UNCONVENTIONAL RESOURCES TECHNOLOGY CONFERENCE
FUELED BY SPE • AAPG • SEG

URTeC: 2671117

Predicting Sweet Spots in Shale Plays by DNA Fingerprinting and Machine Learning

Chris Te Stroet*, Jonathan Zwaan*, Gerard de Jager*, Roy Montijn**, Frank Schuren**
*BIODENTIFY **TNO

Summary

This paper presents a method to generate a >70% accurate predictive map of sweet spots in shale plays prior to drilling. It indicates where to drill, and where not. The approach uses DNA analysis of surface soil samples, to derive information on the mix of microbial species in the samples. Using our database to correlate DNA in soil samples and production data of earlier drilled areas, the new DNA fingerprint is an indicator of the presence of vertical micro-seepage to the surface from hydrocarbon accumulations in the subsurface - including sweet spots in shale plays. In times of low oil and gas prices, stepping away from grid drilling and implementing an iterative procedure of prioritized development of higher profitable areas of a play, could prove a game changing strategy.

First technological break-through: DNA 'fingerprinting', biotechnology
The occurrence of vertical upward micro-seepage has been known for decades and is extensively described in the literature. But the microbial life is much more complicated than the few species that were known as hydrocarbon-oxidizing bacteria. It is necessary to determine the complex composition of microbes -not only those that flourish at micro-seepage sites, but also those that are eliminated under such conditions and are therefore found in reduced concentrations above sweet spots. Recent developments in DNA analysis techniques have made this complex and previously expensive problem efficiently and economically solvable.

Second technological break-through: Big Data, Machine Learning, super computing
The millions of microbes counted in thousands of soil samples by applying 16SrDNA 'fingerprinting' techniques create terabytes of data that must be correlated with the presence of hydrocarbons. This is a huge mathematical and computational big data problem. Advancements in machine learning applications together with parallel computing (Hadoop in the cloud, GPU) have made it possible to construct robust and reliable predictive DNA based models for sweet spot locations.

The combination of both technologies will be illustrated with two case studies: 1) a validation case in the Haynesville shale, an area with known production data, and 2) two areas in The Netherlands where the prospectivity of two shale formations was estimated.

Introduction

The patented technology produces a >70% accurate predictive map of the highest producing areas in a shale play, so called sweet spots, using DNA analysis of surface soil samples. The DNA analysis contains information on the mix of microbial species in the soil samples which is a direct indicator for the presence of vertical micro-seepage (red arrows in Figure 1) from hydrocarbon accumulations in the subsurface. This paper describes this technology and its validation in detail. For exploration, its paves the way to highly effective exploration drilling and investment decision-making. For production, it leads to significant cost reduction by optimizing drilling scenarios and increased

development speed of a play. The accuracy of the predictive map is iteratively increased from about 70% before drilling to up to 85% by incorporating information from the drilled wells with local, field specific microbial information.

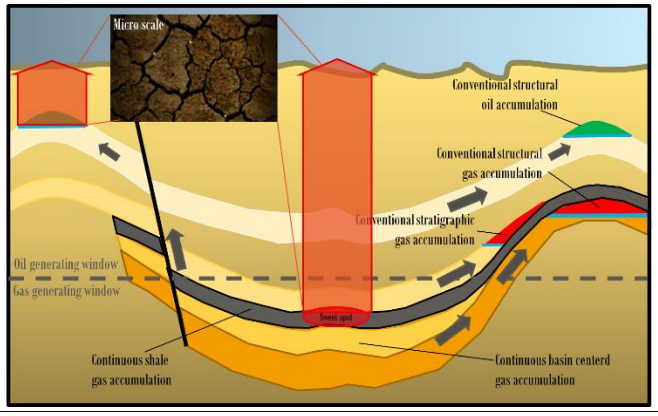

Figure 1: Micro-seepage by buoyancy of colloids from subsurface hydrocarbon accumulations moving vertically upward through micro cracks towards the surface.

The presence of vertical upward micro-seepage is extensively described in the literature. The upward migration of colloids (a few molecules clustered together) of light gaseous hydrocarbons (C1-C5) from subsurface petroleum accumulations, and the use of specific micro-organisms as indicators for this process, was already described by German pioneers in this field (Laubmeyer, 1933) as well as by scientists in the USSR (Horvitz, 1939). Bacteria are present in high numbers and variation of species in soil samples. Their exceptionally high adaptivity to grow on small amounts of nutrient sources form the micro-seepage, produce a very valuable 'soft sensing mechanism'. Some bacteria oxidize the carbon molecules and use the micro-seepage as a carbon source for their metabolic activities and growth. The use of this principle for exploration, was already described within the USSR (Mogilevskii) in 1940 to 1959. After these publications, scientists in the US documented a correlation between higher hydrocarbon concentrations originating from oil and gas fields and hydrocarbon-oxidizing bacteria measured in surface soil samples (Davis, 1956, Sealy, 1974, Miller, 1976). The metabolism of hydrocarbon oxidizing bacteria is the cause of the development of near surface oxidation- and reduction-zones and the alteration of soils and sediments above the reservoir. Differences in indirect measurements indicating the activity of oxidizing bacteria form the basis of geochemical exploration techniques developed during the 90s such as radioactivity, carbonate, pyrite, magnetic, electrical, and satellite-based methods (Schumacher, 1996, Wagner, 2002, Schumacher, 2012, Rasheed et al., 2013). It was shown that clear and sharp anomalies of -geochemical- properties could be indicated very precisely at the border of reservoirs due to vertical upward migration of micro-seepage. In the same period the transport mechanism was also investigated (Klusman and Saeed, 1996, Saunders et al, 1999). The commonly accepted conclusion is that micro-seepage is caused by buoyancy of very small colloids of molecules (nanoscale) through micro-cracks (microscale) that are present even in 'impermeable material' (Figure 2a). Micro-seepage yields sharp anomalies right above hydrocarbon accumulations because of the distribution of a colloid going randomly left or right through a micro-crack is tightly clustered when this process is repeated many times (Figure 2b). Field observations show that these surface anomalies, within a few years after production has been stopped, disappear. This fits the explanation in Klusman and Saeed (1996) that micro-seepage is caused by vertical buoyancy, driven by the difference in gas and water pressure, resulting in a significantly faster process than convection-diffusion as long as the pressure difference is large enough. This also means that distinguishing different vertical payzones above another must come from other geo-knowledge/monitoring.

