Subsurface Trend Analysis

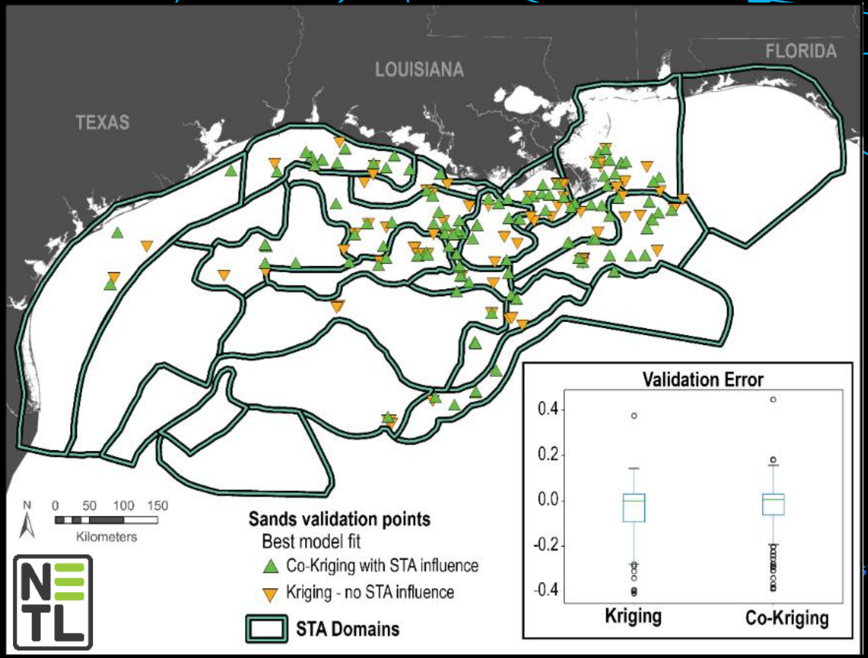
A methodical framework for artificial intelligence subsurface property prediction

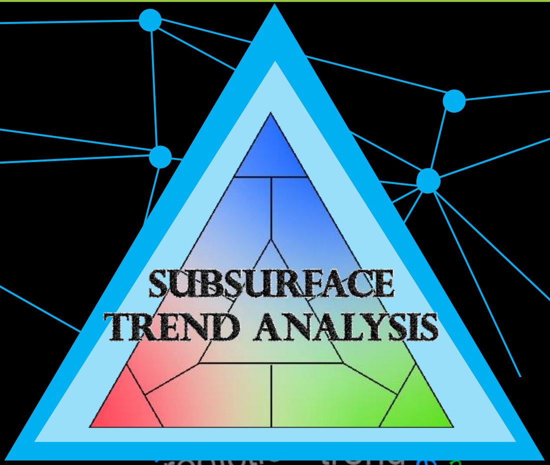
MacKenzie Mark-Moser[1,2], Kelly Rose[1], Jennifer Bauer[1], Patrick Wingo[1,2], Anuj Suhag[1,3]

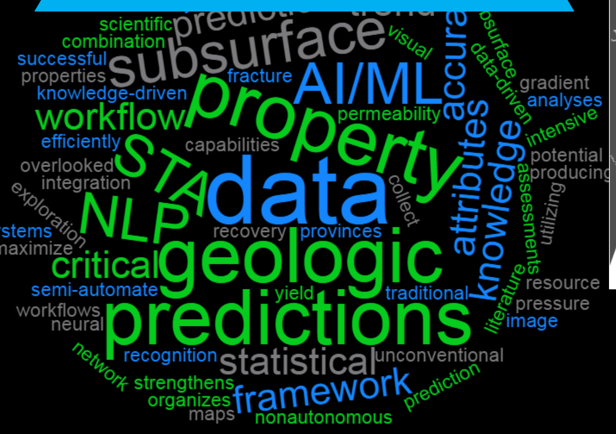
[1]National Energy Technology Laboratory [2]LRST, [3]ORISE Fellow

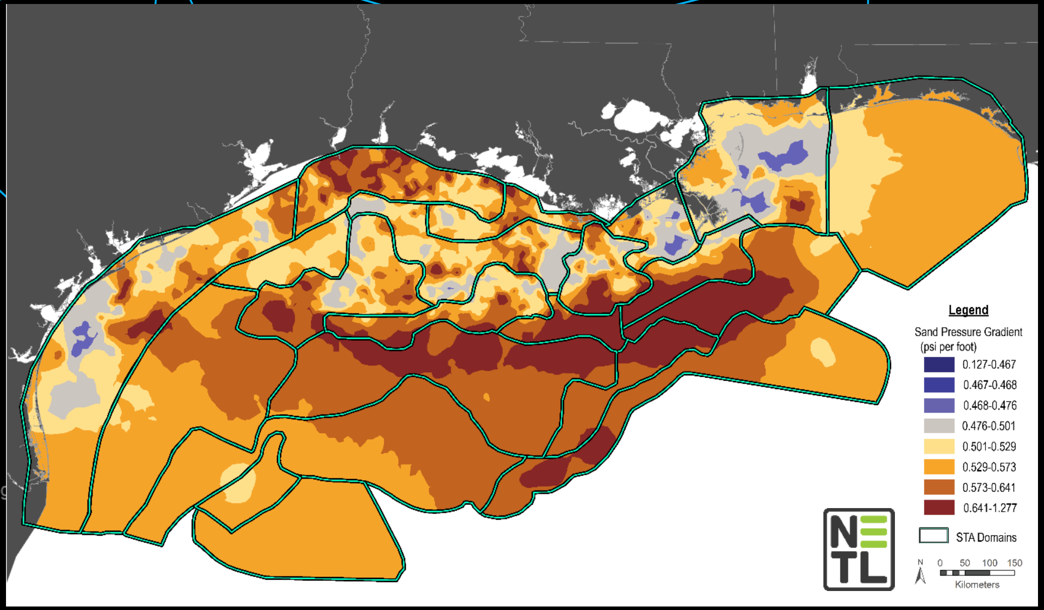
NATIONAL ENERGY TECHNOLOGY LABORATORY

AGU Fall Meeting, Dec. 11th 2019, Moscone Center, San Francisco, CA







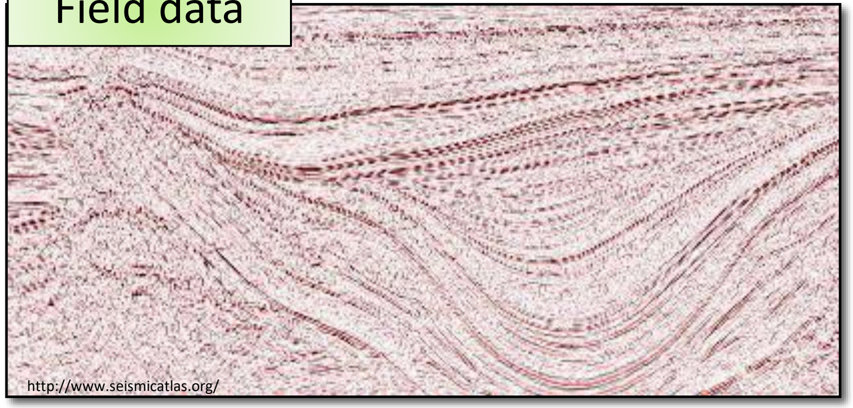


Solutions for Today | Options for Tomorrow

U.S. DEPARTMENT OF ENERGY

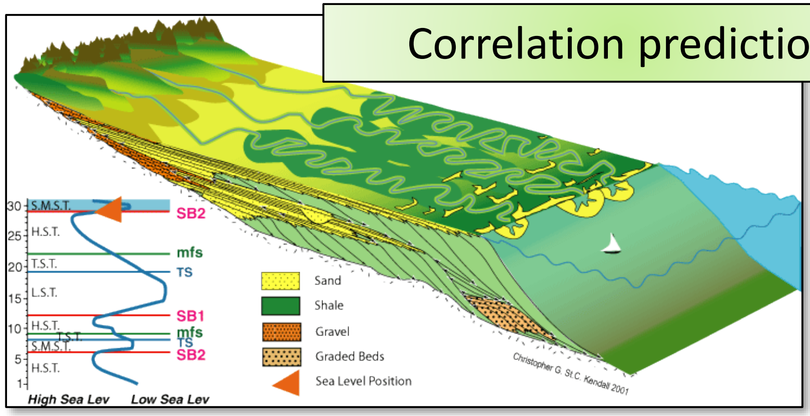
Classic Subsurface Interpretation Approach

Field data

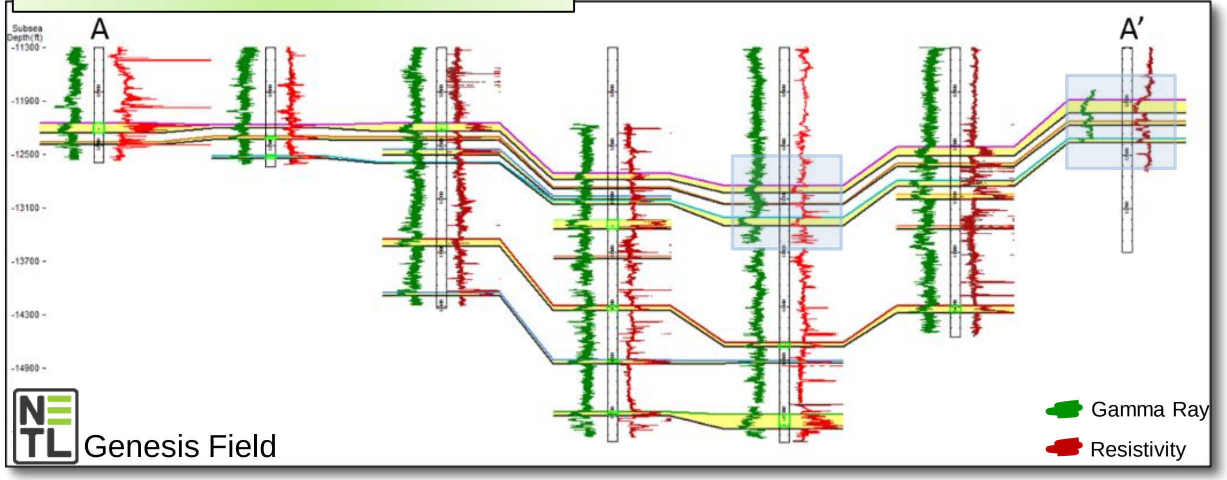


- Driven by field-based observations, interpretations and measurements
- Integrated with contextual geologic systems information *after the fact*
- Often indirect– only direct data are outcrop and core!

