Data Assimilation and Modelling the Carbon Cycle

Philip Aston, Anna Chuter, **Sylvain Delahaies,**
Anne Skeldon, **Nancy Nichols (*Reading*),**
Ian Roulstone
University of Surrey and NCEO

Outline

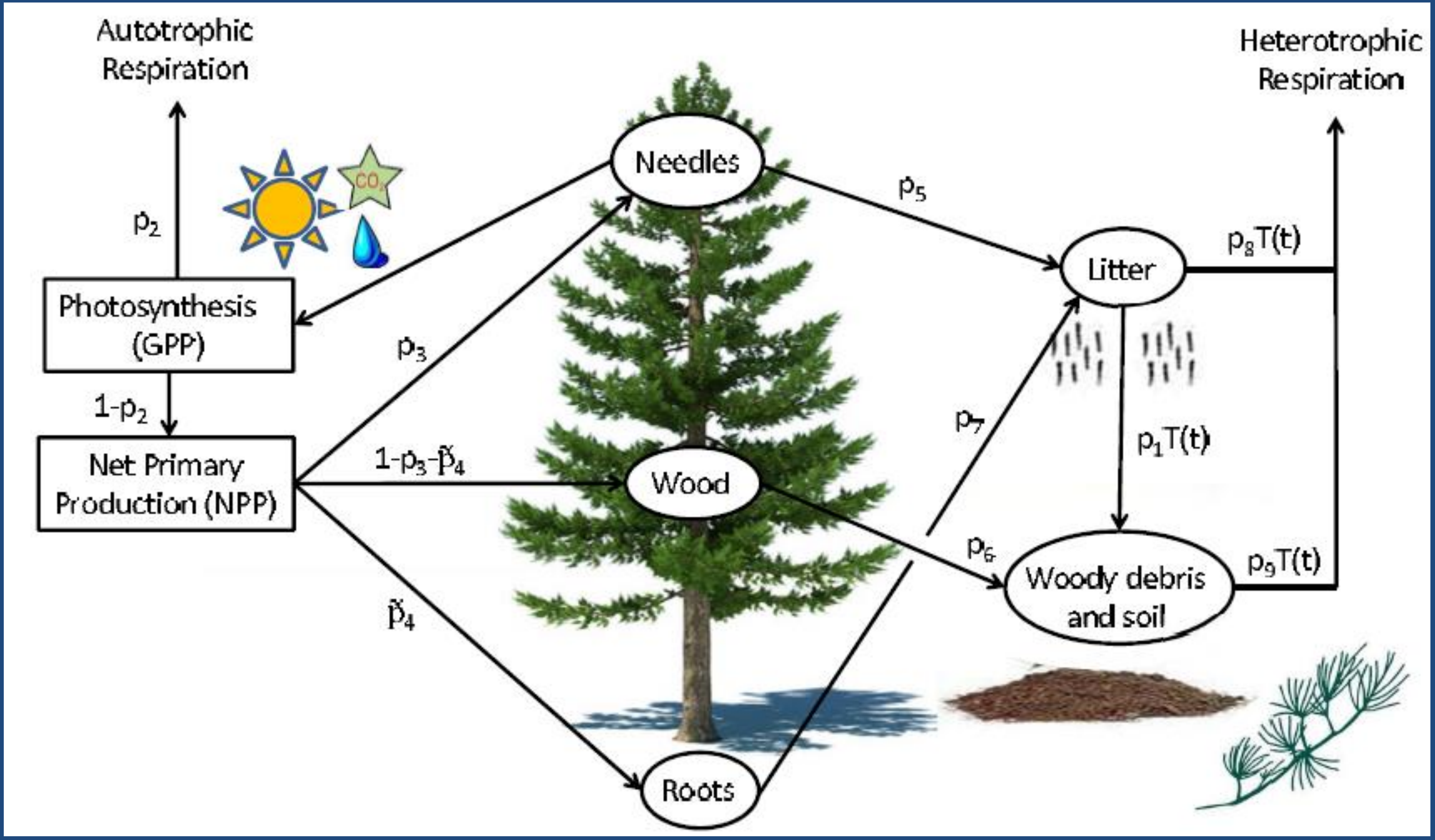
- The Data Assimilation-Linked ECosystem model (DALEC)
- A dynamical systems approach
- Sensitivity analysis
- Data assimilation
- Model and data resolution matrices

Constraining DALEC v2 using multiple data streams and ecological constraints: analysis and application – Delahaies, Nichols and R, Geophys. Model Dev. 2017

UNIVERSITY OF SURREY

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Carbon Cycle



DALEC EV

DALEC Evergreen Model

- The Gross Primary Production (GPP) function $(G(C_\mathrm{f}(t), t))$ represents a daily accumulation of photosynthate which is based on the Aggregated Canopy Model (ACM) (Williams *et al.*, 1997)
- GPP is a complicated function that depends on C_f, a variety of parameters, and on the daily drivers:
  - maximum temperature
  - minimum temperature
  - irradiance
- DALEC is a ‘simple’ model, but is essentially at the heart of all the more complex models
- Our aim is to understand the dynamical behaviour of this model

