HTCondor and Monte Carlo techniques to evaluate the uncertainty of water sources to groundwater wells

Mike Fienen
mnfienen@usgs.gov
Research Hydrologist
Paul Juckem
Hydrologist
USGS Wisconsin Water Science Center

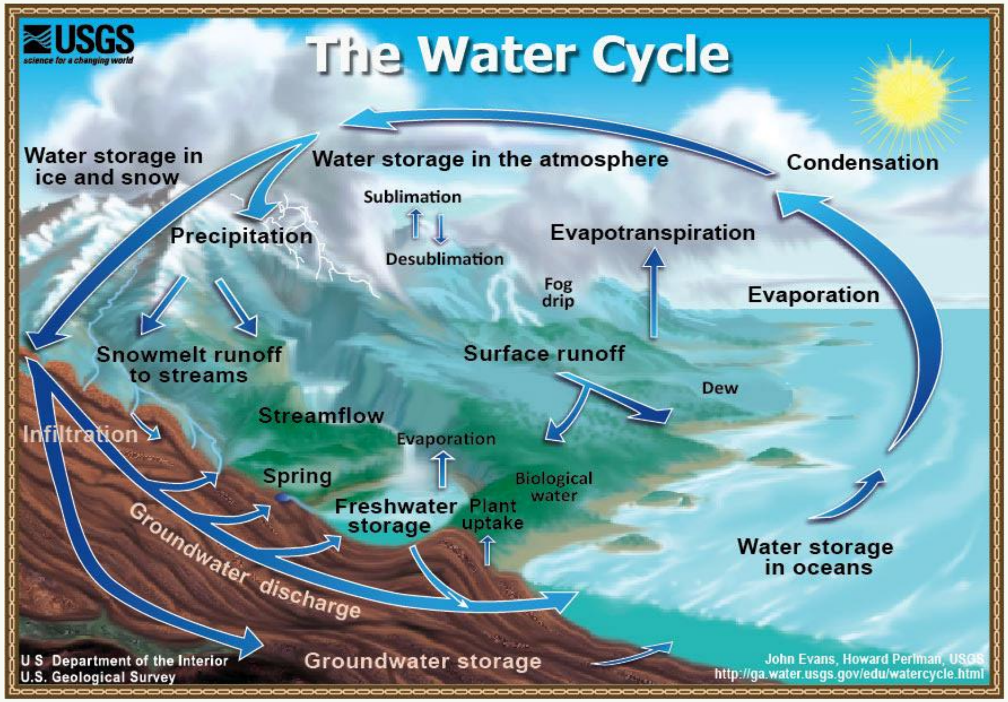
HTCondor Week—2014
Madison, Wisconsin

USGS

The Source of Water to Wells

Groundwater is part of the hydrologic cycle

We can identify the surface expression of where water originates relative to a well



