

Subsurface Trend Analysis

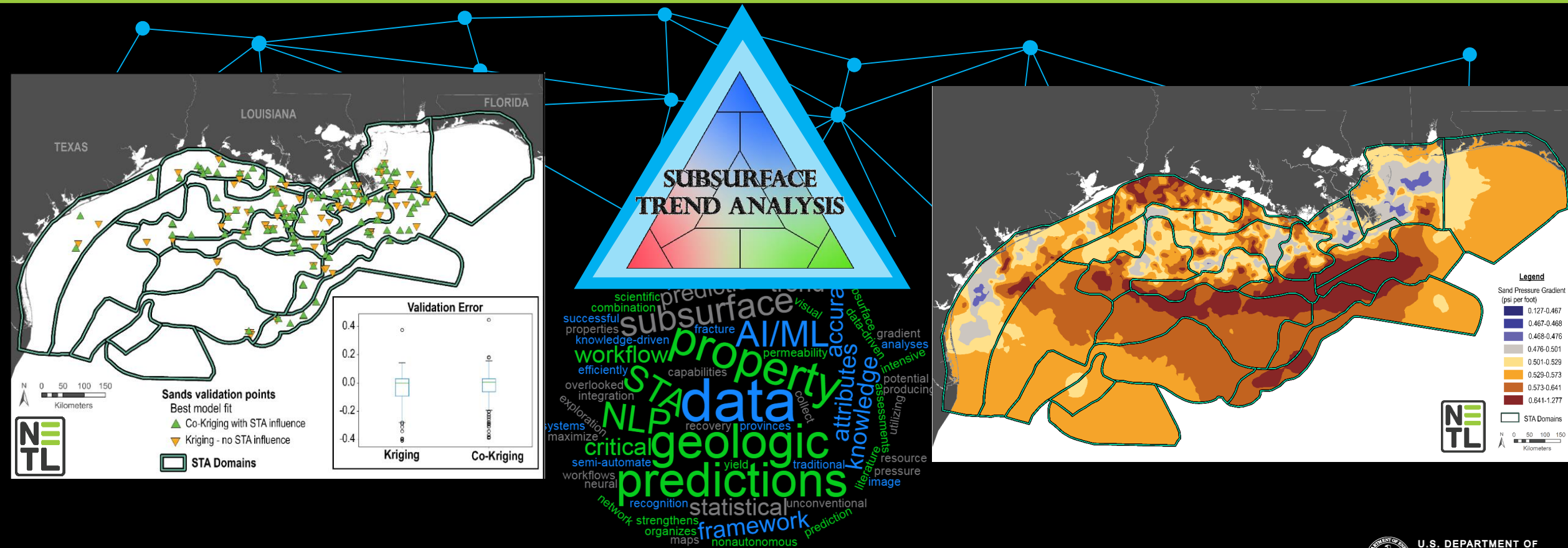
A methodical framework for artificial intelligence subsurface property prediction

MacKenzie Mark-Moser^{1,2}, Kelly Rose¹, Jennifer Bauer¹, Patrick Wingo^{1,2}, Anuj Suhag^{1,3}

¹National Energy Technology Laboratory ²LRST, ³ORISE Fellow



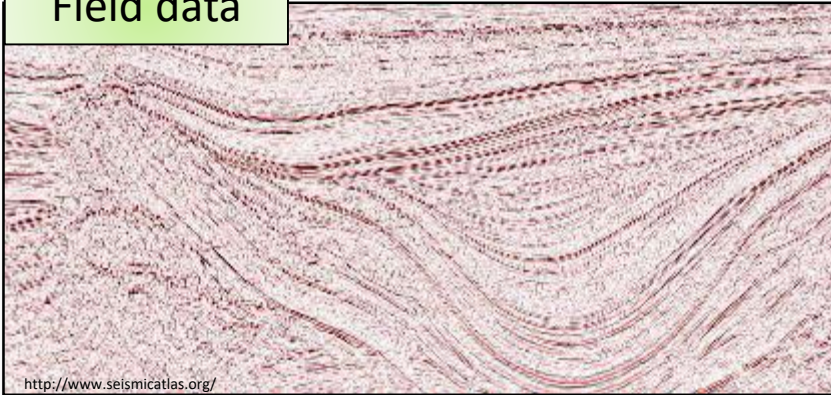
**AGU Fall Meeting, Dec. 11th 2019, Moscone Center,
San Francisco, CA**



Solutions for Today | Options for Tomorrow

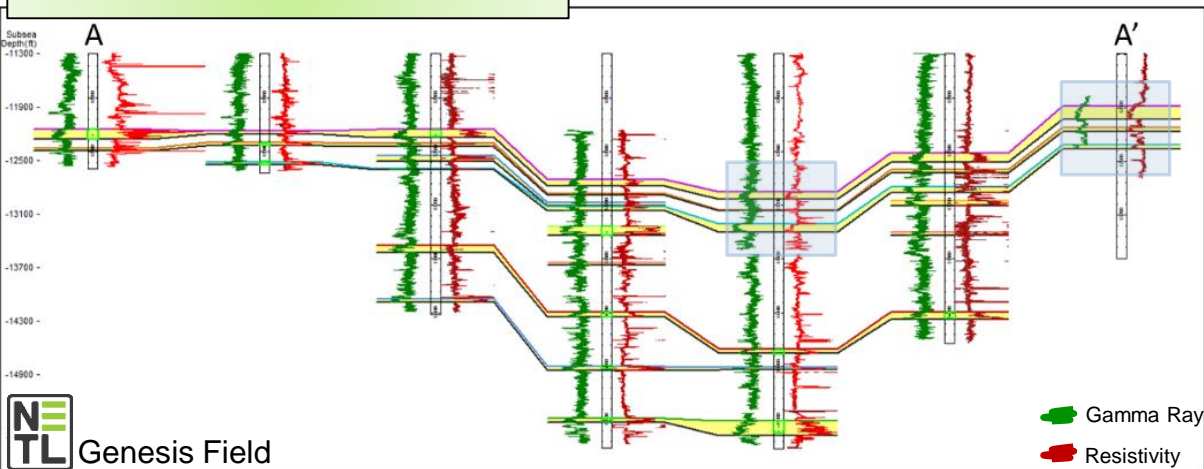
Classic Subsurface Interpretation Approach

Field data

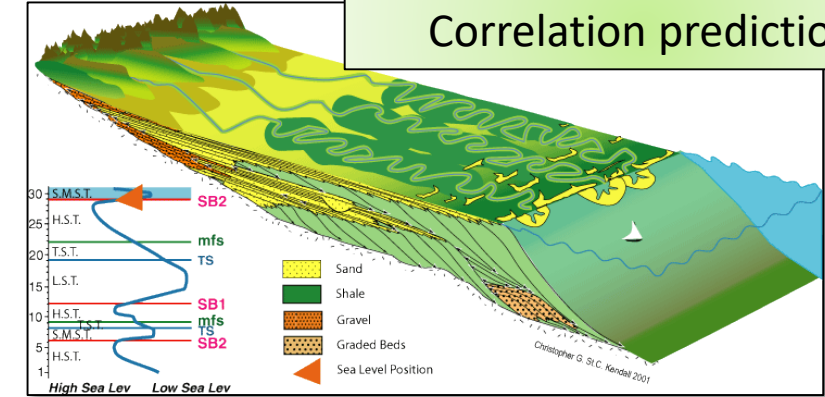


- Driven by field-based observations, interpretations and measurements
- Integrated with contextual geologic systems information *after the fact*
- Often indirect– only direct data are outcrop and core!

