PROFOUND thoughts on Model-Data Integration & Forecasting (pun intended)

MICHAEL DIETZE
ROADMAP: REVISITING 2014

WHERE TO NEXT?

NATURE OF PREDICTABILITY

*FX FIRST PRINCIPLES*

4 WAYS DATA GETS INTO MODELS

*PECAN REVIEW & UPDATE*

MULTIPLE DATA CONSTRAINTS

*TG15*
Our Vision:

ECOSYSTEM SCIENCE, POLICY, AND MANAGEMENT INFORMED BY THE BEST AVAILABLE DATA AND MODELS
DEVELOP AND PROMOTE ACCESSIBLE TOOLS FOR REPRODUCIBLE ECOSYSTEM MODELING AND FORECASTING
4 WAYS DATA GETS INTO MODELS
PECAN REVIEW & UPDATE
4 WAYS DATA GETS INTO MODELS
PECAN REVIEW & UPDATE

Dietze 2017 “Ecological Forecasting” Princeton University Press
4 WAYS DATA GETS INTO MODELS
PECAN REVIEW & UPDATE
Bayesian Parameter Calibration

Fer et al in prep
4 WAYS DATA GETS INTO MODELS

Traits
Informatics
Drivers

MODEL

Data

AGB (megagrams/hectare)

Fx

Models should have CI!!

SIPNET @ Hubbard Brook, NH

Param

Forecast

Dietze 2017

Ecological Forecasting
Princeton University Press
Multivariate Tobit

- Range restrictions
- Zero inflated

Estimated Process Error

Raiho et al. *in prep*
4 WAYS DATA GETS INTO MODELS
PECAN REVIEW & UPDATE

MODEL

Data
Informatics

Traits

Initial Condition

Assimilation

Forecast Cycle

Forecast

Utility

Decision Support

Scenario Development

Drivers

Calibration

Param

Structure

Uncertainty Propagation

Assessment Validation Benchmarking

Sensitivity & Uncertainty Analysis

Priors

Theory
MEDLYN ET AL. 2015 NCC
WHAT OBSERVATIONS DO MODELS NEED TO REPLICATE?
Uncertainty Analysis

Parameter Uncertainty

- Growth Respiration
- Fine Root Allocation
- Leaf Turnover Rate
- Specific Leaf Area
- Vcmax
- Seedling Mortality
- Mortality Coefficient
- Reproductive Allocation
- Photosynthesis min temp
- Quantum Efficiency
- Root Turnover Rate
- Seed Dispersal
- Root Respiration Rate
- Stomatal Slope
- Leaf Width

Sensitivity

- Elasticity

Output Uncertainty

- Partial
- Variance

*can easily be extended to driver and IC uncertainties

SYNTHESIS
CALIBRATION
DATA COLLECTION
ROADMAP: REVISITING 2014

WHERE TO NEXT?

NATURE OF PREDICTABILITY
*FX FIRST PRINCIPLES*

4 WAYS DATA GETS INTO MODELS
*PECAN REVIEW & UPDATE*

MULTIPLE DATA CONSTRAINTS
*TG15*
MULTIPLE DATA CONSTRAINTS

- Autocorrelated error can lead to overconfident DA
  5 min vs 10 min sampling
- Unbalanced data, biased calibrations
- Ad hoc methods unsatisfying ‘choose your own CI’
- Paradox that noisier data ‘appear’ to have more information
- 2014 focus: INFORMATION
TG15: PSEUDODATA EXPERIMENTS

VSEM

- Perfect model, balanced data
- Perfect model, unbalanced data
- Model error, balanced data
- Model error, unbalanced data

Fit | Param
--- | ---
✓ | ✓
✓ | ✓
✓ | ✗
✗ | ✗

Lead: David Cameron
MODEL ERROR, UNBALANCED DATA

NEE

vegetative carbon (Cv)

Observed / predicted values

Observed / predicted values

Time

Time

truth

pseudodata
WHAT IF THERE ARE DATA BIASES

- PM, Balanced, multiplicative data bias: ✔️
- PM, Unbalanced, MB data: ❌
- Model error, Unbalanced, MB data: ❌
MODEL ERROR
UNBALANCED DATA W/ MULTIPLICATIVE BIAS
### WHAT IF THERE ARE DATA BIASES

<table>
<thead>
<tr>
<th>PM, Balanced, multiplicative data bias</th>
<th>Fit</th>
<th>Param</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM, Unbalanced, MB data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model error, Unbalanced, MB data</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### CAN WE CORRECT BIAS W/ STATS?

