GCAGS JOURNAL
A Publication of the Gulf Coast Association of Geological Societies
www.gcags.org

Analytics, Machine Learning, and AI for Independent Petroleum Geologists and Independent Oil and Gas Companies: Successful Case Study, Rodessa Oil Pool, Martinville Field, Simpson County, Mississippi

Bill Fairhurst

Riverford Exploration, L.L.C., The Woodlands, Texas 77382, U.S.A.

ABSTRACT

Petroleum geologists have knowledgeably and successfully applied geostatistics and other analytical tools for decades as one of their many skill sets. Geologists' knowledge and practiced, successful implementation and use of these tools have them well placed for involvement in current oil and gas industry transformation into the broader use of analytics, machine learning, and AI (artificial or augmented intelligence).

This example of the Rodessa Oil Pool, Martinville Field, Simpson County, Mississippi, provides insight into the use of descriptive statistics to advance geologic insight and discovery using machine learning to define and distinguish separate reservoir productivity when completed in non-segmented multiple reservoirs and from multiple wells. This smaller, well-controlled dataset provides an excellent example of and learning set for the powerful discovery using descriptive statistics and the predictive strength of simple, more advanced analytic methodologies.

The statistical models are defined and tested using standard, normal statistics, sample sets, replacement in this case and addition of new data, and review of the validity of the models developed. These tools and methodologies are consistent with current practice in petroleum geology. Discovery of variables and integration of multidisciplinary research will continue to take an ever-increasing role in the work and practice of petroleum geologists.

INTRODUCTION

Geologists have successfully applied geostatistics and other analytical tools for decades. The current oil and gas industry transformation into the broader use of analytics, machine learning, AI (artificial or augmented intelligence) varies only in consistency of application and scope. Petroleum geologists who have this long history of successful use of analytical tools are well suited to be part of the analytical revolution or transformation occurring as part of their own work, within their own firms, and throughout the oil and gas industry.

Geologists have recognized statistics and analytics for what they most truly are, a tool. This is just one tool available for use in specific projects or programs to enhance understanding and communication of concepts for that project, program, or study. It is not universally applicable or available for academic or applied research or for communication of concepts.

Manuscript received April 2, 2020; revised manuscript received August 1, 2020; manuscript accepted August 5, 2020.

GCAGS Journal, v. 9 (2020), p. 93–114.

Analytics is the discovery, interpretation, and communication of meaningful patterns in data and predictive models. This includes the evaluation, synthesis, integration of disorganized or complex data, patterns of data, and interpretations typically from multidisciplinary sources and effectively communicating them into meaningful patterns or results often as predictive models. These, similar steps, and processes are familiar to the petroleum geologist who accomplishes these steps and processes continuously.

The current firm and industry-wide universal application of analytics is being heralded as a needed transformational event in the oil and gas, other industries, and is being pushed down by boards and senior management throughout the oil and gas industry. However firm-wide, early adoption and practice has resulted in a 70–90% failure rate (Overby, 2020; Sood and Coxon, 2020; Taylor, 2020; Bagchi, 2020; Kesari, 2019; Freedman, 2019; Bendor-Samuel, 2019; Stuchfield, 2019; Hinchcliffe, 2018; Walker, 2017). Why do these corporate transformational events such as adoption of analytics have such a high failure rate? The referenced research provides many reasons which can best be summarized by petroleum geologist Andrew Silver's Venn diagram (Silver, 2019) (Fig. 1).

This Venn diagram brings together three parts of the analytical team. The analyst (math and statistics) in the upper right,

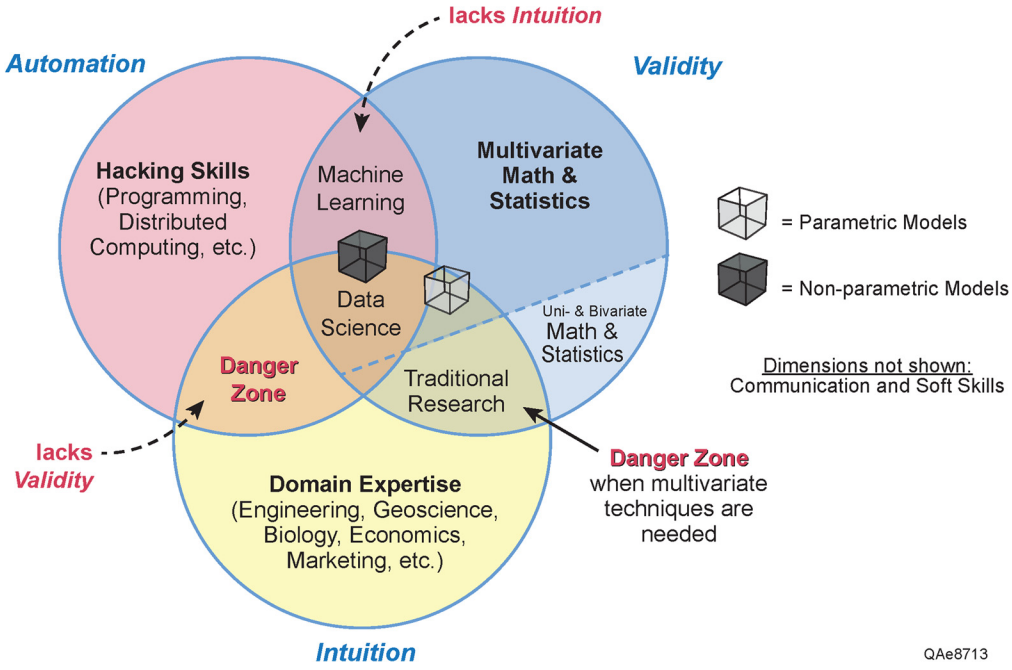


Figure 1. Potential analytical team interactions and performance (Silver, 2019).

programmer in the upper left, and below domain expert (geologist, engineers, economist, and landman). The overlap between the analyst and domain expert is the white box area where we want to be. White boxes are analytical solutions, statistical models, which are understood and applicable by both analyst and domain expert. The machine learning overlap between the analyst and programmer marginal to or excluding domain experts are often black boxes. In Figure 1, those can be non-parametric or non-standard statistical analyses (Hoskin, 2012). These analyses should be or need to be easily explained to the knowledgeable and practiced domain expert but often are not. Not all analytical processes or models developed in this non-intersecting area of this Venn diagram are universally applicable or useable models for all experts represented by the diagram. These models may not have been satisfactorily explained by programmers and data scientists, may be poorly understood solutions, may not be practical for subsequent use by others, or may not be repeatable scientific solutions and by definition, are not science. As Albert Einstein explained "all physical theories, their mathematical expressions apart, ought to lend themselves to so simple a description 'that even a child could understand them.'" (Clark, 1971).

In the universe of Figure 1, analytics are a useful tool. However, they are not universally applicable. There are applications when analytical tools are beneficial in explanation and communication; there are many when they are not. It is up to the team of knowledgeable and practiced domain experts, data scientists, and programmers to distinguish and communicate the applicability or non-applicability of the tool. To expect that analytics, machine learning, or AI is the only approach, the universal key, and useful tool in every application and for every project, universally throughout organizations or an industry, is unrealistic.

If one combines the programmer with the analyst in the upper right-hand portion of the Venn diagram, the multiple factors that have made the analytical transformation so unsuccessful can be explained. The upper left population will now represent management. In this model, the best analytical models are developed in the overlap among manager, analyst, and domain expert. Again, the failure of one of the three results in failure of the process. Why? The more the domain expert understands, works with and practices analytics the greater the overlap in the Venn diagram with the analyst. The more the analyst works with domain experts and takes time to understand their work and variables of their domain, the greater the analyst's circle will overlap with the domain expert's. Both the domain expert and analytic sciences include analytical disciplines and backgrounds so there is great potential for significant overlap in the data science and domain experts area of expertise.

That leaves management. In many organizations the analytical transformation throughout the industry is being charged from upper management and Boards that have observed the use of analytics in other industries. They may be pushing the technical tool down into groups managed by individuals with domain expertise but no longer functioning in those roles. Many of these managers have not incorporated analytics into their own previous work and in practice are unlikely to successfully integrate both the domain expertise, analyst, and the common Venn diagram intersections and overlap between these two groups, and themselves, as managers.

In these cases, regardless of how much overlap exists between the domain expert and analyst, there is little to no overlap with management and the analytical transformation will fail. Organizations should seek management with proven expertise with these processes, practice, and methodologies or make sure management gains the training, knowledge, and experience to increase the overlap with domain experts and analysts in data science, the center of Figure 1. There are cases when analytics will not be applicable. In those cases, other domain expert's tools are likely more applicable to project and company success. The transformation will not be universal. These observations and understanding will assist in making analytics, machine learning, and AI more successful.

ANALYTICS FOR THE INDEPENDENT GEOLOGIST AND INDEPENDENT OIL & GAS FIRM

The more advanced analytic models, machine learning, and AI tools and application are achievable for knowledgeable and practiced domain experts trained in most geologic and engineering disciplines. In fact the logic of the domain expert and even simple analytical models often provide superior results that are,

white boxes. More sophisticated, more advanced pure mathematical, statistical, and programing skill models developed only by data scientists and programmers may be black boxes.

In this study we will examine the use of just a few simple analytic variables and models as one of the tools used by domain experts to enhance the interpretation of a small project and problem. The goals are:

(1) Many mathematical and statistical experts and firms are attempting to penetrate the oil and gas industry and firms but with little to no industry expertise with successful application. Successful application is applied and shown here in a real-world solution.

(2) These variables and models do not need to be extensive or complex to come up with superior results to models and solutions which when used during interactive project development can assist in the discovery, interpretation, and communication of multidisciplinary and intangible concepts.

(3) Finally, most of the observations, variables, and statistical models are within the grasp of most petroleum geologist and Independent oil and gas firms. Practical, successful use and application will assist others in obtaining similar superior results using these analytical tools.

Why do petroleum geologists need to be involved professionally and personally with analytics, machine learning, and AI? As stated by the Wall Street Journal (2017), the future in the Global Economy "it will soon be a matter of those who can and those who cannot."

THE EXAMPLE: RODESSA OIL POOL, MARTINVILLE FIELD, SIMPSON COUNTY, MISSISSIPPI

Martinville Field

Martinville Field, northeastern Simpson County, Mississippi, is located along the northern margin of the Mississippi Salt Basin (Fig. 2). This northern margin of the Mississippi Salt Basin is also known as the Lower Cretaceous (Hauterivian, Barremian, Aptian, and Albian) production trend (Fig. 3).

Field Discovery

After disappointing drilling and production results from Lower Cretaceous reservoirs along the northern flank of the Mississippi Salt Basin through the 1940s (Forgotson, 1963) into the early 1950s (Nunnally, 1954), the Bolton Field (cumulative production of 6.1 million barrels of oil (MMBO) and 12.3 billion cubic ft of gas (BCFG) through 2017 (Mississippi Oil and Gas Board, 2020) was discovered 15 mi west of Jackson, Mississippi, in 1954 (Frascogna, 1957; Davis, 1963). The discovery well was drilled and completed in the Paluxy during July with the deeper Rodessa added during December as the confirmation well.

The next year, 1955, the Citronelle, Rodessa, Oil Field was discovered (Eaves, 1976) in the eastern Mississippi Salt Basin in southwestern Alabama. Citronelle has produced 175 MMBO and 16.0 BCFG through the end of 2019 (Geological Survey of Alabama, 2020). Citronelle Salt Dome forms a topographic high, which led to the drilling of two shallow, dry wells. These wells confirmed the subsurface structure which lead to the drilling of the discovery well drilled deeper into the Rodessa (Eaves, 1976). These two new field discoveries, Bolton and Citronelle, defined the Lower Cretaceous play potential and target areas. Martinville is located near the center of the Lower Cretaceous play on the northern flank of the Mississippi Salt Basin (Fig. 2B); Bolton Field is 48 mi northwest and the Citronelle Field is 110 mi southeast.

Using regional gravity data to identify gravity lows (salt ridges and salt domes, thick salt formed from diapiric movement) and gravity residual mapping to identify "the location of the highest and/or the thickest salt mass" (Karges, 1968) companies shot seismic data to confirm and drill multiple new field discoveries in the Lower Cretaceous Mississippi Salt Basin play (Davis, 1963; Frascogna, 1957) from the mid-1950s (Philpott, 1960) into the 1960s.

In February of 1957, the Martinville Field discovery well, the Central Oil Company #1 Jennings was drilled and tested 246 barrels of oil per day (BOPD) in the Sligo. In this case, the prospect was defined by subsurface mapping (Davies, 1963), similar to the Citronelle discovery. In November 1957, the Hosston was completed in the Central Oil Company #1 Sullivan flowing at 259 BOPD and the Rodessa was added in January 1958 in the Gulf Oil #2 Sullivan well. The field has produced just over 10 MMBO and 2.8 BCFG with the Rodessa in the northeast fault block being the most productive horizon producing 3.7 million barrels of oil equivalent (MMBOE) or 36% of the field total. Lower Cretaceous production has been established in the Martinville Field from Washita-Fredericksburg, Paluxy, Mooringsport, Rodessa, Sligo, and Hosston, the complete suite of Lower Cretaceous traditional reservoir horizons. The only non-producing Lower Cretaceous formations are the Ferry Lake Anhydrite and Pine Island Shale (Fig. 3), immediately above and below the Rodessa, respectfully.

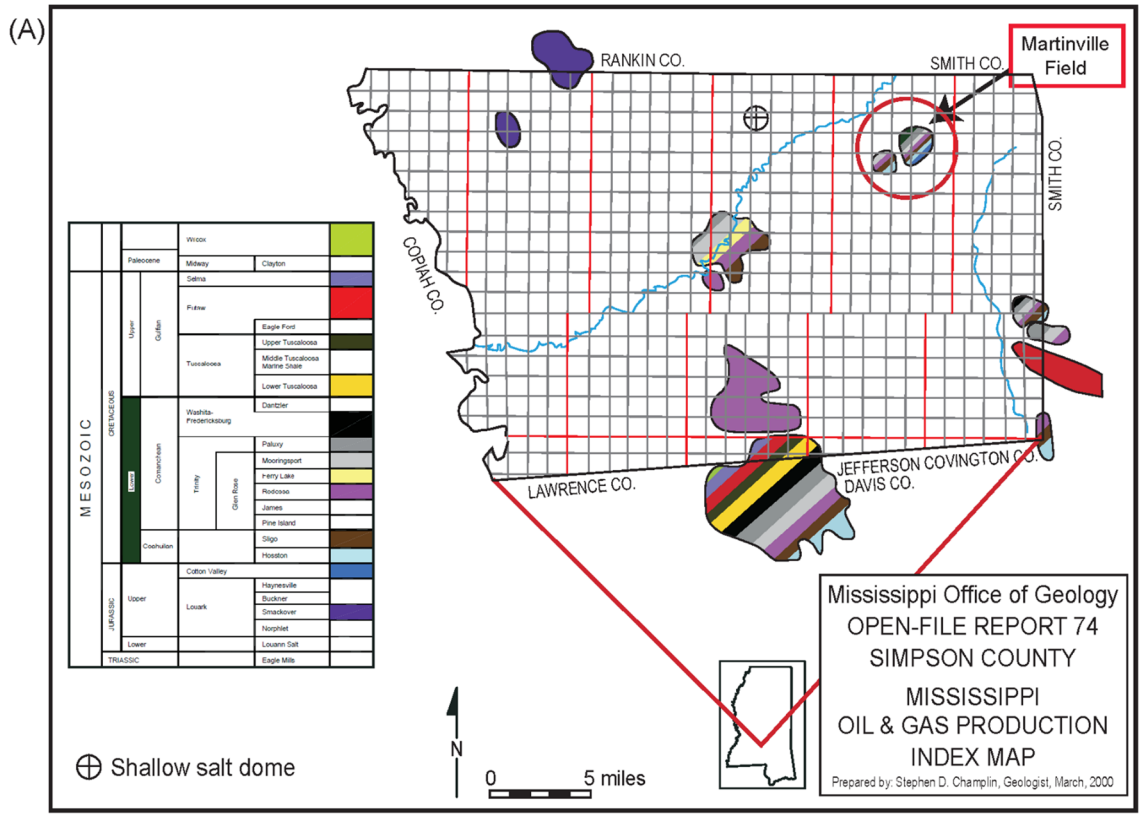
Structural Trap

Martinville Field is a faulted, salt anticline with four-way closure (Fig. 4A). The salt dome is non-piercement at least through the Lower Cretaceous stratigraphic units. The majority of the production is from the high-side, two-way fault closure and two-way dip closure, northeastern fault block. The west-southwestern and southern fault block are identified as Southwestern Martinville Field or by separate pool names (Mississippi Oil and Gas Board, 2020), all part of the single salt dome, structure, and structural origin.

