GEOPHYSICS, VOL. 83, NO. 2 (MARCH-APRIL 2018); P. D27–D39, 11 FIGS.
10.1190/GEO2016-0566.1

A sequential dynamic Bayesian network for pore-pressure estimation with uncertainty quantification

Rachel H. Oughton[1], David A. Wooff[2], Richard W. Hobbs[3], Richard E. Swarbrick[4], and Stephen A. O'Connor[5]

ABSTRACT

Pore-pressure estimation is an important part of oil-well drilling because drilling into unexpected highly pressured fluids can be costly and dangerous. However, standard estimation methods rarely account for the many sources of uncertainty, or for the multivariate nature of the system. We have developed the pore-pressure sequential dynamic Bayesian network (PP SDBN) as an appropriate solution to both these issues. The PP SDBN models the relationships between quantities in the pore-pressure system, such as pressures, porosity, lithology, and wireline-log data, using conditional probability distributions based on geophysical relationships to capture our uncertainty about these variables and the relationships between them. When wireline log data are given to the PP SDBN, the probability distributions are updated, providing an estimate of pore pressure along with a probabilistic measure of uncertainty that reflects the data acquired and our understanding of the system. This is the advantage of a Bayesian approach. Our model provides a coherent statistical framework for modeling the pore-pressure system. The specific geophysical relationships used can be changed to better suit a particular setting, or reflect geoscientists' knowledge. We determine the PP SDBN on an offshore well from West Africa. We also perform a sensitivity analysis, demonstrating how this can be used to better understand the working of the model and which parameters are the most influential. The dynamic nature of the model makes it suitable for real-time estimation during logging while drilling. The PP SDBN models the shale pore pressure in shale-rich formations with mechanical compaction as the overriding source of overpressure. The PP SDBN improves on existing methods because it produces a probabilistic estimate that reflects the many sources of uncertainty present.

INTRODUCTION

Understanding the pore-pressure profile is crucial when drilling, so that the mud-weight profile can be designed appropriately. This mud weight forms a key part of any well plan. An example is shown in Figure 1. In general, the mud weight is designed to be slightly higher than the pore pressure. If the mud weight is too low because the pore pressure has been poorly estimated and a porous and permeable unit (e.g., a sandstone) is suddenly encountered, formation fluids may enter the wellbore (termed an influx) resulting in a kick, causing drilling problems and a well control incident. Conversely, if the mud weight is too high, drilling mud can be lost to the porous unit, again causing well control problems.

Predicting the pressure in sandstone before drilling ("predrill stage") may be achieved by looking at data from sand layers in any neighboring wells. However, because the tools used to measure and record sandstone pressures rely on high permeability, understanding pressure in shale, in which the permeability is low, requires another approach.

The standard shale pore-pressure prediction workflow can be crudely summarized by the following steps:

Manuscript received by the Editor 4 November 2016; revised manuscript received 27 September 2017; published ahead of production 02 November 2017; published online 28 December 2017.
[1]Formerly Durham University, Department of Earth Sciences, Durham, UK; presently Durham University, Department of Mathematical Sciences, Durham, UK. E-mail: r.h.oughton@durham.ac.uk.
[2]Durham University, Department of Mathematical Sciences, Durham, UK. E-mail: d.a.wooff@durham.ac.uk.
[3]Durham University, Department of Earth Sciences, Durham, UK. E-mail: r.w.hobbs@durham.ac.uk.
[4]Swarbrick Geopressure Consultancy Ltd., Durham, UK. E-mail: swarbrick@geopressureconsult.com.
[5]Ikon Science Ltd, Venture House, Aykley Heads Business Centre, Durham, UK. E-mail: soconnor@ikonscience.com.

1) Use the bulk density log (RHOB) to estimate the total vertical stress (TVS) or overburden (S_v) — in practice, resistivity, sonic, and seismic velocity data can be used additionally.
2) Use the gamma ray (GR) and/or a combination of neutron porosity and RHOB to understand the lithology and so restrict the intervals for analysis to shales.
3) Generate a shale normal compaction trend (NCT) in terms of one of the logs, usually sonic transit time (ΔT) or resistivity. This involves specifying matrix and sea-floor values for the log.
4) Use a published pore-pressure prediction formula, such as Eaton (1975), Bowers (1995), or the equivalent depth method (Foster and Whalen, 1966) to estimate vertical effective stress (VES) (the difference between the pore pressure and the overburden). This uses only one log at a time and relies on the NCT (or a slightly different curve for Bowers).
5) If pore-pressure measurements are available that are believed to be in equilibrium with the shale, use these to calibrate the prediction, and repeat steps 3–5.

Existing work on uncertainty in pore pressure

The procedure outlined above is deterministic, and as such it does not include a measure of uncertainty. Wessling et al. (2013) develop an algorithm to automate the pore-pressure estimation process in such a way that uncertainties are accounted for. They focus on two parts of the process in which human interaction is most at work: shale discrimination and the estimation of the NCT. For the NCT, they vary the depth interval considered normally compacted, and therefore over which the data are used, and fit an NCT to the data from every possible interval. Each NCT is then used for pore-pressure prediction, creating a suite of pressure predictions that can be used to understand the uncertainty. Although Wessling et al. (2013) do account for uncertainty in the data, and in decisions over the depth at which overpressure begins, modern Bayesian statistical methods argue against automating out such human interaction. Geoscientists will often have knowledge and experience that may not be reflected in the data. Furthermore, there will not always be sufficient log data at normally pressured depths, and in this situation, the method of Wessling et al. (2013) will be unusable.

Malinverno et al. (2004) and Moos et al. (2004) use Monte Carlo (MC) error-propagation methods to assess parameter uncertainty for this workflow. For each input parameter (the seafloor and matrix log values for the NCT, the Eaton exponent, and so on), one must specify a probability distribution to reflect the uncertainty in that parameter. This may be derived from data, or a geoscientist may draw on knowledge and experience to specify values. For example, for the matrix sonic transit time, the user may choose a normal distribution with mean 110 μs/ft and standard deviation (SD) 10 μs/ft. Using these distributions, a large number of random values is then generated for each input, producing many random "settings" for the workflow. The workflow is then implemented at each of these settings, using some data, to generate a set of pore-pressure predictions. The variation in these pore-pressure predictions reflects the parameter uncertainty represented in the probability distributions.

It is crucial to understand that the method we propose here goes far beyond MC error propagation. Our focus is on inference because prediction is our primary aim. In our formulation, an understanding of error propagation is a quite trivial side benefit, as shown in the sensitivity analysis we conduct.

MC error propagation goes some way to understand parameter uncertainty, but it ignores the uncertainty resulting from the workflow itself, which is arguably much more important. The equations used are simple, usually involving only a small number of variables and ignoring many sources of variation. MC error propagation assumes that the scientific model is perfect and that the only source of uncertainty is in the parameters used. MC methods are limited by the fact that they work with the existing method, which has some weaknesses that we will briefly explore.

The pore-pressure system is such that each log may be affected by several properties of the system, and each property (e.g., lithology, pore fluid, and pore pressure) is likely to influence several logs. It seems reasonable to collect all relevant logs and process them together to learn about these properties, rather than to treat them separately.

Throughout the prediction process, geologists are able to draw on their extensive knowledge gained from experience with similar geologic settings or with nearby fields. This is highly valu-

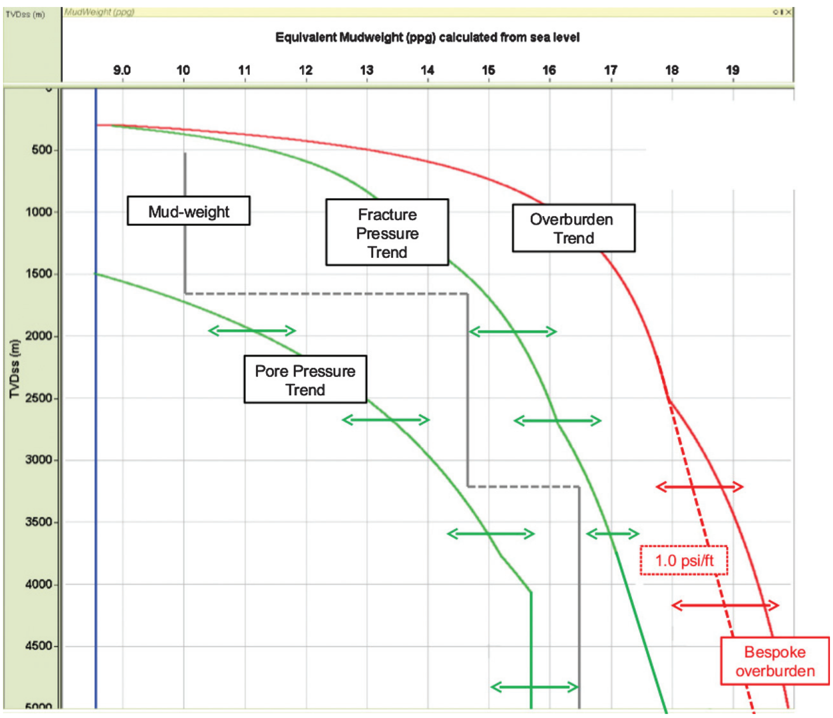

Figure 1. Schematic depth/ppg plot illustrating the key components of a well plan, i.e., pore pressure, fracture pressure and overburden, or TVS. Once these are defined, a mudweight and casing design can be prepared (black lines). The fewer casing strings required, the more quickly/cheaply the well can be drilled. Uncertainty is highlighted on this figure by the double-headed arrows. In red is shown the overburden as generated by using a typically applied 1.0 psi/ft gradient. This may be modified by using high-quality, local density data, as shown by the bespoke overburden. This figure is from the Ikon GeoPressure training manual.

able information, yet there is no structure to include it. Either the geologist adjusts the predictions to better fit their expectations, or their input is ignored as the equations are used without changes. Neither of these approaches will produce the optimum outcome.