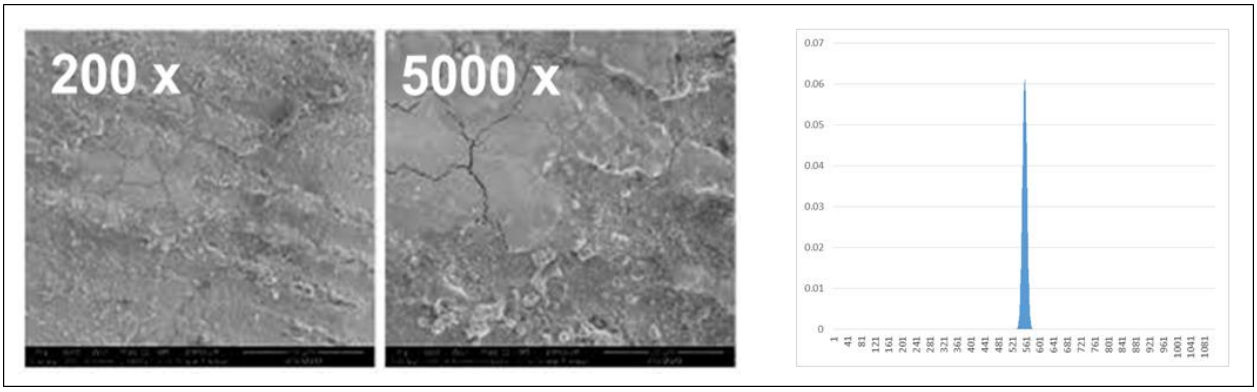

Figure 2: a) micro cracks are present in 'impermeable' layers b) peak distribution of a 'stochastic tree'

Despite the long history of research and publications, problems were encountered with the acceptance of microbial exploration techniques. Microbial life is much more complicated than the few species that are known as hydrocarbon oxidizing bacteria. Furthermore, these bacteria were costly to quantify because a different agar plate culture had to be used for every species. It was even much less accurate when measured indirectly by geochemical methods. Additionally, acceptance was influenced by a competing and superior technology, seismology, developing very rapidly in the same period.

The development of recent DNA analysis techniques makes the microbial quantification of species comprehensive and economically feasible. The millions of counted species in thousands of soil samples by applying 16SrDNA techniques creates an enormous amount of input data (terabytes in size) that must be correlated with the presence of hydrocarbons. This is a serious big-data mathematical and computational challenge. The progress made in supercomputing makes it possible to construct robust and reliable predictive models by applying machine learning techniques derived from pharmaceutical applications. The strict experimental design rules in this field are used to find a complex composition of microbes, not only those who flourish by the micro-seepage, but also those who are terminated by it and that have therefore lower concentrations above sweet spots. The latter category could never be found by the earlier adopted microbial exploration techniques, however have proven to be essential for identification of sweet spots.

Our primary focus for the application of this technology therefore is to identify sweet spots in shale plays where seismic data is insufficiently indicative or too expensive, an area where a robust, economic, and accurate technology is not yet available. It is shown that microbes, sensing vertical micro-seepage, identify sweet spots and low productive areas consistently and adequately.

This article consists of three parts: The first part describes the methodology of the workflow; the second part covers validation cases in the Haynesville shale and a pre-drilling case in shale formations in the Netherlands. Third and last, conclusions together with remaining uncertainties are summarized.

Methodology

To generate an accurate and predictive map with sweet spots highlighting the most prospective areas, the following steps are followed:

1. Soil sampling in the field, followed by DNA analysis to get DNA-fingerprints of the microbes.
2. Selecting a training set from our database with sample data from earlier drilled and known areas. The training set has similar DNA fingerprints compared to the new samples and can correlate DNA from soil samples with production data.
3. Modelling and validating the microbes that determine the sweet spots.
4. Mapping and analyzing the prospectivity / sweet spots / low production areas of the target play.

1. DNA analysis to get DNA-fingerprints

DNA based methodologies are broadly applied in life sciences. The methodology described here is derived from the workflow to determine a DNA fingerprint in life sciences applications (extraction, multiplication with PCR, and sequencing of DNA), modified for this specific Oil & Gas soil application. The methodology is based on identification of bacteria, present in soil samples, based on their 16S ribosomal DNA sequence (hereafter 16S sequence). The 16S sequence of bacteria is generally used as a genetic fingerprint of bacteria and consists of conserved and variable regions. The sequence of one such variable region is determined by Illumina sequencing. Similarities and differences in the 16S sequence are translated to bacteria. In Figure 3 a schematic representation of the steps required for the DNA-based analysis of soil samples is shown.

