DATA VISUALIZATION WITH SHINY

BENJAMIN BEAL

PORTLAND WATER BUREAU

BENJAMIN.BEAL@PORTLANDOREGON.GOV

OUTLINE

- Modeling and the paradox of visualizing big data sets
- What is Shiny, and how can it help
- Example 1 – Stochastic modeling of our water system supply
- Example 2 – In depth analysis of our reservoir model parameters
- Thoughts on Shiny
- Questions

MODELING AND THE PARADOX OF VISUALIZATION

- Models generate a lot of data per run
- Top-down climate modeling approaches generate 100s of simulations for an "ensemble"
- Too much data to plot, it just looks like spaghetti
- Must summarize data to look at it all, but summaries hide the mechanisms that cause interesting simulations
- How to weigh the "big picture" verses the "fine detail"?

WHAT IS SHINY?

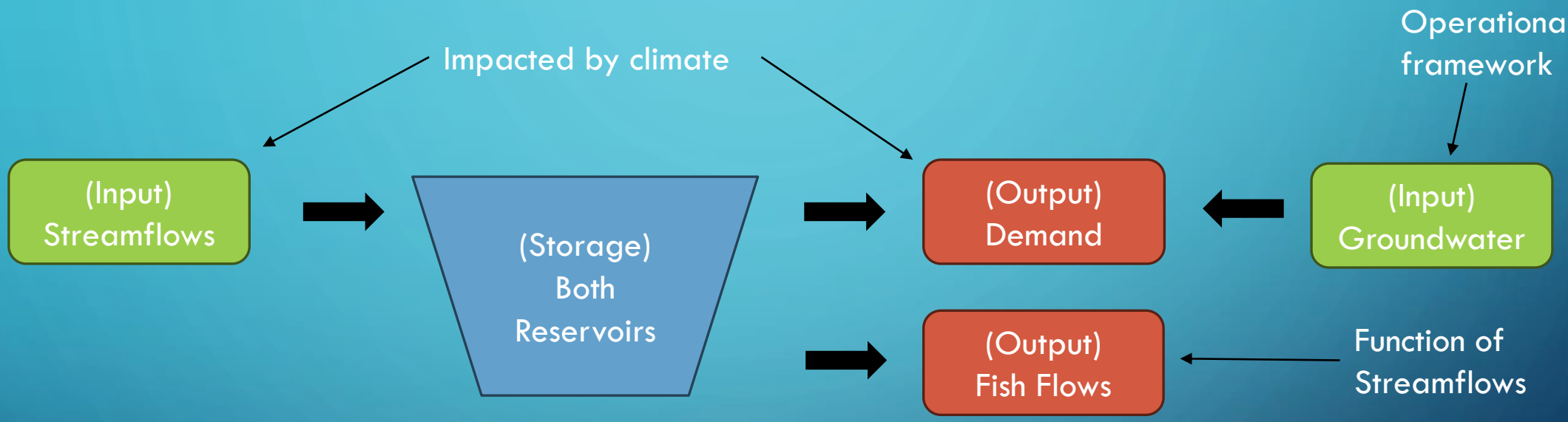
- Shiny is a free library for the R programing language
- Builds web applications by converting code in R into HTML/CSS/JavaScript
- Similar to Power BI and Tableau
- Highly customizable
- Adapts to your existing R code
- Works well with geospatial data using “leaflet” library


Bull Run Watershed
Columbia River
Columbia South Shore Well Field
Powell Butte
Kelly Butte
Willamette River
Mount Tabor
Washington Park
Washington County

EXAMPLE 1: EFFECTS OF DRAWDOWN ON OUR STORAGE DUE TO CLIMATE CHANGE

- Top-down modeling approach utilizing 5 selected GCM models
- Utilizes a hydrologic model (PRMS) for inflows
- Utilizes in house daily demand model for outflows
- Utilizes a set of regulatory commitments to determine “fish flows”
- Utilizes “groundwater curves” to initiate supplemental groundwater supply