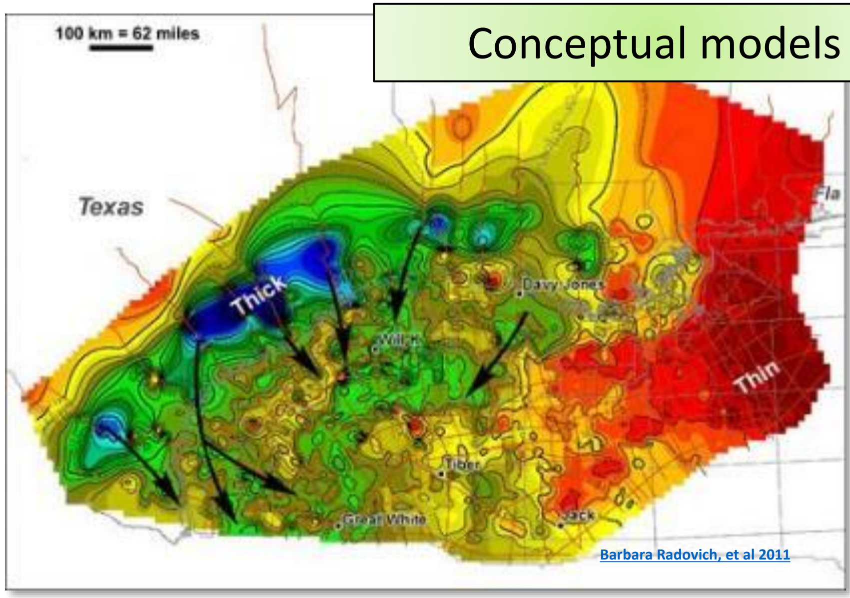
Correlation prediction



Indirect measurements



Conceptual models



Addressing Gaps in Subsurface Prediction

- Subsurface is a **largely unconstrained**
- Highly **heterogeneous**
- Prediction relies on **indirect** data and methods
- Quality and quantity of data is highly **variable**
- Existing interpolation algorithms overlook **contextual information**
- Need to improve prediction for areas of **little or no data**

The STA method seeks to train data and statistical methods to "think" like a geologist

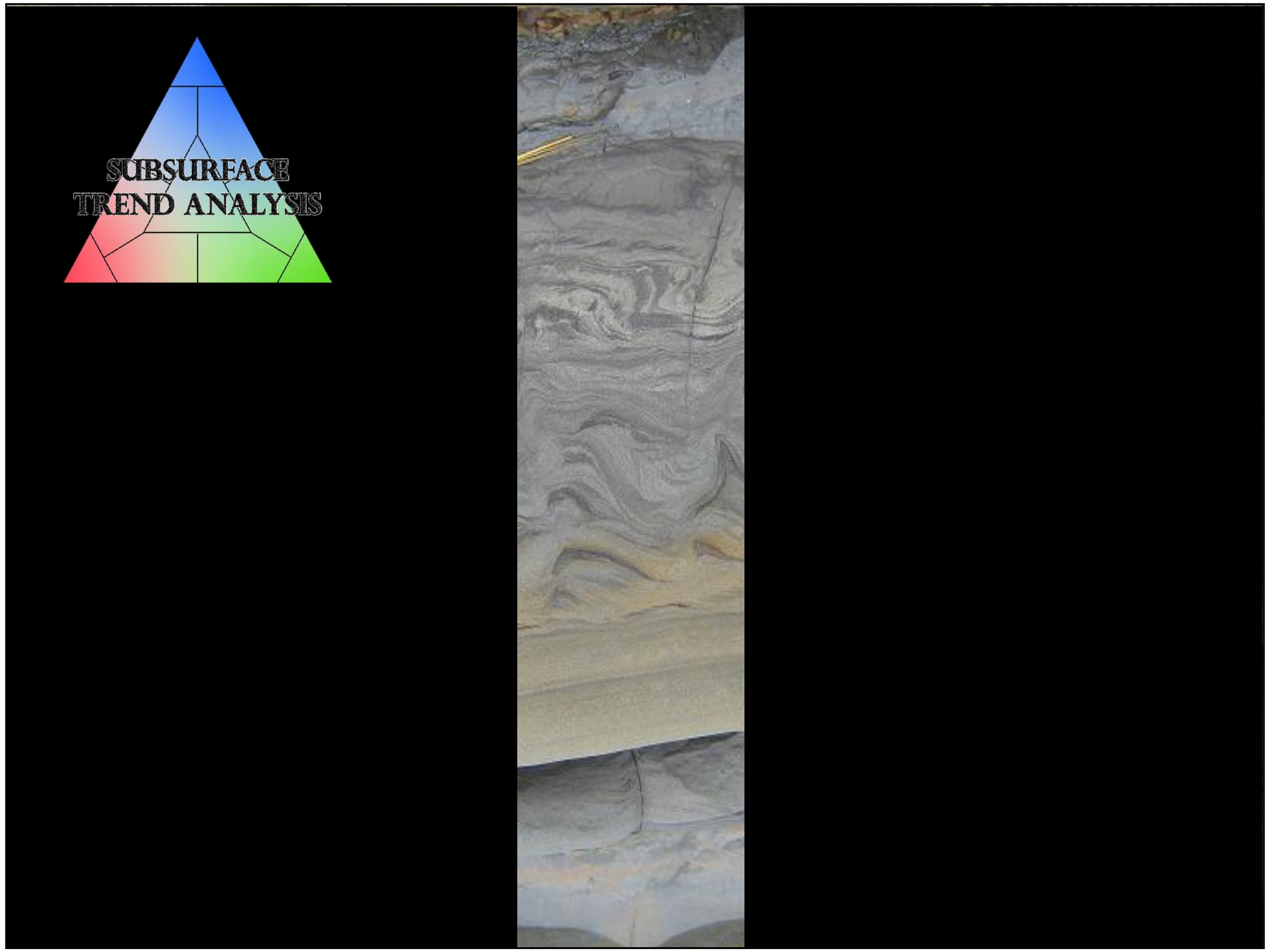
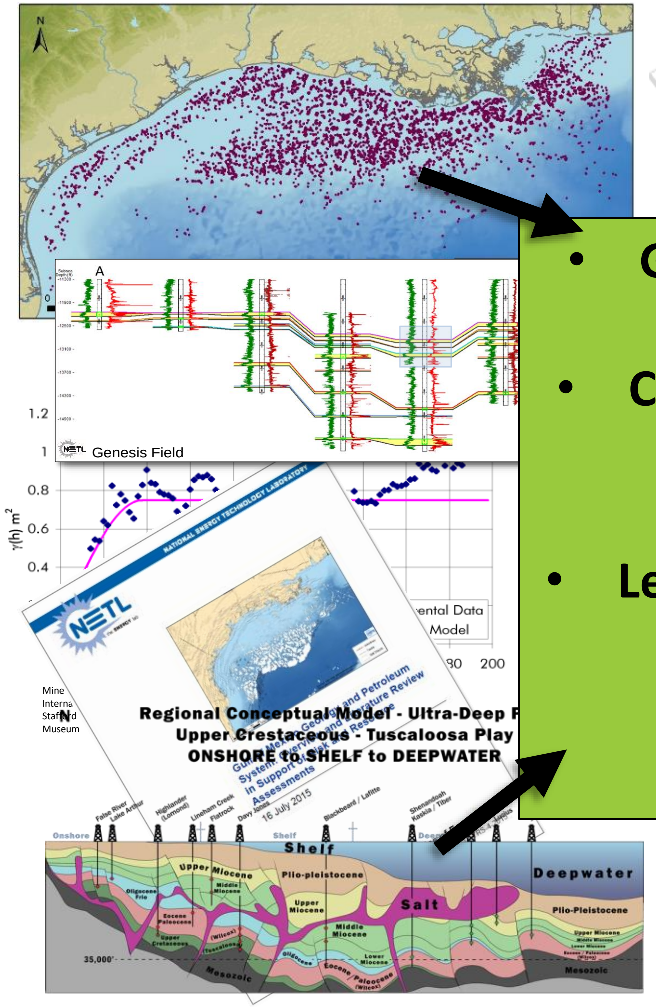
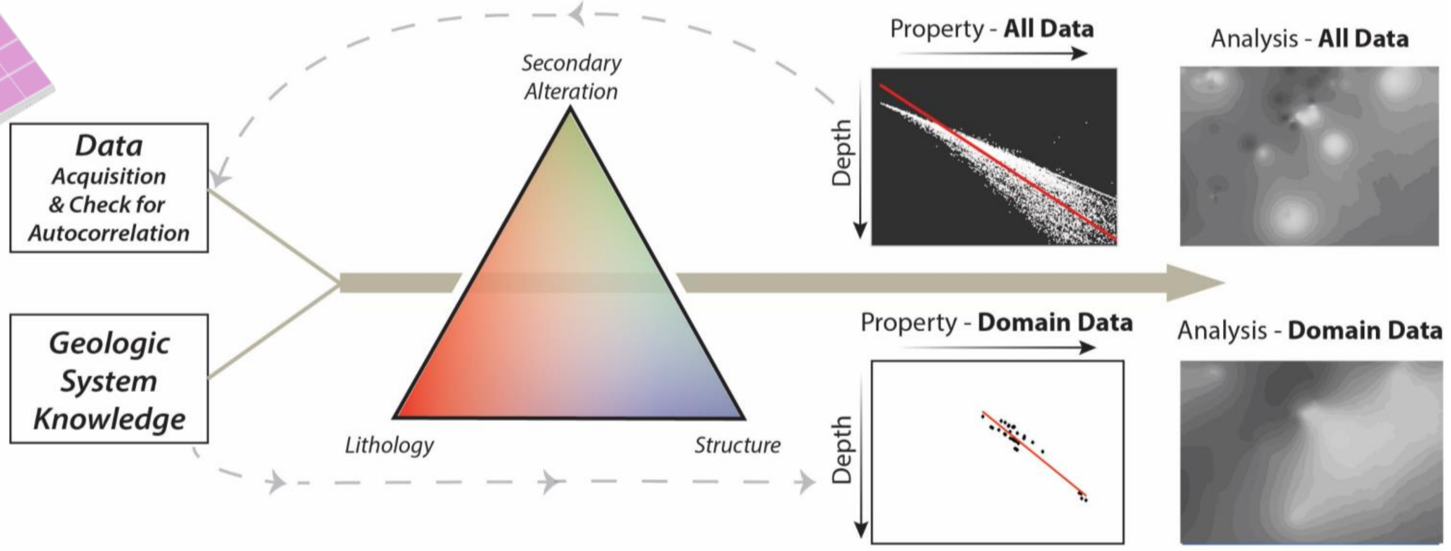


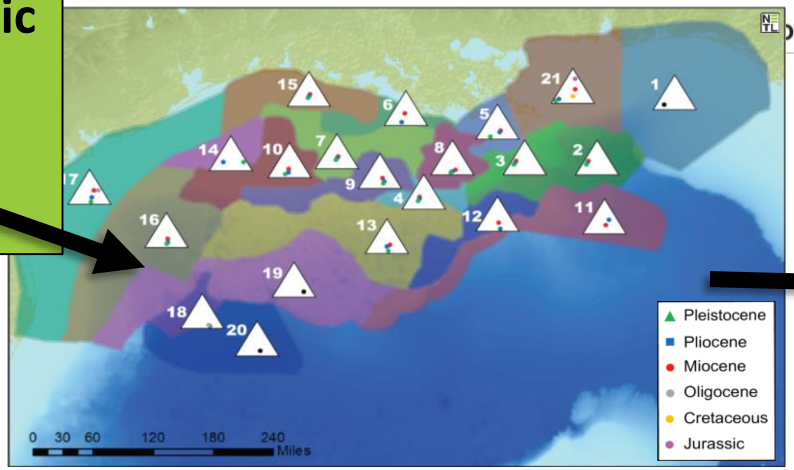
Photo Credit: Kelly Rose

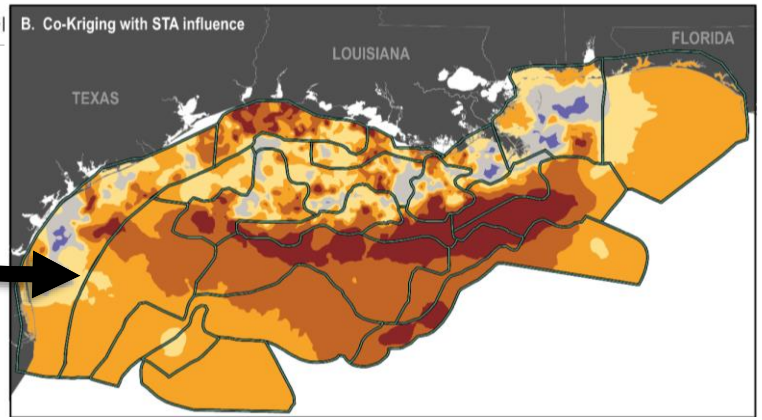
Science-driven workflow for improved subsurface prediction: the Subsurface Trend Analysis Method



