Geogg124: Data assimilation

P. Lewis

What is Data Assimilation?

Optimal merging of models and data

- Models
  - Expression of current understanding about process
  - E.g. terrestrial C model
- Data
  - Observations
  - E.g. EO

Some basic stats

Gaussian PDF:

$$P_b(x) = \frac{1}{(2\pi)^{\frac{\pi}{2}}\sqrt{\det(B)}} \exp\left(-\frac{1}{2}(x_b - x)^T B^{-1}(x_b - x)\right)$$

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

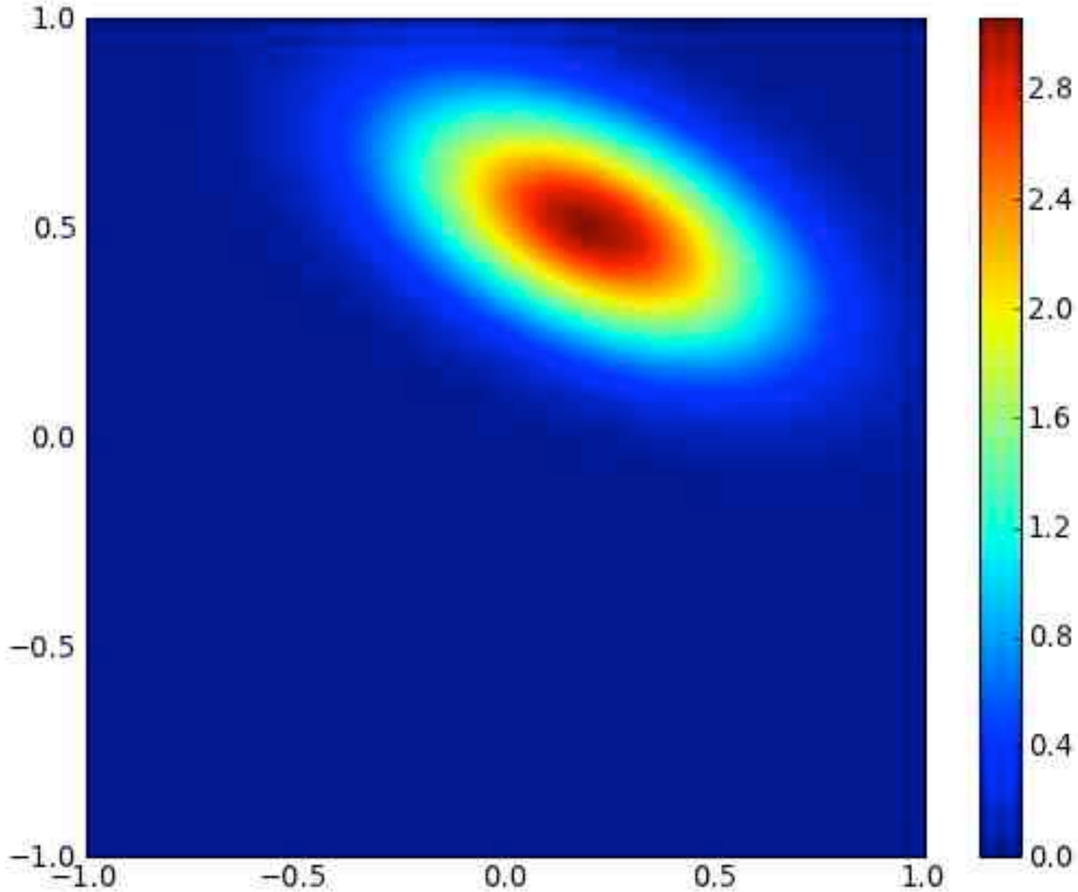
UCL

The uncertainty matrix

$$B = \begin{bmatrix} \sigma_0^2 & \rho_{0,1}\sigma_0\sigma_1 \\ \rho_{0,1}\sigma_0\sigma_1 & \sigma_1^2 \end{bmatrix}$$

$$B = \begin{bmatrix} b_{0,0} & b_{0,1} \\ b_{0,1} & b_{1,1} \end{bmatrix}$$

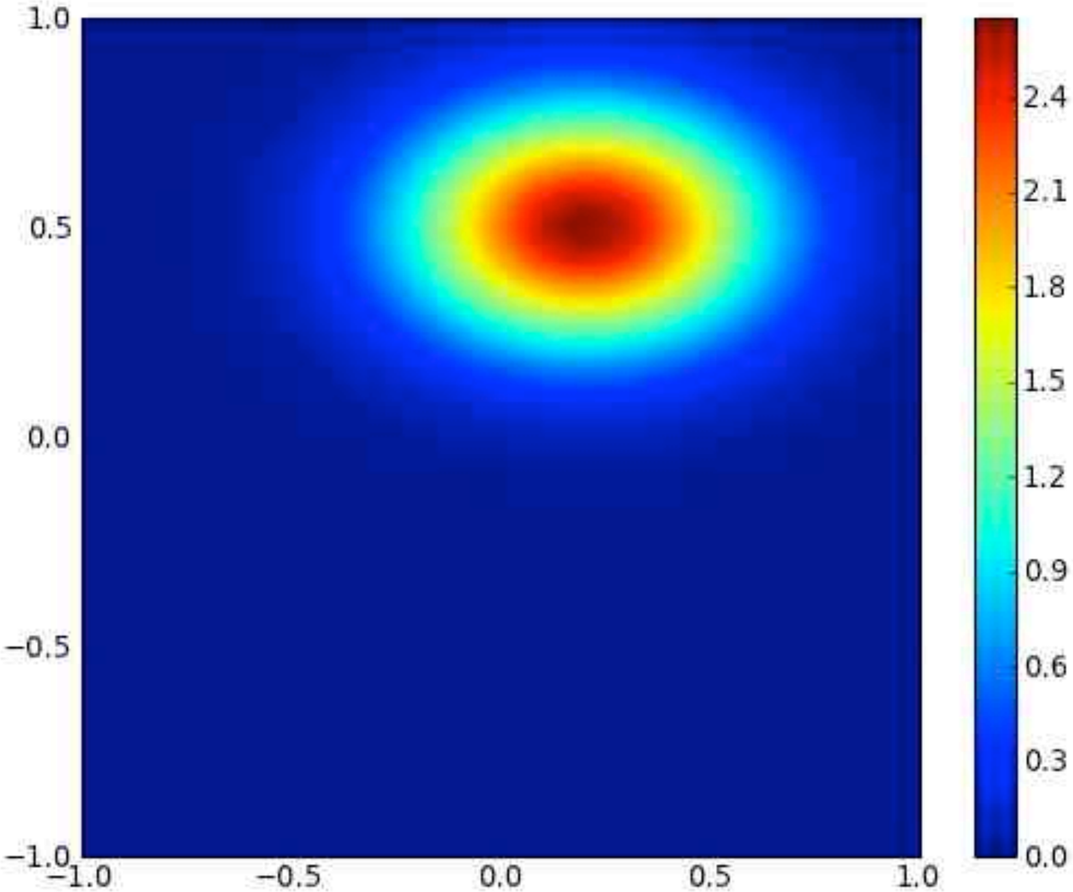
X=(0.2,0.5) sigma=(0.3,0.2), rho=-0.5



National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

X=(0.2,0.5) sigma=(0.3,0.2), rho=0.0



National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

Combining probabilities

Bayes theorum

$$P(b|a) = \frac{P(b)P(a|b)}{P(a)}$$

Observation operator

Model of y:

$$\hat{y} = H(x)$$

$$y = \hat{y} + \epsilon = H(x) + \epsilon$$

The PDF of the observations is the PDF of the observations *given x*

The observation PDF

$$P(y|x) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det(R)}}\exp\left(-\frac{1}{2}(y - H(x))^T R^{-1}(y - H(x))\right)$$

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

Using Bayes theorum

$$P(x|y) = \frac{P(x)P(y|x)}{P(y)}$$

$$P(x|y) \propto P(x)P(y|x)$$

$$P(x') = P(x|y) \propto \exp\left(-\frac{1}{2}(x_b - x)^T B^{-1}(x_b - x)\right)\exp\left(-\frac{1}{2}(y - H(x))^T R^{-1}(y - H(x))\right)$$

So … we can improve on a prior estimate (x_b) by multiplying by the observation PDF

Using Bayes theorum

$$J_b(x) = \frac{1}{2}(x_b - x)^T B^{-1}(x_b - x)$$

$$J_o(x) = \frac{1}{2}(y - H(x))^T R^{-1}(y - H(x))$$

$$J(x) = J_b(x) + J_o(x)$$

$$P(x') \propto \exp(-J(x))$$

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

Differentials of J

Minimisation:

- Need to know the derivatives
- J' : Jacobian

Uncertainty:

- Posterior uncertainty:
  - Curvature of cost function
  - Inverse of 2nd O derivative of J, J''
  - Hessian

Differentials of J

$$J'_b = -B^{-1}(x_b - x)$$

$$J'_o = -H'(x)^T R^{-1}(y - H(x))$$

$$J' = J'_b + J'_o$$

$$J''_b = B^{-1}$$

$$J''_o = H'(x)^T R^{-1} H'(x) - H''(x)^T R^{-1}(y - H(x))$$

$$J'' = J''_b + J''_o$$

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

Posterior uncertainty

Follows from previous that (for linear case)

$$C_{post} = \left(B^{-1} + H^T R^{-1} H\right)^{-1}$$

Univariate example

If H=I (identity operator)

$$C_{post} = \left(B^{-1} + R^{-1}\right)^{-1}$$

Univariate:

$$\frac{1}{\sigma_{post}^2} = \frac{1}{\sigma_a^2} + \frac{1}{\sigma_b^2}$$

Equal variances:

$$\sigma_{post} = \frac{\sigma_a}{\sqrt{2}}$$

National Centre for Earth Observation

NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

Optimal estimate

Solve for univariate case:

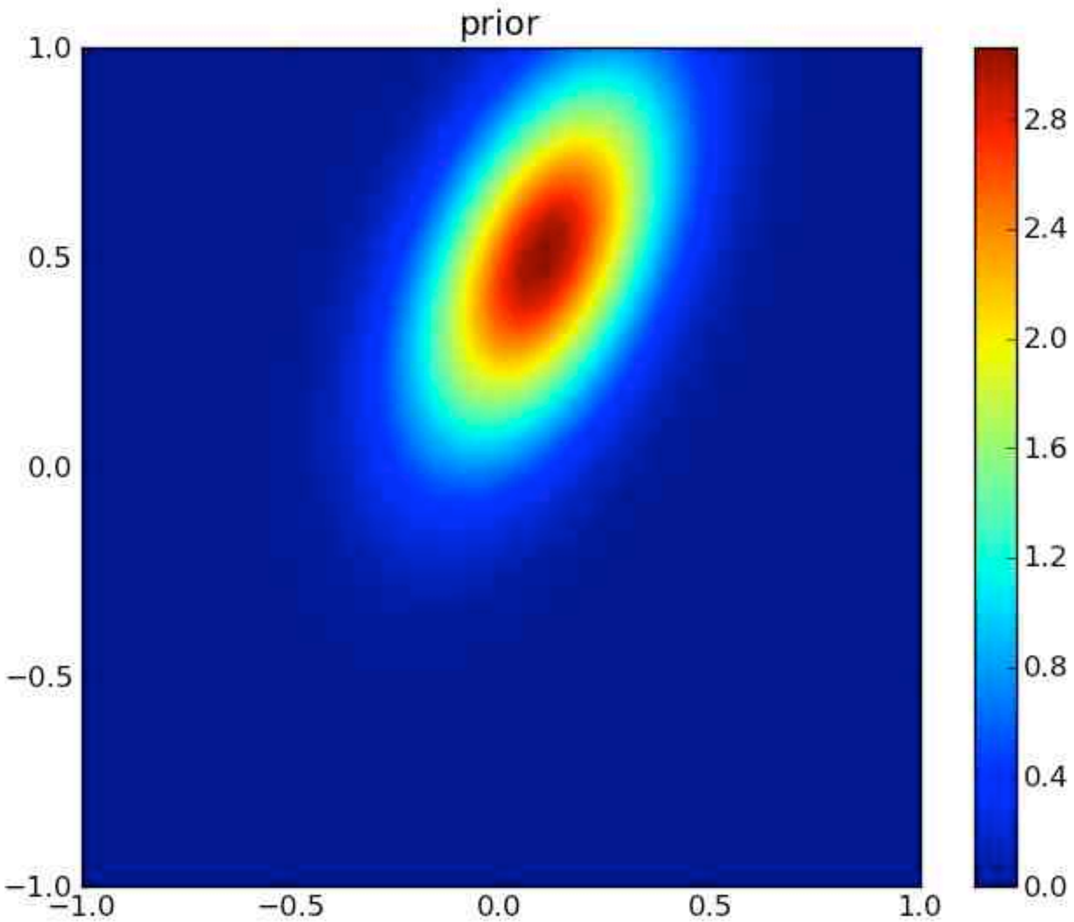
Samples (x_a, x_b) (sigma$_a$, sigma$_b$):

