

# Exploring New Frontiers in Prediction and Evaluation of Virtualized Well Locations Using Geo-spatial and Geo-statistical Analysis\*

Eduard Breuer<sup>1</sup> and Gregory K. Burns<sup>2</sup>

Search and Discovery Article #41219 (2013)

Posted October 29, 2013

\*Adapted from extended abstract prepared in conjunction with poster presentation at AAPG International Conference and Exhibition, Cartagena, Colombia, September 8-11, 2013, AAPG©2013

<sup>1</sup>Northern Illinois University, Department of Geology and Environmental Geosciences, DeKalb, Illinois ([ed@opengeolog.com](mailto:ed@opengeolog.com))

<sup>2</sup>EP Energy, Houston, Texas ([greg.burns@epenergy.com](mailto:greg.burns@epenergy.com))

## Abstract

Shale play gas reservoirs are a critically important domestic source of clean natural gas. To find reliable well locations in a brand new previously unexplored play, one must examine a plethora of core data from separate text files. Such data contain very detailed information, including mineralogy, rock properties, organic content and much more.

In a new frontier area, where detailed geological interpretations are often missing, a different methodological approach is needed. This approach combines relational database system, geo-statistics, and GIS to find reliable exploration well locations. The utilization of a relational database management system allows for custom data searches, minimizing human error and processing time. Data is grouped by horizons, types of data and its ranges. Each data table follows all the rules to be at least in third normal form. Such data can be easily queried, resulting in a custom dataset.

A more localized study area was selected and parameters were investigated for clustering. Further, geostatistics were used to test for normality and to identify extremes that could skew the results. Relationships between parameters and production were identified. A few expected parameters, such as permeability, indicated that there is no correlation between production and permeability of shales (?).

Data passing the test of normality and having correlation relationships were defined and imported into a GIS system. Each individual geochemical parameter and probable production (IP30) of selected horizons between selected wells were spatially extended into a continuous 3D layer using a spatial krigging method. The outputs were stored in a spatial database to be visually inspected. Further, a mesh of "virtual wells" was laid over the study area.

This technique is called a “fishnet”. Next, all the individual layers were spatially joined. The geologist thus benefits, not only by visually seeing the horizon and the continuously interpolated parameter of interest, but can click on any of the “virtual wells” and instantly obtain a full list of all interpreted geochemical and production parameters.

The power of this method is that it can be built on top of any software. The ability to combine custom queries to provide data for the virtual wells can benefit any sized operation, given reliable data is available. The results of this method were highly correlated to findings by Hammes et al. (2011).

## **Introduction**

Fast growing demand for clean energy requires unlocking new types of resources. For the majority of the petroleum era, shale acted only as a source rock and a cap rock. This view changed in the past decade. Due to introduction of horizontal drilling and hydraulic fracturing, shale has become a source and reservoir rock for natural gas. New unconventional shale frontiers are being explored throughout the world. Since shale has not been the center point of the industry in the past, little is known about the new frontier shale play. New wells need to be drilled, cored and analyzed. A new knowledge base needs to be built for these new resources. Luckily, with the quick advancement of the industry in shale exploration there are new wells and new information available on a frequent basis. One of the most traditional ways of studying the subsurface is by coring. The core data provides a vast amount of petrochemical and mineralogical information. Many exploratory wells also have petrophysical wire line information, such as natural gamma, spectral gamma, porosity, density, full wave sonic and so on. A suite of such high quality and wide spectrum data is critical and expensive to acquire. There will often be more production information from the newly drilled wells than exploratory well core and wire line data. The production will vary from well to well, and sometimes abruptly change. To have the ability to unlock the key well attributes, a complex analysis needs to be performed.

This research is focusing on unlocking well data by utilizing simple methods widely available. Inexpensive off the shelf software and open source alternatives make this methodology even more accessible.

## **Key Components**

The key components used in this work are:

- Data
- Database Management System (DBMS)
- Statistical Package
- Geographic Information System (GIS)

## **Data**

Core and geophysical wire line data traditionally come in a standard LAS format. The advantage of this data type is that it is in a plain text format and it is more easily readable, making it possible to either import into a spreadsheet or clean it up and import straight into a DBMS.

## **DBMS**

The backbone of this work is the use of a relational database management system (DBMS), instead of a spreadsheet approach. The advantage of DBMS is that data is organized in tables. Each table holds a subset of the original data, thus making searching and updating data faster and more reliable. When searching data using a relational DBMS one can perform arithmetic calculations such as minimum, maximum, average and so on automatically and with little effort. These calculated values can be further used in other software. A DBMS will store data separately indexed in tables and only data needed will be logically connected together and searched during a query. This will make the search fast and accurate and will return custom data in various data formats that can be imported into many other types of software or used directly for reports. This makes DBMS a very powerful tool.

## **Geo-statistics**

The role of statistics in a geospatial world of geological exploration is to understand complex data. This critical part is often skipped due to heavy reliance on software capabilities. Understanding how many factors influence the data and being able to find relationships to perform the correct predictions adds a higher level of accuracy to the outcome.

## **GIS**

The spatial interpretation and visualization makes it possible for the data to be observed in the real world and can reveal hidden caveats, such as clustering of the wells. Adding subsurface information such as faults, or other geological features, can give a very different perspective in understanding the petrochemical and petrophysical data. This can be easily achieved by utilizing a Geographic Information System (GIS). A GIS system does not only serve as a tool to see well locations, but also has powerful builtin tools and extended scripting languages allowing the ability to write custom programs that can create 3D interpretations. Complicated spatial operations can be performed on the fly, which saves lot of time and helps avoid introduction of human error into the interpretation. We used GIS to build predictive tools for our virtual wells using Krigging, spatial interpolation techniques and a method known as “fishnet”.