Rodessa Formation Reservoir

The Rodessa Formation is 470–520 ft thick in the Martinville Field (Fig. 5). The best developed sandstone facies have well-developed, bell-shaped spontaneous potential (SP) log signature. Core from the Continental Oil Company #1 Central Oil Company drilled in 1955 (northernmost well in the northeastern, upthrown fault block [Figs. 4A, 4B, and 5], a dry hole that may have been a missed discovery well for the field and helped set up the interpretation for the discovery well) was cut and recovered through most of the Ferry Lake Anhydrite, Rodessa, Pine Island, Sligo, and Hosston. Like descriptions of the Rodessa in Citronelle Field (Esposito et al., 2008), the sandstones have intraclastic conglomeritic bases, grading upward into white to light-gray coarse to medium grain sandstones to finer grained, darker-gray sandstones into red-brown and darker gray and black mudstones. These red-brown mudstones frequently have slickensides noted throughout and the darker shales are occasionally interbedded with coal. Fossils are noted in the finer grained mostly upper Rodessa darker-gray sandstone and mudstones and there is mica throughout the section regardless of grain size. Better developed, thicker, higher SP response, higher porosity (core description and sidewall core laboratory measurements) are in the lower half. The upper half appears finer grained, with lower and less distinctive SP response and tighter porosity (core descriptions and sidewall core laboratory measurements). The gross sandstone isopach and isopach maps of individual layers (defined later) are north-south oriented sandstones similar to the Rodessa descriptions at Pelahatchie Field, 22 mi to the north (Karges, 1968; Petkovsek, 2019).

The Rodessa sandstone reservoirs are interpreted to be generally north-south oriented fluvial-deltaic channel complexes



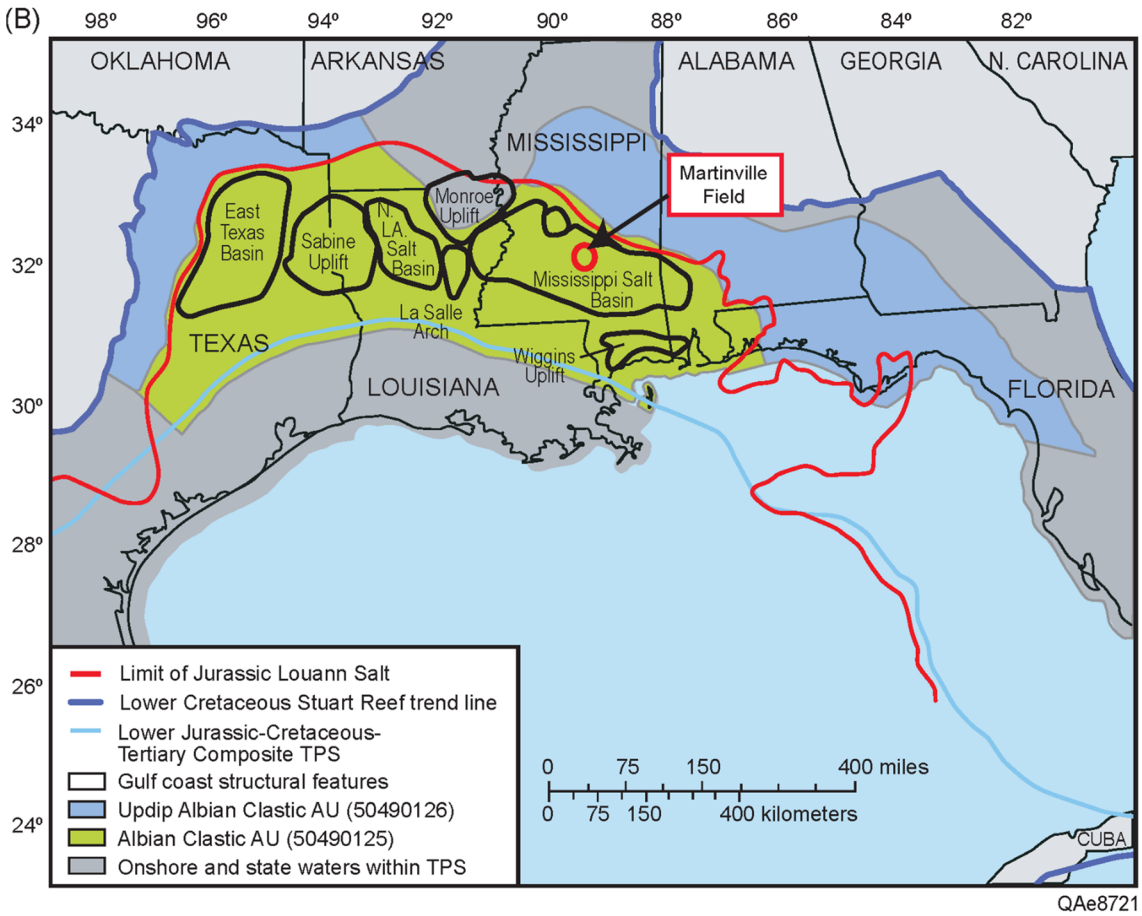


Figure 2. Martinville Field is located in northeastern Simpson County, Mississippi. Figure 2A modified after Champlin (2000) on the northern margin of the Mississippi Salt Basin; and Figure 2B modified after Merrill (2016).

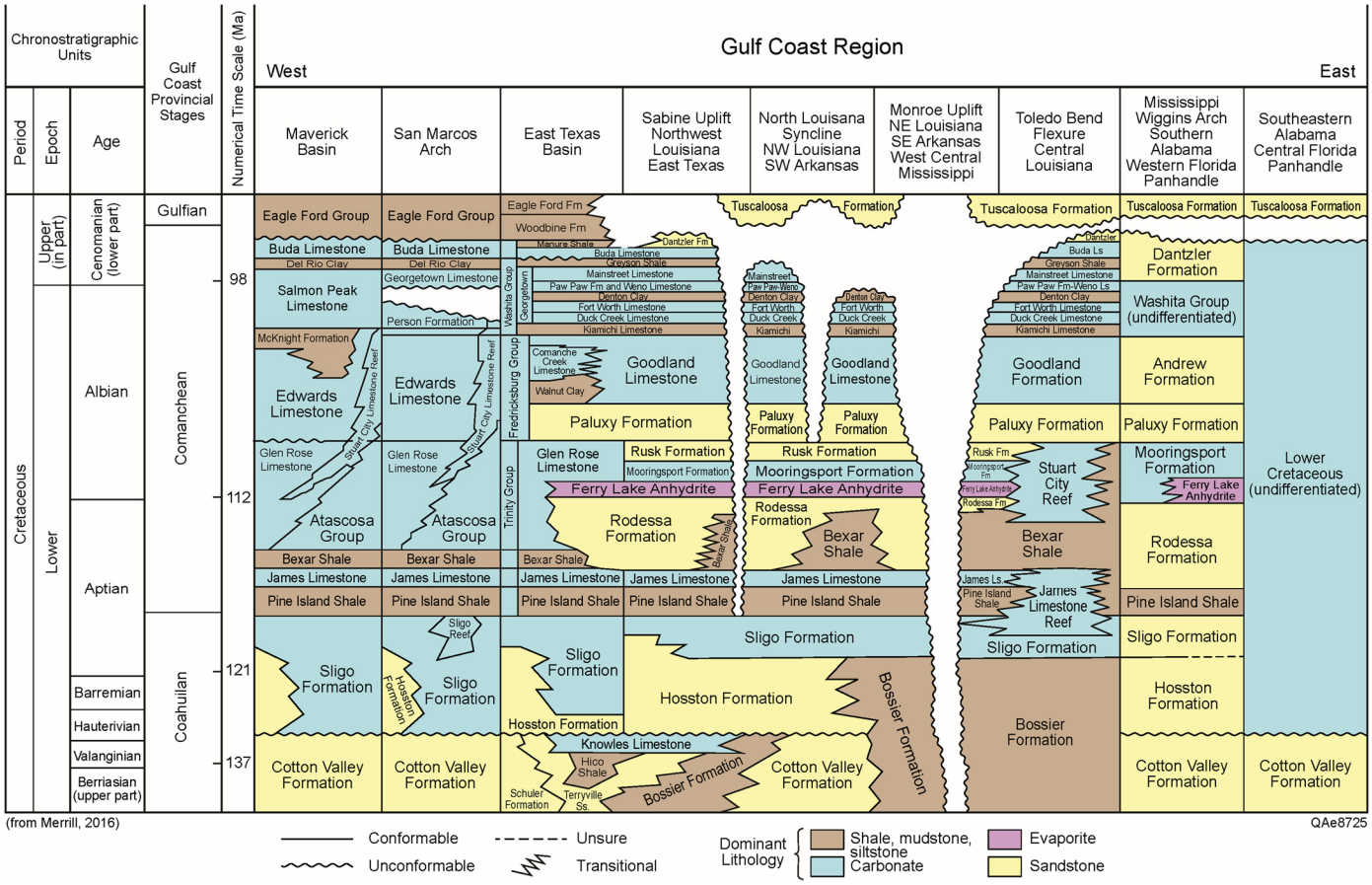


Figure 3. Stratigraphic column for the Lower Cretaceous in the U. S. Gulf Coast region (from Merrill, 2016, courtesy of the U.S. Geological Survey). The figure is based on the American Association of Petroleum Geologists (2002) Gulf Coast regional correlation chart, and modified with information from Ambrose et al. (2009); Anderson (1980, 1989); Chenault and Lambert (2005); Mancini et al. (2008); Salvador (1991); Salvador and Quezada Muñeton (1991); Smith et al. (2000); and Zahm et al. (1995).

in the lower half and distributary channel and littoral environments in the upper half. The mudstones were deposited in levee, flood-plain and bay environments. The slickensides in the reddish-brown mudstones at the top of the fining-upward sequences are likely of pedogenic origin (Fig. 6). The mica rich mineral assemblages likely indicate a southern Appalachian highland source draining into the relatively stable northern margin of the Mississippi Salt Basin (Mancini, 2012; McFarland and Menes, 1991; Forgotsen, 1963). The channelized sandstone forms an anastomosing reservoir system likely forming continuous and non-continuous reservoirs, vertically and horizontally (Pashin, 2013).

Individual sandstone channels or channel systems are thicker and higher quality reservoir in the lower Rodessa. The channelized sandstones generally thin upward and basinward away from the sediment source and supply with the entire section continuing to thicken to the south in the accommodation zone of the subsiding Salt Basin, a basinward thickening interval (Forgotson, 1963). Deeper into the basin, lower energy, environments were present. There finer grained, darker, pelagic shales (McFarland and Menes, 1991) eventually becoming carbonate rich, the lateral equivalent to the lower Glen Rose in East Texas and upper James Limestone in the East Texas and Louisiana salt basins (Merrill, 2016; McFarland and Menes, 1991; Devery, 1982; Nunnally and Fowler, 1954). The upper sequences are finer grained, have a decreased, siliciclastic sediment supply resulting "in the spread of an epineritic biostromal environment over most of the shelf area basinward of the near-shore" siliciclastic facies (Forgotson, 1963).

In the Martinville Field, 6 genetic fining upward sequences were recognized and correlated throughout the field (Fig. 5). These fining upward sequences range from 45 to 120 ft thick. Again, generally, with thicker individual sandstone units with better quality reservoir at the base and updip becoming thinner and lower quality reservoir upsection (Fig. 6) and into the basin.

Seal

Biohermal carbonate facies that developed basinward of the upper Rodessa sequences eventually formed the restriction necessary for the formation of calcium sulfate, deposited as gypsum. Early anhydrite stringers are interbedded with the upper Rodessa basinward of Martinville Field eventually becoming more continuous as a separate formation, the Ferry Lake Anhydrite, sealing the Rodessa. The anhydrite was deposited on the seaward, slightly less stable, platform that accommodated deposition of thick gypsum layers. This restricted environment never gained the salinity high enough for the deposition of salt. With burial compaction pressure, the gypsum was altered to anhydrite (Forgotson, 1963) forming an excellent seal for Lower Cretaceous (Aptian) reservoirs. The base of the Ferry Lake Anhydrite / top of the Rodessa is an excellent local and regional mapping horizon and "a regionally valid time-rock unit representing contemporaneous deposition" (Forgotson, 1963).

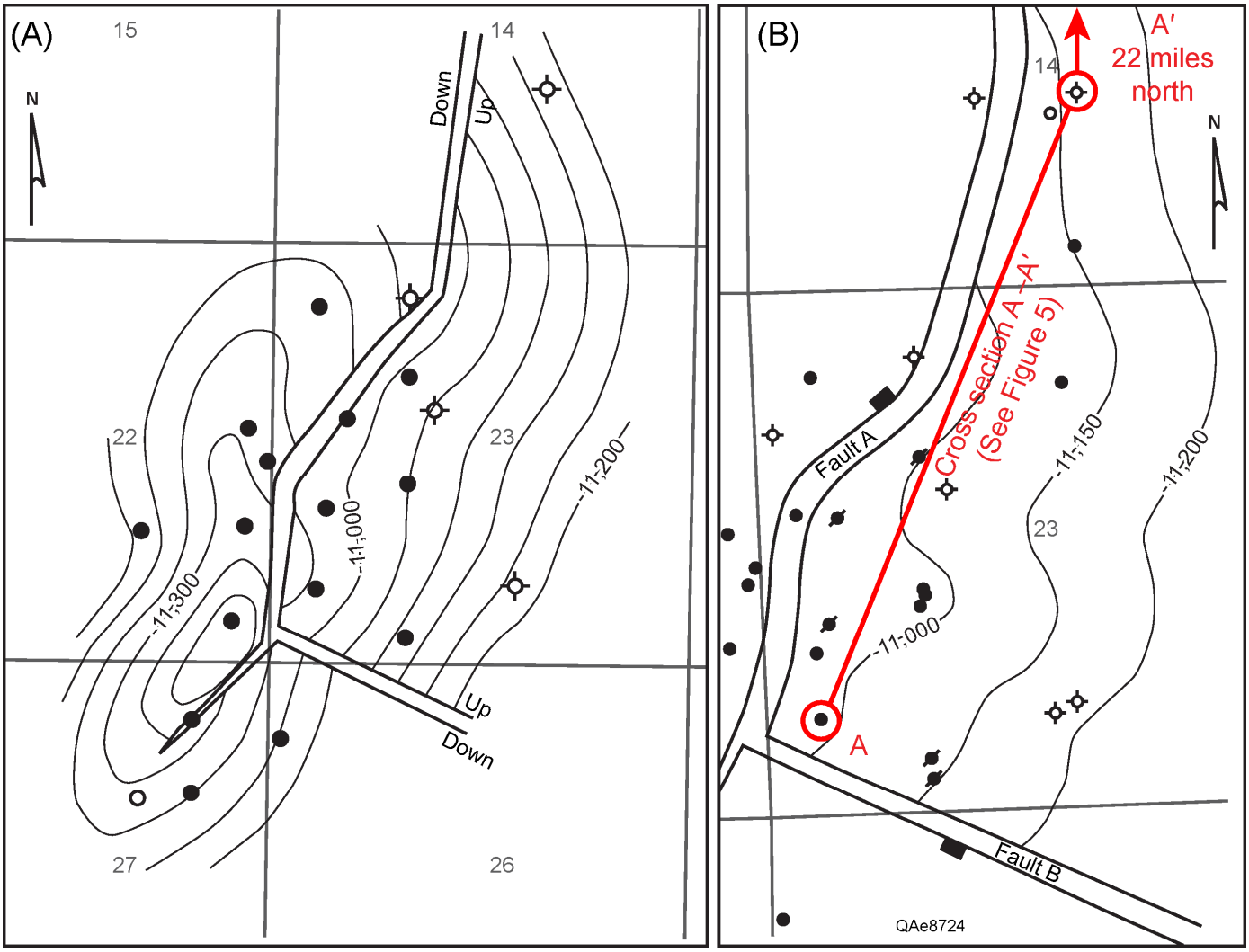


Figure 4. Structure maps on the Base Ferry Lake Anhydrite / Top Rodessa Formation. Figure 4A modified after Davis, (1963); and Figure 4B modified after Fairhurst (1996) with 2017 revisions. One mile section boundaries provide horizontal scale. Cross section A–A' is shown in Figure 5.

