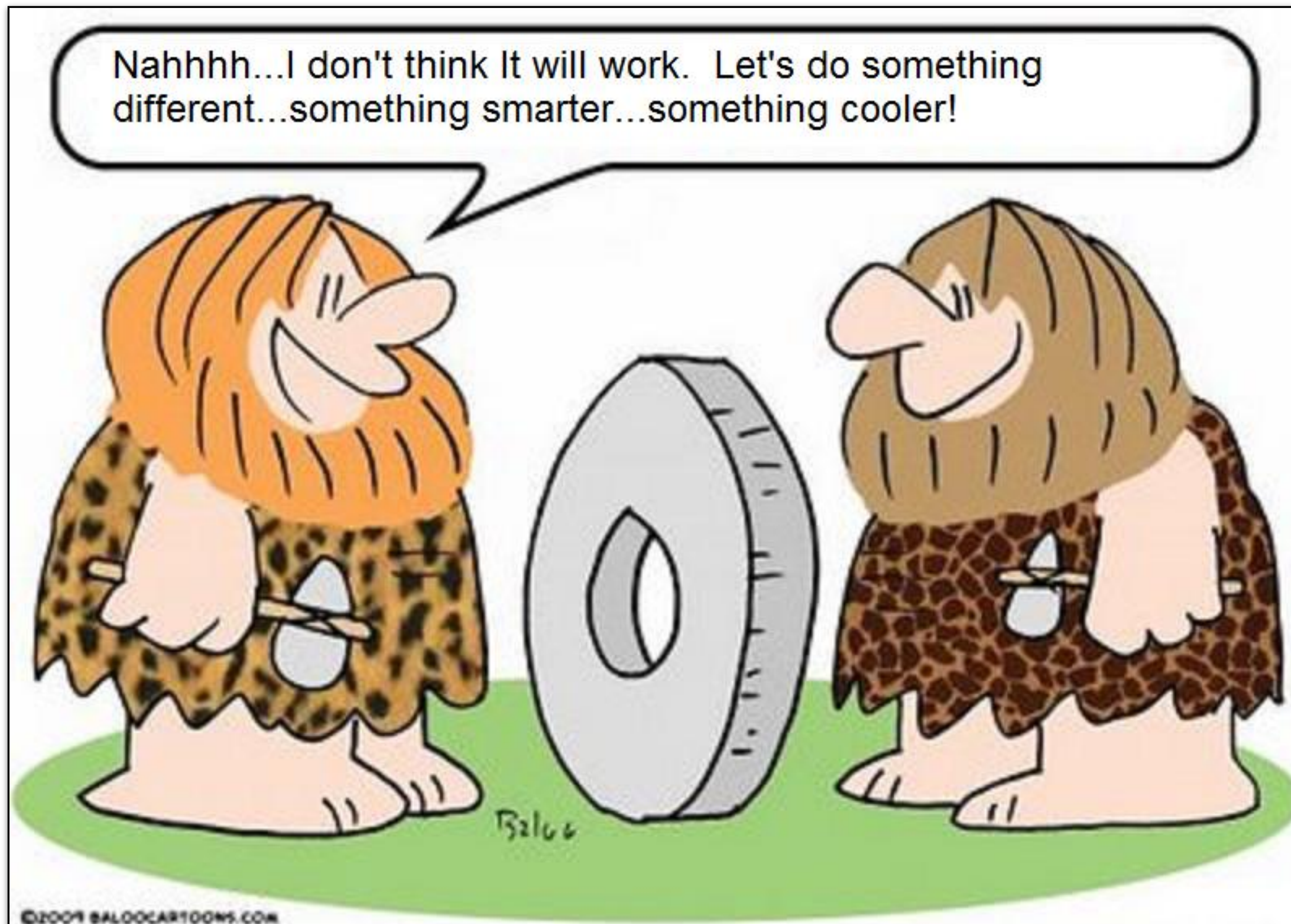
Off-the shelf models

Useful because they usually have had multiple experts developing, testing and refining over time so you don't have to re-invent the wheel



Model Selection Criteria: Appropriate

Does it represent the processes that you are interested in:

- outputs (e.g. streamflow, ET, N-export, forest water use)
- temporal and spatial resolution appropriate for the questions you are asking
- does it account for the mechanisms that are likely to be important in the questions you are asking - (examples)

