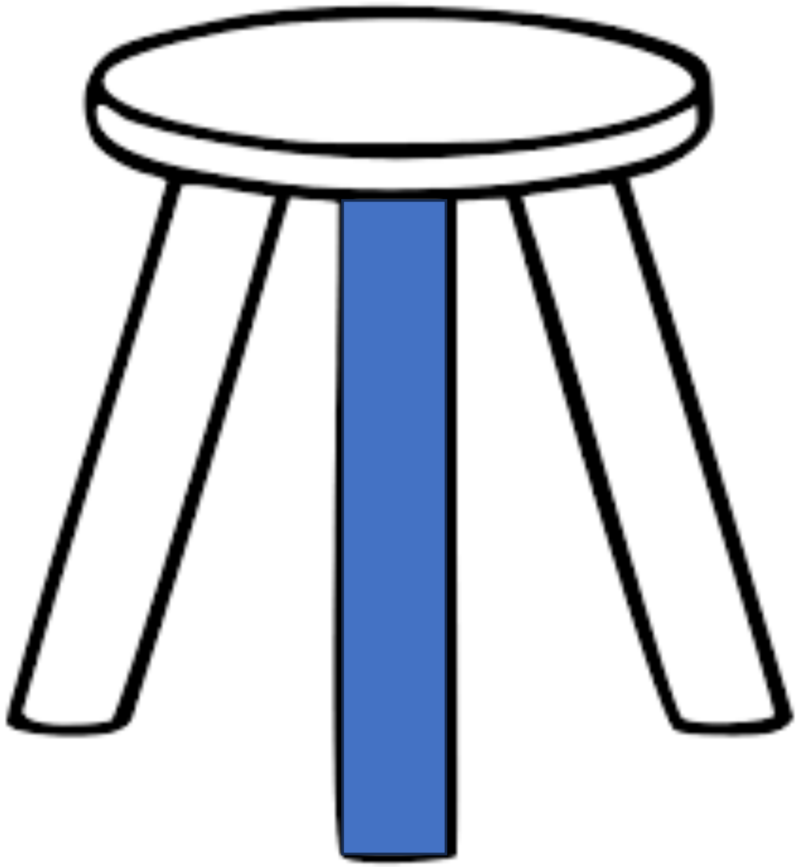
Machine learning and mechanistic modeling in hydrology: successes and ongoing challenges

Laura Condon

The University of Arizona

Outline/Goals:

- Tell you a bit about me and what I do
- Provide a very broad-brush overview of hydrologic challenges*
- Starting point for discussion on interdisciplinary opportunities in IT and Hydrology



Domain Science: Hydrology

* For the record we have solve a lot of problem too… but I’m mostly going to focus on challenges here.

A bit about me

Water sustainability and watershed dynamics

- Large scale distributed physically based models that simulate groundwater and surface water together

Improving the ways we use and learn from process-based models

- What information can we learn from these models that we can't get from other sources?

Building tools and platforms to make modeling and data products more accessible

- Lowering barriers to entry for analysis and expanding the ways we can use these tools

Translating science to practice and learning from water managers and water users

- Finding solutions to problems that matter

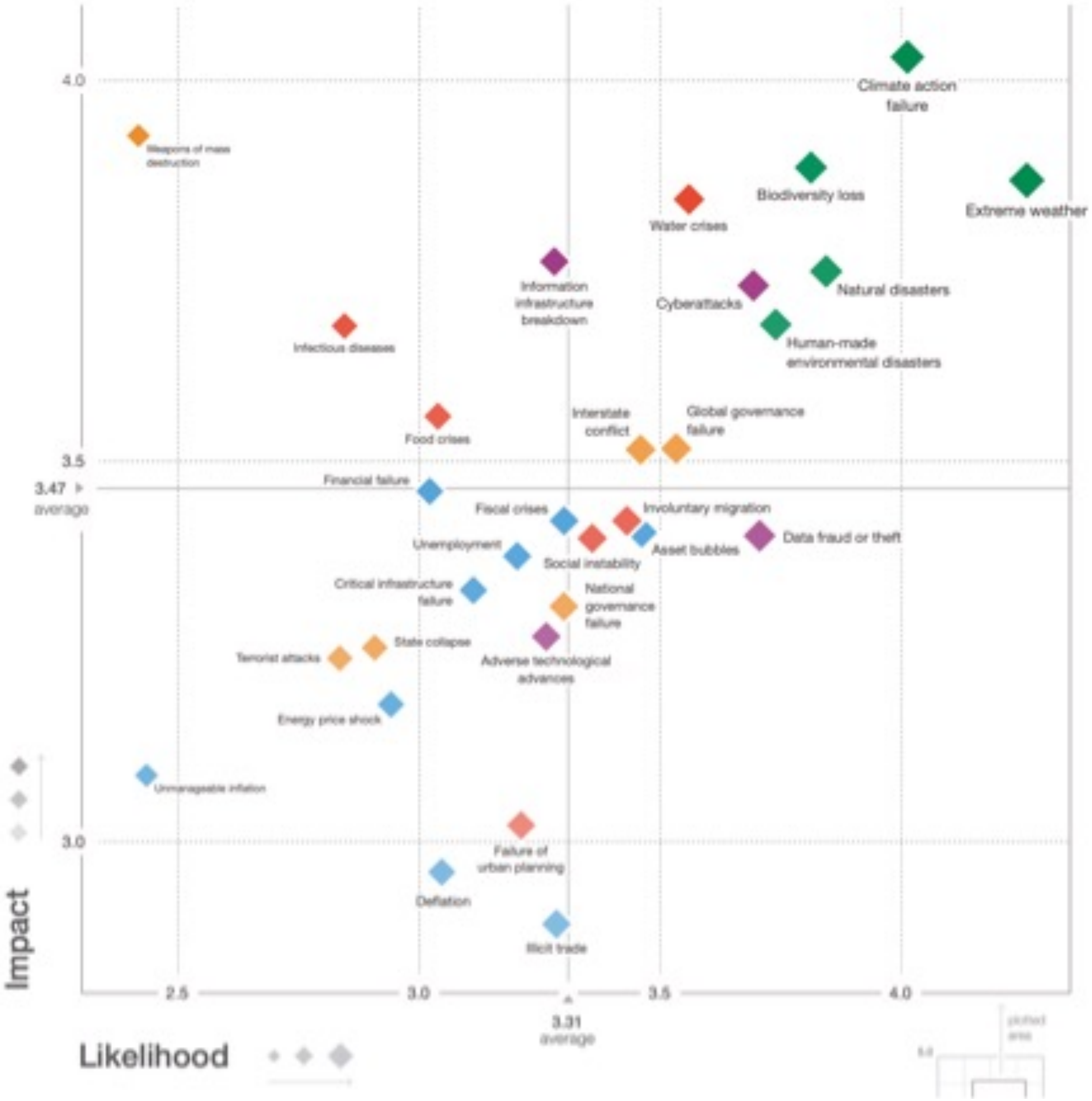
What is the job of hydrologists?

To understand and predict the state of water, and its response to past and future perturbations and forcings across terrestrial systems

This is not an abstract calling:

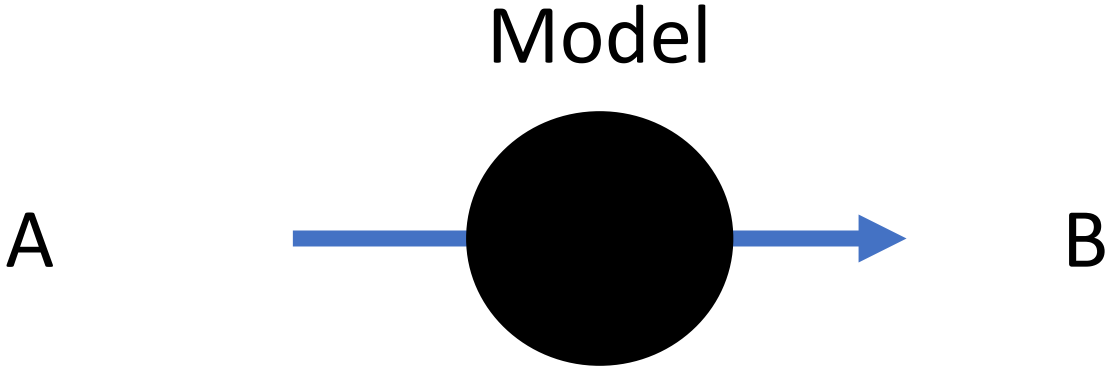
Is prediction all that matters in this case?

What is the role of understanding?

