
  - Observations – quantities
  - Process level information
  - Prior estimates for model quantities
  - ***Uncertainty specifications*** on the data (data uncertainties are as important as data values themselves in determining the final model outcome)
- **Synthesis** approach:
  - Model properties to be adjusted
  - Measure of departure between data and model (cost function)
  - Search strategy for finding the optimum values.

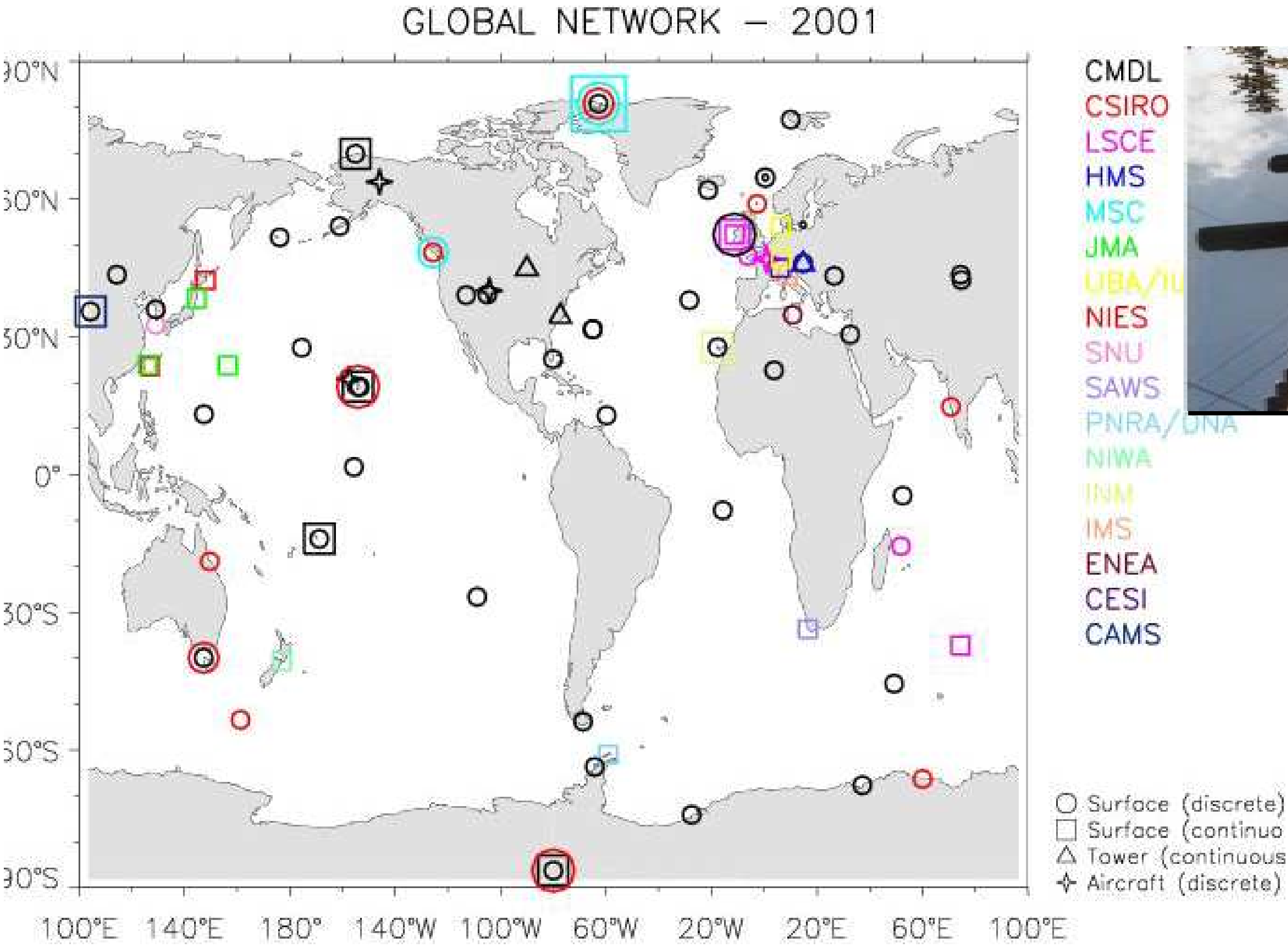
Data components for terrestrial carbon M-D fusion



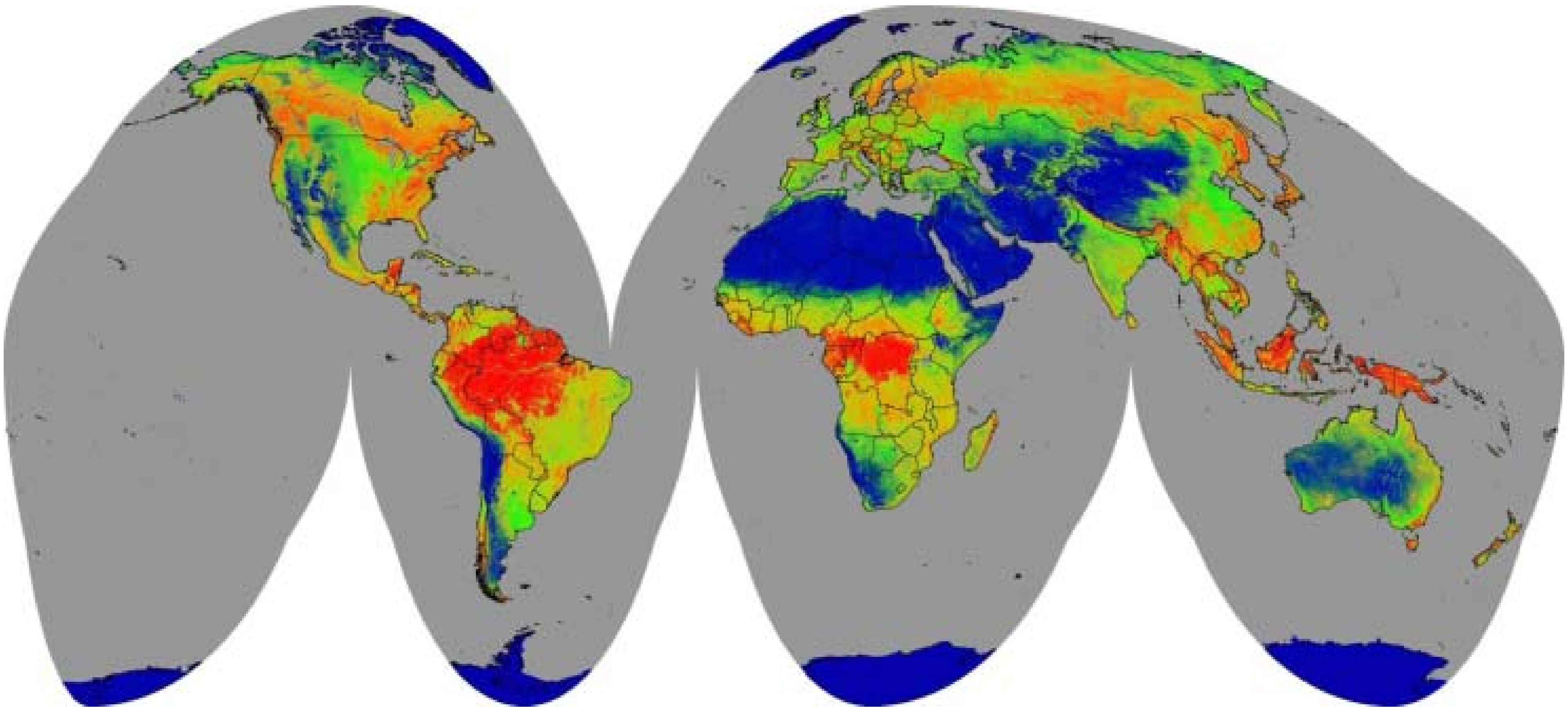
Canadell et al. 2000

- **Atmospheric composition** and isotopes (CO_2, CH_4, O_2/N, ^{13}C, ^{14}C)
- **Remote sensing** (NDVI, phenology, GPP, fire scars, fire area)
- **Forest inventory** and other *insitu* data
- **Process data**, eg. Flux tower and manipulative experiments (FACE, warming)
- **Soil C** content, depth, etc.
- **Land use and disturbance** history