Is it APPROPRIATE

Then pick the simplest model that accounts for your processes of interest

Model Selection Criteria: Appropriate

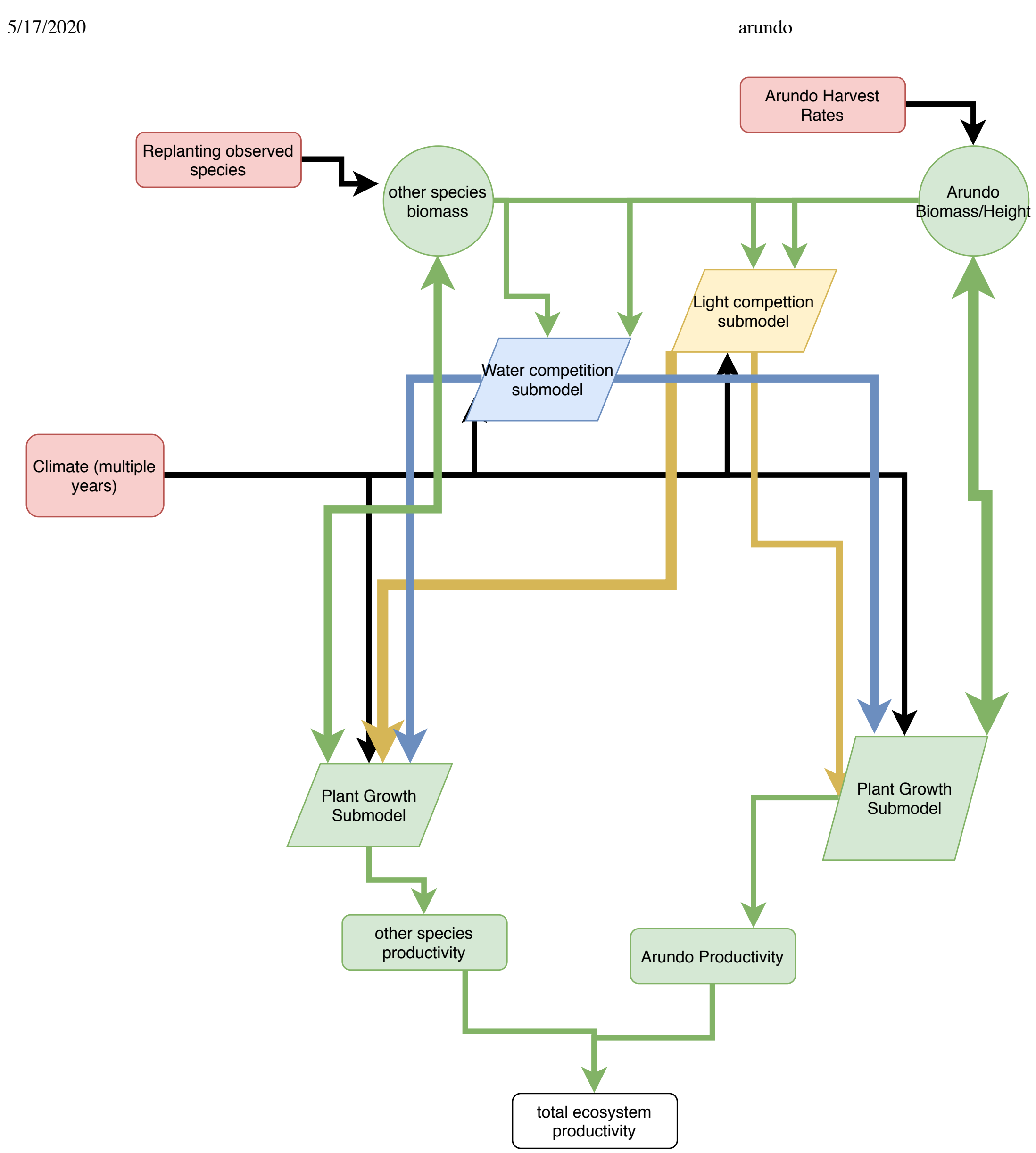
How do I know if the model is appropriate

- Develop your conceptual model first - include inputs/output, spatial temporal resolution, processes/interaction
- Compare candidate models - their conceptual models to the one you developed (does it have the interactions you need)
- Sources: documentation, previous application

Is it APPROPRIATE

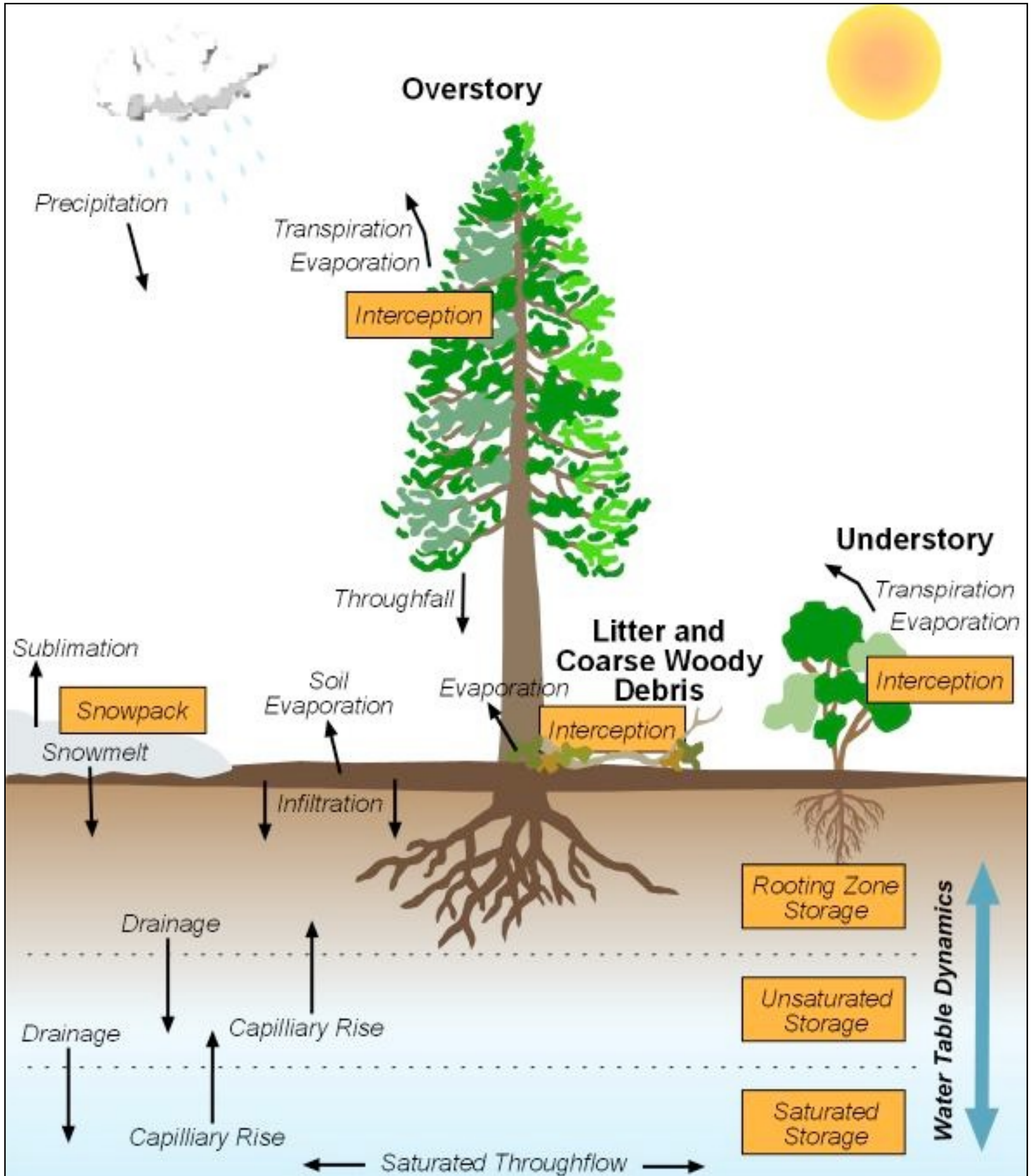
Then pick the simplest model that accounts for your processes of interest