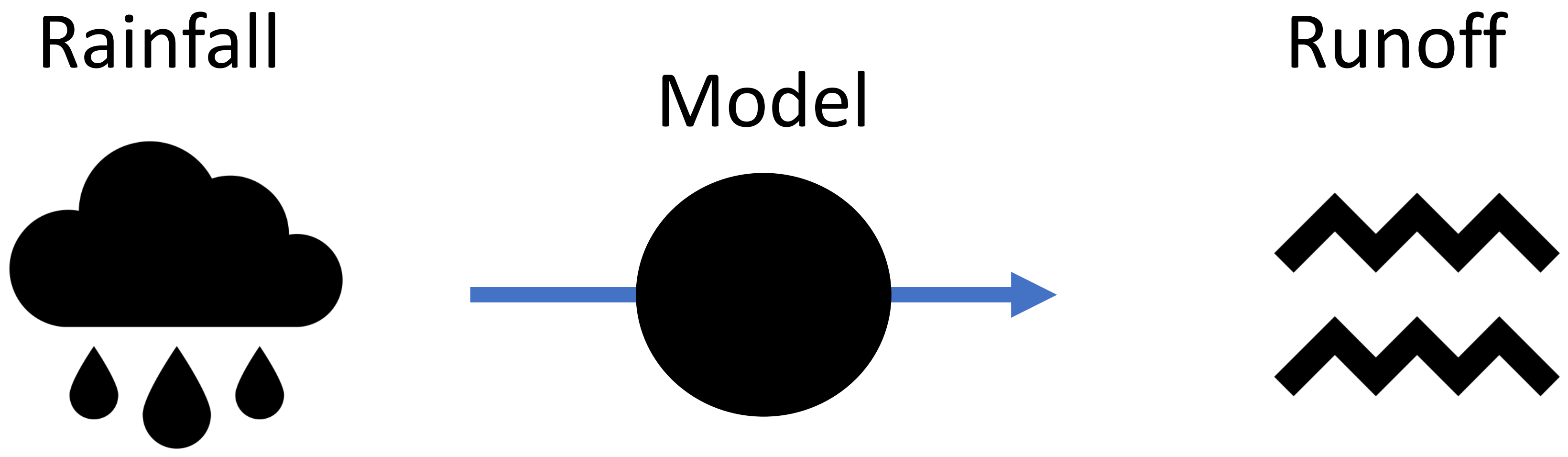


Part 1: Thinking conceptually about some example problems

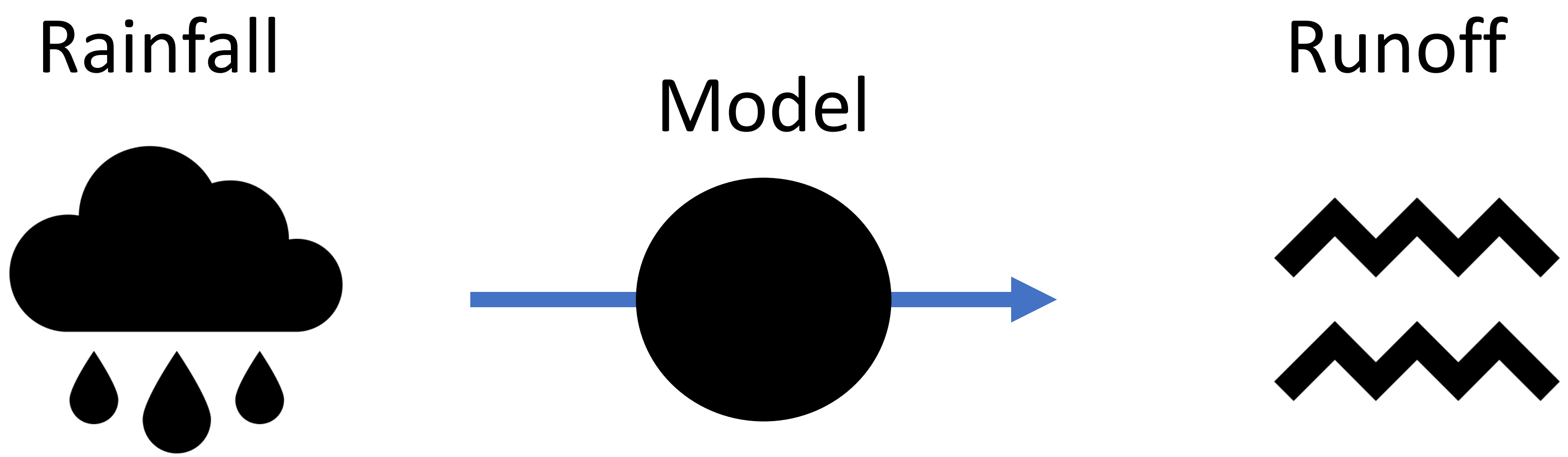
What do we mean by prediction?



Let’s start with our simplest hydrology example:



Let’s start with our simplest hydrology example:



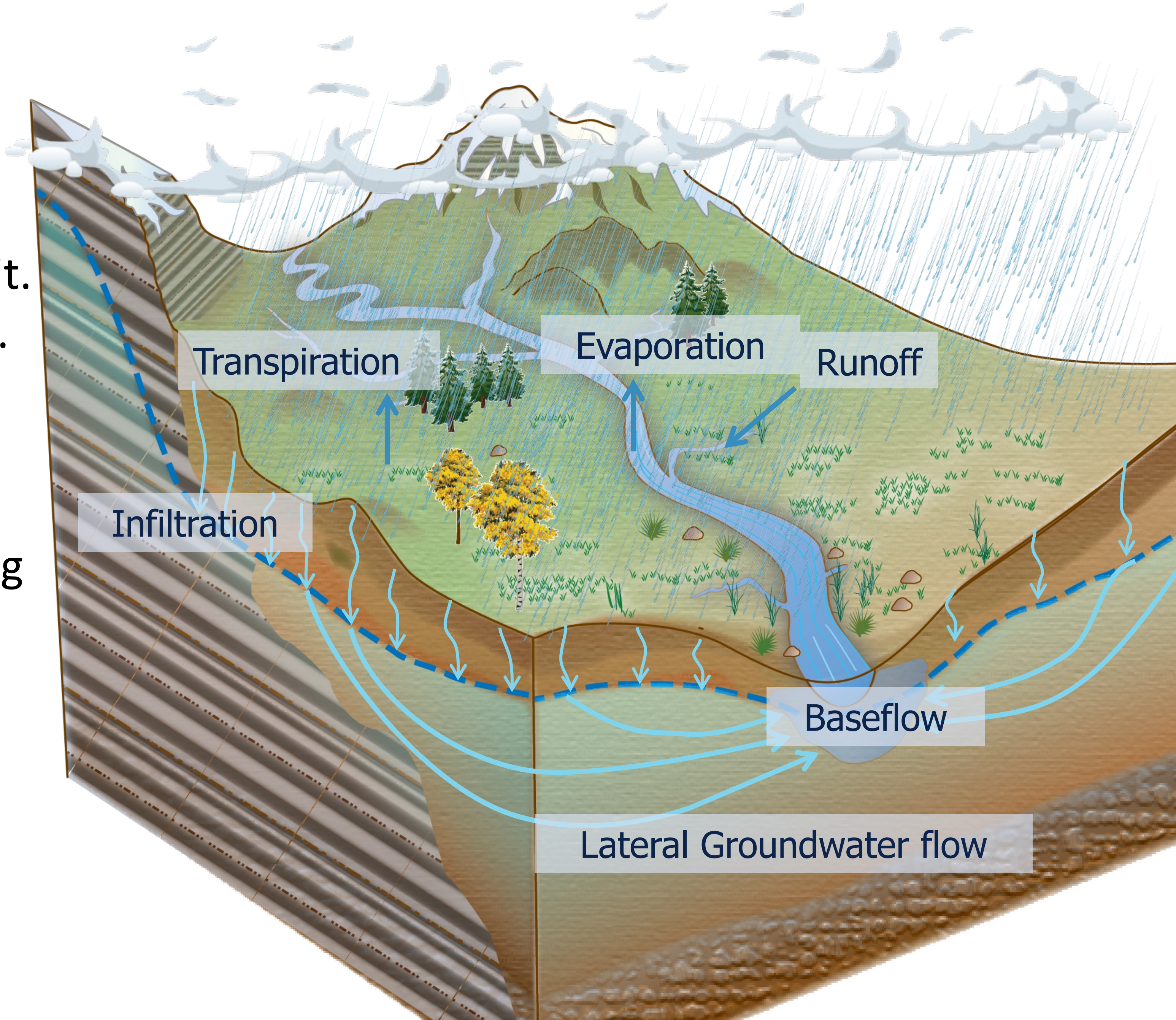
Perfectly measured

Just one watershed

Basically just going from one time series to another… lots of ways to do this some involving more physics than others

Why is this hard?

- Runoff is the result of all of the watershed processes upstream of it.
- Watersheds change seasonally (e.g. plants using more water in the summer)
- Response varies based on the state of the system (e.g. if its been raining for a week)
- We know there are **many** nonstationarities (e.g. land cover change with warming, human development)

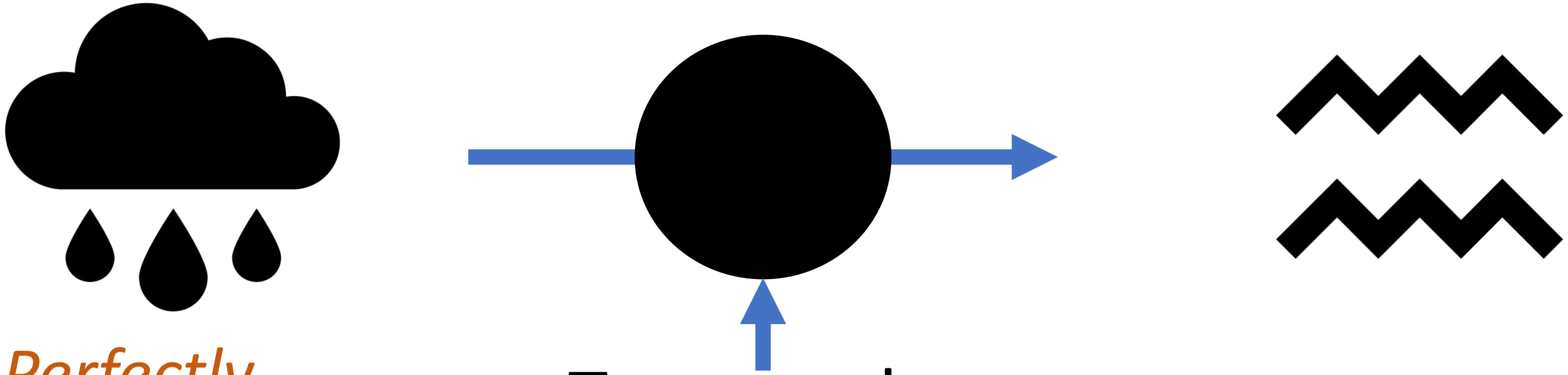


What if we want to do this for more than one watershed ?

