Carbon Data Assimilation

Ian G. Enting

MASCOS

The University of Melbourne

Summary

- Context of carbon data assimilation
  - modelling, applications, data
- Inversion formalisms and techniques
  - beyond flux inversions: adjoints
- Radiocarbon revisited
  - information about gross terrestrial carbon fluxes

Why Carbon?

Calibration for earth system simulator

- ACCESS: CABLE + CASA(CNP?) + LPJ
- Application to projection and detection of feedbacks between physical climate system and the carbon cycle.

Nowcasting

- For natural resource management

Links to water cycle

Interpretation

Two main inverse problems: calibration and data assimilation (deconvolution).

$$C(t) = C(0) + \int_0^t R(t - t')\, S(t')\, dt'$$

$$C(t) = C(0) + \int_0^t R(t'')\, S(t - t'')\, dt''$$

The problems of deducing model response, $R(t)$ (given ($S(.)$ and $C(.)$) and deducing forcing term $S(t)$ (given $R(.)$ and $C(.)$) are formally equivalent, but in practice differ greatly because of the different characteristics of the statistics.

Hybrid

Using observations in an assimilation mode as part of a calibration is a case that sits between the 'pure' calibration and deconvolution formalisms.

Use of satellite vegetation indices is an important case for terrestrial modelling.

This type of problem will almost always involve non-linear estimation.

Information content: e.g. CO_2

Quasi-exponential growth in emissions in 20th century:

$$C(t) \approx C_{\text{equil}} + \int_{-\infty}^{t} A \exp(\alpha t')\, R(t - t')\, dt'$$

Sequence of $C(t)$ values all characterise same projection of the response function. [Happens to be $p = \alpha$ value of Laplace Transform of $R(t)$.]

Information about longer time scales comes from ocean chemistry and natural ^{14}C.
Information about shorter time scales comes from bomb ^{14}C.

Data for Carbon Cycle Studies

- Air sampling networks interpreted by inverse modelling;
- Satellite data, for quantities such as leaf-area index and phenology
- Terrestrial biosphere models;
- Convective boundary layer measurements;
- Stand-level flux networks;
- Ecosystem experiments;
- Small cuvettes.

From Canadell et al., *Ecosystems*, 3:115, 2000.
Satellite CO_2 data is potential addition to this list.

Key characteristics of statistics

- magnitude;
- degree of correlation between components;
- temporal correlation structure;
- spatial correlation structure;
- distribution;
- mismatches in averaging;
- contribution from model representativeness error.

From Raupach et al., *Global Change Biol.*, 11: 378, 2005

Characteristics of terrestrial carbon

- very great spatial heterogeneity
- dominated by local interactions (coupled to atmosphere)
- wide range of time-scales involved

Water in the land-surface has similar characteristics.

Inversion for CO_2 fluxes

Synthesis

- Discretise, and calculate responses to specified set of surface fluxes. Estimate fluxes from best fit to data from linear combinations of responses.

Mass balance

- interpolate data to provide surface concentrations at all points as function of time, and integrate transport equations using this boundary condition – deduce fluxes from surface mass balance.

Gradient methodology

- Direct (iterative) minimisation of cost function

- Doesn't assume linearity so can do parameter estimation (and/or non-Gaussian statistics)

- Can work in large dimension spaces (but requires efficient calculation of gradients using adjoint relations)

- Adjoint relations less important if dimension of parameter space is small

Adjoint transformation

Simplify by separating parametric differentiation from integration of model ($\mathcal{L}\underline{u}(.) = \underline{f}(.)[\underline{a}]$), expressed as Green's function $\underline{u}(.)[\underline{a}] = \mathcal{G}\underline{f}(.)[\underline{a}]$

Then $\nabla_{\underline{a}}\langle \underline{w}(.)|\underline{u}(.)[\underline{a}]\rangle$ transforms as
$\nabla_{\underline{a}}\langle \underline{w}(.)|\underline{u}(.)[\underline{a}]\rangle = \nabla_{\underline{a}}\langle \underline{w}(.)|\mathcal{G}\underline{f}(.)[\underline{a}]\rangle =$
$\nabla_{\underline{a}}\langle \mathcal{G}^{\dagger}\underline{w}(.)|\underline{f}(.)[\underline{a}]\rangle = \nabla_{\underline{a}}\langle \underline{v}(.)|\underline{f}(.)[\underline{a}]\rangle$
where $\underline{v}(.) = \mathcal{G}^{\dagger}\underline{w}(.)$ defines a single function $\underline{v}(.)$ with no dependence on $\underline{a}$