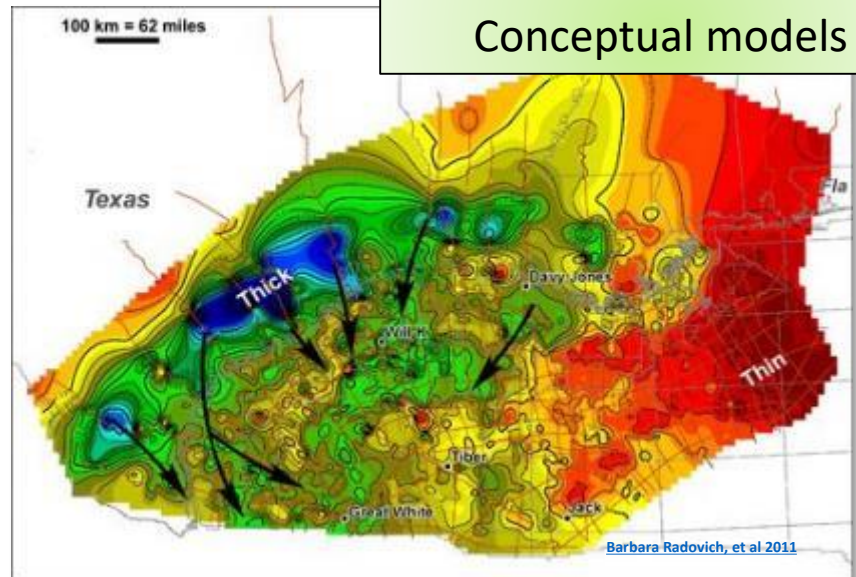
Indirect measurements



Correlation prediction



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

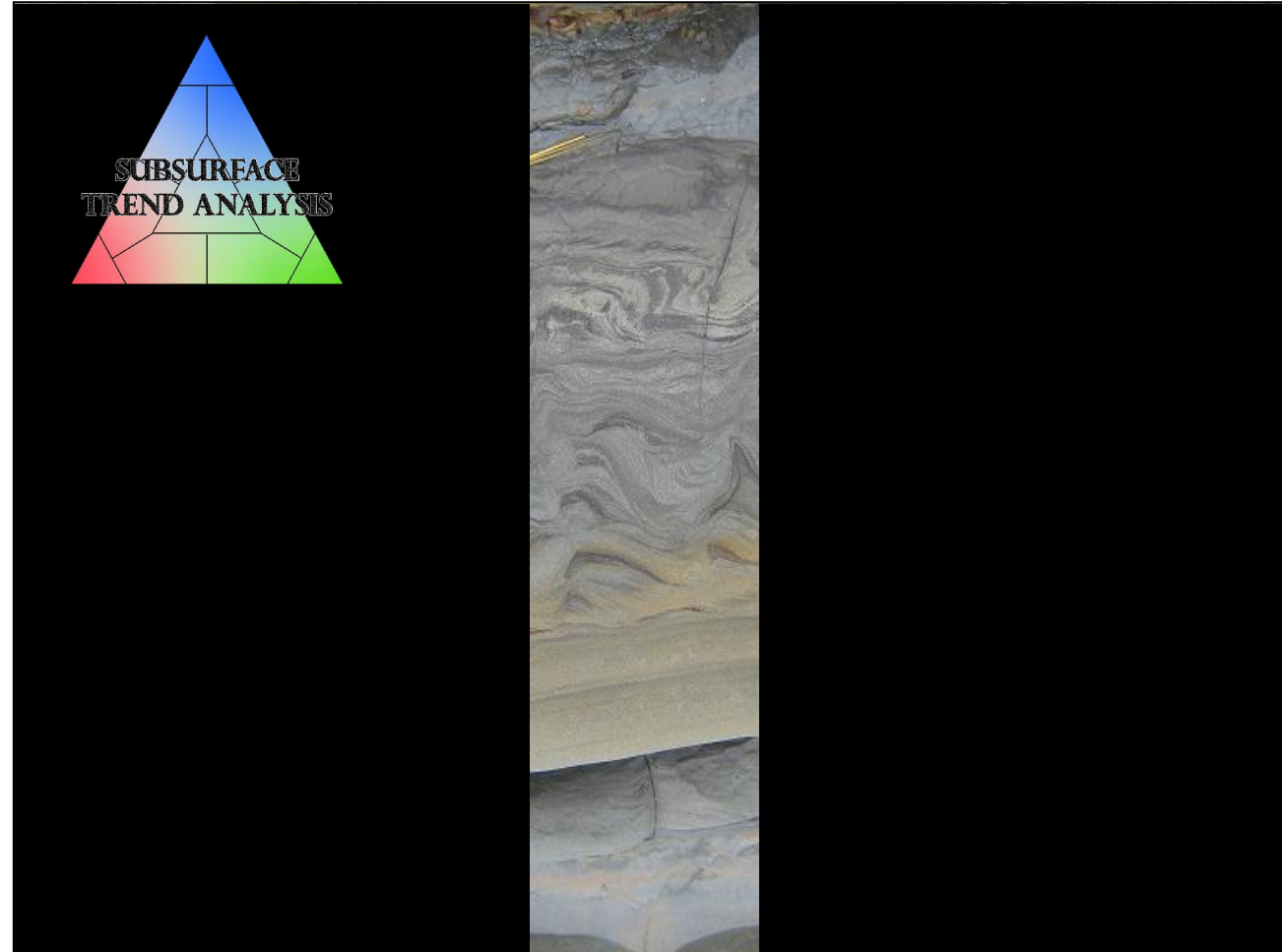
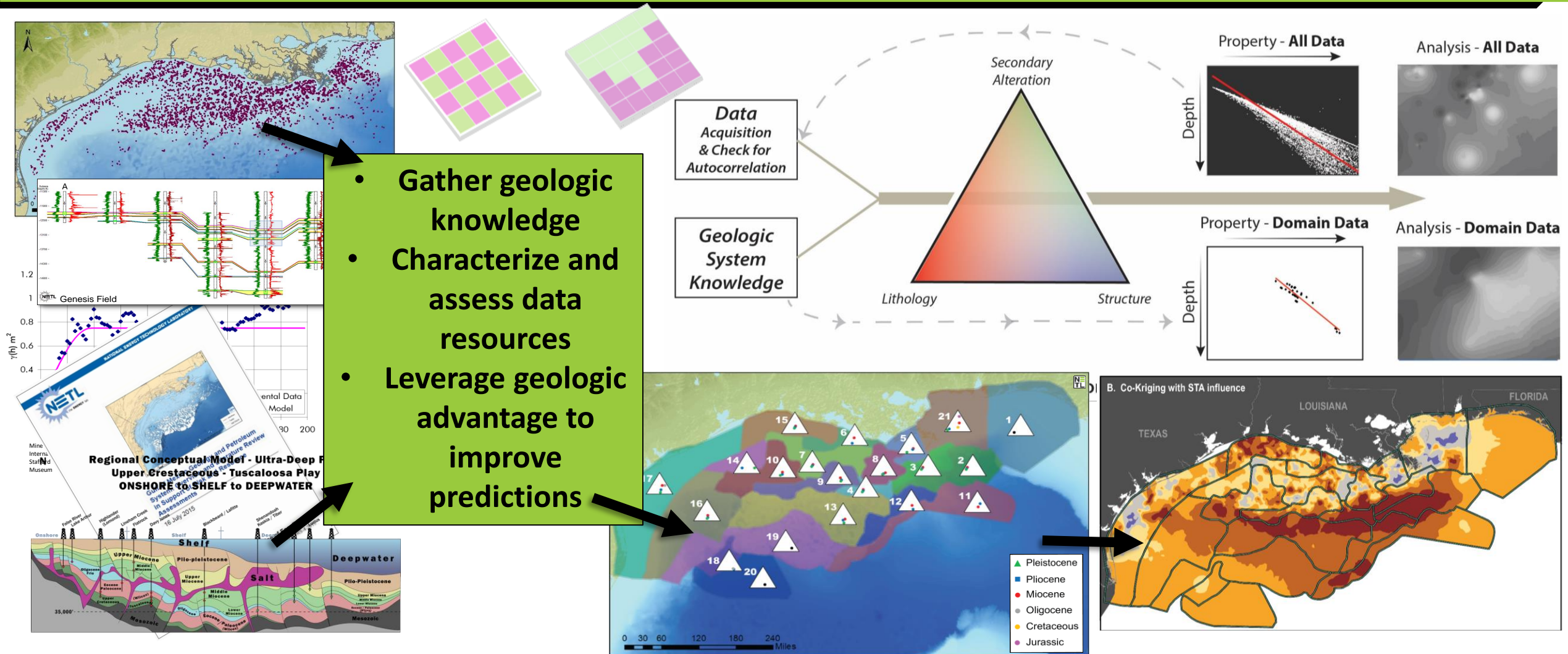
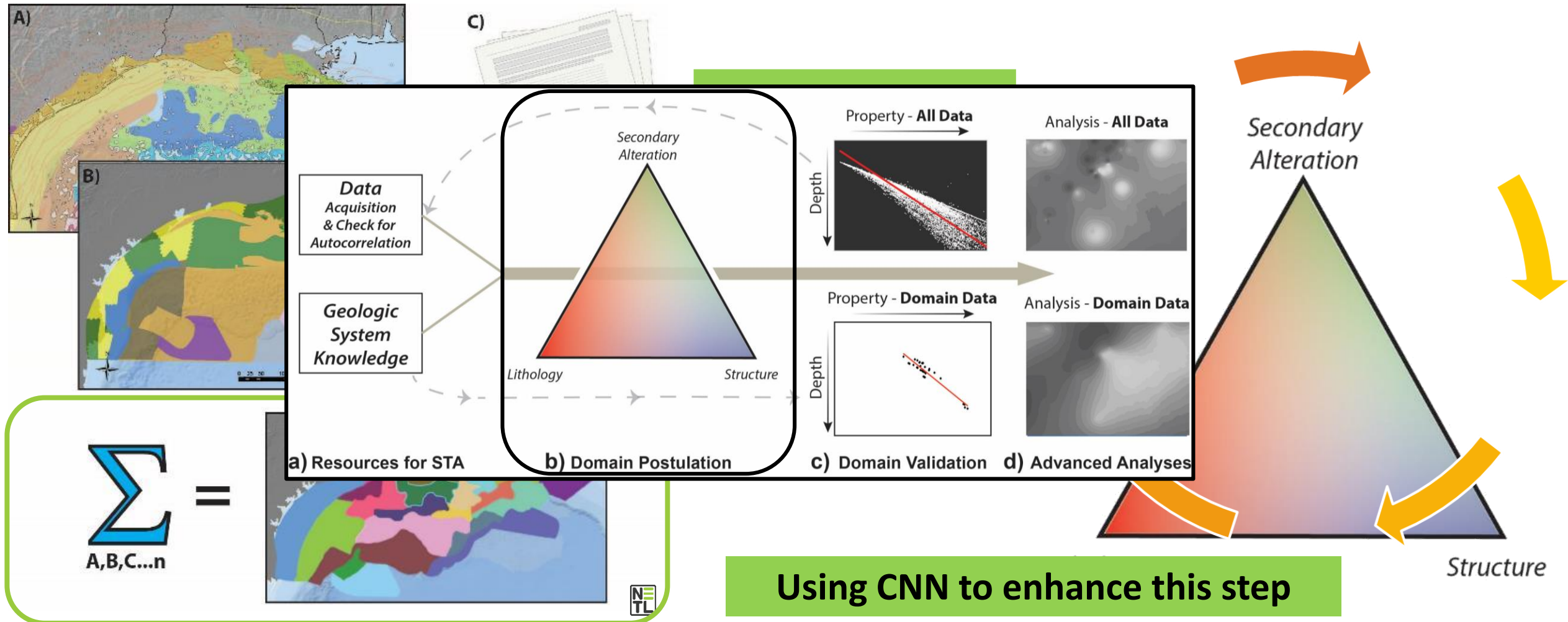