USGS

Methods to determine source of water to a well

The classic method is the fixed-radius method

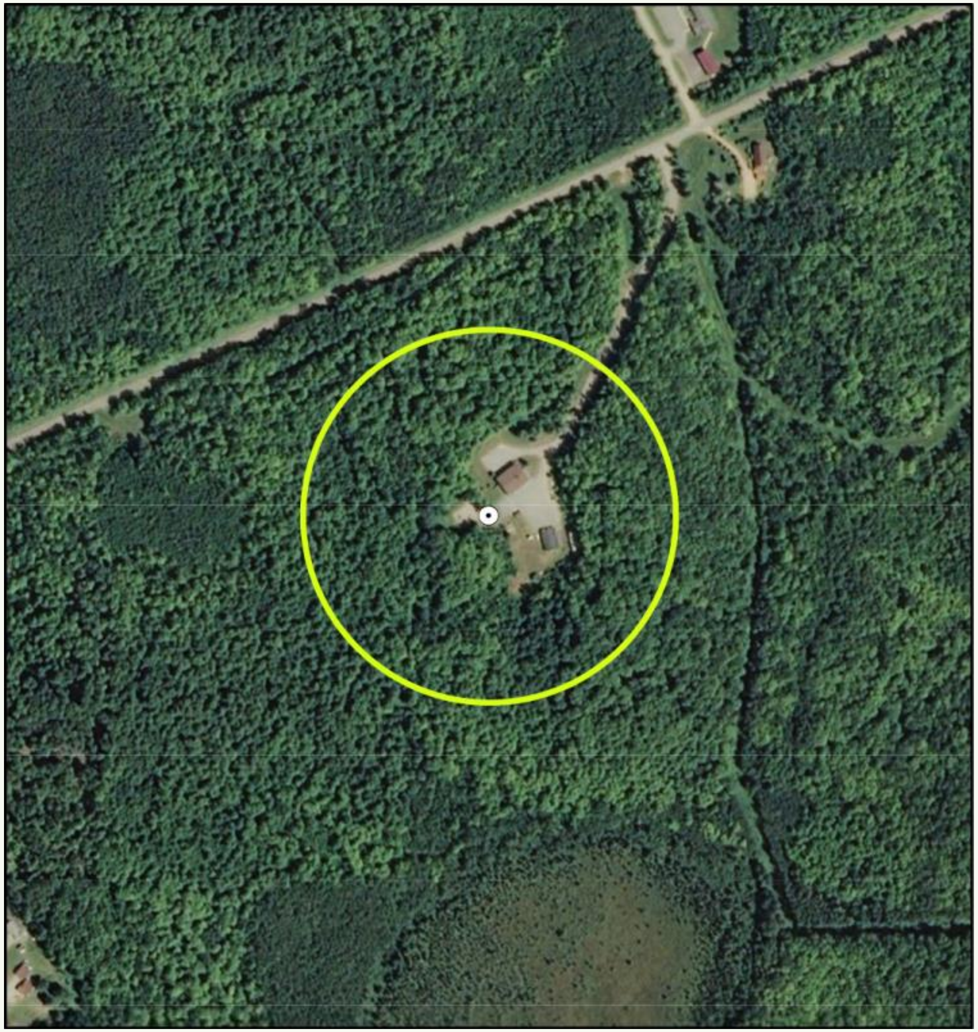
$$r = \sqrt{\frac{Qt}{\pi n H}}$$

where

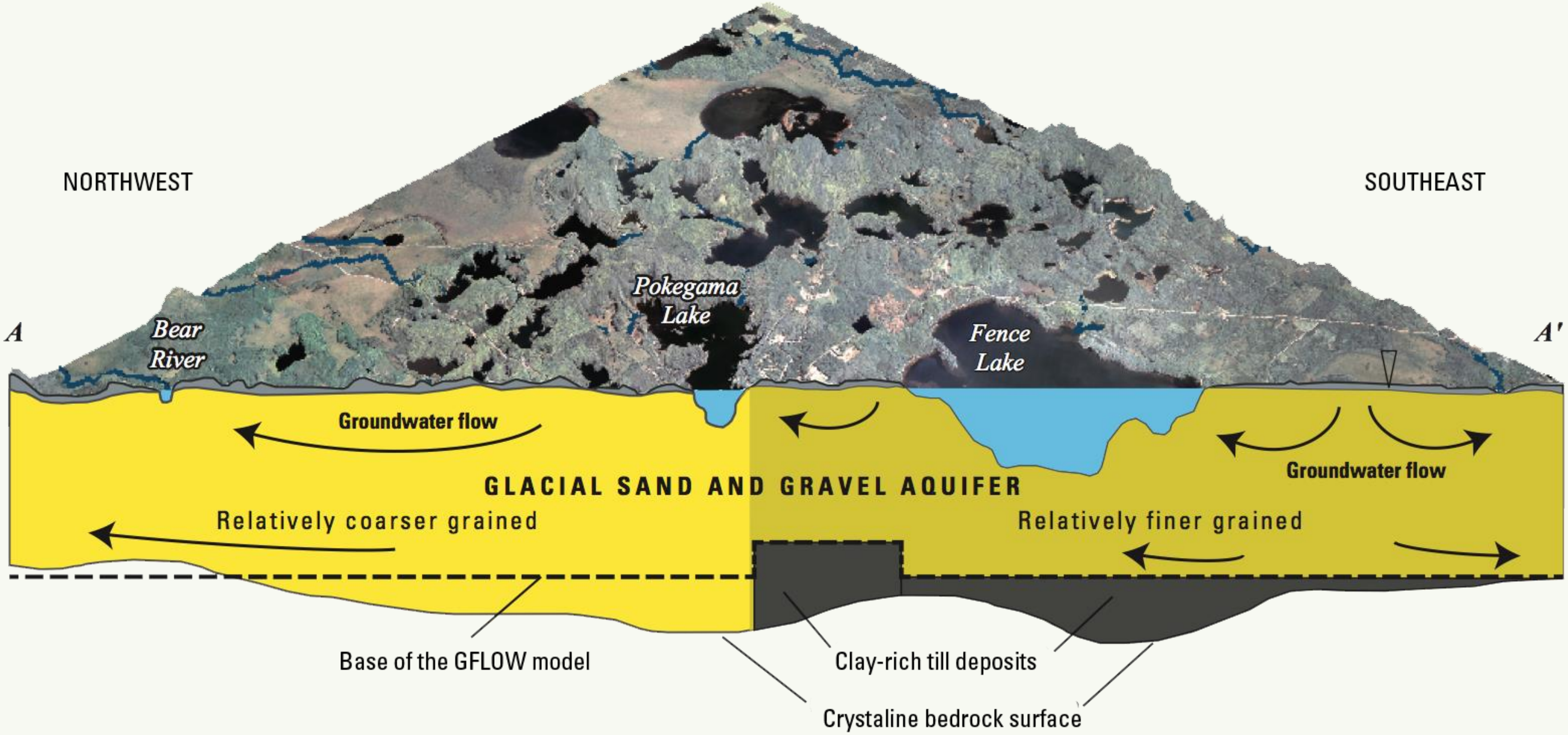
r	is the radius,
Q	is the pumping rate,
t	is the time of pumping,
π	is the mathematical constant,
n	is the porosity, and
H	is the aquifer thickness.