Source and Migration

There are multiple source rocks throughout the stratigraphic section of the Gulf Coast Mesozoic interior salt basins (East Texas, Louisiana, and Mississippi). The majority of those are locally or play specific source units with the lower and middle mudstones of the Upper Jurassic Smackover Formation being the only regionally significant source rocks within each of these interior salt basins (Merrill, 2016; Mancini et al., 2012, 2008, 2005; Wescott and Hood, 1994; Claypool and Mancini, 1989; Sassen et al., 1987; Oehler, 1984). The Upper Cretaceous is only buried deep enough to generate sufficient hydrocarbons in the very southern portions of the East Texas and Mississippi salt basins and the Lower Tertiary is only buried deep enough to generate hydrocarbons in the southern Louisiana Salt Basin (Fig. 7). Mancini (2012) also documented a higher geothermal gradient in the Louisiana Salt Basin compared to the Mississippi Salt Basin associated with Cretaceous volcanic emplaced during extension. Similar volcanic rocks, timing, and genesis have been documented in the Realfoot Aulacogen and South Texas. Chemical analyses of the Smackover source, intra-formational Smackover oils mostly found in the upper Smackover, and many of the hydrocarbons in Cretaceous reservoirs that have been typed and are determined to have been derived from the lower and middle Smackover mudstones (Mancicni, 2012).

In the Martinville Field area, the Smackover source rocks started generating oil during mid-Albian (Figs. 8B and 9) likely after deposition of the Paluxy. Although migration would have occurred slightly later likely during the Late Cretaceous (Fig. 8C). Wescott and Hood (1994) studied the migration of these hydrocarbons in East Texas and Evans (1987) in Mississippi evaluated vertical migration processes under pressure with and without faulting. Under those processes, most traps fill a single reservoir. Migration under pressure occurs until that pressure is relieved in a normal pressure environment. When this normal pressure environment is found by the migrating hydrocarbons it remains in this single, higher porosity, higher permeability reservoir compared to the source, carrier beds, or faults under pressure.

Evans (1987) also noted in the Mississippi Mesozoic Salt Basin that a few, the largest fields, have multiple stacked reservoirs. Each of these have well documented faults through the hydrocarbon reservoir units that continue down to the Smackover providing a migration highway for the hydrocarbons generated in

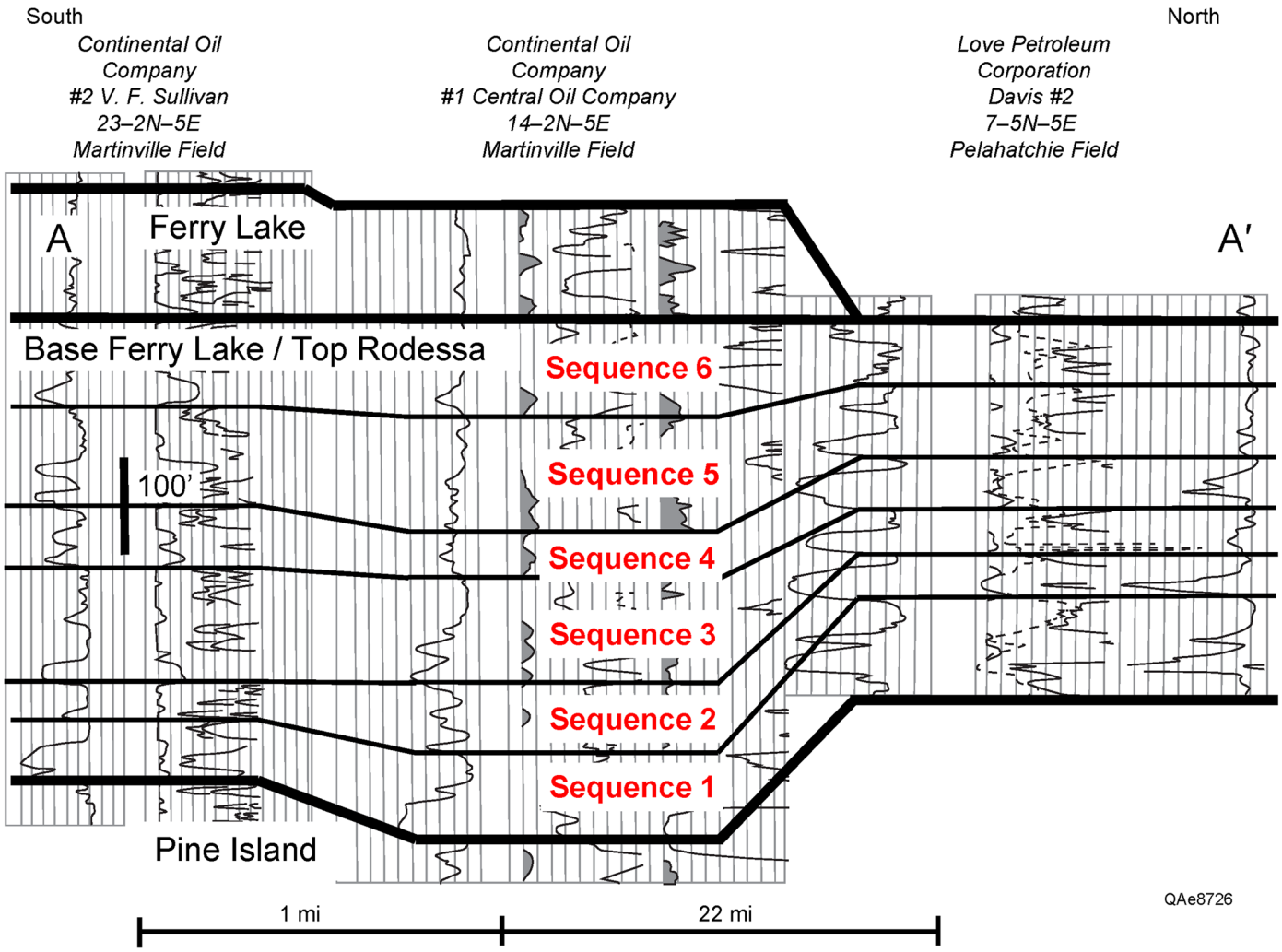


Figure 5. Stratigraphic cross section A–A’, flattened on the base of the Ferry Lake Anhydrite / top of the Rodessa. Line of cross section through the Martinville Field shown in Figure 4B. Ferry Lake Anhydrite pinches out between the Martinville Field and Pelahatchie Field, 22 mi to the north (McFarlan and Menes, 1991; Forgotson, 1963). The Rodessa in Martinville was divided into 6 fining-upward genetic sequences that have thicker and higher quality reservoirs at the base and updip in the Pelahatchie Field.

the lower and middle Smackover (Sassen et al., 1987). Martinville Field is an example, in this case, a salt-cored, faulted anticlinal trap with vertical migration through a significantly thick Upper Jurassic and Lower Cretaceous section into multiple Lower Cretaceous hydrocarbon reservoirs along the fault surface.

APPLICATION OF GEOLOGICAL INTERPRETATION SUPPORTED BY ANALYTICAL DESCRIPTION, MACHINE LEARNING, AND AI FOR THE RODESSA OIL POOL, MARTINVILLE FIELD

Background

During December 1995, Coho Resources petitioned the State Oil and Gas Board of Mississippi to unitize the Martinville Field and Southwestern Martinville, Rodessa into a single unit. At that time, Coho was producing from and controlled individual 40 ac HBP (held by production) units in the southern two-thirds of the northeastern Martinville Field (Fig. 4), multiple units in the Southwestern Martinville Field and two other operators (Marathon Oil Company and Sklar & Phillips) held two HBP 40 ac units in the northern third of the northeastern Martinville Field (see Marathon [TXO] Kennedy 23–2 [subsea elevation, -11,130 ft] and Sklar & Phillips Kenedy 14–15 #1 [subsea elevation, -11,154 ft] in the northern portion of Figure 4B).

The participation formula presented by Coho used 5 variables, four engineering variables; cumulative production from each tract (Fig. 10A), 25% weight, current production (Fig. 10B), 10% weight, estimated remaining reserves (Fig. 10C), 35% weight, and useable wellbores in each tract (25% weight) and one land variable; surface acreage, 5% weight (Coho, 1995). There were no scientific—engineering, geological, or mathematical/statistical/analytical—justifications for the variables or weights and no recognition of the geology or reservoir characteristics of the Rodessa. However, the faulting and separation as mapped (Fig. 4) was accepted and known.

The tract participation would give Coho in the southern two-thirds of the northeastern fault block in the Martinville Field, 79.9% interest, and the other companies in the northern one-third of the productive fault block, 20.1% interest.

Research Question

The focus was on the development, future production, and value of the northeastern fault block, Martinville Field, Rodessa

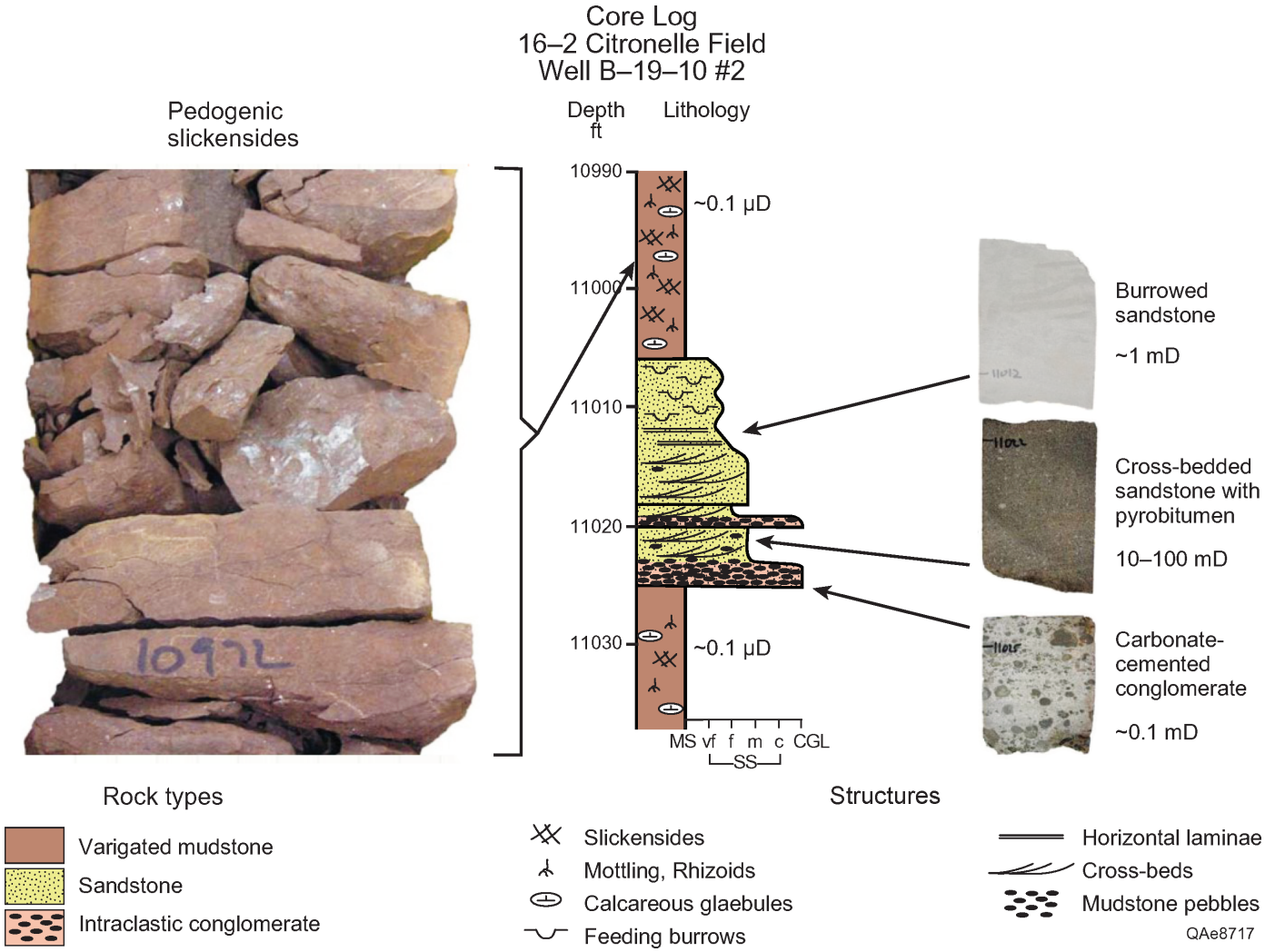


Figure 6. 16–2 sand, well B–19–10 #2. Pedogenic slickensides from Rodessa core in Citronelle Field (modified after Pashin, 2013).

Oil Pool and each 40 ac tract. The research question for geology, because the Rodessa in the northeast fault block was treated as a single reservoir, was there any way to define and communicate the potential for each 40 ac unit and one-third portions of the northeastern Martinville Field separately and differently than the single reservoir model using available data and information?

Geological Interpretation

Plotting the perforated intervals for each of the Rodessa producing wells in the northeastern Martinville Field fault block, it is observed that the most wells produced from sequence 3 through sequence 6 (Fig. 5) with only minor contribution from the upper sandstone in sequence 2 at the structurally highest portion of the northeastern fault block. Even though sequences 1 and 2 typically have some of the better developed and thickest, high-quality reservoir sandstone. All indications are that these sequences (1 and 2) are water wet. In sequences 1 and 2, the deep resistivity is lower with a negative deflection but has good separation between the intermediate and deep resistivity indicating a high-porosity and permeability sandstone reservoir. Well reports show no perforations were attempted in sequence 1, indicating the high-quality sandstone reservoir is water wet as is most of sequence 2. Recent data and interpretations of the Pelahathie Field (Fig. 6 and Petkovsek, 2019) also show a negative deflection in the deep resistivity and no reported production from sequence 1 in that field. The reported production from the Rodessa in the Pelachatchie occurs in sequences 2–4 (Petkovsek, 2019).

The amount of missing section measured in the wellbores along fault A (Fig. 4B) is approximately 400 ft and less along fault B. These are less than the thickness of the Rodessa formation (Fig. 5). It is likely that upper Rodessa sequences 3 through 6 on the upthrown fault block are trapped against the Ferry Lake Anhydrite and overlying lower Mooringsport Shale on the downthrown side of faults A and B. That juxtaposition forms an effective lateral seal. Rodessa sequences 1 and 2 (the thickest and highest-quality sandstone reservoirs) on the upthrown side are likely laterally offset to the upper Rodessa on the downthrown side, which likely formed an ineffective lateral seal for these sequences and reason sequences 1 and 2 are water-wet and non-productive.

A structure map (Fig. 4B) was constructed using the intersections of the fault surface maps with the upthrown and downthrown base of the Ferry Lake Anhydrite / top of the Rodessa surface. The width of each fault mapped, measured parallel to the fault dip direction, corresponds to fault heave (Subsurface Consultants & Associates, Inc., 1993; Tearpock et al., 1991).

Total net productive Rodessa sandstone in sequences 3–6 was interpreted from the available open-hole SP / resistivity logs in the northeastern fault block (Fig. 11A). This map is a statisti-

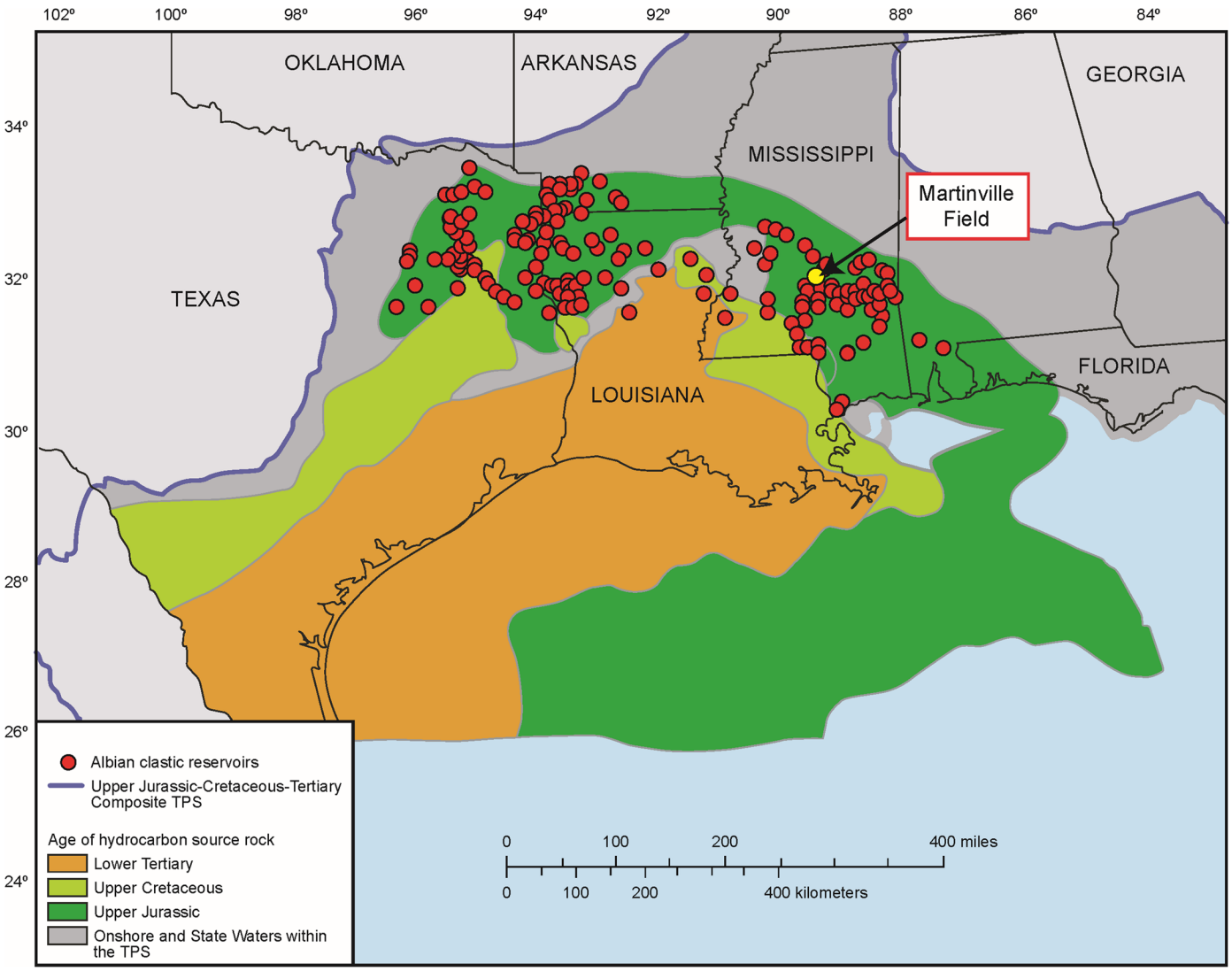


Figure 7. Map showing reservoir oil age across the U.S. Gulf Coast region (from Merrill, 2016, courtesy of the U.S. Geological Survey, as modified and generalized from Hood et al. (2002]). Red colored points indicate the approximate locations of the Albian age clastic reservoirs (Nehring Associates, Inc., 2009).

cal interpretation using Geographix software to eliminate interpreter bias and hand smoothed to eliminate some obvious contouring errors and statistical inaccuracies between datapoint and the statistical interval thickness mapped. Note that not all measured datapoints agree with the isopach contours (uncorrected statistical interpretation) but the interpretation provides a reasonable model of total Rodessa sandstone reservoir quality thickness.