Photo Credit: Kelly Rose

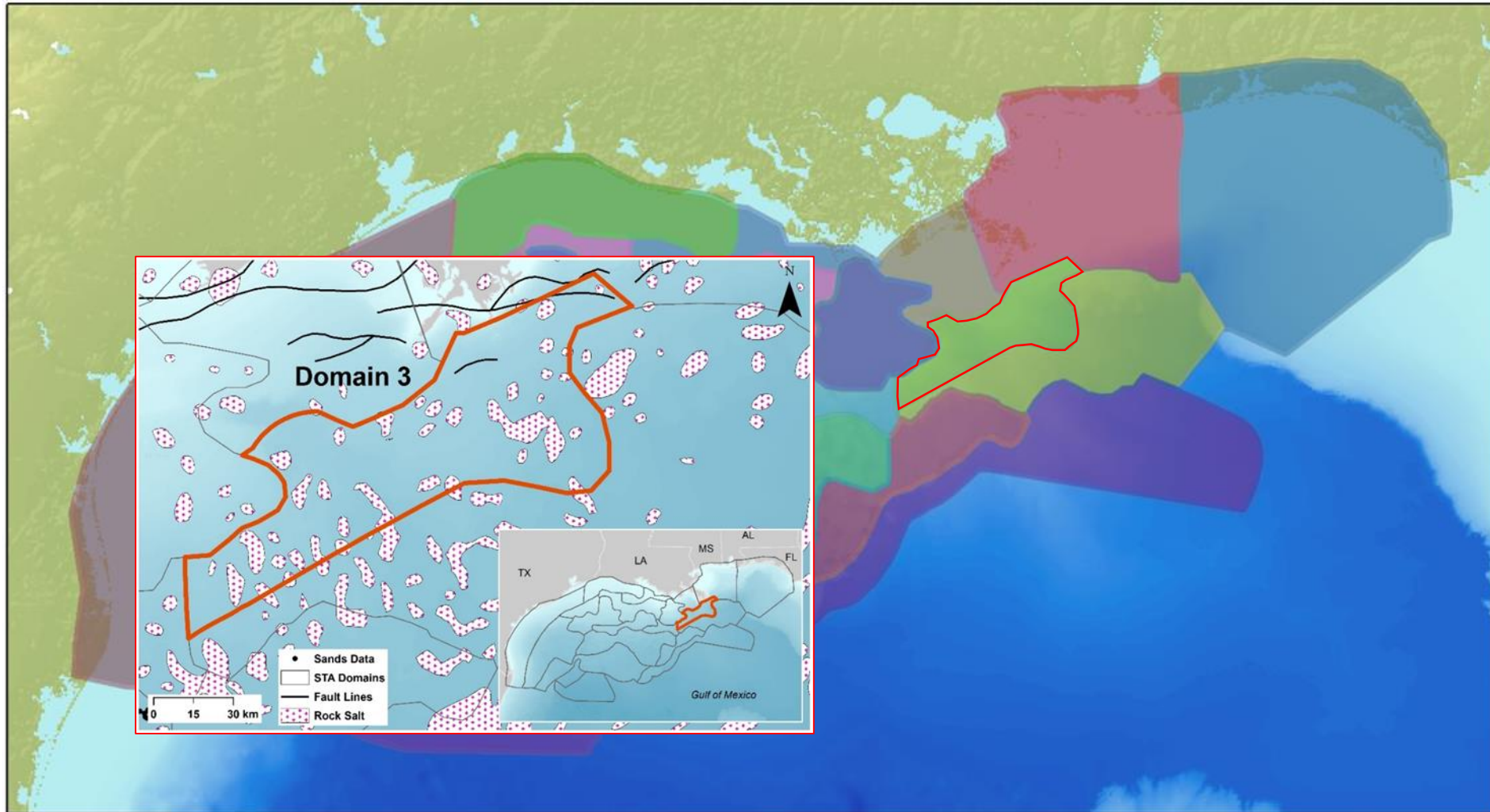
Science-driven workflow for improved subsurface prediction: the Subsurface Trend Analysis Method