Pools and Parameters

- 5 carbon pools $(C_f, C_r, C_w, C_l, C_s)$
- 11 parameters $(p_1, \ldots, p_{11})$
- 3 meteorological drivers $(Temp, rad, CO2)$

Parameters used in DALEC EV

	Description	**Value**
p_1	Daily decomposition rate	0.0000044100
p_2	Fraction of GPP respired	0.52
p_3	Fraction of NPP allocated to foliage	0.29
$\tilde{p}_4$	Fraction of NPP allocated to roots	0.2911
p_5	Daily turnover rate of foliage	0.0028
p_6	Daily turnover rate of wood	0.00000206
p_7	Daily turnover rate of roots	0.003
p_8	Daily mineralisation rate of litter	0.02
p_9	Daily mineralisation rate of soil and organic matter	0.00000265
p_{10}	Parameter in exponential term of temperature dependent parameter	0.0693
p_{11}	Nitrogen use efficiency parameter in ACM	7.4

Canopy Model and Leaf Area Index

The aggregated canopy model (ACM) predicts the gross primary production (GPP) at a daily time step

$$GPP = \mathrm{ACM}(Lai, p_{11}, Temp, rad, CO2),$$

where

$$Lai = C_f / LMA.$$

ACM is based on the more complex soil-plant-atmosphere (SPA) model calibrated across a wide range of driving variables to produce a simple model that maintains the essential behaviour of the fine scale model.

Towards Differential Equations

The daily map for the carbon pools are given by

$$\begin{aligned}
C_f(n+1) &= (1-p_5)C_f(n) + p_3(1-p_2)GPP(C_f(n), p_{11}, n) \\
C_r(n+1) &= (1-p_7)C_r(n) + p_4(1-p_3)(1-p_2)GPP(C_f(n), p_{11}, n) \\
C_w(n+1) &= (1-p_6)C_w(n) + (1-p_4)(1-p_3)(1-p_2)GPP(C_f(n), p_{11}, n) \\
C_l(n+1) &= (1-(p_8+p_1)T(p_{10}, n))C_l(n) + p_5 C_f(n) + p_7 C_r(n) \\
C_s(n+1) &= [1-p_9 T(p_{10}, n)]C_s(n) + p_6 C_w(n) + p_1 T(p_{10}, n)C_l(n)
\end{aligned}$$

The net ecosystem exchange (NEE) defined as the difference between Gross primary production and respiration

