

Geochemical determination of heavy oil viscosity using multivariate statistical algorithms: Procedures, pitfalls and paths to practicality!

Jennifer Adams^{1,2}, Barry Bennett³, Lloyd Snowdon³ and Steve Larter²

¹ConocoPhillips, 600 N Dairy Ashford, Houston, TX (Jennifer.J.Adams@ConocoPhillips.com);

²Petroleum Reservoir Group, Dept of Geosciences, University of Calgary, Canada.

³Gushor Inc., Calgary, Canada.

Summary

Accurate prediction of fluid properties, such as viscosity, using geochemical proxies correlated to physical measurements of viscosity is viable for heavy oil production and exploration activities and allows viscosity assessments of both core and cuttings even when samples are contaminated with well additives. Such predictions rely not only on high precision geochemical measurement to calibrate proxy methods but crucially on accurate, representative, calibration viscosity measurements. Error in measured viscosity data is primarily introduced via sample handling and analysis, rather than via geological processes. Here we present multivariate statistical partial least squares (PLS) and artificial neural network (ANN) oil geochemistry-viscosity prediction methods as well as a storage time viscosity correction (STVC) methodology able to ameliorate core storage effects in large legacy viscosity datasets enabling effective correlations of chemical properties and oil viscosity.

Introduction

Modern reservoir engineering practices, especially those focused at unconventional resources, rely on accurate physical fluid property estimates for resource assessment, risking, and optimal exploration and production schemes. In conventional petroleum fields, in-field variations in oil production as a function of oil mobility are primarily controlled by variance in reservoir permeability, which is more heterogeneous than *in situ* oil-phase fluid properties; however, the opposite is true in heavy oil and super heavy oil fields, where oil viscosity varies laterally and horizontally over short distances by orders of magnitude (Adams, 2008; Larter et al., 2008a). In these resources *in situ* sampling becomes difficult or impossible, horizontal wells are common and cuttings and side wall core become the common sample types.

Pre-drill fluid property predictions are possible using genetic reactive compositional models (e.g., GNAWD-Adams, 2008, MPATH, Trinity or BioClass) or geostatistical extrapolation of large measured API gravity or viscosity data sets, but typically the required data are unavailable for new development and exploration plays. Molecular representation models can also predict equation of state parameters, either directly (e.g., Crampon et al., 2004) or via critical properties (Marreo and Gani, 2001), but required detailed compositional models and skilled modelers. More generally, lack of core or downhole sampled fluids, prohibits collection of a representative (unaltered) oil sample and small sample volumes or sample aging or contamination dictate that API gravity and/or viscosity must be estimated using genetic or statistical models and geochemical analysis. Chemical correlations with measured fluid properties using multivariate methods, e.g., partial least squares (PLS; Rosipal and Kramer, 2006) and neural networks (Qin and McAvoy, 1992) have been applied to complex systems involving oils, coal and other substances by using geochemical components to predict key parameters for many applications. These methods can be applied quickly at low cost using limited measured or analog data and tiny oil samples for the prediction sample set (Koopmans et al, 1998, 2002; Larter et al, 2008).

Fluid property prediction, especially for viscosity is most important in heavy oil reservoirs due to the wide viscosity variations with up to two orders of magnitude change in viscosity at reservoir

conditions observable within a 30 m interval. In most cases, no single parameter can adequately represent the measured viscosity of oil across a heavy oil field. We showed earlier, (Adams, 2008; Larter et al, 2008) that complex relationships between geochemical parameters and dead oil viscosity may be resolved using multivariate methods. This paper addresses best practices for developing correlations and methods for identifying sources of noise in large complex data sets to generate accurate predictions of oil viscosity based on geochemical data. The same methods can be applied to predict API gravity; however, we caution that using standard correlations to compute viscosity from API are totally unreliable for heavy oils! We don't recommend this!

Method

For regression of chemical data, PLS is particularly effective when the number of independent variables (chemical components) is greater than the number of samples, which is often the case in exploration, and when the independent and dependent (predicted) variables share a linear correlation. Viscosity commonly varies highly non-linearly with heavy oil composition due to its complex, often non-Newtonian nature and thus non-linear PLS or non-linear neural networks may best predict heavy oil viscosity from oil composition. Non-linear PLS, however, is very sensitive to starting values and typically does not produce significantly higher quality predictions than conventional PLS but can account for outliers more effectively (Malthouse et al., 1996). For this reason, we commonly apply a simple log transform to the measured viscosity data and use PLS methods and multi layer perceptron or radial basis network neural networks (ANN model), trained on oil component concentration data and dead oil viscosity (20°C in cP). PLS methods work easily with small sample sets while ANN methods require more substantial training sets though we have developed means to apply ANN methods in typical geochemical data sets.