Understanding the regional depositional setting, sedimentological process, facies, and intra-formational fining-upward depositional sequences as presented in Figure 5, were used to assist in interpretation of the productive net sandstone isopach maps for sequence 3 through sequence 6 (Figs. 11B–11E). The individual sequence isopachs created in 1996 (shown in black and white) were statistical interpretations using Geographix software then hand corrected to individual datapoint measurements. The revisions (color contours) were recreated in 2020 to include the additional well and subsurface data interpretations for the Smith 23–6 #2 well drilled and completed in late 1996. The location of the Smith 23–6 #3 well is found in the color contours of Figures 11B–11E; however, it is best identified in Figure 18C. These interpretations (color isopach maps, Fig. 11) were created using Petra software and are not hand corrected other than partial definition of several zero contour placements.

These standard geologic measurements and interpretations (Figures 4B and 11) by themselves do not provide any discernment of individual sandstone reservoir productivity. However, it was observed that sequence 3 appeared vertically thickest in the better, higher rate and higher cumulative, productive wells. This observation lead to investigation of the porosity-permeability relationships measured by sidewall cores (Fig. 12). The regression model between permeability and porosity is statistically valid at the 99% confidence interval (Snedecor, 1946). There are two clouds of data measurements with slightly high and low observations outlined in red in Figure 12. These were observed to be dominantly sequence 3 datapoints.

Separating permeability and porosity measurements by sequence (Fig. 13), it is observed that the majority of the sequences 4–6 and higher porosity sequence 3 (18–27% porosity) measurements follow similar linear models but that the mid-porosity range (6–16% porosity) in sequence 3 measurements have a much different linear relationship showing much higher permeability for porosities in that range than the general population of measurements. Based on these observations, sequence 3 sand-

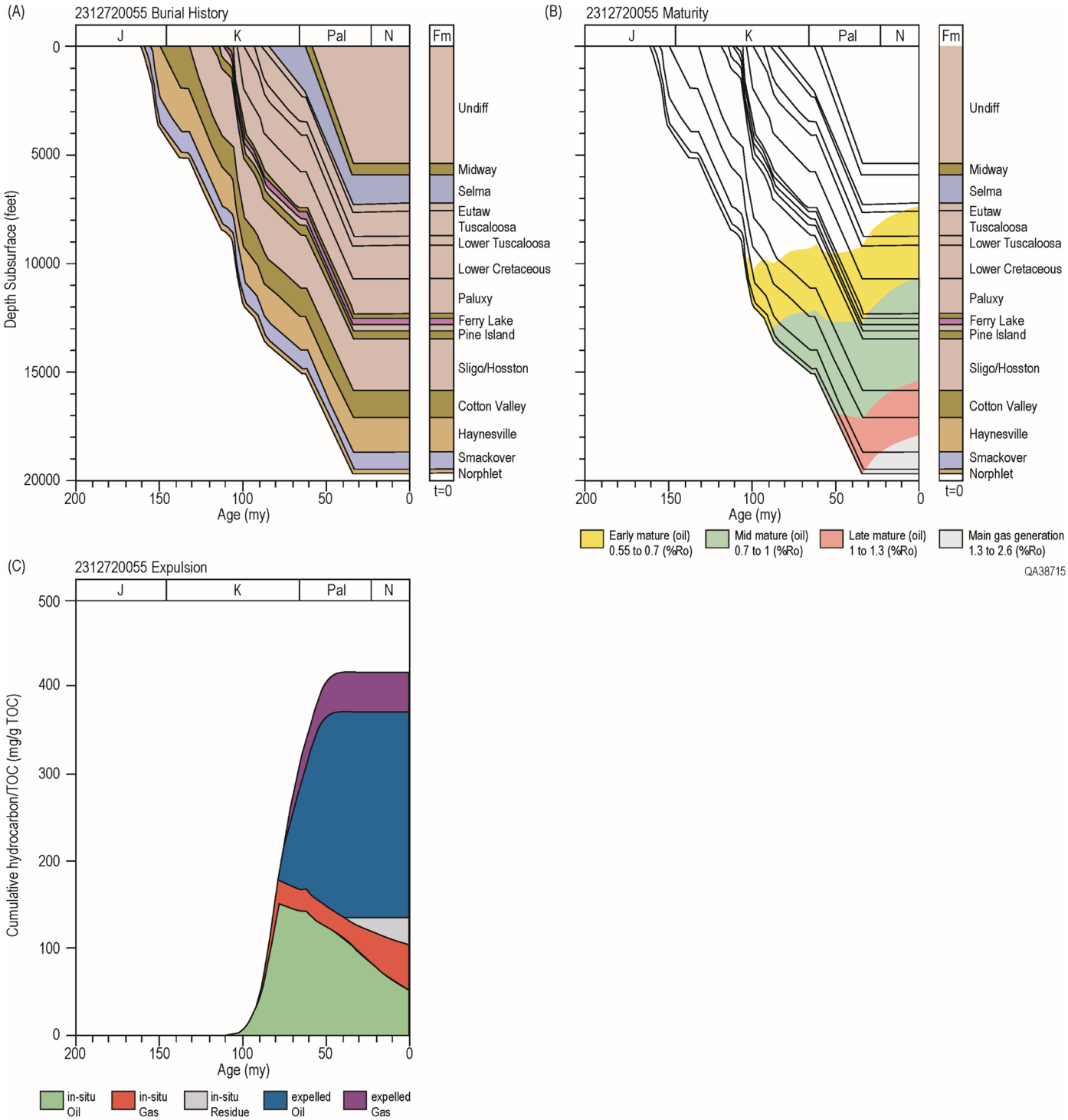


Figure 8. (A) Burial history for the Ram Petroleum (Pruett) Womack 1–3 well, 1–10N–17W, Magee Field, Simpson County, Mississippi. This well is approximately 10 mi southeast of the Martinville Field. The base of the Ferry Lake / top of the Rodessa is 100 ft deeper than at Martinville Field but a good proxy for the burial history for the Martinville. (B) Thermal maturation model. (C) In situ hydrocarbon generation and expulsion models. Modified after Mancini (2012).

stones with 10–16% porosities have much greater deliverability (permeability) than the general sandstone reservoir permeabilities measured. These models are statistically valid at the 99% confidence interval (sequence 5 is only valid at the 95% confidence interval, but the population of sequences 4–6 combined are valid at the 99% confidence interval) (Snedecor, 1946). These observations are highly statistically significant; but how are they quantified with productivity?

Statistics of Model Variables (Reporting and Description)

Descriptive statistics are the exploratory data analyses (EDA) defining the location, spread, and shape of the variable data distributions (univariate statistics). This process is often overlooked in application of machine learning and AI. It is important to build and review these simpler, individual variable

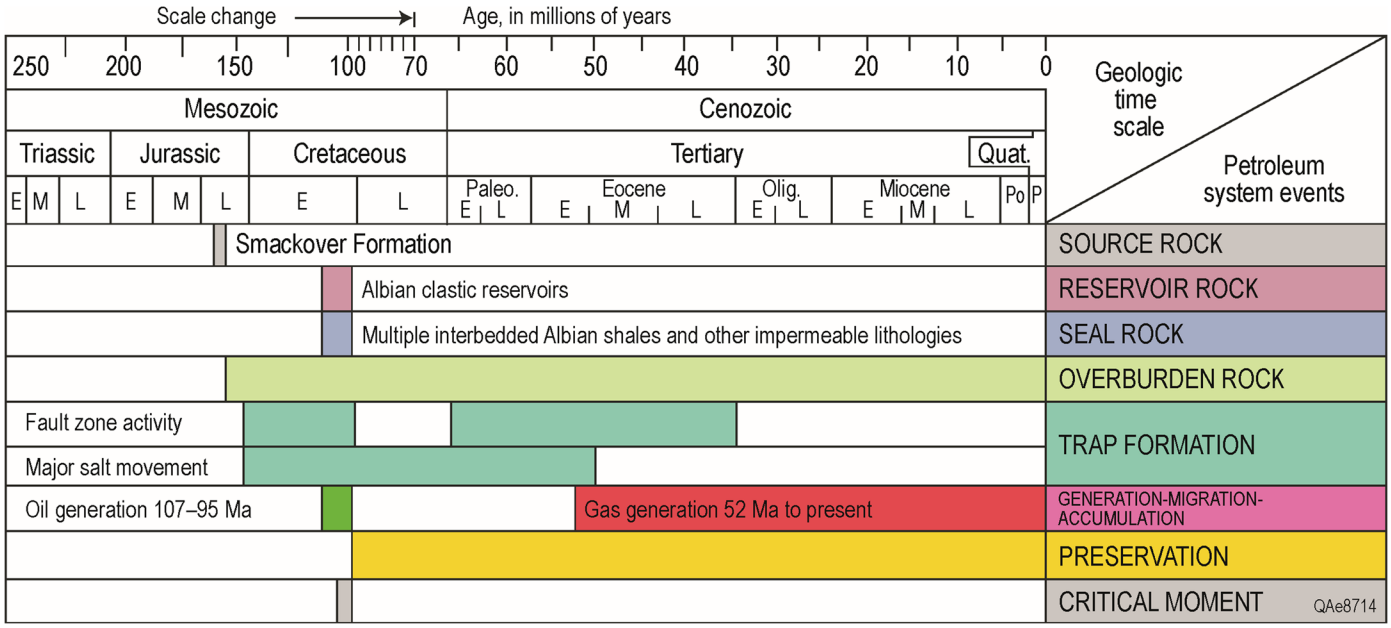


Figure 9. Generalized events chart for Smackover Formation hydrocarbons in Albian clastic reservoirs (from Merrill, 2016, courtesy of the U.S. Geological Survey). More detailed chronology for the Martinville Field area is shown in Figure 8.

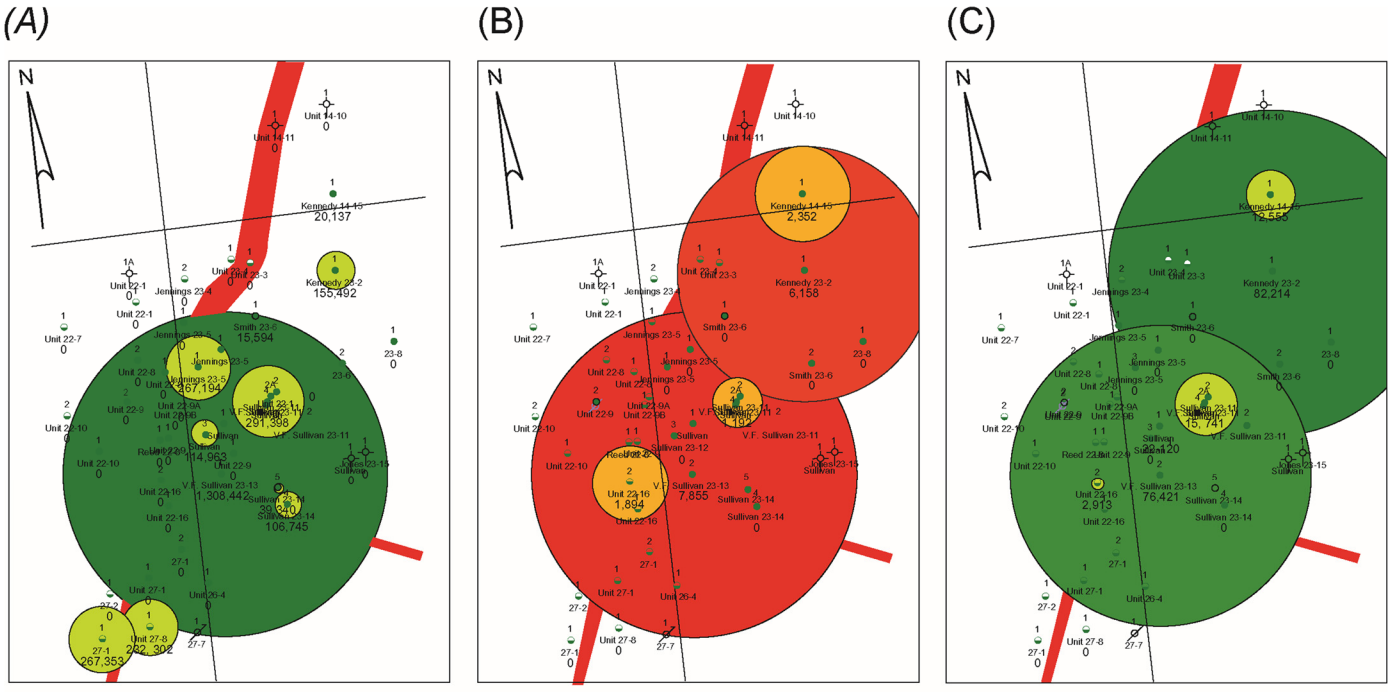


Figure 10. Bubble maps of 3 of the 5 parameters for the participation agreement proposed. (A) Cumulative production (25%) from the Rodessa (agreement production through October 1995; production shown through December 1995). (B) Production rate for the most recent 4 months (10%), July 1–October 31, 1995. (C) Estimated remaining reserves (35%). Green bubbles are static production figures (cumulative production and estimated remaining) and red bubbles are production rates over given periods of time. Bubble maps where scaled for each individual map and not relative size among maps. The focus is on the Rodessa production in the northeastern fault block. Sligo production is shown for two wells in the southern and southwestern fault blocks on the southern portion of (A) only.

models because they are not only the basis of all analyses and analytics, but also once understood bivariate, multivariate, and more complex models are much easier to build and understand by the architect (geoscientist, engineers, and data scientists) and target audience (managers or investors) for whom the models are to be communicated. Understanding the terminology and meaning of these models, makes the conceptual and intangible better understood.

