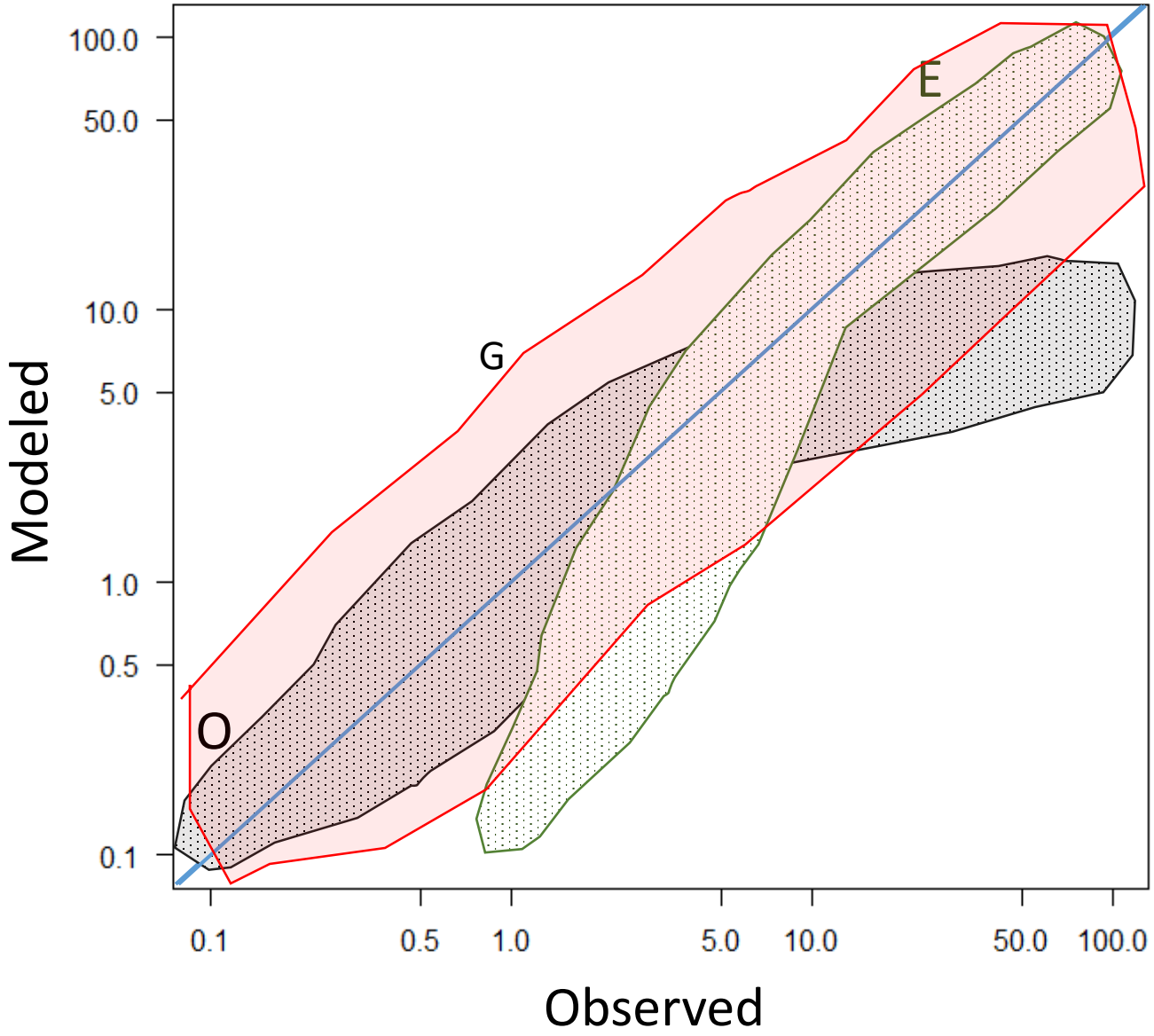
Which model is better?



