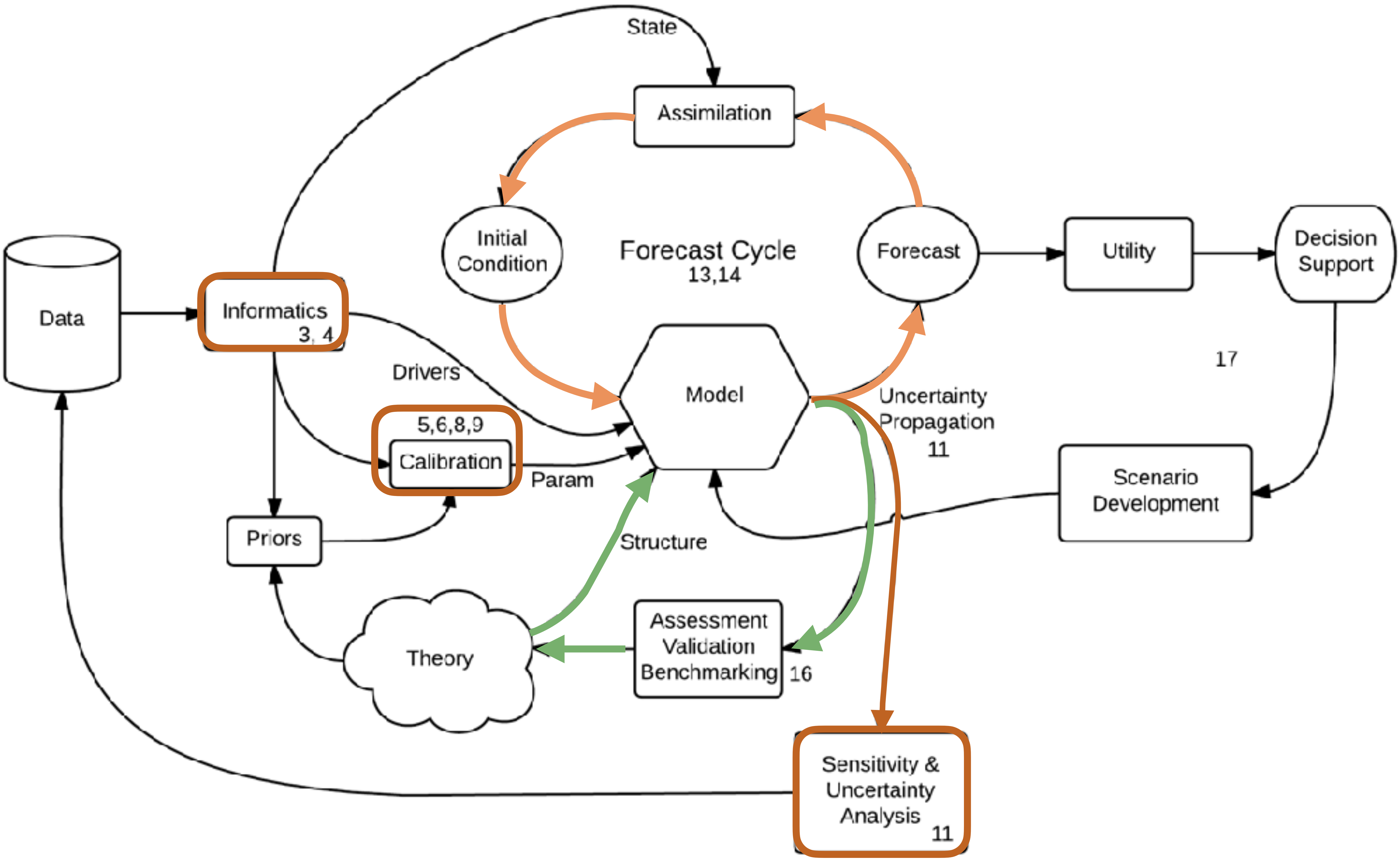
Assessing Model Performance

Lesson 11



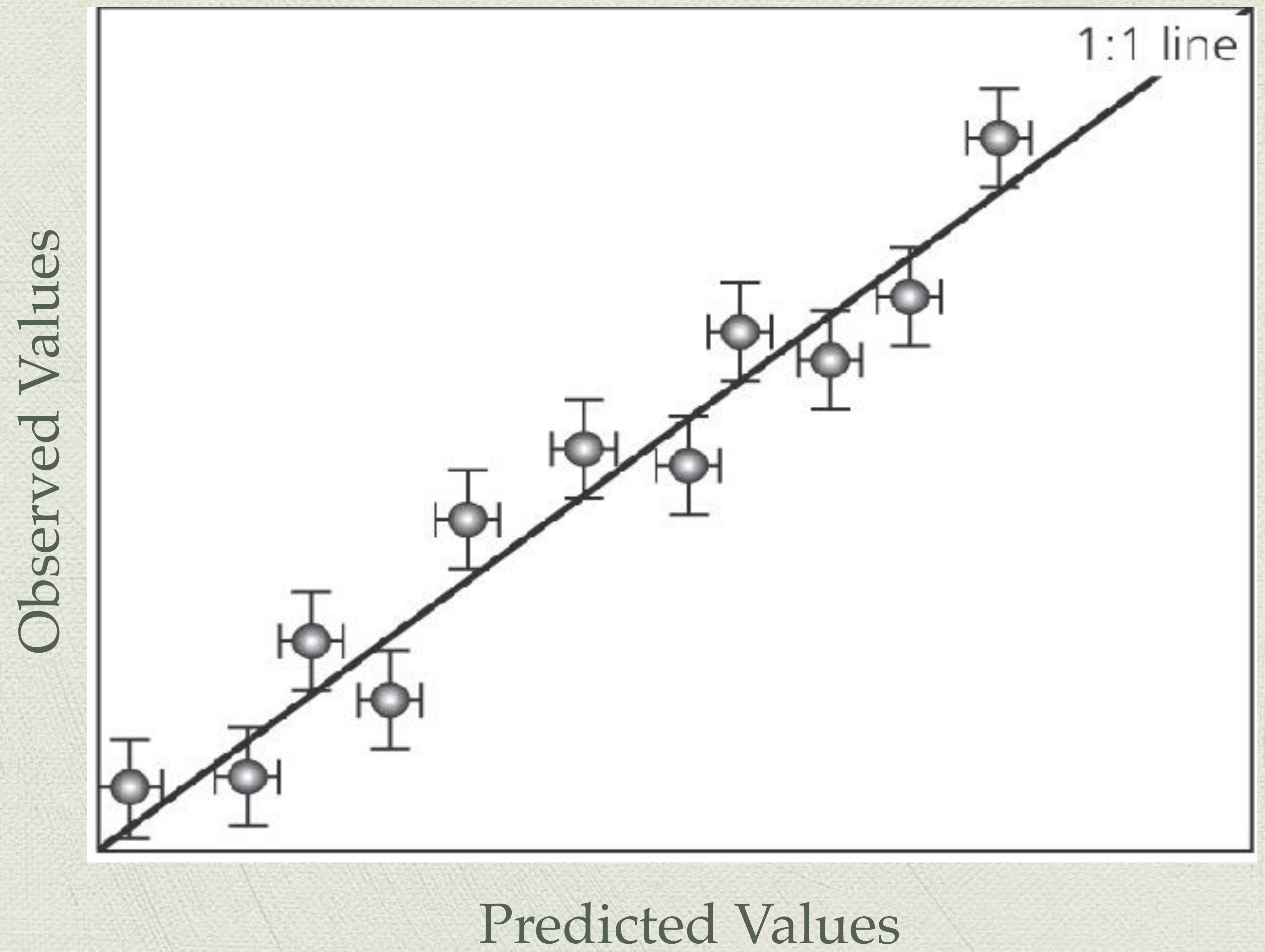
State
Assimilation
Initial Condition
Forecast Cycle
13,14
Forecast
Utility
Decision Support
Data
Informatics
3, 4
Drivers
Model
Uncertainty Propagation
11
17
5,6,8,9
Calibration
Param
Scenario Development
Priors
Structure
Theory
Assessment Validation Benchmarking
16
Sensitivity & Uncertainty Analysis
11

- Range of values
- Units
- General pattern in time & space

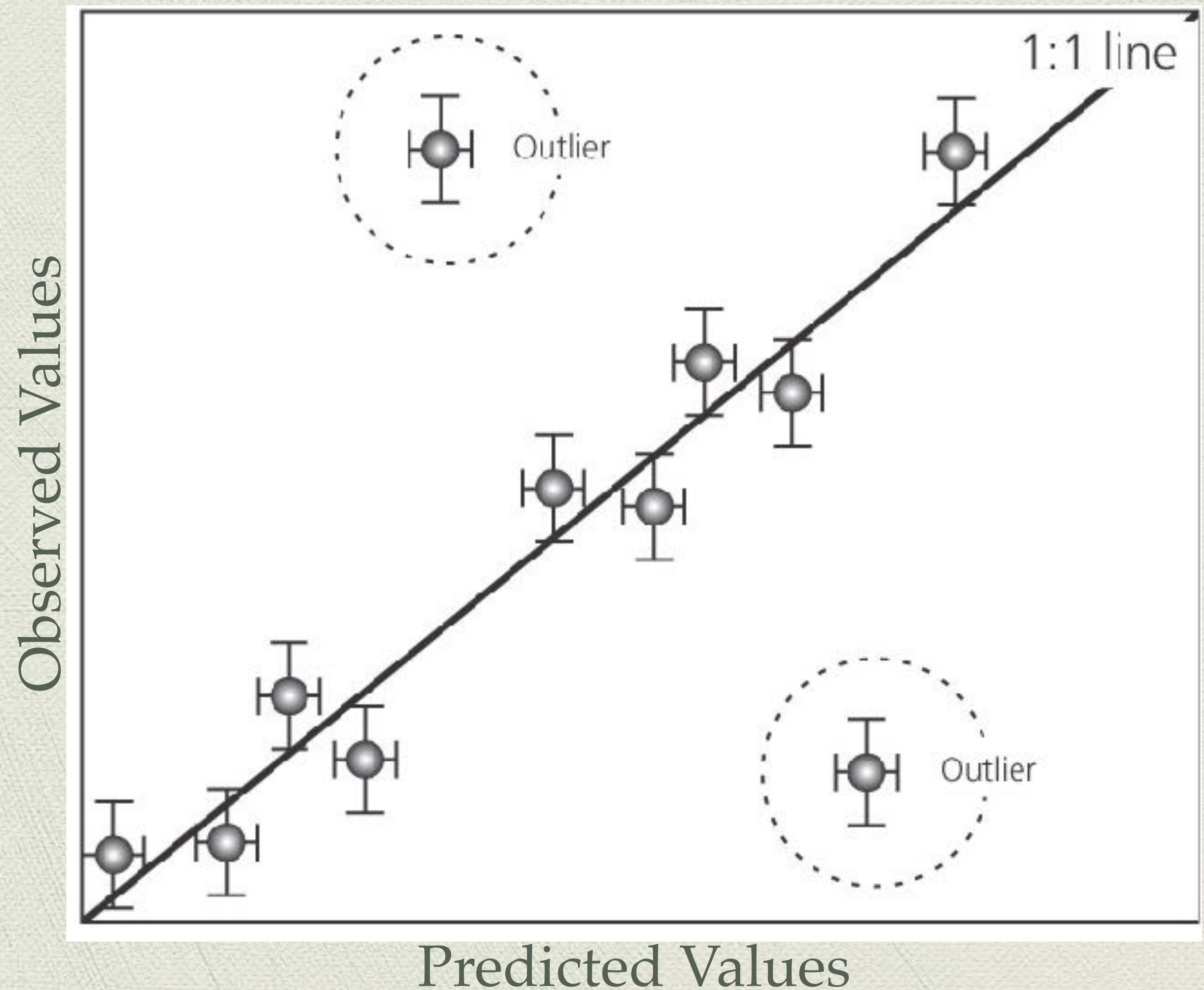
Step 1: Is the model output reasonable?

