AADE-16-FTCE-07

Agile Data-Driven Fluid Design: Predicting the Properties of Drilling, Spacer and Cement Slurry Fluids

Arash Shadravan, ReservoirFocus LLC, Mohammadali Tarrahi, Shell Global Solutions (US), Inc.

Abstract

The rheological compatibility between drilling fluids, spacers and cement slurries is very important. Incompatible fluids cause excessive pressure, fluid channeling and a poor cement job. Fluid compatibility tests are all often conducted by trial and error in the laboratory and it is, in fact, very time consuming. There are several test data that would not be used in this method and it is hard to capture a plethora of information for users and take intelligent and cost-effective decision to design a fluid with the desired properties. Therefore, trial and error method is considered to be very costly and misleading. Today, there is a need for an intelligent system which uses all the available fluid design data stored in a database by which we can benefit from its insights for smart fluid designs. This predictive tool suggests a composition for drilling fluids, spacer fluid or cement slurry by implementing a machine learning algorithm on imported experimental data. Designing these fluids in a wide range of rheological profiles allows minimization of fluid intermixing.

This study demonstrates the implementation of a data-driven predictive tool which uses Gaussian Process Regression (GPR) to design or improve compatibility between fluids. GPR as a machine learning method considerably reduces the costs of testing, optimizes the material use, integrates available experimental data and eliminates the user bias. This practical nonlinear regression method fosters an efficient and fast prediction analysis which do not require including complex physics of the underlying intricate chemical fluid behavior while integrating all available data from different databases. GPR has exceptional advantages over traditional regression methods since it does not require a known form for regression function. This machine learning based tool offers comprehensive insights for intelligent fluid design and considerably reduces the experiment cost. This study showcases an example through which of GPR predicts incompatibility between fluids and helped engineers to maintain the fluid rheological hierarchy for better cement jobs and well integrity.

Introduction

Machine learning methodologies and expert systems have proved to be conducive to digest large data sets and make intelligent decisions for agile productivity and accurate measures. These algorithms have been successfully applied to petroleum engineering fields such as production forecasting, history matching, reservoir characterization and EOR screening (Alvarado *et al.* 2002; Parada *et al.* 2012; Tarrahi *et al.* 2015). The proposed workflow in this paper includes the following tools:

1- ***Data preparation tool:*** preparing the training data through guided experiments

- Determining the most sensitive and effective design parameters (e.g. temperature, density, etc.)
- Design of experiment (DOE), Latin Hypercube Sampling (LHS), or Monte Carlo sampling to produce the most representative and wide enough sample cases

2- ***Prediction tool:*** building an intelligent data-driven prediction tool

- Build a reliable prediction tool (regressor) through Machine Learning algorithms
- Predict (and assess the uncertainty of) any new samples properties (e.g. rheological properties)

3- ***Design tool:*** intelligent fluid design tool

- Designing the desired fluid system based on the requested sample properties
  a. Performing the opposite of prediction tool and building a regressor to predict corresponding parameters of a suitable fluid system (through machine learning algorithms)
  b. Applying inverse problem methods to prediction tool and obtaining the desired fluid system

In the machine learning literature some of the most well-known regression methods are: Linear Regression (Ordinary Least Square), Nonlinear Regression or Curve Fitting (e.g. Polynomial Regression), Logistic Regression, Ridge regression, Artificial Neural Network, Bayesian Network, Radial Basis Function (RBF) Network, Support Vector Machine (SVM), Gaussian Process Regression and Kernel Regression (Duda *et al.* 2012; Theodoridis *et al.* 2010). In this paper we choose Gaussian Process Regression (GPR) algorithm to investigate its application in fluid characteristics estimation and fluid design and build an intelligent multi-purpose prediction and design tool (Shadravan *et al.* 2015). Predicting the rheological profiles mud, spacer and cement slurries and maintaining proper rheological hierarchy will benefit long term zonal isolation, Shadravan *et al.* (2015). **Figure 1** lists various types of fluid designs experiments which usually are pursued by a trial and error methods.