## **Database Management System**

A DBMS provides a platform to obtain custom datasets or sub-datasets in a fast and accurate way. The datasets can serve as inputs to a wide range of processing and analytical software. Usage of DBMS vs. a spreadsheet program has benefits in controlling redundancy, enforcing integrity, avoiding inconsistency, data sharing, enforcing standards, restricting unauthorized access and providing backup and recovery. The cost of development is low, and there is small chance of error and better concurrency control. Data is organized in a relational database in

logically independent tables. Each line of data is called a record. A record could hold a primary key, well ID and individual data entities. Each column is called an attribute. The data in one table may reference similar data in other tables, which maintains the data referential integrity. Operations such as select and join can be performed on these tables.

[Figure 1](#) shows a sample data table called “append\_test\_formation\_names” with attributes “Formation\_ID”, “Depth”, “Formation”, and “API”. The “Formation\_ID” is an auto generated unique key.

The DBMS used in this work is Microsoft Access in the Microsoft Windows environment, which is an inexpensive off-the-shelf standard platform. Original input data provided by Core Lab was in LAS format. Each LAS file was manually separated into individual data tables, such as XRD, Well Location, Mineralogy, IP30, and so on. Each well is identified by API number and depth. At each depth point a unique key was generated by the software, for each record (line of data), to ensure that data is in a normal form. An example of a data query in [Figure 2](#) shows a list of individual data tables of the relational database. Each table has a list of table attributes (columns) such as Depth, API number and Formation. “Formation\_ID” is the unique key of the “append\_test\_formation\_table” data table, that holds core information at a different depth. It is an automatically generated key. The same role is played in the auto generated “XRD\_ID” key in the “XRD\_CLAY\_2” table. The API unique key is used as a primary key in the “Well names” and “Well locations” tables. The difference between these two types of tables demonstrates the unique characteristics of geological data at a depth. Since each table holds a subset of the original LAS file, individual tables of interest need to be connected so that all the customized requested data can be returned by the query. This process is called “join”.

[Figure 2](#) shows individual data tables joined using the “API” and “Depth” attributes as a unique key. This is needed since there is a sample taken at each well at a certain depth. In our work we grouped a certain part of the well by the formation identifier. When the tables are joined, all the attributes in each table are available to be queried. The lines between the boxes represent the joins. Lower part of [Figure 2](#) is a visual built of a query in Microsoft Access. The line “show” visually shows which fields will be selected (with checkmarks) and what operation will be performed on these attributes. For instance “Avg” will return an average value Chlorite, Mixed Chlorite for each well. In a SQL language “where” is a logical operator that will select only values where “BSSR” or “BSSRGZ” is true (not show in the figure). We also returned average values for “Chlorite” and “Mixed\_CL1” attributes in the “XRD\_CLAY\_2” table. Many custom queries can be built, saved and rerun in the future when using this particular DBMS system. Every time the query is run (or re-run) fresh data is returned. The practicality here is that if data is updated, new well data added or some incorrect data purged from database, new data output can be acquired virtually within seconds. A database can grow periodically without too much effort or risk of data contamination due to human error. New tables can be created or added, new joins can be built, and new queries can be performed as quickly as needed.

## Geo-statistics

### Testing Normality

Most of the interpretative software used by the industry runs its interpretations, interpolations, and predictions based on some assumption without really testing these assumptions. The most common assumption is that data used falls into a normal Gaussian distribution. Normality is a most common prerequisite for many spatial interpolations, such as Krigging. Therefore, it is one of the most important questions to ask and

answer. There are two tests for normality that can be run. For a dataset smaller than 2000 elements, the Shapiro-Wilks test can be used. For large datasets the Kolmogorov-Smirnov test can be used. In our case, since the sub-dataset was already large, we used the Kolmogorov-Smirnov test. We used a confidence level of 95%, thus the SIG must be greater than 0.05 to accept the hypothesis.

After we obtained a specific dataset from a custom database query, we inspected the data for possible clustering effects using the GIS. For all the work in this article we used SPSS software, which is widely accepted both commercially and by academia. We ran both the Shapiro-Wilks and the Kolmogorov-Smirnov tests to compare the results. Due to the data size, the Kolmogorov-Smirnov test is a better fit. [Figure 3](#) in this example suggests that IP30 production data is normally distributed. Other attributes, such as the Total Organic content (TOC) and the Chlorite values tested to be normally distributed as well. The same seems to be true for TOC in the BSSR Formation. However, Chlorite in BSSR Formation and Illite in Haynesville Formation are not normally distributed.

It is important to visually inspect the histogram ([Figure 4a](#)) and either a P-P or a Q-Q plot ([Figure 4b](#)). The histogram can tell a lot about the data. The histogram can be right or left skewed, bimodal, plateau, edge peak and many others. Visually inspecting the histogram we learn a lot about the data instantly. The normal plot is designed specifically for judging normality. The center line shows the ideal normal distribution with mean and standard-deviation of the sample. If the points roughly follow the line, the sample has normal distribution. The probability P-P plots are better to spot non-normality around the mean, and normal quantile Q-Q plots to spot non-normality in the tails. The idea here is that the data are plotted against a theoretical normal distribution, in such a way that the points should form an approximately straight line. The Q-Q plot example from our data set in [Figure 5b](#) suggest that IP30 data in the Haynesville Formation are not normally distributed. However, [Figure 5a](#) suggests that distribution of Chlorite in the BSSR Formation is bimodal.

### **Choosing Between Parametric and Nonparametric Methods**

A parametric or a nonparametric test can be used based on what type of distribution the data has, the type of data and how large the dataset is. If the data is of a ratio or interval type, a parametric test is chosen. A non-parametric test is selected if the data is of ordinal or nominal type. The most common parametric assumption is that data are approximately normally distributed. Nonparametric tests do not rely on assumptions about the shape or parameters of the underlying population distribution. If the data deviate strongly from the assumptions of a parametric procedure, using the parametric procedure could lead to incorrect conclusions. As mentioned previously, the parametric assumption of normality is particularly worrisome for small sample sizes ( $n < 30$ ). Non-parametric tests are often a good option for this data. Non-parametric procedures generally have less power for the same sample size than the corresponding parametric procedure if the data truly are normal. Since often we have data like in the BSSR Formation case, which does not have normal behavior, we may want to use nonparametric methods. Nonparametric tests do not deal with parameters of the populations (such as means). They estimate percentiles of a distribution without defining the shape of the distribution by the means of parameters. Nonparametric tests are simple and often they are the only ways of statistical estimations. For the majority of this work our attention is going to be focused on the normally distributed data since our data falls into this category.