USGS

Fixed radius results



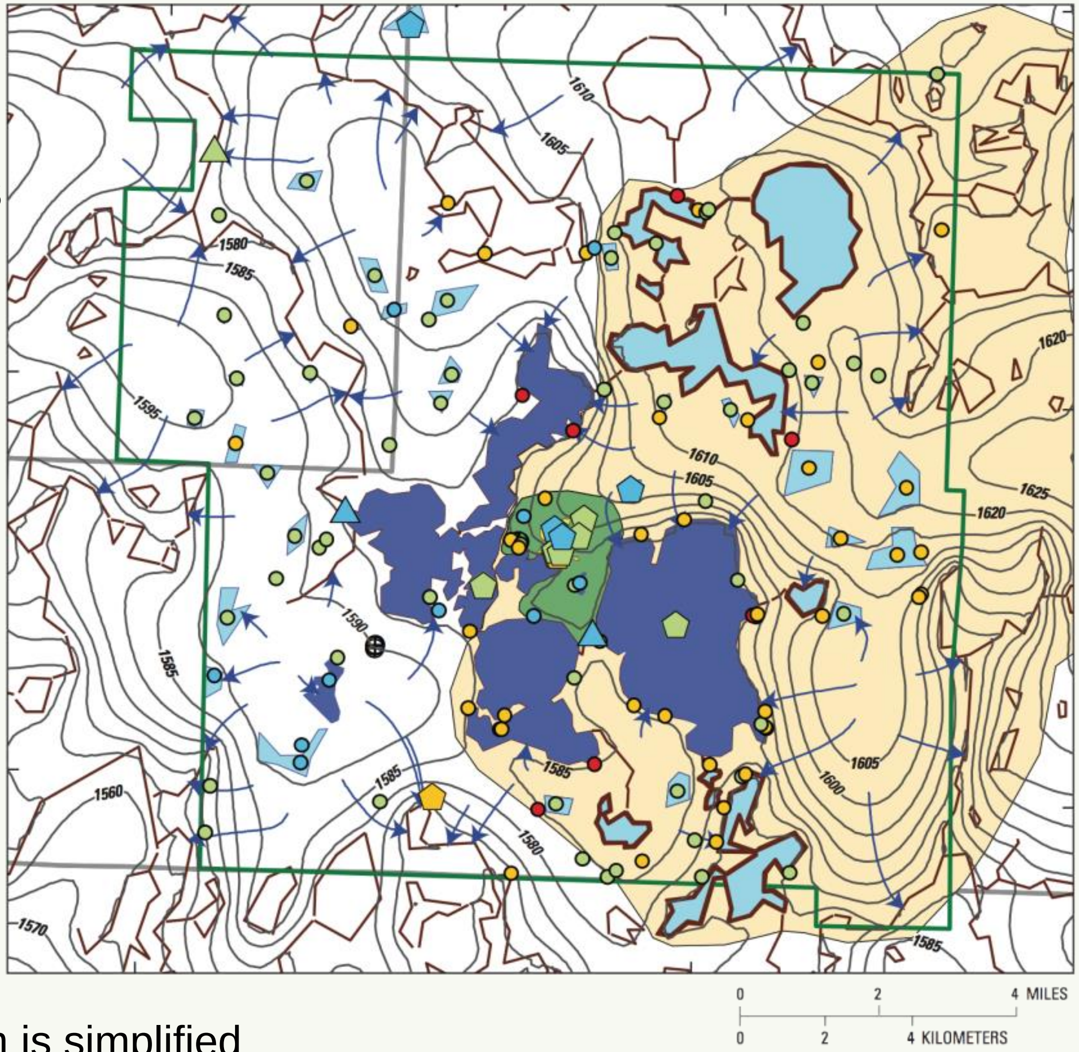
Reality is a bit more complicated



GFLOW is an analytic element model simulating groundwater and surface water*

GFLOW superposes analytic solutions representing elements

No numerical grid (!)



* Surface water system is simplified

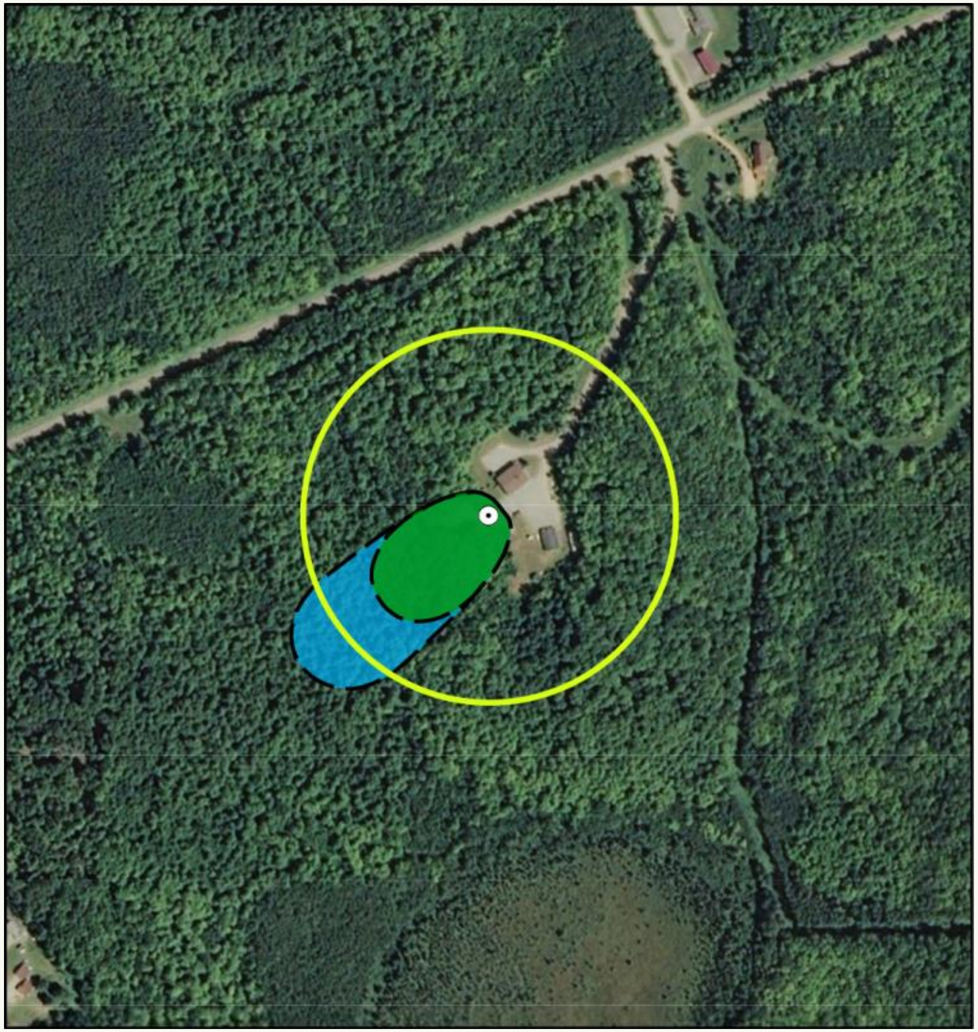
USGS

Using GFLOW to estimate area of capture

This is clearly an improvement but what about uncertainty?

We can vary parameters about their calibrated values and see how that uncertainty propagates into the capture estimates

Monte Carlo with Latin Hypercube sampling strategy



Monte Carlo with Latin Hypercube: Starting with calibration

PEST – *Parameter calibration and covariance*

- Calibrate to heads, gradients, and flows with hydraulic conductivity (K), recharge (R), and sediment resistance (C)
- Covariance matrices for calibrated parameters; estimated variances for non-calibrated parameters

Monte Carlo – *Uncertainty analysis*

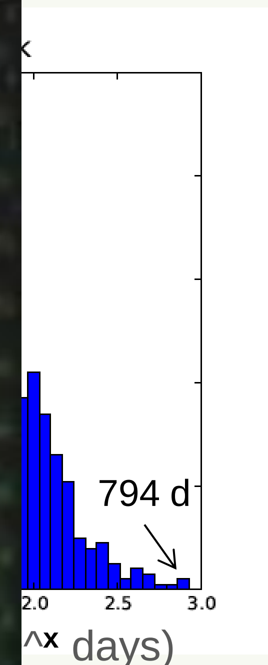
- Latin Hypercube sampling of covariance
- (Starn and Bagtzoglou, 2012)

Custom Python codes

- End-point analysis for Areas Contributing Recharge to wells
- Aquifer Mapping to synthetic grids for source tracking
- Evaluate convergence of the Monte Carlo simulations

