

Enhanced production surveillance using probabilistic dynamic models

Ashutosh Tewari¹, Stijn de Waele², and Niranjan Subrahmanya³

^{1,2} ExxonMobil Research and Engineering Company, Annandale, NJ: 08801, USA

ashutosh.tewari@exxonmobil.com

stijn.dewaele@exxonmobil.com

³ Former employee of ExxonMobil Research and Engineering Company, Annandale, NJ: 08801, USA

niranjan.udupa@gmail.com

ABSTRACT

Production surveillance is the task of monitoring oil and gas production from every well in a hydrocarbon field. A key opportunity in this domain is to improve the accuracy of flow measurements per phase (oil, water, gas) from a multi-phase flow. Multi-phase flow sensors are costly and therefore not instrumented for every production well. Instead, several low fidelity surrogate measurements are performed that capture different aspects of the flow. These measurements are then reconciled to obtain per-phase rate estimates. Current practices may not appropriately account for the production dynamics and the sensor issues, thus, fall far short in terms of achieving a desired surveillance accuracy. To improve surveillance accuracy, we pose rate reconciliation as a *state estimation* problem. We begin with hypothesizing a model that describes the dynamics of production rates and their relationship with the field measurements. The model appropriately accounts for the uncertainties in field conditions and measurements. We then develop robust probabilistic estimators for reconciliation to yield the production estimates and the uncertainties therein. We highlight recent advancements in the area of *probabilistic programming* that can go a long way in improving the performance and the portability of such estimators. The exposition of our methods is accompanied by experiments in a simulation environment to illustrate improved surveillance accuracy achieved in different production scenarios.

1. INTRODUCTION

Production in the context of a hydrocarbon field is the total output from production wells that, apart from oil, contains water, gases and particulate matter. Although accurate information about the total production from a field is usu-

ally available, there exist significant gaps in the per-well production estimates. The latter is valuable for several reasons that include equipment health monitoring, improved resource management, reduced operational cost and ultimately, production optimization. Even with state-of-the art sensors, the accuracy of multi-phase flow metering is quite limited. Because of the limited accuracy and the cost of multi-phase flow sensors, fields are instrumented with more conventional sensors that capture different aspects of the flow. Measurement from these multi-modal sensors are then combined to obtain an estimate of the production rate of each phase per individual well. This task is commonly referred to as *rate reconciliation*. Achieving a desired reconciliation accuracy is hampered by many issues such as sensor failures, evolving and unknown reservoir conditions, time varying sensor errors, sensors with varying sampling rates and duplicate measurements. Current deterministic approaches of surveillance are unable to address these issues in a systematic way, and thus remain sub-optimal at best.

In recent years, data-driven solutions that require minimal knowledge of the subsurface properties are gaining traction for production surveillance (Goh, Moncur, Van Overschee, Briers, et al., 2007; Poulisse et al., 2006). With a similar intent, we appeal to the rich area of *probabilistic state estimation*, which makes our approach the first to tackle the production surveillance task in a fully probabilistic framework. The approach establishes a dynamic production model describing the time evolution of production rates, and a measurement model defining the relationship between rates and the field measurements. Importantly, these models do not require knowledge of the subsurface properties or the physics of fluid flow through porous media. Once the models are specified, the rate reconciliation amounts to inverting the dynamic model for its states (and/or parameters), given the field measurements. An important feature of our approach is that it allows for the quantification of uncertainty in the estimated

Ashutosh Tewari et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

rates, thus leading to better operational decisions via improved risk assessment.

It should be noted that we do not cover the topic of sensor validation, which is a prerequisite to the surveillance approach discussed in this paper. We assume that the sensor data is validated and that the noise characteristics are known. This is a reasonable assumption in practice, since most modern oil fields have either a vendor supplied or a customized solution for validation and noise characterization. Also, this task is very sensor specific and cannot be put into a general framework.

In section 2, we describe a typical sensing environment found in a production field and outline a commonly used deterministic rate reconciliation approach. In section 3, we present a production model that appropriately captures the dynamics of production as well as the noise characteristics of field measurements. Simplifications to this model, for the purpose of inversion, are presented in section 4. Section 5 describes the probabilistic estimators to carry out the inversion of the aforementioned dynamic models. In section 6, we discuss the performance of these estimators in different production scenarios using a simulation study and finally conclude with remarks on future research direction in section 7.

2. MEASUREMENT SETUP

Consider a typical field set-up, where n wells are grouped together in the measurement configuration shown in figure 1. The output from these wells goes to a test separator, where the incoming liquid is separated into oil and water. For ease of exposition, we assume a two-phase flow, although the proposed methods can be extended for the presence of gas as a third phase. The following is a list of typical measurements that are available for reconciliation.

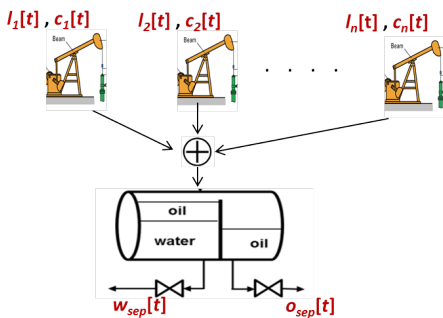


Figure 1. A schematic of typical sensing configuration in a production field, where a group of n wells feed to a separator. Text in red shows different types of measurements that can be available at time t . $w_{sep}[t]$ and $o_{sep}[t]$ are the aggregated water and oil rates from the separator. $l_i[t]$ and $c_i[t]$ are the measured total liquid rate and water-cut (fraction of water in the total flow) for the i^{th} well.

Test-separator rates ($w_{sep}[t]$, $o_{sep}[t]$): Test-separators are

complex devices that are used to measure individual phase rates in a multi-phase flow. They are based on the principle of gravimetric separation of different phases that are present in the incoming feed. Test-separators are quite expensive and cumbersome to instrument and each measurement can take upto a few hours waiting for gravimetric separation. For these reasons, it is very common to have a test-separator that is shared by a group of wells. During the production, each group is diverted to the separator based on a fixed schedule. In this way, one can get a snap-shot estimate of combined water and oil rates every few days.

Total liquid rates ($l_i[t]$): The total fluid production rate from each well is usually inferred using an indirect measurement (such as ultrasound sensing, venturimeter) or some form of soft-sensing model. For example, for wells that are equipped with artificial lift systems such as *sucker-rod* pumps, *electrical submersible* pumps or *gas-lift* systems, a soft-sensing model can estimate the amount of total liquid (oil and water) that is being produced. These measurements are often called *inferred* rates. Although these measurements could be very noisy and not informative about the individual oil/water rates, they are available much more frequently (perhaps hourly) and hence play a key role in a rate-reconciliation process.

Water-Cuts ($c_i[t]$): It is often possible to manually obtain a homogeneous sample of the liquid from a well and measure the fraction of water present in the liquid. This fraction is referred as the *water-cut*. Since the process is laborious, water-cuts are obtained very infrequently (perhaps monthly) but have high accuracy. Moreover, these measurements provide complimentary information that is valuable for rate-reconciliation.

Any reconciliation approach should assimilate these measurements as they become available and appropriately use them based on their information content and measurement noise level. Here we quickly describe a commonly employed *deterministic* reconciliation approach, which would serve as a reference approach for later comparisons. The deterministic approach starts with the total liquid rate measurements that are available for each individual well. These liquid rates are scaled to match the total liquid rate obtained from the last successful well-test from the test separator. A well-test is considered successful if a desired level of phase separation was attained during the test. The water-cuts are determined either by a manual water-cut measurement or a successful well-test, whichever one occurred the latest. All the wells are assigned the same water-cut value, if the value is coming from a well-test. Finally, with the information about the individual liquid rates (scaled) and the water-cut values, the water and oil rates per well are straightforward to estimate. Deterministic rate estimation suffers from many issues. It lacks a systematic framework to handle multitude of sensor issues (failures, varying noise levels, different sampling rates, data gaps etc.),

let alone a mechanism that accounts for declining production rates. In the light of these shortcomings, we propose a dynamic production model, in the next section, that forms the backbone of rate estimation approach presented in this paper.

We use the following notation in the rest of the paper; the lower case letters are reserved for scalar quantities, bold lower case letters for vectors and upper case letters for matrices. All the vectors are column vectors unless otherwise stated. Greek letters are used to denote model parameters. The square parenthesis $[\cdot]$ is used for time indexing.