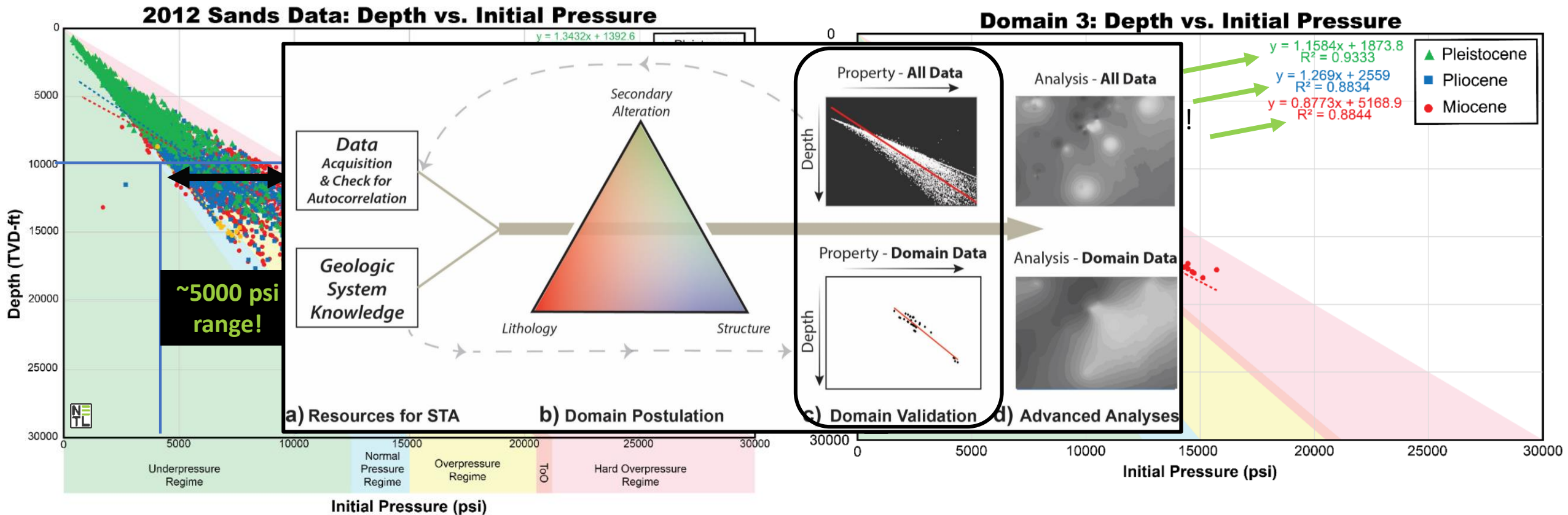
Defining areas with a common history



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

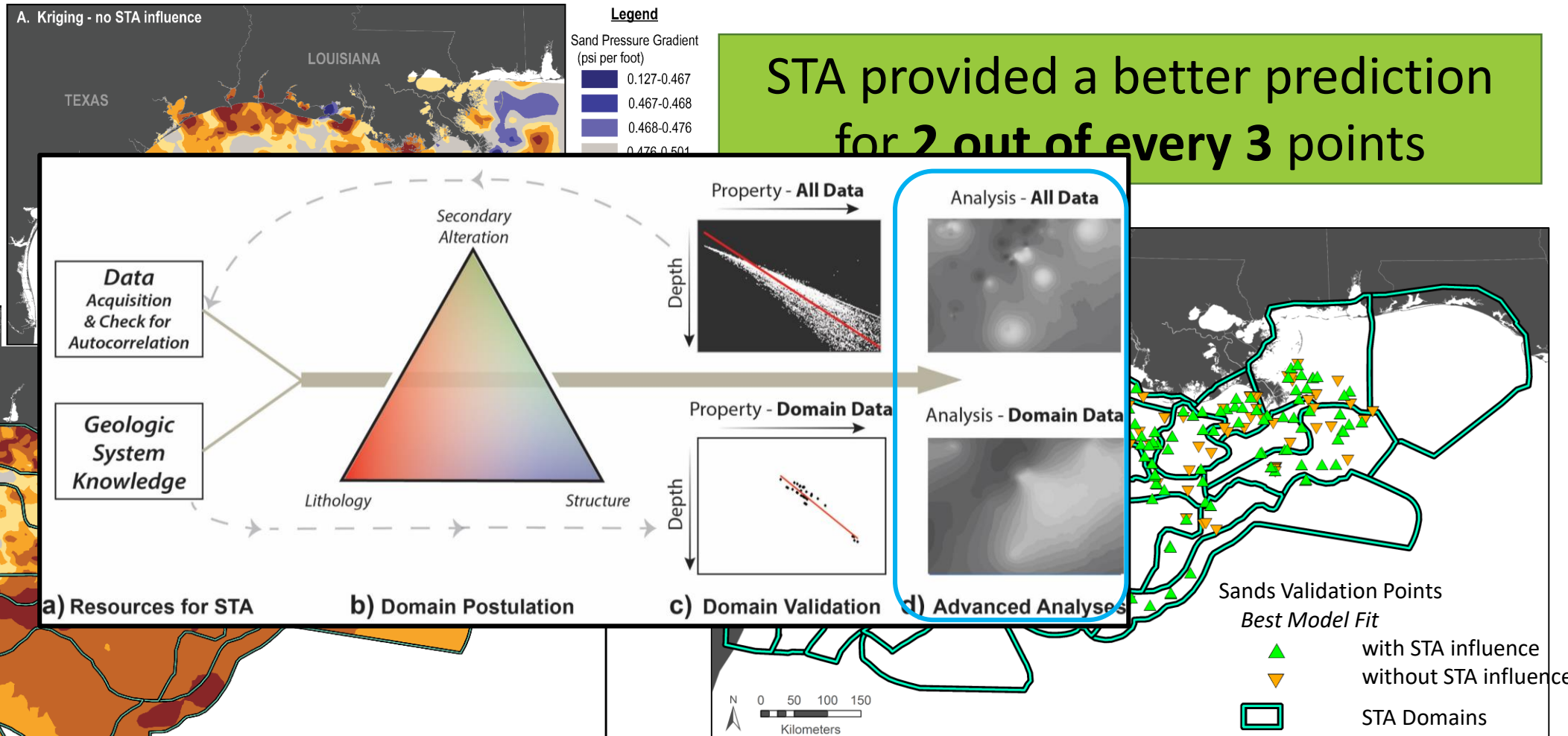
Mark-Moser, M., Miller, R., Bauer, J., Rose, K., and C. Disenhof. 2018, *Analysis of Subsurface Reservoir Properties Using a Novel Geospatial Approach, Offshore Gulf of Mexico*. NETL-TRS-2018

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*



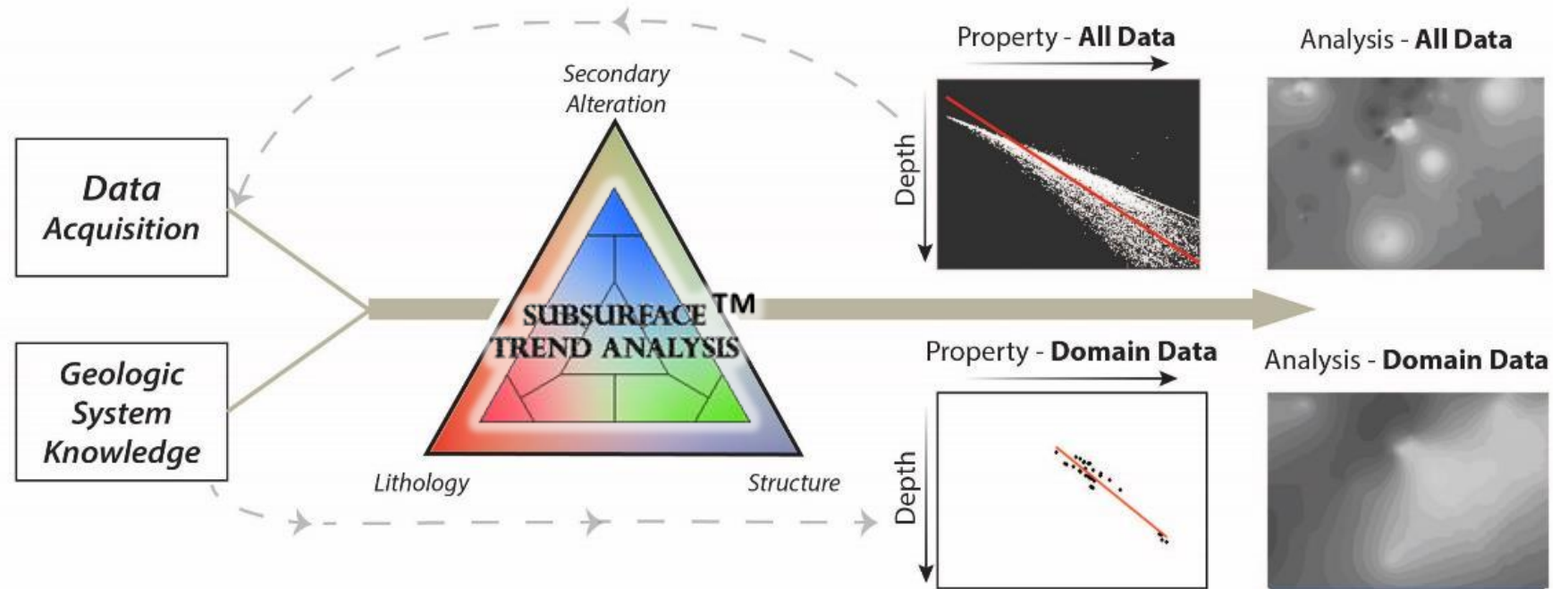
- Defined 21 STA geologic domains
- STA Predictions utilizing >13,000 data points

