Rapid Exploration in a Mature Area in Northwest Kansas: Improving the Definition of Key Reservoir Characteristics using Big Data*

W.E. (Bill) Full[1] and Steve James[1]

Search and Discovery Article #41685 (2015)**
Posted September 28, 2015

*Adapted from oral presentation given at AAPG Education Directorate Geoscience Technology Workshop, Revitalizing Reservoirs-Rocky Mountains, Mid-Continent, Canada, International Focus, Golden, Colorado, August 11-12, 2015

[1]GXSTAT (Bill@GXStat.com, Steve@GXStat.com)

Abstract

With the rapid changing economic environments associated with defining new viable oil and gas plays in a mature region, attention has turned to the statistical analysis of all of the available data (i.e. big data). These studies have been on basin-wide through individual play scales. For the most part, the tools used are statistical, geostatistical and multivariate in nature. Oftentimes, the user either is given a toolset found within a larger program or works with one of the many fine programs available on the market. Even with a deep understanding of the myriad of assumptions associated with of these approaches, it is difficult to extract all but the most obvious results from the data let alone the more thorny questions on how to quantify the economic risk.

A common workflow is to gather in the best data available (e.g. geologic, geophysical, geochemical, log-based), create multiple layers/surfaces of geostatistically-mapped information and then perform some appropriate multivariate analysis. A great many of the assumptions associated with the common multivariate techniques are based on the necessity of the data being derived from either one or a fixed number of known populations. With big data, the verification of these assumptions is often overlooked resulting in statistically ambiguous or difficult to validate results. An extra step in the workflow needs to be added with these cases - partitioning the data in an appropriate way and analyzing each partition separately before recombining to produce a final risk map. Recognizing when this partitioning is needed requires visual, statistical, geostatistical and deterministic techniques, as described below.

This study consisted of a large data set of well-based and geophysical data (gravity and magnetic) in several counties in northwest Kansas. In this area, the early Paleozoic rocks are likely dominated by basement tectonics at the time of deposition and the later Paleozoic formations appear overlie the earlier rocks including their related fault/fracture zones. After recognizing the visual hints that data partition was appropriate, computer programs designed for data partitioning (Polytopic Vector Analysis-based programs including Hyperplanar Vector Analysis, Fuzzy and Hard Clustering including Fuzzy N-Varieties) were applied. The results showed that partitioning the data produced a more refined probability of success than could be defined by the multivariate analysis alone. Key variables tied to reservoir quality were also defined that

have been used to increase understanding of both new prospects and potentially increase reservoir production for known fields. This increased knowledge directly leads to a more confident economic risk assessment in a mature area.

References Cited

Bezdek, J.C., 1981, Pattern Recognition with Fuzzy Objective Function Algorithms: New York, Plenum, 256 p.

Bezdek, J.C., R. Ehrlich, and W.E. Full, 1984, FCM; The FUZZY C Means Clustering Algorithms: Computers and Geosciences, v. 10/2-3, p. 191-203.

Bezdek, J.C., M. Trivedi, R. Ehrlich, and W.E. Full, 1984, FUZZY CLUSTERING: A new approach for geostatistical analysis: Int. Jour. of Sys., Meas. and Decisions, v. 1/2, p. 13-24.

Ehrlich, R., and W.E. Full, 1987, Sorting out geology Unmixing mixtures: IAMG Special Publication, W. Size, ed., Oxford University Press, New York, p. 33-46.

Evans, J.C., R. Ehrlich, D. Krantz, and W.E. Full, 1992, A comparison between polytopic vector analysis and empirical orthogonal function analysis for analyzing quasigeostopic potential vorticity: JGR-Green, v. 97/C2, p. 2365-2378.

Full, W.E., 1982, The EXTENDED CABFAC/QMODEL family algorithms a multivariate pattern classification scheme: Naval Oceanographic Research and Development Activity Symposium on Pattern Recognition in the Marine Environment, Pame Abstracts.

Full, W.E., 1983, Models and underlying assumptions inherent in the analysis of mixtures: Geol. Soc. of Am., Southcentral, Abst.

Full, W.E., 1985, Non constant sum linear unmixing: Petroleum Research Fund Annual Report.

Full, W.E., 1986, Models for unmixing data whose row sums are not constant: Petroleum Research Annual Report.

Full, W.E., 1989, A new implemented approach for non constant sum linear unmixing for geochemical data analysis: GSA Symposium, St. Louis, MO.

Full, W.E., R. Ehrlich, and J.C. Bezdek, 1982, FUZZY QMODEL: A new approach for linear unmixing: Int. Jour. of Math Geol., v. 14/3, p. 257-268.

Full, W.E., R. Ehrlich, and J.E. Klovan, 1981, EXTENDED QMODEL Objective definition of external end members in the analysis of mixtures: Int. Jour. of Math. Geol., v. 13,/4, p. 331-334.

Full, W.E., and C. Gary, 2006, Polytopic Vector Analysis - Non-Constant Sum (PVA-NCS): Confidence and Risk Assessment: AAPG Annual Convention, Houston, TX.

Imbrie, J., 1963, Factor and Vector Analysis Programs for Analyzing Geologic Data: Office of Naval Research. Tech Report No. 6, p. 83.

Imbrie, J. and T.H. Van Andel, 1964, Vector analysis of heavy mineral data: Bull. Geol. Soc. Amen, v. 75, p. 1131-1156.

JMP Pro 12, SAS Institute Inc., 2015, Cary, NC, SAS Institute Inc.

Klovan, J.E. and J. Imbrie, 1971, An algorithm and FORTRAN-IV program for large scale Q-mode factor analysis and calculation of factor scores: Jour. Math. Geol., v. 3/1, p. 61-76.

Klovan, J.E. and A.T. Miesch, 1976, EXTENDED CABFAC and QMODEL computer program for grain-size distributions: Jour. Sed. Pet., v. 36/1, p. 115-125.

Klovan, J.E. and A.T. Miesch, 1976, EXTENDED CABFAC and QMODEL computer programs for Q-mode factor analysis of compositional data: Comput. Geosci., v. L/3, p. 161-178.

Miesch, A.T., 1976a, Q-mode factor analysis of geochemical and petrologic data matrices with constant row sums: U.S.G.S. Prof. Paper 574G.

Miesch, A.T., 1976b, Q-mode factor analysis of compositional data: Comput. Geosci., v. 1/3, p. 147-159.

Rapid Exploration in a Mature Area in Northwest Kansas:

Improving the Definition of Key Reservoir Characteristics using Big Data

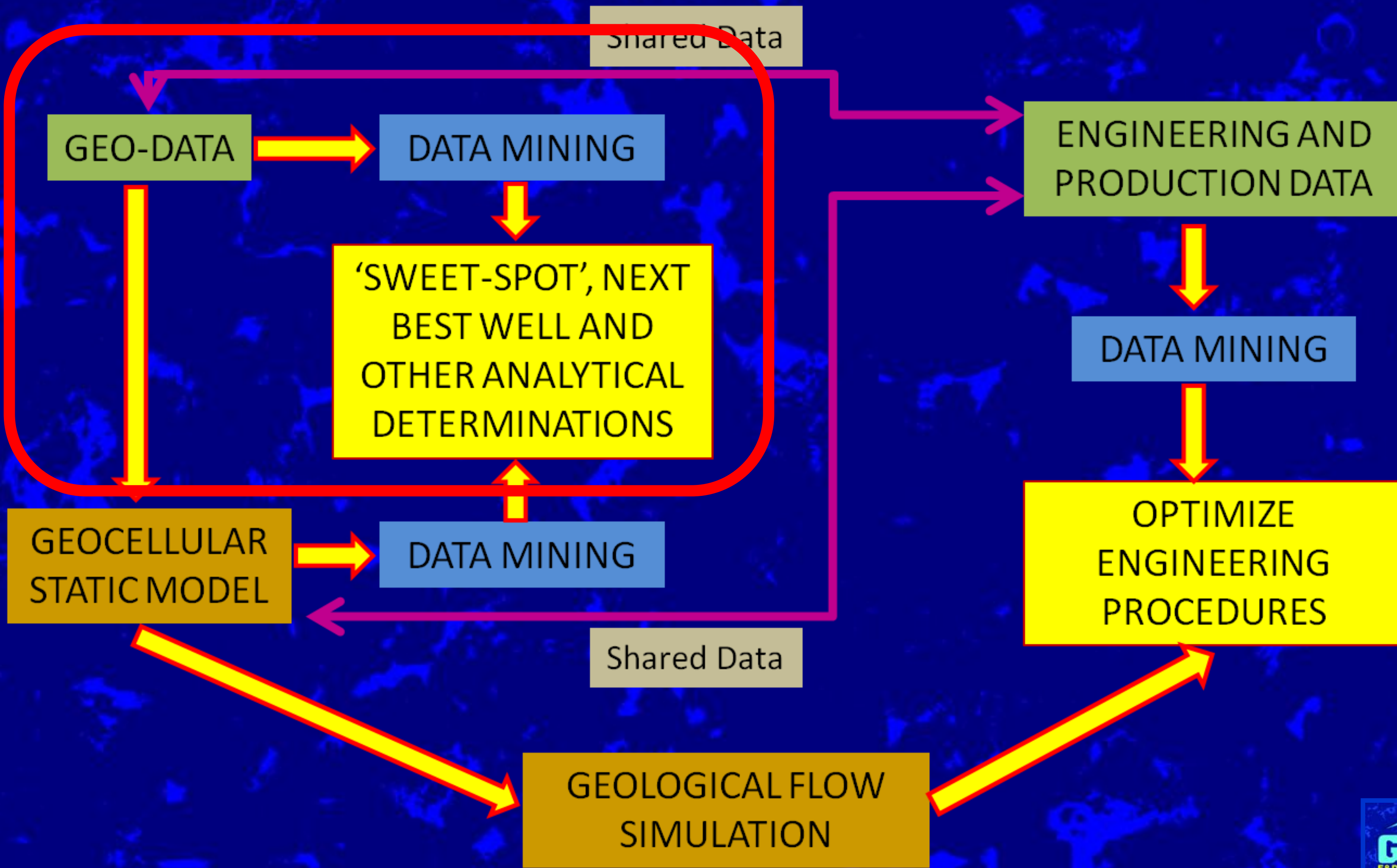
W. E. (Bill) Full
Steve James
www.GXStat.com

