Introduction to Data Assimilation

J Gómez-Dans

\(^1\)NCEO & Dept. of Geography, University College London (UK)

March 2014
Outline

1. What is data assimilation (DA)?
2. From Bayes’ Rule to DA
   - Assessing the update
3. DA methods
   - Variational schemes
   - Sequential methods: Filters & Smoothers
   - MCMC
4. Some DA examples
   - CCDAS
   - EnKF & DALEC model
   - JRC-TIP
   - EO-LDAS
5. Discussion
1. What is data assimilation (DA)?

2. From Bayes’ Rule to DA
   - Assessing the update

3. DA methods
   - Variational schemes
   - Sequential methods: Filters & Smoothers
   - MCMC

4. Some DA examples
   - CCDAS
   - EnKF & DALEC model
   - JRC-TIP
   - EO-LDAS

5. Discussion
We have seen *models* and *observations*

**Models** Can be wrong (missing processes, inadequate parametrisation, uncertainty in drivers, ...)

**Observation** Have uncertainty, may be missing or might not reflect the quantity of interest directly.

DA provides an statistical framework that optimally combines *observations* and *models*, plus all the associated *uncertainties* to infer the most likely state of the land surface, plus uncertainties.
1. What is data assimilation (DA)?

2. From Bayes’ Rule to DA
   - Assessing the update

3. DA methods
   - Variational schemes
   - Sequential methods: Filters & Smoothers
   - MCMC

4. Some DA examples
   - CCDAS
   - EnKF & DALEC model
   - JRC-TIP
   - EO-LDAS

5. Discussion
Typically, the state vector $\vec{x}$ (LAI, soil moisture...) is not directly observable.

Need to map from $\vec{x}$ to the observations (e.g. backscatter, reflectance, temperature brightness) using an Observation Operator $\mathcal{H}(\vec{x})$.

Observations $\vec{y}$ are corrupted by additive Gaussian noise $\epsilon_{obs}$.

The measurement equation is then simply:

$$\mathcal{H}(\vec{x}, I) = \vec{y} + \epsilon_{obs}$$ (1)

We want to estimate the $\vec{x}$ based on the data. How do we go about that?
A way to find the solution to the measurement equation is to find a set of parameters $\vec{x}_0$ that minimise the difference between predicted observation and actual observation, modulated by the uncertainty in the observations:

$$\|H(\vec{x}, I) - \vec{y}\|^2 \leq \epsilon_{obs}$$ (2)

This means that there are infinite solutions that meet the criterion in Eq. 2! The problem is said to be ill posed. The only practical way to solve this underdetermination is by adding extra constraints.
We can’t avoid ill posedness. One way around it is to assume we no longer have discrete parameters, but parameter distributions. The distribution encompasses our current belief in the value of the parameter.

Our problem is now how to express \( p(\vec{x}|\vec{y}) \), the probability of the state \( \vec{x} \) conditioned on the observations, \( \vec{y} \).
Bayes’ Rule

\[
p(\bar{x}|y) = \frac{p(y|\bar{x}) \cdot p(\bar{x})}{p(y)}
\]

The likelihood is derived from the mismatch of the model to the observations using an observation operator.

The prior is a pdf of any knowledge we might have about the state before the observations are considered.
Try to infer $x$ from $y$, using an identity observation operator (i.e., $x = y$) and Gaussian noise:

$$p(y|x) = \frac{1}{\sqrt{2\pi\sigma_{\text{obs}}}} \exp \left[ -\frac{1}{2} \frac{(x - y)^2}{\sigma_{\text{obs}}^2} \right]. \quad \text{Likelihood} \quad (4)$$

Assume that we only know that $x$ is Gaussian distributed with $\mu_p$ and $\sigma_0$:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_0}} \exp \left[ -\frac{1}{2} \frac{(x - \mu_p)^2}{\sigma_0^2} \right] \quad \text{Prior} \quad (5)$$

$$p(x|y) \propto \frac{1}{\sigma_{\text{obs}}\sigma_0} \exp \left[ -\frac{1}{2} \frac{(x - \mu_p)^2}{\sigma_0^2} \right] \cdot \exp \left[ -\frac{1}{2} \frac{(x - y)^2}{\sigma_{\text{obs}}^2} \right]. \quad \text{Posterior} \quad (6)$$
The posterior distribution is indeed a Gaussian, and its mean and std dev can be expressed as an update on the prior values:

\[ \mu_p = \mu_0 + \frac{\sigma_0^2}{\sigma_0^2 + \sigma_{obs}^2}(y - \mu_0). \]  \hspace{1cm} (7)

\[ \sigma_p^2 = \sigma_0^2 \cdot \left(1 - \frac{\sigma_0^2}{\sigma_0^2 + \sigma_{obs}^2}\right), \] \hspace{1cm} (8)
Univariate example (III)

\[ p(x) \sim \mathcal{N}(\mu = -5, \sigma^2 = 1) \quad p(y|x) \sim \mathcal{N}(0, 4) \quad p(x|y) \sim \mathcal{N}(-4, 2/5) \]