Step 2: Graphical comparisons to data

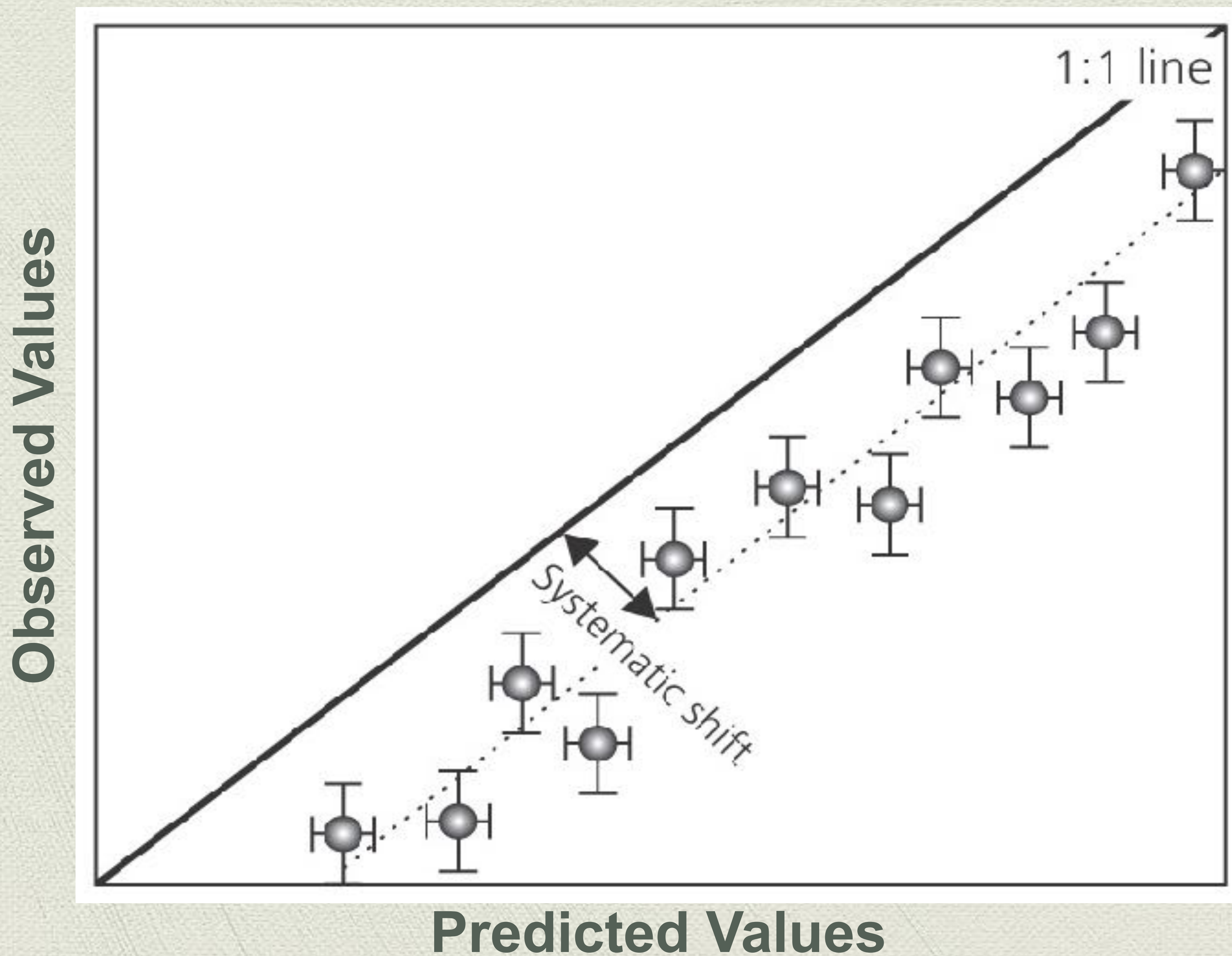
Accuracy of Prediction



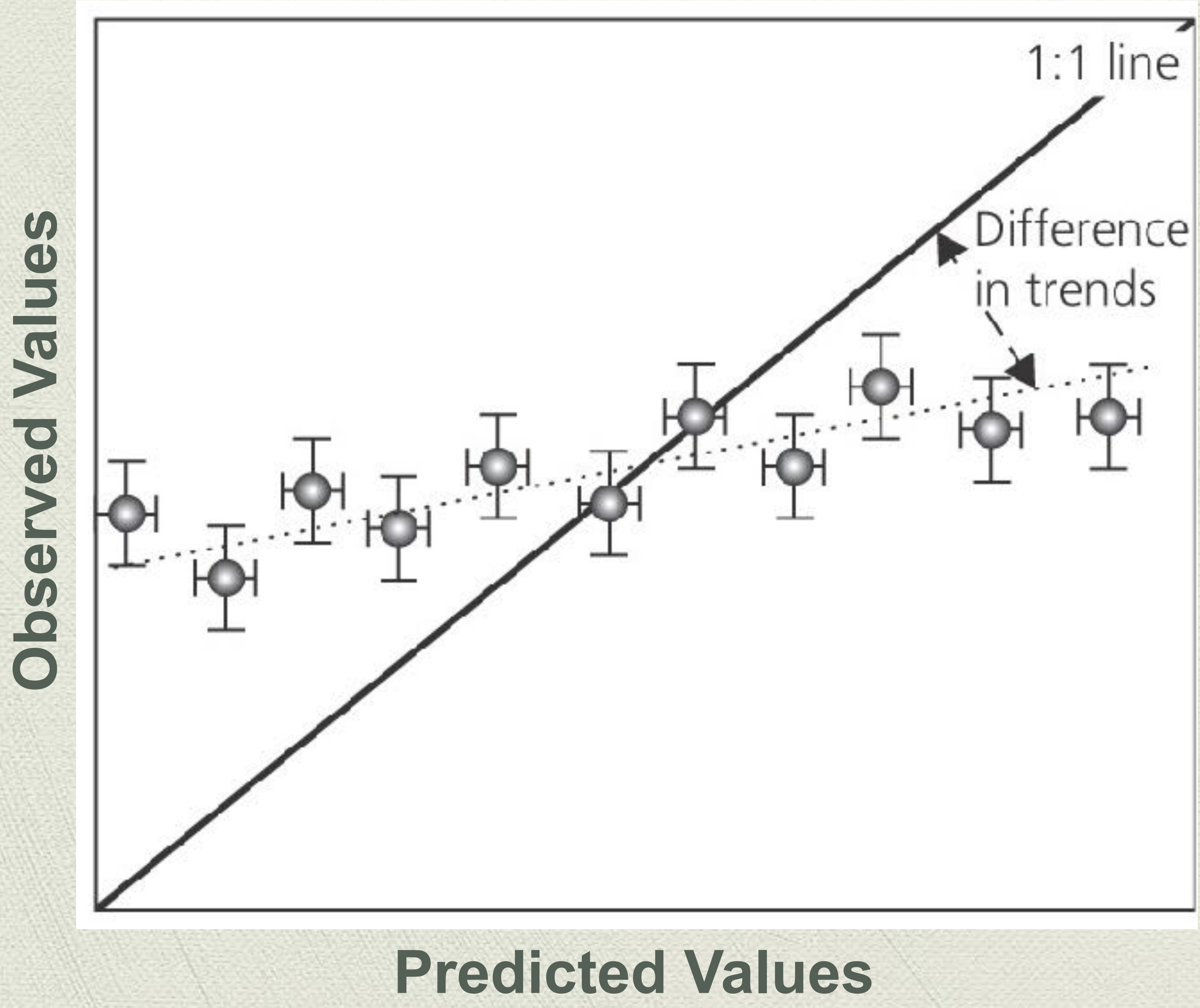
Identify Outliers



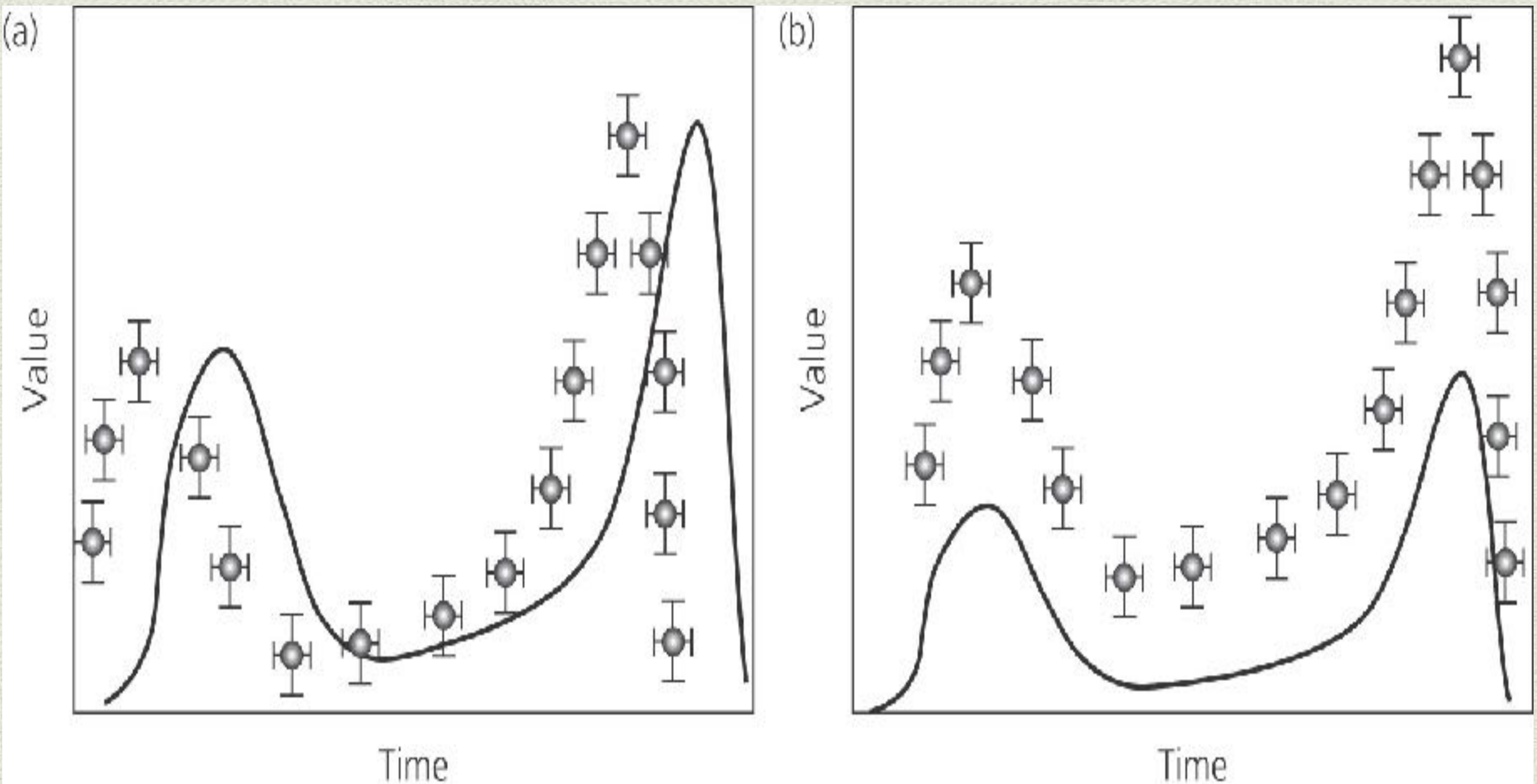
Assess Biases

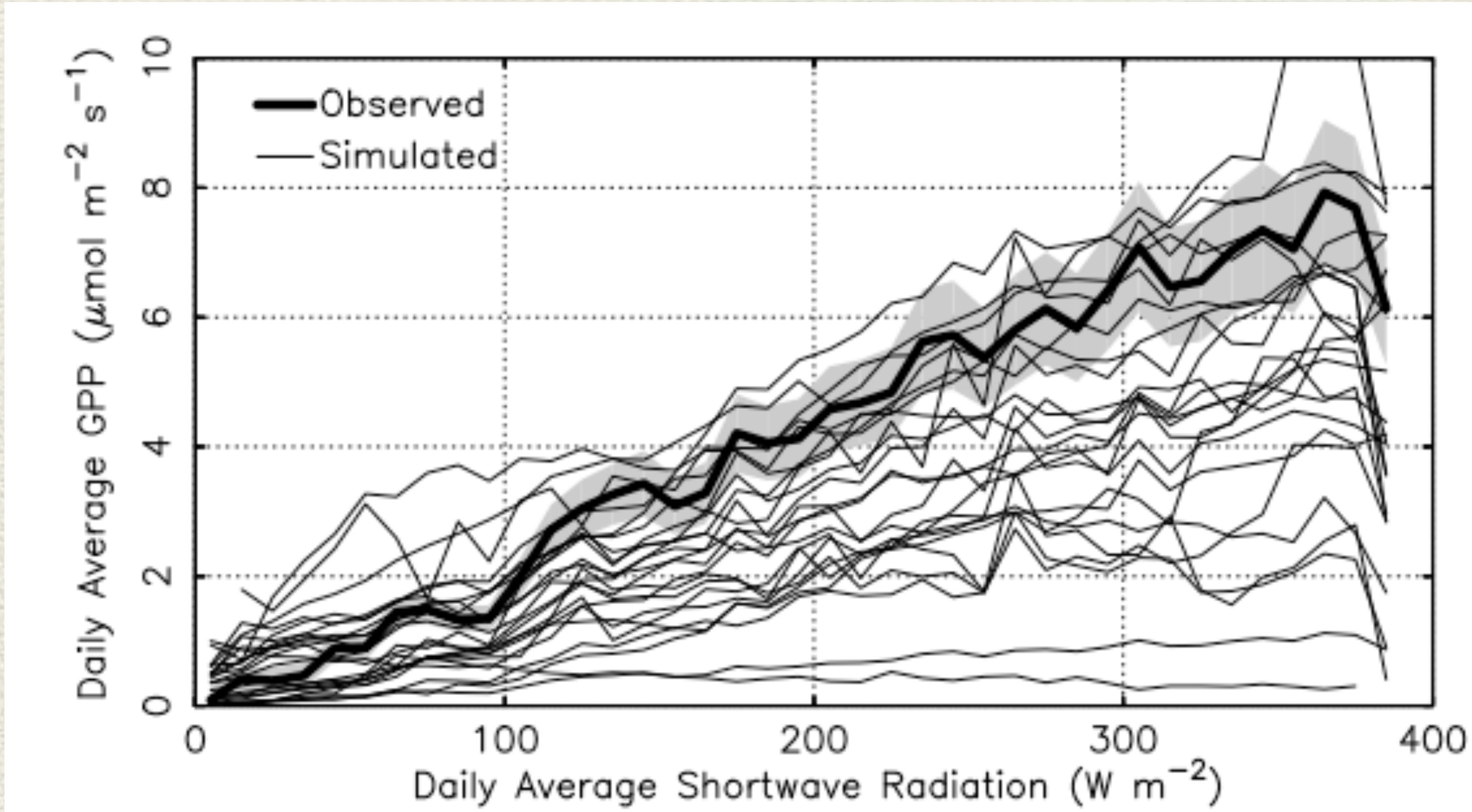


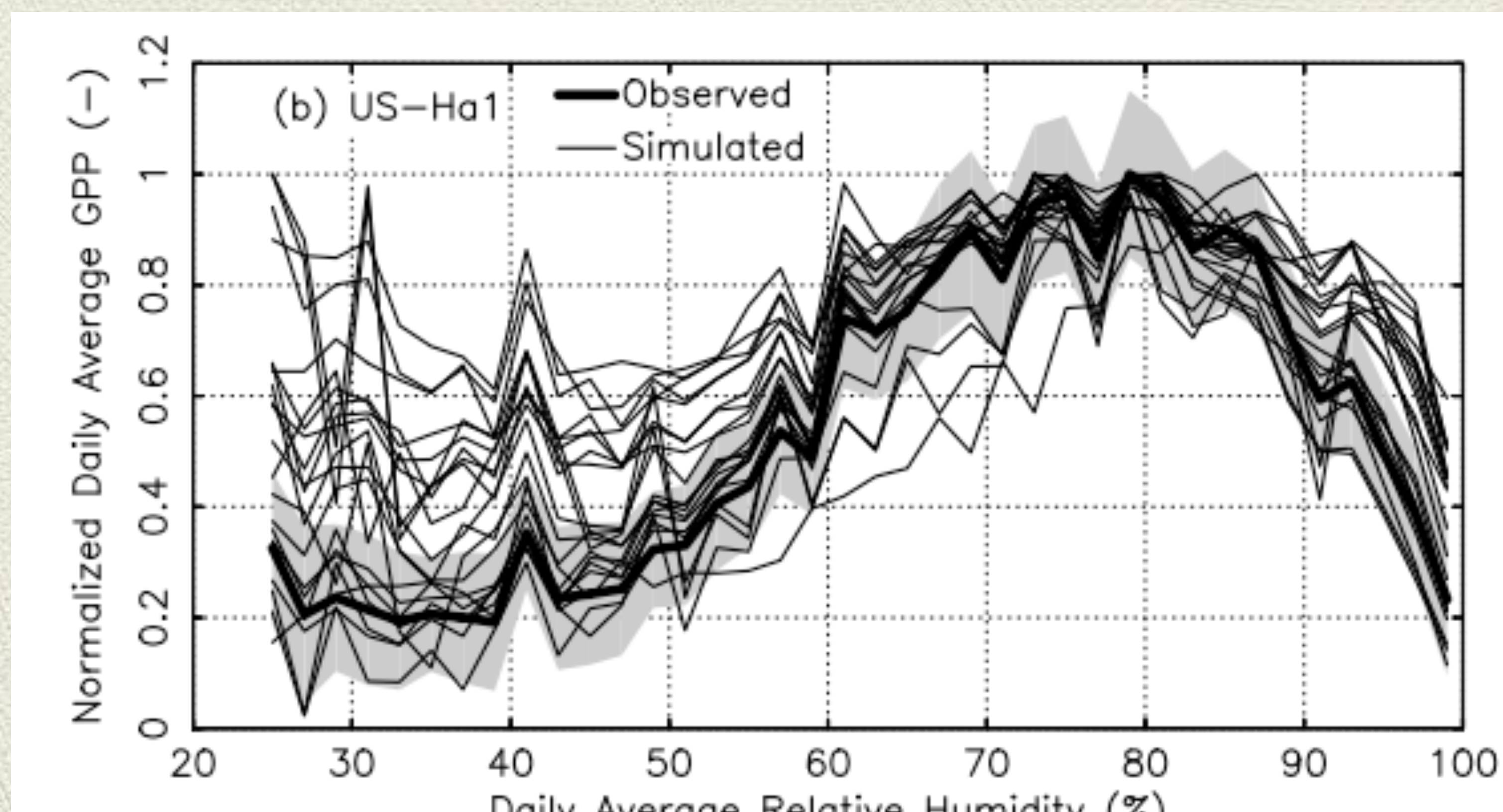