Furthermore, the standard workflow as described is not a faithful representation of geologists' understanding of the system. For example, it is commonly understood that the link between porosity and effective stress is key in understanding compaction and that a wireline log is used as a proxy for the porosity. However, formulas such as Eaton's (1975) relation relate the effective stress (and therefore pore pressure) directly to the wireline log only, so that the effect of other data or parts of the system is either ignored or accounted for in an ad hoc fashion. This makes for a model of the system that is less flexible, more difficult to interrogate, and more like a "black box."

To address these issues, we present a Bayesian network for pore-pressure prediction. A Bayesian network allows us to model the system using our choice of scientific relationships, and to include uncertainty from various sources, including those relationships and their parameters. This method improves on an MC analysis, in which the only uncertainty considered is observation error (or uncertainty about parameter values). Unlike MC, which is a method for analyzing an existing model, a Bayesian network is a complete model in itself, built to best represent scientific understanding of the process; it is far from a black box.

Before proceeding, we should note that we are not suggesting that the underlying physics and chemistry of the processes involved are perfectly captured in the equations we have used. This is a matter for geoscientists to debate, not statisticians. However, the incorporation of uncertainty into those equations does provide for some slack in whether the equations represent an agreed underlying reality. The equations used within the statistical model may be updated as further geophysical research provides more insight into the underlying reality, but the basic statistical approach advised here would be unchanged. Indeed, the PP SDBN as presented in this paper is a preliminary model, including only fairly basic scientific relationships. However, the Bayesian principles underpinning it would remain the same as more complexity is added.

BAYESIAN NETWORKS

The theory of Bayesian networks (Pearl, 1988; Cowell et al., 1999; Jensen, 2007) has led to many new applications of uncertainty modeling, in particular to complex problems, in which a large number of factors contribute to overall uncertainty. A clear and detailed explanation of Bayesian networks, with application to a geologic example, is given by Martinelli et al. (2011). For further examples of Bayesian networks in a geoscience context, see Van Wees et al. (2008) and Martinelli et al. (2014).

Bayesian networks derive from Bayesian statistical methodology, which is characterized by providing a formal framework for the combination of data with the judgements of experts, such as reservoir engineers. A Bayesian network is a formal way of factorizing a multidimensional probability distribution over many variables into a product of simpler conditional distributions, which represent dependencies more directly. This results in a mathematically equivalent, but more tractable, representation of the geophysical variables and their interrelationships.

Unlike many Bayesian methods, Bayesian networks are not expressed in terms of prior distributions and likelihood functions; they are used to model systems in which it is impossible or impractical to specify a prior or likelihood over all the parameters. We instead think in terms of smaller collections of parameters. Human expertise is expressed through (1) defining the qualitative structure, i.e., the dependencies between variables; (2) defining how dependent variables behave given the values of other variables influencing them; and (3) describing how nondependent variables behave in the problem at hand. See Zellner (1995) for a fuller comparison of Bayesian and traditional approaches, and Goldstein (2006) on the central importance of role (3) for uncertainty analysis in complex stochastic systems.

In a dynamic Bayesian network (DBN), the same network structure is repeated to represent a system evolving. Usually, this represents the passage of time, with the network repeated for each time step, but for us, it will represent a change in depth down a borehole. We conceal here some technical difficulties in working with DBNs because they apply to all problems rather than just to pore-pressure estimation. These tend to be mathematical (not all probability distributions are easy to work with) and computational (the factorization for large stochastic systems is difficult). As DBNs become larger, computing with them becomes prohibitively expensive if standard methods are used. Therefore, various authors have proposed schemes for working efficiently with large DBNs (e.g., Berzuini et al., 1997; Wilkinson and Yeung, 2002). Our needs are different, and so we develop a new approach, the sequential DBN (SDBN).

In an SDBN, the DBN is treated as a series of separate Bayesian networks, one for each step. When data are entered, the network is updated at the first step to produce posterior distributions. These are used to inform the nodes at the second step, through the links connecting the two steps, and the model is updated at the second step to produce posterior distributions. These are fed to the third step, and so on. This means that the posterior distributions at each step reflect all data up to that point. This sequential updating makes the SDBN particularly appealing in situations in which data are acquired sequentially, for example, in real-time drilling.

A BAYESIAN NETWORK FOR PORE-PRESSURE ESTIMATION

The pore-pressure system involves quantities of several different types. Some we may measure, such as wireline logs or drilling data. Others we cannot observe directly, such as shale pore pressure, effective stress, or porosity. Some have a definite, measureable (at least in principle) physical meaning, whereas others are more conceptual. Although we do not know all their values, we have some understanding of the relationships between them, which we can represent in the structure of the Bayesian network.

The pore pressure SDBN (PP SDBN) works by modeling the system at each depth, with connections between depths to capture the relationships that act vertically. By "the system," we mean the collection of quantities connected to pore pressure and their interactions. Two consecutive depth levels of the PP SDBN are shown in Figure 2. Only the model's more physical nodes are shown in Figure 2, for clarity. The formulas used in the probability distributions are formed from published information on geophysical relationships (Rider, 1996; Hearst et al., 2000; Hantschel and Kauerauf, 2009) and conversations with various individuals. If an expert were to find them inappropriate, they could change the shapes of distributions or the values of model parameters. This might be the case especially if applying the model to a new location, which is known to be different from the "average" settings given here. In this sense, the model is flexible. As with any model, flexibility is open to abuse, with parameters being "fudged" to give the best fit to data, but the emphasis here is on making judgements about properties of the system. Sensitivity analysis, which we will demonstrate in a later section, enables us to discover to which of the input parameters the PP SDBN output is most sensitive, and therefore which values we should put effort into learning about to reduce uncertainty.

Because the pore pressure is included as a node at each depth, it has a probability distribution that will be updated as data are entered. This gives us an estimate of pore pressure with uncertainty that accounts for each part of the model and all the data we have used. The same is true of any node, and so we also produce estimates (with uncertainty) of lithology, porosity, TVS, matrix density, and every other node included in the PP SDBN.

The model

When describing edges connecting one depth to the next, we use superscripts to denote a variable at a specific depth. For example, $S_v^{(z_i)}$ is the TVS at depth z_i.

For an offshore well, the PP SDBN requires an additional input P_{sea} the pressure contributed by the seawater above the borehole that sits outside the repeating part of the network shown in Figure 2. We model this using the known surface elevation depth and seawater density, which we model as normally distributed. The PP SDBN currently assumes an offshore context. If it were to be applied onshore, the model would need to be adapted to deal with the land mass above sea level and also any "near-surface" issues. Given the vertical depth, we can form a probability distribution for the hydrostatic pressure P_{hyd} using the prior distribution that we have specified for the hydrostatic gradient. This will reflect our understanding of variations in fluid density due to salinity, temperature, and any other factors. Where there is little knowledge of the area, it is possible to use global values to create a less informative prior, whereas an expert in the geology of the region should have more accurate knowledge and would therefore choose a more restrictive prior distribution. In either case, the distribution for the

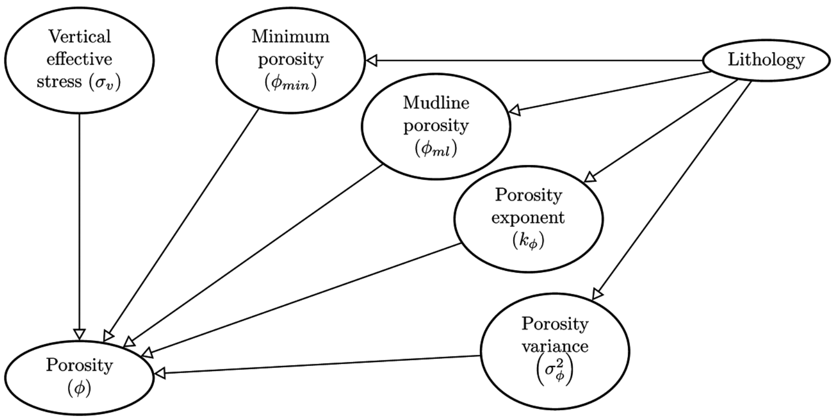

Figure 3. A close-up on the part of the PP SDBN modeling compaction.