3. A DYNAMIC PRODUCTION MODEL

The proposed dynamic model simulates production from a typical hydrocarbon field. Let $w_i[t]$ and $o_i[t]$ represent the water and the oil rates from the i^{th} well ($i \in [1, 2, \dots, n]$) at time t . We focus on the decline phase where the production displays an exponential type of decay (Höök, Davidsson, Johansson, & Tang, 2013). To model observed changes around the exponential decay and to ensure strictly positive rates, we sample the rates from log-normal distributions. Thus the time evolution of the rates can be written as:

$$w_i[t] \sim \text{LogNormal}(\log(\bar{w}_i[0]) - \lambda_{w_i}t, \gamma_{w_i}) \quad (1)$$

$$o_i[t] \sim \text{LogNormal}(\log(\bar{o}_i[0]) - \lambda_{o_i}t, \gamma_{o_i}) \quad (2)$$

$\bar{w}_i[0]$ and $\bar{o}_i[0]$ are the initial ($t = 0$) mean rates of water and oil respectively. The parameters λ_{w_i} and λ_{o_i} are the decay rates of the same. The arguments of the $\text{LogNormal}(\cdot)$ function are the mean and standard deviation of the rates in the log-space. Given the time evolving rates in (1) and (2), the sensor measurements can be simulated as follows:

$$\begin{aligned} w_{sep}[t] &\sim \text{Normal}(\sum_{i=1}^n w_i[t], \sigma_{sep}[t]) \quad \text{T}[0, \infty) \\ o_{sep}[t] &\sim \text{Normal}(\sum_{i=1}^n o_i[t], \sigma_{sep}[t]) \quad \text{T}[0, \infty) \\ l_i[t] &\sim \text{Normal}(w_i[t] + o_i[t], \sigma_{l_i}[t]) \quad \text{T}[0, \infty) \\ c_i[t] &\sim \text{Normal}(w_i[t]/(w_i[t] + o_i[t]), \sigma_c) \quad \text{T}[0, 1] \end{aligned} \quad (3)$$

Where the definitions of $w_{sep}[t]$, $o_{sep}[t]$, $l_i[t]$ and $c_i[t]$ follow from section 2. Note that the measurements are obtained by assuming an additive normally distributed error. Truncation of values between a and b is denoted $\text{T}[a, b]$. The time-dependent nature of measurement noise standard deviation, σ_{sep} and σ_{l_i} , are shown in equations (4) and (5). The choice of Gamma distribution and the subsequent parametrization is motivated by typical noise characteristics of field sensors. $\bar{\sigma}_{sep}$ and $\bar{\sigma}_{l_i}$ can be interpreted as average noise standard deviation of the sensors over a long term. The noise variance in manual water cut measurements, σ_c^2 , is assumed to be a

constant.

$$\sigma_{sep}[t] \sim \text{Gamma}\left(10, \frac{\bar{\sigma}_{sep}}{10}\right) \quad (4)$$

$$\sigma_{l_i}[t] \sim \text{Gamma}\left(10, \frac{\bar{\sigma}_{l_i}}{10}\right) \quad (5)$$

4. STATE-SPACE MODELS FOR RATE ESTIMATION

In the previous section, we presented a forward model that generates the production rates and the sensor measurements, given pre-specified parameters (decay rates, noise variances etc.). The corresponding inverse problem would entail recovery of the production rates as well as the underlying parameters, from the noisy sensor measurements. Even though the forward model in section 3 is a realistic description of the production in a hydrocarbon field, it may not be conducive for inversion (refer section 5.1). Here, we present two models as the alternatives to this forward model, that make some simplifying assumptions and hence allow for efficient inversion. A probabilistic state-space framework is chosen to describe these models. We refer the readers to (Srkk, 2013) and the references therein for more details on the chosen framework.

Let $\mathbf{x}[t]$ and $\mathbf{y}[t]$ be the state and the measurement vectors of the dynamical system at time t . The states correspond to unknown entities such as water and oil rates from all wells, in which case $\mathbf{x}[t] = [w_1[t], \dots, w_n[t], o_1[t], \dots, o_n[t]]^T$. The measurement vector contains all the observed sensor measurements at time t i.e. $\mathbf{y}[t] = [w_{sep}[t], o_{sep}[t], c[t], l_1[t], \dots, l_n[t]]^T$. A state-space model can be represented succinctly using equations 6 and 7. Equation 6 (*dynamic* component) describes the time evolution of the states, whereas equation 7 (*measurement* component) relates the measurements to the states. In the probabilistic setting, both the dynamic and measurement models are encoded as probability distributions $P_{dyn}(\cdot)$ and $P_{mea}(\cdot)$.

$$\mathbf{x}[t] \sim P_{dyn}(\mathbf{x}[t] | \mathbf{x}[t-1]) \quad (6)$$

$$\mathbf{y}[t] \sim P_{mea}(\mathbf{y}[t] | \mathbf{x}[t]) \quad (7)$$

The choices of $P_{dyn}(\cdot)$ and $P_{mea}(\cdot)$ and the subsequent parametrization determines the complexity of inversion. Inversion (with fixed parameters) in a state-space framework is also referred as the *state-estimation* problem. The two proposed state-space models differ only in the component describing the system dynamics, while having the same measurement component.

4.1. State Space Model 1 (SS-1)

SS-1 assumes simple linear dynamics with Gaussian noise as shown in equation (8). Symbols $A, Q[t] \in \mathcal{R}^{2n \times 2n}$ represent the state transition and the process noise covariance matrices respectively. The process noise is time varying as denoted by the time index in $Q[t]$. The measurement model, shown

in equation (9), assumes a Gaussian noise but could be non-linear with respect to the state as denoted by function $g(\cdot)$. Symbol $R[t]$ is the time-varying measurement noise variance, which is assumed given by a separate sensor validation process. Figure 2(a) shows graphical representation of SS-1.

$$P_{dyn}(\mathbf{x}[t] | \mathbf{x}[t-1]) = \text{Normal}(A\mathbf{x}[t-1], Q[t]) \quad (8)$$

$$P_{mea}(\mathbf{y}[t] | \mathbf{x}[t]) = \text{Normal}(g(\mathbf{x}[t]), R[t]) \quad (9)$$

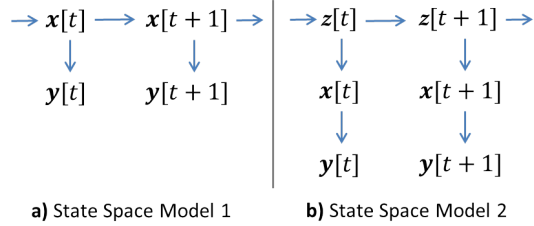


Figure 2. A graphical representation of the two state-space models with Markovian dynamics. For both models, $\mathbf{x}[\cdot]$ and $\mathbf{y}[\cdot]$ denote the unknown rates and the field measurements respectively. In SS-2, the state-space is augmented to include the mean of the rates, $\mathbf{z}[\cdot]$, in an attempt to closely mimic the dynamic model from section 3

4.2. State Space Model 2 (SS-2)

SS-1 is substantially different from the dynamic model presented in section 3. Here, we present another model (SS-2) that is derived by simplifying our dynamic production model. The simplification is due to an important insight about log-normal distributions, i.e. $\text{LogNormal}(\xi, \gamma)$ can be well approximated by $\text{Normal}(e^\xi, \gamma e^\xi)$ for a small value of γ . We argue that the γ values in equations (1) and (2) are indeed small. To support that we refer to an equality that holds true for γ_{w_i} (or γ_{o_i}) in relation to the mean and the variance of the true water rates (or oil rates).

$$\gamma_{w_i} = \sqrt{\log\left(1 + \frac{\text{Var}(w_i)}{\mathbb{E}(w_i)^2}\right)} \quad (10)$$

The ratio on the right hand side is the squared *coefficient of variation* (CV) i.e. the amount of variability in the true water rates, relative to the true mean. In reality, we expect CV for water and oil rates to be small (< 0.2). Using the Taylor series expansion of $\log(1+x)$ and ignoring the higher order terms, the γ value can be interpreted as CV and thus get bounded as shown below. Hence the normal approximation is valid.

$$\gamma_{w_i} \sim CV < 0.2 \quad (11)$$

With the Normal approximation of the LogNormal distribution, the simplified dynamic model can now be written as

shown below.

$$w_i[t] \sim \text{Normal}(\bar{w}_i[t], \gamma_{w_i} \bar{w}_i[t]) \quad (12)$$

$$\bar{w}_i[t] \sim \text{Normal}(e^{-\lambda_{w_i}} \bar{w}_i[t-1], s) \quad (13)$$