Rick Stumpf
IOCS Breakout Session 7
Multi-water algorithms and Algorithm Performance Assessments

Needs and approaches in algorithm assessment

How do I decide to use or recommend a model?

What is the application?

How “good” is this model

Which model is better

- in general
- for my application

How good is the model

- Bias
- Variability/ precision
- Accuracy (bias + variability)

- The choice of which stat to use for these may depend on other conditions
  - Water type, oligotrophic, eutrophic, hyper-eutrophic
  - Seasonality, geography
  - Low/high sediment

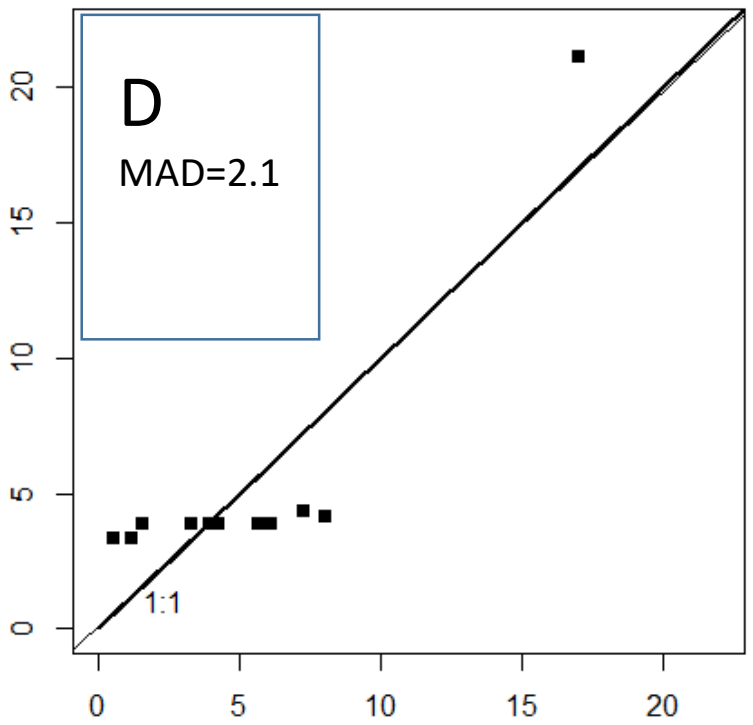
Which is the better model?

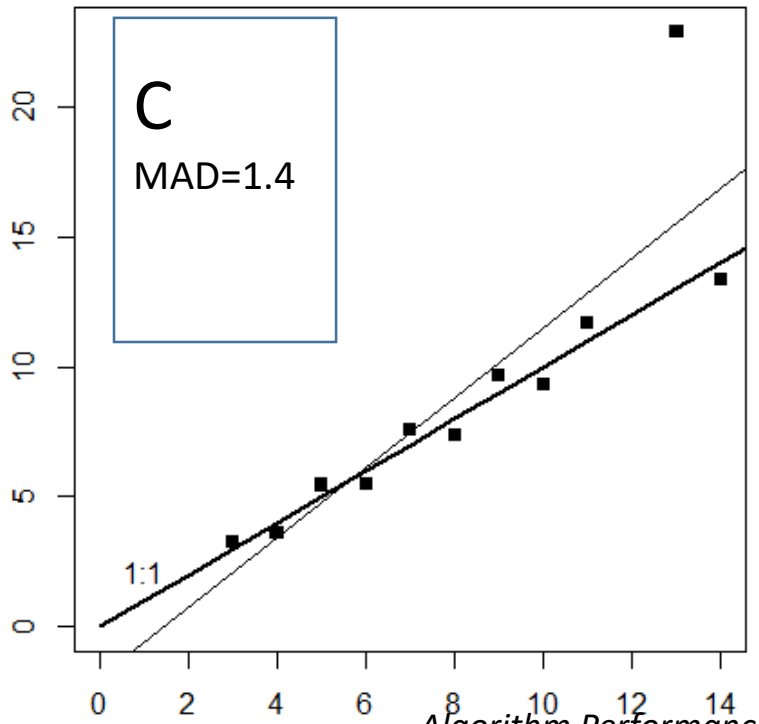
	Model D		Model C
R-squared	0.75		0.75
Slope	1.0		1.3
RMSE	2.5		3.0
bias	0		0.9

Graphics are essential

And

Mean square error stats are not robust





What stats should be used

Is the measurement uncertainty proportional to the measurement?

