Ecological Data Assimilation

The Flux Tower Story

Ankur Desai,
Atmospheric & Oceanic Sciences,
University of Wisconsin-Madison

Ameriflux Meeting, October 2007, Boulder, CO
The Scene

- An ecosystem or land-atmosphere model
  - With parameters, drivers, fluxes, stocks
  - Probably non-linear, might be chaotic
  - Many parameters are not well known (e.g., Q10)
- Driver data (filled?)
- Some observation you want to reproduce (e.g., CO2 flux) - data has noise/uncertainty
- How to minimize model-data difference taking noise into account and estimate “true” parameters and their uncertainty?
  - Partly depends on the questions you were exploring with model
Why are we doing this?

• Prediction / State space exploration
• Spatial scaling (model calibration)
• Parameter estimation / comparison
• Estimate unobserved state variables (GPP)
• Mechanism testing / Model selection
• Observation set consistency / value
• Hypothesis testing

• Flux towers are well suited to helping models do a better job at all of these…
Desai et al., accepted, AgForMet (GPP/RE)
Moffatt et al., in press, AgForMet (Gaps)

- Flux tower gap filling and GPP/RE retrieval are kinds data assimilation
Questions within this framework:
1) How should flux towers be grouped when solving for SiB3 parameters?
2) How should the resulting optimized parameters be mapped across space?
3) How should this framework be linked to the atmospheric framework. Can we solve for model parameters at that stage?
Some solutions (simple data assimilation)

- Manual (guess parameters, run, compare, try again)
  - An army of students helps
- Least squares linear fits
- Maximum likelihood
- Steepest descent and gradient optimizers (e.g., Levenburg-Marquardt, Gauss-Newton)

- Better solution: let’s ask Mathematicians and Meteorologists instead…
Bayes’ Theorem to the rescue

\[ [A|B] = [AB] / [B] \]

\[ [P|D] = ( [D|P] [P] ) / [D] \]

(parameters given data) = \[ (data given parameters) \times (parameters) \] / (data)

Posterior = (Likelihood x Prior) / Normalizing Constraint

In the long run, this is least-squares and Gaussian in the basic setup (can be modified). Main things needed for implementation are Forward operator and Likelihood function
DATA = 1.5, PRIOR $N(0, (1.5)^2)$
Likelihood, POSTERIOR
Leading to advanced data assimilation

- Direct parameter distribution exploration
- Markov Chain Monte Carlo (MCMC) - Metropolis-Hastings Algorithm (Metropolis et al, 1953) and other stochastic techniques
- (Ensemble) Kalman Filters and Smoothers
  - Good for expensive models, multiple datasets
- Genetic Algorithms - e.g., Stochastic Evolutionary Ranking Strategy (SRES)
- Neural networks
- Variational methods* (need to know adjoint)
- Tests with REFLEX, …
Lots of activity

- Big focus on MCMC
  - Big problem with MCMC is need for many iterations to sample parameter space
    - Recall [D|P]
  - But others methods are gaining
- Some models with data assimilation routines developed or in development include SipNET, TREES, ORCHIDEE, BETHY, TRIFFID, ED, Biome-BGC, LoTEC, SiB3
WLEF - Desai et al., in prep
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior</th>
<th>Posterior</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Growth related parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>photosynthetic capacity (amax)</td>
<td>112</td>
<td>58.6 +/- 2.2</td>
</tr>
<tr>
<td>growth respiration fraction</td>
<td>0.33</td>
<td>0.34 +/- 0.06</td>
</tr>
<tr>
<td>VPD modifier slope</td>
<td>0.05</td>
<td>0.066 +/- 0.009</td>
</tr>
<tr>
<td>Half saturation PAR</td>
<td>17</td>
<td>9.0 +/- 0.76</td>
</tr>
<tr>
<td>Light attenuation</td>
<td>0.5</td>
<td>0.67 +/- 0.02</td>
</tr>
<tr>
<td>WUE factor</td>
<td>10.9</td>
<td>13.4 +/- 0.46*</td>
</tr>
<tr>
<td><strong>Decomposition parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lloyd-Taylor E0</td>
<td>309</td>
<td>448 +/- 121</td>
</tr>
<tr>
<td>Lloyd-Taylor T0</td>
<td>-46</td>
<td>-59.5 +/- 10.6</td>
</tr>
<tr>
<td>Turnover rate</td>
<td>0.03</td>
<td>0.19 +/- 0.02</td>
</tr>
</tbody>
</table>
Niwot Ridge - Sacks et al., 2006, GCB