GXSTAT
E&D Big Data Solutions

Overall Goals of Larger Project

- Rapid exploration of a mature areas
  - Define locations in a region known by a company
  - Rapid evaluations in new regions
- Make an assessment of the reservoir quality of the rock and type of reservoir (fractured, stratigraphic, structural, fluid flow parameters)
- Give the engineer as much information as possible for defining a completion plan
- Find the “next-best” well location (forward modeling)
- Determine a realistic risk assessment of success and failure

Generalized Vision of the Larger Analytic Workflow



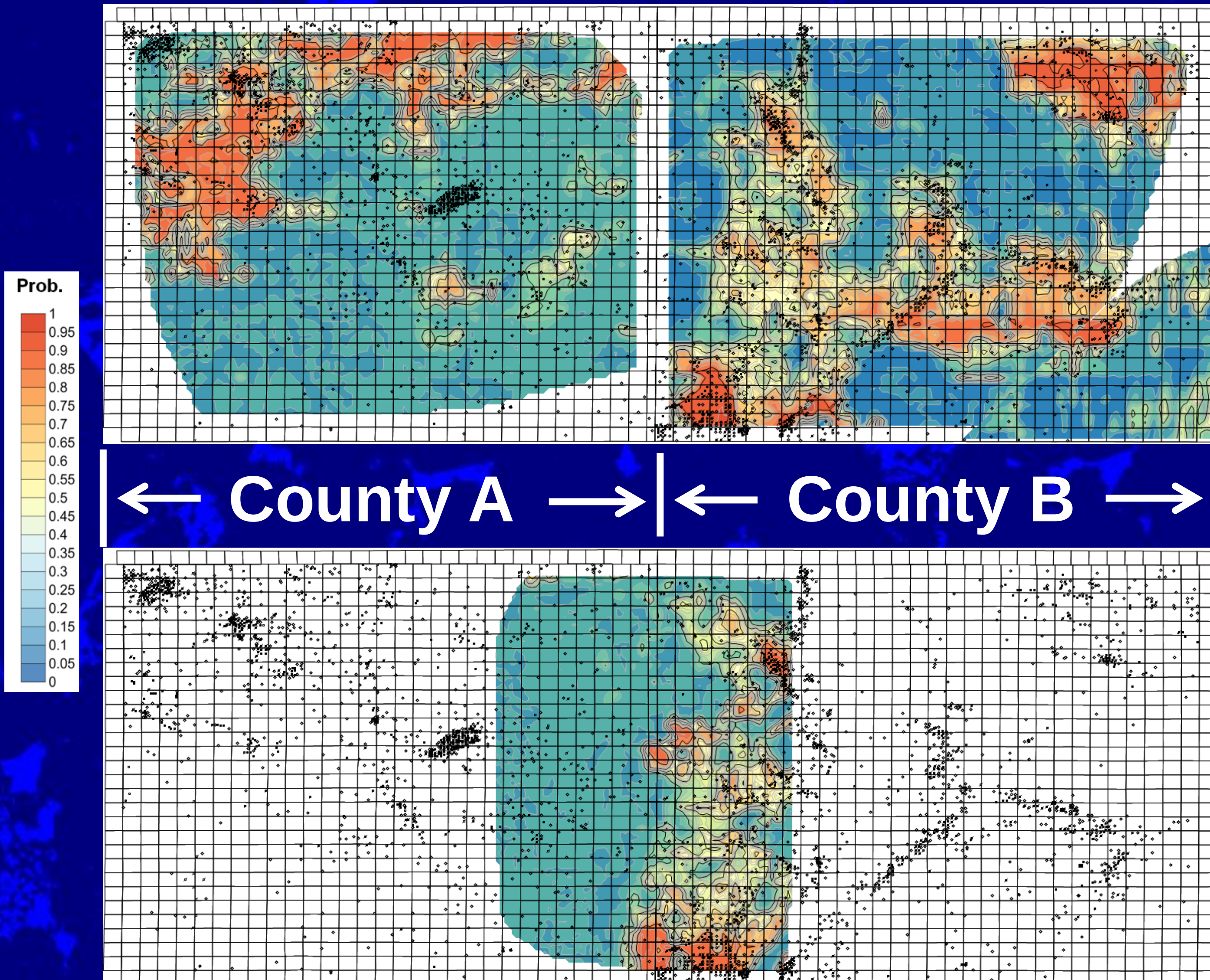
GEO-DATA: Includes geophysical, petrophysical (logs), seismic and other available data
ENGINEERING AND PRODUCTION DATA: All completion and production information

The Problem

• Given a set of data used to determine the probability (risk) of finding a producing well location for any given area, overlapping or adjacent areas do not necessarily correspond to each other

• Put another way, what is the ideal size of a data set to analyze for any given area?
• When and how far can we extend information beyond wells in an area?

The Problem – Map Visual



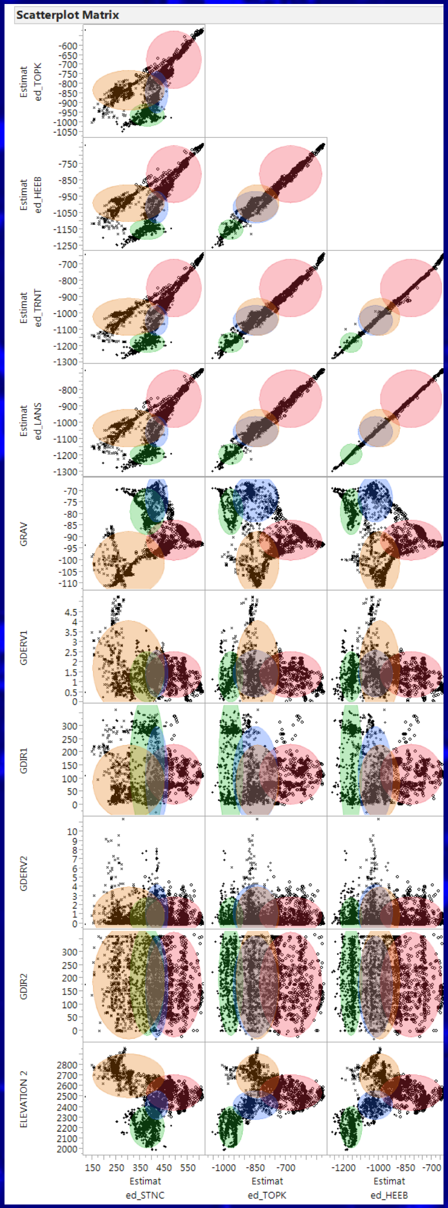
Counties Analyzed Separately

Scale: Each box approx. 1x1 mile

County Overlap Analyzed Separately

Probability (risk) maps of finding a productive Lansing well

The Problem – Data Visual



Subset of cross-plotted variables that show definite data structure (color added for visual clarity)