**Figure:** The prior (red), the likelihood (yellow) and the posterior (green) for the simple 1D case. The black lines indicates the maximum \textit{a posteriori}.
\( \vec{x} \) (the state) is a vector, rather than a scalar as above. The Gaussian pdf for \( \vec{x} \) is given by

\[
P_b(\vec{x}) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(B)}} \exp \left[ -\frac{1}{2} (\vec{x}_b - \vec{x})^\top B^{-1} (\vec{x}_b - \vec{x}) \right],
\]

where \( n \) is the size of \( \vec{x} \), \( \vec{x}_b \) is the mean vector, \( B \) is the covariance matrix and \( ^\top \) indicates transposition. \( B \) indicates the covariance between the elements of \( \vec{x} \). For a two dimensional case, we have

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

(10)

Again, posterior is Gaussian if prior and likelihood are Gaussian and observation operator linear.
Multivariate example (II)

\[ \vec{x} = [0.2, 0.5]^\top \quad \vec{x} = [-0.5, -0.2]^\top. \]

Covariance is

\[ B = \begin{bmatrix}
0.3^2 & \pm 0.5 \times 0.3 \times 0.2 \\
\pm 0.5 \times 0.3 \times 0.2 & 0.2^2
\end{bmatrix}, \] (11)
Since the posterior is Gaussian, we can write update equations:

$$
\vec{x}' = \vec{x}_b + BH^\top \left( H BH^\top + R \right)^{-1} (\vec{y} - H\vec{x})
$$

$$
= \vec{x}_b + K \cdot (\vec{y} - H\vec{x}).
$$

(12)

$$
Q = (I - KH) B.
$$

(13)
Multivariate example (IV)

Figure: The prior (dotted line), the likelihood (dashed line) and the posterior (full line) for the simple experiment presented above. The red symbols identify the mean of each distribution.
Figure: The prior (dotted line, only shown in the first panel), and the posterior (full line) for the sequential experiment. Each panel, from left to right, represents the ingestion of a new observation, drawn from a distribution centred in the location of the green square, and whose value is indicated by the red circle.
Assessing the effectiveness of the update

Use relative entropy: the difference between prior and posterior! Relative entropy can be expressed as dispersion of the posterior $C_{post}$ relative to the prior $C_{prior}$:

$$D = \frac{1}{2} \left[ \ln \left( \frac{\det C_b}{\det C_{post}} \right) + tr(C_{post}C_b^{-1}) - n \right], \quad (14)$$

where $tr(C)$ denotes the *trace* operator (the sum of the diagonal elements) and $n$ is the rank of the matrix (the number of elements in $\vec{x}$).

Another metric is a form of ‘distance’ moved by the mean state relative to the prior uncertainty in going from the prior mean to the posterior (the ‘signal’):

$$S = \frac{1}{2} (\vec{x}_{post} - \vec{x}_b)^\top C_b^{-1} (\vec{x}_{post} - \vec{x}_b) \quad (15)$$

$$E = \frac{D + S}{\ln 2}, \quad (16)$$
A prior is **everything** that is known about a magnitude. This information includes:

- Parameter boundaries
- Expectations of smoothness
- Temporal (or spatial) trajectories for parameters, derived from a DGVM
- Experts choice
- Etc.
DA is basically Bayes’ Rule

The likelihood is how we incorporate observations, using an Observation Operator to map from state to observation space

The prior is typically just an expectation on the value of the parameter, but can be many other things

