Oilfield Analytics: How New Insights Mitigate Subsurface Uncertainty and Unconstrained Risk
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Executive Summary</td>
<td>1</td>
</tr>
<tr>
<td>SAS® for Oilfield Analytics</td>
<td>2</td>
</tr>
<tr>
<td>Exploration Analysis</td>
<td>2</td>
</tr>
<tr>
<td>Appraisal Management</td>
<td>3</td>
</tr>
<tr>
<td>Drilling and Completion</td>
<td>3</td>
</tr>
<tr>
<td>Reservoir Management</td>
<td>4</td>
</tr>
<tr>
<td>Intervention Management</td>
<td>4</td>
</tr>
<tr>
<td>Performance Forecasting</td>
<td>4</td>
</tr>
<tr>
<td>Production Optimization</td>
<td>5</td>
</tr>
<tr>
<td>SEMMA Process in E&amp;P</td>
<td>6</td>
</tr>
<tr>
<td>High-Performance Analytics</td>
<td>6</td>
</tr>
<tr>
<td>In-Memory Analytics</td>
<td>7</td>
</tr>
<tr>
<td>In-Database Analytics</td>
<td>7</td>
</tr>
<tr>
<td>Conclusion</td>
<td>8</td>
</tr>
<tr>
<td>Learn More</td>
<td>8</td>
</tr>
</tbody>
</table>

Keith R. Holdaway provided the content for this paper.
Holdaway is an oil and gas domain expert in the SAS Global Energy Practice. He joined SAS in 1997 after nearly 20 years at Shell, BP and two geophysical services companies, including a decade in the Middle East.
Executive Summary

Oil companies are being forced to explore in geologically complex and remote areas to exploit more unconventional hydrocarbon deposits. In areas with intrinsically poor data quality, problems are convoluted, and the cost associated with poor predictions (dry holes) rises. To counter these issues in petroleum exploration and production (E&P), companies need to integrate disciplines and rely on approaches that include data integration, risk assessment and quantification of uncertainty.

Soft computing methods offer an excellent opportunity to integrate information from various sources with varying degrees of uncertainty. These methods also help to establish relationships between measurements and reservoir properties in a multidimensional, multivariate and stochastic environment, and to assign risk factors to predictions. By applying analytics, you can optimize a wide range of upstream data to empower effective, efficient and timely decisions that affect the intrinsic value of the E&P value chain.

SAS offers a full suite of analytical solutions to address challenges in the oil and gas industry. For the upstream domain, SAS® solutions can address questions related to risk and uncertainty, and can strengthen confidence in reserves information while quantifying the impact of exploitation plans on attaining predefined targets. SAS offers descriptive and predictive modeling, forecasting and advanced data-driven models to better comprehend the facilities, well and reservoir systems. Advanced visualization tools encapsulate exploration and production data across disparate geoscientific disciplines. Advanced analytics, underpinned by high-performance computing, positions upstream companies to act quickly so they can capitalize on opportunities in the face of increasing data volumes.

This paper explores best practices for managing and analyzing a multivariate, multidimensional and multivariate suite of parameters – in a way that surfaces hidden patterns and trends that help to develop predictive models. Along the way, it describes SAS technologies that can effectively address these data management and analytical challenges.

“As complexity increases, precise statements lose meaning and meaningful statements lose precision.”

Attributed to L.A. Zadeh, Professor Emeritus of Computer Science, University of California, Berkeley (he is credited with coining the term “fuzzy logic”).

Key challenges to delivering upstream sustainability

Deterministic model building and interpretation are increasingly supplemented by stochastic and soft computing methods. This shift takes place as companies strive to use more data-driven approaches to:

- Increase reliability in reserves information.
- Expedite processes for reserves exploitation.
- Reduce subsurface uncertainty.
- Manage portfolios of disparate exploration and production data.
- Preclude production loss.
- Anticipate reservoir damage.
- Control cost of current assets.
- Create seamless communication paths among exploration and production silos.
- Generate knowledge from raw data to expedite robust decision-making cycles.
- Ensure production runs according to predefined targets.
Before addressing the analytical opportunities apparent in the upstream data, we must reflect on the importance of data management. Today, upstream operators spend more than 70 to 80 percent of their time managing and organizing an explosion of historical and real-time data. Ensuring the integrity of this big data has become a pressing issue for the industry.