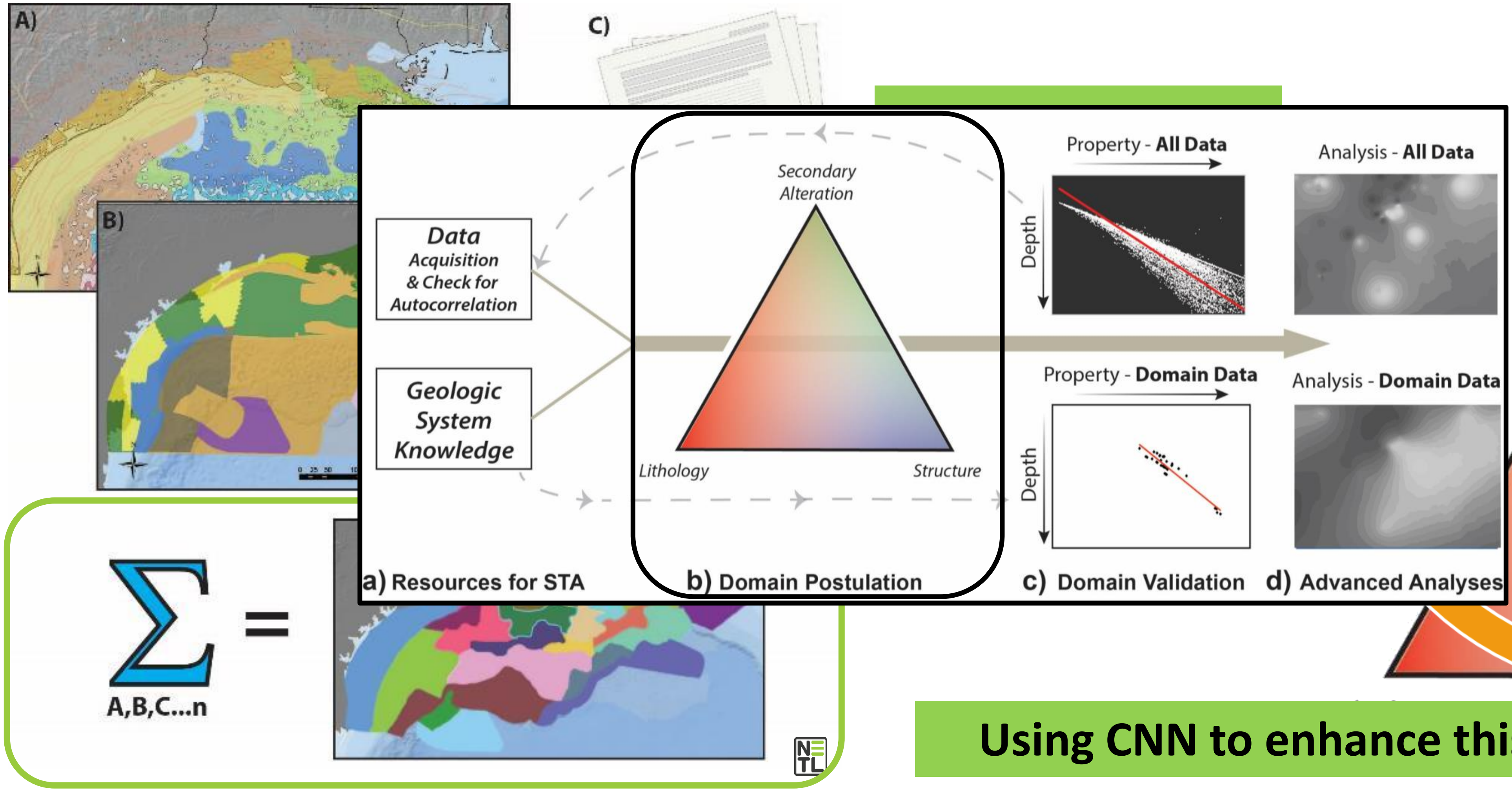
- Gather geologic knowledge
- Characterize and assess data resources
- Leverage geologic advantage to improve predictions

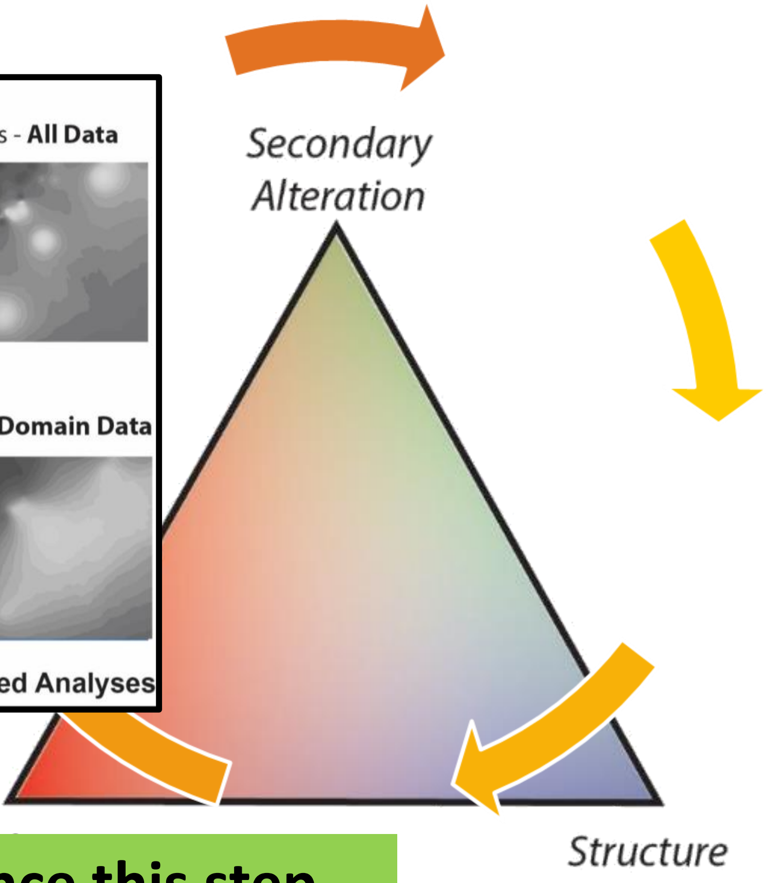






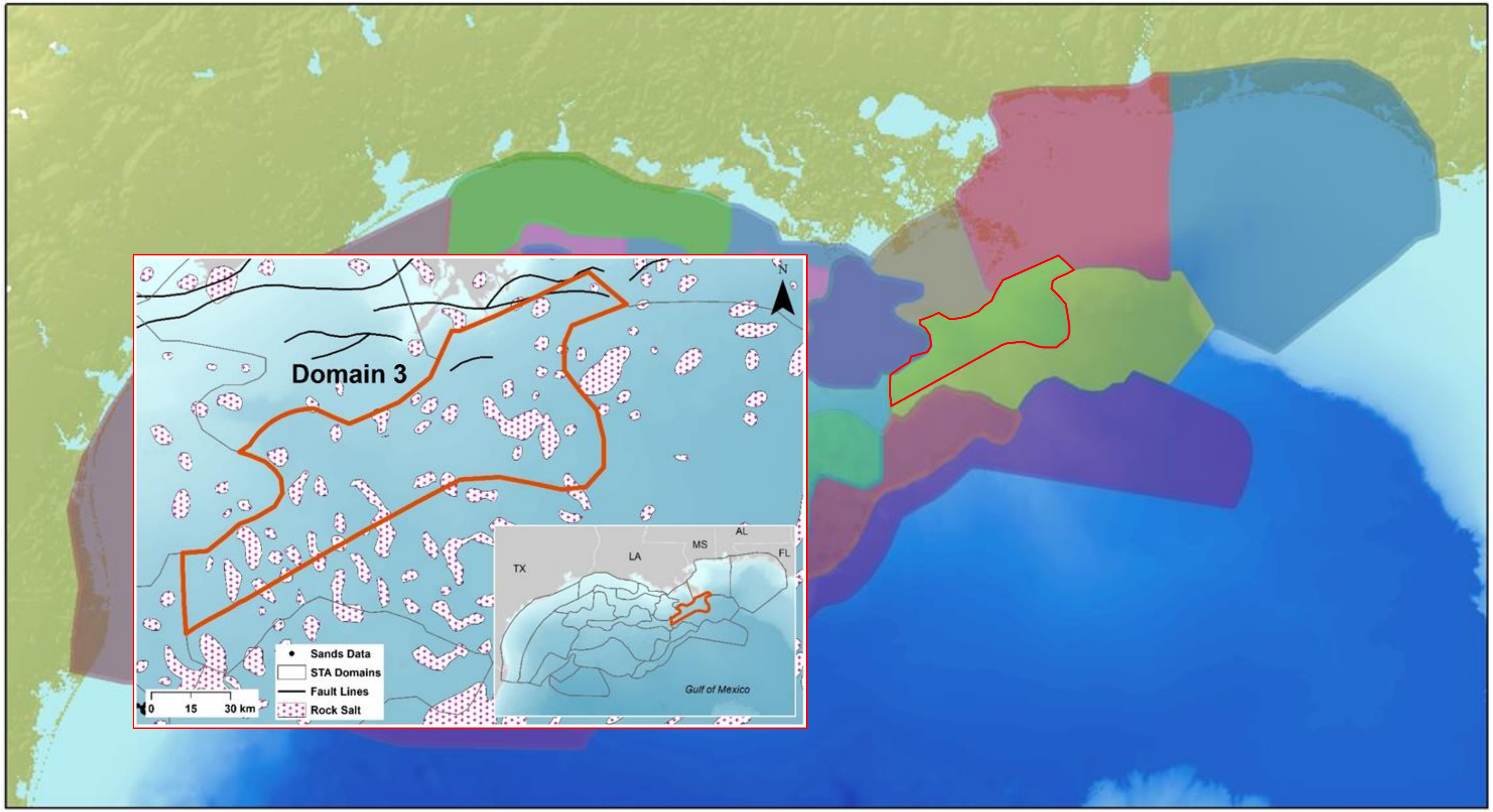
Defining areas with a common history



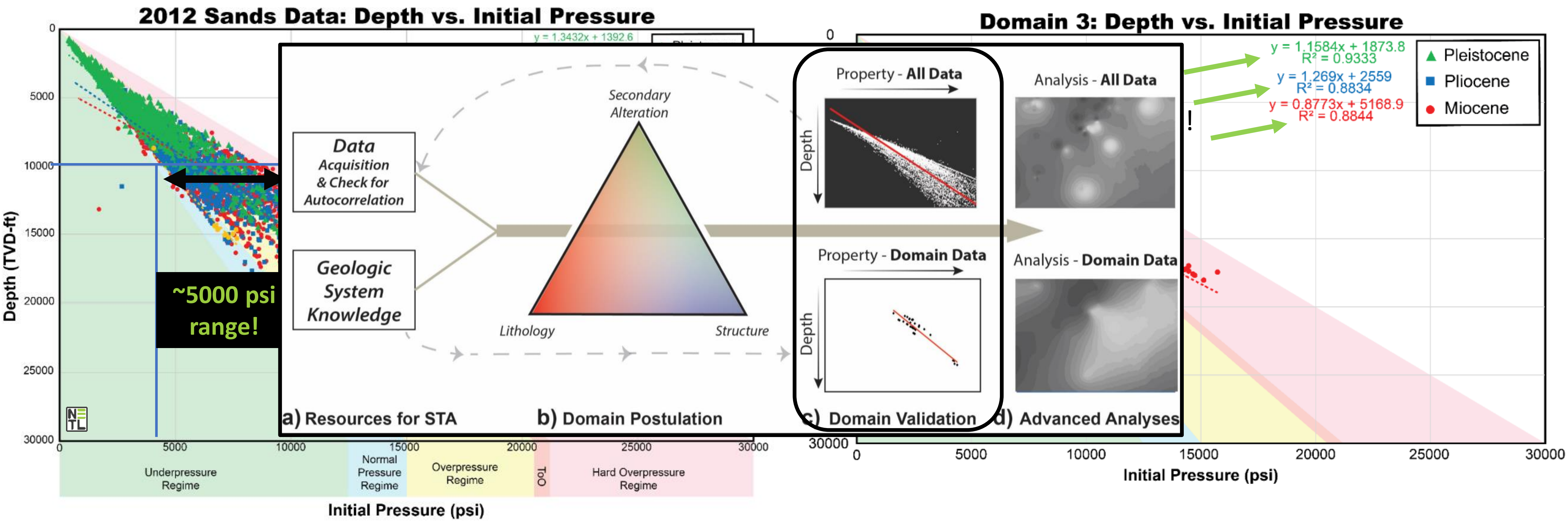


Using CNN to enhance this step

Result: Domains bounding common geology



Contrast against the full suite of data



Use ML methods to mine knowledge resources and autodefine domain clusters

Sand Pressure Gradient prediction in the Offshore GOM

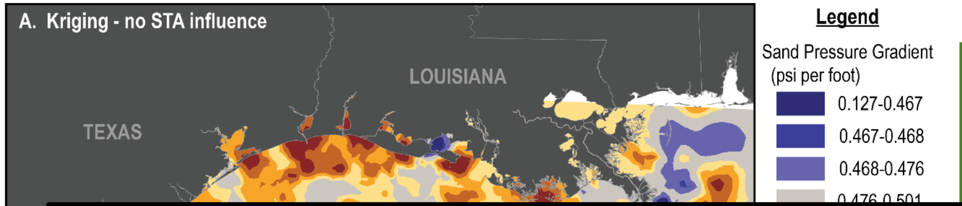
Mark-Moser, M., Miller, R., Bauer, J., Rose, K., and C. Disenhof. 2018, *Analysis of Subsurface Reservoir Properties Using a Novel Geospatial Approach, Offshore Gulf of Mexico*. NETL-TRS-2018