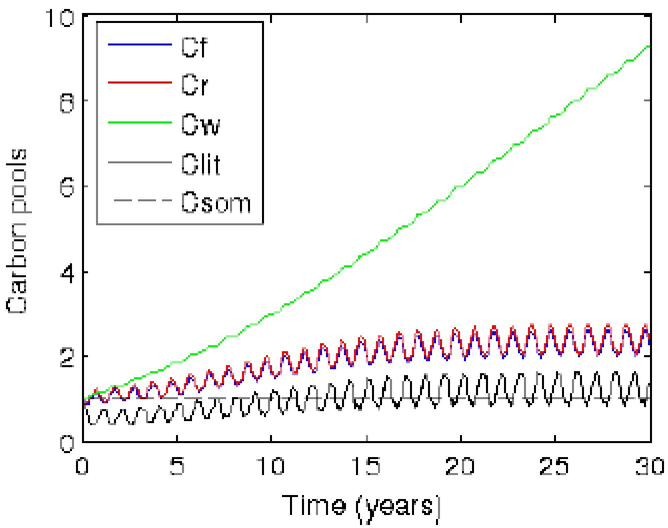
$$NEE = Ra + Rh1 + Rh2 - GPP,$$

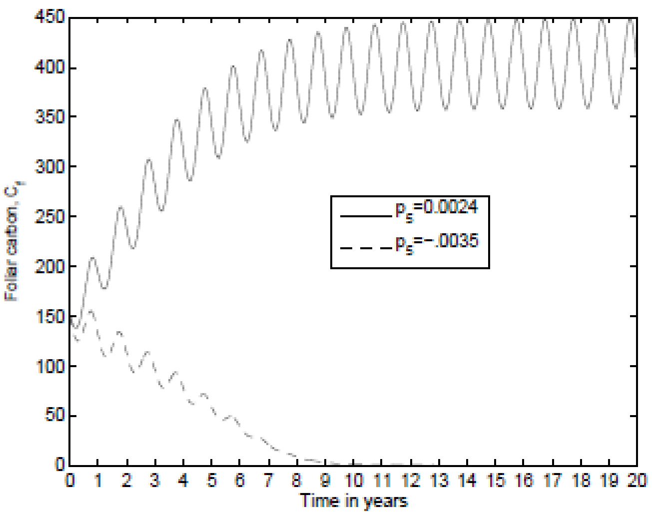
can be expressed as

$$NEE(n) = (1-p_2)GPP - p_8 T(p_{10}, n)C_l(n) - p_9 T(p_{10}, n)C_s(n).$$

DALEC Dynamics

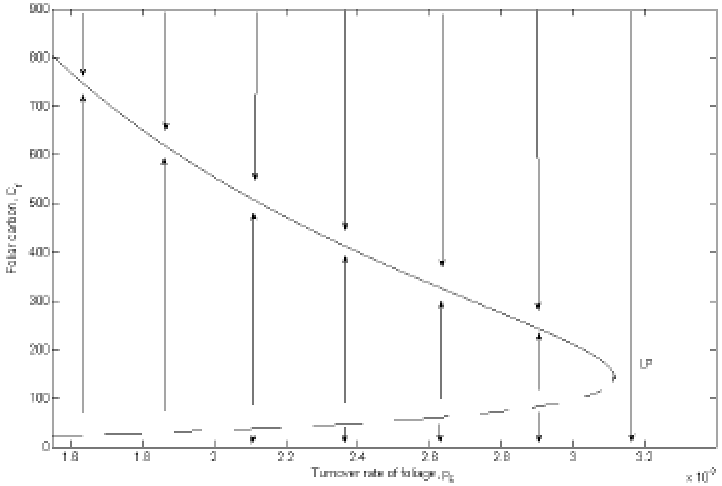
- If the drivers are made periodic with period 1 year, then the carbon pools evolve to a periodic solution
- Note that the pools evolve on different timescales
- The qualitative behaviour is parameter dependent





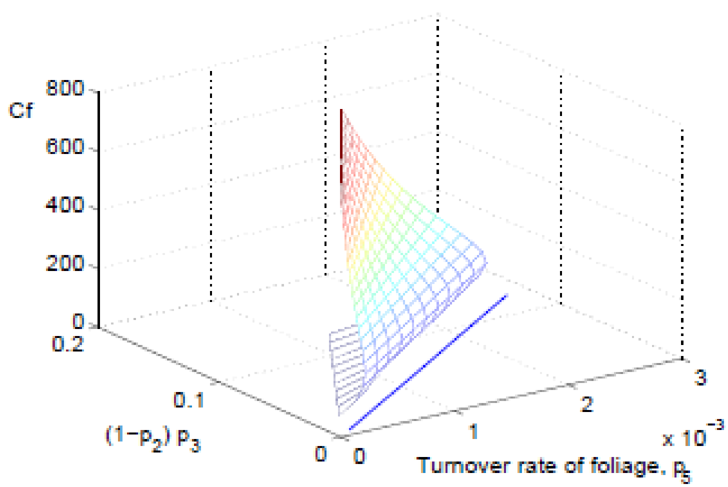
Periodic Solutions

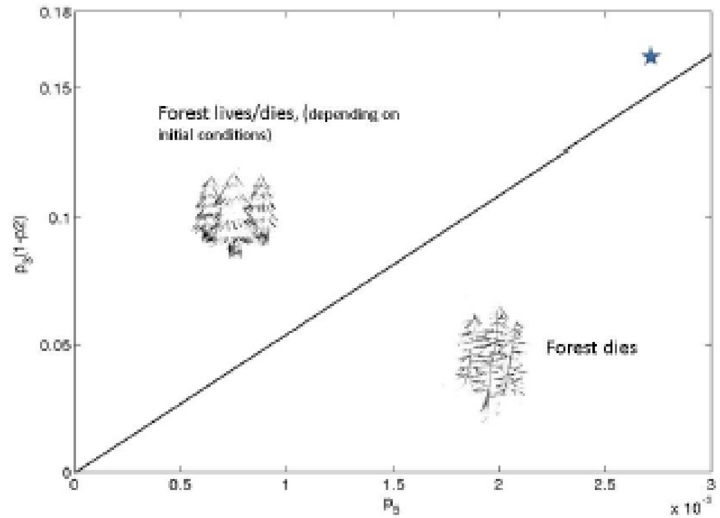
- We find periodic solutions as fixed points of an annual map, which satisfy $C_{\mathrm{f}}(0) = C_{\mathrm{f}}(365)$
- $C_{\mathrm{f}} = 0$ is always a stable fixed point
- If $C_{\mathrm{f}} = 0$, then all the other pools converge to zero as well – one dead forest!
- Non-zero fixed points of C_{f} can be found numerically
- p_5 is the rate at which foliar carbon goes into the litter
- If the needles drop at a high enough rate, then there is not sufficient carbon to sustain the tree and it will die



Limit Points

- The C_f equation depends only on p_5 (the rate at which foliar carbon goes into the litter) and $p_2(1-p_3)$ (the fraction of GPP allocated to the foliar carbon)
- We can find fixed points of C_f as a function of these two parameters
- This gives a line of limit points





- Maximum growth would be achieved by allocating as much carbon as possible to the wood and roots, keeping the foliar carbon to a minimum

Sensitivity

Parameter	Value	$\frac{\delta C_f}{\delta p_i} \frac{p_i}{100}$	$\frac{\delta NEE}{\delta p_i} \frac{p_i}{100}$
p_1	0.0000044	0	-0.00065
p_2	0.52	-1.8	8.2
p_3	0.29	1.7	-2.5
p_4	0.41	0	0.48
p_5	0.0028	-1.6	3.4
p_6	0.0000021	0	0.000038
p_7	0.0030	0	0.95
p_8	0.020	0	0.43
p_9	0.0000027	0	0.11
p_{10}	0.069	0	0.34
p_{11}	7.4	0.83	-3.8

Sensitivity

- From LAI measurements one can only estimate
  - (i) the turnover rate of foliar carbon (p_5),
  - (ii) the fraction of GPP allocated to foliar carbon ($p_3(1-p_2)$).
  - (iii) p_{11} (parameter in GPP).
- NEE is only sensitive to some of the parameters and hardly at all to the decomposition rate of litterfall, p_1, and turnover rate of wood, p_6.