In this paper we discuss the advantages and disadvantages of the various methods on large datasets of penecontemporaneously analysed heavy oil geochemical data and a fluid data set from Canadian heavy oilfields. For example, in one study, over 300 solvent extracted oils from delineation well core from a heavy oil field were analyzed for SARA, and GC-MS of total hydrocarbons. The viscosity of 120 oil samples mechanically extracted from core using a sealed vessel compaction driven recovery device (Larter et al, 2006) were measured using a Brookfield viscometer. The maturity and source rock facies variations of the oil charge are comparatively uniform in this dataset but the reservoir core was stored frozen under standard storage conditions for a variable time of from 2 to ~1800 days prior to mechanical oil extraction and viscosity measurement. Oil API gravities vary from 7.7 to 11 degrees, dead oil viscosities(20C) as vary from ~3000 to 3,000,000 cP and biodegradation levels from 5 to 6 (Peters and Moldowan scale).

Statistical predictions of viscosity from oil chemistry rely on accurate viscosity measurements to calibrate proxy correlations and an adequate number of precise chemical parameters to span the whole range of compositions in the sample set (e.g., biodegradation can remove key components from some oils) and account for processes that have "altered" different compounds in the oil, affecting viscosity (maturity and facies, biodegradation and sample storage effects). Measured oil viscosity is a function of not only intrinsic oil charge chemistry and in reservoir alteration, but also the storage conditions and duration, sample collection, contamination and processing and measurement (Adams, 2008; Adams et al., 2008). Volatilization of light end hydrocarbons (LEH) during core storage, handling, extraction and oil cleaning significantly affects measured viscosity. Heavy oil core, stored frozen, loses LEH systematically with time, causing order of magnitude increases in bitumen viscosity over 1 year, while relative viscosities in a vertical reservoir profile are maintained. As standard centrifugation of core to extract oil often volatilizes LEH, compaction-based bitumen extraction (Larter et al, 2006) was used to obtain viscosity data

In addition, experiments investigating contamination of bitumen samples show that even small quantities of dispersed water and reservoir solids can significantly affect measured viscosity. These findings are consistent with the correlation of heavy oil viscosity with solid content defined as asphaltenes and the proportion of solid maltenes at < 323 K (Hasan et al., 2009) that support the rheological model of heavy oil as “a dispersion of non-interacting uniform rigid spherical particles in a Newtonian fluid”. This non-Newtonian behavior at low temperatures may cause viscosity to vary by a factor of 2 to 5 or more with measurement shear rate due to shear thinning. Geochemists need to be aware of the complexities and analytical difficulties evident in “routinely” produced viscosity datasets. In general geochemical data is not the source of most experimental error in fluid property prediction studies!

Proper selection of oil samples representative of produced oil is crucial for viscosity-chemistry calibration data sets but can be challenging due to different sample types and storage and sample preparation procedures prior to viscosity measurement and geochemical analysis. Oil components that vary in the sample tracking oil maturity, source rock facies, in reservoir alteration and light end recharge, etc. are the dominant parameters in most correlations and thus must be accurately measured. For heavy oils, this means careful LEH retention or clear assessment of LEH loss during sampling, water removal, then accurate hydrocarbon and thioaromatic component separation, with good resolution of key parameters usually by GCMS and accurate quantification with sufficient internal standards to produce accurate, precise concentration data. Our experience is that the thioaromatic compounds are key and thus “total hydrocarbon” analyses have been found to be more effective than analysis of LC separated saturated and aromatic hydrocarbon fractions. Quantitative GCMS analysis of total hydrocarbon fractions were made using internal and recursive standards (Bennett et al., this publication).