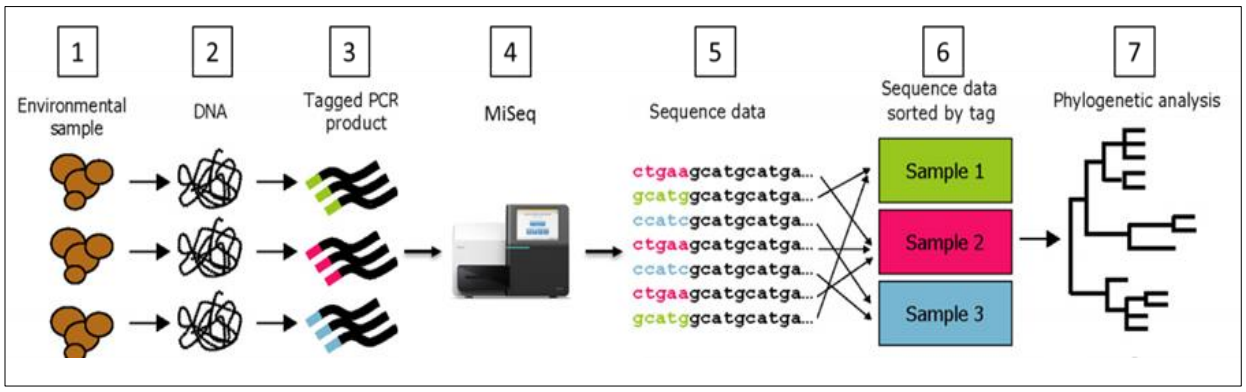

Figure 3: Workflow for getting DNA fingerprints for soil samples: 1) Soil samples preparation for DNA analysis, 2) Extraction of bacterial DNA from soil sample, 3) Amplification of the 16S rDNA by Polymerase Chain Reaction, 4) 16S Sequence analysis using next generation Illumina MiSeq, 5) Processing raw data from MiSeq to verified 16S sequences, 6) Processing verified 16S sequences back to individual soil samples and 7) Interpretation of 16S sequence data: translation to bacterial genera (families); the steps 2 and 3 are tuned by BIODENTIFY to get maximum information on species.

Analyzing the composition of the microbial population present in the soil samples requires extraction of microbial chromosomal DNA from the samples. This is performed by mechanical disruption of the bacteria using small beads and vigorous shaking. To handle the large number of samples in this study, extraction of chromosomal DNA from the micro-organisms present in the soil samples was performed in 96-well deepwell plates (Figure 4).

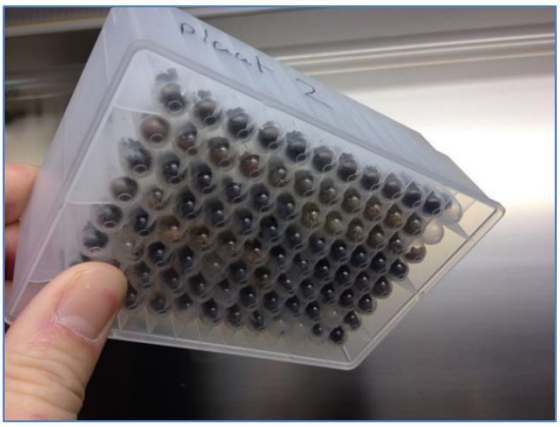

Figure 4: Example of 96-wells plate filled with soil samples, ready for extraction. Samples from different locations sometimes differ in their appearances as is clear from the differing coloration in the wells of this plate.

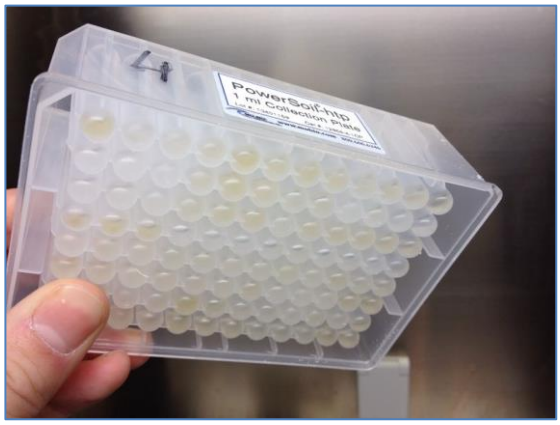

Figure 5: Example of sample-to-sample differences in impurities after DNA extraction.

A specialized, proprietary DNA extraction method was developed that yields maximal successfully amplified samples but also can be automated to prevent extreme costs because of a growing number of samples. It is based on bacterial

cell lysis by bead-beating using zirconium beads, removal of inhibiting compounds using proprietary solutions provided by adapting commercial available kits, binding and washing of the DNA and finally elution of the pure DNA.

After amplification and sequencing the processing is finalized in a series of special designed quality control steps which together are called the 'pipeline' (Figure 6). Initially, the number of DNA segments equals the number of unique sequences (>35 million sequences). During the processing, the total number of sequences (blue bars) slowly decreases to ~29 million sequences. However, the number of unique sequences (crimson bars) diminishes, finally resulting in a fraction (~340 thousand) of the initial number of sequences. These unique sequences form the basis for the identification of biomarkers as they are translated from sequences to families or genera. Figure 7 shows a representation of the relative abundance of bacterial families or genera within 7 different samples.

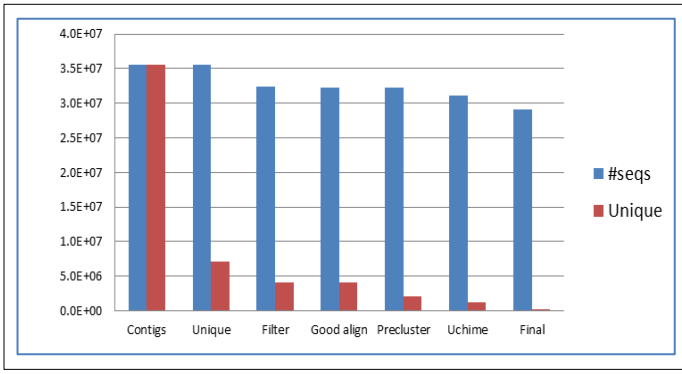

Figure 6: Schematic representation of the effect of the various proprietary filtering steps within the 'pipeline'. Please note, at this scale, the number of unique sequences at the final stage is hardly visible (value ~340 thousand).