<table>
<thead>
<tr>
<th>ME, Unbalanced, linear model</th>
<th>Fit</th>
<th>Param</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM, Unbalanced, MB data, linear model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ME, Unbalanced, MB data, linear model</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
MODEL W/ ERROR
UNBALANCED

DATA W/
MULTIPLICATIVE BIAS

LINEAR BIAS COR
TG15 TAKE HOMES

- More than Information content, issue is **errors & biases** in models and data
- Perfect models can fit unbalanced data
- Building bias correction into calibration can lead to acceptable performance, but not the ‘true’ parameters
ROADMAP: REVISITING 2014

WHERE TO NEXT?

NATURE OF PREDICTABILITY

FX FIRST PRINCIPLES

4 WAYS DATA GETS INTO MODELS

PECAN REVIEW & UPDATE

MULTIPLE DATA CONSTRAINTS

TG15
ON THE NATURE OF THE ECOLOGICAL FORECASTING PROBLEM

Heterogenous  Stochastic  Memory

State Space  ???

Process error?  Updatable hierarchical models?
State convergence?  Fusing multiple data?
HOW DO WE MEASURE PREDICTABILITY?

\[ Y_{t+1} = f(Y_t, X_t | \bar{\theta} + \alpha) + \varepsilon \]
WHAT CAUSES VAR TO INCREASE WITH TIME?

\[
\text{Var}[Y_{t+1}] \approx \left( \frac{\partial f}{\partial Y} \right)^2 \text{Var}[Y_t] \text{ IC uncert} + \left( \frac{\partial f}{\partial X} \right)^2 \text{Var}[X] \text{ driver uncert} + \left( \frac{\partial f}{\partial \theta} \right)^2 \left( \text{Var}[\theta] + \text{Var}[\alpha] \text{ param uncert} \right) \text{ param variability} + \text{Var}[\varepsilon] \text{ process error}
\]

\[
= \text{INTERNAL} + \text{EXTERNAL} + \text{PARAMETERS} + \text{RANDOM EFFECTS} + \text{PROCESS ERROR}
\]
<table>
<thead>
<tr>
<th>Concept</th>
<th>Predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endogenous (internal) stability, $Y$</td>
<td>Grows or declines exponentially, all other terms are linear</td>
</tr>
<tr>
<td></td>
<td>Predictive uncertainty grows without bound or asymptotically</td>
</tr>
<tr>
<td></td>
<td>Determined by classic stability thresholds</td>
</tr>
<tr>
<td></td>
<td>Relative importance increases with larger scales</td>
</tr>
<tr>
<td>Exogenous (external) stability, $X$</td>
<td>Predictability increases when drivers are predictable or Dynamics are insensitive to variation</td>
</tr>
<tr>
<td></td>
<td>Relative importance increases with time</td>
</tr>
<tr>
<td></td>
<td>Covariates useful for prediction may be different from those used for explaining the same process</td>
</tr>
<tr>
<td></td>
<td>Experimental design emphasizes how much $X$ affects $Y$</td>
</tr>
<tr>
<td></td>
<td>Autocorrelation slows how Var[$X$] averages out when scaling</td>
</tr>
<tr>
<td></td>
<td>Model selection chooses models that are overly complex for prediction</td>
</tr>
<tr>
<td>Parameter uncertainty, $\theta$</td>
<td>Dominates data-limited problems and over-parameterized models</td>
</tr>
<tr>
<td></td>
<td>Within a forecast, does not necessarily increase or decrease</td>
</tr>
<tr>
<td></td>
<td>With sampling, declines asymptotically to zero</td>
</tr>
<tr>
<td></td>
<td>Does not average out when aggregating in space or time</td>
</tr>
<tr>
<td></td>
<td>Sampling can be targeted through uncertainty analysis</td>
</tr>
<tr>
<td>Parameter variability, $\alpha$, and</td>
<td>Encompasses heterogeneity, model structural error, and stochasticity</td>
</tr>
<tr>
<td>process error, $\varepsilon$</td>
<td>Does not decline asymptotically with sampling</td>
</tr>
<tr>
<td></td>
<td>Like $X$, declines with scale but rate dependent on autocorrelation</td>
</tr>
<tr>
<td></td>
<td>Declines to non-zero asymptote through model improvement</td>
</tr>
<tr>
<td></td>
<td>Partitioning of parameter variability important for extrapolation from one unit of measure to another to account for potential differences and larger uncertainty</td>
</tr>
</tbody>
</table>
NATURE OF THE PREDICTION PROBLEM...