Global air flask and tall tower network



500m MODIS Vegetation Continuous Fields – 2001



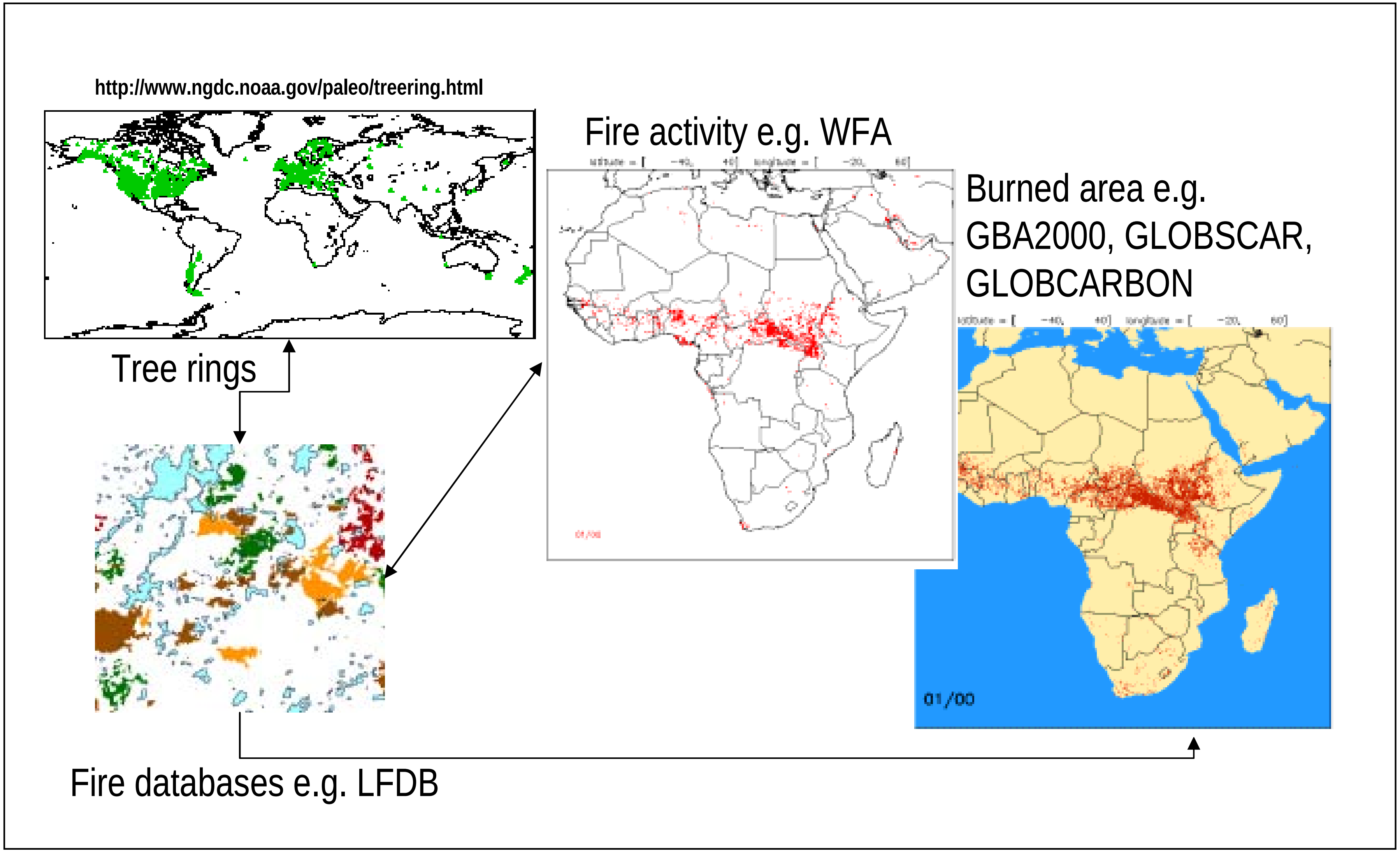
Bare ground

Percent cover 0% Grass/shrubs/moss 100%

Trees

Hansen et al. 2003, in press

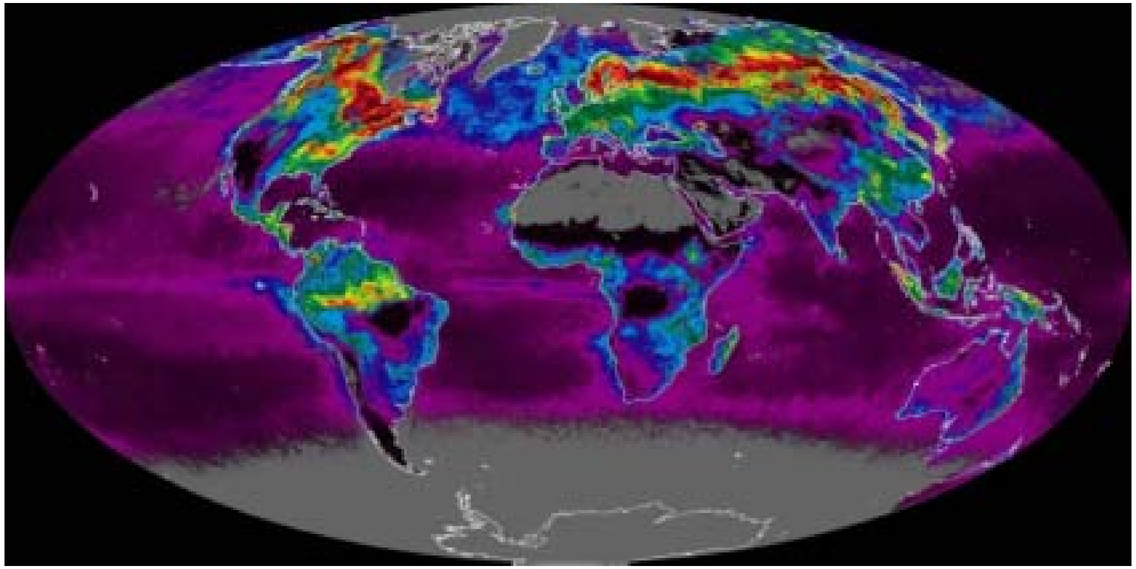
Disturbance/Fire - data sources



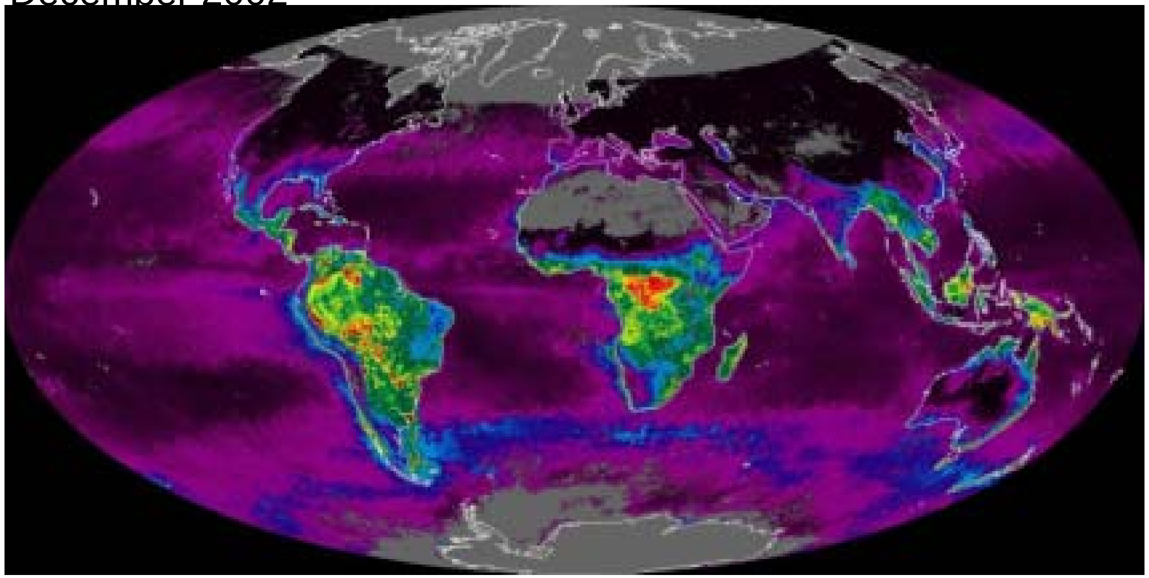
IGBP – ESA. Courtesy of Stephen Plummer, 2003

