Assessing Model Performance

Lesson 11
Step 1: Is the model output reasonable?

- Range of values
- Units
- General pattern in time & space
Step 2: Graphical comparisons to data
Accuracy of Prediction

![Graph showing observed values against predicted values with a 1:1 line]

- Observed Values
- Predicted Values
Identify Outliers
Assess Biases
Miscalibration

Observed Values vs Predicted Values

1:1 line

Difference in trends
Dynamics & Drivers

(a) Value vs Time

(b) Value vs Time
Diagnosing a model is Hypothesis Testing
Why would a model fail at low humidity?

Stomatal sensitivity too low?

Too much soil moisture?

What experiments would I run in the model to test this?
Focus on key assumptions

Walker et al. 2014
Most models treated LMA as a constant

Models need drought effect on both $G_s$ and $A$

Drought

Models disagreed on fraction of rainfall intercepted and canopy gas exchange when wet

Rainfall interception

Increase in LMA

A/$G_s$ should be proportional to $C_a$

Stomatal conductance

Photosynthesis

Models assumed different values for the ratio of light- or CO$_2$- limited photosynthesis

Transpiration

Model sensitivity of canopy transpiration to stomatal conductance varied

Decline in foliar N

Plant tissue turnover

Models should include dynamic allocation constraints

Large uncertainty in model parameter values for wood turnover

Increased belowground C allocation

Priming

Ecosystem N losses

Models underestimated net transfer of N from soil organic matter to vegetation

Declining N availability

Models need constrained estimates of N losses with eC$_a$

Models need to allow for some flexibility in coupling of C and N cycles
“data simulated under a model should look similar to data gathered in the real world.”
Conn et al 2018
IN THE FITTING, WE ASSUMED IID NORMAL ERRORS

Does that seem like an adequate description of the data?
IN THIS FITTING, WE ASSUMED EXPONENTIAL ERRORS WITH NON-CONSTANT VARIANCE

Does that seem like an adequate description of the data?
Bayesian p-value / prediction interval

- Posterior predictive distribution is the uncertainty of the "true" value

- **Prediction interval** is the expected variance of the observed values = PPD + error
  - Shows us what distribution we would expect for the data

- Bayesian p-value is when we use PPD + error to calculate the value of the cdf of the observed data
  - Distribution should be flat (uniform)
  - "Bayesian residuals"
Step 3: Quantitative Skill Assessment
Error Statistics

- Root Mean Square Error (RMSE)
- Bias
- Correlation ($r$)
- $R^2$
- Regression slope

Proper: based on the metric used for calibration
Local: depends on data that could actually be collected

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - x_i)^2}$$
Correlation

Correlation $r = 0$

Correlation $r = -0.3$

Correlation $r = 0.5$

Correlation $r = -0.7$

Correlation $r = 0.9$

Correlation $r = -0.99$
Correlation

Anomaly Correlation of ECMWF 500 hPa Height Forecasts

- Northern Hemisphere (darker)
- Southern Hemisphere (lighter)

Year: 1981-2006

- Day 3
- Day 5
- Day 7
- Day 10

ECMWF
Taylor Diagram

Correlation Coefficient

Standard deviation ratio

Normalized RMSE

Schaefer et al. 2012 JGR-B
Autocorrelation

Correlogram

\[ \rho_s = \frac{\gamma_s}{\gamma_0} \]

R: \text{acf}(x)
“Climatology”

Comparing process model performance to time-scale appropriate summary statistics
## Data mining the residuals

- Wide variety of Data Mining algorithms in use
- Large debate about use in process modeling and forecasting
- Potentially useful for generating hypothesis about when/where model fails

<table>
<thead>
<tr>
<th>Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>CART</td>
</tr>
<tr>
<td>GAM</td>
</tr>
<tr>
<td>Random Forests</td>
</tr>
<tr>
<td>Boosted regression trees</td>
</tr>
<tr>
<td>Artificial Neural Network</td>
</tr>
<tr>
<td>Support Vector Machines</td>
</tr>
</tbody>
</table>
How Good Are FiveThirtyEight Forecasts?

MLB games, 2016-18

We thought the 306 teams in this bin had a 31% chance of winning. They won 28% of the time.

https://projects.fivethirtyeight.com/checking-our-work/