The available data in the evaluation are a subset of the population used to describe the population through statistical infer-

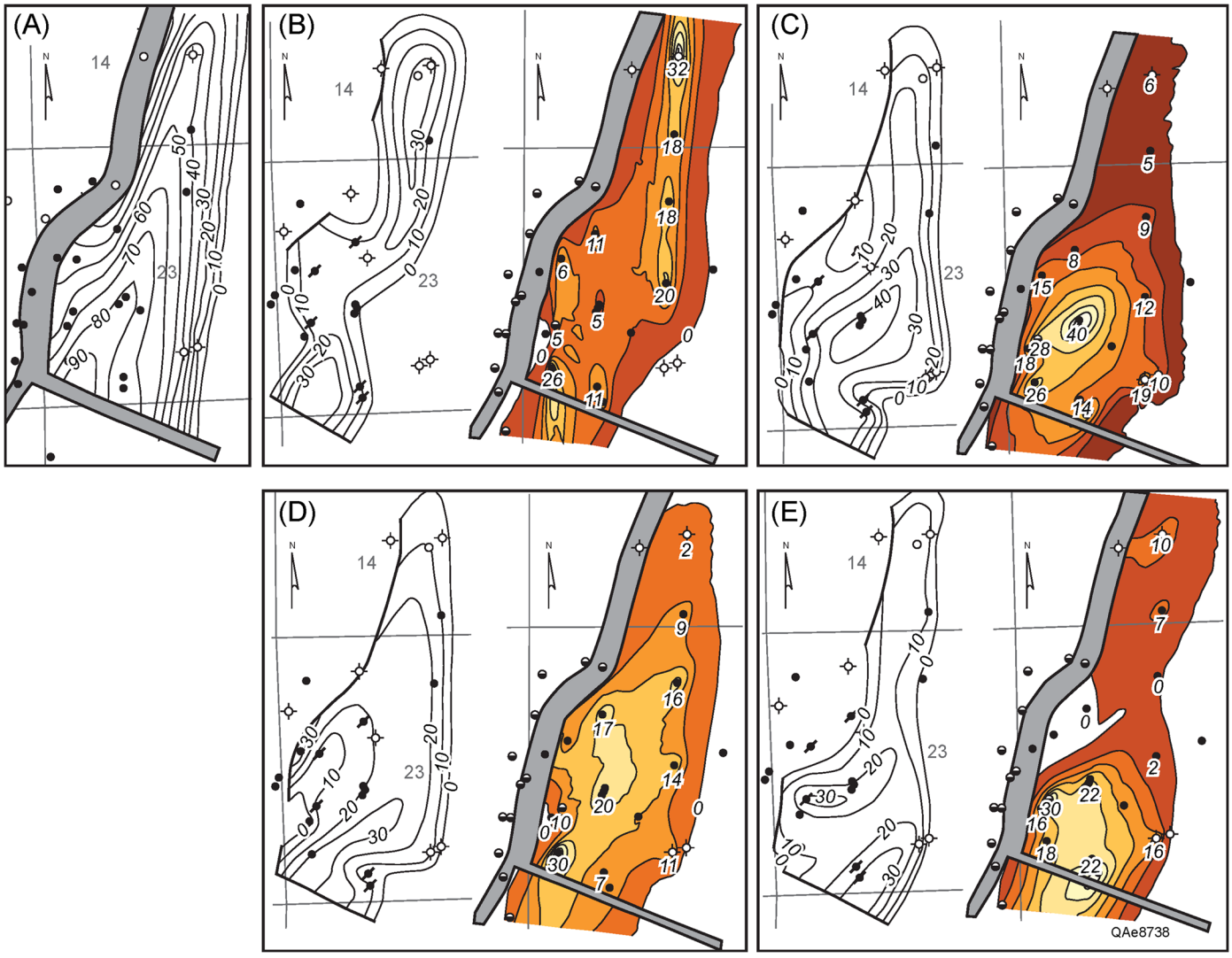


Figure 11. Rodessa sandstone isopach maps, with black & white versions from 1996 (10 ft contour interval) and color versions revised in 2020 (5 ft contour interval) with the color revised versions also including data from the Smith 23–6 #2 well drilled and completed after the 1996 black & white interpretations were made. (A) Net Rodessa sandstone reservoir isopach maps of sequences 3–6. (B) Net reservoir sandstone isopach maps, sequence 3. (C) Net reservoir sandstone isopach maps, sequence 4. (D) Net reservoir sandstone isopach map, sequence 5. (E) Net reservoir sandstone isopach maps, sequence 6.

ence. In standard, normal statistical inference, the samples must be unbiasedly sampled or collected, use a consistent and efficient sampling method, be of sufficient sample size, be randomly selected, and be normally distributed from the normally distributed continuous population being defined and described. Most of the datasets collected and used in this study do not meet all these requirements; they are imperfect. It is important to be aware and understand these imperfections when using statistical inference in attempts to model and define the population sampled.

The sidewall core laboratory measurements of porosity for the Rodessa sandstone reservoir, Martinville Field (Fig. 14), provide an example. Test for normalcy include the three measures of central tendency (mean, median, and mode) each at or close to the same value. In this case the mean and median are both at 16% porosity, whereas the mode is at 22% porosity. In a standard, normal plot, each of those measures should be at or near a single, central measurement point. The inflection point (red circles outlined in black in Figure 14B) of the frequency plot should be at +/- 1 standard deviation but are stretched out because of the flat nature of the mid-range of the frequency plot. All of the measurements are within three to four standard deviations, which is positive (no exaggerated tails). Given the sample size, symmetry of the curve and using standard statistical inference, likelihood or confidence interval that the population mean is captured by this sample mean is defined by the following. There is a 99% confidence interval that the population mean is between 14.6% and 17.4% porosity and at the 80% confidence interval that the population mean is between 15.3% and 16.7% porosity (Daniel, 1977). In medical sciences and pure sciences (physics, chemistry, and mathematics; these are non-complex sciences [McAlester and Hay, 1975]), 99% confidence intervals are typically used. In complex sciences like geology, an 80% confidence interval is more typical. Given these, the sample is not perfect but useful. Using sample size and standard deviation, it is determined that this sample is within 1% of the true population with a 90% confidence interval but not a 95% confidence interval. Through interpolation it is estimated to be closer to a 93.2% confidence interval (Daniel, 1977).

Why is this expected standard, normal sample distribution not more perfectly normal? In this case, it is most likely due to sampling error. That is, when sampling, cutting, and collecting sidewall cores, the samples were not unbiased, random, consist-

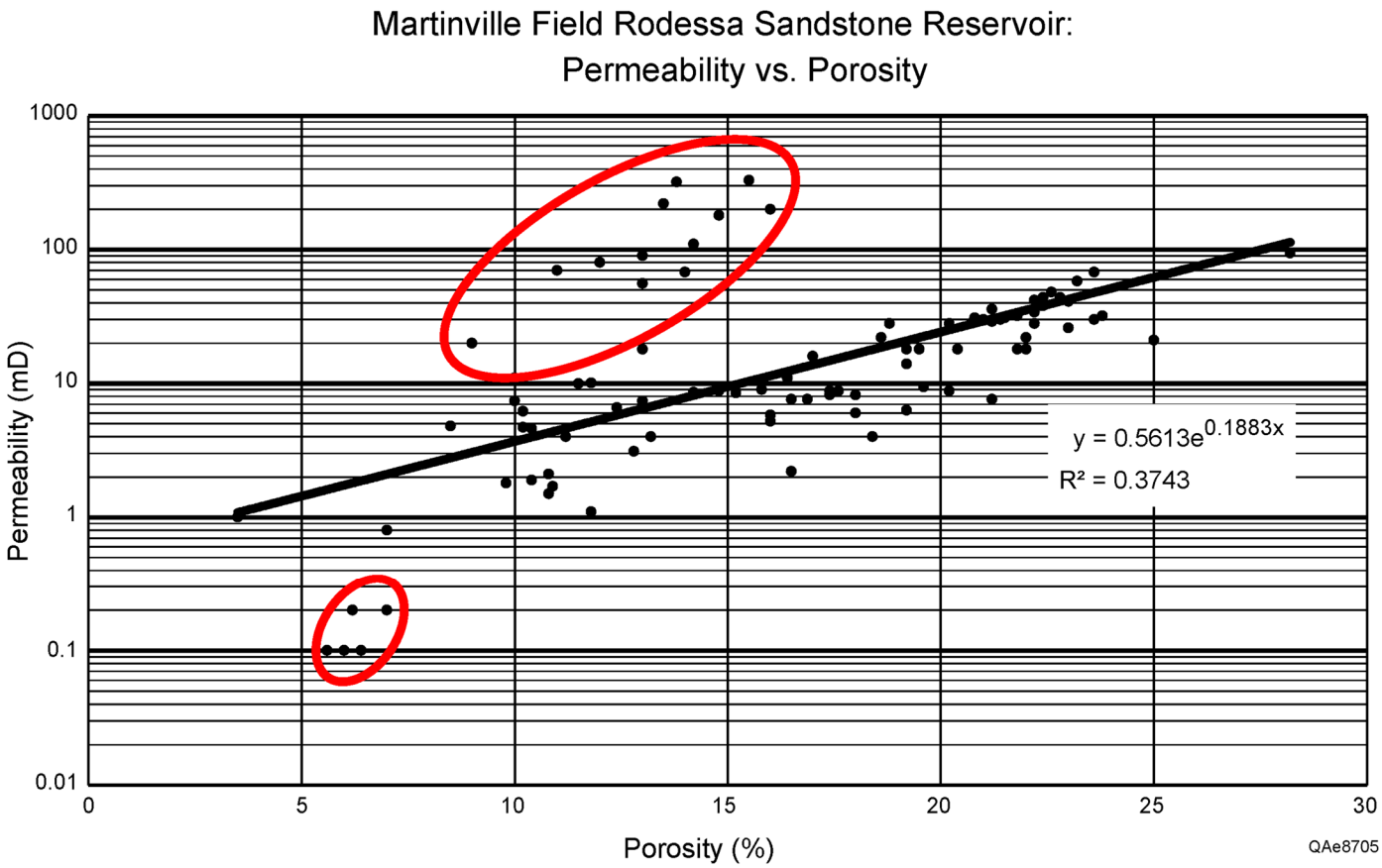


Figure 12. Total Martinville Field Rodessa sandstone reservoir permeability and porosity sidewall measurements. The regression model is statistically valid at the 99% confidence interval. There are two clouds of data measurements with slightly high and low observations (clouds outlined in red).

ently and efficiently selected. Often when sidewall core points are selected, there is an oversampling of higher porosities, undersampling of low porosities, and the method of selection is not consistent. That is, part of the reason why the mode in Figure 14 is at the high end of the porosity range (22%) and not closer to the sample mean and median (16%), thus causing a skewing to the left (longer left tail) because of the sampling error. It does not mean that the selected variable cannot be used, just that the investigator should be aware of the distribution parameters and aware of the bias.

The histogram and frequency plot of permeability (Fig. 15A) illustrates another common observation of the variables in geologic analyses. Permeability is not normally distributed; it is log-normally distributed as are many variables in geoscience and in nature. Why are there so many log-normally distributed naturally occurring variables? It is because these are typically not a simple basic measurement but a product of two or more variables. Porosity described above is a simple static parameter and follows a standard, normal distribution. Permeability measurements are not. Permeability is a dynamic measurement of the movement of a fluid or gas through a pore throat. The measurement is dependent on the type of fluid or gas used, temperature, pressure and possibly other variables (e.g., time and sample size measured against a standard). Another common example is rainfall which is not only dependent on relative humidity, but also temperature change, pressure, wind, and other variables. Fortunately, simple transformations, in this case transforming the measurements into log-normal distribution (Fig. 15B), create the standard, normal distribution that lends itself to standard, normal statistical analysis and treatment. This highlights the importance of reporting and reviewing the simple univariate plot of each variable. With practice and expertise, observations of log-normal, quadratic, and other transformation methodologies gain clarity. The transformation and use of these transformed measurements enhances the application of that variable in standard, normal statistical models and practice.

This parameter becomes even more compact or a tighter statistically and a more standard, normal distribution is obtained by truncating measurements more than three standard deviations from the mean (Fig. 15C). Such data management is useful and correct if used consistently for individual variables and for all variables. It is also important to know the measurement accuracy of the equipment being used. At the time the samples were taken and measured (1958–1990) were measurements of permeability in the 1 millidarcy range as accurate as in the 10–100 millidarcy range? Low or high measurements may be eliminated due to knowledge of the measurements as well as pure statistical reasons. Notice that the measures of central tendency move to the center of the distribution and the variability is reduced (standard deviation is reduced from approximately 0.52 (Fig. 15B) to 0.31 millidarcys (Fig. 15C).

Reviewing Figures 15B and 15C provides a clue to additional analysis (statistical and technical domain expertise) that is needed. These univariate plots show a bimodal distribution. There are many sophisticated test and solutions to treating bimodal distributions statistically/mathematically (Zhang et al., 2003; Van der Eijk, 2001; Ashman et al., 1994; Ellison, 1987) for the data scientist but these are beyond the scope of this research. The significant application here is that bimodal or multimodal distributions are often an indication of the combination of more

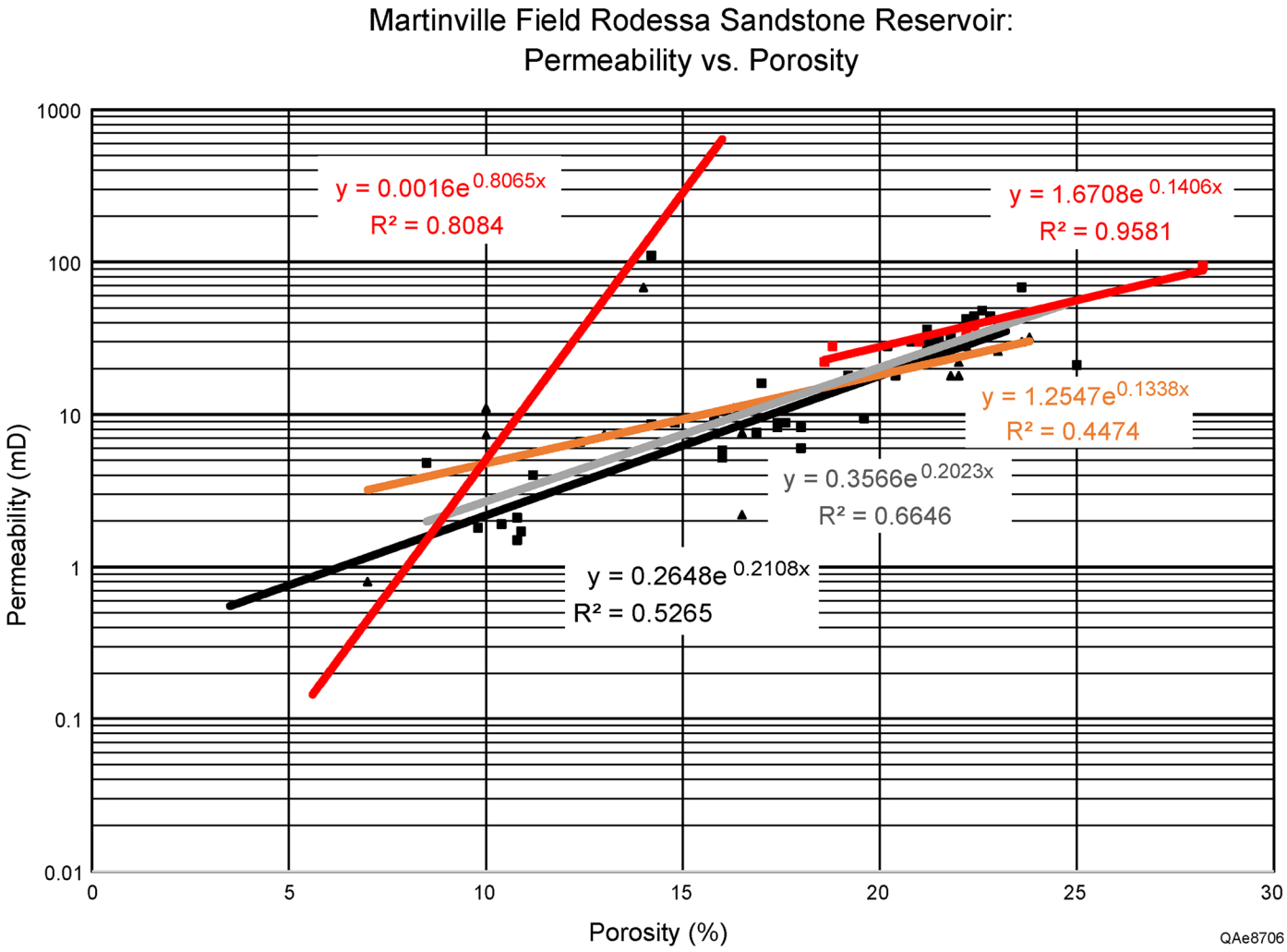


Figure 13. Sequence 3 linear model in red (data are red circles); sequence 4 linear model in gray (data are black squares); sequence 5 linear model in orange (data are black triangles); and sequence 6 linear model in black (data are black circles).