If cross-plots show non-normal relationships, the data structure must be addressed

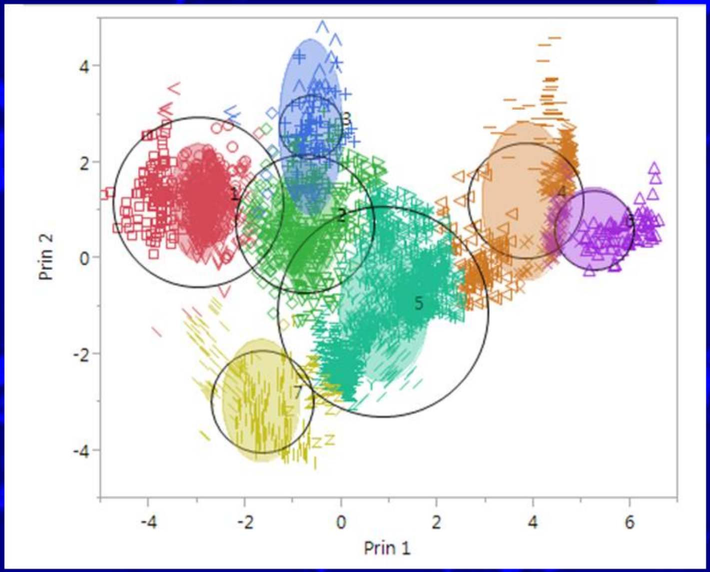
Created using JMP®Pro

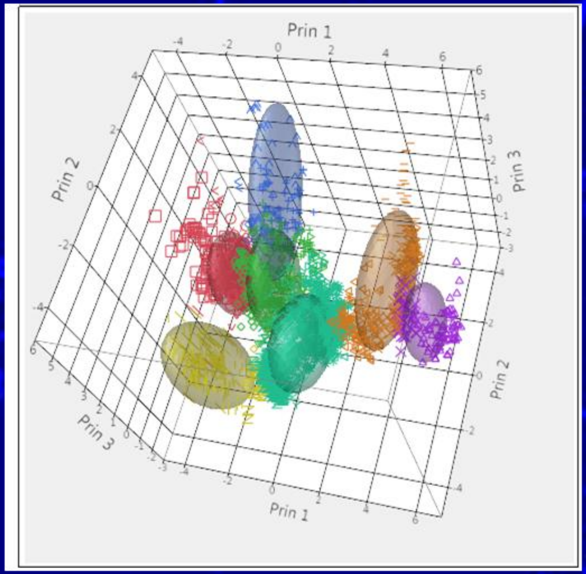
GXSTAT E&D Big Data Solutions

The Problem – Data Visual

Plot of data in Principle Component Space

Created using JMP®Pro





Plot of well data in space defined by first 2 principle components: Note the non-normal data distribution

Plot of well data in space defined by first 3 principle components: Note the multi-planar nature of the data

Color, circles and ellipsoids added for visual enhancement

The Issue – What does it mean?

- The subsurface geology varies across the region (i.e. rocks, geophysics, fluid migration pathways, geologic history, etc.)
- Any defined model must take into account this spatial and temporal variation (i.e. similar zones can be productive for different reasons in a region and multiple models must be created and applied)
- Could also be that the data are of low quality (assumed not to be true as the project has been successful for oil exploration)

Solutions(?): Approaches to Evaluate

All the algorithms will be briefly discussed and all are unbiased unsupervised algorithms. They all address data structure.

- Hard Clustering (k-Means clustering)
  JMP Pro
- Fuzzy (Soft) Clustering (Fuzzy k-Means)
  Revised program from Bezdek, et. al. (1984)
- Fuzzy N-Varieties
  Based on Bezdek (1981)
- Vector Analysis (PVA and HVA)
  Based on multiple sources (see references)

Evaluation Criteria for each Approach

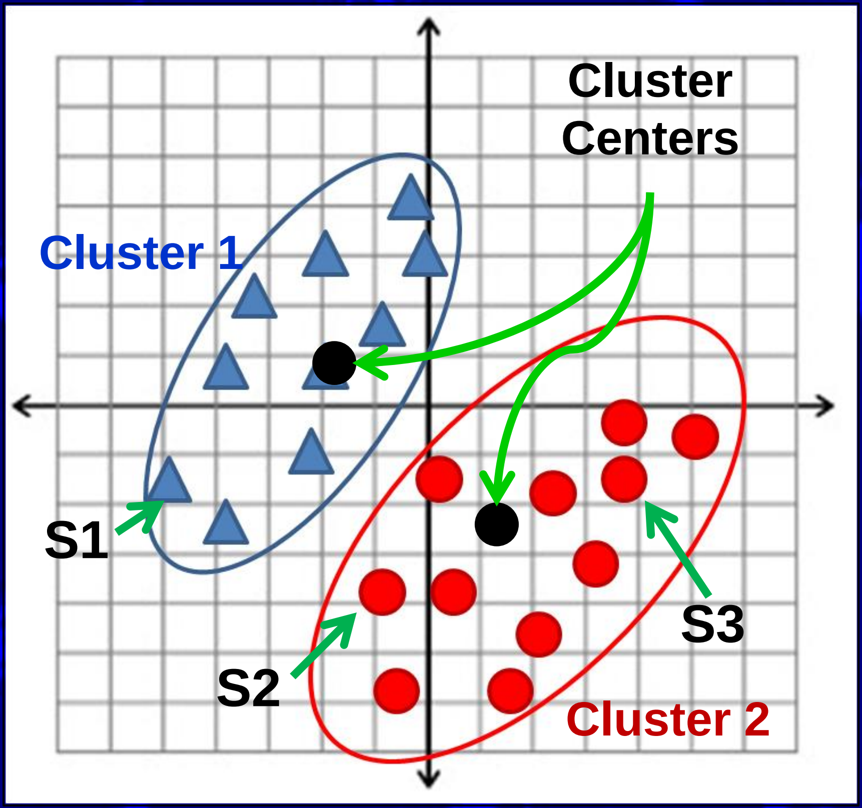
Criteria (statistical) for evaluation of each of these techniques include the following:

- **The percentage of wells that were productive that were identified as productive (goal of this project)**

- **The percentage of wells that were not productive that were identified as non-productive**
  - **Note that the definition of non-productive has changed over the 80+ years of data collection – less so the productive wells**

- **The Relative Model Score = (# correctly identified wells)÷(total number of wells)**

Hard Clustering



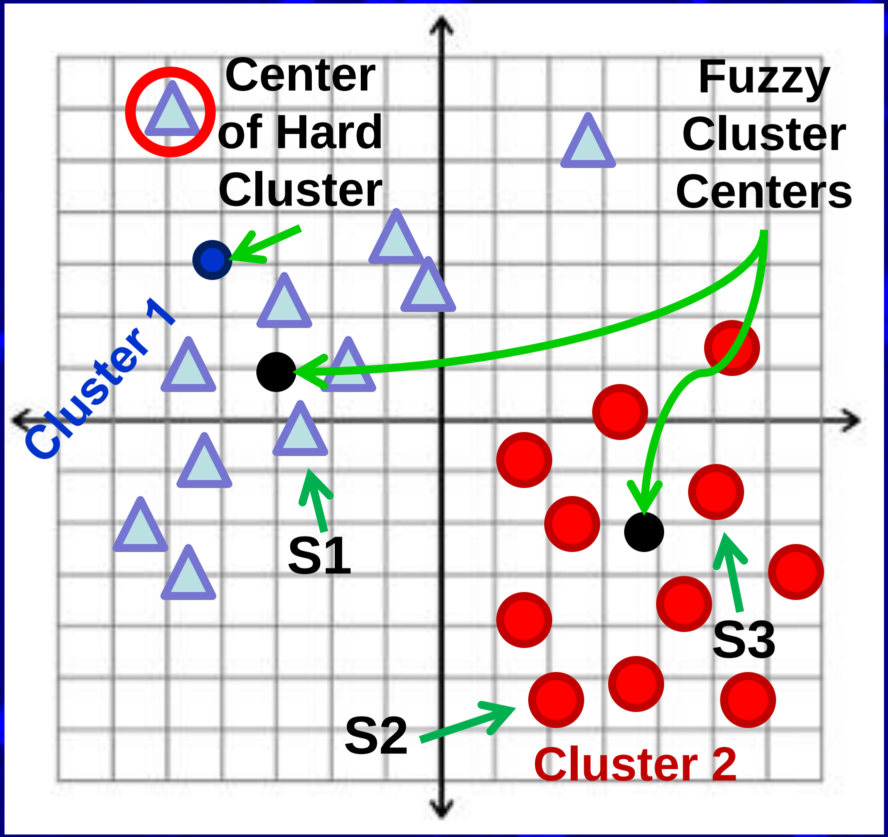
MEMBERSHIP MATRIX		
SAMPLE	Cluster 1	Cluster 2
1	1	0
2	0	1
3	0	1
●●●	1 or 0	1 or 0

Membership function for samples S1, S2 and S3 for hard clustering

Every data point must belong to either Cluster 1 or Cluster 2 based on distance to each cluster center

Note that extreme data can have a major effect on the placement of the cluster centers and it is not easy to determine an extreme point in large dimensional space.

