Gaussian Process Regression Modeling on a Groundwater Plume

Halle Ritter

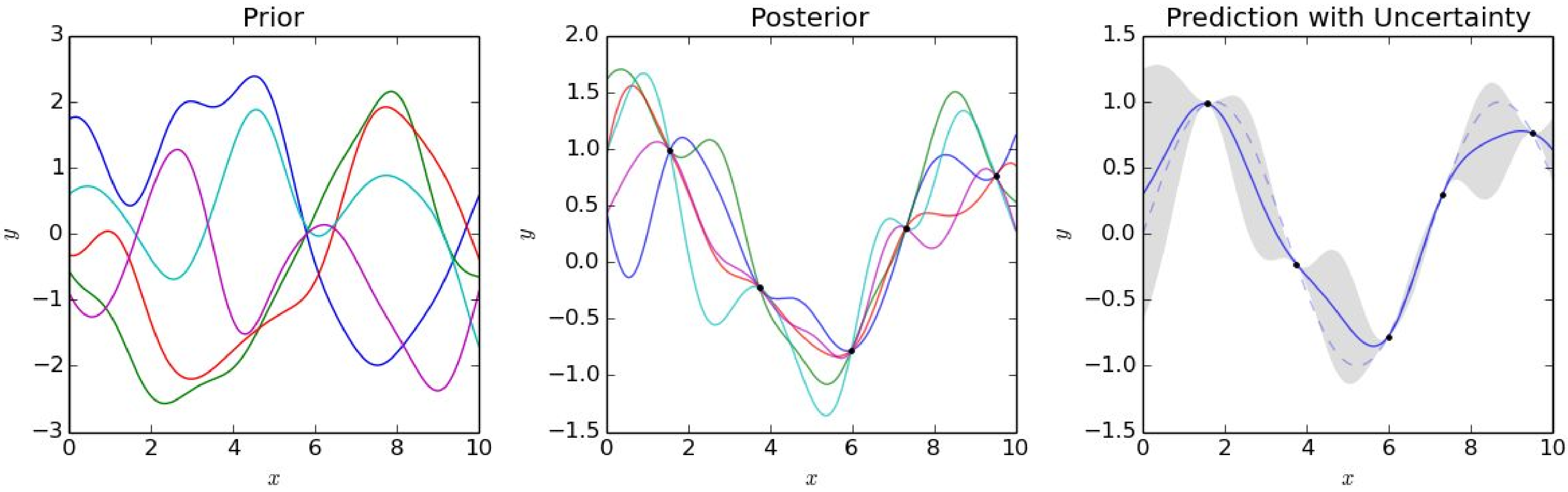
Problem Statement

- Sampling is expensive
- What do you do when geology -> gw flow is complicated?
- What if you could predict where to sample purely statistically?

Concept

- Gaussian process regression
  - Assumptions: There exists an underlying process generating noisy data
  - Using as few input points as possible
- Picks ideal plume sampling locations
  - For max concentrations
  - For plume boundaries
  - For minimizing uncertainty