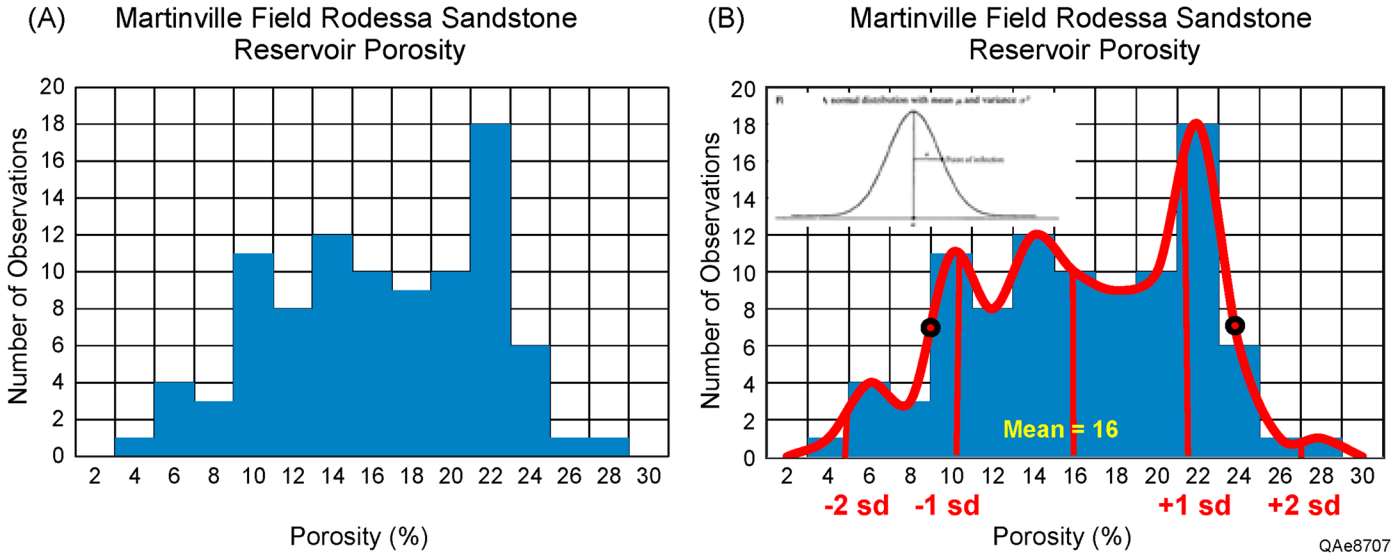


Figure 14. (A) Univariate histogram plot of porosity for the Rodessa sandstone reservoir, Martinville Field. (B) Univariate, histogram, and frequency plot of porosity for the Rodessa sandstone reservoir, Martinville Field

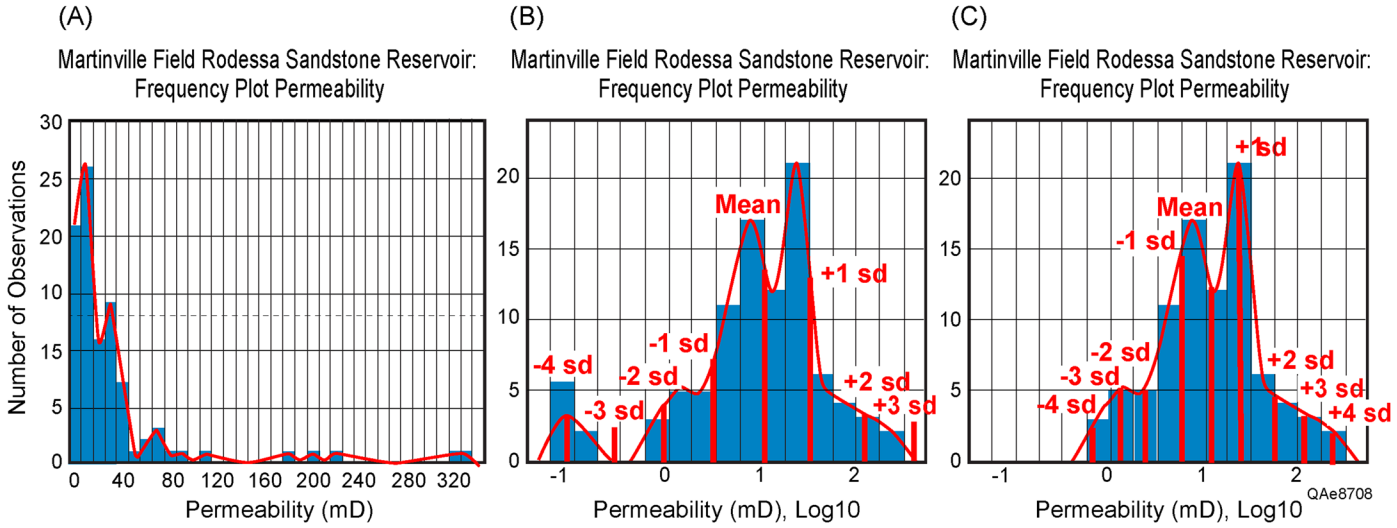


Figure 15. Histogram and Frequency Plot of the permeably of the Rodessa sandstone reservoir, Martinville Field. (A) Standard numeric plot, (B) Log normal plot, and (C) truncated log normal plot.

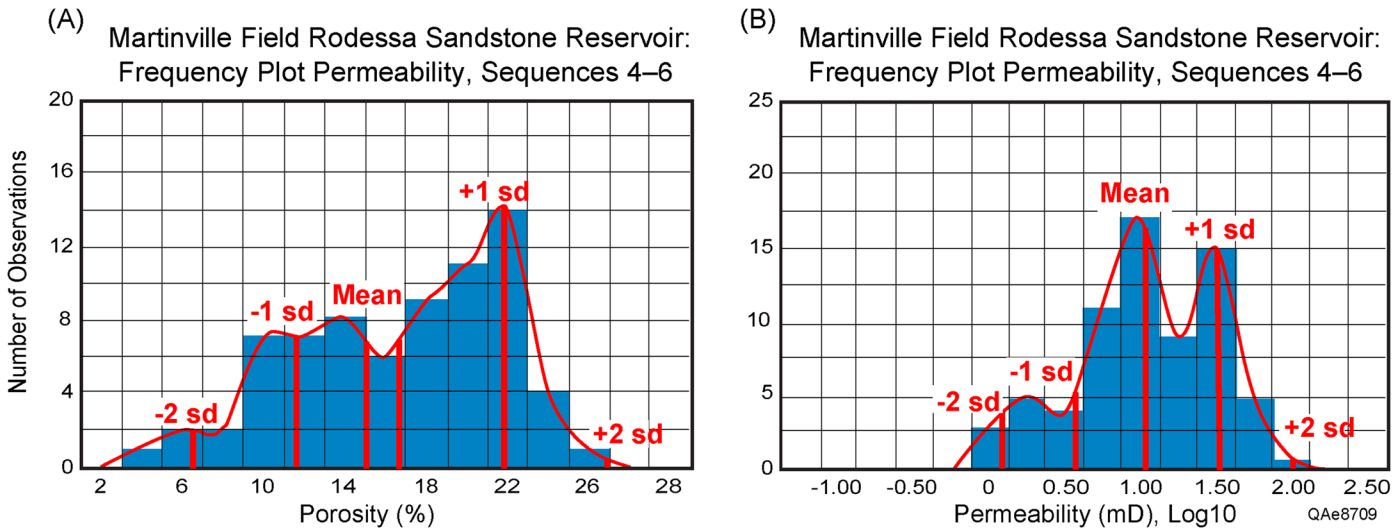


Figure 16. (A) Porosity histogram and frequency plot for Martinville Field, Rodessa sandstone reservoir sequences 4–6. (B) Permeability histogram and frequency plot for Rodessa sandstone reservoir sequences 4–6. Sequence 3 has been removed to test the two population distributions validity.

than one underlying population that should be separated (Chaudhuri et al., 2010; Sturrock, 2008; Lindsay, 2005; Schilling et al., 2002; Sambrook Smith et al., 1997; Wilcock, 1993; Eisenberger, 1964). The observations in Figure 15 are defined and described by the bivariate reporting and description in Figure 13 (and later with multivariate or multilinear regression). Once observed by the domain expert, rather than splitting those distributions with sophisticated statistical or mathematical solutions, the split can be made by recognizing the technical reasons for the bi- or multimodal distribution and related geologic reasoning. That is to determine if there are actually two or more separate populations represented.

In this case it is the higher permeability distribution of sequence 3 compared to sequences 4–6 (Fig. 13). Separating sequence 3 from the population of hydrocarbon productive Rodessa sandstone porosity and permeability are illustrated in Figure 16. Figures 16A and 16B are at the same scales as Figures 14 and 15 for ease in comparison. The resulting porosity distribution (Fig. 16A) reduces the multi-modality of the distribution, but results in a more skewed distribution; again, likely from selection bias. The selection process was heavily weighted toward samples in the 17–23% porosity range over samples in the 6–16% range. The range and bimodality of permeability (Fig. 16B) is reduced supporting the separation of sequences 3 from sequences 4–6. The range is reduced from Log(10) 0–2.5 or 1–316 millidarcys to Log(10) 0–2 or 1–100 millidarcys. Two-thirds of the permeability measurements in sequence 3 are Log (10) 1.5 (32 millidarcys) or greater but only 25–31% of sequences 4–5 and only 14% of sequence 6 has permeability in that upper range. The remaining bimodality is again likely bias in selection. Other technical separation tests did not prove to be justified or statistically valid. The original bimodality indicated or provided evidence of multiple, different populations captured and illustrated in the univariate plot, the geological reasoning and scientific investigation sup-

ported the separate treatment of those populations. Additional treatment and separation follows with development and investigation of multivariate and multilinear regression models.

The critical point for these univariate statistical measurements and plots is to review the data to determine its applicability to treatment as a standard, normal statistical variable. If it is not a standard, normal distribution, then visual observation of the plotted variable may determine if a simple transform can be applied to convert the measurements for standard, normal statistical inference and application.

Basic univariate (Figs. 14–16) and bivariate (Figs. 12 and 13) statistical analysis applied to the Rodessa sandstone hydrocarbon reservoir in the Martinville Field are important descriptive, reporting, and analytical modeling steps (Fig. 17) completed and plotted using basic Excel spreadsheets and plots. These steps in current machine learning and AI are often underutilized and/or underreported. Those analyses are missing these steps and may be using more-sophisticated statistical methodologies without realizing or reporting the potential statistical, analytical errors and moving toward the black box area of Figure 1. The steps shown and reporting here make additional, more sophisticated models and solutions move into the white box, understood, region of analytics, more applicable for the domain expert and petroleum geologist, and are often superior statistical models that also make it easier to communicate to the intended, audience.

Descriptive Analytics, Multivariable Analytics, and Segmentation

With these data, information, and reported univariate and bivariate statistics, we return to the geologic research question: is there any way to define and communicate the potential for each 40 ac unit separately and differently than the single reservoir model for the Rodessa sandstone reservoir using available data and information? The goal is to develop a descriptive analytic model using multivariate analysis to more thoroughly and accurately define the productive potential of each 40 ac tract. In developing and communicating these models, the goal is to develop a more sophisticated solution (x axis of Figure 17) with a greater business impact (y axis of Figure 17).

What is to be modeled is the estimated ultimate recovery for each tract; future production and value, more specifically the 8 tracts with wellbores used to estimate cumulative production, current production, and remaining reserves (value) for the unitization agreement. The estimated ultimate recoveries and estimated remaining production (value) for available wellbores and units were provided by reservoir engineering using standard decline curve analyses. Combining the geologic and engineering variables a first model (model #1) was developed to predict estimated ultimate recovery by total Rodessa sandstone thickness (h, Fig. 11A), year of first production, and structural subsea elevation of the base of the Ferry Lake Anhydrite / top Rodessa sandstone (Fig. 4B). These multilinear regression models were developed in 1996, revised during 2017 using SAS and Minitab, and confirmed using R in 2020.

$$\text{Model \#1} = 36{,}257{,}308 + (16{,}257 * \text{NetSd}) + (11{,}024 * \text{Yr1stProd}) + (5273 * \text{subsea z value})$$

This model has a multiple r^2 of 0.686, adjusted r^2 of 0.4505, F–statistic of 2.913, and p–value of 0.1642. Most reported research incorrectly uses r^2 values (additional explanation is significant and needed but not provided here). The F–statistic is more significant for these multilinear regression models. The F–statistic here does not have a high level of statistical significance (confidence interval is approximately 83.58%; Documenta Geigy, 1970) meaning that just using total net Rodessa sandstone reservoir thickness, structural level, and time variables do not provide a very strong statistical explanation for the estimated ultimate recovery estimates. This model is likely too generalized

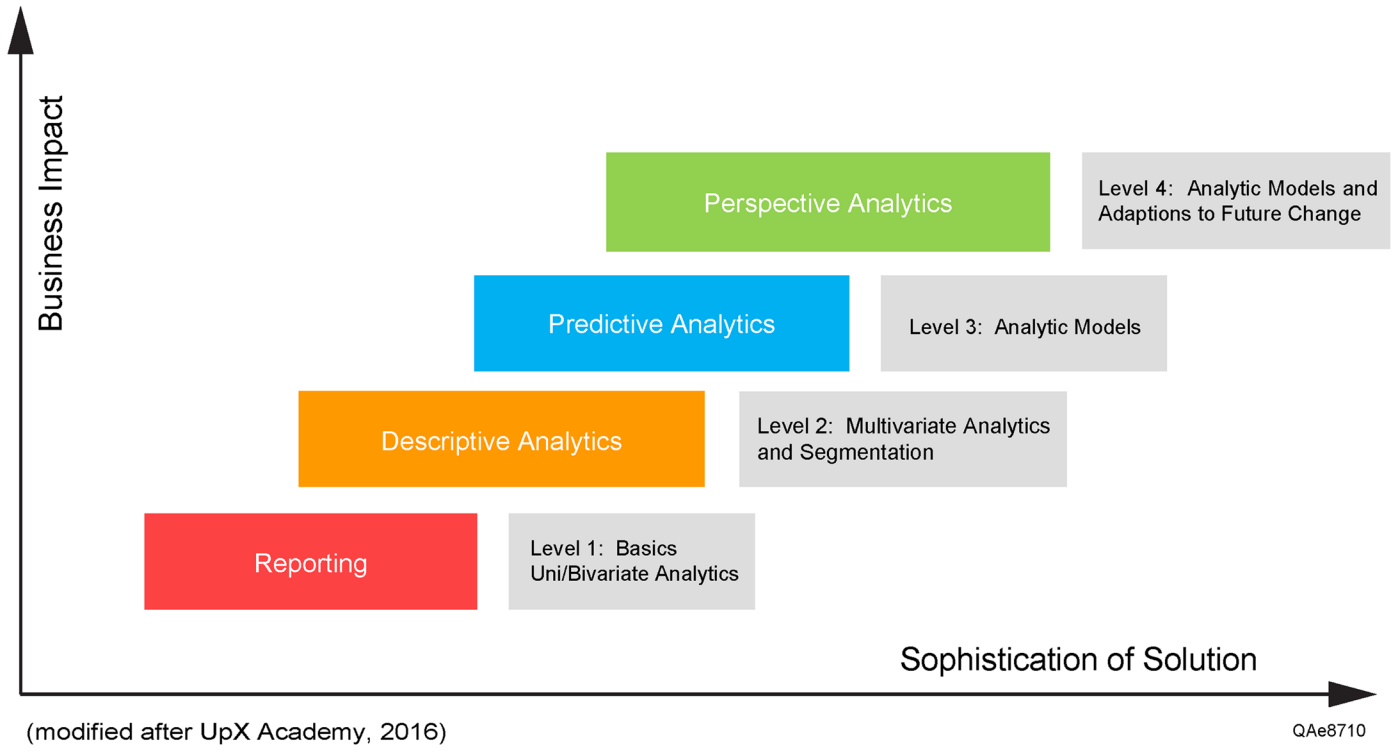


Figure 17. Plot of the business impact of statistical models by the sophistication of the solution.

without enough well-defined variables or observations to accurately model estimated ultimate recoveries. An 80% confidence interval for geologic studies may be acceptable but because better models were developed this model has been eliminated.

Rather than using total net Rodessa sandstone thickness, model #2 uses the thickness of each sequence separately (NetSd6 = net sandstone thickness of sequence 6, etc.):

Model #2 = 27,514,044 – (25,277 * NetSd6) – (26,556 * NetSd5) + (45,585 * NetSd4) + (89,221 * NetSd3) – (14,503 * Yr1stProd)

This model has a multiple r^2 of 0.9972, adjusted r^2 of 0.9903, F–statistic of 144.4, and p–value of 0.006894. Statistically valid at the 99.5% Confidence Interval using r^2 and the F–statistic (Documenta Geigy, 1970). However, this is a classic case of overfitting. Overfitting in this case occurred because the model contains a high number of variables relative to the number of observations (Everitt et al., 2010). Overfitting of a model occurs when "an analysis which corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably" (Lexico, 2020). This can also mean the statistical model is failing to be a good statistical inference for the population being modeled. The "noise or random fluctuations in the training data is picked up and learned as concepts by the model … [and] is more likely with nonparametric and nonlinear models that have more flexibility when learning a target function" (Browniee, 2016). So, a warning with nonparametric and nonlinear black box models (Fig. 1). The domain experts should be aware of overfitting and review all models with data scientists.