Rainfall

Model

Runoff

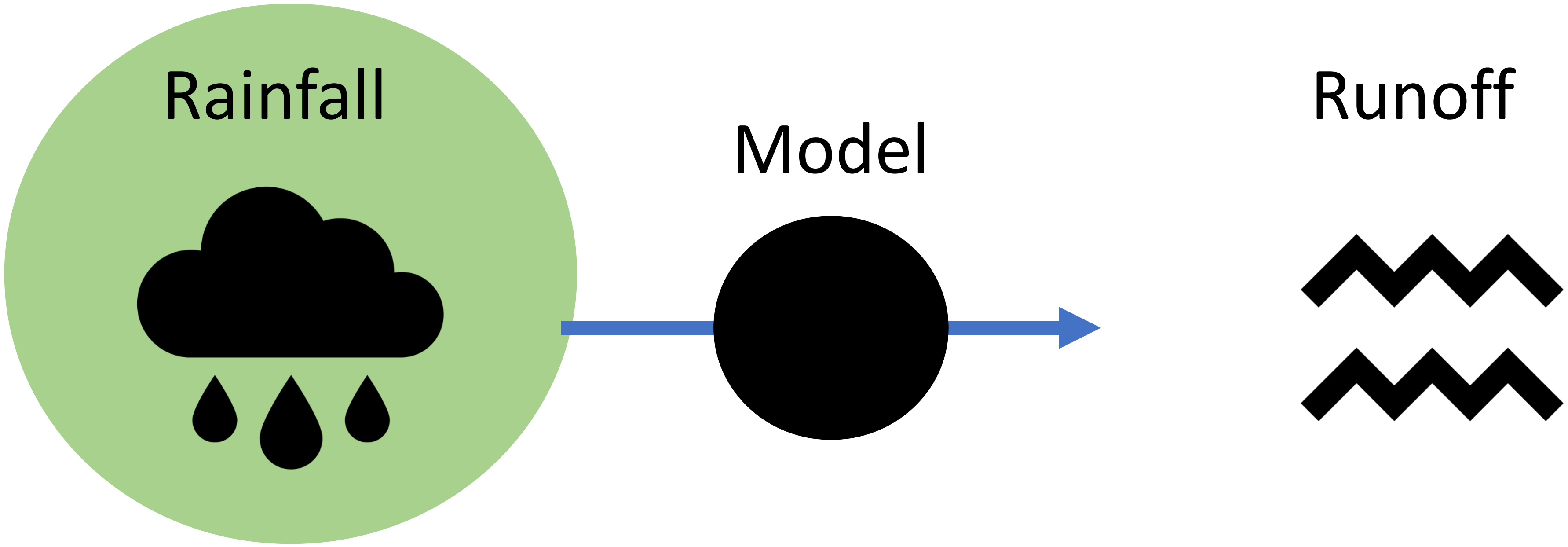


Perfectly measured

- Topography
- Land cover
- Land use
- Water management
- Hydrogeology

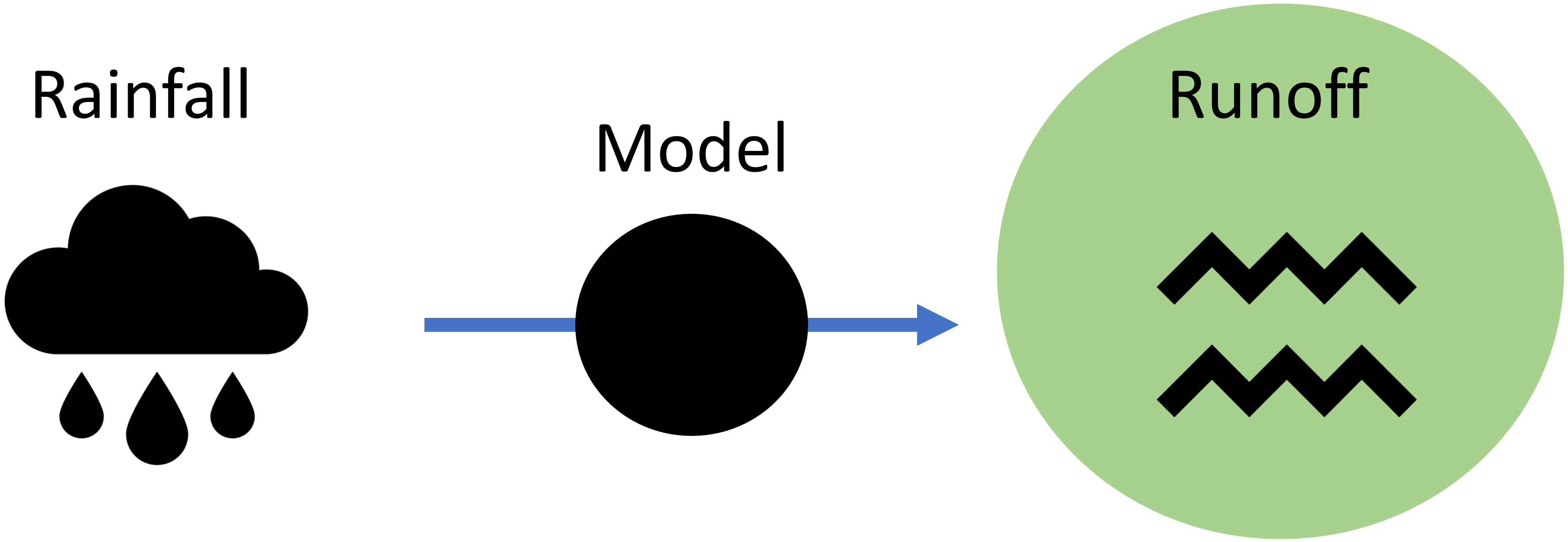
Just one watershed

In reality we don't know the rainfall very well either



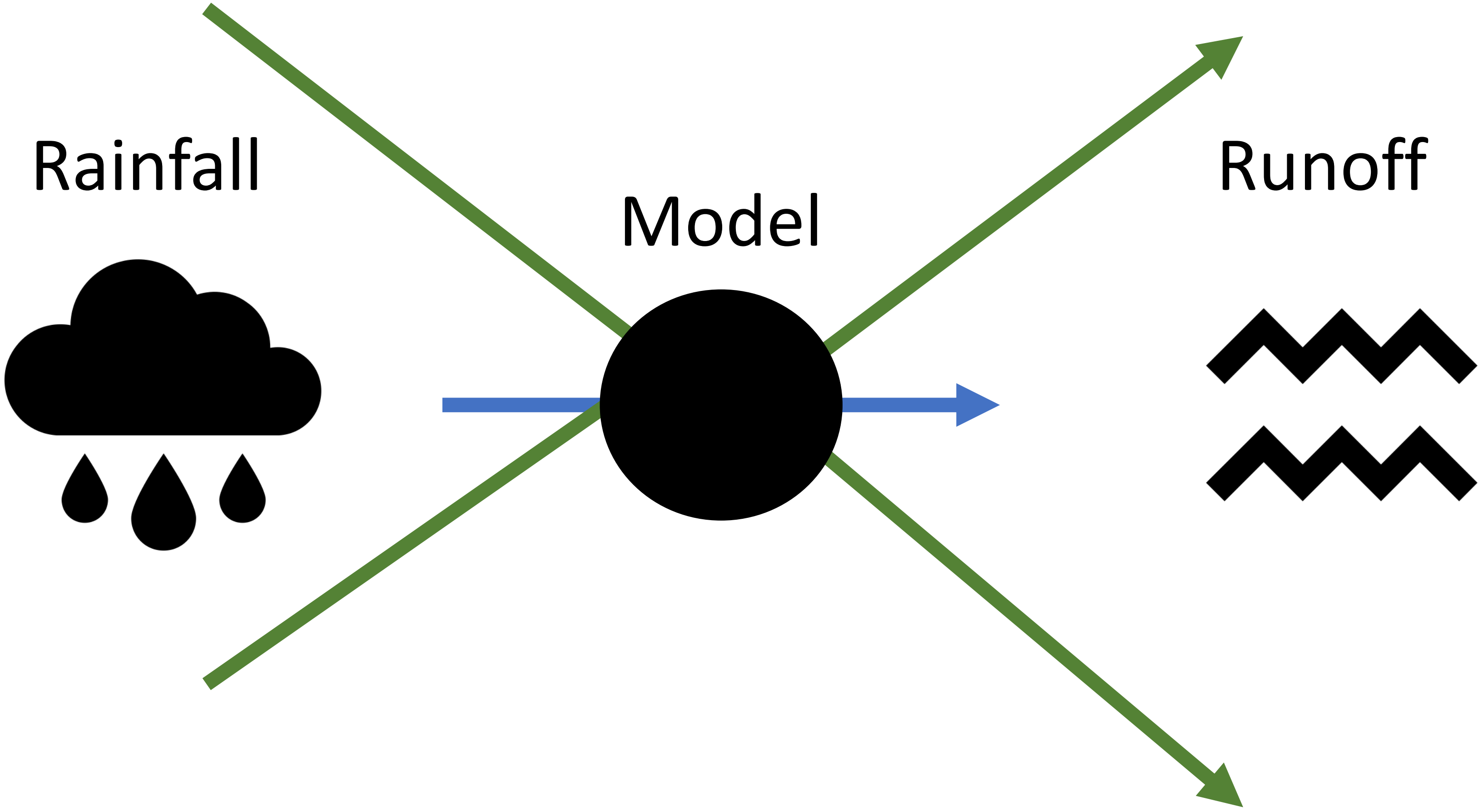
- Lots of uncertainty in measurements (spatially and temporally)
- In most cases we are actually concerned with rainfall forecasts
- There are feedbacks between the land surface and precipitation

Would we all agree about what the 'right' runoff is?