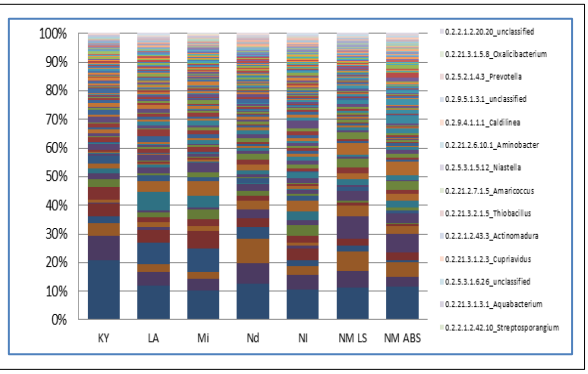

Figure 7: Schematic representation of the bacterial diversity and relative abundance within samples. Each colored bar represents a bacterial genus or family. Please note that the legend on the right only shows a minor fraction of all bacterial diversity.

2. Using our database with samples from drilled and known plays, to correlate DNA in soil samples with production data

When predicting the productivity of undrilled, target locations where only DNA fingerprints of shallow soil samples is available, it is necessary to have a correlation model that connects the new fingerprints of the new target location with a prospectivity index. This correlation model is produced by using our database, containing DNA fingerprints and production data from executed projects. Currently (May 2017), we have well over 2000 samples with known production data from six different and varying shale plays in our database (Figure 8):

1. The Haynesville Shale: in 2013 the highest shale gas producer. Situated in hot and moist climate (bayous).
2. The Bakken Shale: in 2013 the highest shale oil producer. Situated in a land climate with limited vegetation (prairies).
3. The Antrim Shale: one of the oldest shale gas producers and also a -proven- biogenic shale play, situated in forested areas.
4. Avalon and Bone Spring: Situated in a desert environment.
5. Lewis Shale: Best producer in the Rocky Mountain province.
6. Big Sandy: Good gas producer in the North East.

Figure 8: Map of the sampled shale gas and shale oil plays

Each sample contains up to hundreds of thousands counted different species. The correlation of each of these samples (a vector **x**) to the newly sampled DNA fingerprints is a measure for the similarity of the DNA fingerprints in the database with the new one. This correlation, or more precisely a "co-occurrence measure", is estimated by unsupervised k-nearest neighbor techniques (kNN) addressing the high dimensionality of the problem. The co-occurrence is expressed by the cosine similarity. By selecting the samples from the database that have the highest correlation with the DNA fingerprints of the new samples, a training set is generated, where fingerprints and productivity are known. This training set is used to find the correlation model with Machine Learning algorithms to predict the prospectivity at new locations (see next step 3 described in detail below). Because the database has samples from different and varying plays (hydrocarbon source, climate, geology), creation of a training set with a sufficiently strong correlation between samples from drilled areas and new locations can be realized.

3. Modelling and validating the microbes that determine the sweet spots

The next step is to correlate the results of the sequenced analysis (the bacterial diversity in the DNA-fingerprints of all soil samples) with production data in the selected training set. The goal is to find correlations between production and presence or absence of specific bacteria, typically about 50-200 out of the 340,000 species in our database, the so-called biomarkers. A biomarker that is oxidizing the colloids of gas that are transported by micro-seepage is more abundant above sweet spots. But the difference in abundance is not distinctive, it is therefore not an absolute indicator. In fact, one biomarker will result in a map with a lot of noise (compare Figure 9a and 9b). The model must 'find the signal in the noise'. Only when many biomarkers (> 50) are used together, a sufficiently accurate estimate can be made and the location of sweet spots become visible in the map (Figure 9c).

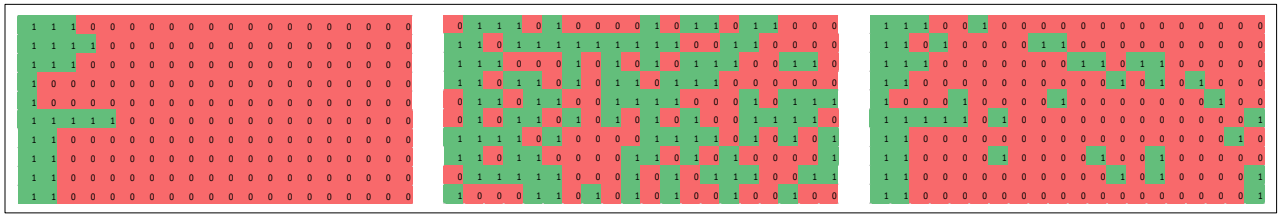

Figure 9: a) area with a sweet spot (in green) b) modelled sweet spots using 1 biomarker c) modelled sweet spots using 70 biomarkers

To build a model that consists of the 50-200 differentiating biomarkers is a real Big Data challenge. It aims to find both those species that are more present above sweet spots because they use the seeped colloids of hydrocarbons for their metabolism, and those species, that are less present above sweet spots because the population is partially eliminated by the hydrocarbons. The number of possible solutions is a power law of the hundreds of thousands of possible species, resulting in an almost infinite solution space. A special Machine Learning algorithm has been

developed to perform this task. State of the art methods were used to deal with 'sparse modelling' issues like non-linearity, the influence of noise and to prevent overfitting (Gaussian kernels for non-linear problems (Cortes, 2012) and L1-based regularization methods (Mosci, 2011)).

To find the most robust and reliable predictive biomarkers, we adapted the above described algorithms in a triple-loop validation/prediction procedure (developed for medical applications by TNO; Figure 10). To ensure that the model correctly captures the information contained in the dataset, samples are randomly shuffled, based on the input data matrix of microbial abundances. Next, the reshuffled matrix is split into a calibration subset (randomly 70% of the training set) and a validation subset (30% of the training set). The calibration set then is subjected to N-cross folds to estimate the modelling parameters N-times. The final selected parameter set minimizes the average misfit in the predictions (inner loop in Figure 9). These parameters are used on the separately reserved validation set to confirm the validity of the model and the identified biomarkers. Since variations inevitably occur within this type of computational modelling, 1000 repetitions are run, thus selecting only the most stable biomarkers over all 1000 x N simulations (second loop in Figure 10). The model is aimed at reducing the number of biomarkers until an optimum in prediction accuracy is achieved where the prediction is at its optimum and the number of biomarkers is at its minimum (expressed in a ROC curve; Figure 11). Only then a prediction can be made for a new area, and a model can be improved with new measurements (outer loop in Figure 9). The 'price' for calculating a robust predictive model is the compute time needed for calibrating 1000 x N models (to achieve the best final prediction). This is solved economically by computing in the cloud using Hadoop technology (32,000 processors) and using GPU clusters in-house to accelerate the highly parallelizable calculations.