- Water temperature, no (water reflectance, probably not)
- Chlorophyll, yes

- Most error distributions are not truly Gaussian distribution
  - outliers are common in optical (e.g. high chlorophyll)
  - Long tails are common in others (e.g. temperature)

Type of uncertainty/error impacts choice of analysis

	Bias	Variability (Precision)	Accuracy
Proportional error	Multiplicative bias (from log) (MeanPE ok for small biases & non-negative)	Remove bias	Absolute multiplicative error (from log) Absolute PE ok if non-negative
Uniform error	ME (mean error)	Remove bias in calculation	Mean Absolute Error

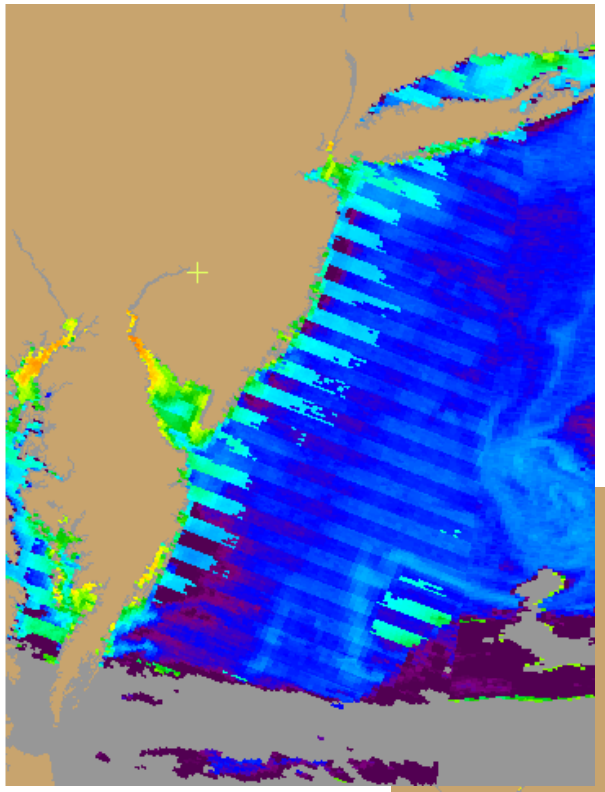
Need to use statistically robust metrics.

Robust statistics work well for a variety of distributions

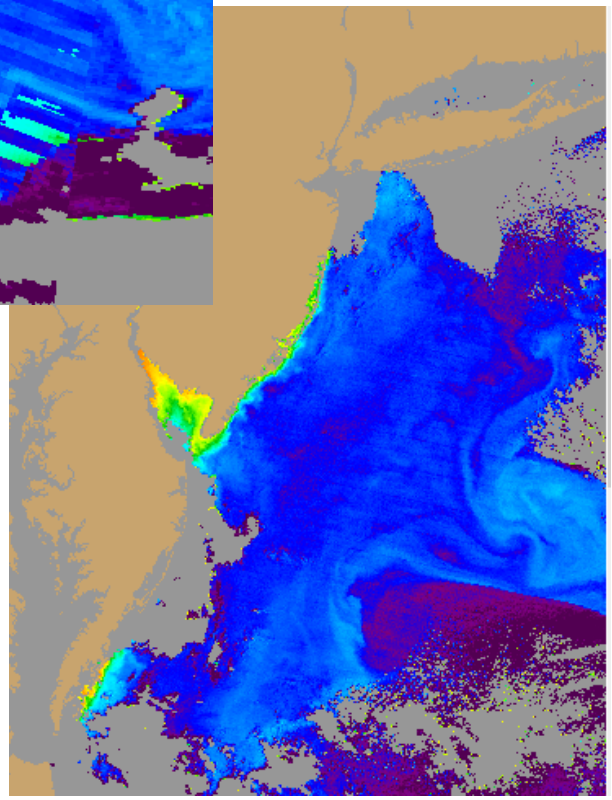
The Gaussian ("normal") distribution is not common in our data (outliers, fat tails)

