

DOE-FIU Cooperative Agreement Annual Research Review – FIU Year 1

Al for Soil & Groundwater Data (SRS F-Area) Aurelien Meray (DOE Fellow)



Advancing the research and academic mission of Florida International University



Task 8: AI for EM Problem Set (Soil & GW):Data analysis and visualization of sensor data from the wells at the SRS F-Area using machine learning

Overall Needs:

 Develop machine learning tools to automate the monitoring and forecasting of contaminant transport dynamics at the Savannah River Site (SRS) F-Area to support DOE-EM's goal for long time monitoring of contaminated groundwater sites.



Objectives:

- Develop data exploration tools for understanding the spatial and temporal distribution of the F-Area dataset.
- Develop a spatial interpolation approach for estimating a plume.
- Examine proxy variables at the site.
- Develop a sensor placement optimization approach for identify a subset of wells that captures the overall plume dynamic.





FIU Year 1 Research Highlights:

- Trend analysis
- Correlation analysis
- **Preprocessing/Transformation functions** (interpolation etc.)
- Insightful visualizations
- Spatial estimation
- pyLEnM Python package for Long-term **Environmental Monitoring**



Prediction: 2013-09-28



FSB 95DR correlation



FSB110D - TRITIUM





Applied Research Center FIU Year 1 Research Highlights:

- Trend analysis
- Correlation analysis
- Preprocessing/Transformation functions (interpolation etc.)
- Insightful visualizations
- Spatial estimation
- pyLEnM Python package for Long-term Environmental Monitoring









PyLEnM package:



Visualization functions:

plot_data

plot_all_data

- plot_correlation_heatmap
 - plot_all_correlation_heatmap
- plot_corr_by_well
 - plot_all_corr_by_well
- plot_corr_by_date_range
- plot_corr_by_year
- plot_MCL
- plot_PCA_by_date
- plot_PCA_by_year
- plot_PCA_by_well
- plot_coordinates_to_map
- plot_date_ranges





Spatial Estimation:

E 3.6820

3.6815

3.6810

(NAD83)

INPUTS PROCESS OUTPUT Regression 4.22 3.6822 Trained on: 4.20 E 3.6820 4.18 X = coordinates & elevation at wells 4.16 õ 3.6818 y = concentration at wells . 9 3.6816 4.12 3.6814 3.6812 4.08 Predict on: entire site coordinates 4.365 4.370 ×10³ Easting (NAD83), m 3.6810 E 3.6820 4.362 4.364 4.366 4.368 4.370 4.372 Easting (NAD83), m **Elevation at entire site** ×10⁵ (SRTM [~30 meters]) **Concentration at wells** $\times 10$ **Gaussian Process** 0.015 4.365 4.370 3.6822 -0.010 Easting (NAD83), m Trained on: 0.005 € 3.6820 0.000 X = coordinates at wells 3.6818 -0.005 3.6816 y = regression residual $R^2 = 0.963$ -0.010 3.6814 -0.015 concentration at wells -0.020 3.6812 -0.025 Predict on: entire site coordinates 3.6810 4.362 4.364 4.366 4.368 4.370 4.372 Easting (NAD83), m $\times 10^{5}$

4.225

4.200 E

1.175

1.150 1.125

4.100

075

×10⁵





Proxy Spatial Estimation:

- Explored proxy variable relationships.
- Observed a consistent correlation between Specific Conductance (SC) and Tritium (R² above 0.8).
- Spatially estimated Tritium by using the in-situ parameter SC.





Test set (20% of data) prediction



Well Optimization:

• The main idea of the optimization is to minimize the overall error (MSE) between the high-quality reference field and the spatial interpolated map with the 15-20 subset wells.



Water Table optimization (single timestep)





Well Optimization:

- Majority voting approach to capture multiple parameters (8 analytes) and multiple time steps.
- Results:
 - 'FSB 97D', 'FSB 95DR', 'FSB 78', 'FSB126D', 'FSB130D', 'FSB 79', 'FEX 4', 'FSB 99D', 'FSB 91D', 'FSB118D', 'FSB124D', 'FSB128D', 'FSB135D', 'FSB132D', 'FSB138D'
 - ALTEMIS Team chose:
 - 13/15 within top 20
 - 2/15 within top 22

Rank	well name	W/O preset Run 1 (Count)	W/O preset Run 2 (Count)	W/O preset Run 1 (%)	W/O preset Run 2 (%)	W/O preset AVGERAGE
1	FEX 4	30	26	93.75%	81.25%	87.50%
2	FSB135D	24	24	75.00%	75.00%	75.00%
3	FSB 91D	24	20	75.00%	62.50%	68.75%
4	FSB 92D	24	20	75.00%	62.50%	68.75%
5	FSB126D	22	21	68.75%	65.63%	67.19%
6	FSB 87D	18	23	56.25%	71.88%	64.06%
7	FSB124D	20	20	62.50%	62.50%	62.50%
8	FSB 93D	22	16	68.75%	50.00%	59.38%
9	FBI 17D	19	16	59.38%	50.00%	54.69%
10	FSB132D	16	19	50.00%	59.38%	54.69%
11	FSB130D	18	16	56.25%	50.00%	53.13%
12	FSB 99D	20	13	62.50%	40.63%	51.56%
13	FSB 97D	17	16	53.13%	50.00%	51.56%
14	FSB128D	14	19	43.75%	59.38%	51.56%
15	FSB 79	16	16	50.00%	50.00%	50.00%
16	FSB112DR	12	20	37.50%	62.50%	50.00%
17	FSB 95DR	15	16	46.88%	50.00%	48.44%
18	FSB 98D	12	19	37.50%	59.38%	48.44%
19	FSB136D	11	20	34.38%	62.50%	48.44%
20	FSB118D	17	13	53.13%	40.63%	46.88%
21	FSB138D	13	17	40.63%	53.13%	46.88%
22	FSB 78	17	12	53.13%	37.50%	45.31%
23	FSB137D	15	14	46.88%	43.75%	45.31%
24	FSB104D	12	16	37.50%	50.00%	43.75%
25	FSB 94DR	15	11	46.88%	34.38%	40.63%
26	FOB 13D	11	15	34.38%	46.88%	40.63%
27	FSB120D	12	13	37.50%	40.63%	39.06%
28	FSB134D	11	14	34.38%	43.75%	39.06%
29	ESB133D	10	15	31 25%	46 88%	39.06%

Multi-analyte optimization (Multiple timesteps)





Acknowledgments

• DOE-FIU Science and Technology Workforce Development Program

- Dr. Leonel Lagos, FIU ARC
- Dr. Ravi Gudavalli, FIU ARC

Mentors and Collaborators

- Dr. Haruko Murakami Wainwright, LBNL
- Dr. Himanshu Upadhyay, FIU ARC
- Dr. Masudur Siddiquee, FIU ARC
- Sponsored by the U.S. Department of Energy, Office of Environmental Management, under Cooperative Agreement #DE-EM00005213.