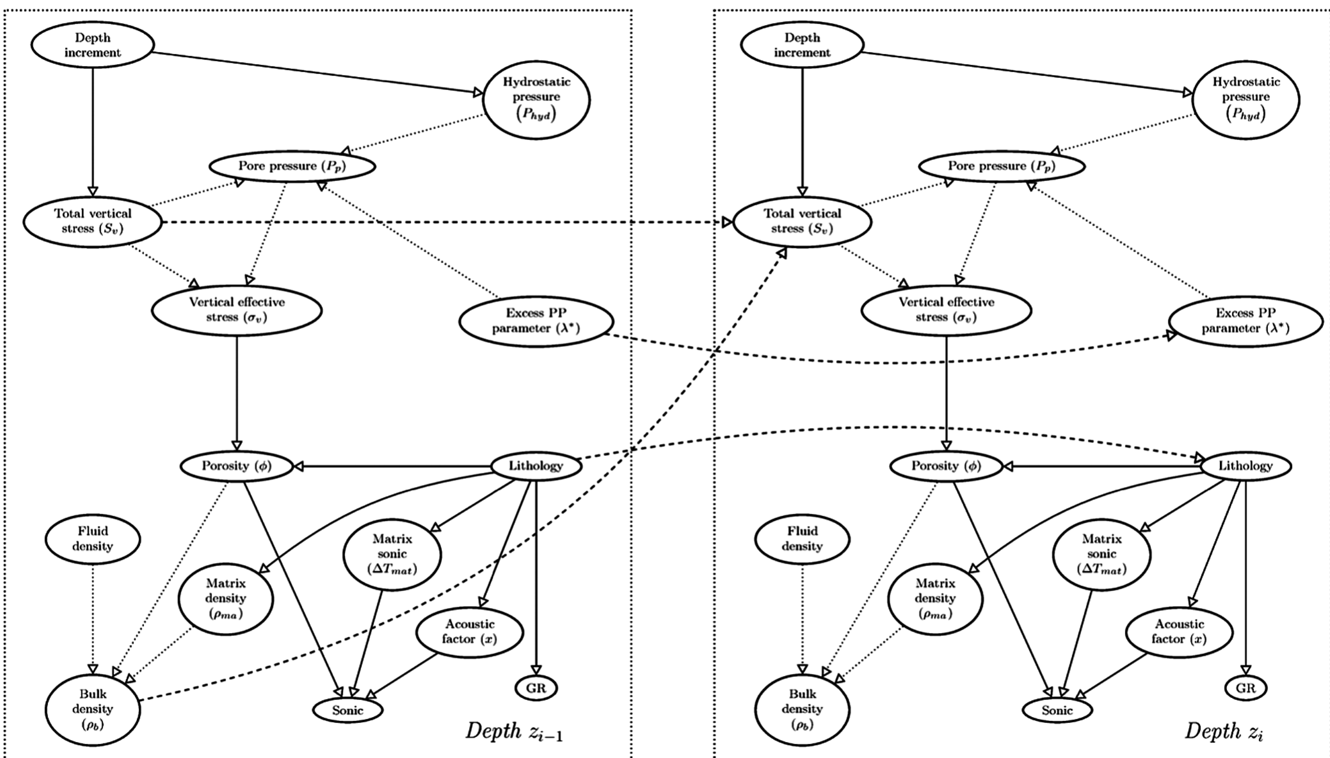

Figure 2. A simplified version of the PP SDBN, showing how the system is modeled at each depth level. The connections between two levels are shown by dashed lines. Deterministic relationships are shown by dotted lines.

hydrostatic gradient is likely to be narrow because water density is well-understood.

At the first depth, the TVS is formed using P_{sea} and a normally distributed bulk density for the rock between the seafloor and the first depth covered by the data. For subsequent depths, we use the depth increment and the TVS and bulk density from the previous depth z_{i-1}. The bulk density data and posterior distribution for TVS at z_i are stored. They are then used to calculate the mean vertical stress at the next depth giving the normal distribution described by

$$S_v^{(z_i)} \sim N(S_v^{(z_{i-1})} + (z_i - z_{i-1}) g \rho_b^{(z_{i-1})}, \sigma^2_{\text{lith}}), \tag{1}$$

where g is the acceleration due to gravity. The variance σ^2_{lith} represents uncertainty in the calculation of S_v even with accurate bulk density data. Using the bulk density log to estimate S_v gives a more accurate estimate than having a prior distribution on the lithostatic gradient, as we are doing with the hydrostatic pressure. If bulk density data are unavailable, then this part of the model still holds, but the bulk density node will pass on a probability distribution rather than a single value. If other wireline-log data are available, then the distribution on the bulk density will be updated to reflect them.

The excess pore-pressure parameter λ^* is a continuous value between zero and one, defined by

$$\lambda^* = \frac{P_p - P_{\text{hyd}}}{S_v - P_{\text{hyd}}}, \tag{2}$$

as in Shi and Wang (1988), with P_p, P_{hyd}, and S_v, as in Figure 2. We model λ^* using a beta distribution. This is a standard way of handling a variable taking values in the interval $[0, 1]$, allowing a variety of shapes. If $\lambda^* = 0$, then there is no overpressure. If, hypothetically, $\lambda^* = 1$, then the pore pressure is the same as the TVS. In the PP SDBN, there is an edge from λ^* at one depth to λ^* at the next, indicating that the prior distribution for one layer comes from the posterior distribution for the previous layer. We also inflate slightly the prior variance for λ^* for the next layer to avoid λ^* converging to a single point, or expressing overconfidence. Because of this dynamic link, the excess pore-pressure parameter is expected to remain the same from one depth to the next, this equates to a slight increase in pore pressure. Although small changes in λ^* are favored in the conditional distributions we choose, we ensure that more dramatic jumps are still possible.

The nodes S_v, P_{hyd}, λ^*, and pore pressure P_p are linked deterministically, through the equation

$$P_p = P_{\text{hyd}} + \lambda^*(S_v - P_{\text{hyd}}). \tag{3}$$

When data are entered into the model, these nodes' distributions will be constrained by information coming from the depth, which mostly constrains S_v and P_{hyd}, and by VES, which will have been constrained by porosity through information from the wireline logs. The posterior distribution of λ^* from the previous depth will influence the current λ^*, and this too will influence the pore-pressure posterior distribution.

The link between porosity ϕ and VES is the most important part of the model. Figure 2 shows lithology and VES as parents of porosity; however, this is a simplification. This part of the model is shown in more detail in Figure 3. The conditional distribution for porosity is

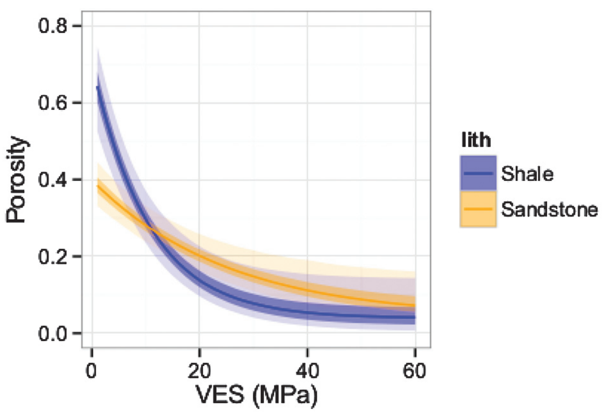

Figure 4. The compaction curves in the model. The mean (in a solid line) and central 50% and 95% intervals (shown by shading) are given to show the spread of the probability distribution of porosity for each value of VES and lithology.

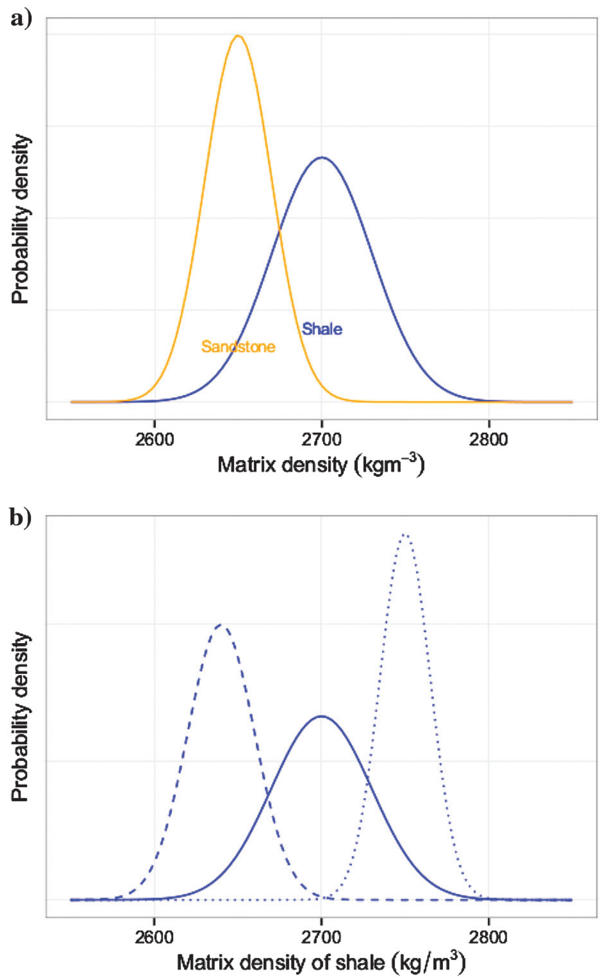

Figure 5. Examples of possible prior distributions for matrix density (ρ_{ma}). (a) Current probability distribution for matrix density (ρ_{ma}) for sandstone and shale. (b) Possible alternative distributions for matrix density for shale. For example, there are lower values for a smectite-rich formation and higher values for an illite-rich shale.

$$\phi|\sigma_v, \phi_{\min}, \phi_{\mathrm{ml}}, k_\phi, \sigma_\phi^2 \sim N(\phi_{\min} + (\phi_{\mathrm{ml}} - \phi_{\min}) \times \exp[-10^{-6} k_\phi \sigma_v], \sigma_\phi^2), \tag{4}$$

based on equations found in Hantschel and Kauerauf (2009). The parameters $\phi_{\min}$, ϕ_{ml}, k_ϕ, and σ_ϕ^2 each depend on the lithology. Again, the actual form of the relationship between porosity, VES, and lithology can be changed as required.

By using this model, the PP SDBN reflects the fact that different lithologies compact differently. For any lithology and VES value, there is a probability distribution for porosity as shown in Figure 4.