Model Selection Criteria: Appropriate

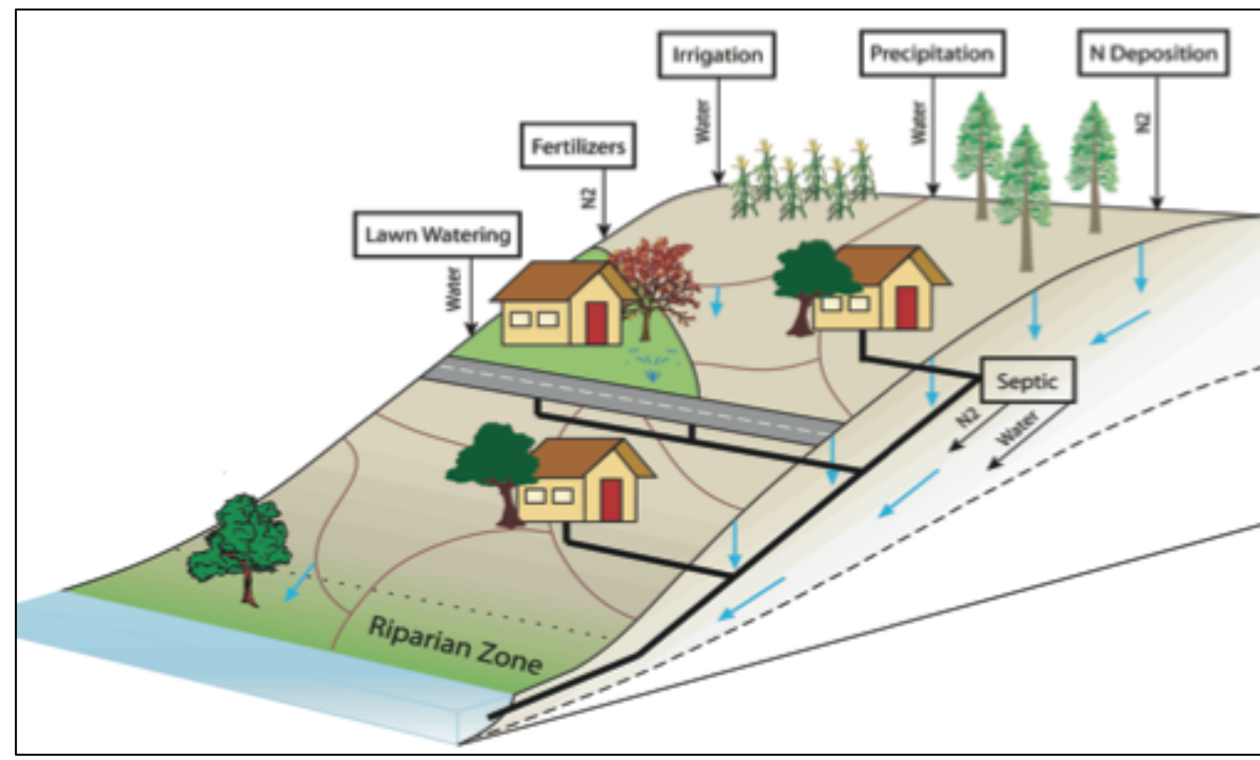


Consider a project that seeks to determine how arundo invasions will change through time with climate and how harvesting and planting other species can influence this

Hydrologic model example



RHESSys based model of water, carbon, and nitrogen cycling model - but no fire!



Fire model

https://www.firelab.org/project/flammap

Model Selection Criteria: Feasible

* Can you obtain (or approximate) required inputs/output
* Complexity - do you have time and expertise to run the model
* Computational resources (do you have computing resources to run it)

Model Selection Criteria: Performance

Does it capture the outputs of interest (and relationships between inputs/parameters and outputs) with sufficient *accuracy* to answer your research questions

Can it do so reliably (across scenarios you will run)

IS IT GOOD ENOUGH?

Model Selection Criteria: Performance

IS IT GOOD ENOUGH?

Challenges - do you have data for performance evaluation?

If NOT

* Performance evaluation in similar location
* Captures expected relationships from science theory
* Comparison with other models
* Science-literature

Model Validation (Is it reasonable)

- Compare model results to simple thought experiments

- Similar to testing: conservation of mass, energy, behaviors under known conditions (zero rain = zero streamflow; zero CO2 change = zero T change)?

- Are the values for outputs physically reasonable (e.g snowpack > 0, reservoir storage < reservoir)

Validation (is it accurate?)

Compare model results to observations
(either local or non-local)

How good is good enough?

Validation (is it accurate?)