$$J' = 0 = \frac{-(x_a - x)}{\sigma_a^2} + \frac{-(x_b - x)}{\sigma_b^2}$$

$$x\left(\frac{1}{\sigma_a^2} + \frac{1}{\sigma_b^2}\right) = \frac{x_a}{\sigma_a^2} + \frac{x_b}{\sigma_b^2}$$

$$x = \frac{x_a\sigma_b^2 + x_b\sigma_a^2}{\sigma_a^2 + \sigma_b^2}$$

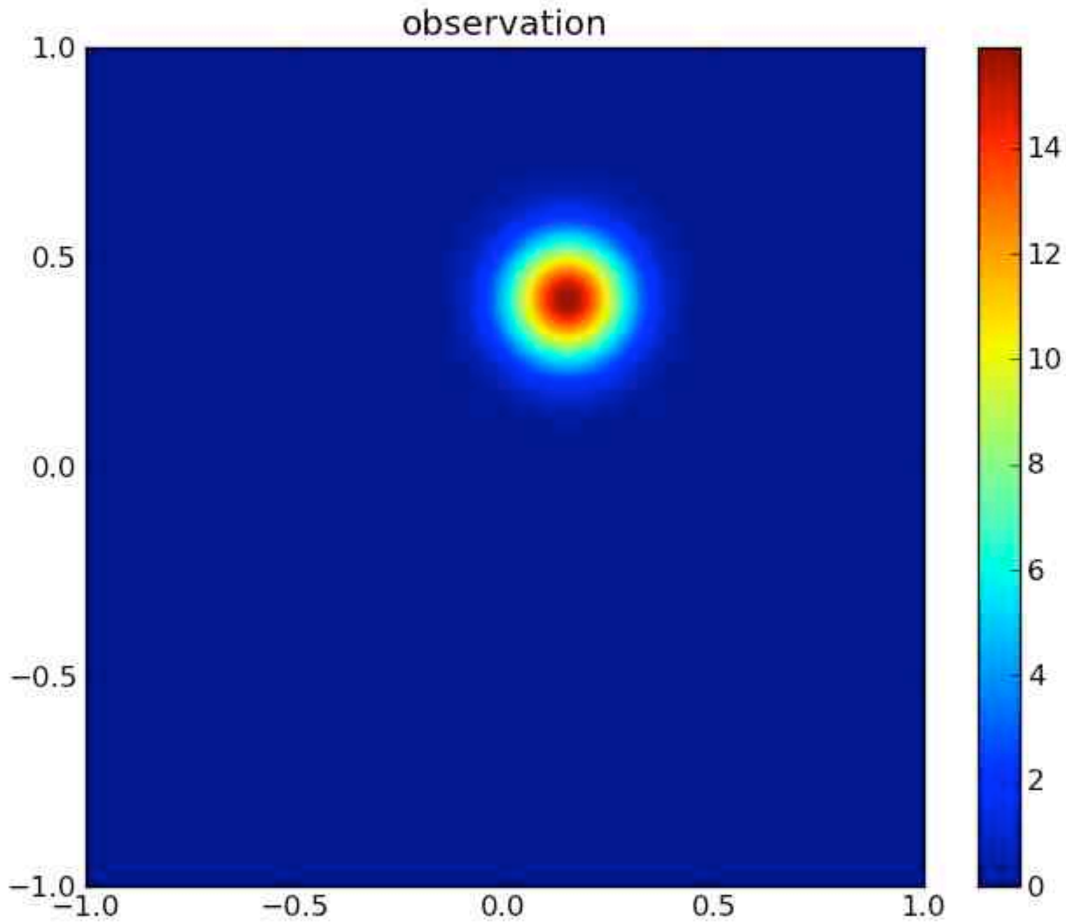
prior



National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

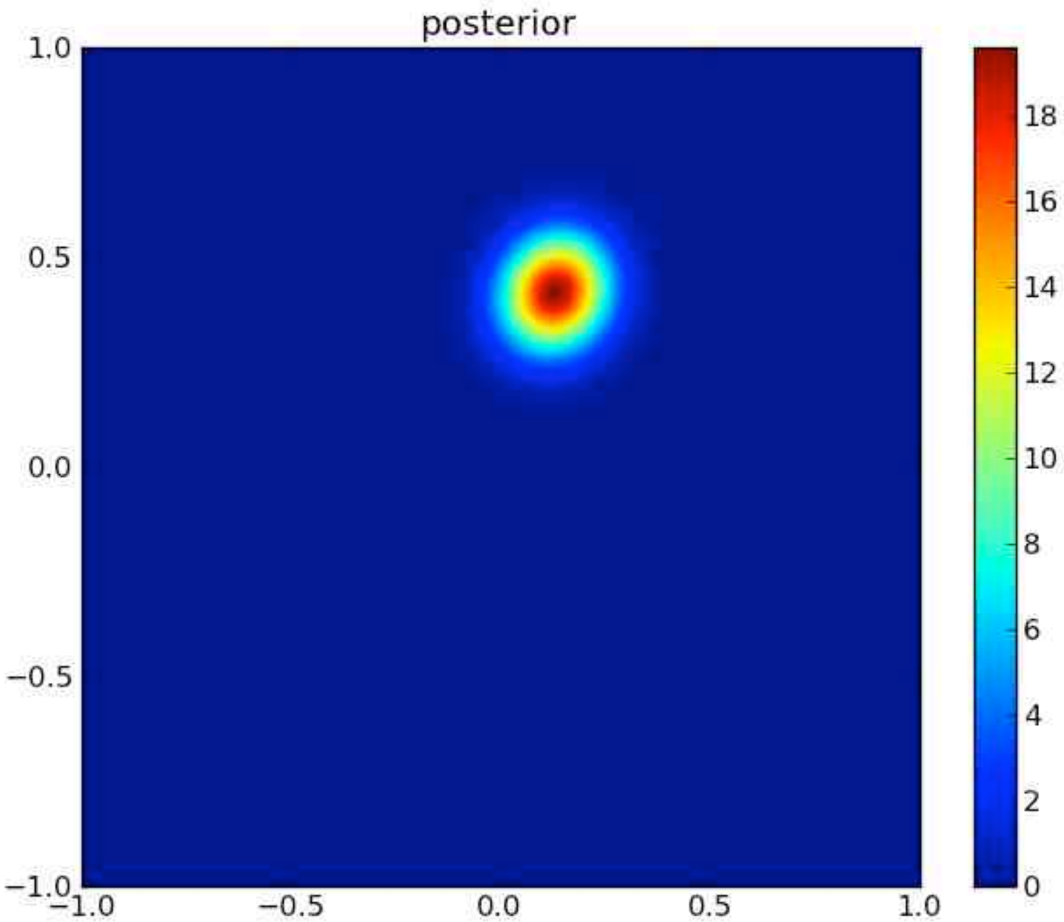
observation



National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

posterior



National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

Finding solutions

DA methods

- Variational (strong, weak constraint)
- Sequential
- MCMC (briefly)

Variational data assimilation: strong constraint

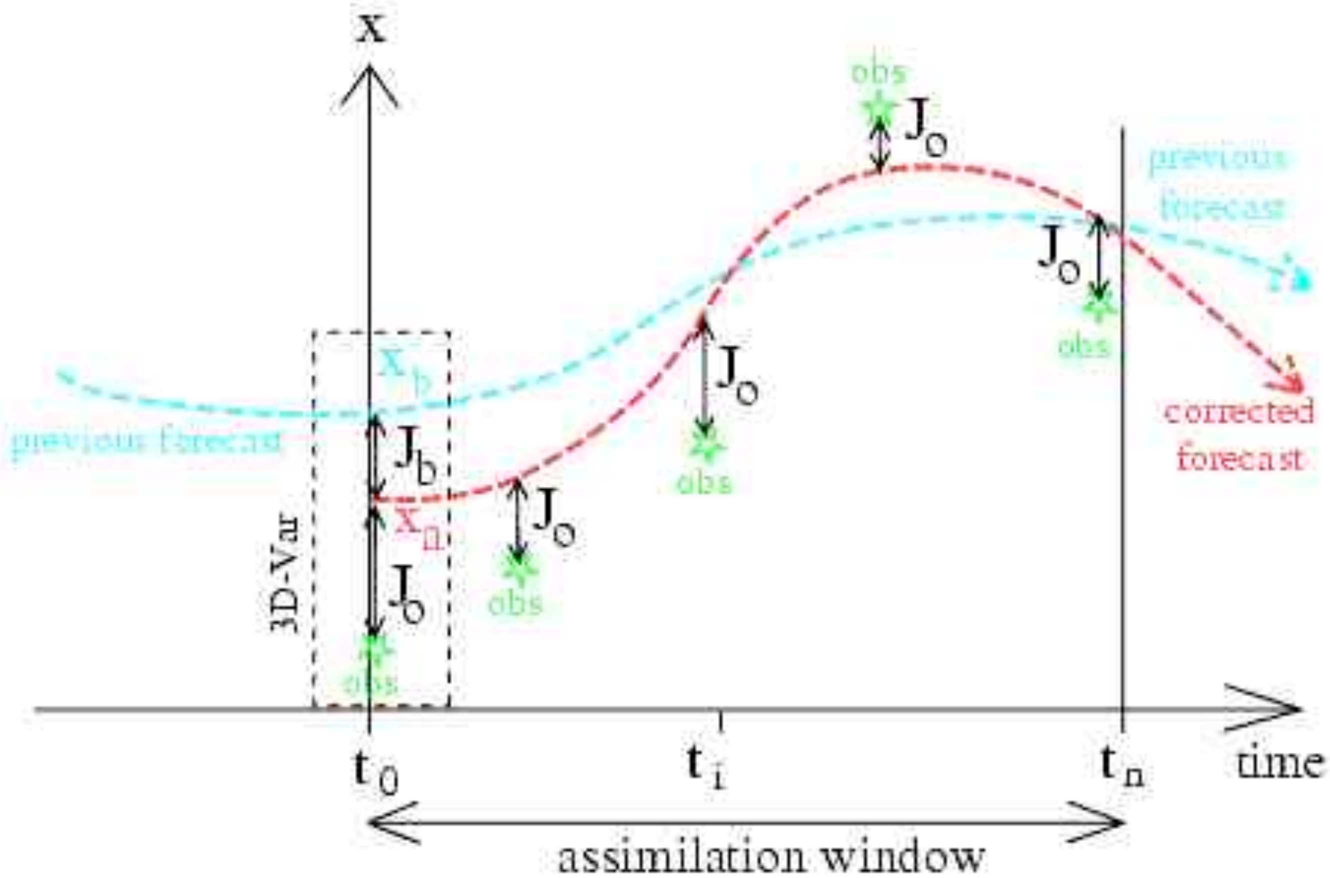
Cost function minimisation

$$J(x) = J_b(x) + \sum_{i=0}^{i=n} J_{oi}(x)$$

- Multiple obs constraints
- Plus prior

- Strong constraint:
  - Prediction must be a valid model state
  - ‘trust the model’

