Gaining Insights into Risks to Groundwater-
using Big Data and Machine Learning

Coping with Complexity

Efforts to gain more insight into potential risks to groundwater that coincide with energy development and production requires us to better understand the dynamics and interactions across the entire engineered-natural system.

Addressing the “Elephant in the Room”

However, the inherent complexity of these systems coupled with heterogeneous and ambiguous data, provide several unique challenges when trying to assess the broad range of potential risks posed to groundwater.

Finding Solutions to Challenges

Successful applications of big data and machine learning in other scientific disciplines suggest these approaches are well suited to coping with these types of challenges.

Buzzwords du jour:

What is “Big Data”?

Big data can refer to a couple of things:
1) Big data - big data often refers to in terms of the 5 V’s: Volume, Velocity, Variety, Veracity, and Value
2) Big data computing – computational logic designed to improve the management, performance, storage and data movement of big data

How about “Machine Learning”?

Gives computers the ability to learn without being explicitly programmed; e.g. we should be able to provide computers access to data and let them learn themselves.

What’s the value of Big Data & Machine Learning?

Data is the new oil, but do you have the resource to refine it?
Application of ML & BD at NETL – Evaluating Subsurface Fluid & Gas Migration Risk

Prior research highlighted the need for a way to identify areas with an increased likelihood for fluid and gas migration to provide additional insights to better support a range of decision making needs.

Any framework or model designed to support this analysis would need to work with a big volume, variety, and veracity of data in space & time.

Prior research highlighted the need for a way to identify areas with an increased likelihood for fluid and gas migration to provide additional insights to better support a range of decision making needs.

Designing the Framework

Wellbores:
- Where are they?
- What condition are they in?
- What are they proximal to?

Natural Pathways:
- Where are they?
- What are they?
- What are they proximal to?

Due to the nature of our problem set & available data, we needed an approach that is flexible, works with imprecise data, and accounts for expert knowledge.

Embracing Subsurface Fuzziness

NETL’s Spatially Integrated Multi-variate Probabilistic Assessment (SIMPA) model is a data-driven framework that couples open data with big data processing and fuzzy logic.

Fuzzy Logic (FL) is a multi-valued logic system.

Fuzzy logic was selected because:
- easy to understand (based on natural language),
- flexible,
- handles highly complex, real world data and uncertainty,
- works with numerical and categorical data inputs,
- can be built on top of the experience of experts,
- can readily couple with traditional statistical, spatio-temporal statistical, and machine learning techniques.

Building a Fuzzy Logic Model

Once key data variables were identified, the challenge was building ‘baseline’ fuzzy rules for SIMPA that would be applicable for analyzing likelihood of fluid &/or gas migration at different locations and scales.

Bringing the Inputs Together in Fuzzy Space:

Wellbore Pathways

IF

Completion Date

Abandonment Date

THEN Relative Wellbore Risk

Bringing the Inputs Together in Fuzzy Space:

Natural Pathways

IF

Magnetic 1 AND Magnetic 2 OR Magnetic 3 OR Gravity

THEN Subsurface Complexity
SIMPA Status

- Identified 'baseline' inputs and fuzzy rules
- Developing an open-source UI
- Release SIMPA beta version via EDX at end of year

For more information on NETL's Carbon Storage portfolio, data, and tools visit: https://edx.netl.doe.gov/carbonstorage

Acknowledgment: This technical effort was performed in support of the National Energy Technology Laboratory's ongoing research under the RES contract DE-FE0004000.

Disclaimer: This project was funded by the Department of Energy, National Energy Technology Laboratory, an agency of the United States Government, through a support contract with AECOM. Neither the United States Government nor any agency thereof, nor any of their employees, nor AECOM, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Jennifer Bauer
jennifer.bauer@netl.doe.gov
NETL Geospatial Researcher

Questions?