Successful and credible advanced analytical methodologies require a robust, consistent suite of data. A stable and flexible data management platform is a prerequisite for any soft computing analysis. Once this foundation is established, multiple analytical data marts can be spawned from the master data management (MDM) environment. With data management technology from SAS, operators can take a comprehensive approach to managing data that spans data integration, data quality, MDM, data governance and enterprise data access.

**Exploration Analysis**

Seismic data is pivotal today, as 3-D and 4-D surveys are accelerated across green and brown fields. In addition to the customary seismic wavelet data processing, it’s increasingly important to fully appreciate the hundreds of seismic attributes that exist, and to build a seismic data mart for advanced analysis.

Analytical workflows (see Figure 2) that map seismic attributes to reservoir properties are a key way to define more credible and reliable reservoir characterization definitions that underpin field (re)development. These workflows should integrate - and

---

**SAS® for Oilfield Analytics**

SAS provides a data-driven framework aligned to support the five key areas across the E&P value chain: exploration, appraisal, development, production and enhancement (see Figure 1). The approach is a simple, flexible structure that incorporates advanced analytical methodologies.

![Figure 1: Oilfield analytics framework.](image-url)

![Figure 2: Seismic analytical workflows.](image-url)
supplement – data from well logs. To maintain the spatial integrity of large-areal reservoirs – and to identify both stratigraphic and structural traps – requires high-resolution seismic attributes, and an in-depth understanding of those attributes.

**Appraisal Management**

The process of characterizing the reservoir(s) of a potential or mature field requires analyzing large data sets collated from well tests, production history and core analysis results, enhanced by high-resolution mapping of seismic attributes to reservoir properties. To comprehend the structure of the data, it’s imperative to capture the more subtle observations inherent in these data sets. You can routinely rely on geostatistical methods to accurately quantify heterogeneity, integrate scalable data and capture the scope of uncertainty. But between 70 and 80 percent of allotted time for any reservoir characterization study worth its investment should be focused on exploratory data analysis (EDA). In combination with spatial analysis, simulation and uncertainty quantification, EDA ensures consistent integration, aggregation and management of data – underpinned by univariate, bivariate and multivariate analysis.

To communicate the insights from EDA, upstream operators are turning to advanced data visualization technologies. In particular, they seek solutions that enable a distributed and mobile workforce to access and analyze information while they are in the field. Solutions such as SAS Visual Analytics (Figure 3) use an in-memory analytic engine to perform descriptive and inferential statistics on upstream data.

**Figure 3**: Visual analytics surfaces hidden patterns in multivariate big data.

**Highlights of SAS® Visual Analytics:**

- Forecasting, scenario analysis and other analytic visualizations.
- Interactive dashboards and reporting.
- Integration with mapping technologies.
- Design reports once, publish anywhere.
- Feature-rich mobile apps for iPad® and Android.

**Drilling and Completion**

The target of many operators, particularly for unconventional assets, is to determine the variables that affect the key performance metric of cost-per-foot-drilled. Other focus areas could be on spud-to-total-depth (TD) and costs of drilling and casing horizontal wells. Historical drilling data can be used to quantitatively identify best and worst practices that affect the target, with the intent of improving future drilling operations in unconventional plays and potentially in conventional fields. Advanced analytical methodologies would ultimately develop a predictive model to provide early warnings of deviations from best practices or other events that would adversely affect drilling time or costs.

The vision of real-time drilling optimization is achievable (see Figure 4). You can use advanced analytical techniques to gauge how real-time data is analyzed relative to past performance or events to predict downhole tool failures. Advanced analytics also enables you to do immediate root-cause activities and implement real-time solutions.
Intervention Management

Intervention optimization remediates wells that have either formation or mechanical issues. Developing a suite of advanced analytical data mining workflows that implement soft computing techniques – such as principal component analysis (PCA), multivariate analyses, clustering, self-organizing maps (SOM) and decision trees – generates descriptive and predictive models that efficiently identify candidate wells for remediation. Implementing an oilfield performance forecasting module to determine robust and reliable probabilistic forecasts for the complete portfolio of wells in an asset is another essential step. Subsequent to this step, you can identify wells suitable for intervention by comparing real-time production data with the type curves determined, with 90 percent confidence limits.