CHAOS **25**, 000000 (2015)

A dynamical systems analysis of the data assimilation linked ecosystem carbon models

Anna M. Chuter, Philip J. Aston,[a)] Anne C. Skeldon, and Ian Roulstone
Department of Mathematics, University of Surrey, Guildford, Surrey GU2 7XH, United Kingdom

REgional Flux Estimation EXperiment (REFLEX)

A. Fox et al. (2009) Agricultural and Forest Meteorology 149.

Aims

- To compare the strengths and weaknesses of various data assimilation techniques for estimating carbon model parameters and predicting carbon fluxes.
- To quantify errors and biases introduced when extrapolating fluxes.

Experiments

- Nine participants using Monte Carlo methods and EnKF,
- Assimilation of both real and synthetic NEE and LAI observations over a two year period.

Results

- parameters directly linked to GPP and respiration were best constrained and characterised,
- parameters related to the allocation to and turnover of fine root/wood pools.

4D VAR

Variational Data Assimilation

UNIVERSITY OF SURREY

$$J(x)=\frac{1}{2}(x_0-x_b)^T B^{-1}(x_0-x_b)+\frac{1}{2}\sum_{i=1}^{N}(y_i-H(x_i))^T R_n^{-1}(y_i-H(x_i))$$

Find

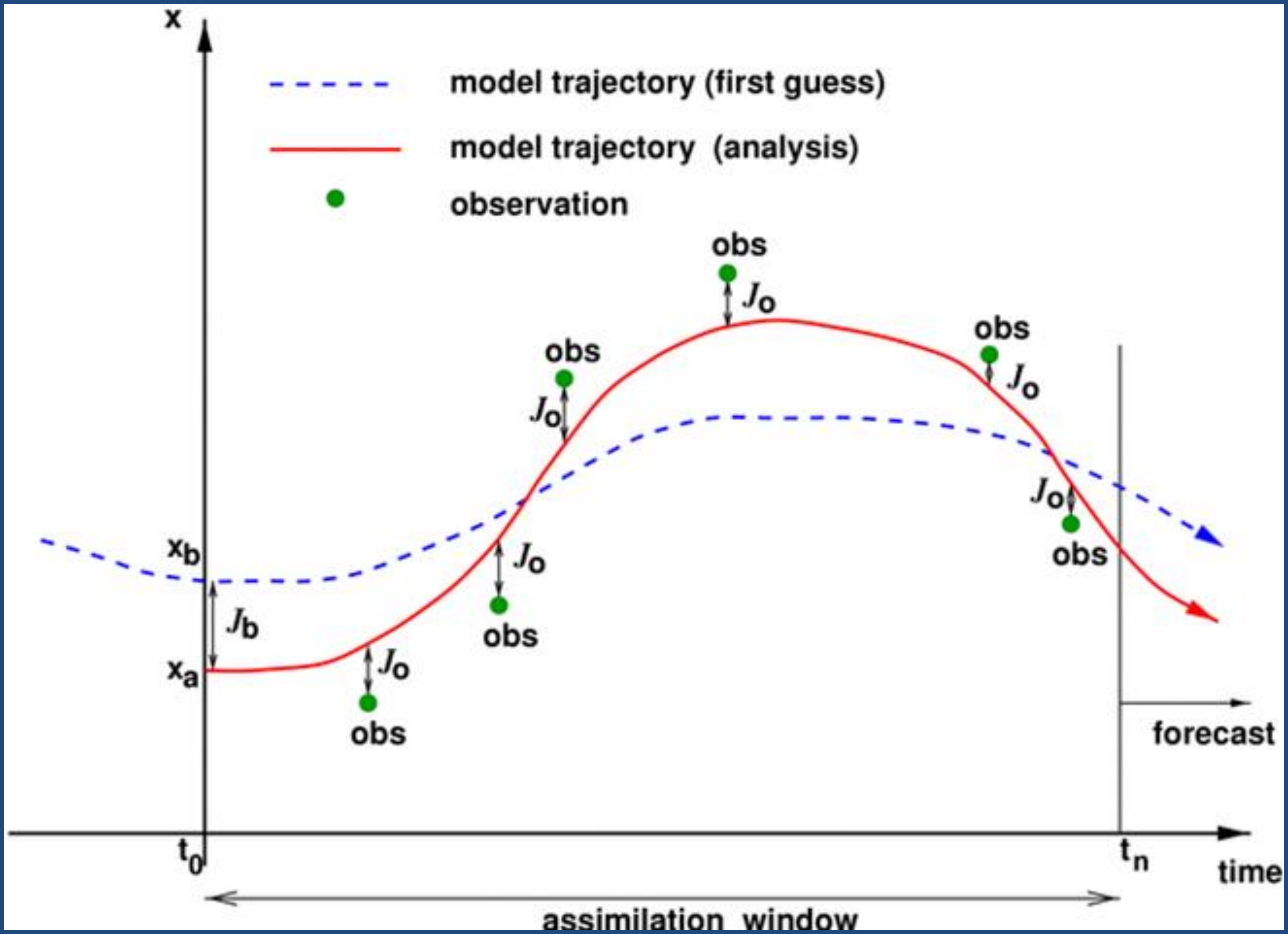
$$\min_{x_0} J(x)$$

Subject to the strong constraint that the model states are a solution to the numerical model and that the tangent linear hypothesis holds.

Adjoint variable λ:

$$\frac{\partial J}{\partial x_0} \Leftarrow -\lambda_0$$

4D VAR



x
model trajectory (first guess)
model trajectory (analysis)
observation
obs
J_o
obs
J_o
obs
J_o
J_o
x_b
J_o
J_b
obs
J_o
obs
x_a
obs
forecast
t_0
t_n
time
assimilation window

Inverse Problem and Data Assimilation

At the heart of incremental-4DVAR lies a linear inverse problem $Ax = b$, the least square solution is given by

$$x^* = \mathrm{argmin}\|Ax - b\|^2 = \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i,$$

using a singular value decomposition $A = U\Sigma V^T$ with $\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_n)$.

When the data b is contaminated with noise the least square solution can be unreliable. Given $\bar{x}$ and $\bar{b}$ such that $A\bar{x} = \bar{b}$ we have

$$\frac{\|\delta x\|}{\|\bar{x}\|} \leq \kappa(A) \frac{\|\delta b\|}{\|\bar{b}\|},$$

where