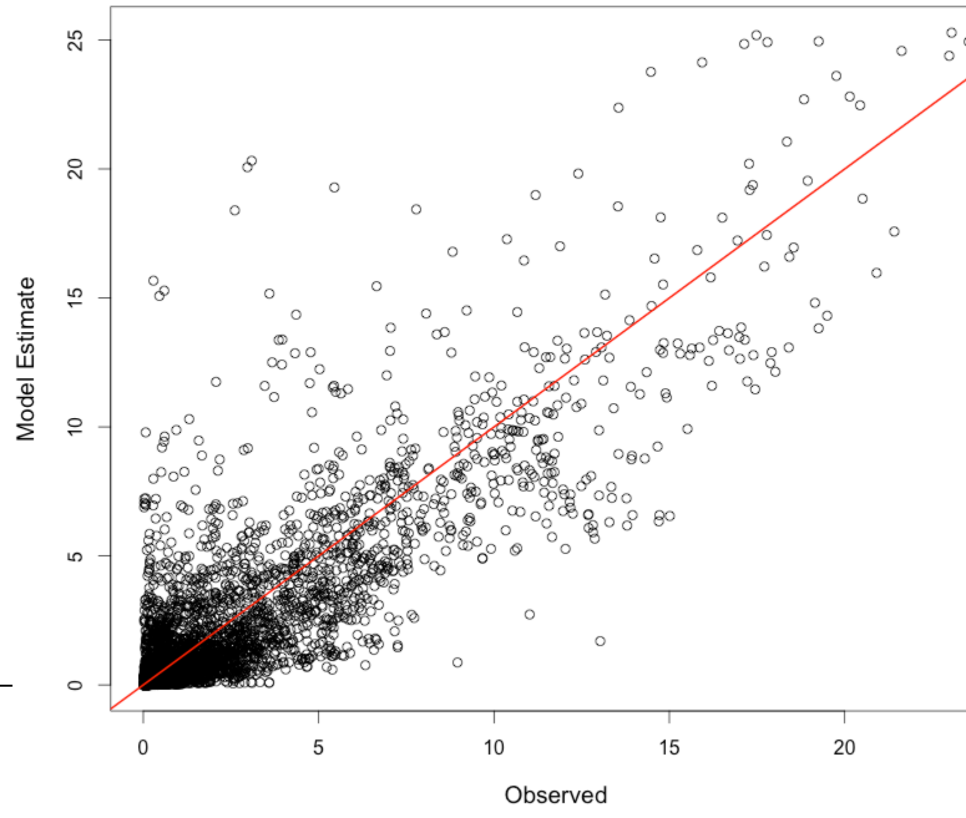
Compare model results to observations

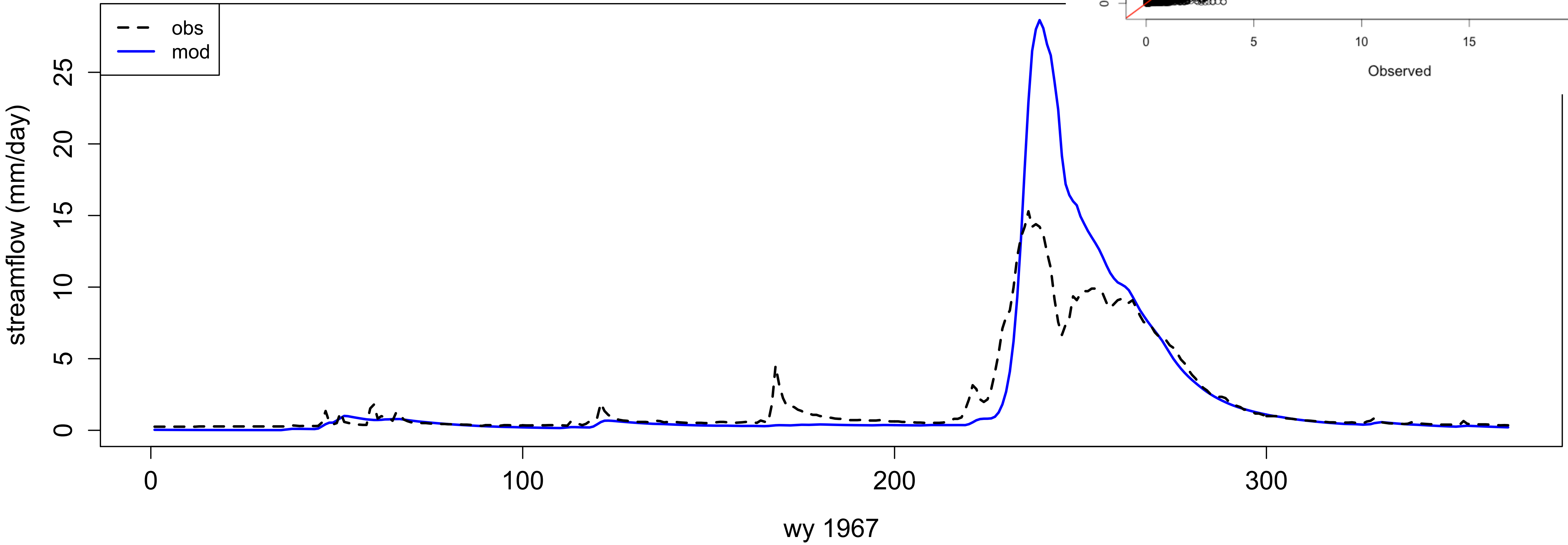
- Observations from the same site/scenario/ circumstance
  - streamflow from a rainfall-runoff model applied to a given watershed
  - estimates of population growth after an actual disturbance
- Observations of general patterns, relationships
  - ranges of precip/streamflow for watersheds in that region
  - ranges of population growth after similar disturbances

Model Performance

PLOT: First Step!!

Model and observed through time
Error through time
Model versus observed





Validation (is it accurate?)

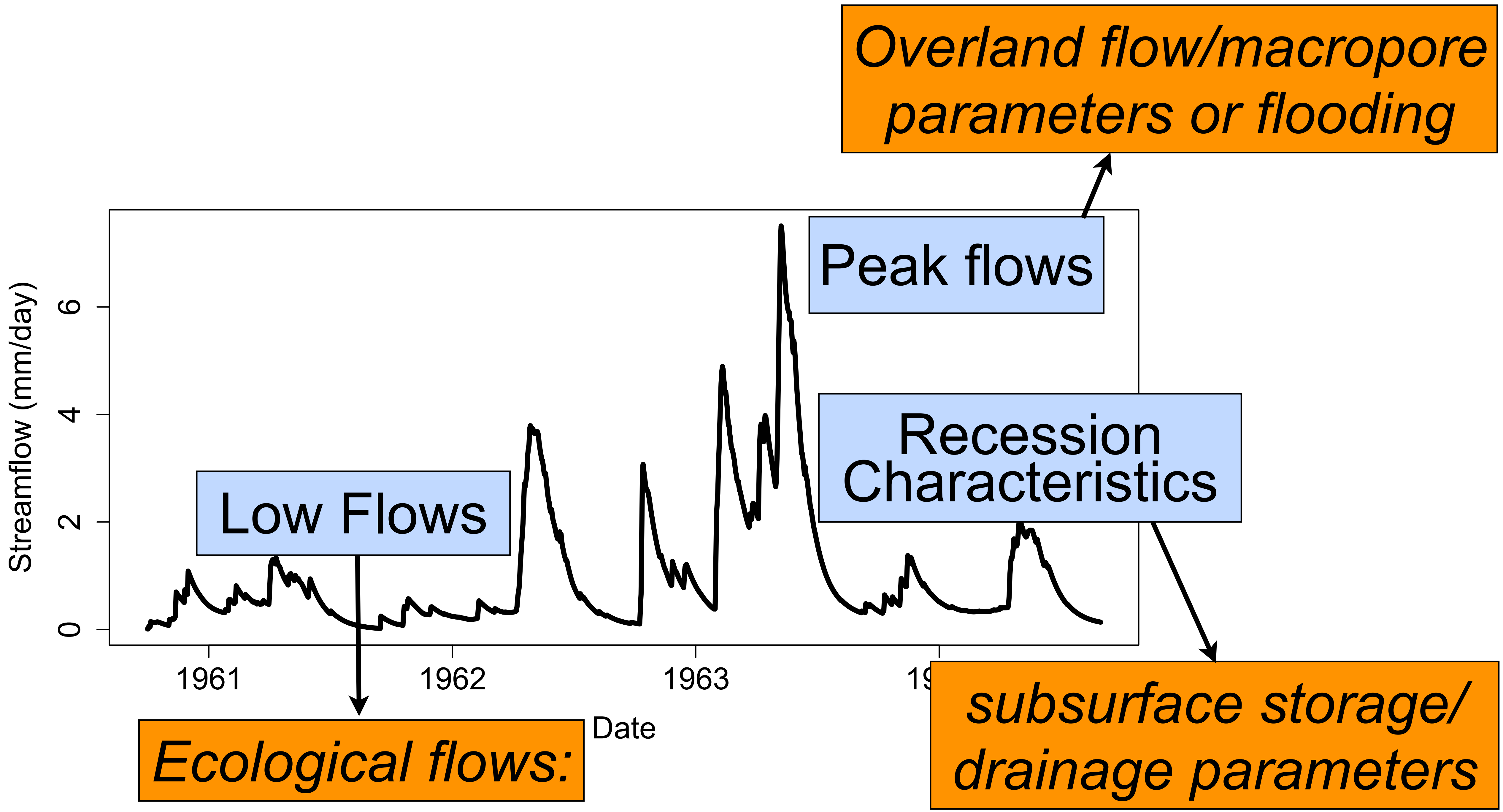
IS IT GOOD ENOUGH?