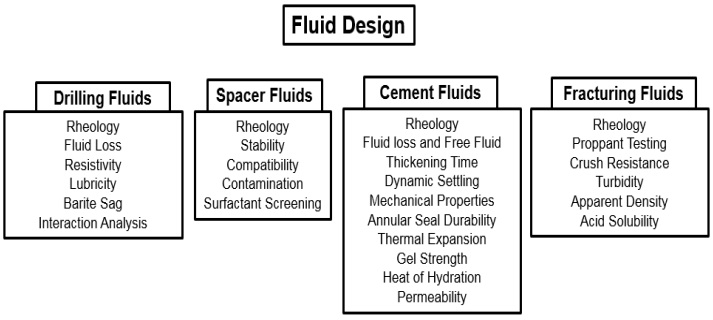

Figure 1: Applications of the intelligent design tool for various experiments

Shadravan *et al.* (2014) presented a model which could predict the rheological profiles of a micro-emulsion spacer fluid up to 300 °F. This model was built based on Design of Experiments (DOE). DOE relies heavily on individual and joint statistics of fluid system parameters, and its implementation in fluid systems with many components requires huge investment in time and cost. DOE along with response surface methodology (RSM) are also not able to control certain variables, problems identifying all the variables that affect the process, and a lack of linearity in variables that makes it hard to identify optimum settings. These drawbacks tend to limit the frequency with which DOE is used to solve problems. Therefore in this paper, GPR method shows how to predict such properties for desired fluid design.

The complexity and variations of downhole conditions, geology, additives (water, polymers, cement types, weighting agents, crosslinker, breaker, etc.), testing procedures, and experiment schedule could be all properly taken into account to construct a useful laboratory database for an intelligent fluid design tool, Lee *et al.* (2013). In this study, separate data silos for drilling fluids, spacers and cement slurries were created and used. Proper equipment calibration and competent laboratory personnel are also evidently valuable elements of successful fluid design. **Figure 2** shows traditional trial and error fluid design and **Figure 3** demonstrates the comprehensive intelligent machine learning fluid design workflow.

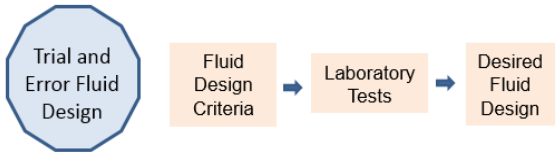

Figure 2: Traditional arbitrary and trial-and-error-based fluid design procedure

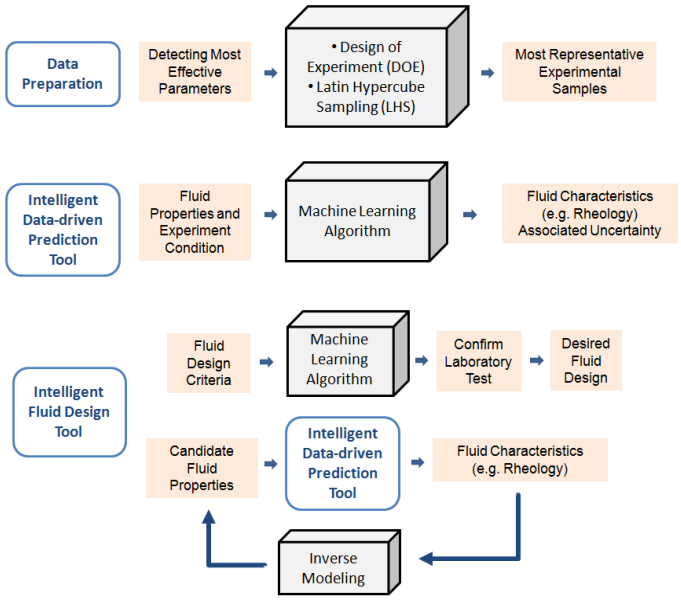

Figure 3: Schematic of the proposed intelligent integrated tools for guided experiment design, characteristic prediction and fluid system design

Data-Driven Analytics for Fluid Characteristics Prediction and Optimal Design

Rheology is a common experiment which is conducted on drilling, spacer and cement slurry fluids. This paper delineates the benefits of GPR as a powerful and robust machine learning algorithm to predict the rheological properties of a fluid as a simple example of the breath of Machine Learning methods application. Predicting such properties can be viewed as a regression analysis (a data-driven approach) where the goal is to design the best linear or nonlinear regression function (or rule).