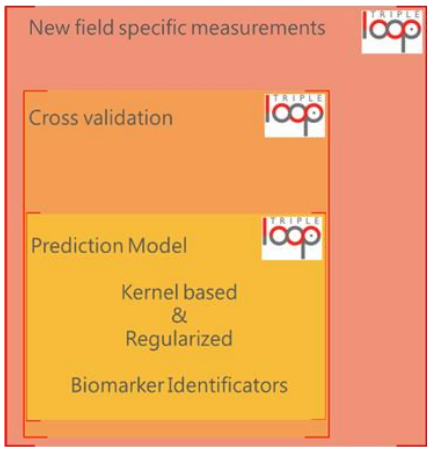

Figure 10: Schematic representation of the machine learning 'triple loop' modelling procedure.

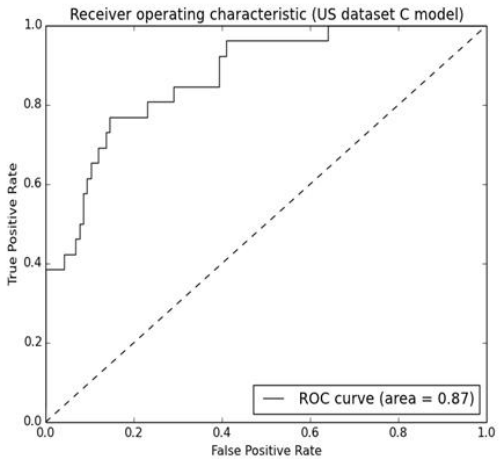

Figure 11: Representation of the performance of a binary classifier system (prospective vs. non-prospective). The fraction of true positives out of the total positives (TPR – true positive rate) is plotted versus the fraction of false positives out of the total negatives (FPR = false positive rate) for all weighted biomarkers. The 'ROC curve' indicates the area under the curve, indicating that the reliability of the predicted biomarkers is 85%.

4. Mapping and analyzing the prospectivity / sweet spots / low production areas of the target play

With 1000 models 'trained' (every time on another random 70% subset and validated on 30% of the samples that are left out from the training set) we can predict a value that indicates whether this location will be a sweet spot or not for each new location with a DNA fingerprint (dimensionless value range out of a Machine Learning classification algorithm: between -1 and 1). These values, on the sampled locations, are then contoured onto a map, showing the estimated sweet spots in the target area.

The use of DNA analysis and the modelling to produce the sweet spots map, is summarized by the workflow in Figure 12:

a) Take shallow soil samples

b) In a grid over an area where the grid distance depends on the expected size of the sweet spot (grid spacing typically in the range of 0.5-1 mile to increase the change a sweet spot is sampled multiple times)
c) Analyze every soil sample in the grid on DNA fingerprints
d) Estimate productivity with a predictive model (generated using a training set from our database), using the DNA fingerprints of soil samples as model input, then draw contour estimates to a map

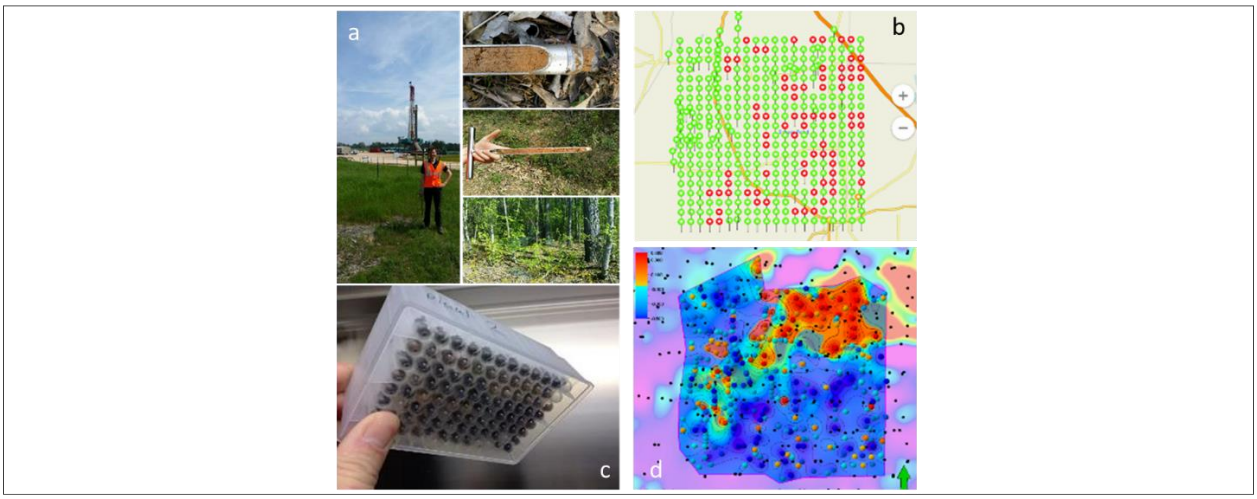

Figure 12: Steps to produce a sweet spot map

Cases

Validation case in Haynesville. In accordance with the workflow described above, the technology is illustrated with an example of a 15x15 miles area in the Haynesville:

1. A grid with 362 locations were sampled, 20-50 cm below surface elevation (Figure 13c on the right). The productivity map (Figure 13b in the middle), from actual well production data, is used to validate the generated model.