A key feature of the distribution shown in Figure 4 is that porosity is more uncertain in sandstones than in shales because shale compaction is generally better understood. This greater uncertainty feeds through the model, and so where the posterior distribution for lithology favors shale, the posterior distribution for pore pressure will be narrower than in what the model estimates to be sandstones.

The logic of the model is similar to that of the equivalent depth method for estimating pore pressure; it is assumed that under the same lithologic conditions, a particular value of VES will lead to a particular value of porosity. Because we know the depth, and therefore have an estimate for S_v, we can use this to estimate pore pressure. The PP SDBN presented here is therefore based on mechanical compaction.

The lithology posterior distribution will take the form of probabilities of sandstone and shale. For example, in the posterior distribution samples for depth z_{i-1}, 10% might be sandstone with the remaining 90% being shale. Therefore, in the posterior distribution at depth z_{i-1}, the probability of the lithology being shale is $p_{\mathrm{sh}}^{(z_{i-1})} = 0.9$. The sequential model includes a lithology transition matrix

$$\begin{pmatrix} p_{\mathrm{sst|sst}} & p_{\mathrm{sst|sh}} \\ p_{\mathrm{sh|sst}} & p_{\mathrm{sh|sh}} \end{pmatrix}, \tag{5}$$

which gives the probability of each lithology at depth z_i given the lithology at depth z_{i-1}, and this, together with the posterior samples from z_{i-1}, is used to generate $p_{\mathrm{sh}}^{(z_i)}$ and $p_{\mathrm{sst}}^{(z_i)}$.

The GR count is strongly influenced by lithology, so in the PP SDBN, the GR variable is represented as a child of lithology. We must therefore define a conditional probability distribution for GR for each kind of lithology considered, as we expect the GR log to behave differently for different lithologies. In the PP SDBN, we use the GR index (I_{GR}), so that this variable is standardized to between zero and one. In practice, GR is observed via wireline-log data. Hence, using the Bayes theorem, we can make inferences about unobserved lithology and any other nodes connected to lithology such as porosity, from the observed GR wireline log.

Bulk density ρ_b is another node in Figure 2 that can often be constrained by observed data. The key equation in understanding its surrounding links is

$$\rho_b = \phi\rho_{\mathrm{fl}} + (1 - \phi)\rho_{\mathrm{ma}}, \tag{6}$$

where ρ_{fl} and ρ_{ma} are the fluid and matrix density, respectively, and ϕ is the porosity. The matrix density depends on the lithology (specifically on the dominant mineral composition), and the fluid density on the pore fluid type. Figure 5a shows the default probability distribution for matrix density for sandstone and shale. One could argue that there is too much overlap between the two; however, this ensures that the model does not "get stuck" in a particular lithology. Figure 5b shows some examples of alternative distributions for shale-matrix density. At present, the pore-fluid type node has no parents, and it is assumed that the rock is predominantly water filled.

We base the conditional distributions for sonic transit time (ΔT) on the equation

$$\Delta T = \frac{\Delta T_{\mathrm{mat}}}{(1 - \phi)^x}, \tag{7}$$

where ΔT_{mat} is the matrix sonic transit time and x is an acoustic formation factor. This relationship is presented by Raymer et al. (1980) for sandstones and by Issler (1992) for shales. The distributions for ΔT_{mat} and x depend on lithology, and ΔT is then normally distributed with equation 7 used as the mean. Equation 7 was developed from deep-borehole data that do not include shallow depths, typically less than 500 m below sea bed, and our model has not been applied to shallow depths on account of the absence of data in the example wells. One could instead use Wyllie et al.'s. (1956) time-average equation, in which case, a fluid sonic transit time node would be introduced and the acoustic formation factor x removed. However, Raymer et al. (1980) and Issler (1992) propose the form in equation 7 as an improvement, stating that it better captures the curvilinear relationship between porosity and ΔT, and it is less prone to producing unrealistic porosity values, or requiring extensive tuning.

The PP SDBN was implemented in R Development Core Team (2011), with links to just another Gibbs sampler' (JAGS). JAGS is Gibbs sampling software, which we use to evaluate the posterior distributions of the nodes in the Bayesian network. Rather than find the posterior distributions analytically, the Gibbs sampler generates samples of values from each posterior distribution, which can then be used to understand the distribution. This is a standard way of approaching Bayesian networks (Bernado and Smith, 1994). A simplified example of how the Gibbs sampler works is provided in the Supplementary Materials (supplementary information can be accessed through the following link: s1.pdf).

Advantages of this method

Unlike traditional methods, the PP SDBN models the interactions between different quantities in the system. For example, we learn about the lithology using the GR and sonic and bulk density logs simultaneously, and the posterior probability distribution for porosity reflects the bulk density and sonic. Therefore, the uncertainty reflects the extent to which different sources of information agree with one another. Because the PP SDBN will learn from whatever set of information it is given, it is not dependent on any particular set of log data being available. If part of a log is missing for some depth range, that node's conditional distribution will be used to learn about its behavior in light of all available data. Therefore, this method is flexible and robust, not requiring a particular log or combination of data to be available at all depths, unlike MC error propagation, which is not robust to missing data.

Because the PP SDBN is a full probabilistic model of the system, we learn about not just pore pressure, but all the unobserved nodes through their posterior distributions. This allows us to more fully assess our model and also to learn more about the system. This is partially the case when using MC with standard pore-pressure estimation; for example, we will generate a sample of lithologies having perturbed the shale cut-off value, or we may have a sample of TVS values by perturbing parameters relating to the estimation.

In contrast though, these samples will reflect only the small set of data types involved in that part of the process, whereas when using a Bayesian network, the posterior distribution reflects all the data that have been used for the model. This reveals the fundamental difference between MC error propagation and the PP SDBN. The former can only assess uncertainty in an existing model, whereas the latter incorporates a fully joint model of the entire system.

There is an equivalence in the results produced by the SDBN approach and MC error propagation, in the sense that if we applied an idealized error-propagation approach to our model, the results of the error propagation would be exactly the same as the prediction uncertainties produced by our model. If we regard the PP SDBN as a gold-standard approach combining the best available synthesis of data and human expertise, we could in principle examine any discrepancies between it and uncertainties produced by MC error propagation applied to other modeling approaches. However, this would need substantial effort to match input choices and would in any case only allow us to conclude that different models can produce different answers.

In the SDBN, expert knowledge and data are combined in a rigorous way using the Bayes theorem. As with any other method developed for pore-pressure prediction, the quality of the results depends on the quality of judgments about model relationships and so forth, but with the advantage for the SDBN approach that we formally quantify the experts' uncertainty in the model via probability distributions. The experts' uncertainty is therefore reflected in the final pore-pressure estimate. As pointed out earlier, the structure of the model is also subjective, and so it should be designed with care (Plummer, 2014; Su and Yajima, 2014). Note however that the traditional workflow and the corresponding "standard" methods chosen are themselves highly subjective, but they are not handled within the rigorous formal statistical framework of a Bayesian network.

EXAMPLES

West Africa 1

Figure 6 shows data from a well in West Africa, with predictions made by the PP SDBN. This interval was chosen because it appeared to be predominantly shale, but it contains several sandstone intervals in which the pore pressure was measured. Of the six sandstone pressure measurements, the shallowest and the deepest three are judged by experts to be in isolated sandstones, and therefore our shale pore-pressure estimate should match them. Because they are taken from sandstones, we can use their depths to assess the performance of the lithology estimation, and indeed they are all matched by regions of sandstone in the lithology plot. The two pressure measurements at approximately 3350 m are thought to be in a thicker and more laterally extensive sandstone that is slightly drained, and the lithology posterior distribution indeed suggests a thicker layer of sandstone here. The estimated pore-pressure trajectory above this is in agreement with our experts' expectations. The