## One Sample t-Test

A one-sample t-test is used to compare the mean value of a sample with a constant value; our value of interest, denoted as  $\mu_0$ . The test has the null hypothesis that the population mean is equal to  $\mu_0$  and that the alternative hypothesis is not equal to  $\mu_0$ . Similarly, a one sided test can be used to determine if the constant value of interest is either smaller or larger than the mean value of the population. For completeness, there are two tests, a t-test and a z-test. A z-test is used when population distribution is normal or sample size is large and the population standard deviation is known. The t-test requires population distributions to be normal, but is robust with departures from normality. As the sample size increases, the t-distribution approaches a normal distribution. Therefore, we used the t-test for our study. The test was used to answer questions such as “will a well drilled in our study area produce well enough to be profitable?”. What the test really does is testing if an “average” well meets the criteria of a profitable well. We arbitrarily set a value of 12 MMCF/day for the t-test. [Figure 6](#) shows that SIG is less than 0.05, which means that an average well in the study area does not produce 12 MMCF/day. In this case the null hypothesis was rejected.

## Correlation - Looking for Relationships

In order to predict values one needs to test if there is a relationship between the attributes. To find relations in statistics, correlation is used. The null hypothesis states: “there is no relationship” and the alternative that “there is a relationship”. There are two basic tests available. The Person (P) and Spearman (S) correlation. For normally distributed data, the Pearson is best used. Pearson benchmarks linear relationships, whereas Spearman benchmarks monotonic relationships. Spearman is computed on ranks and so depicts monotonic relationships, while P is on true values and depicts linear relationships. Pearson is most appropriate for measurements taken from an interval scale, while the Spearman is more appropriate for measurements taken from ordinal scales. Since our data is interval data, population is large enough and the data is normally distributed, we used the Person correlation test. The table in [Figure 7a](#) shows clearly that there is a strong relationship between IP30 production data and TOC in the Haynesville Formation. Looking at a scatter plot graph in [Figure 7b](#) one could draw a line through the plotted data. We came across an interesting result when testing correlation between permeability and IP30 production. The test indicated no relationship, where  $H_0$  was accepted ([Figure 7c](#)). Inspecting the scatter plot in [Figure 7d](#) clearly stated that there must be something wrong with the data and this suggested that the permeability should not be used for predictions or other calculations in this work.

## Regression - Predictions

We used linear regression for this work, which attempts to model the relationship between two variables, by fitting a regression line equation (1) to observed data. Where “a” is the intercept (value of Y when X=0) and “b” is the slope (change in Y per unit increase in X). We used SPSS for our work. [Figure 8a](#) shows the parameters necessary for building the prediction equation (2) for TOC:

$$y = bx + a \quad (1)$$

$$y = 18.492 + 8.583 * \text{TOC} \quad (2)$$

Similarly, equations were built for other attributes of interest. [Figure 8b](#) shows the best fit line through the TOC data and 95% confidence lines. Since clusters of data may change the trend of the equation, we checked for residuals and influential outliers, by visually inspecting all the scatter plots used in our work. [Figure 8b](#) shows a regression scatter plot for TOC and IP30 defined by equation (2).

## Well Core and Production Parameters Prediction in GIS

### GIS Background

A Geographic Information System (GIS) is not just a software, it is a collection of tools that allows us to visualize, interpret, analyze, manipulate and store data, and find relationships, trends, patterns and much more. Use of a GIS has extended from desktops, through networks to clouds. Most popular GIS variants are ArcGIS, and GeoMedia for commercial platforms. The open-source software expands into the multi-operating system (OS) world, including but not limited to Microsoft Windows, Linux and a variety of other specialized OS's. To name a few that can be used both on Microsoft Windows and Linux: Quantum GIS (QGIS) and a much older raster version of GIS called GRASS. For the purpose of this study we used the most common and well accepted by the industry, ArcGIS under Microsoft Windows OS.

### From Data to Virtual Wells

As input we have data that was custom queried with the necessary attributes, statistically tested and normally distributed. To plot the data it is very important to inspect for possible clustering effects during interpolation. Data in [Figure 9a](#) was selected to avoid clustering of points and far outliers to prevent effects of influential observations. The data between each of the wells were spatially continuously interpolated by using a Krigging method for each parameter of the entire selected depth interval. For each depth interval we first created continuous 2D layers that had predicted values for the given attribute everywhere using Krigging. We repeated this process for all the attributes at all depths. This gave us a large number of layers at the same depth which would hold values for different attributes. These layers need to be spatially joined. This gave us a single layer holding all attributes at depth. This process was repeated until we satisfied all the depth ranges we had interest in. This process can be done manually in the GIS or can be automated by writing a script using the Python programming language or a software extension already prepared by the vendor. We used an ArcGIS extension called Premiere Power Attribute Maker (v5.0) that is widely accepted by the oil industry.