Monte Carlo with Latin Hypercube: a few more details

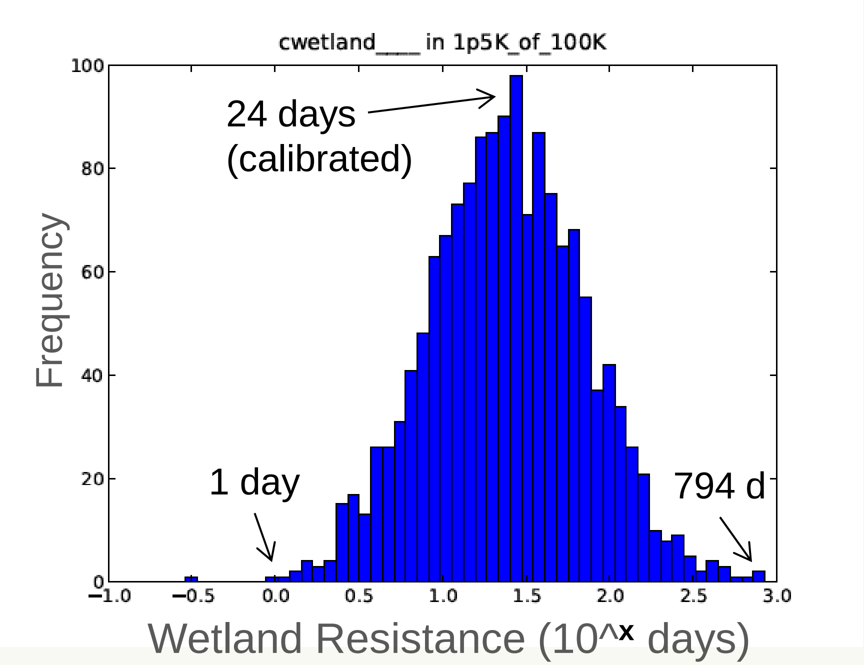
Latin H
sampli
covaria



Monte Carlo with Latin Hypercube: a few more details

Latin Hypercube: Equiprobable sampling that honors parameter covariance and correlation

Model is run for each parameter set (1200 realizations)



Model runs distributed on 150 Linux CPUs using Wine with HTCondor

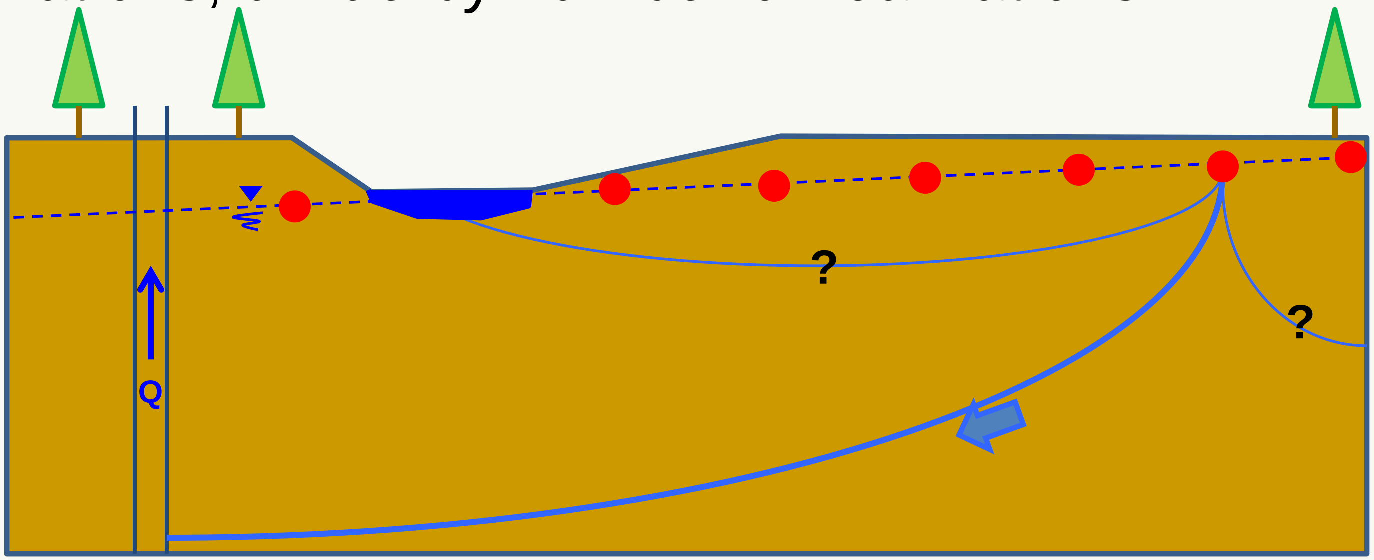
Areas Contributing Recharge to Wells (end-point analysis)