Variational data assimilation: strong constraint



Variational data assimilation: weak constraint

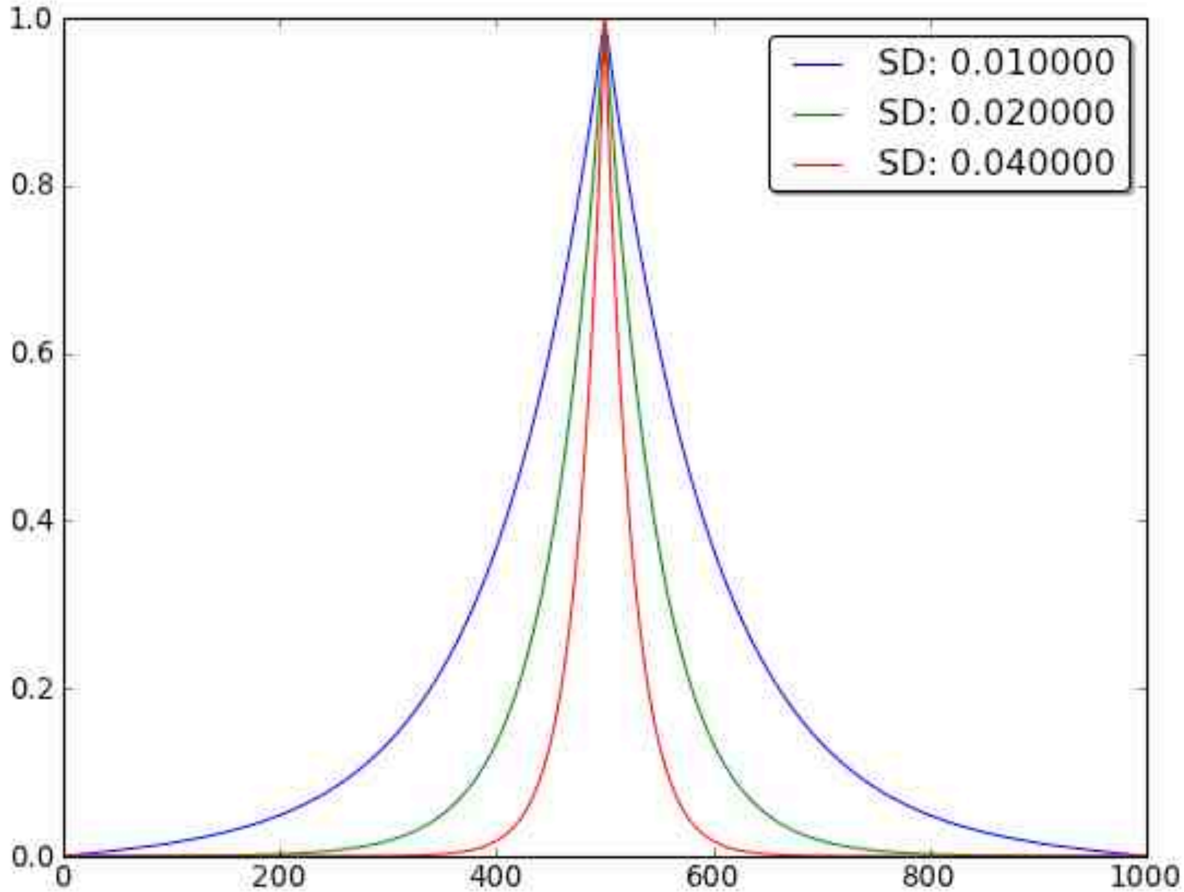
Sometime want model as a 'guide' but don't trust it.

- So define model uncertainty
- Use appropriate cost function for model

- E.g. model $Dx = 0$ $$J_D(x) = \frac{1}{2}(Dx)^T C^{-1}(Dx)$$

$$D = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \dots & -1 & 0 \\ 0 & 0 & 0 & \dots & 1 & -1 \end{bmatrix}$$

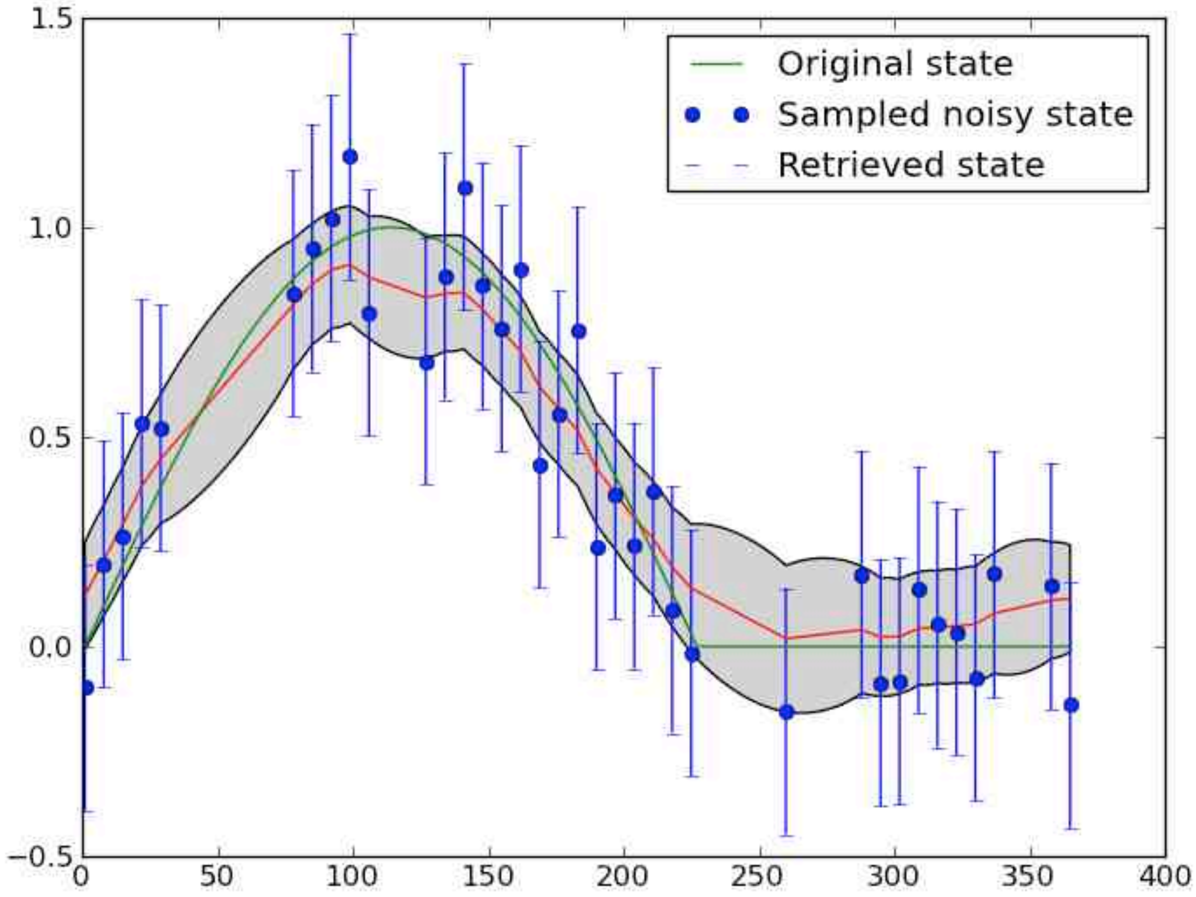
Dx: ~equivalent to convolution with Laplace function: smoother



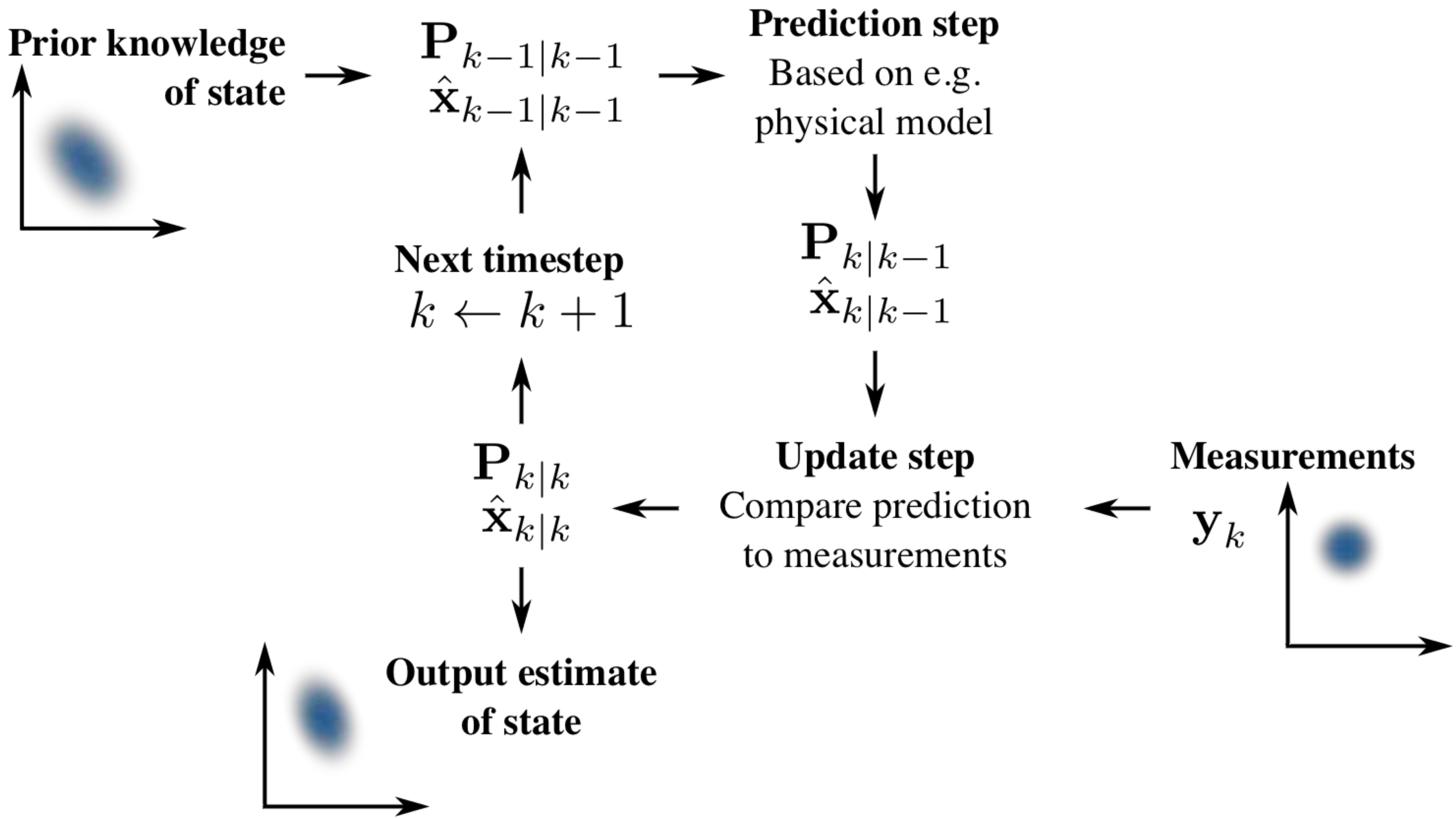
National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

Dx constraint as a smoother



Sequential methods: Kalman Filter



National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

Prediction step

$$x_{k|k-1} = M_k^T x_{k-1|k-1}$$

$$P_{k|k-1} = M_k^T P_{k-1|k-1} M_k + Q_k$$

Update step: 1

Calculate the residual ('innovation')

$$\hat{y}_k = H_k x_{k|k-1}$$

$$r_k = y_k - \hat{y}_k = y_k - H_k x_{k|k-1}$$

Innovation uncertainty

$$S_k = H_k^T P_{k|k-1} H_k + R_k$$

where R_k is the measurement (and observation operator) uncertainty.