To predict petrochemical and petrophysical attributes of a well in unexplored areas we decided to create “virtual wells”. The virtual wells are simply center point nodes of polygons. The locations of these wells can be at either predetermined regularly spaced grid cells, or the nodes can be also randomly distributed in a predetermined area. There are several ways to achieve this, depending on the GIS package used or the version of ArcGIS used. For an older version of ArcGIS, the process would consist of several steps, but can be achieved. We used an ArcGIS version 9.3, which has a built-in tool in the data management “tool box”. This tool is called “Fishnet”. Fishnet has the ability to create nodes which we call “virtual wells”. These point features will hold for each virtual well all the data in a database for all the points where the fishnet was defined. This information can be obtained by using the “info” tool and selecting or clicking on any of the nodes or virtual wells. The brown diamonds in [Figure 9b](#) represent virtual wells. The green lines are boundaries of polygons that these wells are a center point of. The virtual wells do not need to be spread in an equal distance grid. The software can generate a random grid of any  $n$  wells in any defined area. The green

squares with check marks are existing wells. The larger green circle in the upper left corner of the grid in [Figure 9b](#) is a selected virtual well. The info table holding all the data for this point is displayed on the right. This table will provide a list of records for all the attributes that were predicted for this well. It is interesting to notice that this table holds records of predicted estimated core values for multiple formations. In this study we worked on the Haynesville and BSSR formations. GIS predicted petrochemical and estimated IP30 production values at this location. The GIS proved to be a very effective and powerful tool for the prediction process. It is especially useful since GIS can import other data that can be overlaid on top, such as roads, property boundaries, owner data, and so on.

### **Summary - Benefits and Conclusions**

Our method proved to be quick, efficient, cost effective, scalable, and expandable. Since most of the core petrochemical and petrophysical wire line data is supplied in easily readable LAS files, the relational database system proved to be a very versatile system. The most crucial part was the first step: the design. We learned that due to the unique nature of the data used in our field it is necessary to have a unique key generated by the DBMS. This was probably the most important step of the design. We kept our databases in third normal form, where:

1st Normal Form - No Repeating Elements

2nd Normal Form - No Partial Dependencies

3rd Normal Form - No Dependencies on Non-Key Attributes

We found that adding new core data into the database was very quick and accurate. We were able to extract very specific information with an added benefit of being able to perform arithmetic calculations on records for selected attributes. We specified that we desire average TOC, Chlorite, Illite and IP30 data for the wells which are in Haynesville and BSSR formations. We also targeted specific parts of formations, such as the upper section of the Haynesville Formation.

Since Microsoft Access is commercial software there was a small cost associated with it. Microsoft Access is well documented, easily accessible in almost any country and is available in most world languages. Recently we have been testing open-source DBMS such as SQLite. This is an open-source free of charge software. There is no installation needed, it is self-contained, executable and uses standard SQL language. This software can be found in a majority of cell phones running text messages, managing all the contacts, and much more. The software is well documented and there are ports to use it with GIS software such as QGIS to import data directly. SQLite used in Linux OS can be combined with scripting languages such as BASH which makes use of AWK and SED. The two build in line editors have the power to process the new LAS file in the background and supply the data back to the DBMS. This process can be very interactive, simple and fast. All these components are very well documented.

Geostatistical testing gave us many clues of which data in fact belonged to the same geological environment and where the deposition environment changes, without the need to know much about the geology or the processes that formed these environments. We were able to study effects of clustering, its influence on statistical weight and error in predictive interpolation. Statistics also helped to identify the spatial extend to which the data behaved normally and where we started to see influence and change of attributes. Since in geology we work in the real world, statistics can help to identify different depositional environments by simply inspecting histograms, testing for normality, and being able



to visually identify bimodal behavior. When our initial work was done, our study was in the frontier with no articles or studies published yet on the topic. Later an article Hammes et al. (2011) appeared to be based on a much larger dataset than the data set we worked with. However, our results are in excellent correlation with their findings. The objective of our work was to be able to maneuver the frontier and understand the subsurface behavior without being able to know the story of the geology in great detail. We believe we were able to predict the necessary key parameters that these depositional environments would have characteristics for. We were able to predict with confidence what the production could be in areas where we had no data or prior work experience. Our data seems to be again in good correlation with the published data Hammes et al. (2011). We were also able to identify a parameter that might have been influenced by a lab measurement technique used by the industry at the time. The method used for measuring permeability by labs was flawed, which was later discovered by the industry and that method is no longer used. We were able to identify this from our analysis before this flaw was discovered by the industry. We assessed the quality of spatial prediction by testing for correlations and regression. Once we knew our data well we could move on to the phase of spatial prediction and building our predictive virtual well network.

GIS proved to be a very powerful and versatile tool. Using GIS at the time when this area was a frontier was essential to identify and verify possible core parameters and production rates in the areas of interest, where land leases were available. Virtual wells helped to identify the best production areas and steer the focus of the effort to obtain leases in that area. Since the GIS was able to aid production prediction it also could help to provide information on the right price for the land leases in the areas of interest. The virtual wells method can also provide a full suite of predicted attributes for geotechnical, petrochemical, petrophysical, and any other type of data entered and queried out of the relational database. Having this tool does provide a wide range of information to various parts of the exploration team. The petrochemical and core data helped to identify areas with different mineralogy and maturation of which is essential knowledge for frontier exploration areas. We believe that utilization of a GIS needs to be based on a good knowledge of the data. Once that criteria is satisfied in combination with the DBMS and perhaps aided by Python scripting used to automate the laborious part of the process, GIS can provide results in minutes. The speed of this process gives an opportunity to ask new questions and to speed up the exploration with a higher level of confidence in attaining high quality results. Currently, there is additional work in development to have this process to be based on 100% open-source, which would make the method very affordable and even more widely accessible.

The power of this method is that it can be built on top of any software. The ability to combine custom queries to provide data for the virtual wells can benefit any size operation, given reliable data is available.