Miscalibration



Dynamics & Drivers







Schaefer et al. 2012 JGR-B

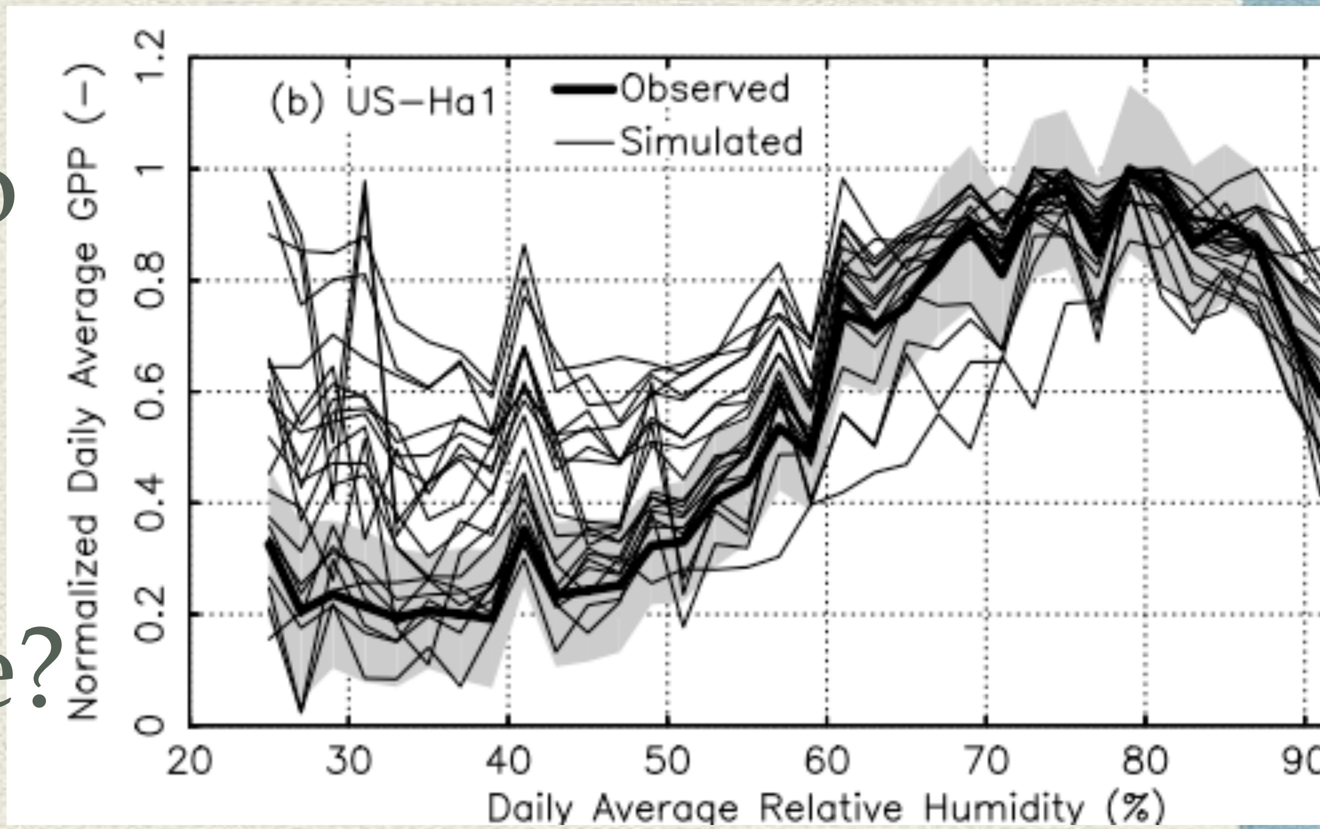
Diagnosing a model is Hypothesis Testing

- Why would a model fail at low humidity?

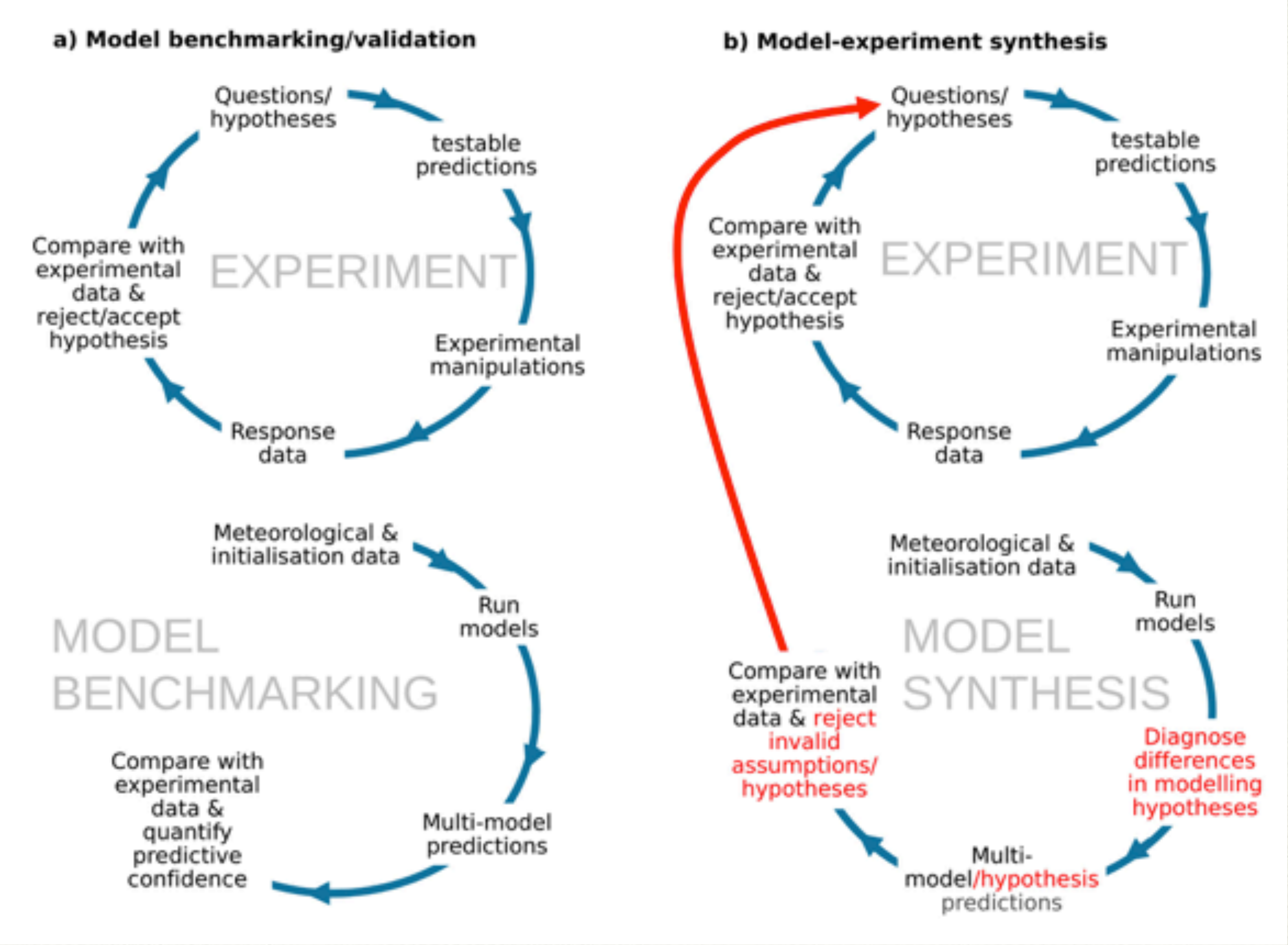
- Stomatal sensitivity too low?