MODEL + INPUTS/
PARAMETERS

What feature of outputs are important to you?

Information content of observations:

What you try to get "right" depends on the processes that the parameters influence and your use of the model



Brutsaert, W., and J. P. Lopez (1998), Basin-scale geohydrologic drought flow features of riparian aquifers in the Southern Great Plains, Water Resour. Res., 34(2), 233–240, doi:10.1029/97WR03068..

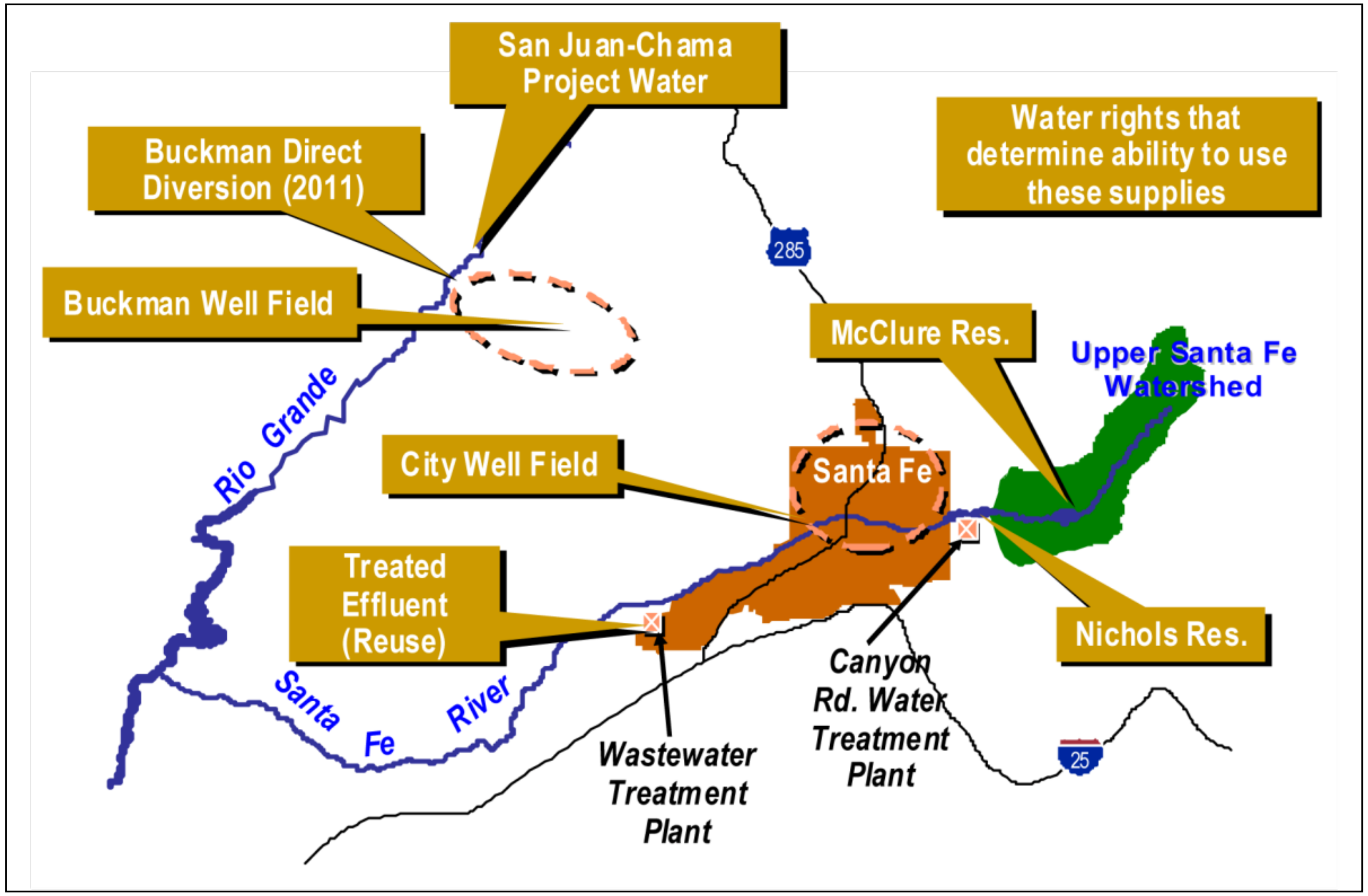
Summary of criteria for model selection

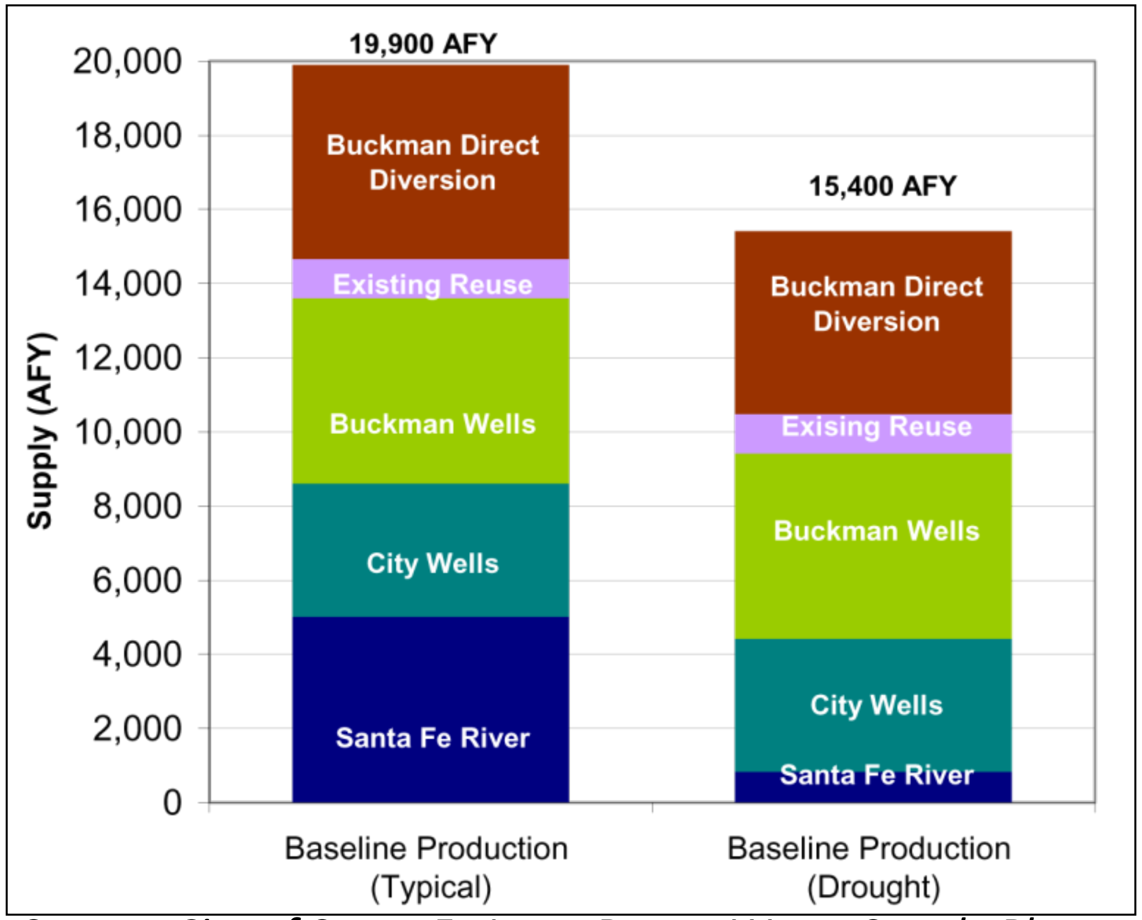