Fuzzy (Soft) Clustering



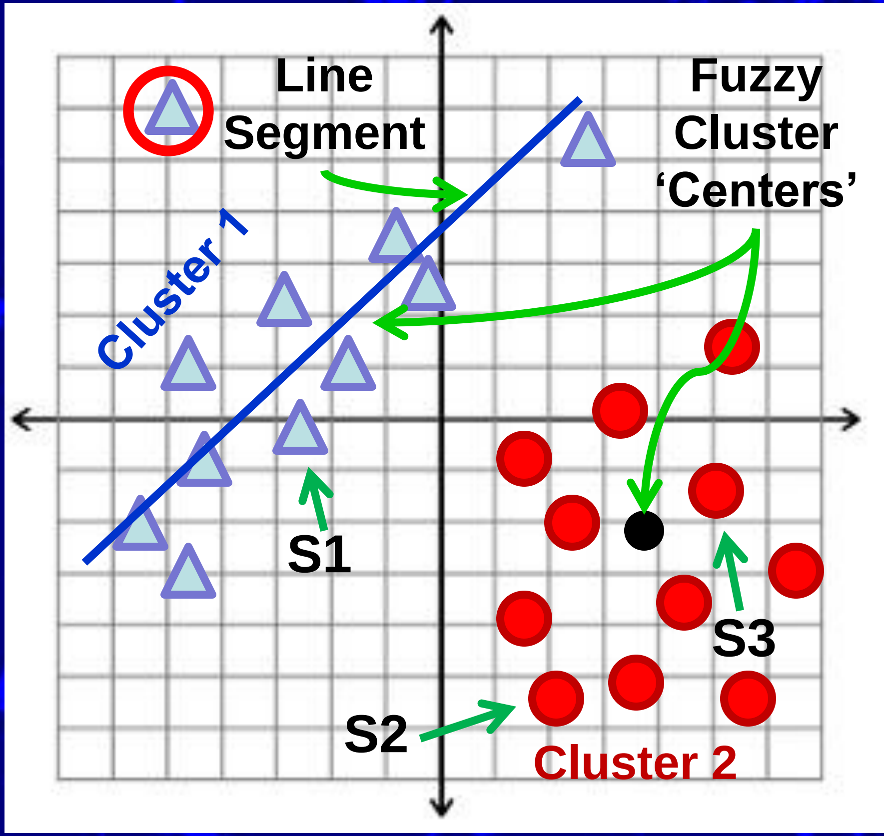
MEMBERSHIP MATRIX		
SAMPLE	Cluster 1	Cluster 2
1	0.95	0.05
2	0.38	0.62
3	0.2	0.8
●●●	[0,1]	[0,1]

Membership function for samples S1, S2 and S3 for hard clustering

Every data point has a ‘membership’ in all the other clusters based on a non-linear distance to each cluster center

Note that extreme data such as the red circled triangle on the upper left have minor effect on the placement of the cluster centers (hard cluster center shown by blue dot)

Fuzzy N-Varieties Clustering



Example in 2D of clustering about a line segment (blue) and a cluster centroid (red)

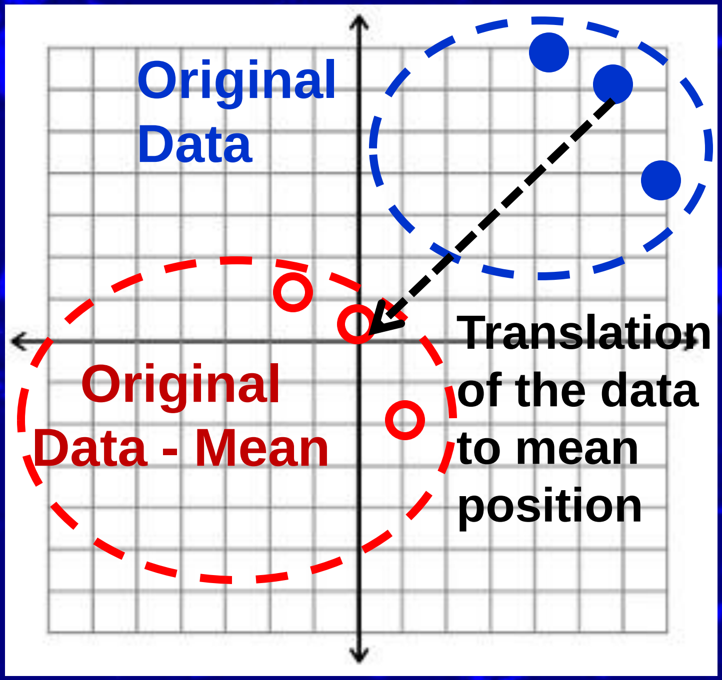
MEMBERSHIP MATRIX		
SAMPLE	Cluster 1	Cluster 2
1	0.98	0.02
2	0.38	0.62
3	0.2	0.8
●●●	[0,1]	[0,1]

Example of a corresponding membership matrix for the data on the left

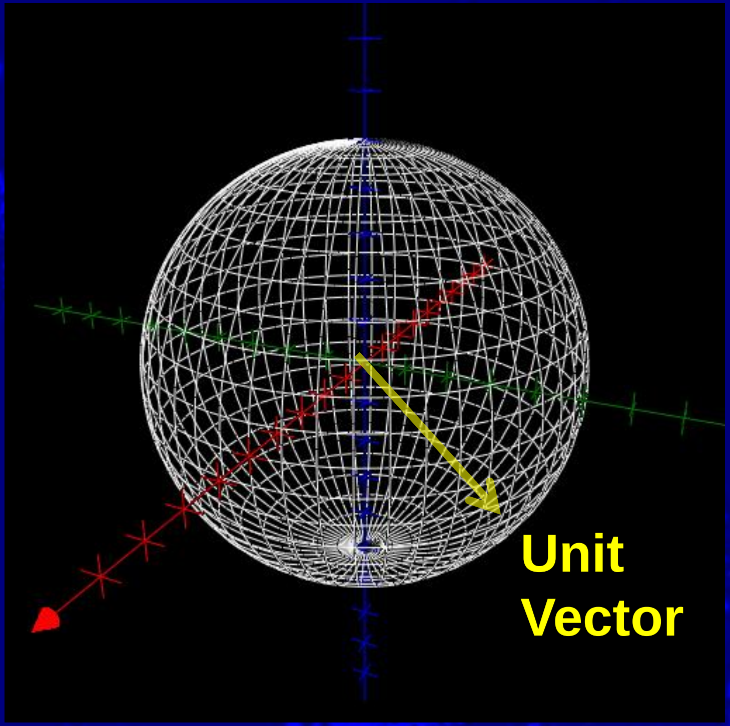
Note that extreme data such as the red circled triangle on the upper left have minor effect on the placement of the line segment

PVA and HVA

Both methods based on <u>vector analysis</u> in unit sphere (hypersphere) as opposed to <u>centroid-based</u> approaches (i.e. PCA, Factor Analysis). The vector analysis approach produces results in the raw measurement units.