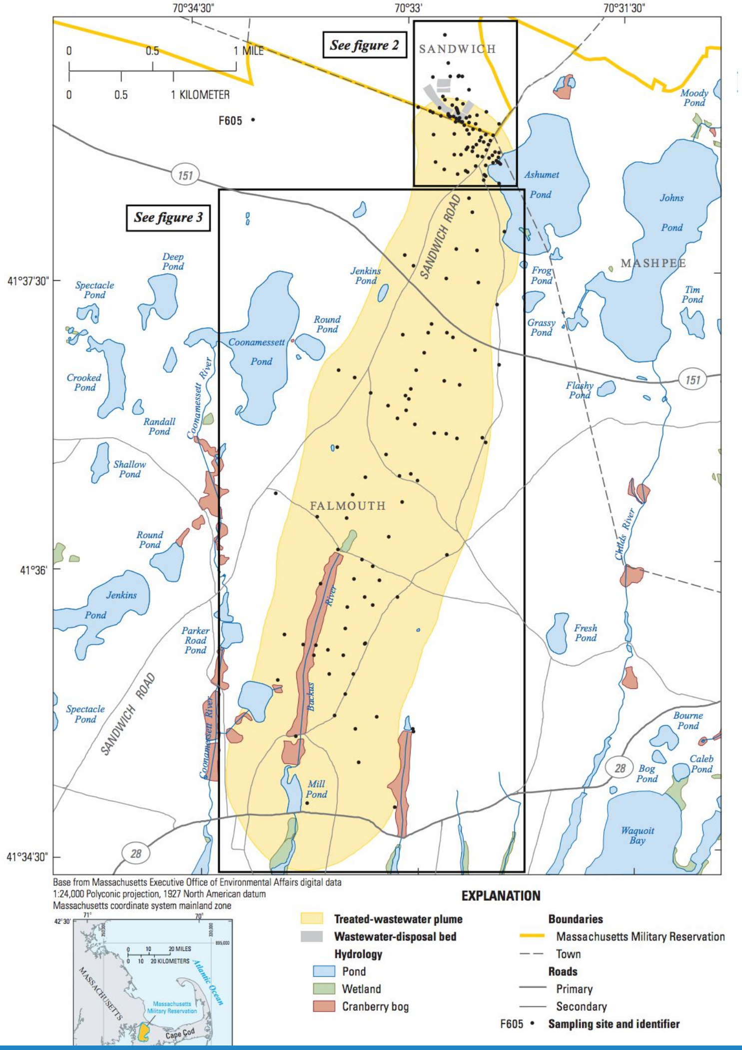
Gaussian Process- Theory Visualized



Source: Wikimedia Commons

Process assumes as a prior some probability density function over all points in space, and at each point a probability density function of certainty. Each data point given to the algorithm narrows the window of uncertainty of function location.

Example Site- Mass. Military Academy



(Savoie, et al.)

3D Model

- Kernel selection
  - *RBF*- general-purpose
  - *Matern/Exponential*: from fractional advection-dispersion equation (power law scaling of variance)
- Validated w/ test/train split to avoid overfitting small data set
- Performs well - two kernels very similar