We solved this problem by GPR machine learning method, which is well-established in computer science literature. The proposed algorithm eliminates arbitrary approach in making decisions, and provides accuracy and fast computation. This approach substitutes the need for days and weeks of testing in the laboratory with an intelligent and agile fluid design which could require only a few number of tests to

confirm the proposed compositions (additives) for the desired properties.

In the parametric regression methods, a known regression function (e.g. polynomial, exponential, etc.) is defined in terms of finite number of unknown parameters which are estimated from the training data. In general since the nature of the underlying physical process (e.g. relation of rheological properties and fluid characteristics) is very complex, assuming a known and closed form for the regression function is a rigorous task. Therefore nonparametric regression methods are proposed that have no (or very little) a-priori knowledge of the form of the function that is being estimated. These methods allow the class of functions which the model can represent to be very broad (Bishop, 2006, Duda, *et al.* 2012, Theodoridis *et al.* 2010). Here we implement and apply Gaussian Process Regression that counts as one the most practical nonparametric regression methods. One of the advantages of Gaussian Processes application for machine learning is that the prediction or estimated value is probabilistic (Gaussian) so empirical confidence intervals and exceedance probabilities can be computed that might be used to refit (online fitting, adaptive fitting) the prediction in some regions of interest. Also the estimation and regression procedure is versatile i.e. different linear regression models and correlation models can be specified.

Gaussian Process Regression for Data-Driven Machine Learning

Gaussian process regression (GPR) is a nonlinear regression or interpolation technique that models the new estimated (interpolated) values based on Gaussian process determined by a covariance function (Williams & Rasmussen 2006). GPR is a statistical machine learning method which is also known as Kriging and is well established in Geostatistics and computer science literature (Deutsch & Journel 1992, Shi & Choi 2011). To estimate the corresponding value for the new input, GPR calculates the weighted average (linear combination) of the known values (training data) based on the correlation (or proximity) of the new input and the training data governed by the covariance function (Rasmussen, 2006). Providing the training data and the prior assumptions (e.g. covariance function) are suitable (Deutsch & Journel 1992), GPR or kriging is the best linear unbiased estimator (BLUE). In Geostatistics literature the spatial dependencies are represented by variaogram function instead of covariance function. In signal processing literature this estimation procedure is known as Kalman filtering which have been also applied to history matching problems (Tarrahi *et al.* 2013, 2015).

In our current experiment, there are 3 input parameters (fluid density, ingredient A content and temperature) and 6 output values (300, 200, 100, 60, 6, 3 RPM viscosities) and we have a set of experimental data points (training samples).

$$training\ data: \{X_i, Y_i\}, \quad i = 1,2, \dots, N$$
$$X_i = i^{th}\ input\ vector\ (n \times 1) \tag{1}$$
$$Y_i = i^{th}\ output\ vector\ (m \times 1)$$

Where N represents the number of experiments or the number of data points. In this study, n and m are 3 and 6, respectively. To perform Gaussian process regression, we consider one output at a time i.e. we construct seven GPR models for seven dial reading outputs. So the training data for each GPR model design is:

$$training\ data\ D: \{X_i, y_i\}, \quad i = 1,2, \dots, N \tag{2}$$

Where y_i is a scalar and represents one of the dial reading outputs. Each nonlinear regression function is a random sample (parametrized by the input variable) drawn from a joint Gaussian probability function given the training data set D. Our goal in GPR is to train a function f from data D. A Gaussian process is a prior (information) over functions $p(f)$ that can be utilized for Bayesian regression (inference):

$$p(f|D) = \frac{p(f)p(D|f)}{p(D)} \tag{3}$$

Where $p(f|D)$ shows the posterior probability of all the regression functions given the training data. $p(f)$ represents the prior probability on all the possible regression functions and $p(D|f)$ is the likelihood function. In Bayesian inference context, $p(D)$ is only a scaling parameter which does not affect the intended posterior probability density function $p(f|D)$.