- Assuming we have an imperfect model then there are different preferences about how to be wrong (e.g. preference for high or low bias, focus on extreme events)

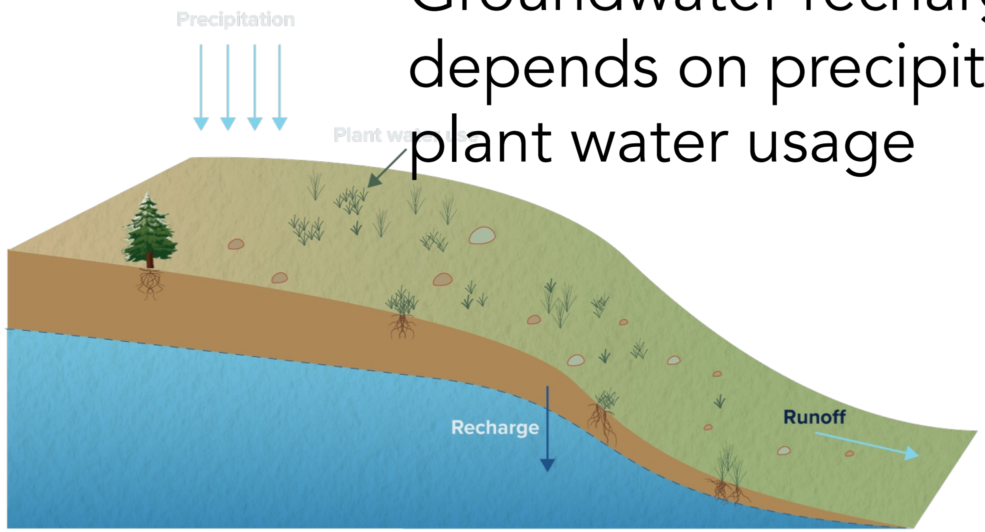
We routinely want to know about the likelihood of **events** or **conditions** that we have never seen before



What about some more complicated examples?

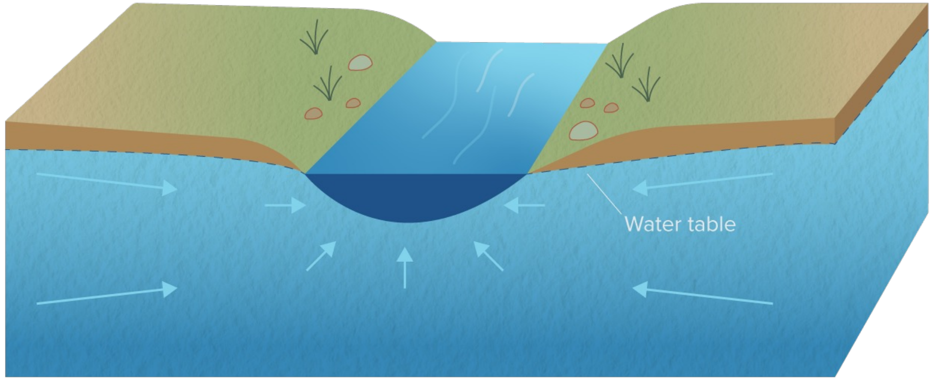
- When and where do we expect to see long term shifts in streamflow regimes with aridification?
- What are the impacts of wetland restoration on quantity and quality?
- Where are we most at risk for compound hazards (e.g. wildfires and landslides)?
- How should we conjunctively manage groundwater and surface water?
- How does groundwater pumping impact drought recovery?

Many of our questions really require a wholistic view of watershed processes

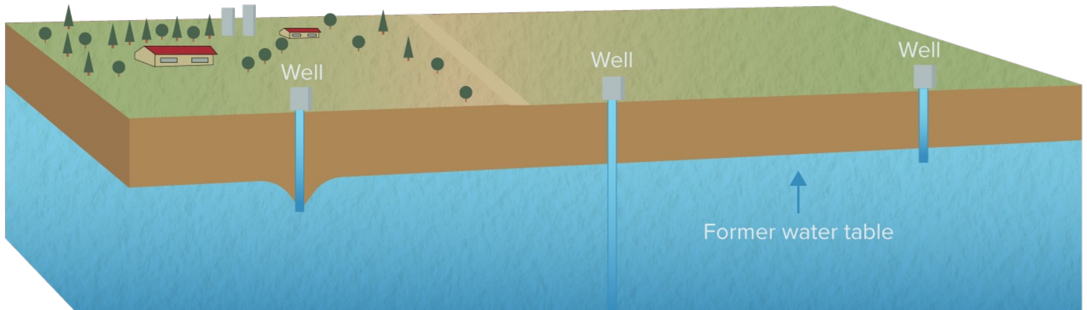


Groundwater recharge depends on precipitation and plant water usage

Streamflow depends groundwater levels which control baseflow



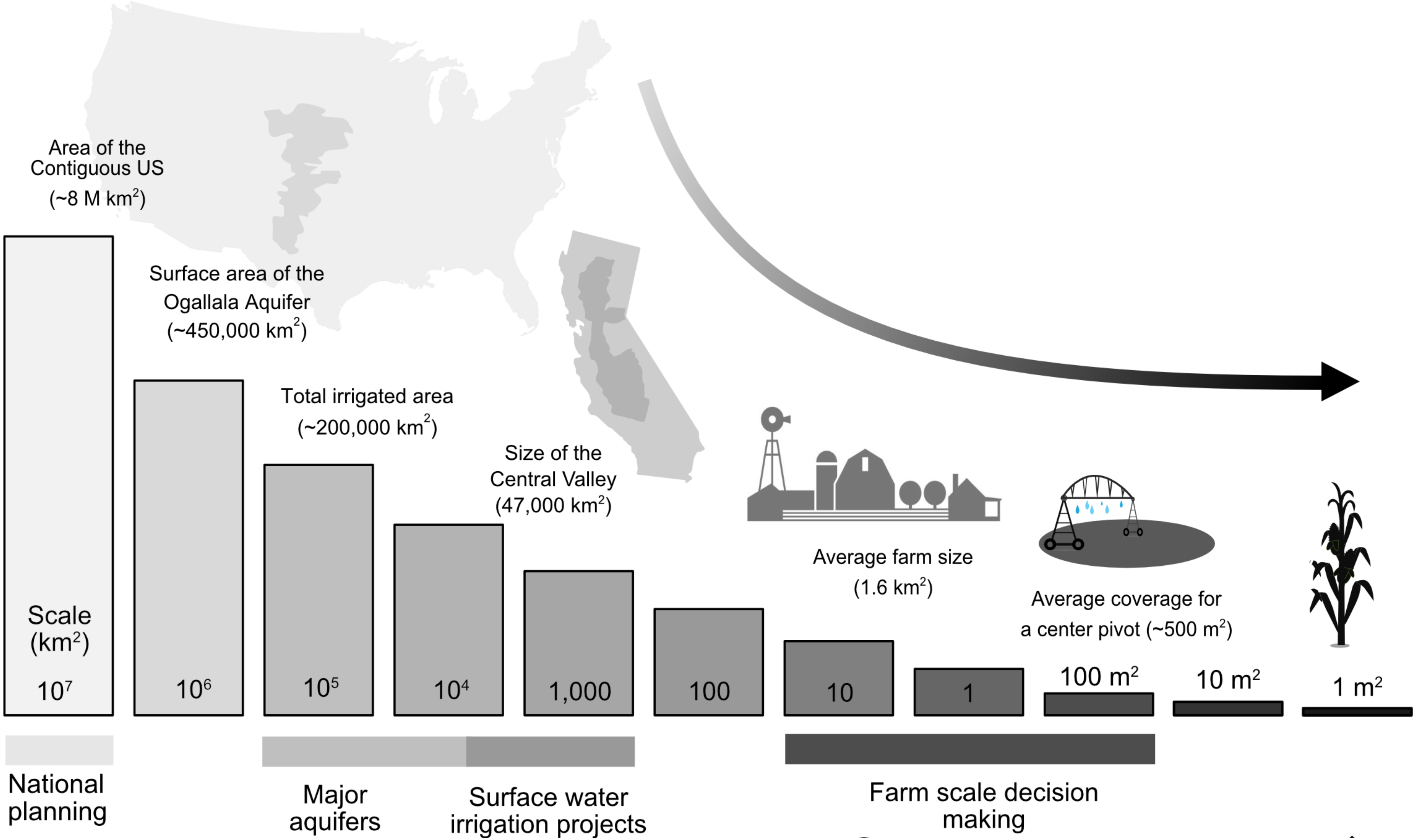
Groundwater levels depend on human water usage



Part 2: Data and models

“Our models have significant structural errors and our data are subject to huge uncertainties”

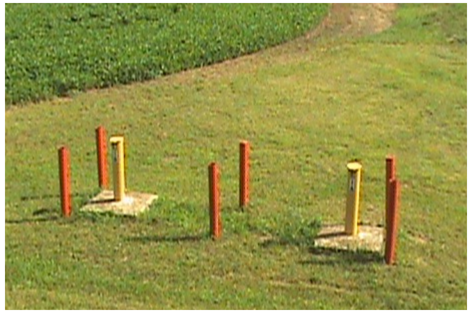
- Someone in this room yesterday


Area of the Contiguous US (~8 M km²)
Surface area of the Ogallala Aquifer (~450,000 km²)
Total irrigated area (~200,000 km²)
Size of the Central Valley (47,000 km²)
Average farm size (1.6 km²)
Average coverage for a center pivot (~500 m²)
Scale (km²)
10⁷
10⁶
10⁵
10⁴
1,000
100
10
1
100 m²
10 m²
1 m²
National planning
Major aquifers
Surface water irrigation projects
Farm scale decision making