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

Update step: 2

Optimal Kalman gain

$$K_k = P_{k|k-1} H_k S_k^{-1}$$

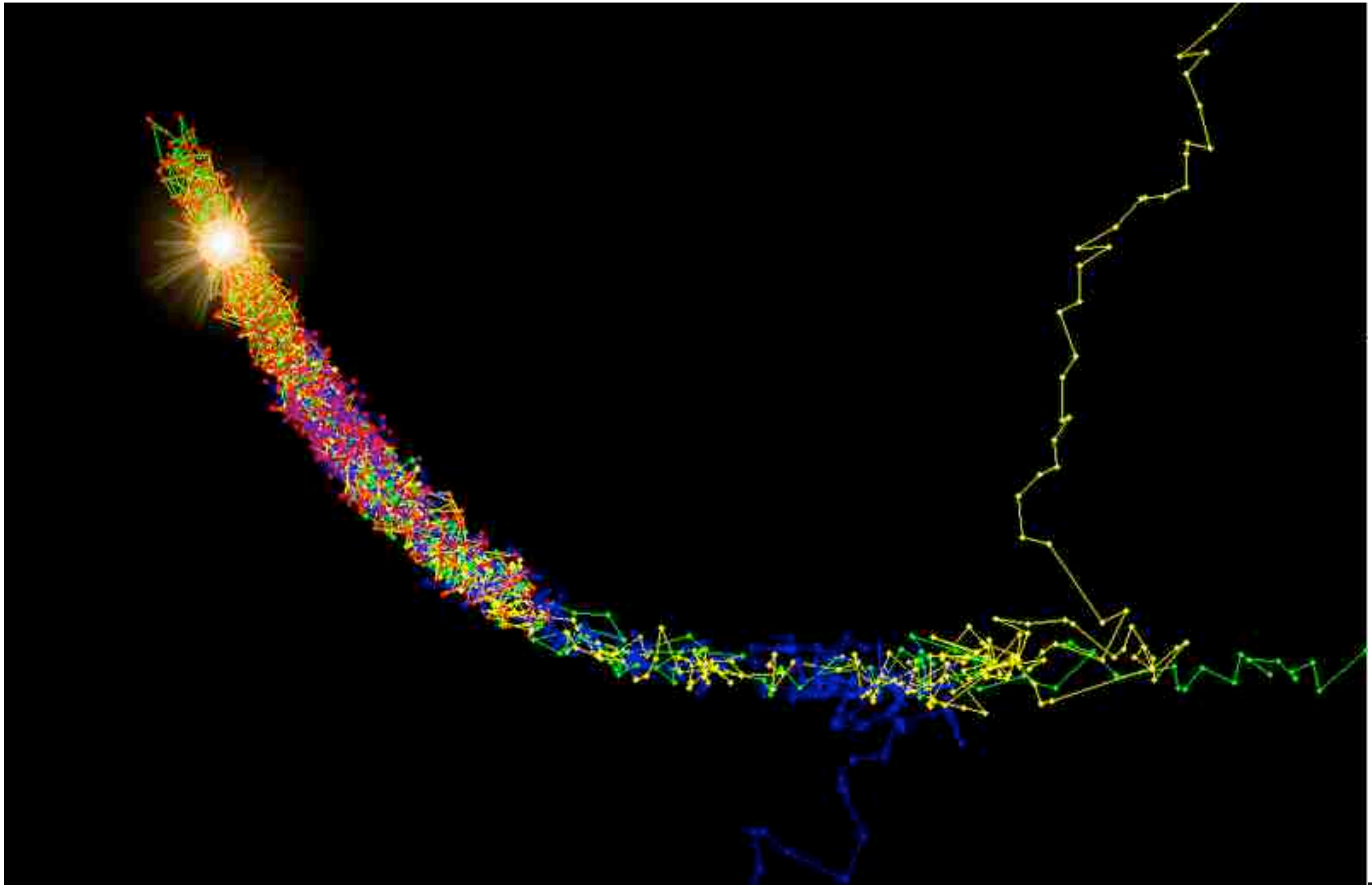
$$x_{k|k} = x_{k|k-1} + K_k r_k$$

$$P_{k|k} = (I - K_k H_k^T) P_{k|k-1}$$

Variants

- **Extended Kalman filter**
  - deal with non-linearity by linearsing the operators
- **Particle filter**
  - Instead of linearising the operators, sample the distributions using Monte Carlo methods
- **Ensemble Kalman filter**
  - Deal with non-linearities by running an ensemble of sample trajectories through the model.
- **MCMC**
  - Use Markov chains to sample space. Clever rules such as MH.
  - MH: propose update of x->x'. Measure improvment *a*
  - Accept is *a>1 or* use *a* as probability of acceptance else