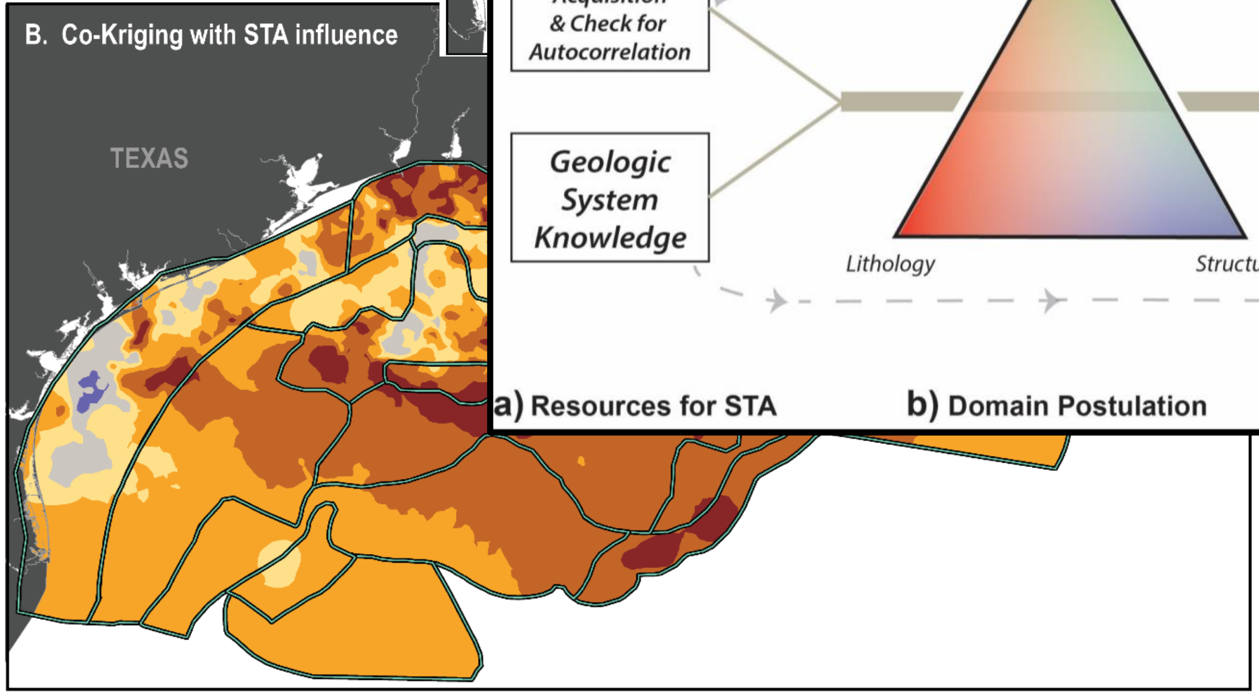
Rose, K., Bauer, J.R., and Mark-Moser, M. **(Expected Feb. 2020)**. Subsurface trend analysis, a multi-variate geospatial approach for subsurface evaluation and uncertainty reduction, *Interpretation*

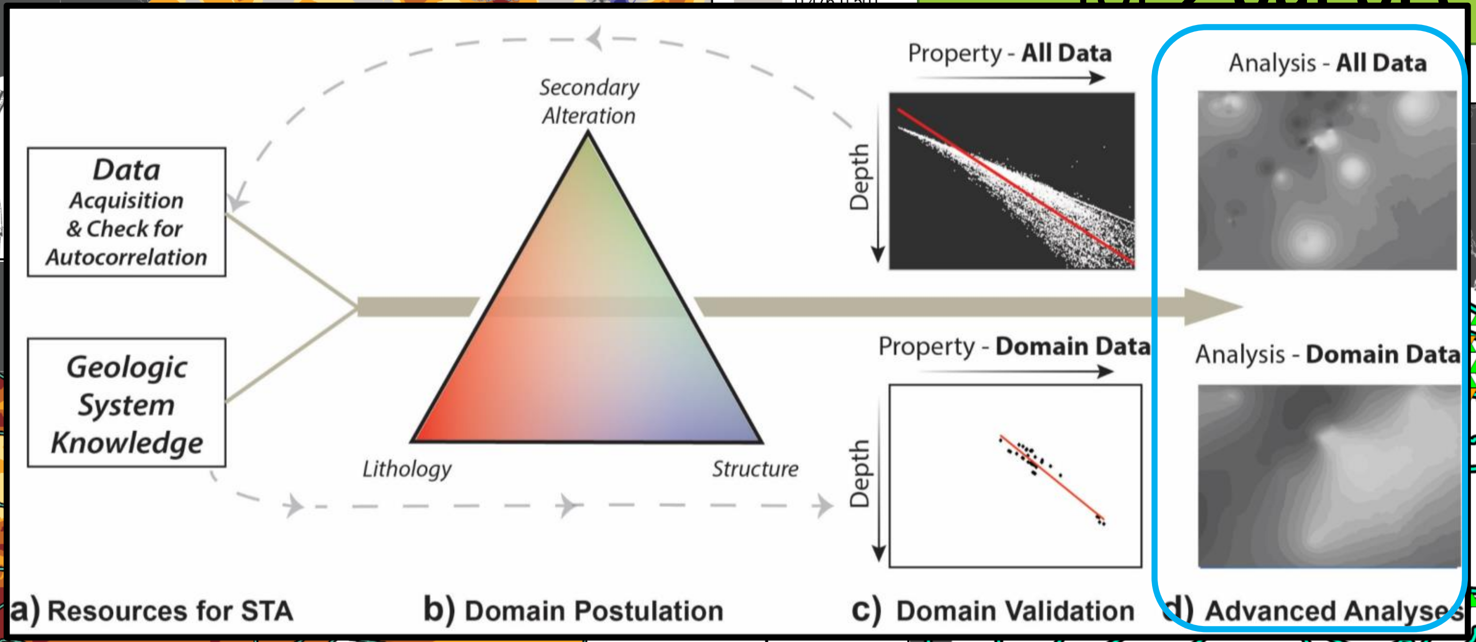
NATIONAL ENERGY TECHNOLOGY LABORATORY

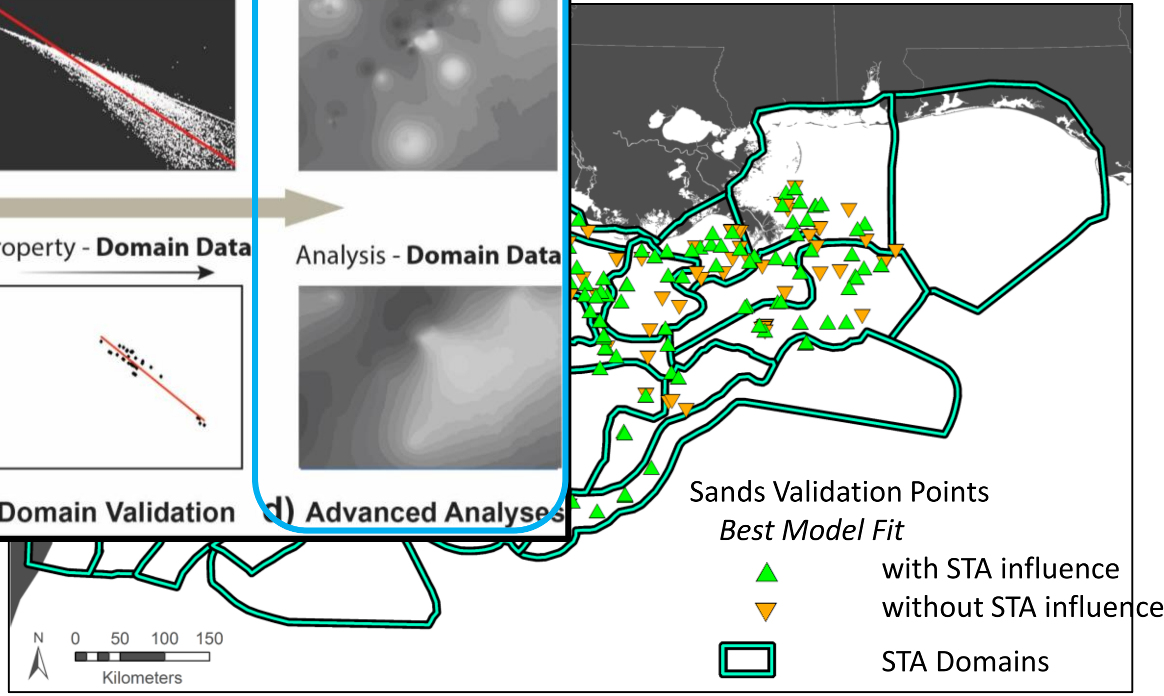
- Defined 21 STA geologic domains
- STA Predictions utilizing >13,000 data points

STA provided a better prediction for **2 out of every 3** points



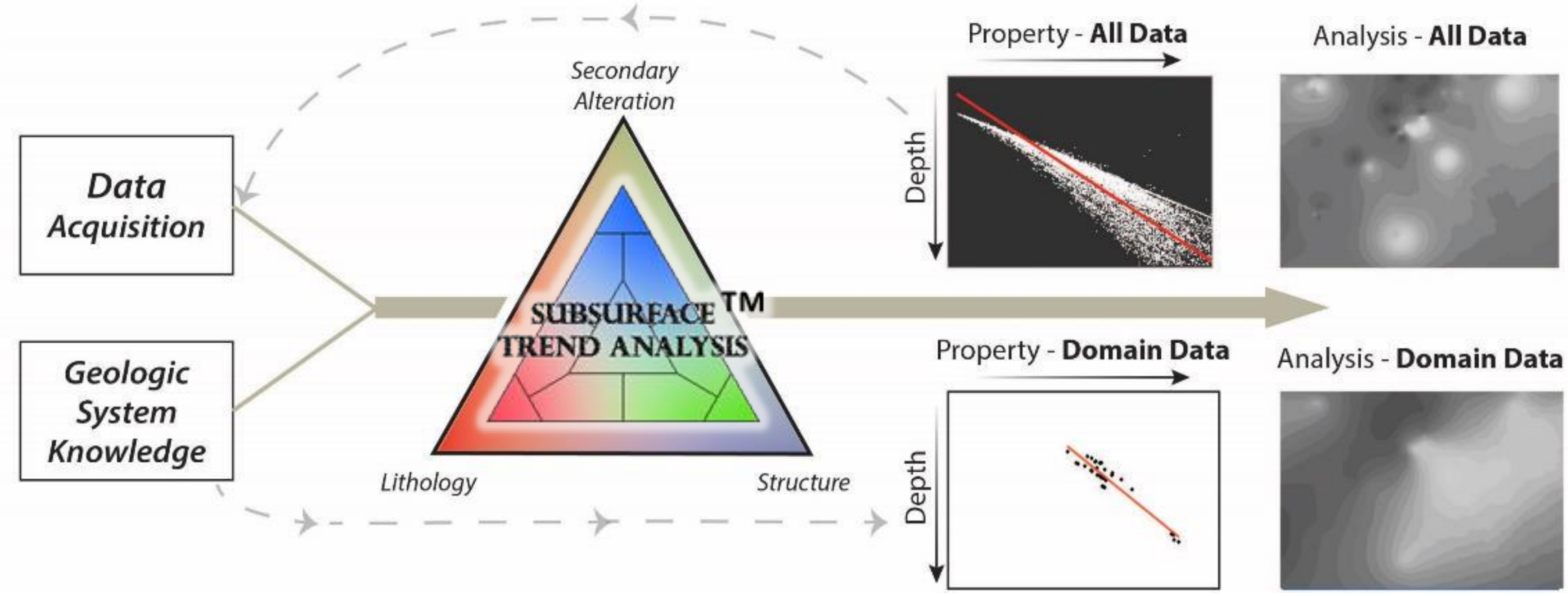






Incorporating ML into the Science-Driven STA Approach

- **Filling the need: real time, adaptive, improved prediction of subsurface properties**
- **A smarter, more efficient way to gain subsurface insights**