Correlations between fluid properties and geochemical data can be developed using component concentration data or parts of, or whole, chromatograms of the key components as time series datasets with similar success. Ratios of geochemical components are widely reported in the literature, but PLS regression assumes linear additivity (i.e., variables or concentrations when summed are additive), a condition ratios do not honor. In general peak ratio geochemistry is of limited value in predictive geochemistry and with routine accurate quantitation now available it is no longer as necessary. Here concentrations of aromatic, thioaromatic hydrocarbons and saturated hydrocarbon biomarkers (C2, C3, C4, and C5 substituted alkylnaphthalenes; C1, C2 methylphenanthrenes; methylthiophenes; C20 to C28 mono- and tri-aromatic steroid hydrocarbons and selected steranes and diasteranes) were used as correlation proxies with the lighter aromatics excluded due to volatility and variable loss during storage. To minimize variability related to experimental drift in large datasets data transformation by autoscaling concentrations (Pirouette Infometrix) or aligning chromatograms is important. To ensure calibration samples span the full compositional range of the prediction set, hierarchical cluster analysis (HCA) and/or Principal Component Analysis (PCA) were applied to define the most characteristic variables. Here only core oil was used, but oils from different sample types can be used in the calibration or prediction data set if they are all part of a single population.

Clearly errors in viscosity data will be present in any large dataset collected over a period of time reflecting variable storage histories and oil extraction at ranges of times after drilling. In this case, individual viscosity values varied by up 30-80% in the worst cases from “true values” equivalent to drill time dead oil viscosity (DTDV) due to long storage times and centrifugation extraction effects for some samples. To compare these data, a correlation curve between equivalent compaction and centrifuge extracted oils was developed and centrifuged oil viscosities corrected to a common basis. A method to correct measured oil viscosities of stored core to equivalent dead oil viscosity at drilling time called the storage time viscosity correction

(STVC) was developed (Adams et al, 2008). The best algorithm for STVC varies with the natural logarithm of storage time assuming binary end member mixing of light end and heavy end fraction viscosities consistent with published algorithms of diffusive mass transport through polymers and oil spill evaporation. A similar correction for viscosity using sample storage time as a dependent variable via PLS regression, as described below, is possible using the fraction of the oil not susceptible to evaporation correlated to measured viscosity of fresh oils (short storage time). This PLS method can correct for storage time, mud contamination and sample type (e.g., produced oil, cuttings, core etc) when estimating viscosity from non-ideal calibration data.

Results

The PLS models using the whole training dataset of oil component concentrations predicted the measured viscosity over the whole viscosity range (5 decades) with RMSE of 139,187 cP. Inclusion of storage time and relative depth from the top of the bitumen pay zone plus removal of the 11 statistical outliers improved prediction accuracy (RMSE = 72,541 cP), but removal of 30 samples with storage time longer than 200 days (old samples) most significantly improved the model (RMSE = 31,503 cP; *best fit model*). The *best fit model* predicts reasonably accurate dead oil viscosity values across this field in the training set (predominantly oils with viscosities $\gg 100,000$ cP) and with the inclusion of storage time allows for prediction of viscosity at 30 days of storage time for all the samples including the additional 200+ chemistry only samples and the 30 old samples. Including storage time corrected viscosities of the oil samples in the training set did not significantly improve predictions. Using chromatograms in place of the concentration data also provided good results for the fresher viscosities because they provide additional chemical information in the small peaks not included in the integrated concentration dataset. Chromatograms can be used more reliably in data sets with uniform maturity and oil sources as well as sample handling and storage time, especially if internal standards have not been used. The PLS models poorly predicted the most degraded, high viscosity samples in the data set because of the non-linear variation of chemical composition with viscosity. In the most degraded oils, the key parameters controlling the PLS model like dibenzothiophene had been removed by biodegradation, such that either additional calibration data were needed to refine the linear regression or a separate PLS model was needed to predict viscosity $> 400,000$ cP.

Generally good linear fits of observed and predicted viscosities are obtained with the ANN model trained on 30-day viscosity corrected data set. The best equivalent model by far included both relative depth and the storage time viscosity correction data and better predicted the highest viscosities (RMSE ~ 9500 cP), possibly due to the model's non-linear basis. The ANN model produced similarly accurate results using only a subset of the aromatic hydrocarbon data, which could save on analytical costs. Comparisons of predicted and measured viscosity in a few vertical well profiles showed that the STVC estimated 30-day viscosities match the PLS 30-day and ANN 30-day estimated viscosity data.