If things are

1. Gaussian
2. Linear

Then the posterior is Gaussian too

We can see DA as an update of the prior when new evidence is included
Outline

1. What is data assimilation (DA)?
2. From Bayes’ Rule to DA
   - Assessing the update
3. DA methods
   - Variational schemes
   - Sequential methods: Filters & Smoothers
   - MCMC
4. Some DA examples
   - CCDAS
   - EnKF & DALEC model
   - JRC-TIP
   - EO-LDAS
5. Discussion
Variational schemes start by writing the $-\log$ posterior, and minimise it to find the most likely value of $\vec{x}$

\[
J(\vec{x}) = \log p(\vec{x}|\vec{y}) = \left(\vec{x} - \vec{x}_b\right)^\top C_b^{-1} \left(\vec{x} - \vec{x}_b\right) + \left(\vec{y} - \mathcal{H}(\vec{x})\right)^\top C_{obs}^{-1} \left(\vec{y} - \mathcal{H}(\vec{x})\right)
\]

(17)

$\vec{x}_b$ Prior mean vector
$\vec{y}$ Observations vector
$\mathcal{H}(\vec{x})$ Observation operator acting on state
$C_b$ Prior covariance matrix
$C_{obs}$ Obs covariance matrix
"Strong" constraint

\[ J_{obs}(\vec{x}) = [\vec{y} - \mathcal{H}(\vec{x})]^\top C_{obs}^{-1} [\vec{y} - \mathcal{H}(\vec{x})] \]  \hspace{1cm} (18)

We assume that the mapping is from \( \vec{x} \) to \( \vec{y} \) using ObsOp \( \mathcal{H} \). However, we assume that this mapping is **perfect**: for a particular state value, the ObsOp will produce the required observation, without the additive noise.

**Strong constraint** The solution **must** be in the space of \( \mathcal{H} \). E.g. if solving for parameters in a DGVM, and then mapping from LAI to reflectance using an RT scheme, the solution will be compatible with the DGVM and RT scheme. This can be a major problem as models are often very wrong!

**Weak constraint** The solution needs not be in the domain of the model. This assumes **model error**
In a weak constraint system, the solution $\mathbf{x}_s$ might not be a value that can be produced by $\mathcal{M}$, but $\mathbf{C}_{model}$ will indicate how much the solution can deviate from the model. Estimating the model error is typically quite hard.
Smoothness priors

- Spatial/temporal evolution of many geophysical fields is smooth. How can we integrate this knowledge into DA?
- We can think of an extra constraint that “penalises“ rough trajectories
- Smooth trajectories characterised by 1st order differences $\sim 0$, rough trajectories 1st order difference are large
- So model $\mathcal{M}(\vec{x}) \leadsto \vec{x}_{k+1} = \vec{x}_k + \epsilon_{mod}$

$$J_{smooth} = \frac{1}{2}(\Delta \vec{x})^\top C_{smooth}^{-1}(\Delta \vec{x})$$  \hspace{1cm} (20)

$$\Delta = \begin{bmatrix} 1 & -1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & -1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -1 & 0 \\ 0 & 0 & 0 & \ldots & 1 & -1 \end{bmatrix}$$  \hspace{1cm} (21)
Smoothness priors

For this example, the mean sum of squared first order differences is 0.07 for the smooth time series, and 1.06 for the rough!
Smoothness priors

For this example, the mean sum of squared first order differences is 0.07 for the smooth time series, and 1.06 for the rough!
Solving the variational problem

Typically, we use gradient descent methods that require the partial derivatives of the cost function.

Uncertainty is calculated by looking at the matrix of second partial derivatives, the Hessian.

For linear ObsOp& Gaussian, the Hessian is the inverse of the posterior covariance matrix!
Variational methods (inc 3dVAR and 4dVAR) minimise a cost function, derived from the logarithm of the posterior:

- Weak & strong constraint flavours to deal with model error
- Very flexible: if sources of information are independent, you just add more terms to the cost function
- Minimisation is computationally expensive (particularly if the state vector is large). 
- Exploit having partial derivatives to use gradient descent methods
- The posterior covariance matrix is approximated by the inverse of the Hessian at the minimum
- Look out for local minima!
Sequential methods

- Use the updating concept: the posterior of time $k - 1$ is the prior of time $k$.
- The state and its associated uncertainty $P_{k-1}$ are propagated using a trajectory model $M \xrightarrow{} \tilde{x}_k = M\tilde{x}_{k-1}$
- Kalman filter is a good example, heritage from position tracking
The Kalman filter

Assumptions:

Distributions All

- Prior $\bar{x}_{k-1}$, $P_{k-1}$
- Obs Unc $R_k$
- Temporal Evolution model Unc $Q_{k-1}$

are Gaussian

Models Both

- Observation Operator $H$
- Temporal Evolution model $M$

are linear
KF workings

Predict state

\[
\tilde{x}_{k|k-1} = M_k^T \tilde{x}_{k-1|k-1} \quad \text{Advance state} \tag{22}
\]

\[
P_{k|k-1} = M_k^T P_{k-1|k-1} M_k + Q_k. \quad \text{Advance state uncert} \tag{23}
\]