- **Train NLP to produce relevant literature & data sources**

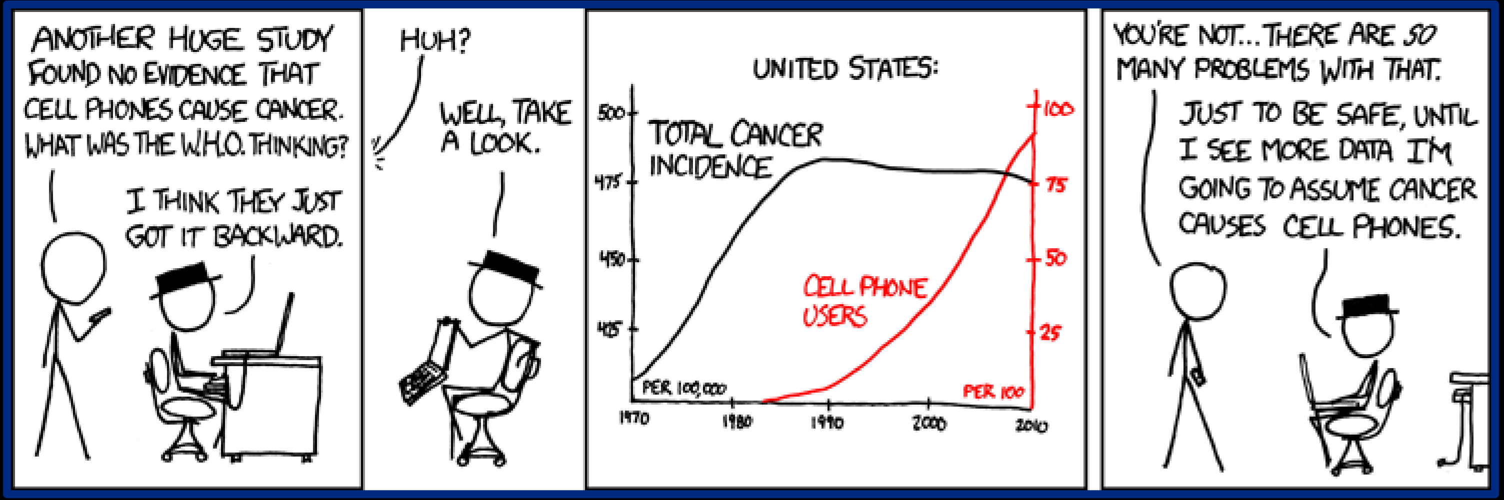
- **Extract geologic contextual information literature**
- **Identify key spatial information (e.g. geologic provinces)**

- **Characterize & analyze data with ML insights**
- **Produce in-depth statistical evaluation**

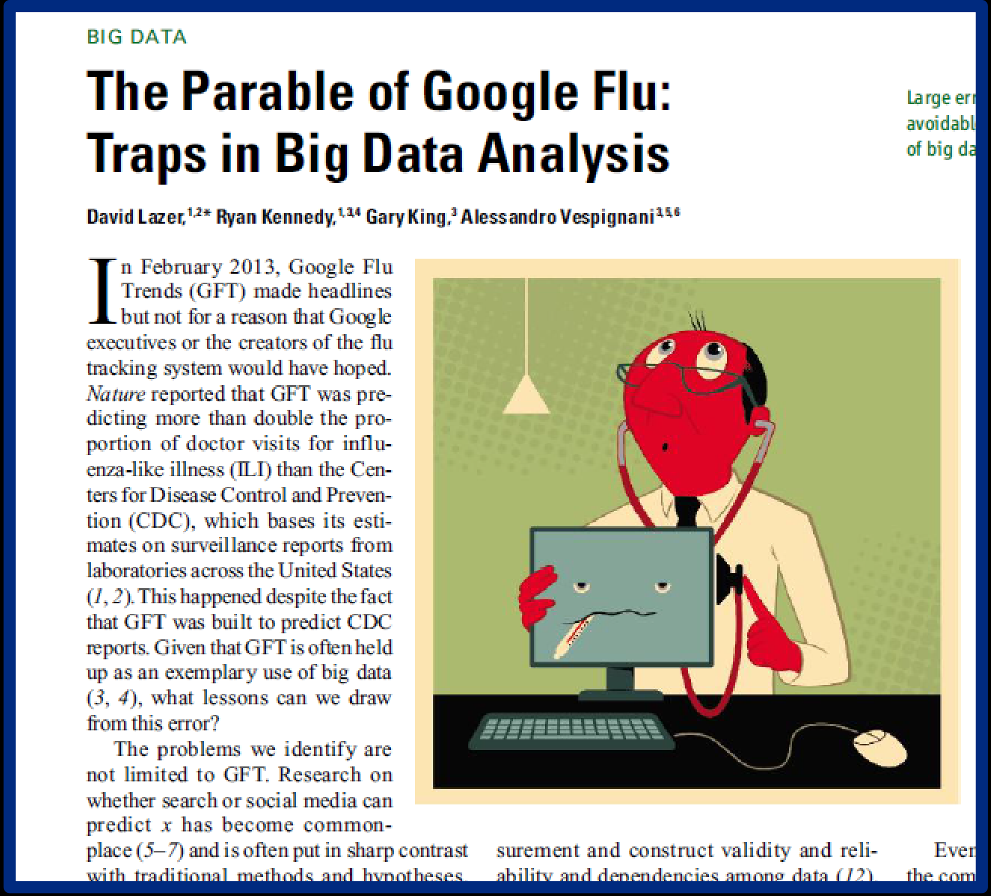
- **Optimize prediction techniques**

Cautions for big data, ML driven analytics

- Correlation does not equal causation
  - Just because you have an analysis doesn't mean the results are meaningful
- Uncertainty is critical
  - Capture, reduce if possible, represent, utilize, quantify
- Data science driven analytics should not be randomly applied, must be guided by conventional science methods



- Analytical methods must be appropriate for the **goal**
- Analytical methods must be appropriate for the **data**
- Analysis must be made in **context** of data collection



BIG DATA

The Parable of Google Flu: Traps in Big Data Analysis

David Lazer,[1,2]* Ryan Kennedy,[1,3,4] Gary King,[3] Alessandro Vespignani[3,5,6]

In February 2013, Google Flu Trends (GFT) made headlines but not for a reason that Google executives or the creators of the flu tracking system would have hoped. *Nature* reported that GFT was predicting more than double the proportion of doctor visits for influenza-like illness (ILI) than the Centers for Disease Control and Prevention (CDC), which bases its estimates on surveillance reports from laboratories across the United States (*1*, *2*). This happened despite the fact that GFT was built to predict CDC reports. Given that GFT is often held up as an exemplary use of big data (*3*, *4*), what lessons can we draw from this error?

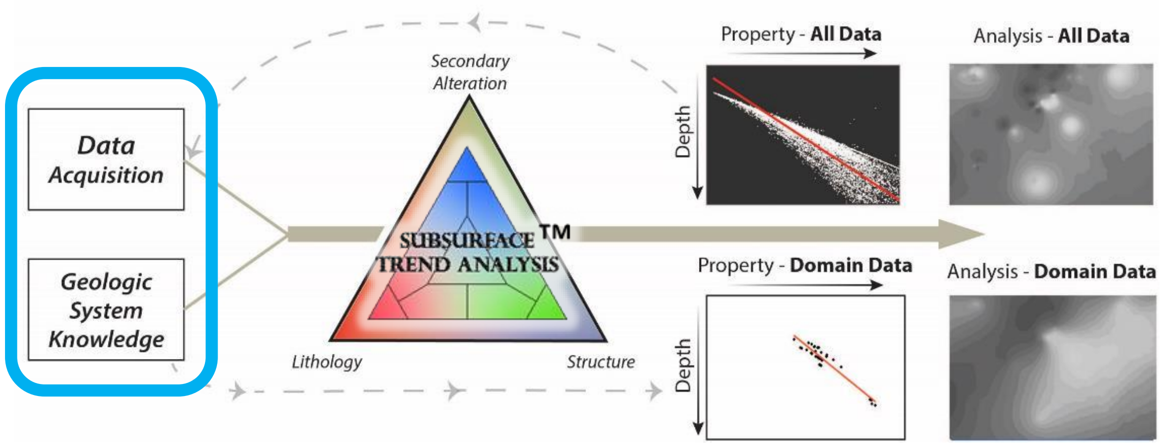
The problems we identify are not limited to GFT. Research on whether search or social media can predict *x* has become commonplace (*5*–*7*) and is often put in sharp contrast

surement and construct validity and reliability and dependencies among data (*12*)

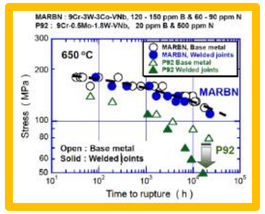
Lazer et al 2014, the parable of google flu: traps in big data analysis, *Science*