Global NPP derived from MODIS

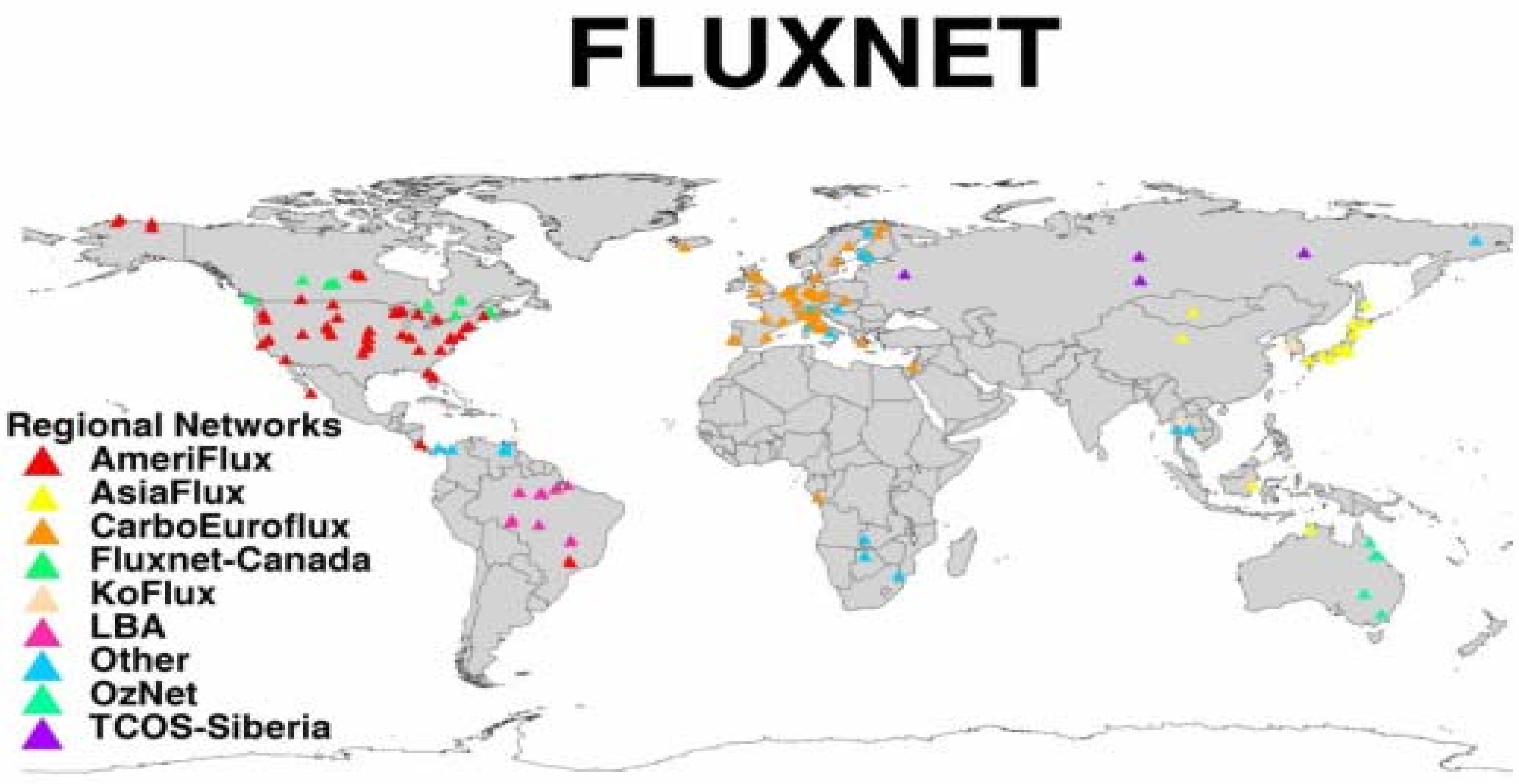
June 2002



December 2002



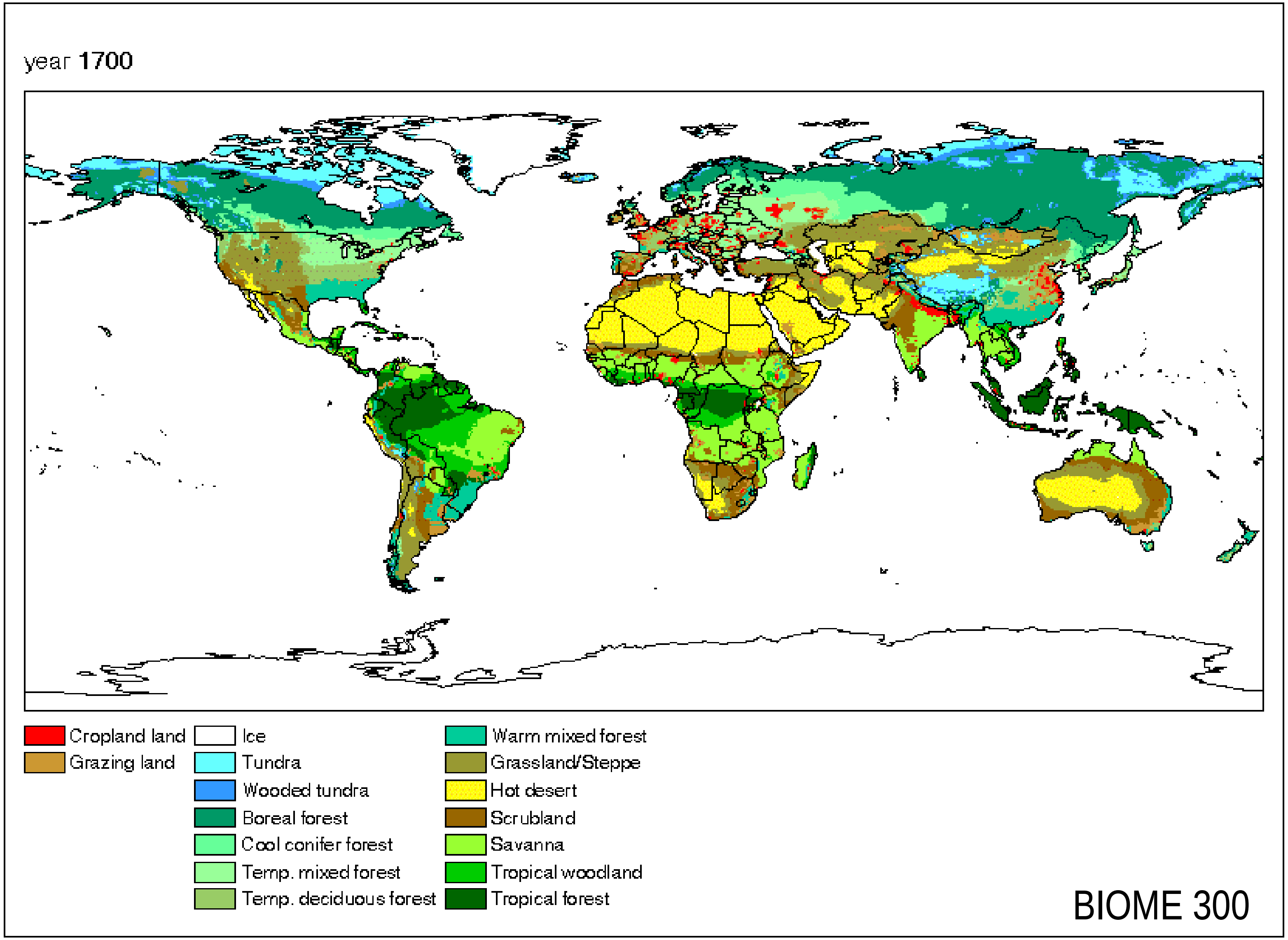
FLUXNET – eddy covariance flux measurement sites



Loblolly Pine Forest, NC, USA

Temperate Grassland (CO_2 x Nutrients x Biodiversity), Wisconsin, USA