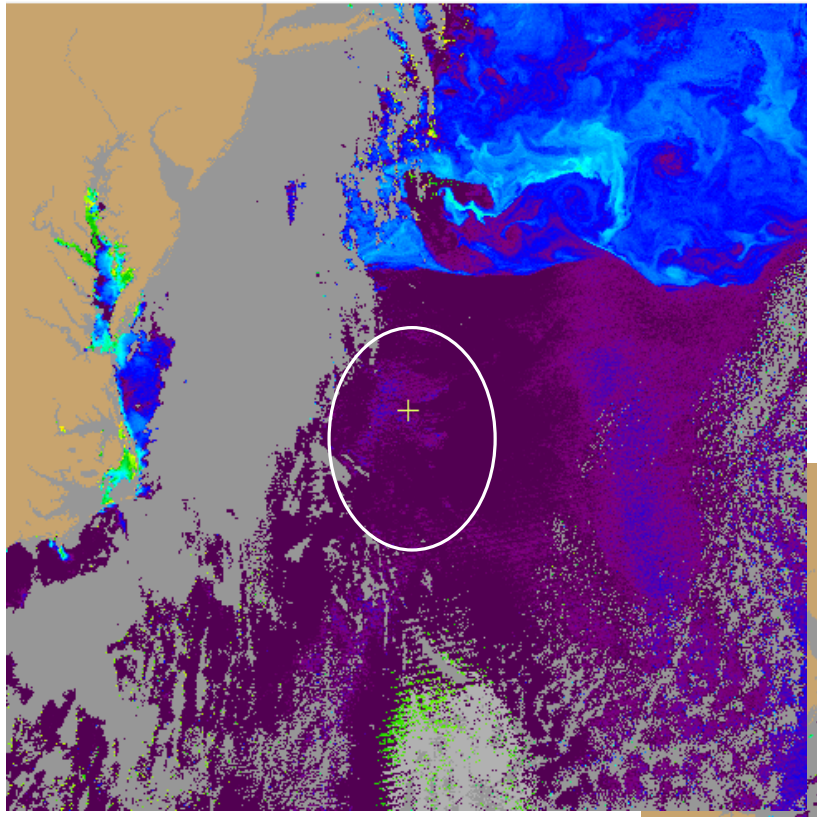
If Bill Gates walked into this room the average income would be $1 million and standard deviation $100 million

