Options for data assimilation

Tristan Quaife

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

COST 734 WG 2.1 Workshop
Salford UK, January 2010

UNIVERSITY OF EXETER

What is Data Assimilation?

- A working definition:

  The set techniques the combine data with some underlying process model to provide optimal estimates of the true state and/or parameters of that model.

What does DA aim to do?

- Use all available information about
  - The underlying model
  - The observations
  - The observation *operator*
- Including estimates of uncertainty and the current state of the system
- To provide a best estimate of the true state of the system with quantified uncertainty

Two questions

The purpose of this presentation is to explore the following:

- What DA technique to use?
- What observations to assimilate?

Nominal classification of DA

- Sequential
- Variational

Sequential methods

- Kalman Filter
  - Variants - EKF
- Ensemble Kalman Filter
  - Variants – Unscented EnKF
- Particle filters
  - Lots of different types
  - true MCMC technique

The Kalman filter

- Propagation step:

$$x = Mx^{-}$$

$$P = P^{-}MP^{-T} + Q$$

- Analysis step:

$$x^{*} = x + K(y - Hx)$$

$$K = PH^{T}(HPH^{T}+R)^{-1}$$

Observation
Model
matrix

The Kalman Filter

- Linear process model
- Linear observation operator
- Assumes normally distributed errors

The Ensemble Kalman filter

- Propagation step:

$$X = m(X^-) + Q$$

no explicit error propagation

- Analysis step:

$$X^* = X + K(D - HX)$$

$$K = PH^T(HPH^T + R)^{-1}$$

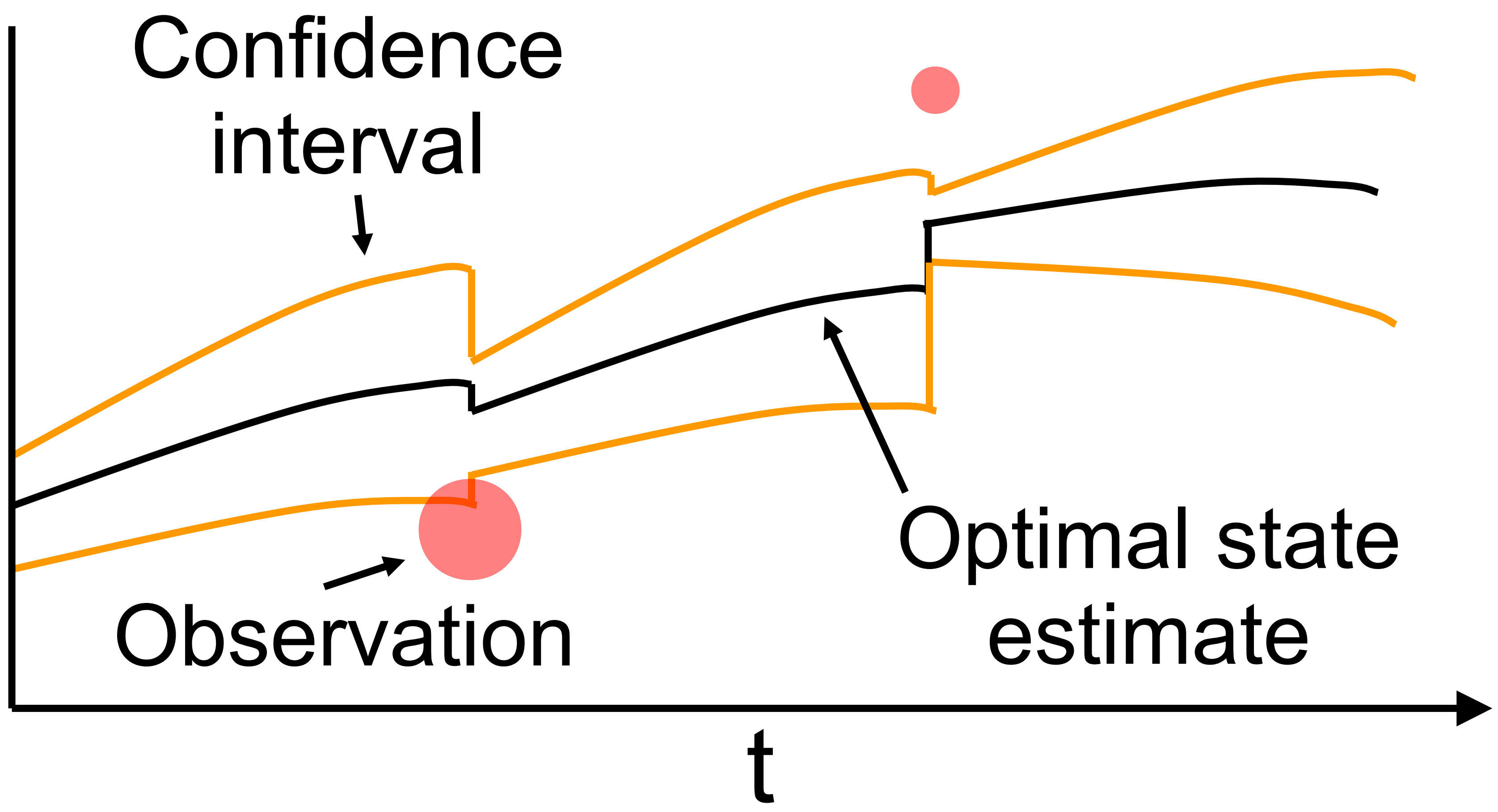
State vector ensemble

Perturbed observations

Sequential techniques

- Designed for real time systems
- Only consider historical observations
- Only assimilates observations in single time step
- Can lead to artificial high frequency components

Sequential techniques



Variational techniques

- Expressed as a cost function
- Uses numerical minimisation