- $\delta x = x^* - \bar{x}$ and $\delta b = b - \bar{b}$,
- $\kappa(A) = \sigma_1 / \sigma_n$ is the condition number of A.

Well-posedness

A generic inverse problem consists in finding a n-dimensional state vector ***x*** such that

$$\mathbf{h}(\boldsymbol{x}) = \boldsymbol{y}$$

for a given N-dimensional observation vector ***y***, including random noise, and a given model **h**. The problem is well posed in the sense of Hadamard (1923) if the three following conditions hold:

1) a solution exists,
2) the solution is unique, and
3) the solution depends continuously on the input data.

An ill-posed problem

	$\mathbf{x}_0$	$\delta\mathbf{x}_0$	RE_0	RE_1	RE_2	RE_3
C_f	298.4	25.0	1.1×10^{-12}	2.3×10^{-9}	1.2×10^{-4}	0.3
C_r	280.1	19.0	3.5×10^{-7}	5.7×10^{-5}	6.3×10^{1}	3.9×10^{4}
C_w	10410.6	920.0	1.8×10^{-6}	7.1×10^{-4}	1.11×10^{3}	3.1×10^{5}
C_l	38.9	16.0	2.8×10^{-8}	4.5×10^{-6}	5.0	3.1×10^{3}
C_s	10631.0	1100.0	8.7×10^{-7}	1.1×10^{-5}	1.0×10^{-5}	1.7×10^{4}
p_1	4.41×10^{-6}	4.0×10^{-7}	5.9×10^{-4}	9.4×10^{-2}	1.0×10^{5}	6.4×10^{7}
p_2	0.51	5.2×10^{-2}	1.3×10^{-10}	1.5×10^{-8}	3.2×10^{-3}	2.7
p_3	0.29	2.9×10^{-2}	8.4×10^{-11}	1.0×10^{-8}	3.5×10^{-3}	1.7
p_4	0.41	4.1×10^{-2}	2.3×10^{-7}	3.7×10^{-5}	4.1×10^{-5}	2.5×10^{4}
p_5	2.8×10^{-3}	2.8×10^{-4}	2.8×10^{-12}	4.6×10^{-9}	9.2×10^{-4}	1.0
p_6	2.06×10^{-6}	2.06×10^{-7}	5.8×10^{-6}	6.0×10^{-4}	4.3×10^{-2}	4.1×10^{5}
p_7	3.0×10^{-3}	3.0×10^{-4}	2.2×10^{-9}	1.1×10^{-6}	1.4	8.2×10^{2}
p_8	2.0×10^{-2}	2.0×10^{-3}	1.1×10^{-7}	1.8×10^{-5}	2.0×10^{-5}	1.2×10^{4}
p_9	2.65×10^{-5}	2.65×10^{-7}	9.0×10^{-6}	1.1×10^{-4}	1.0×10^{2}	1.8×10^{5}
p_{10}	6.93×10^{-2}	6.93×10^{-3}	1.9×10^{-10}	2.1×10^{-8}	2.2×10^{-3}	6.4
p_{11}	7.4	0.74	1.4×10^{-10}	2.4×10^{-8}	1.9×10^{-3}	4.9
$\|RE\|$	–	–	2.4×10^{-24}	1.8×10^{-8}	1.0×10^{2}	2.2×10^{7}

Table: The condition number is the same for each simulation $\kappa(A)\approx 1.1\times10^{23}$. $\delta\mathbf{x}_0$ is the perturbation used to generate the observations. The row RE_0 gives the relative error for observations without noise, the column RE_1 (resp. RE_2, RE_3) gives the relative error for the analysis for observations with a gaussian noise with variance $\sigma = 1.1\times10^{-16}$ (resp. $\sigma = 1.0\times10^{-5}$, $\sigma = 5.0\times10^{-1}$).