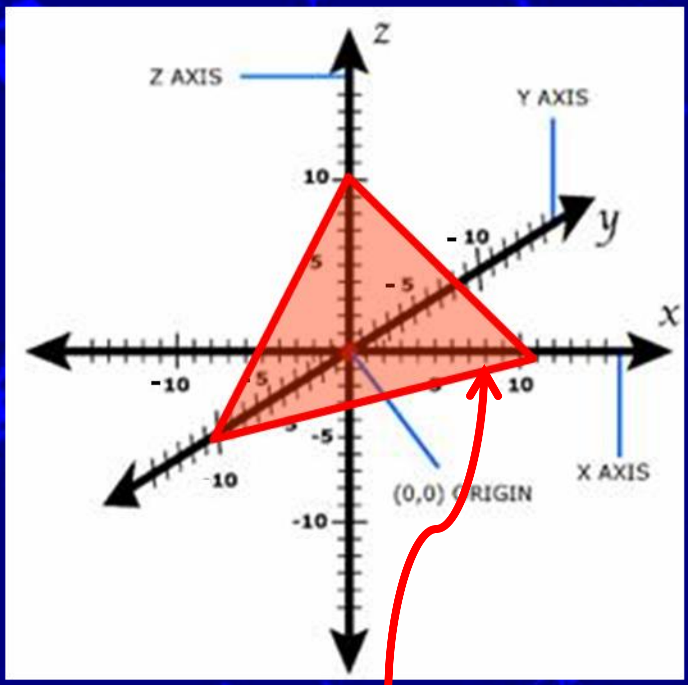
For PCA based techniques, the mean is subtracted from the data and the procedure continues

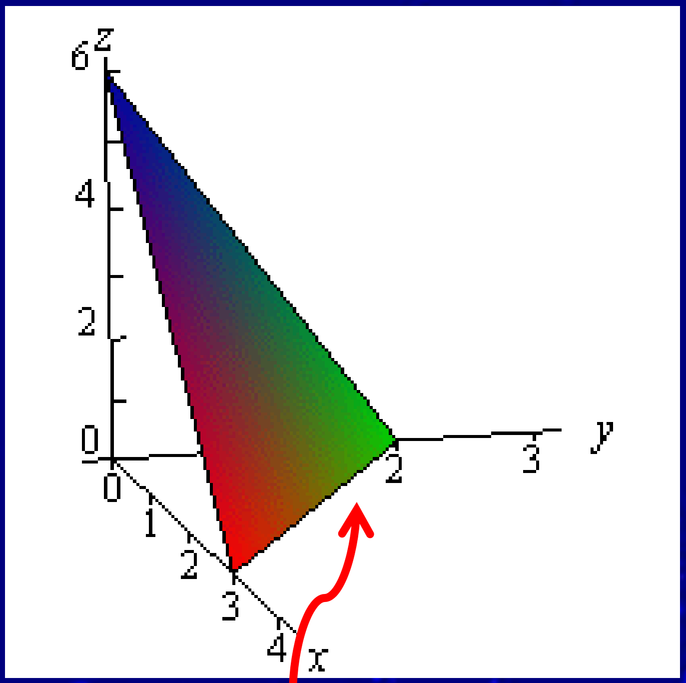


For PVA and HVA, the data are projected into the unit sphere and then the procedure continues

PVA versus HVA

PVA requires row sum to be a constant value (i.e. 100%, 1.0) whereas HVA does not have that requirement. HVA was used in this study.





Constant sum data always fall in a plane (hyperplane) parallel to this plane (hyperplane); examples include % data such as grain size and composition

Non-constant sum data fall on a plane (hyperplane) at an oblique angle to the constant sum plane (hyperplane); examples include seismic, tops, logs.

Data Used in this Study

Data include:

- **Subsurface elevations of key formations**
- **Gravity, magnetic and elevation data**
- **Derivatives (1st and 2nd) of gravity and magnetic data**
- **Directional vector of the derivative gravity and magnetic data**
- **Producing/non-producing**

- ❖ **24 Variables**
- ❖ **Each variable was independently scaled (scaling depended on technique)**
- ❖ **Over 3400 wells in data**

RESULTS

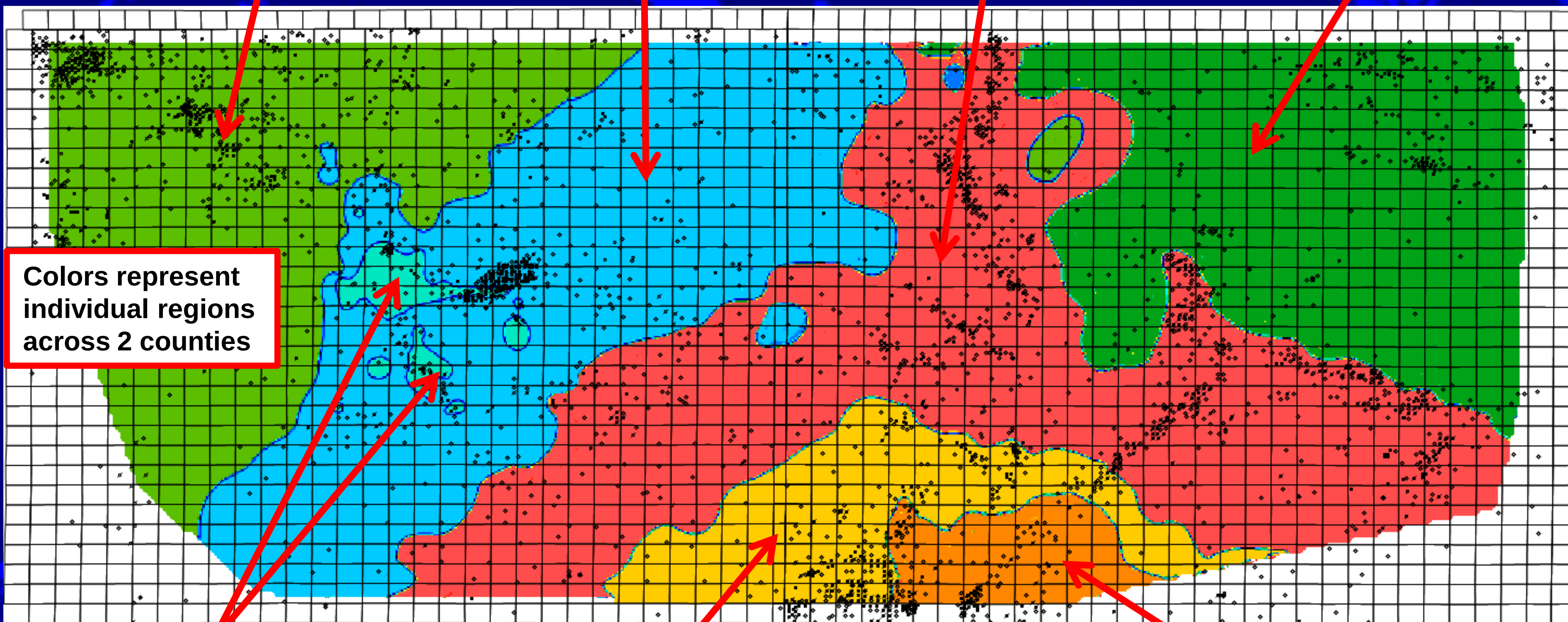
Hard Clustering – Key Variables

1st Derv. Grav.
Stone Corral
2nd Derv. Mag.

Gravity
Magnetic
2nd Derv. Mag.

Gravity
Lansing
Topeka

1st Derv. Grav.
Dir. 1st Derv. Grav.
Magnetic



Dir. 1st Derv. Mag.
2nd Derv. Mag.
1st Derv. Grav.

Lansing
Topeka
Gravity

Base of the Kansas City
Dir. 1st Derv. Mag.
Dir. 2nd Derv. Grav.