Overfitting in this case was identified because of the extremely high statistical fit of the model, negative constraints associated with items that should have positive constraints (NetSd6 and NetSd5), and resultant negative estimated ultimate recovery for one of the wells and units, not a possible outcome or predictor value. Because there are only 8 wells in the dataset providing only 8 observations, too many variables were included for accurate statistical inference.

The failed model does, however, provide useful information. The overfitting by having too many variables for the number of observations together with the observation having better production with a well-developed sequence 3 sandstone reservoir is supported by the higher permeability in that sequence at a given porosity relative to sequences 4–6 (Fig. 13). From these, model #3 is developed to use fewer variables, honoring those relationships, and using the separation developed during description and reporting of univariate statistics and geologic, domain expertise, observation (Table 1):

Model #3 = 64,825,519 + (9,639.5 * NetSd456) + (54,320.5 * NetSd3) – (1270.6 * Yr1stProd) + (5687 * subsea z value).

NetSd 456 is the combined subsets of sequences 4–6 with NetSd3, sequence 3 separated. Model #3 multiple r^2 is 0.9963, adjusted r^2 is 0.9913, F–statistic is 200.1, and p–value of 0.000569. A very good statistical fit, no erroneous directional constraints, and individual variable statistical significance codes at the 99% and 99.5% significance level (excluding Yr1stProd, see asterisks in variable line and explanation in Table 1). The small or tight statistical measures of the individual variables within the model (not provided by r^2 measurements, but provided separately with model output; Table 1) support a good, not overfitted model (Burnham et al., 2002).

What does this analytical, machine learning model #3 tell us that was not known from previous geological and reservoir engineering evaluation? Most significantly, it supports, qualitatively, the observations in Figure 13 that the higher permeability-porosity relationships observed in sequence 3 should be separated from the permeability-porosity relationships in sequences 4–6. These sequences can be defined and separated for evaluation as concluded from statistical and geologic observations from the univariate plot and analyses and multivariate / multilinear modeling of estimated ultimate recovery. If this were not the case,

```
lm(formula = EUR ~ NetSd123 + NetSd4 + YrProd + z, data = df[,
    .(NetSd123 = (NetSd1 + NetSd2 + NetSd3), EUR, YrProd, NetSd4,
        z)])

Residuals:
     1      2      3      4      5      6      7      8
 -8721  15218  50298 -32392  10700 -33337  -8061   6295

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 64825519.0  7710485.5   8.407 0.003530 **
NetSd123        9639.5      895.7  10.762 0.001716 **
NetSd4         54320.5     2537.2  21.409 0.000223 ***
YrProd         -1270.6     1704.6  -0.745 0.510115
z               5687.0      806.6   7.051 0.005864 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 41700 on 3 degrees of freedom
Multiple R-squared:  0.9963,    Adjusted R-squared:  0.9913
F-statistic: 200.1 on 4 and 3 DF,  p-value: 0.000569
```

Table 1. Multivarient / multilinear regression model for net sandstone reservoir sequences, year of first production, and structural position of the Rodessa production, Northeast Martinville Field fault block. Note: there was a different numbering system at the time these models were run than the numbering system used in this paper. Previously, the sequences were numbered (1–6) top to bottom as encountered by drilling; subsequently, they were numbered geologically more correctly in the proper sequence-stratigraphy, order of deposition. NetSd123 and NetSd4 in this table and as run during statistical, multilinear evaluation are sequences 6, 5, 4, and 3, respectfully.

model #3 statistical results would look much like model #1. Because that separation did provide for a more valid and reliable model, the treatment of sequence 3 separate from sequences 4–6 was validated.

The addition of 1 ft, more or less, of reservoir sandstone in sequence 3 has a 5.6 times greater effect on estimated ultimate recovery than a 1 ft change in sequences 4–6 or a 2.4 times greater effect on a percentage basis. Having identified and provided proof of the separation of these sequences and differences in productive potential of each sequence, explains the relative differences in production observed in 1996. These are the Sullivan 23–13 #2 well, 1.3 MMBO production since 1958, the 155 thousand barrels of oil (MBO) in the Kennedy 23–2 well in just 4 yr, more than 3 decades after field discovery, and better production in the Jennings 23–5 well (267 MBO) and Sullivan 23–11 well (291 MBO) than the other three wells in better structural positions in the field (average production of 79 MBO). The year of first production and structural position are contributing variables to the model. The ordered weight of the variables in this model are: NetSd3, Yr1stProd, NetSd456, and structural position. This modeling and proof was not accomplished by geologic or analytical investigation separately but by the interaction of the two domains during scientific investigation and exploratory data analyses (EDA) to analytical modeling interactive processes.

Model Testing

Three types of model testing are presented. First is the statistical validity and confidence intervals defined with the descriptions of the models. Second is to randomly remove small datasets from the model, redevelop models on the majority of the data, and test how well the model defines the data removed or how little the model is changed when the data are reincorporated into the model. This method will be shown later when a new well is drilled subsequent to the initial study in 1996. Any finally, third, is to look back at how well the model did with the test of time. Because these models were first developed in 1996, it is possible to see how well or not these models predicted future production (value) almost a quarter of a century later. This is perhaps the best of the 3 test methodologies, but is typically not available when models are first developed.

Model #3 can be tested against the 3 metrics from the original research question: unitization percentages, estimated ultimate recovery percentages, and current production percentages, each meant to represent proportional contribution and future value to the field-wide, northeastern fault block of the Martinville, Rodessa Oil Pool. What is being estimated is the contribution in production and value of the southern two-thirds of the field compared to the northern one-third of the field. The first test will be made for the next 11 yr of production. The reasoning for the chosen time periods is explained later.

The unitization agreement proposed and accepted by the Mississippi Oil & Gas Board weighted Coho's interest in the southern two-thirds of the field at 79.9% of the field and the other companies on the northern one-third, 20.1%. Model #3 valued the future production 52.63% of the total fault block in the southern two-thirds and 47.37% in the northern one-third. A 23.27% difference in interpretations. Which was a better predictor with known information in 1996 (Fig. 18A) looking back at actual production during the next 11 years (production during 1996 through 2006)? Actual production from known producing wells and units from 1996 through 2006 was 59.15% in the southern two-thirds and 40.85% in the northern one-third. The unit agreement overestimated the southern two-thirds contribution and underestimated the northern one-third contribution by 20.75%. Model #3 underestimated the southern two-thirds of the field contribution and overestimated the northern one-third of the field contribution by 6.55%. The errors were in the opposite direction, but model #3 was 3.2 times more accurate (20.75% / 6.55%) in estimating future value than the arbitrary unitization agreement as proposed and accepted. At the level of the individual units, there was a weighted average error of only 3.27% between model #3 constructed using the geologic, production, and field information available in 1996 in predicating the percentage contribution of each known producing well and 40 ac unit for the next eleven years (1996–2006).

Activity in the field did not remain static. Later during 1996 after the initial study was completed, an additional well was drilled in the central portion of the field, the Smith 23–6 #2 well (Figs. 18B and 18C; and included in the color isopachs of Figures 11B–11E). The new Smith well and open-hole log interpretations provide another test for model #3. Using geologic subsurface data from log (Fig. 18B) interpretation, timing, and the analytical equation developed using multivariate, multilinear regression, model #3 predicted an estimated ultimate recovery for the new Smith 23–6 #2 well of 178,797 BO. The well produced 164,344 barrels of oil (BO) from late 1996 through 2006. Model #3 estimate is within 8% of the actual production over the next 11 yr—an excellent test of the model. The Smith 23–6 #2 well was converted to a salt-water disposal well once depleted to support production via water-drive for the Kennedy 23–2 well, 1320 ft to the north from 2007–2015 as described below.

Denbury Resources, Inc. purchased the field in 2002. Beginning in 2006–2007 changes Denbury made to the water injection program produced another 267,460 BO in the Kennedy 23–2 well representing 51% of total field production during 2007-2019 in the northern one-third of the field. In the southern two-thirds of the field, 225,028 BO was produced from the Sullivan 23–13 #2 well and 27,877 BO from the Jennings 23–11 well together representing 49% of the total field production during 2007–2019 (Fig. 19A). These are within 4% of the proportional field models that are relatively simple geologic and engineering variables and analytic models developed nearly quarter century earlier by model #3.

SUMMARY AND CONCLUSSIONS

The oil and gas industry transformation involving analytics, machine learning, and AI is underway. The petroleum geologist using current skills or skills they are well suited to obtain are capable of being a valued part of this transformation.

Data scientists working in independent oil and gas firms and service organizations have the statistical and mathematical skills to accomplish the analytical modeling of projects. They often lack the industry specific knowledge, skills, and industry experience to provide successful application to real-world oil and gas application. This provides an opportunity for the petroleum geologist to share and train data scientists with our methods, processes, and specific variables in this highly technical industry and applications. With knowledge of analytics and machine learning, the petroleum geologist and other industry domain experts can not only include analytics in their own technical work, but can also provide and fully perform the data science function within their own workflows through to project completion.

The same skills and tools that geologists use in observation, measurement, discovery, insight, and connecting variables and information from multi-disciplinary nature of the petroleum geologist's science are the same skills needed to make analytics and machine learning useful and successful. The process is best when it is an interactive sharing of observations for the discovery of the relationships between the petroleum geologist and data scientist or petroleum geologist knowledgeable in the use of analytics, using analytics themselves as one of their technical tools. This study has documented the understanding and relationships between Rodessa sandstone environments of deposition, facies, sequences, reservoir quality variables, and using geology to define the productivity by sequences through data analytic models. Without superior technical analysis, superior analytical analysis

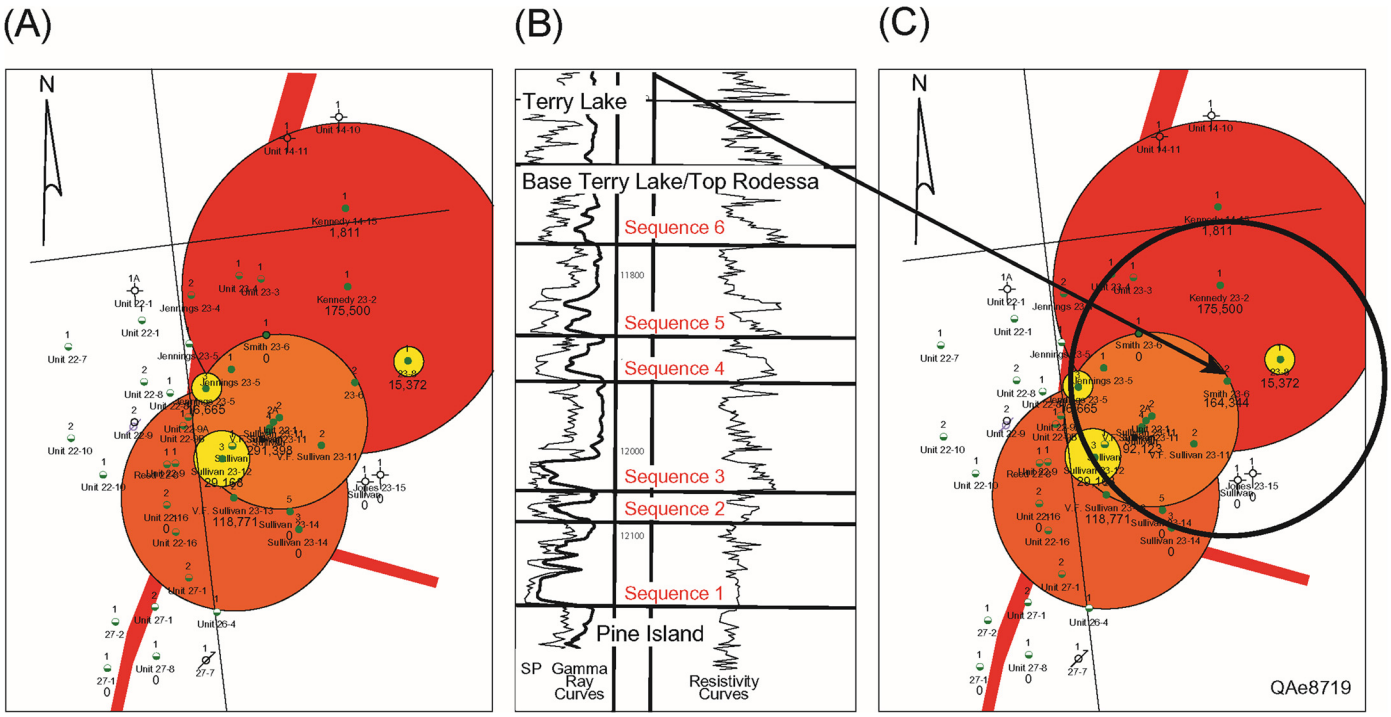


Figure 18. Production 1996–2006. (A) Production 1996–2006 in known wells and units from early 1996. (B) Smith 23–6 #2 drilled and completed late 1996. Note the well-developed reservoir sandstones sequences 1-3, good SP (solid line) and lean Gamma Ray (dashed line) development and separation good porosity and permeability shown by the separation of the medium (dashed line) and deep resistivity (thin solid line) curves. Note the negative deflection and low resistivity, water-wet porosity in sequence 1, developing transitional to oil wet in sequence 2 and well-developed sandstone reservoir and oil wet, positive deflection and 10 times higher resistivity. Deep resistivity is 2 ohms in sequence 1 and 20+ ohms in sequence 3 (oil productive). Sandstone reservoir quality decreases up section. (C) Location and bubble symbol of the Smith 23–6 #2 well production, 1996–2006.

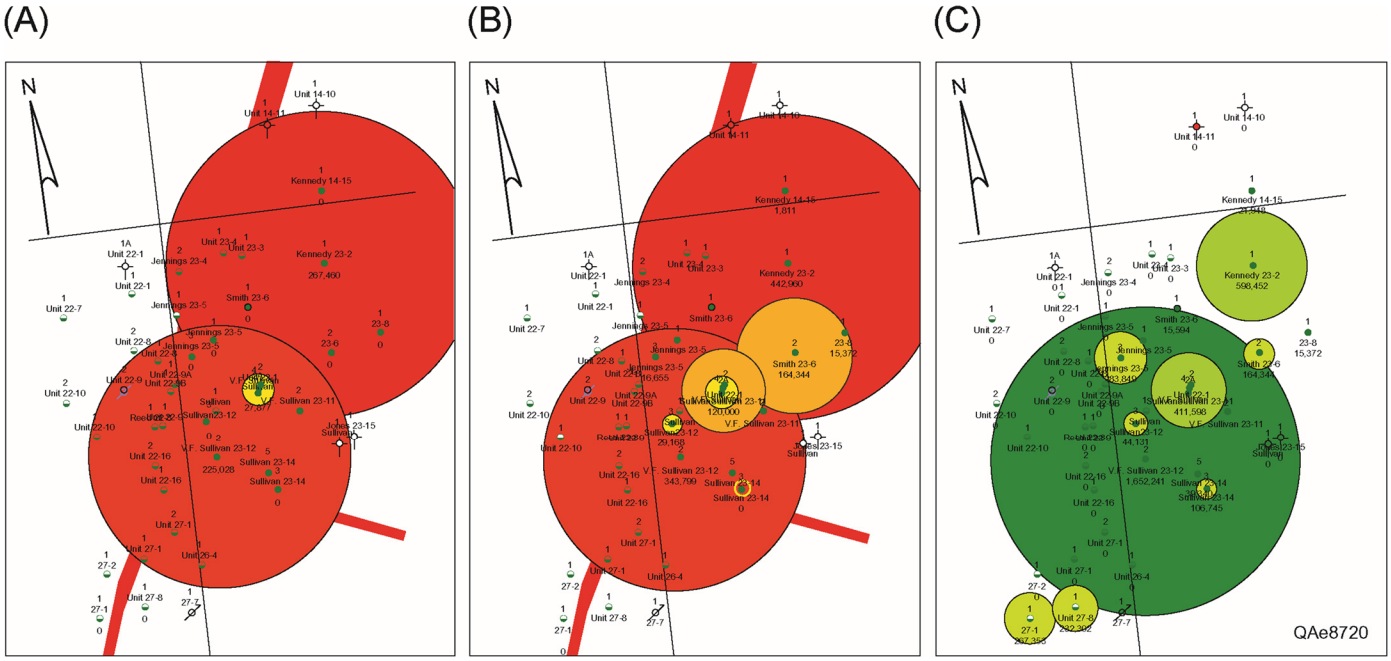


Figure 19. (A) Bubble map of total Rodessa sandstone reservoir production 2007–2019 (last 12 yr). (B) 1996–2019 (last 24 yr). (C) Cumulative field production from discovery through December 31, 2019 (62 yr).

will not be achieved. Analytics without the petroleum geologist and other domain experts input will not be as useful to the organization, industry, and management as investors and organizations need it to be.