Model Reduction?

	RE
C_f	1.2×10^1
C_r	3.9×10^4
C_w	3.1×10^5
C_l	3.1×10^3
C_s	1.7×10^4
p_1	6.4×10^7
p_2	7.4×10^{-3}
p_3	1.1×10^1
p_4	2.5×10^4
p_5	3.6
p_6	2.4×10^5
p_7	8.0×10^2
p_8	1.2×10^4
p_9	1.2×10^5
p_{10}	2.3
p_{11}	11.6

	RE
C_f	6.5
C_l	0.5
C_s	7.4
p_2	1.2
p_3	1.3
p_4	0.2
p_5	3.1
p_8	1.8
p_{10}	3.5×10^{-2}
p_{11}	4.3

Linear Analysis

Considerable theoretical insights into the nature of the inverse problem, and the ill-posedness, can be obtained by studying a linearisation of the operator **h**. A first approximation to the inverse problem consists in finding a perturbation ***z*** which best satisfies the linear equation

$$\mathbf{H}\boldsymbol{z} = \boldsymbol{d}$$

where **H** is the tangent linear operator for **h** and ***d*** is a perturbation of the observations.

- Finding a solution $\boldsymbol{z}$ amounts to constructing a generalized inverse $\mathbf{H}^g$ such that formally