- Gradient descent requires **differentiated model**
- Traditionally used for initial conditions
  - But parameters may also be adjusted

Variational techniques

Background

$$J(x) = (x - x^{-})\, P^{-1}\, (x - x^{-})^{T} \; + \; (y - h(x))\, R^{-1}\, (y - h(x))^{T}$$

Observations

Variational techniques

$$J(x) = (x - x^{-})\, P^{-1} (x - x^{-})^{T} +$$

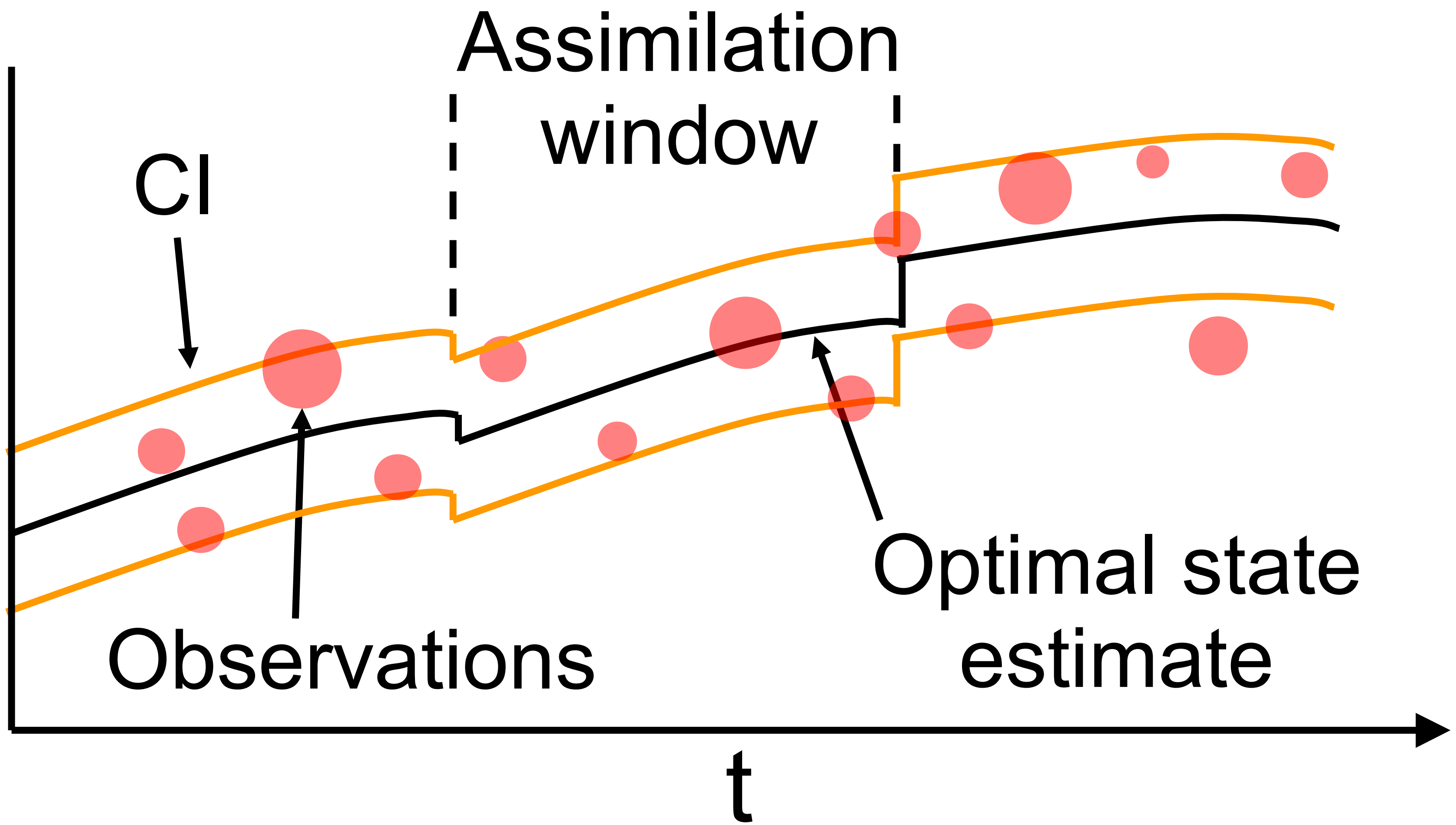
$$\sum_i (y - h(x_i))\, R^{-1} (y - h(x_i))^{T}$$