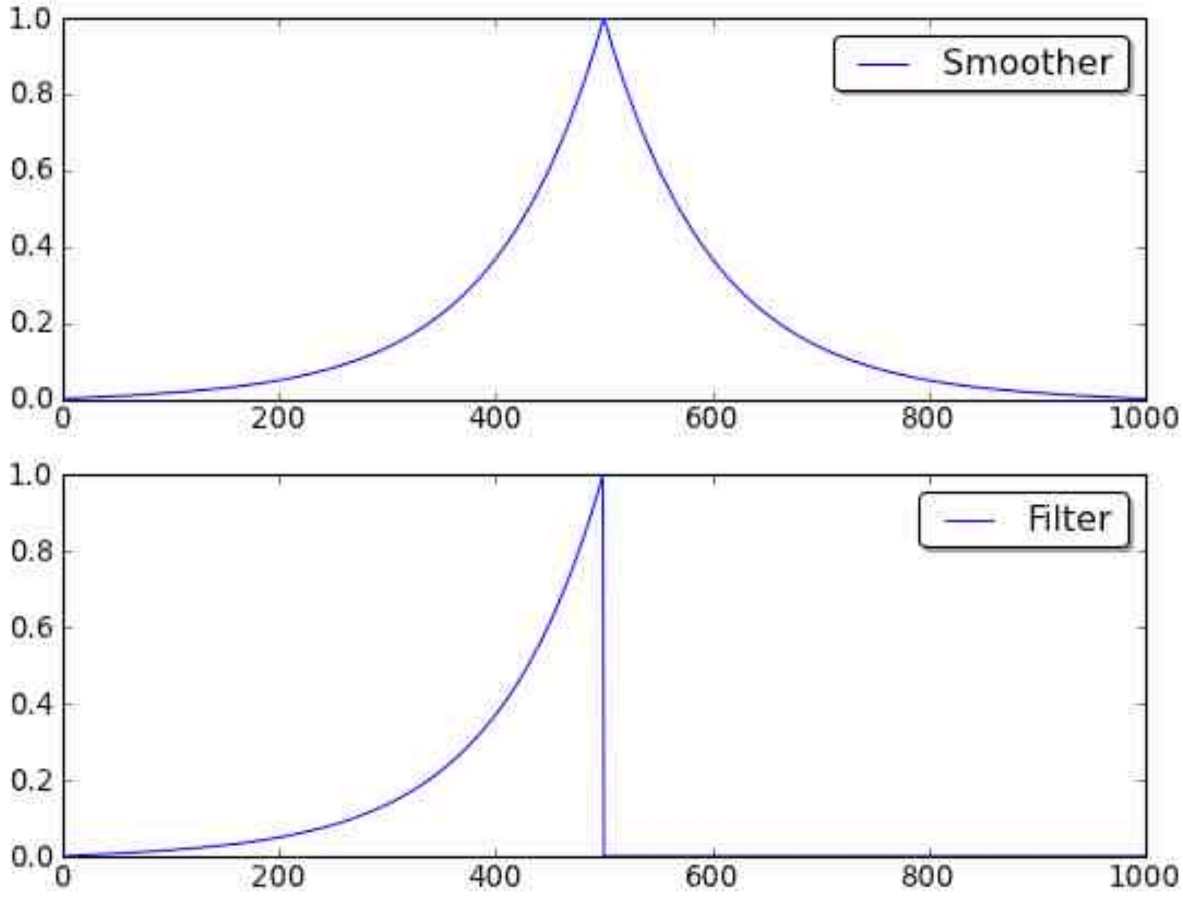
Metropolis Hastings



National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

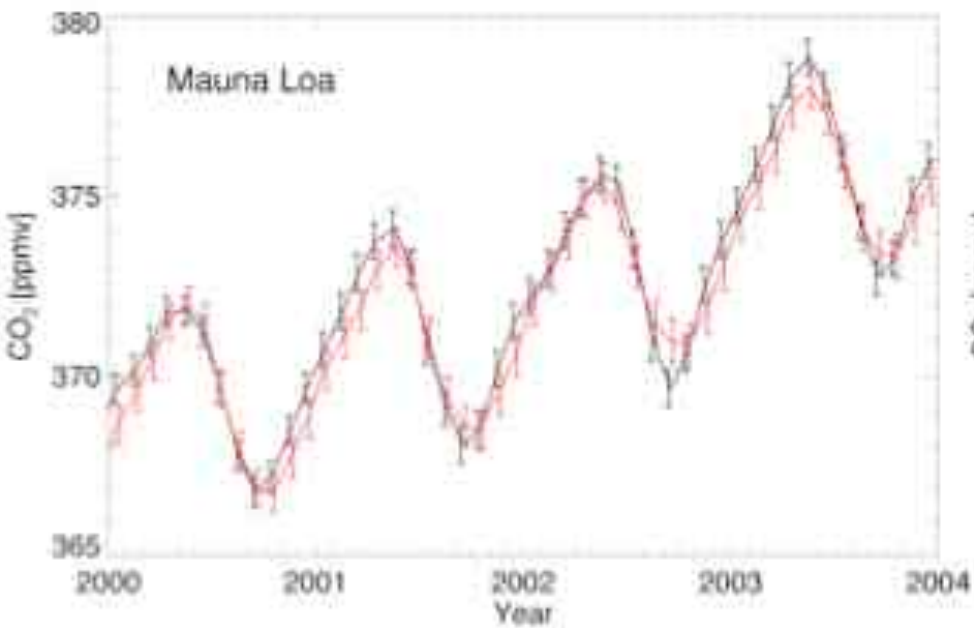
Smoothers and filters

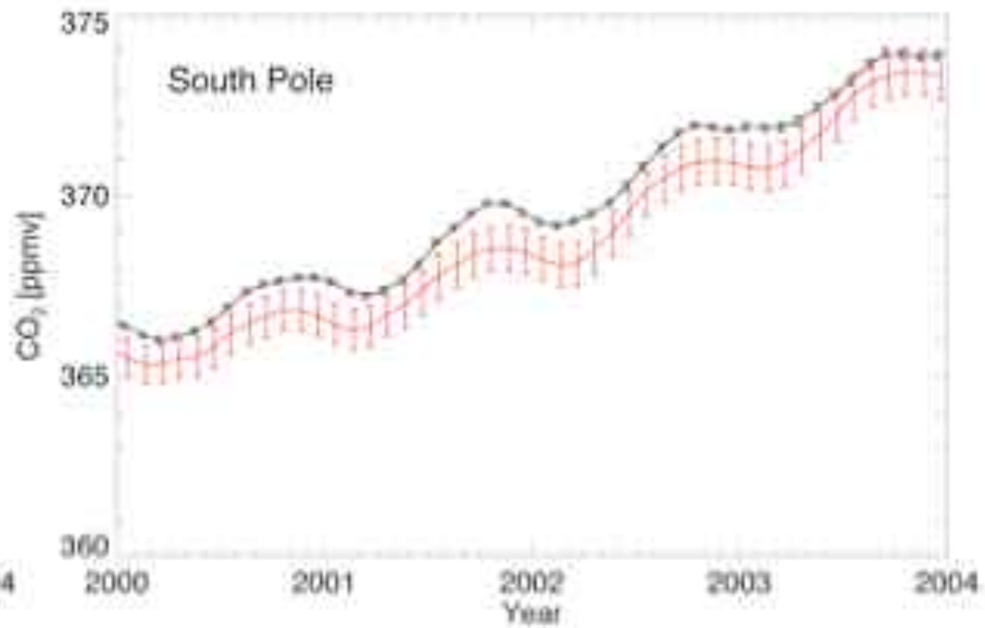


Applications

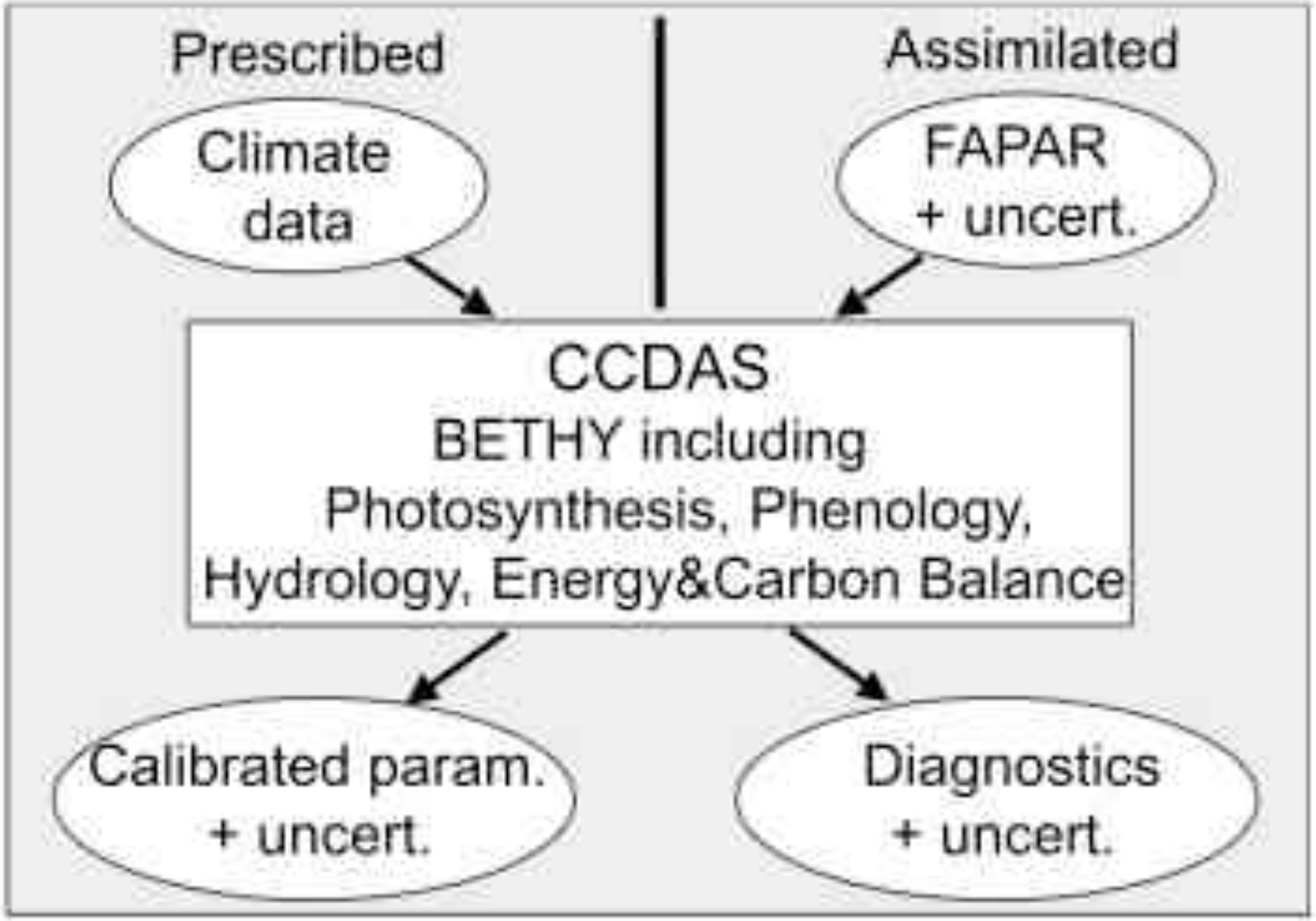
Carbon Cycle Data Assimilation System

- DA system around BETHY model (+adjoint)
- Deal with e.g. atmospheric conc. data. Through coupled atmospheric tracer model





CCDAS



National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

fAPAR assimilation

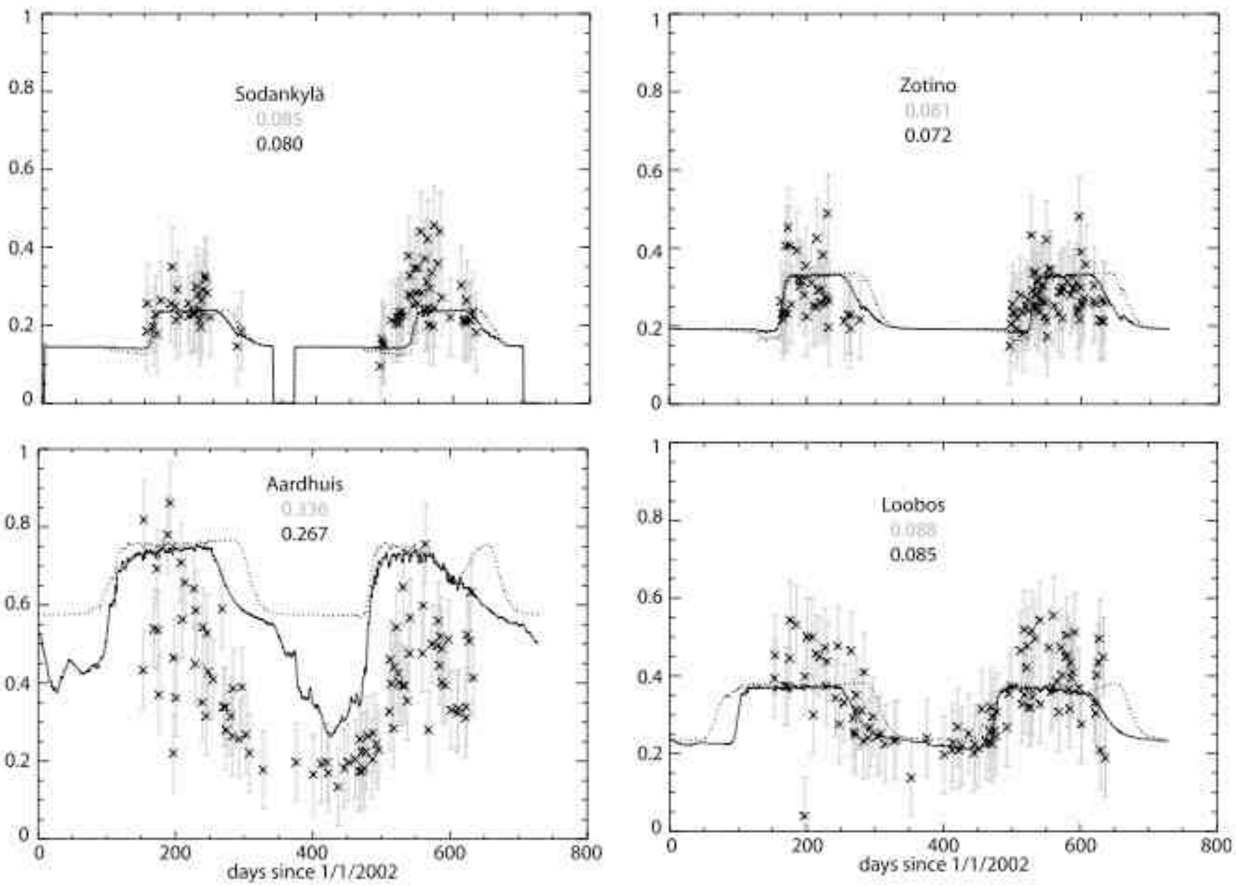


Figure 2. Observed (crosses with uncertainty ranges) and modeled prior (dotted) and posterior (solid line) FAPAR for Sodankylä, Zotino, Aardhuis, and Loobos from north to south. Numbers are root-mean-squared deviation between model and satellite data for the prior (gray) and posterior (black) case.