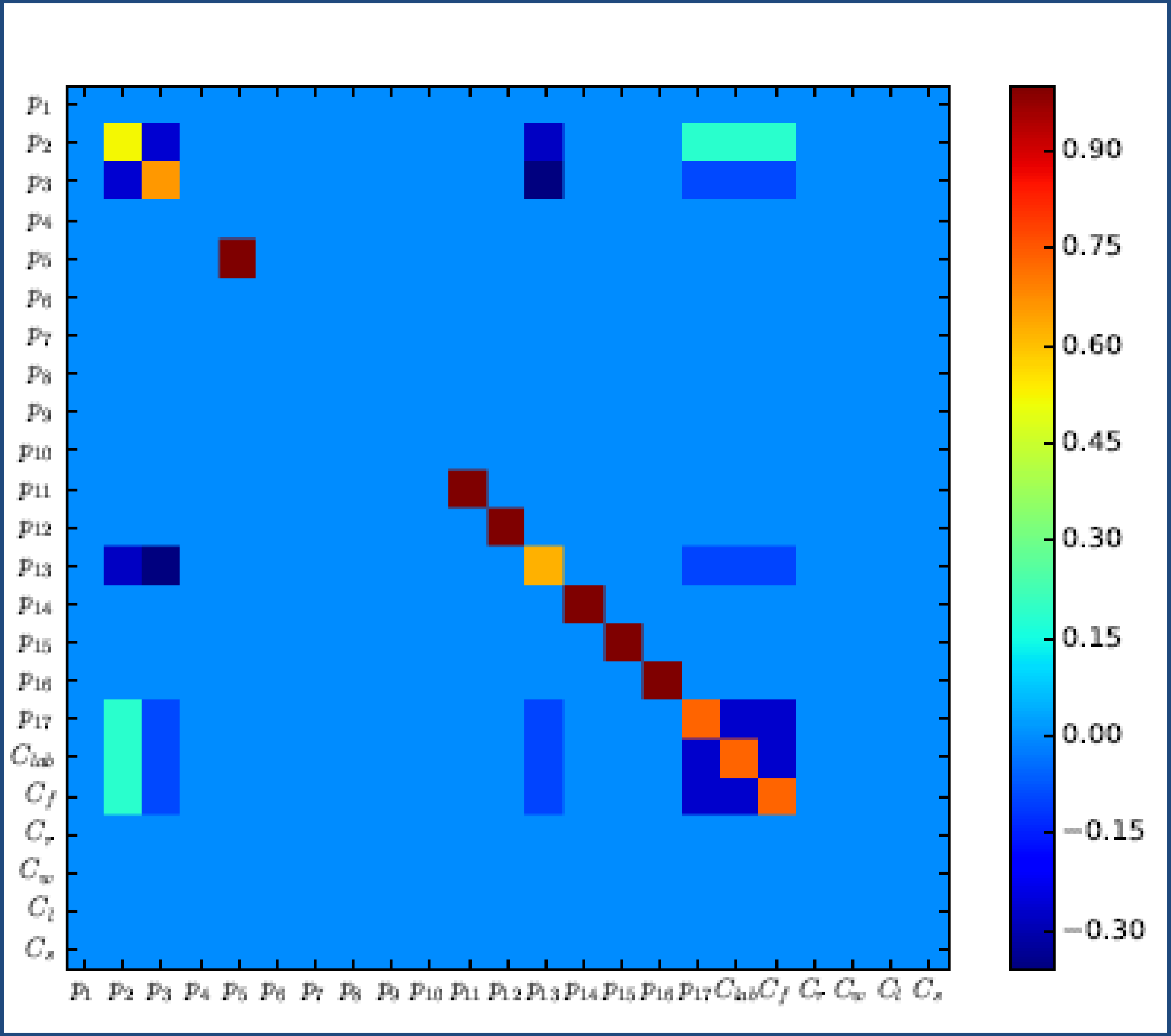
$$z = \mathbf{H}^g \boldsymbol{d}$$

- Assuming a true state $\boldsymbol{z^*}$ exists, possibly unknown, then we can define an operator **N** called the <u>model resolution matrix</u> which relates the solution $\boldsymbol{z}$ to the true state

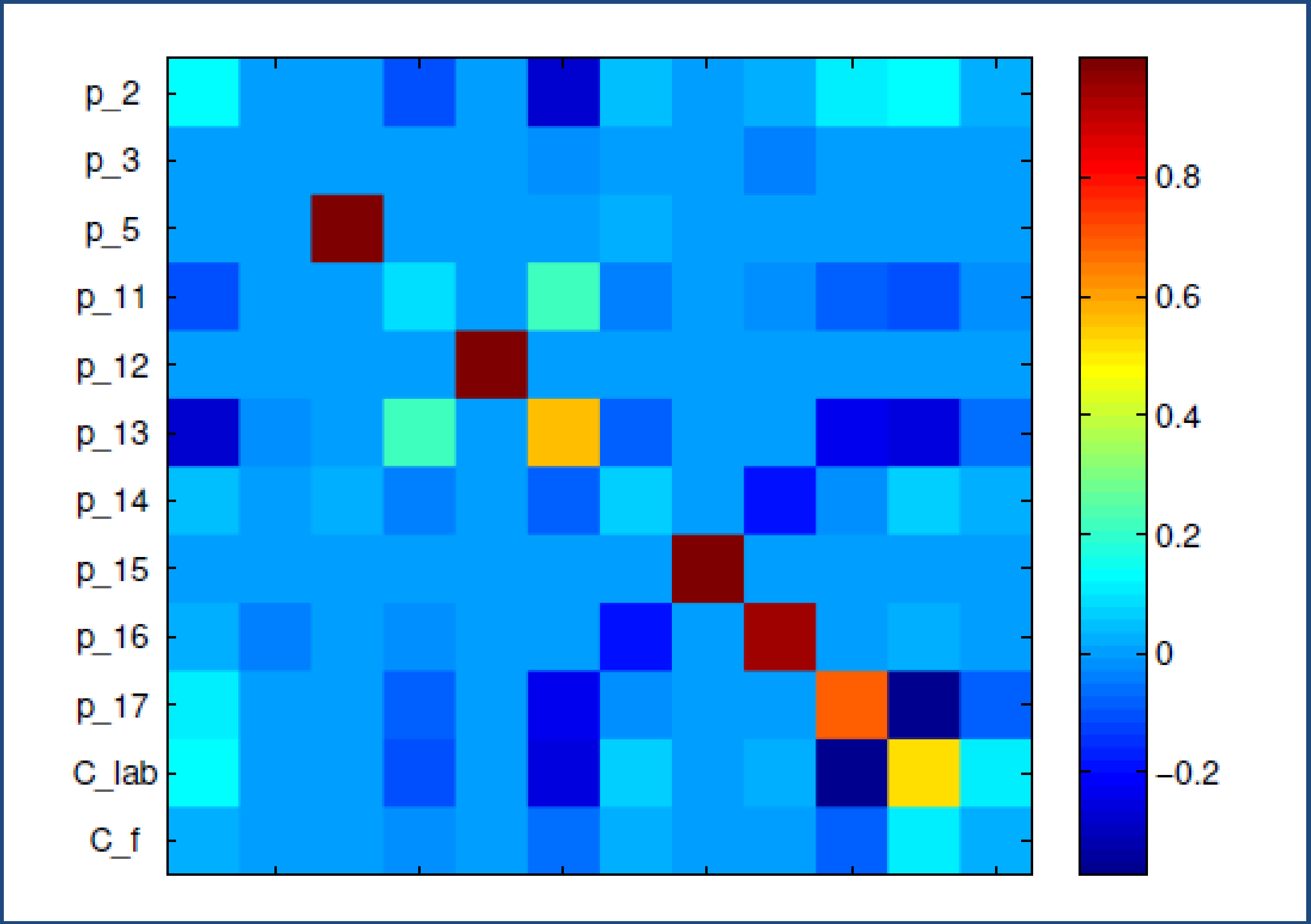
$$\boldsymbol{z} = \mathbf{H}^g \mathbf{H}\boldsymbol{z^*} = \mathbf{N}\boldsymbol{z^*}$$

- This matrix **N** gives a practical tool to analyse the resolution power of an inverse method, that is its ability to retrieve the true state, including or not any regularization method
- The closer **N** is to the identity the better the resolution.
- The trace of the matrix defines a natural notion of information content (IC).
- Similarly a <u>data resolution matrix</u> can be defined to study how well data can be reconstructed and its diagonal elements naturally define a notion of data importance.

