El Dia Project Proposal
How Many Monitoring Wells Are Enough?

Jessica Carvalho

Hypothesis

Groundwater models are commonly used to estimate water levels and support decisions about pumping, water supply, and long-term planning. These models are often "calibrated" by adjusting parameters until simulated groundwater levels match observed levels in monitoring wells. However, monitoring wells are expensive and unevenly distributed, and it is not always clear how many wells are enough (or where they should be located) to make the model trustworthy at places where we do not have measurements.

The Testable hypothesis: As the number of monitoring wells used for calibration decreases, prediction accuracy at unsampled locations will decline in a measurable way, and there will be a threshold well network density and/or spatial configuration below which the model becomes unreliable for predicting groundwater heads away from the remaining wells.

This addresses a common gap in hydrogeological practice. A model can fit the wells used for calibration but still produce unreliable predictions elsewhere. This project focuses on the difference between "a model that matches the data it was trained on" and "a model that can be trusted to predict groundwater levels where we lack direct observations."

Scientific & Societal Value

Scientific value: This project evaluates a core question in hydrologic modeling: "How observation networks control model reliability." It explicitly links calibration performance to predictive performance and helps clarify when a model's apparent success is driven by strong constraints (for example, boundary conditions or smoothing assumptions) rather than by data support. The results will improve understanding of how prediction and uncertainty respond to changes in observation density and spatial coverage.

Societal value/real-world applications: Monitoring wells cost time and money to drill and maintain. Knowing when additional wells yield diminishing returns and when a network is too sparse to support decision-making can help agencies and communities allocate resources more effectively. Outcomes could support groundwater management decisions, monitoring program design, and policy and planning. This connects to practical needs in Arizona and other water-limited regions, where groundwater models are routinely used to inform decisions affecting communities, ecosystems, and agriculture.

Analysis Plan

Models and observations

- A groundwater flow model built in the MODFLOW 6 framework
- Observed groundwater head measurements at wells
- A calibration workflow designed to be repeatable across scenarios
- Model calibration and uncertainty analysis will be implemented using PEST++, with PyCap supporting a Python-based workflow.

Governing equations

Groundwater flow in a saturated porous medium follows conservation of mass combined with Darcy's law. A common transient form of the groundwater flow equation is:

$$S_s \frac{\partial H}{\partial t} = \nabla \cdot (k \nabla H), \qquad q = -k \nabla H$$

Planned analysis:
1) Baseline case (full network): Calibrate the model using the full monitoring-well dataset and record fit metrics (for example, error statistics at calibration wells) as a reference.
2) Progressive downsampling: Create multiple reduced-well scenarios by removing wells in steps (e.g., 100%, 75%, 50%, 25% of the network). If feasible, test different removal patterns (e.g., random removal versus spatially clustered removal) to evaluate the importance of spatial coverage.
3) Recalibrate consistently: Recalibrate the model for each reduced network using the same workflow and assumptions to keep cases comparable.
4) Out-of-sample reliability testing: For each scenario, evaluate predictive skill at withheld wells (wells not used in calibration) to measure how well the model performs at unsampled locations.

Outcomes that would support the hypothesis:
• Prediction error at withheld wells increases as wells are removed.
• A break point where prediction errors rise rapidly once the network becomes too sparse.
• Increasing instability in the simulated head surface

Outcomes that would refute the hypothesis:
• Prediction accuracy at withheld wells remains relatively stable even after removing wells.

Alternative explanations:
• The model may appear robust because boundary conditions or model assumptions heavily constrain results rather than because observation data are sufficient.
• Measurement noise or inconsistent observation periods could mimic reduced reliability.

Milestones March-June 2026
• Define evaluation metrics (e.g., RMSE or MAE) and implement withheld-well testing.
• Create the first draft version of the conceptual workflow figure.
• Evaluate whether reliability depends more on network size or spatial coverage.
• Draft initial Results and Discussion bullets.
• Finalize the reliability-threshold plot and a summary table of results.
• Write a short Methods and Results narrative.