Conclusions

The accurate prediction of oil mobility is crucial for effective exploration and design/optimization of production strategies especially in biodegraded heavy oil and bitumen reservoirs which exhibit strong fluid property heterogeneity. To minimize inclusion of spurious error in the models, calibration data and analytical and data processing procedures for geochemical viscosity predictions must be robust enough to represent subsurface oil variability and not be perturbed by sample handling and analytical methods. Quality control of viscosity data is crucial and legacy viscosity data sets typically need correction before use. Multivariate statistical models of heavy oil composition, based on GCMS molecular geochemistry data and using PLS and ANN regression models accurately predicted a wide range of dead oil viscosities for samples covering a 5 decade

range in viscosity and stored for variable times up to 4 years. Overall, the *best fit model* required storage time and sample relative depth as dependent variables and the exclusion of the samples stored for more than 6 months. The measured viscosities corrected using the STVC model compare well to independently determined PLS storage time viscosity at 30 days of storage time. The approach is sufficiently accurate for initial heavy oil and bitumen field development and production sweet spot targeting, especially for cold production operations viable for dead oil viscosities $\ll 100,000$ cP, especially in areas where insufficient liquid oil can be extracted from the core. The non-linear impact of storage time effects on oil composition and viscosity in this large complex history dataset has however compromised the accuracy of the model but similar studies with smaller datasets and fresh samples typically allow viscosity predictions with accuracies better than 5-10%. In general, long term analytical data quality and reproducibility, the biogeochemical and geological complexity of the system and the large magnitude and variability of sample storage effects are the ultimate limits on viscosity prediction using geochemical data.

Acknowledgements

This work was partly supported by Shell, the Alberta Ingenuity Fund and by NSERC and PRG research funds. The authors acknowledge the Scott Ramos of Infometrix for assistance with Pirouette, and data management support by Ben Hsieh. Haiping Huang and Mark Yunker provided useful commentary.

References

- Adams, JJ 2008 Fluid variation in biodegraded oil reservoirs. PhD Thesis U Calgary.
- Adams, JJ, Jiang, C, Bennett, B, Huang, H, Oldenburg, T, Noke, K, Snowdon, LR, Gates, ID and Larter, SR, 2008. Viscosity determination of heavy oil and bitumen; Cautions and solutions: Heavy Oil Congress, March 9 – 12, Edmonton, Alberta, Paper 2008-443.
- Bennett, B, Huang, H, Fustic, M and Larter, SR, 2010. Geochemical Solutions to Heavy Oil Production Problems AAPG Hedberg Conference, Vail Colorado, June 7-10, 4 pp.
- Crampon, C, Trassy, L, Auaulle, L, Neau E and Coniglio, L, 2004. Simplification and extension of a predictive group contribution method for estimating heavy organic pure compound vapor pressures - I. Hydrocarbons, Fluid Phase Equilibria v. 216, p. 95–109.
- Hasan, A., Fulem, M., Bazyleva, A., and Shaw, J.M. 2009. Rheological Properties of Nanofiltered Athabasca Bitumen and Maya Crude Oil. Energy Fuels, in press.
- Koopmans MP, Larter SR, Zhang CM, et al. 2002. Biodegradation and mixing of crude oils in Eocene Es3 reservoirs Liaohe basin China. AAPG BULL 86 (10): 1833-1843.
- Koopmans, MP, Larter, SR and Tegelaar, EW. 1998, Viscosity prediction of reservoir heavy oils using molecular geochemistry and neural networks; in 215th ACS national meeting., 215; 1-2, Abstracts of Papers – American Chemical Society, National Meeting.
- Larter, S. R., J. J. Adams, I. D. Gates, B. Bennett, and H. Huang, 2008, The origin, prediction and impact of oil viscosity heterogeneity on the production characteristics of tar sand and heavy oil reservoirs: Journal of Canadian Petroleum Technology, v. 47, p. 52-61.
- Larter, S., Jiang, C., Noke, K., Adams, J., Gates, I., Bennett, B. and Oldenburg, T. Method and Apparatus for Obtaining Heavy Oil Samples from Reservoir Samples. Canadian, Patent Pending 2006. Application No. 2,578,319.
- Malthouse, EC, Tamhane, AC and Mah, RSH, 1997. Nonlinear Partial Least Squares, Computers Chemical Engineering, v. 21, p. 875–890.
- Marrero, J and Gani, R, 2001. Group-contribution based estimation of pure component properties Fluid Phase Equilibria, v. 183-184, p. 183–208.
- Qin, S. and McAvoy, T. 1992 Non-linear PLS modeling using neural networks. Computers Chemical Engineering v. 16, 4, p. 379-391.
- Rosipal, R and Kramer, N, 2006. Overview and Recent Advances in Partial Least Squares, in Subspace, Latent Structure and Feature Selection Techniques, Saunders, et al. (Eds), Springer-Verlag, Berlin..p. 34–51.