Time varying state vector

Variational techniques

- Parameters constant over time window
- Non smooth transitions
- Size of time window?
- For zero-order case 3D = 4D
- Absence of **Q** - propagation of **P**?

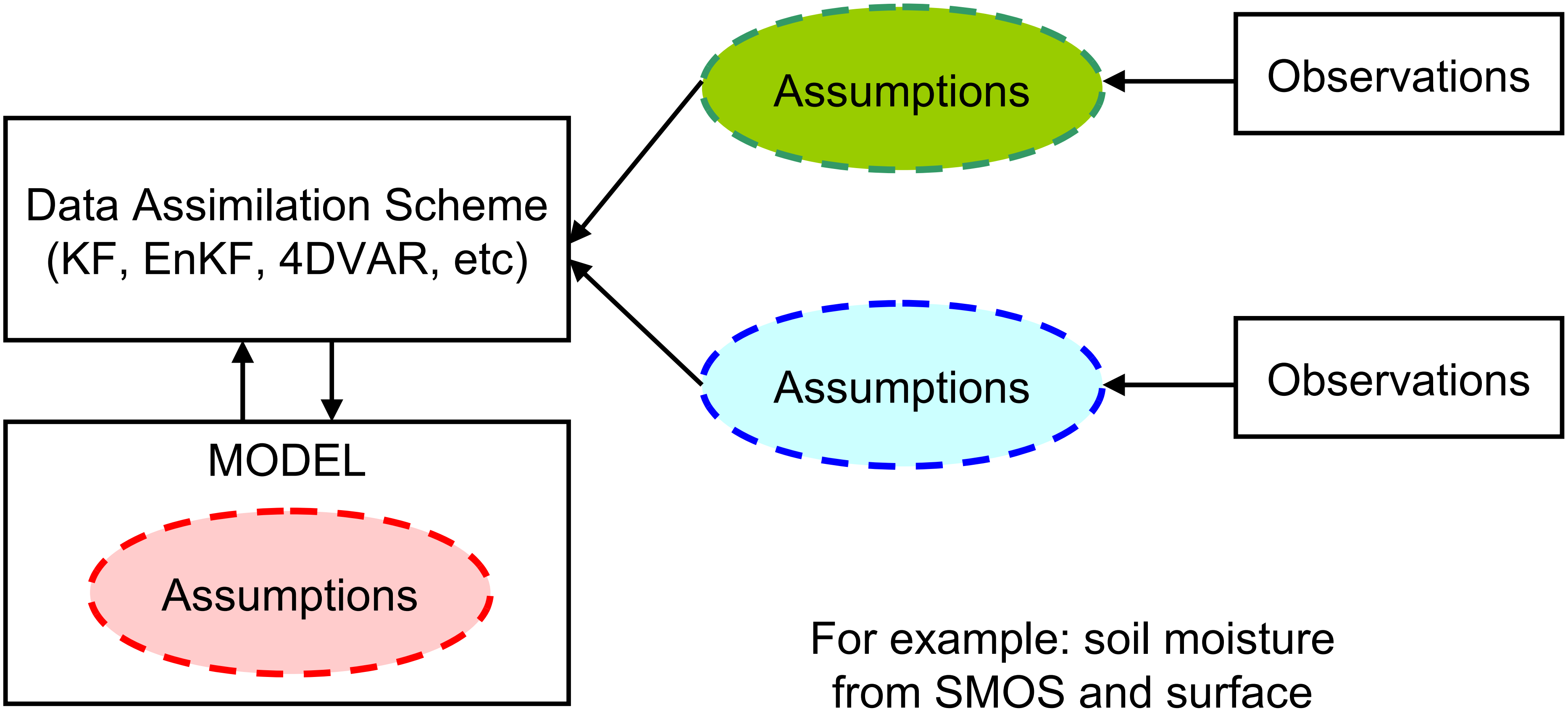
Variational techniques



Observations

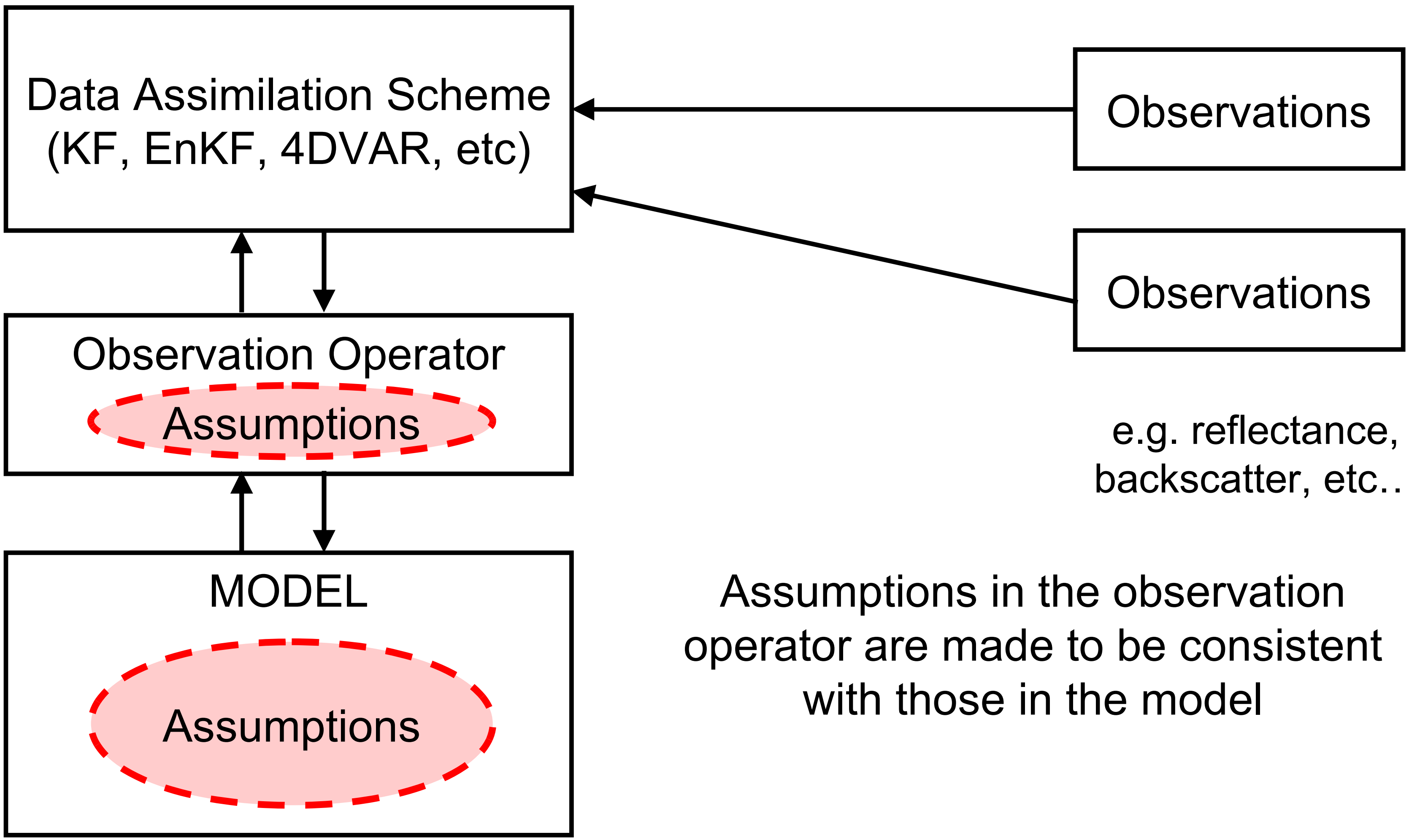
- High level products (e.g. LAI)
  - Easy to assimilate
  - Often difficult to quantify uncertainty
  - Assumptions inconsistent with model
- Low level products (e.g. reflectance)
  - More complex to utilise
  - Uncertainties better understood