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

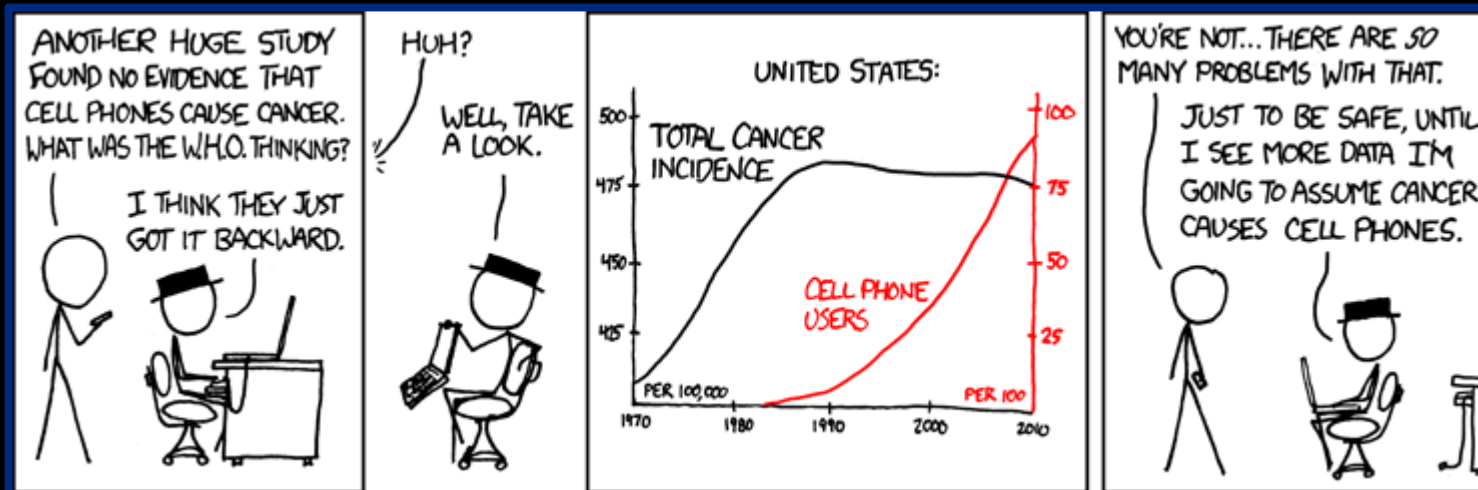
- **Extract** geologic contextual information literature
- **Identify** key spatial information (e.g. geologic provinces)

- **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

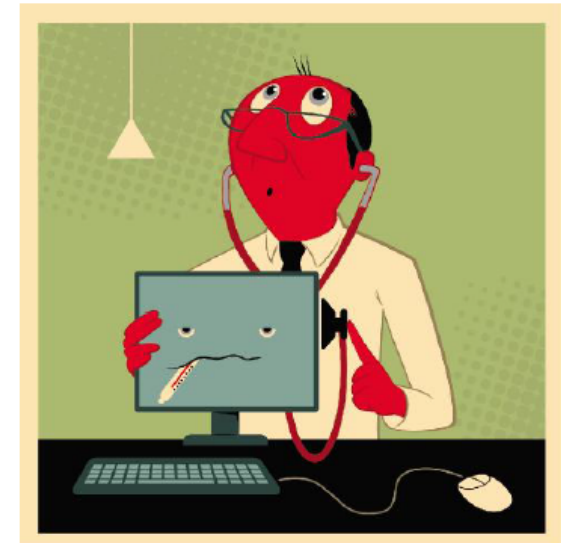
BIG DATA

The Parable of Google Flu: Traps in Big Data Analysis

David Lazer,^{1,2*} Ryan Kennedy,^{1,3,4} Gary King,³ Alessandro Vespignani^{1,5,6}

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?

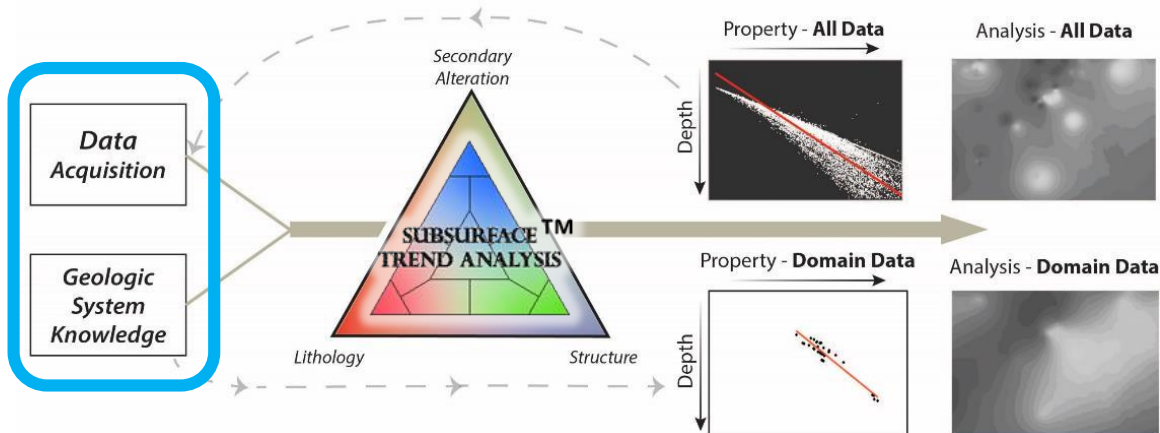
The problems we identify are not limited to GFT. Research on whether search or social media can predict x has become commonplace (5–7) and is often put in sharp contrast with traditional methods and hypotheses. measurement and construct validity and reliability and dependencies among data (12) Even the com



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

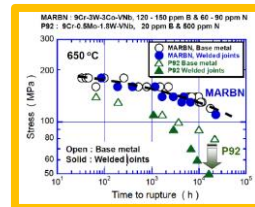
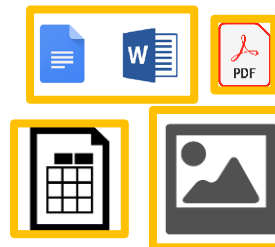
NLP & AI for unstructured data & knowledge