The nonlinear regression function in GPR is a set of random variables indexed by a continuous parameter (e.g. time, space, temperature, etc.) which is also called random function $f(X)$. The assumption in GPR is that any set of regression functions has jointly Gaussian distribution with zero mean:

$$p(f|X) = N\big(0, C(X)\big) = (2\pi)^{-\frac{k}{2}}|C(X)|^{-\frac{1}{2}}e^{-\frac{1}{2}fC(X)^{-1}f} \tag{4}$$

Where $C(X)$ is the covariance matrix and k represents the dimension of function f with respect to continuous parameter. Estimating the unknown target value y_0 given the set of training y presents a conditional probability function:

$$p(y_0|y) = N(\overline{y_0}, \sigma_{y_0}) \tag{5}$$

The GPR estimation value is the mean of the above conditional probability density function. Basically GPR proposes a Gaussian probability density function (PDF) for the new estimated value which can be fully characterized by a mean and a standard deviation value.

As a necessary preprocessing step to avoid scaling issues due to different input and output ranges, we apply data normalization and scaling to input and output values. Target or output values are assumed to be normal standard variables (resulted from a Gaussian distribution with zero mean and unit variance). Moreover the input values (input parameters) are

shifted and scaled to [0,1] interval.

$$X_s = \frac{X - X_{min}}{X_{max} - X_{min}}$$
$$Y_{sn} = \frac{Y - \bar{Y}}{\sigma_Y} \quad (6)$$

To define the spatial relation of the samples in the input space (in this study a 3D space), a covariance function is established:

$$C(X_i, X_j) = vexp(-h^p)$$
$$h = \sqrt{g^T L^{-1} g}$$
$$g = R(X_i - X_j)$$
$$L = \begin{bmatrix} l_1 & 0 & 0 \\ 0 & l_2 & 0 \\ 0 & 0 & l_3 \end{bmatrix} = scaling\ matrix \quad (7)$$
$$X_i, X_j = experiment\ input\ values\ (3 \times 1)$$

l_1, l_2 and l_3 are correlation lengths in three input directions. Matrix R represent the rotation matrix in 3D which can be calculated based on the rotation angles around different axes (D'Orangeville et al. 2003). The combination of L and R matrices determines the spatial correlation of fluid properties in matrix X. v represents the variance parameter. In specific cases, where the parameter p is equal 2, the resulted covariance is the squared exponential covariance function (or Gaussian) and if $p = 1$, the exponential covariance function is reproduced. The parameters involved in defining the covariance function are called hyper parameters. By changing the hyper parameters the GPR estimator quality will change. In this study we tune 6 hyper parameters (three correlation lengths, three rotation angles and the power) through leave-one-out cross validation (LOOCV) to obtain the optimal GPR estimator. In LOOCV, we assume one of the training data points is unknown and use the other 8 training data to design GPR estimator and estimate the unknown data point and then compare the estimated value with the true value. This procedure is repeated for all the outputs (6 viscosity readings) and for all the training data. In cross validation procedure, the performance measure of the GPR estimator is the average relative error.

The following is the relation between semivariogram (popular spatial correlation representation in Geostatistics) and covariance function:

$$\gamma(X_i, X_j) = v - C(X_i, X_j) \quad (8)$$

To estimate the corresponding output y_0 of the new input parameter X_0, we use the Gaussian process interpolation (also called simple Kriging):

$$y_0 = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}^T C_{nn}^{-1} C_{nu} = \sum_{i=1}^{N} \lambda_i y_i \quad (9)$$

$$C_{nn} = \begin{bmatrix} C(X_1, X_1) & \cdots & C(X_1, X_N) \\ \vdots & \ddots & \vdots \\ C(X_N, X_1) & \ldots & C(X_N, X_N) \end{bmatrix}$$
$$= covariance\ matrix\ of\ training\ data$$

$$C_{nu} = \begin{bmatrix} C(X_1, X_0) \\ \vdots \\ C(X_N, X_0) \end{bmatrix}$$
$$= cross\ covariance\ of\ training\ data\ and\ new\ input$$