### **References Cited**

- Hammes, U., H.S. Hamlin, and T.E. Ewing, 2011, Geologic analysis of the Upper Jurassic Haynesville shale in East Texas and west Louisiana: AAPG Bulletin, v. 95/10, p. 1643-1666.
- Sinha, S., E.M. Braun, Q.R. Passey, S.A. Leonardi, A.C. Wood III, T. Zirkle, J.A. Boros, and R. Kudva, 2012, Advances in measurement standards and flow properties measurements for tight rocks such as shales: SPE/EAGE European Unconventional Resources Conference and Exhibition, SPE 152257-MS.

append_test_formation_names			
Formation_ID	Depth	Formation	API
1	10742	BSSR	42365371690000
2	10750	BSSR	42365371690000
3	10758	BSSR	42365371690000
4	10766	BSSR	42365371690000
5	10774	BSSR	42365371690000
6	10782	BSSR	42365371690000
7	10790	BSSR	42365371690000
8	10798	BSSR	42365371690000
9	10806.5	BSSR	42365371690000
10	10814	BSSR	42365371690000
11	10863	BSSR	42365371690000
12	10870.5	BSSR	42365371690000
13	10871	BSSR	42365371690000
14	10878.5	BSSR	42365371690000
15	10893.5	HSVLU	42365371690000
16	10901	HSVLU	42365371690000

Figure 1. Microsoft Access sample data table.

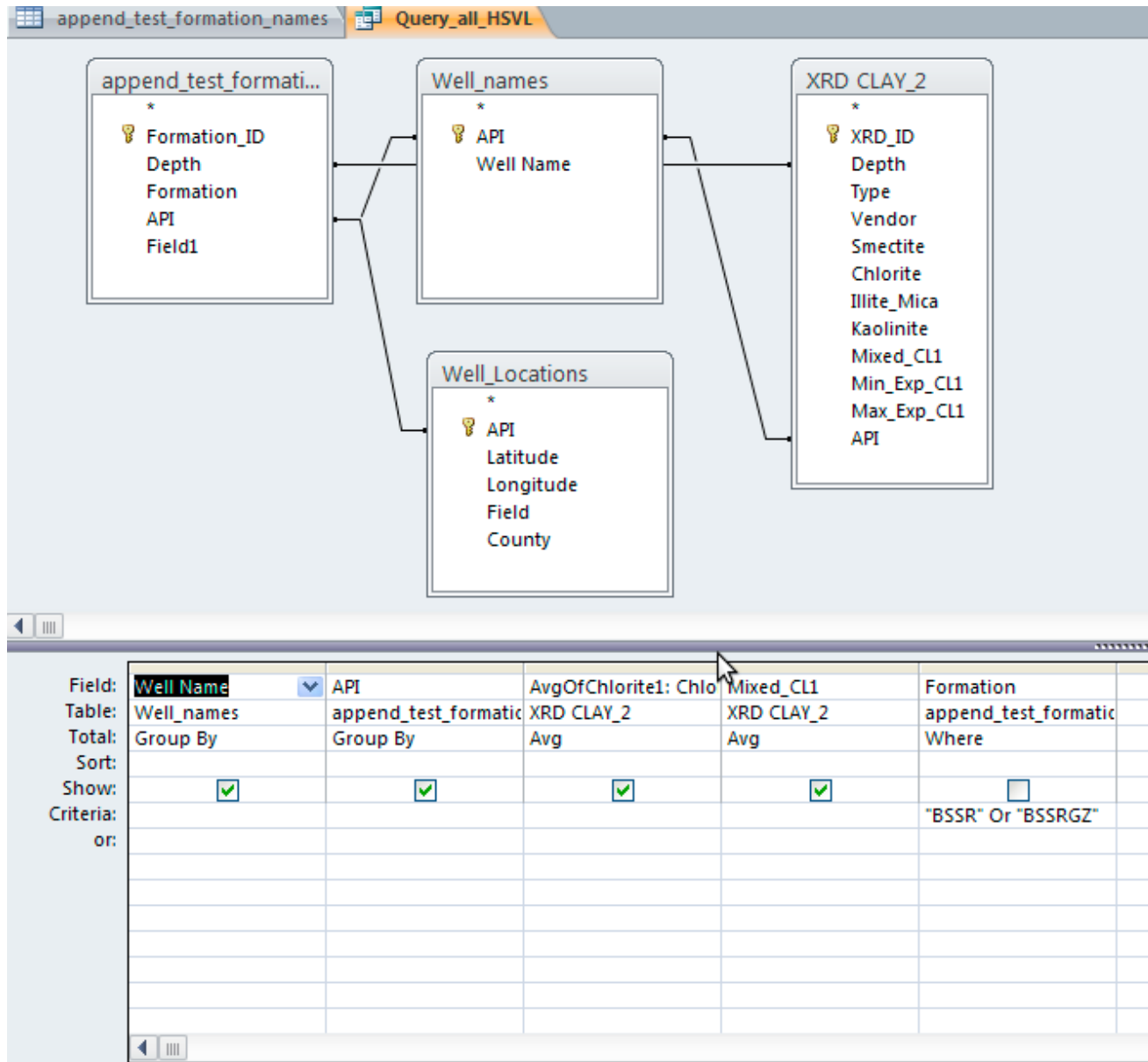


Figure 2. Data table join, sample query.

	Kolmogorov-Smirnov <sup>a</sup>		
	Statistic	df	Sig.
ip30_mmcfp	.076	94	.200
hvl_cl	.069	94	.200
hvl_IL	.114	94	.004
hvl_TOC	.086	94	.085
bss_cl	.176	94	.000
bss_il	.113	94	.005
bss_toc	.081	94	.155

Figure 3. SPSS K-S normality test results.