The example of past and future production (value) from the Rodessa sandstone reservoirs in the northeastern fault block, Martinville Field, Simpson County, Mississippi, was used to illustrate:

(1) Advancement of geological understanding of the trap; reservoir; top, bottom and lateral seals; source; and migration to support the enhanced geologic interpretation and understand of production from these Rodessa sandstone reservoirs.
(2) Using that enhanced understanding to interactively work with statistics and analytical tools to further define the geological models and understanding of porosity, permeability, and the relationships with hydrocarbon production by sequence.
(3) Finally, building a complete analytic, machine learning statistical model to explain past and future hydrocarbon production and potential. That model and individual variables were validated using standard statistical tests, accurately predicting future wells performance, future performance and value of the separate areas of the field down to the individual well and 40 ac units. These models provided analysis in 1996 that proved to be accurate to within 4% by 40 ac units and full-field through the end of 2019, nearly a quarter-century later. Looking back, it passed the test of time using technical variables to predict future outcomes.

The new analytical, machine learning, AI model is successful. It was successful in the overlapping intersection of domain expert and data scientist. Marathon Oil Company understood and valued the advancement of technical understanding provided by the analytics and machine learning multivariate and multilinear regression models and how it could be used to evaluate and predict future field performance.

ACKNOWLEDGMENTS

I would like to acknowledge the financial support and time to complete this research provided by the consulting practice, clients, and producing properties of Riverford Exploration, LLC, and original work completed while working for Marathon Oil Company. The Bureau of Economic Geology, Jackson School of Geosciences, University of Texas at Austin, provided financial and publication support for this work. Review and recommendations on the statistical models were provided by Dr. Frank Male, Department of Petroleum and Geosystems Engineering, University of Texas at Austin, and are most appreciated. Thank you, Robert Merrill, *GCAGS Journal* Editor, and James Willis, GCAGS Managing Editor, for review, suggestions, and support of the work and manuscript.

REFERENCES CITED

Ambrose, W. A., T. F. Hentz, R. Bonnaffe, R. G. Loucks, L. F. Brown Jr., F. P. Wang, and E. C. Potter, 2009, Sequence-stratigraphic controls on complex reservoir architecture of highstand fluvial-dominated deltaic and lowstand valley-fill deposits in the Upper Cretaceous (Cenomanian) Woodbine Group, East Texas Field; regional and local perspectives: American Association of Petroleum Geologists Bulletin, v. 93, p. 231–269.

American Association of Petroleum Geologists, 2002, Correlation of stratigraphic units of North America (COSUNA) charts; Gulf Coast region correlation chart: Tulsa, Oklahoma, CD–ROM publication.

Anderson, E. G., 1980, Sulphur Bluff Field, Hopkins County, Texas, *in* F. A. Herald, ed., Occurrence of oil and gas in northeast Texas: East Texas Geological Society Transactions, v. 30, p. 35–39.

Anderson, M. L., 1989, Stratigraphy of the Fredericksburg Group, East Texas Basin: Baylor Geological Studies Bulletin 47, Waco, Texas, p. 1–38.

Ashman, K. M., C. M. Bird, and S. E. Zepf, 1994, Detecting bimodality in astronomical datasets, The Astronomical Journal, v. 108, p. 2348–2361.

Bagchi, S., 2020, What's derailing digital transformation in 2020?, <https://www.cxotoday.com/newsletter/whats-derailing-digital-transformation-in-2020/>.

Bendor-Samuel, P., 2019, Why digital transformations fail: 3 exhausting reasons <https://enterprisersproject.com/article/2019/8/why-digital-transformations-fail-3-reasons>.

Browniee, J., 2016, Overfitting and underfitting with machine learning algorithms, machine learning mastery: Making developers awesome at machine learning, <https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/>.

Burnham, K. P., and D. R. Anderson, 2002, Model selection and multimodel inference, 2nd ed.: Springer-Verlag, New York, 488 p.

Champlin, S. D., 2000, Simpson County, Mississippi, oil & gas production index map: Mississippi Office of Geology Open-File Report 74, Jackson, 1 sheet.

Chaudhuri, D., and A. Agrawal, 2010, Split-and-merge procedure for image segmentation using bimodality detection approach: Defense Science Journal, v. 60, p. 290–301.

Chenault, D. S., and L. L. Lambert, 2005, Sequence correlation of the mid-Comanche series, South Texas region: Gulf Coast Association of Geological Societies Transactions, v. 55, p. 71–88.

Clarke, R. W., 1971, Einstein: The life and times: The World Publishing Company, New York and Cleveland, Ohio, 718 p.

Claypool, G. E., and E. A. Mancini, 1989, Geochemical relationships of petroleum in Mesozoic reservoirs to carbonate source rocks of Jurassic Smackover Formation, southwestern Alabama: American Association of Petroleum Geologists Bulletin, v. 73, p. 904–924.

Coho Resources, Inc., 1995, Exhibit "D" tract participation formula, Martinville Rodessa Field Unit, Rosessa Oil Pool, Martinville Field, Simpson County, Mississippi to Unit Agreement dated December 12, 1995: Mississippi Oil & Gas Board Docket No. 3–96–114, Exhibit M–8, Jackson, 1 p.

Continental Oil Company, Central Oil Company 14–11 #1, 1955, Electrical log, core report descriptions: Mississippi Oil & Gas Board, Jackson, API 23–127–00027 <www.ogb.state.ms.us/welldata.php>.

Daniel, W. W., 1977, Introductory statistics with applications, Houghton Mifflin Company, Boston, Massachusetts, 475 p.

Davis, D. C., and E. H. Lambert, Jr., eds., 1963, Mesozoic-Paleozoic producing areas of Mississippi and Alabama, v. 2: Mississippi Geological Society, Jackson, 361 p.

Devery, D. M., 1982, Subsurface Cretaceous strata of Mississippi: Mississippi Department of Natural Resources Bureau of Geology and Energy Resources Information Series, Jackson, v. 82, no. 1, p. 1–24.

Documenta Geigy, 1970, Scientific tables, 7th ed., *in* W. W. Daniel, 1977, Introductory statistics with applications: Houghton Mifflin Company, Boston, Massachusetts, p. 441.

Eaves, E., 1976, Citronelle Oil Field, Mobile County, Alabama, *in* J. Braunstein, ed., North American oil and gas fields: American Association of Petroleum Geologists Memoir 24, Tulsa, Oklahoma, 360 p.

Eisenberger, I., 1964, Genesis of bimodal distributions: Technometrics, v. 6, p. 357–363.

Ellison, A. M., 1987, Effect of seed dimorphism on the density-dependent dynamics of experimental populations of Atriplex triangularis (Chenopodiaceae): American Journal of Botany, v. 74, p. 1280–1288.

Esposito, R. A., J. C. Pashin, and P. M. Walsh, 2008, Citronelle Dome: A giant opportunity for multizone carbon storage and

enhanced oil recovery in the Mississippi Interior Salt Basin of Alabama: Environmental Geosciences, v. 15, no. 2, p. 53–62.

Evans, R., 1987, Pathways of migration of oil and gas in the south Mississippi Salt Basin: Gulf Coast Association of Geological Societies Transactions, v. 37, p. 75–76.

Everitt, B. S., and A. Skrondal, 2010, Cambridge dictionary of statistics: Cambridge University Press, U.K., 480 p.

Fairhurst, B., 1996, Geologic and Statistical Exhibits, Marathon Oil Company's Exhibits for its Motion to Re-Open the Hearing and Notice of Contest, Plan of Unitization of the Martinville, Rodessa Field Unit, Martinville Field, Simpson County, Mississippi: Mississippi Oil and Gas Board Docket No. 3–96–114, Jackson, Exhibit M–10, 9 p., and Exhibit M–11, 2 p.

Forgotson, J. M., 1963, Depositional history and paleotectonic framework of Comanchean Cretaceous Trinity Stage, Gulf Coast area: American Association of Petroleum Geologist Bulletin, v. 47. n. 1, p. 69–103.

Frascogna, X. M., ed., 1957, Mesozoic-Paleozoic producing areas of Mississippi and Alabama, v. 1: Mississippi Geological Society, Jackson, 139 p.

Freedman, R., 2019, Executives pursue digital transformation projects despite 80% failure rate, survey finds, <https://www.cfodive.com/news/executives-pursue-digital-transformation-projects-despite-80-failure-rate/562148/>.

Geological Survey of Alabama, State Oil and Gas Board, 2020, online database, <http://gsa.state.al.us>.

Hinchcliffe, D., 2018, Digital transformation in 2019: Lessons learned the hard way, <https://www.zdnet.com/article/the-biggest-lessons-learned-in-digital-transformation/>.

Hoskin, T., 2012, Parametric and nonparametric: Demystifying the terms, <https://www.mayo.edu/research/documents/parametric-and-nonparametric-demystifying-the-terms/doc-20408960>.

Karges, H. E., 1968, Pelahatchie Field—Mississippi Giant?: Gulf Coast Association of Geological Societies Transactons, v. 28, p. 264–274.

Kesari, G., 2019, 5 reasons why analytics projects fail, <https://enterprisersproject.com/article/2019/10/why-analytics-projects-fail-5-reasons>.

Lindsay, R. S., 2005, The topography of multivariate normal mixtures: Annals of Statistics, v. 33, p. 2042–2065.

Mancini, E. A., P. Aharon, D. A. Goddard, M. Horn, and R. Barnaby, 2012, Basin analysis and petroleum system characterization and modeling, interior salt basins, central and eastern Gulf of Mexico, part 1: North Louisiana Salt Basin; part 2: Mississippi Interior Salt Basin; part 3: Tectonic / depositional history, resource assessment: American Association of Petroleum Geologists Search and Discovery Article 10396, Tulsa, Oklahoma.

Mancini, E. A., P. Li, D. A. Goddard, and R. K. Zimmerman, 2005, Petroleum source rocks of the onshore interior salt basins, north central and northeastern Gulf of Mexico: Gulf Coast Association of Geological Societies Transactions, v. 55, p. 486–504.

Mancini, E. A., J. Obid, M. Badali, K. Liu, and W. C. Parcell, 2008, Sequence-stratigraphic analysis of Jurassic and Cretaceous strata and petroleum exploration in the central and eastern Gulf Coastal Plain, United States: American Association of Petroleum Geologists Bulletin, v. 92, p. 1655–1686.

McAlester, A.L. and Hay, E.A., 1975, Physical geology: Principles and perspectives: Prentice-Hall, Englewood Cliffs, New Jersey, 439 p.

McFarlan, E., and S. L. Menes, 1991, Lower Cretaceous, *in* A. Salvador, ed., The geology of North America, v. J: The Gulf of Mexico Basin: Geological Society of America, Boulder, Colorado, p. 181–204.

Merrill, M., 2016, Geologic assessment of undiscovered conventional oil and gas resources in the Albian Clastic and Updip Albian Clastic assessment units, U.S. Gulf Coast: U.S. Geological Survey Open-File Report 2016–1026, 34 p.

Mississippi Oil and Gas Board, 2020, well database, <ogb.state.ms.us>.

Nunnally, J. D., and H. F. Fowler, 1954, Lower Cretaceous stratigraphy of Mississippi: Mississippi State Geological Survey Bulletin 79, Jackson, 45p.

Oehler, J. H., 1984, Carbonate source rocks in the Jurassic Smackover trend of Mississippi, Alabama, and Florida, *in* J. G. Palacas, ed., Petroleum geochemistry and source rock potential of carbonate rocks: American Association of Petroleum Geologists Studies in Geology 18, Tulsa, Oklahoma, p. 63–69.

Overby, S., 2020, 8 reasons why AI projects fail, <https://enterprisersproject.com/article/2020/3/why-ai-projects-fail-8-reasons>.

Lexico, 2020, Overfitting, <https://www.lexico.com/definition/overfitting>.

Pashin, J. C., 2013, Spaghetti with marinara: Tertiary oil recovery in a Cretaceous redbed succession, Citronelle Field, Alabama: American Association of Petroleum Geologists Search and Discovery Article 20207, Tulsa, Oklahoma, 46 p., <http://www.searchanddiscovery.com/pdfz/documents/2013/20207pashin/ndx_pashin.pdf.html>.

Petkovsek, C., 2019, Structural controls and depositional environments of the Glen Rose Subgroup in Pelachatchie Field in Rankin County, Mississippi: M.S. Thesis, University of Mississippi, Oxford, 104 p., <egrove.olemiss.edu/etd/1658>.

Philpott, T. H., 1960, Lower Cretaceous trend of South Arkansas, North Louisiana, Mississippi and Alabama: Gulf Coast Association of Geological Societies Transactions, v. 10, p. 276–278.

Salvador, A., 1991, Origin and development of the Gulf of Mexico Basin, *in* A. Salvador, ed., The geology of North America, v. J: The Gulf of Mexico Basin: Geological Society of America, Boulder, Colorado, p. 389–444.

Salvador, A., and J. M. Quezada-Muneton, 1991, Stratigraphic correlation chart; Gulf of Mexico Basin, *in* A. Salvador, ed., The geology of North America, v. J: The Gulf of Mexico Basin: Geological Society of America, Boulder, Colorado, plate 5.

Sambrook Smith, G. H., A. P. Nicholas, and R. I. Ferguson, 1997, Measuring and defining bimodal sediments: Problems and implications: Water Resources Research, v. 33, p. 1179–1185.

Sassen, R., C. H. Moore, and F. C. Meendsen, 1987, Distribution of hydrocarbon source potential in the Jurrassic Smackover Formation: Organic Geochemistry, v. 11, p. 379–383.

Schilling, M. F., A. E. Watkins, and W. Watkins, 2002, Is human height bimodal?: The American Statistician, v. 56, p. 223–229.

Silver, A., 2019, The essential data science Venn Diagram, <https://www.kdnuggets.com/2019/02/essential-data-science-venn-diagram.html>.

Smith, C. I., J. B. Brown, and F. E. Lozo, 2000, Regional stratigraphic cross sections, Commanche Cretaceous (Fredericksburg-Washita division), Edwards and Stockton plateaus, West Texas; interpretation of sedimentary facies, depositional cycles, and tectonics: Bureau of Economic Geology, Austin, Texas, 39 p., 6 plates.

Snedecor, G. W., 1946, Statistical methods, *in* H. Arkin and R. C. Raymond, Tables for statisticians, 1963, Barnes & Noble, Inc., New York, 168 p.

Sood, R., and M. Coxon, 2020, Radical, sustainable productivity gains with the next generation operating model: Keynote Address at Opex Week, Business Transformation World Summit, Orlando, Florida, January 23, 2020.

Stuchfield, T., 2019, What does digital transformation in oil and gas look like? Is digital transformation in the oil and gas industry as simple as it sounds?, <https://www.oilandgasiq.com/oil-gas/news/what-is-digital-transformation>.

Sturrock, P., 2008, Analysis of bimodality in histograms formed for GALLEX and GNO solar neutrino data: Solar Physics, v. 249, n. 1, p. 1–10.

Subsurface Consultants & Associates, Inc., 1993, Quick look techniques: Mapping throw in place of vertical separation: A costly subsurface mapping misconception: Houston Geological Society Bulletin, v. 36, no. 4, p. 56–57.

Taylor, B., 2020, Why 84% of digital transformations are failing, <https://www.from.digital/is-digital-transformation-worth-it>.

Tearpock, D. J. and R. E. Bischke, 1991, Applied subsurface geological mapping: Prentice Hall, Upper Saddle River, New Jersey, 648 p.

UpX Academy, 2016, <upxacademy.com>.

Van der Eijk, C., 2001, Measuring agreement in ordered rating scales: Quality & Quantity, v. 35, p. 325–341.

Walker, J., 2017, 9 out of 10 digital transformation projects will fail, <http://www.digitaljournal.com/tech-and-science/technology/9-out-of-10-digital-transformation-projects-will-fail/article/499314>.

Wescott, W. A., and W. C. Hood, 1994, Hydrocarbon generation and migrating routes in the East Texas Basin: American Association of Petroleum Geologists Bulletin, v. 78, p. 287–307.

Wilcock, P. R., 1993, The critical shear stress of natural sediments: Journal of Hydraulic Engineering, v. 119, p. 491–505.

Zahm, L. C., C. Kerans, and J. L. Wilson, 1995, Cyclostratigraphic and ichnofacies analysis of the upper Albian Salmon Peak Formation, Maverick Basin, Texas: Gulf Coast Association of Geological Societies Transactions, v. 45, p. 595–604.

Zhang, C., B. E. Mapes, and B. J. Soden, 2003, Bimodality in tropical water vapour: Quarterly Journal of the Royal Meteorological Society, v. 129, p. 2847–2866.