Note that (12) is due to the Normal approximation of a Log-Normal distribution and equation (13), with $s \rightarrow 0$, is a stochastic approximation of an exponentially decaying curve $\bar{w}_i[t] = \bar{w}_i[0]e^{-\lambda_{w_i}t}$. As a result of the aforementioned approximations, SS-2 can be defined in a similar fashion as SS-1, albeit with some modification to its dynamic component. The modification entails augmenting the state-space of SS-1 to include unknown states $\mathbf{z}[t]$ of the same dimension as $\mathbf{x}[t]$ (refer figure 2(b)). $\mathbf{z}[t]$ is the expected water and oil production, at time t i.e. $\mathbf{z}[t] = [\bar{w}_1[t], \dots, \bar{w}_n[t], \bar{o}_1[t], \dots, \bar{o}_n[t]]^T$. The dynamic component of SS-2 is given by equations (14) and (15) which are the multivariate extension of the equations (13) and (12) respectively.

$$\mathbf{z}[t] \sim \text{Normal}(A\mathbf{z}[t-1], S) \quad (14)$$

$$\mathbf{x}[t] \sim \text{Normal}(\mathbf{z}[t], F[t]) \quad (15)$$

The diagonal covariance matrix $S \in \mathcal{R}^{2n \times 2n}$ is constant with diagonal entries equal to s . The covariance matrix $F[t]$ is diagonal and can be written as $F[t] = \text{diag}(\gamma \otimes \mathbf{z}[t])$, where $\gamma = [\gamma_{w_1}, \dots, \gamma_{w_n}, \gamma_{o_1}, \dots, \gamma_{o_n}]$. The transition matrix, A , is defined as $A = \text{diag}(e^{-\lambda})$, where $\lambda = [\lambda_{w_1}, \dots, \lambda_{w_n}, \lambda_{o_1}, \dots, \lambda_{o_n}]$. As for the observation component of SS-2, it remains identical to that of SS-1. Therefore, the difference in the two models exists only in the dynamics, which is shifted from states $\mathbf{x}[t]$ to $\mathbf{z}[t]$ in SS-2. Figure 3 juxtaposes the time evolution of a state under the dynamics specified by SS-1 and SS-2. The initial state distribution is the same for the two models, so are the time varying process noise $F[t]$ and $Q[t]$. The top and the bottom row correspond to case when the system exhibits a random walk dynamics ($A = 1$) and an exponential decline ($A = 0.97$) respectively.

5. RATE ESTIMATION

Here we present estimators for the dynamic models described in previous sections. Section 5.1 describes an approach that attempts to directly invert the model presented in section 3. Because this direct inversion is notably hard, we highlight recent algorithmic and software advancements that have made it possible. Section 5.2 presents more efficient recursive estimators for the simplified state space models from section 4.

5.1. Direct Estimator

Given all the measurements, $\mathbf{y}[1:t]$, from time 1 to t , the direct estimator attempts to invert the model from section 3, for the unknown rates $\mathbf{x}[1:t]$ and the unknown parameters Θ .

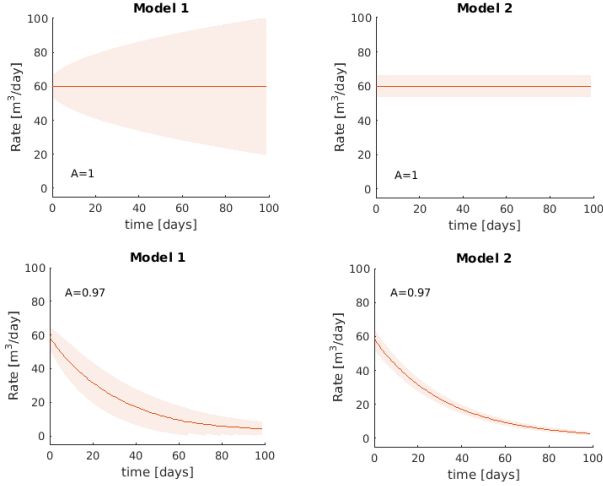


Figure 3. A graphical comparison of the dynamics encoded by the two state-space models. The top row shows the time evolution of a state under random walk dynamics ($A = 1$, half-life = ∞) and the bottom row shows the same for an exponential decay with a half-life of 25 days ($A = 0.97$).

This inversion can be represented as the following probabilistic inference.

$$P(\mathbf{x}[1:t], \Theta \mid \mathbf{y}[1:t]) \quad (16)$$

The parameter set Θ is comprised of $\{\lambda_{w_i}, \lambda_{o_i}, \gamma_{w_i}, \gamma_{o_i}\}_{i=1}^n$. From the production surveillance viewpoint, the benefit of this approach comes from the simultaneous state and parameter estimation, which leads to improved surveillance accuracy. In addition, being an off-line setting, measurement from the future can be used to improve the rates estimates from the past. However, these benefits come at the cost of significant computational challenges. First, because of the non-linear, non-Gaussian model with state constraints, the posterior in equation (16) cannot be obtained analytically. Second, the dimensionality of state-space ($\mathbf{x}[1:t]$), which grows with time, can become prohibitively large for any sampling based method for posterior approximation. For these reasons, in practice, recursive estimators (section 5.2) are carefully designed to overcome these challenges but of course at the cost of reduced estimation accuracy.

In recent years, with the advent of expressive probabilistic programming languages (Gordon, Henzinger, Nori, & Rajamani, 2014), such direct inversion has become practical. The primary motivation behind such languages is to make machine learning applications, rooted in probabilistic modeling, easier to build. The benefits of probabilistic programming are twofold. First, it provides the means for describing complex probabilistic models that would otherwise take many lines of code in traditional programming languages. And second, the inference, such as the one in equation (16), is automatic and does not require additional effort on the part of a modeler.

For example, solving (16) using forward-backward recursion will not only require some simplifying assumptions but also significant effort in algorithm and code development. A probabilistic programming language, with its own inference engine, completely obviates the need of a separate inference capability. We selected Stan (Carpenter et al., 2017) as the probabilistic programming language to implement our model because of its maturity, robustness and speed. Figure 4 shows a snippet of Stan code for illustration. The one to one correspondence between the mathematical representation of the production model (equations (1), (2) and (3)) and the code is specially to be noted. For the inference purpose, Stan offers automatic inference engine based on variational (Kucukelbir, Ranganath, Gelman, & Blei, 2015) and sampling based algorithms (Hoffman & Gelman, 2014). Results presented in the paper uses the latter.

```

for (t in 1:T) {
  for (i in 1:n) {
    # Priors from domain knowledge
    gamma_w[i] ~ halfnormal(0, 0.25);
    gamma_o[i] ~ halfnormal(0, 0.25);
    lambda_w[i] ~ halfnormal(0, 0.1);
    lambda_o[i] ~ halfnormal(0, 0.1);
    # Time evolving water and oil rates
    w[i, t] ~ lognormal(log(w_bar[i]) - lambda_w[i]*t, gamma_w[i]);
    o[i, t] ~ lognormal(log(o_bar[i]) - lambda_o[i]*t, gamma_o[i]);
    # Field measurements: total liquid rate and water-cuts
    l[i, t] ~ normal(w[i, t] + o[i, t], sigma_l[i, t]) T[0,];
    c[i, t] ~ normal(w[i, t] / (w[i, t] + o[i, t]), sigma_c) T[0, 1];
  }
  # Field measurements: test-separator rates
  w_sep[t] ~ normal(sum(col(w, t)), sigma_sep[t]) T[0,];
  o_sep[t] ~ normal(sum(col(o, t)), sigma_sep[t]) T[0,];
}

```

Figure 4. Stan code for the simulation model described in section 3. Only the *model* block of the code is included for illustration. The other required programming blocks *parameters* and *data*, which specify the unknowns and observed entities, are not shown. The inference is carried out on the parameters conditioned on the observed data and the specified model.

The Stan code in figure 4 specifies prior distributions (halfnormal) on the parameters λ and γ . We base the standard deviation for these distributions on knowledge of site operations, then multiply by a factor greater than 1 to obtain weakly informative prior. The impact of the prior is discussed briefly in section 6. Such priors from domain knowledge and other physical constraints, like non-negativity of flow rates, can be easily encoded in Stan. The automatic inference procedure ensures that the posterior distribution/samples honor such knowledge or constraints.