EXAMPLE 1: COMPONENTS OF A SINGLE MODEL



EXAMPLE 1: COUNTING ALL THE DATA POINTS

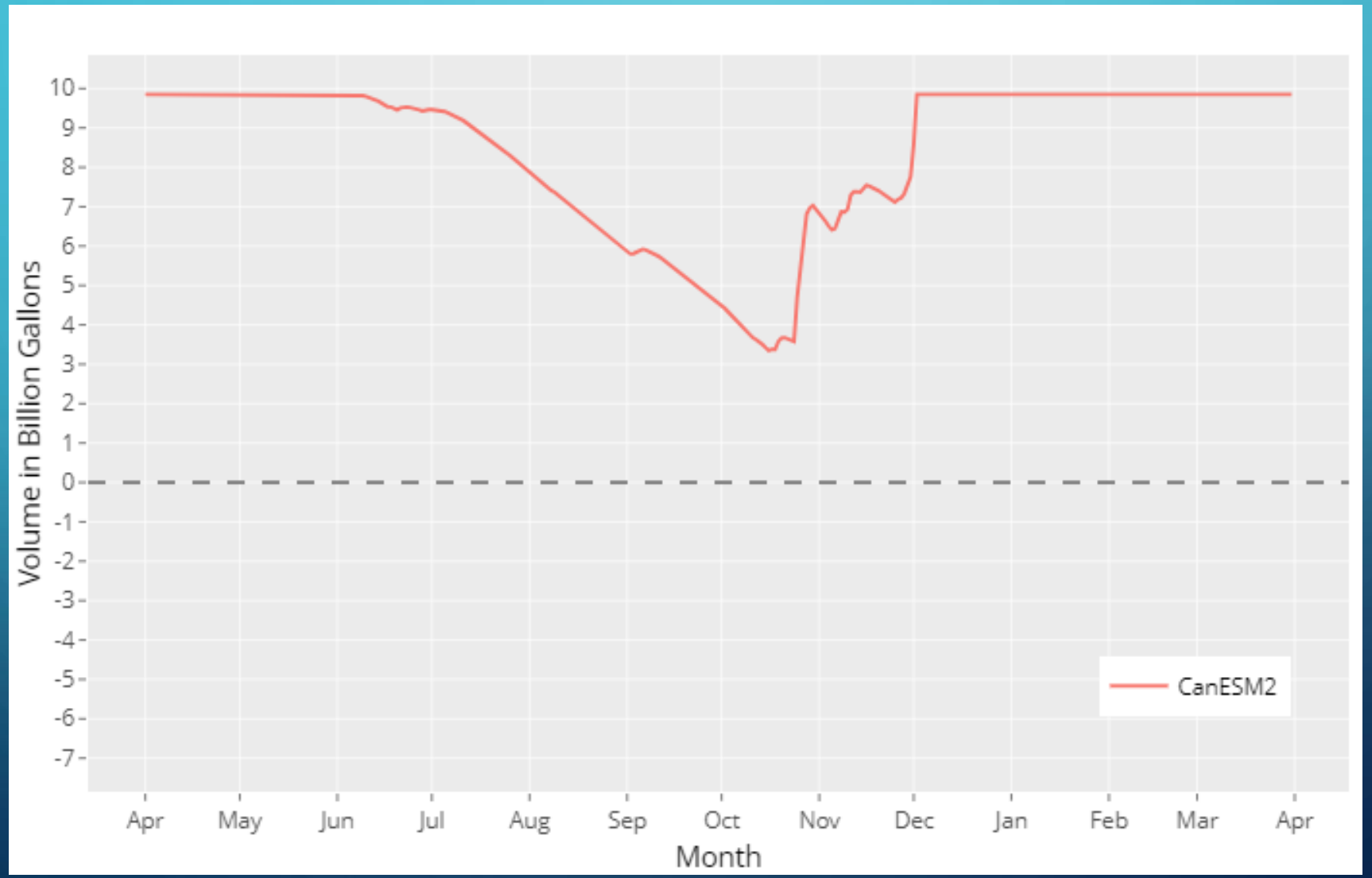
- 5 GCM models
- 4 Different “demand populations”
- 56 Historical years (1950-2005)
- 36 Future years (2030-2059)