Gather, curate & transform



gensim
topic modelling for humans

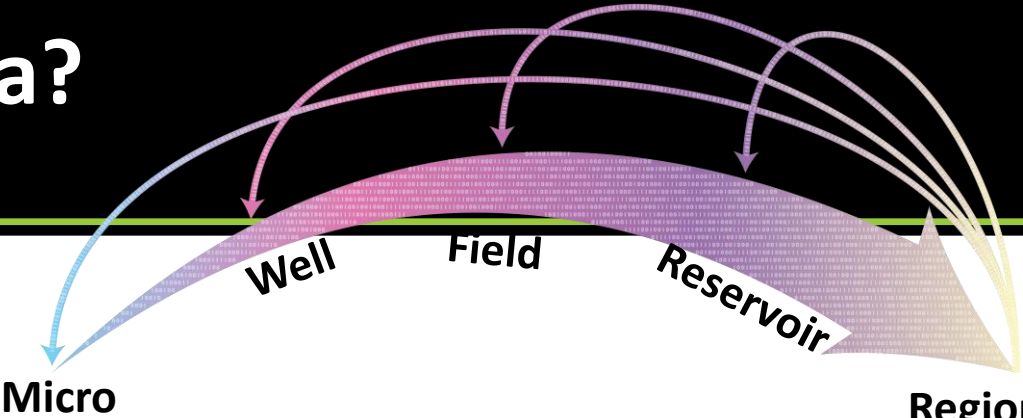
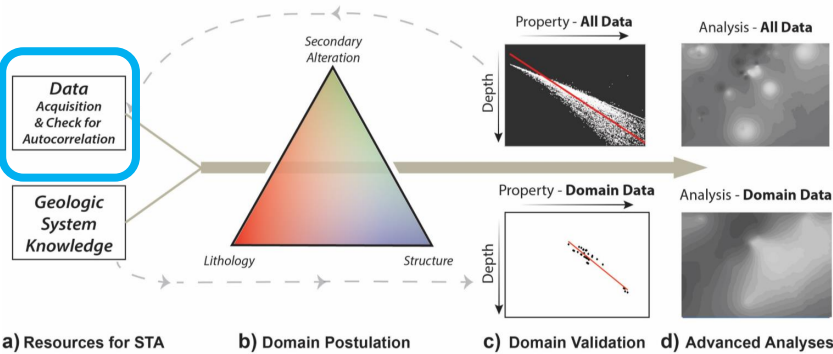
Mining/rescuing old data from documents, R&D products, presentations, etc. with the advantage of NLP and Neural Networks



```
[ -9.83387679e-02 -5.56519488e-04 -1.7448...
-2.03307748e-01 -3.83963548e-02 -7.0448...
-1.27868637e-01 -4.75849025e-02 2.61805...
-2.26884838e-02 -1.24736726e-01 -2.20635176e-01 -7.6324311e-01
2.48466954e-01 6.05324358e-02 -1.36489823e-01 -7.38347322e-02
1.15407623e-01 3.11685920e-01 3.41568053e-01 4.33256000e-01
-6.75646076e-03 2.88540751e-01 -4.58129570e-02 -2.84765691e-01
8.85103270e-02 3.89607370e-01 1.47015020e-01 -2.71234483e-01
-3.14288795e-01 1.66953489e-01 -2.28361294e-01 -3.78020555e-01
-7.28930831e-02 3.00001532e-01 -1.27531856e-01 7.65746981e-02
1.20075434e-01 2.76841551e-01 -2.22608969e-02 2.49355540e-01
3.28565747e-01 -1.33367807e-01 -1.20049961e-01 1.91137344e-02
9.41278636e-02 1.38300151e-01 1.40407398e-01 2.08285078e-01
-2.53193807e-02 -3.50823522e-01 -1.32105038e-01 -1.34713382e-01
2.59572566e-01 -6.11362867e-02 -2.37589628e-01 -1.57545641e-01
-3.32306623e-02 1.00493170e-01 8.00671205e-02 -6.56809136e-02
-3.89025569e-01 2.89072007e-01 1.44446552e-01 -8.57475400e-02
3.49803627e-01 2.13039834e-02 2.74429787e-02 -5.25448509e-02
-1.21638365e-01 1.21457621e-01 1.18799843e-01 1.95591331e-01
-3.04924250e-01 1.62359387e-01 1.73498645e-01 -2.98044860e-01
-1.11183167e-01 -2.11784944e-01 1.25914544e-01 -4.40392137e-01
-6.12635016e-02 3.23301762e-01 -3.21543872e-01 -2.18976587e-01
1.72268063e-01 -4.81558032e-02 6.03809394e-02 1.08830504e-01
4.26880717e-02 4.77878451e-02 -4.44996715e-01 -2.91929096e-01
-6.17334135e-02 -2.71454215e-01 2.70434946e-01 1.65137619e-01
5.15525490e-02 -1.33983821e-01 1.15802862e-01 3.06160748e-01
-5.90535253e-02 2.89506316e-01 -2.95480881e-02 1.59748912e-01
4.36033159e-02 -1.79844834e-02 -3.07084829e-01 1.31049633e-01
-2.53072437e-02 8.73016939e-02 -1.94220737e-01 3.50081921e-02
```

Digital structured data

What is our target data?



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

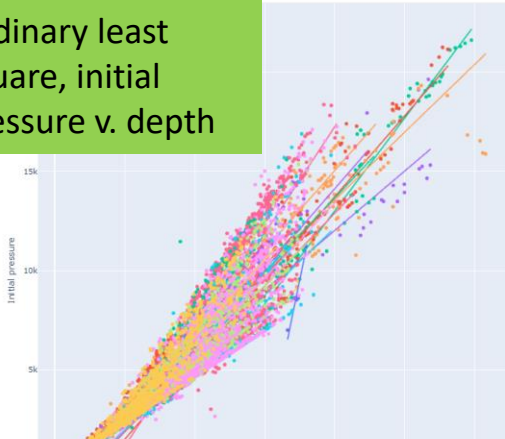
Target data for the STA method is:

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

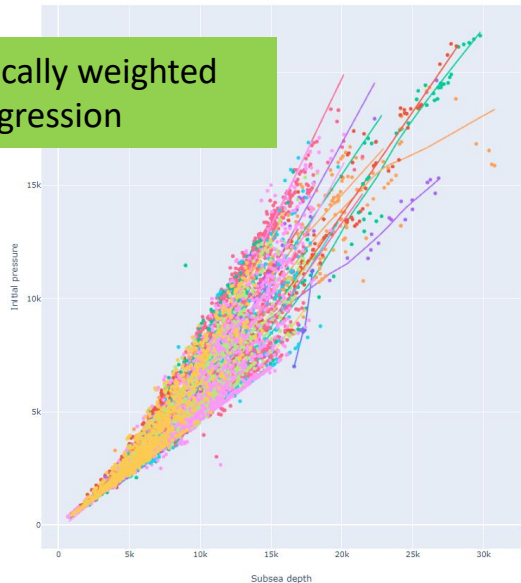


Dimensional analyses of subsurface property data

Ordinary least
square, initial
pressure v. depth

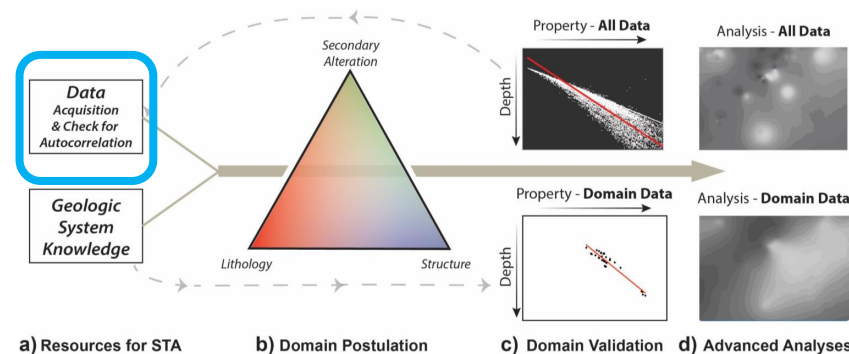


Locally weighted
regression

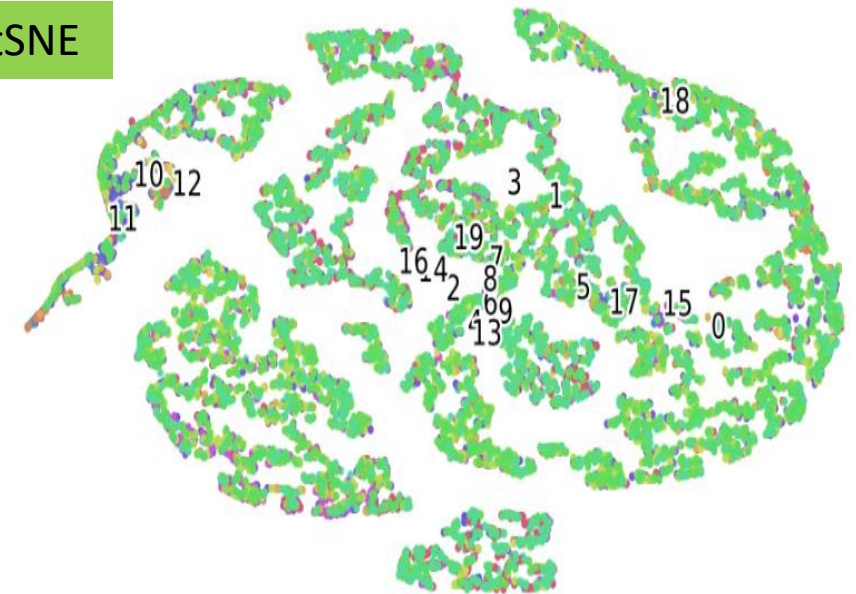


Keys to meaningful ML analysis:

- 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)

-



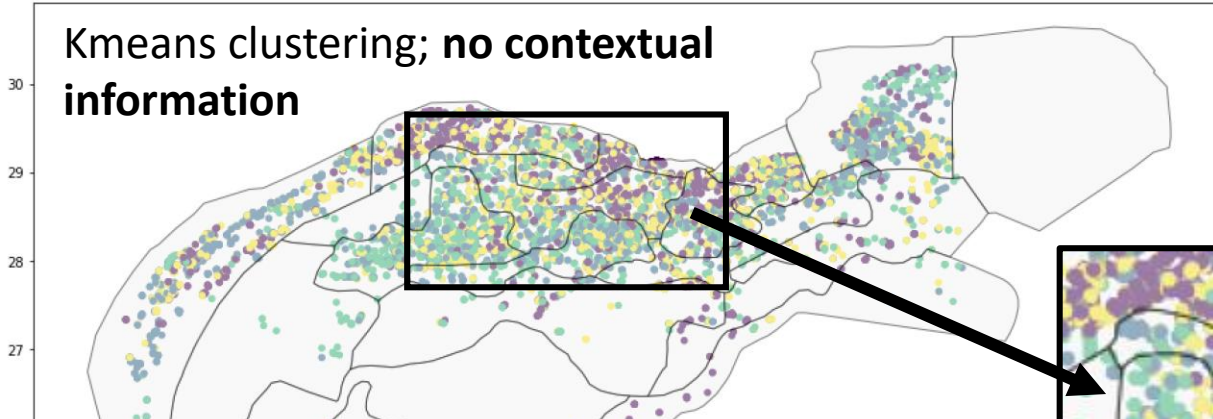
The diagram illustrates the flow of data through a CNN. It starts with an **Input** image of a cat. This is processed by a **Convolutions** layer, which produces a stack of feature maps labeled **f.maps**. These are then passed through a **Subsampling** layer. The resulting feature maps are then processed by another **Convolutions** layer, followed by another **Subsampling** layer. Finally, the output is fed into a **Fully connected** layer, which is represented by a network of nodes and connections.



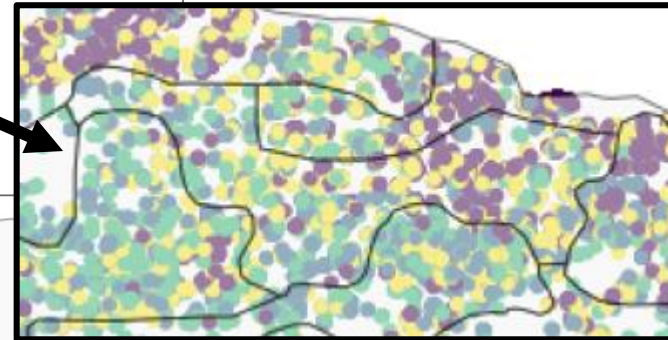
Domain Validation & Universal Clustering Analysis

Gulf of Mexico application

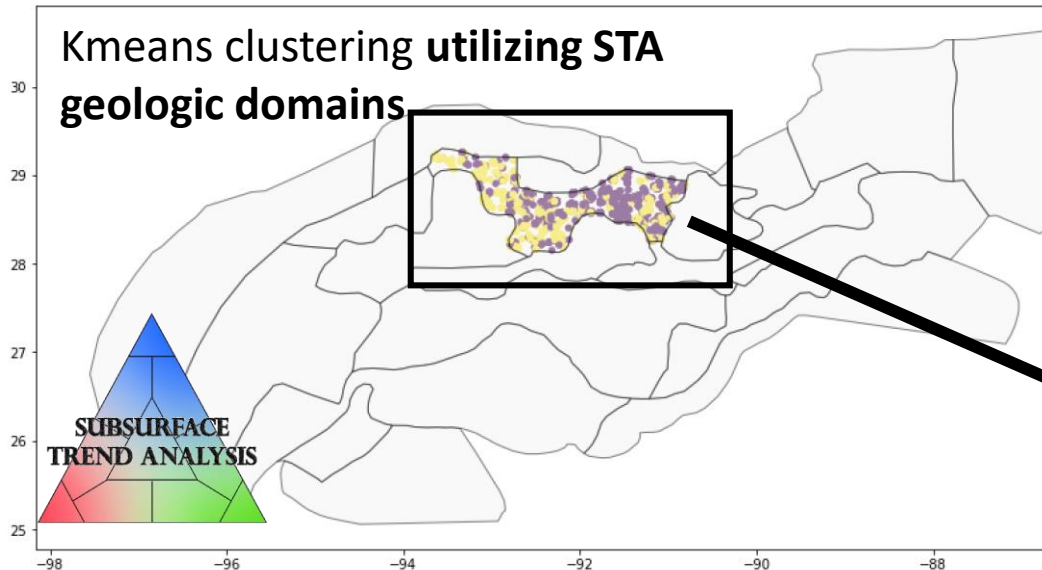
Kmeans clustering; no contextual information



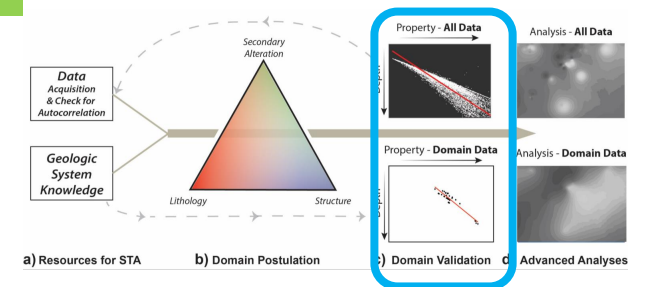
Kmeans dimensional analysis of 6 subsurface properties



Kmeans clustering utilizing STA geologic domains



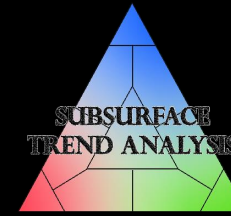
Results are repeatable when utilizing Dbscan!