Note the role of faulting/fracturing in most areas
Number of Clusters = 7; Criteria: Cubic Cluster Criteria (CCC)

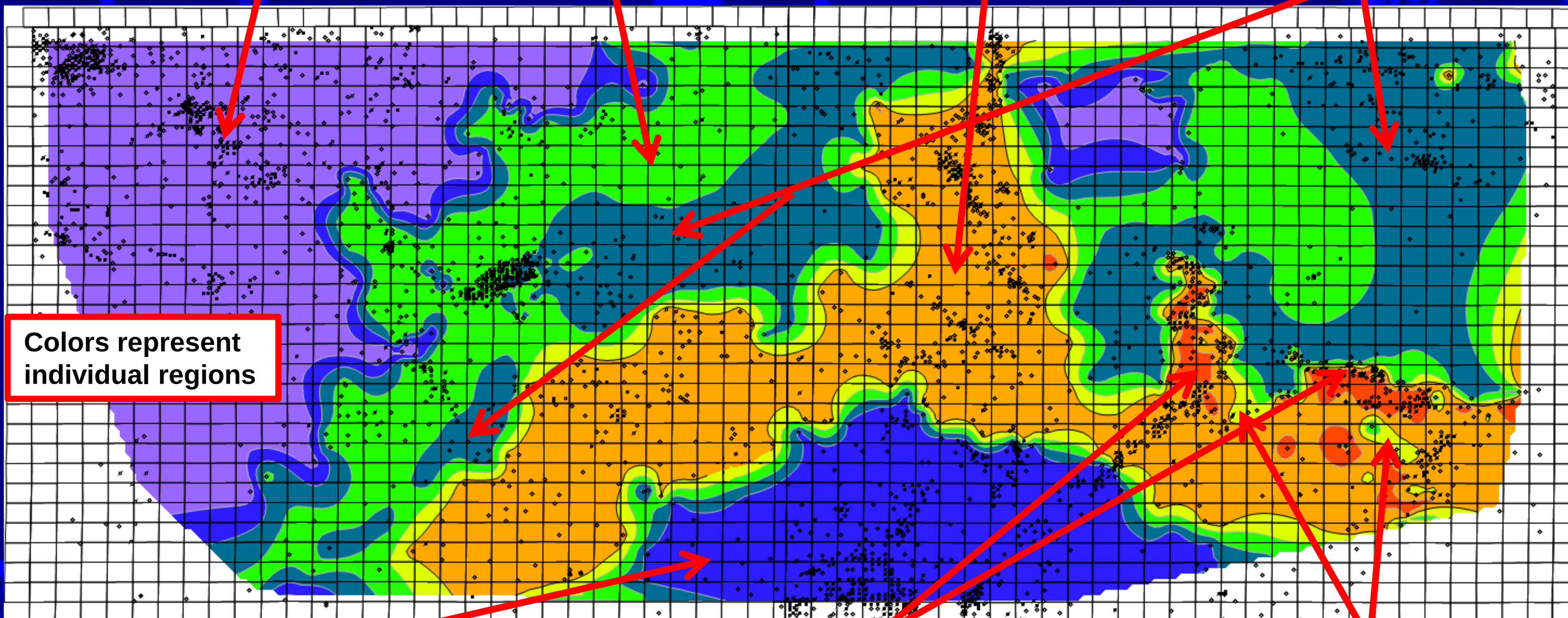
Fuzzy Clustering – Key Variables

1st Derv. Grav.
Heebner
2nd Derv. Mag.

Heebner
Dir. 1st Derv. Mag.
Toronto

Gravity
Lansing
Heebner

Stone Corral
Topeka
Heebner



Base of the Kansas City
Topeka
Toronto

Stone Corral
Topeka
Heebner

2nd Derv. Grav.
Dir. 2nd Derv. Grav
1st Derv. Mag.

Note the role of faulting/fracturing - less than Hard Clustering

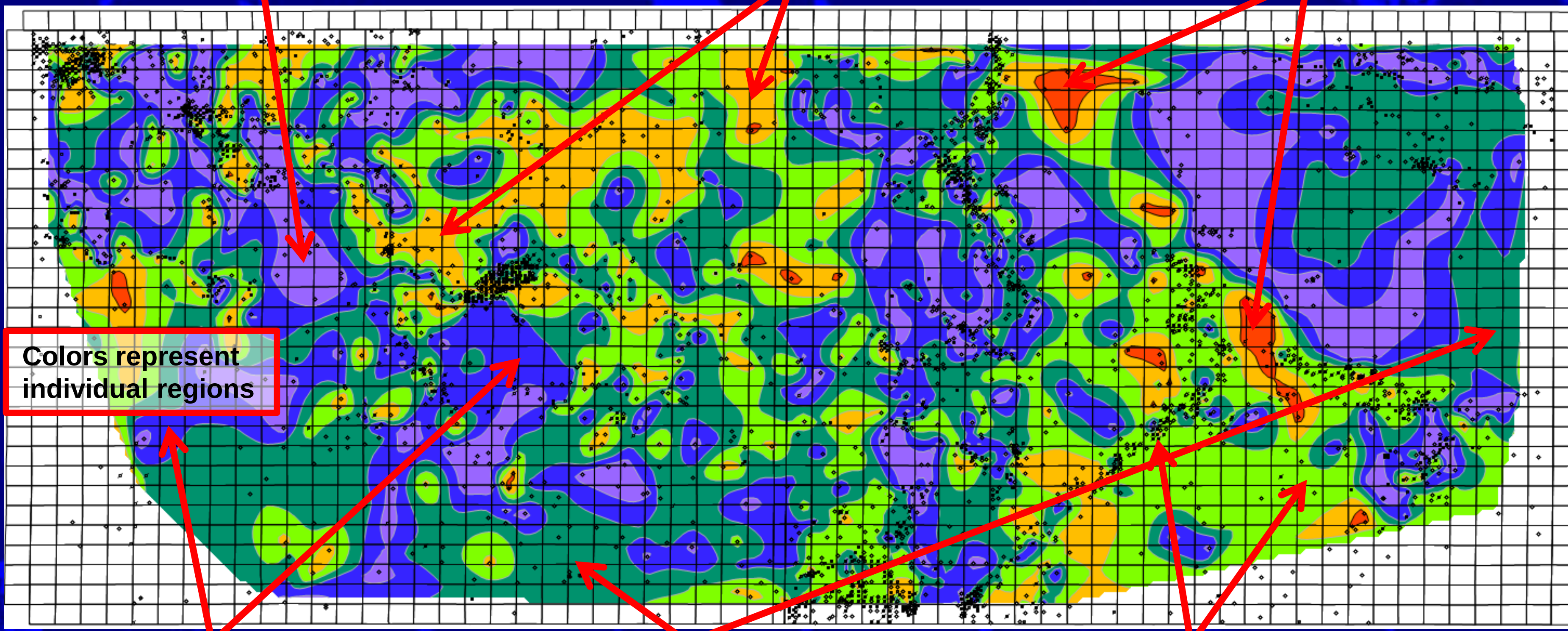
Number of Clusters = 7; Criteria: Entropy, Pseudo-F, Payoff

FNV – Key Variables

Toronto
Heebner
2nd Derv. Grav.

Gravity
1st Derv. Grav.
1st Derv. Mag.