Assimilating products



For example: soil moisture from SMOS and surface temperature from MODIS

Assimilating reflectance



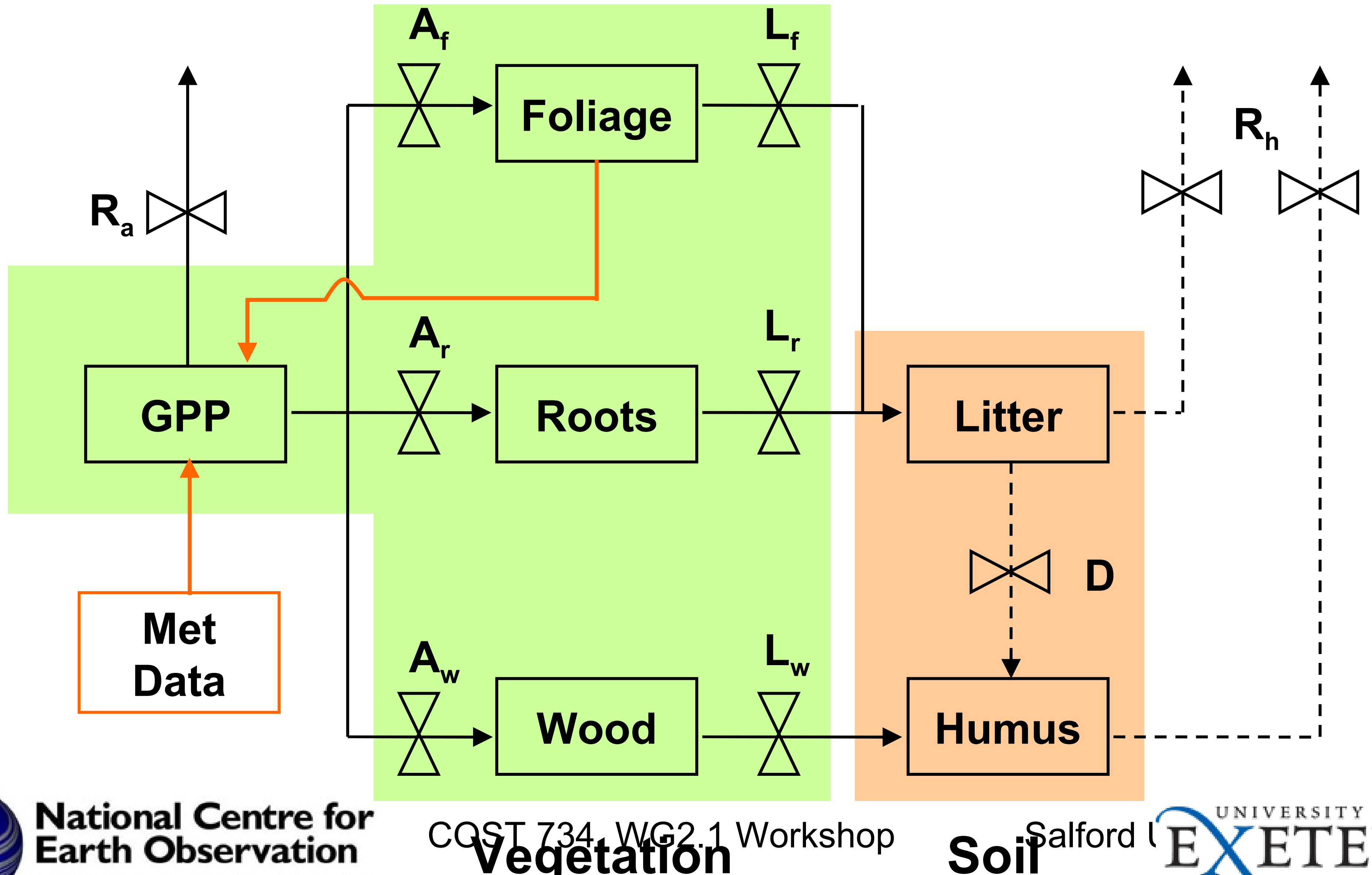
Assumptions in the observation operator are made to be consistent with those in the model

T. Quaife, P. Lewis, M. DE Kauwe, M. Williams, B. Law, M. Disney, P. Bowyer (2008), Assimilating Canopy Reflectance data into an Ecosystem Model with an Ensemble Kalman Filter, Remote Sensing of Environment, 112(4),1347-1364.