We have many observations that can tell us part of the story

Point measurements have limited spatial coverage

Groundwater Well



Flux Tower

SNOTEL Station

Stream gauges are spatially aggregated measurements

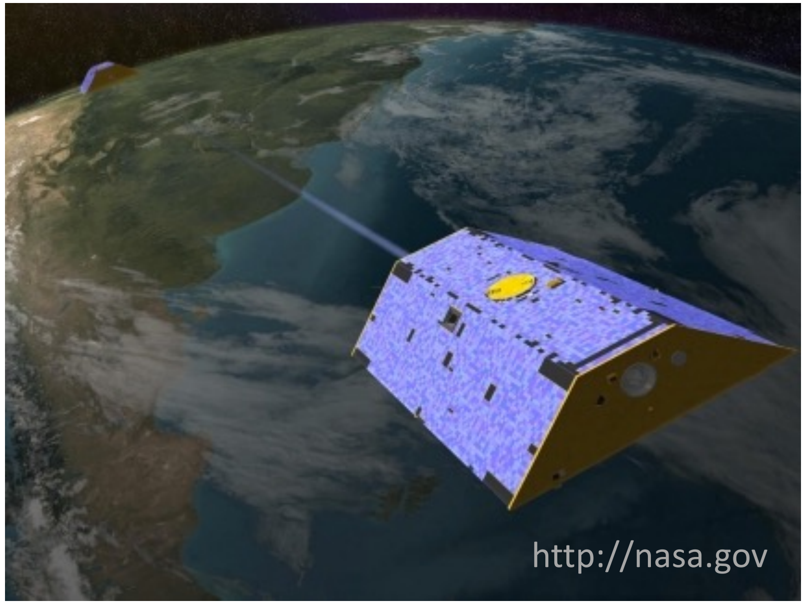
Satellite data may have limited resolution

NASA Earth Observing Missions

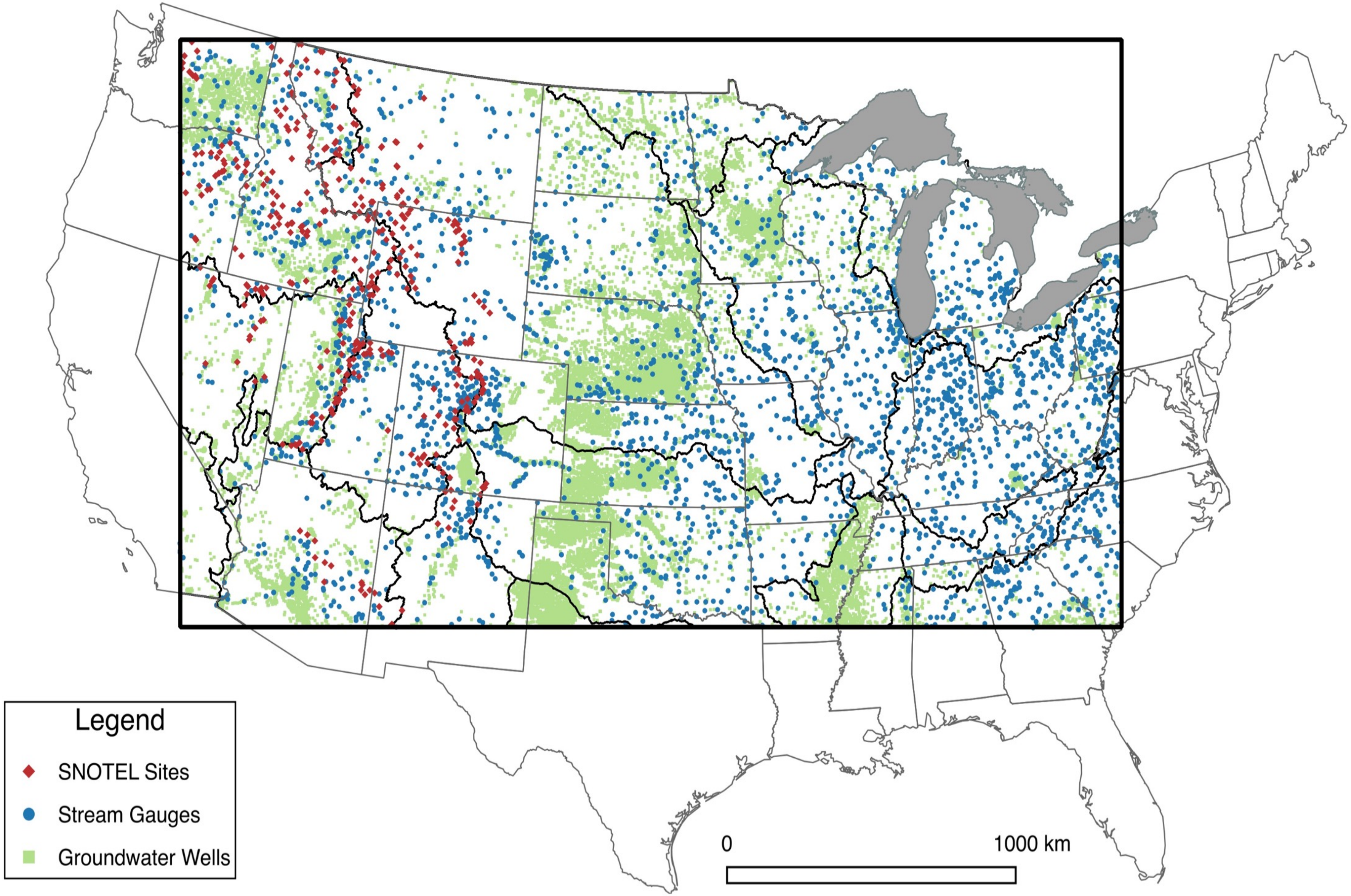
None of them can tell the whole story

Local measurements are difficult to scale

Remote sensing can't see everything

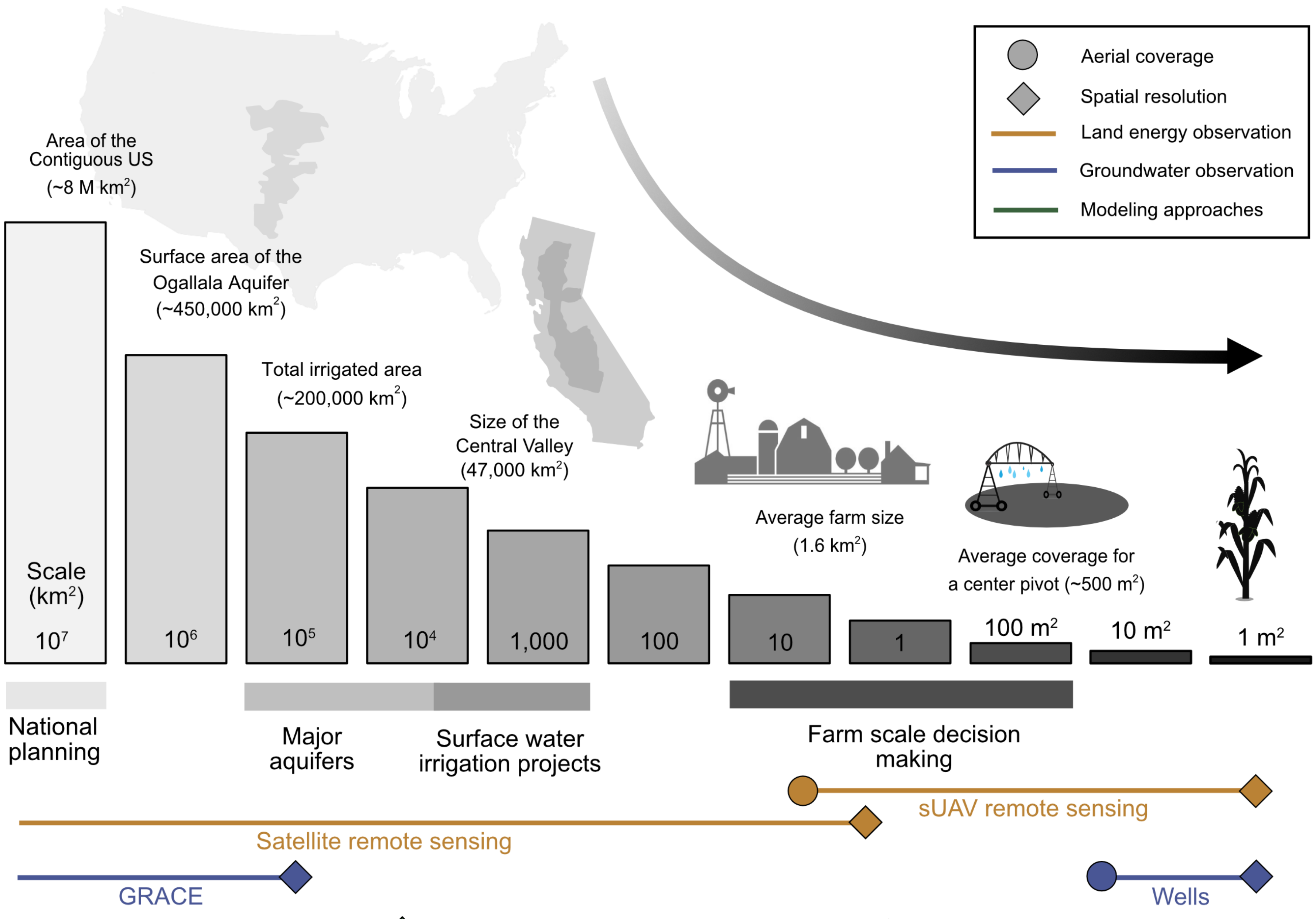


We likely haven't observed many of the events we care most about



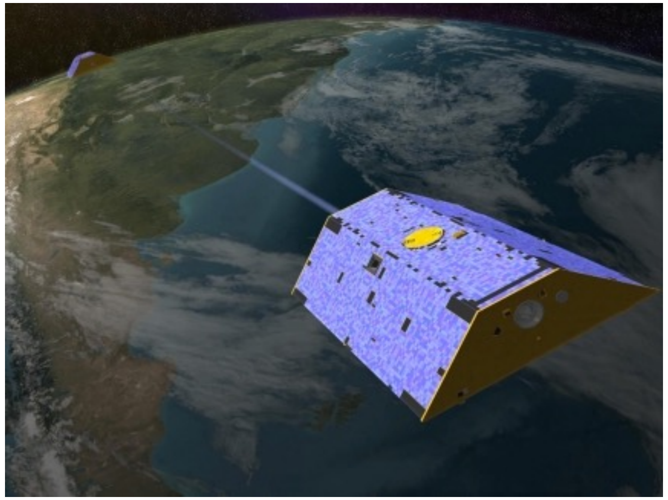
~1.2 million observations available for a one-year simulation