Ingest observations

\[
r_k = y_k - H_k \tilde{x}_{k|k-1} \quad \text{Mismatch between prediction & Obs} \tag{24}
\]

\[
S_k = H_k^T P_{k|k-1} H_k + R_k. \quad \text{Innovation uncertainty} \tag{25}
\]

State update

\[
K_k = P_{k|k-1} H_k S_k^{-1}. \quad \text{Kalman gain} \tag{26}
\]

\[
\tilde{x}_{k|k} = \tilde{x}_{k|k-1} + K_k r_k. \quad \text{Update state conditioned on Obs} \tag{27}
\]

\[
P_{k|k} = \left( I - K_k H_k^T \right) P_{k|k-1}. \quad \text{Update state Unc conditioned on Obs} \tag{28}
\]
A number of extensions to the standard KF have evolved to deal with non-idealities. These are:

**Extended Kalman Filter (EKF)**  Deal with non-linearity by linearsing operators.

**Ensemble Kalman Filter (EnKF)**  Run an ensemble of sample trajectories through the model.

**Particle filter**  Monte Carlo method: uses a finite number of particles to represent the state. Lack of assumptions about linearity and normality.
So far, we assume that we update the state synchronously from $k - 1$ to $k$ to $k + 1$

- We could also update from $k + 1$ to $k$ to $k - 1$
- For time step $k$, we have two independent estimates of the state, $\hat{x}_{k|k-1}$ and $\hat{x}_{k|k+1}$

A smoother combines these two estimates

Equivalent to a 4dVAR!
Sequential methods summary

- Kalman Filter: ideal for linear models and normal statistics
- (Extensions to KF to deal with non-linear models and non-normal stats)
- Main idea is to propagate state vector and uncertainty using a process model and use as prior for new observation assimilation
- Straightforward extension of the simple univariate case at the start!
- Smoothers: Combine fwd & bkwd KF to produce an equivalence to variational scheme
Markov Chain Monte Carlo (MCMC)

- No assumptions! Just solve whatever Bayesian problem you need by Monte Carlo sampling!
- Eg: measurements of exponential decay function
  \[ y_{obs} = A \exp \left( -\frac{t}{\tau} \right) + \epsilon_{obs} \]
  ... Solve for \( p(A, \tau | y_{obs}) \) using MCMC iterative approach and 50 000 iterations (!!)
Because of its lack of assumptions, MCMC is a general method to solve inverse problems. You are guaranteed a solution, but this may require many iterations, so typically MCMC is painfully slow. The algorithm is very simple, but more advanced & efficient versions exist. MCMC might provide a benchmark solution to compare other algorithms and evaluate the effect of assumptions on DA performance.
Outline

1 What is data assimilation (DA)?

2 From Bayes’ Rule to DA
   - Assessing the update

3 DA methods
   - Variational schemes
   - Sequential methods: Filters & Smoothers
   - MCMC

4 Some DA examples
   - CCDAS
   - EnKF & DALEC model
   - JRC-TIP
   - EO-LDAS

5 Discussion
Carbon Cycle Data Assimilation System

- DA system around BETHY model (+partial derivs)
- Deal with e.g. atmospheric conc. data. Through coupled atmospheric tracer model
- Assimilate fAPAR and/or CO$_2$ concentrations
fAPAR assimilation

- Use fAPAR satellite-derived product
- fAPAR \(\rightarrow\) LAI
- Predict \(CO_2\) concentrations
fAPAR assimilation results

Figure 2. Observed (crosses with uncertainty ranges) and modeled prior (dotted) and posterior (solid line) fAPAR for Sodankylä, Zolino, 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.
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.
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\(^a\)

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

\(^a\)Units are in gC m\(^{-2}\) yr\(^{-1}\) or percentage when stated.
It is hard to make e.g. \textit{fAPAR} products consistent with radiative transfer in DGVMs.

Hard to track the noise on the product and how it feeds into the DA system.

So link the DGVM to the “raw” EO data (Surface Directional Reflectance, SDR).