- Error/uncertainty in variable of interest (or response of interest) is small relative to use of model results for decision making

- Errors/uncertainty is small relative to simulated effect of change or relationships of interest

- Similar levels of performance by other models/studies reported in the literature - the “state of the art” argument - *or “at least its better than the other one”*

Change in flow at which Santa Fe water-managers would consider purchasing additional water rights or wells



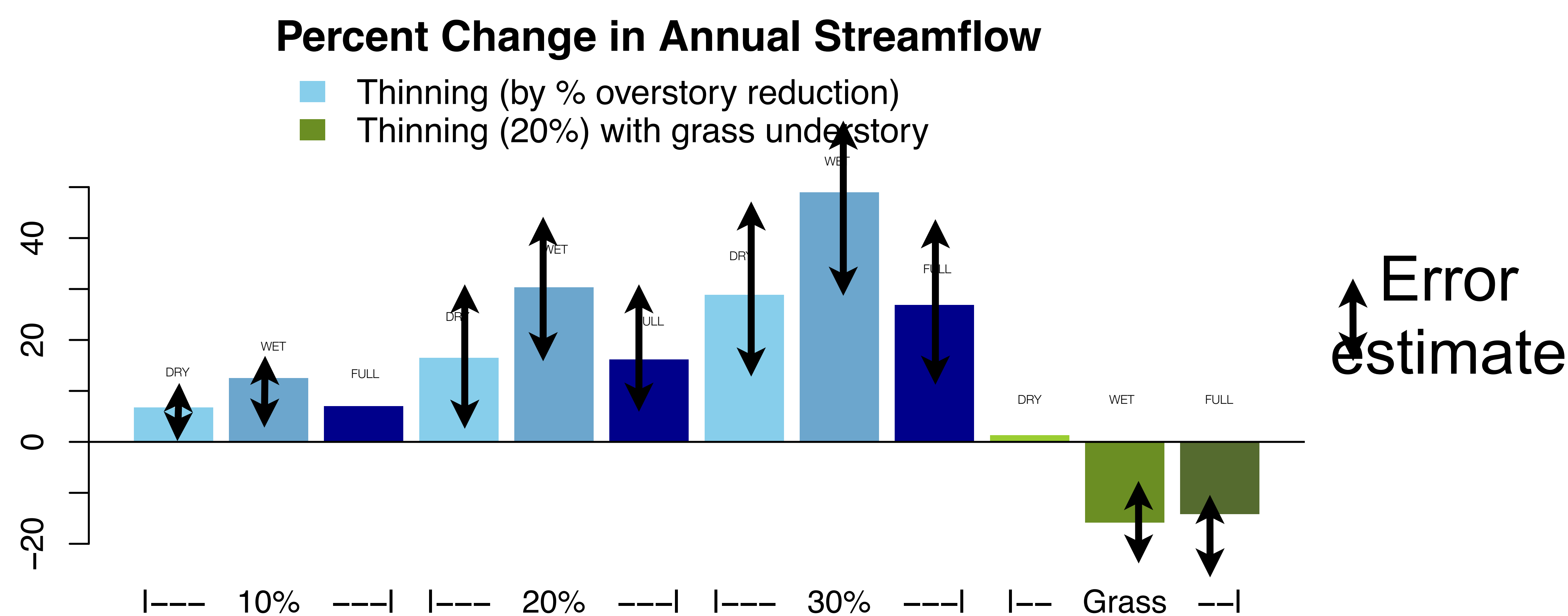


Source: City of Santa Fe Long-Range Water Supply Plan, 2008.