Forward particle tracking

Read pathline output & search for the well ID

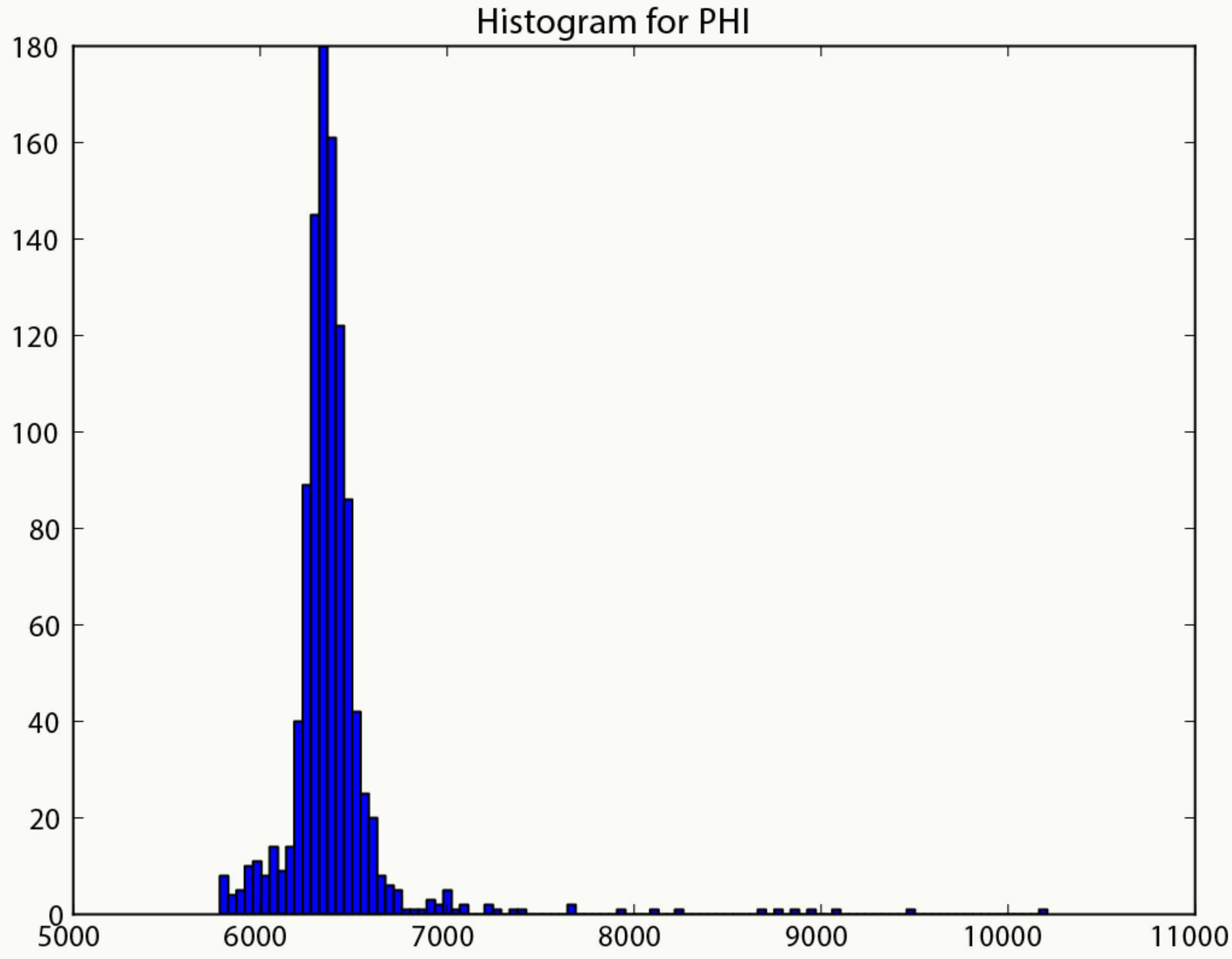
Captured by the well? Yes = assign 1 to the particle starting location for this realization

Sum results for a grid of particles and 1000s of realizations; divide by number of realizations



Evaluate whether realizations kill our calibration

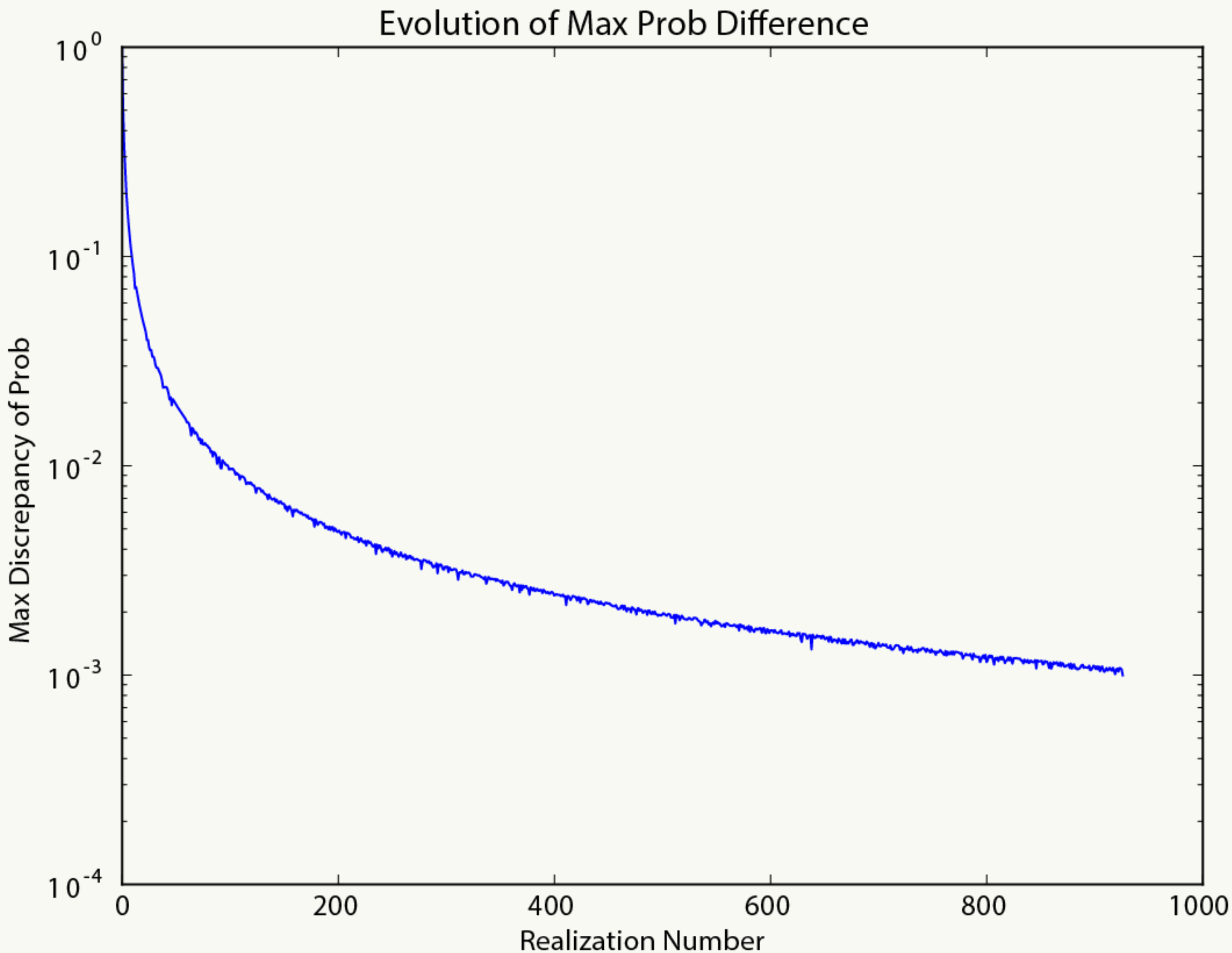
Throw out a few realizations that either failed or where PHI was too high or "uncalibrated"