Cumulative NEE (g C m$^{-2}$)

Days after Nov. 1, 1998

Observations

Model

Cumulative NEE (g C m$^{-2}$)

Days after Nov. 1, 1998

Observations

Model
Sacks et al., 2006, GCB

- **Min. temp. for photosynthesis**
- **Optimum temp. for photosynthesis**
- **PAR attenuation coefficient**
- **Soil respiration Q_{10}**

Initial guess
Base soil respiration rate (g C g⁻¹ C day⁻¹)

Initial soil C content (g C m⁻²)

C content of leaves per unit area (g C m⁻²)

PAR half-saturation point (mol m⁻² day⁻¹)
D. Moore, in prep, AgForMet

- Cumulative NEE (gC m\(^{-2}\))
  - Reduction in Precipitation

0 10 20 30 40 50 60 70

SIPNET NEE & ET
SIPNET NEE only
Ricciuto et al., in press, AgForMet
Ricciuto et al., submitted
Ricciuto et al., submitted

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Published</th>
<th>HV</th>
<th>HW</th>
<th>WL</th>
<th>MM</th>
<th>UM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS</td>
<td>10.0</td>
<td>4.81</td>
<td>4.827</td>
<td>9.81</td>
<td>10.7</td>
<td>7.37</td>
</tr>
<tr>
<td>$Q_{10}$H</td>
<td>2.0</td>
<td>1.47</td>
<td>2.394</td>
<td>3.28</td>
<td>1.92</td>
<td>2.65</td>
</tr>
<tr>
<td>$\theta_{\text{opt}}$</td>
<td>0.55</td>
<td>0.843</td>
<td>0.717</td>
<td>0.745</td>
<td>0.685</td>
<td>0.764</td>
</tr>
<tr>
<td>$\theta_{\text{inc}}$</td>
<td>0.8</td>
<td>1.00</td>
<td>0.320</td>
<td>0.858</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$\theta_e$</td>
<td>0.30</td>
<td>0.247</td>
<td>0.129</td>
<td>0.240</td>
<td>0.285</td>
<td>0.239</td>
</tr>
<tr>
<td>$\theta_w$</td>
<td>0.13</td>
<td>0.00</td>
<td>0.0498</td>
<td>0.0534</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$n_{\text{BEL}}$</td>
<td>0.036</td>
<td>0.149</td>
<td>0.250</td>
<td>0.130</td>
<td>0.100</td>
<td>0.112</td>
</tr>
<tr>
<td>$n_{\text{BNL}}$</td>
<td>0.030</td>
<td>0.140</td>
<td>0.045</td>
<td>0.0430</td>
<td>0.0500</td>
<td>0.0516</td>
</tr>
<tr>
<td>$\alpha_{\text{BEL}}$</td>
<td>0.06</td>
<td>0.0535</td>
<td>0.117</td>
<td>0.0314</td>
<td>0.0498</td>
<td>0.0353</td>
</tr>
<tr>
<td>$\alpha_{\text{NL}}$</td>
<td>0.06</td>
<td>0.0274</td>
<td>0.021</td>
<td>0.483</td>
<td>0.249</td>
<td>0.340</td>
</tr>
<tr>
<td>$Q_{10}^{\text{VM}}$</td>
<td>2.0</td>
<td>2.23</td>
<td>1.74</td>
<td>2.36</td>
<td>2.15</td>
<td>2.04</td>
</tr>
<tr>
<td>$T_{\text{low, BEL}}$</td>
<td>-5.0</td>
<td>1.48</td>
<td>-40.0</td>
<td>-9.70</td>
<td>6.06</td>
<td>-1.83</td>
</tr>
<tr>