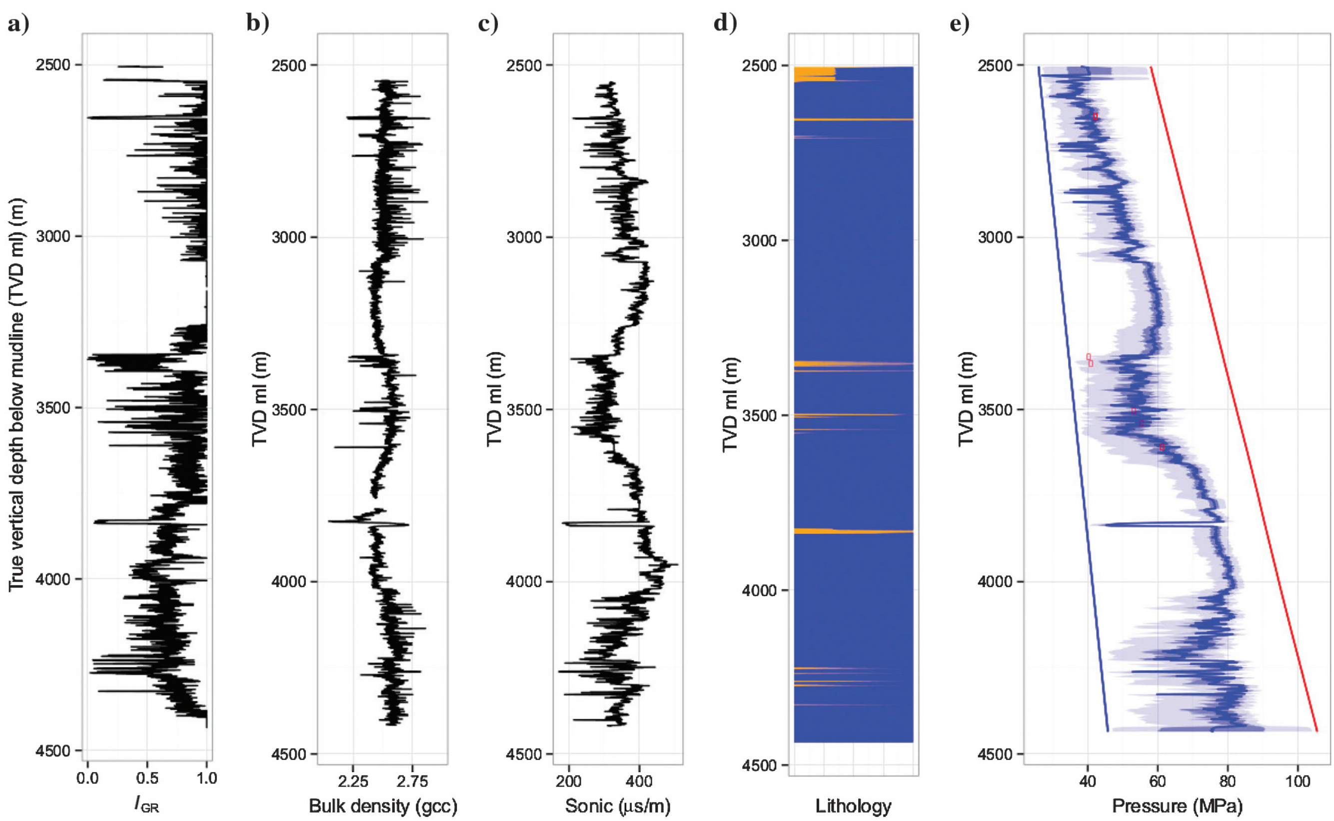

Figure 6. West Africa 1: (a) GR index data; (b) bulk density data; (c) sonic transit time data; (d) lithology estimated by SDBN (blue is shale and orange is sandstone); and (e) pore pressure estimated by SDBN, with mean and central 50% and 95% intervals shown by shading. Mean and central 95% are shown for TVS and hydrostatic pressure, although there is little uncertainty in these compared with the pore pressure, and so they appear as lines. Sandstone pore-pressure data are shown by the red dots in the pressure plot. Note that these pressure observations are not used as input data to the PP SDBN.

sharp spikes approximately 3800 m coincide with a casing point, which has been identified as a sandstone by the SDBN because this knowledge is outside its scope. However, it has not influenced the nearby results, showing that the SDBN is robust to unexpected results.

Figure 7 shows the samples from the posterior distributions of pore pressure and porosity at three particular depths. Such plots can be made for any unobserved node, at any depth, and so they can be useful for developing a greater understanding of uncertainty.

Figure 8 compares the posterior pore-pressure distribution for West Africa 1 for different combinations of input data. This demonstrates the reduction in uncertainty that can come with including additional data. Without the control from the gamma log, the lithology is poorly constrained, leading to a significantly worse pressure estimation.

SENSITIVITY ANALYSIS

The PP SDBN is a statistical model constructed from expressed geophysical relationships and tuning parameters. The default values we supply can be used for the tuning parameters, or a reservoir engineer may supply more carefully considered inputs, depending on their expertise and local knowledge. The PP SDBN lends itself well to sensitivity analysis techniques. The aim of sensitivity analysis is to discover how variation in the output can be explained by variation in the collection of inputs. Variation in inputs can be attributed to several sources. A physical quantity may be subject to measurement error, or there may be a high level of uncertainty about a particular parameter owing to a lack of information or understanding of the system. Sensitivity analysis reveals how this uncertainty propagates to the output, and it therefore indicates the degree of confidence we can have in the model's results. For example, if a model's output is highly sensitive to a physical parameter about which little is known, there is consequential uncertainty surrounding the model output.

Sensitivity analysis allows us to more deeply examine how the model is working, and whether it resembles the real system in the way that we expect. For example, system experts are likely to expect some parameters to be among the most influential. If the sensitivity analysis shows them to be insignificant in the model, this suggests that the model is not representing the system as intended. Learning which are the most crucial parameters for tuning the model can help to improve predictions. If current estimates for the values of these parameters are not sufficiently precise to give confidence in the output values, further research should be conducted into these parameters. Equally, the model can also be simplified by eliminating variables to which the model is not at all sensitive. Saltelli et al. (2000) give a thorough account of the theory and techniques of sensitivity analysis.

Preliminary examples

Here, we demonstrate some simple preliminary techniques, before going on to demonstrate a more comprehensive method. To gain some insight into how influential a parameter is, one can hold all others fixed at their default value, then vary the parameter in question. In the following examples, we do this for West Africa 1, and with each parameter being varied between three values: low, default, and high. The specific values for each parameter were formed by surveying geologist colleagues. This was not a thorough elicitation, but a

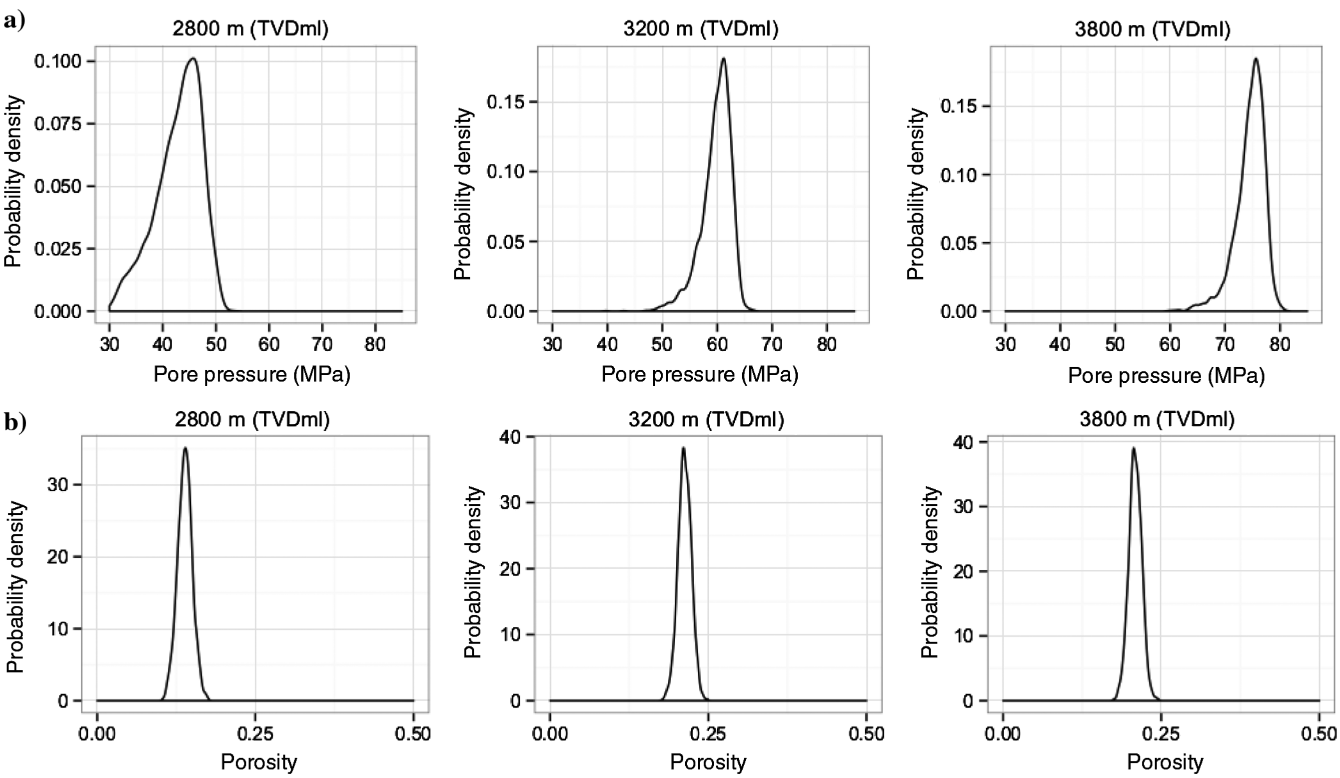

Figure 7. (a) Pore-pressure posterior distributions and (b) porosity posterior distributions.

casual experiment. Nevertheless, the ranges provided should give reasonable results.

Figure 9 compares posterior probability distributions for pore pressure when three different input parameters are varied. This could be done for any other unobserved node, for example, porosity or TVS. The leftmost plot, in which an input (the first scale parameter of a beta distribution) relating to sea-floor porosity in shales has been varied, shows that this parameter has little effect on pore pressure at this depth for this well. The three posterior distributions are very similar. The middle plot, in which the mean matrix sonic transit time of shale (in sm^{-1}) is varied, shows a stronger influence.

The three posterior distributions are clearly separated, and the means differ by approximately 8 MPa. Therefore, for better understanding, the matrix sonic transit time in shale would increase our confidence in pore pressure. The third plot, in which the SD of the matrix sonic transit time is varied, shows different behaviors still. It appears that for some value of matrix sonic SD between the low and default values, there is a discrete change in model behavior. Studying the posterior distributions shows that this relates to lithology; the posterior samples for the default and high values contain much more sandstone than those from the low value, and this difference has manifested itself in the pore-pressure posterior distributions.

Figure 10 summarizes the results of performing this analysis on several input parameters, by plotting the means of the posterior distributions for each parameter that has been varied, in order of range.