Produces 720 future traces = 263,000 days of information

Additionally, 1120 historical traces = 409,000 days of information

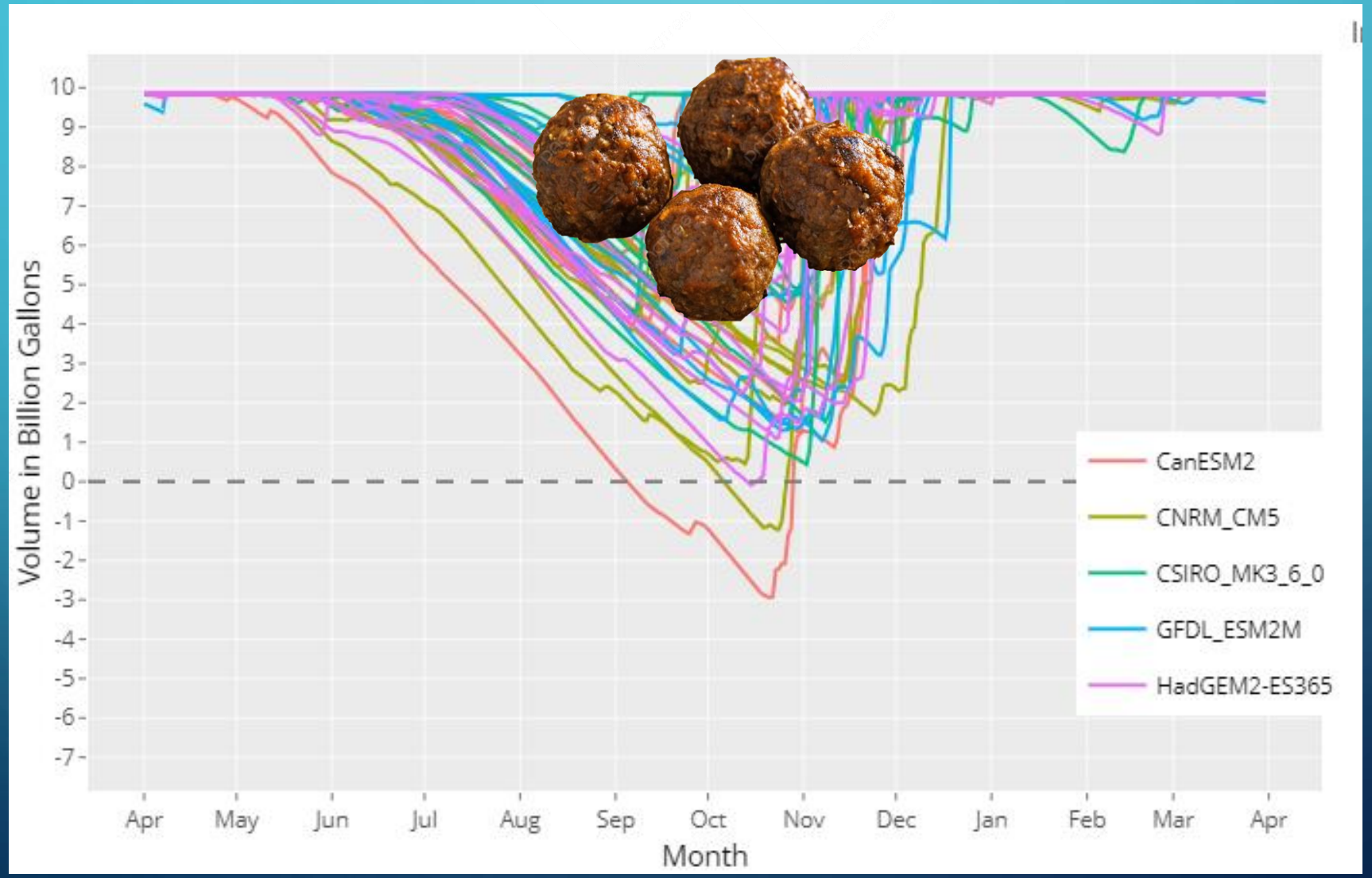
LET'S LOOK AT 1 TRACE

1 GCM, 1 Demand population, 1 years (1 trace of the total 720 future traces)

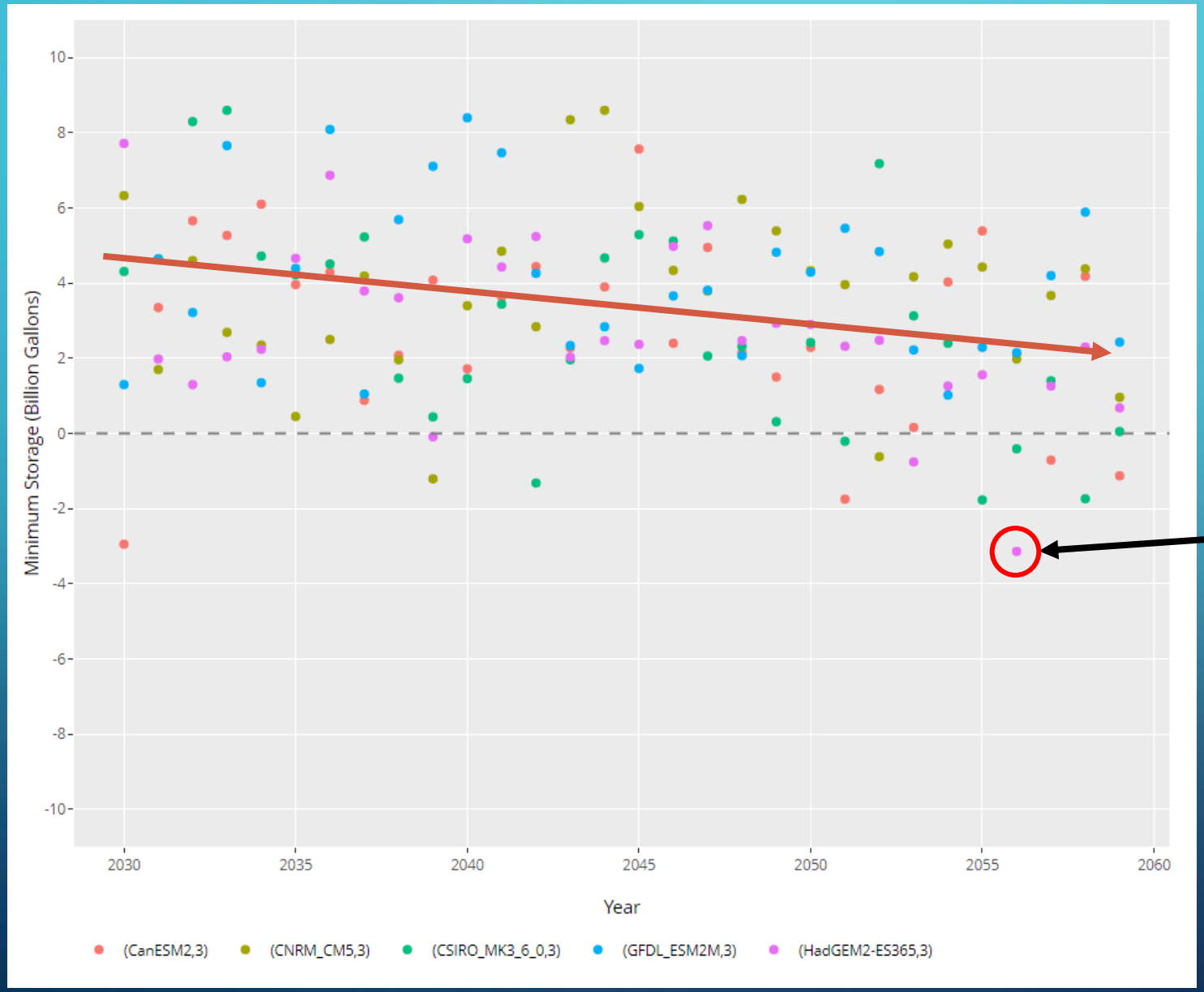


LET'S LOOK AT A FEW MORE

5 GCMs, 1 Demand population, 10 years (50 traces of the total 720 future traces)



HOW ABOUT A DIFFERENT APPROACH?



Summarize by minimum storage

Can now see something of a trend to the statistic, that's useful!

What's up with this one?

SHINY ALLOWS US TO BRIDGE THIS GAP

Live demo of application!

Link to Shinyapps.io

SUMMARY OF THE APPLICATION – EXAMPLE 1

- All data from project resides in one place
- Data is scalable
- Can download subsets of data if needed for other uses
- Sharable via a web address
- No software installation or license required
- Can be connected directly to a database
- Entire code-base can be re-used when we revisit the analysis