- Too much soil moisture?

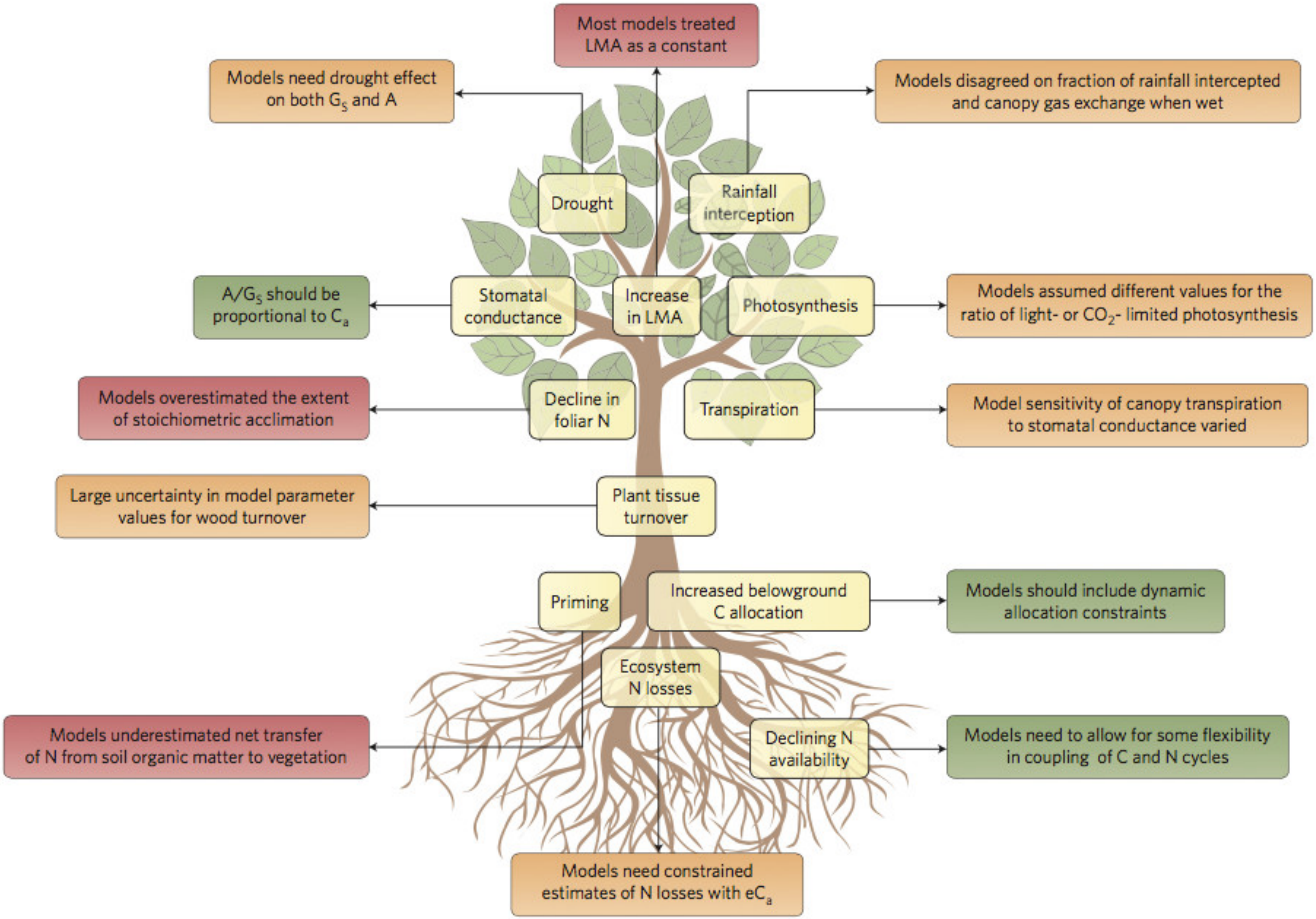
- What experiments would I run in the model to test this?



Focus on key assumptions



Walker et al 2014

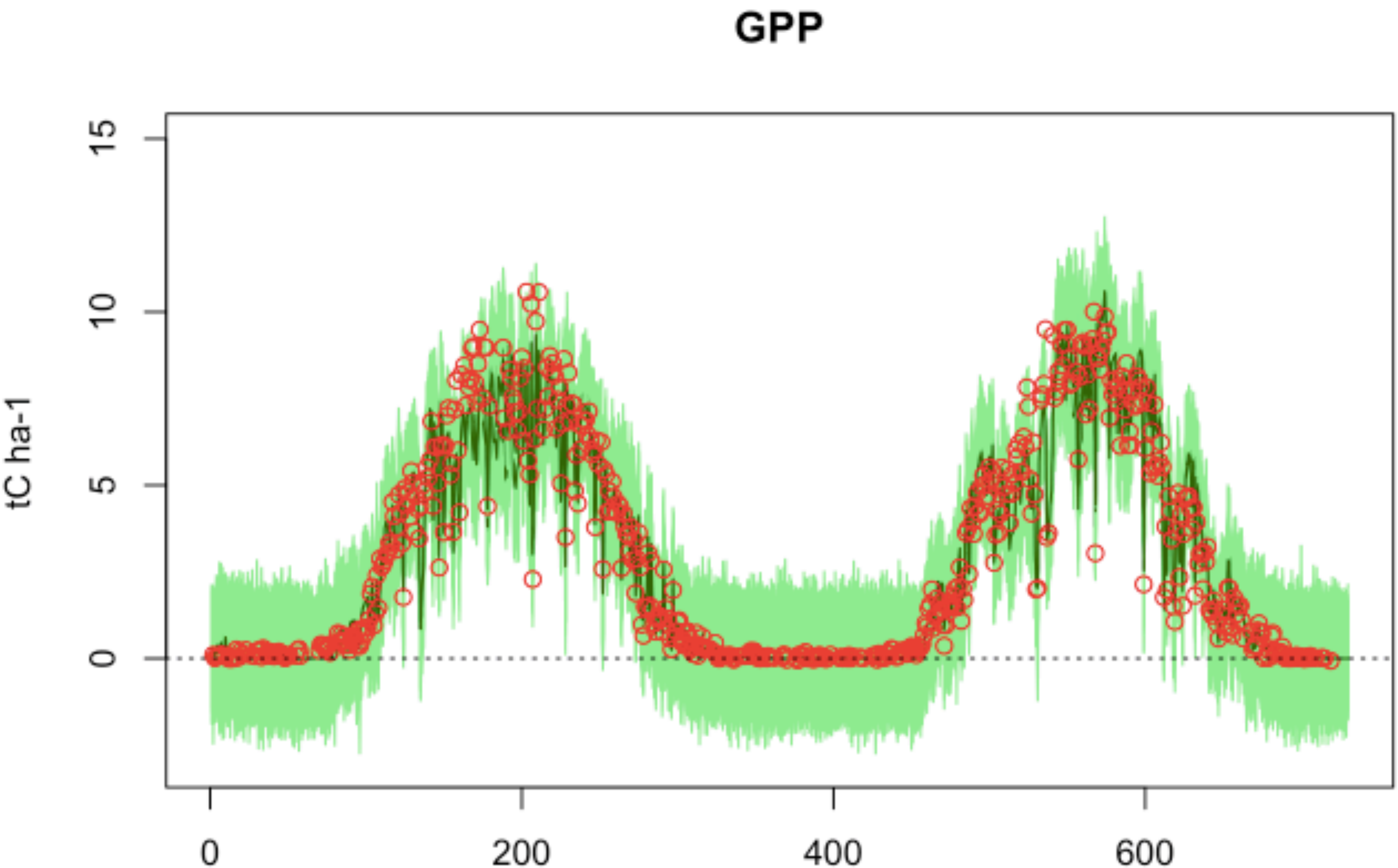


Medlyn et al NCC 2015

“data simulated under a model should look similar to data gathered in the real world.”

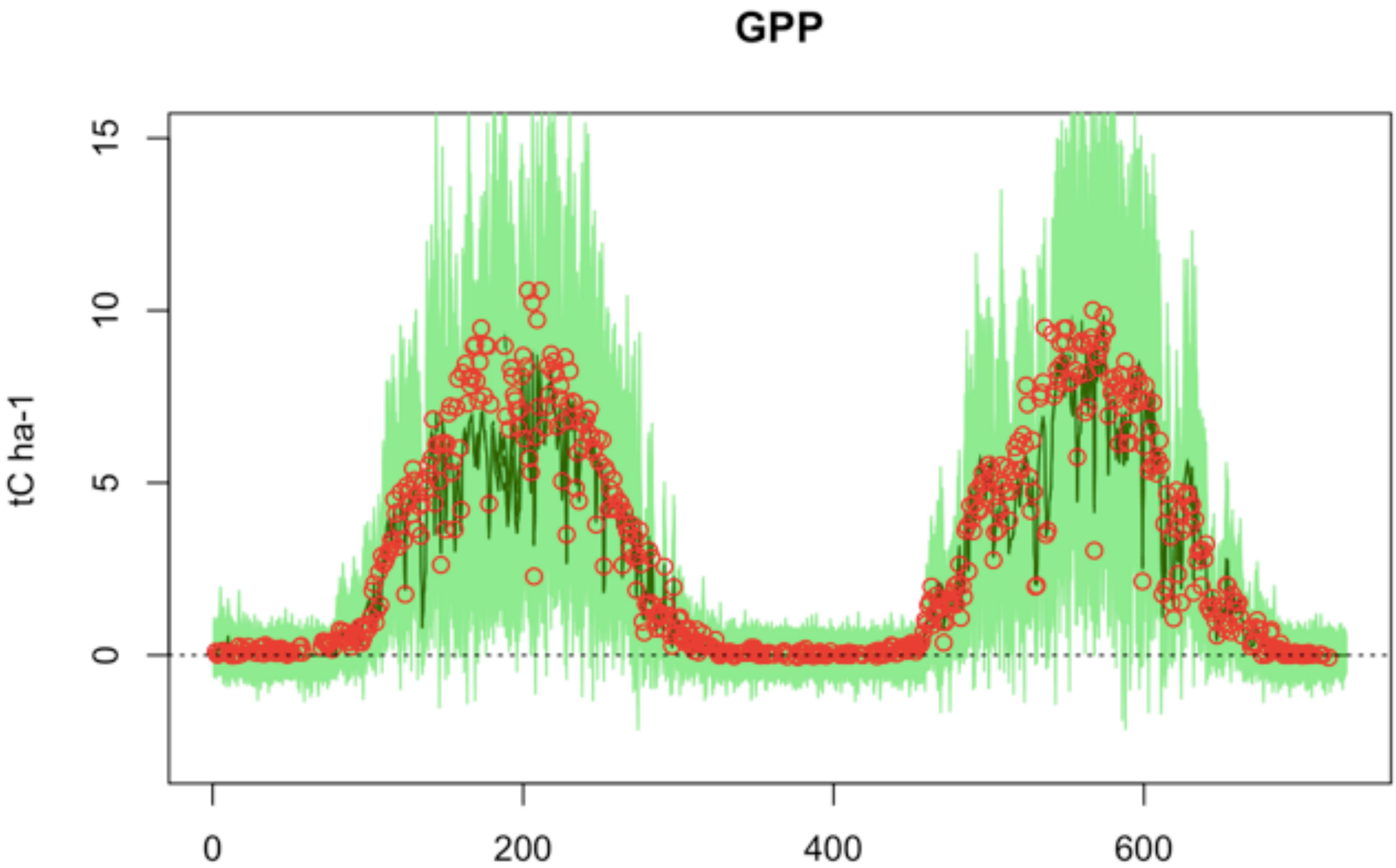
Conn et al 2018

IN THE FITTING, WE ASSUMED IID NORMAL ERRORS



Does that seem like an adequate description of the data?