Performance Forecasting

Analytical workflows can incorporate a decline curve analysis (DCA) step to implement an oilfield production forecast. This identifies short- and long-term forecasts for oil, gas and water production (see Figure 5). Implementing mature forecasting models and first principles, such as Arp’s empirical algorithms, you can estimate accurate well performance and estimated ultimate recovery (EUR) and measure the impact – positive or negative – of well remediation techniques.
Comparing real-time production data rates and type curves against forecasted trends, you can:

- Quickly and efficiently identify wells that require remediation.
- Segment the field via well profile clustering.
- Ratify from a field, reservoir or well perspective whether current production falls within confidence intervals of expectation, and act accordingly.

Production Optimization

With advanced analytical methodologies, you can perform multivariate analysis on disparate upstream data sets, both operational and nonoperational, to evaluate and determine those variables that either inhibit or improve well performance. Combining predictive and descriptive analytical workflows to explore the data, you can surface hidden patterns and identify trends in a complex system.

Adopting data-driven models improves efficiency and enables significant discoveries of influential parameters used to address issues that adversely affect production, without relying solely on first principles. For example, many production inhibitors, such as skin damage and sanding, can be predicted before they happen. However, the predictive model relies on insights from exploratory data analysis, which, in turn, aggregates and integrates data from across geoscientific silos. As such, optimizing production is an iterative process that requires data from multiple disciplines.
SEMMA defines data mining as the process of sampling, exploring, modifying, modeling and assessing inordinate amounts of data to surface hidden patterns and relationships in a multivariate system (see Figure 7). The data mining process provides methodologies for diverse business problems, such as maximizing well location, optimizing production, ascertaining maximum recovery factor, identifying an optimum hydraulic fracture strategy in unconventional reservoirs, field 

**SEMMA Process in E&P**

SEMMA is a proven analytical workflow that generates exploratory analysis and helps reveal insights about data. SEMMA enables you to:

- **Sample the data** by extracting and preparing a portion of the data for model building, using one or more data tables. Sampling includes operations that define or subset rows of data. The samples should be large enough to efficiently contain the significant information. It’s best to include the complete and comprehensive data set for the “explore” step, owing to hidden patterns and trends that are only discovered when all the data is analyzed. Software constraints may preclude such an ideal.
- **Explore the data** by searching for anticipated relationships, unanticipated trends and anomalies. This returns deeper understanding and ideas that insinuate hypotheses worth modeling.
- **Modify the data** by creating, selecting and transforming the variables to focus the model selection process on the most valuable attributes. This focuses the model selection process on those variables displaying significant attributes vis-à-vis the objective function or target variable(s).
- **Model the data** by using analytical techniques to search for a combination of the data that reliably predicts a desired outcome.
- **Assess the data** by evaluating the usefulness and reliability of the findings from the data mining process. Compare different models and statistically differentiate and grade those models to ascertain an optimum range of probabilistic results, delivered under uncertainty.

**High-Performance Analytics**

High-performance analytics enables oil and gas companies to be more nimble and confident in their decision-making cycles as they engage in new ventures and derive new value from a tsunami of data. Even the most challenging fields can be quickly
assessed, generating high-impact insights to transform operations.

In the age of big data, oil and gas companies depend on sophisticated analysis to manage all the volumes and varieties of data collected at ever-increasing frequencies across the siloed geoscientific community. This data that comes from intelligent wells equipped with downhole sensors has added enormous pressures on upstream thinking. Engineers are now asking questions such as:

How can we extract the most knowledge from raw data? How can we impose quality control workflows that filter noise and outliers, impute missing values, and normalize and transform data values?