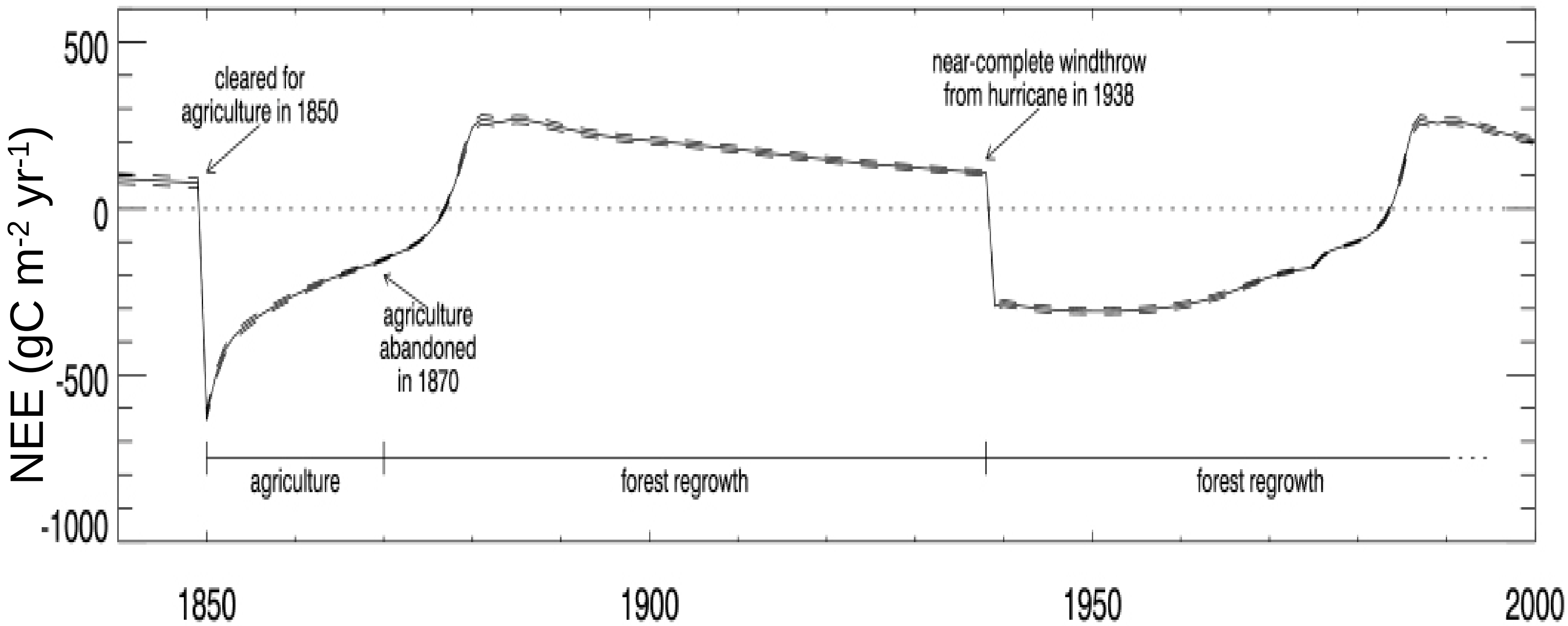
Historical Land Use Change



RIJKSINSTITUUT VOOR VOLKSGEZONDHEID EN MILIEU
NATIONAL INSTITUTE OF PUBLIC HEALTH AND THE ENVIRONMENT

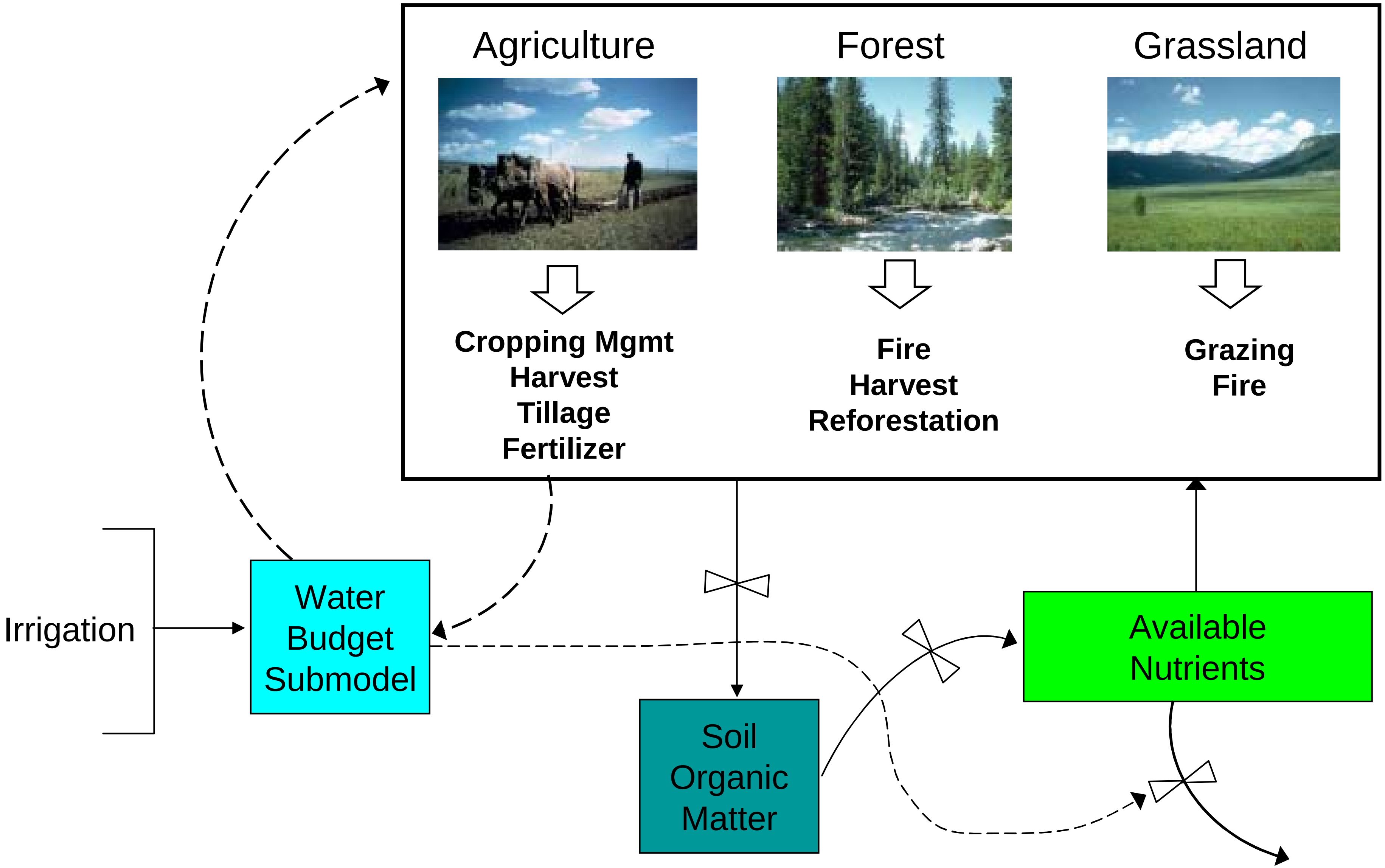
System memory: Effects on NEE of natural and anthropogenic disturbances:

Simulation based on recorded disturbance history at Harvard Forest



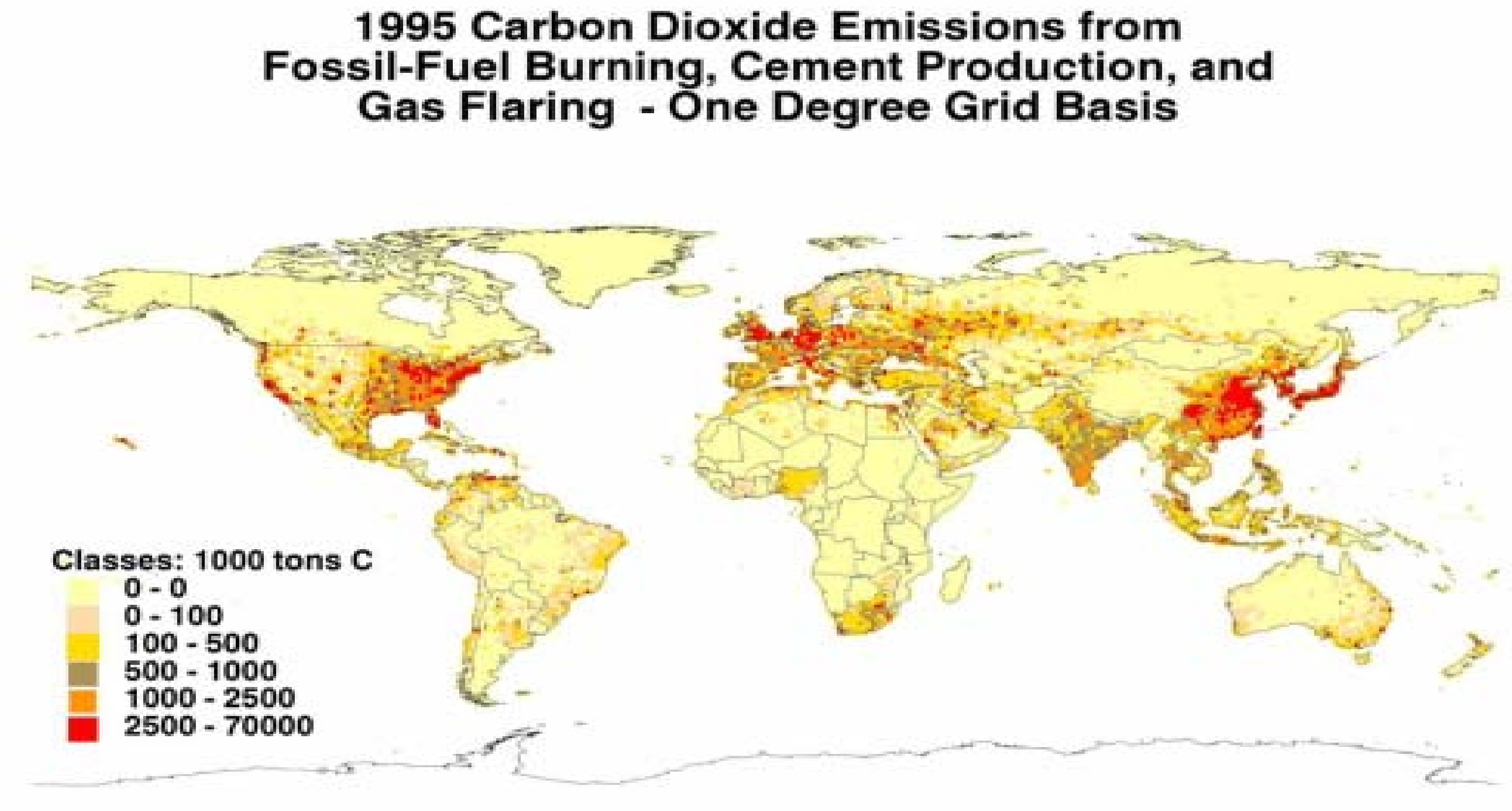
Courtesy of Peter Thornton 2003

Management Practices and Disturbances

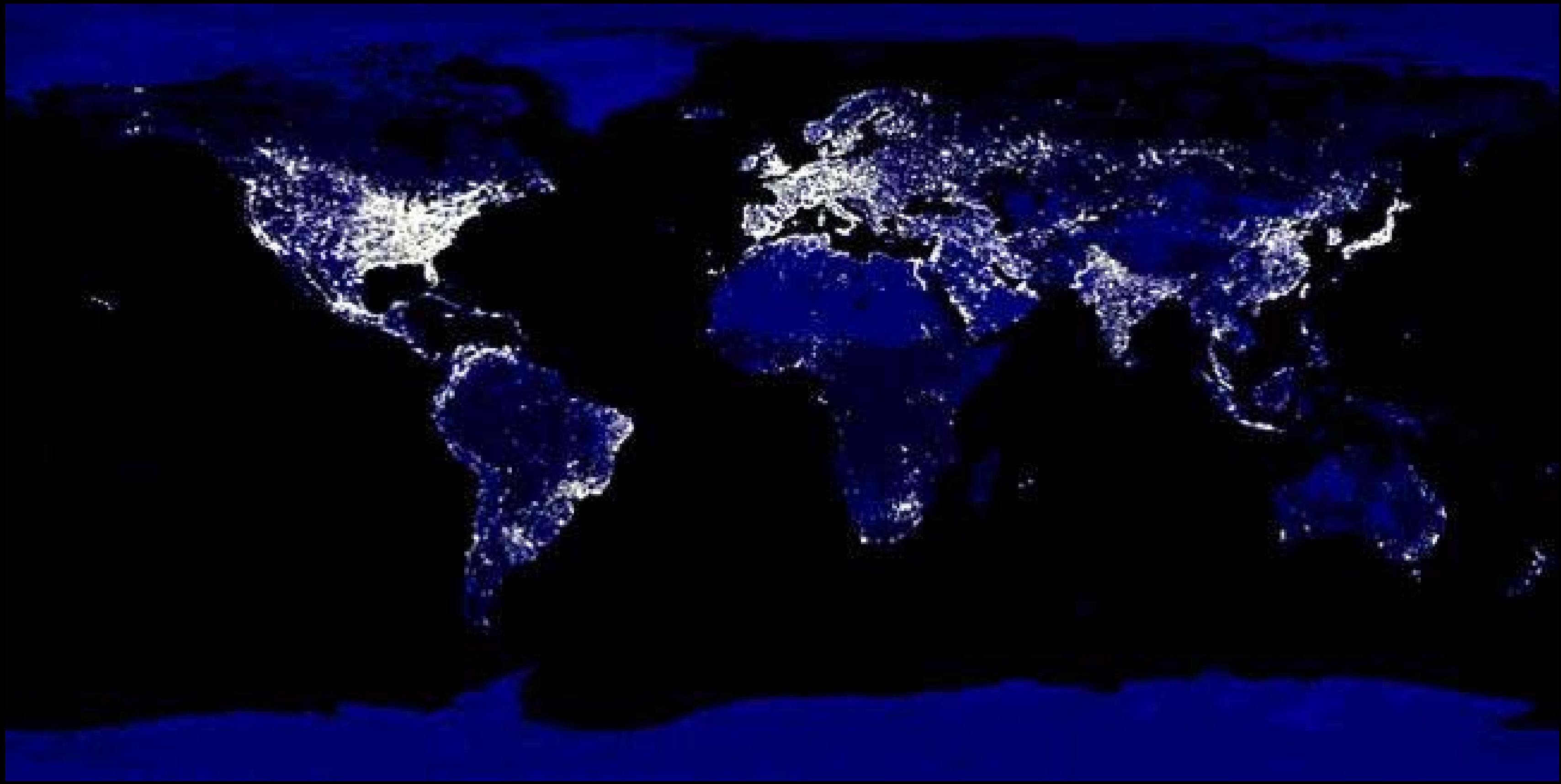


Courtesy of Dennis Ojima, 2003

CO_2 emissions from FF burning, Cement, Gas Flaring



City Lights related to regional energy consumption and income.

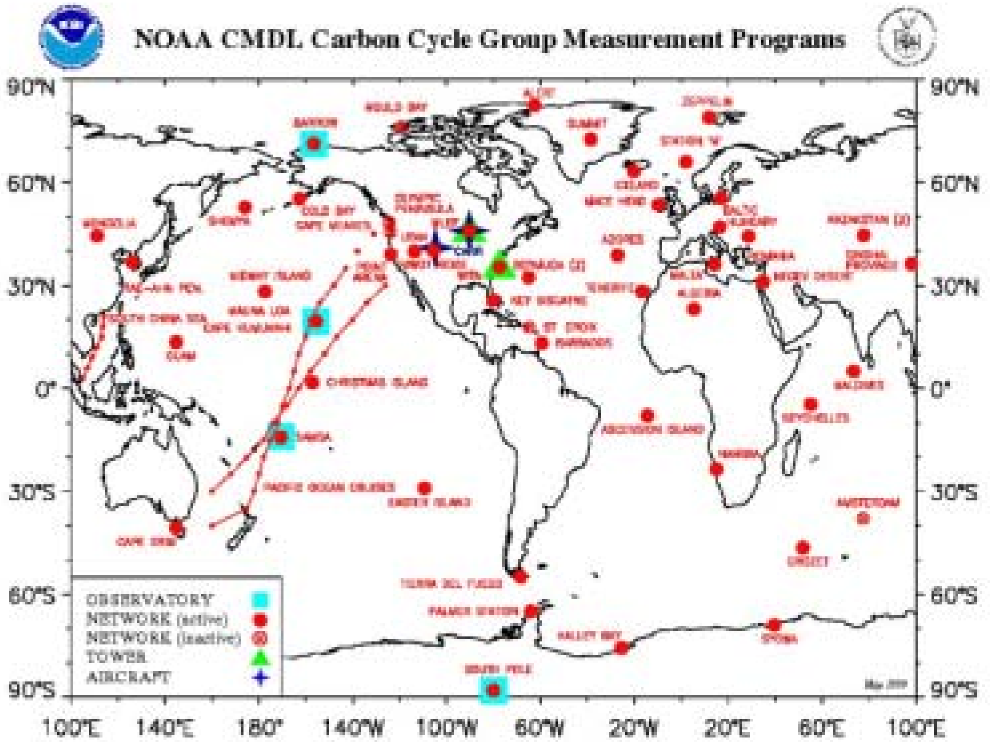


Examples of Model-Data Synthesis

- Parameter estimation
- **Inversion methods**
- **Multiple constraints**
- Data assimilation