- 378 SNOTEL Stations
- 3,050 USGS gages
- 29,385 USGS Wells
- Varying temporal resolution

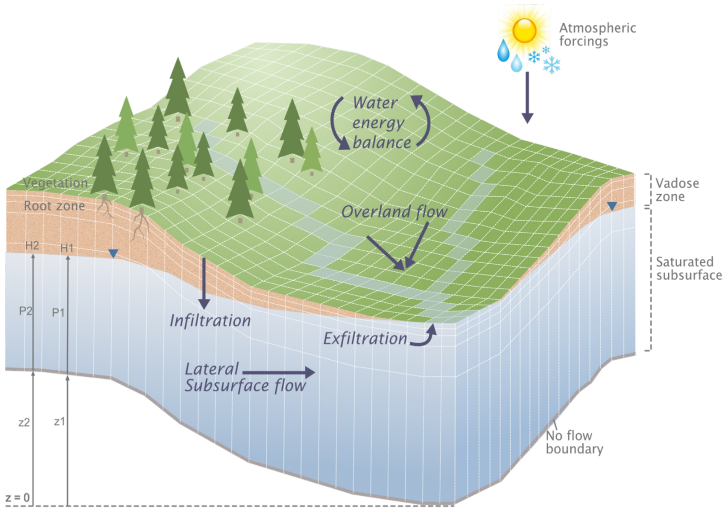


Aerial coverage
Spatial resolution
Land energy observation
Groundwater observation
Modeling approaches
Area of the Contiguous US (~8 M km²)
Surface area of the Ogallala Aquifer (~450,000 km²)
Total irrigated area (~200,000 km²)
Size of the Central Valley (47,000 km²)
Average farm size (1.6 km²)
Average coverage for a center pivot (~500 m²)
Scale (km²)
10⁷
10⁶
10⁵
10⁴
1,000
100
10
1
100 m²
10 m²
1 m²
National planning
Major aquifers
Surface water irrigation projects
Farm scale decision making
sUAV remote sensing
Satellite remote sensing
GRACE
Wells

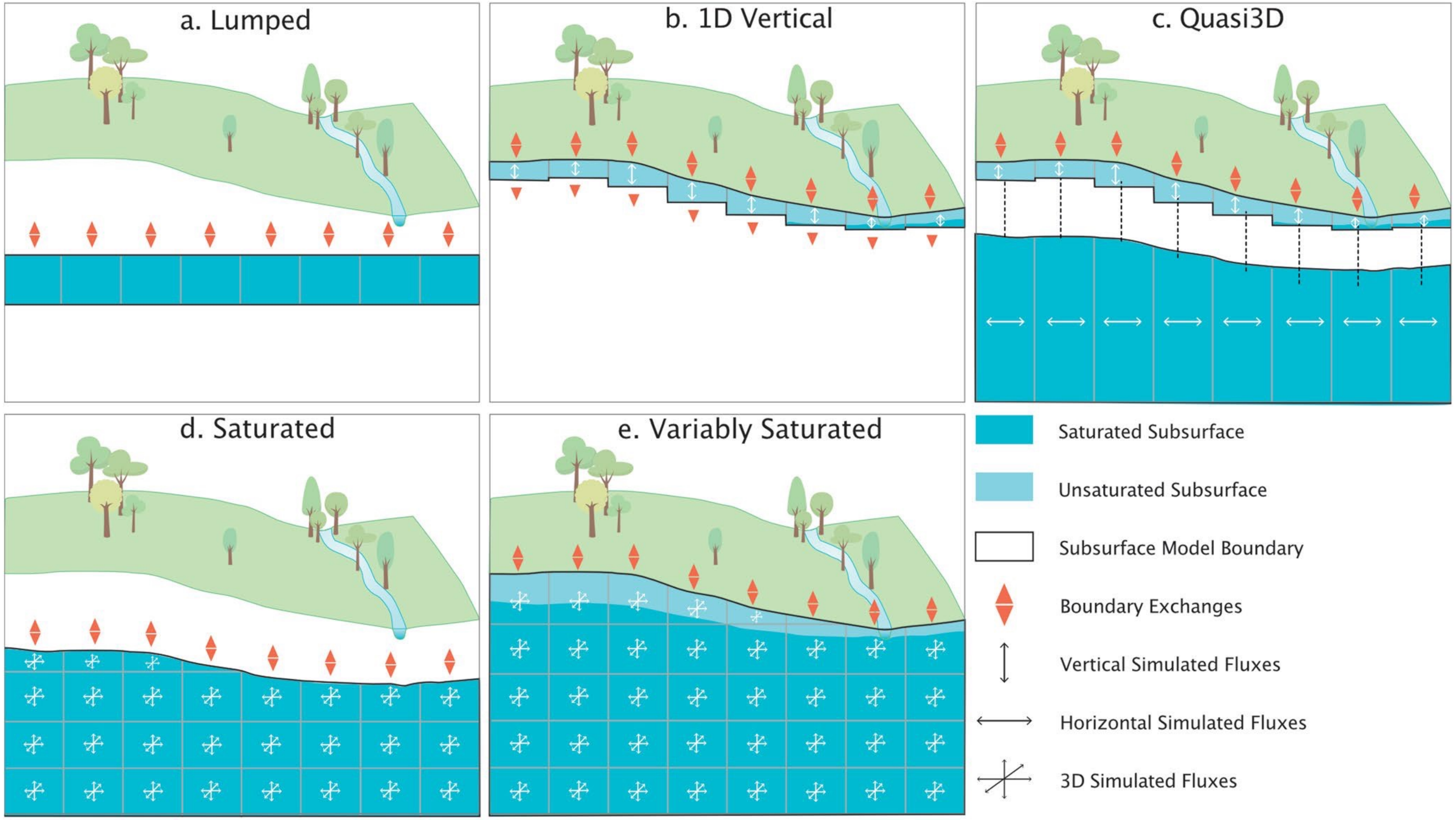
Models can help bridge gaps



What kind of model?