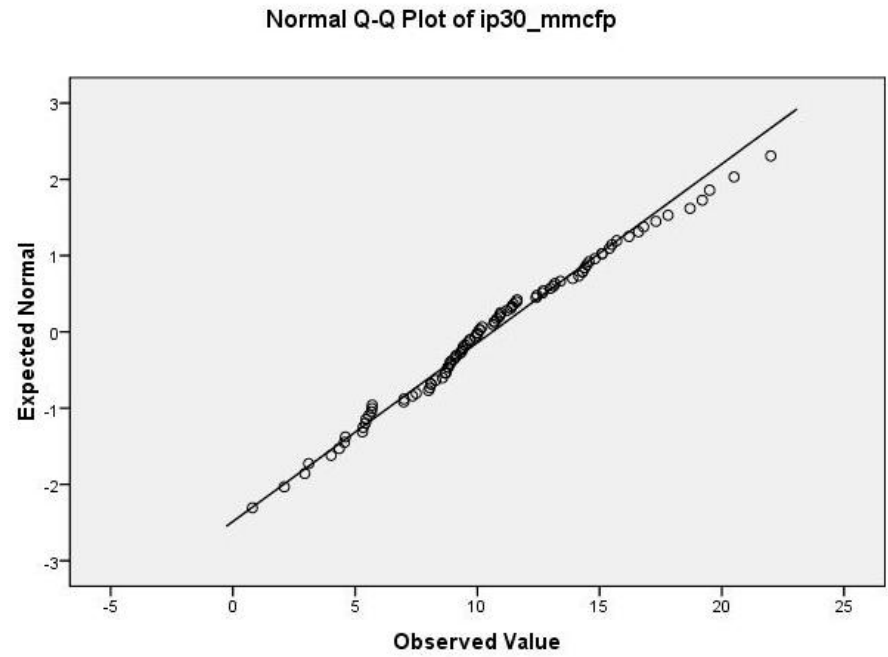
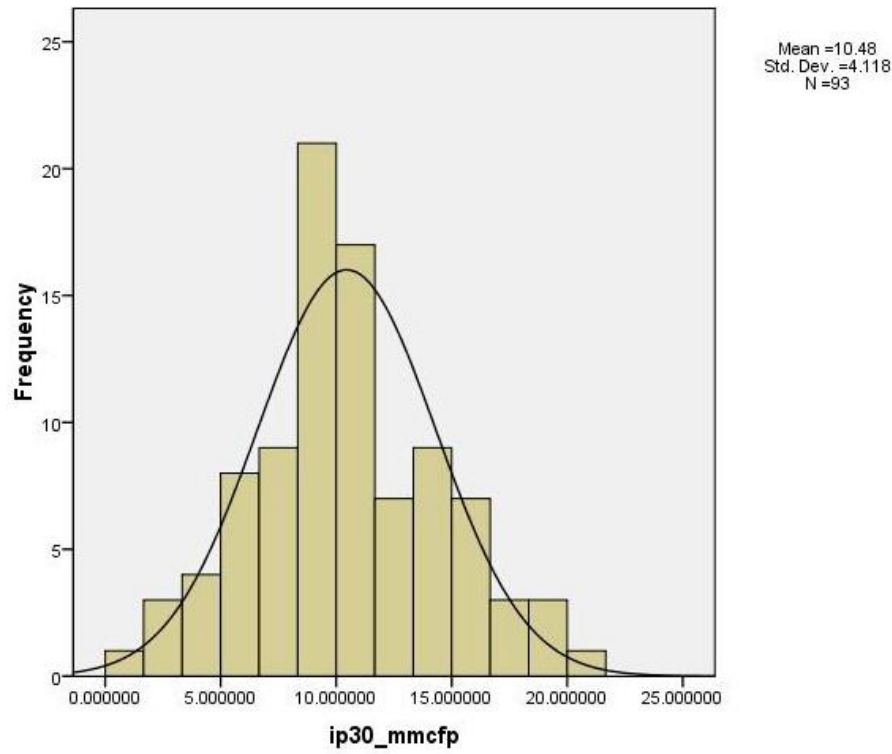


Figure 4. Normally distributed data test example. (a) IP30 production data histogram. (b) IP30 production quantile Q-Q plot.

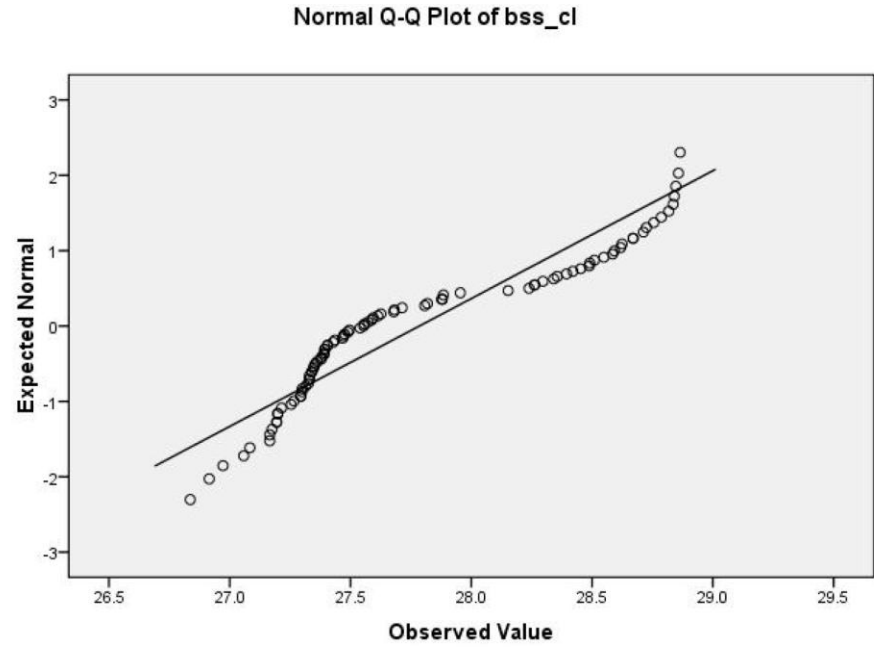
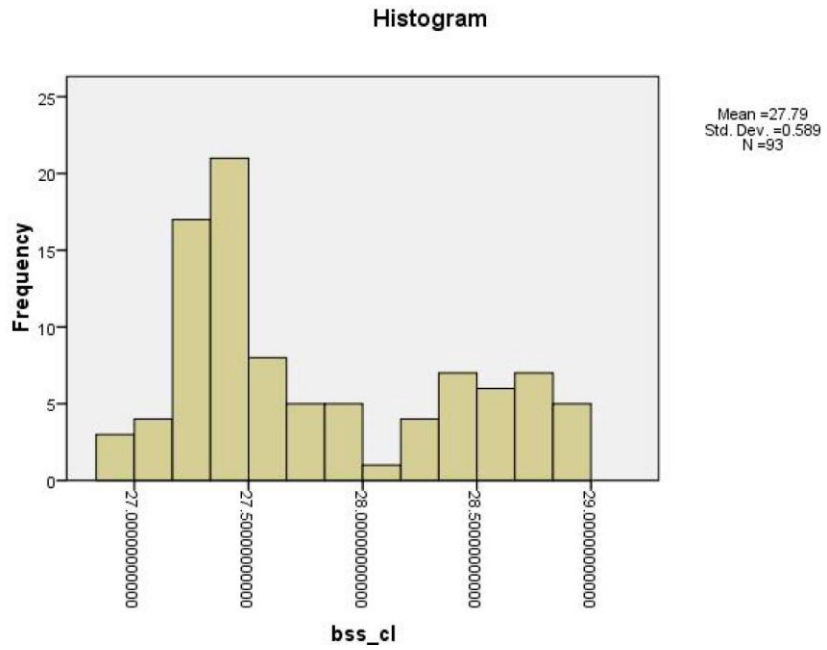