<td>$T_{\text{low, NL}}$</td>
<td>-15.0</td>
<td>3.55</td>
<td>3.41</td>
<td>-1.80</td>
<td>10.0</td>
<td>-20.0</td>
</tr>
<tr>
<td>$T_{\text{app, BEL}}$</td>
<td>33.0</td>
<td>50.0</td>
<td>50.0</td>
<td>50.0</td>
<td>50.0</td>
<td>50.0</td>
</tr>
<tr>
<td>$T_{\text{app, NL}}$</td>
<td>28.0</td>
<td>11.66</td>
<td>50.0</td>
<td>50.0</td>
<td>10.0</td>
<td>49.1</td>
</tr>
<tr>
<td>$R_k$</td>
<td>0.09</td>
<td>0.0192</td>
<td>0.0266</td>
<td>0.0249</td>
<td>0.0371</td>
<td>0.0282</td>
</tr>
<tr>
<td>$R_{\phi}$</td>
<td>0.25</td>
<td>0.107</td>
<td>0.0500</td>
<td>0.122</td>
<td>0.0500</td>
<td>0.0500</td>
</tr>
<tr>
<td>$R_{\text{dk}}$</td>
<td>0.015</td>
<td>0.00822</td>
<td>0.0208</td>
<td>0.0183</td>
<td>0.0019</td>
<td>0.00866</td>
</tr>
<tr>
<td>$Q_{10}^{\text{ERD}}$</td>
<td>2.0</td>
<td>1.49</td>
<td>1.61</td>
<td>2.00</td>
<td>1.100</td>
<td>1.33</td>
</tr>
<tr>
<td>$T_{\text{off}}$</td>
<td>0.0</td>
<td>11.6</td>
<td>0.80</td>
<td>9.30</td>
<td>14.3</td>
<td>11.4</td>
</tr>
<tr>
<td>$\beta_{\text{off}}$</td>
<td>0.90</td>
<td>0.678</td>
<td>0.850</td>
<td>0.672</td>
<td>0.172</td>
<td>0.517</td>
</tr>
<tr>
<td>$\gamma_p$</td>
<td>14.5</td>
<td>18.2</td>
<td>13.4</td>
<td>16.5</td>
<td>20.0</td>
<td>38.7</td>
</tr>
<tr>
<td>-(Log L)</td>
<td>N/A</td>
<td>157209</td>
<td>100989</td>
<td>69511</td>
<td>68554</td>
<td>61217</td>
</tr>
</tbody>
</table>

Well-constrained parameters:
- leaf nitrogen
- quantum efficiency
- $T_{\text{lower}}$
- phenology ($T_{\text{off}}$)

Poorly constrained parameters:
- autotrophic respiration
- soil moisture parameters

Difference from published

Coherence across space
Moore et al., submitted
Samanta et al., 2007, Water Resources Res.
A Request and Future Directions

- For model optimization / data assimilation - missing flux data is not necessarily a big deal
- But: Filled driver data (micromet, soil temp, radiation) are generally needed
- Note: Driving data vs assimilated data is an arbitrary choice (as long as you can build a forward operator)
- We can assimilate not just CO2/H2O flux, but also concentrations, micromet, radiation, wind, isotopes, chamber flux, inventory data, sapflux
- Assimilation of multiple towers promising, but consistent data sets needed
Thanks

• Thanks to the PSU/ UW-Madison/ NCAR/ SUNY-Buffalo/ CSU/ ORNL/ King’s College / WHRC/ ChEAS flux data assimilation collective and funders, DOE, NICCR, NSF, NASA, USDA