Need an (non-linear!) ObsOp to go from \textit{LAI} to SDR.
The ObsOp $\mathcal{H}(\vec{x})$
- Non-linear ObsOp $\rightsquigarrow$ EnKF
- Observations constrain leaf C through LAI
- EO data requires information about other parameters
  - Leaf Chlorophyll concentration, soil reflectance, etc.
  - Assumed known and constant (!)
Non-linear ObsOp $\leadsto$ EnKF

Observations constrain leaf C through $LAI$

EO data requires information about other parameters

- Leaf Chlorophyll concentration, soil reflectance, etc.
- Assumed known and constant (!)
## Results

<table>
<thead>
<tr>
<th>Flux (gC.m$^{-2}$)</th>
<th>Assimilated data</th>
<th>Total</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NEP</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assimilation exc. snow</td>
<td>373.0</td>
<td>151.3</td>
<td></td>
</tr>
<tr>
<td>Assimilation inc. snow</td>
<td>404.8</td>
<td>129.6</td>
<td></td>
</tr>
<tr>
<td>Williams et al. (2005)</td>
<td>406.0</td>
<td>27.8</td>
<td></td>
</tr>
<tr>
<td><strong>GPP</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assimilation exc. snow</td>
<td>2620.3</td>
<td>96.8</td>
<td></td>
</tr>
<tr>
<td>Assimilation inc. snow</td>
<td>2525.6</td>
<td>42.7</td>
<td></td>
</tr>
<tr>
<td>Williams et al. (2005)</td>
<td>2170.3</td>
<td>18.1</td>
<td></td>
</tr>
</tbody>
</table>
Remember Bernard Pinty’s seminar late last year?

Use observations of visible and near infrared albedo from satellite to estimate radiative fluxes & LAI.

Variational scheme: prior on parameters & observations.

Use a fast and simple two stream RT model → effective LAI!

Minimise $J(\vec{x})$ using partial derivatives.

Constant set of prior parameters for everywhere.

Each time step considered individually (i.e. no temporal models)

\[
J(\vec{x}) = [\vec{x} - \vec{x}_0]^\top C_0^{-1} [\vec{x} - \vec{x}_0] + [\alpha_{VIS} - \mathcal{H}(\vec{x})]^\top C_{obs, VIS}^{-1} [\alpha_{VIS} - \mathcal{H}(\vec{x})] + [\alpha_{NIR} - \mathcal{H}(\vec{x})]^\top C_{obs, NIR}^{-1} [\alpha_{NIR} - \mathcal{H}(\vec{x})]
\] (29)
Focus on interpreting EO data

- Simple models (temporal regularisation, i.e., prior on expectation of smooth parameter trajectory)

- Observation Operator: complex non-linear RT model + partial derivatives
Simple DA example

- From single MERIS observation
- Reduces uncertainty
- Models observation well
- Solve for 7 biophysical parameters here
- Uncertainty on most are high
Simple DA example

- Predict two MODIS obs on that same day
- Really poor performance!
- Ill-posed problems, remember?
- Strong correlation between parameters compensate fitting
EO-LDAS Sentinel-2 performance simulation

Spatial resolution 10m for bands in the VNIR, 60m for 3 atmosph correction bands, 20m other bands

Swath 290 km

Accuracy 3-5% absolute radiometric accuracy

Wavabands 13 in total, VIS, NIR & SWIR

Revisit period With 2 sensors, 5 days

Launch 2015?
S2 performance simulation

- Set of synthetic experiments, where land surface parameters are varied
- Simulate effect of cloudiness
- Use regularisation as temporal model ("today same as yesterday, tomorrow same as today")
CCDAS: practical example using complex models and satellite-derived products

CCDAS drawback: inconsistency between EO products and model? Obs Uncertainty?

Quaife et al: Solve previous point by assimilating reflectance directly

JRC-TIP & EO-LDAS concentrate on EO

JRC-TIP & EO-LDAS: variational systems

EO-LDAS uses simple regularisation methods

EO-LDAS is in

http://jgomezdans.github.com/jgomezdans/eoldas_ng
Outline

1. What is data assimilation (DA)?
2. From Bayes’ Rule to DA
   - Assessing the update
3. DA methods
   - Variational schemes
   - Sequential methods: Filters & Smoothers
   - MCMC
4. Some DA examples
   - CCDAS
   - EnKF & DALEC model
   - JRC-TIP
   - EO-LDAS
5. Discussion
- State of the art quite basic (Use NDVI to calibrate phenology)
- CCDAS shows what can be achieved in a more consistent way
- JRC-TIP shows how to go about producing information about the land surface
- EO-LDAS shows a consistent way to use all available data to infer the state of the land surface
- Directly linking observations to models is a good idea as it simplifies uncertainty propagation
- However this is quite complex still!