USGS

Declare convergence when probability from one realization to the next is below a threshold

Threshold here was 0.001 or 0.1%



USGS

Performance

Each forward run takes about 45 minutes

GFLOW has no grid, so the particle tracking work is done by an external Python script – tack on another 15 minutes

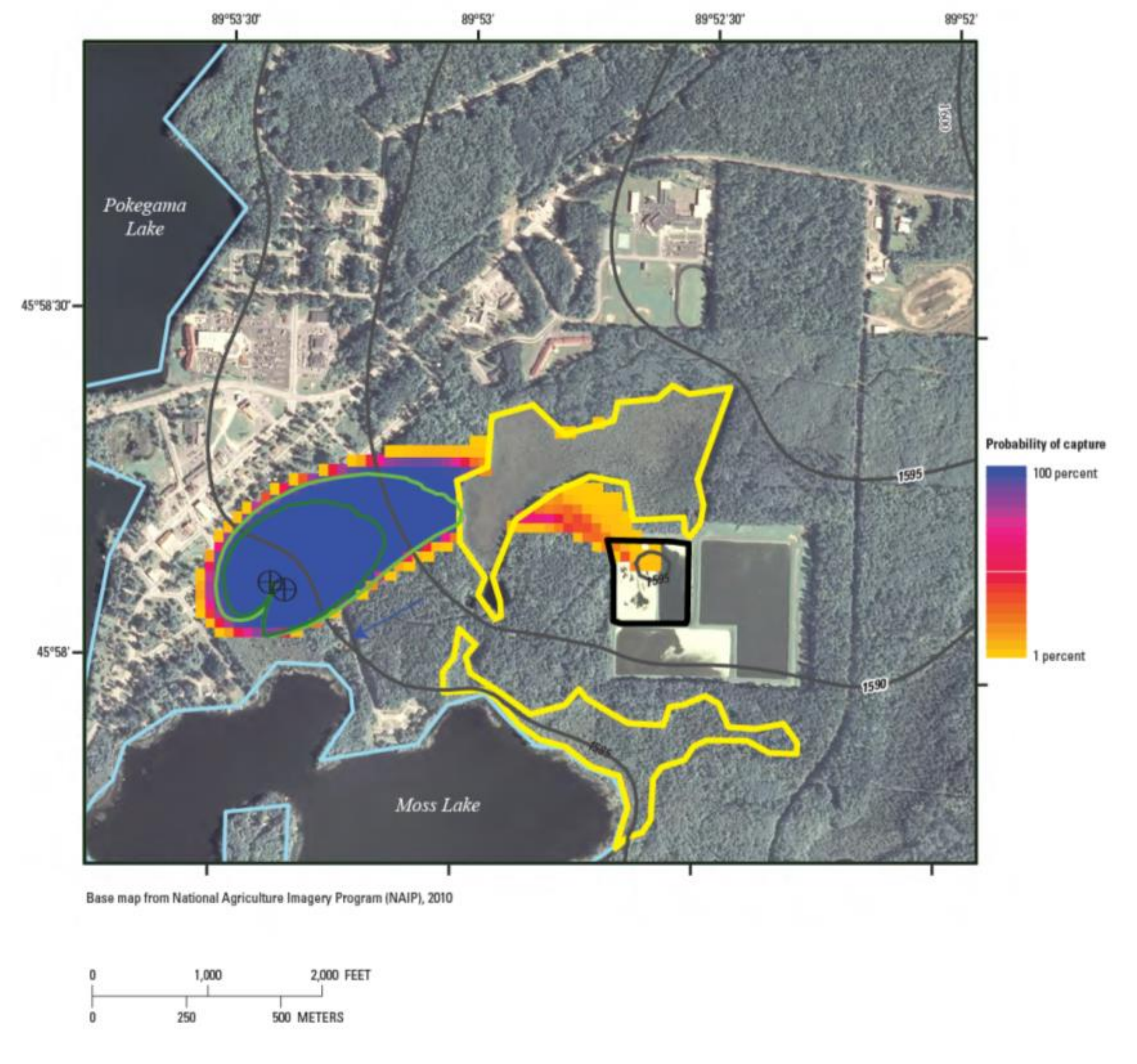
Need roughly 1,200 runs to be sure of convergence

HTCondor makes this all possible!