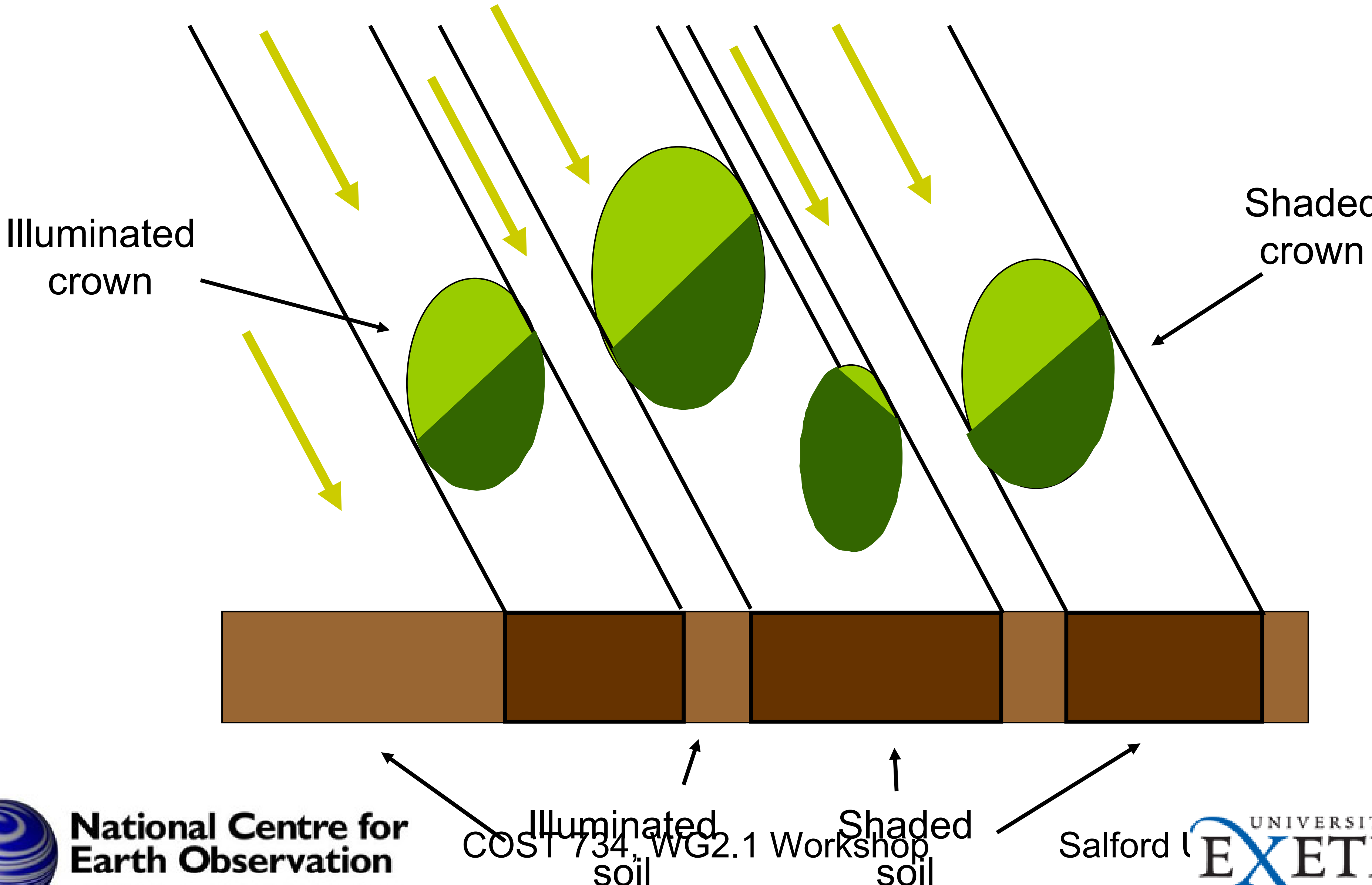
Conclusions

- Not interested in real time applications so variational techniques offer more flexibility.
- Assimilating high-level products (LAI etc) offers a simple and pragmatic way forward.
- Assimilating "raw" observations will ultimately provide better results but may be an unrealistic goal for many projects.