The biggest advantage of probabilistic programming comes from the agile development of the applications. For instance, in another hydrocarbon field, with different production dynamics and sensor configuration, the rate estimator (Stan code) can be readily adapted by changing the production model, the measurement model and the sensor configuration. On the contrary, with a traditional programming language, significant effort will be needed to hand-craft and code an appropos

inference algorithm.

5.2. Recursive Estimators for the State Space Models

Recursive rate estimation amounts to performing the following probabilistic inference.

$$P(\mathbf{x}[t] | \mathbf{y}[1:t], \Theta) \quad (17)$$

Note that this is a much smaller inference problem than the one given in equation (16). It can be argued that this inference can also include the parameter Θ , so as to perform combined state and parameter estimation, along the line of work in (Liu & West, 2001). However, limiting only to state estimation greatly enables real-time implementation of the algorithm in large assets with many producing wells. Nevertheless, simultaneous parameter and state estimation, in a recursive setting, could be an important capability that we may pursue in future research. Henceforth, we will denote the distribution $P(\mathbf{x}[t] | \mathbf{y}[1:t], \Theta)$ as $\pi[t]$ for brevity. The goal of recursive Bayesian estimation is to perform the inference in equation 17, recursively, for each time step by repeated applications of *prediction* and *correction* steps. The literature on recursive estimation is rather rich, ranging from Kalman filters and its variants to numerous Sequential Monte Carlo (SMC) methods (Arulampalam, Maskell, Gordon, & Clapp, 2002). Our goal was to design robust and efficient recursive estimators for our state-space models by leveraging existing methods. To this end, we developed a hybrid estimator which combines analytical approaches with the approximate ones (sampling based), thereby bringing the best of both worlds.

Before presenting the hybrid estimator, an important note on the parametric form of the posterior distribution ($\pi[t]$). We enforce $\pi[t]$ to be a Mixture of Gaussians (MoG). The use of MoG posterior distribution is well documented (Alspach & Sorenson, 1972; Sorenson & Alspach, 1971; Vo & Ma, 2006) (Sondergaard & Lermusiaux, 2013), in diverse problem settings where the Gaussian assumption is severely limiting. In our case, the need for a MoG arises due to the fact that flow distributions may deviate from normality as the flows approach zero. MoG is a suitable choice to model such non-normality mainly for three reasons. First, MoG can characterize arbitrary non-normal distributions; second, a MoG prior is conjugate with respect to a linear-Gaussian observations; and lastly, it provides a mechanism to bridge the analytical and approximate approaches for posterior computation. These reasons are central for the success of our hybrid estimator and would become clearer in the ensuing discussion. Let $\psi(\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote a k component MoG distribution as shown in (18).

$$\psi(\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{c=1}^k \alpha_c \times \text{Normal}(\mu_c, \Sigma_c) \quad (18)$$

where $\boldsymbol{\alpha} = \{\alpha_c\}_{c=1}^k$, $\boldsymbol{\mu} = \{\mu_c\}_{c=1}^k$, $\boldsymbol{\Sigma} = \{\Sigma_c\}_{c=1}^k$

The symbols α_c , μ_c and Σ_c are the mixing proportions, the mean, and the covariance parameters of the c^{th} Gaussian component respectively. Thus $\pi[t]$ becomes $\psi(\boldsymbol{\alpha}[t], \boldsymbol{\mu}[t], \boldsymbol{\Sigma}[t])$. In the following, we estimate the posterior distribution $\pi[t+1]$, for SS-1 and SS-2, given the prior distribution $\pi[t]$ and a new observation vector $\mathbf{y}[t+1]$. For completeness, we explicitly provide the update equations of the *prediction* and *correction* steps. The detailed derivation of the update equations for SS-1 can be found elsewhere (Sorenson & Alspach, 1971). For SS-2, the derivation is included in the appendix. In cases where correction step can not be carried out exactly, we provide suitable approximations.

5.2.1. Prediction:

In the prediction step, we propagate the posterior, $\pi[t]$, at time t to the next time step using the dynamics specified in the state-space model. Let's denote the propagated distribution as $\tilde{\pi}[t+1]$. Since, both SS-1 and SS-2 have linear dynamics and since $\pi[t]$ is a MoG, $\tilde{\pi}[t+1]$ remains a MoG and can be represented as $\psi(\tilde{\boldsymbol{\alpha}}[t+1], \tilde{\boldsymbol{\mu}}[t+1], \tilde{\boldsymbol{\Sigma}}[t+1])$. The expressions for the parameters, $\tilde{\boldsymbol{\alpha}}[t+1]$, $\tilde{\boldsymbol{\mu}}[t+1]$, $\tilde{\boldsymbol{\Sigma}}[t+1]$, for the two models are provided as follows.

SS-1: The propagated mean ($\tilde{\mu}_c[t+1]$) is simply the linear transformation A of the current posterior mean. The covariance ($\tilde{\Sigma}_c[t+1]$) is the linear transformation of the current posterior covariance via A , plus the noise covariance $Q[t]$. The mixing proportions ($\tilde{\alpha}_c[t+1]$) remain unchanged.

$$\begin{cases} \tilde{\alpha}_c[t+1] = \alpha_c[t] \\ \tilde{\mu}_c[t+1] = A\mu_c[t] \\ \tilde{\Sigma}_c[t+1] = A\Sigma_c[t]A^T + Q[t] \end{cases} \quad (19)$$

SS-2: We write the parameters of SS-2 as shown in equation (20). This representation is due to the grouping of the parameters with respect to states \mathbf{z} and \mathbf{x} . Thus, the parameters of $\tilde{\pi}[t+1]$ can be written in terms of the parameter of $\pi[t]$, as shown in equation group (21) (details included in the appendix).

$$\tilde{\mu}_c[t+1] = \begin{bmatrix} \tilde{\mu}_c^z[t+1] \\ \tilde{\mu}_c^x[t+1] \end{bmatrix}, \tilde{\Sigma}_c[t+1] = \begin{bmatrix} \tilde{\Sigma}_c^{zz}[t+1] & \tilde{\Sigma}_c^{zx}[t+1] \\ \tilde{\Sigma}_c^{xz}[t+1] & \tilde{\Sigma}_c^{xx}[t+1] \end{bmatrix} \quad (20)$$

where

$$\begin{cases} \tilde{\mu}_c^z[t+1] = A\mu_c^z[t] \\ \tilde{\mu}_c^x[t+1] = A\mu_c^x[t] \\ \tilde{\Sigma}_c^{zz}[t+1] = A\Sigma_c^{zz}[t]A^T + S \\ \tilde{\Sigma}_c^{zx}[t+1] = A\Sigma_c^{zx}[t]A^T + S \\ \tilde{\Sigma}_c^{xx}[t+1] = A\Sigma_c^{xx}[t]A^T + S + F[t] \end{cases} \quad (21)$$

The mixing proportions remain unchanged i.e. $\tilde{\alpha}_c[t+1] = \alpha_c[t]$.

5.2.2. Correction (exact):

In the correction step, the posterior distribution $\pi[t + 1]$ is obtained by performing the following Bayesian inference,

$$\pi[t + 1] = \frac{\tilde{\pi}[t + 1] \times \phi(\mathbf{y}[t + 1]; g(\mathbf{x}[t + 1]), R[t + 1])}{\int_{\mathbf{x}[t+1]} \tilde{\pi}[t + 1] \times \phi(\mathbf{y}[t + 1]; g(\mathbf{x}[t + 1]), R[t + 1])} \quad (22)$$

where, $\phi(\cdot)$ denotes the Normal density function. Equation 22 holds for both the models as they have the same measurement component. The complexity of this posterior computation depends on the form of $g(\cdot)$. For instance, when $g(\cdot)$ is linear i.e. $g(\mathbf{x}[t + 1]) = H\mathbf{x}[t + 1]$, where $H \in \mathcal{R}^{m \times 2n}$ (m =number of measurements), $\pi[t + 1]$ can be obtained analytically. The Matrix H is expected to be known from the information about the measurement configuration. For instance, consider a group of three wells ($n = 3$), in which case $[w_1[t+1] w_2[t+1] w_3[t+1] o_1[t+1] o_2[t+1] o_3[t+1]]^T$ denotes the state vector $\mathbf{x}[t + 1]$. Let the measurement vector at time $t + 1$ be $\mathbf{y}[t + 1] = [l_1[t+1] l_2[t+1] l_3[t+1] w[t+1] o[t+1]]^T$ (refer section 2). As a result, the matrix H takes the form shown in equation 23.

$$H = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \quad (23)$$

With this information, we can obtain the update equations for the parameters of $\pi[t + 1]$. Since the MoG prior ($\tilde{\pi}[t + 1]$) is conjugate for linear-Gaussian observations, the posterior in equation 22 is again a MoG distribution, i.e. $\pi[t + 1] = \psi(\boldsymbol{\alpha}[t + 1], \boldsymbol{\mu}[t + 1], \boldsymbol{\Sigma}[t + 1])$. In the following, we explicitly write the update equations for $\boldsymbol{\alpha}[t + 1]$, $\boldsymbol{\mu}[t + 1]$ and $\boldsymbol{\Sigma}[t + 1]$, for the two state-space models.