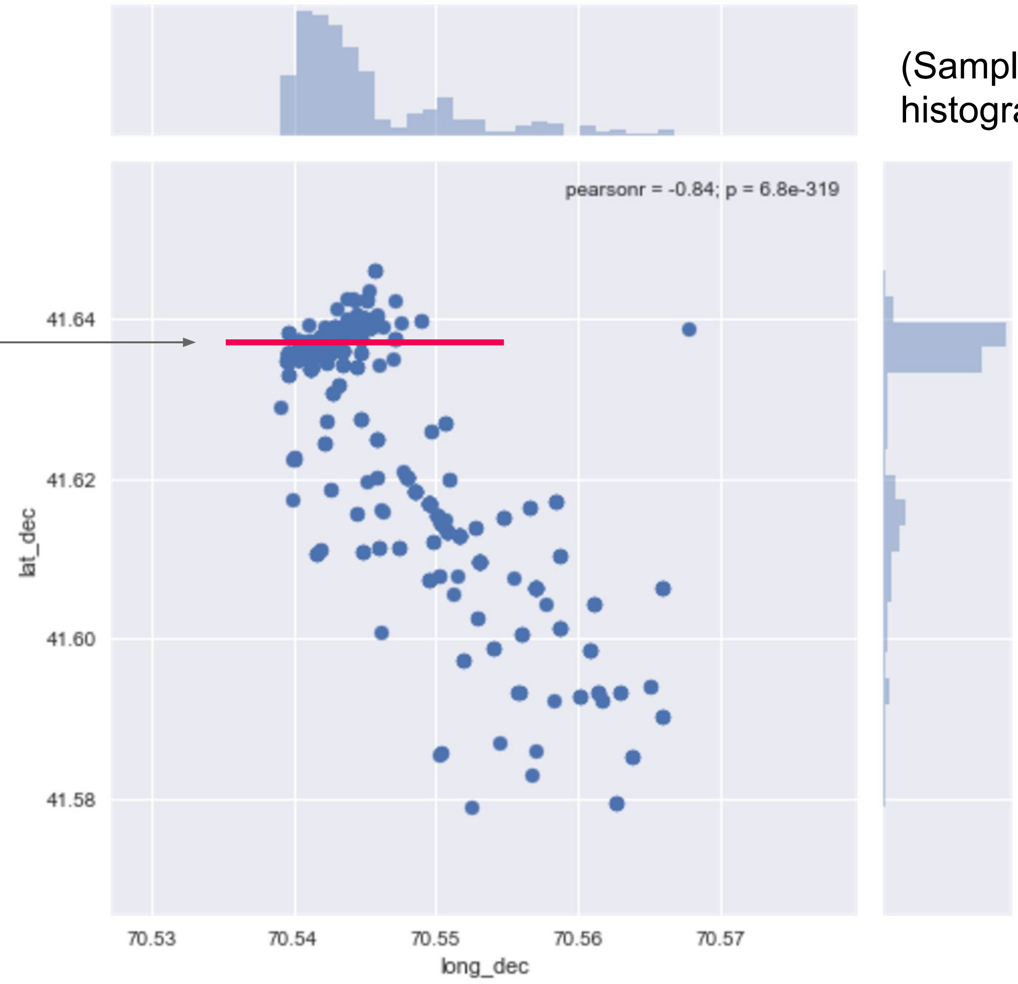
1D Model

- ▷ However: 3D model difficult to visualize
- ▷ Took “1D slice” of data for visualization/demonstration purposes

Sampling Locations

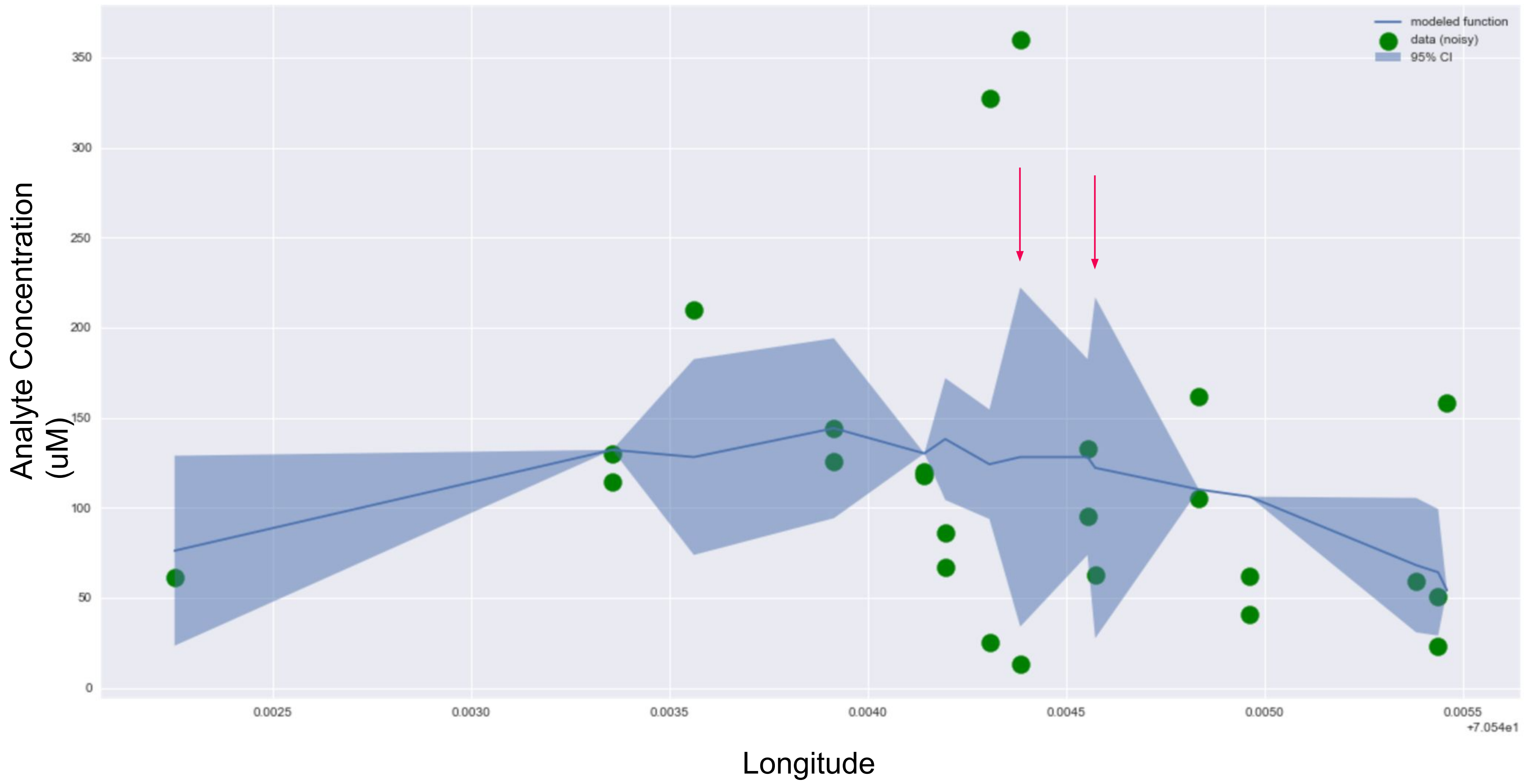
(Sampling density histograms)