Gradients for soft constraints:
$\nabla_{\underline{a}}\langle \underline{H}\underline{u} - \underline{z}|\underline{H}\underline{u} - \underline{z}\rangle = 2\nabla_{\underline{a}}\langle \underline{H}\underline{u}_0 - \underline{z}|\underline{H}\underline{u}\rangle =$
$2\nabla_{\underline{a}}\langle \underline{H}\underline{u}_0 - \underline{z}|\underline{H}\mathcal{L}\underline{f}\rangle = 2\nabla_{\underline{a}}\langle (\underline{H}\mathcal{L})^{\dagger}(\underline{H}\underline{u}_0 - \underline{z})|\underline{f}\rangle$

Adjoints as matrix transpose

For $\nabla_{\underline{a}} \langle \underline{w}(.) | \underline{u}(.)[\underline{a}] \rangle$ with $\underline{u}(.)[\underline{a}] = \mathcal{G}\underline{f}(.)[\underline{a}]$

$$\sum_{j=1}^{J} \sum_{k=1}^{K} w_k G_{kj} \frac{\partial f_j}{\partial a_p} \qquad \text{for } p = 1, P$$

$$\sum_{k=1}^{K} w_k \sum_{j=1}^{J} G_{kj} \frac{\partial f_j}{\partial a_p} \qquad \text{takes } KP + KJP \text{ operations}$$

$$\sum_{j=1}^{J} \left[\sum_{k=1}^{K} w_k G_{kj} \right] \frac{\partial f_j}{\partial a_p} \qquad \text{takes } KJ + JP \text{ operations}$$

$G_{kj} \frac{\partial f_j}{\partial a_p}$ solves TLM, $w_k G_{kj}$ solves its adjoint.

Significance of localisation

For sites $n = 1, N$ with total of NP parameters and NJ forcings

$$\sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{n=1}^{N} w_k G_{k,jn} \frac{\partial f_{jn}}{\partial a_{pn}} \qquad \text{for } p = 1, P, \; n = 1, N$$