SS-1: The parameter update equations of the first model are given in equation group 24.

$$\begin{cases} \alpha_c[t + 1] \propto \tilde{\alpha}_c[t + 1] \times \phi(\mathbf{y}[t + 1]; H\tilde{\boldsymbol{\mu}}_c[t + 1], J_c[t + 1]) \\ \boldsymbol{\mu}_c[t + 1] = \tilde{\boldsymbol{\mu}}_c[t + 1] + K_c[t + 1] \times (\mathbf{y}[t + 1] - H\tilde{\boldsymbol{\mu}}_c[t + 1]) \\ \boldsymbol{\Sigma}_c[t + 1] = \tilde{\boldsymbol{\Sigma}}_c[t + 1] - K_c[t + 1] \times H\tilde{\boldsymbol{\Sigma}}_c[t + 1] \end{cases} \quad (24)$$

where $J_c[t + 1] = (H\tilde{\boldsymbol{\Sigma}}_c[t + 1]H^T + R[t + 1])$ and $K_c[t + 1] = \tilde{\boldsymbol{\Sigma}}_c[t + 1]H^T J_c^{-1}[t + 1]$ is the Kalman gain for the c^{th} Gaussian component. The first equation in the group updates the mixing proportions and the proportionality sign there indicates the need of re-normalization so that the posterior remains a valid MoG distribution.

SS-2: For SS-2, let the mean and the covariance of the c^{th} component be represented as $\boldsymbol{\mu}_c[t + 1] = \begin{bmatrix} \mu_c^z[t + 1] \\ \mu_c^x[t + 1] \end{bmatrix}$, $\boldsymbol{\Sigma}_c[t + 1]$

$\boldsymbol{\Sigma}_c^z[t + 1]$ and $\boldsymbol{\Sigma}_c^{zx}[t + 1]$ respectively. The parameter update equations can be derived to yield equation group (25) (details included in the appendix).

$$\begin{cases} \alpha_c[t + 1] \propto \tilde{\alpha}_c[t + 1] \times \phi(\mathbf{y}[t + 1]; H\tilde{\boldsymbol{\mu}}_c[t + 1], J_c[t + 1]) \\ \mu_c^z[t + 1] = \tilde{\mu}_c^z[t + 1] + K_c^z[t + 1] \times \mathbf{r}[t + 1] \\ \mu_c^x[t + 1] = \tilde{\mu}_c^x[t + 1] + K_c^x[t + 1] \times \mathbf{r}[t + 1] \\ \boldsymbol{\Sigma}_c^{zz}[t + 1] = \tilde{\boldsymbol{\Sigma}}_c^{zz}[t + 1] - K_c^z[t + 1] \times H\tilde{\boldsymbol{\Sigma}}_c^{zz}[t + 1] \\ \boldsymbol{\Sigma}_c^{zx}[t + 1] = \tilde{\boldsymbol{\Sigma}}_c^{zx}[t + 1] - K_c^z[t + 1] \times H\tilde{\boldsymbol{\Sigma}}_c^{zx}[t + 1] \\ \boldsymbol{\Sigma}_c^{xx}[t + 1] = \tilde{\boldsymbol{\Sigma}}_c^{xx}[t + 1] - K_c^x[t + 1] \times H\tilde{\boldsymbol{\Sigma}}_c^{xx}[t + 1] \end{cases} \quad (25)$$

where,

$$\begin{aligned} K_c^z[t + 1] &= \tilde{\boldsymbol{\Sigma}}_c^{zz}[t + 1]H^T(H\tilde{\boldsymbol{\Sigma}}_c^{zz}[t + 1]H^T + R[t + 1])^{-1} \\ K_c^x[t + 1] &= \tilde{\boldsymbol{\Sigma}}_c^{zx}[t + 1]H^T(H\tilde{\boldsymbol{\Sigma}}_c^{zz}[t + 1]H^T + R[t + 1])^{-1} \\ \mathbf{r}[t + 1] &= \mathbf{y}[t + 1] - H\tilde{\boldsymbol{\mu}}_c[t + 1]. \end{aligned}$$

5.2.3. Correction (approximate):

The posterior in equation (22) need not always assume a close-form and hence requires approximation. The need of approximation arises mainly due to the non-linear measurements and/or the physical constraints on the states. We elaborate these two scenarios and propose suitable methods for approximating the posterior.

Scenario 1 (state constraints): Well production rates need to satisfy non-negativity constraint. Additionally, one may also have information about the pump sizes that defines an upper limit on the well rates. However, the exact method for posterior computation from section 5.2.2 does not directly lend itself to such constraints. Numerous options exist to incorporate such inequality constraints that include methods based on sampling, projection, density truncation, unscented transformation, moving horizon estimation etc. Reference (Simon, 2010) covers these options in detail and compares their performance in terms of accuracy and computational overhead. We down-selected *Importance Sampling* (IS) for our purpose. IS works on the premise that there exists an easy to sample *proposal* distribution, samples from which, allow us to approximate the *target* (posterior) distribution. IS works quite well when the target and the proposal distributions are not too different in terms of how the probability mass is distributed in the state-space (Tokdar & Kass, 2010). This condition is very likely to be satisfied if we take the proposal distribution as the unconstrained posterior obtained using the exact method (section 5.2.2). Let's denote this unconstrained posterior as $\pi_{uncon}[t + 1]$. Since $\pi_{uncon}[t + 1]$ is a MoG, sampling from it is efficient. Let \mathbf{s}_i denote the i^{th} sample from a set of total q samples from the chosen proposal distribution. IS produces a set of weighted samples $\{w_i, \mathbf{s}_i\}_{i=1}^q$, where the weights are

obtained as follows

$$w'_i = \frac{\tilde{\pi}[t+1](s_i) \times \phi(\mathbf{y}[t+1]; g(s_i), R[t+1]) \times u(\mathbf{s}_i; \mathbf{l}, \mathbf{u})}{\pi_{uncon}[t+1](s_i)}$$

$$w_i = \frac{w'_i}{\sum_i w'_i} \quad (26)$$

Note that the numerator is the unnormalized posterior from equation 22 with the added uniform density term $u(\cdot)$ to ensure that the lower (\mathbf{l}) and the upper (\mathbf{u}) bounds are accounted for. From the weighted sample set, $\{w_i, \mathbf{s}_i\}_{i=1}^q$, the target distribution, $\pi[t+1]$, can be obtained using the *Expectation-Maximization* (EM) algorithm (Bilmes, 1997) for MoG distribution. The standard EM algorithm assumes equally weighted samples and hence needs a minor modification to account for the sample weights.

Scenario 2 (water-cut measurements): A water-cut measurement adds non-linearity to our model which needs to be addressed during posterior computation. Such non-linear measurements can be handled in different frameworks such as Extended (E) or Unscented (U) Kalman filters (KF). We resort to a sampling-based approach for this purpose because it can yield more accurate posterior distributions than the EKF and UKF, provided a large number of samples can be simulated from the true posterior. One approach could be to use $\pi[t+1]$ (exact posterior without the water-cut measurement) as the proposal distribution in the IS framework, as done previously. However, IS may not be a wise choice because of potentially significant differences in the proposal and target distribution. A water-cut measurement can drastically change the posterior distribution from the prior, in which case IS will require prohibitively large number of samples to get to the target distribution. A Markov Chain Monte Carlo (MCMC) method is more suited for posterior simulation in such situations. There are many MCMC algorithms to choose from, ranging from well known Metropolis-Hastings (MH) (Chib & Greenburg, 1995) to relatively recent Slice-sampling (Neal, 2003) and Hamiltonian Monte Carlo (HMC) (Hoffman & Gelman, 2014) algorithms. For our purpose we used the HMC for posterior simulation. Unlike IS, HMC yields equally weighted (or unweighted) samples. The samples thus generated are used to obtain the MoG distribution using the standard EM algorithm.