The estimated output is basically a linear combination of the training outputs where the coefficients are determined by the spatial configuration of the input (with respect to the training data) and the covariance function shape. One of the unique advantages of GPR for instance comparing to Artificial Neural Networks (ANN) is the ability to obtain the estimation error. GPR not only provides us with the estimated output at the new input value but also it is able to calculate its associated error or variance. With this capability we can specify how good each estimation is or determine the confidence of each estimation and decide if we need to perform new laboratory experiments.

$$estimation\ variance\ of\ y_0 = v - C_{nu}^T C_{nn} C_{nu} \quad (10)$$

This provides us with the estimation standard deviation and consequently the confidence interval associated with each estimation. It should be noted that GPR is an absolute estimator i.e. estimated values for the training data are equal to true outputs. After tuning using LOOCV the average minimum cross validation error is 10%. We obtained different hyper parameters (different covariance functions) for different viscosity outputs. For all the output viscosity readings, the obtained power parameter p through cross validation is 2 so the preferred covariance function is the squared exponential covariance function.

Since there are 3 input parameters and plotting each of the viscosity readings with respect to 3 independent input parameters is not very informative, for illustration purposes, we plotted the viscosity with respect to only two input parameters and kept the third one constant. In **Figure 4** all of the predicted output values are plotted versus two pairs of the input parameters for mud, spacer and cement. For instance, first plot in **Figure 4** shows the 3 RPM predicted viscosity of cement, mud and spacer with respect to temperature and ingredient of different material contents while their density is constant (lb/gal). The plotted surfaces are the nonlinear regression models resulted from GPR. Having this nonlinear regression function, we are able to predict the rheological properties of the new mud, spacer and cement samples prior to performing the confirmation laboratory experiment.