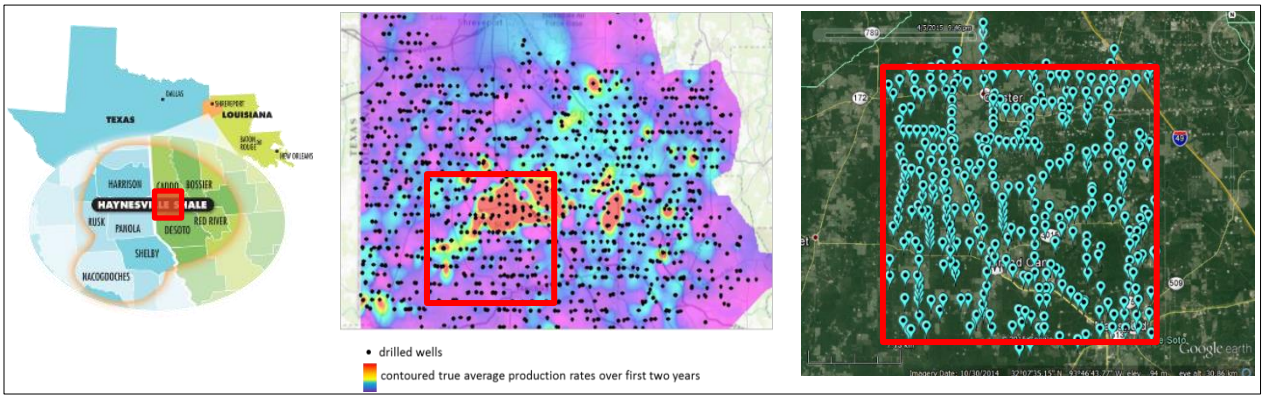

Figure 13: a) pilot Haynessville shale b) drilled wells and contoured production (2 years) c) locations of soil samples orange = top 10% of producers in this play

2. Generate a model with the triple loop modelling procedure applying selected sample data from the database that show sufficient similarity with the local DNA fingerprints from the target area.
3. Predict the chance of a sweet spot at grid sample point location by using the model from the training set and the local DNA-fingerprint data only, and generate a first sweet spot prediction map by interpolation of the predictions. In this case, successful DNA analyses were generated at 314 locations. For those locations, the predictions are shown as colored points on the map which are subsequently contoured (Figure 14). Since this area is used as a validation case, the actual drilled wells with production rates are known (background map of Figure 14) and these

are used to benchmark the DNA based predictions. As mentioned above, the model is trained with data from our database from other shale areas, and the model productivity predictions are made solely by using local soil sample DNA fingerprints (bright colored contours at foreground of Figure 14). The map in Figure 13 provides the predicted sweet spot map before drilling, with 72% accuracy. The highly productive area in the North East is clearly predicted, but there are also some 'probably false sweet spots' (sweet spots predicted by our model but not seen in the validation / productivity map) in the South East.

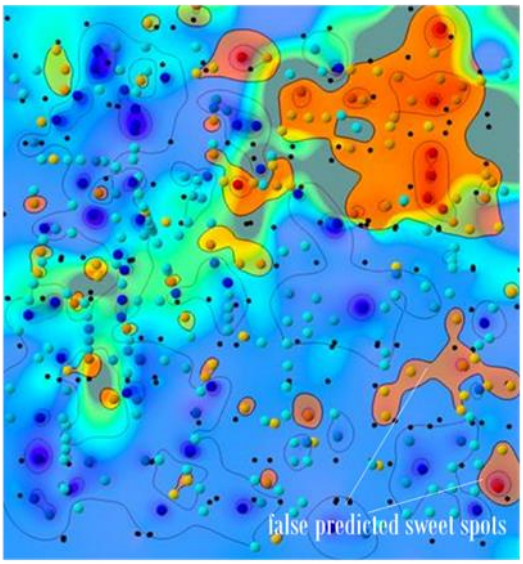

Figure 14: Prediction map of sweet spots without using local production data (can be seen as predicted map before drilling).

4. Wells will be drilled targeting the predicted sweet spots. When drilled, production information from these wells will be used to increase prediction accuracy by adding these locations with known data to the training set, then re-model. This is how we generated our final map, using 'new' and local field information from the validation data. Applying this principle, a new map with increased accuracy, now 86% (Figure 15) was produced. The mismatches on individual isolated locations reflect the 'noise in the data', always present in nature, but as there is no correlation between these points, it will not impact the decision-making based on this predictive map.

   In reality, 97 wells were drilled in this area, with 27 high producers (table with Figure 15). These numbers showcase the value of the methodology: through smart and prioritized drilling, based on updated predictive maps, more hydrocarbons can be produced even while drilling significantly fewer wells.

	actual drilled wells	with DNA information
low producing	70	15
high producing	27	31
total	97	46

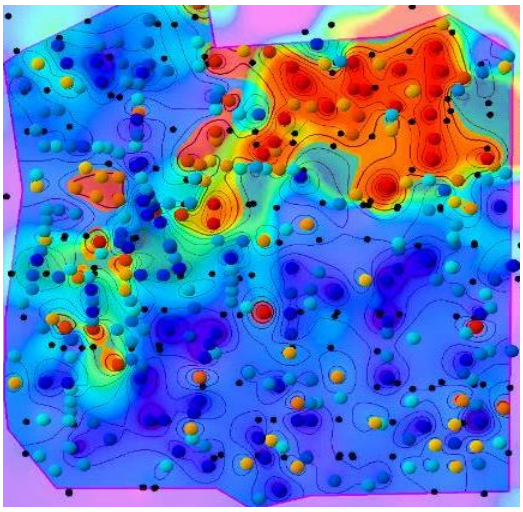

Figure 15: Final estimated sweet spot map after 'virtually drilling' wells and therefore iteratively increasing the prediction accuracy. By using the DNA fingerprints the area is (virtually) produced with significantly fewer wells yet with a higher absolute number of wells in sweet spots