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

fAPAR assimilation

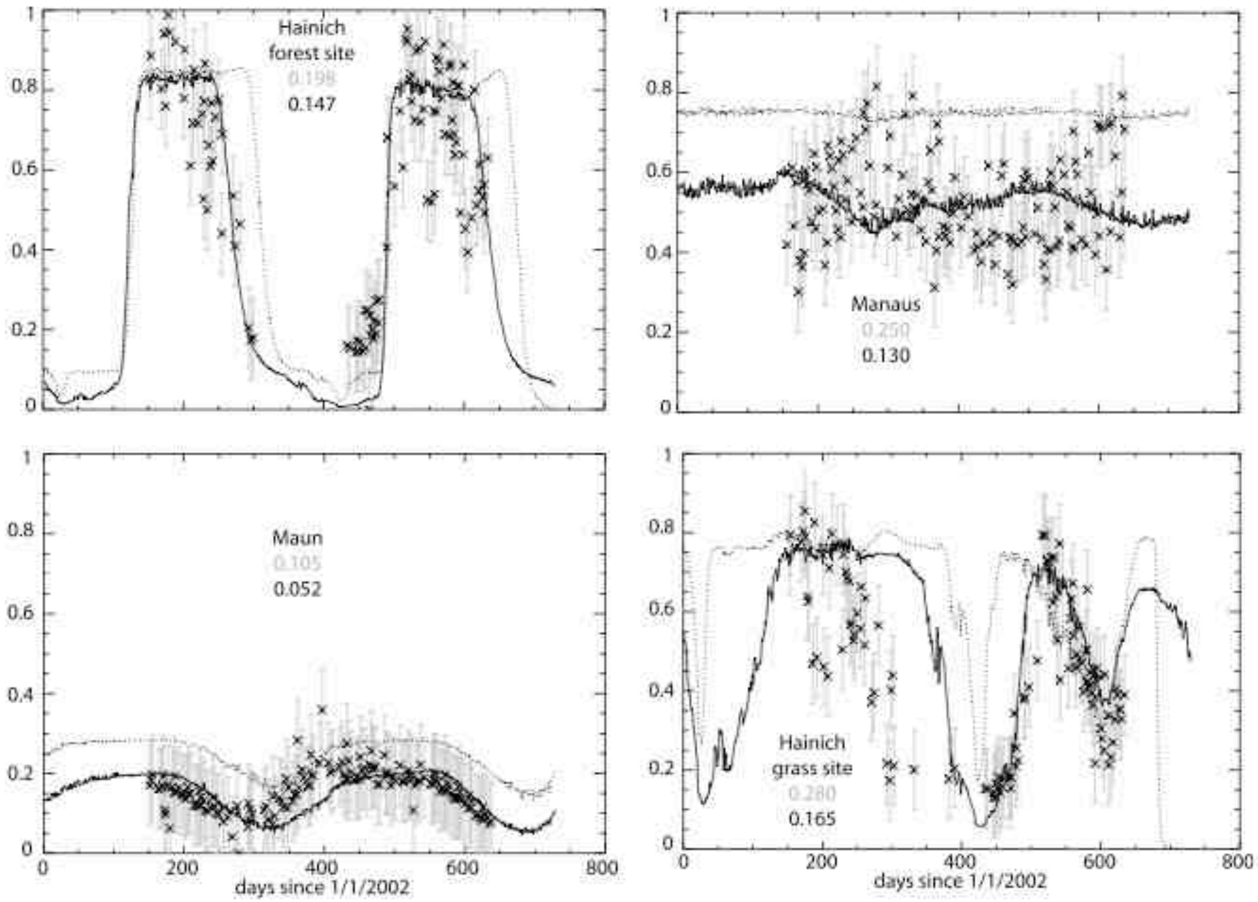


Figure 3. As in Figure 2, but for the Hainich forest site, Manaus, Maun, and the Hainich grass site. The Hainich grass site is shown for validation and not included in the assimilation.

Impact of DA on NPP

Table 7. Mean Annual Prior and Posterior NPP for the Period 2000–2003 (Inclusive) With Uncertainty, Change Relative to Prior Uncertainty, and Relative Uncertainty Reduction[a]

Site	Prior NPP	Posterior NPP	Relative Change (%)	Prior Uncertainty	Posterior Uncertainty	Uncertainty Reduction (%)
Sodankylä	137	151	68	112	98	5
Zotino	201	216	54	28	28	0
Aardhuis	853	842	−7	164	101	38
Loobos	449	424	−40	62	59	5
Hainich forest	689	657	−29	112	98	13
Manaus	1465	964	−196	255	168	34
Maun	350	346	−10	50	46	8
Hainich grass	619	786	97	172	89	48

[a]Units are in gC m^{-2} yr^{-1} or percentage when stated.

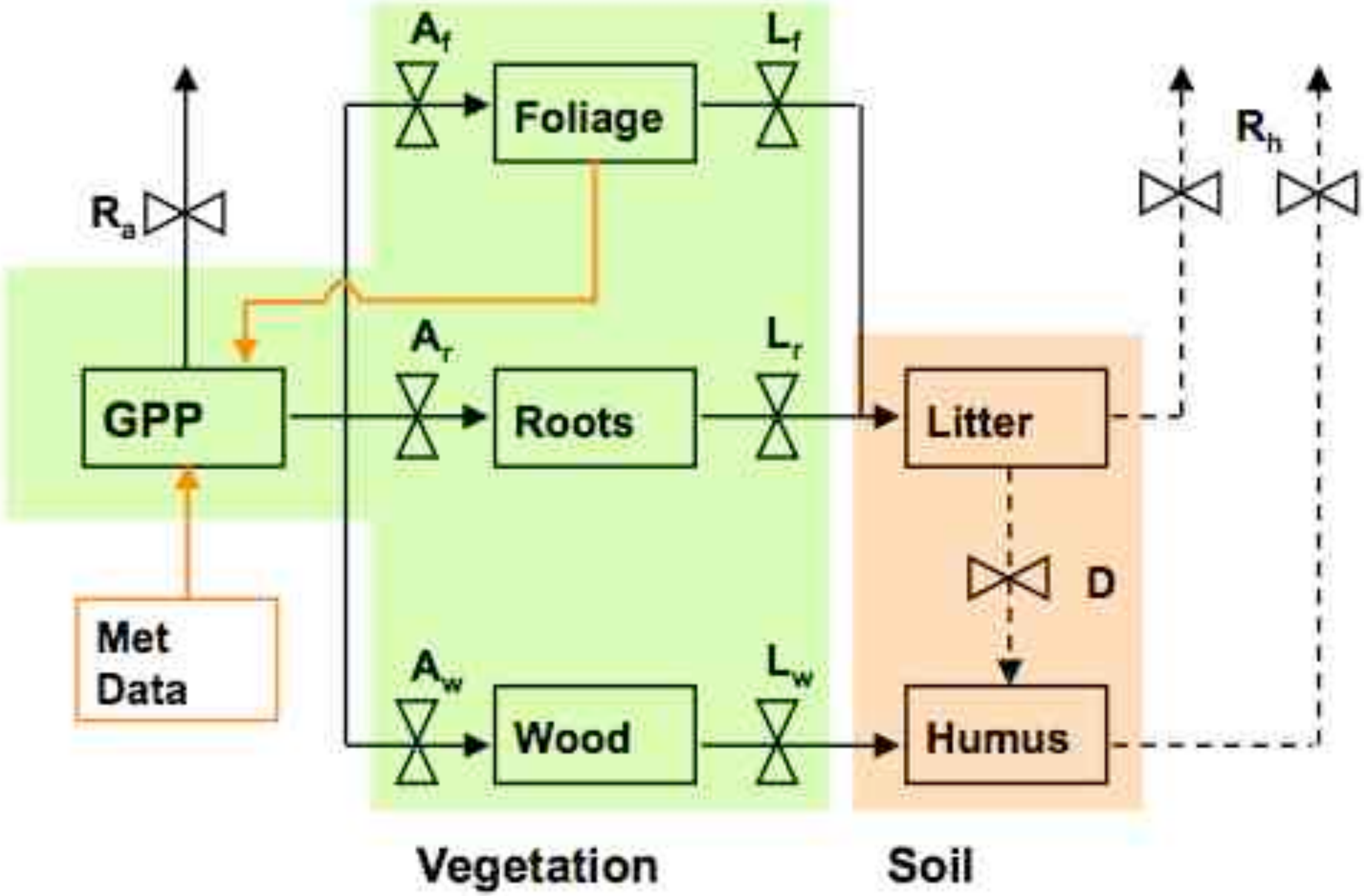
National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

EnKF of surface reflectance data into an ecosystem model

- CCDAS has simple interface to observations
- But difficult to make products (fAPAR) & use in model consistent
- Difficult to track noise in satellite product
- So, Quaife et al. (2008) attempt to link model state (LAI) to EO surface reflectance
    - using NL obs. op.