As a closing remark; we adopt a strategy to combine the exact and approximate methods in our recursive estimators. That is, carry out the correction step exactly, using the approach presented in section 5.2.2, and switch to an approximate method (section 5.2.3) when exact solution is not applicable. Since, both the exact and approximate methods yield a MoG distribution in the end, the transition between the two becomes seamless.

Table 1. The values of dynamic model parameters of the three wells used for generating production rates in the simulation study. The parameter symbols correspond to the ones shown in equations 1 and 2. The unit of initial rates ($\bar{w}[0]$ and $\bar{o}[0]$) is m^3/day . The declines rates (λ_w and λ_o) are specified by their half-life in days.

Well	$\bar{w}[0]$	$\bar{o}[0]$	λ_w	λ_o	γ_w	γ_o
1	90	10	20	10	0.05	0.05
2	50	50	20	10	0.05	0.05
3	14	6	20	10	0.05	0.05

Table 2. Description of the three cases used to evaluate the proposed estimators. These cases differ in the following measurement settings. Gaps: *no* - test separator measurements (w_{sep}, o_{sep}) are available on a daily basis; *yes* - test separator measurements are available once per 3 days. Noise: Measurement noise $\sigma_{sep}, \sigma_{l_i}, \sigma_c$ is low or high. t_{wc} : A water cut measurement (c_i) is available for each well at the start ($t_{wc} = 1$) or after 5 days ($t_{wc} = 5$). For all 3 cases, daily measurements of the per-well liquid rate (l_i) are assumed to be available.)

Case	Gaps	Noise	t_{wc}
A	no	low	1
B	yes	high	1
C	yes	high	5

6. EXPERIMENTS

Experiments were conducted in a simulation environment to assess the performance of different estimators. The need of simulation stems mainly because it provides access to ground truth rates for the validation of the estimators. Similar information from an oil field is not only proprietary but also cost prohibitive to obtain, thus preventing us from using real production data for any meaningful analysis. The simulations are governed by the dynamic model described in section 3. We investigate a configuration of $n = 3$ wells feeding into the test separator, as depicted in figure 1. The true production rates are generated, for a period of 30 days, using the dynamic model parameters specified in table 1. For the measurements, the base sampling interval is one day. However, every measurement need not be available on a daily basis. Three test cases were defined by varying the measurement settings (see table 2 for details) to generate diverse datasets for reconciliation. The idea is to generate measurements such that the reconciliation becomes increasingly challenging as we go from case A to C.

We evaluate the performance of the following estimators on the three test cases. The recursive rate estimator 1 with fixed parameters (SS-1), the direct estimator (DIR), the recursive estimators with estimated parameters (SS-1 E and SS-2 E) and the deterministic estimator (DET) as described in section 2. All the estimators, except for DIR (which uses Stan), are implemented in Matlab. In the fixed parameter cases (SS-1), the transition matrix A is assumed to be an identity matrix and the process noise covariances $Q[t]$ is obtained by assuming

the coefficient of variation of rates as 0.1. Although crude, both these assumptions are somewhat reasonable in the absence of any information about the dynamics of production rates. For the estimated parameter case (SS-1 E and SS-2 E), the recursive estimators are seeded with the posterior mean of the dynamic model parameters obtained from the direct estimator. The goal was to assess the uplift in the estimation accuracy if parameters are also estimated from data. As a reference (REF), we use the true state distributions known from the simulation settings. Two performance metrics, the mean Absolute Error (AE) and the scaled likelihood (L) are defined to quantify the performance of an estimator. The former is a measure of distance between the estimated mean and the ground truth, while the latter assesses the quality the uncertainty estimates. Note that the likelihood metric (L) is the log probability ratio of the true distribution $f_{REF}(\cdot)$ and the estimated distribution $\hat{f}(\cdot)$ by an estimator. Also, both metrics are non-negative with smaller value indicating better performance.

$$AE = \frac{1}{30} \sum_{i=1}^3 \sum_{t=1}^{30} (|\hat{o}_i[t] - o_i[t]| + |\hat{w}_i[t] - w_i[t]|) \quad (27)$$

$$L = \frac{1}{30} \sum_{t=1}^{30} 2(\log f_{REF}(\mathbf{x}[t]) - \log \hat{f}(\mathbf{x}[t])) \quad (28)$$

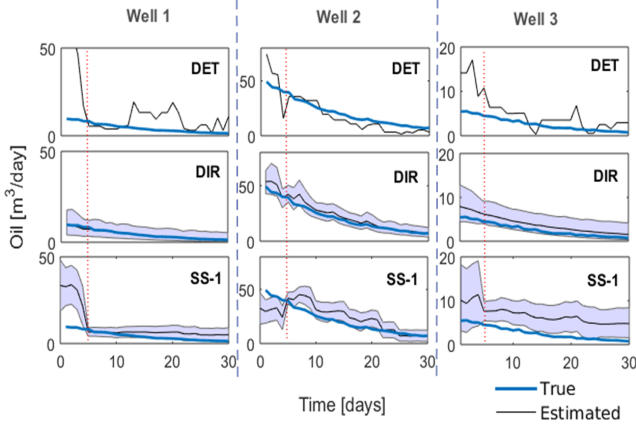


Figure 5. Estimated oil rates obtained from the three estimators (deterministic (DET), direct (DIR) and state-space 1(SS-1)) for simulation case C. The dotted red line shows the location of the only water-cut measurement. The deterministic estimator does not quantify the uncertainty. For the other two estimators, we show the 80 % confidence interval.

Figure 5 shows reconciliation results for a typical simulation run. For conciseness, we have selected three key estimators (DET, DIR and SS-1), and only show estimates for the oil rate. The DET and SS-1 estimators were initialized to have equal water and oil rates at the start. Unlike the other two, the deterministic method (DET) does not yield an uncertainty interval. SS-1 and DET estimators are far off from the ground

truth at the start. Once a water-cut measurement is taken at $t = 5$, accuracy of both these estimators improves markedly. The direct estimator (DIR) performs the best especially early on as it uses the information from future (water-cut at $t = 5$) to improve the estimates during $t < 5$.

A detailed comparison of the performance of the estimators, using metrics AE and L, on the three test cases is shown in figure 6. The following conclusions can be drawn from these results. The state-space model SS-1, despite having crude dynamic model parameters, delivers reasonably accurate rate estimates at a low computational cost. The absolute error is reduced by approximately a factor of 2 compared to the deterministic estimator (DET). Using estimated model parameters in SS-1 E, further improves its performance. Also, it is expected this estimator will remain accurate over a wider range of circumstances. The direct estimator, DIR, delivers the most accurate result by further reducing the absolute error by another factor of 2 compared to SS-1. A part of this increased accuracy can be attributed to the fact that DIR inverts the same model that generates the synthetic data. However, two other factors contribute to the increased accuracy here. First, DIR estimates both the parameters (decline rates, process noise levels) and states (production rates) from the measurements. Second, by virtue of being a batch estimator, DIR also uses future measurements to improve the past rate estimates (*smoothing*), as oppose to the other estimators that operate in an online fashion (*filtering*). Thus, in a way, the evaluation of DIR provides a lower bound on the error (with respect to the ground truth) with which the production rates can be estimated from the measurements of varying fidelities.

The estimation of model parameters ($\lambda_{o_i}, \lambda_{w_i}, \gamma_{o_i}, \gamma_{w_i}$), besides contributing to more accurate rate estimates, can be of great significance from the viewpoints of production optimization and field management in general. Figure 7 shows prior and posterior credible intervals of these parameters obtained from the DIR estimator, for the simulation case A. The true parameter values are also shown for comparison. For most parameters, the data is informative about the parameters, leading to a small posterior uncertainty compared to the prior. The exception is the oil parameters for well 3. This is caused by the fact that the oil production rate for this well is low, resulting in a low signal-to-noise ratio.