Supply < 2000 AFY, costs X dollars to “buy” additional water..., what is the cost effective decision given climate change estimates for the next decade?

What is good enough?

Errors/uncertainty is small relative to simulated effect of change or process of interest



adapted from Dugger et al., 2013

Could a grass understory impact how streamflow changes with overstory reduction (thinning)?

What is good enough?

2b. Demonstration of improved performance

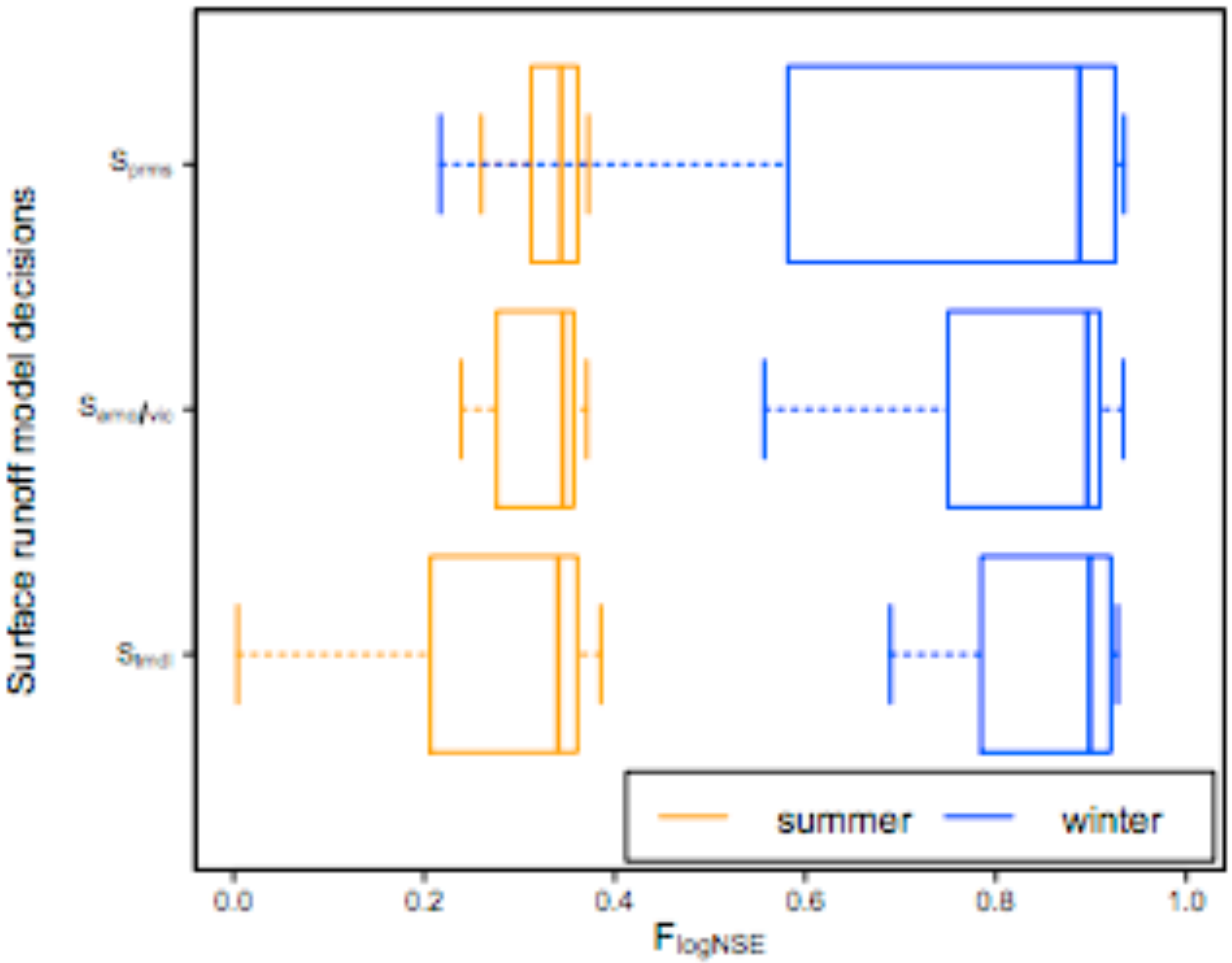


Fig. 8. Boxplots of model performance for summer and winter streamflow simulations for the three surface runoff decision options.

Plot demonstrates the effect of different model structures (surface runoff assumptions adapted from VIC, PRMS, TOPMODEL) on performance

Staudinger, M., Stahl, K., Seibert, J., Clark, M. P., and Tallaksen, L. M.: Comparison of hydrological model structures based on recession and low flow simulations, Hydrol. Earth Syst. Sci. Discuss., 8, 6833-6866, doi:10.5194/hessd-8-6833-2011, 2011.

Model selection criteria: Summary

- Is it appropriate (processes, spatial-temporal scale)
- Is it feasible (time, expertise, data)
- Is its performance good enough to answer your question
  - ‘laugh test’ - makes sense
  - Local evaluation (if possible)
  - Alternatives - Science-literature, evaluation in similar setting, comparison with other models

Model selection criteria: Performance

* Choose metrics for evaluation
    - what ‘matters’ to you - what do you need to get right?
    - average or extremes?
    - total, minimums, maximum?
    - variation? (at what time/space scales)?
    - relationships (e.g temperature vs productivity)

First choose what you care about

Then define your metric (e.g correlation coefficient of what?)