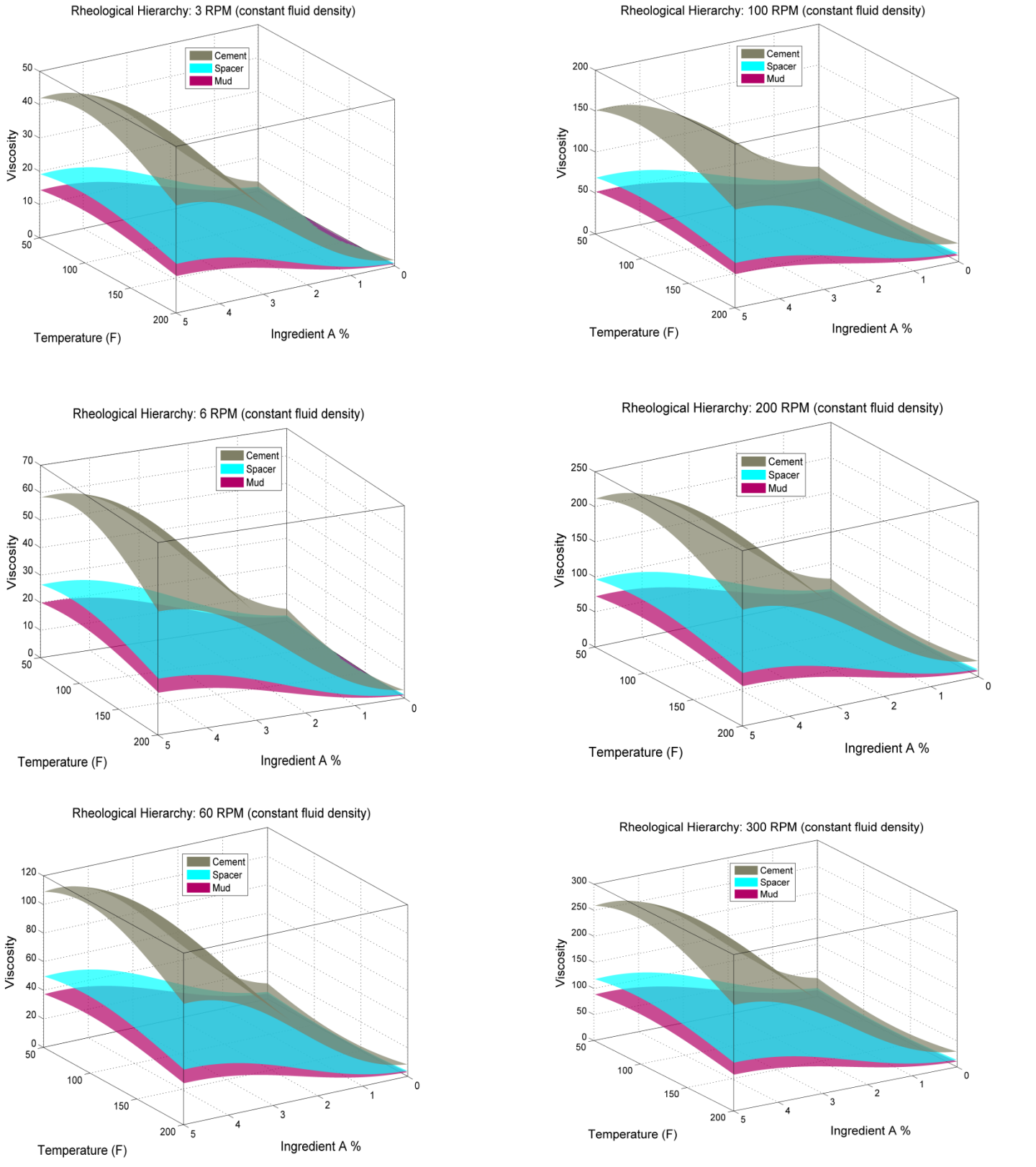

Figure 4: Nonlinear estimator resulted from GPR for different RPMs

Work flow execution for predicting mud, spacer and cement properties

The following figures show the comprehensive procedure of experimental data importing, prediction with machine learning technique, visualization and finally fluid characteristics design. As shown in **Figure 5**, experimental fluid data can be manually or automatically imported to the design tool in the first step. Then in the prediction step (**Figure 6**), GPR based data-driven tool will be trained by the imported data and customized fluid characteristics plots can be visualized and their data can be exported. In the final step in **Figure 7**, the rheological characteristics of mud, spacer and cement are predicted and plotted based on their fluid system properties.

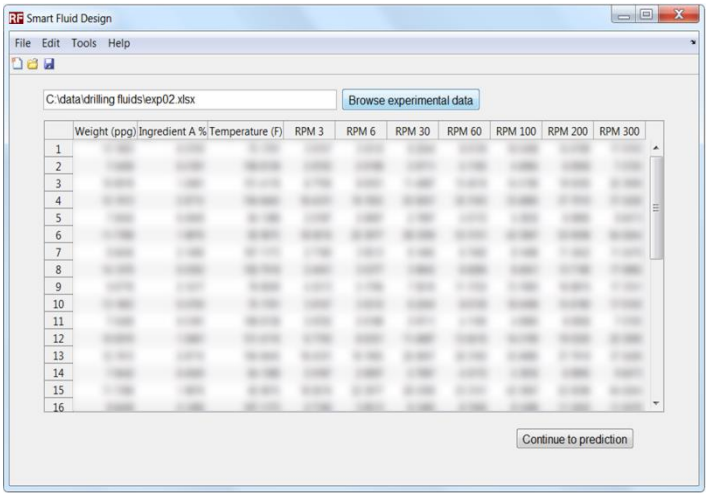

Figure 5: Importing experimental data (training data) to smart data- driven fluid design tool