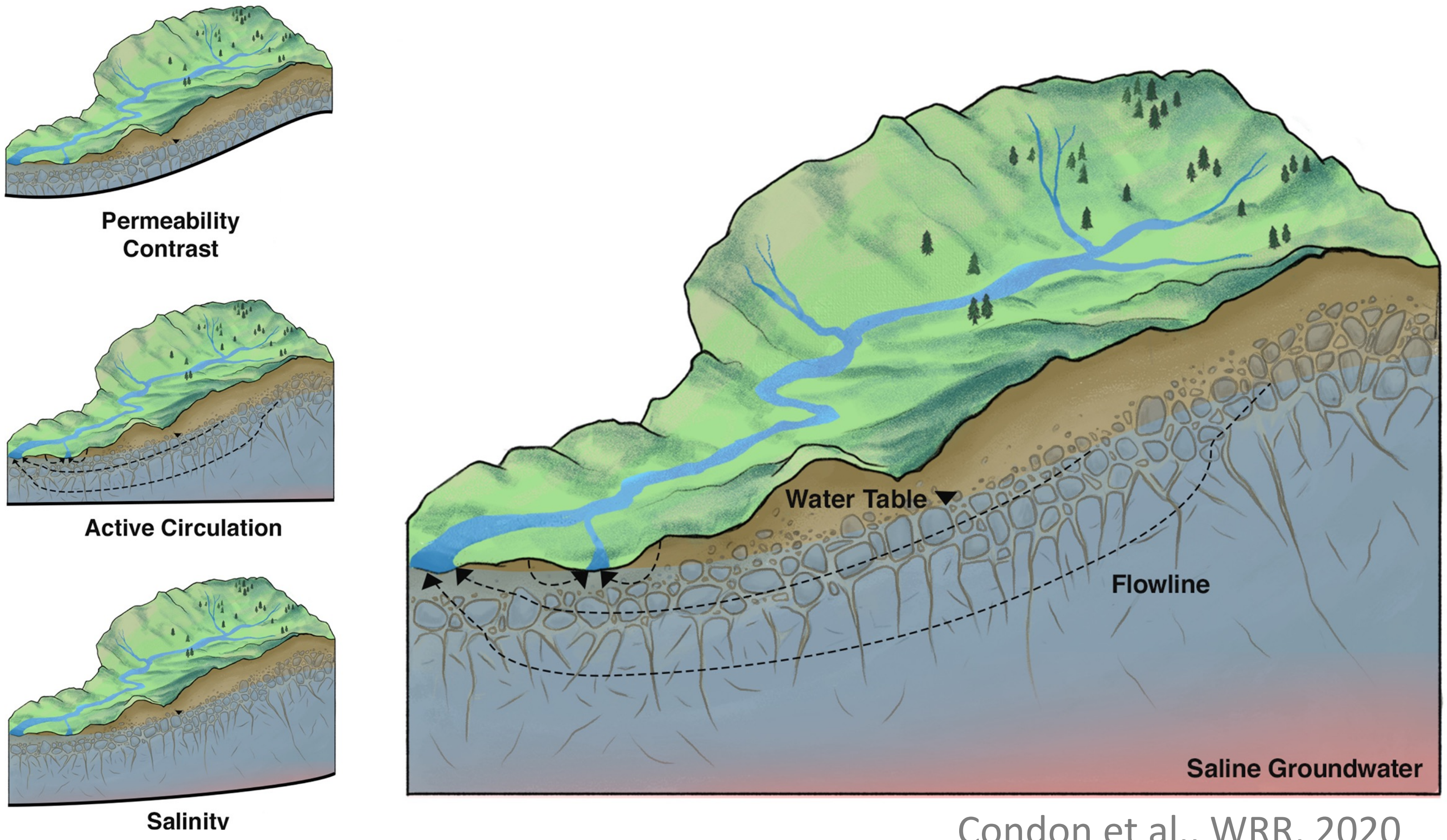


Approaches to physically based modeling vary greatly depending on application



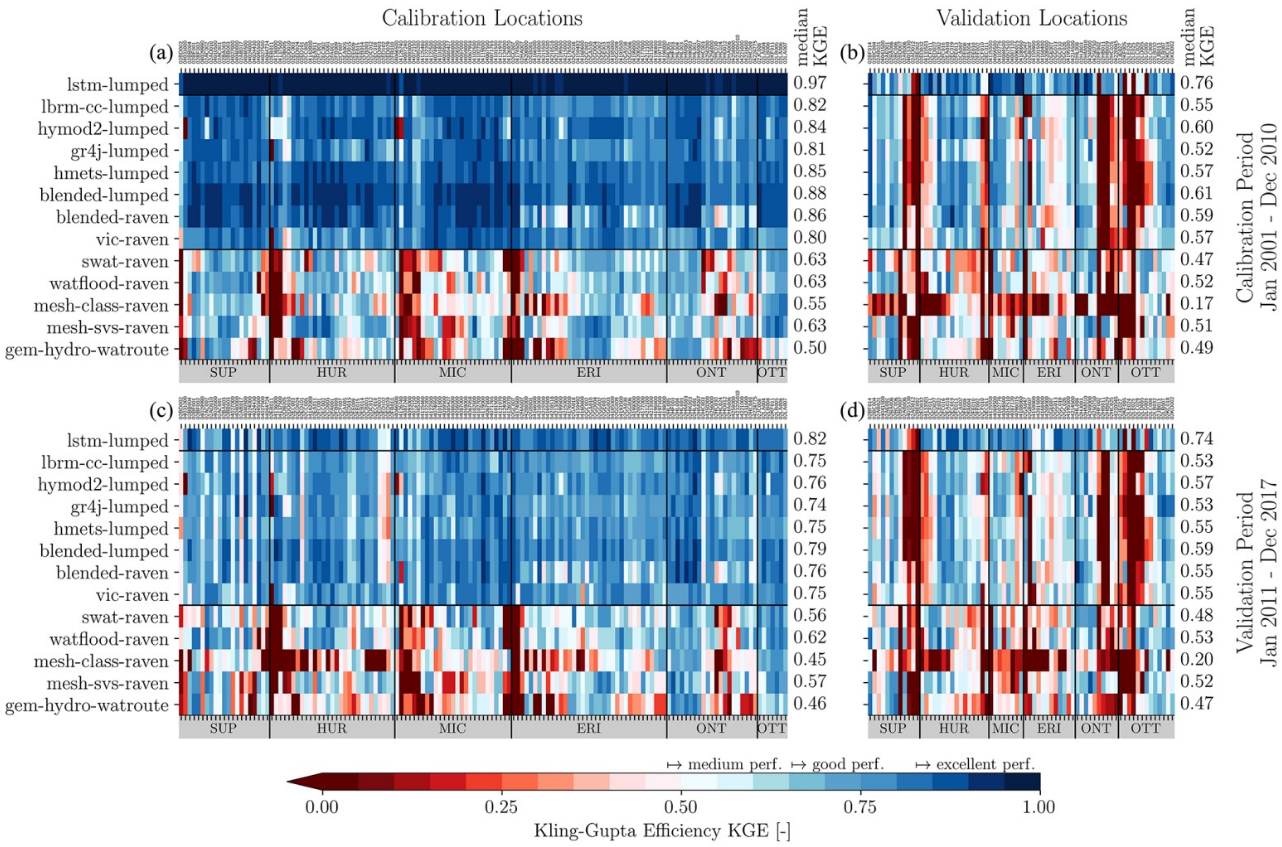
Condon et al., WRR, 2021

There is also a lot of variability in where we place the boundaries of our models



Condon et al., WRR, 2020

Machine learning has really taken off in Hydrology and in many cases outperforms physically based models



Hydrol. Earth Syst. Sci., 26, 3537–3572, 2022
https://doi.org/10.5194/hess-26-3537-2022

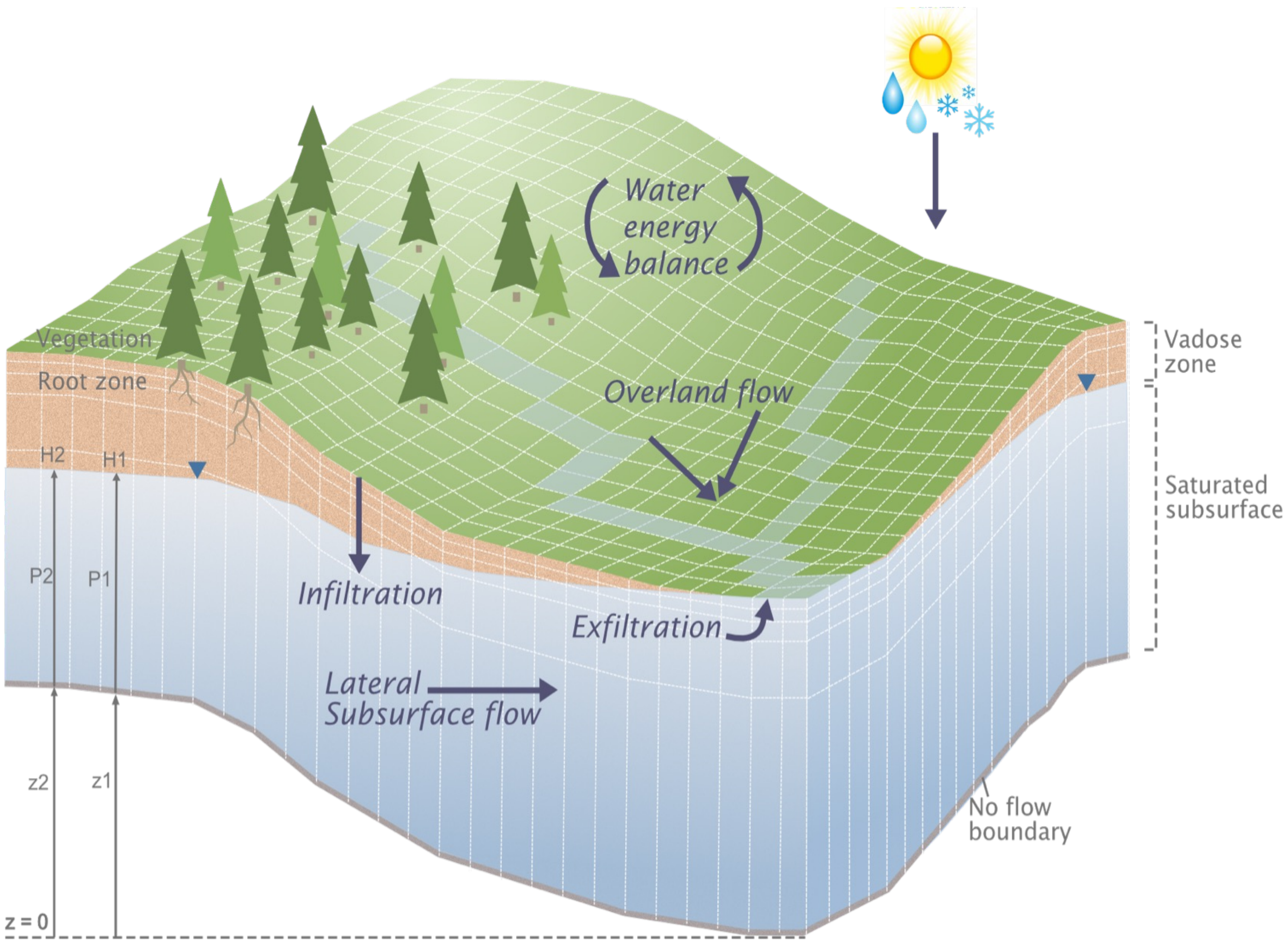
Hydrology and Earth System Sciences
Open Access

The Great Lakes Runoff Intercomparison Project Phase 4: the Great Lakes (GRIP-GL)