- 4 clusters
- Poor continuity among clusters

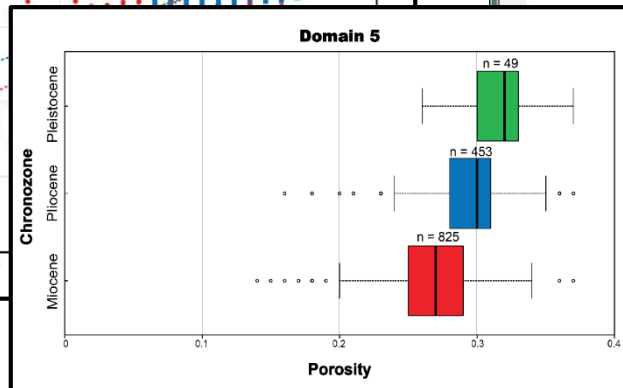
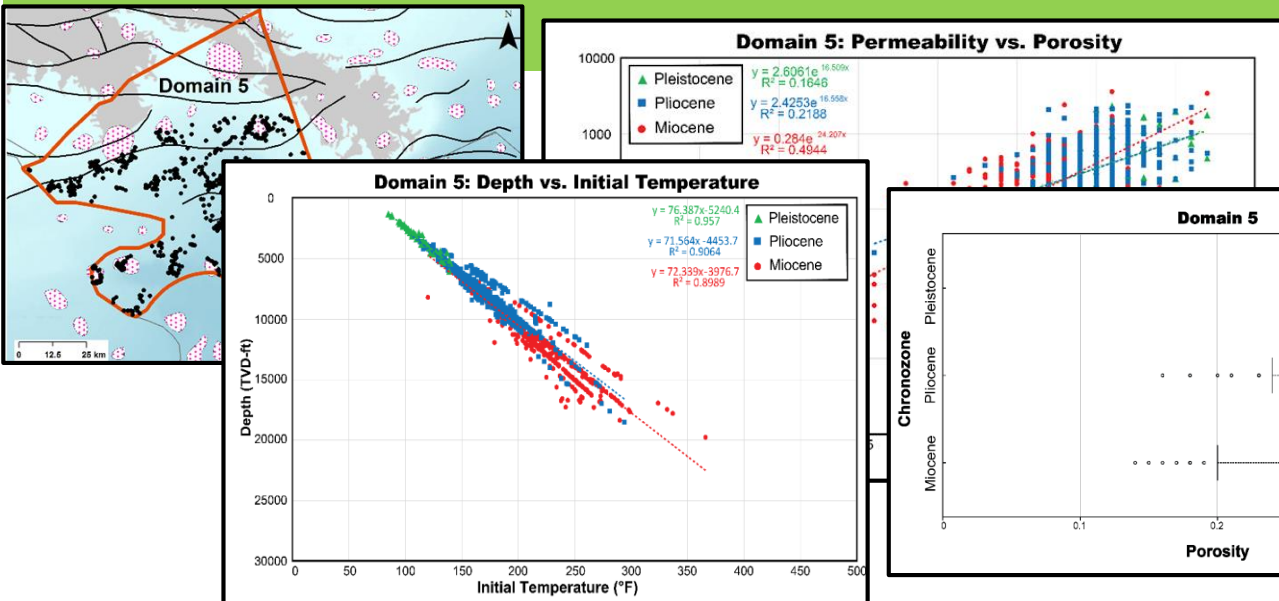
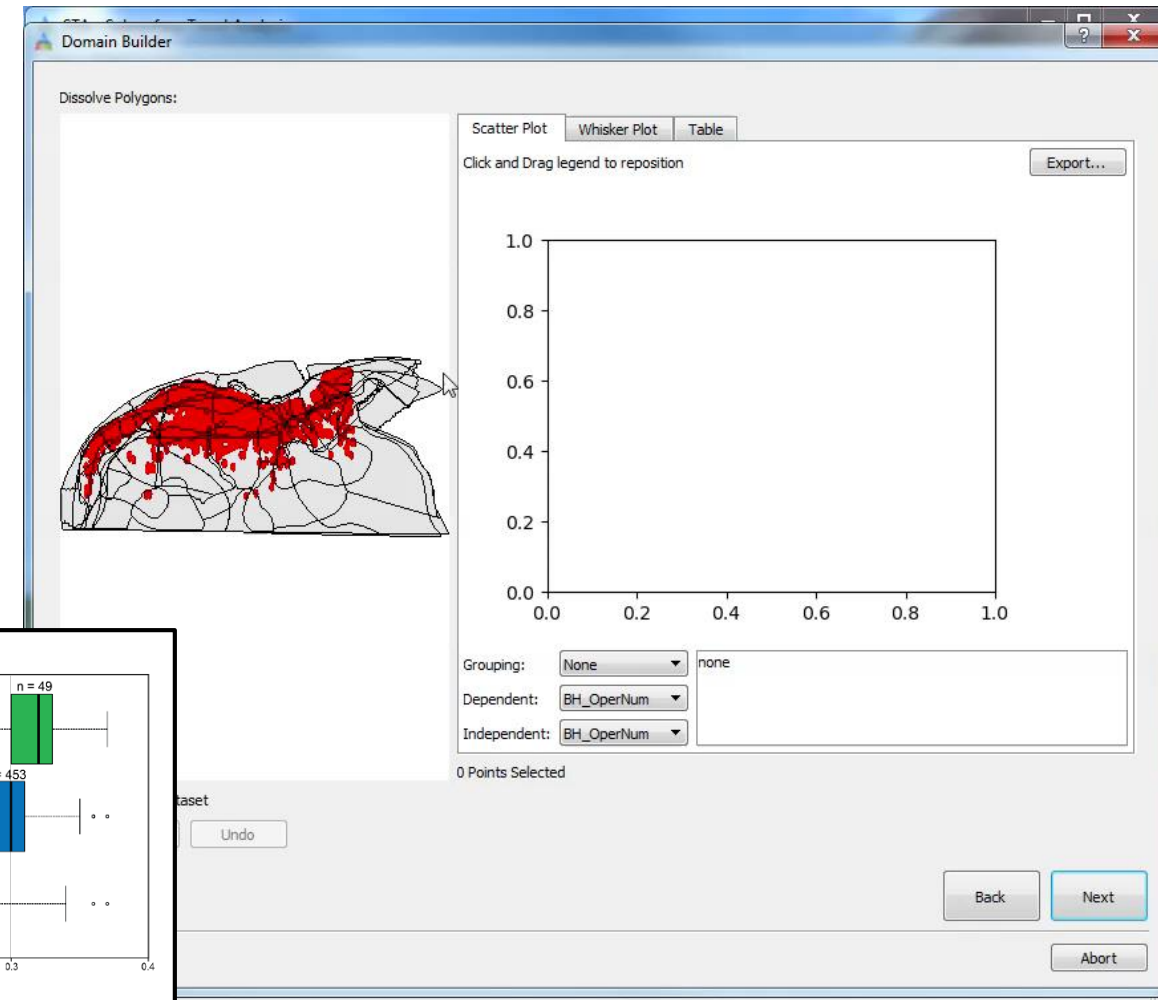
- 2 clusters
- Improved continuity among clusters

Ongoing STA Tool development

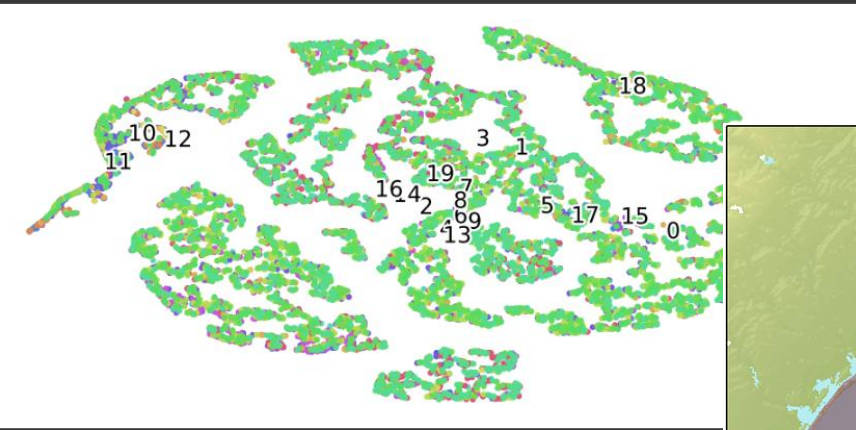
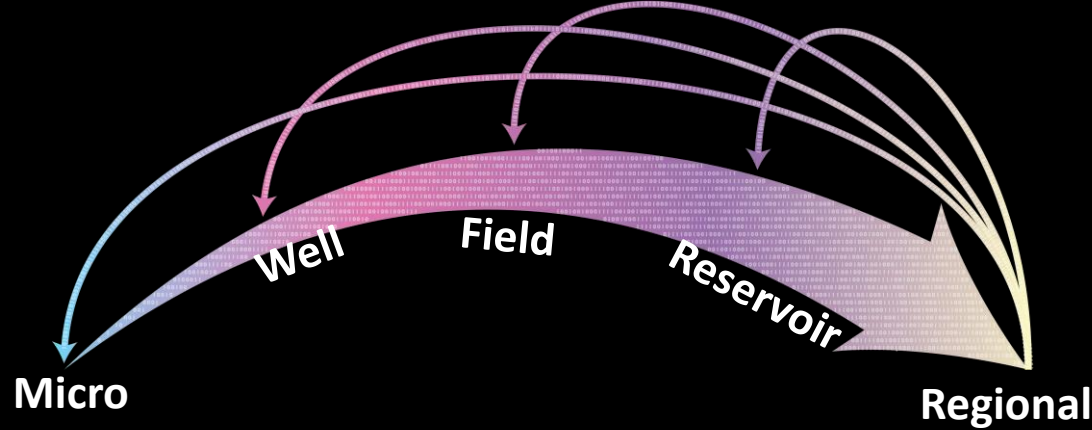


STA Tool– a *virtual research assistant* designed to

- 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?



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.

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.