Inverting CO_2 into into CO_2 Fluxes

Atmospheric Transport Model

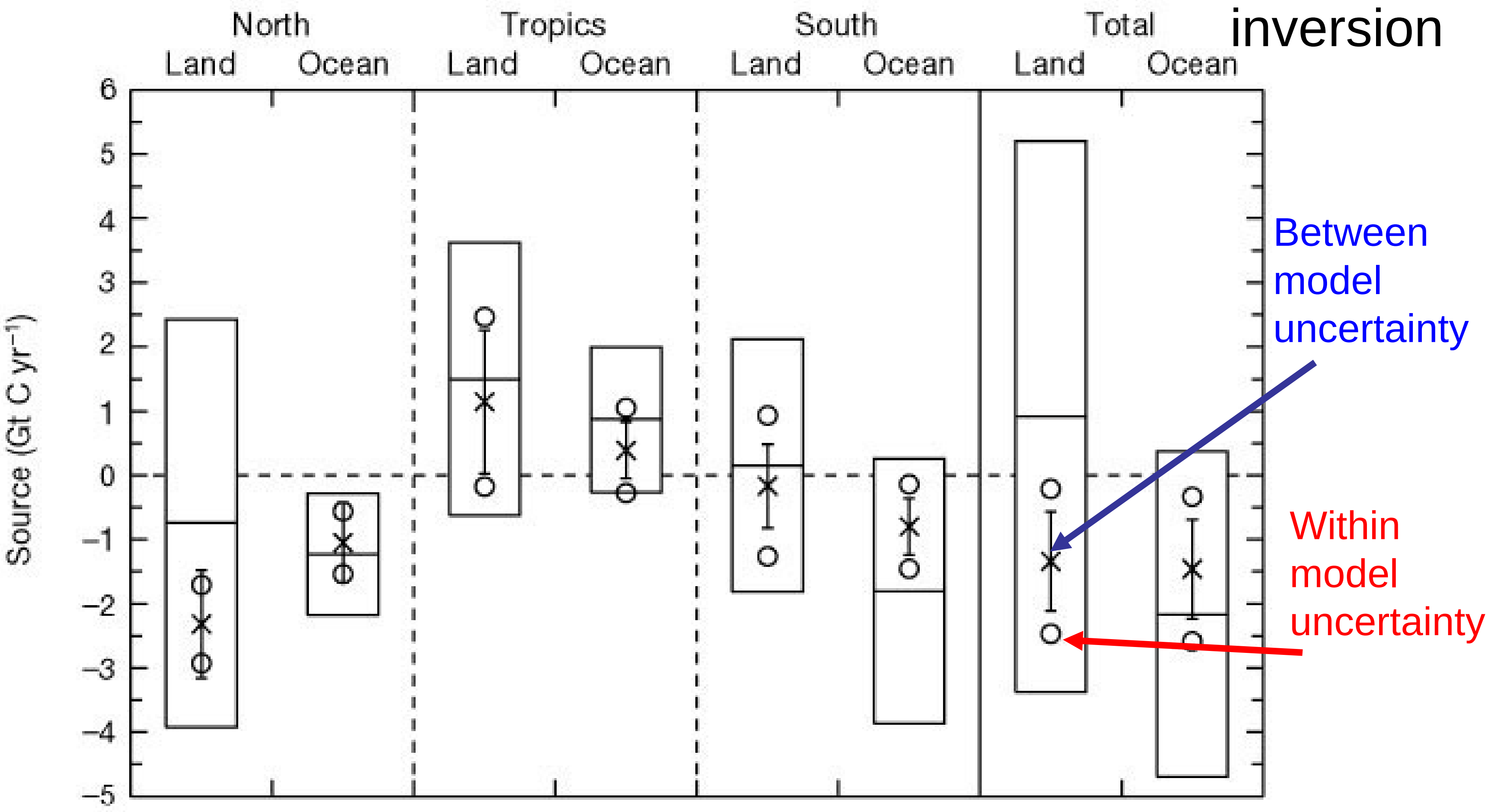


NOAA-CMDL 1999

Inverse Model estimates of the latitudinal carbon sinks

They still have significant uncertainties,
and are not constrained by ecophysiological understanding

Bayesian synthesis inversion



Gurney et al, 2002

Inverse analyses of Eddy-Flux and FACE datasets

Courtesy of Yiqi Lou, 2003

Flux-based inversion -- Method

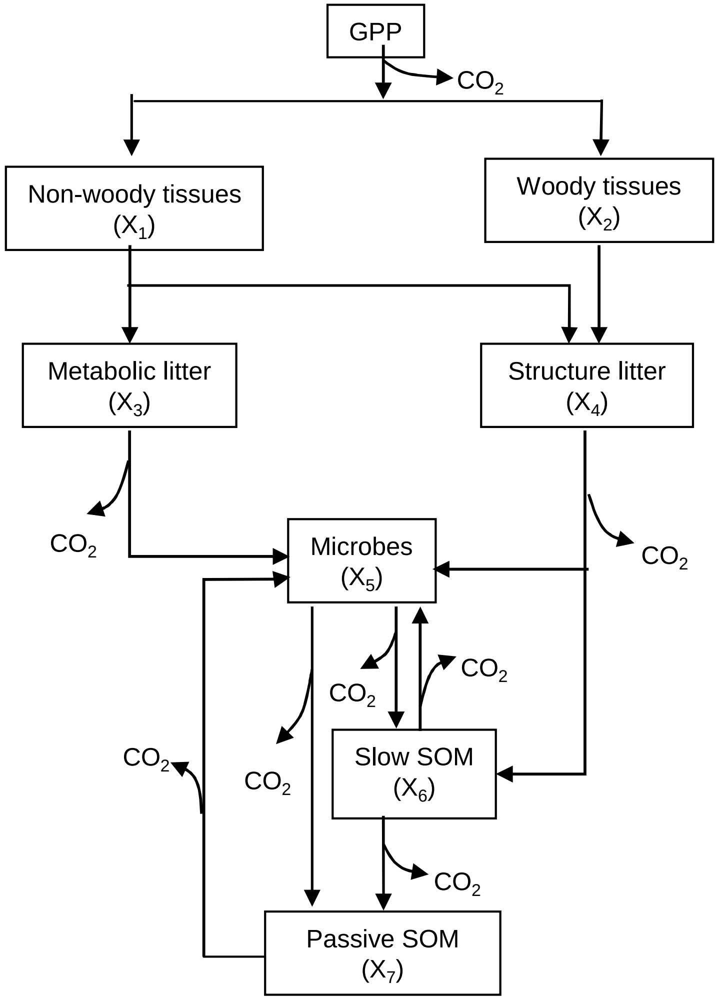
Optimization techniques for searching parameter values

1. Steepest descent