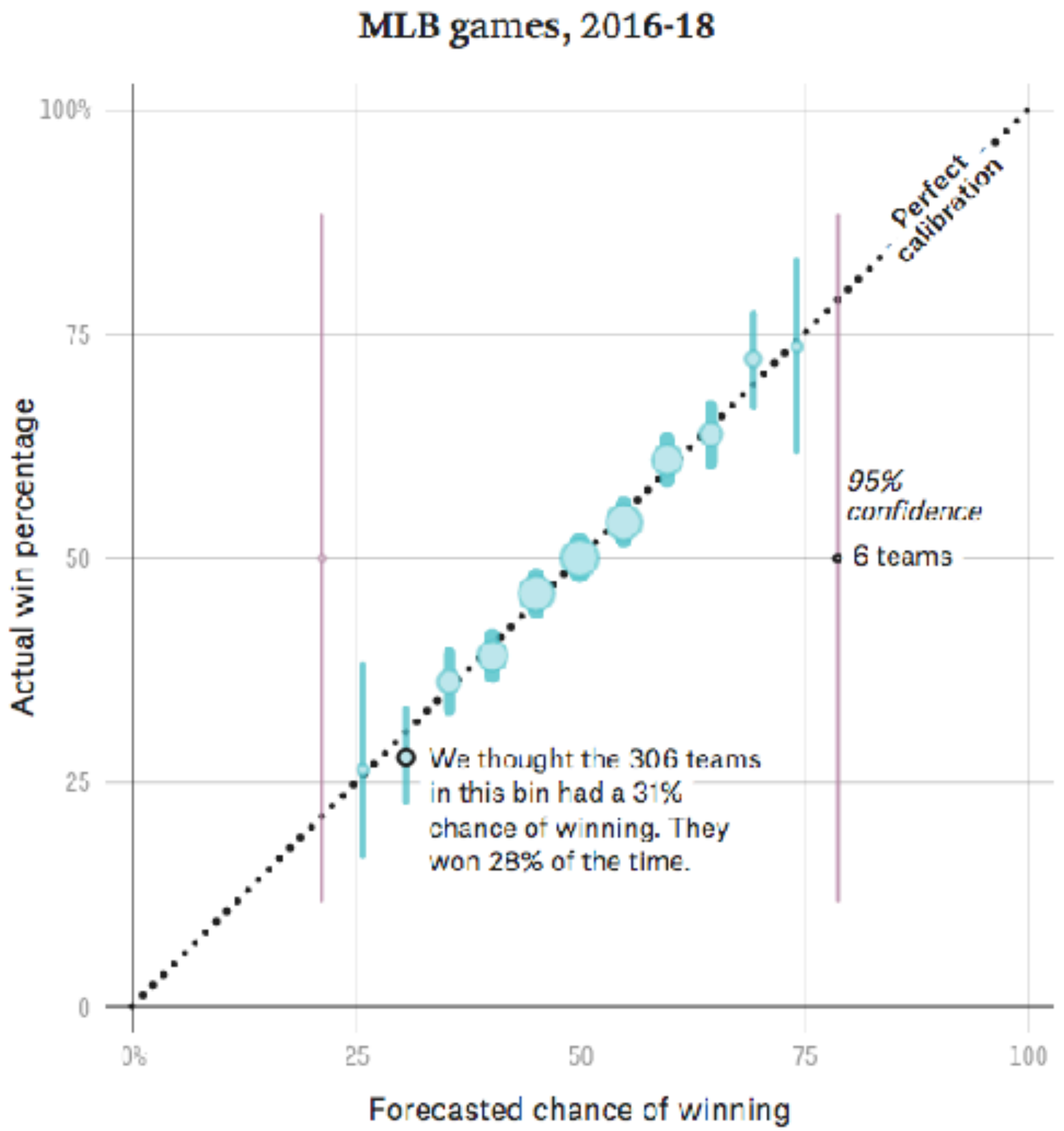
IN THIS FITTING, WE ASSUMED EXPONENTIAL ERRORS WITH NON-CONSTANT VARIANCE



Does that seem like an adequate description of the data?

PUBLISHED APR. 4, 2019, AT 7:00 AM

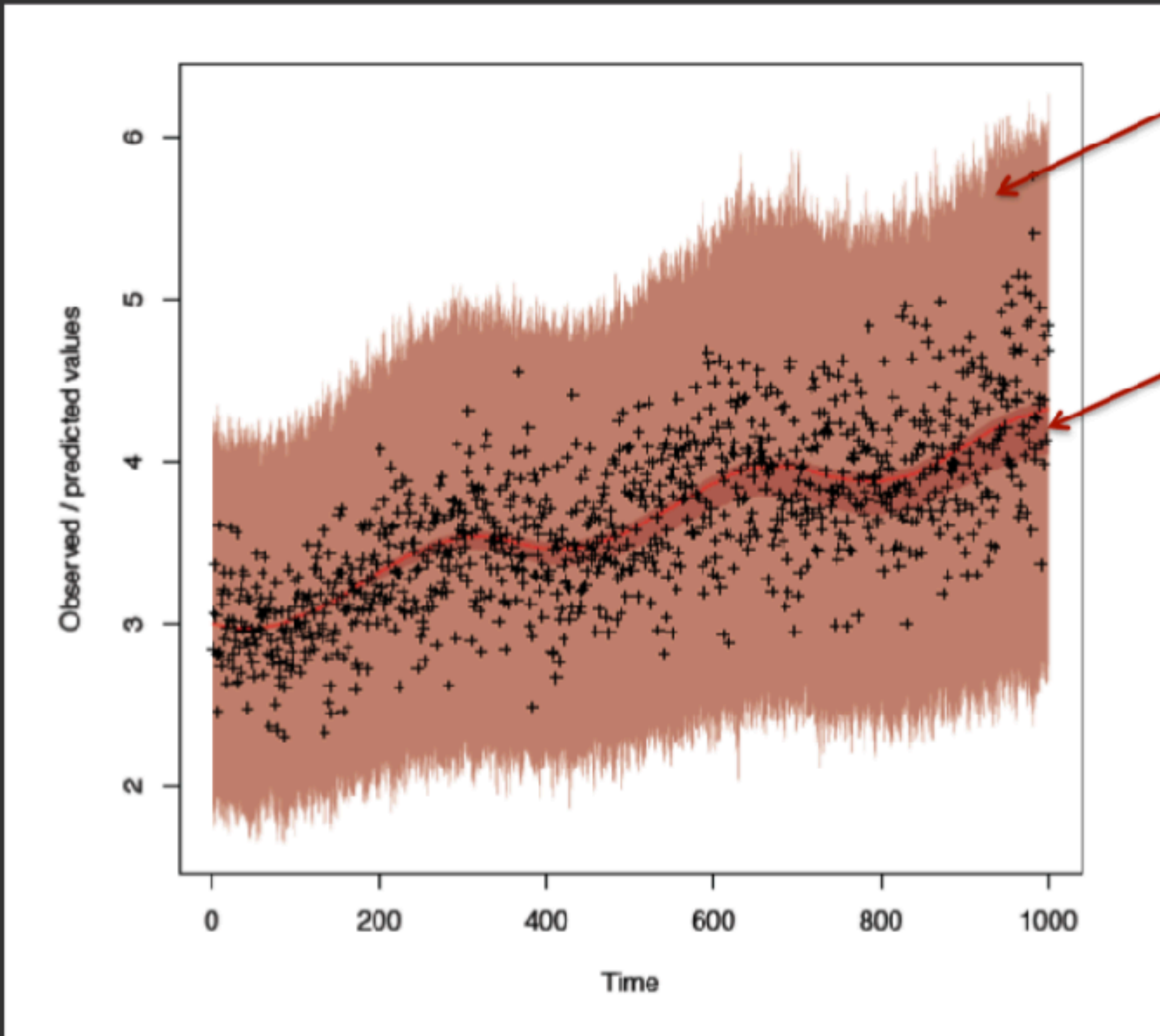
How Good Are FiveThirtyEight Forecasts?



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

Bayesian p-value / prediction interval

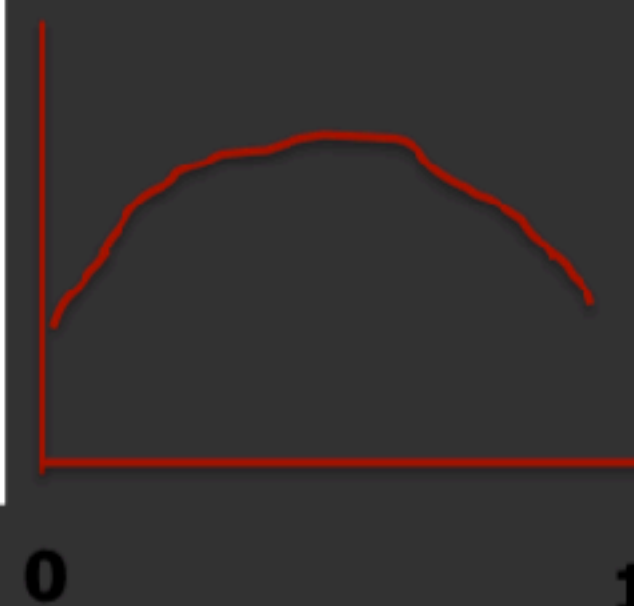
- Posterior predictive distribution is the uncertainty of the „true“ value
- **Prediction interval** is the expected variance of the observed values = PPD + error
  - Shows us what distribution we would expect for the data
- Bayesian p-value is when we use PPD + error to calculate the value of the cdf of the observed data
  - Distribution should be flat (uniform)
  - „Bayesian residuals“