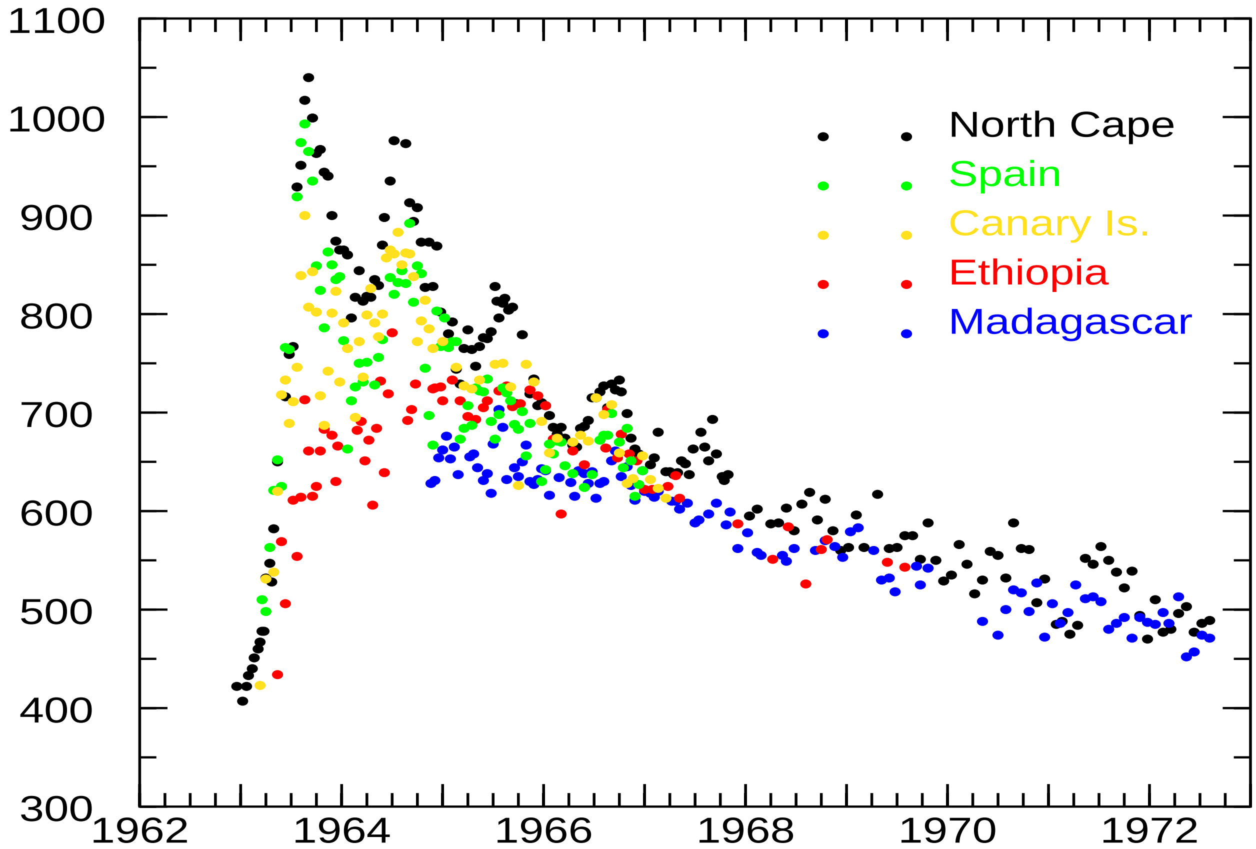
$$\sum_{k=1}^{K} w_k \sum_{j=1}^{J} G_{k,jn} \frac{\partial f_{jn}}{\partial a_{pn}} \qquad \text{takes } KPN + KJPN \text{ ops}$$

$$\sum_{j=1}^{J} \left[\sum_{k=1}^{K} w_k G_{k,jn} \right] \frac{\partial f_{jn}}{\partial a_{pn}} \qquad \text{takes } KJN + JPN \text{ ops}$$

For large N and small P:

comparative advantage of adjoint form is small.