Undrilled pilots in the Netherlands. The second case is carried out prior to drilling. In the Netherlands two pilot areas were sampled and analyzed on DNA fingerprints to produce predictive maps. As a training set, again data from the US database is used, with sufficient, validated similarity in microbial DNA. The maps, from two different formations (Geverik and Posidonia) are showing a significant difference (Figure 16): in the Geverik no shale gas is predicted while in the Posidonia there is a clear indication of sweet spots. Both maps are generated using the same trained model from our database. The absence of shale gas in the Geverik map confirms the existing theory that the circumstances for the formation of shale gas were unfavorable because it is on top of Visean carbonate structures (Harings, 2012). Naturally, the outcome can only be verified after drilling, but the plausibility of the maps is further demonstrated because the DNA fingerprints in the Geverik area were positive in the small zone above a conventional prospect (prospect mapped based on seismic data and geological interpretations), whereas the known and produced (now 'empty') fields in the Posidonia area were not captured. These findings again support the earlier explained concept of micro-seepage with a relatively high velocity of buoyancy where the signal in the DNA fingerprint disappears within a few years after production because of a decrease in reservoir pressure.

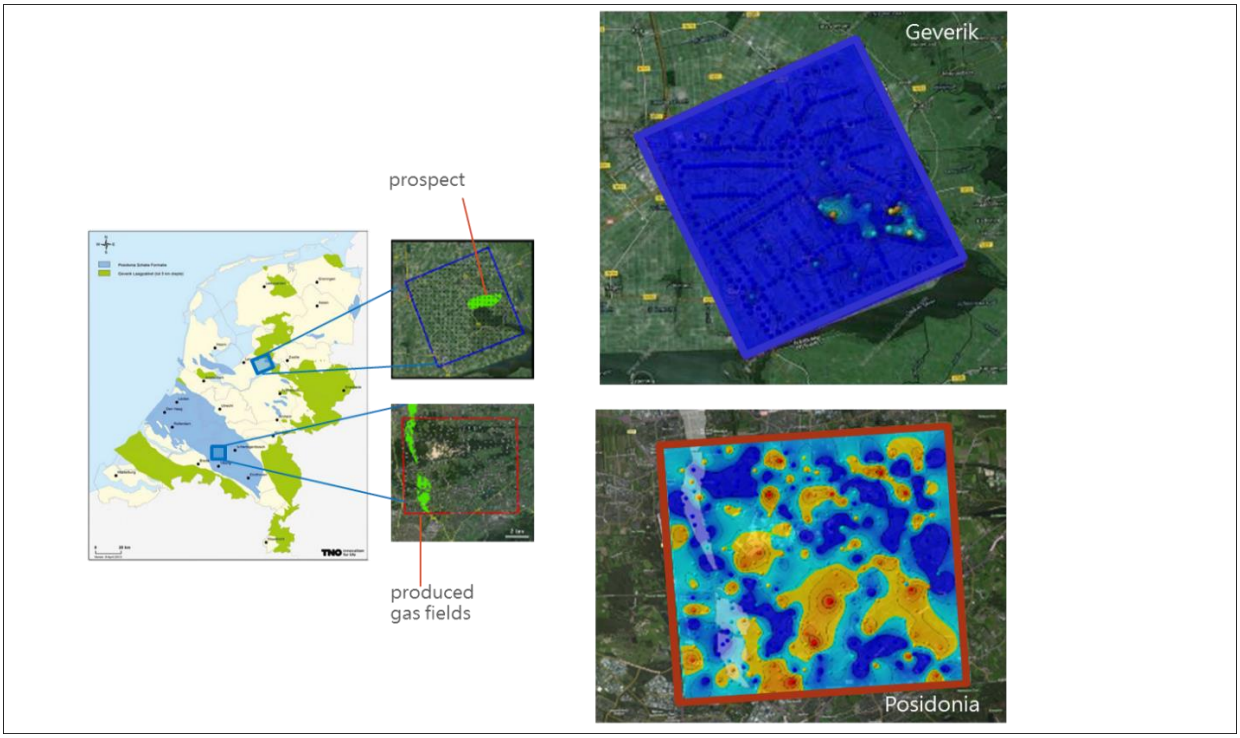

Figure 16: Predicted sweet spot maps of two areas in the Netherlands: one of the Geverik and one over the Posidonia formation.

Discussion and outlook

We propose that sweet spot prediction, with an accuracy of 70% prior to drilling, by using DNA fingerprints from shallow soil samples is a valid additional exploration tool. Our case studies show reproducible results supporting this. The presence of vertical micro-seepage by buoyancy through microcracks is a difficult process to prove because it cannot be measured directly (as opposed to macro-seepage). Although many geochemical anomalies have been detected for conventional fields and distinctive differences in biomarkers above sweet spots in shale are found as described in this article, proof of micro-seepage remains an indirect case.

Some additional remarks:

- What to do in case of stacked layers where also conventional reservoirs are produced in the same area? A good example is the Avalon/Bonespring formations in the Permian basin. Conventional reservoirs are mostly known

and already produced; its projected areas on the surface (usually less than a few % of the total area) can be excluded from the shale grid sampling, but the different sweet spots in stacked shale layers may still overlap. The method in this paper can delineate the summed lateral spreading of the stacked sweet spots. Additional geological (drilling/seismic/other) information or interpretations is needed to determine the vertical origin of the hydrocarbon.

- The quality of completions clearly has a major effect on the productivity of the well. For samples in our reference database, completions influence the actual production rates and thus influence the classification used in the training set that we correlate with. Therefore, classification of well productivity in our database is done on a relative base, e.g. the top 10% producing wells compared to wells in the same area / play, that are drilled in the same period (using comparable technology).

Despite these remarks, the authors believe that the outlook of this methodology is very positive. The accuracy in predictability that can be reached constitutes to very significant cost savings. And with more data becoming available, our database increases in value as it is a learning system: every new sample point and drilled well increases the information content that can be used to predict new wells drilled in the future.