PPD + Error

Posterior Predictive Distribution

Distribution of ecdf values for residuals

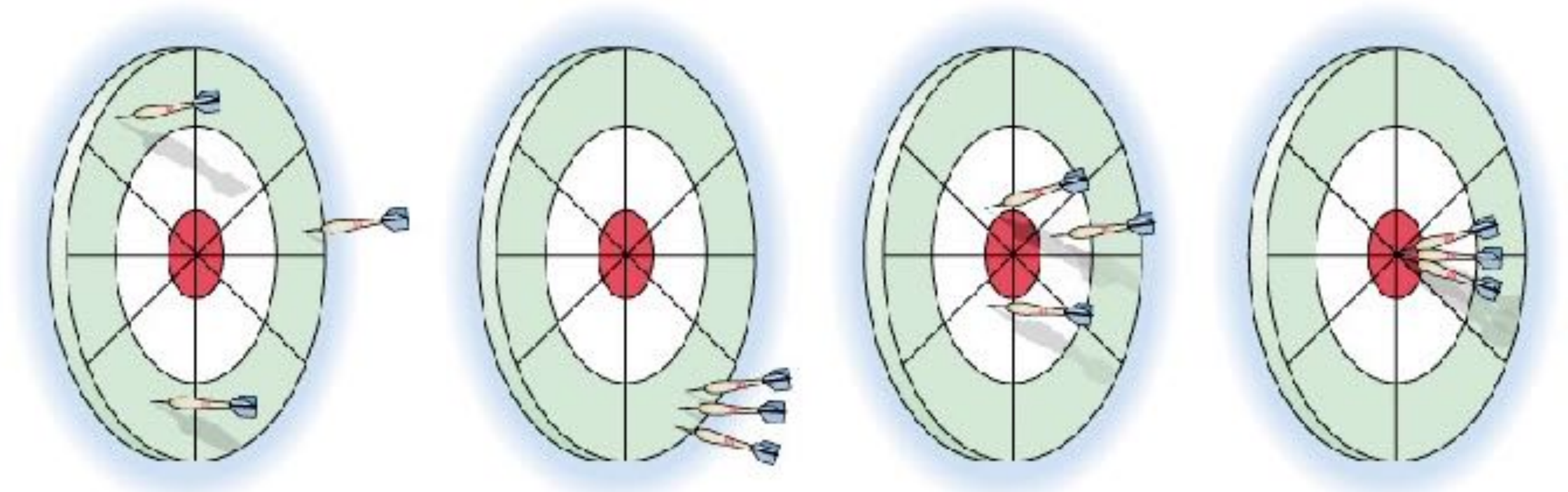


Step 3: Quantitative Skill Assessment

Error Statistics

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

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



(a) Low accuracy Low precision

(b) Low accuracy High precision

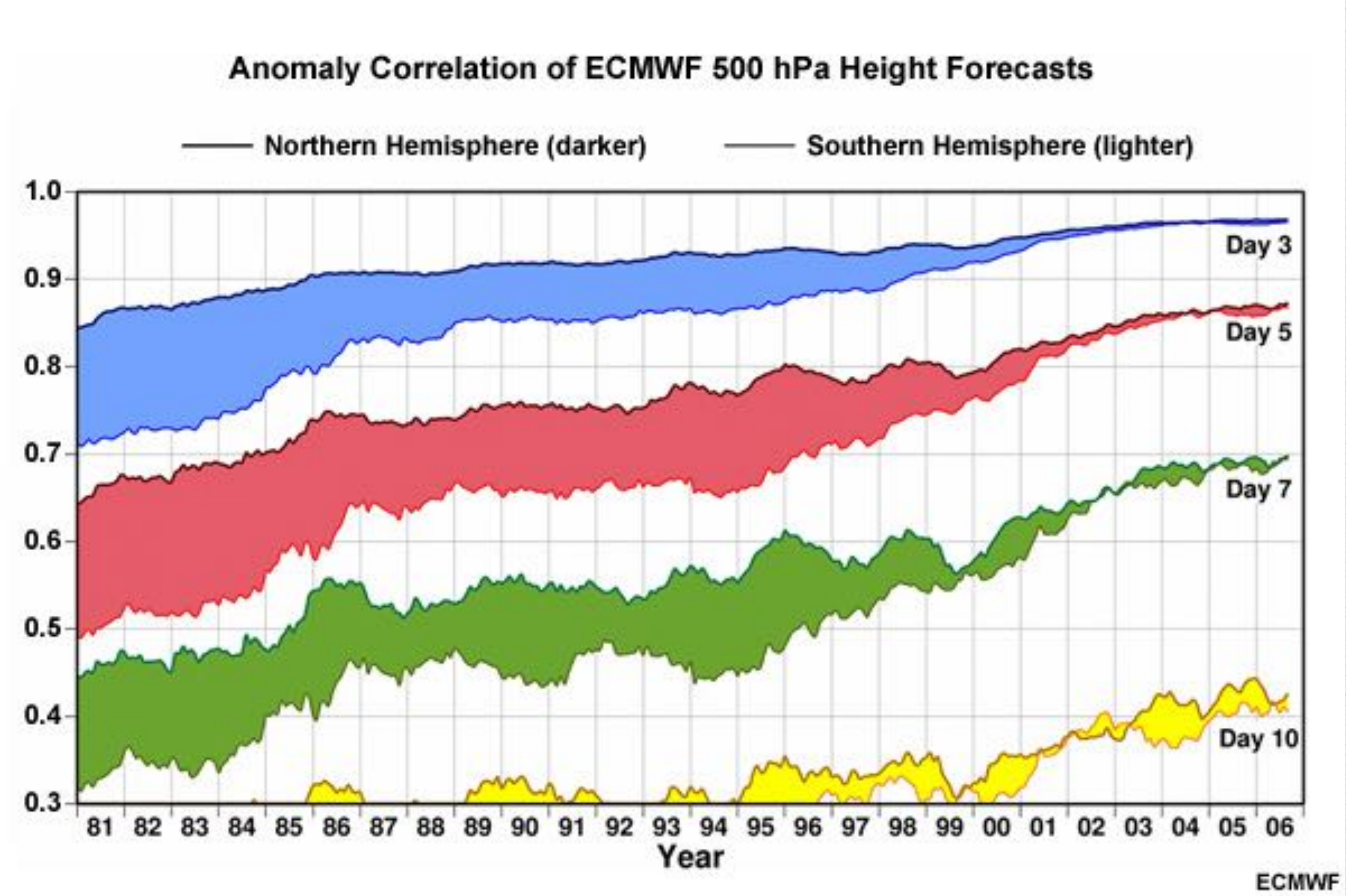
(c) High accuracy Low precision

(d) High accuracy High precision

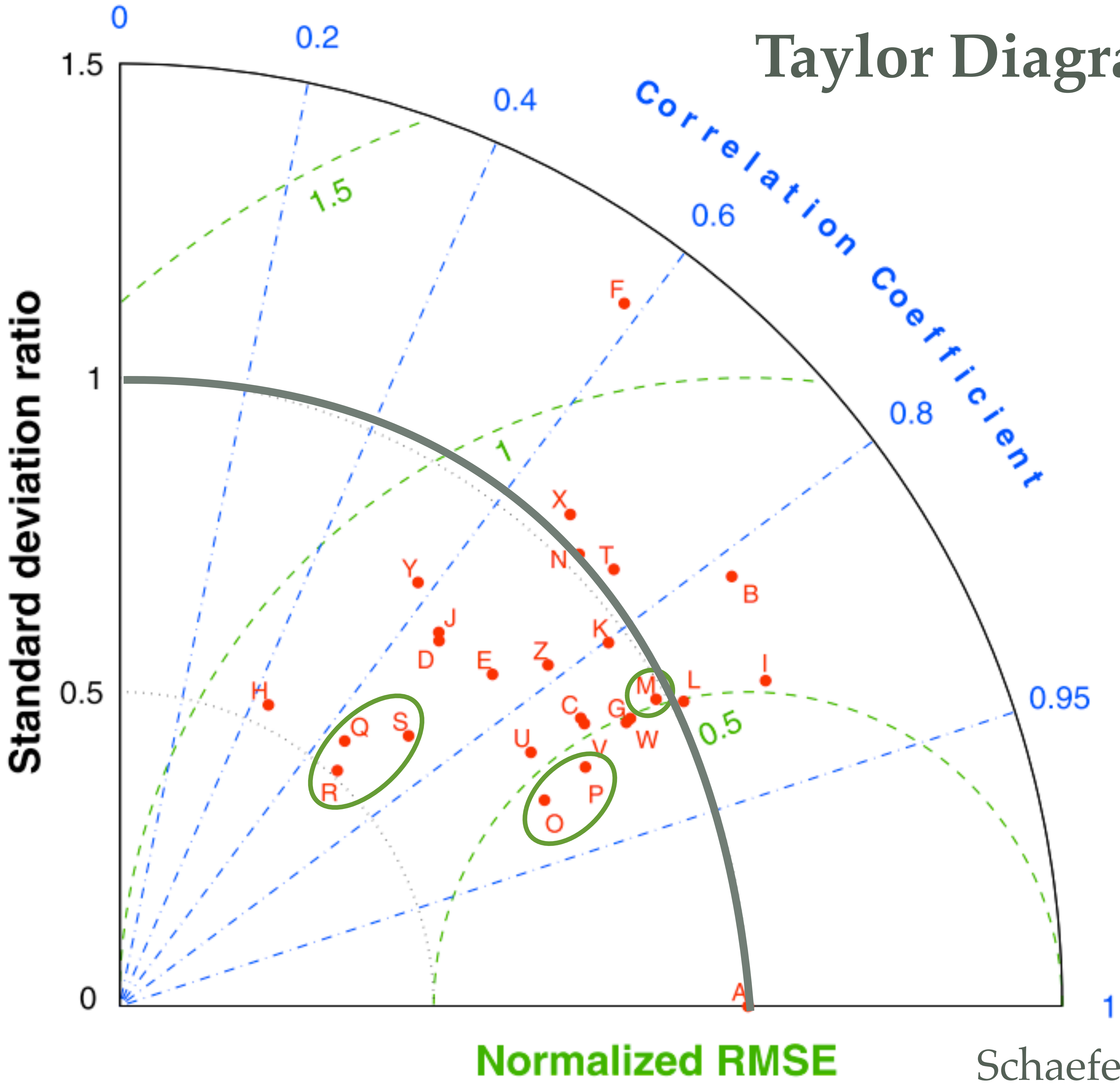
Proper: based on the metric used for calibration

Local: depends on data that could actually be collected

Correlation



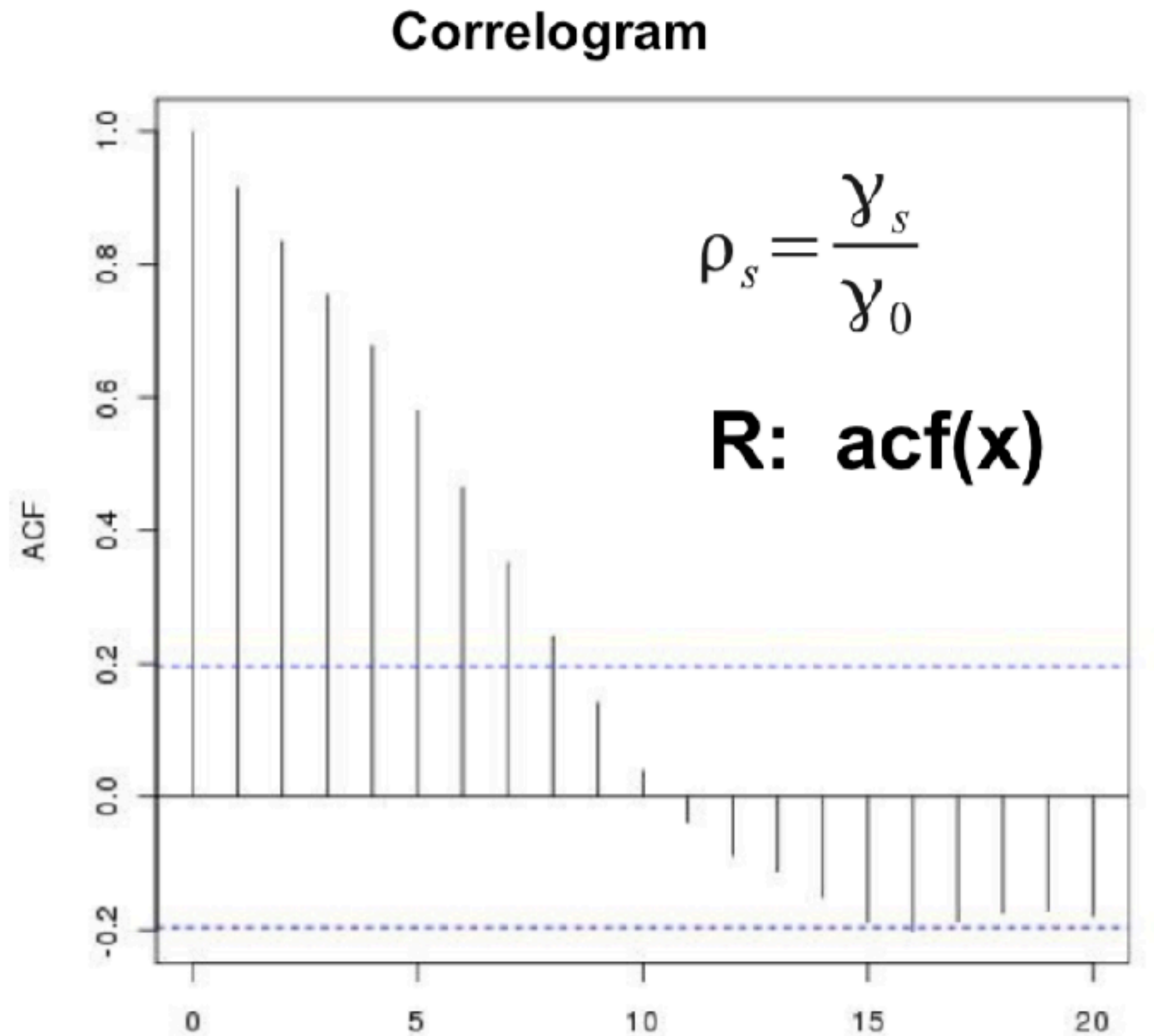
Taylor Diagram



A OBSERVED
B AGROIBIS
C BEPS
D BIOMEBGC
E CAN–IBIS
F CNCLASS
G DLEM
H DNDC
I ECOSYS
J ED2
K ISAM
L ISOLSM
M LOTEC
N LPJ
O MEAN_ALL
P MEAN_DIURN
Q MODIS_ALG
R MODIS_5
S MODIS_5.1
T ORCHIDEE
U SIB
V SIBCASA
W SIBCROP
X SSIB2
Y TECO
Z TRIPLEX