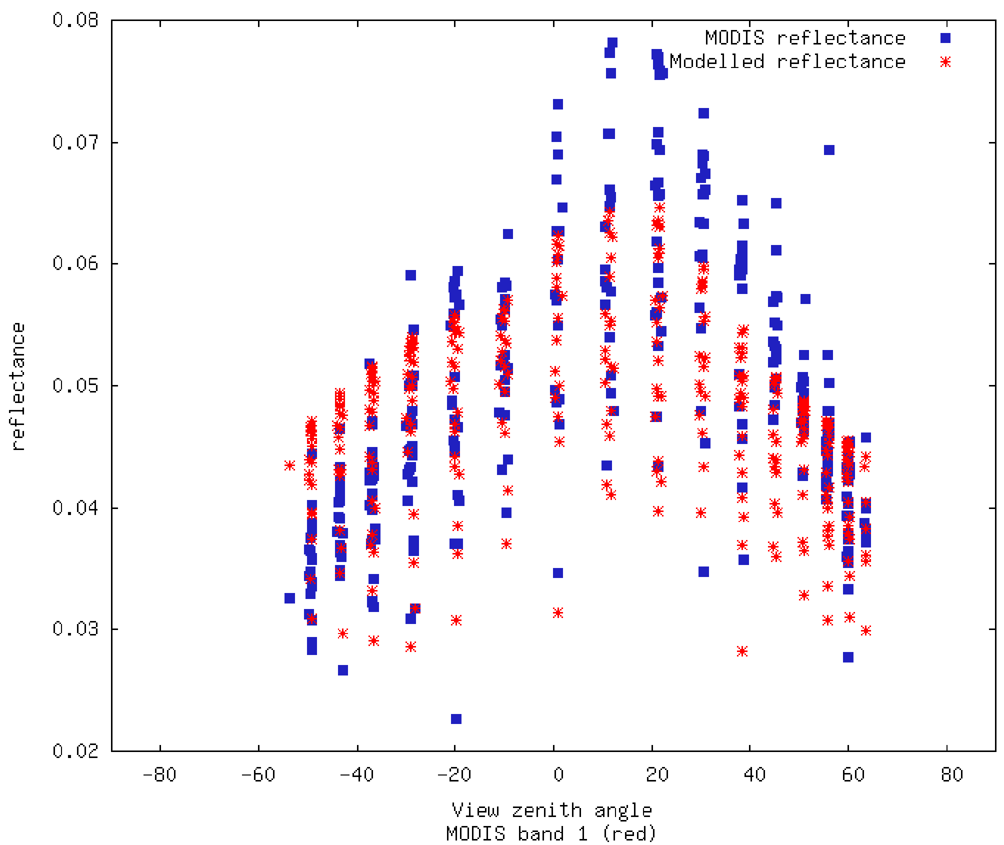
DALEC – ecosystem model



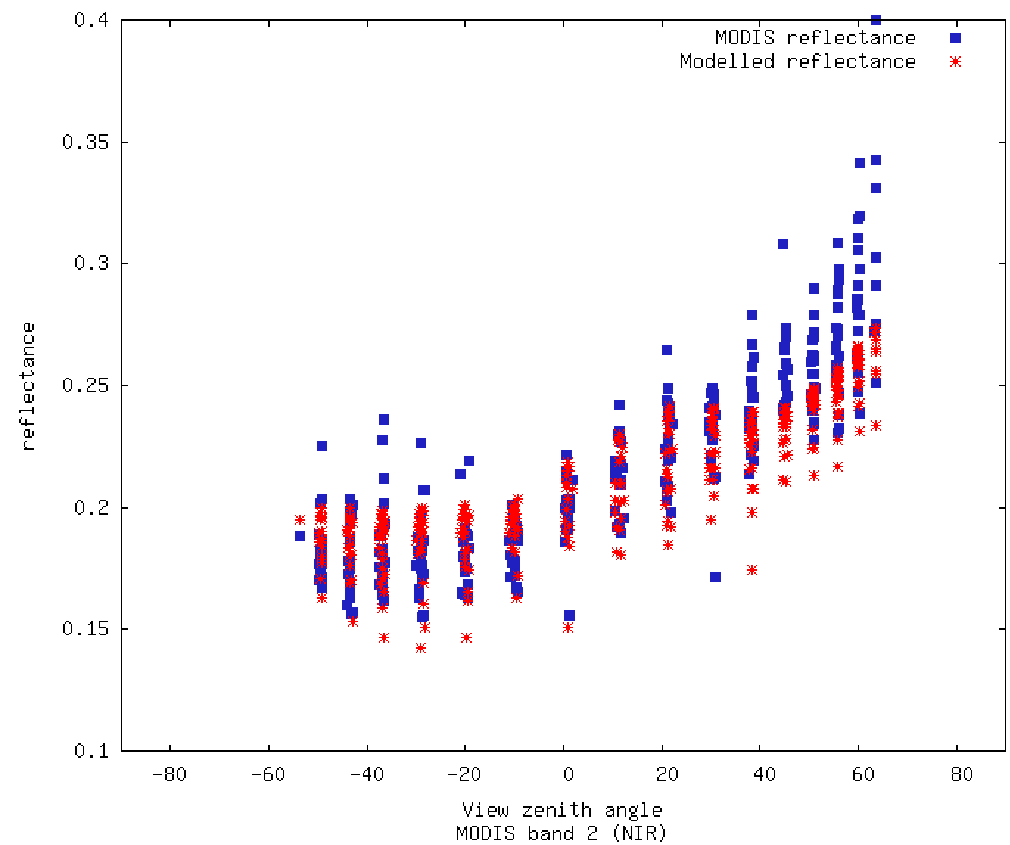
Observation operator - GORT



Modelled vs. observed reflectance