Morris screening design

To gain insight into which are the more influential parameters, we will use a one-at-a-time screening design proposed by Morris (1991). Once complete, we have several elementary effects values for each input parameter. Each one can be thought of as an estimate of the effect of changing that input from its minimum to its maximum, with every other input held the same, somewhat like a partial derivative of the output with respect to that input parameter. Therefore, a large (negative or positive) elementary effect suggests an influential parameter. One close to zero suggests a more negligible parameter. In summary, the following observations can be made:

- If the elementary effects for input i have a mean close to zero and a low variance, input i appears to have little effect.
- If the elementary effects for input i have a high (in magnitude) mean and a low variance, input i appears to have a strong linear effect.

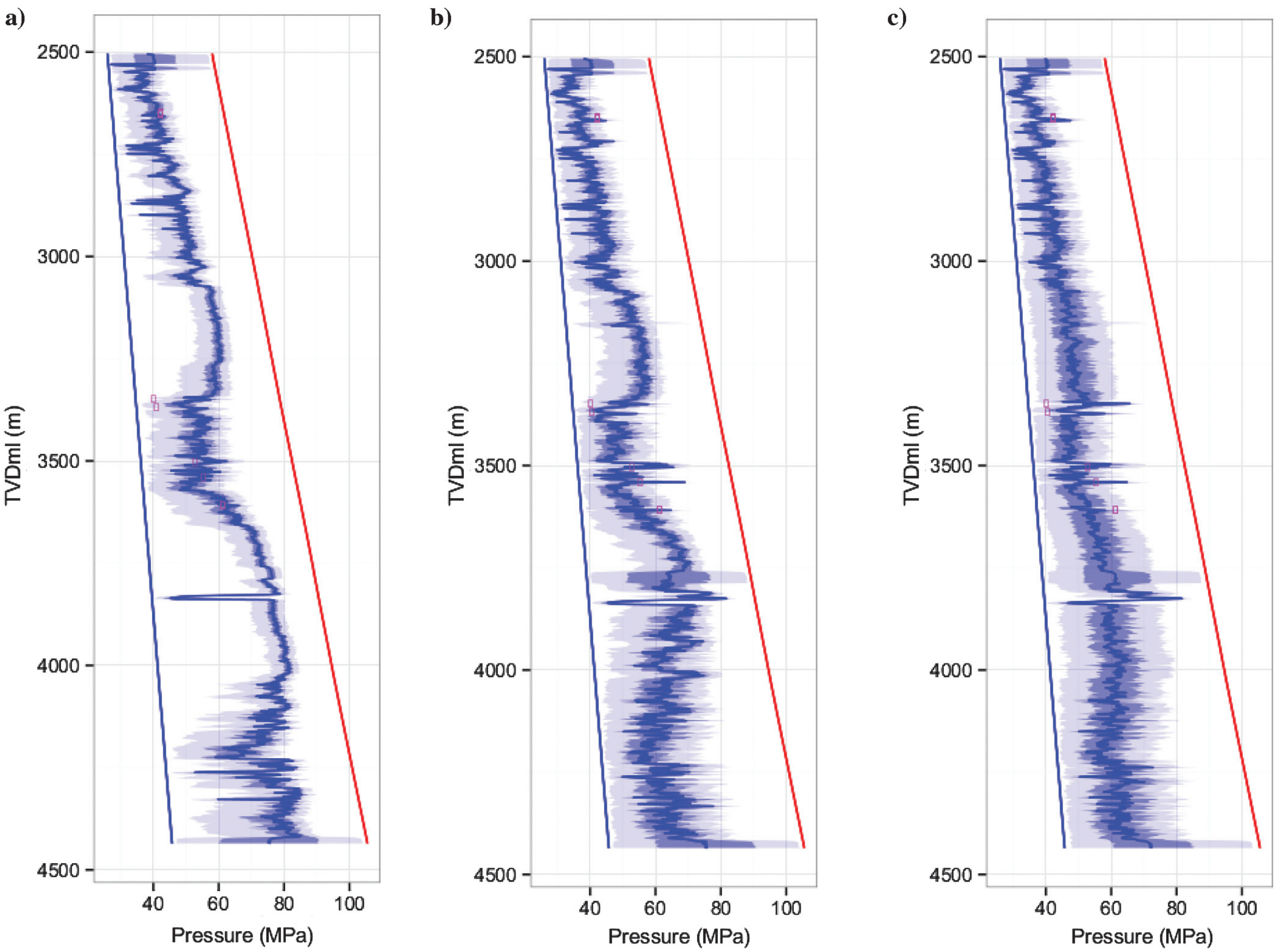

Figure 8. Panels (a) I_{GR}, RHOB, and DT, (b) I_{GR} and RHOB only, and (c) RHOB only.

- If the elementary effects for input i have a high variance, input i appears to be involved in interactions with other inputs or to have a nonlinear effect.

Example: West Africa 1

In the PP SDBN, there are 38 input parameters, each of which we chose initially to vary. We choose five values for each input parameter, and run the model 1000 times in total. We stipulated that we must produce at least 20 elementary effects for each input parameter.

Figure 11 shows the elementary effects for two depths in West Africa 1. The first depth (2670 m) is near the top of the interval in which the experiment was run. Above this point, the borehole is almost entirely shale, and this is true in almost all of the input space. By 2770 m (the depth of the second plot), there has been some sandstone, with the PP SDBN estimating more sandstone at some input value settings than at others. This could account for the higher variability elementary effects at the deeper point.

The most influential parameters remain the same in each plot, with "dtma_mean_sh" (mean matrix sonic transit time in shale) having a negative effect on the mean pore pressure and "porsd_sh_fac" ($\sigma_{\phi}^{\mathrm{fac}}$) having positive effects. The distributions of these elementary effects are similar in each plot. "b_porml_sh" and "b_pormin_sh" (these are shape parameters for the mudline porosity and minimum porosity, respectively, in shales) have a slight positive effect in each case. Otherwise, the elementary effects are centered around zero, some with little spread. Therefore, our primary focus would be on the four parameters already mentioned, and it may well be possible to eliminate some of the consistently negligible input parameters without degrading the result of the pore-pressure prediction.

DISCUSSION

The problem we have addressed in this paper is that of quantifying uncertainty in pore-pressure predictions in a meaningful way. We have approached this problem from an entirely fresh perspective, based on a rigorous formal statistical method and present a proof-of-concept model that is highly adaptable. The appropriate mathematical machinery is the Bayesian network, which allows us to express causal dependencies between the geophysical elements that make up our understanding of the relationships between pressures, lithology, porosity, wireline logs, and so forth. This kind of approach is open and transparent, with all the ingredients (structure, experiential judgements, and data) having a clear role and implication. The Bayesian network allows us to collect together expert knowledge, uncertainty, and data into a rigorous and coherent model, so that the resulting pore-pressure probability distribution makes sense of these.

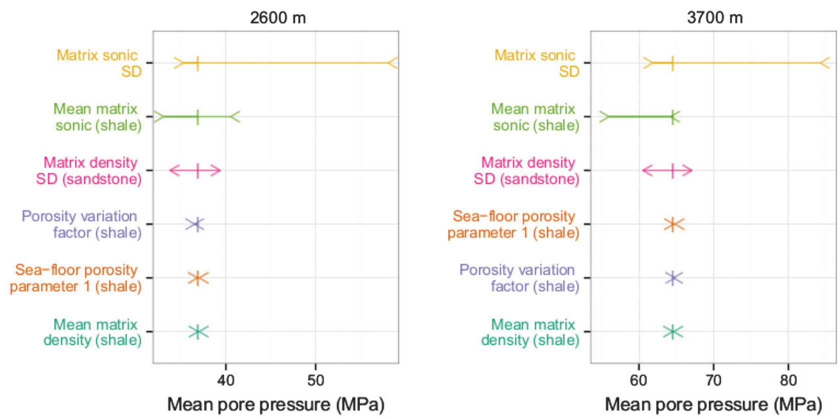

Figure 10. Tornado plots Howard (1988) for some SDBN input parameters, at two depths in West Africa 1. The left pointing arrows mark the mean posterior pore pressure for the lower of the three input values, the right pointing arrows for the upper value, and the vertical bar for the default. Outward-pointing arrows (e.g., as with matrix density SD) imply a positive correlation between that input and the mean pore pressure. Inward-pointing arrows (as with matrix sonic SD) imply a negative effect.