Performance Measures

Root Mean Square Error (RMSE)

$$SSE = \frac{1}{n}\sum_{i=1}^{n}(m_i - o_i)^2$$

$$RMSE = \sqrt{SSE}$$

Nash Sutcliffe Efficiency (NSE)

Nash and Sutcliffe, 1970, J. of Hydrology

Widely used in hydrology

Range – infinity to +1.0

Overly sensitive to extreme values

$$NSE = \frac{\sum_{i=1}^{n}(o_i - m_i)^2}{\sum_{i=1}^{n}(o_i - \bar{o})^2}$$

BIAS or Percent Error (Err)

Useful for determining if there is a long term flow over or under estimation

$$Err = \frac{(\bar{m} - \bar{o})}{\bar{o}} * 100$$

*Others: Cor, R^2

Model Performance: implement as a function in R

```r
#' nse
#'
#' Compute NSE between observation and model
#' @param  m  model estimates
#' @param  o  observations
#' @return nse


nse = function(m,o) {

err = m-o
meanobs = mean(o)
mse = sum(err*err)
ovar = sum((o-meanobs)*(o-meanobs))
nse = 1.0-mse/ovar

return(nse)
}
```

Model Performance

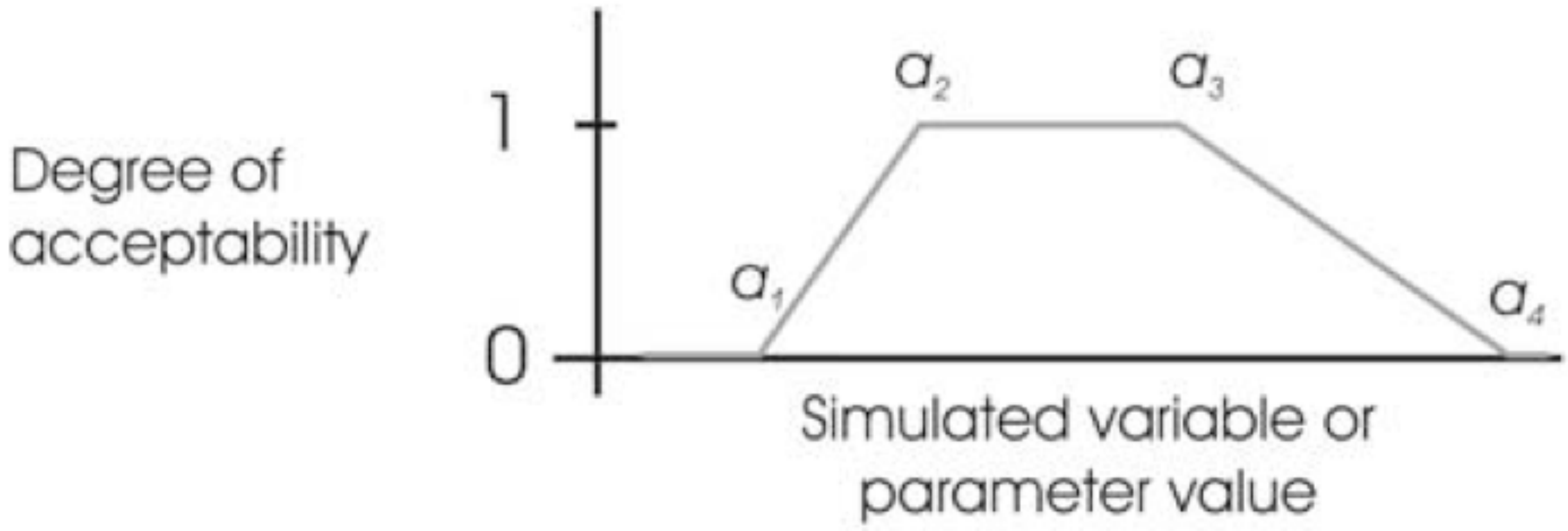
```r
#' relerr
#'
#' Compute percent error between observation and model
#' @param  m  model estimates
#' @param  o  observations
#' @return relerr


relerr = function(m,o) {

err = m-o
meanobs = mean(o)
meanerr = mean(err)

res = meanerr/meanobs
return(res)
}
```

Soft metrics - Fuzzy-Evaluation



$$\mu(x) = \begin{cases} 0 & \text{if } x \le a_1 \\ \dfrac{x - a_1}{a_2 - a_1} & \text{if } a_1 \le x < a_2 \\ 1 & \text{if } a_2 \le x < a_3 \\ \dfrac{a_4 - x}{a_4 - a_3} & \text{if } a_3 \le x < a_4 \\ 0 & \text{if } x > a_4 \end{cases}$$

For data where there is a lot of uncertainty in observed values (imprecise measurements)

Seibert, J., and J. J. McDonnell, On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, Water Resour. Res., 38(11), 1241, doi:10.1029/2001WR000978, 2002.

Performance: Metrics

R also has built in functions that can be helpful

- cor(x,y) - correlation coefficient
- help(cor)

Model Performance

Combining objective functions

Multiplicative approach

Metric A * Metric B …

Metric A * weighting A + Metric B * weighting B…

Requires metrics to be normalized (same range)
- 0-1
- divide by maximum value

Metrics must all work in the same direction
increase = better OR decrease = better

Performance Metrics

- If you are combining = need to increase with "better" models

- Transform metrics like RMSE that work in reverse

$$SSE = \frac{1}{n}\sum_{i=1}^{n}(m_i - o_i)^2$$

$L = (SSE)^{-n}$ where n is a shaping parameter
(Freer et al., (1997)

$L = \exp(-nSSE)$

$L = (\max(RMSE) - RMSE) / (\max(RMSE) - \min(RMSE))$

$$\text{relErr} = \frac{(\bar{m} - \bar{o})}{\bar{o}}$$

Transform to 0-1,
and positive

mErr = 1.0 - min(1.0, abs(relErr))

mErr 1.0 - min(1.0,abs(relErr)/max(abs(relErr))

Combining

cperf = mErr * max(NSE, 0)
cperf = 0.75*mErr * 0.25*max(NSE,0)

Model Performance

```
#' cper
#'
#' Compute a performance measure (0-1) between observation and model
#' based on both NSE and relative error
#' @param  m  model estimates
#' @param  o  observations
#' @param  weight.nse weighting to give NSE metric
#' @param  weight.relerr weighting to give relative error metric
#' @return  combined 0-1 performance measure


cper = function(m,o,weight.nse=0.5, weight.relerr=0.5) {

nse = nse(m,o)
mnse = max(nse,0)

rel.err = relerr(m,o)
merr = 1.0-min(1.0, abs(rel.err)/max(abs(rel.err)))

combined = weight.nse*mnse + weight.relerr*merr

return(combined)

}
```