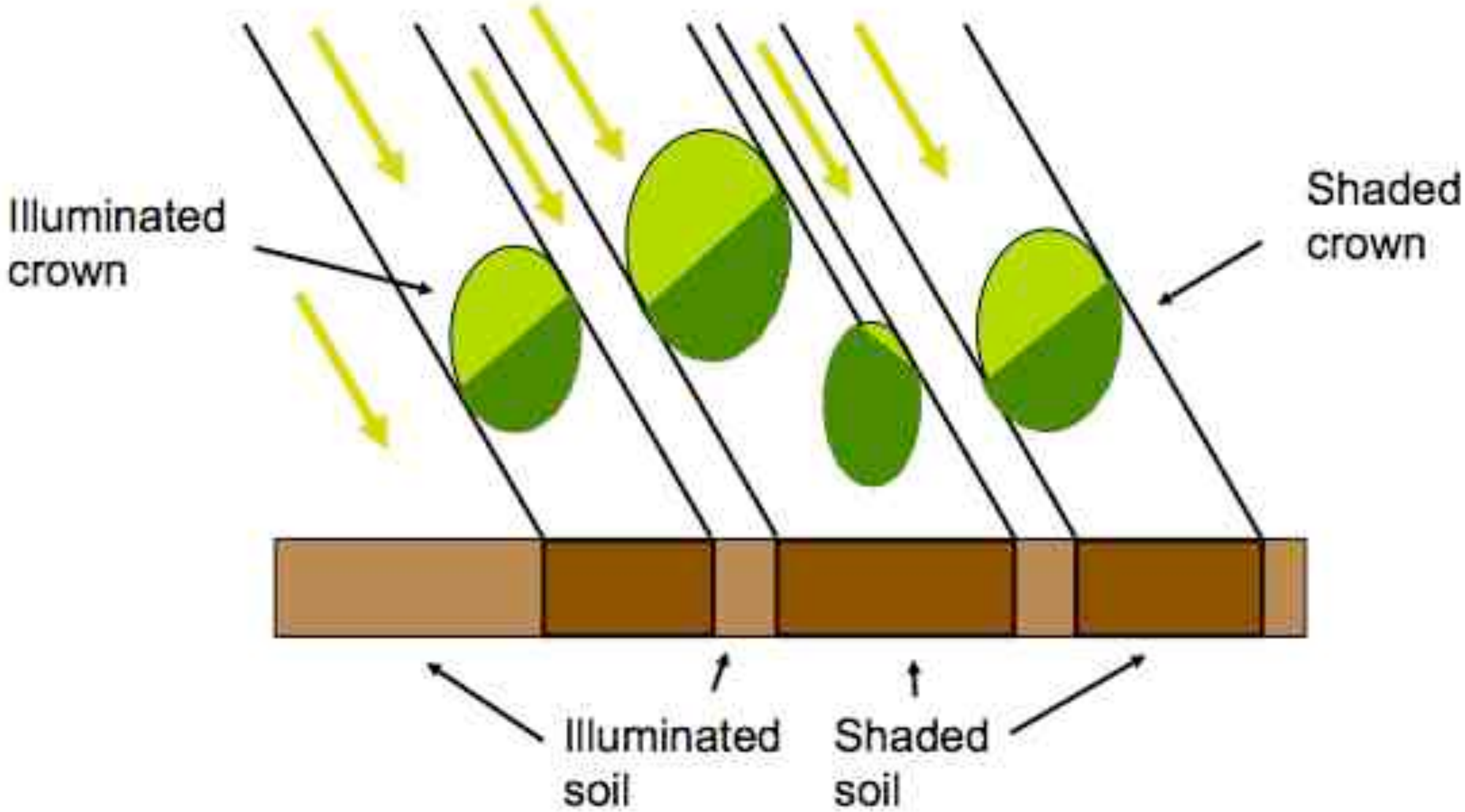
DALEC



National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

Observation operator



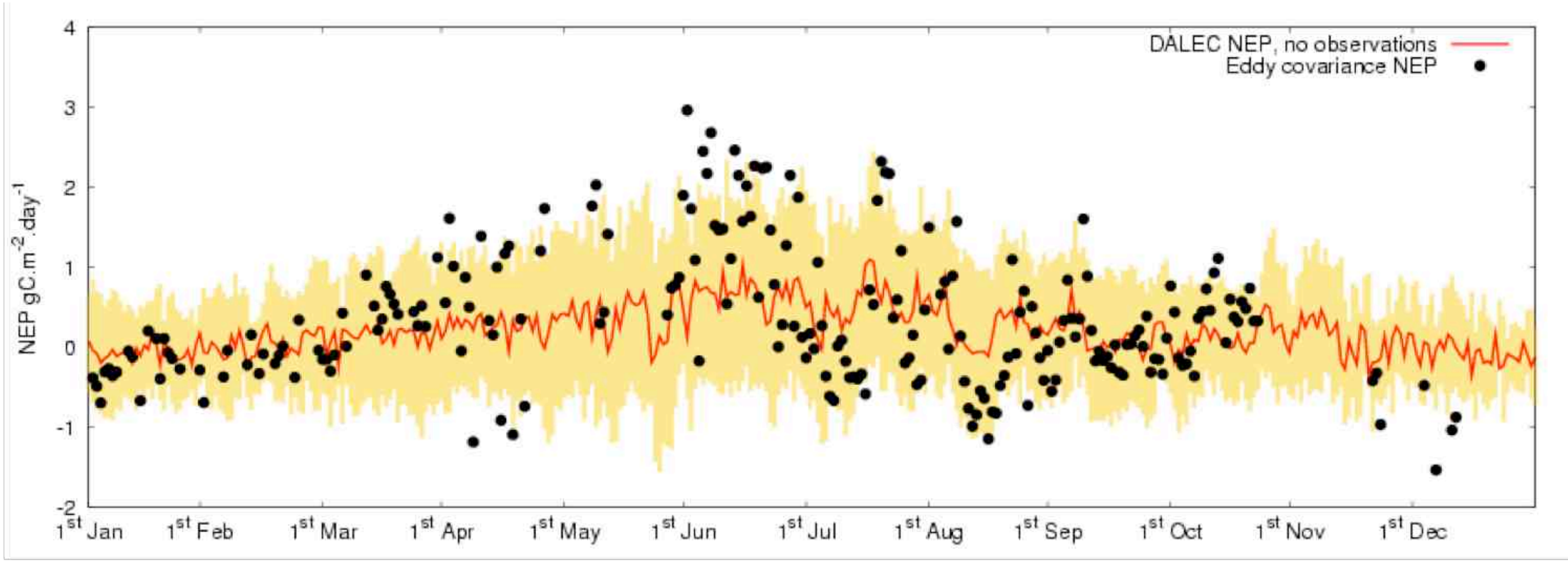
National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

DA

- EnKF
- Used to affect leaf C (LAI)
- EO data fn of ‘ancillary’ terms
  - Leaf chl., water etc.
  - Assumed constant (no uncertainty)

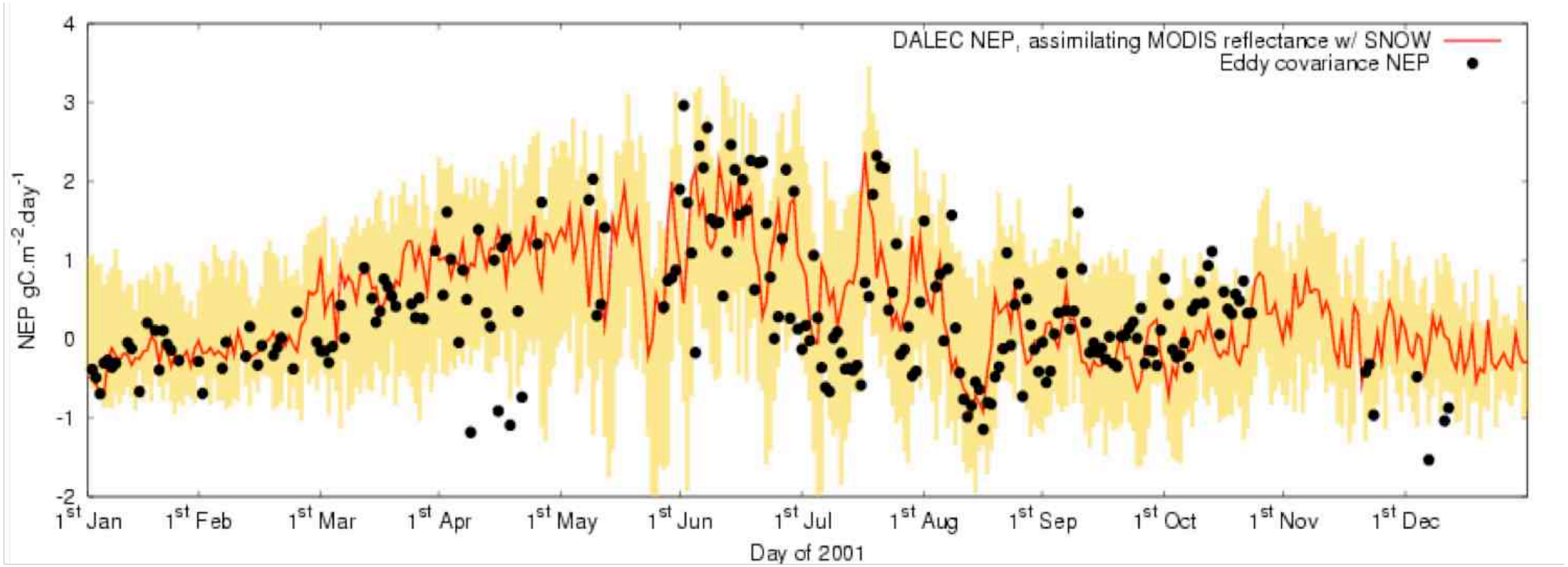
NPP



National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

NPP



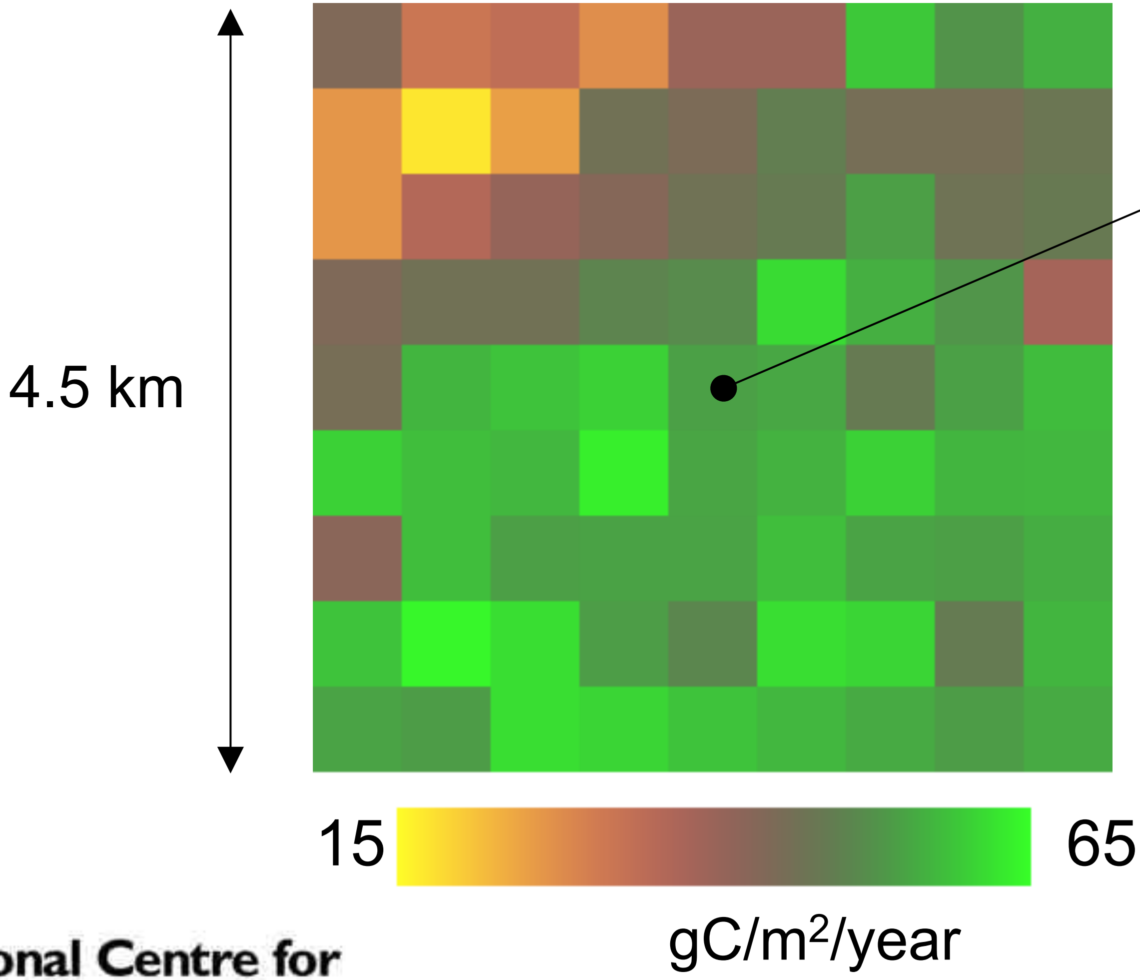
National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

DA

Flux ($gC.m^{-2}$)	Assimilated data	Total	Standard Deviation
NEP	Assimilation exc. snow	373.0	151.3
	Assimilation inc. snow	404.8	129.6
	Williams et al. (2005)	406.0	27.8
GPP	Assimilation exc. snow	2620.3	96.8
	Assimilation inc. snow	2525.6	42.7
	Williams et al. (2005)	2170.3	18.1

Mean NEP for 2000-2002



Spatial average = 50.9

Std. dev. = 9.7

(gC/m^2/year)

National Centre for Earth Observation

NATURAL ENVIRONMENT RESEARCH COUNCIL

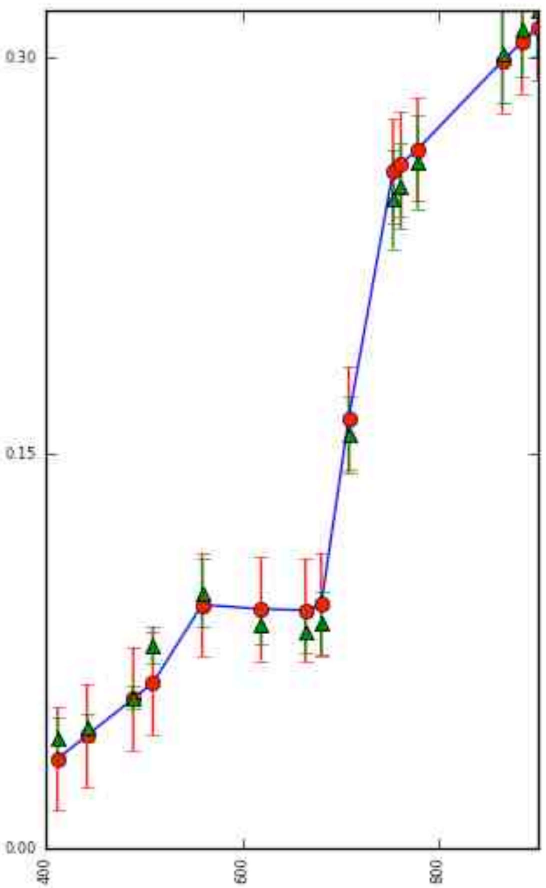
UCL

EO-LDAS

Earth Observation Land Data Assimilation

- Focus on EO
- Simple models only (regularisers, Dx, D^2x …)
- Observation operator
  - Semi-discrete (Gobron) + soil + ~PROSPECT
  - + adjoint!