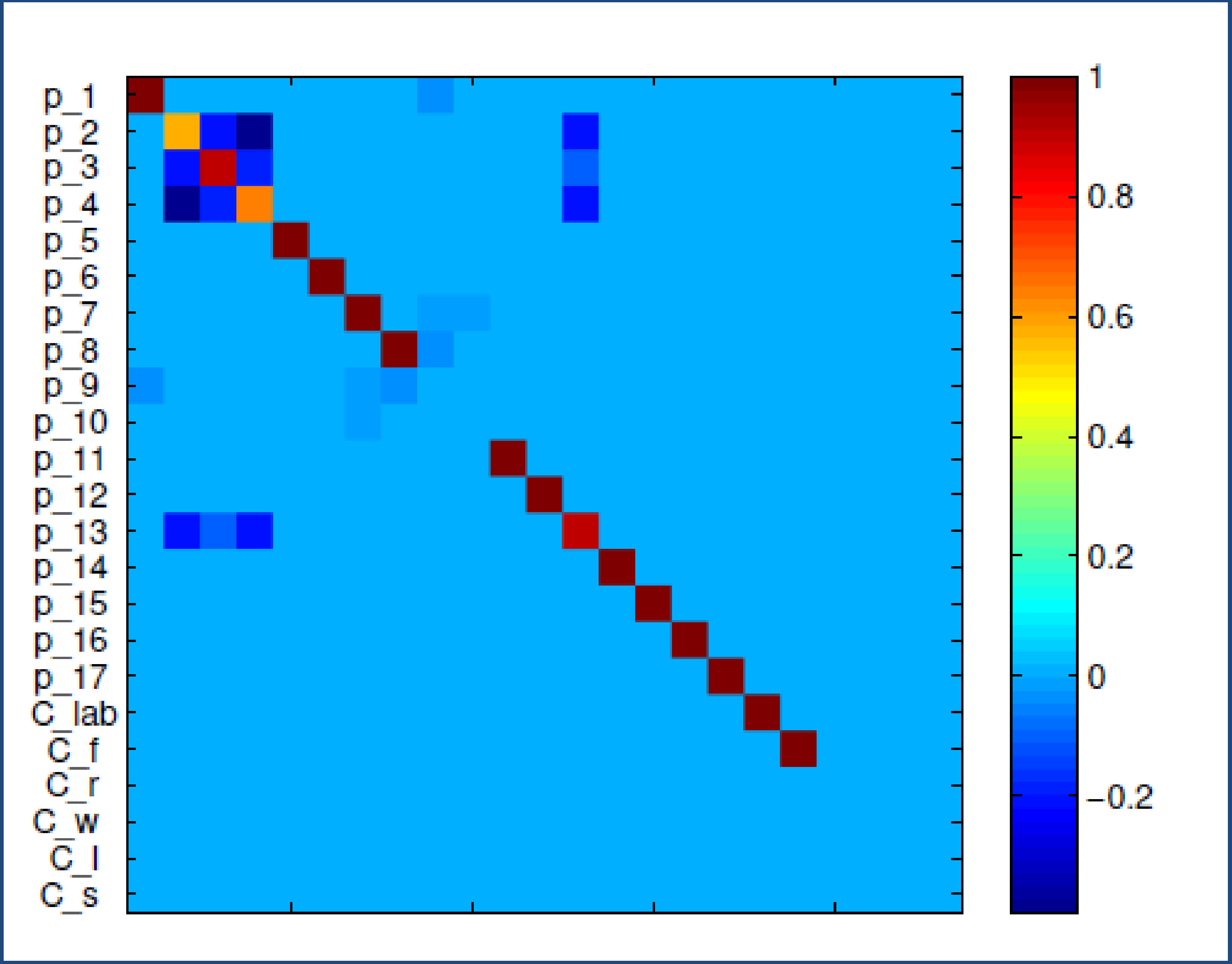
Model Resolution: LAI



Reduced Model



EDCs



p_1
p_2
p_3
p_4
p_5
p_6
p_7
p_8
p_9
p_10
p_11
p_12
p_13
p_14
p_15
p_16
p_17
C_lab
C_f
C_r
C_w
C_l
C_s
1
0.8
0.6
0.4
0.2
0
−0.2

Constraining DALEC2 using multiple data streams and ecological constraints: analysis and application

Sylvain Delahaies[1], Ian Roulstone[1], and Nancy Nichols[2]

[1]Department of Mathematics, University of Surrey, Guildford, UK.
[2]Department of Mathematics, University of Reading, Reading, UK.

Correspondence to: S. Delahaies (s.b.delahaies@surrey.ac.uk)

Future Work

- Sensitivity analysis of TRIFFID/JULES (Met O model)
- Sensitivity analysis of visible radiance, near infrared radiance and vegetation index, to the model parameters of the two-stream Sellers approximation of radiative transfer (with NPL)