NLP & AI for unstructured data & knowledge

Gather, curate & transform

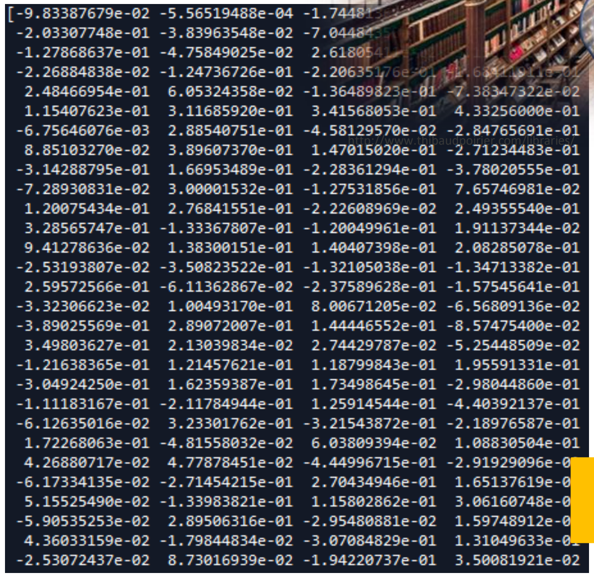


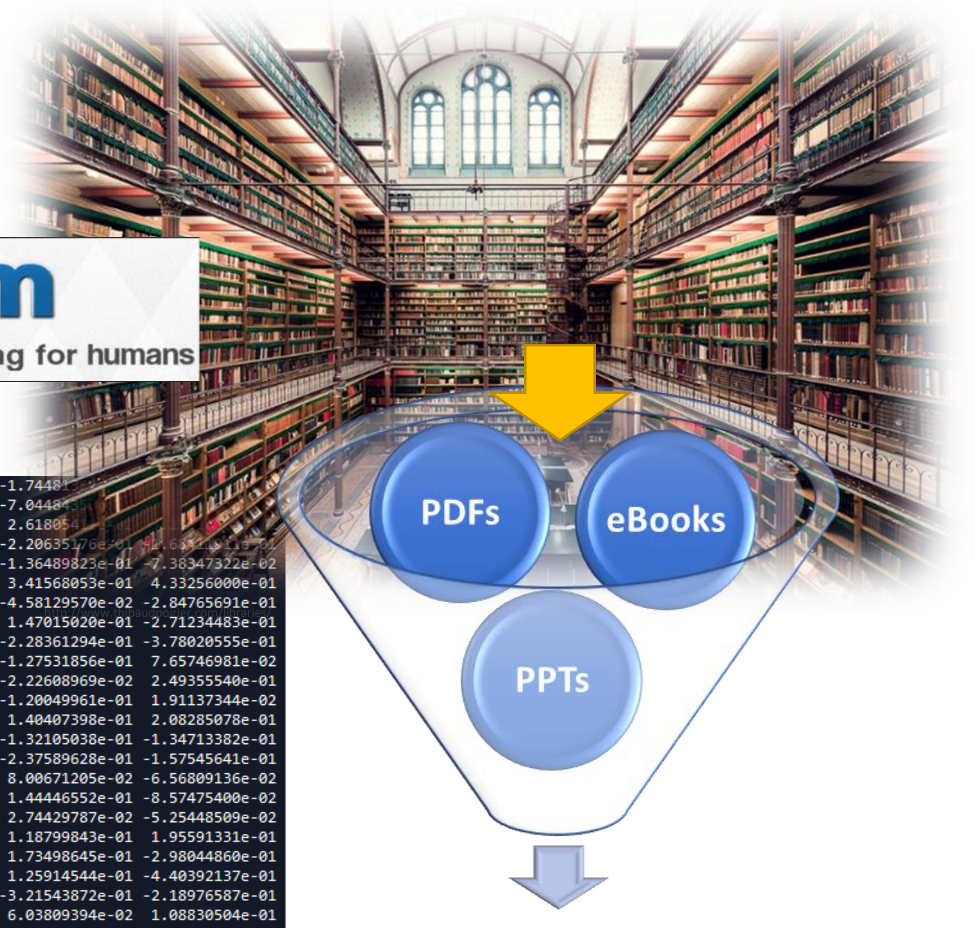
Mining/rescuing old data from documents, R&D products, presentations, etc. with the advantage of NLP and Neural Networks



Apache Tika

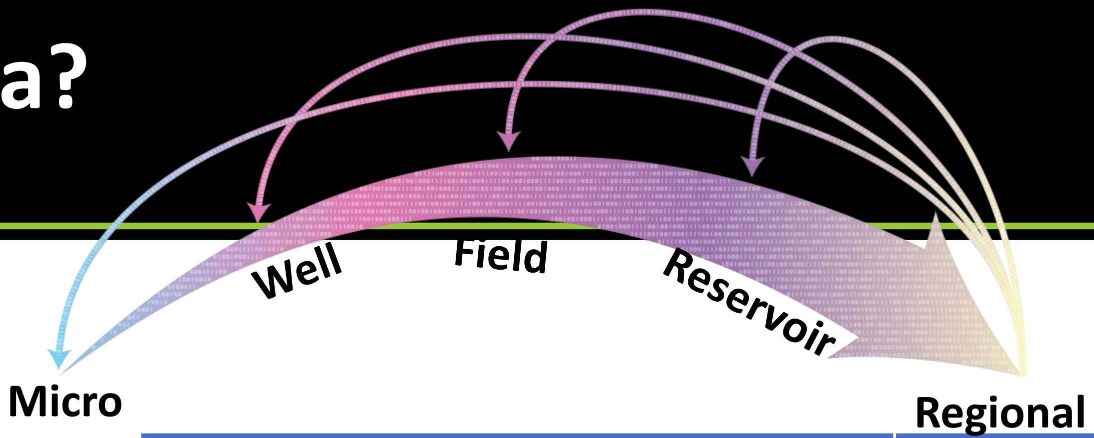
gensim
topic modelling for humans

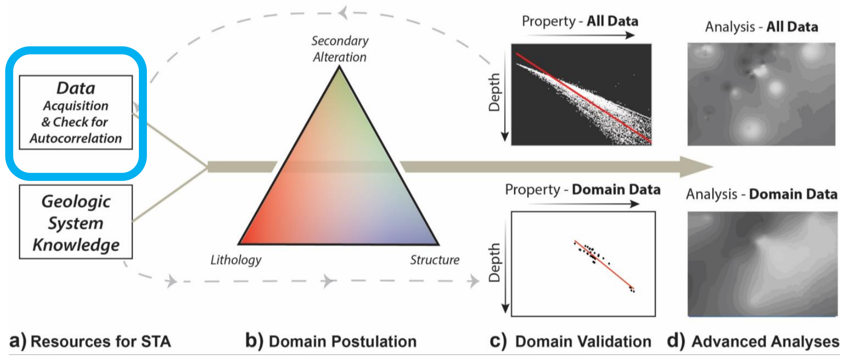




Digital structured data

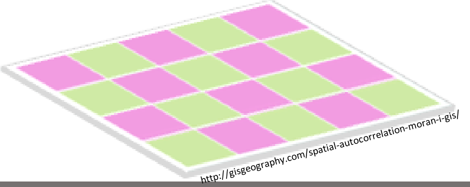
What is our target data?

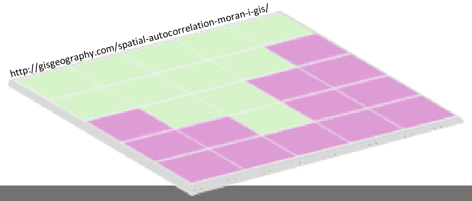




Target data for the STA method is:

- *Driven by geologic processes*
- *Non-random*
- *Autocorrelative (negative or positive)*
- *Single or multi-scale*

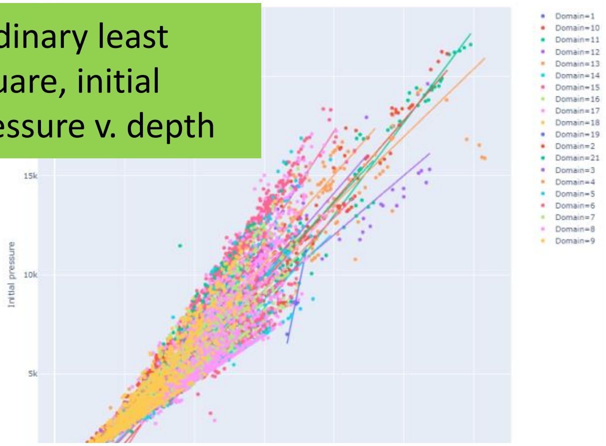




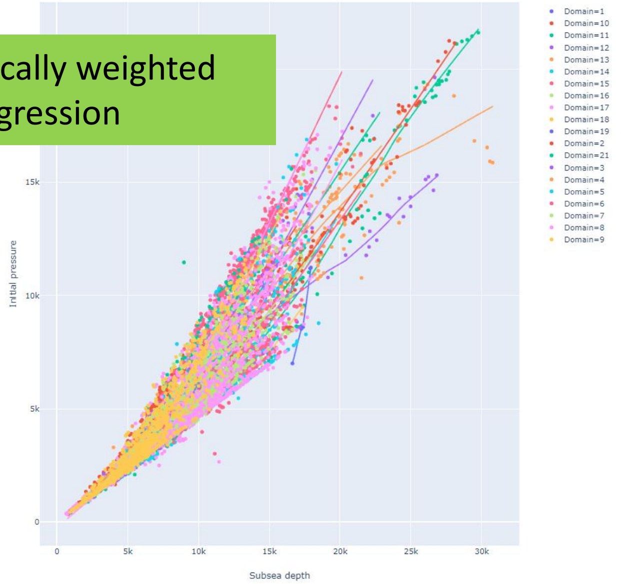
Autocorrelative property	Scale (finer→coarser)
Lithologic thickness	Reservoir, field, region, basin
Lithologic composition	Reservoir, field, region, basin
Porosity	Well, reservoir, field, region, basin
Reservoir pressure	Reservoir, field
In situ pressure	Well, reservoir, field, region, basin
Reservoir temperature	Reservoir, field
In situ temperature	Well, reservoir, field, region, basin
Permeability	Reservoir, field, region, basin
Natural fractures	Reservoir, field
Secondary alteration (e.g., diagenesis, mineralization)	Reservoir, field, region, basin
Volume of shale/clay (Vsh)	Reservoir, field, region, basin

Dimensional analyses of subsurface property data

Ordinary least square, initial pressure v. depth

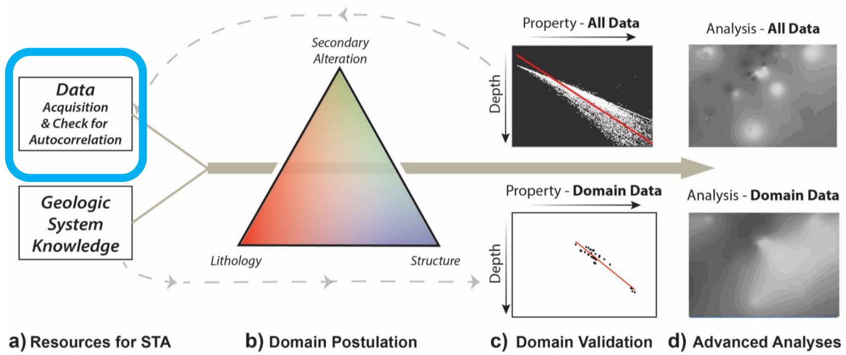


Locally weighted regression

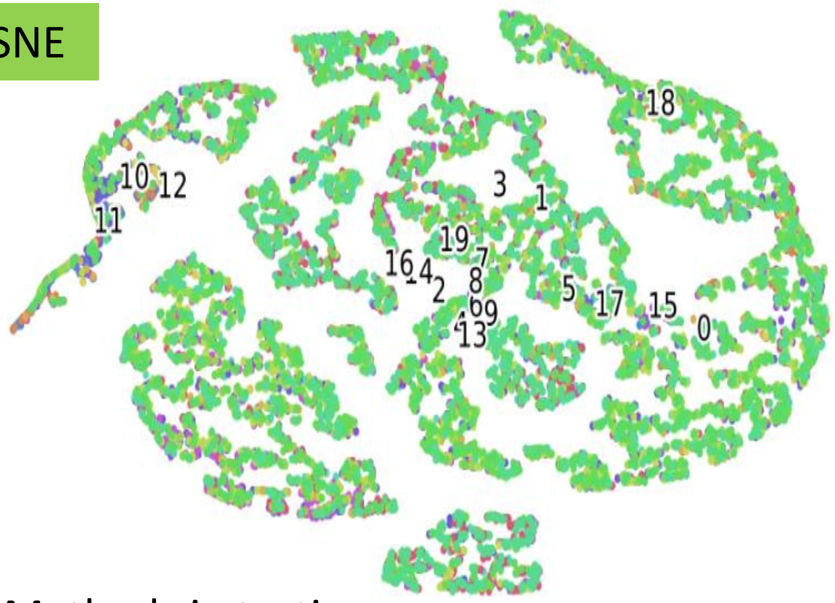


Keys to meaningful ML analysis:

- **Evaluate for logical continuity & efficiency**
- **Clustering must reflect geologic environment**
- **Ground statistical analyses in geologic reality**



tSNE

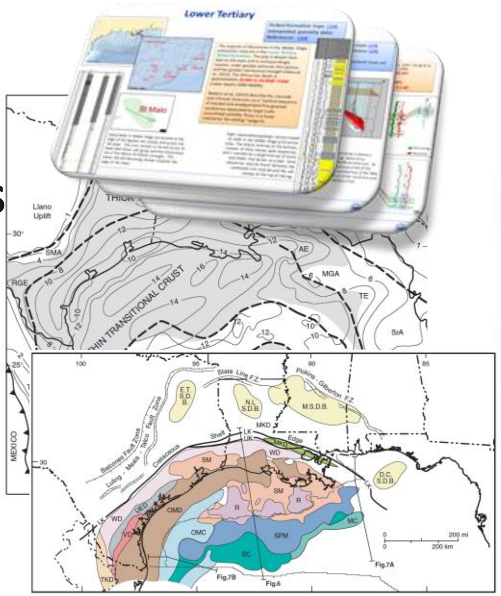


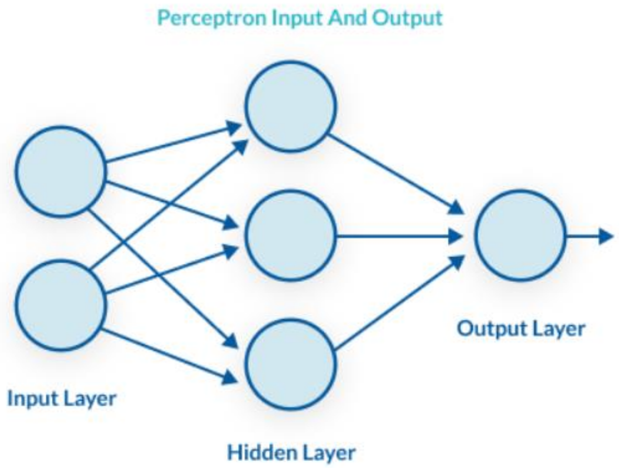
Methods in testing

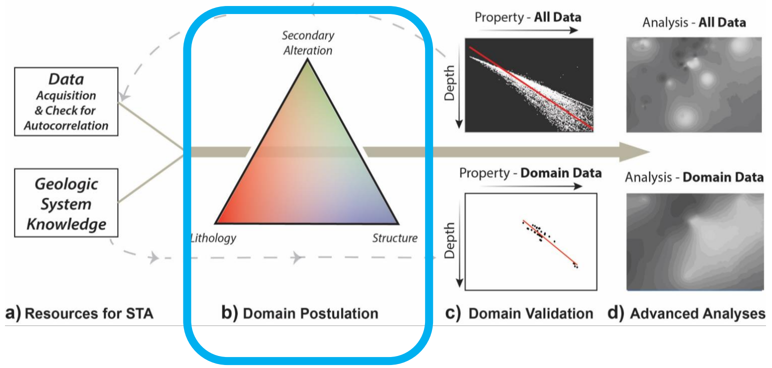
- kMeans
- Principal Component Analysis (PCA)
- tSNE
- Dbscan
- U-Maps
- Topological Data Analysis (TDA)

Neural network & NLP image analysis

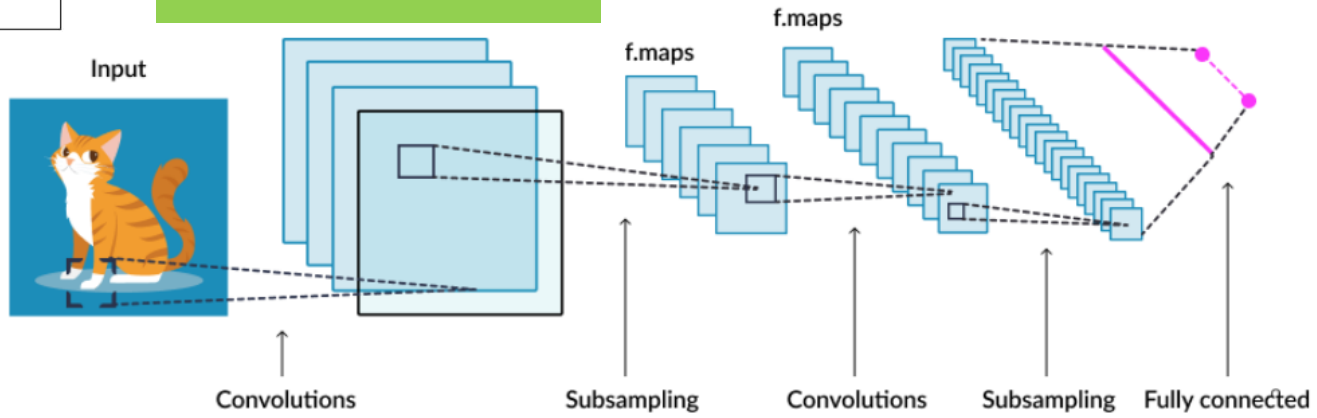
- **Natural language processing to identify content from figure captions**
- **Training convolutional neural network for image identification of spatial geologic information: geologic maps, provinces**







TensorFlow CNN

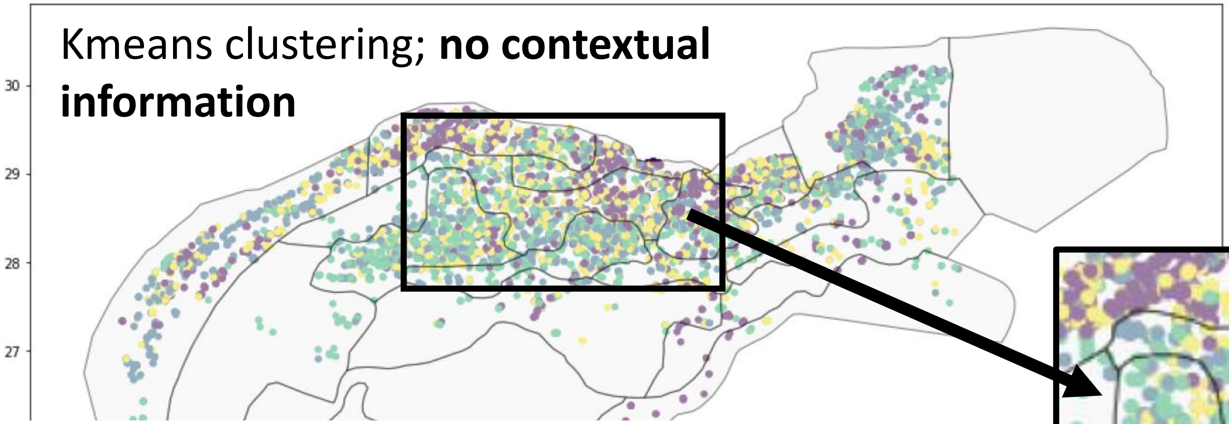


Source: MissingLink.ai

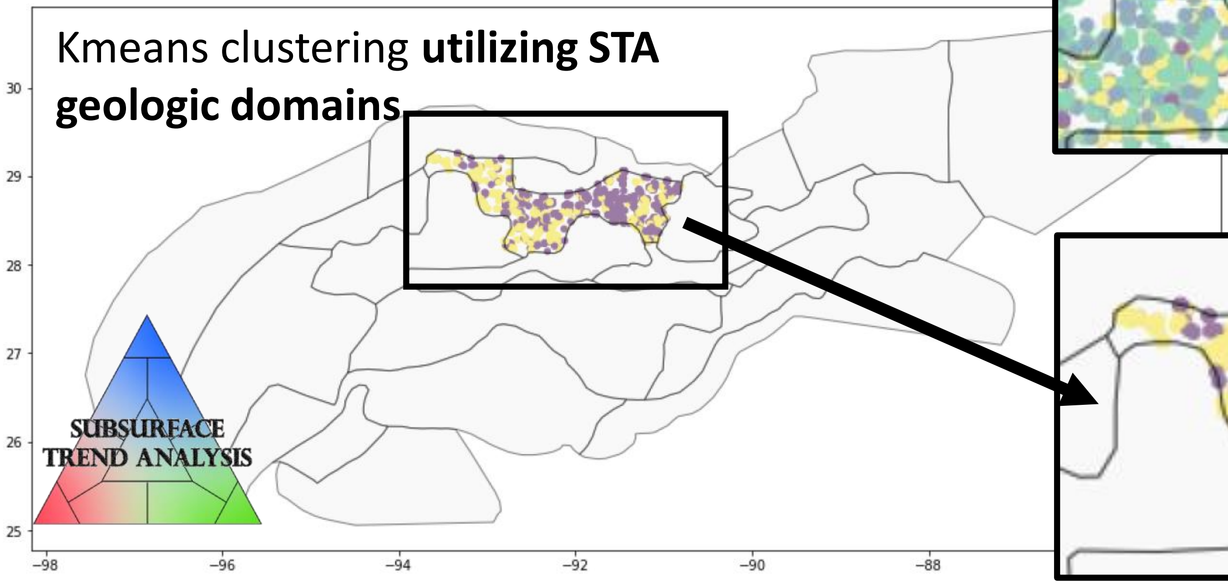
Domain Validation & Universal Clustering Analysis

Gulf of Mexico application

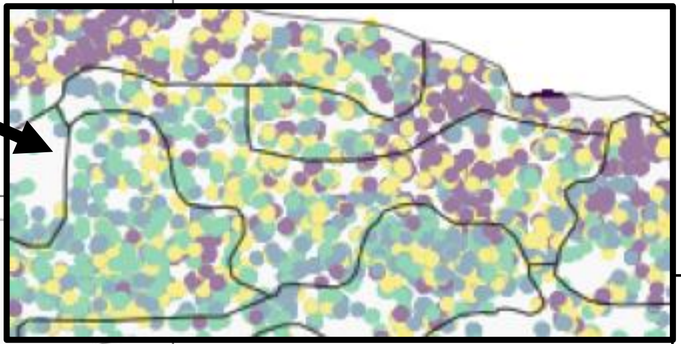
Kmeans clustering; **no contextual information**



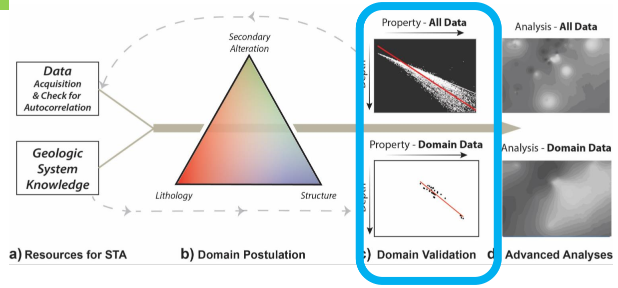
Kmeans clustering **utilizing STA geologic domains**



Kmeans dimensional analysis of 6 subsurface properties



Results are repeatable when utilizing Dbscan!