Simple DA: weak prior + MERIS observation



Reduces uncertainty

Models observation well

Solve for 7 biophysical parameters here

Uncertainty on most are high

From single meris observation

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

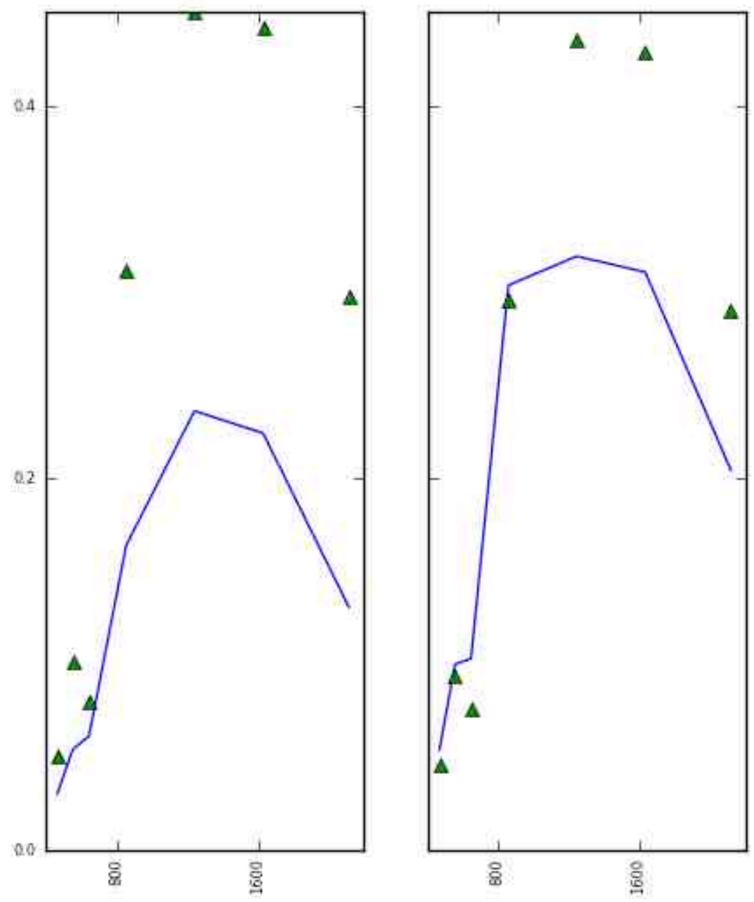
UCL

Predict MODIS

from MERIS parameters

Poor performance in extrapolation

Due to correlations and large uncertainties



National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

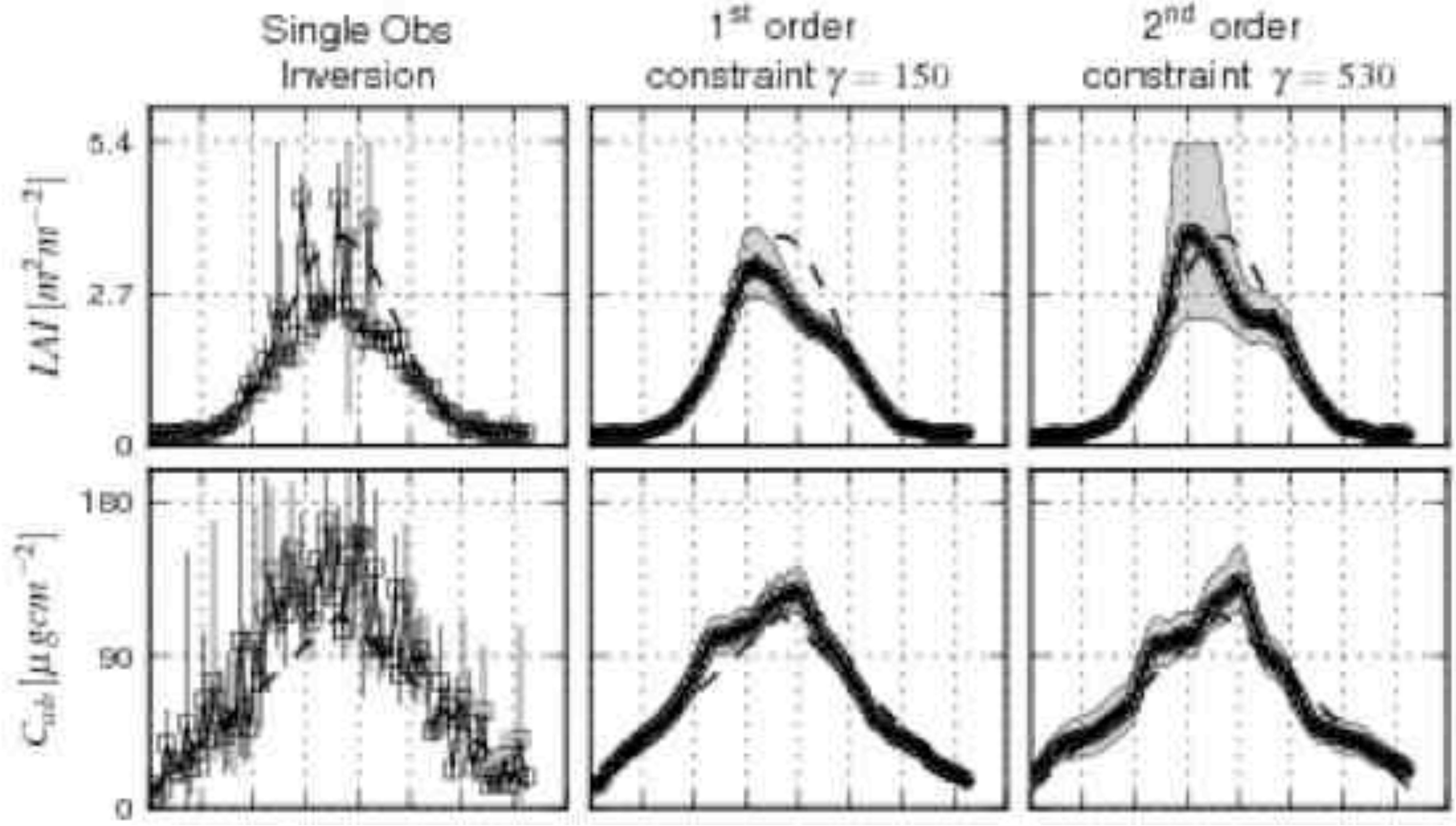
Sentinel-2 MSI

Resolution Summary	10m for 4 bands in VNIR, 60 m for 3 dedicated atmospheric correction bands, 20 m for remaining bands *[Best Resolution: 10m]*
Swath Summary	290 km *[Max Swath: 290 km]*
Accuracy Summary	Absolute radiometric accuracy for Level 1C data: 3-5%
Waveband Summary	13 bands in the VNIR-SWIR *VIS (~0.40 μm - ~0.75 μm)* *SWIR (~1.3 μm - ~3.0 μm)*

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

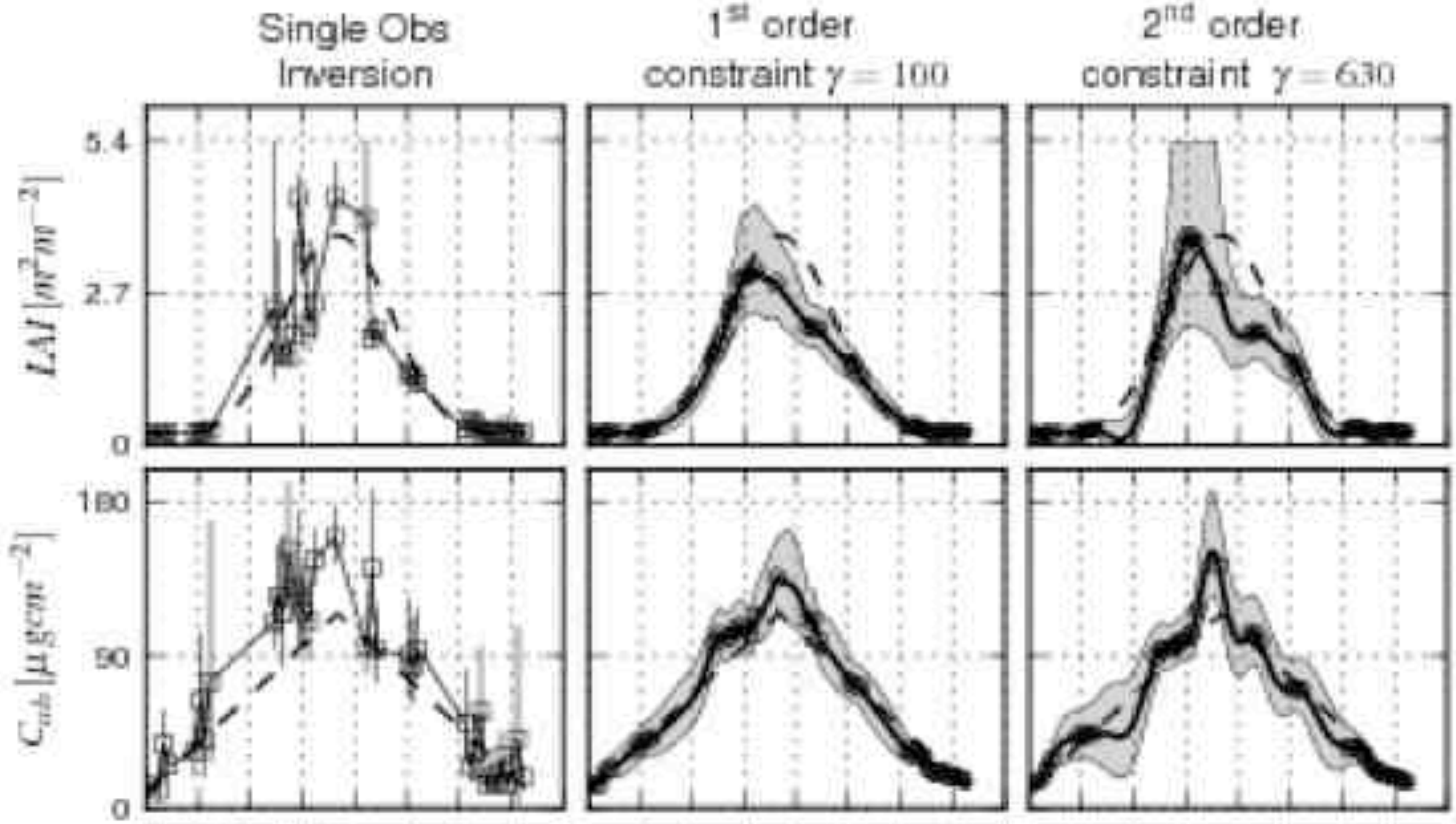
Use Dx, D^2x to constrain



National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

Cloudy case



National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCL

Discussion

- Outlined what DA is
- Presented basic stats
  - How to combine uncertainties
  - Leads to how to combine constraints
- Reviewed some major DA approaches
  - Variational / sequential
  - Weak constraint / strong constraint
  - Filters / smoothers

Discussion

- Applications
  - CCDAS
    - Simple obs. Op.,
    - concentrates more on model (& atmos. transport)
  - Quaife et al
    - Connected refl. data to ecological model
    - Try to make canopy assumptions in M and H consistent
      - But only partially achieved this
  - EO-LDAS
    - Concentrate more on EO (H)
    - Variational system
    - Demonstrates use of regularisation

Discussion

State of the art in combining models/data

- Really quite basic
  - NDVI to calibrate phenology
- DA (CCDAS) shows that can be achieved in much more consistent / scientific manner
  - Incorporate uncertainty … DA
- DA (Quaife/EO-LDAS) show potential for bypassing ‘products’ from EO
  - Directly link observations
  - Many advantages (multiple constraints)
  - But a bit slow/complex now
- “In the future” this is how EO data will be used … (I hope!!)