A good example is our work with subsea samples, now available in the database, used to de-risk offshore prospects and wells: The same methodology is used but grid sampling is replaced by sampling cross-sections of seabed samples over a prospect or around a well (Figure 16). The insight used to apply this methodology is not new. The innovation comes from the ability to couple this with state-of-the-art technologies of recent DNA developments and the progress in supercomputing and machine learning for handling big data challenges.

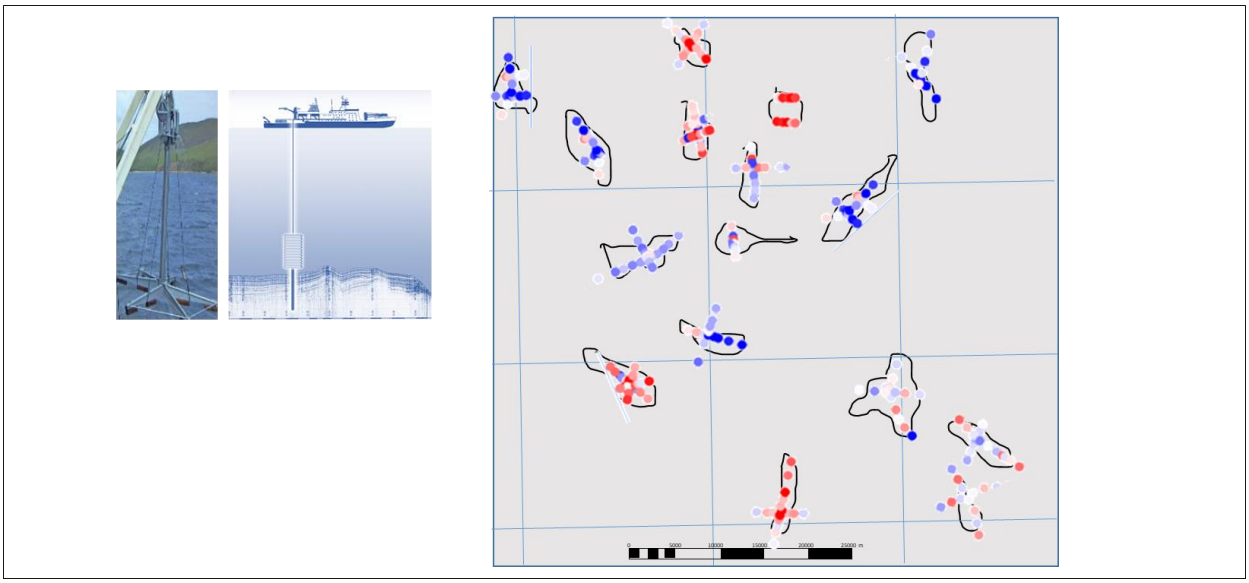

Figure 16: Using DNA fingerprints to de-risk offshore prospects.

References

[1] Caporaso, J.G. et 13 others (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME Journal 6, 1621-1624.
[2] Cortes, C., M. Mohri, A. Rostamizadeh (2012) Algorithms for Learning Kernels based on Centered Alignment. Journal of Machine Learning Research, vol. 3, 1-35.
[3] Harings, M.J., (2014) CAL-GT-01 - Implications for the Distribution of the Early Namurian Geverik Member in the
[4] Netherlands, 76th EAGE Conference Amsterdam.
[5] Horvitz, L., (1939) On geochemical prospecting-I. Geophysics, Vol. 4, 210-228.
[6] Davis, J.B., (1956) Microbial decomposition of hydrocarbons. Industrial and Engineering Chemistry, vol. 48, no. 9, 1444-1448.

[7] Klusman, R.W., M.A. Saeed (1996) Comparison of Light Hydrocarbon Microseepage Mechanisms. in "Hydrocarbon migratin and its near-surface expression", eds. D. Schumacher and M.A. Abrams, AAPG Memoir 66, 157-168
[8] Laubmeyer, G. (1933) A new geophysical prospecting method, especially for deposits of hydrocarbons. Petroleum London, vol. 29, 14.
[9] Miller, G.H., (1976) Microbial survey helps to evaluate oil and gas. Oil and Gas Journal, 4, 192.
[10] Mogilevskii, G.A. (1940) The bacterial method of prospecting for oil and natural gases. Razvedka Nedr, vol 12, 32-43.
[11] Mogilewskii, G.A. (1959) Geochemical methods of prospecting and exploration for petroleum and natural gas. Berkely Univ. of California Press, 349.
[12] Mosci, S., L. Rosasco, M. Santoro, A. Verri, S. Silva (2011) Non Parametric Sparsity and Regularization. MIT Computer Science and Artificial Intelligence Labatory Technical Report, 1-39.
[13] Rasheed, M.A., D.J. Patil, A.M. Dayal (2013) Microbial Techniques for Hydrocarbon Exploration. Chaper 9, in "Hydrocarbon" edts. Vladimir Kutcherov and Anton Kolesnikov, ISBN 978-953-51-0927-3, InTech.
[14] Saunders, D.F., K.R. Burson, C.K. Thompson (1999) Model for Hydrocarbon Microseepage and Related Near-Surface Alterations. AAPG Bulletin, vol. 83, no. 1, 170-185.
[15] Schumacher, D. (1996) Hydrocarbon-induced alteratiojn of soils and sediments. In "Hydrocarbon migration and its near-surface expression: AAPG Memoir 66", eds. D. Schumacher, M.A. Abrams, 71-89.
[16] Schumacher, D. (2012) Pre-Drill Prediction of Hydrocarbon Charge: Microseepage-Based Prediciton of Charge and Post-survey Drilling Results. AAPG Datapages, GeoConvention 2012 (Vision).
[17] Sealy, J.R. (1974) A geomicrobial method of prospecting for oil. Oil and Gas Journal, 15, 98-102.
[18] Wagner, M., J. Piske, R. Smit, (2002) Case histories of microbial prospection for oil and gas. AAPG Studies in Geology 48 and SEG Geophyscial References Series, vol. 11, 453-479.