Schaefer et al. 2012 JGR-B

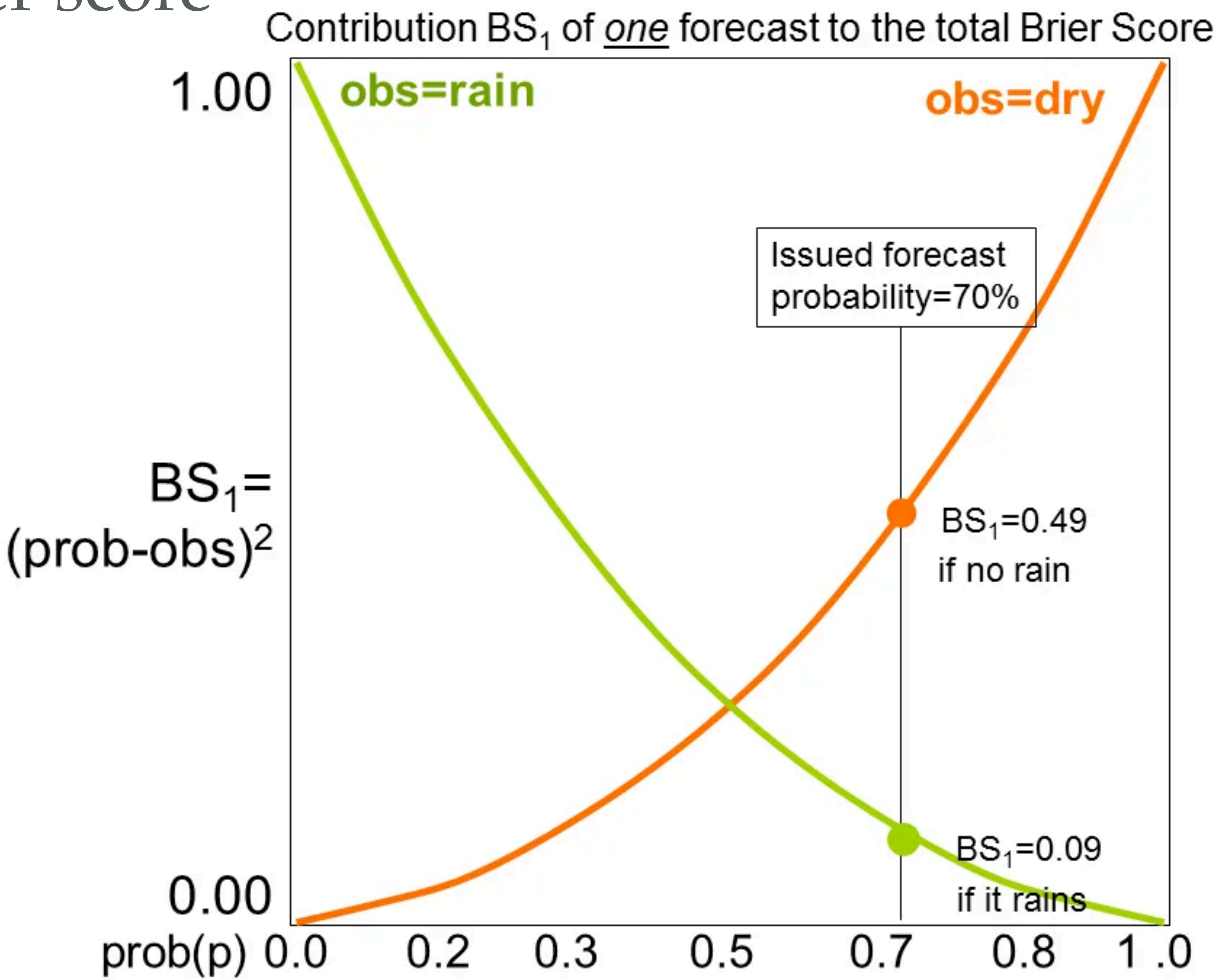
Autocorrelation



$$\rho_s = \frac{\gamma_s}{\gamma_0}$$

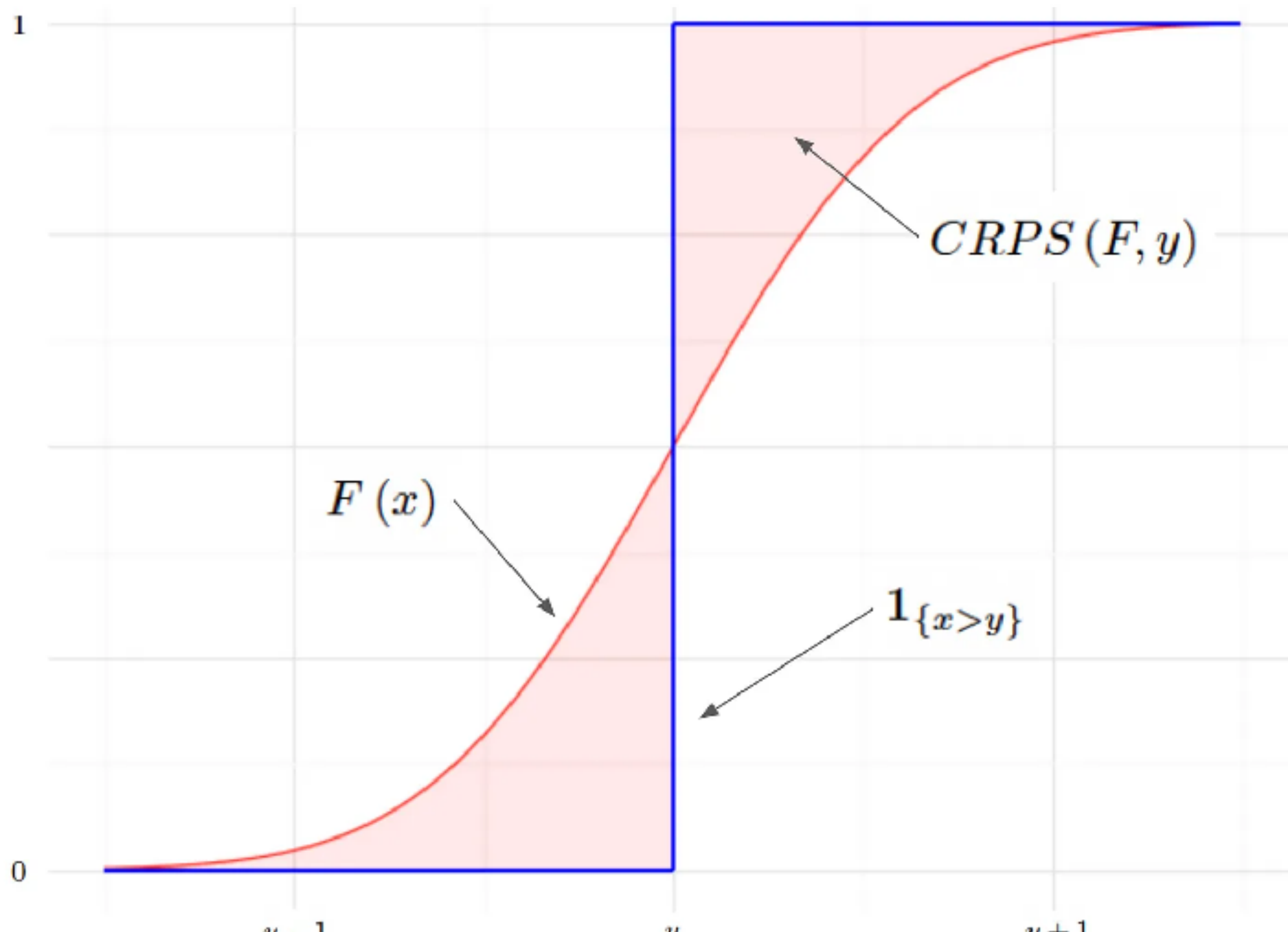
R: acf(x)

Brier score


Contribution BS_1 of *one* forecast to the total Brier Score
1.00
0.00
BS_1=
(prob-obs)^2
obs=rain
obs=dry
Issued forecast probability=70%
BS_1=0.49 if no rain
BS_1=0.09 if it rains
prob(p) 0.0 0.2 0.3 0.5 0.7 0.8 1 .0

Continuous Ranked Probability Score

$$\mathrm{CRPS}(F, x) = -\int_{-\infty}^{\infty} \left(F(y) - \mathbb{1}\{y \geq x\}\right)^2 \mathrm{d}y$$



Continuous Ranked Probability Score

$$\mathrm{CRPS}(F, x) = -\int_{-\infty}^{\infty} (F(y) - \mathbb{1}\{y \geq x\})^2 \, \mathrm{d}y$$

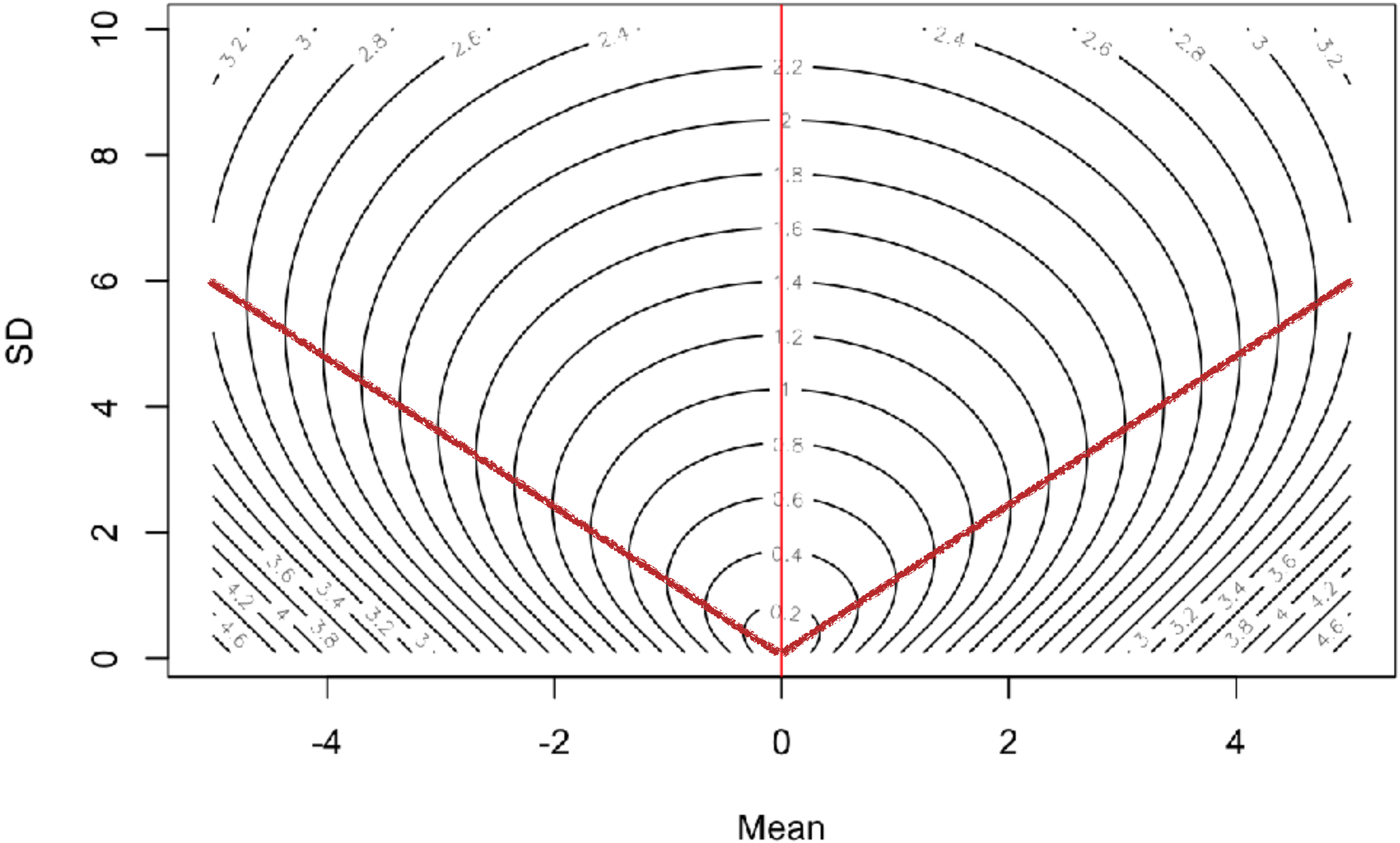
Ensemble member

Data

$$\mathrm{CRPS}(\hat{F}_m, y) = \frac{1}{m}\sum_{i=1}^{m} |X_i - y| - \frac{1}{2m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} |X_i - X_j|$$