Lansing
1st Derv. Grav.
Gravity



Colors represent individual regions

Gravity
Elevation
Toronto

Gravity
1st Derv. Gravity
Lansing

Lansing
Topeka
Magnetic

Note the role of faulting/fracturing in most areas

 # Clus. = 6; Type: Points, lines, planes; Criteria: Entropy, Pseudo-F, Payoff

GXSTAT E&D Big Data Solutions

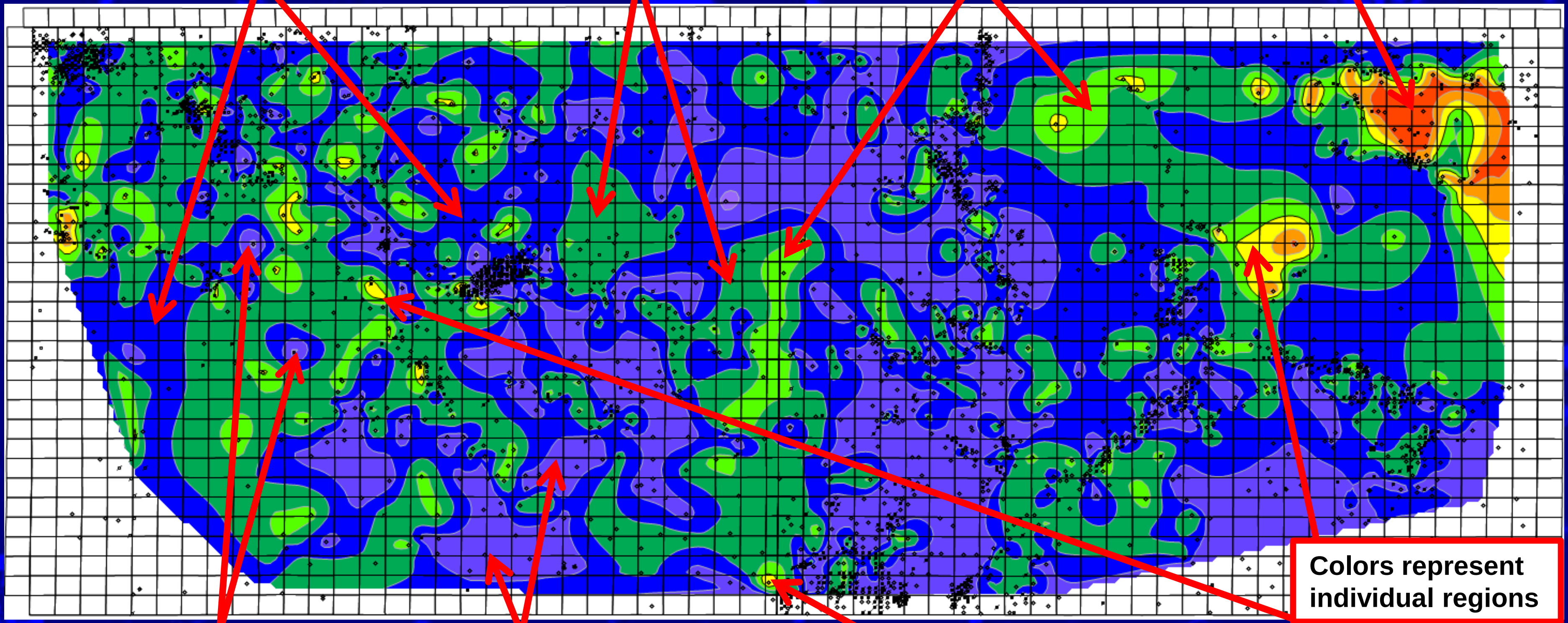
HVA Decomposition – Key Variables

Gravity
1st Derv. Grav.
2nd Derv. Mag.

Gravity
1st Derv. Mag.
1st Derv. Grav.

Gravity
1st Der. Grav.
Elevation

2nd Derv. Mag.
Dir. 1st Derv. Grav.
2nd Derv. Mag.



Colors represent individual regions

Toronto
Heebner
1st Derv. Grav.

Toronto
Gravity
Heebner

Stone Corral
2nd Derv. Mag.
Dir. 1st Derv. Mag.

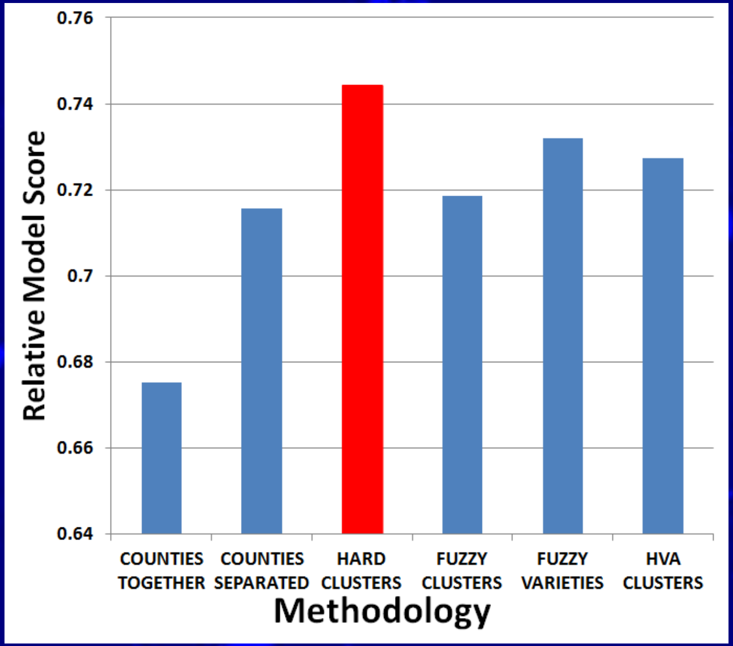
Topeka
Elevation
Dir. 2nd Derv. Grav

Note the role of faulting/fracturing in most areas

Number of 'Clusters' = 8; Criteria: Scree Plot, CD's, Johnson Plots

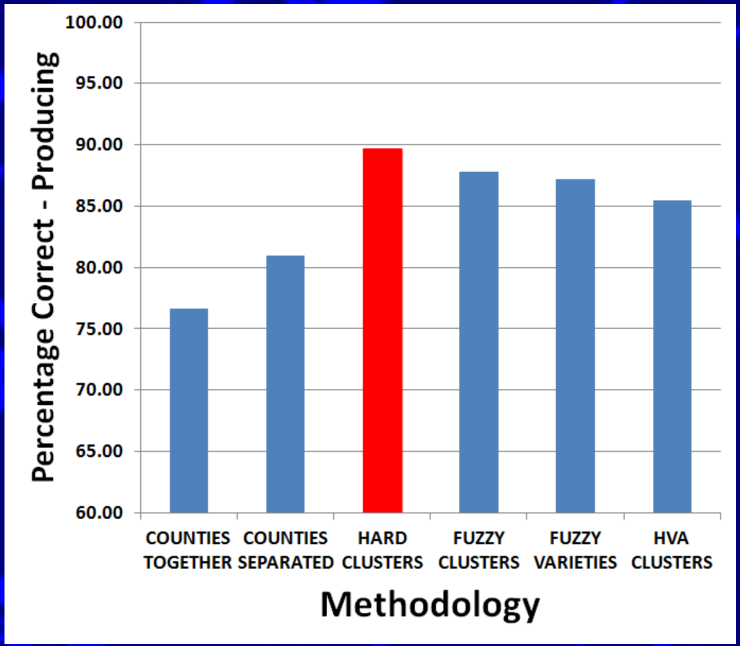
GXSTAT E&D Big Data Solutions

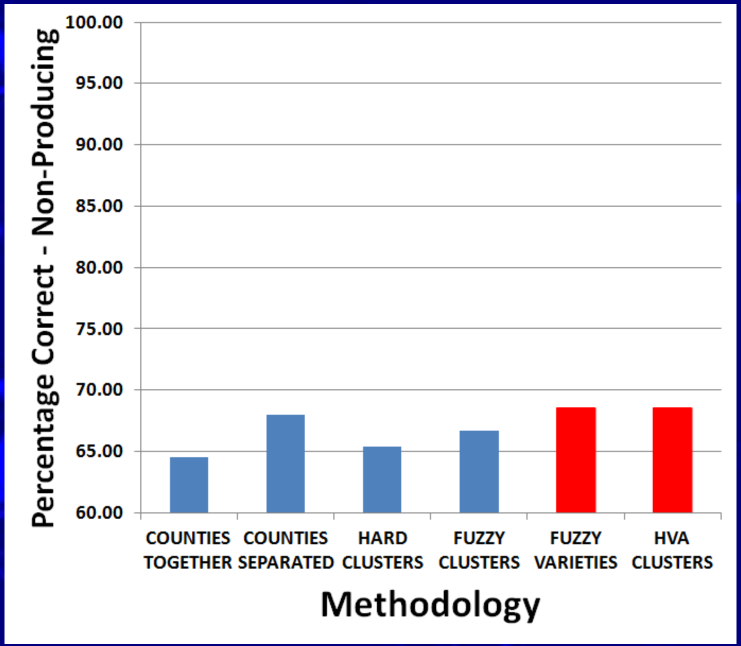
Statistical Results



<u>In terms of identifying both producing and non-producing wells</u>, the data structural decomposition increased the model's ability to identify producing and non-producing wells, thereby reducing future risk. <u>Hard cluster decomposition</u> (red) did the best for this data set (*not necessarily true with other data*).

Statistical Results (con't)