2. Newton
3. Gauss-Newton
4. Marquardt
5. Simulated annealing

Data source: Niwot Ridge Flux site

Yiqi Lou, 2003

Pool-based inversion - Methods



1. Linking NEE with pool changes

$$NEE = \sum_{i=1}^{m} \frac{dX_i}{dt}$$

2. Matrix to describe changes in C pools

$$\frac{dX}{dt} = \tau AX + Bu$$

3. Evaluation criteria

$$J(A) = \sum_{j=1}^{m} v_j \left[\sum_{i=1}^{n_j} (Q^j(A)(t_i) - Q_0^j(t_j))^2 \right]$$

4. Search method

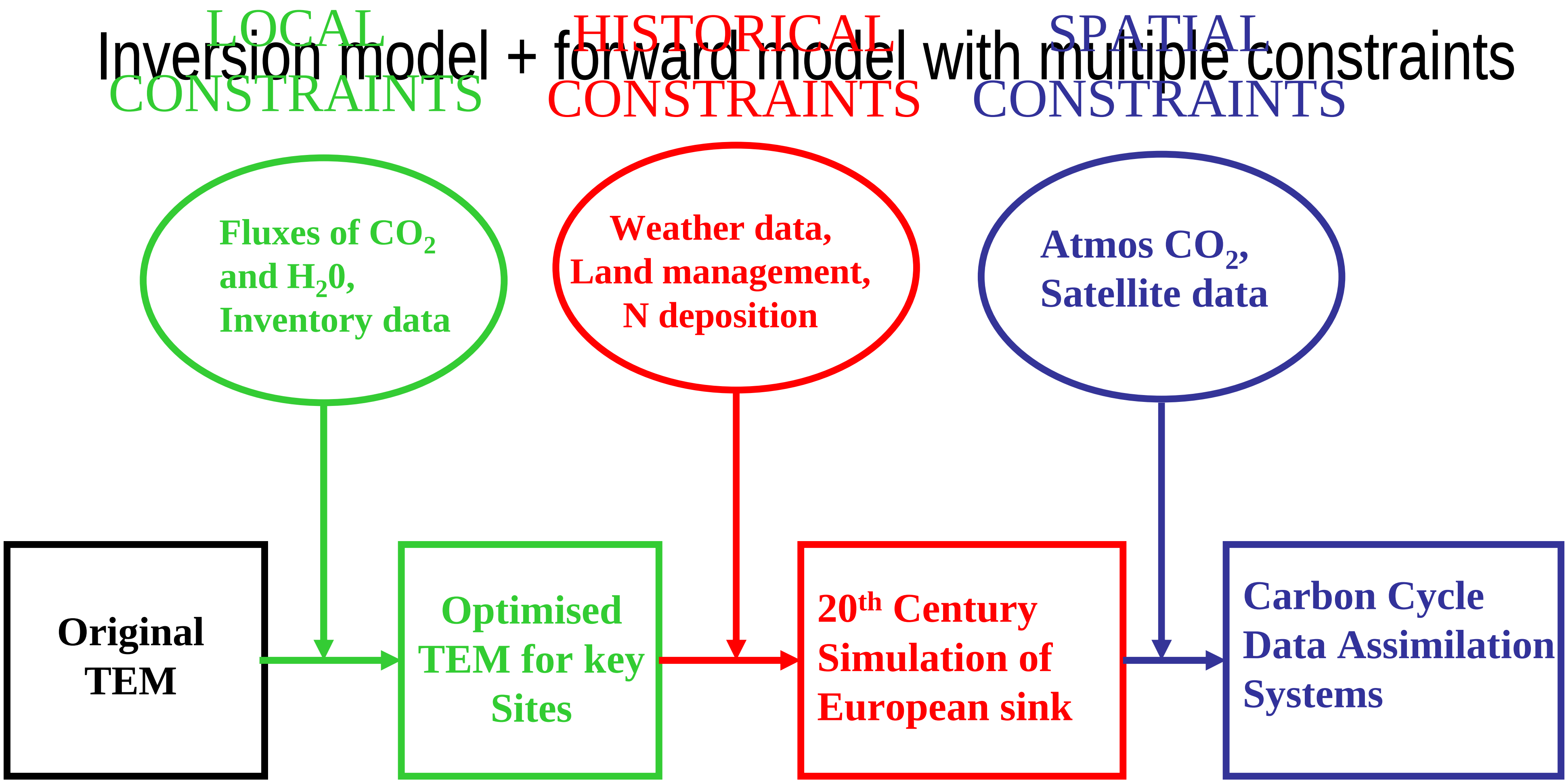
a) Levenburg-Marquardt minimization method with a quasi-Monte Carlo algorithm to search *A*