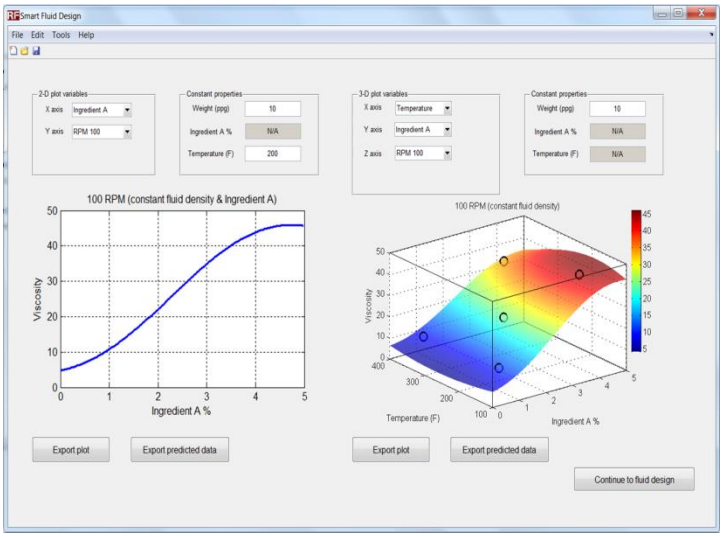

Figure 6: Data-driven rheological properties estimation and visualization

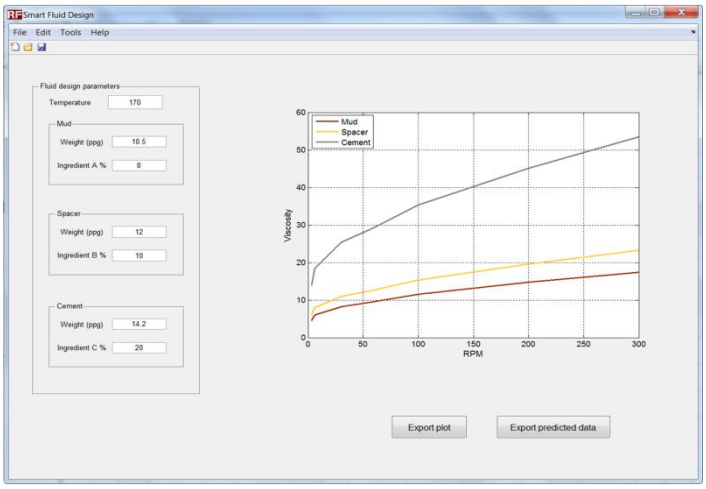

Figure 7: Predicting rheological properties of mud, spacer and cement: smart fluid system design

Discussion

Fluid compatibility is extremely important for cement job designs and especially offshore cementing. Predicting the fluid properties is beneficial to the operations cost effectiveness and success. Traditional fluid design methods (e.g. trial and error) are incapable of offering an agile comprehensive computation capability which is provided by machine learning algorithms. DOE along with RSM is a good method for data point selection to learn more about R&D and new products however it is limited and time consuming method and not flexible enough for agile decision making. GPR is a strong regressor which can suitably estimate the mud, spacer and cement properties. This paper showcased the strength of GPR in predicting the rheological properties of mud, spacer and cement. This is an important example through which GPR assisted maintaining the rheological hierarchy at different RPMs. A good rheological hierarchy would be a contributor to better friction pressure hierarchy and therefore less fluid intermixing and viscous fingering, better cement jobs and well integrity excellence.

Conclusion

Conventional fluid design methods such as trial and error, is not cost effective and lacks long-term vision to accumulate knowledge on mud, spacer and cement fluid systems. There is a need for a tool which is empowered by machine learning methodologies. This paper investigated an intelligent tool which is equipped with flexible machine learning algorithms. GPR unlike ANN honors the training data and offers a prior-free regression implementation. Furthermore GPR is unique in the sense of providing the estimation error and uncertainty assessment while predicting the fluid characteristics and designing a fluid system. The developed intelligent prediction and design tool is able to extract and relate any set of input parameters to output variables based on the provided training samples that can be both experiment and simulation results.

Acknowledgment

Authors would like to thank the management of ReservoirFocus LLC for the permission to publish this work.