Mean Absolute Error

Penalty for ensemble spread



https://github.com/eco4cast/neon4cast-scoring/blob/main/CRPS_example_JRT.Rmd

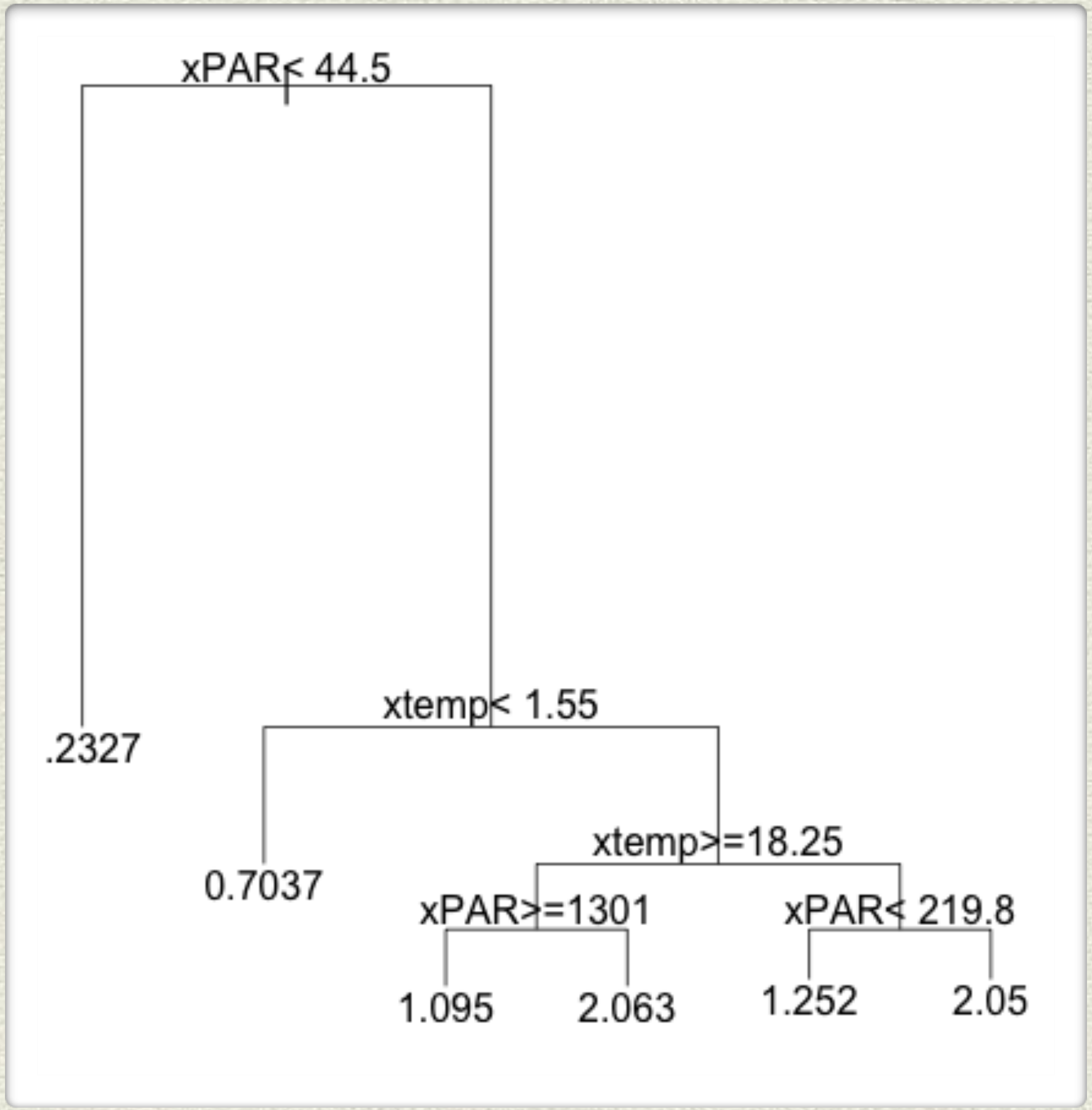
Data mining the residuals

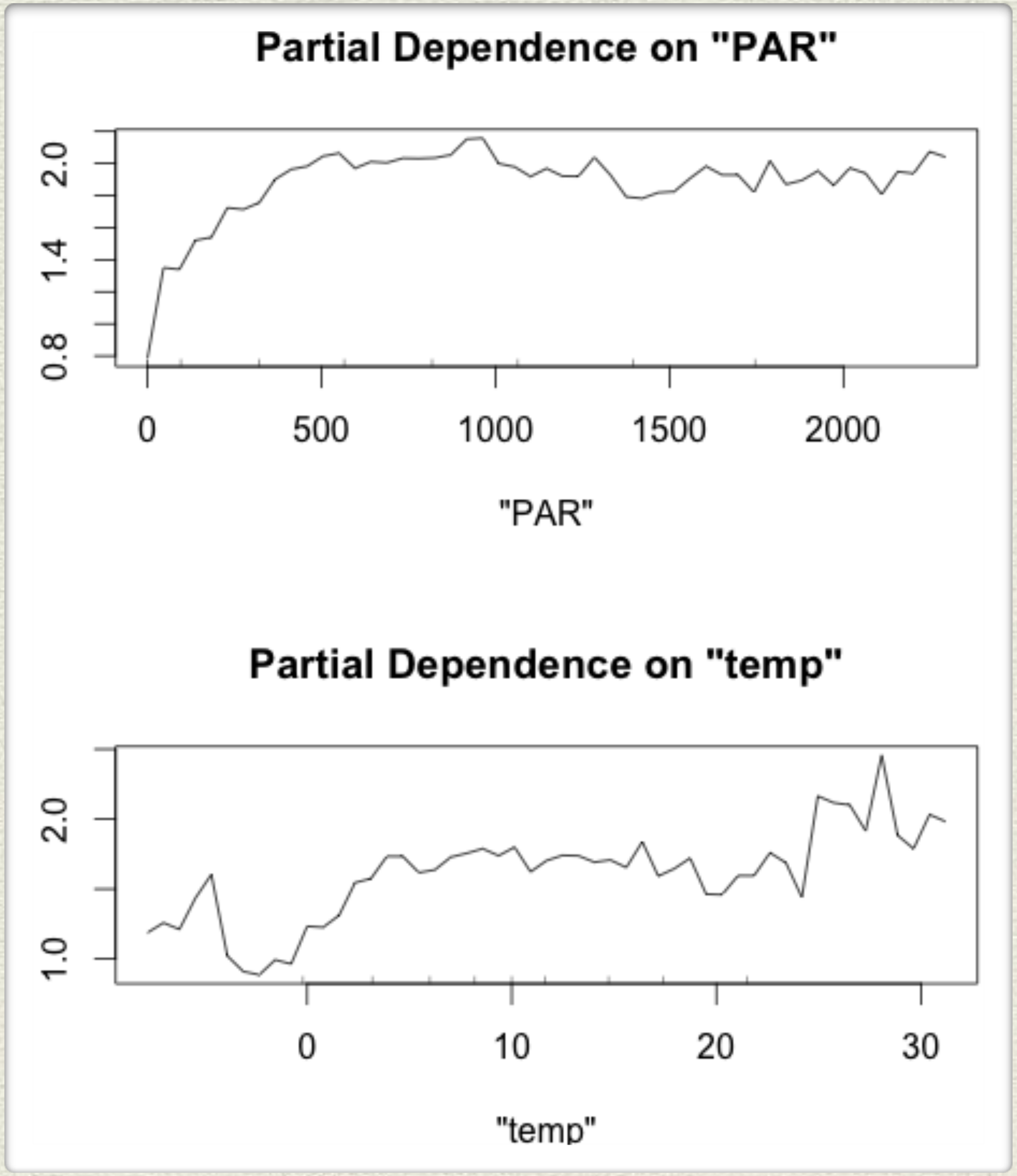
- Wide variety of Data Mining algorithms in use
- Large debate about use in process modeling and forecasting
- Potentially useful for generating hypothesis about when/where model fails

- CART
- GAM
- Random Forests
- Boosted regression trees / XGBoost
- Artificial Neural Network
- Deep Learning

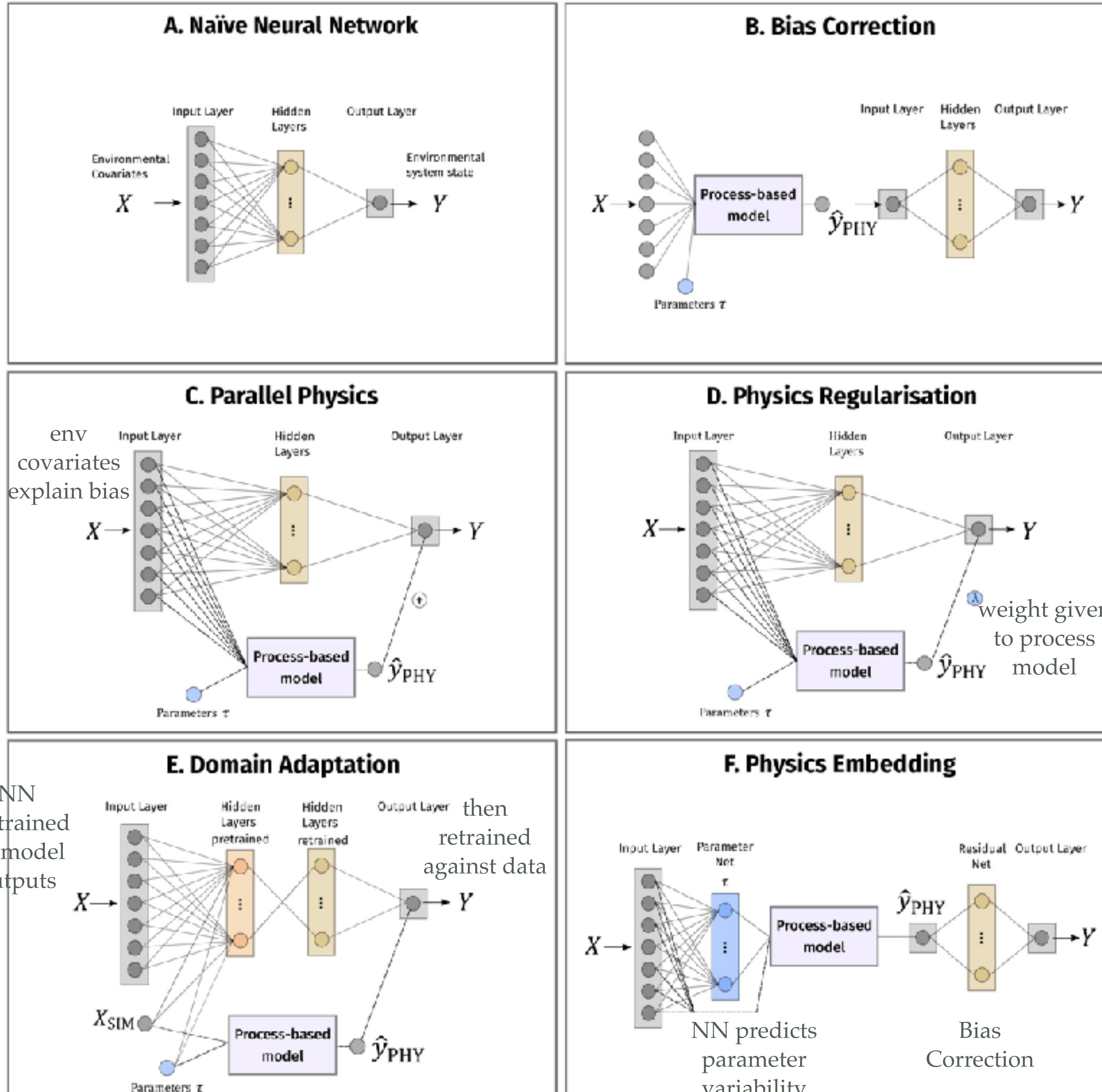
Random Forest

CART

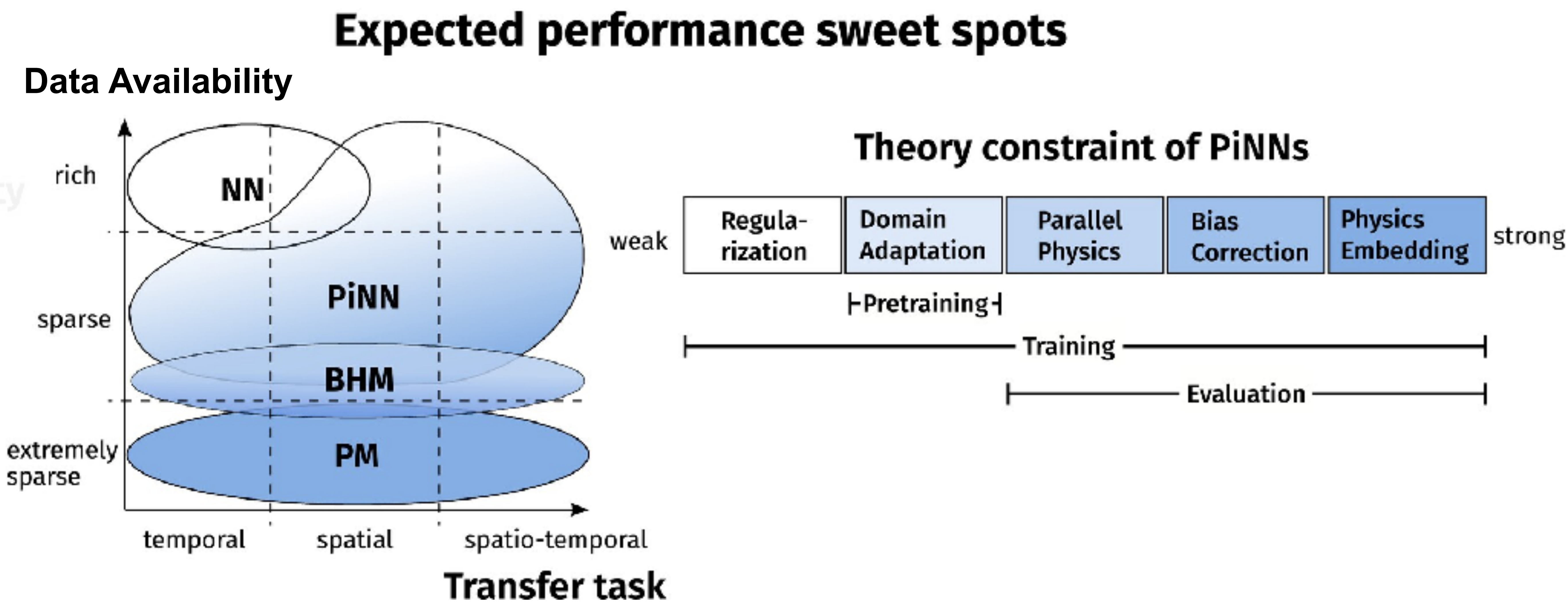




Hybrid Models (Process + NN)



Wesselkamp et al 2024 Ecology Letters



Wesselkamp et al 2024 Ecology Letters