7. CONCLUDING REMARKS

We propose a probabilistic approach for the task of production surveillance in oil fields. Accurate production surveillance is important to ensure productive, profitable and safe operations in oil fields. The incumbent surveillance approaches are ill-equipped to handle a variety of real-world issues such as degrading sensors, sensors with varying sampling rates, low-resolution and duplicate measurements, time-varying sensor noise, communication drops etc. The proposed approach

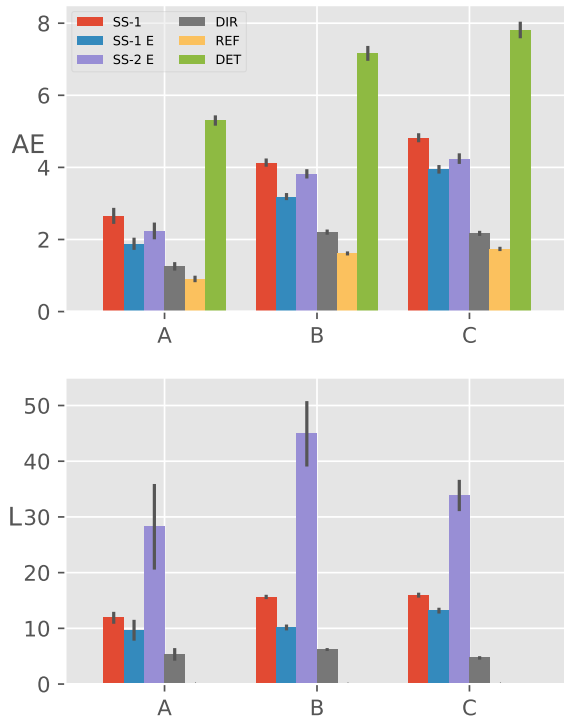


Figure 6. Average error in the estimated production rates per estimator and per simulation case, averaged over 100 simulation runs. Top: Absolute Error (AE); bottom: scaled log likelihood (L). For the scaled log likelihood, DET is omitted because it does not define a distribution; for REF, $L=0$ by definition.

handles these issues in a systematic way in order to yield production estimates of higher accuracy. The underpinnings of our approach are the dynamic production models (of varying complexities) that capture the critical aspects of production in an actual field. We present estimators for these dynamic models by leveraging state-of-the-art methods from the area of probabilistic state estimation, with the intent of meeting the requirements imposed by the operations in a large oil field. Apart from improved surveillance accuracy, another important outcome of the proposed approach is the ability to quantify the uncertainties in production rates; something that is largely ignored in the current practice. This capability opens the door for future research opportunities, for example, field optimization under uncertainty to attain better asset allocation and facility utilization. Another direction of research pertains to the idea *active* surveillance, which aims at collecting measurements, when and where the uncertainty is the highest, by actively perturbing the sensing environment. If successful, such research areas will enable a multitude of tasks geared towards building the next generation of digital oil fields.

REFERENCES

Alspach, D. L., & Sorenson, H. W. (1972, August). Nonlin-

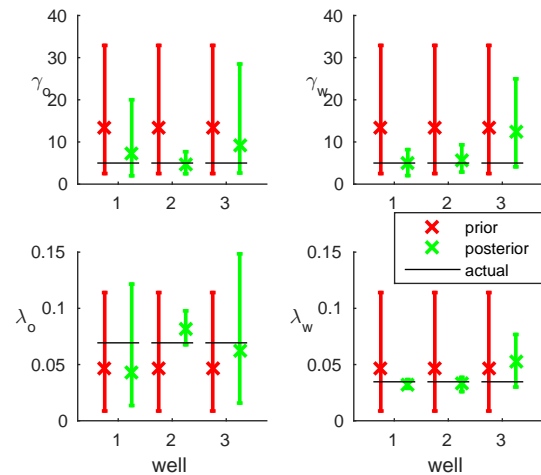


Figure 7. Prior and posterior credible intervals of dynamic model parameters for simulation case A. Top row: coefficient of variation γ (%); bottom row: decay rates λ .

ear Bayesian estimation using Gaussian sum approximations. *Automatic Control, IEEE Transactions on*, 17(4), 439–448. doi: 10.1109/tac.1972.1100034

- Arulampalam, S. M., Maskell, S., Gordon, N., & Clapp, T. (2002, February). A tutorial on particle filters for on-line nonlinear/non-gaussian bayesian tracking. *Trans. Sig. Proc.*, 50(2), 174–188. doi: 10.1109/78.978374
- Bilmes, J. (1997). *A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models* (Tech. Rep. No. TR-97-021). International Computer Science Institute.
- Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, Articles*, 76(1), 1–32.
- Chib, S., & Greenburg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4), 327–335.
- Goh, K.-C., Moncur, C. E., Van Overschee, P., Briers, J., et al. (2007). Production surveillance and optimization with data driven models. In *International petroleum technology conference*.
- Gordon, A. D., Henzinger, T. A., Nori, A. V., & Rajamani, S. K. (2014). Probabilistic programming. In *International conference on software engineering (icse, fose track)*.
- Hoffman, M. D., & Gelman, A. (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1), 1593–1623.
- Höök, M., Davidsson, S., Johansson, S., & Tang, X. (2013). Decline and depletion rates of oil production: a comprehensive investigation. *Philosophical Transactions*

of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 372(2006). doi: 10.1098/rsta.2012.0448

- Kucukelbir, A., Ranganath, R., Gelman, A., & Blei, D. M. (2015). Automatic variational inference in stan. In *Neural information processing systems*.
- Liu, J., & West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In A. Doucet, N. de Freitas, & N. Gordon (Eds.), *Sequential monte carlo methods in practice* (pp. 197–223). New York, NY: Springer New York.
- Neal, R. M. (2003). Slice sampling. *The Annals of Statistics*, 31(3), 705–741.
- Poullisse, H., Van Overschee, P., Briers, J., Moncur, C. E., Goh, K.-C., et al. (2006). Continuous well production flow monitoring and surveillance. In *Intelligent energy conference and exhibition*.
- Simon, D. (2010, August). Kalman filtering with state constraints: a survey of linear and nonlinear algorithms. *IET Control Theory Applications*, 4(8), 1303–1318. doi: 10.1049/iet-cta.2009.0032
- Sondergaard, T., & Lermusiaux, P. F. J. (2013). Data assimilation with gaussian mixture models using the dynamically orthogonal field equations. part i: Theory and scheme. *Monthly Weather Review*, 141(6), 1737–1760. doi: 10.1175/MWR-D-11-00295.1
- Sorenson, H. W., & Alspach, D. L. (1971, July). Recursive bayesian estimation using gaussian sums. *Automatica*, 7(4), 465–479. doi: 10.1016/0005-1098(71)90097-5
- Srkk, S. (2013). *Bayesian filtering and smoothing*. New York, NY, USA: Cambridge University Press.
- Tokdar, S. T., & Kass, R. E. (2010). Importance sampling: a review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(1), 54–60. doi: 10.1002/wics.56
- Vo, B.-n., & Ma, W.-k. (2006). The gaussian mixture probability hypothesis density filter. *IEEE Trans. SP*, 4091–4104.

BIOGRAPHIES



Ashutosh Tewari earned his BS/MS in Material Science & Metallurgical Engineering from Indian Institute of Technology, Bombay in 2003 and PhD in Engineering Science & Mechanics from Penn State University in 2009. He worked as a research scientist at United Technologies Research Center between 2009 and 2015, focusing in the areas of statistical modeling & machine learning with applications to aerospace systems. He joined ExxonMobil's Corporate Strategic Research in February 2015 as a senior researcher and has been primarily working towards furthering technologies pertaining to hydrocarbon exploration & production. His research interests cut across many disciplines of machine learning that include, among others, probabilis-

tic graphical models, Bayesian analysis, kernel methods and copula functions.



Stijn de Waele is research associate at ExxonMobil Corporate Strategic Research in Annandale, New Jersey, USA. He received his Masters degree in Applied Physics in 1998 from Delft University of Technology, and his PhD degree in 2003 from the same university. Before joining ExxonMobil in 2015, he worked as a senior scientist for Philips Research from 2003 to 2015. His research interest include algorithm development for probabilistic parameter estimation, signal processing, and time series analysis. He has applied these algorithms in a range of industrial applications including patient monitoring in the health-care domain, and hydrocarbon production surveillance and optimization.



Niranjana Subrahmanya received B.S. and M.S. degrees in Mechanical Engineering from the Indian Institute of Technology, Mumbai (2003), and a Ph.D. from Purdue University (2009). His dissertation focused on the development of machine learning algorithms for data-driven process modeling, monitoring and optimization. After graduation, he joined ExxonMobil's Corporate Strategic Research Lab as a Senior Engineer in the Data Analytics and Optimization section. There, he led various projects related to the development of novel data-driven methods in production surveillance & optimization, refinery process monitoring & modeling and deep learning applications for energy industry specific problems. He no longer works for ExxonMobil, nevertheless, the work presented in this paper was done during his time at the company.