Figure 5. Bimodally distributed data example. (a) Bimodal distribution histogram. (b) Quantile Q-Q plot indicating a non-normal distribution.

### One-Sample Test

	Test Value = 12			
	t	df	Sig. (2-tailed)	Mean Difference
ip30_mmcfp	-3.548	92	.001	-1.515258065

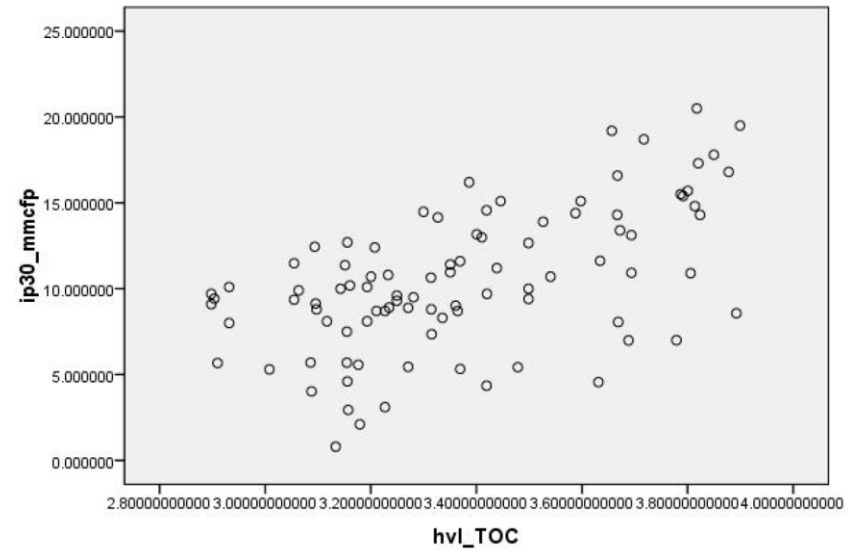
Figure 6. SPSS t-test results for  $\mu_0 = 12$ .

**Correlations**

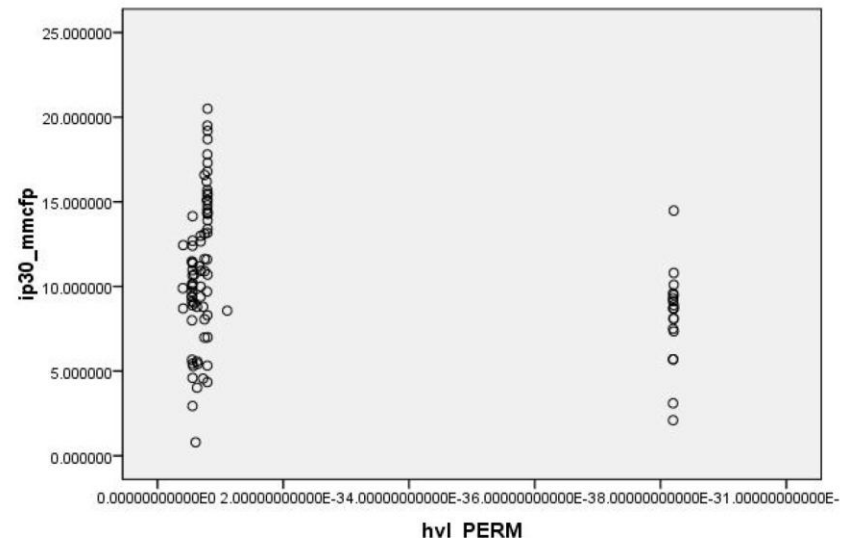
		ip30_mmcfp	hvl_TOC
ip30_mmcfp	Pearson Correlation	1	.571
	Sig. (2-tailed)		.000
	N	93	93
hvl_TOC	Pearson Correlation	.571	1
	Sig. (2-tailed)	.000	
	N	93	93

\*\* . Correlation is significant at the 0.01 level (2-tailed).

Scatter plot for rel. of HSVL TOC and IP30



Scatter plot for rel. of Perm and IP30 form HSVL



**Correlations**

		hvl_PERM	ip30_mmcfp
Spearman's rho	hvl_PERM	Correlation Coefficient	1.000
		Sig. (2-tailed)	.439
		N	93
	ip30_mmcfp	Correlation Coefficient	.081
		Sig. (2-tailed)	.439
		N	93

Figure 7. Correlation test results for Haynsville Formation. (a) Correlation table for IP30 and TOC. (b) Scatter plot between IP30 and TOC. (c) Correlation table for IP30 and permeability. (d) Scatter plot between IP30 and permeability.