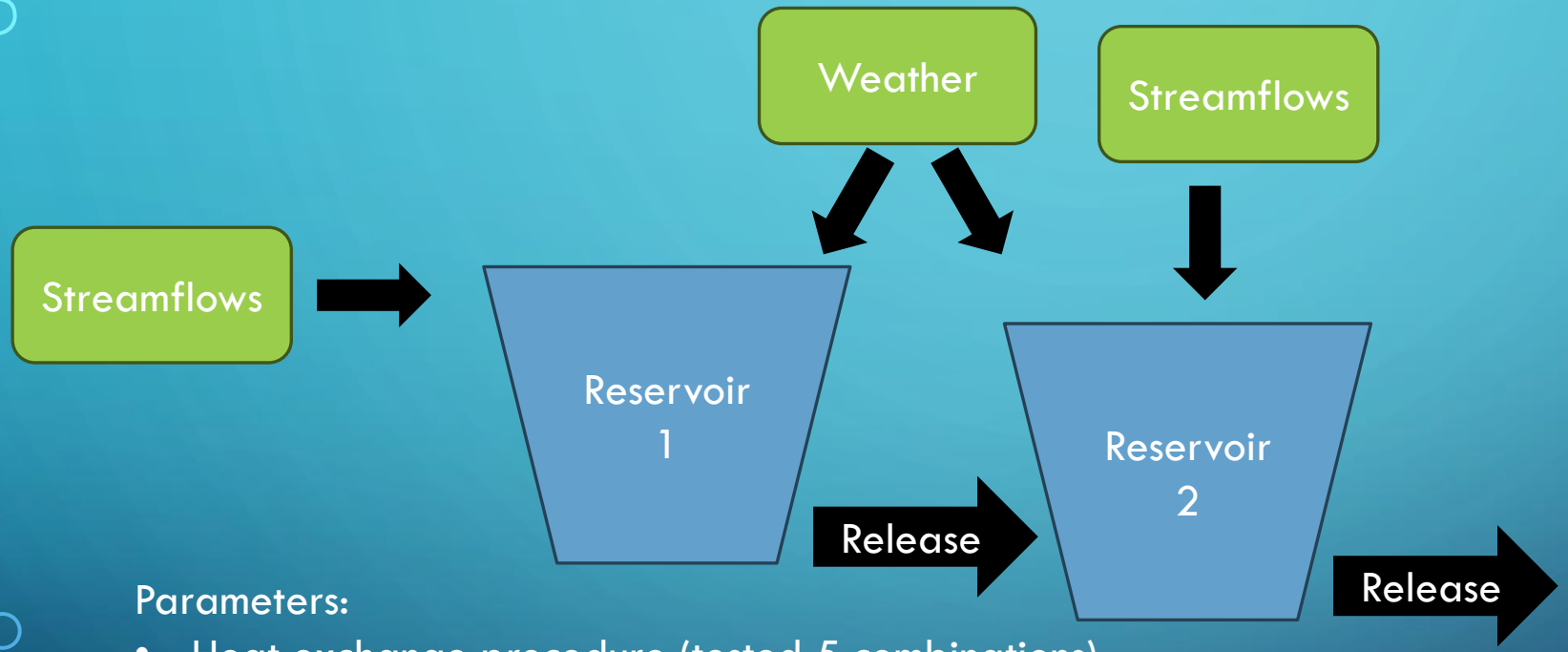
EXAMPLE 2: SENSITIVITY ANALYSIS OF OUR RESERVOIR TEMPERATURE MODEL (CEQUAL-W2)

- Trying to fix a “cold bias”
- Simulate operations and weather impacts
- CEQUAL-W2 has a lot of parameters
- How can we tune the calibration of CEQUAL-W2?

EXAMPLE 2: SENSITIVITY ANALYSIS OF OUR RESERVOIR TEMPERATURE MODEL (CEQUAL-W2)

- "One at a time" sensitivity analysis
- What parameter values "work best"
- Summarized objective functions only tell part of the story
- Requires filtering and comparison amongst different "scenarios" and years
- Results must be sharable with consultants

EXAMPLE 2: CEQUAL-W2 TEMPERATURE MODEL



Parameters:

- Heat exchange procedure (tested 5 combinations)
- Light extinction coefficient (tested 2 values)
- Mixing parameters (tested 10 combinations)
- Wind sheltering (tested 15 values)
- Intake tower partitioning (tested 33 combinations)
- Solar radiation corrections (tested 8 factors)
- Alternative SR datasets (tested 8 alternatives)

EXAMPLE 2: RESERVOIR TEMPERATURE

- 100 different scenarios
- 4 different years
- 2 different reservoirs
- 5 different elevations for looking at temperatures

4000 traces of temperature to consider

SHINY SAVES THE DAY AGAIN

Live demo of application!

Link to Shinyapps.io

SUMMARY OF THE APPLICATION – EXAMPLE 2

- All data from project resides in one place
- Data is scalable, filterable, zoomable
- Allows careful examination between similar (or different) parameter sets
- Allows quick examination between different years
- Allowed us to key in on what parameters where sensitive and insensitive
- All simulation data for each scenario stored in one file
- Generating scenarios mostly automated

SHINY IS RIGHT FOR YOU IF…

- You already use R
- You need to explore your data at different scales
- You need complex filtering that can be done on the fly
- You are going to re-run an analysis
- You need to share results with many stakeholders

SHOUT OUT FOR SHINYAPPS.IO

- Will host your Shiny application, very easy to use
- Has a free tier to get you started
- Paid tiers are incredibly affordable

QUESTIONS?

WHAT IS A PIRATE'S FAVORITE LETTER OF THE ALPHABET?