APPENDIX

In the subsequent derivations we make use of the following two lemmas pertaining to the multivariate Gaussian distribution.

Lemma 1: If random variables $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ have Gaussian probability distributions given as

$$\mathbf{x} \sim \mathcal{N}(\mathbf{m}, P), \text{ and} \\ \mathbf{y} | \mathbf{x} \sim \mathcal{N}(W\mathbf{x} + \mathbf{u}, V),$$

then the joint distribution of \mathbf{x} , \mathbf{y} and the marginal distribution of \mathbf{y} are given as

$$\mathbf{x}, \mathbf{y} \sim \mathcal{N}\left(\begin{bmatrix} W\mathbf{m} + \mathbf{u} \\ \mathbf{m} \end{bmatrix}, \begin{bmatrix} P & PW^T \\ WP & WPW^T + V \end{bmatrix}\right), \\ \mathbf{y} \sim \mathcal{N}(W\mathbf{m} + \mathbf{u}, WPW^T + V),$$

respectively.

Lemma 2: If random variables $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ have a joint Gaussian probability distribution

$$\mathbf{x}, \mathbf{y} \sim \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \begin{bmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{bmatrix} \right)$$

then the distribution of \mathbf{x} conditional on \mathbf{y} is given as

$$\mathbf{x} | \mathbf{y} \sim \mathcal{N} \left(\boldsymbol{\mu}_x + S_{xy} S_{yy}^{-1} (\mathbf{y} - \boldsymbol{\mu}_y), S_{xx} - S_{xy} S_{yy}^{-1} S_{yx} \right).$$

We derive the update equations for the prediction and correction step for the SS-2 model from section 4.2. The derivation corresponds to a single Gaussian distribution. For a mixture of Gaussians, the update equations can be applied to each mixing components separately. Let the state estimate at time t (conditional on all the measurements up to time t) be given as

$$z[t], x[t] | y[1:t] \sim \mathcal{N} \left(\begin{bmatrix} \mu^z[t] \\ \mu^x[t] \end{bmatrix}, \begin{bmatrix} \Sigma^{zz}[t] & \Sigma^{zx}[t] \\ \Sigma^{xz}[t] & \Sigma^{xx}[t] \end{bmatrix} \right). \quad (29)$$

Prediction Step: The state evolves to the next time step based on the dynamics governed by equations

$$z[t+1] | z[t] \sim \mathcal{N}(Az[t], S) \quad \text{and} \quad (30)$$

$$x[t+1] | z[t+1] \sim \mathcal{N}(z[t+1], F[t]). \quad (31)$$

Given the linear dynamics, the prediction step yields a Gaussian distribution of the form

$$z[t+1], x[t+1] | y[1:t] \sim \mathcal{N} \left(\begin{bmatrix} \tilde{\mu}^z[t+1] \\ \tilde{\mu}^x[t+1] \end{bmatrix}, \begin{bmatrix} \tilde{\Sigma}^{zz}[t+1] & \tilde{\Sigma}^{zx}[t+1] \\ \tilde{\Sigma}^{xz}[t+1] & \tilde{\Sigma}^{xx}[t+1] \end{bmatrix} \right). \quad (32)$$

In the following steps we derive the expressions of unknown parameters shown with a circumflex $\tilde{\cdot}$. We invoke lemma 1 first on equations (30), (31) and then again on its result and equation (29), resulting in the following joint distribution.

$$z[t], x[t], z[t+1], x[t+1] | y[1:t] \sim \mathcal{N}(\mathbf{b}, C), \quad (33)$$

where

$$\mathbf{b} = \begin{bmatrix} \mu^z[t] \\ \mu^x[t] \\ A\mu^z[t] \\ A\mu^x[t] \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} \Sigma^{zz}[t] & \Sigma^{zx}[t] & \Sigma^{zz}[t]A^T & \Sigma^{zz}[t]A^T \\ \Sigma^{xz}[t] & \Sigma^{xx}[t] & \Sigma^{xz}[t]A^T & \Sigma^{xz}[t]A^T \\ A\Sigma^{zz}[t] & A\Sigma^{zx}[t] & A\Sigma^{zz}[t]A^T + S & A\Sigma^{zz}[t]A^T + S \\ A\Sigma^{xz}[t] & A\Sigma^{xx}[t] & A\Sigma^{xz}[t]A^T + S & A\Sigma^{xz}[t]A^T + S + F[t] \end{bmatrix}.$$

Marginalizing out $z[t]$ and $x[t]$ from the equation (33), we obtain

$$z[t+1], x[t+1] | y[1:t] \sim \mathcal{N} \left(\begin{bmatrix} A\mu^z[t] \\ A\mu^x[t] \end{bmatrix}, \begin{bmatrix} A\Sigma^{zz}[t]A^T + S & A\Sigma^{zz}[t]A^T + S \\ A\Sigma^{xz}[t]A^T + S & A\Sigma^{xz}[t]A^T + S + F[t] \end{bmatrix} \right) \quad (34)$$

Comparing equation (34) with (32) gives us the update equation of the prediction step.

Correction Step: The correction step assimilates a new measurement $y[t+1]$ to yield state estimate at time $t+1$ i.e.

$$z[t+1], x[t+1] | y[1:t+1] \sim \mathcal{N} \left(\begin{bmatrix} \mu^z[t+1] \\ \mu^x[t+1] \end{bmatrix}, \begin{bmatrix} \Sigma^{zz}[t+1] & \Sigma^{zx}[t+1] \\ \Sigma^{xz}[t+1] & \Sigma^{xx}[t+1] \end{bmatrix} \right). \quad (35)$$

We obtain the expressions of the unknown parameters at $t+1$ as follows. First note that the observation $y[t+1]$ depends only on $x[t+1]$ (refer figure 2b). Therefore,

$$y[t+1] | z[t+1], x[t+1] \sim \mathcal{N} \left(\begin{bmatrix} 0 & H \end{bmatrix} \begin{bmatrix} z[t+1] \\ x[t+1] \end{bmatrix}, R[t+1] \right). \quad (36)$$

Invoking lemma 1 on equations (32) and (36), we get

$$z[t+1], x[t+1], y[t+1] | y[1:t] \sim \mathcal{N} \left(\begin{bmatrix} \tilde{u}^z[t+1] \\ \tilde{u}^x[t+1] \\ H\tilde{u}^x[t+1] \end{bmatrix}, \begin{bmatrix} \tilde{\Sigma}^{zz}[t+1] & \tilde{\Sigma}^{zx}[t+1] & \tilde{\Sigma}^{zx}[t+1]H^T \\ \tilde{\Sigma}^{xz}[t+1] & \tilde{\Sigma}^{xx}[t+1] & \tilde{\Sigma}^{xx}[t+1]H^T \\ H\tilde{\Sigma}^{xz}[t+1] & H\tilde{\Sigma}^{xx}[t+1] & H\tilde{\Sigma}^{xx}[t+1]H^T + R[t+1] \end{bmatrix} \right) \quad (37)$$

Finally, invoking lemma 2 on equation (37) followed by some algebraic manipulations, we obtain the update equations for the correction step.

$$\begin{aligned} \mu^z[t+1] &= \tilde{\mu}^z[t+1] + K^z[t+1] \times r[t+1] \\ \mu^x[t+1] &= \tilde{\mu}^x[t+1] + K^x[t+1] \times r[t+1] \\ \Sigma^{zz}[t+1] &= \tilde{\Sigma}^{zz}[t+1] - K^z[t+1]H\tilde{\Sigma}^{zz}[t+1] \\ \Sigma^{zx}[t+1] &= \tilde{\Sigma}^{zx}[t+1] - K^z[t+1]H\tilde{\Sigma}^{zx}[t+1] \\ \Sigma^{xx}[t+1] &= \tilde{\Sigma}^{xx}[t+1] - K^x[t+1]H\tilde{\Sigma}^{xx}[t+1] \end{aligned}$$

where,

$$\begin{aligned} K^z[t+1] &= \tilde{\Sigma}^{zx}[t+1]H^T(H\tilde{\Sigma}^{xx}[t+1]H^T + R[t+1])^{-1} \\ K^x[t+1] &= \tilde{\Sigma}^{xx}[t+1]H^T(H\tilde{\Sigma}^{xx}[t+1]H^T + R[t+1])^{-1} \\ r[t+1] &= y[t+1] - H\tilde{\mu}^x[t+1]. \end{aligned}$$