Today, organizations must strive to yield a robust collection of disparate data so they will be ready for both deterministic and stochastic workflows. This is because today's hybrid approach to analytics marries new concepts of soft computing with the traditional interpretations ingrained by our geophysics, geology, petroleum and reservoir engineering institutions.

As oil and gas companies adopt this hybrid approach, CIOs are turning to high-performance environments to meet the needs of the business. With high-performance analytics you can:

- Attain timely insights to help you make decisions in a diminishing window of opportunity.
- Speed innovation by surfacing insights in just hours or days instead of months.
- Discover precise answers for complex problems.
- Seize opportunities for growth that might otherwise be missed.
- Achieve tremendously improved performance.

At SAS, high-performance analytics is enabled by two innovative technologies. These technologies increase the speed of insights on massive data sets, either by holding the data in memory or by executing the algorithm within the database.

In-Memory Analytics

In-memory analytics enables analytical workflows on big data to solve complex upstream E&P problems in an unfettered manner. The speed and flexibility of in-memory analysis enables analysis that was previously out of scope due to computing environment constraints.

In-memory analytics scales to your business needs, providing concurrent, in-memory and multiuse access to data, no matter how big or small. The software is optimized for distributed, multithreaded architectures and scalable processing, so requests to run new scenarios or complex analytical computations are handled blazingly fast.

In-memory analytics technologies enable upstream geoscientists to perform analyses that range from data exploration, visualization and descriptive statistics to model building with advanced algorithms.

When it comes to the most common descriptive statistics calculations, SQL-based solutions have a number of limitations, including column limits, storage constraints and limited data type support. In addition, the iterative nature of EDA and data mining operations - such as variable selection, dimension reduction, visualization, complex analytic data transformations and model training - requires multiple concurrent passes through the data. SQL and relational technologies are not well-suited for these operations.

In-Database Analytics

In-database analytics executes an analytic algorithm directly within database engines using native database code. Traditional processing may include copying data to a secondary location, and the data is then processed using E&P upstream products. Benefits of in-database processing include reduced data movement, faster run times and the ability to use existing data warehousing investments.

In-database analytics invariably covers two key areas. It helps you:

- Develop new products that provide access to and process extant functions within the database.
- Enhance existing products to take advantage of database functionality.

In-database processing is a flexible, efficient way to use increasing amounts of data by integrating select upstream technology into databases or data warehouses. It utilizes the massively parallel processing architecture of the database or data warehouse for scalability and better performance. Moving relevant data management, analytics and reporting tasks to where the data resides is beneficial in terms of speed, reducing unnecessary data movement and promoting better data governance. For upstream decision makers, this means faster access to analytical results and more agile and accurate decisions.
Conclusion

Oil companies operate in a highly competitive, changeable global economy. But every problem has an opportunity attached. Most organizations struggle to manage and glean insights from data and use analytic results to improve performance. They often find analytic model development, deployment and management to be a time-consuming, labor-intensive process, especially when it’s combined with excessive data movement and redundancy.

In-database processing is ideal for two key scenarios. The first scenario is for big data enterprise analytics, where the sheer volume of data involved makes it impractical to repetitively copy it over the network. The second scenario is for complex, organizationally diverse environments, where varying business communities need to share common data sources – driving the need for a centralized enterprise data warehouse. Oil companies should implement corporate data governance policies to promote one single version of the truth. This would minimize data inconsistency and redundancy, and align data access needs to common business usage.

One of the best ways to address a multivariate, multivariant, multidimensional and stochastic system such as a hydrocarbon reservoir is to adopt a suite of data-driven methodologies underpinned by the SEMMA process. Implementing such advanced analytical workflows in high-performance environments positions geoscientists to harness big data from across all the disparate engineering silos in exploration and production organizations. As a result, operators can act quickly and decisively to take full advantage of new business opportunities.

Learn More

Find out more about how SAS can help upstream operators: [sas.com/oilgas](http://sas.com/oilgas).

Watch a webinar to learn how you can apply predictive analytics in the upstream field: [Using Predictive Analytics Across the Exploration and Production Value Chain](#).

Read a white paper about creating groundbreaking intelligence from new data sources: [Analytic Innovations Address New Challenges in the Oil and Gas Industry](#).