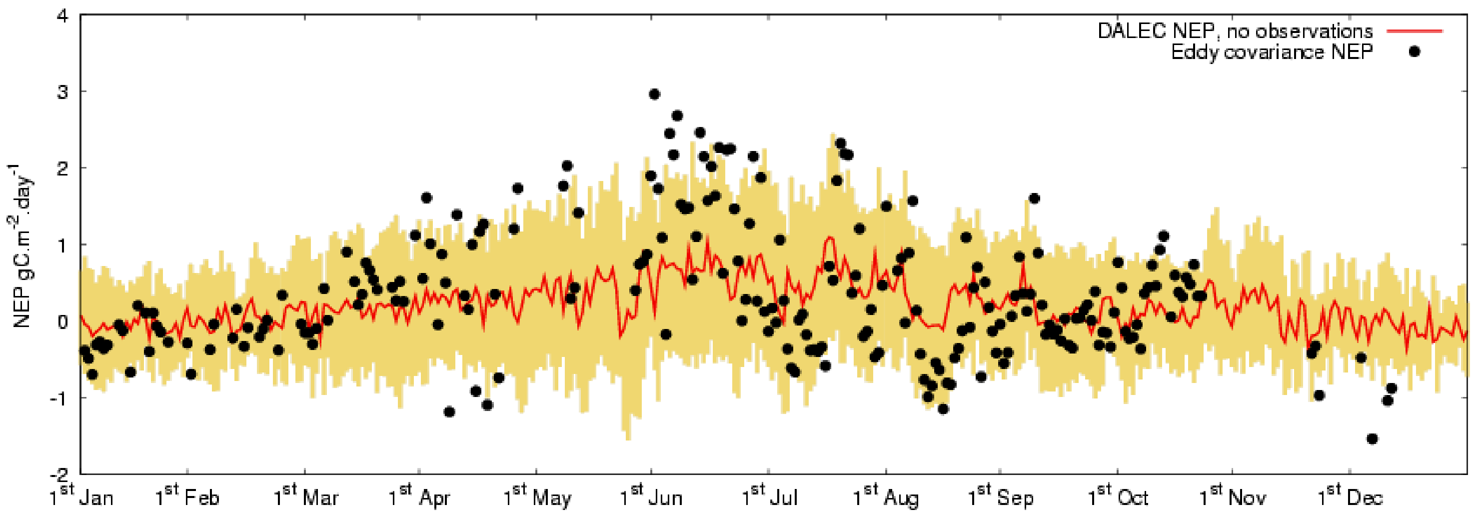
Band 1



Band 2

Net Ecosystem Productivity

No assimilation



Assimilating MODIS surface reflectance bands 1 and 2