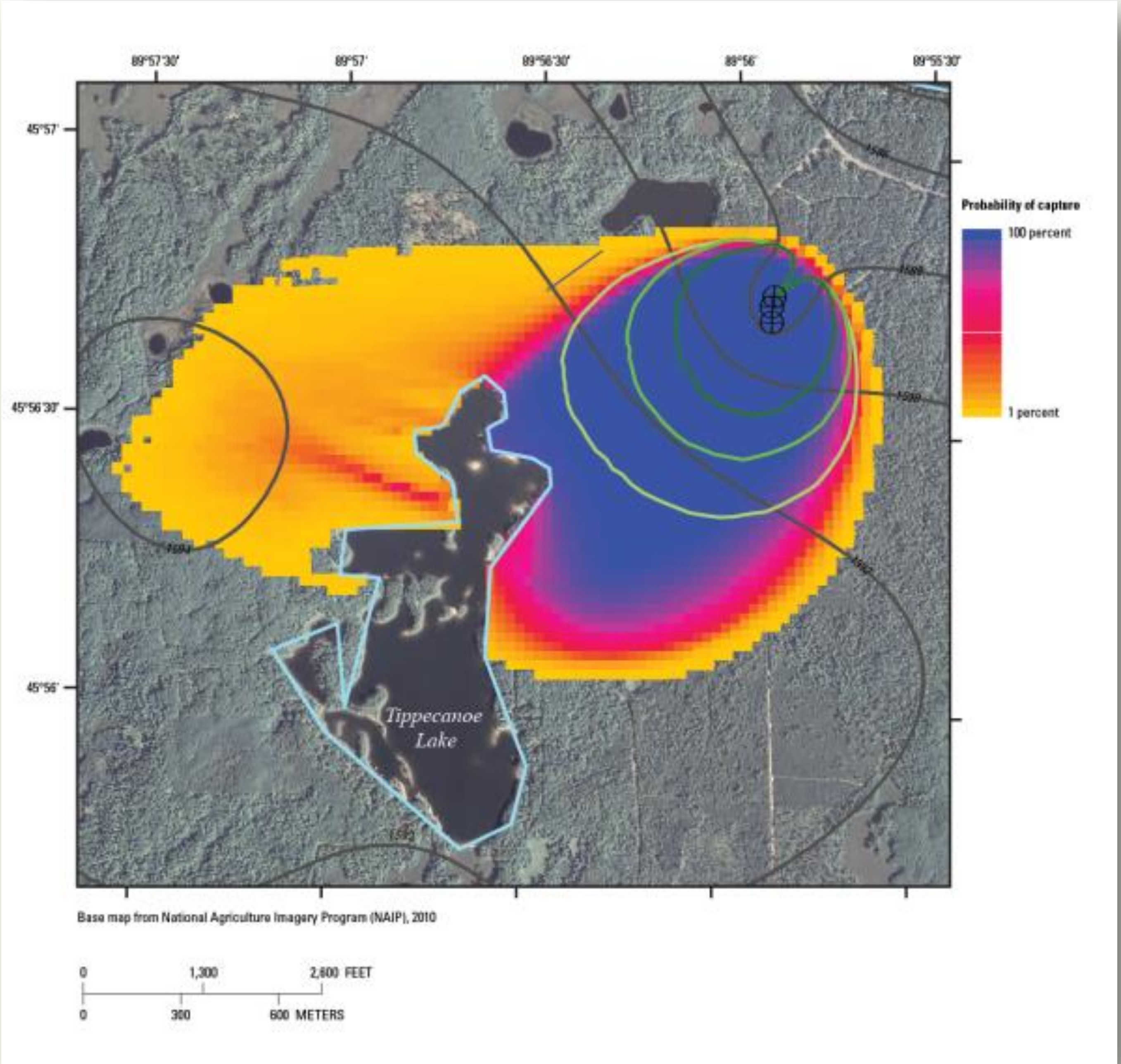
Use Python scripts to assemble files, push to a submit node, launch the HTCondor jobs, and postprocess results.

Maybe this is a good candidate for DAGMan?

Area contributing recharge considering uncertainty

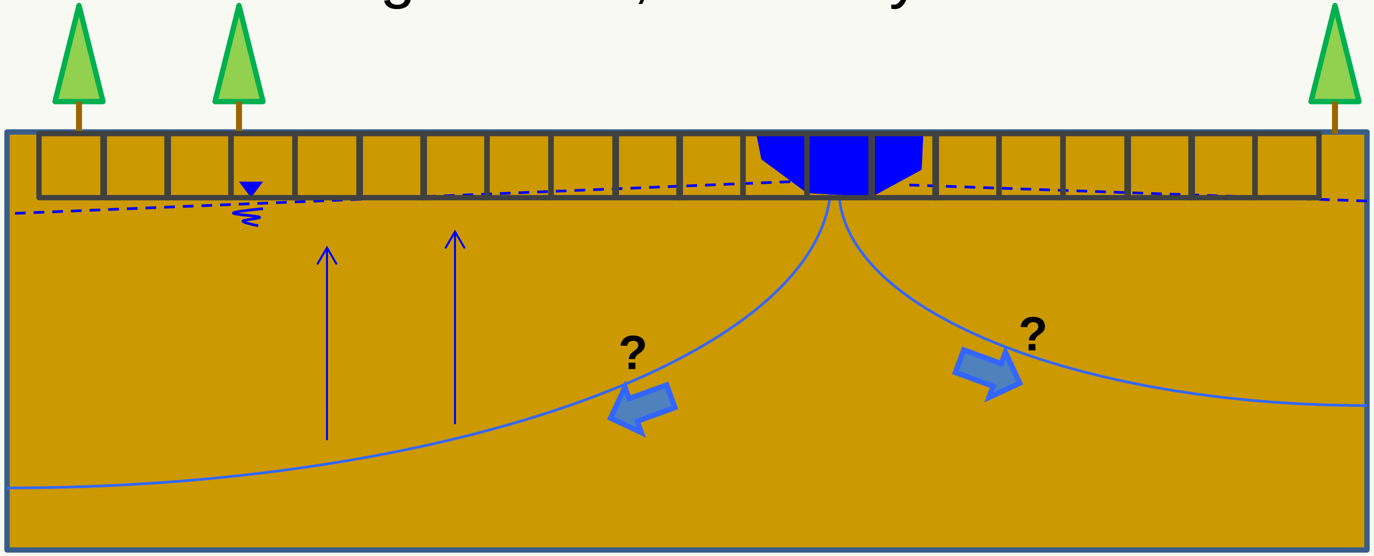


Area contributing recharge considering uncertainty



Wastewater Source Tracking
(where did the pooky go? pathline projection to grids)

- Generate synthetic grid
- Forward particle tracking (track the source)
- Project pathline onto grid
- Does a pathline traverse a specific cell? Yes = assign 1 to the cell for this realization
- Sum results for all grid cells; divide by “n” realizations