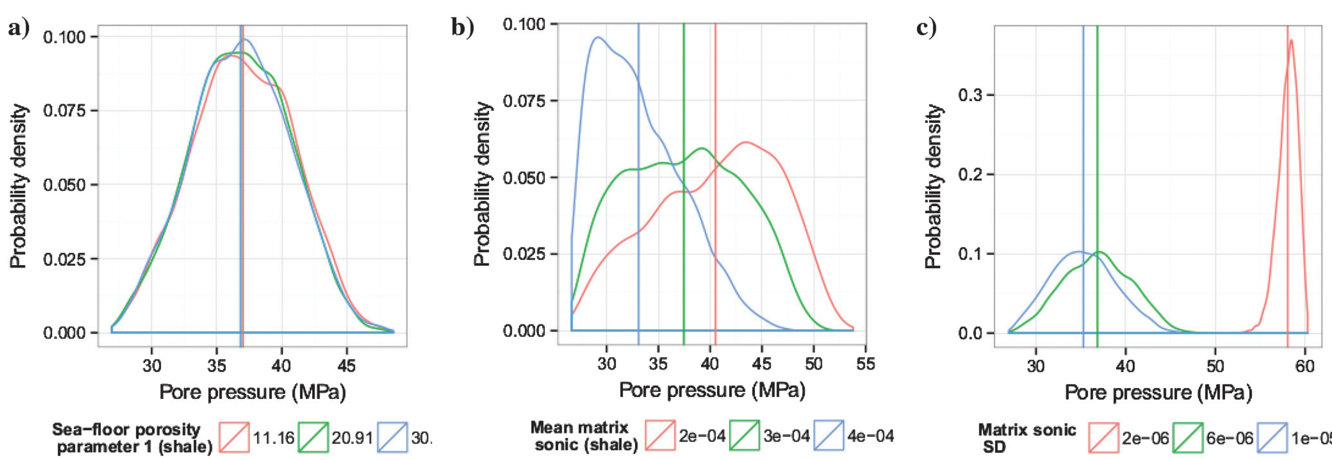

Figure 9. Posterior probability distributions for pore pressure at 2600 m. In each plot, all but one of the input parameters have been held at their default values, and one has been varied, as shown in the legend beneath each plot. The mean for each posterior distribution is also shown by a vertical line in the corresponding color. (a) Sea-floor porosity parameter 1 for shale. This is the first shape parameter of a beta distribution. (b) Mean matrix sonic for shale in s/m. (c) SD of matrix sonic in s/m.

Through sensitivity analysis, we can understand which of the input parameters are the most influential. This can lead to an increased focus in these areas and therefore to a reduction in uncertainty in pore pressure because they are better understood. The PP SDBN is a flexible-core framework that can be extended in many ways to represent the pore-pressure system and to be useful practically in the process of planning and drilling a well.

From a geologic perspective, log and seismic data are inherently less reliable for pressure understanding than direct pressure measurements; thus, any technique that can be developed that helps the geologist to visualize, understand, and therefore reduce the uncertainty in these data types is highly valuable and will result in more accurate pressure prediction. Moreover, if the same approach can also define and quantify prior understanding of how a system behaves and express uncertainty about this understanding, then the final pressure profile will be much more robust.

Geologic basins are complex environments, where multiple factors affect a simple variable such as porosity. Porosity (or often a proxy variable such as sonic transit time or bulk density) is used directly to relate to pore pressure and yet many co-dependent factors influence its value. The PP SDBN allows us to jointly model these factors, their codependence, and our uncertainty. This provides a more holistic way to approach, in this example, porosity. The effects of data gaps or missing logs can be quickly assessed in terms of our ability to define an accurate porosity and subsequently, pressure. Expert knowledge and data are combined in this approach so that it is geologically based. Uncertainty in the expert judgments is captured in the conditional probability distributions, and this is reflected in the posterior probability distributions attached to the pressure estimates.

The PP SDBN as presented in this paper is a preliminary proof-of-concept model, involving a limited selection of data types and assuming disequilibrium compaction as the pressure-generating mechanism. However, the Bayesian network structure lends itself to augmentation. To develop the PP SDBN, more data types would need to be incorporated. This includes additional log measurements, such as resistivity and neutron density. The caliper log could also be introduced to inform the uncertainty based on borehole quality. By learning from the equivalent circulating and static densities, along with any connection gas or kicks, we could put logical constraints on the pore pressure.

An extension to the PP SDBN proposed here for predrill pressure prediction would be the incorporation of velocity information from surface seismic reflection data. As part of the seismic processing workflow, several velocity models can be derived ranging from simple NMO to tomographic inversion and more recently to full-wave inversion. The choice of which to use depends on the complexity of the problem (Cibin et al., 2004). Velocity models and their relationship to pore pressure can be further refined by calibration to offset wells, if available (Den Boer et al., 2006; Sayers et al., 2006). To ensure consistency with the SDBN's modeling of uncertainty, the chosen velocity model should also include an estimate of its own

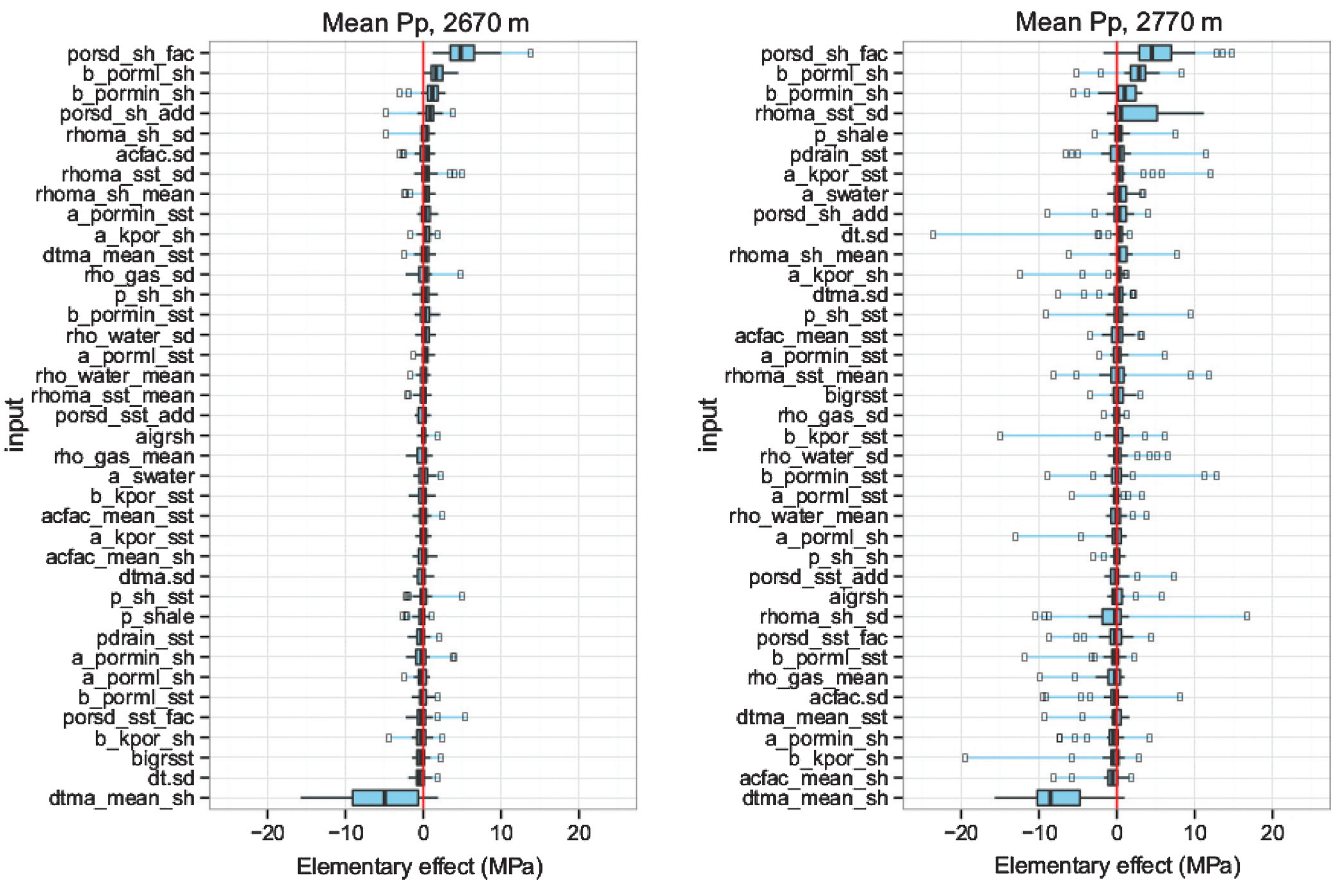

Figure 11. Elementary effects for mean pore pressure at two depths in West Africa 1, ordered by median elementary effect. Points falling outside the interquartile range (IQR) by more than $1.5 \times \text{IQR}$ will be plotted as outliers.

uncertainty, ideally computed in a compatible Bayesian manner, as for example by Caiado et al. (2012). The uncertainty in the velocity model would then be reflected in the predrill pore-pressure distribution. Incorporating this surface-derived velocity information into the PP SDBN would be effected by addition of a node as an alternative for or to complement the wireline-log nodes.

A desirable development would be to extend the SDBN to three dimensions to create a full-3D probabilistic pore-pressure estimate, as for example in Doyen et al. (2003). There are two key challenges here. First, the SDBN is a computationally intensive method and extending to 3D would multiply this problem. Second, extending to 3D would require us to carefully think about horizontal correlation and how to capture features such as lateral transfer and drainage, extensional and compressional stress, and anisotropy. It is worth noting that Doyen et al. (2003) avoid this issue by deriving their probability distributions empirically from well data: Their example uses data from 21 nearby wells. The first issue may be overcome by the use of a large multicore computer, or by the use of a velocity model such as that from Caiado et al. (2012), which could be used to reduce the size of the data set. The second requires more thought. It is difficult to validate the approach of Doyen et al. (2003) inasmuch as linking the empirically derived parameters to physical realities. Our preferred approach would be to model these parameters; in which case, the physical model, the data and judgements used to populate it, and the inferences and predictions drawn from it, are transparent and open to scrutiny.

CONCLUSION

In this paper, we have presented the PP SDBN, a novel, statistically rigorous framework for quantifying uncertainty in pore-pressure estimation. The Bayesian network we have developed allows the geologist to capture their scientific understanding of the pore-pressure system, in order for this to be updated in light of any available data. The PP SDBN as we have presented it is currently applicable to those basins in which mechanical compaction is the generator of pore pressure. The flexible nature of the network means that adapting it to account for more data types in the future (e.g., predrill seismic velocity models or real-time data) or a more complicated scientific model (e.g., including chemical compaction or nonvertical stresses) is feasible.