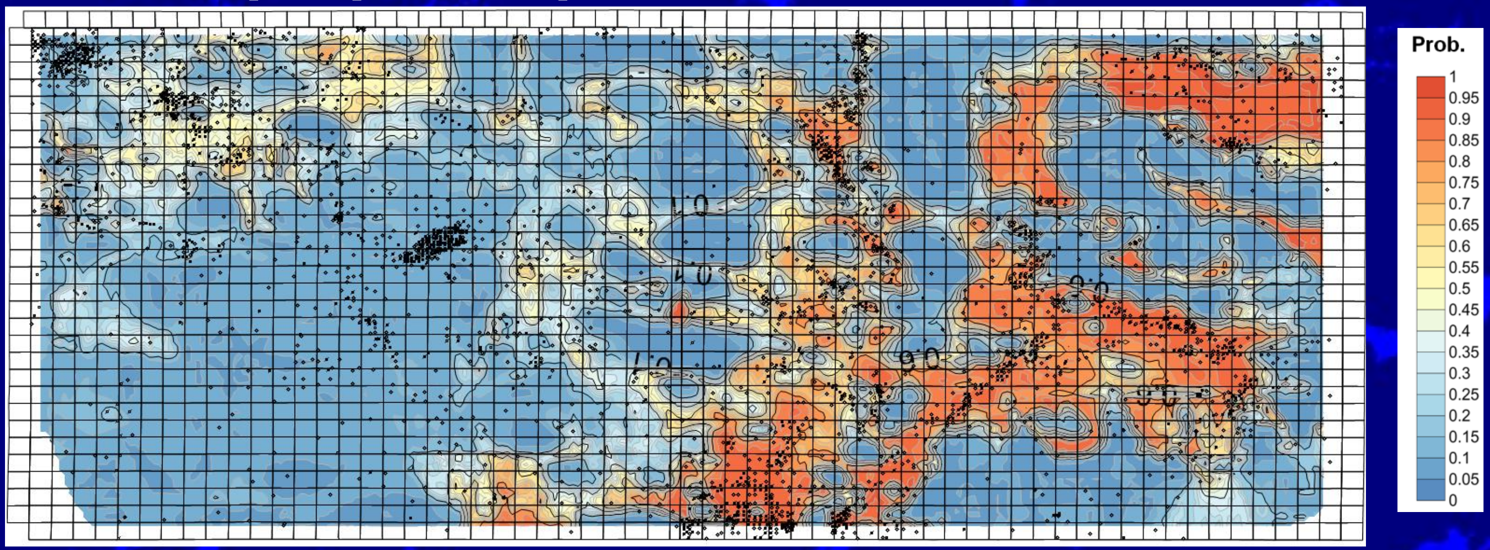




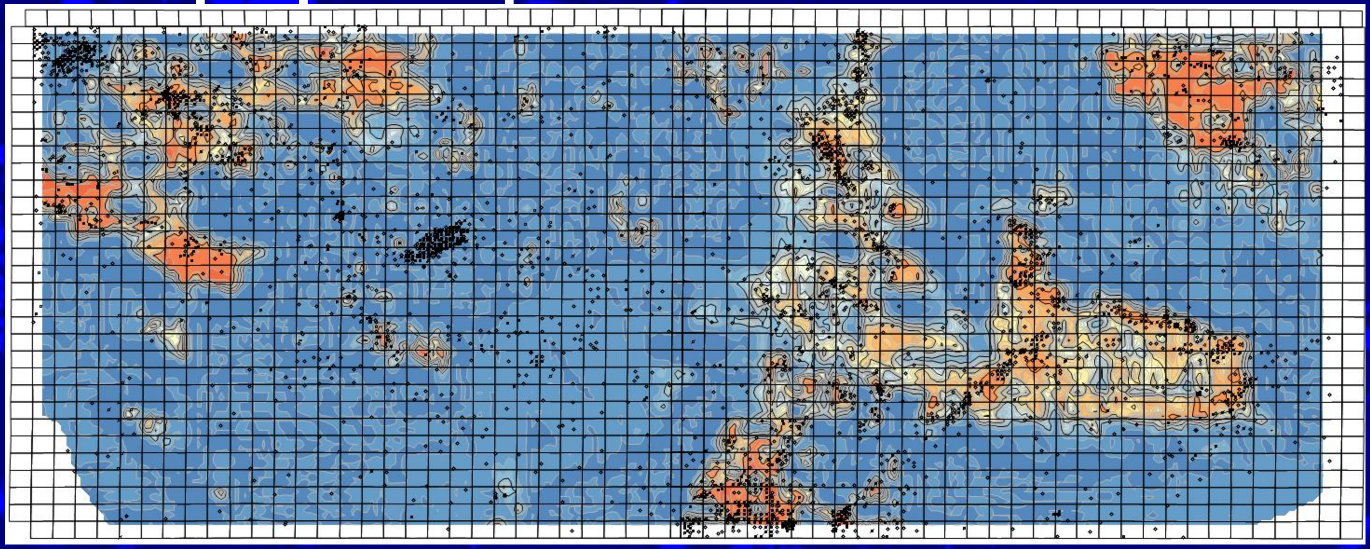
<u>Hard clustering</u> (red on left plot) increased the model's ability to identify <u>producing wells</u>. Non-producing well identification was enhanced by both <u>Fuzzy N Varieties</u> and <u>HVA</u> (red on right plot: a statistical tie). <u>Identification of non-producing wells is a key component for risk analysis.</u>

Before and After

Before proposed procedures:



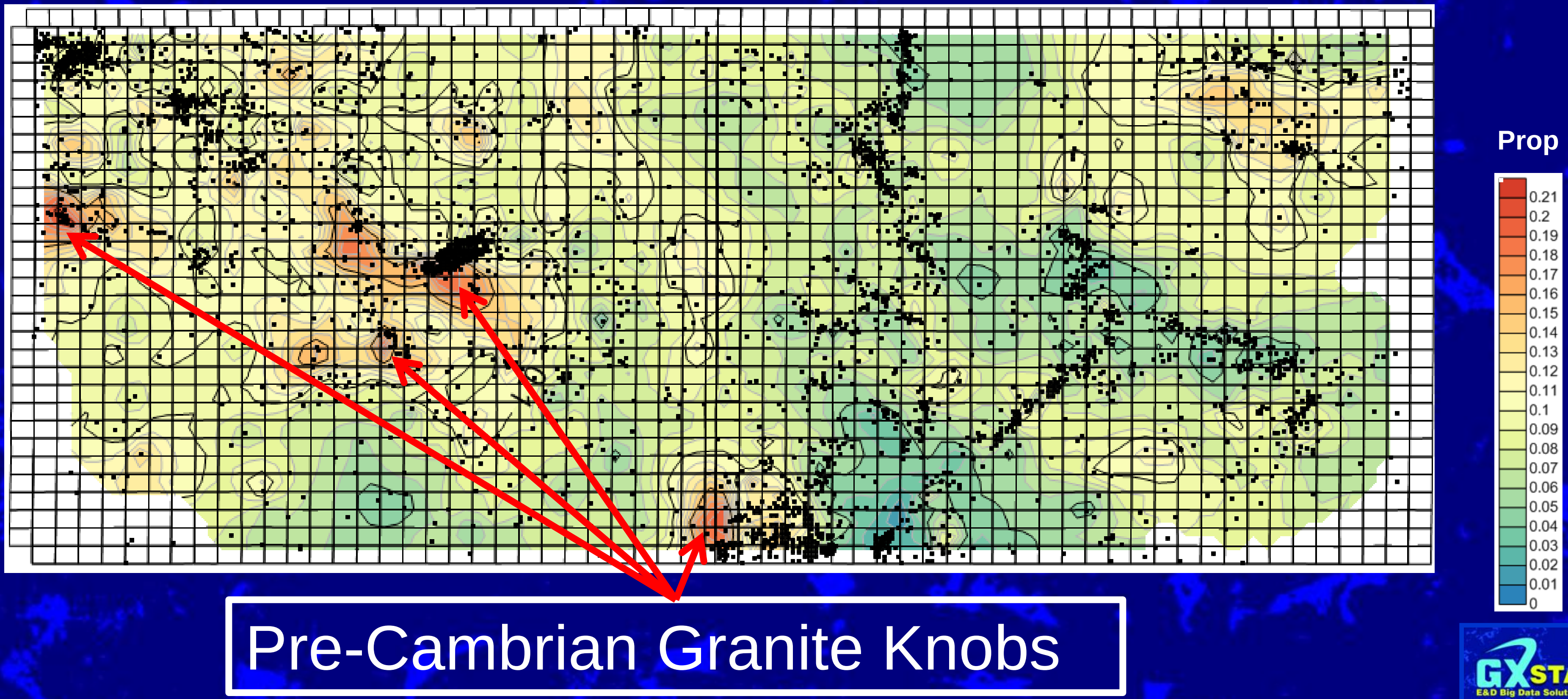
After proposed procedures:



Note the increased resolution of the probability (risk) map after segmentation procedures

Additional Information

The HVA analysis produced mixing proportions for the 8 end members (EM) which were subsequently mapped giving more insight into the geologic history of this area. One example is given below for EM7. Other EMs were tied to block faulting, Laramide fault movements and platform depositional processes.



Conclusions

- The size (geographic area) of a study region is important.
- Every data set needs to be investigated for the presence of data structure.
- The described methodology represents an exciting improvement for big data analysis.

Additionally, important geologic information was gleaned from every step of this decomposition analysis that produced deep insight that is being used for locating economic deposits of oil and gas. This was not fully discussed in this report.

Finally, additional data such as seismic, reservoir quality, petrophysical and similar can be easily added in this analysis.

Thank You!

Bill@GXStat.com
Steve@GXStat.com
Office: (316) 691-5607
www.GXStat.com