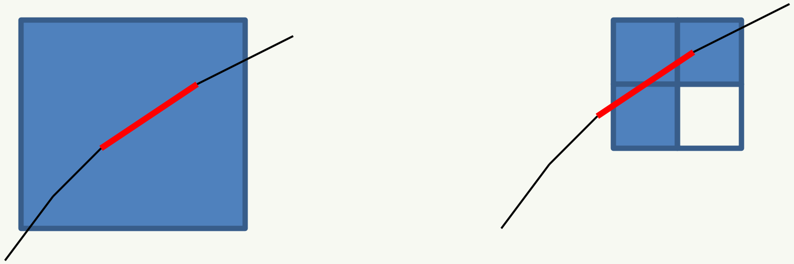
USGS

Projecting a Pathline to a Grid

Process sequentially along each pathline

Identify nearby grid nodes

Make a map of which cells are traversed by a pathline



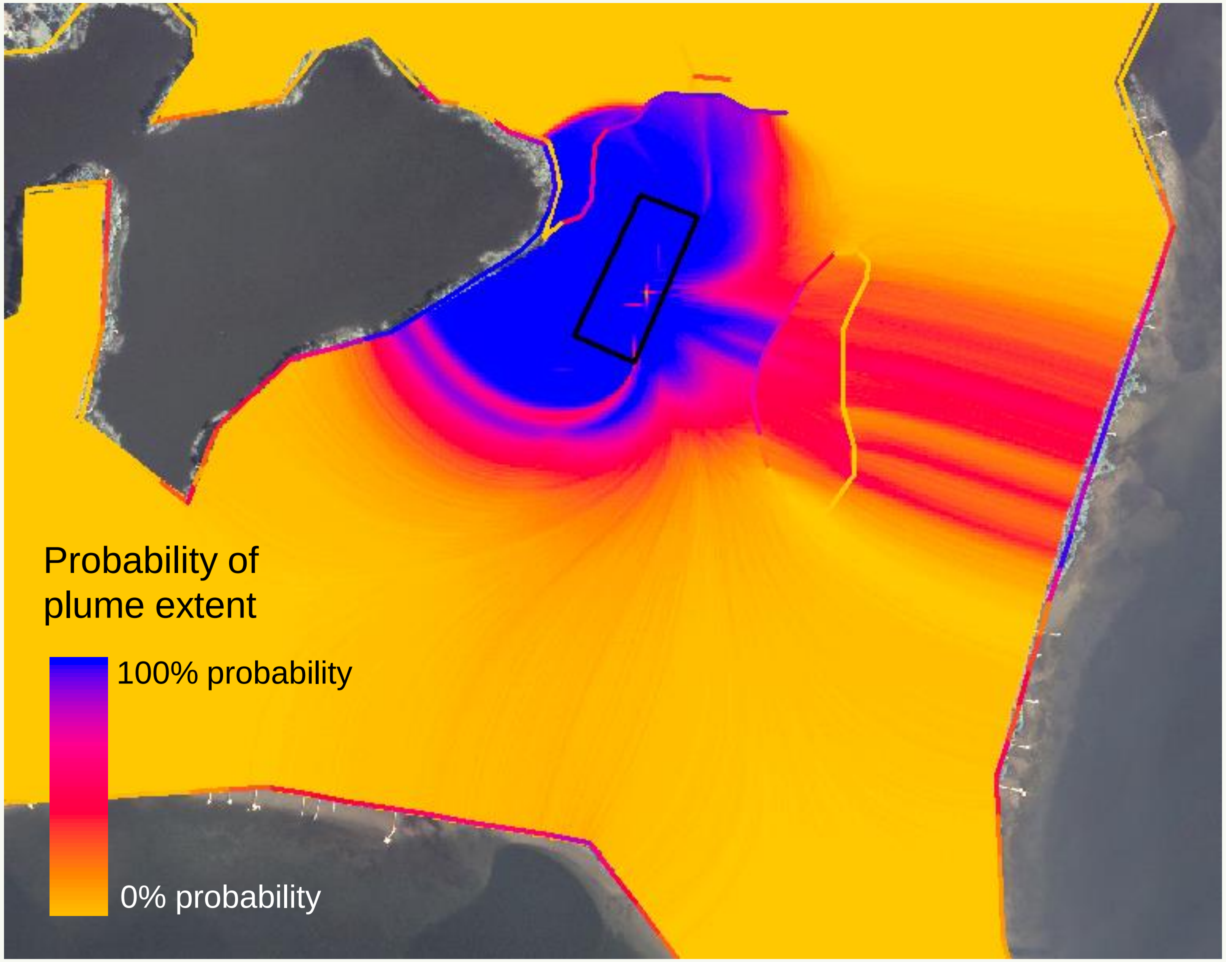
USGS

Wastewater Source Tracking
Grid spacing = particle step size (5m)

1000 particles
(forward tracked)

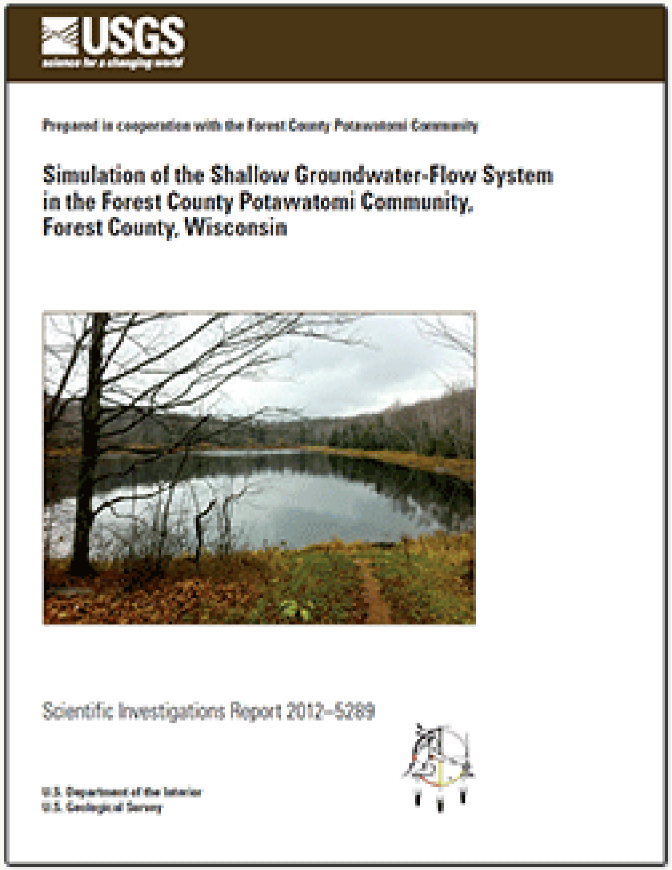
Probability of at least one particle intersecting a cell – “probability of plume extent”

Considerations:
- Number of particles
- Particle step size
- Grid spacing
- Weak sinks



USGS

Any Questions?



For more information:

http://pubs.usgs.gov/sir/2012/5289/ http://pubs.usgs.gov/sir/2014/5020/

mnfienen@usgs.gov

USGS