The PP SDBN is an improvement on methods such as MC; the pore-pressure uncertainty will not necessarily be smaller, but it has a clear meaning, having arisen from a careful specification of the expert's understanding of the system. The uncertainty reflects the data and expert knowledge in a way that is not possible with MC because the PP SDBN is a fully probabilistic model of the system. The accuracy of the PP SDBN's pore-pressure prediction (the mean of the posterior distribution) will depend on the geophysical relationships used, but because posterior distributions (and hence predictions) are produced for all nodes, the model can be interrogated and understood in terms of how it models each part of the system, rather than the pore pressure alone.

Previous approaches to quantifying uncertainty tend to be ad hoc; industry-standard relationships such as Eaton or the equivalent depth method do not easily allow all sources of uncertainty to be easily represented. Pressure is calculated on an increasing depth basis, ignoring the codependency of many of the variables in these algorithms. Our method differs in that it offers a coherent structure for containing the data, geologic knowledge, and physical understanding available, with assessments of uncertainty on each of these elements. The conditional probability distributions are specified to best represent our understanding of how the quantities in the system interact. Because of this, the method is transparent in that uncertainty in the posterior pore-pressure distribution can be understood in terms of uncertainty in the input parameters and scientific relationships used in the PP SDBN. It is therefore a more effective tool for capturing and displaying uncertainty and for indicating deficiencies in understanding, as we have shown through sensitivity analysis. As with any decision-support tool, the quality of the prediction depends on the quality of the model, but also overtly here on the quality of the human expertise supplied to it.

Pore-pressure prediction is inherently uncertain, especially in shale lithologies where the low permeability precludes the use of direct pressure tests. Many assumptions have to be made, and it is typically problematic to test which of these assumptions are the most reliable, and which particular parameter holds the most weight. The PP SDBN allows careful and rigorous analysis of these factors, resulting in a clearer understanding of the geologic system in terms of its influence on the pressure regime. This leads to more accurate pressure prediction and ultimately to the more cost-effective and safe drilling of future wells.

ACKNOWLEDGMENTS

Our thanks go to the sponsors of the GeoPOP3 project, BG, BP, Chevron, ConocoPhillips, DONG Energy, E.ON, Eni, Petrobras, Petronas, Statoil, Total and Tullow Oil, for financial support, and also to GeoPOP3 colleagues for their advice and support.

REFERENCES

Bernado, J. M., and A. F. Smith, 1994, Bayesian theory: Wiley.

Berzuini, C., N. G. Best, W. R. Gilks, and C. Larizza, 1997, Dynamic conditional independence models and Markov chain Monte Carlo methods: Journal of the American Statistical Association, **92**, 1403–1412, doi: 10.1080/01621459.1997.10473661.

Bowers, G., 1995, Pore pressure estimation from velocity data: Accounting for overpressure mechanisms besides undercompaction: SPE Drilling & Completion, **10**, 89–95, doi: 10.2118/27488-PA.

Caiado, C. C., R. W. Hobbs, and M. Goldstein, 2012, Bayesian strategies to assess uncertainty in velocity models: Bayesian Analysis, **7**, 211–234, doi: 10.1214/12-BA707.

Cibin, P., M. D. Martera, M. Buia, D. Calcagni, D. J. Runcer, and T. Talkan, 2004, What seismic velocity field for pore pressure prediction?: 74th Annual International Meeting, SEG, Expanded Abstracts, 1531–1534.

Cowell, R., A. Dawid, S. Lauritzen, and D. Spiegelhalter, 1999, Probabilistic networks and expert systems: Springer.

Den Boer, L. D., C. M. Sayers, Z. R. Nagy, P. J. Hooyman, and M. J. Woodward, 2006, Pore pressure prediction using well-conditioned seismic velocities: First Break, **24**, 43–49, doi: 10.3997/1365-2397.2006012.

Doyen, P., A. Malinverno, R. Prioul, P. Hooyman, S. Noeth, L. den Boer, D. Psaila, C. Sayers, T. Smit, C. van Eden, and R. Wervelman, 2003, Seismic pore pressure prediction with uncertainty using a probabilistic mechanical earth model: 73rd Annual International Meeting, SEG, Expanded Abstracts, 1366–1369.

Eaton, B. A., 1975, The equation for geopressure prediction from well logs: Presented at the Fall Meeting of the SPE of AIME.

Foster, J., and H. Whalen, 1966, Estimation of formation pressures from electrical surveys: Offshore Louisiana: Electrical Logging, **18**, 165–171, doi: 10.2118/1200-PA.

Goldstein, M., 2006, Subjective Bayesian analysis: Principles and practice (with discussion): Bayesian Analysis, **1**, 403–420, doi: 10.1214/06-BA116.

Hantschel, T., and A. I. Kauerauf, 2009, Fundamentals of basin and petroleum systems modeling: Springer.

Hearst, J. R., P. H. Nelson, and F. L. Paillett, 2000, Well logging for physical properties: Wiley.

Howard, R. A., 1988, Decision analysis: Practice and promise: Management Science, **34**, 679–695, doi: 10.1287/mnsc.34.6.679.

Issler, D., 1992, A new approach to shale compaction and stratigraphic restoration, Beaufort-Mackenzie Basin and Mackenzie Corridor, Northern Canada: AAPG Bulletin, **76**, 1170–1189.
Jensen, F., 2007, Bayesian networks and decision graphs: Springer.
Malinverno, A., C. M. Sayers, M. J. Woodward, and R. C. Bartman, 2004, Integrating diverse measurements to predict pore pressure with uncertainties while drilling: SPE Annual Technical Conference and Exhibition, SPE, Extended Abstracts, doi: 10.2118/90001-MS.
Martinelli, G., J. Eidsvik, R. Hauge, and M. D. Forland, 2011, Bayesian networks for prospect analysis in the North Sea: AAPG Bulletin, **95**, 1423–1442, doi: 10.1306/01031110110.
Martinelli, G., J. Eidsvik, K. Hokstad, and R. Hauge, 2014, Strategies for petroleum exploration on the basis of Bayesian networks: A case study: SPE Journal, **19**, 564–575, doi: 10.2118/159722-PA.
Moos, D., P. Peska, C. Ward, and A. Brehm, 2004, Quantitative risk assessment applied to pre-drill pore pressure, sealing potential, and mud window predictions from seismic data: Gulf Rocks 2004, The 6th North America Rock Mechanics Symposium (NARMS), American Rock Mechanics Association, ARMA-04-499.
Morris, M. D., 1991, Factorial sampling plans for preliminary computational experiments: Technometrics, **33**, 161–174, doi: 10.1080/00401706.1991.10484804.
Pearl, J., 1988, Probabilistic reasoning in intelligent systems: Networks of plausible inference: Morgan Kaufmann.
Plummer, M., 2014, rjags: Bayesian graphical models using MCMC. R package version 4-6), https://CRAN.R-project.org/package=rjags, accessed 15 December 2017.
Raymer, L., E. Hunt, and J. Gardner, 1980, An improved sonic transit time-to-porosity transform: 21st Annual Logging Symposium, SPWLA, Extended Abstracts, SPWLA-1980-P.
R Development Core Team, 2011, R: A language and environment for statistical computing: R Foundation for Statistical Computing.
Rider, M., 1996, The geological interpretation of well logs, 2nd ed.: Whittles.
Saltelli, A., K. Chan, and E. Scott, 2000, Sensitivity analysis: Wiley.
Sayers, C. M., L. D. den Boer, Z. R. Nagy, and P. J. Hooyman, 2006, Well-constrained seismic estimation of pore pressure with uncertainty: The Leading Edge, **25**, 1524–1526, doi: 10.1190/1.2405338.
Shi, Y., and C.-Y. Wang, 1988, Generation of high pore pressures in accretionary prisms: Inferences from the Barbados subduction complex: Journal of Geophysical Research, **93**, 8893–8910, doi: 10.1029/JB093iB08p08893.
Su, Y.-S., and M. Yajima, 2014, R2jags: A package for running jags from r (R package version 0.04-01).
Van Wees, J.-D., H. Mijnlieff, J. Lutgert, J. Breunese, C. Bos, P. Rosenkranz, and F. Neele, 2008, A Bayesian belief network approach for assessing the impact of exploration prospect interdependency: An application to predict gas discoveries in the Netherlands: AAPG Bulletin, **92**, 1315–1336, doi: 10.1306/06040808067.
Wessling, S., A. Bartetzko, and P. Tesch, 2013, Quantification of uncertainty in a multistage/multiparameter modeling workflow: Pore pressure from geophysical well logs: Geophysics, **78**, no. 3, WB101–WB112, doi: 10.1190/geo2012-0402.1.
Wilkinson, D. J., and S. K. Yeung, 2002, Conditional simulation from highly structured Gaussian systems, with application to blocking-MCMC for the Bayesian analysis of very large linear models: Statistics and Computing, **12**, 287–300, doi: 10.1023/A:1020711129064.
Wyllie, M., A. Gregory, and L. Gardner, 1956, Elastic wave velocities in heterogeneous and porous media: Geophysics, **21**, 41–70, doi: 10.1190/1.1438217.
Zellner, A., 1995, Bayesian and non-Bayesian approaches to statistical inference and decision-making: Journal of Computational and Applied Mathematics, **64**, 3–10, doi: 10.1016/0377-0427(95)00002-X.