“1D Slice”



Visualization of model results

- Following slide: Concentration of analyte vs longitude
  - “Noisy” data points
  - Predicted model fit for underlying ideal process
  - 95% CI on location of ideal process
- Practical use: e.g. at red arrows, take additional sampling points to narrow uncertainty window


modeled function
data (noisy)
95% CI
350
300
250
200
150
100
50
0
0.0025
0.0030
0.0035
0.0040
0.0045
0.0050
0.0055
+7.054e1
Analyte Concentration (uM)
Longitude

Next steps

- Finding boundaries
- Applying to other plumes
- Gaussian process regression: wide applicability of algorithm
  - Sensor placement
  - Demand modeling
  - Optimizing neural nets & other algorithms

Data Source:

Savoie, J.G., LeBlanc, D.R., Fairchild, G.M., Smith, R.L., Kent, D.B., Barber, L.B., Repert, D.A., Hart, C.P., Keefe, S.H., and Parsons, L.A., 2012, Groundwater-quality data for a treated-wastewater plume near the Massachusetts Military Reservation, Ashumet Valley, Cape Cod, Massachusetts, 2006–08: U.S. Geological Survey Data Series 648, 11 p., at https://pubs.usgs.gov/ds/648.

Contact

Halle Ritter

M.Eng., Environmental Engineering

MIT

hritter@alum.mit.edu

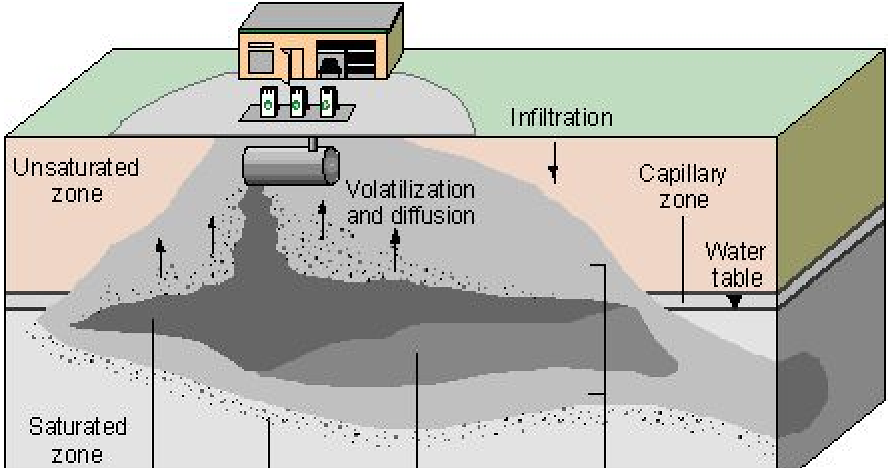
linkedin.com/ritterh

Appendix

(for Tech/ Non-Environmental Audiences)

Groundwater Plumes

- A chemical or waste spill into the water supply forms an aquifer "plume"



Source: US EPA

- We need to know where it is
  - So nobody drinks it
  - So we can clean it up

Evaluating Performance

- Numerical evaluation difficult to interpret
  - Goal: recover underlying process from noisy input data
  - Use process to guide additional sample selection