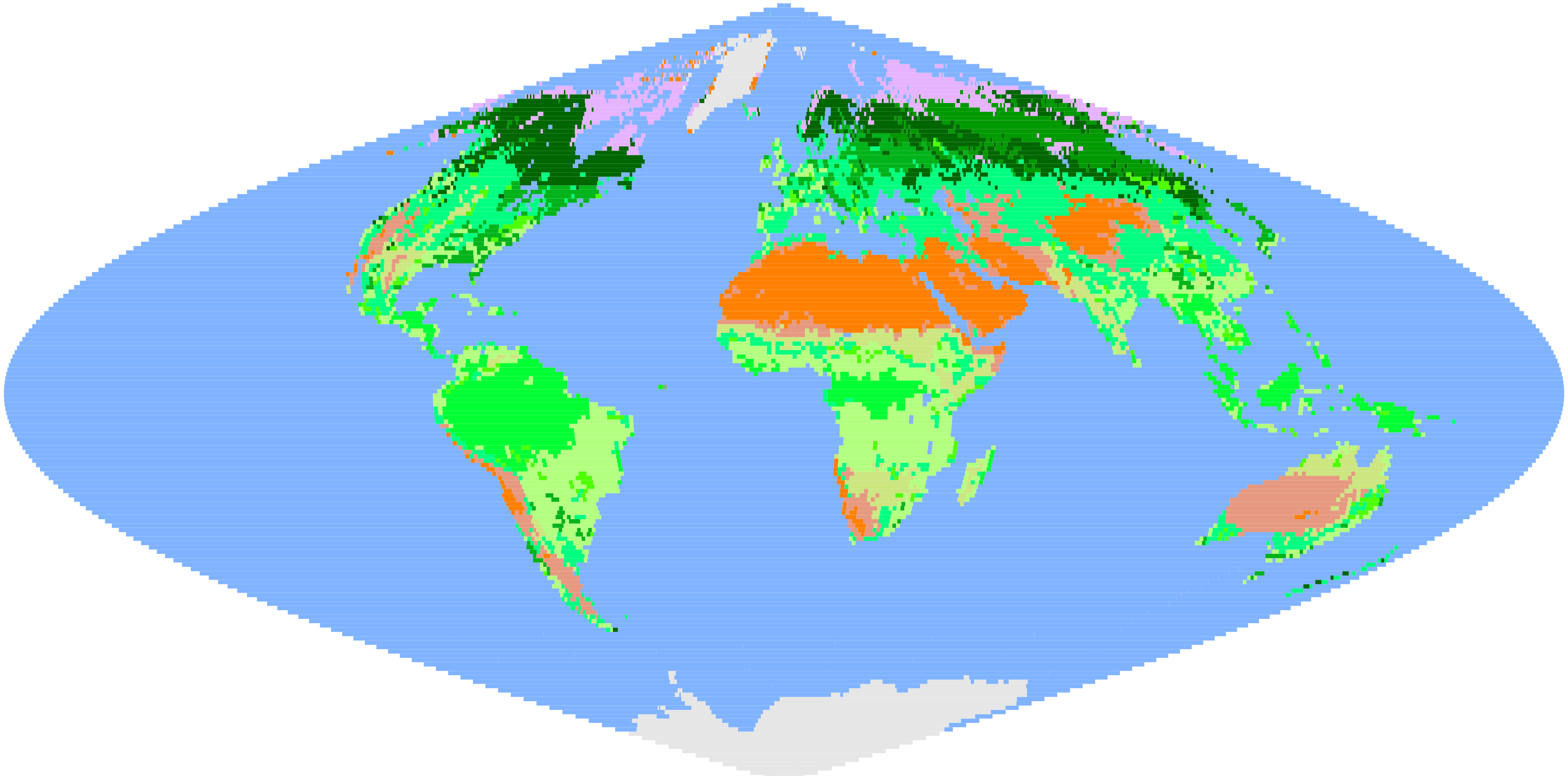
Bomb-^{14}C, with seasonal variation



Surface values, (Δ^{14}C in per mil) from Nydal and Lovseth, JGR, 88C, 3621, 1983.
CDIAC NDP057.

In times of isotopic disequilibrium, ^{14}C data give information about gross terrestrial fluxes. Randerson et al, (2002), analysed these data mainly as a constraint on seasonality of stratosphere-troposphere exchange.

CASA vegetation types



Seasonal modulation of bomb ^{14}C 'spike' gives a low-pass spatial filtering of the age distribution associated with the spatial distribution.

Concluding remarks

Analysis of the structure of inversion problems (including data assimilation) is important for

- using appropriate statistics
- identifying the actual information content
- choosing an appropriate computational formalism

Acknowledgments

The ARC funds the Center of Excellence for Mathematics and Statistics (MASCOS).

My fellowship at MASCOS is supported by CSIRO through a sponsorship agreement.

Collaborators:

Cathy Trudinger and Yingping Wang of CSIRO Marine and Atmospheric Research.

Roger Francey, Denis O'Brien, Peter Rayner, formerly of CSIRO Atmospheric Research.

Inez Fung, lecturers and participants at MSRI-NCAR workshop on Carbon Data Assimilation.

Further Information

- I. Enting: *Characterising the Temporal Variability of the Global carbon Cycle*. CSIRO Atmospheric Research, Technical paper 40.
- I. Enting: *Inverse Problems in Atmospheric Constituent Transport*. 2002, CUP.
- C. Rödenbeck: *Estimating CO_2 sources and sinks . . . .* MPI-BGC Technical Report 6.
- I. Enting: *Statistics of Multiple Constraints: Analysis of Global Change and the Carbon Cycle*. Presentation at MSRI-NCAR workshop on Carbon Data Assimilation, Berkeley, July 2006. On-line at MSRI.
- J. Randerson et al., *Temporal and spatial variability of $\Delta^{14}CO_2$:. . .*, Global Biogeochem Cycles: 16, 2002.