- **Theory**
  - What drives dynamics?
  - Generality across processes and locations

- **Practice**
  - What can we predict?
  - How to tackle new systems

- **Methods**
  - What to measure
  - How we build models
  - How we assimilate data

\[
\text{Var}[Y_{t+1}] \approx \left(\frac{\partial f}{\partial Y}\right)^2 \text{Var}[Y]_{\text{stability}} + \left(\frac{\partial f}{\partial X}\right)^2 \text{Var}[X]_{\text{driver uncertain}} + \frac{\partial f}{\partial \theta} \left(\text{Var}[\theta]_{\text{param uncertain}} + \text{Var}[\alpha]_{\text{param sensitivity}}\right)_{\text{param error}} + \text{Var}[\epsilon]_{\text{process error}}
\]

- INTERNAL + EXTERNAL + ERROR
Near-term Ecological Forecasting Initiative

Dietze & Wheeler: Fluxes & Phenology

Weathers: Aquatic Productivity

LaDeau & Foster: Ticks & Small Mammals

Talbot & Averill: Microbial diversity
• Accelerate basic science
• Make our science more relevant
Postdoc: A Prototype Data Assimilation System for the Terrestrial Carbon Cycle

A post-doctoral position is available in the Ecological Forecasting lab at Boston University as part of a larger project to develop a terrestrial carbon cycle data assimilation system, focused initially on North America, using the PECAn model informatics system (pecanproject.org). This system will employ a formal Bayesian model-data fusion between bottom-up process-based ecosystem models and multiple data sources to estimate key C pools and fluxes.

Duties:

Candidate will work with collaborators at Brockhaven to iteratively extend the PECAn data assimilation system to ingest a wide range of remotely-sensed and ground data with the goal of fusing and reconciling multiple data streams into a continental-extent carbon cycle (pools and fluxes) data product. They will be responsible for incorporating new scaling approaches into the data assimilation system itself and extending the assimilation to work with multiple land surface models. They will contribute to analyses assessing the impacts of different data sources and models on carbon pool and flux estimates and uncertainties, with the aim of improving carbon monitoring, reporting, and verification. Finally, the candidate will also assist collaborators at NOAA who will be incorporating the bottom-up assimilation product into the CarbonTracker-Lagrange (CT-L) inverse modeling framework to help reconcile top-down and bottom-up flux estimates.

Qualifications:

Minimum qualifications are a doctoral degree in a related environmental science (ecology, geography, atmospheric science, earth science, etc.) Experience with R and at least one of the following topics is required(along with interest in learning the others): Bayesian statistics, ensemble filtering approaches (e.g. EnKF), ecosystem or land surface modeling, remote sensing, and ecoinformatics. Salary is commensurate with experience and qualifications. Three years of funding available.

Interested applicants are encouraged to submit a cover letter, CV, and contact info for 3 references to Dr. Michael Dietze (dietze@bu.edu).

Boston University is an Equal Opportunity/Affirmative Action Employer.