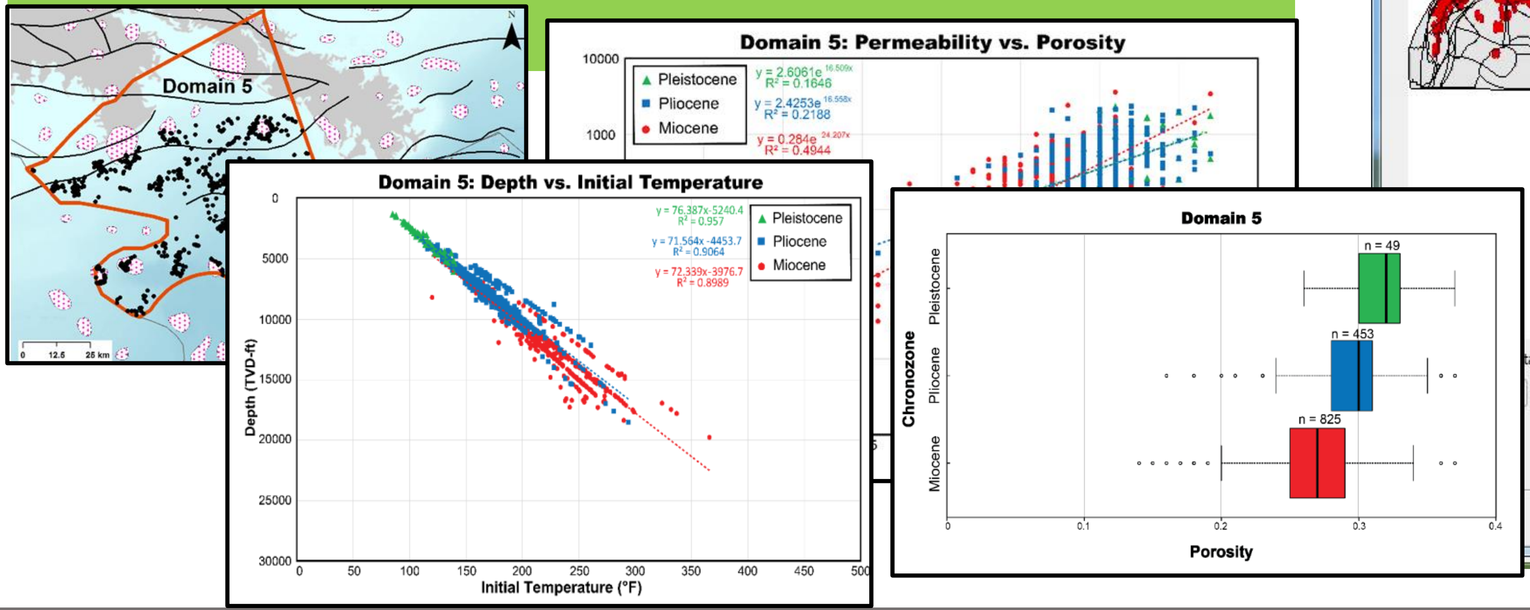
- 4 clusters
- Poor continuity among clusters

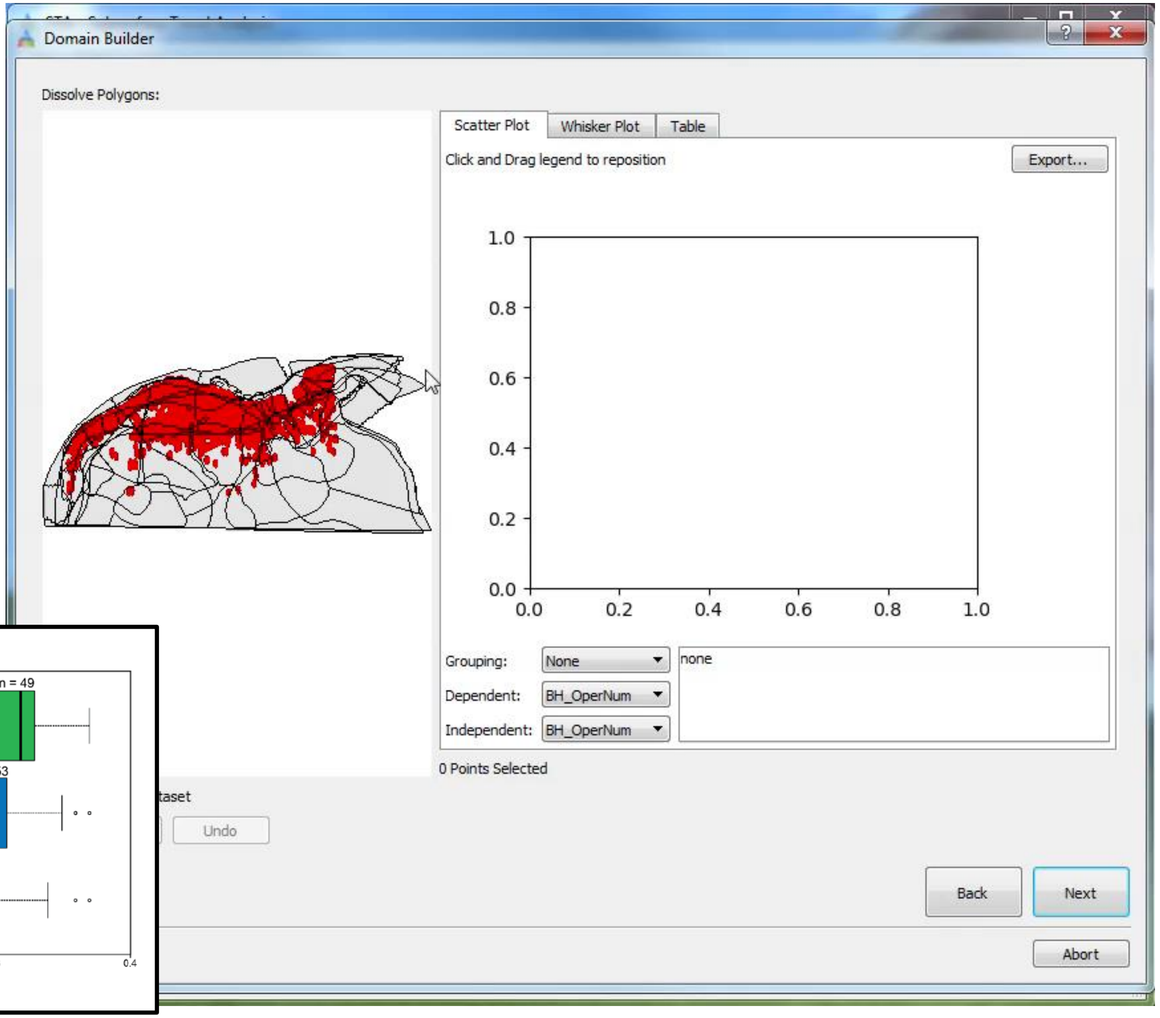
- 2 clusters
- Improved continuity among clusters

Ongoing STA Tool development

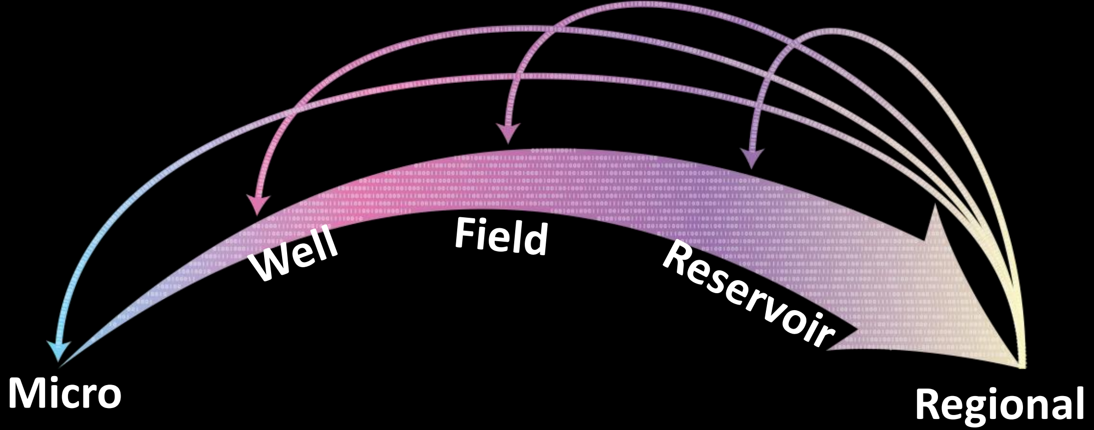
STA Tool– a *virtual research assistant* designed to

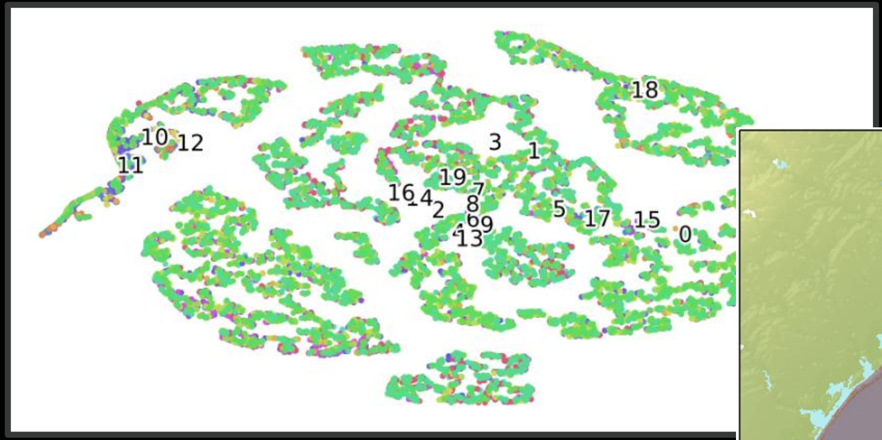
- Organize and visualize disparate big data and knowledge resources
- Simplify and automate geologic domain formation
- Provide and execute statistical analyses and validation
- Utilize machine learning to characterize property trends and predictions



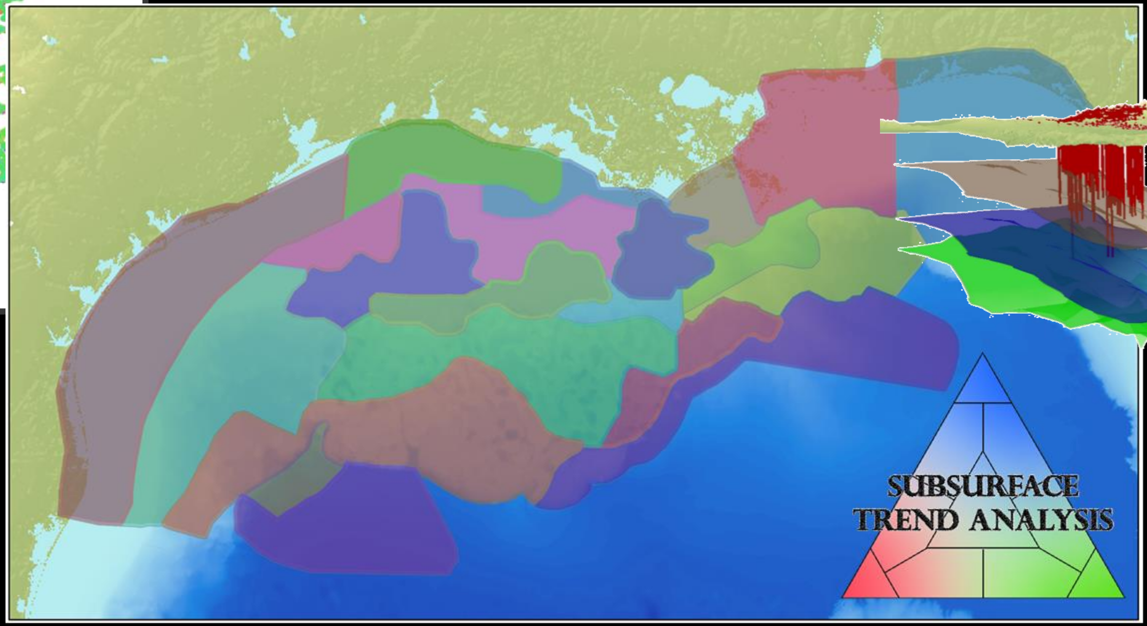


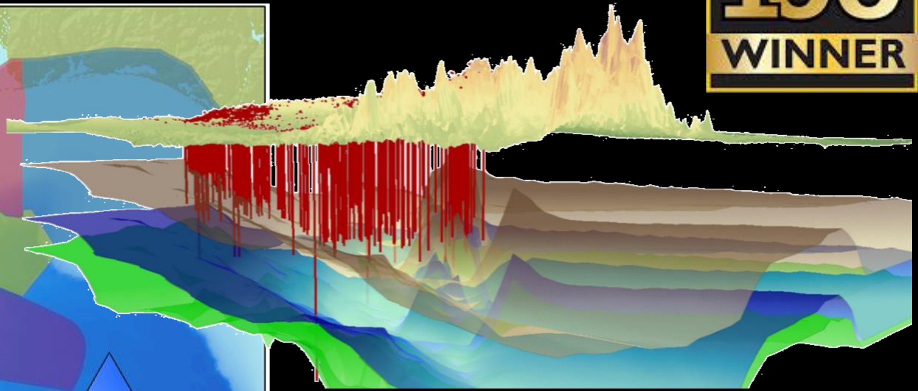
Questions?





Energy Data eXchange





For more information:

- Visit https://edx.netl.doe.gov/offshore
- Email mackenzie.mark-moser@netl.doe.gov, kelly.rose@netl.doe.gov, anuj.suhag@netl.doe.gov

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