Temporal variability also matters

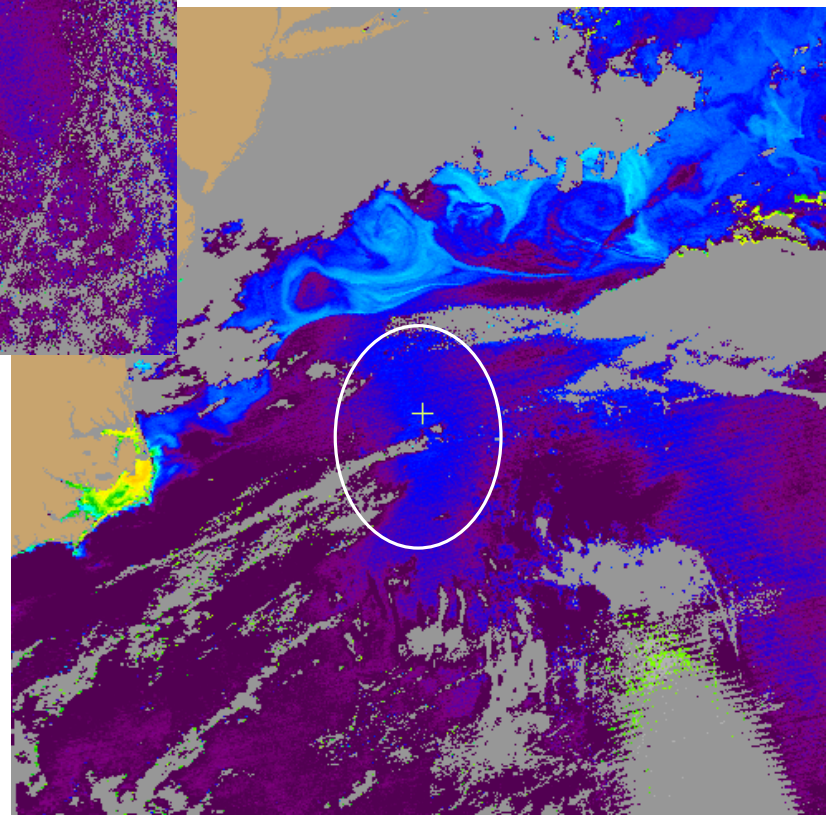


Terra Rrs667
April 18, 192017



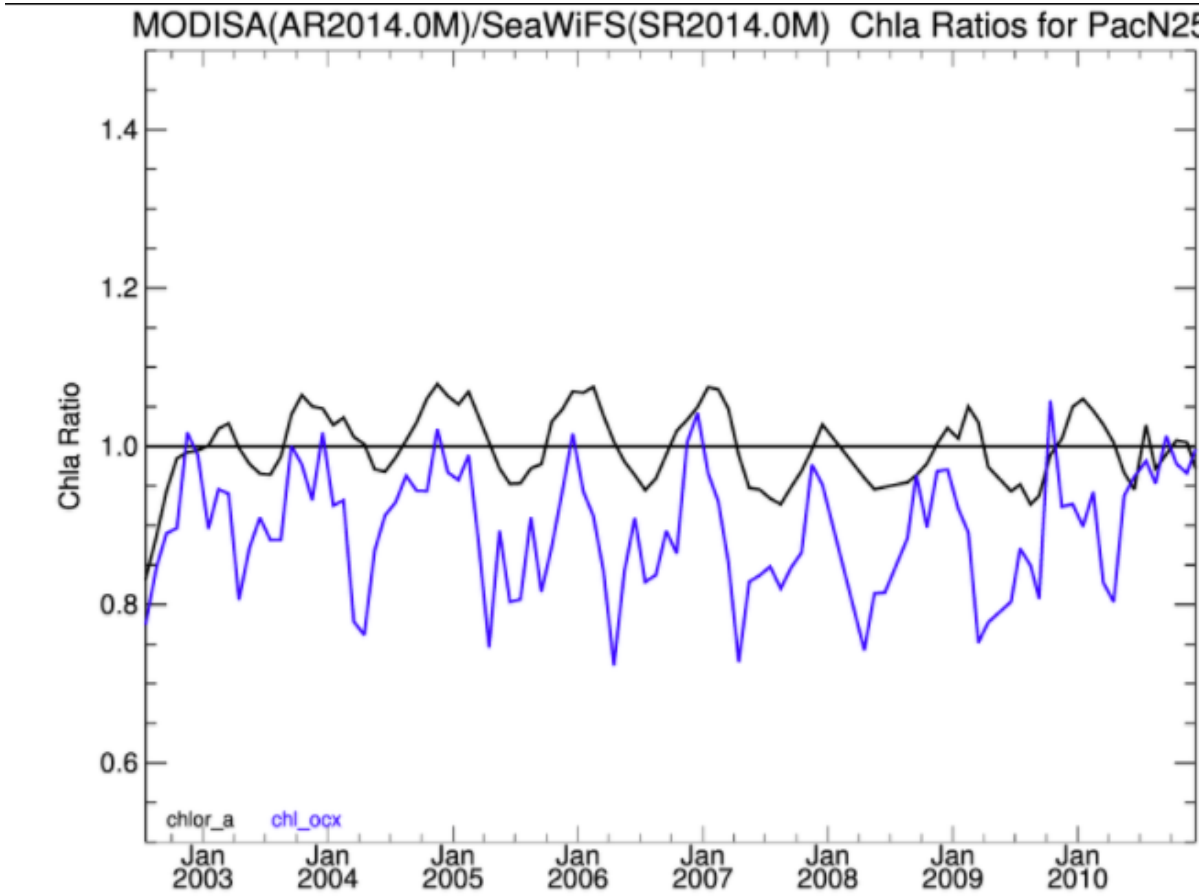


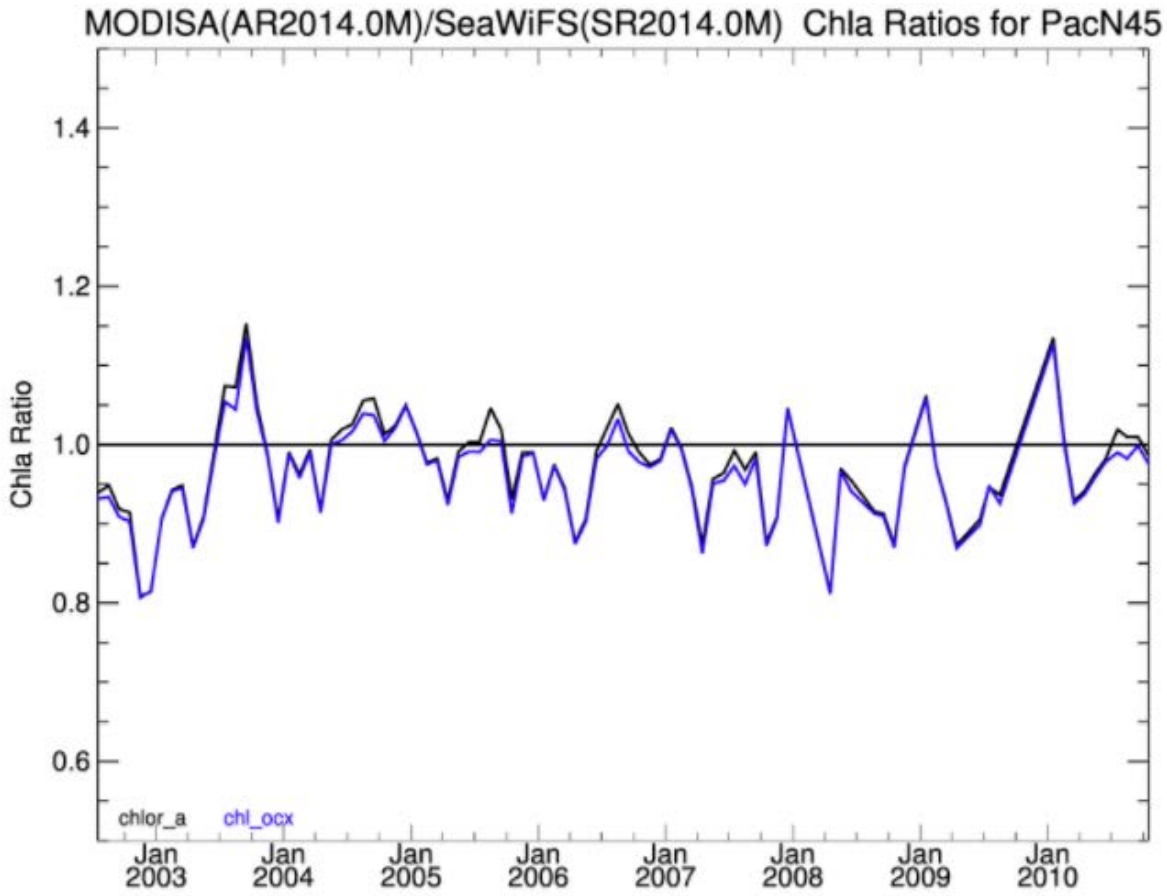
Aqua Rrs667
April 15, 17 2017



Misfit is not random,
Season and latitude can change results substantially

NASA OBPG oceancolor.gsfc.nasa.gov/analysis/global



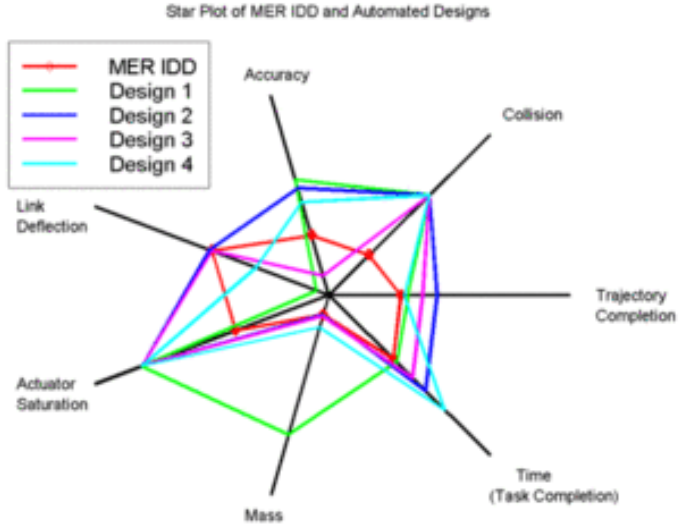


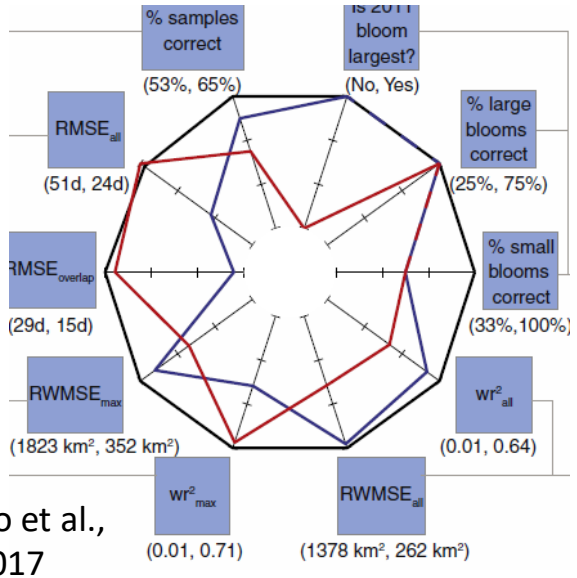
Model comparison

- Direct comparison of observations. Pair-wise comparison (Broomell et al. 2011)
  - Model that has the most wins is better
  - long history in decision metrics.(“voting” method of Condorcet (1785))
  - Addresses model failure directly

- Multi-criteria methods and **graphics**
  - Use the criteria that matter (e.g. errors in different water types)
  - Star plots (also called radar plots) Chambers, Cleveland et al. 1983
    - (Taylor plots have been used for ecological models)
  - Numeric (rank the criteria that matter) **Brewin et al. 2015 RSE**
    - Condorcet (Broomell et al.) quite effective for this purpose

- Avoid redundant metrics
  - one each (bias, precision, accuracy)
  - then type (oligotrophic, eutrophic, etc.)





Ho et al., 2017

Brewin et al., 2015 Remote Sensing of Environment.
Broomell et al., 2011, Pair-wise comparisons of multiple models, Judgment and Decision Making
Chambers, Cleveland Kleiner Tukey, 1983, Graphical Methods for Data Analysis.
Condorcet MD. 1785, Essay on the Application of Analysis to the Probability of Majority Decisions.
Paris: Imprimerie Royale. 1785.
Ho et al., 2017; Remote Sensing of Environment

Which model is better?