b) Stochastic method to evaluate information content

Yiqi Lou, 2003

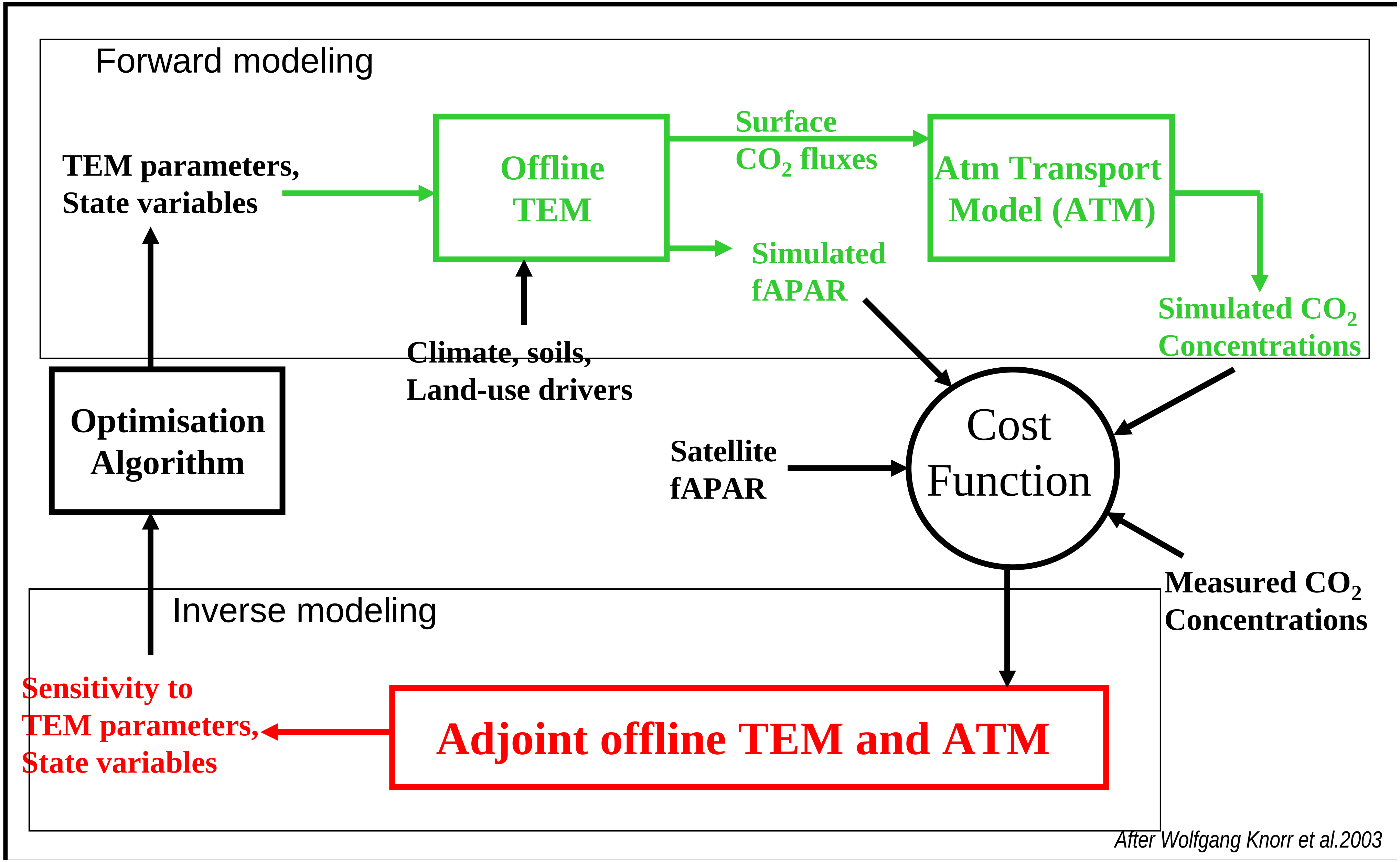
Carbon **A**ssimilation and **M**odelling of the **E**uropean **L**and **S**urface: CAMEL

Inversion model + forward model with multiple constraints



Peter Cox, 2003

Coupling of the models



After Wolfgang Knorr et al.2003

The Case for Data-Model Fusion

- Mechanistic Models are needed to separate contributions to the land carbon sink (e.g. as required by KP: CO_2 & N fertilization, regrowth).
- Large-scale data constraints (from CO_2 and remote-sensing) are required to provide best estimates and error bars at regional and national scales.
- Data-Model Fusion = ecophysiological constraints from forward modelling + large-scale CO_2 constraints from inverse modelling

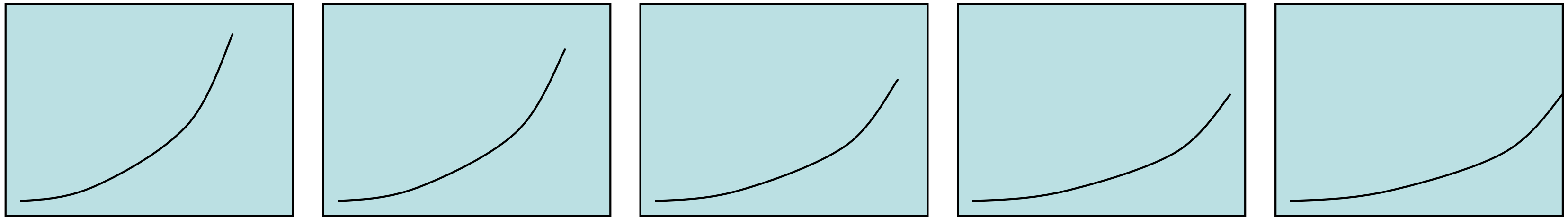
Examples of Model-Data Synthesis

- Parameter estimation
- Inversion methods
- Multiple constraints
- **Data assimilation**

Data assimilation in atmospheric and ocean circulation models

Sequential strategy, involving adjustment to the model state variables as the model is integrated forward in time. This is the model applied routinely in weather forecasting.

(e.g., techniques includes kalman filter)



Conclusions

- **Inverse methods + forward modeling** is a critical approach in model-data fusion
- **Multiple data streams** (providing overlapping information on processes but independently measured).
- **Data uncertainty** is as important value as the data themselves.

Global Carbon Project

www.GlobalCarbonProject.org