**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-18.492	4.378		-4.224	.000
	hvl_TOC	8.583	1.292	.571	6.641	.000

a. Dependent Variable: ip30\_mmcfp

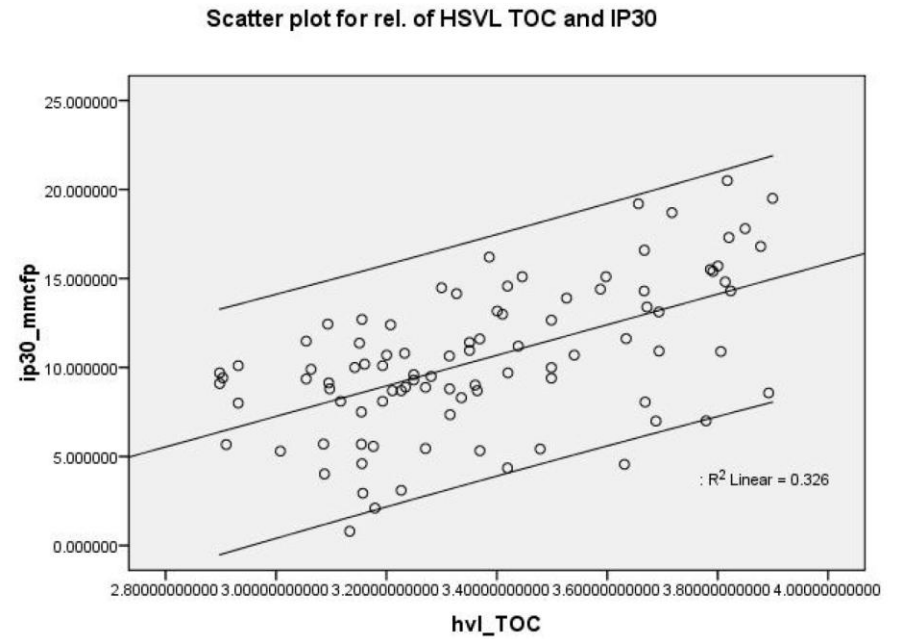


Figure 8. Regression test results example for Haynesville Formation. (a) Linear regression results for IP30 and TOC. (b) Regression scatter plot for IP30 and TOC.

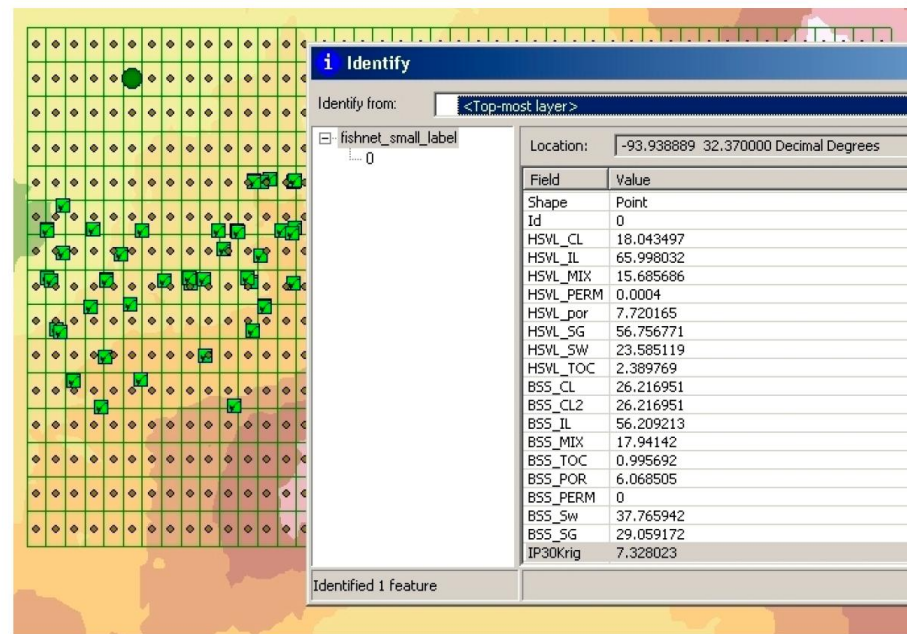
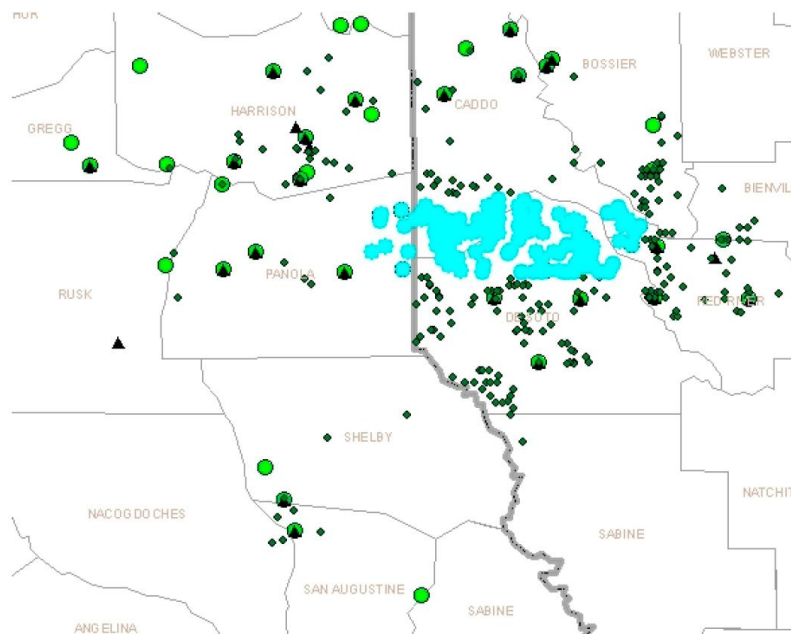


Figure 9. GIS selected wells before interpolation and with interpolated “virtual wells”. (a) Selected wells with normally distributed data with least clustering. (b) Fishnet with selected virtual well and its attributes.