References

Alvarado, V., Ranson, A., Hernandez, K., Manrique, E., Matheus, J., Liscano, T., & Prosperi, N. (2002, January 1). Selection of EOR/IOR Opportunities Based on Machine Learning. Society of Petroleum Engineers. doi:10.2118/78332-MS

Amani, M., Al-Jubouri, M. and Shadravan, A., (2012), Comparative study of using oil-based mud versus water-based mud in HPHT fields, Advances in Petroleum Exploration and Development, Vol. 4, No. 2, 2012, pp. 18-27 DOI:10.3968/j.aped.1925543820120402.987

Bishop, C. M. (2006). Pattern recognition and machine learning. springer.

Deutsch, C. V., & Journel, A. G. (1992). Geostatistical software library and user's guide. New York.

D'Orangeville, C., & Lasenby, A. N. (2003). Geometric algebra for physicists. Cambridge University Press.

Duda, R. O., Hart, P. E., & Stork, D. G. (2012). Pattern classification. John Wiley & Sons.

Lee, J., Shadravan, A., & Young, S. (2012, January 1). Rheological Properties of Invert Emulsion Drilling Fluid under Extreme HPHT Conditions. Society of Petroleum Engineers. doi:10.2118/151413-MS

Parada, C. H., & Ertekin, T. (2012, January 1). A New Screening Tool for Improved Oil Recovery Methods Using Artificial Neural Networks. Society of Petroleum Engineers. doi:10.2118/153321-MS

Rasmussen, C. E. (2006). Gaussian processes for machine learning.

Shadravan, A., Alegria, A., & Castanedo, R. (2015, September 28). Rheological Hierarchy Optimization Improves Fluid Displacement and Well Integrity, 3 Worldwide Case Studies. Society of Petroleum Engineers. doi:10.2118/174773-MS

Shadravan, A., Narvaez, G., Alegria, A., Carman, P., Perez, C., & Erger, R. (2015, March 16). Engineering the Mud-Spacer-Cement Rheological Hierarchy Improves Wellbore Integrity. Society of Petroleum Engineers. doi:10.2118/173534-MS

Shadravan, A., Tarrahi, M., & Amani, M. (2015). Intelligent Tool to Design Fracturing, Drilling, Spacer and Cement Slurry Fluids Using Machine Learning Algorithms. Society of Petroleum Engineers. SPE-175238-MS

Shi, J. Q., & Choi, T. (2011). Gaussian process regression analysis for functional data. CRC Press.

Tarrahi, M., Jafarpour, B., & Ghassemi, A. (2015). Integration of microseismic monitoring data into coupled flow and geomechanical models with ensemble Kalman filter, Water Resour. Res., 51, doi:10.1002/2014WR016264.

Tarrahi, M., Jafarpour, B., & Ghassemi, A. (2013, September 30). Assimilation of Microseismic Data into Coupled Flow and Geomechanical Reservoir Models with Ensemble Kalman Filter. Society of Petroleum Engineers. doi:10.2118/166510-MS

Tarrahi, M., Afra, S., & Surovets, I. (2015). A Novel Automated and Probabilistic EOR Screening Method to Integrate Theoretical Screening Criteria and Real Field EOR Practices Using Machine Learning Algorithms. Society of Petroleum Engineers. SPE-176725-MS

Theodoridis, S., Pikrakis, A., Koutroumbas, K., & Cavouras, D. (2010). Introduction to Pattern Recognition: A Matlab Approach: A Matlab Approach. Academic Press.

Williams, C. K., & Rasmussen, C. E. (2006). Gaussian processes for machine learning. the MIT Press, 2(3), 4.

Nomenclature

BHA	*Bottomhole assembly*
$C(X)$	Covariance matrix of input data
D	The complete training dataset
f	Possible regression function form
m	Dimension of output data vector
n	Dimension of input data vector
N	Total number of training data
X	Training input data
X_i, X_j	Different input values
X_{min}	Minimum of X
X_{max}	Maximum of X
Y	Training output data
$\bar{Y}$	Mean of Y
y_0	Unknown output
$\overline{y_0}$	Estimated mean of y_0
σ_{y_0}	Estimation error of y_0
σ_Y	Standard deviation of Y