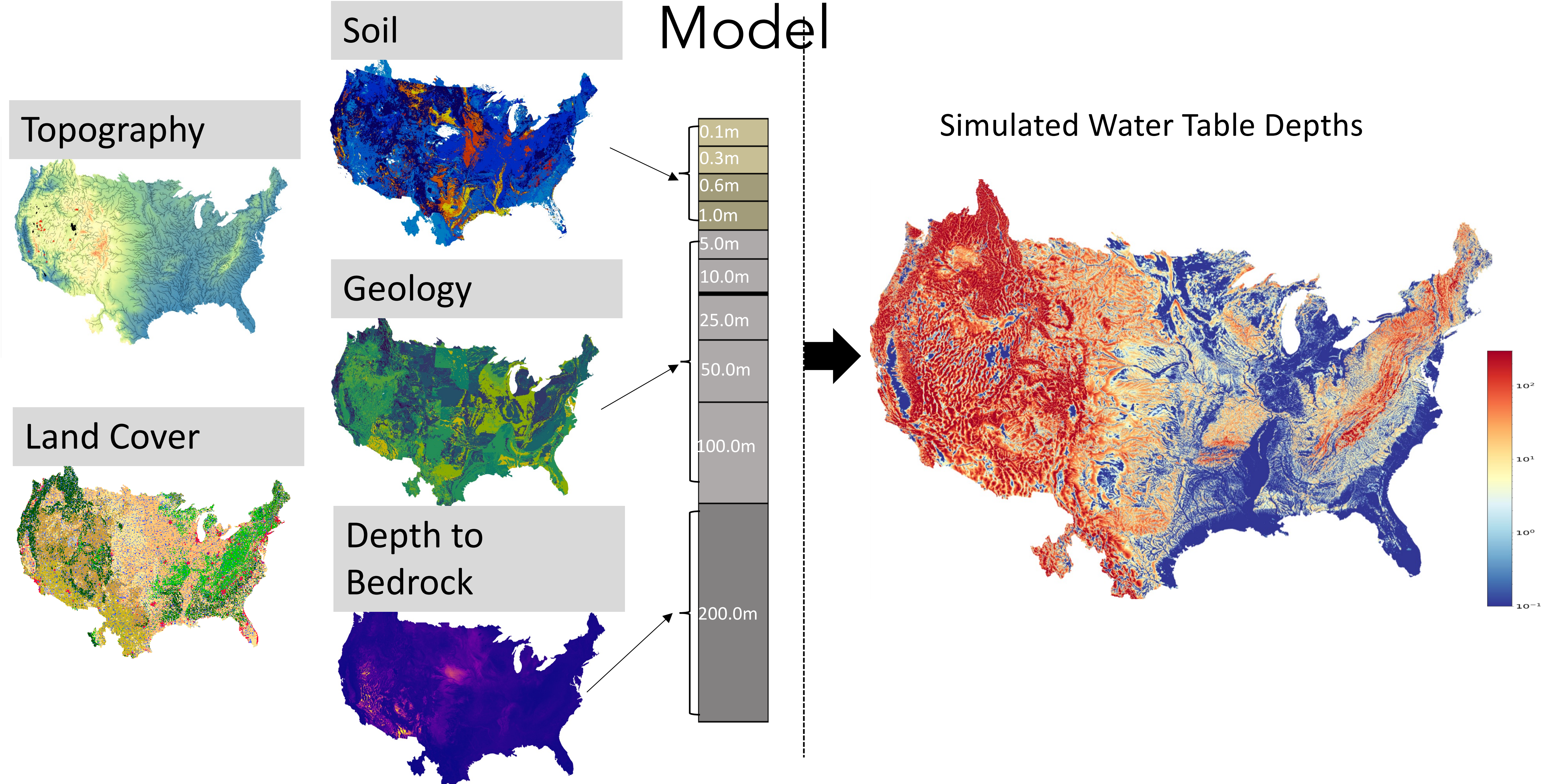
Juliane Mai[1], Hongren Shen[1], Bryan A. Tolson[1], Étienne Gaborit[2], Richard Arsenault[3], James R. Craig[1], Vincent Fortin[2], Lauren M. Fry[4], Martin Gauch[5], Daniel Klotz[5], Frederik Kratzert[5,6], Nicole O'Brien[7], Daniel G. Princz[8], Sinan Rasiya Koya[9], Tirthankar Roy[9], Frank Seglenieks[7], Narayan K. Shrestha[7], André G. T. Temgoua[7], Vincent Vionnet[2], and Jonathan W. Waddell[10]

A bit about how I use models

I use integrated hydrologic models to explore interactions that are hard to see and measure



CONUS-2.0 Second Generation National ParFlow Model



WATER RESOURCES

Connections between groundwater flow and transpiration partitioning

Reed M. Maxwell[1]* and Laura E. Condon[2]

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Quantitative assessment of groundwater controls across major US river basins using a multi-model regression algorithm

Laura E. Condon[a,b,d,*], Amanda S. Hering[c], Reed M. Maxwell[a,b,d]

[a] Department of Geology and Geological Engineering, Colorado School of Mines, United States
[b] Integrated GroundWater Modeling Center, United States
[c] Department of Applied Mathematics and Statistics, Colorado School of Mines, United States
[d] Climate Change Water and Society (CCWAS), Integrative Graduate Education and Research Traineeship (IGERT), United States

ARTICLE

https://doi.org/10.1038/s41467-020-14688-0 OPEN

Evapotranspiration depletes groundwater under warming over the contiguous United States

Laura E. Condon[1✉], Adam L. Atchley[2] & Reed M. Maxwell[3]

OPEN ACCESS
IOP Publishing Environmental Research Letters

Environ. Res. Lett. 9 (2014) 034009 (9pp) doi:10.1088/1748-9326/9/3/034009

Groundwater-fed irrigation impacts spatially distributed temporal scaling behavior of the natural system: a spatio-temporal framework for understanding water management impacts

SCIENCE ADVANCES | RESEARCH ARTICLE

ENVIRONMENTAL STUDIES

Simulating the sensitivity of evapotranspiration and streamflow to large-scale groundwater depletion

Laura E. Condon[1]* and Reed M. Maxwell[2]

Hydrol. Earth Syst. Sci., 21, 1117–1135, 2017
www.hydrol-earth-syst-sci.net/21/1117/2017/
doi:10.5194/hess-21-1117-2017

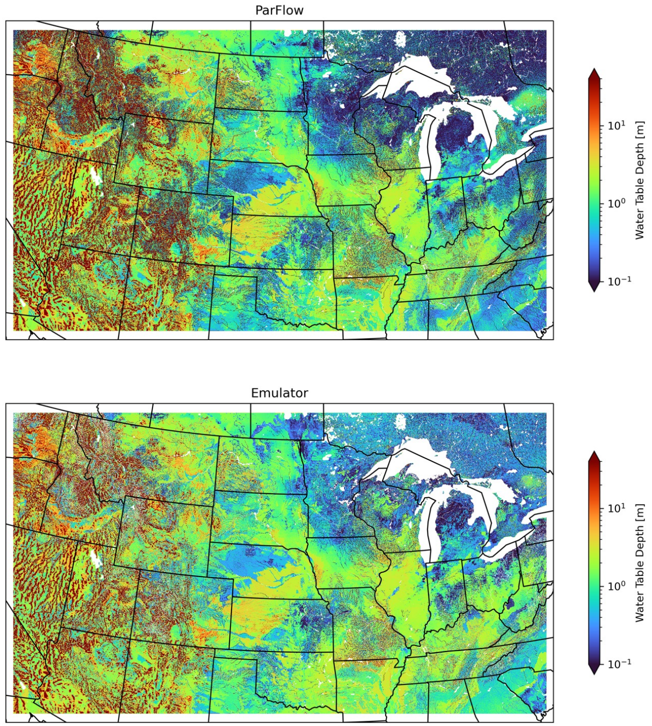
Hydrology and Earth System Sciences
Open Access

Systematic shifts in Budyko relationships caused by groundwater storage changes

Laura E. Condon[1] and Reed M. Maxwell[2]

Machine learning Emulators can run 1000 times faster

- Hydrologic Emulator of the physics-based simulations (emulators for the the 3-D pressure field and the land surface processes)
- Current conditions generators use observations to generate gridded current conditions



Example of emulator performance comparing emulated water table depth to the physics-based simulation

HydroGEN

A Machine Learning Platform for
Hydrologic Scenario Generation

THE UNIVERSITY OF ARIZONA

BUREAU OF RECLAMATION

CYVERSE

PRINCETON UNIVERSITY

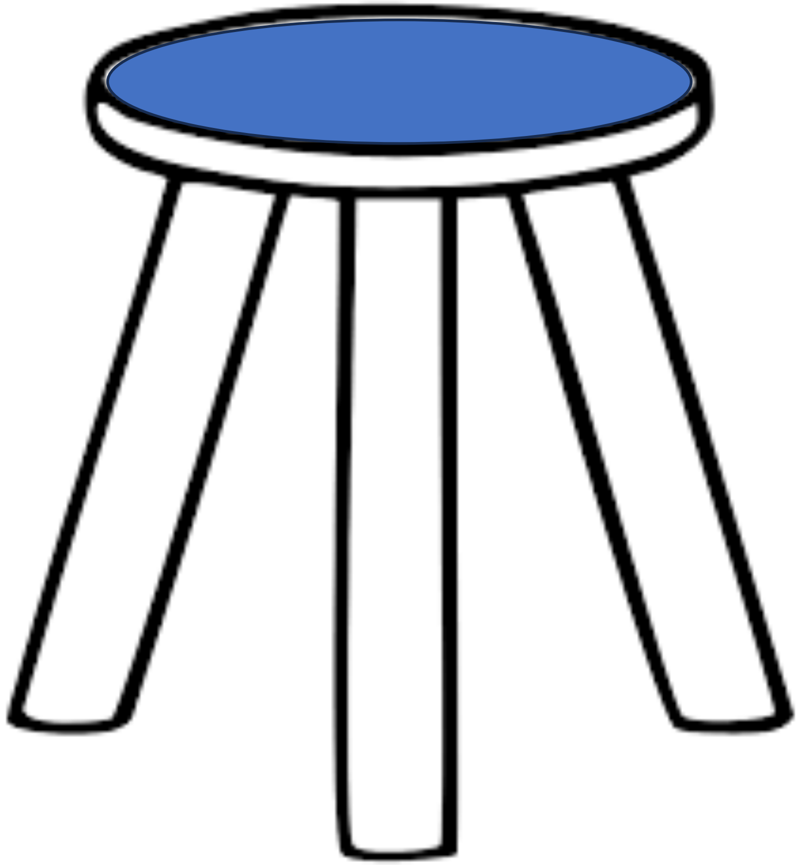
Part 3: Some concluding thoughts from yesterday

Reasons we have trouble fitting our problems into simple boxes

- Nonstationarity is central to many of the most pressing problems we want to solve
- Almost everything we do is multivariate
- Challenges translating from continuous to discrete variables
- Our causal relationships change in both space and time
- We know there are feedbacks (and they often matter a lot for the extreme events we worry a lot about)
- We are not able to consistently observe the inputs or the outputs of our systems. To do our best job we need to combine many pieces of information from different parts of the system that are often measured in different locations, and at different spatial and temporal scales
- We are very uncertain about many of the parameters we need to solve our physical equations

Where do we go from here?

- Progress in mathematical approaches that allow us to relax some of the assumptions that don't fit our systems
- Do a more thoughtful job of figuring out how to fit into the requirements of the methods we adopt
- Better understand how the assumptions we make bias our solutions



Thank you!